Colorado Ongoing Basin Emissions (COBE) Updated Final Report

Jenna A. Brown[1], Michael Moy[1], Arthur Santos[1], Ethan Rimelman[1], Winrose Mollel[1], Olga Khaliukova[2], Callan Okenberg[2], William S. Daniels[2], Dorit M. Hammerling[2], Daniel Zimmerle[1], and Anna L. Hodshire[1]

[1]Energy Institute, Methane Emissions Technology Evaluation Center, Colorado State University, Fort Collins, Colorado, USA
[2]Dept. of Applied Mathematics and Statistics, Colorado School of Mines, Golden, Colorado, USA

November 20, 2025

Abstract

The Colorado Ongoing Basin Emissions (COBE) project was jointly developed between teams at Colorado Department of Public Health and Environment (CDPHE)'s Air Pollution Control Division (APCD) and Colorado State University (CSU)'s Methane Emissions Technology Evaluation Center (METEC) to help inform the 2026 Colorado greenhouse gas (GHG) Intensity Verification Rule. The project is also intended to help inform the implementation of the GHG Intensity Verification Rule for calendar year 2026 and beyond. COBE had three primary objectives:

- Collect representative measurements of methane emissions from upstream oil and gas facilities throughout the state of Colorado via anonymous aerial campaigns.
- Develop measurement informed inventory (MII)s using the aerial emissions data.
- Compare the MIIs to operator-reported emissions in the Oil and Natural Gas Annual Emission Inventory Reporting (ONGAEIR) to provide recommended ratios of modeled total emissions to corresponding reported emissions.

To collect aerial measurements, the project worked with Bridger Photonics, Inc. (Bridger), GHGSat, and Insight M. METEC formed a scientific modeling team with Colorado School of Mines (CSM). METEC's modeling approach used a discrete event simulation tool via the Mechanistic Air Emissions Simulator (MAES). MAES is intended to first match a reported inventory, here ONGAEIR [1], and then add in any measurements of emissions that are determined to not be included in the reported emissions. If there is missing key information in ONGAEIR the facility cannot be modeled in MAES, which was the case for 19% of facilities for this study. While 81% of ONGAEIR upstream facilities that were operating, or partially operating, were modeled

in MAES. MAES allows understanding of emissions at the emitter level (most often, equipment level). CSM concurrently developed a statistical model that relied only on the emissions detections by the measurement technologies, using various data sets to inform emissions below the detection limits of the aerial companies, including one of emission estimates derived from continuous monitoring systems at facilities included in the study and two from the recent literature. Both models developed emissions totals and estimated ratios of total modeled emissions to reported emissions. These ratios were further split out by major basins and major facility classification. The CSM statistical model predicted higher state-wide emissions totals and ratios than the MAES model. It estimated emissions between 87,210 and 134,352 mt/y and ratios of 3.30 to 5.09 (depending on the below-threshold dataset used) when using the same subset of ONGAEIR facilities as the MAES model, and emissions of between 109,364 and 167,848 mt/y with ratios of 3.81 to 5.85 when using all ONGAEIR facilities. In comparison, MAES estimated emissions of 38,936 mt/y and a ratio of 1.47. These results are based on the 2024 ONGAEIR dataset and provide an update to a previous version of this report based on the 2022 ONGAEIR dataset.

In addition to updating MII results to the 2024 ONGAEIR, this updated report includes:

- The contribution of various emission rates to the MAES model total, showing the importance of small emissions (<5 kg/h)
- Additional methods for estimating emissions below aerial threshold in the CSM model

More work will be done by the science team in COBE-2 to provide a comprehensive method reconciliation between the two models developed in COBE. COBE-2, funded via the Mark Martinez and Joey Irwin Memorial fund, will additionally develop recommended default factors for 2027. Similar to COBE, a public report will be disseminated near the end of 2026.

Contents

1 Introduction

This report is an updated version of the report submitted to CDPHE on June 30, 2025. Updates were determined and communicated between the COBE science team and CDPHE. COBE represented the largest data collection of its kind (aerial data over upstream facilities) and analysis of data and modeling results, including comprehensive reconciliation between the two models, will be continued in the recently funded COBE-2 project, anticipated to run between 2026 and early 2027.

1.1 Project Overview

Anthropogenic methane emissions originate from several major sectors, including agriculture through livestock digestion and manure management, energy systems, waste management facilities such as landfills and wastewater treatment, and various industrial processes. Natural gas and petroleum systems are the second largest source of methane emissions in the United States after agricultural sources, contributing almost one-third (30%) of anthropogenic methane emissions [2]. Methane is a potent, short-lived GHG and a pollutant of concern. During a 20-year period, it has a global warming potential (GWP) of 86 times that of carbon dioxide, making the assessment and mitigation of methane emissions especially important to achieve near-term climate goals [3].

Natural gas operations span several distinct phases, from upstream exploration and production at well pads to processing, midstream transport, and distribution to end users. The work that follows focuses specifically on upstream production activities, which encompass wellhead facilities, associated equipment, and on-site operations that extract and initially process natural gas before it enters the broader supply chain. Upstream facilities represent a high impact area for the measurement and control of methane emissions, as the production segment accounts for 60% of the total methane emissions from the United States oil and natural gas industry, according to estimates from the EPA [2].

The traditional approach to quantifying methane emissions from oil and gas facilities is the development of a bottom-up (BU) inventory. These inventories form the backbone of official regulatory frameworks, including the EPA's Greenhouse Gas Inventory (GHGI) [2]. BU inventories estimate emissions by multiplying measured emission rates from individual sources by activity factors that represent how frequently those emission rates occur. When summed across all equipment at a facility or region, this methodology produces aggregate emission estimates. However, limitations in traditional BU approaches drive the need for measurement integration to improve inventory accuracy [4, 5, 6, 7].

Quantifying total methane emissions from producing basins is a topic of interest for both operators and policymakers at the federal and state level in the United States. Colorado in particular has advanced regulations designed to limit methane emissions during production. December 2021 rulemaking created a framework for a program that included the intensity thresholds in Kg CO_2e/kBOE beginning in CY2025. In 2023, the Colorado Air Quality Control Commission (AQCC) adopted its GHG Intensity Verification Rule, which defines intensity as the ratio of facility GHG emissions to oil and gas production volume [8]. Before calculating intensity, the emissions submitted for a given development are multiplied by a distinct intensity factor [8]. Operators in the state are required to either use the default intensity factor provided by CDPHE, or follow an outlined methodology to calculate their own, by developing an operator-specific measurement informed inventory (MII) [8]. To support accurate implementation of this rule, updated emissions measurements and more accurate intensity factors are needed for each basin.

The COBE project is an environmental initiative to create and refine a regional model of the methane emissions of Colorado's upstream oil and gas facilities. COBE is led by CSU's METEC group with significant modeling support from CSM. By providing an updated inventory of methane emissions, the CDPHE and its APCD can better implement and enforce the state's GHG Intensity Verification Rule [8] and other air quality regulations. COBE helps develop the methodology for the calculation of the default intensity factor by comparing emissions reported to the Oil and Natural Gas Annual Emissions Inventory Reporting (ONGAEIR) by upstream operators with the emissions measured by aerial measurement campaigns throughout Colorado. There are numerous oil fields in the state; in this project, production activities are grouped into three main basins: the Denver-Julesberg, Piceance, and "Other", which includes the Raton, North Park, and other smaller reserves. For this project, three aerial methane detection companies were contracted to fly aerial campaigns to find methane emissions: Bridger, Insight M, and GHGSat. Each company uses different sensor technologies and detection methodologies to quantify methane emissions, providing independent datasets for emission measurements and uncertainty assessment [9, 10, 11, 12, 13, 14].

To develop ratios that compare modeled to reported total emissions that will be used by CDPHE APCD to develop intensity factors to support the GHG Intensity Verification Rule, the modeling team (METEC and CSM) focused on measurement-informed inventory MII methods. MIIs are an approach to regional emission modeling that combine BU estimates with spatially and temporally overlapping measurements. Currently, top-down (TD) methods generally suggest that bottom-up estimates based on traditional inventories underestimate emissions [4, 5, 6, 7]. There are several reasons for this: one such reason for this is that large, rare emissions are difficult to capture in brief measurement campaigns, which means that emission factors used in the inventories do not adequately represent the full distribution of emission sources [4]. Among these emitters not captured within BU emission factors, "super-emitters" are significant sources of methane that are often revealed by TD methods [15].

Additionally, BU modeling relies on activity data which is often incomplete; reporting programs such as ONGAEIR only represent known frequency, not the true prevalance of emission events. In contrast, TD measurements do not describe behaviors at the emitter level, which are useful to assess whether leaks can be prevented or mitigated [16].

This report details the aerial campaigns, aerial results, and MIIs for Colorado's upstream sector. METEC and CSM developed two independent models, each with strengths and limitations, to determine the MIIs and ratios of modeled to reported total emissions. METEC uses a discrete event simulation tool called MAES. MAES uses site characteristics and emission factor data to generate transient emissions expected for oil and gas facilities across the state. For a given oil/gas facility, MAES is used to model a profile of "normal" emissions, or essentially a BU inventory of expected emissions. In COBE, the model is compared to ONGAEIR as a check on whether MAES accurately represents a facility's "normal" or expected emissions. Then, to address the shortcomings of BU methods, aerial measurements are incorporated into the model to capture emissions not reported in ONGAEIR. Post-completion of the measurement surveys, emission detections were analyzed and categorized in conjunction with each site's operator (when available). This process allowed the modeling team to attribute a source cause for the results, which is necessary to exclude emissions that are already reported within ONGAEIR or caused by equipment maintenance. Emissions deemed unlikely to be in the inventory are then integrated into updated MAES models: the final modeled emissions are then an estimate of those in the existing inventory plus unreported emissions observed in the measurement campaign.

While the MAES-based approach incorporates operational data and mechanistic simulations to predict emissions on a site, the statistical model by the CSM team relies solely on rates estimated by measurement technologies: it assumes no prior knowledge of typical facility emissions. This approach first fits a probability distribution to site-level emission rate estimates from all three aerial vendors, taking into account the differences in detection sensitivity across vendors. Unlike the MAES modeling approach, the statistical model does not differentiate between abnormal emissions and normal process emissions. It assumes there is enough aerial data to fully capture the relative rate of occurrence and emission characteristics of the abnormal emissions when fitting an overall emission distribution. A separate distribution is used to model emissions below the aerial detection thresholds, which differ by vendor. Three methods are proposed to this end: one that is informed by continuous monitoring systems (CMS)-derived rates, and two that are informed by previous work by Williams [15] and Sherwin [17], respectively. These two distributions are then repeatedly sampled from to provide state- and basin-wide emissions estimates.

2 Data and Measurement Methods

The 2022 ONGAEIR dataset—the most recent publicly available inventory at the project's inception in March 2024—served as a foundational resource. Maintained by the CDPHE, ONGAEIR is a database of annual GHG emission estimates submitted by oil and gas operators per state regulations [1]. It provides detailed, facility-level information on equipment types known to be potential sources of methane and other GHGs. Updated annually and made publicly available, the ONGAEIR database plays a critical role in supporting regulatory oversight and emissions reduction goals in Colorado. Its comprehensive scope and standardized reporting structure made it essential to the design of the sampling plan and modeling approach in this study. Although the 2023 ONGAEIR dataset became available midway through the project, the team proceeded with the 2022 inventory due to the absence of quality control in the newer dataset and because flight planning had already been based on 2022 data. The 2022 dataset includes records for 11,473 upstream oil and gas facilities in Colorado that were fully or partially operational during 2022, according to information provided by operators to CDPHE.

The 2024 inventory, released around September 2025, was incorporated retrospectively to align with the timing of the flight campaigns. There is a challenge in merging ONGAEIR data across reporting years because facility names often change, reported lat/long locations shift, and no unique identifier carries over from one year to the next. Due to flight planning using the 2022 ONGAEIR, approximately 8% of facilities in 2022 were not present in the

2024 database, which could be due to facilities being shut-in, reported in a different sector (i.e. midstream) or an unknown reason. This version of the report uses the new 2024 inventory as new counts of facilities and the updated base of emissions. However, all measurement figures and statistics are relative to ONGAEIR 2022. In ONGAEIR 2024, 11,681 production facilities were operational or partially operational. There were a few facilities with egregiously high reported methane emissions in ONGAEIR 2024 and therefore, these facilities were not included in this analysis, see Section A.14.

In addition to the difference in facility counts, the annual gas production and the total annual methane emissions reported in ONGAEIR 2024 are roughly half of what was reported in ONGAEIR 2022, despite 2,871 additional wells in 2024. One operator with about 2,000 wells did not report in 2022, which contributes to the well discrepancy. In addition, ONGAEIR 2022 included approximately 64,000 more pieces of equipment than 2024.

2.1 Measurement Campaign Criteria & Prototypical Sites

COBE deployed three aerial companies, Bridger, GHGSat, and Insight M (formerly Kairos Aerospace), to collect a representative sample of measurements of methane emissions from operating upstream facilities in Colorado. For sample planning, the METEC team considered several key stratification variables, including the number of wells per facility, production levels, operator diversity, and representative facility types. These representative facility types are addressed, following previous studies ([18], [19]), by classifying facilities into categories with common equipment groupings, called prototypical sites (PSs); see Table 1. Based on the classification of Winrose et al. [18], which defined prototypical sites for the Colorado Coordinated Campaign project [20], we developed a simplified classification that also accounts for the impact of fluid flow on equipment-level emissions. Specifically, we considered the influence of gas lifts, tank batteries, flares, and vapor recovery units, which are known to significantly affect site emission profiles. The PSs classifications were made using reported equipment from ONGAEIR, and Figure 1 shows the determination scheme.

Sampling criteria were communicated to the aerial vendors and iterative adjustments were made until acceptable sample plans were established. An additional component of the sampling strategy included reflights, in which aerial vendors were instructed to re-survey 20–25% of the selected facilities, ensuring a minimum interval of 24 hours between flights. The facilities to resample were pre-determined before flights to not bias towards facilities that did or did not have emissions on the first fly-over. Flight scheduling was left to the discretion of the aircraft companies, who coordinated operations independently.

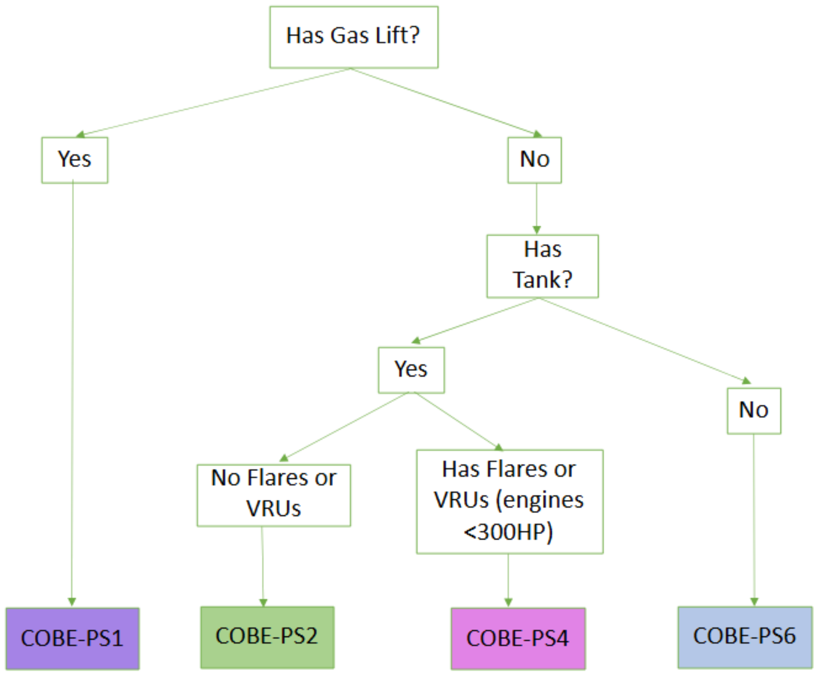


Figure 1: Facilities were categorized into common equipment groupings based on whether they contained gas lifts, tank batteries, flares, or vapor recovery units (VRUs). This diagram shows how PSs were determined.

Our later estimates of emissions by equipment type will require a count of equipment at each facility, which we take primarily from ONGAEIR. Using the ONGAEIR database for equipment counts is not without limitations. Notably, operators are only required to submit records of equipment that have known associated emission events or activities in their ONGAEIR submissions. This is because ONGAEIR was designed as an inventory of emission sources, rather than a comprehensive inventory of all equipment present at a facility, regardless of whether the equipment is expected to have "as-designed" emissions. However, as borne out in many recent studies (e.g. [19]), emissions frequently happen unbeknownst to the operator. This is particularly pronounced in separators and heaters, which are underreported in ONGAEIR compared to alternative data sources such as aerial imagery and operator records. For example, these components may emit during failure conditions, yet such emissions would be absent from BU inventories and may or may not be captured by aerial surveys, depending on time and detectability. This highlights the importance of inventorying all equipment with emission potential, not just those with operator-reported leaks. To address the specific issue of missing heaters and separators, a decision tree was developed and used to adjust equipment counts based on facility characteristics. The logic begins by evaluating whether heaters are reported at each facility. When heaters are present but separators are absent, separator counts are set equal to the number of heaters. Coalbed methane wells retain their original equipment counts due to distinct operational requirements. Non-coalbed wells producing only gas maintain original counts, while oil-producing facilities have both heater and separator counts adjusted to match well counts to reflect common operational practices.

2.2 Measurement Campaigns

The three aircraft companies were deployed across three distinct project phases. The first phase occurred from May to July 2024, the second spanned from late July through the end of August 2024, and the final phase extended from September 2024 to February 2025. Approximately 75% of flights occurred on weekdays and 25% on weekends, with all flights conducted between 6:50 AM and 4:50 PM. Gas Mapping LiDAR (GML) data from Bridger of concurrent measurements within the Site-Aerial-Basin Emissions Reconciliation (SABER) project in the Denver-Julesberg (DJ) basin were incorporated into COBE total emissions analysis to increase the available dataset. Due to coordination between the two projects, flight data from Bridger in the DJ basin campaign were shared with COBE.

The objective of the flight campaigns was to ensure broad representation across the dataset, with approximately one-third of the samples allocated to each of the major regions: the Piceance Basin, the DJ Basin, and the "Others" region, or all remaining regions combined, as seen in Figure 2. These basin outlines were provided by CDPHE. The number of PSs per basin are shown in Table 1.

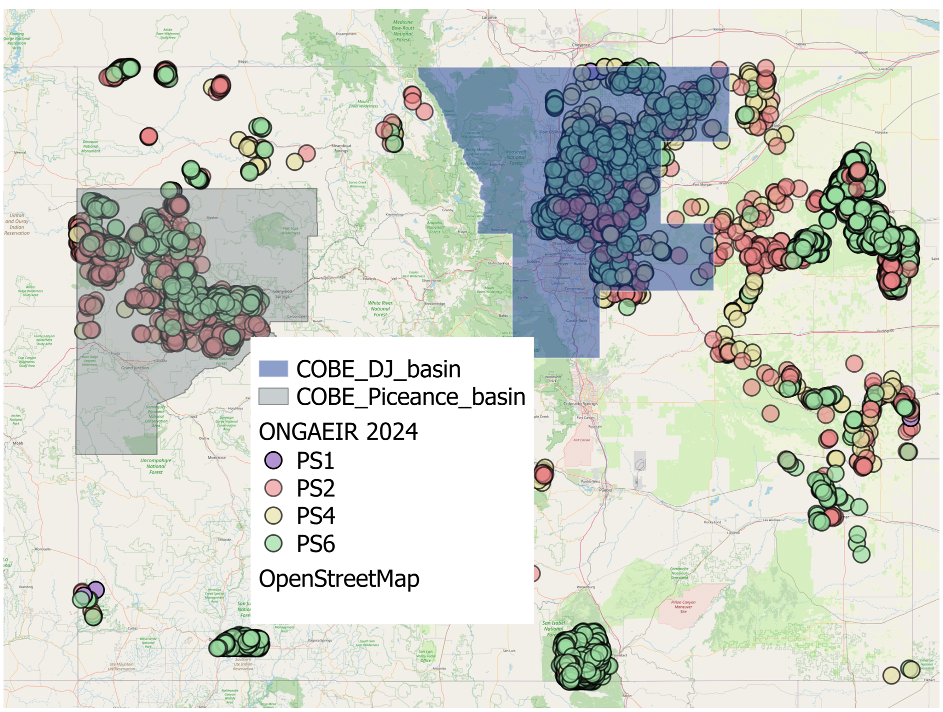


Figure 2: Map of Colorado showing the spatial distribution of PS across major oil and gas basins using ONGAEIR 2024. The state is divided into two primary basins: the Piceance Basin (gray) and the Denver-Julesberg (DJ) Basin (blue), with all other basins grouped as "Other." These basin outlines were provided by CDPHE. Prototypical sites are color-coded by class: PS1 (purple), PS2 (red), PS4 (yellow), and PS6 (green). Section 2.1 details the specifics for each PS.

Operators across the state were informed of the project through email communications and public informational sessions were held prior to the start of flight operations. Operators were invited to participate and those who chose to participate received all aerial overflight data

Basin	PS1	PS2	PS4	PS6
DJ	270	378	1,502	591
Piceance	10	1,937	410	744
Others	12	1,187	447	4,193

Table 1: Production facility counts by PS classification across different basins present in the 2024 ONGAEIR data.

related to their assets. This included a full list of all assets that were flown and all detected emissions. In return, participating operators supported the METEC team by providing cause analyses for detected emissions. A total of 12 operators participated in the project, collectively representing approximately 70% of facilities in the 2024 ONGAEIR dataset (and 77% of the facilities in the 2022 ONGAEIR dataset).

A structured process was implemented for participating operators: 1) The aerial companies conducted flyovers of oil and gas facilities. 2) Emissions data were received by the METEC team. 3) METEC organized and sorted all data—both detections and non-detections—by operator. 4) Operator-specific datasets were then sent to each participating company, along with a set of questions to complete about each emission detection, referred to as a "cause analysis". The project was structured such that all cause analyses reported to the METEC team were voluntary and were not required to follow the rigor of more formal processes, such as root cause analysis, in order to reduce burden on the operators.

2.3 Measurement Methods and Data by Vendor

The three aerial companies use distinct remote sensing technologies for detecting methane emissions. All three have participated in controlled testing and field validation studies and have demonstrated strong methane emissions localization and quantification capabilities. While their measurement systems differ, each has shown the ability to accurately detect methane sources under a range of environmental conditions. A summary is provided below, and a more detailed description of each aerial technology is presented in Section A.2 of the appendix.

Bridger relies on a GML system that enables high-precision localization and quantification of methane plumes by combining a cross-sectional flux estimation method [21] with atmospheric data [22]. Specifically, the GML 2.0 system was used in COBE, and its performance has been evaluated in a controlled release study performed by Thorpe et al. [23]. A detailed description of the GML system is provided in Section A.2.1 of the appendix. Bridger's localization capabilities enable attribution of emissions to specific equipment, with reported measurements including both the emission rate and the associated equipment type. Site-level emissions are calculated by aggregating the daily average emissions from all point sources at a given site.

GHGSat's aerial measurement technology uses shortwave infrared (SWIR) spectrometry to detect methane by analyzing reflected sunlight for gas-specific absorption signatures [24]. During the measurement campaign, GHGSat deployed three sensors from two generations of its technology, with reported detection limits of 10 kg/h and 5 kg/h, respectively. GHGSat reports emission rates at specific, geolocated points within the scanned site. In most detections during the measurement campaign, the location was only specific enough to treat the emission as a facility-level estimate, but some measurements showed multiple clearly defined plumes, which were identified as separate emissions. GHGSat emission detection and localization capabilities have been tested in various studies [25, 26, 27].

Insight M uses LeakSurveyor technology: an aircraft-mounted hyperspectral infrared system designed to measure patterns of sunlight energy absorbed by methane [11]. Insight M used two different sensors during the measurement campaign, with reported detection limits of 25 kg/h and 10 kg/h. Its measurement systems have been tested in several controlled release studies, providing accurate plume detection and emission rate estimation [26, 27]. Insight M reports emission rate estimates at a facility level.

2.4 Measurement Uncertainty

The three different aerial companies and the different sensors used have variable probabilities of detection and measurement uncertainties. The differences in probabilities of detection are especially evident in the data from the measurement campaign (for instance, see Figure 36 in the appendix). To account for these differences in aerial technologies, our analysis makes use of previously published results from controlled release testing involving the three companies, which provide estimates of measurement uncertainties and probability of detection curves.

Bridger provided us with a copy of their error model, which models the relative error ratio for each measurement according to a log-logistic distribution. For consistency, we chose to model the errors in GHGSat and Insight M measurements by log-logistic distributions as well. Based on publicly available data [25, 27] from controlled release tests, we fit a distribution for each of the two different sensors flown by Insight M. GHGSat reports a standard deviation for each of their measurements, estimated from multiple sources of error [28], so we used log-logistic distributions with these reported standard deviations to model the errors. The resulting combination of error models accounts for the differences between the companies, and they are incorporated into the analysis and modeling described in later sections. Further details on the error models can be found in Section A.2 of the appendix.

In addition to the error models, probability of detection curves were estimated using a combination of the data from the measurement campaign and previously published data from controlled release experiments. The probability of detection curves were used in the analysis of the data, but to avoid making a direct comparison of the technologies, the curves themselves are not presented here. The controlled release experiment of [25] and [29] provided enough data to estimate probability of detection for Bridger and for Insight M's 10 kg/hr sensor; in these cases, we fit logistic curves estimating the probability of detection as a function of emission rate. Insight M's 25 kg/hr sensor was assumed to reach a given probability of detection at 2.5 times the emission rate needed for the 10 kg/hr sensor. For GHGSat's three sensors, we approximated probability of detection curves by comparing the emission rates seen during the measurement campaign with those seen by Insight M's 10 kg/hr sensor. Further information is given in Section A.8 of the appendix. These probability of detection curves were taken into account when estimating distributions of emissions attributed to specific sources, described in the following section.

3 Modeling Methods

This section presents two distinct methods for modeling methane emissions from production sites and estimating state-wide annual emissions. One approach, developed by the METEC team, analyzes measurement data in detail to determine emissions that are likely not already reported in ONGAEIR, often due to abnormal conditions or equipment failures. Operational changes due to the addition of these unreported emission sources are simulated using MAES to generate facility-level MIIs. Running these simulations across all sites provides an updated annual emissions estimate for Colorado that can be broken down by equipment type and site classification.

The CSM team pursued a statistical approach to provide an independent estimate of the average emissions from all production sites as a whole by fitting a distribution to the emissions measured by the aerial companies. For the CSM's statistical model, the aerial data, which provides a representative sample of "large enough" emissions (those detectable by aircraft) is combined with various datasets to characterize the remaining smaller emissions, producing facility-level estimates that account for the full range of emission rates. In particular, three below-threshold datasets are tested and compared: one using continuous monitoring data from a very small sample of homogeneous sites, and two from the literature in papers by Williams [15] and Sherwin [17] that both aim to characterize emissions distributions in the DJ Basin. Throughout this section and our results and discussion sections, we will make a clear distinction between the two approaches, as they provide different perspectives on how measurement data can be used to improve inventory emission estimates.

3.1 METEC Modeling: the Mechanistic Air Emissions Simulator (MAES)

While various measurement-based approaches exist for quantifying methane emissions from oil and gas facilities, an alternative method involves modeling emissions based on facility-specific operational data. The Mechanistic Air Emissions Simulator (MAES) is a model developed by the METEC team to simulate process flows and associated emissions from oil and gas infrastructure at the equipment- and failure-level. Examples of its use in simulating oil and gas facilities can be found in [18] and [19]. MAES uses Monte Carlo (MC) methods to capture the variability in facility operations and is based on the discrete event simulator (DES) method with a time resolution of 1 second. Individual pieces of equipment are simulated as state machines, while simulated fluid flows between equipment provide a cohesive model of an entire facility; see [30, 31] for further explanation of these modeling approaches. Multiple facilities are individually simulated with site-specific parameters, and results are combined to derive regional emission estimates. A single simulation of a facility over a period of time (typically weeks to years) is referred to as an MC iteration, and the results from a collection of MC iterations can be used to approximate a distribution of emissions produced by the facility. Typical simulation parameters include a one-year time frame and 100 MC iterations, but these may vary depending on the event types users aim to capture. For example, a failure event with a probability of 0.001 is expected to occur once, on average, every 1,000 MC iterations; a larger number of runs increases the likelihood of observing such rare events.

MAES estimates emissions using two different types of models, mechanistic and traditional.

Mechanistic models focus on how fluids move through equipment by modeling the physical processes and interactions that govern emissions at each stage of the system. Since they model the mechanisms that lead to emissions, they provide a way to model emissions from facility characteristics rather than empirical emission data (e.g. emissions factors, campaign data). Traditional models use activity data multiplied by emission factors to estimate emissions. Emission factors are input to MAES as distributions specifying the frequency of a given emission rate. These are determined from emission measurements at oil and gas facilities, both from preexisting datasets [32, 33] and from the specific datasets to be studied, in our case the data from the COBE measurement campaign. Methods for determining emission factors from the observed data are described in more detail in Section 4.2.

To accurately represent each facility's unique configuration, MAES requires several key inputs, including gas composition, facility configuration, and equipment counts (see Figure 40 in the appendix). For this study, facility-specific data for use in MAES were obtained from the ONGAEIR database using the calendar year 2022 report, but were then updated with ONGAEIR 2024 facility information when it became available. However, some critical parameters—such as facility-specific gas composition and detailed process connectivity between equipment—were not available in public datasets. In such cases, reasonable assumptions were made to fill these data gaps, based on engineering judgment and typical facility design practices, using findings from Mollel et al. [18]. To simplify this process, the prototypical sites defined above were used to determine the connections between equipment.

To model fluid flows through a facility in MAES, another key requirement is reported gas or liquid production. If there is no reported production or there is missing facility information, the facility will not be modeled. In ONGAEIR 2022, there were 10,144 production facilities that were partially operating or operating that met this criteria and were therefore modeled in MAES. This number was reduced to 9,411 using ONGAEIR 2024, which is roughly 81% of the 11,681 upstream facilities that were operating or partially operating. See Appendix Section A.15 for more information on the criteria for MAES to model a facility. Section 4.5 investigates the difference in the ONGAEIR reported emissions of the modeled sites versus the unmodeled sites.

For MAES to generate a baseline inventory of Colorado's many production sites, the counts of equipment by type must be input for all facilities. MAES has models for simulating various equipment such as wells, tanks, flares, separators, heaters, compressors, dehydrators, pneumatics, and miscellaneous equipment. Facility equipment data was taken primarily from ONGAEIR, as described in Section 2.1.

From the inputs described above, MAES outputs a record of each MC iteration for each facility. Emissions by each piece of equipment are recorded by start time, duration, and emission rate with a time resolution of one second to capture the temporal variability of emissions. Results for the entire collection of facilities simulated are combined to produce annual emission estimates, broken down by site, equipment, or emission type. These estimates can also be made for subsets of the facilities: in our case, we generate separate estimates by basin and by prototypical site.

3.2 METEC Modeling: Building a Measurement-Informed Inventory (MII) with MAES

We follow the process outlined in [19] to create a facility-level MII using MAES. Beginning with a given inventory, in our case ONGAEIR, the process identifies emission sources detected in the measurement campaign that are likely not accounted for in the inventory. This requires a combination of discussions with participating operators and comparisons of measured emissions with those from initial MAES model outputs. Once these sources are identified, their contribution to the inventory is estimated through updated MAES models that include these sources, thereby adding emissions from these sources into the inventory. This process is divided into the following steps.

- A) Inventory matching: normal emissions, including both vented and combusted sources, are simulated in MAES. The modeled emissions are compared to reported annual emissions from the inventory, providing a diagnostic check on whether the model accurately represents the facility's typical emission behavior. When discrepancies arise, both the MAES model and the inventory-reported emissions are examined to identify potential causes and resolve inconsistencies. The result is a MAES model that can accurately simulate emissions currently reported in the inventory. These initial models are called MAES inventory models.
- B) Emissions Survey and Classification: analysts use the cause analysis (see Section 3.2.1) and preliminary MAES models to determine whether measured emissions were related to maintenance activities, already reported in the inventory, or unreported. Unreported emissions are further classified by their sources (see Section 3.2.2) for use in simulations in step D.
- C) Maintenance Emissions: operator cause analysis or aerial imagery is used to identify emissions due to maintenance events. These emissions are not modeled in MAES, and the inventory estimates of maintenance emissions are used in the final results.
- D) MII: based on the classification in the previous steps, emissions that were unreported are incorporated into an updated MAES model. These additional emissions are simulated as abnormal conditions in the identified sources with the frequencies observed in measurements (see Section 3.2.3). These models are called MAES MII models.
- E) Results: the MAES MII models produce a detailed MII, with annual emissions estimated by equipment type and site classification.

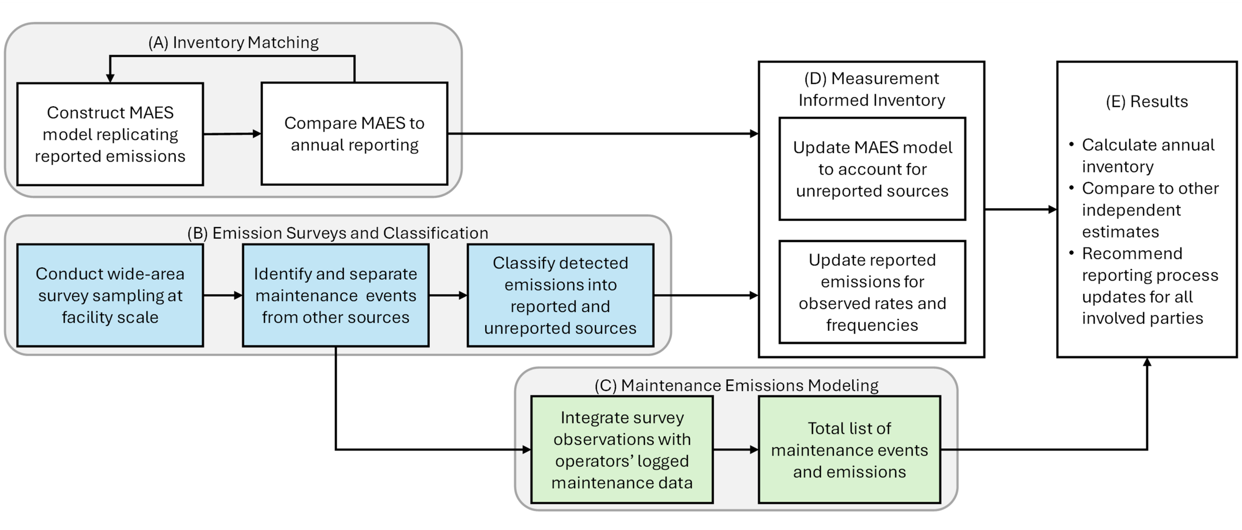


Figure 3: An outline of the MII approach using MAES (reproduced from [19]).

In step A, when the MAES model is compared to ONGAEIR, emission categories not modeled within the MAES framework are excluded from the analysis. The following categories are excluded:

- Well Maintenance
- Loadout
- Venting and blowdowns - with exception of compressor blowdowns, which are simulated by MAES[1]
- Well Bradenhead

Compressor venting and blowdowns are modeled in MAES whereas other equipment blowdowns are not modeled and therefore excluded from further comparison to MAES. Equipment blowdowns (with the exception of compressor engines) and other maintenance-related events, as described above, are not included in MAES due to their highly episodic nature and the current lack of sufficient data to characterize their frequency and magnitude reliably.

In the early steps of the process, meetings were held with participating operators to ensure accurate interpretation and model alignment. The first meeting reviewed results from step A, to confirm that the inventory data were correctly represented in the model. A follow-up meeting focused on step B to address questions related to the detected emissions, the operator-provided cause analysis responses, and any remaining uncertainties regarding specific emission events. This process provided valuable insight into the likely causes of each emission event, allowing the team to determine whether the source was already accounted for in the reported inventory, missing and therefore requiring modeling, or associated with a maintenance activity.

[1]Emissions from venting and blowdowns were further categorized: venting from compressors was classified under Compressor Venting, while blowdowns from compressors were placed under Compressor Blowdowns.

MAES outputs are used at two distinct points in the MII process, in steps A and E. In step A, normal operating conditions are simulated in MAES. Here, we assume that the inventory (ONGAEIR) provides a reasonable baseline estimate of normal emissions. Activity data from the inventory are used to build the MAES inventory models, and the resulting emission estimates serve as a diagnostic tool to evaluate the consistency of reported values. Rather than adjusting the model to force agreement, discrepancies between simulated and reported emissions are investigated to identify potential issues in either the model assumptions or the inventory data.

To evaluate discrepancies between simulated and reported emissions, we applied operator-specific thresholds based on the magnitude and context of the observed differences. For major participating operators, facilities with absolute differences exceeding 20 metric tons per year were flagged for review. For a specific company, a higher threshold of 40 mt/y was applied due to broader variability. For smaller operators, a lower threshold of 3-6 mt/y was used, given the facility types and smaller sample size, with generally lower errors. With participating operators, a discussion about these facility discrepancies between MAES inventory and ONGAEIR was covered in the first meeting. For non-participating operators, where errors were consistently larger and more systematic, we adopted a higher threshold of 100 mt/y to identify the most significant anomalies. The analysis overall revealed several instances where discrepancies appeared to stem from issues within the ONGAEIR database or from operator-reported data entry errors. This process was simplified when the model was rerun using 2024 data, due to time constraints.

To compare the best-estimate inventory with the model, one adjustment was made to ONGAEIR. It was determined that some operators used Subpart C methane emission factors to estimate combustion emissions from stationary engines and turbines. Because these factors underestimate emissions from natural gas engines [34], they were scaled to align with the updated Subpart W emission factors. This increases the ONGAEIR total methane emissions by approximately 2,300 mt/y.

This iterative process supports mutual validation of both the simulation framework and the reported emissions, assuming the activity data is correct. Once the inventory model is close to the reported ONGAEIR annual emissions (approximately within a 15% error), then the MII model can be run, using updated inputs that reflect the emissions classified as unreported. Multiple MC iterations were used to approximate the distribution of emission estimates for each facility—100 iterations per facility for the inventory model, and a variable number of iterations in the MII model determined by 1/probability of leak (pLeak). From these distributions, 95% confidence intervals are reported to indicate the variability in these estimates. The outputs from the MII model were compared with those from the inventory model and estimates from ONGAEIR to determine the change in emissions.

Sections 3.2.1, 3.2.2, and 3.2.3 elaborate on the more intricate parts of the process. The MAES inventory and MII models constructed by this process, along with comparisons ONGAEIR, are presented in Section 4.3.

3.2.1 Operator cause analysis

For the MAES MII process, it was necessary to parse emissions data and label events with a suspected mechanism/cause. Certain types of equipment failure (i.e., flare malfunction) can be modeled mechanistically within MAES once their frequency and emission characteristics are understood from the measurement data. Maintenance-related emissions, however, must be excluded from simulation, as they are operator-controlled and don't follow predictable emission rate patterns that can be captured in the modeling framework (see below). To properly classify each detected emission, we engaged in a structured cause analysis process with participating operators. Each operator received a specific dataset of all aerial detections at their facilities, including both detections and non-detections, and a meeting was held to determine a plausible explanation for each emission event. This process aimed to determine whether the emission was due to normal operations, equipment failure, or maintenance activities. It also served to assess whether the emission was already captured in their ONGAEIR reporting and to identify the likely equipment source. This operator feedback was important for accurately categorizing emissions and ensuring the MII properly captured only those emissions missing from ONGAEIR.

Maintenance activities are highly transient, so aerial methods, which see a snapshot of emissions at a particular time, cannot reliably quantify emissions from these events. For this reason, these emissions were not modeled in the MAES framework and were excluded from the emission distribution. Since maintenance still contributes to total emissions, we add maintenance emissions reported in ONGAEIR back to the final MII totals to ensure complete emissions accounting.

In total, 42 measured emissions were attributed to maintenance activities, 38 of which were identified by the operator, and 4 of which were identified by the analyst of this team. The analyst identified a maintenance activity if there was a truck on site near the emission event, or if the same source was emitting within one work week of the operator reporting a maintenance activity. These events included liquid unloading, blowdowns, engine startups, bradenhead venting, swabbing, and open thief hatches. If there were multiple detects on the site during a maintenance event, all emissions from that day were excluded from the MAES MII modeling. The probability of detecting a maintenance event determined from the COBE aerial campaigns is 0.00127 for the state of Colorado. At a more granular level, Bridger detected 33 emission events that were classified as maintenance activities across 26 facilities, with an average emission rate of 12 kg/h. GHGSat detected 4 maintenance activities at 4 facilities, averaging 82.2 kg/h. Insight M detected 5 maintenance activities at 5 facilities, with an average emission rate of 36.7 kg/h.

Emissions that are excluded from the MII model, shown in Table 2, include those due to maintenance activities, pre-production activities, midstream site identity, and misalignment between the detection location and the reported coordinates. Of 2,102 nonzero emission measurements, Bridger recorded 44 that were determined by the modeling team to be from pre-production activities, while GHGsat and InsightM did not pick up any emissions at these sites. 96 emission detections from midstream facilities were identified (by operators): 80 from Bridger, 6 from GHGSat, and 10 by InsightM. 40 emission detections were spatially offset from the facility coordinates reported in ONGAEIR. Operators informed the team that the associated facility names were either incorrect or that the facilities no longer belonged to them; all but two (one from InsightM, one from GHGSat) were detected by Bridger. This is all reflected in the anonymized dataset [35].

Table 2: Summary of emissions excluded from MII modeling by category and aerial vendor

Category	Bridger	GHGSat	Insight M	Total
Pre-production activities	44	0	0	44
Midstream facilities	80	6	10	96
Location misalignment	38	1	1	40
Maintenance	33	4	5	42

3.2.2 Classifying emissions

As described above, one of the benefits of incorporating MAES simulations into an MII is the ability to model the emissions contributed by various types of equipment. Here we detail the process of classifying unreported emissions observed by aircraft according to their sources. Emissions are considered at the equipment level, when possible; in cases where multiple measurements of emissions from the same equipment were recorded in a single day, they were recorded as a single detection and the emission rates were averaged. Insight M and GHGSat typically reported at the facility level. When successive observations are made within minutes of each other, they are counted as a single observation, and the associated emission rates are averaged. Based on the simulation abilities of MAES, for this study we have attributed emissions to the following "failure types", each of which is simulated in an associated type of equipment in MAES.

- Compressors - rod packing failures are modeled by MAES. Observed emissions likely include a combination of combustion slip, crankcase emissions, and rod packing emissions, which cannot be measured separately by aircraft. To assess whether observed emissions are consistent with normal operation or indicative of a failure, we first reviewed the operator's cause analysis. Next, for each facility, compressor-specific information (brake horsepower, engine class, etc.) from ONGAEIR was used in MAES to estimate crankcase, driver exhaust, and rod packing emissions for all compressors. Due to the lack of operational data at the time of the flyover, we assumed all compressors were active. To isolate measured rod packing emissions, the MAES estimates for driver exhaust and crankcase emissions were subtracted from the total measured compressor emissions, and the remaining emissions were attributed to rod packing. This value was then compared to the expected rod packing emissions from MAES: if the measured rod packing emissions exceeded the MAES estimate, the excess was attributed to potential rod packing failures at the facility.

- Flares - failures include both malfunctioning and unlit flares. To identify these cases, the process below was used to determine whether the measured emissions exceeded normal emissions estimated by MAES. The mechanistic MAES model for flares only requires an estimate of frequency of failures, so estimates of emission rates are not needed.

- Heaters - failures are heater malfunctions resulting in incomplete combustion. To identify when a heater was malfunctioning, the process below was used to determine whether the measured emissions exceeded normal emissions estimated by MAES. As with flares, the MAES model for heaters only requires an estimate of frequency of failures, so estimates of emission rates are not needed.

- Tanks - controlled and uncontrolled tanks are modeled separately. Since emissions seen by aircraft often have an unknown cause (uncontrolled tanks, stuck dump valves, open thief hatch, etc.), they are grouped into a single emission factor. Because the specific cause can often not be determined, tank emissions modeled in the MAES MII models include all emissions from tanks greater than 2 kg/hr, regardless of failure or normal operations. Emissions from tanks below 2 kg/hr are already modeled as tank component leaks by MAES, and are matched to inventory emissions in step A.

- Miscellaneous emissions - these are emissions classified by Bridger as “Other” or “Facility Piping”, and emissions where the source is unknown. These are modeled by a single “miscellaneous” emitter in MAES. If aircraft measurements include multiple simultaneous emissions attributed to the miscellaneous category, these are summed to be modeled as a single emitter in MAES. To identify when the miscellaneous emissions exceeded expected MAES estimates, the process below was used.

Each detected emission was attributed to one of these failure types. Bridger reports associated equipment for their emission measurements, which were used in the absence of other information from the cause analysis for these cases. Equipment are not reported by GHGSat and Insight M, so emissions they report must be assigned a failure type separately. The cause analysis and aerial imagery were used in these cases to determine the likely source of the emission and assign a failure type. Uncertainty in this process is reflected through the probability scores described below.

To determine whether the detected emitters exceeded levels consistent with normal operations, the following process was used (see Figure 4 for an example). The MAES results from step A (i.e. expected inventory emissions) generated both a probability distribution function (PDF) and a cumulative distribution function (CDF) for each facility. Each detected emission event was overlaid on the corresponding facility-specific CDF to determine whether it fell within the expected range of emissions. If the detected emission was within the modeled CDF range, it was considered consistent with expected emissions. If it fell outside the modeled distribution—particularly in the upper tail, like in Figure 4—analysts assigned a MAES failure type using any operator notes and aerial imagery. In the above step, if there were questions regarding an emission detection, this was covered in the second operator meeting.

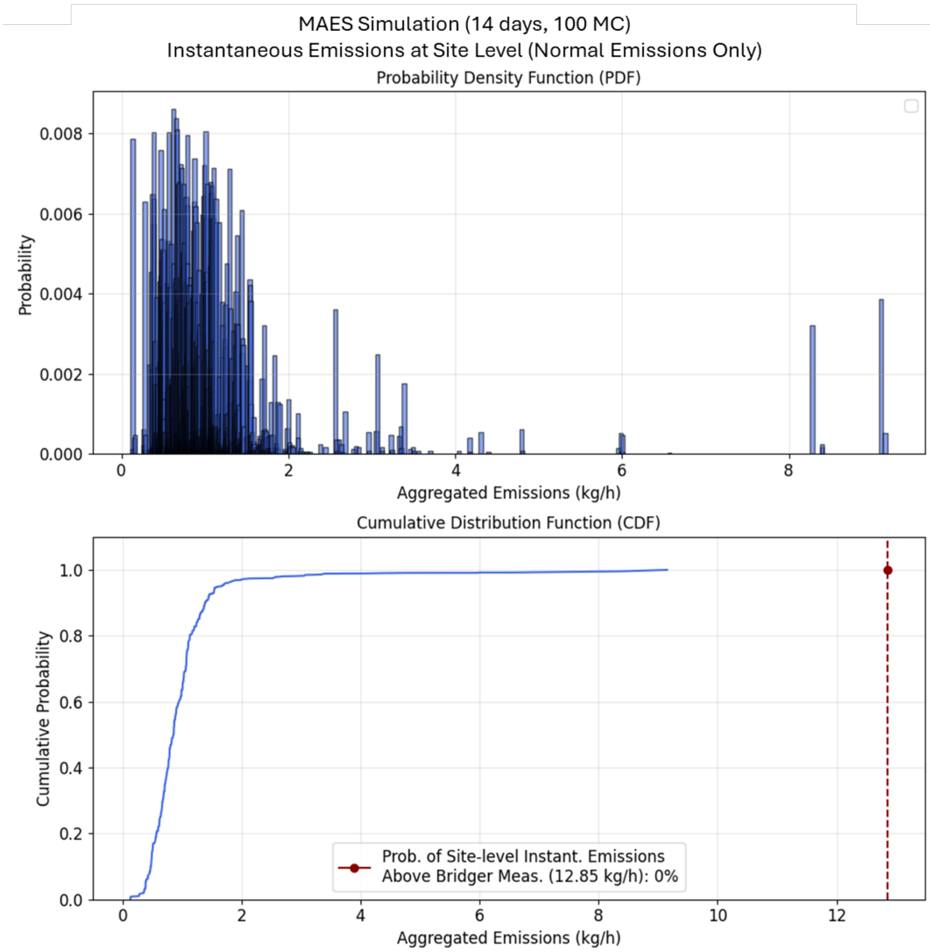


Figure 4: The 1-Hz simulated results for a facility were converted to a probability distribution function (PDF) (top) and cumulative distribution function (CDF) (bottom). This facility was simulated for 14 days, with 100 MC iterations. The aerial estimate is overlaid on the CDF as a red dot. In this example, the emission is unlikely to be due to normal operations, as it falls above the distribution of emissions simulated in MAES.

Identification of emission sources comes with uncertainty, as aerial measurements are frequently not precise enough to identify a source with absolute certainty. As such, for each emission classified into a MAES failure type, a probability was assigned to indicate the likelihood that the event represented a failure and the likelihood the emission was from the correct location. Analysts reviewed aircraft-provided imagery and evaluated each detection based on the following criteria, assigning each a probability score ranging from 0 to 1:

LOCATION PROBABILITIES

- Emission onsite (binary):
  - 0 = No (emission does not appear to be on facility site).
  - 1 = Yes (emission is clearly located on the facility site).
- Plume origin at specific equipment source:

  - 0 = No identifiable concentration near specific source equipment.
  - 0.4 = Diffuse concentration observed, not clearly associated with a specific source.
  - 0.6 = Emission plume is visible but has drifted or is not clearly traceable to a specific piece of equipment.
  - 0.8 = Plume appears likely to originate from equipment source, though the source may be shared among multiple units or is somewhat ambiguous.
  - 1 = Clear emission concentration from a specific source (e.g., tank, separator, compressor).

- Plume transport quality:
  - 0 = Poor or no visible transport
  - 0.4 = Plume is present but poorly defined, with a wide, irregular, or unstable shape. Transport direction is unclear or inconsistent.
  - 0.6 = Plume is somewhat visible, though still lacking clear definition. Transport direction may be inferred but is uncertain.
  - 0.8 = Plume is faint or somewhat dispersed, but transport is still reasonably directional and consistent.
  - 1 = Plume is clearly visible, with a narrow and coherent structure that reflects strong, directional atmospheric transport.

FAILURE PROBABILITIES

- Classification of failure or normal:
  - 0 = Normal emissions, falls within the MAES CDF and/or operator noted normal operations.
  - 0.4 = Outside of MAES CDF, but measurement uncertainty could indicate within CDF.
  - 0.6 = Emission is outside the MAES CDF, but the operator reported normal operations, or emission is within the CDF, but the operator reported a failure.
  - 0.8 = Outside of MAES CDF, no operator information to confirm.
  - 1 = Operator noted failure and outside of MAES CDF.

The product p of these scores is an estimate of the probability that the source is identified correctly and is a failure event. An emission identified as a failure type with probability p is counted as p emissions when estimating the frequency of this failure type (details will be described in Section 3.2.3). In cases where the source is not confidently identified, the observed emissions are not discarded; instead, they are modeled as originating from an unknown source. Specifically, if $p < 1$, then the remaining probability $1 - p$ is used as the probability the emission is attributed to the miscellaneous emitter category. For example, if an emission is attributed to tanks with a probability of $p = 0.6$, then the remaining probability of 0.4 is assigned to miscellaneous emissions.

While the assignment of these probabilities involves some degree of subjectivity, it offers a more realistic representation of uncertainty compared to treating all detections as fully certain. This approach acknowledges the inherent variability in observational data and addresses limitations in confidently attributing emissions to specific sources. In practice, probabilities were rarely assigned a value of zero, reflecting the presence of at least some supporting evidence in most cases. Additionally, the classification of emission events as either normal or indicative of failure further illustrates how the MAES framework integrates with and depends on information reported in ONGAEIR, as the CDFs used above are from the MAES inventory model that has been designed to match ONGAEIR. The probabilities determined in this step are taken into account when estimating the frequency of emissions for use in the MAES MII model, as described in the following section.

It should be noted that this emission classification step was done for the MAES inventory models using ONGAEIR 2022. Due to time limitations, we could not go back through this step using the updated MAES inventory model with ONGAEIR 2024.

3.2.3 Estimating distributions of emissions from failures

Based on the classification of emissions described above, we estimate a probability of observing each failure type along with a distribution of the resulting emissions, both used as inputs to MAES. The use of aircraft measurements to simulate emissions in MAES makes the common "ergodic assumption" of emissions: that the distribution of emissions observed across many facility/equipment samples provides an accurate estimate of the emissions expected from a single facility/piece of equipment over a long period of time.

The probability we estimate, called pLeak, is the probability a piece of equipment is leaking at any given time. This is a useful statistic since it can be estimated from observations, and MAES simulates these leaks according to a Poisson process that ensures the portion of time spent in a failing state matches this probability. In previous studies, pLeak for a given failure type was estimated as the number of times this failure type was observed within a measurement campaign divided by the total count of equipment observed; the distribution of emissions from the failure type was approximated by the observed distribution of emissions (taking into account uncertainties from aerial measurements). Because the present data comes from six sensors across three different aerial companies, we found it necessary to use a more detailed process to estimate these probabilities and distributions, so that the different detection limits were considered. That is, because of Bridger's lower detection limit relative to the other aerial companies, we expect Bridger to have a much more accurate estimate of the frequency of low emission rates, whereas all three companies should be used to estimate high emission rates.

Rather than establish a hard cutoff of an emission rate under which only Bridger's data is used, we use the probability of detection curves for the different companies and sensors to weight the observations appropriately based on emission rate. For a small range of emission rates, the number of "effective samples" taken by a sensor is the total number of samples times the probability of detection in this range; the probability of a failure in this range is then estimated by the number of failures observed (weighted by the probability scores assigned above) divided by the total number of effective samples across all sensors. From the estimates of the probabilities in each range of emission rates, pLeak is estimated as the sum of the probabilities and the distribution of emissions is approximated as the normalized histogram of numbers of failures in these ranges. The aircraft measurement uncertainties described in Section 2.4 are used throughout by replacing each measured emission rate with its modeled distribution for the true emission rate. The end result is distributions that rely mostly on Bridger at low emission rates and gradually incorporate Insight M and GHGSat measurements as the emission rate increases. Details on the method are given in Section A.9 of the appendix, and the resulting distributions are pictured in Figure 38.

Heaters and flares are modeled mechanistically in MAES, so an emission factor is not used. Instead, the pLeak calculated for heaters and flares is used in a Markov transition matrix that calculates the probability of malfunctioning, as required by MAES (see the Supplementary Information of [18]). Because of this difference in modeling, which does not require a distribution of emission rates, and because GHGSat and Insight M observed only small numbers of heater and flare failures, we simply used Bridger's detections and sample size to estimate pLeak for these equipment types, rather than the method above. That is, in these cases, pLeak was computed as the number of failures observed by Bridger (weighted by the probabilities assigned above) divided by the number of samples of the equipment type taken by Bridger. The remaining equipment types are modeled traditionally in MAES. The distribution from the associated failure types are used as emission factors for abnormal emissions, which are simulated in the MAES MII model.

Table 3 shows the estimated values of pLeak, along with the sample sizes observed during the measurement campaign. Equipment counts for each site were obtained primarily from ONGAEIR, as described in Section 2.1. In cases where the same site was scanned multiple times, the equipment was counted once for each day scanned: reflights of facilities were predefined and therefore counted only once per day, even if a facility was captured multiple times within a short time span. This approach accounts for the fact that some aerial methods have wide scan widths, which can result in multiple detections of the same facility within minutes. For the miscellaneous category, one sample was counted for each site for each day scanned, as this agrees with the modeling of miscellaneous emissions in MAES. While the values of pLeak are dependent on the manual classification of emissions described in Section 3.2.2, a sensitivity study showed that errors in the pLeak values produced proportionally smaller errors in the final MII results; see Section 5.3.1 for a summary.

Table 3: Equipment samples and estimated values of pLeak. Here equipment has been counted once for each day scanned. For flares and heaters, only Bridger samples were used to compute pLeak: Bridger sampled 10857 flares and 35064 heaters.

	Sample size	**pLeak**
Compressors	11,015	0.0160
Miscellaneous emissions	32,865	0.0368
Flares	23,941	0.0038
Heaters	118,799	0.0026
Controlled Tanks	74,051	0.0028
Uncontrolled Tanks	26,854	0.0076

3.3 Colorado School of Mines Modeling: Measurement Based Inventory Using a Statistical Model

We now pivot to describe a fundamentally different approach to building a Measurement Based Inventory (MBI), developed by the CSM team. Unlike the MAES approach, which uses detailed facility information from ONGAEIR, we now assume no prior knowledge about site emissions and instead base the statistical model on measurement data. At a high level, this approach uses two distributions of facility-level emission rates in Colorado: one fit using aerial emissions estimates, adjusting for the differing detection sensitivities between vendors, and one that represents emission rates below the aerial detection thresholds, which is estimated in a few different ways: one using Continuous Monitoring System (CMS)-derived emissions and two additional approaches based on prior work by Williams (2025) [15] and Sherwin (2024) [17]. Details of how the emissions were derived from CMS data are provided in Section A.3 of the Appendix. We then repeatedly sample from these distributions to generate state-wide emissions estimates on any desired timescale. This work represents an early iteration of our conceptual approach; future research will investigate alternative methodologies and examine several of the assumptions currently being made in more detail.

3.3.1 Distribution modeling

As a first step in our statistical MBI model, we aim to build a facility-level emission rate distribution for oil and gas production sites in Colorado, combining data from all three available vendors. However, we cannot simply fit a distribution to all three datasets combined since each vendor has a different detection sensitivity and multiple vendors flew systems with differing sensitivities, meaning that emission rates that all three vendors are likely to see would be overrepresented in comparison to emission rates that only one or two of the vendors would be sensitive enough to detect. To solve this issue, we draw inspiration from Kunkel et al. [36], who fit an emission rate distribution to data provided by Bridger and Carbon Mapper, taking into account the varying detection sensitivities of the two technologies. A core idea behind their methodology is choosing a distribution matching cutoff (DMC) for each vendor: a facility-level emission rate above which we expect that vendor to detect all emissions, i.e. where probability of detection approaches 100%. The emission rate distribution is then fit only to rates above this DMC for each vendor. In this project, we use 5 kg/hr for anonymized company code L (Company L), 51 kg/hr for anonymized company code H (Company H), and 49 kg/hr for anonymized company code Q (Company Q). Note that while some vendors (specifically GHGSat and Insight M) use multiple sensors which in reality likely have different DMCs, we opt to find DMCs on the vendor level due to the small sample sizes of positive detections, an issue which would be exacerbated when dividing further by sensor. The DMCs for Company H and Company Q were determined by examining how well the distributions of their observed facility-level rates align with those observed by Company L above a range of cutoffs. This method is based on the assumption that all the vendors sample from the same underlying facility-level emission rate distribution, just at different detection sensitivities. If that assumption is met, their observed emission rate distributions should align above an appropriate DMC, in the emission rate regime where both vendors are detecting all occurring emissions. This assumption was tested using two-sample

Anderson-Darling tests, which test the null hypothesis that two samples come from different distributions. To account for the small sample sizes, permutation-based tests were used, and found no significant differences in distribution between facility-level emission rates observed by Company L and Company H/Company Q above their respective DMCs (p-values of 0.26 and 0.81, respectively, with significance defined as $p < 0.05$). Note that decreasing the DMCs does not immediately result in significantly different distributions, and these higher DMCs with larger p-values were chosen as conservative estimates: DMCs are not meant to represent detection thresholds, especially since they aggregate together vendor systems flown with different sensitivities. Rather, they are intended to provide a cutoff to ensure all emission rates used in the distribution-fitting process are being sampled at their true frequency, and not impacted by detection sensitivity. Even though a DMC cutoff was applied, the measured surveys did capture many emissions below the DMC threshold for all three of the aerial vendors. Determining an appropriate facility-level DMC for Company L is more challenging, as there is no reference distribution with a lower detection threshold to compare against. Instead, we choose 5 kg/hr as a reasonable estimate based on their probability of detection curves for equipment-level detections, increasing the cutoff to adjust for our DMC being on the facility-level, and the probability of detection curves being on the equipment-level (see Section A.8 of the Appendix), and provide analysis on the effects of that choice on the results in the form of a sensitivity study, see Figure 6. Note that in future iterations of this work, we will investigate more robust methods for selecting DMCs, as well as alternatives for the combination of data across vendors more generally.

We use a lognormal distribution to model facility-level emission rates, as it handles nonnegative data that is right-tailed, both of which are true of the observed emission rates. Note that a more flexible generalized lognormal distribution was also tested, but via Akaike information criterion (AIC) and Bayesian information criterion (BIC) testing, the traditional lognormal was found to perform better. The lognormal distribution has two parameters, b and x_0, and follows the density

$$p(x; b, x_0) \propto \frac{1}{x}\exp\left(-\frac{(\log_{10}x - x_0)^2}{b^2}\right).$$

The parameters are estimated via maximum likelihood estimation, and in order to fit to all three vendors' data simultaneously, their datasets are assumed to be independent, and their respective log-likelihoods are summed. Note that these log-likelihoods are calculated only using emission rates above each vendor's DMC. This fitting process results in estimated parameters of a lognormal distribution that represents the relative frequency of emission rates above the lowest DMC, in this case Company L's, 5 kg/hr. The resulting distribution can be seen in Figure 5, where observed frequencies are shown by colored shapes that differ by vendor, and the model is shown with a black line.

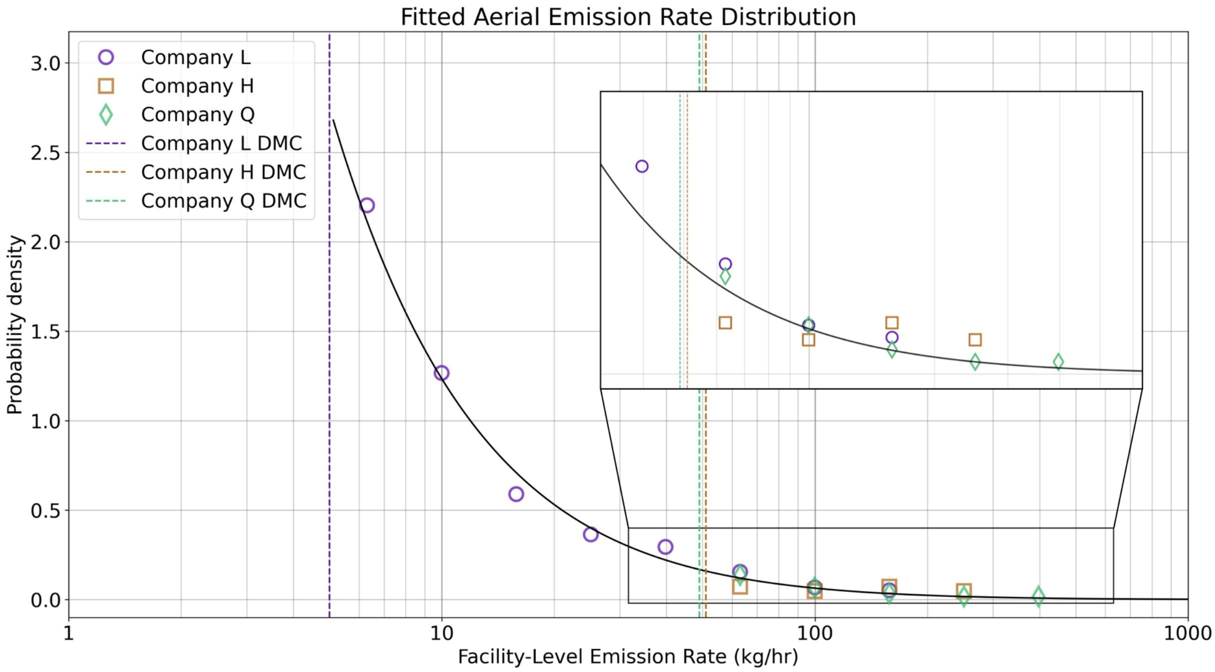


Figure 5: Estimated distribution of facility-level emission rates above 5 kg/hr for all sites in ONGAEIR. The dots for each vendor represent a histogram of the observed emission rates above the selected DMC threshold, plotted as symbols instead of bars for visual clarity. The black line indicates the fitted lognormal density. An inset is shown for the higher rates. Note the logarithmic scale of the horizontal axis.

3.3.2 Aggregation

With an estimated facility-level emission rate distribution, additional steps must be taken to arrive at a state-wide emission rate and/or mass estimate. On a high level, our approach involves segmenting the time-frame and facilities of interest into a number of "facility-hours" (the number of facilities multiplied by the number of hours in that time-frame, e.g. for an annualized inventory, 8760 hours), sampling an emission rate for each of these facility-hours, and summing the resulting rates to get a total mass estimate, which can then easily be converted into a rate if needed. Implicit in this method is the assumption that each emission lasts for an hour. However, this does not mean we think that that assumption is necessarily reflective of actual emission durations: it is simply a discretization choice, and the method is invariant to that choice: using an assumed duration of 1 minute (i.e. sampling 60 times the number of rates but dividing by 60 to get the mass emitted by each rate) led to nearly identical results, differing slightly only because of the stochasticity of the method. This method has another implicit assumption: that emissions are an ergodic process, meaning that we can use our distribution estimated from emission rates at many facilities equivalently as a distribution for a single facility over time. This is a common assumption in the methane emission aggregation literature [37, 36], and will be tested using CMS data in future work.

Since our emission rate distribution inferred from the aerial data is only valid down to a threshold of 5 kg/hr, there are two additional components we need for this method: a way to estimate the probability of an emission above 5 kg/hr occurring at any given time and a way to sample emission rates below 5 kg/hr. The first is quite straightforward: we can estimate this probability simply as

$$\frac{\text{\# of Company L estimates} > 5\text{ kg/hr}}{\text{\# of total Company L estimates, including non-detects}}.$$

Since we are treating 5 kg/hr as Company L's facility-level DMC, the assumption is that any rate at 5 kg/hr or above will be observed properly, so if that assumption is met, we should have an unbiased estimate of the desired probability. However, this method also depends on the ergodic assumption: this determination method assumes that the probability is the same for a single site over time as it is across multiple sites. The resulting probability is just over 4%.

A way to sample from below-threshold emission rates is more challenging to obtain. We consider five methods for below-threshold sampling.

1. The first is simply sampling "zeroes" (i.e. each non-detect represented zero emissions from a given facility) for all rates below 5 kg/hr. This will clearly result in an underestimate for the total mass/rate, as not all rates below 5 kg/hr are identically zero, but it serves to provide a lower bound on our estimate.

2. The second method is sampling from a uniform distribution between 0 and 5 kg/hr. This method is almost certainly an overestimate, as the literature indicates that emission rate distributions are heavily right-skewed, meaning lower emission rates are much more common than higher emission rates [38]. However, analogous to the method that samples only zeroes, this is helpful in providing an upper bound on our estimate.

3. The third method we propose for below-threshold sampling involves sampling (with replacement) from rate estimates based on CMS data. Further details about the CMS-derived emission rates in this study are provided in Section A.3 of the appendix. Note that we only sample from emission rates not captured by the aerial distribution, i.e. rates between 0 and 5 kg/hr. This method also has its downsides, specifically that the CMS data comes from only 5 facilities, all owned by the same operator, all in the Piceance basin, and all of class PS6, which likely does not generalize well to the entire state of Colorado. However, given the available data, this CMS-informed method represents a first estimate for an entirely measurement-based inventory using timely data from within the study region. Future work will involve investigation into methods to integrate these CMS-derived emissions distributions more rigorously with the aerial data, as well as the conduction CMS inference data on more sites to more accurately capture the below-threshold emissions distribution across Colorado.

4. The fourth method samples from the Denver-Julesburg-specific emission rate distribution from Williams et al. [15]. This emission rate distribution was created by assimilating many methane measurements from technologies with low detection limits ($\sim$ 0.1 - 1.0 kg/hr) within a probabilistic framework. As with method 3, we only sample from the [0, 5] kg/hr regime of this distribution. Importantly, this distribution includes emissions from both upstream and midstream facilities. As such, it very likely overestimates

emissions from just the production facilities in the DJ basin. However, this bias is likely mitigated somewhat by the fact that we only sample from the [0, 5] kg/hr regime, which is a regime more common to production facilities [15]. Nevertheless, emissions in this range do occur on midstream sites, and these emission likely tend to be larger than those on production sites, which could bias our estimates high. Furthermore, the Denver-Julesburg-specific distribution from Williams et al. is informed by methane measurements across the continental United States; it is specific to the DJ basin only through facility counts, which are used to extrapolate emissions from the site-level to the basin-level. This is an another limitation of this data source for below DMC emissions, but a sensitivity study revealed that any potential biases introduced by this assumption are minimal [15].

5. The fifth method samples from one of the Denver-Julesburg-specific emission rate distribution from Sherwin et al. [17]. Specifically, we use the Carbon Mapper Summer 2021 distribution. This distribution was created by assimilating aerial data from the very-short-wavelength infrared imaging spectrometer on the Global Airborne Observatory (GAO) with simulated emissions from the bottom-up simulation framework described in Rutherford et al. [5] to account for below detection threshold emissions on production sites and with midstream emissions information from the US Greenhouse Gas Inventory to account for below detection threshold emissions on midstream sites. As with the previous methods, we only sample from the [0, 5] kg/hr regime of this distribution. Because this distribution transitions from the bottom-up simulation tool to the aerial data at 73.0 kg/hr [17] for production sites, all of our samples from the distribution come from the bottom-up data sources rather than the Carbon Mapper aerial data. As with method 4, this distribution includes emissions from both upstream and midstream facilities. As such, it very likely overestimates emissions from just the production facilities in the DJ basin. However, as discussed for method 4, this bias is likely mitigated somewhat by the fact that we only sample from the [0, 5] kg/hr, but could contribute to the higher emissions estimates from the statistical model.

The relationship between below-threshold sampling method, choice of Company L's DMC, and estimated total emissions is shown in Figure 6, with Company L DMC on the horizontal axis and estimated methane emissions on a state-wide annual basis on the vertical axis. We include the mean annual emissions, in metric tons per year, and 95% confidence intervals (CIs) for each below-threshold method listed above. Different below-threshold sampling methods are indicated by different colored lines: sampling from a uniform distribution (a known overestimate) is indicated by a green line at the top, sampling from a CMS-informed lognormal is shown with a blue line toward the middle, sampling from a lognormal fit to the Williams data is shown with an orange line, directly sampling from the Sherwin dataset is shown with a purple line, and sampling all zeros (a known underestimate) is represented by a red line at the bottom. The methods and DMC that will be used in results figures are indicated by a dashed black box. While we do see a dependence of estimated emissions on DMC for the CMS-informed sampling method, highlighting the need for robust determination of DMC in future work, it is much less sensitive to the choice of DMC than sampling from a uniform distribution. The estimated emissions using the Williams and Sherwin datasets to inform below-threshold are very similar and show much less dependency than other methods on Company L DMC, indicating that these datasets align better with the aerial data. We also see the expected behavior of the three below-threshold sampling methods: the known underestimate is lower than our best estimates, while the known overestimate is higher.

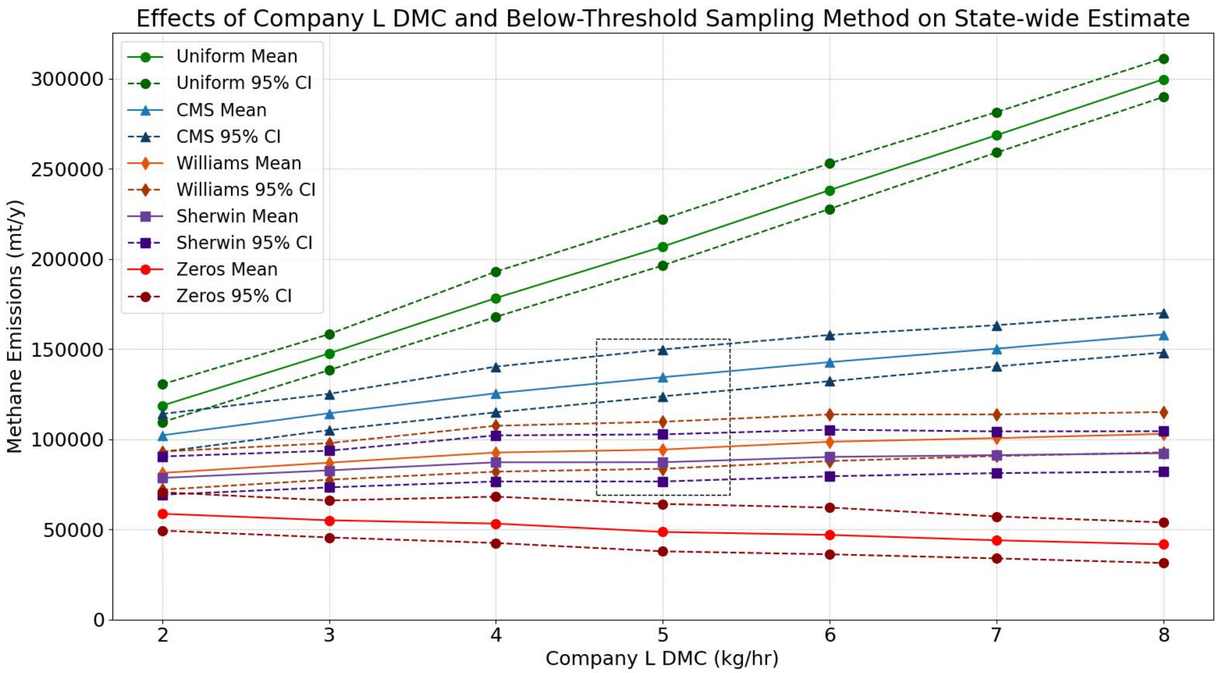


Figure 6: Effects of Company L DMC and below-threshold sampling method on estimated methane emissions for production sites with positive oil or gas production in the state of Colorado. Company L DMC is on the horizontal axis, with estimated methane emissions in kg/hr/facility shown on the vertical axis. Different below-threshold sampling methods are indicated by different colors, and a dashed black box shows the method/DMC combinations that will be shown in results figures.

Given our estimated distribution above 5 kg/hr, an estimated probability of observing a rate in that regime, and an estimated distribution below 5 kg/hr, we can now aggregate our emissions distributions into a state-wide mass estimate using the following algorithm:

Algorithm 1

Calculate n as `n_facilities` $\times$ `n_hours`
Let $\hat{p}$ be the estimated probability of observing a rate above 5 kg/hr
Let $\hat{b}, \hat{x}_0$ be the estimated lognormal parameters from the aerial data
Let $\pi(\hat{\theta})$ be the estimated distribution for below-threshold rates
Initialize `sum` $= 0$
for $i = 1, ..., n$ **do**
    Draw $P \sim \text{Bernoulli}(\hat{p})$
    if $P = 1$ **then**
        Draw $X \sim \text{Lognormal}(\hat{b}, \hat{x}_0)$
        `sum` $=$ `sum` $+$ X
    else
        Draw $X \sim \pi(\hat{\theta})$
        `sum` $=$ `sum` $+$ X
    end if
end for
return `sum`

At the end of the above algorithm, `sum` represents the total estimated emitted methane, in kg, for `n_facilities` over the course of `n_hours`. Note that `n_facilities` is derived from the ONGAEIR 2024 data, excluding facilities that MAES cannot model to ensure the MAES and statistical models are aligned, which results in 9,411 facilities. This excludes some emitting, non-producing facilities, meaning that these results are not representative of the entire state, and as such cannot be directly compared to, for example, satellite emissions estimates. Therefore, we also show results using all facilities in ONGAEIR 2024. Also note that `n_hours` can be adjusted based on the desired time-frame: for an annualized inventory estimate it is set to 8760, the number of hours in a year. Once again, this segmentation into hour-long time chunks is a discretization tool rather than a judgment on actual event durations, and our method is insensitive to the choice of an hour. To convert the resulting mass to a rate, we can simply divide by `n_hours` to achieve an estimated rate in kg/hr, which we can also convert to an estimated average facility-level rate by dividing by `n_facilities`. To account for uncertainty, we perform this process many different times within a Monte Carlo framework. Each time, we resample our aerial rate estimates with replacement to obtain different estimates for the parameters of a lognormal and the probability of observing a rate above 5 kg/hr. We then run the algorithm on every combination of these estimated parameters and probabilities, and the spread of the resulting total estimates gives us an estimate of uncertainty.

Once we have aggregated mass/rate estimates, there is an important final adjustment step. Since most of the aerial data was recorded during daytime hours, it captured maintenance events at a higher frequency than they occur when scaling to a time-frame that includes nights. Since maintenance events tend to be accompanied by higher emissions, simply extrapolating rates recorded in the daytime to an entire 24-hours period will result in an overestimation. To account for this, we adjust our final rate estimates down according to the results of a 2025 study by Barkley, et al. [39], which estimated that an extrapolation of daytime emissions to a longer time-frame results in about a 25% overestimation (with a sensitivity study indicating a reasonable range of 15%-35%).

Given enough data, this methodology can easily be used to generate emissions estimates for subsets of the state of Colorado. For example, to generate an emissions estimate for only the DJ basin, we restrict the aerial data used to rates observed on facilities in the DJ basin, and adjust the `n_facilities` input to the algorithm to reflect the number of facilities in the DJ. Ideally, we would also subset the below-threshold rate estimates to use only those generated within the desired basin, but this is not possible with our current datasets – the CMS-derived emission estimates all come from the same basin, and both Williams and Sherwin aimed only to estimate distributions in the DJ – so below-threshold distributions remain the same for all subsets of Colorado. We can make the exact same adjustments if we want an estimate for a specific PS class, restricting our aerial data to facilities of that PS class, and updating `n_facilities` to align with the number of facilities classified as the desired PS class. Note that this requires enough aerial data to adequately fit a lognormal distribution, which is not always the case. Specifically, in this report we fit to three agglomerated basins: DJ, Piceance, and Others, and to only two PS classes: PS2 and PS4, as we do not have sufficient positive aerial rate estimates for the other classes. Also note that while ONGAEIR 2024 data are used for facility counts and reference emissions estimates, ONGAEIR 2022 data are used for classifying aerial measurements into basins/PS classes, as the intensive matching process between the aerial and ONGAEIR datasets was performed before 2024 data were available.

4 Results and Discussion

This section presents a summary of the data collected in the measurement campaign and the results of the two MII processes. A previous version of this report presented results based on the 2022 ONGAEIR dataset; the MII model results given here have been updated to the 2024 ONGAEIR dataset. Additionally, methane emission data from the measurement campaign is made publicly available in an anonymized dataset, which lists the detected emissions with facility names and locations removed; see Section A.13 of the appendix. This dataset also presents results from the operator cause analysis (see Section 3.2.1), including the sources of emissions when they were identified.

4.1 Overall Campaign Data

Approximately 94% of production facilities that were operating or partially operating in the 2022 ONGAEIR dataset were scanned by at least one aerial measurement company. The breakdown by PS classification for all considered basins is shown in Figure 7. Refer to Section A.1 of the appendix for similar figures for the DJ, Piceance, and other basins. While the majority of PS1, PS2, and PS4 facilities were scanned by GHGSat, most PS6 facilities were scanned by Insight M. Bridger accounted for the majority of positive emission detections across all classes.

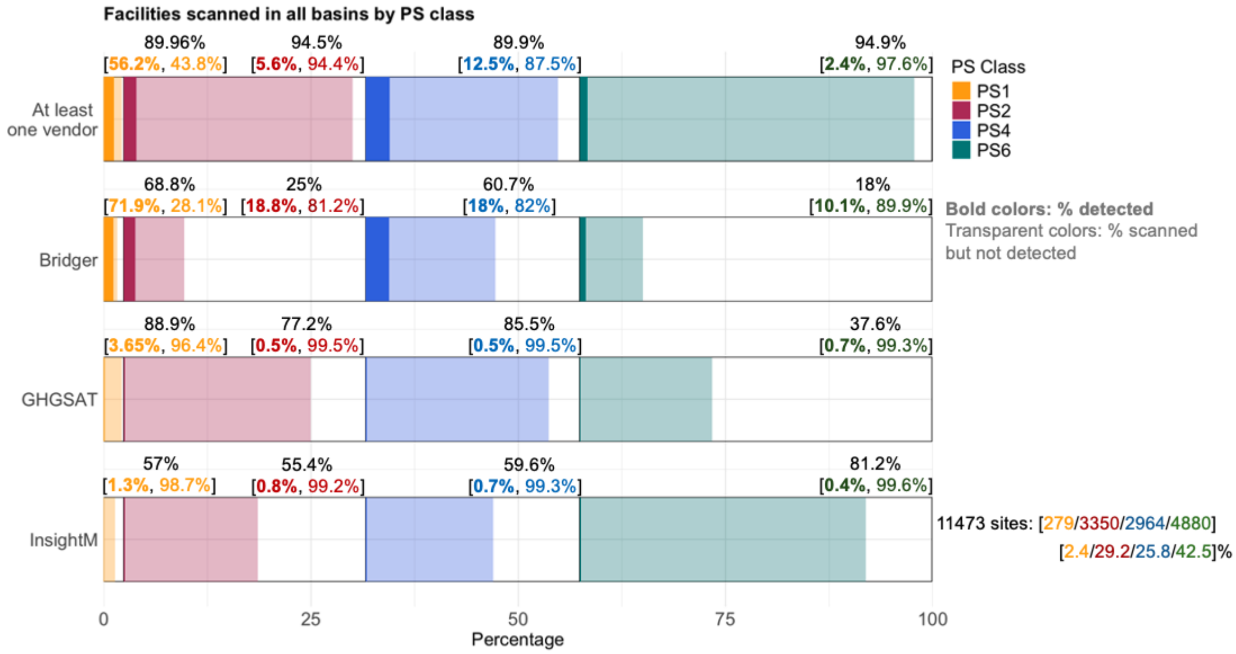


Figure 7: Percentage of facilities in all basins scanned by at least one vendor (top row) and by each vendor (subsequent rows). The percentage in black indicates the overall proportion of facilities scanned within each PS class. The bold percentage in parentheses represents the share of scanned facilities where emissions were detected, while the regular-font percentage shows the share of scanned facilities with no detected emissions. Percent colors correspond to the associated PS classes.

Insight M surveyed the largest share of ONGAEIR facilities among the aerial companies, surveying 7,749 sites, representing 68% of all sites in the ONGAEIR dataset. GHGSat scanned about 63% (7,209 sites), while Bridger covered approximately 32% (3,708 sites) of the ONGAEIR facilities. All three companies covered the DJ, Piceance, and other basins. Table 4 indicates the number of total scans broken out by unique facilities and repeat facilities per aerial company.

Aerial Company	**Total Scans**	**Unique Facilities**	**Repeat Facilities**
Bridger	7,043	3,708	1,836
GHGSat	10,915	7,209	3,057
Insight M	15,127	7,749	4,296
Campaign Total	33,085	10,771	7,732

Table 4: Summary of facility scans by aerial company

In total, 2,102 emissions events were detected in the COBE measurement campaigns. Emission events are reported differently across the measurement platforms; we summarize here, and more details are given in Section A.2 of the appendix. For Bridger, emissions are reported at the source level. If multiple emissions are detected from the same source within a single day, they are averaged to generate a single source-level emission rate for that day. A single facility may have multiple emission sources, and each source is treated as a separate emission event in the dataset. Insight M reports emissions at the facility level. GHGSat primarily reports emissions at the facility level, although in a few cases, multiple clearly distinguishable plumes were detected and reported as separate emission events. In cases where GHGSat or Insight M detected facility-level emissions more than once in a single day, each emission event is retained as a separate entry in the dataset. See Figure 8 for the number of emission events by facility class and aircraft company.

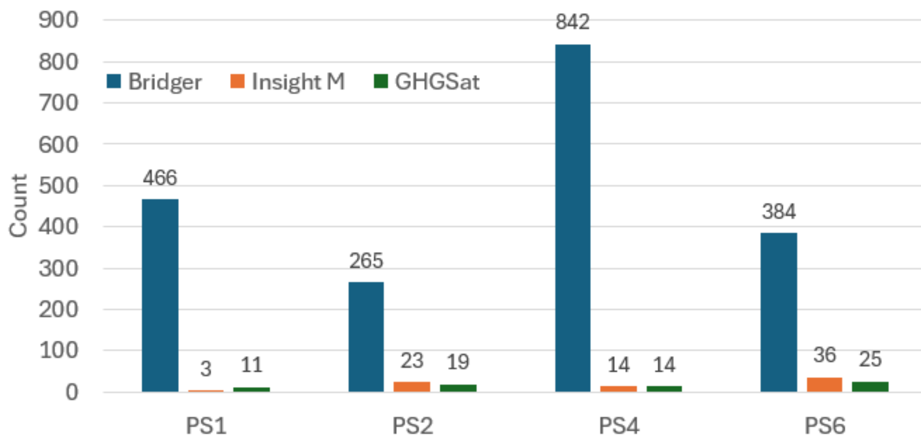


Figure 8: Summary of count of emission detections by PS and aircraft company.

At the facility level, Bridger reported more facilities with positive emissions than the other two companies, which can be explained by Bridger's lower detection limits. Facility-level detected emission rates varied by company and basin. When scanned by Bridger, 88.6% of surveyed facilities had no emission detected in the DJ basin, 69.1% in the Piceance basin, and 95.3% in other basins. More facilities were reported as having no detected emissions by GHGSat with 99.6% in the DJ, 99.7% in the Piceance basin, and 97.8% in other basins. Most of the aerial measurements conducted by Insight M resulted in no emissions detected, accounting for 99.6% of surveyed facilities in the DJ basin, 99.3% Piceance basin, and 99.4% in other basins. Summary statistics of facility-level detected emission rates in three basins by vendors is shown in Table 5.

4.2 Emission Factors

Emission factors incorporating the aerial measurements were developed from the MAES MII results. Emission categories were disaggregated to align with Bridger's major equipment groups: flares, heaters, compressors, separators, and a "miscellaneous" category. Each emission category encompasses multiple emission sources. For each facility within a given PS class and equipment type, total emissions were aggregated across each unit of that equipment type for each MC iteration. The summed values that were positive were then used to construct the distribution of annual emissions for that facility-equipment-PS combination, giving the equipment group's emission factor. The distributions for each PS and equipment type are shown in Figure 9 as violin plots, with embedded mini box plots indicating the median and

Table 5: Summary of facility-level detected emission rates measured in kg/hr by aerial measurement company and basin.

Company	Basin	Median	Average	Min	Max	Range
Bridger	DJ	2.13	5.33	0.203	189	188
	Piceance	1.53	3.96	0.135	81.9	81.7
	Other	2.09	5.39	0.203	43.7	43.5
GHGSat	DJ	105	118	34	248	214
	Piceance	24	57.3	10	157	147
	Other	29	46.5	8	285	277
Insight M	DJ	36	113	7	353	346
	Piceance	43	49.4	3	143	140
	Other	17	33	3	114	111

interquartile range. The distributions tend to be heavily skewed to the right. See Table 17 in Section A.11 of the appendix for the mean and quartiles.

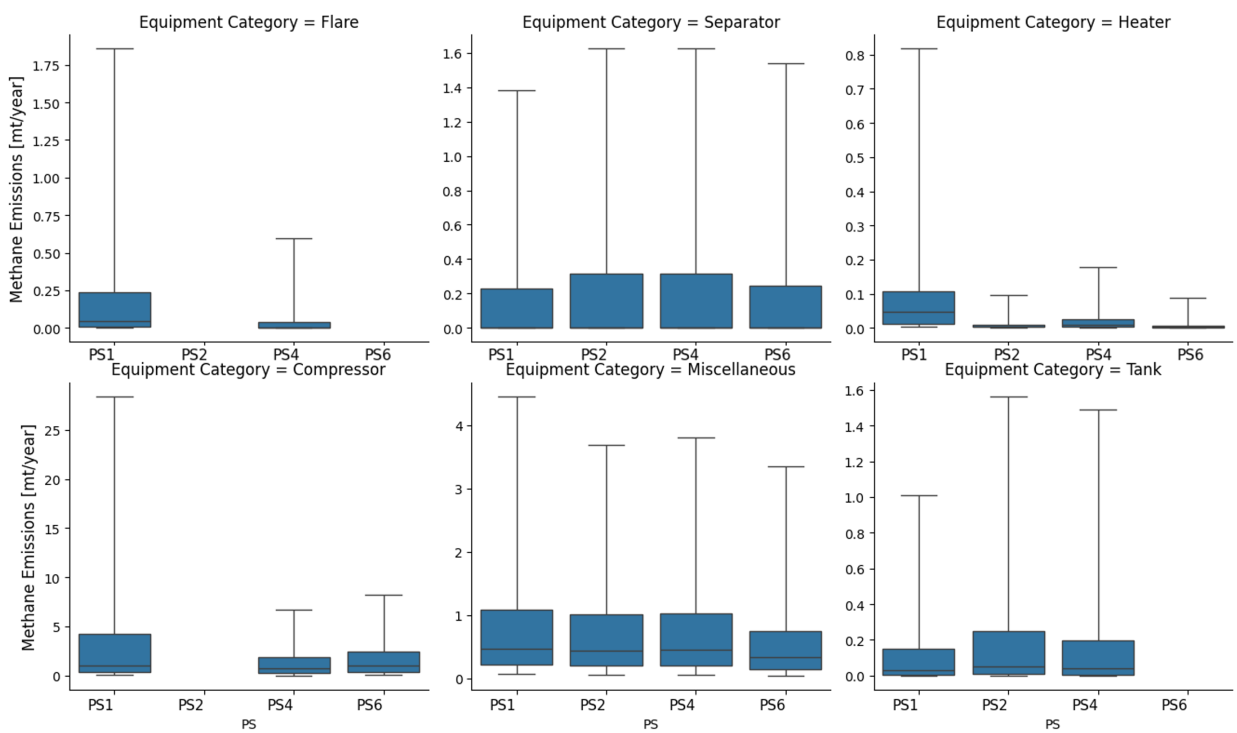


Figure 9: The distributions for each PS and equipment type are shown as box plots.

The emission factor for PS1 compressors is higher compared to other PS classes. This difference can be attributed to the use of gas lift systems within PS1, which involve larger horsepower engines. This category also had the highest number of 4-stroke-lean-burn (4SLB) engines, which are known to emit more than 4-stroke-rich-burn (4SRB) engines: according to AP-42 emission factors, 4SLB engines have an emission factor 5.4 times higher than 4SRB engines [34].

4.3 MAES Model MII Results

The MAES MII model produces emission estimates informed by aerial measurements, which we compare to the MAES inventory model and to ONGAEIR. Emissions data in this section are taken from the 2024 ONGAEIR dataset, which became available during the writing of this report. A total of 9,411 sites were modeled in MAES, roughly 81% of the 11,681 upstream sites reported in ONGAEIR that were operating of partially operating. For discussion of the unmodeled sites, see Sections A.15 and A.16 of the appendix. We begin with comparisons that exclude maintenance-related emissions, as these are not modeled in MAES. As described in Figure 3 and Section 3.2, the first step in the MAES MII process is to compare the inventory model to the reported inventory (ONGAEIR, adjusted by removing those emission categories not modeled in MAES – see Section 3.2). The MAES inventory model total is 27,181 mt/y compared to the adjusted ONGAEIR total of 26,415 mt/y (with maintenance equipment emissions of 2,339 mt/y excluded). The MAES MII model total is 36,597 mt/y, which indicates an increase of 52% from the adjusted ONGAEIR, attributable to failure events.

These results are summarized in Figures 10 and 11. The brackets in the figures show the 95% confidence intervals for the distributions of values across the multiple MC iterations in the MAES simulations.

Since MAES does not estimate emissions from maintenance events, to get a total estimate for Colorado, the total ONGAEIR emissions from maintenance, 2,339 mt/y, were added to the MAES MII. These maintenance-related emissions increase emissions by 9% in the ONGAEIR inventory and by 6% in the MAES MII model. Additionally, emissions from dehydrators, NR internal combustion engines, and pneumatic pumps were not modeled and are added to both the ONGAEIR and MAES model totals. The total MAES MII estimate plus ONGAEIR maintenance emissions is 38,936 mt/y. This leads to a state-wide ratio of **1.47** when compared to the ONGAEIR total of 26,415 mt/y. When broken down by basin, emissions totals due to failure-related events increase by 58% in the DJ Basin, 27% in the Piceance Basin, and 53% across all other basins.

The following sections summarize the MAES MII model results by equipment type, basin, and PS class. For these more detailed analyses, maintenance events are again excluded to provide direct comparisons of the types of emissions simulated by MAES.

4.3.1 Comparison by equipment

An important step in the MAES process is to evaluate whether the model accurately represents emissions at the equipment level. To do this, emissions from the MAES inventory are compared to the adjusted ONGAEIR data, grouped by equipment category. Figure 10 shows that the MAES inventory and adjusted ONGAEIR agree by equipment categories with a few exceptions, discussed below. One of the largest emission sources in both the adjusted ONGAEIR and the MAES model is pneumatic controllers. In the MAES model, 68% of emissions attributed to pneumatic controllers are associated with separator pneumatic emissions. The next largest contributors in both the adjusted ONGAEIR and the MAES model are fugitive emissions and compressor-related sources.

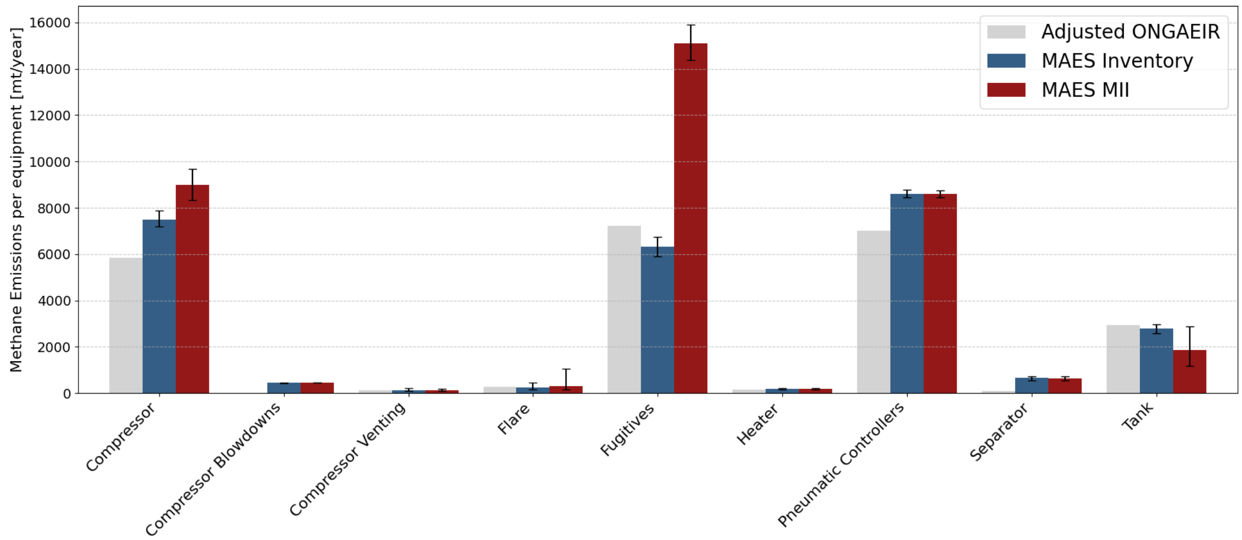


Figure 10: MAES MII and inventory results: state-wide annual emissions by equipment type. Adjusted ONGAEIR and the MAES inventory model are compared to evaluate how MAES models normal emissions, while the MAES MII model shows the increased emissions resulting from incorporating aircraft measurements into the model.

Another difference between ONGAEIR and MAES results relates to compressor emissions. As noted in Section 3, one adjustment was applied to ONGAEIR to enable a consistent comparison: when operators used the Subpart C emission factor for combustion, those values were scaled to the updated Subpart W emission factor, increasing the ONGAEIR estimate by 2,163 mt/y. In addition, the 2024 ONGAEIR reporting requirements did not include crankcase vent emissions, which are incorporated in the MAES model. When comparing ONGAEIR compressor totals to MAES (inventory) combustion and seal-vent estimates, the two agree within 13%.

The fugitive emissions category differs between models: in the inventory-based approach, emissions include all component leaks and other miscellaneous leaks, whereas the MII model additionally incorporates large emitters from the wellpad, which are informed by miscellaneous emitter data derived from aircraft observations. This highlights that the majority of the aircraft observations were from fugitive emissions and they are likely underreported in ONGAEIR.

Tank emissions in MAES inventory were aligned to match ONGAEIR emissions, and in the MAES MII model, tank emissions were informed by aerial emissions. It could be that tank emissions are low and therefore were missed by the aerial methods due to their detection limits. Tank emissions in MAES include contributions from both controlled and uncontrolled tanks, modeled using the traditional emission factor (EF) times activity factor (AF) approach. As previously noted, these EFs were developed based on COBE campaign measurements for emission rates exceeding 2 kg/hr. Separate EFs for controlled and uncontrolled tanks may capture emissions from both routine operations (e.g., tank flashing) and upset conditions (e.g., overpressure events, dump valve releases). Emissions from controlled tanks are likely underestimated, as MAES currently simulates only direct tank venting. In reality, overpressure events may also lead to excess gas being routed to the flare, depending on the volume of gas released upstream during the upset. This may increase combustion slip from the flare due to higher gas throughput. The modeling team is actively working to resolve this limitation.

4.3.2 Results by basin and prototypical site class

Figure 11 presents the results of the MAES inventory and MII model as stacked bar charts, with emissions aggregated by equipment type. These are compared to the ONGAEIR reported emissions for each basin and PS, all expressed in metric tons per year. The distinguished equipment types from the MAES results include several different emission sources, as follows:

- Compressor-related emission sources in the MAES model include: compressor blowdown events; blowdown vent leaks; component leaks; pneumatic emissions; rod packing emissions from large emitters (included only in the MII model); rod packing venting during non-operating depressurized (NOD), non-operating pressurized (NOP), and normal operating (OP) conditions; crankcase emissions; and emissions from compressor driver exhaust.
- Flare emissions are attributed to component leaks, flared gas during malfunction and normal operations, and unflared gas.
- Heater emissions originate from both operating and malfunctioning heaters.
- Fugitive emissions include leaks from miscellaneous equipment, pneumatic emissions from miscellaneous sources, and, in the MII model, wellpad large emitters.
- Separator emissions consist of component leaks and pneumatic emissions.
- Tank emissions include component leaks, pneumatic, tank flash and overpressure venting (the latter included only in the MII model).
- Wellhead emissions include component leaks and pneumatic emissions at the wellhead.
- Other category are emissions that are in ONGAEIR that are not modeled in MAES and therefore are added onto the MAES results. This includes emissions from dehydrators, NR internal combustion, and pneumatic pumps.

In the ONGAEIR hatched bar in Figure 11, pneumatics are shown as a standalone category. In the MAES results, however, pneumatic emissions are incorporated into their respective equipment groups (compressors, separators, wellheads, and tanks). Most pneumatic emissions in MAES fall within the separators category.

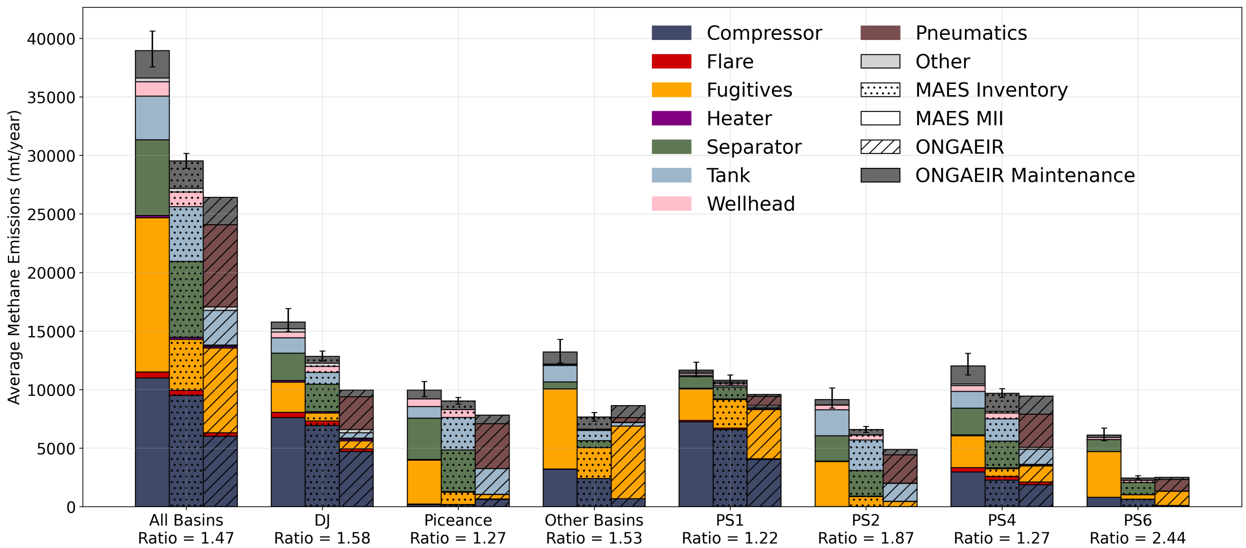


Figure 11: MAES inventory and MII model results by basin and PS compared to reported ONGAEIR values. ONGAEIR (without maintenance) totals are the hatched bars, the MAES inventory (hatched dots), and MII model, all broken out by equipment type. The ONGAEIR maintenance total (dark grey) is added to all estimates. All estimates are shown as annual estimates in metric tons per year.

The ratios between the MAES MII estimates and ONGAEIR totals vary across basins and PS classes, ranging from 1.22 to 2.44. These ratios are calculated by dividing the total MAES MII emissions by the total ONGAEIR emissions for each subset of facilities. As shown in Table 1, the Piceance Basin is composed primarily of PS2 facilities and has the lowest number of PS1 facilities. Only 99 stationary natural gas engines were reported in the Piceance Basin in the 2024 ONGAEIR dataset, likely contributing to the comparatively low emissions from the MAES inventory observed in this basin. In contrast, the DJ Basin contains the highest number of compressors and is predominantly composed of PS4 facilities, leading to higher MAES annual emissions. Statewide, fugitive emissions show the largest increase between ONGAEIR and the MII, increasing by approximately 6,000 mt/y (an 80% increase). This highlights that in ONGAEIR, fugitive emissions are the category that is the most under-reported.

Since MAES simulates duration and rate for each emission, we also summarize the MII model results by emission rate. Figure 12 is a stacked bar graph showing how three ranges of emission rates contributed to the total amount emitted; results are averaged to an hourly emission rate per site. The distribution of rates suggests that emissions in Colorado are dominated by relatively small emission rates. The figure also shows that PS1 sites have the highest average emission rate, which is expected given the use of gas-lift compression. There are fewer than 300 of these sites in the state, so they contribute little to the statewide total. Overall, the figure highlights that the average production facility in Colorado emits very little methane, typically less than 1 kg/h. For a more detailed view of the emissions distributions estimated by MAES, Section A.7 of the appendix shows CDFs of emission rates.

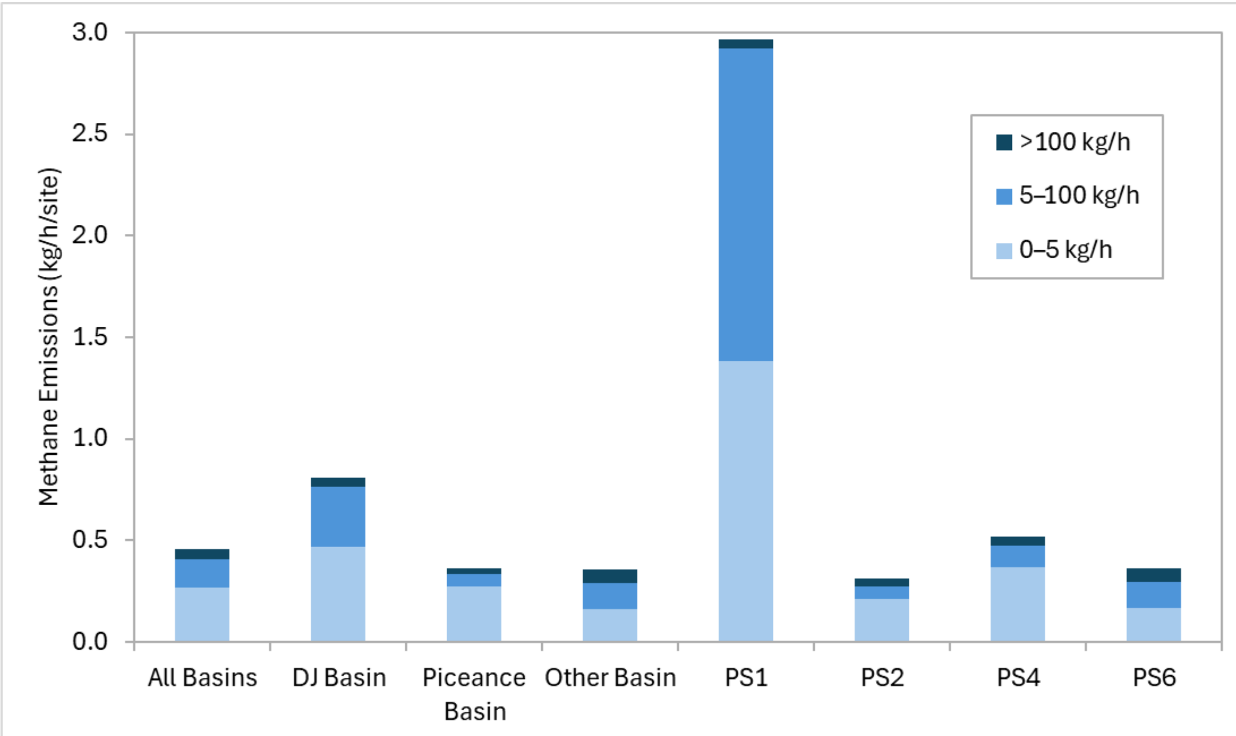


Figure 12: Average site MAES MII model results by basin and PS divided into the contributions from rates above 100 kg/h (dark blue), from rates between 5 and 100 kg/h (blue), and from rates less than 5 kg/h (light blue). The large rates of PS1 are attributed to gas lifts.

4.4 Statistical Model MBI Results

The results of the statistical MBI model yield emissions estimates for the state of Colorado (specifically for upstream facilities in the ONGAEIR database), as well as for certain subsets: facilities in the DJ basin, Piceance basin, and other basins, and facilities classified as PS2 and PS4. Note that these results were calculated using ONGAEIR 2024 facility counts and are compared against emissions reported in ONGAEIR 2024, but ONGAEIR 2022 data are used to classify aerial measurements into basins/PS classes. For ease of comparison between subsets, we report estimated emissions in units of kg on a per-facility, per-hour basis. ONGAEIR-reported emissions are also converted into the same units by taking the total amount of methane emissions reported in ONGAEIR (or for the relevant subset of ONGAEIR), converting these to kilograms, normalizing by the number of facilities, and dividing by the number of hours in a year. A comparison of measurement-derived rate estimates with ONGAEIR-reported rate estimates is shown in Figure 13, with estimates provided both on the state level and for the subsets described above. An alternate version is provided in Figure 14 in units of metric tons per year, and on a state/basin level instead of on a facility level, for direct comparison to MAES estimates.

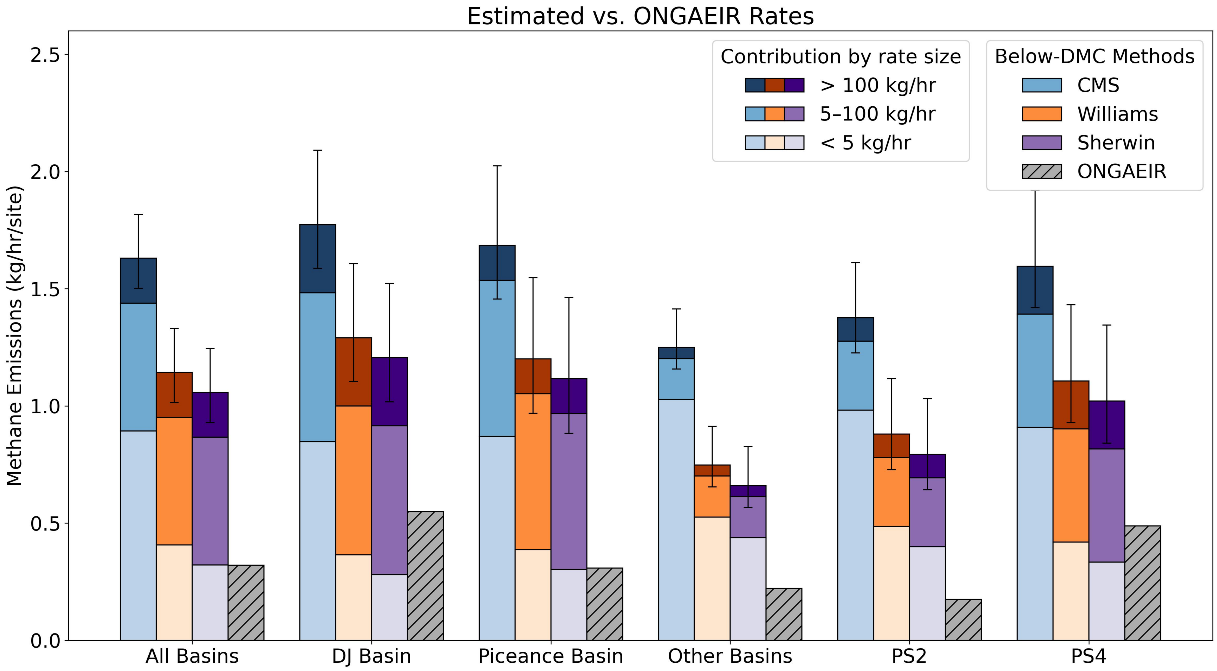


Figure 13: Comparison of measurement-derived rate estimates with those reported in ONGAEIR. Facility counts and ONGAEIR emissions are from ONGAEIR 2024 data, excluding facilities that MAES cannot model. The horizontal axis indicates the subset that the corresponding bars represent, with each subset containing four bars: 3 that represent our estimates using different below-threshold sampling methods (CMS, Williams, and Sherwin, respectively), and one grey hatched bar that represents the ONGAEIR-reported rate. The measurement-derived rates are further divided into the contribution from rates above 100 kg/hr, from rates between 5 and 100 kg/hr, and from rates less than 5 kg/hr. The 95% confidence interval for the measurement-derived rates is represented by a black interval at the top of the measurement-derived rate estimates.

Figure 13 shows that the measurement-derived emissions estimates are consistently higher than those reported in ONGAEIR, with overall ratios of 5.09, 3.57, and 3.30 using the CMS-based, Williams, and Sherwin distributions, respectively, meaning that the average per-facility measurement-derived rate is approximately 3 to 5 times as large as the rate reported in ONGAEIR. These ratios vary across basins, with a range of 3.23 to 5.64 when using a CMS-informed distribution, and a range of 2.19 to 3.90 when using the other two distributions. The ratio differs notably for the two PS classes, with PS2 showing much higher ratios between 4.54 and 7.87, whereas PS4 shows lower ratios between 2.09 and 3.27. This difference in ratios is primarily due to the much lower ONGAEIR-reported rate present in the PS2 class. Note that results for PS1 and PS6 are not shown here as there were not enough positive detections available to reliably model these classes. See Table 6 for ratios for every below-threshold distribution and subset combination.

	CMS	Williams	Sherwin
All Basins	5.09	3.57	3.30
DJ Basin	3.23	2.35	2.19
Piceance Basin	5.46	3.90	3.62
Other Basins	5.64	3.38	2.98
PS2	7.87	5.04	4.54
PS4	3.27	2.27	2.09

Table 6: Ratios between estimated emissions using different below-threshold sampling distributions and ONGAEIR reported emissions for Colorado and subsets of Colorado. For example, 5.09 in the upper left cell of the table indicates that estimated emissions when using CMS-informed rate estimates were 5.09 times higher than ONGAEIR-reported emissions.

Also of note is the distribution of the contributions of different rate magnitudes within the measurement-derived rate estimates. We see that across the board, below-threshold rates (i.e. below 5 kg/h) contribute a large portion of emissions, although the proportion varies between sampling methods, highlighting the importance of developing robust methods for estimating the distribution of these below-threshold emissions in future work. We also see that rates above 100 kg/hr contribute approximately 1/5 to 1/3 of the emissions for above-threshold rates, varying slightly across subsets of Colorado. Tables corresponding to these results can be found in the Appendix, Tables 8 through 16: one table per below-threshold sampling distribution.

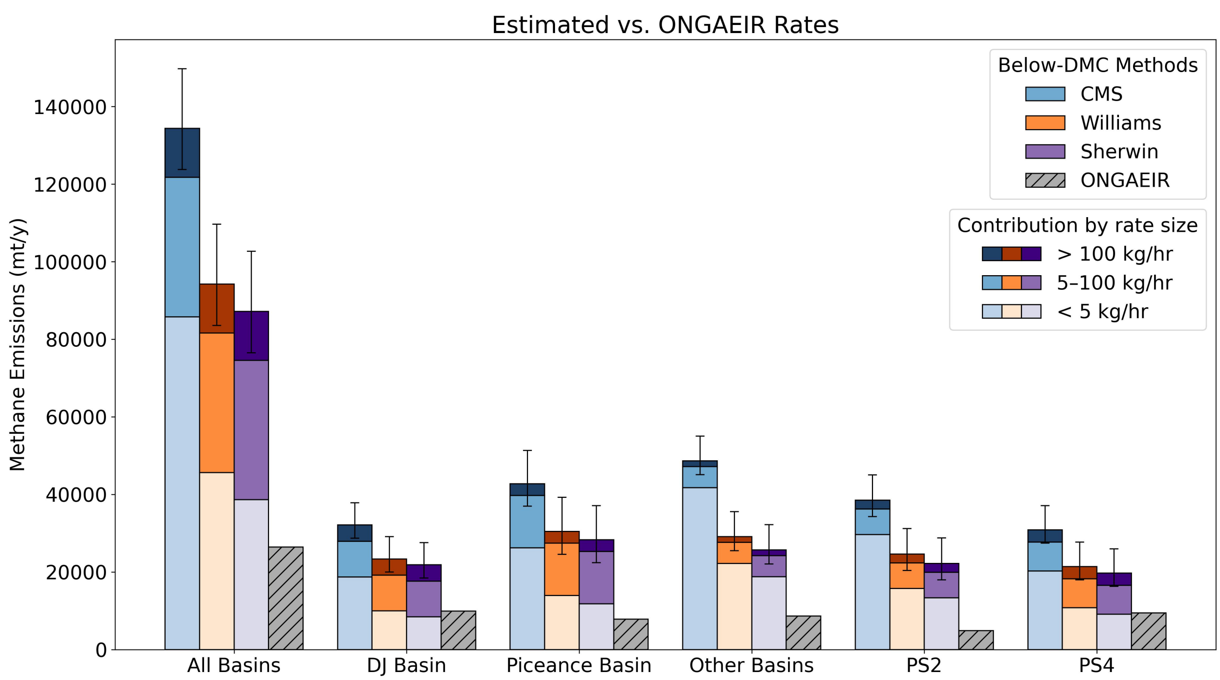


Figure 14: Analagous to Figure 13 but on the state/basin level instead of on the facility level and with vertical axis units of metric tons per year.

Figure 14 shows the same results as Figure 13, but scaled according to the number of sites in each subset. Note that these results are based on a filtered ONGAEIR dataset containing only facilities that can be modeled by MAES, and as such are not representative of the entire state or basins: there exist emitting, non-producing facilities in Colorado that are not captured here. We see that each the Pieceance basin contributes more emissions than either of the other two agglomerated basins, with other basins contributing the least, and that the distribution of sizes of rates within these contributions differs notably. For example, the emissions from the DJ and Piceance basins are made up of more rates above 5 kg/hr and above 100 kg/hr compared to those from other basins. Tables corresponding to these results can be found in the Appendix, Tables 11 through 13: one table per below-threshold sampling method.

These results (and those shown in this figure) are in the same units as, and can be directly compared to, those in Figure 11. A version of Figure 14 normalized by natural gas and oil production can be found in Section A.5 of the appendix.

4.4.1 Results using all ONGAEIR facilities

Previous figures have used facility counts and reference emissions from the ONGAEIR 2024 dataset, excluding facilities that cannot be modeled by MAES. Here, we present results using the full ONGAEIR 2024 dataset. Note that we do not individually model these previously excluded facilities; rather, we increase the facility counts accordingly and add their reported emissions into the ONGAEIR-reported total. Figure 15 shows these results in kg on a per-facility, per-hour basis, while Figure 16 shows them in metric tons per year on a state/basin level.

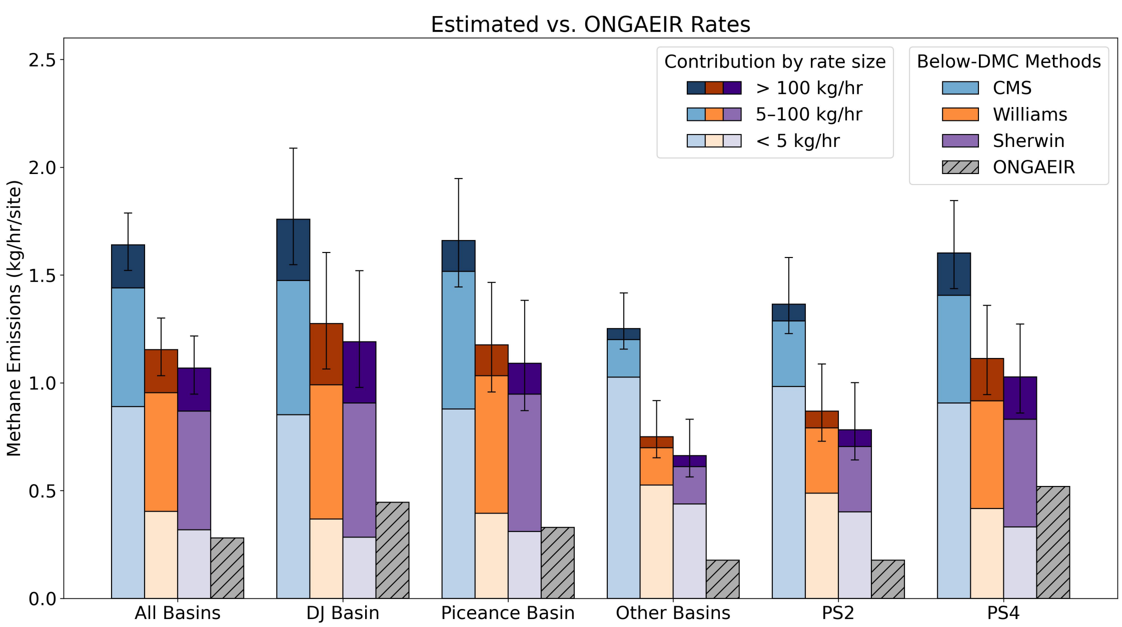


Figure 15: Comparison of measurement-derived facility-level emission rate estimates with those reported in ONGAEIR. Facility counts and ONGAEIR emissions are from the full ONGAEIR 2024 dataset, including facilities that MAES cannot model.

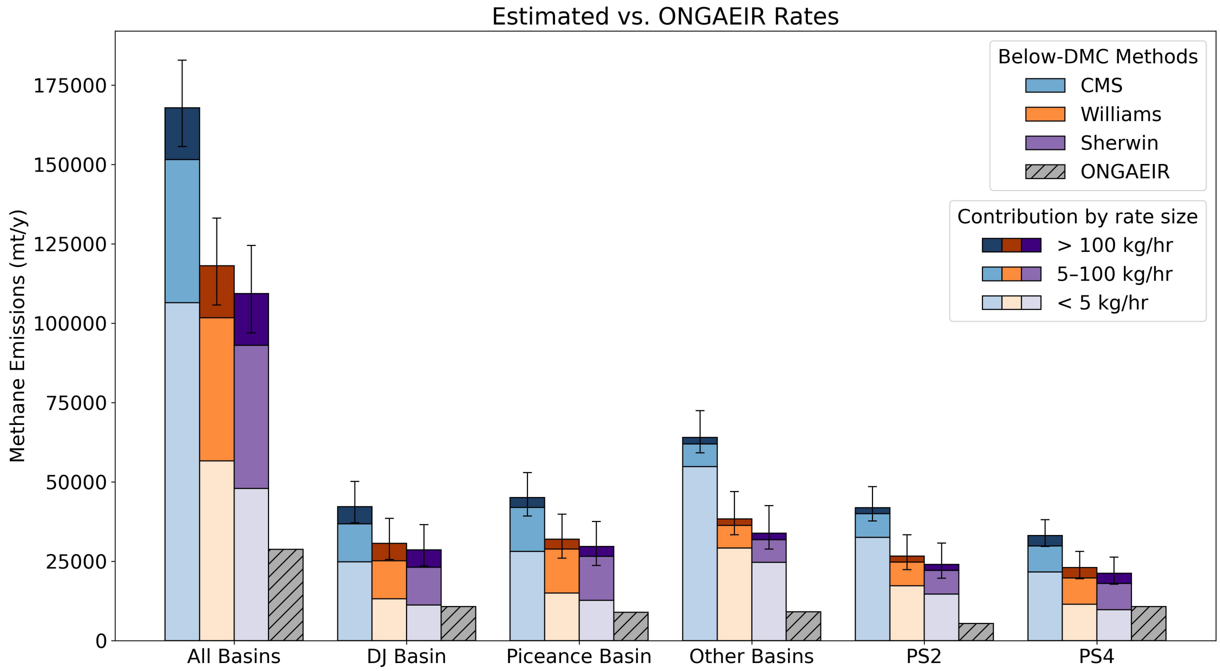


Figure 16: Analagous to Figure 15 but on the state/basin level instead of on the facility level and with vertical axis units of metric tons per year.

Figures 15 and 16 show somewhat similar results to their analogues using only the ONGAEIR facilities that MAES can model, but with some key differences. The total emissions estimates are notably larger in Figure 16 than in Figure 14, since 2,270 more facilities are being modeled. The ratios, shown using all ONGAEIR 2024 data in Table 7, also change notably. Overall ratios increased, indicating that the facilities unable to be modeled by MAES had lower emissions on average. However, this is not the case for all subsets: for example, ratios in the Piceance basin decreased when including all ONGAEIR 2024 facilities, meaning that the facilities excluded from the Piceance had higher emissions than the other facilities in the Piceance. For some other subsets, for example PS2, ratios did not change significantly.

	CMS	Williams	Sherwin
All Basins	5.85	4.11	3.81
DJ Basin	3.94	2.86	2.67
Piceance Basin	5.04	3.57	3.31
Other Basins	7.06	4.23	3.74
PS2	7.70	4.90	4.41
PS4	3.09	2.14	1.98

Table 7: Ratios between estimated emissions using different below-threshold sampling distributions and ONGAEIR reported emissions for Colorado and subsets of Colorado. These ratios use facility counts and reference emissions from the full ONGAEIR 2024 dataset.

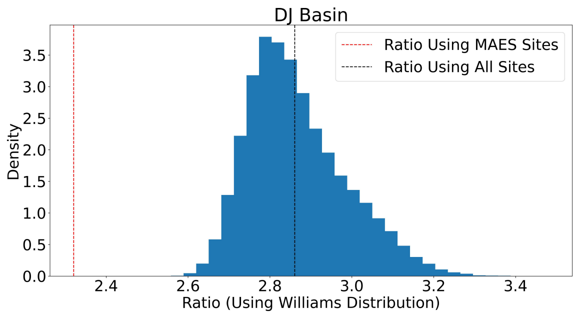
4.5 Influence of sites not modeled in MAES

Since MAES was unable to model a proportion of ONGAEIR facilities (due to a lack of information), we perform a comparison between the ONGAEIR-reported emissions from modeled and unmodeled facilities to examine the differences between these sites. Figure 17 shows the results of this comparison for the DJ Basin; figures for other subsets, as well as for all basins, can be found in the Appendix, section A.16.

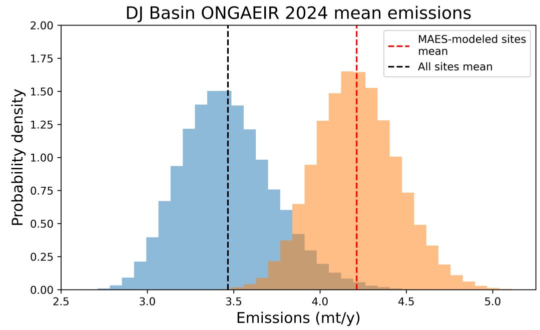
Figure 17b shows the mean emissions reported in ONGAEIR 2024 of all facilities in the DJ Basin, as well as the mean emissions of those modeled in MAES. The distributions shown are bootstrapped distributions for these means, found by resampling the data 50,000 times. Using the bootstrapped distributions, the probability for the basin mean exceeding the MAES-modeled mean was estimated to be $p = .022$. This suggests a significant difference in **reported** emissions, with MAES-modeled sites reporting higher on average.

Figure 17a shows the distribution of ratios between the statistical model's estimated emissions and ONGAEIR's reported emissions for different random subsamples of the DJ Basin. In the DJ, 2,068 out of 2,741 were able to be modeled by MAES. Therefore, an estimate for the DJ Basin is calculated using the statistical model (specifically using the Williams distribution for below-threshold rates) assuming 2,068 sites, and then for each repetition, 2,068 random sites are selected to calculate the ONGAEIR-reported emissions, and a ratio between the two is calculated. The distribution of 50,000 of these ratios is shown in the blue histogram, with the ratio using all 2,741 sites shown as a black dashed line and the ratio using the specific 2,068 sites modeled by MAES shown as a red dashed line. The ratio using the MAES-modeled sites is lower than the ratios from any of the randomly sampled subsets of the same size. This indicates that in terms of the effect on the statistical model's ratio (which is a function of ONGAEIR-reported emissions), the unmodeled sites are significantly different from the sites as a whole in the DJ Basin.

For the MAES model, the effect of the unmolded sites is less clear. In MAES, reported emissions and modeled emissions are not directly correlated: because MAES relies on mechanistic models to estimate emissions from fluid flows and equipment states, we cannot determine how the differences in reported emissions above would influence modeled emissions. Therefore the effect of the unmodeled sites on the emission ratio produced by MAES cannot be determined, as the effect on the numerator is unknown.



(a) Distribution of statistical MBI ratios for the DJ Basin resulting from random samples of facilities of the same size as the number of facilities modeled by MAES. The ratio using all sites is shown with a dashed black line, and the ratio using the MAES-modeled sites is shown with a dashed red line.



(b) Mean emissions as reported in the ONGAEIR 2024 dataset for the DJ basin, shown for both all sites and the subset that were modeled in MAES. The blue distribution is a bootstrapped distribution for the mean for the whole basin, and similarly the orange distribution is for the subset modeled in MAES. The probability for the basin mean exceeding the MAES-modeled mean was estimated to be $p = .022$.

Figure 17: Subsampling study results for the DJ Basin.

5 Cohesive Analysis and Future Work

COBE's project design has enabled the development of large-scale, high-quality MIIs. Novel features of the project include the largest dataset of upstream facilities collected via aerial measurements in Colorado, contracting with multiple aerial companies, a blend of participating and non-participating operators, and development of two models that use the same underlying measurement data. This section will take a step back from the detailed methods and results of the report to consider how the project worked as a whole. We will discuss strengths and suggestions for changes and improvements for future campaigns similar in scope conducted by the state or other entities, including opportunities for future work.

5.1 Measurements

COBE funded three aerial platforms, Bridger, GHGSat, and Insight M, to conduct the project's measurements. This represents the first time that multiple aerial imagers were deployed on such a large scale. In lieu of different information, APCD and the project team agreed to provide equal funds to each company. Each company has different business models and flight capabilities per dollar, and as expected each company flew different numbers of unique and repeated facilities (Table 4).

The key differences in technological capabilities with respect to the way the modeling team used the data were detection limits, aerial imagery quality, and total facility coverage per aerial company. As the modeling team used two different modeling approaches (the METEC MAES model and the CSM statistical model), we will break out our discussion for each model, as needed.

- **Detection limits:** Bridger had the publicly reported lowest lower detection limit (LDL) of the three aerial companies. The majority of emissions detected in the aerial campaigns were detected by Bridger (Figure 7), indicating that the majority of emitters in Colorado are relatively small. Approximately 93% of the emissions detected by Bridger that were categorized into a MAES failure type were below 10 kg/hr, indicating that many upset conditions are relatively small emitters.

- **Aerial imagery quality & data reporting:** The CSM statistical model did not use aerial imagery as no emission classification was used in their methods. The MAES model used the aerial imagery extensively for emissions classification purposes, as MAES is intended to model emissions at the emitter level with as much specificity as possible. The imagery was shared with participating operators to help operators narrow down potential emission causes. The METEC team used the imagery to assist with further validation of operator notes. Bridger had the highest-quality aerial imagery at the time of the aerial campaigns and was most often able to assign emissions down to equipment level (only 15% of detections were assigned to an "other" equipment category). GHGSat and Insight M did not include equipment localization in their detections, although participating operators were able to determine the emitter down to the equipment level in some of their cause analyses.

- **Total facility coverage:** GHGSat and Insight M were able to scan significantly higher numbers of facilities than Bridger. As a result, GHGSat and Insight M were more likely to catch large, rare emitters, and this was borne out in the data: Table 5 shows that for each scanned region, GHGSat and Insight M consistently saw larger emissions. Additionally, GHGSat saw an emission rate of over 3,000 kg/hr on a facility that did not report to ONGAEIR (Section A.10).

In addition to these considerations above, a key output of the CSM's statistical model is the prediction that well over half of total emissions in Colorado are from <5 kg/hr emitters. There are two competing factors to consider here:

- The CMS-derived emissions distribution used was likely not representative of true <5 kg/hr emissions rates for Colorado. The CMS dataset was limited in statistical representation in number (5 facilities), location (only from one basin), and facility representation. To attempt to address this limitation for COBE, the Mines team updated their analysis to consider the Williams and Sherwin papers. The results of using these two studies led to lower predicted contributions from <5 kg/hr emissions, indicating that the original Mines model using CMS-derived rates may be overestimating this contribution. In COBE-2, additional CMS data is anticipated to be collected and used to derive rate estimates that better represent <5 kg/hr emissions in Colorado across site types.

- ONGAEIR may be under-estimating (under-reporting) the smallest (< 5 kg/hr) emissions (Figure 13, 14). The METEC team chose to use ONGAEIR as MAES's base for reported emissions and classified most aerial emission detects that align with reported emissions as already within the inventory. This assumption means that the METEC team may have discarded emissions detections that were not actually reported to ONGAEIR. This is a limitation in general of the MAES approach - emissions may be mis-classified as being in the inventory if the emission is within the range of expected, reported emissions. To this point, COBE-2 will include working on full emission range distribution comparisons to determine if there are emission ranges detected via aerial that are currently classified as "in the inventory" that may be partially "out of the inventory". This increase in sophistication in analysis can be carried forward into all other MII work for all measurement types, including continuous monitors and satellites.

The result of these two factors is that the statistical model's methods may be over-estimating OR under-estimating < 5 kg/hr emissions while the MAES model may be underestimating them.

To try to better capture < 5 kg/hr emitters, a future version of the campaign should consider including a representative sampling plan of CMS. Many Colorado operators, including COBE participants, already deploy monitors at select sites. Due to time constraint within COBE, we did not attempt to request data from the majority of these deployed monitors. A limitation of requesting data from participating operators is that it would be limited to participating operators, as non-participating operators would presumably not be willing to share their CMS with the science team. We do not have sufficient evidence as to whether limiting additional data collection for the smallest emitters would be skewed when only using participating operators, given that the DMCs of all vendors was 5 kg and above. However, it is clear that understanding < 5 kg/hr emissions and how they relate to ONGAEIR reporting is a critical next step for developing accurate and defensible MIIs.

Additionally, even though 5 kg/hr was applied as the DMC for Company L facility-level emissions, the majority of Company L's detections were <5 kg/hr, and these measurements were over a much larger number of facilities and wider range of facility types. Future work is needed to assess the two datasets to each other to gain further insights within the measurements already available. This assessment will also include determining if clues exist for differences between participating and non-participating operators at the lowest emission rates.

Viewing these trade-offs for each aerial company along with supplemental CMS-derived emissions rates holistically indicates that the combination of a higher-resolution aerial data source (here, Bridger), a lower-resolution aerial data source (here, GHGSat, Insight M), and statistically representative CMS-derived emission rates could provide the strongest stack of data currently available. Both aerial data sources still have the necessary ability to scan both participating and non-participating operators. The higher-resolution aerial data source could be an aerial or satellite measurement method: the largest emissions detected by GHGSat and Insight M exceeded 100 kg/hr (Table 5), which tends to be within the lower detection limits of current satellite technology. There would be greater risk that the near-100 kg/hr detections might be missed by satellite, however. Given that the majority of emissions were detected by Bridger (Figure 8), the higher-resolution aerial data source is essential in a relatively clean location, such as Colorado, where mitigation opportunities lie more within these relatively smaller emitters. And the CSM results clearly indicate that more investigation is needed to determine the significance of < 5 kg/hr emitters to total emissions within the state of Colorado.

5.2 Operator Participation

It is important to determine whether non-participating operators "look" like participating operators in large studies like COBE in terms of emissions and facility profiles. This question is still an open area of research for the COBE modeling team. We developed CDFs of participating and non-participating operators, using the combined aerial distribution technique described in A.8. Nondetections are included in these distributions at an emission rate of 0 kg/hr. Minor differences are noted between the distributions but this is an area of additional research that the modeling team will continue to pursue.

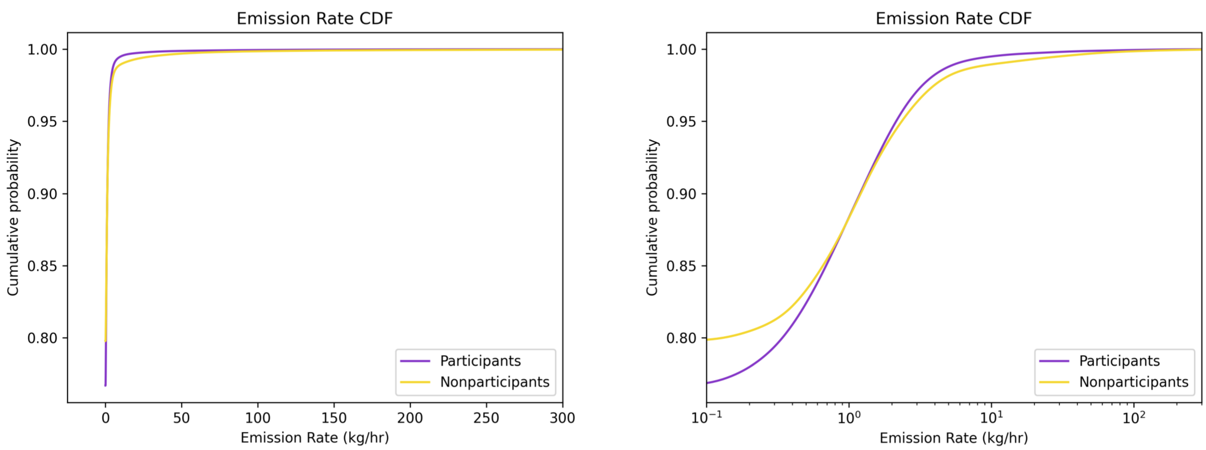


Figure 18: Cumulative distribution functions (CDFs) of participating and non-participating operators. The right plot shows emission rates on a log scale for visual clarity.

5.3 Model Limitations

5.3.1 MAES

A key limitation of the MAES MII approach is the reliance on inventory data to determine facilities' normal operating conditions, as noted above. The initial MAES inventory model assumes emissions reported in the inventory provide a reasonably accurate estimate of the emissions from normal conditions, to which unreported emissions will be added. If the inventory underestimates normal emissions, as the prevalence of small emission rates in the CSM model suggests it might, then the final MAES MII model likely will as well. The classification of emissions due to failure types (Section 3.2.2) is also influenced by the inventory, since the MAES-simulated emissions against which observed emissions are compared is largely based on the inventory. Separately, the detailed facility information required by MAES, including equipment types and counts, also depends on the accuracy of the inventory. Some of this dependence may be relieved by improving aircraft measurement technologies – for instance, using a company like Bridger that can provide estimated facility equipment counts – but the inventory data remains an essential part of this method.

In particular, this study relied on the most recent publicly available ONGAEIR inventory dataset from 2022 for planning and initial analysis, and only updated results to the 2024

ONGAEIR dataset when it became available. The sampling plan for aerial measurements was based on the 2022 dataset, which means that facilities constructed after 2022 were not included, and facilities that ceased operation after 2022 were still scanned due to the absence of updated facility-level information. The classification of emissions described in Section 3.2.2 was carried out using MAES inventory models based on 2022 ONGAEIR data; as this is a manual process, it was not repeated after the 2024 ONGAEIR data became available due to time constraints. Key variables affecting model estimates that are likely to have shifted between 2022 and 2024 include oil and gas production volumes, facility equipment counts, and equipment operating hours. While the METEC team recognizes that ONGAEIR represents the most comprehensive emissions inventory currently available for Colorado, it remains subject to reporting gaps and temporal limitations.

A further limitation of the MAES model is its reliance on manual classification of emissions into predefined failure types. This process is not automated and requires human interpretation of aerial imagery for each detection to assign a probability to the emission source location. While this introduces a degree of subjectivity, it ensured a consistent and standardized approach was applied across all measurement solutions. This classification process then influences the pLeak value and may affect the extent to which each traditionally modeled emission source is impacted. Furthermore, in the current implementation, MAES uses these pLeak inputs as fixed values without accounting for uncertainty, thereby not accounting for some of the variability that would be expected at an actual site. A sensitivity study was conducted on the pLeak value using the MII results based on the 2022 ONGAEIR data. It was found that multiplying pLeak values by .5 and by 2 resulted in a 24% decrease and a 26% increase respectively to the final MAES MII model total emissions (maintenance excluded). Multiplying pLeak values by .1 and 10 resulted in a 36% decrease and a 77% increase respectively.

Additional limitations of the MAES model include the exclusion of pre-production and maintenance-related emissions, as well as an incomplete representation of controlled tanks. While MAES estimates direct tank emissions, it does not account for excess gas routed to the flare during overpressure events. The modeling team is actively working on these limitations, and the development of MAES continues as separate work by METEC.

5.3.2 Statistical model

The statistical model does not currently account for frequency and duration. Instead, emissions are collapsed into a single point in time, and the ergodic assumption is used to translate these emissions into a distribution over time. However, despite the frequent use of the ergodic assumption in methane emissions literature, more investigation is needed to verify how well this assumption is satisfied in this context.

Another key limitation of the statistical model comes from the estimation of a below-threshold distribution. Currently, the estimates of this distribution come either from pre-existing literature that aims to characterize the DJ basin rather than the entire state, or from using CMS-derived inference from only 5 facilities, all owned by the same operator and all in the Piceance Basin. Not having had the opportunity to conduct further testing of any such assumption, it would be ludicrous to assume that rates on these sites are representative of the entire state. Regardless, these methods are currently the best estimates of emissions below Company L's DMC of 5 kg/hr available for this study without relying on a bottom-up inventory or an emission simulation tool. This limitation could be addressed in the future with more by conducting more CMS-derived inference on a larger number of facilities across different operators and basins, as discussed above. The lack of a robust method for estimating Company L's DMC is another limitation of this model, although a sensitivity analysis showed only a minor dependency on this DMC when sampling from either a CMS-informed distribution or those from the literature for below-threshold rates.

5.3.3 Comparison and directions for future work

It is the METEC and CSM science team's opinion that the most important next step is to determine why the two models have such different results. As an initial direction, when comparing the two models, the <5 kg/hr emissions stand out as a key discrepancy. As discussed above, the MAES MII approach relies on current inventory activity data, which may be inaccurate, to estimate emissions in this range. On the other hand, the statistical MBI approach estimates these emissions based either on a limited data source, the small sample of CMS-derived rates, or preexisting distributions with their own limitations. More comprehensive data on this range of emission rates will improve both approaches and help reduce the uncertainty in future MIIs. In particular, producing representative CMS-derived rate estimates as part of future measurement campaigns will greatly improve data on <5 kg/hr emissions. A challenge will lie in understanding how these smallest emissions compare to ONGAEIR reporting, as the smallest emissions will often overlap in size with the reported emissions.

For future work, there are opportunities to continue adjusting the two models to move closer to consensus. MAES has significant value in being able to predict source-level emissions, rather than facility-level, and can also be used for other direct MII reporting needs, such as the Oil and Gas Methane Partnership (OGMP2.0) voluntary reporting program. In future iterations of this work, working to additionally inform MAES in the <5 kg/h category may bring the two models into closer agreement. However, MAES is limited by its inputs and modules: if a process is not correctly modeled or is missing in MAES, it will lead to incorrect or missing emissions estimates. Since MAES relies on ONGAEIR for the inputs, when key facility information is missing from ONGAEIR, MAES cannot model these sites. The team is working with CDPHE to improve future iterations of ONGAEIR so all facility information is available.

The CSM statistical model alternatively considers all available measurement data but assumes that the measurement data is statistically representative and ergodic. It also relied on a highly limited data set of CMS-derived rates for emissions below 5 kg/hr. Future iterations must take into account a better representative sample of CMS-derived rates, given that well over half of predicted total emissions from the statistical model are <5 kg/hr.

METEC and CSM will continue to collaborate in the COBE-2 project to develop peer-reviewed papers that will be published and communicated to the CDPHE APCD team for dissemination. As appropriate, the science team will likely issue an update to this report noting any major findings or updates to results.

6 Summary

The 2024-2025 COBE project was contracted between CDPHE's APCD and CSU's METEC to develop estimates of total emissions and ratios between these estimates and reported emissions (via ONGAEIR) to assist with the 2026 Colorado GHG Intensity Verification Rule. COBE was intended to obtain aerial emission detections for the entire state to develop MIIs. Working with Bridger, GHGSat, and Insight M, the COBE science team (METEC and CSM) obtained over 30,000 individual scans of facilities from aerial overflights. These scans detected approximately 2,000 emissions from upstream facilities that report to the Colorado ONGAEIR, spanning from < 1 kg/hr to upwards of 350 kg/hr (Table 5). METEC and CSM developed independent models, each with strengths and limitations, to determine total emissions and ratios of total emissions to reported emissions. The two models made key different assumptions about incorporating the measurement data, and came up with different sets of state-wide emissions totals and ratios: between 87,210 and 134,352 mt/y and ratios of 3.30 to 5.09 for the statistical model (when filtered down to sites in ONGAEIR modeled by MAES) vs 38,936 mt/y and a ratio of 1.47 for MAES. When including all ONGAEIR facilities, the statistical model estimates emissions between 109,384 and 167,848 mt/y and ratios of 3.81 to 5.85. The ratios developed in this study are specific to the ONGAEIR data and should not be interpreted as methane ratios related to total production in Colorado. This report is an update to the originally submitted report on June 30, 2025. The updates are focused on the model results and include:

- The contribution of various emission rates to the MAES model total, showing the importance of small emissions
- Additional methods for estimating emissions below aerial threshold in the CSM model

The modeling teams will continue to collaborate in the recently funded COBE-2 project to determine specific causes for the discrepancies in model results. Additionally, COBE-2 will develop recommended default factors for 2027 and will continue to work with operator participants. These findings will be communicated regularly to the APCD team through peer-reviewed journal articles.

7 Project Team Contributions

COBE had two primary funded project teams: the METEC and CSM modeling team, and the aerial measurement companies, Bridger, GHGSat, and Insight M.

METEC was the overall project lead (PI: Hodshire) and was responsible for overall direction, project management, and execution of all deliverables for CDPHE's APCD. They also led flight planning with each aerial company and led all participating operator engagement.

METEC and CSM each developed separate models to estimate total emissions and ratios of modeled to reported emissions and collaborated closely on data sharing and additional methodological and results discussions.

The aerial teams each provided measurements and participated in the following roles:

- **Bridger** participated in emission data collection for the Piceance, DJ, and other basins within the COBE project, as well as assisted in site selection and sample planning for its aerial measurement campaigns. Bridger provided a preliminary unpublished quantification error (QE) model and advised the COBE team on best practices for implementing the QE model. Bridger did not participate in total emissions estimation model development or integration of measurement data and models into total emissions estimates.
- **GHGSat** participated in emission data collection for the Piceance, DJ, and other basins within the COBE project, as well as assisted in site selection and sample planning for its aerial measurement campaigns. GHGSat did not participate in total emissions estimation model development or integration of measurement data and models into total emissions estimates.
- **Insight M** participated in emission data collection for the Piceance, DJ, and other basins within the COBE project, as well as assisted in site selection and sample planning for its aerial measurement campaigns. Insight M did not participate in total emissions estimation model development or integration of measurement data and models into total emissions estimates.

8 Funding

Funding for COBE was provided by the Colorado Department of Public Health and Environment Agreement #2024*3364.

9 Competing Interests

The authors of this report have no competing interests to declare.

References

[1] "Oil and Natural Gas Annual Emission Inventory Reporting | Colorado Department of Public Health and Environment." https://cdphe.colorado.gov/ongaeir.

[2] "Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2022," Tech. Rep. EPA 430R-24004, Environmental Protection Agency, 2024.

[3] R. A. Alvarez, S. W. Pacala, J. J. Winebrake, W. L. Chameides, and S. P. Hamburg, "Greater focus needed on methane leakage from natural gas infrastructure," Proceedings of the National Academy of Sciences, vol. 109, pp. 6435–6440, Apr. 2012.

[4] D. Zavala-Araiza, D. R. Lyon, R. A. Alvarez, K. J. Davis, R. Harriss, S. C. Herndon, A. Karion, E. A. Kort, B. K. Lamb, X. Lan, A. J. Marchese, S. W. Pacala, A. L. Robinson, P. B. Shepson, C. Sweeney, R. Talbot, A. Townsend-Small, T. I. Yacovitch, D. J. Zimmerle, and S. P. Hamburg, "Reconciling divergent estimates of oil and gas methane emissions," Proceedings of the National Academy of Sciences, vol. 112, pp. 15597–15602, Dec. 2015.

[5] J. S. Rutherford, E. D. Sherwin, A. P. Ravikumar, G. A. Heath, J. Englander, D. Cooley, D. Lyon, M. Omara, Q. Langfitt, and A. R. Brandt, "Closing the methane gap in US oil and natural gas production emissions inventories," Nature Communications, vol. 12, p. 4715, Aug. 2021.

[6] G. Pétron, A. Karion, C. Sweeney, B. R. Miller, S. A. Montzka, G. J. Frost, M. Trainer, P. Tans, A. Andrews, J. Kofler, D. Helmig, D. Guenther, E. Dlugokencky, P. Lang, T. Newberger, S. Wolter, B. Hall, P. Novelli, A. Brewer, S. Conley, M. Hardesty, R. Banta, A. White, D. Noone, D. Wolfe, and R. Schnell, "A new look at methane and nonmethane hydrocarbon emissions from oil and natural gas operations in the Colorado Denver-Julesburg Basin," Journal of Geophysical Research: Atmospheres, vol. 119, pp. 6836–6852, June 2014.

[7] S. N. Riddick, M. Mbua, A. Anand, E. Kiplimo, A. Santos, A. Upreti, and D. J. Zimmerle, "Estimating Total Methane Emissions from the Denver-Julesburg Basin Using Bottom-Up Approaches," Gases, vol. 4, pp. 236–252, Aug. 2024.

[8] "Colorado Greenhouse Gas Intensity Verification Rule," July 2023.

[9] M. R. Johnson, D. R. Tyner, and A. J. Szekeres, "Blinded evaluation of airborne methane source detection using Bridger Photonics LiDAR," Remote Sensing of Environment, vol. 259, p. 112418, June 2021.

[10] "Insight M Methane Emissions Quantification Methodology," white paper, Insight M, Apr. 2024.

[11] Insight M, "Alternative test method (matm-004) for aerial methane detection using LeakSurveyor™," tech. rep., U.S. Environmental Protection Agency, 2025.

[12] Insight M, "Insight M: Methane emissions monitoring." `https://www.insightm.com`, 2024. Accessed: 2025-06-09.

[13] C. A. McLinden, D. Griffin, Z. Davis, C. Hempel, J. Smith, C. Sioris, R. Nassar, O. Moeini, E. Legault-Ouellet, and A. Malo, "An Independent Evaluation of GHGSat Methane Emissions: Performance Assessment," Journal of Geophysical Research: Atmospheres, vol. 129, p. e2023JD039906, Aug. 2024.

[14] GHGSat, "GHGSat: Greenhouse gas emissions monitoring." `https://www.ghgsat.com/en/`, 2025. Accessed: 2025-06-09.

[15] J. P. Williams, M. Omara, A. Himmelberger, D. Zavala-Araiza, K. MacKay, J. Benmergui, M. Sargent, S. C. Wofsy, S. P. Hamburg, and R. Gautam, "Small emission sources in aggregate disproportionately account for a large majority of total methane emissions from the US oil and gas sector," Atmospheric Chemistry and Physics, vol. 25, pp. 1513–1532, Feb. 2025.

[16] W. S. Daniels, J. L. Wang, A. P. Ravikumar, M. Harrison, S. A. Roman-White, F. C. George, and D. M. Hammerling, "Toward Multiscale Measurement-Informed Methane Inventories: Reconciling Bottom-Up Site-Level Inventories with Top-Down Measurements Using Continuous Monitoring Systems," Environmental Science & Technology, vol. 57, pp. 11823–11833, Aug. 2023.

[17] E. D. Sherwin, J. S. Rutherford, Z. Zhang, Y. Chen, E. B. Wetherley, P. V. Yakovlev, E. S. F. Berman, B. B. Jones, D. H. Cusworth, A. K. Thorpe, A. K. Ayasse, R. M. Duren, and A. R. Brandt, "US oil and gas system emissions from nearly one million aerial site measurements," Nature, vol. 627, pp. 328–334, Mar. 2024.

[18] W. Mollel, D. Zimmerle, A. Santos, and A. Hodshire, "Using Prototypical Oil and Gas Sites to Model Methane Emissions in Colorado's Denver-Julesburg Basin Using a Mechanistic Emission Estimation Tool," ACS ES&T Air, vol. 2, pp. 723–735, May 2025.

[19] A. Santos, W. Mollel, J. Duggan, A. Hodshire, P. Vora, and D. Zimmerle, "Using Measurement-Informed Inventory to Assess Emissions in the Denver-Julesburg Basin," Dec. 2024.

[20] D. Zimmerle, S. Dileep, and C. Quinn, "Unaddressed Uncertainties When Scaling Regional Aircraft Emission Surveys to Basin Emission Estimates," Environmental Science & Technology, vol. 58, pp. 6575–6585, Apr. 2024.

[21] D. J. Varon, D. J. Jacob, J. McKeever, D. Jervis, B. O. A. Durak, Y. Xia, and Y. Huang, "Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes," Atmospheric Measurement Techniques, vol. 11, pp. 5673–5686, Oct. 2018.

[22] Bridger Photonics, "Performance of Gas Mapping LiDAR™ for quantification of very high methane emission rates," technical report, Bridger Photonics, 2021. Single-blind controlled-release study; available as PDF from Bridger Photonics website.

[23] M. J. Thorpe, A. Kreitinger, D. T. Altamura, C. D. Dudiak, B. M. Conrad, D. R. Tyner, M. R. Johnson, J. K. Brasseur, P. A. Roos, W. M. Kunkel, A. Carre-Burritt, J. Abate,

T. Price, D. Yaralian, B. Kennedy, E. Newton, E. Rodriguez, O. I. Elfar, and D. J. Zimmerle, “Deployment-invariant probability of detection characterization for aerial LiDAR methane detection,” Remote Sensing of Environment, vol. 315, p. 114435, 2024.

[24] U.S. Environmental Protection Agency, “Approval of GHGSat’s alternative test method for airborne methane detection technology (MATM-007),” EPA Approval Letter / Alternative Test Method MATM-007, U.S. EPA Office of Air Quality Planning and Standards, Washington, DC, Feb. 2025.

[25] S. H. El Abbadi, Z. Chen, P. M. Burdeau, J. S. Rutherford, Y. Chen, Z. Zhang, E. D. Sherwin, and A. R. Brandt, “Technological Maturity of Aircraft-Based Methane Sensing for Greenhouse Gas Mitigation,” Environmental Science & Technology, vol. 58, pp. 9591–9600, June 2024.

[26] S. El Abbadi, Z. Chen, P. Burdeau, J. Rutherford, Y. Chen, Z. Zhang, E. Sherwin, and A. Brandt, “Comprehensive evaluation of aircraft-based methane sensing for greenhouse gas mitigation,” preprint, Engineering, June 2023.

[27] E. D. Sherwin, Y. Chen, A. P. Ravikumar, and A. R. Brandt, “Single-blind test of airplane-based hyperspectral methane detection via controlled releases,” Elementa: Science of the Anthropocene, vol. 9, p. 00063, Mar. 2021.

[28] D. J. Varon, J. McKeever, D. Jervis, J. D. Maasakkers, S. Pandey, S. Houweling, I. Aben, T. Scarpelli, and D. J. Jacob, “Satellite Discovery of Anomalously Large Methane Point Sources From Oil/Gas Production,” Geophysical Research Letters, vol. 46, pp. 13507–13516, Nov. 2019.

[29] C. Bell, J. Rutherford, A. Brandt, E. Sherwin, T. Vaughn, and D. Zimmerle, “Single-blind determination of methane detection limits and quantification accuracy using aircraft-based LiDAR,” Elementa: Science of the Anthropocene, vol. 10, p. 00080, Nov. 2022.

[30] D. Zimmerle, G. Duggan, T. Vaughn, C. Bell, C. Lute, K. Bennett, Y. Kimura, F. J. Cardoso-Saldaña, and D. T. Allen, “Modeling air emissions from complex facilities at detailed temporal and spatial resolution: The Methane Emission Estimation Tool (MEET),” Science of The Total Environment, vol. 824, p. 153653, June 2022.

[31] P. Vora, A. Hodshire, G. P. Duggan, and D. Zimmerle, “Use of mechanistic modeling to improve failure mode representation in oil and gas emission studies.” In progress, 2025.

[32] D. Zimmerle, T. Vaughn, B. Luck, T. Lauderdale, K. Keen, M. Harrison, A. Marchese, L. Williams, and D. Allen, “Methane Emissions from Gathering Compressor Stations in the U.S.,” Environmental Science & Technology, vol. 54, pp. 7552–7561, June 2020.

[33] D. T. Allen, A. P. Pacsi, D. W. Sullivan, D. Zavala-Araiza, M. Harrison, K. Keen, M. P. Fraser, A. Daniel Hill, R. F. Sawyer, and J. H. Seinfeld, “Methane emissions from process equipment at natural gas production sites in the united states: Pneumatic controllers,” Environmental science & technology, vol. 49, no. 1, pp. 633–640, 2015.

[34] T. L. Vaughn, B. Luck, L. Williams, A. J. Marchese, and D. Zimmerle, “Methane Exhaust Measurements at Gathering Compressor Stations in the United States,” Environmental Science & Technology, vol. 55, pp. 1190–1196, Jan. 2021.

[35] J. A. Brown and A. Hodshire, “Colorado Ongoing Basin Emissions Study (COBE) anonymized final data set of emissions measurements,” Aug. 2025.

[36] W. M. Kunkel, A. E. Carre-Burritt, G. S. Aivazian, N. C. Snow, J. T. Harris, T. S. Mueller, P. A. Roos, and M. J. Thorpe, “Extension of Methane Emission Rate Distribution for Permian Basin Oil and Gas Production Infrastructure by Aerial LiDAR,” Environmental Science & Technology, p. acs.est.3c00229, Aug. 2023.

[37] K. J. Biener, A. M. Gorchov Negron, E. A. Kort, A. K. Ayasse, Y. Chen, J.-P. MacLean, and J. McKeever, “Temporal variation and persistence of methane emissions from shallow water oil and gas production in the gulf of mexico,” Environmental Science & Technology, vol. 58, p. 4948–4956, Mar 2024.

[38] A. R. Brandt, G. A. Heath, and D. Cooley, “Methane leaks from natural gas systems follow extreme distributions,” Environmental Science & Technology, vol. 50, p. 12512–12520, Oct 2016.

[39] Z. R. Barkley, K. J. Davis, N. L. Miles, and S. J. Richardson, “Examining daily temporal characteristics of oil and gas methane emissions in the delaware basin using continuous tower observations,” Journal of Geophysical Research: Atmospheres, vol. 130, Mar 2025.

[40] Bridger Photonics, “Gas Mapping LiDAR™ for aerial methane detection.” `https://www.bridgerphotonics.com/gas-mapping-lidar-for-aerial-methane-detection`, 2025. Accessed: 2025-06-09.

[41] W. S. Daniels, D. W. Nychka, and D. M. Hammerling, “A Bayesian hierarchical model for methane emission source apportionment,” June 2025. arXiv:2506.03395 [stat].

[42] W. S. Daniels, M. Jia, and D. M. Hammerling, “Estimating Methane Emission Durations Using Continuous Monitoring Systems,” Environmental Science & Technology Letters, vol. 11, pp. 1187–1192, Nov. 2024. Publisher: American Chemical Society.

A Appendix

A.1 Facilities Scanned in Basins by PS Class

Approximately 91.4% of the facilities in the DJ basin that are included in 2022 ONGAEIR were scanned by at least one aerial vendor. The breakdown by PS classification is shown in Figure 19. GHGSat scanned the majority of the facilities in each PS class in the DJ basin. Most of the positive detections were reported by Bridger.

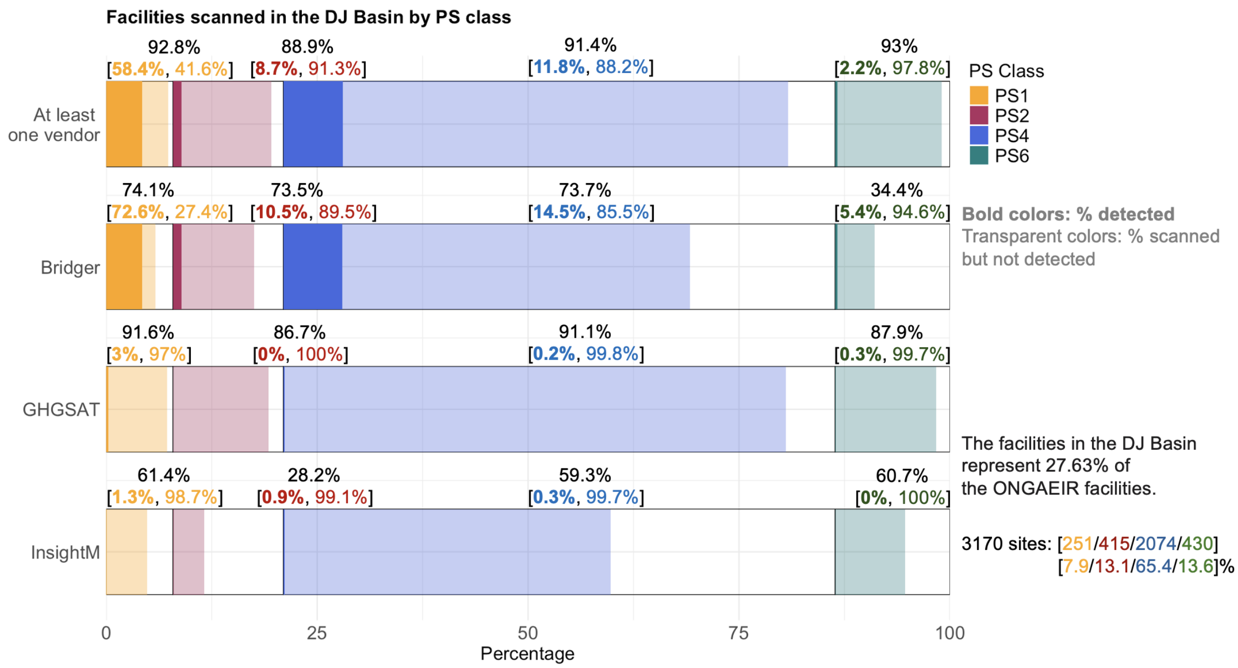


Figure 19: Percentage of facilities in the DJ basin scanned by at least one vendor (top row) and by each vendor (subsequent rows). The percentage in black indicates the overall proportion of facilities scanned within each PS class. The bold percentage in parentheses represents the share of scanned facilities where emissions were detected, while the regular-font percentage shows the share of scanned facilities with no detected emissions. Percent colors correspond to the associated PS classes.

Approximately 96.8% of the facilities in the Piceance basin that are included in 2022 ONGAEIR were scanned by at least one aerial vendor. The breakdown by PS classification is shown in Figure 20. Similarly to the DJ basin, GHGSat scanned the majority of the facilities in each PS class in the Piceance basin as well. Most of the positive detections were reported by Bridger.

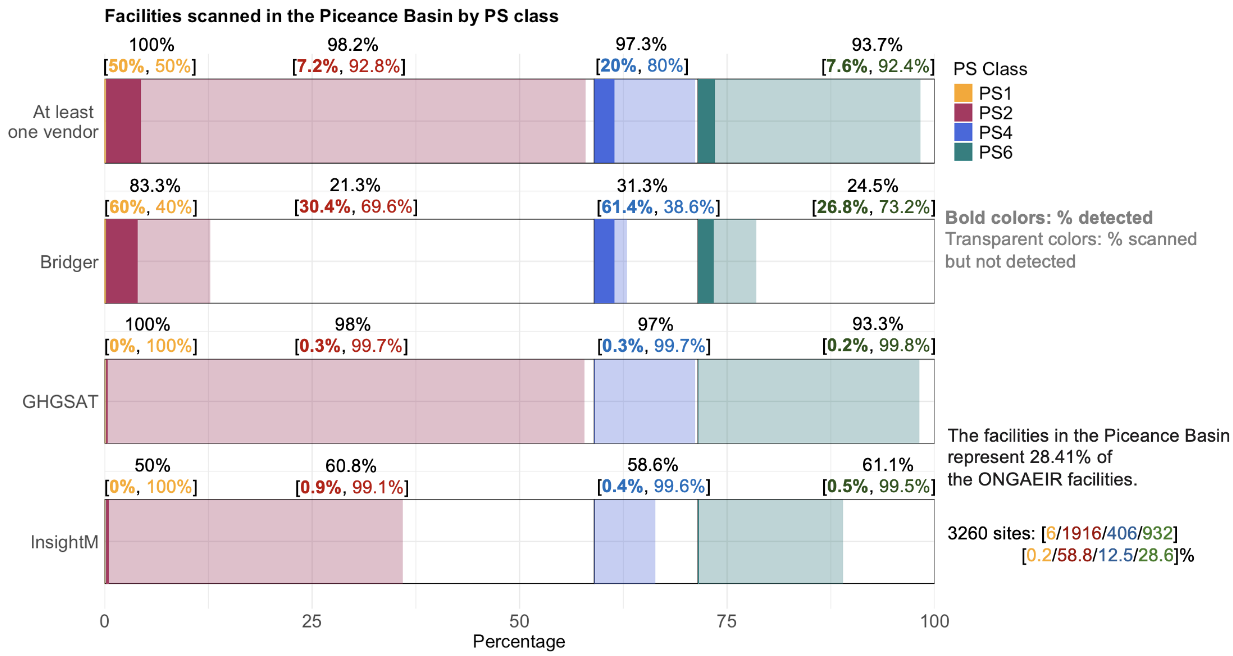


Figure 20: Percentage of facilities in the Piceance basin scanned by at least one vendor (top row) and by each vendor (subsequent rows). The percentage in black indicates the overall proportion of facilities scanned within each PS class. The bold percentage in parentheses represents the share of scanned facilities where emissions were detected, while the regular-font percentage shows the share of scanned facilities with no detected emissions. Percent colors correspond to the associated PS classes.

Approximately 92.4% of the facilities in the other basins that are included in 2022 ONGAEIR were scanned by at least one aerial vendor. The breakdown by PS classification is shown in Figure 21. Insight M scanned the majority of the facilities in PS2, PS4, and PS6 classes in other basins. Bridger scanned the least number of facilities in the other basins, not capturing any of PS1. Most of the positive detections on PS4 and PS6 facilities were reported by Bridger, while more facilities of class PS1 and PS2 had positive emissions according to GHGSat reports.

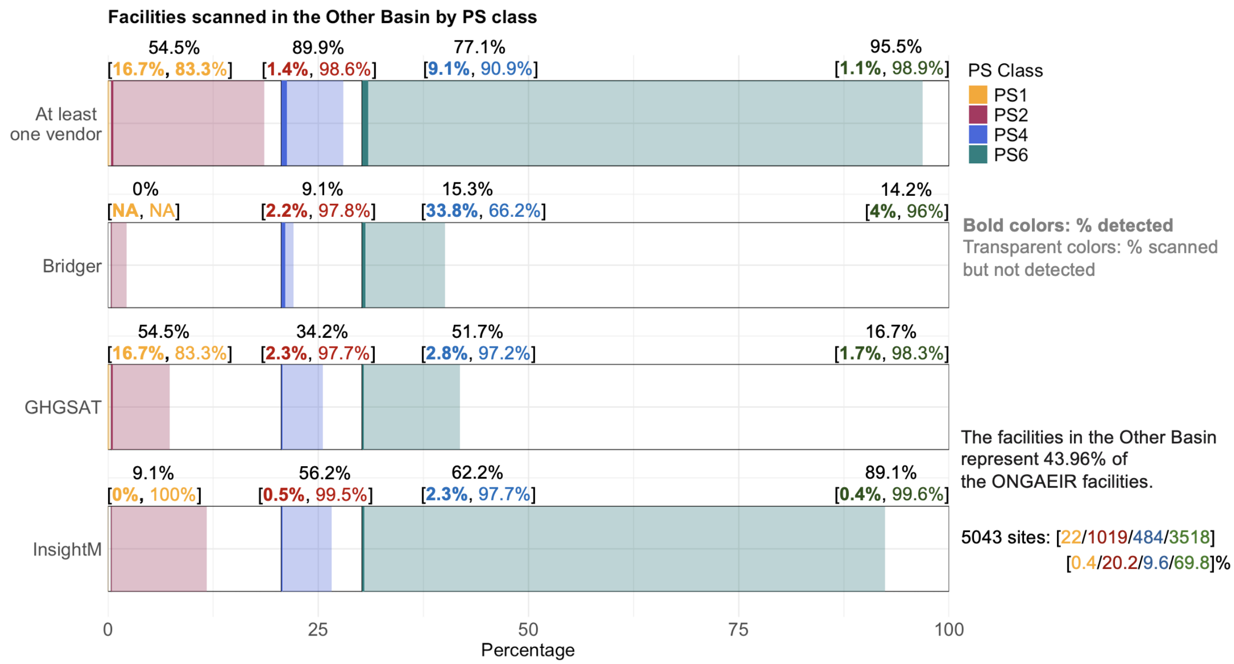


Figure 21: Percentage of facilities in other basins scanned by at least one vendor (top row) and by each vendor (subsequent rows). The percentage in black indicates the overall proportion of facilities scanned within each PS class. The bold percentage in parentheses represents the share of scanned facilities where emissions were detected, while the regular-font percentage shows the share of scanned facilities with no detected emissions. Percent colors correspond to the associated PS classes.

A.2 Details on Aerial Measurement Technologies

A.2.1 Bridger

Bridger Photonics (Bridger) specializes in aerial methane detection and quantification using its proprietary Gas Mapping Light Detection and Ranging (LiDAR) (GML) technology [40]. GML is a high-resolution, aircraft-mounted remote sensing technology that scans facilities to produce a fine-scale (2 m resolution [36]) grid of methane observations. Gas Mapping LiDAR uses laser spectroscopy lidar measurements of the methane absorption line at 1651 nm to determine the methane concentration between the sensor and the objects on the ground illuminated by the GML LiDAR beam. Individual LiDAR point measurements are rasterized to create geo-registered methane concentration imagery, which enables high-sensitivity detection of methane plumes and precise localization and quantification of emission sources as described in Johnson et al. [9]. For analysis in COBE, facility-level emissions were calculated by aggregating the daily average emissions from all sources within the facility.

Bridger is continuously refining and enhancing its quantification model to improve the accuracy of their emission estimates. Bridger developed a GML quantification error model to account for the bias and uncertainty of single-pass emission rate estimates. A preliminary, unpublished version of the model was provided to the COBE team and was used to correct measurement bias and perform uncertainty analysis. The error model relies on a single input parameter, the average signal-to-noise ratio (SNR). Average SNR represents the enhancement of the plume signal above the noise floor during measurement conditions. In general, bias and uncertainty are highest when detections barely exceed the noise floor (smaller average SNR) and decrease as the plume enhancement increases (higher average SNR). For each detection event, the average SNR is computed by averaging the SNR of all pixels within the enhanced region of the detected methane plume.

The error rate is modeled by a log-logistic distribution, with probability density function

$$pdf := f(\alpha, \beta; R) = \frac{(\frac{\beta}{\alpha})(\frac{R}{\alpha})^{\beta-1}}{(1 + (\frac{R}{\alpha})^{\beta})^2}.$$

where R is the relative error ratio, $R = \frac{\text{Actual emission rate}}{\text{Estimated emission rate}}$. The scale, α, and shape, β, parameters vary based on the magnitude of the average SNR and define the distribution of relative error ratios. The distribution average (mean) for R, the bias correction factor, is then used to scale the original estimated emission rate, yielding the bias corrected emission rate. The corrected rates were predominantly lower than the original reported (estimated) rates with an average decrease of -26.7% and range of decrease of -32.78% to -0.002%. Only a few observations were increased, by an average of 0.14%. The full distribution for each detection was used in uncertainty analysis, for instance to provide confidence intervals in the accompanying anonymized dataset and for the estimation of distributions as described in Section A.9. We also mimic the use of log-logistic distributions in error models for the other aerial companies, as described in the following sections.

Sometimes, Bridger detects elevated methane concentrations that signal the presence of an emission, but no corresponding emission rate estimate is generated. This can occur when the methane plume is at the edge of Bridger's survey swath or when methane transport conditions limit the accuracy of plume quantification. Before aggregating emission rate estimates at the facility level, the CSM team imputed emission estimates in cases where elevated methane concentrations were detected but no emission rate was reported (the statistical MII approach uses these imputed values, whereas the MAES MII approach does not). When available, source-level daily mean of positive emission rates were used. If these were unavailable, the project overall source-level mean was applied; otherwise, a default value based on the 90% probability of detection (1.27 kg/h) was used, as reported by Thorpe et al. [23]. The impact of imputation on the distribution is subtle. After imputation, the lower first quartile and median of the data slightly increased at the first decimal level: this indicates that the central mass of the data is slightly shifted upward. However, the higher mean and third quartile indicate that the upper tail contains higher values in the non-imputed data, again at the first decimal place level. Consequently, imputation appears to slightly dampen upper-end variability while elevating mid-range values.

A.2.2 GHGSat

GHGSat High-Resolution Airborne Methane Monitoring, known as DATA AIR [14], uses a high-resolution spectrometer mounted on aircraft to detect and quantify methane emissions, mostly at the facility level. Two generations of products were used during COBE: Gen1, which is capable of detecting emissions above 10 kg/hr and Gen2 with emission detection as low as 5 kg/hr at a 3 m/s wind speed when flying at 10,000 ft above ground level at a nominal speed of 140 knots. One Gen1 sensor, referred to as AV1, was used, and two Gen2 sensors, referred to as AV3 and AV5, were used. The sensors are engineered to detect elevated methane concentrations from local sources by comparing them to the surrounding background levels within the observed scene [24]. All collected imagery is processed with GHGSat proprietary toolchain software. Most detections by GHGSat were only precise enough in location to be treated as facility-level emissions, but in some specific cases, multiple clearly defined plumes were detected at the same time and treated as separate emissions. When facility-level estimates were needed, these emissions reported at the same timestamp for a given facility were first summed, and the facility-level emission rate was then calculated by averaging these totals on a per-day basis.

GHGSat reports uncertainty as a standard deviation for each individual measurement, based on their analysis of multiple sources of error [28]. For consistency with the error models of the other aerial companies, we used a log-logistic distribution with the reported standard deviation to model the error of each measurement.

A.2.3 Insight M

Insight M (formerly Kairos Aerospace) uses a proprietary aerial methane detection system called LeakSurveyor [12], which combines spectral imaging sensors, high-resolution optical imaging, GPS locations, and inertial measurement units mounted on small fixed-wing aircraft. Insight M airplanes fly in a lawnmower pattern to ensure full coverage of the area of interest. The system is designed to detect emissions as low as 10–50 kg/hr, with 90% probability of detection at 10 kg/hr under optimal conditions. Insight M's data processing pipeline converts raw spectral and meteorological data into plume detections and emission rate estimates at the facility level. Two sensor types with 10 kg/hr and 25 kg/hr detection limits were used during COBE. Multiple emission rates reported for the same facility were averaged on a per-day basis.

Insight M cites a 40% standard deviation for uncertainty in all measurements, found in [27]. For this study, we fit log-logisitic distributions to give a more precise error model, and one consistent with the other aircraft companies. A log-logisitic distribution for the relative error ratio with median 1 was fit for Insight M's 25 kg/hr sensor using the data from [27], and another was fit for Insight M's 10 kg/hr sensor using the data from [25]. The distributions are shown in Figure 22.

A.3 Details on Continuous Monitoring Systems (CMS)

CMS data for this study come from five sites in the Piceance basin that are all owned by the same operator. The data was shared confidentially and specific details on facility locations and sensor types remain confidential. The CMS are point-sensor networks, meaning that methane concentrations are measured by a network of in situ point sensors that are arranged around the perimeter of each site. Each of the five sites in this study is equipped with three or four CMS point sensors, all provided by the same CMS vendor. The amount of data varies per site, with one site having 16 months of data, another having 12 months of data, two having 10 months, and the last having 6 months. The fact that emissions below 5 kg/hr

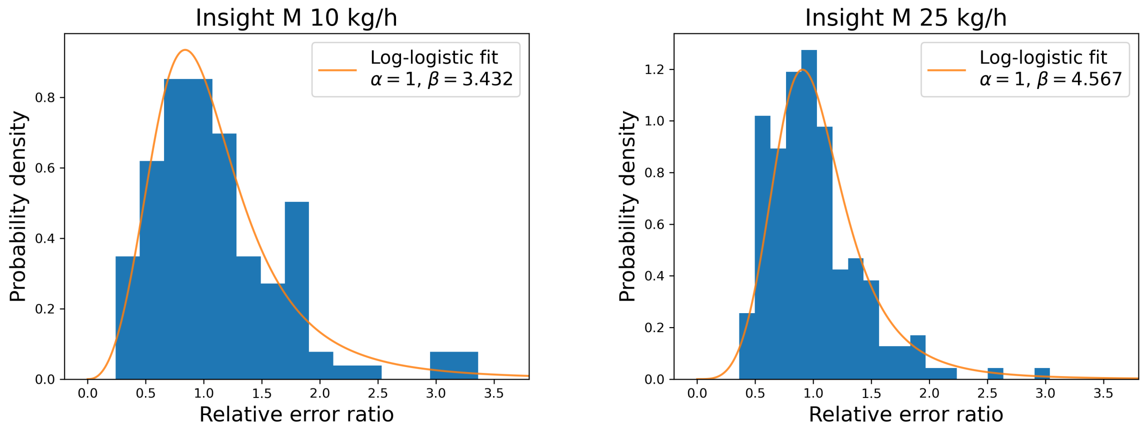


Figure 22: Log-logistic distributions were fit to publicly available controlled release data to model the errors for Insight M's two sensors.

across Colorado are estimated using CMS data from only five sites is a clear limitation of this approach, but it provides a starting point for a completely measurement-based method for estimating emissions below the detection limit of aerial technologies. It has the advantage of not relying on a bottom-up inventory to estimate small emissions, which are known to underestimate emissions [15]. Future work will extend this analysis to many more sites across basins and operators.

An analytical framework is required to translate the raw CMS concentration measurements to estimates of emission source and rate, which are necessary to fill in the distribution of emission rates below Bridger's DMC of 5 kg/hr (discussed in the main body of the report). We use the Bayesian hierarchical model described in [41] to perform this task. At a high level, this model estimates multi-source emissions by combining two separate models within a Bayesian framework: an atmospheric transport model and a time series model for the sensor data. The model uses a spike-and-slab prior for the emission rate parameters, which allows them to be estimated as identically zero, as there are often times when equipment groups are not emitting on oil and gas sites. Furthermore, this model accounts for periods of "no information," or the times when wind blows emitted methane between the CMS point sensors, by using the method described in [42]. In short, this method identifies periods of no information for each source via an atmospheric dispersion model and removes them from subsequent analysis. Finally, to aggregate the source-level emission rate estimates from this model to the site level (to match the aerial analysis described in the main text), we simply sum across the source-level estimates at each time step. Importantly, we only do this for time steps where there is "information" for each source, meaning that there is a downwind sensor for each source. This results in 3,586 site-level emission rate estimates.

Figure 24 shows the distribution of site-level CMS emission rate estimates across the five sites used in this study. The red curve on the left-most plot shows a truncated lognormal fit to the data below 5 kg/hr, that is the data below Company L's DMC: the data that we sample from. The right-most plot shows a quantile-quantile (QQ) plot of the log of the site-level emission rate estimates, which justifies the lognormal fit.

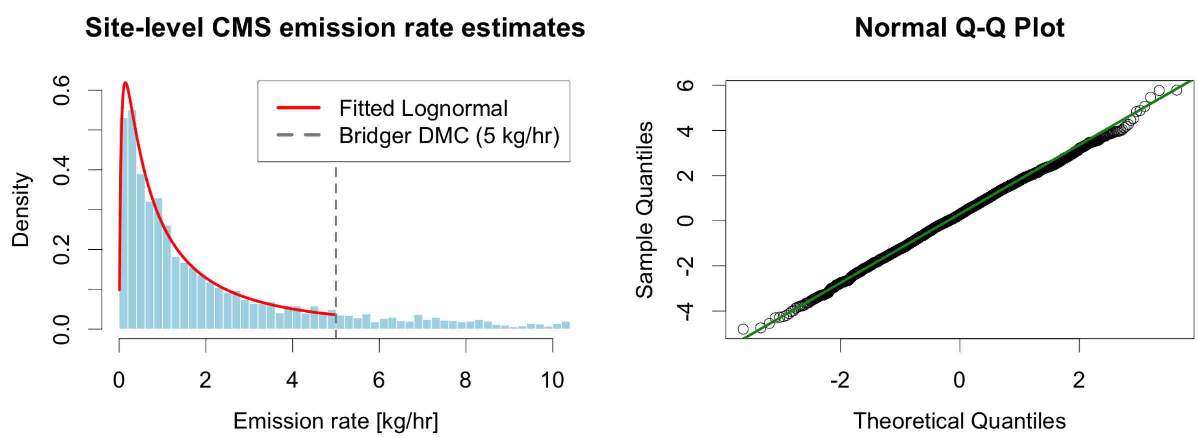


Figure 23: Left: Distribution of site-level CMS emission rate estimates with fitted lognormal shown in red. The vertical dashed line shows where the distribution is truncated when paired with the distribution of aerial rates. Right: QQ plot of the log of the CMS emission rate estimates.

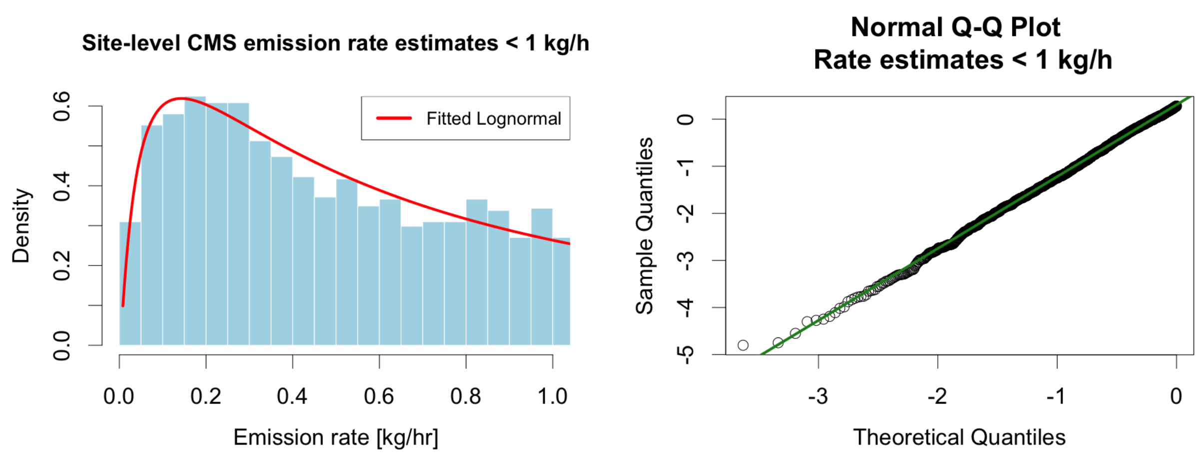


Figure 24: Left: Distribution of site-level CMS emission rate estimates below 1 kg/h with fitted lognormal shown in red. Right: QQ plot of the log of the CMS emission rate estimates below 1 kg/h.

A.4 Comparison of Below-threshold Distributions

Here we show histograms and CDF plots for the three below-threshold distributions used in the statistical model. Only the section of each distribution that is sampled from is shown, that is, only rates below 5 kg/hr. The distribution of CMS-derived rates has much more density at higher emission rates compared to the two distributions from the literature, which explains why estimates using the CMS-derived rates are notably higher.

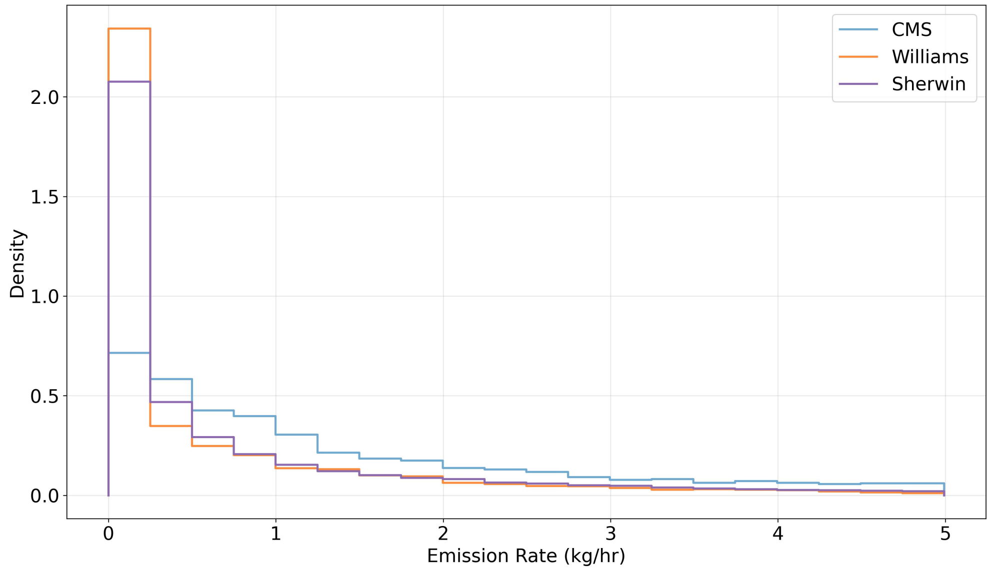


Figure 25: Histograms of the three below-threshold distributions used in the statistical model: CMS-informed, Williams, and Sherwin. All three are truncated at 5 kg/hr, as that is the regime sampled from.

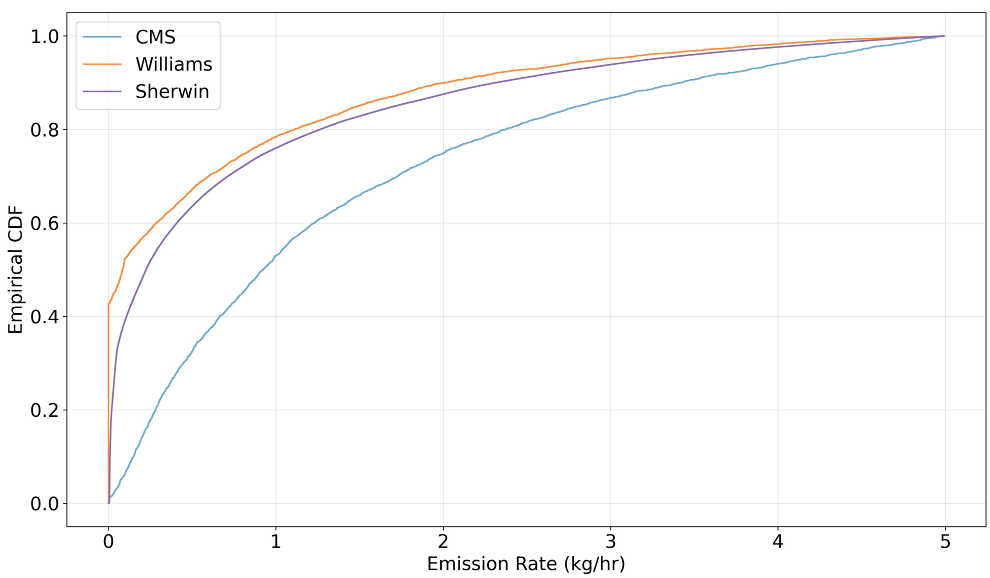


Figure 26: Empirical CDFs of the three below-threshold distributions used in the statistical model: CMS-informed, Williams, and Sherwin. All three are truncated at 5 kg/hr, as that is the regime sampled from.

A.5 Normalized Statistical MBI Results

Here we present the results of the statistical MBI model, normalized by oil and gas production across Colorado and subsets of Colorado. Figure 27 shows these results normalized by natural gas and oil production (in barrel of oil equivalent (BOE)), respectively. A clear trend in this figure is that the normalized emissions in the DJ basin are much lower compared to the Piceance and other basins. This is also the case for PS4 compared to PS2. Note that BOE numbers were calculated as:

$$\text{BOE} = \text{Gas Production [Mcf]}/5.8 + \text{Oil Production [BBL]}$$

Tables corresponding to these results are shown in Tables 14 - 16: one per below-threshold sampling method.

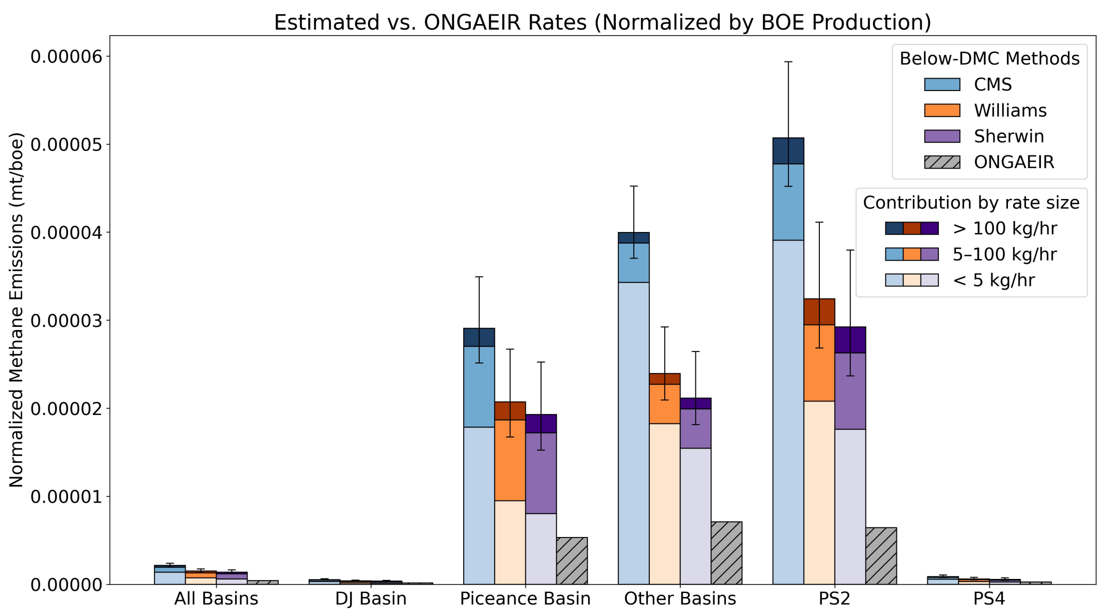


Figure 27: Summary of statistical MBI results, comparable to Figure 14, normalized by oil and gas production. The vertical axis is in units of metric tons of methane emitted per BOE produced.

A.6 Tabulated Version of Statistical MBI Results

	Estimated Rate	95% CI	Rates $<$ 5 kg/hr	Rates 5 - 100 kg/hr	Rates $>$ 100 kg/hr	ONGAEIR Rate	Ratio
All Basins	1.63	(1.50, 1.82)	0.89	0.54	0.19	0.32	5.09
DJ Basin	1.77	(1.59, 2.09)	0.85	0.63	0.29	0.55	3.23
Piceance Basin	1.68	(1.46, 2.02)	0.87	0.67	0.15	0.31	5.46
Other Basins	1.25	(1.16, 1.41)	1.03	0.18	0.05	0.22	5.64
PS2	1.38	(1.23, 1.61)	0.98	0.29	0.10	0.17	7.87
PS4	1.60	(1.42, 1.92)	0.91	0.48	0.20	0.49	3.27

Table 8: Tabulated version of results in Figure 13, specifically using the CMS-informed distribution for below-threshold sampling. All numbers are normalized to a per-facility, per-hour level, and units are kg/hr.

	Estimated Rate	95% CI	Rates < 5 kg/hr	Rates 5 - 100 kg/hr	Rates > 100 kg/hr	ONGAEIR Rate	Ratio
All Basins	1.14	(1.01, 1.33)	0.41	0.54	0.19	0.32	3.57
DJ Basin	1.29	(1.10, 1.61)	0.36	0.63	0.29	0.55	2.35
Piceance Basin	1.20	(0.97, 1.55)	0.39	0.67	0.15	0.31	3.90
Other Basins	0.75	(0.65, 0.91)	0.53	0.18	0.05	0.22	3.38
PS2	0.88	(0.73, 1.12)	0.49	0.29	0.10	0.17	5.04
PS4	1.11	(0.93, 1.43)	0.42	0.48	0.20	0.49	2.27

Table 9: Tabulated version of results in Figure 13, specifically using the Williams distribution for below-threshold sampling. All numbers are normalized to a per-facility, per-hour level, and units are kg/hr.

	Estimated Rate	95% CI	Rates < 5 kg/hr	Rates 5 - 100 kg/hr	Rates > 100 kg/hr	ONGAEIR Rate	Ratio
All Basins	1.06	(0.93, 1.25)	0.32	0.54	0.19	0.32	3.30
DJ Basin	1.21	(1.02, 1.52)	0.28	0.63	0.29	0.55	2.19
Piceance Basin	1.12	(0.88, 1.46)	0.30	0.67	0.15	0.31	3.62
Other Basins	0.66	(0.57, 0.83)	0.44	0.18	0.05	0.22	2.98
PS2	0.79	(0.64, 1.03)	0.40	0.29	0.10	0.17	4.54
PS4	1.02	(0.84, 1.34)	0.33	0.48	0.20	0.49	2.09

Table 10: Tabulated version of results in Figure 13, specifically using the Sherwin distribution for below-threshold sampling. All numbers are normalized to a per-facility, per-hour level, and units are kg/hr.

	Estimated Rate	95% CI	Rates < 5 kg/hr	Rates 5 - 100 kg/hr	Rates > 100 kg/hr	ONGAEIR Rate	Ratio
All Basins	134,351.93	(123,760.01, 149,769.76)	85,782.75	35,935.24	12,633.93	26,410.65	5.09
DJ Basin	32,123.54	(28,747.12, 37,874.75)	18,714.58	9,200.21	4,208.75	9,955.14	3.23
Piceance Basin	42,746.42	(36,956.97, 51,360.31)	26,220.24	13,504.96	3,021.22	7,825.36	5.46
Other Basins	48,661.74	(45,085.55, 55,045.82)	41,731.48	5,453.41	1,476.84	8,630.15	5.64
PS2	38,486.37	(34,289.31, 45,046.71)	29,658.19	6,586.56	2,241.62	4,888.36	7.87
PS4	30,859.13	(27,430.92, 37,119.98)	20,231.11	7,460.85	3,167.17	9,430.86	3.27

Table 11: Tabulated version of Figure 14, specifically using the CMS-informed distribution for below-threshold sampling. All numbers are on a basin or state-wide level, and units are mt/y.

	Estimated Rate	95% CI	Rates < 5 kg/hr	Rates 5 - 100 kg/hr	Rates > 100 kg/hr	ONGAEIR Rate	Ratio
All Basins	94,228.17	(83,590.13, 109,668.13)	45,658.99	35,935.24	12,633.93	26,410.65	3.57
DJ Basin	23,370.31	(19,990.91, 29,114.29)	9,961.35	9,200.21	4,208.75	9,955.14	2.35
Piceance Basin	30,481.29	(24,587.89, 39,248.04)	13,955.10	13,504.96	3,021.22	7,825.36	3.90
Other Basins	29,142.82	(25,495.42, 35,581.00)	22,212.57	5,453.41	1,476.84	8,630.15	3.38
PS2	24,614.47	(20,366.35, 31,214.30)	15,786.29	6,586.56	2,241.62	4,888.36	5.04
PS4	21,395.84	(17,950.42, 27,680.68)	10,767.83	7,460.85	3,167.17	9,430.86	2.27

Table 12: Tabulated version of Figure 14, specifically using the Williams distribution for below-threshold sampling. All numbers are on a basin or state-wide level, and units are mt/y.

	Estimated Rate	95% CI	Rates < 5 kg/hr	Rates 5 - 100 kg/hr	Rates > 100 kg/hr	ONGAEIR Rate	Ratio
All Basins	87,209.80	(76,538.70, 102,670.46)	38,640.62	35,935.24	12,633.93	26,410.65	3.30
DJ Basin	21,839.95	(18,430.55, 27,582.62)	8,430.99	9,200.21	4,208.75	9,955.14	2.19
Piceance Basin	28,335.90	(22,422.15, 37,125.64)	11,809.72	13,504.96	3,021.22	7,825.36	3.62
Other Basins	25,729.38	(22,086.96, 32,189.49)	18,799.13	5,453.41	1,476.84	8,630.15	2.98
PS2	22,189.19	(17,965.77, 28,807.93)	13,361.01	6,586.56	2,241.62	4,888.36	4.54
PS4	19,741.15	(16,267.51, 25,993.67)	9,113.13	7,460.85	3,167.17	9,430.86	2.09

Table 13: Tabulated version of Figure 14, specifically using the Sherwin distribution for below-threshold sampling. All numbers are on a basin or state-wide level, and units are mt/y.

	Estimated Rate	95% CI	Rates < 5 kg/hr	Rates 5 - 100 kg/hr	Rates > 100 kg/hr	ONGAEIR Rate	Ratio
All Basins	2.14×10^{-6}	$(1.98 \times 10^{-6}, 2.39 \times 10^{-6})$	1.37×10^{-6}	5.74×10^{-7}	2.02×10^{-7}	4.22×10^{-7}	5.09
DJ Basin	5.36×10^{-7}	$(4.8 \times 10^{-7}, 6.32 \times 10^{-7})$	3.12×10^{-7}	1.53×10^{-7}	7.02×10^{-8}	1.66×10^{-7}	3.23
Piceance Basin	2.91×10^{-5}	$(2.51 \times 10^{-5}, 3.49 \times 10^{-5})$	1.78×10^{-5}	9.19×10^{-6}	2.06×10^{-6}	5.32×10^{-6}	5.46
Other Basins	4×10^{-5}	$(3.7 \times 10^{-5}, 4.52 \times 10^{-5})$	3.43×10^{-5}	4.48×10^{-6}	1.21×10^{-6}	7.09×10^{-6}	5.64
PS2	5.07×10^{-5}	$(4.52 \times 10^{-5}, 5.94 \times 10^{-5})$	3.91×10^{-5}	8.68×10^{-6}	2.95×10^{-6}	6.44×10^{-6}	7.87
PS4	8.72×10^{-7}	$(7.75 \times 10^{-7}, 1.05 \times 10^{-6})$	5.72×10^{-7}	2.11×10^{-7}	8.95×10^{-8}	2.66×10^{-7}	3.27

Table 14: Tabulated version of Figure 27, specifically using the CMS-informed distribution for below-threshold sampling. All numbers are on a basin or state-wide level, and units are mt/boe production.

	Estimated Rate	95% CI	Rates < 5 kg/hr	Rates 5 - 100 kg/hr	Rates > 100 kg/hr	ONGAEIR Rate	Ratio
All Basins	1.5×10^{-6}	$(1.33 \times 10^{-6}, 1.75 \times 10^{-6})$	7.29×10^{-7}	5.74×10^{-7}	2.02×10^{-7}	4.22×10^{-7}	3.57
DJ Basin	3.9×10^{-7}	$(3.33 \times 10^{-7}, 4.86 \times 10^{-7})$	1.66×10^{-7}	1.53×10^{-7}	7.02×10^{-8}	1.66×10^{-7}	2.35
Piceance Basin	2.07×10^{-5}	$(1.67 \times 10^{-5}, 2.67 \times 10^{-5})$	9.49×10^{-6}	9.19×10^{-6}	2.06×10^{-6}	5.32×10^{-6}	3.90
Other Basins	2.39×10^{-5}	$(2.09 \times 10^{-5}, 2.92 \times 10^{-5})$	1.82×10^{-5}	4.48×10^{-6}	1.21×10^{-6}	7.09×10^{-6}	3.38
PS2	3.24×10^{-5}	$(2.68 \times 10^{-5}, 4.11 \times 10^{-5})$	2.08×10^{-5}	8.68×10^{-6}	2.95×10^{-6}	6.44×10^{-6}	5.04
PS4	6.04×10^{-7}	$(5.07 \times 10^{-7}, 7.82 \times 10^{-7})$	3.04×10^{-7}	2.11×10^{-7}	8.95×10^{-8}	2.66×10^{-7}	2.27

Table 15: Tabulated version of Figure 27, specifically using the Williams distribution for below-threshold sampling. All numbers are on a basin or state-wide level, and units are mt/boe production.

	Estimated Rate	95% CI	Rates < 5 kg/hr	Rates 5 - 100 kg/hr	Rates > 100 kg/hr	ONGAEIR Rate	Ratio
All Basins	1.39×10^{-6}	$(1.22 \times 10^{-6}, 1.64 \times 10^{-6})$	6.17×10^{-7}	5.74×10^{-7}	2.02×10^{-7}	4.22×10^{-7}	3.30
DJ Basin	3.64×10^{-7}	$(3.07 \times 10^{-7}, 4.6 \times 10^{-7})$	1.41×10^{-7}	1.53×10^{-7}	7.02×10^{-8}	1.66×10^{-7}	2.19
Piceance Basin	1.93×10^{-5}	$(1.53 \times 10^{-5}, 2.53 \times 10^{-5})$	8.03×10^{-6}	9.19×10^{-6}	2.06×10^{-6}	5.32×10^{-6}	3.62
Other Basins	2.11×10^{-5}	$(1.81 \times 10^{-5}, 2.64 \times 10^{-5})$	1.54×10^{-5}	4.48×10^{-6}	1.21×10^{-6}	7.09×10^{-6}	2.98
PS2	2.92×10^{-5}	$(2.37 \times 10^{-5}, 3.8 \times 10^{-5})$	1.76×10^{-5}	8.68×10^{-6}	2.95×10^{-6}	6.44×10^{-6}	4.54
PS4	5.58×10^{-7}	$(4.6 \times 10^{-7}, 7.34 \times 10^{-7})$	2.57×10^{-7}	2.11×10^{-7}	8.95×10^{-8}	2.66×10^{-7}	2.09

Table 16: Tabulated version of Figure 27, specifically using the Sherwin distribution for below-threshold sampling. All numbers are on a basin or state-wide level, and units are mt/boe production.

A.7 MAES MII Emission Distributions

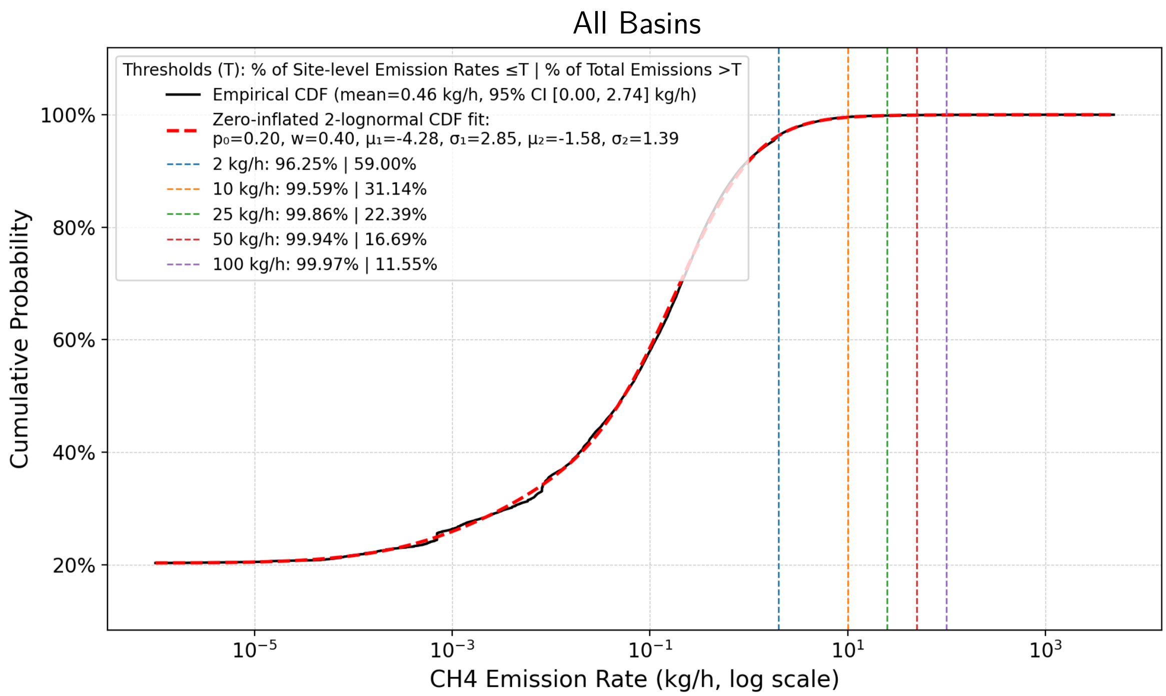


Figure 28: CDF of MAES MII results for all basins

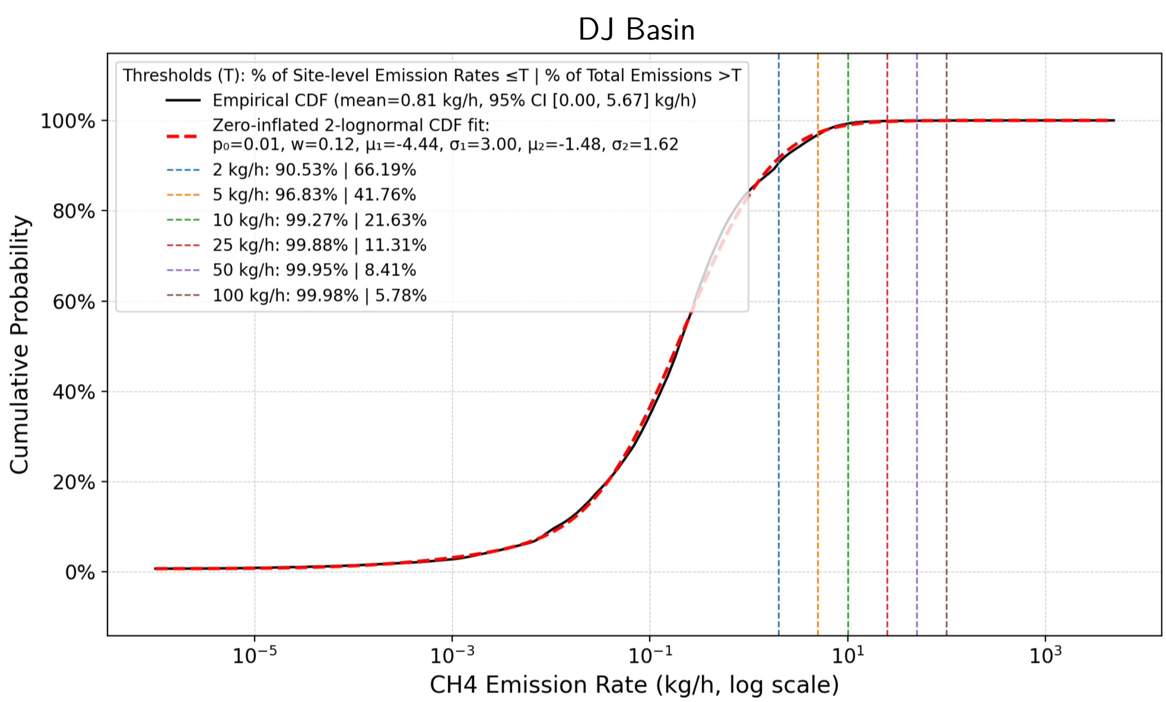


Figure 29: CDF of MAES MII results for the DJ basin

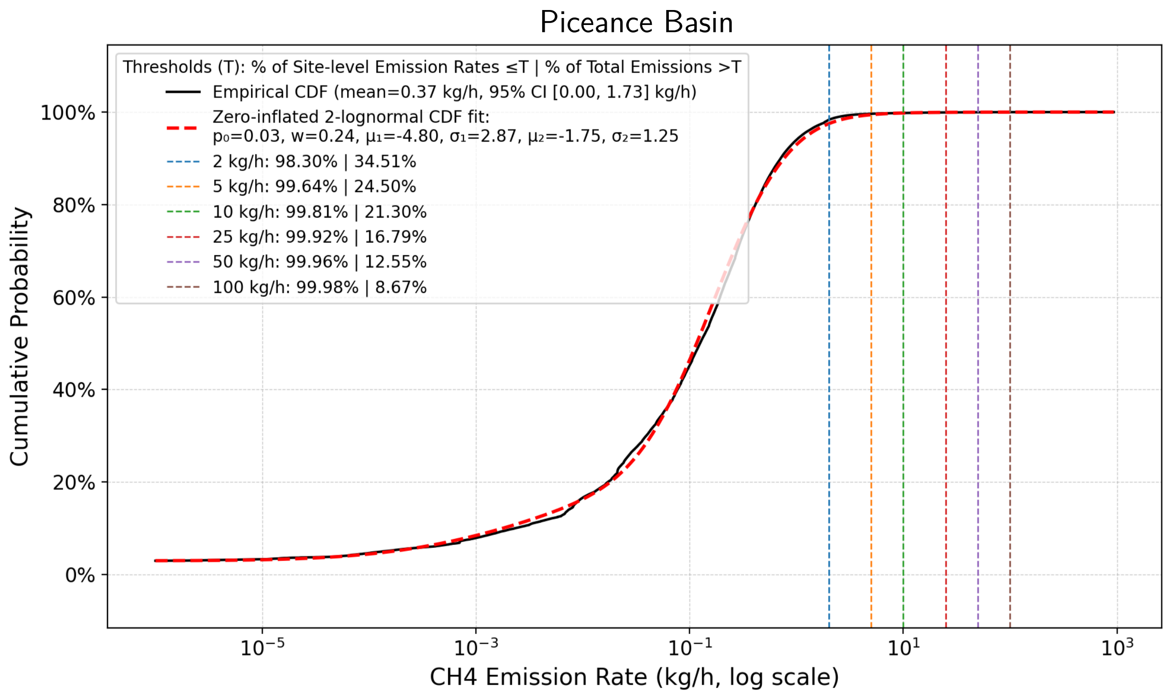


Figure 30: CDF of MAES MII results for the Piceance basin

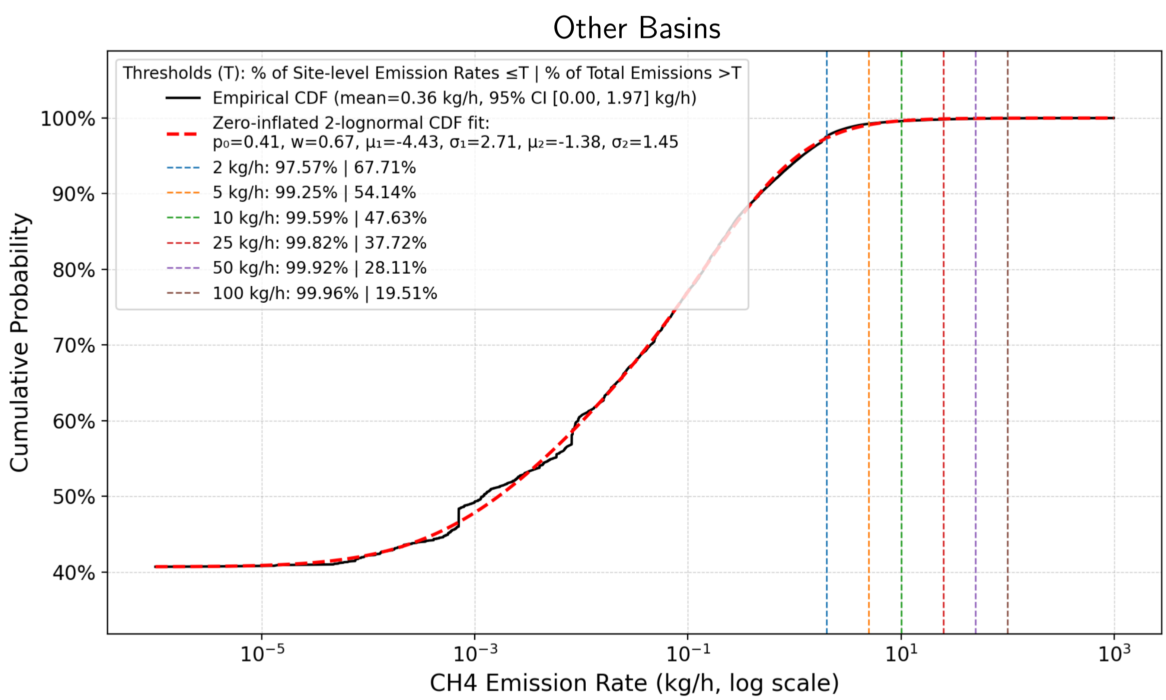


Figure 31: CDF of MAES MII results for other basins

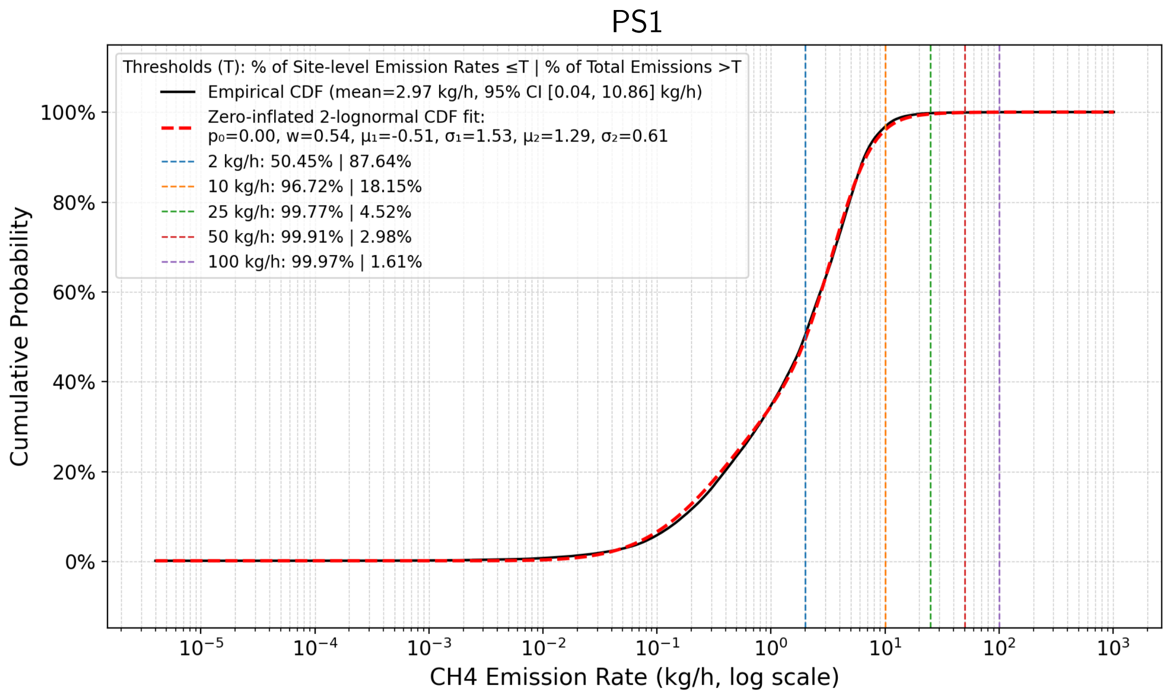


Figure 32: CDF of MAES MII results for PS1 sites

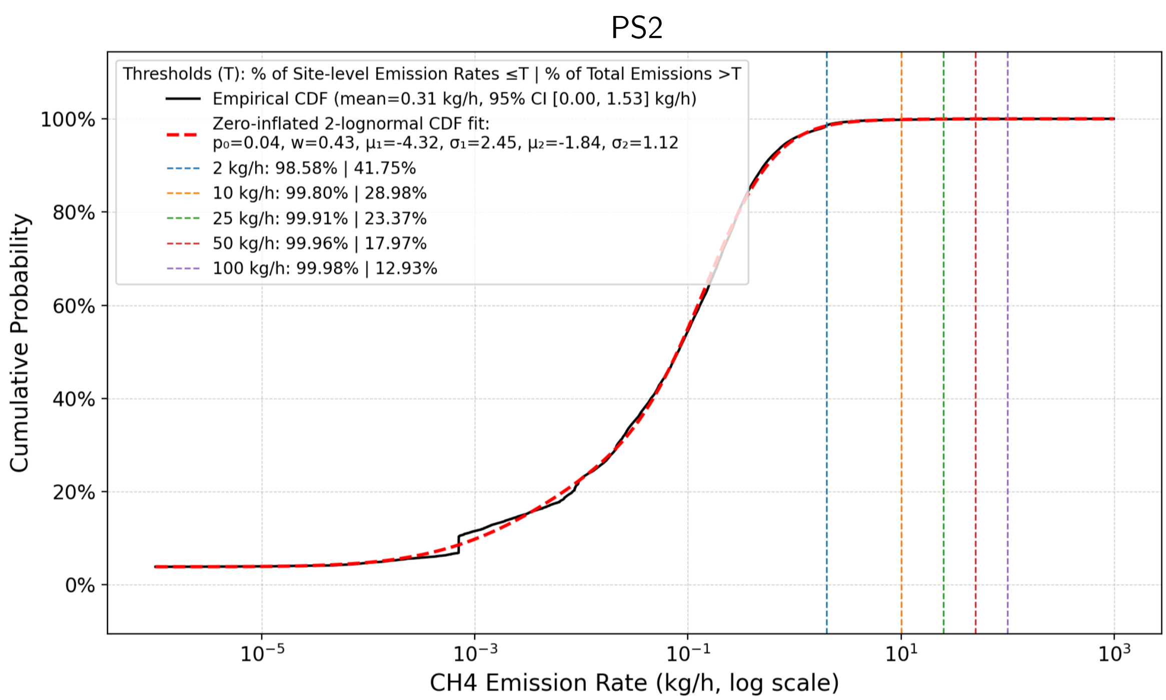


Figure 33: CDF of MAES MII results for PS2 sites

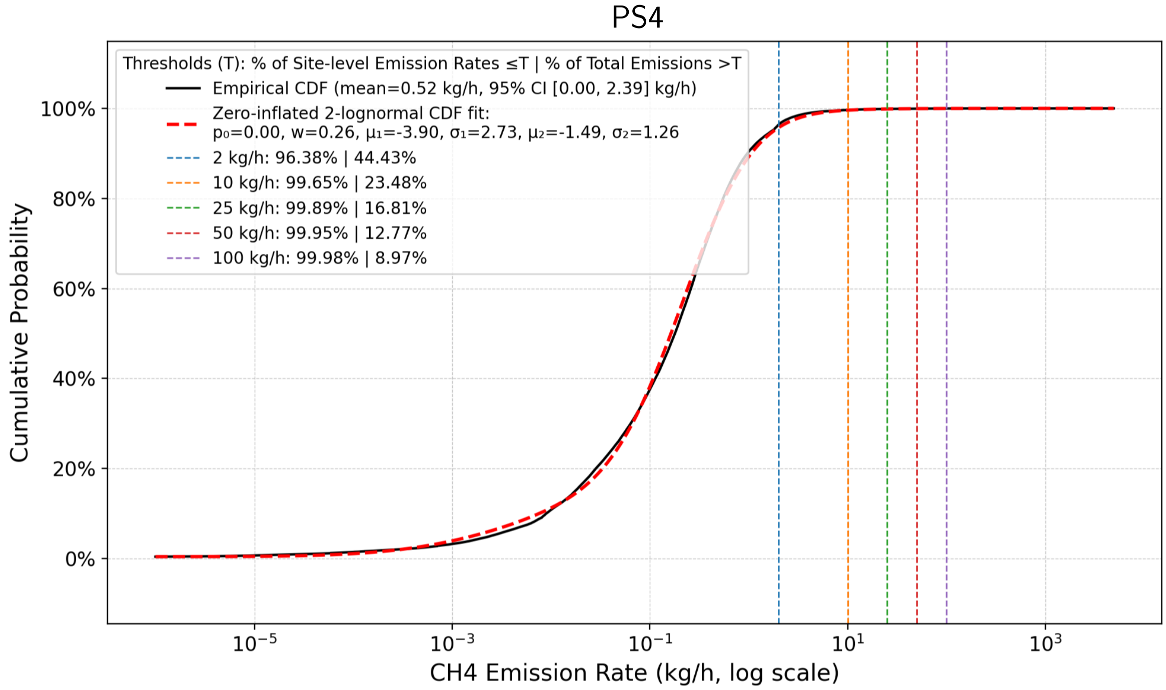


Figure 34: CDF of MAES MII results for PS4 sites

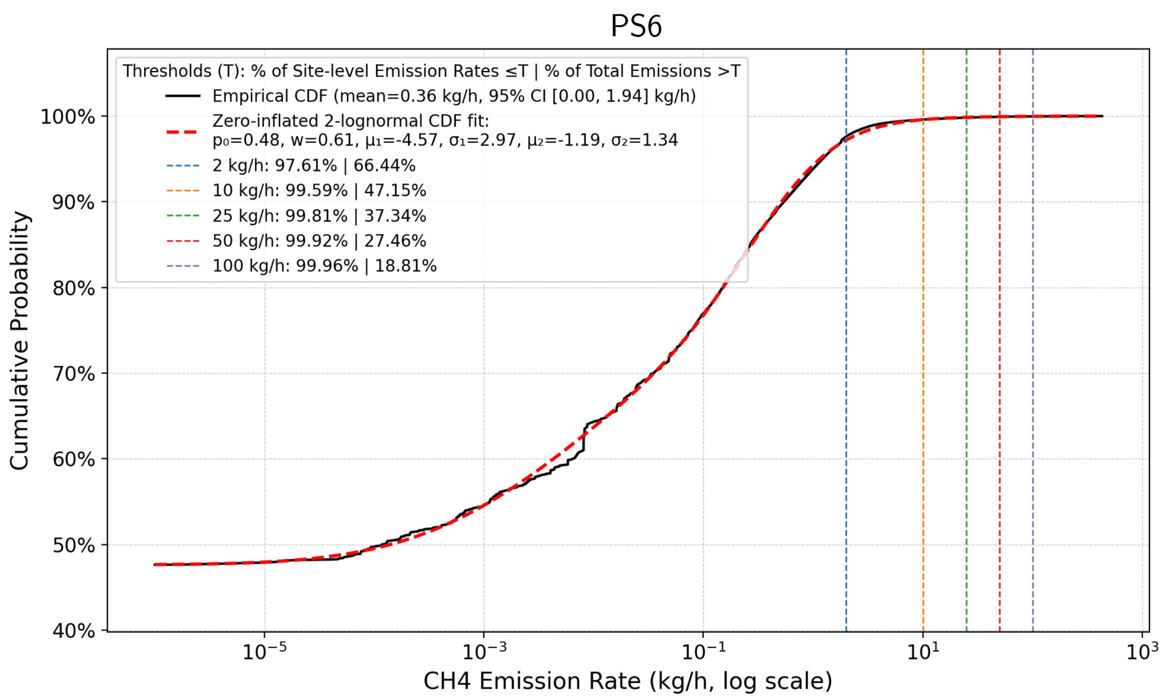


Figure 35: CDF of MAES MII results for PS6 sites

A.8 Estimating Probability of Detection Curves

The difference in technologies and measurement methods between the three aerial companies and their various sensors play an important role in our study, as the separately collected datasets must be combined for analysis. In this section, we estimate probability of detection curves that are used in Section A.9 to combine the distributions of emissions viewed by the various sensors. These probability of detection curves are not meant as a comparison of the aerial vendors and are only intended to assist in the analysis of the present data. As such, while the methods and data sources are described here, the final curves used for the study are not presented.

Any characterization of the state-wide distribution of emissions from the aerial data must be made with the knowledge that a sensor can only see a representative sample of this distribution at sufficiently high emission rates, and as such, sensors with different capabilities will provide different views of the distribution. The differences in sensor capabilities can be seen from the data collected during the measurement campaign: in Figure 36, each sensor exhibits a clear increase in the number of detections through a lower range of emission rates, before reaching a peak and decreasing. The increasing range indicates a gradual increase in the probability that the sensor successfully detects an emission.

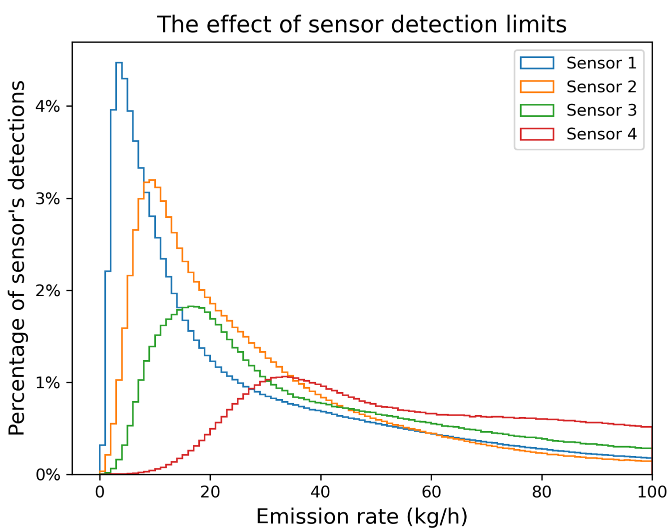


Figure 36: Distributions of detections in the measurement campaign for four of the sensors used; sensors are anonymized in this figure. To account for error, the emission rate for each detection is distributed according to the probability density function from the respective error model. For each sensor, the value on the y-axis shows the approximate percentage of this sensor's detections in a given 1 kg/hr range. The shapes of the distributions, in particular the different locations of the peaks, indicate different sensor detection limits.

To begin, there is publicly available controlled testing data detailed enough to fit a probability of detection curve for Insight M's 10 kg/hr sensor [27]. We fit a logistic curve to this data, estimating probability of detection as a function of emission rate (see Figure 37). For Insight M's 25 kg/hr sensor, we make the simplistic assumption that a given probability of detection is reached at 2.5 times the emission rate needed for the 10 kg/hr sensor. For Bridger, probability of detection curves have been fit in previous papers [23, 29]. Bridger reaches a high probability of detection at lower rates, for instance achieving a 90% probability of detection around 1.27 kg/h [23]. As such, we simply treat the probability of detection for Bridger measurements as 1, with the acknowledgment that very low emission rates are likely underrepresented in the sample.

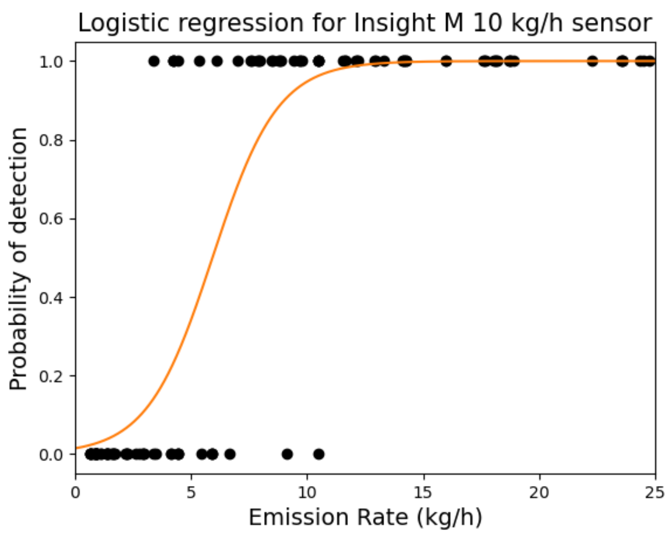


Figure 37: Logistic probability of detection curve for Insight M's 10kg/h sensor fit to controlled release data [27], expressing the probability of detection as a function of emission rate. Probability of detection curves for other sensors were approximated by comparing to this curve.

For the three GHGSat sensors, controlled release data was either not available or was not detailed enough to fit probability of detection curves. We estimate probability of detection curves by comparing their detections during the measurement campaign with those of Insight M's 10 kg/hr sensor. We divide the number of detections of a sensor at a particular emission rate by that for Insight M's 10 kg/hr sensor, then rescale by a linear function so that the resulting curve reaches a peak at 1. This serves as an estimate of the portion of detections seen by Insight M's 10 kg/hr sensor that would be seen by the other sensor, so multiplying by the probability of detection curve for Insight M's 10 kg/hr sensor gives an approximate probability of detection curve. While these are not replacements for probability of detection curves found directly from controlled release data, they provide a rough estimate based on the data available in this study and allow for a more informed analysis than could be done without attention to the different sensors' capabilities.

A.9 Combined Distributions for Failure Types

Here we describe in more detail how data from the three aerial companies are combined into a single distribution for emission rates from a given failure type (for use in MAES models; see Section 3.2.3). While the goal is similar to the distribution combining of Section 3.3, that section considered facility-level emission rates, whereas the techniques for this section are applied to equipment-level emission rates for use in the MAES MII approach.

Partition the range of emission rates to be modeled into "bins," narrow ranges of emission rates with endpoints $0 = e_0 < e_1 < e_2 < \ldots$. If we can estimate the probability that an emission rate x is in bin $[e_{j-1}, e_j)$, dividing this probability by the length of the bin gives an estimate of the probability density in this bin. We therefore describe a procedure to estimate the probability that an emission rate falls in each bin, given samples taken by multiple sensors.

Let M be the number of aircraft measurements taken by a mix of sensors and let i index all measurements. Let $p_i(x)$ be the probability of detection curve for the sensor that took the i^{th} measurement, as a function of emission rate x. For the i^{th} measurement, let x_i be the actual emission rate and y_i the observed emission rate, where nondetection is recorded as $y_i = 0$. We let $b_i = 1$ if the emission was successfully detected and $b_i = 0$ if not. The variable b_i is modeled as a random variable drawn from a Bernoulli distribution with probability equal to $p_i(x_i)$ (note that probability of detection is typically measured in terms of the actual emission rate, not the observed rate). When $b_i = 1$, we let r_i be the ratio of the actual emission rate to observed, so that $y_i = b_i \frac{x_i}{r_i}$. We can thus model r_i as a random variable whose distribution is determined by the aircraft error models discussed above. Fixing a large number N, for each observed y_i (including zeros), we take N samples from the distribution for r_i and multiply by y_i to generate N samples of $y_i r_i = b_i x_i$. We now count each as $1/N$ samples and group into the bins above: let s_j be the resulting number of samples in the bin $[e_{j-1}, e_j)$. Then s_j is approximately the number of successful detections that are expected to have true emission rates in the bin $[e_{j-1}, e_j)$ when M measurements are taken. From this, we wish to estimate the probability that a true emission rate is in this bin, so we divide by the number of measurements out of the total M that we would expect to be successful when applied to emission rates in this bin, the "effective samples" for this bin. This number of effective samples is approximated by $\sum_{i=1}^{M} p_i(m_j)$, where $m_j = \frac{1}{2}(e_{j-1} + e_j)$ is the midpoint of the bin, so the final estimate of true emission rates in the bin $[e_{j-1}, e_j)$ is $s_j / \sum_{i=1}^{M} p_i(m_j)$.

The procedure described above applies generally to any subset of the aerial measurements for which we wish to create a distribution. For samples by Insight M and GHGSat sensors, the probability of detection curves described in Section A.8 were used. While the Insight M probability of detection is based on controlled release testing, future versions of this analysis will hopefully be able to replace the GHGSat curve with updated probability of detections curves from controlled release tests. For samples by Bridger, the probability of detection was set to 1 throughout, so the number of effective samples is always at least the number of Bridger samples. As a result, the estimated probabilities of low emission rates (where Bridger has decreased probability of detection) are expected to be underestimates; this choice was made in acknowledgment of the limitations in measuring these small emission rates.

For use in MAES, we created such a distribution for each of the following failure types: compressors, miscellaneous emitters, controlled tanks, and uncontrolled tanks (the remaining failure types, flares and heaters, are modeled mechanistically in MAES and do not require a distribution). The distributions, along with the distributions estimated from the individual aerial companies, are shown in Figure 38. Each distribution was created from the collection of measurements that were classified as the given failure type following the procedure in Section 3.2.2. Each measurement was counted with the weight of the probability score it was assigned: for instance, if a measurement was assigned a probability score of .7, it was counted as .7 samples in the procedure above. In each case, we used bins of width .01 kg/hr for emissions between 0 and 5 kg/hr and bins of width increasing on a log scale for emissions above 5 kg/hr. We used $N = 250,000$ random samples per detection following the procedure above to create distributions for each failure type. The number M of aircraft measurements (including nondetects) for each failure type was estimated by summing the total number of associated equipment scanned over all aircraft measurements. As described in Section 2.1, equipment counts for the facilities scanned were taken primarily from ONGAEIR: the counts used are shown in Table 3.

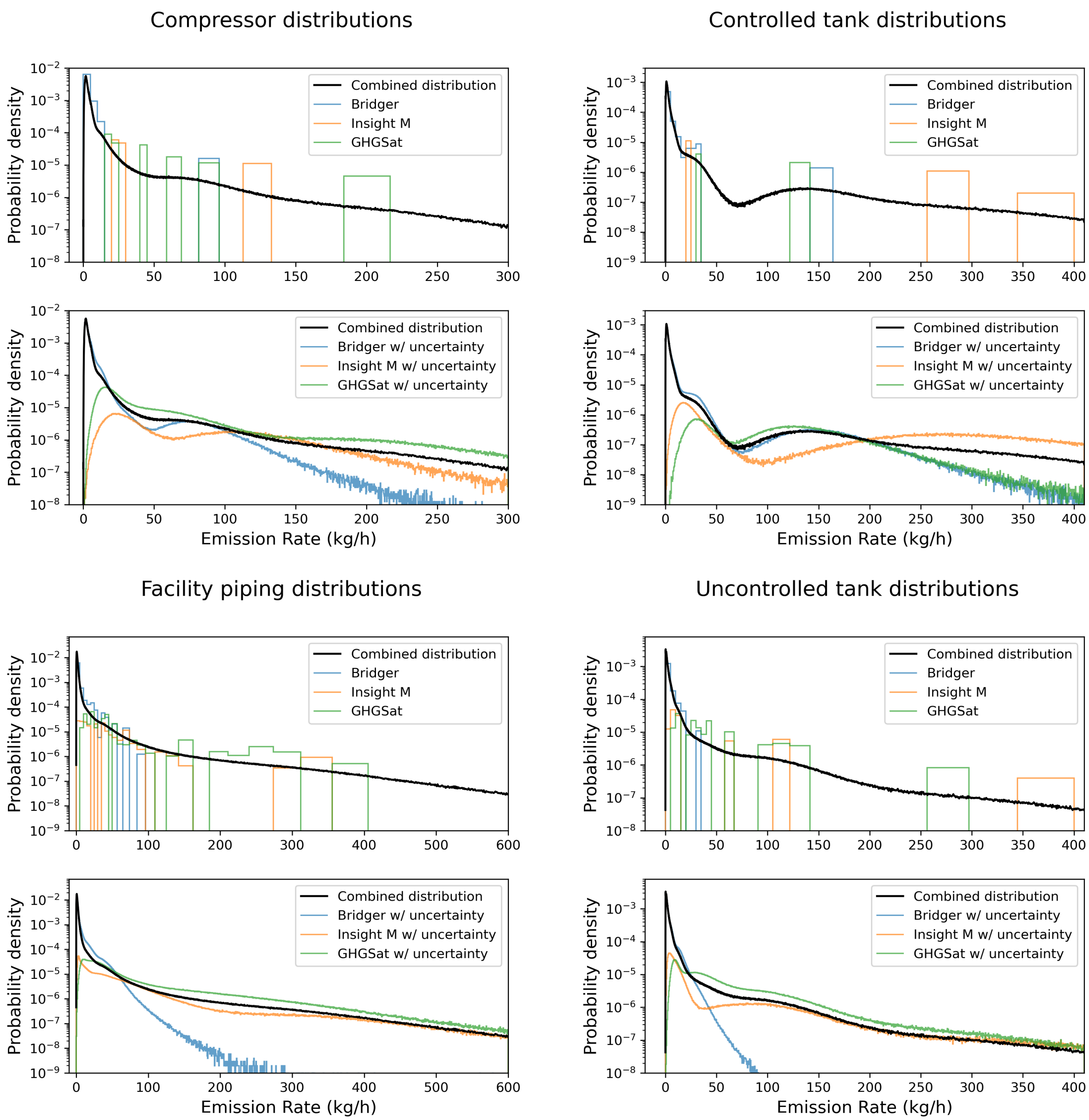


Figure 38: Distributions for emissions observed by aircraft. The total probability in each distribution is the probability of detecting an emission on the given equipment; nondetections are not pictured but account for the remaining probability. The four equipment types shown here are the ones modeled in MAES by emission factor distributions.

A.10 Additional Data Sources

Additional facilities not reported in the 2022 ONGAEIR dataset were scanned by Bridger and GHGSAT in both the DJ and other basins, with the vast majority (99.9%) located outside the DJ Basin. These facilities are present in Colorado Energy and Carbon Management Commission (ECMC) Database. GHGSAT scanned 3,376 ECMC sites while Bridger scanned 343 such sites. Only 2% of the additional data had a record of a positive emission rate. All

Bridger scanned sites are located in the other basins. The average positive detected emission rate reported by Bridger is 3.67 kg/hr with a minimum rate of 0.115 kg/hr and maximum rate of 102 kg/hr. The two facilities scanned by GHGSAT in the DJ basin had no detected emissions. The average positive detected emission rate reported by GHGSAT is 170 kg/hr with a minimum rate of 6 kg/hr and maximum rate of 3,242 kg/hr. This additional data wasn't used in the development of either MII model.

A.10.1 Equipment Count Validation

Information for frequency of different types of modeled equipment is obtained primarily through ONGAEIR reporting, as operators are required to specify the source equipment when reporting emissions. However, additional sources of data on this were explored.

The primary tool used was a machine learning (ML) image classification model used to identify oil/gas equipment. Using satellite imagery, this tool identifies tanks, flares, and separators. This model was deployed on 7,015 facilities throughout the state, of which 7,001 had corresponding ONGAEIR submissions to compare to. Limitations of the satellite imagery approach was age of the images (not necessarily corresponding to 2024) and that it didn't cover all facilities.

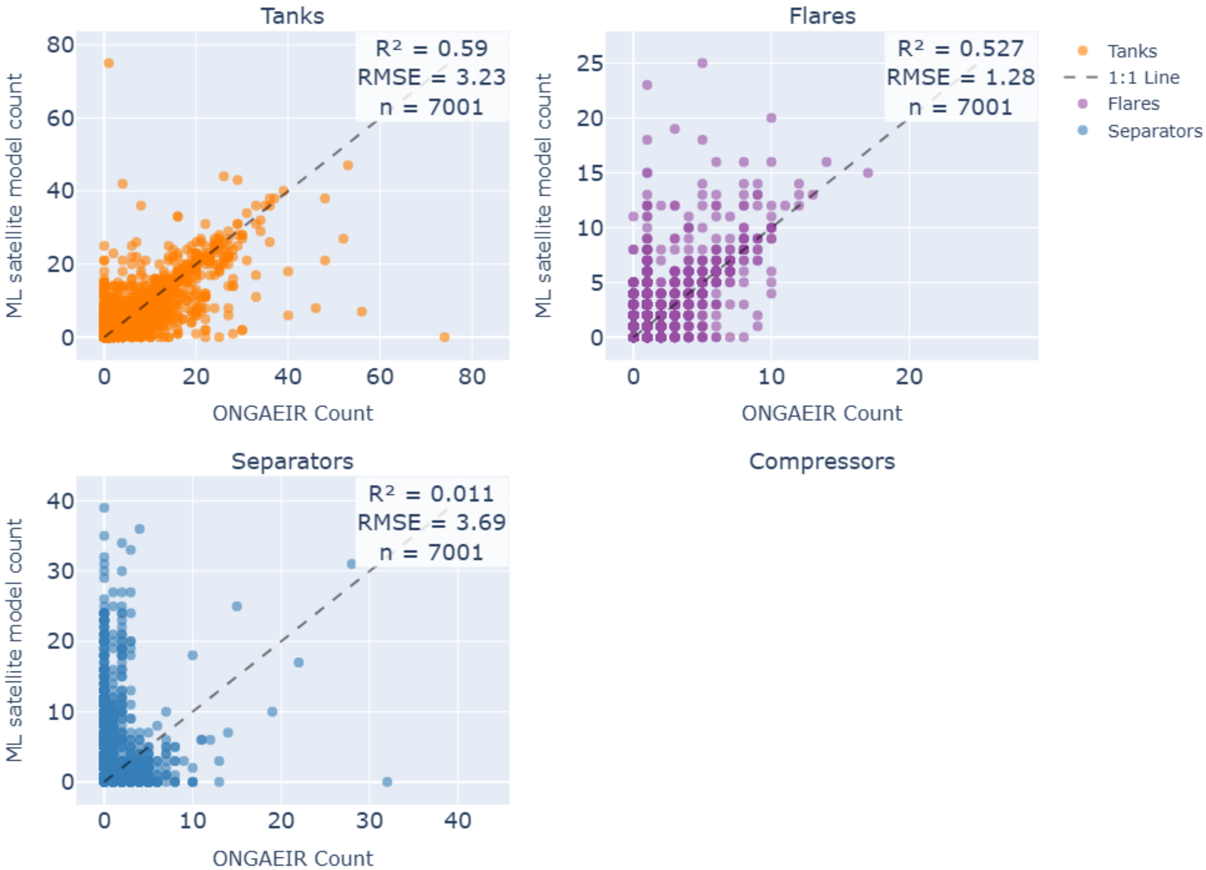


Figure 39: Comparison of equipment counts between ONGAEIR reporting and ML model identification across different facility types.

We additionally used aerial imagery when the imagery was sufficiently high-resolution enough to determine on-site equipment as additional checks.

The results from each of these sources can only be used selectively, where the data from ONGAEIR is clearly inaccurate. An example of this is one site that reported 74 tanks. Upon inspection of satellite imagery, there were many objects (that resembled tanks) clearly not associated with oil/gas production. Both the aerial imagery at this site, as well as the ML model, observed zero tanks at this location, so the count of tanks was updated to zero. This process was conducted to correct tank counts on 16 sites, and flare counts on 47 sites. The count of separators saw far more deviation across the three datasets, prompting the application of assumptions described in Section 2.1.

A.11 Emission Factor Summaries

Table 17: Methane emission factors by equipment group and PS class, reflecting the nonzero samples of equipment annual emissions produced by the MAES MII model. The distributions are consistently skewed right. The mean, 25th percentile, median, and 75th percentile are given in units of mt/year.

PS	Equipment Type	Mean	25%	Median	75%
PS1	Compressor	5.294	0.329	1.024	4.275
PS4	Compressor	1.947	0.246	0.690	1.890
PS6	Compressor	2.393	0.386	0.994	2.429
PS1	Flare	0.553	0.008	0.043	0.236
PS4	Flare	0.228	0.000	0.003	0.040
PS1	Heater	0.202	0.012	0.046	0.108
PS2	Heater	0.021	0.002	0.005	0.010
PS4	Heater	0.053	0.004	0.008	0.025
PS6	Heater	0.021	0.000	0.003	0.007
PS1	Miscellaneous	9.636	0.219	0.476	1.089
PS2	Miscellaneous	1.212	0.199	0.444	1.016
PS4	Miscellaneous	1.223	0.202	0.450	1.024
PS6	Miscellaneous	1.051	0.152	0.332	0.744
PS1	Separator	0.274	0.000	0.004	0.231
PS2	Separator	0.274	0.000	0.001	0.314
PS4	Separator	0.276	0.000	0.001	0.314
PS6	Separator	0.259	0.000	0.001	0.243
PS1	Tank	0.271	0.005	0.032	0.147
PS2	Tank	0.540	0.007	0.052	0.249
PS4	Tank	0.484	0.005	0.038	0.198

A.12 MAES Inputs

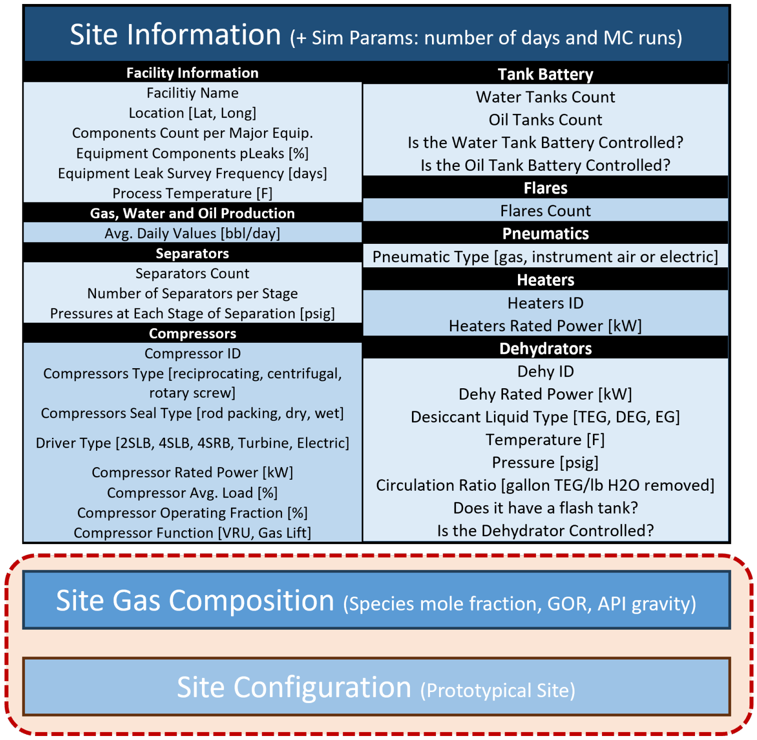


Figure 40: This image shows the equipment and facility information required for the MAES model.

A.13 Anonymized Aerial Dataset

The anonymized dataset is published on Dryad [35], and includes a Comma Separated Variable (CSV) file containing emissions measurements for each aerial vendor and campaign and a README text file with further explanation. This data has been anonymized (by removing any facility-identifying information) to ensure confidentiality for site operators. Some metadata is included with each measurement: the aerial vendor and product used to measure, the campaign (season), and the PS assigned to the facilities. Where available, there is also assigned emission type and a determined cause. This tranche includes all emissions detected, including maintenance emissions. Some measurements were determined to be outside of the modeling scope for one of the following reasons: the site is using apparent pre-production equipment (such as a drilling rig observed from aerial imagery), the site is associated with midstream activities, or the emission recorded does not align with any associated facility. The counts of measurements that fall into each of these categories are described in Table 2, as well as included in the separate dataset.

A.14 ONGAEIR 2024 - Errors

Facilities with reported methane emissions errors. At the time of analysis, these errors were flagged and were left out of this analysis. CDPHE is in contact with these operators to request them to resubmit.

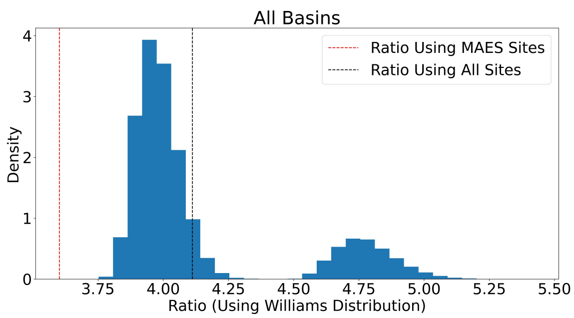
- Island Butte - B... reported 15,000 mt/y reported from tanks.
- Bret Grandbouche 24-02H - reported 231 mt/y in fugitives, would indicate a loss rate of 78%.
- Dawson Creek - reported 231 mt/y in fugitives, would indicate a loss rate of 78%.
- Dill Gulch 1-22 - reported 231 mt/y in fugitives, would indicate a loss rate of 78%.
- Gnat Hill - reported 231 mt/y in fugitives, would indicate a loss rate of 78%.
- Welker 6-92 1-2H11 - reported 231 mt/y in fugitives, would indicate a loss rate of 78%.

A.15 MAES Modeled Criteria

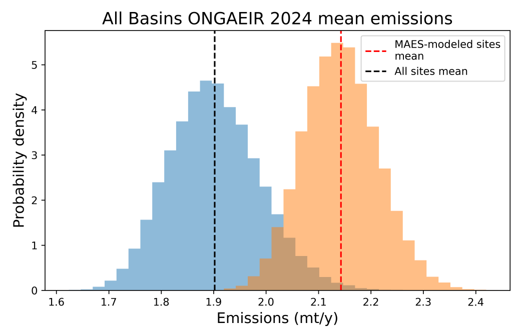
In this project, MAES models facilities with either non-zero hydrocarbon liquid production (oil and water) or gas-only production exceeding 1 (MMscf/year). The liquid production requirement reflects fundamental dependencies in emission quantification algorithms where key sources require liquid production as input parameters. For example, tank emissions depend on gas-liquid phase equilibrium and flashing processes during pressure reduction in separator-tank systems; without liquid throughput data, the volume of liberated gas cannot be estimated. Similar dependencies exist for other liquid-handling equipment where emissions are intrinsically linked to liquid production rates and compositions. Gas-only facilities above the threshold can be modeled using gas throughput alone, as their equipment configurations typically exclude liquid-dependent emission sources. Of the 3,008 facilities in this (gas-only) category, 68% have wellheads only and 32% report additional equipment, though the absence of liquid production at these sites remains unclear. Several facilities were excluded from modeling: 1,463 facilities reported neither gas nor liquid production; one operator reported 700 individual compressor sites with all other equipment aggregated at the basin level (these compressors were consolidated into a single basin-level site for modeling); 95 facilities lacked sufficient compressor data; and 12 duplicate facilities reported by multiple operators were modeled only once.

A.16 Comparison of MAES-modeled and -unmodeled Sites

Here we present the results of subsampling studies on the ONGAEIR 2024 dataset, such as those found in Figure 17, for all subsets of Colorado. The bimodal behavior visible in Figures 41a and 43a is due to the presence of an extreme outlier (one operator reported all their fugitive emissions at a single facility). The clusters of lower ratios consist of samples including this facility, while the clusters of higher ratios consist of samples that do not.

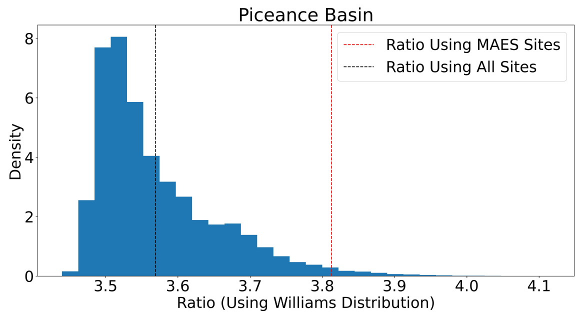


(a) Distribution of statistical MBI ratios for all basins resulting from random samples of facilities of the same size as the number of facilities modeled by MAES. The ratio using all sites is shown with a dashed black line, and the ratio using the MAES-modeled sites is shown with a dashed red line.

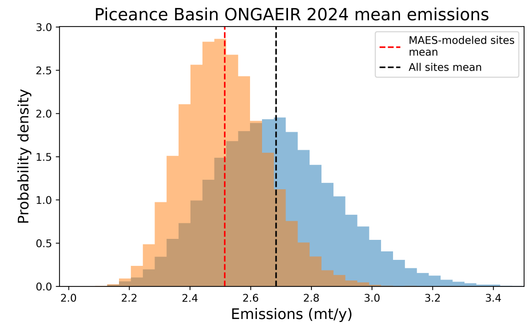


(b) Mean emissions as reported in the ONGAEIR 2024 dataset for all basins, shown for both all sites and the subset that were modeled in MAES. The blue distribution is a bootstrapped distribution for the mean for the whole basin, and similarly the orange distribution is for the subset modeled in MAES.

Figure 41: Subsampling study results for all basins.

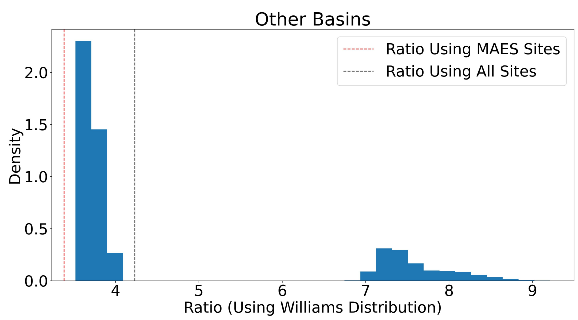


(a) Statistical MBI subsampling study results for the Piceance Basin. Analagous to Figure 41a.

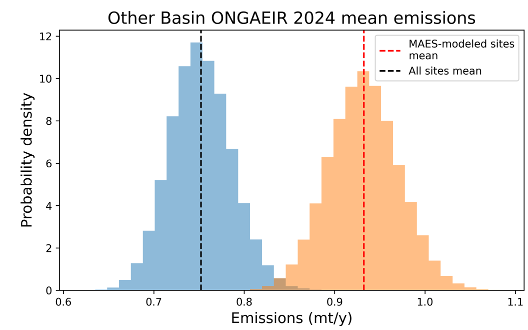


(b) Mean emissions as reported in the ONGAEIR 2024 dataset for the Piceance basin. Analogous to Figure 41b.

Figure 42: Subsampling study results for the Piceance Basin.



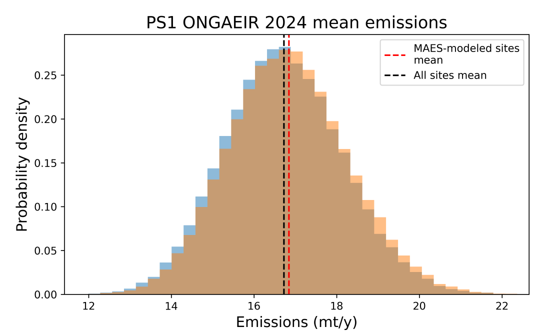
(a) Statistical MBI subsampling study results for other basins. Analagous to Figure 41a.



(b) Mean emissions as reported in the ONGAEIR 2024 dataset for other basins. Analogous to Figure 41b.

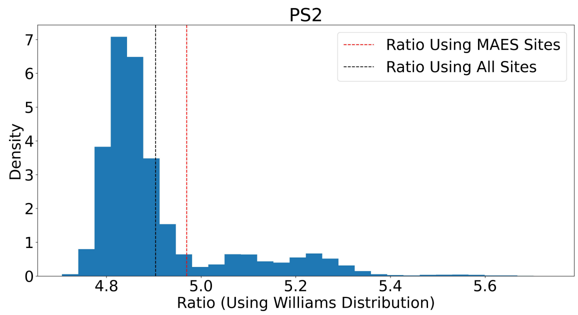
Figure 43: Subsampling study results for other basins.

(a) Statistical MBI was not performed for PS1 sites, so no distribution of ratios is shown.

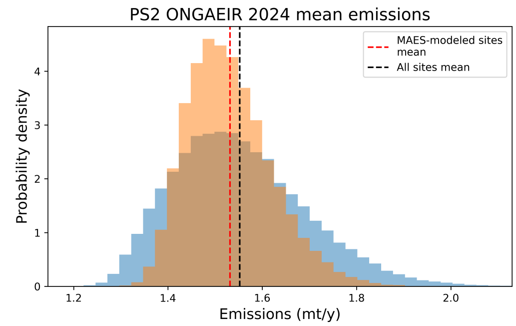


(b) Mean emissions as reported in the ONGAEIR 2024 dataset for PS1 sites. Analogous to Figure 41b.

Figure 44: Subsampling study results for sites in the PS1 class.

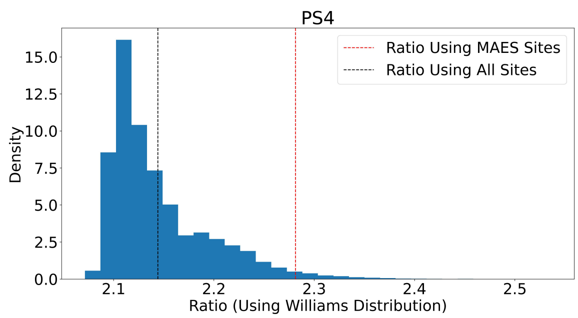


(a) Statistical MBI subsampling study results for sites of class PS2. Analagous to Figure 41a.

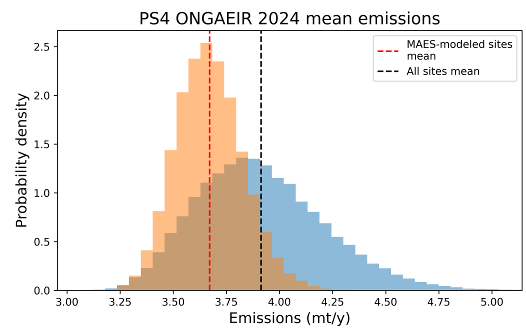


(b) Mean emissions as reported in the ONGAEIR 2024 dataset for PS2 sites. Analogous to Figure 41b.

Figure 45: Subsampling study results for sites in the PS2 class.



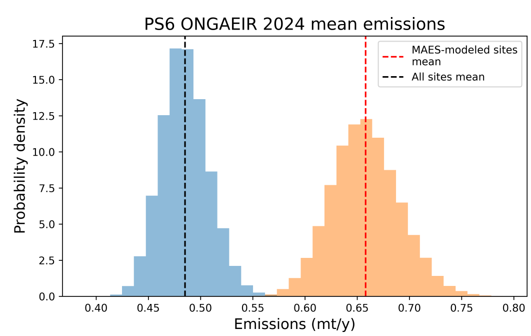
(a) Statistical MBI subsampling study results for sites of class PS4. Analagous to Figure 41a.



(b) Mean emissions as reported in the ONGAEIR 2024 dataset for PS4 sites. Analogous to Figure 41b.

Figure 46: Subsampling study results for sites in the PS4 class.

(a) Statistical MBI was not performed for PS6 sites, so no distribution of ratios is shown.



(b) Mean emissions as reported in the ONGAEIR 2024 dataset for PS6 sites. Analogous to Figure 41b.

Figure 47: Subsampling study results for sites in the PS6 class.

A.17 Previous results based on 2022 ONGAEIR

The previous version of this report, completed June 2025, gave results based on the 2022 ONGAEIR dataset. Figures from that report are reproduced in Figures 48 and 49 for reference. Figures 50 and 51 are updated versions, still based on the 2022 ONGAEIR dataset. The main differences between the 2022 and 2024 ONGAEIR datasets are summarized in Section 2. Results are summarized below, all comparing to the adjusted ONGAEIR totals described in the final report.

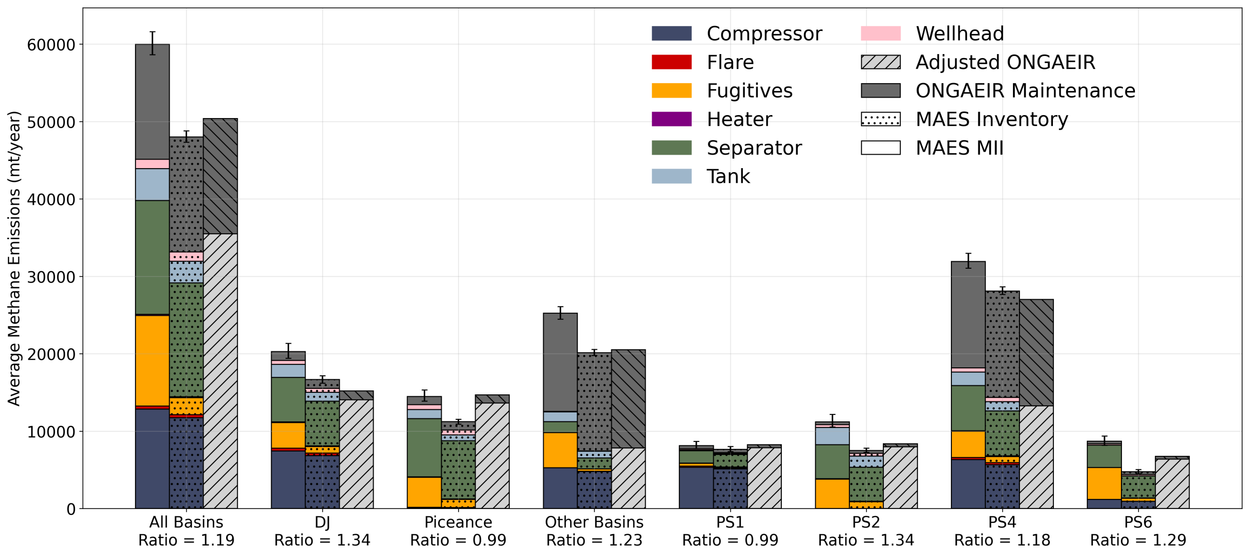


Figure 48: MAES MII results based on the 2022 ONGAEIR dataset, reproduced from the June 2025 COBE Final Report

Using the 2022 ONGAEIR data, the MAES inventory model total was 33,140 mt/y compared to the adjusted ONGAEIR total of 35,508 mt/y, with maintenance emissions excluded. The MAES MII model total was 45,207 mt/y. With ONGAEIR maintenance emissions of 14,880 mt/y added in, the MAES MII model total increased to 60,087 mt/y. This produced a state-wide ratio of 1.19 when compared to the ONGAEIR total (with maintenance emissions) of 50,388 mt/y. Results are summarized in Figure 48.

The statistical model based on the 2022 ONGAEIR dataset estimated statewide emissions of 145,766 mt/y using the CMS-informed distribution for the below-threshold rates, 102,554 mt/y using the Williams distribution, and 94,994 mt/y using the Sherwin distribution. These produced ratios of 2.89, 2.04, and 1.89, respectively when compared with the ONGAEIR total of 50,388 mt/y. Note that these results differ from those in the original report, as the statistical model now uses the same subset of ONGAEIR 2022 as the MAES model for consistency, which was not previously the case.

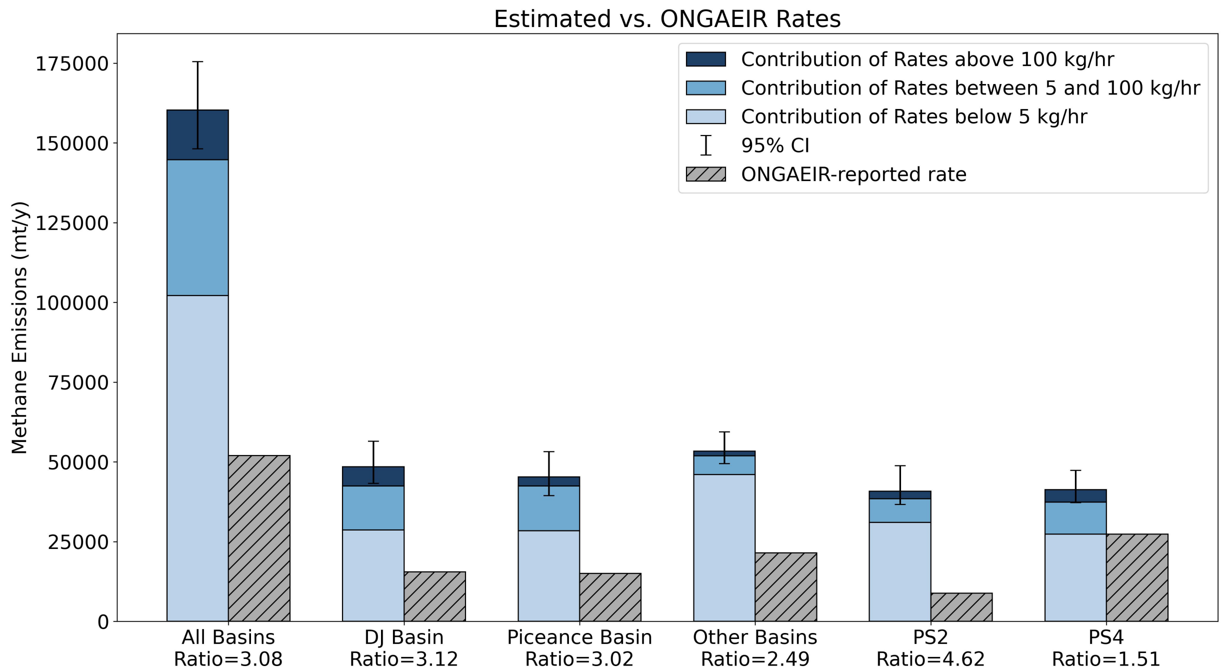


Figure 49: Statistical model MBI results based on the 2022 ONGAEIR dataset, reproduced from the June 2025 COBE Final Report

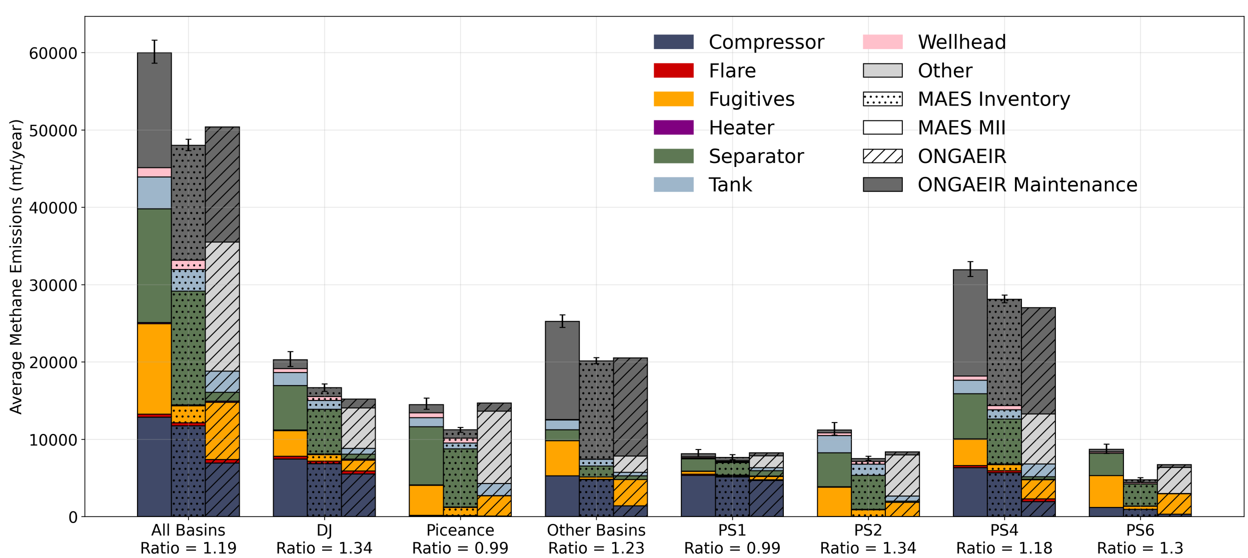


Figure 50: Update to Figure 48: MAES MII results based on the 2022 ONGAEIR dataset, showing contributions of equipment types to ONGAEIR emissions

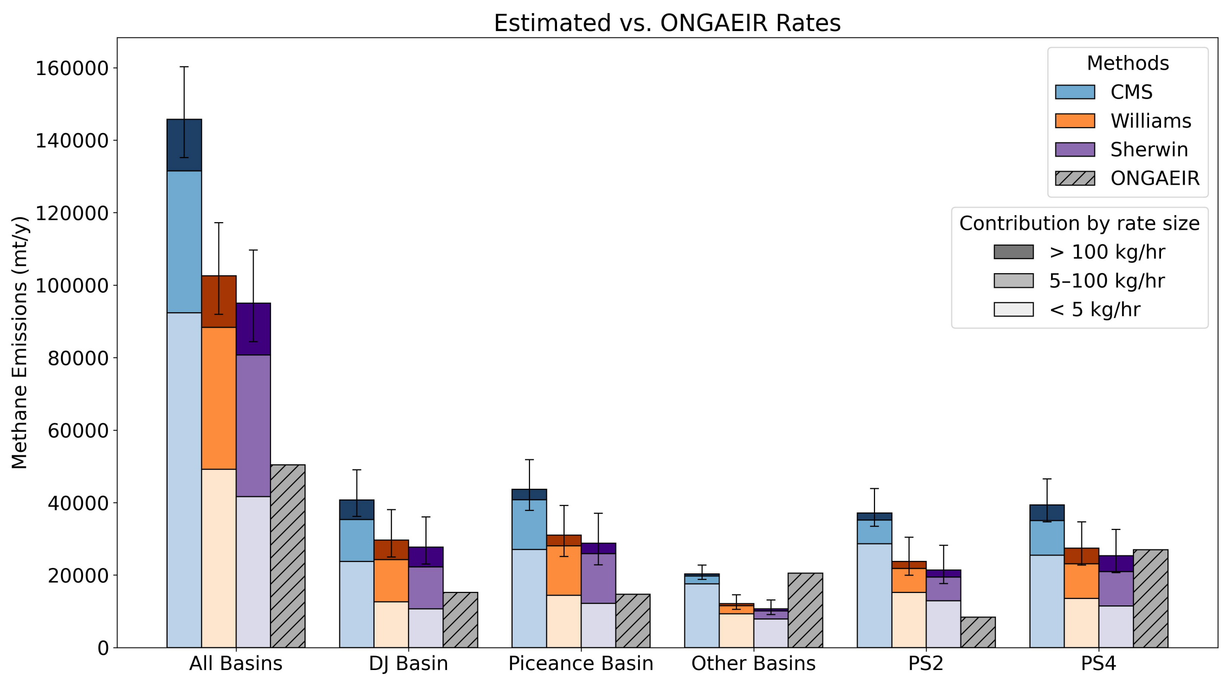


Figure 51: Update to Figure 49: statistical model MBI results based on the 2022 ONGAEIR dataset, including estimates based on Williams and Sherwin below-threshold distributions.