THESIS

INFORMING METHANE EMISSIONS INVENTORIES USING FACILITY AERIAL MEASUREMENTS AT MIDSTREAM NATURAL GAS FACILITIES

Submitted by
Jenna A. Brown
Department of Mechanical Engineering

In partial fulfillment of the requirements
For the Degree of Master of Science
Colorado State University
Fort Collins, Colorado
Fall 2023

Master's Committee:

Advisor: Bret Windom
Co-Advisor: Daniel Zimmerle

Nathaniel Blanchard

ABSTRACT

INFORMING METHANE EMISSIONS INVENTORIES USING FACILITY AERIAL MEASUREMENTS AT MIDSTREAM NATURAL GAS FACILITIES

Increased interest in greenhouse gas (GHG) emissions, including recent legislative action and voluntary programs, has increased attention on quantifying, and ultimately reducing, methane emissions from the natural gas supply chain. While inventories used for public or corporate GHG policies have traditionally utilized bottom-up (BU) methods to estimate emissions, the validity of such inventories has been questioned. To align with climate initiatives, multiple reporting programs are transitioning away from BU methods to utilizing full-facility measurements using airborne, satellite or drone (top-down (TD)) techniques to inform, improve, or validate inventories.

This study utilized full-facility estimates from two independent TD methods at 15 midstream natural gas facilities in the U.S.A., and were compared with a contemporaneous daily inventory assembled by the facility operator, employing comprehensive inventory methods. Methods produced multiple full-facility methane estimates at each facility, resulting in 801 individual paired estimates (same facility, same day), and robust mean estimates for each facility. Mean estimates for each facility, aggregated across all facilities, differed by 28% [10% to 43%] for the first deployment and nearly 2:1 (49% [32% to 68%]) the second deployment. Estimates from the two TD methods statistically agreed in 12% (97 of 801) of the paired measurements. These data suggest that one or both methods did not produce accurate facility-level estimates, at a majority of facilities and in aggregate across all facilities. Operator inventories, which included extensions to capture sources beyond regular inventory requirements and to integrate local measurements, estimated significantly lower emissions than the TD estimates for 96% (1535 of 1589) of the paired comparisons. Significant disagreement is observed at most facilities, both between the two TD methods and between the TD estimates and operator inventory.

Overall results were coupled with two case studies where TD estimates at two pre-selected facilities were coupled with comprehensive onsite measurements to understand factors driving the divergence between TD and BU inventory emissions estimates. In 3 of 4 paired comparisons between the intensive onsite estimates and one of the TD methods, the intensive on-site work did not conclusively diagnose the difference in estimates. In these cases, the preponderance of evidence suggests that the TD methods mis-estimate emissions an unknown fraction of the time, for unknown reasons. While two methods were selected for this study, it is unlikely that the issues identified here are confined to these two methods; similar issues may exist for other similar full-facility methods, on midstream and/or other facility types.

The results presented here have two implications. Firstly, these findings have important implications for the construction of voluntary and regulatory reporting programs that rely on emission estimates for reporting, fees or penalties, or for studies using full-facility estimates to aggregate TD emissions to basin or regional estimates. Secondly, the TD full-facility measurement methods need to undergo further testing, characterization, and potential improvement specifically tailored for complex midstream facilities.

ACKNOWLEDGEMENTS

I would like to thank the following people for their support and encouragement throughout this journey.

Dan Zimmerle: I want to express my gratitude for giving me the opportunity to work as a research assistant and for sharing you invaluable knowledge. I feel lucky to have worked under the godfather of methane.

My Family: For always believing in me and supporting me in this crazy journey.

Laurie Williams: For being a great mentor, friend, and professor, I don't know where I would be without you. Thank you for connecting me with Dan.

Rachel Day: For being a great friend and going through it all with me, we did it!!

Ben Luck: For blazing a trail that I've had the privilege to follow in.

Powerhouse colleagues: For creating a great work environment and helping me when needed.

Bret Windom and Nathaniel Blanchard: For serving on my committee.

Cheniere and SLR colleagues: For making this study possible and being a great team to work with!

DEDICATION

I would like to dedicate this to my Niece, Nina Rebecca Luck. May we address the pressing climate challenges today to ensure a more sustainable world for your generation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction and Background

Climate change poses an increasingly urgent challenge as our planet warms at an unprecedented rate. One significant contributor to this global crisis is greenhouse gas (GHG) emissions from human activities [1]. These emissions include a variety of gases, and among them, carbon dioxide (CO_2) and methane (CH_4). Methane is the second most prevalent GHG after carbon dioxide [2]. The main anthropogenic sources of methane emissions in the United States are livestock, transport and production of natural gas, landfills, and coal mining [3]. Understanding and accurately estimating methane emissions from these sources are critical steps in addressing this environmental challenge.

This study focuses on estimating methane emissions from natural gas systems using various techniques and methodologies to enhance the understanding of these emissions and contribute to effective mitigation efforts. Estimating methane emissions from natural gas systems can be approached through two distinct methods: bottom-up (BU) and top-down (TD). The BU methods use activity-based emission factors, which are data-driven estimates of emissions based on source-specific characteristics (e.g., equipment type, operating hours). They often rely on direct measurements or engineering calculations to estimate emissions at the component or source level. TD methods, in contrast, take a broader perspective, estimating emissions at the facility, regional, or basin level. These methods, employing techniques like aerial, satellite, or drones, are valuable for estimating emissions over larger areas but may have limitations in source-specific accuracy due to the broader spatial scale.

1.1 Study Design and Aims

This research was a subset of the Quantification, Monitoring, Reporting and Verification (QMRV) research and development (R&D) program which was funded by Cheniere Energy, Inc. Colorado State University (CSU) was a subcontractor for the program working under the engineering research firm, SLR Consulting, Ltd. The overall goal of this program was to reduce emissions across the natural gas supply chain, verify accuracy of reported emissions using multi-scale, multi-technology measurement methodologies, and improve emission transparency. Within the comprehensive program, the results discussed here are from the midstream sector of the supply chain.

The primary objective of this study is to investigate two fundamental questions: Do the estimates generated by using the BU methodology align with those produced by the TD methods? Additionally, is there concurrence between the estimates derived from two distinct contracted TD technologies across a diverse set of facilities? The answers to these questions hold significant relevance, as they directly influence regulatory decisions and methane fee assessments.

1.2 Background of Methane Emission Estimates

In 2021, over one hundred countries committed to the Global Methane Pledge, a collective effort aimed to achieve a 30% reduction in methane emissions by 2030 compared to 2020 levels [4]. In the United States, a significant step towards meeting climate goals was taken in 2022 with the signing of the Inflation Reduction Act (IRA) [5]. This landmark legislation represents one of the most ambitious actions ever taken to address the climate crisis within the United States.

The IRA relies on inventories for GHG emission accounting; the U.S. Environmental Protection Agency (EPA) monitors GHG emissions through two programs: the Greenhouse Gas Reporting Program (GHGRP) [6] and the Greenhouse Gas Inventory (GHGI) [7]. The GHGI, established in 1990, provides a comprehensive national overview of GHG emission estimates on a annual basis while the GHGRP, initiated in 2010, offers facility-specific GHG data annually [8]. The GHGRP offers a valuable tool for assessing emissions at the facility level, facilitating targeted regulatory efforts and emissions reduction strategies. In contrast, GHGI, with its national scope and comprehensive approach, plays a crucial role in shaping broader climate policy and tracking the overall progress of the United States in meeting its emission reduction commitments. Both programs are considered activity-based BU inventories, however, the validity of such inventories has been questioned. Emission factors are typically based on prior field studies carried out on facilities managed by various oil and gas (O&G) operators, often conducted several years earlier. It is widely recognized that the traditional BU inventories often underestimates emissions relative to the TD methods [9, 10, 11, 12, 13, 14, 15, 16].

Reporting programs have continually been forced to evolve in response to legislative changes and an expanding comprehension of GHG emissions' ramifications. Such progress reflects the dynamic nature of GHG monitoring and the imperative to adapt to emerging challenges and discoveries in the field. For instance, the IRA, mandates that the EPA revise its GHGRP methods by 2024 to utilize empirical data that can more accurately represent total methane emissions at O&G facilities [17]. In anticipation of this requirement, in 2022, the EPA proposed revisions for reporting emissions with a focus on incorporating empirical data obtained from TD methods and enhancing existing methodologies for improved accuracy [18]. TD technologies have seen advancements and the technologies "provide reliable methane emission monitoring and quantification" [19]. While these technological advancements are essential for improving the representation of facility methane emissions concerns regarding the accuracy and uncertainty of TD methods persist.

TD measurements are typically conducted within minutes to hours, whereas inventories traditionally provide annual estimates of emissions. The spatial and temporal scale of TD measurements differ from traditional inventories, posing a challenge for direct comparisons. The process of mapping these short-duration measurements to inventory estimates has not been previously demonstrated for midstream facilities.

One approach to use empirical data to inform inventories is through *measurement informed inventory* methods. While variations exist between methods, this represents a significant evolution in the field of emissions estimation, offering a more data-driven and comprehensive approach to quantify GHG emissions. The method requires (a) a **daily** BU inventory estimate, (b) TD full-facility estimates, and (c) aligning the inventory to the same temporal basis of the snapshot estimate. For (a), the inventory is estimated by multiplying activity counts with emission factors while integrating supplemental measurements to maximize accuracy. For (b), independent TD methods are employed, utilizing advanced technologies to capture full-facility emissions. These TD methods, may have limitations such as method detection limits (MDLs) that require supplementation with additional data, often drawn from inventories. For (c), this requires detailed facility operational information for the measurement day(s) to ensure the two estimates represent the same facility emissions profile. With this information, the TD estimate can identify areas where the inventory may not be capturing or is underestimating emissions from certain sources. By conducting these assessments multiple times throughout the year, the inventory should progressively improve its accuracy, resulting in a more robust and reliable emission estimate.

This research aims to explore these BU and TD methodologies further by deploying them at various facilities, and represents the first rigorous test of *measurement informed inventories* for midstream facilities.

1.3 Document Structure

This thesis is structured as follows:

- Chapter 2 presents the results from the baseline phase. This section focuses on the disagreement between the two TD methods and the TD/BU disagreement.
- Chapter 3 describes the results from the End-of-Project (EOP). This section focuses on a more in depth analysis of the disagreement between the two TD methods.
- Chapter 4 discusses the conclusions from the study and recommendations for future studies.

Chapter 2

Informing methane emissions inventories using facility aerial measurements at midstream natural gas facilities

This chapter was published in the *Environmental Science and Technology* journal [20].

2.1 Introduction

Numerous U.S. domestic and international initiatives [21, 22] have increased focus on GHG emissions reductions. In the U.S., the National Climate Task Force has set groundbreaking goals to reduce domestic GHG emissions 50-52% below 2005 levels by 2030 [23]. These efforts have focused attention on reducing methane emissions; methane is the second most common greenhouse gas after carbon dioxide (CO_2), and has a global warming potential (GWP) 86 [2] times that of CO_2 on a 20 year basis.

The O&G supply chain represents a substantial source of methane emissions, notably from the production, transport, and use of natural gas [7, 24, 12]. The natural gas supply chain is commonly divided into production, midstream, and distribution sectors; this study considers only facilities in the midstream sector; a related paper discusses similar work in production [25]. The midstream sector is commonly further divided into gathering and processing (G&P) and transmission and storage (T&S) segments. Midstream facilities are more complex, and often have larger structures and buildings, than production and distribution facilities.

Nearly all midstream facilities include gas compression equipment augmented by inlet and inter-stage separators that remove liquids from gas streams and tanks to store liquids. In many cases, the largest methane emitters at midstream facilities are from compressors and compressor drivers [7, 26, 27]. Gas processing plants and some gathering compressor stations include additional processing equipment to upgrade gas to pipeline quality, such as dehydrators, acid gas removal units, and associated flares and tanks. Most midstream facilities also include miscellaneous equipment to support pipelines, fuel systems, and similar functions [28]. Finally, storage facilities also include wells and wellhead equipment to store gas in underground reservoirs. This study includes eight transmission compressor stations, five gathering compressor stations, one gas processing plant, and one underground storage facility.

Multiple recent studies have characterized methane emissions from the O&G sector, primarily focusing on emissions in production basins. Broadly, these studies indicate a tendency for inventories to underestimate emissions relative to facility-scale or regional estimates of emissions [9, 10, 11, 12, 13, 14, 15, 16]. The persistent disagreement between inventories and measurement-based estimates has placed additional focus on changing traditional inventory methods to improve emission estimates. New regulatory measures that include methane fees, coupled with public commitments by companies to environmental, social and governance (ESG) programs, have also raised pressure on inventories to produce accurate and defensible results.

Recent U.S. legislative action includes the 2022 Inflation Reduction Act (IRA)[17], which funds a number of initiatives to reduce GHG emissions. As with most governmental initiatives, the IRA relies on the inventories as the basis for GHG accounting; for the U.S. these inventories are the EPA GHGRP [6] and the GHGI [7]. Both are considered activity-based BU inventories, and have been used as basis for public policy decisions [29]. By 2024, the IRA requires the EPA to revise its GHGRP reporting methods and establish methane emission fees based on empirical data that "accurately reflect the total methane emissions." [17] The IRA also allows owners and operators of applicable facilities to submit empirical emissions data for compliance with the methane fees. As an indicator of broad support for direct measurements, in 2023, corporate investors filed 10 resolutions [30] calling for direct measurements instead of conventional inventory estimates of emissions. The combination of investor action and regulatory changes indicates broad support for direct measurements of emissions.

One promising approach in developing empirical datasets to enhance inventories at a facility-scale is to use multi-scale measurements [25]. These approaches are broadly identified as *measurement informed inventory (MII)* methods. While variations exist among proposed methods [31, 32, 33], all typically include a three-step process:

1. Traditional BU inventory methods (activity counts × emission factors) and supplemental measurements are used to produce a reference estimate of emissions, herein termed the 'operator estimated inventory (OEI)'. In most cases, inventories are enhanced by comprehensively identifying all possible sources and utilizing per-source measurements, when available, in place of emission factors.

2. Independent TD estimates are performed using methods that measure full-facility emissions (either directly or by summing emissions from major sources). Since these methods have limitations, including MDLs that vary with environmental conditions, some of theses methods may not necessarily capture all emissions from a facility. Therefore, some of these methods are often supplemented with other information, such as per-source measurements and/or estimations based on BU emission factors, to produce a complete estimate of emissions, or an *adjusted* TD estimate.

3. The OEI and TD estimates are aligned to the same time basis and compared to assess inventory completeness and inform changes to improve inventory results.

In this work we will examine these three steps by focusing on the time period when measurements underlying the TD estimates were conducted - typically one working day (the time basis in step 3, above). This represents a subset of a complete MII process, which must also account for source intermittency and episodic events, like blowdowns or upset conditions, changes in the operational state of facilities, etc. By focusing on the measurement period, TD and BU methods can be compared with minimal complications, highlighting how information can be exchanged between these methods to improve emissions estimates.

While theoretically estimating the same emissions, fundamental differences between TD and OEI methods complicate comparisons for midstream facilities. Three key differences need to be considered:

First, the TD methods typically measure over short time periods – from seconds to hours. In contrast, OEI methods sum a set of point-source estimates, which are inherently based upon observations made over different time frames - from hourly logs of operating conditions to short-duration point-measurements of individual sources. The averaging in inventory methods works well for long-duration (typically annual) emissions estimates, but may not directly account for the specific conditions when TD methods were sampling.

Second, OEI methods rely on emission factors when measurements cannot be made. These factors typically originate from prior field studies conducted years earlier on facilities operated by multiple O&G operators. While well-performed studies include the full range of emissions from each source category, emission distributions are inherently averages across facility types, operating methods, failure modes, etc. Further, OEI methods use only the mean emission factor, and therefore do not account for extremes in emission rates which may exist on any one facility or for any one source.

Third, midstream facilities frequently change *operational state* – primarily by starting, stopping or changing load on compressors or processing equipment. These state changes drive substantial variation in emissions from the facility [34]. For example a state simulation conducted on a mid-sized gathering compressor station could see methane emissions vary by 350% due solely to compressors changing operating mode (Section A.3.3). State changes may also occur frequently, often multiple times per day.

The work conducted for this study was part of a larger QMRV program [35]. The QMRV R&D program includes modules for all sectors of the natural gas supply chain, and each module defines three phases:

- A **baseline phase** resulting in an TD-OEI comparison, as described above.

- An **enhanced monitoring phase**, where operators monitor emissions, estimate/update monthly OEIs, and monthly MIIs are computed.
- An end-of-project **verification phase** when another TD-OEI facility comparison is done, and final MII analysis is completed.

Emissions data were collected at 15 facilities, dispersed across four states, and operated by six midstream companies. This paper analyzes results from the completed baseline phase.

2.2 Methods

All facilities in this study were measured *as found*, during a pre-planned measurement period, typically one work day. Facilities were operated as if measurements were not occurring, with the measurement teams working around any on-site activities, including maintenance work and operational changes.

Reconciliation is a process whereby the TD estimates are compared to the OEI. This requires three methodological elements: Characterizing the TD measurement methods, estimating uncertainty, and adjusting TD estimates to fully estimate facility emissions.

2.2.1 Measurement Methods

Solution 1 [36] and Solution 2 [37] were contracted to provide TD estimates at all facilities.

Solution 1 utilizes a downward looking laser system that sweeps perpendicularly across the direction of flight of an aircraft, and uses differential absorption to compute path-integrated methane concentration (ppm-m) from the aircraft to the ground. This system can detect concentrated (point source) emissions that produce sufficient imaging contrast to separate the emissions plume from background methane concentrations. Solution 1 utilizes wind data, particularly wind speed, obtained from nearby weather station(s) to compute an emission rate from the plume image (i.e. 2-dimensional concentration map). This approach eliminates the need to install an anemometer on the facility. Solution 1 also collects visual imagery and superimposes plume data on the photos to provide context for the detections.

In this study, Solution 1 typically screened (hereafter termed an ‘overflight’) each facility twice during one day, typically making multiple passes to complete each overflight. Each facility overflight takes less than an hour, while individual plumes are scanned in a few seconds. Since each overflight has multiple passes, an individual emitter may be characterized by several plumes. If the emitter appears in more than one pass, the emissions from that source are averaged and included only once. Therefore, the estimated facility emissions for Solution 1 are the result of adding estimates from multiple distinct plumes into one overflight estimate. In this study, each overflight, typically separated by several hours, is treated as an independent estimate.

Solution 2 uses a flux plane method with concentration data from a miniature, tunable-laser spectroscopy sensor on a drone platform. The flux plane methodology is thoroughly discussed in literature [38, 39]. Briefly, the drone-mounted sensor measures methane concentration while making multiple downwind passes through the facility’s emission plume at different heights, typically from near ground level to above and outside the plume – i.e., above any methane enhancements from the equipment being measured. Concentration measurements, multiplied by the normal of the wind speed through the plane of flight, are integrated across the flight plane to calculate an emission rate.

Solution 2 deploys a ground meteorological station to collect wind speed data, which is then utilized to apply directional corrections to the drone while in flight. Solution 2 typically flies multiple flight planes within the facility to estimate emissions from subsets of equipment; the total facility estimate is the sum of all subsets. While each flux plane is completed in 20-40 minutes, it takes several hours to collect a full-facility estimate when measuring multiple subsets of the facility.

2.2.2 Solution Uncertainty

Both TD methods utilize proprietary algorithms to translate measurements from onboard instruments into emissions estimates. Uncertainty of each estimate is a function of multiple input uncertainties, ranging from instrument uncertainty (typically small) to larger wind field and algorithmic uncertainties. Therefore, this study uses uncertainty estimates from controlled testing of the integrated solution.

Controlled release data from studies by Bell et al. [40] and Corbett et al. [38] were used to create an uncertainty model for Solution 1 and Solution 2 respectively. See Sections A.2.3 and A.2.5.

When Solution 2 subdivides the facility to create partial-facility estimates, analysts may subtract an emission estimate from one flux plane estimate from another flux plane estimate to account for upwind emissions. Subtraction impacts uncertainty of the estimates. Since reports from the vendor do not indicate when subtraction was used, uncertainty was not adjusted for this impact.

Both methods are dependent on environmental and atmospheric conditions resulting in detection and quantification limits, see Section A.2. For example, methane in compressor driver exhaust (‘methane slip’) may be too dilute to be visible to Solution 1’s imaging or emissions may pool or recirculate near large compressor buildings, possibly complicating or distorting either solution’s quantification estimates. The controlled testing used single point sources in known locations that had less complex configurations than at midstream facilities, which have multiple, potentially overlapping and/or intermittent, emission sources and large structures that produce complex near-field winds. Therefore, the uncertainty models used here should be considered minimum estimates of uncertainty; in field conditions at complex midstream facilities, uncertainty is likely larger.

Finally, the methods differ on how brief, small, emission events (e.g. intermittent gas releases from pneumatic controllers) are quantified. Timing of these events is unknown. Assuming random timing, some portion of these emissions will be transported by the wind and increase methane concentrations seen by the Solution 2 sensor, increasing Solution 2 estimates. The impact is unknown; in this study we assume these emissions are sufficiently random and mixed that Solution 2 detects methane enhancements that are representative of mean emissions from these sources. In contrast, these emission events create insufficient plumes to be detected during a Solution 1 overflight, and have minimal impact on Solution 1 estimates. Therefore, to compare the two solutions, we add an estimate of these emissions to Solution 1.

In contrast, large episodic emissions, such as compressor unit blowdowns (depressurization of a compressor), can be identified by onsite observers. Since neither method can accurately estimate emissions from these events, estimates made during large episodic events are discarded, see Section A.3.1.

2.2.3 Aligning TD and OEI estimates

A series of adjustments is required to bring the OEI and TD estimates to the same temporal basis and to accommodate differences in detected emitters.

OEI methods estimate emissions by identifying emission sources, and estimating emissions by one of two methods: (1) estimating activity data for each source and multiplying by an appropriate emission factor (e.g. pneumatic controller emissions), or, (2) using a source-level on-site measurement (e.g. component leaks). Sources were either known (e.g. compressor vents, blowdown stacks, etc.) or are discovered via leak surveys; optical gas imaging (OGI) surveys and a Hi-Flow sampler were used in this phase of the project. In this study, source-level measurements were made where possible (Section A.1); other emitters were estimated using emission factors. OGI surveys were conducted throughout the day, typically measuring a source only once (with the exception of the compressor-dependent emitters discussed below), therefore, all emitters were assumed to emit at a constant rate throughout the measurement period.

As noted earlier, a change in operating state of the facility may change emissions significantly. The primary factors driving emissions at a facility are the online status of compressor or processing units, the load on these units, and changes to process settings. Emissions at the facility may also be influenced, typically to a lesser degree, by other process equipment, such as gas upgrading equipment. Prior studies have indicated that emissions driven directly by the operating state of compressor units often dominate overall emissions from midstream facilities [27, 26]. Therefore, compressor operating state provides a useful surrogate for the facility state at midstream facilities, and are, of course correlated with the throughput of compressor stations.

To align estimates, three steps were taken. First, operators were requested to calculate a daily OEI estimate so that individual operating states could be extracted. Operators also provided a log of when compressor units were operating and/or pressurized. This allowed the OEI to be decomposed into one OEI estimate for each operating state. Note that OEI estimates are assumed constant for each operating state; this effectively averages estimates for variable emitters (e.g. intermittent pneumatics or dump valves on compressor interstage separators).

Second, an independent observer from the study team noted compressor states and the time of large episodic emissions, such as blowdowns, compressor starts, etc. While episodic emissions need to be tracked and included in long-duration emissions totals, comparisons performed here exclude episodic events from OEI estimates. TD estimates made during these episodic events were eliminated from resulting comparisons. This simplification has no impact on TD-OEI comparisons; TD methods cannot reliably measure total emissions from episodic events, and the same engineering calculations would be used to estimate these emissions for both the OEI and TD estimates.

Third, TD estimates were compared to OEI estimates in the same compressor operating state. For Solution 1 overflights, comparable states were identifiable. Solution 2 measured over longer periods, increasing the likelihood that operating state would change during measurements. This type of impact was seen at 3 of 15 facilities. Since operating state changes may have created substantial variations in emissions, the comparison weighted the OEI estimates by the time in each state over the measurement period (Section A.2.6). Using a weighted average OEI assumes that total estimate of emissions from Solution 2 effectively averages emissions in a manner similar to the weighted average OEI. The validity of this assumption is impossible to test and is indicative of the challenges of comparing TD estimates to complex time-averaged OEIs in any MII program.

In addition to temporal alignment, estimates need to be adjusted to represent, but not double-count, all emissions from the facility. Adjustments are method-specific. For example, Solution 1's method typically omits emitters below the method's MDL; these need to be added to the Solution 1 estimate. Additionally, quality review indicated that Solution 1 did not consistently detect and quantify methane in compressor driver exhaust (compare Figure A.2 versus A.3). An expert panel was utilized to estimate whether Solution 1's estimate included combustion slip (Section A.2.2). Results indicated that Solution 1 detected exhaust plumes at 3 of 8 facilities where substantial combustion slip would be expected due to the type and size of compressor drivers on the facility. If Solution 1 did not detect exhaust at a facility, emissions from combustion slip were estimated using recent stack tests, if available, or emission factors if not.

In theory, Solution 2 captures emissions from all emitters upwind of the flux plane that were active within the transport time of the emissions. However, exceptions exist for flights that did not extend high enough to traverse exhaust plumes from compressor drivers (Facility C). Since the drone transects are fast, additional uncertainty exists for unsteady emission sources near the flux plane (e.g. intermittent gas pneumatics) or in cases where wind near a building may attenuate or recirculate emissions. Unfortunately, no data exist to quantify these uncertainties. Finally, as noted earlier, multiple flux planes within one facility require adjustments by Solution 2's analysts to avoid double-counting emissions. In this study we assume these corrections were perfect, and no emissions were transported through two flux planes without correction. (Section A.2.4)

The adjustment process outlined here describes the necessary TD adjustments for midstream MII methods to construct complete estimates of emissions. Similar methods are likely required for other sectors with highly variable operating states.

2.2.4 Calculating the MEC

After the TD methods were adjusted to represent a full-facility estimate, all independent estimates were averaged to produce the measurement emissions check (MEC) – typically one Solution 2 estimate and two *adjusted* Solution 1 estimates for each facility (see Section A.3.1 for exceptions). The MEC serves as a useful basis for plotting (X-axis in figures below) and for normalizing comparisons (Y-axis). However it is *not* used for comparisons. When comparing between any pair of estimates, each comparison was done with one TD method individually to highlight differences between the TD methods.

2.3 Results & Discussion

Results are presented as a series of full-facility emission comparisons, with all independent estimates brought to a comparable basis as described in *Methods*. For these comparisons, *adjusted* Solution 1 and Solution 2 estimates are used, compared to each other and to the OEI. In the following discussion, the reader should note that any mention of Solution 1 refers to the *adjusted* estimate. Table 3.1 provides information about all the enrolled facilities.

Table 2.1: Facility Information

Facility Characteristics					**Compressor State** in meas. period		**Estimates (kg/h)**	
Facility ID	Supply Chain Sector[†]	Compressor Driver Type[‡]	Number of Compressor Units at Facility	Engine Class[*]	Units Operating[#]	State Change[+]	Operator Estimated Inventory (OEI)	Measurement Emissions Check (MEC)[‖]
A	G&P	Recip	5	4SLB	Yes	No	93.4	263 [193 to 331]
B	G&P	Electric	3	–	Yes	No	26.2	545 [403 to 724]
C	T&S	Turbine	4	–	Yes	No	23	141 [115 to 165]
D	G&P	Recip	11	4SRB	Yes	No	59.3	63 [52 to 80]
E	G&P	Recip	15	4SLB	Yes	Yes	117	694 [574 to 818]
F	G&P	Recip	12	4SLB	Yes	No	95.6	172 [124 to 216]
G	T&S	Turbine	5	–	Yes	No	18.6	57 [46 to 72]
H	G&P	Recip	10	4SLB	Yes	Yes	58.6	70 [52 to 88]
I	T&S	Turbine	1	–	Yes	No	0.85	41 [30 to 54]
J	T&S	Turbine	1	–	No	No	2.14	17 [13 to 21]
K	T&S	Recip	8	2SLB	No	No	6.4	79 [65 to 97]
L	T&S	Recip	5	4SLB	Yes	Yes	97.3	866 [689 to 1,144]
M	T&S	Recip	6	2SLB	Yes	No	21.8	75 [59 to 92]
N	T&S	Turbine	2	–	Yes	No	13.8	40 [31 to 53]
O	T&S	Recip	8	2SLB	No	No	62.4	300 [230 to 383]

[†] Supply chain sector of the facility: G&P = Gas Processing, T&S = Transmission and Storage

[‡] The type of mover driving the compressor(s) at the facility: *Recip* = reciprocating (piston) engine, *Turbine* = combustion turbine

[*] For facilities with reciprocating (piston) engines, code indicates the type of engine: *2SLB* = two-stroke, lean-burn, *4SLB* = four-stroke, lean-burn, *4SRB* = four stroke, rich-burn

[#] Yes if any compressor units were operating during the measurement period

[+] Yes if any compressors changed state during the measurement period

[‖] The MEC is typically the average of one Solution 2 estimate and two *adjusted* Solution 1 estimates

2.3.1 Solution Comparison

While most MII efforts would deploy only one TD method, this study had access to two TD technologies to support additional analysis. The methods selected for the study both create facility estimates, but use distinctly different measurement instruments and analysis algorithms. Since both were deployed contemporaneously (same day in most cases, see Section A.3.1), these estimates can be compared to better inform the uncertainty of both methods.

Figure 2.1 uses a Bland-Altman difference plot to compare the two TD estimates. This analysis plots relative difference between two estimates against a common estimate of emission rate. The facility MEC was used for both the X axis and as an average estimate to normalize relative differences. Each point represents one estimate at one facility. Error bars indicate a 95% confidence interval for each estimate; see Section A.2.3 and A.2.5. For each facility, there are two independent Solution 1 estimates, completed at different times, which appear as points immediately above each other. As depicted in the figure, as the MEC increases the difference between the two Solution 1 estimates becomes more pronounced.

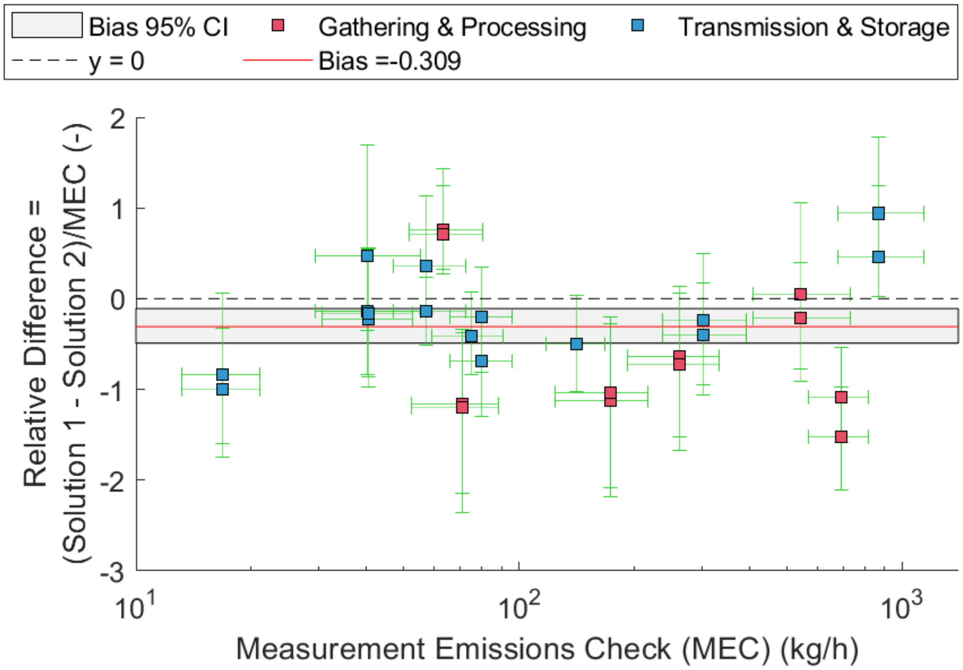


Figure 2.1: The horizontal axis is the MEC, an average of the TD estimates, the vertical axis is the relative difference between the *adjusted* Solution 1 and Solution 2 estimates for each facility. The gray box displays the 95% confidence interval over all facilities. The dashed line (y = 0) displays perfect agreement.

Aggregated across all facilities (gray box in figure), *adjusted* Solution 1 estimated emissions were 1,073 [387 to 1,586] kg/h lower than Solution 2, a statistically significant difference of 31%

[12% to 47%]. For pairwise comparisons at individual facilities, the two TD estimates statistically agree in 2 of 28 comparisons (Kolmogrov-Smirnov 2-sided, $\alpha = 0.05$), with Solution 1 reporting emissions that were statistically lower than Solution 2 in 20 comparisons and higher in 6 comparisons. Results are compared using other methods in the Section A.3.2.

Method disagreement could be compounded by changing operating states during the measurement period. Eliminating the 3 facilities where the operating state changed during measurement increased the fraction of comparisons that agree by ≈3%, from 2 of 28 to 2 of 20, and decreased the fraction of comparisons that disagree by a similar amount; this is likely not a significant source of disagreement between methods. See Table 3.1, column "State Change" for the specific facilities.

While emissions *could* change while the facility is in one operating state, operational knowledge indicates that large changes were unlikely. Therefore, since both methods measured when emissions were essentially stable, it is likely that the TD methods, in field conditions at midstream facilities, have uncertainties larger than indicated by controlled testing, likely due to multiple emission sources and complex wind fields at these facilities.

The above analysis of TD method uncertainty impacts the confidence in the TD estimation component of MII methods, and offers guidance for calculating these results. Outside an R&D program, most MII programs would utilize only one TD estimate at a time. Therefore, uncertainty in the TD estimate would be inherently difficult to assess using one estimate at each facility. Inventories could be in or out of agreement with TD estimates due solely to uncharacterized uncertainty in the chosen TD method. Further, most reporting programs report only mean values without uncertainty, and have no systematic method to capture uncertainty. Taking the current study as a test case, had all 15 facilities been part of a single report, the aggregate uncertainty in the TD methods, without reference to any inventory methods, would be at least 31% – the difference in the mean estimate of total emissions between the two TD methods.

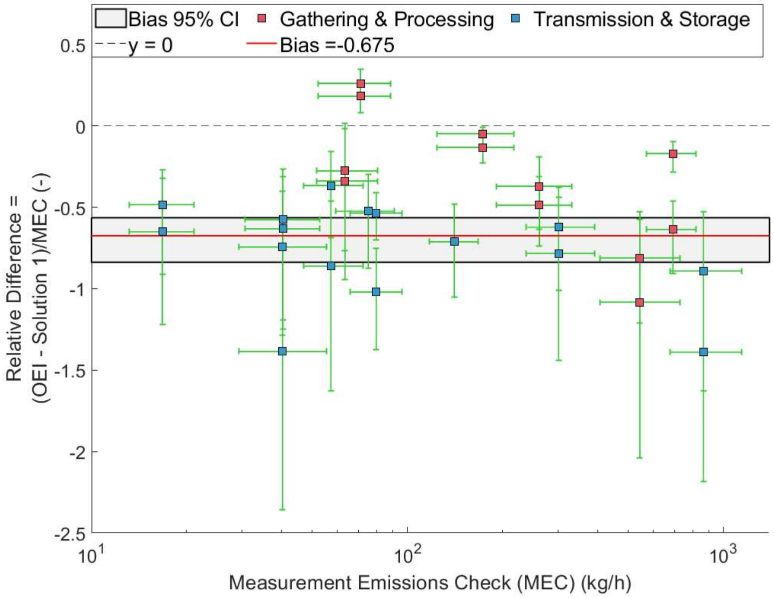
These results indicate a need for better controlled release and field testing of these methods at complex facilities to better characterize TD method uncertainties, and a need for MII protocols to consider further assessment and reporting of TD uncertainty.

2.3.2 OEI Solution Comparison

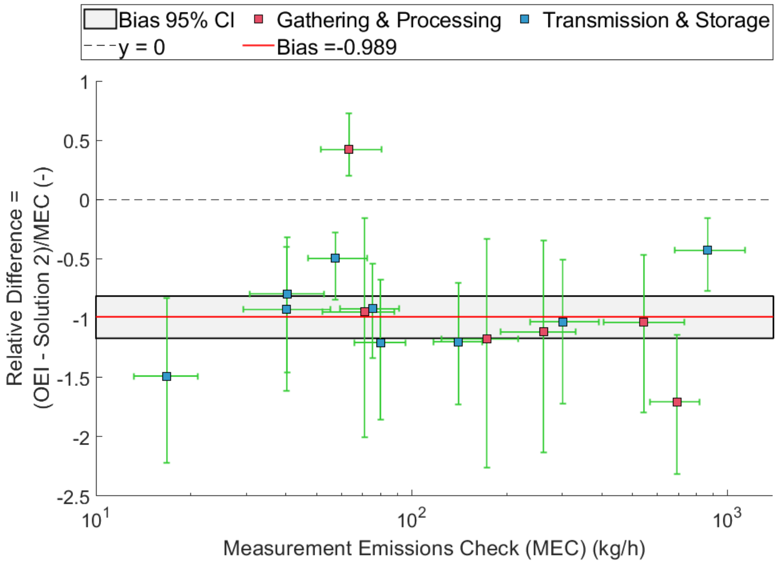
The second analysis of importance is the comparison between the OEI – i.e. the BU inventory – and TD estimates for each facility and for all facilities aggregated together. This comparison uses the adjusted full-facility estimate from each TD method, compared to the calculated OEI.

An OEI was estimated for each of the 15 facilities using traditional inventory methods augmented with supplemental sources, emission factors from recent studies, and other modifications to assure the inventory captured all known sources (see the Emission Calculation Guidance Tool attached to the SI). The guidance tool informs how to augment routine emission factors with direct measurements for significant sources. For example, stack tests were used to directly measure methane slip in combustion exhaust, in an as-found conditions.

Therefore, the OEI process used in this study is among the most robust used in any program. Figure 2.2 illustrates the difference between the OEI and the TD estimates, using the same plot format as Figure 2.1.



(a) Comparison of the OEI to the *adjusted* Solution 1. Where two points are vertically aligned, each represents a Solution 1 overflight, typically two per facility.



(b) Comparison the OEI to Solution 2

Figure 2.2: Comparison of the BU inventory (OEI) to each of the two TD methods, Panel (a) for *adjusted* Solution 1 and Panel (b) for Solution 2. The horizontal axis is the MEC, an average of the TD estimates, the vertical axis is the relative difference between the OEI and *adjusted* TD estimate for each facility. The gray box displays the 95% confidence interval over all facilities. The dashed line (y = 0) displays perfect agreement.

When comparing OEI to TD methods, it is important to note that OEI estimates have no stated uncertainty. Therefore, uncertainty in all comparisons is exclusively from the uncertainty estimate of the TD method. For both TD methods, the operator's inventory is significantly lower than either of the TD methods at the vast majority of facilities; the OEI is statistically higher than Solution 2 in 1 of 15 comparisons, and statistically higher than Solution 1 in 2 of 28 comparisons. Although TD estimates may have higher uncertainty than shown, logical increases in uncertainty are unlikely to close the gap between the TD and OEI estimates.

Given the diversity of reporters and facility types, the significant disagreement between BU inventories and TD estimates indicates systematic under reporting in the OEI and/or issues in the TD measurement. Disagreements are neither company-specific nor facility-type specific.

The intent of the MII process is to utilize measurement-based estimates to improve the inventory process for a specific facility (or small group of facilities) over results using traditional inventory methods. If successful, the improved inventory process would then be used to accurately estimate emissions for each of the facilities' operating states, over extended periods – the result needed for regulatory reporting, voluntary initiatives, and similar purposes. The question then, is how *do* TD measurements inform inventories? In this study, the TD/BU disagreement could originate from three potential causes (or any combination of the three):

1. *Known sources are systematically underestimated in BU inventories.* Midstream facility emissions are primarily dominated by a limited number of well-known sources (e.g. leakage through large valves, venting from compressor seals, etc.). [26, 27, 41] All such sources are included in the QMRV OEI calculation, using on-site measurements wherever possible, or best-in-class emission factors when measurement was not possible. Therefore, for known sources to drive the TD/BU disagreement, both measurements and/or emission factors for a large number of key midstream sources would need to be systematically underestimated relative to common field conditions.

   While there is little evidence for systematic low bias, it cannot be ruled out. Some sources at facilities are difficult to measure, due to the size of equipment (e.g. a large diameter blowdown stack), accessibility of an emission location, or safety issues. For on-site measurements, the largest emitters often prove the hardest to measure [27]. Systematic errors in these measurements are possible, and could contribute to the TD/BU disagreement. BU emission factors are typically developed from the results of field studies, and any given facility may have a different emissions profile than the field study. While this should not lead to systematic bias, bias cannot be ruled out. Additional on-site measurements and/or continuous metering of some sources may be required to characterize major emission sources across multiple operating states.

2. *Un-inventoried, large, emission sources exist on many facilities.* One emitter of this type was discovered on one facility during TD measurements (a leak in a fuel gas system, included in data presented here and corrected the day after identification). No similar large sources were found on the other 14 facilities. Therefore, *if* large emitters were to explain the gap between the TD estimates, multiple such sources must exist on most facilities *and* these sources must have unknown characteristics that made them undetectable by Solution 1 overflights. While there is no evidence for this type of large, systematic, emission source on midstream facilities, it cannot be ruled out. Per-source screening and measurements would likely be required to either identify these sources or increase confidence that they do not exist.

3. *TD estimates are biased high, and this bias is systematic for this type of facility.* As with the other potential causes, there is little evidence of systematic high bias in TD methods, in part because these methods have not been extensively tested on complex facilities like compressor stations under controlled conditions. Current controlled testing was performed at near-ideal conditions - single source, no nearby structures, etc. Winds may recirculate emissions near large structures (where both TD methods estimate emissions), or multiple nearby emission sources may complicate rate recovery algorithms.

Given the size of the TD/BU disagreement, all three of the above causes deserve evaluation. The above analysis indicates that, for midstream facilities, an ‘informed’ inventory will likely require a more comprehensive program than simply performing periodic TD measurements at a facility; full-facility estimates need to be augmented with additional, on-site, diagnostic work and measurements. For example, many vents or other sources on midstream facilities could be metered and monitored over extended periods, and compared to TD estimates.

2.3.3 Extending to Long-Duration Estimates

The ultimate goal of the MII is to establish total facility emissions over extended periods; monthly or annual estimates are common. All analysis above considers the simplified case where (nearly) contemporaneous TD estimates were compared to an OEI estimate made for the facility. This comparison was typically limited to emissions measured in a single operating state, even when a operating state changed during the day (3 facilities), full estimates in each state were not possible. Emissions in operating states that were not active during measurements were not characterized by TD estimates nor measured by onsite methods. For example, a single compressor unit may be operating, pressurized not operating, or depressurized. In each mode, different valves are open and shaft or rod packing seals operate differently. Considering only compressors at a typical station, emissions may vary by 2.5-3.5 times due *solely* to changes in operating mode (see simulation in Section A.3.3).

As a result of state changes, MII methods must estimate emissions from operating conditions which were inactive during TD surveys or design TD surveys to better study the facility in *all* operating conditions. This implies that single snap-shot TD estimates are unlikely to replace annual BU inventory methods for midstream facilities until additional testing is completed, and that midstream facilities will require high-quality logging of operating state, per-unit and per-source emission factors and/or per-source measurements for the foreseeable future. This can be in the form of advanced monitoring and tracking systems of a facility's operational states, supplemented by identifying, measuring and including any unplanned emissions.

2.4 Implications

Inventories are an important tool for making policy and industrial decisions, and there needs to be confidence that these inventories are accurate. MII methods suggest that inventories can be verified by TD estimates that capture all emissions at a facility [42]. Results from this study confirm this reconciliation may be possible, but indicate several nuances that must be considered when using TD estimates to verify inventories at midstream facilities. First, for the diverse set of facilities considered here, a simple comparison of TD estimates to a contemporaneous BU inventory yields the unsatisfying result that these estimates disagree, with little guidance on how to eliminate the disagreement. Therefore, measurement informed inventories will likely require both full-facility TD estimates *and* a battery of diagnostic measurements or monitoring on-site – both to identify causes of disagreement between TD and BU estimates made in the same operating state and to capture operating states not included in the TD estimates.

Second, in this study, the TD/BU disagreement is systematic – at 40 of 43 TD-OEI comparisons, the TD estimates are statistically higher than the contemporaneous inventory. For midstream facilities there is no ready explanation of *why* this disagreement exists. While per-source measurements at facilities may under-report emissions due to challenges mentioned earlier, it is unlikely that these issues would explain the TD/BU disagreements seen in the study - at 35 of 43 comparisons, there is a disagreement by more than a factor of two. All known large sources are included in the enhanced inventory process used here, and additional large sources were not identified by the source-locating TD method (Solution 1) in sufficient quantity and size to account for the difference. Therefore, results of this study indicate the need for more extensive on-site identification and measurement of midstream sources and/or more complete and representative emission factor data, coupled with better characterization of TD method uncertainty for midstream facilities.

Chapter 3

Evaluating the divergence of two top-down methods at midstream natural gas facilities

This chapter is submitted to *Environmental Science and Technology* journal. But currently is in preprint. [43]

3.1 Introduction

The global initiative to reduce methane emissions from the energy sector has gained substantial momentum because of its pivotal role in combating global warming [4]. Methane emissions are the second most prevalent GHG after CO_2, with a GWP that is 86 [2] times higher than CO_2, over a 20-year time frame. Accurately quantifying these emissions stands as a fundamental prerequisite in the pursuit of effective methane emission reduction policies.

The natural gas system is the second largest anthropogenic source of methane in the U.S., spanning from production and processing to transmission and distribution[7]. This study focuses on the midstream sector which includes G&P as well as T&S segments. Midstream facilities are complex in nature and typically include gas treating and compression equipment, with the largest source of methane emissions being from compressors and compressor drivers, particularly reciprocating engines [44, 27]. This study enrolled five G&P and nine T&S facilities operated by six different midstream companies.

Methane emissions on midstream facilities are typically divided into three categories [27, 26, 6]. *Fugitive* emissions, at times called 'leaks' are unintended, unknown releases of unburned gas, *venting* is the release of unburned gas for known and planned reasons, such as maintenance activities or to drive valve controllers or pumps. *Combustion* or *combustion slip* refers to the fraction of fuel gas that remains uncombusted during any combustion process, such as engine or turbine exhaust, heaters, or other process equipment. All combustion processes have some slip.

For the last decade the EPA has required mandatory annual reporting of GHG emissions using the GHGRP [6], which relies on activity-based or BU inventory methods. Policy making decisions frequently hinge on these reported inventory estimates [17]. Recent regulations are shifting to require empirical data at facilities. In 2022, the EPA proposed the incorporation of empirical data to improve the accuracy of inventories, recognizing the advancements in TD technologies for reliable methane emission monitoring and quantification [45].

Various measurement methods have been developed to assess emissions using remote sensing (TD) techniques, including drones, manned aircraft, and satellite technologies. These methods are often contracted to conduct emission assessments at facilities, providing a comprehensive "snapshot" of whole-facility emissions. In the past, there existed no direct means to validate the accuracy of the BU inventory methodology. Recently, it is widely recognized that the traditional BU inventories often underestimate emissions relative to the TD methods [9, 10, 11, 12, 13, 14, 15, 16]. While it is evident that BU inventories might miss or underestimate certain emission sources, a critical question arises: Can we place full confidence in the accuracy of TD technologies in capturing the emissions profile of a facility? Many of these technologies have undergone controlled testing [40, 38, 46], but these tests are single point sources with near ideal conditions, which may not faithfully replicate the complexities of midstream facilities.

Other relevant studies have also explored the variability of methane emission estimation methodologies. For instance, Daniels et al. conducted research that utilized multi-scale technologies in the production segment and discovered substantial discrepancies in TD methods, with variations spanning over three orders of magnitude [47, 25]. Similarly, Stokes et al. deployed two distinct aircraft-based emission measurement systems at tank battery sites and observed discrepancies in their respective estimates of total emissions [48]. However, a notable gap remains in the literature when it comes to extensive comparisons between different TD methods. In-depth analyses directly comparing the outcomes of various TD approaches are limited, especially in terms of investigating the factors contributing to the observed differences. This research aims to bridge that gap by conducting a detailed examination of two widely used TD methods – Solution 1[36] and Solution 2[37] – systematically dissecting the underlying sources of variation between their estimates.

As new regulations mandate empirical data and levy methane fees, the accuracy of methane estimates gains heightened significance. Disagreements between TD methods measuring the same facility at the same time, indicates inaccuracy in one or both of the TD methods. Resolution of this issue is paramount to ensure that the site's emissions are both accurate and comparable across different methods, for all types of facilities.

This study is part of the QMRV R&D program [35], which is structured into three distinct phases: baseline, enhanced monitoring, and EOP verification. This program deploys multiscale measurement technologies to quantify emissions and spans all sectors of the natural gas supply chain. Recent studies, Daniels et al. [47] and Wang et al. [25], present results from the QMRV program focused on the production sector.

In the baseline phase, operators estimated facility emissions (for a 24-hour period) using BU inventory methodologies, while simultaneously two TD technologies conducted whole-facility emissions estimations. This setup facilitated a TD - BU comparison per facility. Data was collected over 15 facilities, all within the midstream sector. The results from the baseline phase are presented in Brown et al. [20] and a short summary is provided below:

- Analysis revealed a discrepancy of approximately 28% [10% to 43%] between two commonly deployed TD methods measuring the same 15 facilities. In aggregate the methods differed by 958 [368 to 1,455] kg/h.
- For pairwise comparisons at individual facilities, the two TD methods demonstrated statistical agreement (Kolmogrov-Smirnov 2-sided, $\alpha = 0.05$) in only 2 of 28 paired measurements.
- There was systematic disagreement between the TD/BU methods in 40 of the 43 comparisons, with the TD estimates being statistically higher than inventory estimates.

Despite the relatively small sample size of 28 paired measurements, the prior baseline study provided sufficient data to identify a significant and non-trivial issue: the substantial disagreement between the two popular, whole-facility, TD methods for these midstream facilities. This research aims to examine the divergence between these TD methods. To achieve this, we engaged the same two TD methods to perform more extensive repeat measurements at each facility during the EOP project phase. Additional repeat estimates address a fundamental experimental question from the prior study: Is the inconsistency observed between the TD methods a consequence of limited sample size, or a more substantial disagreement between the methods? It is worth highlighting that one facility was sold during the QMRV project term; thus, the facility did not participate in EOP measurements, leading to a reduction in the total number of facilities from 15 to 14.

3.2 Methods

This analysis focuses on two datasets. First, the two TD methods were deployed contemporaneously and provided multiple whole-facility estimates of emissions. Throughout this analysis, we refer to two separate deployments of these (and auxiliary) measurement methods. The *baseline* deployment was conducted early in the QMRV project. Results of this phase were presented in Brown et al. [20], and will be utilized here as well. The second deployment was the end of the project (EOP), 9 months after the baseline. The EOP deployment made adjustments relative to the baseline deployment to support the study presented here.

The second data source is an operator-provided BU inventory estimate that corresponds to the facilities operational state during the measurement period, using either the same methodology as in the baseline deployment, or an updated methodology including more measurement data. The inventory estimate included information about the facility operations during the measurement day - operational configuration for the 24-hour period and the timing of any episodic or upset events.

Although both methane and carbon dioxide are important to the overall study and QMRV program, this study (EOP estimates) focuses on methane emissions only; unless specifically indicated, this paper uses the term emission to mean methane emissions.

3.2.1 Field Deployment for EOP

All methods were deployed simultaneously on a single day, with two exceptions: adverse weather conditions prevented Solution 1 from conducting simultaneous flights at two facilities which were completed on the following day without any noted changes in facility operations. Other exceptions are noted in SI Section B.2. A typical measurement day follows Brown et al.[20] and is outlined in SI Section B.1.

3.2.2 BU Estimate

Operators were instructed to operate the facility *as normal*, so that the EOP estimates would provide a "snapshot" of methane emissions during typical operations. Operators of the enrolled assets estimated the facility emissions for the 24-hour day using BU methodologies. Operators typically based inventories on equipment count and GHGRP factors plus supplemental emissions from either calculation or additional measurements[20].

Although these inventory methods often result in underestimations of emissions compared to TD methods, they serve as a valuable baseline for understanding a facility's emissions profile. Given that these facilities had 1 to 15 compressors, combustion slip is a substantial emission source, particularly if reciprocating engines are used to drive operational compressors [44, 14]. When a compressor is in a pressurized standby mode (prepared for operation but not actively running), compressor vents may also be a large source of methane emissions. Except in rare cases where there is an usual process failure or large leak, emissions from compressors are often a more substantial contributor of emissions at midstream facilities than fugitive emission sources [26]. Midstream facilities may undergo operational state changes in response to varying demands throughout the day, underscoring the significance of monitoring compressor states for the inventory during the measurement period. A significant effort was made to avoid measuring temporary blowdown events by the TD technologies.

The prior study discusses a comparison between the BU estimate and TD estimates [20]. For this study, the primary use of the BU estimate was to estimate the change in facility emissions due to operational state differences between the baseline and EOP deployments. These estimates illuminate whether emissions should have increased or decreased between field deployments, which is critical for interpreting the results from the TD methods.

3.2.3 Top-down Methods

Two aerial methods, Solution 1 and Solution 2, were contracted to estimate a snapshot of emissions at the enrolled facilities during both baseline and EOP deployments. This study did not assess the methodologies of the two technologies; instead, it simply evaluates the output of these solutions as an operator would in real-world scenarios.

Solution 1 operates a remote sensing technology that is deployed via aircraft for methane emission detection and quantification. Using a laser they measure path-integrated methane gas concentration between the aircraft and ground. This technology differentiates emission plumes from background methane levels to detect and quantify concentrated point source emissions. Their proprietary algorithms utilize wind data from a near-by weather station to compute emission rates.

Solution 1 was instructed to perform as many whole-facility surveys as possible, in a single day, resulting in an increased number of estimates in the EOP relative to the baseline deployment. The method for grouping data into whole-facility estimates is described in SI Section B.1.1. The uncertainty for Solution 1 was estimated using the relative error in (857) controlled release tests by Bell et al. [40]. See SI Section B.1.2.

Solution 1 states a MDL of 3 kg/h with 90% probability of detection, but this varies with wind speed [40]. Therefore we need to adjust Solution 1 estimates to include emissions that were known to exist, but were below the MDL and may not have been detected by Solution 1's measurement technology. Emissions from a range of sources, including fugitive leaks < 1 kg/h, combustion slip from turbines and non-compressor sources like heaters and flares, typically register below Solution 1's MDL and were added to facility-scale estimates from Solution 1. The quantification of these additional sources was achieved through the BU methodology, following the same methodology as in Brown et al. [20]. Added sources varied between facilities (SI Section B.2). All references to Solution 1 results in this paper imply that the these adjustments have supplemented Solution 1's estimates. The exact uncertainty related to these adjustments was not available in the BU estimate, and these adjustments were made without uncertainty estimates.

Solution 2 uses a drone-mounted spectroscopy sensor to fly downwind passes through emission plumes at varying heights, a technique commonly known as a 'flux plane'. Estimated emission rates for all sources upwind of the flux plane are calculated by multiplying concentration estimates throughout the flux plane with the normal of wind speed through the flux plane, and integrating all estimates across the flux plane. Wind data originated from a portable onsite anemometer with additional data from corrections applied by the drone's autopilot system. In principle, Solution 2 detects emissions from all sources located upwind of the flux plane that were active within the time frame of emission transport. The uncertainty for Solution 2 was estimated using relative error in (12) controlled release tests by Corbett and Smith [38]. See SI Section B.1.2.

In a typical deployment, Solution 2 estimates emissions from individual "zones" or equipment groups, which would then be summed to create a whole-facility estimate; this method was utilized for the baseline deployment. For the EOP estimates, Solution 2 was instructed to make more comprehensive whole-facility surveys—capturing all facility emissions—rather than summing individual zones. In some cases, the meteorological conditions and/or the footprint of the enrolled facility prevented Solution 2 from estimating the entire facility with one flux plane. It was left to the discretion of the Solution 2 field team whether they could fly a whole-facility estimate in one flux plane flight, or if multiple flights (all downwind of the facility) were needed to capture the entire facility. Therefore, at some facilities (SI Section B.2), a single Solution 2 whole-facility estimate may include multiple flux plane flights.

3.3 Results & Discussion

We address comparison of the TD methods in four analyses:

1. Method consistency: Do the methods exhibit similar changes between baseline and EOP deployments?

2. Per-facility results: How do the methods compare when evaluated at each facility? Did the methods produce repeatable results during the EOP?

3. Aggregate comparison: If all 14 facilities were owned by one operator, what would be the difference between the TD methods?

4. Analyzing disagreement: We present two case studies of intensive contemporaneous ground estimates to analyze disagreements between the methods.

3.3.1 Method Consistency

Table 3.1 provides information about the facilities enrolled, including the operational state and the BU inventory estimate for the baseline and EOP measurement days. This insight is crucial for interpreting the expected difference in emissions between the two field deployments. While the daily BU estimates may not accurately reflect emissions at the facility (see Brown et al. [20]), *changes* in the BU estimate provide an estimate of how emissions likely differed between baseline and EOP based on operational conditions which is independent of either TD method. Most BU estimates in this study utilized prior onsite direct measurements of emissions (stack tests, high-flow sampler, etc.) rather than emission factors, and are therefore site-specific estimates. As a result, facilities may have the same number of compressors operating in both baseline and EOP deployments, but have different compressors running, creating substantially different BU emissions estimates.

We used the change in BU estimates ('Delta BU') between baseline and EOP to improve comparisons: Subtracting Delta BU from the EOP estimates corrects for expected differences in emissions between the two deployments. In the baseline, the per-facility estimate from Solution 2 is the sum of the zones and in the EOP it is the average of all whole-facility estimates. In both baseline and EOP cases, the per-facility mean estimate from Solution 1 is the average of all whole-facility estimates. The difference between the baseline and EOP estimates for each TD method will be referred to as the 'delta difference'. Figure 3.1 displays the delta difference for estimates from both TD methods.

Table 3.1: Facility Information

	Facility Characteristics				Number of Compressor's Operating‖		BU Estimates (kg/h)		
Facility ID	Supply Chain Sector†	Compressor Driver Type‡	Engine Class*	Number of Compressor Units at Facility	Baseline	EOP	Baseline	EOP	Delta BU
A	G&P	Recip	4SLB	5	3	3	93.4	58.3	-35.1
B	G&P	Electric	–	3	2	3	26.2	53.8	27.6
C	T&S	Turbine	–	4	3	4	23	25.2	2.2
D	G&P	Recip	4SRB	11	8	10	59.3	1.12	-58.2
E	G&P	Recip	4SLB	15	13	12	117	209	92
G	T&S	Turbine	–	5	3	3 to 2	18.6	31	12.4
H	G&P	Recip	4SLB	10	8	8 to 9	58.6	44.9	-10.2
I	T&S	Turbine	–	1	1	1	0.85	1.68	0.83
J	T&S	Turbine	–	1	0	0	2.14	1.96	-0.18
K	T&S	Recip	2SLB	8	0	0	6.4	10.7	4.3
L	T&S	Recip	4SLB	5	4 to 3	1	97.3	45.9	-51.4
M	T&S	Recip	2SLB	6	1	1 to 2	21.8	31.9	10.1
N	T&S	Turbine	–	2	2	2	13.8	12	-1.8
O	T&S	Recip	2SLB	8	4	0	62.4	60.7	-1.7

† Supply chain sector of the facility: G&P = Gas Processing, T&S = Transmission and Storage

‡ The type of mover driving the compressor(s) at the facility: *Recip* = reciprocating (piston) engine, *Turbine* = combustion turbine

* For facilities with reciprocating (piston) engines, code indicates the type of engine: *2SLB* = two-stroke, lean-burn, *4SLB* = four-stroke, lean-burn, *4SRB* = four stroke, rich-burn

‖ If there are two numbers that indicates the facility changed from state 1 to state 2 during the measurement period.

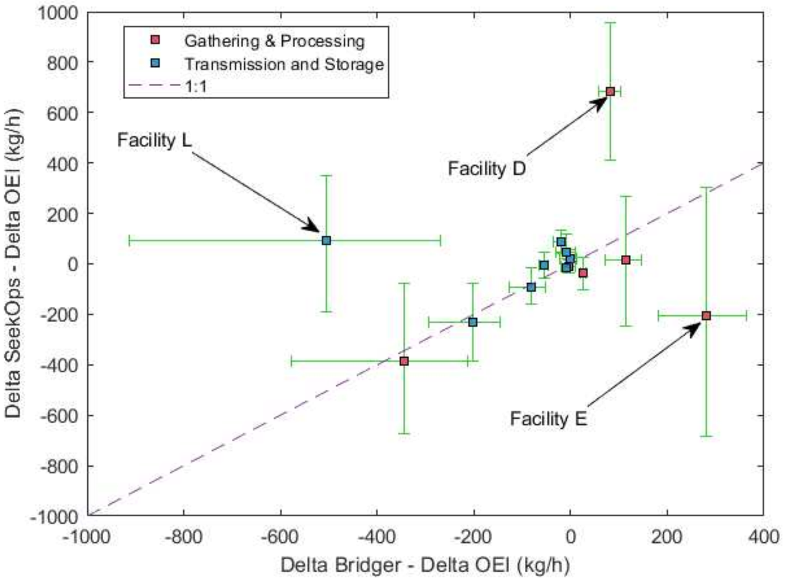


Figure 3.1: The horizontal axis is the difference in Solution 1 estimates from the EOP to the baseline, the vertical axis is the difference in Solution 2 estimates from the EOP to the baseline. Both axes are corrected by the difference in the BU estimate. The red points represent G&P facilities and the blue points represent T&S facilities. Error bars indicate 95% confidence interval for each estimate.

At 11 of the 14 facilities the methods closely followed the parity line (within +/- 20%), indicating both methods exhibited similar delta difference, i.e. both methods displayed similar changes in emissions between baseline-to-EOP. The three outliers, labeled on the Figure 3.1, draw attention to instances where one method exhibited minimal change baseline-to-EOP, while the other experienced a substantial change.

The most significant disagreement in delta difference occurred at facilities D and L:

- Facility D: Solution 2 estimated an increase of 624 kg/h emissions baseline-to-EOP. During the EOP, the onsite observer noted that wind speeds picked up early in the morning. Solution 2 managed to complete two estimates before the winds became too strong, and also shifted from due south to due west between the two estimates. Results changed by 356 kg/h between the two estimates. No noticeable change occurred in the facility operation during this period. These factors suggest that wind conditions may impact TD results.

- At facility L, a T&S facility, fewer compressors were in operation on the EOP than baseline deployment. Solution 1 estimated a delta difference in emissions of -549 kg/h. In contrast, Solution 2's estimates remained consistent between the two days, with a delta difference of less than 50 kg/h. During the baseline measurement survey, Solution 1 detected a substantial emitter with an average emissions rate of 410 kg/h. This same source location was detected during the EOP survey at a mean rate of 200 kg/h. Additionally, another significant emitter was detected during the baseline survey, with an estimated emission rate of approximately 300 kg/h, but it was not detected during the EOP survey. Theses results reflect either marked changes in Solution 1's detection and quantification, or changes in actual emission rates, which did not impact Solution 2's estimates for unknown reasons.

These two cases suggest that delta difference in measured emissions may be due to: (a) changes in facility operations or emission sources, (b) variations in the performance of the TD methods in varying environmental conditions or, (c) inherent differences between the two TD methods regardless of wind conditions. Explanation (a) requires coordinated observation, likely with robust measurement capability, on the facility to identify changes in emission sources or rates. Explanation (b) would require robust quality indicators incorporated into reported results from the TD methods; as part of this QMRV experiment or deployment neither solution provides these indicators. A quality indicator could identify measurements with elevated uncertainty levels due to sensor and/or wind speed not within a specified range. Explanation (c) would highlight a need to improve method performance, including corrections for conditions that are poorly estimated by a solution.

More generally, an operator deploying only one TD method will have no information to determine if a measured facility is one of the 11 (79% of facilities) which showed similar delta differences between methods, or one of the three (21%) that showed dissimilar delta differences between facilities, and therefore would have few clues to differentiate between a decrease in the quality of TD method results, or a major change in facility emissions.

3.3.2 Per-Facility Analysis

During the baseline phase, the methods statistically agreed in 2 of 28 comparisons using the 2-sided ks test, $\alpha = 0.05$. The first analysis compares the mean of all measurements by each method, during the EOP, at each facility. SI Table B.1 presents all statistical test results; highlights are below:

- The methods agree by the 2-sided ks test (distribution shape) and t-test (distribution mean) at one of 14 facilities, and they agree by the Wilcoxon test (distribution median) at three facilities.
- The mean values from the methods overlap in the 95% confidence interval (CI) at 6 of 14 facilities, however, it is important to note that this comparison does not imply equality, see SI Section B.3.

Outside of research projects, it is improbable that any operator would engage a TD method to conduct multiple measurements at a single facility, as observed in this study. Therefore, it is crucial to focus on comparing mean values of individual estimates, each of which would represent a single measurement at a facility. For regulatory or voluntary reporting, operators will likely use only one TD estimate and report one mean value. To explore agreement between estimates, the same statistical tests were performed comparing each individual estimate at a facility from one method to all individual estimates from the other method, resulting in 773 pairwise comparisons. In the pairwise testing, using the 2-sided ks test (α = 0.05) as an example, at half (7) of all facilities, no paired tests showed agreement, at 4 facilities greater than 15% of paired tests showed agreement, and at one facility 40% of paired tests agreed. See SI Table B.2 for the number of comparisons and results for each facility.

Had an operator deployed either method at any given facility, the mean of any one of the estimates from either TD method would represent a valid number to report, under regulatory reporting. To assess the range of reported estimates we simulated 1,000 possible field campaigns. In each simulation, one estimate from each method was randomly selected and the difference in the mean estimates between the two methods was calculated for each iteration. The resulting 95% CI of these differences provides an estimate of how much the reported emissions could vary, based solely upon the choice of measurement vendor ('Reporting Range' in Table 3.2).

Table 3.2: EOP facility estimates and variability between methods

Facility ID	EOP Estimates (kg/h) Solution 1	 Solution 2	Ratio $\frac{\text{Solution 2}}{\text{Solution 1}}$	Reporting Range (kg/h) 95% CI	Variation $\frac{Reporting\ Range}{Mean\ of\ TD\ Estimates}$
A	281 [+6.4%/-5.8%]	357 [+43%/-38%]	127% [79% to 183%]	18 - 109	25% [6.1% to 38%]
B	211 [+7.8%/-6.7%]	226 [+31%/-30%]	107% [74% to 141%]	2 - 91	15% [0.76% to 41%]
C	42 [+20%/-15%]	103 [+26%/-25%]	248% [180% to 325%]	21 - 115	85% [27% to 170%]
D	103 [+15%/-12%]	657 [+42%/-40%]	640% [383% to 938%]	307 - 770	151% [73% to 269%]
E	764 [+7.9%/-6.9%]	1,178 [+31%/-30%]	154% [106% to 202%]	53 - 687	42% [5.5% to 77%]
G	73 [+7.2%/-6.9%]	80 [+38%/-33%]	110% [73% to 151%]	1 - 49	26% [1.1% to 71%]
H	59 [+7.6%/-7.3%]	79 [+34%/-33%]	133% [89% to 180%]	1 - 53	29% [1.3% to 75%]
I	41 [+8.7%/-8%]	28 [+36%/-34%]	69% [45% to 95%]	1 - 23	38% [2.6% to 71%]
J	5 [+22%/-18%]	9 [+31%/-29%]	192% [129% to 265%]	0 - 7	62% [3.4% to 115%]
K	17 [+12%/-10%]	102 [+30%/-28%]	590% [410% to 793%]	55 - 106	145% [81% to 213%]
L	512 [+17%/-12%]	496 [+31%/-30%]	97% [66% to 133%]	2 - 203	15% [0.42% to 38%]
M	63 [+6.4%/-6%]	148 [+43%/-39%]	235% [144% to 339%]	43 - 117	82% [43% to 131%]
N	16 [+12%/-10%]	130 [+32%/-32%]	795% [527% to 1093%]	104 - 120	159% [119% to 221%]
O	66 [+4.4%/-4.1%]	136 [+28%/-30%]	207% [147% to 265%]	42 - 106	70% [42% to 109%]

The variability in the reporting range at individual facilities spans a wide range, from as low as 0 to 7 kg/h up to as high as 53 - 687 kg/h. This range is substantial, particularly considering that a single mean value from either TD method has significant implications for methane fee calculations under future U.S. regulations [17]. Additionally, should one of these facilities be included in a supply route calculation for life cycle or supply chain assessments, changes in reported emissions could have a material impact on the resulting emission intensity, with resultant contractual or financial issues.

Given the observed disagreement between the methods, it is also interesting to look at the agreement within one method's estimates. For this analysis, we consider the set of all estimates made by one method at one facility. If this set of estimates agree within the set, the method produces repeatable results with relatively small variations in the mean. When averaged, the average will exhibit reduced uncertainty bounds relative to the individual estimates. Excluding the facilities that underwent operational changes during the measurement period (3 facilities), this is true at 11 of 11 facilities for Solution 1 and 9 of 11 facilities for Solution 2, see SI Table B.3. At the remaining facilities (2 for Solution 2), relative uncertainty of the average exceeds relative uncertainty of the individual estimates. This result is counter to conventional expectations that increased measurements will reduce uncertainty.

3.3.3 Aggregate Method Comparison

The extended measurements conducted during the EOP deployment provided more robust comparisons between the two TD methods. Solution 1 produced two to 31 whole-facility emissions estimates at a single facility, at 12 facilities Solution 1 produced more than 10 whole-facility emissions estimates, while Solution 2 produced two to six whole-facility emissions estimates. In contrast, during the baseline period, Solution 1 produced one or two whole-facility estimates, while Solution 2 produced one.

Figure 3.2 presents a Bland-Altman difference plot, illustrating a comparison between the two TD methods. The MEC represents the average of the mean values from the TD methods and is utilized as the X-axis and to normalize the relative difference on the Y-axis. Each data point corresponds to a mean estimate made at one facility, and error bars denote a 95% empirical confidence interval associated with each estimate.

The concurrent deployment of the two methods at the same facilities assumes that they (a) captured the same emissions profile, and (b) should produce the same estimated emissions.

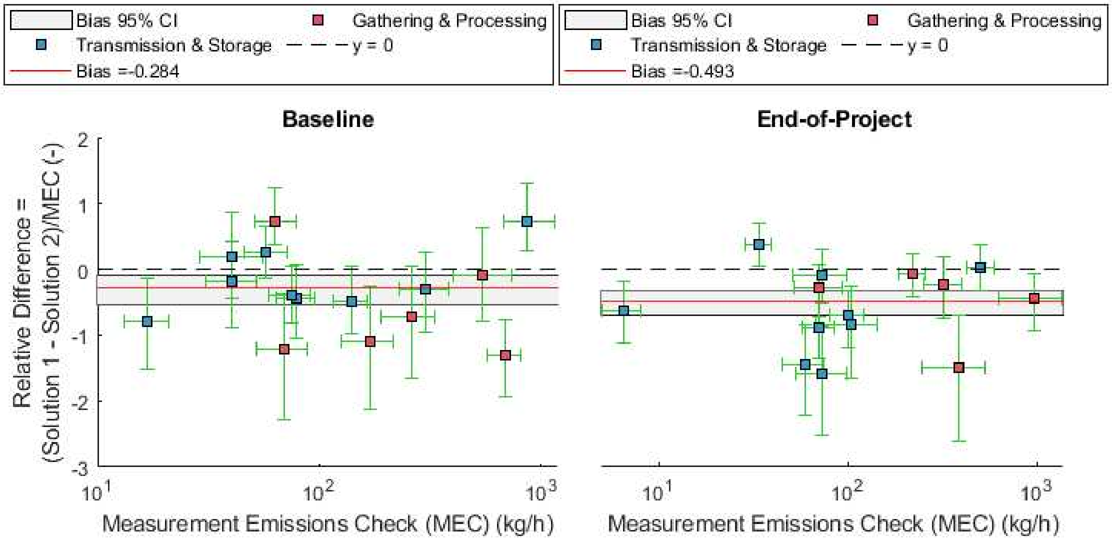


Figure 3.2: Comparison of Solution 1 and Solution 2 estimates. Left plot displays baseline data, right plot EOP data. The horizontal axis is the MEC, an average of the TD estimates. Vertical axis is the relative difference between Solution 1 and Solution 2's estimates for each facility. The gray box displays the 95% confidence interval over all facilities. Dashed line at y = 0 displays perfect agreement.

Solution 2 exhibited a 28% [10% to 43%] higher mean compared to Solution 1 during the baseline deployment, and in the EOP, Solution 2 displayed a mean 49% [32% to 68%] relative to Solution 1. These data indicate that the difference seen in the baseline was not due to a limited number of samples. Instead, the increased number of measurements from both TD methods demonstrates that with repeat measurements there is a larger disagreement between methods.

The aggregate of all 14 facilities is indicative of what an operator would report, had this been one operator's asset base. In aggregate, the discrepancy between the two methods increased from the baseline – 958 [368 to 1,455] kg/h over 15 facilities, or 64 [25 to 97] kg/h per facility, to the EOP – 1,475 [947 to 1,977] kg/h over 14 facilities, or 105 [68 to 141] kg/h per facility. In both project deployments, in aggregate Solution 1 consistently estimated lower emissions than Solution 2. During the EOP, Solution 2's mean estimate was 2X or greater than Solution 1's mean estimate at 6 of 14 facilities.

Given the magnitude of disagreement comparing pairs of single estimates, means of multiple estimates at a facility, and aggregate mean estimates across multiple facilities, one or both methods cannot be providing accurate estimates of emissions at these midstream facilities. While this study focused on two methods, few similar comparisons have been completed with other methods in the field or in controlled test conditions. Therefore, this type of unpredictable inaccuracy in either single estimates or aggregate results may occur with other solutions, in other facility types, and/or in other environmental conditions.

3.3.4 Identifying Sources of Disagreement

To understand the disagreement between TD methods, the EOP deployments included enhanced on-ground emissions detection and quantification efforts, as suggested in Brown et al.[20].

Two facilities (E & N) were volunteered by their operators for enhanced diagnostics after the baseline deployment and before start of the EOP deployment. At these facilities, intensive BU screening was performed to supplement the TD methods. Facility N is a simple transmission site featuring two turbines (32,000 HP/24 MW) driving centrifugal compressors and minimal other equipment, including inlet scrubbers and yard piping. Facility E is a complex gas processing plant with 15 compressors (total of >40,000 horsepower/35 MW) and processing equipment to upgrade gas to pipeline quality, including dehydrators, heaters, acid gas removal units, and associated flares and tanks. These two facilities offer a unique perspective as they represent opposite ends of the midstream sector complexity. In both cases, the enhanced surveillance measured all known sources, using a ground team performing OGI and per-source measurements of all detected sources using high-flow samplers. Additionally, compressor engines on facility E had stack testing and crankcase vent measurements conducted on the measurement day. These two facilities provide compact case studies of issues when using TD methods to estimate emissions for midstream facilities.

At facility N, the two turbine-driven compressors were operating during measurements. Major expected emissions were compressor centrifugal venting and other fugitive leak emissions (e.g. isolation valve leaks); SI Section B.5.1. Turbine exhaust has minimal combustion slip [26].

On the compressor building, Solution 1 detected emissions in a location suggestive of emissions from the centrifugal compressors' seal vents. The average Solution 1 estimate for the each seal vent was 6 kg/h, in comparison to the BU estimate for each seal vent of 5.15 kg/h. Solution 1 detected emissions in the area of the blowdown stacks and reported an average of 2 kg/h, roughly four times the high-flow sampler estimate of 0.45 kg/h for the two compressor blowdown stacks. Combined, the BU estimated 12 kg/h for total facility emissions, in relatively good agreement with Solution 1's mean emission estimate of 17 [15.2 to 19.2] kg/h, but still outside the CI of the TD method. In contrast, Solution 2 conducted three separate downwind flux measurements at this facility, yielding a mean estimate of 130 [88.7 to 180] kg/h, over seven times the Solution 1 and BU estimates.

We consider two possibilities for the disagreement between the two TD methods.

1. Solution 2 either detected emissions transported from upwind, or encountered difficulties in capturing downwind emissions on multiple occasions. Probability and impact of these issues remains unknown.

2. Both Solution 1 and on-ground teams missed onsite sources totaling 113 to 118 kg/h (mean). Solution 1 would miss sources if they were: (a) exceptionally diffuse and did not form visible plumes; sources of this type were not observed by the ground based monitoring methods or crew. Or, (b) sources were all below Solution 1's MDL (1-3 kg/h). This would require between 40, 3 kg/h sources, or 120, 1 kg/h sources, all of which were not detected by on-ground teams. Both of these scenarios appear improbable.

At facility E, 12 compressors were operating during the EOP deployment. While many emission sources *could* exist, major sources from compressors include compressor driver exhaust, compressor rod packing vents, and crankcase vents. (SI Section B.5.2) The BU estimate utilized extensive onsite measurements rather than emission factors.

Given the large size of the facility, Solution 2 utilized three flux planes due to drone battery limitations. These zones were designated as follows: Zone 1 encompassing six compressors (five operating), Zone 2 comprising nine compressors (seven operating), and Zone 3 housing a flare and stabilizing equipment. SI Figure B.2 represents the layout of facility E with major equipment outlined and Solution 2's flux planes for each zone.

Solution 1 estimated emissions on the 12 operating units; we assume these estimates include emissions from compressor combustion exhaust and nearby compressor rod packing vents and crankcase vents, see SI Table B.4. Solution 1 and the BU estimate's results were partitioned according to the specific zones flown by Solution 2 in order to make zonal comparisons. SI Figure B.3 compares each method's estimate by zones; key results are noted here:

- In zone 1, there was agreement overall between Solution 1 and the BU estimate, with the estimates differing by 12 kg/h. Solution 1 estimated 78 kg/h, while the BU estimate was 66 kg/h. Solution 2's estimate was on average 274 kg/h higher than Solution 1's estimate and 286 kg/h higher than the BU estimate.
- In zone 3, where there were no compressors, Solution 1 and the BU estimates agree to within 2 kg/h of each other. Solution 2 on average was 95 kg/h higher than both Solution 1 and the BU estimates.
- In zone 2, which includes seven operating compressors, a significant discrepancy is evident, where Solution 1's estimate surpasses the BU estimates by more than 450 kg/h. Solution 2 and Solution 1's estimates relatively agree in this zone, with a difference of 50 kg/h between the two methods.

The same supplemental onsite measurements were made in all three zones. The bulk of the BU-Solution 1 disagreement in zone 2 originates with four compressors estimated at 120 kg/h, substantially in excess of stack tests and nearby point-source measurements, which agreed with Solution 1 estimates in zone 1, on similar equipment. No other large sources are in this vicinity. Given this extensive on-ground measurement work, it is unlikely that any persistent emitter or emitters, large enough to account for the BU-Solution 1 disagreement near these compressors would have gone undetected by the BU team making supplemental measurements.

Stack testing and extensive ground measurements occurred over an extended period. Concurrently, Solution 1 flew 18 overpasses and produced consistent estimates with a mean of 764 [+7.9%/-6.9%] kg/h. No large transient emitters were identified by Solution 1 that may have skewed the TD results, and none were noted by the ground team. Therefore, a large, transient, emitter(s) does not explain the disagreement.

The onsite observer noted that zone 2 was in a depression relative to the other two zones. Topology may have impacted wind speeds and emission transport, but there is insufficient data to validate this hypothesis.

In summary, intensive onsite work at Facility E included measurements of all known emission sources with the potential to explain the BU-Solution 1 disagreement on the compressors in zone 2. The same methods on similar equipment agree relatively well in zone 1. Additionally, the TD methods disagree substantially with each other in all three zones. Taken together, this analysis suggests that the discrepancy in measured emissions, particularly in zone 2, is more likely attributable to inaccuracies in the TD method(s) rather than the omission of a emission source by the onsite measurement teams. However, this cannot be conclusively proven.

The above analysis indicates that, for unknown reasons, a whole-facility method may fail, and current quality-control measures are not mature enough to detect the failure. While this study deployed and compared two specific methods, it is highly unlikely that this type of issue is confined to any one solution; any solution could have similar issues under currently unknown conditions. Therefore, data and analysis suggest that further development of TD analysis algorithms is needed.

Additionally, the extensive onsite measurement work performed at these two facilities is not characteristic of normal operating practice or required by any regulatory or voluntary programs, including BU estimations at the other 12 facilities in this study. This type of diagnostic is both expensive and cumbersome to perform. In Brown et al.[20], the authors suggested that intensive onsite measurements were needed to diagnose disagreements between BU and TD estimates of emissions. Experimentally conducting this type of intensive onsite work at two facilities resulted in 3 of 4 BU-TD comparisons (BU-Solution 2 at both facilities, BU-Solution 1 at Facility E) not conclusively diagnosing disagreements. This illustrates the complexity of using measurements to inform inventories.

3.4 Implications

While inventories are evolving to incorporate empirical data it is important that the data being used is defensible. Results from this study confirm our previous finding that the TD estimates disagree. With an increase of sample size by repeated measurements, compared to the baseline, the disagreement between the methods expanded. Across all 15 facilities in the baseline deployment, the methods exhibited a relative difference of 28% [10% to 43%] in the mean, which increased to 49% [32% to 68%] over 14 facilities during the EOP, despite a dramatic increase in estimates by both methods (from 43 total estimates in the baseline to 773 estimates). This becomes particularly significant when considering methane or other GHG fees, as the choice of TD measurement method can lead to substantial variations in the fees assessed.

Increasing the number of TD measurements at a facility presented both methods with more opportunities to capture facility emissions accurately, and also allowed analysis to better control for changes in operations at the facility and/or transient events, than could be done during the baseline analysis [20]. At 6 of the 14 facilities, the mean of the TD method estimates differed by at least 2:1; i.e. one or both methods did not provide accurate estimates for nearly half of these midstream facilities. This result has a direct impact on any voluntary or regulatory reporting program focused on per-facility reporting. For example, EPA's recently proposed 'super-emitter' reporting program and the ‘other large release event’ reporting [18] would rely on approved anonymous surveillance by third parties to detect and report “large release event of at least 250 mt CO_2e per event or have a methane emission rate of 100 kg/h or greater at any point in time" [45] at oil and gas facilities, including midstream facilities such as a compressor station, or natural gas processing plant. Reported emission rates for these large emitters would be used to estimate total emissions, with emissions duration assumed to be a default of 182 days unless the operator can prove otherwise [45]. Over or under estimation of emissions by 10s to 100s of kg/h – difference seen multiple times in this analysis – would result in substantial errors in quantification, and given methane fees of 1500 USD per tonne methane (CH_4), significant financial impact for individual facilities and their operators. Similarly, the European Union (E.U.) is considering methane emissions rules covering operators within the E.U., and also additional reporting requirements on emissions data for importers of natural gas and liquified natural gas (LNG) from the exporting suppliers. Voluntary programs do not address these inter-measurement technology considerations presented in this work.

Given longstanding issues with inventory estimates [12, 41, 49, 9], there is a strong case for utilizing measurements to inform, supplement, or potentially replace inventory estimates. However, to rely on measurements, results of this study indicate methods must improve quantification accuracy, produce more representative uncertainty estimates, and provide robust quality control indicators to identify when estimates are suspect or potentially in error.

Chapter 4

Conclusion

This study represents a comprehensive evaluation of measurement informed inventories, featuring an in-depth examination of two commonly deployed TD estimation methods across a diverse range of midstream facilities.

It is essential to acknowledge certain limitations and challenges associated with this study. Firstly, the sample size, comprising 15 to 14 facilities, although valuable in its diversity - spanning from large processing plants to small transmission facilities, may not be fully representative of the entire midstream sector. Secondly, for a comprehensive view of the entire supply chain, a consistent analysis approach should be employed in all sectors. While this program extended its measurement technologies to the production sector, it's important to note that the analysis conducted here differs from that applied to the production sector data.

A pivotal aspect of this research was the presence of an onsite observer. Their role was instrumental in not only data collection but also in providing critical insights for interpreting real-time conditions during the measurement period. The baseline deployment revealed distinct challenges in deploying snapshot TD estimation methods at midstream facilities. With enough time between the two deployments, adjustments were made to the study to address the issues determined in the baseline.

The results of both deployments in this study emphasize several key findings, the TD methods:

- Do not agree in aggregate or on a per-facility basis: Mean estimates for each facility, aggregated across all facilities, differed by 28% [10% to 43%] during the first deployment and the second deployment nearly 2:1 (49% [32% to 68%]). In terms of individual facility comparisons, estimates from the two TD methods statistically agreed in 12% (97 of 801) of the paired measurements.

- Estimate higher emissions compared to the BU (OEI) estimate: Operator inventories, which included extensions to capture sources beyond regular inventory requirements and to integrate local measurements, estimated significantly lower emissions than the TD estimates for 96% (1535 of 1589) of the paired comparisons.
- Need better characterization of uncertainty specifically for midstream facilities, and a quality indicator for when estimates may be in error.

The primary concern arising from these findings is that, as reporting programs begin to require empirical data, the necessity for this data to be both reliable and accurate becomes eminent. Additionally, there is a need for characterization of uncertainty associated with these data, specifically for midstream facilities. The results presented in this study should inform operators on the importance of employing multiple TD technologies to comprehensively compare and contrast their results, aiding in the determination of which method proves more beneficial for their specific circumstances.

In 2024, as part of evolving climate change regulations, operators within the natural gas sector will be subject to substantial charges of $900 for every metric ton of methane emissions, and this amount is set to increase further to $1,500 by 2026 [17]. Opting for one measurement method over the other when reporting methane emissions could yield a substantial difference of millions of dollars. The financial consequences of this choice are far from insignificant. This study highlights that these methods differ by almost 50%, emphasizing the high-stakes nature of this decision. It's not merely a financial decision but one with implications that resonate throughout the industry and the broader environmental landscape.

Operators find themselves in a challenging position, having to select the most suitable emission estimation method that aligns with their financial interests. Consequently, many operators will be driven by the desire to minimize their methane fees, thereby seeking the method that results in the lowest financial liability. This transition towards measurement-informed inventories introduces a new challenge for technologies in the industry. These technologies need to prove their worth by delivering consistently accurate and reliable results, or they risk being phased out. It's no longer sufficient for these technologies to be novel; they must demonstrate repeatability and reliability to continue playing a pivotal role in the evolving landscape of methane emissions reporting. As 2024 looms on the horizon, the current state of methane emission estimation technologies suggests that they may not be fully prepared to meet the impending challenges. The clock is ticking, and there's a growing sense of urgency.

As we move toward greater reliance on measurements for emission reporting, the path forward requires careful consideration. Reporting programs that mandate empirical data for inventory purposes must acknowledge and address the challenges identified in this study. While measurement informed inventories represent the future, there remains a significant gap in aligning current technologies with these methodologies.

4.1 Recommendations for Future Studies

- **Expanding the Sample Size:** A more extensive and diverse sample of midstream facilities could provide a more comprehensive understanding of emissions patterns within this sector.

- **Comparative Analysis with Production Sector:** Revisiting the production sector data to determine if similar patterns emerge, thus contributing to a broader understanding of emissions across the supply chain.

- **Enhanced Diagnostics at all Facilities:** Implement additional on-ground detection and quantification efforts at all facilities to facilitate comprehensive source-by-source comparisons with contemporaneous TD measurements. This highlights the circumstances under which a method might fail or demonstrate reasonable agreement with another method.

- **Enhancing Uncertainty Characterization:** Developing improved methodologies for characterizing uncertainty specifically tailored to midstream facilities, enhancing the accuracy and reliability of emission estimates. Conduct controlled testing at midstream facilities, considering their complexity with factors like compressors operating intermittently, frequent shutdowns, large buildings, and onsite processing equipment. Prior to testing, a baseline estimate should be established through stack testing and on-ground measurement methods (high-flow sampler). This baseline would provide a comprehensive understanding of the emission profile, allowing the TD methods to better diagnose issues, such as potential double counting and accounting for intricate wind patterns.

- **Operator Decision-Making:** Investigating how operators select between different measurement technologies when calculating methane fees, considering factors beyond mere cost-effectiveness.

These potential areas of research can contribute to a more complete understanding of methane emissions in the midstream sector while addressing the limitations encountered in this study.

When we extrapolate the findings of this study to the entire population of compressor stations in the United States, estimated at approximately 1400 [50], the implications become even more evident. The average range of difference in mean estimates (reporting range), as uncovered in this study, implies that the emissions inventory for these stations could fluctuate significantly, spanning from 2.7 to 5 Gg. This nearly twofold difference in estimates underscores the substantial impact that this research has on a larger scale. These variations in emissions estimates, when extended to a national level, have significant implications for regulatory decisions, policy-making, and environmental management, emphasizing the critical role of accurate methane emissions estimates.

Bibliography

[1] OAR. US EPA, “Climate Change Indicators: Greenhouse Gases.” https://www.epa.gov/climate-indicators/greenhouse-gases, Dec. 2015.

[2] G. Myhre, D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, H. Zhang, B. Aamaas, O. Boucher, S. B. Dalsøren, J. S. Daniel, P. Forster, C. Granier, J. Haigh, Ø. Hodnebrog, J. O. Kaplan, G. Marston, C. J. Nielsen, B. C. O’Neill, G. P. Peters, J. Pongratz, V. Ramaswamy, R. Roth, L. Rotstayn, S. J. Smith, D. Stevenson, J.-P. Vernier, O. Wild, P. Young, D. Jacob, A. R. Ravishankara, and K. Shine, “Anthropogenic and Natural Radiative Forcing (Chapter 8). In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,” 2013.

[3] S. M. Miller, S. C. Wofsy, A. M. Michalak, E. A. Kort, A. E. Andrews, S. C. Biraud, E. J. Dlugokencky, J. Eluszkiewicz, M. L. Fischer, G. Janssens-Maenhout, B. R. Miller, J. B. Miller, S. A. Montzka, T. Nehrkorn, and C. Sweeney, “Anthropogenic emissions of methane in the United States,” *Proceedings of the National Academy of Sciences*, vol. 110, pp. 20018–20022, Dec. 2013.

[4] “Homepage | Global Methane Pledge.” https://www.globalmethanepledge.org/.

[5] J. D. Podesta, “Inflation-Reduction-Act-Guidebook.pdf,” tech. rep.

[6] OAR. US EPA, “Greenhouse Gas Reporting Program (GHGRP).” https://www.epa.gov/ghgreporting, June 2014.

[7] E. (2022), “Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2020..” https://www.epa.gov/ghgemissions/draft-inventory-us-greenhouse-gas-emissions and-sinks-1990-2020., 2022.

[8] OAR. US EPA, "GHG Emissions by Industry – GHGRP vs. U.S. GHG Inventory." https://www.epa.gov/ghgreporting/ghg-emissions-industry-ghgrp-vs-us-ghg-inventory, July 2015.

[9] A. R. Brandt, G. A. Heath, E. A. Kort, F. O'Sullivan, G. Petron, S. M. Jordaan, P. Tans, J. Wilcox, A. M. Gopstein, D. Arent, S. Wofsy, N. J. Brown, R. Bradley, G. D. Stucky, D. Eardley, and R. Harriss, "Methane Leaks from North American Natural Gas Systems," *Science*, vol. 343, pp. 733–735, Feb. 2014.

[10] 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.

[11] J. Rutherford, E. 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 anad natural gas production emissions inventories," 2021.

[12] R. A. Alvarez, D. Zavala-Araiza, D. R. Lyon, D. T. Allen, Z. R. Barkley, A. R. Brandt, K. J. Davis, S. C. Herndon, D. J. Jacob, A. Karion, E. A. Kort, B. K. Lamb, T. Lauvaux, J. D. Maasakkers, A. J. Marchese, M. Omara, S. W. Pacala, J. Peischl, A. L. Robinson, P. B. Shepson, C. Sweeney, A. Townsend-Small, S. C. Wofsy, and S. P. Hamburg, "Assessment of methane emissions from the U.S. oil and gas supply chain," *Science*, p. eaar7204, June 2018.

[13] D. R. Lyon, D. Zavala-Araiza, R. A. Alvarez, R. Harriss, V. Palacios, X. Lan, R. Talbot, T. Lavoie, P. Shepson, T. I. Yacovitch, S. C. Herndon, A. J. Marchese, D. Zimmerle, A. L. Robinson, and S. P. Hamburg, "Constructing a Spatially Resolved Methane Emission Inventory for the Barnett Shale Region," *Environmental Science & Technology*, vol. 49, pp. 8147–8157, July 2015.

[14] T. L. Vaughn, C. S. Bell, C. K. Pickering, S. Schwietzke, G. A. Heath, G. Pétron, D. J. Zimmerle, R. C. Schnell, and D. Nummedal, “Temporal variability largely explains top-down/bottom-up difference in methane emission estimates from a natural gas production region,” *Proceedings of the National Academy of Sciences*, vol. 115, pp. 11712–11717, Nov. 2018.

[15] D. H. Cusworth, R. M. Duren, V. Yadav, A. K. Thorpe, K. Verhulst, S. Sander, F. Hopkins, T. Rafiq, and C. E. Miller, “Synthesis of Methane Observations Across Scales: Strategies for Deploying a Multitiered Observing Network,” *Geophysical Research Letters*, vol. 47, Apr. 2020.

[16] Z. R. Barkley, K. J. Davis, S. Feng, Y. Y. Cui, A. Fried, P. Weibring, D. Richter, J. G. Walega, S. M. Miller, M. Eckl, A. Roiger, A. Fiehn, and J. Kostinek, “Analysis of Oil and Gas Ethane and Methane Emissions in the Southcentral and Eastern United States Using Four Seasons of Continuous Aircraft Ethane Measurements,” *Journal of Geophysical Research: Atmospheres*, vol. 126, May 2021.

[17] J. Ramseur, “Inflation Reduction Act Methane Emissions Charge: In Brief.” https://crsreports.congress.gov/product/pdf/R/R47206, Aug. 2022.

[18] “EPA’s Supplemental Proposal to Reduce Pollution from Oil and Natural Gas Operations: Key Changes Since the November 2021 Proposal and Areas Where EPA is Seeking Comment,” Nov. 2022.

[19] “Greenhouse Gas Reporting Rule: Revisions and Confidentiality Determinations for Petroleum and Natural Gas Systems,” 2023.

[20] J. A. Brown, M. R. Harrison, T. Rufael, S. A. Roman-White, G. B. Ross, F. C. George, and D. Zimmerle, “Informing methane emissions inventories using facility aerial measurements at midstream natural gas facilities,” *Environmental Science & Technology*, vol. 0, p. null, Sept. 2023.

[21] U. N. E. Programme, "Oil and Gas Methane Partnership: Reconciliation and Uncertainty in Methane Emissions Estimates for OGMP2.0.,"

[22] U. N. Environment, "International Methane Emissions Observatory (IMEO)." http://www.unep.org/explore-topics/energy/what-we-do/imeo, Wed, 10/11/2017 - 13:01.

[23] "National Climate Task Force." https://www.whitehouse.gov/climate/, Jan. 2021.

[24] J. A. Littlefield, J. Marriott, G. A. Schively, and T. J. Skone, "Synthesis of recent ground-level methane emission measurements from the U.S. natural gas supply chain," 2017.

[25] J. L. Wang, W. S. Daniels, D. M. Hammerling, M. Harrison, K. Burmaster, F. C. George, and A. P. Ravikumar, "Multiscale Methane Measurements at Oil and Gas Facilities Reveal Necessary Frameworks for Improved Emissions Accounting," *Environmental Science & Technology*, vol. 56, pp. 14743–14752, Oct. 2022.

[26] D. J. Zimmerle, L. L. Williams, T. L. Vaughn, C. Quinn, R. Subramanian, G. P. Duggan, B. Willson, J. D. Opsomer, A. J. Marchese, D. M. Martinez, and A. L. Robinson, "Methane Emissions from the Natural Gas Transmission and Storage System in the United States," *Environmental Science & Technology*, vol. 49, pp. 9374–9383, Aug. 2015.

[27] 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.

[28] "eCFR :: 40 CFR Part 98 – Mandatory Greenhouse Gas Reporting." https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-98, Oct. 2009.

[29] National Academies of Sciences, Engineering, and Medicine, *Improving Characterization of Anthropogenic Methane Emissions in the United States*. Washington, D.C.: National Academies Press, July 2018.

[30] H. Welsh and M. Passoff, "Proxy Preview 2023," tech. rep., Jan. 2023.

[31] “Veritas.” https://www.gti.energy/veritas-a-gti-methane-emissions-measurement-and-verification-initiative/, Sept. 2021.

[32] “Oil and Gas Methane Partnership (OGMP) 2.0 Framework.” https://www.ccacoalition.org/en/resources/oil-and-gas-methane-partnership-ogmp-20-framework.

[33] “MIQ.” https://miq.org/.

[34] B. J. Nathan, L. M. Golston, A. S. O’Brien, K. Ross, W. A. Harrison, L. Tao, D. J. Lary, D. R. Johnson, A. N. Covington, N. N. Clark, and M. A. Zondlo, “Near-Field Characterization of Methane Emission Variability from a Compressor Station Using a Model Aircraft,” *Environmental Science & Technology*, vol. 49, pp. 7896–7903, July 2015.

[35] “A closer look at Quantification, Monitoring, Reporting and. . . ” https://www.cheniere.com/our-responsibility/climate/closer-look-qmrv, 2022.

[36] P. Bridger, “Gas Mapping LiDAR for Methane Detection.” https://www.bridgerphotonics.com/gas-mapping-lidar-for-aerial-methane-detection.

[37] SeekOps, “Technology - SeekOps.” https://seekops.com/technology/.

[38] A. Corbett and B. Smith, “A Study of a Miniature TDLAS System Onboard Two Unmanned Aircraft to Independently Quantify Methane Emissions from Oil and Gas Production Assets and Other Industrial Emitters,” *Atmosphere*, vol. 13, p. 804, May 2022.

[39] C. W. Rella, T. R. Tsai, C. G. Botkin, E. R. Crosson, and D. Steele, “Measuring Emissions from Oil and Natural Gas Well Pads Using the Mobile Flux Plane Technique,” *Environmental Science & Technology*, vol. 49, pp. 4742–4748, Apr. 2015.

[40] 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.

[41] A. J. Marchese, T. L. Vaughn, D. J. Zimmerle, D. M. Martinez, L. L. Williams, A. L. Robinson, A. L. Mitchell, R. Subramanian, D. S. Tkacik, J. R. Roscioli, and S. C. Herndon, "Methane Emissions from United States Natural Gas Gathering and Processing," *Environmental Science & Technology*, vol. 49, pp. 10718–10727, Sept. 2015.

[42] G. Heath, E. Warner, D. Steinberg, and A. Brandt, "Estimating U.S. Methane Emissions from the Natural Gas Supply Chain. Approaches, Uncertainties, Current Estimates, and Future Studies," Tech. Rep. NREL/TP–6A50-62820, 1226158, Joint Institute for Strategic Energy Analysis, Aug. 2015.

[43] J. Brown, M. Harrison, T. Rufael, S. Roman-White, G. Ross, F. George, and D. Zimmerle, "Evaluating development of empirical estimates using two top-down methods at midstream natural gas facilities," Oct. 2023.

[44] 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.

[45] "Federal Register / Vol. 88, No. 146 / Tuesday, August 1, 2023 / Proposed Rules," Aug. 2023.

[46] 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.

[47] 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*, July 2023.

[48] S. Stokes, E. Tullos, L. Morris, F. Cardoso-Saldaña, M. Smith, S. Conley, B. Smith, and D. T. Allen, "An aerial field trial of methane detection technologies at oil and gas production sites," tech. rep.

[49] D. H. Cusworth, A. K. Thorpe, A. K. Ayasse, D. Stepp, J. Heckler, G. P. Asner, C. E. Miller, V. Yadav, J. W. Chapman, M. L. Eastwood, R. O. Green, B. Hmiel, D. R. Lyon, and R. M. Duren, "Strong methane point sources contribute a disproportionate fraction of total emissions across multiple basins in the United States," *Proceedings of the National Academy of Sciences*, vol. 119, p. e2202338119, Sept. 2022.

[50] "EIA - Natural Gas Pipeline Network - Transporting Natural Gas in the United States." https://www.eia.gov/naturalgas/archive/analysis_publications/ngpipeline/index.html.

Appendix A

Supplementary Material for Informing methane emissions inventories using facility aerial measurements at midstream natural gas facilities

A.1 QMRV Protocol Overview

A short overview of the QMRV protocol is provided here. QMRV utilizes a MII approach to better estimate emissions from facilities or groups of facilities; where full-facility estimates are combined with conventional inventory methods to estimate facility emissions over extended time periods.

To develop inventory elements, the QMRV protocol outlines estimation methods for an operator to estimate GHG emissions. An emissions calculation guidance document was also provided to the operators and is attached here. Fundamental methods are based upon GHGRP methodology (activity counts × emission factors), supplemented by additional categories of emissions not included in the GHGRP and by allowing equipment-level measurements to be used in place of emission factors. Supplemental emission sources included - more accurate estimates of methane slip in compressor engines, compressor starts, combustion emissions from small sources, flares and glycol dehydrators, waste gas recovery systems, tank and compressor vents in all operating modes. See the attached Emission Calculation Guidance Tool which outlines emission sources comparing the EPA GHGRP methodology and an improved calculation methodology.

A.2 Methods

The purpose the MEC is to develop a 'best estimate' of emissions from a facility by combining all available sources of information. For this study, full-facility estimates are based upon aerial TD estimation methods (see main paper). Since every method has conditions or emissions that cannot be completely sensed, TD estimates are augmented by adding or subtracting other estimates to produce a complete estimate of emissions for the facility. Figure A.1 displays a flow chart of how the estimates were adjusted to be comparable to the OEI.

Each TD method has detection and quantification limits, requiring adjustments to method estimates to completely estimate emissions. For example, emissions may be missed by a method if:

- Nature of the emissions is unsuitable for the methods. For example, exhaust methane (methane slip) may be too dilute for plume detection in some cases.
- An emitter is below the MDL of the method.
- An intermittent emitter may emit too infrequently, or for too short a period, for a method to reliably detect that emitter (e.g. intermittent pneumatic controllers).

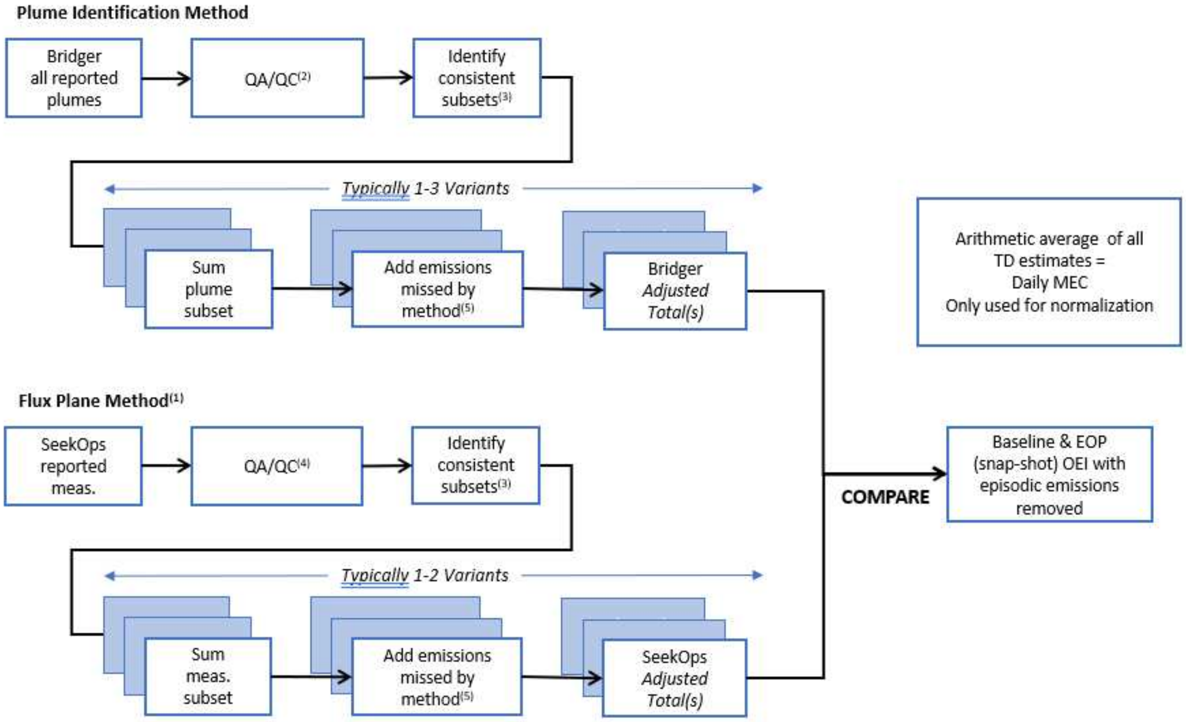
Table A.1 provides an example of adjustments required for one facility. Section A.2.2 and A.2.4 discuss methods specific to the chosen TD estimation methods.

Table A.1: An example of adjustments made to the TD estimates for one facility

		Corrections			
Method Estimate	Facility Estimate (kg/h)	OnSite Dectection (kg/h) †	Compressor Exhaust (kg/h)	Other Exhaust (kg/h) *	Total (kg/h)
Solution 1	553 [456 to 691]	14	130	8	685 [588 to 824]
Solution 2	1290 [963 to 1620]				1290 [963 to 1620]

† 'OnSite Detection' refers to pneumatic emissions

* 'Other Exhaust' was emissions from flares and heaters



Notes:

1. Generally, flux plane methods integrate emissions over larger areas, for a longer times, and therefore require few adjustments to produce an MII
2. QC actions include general data quality checks, including rejecting offsite plumes, assessing, plumes near engine exhaust locations, etc.
3. Subsets control for different overflight times, overlaps between measurements, or similar causes so that each measurement sum is representative.
4. QC actions include identification of overlapped measurements, multiple measurements of same equipment groups, or measurement problems.
5. Measurements added or subtracted from method vary based upon the method and individual measurements. See text.

Figure A.1: Flowchart to make the TD estimates comparable to the daily OEI.

A.2.1 Quality Control Procedures

Field conditions occasionally deviated from the protocol requirements. Additionally, since operations at the facilities were not modified to support measurements, facilities could have changed operating conditions at any time during the measurement period. Specific quality control actions include:

- Changes in operating state. Since emissions at midstream facilities are dominated by compressor or processing equipment operation, changes in the operation of major equipment will change emissions at a facility. Therefore, an on-site observer recorded (or recovered from operational logs) the operating state of the facility during the measurement period. These data were used to identify potential issues with emissions estimation methods:
  - Solution 2 generally subdivides a facility into multiple zones and reports estimates for each zone. While each estimate takes 15-30 minutes, all estimates at a facility take 4-8 hours to complete. If a facility operating state changes between, or during, an estimate, this is noted, and if the change appeared to have significant impact on emissions, the Solution 2 (partial site) estimate may be modified or removed from analysis.
  - Individual Solution 1 plumes were acquired in seconds and represent a snapshot of operating conditions at the time. Therefore, the key quality control action is to identify episodic emissions (e.g. equipment blowdown) that may have occurred during an overflight; these estimates were removed from the Solution 1 total.
- Identify and remove estimates that included emissions outside the facility limits. Note that a facility may be adjacent to equipment owned by another company that is not part of the facility.
  - Since Solution 2 integrates all emissions arriving from upwind, in rare cases where another facility was upwind, a qualitative analysis was conducted with Solution 2 to determine if adjacent facility emissions were properly excluded from Solution 2's partial-facility estimates. Several facilities were adjacent to upwind facilities where emissions

  were sufficiently dispersed to be considered background at the target facility. In one case, emissions from an adjacent equipment yard (metering station) were eliminated by using data from Solution 1 plumes.
  - Since Solution 1 plumes can be located to major equipment units, Solution 1 plumes outside facility boundaries were not included in facility estimates.

- Identify portions of a facility that were not measured. If minor portions of the facility were not measured, the scientific team added estimates for likely missed emissions. If major portions of a facility were not measured, additional measurements were required, typically on a later day.

A.2.2 Solution 1 Estimates

For Solution 1, three specific estimation concerns were addressed: Visibility of combustion slip (methane in combustion exhaust), detectability of gas pneumatic emissions, and adjustments for fugitive emissions below the method's detection limit.

Reciprocating Engine Exhaust: For midstream facilities, methane slip in reciprocating engine exhaust is often the largest source on a facility [27, 41]. It is therefore critical to know if/when a method detects and quantifies exhaust emissions.

Qualitative evaluation early in the study indicated that Solution 1's method may intermittently detect methane slip emissions from reciprocating engines (typically compressor drivers), even though these emissions are often higher than Solution 1's stated MDL. Therefore, the scientific team assessed whether Solution 1 detected exhaust emissions by considering the positioning of exhaust stacks, clarity of the plumes from stacks (i.e. whether the plumes originated at the stacks), and other factors. This analysis was inherently qualitative; for a given engine type and size, Solution 1 may 'see' emissions at one facility but not at another (even though stack tests indicate similar emission rates), based upon weather conditions, ground reflectivity, and/or algorithmic issues.

To distinguish if Solution 1 saw reciprocating exhaust at all facilities with reciprocating engines that were operating on the measurement day - a small number of reviewers classified whether a facility clearly captured exhaust. Solution 1's google earth files were decomposed into individual images with each detection plume. These images along with the name of the facility, date of measurement and units that were operating were presented to the reviewers independently. Each image was analyzed to see if there was a clear exhaust plume based on position, shape and location. The reviewer then classified the plume as 'yes' - exhaust is visible, 'no' - exhaust not visible or 'unclear' - the exhaust is not a clear plume. The final decision on whether or not exhaust was detected was based on if Solution 1 captured the correct operating units and there was a clear plume. The results are presented in the table below.

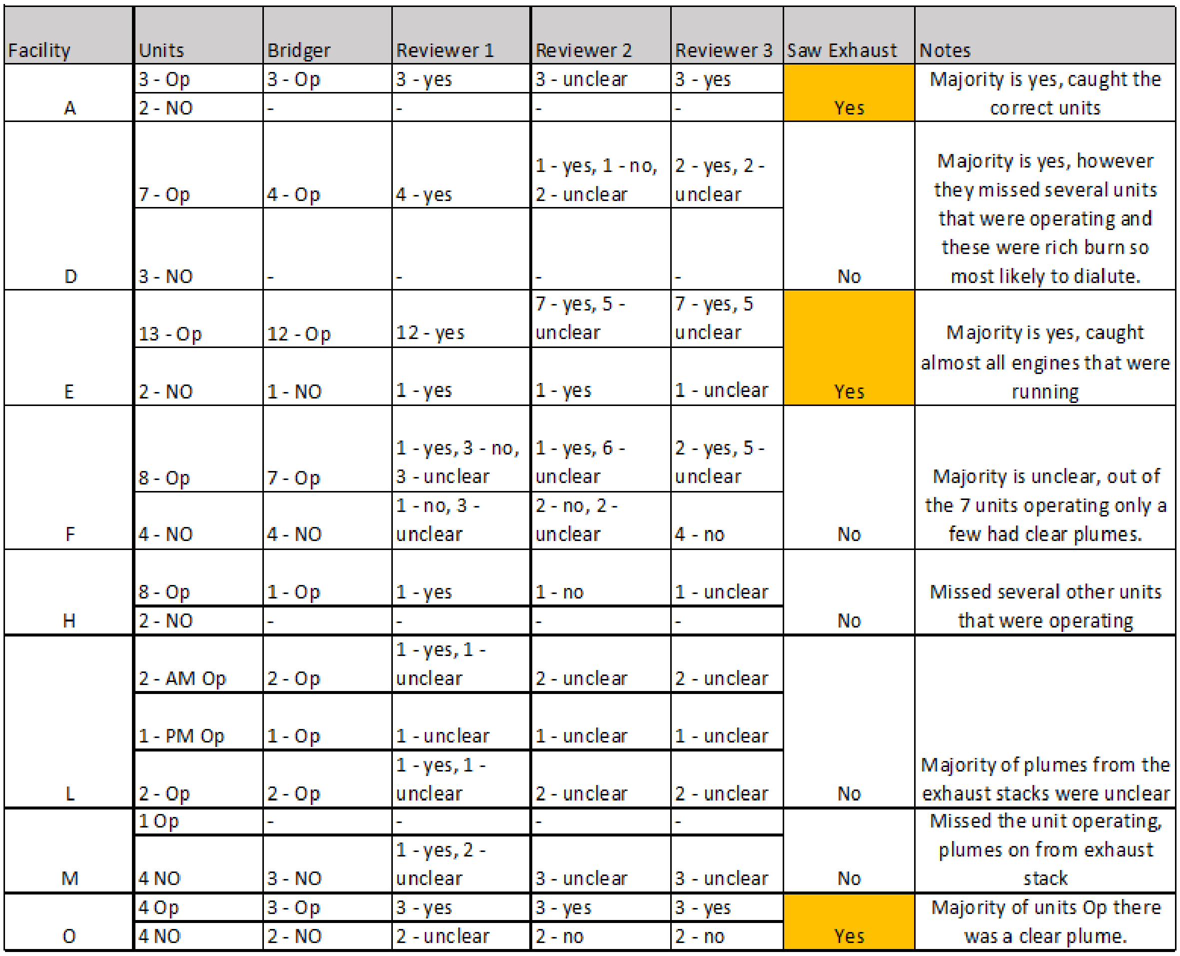
After evaluation, the following rule set was utilized:

- Solution 1 estimates for exhaust emissions were used if three criteria were met: (a) Solution 1 detected an emission plume on all reciprocating exhaust stacks for running units where nominal emissions (from emission factors) are above Solution 1's MDL, *and* (b) these stacks were separated from other potential emitters to avoid confusion, *and* (c) Solution 1 measurements are based upon distinct plumes, i.e. per-stack estimates were not based upon an average estimate for the area, divided between visually identified stacks.

  Figure A.2 provides an example of clearly-identified exhaust stack plumes while Figure A.3 provides an example of an unclear identification.

- In all other cases, exhaust emissions were estimated using stack tests.

Table A.2: Panel review of Solution 1 exhaust at select facilities with reciprocating engines.



Facility	Units	Bridger	Reviewer 1	Reviewer 2	Reviewer 3	Saw Exhaust	Notes
A	3 - Op	3 - Op	3 - yes	3 - unclear	3 - yes	Yes	Majority is yes, caught the correct units
	2 - NO	-	-	-	-		
D	7 - Op	4 - Op	4 - yes	1 - yes, 1 - no, 2 - unclear	2 - yes, 2 - unclear	No	Majority is yes, however they missed several units that were operating and these were rich burn so most likely to dialute.
	3 - NO	-	-	-	-		
E	13 - Op	12 - Op	12 - yes	7 - yes, 5 - unclear	7 - yes, 5 unclear	Yes	Majority is yes, caught almost all engines that were running
	2 - NO	1 - NO	1 - yes	1 - yes	1 - unclear		
F	8 - Op	7 - Op	1 - yes, 3 - no, 3 - unclear	1 - yes, 6 - unclear	2 - yes, 5 - unclear	No	Majority is unclear, out of the 7 units operating only a few had clear plumes.
	4 - NO	4 - NO	1 - no, 3 - unclear	2 - no, 2 - unclear	4 - no		
H	8 - Op	1 - Op	1 - yes	1 - no	1 - unclear	No	Missed several other units that were operating
	2 - NO	-	-	-	-		
L	2 - AM Op	2 - Op	1 - yes, 1 - unclear	2 - unclear	2 - unclear	No	Majority of plumes from the exhaust stacks were unclear
	1 - PM Op	1 - Op	1 - unclear	1 - unclear	1 - unclear		
	2 - Op	2 - Op	1 - yes, 1 - unclear	2 - unclear	2 - unclear		
M	1 Op	-	-	-	-	No	Missed the unit operating, plumes on from exhaust stack
	4 NO	3 - NO	1 - yes, 2 - unclear	3 - unclear	3 - unclear		
O	4 Op	3 - Op	3 - yes	3 - yes	3 - yes	Yes	Majority of units Op there was a clear plume.
	4 NO	2 - NO	2 - unclear	2 - no	2 - no		

Column Units - the number corresponds to the total units that were in either operating mode (Op) or not operating mode (NO). The AM and PM represent a facility operating state change between the two measurements.
Column Bridger - the number is the detected emission plume on exhaust stacks and the corresponding operating mode.

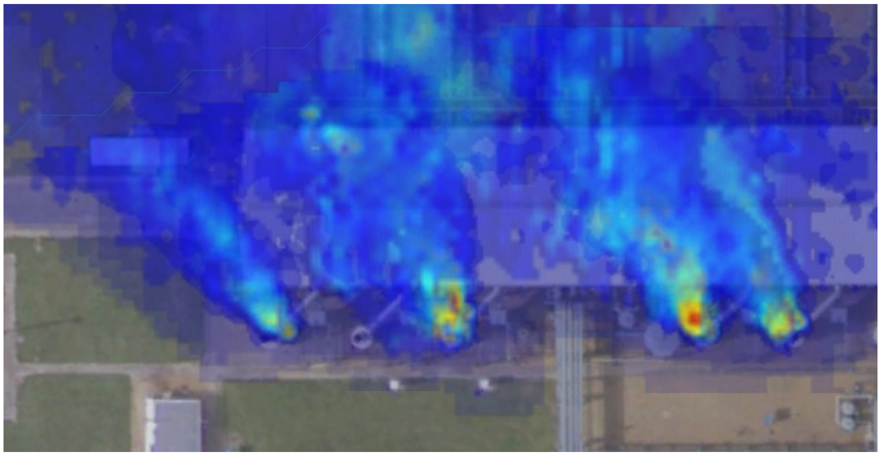


Figure A.2: Solution 1 plume images were provided as part of the survey and overlaid with context camera imagery in Google Earth. This represents a case where clear exhaust plumes were detected. Each exhaust-related plume is located at/from the exhaust stacks, and plumes are detected on all units that were operating at the time of measurement.

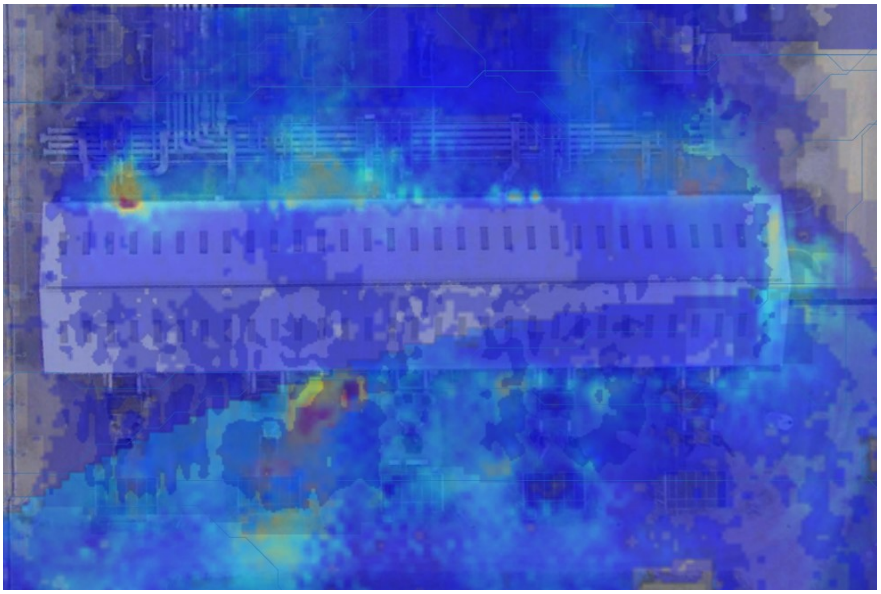


Figure A.3: Solution 1 plume images were provided as part of the survey and overlaid with context camera imagery in Google Earth. This represents a case where exhaust plumes are non-distinct. Identified plumes are not directly associated with exhaust stacks, and it is unclear if plumes were visible from all operating units where plumes should have been visible, given expected exhaust emissions and the method's MDL. The emissions are being obstructed or recirculated by the compressor house.

Turbine Exhaust: For all facilities in this study, combustion slip emissions from turbines were below Solution 1 MDL; stack tests were used.

Other Exhaust: Analyses similar to those outlined for reciprocating engines were used for other combustion sources (e.g. reboiler stacks).

Gas Pneumatics: Solution 1 is unlikely to see emissions from intermittent gas pneumatics; while emitters are technically above Solution 1's MDL (3 kg/h is a common peak emission rate), emissions are too short for a detectable plume to be reliably identified and quantified. Additionally, pneumatics may not activate during an overflight. All continuous bleed pnematics encountered in this study emitted below Solution 1's MDL. Emissions were estimated from OEI device counts and emission factors from either OEI or GHGRP (scientific team choice), and added to Solution 1 totals.

Component Fugitives: Component fugitives below Solution 1's MDL were added to Solution 1 measurements to compute an *Adjusted* Total. If a large component leak was identified and measured by the ground team, and Solution 1 clearly detected the same large emitter, that fugitive detection was eliminated from the set of detection's and the Solution 1 measurement was retained.

A.2.3 Solution 1 Uncertainty

Bell et al. [40] performed controlled testing with Solution 1 to assess the quantification accuracy and detection limits. The dataset included 650 individual measurement passes - the controlled release rates, Solution 1's estimate and the flow error were used in this uncertainty model. The controlled release rate was binned (0 to 10 kg/h and 10 - 2500 kg/h) to create two relative uncertainty models, since relative uncertainty was substantially higher below 10 kg/h.

Note that the controlled test data was inverted from the controlled test environment. During the controlled test, a known emission rate was compared to an unknown estimate from Solution 1. In this study, relative errors from controlled testing were utilized to estimate possible actual emission rates from each Solution 1 estimate. Relative error models were applied to each Solution 1 estimate and Monte Carlo (MC) methods were used to propagate uncertainty through calculations.

An example of adjustments at one facility are in Table A.1.

A.2.4 Solution 2

For Solution 2 two estimation concerns were addressed.

Repeat measurements: On occasion, Solution 2 made multiple estimates of one area for operational reasons, such as access issues, change in wind direction, to improve measurements, etc. If Solution 2 indicated a duplicate estimate was not preferred or fails quality control checks, that estimate was discarded. In several cases, Solution 2 did one downwind flux due to difficulties in dividing the equipment subsets. When this occurred any subset estimates were discarded in favor of the full-facility estimate.

Exhaust plume height: On at least one facility, Solution 2 did not fly high enough to capture emissions from the exhaust plumes of the compressor drivers. In that case exhaust estimates from stack tests were utilized for this source.

A.2.5 Solution 2 Uncertainty

One relative uncertainty model was developed from data in Corbett et al. [38], Table 8, which included 12 non-zero controlled release tests. The error model was applied to each subset estimate and uncertainty was propagated through calculations using MC methods. The facility total is the sum of all the subsets.

A.2.6 Variable Operating State

For most facilities, Solution 1 made two separate passes, separated by several hours, or possibly different days. Each of these covered the entire facility. Therefore, often there are two estimates of emissions, taken at different times, and possibly with different facility operating conditions. Flux plane techniques generally take 10s of minutes to complete, and naturally integrate emissions from all variable sources.

If the operating state of the facility changed more than once and there is no similar state when Solution 1 and Solution 2 measured, then a weighted average was used for the comparison. The weighting is based on the operating time of each unit during the measurement period.

Three representative examples:

1. Solution 1 completed an overflight at 10:00 and 18:00. Solution 2 completed measurements between 8:00 and 14:00. One of four compressors ceased operation at 15:00. In this case:

   OEI is adjusted to produce a weighted estimate of emissions during the measurement period (8:00 - 18:00). As an example, the compressor driver exhaust hourly emissions (in kg/hr) in the OEI represent the rate when the compressor driver was operating during the 10-hour measurement period, rather than an average of the total exhaust emissions over the entire baseline calendar day (24 hours).

   Table A.3: Weighted OEI estimate.

Engine	Operating hours	Not Operating hours	Methane (kg/h)	*Weighted*Methane (kg/h)
Unit 1	5	5	15	7.5
Unit 2	8	2	18	14.4
Unit 3	10	0	12	12

   The *weighted* methane value equals the methane * (Operating hours / hours in period = 10 hours)

   Solution 1's two estimates are treated as independent. Emissions that Solution 1 missed that may change with variable operating states (compressor driver exhaust) are weighted to represent the measurement period.

   Solution 2's estimate is assumed to be comparable to the weighted OEI and other estimates. This represents the 'best estimate' of emissions from Solution 2 during the measurement period, but since individual subsets of the facility were not measured in every operating state of the equipment, the weighting created by summing Solution 2's subset estimates may differ from the weighting applied to the OEI.

2. Solution 1 completed an overflight at 8:00 and 14:00. Solution 2 completed measurements between 10:00 and 14:00. One of four compressors ceased operation at 9:00. In this case:

   MEC averages one Solution 1 estimate (14:00) and one Solution 2 estimate - these estimates are comparable to the weighted OEI. The Solution 1 estimate at 8:00 is not comparable to Solution 2.

3. The same example as above would be valid if one or more of the measurements happened on another day.

A.2.7 MEC

For the purposes of comparing all estimates, episodic emissions were subtracted (removed) from the OEI for the comparison. Note that episodic emissions are needed for annualized emissions, but not for a single-day comparison between the OEI and TD solutions.

The MEC were the arithmetic average of all independent adjusted totals of the facility's emissions for the measurement period.

Example:

- Solution 1 completed an overflight at 10:00 and another at 16:00, Solution 2 completed one complete measurement of the facility the same day, and the facility operating state was near-constant that day. In this case, three adjusted totals would be averaged to develop the MEC – two from Solution 1, one from Solution 2 – each of which is an independent, full-facility, estimate of emissions that would be compared with the OEI for that day. The resulting MEC is used only for normalization purposes.

A.2.8 MEC Uncertainty

MC methods were used to propagate uncertainty from the measurement methods through calculations to determine the MEC. The CI was computed from the empirical distributions using the 2.5 and 97.5 percentile values. The emission rates below are shown as a range *x [l to u]* where *x* is the mean of the distribution, *l* is the lower CI, and *u* is the upper CI.

A.3 Results and Discussion

A.3.1 Facilities

Exceptions to the methods during the baseline deployment were:

- *Facility A and B:* Solution 2 performed one downwind flux of the whole facility. The study team analyzed all Solution 2 data and decided this estimate was the best estimate of the facility emissions.
- *Facility C:* Solution 1 did only one overflight, in the morning.
- *Facility G:* Solution 1 missed a portion of the site on the initial measurement day. The missing area was overflown at a later date.
- *Facility J:* Solution 1 flew over three times during the measurement period. All three were used in the MEC calculation, but for comparison purposes the first two are used.
- *Facility L:* Solution 1 detected a large blowdown that was logged as a episodic event. This estimate was taken out of Solution 1's facility estimate.
- *Facility M:* Solution 1's morning flight captured the facility in a different operating state than other estimates and was not used for comparison.

With those exceptions, there are two Solution 1 measurements and one Solution 2 measurement for each facility these adjusted estimates were averaged to produced one MEC for the facility.

A.3.2 Solution Comparison

A two sample Kolmogorov-Smirnov test (kstest) was used to determine if the two samples were from the same distribution. The test was run at a 5% and 10% significance level and produced the same results, 2 of 28 comparisons statistically agree. Additionally, the solutions in aggregate do not statistically agree at 5% and 10% significance level.

For comparison purposes two other statistical tests were run - the Mann-Whitney (Wilcoxon) and t test. The Wilcoxon test compares the medians of the two samples while the t test compares the mean. The results from the Wilcoxon test show 2 of 28 TD comparisons statistically agree - these are the same facilities (B and I) determined from the kstest. The t test only determined 1 of 28 TD comparisons to be statistically in agreement - facility B.

The aggregate of the solutions is shown below in the last row. The difference in the TD methods (when *Adjusted* Solution 1 1st and 2nd were averaged) is 1,073 [387 to 1,586] or 31% [12% to 47%].

Table A.4: Solution comparison using the two sample Kolmogorov-Smirnov test (kstest).

Adjusted Solution 1 - 1st (kg/h)	*Adjusted* Solution 1 - 2nd (kg/h)	Solution 2 (kg/h)	kstest 1st†	kstest 2nd*
218 [183 to 265]	188 [149 to 247]	376 [192 to 547]	0	0
457 [361 to 628]	601 [380 to 1,074]	578 [294 to 842]	0	1
122 [93 to 166]	N/A	192 [128 to 254]	0	0
80 [60 to 117]	77 [58 to 108]	33 [18 to 45]	0	0
233 [185 to 310]	553 [456 to 691]	1,291 [963 to 1,618]	0	0
104 [97 to 114]	118 [110 to 130]	294 [156 to 425]	0	0
67 [46 to 106]	39 [28 to 56]	47 [36 to 63]	0	0
46 [42 to 53]	41 [41 to 41]	125 [72 to 178]	0	0
55 [36 to 90]	30 [19 to 51]	36 [17 to 55]	0	1
10 [7 to 16]	13 [8 to 22]	27 [18 to 35]	0	0
49 [42 to 57]	87 [71 to 107]	102 [66 to 142]	0	0
856 [584 to 1,511]	1,279 [948 to 1,906]	459 [240 to 654]	0	0
N/A	61 [45 to 83]	90 [66 to 111]	0	0
37 [25 to 59]	39 [27 to 61]	45 [27 to 66]	0	0
246 [185 to 335]	295 [202 to 462]	364 [227 to 495]	0	0
2,580 [2,246 to 3,243]	3,419 [2,932 to 4,135]	4,057 [3,688 to 4,361]	0	2

† The kstest results from comparing Solution 1 - 1st to Solution 2

∗ The kstest results from comparing Solution 1 - 2nd to Solution 2

A 0 from the kstest represents the null hypothesis is accepted at the 5% significance level - meaning the estimates are from two distributions that are the same.

A 1 represents the test rejected the null hypothesis at the 5% significance level - the two estimates are from different underlying distributions.

The last row is the aggregate.

A.3.3 State Simulation

A transmission station pressurizes the natural gas that is then transported further down the supply chain. Figure A.4 displays a simplified schematic of a transmission station. The natural gas enters the station through pipelines into a suction header, is compressed by one or more compressors, and exits through a discharge header connected to the outgoing pipeline. Isolation valves separate the compressor from the suction and discharge headers, and isolate the compressor station from the pipeline. Inlet gas flow is routed to station separators, which remove any liquids from the natural gas, before being routed to the compressors. The gas then enters the compressor unit, each unit equip with: suction and discharge isolation valves, one or more blowdown valve, seal or rod packing vents. When opened, a blowdown valve depressurized a compressor by routing gas inside the compressor to a blowdown vent to the atmosphere. For a unit in operating (OP) - blowdown valve is closed, seal vents are in operating mode. For a unit in not operating pressurized (NOP) - blowdown valve is closed, one isolation valve is closed, seal vents are in NOP. For a unit in not operating depressurized (NOD) - both isolation valves are closed and seal vents are inactive.

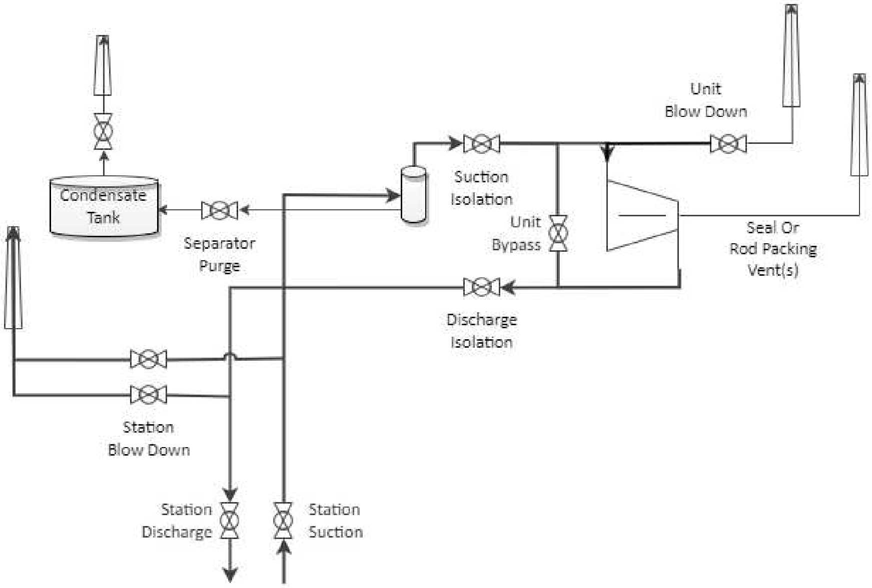
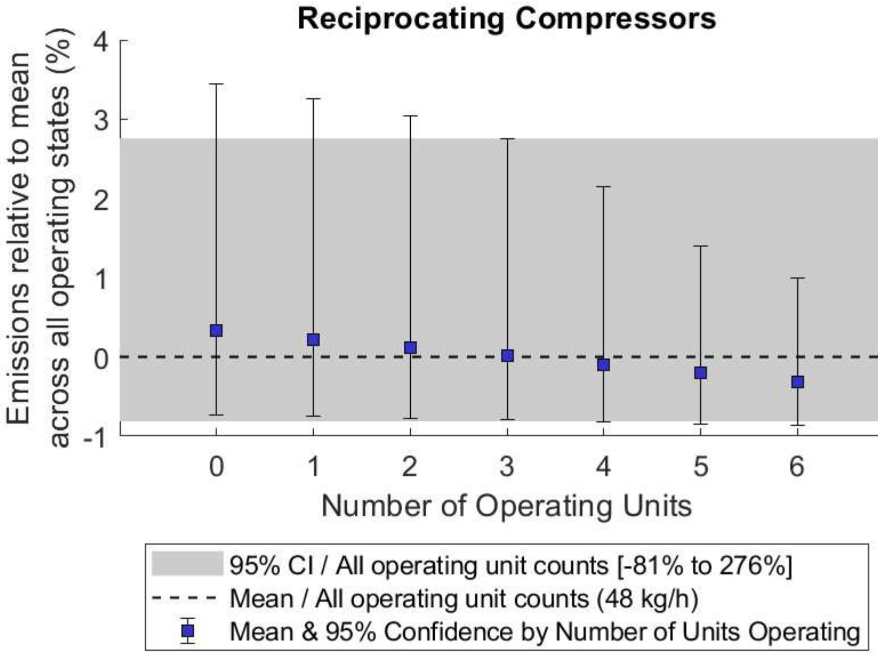


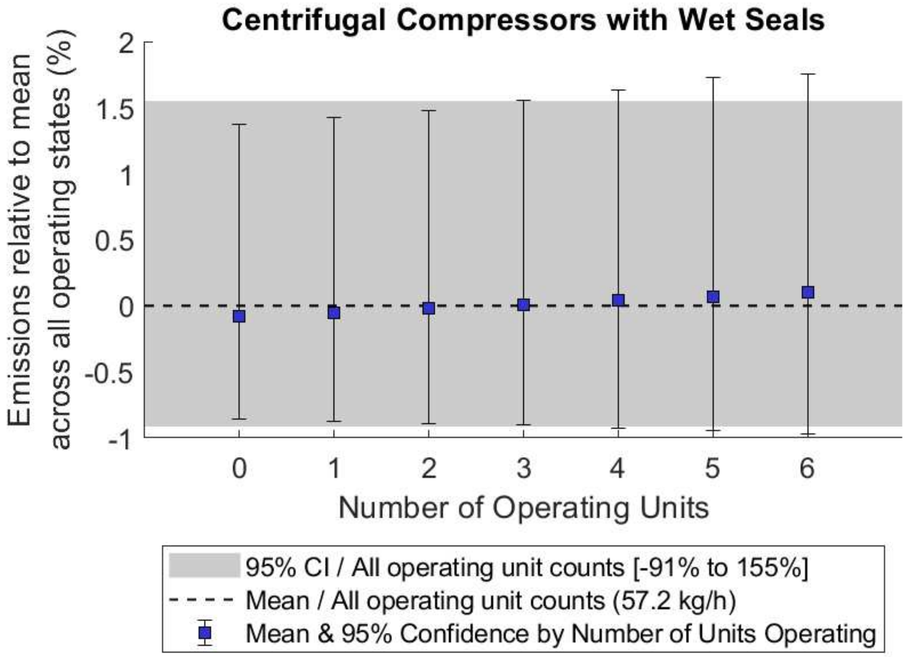
Figure A.4: Transmission station schematic with one compressor unit and major equipment.

A state simulation was run by randomizing units in OP, NOP and NOD. Emission factors from the T&S study [26] were then used to simulate principal compressor emissions in each state. To simulate the different compressor types (reciprocating or centrifugal) the emissions from isolation valves, seal vents and blowdown vents were compiled based on the state of each unit. The total variation is defined as the difference between the upper and lower 95% CI.

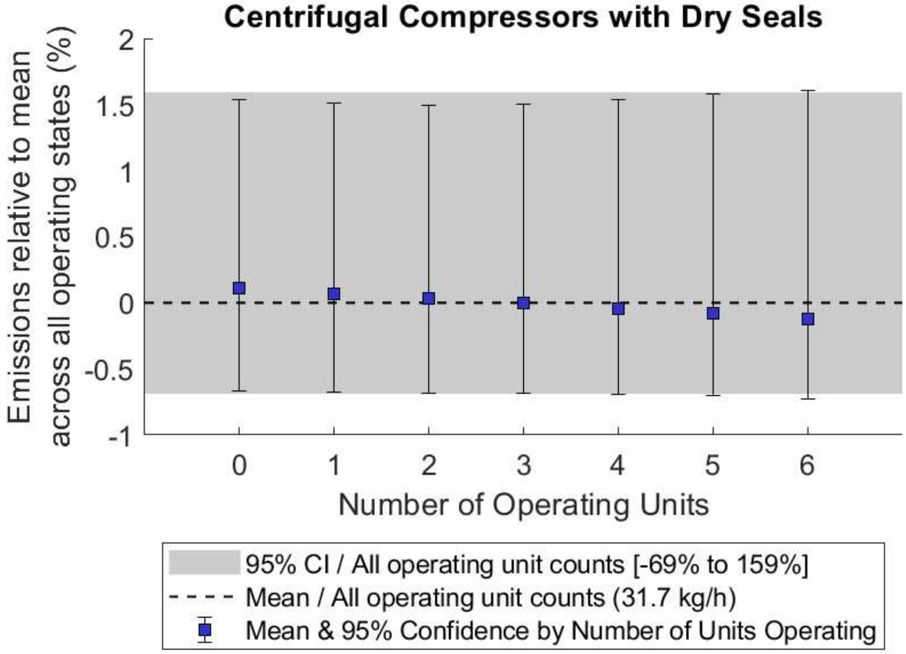


(a) Emissions vary by 357%

Figure A.5: Simulated emissions from reciprocating compressors, in all possible operating states, at a station with six compressor units.



(a) Centrifugal compressors with wet seals. Emissions vary by 246%.



(b) Centrifugal compressors with dry seals. Emissions vary by 228%.

Figure A.6: Simulated emissions for centrifugal compressors in all possible operating states, at a station with six compressor units. Results indicate emissions from units with wet seals (Panel A) and dry seals (Panel B).

Appendix B

Supplementary Material for Evaluating the divergence of two top-down methods at midstream natural gas facilities

B.1 Methods

A typical measurement day included:

- The two TD methods (Solution 1 and Solution 2) were instructed to focus on producing multiple whole-facility estimates, rather than focusing on identification of emission sources from subsets of the facility. These instructions represent an intentional change from the baseline phase [20]. In the baseline phase, TD methods typically performed one or two whole-facility estimates, with substantial measurement time and analysis focused on isolating emitters to subsets of the facility or performing estimates at other facilities. In this study, the TD methods made as many whole-facility estimates as possible throughout the measurement day.

- The facility operator was instructed to:

  - Provide a contemporaneous BU inventory estimate that considers the facility's operational state, employing GHGRP methodologies plus supplemental emissions, including any available direct measurements.

  - Provide a log of when compressor units were operating and/or pressurized.

  - Estimate of the total emissions from each episodic emission event, typically maintenance activities like blowdowns (depressurization of equipment).

- An independent observer from the study team noted compressor states, state changes, and the timing of episodic events, such as blowdowns, compressor starts, etc. The observer also noted any unusual aspects of the environmental conditions, TD methods' flights, etc.

Since neither method can accurately estimate emissions from large episodic events, such as compressor blowdowns (depressurization of a compressor), the study team noted the timing of these events and discarded estimates made during large episodic events, if possible. Corrections at specific facilities are discussed in SI Section B.2.

B.1.1 Solution 1

Data was delivered as a set of 'flight swaths' across the facility. A flight swath represented the area covered by the aircraft during a single pass, with each flyover generating a corresponding swath. To create "whole-site" estimates, study team reviewers examined the flight swath coverage of the facility. The facility was deemed fully covered when all of its premises were encompassed by a set of sequential swaths (SI Figure B.1 provides an example), termed here a 'flyover.' A typical whole-site estimate was usually complete within approximately 10 minutes. During each flyover, Solution 1 detected multiple emission sources, which were summed together to calculate the total facility emission rate. Each whole-site flyover was treated as an independent estimation of the entire facility's emissions.

B.1.2 Method Uncertainty

This analysis uses the relative error in controlled release data from recent peer reviewed studies, [40] and [38] to estimate method uncertainties for Solution 1 and Solution 2 respectively. In both scenarios, the controlled tests were performed under near-ideal conditions and employed a single blind test methodology where the release location was known, but the release rate, including zero releases, were unknown.

For Solution 1 there were sufficient (857) controlled release points to divide controlled releases into two ranges (0 to 10 kg/h and 10 - 2500 kg/h) and develop one relative uncertainty model for

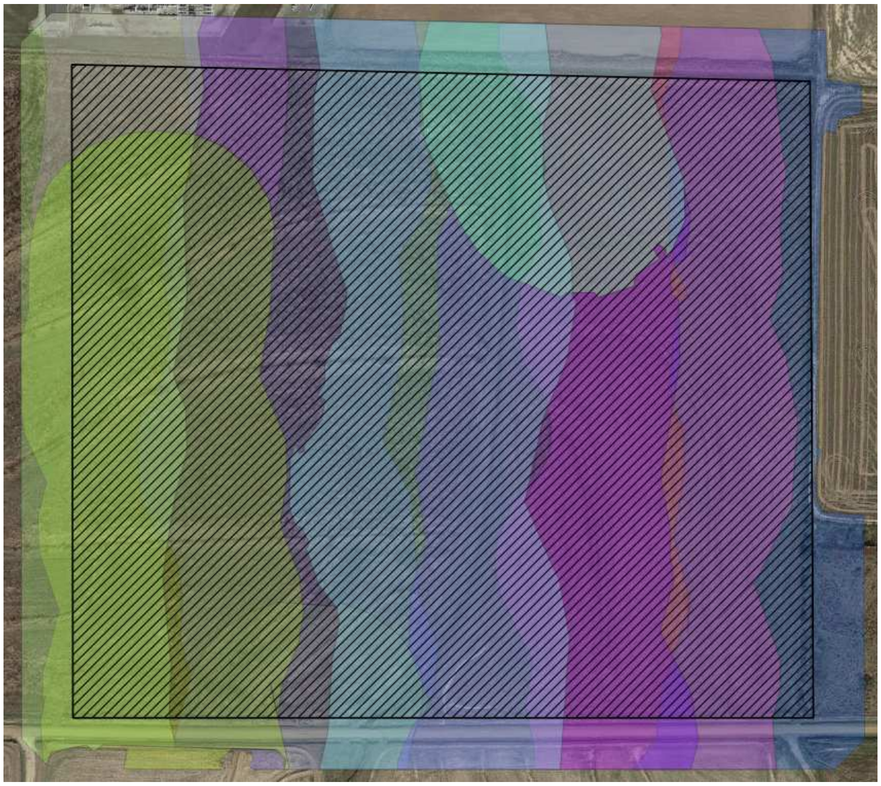


Figure B.1: The facility boundary is the striped area with the black outline. Each swath pass represents a different color, this would represent one whole-site estimate.

each range. Uncertainty is higher for the <10kg/h bin in comparison to the >10kg/h bin. The error model was applied to each Solution 1 plume.

For Solution 2, there were 12 data points in the controlled test, and one relative error model was developed for all observed emission rates. The error model was applied to each Solution 2 observation.

For both methods, test conditions during controlled release testing were substantially different from conditions encountered on midstream facilities: All controlled releases consisted of a single emission point, isolated from other sources, and distant from structures that would disrupt the near-field wind transport. Additionally, weather conditions were constant (Solution 2) to near-constant (Solution 1) during controlled testing. This differs from midstream facilities, which often have complex configurations, including large structures that disrupt or distort the wind field, as well as overlapping and intermittent sources of emissions. Therefore, the uncertainty model used here should be considered a lower bound for uncertainty of each plume or observation.

It's worth emphasizing that uncertainty is considered for each estimate. In the case of Solution 1, uncertainty is applied to each plume, whereas for Solution 2, uncertainty is applied to each whole-site estimate (or each zonal estimate in the case of multiple flights). In Solution 1's methodology, all detection's identified in one whole-site flyover are combined (with associated uncertainty) to calculate the total facility emissions. Subsequently, the mean of all whole-site estimates was computed with Monte Carlo methods, propagating uncertainty through each calculation. Controlled test results showed these aerial methods have some bias in the mean. Therefore, when the uncertainty model is applied to the estimate reported by the solution, the mean value is shifted by that bias.

All CIs were computed from the empirical distributions using the 2.5 and 97.5 percentile values. The emission rates below are shown as a range *x [l to u]* where *x* is the mean of the distribution, *l* is the lower CI, and *u* is the upper CI.

B.1.3 Statistical Approach

Comparisons between methods utilized three statistical tests with $\alpha = 0.05$: the Kolmogorov-Smirnov (ks) test, the t-test, and the Wilcoxon signed-rank test. The two-sided KS test serves as a pivotal tool in gauging whether the two datasets originate from the same underlying distribution or if significant differences are present. Importantly, this test operates without requiring predefined assumptions about the underlying distribution's shape, ensuring that potential variations are not concealed.

The well-known two-sided t-test was employed to compare the means of emissions distributions obtained from Solution 1 and Solution 2. This test assumes that the data follows a normal distribution and facilitates an assessment of whether the observed mean values exhibit statistically significant disparities. Note that neither of the uncertainty models developed from controlled testing were normally distributed. The Wilcoxon sign-rank test was utilized to compare the medians of the emissions datasets. This test is particularly advantageous for small sample sizes and does not rely on specific assumptions about distribution shape.

B.2 Facility Information

Exceptions to facilities during the baseline deployment are outlined in A.3.1.

With those exceptions, there are two Solution 1 measurements and one Solution 2 measurement for each facility. These adjusted estimates were averaged to produced one MEC for the facility. In the initial deployment, there were only three estimates available. In a previous study (Brown et al. [20]), all three estimates were treated as independent, introducing some bias to the Solution 1 results. For the analysis in this paper, we adjusted the methodology. We first averaged the two Solution 1 estimates and then combined this average with the Solution 2 estimate. This adjustment aligns with the EOP methodology, introducing equal bias to both methods.

Exceptions to facilities during the EOP deployment:

- *Facility A:* Solution 1 unable to fly on the EOP day and flew the next day - the compressor configurations was consistent between the two days.
- *Facility A and B:* Solution 2 flew the downwind measurement in two flux plane flights.
- *Facility C:* Solution 1 deployed a helicopter-mounted Light Detection and Ranging (LiDAR) system for EOP estimates rather than the fixed-wing platform used at all other facilities. Comparing the two platforms, the MDL of the helicopter platform is lower (flying lower and slower). However, as there is no controlled release study for the helicopter platform, quantification uncertainty estimates could not be adjusted.
- *Facility D:* A compressor start was noted in the morning, Solution 1's estimates during this time was not included in the analysis. Solution 2 was not flying during this time.
- *Facility E:* Solution 1, Solution 2 measured over two days. The same number of compressors were operating on both days. Solution 2 flew this facility in three different flux plane flights.

- *Facility G:* A compressor unit blew down - Solution 2 stopped measuring during this event (approx. 15 minutes), Solution 1's estimates made during this time were not included in the analysis.
- *Facility I:* The wind direction shifted during Solution 2 last flight causing this estimate to include enrolled and non-enrolled emissions. To correct for non-enrolled emissions, the emissions detected by Solution 1 in the non-enrolled section were subtracted to adjust the estimate.
- *Facility K:* Solution 1 unable to fly on the EOP day and flew the next day - the compressor configurations was consistent between the two days.
- *Facility N:* Solution 2 flew the downwind measurement in two flux plane flights.

Sources below Solution 1's MDL per facility:

- Facility A: Exhaust from heaters and flares
- Facility B: Exhaust from heaters and flares, fugitive leaks
- Facility C: Combustion exhaust from turbines and pneumatic emissions
- Facility D: Combustion exhaust from turbines
- Facility E: Fugitive leaks
- Facility G: Combustion exhaust from turbines
- Facility H: Fugitive leaks and exhaust from reboilers
- Facility I: Fugitive leaks
- Facility L: Fugitive leaks
- Facility M: Fugitive leaks and pneumatic emissions

- Facility N: Fugitive leaks, pneumatic emissions, combustion exhaust from turbines
- Facility O: Fugitive leaks and pneumatic emissions

B.3 Statistical Tests

Tables B.1 and B.2 provide an overview of statistical tests for mean and paired estimates, respectively.

An approach commonly utilized for comparison of estimates is to ask whether the 95% CI of the estimates overlaps. To estimate the strength of this comparison, we use the empirical distribution of estimates to calculate the probability that the distribution with a smaller mean could be larger than the distribution with the larger mean – i.e. the overlap in the distributions. A symmetric (non-skewed) distribution compared to itself would have a score of 50%; this represents the maximum possible score. This metric is included in the columns labeled ‘smaller mean > larger mean’. Scores near 50% represent similarity; lower sores represent less similarity.

Table B.1: Statistical tests comparing mean estimates for the baseline and EOP deployments.

	Statistical Test - Baseline (%)				**Statistical Test - EOP (%)**			
Facility ID	Avg ks-test True	Avg t-test True	Avg Wilcoxon test True	smaller mean > larger mean	Avg ks-test True	Avg t-test True	Avg Wilcoxon test True	smaller mean > larger mean
A	0	0	0	6	0	0	0	17
B	1	1	1	43	0	0	1	35
C	0	0	0	4	0	0	0	0
D	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	1
G	0	0	0	29	0	0	1	33
H	0	0	0	0	0	0	0	8
I	0	0	0	30	0	0	0	1
J	0	0	0	0	0	0	0	0
K	0	0	0	6	0	0	0	0
L	0	0	0	0	1	1	1	42
M	0	0	0	6	0	0	0	0
N	0	0	0	32	0	0	0	0
O	0	0	0	20	0	0	0	0

If Solution 1 produced 10 estimates at one facility and Solution 2 produced 5, this would result in a total of 10 * 5 = 50 pairwise comparisons for that specific facility. In total, across all facilities, there were 773 such pairwise comparisons for the EOP.

Table B.2 presents the comparison of individual estimates at each facility. Highlights of comparisons are:

- There is no observable correlation between the number of comparisons and the degree of agreement or disagreement between the methods.
- The highest agreement at one facility, based on the 2-sided ks test (α = 0.05), shows the methods agree in 40% (31 of 78) of comparisons.
- At another facility, based on the 2-sided t-test (α = 0.05) the methods agree in 29% of comparisons (20 of 68). At the same facility, the Wilcoxon test (α = 0.05) exhibits the highest agreement between the methods with 68% of comparisons (46 of 68).
- On average between all 773 pairwise comparisons, the methods display an overlap within the 95% CI 81% of the time. However, this does not imply equality, as the smaller *mean* estimate is only larger than the lager *mean* estimate 15% of the time.

Table B.2: Statistical tests on all EOP individual comparisons

		Statistical Test (%)			
Facility ID	Num. of Comparisons	Avg ks-test True	Avg t-test True	Avg Wilcoxon test True	smaller mean > larger mean (%)
A	24	0	4	33	29
B	68	37	29	68	37
C	12	0	0	0	4
D	44	0	0	0	0
E	72	10	3	11	17
G	45	27	20	49	28
H	78	40	19	42	26
I	69	12	12	14	17
J	64	5	3	3	12
K	40	0	0	0	0
L	24	38	33	58	36
M	52	0	0	0	3
N	57	0	0	0	0
O	124	0	0	0	3

B.4 Baseline vs. EOP

Comparing the mean's of the methods between the two deployments, in aggregate, Solution 1 estimated a reduction of 740 [414 to 1260] kg/h in emissions from baseline to EOP, while Solution 2 estimated a reduction of only 40.4 [-783 to 836] kg/h. On a per-facility basis, both Solution 1 and Solution 2 estimated lower emissions at a total of 8 facilities from baseline to EOP.

Table B.3: Facility estimates and variability between methods

	Baseline Estimates (kg/h)		EOP Estimates (kg/h)			Uncertainty Reduction [*]	
Facility ID	Solution 1	Solution 2	Solution 1	Solution 2	EOP Operating State Change [‡]	Solution 1	Solution 2
A	203 [+19%/-14%]	376 [+46%/-49%]	281 [+6.4%/-5.8%]	357 [+43%/-38%]	0	1	1
B	529 [+48%/-24%]	578 [+46%/-49%]	211 [+7.8%/-6.7%]	226 [+31%/-30%]	0	1	1
C	122 [+36%/-24%]	192 [+32%/-33%]	42 [+20%/-15%]	103 [+26%/-25%]	0	1	1
D	79 [+30%/-19%]	33 [+37%/-44%]	103 [+15%/-12%]	657 [+42%/-40%]	0	1	0
E	393 [+19%/-14%]	1,291 [+25%/-25%]	764 [+7.9%/-6.9%]	1,178 [+31%/-30%]	0	1	0
G	63 [+34%/-22%]	47 [+34%/-24%]	73 [+7.2%/-6.9%]	80 [+38%/-33%]	1	1	0
H	43 [+7.9%/-4.2%]	125 [+42%/-42%]	59 [+7.6%/-7.3%]	79 [+34%/-33%]	1	0	1
I	43 [+45%/-28%]	36 [+53%/-52%]	41 [+8.7%/-8%]	28 [+36%/-34%]	0	1	1
J	14 [+37%/-24%]	27 [+30%/-34%]	5 [+22%/-18%]	9 [+31%/-29%]	0	1	1
K	68 [+16%/-13%]	102 [+39%/-35%]	17 [+12%/-10%]	102 [+30%/-28%]	0	1	1
L	1,068 [+38%/-21%]	459 [+43%/-48%]	512 [+17%/-12%]	496 [+31%/-30%]	0	1	1
M	61 [+36%/-26%]	90 [+23%/-26%]	63 [+6.4%/-6%]	148 [+43%/-39%]	1	1	0
N	38 [+39%/-23%]	45 [+46%/-39%]	16 [+12%/-10%]	130 [+32%/-32%]	0	1	1
O	270 [+34%/-21%]	364 [+36%/-37%]	66 [+4.4%/-4.1%]	136 [+28%/-30%]	0	1	1

[‡] A 1 indicates that the operational state changed during the EOP measurement period, a 0 means the operational state was constant during the EOP measurement period.

[*] A 1 indicates uncertainty decreased from baseline to EOP, a 0 indicates uncertainty increased.

B.5 Case Studies

B.5.1 Facility N

Major equipment at this facility is listed below.

- Two (2) Natural gas fired turbines (16,000 horsepower [hp] each)

B.5.2 Facility E

Major equipment at this facility is listed below.

- Two (2) Amine Treaters controlled by Thermal Oxidizer
- Fifteen (15) Compressor Driver Engines (four-stroke lean burn [4SLB], natural gas-fired)
- Two (2) Electric motor driven regen. gas compressors
- Two (2) Electric motor driven compressors to stabilize overhead gas stream
- Three (3) Generators
- Six (6) Dessicant Dehydrators
- One (1) Process/Emergency Flare
- Four (4) Heaters
- Six (6) Condensate and Water Tanks

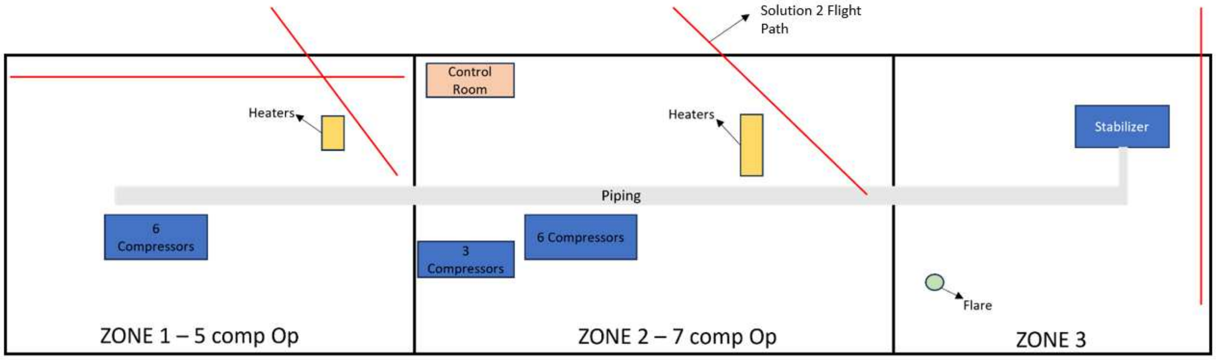


Figure B.2: The black outlined areas represent the zones by which Solution 2 divided the facility and the red lines are the flux plane for each zone. Major equipment is laid out.

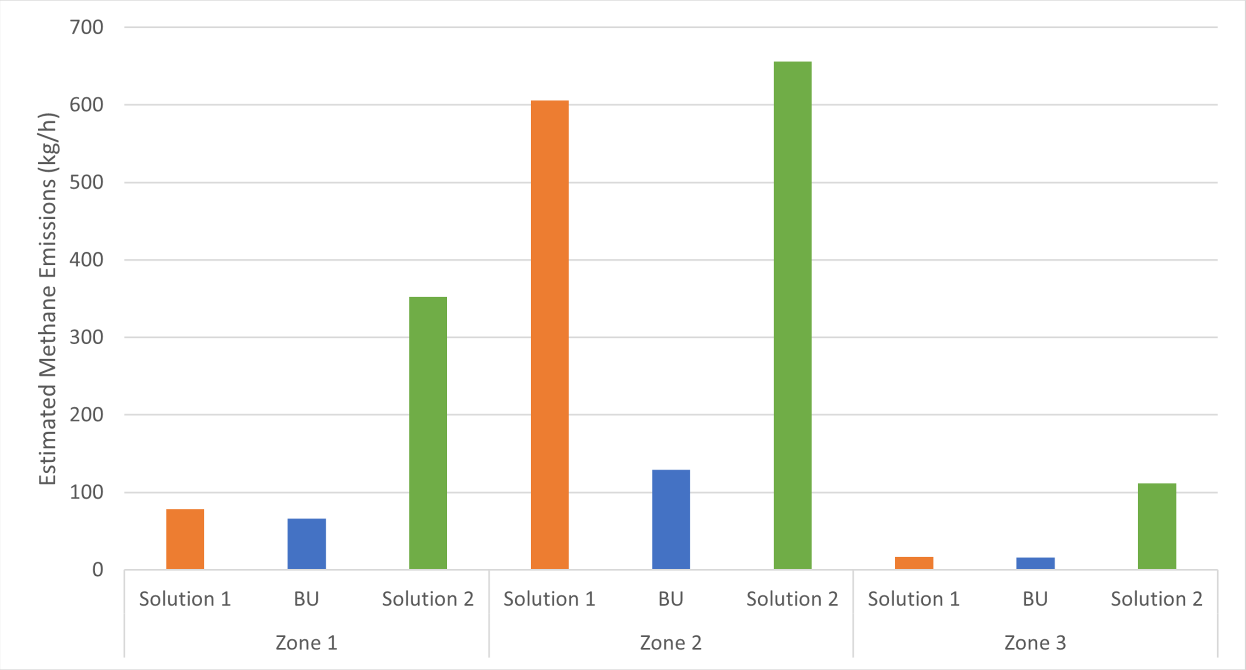


Figure B.3: The Solution 1, BU, and Solution 2 estimate's are compared by zone.

Table B.4: Solution 1 estimates per compressor compared to stack testing, crankcase vent (CCV) and rod packing vent (RPV) estimates.

Compressor ID	Solution 1 Avg. Estimate (kg/h)	Stack Test Avg. Estimate (kg/h)	CCV Estimate (kg/h)	RPV Estimate (kg/h)
C-1	18.0	14.1	4.6	2.8
C-2	18.3	7.3	3.0	2.6
C-3	–	–	–	–
C-4	12.0	6.2	0.8	1.3
C-5	11.7	6.7	1.7	1.6
C-6	12.6	6.7	2.1	1.0
C-7	24.6	23.0	0.8	2.5
C-8	–	–	–	–
C-9	119.5	8.7	3.6	3.4
C-10	119.5	9.4	3.8	0.1
C-11	119.5	15.7	1.5	0.0
C-12	119.5	7.2	2.0	2.0
C-13	9.0	5.3	5.0	13.9
C-14	11.9	6.2	5.8	9.6
C-15	–	–	–	–

C-3, C-8 and C-15 were not operational on the EOP day