DISSERTATION

CONTROLLED TESTING OF NEXT GENERATION LEAK DETECTION AND QUANTIFICATION SOLUTIONS TO EVALUATE PERFORMANCE AND DEVELOP CONSENSUS ASSESSMENT METRICS

Submitted by

Chiemezie Okechukwu Ilonze

Department of Mechanical Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2025

Doctoral Committee:

Advisor: Bret C. Windom
Co-Advisor: Daniel Zimmerle

Daniel B. Olsen
Ezra Levin
Jeffrey Pierce

ABSTRACT

CONTROLLED TESTING OF NEXT GENERATION LEAK DETECTION AND QUANTIFICATION SOLUTIONS TO EVALUATE PERFORMANCE AND DEVELOP CONSENSUS ASSESSMENT METRICS

Reducing methane emissions, a potent short-term climate forcer, is critical for mitigating global warming. The oil and gas (O&G) industry is a major source of anthropogenic methane emissions, and regulations in the U.S. and Canada mandate leak detection and repair (LDAR) programs to mitigate these emissions. Traditional LDAR methods, which includes manually surveying O&G assets with handheld optical gas imaging (OGI) cameras or portable organic vapor analyzers, can be costly and labor-intensive given the vast spatial extent of O&G facilities. However, emerging, next-generation leak detection and quantification (LDAQ) solutions promise a more cost-effective alternative but must demonstrate equal or superior emissions mitigation potential to gain regulatory approval. Standardized controlled testing is essential for verifying this equivalence, yet no widely accepted framework currently exists to achieve this goal. This study evaluates and improves the first known standardized controlled testing protocols designed to address this gap. Two test protocols were developed for the two broad categories of LDAQ solutions: continuous monitors, which operate autonomously over extended durations, and survey solutions, which function over shorter durations with human supervision. These protocols, developed through multi-stakeholder collaboration, were used to test 29 LDAQ solutions (some tested multiple times) at the Methane Emissions Technology Evaluation Center (METEC). METEC is an 8-acre outdoor controlled testing facility that simulates methane emissions from North American onshore O&G equipment. Each survey solution and continuous monitor was tested for a minimum of 3 days and 11 weeks, respectively. Tested controlled release rates were up to 7100 g CH_4/h for continuous monitors and 2100 g CH_4/h for survey solutions. Key performance metrics, including probability of detection (POD), local-

ization accuracy and precision, quantification accuracy, and survey times, were assessed. Seven solutions were retested 3 to 13 months after their initial tests to examine performance changes over time. Results showed that no single LDAQ solution or solution category achieved optimal performance across all the metrics evaluated. For continuous monitors, only two solutions achieved 90% POD within the tested range, failed to detect $\leq$ 40% of the controlled releases, and had $\leq$ 40% of their reported detections classified as false alerts. Camera- and laser-based continuous monitors demonstrated the highest emissions source localization accuracy, with most of them attributing $\geq$49% of their detection reports to the correct emission source. Quantification uncertainty varied widely, with solutions underestimating and overestimating actual emission rates by factors up to 50 and 46, respectively. For survey solutions, handheld OGI cameras exhibited better accuracy and repeatability in detecting and localizing small fugitive emissions compared to mobile (automobile-/drone-based) survey solutions, although the latter completed emission surveys faster. Additionally, performance improvements were observed with repeated testing, emphasizing the likely importance of regular, independent, and comprehensive evaluations in advancing LDAQ solutions. Findings from these controlled tests, combined with stakeholders feedback and insights from parallel field testing, informed the revision of the protocols to better reflect the application of LDAQ solutions at real O&G facilities. Study findings demonstrates that integrating multiple solutions can complement the limitations of any individual or category of LDAQ solutions. Continuous monitors and automobile-/drone-based survey LDAQ solutions can rapidly detect and narrow-down sources of emissions, enabling targeted follow-up investigations with handheld LDAQ survey solutions. In general, this work contributes significantly to efforts aimed at accelerating regulatory approval and adoption of next-generation LDAQ solutions for methane emissions mitigation through transparent and rigorous controlled testing.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to the entire Methane Emissions Technology Evaluation Center team led by Daniel Zimmerle for their continuous support, invaluable guidance, and constructive feedback throughout the course of my research and the preparation of this dissertation. Their expertise and encouragement have been instrumental in shaping the work presented here.

I sincerely appreciate the members of my dissertation committee for their insightful comments, valuable suggestions, and thoughtful critiques, which greatly enhanced the quality of my work. I am also deeply grateful to the United States Department of Energy's Office of Fossil Energy and Carbon Management for providing the funding (award number DE-FE0031873), without which this work would not have been possible.

Special thanks go to my family for their unconditional love, patience, and encouragement. To my parents (Chief Okechukwu and Martina Ilonze) and siblings (Chike, Nnee, Ekene, Ezinne, and Chidimma), your belief in me has been a constant source of motivation and strength. I would like to extend my heartfelt appreciation to my friends and colleagues for their camaraderie, stimulating discussions, and unwavering support during both the challenging and rewarding phases of this journey.

May God bless you all.

DEDICATION

I dedicate this work to every curious mind in a quest for knowledge on how to fight climate change through the mitigation of methane emissions from the oil and natural gas sector.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

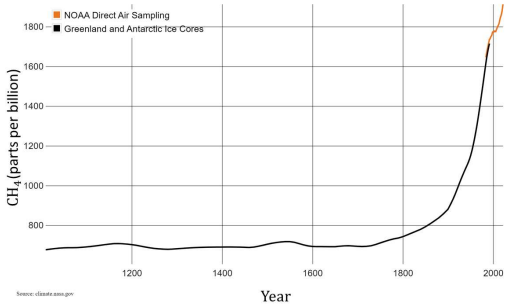
Chapter 1

INTRODUCTION

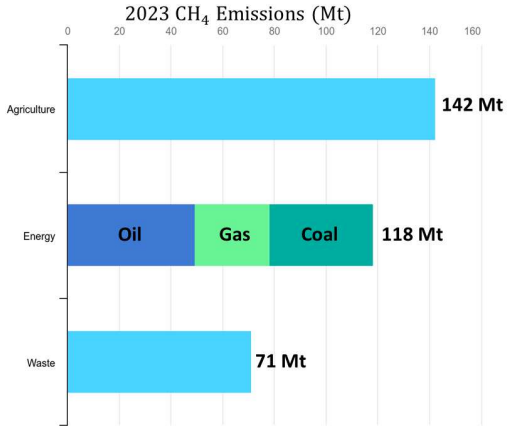
1.1 Methane Emissions and the Oil and Gas Industry.

Methane, a powerful greenhouse gas (GHG) with a short atmospheric lifespan (≈ 12 years), is responsible for about 30% of the rise in global temperatures, with current atmospheric concentration more than twice pre-industrial levels (Figure 1.1a) [1–3]. Methane, with an estimated 20 year global warming potential (GWP) 82 to 87 times that of CO_2, is commonly emitted from the energy sector, landfills, wetlands, and agricultural processes (e.g. domestic livestock) [4]. Globally, the energy sector, including the oil and natural gas (O&G) industry, is the second largest contributor of anthropogenic methane emissions (≈40%) behind agricultural sources (Figure 1.1b) [5]. The O&G industry is the largest industrial source of anthropogenic methane emissions(≈30%) in the united states (Figure 1.1c) and accounts for ≈66% of the methane emissions from the energy sector globally (Figure 1.1b) [3,5]. Reducing methane emissions from the O&G sector (and other sectors as shown in Figure 1.1d) is critical to achieving the Paris Agreement of limiting global mean temperature rise to within 1.5 to $2\,^{\circ}\mathrm{C}$ with significant gas savings, often at "near-zero net cost" due to the captured methane being sold as natural gas [6–8].

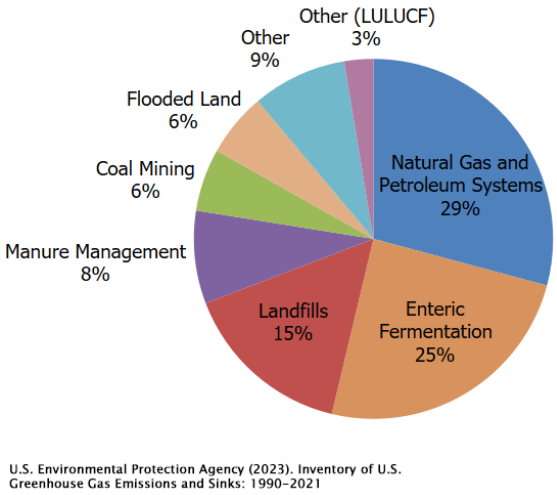
In the O&G industry, methane is emitted across the production and processing [9–22], transmission and storage [23, 24], and distribution [25, 26] sectors of the supply chain. Generally, methane emissions mitigation efforts from these sectors can be applied at five spatial levels as shown in Figure 1.2. Equipment component-level mitigation involves monitoring, identifying, and reducing emissions from components like connectors, valves, thief hatches, etc., found on O&G equipment like a condensate tank, separator, etc. Equipment unit-level mitigation involves monitoring, identifying, and reducing emissions from a single unit of equipment like a wellhead, separator, etc.



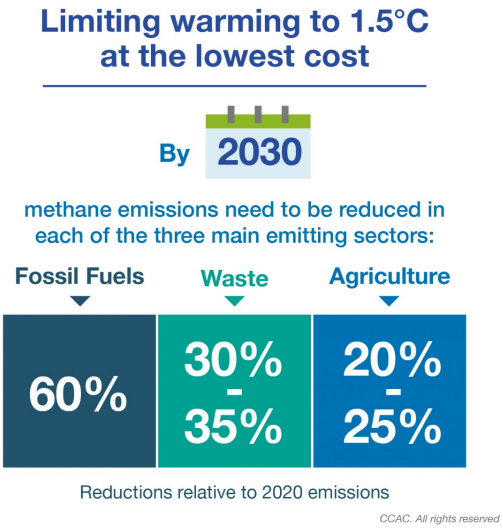
(a) Historic trend in atmospheric methane concentration.



(b) Global methane emissions by sectors



(c) The US methane emissions by source



(d) Methane emissions reduction to achieve climate goals

Figure 1.1: The methane problem: Figure a (Image credit: NASA.) shows the trend of atmospheric methane concentration since pre-industrial times. Figure b (Image credit: IEA.) shows the global methane emissions by sectors in 2023. Figure c (Image credit: US EPA.) shows the U.S. methane emissions contribution by source (1990-2021). Figure d shows desired methane emissions reduction by emitting sectors to limit earth's mean temperature rise to $1.5\,^{\circ}\mathrm{C}$ by 2030 [2, 3, 5, 6].

The equipment group-level mitigation involves inspecting, identifying, and reducing emissions from an aggregation of similar, adjacent equipment units, e.g., a tank battery, etc. Facility-level emissions mitigation involves surveying, identifying, and reducing emissions from any building, structure, or installation of various O&G equipment that is under the same ownership or operation and located on one or more adjacent properties. Basin/regional-level emissions mitigation involves surveying, identifying, and abating emissions from an extensive geological area consisting of several O&G facilities. For example, the Permian Basin in the US is spread between Texas and New Mexico and contains thousands of well pads that are owned by several O&G companies/operators.

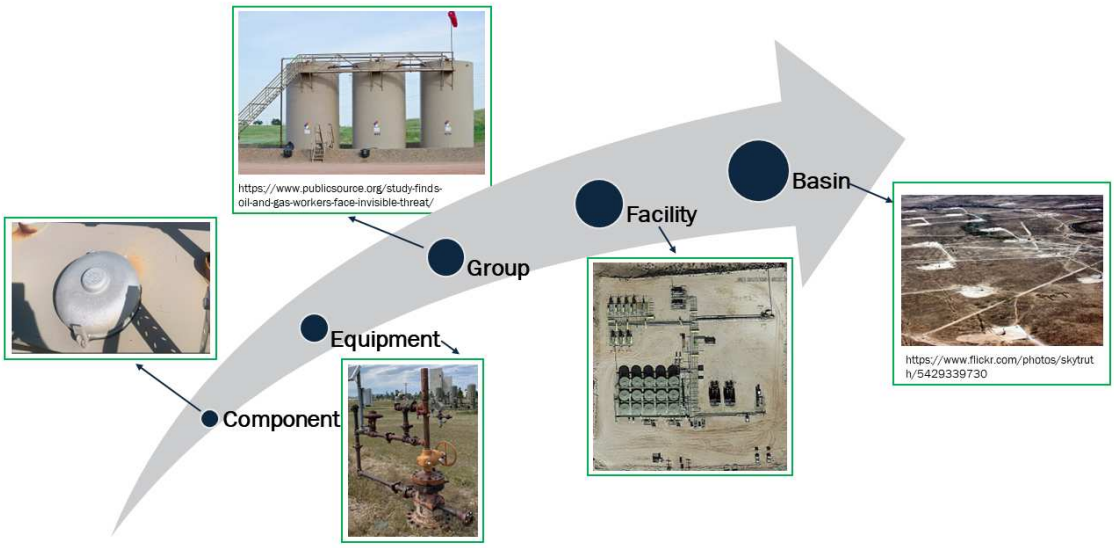


Figure 1.2: The figure shows the 5 tiers of emissions mitigation across the oil and gas industry.

Methane emissions in the O&G sector are generally categorized into fugitive, vent, and combustion slip [27]. Fugitive emissions, or equipment leaks, are unintentional emissions typically caused by process or equipment component failures (e.g., damaged valves, corrosion, loose fittings and connections, etc.). Vent emissions are intentional releases of combustible gases for maintenance activities (e.g., equipment depressurization, liquid unloading of wells, etc.) or the non-combustion use of pressurized gas in process control (e.g., gas-driven pneumatic controllers and

pumps, compressor seals, etc.) [27, 28]. Combustion slip is gas that escape intentional combustion processes partially/fully unburned (e.g., natural gas-fired turbines and engines, flares, etc.) [29]. To meet the world's climate goal, several emissions mitigation efforts, including leak detection and repair (LDAR) have been identified as promising strategies to curb emissions, especially fugitive emissions and combustion slip across the O&G industry [30–32]. Figure 1.3 shows the emissions reduction by LDAR and other strategies under the International Energy Association (IEA)'s net zero emissions (NZE) by 2050.

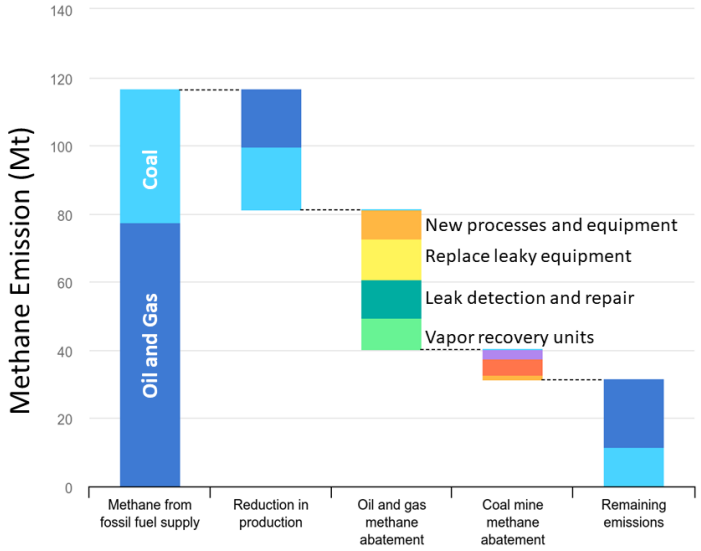


Figure 1.3: The figure shows the methane emissions reduction under the IEA's NZE by 2050 scenario [33].

1.2 Overview of Next Generation Leak Detection and Quantification Solutions and their Applications

LDAR, a work practice designed to identify leaking equipment so that emissions (methane and volatile organic compounds) can be reduced through repairs, is currently implemented in the U.S. and Canada [34–39] with legislative proposals close to adoption by the European Union Parliament [40]. The current regulatory-approved LDAR techniques are handheld OGI cameras and portable organic vapor analyzers, also known as the U.S. Environmental Protection Agency (EPA)

method 21, both of which are used for periodic surveys of O&G facilities [41, 42]. Although these techniques can precisely localize emitting sources, they are labor-intensive, do not estimate emission rates, and have a limited scale of application (component-level for method 21 and up to equipment unit level for OGI cameras). Also, their emissions reduction potential depends on survey frequency, surveyor experience, and scale of application [43, 44]. For example, to be regulatory-compliant in the U.S. and Canada, LDAR must be applied to millions of components spread over a large spatial extent (100s to 1000s of facilities) with variable density of O&G assets, which has significant time, human labor, and cost implications for O&G operators [45]. In addition, several methane emissions measurement studies at O&G facilities in North America have shown that fugitive emissions are temporally and spatially variable [9, 15, 46–50], unpredictable [32, 51], with a small number of facilities and emitters often responsible for a disproportionately large portion of total emissions (skewed and log-normal distribution) [49, 52–56], hence making it more difficult to reduce methane emissions using existing LDAR methods.

Over the last 2 decades, there have been emergence of several *next generation* (i.e., advanced) LDAQ *solutions* deployed on platforms such as vehicles [57,58], aircrafts [27,59], drones/unmanned aerial vehicle (UAV)s [56, 60], satellites [61, 62], stationary towers (i.e. continuous monitors) [63, 64], or handheld [65, 66], which promise to improve LDAR efficiency in a faster, cheaper, and less labor-intensive way compared to existing LDAR techniques. An LDAQ *solution* can be used as emissions monitoring and measurement methods in LDAR programs. An LDAQ solution is an integrated system consisting of an LDAQ technology (methane sensors using one, or a combination of, sensing methodologies), the deployment platform, and data analytics/algorithms used to interpret raw measurement data. According to Aldhafeeri et al. [67], the most common types of methane sensors include optical sensors [68–70], calorimetric gas sensors [71–73], pyroelectric sensors [74–76], semiconducting metal oxide sensors (SMO) [77–79], and electrochemical sensors [80–82]. Optical sensors operate based on emission and absorption spectroscopy [83] including tunable diode laser absorption spectroscopy (TDLAS) [84], light detection and ranging (LIDAR) [85], differential absorption LiDAR (DIAL) [86], cavity ring-down spectroscopy

(CRDS) [87], fourier transform infrared spectroscopy (FTIR) [88], and differential optical absorption spectroscopy (DOAS) [89]. Other sensing methodologies for optical sensors are short-wave ($0.9\,\mu\text{m}$ - $2.5\,\mu\text{m}$), mid-wave ($3.0\,\mu\text{m}$ - $5.0\,\mu\text{m}$) and long-wave ($8.0\,\mu\text{m}$ - $14.0\,\mu\text{m}$) infrared imaging [43,90–92], multispectral and hyperspectral imaging [92–94], and non-dispersive infrared spectroscopy [94]. For calorimetric sensors, the sensing mechanism is based on the effect of heat changes (exothermic or endothermic) associated with gas adsorption, desorption, or chemical reaction on the sensor surface [67]. Catalytic and thermal conductivity gas sensors are the most commonly used calorimetric gas sensors [67, 95]. Pyroelectric sensors, through the pyroelectric effect, convert thermal energy into electrical energy, which is used to identify a target gas by evaluating its thermal diffusivity and conductivity [96, 97]. While SMO sensors identify gas species through the change in electrical resistance due to the reduction-oxidation (redox) reaction between the active sensing layer of the sensor and the gas species [67,98,99], electrochemical sensors identify gas species through the electrical signal produced by the chemical interaction of the target gas and the sensor's receptor [67, 100].

When deployed at O&G facilities, LDAQ solutions can *detect*, *localize*, and/or *quantify* emissions at different spatial scales as shown in Figure 1.2 either through in-situ measurements or remote sensing. Emissions detection is defined as identifying change (enhancements) in the ambient concentration of a target gas specie(s) relative to a predefined baseline. Emission source localization is defined as the spatial resolution/scale at which an emission is detected, while emissions quantification is defined as the estimation of the volumetric/mass flow rate of detected gas species(s). Some of the arbitrarily selected, commercially available next-generation LDAQ solutions with varying localization levels are as follows:

1. The Tropospheric Monitoring Instrument (TROPOMI) is a satellite-based solution launched aboard the Copernicus Sentinel-5P satellite in October 2017 in Russia. TROPOMI maps the global atmosphere every day with a resolution up to 5.5 km $\times$ 3.5 km (7 km $\times$ 7 km at nadir pixel resolution) and a swath of $\approx$ 2600 km [101]. TROPOMI detects and quantifies methane emissions using the absorption data from the oxygen-A band ($0.76\,\mu\text{m}$) and the SWIR spec-

tral range, providing an overall view of emissions at a regional scale as demonstrated by several studies [62, 102, 103].

2. Gas Mapping LiDAR (GML)™ is an aircraft-based solution by Bridger Photonics, Inc. that monitors for methane emissions by beaming continuous-wave lasers (LiDAR) sweeping the ground at a frequency within the absorption spectrum of methane. The solution localizes and quantifies detected emissions using proprietary analytics furnished with geo-registered images of the gas plume generated and other data (meteorological, etc.) [104], and has been used in measurement studies to characterize O&G facility-to-subregional methane emissions [59, 105, 106].

3. SeekIR® is a drone-based solution by SeekOps Inc. that detects methane using methane-specific open-cavity TDLAS [107]. The solution collects a high-density, 3-D point cloud of atmospheric methane concentration to make detection and can also localize and quantify emission sources [108] as demonstrated in field measurement studies [50, 59, 109].

4. MobileGuard™ is a vehicle-based solution by ABB based on ABB's patented off-axis integrated cavity output spectroscopy (OA-ICOS), which is a cavity-enhanced absorption technology and a variation of the TDLAS technology [110]. The solution applies a proprietary algorithm to raw measurement data to detect, localize, and quantify methane emissions while driving. The solution has been used in methane emissions measurement studies, including the study by the European Gas Research Group [111].

5. Canary X is a stationary, point sensor solution by Project Canary that uses near-infrared TDLAS technology to detect emissions when installed downwind of an emission source. To increase the chances of a sensor being downwind of an emission, multiple sensors are installed along the fenceline/boundary of an O&G facility or specific assets for continuous monitoring. The solution uses proprietary analytics to transform raw sensor measurements, meteorological data, information of potential leak sources, and an inverse solver into emission rate estimates and source localization predictions [112].

6. FLIR™ QL320 is a handheld quantitative optical gas imaging (QOGI) solution developed by TELEDYNE FLIR to estimate the rate of methane and other hydrocarbon emissions detected and localized by FLIR's G-series OGI cameras [113]. QOGI analyzes plume pixels from videos of hydrocarbon emissions captured by the OGI camera and quantifies them using proprietary algorithms [114] as shown in methane emissions measurement studies like [115].

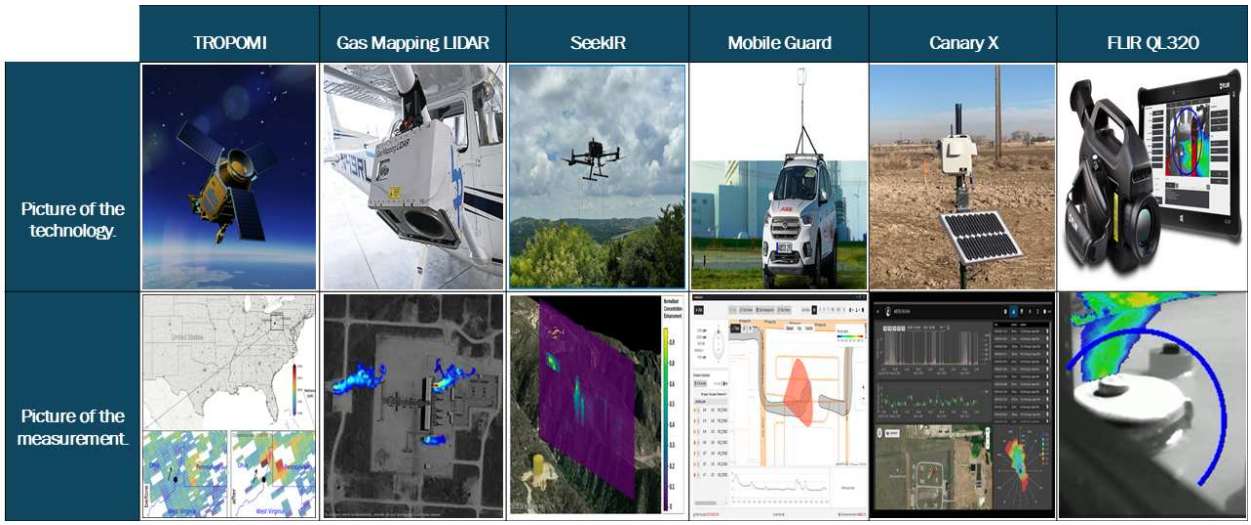


Figure 1.4: The figure shows examples of next-generation LDAQ solutions with different modes of deployment and emission localization levels. The top panel shows pictures of the solutions while the bottom panel illustrates the presentation of the measurement data [101, 104, 108, 110, 112, 113].

1.3 Demonstrating Equivalence and Controlled Testing of Leak Detection and Quantification Solutions

Despite the broad application of next-generation LDAQ solutions for emissions measurement studies and by O&G operators for voluntary and internal emissions mitigation programs [116–118], they still cannot be used for regulatory-approved LDAR programs [34, 38, 39]. To be regulatory approved, LDAR programs based on next-generation LDAQ solutions must demonstrate equivalent or better emissions reductions relative to LDAR programs based on regulatory-approved techniques (i.e., EPA method 21 and OGI camera surveys) [119]. However, demonstrating equivalence requires a standardized, scientifically rigorous, and transparent framework to support such

comparison [119]. A framework for that purpose developed by stakeholders consisting of scientists, regulators, consultants, O&G operators, and non-profit organizations (NGOs) from the U.S. and Canada identified standardized controlled testing as a critical step [119]. A standardized, third-party, single-blind, controlled testing of solutions should develop performance metrics across a range of operational conditions, support the development of reproducible test standards, and contribute to the development of solutions (i.e., the technology, data analytics, and work practice for the technology) with the development of the testing protocol for such controlled testing having consensus support from stakeholders [119].

Several controlled tests of next-generation LDAQ solutions evaluating their performance across several metrics have been conducted. However, they have used study-specific protocols or tested with limited scope, complexity, and environmental/operational conditions [85, 120–127]. Ravikumar et al. [120] assessed and compared the performance of 10 vehicle-, drone-, and aircraft-based LDAQ solutions under a single-blind controlled test condition during the Stanford Environment Defense Fund Mobile Monitoring Challenge using an ad hoc protocol designed solely for the study. Similarly, Johnson et al. [85] and Sherwin et al. [121] used study-specific protocols in the single-blind controlled testing of the performance of aircraft-based solutions (i.e., Bridger Photonics GML™ and Kairos Aerospace hyperspectral imaging). Bell et al. [122] tested 12 next-generation mobile and stationary LDAQ solutions developed by the ARPA-E MONITOR program [128]. Testing was single-blind, but the protocol (an early version of the protocol used in this work) used was not comprehensive, as testing was conducted over very limited emission scenarios and meteorological conditions. Siebenaler et al. [123] tested 4 LDAQ solutions for 75 days during the Methane Detectors Challenge (MDC) in 2014. However, testing was not blinded, and the protocol was study-specific with limited test complexity. Studies by Gardiner et al. [124] and Titchener et al. [125] in collaboration with the U.K. National Physical Laboratory had similar limitations: study-specific protocol, small sample size, and limited testing complexity. Recently, Chen et al. [126] tested 8 *continuous monitoring* LDAQ solutions, but the study used only 1 emission point at any instant and was not single-blind, as participating solutions knew the height and loca-

tions of the mission source. In addition, Liu et al. [127] evaluated the quantification performance of 10 solutions composed of mobile (drone-, automobile-, and aircraft-based), ground-based (fixed camera system), and handheld (OGI camera) technologies at a non-operational compressor station in Spain. The study was partially blind (i.e., only release rate was unknown to participants) and implemented a study-specific test protocol (i.e., timing of controlled releases was constrained to $\approx$2 hours, etc.) resulting in small data sample size. The key point from these studies is that with study-specific protocols, achieving reproducible test standards is elusive, and testing over a short range of environmental conditions and small sample data sizes produces less-reliable and robust performance results.

1.4 Research Motivation

To bridge this gap, the U.S. Department of Energy funded the Advancing Development of Emissions Detection (ADED) project with an objective of advancing the state of LDAQ solutions testing to be scientifically rigorous, affordable, repeatable, and adaptable to field conditions [129]. The controlled testing standards and performance metrics developed will facilitate the characterization and improved understanding of next-generation LDAQ solutions for field deployment. With effective regulation and policy formulation on methane emissions mitigation across the O&G sector relying more on measurement data [130], and the need for rapid and frequent emissions monitoring across a broad coverage of facilities driving interest in next-generation LDAQ solutions, there is a need for a comprehensive understanding of the emissions mitigation performance (e.g., detection, localization, and quantification accuracies, etc.) of solutions. The ADED project broadly categorized next-generation LDAQ solutions into *continuous monitoring* and *survey* solutions to develop the first known standardized and consensus-controlled testing protocols for each category. Continuous monitoring LDAQ solutions (also known as continuous monitors) are autonomous systems (i.e., not continuously supervised) that provide near-continuous monitoring of emissions, providing high-frequency measurement data over extended observation periods (i.e., months to years), and can identify fugitive emissions sooner than other methods. Survey LDAQ solutions

are systems that require human supervision to operate (i.e., handheld, drone-based, automobile-based, and aircraft-based) and can rapidly provide "snapshot" emissions measurements of a vast extent of O&G facilities through periodic surveys lasting minutes to a few hours. A protocol development committee (PDC) of more than 76 members, consisting of environmental non governmental organization (NGO)'s, U.S. federal and state regulators, LDAQ technology developers, scientists, environmental consultants, and O&G operators, peer-reviewed the draft of the test protocols developed by the team from Colorado State University (CSU). The protocols went through at least 1 PDC peer-review cycle, accumulating more than 500 comments to produce the published versions of the protocol [131, 132]. While the test protocol for continuous monitors has not been implemented until this work (dissertation), the protocol for survey solutions has already been used in the controlled testing of some aircraft-based solutions, which is not captured in this work [133–135]. Additionally, with the EPA's final New Source Performance Standards (NSPS) OOOOb rules establishing pathways for the use of next-generation LDAQ solutions in alternative, regulatory-compliant LDAR programs, there is a critical need for robust, consensus controlled testing protocols. Such protocols are essential for benchmarking and comparing vendor performance claims, generating high-quality third-party ground truth data to support technology development, and reliably evaluating LDAQ solution performance to inform regulatory approval decisions [136].

1.5 Dissertation Overview

To date, the literature review and the research motivation, above, have shown that next-generation LDAQ solutions can substantially help mitigate methane emissions from the O&G sector; however, a trusted and reliable framework is needed to demonstrate and characterize the performance of these solutions. This demonstration requires a controlled environment, where all ongoing emissions are known, that is reflective of actual deployments. This dissertation aims to address this concern by answering the following questions:

1. What is the state of the art of next-generation LDAQ solutions? Do these solutions work, and how does performance vary based on the category and type of LDAQ solutions?

2. What is the trend in the development of these solutions with time? How does regular, independent third-party assessment of solutions with standardized testing protocols support this process?

3. How can we improve the protocols to have clear and reproducible test standards with performance metrics and results more representative of expected field application and performance?

To answer these questions, different types of next-generation LDAQ solutions will be selected and tested across a wide range of emissions scenarios, meteorological conditions, and measurement conditions at the CSU METEC facility using the protocols developed by the ADED project to assess the state of the art of LDAQ solutions. This will be followed by retesting solutions (through voluntary participation) that had been tested before, months after their initial tests, at the same facility, using the same protocol to assess the trend in the development and performance of the solutions. Then, using a selected solution as an example, a detailed investigation will be conducted to assess the impact of metrological and measurement conditions on its performance. The findings will provide guidance for the practical field application of the solution. Finally, the limitations from the controlled testings highlighted above and the feedback from stakeholders (i.e., PDC) will be combined with key findings from a parallel field-testing of LDAQ solutions previously involved in the controlled testing, to revise and improve the ADED controlled testing protocols. The dissertation is summarized in the following chapters:

1. Chapter 1 discusses the methane emissions problem, the O&G industry as a significant source of methane emissions, LDAR with next-generation LDAQ solutions as a promising methane emissions mitigation approach, and the need for a consensus/standardized controlled testing protocol for characterizing the performance of LDAQ solutions. The chapter concludes with an overview of the dissertation work and the research questions to be answered.

2. Chapter 2 presents the experimental methodology used for the work which is derived from published manuscripts [137–140]. This consists of a description of the test facility, the test protocols including the testing process and performance metrics, and the selection and deployment guidelines for LDAQ solutions. The chapter also details the data analysis methodologies applied throughout the study, including key statistical techniques such as bootstrapping and Monte Carlo simulation.

3. Chapter 3 is an excerpt of published articles [137, 138] discussing the single-blind controlled testing of sixteen LDAQ continuous monitors using the standardized controlled testing protocol developed for continuous monitors. The discussion includes tested capabilities and test conditions, results based on defined performance metrics, and concludes with the study's implications.

4. Chapter 4 is an excerpt of an article under peer-review [140] discussing the single-blind controlled testing of twelve survey LDAQ solutions using the standardized controlled testing protocol developed for survey solutions. The discussion includes tested capabilities and test conditions, results based on defined performance metrics, study limitations, and concludes with the study's implications.

5. Chapter 5 is an excerpt from a journal publication [139] focusing on the single-blind controlled testing and performance evaluation of an LDAQ solution (QOGI systems) to provide guidance on the field application of the solution.

6. Chapter 6 presents excerpts from published articles [138, 140] that examine changes in the performance of continuous monitors and survey LDAQ solutions retested several months after their initial evaluation. These follow-up tests were conducted using the same standardized test protocol and at the same test facility, enabling a consistent assessment of performance evolution over time. The discussion includes a comparison of results using the same performance metrics and concludes with the implications of the study.

7. Chapter 7 focuses on the protocol's revision to make the testing process and performance metrics/results more reflective of the application of LDAQ solutions at real O&G facilities. The chapter discusses the limitations of existing protocols, the protocol's revision process, and summarizes the revised testing methods and performance metrics and how they are improvements from the existing protocols.

8. Chapter 8 provides a comprehensive conclusion, integrating the findings and discussions from the preceding chapters. It emphasizes the main insights, key takeaways, and contributions made to the existing body of literature and ends by suggesting future research directions.

Chapter 2

EXPERIMENTAL METHODS[1]

2.1 Test Facility

Testing was conducted at METEC (Figure 2.1), an open-air controlled testing and research facility (GPS coordinates: 40.59559, -105.13984) located at CSU, Fort Collins, CO, U.S. that simulates fugitive and vent emissions behavior associated with typical North American production facilities using more than 200 representative emission points (e.g., flanges, connectors, valves, etc.). METEC is furnished with decommissioned surface O&G equipment (e.g., wellheads, separators, condensate tanks, flare stacks, a compressor, and dehydrator unit) embedded with strategically hidden emission points (Figure 2.2), and arranged into five well pads, a dehydrator, and a small compression station. A wellhead is an O&G infrastructure installed at the surface of an O&G well to ensure a secured and controlled flow of hydrocarbons and other fluids from the wellbore to the surface production facilities - Figure 2.2 (a). A separator filters and isolates the components of fluids from wellheads into their component phases - typically oil, gas, and water - Figure 2.2 (b). A condensate tank collects and stores hydrocarbon liquids obtained during the production process - Figure 2.2 (c). The dehydrator removes water vapor from natural gas to prevent pipeline corrosion,

[1]Chapter 2 is extracted from published articles:
[137] Reprinted with permission from C. Bell, C. Ilonze, A. Duggan, and D. Zimmerle, "Performance of Continuous Emission Monitoring Solutions under a Single-Blind Controlled Testing Protocol," *Environmental Science & Technology*, vol. 57, no. 14, pp. 5794–5805, 2023. DOI: 10.1021/acs.est.2c09235.Copyright 2023 American Chemical Society.
[138] Reprinted with permission from C. Ilonze, E. Emerson, A. Duggan, and D. Zimmerle, "Assessing the progress of the performance of continuous monitoring solutions under a single-blind controlled testing protocol," *Environmental Science & Technology*, vol. 58, no. 25, pp. 10941–10955, 2024. DOI: 10.1021/acs.est.3c08511. Copyright 2024 American Chemical Society.
[139] C. Ilonze, J. L. Wang, A. P. Ravikumar, and D. Zimmerle, "Methane quantification performance of the quantitative optical gas imaging (qogi) system using single-blind controlled release assessment," *Sensors*, vol. 24, no. 13, 2024. DOI: 10.3390/s24134044.
[140] C. Ilonze, R. Day, E. Emerson, A. Duggan, R. Brouwer, and D. Zimmerle, "Performance of survey solutions under single-blind controlled testing protocol," *ChemRxiv*, 2025. Preprint URL: 10.26434/chemrxiv-2025-5ff89.

hydrate formation, and inefficiencies in the gas processing systems - Figure 2.2 (d). A compressor station maintains or increases the pressure of natural gas to the pipeline pressure - Figure 2.2 (e).

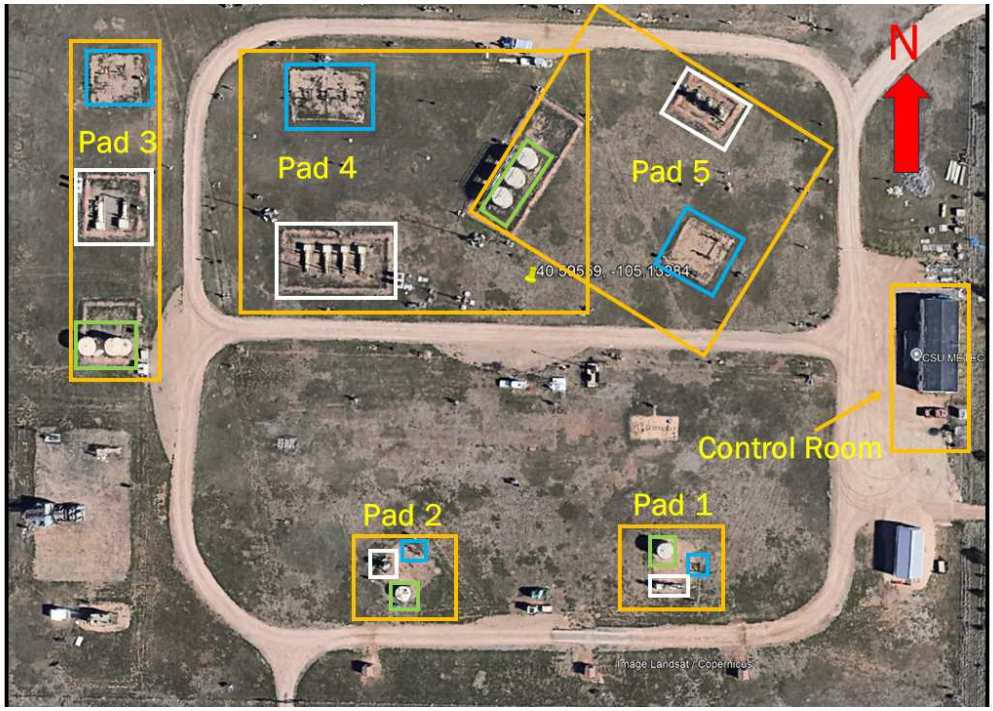


Figure 2.1: The aerial view of METEC shows the layout of the various well pads (the boundary is marked in orange) used for the study. The boxes with green-, white-, and blue-colored boundaries show each well pad's condensate tank (i.e., tank battery), separator, and wellhead equipment group. The dimensions of well pads 1, 2, 3, 4, and 5 are 10 m $\times$ 10 m, 10 m $\times$ 10 m, 10 m $\times$ 60 m, 60 m $\times$ 60 m, and 60 m $\times$ 60 m, respectively.

In general, METEC's design was intended to simulate a wide range of natural wind patterns associated with real upstream production facilities as much as possible by varying the characteristics of the well pads. Each well pad (pads 1 to 5 in Figure 2.1) consists of at least one condensate tank, separator, and wellhead grouped and arranged into clusters with varying layout configurations and orientations, mimicking a wide variety of real O&G production well pads. Each cluster of similar, adjacent equipment units of the same type is an equipment group. Each unit of equipment (i.e., equipment unit) is assigned a unique 3-letter code identifier ("XY-Z"); "X" is *1*, *2*, *3*, *4*, or *5* for the well pad; "Y" is *F* for flares, *S* for separators, *T* for tanks, and *W* for wellheads; followed by "Z", the equipment unit number as summarized in Table 2.1. For example, 4S-1 is the 1st separator in

the separator equipment group on well pad 4. A perimeter roadway ensures mobile LDAQ solutions can access measurement points and simplifies the deployment of fixed LDAQ systems across the facility. The characteristics of each well pad are as follows:

Figure 2.2: Figures (a), (b), (c), (d), and (e) show examples of decommissioned wellheads, separators, condensate tanks, a dehydrator, and a compression station available at METEC. The equipment units were donated by METEC's Industry Advisory Board (IAB) during the construction of the facility in 2016.

1. **Wellpad 1:** Well pad 1, as shown in Figure 2.3(a), is approximately a 10 m $\times$ 10 m mock wet gas O&G upstream production facility consisting of a wellhead, a horizontal separator, and a 150 bbl fiberglass liquid storage tank fitted with strategically hidden emission points. Unlike dry gas, which consists almost entirely of methane, wet gas contains a substantial amount of ethane and other higher hydrocarbons (typically less than 85% methane).

2. **Wellpad 2:** Well pad 2, as shown in Figure 2.3(b), is approximately a 10 m $\times$ 10 m mock dry gas O&G upstream production facility consisting of a wellhead, a vertical separator, and a 70 bbl fiberglass liquid storage tank fitted with strategically hidden emission points.

Table 2.1: Summary of the equipment units in all the well pads used for the study.

Pad	Equipment Type	Equipment Group ID	Equipment Unit ID
1	Separators	1S	1S-1
	Tanks	1T	1T-1
	Wellheads	1W	1W-1
2	Separators	2S	2S-1
	Tanks	2T	2T-1
	Wellheads	2W	2W-1
3	Separators	3S	3S-1, 3S-2
	Tanks	3T	3T-1, 3T-2
	Wellheads	3W	3W-1, 3W-2, 3W-3
4	Separators	4S	4S-1, 4S-2, 4S-3, 4S-4
	Tanks	4T	4T-1, 4T-2, 4T-3
	Wellheads	4W	4W-1, 4W-2, 4W-3, 4W-4, 4W-5
5	Separators	5S	5S-1, 5S-2, 5S-3
	Wellheads	5W	5W-1, 5W-2, 5W-3

Figure 2.3: (a) Well pad 1 is a 10 m $\times$ 10 m wet gas production facility consisting of 1 wellhead, horizontal separator, and liquid storage tank. (b) Well pad 2 is a 10 m $\times$ 10 m dry gas production facility consisting of 1 wellhead, vertical separator, and liquid storage tank.

3. **Wellpad 3:** Well pad 3 measuring approximately 60 m $\times$ 10 m, as shown in Figure 2.4, mimics a wet gas O&G upstream production facility equipped with three adjacent wellheads (10 m $\times$ 10 m), two adjacent horizontal separators (10 m $\times$ 10 m), and two adjacent 300

bbl liquid storage tank (10 m $\times$ 10 m), all fitted with strategically hidden emission points. Earthen barricades support the separators and tanks to replicate the dispersion effects of secondary containment systems typical of real O&G operations. Unlike other well pads at METEC, equipment groups are arranged North to South of the well pad, allowing more complex plume dispersion in that direction for next-generation LDAQ solutions that detect fugitive emissions (i.e., leaks) using downwind transport of emissions.

Figure 2.4: Well pad 3 is a 60 m $\times$ 10 m wet gas production facility consisting of 3 wellheads, 2 horizontal separators, and 2 liquid storage tanks.

4. **Wellpad 4:** Well pad 4, as shown in Figure 2.5, simulates a wet gas production facility of dimension 60 m $\times$ 60 m made up of five adjacent wellheads (10 m $\times$ 15 m), four adjacent horizontal separators (10 m $\times$ 20 m), and three (10 m $\times$ 25 m) 300-barrel liquid storage tanks

with an access catwalk shared with well pad 5. All the equipment is fitted with strategically hidden emission points. This design is intended to replicate the scale, equipment spacing, and operational complexity of real upstream production facilities. As with pad 3, earthen berms surround the separators and tanks to replicate the dispersion effects of secondary containment systems typical of real O&G operations. These berms were characteristic of production facilities when METEC was designed in 2016.

Figure 2.5: Well pad 4 is a 60 m $\times$ 60 m wet gas production facility consisting of 5 wellheads, 4 horizontal separators, and 3 liquid storage tanks.

5. **Wellpad 5:** Well pad 5, as shown in Figure 2.5, is a 60m $\times$ 60m mock dry gas production facility including a 10m $\times$ 10m section with three wellheads, a 10m $\times$ 15m area holding three vertical separators, and a shared 10m $\times$ 25m zone housing the same three 300-barrel

liquid storage tanks from Pad 4. As with other well pads, All the equipment is fitted with strategically hidden emission points. This configuration mirrors well pad 4's structure while introducing equipment diversity reflective of regional production site variations. As with well pads 3 and 4, raised soil ridges encircle the separators and tanks to simulate dispersion patterns caused by secondary containment systems common in actual oil and gas facilities.

Figure 2.6: METEC pad 5 is a 60 m $\times$ 60 m production well pad consisting of 3 wellheads, 3 vertical separators, and 3 liquid storage tanks.

Figure 2.7 shows the piping and instrumentation diagram (P&ID) illustrating the control, metering, and transportation of gas from the onsite gas storage cylinders to emission points for a well pad at METEC. The P&ID is segmented into five sections, which include (1) the gas storage system, (2) the gas flow metering system, (3) the control and supply system from the gas supply house (GSH) to the control boxes at the intended well well pad where controlled releases are routed to, (4) the system of precision orifices and solenoid valves that regulate gas flow to emission points, and (5) the emission points. Each well pad at METEC has either the same or slightly different P&ID but the core operational structure is the same as described below.

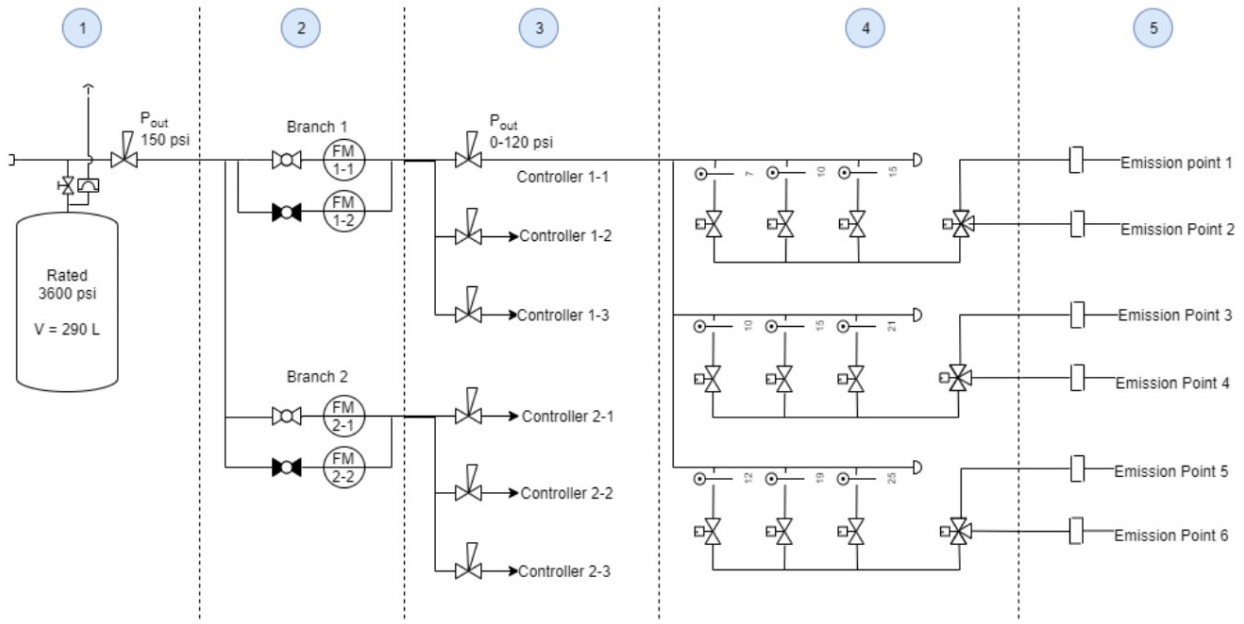


Figure 2.7: The piping and instrumentation diagram of a typical gas metering and transport system at METEC [122]. The figure shows (1) the gas storage system, (2) the gas flow metering system, (3) the control and supply system from the gas supply house to the control boxes at the intended well pad controlled releases are routed to, (4) the system of precision orifices and solenoid valves that regulate gas flow to emission points, and (5) the emission points.

1. **Section 1:** All the well pads at METEC are equipped with GSHs (Figure 2.8(a)); shed-like structures that house the flow meters, manual pressure regulators and valves, and heater (to keep flow meters within the desired operating temperature during low-temperature conditions) used to manage controlled releases to emission points. A truck (Figure 2.9(a)) with a retrofitted fuel system supplies compressed natural gas (CNG) from the local gas distribution network to a set of onsite 290 liters, 3600 psi-rated steel gas storage cylinders (Figure 2.9(b)) attached to each GSH regularly.

Figure 2.8: (a) The gas supply house (GSH) that regulates the controlled release of gas to emission points embedded within equipment units in well pads 4 and 5. The figure also shows the gas storage bottles provided to augment the onsite gas storage capacity for releases to the well pads. (b) The ≈7m weather station located near the western end of the test center.

Gas samples are obtained after every refill of the cylinders for gas composition analysis. Gas flow from the cylinder outlet to the gas flow metering system (section 2) is maintained at a pressure of 150 psig with a manual pressure regulator (0 - 200 psig). A heated 50/50 propylene glycol loop from the heater in the GSH averts the freezing of the gas supply line to the gas flow metering system because of the Joule-Thomson effect caused by the significant drop in gas pressure. An adjustable regulator further regulates the gas pressure between 0-125 psig, reducing drift and improving flow stability as gas flows into section 2.

Figure 2.9: The top panel (a) shows the truck used for supplying compressed natural gas (CNG) to the onsite gas storage canisters/cylinders at METEC. The bottom panel (b) shows a set of the 290 liters, 3600 psi-rated steel canisters attached to each gas supply house at METEC, used for storing CNG.

2. **Section 2:** The gas flow is then split according to the configuration of the manual valves at the inlet of the thermal mass flow meters as shown in Figure 2.10. The selection of flow meters during a test depends on the flow meter's span, intended emission rate and points, and the number of simultaneous controlled releases planned for the test. The electronic gas mass flow meters (OMEGA FMA-17xx series) utilize heat transfer through a heated tube to measure molecular gas flow rate and operate normally up to a maximum gas pressure of 500 psig and a temperature range of 0 to 50 °C. The four spans of the flowmeters installed across the GSHs at METEC are 0-7.5 lpm methane, 0-75 lpm methane, 0-150 lpm methane, and

0-375 lpm methane with an accuracy of ± 0.075 lpm, ± 0.75 lpm, ± 2.25 lpm, and ± 5.625 lpm, respectively.

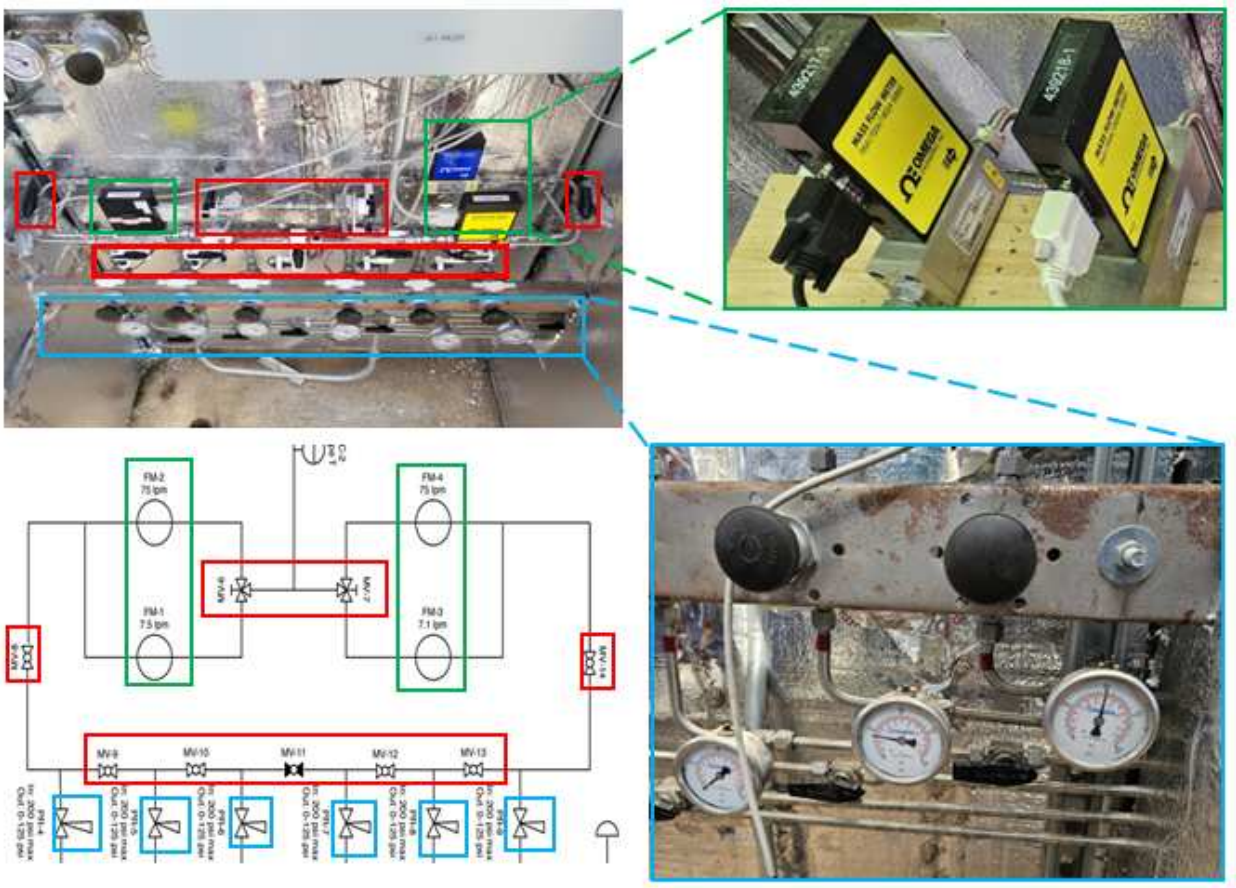


Figure 2.10: Figure shows the interior of a gas supply house at METEC. The green boxes highlights the thermal mass flow meters (of different spans) used to measure flow rates. The red boxes highlights the manual valves that select which emission point gas is routed to and which flow meter gas is measured from. The blue boxes shows the manual pressure regulator used to adjust pressure differentials needed to achieve a desired flow rate.

3. **Section 3:** The gas flow branches from the flow meters are divided using directional manual valves (Figure 2.10) and then routed through buried, small diameter steel piping to the intended controllers at well pads as shown in Figure 2.11(a). Pressure regulators in the GSH but downstream of the flow meters independently regulate the pressure (0 - 125 psig) of gas flows to the controllers. Each equipment group across the well pads is assigned a controller

that uses a series of solenoid valves and precision orifices to control the gas flow rate to each emission point.

4. **Section 4:** METEC is equipped with twenty-one flow controllers. Each controller (Figure 2.11(a)) has an inlet manifold and two or three rows of solenoid gas shut-off valves. The gas flow from the valves in each row flows into an outlet manifold and is directed to one of two emission points by directional solenoid-actuated and manual valves. All the gas shutoff valves on each row are fed from the inlet manifold by different sizes of precision orifices. Release rates using the precision orifices at a fixed upstream pressure are actualized through choked flow conditions.

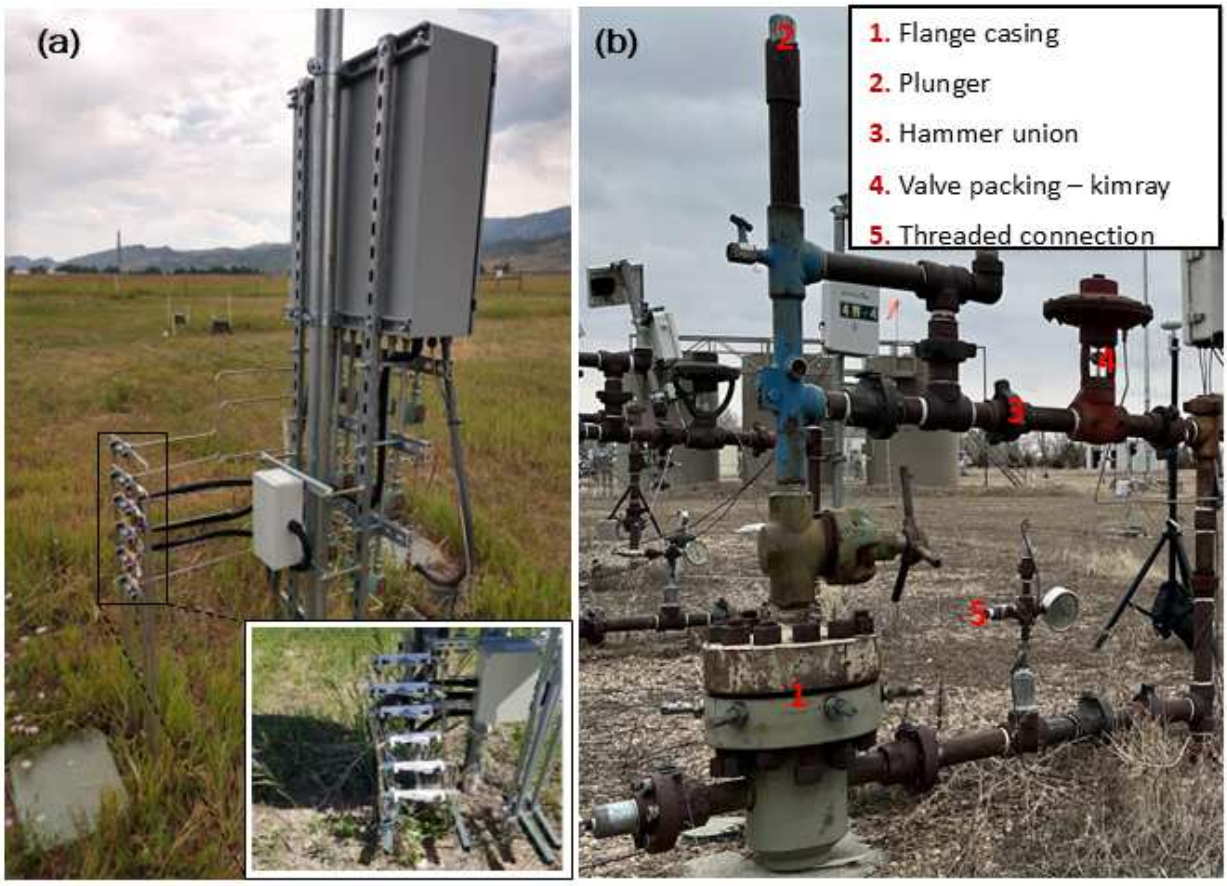


Figure 2.11: The right panel (a) shows the manual valves downstream of the gas supply house used to route gas flow to desired emission points. The left panel (b) shows some of the emission points strategically hidden at representative sources known to leak on a wellhead to mirror real emission scenarios.

By changing pressure conditions at the inlet manifold using the manual pressure regulators downstream of the flow meters in section 3, the release rates span of each row of the controller is adjusted. The multiple rows on the controllers allow the controller to achieve simultaneous controlled releases through more than one emission point.

5. **Section 5:** Controlled releases from the flow controllers are transported to at least 1 of the 198 above-ground emission points hidden at locations on a piece of equipment known to leak or vent from field measurement studies as shown in Figure 2.11(b). This setup allows for "blind" testing for remote LDAQ solutions and survey solutions at METEC.

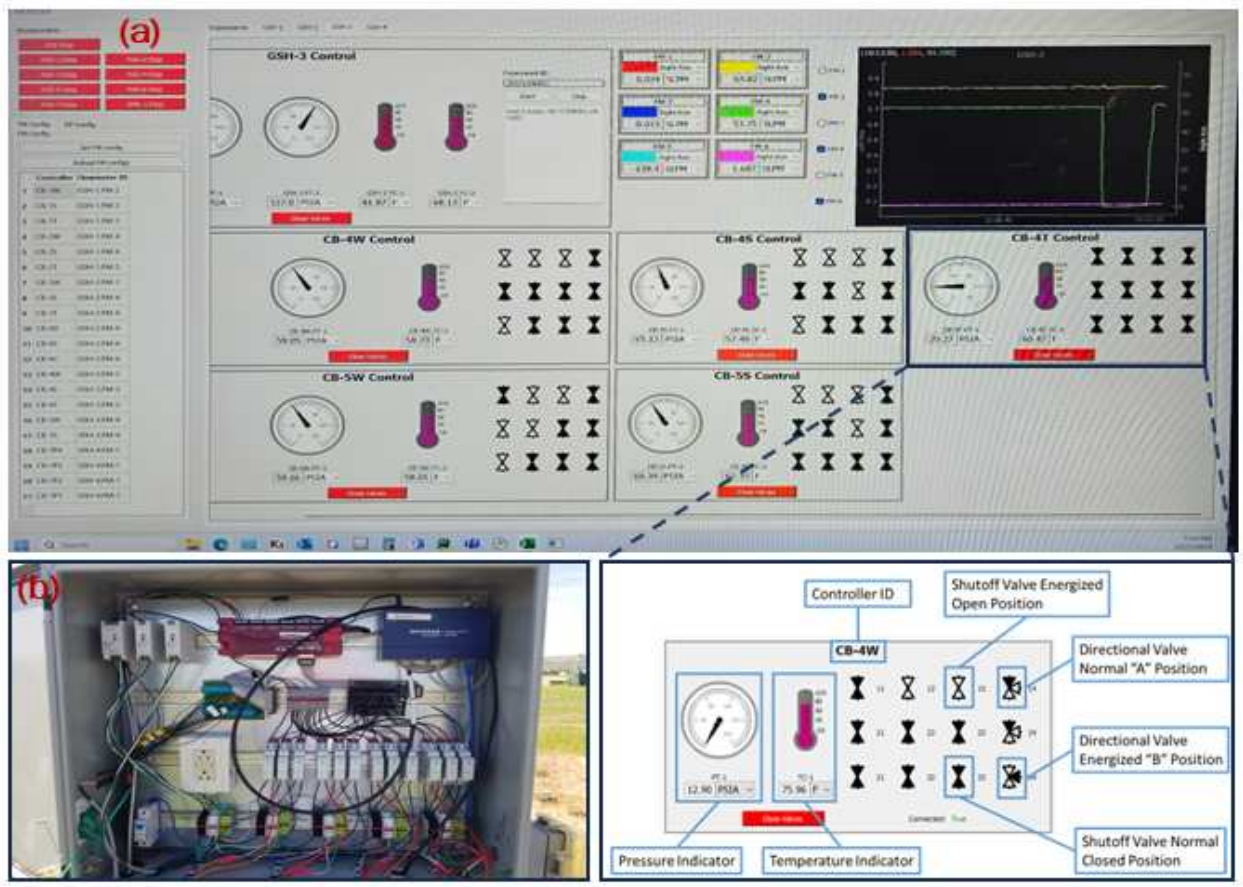


Figure 2.12: (a) The control panel of the electronics that programmatically automates controlled release of gas and data collection at METEC. The figure shows buttons for controlling valves that control gas flow and the indicators monitoring the pressure and temperature of the gas. (b) The control box contains the electronics that translate computer commands from the control panel to instructions to the solenoid valves managing gas flow to emission points.

Gas releases and data collection at METEC are automated and controlled from a computer operated by the facility operator. The control software allows the operator to actuate any electronic valve across the GSHs and controllers in the facility, and monitor readings from flowmeters, pressure transducers, and thermocouples throughout the system on the software's graphical user interface as shown in Figure 2.12(a). Computer-based instrumentations (Figure 2.12(b)) translate programmed instructions from the control software to actionable commands to electronic valves, gas flow meters, etc., and subsequently relay sensor and state readings back to the computer. Meteorological data (e.g., relative humidity, wind speed/direction, ambient temperature, and barometric pressure) measured using an ≈7m stationary 3-D sonic anemometer (Figure 2.8(b)) and controlled release data (e.g., flow rates and valve positions) are collected at a frequency of 1 Hz.

2.2 Testing Process

The protocols for continuous monitors and survey LDAQ solutions [131, 132] were designed to test LDAQ solutions as integrated systems and not the individual subsystems (e.g., sensors, deployment approach, or algorithm/analytics) to evaluate the performance of LDAQ solutions as deployed at O&G facilities. The analytics/algorithms of LDAQ solutions transform raw sensor measurements (e.g., ambient gas species concentration readings, meteorological, and geographical data) into actionable data such as the presence and absence of emissions (i.e., detections), emission rate estimates (i.e., quantification), and emission source attribution (i.e., localization). Performance evaluation is based upon these actionable inferences using the metrics to be discussed.

In general, the protocols divide testing into a set of discrete *experiments*. Each experiment consists of one or multiple simultaneous controlled releases of CNG gas, each emitting at a steady emission rate for a specified duration. For experiments with multiple controlled releases flowing through the same gas flow meter, a pre-calibration was done before releases began to correctly meter and log the rate of each controlled release. Experiments with multiple controlled releases were intended to evaluate an LDAQ solution's ability to characterize, isolate, and attribute emissions to individual source locations under prevailing meteorological conditions. To give LDAQ solu-

tions the best opportunity to isolate and estimate the characteristics of each emission source, the study team ensured that no two controlled releases within an experiment flowed through the same equipment unit and drastically limited scenarios where two consecutive experiments had controlled releases flowing to the same equipment unit. This simplifies observed emissions behavior in real O&G facilities where emitters may follow random patterns or emit at variable rates. Experiments were separated by a period with no emissions to allow LDAQ solutions to identify the start and end of each controlled release event by recognizing a return to background atmospheric concentration levels.

Onsite gas chromatography (Figure A.1) is used to identify and analyze the species and chemicals that make up the CNG used for testing, with which the methane equivalent of the gas is calculated. All controlled gas releases during testing had a mean gas composition by volume of 76%-88% of methane, 12%-20% of ethane, 1%-4% of propane, and a trace amount of heavier hydrocarbons and other gases. For each controlled release, METEC logged the timing, location, release rates with their associated uncertainties, gas composition, and prevailing meteorological data, which were time-averaged over the release duration. METEC kept a maintenance record, documenting facility downtime and the timing of faulty experiments and controlled release events non-compliant with the protocols (e.g., venting gas supply lines, controlled releases on well pads not used for the study, etc.) so they could be excluded during data analysis. All tests were conducted between 2021 and 2023.

2.2.1 Continuous Monitoring Solutions

Testing was conducted exclusively on pads 4 and 5 at METEC covering $\approx 8450\,\mathrm{m}^2$ due to the complexity and the large spatial extent of the combined well pads (Figure 2.1). Given that continuous monitors operate near-continuously regardless of the prevailing weather conditions for an extended period (e.g., months to years), testing was conducted 24 hours per day, every day, exclusively for at least 11 weeks. Exceptions included winter conditions with temperatures below the operating specifications of METEC's thermal flow meters (OMEGA FMA-17xx series).

Testing was performed single-blind: LDAQ solutions were unaware of the timing, number, location(s), durations, and release rates of controlled releases by the test center (METEC). Experiments were designed to sweep the range of tests (e.g., emission rate, release duration, etc.) and meteorological (e.g., wind speed, temperature, etc.) conditions needed to characterize the *probability of detection* curves of the LDAQ solutions tested. The study team periodically analyzed the performance of solutions during testing to choose the emission rates and release durations for subsequent experiments. This was intended to populate test conditions with a small sample size (e.g., larger rate and longer duration controlled releases) to "map" the *probability of detection* curve of solutions.Essentially, the detection performance of LDAQ solutions tends to improve with increasing controlled release rates. Some solutions require extended surveying or averaging of sensor readings to detect emissions. Testing in this manner allows for a comprehensive characterization of LDAQ solutions' detection behavior across a wide range of testing conditions using a *probability of detection* curve. The release rate of controlled releases and experiment duration were selected considering METEC's operational constraints, e.g., available gas supply, emission point orifice size, and safety.

Participating continuous monitors were recruited through an open invitation advertised on METEC's website and through contacts in the PDC. Vendors of participating LDAQ solutions were required to install their systems at least 3 days before testing begins to participate in the mock testing, which allows both METEC and vendors to troubleshoot their respective setups. Vendors decided on the number of sensors, positioning of sensors, and the equipment units to monitor; the only limitations imposed by the study team were related to safety (e.g., trips and falls) and obstructions (e.g., system installation near or along driveways). Vendors were asked to install as they would at real O&G facilities. This implied that some vendors either installed their solutions along the fenceline of the pads or around the equipment units monitored (Figure 2.13). Only controlled releases from the designated test area monitored were considered in the performance analysis of each LDAQ solution. In many field applications, sensor locations are likely restricted to the periphery of the facilities, while the number of sensors installed per facility largely depends

on the cost of deployment and the size of the O&G facility. Vendors were responsible for all auxiliary instruments, like weather stations, needed by their systems to operate. This also included the communication systems connecting their on-site hardware to backend servers and algorithms operating offsite; most LDAQ solutions utilized cellular data for this purpose. After installation and mock testing, the vendors left the test center and operated their systems remotely, except to conduct repairs or maintenance activity on their system.

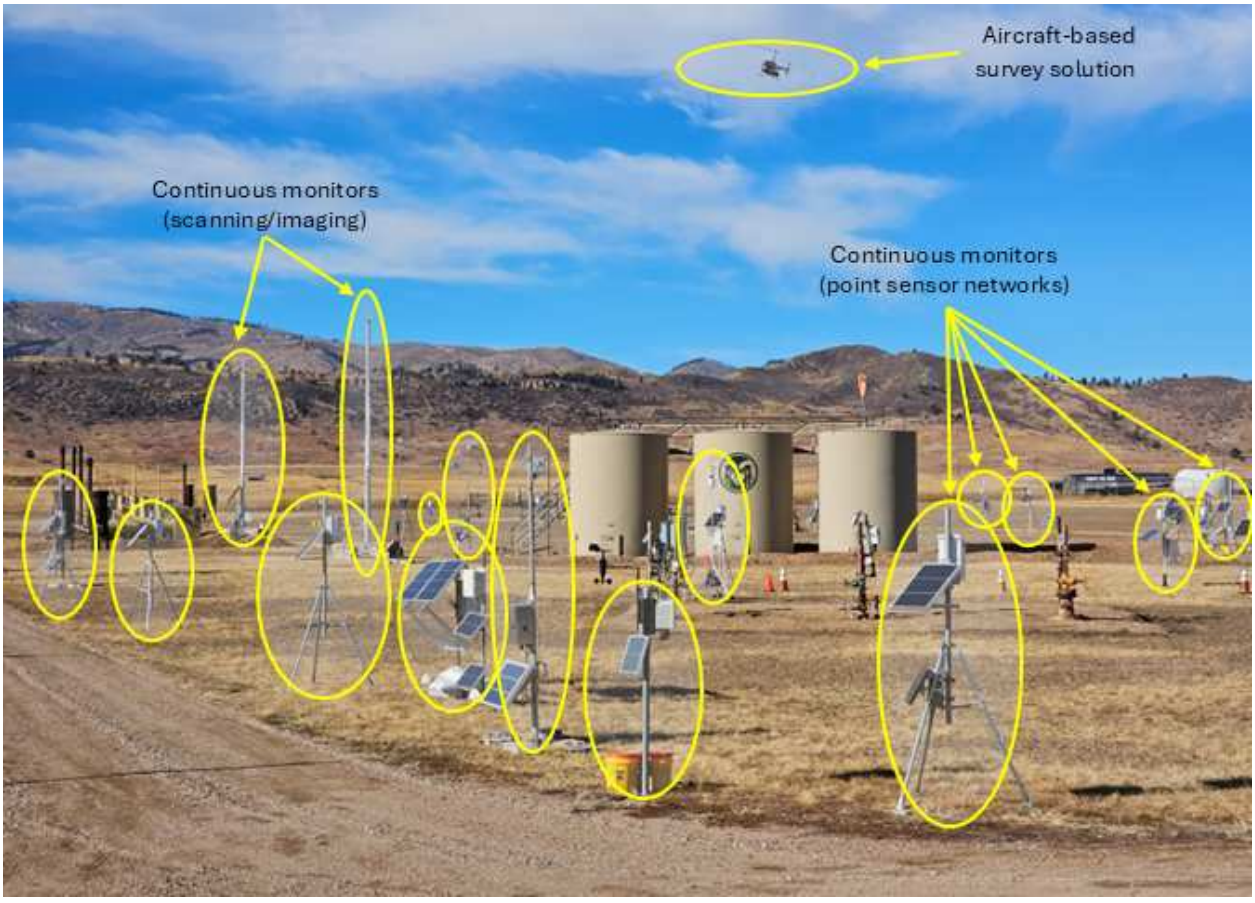


Figure 2.13: The deployment of continuous monitors at METEC during testing.

In this study, continuous monitors are broadly categorized into two categories based on ambient concentration sensing methodology, and they are:

- ***Point sensor network:*** These are continuous monitors deploying multiple point concentration measurement sensors that sense hydrocarbons and use proprietary algorithms that

include inversion modeling to combine meteorological and concentration readings to infer detections at minimum.

- ***Scanning/imaging:*** These continuous monitors utilize scanning/line lasers or short-/mid-wave infrared cameras to detect emissions. They either perform open-path integration to measure ambient emission concentrations or use ambient illumination to generate and visualize two-dimensional images of hydrocarbon plumes. These visualizations are then combined with meteorological data to infer detections at minimum.

A total of 16 continuous monitors were tested; four of the solutions were tested twice approximately one year apart. The participating LDAQ solution developers, in alphabetical order, are developed by the following companies: Baker Hughes, CleanConnect.ai, Earthview.io, Honeywell, Honeywell Rebellion, Kuva Systems, Luxmux Technology, Molex, Pergam Technical Services, Project Canary, QLM Technology, Qube Technologies, Sensia-solutions, Sensirion, and SLB. Due to confidentiality agreements, each solution is arbitrarily identified here by a unique identifier. While all LDAQ solutions tested detection and emission source localization capabilities, not all the solutions tested emissions quantification capability. Table 2.2 summarizes the tested capabilities and conditions of solutions that participated in the study. During testing, each LDAQ solution vendor sent *detection reports* containing data inferred from sensed emission (e.g., emission rate, emitting source, etc.) to a unique email address provided by METEC. While this process was automated for some solutions, others required human support to interpret and prepare reports according to the template in the protocol. Such human engagement delayed detection reporting to the test center by days or weeks; likewise, LDAQ solutions with automated reporting that required varying levels of human support when their data transmission system failed. The email setup at METEC parsed through reports as they arrived and automatically rejected those non-compliant with the protocol's reporting template [131]. This contrasts with field deployments, where operators obtain measurement data through web-based dashboards—such as time series of methane concentrations or emission rates—transmitted by LDAQ solutions installed in O&G facilities. Operators then interpret this data based on predefined alert thresholds to make actionable detection decisions.

Table 2.2: Characteristics of participating continuous monitors and testing conditions.

Solution			Tested Capability	Test Conditions[1]	
ID	**Type**	**Count**	**Quantification**	**Release Rate (g CH_4/h)**	**Wind speed (m/s)**
Testing under the protocol for the first time.					
A†	Point sensor network	8	✓	1053[0.4, 6390]	3.3[0.8, 13.5]
B†	Scanning/imaging	1	✓	1000[0.4, 6390]	3.3[0.8, 13.5]
C†	Point sensor network	6	✓	1004[0.4, 6390]	3.3[0.8, 13.5]
D†	Point sensor network	8	✓	1053[0.4, 6390]	3.3[0.8, 13.5]
E†	Point sensor network	16	✓	1059[0.4, 6390]	3.3[0.8, 13.5]
F†	Point sensor network	8	✓	1053[0.4, 6390]	3.3[0.8, 13.5]
G†	Scanning/imaging	1	✓	788[0.4, 5370]	4.0[0.8, 11.6]
H†	Scanning/imaging	1	X	1246[0.4, 6390]	3.8[0.8, 13.5]
I*	Scanning/imaging	1	X	813[0.3, 1960]	2.8[0.3, 5.8]
J*	Scanning/imaging	2	X	1109[4.7, 4990]	2.7[0.1, 10.0]
K†	Point sensor network	7	X	668[0.3, 2980]	2.7[0.1, 12]
L‡	Scanning/imaging	1	✓	1913[11, 7098]	3.3[0.8, 10.0]
N‡	Point sensor network	18	✓	1516[8, 7098]	3.1[0.7, 10.0]
O‡	Scanning/imaging	1	✓	1656[6, 7098]	3.3[0.8, 10.0]
P‡	Point sensor network	6	X	1518[6, 7098]	3.1[0.7, 10.0]
Q‡	Point sensor network	13	X	1518[6, 7098]	3.1[0.7, 10.0]
Testing under the protocol for the second time.					
A‡	Point sensor network	8	✓	1518[6, 7098]	3.1[0.7, 10.0]
B‡	Scanning/imaging	1	✓	1518[6, 7098]	3.1[0.7, 10.0]
D‡	Point sensor network	8	✓	1518[6, 7098]	3.1[0.7, 10.0]
F‡	Point sensor network	10	✓	1518[6, 7098]	3.1[0.7, 10.0]

[1] The value outside the bracket indicate the mean value, while the minimum and maximum values of the test condition are enclosed within the square brackets.
* LDAQ solutions tested in 2021.
† LDAQ solutions tested in 2022.
‡ LDAQ solutions tested in 2023.

The detection reporting approach used in this study eliminated inference errors and biases associated with the study team interpreting raw measurement readings of solutions. According to the protocol, each detection report, which either identifies a fresh emission or updates previous reports, contained at minimum the following: *DetectionReportID* - an incremental unique identifier of each detection report. *EmissionSourceID* - a unique identifier referencing the emitter the detection report identifies. *EmissionStartDateTime* - the estimated time and date a detected emission started emitting. *EquipmentUnit* - the identifier of the equipment unit on which an emission was detected. *Gas* - the gas species measured to infer a detection. LDAQ solutions also reported system *downtime* which captures periods during testing when LDAQ solutions were offline (e.g., not taking measurements), which should be ignored by the study team during result analysis. Before the performance analysis for each LDAQ solution, the study team excluded detections (1) reported during METEC's maintenance and LDAQ solutions' downtime periods, (2) reports with *EmissionStartDateTime* before and after the analysis window of each solution, and (3) reports identifying *EquipmentUnit* outside the fence line of METEC (OFF FACILITY) in the latest detection report. These exclusions were done to avoid bogus *false positive* detections. Similarly, the study team excluded controlled releases (1) conducted during METEC's maintenance and solution downtime periods, (2) conducted outside the analysis window of the solution, and (3) with durations shorter than required to get a stable flow meter reading. These exclusions were done to avoid spurious *false negative* detections. All detection reports referencing the same *EmissionSourceID* were grouped as one report: for the same *EmissionSourceID*, the time at which the first detection report (smallest *DetectionReportID*) was received by METEC was paired with the data contained in the last detection report (largest *DetectionReportID*). A buffer time of 20 minutes before and after the timing of each controlled release while pairing controlled releases to detection reports (*detection classification*). The buffer time accounted for emissions during experiment pre-calibration periods and the residual emissions detected by solutions after the end of a controlled release.

2.2.2 Survey Solutions

The survey protocol [132] divides testing into 3 activities:

1. **Documentation of the system under test:** Survey LDAQ solutions were encouraged to document descriptions of the system configuration (e.g., model number of each hardware, revision number of software, etc.), components (e.g., sensor, deployment platform, etc.), methodology applied (e.g., number of passes, flight speed and height, etc.), and the personnel needed to perform emissions surveys at O&G facilities.

2. **Emission surveys:** Testing was conducted during the day (i.e., typically between 8 a.m. and 5 p.m. U.S. mountain time) for 3 to 5 days. Testing for a significantly shorter number of days and only during the day compared to continuous monitors was because survey solutions operate periodically, typically for a few minutes to hours during the daytime. Testing was single-blind, as the solutions were unaware of controlled releases' timing, location, and release rate. For each experiment, METEC defined a *survey area* that consisted of either a single or two adjacent well pads and initiated controlled gas releases. The well pads at METEC were divided into 3 survey areas: well pads 1 & 2, well pad 3, and and wellpads 4 & 5 to achieve different site sizes and operational configurations as shown in Figure 2.1. While testing across multiple O&G assets and well pads facilitated the evaluation of *survey time*, a key performance metric for solutions, differences in the surface area, layout, and equipment unit characteristics (e.g., type, count, orientation, etc.) of the well pads also varied testing complexity and rigor. Additionally, to further vary test complexity, each experiment's controlled release(s) were restricted to a well pad or distributed across all or a portion of the survey area defined. The release rates of gases tested match within an order of magnitude the component-level leak sizes previously measured by METEC at natural gas production sites in the U.S., excluding liquid unloading or major failure conditions that produced exceptionally high-rate emission events [11, 65, 141], which the type of LDAQ solutions tested would typically encounter during actual field deployments. The test center recorded the timing, emission rate, location (i.e., emission points and GPS coordinates), and gas composition of

controlled releases. As with Ravikumar et al. [120], the release schedules in this study were intentionally complicated (i.e., variation in the position and number of emission sources per experiment and the *survey areas* used during such experiments) not only to improve testing complexity but also to minimize scenarios where gas odor provided consistent cues to surveyors during testing.

At the start of each test day, the study team briefed solutions surveyors about safety and how the test would be conducted. During each experiment, solutions conducted emissions inspection based on their survey protocol(s), staying within the boundary of the survey area defined for the experiment. Multiple LDAQ solutions were sometimes deployed in the same survey area during an experiment and therefore had overlapping *survey times*. In such cases, solutions alternated while surveying each equipment unit with instructions to avoid communicating with one another to prevent one team from learning about emission points or rates from another team. Solution surveys were sequenced to limit the impact (if any) of one system's detection methodology on others. The only exception was a scenario that included both a drone-based and an automobile-based LDAQ solution. Drone flights generate turbulence and disturbances that can affect the formation and downwind transport of gas plumes measured by the automobile-based solution, typically driving at the edge of the survey area. The end of the emissions survey by the LDAQ solution(s) signaled the end of each experiment, and then the test center stopped all controlled releases.

3. **Detection reporting:** The test protocol stipulates a reporting template for solutions to record experimental and detected emissions data [132]. Vendors of LDAQ solutions conducted quality control and validation of their data before reporting to the test center. Vendors were encouraged to submit recorded data to the test center at the end of each test day containing at minimum the detection report ID, timing of surveys, source location, and the gas species measured to make detection. Reporting measured gas species allowed performance analysis of solutions with either methane-specific or multi-species sensors to be on the same basis.

As with continuous monitors, participating survey LDAQ solutions were selected through contacts in the PDC and an open invitation advertised on METEC's website. Twelve survey solutions were tested for the first time under the protocol, with three solutions tested again within 3 months to 1 year of their initial tests. LDAQ solutions were required to deploy as they would in actual field measurement campaigns (i.e., number of personnel, survey speed/pace, survey height/distance, etc.), providing all auxiliary equipment (e.g., anemometers, battery packs, power cables, internet connection, etc.) needed to perform emissions surveys during testing. The list of vendors of participating LDAQ solutions arranged in alphabetical order is as follows: ABB HoverGuard, ABB MicroGuard, ABB MobileGuard, ChampionX AURA OGI, Cimarex FLIR OGI, Cimarex Ventus, Distran Ultra, Heath EyeCGas, Heath RMLD-CS, Konica Minolta GMP02, Montrose Pergam Laser Falcon, and Xplorobot Laser OGI. Given the confidentiality agreement, solutions were identified with arbitrarily selected, alphabetic identifiers. For this study, participating survey solutions were grouped into three categories as shown in Figure 2.14:

- ***Mobile solutions***: LDAQ survey solutions that are automobile- and UAV/drone-based.
- ***Handheld OGI solutions***: Traditional handheld OGI camera solutions.
- ***Advanced handheld solutions***: Includes other handheld solutions that do not detect emissions using OGI technology alone (if at all). These include sensing techniques like acoustic sensing, infrared absorption spectroscopy, tunable diode laser absorption spectroscopy (TDLAS) technology, etc.

Table 2.3 summarizes the solutions that participated in the study with their deployment characteristics and selected test conditions. While all solutions tested detection and source localization capabilities, only solutions A and G quantified emissions.

For reported data to be considered valid, solutions submitted their survey and detection reports within one week of test completion. This time limit reflects common field practice [50, 59, 106]. The study team collated and quality-controlled all data, including release rates and meteorological

data collected by the test center, and all data reported by LDAQ solutions to perform *detection classification*.

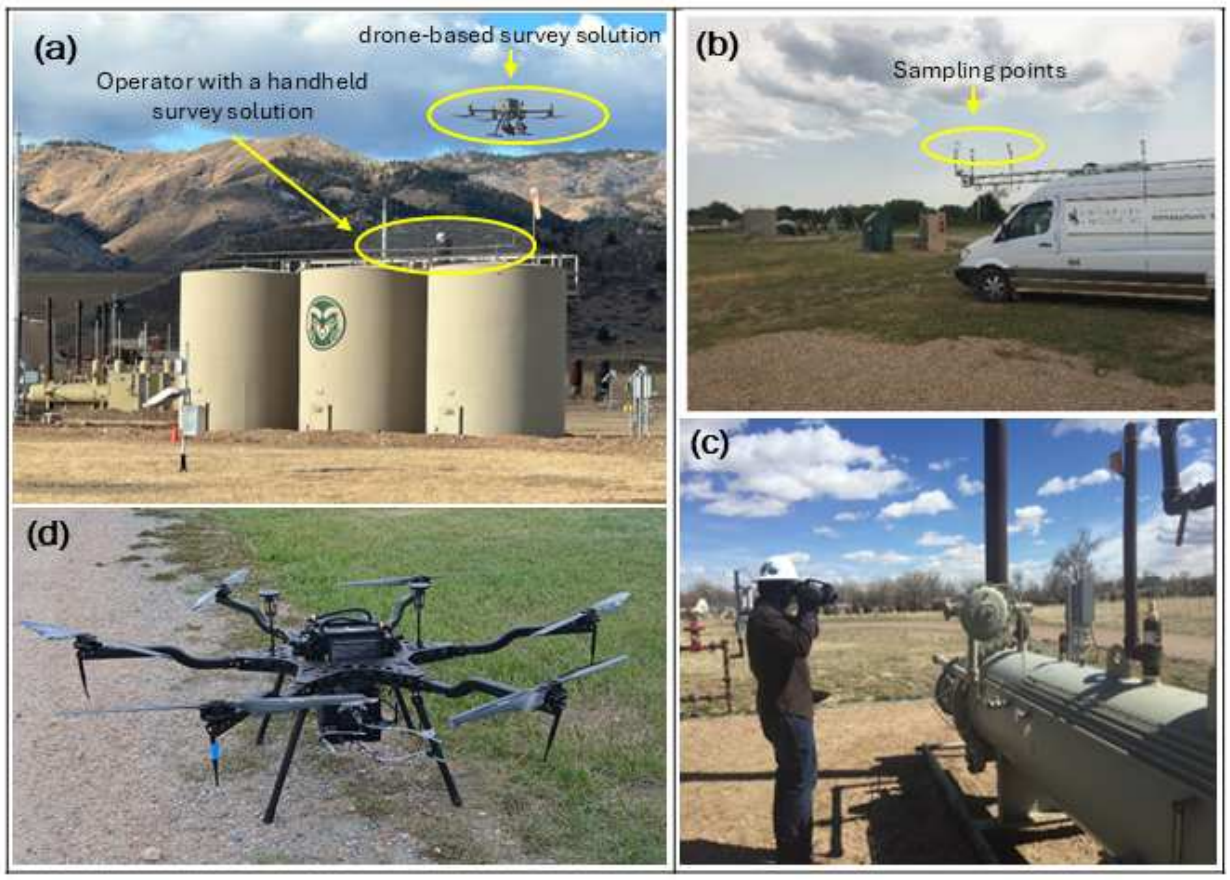


Figure 2.14: The deployment of survey LDAQ solutions at METEC. (a) shows surveys using drone-based and handheld solutions. (b) shows survey using a vehicle based solution. (c) shows a survey using a handheld solution. (d) shows a close-up shot of a drone-based solution.

The test center kept an operator log and a maintenance record during the testing period to facilitate the *detection classification* process and the exclusion of controlled releases and detection reports data invalidated by the requirements of the test protocol [132]. For each LDAQ solution, the operator log documented the unique ID and characteristics (i.e., timing, number of emission points, etc.) of each experiment; the ID was communicated to the solution during testing for documentation. The ID linked the detection and survey data reported during an experiment with the test center's record of the experiment. The maintenance record documented the list of controlled re-

leases or periods to be excluded from the result analysis, either because the controlled release was non-compliant with the testing protocol or due to an unplanned release (i.e., venting an emission point gas supply line).

Table 2.3: Characteristics of participating survey LDAQ solutions and testing conditions.

Solution			Tested Capability	Test Conditions[1]	
ID	**Platform**	**Category**	**Quantification**	**Release Rate (g CH_4/h)**	**Wind speed (m/s)**
Testing under the protocol for the first time.					
A*	Vehicle	Mobile	X	214 [26, 895]	5.3 [1.1, 13.4]
D*	Drone	Mobile	X	73 [3, 297]	3.0 [0.9, 5.7]
F‡	Drone	Mobile	X	471 [30, 2027]	2.5 [1.2, 4.1]
I*	Drone	Mobile	X	175 [22, 586]	4.5 [1.1, 9.3]
B*	Handheld	Handheld OGI	X	76 [4, 297]	3.3 [1.3, 5.7]
E*	Handheld	Handheld OGI	X	198 [4, 808]	3.0 [0.8, 4.6]
G‡	Handheld	Handheld OGI	✓	500 [22, 2110]	3.6 [0.9, 9.0]
L‡	Handheld	Handheld OGI	X	553 [23, 2586]	3.6 [0.9, 8.7]
C*	Handheld	Advanced Handheld	X	198 [4, 808]	3.0 [1.1, 4.6]
H‡	Handheld	Advanced Handheld		471 [3, 2106]	3.4 [0.9, 8.8]
J‡	Handheld	Advanced Handheld	X	464 [23, 1651]	2.9 [0.7, 8.0]
K*	Handheld	Advanced Handheld	X	194 [22, 640]	5.2 [1.2, 12.8]
Testing under the protocol for the second time.					
A†	Vehicle	Mobile	✓	164 [11, 982]	4.3 [0.8, 13.3]
K†	Handheld	Advanced Handheld	X	164 [11, 982]	4.3 [0.8, 13.3]
H‡	Handheld	Advanced Handheld	X	355 [4, 1934]	2.6 [0.9, 3.6]

[1] Indicates the mean, minimum, and maximum values of the test condition.
*, †, ‡ LDAQ solutions tested in 2021, 2022, and 2023, respectively.

2.2.3 Quantitative Optical Gas Imaging (QOGI) System

The OGI camera is widely used for emissions monitoring surveys at O&G facilities, as it is anecdotally effective and regulatory-approved in the U.S. and Canada. However, OGI cameras do

not quantify detected leaks. In studies that required both detection and quantification, emission quantification was performed as an additional measurement step using an auxiliary method (Hi-flow sampler [11,20,142], downwind tracer flux [11,65,143], and the other test method [144–146]) that requires additional labor, is time-consuming, depends on direct contact with the emitting source, which might be inaccessible, and/or depends upon favorable wind transport to perform measurements. The QOGI technology, an add-on device to the OGI camera, can address these challenges; however, comprehensive testing that is more representative of field (i.e., actual O&G facilities) deployment and conditions is needed to understand quantification performance and characterize limiting environmental and operational factors, unlike prior studies [109, 114].

Testing in this study was conducted from June 20th to June 24th, 2022. The QOGI system tested consisted of a FLIR™ GF320 OGI camera and a FLIR™ QL320 QOGI tablet (henceforth "FLIR tablet"). The Providence Photonics QL320 QOGI tablet (henceforth "legacy tablet"), which operates using an older version of the quantification algorithm used in the FLIR tablet, was used as a backup whenever the FLIR tablet had a flat battery or became inoperable. A member of the measurement crew (2 persons) had attended an in-person QOGI training session by Providence Photonics, while the other crew member had previously screened < 60 facilities with an OGI camera. The field crew followed the user manual provided by FLIR. The tablets quantify emissions by analyzing the image of the plume captured by the OGI camera. The tablets can be used in a "tethered" or "Q-mode" configuration. Under the "tethered" configuration, the camera and the tablet are deployed together in the field and connected using a USB cable such that live feed video from the camera is transferred to the QOGI tablet for quantification while the emission is under observation. Under the "Q-mode" configuration, the OGI camera records emission videos together with required input parameters (e.g., wind speed, etc.), and quantification is performed later by importing and analyzing the videos on the QOGI tablet. In this study, the "tethered" configuration was used as it reflects the preferred deployment approach in actual O&G facilities. This study evaluated only the quantification performance, i.e., it did not test the detection performance of OGI camera surveys, as comprehensive studies on the subject exist in the literature [43, 44, 147].

Therefore, testing was performed partially blind: the emission rates of releases conducted by the test center (METEC) were unknown to the measurement crew; however, the location of the leak source was known. The testing process is summarized as follows:

1. The METEC facility operator selected an emission source, initiated a controlled release, waited until the release rate was steady, and then informed the measurement crew of the location of the emission. The METEC facility operator assigned each experiment a unique numeric identifier (ID) and communicated that (without the emission rate) to the field crew for documentation.

2. The field crew identified an unobstructed view of the leak source with the gas plume, considering wind direction, measurement distance, and plume background. This was done by observing the emission source from multiple perspectives and distances to find the "best view" of the plume (as adjudged by the measurement crew) with enough space to accommodate the camera tripod and camera operator [113].

3. Parameter data required for quantification were inputted into the tablet which included wind speed (calm (0-1mph), normal (2-10mph), or high (>10mph)), distance to emitting source, leak type (point or diffuse), and ambient temperature. Ambient wind speed and temperature were measured using a handheld digital anemometer while distance was measured with a measurement tape. While handheld digital anemometers can be highly uncertain and less accurate compared to fixed-mount 2-D and 3-D anemometers, it is more convenient and faster to use digital anemometers in the field during measurements. Depending on the geometry of an emission point, the field crew selected either "point" (sources with a diameter < 2 inches, e.g., connectors, valve packing, etc.) or "diffuse" (sources with a diameter > 2 inches, e.g., flanges, thief hatches, etc.) as the leak type on the device before taking measurements. The overlay function was enabled on the tablet to colorize the plume to increase visibility. The field crew also ensured that only the streaming image of the gas plume interacted with the measurement boundary and used the masking feature to remove other areas of visual distur-

bance (e.g., vegetation on the ground, etc.) when necessary, as shown in Figure 2.15.. The field crew selected the viewing angle and distance from the emission source considering the minimum and maximum distance requirements as specified in the manuals for the 23 mm (24°FOV) camera lens—1.5m and 16.5m from the emission source [113].

4. At each camera position, at least 3 consecutive, individual measurements were taken on the QOGI tablet when stable. A stable measurement was defined in the user manual as when the 10-second quantification result was within 10% of the 1-minute quantification result. At least 3 measurements were attempted for each 'measurement opportunity' to assess measurement precision without modifying the observation position or allowing environmental conditions to change substantially. For each measurement, the field crew documented the background of the plume (sky, equipment, or ground). In some instances, 3 successful measurements could not be completed from a selected location due to rapidly changing meteorological conditions. At the end of measurements for each camera position, the device was reset ahead of measurements from the next camera position.

5. For each experiment, the field crew attempted measurements from 3 different camera positions by repeating steps 2-4. The controlled release rate during each experiment was approximately the same for all the measurements conducted for an experiment. Each camera position was assigned a unique ID, as no two positions had the same measurement conditions. Approximately 10 minutes elapsed between camera positions to locate clear views of the plume. Measurement duration varied substantially (a few minutes to >20 minutes) as highly variable meteorological conditions adversely affected measurement efforts in some cases. In addition to a different location, each new camera position had a different measurement distance and/or plume background/perspective. In some instances, fewer than 3 camera positions were identified for an experiment due to obstructions in the field of view or limitations of environmental and/or meteorological conditions.

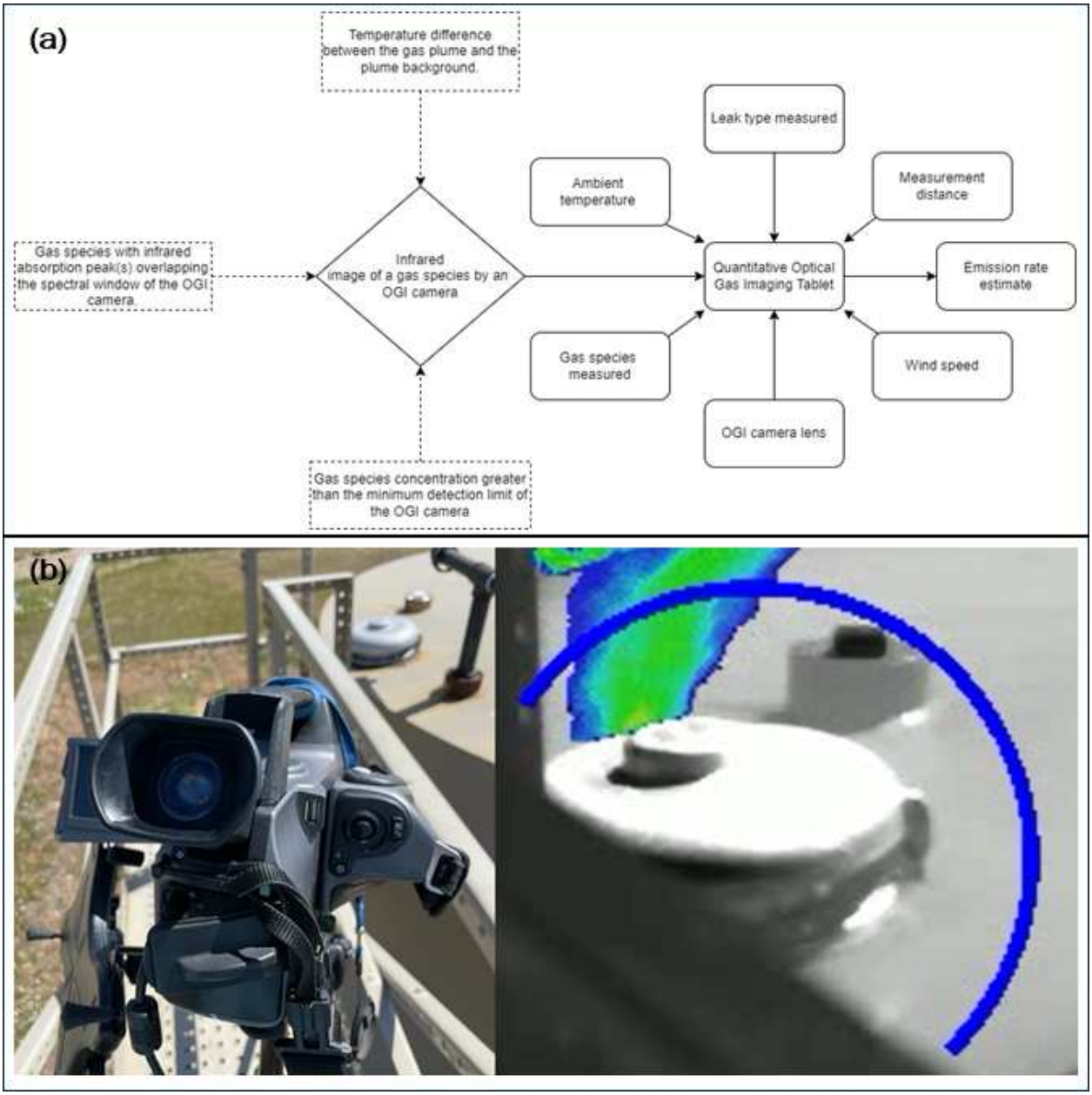


Figure 2.15: (a) A schematic of the QOGI system (OGI camera and QOGI tablet) that was tested in this study. The diagram elements with broken lines or edges represent the conditions required to visualize gas species with an OGI camera to be quantified. The figure also shows input parameters (setups and measurement conditions) into the QOGI device to generate emissions estimates. (b) The figure shows a typical deployment of the OGI camera during the study (left side) and a snapshot of the infrared image of the observed gas in the QOGI tablet (right side). The left side of the figure shows the OGI camera observing emissions from a tank's thief hatch. The right side of the figure shows colorized enhancements overlaid onto the infrared image of the gas plume with a masking feature activated to remove measurement boundaries with visual noise.

6. After completing all measurements for an experiment, the measurement crew notified the METEC facility operator to stop the controlled release to conclude the experiment. The next

experiment was then conducted following the same steps, either using the same emission source at a different emission rate or an emission source in a different location. In total, the field crew attempted 442 measurements, of which 85 of them were excluded due to missing controlled release data (27), Q-mode measurements (11), and wrong leak type selection (47). Wrong leak type exclusions included cases where emissions from sources like flanges and thief hatch (diameter > 2 inches) were wrongly interpreted as "Point" (diameter < 2 inches) sources instead of "Diffuse" leak type during measurements. In summary, 357 individual measurements across 26 controlled releases and 71 camera positions for release rates between 0.1 $kgch_4/h$ and 2.9 $kgch_4/h$.

2.3 Performance Metrics

The detection classification process involved pairing controlled releases with detection reports/measurement data. Therefore, for each experiment and the detection reports attributed to that experiment (i.e., for continuous monitors, this included reported data with *EmissionStartDateTime* within the controlled release start and end times $\pm$ buffer time, while for survey/QOGI solutions, it involved using the unique ID documented by the METEC operator that linked each experiment with associated detection/measurement data), the pairing steps are summarized as follows:

1. The study team filtered all controlled releases and detection reports that attributed an emitter to the same equipment unit tag (as assigned by the test center) and sorted them by controlled release rate and reported emission rates, respectively, in descending order. Each controlled release and detection report is paired as true positive (TP) at the equipment unit level (Figure 2.16(a)) and removed from the unpaired data lists.

2. If the unpaired controlled release and detection report lists from step (1) were not empty, the study team filtered the lists that attributed an emitter to the same equipment group tag (as assigned by the test center) and sorted them by controlled release rate and reported emission rates, respectively, in descending order. Each controlled release and detection report is paired as TP at the equipment group level (Figure 2.16(b)) and removed from the unpaired data lists.

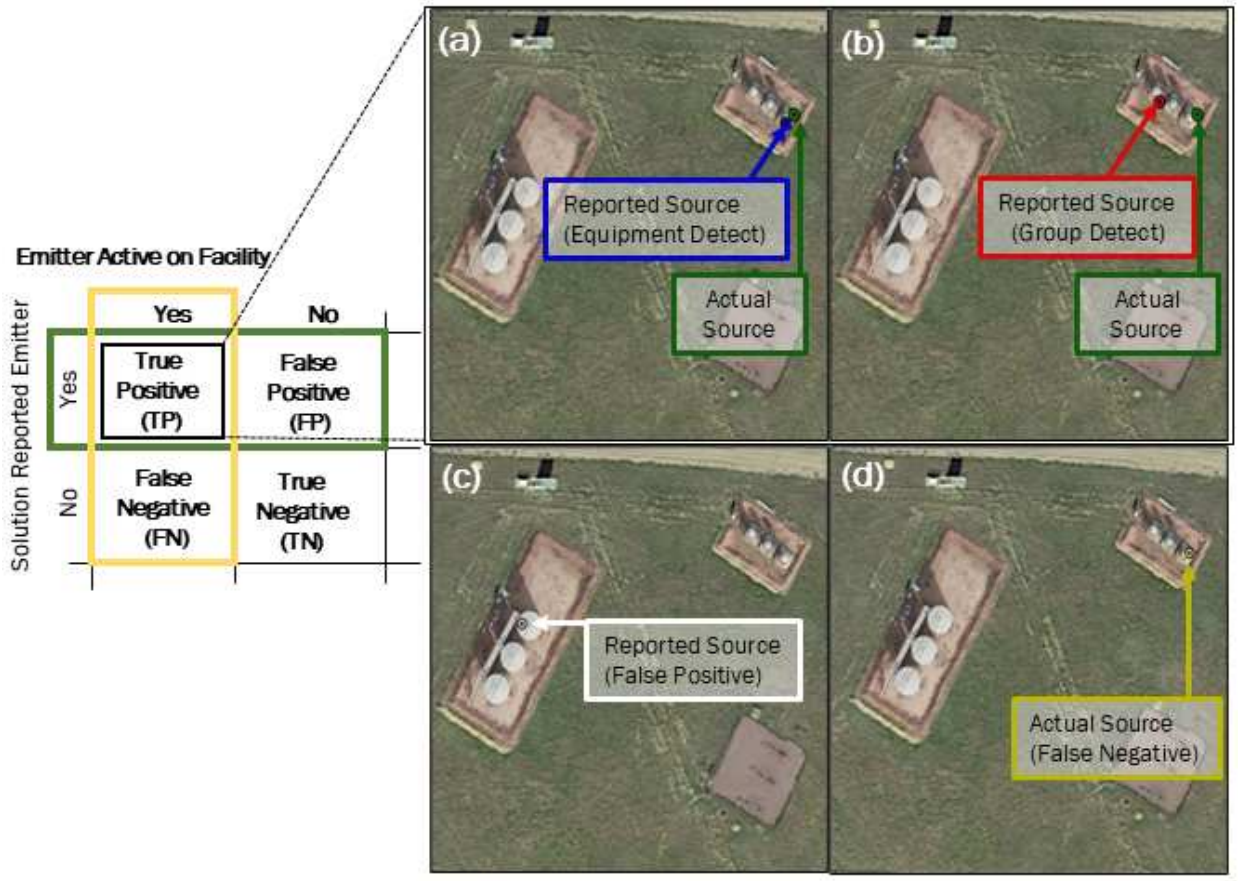


Figure 2.16: The figure shows (a) True Positive detection classification at the same equipment unit level. (b) True Positive detection classification at the same equipment group level. (c) False Positive detection classification. (d) False Negative detection classification

3. Each controlled release and detection report in the sorted (by emission rate in the descending order) unpaired lists remaining from step (2) is paired as TP at the facility level and removed from the unpaired data lists.

4. Each controlled release and detection report left unpaired after step (3) is classified as FP (Figure 2.16(c)) and false negative (FN)(Figure 2.16(d)), respectively.

The result of the detection classification process was used to evaluate performance metrics as specified in the protocols. Key performance metrics are summarized below:

- ***Probability of Detection (POD):*** This represents the probability of detecting an emission over a set of environmental and measurement conditions (e.g., wind speed, emission rate,

release duration, etc.). This is evaluated as the fraction of the count of TP detections to the sum of the counts of TP and FN detections over a set of conditions as shown in equation 2.1.

$$\text{POD}|_{\text{x}} = \left.\frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}\right|_{\text{x}} \tag{2.1}$$

Where x is the set of measurement conditions at which the POD is assessed.

- ***False Positive Fraction (FPF):*** This is the fraction of the count of FP detections to the sum of the counts of TP and FP detections as shown in equation 2.2.

$$\text{FPF} = \frac{N_{\text{FP}}}{N_{\text{FP}} + N_{\text{TP}}} \tag{2.2}$$

- ***False Negative Fraction (FNF):*** This is the fraction of the count of FN detections to the sum of the counts of TP and FN detections as shown in equation 2.3. True Positive Fraction (TPF) = 1 - FNF.

$$\text{FNF} = \frac{N_{\text{FN}}}{N_{\text{FN}} + N_{\text{TP}}} \tag{2.3}$$

- ***Survey Time:*** This is the time taken to complete a survey and is evaluated as the difference between the reported survey start- and end-date times.

- *Localization Precision (Equipment Unit):* This is the fraction of TP detections at the equipment unit, equipment group, and facility levels, respectively.

- *Localization Accuracy (Equipment Unit):* This is the fraction of detection reports at each localization precision level (equipment unit) or better as shown in equations 2.4, 2.5, and 2.6.

$$\text{Correct unit } (\text{LA}_{\text{unit}}) = \frac{N_{\text{TP}_{\text{unit}}}}{N_{\text{TP}} + N_{\text{FP}}} \tag{2.4}$$

$$\text{Correct group } (\text{LA}_{\text{group}}) = \frac{N_{\text{TP}_{\text{group}}} + N_{\text{TP}_{\text{unit}}}}{N_{\text{TP}} + N_{\text{FP}}} \tag{2.5}$$

$$\text{Correct facility } (\text{LA}_{\text{facility}}) = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \tag{2.6}$$

- ***Quantification Accuracy:*** For solutions that estimated the rate (g/h) of the gas species measured, the absolute quantification relative error for each TP detection was evaluated as the difference between the reported emission estimate and the controlled release rate. The relative error was assessed by normalizing absolute error by the controlled release rate. Facility-level quantification relative error was evaluated using all controlled release rates and reported emission estimates considered in the analysis of each LDAQ solution, respectively aggregated over the solution's study duration.

- ***Time to Detection:*** For each TP detection, this is the time difference between when the test center received the first detection report identifying an emission source (*EmissionSourceID*) and the start time of the controlled release it paired with.

- ***Operational factor:*** The fraction of time an LDAQ was operational relative to the total deployment time.

2.4 Data Analysis

The detection classification framework of the protocols penalized extra detection reports (with unique *EmissionSourceID*s) identifying controlled releases already paired with another detection report or emission sources not emitting during an experiment, as FP detections. Hence, the FP detections obtained are categorized into:

1. FP due to no ongoing controlled release—a detection reported when there was no controlled release at the test center.

2. FP due to excess detection report—a detection report (with unique *EmissionSourceID*) identifying a controlled release already paired with another detection report as a new and/or different emission source. For example, reporting multiple detections with different *EmissionSourceIDs* during an experiment with one controlled release.

To evaluate the POD curves of solutions, six functions, including power function (Equation 2.7), binary logistic regression (Equation 2.8), Burr cumulative distribution function (CDF) (Equation 2.9), Frechet CDF (Equation 2.10), Weibull CDF (Equation 2.11), and loglogistic CDF (Equation 2.12)), were assessed and compared using the resulting data of the detection classification process. The binary logistic regression sigmoid function (Equation 2.8) was disqualified because it predicted non-zero POD at zero controlled release rates in some instances, which is physically unrealistic. However, binary logistic regression was used to evaluate the impact of test and measurement conditions on the emissions detectability (TP and FN classification $\rightarrow$ POD) of LDAQ solutions. The power function (Equation 2.7) with the intercept set to zero was used when there were sufficient data points to generate statistically significant bin sizes for the curve. The power curve fitted POD data evaluated from equal-sized sets (quantile-based discretizations) of test conditions (i.e., release rate) derived from the detection classification results. The number of points per bin (N_p) used for POD data evaluation was constrained to $30 < N_p < 50$ and was based on the quantile-based discretization that produced the highest goodness of fit (R^2) value. The lower bound of 30 data points per bin was selected as sufficient for statistically significant analysis based on the central limit theorem. When the detection classification process yielded insufficient data points for binning as required for the power curve, the model from the remaining functions that minimized the corrected (for small sample sizes) Akaike Information Criterion (AICc) was selected.

$$\text{pod} = -a \cdot x^b \Big|_{\text{a,b are curve fitting parameters}} \tag{2.7}$$

$$\text{pod} = \frac{1}{1 + e^{a+b \cdot x}} \Bigg|_{\text{a,b are curve fitting parameters}} \tag{2.8}$$

$$\text{pod} = 1 - \left(1 + (bx)^c\right)^{-a} \Big|_{\text{a, b, c are curve fitting parameters}} \tag{2.9}$$

$$\text{pod} = e^{-(x/s)^{-a}} \Big|_{\text{s, a are curve fitting parameters}} \tag{2.10}$$

$$\text{pod} = 1 - e^{-(x\lambda)^k}\Big|_{\lambda,\text{ k are curve fitting parameters}} \tag{2.11}$$

$$\text{pod} = \frac{1}{1 + (x/b)^{-a}}\Big|_{\text{a, b are curve fitting parameters}} \tag{2.12}$$

For each POD curve function, the curve-fitting parameters are estimated and optimized using maximum likelihood estimation with the detection classification results as input data. This optimization is performed by minimizing the negative logarithm of the likelihood function, as shown in Equation 2.13.

$$\mathcal{L}(y, p) = -\sum_{i=1}^{n} \left[y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right] \Big|_{\substack{y_i \text{ are binary detection results} \\ p_i \text{ are the candidate POD curve results.}}} \tag{2.13}$$

The original data was bootstrapped 500 times to generate the cloud of curves that visually indicated uncertainty on the main POD curve, with the 95% empirical confidence interval (CI) captured as the 2.5 and 97.5 percentiles. Specifically for power curves (equation 2.7), each POD data point (0 to 1) was bootstrapped by random sampling (p) from a uniform distribution ($0 < p < 1$) $N_p\times$ to obtain the fraction of N_p that are TP detections. A TP detection is adjudged when p $\leq$ POD data point for a bin. This process is iterated 500$\times$ for each bin across all bins to obtain the cloud of POD curves. Five hundred iterations were selected to balance computational efficiency with sufficient data resampling.

Quantification accuracy was assessed using relative error. When applicable, the influence of test conditions (e.g., release rate, wind speed) was analyzed using a Gamma generalized linear model (Equation 2.14) with a log link function, where relative error served as the dependent variable. Given the non-symmetric distribution of quantification relative errors, analysis is conducted for data when relative error is > 0 and < 0, separately. Just like the multivariable logistic regression analysis, test conditions included in the models are those with a variation inflation factor (VIF) $<$ 10, which indicates moderate correlation among selected test conditions.

$$\log(\text{quantification relative error}) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i \Bigg|_{\substack{\beta_i \text{ are regression coefficients} \\ X_i \text{ are independent variables}}} \tag{2.14}$$

Four statistical tests were used to evaluate the statistical significance of the correlation between 2 variables (rank-biserial correlations—a binary and a continuous variable) and the difference between groups (2-sample t-test, Mann-Whitney U, and Kolmogorov-Smirnov tests) at a significance level (p) of 0.05. Monte Carlo analysis was used to extrapolate controlled testing results of continuous monitors in this study to expected emission detection and quantification performance if deployed for a field emissions measurement study by Vaughn et al. [19]. The steps below summarize the Monte Carlo analysis involving N simulations of the field campaign for a selected continuous monitor:

- The relative errors of quantification estimates for selected continuous monitors were grouped based on the three orders of magnitude of controlled release rates the errors were associated with (i.e., (0 - 0.1] kg ch_4/h, (0 - 0.1] kg ch_4/h, and (0 - 0.1] kg ch_4/h). This was intended to minimize sampling bias as the *quantification accuracy of LDAQ solutions depends on the size of emission rate*.

- The POD of detecting a random sample (x) drawn from field measurement data of size L is assessed using the POD curve functions derived for the selected continuous monitors.

- A random value from a uniform distribution (between 0 and 1) was compared with the POD data point (p) evaluated above and adjudged TP detection if the random value was $\leq$ p otherwise, an FN detection.

- if TP detection, a random relative error (e) was selected from the appropriate quantification estimation relative error bin generated with controlled release rates within which x lies. The estimated emission rate for the TP detection was evaluated as $x * (e + 1)$.

- Repeat the above processes for all x) drawn from field measurement data for N $\times$ to generate an array of N columns $\times$ L rows.

- The sum of quantification estimates and the fraction of TP detections for all x in each simulation/iteration were evaluated to mimic facility-level quantification estimation and TP fraction by the continuous monitor during the field campaign. This is repeated for all the N simulations, after which the mean of both variables of all simulations was evaluated and the CI assessed as the 2.5 and 97.5 percentiles.

Chapter 3

EVALUATING THE PERFORMANCE OF CONTINUOUS MONITORING LDAQ SOLUTIONS UNDER SINGLE-BLIND CONTROLLED TESTING[2]

3.1 Overview

Continuous monitoring LDAQ solutions (i.e., continuous monitors) promise to detect fugitive methane emissions across O&G infrastructures/assets sooner than traditional leak surveys like EPA method 21 since continuous monitors operate autonomously and near-continuously measure atmospheric methane concentration. Emissions quantification by continuous monitors has been proposed as the foundation of measurement-based emission inventories [116], given the rapid development and regulatory spotlight on continuous monitors for LDAR programs [136]. The literature review has established the need to characterize the performance of LDAQ solutions through regular and rigorous testing using consensus protocols that support repeatable testing across both time and test facilities.

This chapter discusses the single-blind controlled testing of 16 continuous monitors broadly categorized as point sensor networks and scanning/imaging LDAQ solutions based on the emissions measurement methodology. Test results are evaluated based on the following performance

[2]Chapter 3 is extracted from published articles:
[137] Reprinted with permission from C. Bell, C. Ilonze, A. Duggan, and D. Zimmerle, "Performance of Continuous Emission Monitoring Solutions under a Single-Blind Controlled Testing Protocol," *Environmental Science & Technology*, vol. 57, no. 14, pp. 5794–5805, 2023. DOI: 10.1021/acs.est.2c09235.Copyright 2023 American Chemical Society.
[138] Reprinted with permission from C. Ilonze, E. Emerson, A. Duggan, and D. Zimmerle, "Assessing the progress of the performance of continuous monitoring solutions under a single-blind controlled testing protocol," *Environmental Science & Technology*, vol. 58, no. 25, pp. 10941–10955, 2024. DOI: 10.1021/acs.est.3c08511. Copyright 2024 American Chemical Society.

metrics: (1) detection classification and POD, (2) emission source localization accuracy/precision, (3) emission rate quantification accuracy (single estimate and facility level), and (4) operational factor. The detection classification and POD subsection of the chapter focuses on the emissions detection performance of continuous monitors, the impact of environmental and deployment conditions on emissions detection performance, how it varies by the type of continuous monitor, and how these performances compare against recent federal regulations. The emission source localization accuracy/precision subsection discusses the localization performance of solutions and how it varies with the type of continuous monitor and deployment characteristics. The emission rate quantification accuracy subsection of the chapter investigates the emission rate estimation performance of continuous monitors for individual and facility-level sources and the variation with emission rate size and continuous monitor type. The bootstrapping method is used to evaluate the uncertainty of evaluated facility-level quantification accuracy of solutions. Finally, the operational factor subsection of the chapter explores the operational downtime of continuous monitors and how it compares with U.S. federal regulations.

The chapter also discusses the impact of variation in testing complexity and the likely performance level of continuous monitors in an actual field deployment analyzed using Monte Carlo simulations. The chapter concludes with the implication of the test results on the actual application of continuous monitors for emissions mitigation across the O&G value chain.

3.2 Results and Discussion

3.2.1 Emissions Detection and Probability of Detection

A POD curve relates the probability that an LDAQ solution detects an emission of a given rate as composite performance over all other test conditions, like release duration, emission flow rate, wind speed, temperature, etc., that could affect the POD of an LDAQ solution. A multivariable logistic regression analysis of the impact of these factors on POD over the tested range showed varying statistical significance across all continuous monitors as shown in Table 3.1.

Table 3.1: Table summarizing the multivariable logistic regression analysis evaluating the impact of selected test conditions on the emissions detectability/POD of tested continuous monitors

Solution		Release rate		Release duration		Wind speed		Temperature		Release height	
ID	Type	p value	Odds ratio†	p value	Odds ratio†	p value	Odds ratio†	p value	Odds ratio†	p value	Odds ratio†
A	Point sensor network	0.000	1.001	0.001	1.260	0.000	0.701	0.499	1.008	0.878	0.990
B	Scanning/imaging	0.008	1.000	0.044	1.129	0.000	0.834	0.018	0.972	0.654	0.971
C	Point sensor network	0.000	1.001	0.000	1.492	0.003	0.869	0.233	1.014	0.340	1.052
D	Point sensor network	0.000	1.001	0.000	1.316	0.315	0.960	0.675	1.005	0.056	1.116
E	Point sensor network	0.017	1.001	0.433	1.079	0.000	0.638	0.000	1.103	0.142	0.879
F	Point sensor network	0.000	1.001	0.000	1.259	0.002	0.875	0.280	0.987	0.617	1.032
G	Scanning/imaging	0.026	1.001	0.078	1.500	0.223	0.786	0.394	0.952	0.254	0.694
H	Scanning/imaging	0.399	0.100	0.389	1.094	0.010	0.738	0.416	1.021	0.087	0.762
I	Scanning/imaging	0.002	1.002	0.178	1.497	0.530	0.881	0.358	1.034	0.553	0.930
J	Scanning/imaging	0.000	1.001	0.087	0.521	0.161	0.830	0.869	0.996	0.000	2.157
K	Point sensor network	0.904	0.100	0.849	0.861	0.219	1.634	0.179	0.785	0.246	1.608
L	Scanning/imaging	0.000	1.001	0.000	1.324	0.772	0.983	0.276	0.973	0.303	0.923
N	Point sensor network	0.000	1.001	0.297	1.051	0.000	0.734	0.000	0.935	0.155	1.095
O	Scanning/imaging	0.000	1.000	0.002	1.197	0.000	0.806	0.192	1.020	0.062	1.135
P	Point sensor network	0.000	1.001	0.000	1.202	0.034	0.899	0.583	0.993	0.034	0.885
Q	Point sensor network	0.000	1.001	0.027	1.112	0.000	0.498	0.430	0.988	0.044	0.873

† If the odds ratio is > 1, the higher the value of the test condition, the greater the tendency for a TP detection. If the odds ratio is < 1, the higher the value of the test condition, the greater the tendency for a FN detection. If the odds ratio equals 1, the test condition has no effect on the tendency for a TP or FN detection. Note: The odds ratio is evaluated as the exponential of the coefficients of predictor variables in the multi-variable logistic regression. Values highlighted in yellow are statistically significant ($p < 0.05$).

Results indicate that emission rates significantly ($p < 0.05$) affected the POD of all but 2 continuous monitors (solutions H and K), with other variables affecting only at most $2/3^{rd}$ of tested solutions. In addition, with POD curves as a function of emission rate more useful for practical application in internal or regulatory emissions mitigation programs and as a key metric to model the emission mitigation potential of solutions using emissions modeling tools like FEAST [148] or LDAR-Sim [149], this study presents the POD curves of continuous monitors based on emission rate as shown in Figures 3.1, 3.2, and 3.3. The detection sensitivity of LDAQ solutions is investigated by evaluating the method detection limit (DL90) of each solution, which is the minimum emission rate at which the LDAQ solution, as deployed (i.e., the method), detected emitters 90% of the time over the range of tested meteorological conditions (i.e., the emission rate the solution has 90% POD). The lower the DL90, the more likely an LDAQ solution detects a wider range of emission rates at O&G facilities, although the desired DL90 depends on field application. For LDAQ solutions intended for follow-up investigations and fugitive emissions detections, their DL90 should be significantly lower than those used for facility-level or screening purposes (i.e., identifying large emission events).

Figures 3.1 shows the POD curves of the scanning/imaging continuous monitors tested, while Figures 3.2 and Figures 3.3 show that of point sensor network continuous monitors. Table 3.1 and the figures indicate that 14 of the 16 tested continuous monitors had a statistically significant increased tendency at detection (TP classification) as release rate increased. As indicated by the test protocol, the test center should attempt to map the entire POD curve while testing, however, since many solutions' POD are dependent on the release rate - in some cases very high rate as the DL90 of some solutions shows, this was not always possible and had an impact on the uncertainty of the POD curves (i.e., the cloud of curves) and the associated DL90s. Figures 3.1, 3.2, and 3.3 indicate that the POD curves predicted the DL90s of 13 of the 16 continuous monitors ranging from 1.3 [0.6, 15.6] kg CH_4/h to 64.4 [17.9, NA] kg CH_4/h with that of 5 (with 4 of them categorized as point sensor network) of the 13 continuous monitors within tested range (Table 3.2). For the remaining 3 solutions (G, H, and K), their DL90s could not be estimated due to the limited distribution of TPs

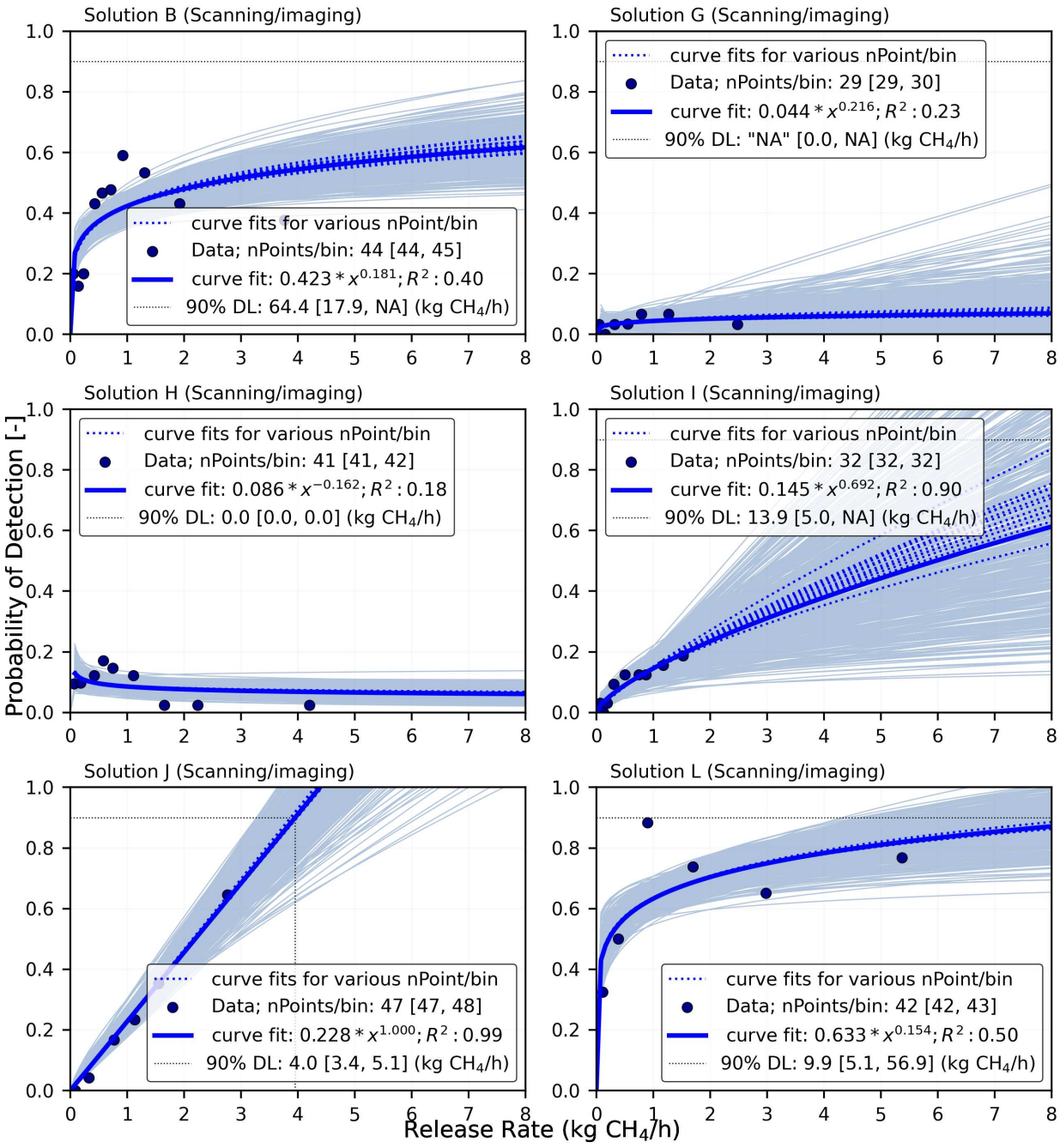


Figure 3.1: The probability of detection (POD) versus emission rate (kg CH_4/h) for solutions B, G, H, I, J, and L (all scanning/imaging) fitted using a power function. The x-axis of each plot is divided into equal-sized bins, with each marker (pod) calculated as the fraction of controlled releases in a bin classified as true positives. Each pod data point is bootstrapped to produce a cloud of curves illustrating associated uncertainty. When the bootstrapping could not evaluate the lower and upper empirical Confidence Interval (CI) on the best estimate of a solution's DL90, they are given as 0 and NA, respectively. Curve fits (dotted colored lines) obtained using other quantile-based discretizations are shown for comparison. The emission rate at which the POD reaches 90% is indicated as the method detection limit (DL90) for each solution. The mean count of points per bin along with the minimum and maximum counts across all bins is also shown in the legend of the figures.

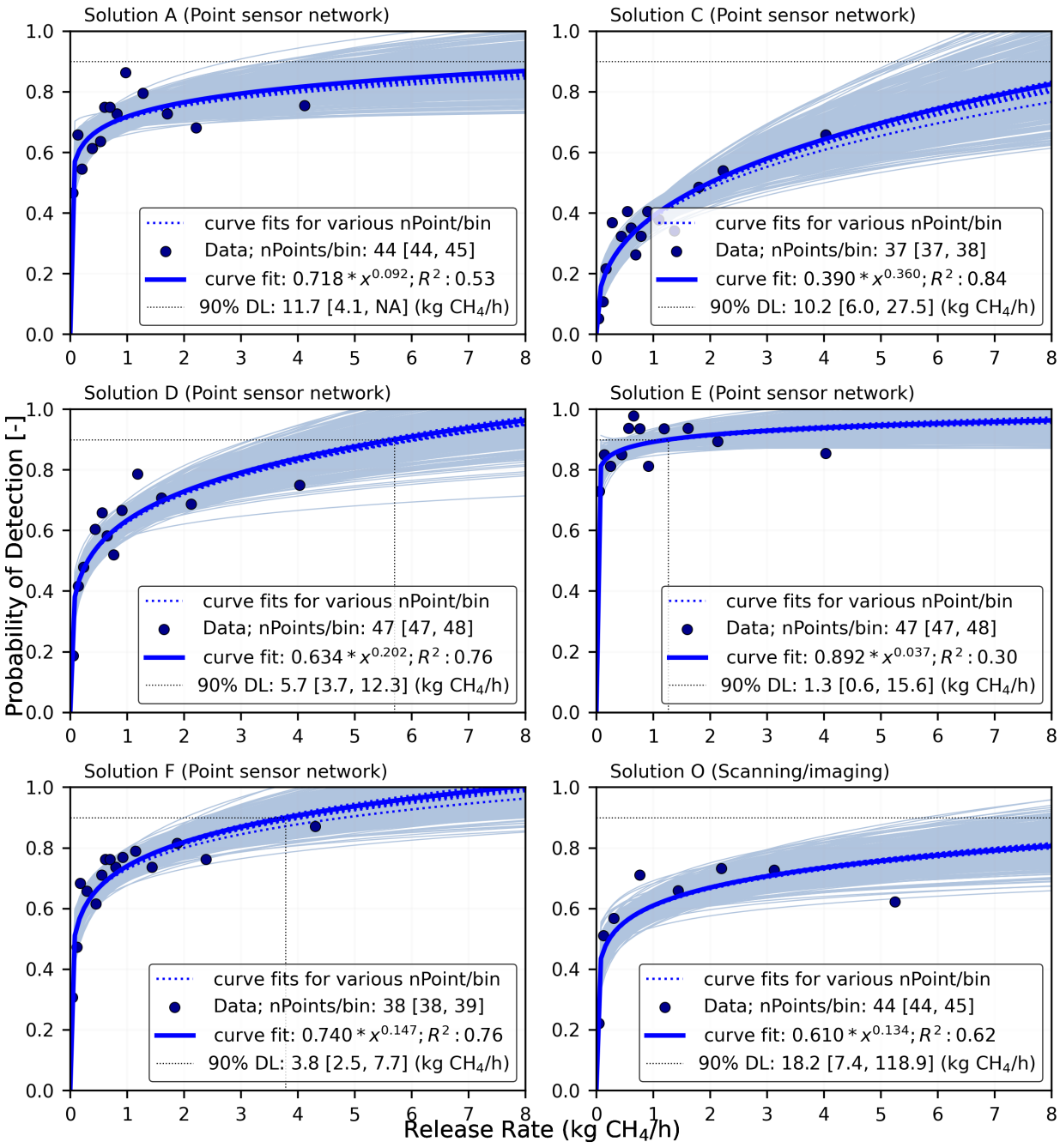


Figure 3.2: The probability of detection (POD) versus emission rate (kg CH_4/h) for LDAQ solutions A, C, D, E, F (all point sensor network), and O (a scanning/imaging) fitted using a power function. The x-axis of each plot is divided into equal-sized bins, with each marker (pod) calculated as the fraction of controlled releases in a bin classified as true positives. Each pod data point is bootstrapped to produce a cloud of curves illustrating associated uncertainty. When the bootstrapping could not evaluate the lower and upper empirical Confidence Interval (CI) on the best estimate of a solution's DL90, they are given as 0 and NA, respectively. Curve fits (dotted colored lines) obtained using other quantile-based discretizations are shown for comparison. The emission rate at which the POD reaches 90% is indicated as the method detection limit (DL90) for each solution. The mean count of points per bin along with the minimum and maximum counts across all bins is also shown in the legend of the figures.

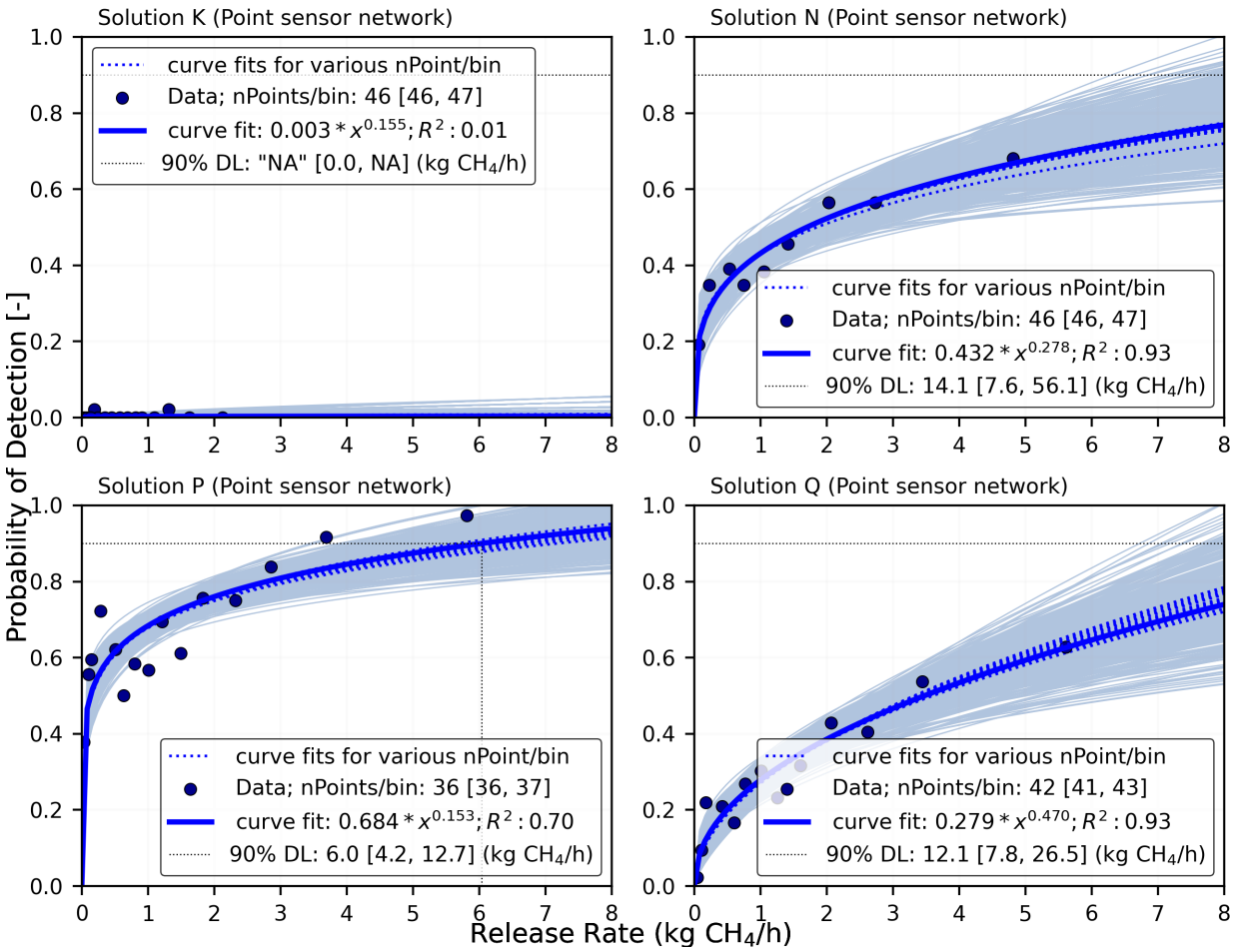


Figure 3.3: The probability of detection (POD) versus emission rate (kg CH_4/h) for LDAQ solutions K, N, O, and Q (all point sensor network) fitted using a power function. The x-axis of each plot is divided into equal-sized bins, with each marker (pod) calculated as the fraction of controlled releases in a bin classified as true positives. Each pod data point is bootstrapped to produce a cloud of curves illustrating associated uncertainty. When the bootstrapping could not evaluate the lower and upper empirical Confidence Interval (CI) on the best estimate of a solution's DL90, they are given as 0 and NA, respectively. Curve fits (dotted colored lines) obtained using other quantile-based discretizations are shown for comparison. The emission rate at which the POD reaches 90% is indicated as the method detection limit (DL90) for each solution. The mean count of points per bin along with the minimum and maximum counts across all bins is also shown in the legend of the figures.

compared to FNs either because of poor detection performance or the majority of the controlled releases in this study substantially below the detection sensitivity of the continuous monitors. This level of variability highlights reasons why the detection performance of continuous monitors must be understood before being used for field measurement studies or LDAR deployment.

Table 3.2: Summary of the detection performance of tested continuous monitors sorted in order of increasing DL90s predicted by the solutions.

Solution		Count		FPF (%)[†]				Method Detection Limit	
ID	Type	Controlled Release	Detection Reports	All	No Controlled Release	Excess Detections	FNF (%)	DL90[‡] (kg CH_4/h)	Within Tested Rates
E	Point sensor network	570	2875	82.6	24.7	75.3	12.3	1.3 [0.6, 15.6]	Yes
F	Point sensor network	574	516	22.5	39.7	60.3	30.3	3.8 [2.5, 7.7]	Yes
J	Scanning/Imaging	287	69	0.0	0.0	0.0	76.0	4.0 [3.4, 5.1]	Yes
D	Point sensor network	574	376	10.4	79.5	20.5	41.3	5.7 [3.7, 12.3]	Yes
P	Point sensor network	547	423	13.2	23.2	76.8	32.9	6.0 [4.2, 12.7]	Yes
L	Scanning/Imaging	256	254	35.0	95.5	4.5	35.5	9.9 [5.1, 56.9]	No
C	Point sensor network	560	246	20.7	68.6	31.4	65.2	10.2 [6.0, 27.5]	No
A	Point sensor network	574	986	59.8	26.9	73.1	31.0	11.7 [4.1, NA]	No
Q	Point sensor network	547	260	38.1	21.2	78.8	70.6	12.1 [7.8, 26.5]	No
I	Scanning/Imaging	352	34	17.6	83.3	16.7	92.0	13.9 [5.0, NA]	No
N	Point sensor network	417	223	18.4	29.3	70.7	56.4	14.1 [7.6, 56.1]	No
O	Scanning/Imaging	357	324	34.6	33.0	67.0	40.6	18.2 [7.4, 118.9]	No
B	Scanning/Imaging	445	250	31.2	61.5	38.5	61.3	64.4 [17.9, NA]	No
G	Scanning/Imaging	209	20	60.0	91.7	8.3	96.2	NA	No
H	Scanning/Imaging	371	56	39.3	95.5	4.5	90.8	NA	No
K	Point sensor network	748	2	0.0	0.0	0.0	99.7	NA	No

[†] **All** is the percentage of all detections classified as FP based on the protocol. **No controlled release** is the fraction of all false positives due to detection reports sent when no ongoing controlled release exists. **Excess TP Detections** is the fraction of all FPs due to excess detections identifying controlled releases already paired as a new and/or a different emitter.

[‡] When DL90s cannot be evaluated, they are given as NA. Similarly, the lower and upper empirical 95% CI on DL90 are given as 0 and NA, respectively, when they cannot be evaluated.

A continuous monitor may be designed to provide only full facility-level monitoring or monitor a specific area of a facility for large emissions (a common mode for imaging solutions). This setup might require tuning the solution's detection sensitivity high (i.e., higher DL90 values) to ignore small-sized emissions, unlike another continuous monitor that might focus on fugitive emissions (which typically falls within tested rates in this study). However, considering only the POD metric presents a limited view of practical detection performance as some continuous monitors trade-off detection sensitivity with false positive fraction (FPF) and FNF, which has cost implications depending on the field application of the continuous monitors.

Table 3.2 shows that the FPF and FNF of tested continuous monitors ranged from 0.0% (solutions J and K) to 82.6% (solutions E), and 12.3% (solutions E) to 99.7% (solutions K), respectively. Of the 5 continuous monitors with DL90s within tested release rate ranges ($\leq$ 6.0 [4.2, 12.7] kg CH_4/h) and hence the lowest DL90s in the study, 4 solutions had 0% $\leq$ FPF $\leq$ 23% and 30% $\leq$ FNF $\leq$ 41%. Results indicate that tested continuous monitors (e.g., solution E) struggled in balancing high detection sensitivity (low DL90) with relatively low FPF and FNF with only 1 solution attempting to balance the 3 detection metrics. Low DL90 may present issues in field deployment if FPF is high as it may force unacceptably high follow-up costs as close inspection with OGI camera(s) investigating false detection alerts can be expensive and time-consuming. Given that solutions' algorithms transform raw sensor measurement to detection alerts/reports, high FPF at low DL90 suggests challenges in separating emissions signal from ambient methane concentration. Continuous monitors with relatively higher DL90s (i.e., outside the tested release rate range), FPF (all but 1 solution $\geq$ 25%), and FN rates ($\geq$ 45%) likely indicate struggles at emissions detection due to poor sensors, ineffective algorithms/data analytics, or low readiness-level of the continuous monitor for field deployment. In general, at a minimum detection threshold of at most 0.40 kg CH_4/h proposed as the detection threshold requirement for the use of continuous monitors for EPA regulatory-approved emissions monitoring [136], 7 of the 16 tested continuous monitors had $\geq$ 50% POD but less than the 99% confidence as defined by the U.S. EPA [150] which suggests

likely immaturity of the detection performance of continuous monitors for regulatory-approved emissions mitigation programs as at the time of this study.

A review of the percentage of FPs due to excess detection reports (0–77% with that of 7 of the 16 solutions ≥50%; 6 of them were point sensor network continuous monitors) as shown in Table 3.2 and Figure 3.4 suggests that if the intended application of several continuous monitors was to correctly alert operators of ongoing emissions with less priority on what is emitting and the number of emitters, then these continuous monitors would have much lower FPF than assessed by the protocol's detection classification framework.

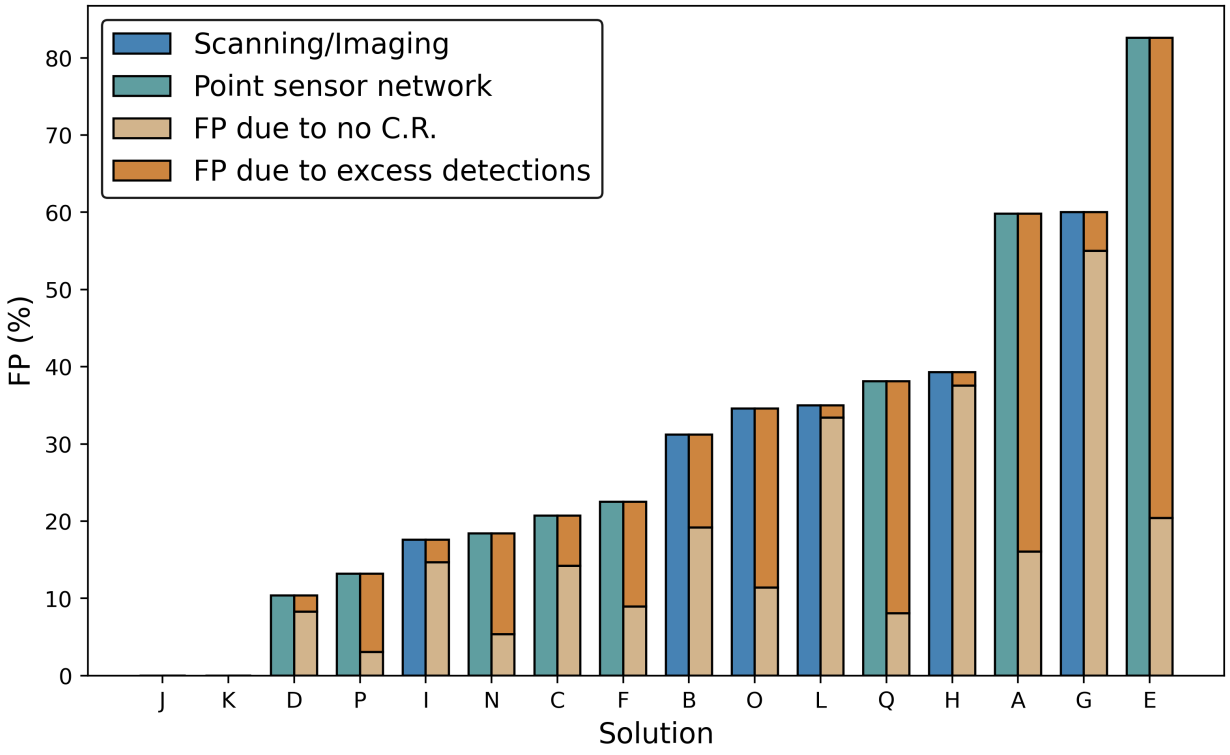


Figure 3.4: This figure shows the categorization of the FPs evaluated in this study for all tested continuous monitors arranged in increasing order of FPF. FPF due to no controlled release is the fraction of all FPs due to detection reports sent when no ongoing controlled release exists. FPF due to excess TP detections is the fraction of all FPs due to excess detections identifying controlled releases already paired as a new and/or a different emitter.

The high contribution of excess detection reports to the FPF of point sensor network continuous monitors highlights likely wide uncertainty and relatively lower precision associated with

attributing emissions to the correct source (i.e., source attribution confusion) compared to scanning/imaging continuous monitors (i.e., $< 20\%$ of the FPF of 5 of 7 of them were due to excess detection reports). Results also showed no statistically significant difference in the detection performance (FPF, FNF, and DL90s) of the point sensor network and scanning/imaging continuous monitors. While deploying fewer sensors can result in a drop in the performance of continuous monitors, especially point sensors (i.e., increases the chances of sensors being downwind of emissions), a Spearman's rank correlation analysis showed that the count of sensors did not statistically affect assessed FNFs, FPFs, or DL90s ($p > 0.5$), as solutions that deployed more sensors did not always have a lower FNF, FPF, or DL90 compared to those that installed fewer sensors. Aside from the difference in the type and quality of sensors and proprietary algorithms, which can heavily influence the performance of individual sensor elements, this observation indicates that over-deployment of sensors yields only a marginal increase in detection performance. Note that the reporting process used in this study did not support continuous monitors attributing specific detections to specific sensor(s), which impedes any systematic assessment of sensor over-deployment.

To investigate the impact of other test conditions on the POD continuous monitors, Table 3.1 indicates most continuous monitors (11 solutions), especially point sensor networks, were significantly more likely ($p < 0.05$) to not detect emissions as windspeed increased. This observation could be attributed to a few reasons. First, gas plumes are narrower when winds are higher—faster vertical transport reduces the time for the plume to widen horizontally, producing a narrower plume more likely to miss all sensors for a point sensor network LDAQ solution. This is particularly challenging for slower sensors, such as catalytic or thermal sensors. Second, turbulent emissions dispersion could inhibit stable/steady measurements needed to infer detections. Similarly, the tendency to detect emissions statistically increased ($p < 0.05$) with release duration for the majority (9 solutions) of the tested continuous monitors, mostly point sensor networks. Longer release durations afford point sensor networksm longer averaging times with which to make detections. Other test conditions, like ambient temperature and controlled release height, each affected only 3 continuous monitors (2 point sensor networks and 1 scanning/imaging).

The key finding from these results is that tested continuous monitors had varying performance levels, indicating their different field-readiness levels and designed field applications. Assuming a continuous monitor is properly deployed for its intended purpose, true detection performance should be judged by the balance between DL90, FNF, and FPF. Also, the performance of scanning/imaging continuous monitors is generally less likely than point sensor network continuous monitors to be affected by emissions and environmental conditions.

3.2.2 Emissions Source Localization

Emission source localization level is often dictated by the target market of an LDAQ solution: The solution's algorithm and data analytics may vary depending upon the intended field application of the solution. Some continuous monitors provide facility-level emissions signals with no capability to localize individual emitters within the facility, while others prioritize the localization of specific emission sources. The test protocol in this study required continuous monitors to report the equipment unit for any identified emission sources, and the protocol's classification framework evaluated localization performance based on how close identified emission location was for TP detections relative to the emitting equipment unit.

Data indicate that 7 of 16 continuous monitors, including 5 scanning/imaging solutions, achieved high equipment unit localization accuracy and precision, as shown in Figure 3.5. Except for solution J, scanning/imaging solutions deployed the fewest sensors per monitored test center surface area (i.e., had the lowest sensor densities), as summarized in Table 3.3. In other words, scanning/imaging continuous monitors showed a higher tendency to successfully localize emitters to the correct equipment unit, despite installing a fewer number of sensors than point sensor network continuous monitors. Scanning and imaging solutions survey, monitor, and *visualize* emissions within their field of view for a given period before changing view, hence minimizing monitoring "blind spots" provided the duration of the emitter is sufficiently long enough that the imager views the emission location before it ends. For operators deploying continuous monitors at multiple (sometimes in the hundreds) and bigger facilities, better localization of emitters reduces the

time required for diagnostic OGI camera surveys by focusing surveys on fewer equipment units; a definite cost-saving measure for operators.

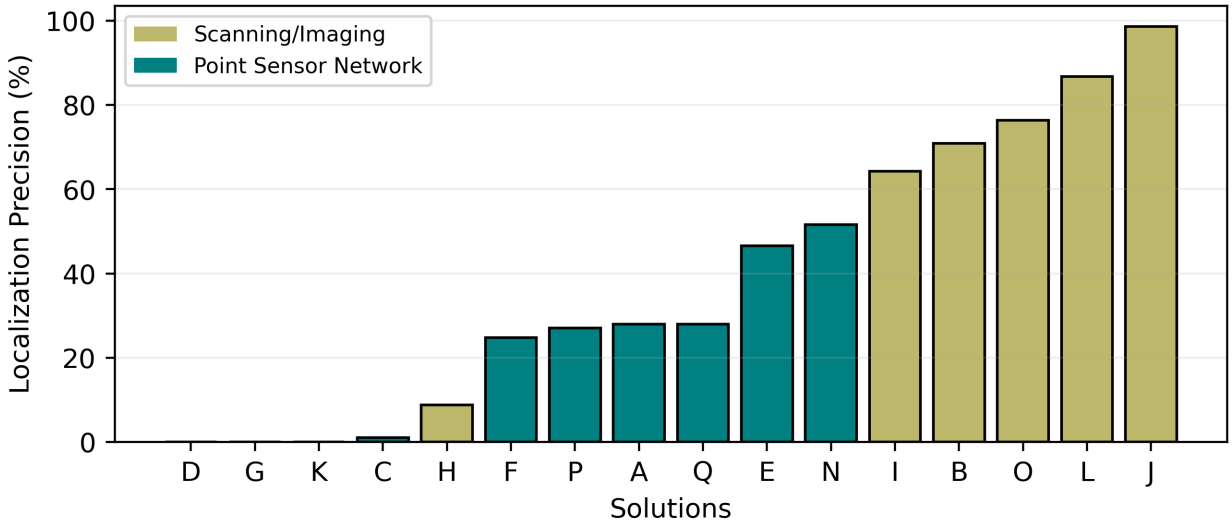


Figure 3.5: The figure presents the equipment unit-level localization precision of the tested continuous monitors, arranged in ascending order of localization precision.

Note that this functionality was not assessed for solutions that provided only facility-scale detection and localization. For solutions that primarily localized emissions at the equipment group and facility levels, a fundamental limitation of the test protocol—not assessing performance based on the LDAQ solution's defined field use case—makes it difficult to determine whether the observed performance results from their intended field application or an ineffective emission source localization algorithm.

Unlike the point sensor network continuous monitors, a change in sensor density did not produce a statistically significant variation (Spearman rank correlation; $p < 0.05$) in equipment unit localization by scanning and imaging solutions. These data suggest that an increase in sensor density likely improves localization performance of point sensor networks, but does not impact scanning and imaging solutions.

Table 3.3: Summary of emission source localization (equipment unit) precision and accuracy for all participating solutions arranged in decreasing localization precision equipment unit level.

Solution				Source Localization (Equipment Unit)					
				Precision (%)			Accuracy (%)		
ID	Category	Sensor Density (sensors/m^2)	Count of TPs	Unit Level	Group Level	Facility Level	Unit	Group Level	Facility Level
J	Scanning/Imaging	0.00416	69	98.6	0.0	1.4	98.6	98.6	100.0
L	Scanning/Imaging	0.00012	165	86.7	10.9	2.4	56.3	63.4	65.0
O	Scanning/Imaging	0.00012	212	76.4	12.7	10.9	50.0	58.3	65.4
B	Scanning/Imaging	0.00012	172	70.9	15.7	13.4	48.8	59.6	68.8
I	Scanning/Imaging	0.00012	28	64.3	14.3	21.4	52.9	64.7	82.4
N	Point sensor network	0.00213	182	51.6	41.8	6.6	42.2	76.2	81.6
E	Point sensor network	0.00189	500	46.6	41.6	11.8	8.1	15.3	17.4
Q	Point sensor network	0.00154	161	28.0	54.0	18.0	17.3	50.8	61.9
A	Point sensor network	0.00095	396	28.0	39.4	32.6	11.3	27.1	40.2
P	Point sensor network	0.00071	367	27.0	56.9	16.1	23.4	72.8	86.8
F	Point sensor network	0.00095	400	24.8	50.2	25.0	19.2	58.1	77.5
H	Scanning/Imaging	0.00012	34	8.8	50.0	41.2	5.4	35.7	60.7
C	Point sensor network	0.00071	195	1.0	0.5	98.5	0.8	1.2	79.3
K	Point sensor network	0.00083	2	0.0	50.0	50.0	0.0	50.0	100.0
G	Scanning/Imaging	0.00012	8	0.0	25.0	75.0	0.0	10.0	40.0
D	Point sensor network	0.00095	337	0.0	52.8	47.2	0.0	47.3	89.6

3.2.3 Emission Rate Quantification Accuracy

Table 3.4 summarizes the impact (evaluated using a generalized linear model) of selected test conditions on the quantification relative error of continuous monitors. Results indicate that release rates significantly ($p < 0.05$) affected the quantification relative error of the majority (all solutions for quantification relative error > 0; 5 solutions for quantification relative error < 0) of continuous monitors, with other variables affecting only at most $1/3^{rd}$ of tested solutions. The dependence of quantification accuracy on the controlled release emission rate is a key variable to understand field performance and is typically the parameter of interest in modeling and regulatory program analysis. Figures 3.6 and 3.7 show the zero-intercept plots of estimated emission rate against controlled release rate for TP detections for all the 9 continuous monitors that tested quantification capability (solution G was excluded due to poor data quality). Although informative, the results and data in Figures 3.6 and 3.7 should be extrapolated and applied to real O&G facilities with caution for three reasons. First, residual errors are strongly heteroskedastic (nonuniform variance of the residual), which makes the regression analysis highly uncertain. Second, the emission rates used in this study were manipulated to map the entire probability of detection curve and are therefore not reflective of the mix of emission rates likely to be encountered at an O&G facility. Finally, many decisions dependent on emission estimates utilize a much smaller number of estimates than used to fit the linear regression in these figures. In practice, operators must respond to detections quickly, requiring them to make decisions based upon a single or a small number of observations. In these practical cases, the accuracy of a single estimates is more indicative of uncertainty in emission rate than the regression presented in the figure.

While Table 3.4 shows a general strong systematic bias of decrease in quantification relative error (from overestimation to underestimation) as release rate increased, data also indicated varying accuracy levels across individual continuous monitors and their types based on the R^2 and fraction of estimates within a quantification factor of 3 $(-67\%|\frac{1}{3}\times,\ +200\%|3\times)$ of actual release rates.

Table 3.4: Table summarizing the gamma generalized linear model with a log-link analysis evaluating the impact of selected test conditions on the quantification accuracy of tested continuous monitors

Solution		Release rate		Release duration		Wind speed		Temperature		Intercept	
ID	**Type**	**p value**	**Odds ratio†**	**p value**	**Odds ratio†**	**p value**	**Odds ratio†**	**p value**	**Odds ratio†**	**p value**	**Odds ratio†**
For quantification relative error > 0											
A	Point sensor network	0.000	0.424	0.014	0.730	0.000	1.759	0.150	1.102	0.000	368.975
B	Scanning/imaging	0.000	0.363	0.277	0.727	0.043	1.695	0.593	0.897	0.000	1076.430
C	Point sensor network	0.000	0.419	0.248	0.707	0.919	1.039	0.062	0.632	0.000	2598.180
D	Point sensor network	0.000	0.527	0.167	0.592	0.954	1.018	0.077	1.600	0.001	68.007
E	Point sensor network	0.000	0.382	0.007	1.240	0.039	1.154	0.460	0.967	0.000	1438.540
F	Point sensor network	0.000	0.395	0.281	0.799	0.562	0.886	0.676	0.951	0.000	1151.640
L	Scanning/imaging	0.003	0.626	0.384	0.805	0.390	1.192	0.793	1.031	0.005	46.856
N	Point sensor network	0.000	0.380	0.090	0.715	0.035	1.749	0.395	1.102	0.000	3359.670
O	Scanning/imaging	0.000	0.290	0.813	1.046	0.308	1.215	0.270	0.912	0.000	11811.000
For quantification relative error < 0											
A	Point sensor network	0.000	0.664	0.781	0.971	0.730	1.043	0.625	1.042	0.000	45.113
B	Scanning/imaging	0.995	1.000	0.943	0.988	0.267	0.829	0.154	0.899	0.326	2.964
C	Point sensor network	0.008	0.611	0.197	0.782	0.861	1.020	0.157	1.101	0.003	79.193
D	Point sensor network	0.021	0.893	0.484	0.952	0.696	0.976	0.017	1.086	0.002	3.161
E	Point sensor network	0.658	1.040	0.274	0.877	0.029	0.782	0.022	0.795	0.094	3.177
F	Point sensor network	0.000	0.602	0.311	1.109	0.000	0.702	0.667	1.026	0.000	157.091
L	Scanning/imaging	0.062	1.6266	0.132	1.576	0.399	1.273	0.607	0.893	0.097	0.037
N	Point sensor network	0.231	0.830	0.920	0.985	0.540	0.855	0.465	1.052	0.046	13.962
O	Scanning/imaging	0.000	0.142	0.518	1.134	0.759	0.957	0.713	1.028	0.000	3.9E+08

† The odds ratio is the exponentiated coefficient of the GLM model. If the odds ratio > 1, then a unit increase in the independent variable will increase the expected relative error by (odds ratio - 1) $\times$ 100%. If the odds ratio < 1, then a unit increase in the independent variable will decrease the expected relative error by (1 - odds ratio) $\times$ 100%. Note: Odds ratio is evaluated as the exponential of the coefficients of predictor variables in the generalized linear model. Values highlighted in yellow are statistically significant ($p < 0.05$).

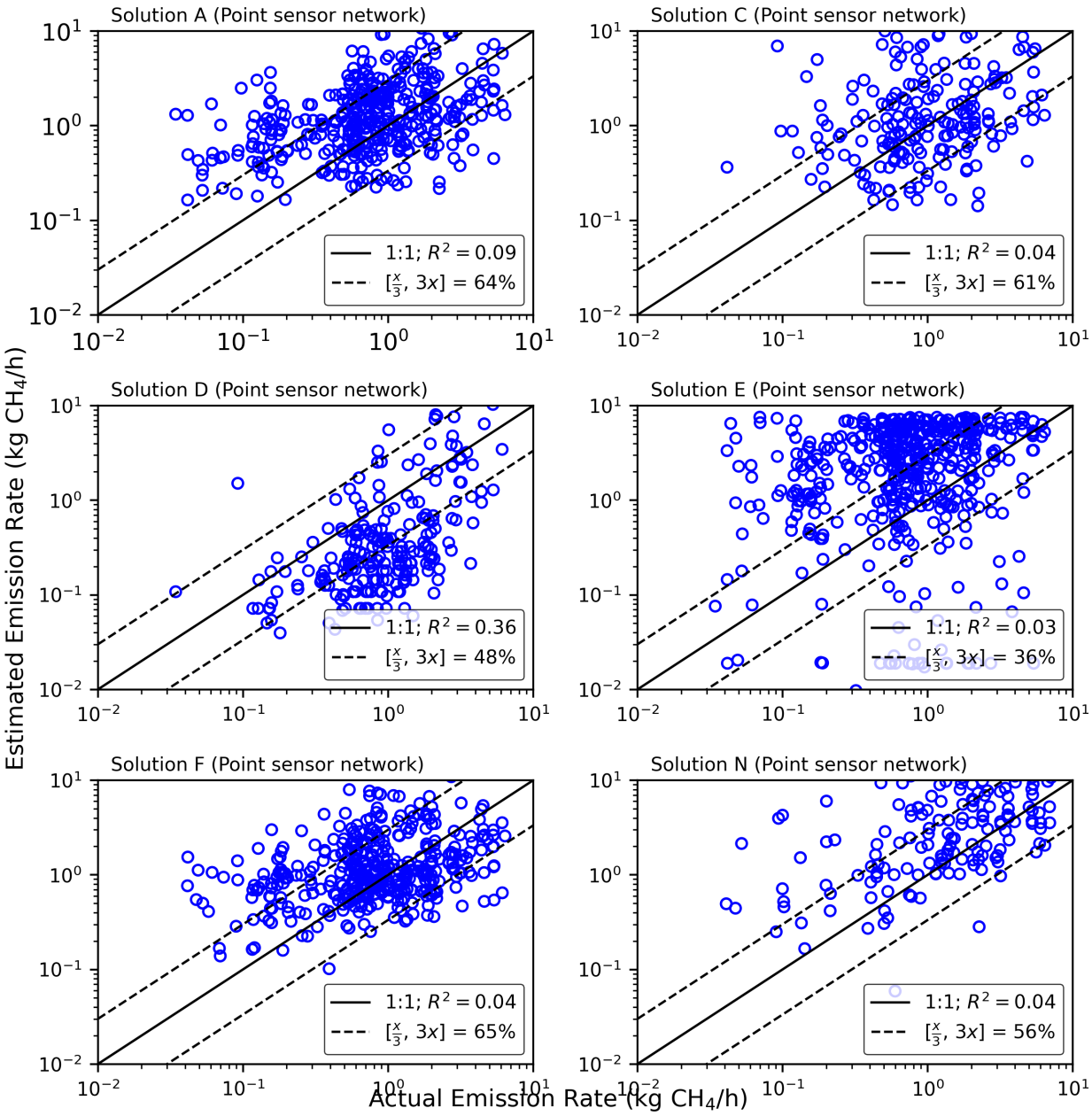


Figure 3.6: Quantification accuracy of all point sensor network continuous monitors. Reported emission rates are plotted against actual (or controlled) emission rates for all point sensors network continuous monitors that tested quantification capability. The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%). The R^2 illustrates the correlation between the reported and actual emission rates.

Tested continuous monitors (except solutions E and L) did not consistently improve quantification accuracy with increasing emission rates. These solutions are likely to substantially underestimate

emissions, even for emission rates exceeding those tested in this study. Overall, scanning/imaging continuous monitors performed comparatively better than point sensor networks, even though all continuous monitors showed wide single estimate quantification uncertainty ($\leq \frac{\times}{30}$ to $\geq 40\times$) as shown in Table 3.5. Table 3.5 breaks down quantification performance of continuous monitors into controlled release rate ranges of (0.1 - 1] kg CH_4/h and $>$1 kg CH_4/h, respectively.

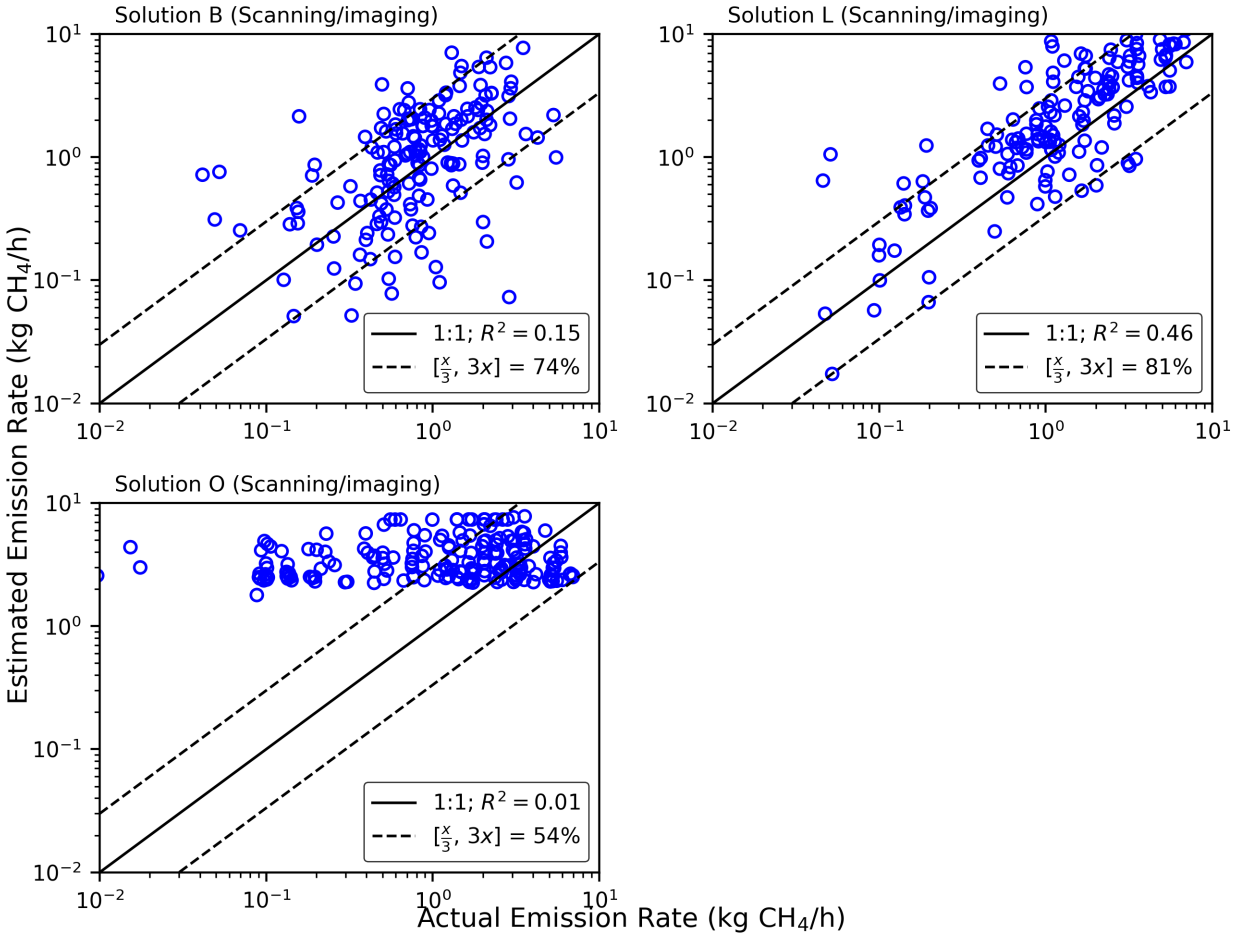


Figure 3.7: Quantification accuracy of all scanning/imaging continuous monitors. Reported emission rates are plotted against actual (or controlled) emission rates for solutions that tested quantification capability. The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%). The R^2 illustrates the correlation between the reported and actual emission rates.

Table 3.5: Summary of single-estimate quantification for continuous monitors along with their 95% empirical confidence limits arranged in increasing the percentage of measurements within a factor of 3.

Solution		Estimates within a Factor of 3			Relative Quantification Error (%)					
		$(-67\%\|_{\frac{1}{3}}\times,\ +200\%\|3\times)$			Controlled Release (0.1 – 1] kg CH_4/h			Controlled Release > 1 kg CH_4/h		
ID	**Type**	**All**	**(0.1 – 1]**	**> 1**	**Mean**	**Median**	**95% C.I.**	**Mean**	**Median**	**95% C.I.**
L	Scanning/Imaging	81	84	81	126.2	91.2	[-49.5, 546.6]	90.4	50.0	[-70.7, 402.2]
B	Scanning/Imaging	74	76	80	74.6	39.5	[-81.1, 343.2]	41.9	25.3	[-90.2, 268.8]
F	Point sensor network	65	62	75	202.2	110.9	[-39.7, 933.2]	9.2	-40.5	[-82.5, 373.6]
A	Point sensor network	64	65	73	211.3	134.2	[-60.9, 946.8]	27.1	-24.2	[-85.6, 338.5]
C	Point sensor network	61	58	67	268.9	76.2	[-73.9, 1875.0]	17.2	-30.8	[-88.6, 369.3]
N	Point sensor network	56	48	62	1036.6	212.8	[-36.2, 2900.9]	256.0	72.0	[-68.0, 1671.6]
O	Scanning/Imaging	54	1	86	1751.2	1074.9	[235.9, 4071.5]	73.4	58.8	[-60.4, 347.0]
D	Point sensor network	48	60	34	-43.0	-60.1	[-92.6, 141.4]	-40.0	-77.0	[-99.9, 242.4]
E	Point sensor network	39	28	64	584.9	411.0	[-96.7, 2078.7]	93.1	51.4	[-99.1, 448.3]

While release rates in the range (0.1 - 1] kg CH_4/h roughly represent equipment component leak rates typically identified through OGI camera surveys [20, 24, 65], rates in the range $>$1 kg CH_4/h represent relatively larger leak rates due to process upset conditions and failures at upstream facilities [11, 65]. For the tested release rate range, data indicate less accurate quantification performance at small release rates (0.1 - 1] kg CH_4/h compared to higher rates $>$1 kg CH_4/h. For rates in (0.1 - 1] kg CH_4/h, quantification uncertainty on estimates of single sources ranged from $\leq \frac{\times}{30}$ to $\geq 40\times$ with 5 of 9 solutions with a quantification fraction within a factor of 3 $\geq$ 60%. Nearly all emission sources from facilities with little venting and no atmospheric tanks likely fall within this emission rate range. For rates $>$1 kg CH_4/h, quantification uncertainty on estimates of single sources ranged from $\leq \frac{\times}{10}$ to $\geq 17\times$ with 8 of 9 solutions with a fraction within a quantification factor of 3 $\geq$ 60%. The quantification results for release rates within (0.1–1] kg CH_4/h are unsurprising given the range of DL90 values evaluated for continuous monitors in this study. A minimum DL90 of 1.3 [0.6, 15.6] kg CH_4/h suggests that the tested continuous monitors were likely to estimate rates within this range poorly—if detected at all—at uncertainty levels dependent on the effectiveness of their data analytics. In general, for all release rates, uncertainty in single estimates ranged from $\leq \frac{\times}{50}$ to $\geq 46\times$, a problematic outcome if extrapolated to field applications. For instance, overestimating a relatively large emission (e.g., a leak rate of 7.1 kg CH_4/h—the maximum rate tested in this study) by 17$\times$ could result in a misleading alert suggesting emissions at the scale of a super-emitter ($\geq$100 kg CH_4/h). Similarly, grossly underestimating a super-emitter could undermine mitigation efforts that rely on these estimates to initiate remediation actions, potentially leading to adverse environmental consequences.

To evaluate each continuous monitor's aggregate/facility-level quantification accuracy, the sum of the release rates of all valid controlled releases during each continuous monitor's testing was compared to the sum of all valid detection reports sent within the same period. Facility-level quantification relative error was assessed for each solution with the associated CI on the error obtained using Monte Carlo simulation and bootstrapping. Figure 3.8 is an error bar plot showing facility-level quantification relative errors (actual and simulated mean) for continuous monitors

aggregated over the test duration. The whiskers represent uncertainty on the relative errors, while the grey shaded region shows the emission rate estimation range within a quantification factor of 3 $(-67\%|\frac{1}{3}\times,\ +200\%|3\times)$ of actual release rates. Results show that at the facility level, 7 of 9 continuous monitors estimated facility-level emissions within a factor of 2 $(-50\%|\frac{1}{2}\times,\ +100\%|2\times)$ with their respective simulated lower and upper limits within a factor of ≈3.

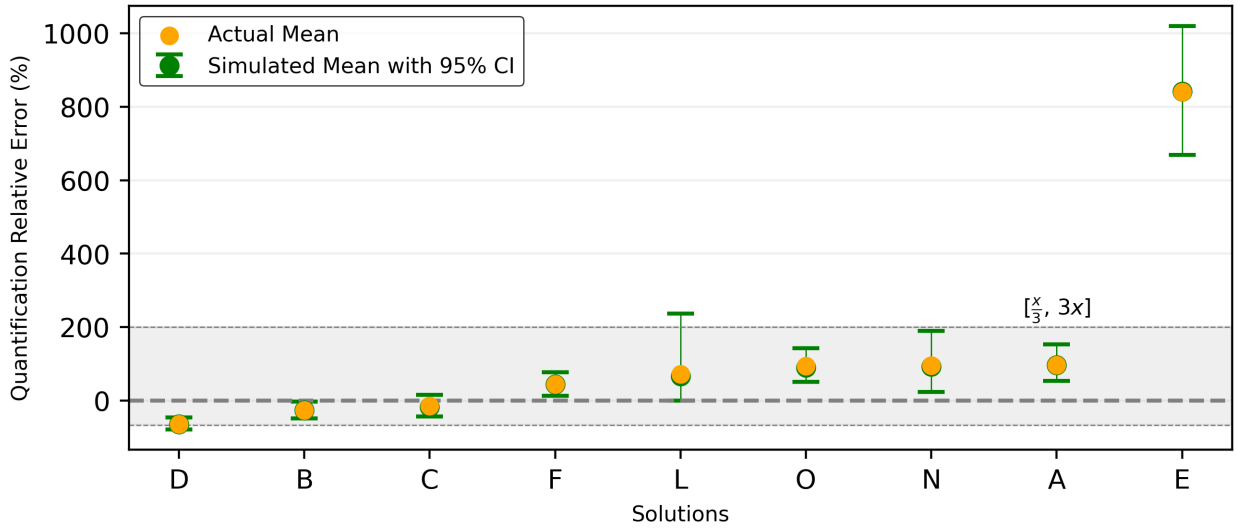


Figure 3.8: Summarizes the facility-level relative error for each continuous monitor arranged in increasing order of mean quantification relative error. Facility-level relative error is bootstrapped to estimate the uncertainty on the actual error. Markers represent bootstrapped facility-level mean relative error (green) and the actual facility-level relative error (orange), respectively. Whiskers represent the 95% CI on the bootstrapped mean relative error.

Given the increasing interest in facility-level quantification, as inferred from the U.S.EPA final rule [136] (NSPS OOOOb), this data indicates that continuous monitors are likely to provide facility-level emissions estimation with higher accuracy and narrower uncertainty than single estimates.

As shown in Table 3.4, emission rate is just one of several independent variables that may have an appreciable impact on the quantification performance of continuous monitors. While release duration was only statistically significant to the quantification performance of 2 solutions, the relative error of most of the tested continuous monitors tended to decrease (i.e., shift from overestimation to underestimation) with increasing release duration. This observation could be attributed to issues

with LDAQ solutions' data analytics given that plume inversion models (e.g., Gaussian plume, Lagrangian) used by several continuous monitors for emissions quantification typically benefit from longer release durations, allowing for extended averaging of ambient data to improve the accuracy of emission estimates [120]. Data also show a decrease in quantification accuracy as wind speed increased for the majority of continuous monitors, with this effect being statistically significant for 5 out of 9 solutions. This result is likely due to the impact of high wind speed on plume shape during transport, making it more challenging for sensors to detect and quantify such gas plumes accurately.

A key takeaway is that most continuous monitors exhibited a statistically significant decrease in quantification relative error as the emission rate increased, following a trend from overestimation to underestimation across the tested range. Only two continuous monitors (E and L) demonstrated improved quantification accuracy with increasing emission rates. There was no clear distinction in the impact of test conditions on the quantification performance of point sensor networks versus scanning/imaging continuous monitors. Overall, quantification algorithms and the associated inversion models of LDAQ solutions need refinement to reduce uncertainties and improve the accuracy of emission estimates.

3.2.4 Operational Factors

The U.S. EPA, in the NSPS OOOOb final rule on the use of continuous monitors for regulatory-compliant emissions mitigation, defines operational downtime as the period during which a continuous monitor fails to collect or transmit data [136]. The rule stipulates a rolling 12-month average operational downtime of $\leq$10% (operational factor $\geq$90%) as a requirement for adopting continuous monitors in regulatory-compliant emissions mitigation programs. Study findings showed that the operational factor of continuous monitors ranged from 0.45 to 1.00, with 10 of the 16 continuous monitors operational at least 90% of the deployment time. If these test data are indicative of performance in real O&G facilities, then the majority of tested continuous monitors are likely robust enough to operate near-autonomously for extended periods without failure.

3.3 Performance Variation with Testing Complexity

Testing for 24 hours with one to six simultaneous controlled releases was designed to increase testing complexity (e.g., different meteorological conditions and atmospheric stability), mimic emissions behavior at real O&G facilities, and assess continuous monitors' ability to isolate individual emission sources.

An investigation into the detection performance of solutions based on whether an experiment contained single or multiple controlled releases showed that, unlike scanning/imaging solutions, most ($2/3^{rd}$) of the point sensor network continuous monitors exhibited a statistically significant ($p<0.05$) higher FNF during multiple controlled release experiments compared to single controlled release experiments (Table B.1). Similarly, the FNFs of continuous monitors for controlled releases conducted during the daytime (between sunrise and sunset) were generally higher than those performed at nighttime, although this difference was not statistically significant (Table B.2).

When quantification performance was evaluated, results indicated that solutions generally exhibited higher quantification accuracy during experiments with a single controlled release (Figures B.1 and B.2) compared to experiments with multiple controlled releases (Figures B.5 and B.6). Similarly, quantification performance was generally better at nighttime (Figures B.3 and B.4) than during the daytime (Figures B.7 and B.8). However, these differences were not statistically significant, and the sample size for nighttime experiments was substantially smaller than that for daytime experiments.

The comparatively better detection and quantification performance for experiments with a single controlled release underscores the challenges associated with isolating and characterizing individual emission sources when multiple, simultaneous emission events occur, particularly for point sensor network solutions. This issue may arise from cross-source/plume interferences that lead to confusion between emission sources by LDAQ solution algorithm. For instance, when plumes from two or more separate sources merge as they intersect with a sensor downwind, it becomes difficult to distinguish the sources accurately. Detection and quantification performance was gen-

erally better at night, likely due to weaker turbulent mixing, a more stable atmospheric class, and more predictable plume dispersion behavior.

3.4 Performance in Field Deployments

The task of the LDAR programs is to alert the operator to fugitive (unplanned) emissions while ignoring non-fugitive (planned combusted or vented) emissions, which do not require the operator to dispatch personnel to the facility for repairs. Real O&G facilities are typically characterized by these planned emissions, which form a highly variable and noisy baseline emission signals that was not present during the controlled testing in this study. In effect, this study treated every emission source as a valid fugitive detection, which is a substantially simplified scenario relative to field conditions. As a result, controlled test results provide a first-order representation, and likely best-case, estimate of field performance.

Comparing detection probabilities to emission sizes in field studies provides an estimate of how solutions would perform in field conditions. To make this comparison, the study used component-level measurements from three measurement field studies at upstream [65] and gathering [19, 20] facilities. These distributions were overlaid with the range of DL90s observed in this study, see Figure 3.9. Using these data, the solutions with the lowest (1.3 kg CH_4/h) DL90 would detect <10% of the emitters found in the field study; this decreases to <1% for the solution with the highest DL90 (64.4 kg CH_4/h).

In addition, a Monte Carlo simulation of the emissions detection and quantification performance of selected tested continuous monitors using real emissions measurement data [19] is summarized in Table 3.6. Simulation results suggest that quantification estimates from continuous monitors for measurement-based inventories may substantially misstate the number of emitters and emission rate. Therefore, these estimates should be used cautiously until both detection limits and quantification accuracy are improved and uncertainties are better characterized.

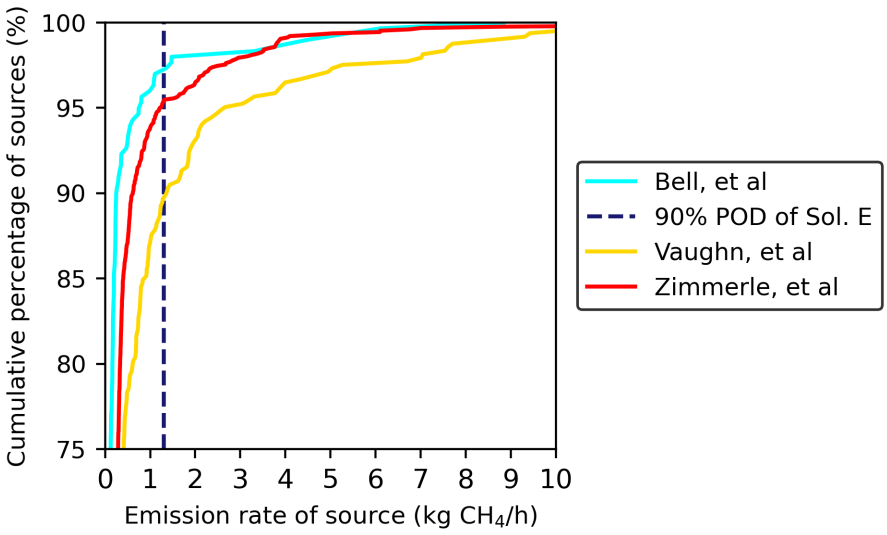


Figure 3.9: The range of DL90s observed in this study was overlaid on a cumulative distribution of direct measurements of component-level fugitive emissions from three studies across the O&G value chain. While the lowest DL90 is shown in the figure, the highest value (64.4 kg CH_4/h) is not shown as the x-axis is trimmed at 10.0 kg CH_4/h. Also, the y-axis is trimmed at 75.0% for improved readability. For all three studies, $\leq$10 of the measured sources exceed the *lowest* DL90 observed in the controlled testing result.

Table 3.6: Summary of the Monte Carlo analysis of the field performance of tested continuous monitors. The Monte Carlo simulation involved 1000 runs, sampling a total emission rate of 304.0 kg/h from 456 sources from the study by Vaughn et al. [19]

Solution		**TPF (%)**		**Quantification Ratio:** $\left[\frac{\text{Estimates}}{\text{Actual}}\right]$	
ID	Type	Mean	95% CI	Mean	95% CI
A	Point sensor network	67.1	[62.9, 70.8]	3.6	[1.9, 5.9]
B	Scanning/Imaging	37.9	[33.6, 42.3]	7.2	[2.6, 13.0]
C	Point sensor network	29.6	[25.9, 33.8]	1.2	[0.9, 1.7]
D	Point sensor network	53.9	[49.3, 58.8]	0.8	[0.6, 1.1]
E	Point sensor network	86.8	[83.8, 90.1]	14.3	[7.5, 23.1]
F	Point sensor network	64.3	[60.1, 68.6]	2.3	[1.7, 3.0]

3.5 Implications of Study Results

The state of the art of continuous monitors has improved substantially since the publication of the first single-blind controlled testing studies [65, 120]. In contrast with these first studies, the current study represents the first study completed using a consensus protocol [131] that supports repeatable testing across both time and test centers. While the current testing does not fully simulate field conditions, the duration, scope, and complexity of testing performed here represent substantially more rigorous and comprehensive testing than previous efforts.

As the stakes in next-generation LDAQ solutions rise dramatically given the increased interest in adopting results from these solutions into programs ranging from company-internal emissions mitigation efforts to regulatory programs with financial penalties, testing has also increased in rigor and quality to improve understanding of solutions performance and characterization of measurement uncertainties.

The results presented here indicate that users should utilize continuous monitors with caution. Detection limits, probability of detection, localization, and quantification may or may not be fit for purpose for all applications. For example, tested continuous monitors had better quantification accuracy with narrower uncertainty at the facility level compared to single sources. If performance is clearly understood and uncertainties are robustly considered, the continuous monitors tested here, as a group, work and need not have excellent performance across all metrics assessed in this study to be useful. For example, rapid detection of large emission sources to inform the urgent field response might not require accurate quantification. However, relying on single-source quantification estimates from these continuous monitors for emissions reporting is likely premature at the time of testing. Also, while a higher sensor count correlates with increased performance across some metrics, users must beware of the cost implications when scaled across multiple O&G facilities, especially when aiming to maximize performance. Therefore, depending on the desired application of a continuous monitor, determining the appropriate and economically feasible number of sensors to install per facility and where they are installed remains critical to the field performance of continuous monitors.

Finally, given the current rate of investment, the performance of continuous monitors is changing rapidly; hence, periodic retesting will be required to track development and ensure performance metrics are current.

Chapter 4

EVALUATING THE PERFORMANCE OF SURVEY LDAQ SOLUTIONS UNDER SINGLE-BLIND CONTROLLED TESTING[3]

4.1 Overview

Next-generation survey LDAQ solutions (e.g., drone-based, automobile-based, and aircraft-based technologies) can screen or survey large spatial areas of O&G facilities faster than other methods like continuous monitors, OGI camera surveys, and U.S.EPA method 21 [41]. Just like continuous monitors (Chapter 3), next-generation survey LDAQ solutions are generally equipped with data analytics that process raw sensor readings to infer detections, unlike existing regulatory-approved methods. Unlike continuous monitors, survey solutions provide periodic surveys at specified intervals/frequencies, do not operate continuously (a few minutes to hours), and require human supervision to work. As in the previous chapter, the literature review highlights the importance of comprehensive controlled testing using standardized test protocols to evaluate the state of the art of next-generation survey solutions and assess their readiness for field deployment.

This chapter examines the first-known single-blind controlled testing results of 12 survey LDAQ solutions that are not aircraft-based using a standardized controlled testing protocol [132]. The survey solutions are categorized into mobile solutions (i.e., automobile-based and drone-based solutions), handheld OGI solutions (i.e., infrared OGI cameras), and advanced handheld solutions (i.e., non-OGI camera solutions). Study results are evaluated based on the following performance metrics: detection classification and POD, (2) emission source localization accuracy/precision, and

[3]Chapter 4 is extracted from the published article.
[140] C. Ilonze, R. Day, E. Emerson, A. Duggan, R. Brouwer, and D. Zimmerle, "Performance of survey solutions under single-blind controlled testing protocol," *ChemRxiv*, 2025. Preprint URL: 10.26434/chemrxiv-2025-5ff89.

(3) survey time. Quantification performance is excluded since only 1 survey solution tested that capacity.

The chapter structures the discussion by first comparing the performance of handheld OGI camera solutions against other categories of solutions based on the performance metrics highlighted earlier. This also includes an investigation of the impact of test conditions on the detection performance of solutions. Secondly, the chapter evaluates the sensitivity of detection performance to changes in testing complexity. This section focuses on the variation in detection performance based on emission point height (tank and non-tank sources) and the number of sources per emissions event (i.e. per experiment). Just as in the previous chapter, the chapter concludes with the implication of the test results on field deployment.

4.2 Results and Discussion

4.2.1 Handheld OGI Camera Solutions vs Other Categories of Solutions.

Overall, handheld OGI camera solutions had comparable or better performance than other categories of solutions across most metrics assessed. A multivariable logistic regression analysis (Table 4.1) evaluating the impact of release rate, wind speed, and wind direction on the emissions detectability (TP and FN classification $\rightarrow$ POD) of tested survey solutions was conducted. Results indicated that an increase in release rate (over tested range) had a weak effect (odds ratio $\approx$1) on the tendency of tested survey solutions to detect emissions despite being statistically significant ($p < 0.05$) for 5 (C, D, E, H, and J) of the 12 tested solutions. Wind speed and direction were separately statistically significant to just one survey solution. Figures 4.1, 4.2, and 4.3 show the POD curves of tested survey solutions as a function of controlled release rate for handheld OGI camera solutions, advanced handheld solutions, and mobile solutions, respectively. The POD curves of solutions characterized the likelihood of detecting emission at various release rates over a wide range of environmental conditions. Measurement conditions like survey height and speed were not input parameters in evaluating the POD curves since they were either approximately constant during each solution's survey or were not documented by the surveyors (i.e., the personnel operat-

ing the solutions during the study). The POD curves of the remaining (7 of 12) survey solutions that were not statistically influenced by release rate and/or wind speed/direction may significantly depend on other factors that could not be investigated (out of scope) in this study, like variations in the survey (driving/flight) speed, height/distance from the emitting source, etc. For example, von Fischer et al. [58] showed that the detection limits of automobile-based LDAQ solutions tend to increase with measurement distance.

Table 4.1: Table summarizing the multivariable logistic regression analysis evaluating the impact of selected test conditions on the emissions detectability/POD of tested survey solutions.

Solution		Release rate		Wind speed		Wind direction	
ID	**Category**	**pValue**	**Odds ratio**	**pValue**	**Odds ratio**	**pValue**	**Odds ratio**
A	Mobile	0.935	1.000	0.418	1.057	0.324	1.002
B	Handheld OGI	0.087	1.008	0.348	0.830	0.560	1.001
C	Advanced Handheld	0.006	1.004	0.575	1.106	0.836	1.000
D	Mobile	0.029	1.009	0.003	0.631	0.049	1.003
E	Handheld OGI	0.000	1.008	0.978	1.005	0.534	1.001
F	Mobile	0.118	1.001	0.534	0.876	0.917	0.100
G	Handheld OGI	0.559	1.000	0.758	1.081	0.601	1.002
H	Advanced Handheld	0.019	1.004	0.758	1.050	0.369	1.003
I	Mobile	0.932	1.000	0.204	1.166	0.103	1.006
J	Advanced Handheld	0.023	1.002	0.988	0.998	0.066	1.004
K	Advanced Handheld	0.527	1.000	0.314	0.835	0.843	1.001
L	Handheld OGI	0.250	1.001	0.418	1.182	0.829	0.999

The p-values that are highlighted indicate statistical significance. Note that the results from a rank biserial correlation analysis did not align with the observed trend regarding the impact of emission rate on the emissions detectability of solutions A, I, and K.

Figures 4.1, 4.2, and 4.3, along with Table 4.2, present the predicted DL90 values for the tested survey solutions. The DL90s of the eight solutions shown were within two orders of magnitude of their respective maximum tested release rates. The predicted DL90s ranged from 0.01 [0.00,

0.21] kg CH_4/h to 1.31 [0.61, 3.98] kg CH_4/h, while the remaining four solutions had their DL90s reported as "NA".

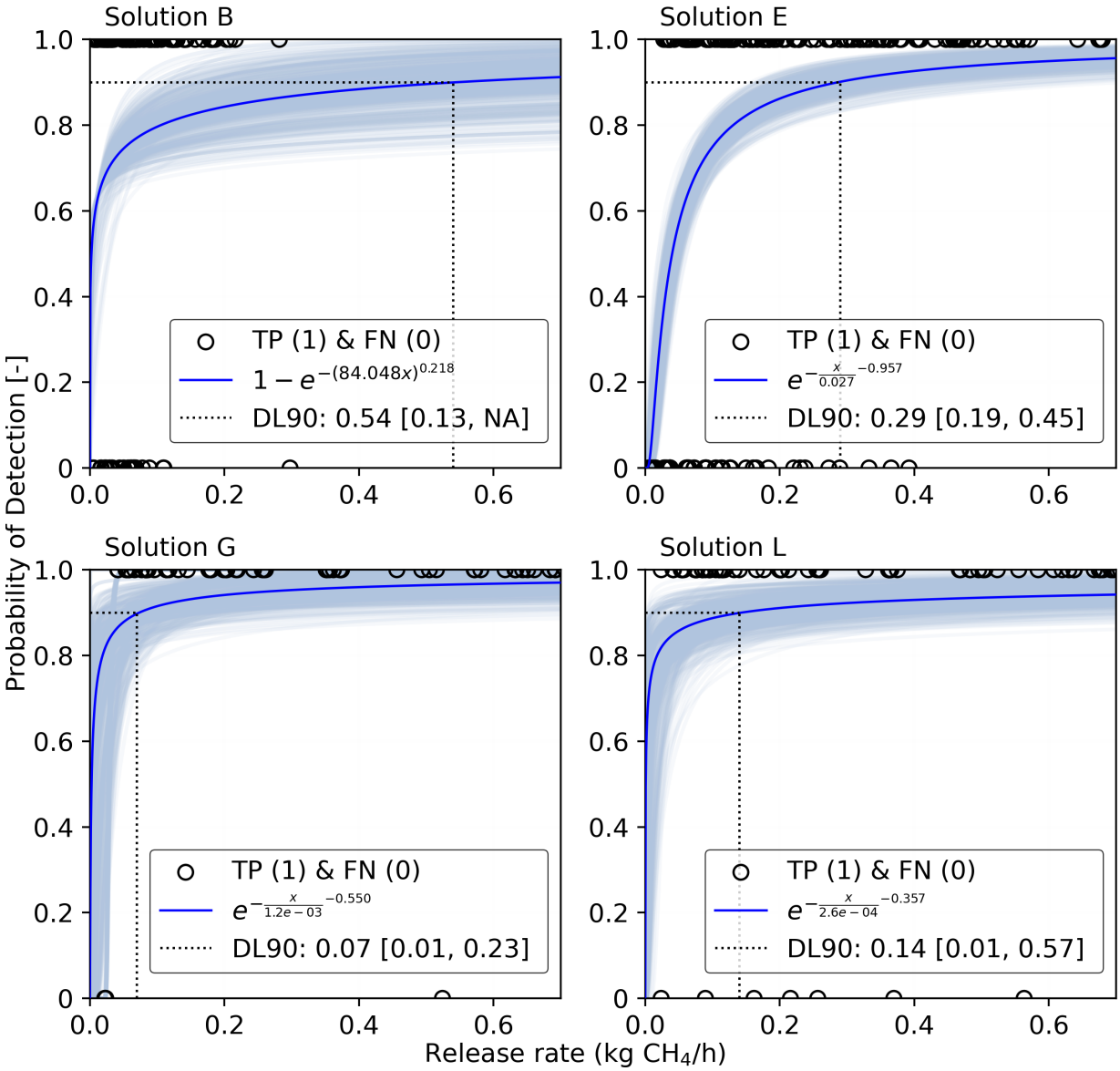


Figure 4.1: Probability of detection (POD) as a function of controlled release rate (kg CH_4/h) for all hand-held OGI camera solutions. The DL90, which is the emission rate at which a solution achieves 90% POD, is shown together with the confidence interval. The uncertainty on the POD curve is illustrated by a cloud of curves obtained through bootstrapping, with the confidence interval evaluated as the 2.5 and 97.5 percentiles. The value "NA" is given when the DL90 or the associated uncertainty is more than $\times 20$ the maximum release rate tested or cannot be evaluated. A scatter plot with circular markers shows binary detection classification results: TP detections (1) and FN detections (0). The upper limit of the x-axis has been trimmed at 0.7 kg CH_4/h to maintain the same axis limit for all survey solutions.

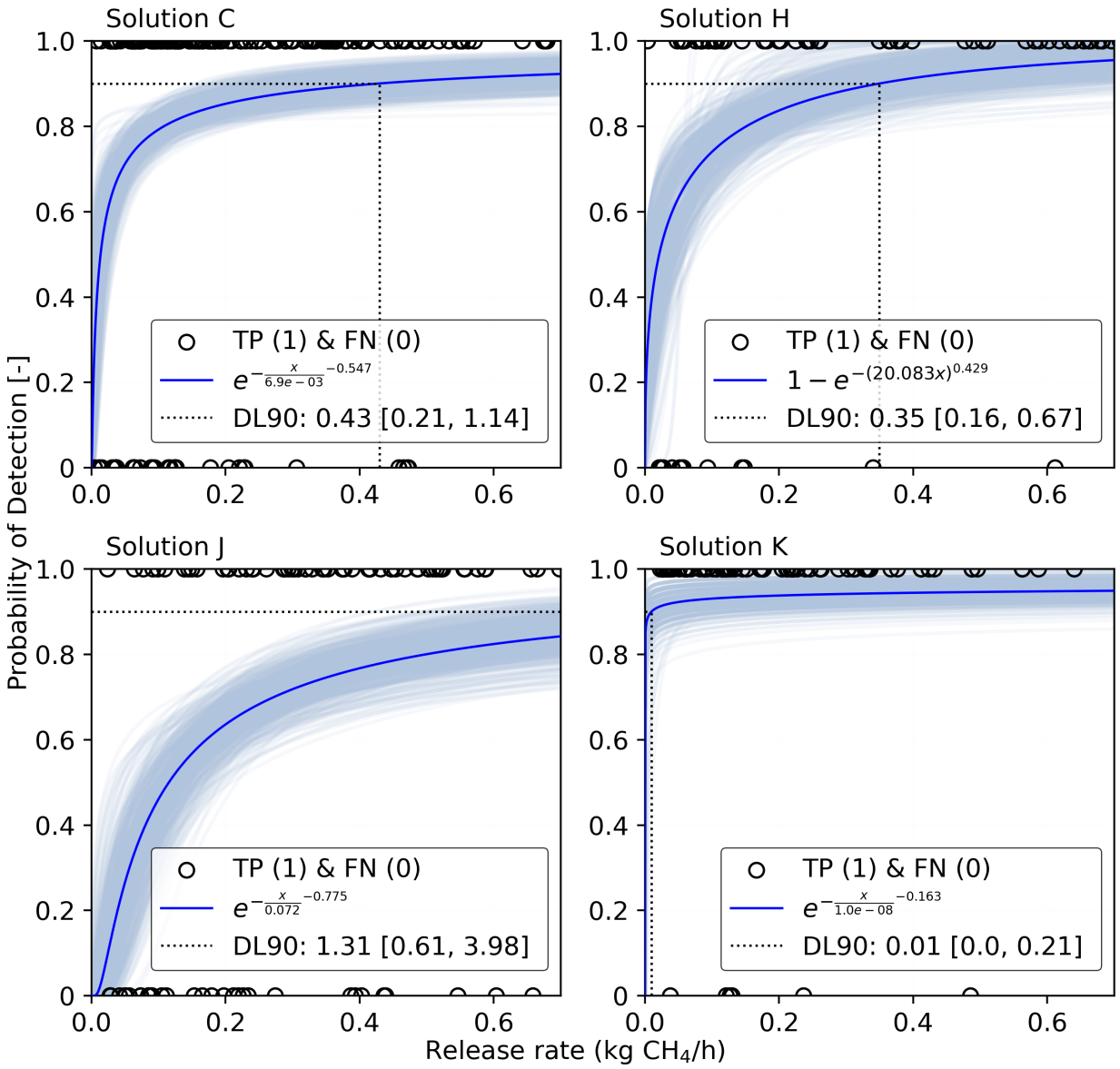


Figure 4.2: Probability of detection (POD) as a function of controlled release rate (kg CH_4/h) for advanced handheld solutions. The DL90, which is the emission rate at which a solution achieves 90% POD, is shown together with the confidence interval. The uncertainty on the POD curve is illustrated by a cloud of curves obtained through bootstrapping, with the confidence interval evaluated as the 2.5 and 97.5 percentiles. The value "NA" is given when the DL90 or the associated uncertainty is more than $\times 20$ the maximum release rate tested or cannot be evaluated. A scatter plot with circular markers shows binary detection classification results: TP detections (1) and FN detections (0). The upper limit of the x-axis has been trimmed at 0.7 kg CH_4/h to maintain the same axis limit for all survey solutions. Note: Solution K has near-uniform POD across tested emission rates with odds ratios of ≈ 1.

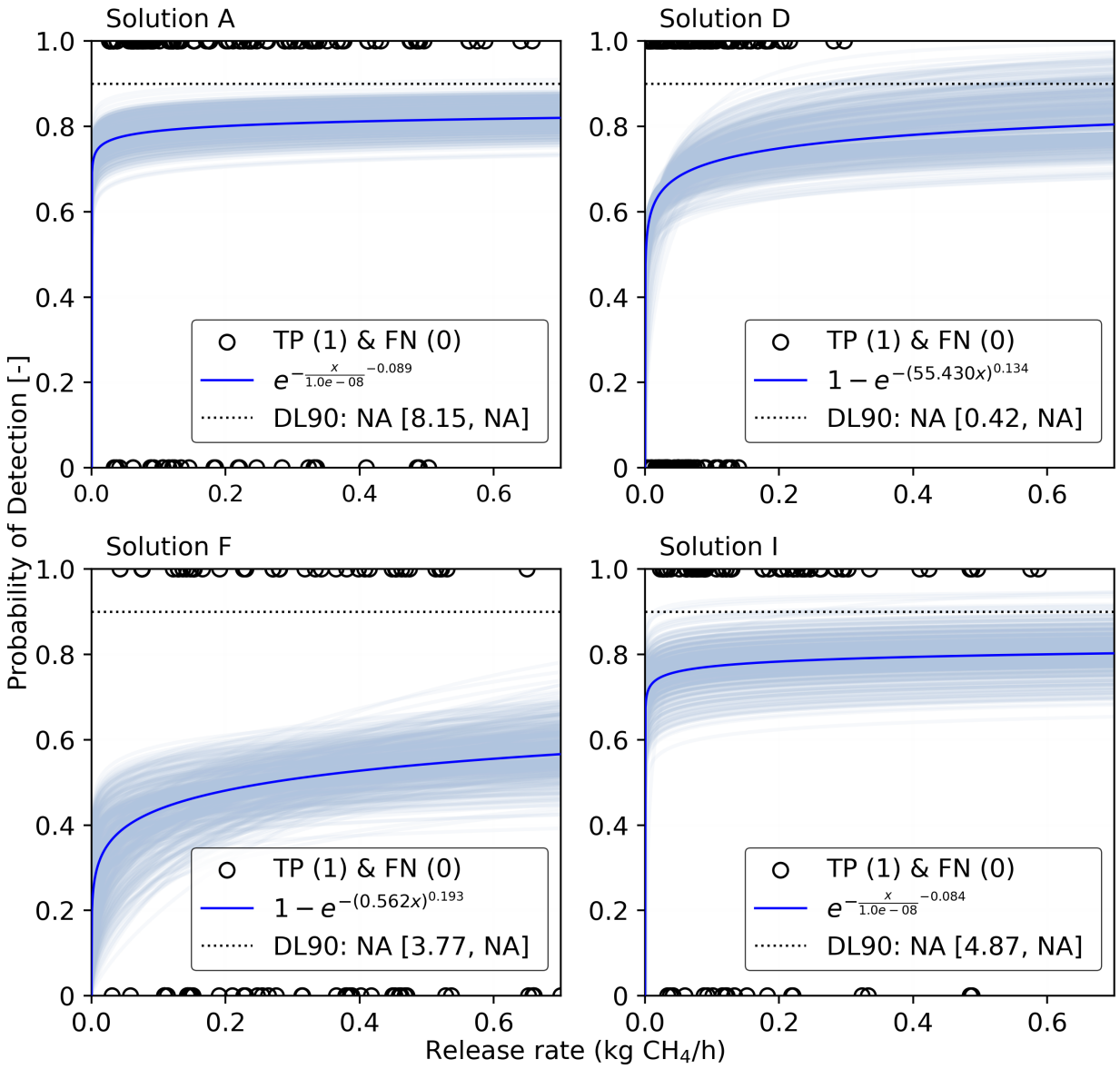


Figure 4.3: Probability of detection (POD) as a function of controlled release rate (kg CH_4/h) for mobile solutions. The DL90, which is the emission rate at which a solution achieves 90% POD, is shown together with the confidence interval. The uncertainty on the POD curve is illustrated by a cloud of curves obtained through bootstrapping, with the confidence interval evaluated as the 2.5 and 97.5 percentiles. The value "NA" is given when the DL90 or the associated uncertainty is more than $\times 20$ the maximum release rate tested or cannot be evaluated. A scatter plot with circular markers shows binary detection classification results: TP detections (1) and FN detections (0). The upper limit of the x-axis has been trimmed at 0.7 kg CH_4/h to maintain the same axis limit for all survey solutions. Note: Solutions A and I have near-uniform POD across tested emission rates with odds ratios of ≈ 1.

Caution must be exercised when interpreting inferences drawn from the POD curve assessed for solution K, which recorded the lowest DL90 (0.01 [0.00, 0.21] kg CH_4/h) in the study. A rank-

biserial correlation analysis on the impact of release rate on the emissions detectability of solutions showed that the POD of solution K tended to decrease with release rate over the tested release rate range, which contradicts the results presented in Table 4.1. Similar trends were observed for solutions A and I. Analysis indicates that all mobile survey solutions had DL90 values exceeding two orders of magnitude of their respective maximum tested release rates. Additionally, they exhibited some of the highest FPF (19.2% to 50.0%) and FNF (20.3% to 48.5%) values observed in this study as shown in Table 4.2 and Figure 4.4. These results indicate mobile survey solutions' struggles at detecting fugitive emissions within tested rates. Recall that the release rates tested were within an order of magnitude of component-level leak sizes measured at natural gas production sites in the U.S., excluding liquid unloading or major failure conditions that produced exceptionally high-rate emission events [11, 65, 141].

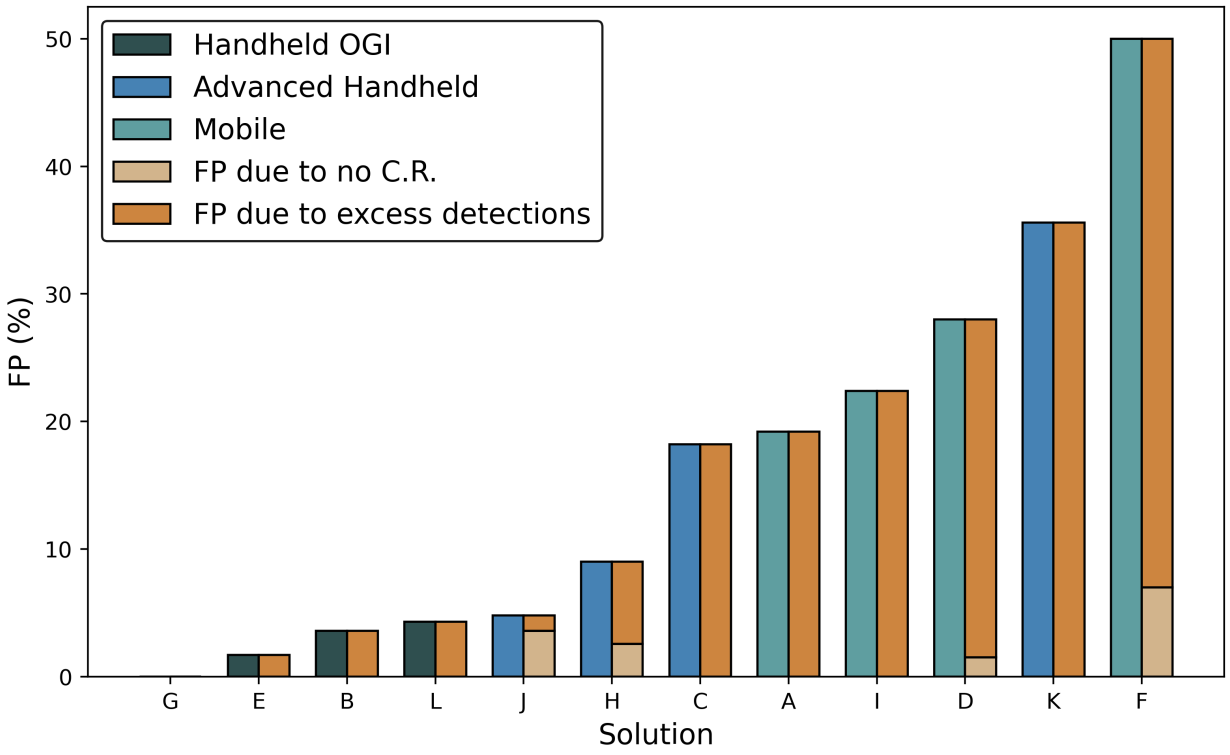


Figure 4.4: Categorization of the FPs evaluated in this study for all survey solutions arranged in increasing order of FPF. FPF due to no controlled release is the fraction of all FPs due to detection reports sent when no ongoing controlled release exists. FPF due to excess TP detections is the fraction of all FPs due to excess detections identifying controlled releases already paired as a new and/or a different emitter.

Table 4.2: Results of the detection classification scheme (FPF and FNF) and the 90% POD predicted by each solution sorted in order of increasing 90% POD (DL90)

Solution		Survey Time[†]	Localization (Unit)		FPF (%)				DL90	
ID	Category	per Unit (minutes)	Accuracy (%)	Precision (%)	All	No Controlled Release	Excess TP Detection	FNF (%)	Estimate[‡] (kg CH_4/h)	Within Tested Range
K[*]	Advanced Handheld	2.53 [1.14, 5.00]	62.9	97.7	35.6	0.0	100.0	6.6	0.01 [0.00, 0.21]	Yes
G	Handheld OGI	5.86 [1.11, 15.0]	98.5	98.5	0.0	0.0	0.0	5.7	0.07 [0.01, 0.23]	Yes
L	Handheld OGI	2.30 [0.72, 4.86]	95.7	100.0	4.3	0.0	100.0	8.2	0.14 [0.01, 0.57]	Yes
E	Handheld OGI	1.91 [1.00, 6.17]	97.1	98.8	1.7	0.0	100.0	24.1	0.29 [0.19, 0.45]	Yes
H	Advanced Handheld	5.09 [0.78, 14.0]	91.0	100.0	9.0	28.6	71.4	16.5	0.35 [0.16, 0.67]	Yes
C	Advanced Handheld	1.89 [0.94, 4.39]	79.5	97.2	18.2	0.0	100.0	19.6	0.43 [0.21, 1.14]	Yes
B	Handheld OGI	1.66 [0.67, 5.00]	92.8	96.2	3.6	0.0	100.0	25.2	0.54 [0.13, NA]	No
J	Advanced Handheld	1.03 [0.39, 2.86]	95.2	100.0	4.8	75.0	25.0	31.0	1.31 [0.61, 3.98]	Yes
D[‡]	Mobile	2.02 [0.94, 8.00]	62.9	87.4	28.0	5.4	94.6	31.7	NA [0.42, NA]	No
I[‡]	Mobile	1.63 [0.56, 3.57]	61.2	78.8	22.4	0.0	100.0	22.4	NA [4.87, NA]	No
A[‡]	Mobile	0.94 [0.44, 2.57]	55.0	68.0	19.2	0.0	100.0	20.3	NA [8.15, NA]	No
F[‡]	Mobile	1.73 [0.50, 3.67]	32.0	64.0	50.0	14.0	86.0	48.5	NA [3.77, NA]	No

[†] This is time taken to survey an equipment unit in a survey area.
[‡] When the DL90 is more than twice the order of magnitude of the maximum release rate tested, the value is given as "NA".
[*] The DL90 is almost 0 because the POD curve is approximately constant at POD $> 90\%$.

Handheld OGI camera solutions had 3 of the 4 lowest DL90s obtained in the study. Unlike advanced handheld solutions, emission rate and other test conditions were not statistically significant to the POD of most of the OGI camera solutions. This observation aligns with data from Zimmerle et al. [43] which demonstrated that the experience level of an OGI camera surveyor had a greater impact on successful emissions detection than other tested variables. Similarly, Ravikumar et al. [120] identified the distance between the OGI camera and an emitting source as the most important parameter affecting the detection effectiveness of handheld OGI camera solutions. Since solution vendors conducted the emissions surveys, it is likely that surveyors sent by the solutions were highly experienced and used that expertise to maximize survey solution performance; deployment of the same solutions at real O&G facilities may not have access to that level of experience. Emission rate was statistically significant to the emissions detectability/POD of 3 of the 4 advanced handheld solutions tested likely because advanced handheld solutions, similar to mobile solutions, measure ambient emissions concentrations. However, because they are typically deployed closer to emission sources, their performance relies less on favorable plume transport conditions (unlike mobile solutions) but more on emission rate for successful detection. Additionally, handheld OGI camera solutions exhibited some of the lowest FPF values in the study ($\leq$4.3%), which were statistically lower than those of advanced handheld and mobile solutions. While two out of four handheld OGI camera solutions had FNF values below 10.0%, the FNF values of the handheld OGI solutions were not statistically different from those of mobile and advanced handheld solutions.

If a field emissions measurement application defined a false positive (FP) strictly as an instance where emissions were erroneously detected despite no ongoing emissions, then the FPF evaluated in this study would decrease by 75% to 100% for all solutions except solution J, as illustrated in Table 4.2 and Figure 4.4. Overall, only two handheld OGI camera solutions demonstrated a balance of low DL90 values ($\leq$ 0.14 kg CH_4/h) with low FPF ($\leq$ 9%) and FNF ($\leq$ 5%). This performance underscores why handheld OGI camera solutions are widely used for follow-up investigations and regulatory-compliant leak detection and repair (LDAR) inspections at O&G facilities. Additionally, study data revealed that the DL90s of all handheld solutions (both OGI and advanced), except

for solution J, met the U.S. EPA's NSPS OOOO(b) final rule's minimum detection threshold requirement for quarterly monitoring ($\leq$ 1 kg CH_4/h) at well sites and centralized production facilities using alternative technologies [136]. For mobile solutions, the observed detection performance could be attributed to ineffective data analytics (e.g., poorly defined lower detection limits) or because their intended applications were beyond the scope of this test. Some mobile solutions may have been designed with intentionally high detection limits to focus on identifying high-emission events that were not tested in this study.

To assess the localization performance of the solutions, the equipment unit-level localization precision and accuracy of participating survey LDAQ solutions were evaluated. Table 4.2 indicates that the equipment unit-level localization accuracy (Figure 4.5) and precision of both handheld OGI cameras and advanced solutions were statistically higher ($p < 0.05$) than those of mobile solutions.

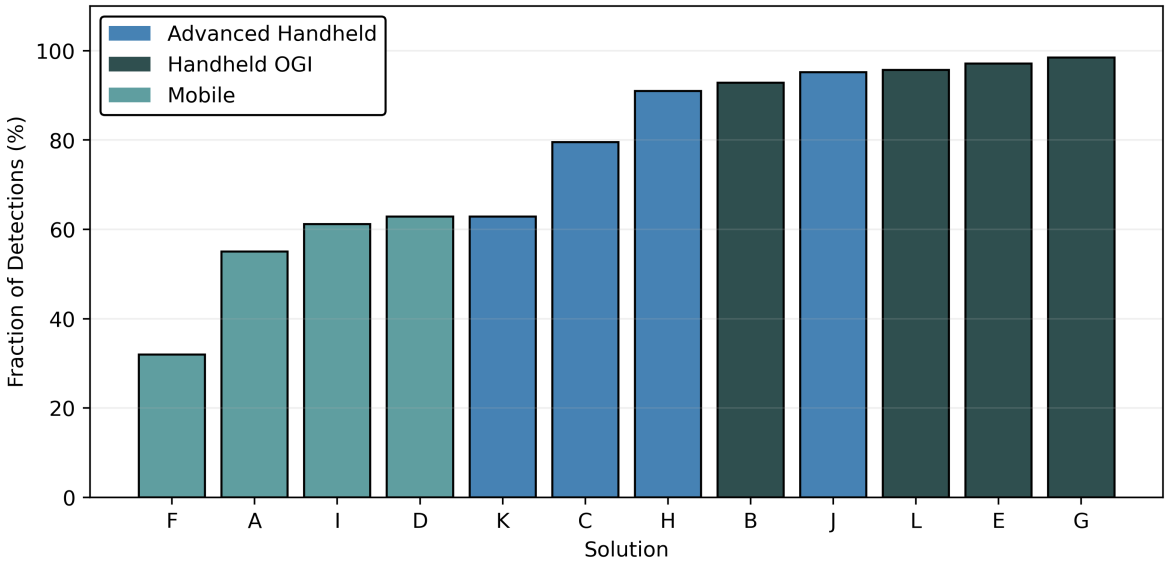


Figure 4.5: Unit-level localization accuracy for all survey LDAQ solutions arranged in increasing order of localization accuracy. The y-axis shows the fraction of detection reports that identified a leak source at the correct equipment unit. The marker colors identify handheld OGI, advanced handheld, and mobile solutions.

Although some mobile survey solutions demonstrated promising equipment unit-level localization performance (accuracy and precision > 50% and 60%, respectively), these findings highlight

the superior effectiveness of handheld solutions in accurately pinpointing leak sources for repairs or reporting. One explanation for the localization performance of mobile solutions is that their data analytics typically require favorable wind conditions to predict emissions location accurately. Generally, the sensor must be positioned directly downwind, winds must have steady speed and direction, and the atmospheric stability class must be well-characterized. This hypothesis is demonstrated by solution D, the only solution for which wind conditions statistically impacted detection performance (Table 4.2), as it exhibited the highest equipment unit-level localization accuracy and precision among mobile solutions.

To characterize the survey time of the solutions, surveyors documented the time required to complete emissions surveys during each experiment. However, as discussed in Subsection 2.2.2, the survey area defined for each experiment differed as testing complexity varied. To enable comparison, as summarized in Table 4.2, the recorded survey time for each experiment was normalized by the number of equipment units within the survey area defined for that experiment. This normalization accounts for variations in pad sizes and configurations, ensuring a consistent basis for evaluating survey efficiency across LDAQ solutions. Figure 4.6 presents the distribution of survey times per equipment unit for each survey solution tested. The face color of each violin plot represents the solution category (mobile, handheld OGI, or advanced handheld). Results indicate that mobile survey solutions generally surveyed O&G assets significantly faster than both handheld OGI and advanced handheld solutions. The result remained the same even when solutions G and H were excluded from the analysis. These two LDAQ solutions exhibited greater skewness in their survey time per equipment unit distribution (mean > median), likely due to inefficiencies in their survey methodologies or challenges in their implementation during testing. However, the survey times of the two handheld solution categories were not statistically different.

As suggested by Fox et al. [45], faster survey times have important cost implications for field deployments, as longer survey durations can increase an operator's emissions monitoring expenses when scaled across multiple O&G facilities. While mobile survey solutions, as a group, exhibited faster survey times per equipment unit compared to handheld solutions in this study, some handheld

solutions individually achieved shorter survey times per equipment unit than certain mobile survey solutions. For example, Figure 4.6 indicates that only two mobile solutions ranked among the top four solutions with the lowest mean survey time per equipment unit.

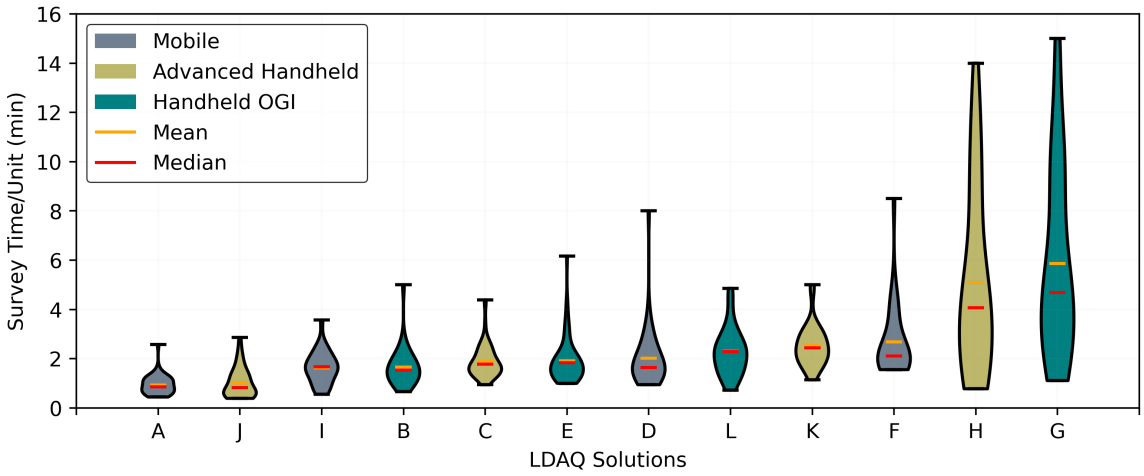


Figure 4.6: A violin plot showing the distribution of survey time per equipment unit arranged in increasing order. The survey time per equipment unit was evaluated by dividing the survey time by the number of equipment units in the survey areas within which an experiment was conducted. The face color of each violin indicates if a solution is a mobile, a handheld OGI, or an advanced handheld solution. The violin whiskers represent the minimum and maximum survey times per equipment unit, while the blue (dashed) and orange (solid) lines indicate the median and mean survey times per equipment unit, respectively.

This observation could be attributed to several factors. First, the noticeable odor and audible sound produced by gas releases during testing may have provided additional olfactory and auditory cues to surveyors using handheld solutions, potentially giving them an advantage over those using mobile solutions in identifying emission sources. Unlike mobile survey solutions, handheld solutions require surveyors to operate in close proximity to potential emitters. This positioning allows surveyors using handheld solutions, especially when downwind of a plume, to leverage sensory cues that help them quickly pinpoint leak sources. In contrast, surveyors using mobile solutions—typically positioned at the edge of a pad—often lack this sensory advantage. This limitation arises because plumes diffuse before reaching the surveyor (e.g., a drone controller) or because the surveyor is physically separated from the surroundings (e.g., inside a vehicle). Although this sensory advantage with handheld solutions is also observed in real facilities, the test

facility (METEC) lacks the sensory distractions typically present at real O&G facilities, which can make emission surveys more challenging. Real O&G facilities often have sensory distractions that can produce misleading auditory and olfactory signals, potentially prolonging survey durations compared to the controlled conditions of this study. These distractions include operational noise from equipment, the movement of trucks, and baseline emissions from routine venting events.

In general, the key takeaway from this study is that more handheld OGI camera solutions demonstrated better emissions detection (i.e., lower DL90, FNF, and FPF) and localization (accuracy and precision) for small leak sources compared to advanced handheld and mobile solutions. While some tested mobile solutions performed comparably to certain handheld solutions on specific metrics, study findings indicate that mobile survey solutions are generally likely optimized for quicker surveys and initial source identification.

4.2.2 Performance Variation with Testing Complexity.

Study data showed that the detection tendency of mobile survey solutions was generally more sensitive to changes in testing complexity than other solution categories. Recall that experiments in this study involved 1 to 6 simultaneous controlled releases from emission sources at varying heights. Findings indicate that mobile solutions faced greater difficulty in isolating and detecting individual emissions during experiments with multiple controlled releases compared to other solution categories. Specifically, more mobile solutions (2 out of 4) exhibited statistically (Mann-Whitney U test) higher FNF per experiment in multi-source release scenarios than in single-source release events, as shown in Table B.1. Additionally, significant differences were observed between the FNFs of mobile solutions and handheld survey solutions for all experiments involving multiple controlled releases. In contrast, these differences were not statistically significant when considering experiments with a single controlled release. These results highlight a key challenge commonly faced by LDAQ solutions that depend on wind transport for detection (e.g., mobile solutions and continuous monitors). The data analytics and algorithms of these solutions often struggle

to robustly isolate and differentiate individual emission sources during multiple-source emission events.

When controlled releases for solutions were aggregated based on emission point heights by categorizing them as tank and non-tank sources, statistical analysis suggested that no solution group exhibited significant differences in FNFs between these source categories. However, handheld solutions demonstrated statistically significantly lower FNF for tank sources compared to mobile solutions. This result is expected, given that handheld solutions involve close component-level emissions inspections, whereas mobile solutions are typically positioned farther from the emission source (e.g., at the edge of a well pad). Despite this result, direct proximity surveys may not always be feasible in real-world applications due to inaccessibility, high-temperature surfaces, or moving parts, necessitating the use of alternative emissions measurement approaches like mobile solutions.

4.2.3 Implications of Study Results

Unlike prior studies [120, 122, 127], this work represents a significant step forward in the controlled testing of next-generation handheld, automobile-based, and drone-based survey solutions using a consensus test protocol [132].

The study findings have shown the following: First, no survey solution or category of solutions (i.e., handheld OGI, advanced handheld, or mobile) achieved optimal performance across all assessed metrics. All solutions demonstrated the ability to systematically detect emissions, albeit with varying strengths and limitations. Handheld OGI camera solutions exhibited high accuracy and repeatability in detecting and localizing small fugitive emissions, while mobile survey solutions generally surveyed O&G assets quicker than other solution categories. This implies that mobile survey solutions can rapidly identify potential leak sources, enabling targeted inspections and measurements with handheld solutions for cost-effective emissions mitigation.

Secondly, isolating and differentiating individual sources from multiple simultaneous emitters remains a significant challenge for solutions that rely on wind transport for detection (e.g., mobile solutions, continuous monitors). Unless these solutions are specifically intended for site-level

emissions monitoring, enhancing their data analytics and algorithms is crucial to overcoming this limitation.

This study's results should be applied with caution for two reasons. First, since the test facility used for the study mimics near-ideal real facility operational conditions, solutions' results likely represent best-case scenario performance. Second, robust field testing of the solutions is needed to validate and build confidence in assessed controlled testing performances.

Chapter 5

EVALUATING QUANTIFICATION PERFORMANCE OF THE QUANTITATIVE OPTICAL GAS IMAGING SYSTEM USING SINGLE-BLIND CONTROLLED TESTING[4]

5.1 Overview

The previous two chapters have examined the state of the art in detection, localization, and quantification performance of a broad range of LDAQ solutions. However, this investigation did not provide an in-depth analysis or guidelines on how these solutions should be deployed at real O&G facilities for near-optimal performance. Among these solutions, the OGI infrared camera—capable of visualizing methane plumes and volatile organic compounds (VOCs) invisible to the naked eye—is widely adopted for emissions monitoring and detection across the O&G value chain in the U.S. and Canada. This widespread use is due to its demonstrated effectiveness, both anecdotally and scientifically (as shown in the previous chapter), as well as its regulatory backing in these regions [151]. However, the OGI camera does not quantify detected emissions but relies on additional, supplementary steps (Hi-flow sampler [11, 20, 142], downwind tracer flux [11, 65, 143], and the other test method [144–146]) to perform that operation. This additional step has recently been made possible with the OGI camera using *quantitative* OGI (QOGI) technology, which quantifies leak rates using the video images stream with the camera. The QOGI technology is commercialized as a tablet to be attached to the OGI camera (e.g., Providence Photonics QL100, Provi-

[4]Chapter 5 is extracted from published article:
[139] C. Ilonze, J. L. Wang, A. P. Ravikumar, and D. Zimmerle, "Methane quantification performance of the quantitative optical gas imaging (qogi) system using single-blind controlled release assessment," *Sensors*, vol. 24, no. 13, 2024. DOI: 10.3390/s24134044

dence Photonics QL320, FLIR QL320, etc.) or integrated directly into the camera (recent development). Essentially, the operationalization involves quantifying leaks already detected with an OGI camera. While the detection performance of OGI cameras have been richly studied [43, 44, 152], the quantification performance and the impact of measurement conditions on the results of the QOGI technology are poorly understood and characterized.

This work examines the performance of the QOGI system by conducting controlled testing of FLIR QL320 under near-field conditions at METEC. Study results will highlight the accuracy and precision of the system along with the effect of emission rate, plume background, wind speed, and measurement distance when observing typical fugitive emission rates ($\leq$ 3.0 kg CH_4/h). Measurement distance, plume background, and leak type (i.e., emission source category) rely on the user's judgment, while emission rates and meteorological conditions (e.g., wind speed) are uncontrollable factors. This chapter combines findings from the investigations to outline favorable measurement scenarios using the QOGI tool, offering guidelines for obtaining accurate measurements at real O&G facilities.

5.2 Results and Discussion

A gamma generalized linear model with a log link function was used to relate quantification relative error to emission rate, measurement distance, plume background (ordinal: sky = 1, ground = 2, equipment = 3), and wind speed (ordinal: calm = 1, normal = 2, high = 3). Due to the non-symmetric distribution of quantification relative errors, the model was divided into two categories (equations 5.1 and 5.2). Ambient temperature was excluded due to high multicollinearity with other variables. Equations 5.1 and 5.2 describe relationships for quantification relative errors > 0 and < 0, respectively.

$$\text{ln(Relative error)} = \ln(0.986)X_1 + \ln(1.215)X_2 + \ln(2.727)X_3 + \ln(0.774)X_4 + 0.248 \quad (5.1)$$

$$\text{ln(Relative error)} = \ln(1.014)X_1 + \ln(0.823)X_2 + \ln(0.367)X_3 + \ln(1.292)X_4 + 4.029 \quad (5.2)$$

$$X_1 = \text{Emission rate} \quad (p < 0.05); X_2 = \text{Plume background} \quad (p < 0.05) \quad (5.3)$$

$$X_3 = \text{Wind speed} \quad (p < 0.05); X_4 = \text{measurement distance} \quad (p < 0.05) \tag{5.4}$$

5.2.1 Quantification Accuracy and the Impact of Emission Rate

Tested controlled release rates ranged from 2.2 slpm to 88.0 slpm (0.1 kg CH_4/h to 2.9 CH_4/h), which is within 90% of component-level measurements from production O&G as shown in Figure 3.9. Data as shown in Figure 5.1 indicates a general overestimation bias of 27% with quantification relative errors ranging from -90% to +831% ($\approx \frac{1}{10}\times$ to $\approx 10\times$). Results show that 46% and 75% of estimates were within a quantification factor of 2 $(-50\%|\frac{1}{2}\times,\ +100\%|2\times)$ and $(-67\%|\frac{1}{3}\times,\ +200\%|3\times)$ of actual rates, respectively. The findings from Figure 5.1 and Equations 5.1 and 5.2 indicate that quantification accuracy (relative error $\rightarrow 0\%$) statistically improved slowly as emission rate increased. This observed improvement is likely because a large emission rate implies a higher path-integrated concentration, which can increase plume image contrast for the same ΔT and enhance the signal-to-noise ratio for more accurate estimates. Note, ΔT represents temperature difference between gas plume and gas plume background.

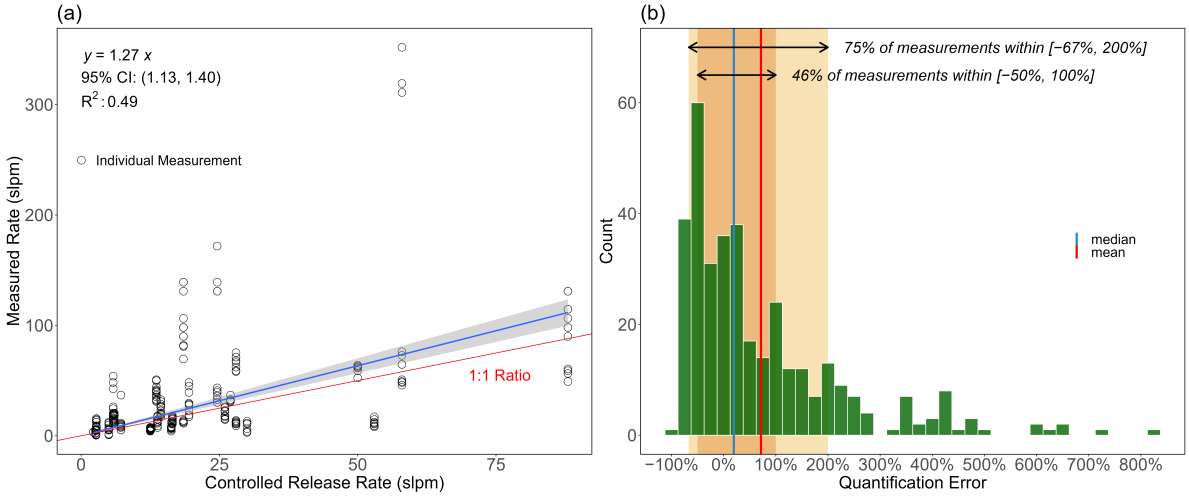


Figure 5.1: Quantification accuracy of individual estimates: (a) measured rates versus controlled release rate and (b) distribution of quantification error of individual estimates. In (a), the blue line represents linear regression through the origin with the gray shading showing the 95% confidence interval of the regression when bootstrapped. The red line represents the 1:1 ratio, where the measured rate matches the controlled release rate. In both cases, the intercept was forced to zero. In (b), the orange shading represents the measured rate within a factor of two of the controlled release rates (-50% to +100% quantification error), and the yellow shading represents the measured rate within a factor of three of the controlled release rates (-67% to +200% quantification error)

5.2.2 Quantification Accuracy and the Impact of Plume Background

According to the FLIR QL320 user manual, a gas plume background with sufficient thermal contrast (at least $\Delta T \geq 2°\text{C}$) is needed for successful emission rate estimation with the QOGI system [113]. Given that the QOGI tablet's quantification method tracks changes in the pixel intensity of infrared images, apparent temperature changes or disturbances in the plume background can interfere with plume boundary identification and consequently quantification accuracy. These disturbances include shadows, glints, reflections of heat sources on any metallic equipment, and motion like cloud cover or ground vegetation near the equipment. This study, like Zimmerle et al. [43], grouped plume backgrounds into equipment, ground, and sky background; quantification performance based on these categories is shown in Figure 5.2.

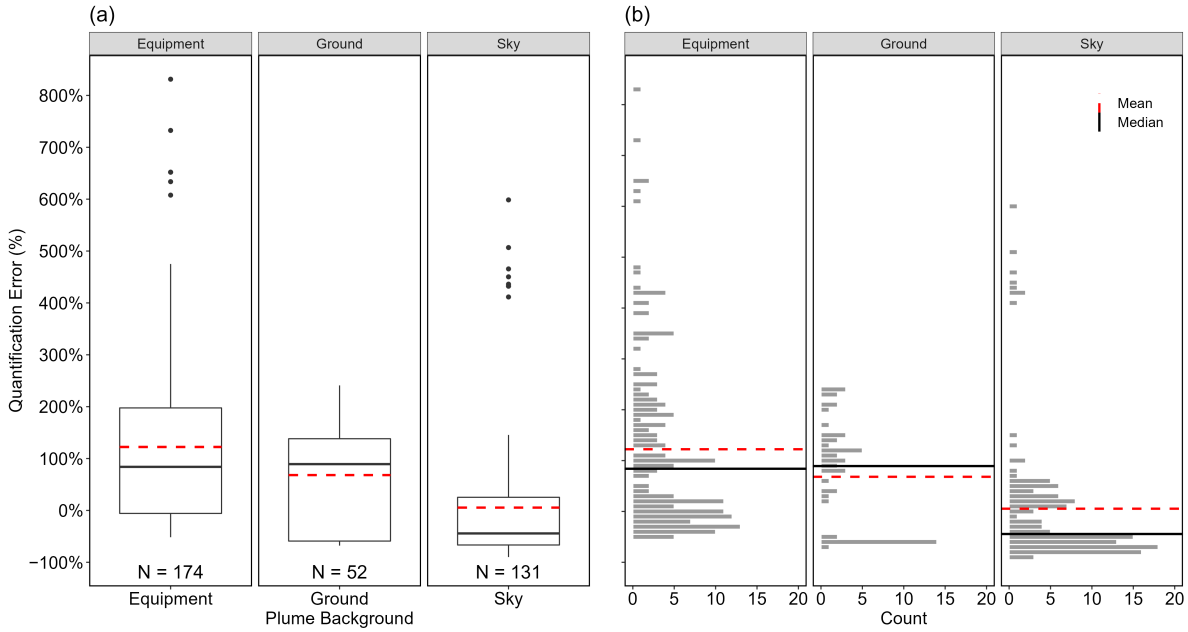


Figure 5.2: Box plots (a) and distribution (b) of the quantification errors of individual estimates based on quantification background (Equipment, Ground, and Sky). The black, circular markers in (a) represent outliers as data points outside 1.5 times the inter quartile range of the data in each group. The horizontal, black solid, and red dotted lines in both (a) and (b) indicate the median and mean of the distribution in each group, respectively. The x-axis represents the groups within each parameter, and the y-axis shows the quantification error in percentage. The numbers at the bottom of the box plots represent the sample sizes of individual estimates within each group.

A background was classified as "equipment" when the gas plume was viewed against either a different part of the same equipment (e.g., a wellhead casing) or nearby equipment (e.g., a neighboring wellhead unit). A background was classified as "sky" when the plume was viewed against the sky, which may or may not have included cloud cover (e.g., when an elevated emission source like the thief hatch of a tank was viewed against the sky). A background was classified as "ground" when the gas plume was seen against the ground (i.e., sand, stones, gravel, vegetation).

Experimental results and Equations 5.1 and 5.2 suggest that measurements taken with the sky as a plume background (mean error: +5% (95% CI [-13%, +32%])) were more accurate than other backgrounds. This is likely because a clear sky with low apparent temperature provided the best thermal contrast for the QOGI system's quantification algorithm, although cloud cover made quantification very challenging. Note that all the outliers (quantification errors between 400% and 600%) for the sky background in Figure 5.2 were estimates conducted under high wind speeds, which will be investigated in the next subsection. Quantification with equipment as background was generally challenging due to visual interferences and poor thermal contrast: unlike most equipment units at real O&G facilities, equipment at METEC was non-operational, hence no fluid flow, which could have likely increased thermal contrast between equipment walls with the gas plume for more accurate estimates.

5.2.3 Quantification Accuracy and the Impact of Wind Speed

Prevailing wind speed was a categorical input parameter in the QOGI devices tested, with three defined levels: calm (0–1 mph), normal (2–10 mph), and high (>10 mph). Analysis in Figure 5.3 and Equations 5.1 and 5.2 reveals that quantification became challenging and accuracy worsened as wind speed increased. This is likely because of the increased turbulent, variable, and unsteady plume dispersion as wind speed increases, which can adversely affect the quality of plume detection. In other words, calm wind speed (mean error: -29% (95% CI [-35%, -21%]) provided the best wind condition for quantification, although with a strong underestimation tendency.

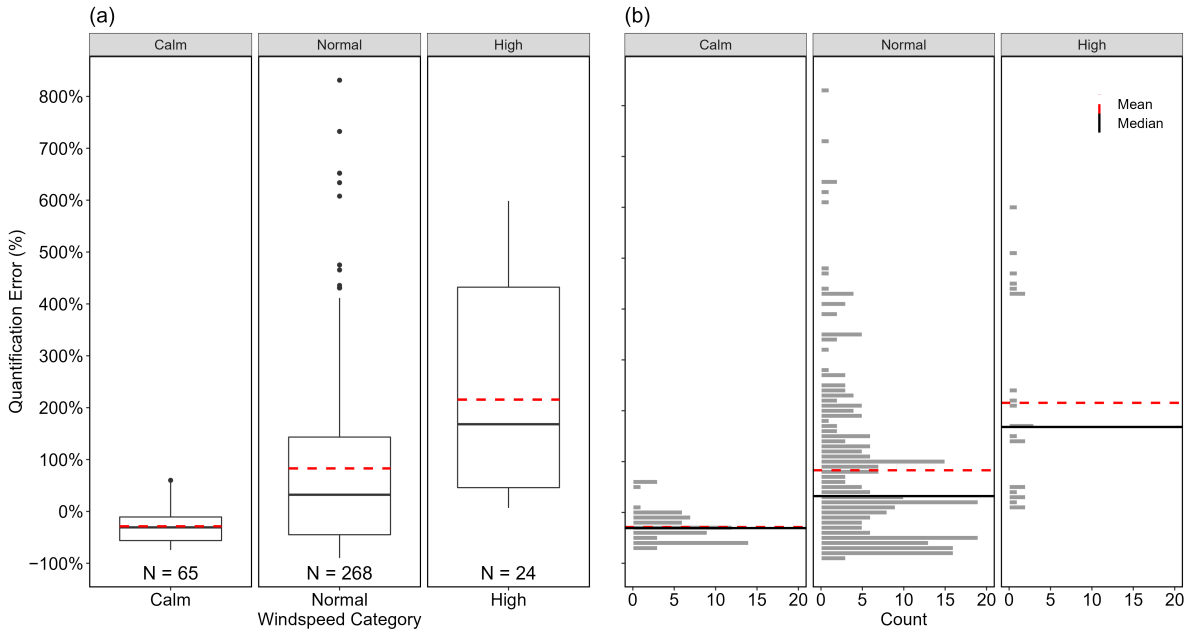


Figure 5.3: Box plots (a) and distribution (b) of quantification errors of individual estimates based on wind speed categories (calm: 0–1 mph; normal: 2–10 mph; or high: >10 mph). The black, circular markers in (a) represent outliers as data points outside 1.5 times the inter quartile range of the data in each group. The horizontal, black solid, and red dotted lines in both (a) and (b) indicate the median and mean of the distribution in each group, respectively. The x-axis represents the groups within each parameter, and the y-axis shows the quantification error in percentage. The numbers at the bottom of the box plots represent the sample sizes of individual estimates within each group.

5.2.4 Quantification Accuracy and the Impact of Measurement Distance

For the QOGI system tested, the acceptable measurement distance from the emitting source is a function of the OGI camera lens [113]. This study used a 23 mm (24° FOV) lens, which limited the measurement distance to ≈1.5 m to ≈16 m (5 feet to 54 feet). Measurement distance and the choice of camera positions were dependent on space availability for the OGI camera's tripod stand and camera perspectives adjudged by the field crew to be the "best" view of the plume during measurements. Also, small and low-pressure emissions tend to equilibrate quickly with the atmosphere as they exit the source. In such cases, the camera was positioned closer to increase the visibility of the smaller plume in the camera image. For example, more than half of the measurements of rates <10 slpm were performed from distances between 1.5 and 2.0 m. Therefore, the measurement distances in this study were unintentionally skewed towards relatively short distances, a scenario

likely to be observed in real O&G facilities, especially in newer production facility designs, where equipment units are clustered closer than in older production facility designs.

Results as shown in Figure 5.4 and Equations 5.1 and 5.2 indicate that quantification accuracy improved with increasing measurement distance. This is likely because a longer measurement distance captures more of the plume's dynamics within the image frame, leading to improved quantification accuracy compared to shorter distances. However, caution is needed to prevent visual noise from the background or adjacent components from entering the gas plume image. This issue becomes more prevalent as the measurement distance from the emission source increases and can negatively impact quantification performance.

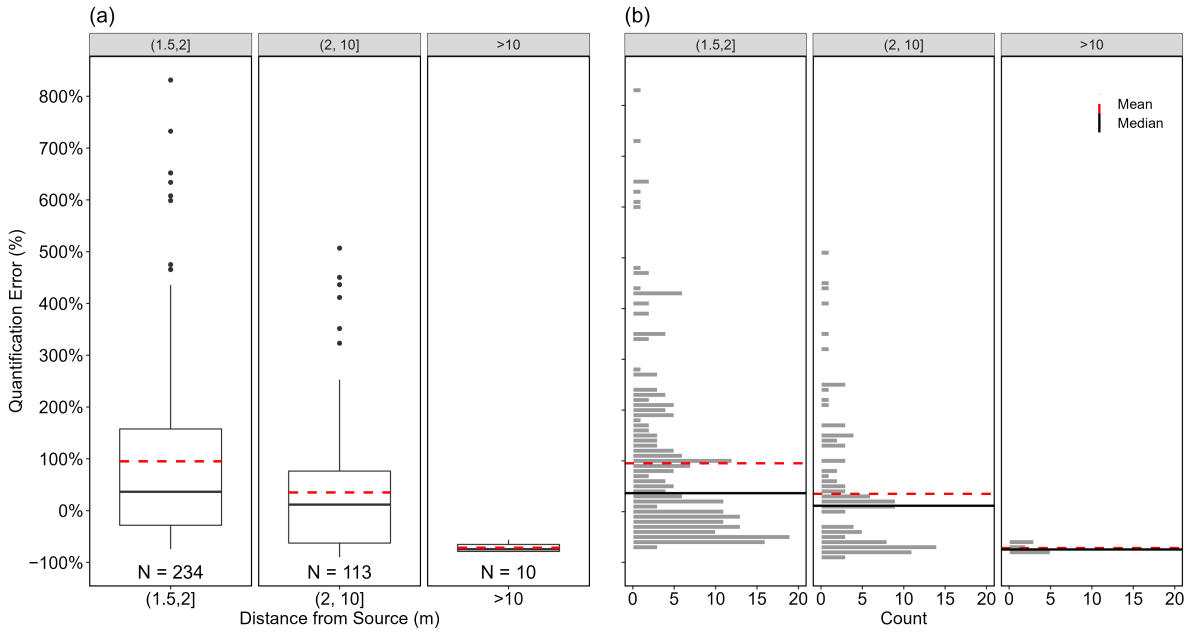


Figure 5.4: Box plots (a) and distribution (b) of the quantification errors of individual estimates based on measurement distance ((1.5, 2] m, (2, 10] m, and >10 m). The black, circular markers in (a) represent outliers as data points outside 1.5 times the inter quartile range of the data in each group. The horizontal, black solid, and red dotted lines in both (a) and (b) indicate the median and mean of the distribution in each group, respectively. The x-axis represents the groups within each parameter, and the y-axis shows the quantification error in percentage. The numbers at the bottom of the box plots represent the sample sizes of individual estimates within each group.

5.2.5 Diffuse vs Point Leak Type

Based on the measurement crew's assessment of the emission point geometry, the leak type was classified as either "point" or "diffuse" in the QOGI device before measurements. Emission sources

with a diameter of less than 2 inches (e.g., connectors, valve packing) were labeled as "point," while those with a diameter greater than 2 inches (e.g., flanges, thief hatches) were categorized as "diffuse". Statistical analysis showed no significant difference in quantification performance when "point" was mistakenly selected instead of "diffuse" as the leak type during measurement. However, the "point" category had a narrower error range (-21% to 155%; 90% of estimates within a quantification factor of 2) compared to "diffuse" (-52% to 600%; 88% within a quantification factor of 2). This observation suggests the potential immaturity of this feature's algorithm when this testing was conducted.

5.2.6 Observed Favorable Measurement Scenario

In actual O&G facilities, the emission rate is always unknown until estimated. During measurements, the field crew intentionally chose the plume background and measurement distance, unlike the prevailing wind condition, which is beyond human control. Table 5.1 summarizes the quantification performance of the QOGI tablets under different measurement scenarios (A–E) irrespective of the prevailing wind condition and release rate. A clear sky background and longer measurement distance (scenario E) were separately identified as favorable for quantification with the FLIR QOGI device. However, study data showed that measurements with equipment as the plume background at distances of 1.5 to 2 meters (scenario A) had the highest fraction of estimates within a factor of 2 (60%) compared to other scenarios.

This observation could be attributed to the "coupling effect" of measurement conditions on quantification performance, as most estimates in scenario A were obtained under calm wind speed (0–1 mph) and for sources within an enclosed chamber. This setup limited the rapid variability of wind conditions/plume dispersion during measurements for improved quantification performance. For scenarios D and E, lower wind speed did not necessarily improve quantification performance, as the fraction of estimates within a factor of 2 remained almost the same ($\pm$2%). However, it is important to note that few estimates were obtained under calm wind conditions. The key observation from the above discussion is that the QOGI tablets tested can obtain relatively accurate

estimates in special cases even when all favorable measurement conditions identified do not co-exist.

Table 5.1: Table summarizes quantification performance under different measurement scenarios (plume background and measurement distance) in this study with a sample count greater than 20. For each measurement scenario, quantification performance is illustrated with the 95% empirical confidence interval and the percentage of estimate within a quantification factor of 2 (-50%, 100%)

		Measurement conditions		**Quantification relative error**	
Scenario	**Sample size**	**Plume background**	**Measurement distance (m)**	**Percentage within** $(-50\%\|\frac{1}{2}\times,\ +100\%\|2\times)$	**95% C.I.**
A	149	Equipment	(1.5, 2.0]	60.0	[-47.0, 639.0]
B	25	Equipment	(2.0, 10.0]	24.0	[7.0, 335.0]
C	52	Ground	(1.5, 2.0]	21.0	[-62.0, 241.0]
D	33	Sky	(1.5, 2.0]	49.0	[-68.0, 492.0]
E	88	Sky	(2.0, 10.0]	49.0	[-87.0, 432.0]

5.2.7 Quantification Precision

Quantification precision was evaluated by comparing the quantification errors of estimates obtained at the same camera position and experiment. The study defined precision range at either the camera position or the experiment level as the difference between the minimum and maximum quantification error evaluated for that level. Ideally, since the controlled release rate of each experiment remained approximately constant throughout all measurements, the quantification errors of estimates from the same camera position were expected to be similar (i.e., precision range $\approx 0\%$), assuming the prevailing measurement conditions remained unchanged. This was not the case in reality, as the precision range for the same camera position spanned from 2% to 439%, and was even worse at the experiment level, ranging from 17% to 690%.

Analysis showed that measurements at the same camera position (Figure 5.5a) had better quantification precision, with 75% of camera positions having a precision range $\leq$50%, compared to the experiment level (Figure 5.5b), where only 35% of experiments had a precision range $\leq$50%. This wide range in quantification precision, especially at the experiment level, could often be attributed to rapid variations in meteorological conditions between when measurement parameters were inputted into the QOGI tablets and the start of actual measurements. Additionally, changes in camera position/perspective could have contributed to the variability. The different camera positions reflect possible decisions made by surveyors with varying levels of experience. More experienced surveyors are more likely to select the "best" available camera perspective for visualizing and quantifying a plume. In other words, this result shows the likely influence of surveyor experience on quantification performance.

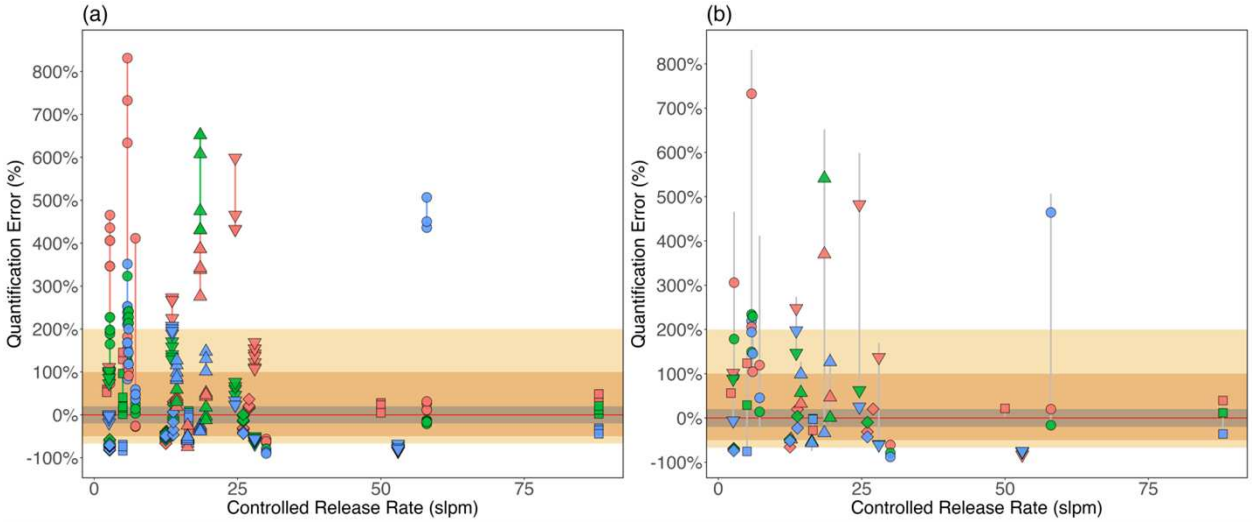


Figure 5.5: Quantification precision versus emission rate at (a) the camera position level and (b) the experiment level. The markers in (a) represent the errors of individual estimates/measurements. Markers of the same type and color show quantification errors of estimates obtained at the same camera positions. The whiskers connecting markers of the same color and shape show the range of quantification errors (precision range) observed at each camera position. The markers in (b) represent the mean error for each camera position. The whiskers connecting markers of the same shape but of different colors (mean error at camera positions) represent the precision range observed during each experiment. The gray shading represents estimates within 20% of the controlled release rate (-20% to +20% quantification error). The dark orange shading represents estimates within a quantification factor of two of the controlled release rates (-50% to +100% quantification error), and the light orange shading represents estimates within a factor of three of the controlled release rates (-67% to +200% quantification error).

To investigate appropriate measurement practices for more accurate estimates, findings from Table 5.2 suggest that using the mean of estimates from multiple camera positions produced more accurate and reliable results. This approach yielded a tighter error range (-79%, 297%) and a higher percentage of estimates within a factor of 2 and 3 of the actual rates compared to using the mean of estimates from the same camera position or single estimates.

Table 5.2: Table summarizes and compares quantification performance when (1) individual estimates are used, (2) the mean estimates at camera position are used, and (3) the mean estimates for any controlled release (experiment) including measurement from at least 1 camera position.

Aggregation level	**Error range (%)**	**Percentage within** $(-50\%\|\frac{1}{2}\times,\ +100\%\|2\times)$	**Percentage within** $(-67\%\|\frac{1}{3}\times,\ +200\%\|3\times)$	**Sample size**
Individual estimates	[-90, 831]	46.0	75.0	357
The mean of estimates at a camera position	[-88, 733]	45.0	72.0	73
The mean of estimates at the experiment level	[-79, 297]	54.0	77.0	26

While all the results discussed so far highlight the wide uncertainty in single estimates, which can significantly impact the implementation of emissions mitigation programs, some applications prioritize quantification accuracy and the associated uncertainty when source-level estimates are aggregated at the facility or asset level. When all individual estimates and controlled releases in this study were aggregated, the QOGI system overestimated the total controlled release rate by 43% (95% CI [+23%, +55%]), which falls within the range of values obtained for continuous monitors in chapter 3.

5.3 Implications of Study Results

This study systematically investigates the impact of release rate, plume background, and selected user input data on the quantification performance of the FLIR QL320 QOGI device. Results indicate a wider quantification error range $(-90\%|\frac{1}{10}\times,\ +800\%|9\times)$ than the prior study (-90% to +330%) that tested a similar but older QOGI device. However, the maximum rate in the current study was about an order of magnitude less than that of the prior study [109]. While the current study result showed improved quantification accuracy and precision as the release rate increased, a similar systematic investigation will be needed to understand quantification performance for larger rates (i.e., $>>$ 3 kg CH_4/h), which is an important emission source category.

Study results also showed that calm wind speed ($<$1 mph), farther measurement distance, and viewing emissions against a clear sky background improved quantification performance. Since computational algorithms are proprietary, the cause of improved performance cannot be stated. However, less turbulent plume dispersion at low wind speeds ensures stable plume behavior, which is favorable for quality plume imaging and quantification. Additionally, viewing the plume against a clear sky, where the sky's apparent temperature is typically low, provides sufficient thermal contrast for more accurate estimates. Longer measurement distance likely ensured that all the plume dynamics were captured within the image, further enhancing quantification accuracy. Conversely, cloudy skies, vegetation on the ground, and/or backgrounds with visual noise and poor thermal contrast/ΔT were unfavorable for quantification.

The key control element for the study was the methodology applied by the OGI surveyors: the same method was used for all positions and all conditions. Controlling the measurement method removed the field crews' experience and bias from the study design. Given that the same measurement methodology was replicable across all experiments, the wide variation (up to 690%) in quantification error as camera position changed highlights that results are highly variable based on camera position and potentially subtle changes in measurement conditions. Therefore, in addition to selecting measurement conditions favorable for quantification as identified in this study, users

are likely to obtain more reliable and accurate estimates by taking measurements from more than one camera position and using the mean of the estimates.

Finally, the different camera positions in this study reflect the likely decisions made by surveyors based on their experience level during measurement. As a result, the experience of the surveyor handling the OGI camera—identified by Zimmerle et al. [43] as the most significant factor influencing detection performance—could significantly impact quantification accuracy of the QOGI system. Therefore, further studies would be needed to assess the impact of surveyor experience on quantification performance.

Chapter 6

ASSESSING THE PROGRESS IN THE PERFORMANCE OF LDAQ SOLUTIONS UNDER A SINGLE-BLIND CONTROLLED TESTING PROTOCOL[5]

6.1 Overview

Chapters 3 and 4 examined the first controlled testing of continuous monitors and survey (i.e., handheld, drone-based, and automobile-based) LDAQ solutions using standardized controlled testing protocols [131, 132]. As highlighted in the literature review, there is growing interest from stakeholders such as OG operators, regulators, and voluntary organizations like the Oil and Gas Methane Partnership (OGMP) [116] in using these next-generation LDAQ solutions for methane emissions mitigation programs. This interest is based on the underlying assumption that these next-generation LDAQ solutions are rapidly evolving over time [136]. Until now, there has been no standardized, systematic and third-party investigation of this assumption, as there was no consensus-controlled testing protocol to support such assessment. A critical objective of such standardized protocols is to reliably characterize and compare performance levels (intra- and inter-LDAQ solutions) across space and time.

---

[5]Chapter 6 is extracted from published articles

[137] Reprinted with permission from C. Bell, C. Ilonze, A. Duggan, and D. Zimmerle, "Performance of Continuous Emission Monitoring Solutions under a Single-Blind Controlled Testing Protocol," *Environmental Science & Technology*, vol. 57, no. 14, pp. 5794–5805, 2023. DOI: 10.1021/acs.est.2c09235.Copyright 2023 American Chemical Society.

[138] Reprinted with permission from C. Ilonze, E. Emerson, A. Duggan, and D. Zimmerle, "Assessing the progress of the performance of continuous monitoring solutions under a single-blind controlled testing protocol," *Environmental Science & Technology*, vol. 58, no. 25, pp. 10941–10955, 2024. DOI: 10.1021/acs.est.3c08511. Copyright 2024 American Chemical Society.

To this end, this study evaluated the change in the performance of four continuous monitors (A, B, D, and F) that tested approximately one year after their initial controlled testing, as discussed in chapter 3. The evaluation followed the same protocol at the same test facility (METEC). Similarly, three survey solutions (A, H, and K) discussed in chapter 4 were tested again approximately three to twelve months using the same protocol at the same test center. All continuous monitors that were retested were deployed similarly to their initial test, using the same number of sensors to monitor the same test area. The only exception was solution F, which deployed two additional sensors compared to its first deployment. There was no significant difference ($p < 0.05$) in the wind speed distribution between the first and second test programs for each continuous monitor, except for solution B, which experienced calmer wind conditions during the second test. During the second testing, the controlled release rate and duration distributions were significantly different, skewing towards larger rates and longer durations compared to the first test. This adjustment aimed to better characterize the POD curve of solutions, particularly those that were not robustly characterized in the first test (e.g., solution B). Additionally, this adjustment was made in response to vendor requests for larger rates and longer-duration controlled releases. Increasing these test conditions statistically improved the detection performance of the continuous monitors testing again (Table 3.1 in Chapter 3), though the impact on quantification accuracy varied in statistical significance (Table 3.4 in Chapter 3). One of the sensors installed by solution A failed during the second testing. However, assessing the impact of this failure on performance was challenging since continuous monitors did not attribute detection reports to specific sensors at the test facility during the study.

For the survey solutions (A, H, and K), the distributions of relevant test conditions (wind speed and release rate) in the second test differed significantly ($p<0.05$) from those in the first test. However, only controlled release rate statistically influenced the detection tendency of solution H, while wind speed had no significant impact on any solution tested, as shown in Table 4.1 in Chapter 4. As was done in Chapters 3 and 4, metrics like the POD curve, true positive fraction (TPF)/FNF, FPF, localization accuracy/precision, quantification accuracy, survey times, and operational factor were used to evaluate performance and to determine if solutions progressed between test programs.

The chapter concludes with the implications of the study, including the impact of regular controlled testing on the development of LDAQ solutions.

6.2 Results and Discussion

6.2.1 Change in the Performance of Continuous Monitors

Overall, most of the tested continuous monitors demonstrated performance improvements across multiple assessed metrics during repeat testing. Figure 6.1 and Table 6.1 compares the detection performance of the four continuous monitors that were tested twice. Blue and crimson-red colored data in the figure represent the first and second test results, respectively. Three of the four solutions showed the same or reduced DL90 with FNF and FPF in the second test relative to the first test. Also, the same continuous monitors balanced relatively low DL90 with relatively low FPF, unlike in the first test, although the FNF was still high as these solutions missed at least a quarter of controlled releases conducted. Recall that, as discussed in Chapter 3, the increased sensitivity (i.e., low DL90) of some continuous monitors also led to more false positive alerts. Their algorithms struggled to differentiate smaller fugitive emissions from background noise (e.g., sensor or algorithmic noise), frequently mistaking background fluctuations for active emissions and issuing excessive false alerts. Conversely, low sensitivity (i.e., higher DL90) can result in continuous monitors missing relatively smaller emission rates, which typically make up the majority of field measurement studies (by count), resulting in high FNF.

For localization performance, most (3 of 4) of the continuous monitors demonstrated improved localization precision and accuracy at both the equipment unit level and equipment group level, during the second testing compared to the first test, as shown in Table 6.2.

Table 6.3 and Figure 6.2 compare the quantification performance of continuous monitors in quantify individual sources over the two test programs. Blue and crimson-red colored data in the figure represent the first and second test results, respectively. Analysis reveals that most continuous monitors showed improved quantification performance, although the wide uncertainty observed in the first test persists.

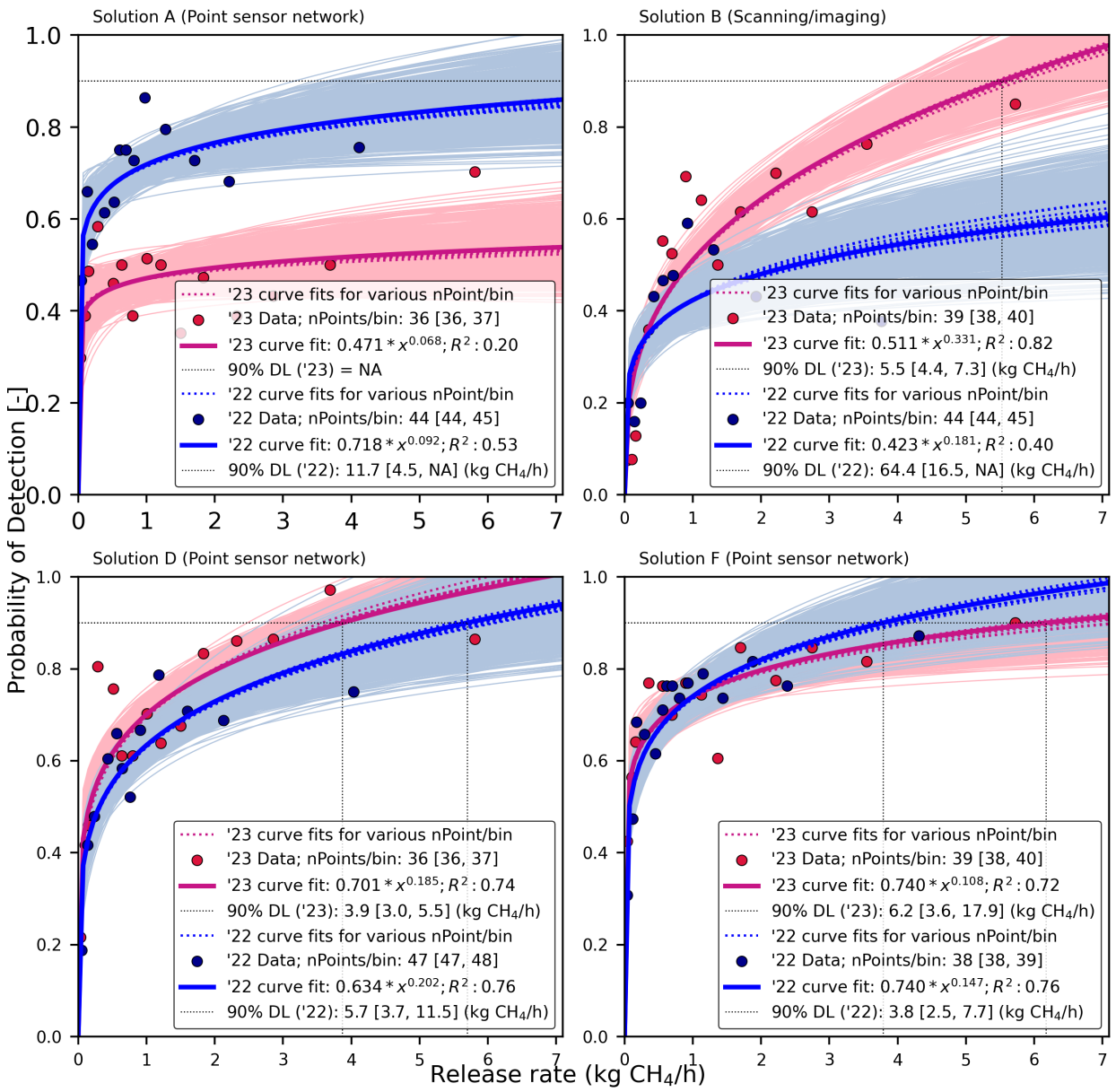


Figure 6.1: The figure compares the POD curve as a function of emission rate (kg CH_4/h) fitted using power functions for the four continuous monitors that tested twice (A, B, D, and F). The x-axis is divided into equal-sized bins with each marker (pod) as the fraction of controlled releases in a bin classified as true positives. Data from the first tests (2022 - blue color) is overlaid on the second test results (2023 - crimson red) for comparison. The emission rate at which the POD reaches 90% is indicated as the method detection limit (DL90) for each solution. Each pod data point is bootstrapped to produce a cloud of curves illustrating associated uncertainty. When the bootstrapping could not evaluate the lower and upper empirical Confidence Limit (CL) on a solution's DL90 best estimate, they are given as 0 and NA, respectively. Curve fits (dotted colored lines) obtained using other quantile-based discretizations are shown for comparison. The DL90s of 3 of the 4 solutions (B, D, and F) in the current study were within the tested emission rate range. The mean count of points per bin along with the min. and max. counts across all bins is also shown in the figure.

Table 6.1: Comparison of the detection performance of continuous monitors that tested twice, sorted in order of increasing solutions' DL90s.

Solution		Count		FPF (%)[†]				Method Detection Limit	
ID	Type	Controlled Release	Detection Reports	All	No Controlled Release	Excess Detections	FNF (%)	DL90[‡] (kg CH_4/h)	Within Tested Rates
Results from the second testing									
D	Point sensor network	547	403	6.9	28.6	71.4	31.4	3.9 [3.0, 5.5]	Yes
B	Scanning/Imaging	547	300	7.7	39.1	60.9	49.4	5.5 [4.4, 7.3]	Yes
F	Point sensor network	547	444	10.6	8.5	91.5	27.4	6.2 [3.6, 17.9]	Yes
A[*]	Point sensor network	547	487	47.8	61.8	38.2	53.6	NA	No
Results from the first testing									
F	Point sensor network	574	516	22.5	39.7	60.3	30.3	3.8 [2.5, 7.3]	Yes
D	Point sensor network	574	376	10.4	79.5	20.5	41.3	5.7 [3.7, 12.3]	Yes
A	Point sensor network	574	986	59.8	26.9	73.1	31.0	11.7 [4.1, NA]	No
B	Scanning/Imaging	445	250	31.2	61.5	38.5	61.3	64.4 [17.9, NA]	No

[†] **All** is the percentage of all detections classified as false positives based on the protocol. **No controlled release** is the fraction of all false positives due to detection reports sent when no ongoing controlled release exists. **Excess TP Detections** is the fraction of all false positives due to excess detections identifying controlled releases already paired as a new and/or a different emitter.

[‡] When DL90s cannot be evaluated, they are given as NA. Similarly, the lower and upper empirical 95% confidence intervals on DL90 are given as 0 and NA, respectively, when they cannot be evaluated.

[*] One of the sensors installed failed during testing.

Table 6.2: Comparison of the emission source localization (equipment unit) precision and accuracy for continuous monitors that tested twice, arranged in decreasing localization precision at the equipment unit level.

Solution				Source Localization (Equipment Unit)					
				Precision (%)			Accuracy (%)		
ID	Category	Sensor Density (sensors/m^2)	Count of TPs	Unit	Group	Facility	Unit	Group	Facility
Results from the second testing									
B	Scanning/Imaging	0.00012	277	89.5	9.4	1.1	82.7	91.3	92.3
F	Point sensor network	0.00118	397	40.8	53.9	5.3	36.5	84.7	89.4
D	Point sensor network	0.00095	375	27.2	68.8	4.0	25.3	89.3	93.1
A	Point sensor network	0.00095	254	26.0	49.6	24.4	13.6	39.4	52.2
Results from the first testing									
B	Scanning/Imaging	0.00012	172	70.9	15.7	13.4	48.8	59.6	68.8
A	Point sensor network	0.00095	396	28.0	39.4	32.6	11.3	27.1	40.2
F	Point sensor network	0.00095	400	24.8	50.2	25.0	19.2	58.1	77.5
D	Point sensor network	0.00095	337	0.0	52.8	47.2	0.0	47.3	89.6

Table 6.3: Comparison of single-estimate quantification performance of continuous monitors that tested twice along with their 95% empirical confidence limits arranged in increasing the percentage of measurements within a factor of 3.

Solution		Estimates within a Factor of 3			Relative Quantification Error (%)					
		$(-67\%\|\frac{1}{3}\times,\ +200\%\|3\times)$			Controlled Release (0.1 – 1] kg CH_4/h			Controlled Release > 1 kg CH_4/h		
ID	**Type**	**All**	**(0.1 – 1]**	**> 1**	**Mean**	**Median**	**95% C.I.**	**Mean**	**Median**	**95% C.I.**
Results from the second testing										
B	Scanning/Imaging	90.0	96.0	89.0	37.4	28.1	[-65.0, 168.5]	55.7	31.5	[-62.3, 339.4]
F	Point sensor network	78.0	90.0	71.0	15.7	-9.6	[-80.4, 195.5]	-12.8	-41.8	[-89.6, 232.3]
D	Point sensor network	67.0	79.0	61.0	64.5	-3.0	[-75.8, 729.0]	30.8	-45.8	[-92.5, 395.7]
A	Point sensor network	60.0	55.0	72.0	330.4	138.4	[-62.2, 1803.4]	57.6	-18.3	[-86.3, 612.5]
Results from the first testing										
B	Scanning/Imaging	74.0	76.0	80.0	74.6	39.5	[-81.1, 343.2]	41.9	25.3	[-90.2, 268.8]
F	Point sensor network	65.0	62.0	75.0	202.2	110.9	[-39.7, 933.2]	9.2	-40.5	[-82.5, 373.6]
A	Point sensor network	64.0	65.0	73.0	211.3	134.2	[-60.9, 946.8]	27.1	-24.2	[-85.6, 338.5]
D	Point sensor network	48.0	60.0	34.0	-43.0	-60.1	[-92.6, 141.4]	-40.0	-77.0	[-99.9, 242.4]

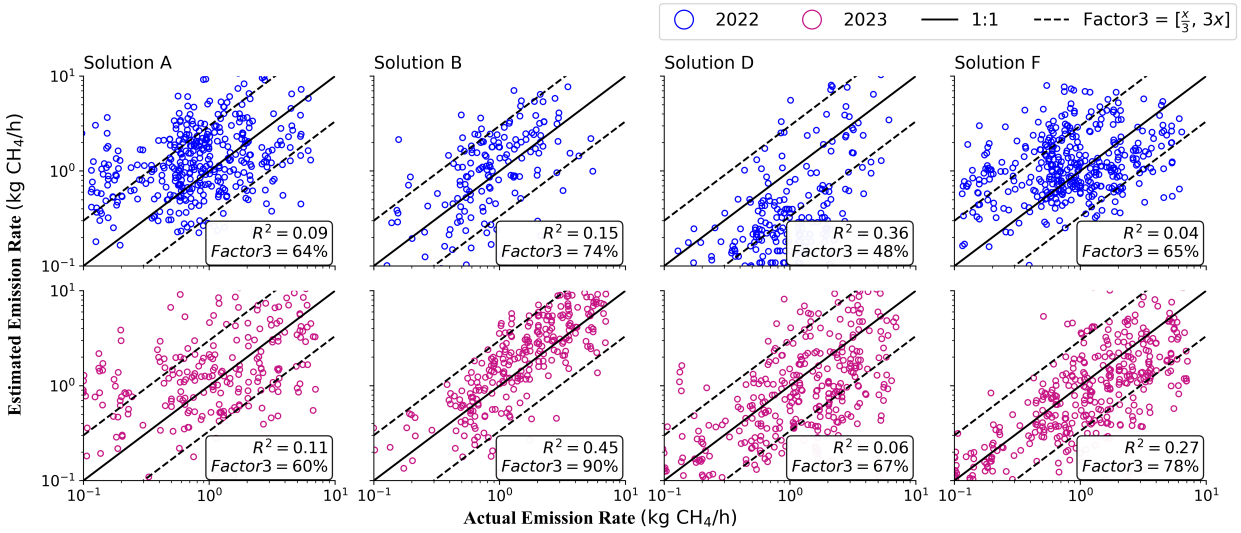


Figure 6.2: Comparison of the 1:1 and scatter plots of continuous monitors that tested twice (solutions A, B, C, and D). The x and y axis indicate reported and actual emission rates respectively. The top row data (blue) shows quantification data from the first test (2020) while the bottom row (crimson red) indicate data from the second test (2023). The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%). The R^2 illustrates the correlation between the reported and actual emission rates.

While performance varied depending on whether release rates were within (0.1–1] kg CH_4/h or >1 kg CH_4/h, all four continuous monitors demonstrated improved quantification accuracy at the facility level, where all releases and estimates were aggregated over the study period. Similarly, the operational factor of the majority of the continuous monitors also improved across both test programs. Recall that this metric captures the fraction of deployment time during which continuous monitors collect and transmit data for post processing. These improvements in performance or the lack thereof can be attributed to a few reasons.

First, some continuous monitors improved the quality of the sensors used by their solutions between the test programs, which resulted in improved performance. For example, solution D used proprietary photoacoustic-based laser absorption spectroscopy sensors in the second test to measure ambient methane concentration, unlike metal oxide sensors used in the first test. Unlike the metal oxide sensor, the laser-based sensor experiences very low sensitivity drift over time, has a quicker response time, and is highly selective/specific to methane with little or no cross-

sensitivity [153]. Cross-sensitivity by gas species (i.e., CH_4, C_2H_6, CO_2, etc.) is often responsible for FP detections by LDAQ solutions at O&G facilities.

Secondly, some continuous monitors likely upgraded their analytics and algorithms used to infer detections and estimate emissions. Third-party controlled testing allows vendors to access comprehensive ground-truth data used to troubleshoot their system and optimize their data analytics. In some cases, solutions leveraged this data to "learn" the test facility, gaining insights into potential emission sources and optimizing their performance accordingly.

Thirdly, since the detection performance of several continuous monitors generally improved with release duration and rate, longer duration and higher-rate controlled releases in the second round likely helped map the POD curve of some monitors (e.g., solution B). Longer release durations likely provided scanning/imaging monitors multiple opportunities to visualize and detect emissions, while point sensor network solutions benefited from extended averaging times of ambient concentration measurements to infer detections more accurately.

Finally, the increased sensor density in the second round of testing by solution F likely contributed to the improved performance of the solution. In contrast, the failure of one of the sensors installed by solution A, among other factors, may have contributed to the decline in its overall performance in the second test program.

6.2.2 Change in the Performance of Survey Solutions

Overall, most of the tested survey solution demonstrated performance improvements across multiple assessed metrics during repeat testing. Figure 6.3 compares the POD curves from different test programs of the 3 survey LDAQ solutions that tested twice. The blue-colored and red-crimson-colored curves indicate the first and second test detection results. Similarly, Table 6.4 compares the detection, localization, and survey times of the solutions across both test programs. As discussed in Section 2.4, a wide range of models were considered when fitting the POD curve of every LDAQ solution tested, with only the model that minimized the AICc selected. Consequently, the

POD curve functions sometimes varied across solutions and between testing rounds for the same solution (e.g., solution H).

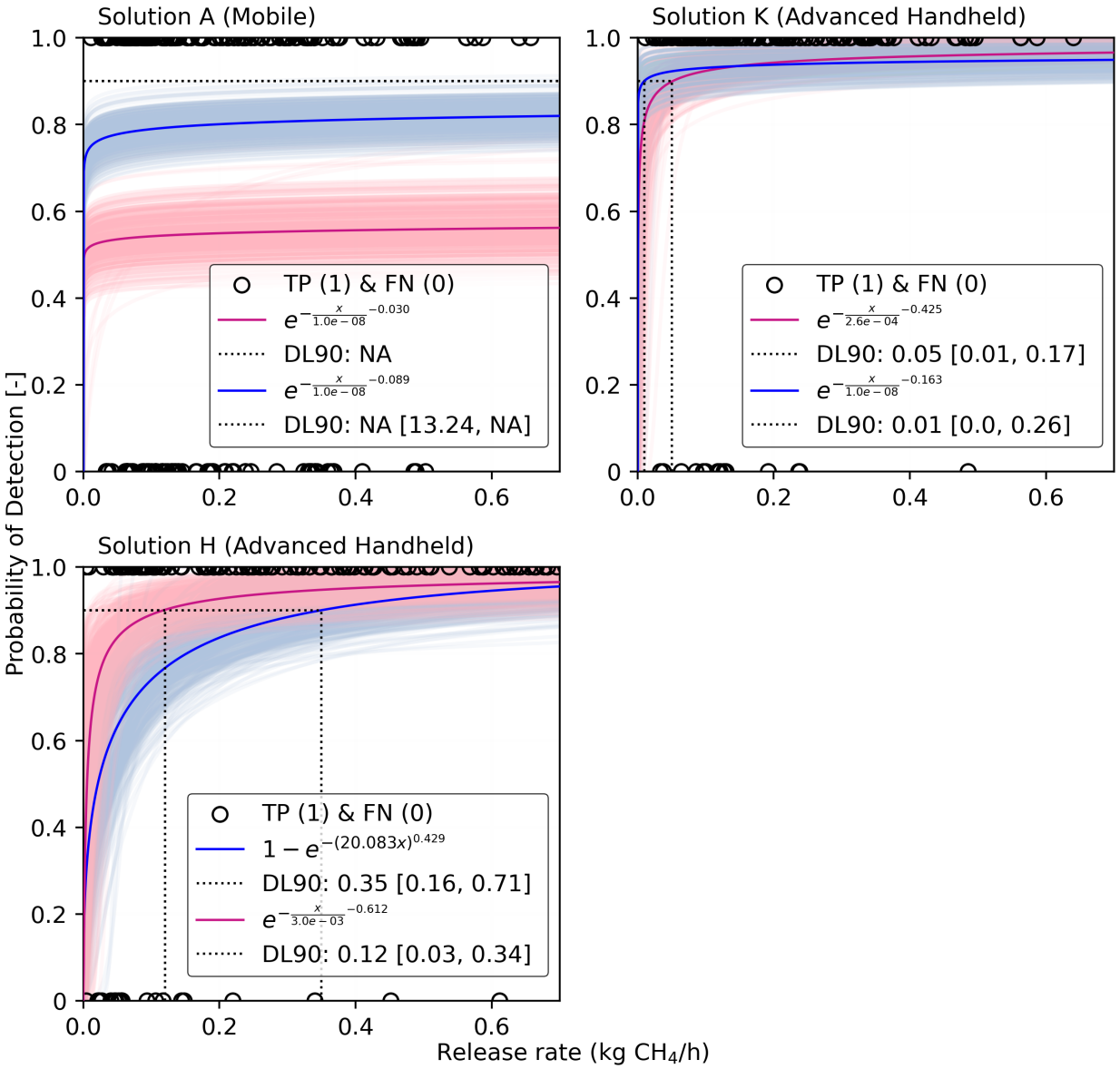


Figure 6.3: Comparison of the POD curve as a function of emission rate (kg CH_4/h) fitted using exponential functions for the three survey solutions that tested twice (A, K, and D). The DL90, which is the emission rate at which a solution achieves 90% POD, is shown together with the confidence interval. The uncertainty on the POD curve is illustrated by a cloud of curves obtained through bootstrapping, with the confidence interval evaluated as the 2.5 and 97.5 percentiles. The value "NA" is given when the DL90 or the associated uncertainty is more than $\times 20$ the maximum release rate tested or cannot be evaluated. A scatter plot with circular markers shows binary detection classification results: TP detections (1) and FN detections (0). The upper limit of the x-axis has been trimmed at 0.7 kg CH_4/h to maintain the same axis limit for all survey solutions. Note: Solution A (both first and second tests) and K (first test) have near-uniform POD across tested emission rates with odds ratios of ≈ 1.

Table 6.4: Comparison of the detection performance (FPF and FNF) and the 90% POD predicted by solutions that tested twice sorted in order of increasing 90% POD (DL90).

Solution		Survey Time[†]	Localization (Unit)		FPF (%)				DL90	
ID	Category	per Unit (minutes)	Accuracy (%)	Precision (%)	All	No Controlled Release	Excess TP Detection	FNF (%)	Estimate[‡] (kg CH_4/h)	Within Tested Range
Results from the second testing										
K	Advanced Handheld	0.51 [0.06, 1.50]	92.2	92.2	0.0	0.0	0.0	7.3	0.05 [0.01, 0.17]	Yes
H	Advanced Handheld	2.34 [0.72, 6.67]	97.9	100.0	2.1	0.0	100.0	8.7	0.12 [0.03, 0.34]	Yes
A	Mobile	1.73 [0.50, 3.67]	55.1	71.1	23.1	72.2	27.8	45.5	NA	No
Results from the first testing										
H	Advanced Handheld	5.09 [0.78, 14.0]	91.0	100.0	9.0	28.6	71.4	16.5	0.35 [0.16, 0.67]	Yes
K[*]	Advanced Handheld	2.53 [1.14, 5.00]	62.9	97.7	35.6	0.0	100.0	6.6	0.01 [0.00, 0.21]	No
A	Mobile	0.94 [0.44, 2.57]	55.0	68.0	19.2	0.0	100.0	20.3	NA [8.15, NA]	No

[†] This is time taken to survey an equipment unit in a survey area.

[‡] When the POD curve cannot evaluate the DL90 or the DL90 is $\times 20$ of the maximum release rate tested, its value is "NA". Similarly, when the lower and upper empirical 95% confidence intervals on a solution's DL90 could not be evaluated, they were given as 0 and NA, respectively.

[*] The DL90 is 0 because the POD curve is approximately constant at POD $> 90\%$.

Findings indicate that 2 (H and K) of the 3 survey solutions generally improved across all metrics assessed. Performance improvements were assessed based on lower or approximately the same DL90, FNF, FPF, and survey times per equipment unit, as well as higher or approximately the same equipment unit-level localization accuracy and precision in the second test compared to the first test. Note that the DL90 of solution K in the second round was considered an improvement despite being higher than the first-round test result. This is because, unlike in the first test (Table 4.1), a rank-biserial correlation analysis assessing emissions detectability (i.e., tendency for a TP detection) as release rate increased aligned with the findings from multivariable logistic regression. This alignment of results reinforces confidence in the inferences made from the POD curve in the second round (i.e., DL90), unlike the POD curve from the first test.

These improvements could be attributed to advancements in sensor type/quality and data analytics/algorithms. However, other factors, such as favorable (or unfavorable) measurement conditions and improved survey methodology during the second test, may also have contributed to the changes in performance. For example, Solutions A and K were retested under temperature conditions between $3°C$ and $11°C$ compared to $8°C$ and $22°C$ during the first test. A surveyor handling solution A (i.e., automobile-based) is less likely to be adversely affected by the relatively low temperature condition, unlike solution K (i.e., handheld), since the latter's operationalization involves direct inspection of suspected leaking components in open air. This hypothesis is further supported by solution K recording the lowest survey times per equipment unit obtained in the study. This performance could also be attributed to improved surveying methodology given that the walking procedure used by solution K during the second test was a modified version of the approach used in the first test.

6.3 Implications of Study Results

Comprehensive third-party controlled testing with standardized protocols is critical for robustly validating performance claims by LDAQ solution vendors. With the proliferation of various LDAQ solutions at different field-readiness levels, applications, costs, and performance trade-offs in re-

cent years, a standardized and consensus framework is needed. Such a framework would support robust inter-LDAQ solution performance comparisons and track changes in performance over time.

This study represents the first known, third-party evaluation of the change in performance of LDAQ solutions over time using the same standardized controlled testing protocol and at the same test facility. Research findings demonstrate that LDAQ solutions tend to improve with repeat testing, highlighting the likely importance of regular, extensive testing to the development of LDAQ solutions. Regular and systematic controlled testing provides high-quality ground truth data that LDAQ solution vendors and technology developers can use to optimize sensor performance and enhance data analytics/algorithms. It also helps evaluate survey and deployment methodologies, as well as the implementation processes, to achieve improved overall performance (detection, localization, and quantification) in a cost-effective manner.

However, while regular, robust controlled testing is encouraged and needed to track the rapid development of next-generation LDAQ solutions, testing methodology must also increase in rigor and quality. Standardized testing protocols like those used in this study should be regularly revised and improved to ensure test results and performance metrics are current and representative of expected performance at real O&G facilities.

Chapter 7

REVISING THE STANDARDIZED CONTROLLED TESTING PROTOCOLS[6]

7.1 Overview

Previous chapters of this work have examined and discussed the controlled testing of various next-generation LDAQ solutions, broadly categorized as continuous monitors and survey solutions, using standardized testing protocols [131, 132]. Continuous monitors operate autonomously, collecting sensor readings near-continuously over extended periods, whereas survey solutions require human support and function over shorter time frames compared to continuous monitors. These studies aim to demonstrate the application of the test protocols developed through multi-stakeholder collaboration (the first of its kind) and to learn from the process to improve the protocols. Concomitantly, the studies demonstrated the state of the art of next-generation LDAQ solutions, through trusted and independent assessment to help guide field deployment decisions of LDAQ solutions.

While testing with these standardized protocols in this study is a significant upgrade to previous controlled testing studies [85, 120–127] in terms of testing complexity, consensus, rigor, and quality, testing is still substantially non-representative of real O&G operations and applications, as demonstrated by Day et al. [154]. In summary, the study by Day et al. tested four of the continuous

[6]Chapter 7 is partially extracted from published articles:
[137] Reprinted with permission from C. Bell, C. Ilonze, A. Duggan, and D. Zimmerle, "Performance of Continuous Emission Monitoring Solutions under a Single-Blind Controlled Testing Protocol," *Environmental Science & Technology*, vol. 57, no. 14, pp. 5794–5805, 2023. DOI: 10.1021/acs.est.2c09235.Copyright 2023 American Chemical Society.
[138] Reprinted with permission from C. Ilonze, E. Emerson, A. Duggan, and D. Zimmerle, "Assessing the progress of the performance of continuous monitoring solutions under a single-blind controlled testing protocol," *Environmental Science & Technology*, vol. 58, no. 25, pp. 10941–10955, 2024. DOI: 10.1021/acs.est.3c08511. Copyright 2024 American Chemical Society.
[139] C. Ilonze, J. L. Wang, A. P. Ravikumar, and D. Zimmerle, "Methane quantification performance of the quantitative optical gas imaging (qogi) system using single-blind controlled release assessment," *Sensors*, vol. 24, no. 13, 2024. DOI: 10.3390/s24134044

monitors discussed in Chapter 3 at real O&G facilities and showed wide performance discrepancy between both studies. The continuous monitors assessed did not perform as well in the field as during controlled testing. The discrepancy was attributed to the gap in deployment strategy, facility operational conditions, and the primitiveness of solutions' data analytics for field conditions.

Many stakeholders are skeptical and have raised concerns about extrapolating controlled test results to field performance due to the protocols' simplistic controlled testing and reporting framework. For example, in real O&G facilities, the baseline emissions concentration are typically elevated and variable compared to regional or global atmospheric backgrounds; the controlled testing described previously did not simulate baseline emissions. Vented (i.e., planned/intentional) emissions from maintenance operations (e.g., equipment depressurization) or the use of gas for process controls (e.g., gas-powered pneumatic actuators) typically constitute baseline emissions. LDAQ solutions are expected to distinguish and isolate (i.e., detect, localize, and quantify) fugitive emissions or leaks from baseline emissions. The U.S. EPA's NSPS OOOOb final rule also acknowledges that LDAQ solutions intending to be approved for regulatory-compliant LDAR programs must be able to make that distinction [136]. These developments have necessitated a revision of the protocols so that testing can reflect emission scenarios and behavior expected at real O&G facilities.

This chapter discusses the limitations of the existing testing protocols, the protocols revision process, and a description of the revised protocol. This includes the revised testing methodology, data reporting, detection classification framework, and performance metrics. The chapter concludes with the implications of the revised standardized test protocol for future controlled testing programs and how test results are translated to real O&G facility applications.

7.2 Limitations of the Testing Protocol

7.2.1 Testing Process

As discussed extensively in Chapter 2, the current protocol divided testing into experiments containing one or multiple steady-rate controlled releases occurring synchronously over a specified

duration, as shown in Figure 7.1. Consecutive experiments are separated by sterile periods, with no controlled releases. This allows LDAQ solutions to identify the start and end of each controlled release event by recognizing a return to background levels for an extended period, typically a significant fraction of an hour to multiple hours.

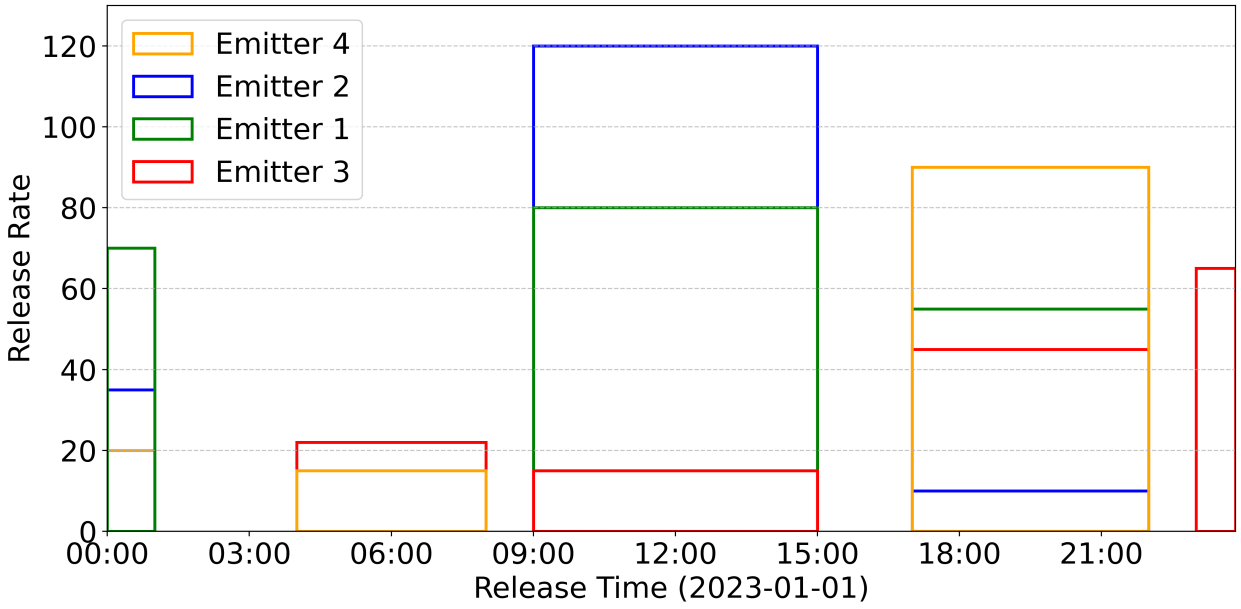


Figure 7.1: The figure shows the structure of controlled releases using the current testing protocol. Multiple, synchronous controlled releases from different emission points each have a constant release rate over the release duration. Consecutive controlled releases are separated by breaks of varying durations.

The problem with this setup is that it violates several field measurement study observations: emissions at O&G facilities are unpredictable [32,51], and temporally and spatially variable [9,15, 46–50]. While some process failures, like the breakdown of tank control systems, can persist for long durations, albeit with variable emission rates, until repaired, most fugitive leaks are intermittent and are short-lived. Emissions typically transition between on and off states asynchronously and unpredictably, unlike the cleanly demarcated emission events shown in Figure 7.1. In addition, real O&G facilities typically have largely known operational emissions from maintenance activities, combustion slip, or gas-powered process controls [27] that overlap with fugitive emissions. LDAQ solutions are expected to differentiate and isolate these two emission types—a capability

not captured in the current test protocol. These emissions characteristics make it more algorithmically challenging for LDAQ solutions to differentiate, detect, localize, and quantify emissions.

Additionally, the event-based nature of controlled releases under the current protocol required several vendors to modify their analytics to match the test conditions. This deviation from the settings used in real O&G facility deployments involves identifying the start and end times of emission events and obtaining a best constant rate estimate for each monitored asset (i.e., equipment unit). Ordinarily in the field, solution analytics must factor in that emissions start and stop asynchronously, and emission rates can vary rapidly [155].

7.2.2 Performance Metrics

One key lesson learned from the previous chapters of this work is that LDAQ solutions have different field applications, depending on their type, category, deployment, and field-readiness level. Therefore, the key challenges identified in the current protocol concentrated at the intersection between the maturity of LDAQ solutions, their use cases for field deployments, and variations in reported data. Under the current protocol, it is challenging to determine whether assessed performance reflects the immaturity of the tested system or if the test conditions simply do not replicate the solution's intended application. For example, a solution may exhibit a high FNF not necessarily because its algorithm is ineffective but because it is designed to detect only emissions above a emission rate threshold larger than most of the release rates tested. Additionally, several solutions provide facility-level monitoring and time-resolved quantification estimates, rather than detecting and resolving individual sources. That is to say, for many LDAQ solutions, performance interpretation and metrics using the current protocol are not reflective of field performance as their application is different from what was tested.

Figure 7.2 shows the implication of the current protocol by examining the pairing of the controlled releases with detection reports by a solution from a day of experiments during the study in Chapter 3. Since the current protocol allowed each controlled release to be paired with only one *EmissionSourceID* reported by each LDAQ solution—and each *EmissionSourceID* to be

paired with only one controlled release—the detection classification process (Figure 7.3) could increase/decrease the number of TP, FN, and FP notifications. This increment/decrement may occur relative to other classification methodologies that more accurately reflect the intended applications of an LDAQ solution. The caveat is that such detection reports reflect expected field performance and are not due to the inefficiency of the LDAQ solutions.

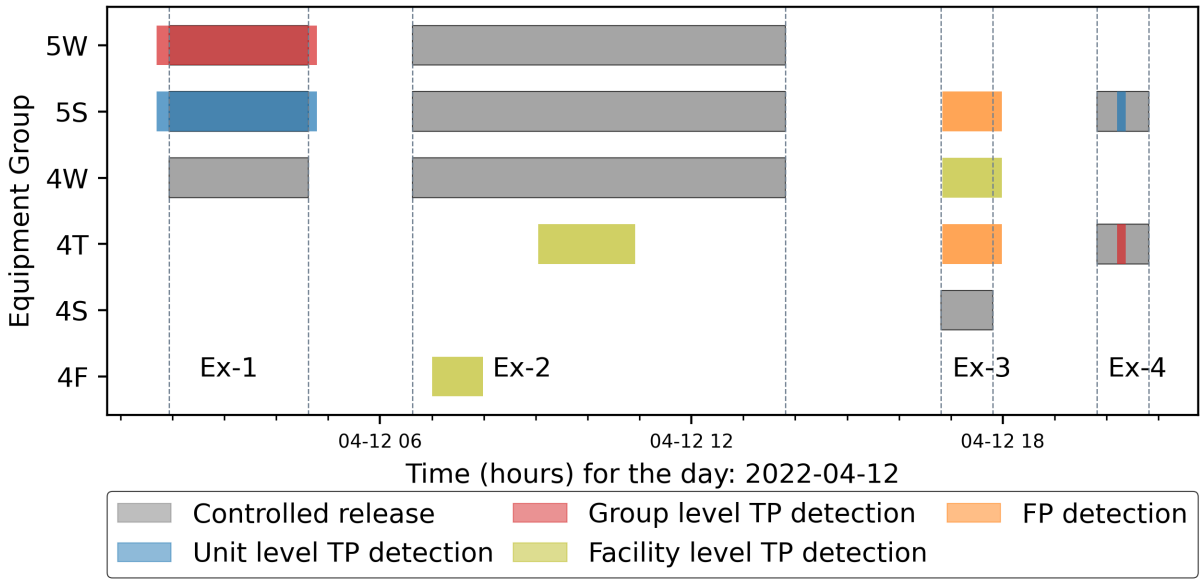


Figure 7.2: The figure shows the timing and location of controlled releases conducted by the test center and detection reports received from solution A during a 24-h period that included four experiments (Ex-1, Ex-2, Ex-3, and Ex-4).

The first experiment (**Ex-1**) illustrates how the classification methodology was intended to work. Three controlled releases from equipment groups 5W, 5S, and 4W were performed. See Section 2.1 in Chapter 2 on how to interpret the equipment group labels. The LDAQ solution reported two detection reports with *EmissionSourceID*s localized to 5W and 5S with estimated start and end times closely matching the controlled releases. These were classified as TP detections at the equipment group and unit level, respectively. The third controlled release at 4W was classified as FN since it could be paired with any detection report. The detection report was classified as a

TP at the equipment unit level if it correctly localized the emission source to the correct equipment unit. If the detection report failed to localize the emission source to the correct equipment unit but instead identified the emission on an adjacent non-emitting equipment unit, it was classified as an equipment group TP detection, as described in Figure 7.3.

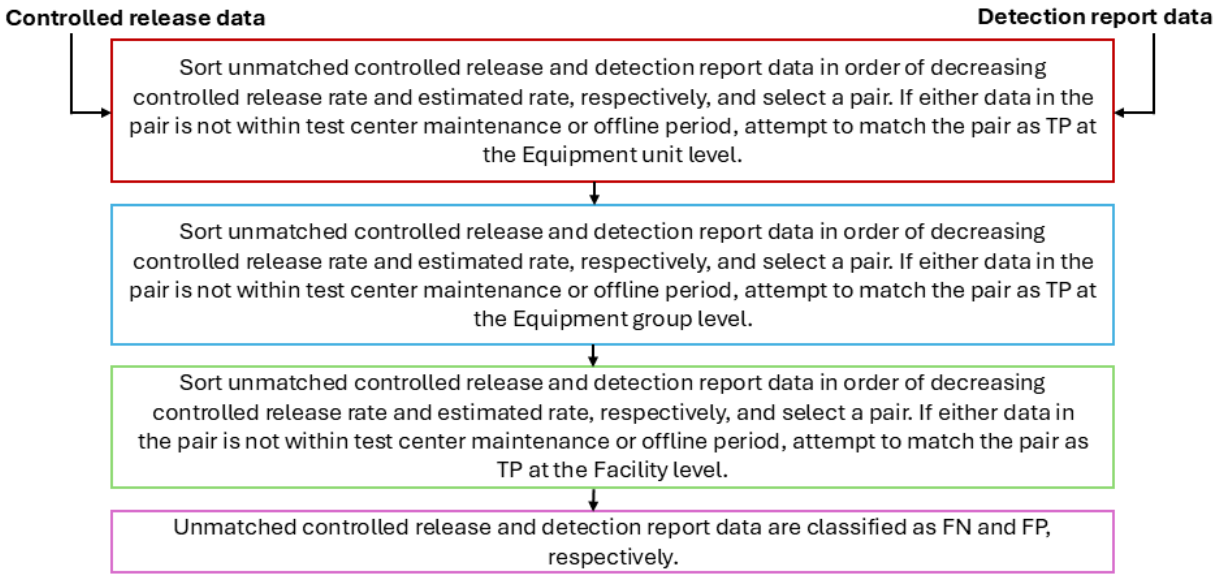


Figure 7.3: The figure shows the methodology for pairing controlled release data with detection report under the existing protocol.

The second experiment (**Ex-2**) shows a scenario when the classification methodology inflates TP detection and deflates FP detection classification. In this experiment, three controlled releases were performed, again located at 5W, 5S, and 4W. The test center received two detection reports attributing emissions to 4T and 4F. Although these emission sources were not correctly identified to the same equipment groups, the protocol matched these reports with two of the three controlled releases as facility-level TP detection classification, with the remaining controlled release classified as an FN detection. Observe that the reported detections did not accurately estimate the controlled release duration/emission event but were classified as TPs anyway. For stricter and more accurate applications and performance evaluation, only portions of the controlled release duration correctly identified would have been classified as TP detection, with the remaining controlled release du-

ration classified as FN detection classification. An alternative classification procedure for field applications in which accurately localizing identified emission sources is priority might have resulted in two FP and three FN detection classifications. Conversely, a facility-level monitoring application would have yielded only TP detection classification.

The third experiment (**Ex-3**) had one controlled release on 4S, but three emission sources were reported. The protocol's classification process produced one facility-level TP and two FP detection classifications. Looking closely at the timing, all three reported detections clearly overlapped with the controlled release period. If an alternative classification methodology had been used—one that allows multiple detection reports to be paired with a single controlled release (e.g., facility-level monitoring)—the result would have been three TP detection classifications. However, none of these detections would have correctly identified the emitting equipment group or unit.

The last experiment (**Ex-4**), with two controlled releases and two detection reports, produced two TP detection classifications: one at the equipment group level and the other at the equipment unit level. Given that the equipment unit level is the highest spatial resolution at which emissions sources can be localized in the study, an application that prioritizes accurate localization performance along with accurate detection in time might have produced a different classification result.

Another limitation of the current protocol is its assumption that LDAQ solutions infer detections whenever ambient emission concentrations exceeds a specified threshold (in this case, atmospheric concentration). While this might be true for some LDAQ solutions, others implement a wide variety of detection criteria like rolling average measurement over a specified measurement window. For example, an LDAQ solution might attempt to perform detection after averaging 24 hours' worth of measurement data on a rolling average basis instead of using instant concentration enhancement measurements.

7.3 Protocol Revision Process

Recall that the current protocols for continuous monitors and survey solutions were developed through a collaborative effort between CSU and a multistakeholder PDC. The PDC comprises

of at least 75 members representing various sectors, including LDAQ solution developers, O&G operators, academia, regulatory agencies, O&G associations, environmental NGOs, and consultants. Figure 7.4 summarized the process and timeline exhausted in revising the protocols into one standardized document. The process includes multiple virtual group meetings with the PDC, one separate in-person working session with LDAQ solution developers and operators, two in-person working sessions with the U.S. EPA, one in-person working session with TotalEnergies Transverse Anomally Detection Initiatives (TADI), multiple working sessions within the CSU team, and one protocol public review cycle. The structure of the meetings was intended to limit bias, encourage honest discussions and feedback, and allow for better management of deliberations during the meetings.

The protocol revision process officially began with virtual meetings that took place between 13 and 31 November 2023 to reactivate the PDC and accept new members. Discussions during these meetings centered on updates of test results using the current protocol, observed challenges and limitations of the protocol, and general feedback from the PDC on ways to improve testing. Figure 7.5 shows the number of participants from each stakeholder group in all these meetings.

The CSU team met with approximately 30 O&G operators in Houston, Texas, on December 4, 2023, to discuss their current application of LDAQ solutions in their facilities and their expectations of the solutions. Specifically, the meeting explored how next-generation LDAQ solutions were deployed by operators to achieve their desired objectives. It also focused on identifying key performance metrics, the types of data that support emissions mitigation goals, and potential improvements to test protocols to better assist operators in achieving their intended emissions reduction objectives. The meeting was organized into a presentation followed by 4 break-out groups of about 7 members, each group moderated by a member of the CSU team. Similar to Fox et al. [119], the presentation was used to introduce the current protocol (i.e., testing and performance metrics), test results (both controlled and field testing), and the gap between both test results with possible reasons. Each group, deliberated on the aims of the meeting and concluded with suggestions

on bridging the gap between controlled and field testing results and the associated performance metrics.

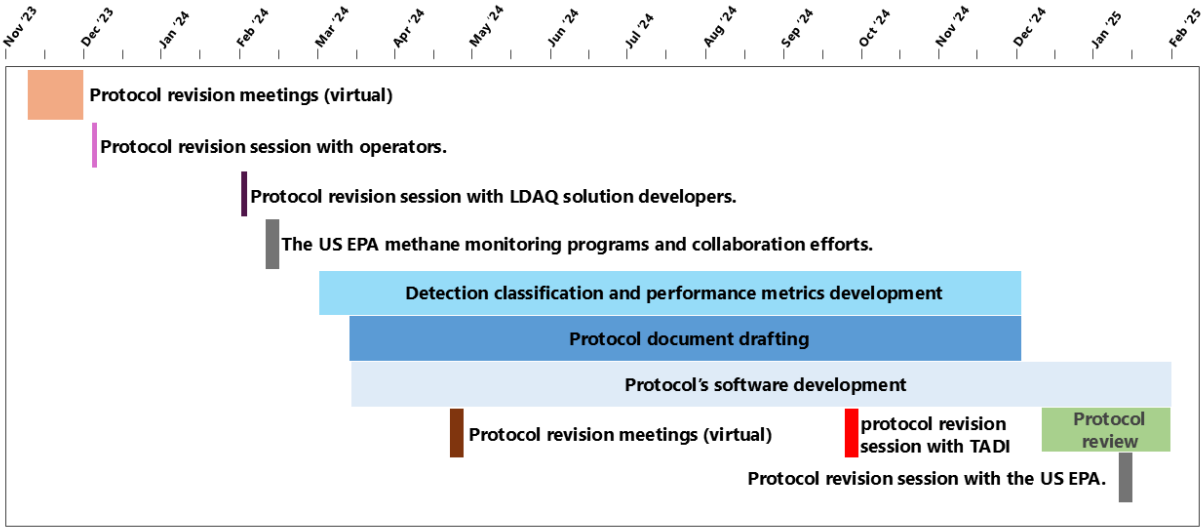


Figure 7.4: The figure shows the timeline and summarizing the protocol revision process

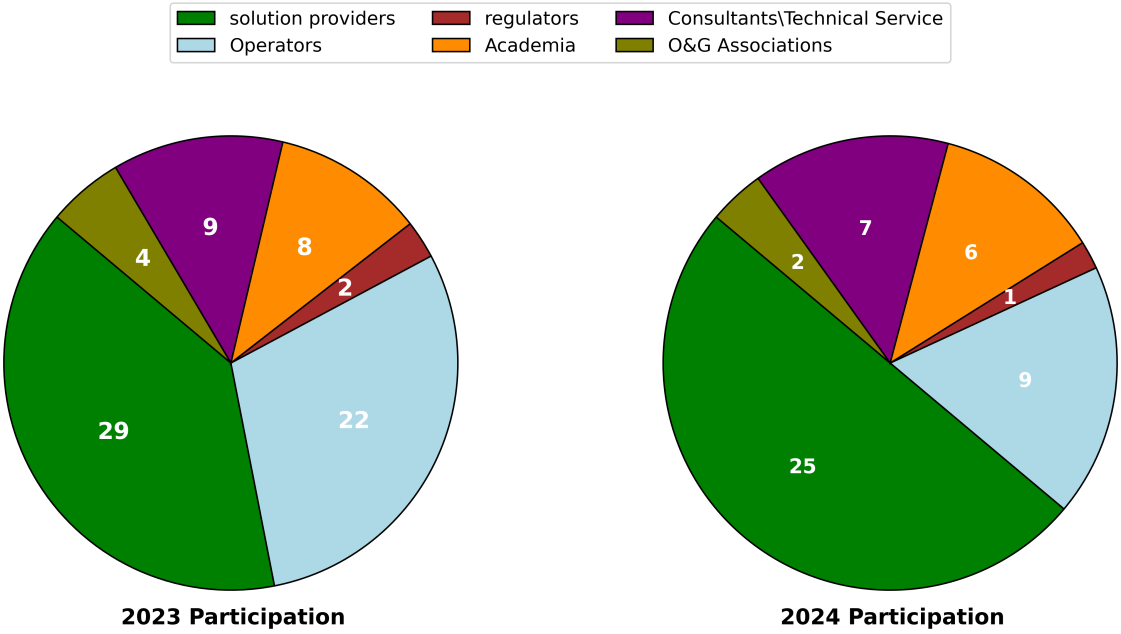


Figure 7.5: The figure shows the number of protocol revision meeting participants by stakeholders.

At the end of the deliberations, all suggestions were collated and discussed with every participants in an open forum for feedback and clarification.

A similar meeting was held with approximately 20 LDAQ solution developers and vendors at the University of Texas in Austin, Texas, on February 2, 2024. Discussions focused on 3 areas: LDAQ solutions deployment and objectives, performance metrics, and improving testing complexity. Similar to the meeting with operators, the session began with a presentation before breaking into 3 working groups of about 7 members each. Unlike the operators, after every working session on each topical area, all the participants reconvened to be briefed about the ideas and suggestions from each group by their representatives. An open forum was opened to discuss the ideas and suggestions while the CSU team documented consensus key points. The working session on deployment modalities and objectives aimed to understand how LDAQ solution developers utilize their systems in the field and the type/format of data typically provided to operators. The session on performance metrics focused on developing a standardized detection classification methodology and metrics that account for the diverse use cases of these solutions. Lastly, discussions on testing complexity highlighted various testing configurations to better assess the expected field applications of solutions.

The two-day workshop with the US EPA on February 14–15, 2024, included discussions on the current protocol's metrics, potential revisions, and the metrics used by the EPA for reviewing LDAQ solutions. Additionally, the CSU team organized virtual meetings with the PDC on April 23–24, 2024, to provide updates on stakeholder discussions and gather perspectives. To advance internationally recognized test protocols, the CSU team also met with representatives from TotalEnergies TADI in Pau, France, from September 23–27, 2024. The meeting focused on variations in testing protocol and performance across test centers and how to develop a unified and standardized testing and assessment framework.

On January 14 and 15, 2025, the CSU team had a two-day working session with a 5-man team from the U.S. EPA at CSU, Fort Collins, Colorado. Discussions primarily centered on reviewing the revised standardized testing protocol draft developed by CSU. The objective was to evaluate the

practicality of the proposed testing methodology and determine how the protocol could be applied to assess LDAQ solutions for approval in line with the NSPS OOOOb rules for advanced LDAQ solutions [136].

The feedback, suggestions, and ideas from all these meetings were harmonized through multiple internal working sessions by the CSU team to produce the draft document of the revised testing protocol. The CSU team developed the document concurrently with the detection classification algorithm, the associated performance metrics, and the software for data analysis. The protocol document was made available on METEC's website for 46 days to gather public contributions and peer reviews. During this period, the protocol document received approximately 40 comments for revision.

7.4 The Revised Protocol Framework

The revised standardized test protocol evaluates the performance of LDAQ solutions and methodologies designed to detect leaks/fugitive emissions. The current application of the revised protocol is grouped into survey solutions and continuous monitors (or methods) with possible extensions to other methods. The performance of LDAQ solutions can be assessed at two testable interfaces. Testable interface 1 allows for a direct comparison of known emissions at the test facility with what was reported by an LDAQ solution. Testable interface 2 allows for the assessment of an LDAQ solution's ability to alert an operator of specific emission events that might indicate a failure.

7.4.1 Definitions

The revised protocol introduced new terminologies to the current protocol; some of the definitions are as follows:

1. **Simulated Facility:** Include all or a portion of a test facility, at which, during testing, the emissions behavior and configuration of an operational real facility are simulated.

2. **Baseline Emissions:** Emissions that simulate normal operating conditions at the simulated facility, i.e., emissions during no-failure conditions.

3. **Baseline Controlled Release (BCR):** Controlled releases simulating baseline emissions during testing.

4. **Baseline Period:** The period over which all simulated emissions are considered simulated baseline emissions.

5. **Detection:** Estimated quantities indicating that an emission has been observed above an LDAQ solution-defined threshold.

6. **Analysis Spatial Extent (ASE):** The spatial boundary within the test facility which will be used for performance analysis.

7. **Estimated Quantity (EQ):** Direct measurements and derived quantities obtained by the LDAQ solution.

8. **Performer:** A single solution testing under the protocol, including all elements in the definition of an LDAQ solution, supporting performer personnel, logistics, and other equipment.

9. **Threshold:** A set of criteria specified by a performer that is required to convert measured quantities into leak detections.

10. **Simulation Period:** The duration when emissions include both baseline emissions and simulated failure conditions.

11. **Failure Controlled Release (FCR):** Controlled release which simulates a failure condition.

12. **Test Program:** A testing project where multiple LDAQ solutions are tested under similar conditions by one test center to produce comparable results across multiple Solutions. A program may include periods where multiple LDAQ solutions are deployed at the test center simultaneously, or a series of deployments of LDAQ solutions.

7.4.2 Testing Process

Testing is divided into 3 activities: setup, operation, and reporting. The setup phase is targeted at analogously recreating typical field deployment of LDAQ solutions at the test facility. There-

fore, such a setup must be clearly and completely documented to allow easy comparison with field deployments. This is because LDAQ solutions' test results apply and extrapolate to real facilities with context to documented and tested conditions. In other words, the use of test results to inform field deployment decisions will rely on whether the same sensor density, sensor type, onsite personnel, version of analytics/algorithm, etc., deployed during testing is deployed in the field. At a minimum, the performer must document the following: the model number and power configuration of each component used onsite, the height and GPS coordinates of all fixed components/sensors, the version number of the software installed in the components, and the thresholding conditions for evaluating LDAQ solution data. For continuous monitors, additional information like the LDAQ solution's data reporting format and strategy will be required. For survey solutions, extra information like the detailed descriptions of the system, deployment methodology, and the personnel used by the performer during emissions surveys will be needed.

During the operation period, which can be divided into baseline and simulation periods, testing will acquire all data necessary to calculate required and optional metrics. The protocol defined separate timeframes for baseline and simulation periods, aligning with the testing modes required by the NSPS OOOOb rules [136]. However, not all test programs will need baseline periods. During the baseline period, both the test center and the performer will independently characterize the baseline typically as a mean value. The test center will strive to simulate baseline emissions in terms of rates, patterns, timing, and frequency from continuous and intermittent sources to mirror emissions conditions at similar facilities or facility types in field deployment. Testing during the simulation period allows for the assessment of LDAQ solutions' ability to identify failure conditions at facilities with non-trivial baseline emissions, a characteristic not captured in the current test protocol. Unlike the baseline emissions, simulated FCRs will cover a broader range of emitter locations, release durations, rates, and frequencies than typically observed in the field. This approach aims to evaluate LDAQ solutions under diverse test conditions and environmental variables while considering the complexity constraints of the test facility and test program duration. FCRs will aim to address the requirements of three key performance metric categories:

- **POD:** Controlled releases should vary emission rates and duration sufficiently such that testing extends from near-zero to near-100% detection probability for all LDAQ solutions participating in a test program.

- **Quantification accuracy:** Controlled releases should vary across the release rates range required to meet the test program's objectives.

- **Localization accuracy:** If localization metrics below the facility level are to be assessed, controlled releases should be clustered in specific sub-sections of the simulated facility to allow localization metrics computation.

FCRs during the simulation periods will be conducted single-blind (i.e., emitter location, timing, and release rate are unknown to the performers) and may include single or multiple controlled releases that may be asynchronous, intermittent, and with variable emission rates as illustrated in Figure 7.6. Simulation periods will also include times with no FCRs, and FCRs will be additive to ongoing BCRs, as the commencement of an FCR will not cause a BCR to stop emitting.

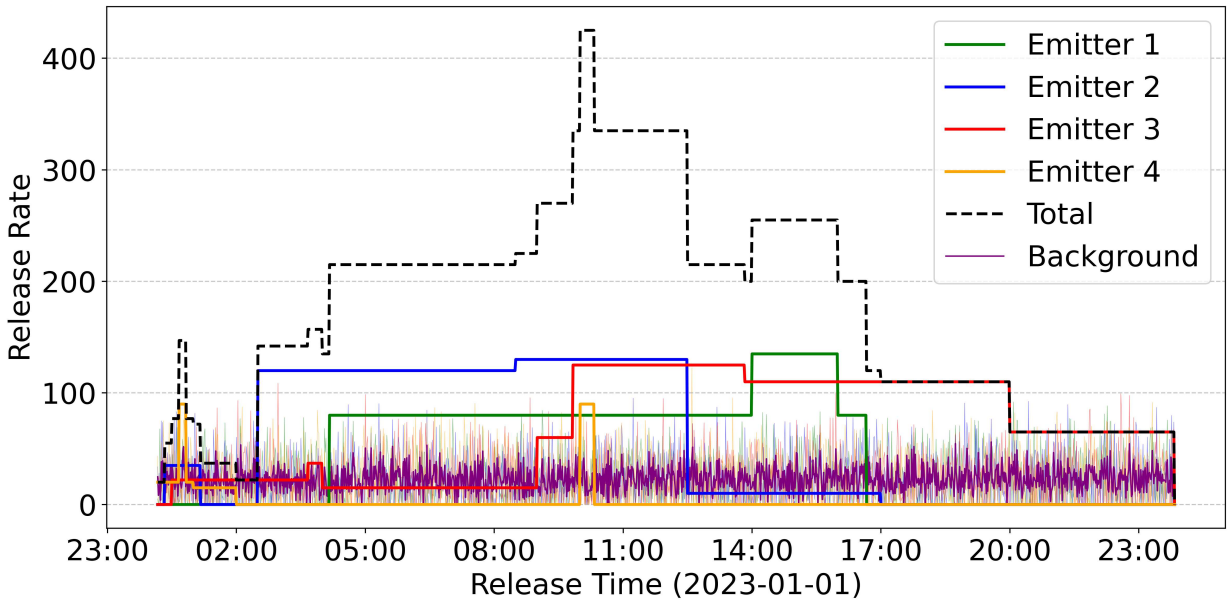


Figure 7.6: The figure shows the structure of controlled releases using the revised testing protocol. Controlled releases are asynchronous with varying release rates over each release duration. The total FCR rate at any time and simulated baseline emission are also shown.

This implies that the total release rate at any time during testing is the sum of ongoing FCRs and BCRs at that given time. This controlled release profile mirrors the key emissions behaviors common across the natural gas supply chain, which were previously overlooked in the current testing protocol. They are O&G facilities typically exhibit normal operational non-zero emissions, and emissions are highly variable in both emission rate and timing. To be able to account for interruptions or system failures while testing during performance analysis, the test center must maintain a record of operational exclusions, documenting the timing of controlled releases that do not comply with the revised protocol's standards, as well as facility downtime due to maintenance, planned outages, or extreme weather conditions.

7.4.3 Test Facility Data Exclusion

To account for interruptions or system failures during performance analysis, the test center must maintain a detailed record of operational exclusions. This includes documenting the timing of controlled releases that do not comply with the revised protocol's standards, as well as any facility downtime caused by maintenance, planned outages, or extreme weather conditions.

7.4.4 Data Reporting

Performers must electronically submit observation reports for each observation to the test center during testing, using computer-readable format(s) specified by the test center. Under the revised protocol, observation reports can be periodic or episodic. Periodic observation reports are continuous time series data analyzed over a defined period, while episodic observation reports capture discrete time intervals with gaps in-between. Most continuous monitors, particularly point sensor networks, provide continuous time series data (disregarding system downtime), often displayed on a web-based dashboard showing concentration or volumetric measurements over time. In contrast, survey- and camera-based LDAQ solutions typically report observations episodically, either through snapshots or short-duration emission measurements which may be arbitrarily spaced. To support the wide range of LDAQ solution types testable under the revised protocol, two reporting modes are supported, as shown in Figure 7.7:

- **Mode 1:** In this mode, LDAQ solutions report estimated quantities. This is typically useful for emissions accounting applications where estimated quantities like quantified emission rates are important or a priority.

- **Mode 2:** In this mode, LDAQ solutions report yes/no based on the outcome of performer-defined thresholding criteria. Most detection alerts by LDAQ solutions in the field operate in this mode.

Each observation report is independent with a unique identifier (*ObservationReportID*), includes a location specifier (a 3D enclosed volume) that indicates the spatial extent over which the report is valid, and may overlap in time and space with other observation reports. Observation reports must also include the start and end time (*StartEnd*), the gas species measured for detection (*Gas*), and whether the report will be used for sample-based or integrative metrics or both.

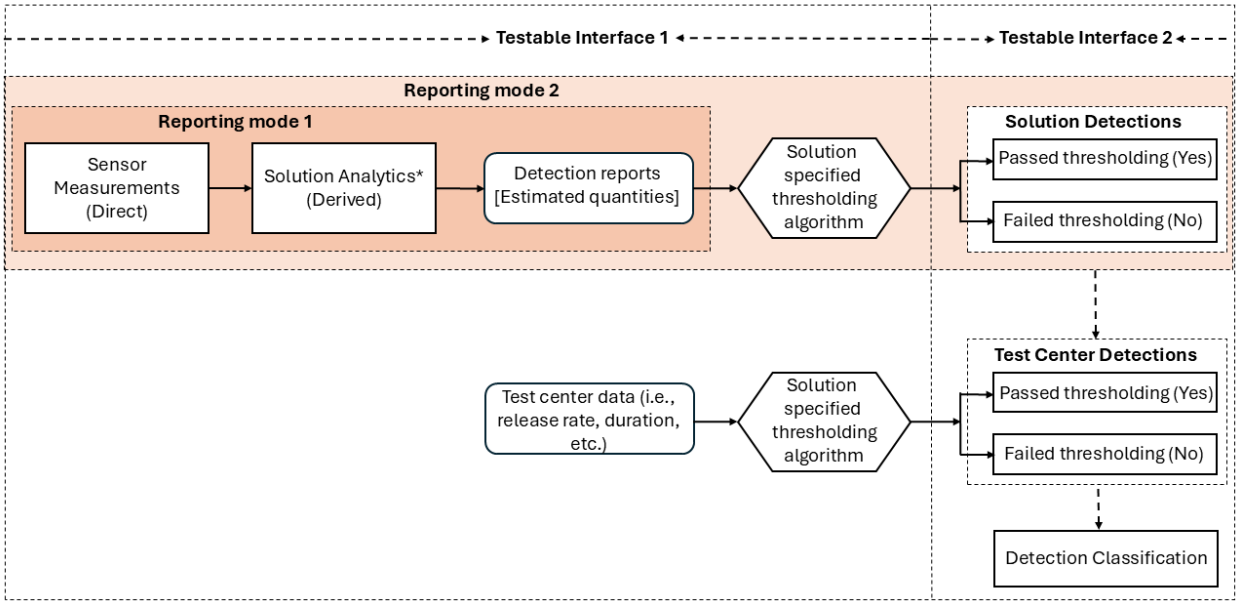


Figure 7.7: The figure shows the revised protocol data reporting and thresholding application process.

Sample-based metrics compare observations over a performer-specified period with test center data from the same period. This approach is useful for applications that rely on a small number of

observations to trigger actionable decisions, such as dispatching a field crew with OGI cameras to investigate a detection alert. Integrative-based metrics compare the time-weighted sum of observation reports with the time-weighted sum of test center data over a defined period. Many O&G operators use time-integrated emissions data, often on an annual basis, to comply with emissions regulations in the U.S. and Canada [156], hence the potential interest in this metrics category.

Under reporting mode 1, performers provide the estimated emission rate (*EmissionRate*) for each observation. Under reporting mode 2, a yes/no indicator (*Detection*) is submitted based on the thresholding result from Testable Interface 2 in Figure 7.7. As highlighted earlier, unlike the current protocol, the revised protocol requires performers to submit their detection criteria or algorithms to the test center. These criteria are then used to apply thresholding to both the observation reports (Mode 1) and the test center's data (regardless of the reporting mode). For example, as a thresholding criteria, the detection value of an observation report or test center data is "yes" if the estimated emission rate/release rate is $>$ x kg CH_4/h above baseline emissions; otherwise, the value is "No".

When necessary, performers are also required to report offline records stipulating the period during which an LDAQ solution is unable to detect and/or estimate emissions. Operating parameters differ substantially between solutions, resulting in both planned and unplanned offline periods. Typical offline reasons include system failure, loss of data communication, or inoperable environmental conditions including insufficient light (imaging solutions), winds outside operational parameters, or precipitation. For LDAQ solutions reporting episodically, offline reports must be provided for gaps in between observations. This ensures that all metrics reflect the performance of the tested LDAQ solution only when it is online and the test center is operational.

7.4.5 Detection Classification

Metrics require comparison, at some time interval, between observations made by the solution and release measurements made by the test center. Additionally, development of *probability of detection* requires that both the test center data and Mode 1 observation reports must be thresholded

based on the performer-provided criteria, as shown in Figure 7.7. Unlike the current protocol, this approach is designed to reflect the detection inference methodology used by LDAQ solutions during actual field deployments, allowing for more accurate extrapolation of test results to their typical operational objectives. To standardize the detection classification process and subsequent metric calculations, several data analysis requirements and steps are imposed on all input data, as illustrated in Figure 7.8.

1. **Spatial filtering:** An ASE is defined for every analysis, representing either the entire test facility or a subset. Only test center data and observation reports with location specifiers that overlap with the ASE are included in the analysis. While metrics are defined for one ASE, results from multiple ASEs may be aggregated to report metrics for similar applications. For example, the performance of an equipment unit-specific LDAQ solution could be evaluated for multiple ASEs, each of which encloses exactly one equipment unit.

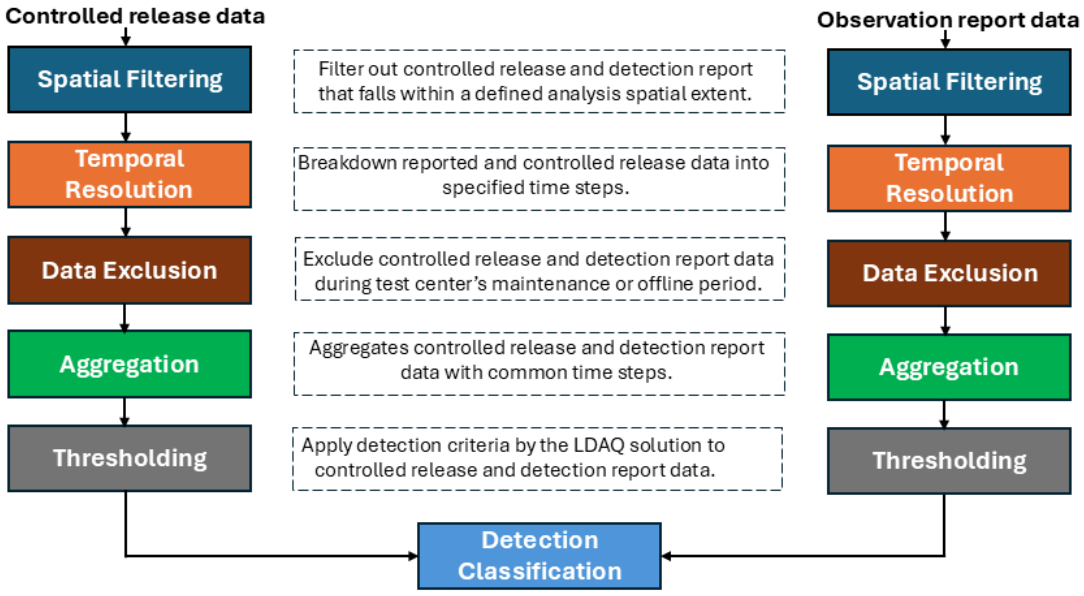


Figure 7.8: The figure shows the methodology for pairing controlled release data with detection report under the revised testing protocol.

2. **Temporal resolution:** The test center determines the time step or resolution for the testing program and analysis (typically set to 1 minute). Observation report start and end times from step 1 are rounded to the nearest test center timestamps before being segmented into the same time steps as the test center data. The test center generates data for each emission location and time step defined for the test program/analysis. For example, if the test center conducted a 1000-minute test program at 1-minute resolution using 10 emission locations, then 10,000 test center data points will be generated.

3. **Data exclusion:** Time-resolved test center data and observational reports from step 2 are processed to eliminate offline periods and test center downtimes. Every observation report that temporally overlaps with any offline report is completely removed, while only the portion of the observation report duration overlapping the test center downtime is removed. For test center data, only the portion of the time interval overlapping the test center downtime or solution offline period is excluded. As stipulated earlier, performers reporting episodically must provide offline records for periods in between observations for accurate detection classification and metrics assessment.

4. **Aggregation:** Since some test center data points from Step 3 may share common time series due to some FCRs having overlapping release durations, the overlapping time steps are aggregated into a single, unified time series with unique timestamps. The same process is applied to observation reports from Step 3, ensuring a unified and comprehensive time series dataset. While aggregation typically involves summation for test center emission rates and estimated emission rates (for Mode 1 reporting), alternative statistical operations — such as computing the mean, minimum, or maximum values — may be applied depending on the data of interest. For example, if an analysis requires test center wind speed data during testing, then averaging wind speed values associated with overlapping time steps may be the most appropriate method to obtain representative wind data for each unique test center time step.

5. **Thresholding:** This applies performer-defined detection criteria to classify test center data and mode 1 observational reports from Step 4 as either detected (yes) or not detected (no). This process is essential for computing all detection-based metrics. While the test center can provide default thresholding criteria, performers are expected to submit criteria of reasonable complexity. These criteria may be provided as a set of conditions or as software integrable with the test center's data analysis infrastructure. To ensure consistency and prevent data mismatches during detection classification, the same thresholding criteria must be applied to both test center data and observation reports at the same stage of analysis. Examples are informative:

    - *Simple thresholding:* Each observation is examined separately, and an emitter is detected if the reported emission rate is above a set value; the same criterion can be applied to the test center data. For a solution reporting on a 15 minute interval, this produces a series of 15-minute detect/non-detects from observations that can be matched with thresholded test center data at the processing data rate - typically 1 Hz.
    - *Persistence thresholding:* The solution defines a more complex classification algorithm that requires examination of multiple reports. For example, a *detection* may be defined as 3 hours where every emission report (12 total if reporting at 15 minute intervals) is two standard deviations above the mean baseline emission rate estimated by the solution. In this case both the baseline mean and the baseline standard deviation were set during examination of the baseline period. Periodic observations are marked as detected retroactively: The detection is calculated 3 hours after the start of elevated emissions, and the 12 reports that contributed to that determination are marked as 'detected' retroactively when a detection is triggered. Again, the same algorithm can be applied to the test center data.

6. **Detection classification:** Thresholded test center data and observation reports from step 5 are compared for every timestep to produce detection classification results summarized in Table 7.1.

Table 7.1: Summary of detection classification.

Thresholded Test Center Data	**Thresholded Observation Report**	
	Not Detected (No) or Null	**Detected (Yes)**
Not Detected (No)	True Negative (TN)	False Positive (FP)
Detected (Yes)	False Negative (FN)	True Positive (TP)

Since observation report timestamps are rounded up to the nearest test center data timestamp, edge cases may arise where certain time steps present in the test center data are absent from the observation reports, resulting in null thresholding values.

7.4.6 Performance Metrics

One key improvement in the metrics under the revised protocol is their alignment with the operational objectives defined by LDAQ solutions rather than those set by the test center. This shift helps explain the gap between controlled test results under the current protocol and field trial outcomes of the same LDAQ solutions [154]. By emphasizing performer-defined operational objectives, the revised protocol improves the applicability of test results and ensures more reliable and representative extrapolation to field conditions, assuming similar deployment characteristics and facility configurations between the test center and the field. For example, the revised protocol improves field representation by decoupling localization from the detection classification algorithm, which previously led to non-representative localization estimates. Instead, localization is now based on the ASE defined for a given analysis, allowing for a more robust assessment of localization performance independent of the detection classification process. Some of the metrics captured in the protocol are summarized below:

1. **Sample-based metric**

- **POD:** This evaluates the ability of an LDAQ solution to detect emissions and alert a user of the solution that a response may be required. POD will be represented as a curve (for a single independent variable) or a surface (for multiple independent variables), fitted to binary detection outcomes (e.g., TP/FN) on the dependent axis. The curve should asymptotically approach 100% detection as emission rate tends to infinity.
- **Quantification Accuracy:** This evaluates the accuracy of emission rate estimates by comparing each observation report estimate with the controlled release rate it is paired with during the detection classification process.
- **Time to Detection (TTD):** This metric evaluates the delay between the start of a detected emission event and the moment an LDAQ solution identifies it. Unlike the current protocol, it is based solely on the detection timing of the LDAQ solution and does not account for transmission delays to the test center. The methodology, illustrated in Figure 7.9, compares the start of a detection by the test center and an LDAQ solution. The analysis produces a list of TTD values, which can be presented as a histogram.

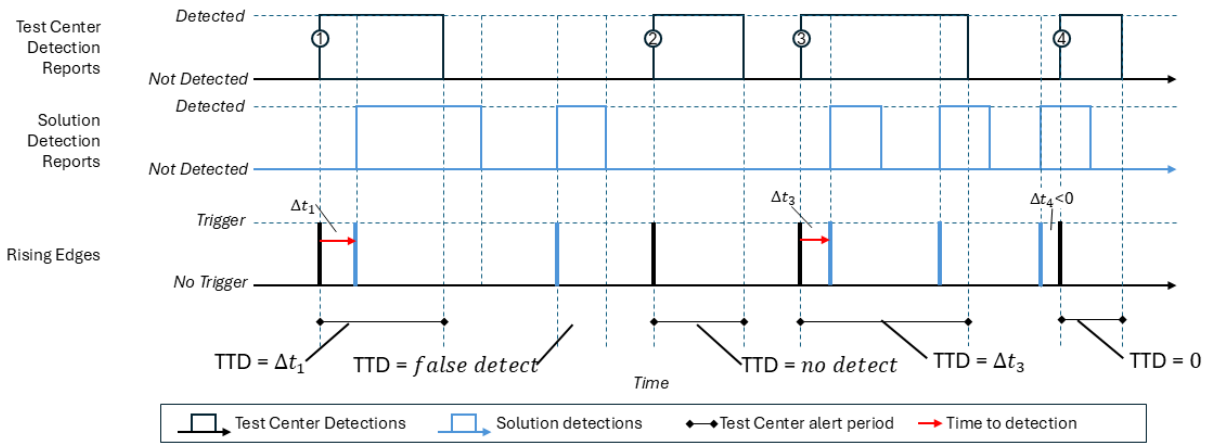


Figure 7.9: The figure shows the evaluation of time to detection under the revised protocol.

- **Time to Alert (TTA):** Evaluates the delay between the start of a detected emission event and when the test center receives the first observation report identifying it.

2. **Integrative metrics**

    - **Total Quantification over Time:** Compares aggregate emissions estimated by an LDAQ solution to those released by the test center over a specified duration.

3. **Operational metrics**

    - **Offline Time:** The fraction of the test center's total online (i.e., active and operational) time during which an LDAQ solution was offline.
    - **Reporting Delay:** The delay between the end of an observation and the test center's receipt of that observation report.
    - **Survey Time:** For survey LDAQ solutions only, survey time is the duration between the test center's initiation of FCRs and the performer's completion of a survey.

4. **Localization metrics:** The localization performance of LDAQ solutions will be evaluated by computing the same metrics for different ASEs within the simulated facility. Metrics can be aggregated across multiple ASEs of a similar type to assess localization characteristics of interest to performers.

7.5 Challenges and Limitations of the Revised Protocol

1. The revised protocol is designed to assess the performance of LDAQ solutions deployed at O&G facilities across all sectors of the natural gas supply chain. However, like the current protocol, it does not evaluate individual system components—such as sensors or analytics—in isolation. Instead, it assesses LDAQ solutions as integrated systems, which includes sensing infrastructure, analytics, deployment, data communication, and human support required for operationalization.

2. Since the protocol accounts for all emissions during implementation, it does not cover challenge testing or field trials where unknown operational emissions may be present at real O&G facilities.

3. The extrapolation of test results obtained using this protocol is valid and applicable only if the LDAQ solution proposed for field deployment matches the tested configuration. This includes maintaining the same sensor density, sensor type, deployment approach, on-site personnel, algorithm revision, and other key factors. Deploying additional or higher-performance sensors, improved analytics, increased human supervision, or slower survey speeds during testing may compromise the applicability of results. Such deviations from typical field deployment conditions could invalidate the results for method certifications, contracting, regulatory applications, or other intended uses.

7.6 Requirements of the Test Center

Controlled test results from any given test center are only as relevant and applicable as the extent to which the center accurately simulates conditions at a real O&G facility. If the test center fails to replicate field conditions, the results—though computed for the tested scenario—will have limited applicability to actual field deployments. Therefore, accurately reflecting field conditions during testing is essential; however, it requires careful consideration of both cost implications and the capabilities of the test center. Establishing baseline and simulation periods demands extensive and accurately metered gas releases, as well as complex logistics, which can be cost-prohibitive, especially over long durations. These challenges may be more manageable for test centers equipped with advanced, automated emissions control systems than those without such capabilities.

7.7 Implications of the Protocol Revision

While testing under the current protocol has significantly improved the understanding of next-generation LDAQ solutions, controlled testing results still do not represent performance during deployments at real O&G facilities. The revised standardized protocol developed in this study marks a significant step toward bridging the gap between controlled testing and field performance of LDAQ solutions.

To address key limitations in the current protocol, the revised protocol introduces several fundamental changes. Most notably, performers are now required to explicitly define how their solutions are intended to be used so that testing can be tailored accordingly. This includes determining the criteria for defining a detection, the spatial scale for emission source localization, and the frequency of reporting—responsibilities that previously fell to the test center. Shifting these responsibilities to the performer helps in addressing a major shortcoming of the current protocol: the difficulty in determining whether poor performance reflects the actual capability of the solution or due to a mismatch between how the solution was tested and how it is intended to be used.

Since the revised protocol tests LDAQ solutions in a manner that mirrors field applications and under conditions representative of real O&G facilities, O&G operators can reliably evaluate cost-performance trade-offs—an inevitable consideration in actual field deployment—based on controlled test results. By contextualizing performance with deployment characteristics such as the number and type of sensors and auxiliary equipment deployed, personnel requirements, frequency of survey flights, etc., users including O&G operators can better determine whether a solution offers an attractive balance between cost and performance.

For test centers, capability requirements have not increased dramatically to support more realistic emissions and potentially much more extended test periods to accumulate sufficient data for metrics. However, test programs must consider the additional cost implications and test center limitations before implementing the revised test protocol. If a test center is unable to replicate real O&G facility conditions, then the controlled test results generated cannot be reliably extrapolated to reflect actual performance in real O&G facilities.

As LDAQ solutions evolve and improve, there is growing advocacy for using these solutions as alternatives to traditional, regulatory-backed techniques like OGI camera surveys for emissions mitigation programs. Currently, the stakes are high in reliably predicting the likely performance of LDAQ solutions in field deployments based on controlled test results. While this revised standardized protocol aims to achieve this objective, regular protocol revisions are encouraged to keep pace with the rapidly changing O&G and LDAQ solution development sectors.

Chapter 8

CONCLUSIONS AND RECOMMENDATIONS

8.1 Overview

This study conducted controlled testing of next-generation LDAQ solutions using standardized protocols to evaluate their performance. Based on observed limitations and feedback from stakeholders in the O&G industry, the test protocols were subsequently revised to better reflect field conditions and deployment applications. In detail, this work started (as discussed in Chapter 3) with the use of a standardized testing protocol developed for continuous monitors to test 16 LDAQ solutions for the first time at METEC. Recall, continuous monitors are LDAQ solutions that operate autonomously and collect sensor readings of ambient data near-continuously for an extended period. METEC is an outdoor testing facility at CSU that mimics emissions behavior from O&G production facilities. Tested continuous monitors were broadly categorized as point sensor network and scanning/imaging LDAQ solutions depending on the ambient concentration sensing methodology. Point sensor network continuous monitors consist of multiple point concentration measurement sensors, typically positioned along the fenceline of an O&G asset. Scanning/imaging continuous monitors, on the other hand, utilize open-path integration, scanning lasers, or short- and mid-wave infrared cameras to monitor for emissions.

Performance was assessed based on the ability to detect (i.e., POD), quantify (i.e., quantification accuracy), and localize emissions (i.e., localization accuracy). The study results indicated that the performance of tested continuous monitors varied widely across assessed metrics. For detection performance, results indicated that only two solutions achieved DL90s within the tested range, failed to detect $\leq 40\%$ of the controlled releases, and had $\leq 40\%$ of their reported detections classified as false alerts. More than half of the continuous monitors failed to detect at least 50% of the controlled releases (i.e., FNF $\geq 50\%$). Similarly, more than half of the continuous monitors had over 25% of their detection reports classified as false alerts (i.e., FPF $\geq 25\%$). For

localization performance, scanning/imaging continuous monitors outperformed other continuous monitors in identifying emission sources. For quantification performance, most solutions estimated the total controlled releases aggregated over the test durations with low relative error and associated uncertainty within a quantification factor of 2. However, estimates of single sources showed high inaccuracies and wide uncertainties, ranging from underestimations by a factor of 50 to overestimations by a factor of 46, for the tested range.

Chapter 4 examined the use of standardized testing protocol designed for survey LDAQ solutions to evaluate 12 survey LDAQ solutions at METEC. These solutions consisted of 4 handheld OGI camera solutions, 4 advanced handheld solutions, and 4 mobile solutions (i.e., drone- and automobile-based). Only detection and localization performances were assessed using the same metrics defined for continuous monitors. An additional metric, survey time, was also investigated. It captures the duration required for a survey solution to complete an emissions inspection survey of a defined area. In contrast to continuous monitors, survey LDAQ solutions exhibited more performance variation based on the solution category (e.g., handheld or mobile) than between the individual survey LDAQ solutions tested. Essentially, Handheld OGI camera solutions exhibited high accuracy and repeatability in detecting and localizing small fugitive emissions, while mobile survey solutions generally surveyed O&G assets quicker than other solution categories.

Chapter 5 focused on the performance assessment of the QOGI system (i.e., by FLIR and Providence Photonics) to recommend a deployment approach to achieve near-optimal performance with the solution at real O&G facilities. Until this study, previous chapters of this work only examined the state of the art of LDAQ solutions based on defined metrics without recommending how these solutions should be deployed in the field for near-optimal performance. The QOGI system was selected for this study because it addressed a major limitation of traditional OGI infrared cameras at the time of testing—the inability to quantify detected emissions. The QOGI system, an add-on tool to the OGI camera, addresses this limitation by enabling quantification of detected emissions. Study results showed that low wind speeds, viewing the emissions plume against a clear sky, and observing the plume from a distance were favorable conditions for obtaining relatively accurate

quantification results. Findings also suggested that the experience level of the OGI camera user could influence quantification performance. However, to increase the reliability of estimates during field deployments, it was recommended to use the mean of estimates obtained from multiple, carefully selected perspectives of the plume.

Chapter 6 examined how the performance of LDAQ solutions evolved over time. Four continuous monitors and three survey LDAQ solutions, previously tested in Chapters 3 and 4, were tested again at METEC between 3 and 13 months after their initial tests using the same protocol. The same performance metrics from the initial tests were applied to assess and compare changes in performance. Study results indicated that LDAQ solutions tend to improve with repeat testing, suggesting the likely importance of regular, third-party testing in their development. Regular testing often provide high-quality ground truth data used by LDAQ solution developers to refine and enhance the performance of their systems.

Chapter 7 discussed the revision of the test protocols to address limitations identified in the current protocol. These include, but are not limited to, discrepancies between controlled test conditions and those at real O&G facilities, as well as misalignments between the tested use cases of LDAQ solutions and their actual field applications. These limitations were underscored by findings from an independent field trial [154] involving some of the solutions tested in Chapter 3. The revised protocol introduced three critical improvements to the current protocol. First, the revised protocol now includes baseline emissions into the testing process to simulate routine operational emissions typically observed at real O&G facilities. Second, the revised protocol accounts for the rapid variability in emission rates, locations, and durations during testing, better reflecting the emissions dynamics at actual O&G facilities. Finally, performance metrics in the revised protocol are based on the operational objectives defined by performers rather than those set by the test center. With these revisions, test results obtained using the revised protocol can be reliably extrapolated to real O&G facility operations if test and field deployment conditions are similar.

8.2 Conclusions

Findings from this work showed that LDAQ solution performance varied substantially between individual solutions (for continuous monitors, except in localization) and across solution categories (for survey solutions), and generally improved with repeat testing. Assessed performance, likely reflect best-case scenario outcome due to the relatively simplistic nature of controlled testing conditions, or may not accurately represent some solutions' capabilities if test conditions did not align with the solutions' intended field application.

Recent NSPS OOOOb regulations, which established performance criteria (e.g., detection thresholds, action levels, operational downtime, etc.) for granting regulatory approval to LDAQ solutions, were not in effect when the controlled tests in this study were conducted. However, they now provide a benchmark for interpreting performance results of LDAQ solutions. If detection thresholds are interpreted as equivalent to the DL90 values evaluated in this study—and assuming solutions were tested in alignment with their intended field deployment—then none of the continuous monitors tested would meet the regulatory detection threshold requirement of 0.4 kg CH_4/h or lower if controlled test results are extrapolated as best-case performance achievable at real O&G facilities. Performance in terms of action level—a time-weighted average that triggers an investigative analysis to identify the cause(s) of an exceedance—could not be assessed, as the release rates tested were below the action level defined for the relevant period (i.e., a rolling 7-day average of 15 kg CH_4/h above the site-specific baseline). More importantly, the *intended* application of continuous monitors has evolved with the new regulations. Whereas the current protocol was designed to assess source-level detection and localization capabilities, the regulations now emphasize facility-level monitoring. As discussed in Chapter 7, this shift underscores the need for a revised testing protocol. Despite these limitations, the controlled test results from this study remain informative. For instance, findings from this study suggest that scanning/imaging continuous monitors are well-suited for facility-level monitoring of large emitters (e.g., super-emitters: $\geq$ 100 kg CH_4/h) and can reliably identify source(s) of identified emissions, enabling more effective and timely field responses. In addition, the majority of the tested continuous monitors can oper-

ate and collect ambient data near continuously for extended periods with less than 10% operation downtime (i.e., 90% operational factor) as specified in the regulation.

Despite the emergence of several next-generation survey LDAQ solutions with varying performance claims, handheld OGI cameras still demonstrate superior performance in detecting and localizing small emission sources compared to next generation handheld and mobile LDAQ solutions with similar value propositions. Also, the integration of quantification capability through the QOGI tool, which demonstrated narrower quantification uncertainty compared to continuous monitors, underscores the potential of handheld OGI cameras as a reliable method for characterizing small fugitive leaks at O&G facilities. Depending on the application, other next-generation handheld LDAQ solutions (i.e., advanced handhelds) that did just as well as handheld OGI cameras in some areas—like localization and survey speed—and have similar deployment demands (e.g., survey period, personnel, etc.), might have a marginal deployment cost advantage over handheld OGI camera solutions. In addition, just like the handheld OGI cameras, the DL90 of most advanced handheld LDAQ solutions met the NSPS OOOOb final rule's minimum detection threshold requirement for quarterly monitoring ($\leq$ 1 kg CH_4/h) at well sites and centralized production facilities using alternative technologies. On-site survey LDAQ solutions (i.e., mobile solutions) with deployment demands—including cost—comparable to handheld OGI cameras may be more advantageous when faster emissions survey is a priority, offering potential cost-saving benefits during field deployment. Despite the promising performance of handheld OGI camera solutions, limitations in terms of scale of application persists, which carries high cost implications. Surveying for emissions using handheld OGI cameras—which are expensive (normally $\geq$ \$100,000 for regulatory-approved models)—typically requires facility access and proximity to potential emitters, often spread across extensive spatial scales. Hence, achieving meaningful emissions mitigation with OGI cameras necessitates deploying thousands of personnel equipped with these cameras at regular intervals which can be labor-intensive and cost-prohibitive.

Since no individual LDAQ solution or category of LDAQ solutions achieved all desired performance levels across the metrics assessed in this study, selecting a LDAQ solution or combination of

LDAQ solutions will depend on those that demonstrates favorable trade-offs between performance and the following factors:

1. **Coverage and Access:** As highlighted earlier, the use of handheld LDAQ solutions—including traditional LDAR techniques—requires surveyors to be near potential leak sources, making these methods effective for source-level emissions detection and localization. In contrast, continuous monitors are not subject to such proximity constraints; however, their deployment over large spatial areas can be limited by cost. At scales beyond a single facility or a few small facilities, deploying continuous monitors may become cost-prohibitive. Automobile-based LDAQ solutions are capable of covering larger areas, but their effectiveness can be limited by the availability and accessibility of roads in and around target sites. Drone-based LDAQ solutions, which are not dependent on access roads, can be used to survey inaccessible areas and assets that may be unsafe for other methods, such as sources on tall platforms. However, the spatial coverage of drone-based LDAQ solutions is constrained by drone specifications—including type (e.g., fixed-wing vs. rotary), weight, and power requirements.

2. **Observation Time:** Although most continuous monitors tested in this study had higher DL90s, FNF, and FPF than handheld survey solutions, continuous monitors can detect large emissions—an important source category—faster and often at a higher success rate than other methods. This helps reduce the risk of prolonged emissions that might otherwise be missed between periodic surveys with mobile solutions and regulatory LDAR inspections. Field studies have shown that large, episodic, and unpredictable emission events often account for a disproportionately large share of total emissions at O&G facilities. While survey solutions and traditional LDAR methods offer short-duration observations, often dependent on timing and prevailing weather conditions, continuous monitors can provide near-continuous emissions observations at all times and across most weather conditions.

3. **Observation or Survey Speed:** Faster survey or observation times can result in significant cost savings during field deployments, particularly when operations are scaled across

multiple facilities. While a trained or certified surveyor using an OGI camera or other handheld LDAQ solution may achieve survey times comparable to some mobile solutions, this performance is highly dependent on the size of the facility and the number of facilities surveyed. As facility size and survey workload increase, detection performance and survey speed for handheld surveyors often decline due to fatigue. In contrast, mobile solutions can cover larger spatial areas faster with more consistent performance, even when operated by a similar number of personnel as handheld solutions.

4. **Personnel Requirement:** When deployed, continuous monitors operate autonomously and generally require human intervention only for maintenance or during periods of system downtime. In contrast, survey solutions—including handheld/traditional LDAR and mobile solutions—typically require a two-person crew, with responsibilities divided between driving or flying and operating the LDAQ solution during surveys. In some cases, additional personnel may be involved to assist with miscellaneous tasks and the documentation of measurement results.

5. **Deployment Cost:** Next-generation LDAQ solutions offer the potential for more cost-effective emissions mitigation compared to traditional LDAR methods, such as handheld OGI camera surveys. These cost benefits are primarily realized through reduced labor costs when deployments are scaled across multiple facilities or over extended durations. For instance, automobile- or drone-based (i.e., mobile) LDAQ solutions may provide a non-trivial cost advantage over handheld OGI camera deployments due to their lower labor intensity and faster survey completion times, even though both approaches typically require comparable initial capital investments. This cost gap could widen when comparing other handheld LDAQ solutions or traditional LDAR methods, as their initial capital investments are often substantially lower than those of OGI cameras. For continuous monitors, the cost benefit—depending on the scale of deployment—is realized through savings on personnel and operational costs that would otherwise be incurred if traditional LDAR methods or handheld/mobile LDAQ solutions were deployed frequently to match the monitoring capabilities of continuous systems.

Overall, this work has demonstrated that next-generation LDAQ solutions are capable of detecting, localizing, and quantifying emissions, though performance varies based on the field readiness level and intended use case of LDAQ solutions. Field readiness level reflects the maturity of an LDAQ solution in terms of sensor performance, effectiveness of data analytics, and deployment approach. Variation in performance based on field readiness was evident across all tested LDAQ solution. To illustrate these variations, results have been categorized into three scenarios:

1. **LDAQ solutions with poor performance across all or the majority of assessed metrics:** Solutions in this category are typically at an early developmental stage at the time of testing. Poor sensor quality, underdeveloped data analytics, and an unoptimized deployment strategy often lead to poor detection performance (i.e., high FNF and FPF values) and significant quantification errors with wide uncertainties. These results indicate that substantial development efforts are required to enhance the readiness of this class of solutions for field deployment.

2. **LDAQ solutions with performance trade-offs across assessed metrics:** Solutions in this category often exhibit performance trade-offs due to ineffective data analytics. For example, the modification of some these LDAQ solutions' algorithm by increasing detection sensitivity (i.e., likely reducing FNF) could increase the difficulty in distinguishing small fugitive emissions from background noise, thereby increasing false detections (i.e., likely increasing FPF). Conversely, reducing detection sensitivity (i.e., likely increasing FNF) could result in fewer false detections (i.e., likely reducing FPF). These trade-offs in performance can have significant cost implications, especially for continuous monitors with high FPF, if follow-up investigations by a ground team are initiated for every detection alert received by the user.

3. **LDAQ solutions with good performance across all or the majority of assessed metrics:** Solutions in this category demonstrate high field readiness for deployment—provided that their test results are replicable in real-world conditions and applicable to facilities similar to METEC.

Likely variation in performance based on the intended use case was evident in the differences in detection, localization, and survey time between handheld and mobile survey solutions. As the results showed, tested handheld solutions generally exhibited higher detection sensitivity and lower FPF values but required more time to complete surveys. In contrast, mobile solutions typically had lower detection sensitivity and higher FPF values, but they completed surveys more quickly. It is important to note, however, that the poor performance of some LDAQ solutions may not necessarily stem from their immaturity, but could instead be influenced by testing conditions that did not align with their typical operational objectives. The inability to make this distinction is a key limitation of the current protocol as highlighted and discussed in Chapter 7.

With numerous LDAQ solutions competing for commercial market share, independent, comprehensive, and transparent validation of vendor performance claims is essential for supporting decision-making regarding field deployment of LDAQ solutions. Given the growing interest in using LDAQ solutions for emissions mitigation, there is an increasing reliance on measurement results from these solutions in various programs. These range from O&G operator-internal emissions mitigation efforts to regulatory programs with financial penalties, underscoring the need for a field-representative, standardized controlled testing protocol. Such a protocol is crucial for accelerating industry-wide adoption and regulatory approval of next-generation LDAQ solutions.

As next-generation LDAQ solutions continue to evolve, testing must also advance in rigor and complexity to accurately characterize their current use and state of the art capabilities. Therefore, if controlled testing effectively mirrors field deployment conditions, test results can be reliably extrapolated to field applications. However, this is only possible if the operational objectives of LDAQ solutions are well-characterized through performance metrics and the intended deployment environment closely aligns with tested conditions.

8.3 Recommendations and Future Work

The first recommendation based on the findings of this work is to evaluate the same LDAQ solutions in controlled testing using the revised standardized protocol, followed by field trials under

similar deployment conditions to compare performance. The reliability of extrapolating test results obtained using the revised protocol to real-world facilities will ultimately depend on the extent to which the gap between these test results is minimized. For a scenario where the gap remains wide, further comprehensive protocol revisions is encouraged to enhance alignment between controlled testing and real-world application of LDAQ solutions.

Other frontiers to be explored include assessing the impact of deployment strategies on solution performance. This involves varying factors such as survey height, survey speed, the number of sensors installed, and the proximity of detection sensors to emission sources. Additionally, comprehensive research on improving the inversion models used for emissions quantification is crucial. These models are integral to any LDAQ solution's analytics. Given the persistent wide uncertainty in single estimates observed throughout this work, enhancing the quantification accuracy of LDAQ solutions remains a key area for further research.

Bibliography

[1] "Methane and climate change." [Online]. Available: https://www.iea.org/reports/global-methane-tracker-2022/methane-and-climate-change. [Accessed January 21, 2025].

[2] "Greenhouse gases continued to increase rapidly in 2022." [Online]. Available: https://www.noaa.gov/news-release/greenhouse-gases-continued-to-increase-rapidly-in-2022. [Accessed January 21, 2025].

[3] "Methane emissions." [Online]. Available: https://www.epa.gov/ghgemissions/methane-emissions. [Accessed January 21, 2025].

[4] "Understanding methane emissions." [Online]. Available: https://www.iea.org/reports/global-methane-tracker-2024/understanding-methane-emissions. [Accessed January 22, 2025].

[5] "Understanding methane emissions." [Online]. Available: https://www.iea.org/reports/global-methane-tracker-2024. [Accessed January 21, 2025].

[6] "Benefits and costs of mitigating methane emissions." [Online]. Available: https://www.ccacoalition.org/content/benefits-and-costs-mitigating-methane-emissions. [Accessed January 22, 2025].

[7] "Natural gas explained." [Online]. Available: https://www.eia.gov/energyexplained/natural-gas/. [Accessed January 22, 2025].

[8] "Summary for policymakers." [Online]. Available: https://www.ipcc.ch/sr15/chapter/spm/. [Accessed January 22, 2025].

[9] D. Johnson and R. Heltzel, "On the long term temporal variations in methane emissions from an unconventional natural gas well site," *ACS Omega*, vol. 6, no. 22, pp. 14200–14207, 2021. DOI: 10.1021/acsomega.1c00874.

[10] M. Omara, D. Zavala-Araiza, D. R. Lyon, B. Hmiel, K. A. Roberts, and S. P. Hamburg, "Methane emissions from US low production oil and natural gas well sites," *Nature Communications*, vol. 13, no. 1, p. 2085, 2022. DOI: 10.1038/s41467-022-29709-3.

[11] D. T. Allen, V. M. Torres, J. Thomas, D. W. Sullivan, M. Harrison, A. Hendler, S. C. Herndon, C. E. Kolb, M. P. Fraser, A. D. Hill, B. K. Lamb, J. Miskimins, R. F. Sawyer, and J. H. Seinfeld, "Measurements of methane emissions at natural gas production sites in the United States," *Proceedings of the National Academy of Sciences*, vol. 110, no. 44, pp. 17768–17773, 2013. DOI: 10.1073/pnas.1304880110.

[12] D. T. Allen, D. W. Sullivan, D. Zavala-Araiza, A. P. Pacsi, M. Harrison, K. Keen, M. P. Fraser, A. Daniel Hill, B. K. Lamb, R. F. Sawyer, and J. H. Seinfeld, "Methane Emissions from Process Equipment at Natural Gas Production Sites in the United States: Liquid Unloadings," *Environmental Science & Technology*, vol. 49, no. 1, pp. 641–648, 2015. DOI: 10.1021/es504016r.

[13] Y. Chen, E. D. Sherwin, E. S. Berman, B. B. Jones, M. P. Gordon, E. B. Wetherley, E. A. Kort, and A. R. Brandt, "Quantifying Regional Methane Emissions in the New Mexico Permian Basin with a Comprehensive Aerial Survey," *Environmental Science & Technology*, vol. 56, no. 7, pp. 4317–4323, 2022. DOI: 10.1021/acs.est.1c06458.

[14] T. N. Lavoie, P. B. Shepson, M. O. L. Cambaliza, B. H. Stirm, A. Karion, C. Sweeney, T. I. Yacovitch, S. C. Herndon, X. Lan, and D. Lyon, "Aircraft-Based Measurements of Point Source Methane Emissions in the Barnett Shale Basin," *Environmental Science & Technology*, vol. 49, no. 13, pp. 7904–7913, 2015. DOI: 10.1021/acs.est.5b00410.

[15] T. N. Lavoie, P. B. Shepson, M. O. L. Cambaliza, B. H. Stirm, S. Conley, S. Mehrotra, I. C. Faloona, and D. Lyon, "Spatiotemporal Variability of Methane Emissions at Oil and Natural Gas Operations in the Eagle Ford Basin," *Environmental Science & Technology*, vol. 51, no. 14, pp. 8001–8009, 2017. DOI: 10.1021/acs.est.7b00814.

[16] 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, no. 17, pp. 10718–10727, 2015. DOI: 10.1021/acs.est.5b02275.

[17] J. Peischl, T. B. Ryerson, K. C. Aikin, J. A. de Gouw, J. B. Gilman, J. S. Holloway, B. M. Lerner, R. Nadkarni, J. A. Neuman, J. B. Nowak, M. Trainer, C. Warneke, and D. D. Parrish, "Quantifying atmospheric methane emissions from the Haynesville, Fayetteville, and northeastern Marcellus shale gas production regions: CH4 emissions from shale gas production," *Journal of Geophysical Research: Atmospheres*, vol. 120, 2015. DOI: 10.1002/2014JD022697.

[18] E. D. Sherwin, J. S. Rutherford, Z. Zhang, Y. Chen, E. B. Wetherley, P. V. Yakovlev, E. S. F. Berman, B. B. Jones, D. H. Cusworth, A. K. Thorpe, A. K. Ayasse, R. M. Duren, and A. R. Brandt, "US oil and gas system emissions from nearly one million aerial site measurements," *Nature*, vol. 627, no. 8003, pp. 328–334, 2024. DOI: 10.1038/s41586-024-07117-5.

[19] T. L. Vaughn, C. S. Bell, T. I. Yacovitch, J. R. Roscioli, S. C. Herndon, S. Conley, S. Schwietzke, G. A. Heath, G. Pétron, and D. Zimmerle, "Comparing facility-level methane emission rate estimates at natural gas gathering and boosting stations," *Elementa: Science of the Anthropocene*, vol. 5, p. 71, 2017. DOI: 10.1525/elementa.257.

[20] 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, no. 12, pp. 7552–7561, 2020. DOI: 10.1021/acs.est.0c00516.

[21] D. Johnson, N. Clark, R. Heltzel, M. Darzi, T. L. Footer, S. Herndon, and E. D. Thoma, "Methane emissions from oil and gas production sites and their storage tanks

in West Virginia," *Atmospheric Environment: X*, vol. 16, p. 100193, 2022. DOI: 10.1016/j.aeaoa.2022.100193.

[22] J. R. Roscioli, T. I. Yacovitch, C. Floerchinger, A. L. Mitchell, D. S. Tkacik, R. Subramanian, D. M. Martinez, T. L. Vaughn, L. Williams, D. Zimmerle, A. L. Robinson, S. C. Herndon, and A. J. Marchese, "Measurements of methane emissions from natural gas gathering facilities and processing plants: Measurement methods," *Atmospheric Measurement Techniques*, vol. 8, no. 5, pp. 2017–2035, 2015. DOI: 10.5194/amt-8-2017-2015.

[23] R. Subramanian, L. L. Williams, T. L. Vaughn, D. Zimmerle, J. R. Roscioli, S. C. Herndon, T. I. Yacovitch, C. Floerchinger, D. S. Tkacik, A. L. Mitchell, M. R. Sullivan, T. R. Dallmann, and A. L. Robinson, "Methane Emissions from Natural Gas Compressor Stations in the Transmission and Storage Sector: Measurements and Comparisons with the EPA Greenhouse Gas Reporting Program Protocol," *Environmental Science & Technology*, vol. 49, no. 5, pp. 3252–3261, 2015. DOI: 10.1021/es5060258.

[24] 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, no. 15, pp. 9374–9383, 2015. DOI: 10.1021/acs.est.5b01669.

[25] B. K. Lamb, S. L. Edburg, T. W. Ferrara, T. Howard, M. R. Harrison, C. E. Kolb, A. Townsend-Small, W. Dyck, A. Possolo, and J. R. Whetstone, "Direct Measurements Show Decreasing Methane Emissions from Natural Gas Local Distribution Systems in the United States," *Environmental Science & Technology*, vol. 49, no. 8, pp. 5161–5169, 2015. DOI: 10.1021/es505116p.

[26] Z. D. Weller, S. P. Hamburg, and J. C. von Fischer, "A national estimate of methane leakage from pipeline mains in natural gas local distribution systems," *Environmental Science & Technology*, vol. 54, no. 14, pp. 8958–8967, 2020. DOI: 10.1021/acs.est.0c00437.

[27] D. Zimmerle, S. Dileep, and C. Quinn, "Unaddressed uncertainties when scaling regional aircraft emission surveys to basin emission estimates," *Environmental Science & Technology*, vol. 58, no. 15, pp. 6575–6585, 2024. DOI: 10.1021/acs.est.3c08972.

[28] "Equipment Leaks." [Online]. Available: https://www.epa.gov/natural-gas-star-program/equipment-leaks. [Accessed January 21, 2025].

[29] "Engine Exhaust." [Online]. Available: https://www.epa.gov/natural-gas-star-program/engine-exhaust. [Accessed January 23, 2025].

[30] International Energy Agency, *Curtailing Methane Emissions from Fossil Fuel Operations: Pathways to a 75% cut by 2030*. OECD, 2021. DOI: 10.1787/1616ff90-en.

[31] "Identifying and Evaluating Opportunities for Greenhouse Gas Mitigation & Operational Efficiency Improvement at Oil and Gas Facilities," [Online]. Available: https://www.ccacoalition.org/resources/identifying-and-evaluating-opportunities-greenhouse-gas-mitigation-operational-efficiency-improvement-oil-and-gas-facilities. [Accessed January 22, 2025].

[32] A. P. Ravikumar, D. Roda-Stuart, R. Liu, A. Bradley, J. Bergerson, Y. Nie, S. Zhang, X. Bi, and A. R. Brandt, "Repeated leak detection and repair surveys reduce methane emissions over scale of years," *Environmental Research Letters*, vol. 15, no. 3, p. 034029, 2020. DOI: 10.1088/1748-9326/ab6ae1.

[33] "Net Zero Emissions by 2050 Scenario (NZE) – Global Energy and Climate Model – Analysis." [Online]. Available: https://www.iea.org/reports/global-energy-and-climate-model/net-zero-emissions-by-2050-scenario-nze. [Accessed January 22, 2025].

[34] "Leak Detection and Repair: A Best Practices Guide." [Online]. Available: https://www.epa.gov/compliance/leak-detection-and-repair-best-practices-guide. [Accessed January 23, 2025].

[35] “Bureau of Air Quality.” [Online]. Available: https://www.dep.pa.gov:443/Business/Air/BAQ/Pages/default.aspx. [Accessed January 22, 2025].

[36] “Oil and Gas Regulation | California Air Resources Board.” [Online]. Available: https://ww2.arb.ca.gov/resources/documents/oil-and-gas-regulation. [Accessed January 22, 2025].

[37] “Code of colorado regulations.” [Online]. Available: https://www.coloradosos.gov/CCR/DisplayRule.do?action=ruleinfo&ruleId=2341&deptID=16&agencyID=7&deptName=Department. [Accessed January 22, 2025].

[38] “Manual 016: How to Develop a Fugitive Emissions Management Program,” [Online]. Available: https://static.aer.ca/prd/documents/manuals/Manual016.pdf. [Accessed January 22, 2025].

[39] “Regulations Respecting Reduction in the Release of Methane and Certain Volatile Organic Compounds (Upstream Oil and Gas Sector).” [Online]. Available: https://laws-lois.justice.gc.ca/eng/regulations/SOR-2018-66/. [Accessed January 24, 2025].

[40] “Proposal for a Regulation of the European Parliament and of the Council on methane emissions reduction in the energy sector and amending Regulation (EU) 2019/942,” 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0805. [Accessed January 22, 2025].

[41] “Method 21 - volatile organic compound leaks.” [Online]. Available: https://www.epa.gov/emc/method-21-volatile-organic-compound-leaks. [Accessed January 29, 2025].

[42] “Federal register: Alternative Work Practice To Detect Leaks From Equipment.” [Online]. Available: https://www.federalregister.gov/documents/2008/12/22/E8-30196/alternative-work-practice-to-detect-leaks-from-equipment. [Accessed January 22, 2025].

[43] D. Zimmerle, T. Vaughn, C. Bell, K. Bennett, P. Deshmukh, and E. Thoma, "Detection Limits of Optical Gas Imaging for Natural Gas Leak Detection in Realistic Controlled Conditions," *Environmental Science & Technology*, vol. 54, no. 18, pp. 11506–11514, 2020. DOI: 10.1021/acs.est.0c01285.

[44] A. P. Ravikumar, J. Wang, and A. R. Brandt, "Are Optical Gas Imaging Technologies Effective For Methane Leak Detection?," *Environmental Science & Technology*, vol. 51, no. 1, pp. 718–724, 2017. DOI: 10.1021/acs.est.6b03906.

[45] T. A. Fox, T. E. Barchyn, D. Risk, A. P. Ravikumar, and C. H. Hugenholtz, "A review of close-range and screening technologies for mitigating fugitive methane emissions in upstream oil and gas," *Environmental Research Letters*, vol. 14, no. 5, p. 053002, 2019. DOI: 10.1088/1748-9326/ab0cc3.

[46] 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, no. 46, pp. 11712–11717, 2018. DOI: 10.1073/pnas.1805687115.

[47] A. M. Robertson, R. Edie, D. Snare, J. Soltis, R. A. Field, M. D. Burkhart, C. S. Bell, D. Zimmerle, and S. M. Murphy, "Variation in Methane Emission Rates from Well Pads in Four Oil and Gas Basins with Contrasting Production Volumes and Compositions," *Environmental Science & Technology*, vol. 51, no. 15, pp. 8832–8840, 2017. DOI: 10.1021/acs.est.7b00571.

[48] A. Heimburger, R. Harvey, P. Shepson, B. Stirm, C. Gore, J. Turnbull, M. Cambaliza, O. Salmon, A.-E. Kerlo, T. Lavoie, K. Davis, T. Lauvaux, A. Karion, C. Sweeney, W. Brewer, R. Hardesty, and K. Gurney, "Assessing the optimized precision of the aircraft mass balance method for measurement of urban greenhouse gas emission rates through averaging," *Elem Sci Anth*, vol. 5, p. 26, 2017. DOI: 10.1525/elementa.134.

[49] D. Zavala-Araiza, D. R. Lyon, R. A. Alvarez, K. J. Davis, R. Harriss, S. C. Herndon, A. Karion, E. A. Kort, B. K. Lamb, X. Lan, A. J. Marchese, S. W. Pacala, A. L. Robinson, P. B. Shepson, C. Sweeney, R. Talbot, A. Townsend-Small, T. I. Yacovitch, D. J. Zimmerle, and S. P. Hamburg, "Reconciling divergent estimates of oil and gas methane emissions," *Proceedings of the National Academy of Sciences*, vol. 112, no. 51, pp. 15597–15602, 2015. DOI: 10.1073/pnas.1522126112.

[50] 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, no. 20, pp. 14743–14752, 2022. DOI: 10.1021/acs.est.2c06211.

[51] D. Zavala-Araiza, R. A. Alvarez, D. R. Lyon, D. T. Allen, A. J. Marchese, D. J. Zimmerle, and S. P. Hamburg, "Super-emitters in natural gas infrastructure are caused by abnormal process conditions," *Nature Communications*, vol. 8, no. 1, p. 14012, 2017. DOI: 10.1038/ncomms14012.

[52] A. Brandt, G. Heath, E. Kort, F. O'Sullivan, G. Petron, S. Jordaan, P. Tans, J. Wilcox, A. Gopstein, D. Arent, S. Wofsy, N. Brown, R. Bradley, G. Stucky, D. Eardley, and R. Harriss, "Methane Leaks from North American Natural Gas Systems," *Science (New York, N.Y.)*, vol. 343, pp. 733–5, 2014. DOI: 10.1126/science.1247045.

[53] D. Zavala-Araiza, D. Lyon, R. A. Alvarez, V. Palacios, R. Harriss, X. Lan, R. Talbot, and S. P. Hamburg, "Toward a Functional Definition of Methane Super-Emitters: Application to Natural Gas Production Sites," *Environmental Science & Technology*, vol. 49, no. 13, pp. 8167–8174, 2015. DOI: 10.1021/acs.est.5b00133.

[54] 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, 2018. DOI: 10.1126/science.aar7204.

[55] S. Schwietzke, G. Pétron, S. Conley, C. Pickering, I. Mielke-Maday, E. J. Dlugokencky, P. P. Tans, T. Vaughn, C. Bell, D. Zimmerle, S. Wolter, C. W. King, A. B. White, T. Coleman, L. Bianco, and R. C. Schnell, "Improved Mechanistic Understanding of Natural Gas Methane Emissions from Spatially Resolved Aircraft Measurements," *Environmental Science & Technology*, vol. 51, no. 12, pp. 7286–7294, 2017. DOI: 10.1021/acs.est.7b01810.

[56] D. H. Cusworth, R. M. Duren, A. K. Thorpe, W. Olson-Duvall, J. Heckler, J. W. Chapman, M. L. Eastwood, M. C. Helmlinger, R. O. Green, G. P. Asner, P. E. Dennison, and C. E. Miller, "Intermittency of Large Methane Emitters in the Permian Basin," *Environmental Science & Technology Letters*, vol. 8, no. 7, pp. 567–573, 2021. DOI: 10.1021/acs.estlett.1c00173.

[57] Z. D. Weller, J. R. Roscioli, W. C. Daube, B. K. Lamb, T. W. Ferrara, P. E. Brewer, and J. C. von Fischer, "Vehicle-based methane surveys for finding natural gas leaks and estimating their size: Validation and uncertainty," *Environmental Science & Technology*, vol. 52, no. 20, pp. 11922–11930, 2018. DOI: 10.1021/acs.est.8b03135.

[58] J. C. von Fischer, D. Cooley, S. Chamberlain, A. Gaylord, C. J. Griebenow, S. P. Hamburg, J. Salo, R. Schumacher, D. Theobald, and J. Ham, "Rapid, vehicle-based identification of location and magnitude of urban natural gas pipeline leaks," *Environmental Science & Technology*, vol. 51, no. 7, pp. 4091–4099, 2017. DOI: 10.1021/acs.est.6b06095.

[59] 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. 57, no. 39, pp. 14539–14547, 2023. DOI: 10.1021/acs.est.3c01321.

[60] L. M. Golston, N. F. Aubut, M. B. Frish, S. Yang, R. W. Talbot, C. Gretencord, J. McSpiritt, and M. A. Zondlo, "Natural Gas Fugitive Leak Detection Using an Unmanned Aerial Vehicle: Localization and Quantification of Emission Rate," *Atmosphere*, vol. 9, no. 9, p. 333, 2018. DOI: 10.3390/atmos9090333.

[61] I. Irakulis-Loitxate, J. Gorroño, D. Zavala-Araiza, and L. Guanter, "Satellites Detect a Methane Ultra-emission Event from an Offshore Platform in the Gulf of Mexico," *Environmental Science & Technology Letters*, vol. 9, no. 6, pp. 520–525, 2022. DOI: 10.1021/acs.estlett.2c00225.

[62] L. Dubey, J. Cooper, I. Staffell, A. Hawkes, and P. Balcombe, "Comparing satellite methane measurements to inventory estimates: A Canadian case study," *Atmospheric Environment: X*, vol. 17, p. 100198, 2023. DOI: 10.1016/j.aeaoa.2022.100198.

[63] R. IJzermans, M. Jones, D. Weidmann, B. Kerkhof, and D. Randell, "Long-term continuous monitoring of methane emissions at an oil and gas facility using a multi-open-path laser dispersion spectrometer," *Scientific Reports*, vol. 14, no. 1, p. 623, 2024. DOI: 10.1038/s41598-023-50081-9.

[64] W. S. Daniels, J. L. Wang, A. P. Ravikumar, M. Harrison, S. A. Roman-White, F. C. George, and D. M. Hammerling, "Toward Multiscale Measurement-Informed Methane Inventories: Reconciling Bottom-Up Site-Level Inventories with Top-Down Measurements Using Continuous Monitoring Systems," *Environmental Science & Technology*, vol. 57, no. 32, pp. 11823–11833, 2023. DOI: 10.1021/acs.est.3c01121.

[65] C. S. Bell, T. L. Vaughn, D. Zimmerle, S. C. Herndon, T. I. Yacovitch, G. A. Heath, G. Pétron, R. Edie, R. A. Field, S. M. Murphy, A. M. Robertson, and J. Soltis, "Comparison of methane emission estimates from multiple measurement techniques at natural gas production pads," *Elementa: Science of the Anthropocene*, vol. 5, p. 79, 2017. DOI: 10.1525/elementa.266.

[66] S. N. Riddick, M. Mbua, A. Santos, E. W. Emerson, F. Cheptonui, C. Houlihan, A. L. Hodshire, A. Anand, W. Hartzell, and D. J. Zimmerle, "Methane emissions from abandoned oil and gas wells in Colorado," *Science of The Total Environment*, vol. 922, p. 170990, 2024. DOI: 10.1016/j.scitotenv.2024.170990.

[67] T. Aldhafeeri, M.-K. Tran, R. Vrolyk, M. Pope, and M. Fowler, "A Review of Methane Gas Detection Sensors: Recent Developments and Future Perspectives," *Inventions*, vol. 5, no. 3, p. 28, 2020. DOI: 10.3390/inventions5030028.

[68] Y. Wu, B. Yao, C. Yu, and Y. Rao, "Optical Graphene Gas Sensors Based on Microfibers: A Review," *Sensors*, vol. 18, no. 4, p. 941, 2018. DOI: 10.3390/s18040941.

[69] J. Yang, X. Che, R. Shen, C. Wang, X. Li, and W. Chen, "High-sensitivity photonic crystal fiber long-period grating methane sensor with cryptophane-A-6Me absorbed on a PAA-CNTs/PAH nanofilm," *Optics Express*, vol. 25, no. 17, pp. 20258–20267, 2017. DOI: 10.1364/OE.25.020258.

[70] C. Zheng, W. Ye, N. P. Sanchez, C. Li, L. Dong, Y. Wang, R. J. Griffin, and F. K. Tittel, "Development and field deployment of a mid-infrared methane sensor without pressure control using interband cascade laser absorption spectroscopy," *Sensors and Actuators B: Chemical*, vol. 244, pp. 365–372, 2017. DOI: 10.1016/j.snb.2016.12.146.

[71] H. Zhang, B. Shen, W. Hu, and X. Liu, "Research on a Fast-Response Thermal Conductivity Sensor Based on Carbon Nanotube Modification," *Sensors*, vol. 18, no. 7, p. 2191, 2018. DOI: 10.3390/s18072191.

[72] A. Mahdavifar, R. Aguilar, Z. Peng, P. Hesketh, M. Findlay, J. Stetter, and G. Hunter, "Simulation and Fabrication of an Ultra-Low Power Miniature Microbridge Thermal Conductivity Gas Sensor," *Journal of the Electrochemical Society*, vol. 161, pp. B55–B61, 2014. DOI: 10.1149/2.032404jes.

[73] P. N. Bartlett and S. Guerin, "A micromachined calorimetric gas sensor: An application of electrodeposited nanostructured palladium for the detection of combustible gases," *Analytical Chemistry*, vol. 75, no. 1, pp. 126–132, 2003. DOI: 10.1021/ac026141w.

[74] L. Jun, T. Qiulin, Z. Wendong, X. Chenyang, G. Tao, and X. Jijun, "Miniature low-power IR monitor for methane detection," *Measurement*, vol. 44, no. 5, pp. 823–831, 2011. DOI: 10.1016/j.measurement.2011.01.021.

[75] T. Qiu-lin, Z. Wen-dong, X. Chen-yang, X. Ji-jun, L. Jun, L. Jun-hong, and L. Ting, "Design, fabrication and characterization of pyroelectric thin film and its application for infrared gas sensors," *Microelectronics Journal*, vol. 40, no. 1, pp. 58–62, 2009. DOI: 10.1016/j.mejo.2008.07.011.

[76] M. E. Edwards, A. K. Batra, A. K. Chilvery, P. Guggilla, M. Curley, and M. D. Aggarwal, "Pyroelectric Properties of PVDF:MWCNT Nanocomposite Film for Uncooled Infrared Detectors," *Materials Sciences and Applications*, vol. 3, no. 12, pp. 851–855, 2012. DOI: 10.4236/msa.2012.312124.

[77] N. Lu, S. Fan, Y. Zhao, B. Yang, Z. Hua, and Y. Wu, "A selective methane gas sensor with printed catalytic films as active filters," *Sensors and Actuators B: Chemical*, vol. 347, p. 130603, 2021. DOI: 10.1016/j.snb.2021.130603.

[78] Y. Zhao, S. Wang, W. Yuan, S. Fan, Z. Hua, Y. Wu, and X. Tian, "Selective detection of methane by Pd-In2O3 sensors with a catalyst filter film," *Sensors and Actuators B: Chemical*, vol. 328, p. 129030, 2021. DOI: 10.1016/j.snb.2020.129030.

[79] D. Xue, P. Wang, Z. Zhang, and W. Yan, "Enhanced methane sensing property of flower-like SnO2 doped by Pt nanoparticles: A combined experimental and first-principle study," *Sensors and Actuators B: Chemical*, vol. 296, p. 126710, 2019. DOI: 10.1016/j.snb.2019.126710.

[80] Z. Wang and X. Zeng, "Bis(trifluoromethylsulfonyl)imide (NTf2)-Based Ionic Liquids for Facile Methane Electro-Oxidation on Pt," *Journal of The Electrochemical Society*, vol. 160, no. 9, p. H604, 2013. DOI: 10.1149/2.039309jes.

[81] Z. Wang, M. Guo, G. Baker, J. Stetter, L. Lin, A. Mason, and X. Zeng, "Methane-oxygen electrochemical coupling in an ionic liquid: A robust sensor for simultaneous quantification," *The Analyst*, vol. 139, 2014. DOI: 10.1039/c4an00839a.

[82] H. Yin, H. Wan, L. Lin, X. Zeng, and A. J. Mason, "Miniaturized planar RTIL-based eletrochemical gas sensor for real-time point-of-exposure monitoring," in *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*, pp. 85–88, 2016. DOI: 10.1109/HIC.2016.7797703.

[83] X. Liu, S. Cheng, H. Liu, S. Hu, D. Zhang, and H. Ning, "A Survey on Gas Sensing Technology," *Sensors (Basel, Switzerland)*, vol. 12, no. 7, pp. 9635–9665, 2012. DOI: 10.3390/s120709635.

[84] Y. Diao and H. Yang, "Chapter 5 - Gas-cleaning technology," in *Industrial Ventilation Design Guidebook (Second Edition)* (H. D. Goodfellow and Y. Wang, eds.), pp. 279–371, Academic Press. DOI: 10.1016/B978-0-12-816673-4.00007-9.

[85] 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, 2021. DOI: 10.1016/j.rse.2021.112418.

[86] F. Innocenti, R. Robinson, T. Gardiner, A. Finlayson, and A. Connor, "Differential Absorption Lidar (DIAL) Measurements of Landfill Methane Emissions," *Remote Sensing*, vol. 9, no. 9, p. 953, 2017. DOI: 10.3390/rs9090953.

[87] A. Maity, S. Maithani, and M. Pradhan, "Cavity Ring-Down Spectroscopy: Recent Technological Advancements, Techniques, and Applications," *Analytical Chemistry*, vol. 93, no. 1, pp. 388–416, 2021. DOI: 10.1021/acs.analchem.0c04329.

[88] A. Forstmaier, J. Chen, F. Dietrich, J. Bettinelli, H. Maazallahi, C. Schneider, D. Winkler, X. Zhao, T. Jones, C. van der Veen, N. Wildmann, M. Makowski, A. Uzun, F. Klappenbach, H. Denier van der Gon, S. Schwietzke, and T. Röckmann, "Quantification of methane emissions in Hamburg using a network of FTIR spectrometers and an inverse modeling approach," *Atmospheric Chemistry and Physics*, vol. 23, no. 12, pp. 6897–6922, 2023. DOI: 10.5194/acp-23-6897-2023.

[89] S. Lambert-Girard, M. Allard, M. Piché, and F. Babin, "Differential optical absorption spectroscopy lidar for mid-infrared gaseous measurements," *Applied Optics*, vol. 54, no. 7, pp. 1647–1656, 2015. DOI: 10.1364/AO.54.001647.

[90] J. Veefkind, I. Aben, K. McMullan, H. Förster, J. de Vries, G. Otter, J. Claas, H. Eskes, J. de Haan, Q. Kleipool, M. van Weele, O. Hasekamp, R. Hoogeveen, J. Landgraf, R. Snel, P. Tol, P. Ingmann, R. Voors, B. Kruizinga, R. Vink, H. Visser, and P. Levelt, "TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications," *Remote Sensing of Environment*, vol. 120, pp. 70–83, 2012. DOI: 10.1016/j.rse.2011.09.027.

[91] D. Tratt, K. Buckland, J. Hall, P. Johnson, E. Keim, I. Leifer, K. Westberg, and S. Young, "Airborne visualization and quantification of discrete methane sources in the environment," *Remote Sensing of Environment*, vol. 154, pp. 74–88, 2014. DOI: 10.1016/j.rse.2014.08.011.

[92] R. D. M. Scafutto, C. R. de Souza Filho, D. N. Riley, and W. J. de Oliveira, "Evaluation of thermal infrared hyperspectral imagery for the detection of onshore methane plumes: Significance for hydrocarbon exploration and monitoring," *International Journal of Applied Earth Observation and Geoinformation*, vol. 64, pp. 311–325, 2018. DOI: 10.1016/j.jag.2017.07.002.

[93] M. Gålfalk, G. Olofsson, P. Crill, and D. Bastviken, "Making methane visible," *Nature Climate Change*, vol. 6, no. 4, pp. 426–430, 2016. DOI: 10.1038/nclimate2877.

[94] T. Ehret, A. De Truchis, M. Mazzolini, J.-M. Morel, p. u. family=Aspremont, given=Alexandre, T. Lauvaux, R. Duren, D. Cusworth, and G. Facciolo, "Global Tracking and Quantification of Oil and Gas Methane Emissions from Recurrent Sentinel-2 Imagery," *Environmental Science & Technology*, vol. 56, no. 14, pp. 10517–10529, 2022. DOI: 10.1021/acs.est.1c08575.

[95] M. Kwaśny and A. Bombalska, "Optical Methods of Methane Detection," *Sensors*, vol. 23, no. 5, p. 2834, 2023. DOI: 10.3390/s23052834.

[96] A. Bondarenko, *Mathematical Processing of a Pyroelectric Detector: Time Domain*. Ivanchenko I.S., 2019.

[97] L. M. Dorojkine, "The non-catalytic thermal wave-based chemical gas sensor for methane and natural gas," *Sensors and Actuators B: Chemical*, vol. 89, no. 1, pp. 76–85, 2003. DOI: 10.1016/S0925-4005(02)00431-8.

[98] R. Wu, L. Tian, H. Li, H. Liu, J. Luo, X. Tian, Z. Hua, Y. Wu, and S. Fan, "A selective methane gas sensor based on metal oxide semiconductor equipped with an on-chip microfilter," *Sensors and Actuators B: Chemical*, vol. 359, p. 131557, 2022. DOI: 10.1016/j.snb.2022.131557.

[99] D. Furuta, T. Sayahi, J. Li, B. Wilson, A. A. Presto, and J. Li, "Characterization of inexpensive metal oxide sensor performance for trace methane detection," *Atmospheric Measurement Techniques*, vol. 15, no. 17, pp. 5117–5128, 2022. DOI: 10.5194/amt-15-5117-2022.

[100] J. Baranwal, B. Barse, G. Gatto, G. Broncova, and A. Kumar, "Electrochemical Sensors and Their Applications: A Review," *Chemosensors*, vol. 10, no. 9, p. 363, 2022. DOI: 10.3390/chemosensors10090363.

[101] "Sentinel-5p." [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/copernicus. [Accessed January 24, 2025].

[102] Y. Zhang, R. Gautam, S. Pandey, M. Omara, J. D. Maasakkers, P. Sadavarte, D. Lyon, H. Nesser, M. P. Sulprizio, D. J. Varon, R. Zhang, S. Houweling, D. Zavala-Araiza, R. A. Alvarez, A. Lorente, S. P. Hamburg, I. Aben, and D. J. Jacob, "Quantifying methane emissions from the largest oil-producing basin in the United States from space," *Science Advances*, vol. 6, no. 17, p. eaaz5120, 2020. DOI: 10.1126/sciadv.aaz5120.

[103] O. Schneising, M. Buchwitz, M. Reuter, S. Vanselow, H. Bovensmann, and J. P. Burrows, "Remote sensing of methane leakage from natural gas and petroleum systems revisited," *Atmospheric Chemistry and Physics*, vol. 20, no. 15, pp. 9169–9182, 2020. DOI: 10.5194/acp-20-9169-2020.

[104] "Gas Mapping LiDAR™ Explained | Bridger Photonics." [Online]. Available: https://www.bridgerphotonics.com/blog/gas-mapping-lidartm-explainedf. [Accessed January 24, 2025].

[105] D. R. Tyner and M. R. Johnson, "Where the Methane Is—Insights from Novel Airborne LiDAR Measurements Combined with Ground Survey Data," *Environmental Science & Technology*, vol. 55, no. 14, pp. 9773–9783, 2021. DOI: 10.1021/acs.est.1c01572.

[106] W. M. Kunkel, A. E. Carre-Burritt, G. S. Aivazian, N. C. Snow, J. T. Harris, T. S. Mueller, P. A. Roos, and M. J. Thorpe, "Extension of Methane Emission Rate Distribution for Permian Basin Oil and Gas Production Infrastructure by Aerial LiDAR," *Environmental Science & Technology*, vol. 57, no. 33, pp. 12234–12241, 2023. DOI: 10.1021/acs.est.3c00229.

[107] W. Gong, Q. Zhang, T. Zhang, T. Liu, Z. Wang, and Y. Wei, "Study on laser methane remote sensor based on TDLAS," in *AOPC 2022: Atmospheric and Environmental Optics*, vol. 12561, pp. 67–71, SPIE, 2023. DOI: 10.1117/12.2651953.

[108] "Methane Leak Detection With Drone-Mounted Sensor." [Online]. Available: https://seekops.com/technology/. [Accessed January 24, 2025].

[109] D. Singh, B. Barlow, C. Hugenholtz, W. Funk, C. Robinson, and A. P. Ravikumar, "Field Performance of New Methane Detection Technologies: Results from the Alberta Methane Field Challenge," 2021. Preprint URL: https://eartharxiv.org/repository/view/1860/.

[110] "ABB Ability™ Mobile Gas Leak Detection Systems." [Online]. Available: https://new.abb.com/products/measurement-products/analytical/laser-gas-analyzers/advanced-leak-detection/abb-ability-mobile-gas-leak-detection-system. [Accessed January 24, 2025].

[111] "GERG 'Technology Benchmark for site level methane emissions quantification' – Phase II.B." [Online]. Available: https://www.gerg.eu/projects/methane-emissions/gerg-technology-benchmark-for-site-level-methane-emissions-quantification-phase-ii-b/. [Accessed April 16, 2024].

[112] P. Canary, "Differentiated Emissions Quantification for Continuous Monitoring Systems." [Online]. Available: https://www.projectcanary.com/blog/differentiated-emissions-quantification-for-continuous-monitoring-systems/. [Accessed April 16, 2024].

[113] "Quantitative Optical Gas Imaging System | Teledyne FLIR." [Online]. Available: https://www.flir.com/products/flir-ql320?vertical=optical+gas&segment=solutions. [Accessed January 24, 2025].

[114] H. Abdel-Moati, J. Morris, Y. Zeng, P. Kangas, and D. McGregor, "New Optical Gas Imaging Technology for Quantifying Fugitive Emission Rates," OnePetro, 2015. DOI: 10.2523/IPTC-18471-MS.

[115] J. L. Wang, B. Barlow, W. Funk, C. Robinson, A. Brandt, and A. P. Ravikumar, "Large-Scale Controlled Experiment Demonstrates Effectiveness of Methane Leak Detection and Repair Programs at Oil and Gas Facilities," *Environmental Science & Technology*, vol. 58, no. 7, pp. 3194–3204, 2024. DOI: 10.1021/acs.est.3c09147.

[116] "OGMP 2.0 – The Oil & Gas Methane Partnership 2.0." [Online]. Available: https://ogmpartnership.com/. [Accessed January 24, 2025].

[117] "The MiQ Standard." [Online]. Available: https://miq.org/the-technical-standard/. [Accessed January 24, 2025].

[118] S. Brennan, "Veritas: GTI Energy's Methane Emissions Measurement and Verification Initiative," 2021. [Online]. Available: https://www.gti.energy/veritas-a-gti-methane-emissions-measurement-and-verification-initiative/. [Accessed January 24, 2025].

[119] T. A. Fox, A. P. Ravikumar, C. H. Hugenholtz, D. Zimmerle, T. E. Barchyn, M. R. Johnson, D. Lyon, and T. Taylor, "A methane emissions reduction equivalence framework for alternative leak detection and repair programs," *Elementa: Science of the Anthropocene*, vol. 7, 2019. DOI: 10.1525/elementa.369.

[120] A. P. Ravikumar, S. Sreedhara, J. Wang, J. Englander, D. Roda-Stuart, C. Bell, D. Zimmerle, D. Lyon, I. Mogstad, B. Ratner, and A. R. Brandt, "Single-blind inter-comparison of methane detection technologies – results from the Stanford/EDF Mobile Monitoring Challenge," *Elementa: Science of the Anthropocene*, vol. 7, p. 37, 2019. DOI: 10.1525/elementa.373.

[121] E. D. Sherwin, Y. Chen, A. P. Ravikumar, and A. R. Brandt, "Single-blind test of airplane-based hyperspectral methane detection via controlled releases," *Elementa: Science of the Anthropocene*, vol. 9, no. 1, p. 00063, 2021. DOI: 10.1525/elementa.2021.00063.

[122] C. S. Bell, T. Vaughn, and D. Zimmerle, "Evaluation of next generation emission measurement technologies under repeatable test protocols," *Elementa: Science of the Anthropocene*, vol. 8, p. 32, 2020. DOI: 10.1525/elementa.426.

[123] S. P. Siebenaler, A. M. Janka, D. Lyon, J. P. Edlebeck, and A. E. Nowlan, "Methane Detectors Challenge: Low-Cost Continuous Emissions Monitoring," in *Volume 3: Operations,*

Monitoring and Maintenance; Materials and Joining, p. V003T04A013, American Society of Mechanical Engineers, 2016. DOI: 10.1115/IPC2016-64670.

[124] T. Gardiner, J. Helmore, F. Innocenti, and R. Robinson, "Field Validation of Remote Sensing Methane Emission Measurements," *Remote Sensing*, vol. 9, no. 9, p. 956, 2017. DOI: 10.3390/rs9090956.

[125] J. Titchener, D. Millington-Smith, C. Goldsack, G. Harrison, A. Dunning, X. Ai, and M. Reed, "Single photon Lidar gas imagers for practical and widespread continuous methane monitoring," *Applied Energy*, vol. 306, p. 118086, 2022. DOI: 10.1016/j.apenergy.2021.118086.

[126] Z. Chen, S. H. El Abbadi, E. D. Sherwin, P. M. Burdeau, J. S. Rutherford, Y. Chen, Z. Zhang, and A. R. Brandt, "Comparing continuous methane monitoring technologies for high-volume emissions: A single-blind controlled release study," *ACS ES&T Air*, vol. 1, no. 8, pp. 871–884, 2024. DOI: 10.1021/acsestair.4c00015.

[127] Y. Liu, J.-D. Paris, G. Broquet, V. Bescós Roy, T. Meixus Fernandez, R. Andersen, A. Russu Berlanga, E. Christensen, Y. Courtois, S. Dominok, C. Dussenne, T. Eckert, A. Finlayson, A. Fernández de la Fuente, C. Gunn, R. Hashmonay, J. Grigoleto Hayashi, J. Helmore, S. Honsel, F. Innocenti, M. Irjala, T. Log, C. Lopez, F. Cortés Martínez, J. Martinez, A. Massardier, H. G. Nygaard, P. Agregan Reboredo, E. Rousset, A. Scherello, M. Ulbricht, D. Weidmann, O. Williams, N. Yarrow, M. Zarea, R. Ziegler, J. Sciare, M. Vrekoussis, and P. Bousquet, "Assessment of current methane emission quantification techniques for natural gas midstream applications," *Atmospheric Measurement Techniques*, vol. 17, no. 6, pp. 1633–1649, 2024. DOI: 10.5194/amt-17-1633-2024.

[128] "Advanced Research Projects Agency – Energy (ARPA-E): Methane Observation Networks with Innovative Technology to Obtain Reduction (MONITOR) – Policies." DOI: https://www.iea.org/policies/13480-advanced-research-projects-agency-energy-arpa-e-methane-observation-networks-with-innovative-technology-to-obtain-reduction-monitor.

[129] "Advancing development of emissions detection (aded)." [Online]. Available: https://netl.doe.gov/node/10981. [Accessed January 24, 2025].

[130] M. Olczak, A. Piebalgs, and P. Balcombe, "A global review of methane policies reveals that only 13*One Earth*, vol. 6, no. 5, pp. 519–535, 2023. DOI: 10.1016/j.oneear.2023.04.009.

[131] C. Bell and D. Zimmerle, "METEC controlled test protocol: Continuous monitoring emission detection and quantification," 2020. DOI: 10.25675/10217/235364.

[132] C. Bell and D. Zimmerle, "METEC controlled test protocol: Survey emission detection and quantification," 2022. DOI: 10.25675/10217/235363.

[133] S. H. El Abbadi, Z. Chen, P. M. Burdeau, J. S. Rutherford, Y. Chen, Z. Zhang, E. D. Sherwin, and A. R. Brandt, "Technological maturity of aircraft-based methane sensing for greenhouse gas mitigation," *Environmental Science & Technology*, vol. 58, no. 22, pp. 9591–9600, 2024. DOI: 10.1021/acs.est.4c02439.

[134] J. S. Rutherford, E. D. Sherwin, Y. Chen, S. Aminfard, and A. R. Brandt, "Evaluating methane emission quantification performance and uncertainty of aerial technologies via high-volume single-blind controlled releases," 2023. Preprint URL: https://eartharxiv.org/repository/view/5113/.

[135] 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, no. 1, 2022. DOI: 10.1525/elementa.2022.00080.

[136] "Standards of Performance for New, Reconstructed, and Modified Sources and Emissions Guidelines for Existing Sources: Oil and Natural Gas Sector Climate Review." [Online]. Available: https://www.federalregister.gov/documents/2024/03/08/2024-00366/standards-of-performance-for-new-reconstructed-and-modified-sources-and-emissions-guidelines-for. [Accessed January 26, 2025].

[137] C. Bell, C. Ilonze, A. Duggan, and D. Zimmerle, "Performance of Continuous Emission Monitoring Solutions under a Single-Blind Controlled Testing Protocol," *Environmental Science & Technology*, vol. 57, no. 14, pp. 5794–5805, 2023. DOI: 10.1021/acs.est.2c09235.

[138] C. Ilonze, E. Emerson, A. Duggan, and D. Zimmerle, "Assessing the progress of the performance of continuous monitoring solutions under a single-blind controlled testing protocol," *Environmental Science & Technology*, vol. 58, no. 25, pp. 10941–10955, 2024. DOI: 10.1021/acs.est.3c08511.

[139] C. Ilonze, J. L. Wang, A. P. Ravikumar, and D. Zimmerle, "Methane quantification performance of the quantitative optical gas imaging (qogi) system using single-blind controlled release assessment," *Sensors*, vol. 24, no. 13, 2024. DOI: 10.3390/s24134044.

[140] C. Ilonze, R. Day, E. Emerson, A. Duggan, R. Brouwer, and D. Zimmerle, "Performance of survey solutions under single-blind controlled testing protocol," *ChemRxiv*, 2025. Preprint URL: 10.26434/chemrxiv-2025-5ff89.

[141] A. Pacsi, T. Ferrara, K. Schwan, P. Tupper, M. Lev-On, R. Smith, and K. Ritter, "Equipment leak detection and quantification at 67 oil and gas sites in the Western United States," *Elem Sci Anth*, vol. 7, p. 29, 2019. DOI: 10.1525/elementa.368.

[142] J. Kuo, T. C. Hicks, B. Drake, and T. F. Chan, "Estimation of methane emission from California natural gas industry," *Journal of the Air & Waste Management Association (1995)*, vol. 65, no. 7, pp. 844–855, 2015. DOI: 10.1080/10962247.2015.1025924.

[143] M. Omara, M. R. Sullivan, X. Li, R. Subramanian, A. L. Robinson, and A. A. Presto, "Methane emissions from conventional and unconventional natural gas production sites in the marcellus shale basin," *Environmental Science & Technology*, vol. 50, no. 4, pp. 2099–2107, 2016. DOI: 10.1021/acs.est.5b05503.

[144] R. S. Heltzel, M. T. Zaki, A. K. Gebreslase, O. I. Abdul-Aziz, and D. R. Johnson, "Continuous otm 33a analysis of controlled releases of methane with various time periods, data rates and wind filters," *Environments*, vol. 7, no. 9, 2020. DOI: 10.3390/environments7090065.

[145] R. Heltzel, D. Johnson, M. Zaki, A. Gebreslase, and O. I. Abdul-Aziz, "Understanding the accuracy limitations of quantifying methane emissions using other test method 33a," *Environments*, vol. 9, no. 4, 2022. DOI: 10.3390/environments9040047.

[146] H. L. Brantley, E. D. Thoma, W. C. Squier, B. B. Guven, and D. Lyon, "Assessment of methane emissions from oil and gas production pads using mobile measurements," *Environmental Science & Technology*, vol. 48, no. 24, pp. 14508–14515, 2014. DOI: 10.1021/es503070q.

[147] Y. Zeng and J. Morris, "Detection limits of optical gas imagers as a function of temperature differential and distance," *Journal of the Air & Waste Management Association (1995)*, vol. 69, no. 3, pp. 351–361, 2019. DOI: 10.1080/10962247.2018.1540366.

[148] C. E. Kemp and A. P. Ravikumar, "New Technologies Can Cost Effectively Reduce Oil and Gas Methane Emissions, but Policies Will Require Careful Design to Establish Mitigation Equivalence," *Environmental Science & Technology*, vol. 55, no. 13, pp. 9140–9149, 2021. DOI: 10.1021/acs.est.1c03071.

[149] T. A. Fox, M. Gao, T. E. Barchyn, Y. L. Jamin, and C. H. Hugenholtz, "An agent-based model for estimating emissions reduction equivalence among leak detection and repair programs," *Journal of Cleaner Production*, vol. 282, p. 125237, 2021. DOI: 10.1016/j.jclepro.2020.125237.

[150] "Method detection limit - frequent questions." [Online]. Available: https://www.epa.gov/cwa-methods/method-detection-limit-frequent-questions. [Accessed February 09, 2025].

[151] "Alternative work practice to detect leaks from equipment." [Online]. Available: https://www.federalregister.gov/documents/2008/12/22/E8-30196/alternative-work-practice-to-detect-leaks-from-equipment. [Accessed February 2, 2025].

[152] A. P. Ravikumar, J. Wang, M. McGuire, C. S. Bell, D. Zimmerle, and A. R. Brandt, ""Good versus Good Enough?" Empirical Tests of Methane Leak Detection Sensitivity of a Commercial Infrared Camera," *Environmental Science & Technology*, vol. 52, no. 4, pp. 2368–2374, 2018. DOI: 10.1021/acs.est.7b04945.

[153] "Bringing lab performance to the field: purpose-built sensor technology, proven in the field." [Online]. Available: https://sensirion-connected.com/emissions-monitoring/sensor-technology. [Accessed February 25, 2025].

[154] R. Day, E. Emerson, C. Bell, and D. Zimmerle, "Point Sensor Networks Struggle to Detect and Quantify Short Controlled Releases at Oil and Gas Sites," *Sensors*, vol. 24, p. 2419, 2024. DOI: 10.3390/s24082419.

[155] P. Canary, "Single-blind testing of continuous monitoring system with realistic emission timeseries." [Online]. Available: https://www.projectcanary.com/learn/white-papers-guides/. [Accessed February 26, 2025].

[156] "Subpart w—petroleum and natural gas systems." [Online]. Available: https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-98/subpart-W. [Accessed February 27, 2025].

Appendix A

Equipment and Setup at METEC

Figure A.1: Shows the gas chromatography device used to identify and analyze the species and chemicals that make up the gas supply at METEC. The figure also shows the canisters used to collect gas samples from the onsite gas storage cylinders for species composition analysis.

Appendix B

Variations in Performance of Continuous Monitors With Testing Complexity.

Table B.1: The table summarizes the rank-biseral correlation between FNFs evaluated per experiment and if an experiment had single or multiple controlled releases.

	Continuous Monitors			Survey Solutions		
ID	**Category**	**Coefficient**	**pValue**	**Category**	**Coefficient**	**pValue**
A	Point sensor network	0.26605	0.00000	Mobile	0.46939	0.08903
B	Scanning/imaging	0.04352	0.50259	Handheld OGI	0.64706	0.27958
C	Point sensor network	0.31990	0.00000	Advanced Handheld	0.23239	0.30607
D	Point sensor network	0.50031	0.00000	Mobile	0.67442	0.25429
E	Point sensor network	-0.05325	0.15594	Handheld OGI	0.29861	0.20175
F	Point sensor network	0.39845	0.00000	Mobile	0.90625	0.00296
G	Scanning/imaging	0.12476	0.04145	Handheld OGI	0.18182	0.39591
H	Scanning/imaging	-0.15408	0.00042	Advanced Handheld	0.37037	0.12110
I	Scanning/imaging	-0.06437	0.29858	Mobile	0.50000	0.08427
J	Scanning/imaging	-0.36549	0.00000	Advanced Handheld	0.03326	0.85973
K	Point sensor network	-0.00990	0.13423	Advanced Handheld	0.21429	0.40982
L	Scanning/imaging	0.12915	0.13280	Handheld OGI	0.18182	0.27790
N	Point sensor network	0.35290	0.00000			
O	Scanning/imaging	0.09886	0.20391			
P	Point sensor network	0.39563	0.00000			
Q	Point sensor network	0.11035	0.10273			

Table B.2: Summary of the FNF evaluated for all continuous monitors under different testing complexity.

Solution		Time		C.R. Count		Height	
ID	**Category**	**Night**	**Day**	**Single**	**Multiple**	**Tanks**	**Non tanks**
A	Point sensor network	20.69	32.17	19.3	35.98	26.13	32.18
B	Scanning/imaging	47.27	63.33	55.47	63.96	54.88	62.81
C	Point sensor network	53.45	66.53	42.04	74.19	46.73	69.54
D	Point sensor network	29.31	42.64	17.54	51.36	22.52	45.79
E	Point sensor network	0.00	13.67	16.37	10.53	12.61	12.20
F	Point sensor network	20.69	31.40	12.87	37.72	19.82	32.83
G	Scanning/imaging	93.10	96.67	84.21	97.37	97.14	95.98
H	Scanning/imaging	91.30	90.77	93.92	88.79	96.15	89.42
I	Scanning/imaging	0.00	92.05	91.11	92.18	90.91	92.36
J	Scanning/imaging	0.00	75.96	82.35	74.58	58.50	94.29
K	Point sensor network	100.00	99.73	100.00	99.62	99.46	99.82
L	Scanning/imaging	18.18	37.18	31.88	36.90	33.33	36.27
N	Point sensor network	57.14	56.30	21.21	62.96	42.06	61.29
O	Scanning/imaging	62.86	38.20	35.53	41.99	29.11	43.88
P	Point sensor network	32.61	32.93	11.69	36.38	30.97	33.41
Q	Point sensor network	76.09	70.06	59.74	72.34	73.45	69.82

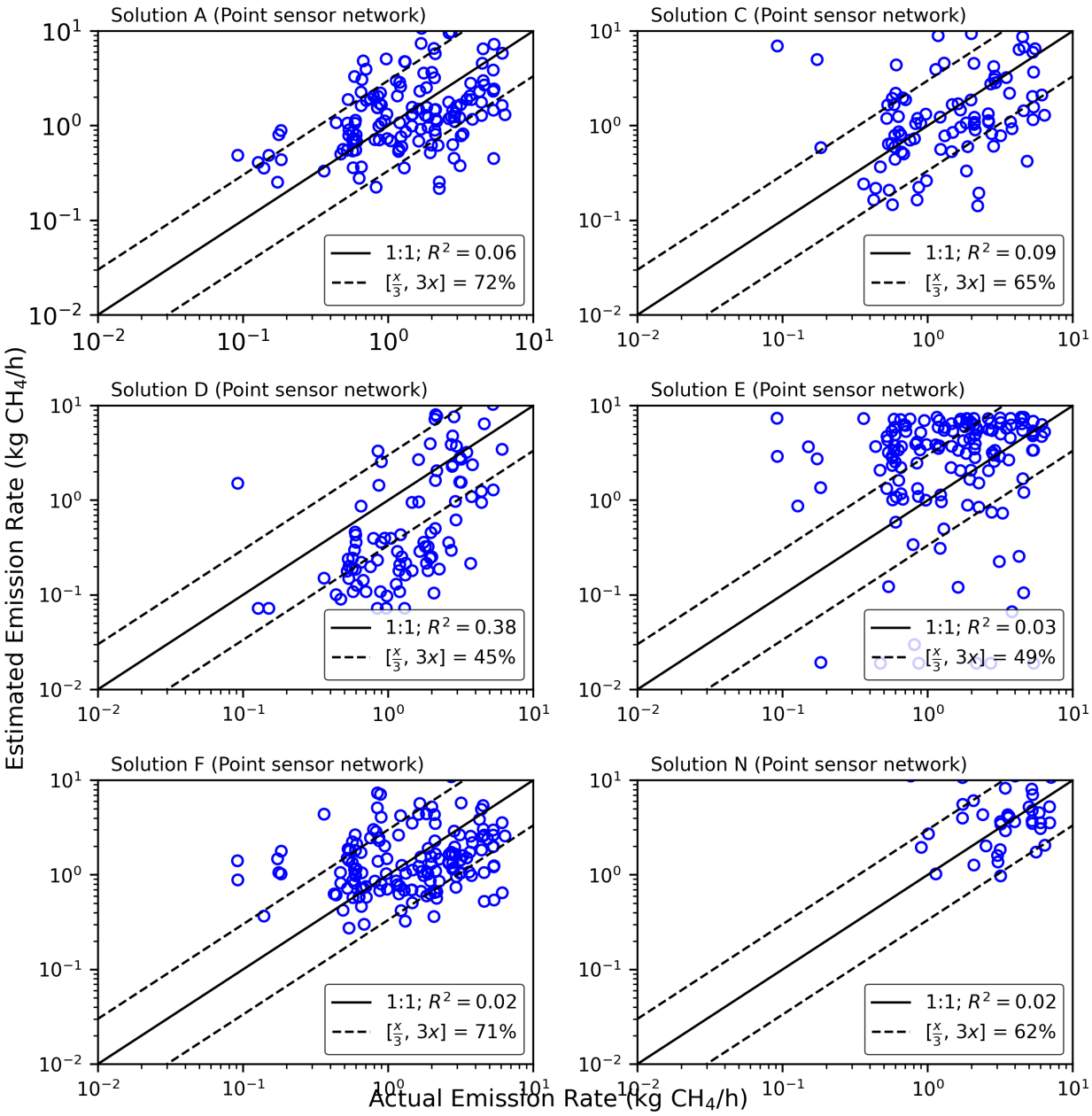


Figure B.1: Shows the plot of reported emission rates against actual (or controlled) emission rates for all point sensors network continuous monitors that tested quantification capability. The plot focuses on estimates when an ongoing emission event contains one controlled release. The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%)

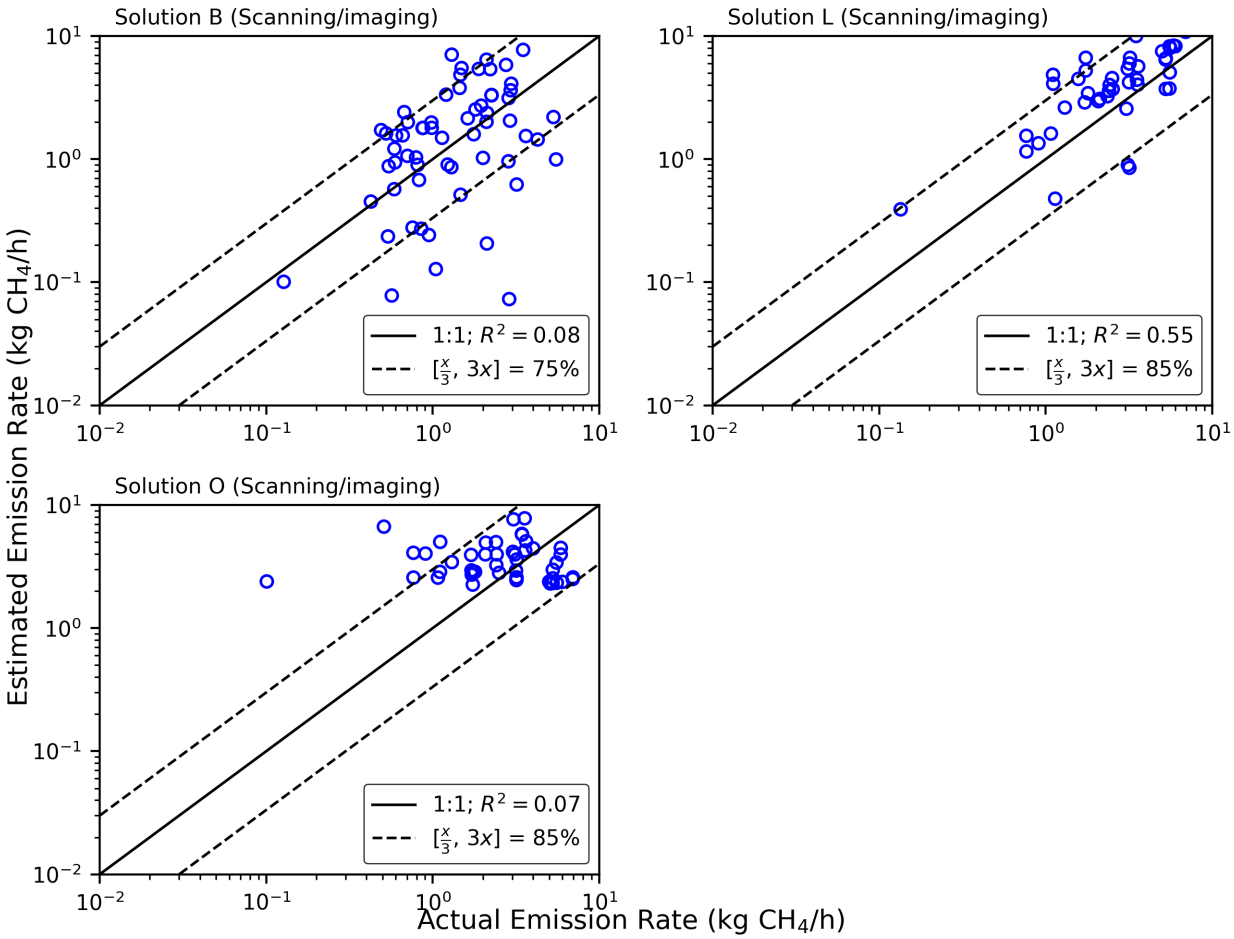


Figure B.2: Shows the plot of reported emission rates against actual (or controlled) emission rates for all scanning/imaging continuous monitors that tested quantification capability. The plot focuses on estimates when an ongoing emission event contains one controlled release. The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%). The R^2 illustrates the correlation between the reported and actual emission rates.

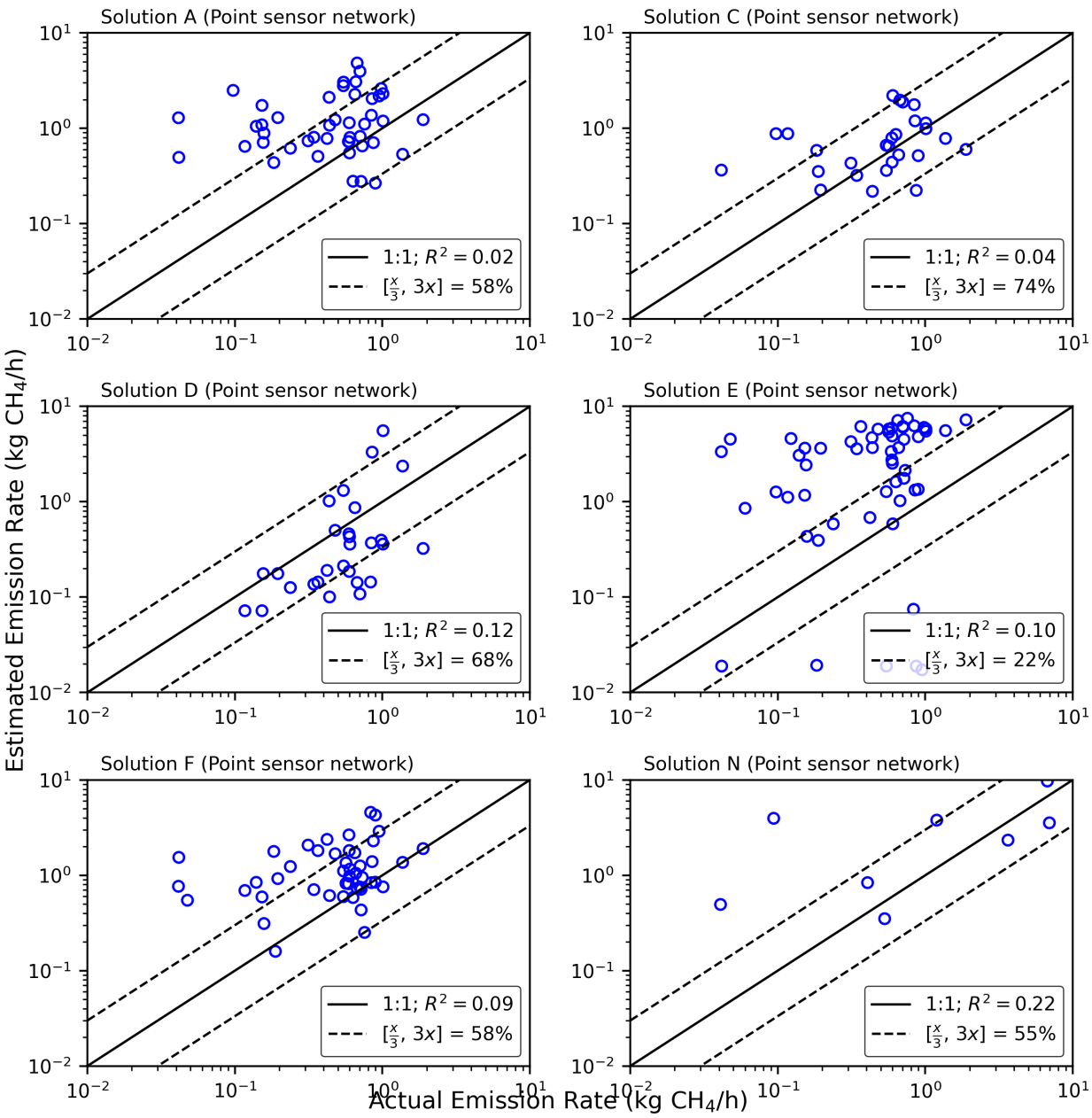


Figure B.3: Shows the plot of reported emission rates against actual (or controlled) emission rates for all point sensors network continuous monitors that tested quantification capability. The plot focuses on estimates of controlled releases that occurred during night time (Fort Collins Colorado, USA). The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%). The R^2 illustrates the correlation between the reported and actual emission rates.

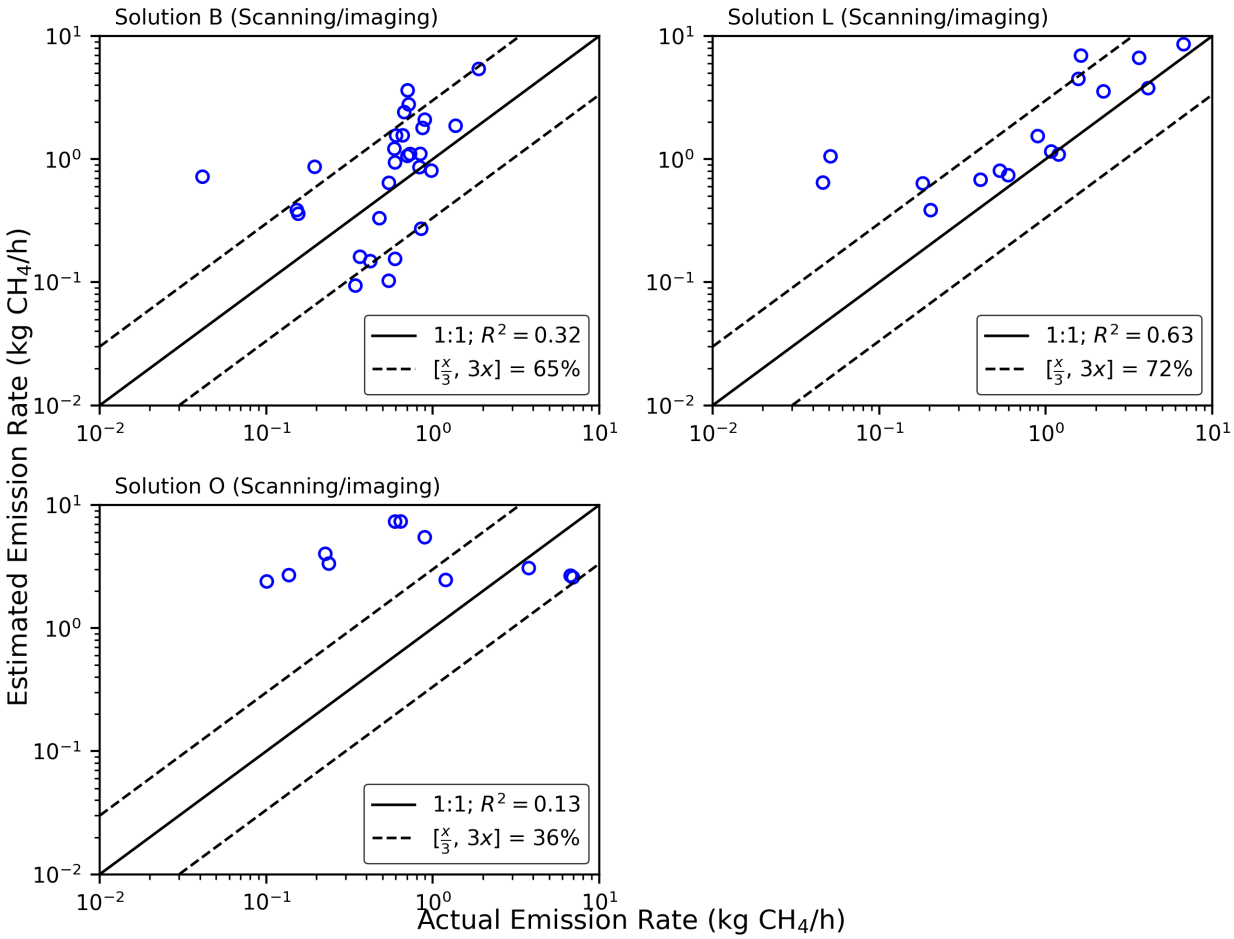


Figure B.4: Shows the plot of reported emission rates against actual (or controlled) emission rates for all scanning/imaging continuous monitors that tested quantification capability. The plot focuses on estimates of controlled releases that occurred during night time (Fort Collins Colorado, USA). The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%). The R^2 illustrates the correlation between the reported and actual emission rates.

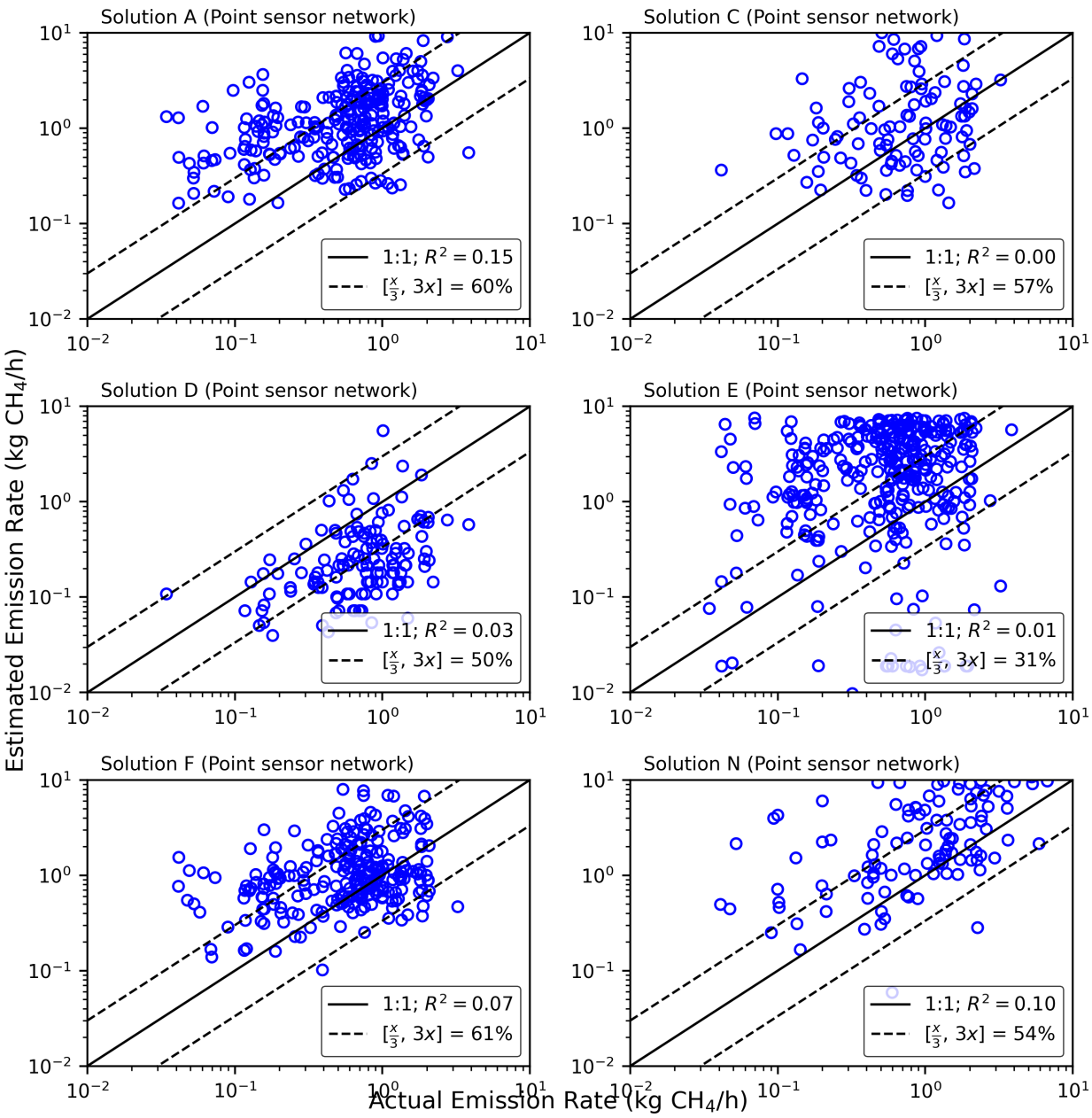


Figure B.5: Shows the plot of reported emission rates against actual (or controlled) emission rates for all point sensors network continuous monitors that tested quantification capability. The plot focuses on estimates when an ongoing emission event contains multiple controlled releases. The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%). The R^2 illustrates the correlation between the reported and actual emission rates.

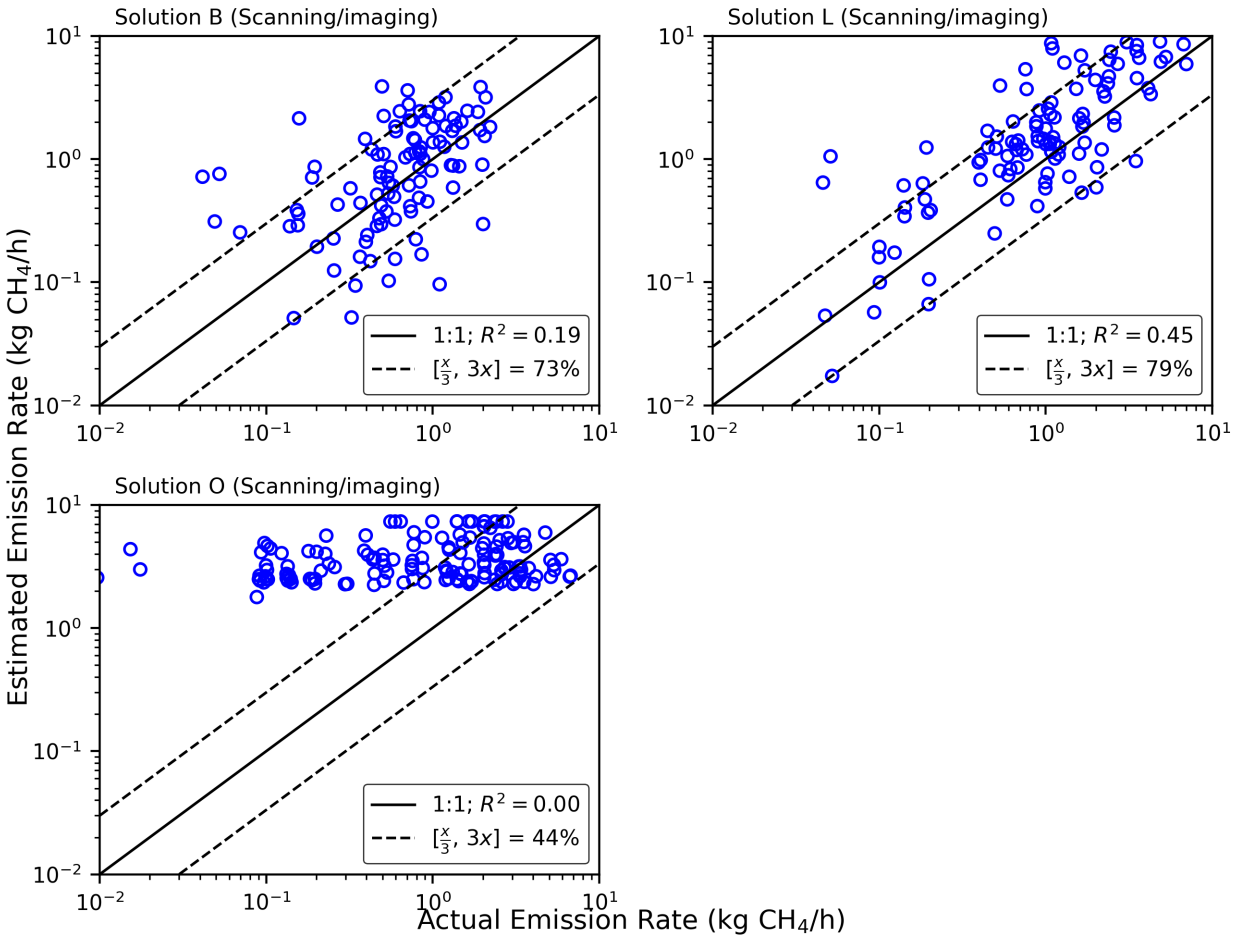


Figure B.6: Shows the plot of reported emission rates against actual (or controlled) emission rates for all scanning/imaging continuous monitors that tested quantification capability. The plot focuses on estimates when an ongoing emission event contains multiple controlled releases. The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%). The R^2 illustrates the correlation between the reported and actual emission rates.

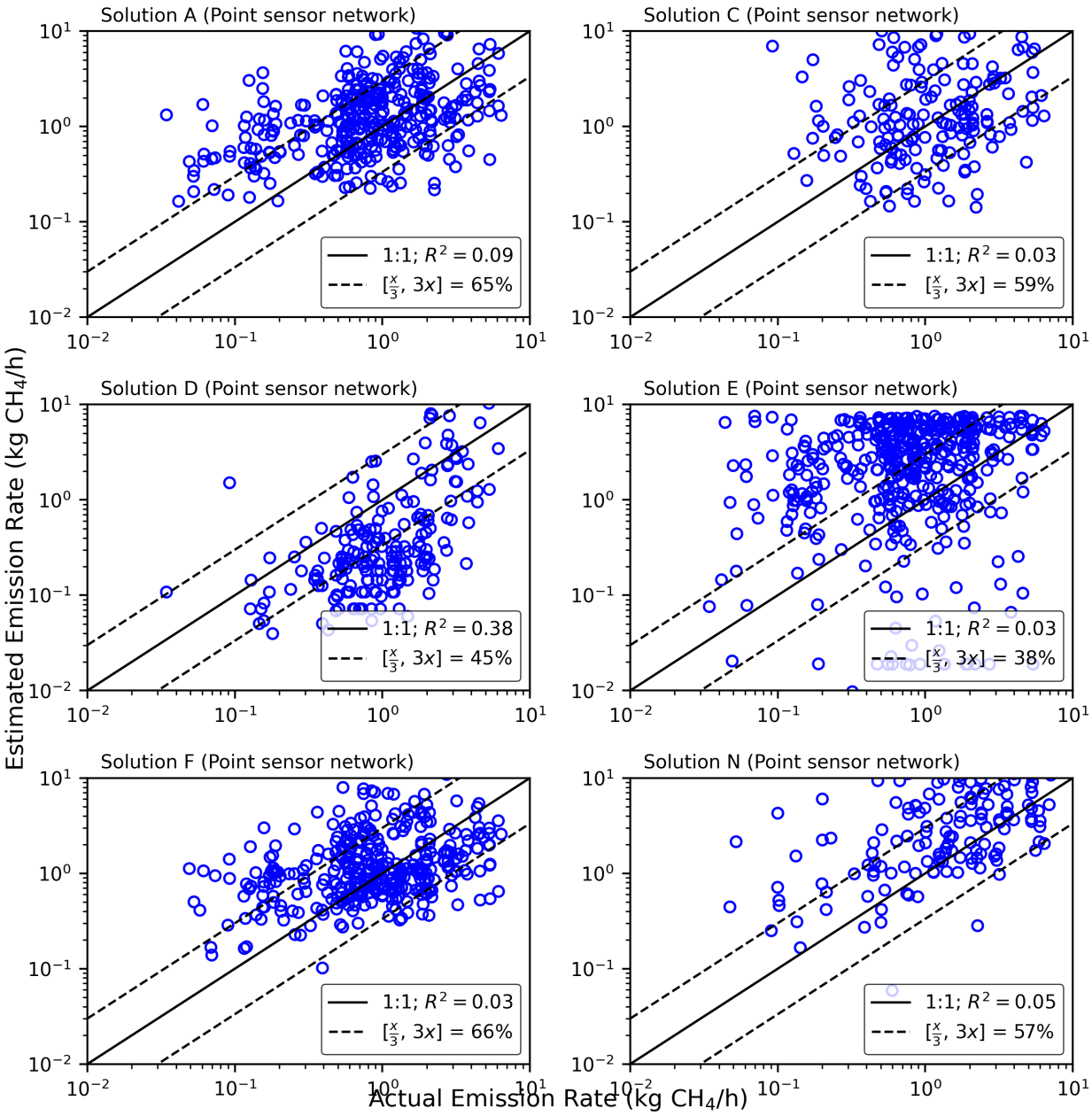


Figure B.7: Shows the plot of reported emission rates against actual (or controlled) emission rates for all point sensors network continuous monitors that tested quantification capability. The plot focuses on estimates of controlled releases that occurred during day time (Fort Collins Colorado, USA). The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%). The R^2 illustrates the correlation between the reported and actual emission rates.

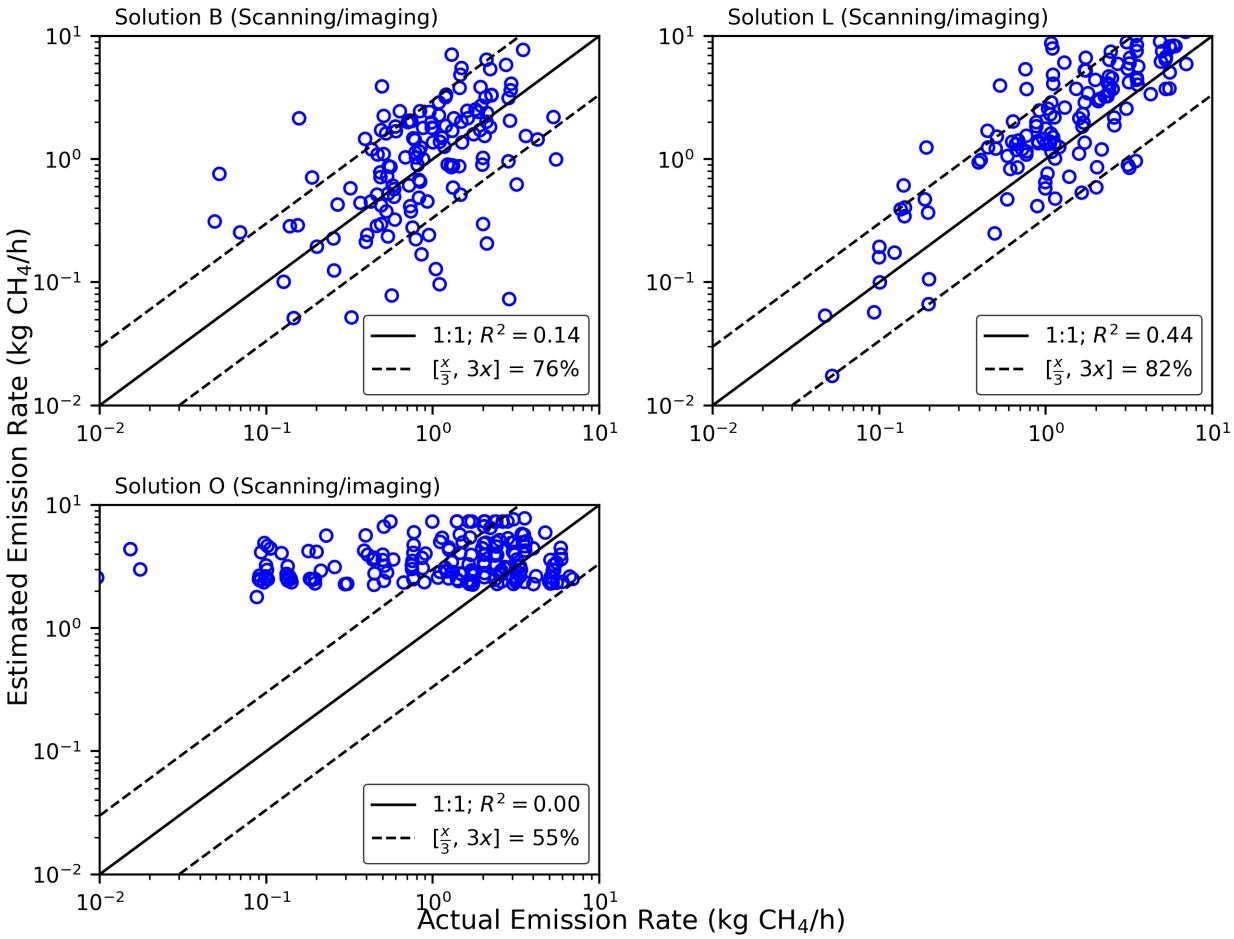


Figure B.8: Shows the plot of reported emission rates against actual (or controlled) emission rates for all scanning/imaging continuous monitors that tested quantification capability. The plot focuses on estimates of controlled releases that occurred during day time (Fort Collins Colorado, USA). The black solid line represents the 1:1 line; data points along this line illustrate that the reported equals the actual emission rate. The black dotted lines highlight the region where the single estimates are within a quantification factor of 3 of the actual emission rates (i.e., relative error from -67% to 200%). The R^2 illustrates the correlation between the reported and actual emission rates.