Performance of survey solutions under single-blind controlled testing protocol

Chiemezie Ilonze,† Rachel Day,‡ Ethan Emerson,‡ Aidan Duggan,‡ Ryan Brouwer,‡ and Daniel Zimmerle*,‡

†*Department of Mechanical Engineering, Colorado State University, Fort Collins, CO 80523, U.S.A*

‡*Energy Institute, Colorado State University, Fort Collins, CO 80524, U.S.A*

E-mail: Daniel.Zimmerle@colostate.edu

Abstract

Standardized controlled testing of advanced methane detection technologies (solutions) has been identified as a step in demonstrating the emissions mitigation equivalence between these solutions and existing regulatory-approved leak detection and repair methods (e.g., ground-based optical gas imaging [OGI] camera survey) in the US and Canada. In this study, 12 solutions consisting of 4 handheld OGI cameras, 4 advanced handheld systems, and 4 mobile (automobile- and drone-based) solutions were tested under a single-blind controlled testing protocol at different periods between 2021 and 2023 at an outdoor test facility that simulates emissions from a simple, onshore North American production oil and gas (O&G) facility. Three solutions were tested again 3–12 months after the first test, using the same test protocol and facility to assess how performance changed over time. Results showed that handheld OGI cameras had comparable or better performance in terms of lower 90% probability of detection (DL90), false positive and negative fractions, and higher equipment unit-level localization performance compared to other categories of solutions tested. Advanced handheld solutions had comparable performance with the OGI cameras across all metrics except false positive fraction (much higher), while mobile solutions generally had shorter survey durations compared to other solutions. For solutions that tested twice, the performance of 2 of 3 solutions generally improved, illustrating the benefit of regular, comprehensive testing in the development of solutions. The different value propositions inferred from the various categories of solutions tested suggest that mobile solutions can rapidly survey larger areas to inform more targeted, follow-up inspections with handheld solutions. While advanced handheld solutions can be as effective as OGI cameras, mobile solutions have some difficulties to overcome to make them directly comparable or need to have a different use case.

Synopsis

To achieve further emissions reductions across the O&G industry, advanced methane detection solutions are potential alternative methods for regulatory LDAR. This study assesses these solutions' performance through robust, controlled testing.

Keywords

Methane, survey solutions, emissions mitigation, detection limit, source attribution, natural gas, survey time, localization accuracy and precision, leak detection and repair

Introduction

The increase in human-induced methane emissions over the last 150 years has contributed to climate change, and mitigating these emissions is an immediate target to limit the planetary mean temperature rise to 1.5 °C to 2.0 °C above the pre-industrial level this century.[1–4] Methane, a short-lived (≈12 years) potent greenhouse gas with a 100-year global warming potential of 28 times that of CO_2 is the primary component of natural gas commonly emitted from the oil and natural gas (O&G) industry.[4] O&G emissions are the second largest source of methane emissions in the United States (US), and it is estimated that methane emissions could be reduced by 75% from the O&G sector globally by implementing already established mitigation approaches, including leak detection and repair (LDAR).[5]

LDAR programs are implemented in the O&G sector to monitor and identify equipment component leaks (fugitive emissions) to reduce emissions through leak repairs.[6,7] Current regulations in the US and Canada require periodic ground-based LDAR using handheld optical gas imaging (OGI) cameras or portable volatile organic compound analyzers (US Environmental Protection Agency (EPA) Method 21)[8,9] with their emission reduction potential depending on survey frequency,[10] surveyor's experience,[11] and scale of application.[12] Although these approaches can precisely localize sources, they are labor-intensive due to their small scope of application (component level) compared to the extensive spatial scale of O&G infrastructure. To be regulatory-compliant in the US and Canada, LDAR must be applied to millions of equipment components spread over large geographical areas with variable OGI assets, which has significant time, human labor, and cost implications for O&G operators.[12]

To address some of these limitations, there has been rapid and continued development of advanced leak detection and quantification (LDAQ) solutions and the increased use of these solutions by O&G operators to internally supplement existing LDAR inspection methods.[12,13] These solutions are attractive given that several advanced LDAQ technologies can detect emissions quickly and cost-effectively (when deployed to monitor large geographical areas) with less susceptibility to human errors, thus representing a promising opportunity to curb methane emissions from the O&G industry.[14–16] LDAQ solutions are either autonomous, fixed, ground-based sensors (continuous monitors) or mobile technologies (handheld, drone-based, automobile-based, and aircraft-based) that require human supervision to operate (survey solutions). Some survey solutions (e.g., aircraft-based technologies) can screen larger spatial areas of O&G facilities to prioritize ground-based inspections of identified emissions using traditional LDAR methods (e.g., OGI camera surveys), rather than requiring those LDAR methods to be applied to every facility. Advances in the spatial and temporal extent of surveys is essential for emission reductions, given that emissions are spatially and temporally variable[17–23] with a small number of facilities and emitters often responsible for a disproportionately large amount of total emissions.[21,24–28] The recent US EPA final New Source Performance Standards (NSPS) rules on alternative methane detection approaches have stipulated work practices for the use of advanced methane detection technologies as an alternative to existing regulatory-approved LDAR approaches at well sites, centralized production facilities, and compressor stations, if the technologies meet certain performance standards and are regulatory-approved.[13] Essentially, LDAR based on advanced survey solutions must demonstrate equivalent or better emissions reduction than that based on traditional approaches (OGI camera surveys, etc.). A consensus framework by Fox et al. identified rigorous controlled testing of solutions using standardized protocols as a step in achieving this objective.[29]

Previous controlled testing of survey solutions used study-specific protocols with limited test complexity[30–35] or standardized protocols with narrow scope of application.[32] These studies assessed the performance of aircraft-based,[30–34] drone-based,[31,35] automobile-based,[31,32] and handheld solutions.[32] Studies by Ravikumar et al.[31] and Bell et al.[32] during the ARPA-E MONITOR[36] and the Stanford - Environment Defense Fund Mobile Monitoring Challenge programs, respectively, are notable for testing and comparing the performance of multiple, different survey solutions. While Ravikumar et al. used a series of non-standardized protocols with varying complexity and scope during testing, Bell et al. applied a standardized protocol (an early version of the protocol used in this paper - SI Section S-5) that was not representative of expected field application of survey solutions. The protocol used by Bell et al. limited the emissions scenarios and the number of controlled gas releases conducted during the study, influencing tested solutions' survey methodology and pace. In addition, a recent study by Liu et al. evaluated the quantification performance of 10 solutions composed of mobile (drone, automobile, and aircraft), ground-based (fixed camera system), and handheld (OGI camera) technologies at a non-operational compressor station in Spain. While the study tested a wide range of controlled release rates (0.01–50.0 kg CH_4/h) and emission point heights (1–28 m), the study was partially blind (i.e., only release rate was unknown to participants) and implemented study-specific test protocol (i.e., timing of controlled releases was constrained to ≈2 hours, etc.) with a small number of experiments (17).[37] These studies therefore suggest the need for consensus, standardized, and field-representative testing protocols for assessing the performance of survey solutions.

In this paper, we present results of the first controlled testing of a variety of survey solutions using a consensus and field-representative protocol developed by the Advancing Development of Emissions Detection (ADED) project.[38] Although the survey test protocol has been used for aircraft-based solutions,[39–41] this study is its first implementation for automobile-based, drone-based, and handheld (including OGI cameras) solutions. In this study, 12 different solutions were tested at different periods between May 2021 and November 2023 (3 solutions tested twice) at the Methane Emissions Technology Evaluation Center (METEC), Colorado State University (CSU), Fort Collins, Colorado. For each solution, the study evaluated metrics such as the probability of detection (POD) curve, false positive fraction (FPF), false negative fraction (FNF) (i.e., 1 - true positive fraction (TPF)), survey time, and localization accuracy and precision, while also assessing how the performance of solutions that were tested twice changed based on these metrics.

Methodology

Test Facility

Controlled-release testing was conducted at METEC at different periods between May 3rd, 2021, and November 5th, 2023. METEC is an open-air test and research facility (GPS coordinates: 40.59559, -105.13984) located at CSU, Fort Collins, Colorado, 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, etc.). METEC is composed of decommissioned surface O&G equipment (e.g., wellheads, separators, condensate tanks, flare stacks, and a compressor and dehydrator unit) embedded with strategically hidden emission points, which are arranged into five well pads and a small compression station. Each equipment unit and group is identified with unique tags, respectively (SI Table S-1). The test center defines an equipment group as a cluster of adjacent, similar equipment units in a well pad. Controlled releases (emissions) are actuated electronically to transport gas from onsite compressed natural gas (CNG) storage tanks through buried, small-diameter steel tubing to the emission points. A matrix of thermal mass flow meters (OMEGA FMA-17xx series) and manual pressure regulators downstream of the CNG storage tanks control gas flow to emission points, and an onsite gas chromatography device is used to determine the composition of gases released at the site at all times. METEC also has a ≈7 m stationary 3-D sonic anemometer that records meteorological data (e.g., relative humidity, wind speed, etc.) at a frequency of 1 Hz. See Zimmerle et al., Bell et al., and Ilonze et al. for more details about the test center and the gas transport system.

Testing Process

The controlled testing protocol used in this study was a result of the same multi-stakeholder collaboration under the ADED project that produced the protocols used by Bell et al., Ilonze et al. to test continuous monitors.[42,43] A protocol development committee (PDC) consisting of more than 75 members consisting of environmental NGOs, state and federal regulators, LDAQ technology developers, academic institutions, and O&G companies worked on the protocol draft developed by CSU. The protocol consisted of testing methodology and performance metrics, and went through 1 PDC review cycle to obtain the currently published version.[44] The protocol tests and evaluates a survey solution's system-level performance: the performance of the methane sensor, data analytics, and deployment (e.g., number of passes, flight speed, height, etc.) as an integrated unit. The sensor component of a solution collects raw measurement data (e.g., ppm, ppm-m, etc.) while aboard a deployment platform (e.g., aircraft, automobile, handheld, etc.) during a flying, walking, or driving survey at a preplanned pace, distance, and height over or around O&G asset(s). The raw data gathered is then processed by solutions' data analytics/algorithm, which might be proprietary, to produce actionable data (e.g., identifying a leak, localizing an emitter, estimating emission rate, etc.). The survey protocol divides testing into 3 activities:

- **Documentation of the system under test:** 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. See details in Section 5.1 of the survey protocol.[44]

- **Emission surveys:** The survey protocol divided testing into *experiments* conducted during the day (i.e., typically between 8 a.m. and 5 p.m. US mountain time) for 3 to 5 days. An experiment consisted of one or multiple, simultaneous controlled releases of CNG, each emitting at a steady emission rate for minutes to hours with longer-duration experiments designed to investigate performance variation due to changing meteorological conditions. Testing was single-blind, as the solutions were unaware of the timing, location, and release rate of controlled releases. 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, 3, and 4 & 5 to achieve different site sizes and operational configurations as shown in SI Figure S-1.

While testing across multiple O&G infrastructures and wellpads 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 wellpads also varied testing complexity and rigor. For this study, well pad 1 & 2 (20 m $\times$ 10 m) contained 6 equipment units, well pad 3 (10 m $\times$ 60 m) had 7 equipment units, while well pad 4 & 5 (60 m $\times$ 120 m) contained 18 equipment units. Additionally, to further vary test complexity, each experiment's controlled release(s) were either restricted to a well pad or was distributed across all or a portion of the survey area defined by METEC for this study.

All controlled releases were CNG, with a mean gas composition by volume of 76.0%–88.0% methane, 11.6%–20.1% ethane, 1.4%–3.6% propane, and trace amounts of other gases. The release rates of gases tested match to within an order of magnitude the component-level leak sizes previously measured by the study team at natural gas production sites in the US, excluding liquid unloading or major failure conditions that produced exceptionally high-rate emission events,[45–47] which the type of 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., 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 experiment) not only to improve testing complexity but 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 will be conducted for the day. 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 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 and had instructions to avoid communicating with one another to prevent one team from learning about emission points or rates from another team. Measurements by solutions were sequenced to limit the impact (if any) of one system's detection methodology on others. The one exception to this goal was a scenario that included both a drone-based solution and an automobile-based solution. Drone flights generate turbulence and disturbances that can affect the formation and downwind transport of gas plumes to be measured by the automobile-based solution typically driving at the edge of the survey area. The end of the emissions survey by solution(s) signaled the end of each experiment, then the test center stopped all controlled releases.

- **Detection reporting:** The test protocol stipulates a reporting template for solutions to record data of experiments and detected emissions. Solutions conducted quality control and validation of their data before reporting to the test center. Solutions 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 (SI Section S-6). Reporting measured gas species allowed performance analysis of solutions with either methane-specific or multi-species sensors to be on the same basis. See sections 5.3.1 to 5.3.3 of the protocol for comprehensive details of all the data that solutions can report during the study.[44]

Performance Metrics

For data to be considered valid, solutions were required to submit their survey and detection reports within one week of test completion. This time limit reflects common field practice.[22,48,49] 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 solutions to perform *detection classification*. 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 data (i.e., controlled releases and detection reports) invalidated by the requirements of the test protocol.[44] For each 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 releases 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). The result of the detection classification process involving pairing controlled releases with detection reports, as summarized in SI Section S-4, was used to evaluate the performance metrics in sections 6.2 and 6.3 of the protocol. Key performance metrics for this study 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 true positive (TP) detections to the sum of the counts of TP and false negative (FN) detections over a set of conditions as shown in equation 1.

$$\mathrm{POD}\big|_{\mathrm{x}} = \left.\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FN}}}\right|_{\mathrm{x}} \tag{1}$$

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

- *False Positive Fraction (FPF):* This is the fraction of the count of false positive (FP) detections to the sum of the counts of TP and FP detections as shown in equation 2.

$$\text{FPF} = \frac{N_{\text{FP}}}{N_{\text{FP}} + N_{\text{TP}}} \tag{2}$$

- *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 3.

$$\text{FNF} = \frac{N_{\text{FN}}}{N_{\text{FN}} + N_{\text{TP}}} \tag{3}$$

- *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 4, 5, and 6.

$$\text{Correct unit } (\text{LA}_{\text{unit}}) = \frac{N_{\text{TP}_{\text{unit}}}}{N_{\text{TP}} + N_{\text{FP}}} \tag{4}$$

$$\text{Correct group } (\text{LA}_{\text{group}}) = \frac{N_{\text{TP}_{\text{group}}} + N_{\text{TP}_{\text{unit}}}}{N_{\text{TP}} + N_{\text{FP}}} \tag{5}$$

$$\text{Correct facility } (\text{LA}_{\text{facility}}) = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \tag{6}$$

Participating Solutions

Participating 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 at different times within the study period, with three solutions tested again within 2 months to 1 year of their initial tests. 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.). Solutions provided all auxiliary equipment (e.g., anemometers, battery packs, power cables, internet connection, etc.) needed to perform emissions survey during testing. The list of participating 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 between the study team and the solutions, solutions were identified with arbitrarily selected, unique alphabetic identifiers.

For this study, participating survey solutions were grouped into three categories:

1. *Mobile solutions*: Survey solutions that are automobile- and unmanned aerial vehicle (UAV)/drone-based.

2. *Handheld OGI solutions*: Traditional handheld OGI camera solutions.

3. *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 1 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. The SI contains reports that comprehensively describes all the tested capabilities evaluated for each solution.

Table 1: Characteristics of participating solutions and testing conditions.

Solution			**Test Conditions**		**Test Year**		
ID	**Platform**	**Category**	**Release Rate ($g\ CH_4/h$)**	**Wind speed (m/s)**	**2021**	**2022**	**2023**
Solutions that participated in the first round of testing.							
A†	Vehicle	Mobile	214 [26, 895]	5.3 [1.1, 13.4]	✓	X	X
D	Drone	Mobile	73 [3, 297]	3.0 [0.9, 5.7]	✓	X	X
F	Drone	Mobile	471 [30, 2027]	2.5 [1.2, 4.1]	X	X	✓
I	Drone	Mobile	175 [22, 586]	4.5 [1.1, 9.3]	✓	X	X
B	Handheld	Handheld OGI	76 [4, 297]	3.3 [1.3, 5.7]	✓	X	X
E	Handheld	Handheld OGI	198 [4, 808]	3.0 [0.8, 4.6]	✓	X	X
G	Handheld	Handheld OGI	500 [22, 2110]	3.6 [0.9, 9.0]	X	X	✓
L	Handheld	Handheld OGI	553 [23, 2586]	3.6 [0.9, 8.7]	X	X	✓
C	Handheld	Advanced Handheld	198 [4, 808]	3.0 [1.1, 4.6]	✓	X	X
H‡	Handheld	Advanced Handheld	471 [3, 2106]	3.4 [0.9, 8.8]	X	X	✓
J	Handheld	Advanced Handheld	464 [23, 1651]	2.9 [0.7, 8.0]	X	X	✓
K	Handheld	Advanced Handheld	194 [22, 640]	5.2 [1.2, 12.8]	✓	X	X
Solutions that participated in the second round of testing.							
A†	Vehicle	Mobile	164 [11, 982]	4.3 [0.8, 13.3]	X	✓	X
K	Handheld	Advanced Handheld	164 [11, 982]	4.3 [0.8, 13.3]	X	✓	X
H‡	Handheld	Advanced Handheld	355 [4, 1934]	2.6 [0.9, 3.6]	X	X	✓

† The solution tested its quantification capability in 2022.
‡ The solution was tested twice in 2023.

Study Limitations

The limitations of the test center in comparison to actual onshore production O&G facilities have been extensively discussed by Zimmerle et al. and Ilonze et al.. In summary, the test center represents near-ideal operational field conditions with little ongoing operational activity that produces vibration, noise, heated sources, etc. For this testing, baseline emissions – commonly defined as routine, planned emissions, including combustion slip, gas pneumatics, periodic venting, and similar sources - were not included. As a result, certain failure conditions, such as a gas pneumatic venting in excess of designed rates, were not included as potential sources. Also, controlled testing may not completely capture the performance of solutions at actual O&G facilities, as Day et al. demonstrated for continuous monitors.

The emissions scenarios simulated for this study did not include large emission rate events, often responsible for a highly skewed share of total emission rates at O&G facilities.[21,24–27] This limitation was intentional. The objective of the study was to characterize the POD of solutions, and since all solutions tested here have a high probability of detecting emitters in excess of 3 kg/h, testing at higher rates provides little additional information. Additionally, the US EPA OOOO(b) regulation[52] was not published prior to these tests, therefore testing was not targeted at the rates specified in that regulation.

The test results described here are for *integrated solutions*, consisting of the sensor, deployment method, and analytics. Field and regulatory application of these results should not apply results from one solution developer to another using the same sensor and deployment method, as differences in analytics may produce significant differences in solution performance. As with other testing programs,[42,43] classification of excess or *redundant* TP detections (detections identifying controlled releases that have been matched already, as new and/or different emitters) for any single controlled release may be interpreted in multiple ways. Since close-range survey solutions (e.g., handheld emissions detection technologies) are generally used for follow-up investigations due to their high localization precision, a surveyor could interpret an extra detection report that incorrectly identified emissions (using GPS coordinates, etc.) from a component some distance from an actual emitter as an FP detection. Conversely, in a different application, such an extra report might not be considered an FP detection if it correctly attributes emission to the emitting equipment unit. Given these possible interpretations of *redundant* TP detections, Table S-7 in the SI breaks down the FPF of solutions shown in Table 2 into FPs due to (1) no-ongoing controlled releases, and (2) redundant TP detections. See Bell et al. for other potential limitations of the protocol's detection classification scheme.

Data Analysis

To evaluate the POD curves of solutions, three curve-fitting models (power function, binary logistic regression function, and exponential function) were assessed and compared using the resulting data of the detection classification framework. The exponential function (equation 7) was selected for two reasons. First, in several cases, the binary logistic regression function predicted non-zero POD at zero controlled release rates, which is unrealistic (see Bell et al.). Second, there were insufficient experimental points to bin data as required for a power curve fit as used in other studies (e.g. Zimmerle et al., Ilonze et al.). For the solutions tested, the range of valid classified data (TP and FN) points for each solution was 70 to 224, with 11 of the 15 solutions controlled testing producing less than 120 data points. The exponential link function selected was:

$$\text{pod} = 1 - \exp(-a \cdot x^b)\big|_{\text{a,b are curve fitting parameters}} \tag{7}$$

The curve fitting parameters for the POD curve (equation 7) were selected using least squares methods (SI Section S-3). The original data was bootstrapped 500 times to generate the uncertainty on any POD curve, with the 95% empirical confidence interval (CI) captured as the 2.5 and 97.5 percentiles. The study used 4 statistical tests to evaluate statistical significance in terms of correlation between 2 variables (Rank-Biseral correlations) and the difference between groups (2-sample T-test, Mann-Whitney U, and Kolmogorov-Smirnov tests) as described in SI section S-3.

Results and Discussions

This section covers results and discussions about solutions' performance and change in performance (i.e., for solutions that tested multiple times) based on the following metrics: probability of detection, emission source localization, and survey time. Analysis on emissions quantification performance was excluded as only two solutions tested that capability over two different rounds of testing. However, quantification performance and other metrics not captured in this paper are summarized in each solution's test report attached with the SI.

Handheld OGI Camera Solutions vs Other Categories of Solutions

Overall, handheld OGI camera solutions had comparable or better performance than other solution categories across most metrics assessed. Figures 1, 2, and 3 shows the POD curves of solutions as a function of controlled release rate. A multivariable logistic regression analysis evaluating the impact of release rate, wind speed, wind direction, and ambient temperature on the emissions detectability (TP and FN classification $\rightarrow$ POD) of solutions indicated that emission rate was statistically significant ($p < 0.05$) for 5 (C, D, E, H, and J) of 12 solutions that tested in the first round (SI Table S-4). Other variables (temperature, wind speed, and wind direction) were separately statistically significant to just one solution each. Measurement conditions like survey height and speed were not input parameters in evaluating the POD curves of solutions since they were either approximately constant during each solution's survey or were not documented by the surveyors. The POD curves of solutions characterized the likelihood of detecting emission at various release rates over a wide range of environmental conditions.

In Figures 1, 2 and 3, for each solution that tested twice (A, H, and K), the POD curves from all rounds of testing are shown in one plot to facilitate performance comparison. Each curve is identified by the year of testing (2021, 2022, or 2023) and the round of testing for a solution during the year (first round — 1 or second round — 2). Similar to Bell et al. and Ilonze et al., this study evaluated the detection sensitivity of solutions by defining each solution's detection limit (DL90) as the minimum emission rate the solution can detect 90% of the time across multiple observations over a wide range of weather conditions (i.e., the emission rate the solution has 90% POD). The lower the DL90, the more likely a solution detects a wider range of emission rates at O&G facilities, although the desired DL90 depends on field application. For solutions intended for follow-up investigations and fugitive emissions detections, their DL90 should be significantly lower than those used for screening purposes.

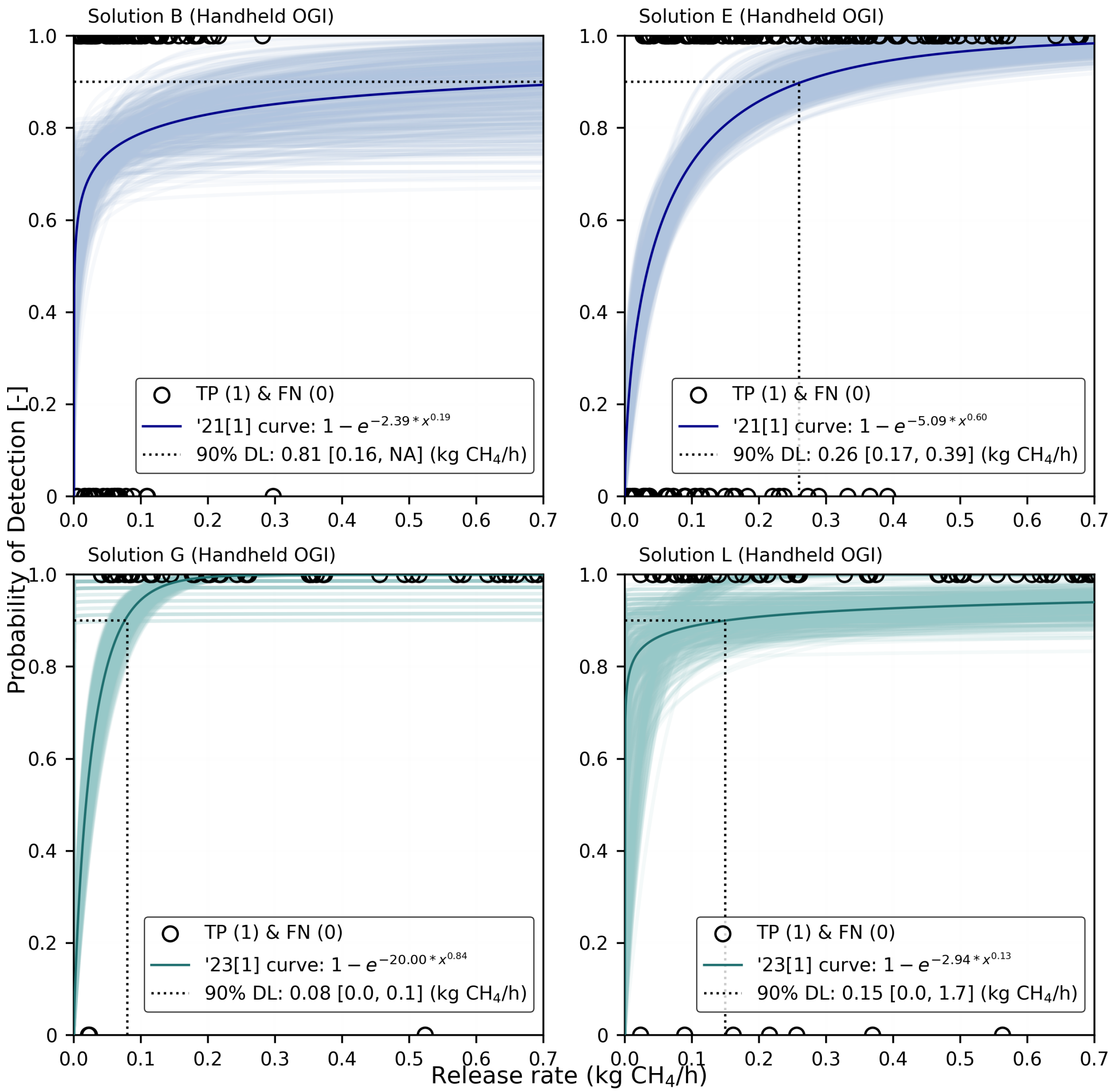

Figure 1: Probability of detection (POD) as a function of controlled release rate (kg CH_4/h) for all handheld OGI camera solutions. The blue and green colors illustrate POD curves from the first round of tests conducted in 2021 and 2023. The DL90 is the emission rate at which the solution achieves 90% POD. 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 ×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 solutions.

Of the 5 solutions whose emissions detectability (i.e., POD) was statistically influenced by release rate, only solution D did not have a statistically significant correlation (Rank-Biseral correlation, $p < 0.05$) between release rate and emissions detectability (SI Section S-7.1). This suggests that the solution's POD curve will likely benefit from accounting for windspeed (which was statistically significant to the solution's emissions detectability), as shown in SI Figure S-7. Similarly, the POD curves of the remaining (7 of 12) solutions that were not statistically influenced by release rate, wind speed, or/and temperature 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 leak source, etc. For example, von Fischer et al. showed that detection limits of automobile-based solutions tend to increase with measurement distance. Results also indicated that the POD curves (i.e., the best estimate curve with the lower and upper confidence limit curves obtained through bootstrapping as shown in Figures 1, 2, and 3) of all but solutions J and D (with that of A and I unclear) agree closely with the curves obtained by scaling POD curves normalized by windspeed by the mean, minimum, and maximum wind speed tested for each solution (SI Figures S-8, S-9, and S-10). This result suggests that the POD curves of most of the tested solutions (especially OGI camera solutions) can also be approximately expressed as a function of emission rate normalized by windspeed as shown in SI Figures S-5, S-6, and S-7.

In general, Figures 1, 2, and 3, and Table 2 shows that the POD curves of solutions could only predict the DL90 of 7 of 12 solutions with values ranging from 0.08 [0.00, 0.10] kg CH_4/h to 0.81 [0.16, NA] kg CH_4/h with 6 solutions within their tested emission rate range. The five solutions that could not predict their respective DL90s or had large DL90 values ($> 70\times$) relative to the maximum release rates tested consisted of all 4 mobile solutions (drone- and automobile-based) tested in the study. Table 2 also shows that mobile solutions had some of the largest FPF (19.2% to 50.0%) and FNF (20.3% to 48.5%) values evaluated in this study. This result is likely due to potential difficulties at either distinguishing methane (or other gas species of interest) signals from background noise by these solutions (increases FPF), high detection limits (increases FNF), or redundant TP detections due to the limitations of the test protocol (inflates FPF) (SI Table S-7).

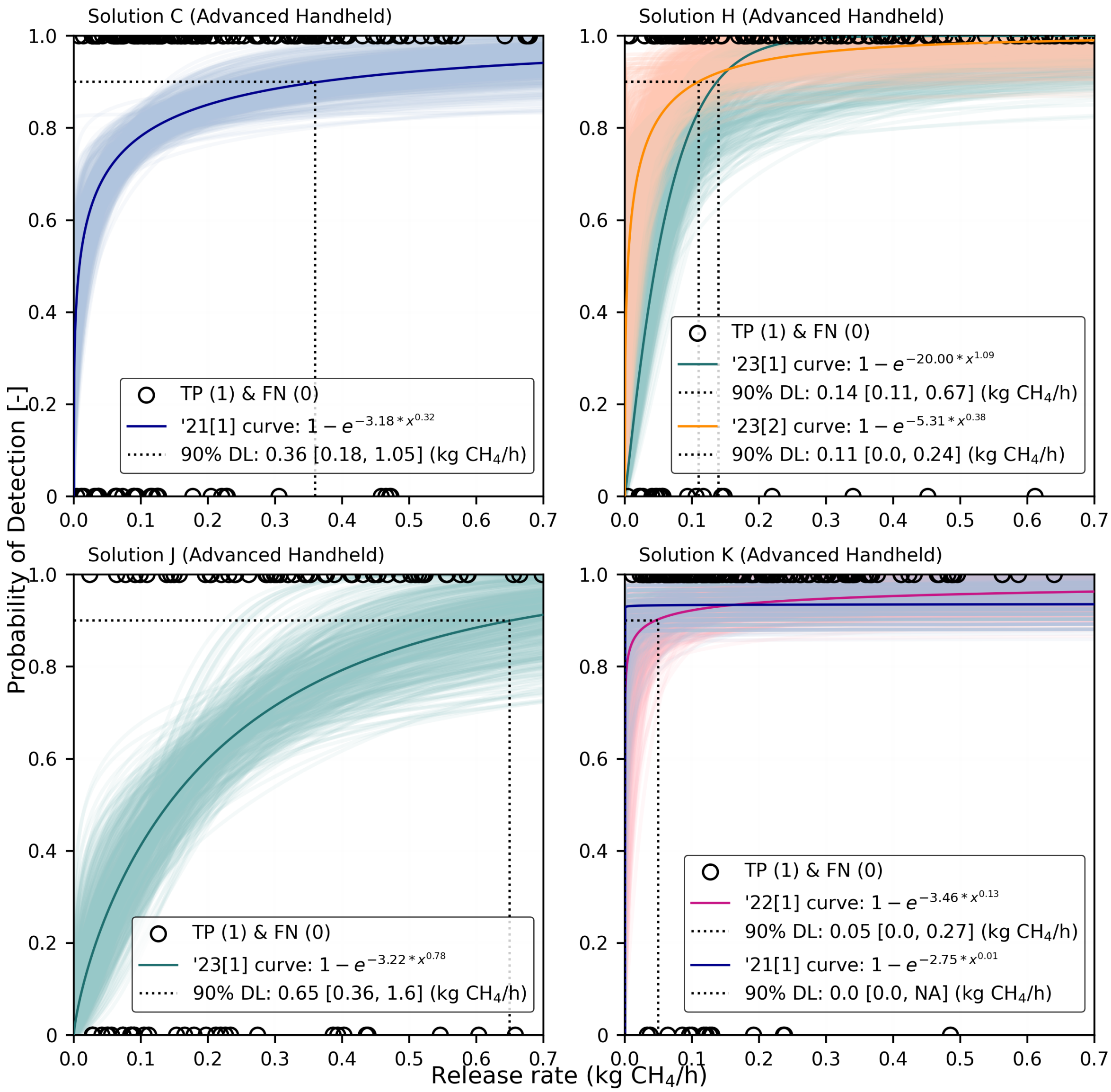

Figure 2: Probability of detection (POD) as a function of controlled release rate (kg CH_4/h) for all Advanced handheld solutions. For solutions that tested twice, the POD curve from each test period is overlaid and identified by the test year (2021, 2022, or 2023) and iteration number (1 or 2) for solutions that tested twice within a year. The blue and red colors illustrate POD curves from the first round of tests conducted in 2021 and 2022, while the green and orange colors are for the first and second rounds of tests conducted in 2023. The DL90 is the emission rate at which the solution achieves 90% POD. 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 ×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). Solution K (2021) has near-uniform POD across tested emission rates (rank-biserial correlation coefficient ≈ -0.05). 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 solutions.

The DL90s of 3 of the 4 handheld OGI camera solutions tested were among the top 4 lowest DL90s obtained in the study. Unlike advanced handheld solutions, emission rate and other test conditions were not statistically significant to the emissions detectability of the 3 camera solutions which aligns with data from Zimmerle et al. that showed that the experience level of an OGI camera surveyor had a more statistically significant impact on emissions detection rate than other variables tested. Similarly, Ravikumar et al. identified the distance between an OGI camera and a leak source as the most important parameter affecting the detection effectiveness of handheld OGI camera solutions.[54] Since solution vendors conducted the emissions surveys in this study, it is likely that surveyors sent by the solutions were highly experienced and used that expertise to maximize solution performance; Deployment of the same solutions in field conditions may not have access to this level of experience. Advanced handheld solutions use instruments that measure ambient emission concentration (like mobile solutions) but are close enough to sources that there is less dependence on favorable plume transport (unlike mobile solutions). As a result, the POD of advanced handeld solutions is dependent on emission rate for 3 of the 4 advanced handheld solutions tested.

Additionally, handheld OGI camera solutions had the lowest FPF values in the study (0.0% to 4.3%) which were statistically significantly lower (T and Kolmogorov-Smirnov tests; $P < 0.05$) than that of mobile solutions but comparable to advanced handheld solutions (4.8% to 35.6%). Although the FNF values of 2 of 4 handheld OGI camera solutions were less than 10%, the FNF values of handheld OGI camera solutions were not statistically different from that of mobile and advanced handheld solutions, respectively.

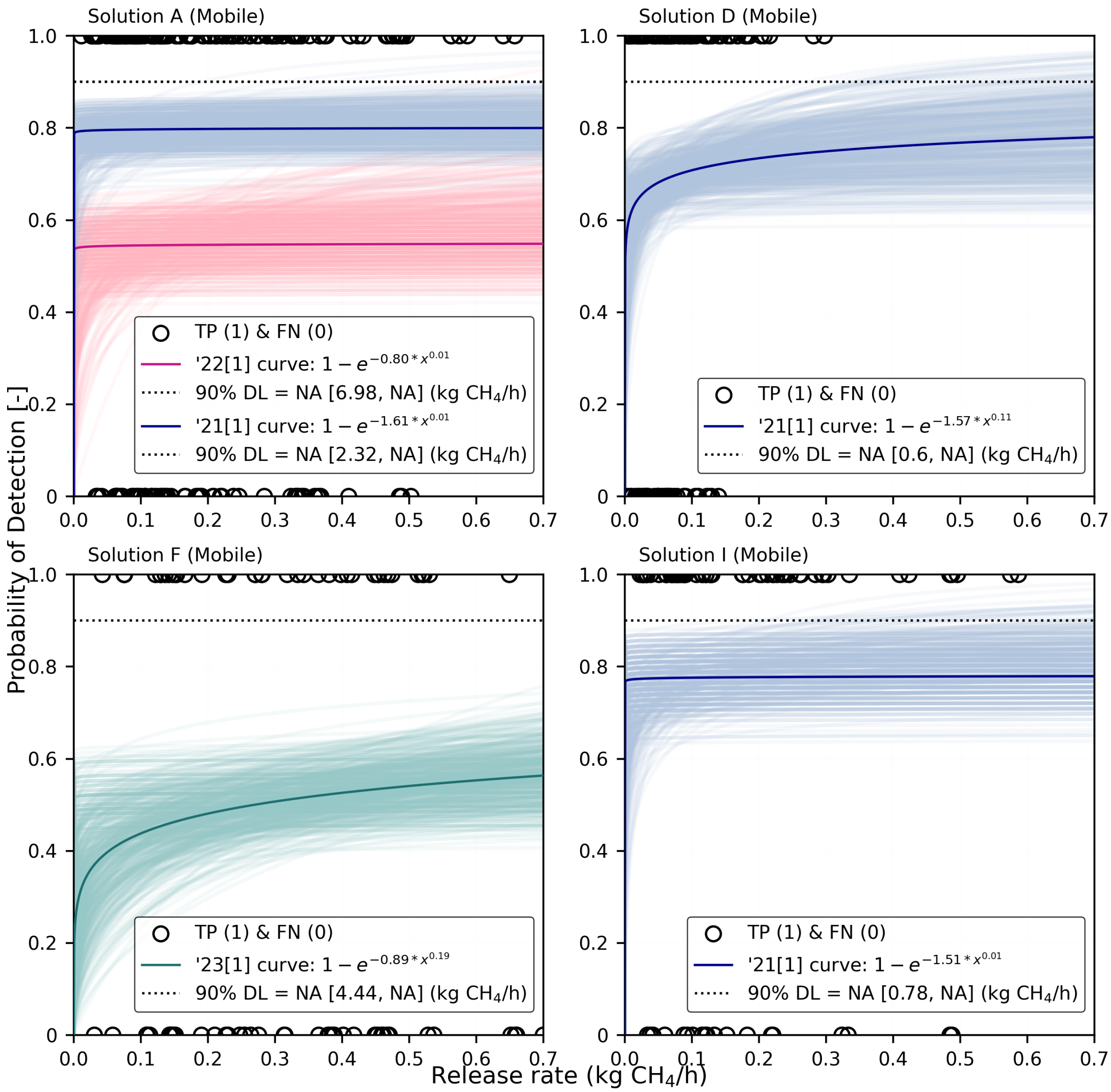

Figure 3: Probability of detection (POD) as a function of controlled release rate (kg CH_4/h) for all mobile solutions. For solutions that tested twice, the POD curve from each test period is overlaid and identified by the test year (2021, 2022, or 2023) and iteration number (1 or 2) for solutions that tested twice within a year. The blue, red, and green colors illustrate POD curves from the first round of tests conducted in 2021, 2022, and 2023. The DL90 is the emission rate at which the solution achieves 90% POD. 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 ×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). Solutions A and I have near-uniform POD across tested emission rates (rank-biserial correlation coefficient between -0.01 and -0.05). 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 solutions.

While these results highlight why handheld OGI camera solutions are typically used, and are anecdotally effective, for follow-up investigations and regulatory-compliant LDAR inspections at O&G facilities, results also showed that the DL90s of all advanced handheld solutions were within the minimum detection threshold stipulated in the EPA's final rule (NSPS OOOO(b) rules) for quarterly monitoring (≤ 1 kg CH_4/h) of well sites and centralized production facilities with alternative technologies.[13] For reference, see SI Figures S-11, S-12, and S-13 for POD curves and DL90s derived using the power function, and SI Figures S-14, S-15, and S-16 for POD curves and DL90s based on binary logistic regression for all solutions.

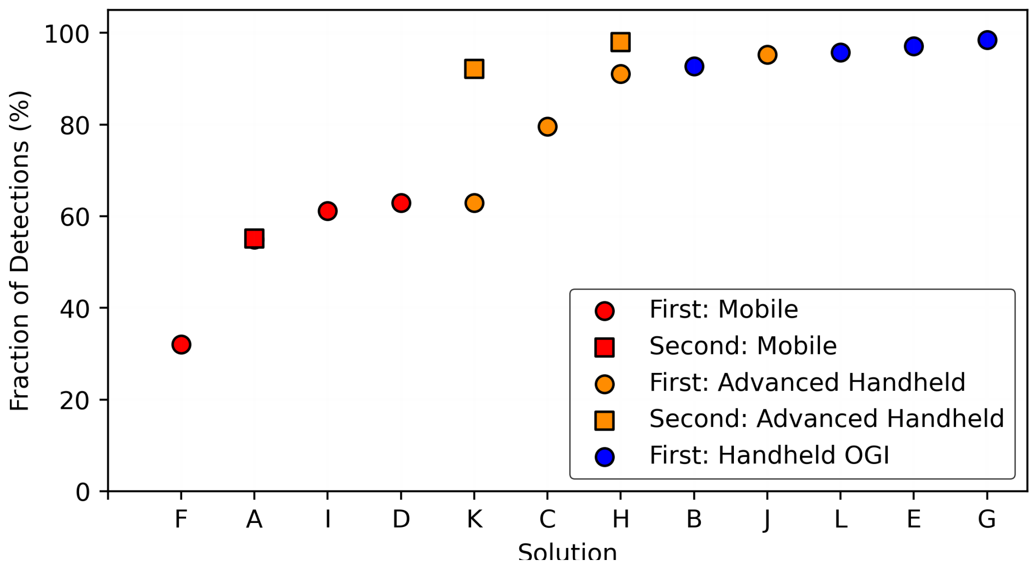

Figure 4: A scatterplot of equipment unit-level localization accuracy for all solutions arranged in increasing order. The y-axis shows the fraction of detection reports that identified a leak source at the correct equipment unit. The marker colors in both figures identify handheld OGI, advanced handheld, and mobile solutions. The circular-shaped marker represents each solution's first-round test results. In contrast, the square-shaped markers identify results from their second round of testing (for solutions tested multiple times).

Figure 4 and Table 2 shows the equipment unit-level localization accuracy and precision of all tested solutions from both the first round and second round of testing. For the first round of testing (circular markers), we find that the equipment unit-level localization accuracies of handheld solutions (i.e., OGI cameras and advanced handheld solutions) were statistically significantly (T and Kolmogorov-Smirnov tests; $p < 0.05$) higher than those of mobile solutions, with the two categories of handheld solutions having comparable localization performance. Mobile solutions attributed emission sources at the equipment group level ($13m^2$ to $88m^2$) more than handheld solutions (SI Table S-8 and Figure S-18). These findings highlight the effectiveness of handheld solutions at pinpointing leak sources for repairs or reporting, even though cost and labor requirements are major concerns when scaled over thousands of O&G facilities. Also, solutions showed a similar trend in localization precision performance as observed for localization accuracy (SI Table S-8 and Figure S-17).

Table 2: Results of the detection classification scheme (FPF and TPFs) and the 90% POD predicted by each solution sorted in order of increasing 90% POD.

Solution		Survey Time†	Localization (Unit)		Detection Classification		
ID	Category	per Unit (minutes)	Accuracy (%)	Precision (%)	FPF (%)	FNF (%)	DL90‡ (kg CH_4/h)
Solutions' results from the first round of testing.							
G	Handheld OGI	5.86 [1.11, 15.0]	98.5	98.5	0.0	5.7	0.08 [0.00, 0.10]
H	Advanced Handheld	5.09 [0.78, 14.0]	91.0	100.0	9.0	16.5	0.14 [0.11, 0.67]
L	Handheld OGI	2.3 [0.72, 4.86]	95.7	100.0	4.3	8.2	0.15 [0.00, 1.70]
E	Handheld OGI	1.91 [1.0, 6.17]	97.1	98.8	1.7	24.1	0.26 [0.17, 0.39]
C	Advanced Handheld	1.89 [0.94, 4.39]	79.5	97.2	18.2	19.6	0.36 [0.18, 1.05]
J	Advanced Handheld	1.03 [0.39, 2.86]	95.2	100.0	4.8	31.0	0.65 [0.36, 1.60]
B	Handheld OGI	1.66 [0.67, 5.0]	92.8	96.2	3.6	25.2	0.81 [0.16, NA]
K*	Advanced Handheld	2.53 [1.14, 5.0]	62.9	97.7	35.6	6.6	0.00 [0.00, NA]
D	Mobile	2.02 [0.94, 8.0]	62.9	87.4	28.0	31.7	NA [0.60, NA]
I	Mobile	1.63 [0.56, 3.57]	61.2	78.8	22.4	22.4	NA [0.78, NA]
A	Mobile	0.94 [0.44, 2.57]	55.0	68.0	19.2	20.3	NA [2.32, NA]
F	Mobile	2.68 [1.56, 8.5]	32.0	64.0	50.0	48.5	NA [4.44, NA]
Results from the second round of testing for solutions that participated in the first round.							
K*	Advanced Handheld	0.51 [0.06, 1.5]	92.2	92.2	0.0	7.3	0.05 [0.00, 0.27]
H	Advanced Handheld	2.34 [0.72, 6.67]	97.9	100.0	2.1	8.7	0.11 [0.00, 0.24]
A	Mobile	1.73 [0.5, 3.67]	55.1	71.7	23.1	45.5	NA [6.98, NA]

† 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 ×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%.

One explanation for the localization performance of mobile solutions is that their data analytics typically requires favorable wind conditions - accessed directly downwind, steady wind direction and speed, well understood stability class - and an accurate estimate of the emission location. Understanding of wind would be improved by averaging over enough time to extract both mean and variability in wind direction, and to estimate stability parameters. However, qualitative observation of solution methods indicate that during this study, most methods did not measure long enough to extract average wind conditions from data, as the selling point of the method is providing faster screening of facilities.

Although some mobile solutions in this study presented promising equipment unit-level localization performance (accuracy and precision $> 50\%$ and 60%, respectively), with additional development, mobile solutions may currently be best deployed to rapidly screen for general emissions locations, followed by investigations with handheld solutions (e.g., OGI camera surveys) to identify emitters for repairs and reporting. In some field applications, a high spatial resolution of detected sources might be less of a priority than the ability to screen a large area at a time with less labor compared to surveying with handheld solutions.[12] For regulatory compliance, the EPA's final rule on advanced methane detection technologies has stipulated secondary inspection work practices at O&G facilities based on the spatial resolution of emission sources by LDAQ solutions.[13]

To characterize the time expended by solutions to complete emissions survey during each experiment, reported survey times by solutions were normalized by the number of equipment units (wellheads, separators, and tanks) in the *survey areas* (which are unequal in size) defined by the study team during experiments, as summarized in Table 2 and Figure 5. Each violin plot in Figure 5 represents the distribution of the survey times per equipment unit for each solution. While the face color of each violin indicate if a solution is categorized as mobile, handheld OGI, or advanced handheld, the edge color shows if the result is from the first or second round of testing of the solution. From Table 2 and Figure 5, results show that the survey times per unit for mobile solutions as a group were statistically significantly lower

(Mann-Whitney U and Kolmogorov-Smirnov tests; $p < 0.05$) lower than those of handheld OGI and advanced handheld solutions, respectively, even though there were no statistically significant difference between the 2 categories of handheld solutions. Faster survey times, as suggested by Fox et al., have important cost implications during field deployments; the longer it takes to complete surveys, the more likely it is to increase an operator's emissions monitoring expenses when scaled over several O&G facilities.

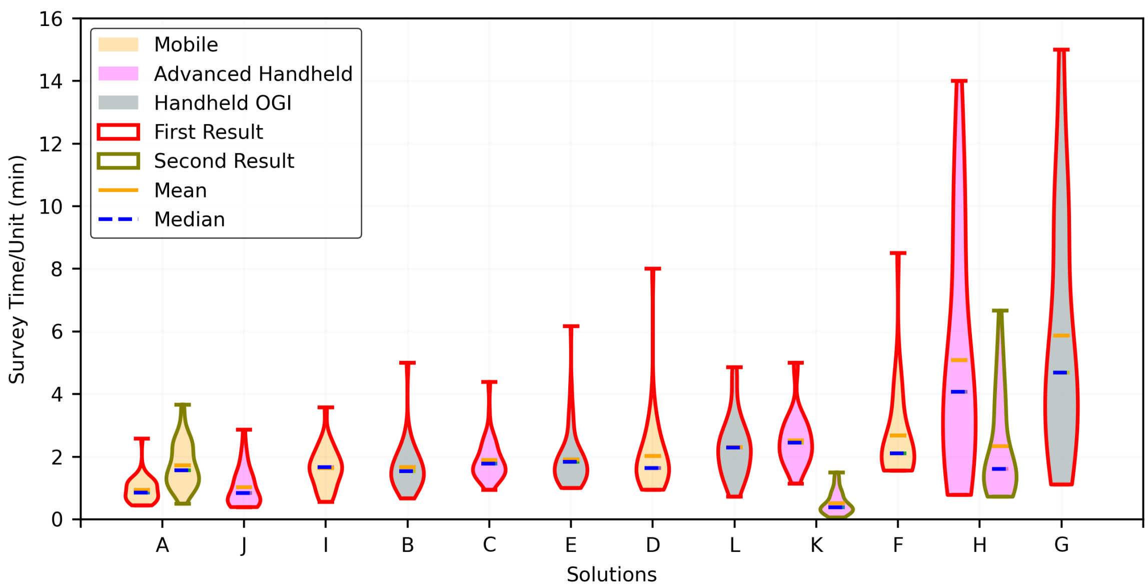

Figure 5: A violin plot showing the distribution of survey time per equipment unit arranged in increasing order of solutions' first-round testing mean survey times. The survey time per equipment unit was evaluated by dividing the survey time by the number of equipment units in the survey areas (as defined by the test center) within which an experiment was conducted. The face color of each violin indicates if a solution is a mobile (i.e., vehicle- or drone-based), 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. The results of solutions that have been tested twice (A, H, and K) are also compared; second test results are differentiated from the first round test results by the edge color of each violin plot. The violins for the first round results have red edge colors while those for the second round have a dark slate gray color.

While mobile solutions had quicker survey times per equipment unit as a group compared to handheld solutions, the reverse was true for a few individual solutions. We hypothesize this was due to a few reasons. First, the noticeable odor and the audible sound made by gas releases during this study could have unblinded some controlled releases in some cases (e.g., surveyors directly downwind of an emission source) and likely aided surveyors with handheld technologies in narrowing down leak locations quicker than expected. Although this survey behavior is also observed in real facilities, the test facility in this study (METEC) is not characterized by O&G facility sensory distractions, which could make emissions survey more challenging. Real O&G facilities are typically characterized by operational noise (e.g., the movement of trucks or other vibrational noise sources) or background emissions due to routine venting events (i.e., from gas-driven pneumatic controllers and pumps, compressor seals, etc.), which can produce misleading auditory and olfactory signals to surveyors, thereby increasing the survey times with these solutions in the field. Second, there were cases where surveys with drone-based solutions were suspended midway through flights to replace flat batteries or to fix impromptu operational challenges like poor wireless communication between the drone system and the controller. It was, however, difficult to assess the extent to which these breaks between flights (if at all) increased the survey times due to unavailable data.

Results also showed that for the range of survey times per equipment unit obtained in this study, the TPFs evaluated per experiment did not statistically significantly correlate (SI section S-7.4) with the associated survey times per equipment unit (except for solution I). This indicates that high TP detection performance is possible at relatively shorter survey times if solutions' survey methodologies are optimized, weather and environmental conditions are favorable (i.e., wind, road accessibility, etc.), and solutions are at high readiness level for field deployment. In general, this study's detection, localization, and survey times data indicate that tested mobile and handheld solutions likely should have different value propositions: perform quicker surveys and narrow down emission sources with mobile solutions to facilitate closer inspections and measurement with handheld solutions - which showed better detection and localization performance.

How Solutions' Performance Changed with Repeat Testing

Our data showed that solutions' performance tends to improve with repeat testing. Three solutions (A, H, and K) participated in the second round of testing 3 to 12 months after they tested for the first time at the same facility with the same protocol. Our data indicate that the performance of solutions H and K generally improved across all metrics (i.e., DL90, FPF, FNF, equipment unit level localization accuracy and precision, and survey time per equipment unit) except localization precision for solution K. In contrast, the performance of solution A declined across all metrics except for localization accuracy and precision. These improvements are unsurprising as there has been rapid and continued development of next-generation emissions detection technologies in terms of sensor type/quality, hardware, data analytics/algorithms, and survey methodology given the growing interest in advanced methane detection technologies by operators and regulators.[13,43] For example, solution H maintained 100% localization precision at the equipment unit level (as with the first round test result) while also increasing localization accuracy to 98%, suggesting likely near-maturity of the solution's localization algorithm.

Other possible reasons for the improvement (or lack thereof) in the performance of solutions include relatively favorable (or unfavorable) test and measurement conditions. For example, Solutions A and K were retested in December 2022 (SI section S-2) under temperature conditions between 3°C and 11°C compared to 8°C and 22°C in 2021 during the first round of testing. Surveyors with solution A (i.e., automobile-based) are less likely to be adversely influenced by relatively cold weather, unlike solution K (i.e., handheld), since the latter's operationalization involves direct inspection of suspected leaking components. This hypothesis is further supported by solution K recording the lowest survey times per equipment unit obtained throughout the study during the second round of testing. Secondly, some solutions might have improved their survey methodology or the implementation process. For example, solution K's walking procedure in the last test was a modified version of the one they used in the first round of testing (see reports attached with the SI).

Sensitivity of Solution's Detection Performance to Testing Complexity

Study data showed that the detection capacity of mobile solutions was generally more sensitive to changes in testing complexity than other solution categories. Recall that experiments conducted during this study had 1 to 6 simultaneous controlled releases from different emissions sources of different heights. Results show that mobile solutions had more difficulty isolating and detecting individual emissions during experiments with multiple controlled releases than other categories of solutions. More mobile solutions (2 of 4) had statistically significant (Rank Biseral Correlation; $p < 0.05$) lower TPF per experiment for experiments with multiple controlled releases compared to experiments with 1 controlled release (SI section S-7.5). Additionally, there were statistically significant differences between the TPFs of mobile solutions and handheld (i.e., OGI and advanced) solutions for all experiments with multiple controlled releases, unlike when experiments with 1 controlled release were considered. These results highlight a key issue generally encountered by solutions that rely on wind transport to make detections (i.e., mobile solutions and continuous monitors): the struggle by solutions' data analytics and algorithms to robustly isolate and differentiate individual emission sources during experiments with multiple, simultaneous controlled releases.[32,42,43]

When controlled releases for solutions were aggregated based on emission point heights by generally categorizing them as tank (height > 12 ft) and non-tank (< 12 ft) sources, our data showed that no group of solutions showed a statistically significant difference in the TPFs between the source categories. However, handheld solutions showed statistically significantly higher TPF for tank sources than mobile solutions (2 sample t-tests; $p < 0.05$), which is unsurprising given that handheld solutions involved close component-level emissions inspections, unlike mobile solutions, which were typically positioned farther away from the emission source (e.g., the edge of a well pad). Note that in several field cases, direct, proximity surveys of emitting sources can be challenging due to inaccessibility, high-temperature surfaces, or moving parts; hence, such scenarios might require alternative emissions measurement methods such as mobile solutions. Therefore, users must understand the limitations of solutions, consider if a solution is fit for purpose for an application during field deployment (i.e., facility-level emissions screening, component-level inspection, etc.), and assess how the trade-offs in the performance of the different categories of solutions can be leveraged for effective emissions mitigation.

Implications

Recent rules and regulatory policies, such as the US EPA final NSPS rules on alternative methane detection approaches[13] and the Inflation Reduction Act Waste Emission Charge[55] aimed at further reduction in methane emissions across the US O&G supply chain, have spurred increased attention to advanced emissions detection solutions. With standardized controlled testing identified as one of the steps involved in demonstrating emissions mitigation equivalence between these solutions and traditional LDAR approaches,[29,31] this study, unlike prior works,[31,32,37] represents a significant step forward in the controlled testing of next-generation survey solutions using a consensus test protocol.[38]

The study results have shown the following: First, while handheld OGI camera solutions had comparable or better performance than other categories of solutions across most performance metrics, no solution or category of solutions achieved all the desired performance levels across all metrics assessed. This implies that no category of solutions tested is superior as all solutions can systematically detect emissions successfully (atleast significantly better than random guessing as shown in SI section S-7.5), albeit with varying strengths and limitations hence the need for different value propositions for mobile solutions and handheld solutions.

Secondly, solutions' performance tends to improve with repeat testing, which highlights the importance of regular, comprehensive testing to the development of solutions, as shown in a prior similar study.[43] Regular and robust controlled testing provide quality ground truth data used by solutions to optimize sensor performance and algorithms, and evaluate survey/deployment methodology and the implementation processes for easier interpretation in the field and improved detection performance

Finally, isolating and differentiating individual sources from multiple, simultaneous emitters is still a persistent challenge for the data analytics/algorithm of solutions that generally rely on wind transport for detection (i.e., mobile solutions, continuous monitors, etc.). However, given the rapid development of advanced methane detection technologies, there is potential that this challenge could be addressed significantly in the near future.

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 (unless detection performance depends on high thermal contrast between the equipment surface and the gas plume). Second, as shown by Day et al. for continuous monitors, field testing performance can vary substantially from that of controlled testing; therefore, robust field testing of the solutions is needed to validate and build confidence in assessed controlled testing performances.

Supporting Information

Zip folder of solutions' performance reports (PDF), data tables (XLSX), and data tables guide (XLSX). Detailed description of the test facility, solutions deployment, additional results, guide to the performance reports, and bootstrapping methodology (PDF).

Acknowledgment

This work was funded by the Department of Energy (DOE): DE-FE0031873 which supported the test protocol development and subsidized the cost of testing for participating solutions at METEC. Industry partners and associations provided matching funds for the DOE contract. The authors acknowledge that the solution vendors paid to participate in the study and we also acknowledge their efforts and commitment in kind to the test program. The authors acknowledge the contributions of the protocol development committee to the development of the test protocol.The authors appreciates Clay Bell for supporting the study's conceptual design and execution. The authors also appreciates Ezra Levin and Anna Hodshire for helping with the Manuscript review process.

Author Information

References

(1) Intergovernmental Panel On Climate Change (Ipcc), Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1st ed.; Cambridge University Press, 2023

.

(2) Methane and Climate Change – Global Methane Tracker 2022 – Analysis. https://www.iea.org/reports/global-methane-tracker-2022/methane-and-climate-change, (Accessed July 20, 2023)

.

(3) Greenhouse Gases Continued to Increase Rapidly in 2022. https://www.noaa.gov/news-release/greenhouse-gases-continued-to-increase-rapidly-in-2022, (Accessed September 3, 2023)

.

(4) Agency, U. S. E. P. Overview of Greenhouse Gases. https://www.epa.gov/ghgemissions/overview-greenhouse-gases, (Accessed July 20, 2023)

.

(5) Global Methane Tracker 2023 – Analysis. 2023; https://www.iea.org/reports/global-methane-tracker-2023, (Accessed March 31, 2024)

.

(6) Agency, U. S. E. P. Leak Detection and Repair: A Best Practices Guide. https://www.epa.gov/compliance/leak-detection-and-repair-best-practices-guide, (Accessed July 20, 2023)

.

(7) Canadian Environment Protection Act, Consolidated federal laws of Canada, Regulations Respecting Reduction in the Release of Methane and Certain Volatile Organic Compounds (Upstream Oil and Gas Sector). **2018**,

(8) METHOD 21 - DETERMINATION OF VOLATILE ORGANIC COMPOUND LEAKS. https://www.epa.gov/sites/default/files/2017-08/documents/method_21.pdf, (Accessed June 26, 2024)

.

(9) Alternative Work Practice To Detect Leaks From Equipment. https://www.federalregister.gov/documents/2008/12/22/E8-30196/alternative-work-practice-to-detect-leaks-from-equipment, (Accessed June 26, 2024)

.

(10) Ravikumar, A. P. Impact of survey frequency on emissions mitigation at oil and gas sites. https://www.edf.org/sites/default/files/content/Appendix_D_Ravikumar_and_Lyon_Impact_of_Sur (Accessed October 26, 2024)

.

(11) Zimmerle, D.; Vaughn, T.; Bell, C.; Bennett, K.; Deshmukh, P.; Thoma, E. Detection

Limits of Optical Gas Imaging for Natural Gas Leak Detection in Realistic Controlled Conditions. Environmental Science & Technology **2020**, 54, 11506–11514
.

(12) Fox, T. A.; Barchyn, T. E.; Risk, D.; Ravikumar, A. P.; Hugenholtz, C. H. A review of close-range and screening technologies for mitigating fugitive methane emissions in upstream oil and gas. Environmental Research Letters **2019**, 14, 053002
.

(13) Standards of Performance for New, Reconstructed, and Modified Sources and Emissions Guidelines for Existing Sources: Oil and Natural Gas Sector Climate Review. https://www.federalregister.gov/documents/2024/03/08/2024-00366/standards-of-performance-for-new-reconstructed-and-modified-sources-and-emissions-guidelines-for, (Accessed April 17, 2024)
.

(14) Wilde, S.; Tyner, D.; Conrad, B.; Johnson, M. A Comparison of the Effectiveness of Regulated OGI Leak Detection and Repair (LDAR) Surveys and Aerial Measurements in the Real-World. https://meetingorganizer.copernicus.org/EGU24/EGU24-6658.html
.

(15) Schwietzke, S.; Harrison, M.; Lauderdale, T.; Branson, K.; Conley, S.; George, F. C.; Jordan, D.; Jersey, G. R.; Zhang, C.; Mairs, H. L.; Pétron, G.; Schnell, R. C. Aerially Guided Leak Detection and Repair: A Pilot Field Study for Evaluating the Potential of Methane Emission Detection and Cost-Effectiveness. 69, 71–88
.

(16) Kemp, C. E.; Ravikumar, A. P. New Technologies Can Cost Effectively Reduce Oil

and Gas Methane Emissions, but Policies Will Require Careful Design to Establish Mitigation Equivalence. 55, 9140–9149, DOI: 10.1021/acs.est.1c03071

.

(17) Lavoie, T. N.; Shepson, P. B.; Cambaliza, M. O. L.; Stirm, B. H.; Conley, S.; Mehrotra, S.; Faloona, I. C.; Lyon, D. Spatiotemporal Variability of Methane Emissions at Oil and Natural Gas Operations in the Eagle Ford Basin. Environmental Science & Technology **2017**, 51, 8001–8009

.

(18) Vaughn, T. L.; Bell, C. S.; Pickering, C. K.; Schwietzke, S.; Heath, G. A.; Pétron, G.; Zimmerle, D. J.; Schnell, R. C.; Nummedal, D. 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 **2018**, 115, 11712–11717

.

(19) Robertson, A. M.; Edie, R.; Snare, D.; Soltis, J.; Field, R. A.; Burkhart, M. D.; Bell, C. S.; Zimmerle, D.; Murphy, S. M. Variation in Methane Emission Rates from Well Pads in Four Oil and Gas Basins with Contrasting Production Volumes and Compositions. Environmental Science & Technology **2017**, 51, 8832–8840

.

(20) Heimburger, A. M. F. et al. Assessing the optimized precision of the aircraft mass balance method for measurement of urban greenhouse gas emission rates through averaging. Elementa: Science of the Anthropocene **2017**, 5, 26

.

(21) Zavala-Araiza, D. et al. Reconciling divergent estimates of oil and gas methane emissions. Proceedings of the National Academy of Sciences **2015**, 112, 15597–15602

.

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

.

(23) Johnson, D.; Heltzel, R. On the Long-Term Temporal Variations in Methane Emissions from an Unconventional Natural Gas Well Site. ACS Omega **2021**, 6, 14200–14207

.

(24) Brandt, A. R. et al. Methane Leaks from North American Natural Gas Systems. Science **2014**, 343, 733–735

.

(25) Zavala-Araiza, D.; Lyon, D.; Alvarez, R. A.; Palacios, V.; Harriss, R.; Lan, X.; Talbot, R.; Hamburg, S. P. Toward a Functional Definition of Methane Super-Emitters: Application to Natural Gas Production Sites. Environmental Science & Technology **2015**, 49, 8167–8174

.

(26) Alvarez, R. A. et al. Assessment of methane emissions from the U.S. oil and gas supply chain. Science **2018**, eaar7204

.

(27) Schwietzke, S. et al. Improved Mechanistic Understanding of Natural Gas Methane Emissions from Spatially Resolved Aircraft Measurements. 51, 7286–7294, DOI: 10.1021/acs.est.7b01810

.

(28) Cusworth, D. H.; Duren, R. M.; Thorpe, A. K.; Olson-Duvall, W.; Heckler, J.; Chapman, J. W.; Eastwood, M. L.; Helmlinger, M. C.; Green, R. O.; Asner, G. P.; Dennison, P. E.; Miller, C. E. Intermittency of Large Methane Emitters in the Permian Basin. Environmental Science & Technology Letters **2021**, 8, 567–573
.

(29) Fox, T. A.; Ravikumar, A. P.; Hugenholtz, C. H.; Zimmerle, D.; Barchyn, T. E.; Johnson, M. R.; Lyon, D.; Taylor, T. A Methane Emissions Reduction Equivalence Framework for Alternative Leak Detection and Repair Programs. 7, DOI: 10.1525/elementa.369
.

(30) Thorpe, A. K. et al. Mapping Methane Concentrations from a Controlled Release Experiment Using the next Generation Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG). 179, 104–115
.

(31) Ravikumar, A. P.; Sreedhara, S.; Wang, J.; Englander, J.; Roda-Stuart, D.; Bell, C.; Zimmerle, D.; Lyon, D.; Mogstad, I.; Ratner, B.; Brandt, A. R. Single-Blind Inter-Comparison of Methane Detection Technologies – Results from the Stanford/EDF Mobile Monitoring Challenge. 7, 37, DOI: 10.1525/elementa.373
.

(32) Bell, C. S.; Vaughn, T.; Zimmerle, D. Evaluation of next generation emission measurement technologies under repeatable test protocols. Elementa: Science of the Anthropocene **2020**, 8, 32
.

(33) Sherwin, E. D.; Chen, Y.; Ravikumar, A. P.; Brandt, A. R. Single-Blind Test of

Airplane-Based Hyperspectral Methane Detection via Controlled Releases. 9, 00063, DOI: 10.1525/elementa.2021.00063

.

(34) Johnson, M. R.; Tyner, D. R.; Szekeres, A. J. Blinded Evaluation of Airborne Methane Source Detection Using Bridger Photonics LiDAR. 259, 112418, DOI: 10.1016/j.rse.2021.112418

.

(35) Morales, R.; Ravelid, J.; Vinkovic, K.; Korbeń, P.; Tuzson, B.; Emmenegger, L.; Chen, H.; Schmidt, M.; Humbel, S.; Brunner, D. Controlled-Release Experiment to Investigate Uncertainties in UAV-based Emission Quantification for Methane Point Sources. 15, 2177–2198

.

(36) ARPA-E's MONITOR Program | Methane Observation Networks with Innovative Technology to Obtain Reductions. http://arpa-e.energy.gov/technologies/programs/monitor, (Accessed July 22, 2023)

.

(37) Liu, Y. et al. Assessment of current methane emission quantification techniques for natural gas midstream applications. Atmospheric Measurement Techniques **2024**, 17, 1633–1649

.

(38) Bell, C.; Zimmerle, D. METEC controlled test protocol: Continuous monitoring emission detection and quantification. **2022**, Publisher: Mountain Scholar tex.copyright: Copyright and other restrictions may apply. User is responsible for compliance with all applicable laws. For information about copyright law, please see https://libguides.colostate.edu/copyright.

(39) Bell, C.; Rutherford, J.; Brandt, A.; Sherwin, E.; Vaughn, T.; Zimmerle, D. Single-Blind Determination of Methane Detection Limits and Quantification Accuracy Using Aircraft-Based LiDAR. 10, DOI: 10.1525/elementa.2022.00080
.

(40) Abbadi, S. H. E.; Chen, Z.; Burdeau, P. M.; Rutherford, J. S.; Chen, Y.; Zhang, Z.; Sherwin, E. D.; Brandt, A. R. Comprehensive Evaluation of Aircraft-Based Methane Sensing for Greenhouse Gas Mitigation.

(41) Rutherford, J. S.; Sherwin, E. D.; Chen, Y.; Aminfard, S.; Brandt, A. R. Evaluating Methane Emission Quantification Performance and Uncertainty of Aerial Technologies via High-Volume Single-Blind Controlled Releases.

(42) Bell, C.; Ilonze, C.; Duggan, A.; Zimmerle, D. Performance of Continuous Emission Monitoring Solutions under a Single-Blind Controlled Testing Protocol. Environmental Science & Technology **2023**, 57, 5794–5805
.

(43) Ilonze, C.; Emerson, E.; Duggan, A.; Zimmerle, D. Assessing the Progress of the Performance of Continuous Monitoring Solutions under a Single-Blind Controlled Testing Protocol.

(44) Bell, C.; Zimmerle, D. METEC controlled test protocol: survey emission detection and quantification. **2022**,

(45) Bell, C. S.; Vaughn, T. L.; Zimmerle, D.; Herndon, S. C.; Yacovitch, T. I.; Heath, G. A.; Pétron, G.; Edie, R.; Field, R. A.; Murphy, S. M.; Robertson, A. M.; Soltis, J. Comparison of methane emission estimates from multiple measurement techniques at natural gas production pads. Elementa: Science of the Anthropocene **2017**, 5, 79
.

(46) Allen, D. T.; Torres, V. M.; Thomas, J.; Sullivan, D. W.; Harrison, M.; Hendler, A.; Herndon, S. C.; Kolb, C. E.; Fraser, M. P.; Hill, A. D.; Lamb, B. K.; Miskimins, J.; Sawyer, R. F.; Seinfeld, J. H. Measurements of Methane Emissions at Natural Gas Production Sites in the United States. 110, 17768–17773, DOI: 10.1073/pnas.1304880110
.

(47) Pacsi, A.; Ferrara, T.; Schwan, K.; Tupper, P.; Lev-On, M.; Smith, R.; Ritter, K. Equipment Leak Detection and Quantification at 67 Oil and Gas Sites in the Western United States. 7, 29
.

(48) Brown, J.; Rufael, T.; Harrison, M.; Roman-White, S.; Ross, G.; George, F.; Zimmerle, D. Informing methane emissions inventories using facility aerial measurements at midstream natural gas facilities; Preprint, 2023
.

(49) Kunkel, W. M.; Carre-Burritt, A. E.; Aivazian, G. S.; Snow, N. C.; Harris, J. T.; Mueller, T. S.; Roos, P. A.; Thorpe, M. J. Extension of Methane Emission Rate Distribution for Permian Basin Oil and Gas Production Infrastructure by Aerial LiDAR. 57, 12234–12241
.

(50) Ilonze, C.; Wang, J. L.; Ravikumar, A. P.; Zimmerle, D. Methane Quantification Performance of the Quantitative Optical Gas Imaging (QOGI) System Using Single-Blind Controlled Release Assessment. 24, 4044
.

(51) Day, R.; Emerson, E.; Bell, C.; Zimmerle, D. Point Sensor Networks Struggle to Detect and Quantify Short Controlled Releases at Oil and Gas Sites. 24, 2419
.

(52) EPA, 40 CFR Part 60 Subpart OOOOb – Standards of Performance for Crude Oil and Natural Gas Facilities for Which Construction, Modification or Reconstruction Commenced After December 6, 2022. **2024**, Accessed: 2024-12-16. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-60/subpart-OOOOb

.

(53) von Fischer, J. C.; Cooley, D.; Chamberlain, S.; Gaylord, A.; Griebenow, C. J.; Hamburg, S. P.; Salo, J.; Schumacher, R.; Theobald, D.; Ham, J. Rapid, Vehicle-Based Identification of Location and Magnitude of Urban Natural Gas Pipeline Leaks. Environmental Science & Technology **2017**, 51, 4091–4099, DOI: 10.1021/acs.est.6b06095

.

(54) Ravikumar, A. P.; Wang, J.; Brandt, A. R. Are Optical Gas Imaging Technologies Effective For Methane Leak Detection? 51, 718–724, DOI: 10.1021/acs.est.6b03906

.

(55) United States Environmental Protection Agency: Waste Emissions Charge. https://www.epa.gov/inflation-reduction-act/waste-emissions-charge, (Accessed August 18, 2024)

.

For Table of Contents Only