

Performance Evaluation of Multi-Source Methane Emission Quantification Models Using Fixed-Point Continuous Monitoring Systems

David Ball¹, Umair Ismail¹, Nathan Eichenlaub¹, Noah Metzger¹, and Ali Lashgari¹

¹Project Canary, Denver, CO

Correspondence: Ali Lashgari (ali.lashgari@projectcanary.com)

Abstract. Quantifying methane emissions from oil and gas facilities is crucial for emissions management and accurate facility-level GHG inventory development. This paper evaluates the performance of several multi-source methane emission quantification models using the data collected by fixed-point continuous monitoring systems as part of a controlled release experiment. Two dispersion modeling approaches (Gaussian plume, Gaussian puff) and two inversion frameworks (least-squares optimization and Markov-Chain Monte-Carlo) are applied to the measurement data. In addition, a subset of experiments are selected to showcase the application of computational fluid dynamic (CFD) informed calculations for direct solution of the advection-diffusion equation. This solution utilizes a three-dimensional wind field informed by solving the momentum equation with the appropriate external forcing to match on-site wind measurements. Results show that the Puff model, driven by high-frequency wind data, significantly improves localization and reduces bias and error variance compared to the Plume model. The Markov-Chain Monte-Carlo (MCMC) based inversion framework further enhances accuracy over least-squares fitting, with the Puff MCMC approach showing the best performance. The study highlights the importance of long-term integration for accurate total mass emission estimates and the detection of anomalous emission patterns. The findings of this study can help improve emissions management strategies, aid in facility-level emissions risk assessment, and enhance the accuracy of greenhouse gas inventories.

1 Introduction

Quantification of methane emissions from oil and gas facilities is crucial for facility-level emissions management and accurate greenhouse gas (GHG) inventory development (Sharafutdinov 2024). Understanding the contribution of different emission sources to overall site emissions allows operators to improve asset risk management and prioritize mitigation efforts. Currently, the US Environmental Protection Agency (EPA) and many other entities use bottom-up GHG emissions inventories, which mainly rely on activity rate and emission factors (Allen, Zimmerle, and Dabbar 2024).

Several studies have highlighted major shortcomings of bottom-up inventories (Riddick and Mauzerall 2023; Riddick, Mbua, Anand, et al. 2024; Riddick, Mbua, Santos, et al. 2024). In 2021, 70% of methane emissions came from facilities emitting less than 100 kg/hour, with 30%, 50%, and approximately 80% coming from facilities emitting less than 10 kg/hour, 25 kg/hr, and 200 kg/hour, respectively (Williams et al. 2025), demonstrating that low-emitting facilities, particularly those below the detection limit of most point-source remote sensing platforms, contribute significantly to total oil and gas methane emissions. Therefore, it is essential to employ approaches that accurately account for the substantial impact of these small sources.

The substantial variability in methane emission intensity across geographic regions, facility types, and operators necessitates a comprehensive characterization of emission events. Ideally, by applying methods that enable emission event detection, localization, and quantification (DLQ), the distributions of rates, durations, and frequencies can be inferred to provide a deeper understanding of site-specific emission patterns, and may bring to light any underlying issues, and aid in root cause analysis.

Direct measurement is essential for a more thorough characterization of emission events. A wide variety of methods can be used to collect various data related to methane emissions. Fixed-point continuous monitoring systems (CMS) have been widely deployed to monitor emissions from oil and gas production facilities for several years. These systems were initially deployed as a means for emissions anomaly detection, and as such, were commonly referred to as “smoke alarms” (IJzermans et al. 2024; Gosse 2023). They were intended to provide timely alerting of elevated emissions by processing raw concentration signals into alerts via a variety of anomalous event detection algorithms, ranging from static concentration thresholds to more sophisticated approaches employing signal processing and/or machine learning methods (Gosse 2023). While anomaly detection is a useful function of CMS, providing additional information regarding the source locations and magnitude of emission events would enhance the actionable insights provided by these systems. If these additional features can be developed, validated, and proven to be reliable, some of the key benefits that CMS could offer include (i) providing a comprehensive picture of site-level emissions for the entire period of deployment, (ii) rapid detection of emissions ranging from relatively low rates to super emitting events, (iii) capturing both short-duration/intermittent and continuous events, (iv) accurate time-bounding of intermittent emission events, (v) providing equipment-specific emissions insights that can aid in root cause analysis and provide strategically relevant information for targeted mitigation efforts, and (vi) complimenting other measurement methods using a continuous stream of site-specific data on emission estimates, direct concentration measurements, and meteorology. To expand the application of CMS, it is crucial to improve the existing understanding of its quantification performance, including its accuracy and uncertainty. This will further demonstrate the value of these systems as measurement tools that can not only detect and time-bound anomalous emission events but also provide insight into the total emissions originating from a given facility over time and the contribution of different sources to the facility-level emissions.

Several studies have independently evaluated the efficacy of CMS in quantifying emissions, suggesting promising advancements in recent years (Bell et al. 2023; Ilonze et al. 2024; Cheptonui et al. 2025). Although technologies have demonstrated

marked improvements in quantification accuracy, the algorithms behind these results are proprietary, making it difficult to compare the results from different technologies of the same sensing modality. Proprietary DLQ algorithms, while understandable in a competitive business environment, are often seen as black boxes, requiring different, more involved methods to evaluate their uncertainties and ensure that their performance is fully understood across various environments. In addition, evaluating methane emission solutions as a single package (combining measurement, data collection strategy, and data processing) yields an inseparable uncertainty value that reflects the combined uncertainties from all three processes. This makes it impossible to differentiate the uncertainty arising from the measurement (i.e., hardware), data collection (i.e., deployment strategy and sensor configuration), and data processing (the application of DLQ algorithms).

While extensive research exists addressing pollutant transport (including long-range dispersion, localization, and quantification at much larger scales for other applications), (Chen, Modi, et al. 2022; Schade and Gregg 2022; Karion et al. 2019; Peischl et al. 2016) relatively little literature focuses on methane emission quantification using near-source point sensor measurements, i.e., measurements within the boundaries of upstream oil and gas facilities. Most of the literature employs existing dispersion modeling tools and methods, such as AERMOD (Cimorelli et al. 2005) and CALPUFF (Allwine, Dabberdt, and Simmons 1998), or quantify single-source emissions (Sharan et al. 2009; Zhang et al. 2019; Kumar et al. 2022; Daniels, Jia, and Hammerling 2024a; Chen, Schissel, et al. 2023; Chen, Kimura, and Allen 2024).

In a study published in 2019, a steady-state Gaussian plume model was employed to estimate emission rates from point sources (Zhang et al. 2019). In this method, the plume spread parameters are simplified for short distances, and a heuristic dispersion modifier is introduced to account for non-ideal measurement conditions. This quantification method is intended to be used in conjunction with source localization techniques. A more recent study (Daniels, Jia, and Hammerling 2024b) provides a framework (and open-source implementation) for single source detection, localization, and quantification, with promising results for cases where only a single emission source is present. However, the fact that the algorithm only identifies a single source per emission event renders the algorithm inapplicable to general use cases. Comprehensive reviews of advanced detection and quantification methods can be found elsewhere (Hollenbeck, Zulevic, and Chen 2021; Yang et al. 2023).

Dispersion models and inversion frameworks are essential tools for translating ambient methane concentration measurements (e.g., ppm readings) into source flux rates (mass of pollutant emitted per unit of time). Forward-running dispersion models simulate how methane released from a source disperses in the atmosphere based on a number of meteorological variables including wind speed, direction, and atmospheric stability. On the other hand, inversion models use mathematical techniques to estimate the source flux(es) that would have resulted in the observed ambient concentrations at the sensor location. This often involves solving an optimization problem, where the inverse model adjusts the source strengths and locations until the simulated concentrations best match the observed amounts.

90 This paper aims to address the critical need for developing a more comprehensive understanding of the performance and robustness of various multi-source methane quantification methods by evaluating the performance of several established atmospheric dispersion modeling and inversion frameworks within a controlled, multi-leak experimental setting with synchronous emission sources and constant rates. This study leverages data collected from a fixed-point CMS deployed at a simulated oil and gas site with multiple simultaneous methane releases of varying magnitudes and locations. The accuracy and reliability of these models are evaluated with respect to several key metrics related to localization accuracy and total-facility (i.e., source-integrated) quantification accuracy.

Three key questions will be addressed in this study: (i) under an optimum sensor density and placement, how effectively can a CMS pinpoint emissions to the correct equipment group? (ii) what is the accuracy of the total site-integrated emissions estimates for such CMS network? And, (iii) How well can an advanced CFD-based forward model, coupled with various inversion frameworks perform in predicting emission rates compared to traditional plume and puff models?

Accurate emissions quantification using CMS can enhance the reliability and robustness of GHG emissions inventory development. Traditional inventory methods often rely on activity data and generic emission factors, which fail to capture the dynamic nature of emissions from individual sources or facilities. By providing continuous, real-time measurements source-specific emissions, CMS offers a direct and empirically driven approach to quantify actual emissions. High temporal resolution of the CMS measurement allows for the identification and characterization of gas releases, including the duration and frequency of emission events. In addition, accurate quantification offers a more in-depth understanding of the magnitude of fugitive emissions, intermittent events, and variations in operational performance that are often missed by periodic or estimation-based methods. Integrating CMS data into GHG inventories leads to a more comprehensive understanding of emission sources, enables the tracking of emission reduction efforts with greater confidence, and supports the development of more granular and verifiable inventories, informing climate policies, and tracking progress towards decarbonization goals.

In this study selected quantification algorithms are evaluated using the data from controlled release experiments featuring constant-rate emission events with known start and end times. However, it's crucial to recognize that these controlled release scenarios are highly idealized, as they involve constant release rates and simultaneous emissions from all active sources. This idealization may impact the practical applicability of these algorithms in more complex, real-world conditions. A more in-depth evaluation of the performance of fixed-point CMS in complex emission environments is provided in a separate study (Ball, Eichenlaub, and Lashgari 2025).

120

This work offers a novel contribution by evaluating several multi-source methane quantification techniques using multi-leak, controlled-release data. Unlike previous studies that often rely on simulations, this study leverages a fixed-point CMS to cap-

ture the complexities of overlapping plumes from simultaneous releases. This approach provides a unique opportunity to assess model accuracy and reliability under semi-realistic field conditions representative of relatively simple upstream oil and gas facilities. By emphasizing the strengths of each technique, this study offers crucial insights for improving methane emission quantification strategies, including guidance for selecting appropriate dispersion models and inversion tools, ultimately informing the development of more effective methane mitigation in the oil and gas industry.

2 Data

The data presented in this work are all collected with the Canary X integrated device, which includes a tuneable laser diode spectroscopy (TDLAS) methane sensor and can additionally be mounted with an ultrasonic anemometer (at least one of which is required for the sensor network to perform quantification). The Canary X integrated monitoring devices use TDLAS technology coupled with other necessary components to serve as an IoT-enabled stand-alone monitoring device with high sensitivity to perform high-fidelity measurement of methane concentrations, crucial for accurately quantifying emissions in the field. The methane measurement sensors have 0.4 ppm sensitivity, $\pm 2\%$ accuracy, and a precision of ≤ 0.125 ppm with 60-second averaging. This integrated measurement device is capable of 1 Hz sampling, although the measured quantities are often aggregated to 1-minute averages for the purposes of analysis and applying quantification algorithms. Throughout this work, a note will be made any time 1Hz data is used for a specific piece of an algorithmic workflow: if not otherwise stated, quantities are assumed to be minute-averaged aggregates.

140

Methane concentration measurements are complemented with meteorological data collected on-site using RM Young 2D ultrasonic anemometers. Manufacturer specifications indicate the anemometers have an accuracy of $\pm 2\% \pm 0.3$ m/s for wind speed and $\pm 2^\circ$ for wind direction, with resolutions of 0.01 m/s and 0.1° , respectively. Methane concentrations and meteorological data are continually published to a cloud server using cellular networks.

145

The data was collected over 82 days (February to April of 2024) as part of an independent, single-blind controlled release study performed at the Colorado State University (CSU) Methane Emission Technology Evaluation Center (METEC) facility in Fort Collins, Colorado. METEC is a research facility hosted by the CSU Energy Institute that facilitates the study of methane leaks from oil and gas infrastructure as part of the Advancing Development of Emissions Detection (ADED) project, funded by the Department of Energy's National Energy Technology Laboratory (NETL). Ten Canary X integrated devices were installed within the METEC site perimeter to measure ambient methane concentrations. All of the Canary X devices used for this study were equipped with an anemometer. More details on the data collected as part of the 2024 CSU METEC controlled release study can be found elsewhere (Cheptonui et al. 2025).

155 Controlled release “experiments” at the METEC facility include between 1 and 5 releases that are synchronously turned on and off at the start and end of each unique experiment. The rate of each source is held approximately constant during an individual experiment. Experiment durations ranged between 30 minutes and 8 hours while individual source rates ranged from 0.081 to 6.75 kg/hr. Figure 1 overviews the number of active releases per experiment and source release rates. The experiments are designed such that only one release point is active per equipment group at the METEC facility. Each equipment group is composed of numerous “equipment units” (i.e., individual tanks, wellheads, or separators) and each equipment unit may have multiple potential release points on it. In other words, each equipment group has numerous *potential* release points, but only one is ever active at a time for a given experiment. In this study, we focus on the ability of the system to correctly detect, localize, and quantify to the equipment group level. As such, the centroid of each equipment group is computed and these 5 coordinate pairs (corresponding to the 5 equipment groups at the facility) are used as the potential source locations as an input to the localization and quantification (LQ) algorithms. The heights of the release locations are unknown but assumed to be 2 meters tall for all sources except for the group of tanks in the middle of the facility, for which a height of 4.5 meters is assumed and used as input to the LQ algorithms.

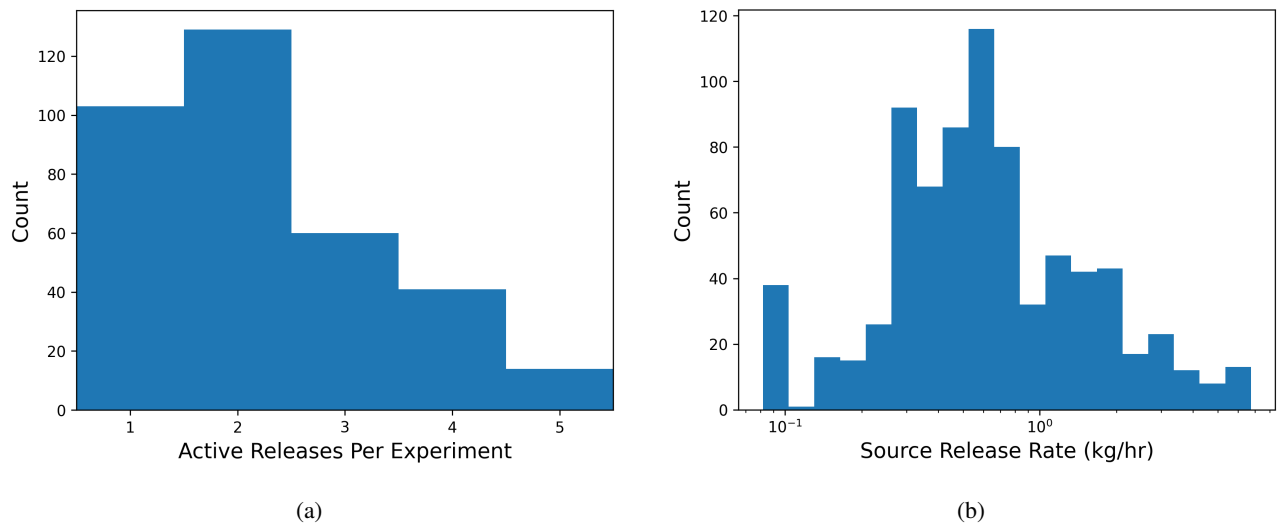


Figure 1. (a) Histogram of active releases per experiment, and (b) Histogram of source release rates (kg/hr).

Figure 2 offers a visual illustration of the layout of the controlled release facility (left), including bounding boxes around each of the 5 equipment groups (left) and sensor locations (x’s). It also shows measurement data from a randomly selected controlled release experiment, including concentration measurements from individual sensors (top right) and the QU and V components of the anemometer measurements (with solid and dotted lines, respectively, bottom right). The colors of the curves in the right panels here correspond to the colored x’s in the left panels. This figure encapsulates all of the data necessary to run quantifica-

tion algorithms (sensor locations, source locations, concentration timeseries data, and wind timeseries data).

175

For the purposes of this study, the known release start and end times are used to segment out the relevant measurement data for each experiment; detecting and time-bounding each unique experiment is outside the scope of this work. Therefore, by applying several selected quantification algorithms to data from individual experiments, this study evaluates how well the quantification algorithms perform on constant-rate emissions events with known start and end times. It's important to note that these controlled release scenarios are highly idealized, featuring constant release rates and simultaneous emissions from all active sources, which may limit the practical application of these algorithms. However, it is still useful to evaluate the efficacy of various quantification algorithms under idealized setups to lay the groundwork for future development and studies in which these underlying assumptions and simplifications will be relaxed to more accurately reflect real-world conditions.

180

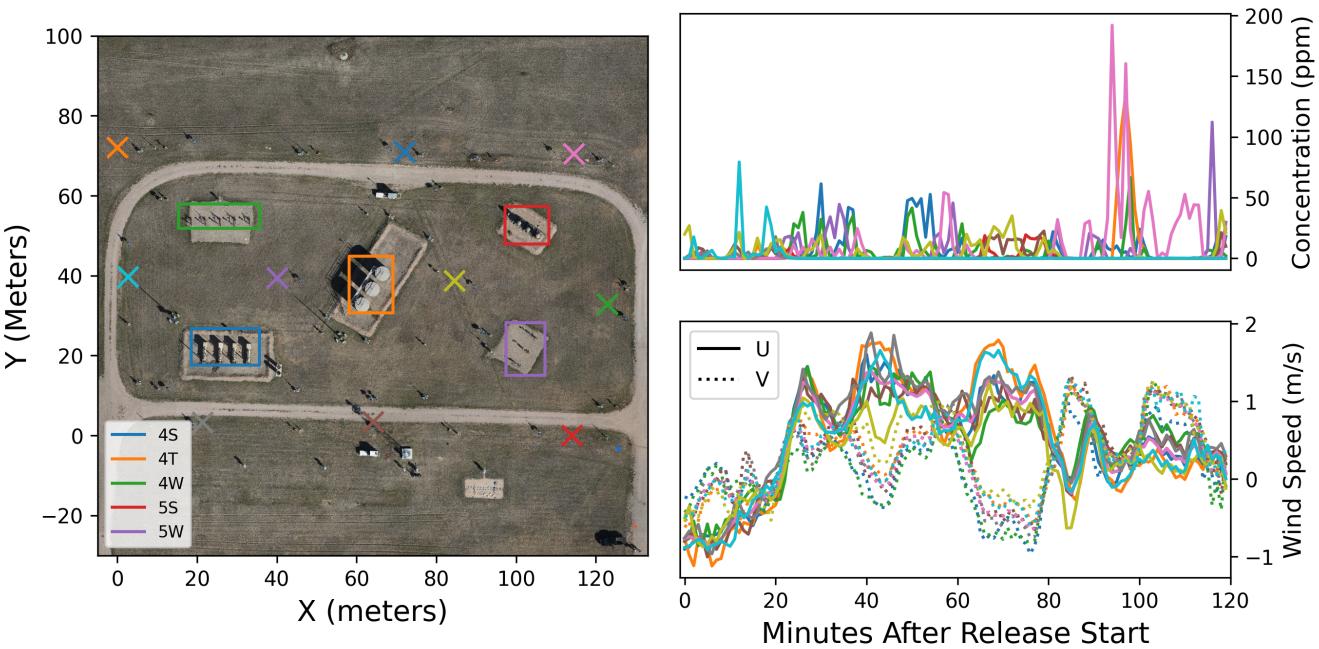


Figure 2. Facility layout including source and sensor positioning (left) and example of data taken during a single experiment (top right: concentration measurements, bottom right: wind measurements). The colored boxes in the left panel show the spatial extent of each of the 5 equipment groups. Each color in the concentration and wind speed curves (right panels) corresponds to a colored x representing sensor location in facility layout (left) panel. Sensor locations are shown in a zero-referenced easting/northing projected coordinate system as opposed to latitudes and longitudes so that the relative spatial distances are more visually interpretable.

This section along with Appendix B detail several dispersion models and inversion frameworks as options to quantify methane emissions based on ambient concentration measurements using fixed-point CMS.

3.1 Dispersion Models

Two distinct methods, the Gaussian plume and Gaussian puff models, for predicting concentrations at receptor locations given a set of sources and associated rates are detailed in this section. The theoretical and fundamental aspects as well as the underlying assumptions of each method are described and in-depth discussions of various aspects of implementation to establish a robust foundation for their use is offered. In addition to these two most commonly used forward dispersion modeling methods, a more novel approach, a CFD-informed calculation is included in Appendix B, that directly solves the advection-diffusion equation with a three-dimensional wind field informed by solving the momentum equation with the appropriate external forcing to match on-site wind measurements. All of these three methods rely on (or are derived from) the advection-diffusion equation, also commonly referred to as the scalar transport equation. For an incompressible flow with homogeneous and isotropic diffusion, this equation can be written as:

$$\frac{\partial C}{\partial t} + \mathbf{u}(\mathbf{x}, t) \cdot \nabla C - D \nabla^2 C = Q(\mathbf{x}, t) \quad (1)$$

where C represents the concentration, $\mathbf{u}$ is the wind vector (which may vary as a function of both space and time), D is the diffusion coefficient, and Q represents emission sources (which can also vary as a function of both space and time). It is important to note that unless the treatment of the wind field, $\mathbf{u}$, explicitly accounts for chemical buoyancy, the resulting solution of the advection-diffusion equation will not capture this effect.

For the purposes of this study, we will focus on cases with constant emission rates, i.e. $Q(\mathbf{x}, t) = Q(\mathbf{x})$. Inversion frameworks to infer time-varying source rates will be addressed in future work. Furthermore, we will focus on emissions from discrete point sources, where the sizes of the orifices of the controlled release systems are of the order $\sim$ centimeters, much smaller than source-receptor distances of the order ~ 10 meters. As such, the constant-rate source function $Q(\mathbf{x})$, can be expressed as the summation of discrete point sources of varying rates:

$$Q(\mathbf{x}) = \sum_{i=1}^n Q_i \delta(x - x_i) \delta(y - y_i) \delta(z - z_i). \quad (2)$$

Here, Q_i represents the emission rate of the i th source, (x_i, y_i, z_i) represent its three-dimensional coordinates, and δ is the Dirac delta function. For the rest of this work, x and y will be reserved for describing horizontal coordinates while z will refer to height.

215 Two important features to note about this equation are its scale invariance and linearity. All the terms on the left-hand side of Equation 1 are linear in C and all of the operators (time derivatives, gradients, dot products, and scalar multiplication) can be distributed across addition. As a result, the solution to the advection-diffusion equation with a set of constant-rate sources can be expressed simply as the sum of the solutions to the partial differential equations associated with each individual source. In other words, the solution to:

$$220 \quad \frac{\partial C}{\partial t} + \mathbf{u}(\mathbf{x}, t) \cdot \nabla C - D \nabla^2 C = \sum_{i=1}^n Q_i \delta(x - x_i) \delta(y - y_i) \delta(z - z_i) \quad (3)$$

can be expressed as $C = \sum_{i=1}^n C_i$ where C_i is the solution to the advection-diffusion equation applied to the i th point source:

$$\frac{\partial C_i}{\partial t} + \mathbf{u}(\mathbf{x}, t) \cdot \nabla C_i - D \nabla^2 C_i = Q_i \delta(x - x_i) \delta(y - y_i) \delta(z - z_i) \quad (4)$$

Finally, note that all of the terms on the left-hand side of Equation 4 are linear with respect to C and the operators commute with scalar multiplication: the result of this is that C and Q_i are directly proportional to one another. Therefore, the solution for
 225 an arbitrary emission rate can be obtained by solving this equation once for a unit impulse and then normalizing the concentrations accordingly. This can also be thought of as solving the equation for C/Q_i and then multiplying it back in the desired rate.

Due to the linear scaling of concentrations with rate and additive nature of discrete point sources, predicted concentrations can
 230 be expressed via a simple linear system that sums up the concentration from every source via $\mathbf{b} = \mathbf{S}\mathbf{Q}$. Here, $\mathbf{b}$ represents a vector of simulated concentrations, $\mathbf{Q}$ is the vector of source rates, and $\mathbf{S}$ is the “sensitivity matrix” that describes the transport of gas from every source to every virtual measurement point. Each row of this matrix corresponds to a simulated methane measurement at a given time and location under measured meteorological conditions. Each column of the sensitivity matrix corresponds to a source that is being modeled. The following subsections describe how this sensitivity matrix is calculated for
 235 three different dispersion models for later use in the inversion process.

The following subsections provide brief overviews of the theory underpinning the dispersion models, followed by more specific implementation details. Note that there are myriad small choices (e.g., stability class calculations, dispersion parametrization)

that must be made in the data processing and algorithmic workflow when it comes to running these dispersion models. It is outside the scope of this study to enumerate and present results from every combination of valid choices. Instead, we will provide clear justifications for the specific choices made in this study and demonstrate the efficacy of the models under these specific implementations. It should be noted that the impact of most of these higher-order decisions on the results is minimal, as they are often different approaches of approximating the same underlying phenomena. For example, there are several commonly-used functional forms and associated coefficients to describe how the dispersion of a gas plume scales with distance. While these empirical formulae may look very different (e.g., some utilize power laws while others employ logarithms), they are generally inferred by fitting these functional forms to the same underlying data, and result in similar general characteristics despite the sometimes dramatically different functional forms.

3.1.1 Gaussian Plume

The commonly used Gaussian Plume model (GPM) provides a closed-form solution to the steady-state advection-diffusion equation for a single point source emitting at rate Q from height H . As a steady-state model, GPM assumes constant wind speed, wind direction, and source rate over time. Pollution concentration, C at a given location (x, y, z) is calculated as:

$$C(x, y, z) = \frac{Q}{2\pi\bar{u}\sigma_y\sigma_z} \exp \left[-\frac{y^2}{2\sigma_y^2} - \frac{(z-H)^2}{2\sigma_z^2} - \frac{(z+H)^2}{2\sigma_z^2} \right] \quad (5)$$

In this coordinate system, x is defined as the downwind distance (aligned with the average wind velocity, the magnitude of which is denoted by $\bar{u}$), y represents the crosswind distance (perpendicular to the wind direction), and z is the receptor height above the ground. The dispersion coefficients σ_y and σ_z represent the horizontal and vertical spread of the plume, respectively. The rate at which plume spread increases with downwind distance is influenced by the stability class, a metric that represents the degree of turbulence in the atmosphere. The approximation of dispersion coefficients can be achieved using various established methods (Carruthers, Weng, et al. 2009). The specific implementation details used for this study are presented later in this Section. The concentration levels modeled by the GPM represent temporal averages that capture the statistics of the turbulent dispersion by small variations in the wind field. As such, an appropriate averaging time must be chosen that is sufficiently long for small-scale turbulence effects to average out, but not so long that coherent changes in the wind direction occur.

It is important to note how different turbulent wavenumbers (k) affect a plume at different characteristic scales (L). For reference, k can be thought of simply as the inverse spatial scale of a turbulent eddy, $k = 2\pi/L_{eddy}$, where L_{eddy} represents the characteristic length scale associated with a particular turbulent eddy. Turbulence scales where $k > 1/L$ (i.e., when turbulent fluctuations are smaller than the length scale of the plume) cause plume dispersion. In other words, small-scale eddies disperse the plume randomly in a way that is captured by σ_y and σ_z . If a turbulent eddy has wavenumber comparable to (or less) than $1/L$, however, this results in meandering motions (or coherent changes in wind direction) for extended periods of time that do

270 not result in statistical dispersion. In this case, the Gaussian Plume will fail to capture the effects of these variations if they are averaged over.

Many applications of the Gaussian Plume are tuned to model dispersion on long length scales (kilometers or greater), and hence require long (> 30 minute) averaging periods. However, when GPM is applied to fenceline monitors placed around
275 facilities with scales of around 100 meters, much shorter averaging periods must be used. This ensures that turbulent eddies on scales comparable to or larger than the source-sensor distance are captured temporally as coherent changes in wind direction, rather than being averaged over. As such, shorter averaging times (1-5 minutes) are more appropriate for near-source dispersion modeling applications.

280 When employing Gaussian Plume models, any violation of the assumptions that go into the derivation of the GPM results in unreliable concentration predictions from this dispersion model. These effects can be especially pernicious at low wind speeds. The underlying assumption of steady-state wind is often violated during low-wind conditions because the ratio between the reference velocity scale of turbulence u_{ref} , which is typically taken as the root-mean-square of the horizontal wind fluctuations u'_{rms} , scales with the mean wind speed used in Equation 5, i.e. $\frac{u_{ref}}{\bar{u}} \sim 1$. Practically, this implies that the underlying assumption for a Gaussian concentration distribution is no longer valid. In addition, frequent and significant changes in wind direction
285 are more common at low wind speeds, making the application of GPM less optimal, as this model requires that the mean wind direction should align with the x (downwind) axis. Also note that since $\bar{u}$ is in the denominator of Equation 5, in low wind speed conditions, concentration predictions will become nonphysically high. Different corrections exist to the GPM under low wind conditions. The approach by (Carruthers, Holroyd, et al. 1994) reconstructs concentration under low wind conditions by
290 using a weighted average between a Gaussian-shaped plume state and a random-walk state in which the release spreads in all horizontal radial directions.

Many of the standard methods for computing the dispersion coefficients σ_y and σ_z rely first on an approximation of the Pasquill atmospheric stability class (ASC) (Mohan and Siddiqui 1998; Venkatram 1996). Several approaches exist for ASC approxi-
295 mation, including using near-surface vertical temperature gradient (Randerson 1984) or inferring them from estimates of solar irradiation (based on cloud cover and time of day), and combining this information with wind speed to infer the stability class via a look-up table (Pasquill 1961). The current study uses the statistics of the fluctuating horizontal wind field to estimate the ASC at minute t by computing the circular standard deviation of horizontal wind direction (σ_θ) over the trailing 15 minutes of wind data and mapping these values to a stability class via Tables 6-8a and 6-8b found in USEPA 2000. This method is chosen
300 over the other approaches because it utilizes a direct measurement of the degree of turbulence in the atmosphere via σ_θ rather than exclusively indirect indicators (solar irradiation, wind speed, cloud cover).

The majority of the methods for calculating dispersion parameters σ_y and σ_z as a function of downwind distance and stability class provide comparable dispersion profiles that are approximated via different empirically-estimated functional forms and associated coefficients (Randerson 1984; R. P. Hosker 1974). This study employs the “Martin Method” (first proposed in Martin 1976) which utilizes power law functions to describe both the horizontal and vertical spread of the plume via:

$$\begin{aligned}\sigma_y &= ax^b \\ \sigma_z &= cx^d + f\end{aligned}\tag{6}$$

where the parameters a, b, c, d and f depend on the ASC and can be found in a look-up table in Martin 1976. Using these functional forms for the dispersion parameters, Equation 5 is applied to every minute of measurement data for every source-sensor combination for the time period associated with a given controlled release experiment. This results in generating the previously-described sensitivity matrix $\mathbf{S}$, the elements of which represent the predicted concentration via Equation 5 (for a unit rate). In this matrix, columns correspond to individual sources and rows correspond to the location and height of a specific measurement device at a specific time. Due to the failure of the GPM to accurately capture pollutant transport under low wind speeds, concentrations from time periods with wind speeds of less than 0.5 meters per second are excised from the sensitivity matrix.

Since GPM utilizes a closed-form algebraic solution (no time-marching is needed), it is computationally fast. Several hours of measurement data can be processed via GPM and corresponding concentration predictions can be made for every source for every sensor location in milliseconds when the calculations are vectorized properly.

3.1.2 Gaussian Puff

The Gaussian Puff model is a Lagrangian approach to approximating the solution to the advection-diffusion equation that makes fewer assumptions than the Gaussian Plume. More specifically, this method can capture the relevant physical effects embedded in spatially varying wind fields (i.e., it does not assume homogeneous wind fields), can handle time-varying emission rates (does not assume steady-state emission rate), and also more properly account for low wind speeds and unstable conditions when the wind vector rapidly changes (does not assume steady-state wind fields). Because of these features, Gaussian Puff is more broadly applicable than GPM and has fewer potential failure modes (e.g., low wind speeds and improper averaging times). The Gaussian Puff method approximates advection and diffusion by emitting a series of “puffs” that are advected with the wind field. The position of each puff, $\mathbf{x}$, as a function of time is simply:

$$330 \quad \mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \mathbf{u}(\mathbf{x}, t') dt' \quad (7)$$

where $\mathbf{x}_0$ represents the puff's initial location (given by the emission point from which the puff originated). Note that the wind vector ($\mathbf{u}$) used to advance the position of the puff can be a function of both position and time. When high-frequency spatial information about the wind field is available (e.g., via multiple anemometers), these effects can be included in the evolution of the puff positions to obtain a more accurate puff trajectory. In this method, puffs are emitted at some fixed interval Δt from
335 each point source, with each puff containing total mass $M = Q_i \Delta t$, where Q_i represents the emission rate of source i from which the puff was emitted. This mass is distributed across a three-dimensional Gaussian profile such that the concentration at a given point can be expressed via:

$$C(x, y, z, t) = \frac{Q \Delta t}{(2\pi)^{3/2} \sigma_x \sigma_y \sigma_z} \exp \left[-\frac{y^2}{2\sigma_y^2} - \frac{(z-H)^2}{2\sigma_z^2} - \frac{(z+H)^2}{2\sigma_z^2} - \frac{x^2}{2\sigma_x^2} \right]. \quad (8)$$

Note that each coordinate (x, y, z) and associated dispersion coefficients (σ 's) are functions of time (the explicit time dependence of these variables in Equation 8 is omitted for the sake of readability). In the implementation of the Gaussian Puff
340 algorithm used in this study, higher frequency temporal wind measurements are used to advance the positions of the puffs via Equation 7. More specifically, 1Hz sampled wind measurement are interpolated to the position of each individual puff using a distance-weighted average with inverse square weights to approximate the wind field at a specific location at a specific time. In principle, this should result in more physically accurate puff trajectories that account for both high-frequency spatial and
345 temporal variations. In other words, $\mathbf{u}(\mathbf{x}, t)$ is computed via:

$$\mathbf{u}(\mathbf{x}, t) = \frac{\sum_i \mathbf{u}(\mathbf{x}_i, t) / d_i^2}{\sum_i 1 / d_i^2} \quad (9)$$

where $\mathbf{u}(\mathbf{x}_i, t)$ represents an anemometer measurement at time t from location $\mathbf{x}_i$ and d_i is the Euclidean distance between the puff location $\mathbf{x}$ and the location associated with a given anemometer measurement, $\mathbf{x}_i$. Note that in this study, each Canary X device was equipped with an anemometer, so the distance-weighted average of 10 anemometers is used to advance the position
350 of the puffs. Wind measurements from 10 devices is highly redundant at such a small and simple facility: as shown in the bottom right hand corner of Figure 2, the horizontal and wind directions measured by all the devices on the facility are highly correlated and differ by only a small amount. At more complex facilities where more spatial variation of the wind field may be expected (e.g., due to significant topographical variations or obstructive complexity), then the use of multiple anemometers may be more advantageous. One puff is emitted for every second of the simulation (i.e., $\Delta t = 1$ second) and every puff's
355 position is advanced every second.

In contrast to GPM, the Gaussian puff framework requires time advancement of each puff’s spatial positioning and associated modulation of the concentration profile of each as the simulation advances. This is further coupled with processing procedures needed for the addition of new puffs and their proper accounting. Consequently, this makes the Gaussian Puff simulation for
 360 an equal number of measurement hours much slower than the GPM.

3.2 Inversion Frameworks

The primary objective of an inversion framework is to utilize ambient concentration measurements represented as the mass of a pollutant per unit volume of air (e.g., parts-per-million) to estimate the locations and rates of the emission sources as the
 365 mass of pollutant per unit time (e.g., kg/hr). The output of dispersion modeling can be expressed as a sensitivity matrix, $\mathbf{S}$, representing the response of every sensor to every potential source. Under the simplifying assumptions of constant emission rates and linear scaling between emission rate and concentration predictions (as implied by Equation 1), inferring the source rates can be expressed as an optimization problem that seeks the vector $\mathbf{Q}$ that minimizes an objective function of the residuals:

$$\min_{\mathbf{Q}} f(\mathbf{S}\mathbf{Q} - \mathbf{b}). \quad (10)$$

370 where $\mathbf{S}\mathbf{Q}$ depicts the predicted concentration vector calculated by summing the contribution of all emission sources $\mathbf{Q}$ at the measurement locations and times corresponding to the measured concentration vector, $\mathbf{b}$.

Selecting an appropriate inversion framework involves balancing computational cost with desired accuracy and control, which all depend on the application’s objective. This section presents two contrasting options, representing extremes in computational complexity, including a computationally efficient least-squares optimizer along with a more computationally expensive
 375 Markov-Chain Monte-Carlo (MCMC) approach. The MCMC inversion method approximates the full posterior distribution function in the high-dimensional parameter space of rate vector $\mathbf{Q}$ with more granular control over prior information and the selected objective function.

It should be emphasized that these are only two of many available methods for performing rate inference, which include ge-
 380 netic algorithms, stochastic variational inference, among many others. Rather than implementing an exhaustive list of inversion solvers, this section aims to apply two example inversion frameworks, spanning a range of complexity, to demonstrate the impact of method selection on the performance of emissions quantification algorithms, as evaluated using several key metrics. Note that there is no one-size-fits-all “best” framework for this problem. The optimal solution will depend on practical constraints (e.g., computational resources and required latency) and desired outcomes, such as a highly responsive leak detector
 385 that prioritizes detecting emission events of various sizes (even at the cost of false positives), or focusing on accurate estimation of cumulative emissions, even if some smaller leaks are potentially overlooked, or “rolled up” into a smaller number of

larger-rate emission points.

3.2.1 Least-Squares Optimization

390 The commonly used method of least-squares optimization (LSQ) seeks to minimize the sum of the squared differences between observed and predicted values (residuals). In this application, using the sensitivity matrix $\mathbf{S}$ and the measured concentrations $\mathbf{b}$, this optimization problem is formulated as:

$$\min_{\mathbf{Q}} \sum_i^N (\mathbf{S}\mathbf{Q} - \mathbf{b})^2 : \mathbf{Q} \in \mathbb{R}^+ \quad (11)$$

where there are N measurements and $\mathbf{Q}$ cannot be negative. In this context, the system is consistently over-determined, with the sensitivity matrix $\mathbf{S}$ having more linearly independent rows (measurements) than columns (sources). The physical interpretation of this system being overdetermined is that there is sufficient sensor coverage, wind variability, and duration of experiment to provide enough information to enable robust emission rate estimation for every source.

400 However, under unfavorable atmospheric conditions (such as a sustained wind direction with low variability due to horizontal turbulence during an emission event), linear independence is not guaranteed. In such cases, even with a large number of measurement data rows, the lack of linear independence may lead to an under-determined system. Furthermore, there may be cases when the wind never carries emissions from a specific source, i , towards a sensor. In this case, the i th column of $\mathbf{S}$ will entirely consist of zeros. As a result, any arbitrary rate for source i would equally satisfy the measurement data.

405 The data collected for this study exhibits sufficient wind variability, high enough sensor density, and long enough event duration to ensure that the system is always overdetermined. For example, a 30-minute experiment (the shortest duration in this study) that is monitored by 10 sensors generates 300 sensor minutes of data ($\mathbf{S}$ has 300 rows). In an extremely unlikely scenario with minimal wind variability, 95% of the rows of $\mathbf{S}$ may not be linearly independent, and the rank $\mathbf{S}$ is still 15, substantially larger than the five sources. Future work will explore the required information density and distribution to make robust emission rate estimates for CMS deployments that more closely resemble real-world conditions, encompassing sensor density, placement, and source count.

In practice, adding a degree of regularization that encourages sparsity in the inferred rate vector can significantly improve the detection and localization statistics. It will also prevent unrealistically high emission rates by requiring stronger evidence (in the goodness of fit) before assigning a high rate. For this purpose, a standard Lasso regression is implemented which incorporates an L1 norm penalty on the rate vector in the optimization problem:

$$\min_Q \sum_{i=1}^N (\mathbf{S}Q - \mathbf{b})^2 + \sum_{j=1}^M \alpha |Q_j| : Q \in \mathbb{R}^+ . \quad (12)$$

This approach introduces a hyperparameter α that penalizes nonzero rate entries with a magnitude proportional to the rate. Consequently, the solver prioritizes sparser solutions, assigning high-rate values only when they lead to a significant improvement in data fitting, thereby avoiding unphysically high rate estimates.

3.2.2 Markov-Chain Monte-Carlo

While a computationally complex and expensive approach to a linear optimization problem may seem unnecessary, it offers distinct benefits for specific applications, warranting its inclusion. In the context of the current study, this approach offers three key advantages. Unlike simpler optimizers like LSQ which is constrained to only minimizing the sum of squared residuals, the use of various error evaluation functions is permitted in this approach. Second, this method offers greater control over the use of prior information in the context of a rate inference. This enables more explicit promotion of sparsity, rate regularization, and even the incorporation of operational or independently gathered data to further inform estimates. For instance, SCADA data combined with emissions factors can generate time-dependent priors, favoring higher rates when a known equipment piece is in operating mode. Finally, a full posterior distribution approximation is crucial for continuous estimators. These estimators compute rates as a function of time on streaming data. They recursively take in the last known rate estimate and combine it with new incoming data to make an updated rate estimate. The posterior distribution of a given timestep can be used as the prior information in the subsequent timestep. This approach enhances emission rate estimation accuracy during periods of limited information. Prior information can be used to propagate source-specific rates through these periods until the proper signal is delivered from the given source to the system (the CMS observes a given source). A full description of a continuous state estimator employing a recursive Bayesian framework is beyond the scope of this paper and will be addressed in future work.

In the context of a Bayesian parameter inference problem, Markov-Chain Monte-Carlo technique (MCMC) constructs a Markov chain by drawing samples from a probability distribution and performing acceptance/rejection sampling. This chain samples the posterior distribution of the model parameters with a frequency that is proportional to the posterior probability. The objective is to estimate the posterior probability density function $P(D|Q)$ of rate vector Q , given some input data D . This can be achieved by Bayes' theorem:

$$P(Q|D) = \frac{P(D|Q)P(Q)}{P(D)} \quad (13)$$

where $P(D|Q)$ is the likelihood, representing how well the data fits a given set of model parameters, $P(Q)$ is the prior, and $P(D)$ is the evidence. Note that $P(D)$ is inconsequential in the context of the existing problem, because this calculation only considers relative probabilities between different Q when performing acceptance sampling, and $P(D)$ does not depend on Q , so it cancels out. As evident in Equation 13, the likelihood function and the prior must be defined in order to run this calculation. An appropriate prior to promote sparsity is the spike and slab prior, with a higher probability (the spike) at zero rate. The distribution of nonzero rates can be specified as $\beta(Q_i)$:

$$P(\mathbf{Q}) = \prod_i S(Q_i) \quad (14)$$

$$S(Q_i) = \begin{cases} H & \text{if } Q_i = 0 \\ \beta(Q_i) & \text{if } Q_i > 0 \end{cases} \quad (15)$$

Here, H represents the height of the spike. The relative magnitude of H compared to the characteristic value of $\beta(Q_i)$, defines the strength of the sparsity bias in the rate inference. The function $\beta(Q_i)$ can be defined using any other prior information available. For the present study, a constant value of $\beta(Q_i) = 1$ is employed, and H is set to 5. In other words, the algorithm favors zero rates over nonzero rates by a factor of 5, without imposing any additional prior on the nonzero distribution. A likelihood function is then defined using a standard approach of employing the chi-squared statistic via:

$$P(D|Q) = \exp[-\chi^2/2] \quad (16)$$

where

$$\chi^2(\mathbf{S}, \mathbf{Q}, \mathbf{b}) = \sum_i^N \frac{(\mathbf{S}\mathbf{Q} - \mathbf{b})^2}{\sigma_i^2}. \quad (17)$$

Here, $\mathbf{S}$, $\mathbf{Q}$, and $\mathbf{b}$ represent the sensitivity matrix, rate vector, and measurement vector, respectively and σ_i denotes the uncertainty associated with each measurement-prediction combination. In principle, σ_i can be a function of the particular dispersion model's error characteristics under certain conditions (i.e., a function of wind speed, stability class, etc.)

A range of sampling schemes with various complexity levels be can employed for approximating $P(Q|D)$. Examples include Metropolis-Hastings, Gibbs, and Hamiltonian (see Dunson and Johndrow 2019 for a recent review and history of MCMC sampling schemes). In this study, we implement the basic Metropolis-Hasting algorithm. A concise description is outlined here for

clarity.

470 First, a random rate vector is chosen as an initial starting point. Next, a random step is taken in the rate vector. The defined likelihood and prior functions are applied to both the old Q and the new Q to compute the ratio of new to old posterior probability. Then, a random number between 0 and 1 is generated. If this value is less than the ratio of the new to the old posterior probabilities, the new sample is accepted and added to the chain. Otherwise, the sample is rejected and the old rate estimate is retained as the most recent sample. This iterative process continues until pre-defined stopping criteria are met, which may include a maximum number of iterations, convergence of the distribution, or other criteria.

475

The result of this calculation is a “chain” of rate vectors that can then be analyzed further for the given purpose. Means or medians across marginalized distributions can be determined to infer the most common or best-fit rate values for each source. In addition, the standard deviations or inner-quartile ranges of the marginalized distributions can be calculated as a measure of uncertainty per source, and the covariance between different source rate estimates can be computed to gain a deeper understanding of the underlying distribution. For example, closely located sources may demonstrate a negative covariance. This indicates that the algorithm recognizes the need for rate allocation to one of the sources, but struggles to differentiate them. In this example, an equivalent data fit can be achieved by distributing the rate arbitrarily between these two sources.

485 Further processing of the rate estimates based on the covariances between sources can be applied to promote sparsity or to highlight potential instances of source confusion or mislocalization. It can effectively act as a quality control measure to flag rate inferences that may be error-prone. In this work, the best-fit rate from a Markov chain of rate vectors is inferred by computing the median of each element of Q . If the median is 0 (more than 50% of the samples for a given source are at 0 rate), that source rate is set to 0. If the median value is nonzero, then the zero-rate samples are excised and the median of the “slab” portion of the distribution is recomputed as the best-fit rate.

490

3.3 Evaluative Metrics

This study aims to answer the following questions: (i) How well can a CMS, under favorable network configuration conditions (high sensor density) localize emissions to the proper equipment group? (ii) How accurate are the total site-integrated emissions estimates? (iii) Can an advanced CFD-based forward model, where the wind field is first resolved, outperform the canonical plume and puff models in predicting the concentration field at CMS stations and site-integrated emission rates for these relatively simple controlled release experiments when combined with different inversion frameworks? To this end, the error metrics and evaluation of a given set of rate estimates in comparison to the “ground truth” are tailored to address these specific questions. Although the focus of this study is to investigate the accuracy of different combinations of the forward-

inversion frameworks, a more direct comparison of forward models using known emission rates and locations is investigated
 500 by computing several statistical error metrics on the predicted and measured concentrations across all forward models in Appendix C.

The result of an individual rate calculation (i.e., a specified dispersion model and inversion framework applied to data from an individual “experiment”) will be rate vector $\mathbf{Q}$, where each element of the vector represents the estimated rate associated with equipment group i . For a given experiment, the ground truth rate vector can be equivalently expressed and will be denoted
 505 as $\mathbf{Q}'$. For the remainder of this document, any primed values will indicate the actual ground-truth release information, while unprimed values represent the estimated quantities.

3.3.1 Localization Metrics

A binary classification scheme is employed to provide a proxy for localization accuracy. In this approach, a given rate vector $\mathbf{Q}$
 510 is processed into a binary vector ($\mathbf{D}$) representing the emission status of a given equipment group. If the rate of a given element is 0, then the associated binary element is set to 0 (not emitting), and if the rate is nonzero, the associated binary element is set to 1 (emitting):

$$D_i = \begin{cases} 1 & \text{if } Q_i > 0 \\ 0 & \text{if } Q_i = 0 \end{cases} \quad (18)$$

These binary values are then compared to the ground truth binary values and classified as True Positives (TP), False Positives
 515 (FP), False Negatives (FN), and True Negatives (TNs). A TP occurs when both the estimated and actual binary elements are 1 (the equipment group was emitting and properly estimated as emitting). FP indicates that the estimated binary element is 1 but the actual is 0 (an equipment group was estimated to be emitting but was not). FN occurs when the estimated binary is 0 and the actual is 1 (the equipment group was emitting but was not estimated to be emitting), and a TN indicates that both binary elements are 0 (the equipment group was not emitting and was not estimated as emitting). These designations effectively
 520 represent the capability of the system to parse out information from the concentration measurements and localize that source to the correct group. For each experiment, the number of correctly identified sources (i.e., the addition of TNs and TPs) is computed to give a localization score (L):

$$L = \sum_{j=1}^M \begin{cases} 1 & \text{if } D_j = D'_j \\ 0 & \text{if } D_j \neq D'_j. \end{cases} \quad (19)$$

A perfect localization score, $N_{L=5}$, is achieved when the emission status of each equipment group is correctly identified (as a TN or TP), resulting in a value of L that equals the number of equipment groups in the experiment (the length of $\mathbf{Q}$). In addition, the total number of false positives and false negatives is recorded for each quantification approach (dispersion model plus inversion method) across all experiments. This enables an analysis of each system's tendency to either over-predict or under-predict the number of active sources as a function of dispersion models and inversion frameworks.

3.3.2 Quantification Metrics

The metrics in this section are developed to evaluate the accuracy of total site emissions quantification using CMS by comparing estimates to the ground-truth total emissions from the facility. First, total emissions estimates for every single experiment are calculated. The rate vectors for a given experiment are summed before error metrics are computed such that the total emission estimate during an experiment (Q_{tot}) is simply:

$$Q_{tot} = \sum_j Q_j. \quad (20)$$

A set of quantities are then calculated across all experiments. First, the mean error ($\overline{E}$), which is a direct measurement of the system's bias (i.e., the mean of the error distribution of facility-level quantified rates) is computed as follows:

$$\overline{E} = \frac{1}{N} \sum_{i=1}^N Q_{tot} - Q'_{tot}, \quad (21)$$

where N is the total number of experiments. As a proxy for the uncertainty of the rate estimates, the mean absolute error ($|\overline{E}|$) is then calculated. It is a measure of how far off in total the emissions estimates are, on average ($\pm |\overline{E}|$ kg/hr):

$$|\overline{E}| = \frac{1}{N} \sum_{i=1}^N |Q_{tot} - Q'_{tot}| \quad (22)$$

Analogous quantities (the mean percent error and mean absolute percent error) are computed for the normalized error ($(Q_{tot} - Q'_{tot})/Q'_{tot}$) to better account for low-rate experiments that have less influence on the raw unnormalized error metrics, denoted $\overline{E}_{rel}$ and $|\overline{E}|_{rel}$. Finally, the fraction of rate estimates that are within a factor of two of the actual rate ($F2$) is computed via:

$$F2 = \frac{1}{N} \sum_{i=1}^N \begin{cases} 1 & \text{if } 0.5 < \frac{Q_{tot} - Q'_{tot}}{Q'_{tot}} < 2 \\ 0 & \text{else} \end{cases} \quad (23)$$

In addition to these summary statistics of the error distribution, the total cumulative estimated emissions (in terms of total mass) is calculated across all experiments. The cumulative emission at time t , is determined by aggregating total emissions through experiments up until that time is computed via:

$$C(t) = \sum_{i=1}^{N_t} Q_{tot} \Delta t_i. \quad (24)$$

550 where Δt_i is the duration of Experiment i . The cumulative error, $\Delta C = C(t_f) - C'(t_f)$, is reported at the end of a set of experiments (i.e., the end of the last experiment). Note that cumulative error should be very similar, although not identical, to the mean error multiplied by the total duration of the experiments, as this metric accounts for the duration of each unique experiment, whereas the mean of the raw error distribution does not factor in experiment duration.

555 To aid in the understanding of these metrics, an illustrative example of these evaluative metrics applied to a single experiment from the study is shown in Figure 3. This figure shows an image and two tables that summarize the output of the system (rate estimates for each equipment group) alongside the ground-truth release rates (top table) and computes the relevant evaluative metrics for this single experiment (bottom table). During this experiment, there were three active release sources: the tanks (group 4T), the western separators (group 4S), and the eastern separators (group 5S). The western and eastern wellheads (groups 4W and 5W, respectively) were not emitting. The quantification estimates are shown in the 2nd column of the top table, while the ground-truth release rates are shown in the 3rd column of the top table. The final column shows the classification of the estimate as either a TP/FP/FN/FP as previously described. The table on the bottom shows the relevant evaluative metrics applied to the estimated and ground-truth rates in the top table. We see that, for this example experiment, there were 2 true positives (the system accurately identified that both the 4S and the 5S groups were emitting), one false negative (the system missed that the tanks were emitting), one false positive (the system assigned a small but nonzero rate to the 4W group, which was not emitting), and one true negative (the system accurately identified that the 5W group was not emitting). These statistics are summarized in the bottom table, along with the overall “localization score”, which in the case, was 3 (i.e., the emission status of 3 out of the 5 equipment groups were correctly identified). The total estimated and actual facility-level emission rate is shown in the bottom table as Q and Q' (these are computed as the sum down the “Estimated Rate” and “Actual Rate” columns, respectively). In this example, the estimated facility rate is 1.73 kg/hr while the actual emission rate is 1.83, representing an error of -0.1 kg/hr (E) and a relative error of -0.055 (i.e., -5.5% error, E_{rel}). In terms of the other quantification-related metrics ($F2$ and ΔC), this experiment’s estimated facility-level rate is within a factor of 2 of the actual rate (so it would positively contribute to the fraction of estimates that were within this factor, when summing over all experiments), and the contribution to the cumulative error from this experiment would simply be $E\Delta t$, where Δt is the duration of this experiment. The duration of this particular experiment is 30 minutes, so the contribution to ΔC is -0.05 kg.

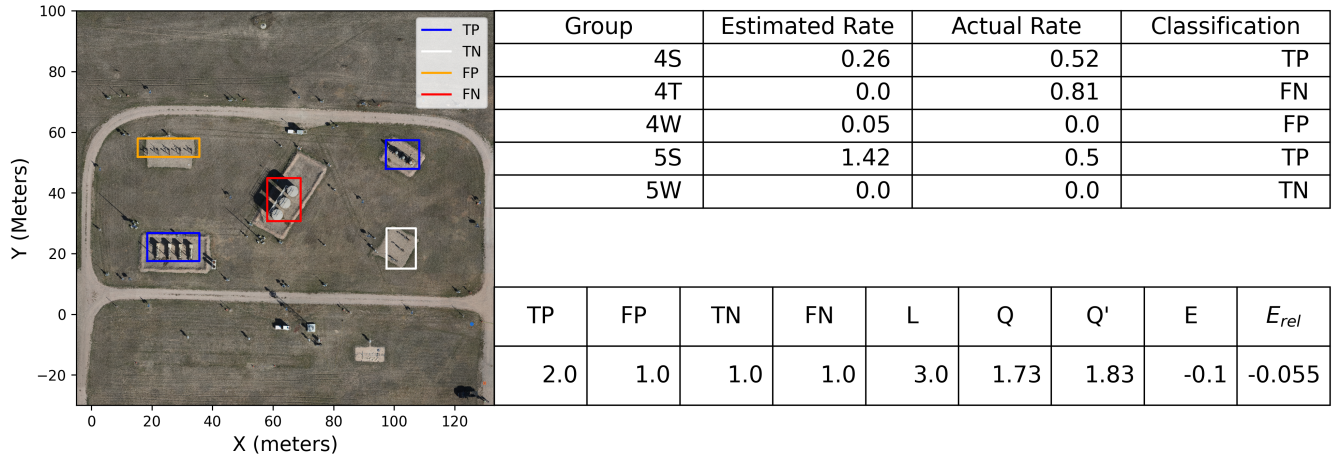


Figure 3. Example experiment to illustrate the evaluation of the output of the system with respect to ground truth rates. The image on the left shows each equipment group’s estimate classified as either a TP/FP/FN/FP. The upper table summarizes the estimated rates, actual rates, and the detection classification, while the lower table applies the evaluative metrics described above to the data from the upper table.

4 Results

Subsection (4.1) details the application of each unique combination of the Puff and Plume dispersion models and inversion framework to the set of 347 experiments. Due to the high computational cost of performing CFD across the entire set of experiments, only a small number of representative cases are computed, the results of which are discussed further in subsection 580 4.2. The evaluation metrics associated with each combination of Plume/Puff/CFD and LSQ/MCMC are then computed and discussed.

4.1 Localization and Quantification Using Gaussian Models

The results obtained by employing each combination of the Gaussian dispersion model and each inversion (LSQ and MCMC) framework across all 347 controlled release experiments are presented in this section. Table 1 details the summary statistics 585 for each combination. In general, the more complex combinations (i.e., puff over plume and MCMC over LSQ) result in better error statistics across the majority of metrics. These improvements are especially prevalent in the localization-related statistics ($N_{L=5}$ and $\bar{L}$) and the variance of quantification errors (e.g., $|\bar{E}|$ and F2). For example, consider the combination of the GPM (Plume) and Least Squares (LSQ) fitting as the simplest combination. In this case, the number of experiments where the emission status of each equipment group are all correctly identified ($N_{L=5}$) is 85 (out of 347). When applying the same LSQ 590 inversion framework but increasing the complexity of the dispersion model to the Gaussian Puff (Puff), this number goes up to 116. In contrast, holding the dispersion model constant (Plume), but applying the MCMC inversion results in this number increasing to 149. Finally, when using the more sophisticated Puff dispersion model **and** MCMC inversion, the number of

cases where all 5 equipment groups' emission status is correctly identified increases to 184.

595 The improvement in localization statistics when employing the Gaussian Puff instead of the Gaussian plume can be explained
by the difference in the fidelity of the temporal modeling of the dispersion. The GPM computes minute-averaged velocity
fields, which are assumed to be spatially homogeneous (an assumption that underpins the derivation of the Gaussian Plume)
and uses this singular mean value to approximate the dispersion of gas on minute-averaged timescales. In contrast, the Gaussian
Puff model directly integrates the spatially and temporally varying wind field on much finer timescales (using 1-Hz wind data),
600 resulting in more accurate dispersion trajectories that take into account the spatial and temporal variation of the wind field.

Improvement in localization statistics when going from the simple LSQ fitting to the MCMC inversion is a direct consequence
of the more aggressive sparsity promotion employed in the MCMC algorithm and more sophisticated postprocessing of the
iterative 5-dimensional chain of rate vectors and associated probabilities. This results in a significantly smaller number of false
605 positives and also a noticeable decrease in the number of false negatives. This highlights the importance of choosing the ap-
propriate inversion framework for the desired outcome. For instance, if over-localizing (producing nonzero emissions where
no emissions were occurring) is not a concern for the application at hand, and only a rough estimate of cumulative emissions
is desired, then using the simple LSQ inversion may be appropriate. If, however, the localization output is being used to guide
manual detection and remediation efforts (e.g., OGI inspections), then reducing the potential search area via more accurate
610 localization is of critical importance. Therefore, a framework that produces fewer false positives (while maintaining a low false
negative count) may be worth the additional computational cost.

Histograms of the number of correctly identified emitters (L) for all 4 combinations of the dispersion model and inversion
framework are shown in Figure 4. Note that the number of cases with poor localization results (where only 1 or 2 of the equip-
615 ment groups' emission status is identified as correct) is significantly lower for the MCMC inversion than the LSQ. Employing a
combination of Puff and MCMC results in only 12 (out of 347) experiments with poor localization (localization scores smaller
than 3), whereas the combination of Plume and LSQ has 66 cases with poor localization. This highlights the advantage of em-
ploying an inversion framework with more aggressive and controllable sparsity promotion. The majority of this improvement
is driven by reducing the false positive count, which is achieved by more strictly penalizing nonzero rates.

620

Figure 5 shows the actual vs estimated facility-integrated rates across all 347 experiments for all 4 combinations of the dis-
persion model and inversion framework on a logarithmic scale. The black dashed line indicates the parity ($y = x$) relation.
The bounding dotted black lines show a factor of 2 above and below parity ($y = 2x$ and $y = 0.5x$) for reference. Note that
there is significantly less scatter in the red dots about the dashed black line than there are any other color, indicating that the
625 combination of Puff MCMC yields the tightest distribution of rate estimates about the parity line. This is supported by the F2

Table 1. Summary statistics of all 4 combinations of dispersion model and inversion calculation.

Method	TP	FP	FN	TN	$\bar{E}$	$ \bar{E} $	$N_{(L=5)}$	$\bar{L}$	F2	$\bar{E}_{rel}$	$ \bar{E} _{rel}$	ΔC
Plume LSQ	713	404	62	556	-0.1	1.13	85	3.66	73	0.13	0.63	-135.02
Puff LSQ	718	319	57	641	-0.02	0.96	116	3.92	79	0.13	0.54	-55.89
Plume MCMC	729	242	46	718	-0.11	0.87	149	4.17	85	0.12	0.48	-117.44
Puff MCMC	743	192	32	768	0.02	0.8	184	4.35	89	0.13	0.42	12.44

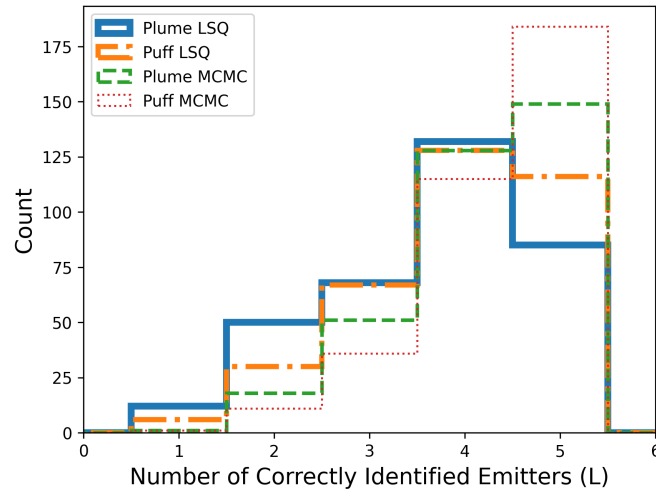


Figure 4. Histograms of the number of correctly identified emitters across all 4 combinations of dispersion model and inversion calculation.

statistic from Table 1, which shows that the combination of the Puff MCMC quantification algorithm has the highest percent of estimates within a factor of 2, and the smallest absolute relative error, $|\bar{E}|_{rel}$.

Figure 6 shows the actual vs estimated facility-integrated rates across all 347 experiments for various combinations of the dispersion model and inversion framework all together on a linear scale. In this panel, linear fits to the data are shown with the slope and associated R^2 shown in the legend. The linear fits indicate that the quantification estimates that utilized Puff have slopes closer to 1 (0.87 and 0.89 for the Puff LSQ and Puff MCMC, respectively) compared to the quantification estimates that utilized the Plume (0.8 and 0.82 for Plume LSQ and Plume MCMC, respectively). It is worth noting that the slope of this line is not a direct measurement of the bias. This is because these linear fits are generally computed by minimizing the squared error, and as such, a single outlying event with a relatively high error can have an outsized effect on the inferred slope. The slopes of these lines are more directly related to the average signed squared error than the bias. This being said, the slopes of these lines are often interpreted as a rough proxy for the bias of a system, and as such, are worth considering with the appropriate context. The trends evident in the linear fits across different quantification estimates are mirrored in the two evaluation metrics

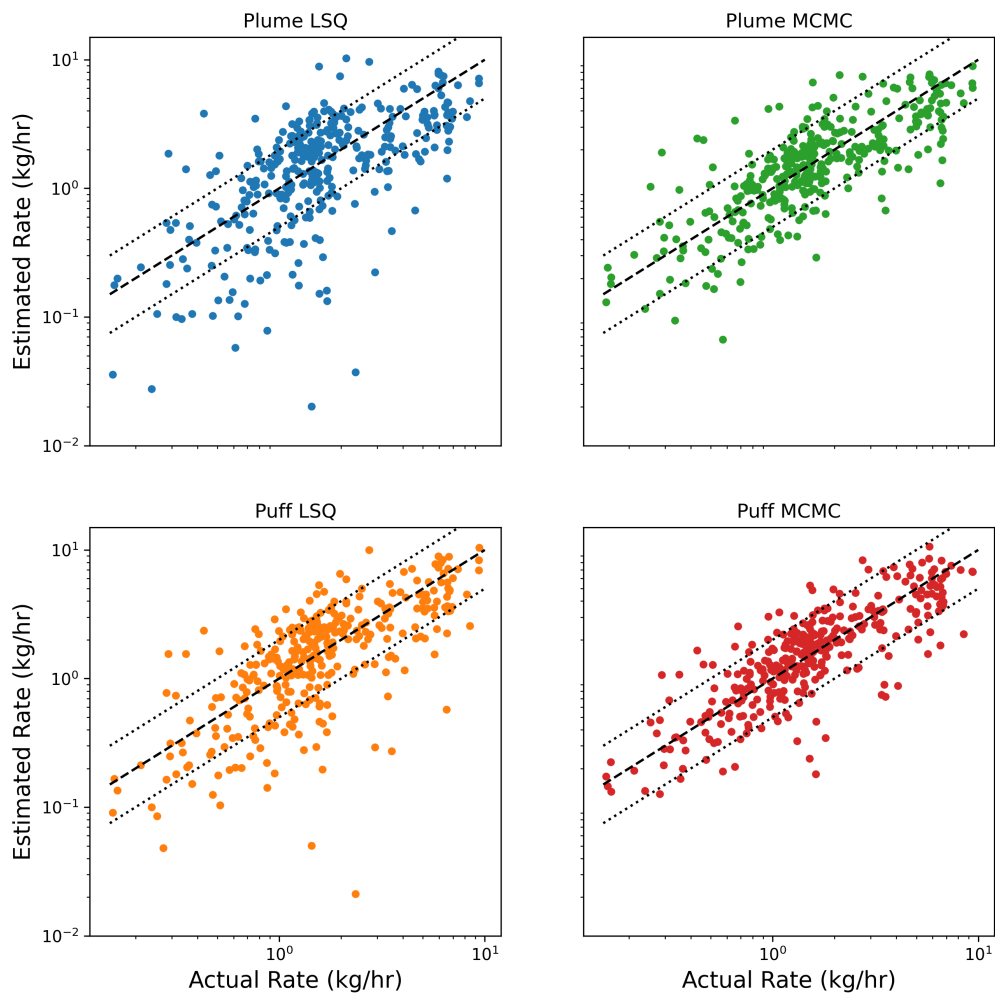


Figure 5. Actual vs. estimated facility emissions for 347 experiments on a logarithmic scale. The dashed black line depicts the parity relation ($x = y$). The dotted lines indicate a factor of 2 lower and higher than the parity relation.

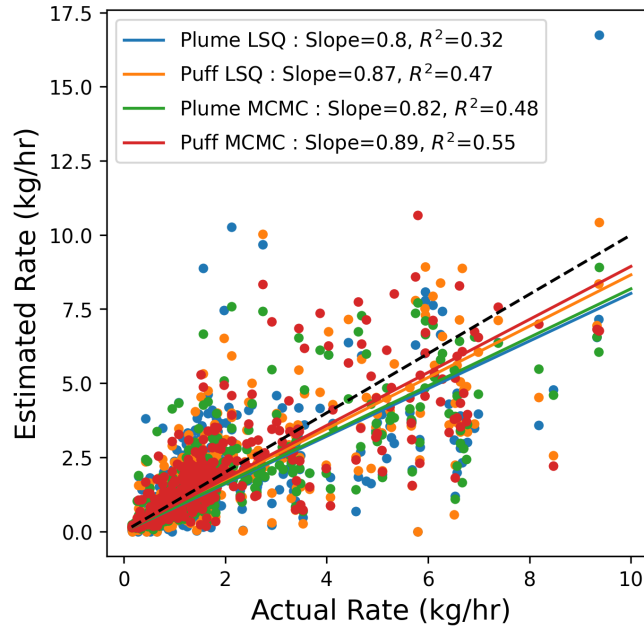


Figure 6. Actual vs. estimated facility emissions for 347 experiments on a linear scale. The dashed black line depicts the parity relation ($x = y$). The lines correspond to the linear fit to each quantification method's respective actual-vs-estimated pairs. The slopes of these lines and the associated R^2 value are shown in the legend of the right panel.

that best relate to the bias of the system in Table 1: the average error, $\overline{E}$, and the cumulative error, ΔC . More specifically, these
640 quantities show the lowest bias (closest to 0 values) for the estimates from the Puff models, which is reflected in Figure 6 (the
linear red fit for the quantification estimates using Puff MCMC and orange for the Puff LSQ that have slopes in the parity plots
that are the closest to 1).

The coefficient of determination (R^2) is shown for each linear fit in the right panel of Figure 6. Similar to how the slope,
645 while related, does not directly measure bias, R^2 reflects the variance of the distribution about the linear fit. In other words,
it can be used as an indicator for the statistics from Table 1 related to the variance of the error distribution (F^2 , $|\overline{E}|$, $|\overline{E}|_{rel}$).
Similar trends are evident in the R^2 values inferred from the linear fits, with the coefficient of determination of the linear fit
getting closer to 1 for increasing complexity in the forward modeling ($R^2_{Puff} > R^2_{Plume}$), as well as in the inverse solver
($R^2_{MCMC} > R^2_{LSQ}$).

650

Error histograms for these 4 quantification calculations are shown in Figure 7 to illustrate that the near-zero-error peak is
significantly higher for the combination of Puff and MCMC than it is for any of the other quantification methods and drops
off more quickly towards higher error. The Plume LSQ results show the most high-error rate estimates, and the Plume MCMC

combination generally shows a marginal improvement over the Puff LSQ. The box-and-whisker plots of the relative error
655 distribution associated with each quantification method are shown in Figure D1.

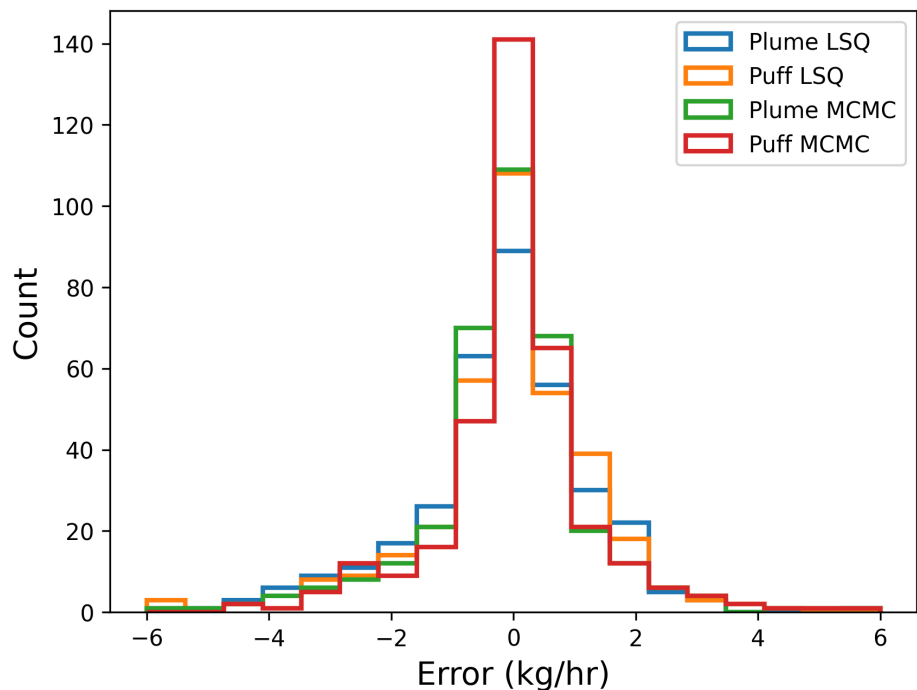


Figure 7. Error histograms for each quantification method across all controlled release experiments.

A pairwise Kolmogorov-Smirnov (KS) test was employed to statistically test the significance of the differences in relative error
distributions among the four quantification methods for all 6 combinations of distributions. Table 2 presents the results of the
KS statistic and associated p-value for the two error distributions from each combination of quantification methods. The KS
statistic represents the maximum difference between cumulative distribution functions. It can be used as a measure of distribu-
660 tion similarity, with a smaller value indicating greater similarity in distributions. The p-value represents the probability of the
two sample distributions being drawn from the same underlying probability distribution. Unsurprisingly, the highest degree of
distinction between relative error distributions is observed between Plume LSQ and Puff MCMC methods, the two methods
that are respectively identified as the worst and best-performing methods.

665 The pairwise comparison of Plume LSQ and Puff MCMC error distributions has a KS statistic of 0.17, the highest of any
other combination, as well as the lowest p-value of 0.00009. It indicates that the null hypothesis that the two samples could be
drawn from the same underlying distribution can be rejected to a very high degree of certainty. In contrast, the two LSQ-based
inversions have a significantly smaller KS statistic, indicating that the two distributions are more similar than any of the other
combinations listed in Table 2. A p-value of 0.208, indicates a substantially higher probability that the two samples could have

670 been drawn from the same underlying distribution.

It worth noting that 4 out of the 6 combinations of distributions have statistically significant differences (p-values < 0.05). The other combination that was not different to a high degree of statistical significance is the Plume MCMC and Puff LSQ methods, showing a KS statistic of 0.1 and a p-value of 0.086. The error metrics related to quantification accuracy indicate improved error distributions for the higher-fidelity modeling/inversion choices (Puff over Plume and MCMC over LSQ). Any quantification method that combines a high-performing model (e.g., Puff for dispersion modeling or MCMC for inversion analysis) with a lower-performing one (Plume for dispersion modeling or LSQ for inversion analysis) yields an intermediately performing algorithm, resulting in relatively comparable error distributions when considering the Puff LSQ / Plume MCMC combination.

Table 2. Table of statistical significance in error distribution differences between every combination of quantification method.

Methods	KS Statistic	p-value
Puff LSQ / Plume LSQ	0.08	0.208
Plume MCMC / Plume LSQ	0.14	0.003
Plume LSQ / Puff MCMC	0.17	0.00009
Plume MCMC / Puff LSQ	0.1	0.086
Puff LSQ / Puff MCMC	0.13	0.007
Plume MCMC / Puff MCMC	0.12	0.019

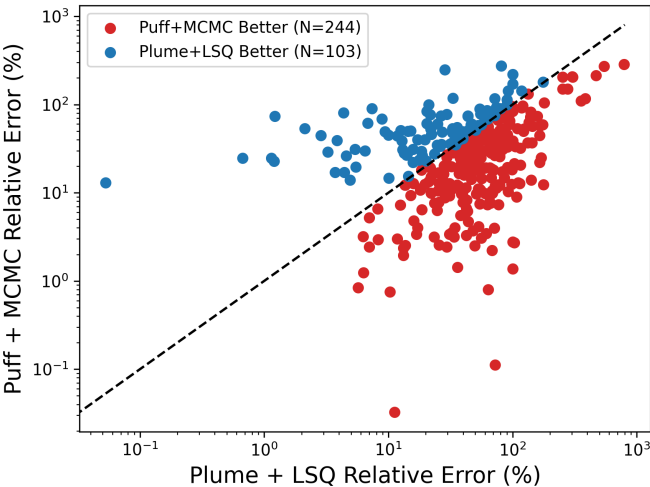


Figure 8. Scatter plot of Plume LSQ versus Puff MCMC relative errors. Red dots denote experiments where the Puff MCMC outperformed the Plume LSQ calculation, while blue dots denote the opposite.

Figure 8 further illustrates the improvements that can be realized by employing more sophisticated dispersion modeling and inversion frameworks by comparing the relative error associated with each experiment for the Plume LSQ against the Puff MCMC emissions quantification methods. The dashed black line depicts $x = y$, denoting equal errors, and the points are colored corresponding to whether they fall above the line (higher error from Puff MCMC colored in blue) or below the line (higher error from Plume LSQ colored in red). The number of experiments for which the Puff MCMC method outperformed the Plume LSQ was 244 out of 347. In other words, the Puff MCMC rate inference was superior to the simpler Plume LSQ 70% of the time.

4.2 The Application of CFD-Driven Localization and Quantification for a Subset of Experiments

Compared to the plume and puff forward models, performing CFD simulations is considerably more computationally expensive, especially when the simulations target resolving the relevant scales embedded inside the surface layer. This cost increases further setting up the sensitivity matrix, which is the required input for the inversion frameworks. This requires solving five additional scalar transport equations where for each equation only one of the five emission sources at the ADED facility is actively emitting at a unit rate. Therefore, quantification and localization estimates using the CFD framework as the forward model were accomplished for a small subset of experiments. As such, 4 experiments are randomly selected. The CFD framework described in Appendix B is applied to generate sensitivity matrices for each of the 4 experiments. Then, the inversion process is performed using LSQ and MCMC, and the error metrics described in Section 3.3 are computed. Table 3 highlights the results of the comparison among the three forward models, indicating the performance improvement realized when employing CFD compared to both Plume and Puff. Only the MCMC results are shown in this table for clarity. Note that the Plume and Puff results presented in this section are also derived from the same four selected experiments to permit a fair comparison.

Table 3. Table of error metrics across all dispersion models, including CFD, using MCMC inversion applied to the smaller subset of experiments that CFD was performed on.

Method	TP	FP	FN	TN	$\bar{E}$	$ \bar{E} $	$N_{(L=5)}$	$\bar{L}$	F2	$\bar{E}_{rel}$	$ \bar{E} _{rel}$	ΔC
Plume MCMC	7	8	1	4	0.58	0.58	1	2.75	1.0	0.28	0.28	1.67
Puff MCMC	8	3	0	9	0.6	0.92	2	4.25	1.0	0.3	0.45	1.56
CFD MCMC	8	3	0	9	-0.14	0.28	1	4.25	1.0	-0.06	0.13	-0.61

Table 3 indicates comparable localization statistics (TP, FP, FN, TN and $\bar{L}$)) for the CFD and Puff models. However, quantification statistics for the CFD models show significant improvements over the other dispersion models: $\bar{E}$ for CFD is -0.14 while the Plume and Puff are 0.58 and 0.6, respectively, a factor of ~ 4 farther from 0. Similarly, the average absolute error, $|\bar{E}|$ is substantially better (0.28 compared to 0.58 and 0.92 for Plume and Puff, respectively). Relative metrics also show analogous improvements.

705 Figure 9 shows parity plots for all dispersion models using the MCMC and LSQ inversion frameworks on the left and right, respectively. This figure further emphasizes the performance advantage that CFD-based quantification offers over the Plume and Puff models. While there are some instances where either the Plume or Puff performs better than the CFD, the CFD-based inversion shows an obviously better fit to the parity line than the other methods, on average. As discussed in the appendix, these improvements result from the CFD approach’s ability to reproduce the underlying unsteadiness and the near-surface complex
 710 flow effects with greater accuracy and detail.

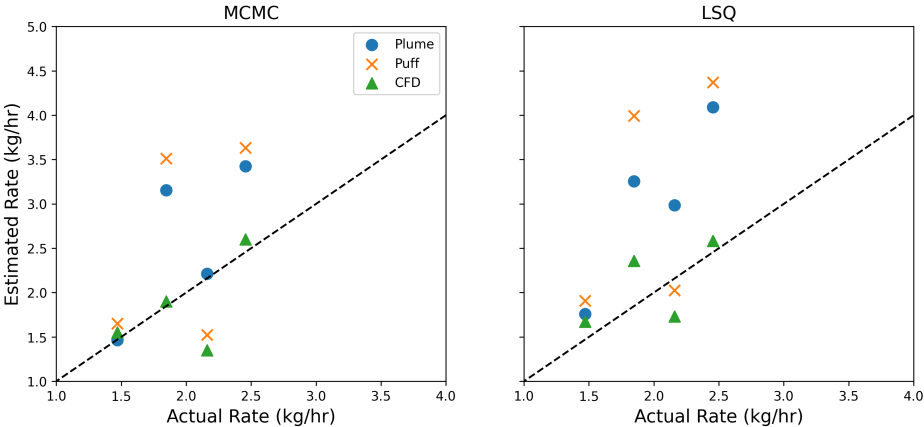


Figure 9. Parity plots for all three dispersion models using MCMC (left) and LSQ (right) for the small sample of CFD-computed experiments

5 Discussion

The development of multi-source methane emission DLQ algorithms is essential for accurate detection and quantification of oil and gas methane emissions. In these facilities, multiple emissions from different sources, varying in magnitude and location, can occur due to the complex infrastructure and operational processes. The shortcomings of single-source models in
 715 disentangling the overlapping plumes from these multiple leaks can lead to significant errors in both the estimated emission rates and the identified leak locations. Multi-source approaches, on the other hand, enable the independent quantification and localization of each individual leak. This capability is crucial for effective facility-level risk assessment and mitigation strategies, as it allows operators to prioritize repairs and address the most significant emission sources. In addition, a comprehensive understanding of the temporal and spatial distribution and magnitude of simultaneous leaks provides a more comprehensive
 720 picture of overall site emissions, which is critical for regulatory compliance and accurate emissions inventory development.

Multi-source methane emission DLQ algorithms require advanced dispersion and inversion methods that account for different aspects of short-range plume transport and inversion. This study represents an initial step toward developing more sophisticated solutions to enable multi-source methane emissions DLQ. However, several simplifying features were implemented in

725 this work, primarily imposed by the data constraints inherent in these specific controlled release studies. Key simplifications include: (i) facility complexity level, (ii) lack of terrain complexity, (iii) lack of complex, time-varying baseline emissions, (iv) constant emission release rates, (v) synchronous emission events during each experiment, and (vi) absence of higher-rate (>10 kg/hr) releases. Furthermore, the focus of this work was on the localization and quantification of constant-rate sources for **known** emission start and end times: the detection and time-bounding of emission events were not a part of this study.

730

CSU's METEC could be a good representation of relatively simple real-world operational upstream oil and gas facilities. However, other types of facilities, including midstream sites may be more congested, representing an additional complexity level in terms of the number of sources, higher and more fluctuating baseline emissions, emission patterns, and obstructive complexity that may render certain dispersion models inapplicable. Also, the METEC facility is located in an area with fairly simple terrain. However, facilities in other regions with more complex terrain, such as Appalachia, can present other challenges related to natural obstacles. This aspect may require the consideration of alternative dispersion modeling techniques that account for the impact of complex terrain and obstacles, such as the CFD simulations informed by on-site wind measurements presented in this study.

735

740 Baseline emissions often depend on many factors, including facility type, site-specific operational activities, facility size, facility age, maintenance practices, and many other considerations. The measurement data collected during the 2024 CSU METEC study did not include any baseline emissions. However, the magnitude of baseline emissions as well as the magnitude of their fluctuations can significantly impact the application of any DLQ solution.

745 The 2024 CSU METEC study featured only constant emission release rates within each experiment, with simultaneous activation and deactivation of the emitting sources at the experiment's start and end times. Consequently, for each experiment, the facility alternated between a sterile "off" state and an "on" state with constant emission rates. These simplified and known patterns of emissions constitute "prior" information that algorithms can, in principle, exploit. In addition, the experimental design required an event-based quantification reporting. This approach may be less practical, and as a result not suited for real-world applications. In the presence of asynchronously changing time-varying source rates (as expected at operational sites), event-based quantification will not properly capture the relevant features in emission timeseries. As a result, the 2024 CSU METEC controlled release testing performance may not fully generalize to the complex emission patterns prevalent in real-world operational settings. It should be noted that to address these concerns, the CSU METEC has developed a more advanced testing protocol that more accurately replicates the complex emissions found at operational facilities, including simulating operational background emissions.

755

The release rates employed for the controlled releases in this study are not sufficiently high for chemical buoyancy to be relevant. However, for large emission events (e.g., super-emitters > 100 kg/hr) neglecting the chemical buoyancy will lead to an overestimation of concentration enhancements for a given source rate, and hence an underestimation of the release rate in the inversion of measurements to rate. Future work with higher rate controlled releases will explore how different approaches to approximating the effects of chemical buoyancy affect the resulting quantification estimates from CMS.

The current study focuses on a small subset of dispersion models and inversion frameworks that are well-established and commonly used for emissions quantification. This deliberate choice was driven by a key objective: to provide a transparent comparison of the performance of methodologies commonly used in atmospheric dispersion modeling and emission quantification. Applying these methods to measurement data with high quality ground-truth releases helps quantify the uncertainty associated with rate estimates. In principle, by applying these same algorithms across different sensor configurations, specific hardware, and sensing modalities (e.g., metal oxide vs. TDLAS), the uncertainties associated with algorithms could be disentangled from the uncertainties inherited from specific deployment strategies and hardware specifications.

In this study, the Gaussian models (plume and puff) were selected for their computational efficiency and widespread application, providing a baseline for comparison. At the opposite end of the complexity spectrum, the CFD modeling was selected for its capability to provide a high-resolution, 3D representation of atmospheric dispersion. This method can capture complex flow patterns that alternative models often overlook. As a result, the CFD modeling allows for detailed simulations of plume behavior, particularly in scenarios involving complex terrain or variable wind fields, where accurate representation of turbulent mixing is crucial. Moreover, CFD models can offer additional benefits to integrate facility-specific data, like site-specific temperature measurements, enabling a more tailored and accurate simulation compared to alternative dispersion models. The MCMC inversion framework was chosen as a more computationally intensive alternative to LSQ for its ability to handle complex, non-linear problems and provide the full high-dimensional posterior probability distribution, which can enable recursive Bayesian estimation for at-scale continuous deployment of these systems (i.e., non-event-based quantification).

Note that the landscape of dispersion modeling and inversion frameworks is far more extensive. The exploration of alternative and often more complex methods, such as Lagrangian stochastic models or more sophisticated computational fluid dynamics (CFD) approaches, could offer valuable insights into the behavior of emissions under complex terrain or highly variable atmospheric conditions. However, for facilities with relatively simple setups and emissions patterns, improvements to the results by employing more sophisticated techniques may be marginal. Therefore, the selection of the most appropriate approach for quantification depends on the objectives and expected accuracy levels.

While a detailed analysis of how meteorological conditions affect the accuracy of the detection, localization, and quantification algorithms is deferred to a future study, there are several expectations that can be provided based on the underlying physics of gas dispersion and assumptions utilized by Gaussian models that may provide some insight into how they will perform under certain conditions. For instance, the utility of these systems may be significantly decreased during time periods with extremely high wind speeds. This is because the measured concentrations scale inversely with wind speed, so if the wind speed is sufficiently high such that the measured concentration enhancements are within noise of the measurement device, then the system's reliability in terms of converting these concentration enhancements to localized source rates will be significantly inhibited, and in some cases, impossible. The precise wind speed cutoff for this depends on the characteristic source-sensor distances, release rates, atmospheric stability, and the sensitivity of the hardware. In addition to high wind speeds having the potential to negatively impact the performance of CMS-based estimates, extended periods of time with extremely low wind speeds can also pose challenges. In the plume-based implantation presented in Section 3.1.1, measurement data points with corresponding wind speeds of less than 0.5 m/s are excised from the analysis due to the Plume's inapplicability under these circumstances. In practice, this means that if there is a period of time when the wind speed is always less than this threshold, then the plume model, as implemented and presented here, will not be able to quantify emissions from this time period. In contrast, the puff-based model can capture these low wind speed time periods, however the standard dispersion coefficients that are employed may not be as accurate during extremely low-wind speed conditions, when gas pools in place, and as such, the accuracy of puff-based quantification estimates will likely be negatively impacted. Finally, time periods with little variability in wind direction are prone to source confusion (see, e.g., Ball, Eichenlaub, and Lashgari 2025), and as such, the accuracy of the system during these time periods will be negatively impacted. Future work will more quantitatively explore how the output of CMS-based quantification estimates is affected by these various meteorological conditions.

These results represent something of a best-case-scenario in terms of the relative simplicity of the facility as well as the over-dense network of sensors that is deployed for this study. In general, the accuracy of the system will likely decrease with lower sensor density. How, exactly, the performance is affected by varying the number of sensors and their configuration will likely depend on the details of the specific facility (number and layout of emission points) as well as the typical variability in the wind direction. In general, we expect the impact of sensor density on DLQ accuracy to be independent of specific model choices (in terms of inverse solvers and forward models). However, future research should explore more quantitatively how the sensor deployment strategy, in terms of both density and configuration, affects the accuracy of various DLQ algorithms from CMS.

Several more in-depth analyses of the performance of these systems and associated and algorithms are possible with this data: generating detection curves as a function of emission rate and inference of 90% detection limits, investigating the per-group detection statistics, rerunning algorithms with different subsets of the underlying sensor data, investigating how well the system is able to detect small leaks in the presence of larger simultaneous emissions, as well as the impact of emission duration on the DLQ statistics. While the focus of this paper was on some relatively simple evaluative statistics related to the total site-level

emissions, and detection and localization accuracy, these more in-depth analyses will be investigated in future work.

825 It is worth noting that while the puff model, driven by high-frequency spatially-informed wind measurements, outperformed
the plume across all metrics, the decision of which model to employ should be informed by the needs of the specific application.
For instance: in many cases, high-frequency wind data may not be available, which may render some of the advantages of the
puff model moot. Additionally, with the same inversion framework, the plume model's overall quantification estimates were
not dramatically worse than the puff: for instance, when comparing the Plume+MCMC and Puff+MCMC models, the fraction
830 of estimates within a factor of 2 was only 4% lower, and the mean relative absolute error was only 6% higher when using the
plume model. Additionally, the cumulative mass estimate, while showing more negative bias than the puff model, was only
off from the true cumulative mass by about 5%. In many cases, such as deployment of these algorithms at scale, especially on
facilities without high-frequency wind data (or at extremely simple facilities with no obstructions where the wind field is more
homogeneous), the additional computational cost of employing the puff model may not be worth the marginal gains. In cases
835 with more complex wind fields, available high-frequency wind data, and a need for accurate localization, then the puff model
should likely be implemented.

Future research should prioritize the evaluation of various quantification methods to refine our understanding and improve the
accuracy of emission estimates across diverse operational settings with more complex operational emissions scenarios. This
840 could include more complex facility layouts with a larger number of sources and obstacles, higher baseline emissions with
increased fluctuations, more complicated emission patterns (e.g., time-varying emission rates), and case studies located in var-
ious regions to account for atmospheric diversity. This highlights the value of conducting more controlled release studies to
generate datasets that are representative of various real-world scenarios.

845 This study underscores one of the primary applications of CMS as long-term integration of emissions for accurate estimates
of the total mass emitted. These long-timescale estimates enable the detection of anomalous emission patterns, such as an in-
creased weekly-averaged facility-level emission rate, potentially indicating anomalous events such as a persistent fugitive leak
or higher-than-average operational emissions. Future research will examine the impact of sensor density and configuration on
quantification accuracy.

850

6 Conclusions

This paper presents algorithms for inferring constant-rate emissions from temporally distinct emission events from multiple
synchronized emission points. The Gaussian Plume (Plume), Gaussian Puff (Puff), and a CFD-based approach are detailed
and coupled with two inversion frameworks, including a simple least-squares estimator with L1 regularization (LSQ) and a

855 Markov Chain Monte Carlo (MCMC) based approach with a spike-and-slab prior. Each combination of the forward model and inversion framework is applied to a set of multi-source constant-rate controlled release experiments. A set of evaluation metrics is presented to investigate the performance of each quantification method for the localization and quantification accuracy. This approach demonstrates the functional dependence of selected key metrics on both forward and inverse modeling techniques.

In general, utilizing Puff, driven by high-frequency 1-Hz wind data and accounting for the spatial inhomogeneity of the wind
860 field results in significantly improved localization statistics when compared to Plume-based estimates. The Puff-based estimates also exhibit a closer-to-zero bias than the Plume-based estimates and reduced variance in the error distribution. Similarly, employing the more sophisticated MCMC-based inversion results in better localization and quantification estimates compared to the simple LSQ fitting. These differences are most stark when comparing the Puff MCMC (most complex) approach to the Plume LSQ (simplest). More specifically, when considering the localization statistics, the number of experiments for which
865 the algorithm correctly identifies the emission status of all of the 5 groups, $N_{(L=5)}$ in Table 1, increases from 85 to 184 (out of a total of 347 experiments) for the Plume LSQ and Puff MCMC approaches, respectively. Similarly, the mean quantification error improves from -0.1 kg/hr to 0.02 kg/hr, a 5-fold reduction in bias, while the fraction of estimates that were correct to within a factor of 2 (a commonly-used statistic to assess the variance of the error distribution) increases from 73% to 89%.

Under ideal CMS deployment (high sensor density and near-optimal placement), relatively simplified emissions scenarios
870 (constant emission rates, synchronous emissions events during each experiment), and a relatively simple (flat terrain, few obstructions) facility, quantification algorithms applied to data from point sensors can achieve low-bias emissions estimates, leading to accurate long-term estimates of total site emissions. While these systems can achieve near-zero bias, significant uncertainties remain in individual event-based emission rate estimates; the best-performing algorithm studied here (Puff MCMC) still had an average absolute relative error of 42%. Therefore, emissions estimates for any given short timeframe should be
875 interpreted with caution, considering that there is significant uncertainty associated with an individual estimate. However, for cumulative metrics, all models performed reasonably well: as shown in Table 1, the cumulative mass error for the Plume LSQ, Puff LSQ, Plume MCMC, and Puff MCMC were -135, -55, -117, and 12 kilograms, respectively, out of a total of 2,284 kg actually emitted, corresponding to percent errors in cumulative mass estimates of -6%, -2%, -5%, and 0.5%, respectively. This demonstrates that CMS systems, under the conditions present during this testing (sensor deployment and configuration, release
880 rates and patterns, environmental conditions) are capable of highly-accurate cumulative emission estimation, even when using lower-fidelity and simple models such as the Gaussian Plume and simple least-squares based rate inference.

This investigation provides further evidence and confirmation that advanced three-dimensional dispersion modeling approaches, e.g., the large-eddy simulation (LES) type companion CFD numerical experiments carried out in this investigation, can consistently predict more accurate time-varying concentration profiles than plume and puff models across a variety of surface
885 meteorological conditions. This however requires ‘nudging’ of the momentum and buoyancy transport equations to ensure agreement with the local observations of wind speed and direction. While it may be overly restricting, the spectral profiles presented in figure B2b depict a decent agreement with the spectral content of onsite anemometers. The current results show that while all three models underestimate the concentration field as indicated by FB, NMSE for the CFD model was 25% lower

than the puff model. More significantly, the CFD outperformed the puff and plume in 72% and 89% of selected experiments, respectively, in predicting the observed concentration traces. The results obtained via the CFD-based forward model coupled with the two inversion approaches for four of the selected experiments are equally encouraging. As evident in Figure 9, the CFD-based estimates of the inferred emission rates are notably closer to the parity line. Separately, both the absolute and relative error metrics in Table 3 show a marked improvement by the CFD MCMC combination over the other two approaches, with the mean error $\overline{E}$ of the test-aggregated emission rate showing a near 4-fold improvement. While there is room for improvement as it is a topic of active research and these simulations are certainly more expensive in terms of the computational cost, the CFD results discussed herein offer a proof-of-concept of employing such unsteady tools for use in conjunction with CMS networks on operational and experimental sites. Such advanced tools are expected to find increasing value in setups with numerous obstacles (e.g., power plants and compressor stations), undulating terrain, complex emission profiles, higher release rates, elevated release points and under scenarios where an operational site is only metered by 2-3 continuous monitoring sensors and may not even have an onsite anemometer, thus requiring the use of forecasting tools like the Weather Research and Forecasting (WRF) Model as a surrogate for onsite anemometer(s).

Crucially, the current study demonstrates the significant gains in quantification accuracy achievable with advanced emissions quantification methodologies using fixed-point continuous monitoring systems, particularly for long-timescale cumulative mass estimates, validating their potential for reliable facility-level emissions management.

905 **Appendix A: Extended Introduction to Measurement Informed Inventories and Methane Emissions Measurement** **And Quantification Techniques**

Recent studies employing various emissions measurement techniques indicate that existing emission inventories underestimate actual methane emissions from the oil and gas sector (Johnson et al. 2017; Alvarez et al. 2018; Robertson et al. 2020; Maasackers et al. 2021; Plant et al. 2022; Shen et al. 2022; Lu, Jacob, Zhang, et al. 2023; Tibrewal et al. 2024; Omara et al. 910 2024; Williams et al. 2024; Lu, Jacob, Wang, et al. 2022; Worden et al. 2022), with significant regional variations in the discrepancies between bottom-up inventories and measurement-based estimates (Ravikumar, Li, et al. 2025; Vallejo, Nguyen, and Ravikumar 2024). A recent study found methane emission loss rates ranging from approximately 0.75% to 9.63% of natural gas production across different regions of the US (Sherwin, Rutherford, Zhang, et al. 2022). Despite variations in emission rate distributions among different oil and gas basins, recent research indicates that the majority of methane emissions are originated 915 from facilities that emit at rates lower than 100 kg/hr (Williams et al. 2025).

Traditional approaches for detecting methane emissions often rely on human senses (auditory, visual, and olfactory (AVO) inspections) or portable sensors used in close proximity to potential sources. These include AVO inspections, EPA Method 21 (a sensitive but labor-intensive EPA-approved method), and optical gas imaging (OGI), which uses infrared cameras to visualize 920 methane leaks (Xia, Strayer, and Ravikumar 2024).

Advances in communication technologies, the Internet of Things (IoT), and reduced sensor costs have facilitated the development of next-generation emission measurement (NGEM) technologies. Unlike traditional methods, NGEM technologies operate at a distance from the source, attempting to remotely detect, localize, and quantify emission events. These technologies 925 operate from space (satellites), air (aircraft and UAVs), or the ground (fixed, mobile, and handheld sensors). Effective and efficient methane emission detection and quantification often requires a combination of measurement technologies, commonly referred to as a multiscale measurement approach (Wang et al. 2022; Daniels, Jia, and Hammerling 2024b).

Satellite and aerial remote sensing techniques can detect emissions from specific sources, with aerial methods capable of detecting emissions as low as 1-3 kg/hr, while satellites have minimum detection limits of to approximately 200 kg/hr or higher 930 (Sherwin, Rutherford, Chen, et al. 2023). Simulation studies based on the Fugitive Emissions Abatement Simulation Toolkit (FEAST), suggest that a minimum detection limit of 0.1–1 kg/hour is sufficient to capture almost all emissions from the oil and gas sector (Ravikumar, Wang, et al. 2018). Based on the results of this study, quantifying emissions below this threshold would not substantially improve mitigation efforts.

935

Satellite measurements offer remote detection of large methane releases, improving transparency and independent monitoring, particularly in remote or under-studied areas. Although, their high detection thresholds and infrequent overpasses limit their utility. Aerial platforms (piloted aircraft and UAVs) in general offer lower detection limits compared to satellites, but may introduce spatial extrapolation errors when regional inventory development is the objective. Because a considerable portion of methane emissions from oil and gas operations are intermittent, and snapshot methods only provide an instantaneous measurement (Santos et al. 2024), they are unable to provide robust estimates of the total mass emitted from a detected plume due to the inherent uncertainty in the required temporal extrapolation. In addition to higher detection limits, common limitations of all snapshot measurement techniques are the lack of site-specific meteorological data and information on emission event timing (e.g., duration and frequency of intermittent releases), which pose a significant challenge to the temporal extrapolation of results.

AERMOD (a Gaussian plume model) and CALPUFF (a Lagrangian puff model) are two air dispersion models commonly used for regulatory purposes. AERMOD is the EPA's recommended model for near-field applications (up to 50 km) (Houweling et al. 1999). This model relies on the assumption of steady-state conditions, which is often unrealistic in the presence of variable wind conditions (Jia, Daniels, and Hammerling 2023). Some of the previous studies employed longer time averaging of wind data to satisfy this steady-state requirement. However, this approach results in masking important short-term wind variations and associated concentration enhancements that are crucial for accurate dispersion modeling at scales relevant to oil and gas facilities.

CALPUFF has not been designated as an EPA-preferred model for near-field applications. However, it may be considered as an alternative dispersion modeling method on a case-by-case basis for near-field applications involving complex winds (Wayland 2008). CALPUFF is often preferred for long-range transport and complex terrain.

Both AERMOD and CALPUFF have limitations when applied to methane dispersion in near-source applications, such as upstream oil and gas. These limitations stem from the models' underlying assumptions and simplifications, which may not be appropriate in tens of meters-long source-to-sensor distances and lack accurate capturing of the complex atmospheric processes governing methane dispersion in upstream facilities. In a recent attempt, (Jia, Fish, et al. 2024) presented a computationally efficient and scalable implementation of the Gaussian puff model for atmospheric dispersion. This model incorporates dynamic spatiotemporal thresholding to achieve a shorter runtime, making it suitable for real-time applications and large-scale deployments.

Appendix B: Computational Fluid Dynamics

The choice of employing CFD simulations is motivated and constrained by two primary reasons: (i) a desire to perform three-dimensional (3D), unsteady modeling of multi-scale emissions from oil & gas facilities where the underlying wind field is directly modeled/resolved onto a computational grid while accounting for the complex effects induced by the presence of obstructions, that are otherwise not included in the previously discussed dispersion models; and (ii) to have a numerical framework that offers a graduated step up from the Gaussian models presented earlier while operating in effectively the same vein to permit a direct comparison of model performance.

The flow solver makes use of the finite-difference framework on a staggered Cartesian mesh and a fractional step approach, alternatively known as the predictor-corrector method, to solve the low-pass filtered incompressible Navier-Stokes equation:

$$\begin{aligned} \frac{\partial \tilde{u}_i}{\partial x_i} &= 0, \\ \frac{\partial \tilde{u}_i}{\partial t} + \frac{\partial \tilde{u}_i \tilde{u}_j}{\partial x_j} &= -\frac{1}{\rho_0} \frac{\partial \tilde{p}}{\partial x_i} - \frac{\partial \tau_{ij}}{\partial x_j} + b\delta_{i3}. \end{aligned} \quad (\text{B1})$$

Solution of the momentum equation requires inverting the following Poisson equation prior to the corrector step to enforce the divergence-free condition on the 3D velocity field:

$$\frac{\partial^2(\delta \tilde{p})}{\partial x_i^2} = \frac{\rho_0}{\Delta t} \frac{\partial \tilde{u}^*_i}{\partial x_i}. \quad (\text{B2})$$

Here, $x_i = (x, y, z)$ represents the horizontal (streamwise: x and spanwise: y) and the vertical z directions of the spatial coordinate system used in the CFD simulations. The asterisk on the right of Equation B2 indicates the intermediate velocity field obtained after the predictor step and does not satisfy the continuity equation. Time advancement is accomplished via a Newton-Raphson based iterative method and spatial derivatives in the momentum equation are discretized using central differences. The overall numerical approach is globally second-order accurate in both space and time. Further details on the numerical approach can be found in (Pierce 2001). Physical obstructions are mapped onto the Cartesian grid via an immersed boundary method (IBM) (Kim, Kim, and Choi 2001). The tool allows the ingestion of a digital elevation model to reproduce the undulating surface topography of complex industrial sites for use in CFD simulations. The tool is parallelized using domain decomposition coupled with the Message Passing Interface (MPI), and was employed successfully in the past to perform direct numerical simulations of both equilibrium and non-equilibrium turbulent channels flows and boundary layers over explicitly resolved rough surfaces on massively-parallel computing architectures (Ismail 2023).

In Equations B1 and B2, $\tilde{\cdot}$ represents the filtered fields whereas $u_i = (u, v, w)$, p and ρ are the instantaneous wind velocity, pressure and density, respectively. The subgrid stress τ_{ij} is modeled using the Smagorinsky approach: $\tau_{ij} \equiv \widetilde{u_i u_j} - \tilde{u}_i \tilde{u}_j =$
 995 $-2\nu_{SGS}\widetilde{S_{ij}}$. The linear constitutive relation between τ_{ij} and the filtered strain rate,

$$S_{ij} \equiv \frac{1}{2} \left(\frac{\partial \tilde{u}_i}{\partial x_j} + \frac{\partial \tilde{u}_j}{\partial x_i} \right) \quad (\text{B3})$$

embedded in the Smagorinsky approach is modulated by the eddy viscosity ν_{SGS} . The eddy viscosity is related to the local characteristic length and time scales using the following mixing-length type formulation: $\nu_{SGS} = (C_S \Delta)^2 |\tilde{S}|$, where $|\tilde{S}| = \sqrt{2\widetilde{S_{ij}S_{ij}}}$ and Δ is local characteristic length scale taken as $\Delta \equiv (\text{vol})^{1/3}$ (vol = local grid cell volume). We deter-
 1000 mine the Smagorinsky constant C_S using the dynamic procedure proposed by (Lilly 1992). A wall model is included that promotes the modulation of the boundary layer at high Reynolds numbers due to surface roughness. The wall model, following (Mukha, Rezaeiravesh, and Liefvendahl 2019), imposes a no-slip condition for the horizontal velocity components and forces the computed wall shear stress by requiring the eddy viscosity be determined by,

$$\nu_{SGS} = \frac{u_*^2}{[(\widetilde{u^P}/h_p)^2 + (\widetilde{v^P}/h_p)^2]^{1/2}} \quad (\text{B4})$$

1005 The superscript p in equation B4 refers to the cell at a wall-normal distance h_p from the surface. The filtered horizontal velocity field at this height is used to estimate the local wall shear stress u_*^2 in the above equation using the ‘law of the wall’ for fully rough surfaces: $u_*^{\text{est}} = \kappa \sqrt{\widetilde{u^2} + \widetilde{v^2}} / \ln(h_p/z_0)$. The METEC facility is approximated as belonging to the open-terrain category and thus the roughness length z_0 is chosen as 0.03 m. The von-Karman constant in the law of the wall is 0.4.

1010 Contaminants like methane are treated as passive tracers and their transport is modeled by the following filtered advection-diffusion equations:

$$\frac{\partial \tilde{c}}{\partial t} + \frac{\partial \tilde{c} \tilde{u}_j}{\partial x_j} = \frac{\partial \pi_j}{\partial x_j} + S_c. \quad (\text{B5})$$

Spatial derivatives in equation B5 are discretized by the Quadratic Upstream Interpolation for Convective Kinematics (QUICK) scheme. In the filtered scalar transport equation (Equation B5), $S_c(t)$ represents the point sources inside the fluid domain, which
 1015 are approximated as 3D Gaussian functions following the approach by (Ražnjević, Heerwaarden, and Krol 2022). While this equation is written for a single tracer, the tool permits the inclusion of a user-specified number of scalar fields all of which, in principle, can have point sources of varying strength placed at arbitrary positions. The closure term due to the subgrid

scalar flux π_j is parametrized using the following gradient-diffusion procedure: $\pi_j \equiv \widetilde{c}u_j - \widetilde{c}\widetilde{u}_j = -\nu_{SGS}/Sc_T(\partial\widetilde{c}/\partial x_j)$. The turbulent Schmidt number Sc_T is fixed at 0.7 (Qian and Venkatram 2011).

$$1020 \quad \frac{\partial b}{\partial t} + \frac{\partial b\widetilde{u}_j}{\partial x_j} = \alpha_b \frac{\partial^2 b}{\partial x_j^2} \quad (B6)$$

The effect of thermal buoyancy is modeled via the Boussinesq Approximation and the equation for conservation of energy is instead replaced by the buoyancy variable: $b \equiv (\theta'_v/\theta_{v0})g$ (Van Heerwaarden et al. 2017). Here, θ'_v , θ_{v0} and g are the fluctuating potential temperature, the reference virtual reference potential temperature and the acceleration due to gravity. The diffusivity coefficient for the buoyancy field α_b is treated like Sc_T .

1025

No-slip and impermeability conditions are applied on the bottom surface for the horizontal and vertical components of the wind field, respectively. A zero-flux condition is used for the scalar fields at the lower surface. Periodic and zero-flux boundary conditions are imposed on the lateral boundary surfaces for velocity and scalar fields, respectively.

1030 As our focus is on near-surface pollutant dispersion, we do not construct the entire boundary layer. Instead, the ceiling of the domain is capped at 200 m, and a no-flux condition is imposed at this height. The horizontal mesh is uniformly spaced with a resolution of 1.5mx1.5m. The mesh is stretched vertically to accommodate the ground-imposed anisotropy: the grid spacing at the ground is 0.2 m, and is gradually relaxed to 2 m towards the top boundary, with the maximum expansion ratio between successive grid points in the z direction remaining below 1.10.

1035

We further leverage the periodicity and uniform grid spacing in the horizontal direction to perform a discrete Fourier transform of the Poisson equation (Equation B2) into a set of linearly independent 1D modified Helmholtz equations that are then solved using the tridiagonal algorithm. Figure B1a shows the 3D digital elevation model (DEM) of the METEC facility after it is mapped onto the Cartesian grid. The positions of the virtual sensors are highlighted using purple spheres. It should be noted that the METEC site exhibits an effectively flat surface, that justifies its earlier classification as open terrain: the only significant obstructions are the three tanks at the center of the facility that have a height of $\sim 4m$.

Preliminary simulations were performed to dissect the signal detected at virtual sensor stations both with the effect of the DEM included and without it, and found negligible value in modeling the tanks as obstacles for the present scenario. The downwash effects presumably induced by the obstacles typically extend less than 5 times the mean obstacle height (LEONARDI et al. 2003). The closest sensors to the tanks are however at least over 20m from the centroid of the tanks. Regardless, the simulations

1045

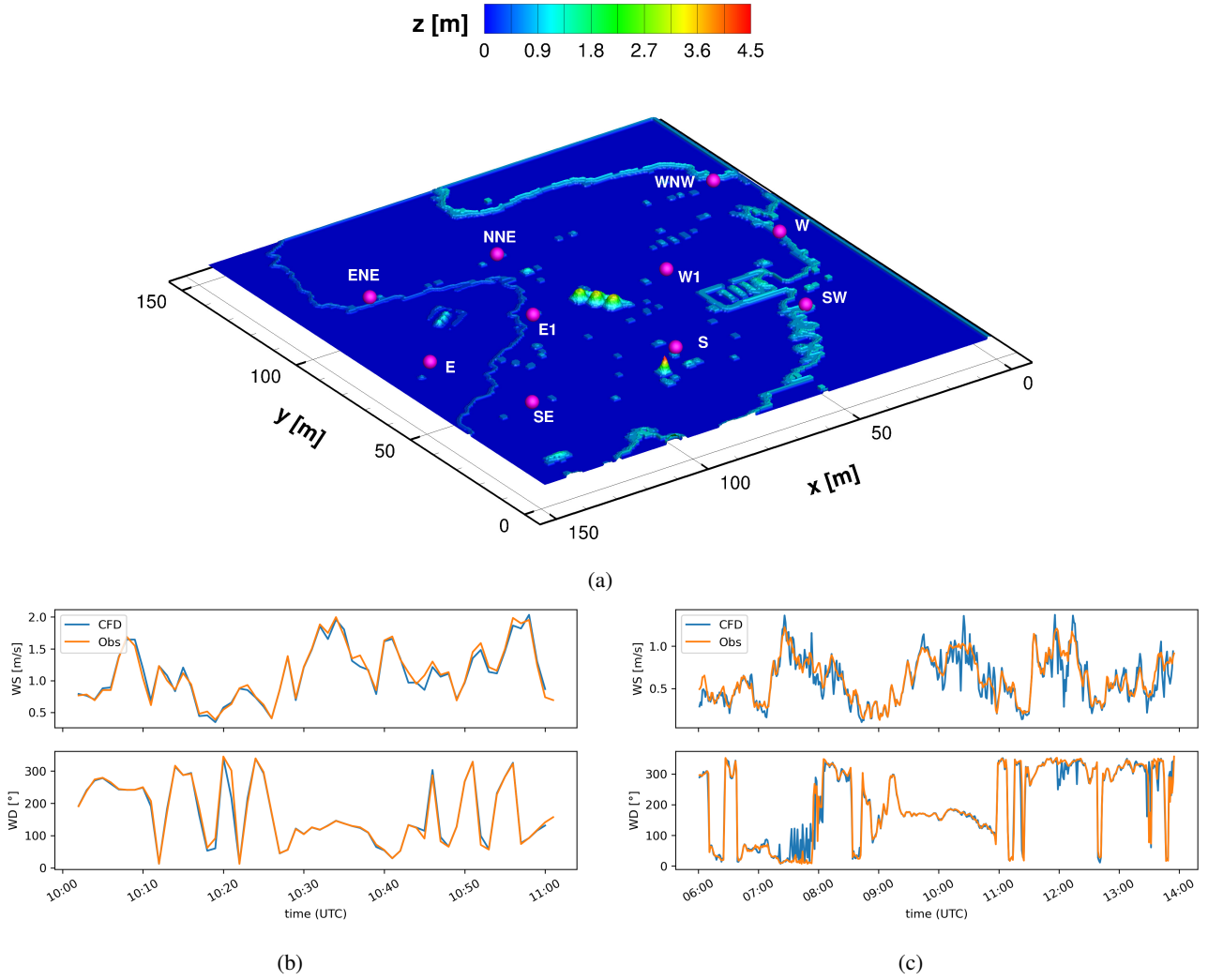


Figure B1. (a) Visualizing the 3D surface terrain after mapping the digital elevation model of the facility onto the CFD mesh. The color map indicates the local elevation z in meters and the magenta spheres identify the position of the point sensors. (b,c) Comparison of the minute-averaged computed wind speed and direction at the virtual NE sensor height from the CFD simulations with onsite observations for two experiments: one from (b) Feb. 08, 2024 and another from (c) April. 11, 2024.

presented later in this investigation include the effect of the 3D DEM.

A novelty of these CFD simulations rests in the use of facility-specific wind measurements collected by an on-site anemometer to reconstruct the horizontally-averaged boundary layer profile. This in turn is employed to interpolate the horizontal forcing terms needed at each time step on the right-hand side of the u and v components of the momentum equation and the buoyancy

transport equation. This approach bears some resemblance to the method used by (Wiersema, Lundquist, and Chow 2020) for their forced microscale simulations. The on-site measurements by the anemometer are combined with the local estimate of the Obhukov lengthscale L provided by the High-Resolution Rapid Refresh (HRRR) atmospheric forecasting model (Dowell et al. 2022). Vertically varying mean velocity and thermal buoyancy profiles mapped to the parametric equations of the Monin-Obhukov similarity theory (MOST) are constructed using L and onsite measurements referenced earlier at the sensor height as inputs. The MOST matching profiles that are used to infer the forcing term are periodically updated during the course of the simulation using time-evolving onsite information. This allows the simulated boundary layer to respond to changing on-site conditions while still retaining the resolved component acquired from solving the filtered momentum and buoyancy transport equations.

The time traces of the instantaneous and minute-averaged horizontal wind speed ($WS = \sqrt{U^2 + V^2}$) and wind direction ($WD = 270^\circ - \tan^{-1}(V/U)$) on the virtual north-east (NE) sensor from simulations of two different cases are compared with onsite measurements in Figures B1b and B1c. It is evident that the simulations are able to faithfully reproduce the varying wind field as measured by onsite monitors. The wall-normal variation of the mean wind speed for three selected test cases is presented in Figure B2a. The three cases are chosen such that each of them falls within the stable (red), neutral (blue) and unstable (green) regimes based on extracted L from the HRRR model. The gray profiles represent several instantaneous transects extracted from the simulations. The qualitative trend of the vertical variation in mean wind speed reproduced by the simulations under different atmospheric stability regimes compares favorably with the results reported by (“Monin-Obukhov similarity theory and its application to wind flow modelling over complex terrain” 2018).

To further demonstrate that the CFD simulations are indeed reproducing eddies representative of the expected turbulence cascade, Figure B2b compares the mean one-dimensional power spectrum E_{WS} of the horizontal wind speed for the neutral test case presented in Figure B2a. Since our simulations are performed using a constant time step of $\Delta t = 0.1s$, the WS at the virtual sensor locations is averaged up to the resolution of $1Hz$ before computing the power spectrum to allow a one-to-one comparison with the power spectrum computed via onsite measurements. Figure B2b identifies a five-order reduction in energy across the range of frequencies captured in this simulation case. The presence of the turbulence cascade with a distinct inertial subrange of scales was identified using Kolmogorov’s -5/3 power law (Pope 2000). The inset of Figure B2b presents the premultiplied power spectrum fE_{WS} for the same simulation. There is a decent qualitative correspondence with the fE_{WS} reported by (Schalkwijk, Jonker, and Siebesma 2014) from their year-long LES of an ABL. Specifically, the range of frequencies that can be classified as being part of the so-called spectral gap extends below the 1-minute level, which is typically above the general upper limit of frequencies typically assumed for the validity spectral gap. Additionally, there is a precipitous drop in spectral power for the range of scales above 0.05Hz, which is in agreement with the simulation by (Schalkwijk, Jonker, and Siebesma 2014). These features in the CFD simulations give confidence that the more salient pieces of flow physics relevant

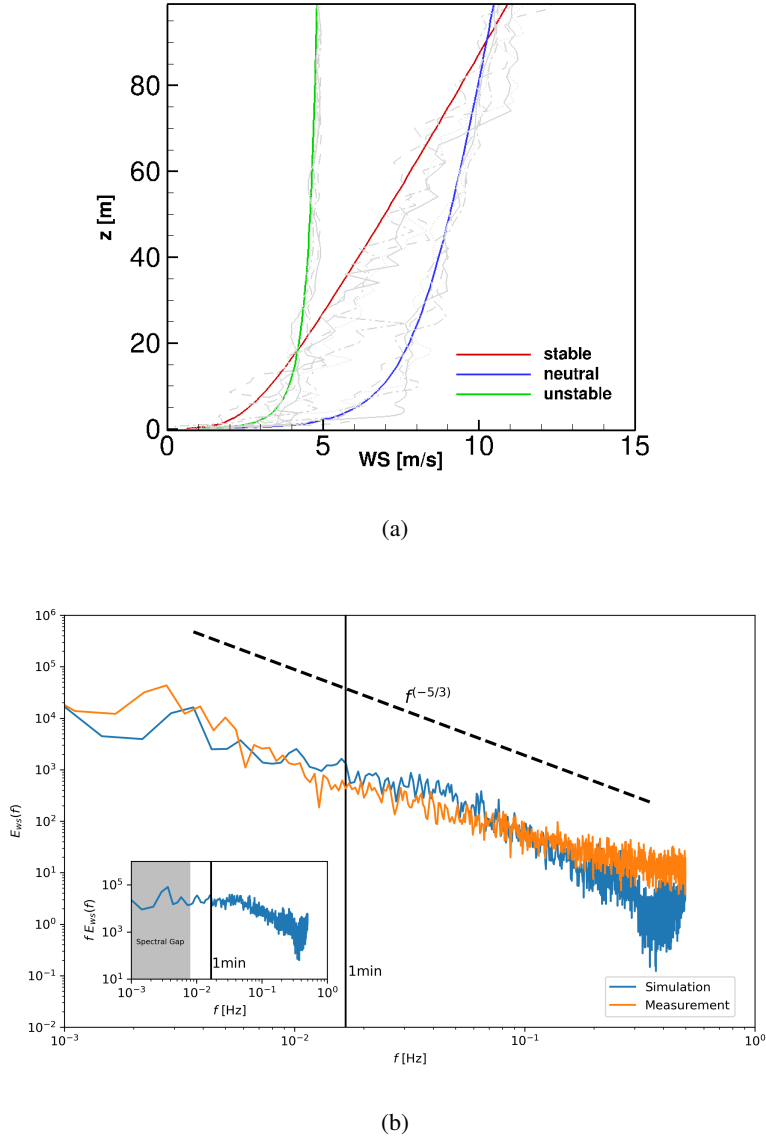


Figure B2. (a) Variation of the mean wind speed in the vertical z direction for three representative cases from the CSU METEC study dataset: stable (red), neutral (blue) and unstable (green). Gray lines indicate instantaneous traces for each of the three cases. (b) Power spectrum profiles of the horizontal wind speed from the simulation for the neutral case in (a) compared against spectral profiles of the measured on-field data at 2m height. The inset shows the premultiplied power spectrum for this simulation case.

1085 to the turbulent dispersion of gas are being captured numerically.

Appendix C: Evaluation of Forward Model Accuracy

In general, it is expected that the more accurate forward model for concentration predictions should result in the best quantification estimates. This Appendix tests whether this is indeed the case. To directly compare the accuracy of the three dispersion modeling techniques (Plume, Puff, CFD), each forward model is applied to actual release rates and locations. Predicted and measured concentrations are then compared using several error metrics. A random set of 60 experiments are chosen for this analysis. This limited subset is used due to the high computational cost of running CFD against all 347 experiments.

C1 Metrics

In order to compare the predicted concentration estimates against the measurement data, we apply the following performance metrics suggested by Chang and Hanna 2004 for this exact purpose. More specifically, the normalized mean squared error (NMSE), fractional bias (FB), geometric mean bias (MG), geometric variance (VG), normalized standard deviation (NSD), and the fraction of predictions within a factor of x (Fac x). Mathematically, these are expressed as

$$NMSE = \frac{\overline{(X_o - X_p)^2}}{X_o X_p}, \quad (C1)$$

$$FB = \frac{2(\overline{X_o} - \overline{X_p})}{(\overline{X_o} + \overline{X_p})}, \quad (C2)$$

$$MG = \exp(\overline{\ln X_o} - \overline{\ln X_p}), \quad (C3)$$

$$VG = \exp(\overline{(\ln X_o - \ln X_p)^2}), \quad (C4)$$

$$NSD = \frac{\sigma_p}{\sigma_o}. \quad (C5)$$

In these equations, X_o represents the observed concentration, X_p refers to the predicted concentrations (computed via $\mathbf{S}\mathbf{Q}'$, where $\mathbf{S}$ is the sensitivity matrix associated with a given model and $\mathbf{Q}'$ is the actual rate vector), and σ is the standard deviation of either the observed or predicted values. Fac x is simply the fraction of predicted values that are within a factor of x of the measurements. In order to compensate for the fact that a few outliers can significantly impact the NMSE and FB, MG and VG

are introduced using logarithmically-scaled data to offer error statistics that are less sensitive to a small number of outliers. For the purposes of computing MG and VG, measurement/prediction concentration pairs where at least one of the values is less than 1 ppm are removed from the calculation. This prevents contamination of mean MG and VG by predictions/measurements by sensors upwind of the source or under conditions when the predictions/measurements are near the instrument sensitivity. In other words, the MG and VG statistics are computed only on prediction/measurement pairs for which both “agree” that the sensor in question is receiving an appreciable amount of signal from a source. This helps balance the statistics from being skewed by an overabundance of $X_o = 0, X_p = 0$ cases, which are the majority of sensor readings.

C2 Results

The results of applying all three forward models to the same subset of 60 experiments and computing the metrics defined in C1 are shown in Figure C1.

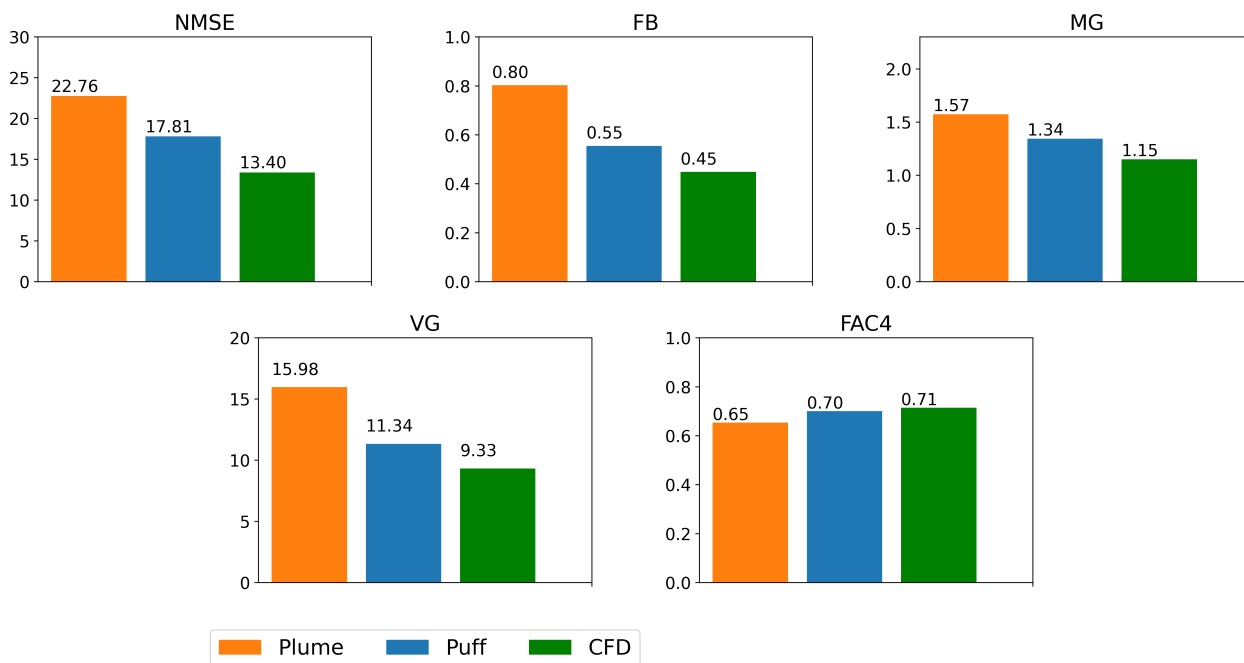


Figure C1. Comparison of the mean measures of the model skill for the Gaussian plume (orange), Gaussian puff (blue) and CFD (green) techniques for the selected 60 experiments.

It is evident considering the results in Figure C1 that increasingly complex dispersion modeling (Plume, Puff, CFD, in that order) results in uniformly and monotonically improving error metrics. In other words, it is the case for each of these 5 metrics that the Puff outperforms Plume, but is outperformed by CFD. More specifically, the NMSE for the CFD model shows an improvement of 45% and 25% over the plume and puff models, respectively. To further illustrate this, Figure C2a shows the ratio of $NMSE_{cf}/NMSE_{puff}$ for the 60 experiments, and demonstrates that the CFD outperforms the Puff model in terms

of NMSE for 72% of cases and a lower MG in 63% and 86%, of the experiments when compared to the puff and plume models, respectively. A large ($\gg 1$) NMSE indicates that the associated error distribution is more likely to be lognormal than normal (Chang and Hanna 2004)). The higher values of geometric variance (VG) for the plume model indicate that the lognormal distribution of the error has comparatively much larger scatter than both puff and CFD, which show much more comparable variance in error, as is evident in Figure C2a, showing that 95% of the experiments have $NMSE_{\text{cfd}}/NMSE_{\text{puff}}$ within a factor of 2.

Both fractional bias (FB) and mean geometric bias (MG), are indicators of systematic error. Considering these metrics across all dispersion models, we see that the models generally underestimate the measured concentrations, although the magnitude of the bias decreases with increasing complexity of the forward model. The systematic underestimation of concentrations across all three approaches indicates a combination of two potential underlying issues: first, that the simulated plumes are overly diffusive, resulting in smaller than actual concentrations. Second, it is also possible that there are more complex wind dynamics that are not being properly captured by any of the dispersion models which could result in concentration enhancements on sensors that the dispersion models think are being “missed” by the plume. Either one of these effects could plausibly cause a systematic underestimation of concentrations. With this said, it is evident that the CFD is more properly capturing salient features in the plume physics that result in substantially better concentration predictions, however there is still significant room for improvement.

In order to further understand the predictive capabilities of the CFD dispersion model under different atmospheric stability classes, we stratified the experiments by stability class and computed the error metrics in each ASC. In general, we found that CFD consistently outperforms the puff model for experiments falling in the unstable and neutral ASC regimes. The relative performance of the CFD is the poorest for the stable regime: CFD only outperformed the Puff in 54% of these cases. It is worth noting that while the dispersion profiles in the plume and puff models are determined by the ASC alone, the structures produced by CFD simulations are far more complex and depend on interactions between particular details of the modeling approach including the local grid resolution, the dynamic balance between thermal and momentum forcing, and potentially many other factors. As such, it is difficult to pin down the exact physical reasons why the CFD modeling results in such an improvement.

Of all the metrics shown in Figure C1, FAC4 shows the lowest progressive improvement from plume to puff to CFD; in fact, the FAC4 value for the CFD (0.71) is nearly identical to that for puff (0.70). This is due to the well recognized challenge of combining observed and simulated plumes both in time and space. As Weil, Sykes, and Venkatram 1992 note, variations in wind direction of 20-40 degrees may result in complete failure of overlap of simulated and observed plume despite manifesting similar magnitudes and shapes. Nevertheless, the current FAC4 results show a favorable comparison with the FAC2 and FAC5 numbers reported by Wiersema, Lundquist, and Chow 2020 from their microscale and multiscale CFD simulations of the Join Urban 2003 experiment, which modeled the single-point release of a tracer in an urban environment. The relative comparison across the following three performance metrics: NSD, correlation coefficient R and normalized root mean square error (NRMSE), which are related by the law of cosines, can be performed using the single diagram method of Taylor 2001. An

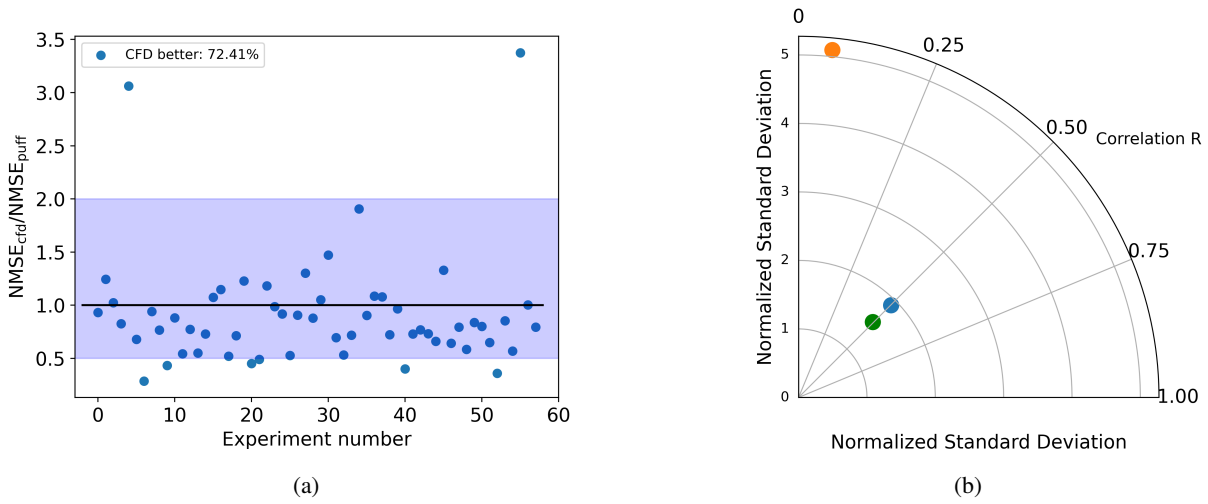


Figure C2. (a) $NMSE_{cfd}/NMSE_{puff}$ from the experimental dataset for CFD. The shaded region indicates the range of cases within a factor of 2 of the scenario with $NMSE=1$. (b) Visualizing the model performance metrics via the single nomogram method proposed by Taylor 2001. Each dot corresponds to the mean value across all experiments used in the CFD dataset. Colors in (b) follow the labeling identified in figure C1.

ideal case for these metrics would be located on the horizontal axis at the origin of the polar plot shown in Figure C2b. For the current experiments, it is clear that both puff and CFD vastly outperform the plume model, the CFD slightly edging out puff with near identical correlation coefficient R but an appreciably lower NSD.

Appendix D: Relative Error Distribution Associated With Quantification Methods

1160 Figure D1 shows box-and-whisker plots of the relative error distribution associated with each quantification method. In this Figure, the orange lines show the median of the relative error distribution while the blue-shaded region shows the two inner quartile regions (IQR) of the distribution, encompassing the middle 50% of the data. The whiskers indicate the first and third quartiles, extending to 1.5 times the IQR, and the circles denote data points outside this range, representing outliers. Several key features are apparent in this figure. First, the Plume-based quantification methods (first and third columns) show a tendency
1165 toward lower median relative errors suggesting a more pronounced low bias compared to Puff-based models. This observation is consistent with the $\overline{E}$ and ΔC statistics represented in Table 1. Second, the IQR narrows when transitioning from the LSQ inversion methods to MCMC, indicating less scatter in the relative error distributions for the MCMC inversion method. This trend is also reflected in the $F2$ statistic from Table 1. Finally, the combination of Puff MCMC has a smaller upper range of percent errors, with the maximum relative error being around 300%, compared to all of the other methods that have outlying
1170 relative error points above 400%.

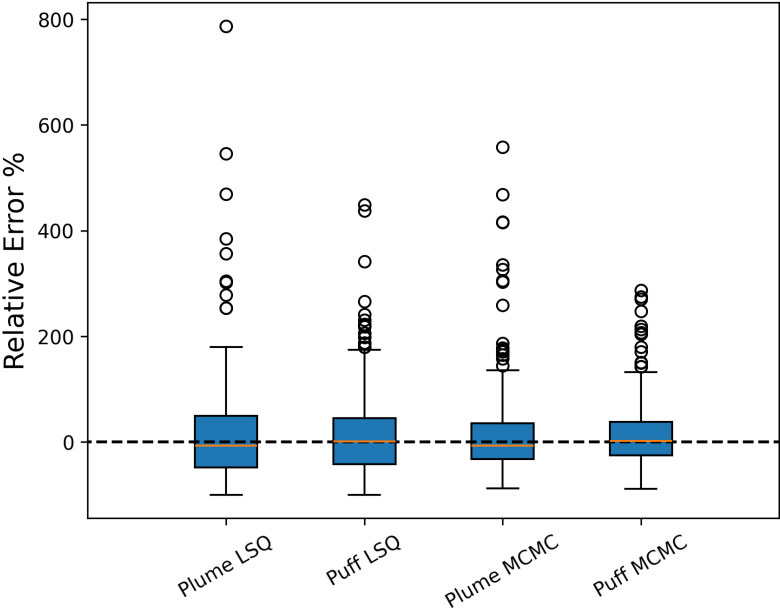


Figure D1. Box and whisker plots of relative error distributions for 4 different quantification methods.

Data availability. Data for selected experiments is available upon request.

Author contributions. All authors contributed to conceptualizing the project; DRB and UI developed and implemented the models, and analyzed the data; DRB, AL, and UI, wrote the manuscript; all authors revised and edited the manuscript.

1175 *Competing interests.* The authors declare that they have no conflict of interest.

Acknowledgements. We acknowledge Project Canary for support. We thank the Colorado State University (CSU) Methane Emission Technology Evaluation Center (METEC) for the data collection as part of the Advancing Development of Emissions Detection (ADED) project, funded by the Department of Energy's National Energy Technology Laboratory (NETL).

References

- 1180 Allen, David T, Daniel Zimmerle, and John Dabbar (2024). "Charting the Course: The National Petroleum Council's Advice on the Role of Emission Detection and Quantification in Reducing Greenhouse Gas Emissions from the US Natural Gas Supply Chain". In: *ACS Sustainable Resource Management* 1.6, pp. 1035–1036.
- Allwine, K Jerry, Walter F Dabberdt, and Larry L Simmons (1998). "Peer review of the CALMET/CALPUFF modeling system". In: *EPA contract N 68-D-98 92*, pp. 1–03.
- 1185 Alvarez, Ramón A. et al. (2018). "Assessment of methane emissions from the U.S. oil and gas supply chain". In: *Science* 361.6398, pp. 186–188. DOI: 10.1126/science.aar7204. eprint: <https://www.science.org/doi/pdf/10.1126/science.aar7204>. URL: <https://www.science.org/doi/abs/10.1126/science.aar7204>.
- Ball, David, Nathan Eichenlaub, and Ali Lashgari (2025). "Performance Evaluation of Fixed-Point Continuous Monitoring Systems: Influence of Averaging Time in Complex Emission Environments". In: *Sensors* 25.9, p. 2801.
- 1190 Bell, Clay et al. (2023). "Performance of Continuous Emission Monitoring Solutions under a Single-Blind Controlled Testing Protocol". In: *Environmental Science & Technology* 57.14. PMID: 36977200, pp. 5794–5805. DOI: 10.1021/acs.est.2c09235. eprint: <https://doi.org/10.1021/acs.est.2c09235>. URL: <https://doi.org/10.1021/acs.est.2c09235>.
- Carruthers, D.J., R.J. Holroyd, et al. (1994). "UK-ADMS: A new approach to modelling dispersion in the earth's atmospheric boundary layer". In: *Journal of Wind Engineering and Industrial Aerodynamics* 52, pp. 139–153. ISSN: 0167-6105. DOI: [https://doi.org/10.1016/0167-6105\(94\)90044-2](https://doi.org/10.1016/0167-6105(94)90044-2). URL: <https://www.sciencedirect.com/science/article/pii/0167610594900442>.
- 1195 Carruthers, DJ, WS Weng, et al. (2009). "Plume/puff spread and mean concentration module specifications". In: *ADMS 2 Technical Specification* 3.
- Chang, J. C. and S. R. Hanna (Oct. 2004). "Air quality model performance evaluation". In: *Meteorology and Atmospheric Physics* 87 (1-3), pp. 167–196. ISSN: 01777971. DOI: 10.1007/s00703-003-0070-7.
- 1200 Chen, Qining, Yosuke Kimura, and David T Allen (2024). "Defining Detection Limits for Continuous Monitoring Systems for Methane Emissions at Oil and Gas Facilities". In: *Atmosphere* 15.3, p. 383.
- Chen, Qining, Mrinali Modi, et al. (2022). "Simulated methane emission detection capabilities of continuous monitoring networks in an oil and gas production region". In: *Atmosphere* 13.4, p. 510.
- Chen, Qining, Colette Schissel, et al. (2023). "Assessing detection efficiencies for continuous methane emission monitoring systems at oil and gas production sites". In: *Environmental Science & Technology* 57.4, pp. 1788–1796.
- 1205 Cheptonui, Fancy et al. (2025). "Assessing the Performance of Emerging and Existing Continuous Monitoring Solutions under a Single-blind Controlled Testing Protocol". In.
- Cimorelli, Alan J et al. (2005). "AERMOD: A dispersion model for industrial source applications. Part I: General model formulation and boundary layer characterization". In: *Journal of Applied Meteorology and Climatology* 44.5, pp. 682–693.
- 1210 Daniels, William, Meng Jia, and Dorit Hammerling (May 2024a). "Estimating methane emission durations using continuous monitoring systems". In: *ChemRxiv*. DOI: 10.26434/chemrxiv-2024-f921j. eprint: <https://chemrxiv.org/engage/chemrxiv/article-details/6633e52891aefa6ce1ffaf3b>. URL: <https://doi.org/10.26434/chemrxiv-2024-f921j>.
- Daniels, William S, Meng Jia, and Dorit M Hammerling (2024b). "Detection, localization, and quantification of single-source methane emissions on oil and gas production sites using point-in-space continuous monitoring systems". In: *Elementa: Science of the Anthropocene* 12.1.
- 1215

- Dowell, David C. et al. (2022). “The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part I: Motivation and System Description”. In: *Weather and Forecasting* 37.8, pp. 1371–1395. DOI: 10.1175/WAF-D-21-0151.1. URL: <https://journals.ametsoc.org/view/journals/wefo/37/8/WAF-D-21-0151.1.xml>.
- Dunson, D B and J E Johndrow (Dec. 2019). “The Hastings algorithm at fifty”. In: *Biometrika* 107.1, pp. 1–23. ISSN: 0006-3444. DOI: 1220 10.1093/biomet/asz066. eprint: <https://academic.oup.com/biomet/article-pdf/107/1/1/32450930/asz066.pdf>. URL: <https://doi.org/10.1093/biomet/asz066>.
- Gosse, Colleen (2023). “PERFORMANCE EVALUATION OF A LOW-COST SENSOR FOR CONTINUOUS METHANE MONITORING.” PhD thesis. Dalhousie University.
- Hollenbeck, Derek, Demitrius Zulevic, and Yangquan Chen (2021). “Advanced leak detection and quantification of methane emissions using sUAS”. In: *Drones* 5.4, p. 117. 1225
- Houweling, Sander et al. (1999). “Inverse modeling of methane sources and sinks using the adjoint of a global transport model”. In: *Journal of Geophysical Research: Atmospheres* 104.D21, pp. 26137–26160.
- IJzermans, Rutger et al. (2024). “Long-term continuous monitoring of methane emissions at an oil and gas facility using a multi-open-path laser dispersion spectrometer”. In: *Scientific Reports* 14.1, p. 623.
- Ilonze, Chiemezie et al. (2024). “Assessing the Progress of the Performance of Continuous Monitoring Solutions under a Single-Blind Controlled Testing Protocol”. In: *Environmental Science & Technology* 58.25. PMID: 38865299, pp. 10941–10955. DOI: 10.1021/acs.est.3c08511. eprint: <https://doi.org/10.1021/acs.est.3c08511>. URL: <https://doi.org/10.1021/acs.est.3c08511>. 1230
- Ismail, Umair (2023). “Direct Numerical Simulation of a Turbulent Boundary Layer Encountering a Smooth-to-Rough Step Change”. In: *Energies* 16.4. ISSN: 1996-1073. DOI: 10.3390/en16041709. URL: <https://www.mdpi.com/1996-1073/16/4/1709>.
- Jia, Meng, William Daniels, and Dorit Hammerling (2023). “Comparison of the Gaussian plume and puff atmospheric dispersion models for methane modeling on oil and gas sites”. In. 1235
- Jia, Meng, Ryker Fish, et al. (2024). “Filling a critical need: a lightweight and fast Gaussian puff model implementation”. In.
- Johnson, Matthew R. et al. (2017). “Comparisons of Airborne Measurements and Inventory Estimates of Methane Emissions in the Alberta Upstream Oil and Gas Sector”. In: *Environmental Science & Technology* 51.21. PMID: 29039181, pp. 13008–13017. DOI: 10.1021/acs.est.7b03525. eprint: <https://doi.org/10.1021/acs.est.7b03525>. URL: <https://doi.org/10.1021/acs.est.7b03525>. 1240
- Karion, Anna et al. (2019). “Intercomparison of atmospheric trace gas dispersion models: Barnett Shale case study”. In: *Atmospheric chemistry and physics* 19.4, pp. 2561–2576.
- Kim, Jungwoo, Dongjoo Kim, and Haecheon Choi (2001). “An Immersed-Boundary Finite-Volume Method for Simulations of Flow in Complex Geometries”. In: *Journal of Computational Physics* 171.1, pp. 132–150. ISSN: 00219991. DOI: 10.1006/jcph.2001.6778.
- Kumar, Pramod et al. (2022). “Near-field atmospheric inversions for the localization and quantification of controlled methane releases using stationary and mobile measurements.” In: *Quarterly Journal of the Royal Meteorological Society* 148.745. 1245
- LEONARDI, S. et al. (2003). “Direct numerical simulations of turbulent channel flow with transverse square bars on one wall”. In: *Journal of Fluid Mechanics* 491, pp. 229–238. DOI: 10.1017/S0022112003005500.
- Lilly, D. K. (Mar. 1992). “A proposed modification of the Germano subgrid-scale closure method”. In: *Physics of Fluids A: Fluid Dynamics* 4.3, pp. 633–635. ISSN: 0899-8213. DOI: 10.1063/1.858280. eprint: https://pubs.aip.org/aip/pof/article-pdf/4/3/633/12443162/633_1_online.pdf. URL: <https://doi.org/10.1063/1.858280>. 1250
- Lu, X., D. J. Jacob, H. Wang, et al. (2022). “Methane emissions in the United States, Canada, and Mexico: evaluation of national methane emission inventories and 2010–2017 sectoral trends by inverse analysis of in situ (GLOBALVIEWplus CH4 ObsPack) and satellite

- (GOSAT) atmospheric observations”. In: *Atmospheric Chemistry and Physics* 22.1, pp. 395–418. DOI: 10.5194/acp-22-395-2022.
 1255 URL: <https://acp.copernicus.org/articles/22/395/2022/>.
- Lu, Xiao, Daniel J. Jacob, Yuzhong Zhang, et al. (2023). “Observation-derived 2010–2019 trends in methane emissions and intensities from US oil and gas fields tied to activity metrics”. In: *Proceedings of the National Academy of Sciences* 120.17, e2217900120. DOI: 10.1073/pnas.2217900120. eprint: <https://www.pnas.org/doi/pdf/10.1073/pnas.2217900120>. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.2217900120>.
- 1260 Maasakkers, J. D. et al. (2021). “2010–2015 North American methane emissions, sectoral contributions, and trends: a high-resolution inversion of GOSAT observations of atmospheric methane”. In: *Atmospheric Chemistry and Physics* 21.6, pp. 4339–4356. DOI: 10.5194/acp-21-4339-2021. URL: <https://acp.copernicus.org/articles/21/4339/2021/>.
- Martin, Delance O. (1976). “Comment On ‘The Change of Concentration Standard Deviations with Distance’”. In: *Journal of the Air Pollution Control Association* 26.2, pp. 145–147. DOI: 10.1080/00022470.1976.10470238. eprint: <https://doi.org/10.1080/00022470.1976.10470238>.
 1265 URL: <https://doi.org/10.1080/00022470.1976.10470238>.
- Mohan, Manju and T.A. Siddiqui (1998). “Analysis of various schemes for the estimation of atmospheric stability classification”. In: *Atmospheric Environment* 32.21, pp. 3775–3781. ISSN: 1352-2310. DOI: [https://doi.org/10.1016/S1352-2310\(98\)00109-5](https://doi.org/10.1016/S1352-2310(98)00109-5). URL: <https://www.sciencedirect.com/science/article/pii/S1352231098001095>.
- “Monin-Obukhov similarity theory and its application to wind flow modelling over complex terrain” (2018). In: *Journal of Wind Engineering and Industrial Aerodynamics* 182, pp. 308–321. ISSN: 01676105. DOI: 10.1016/j.jweia.2018.09.026.
 1270
- Mukha, T., S. Rezaeiravesh, and M. Liefvendahl (2019). “A library for wall-modelled large-eddy simulation based on OpenFOAM technology”. In: *Computer Physics Communications* 239, pp. 204–224. ISSN: 00104655. DOI: 10.1016/j.cpc.2019.01.016. arXiv: 1807.11786.
- Omara, M. et al. (2024). “Constructing a measurement-based spatially explicit inventory of US oil and gas methane emissions”. In: *Earth System Science Data Discussions* 2024, pp. 1–25. DOI: 10.5194/essd-2024-72. URL: <https://essd.copernicus.org/preprints/essd-2024-72/>.
- 1275 Pasquill, Frank (1961). “The estimation of the dispersion of windborne material”. In: *Meteoro. Mag.* 90, pp. 20–49.
- Peischl, Jeff et al. (2016). “Quantifying atmospheric methane emissions from oil and natural gas production in the Bakken shale region of North Dakota”. In: *Journal of Geophysical Research: Atmospheres* 121.10, pp. 6101–6111.
- Pierce, Charles David (Jan. 2001). “Progress-variable approach for large-eddy simulation of turbulent combustion”. PhD thesis. Stanford University, California.
- 1280 Plant, Genevieve et al. (2022). “Inefficient and unlit natural gas flares both emit large quantities of methane”. In: *Science* 377.6614, pp. 1566–1571. DOI: 10.1126/science.abq0385. eprint: <https://www.science.org/doi/pdf/10.1126/science.abq0385>. URL: <https://www.science.org/doi/abs/10.1126/science.abq0385>.
- Pope, Stephen B. (2000). *Turbulent Flows*. Cambridge University Press.
- Qian, Wenjun and Akula Venkatram (Mar. 2011). “Performance of Steady-State Dispersion Models Under Low Wind-Speed Conditions”.
 1285 In: *Boundary-Layer Meteorology* 138.3, pp. 475–491. DOI: 10.1007/s10546-010-9565-1.
- R. P. Hosker, Jr. (1974). “A Comparison of Estimation Procedures for Over-Water Plume Dispersion”. In: *Symposium on Atmospheric Diffusion and Air Pollution held September 9-13, 1974 in Santa Barbara, California*. Vol. 60. 38, pp. 669–669. DOI: 10.1029/eo060i038p00669-03.
- Randerson, D. (July 1984). *Atmospheric science and power production*. Tech. rep. USDOE Technical Information Center, Oak Ridge, TN.
 1290 DOI: 10.2172/6503687. URL: <https://www.osti.gov/biblio/6503687>.

- Ravikumar, Arvind P, Hugh Li, et al. (2025). “Developing Measurement-Informed Methane Emissions Inventory Estimates at Midstream Compressor Stations”. In: *ACS ES&T Air* 2.3, pp. 358–367.
- Ravikumar, Arvind P, Jingfan Wang, et al. (2018). ““Good versus good enough?” Empirical tests of methane leak detection sensitivity of a commercial infrared camera”. In: *Environmental science & technology* 52.4, pp. 2368–2374.
- 1295 Ražnjević, A., C. van Heerwaarden, and M. Krol (2022). “Evaluation of two common source estimation measurement strategies using large-eddy simulation of plume dispersion under neutral atmospheric conditions”. In: *Atmospheric Measurement Techniques* 15.11, pp. 3611–3628. DOI: 10.5194/amt-15-3611-2022. URL: <https://amt.copernicus.org/articles/15/3611/2022/>.
- Riddick, Stuart N, Mercy Mbua, Abhinav Anand, et al. (2024). “Estimating Total Methane Emissions from the Denver-Julesburg Basin Using Bottom-Up Approaches”. In: *Gases* 4.3, pp. 236–252.
- 1300 Riddick, Stuart N, Mercy Mbua, Arthur Santos, et al. (2024). “Potential Underestimate in Reported Bottom-up Methane Emissions from Oil and Gas Operations in the Delaware Basin”. In: *Atmosphere* 15.2, p. 202.
- Riddick, Stuart N. and Denise L. Mauzerall (2023). “Likely substantial underestimation of reported methane emissions from United Kingdom upstream oil and gas activities”. In: *Energy Environ. Sci.* 16 (1), pp. 295–304. DOI: 10.1039/D2EE03072A. URL: <http://dx.doi.org/10.1039/D2EE03072A>.
- 1305 Robertson, Anna M. et al. (2020). “New Mexico Permian Basin Measured Well Pad Methane Emissions Are a Factor of 5–9 Times Higher Than U.S. EPA Estimates”. In: *Environmental Science & Technology* 54.21. PMID: 33058723, pp. 13926–13934. DOI: 10.1021/acs.est.0c02927. eprint: <https://doi.org/10.1021/acs.est.0c02927>. URL: <https://doi.org/10.1021/acs.est.0c02927>.
- Santos, Arthur et al. (2024). “Using Measurement-Informed Inventory to Assess Emissions in the Denver-Julesburg Basin”. In.
- Schade, Gunnar W and Mitchell L Gregg (2022). “Testing HYSPLIT plume dispersion model performance using regional hydrocarbon monitoring data during a gas well blowout”. In: *Atmosphere* 13.3, p. 486.
- 1310 Schalkwijk, Jérôme, Harm Jonker, and Pier Siebesma (2014). “Turbulent Scales in the Boundary Layer : A Year-Long Large-Eddy Simulation YOGA : A Year-Long Large-Eddy Simulation”. In.
- Sharafutdinov, Eldar (2024). “Spatiotemporal variation in anthropogenic methane emissions in the Appalachian Basin”. PhD thesis.
- Sharan, Maithili et al. (2009). “An inversion technique for the retrieval of single-point emissions from atmospheric concentration measurements”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 465.2107, pp. 2069–2088.
- 1315 Shen, L. et al. (2022). “Satellite quantification of oil and natural gas methane emissions in the US and Canada including contributions from individual basins”. In: *Atmospheric Chemistry and Physics* 22.17, pp. 11203–11215. DOI: 10.5194/acp-22-11203-2022. URL: <https://acp.copernicus.org/articles/22/11203/2022/>.
- Sherwin, Evan, Jeffrey Rutherford, Zhan Zhang, et al. (Dec. 2022). “Quantifying oil and natural gas system emissions using one million aerial site measurements”. In: DOI: 10.21203/rs.3.rs-2406848/v1.
- 1320 Sherwin, Evan D, Jeffrey S Rutherford, Yuanlei Chen, et al. (2023). “Single-blind validation of space-based point-source detection and quantification of onshore methane emissions”. In: *Scientific Reports* 13.1, p. 3836.
- Taylor, Karl E. (2001). “Summarizing multiple aspects of model performance in a single diagram”. In: *Journal of Geophysical Research: Atmospheres* 106.D7, pp. 7183–7192. DOI: <https://doi.org/10.1029/2000JD900719>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2000JD900719>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD900719>.
- 1325 Tibrewal, Kushal et al. (Jan. 2024). “Assessment of methane emissions from oil, gas and coal sectors across inventories and atmospheric inversions”. In: *Communications Earth & Environment* 5.1, p. 26. DOI: 10.1038/s43247-023-01190-w. URL: <https://hal.science/hal-04387996>.

- USEPA (2000). *Meteorological Monitoring Guidance for Regulatory Modeling Applications*. Tech. rep., p. 171. URL: <http://www.epa.gov/scram001/guidance/met/mmgrma.pdf>.
- Vallejo, Valeria, Quoc Nguyen, and Arvind P Ravikumar (2024). “Geospatial variation in carbon accounting of hydrogen production and implications for the US Inflation Reduction Act”. In: *Nature Energy* 9.12, pp. 1571–1582.
- Van Heerwaarden, Chiel C. et al. (2017). “MicroHH 1.0: A computational fluid dynamics code for direct numerical simulation and large-eddy simulation of atmospheric boundary layer flows”. In: *Geoscientific Model Development* 10.8, pp. 3145–3165. ISSN: 19919603. DOI: 10.5194/gmd-10-3145-2017.
- Venkatram, Akula (1996). “An examination of the Pasquill-Gifford-Turner dispersion scheme”. In: *Atmospheric Environment* 30.8, pp. 1283–1290. ISSN: 13522310. DOI: 10.1016/1352-2310(95)00367-3.
- Wang, Jiayang Lyra et al. (2022). “Multiscale Methane Measurements at Oil and Gas Facilities Reveal Necessary Frameworks for Improved Emissions Accounting”. In: *Environmental Science & Technology* 56.20. PMID: 36201663, pp. 14743–14752. DOI: 10.1021/acs.est.2c06211. eprint: <https://doi.org/10.1021/acs.est.2c06211>. URL: <https://doi.org/10.1021/acs.est.2c06211>.
- Wayland, RA (2008). “Clarification of Regulatory Status of CALPUFF for Near-Field Applications”. In: *US Environmental Protection Agency, Research Triangle Park EPA, NC, accessed Mar 24, p. 2020*.
- Weil, J. C., R. I. Sykes, and A. Venkatram (1992). “Evaluating Air-Quality Models: Review and Outlook”. In: *Journal of Applied Meteorology and Climatology* 31.10, pp. 1121–1145. DOI: 10.1175/1520-0450(1992)031<1121:EAQMRA>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/apme/31/10/1520-0450_1992_031_1121_eaqmra_2_0_co_2.xml.
- Wiersema, David J., Katherine A. Lundquist, and Fotini Katopodes Chow (2020). “Mesoscale to microscale simulations over complex terrain with the immersed boundary method in the weather research and forecasting model”. In: *Monthly Weather Review* 148.2, pp. 577–595. ISSN: 15200493. DOI: 10.1175/MWR-D-19-0071.1.
- Williams, J. P. et al. (2024). “Small emission sources disproportionately account for a large majority of total methane emissions from the US oil and gas sector”. In: *EGUsphere* 2024, pp. 1–31. DOI: 10.5194/egusphere-2024-1402. URL: <https://egusphere.copernicus.org/preprints/2024/egusphere-2024-1402/>.
- Williams, James P et al. (2025). “Small emission sources in aggregate disproportionately account for a large majority of total methane emissions from the US oil and gas sector”. In: *Atmospheric Chemistry and Physics* 25.3, pp. 1513–1532.
- Worden, J. R. et al. (2022). “The 2019 methane budget and uncertainties at 1 degree resolution and each country through Bayesian integration Of GOSAT total column methane data and a priori inventory estimates”. In: *Atmospheric Chemistry and Physics* 22.10, pp. 6811–6841. DOI: 10.5194/acp-22-6811-2022. URL: <https://acp.copernicus.org/articles/22/6811/2022/>.
- Xia, Haojun, Alan Strayer, and Arvind P Ravikumar (2024). “The role of emission size distribution on the efficacy of new technologies to reduce methane emissions from the oil and gas sector”. In: *Environmental Science & Technology* 58.2, pp. 1088–1096.
- Yang, Xinxiang et al. (2023). “Direct measurement of methane emissions from the upstream oil and gas sector: Review of measurement results and technology advances (2018–2022)”. In: *Journal of Cleaner Production* 414, p. 137693.
- Zhang, Eric J et al. (2019). “Field deployment of a portable optical spectrometer for methane fugitive emissions monitoring on oil and gas well pads”. In: *Sensors* 19.12, p. 2707.