

Assessing the Progress of the Performance of Continuous Monitoring Solutions under a Single-Blind Controlled Testing Protocol

Chiemezie Ilonze, Ethan Emerson, Aidan Duggan, and Daniel Zimmerle*



Cite This: *Environ. Sci. Technol.* 2024, 58, 10941–10955



Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The recent regulatory spotlight on continuous monitoring (CM) solutions and the rapid development of CM solutions have demanded the characterization of solution performance through regular, rigorous testing using consensus test protocols. This study is the second known implementation of such a protocol involving single-blind controlled testing of 9 CM solutions. Controlled releases of rates (6–7100 g) CH₄/h over durations (0.4–10.2 h) under a wind speed range of (0.7–9.9 m/s) were conducted for 11 weeks. Results showed that 4 solutions achieved method detection limits (DL90s) within the tested emission rate range, with all 4 solutions having both the lowest DL90s (3.9 [3.0, 5.5] kg CH₄/h to 6.2 [3.7, 16.7] kg CH₄/h) and false positive rates (6.9–13.2%), indicating efforts at balancing low sensitivity with a low false positive rate. These results are likely best-case scenario estimates since the test center represents a near-ideal upstream field natural gas operation condition. Quantification results showed wide individual estimate uncertainties, with emissions underestimation and overestimation by factors up to >14 and 42, respectively. Three solutions had >80% of their estimates within a quantification factor of 3 for controlled releases in the ranges of [0.1–1] kg CH₄/h and > 1 kg CH₄/h. Relative to the study by Bell et al., current solutions performance, as a group, generally improved, primarily due to solutions from the study by Bell et al. that were retested. This result highlights the importance of regular quality testing to the advancement of CM solutions for effective emissions mitigation.

KEYWORDS: methane, emissions mitigation, detection limit, emissions quantification, source attribution, natural gas



INTRODUCTION

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 concentrations more than twice preindustrial levels.^{1–3} As the major component of natural gas commonly emitted across production, processing, and distribution sectors, mitigating natural gas emissions has economic, safety, and environmental benefits.^{4,5} The oil and gas (O&G) sector is the largest industrial source ($\approx 30\%$) of anthropogenic methane emissions in the United States. Several studies have shown that fugitive (unplanned) methane emissions are stochastic, temporally and spatially variable with large emitters typically responsible for a substantial portion of unplanned emissions.^{6–14} Continuous monitoring (CM) can improve emissions detection since these solutions near-continuously monitor entire facilities (e.g., an entire wellpad) and can identify fugitive emissions faster than existing survey methods (e.g., optical gas imaging camera surveys).^{15,16}

A CM leak detection and quantification (LDAQ) solution is a technology that measures ambient emissions concentration

continuously using one, or a combination of, sensing methodologies (e.g., tunable diode laser absorption spectroscopy, light detection and ranging, etc.) and interprets readings using proprietary algorithms to generate actionable results (e.g., using a gas plume image to estimate the location and size of an emitter).¹⁷ Recently, the United States Environment Protection Agency (USEPA) proposed new pathways for CM solutions to be utilized for regulatory-compliant leak detection and repair (LDAR) programs.¹⁸ Studies have shown that large emission events, including super emitters (large, episodic emissions ≥ 100 kg CH₄/h),^{19–21} contribute to the observed gap between direct emission measurements and the USEPA Greenhouse Gas Reporting Program (GHGRP) estimates^{22–25} and other

Received: October 13, 2023

Revised: May 22, 2024

Accepted: May 31, 2024

Published: June 12, 2024



Table 1. Characteristics of Participating Solutions

ID	sensor		reported data ^a			
	type	count 2022	count 2023	detection	quantification	GPS localization ^b
Participated in Bell et al.						
A ^c	point sensor network	8	8	✓	✓	✓
B	scanning/imaging	1	1	✓	✓	✓
D	point sensor network	8	8	✓	✓	✓
F	point sensor network	8	10	✓	✓	X
Did not Participate in Bell et al.						
L	scanning/imaging		1	✓	✓	✓
N	point sensor network		18	✓	✓	✓
O	scanning/imaging		1	✓	✓	X
P	point sensor network		6	✓	X	X
Q	point sensor network		13	✓	X	X

^a✓ indicates the parameter of interest was reported by the solution. 'X' indicates that it was not reported. ^bIndicates if the solution localized emitters by GPS coordinates. ^cOne of the sensors installed failed during the study.

reporting programs.^{13,14} USEPA has proposed amendments of the Subpart W of the GHGRP²⁶ and the Super Emitter Response Program²⁷ to close the data gap using selected top down approaches (satellite, aerial, etc.),^{28–32} among other methods for measurements. Surveys using top down approaches are typically brief (seconds to minutes), and performance depends on the time of the day and the prevailing meteorological conditions: clear skies for satellites or a specified range of atmospheric stability conditions for aerial surveys. CM solutions can provide time-resolved monitoring across a wider, but not unlimited, range of meteorological conditions to promptly alert operators when facility emissions begin to rise to abnormal levels.

To characterize detection efficacy, CM solutions must be tested to understand the probability of detection (POD), quantification accuracy and associated uncertainties, emission source localization, time to detection, operational downtime, and false positive (FP) and false negative (FN) rates. CM solutions consist of three components: sensing, deployment on a facility, and proprietary algorithms that process sensed data. These three components cannot be tested independently. Therefore, testing must assess the performance of a CM solution as an integrated system that includes sensors, data acquisition and communication, proprietary algorithms, hardware, and mode of installation on the facility. The goal of testing is to ascertain the performance level of CM solutions as deployed, with specific interest in the functionalities highlighted earlier (detection, etc.), each of which can affect the detection or quantification efficacy of CM solutions. Therefore, clear testing using consensus, technology-neutral protocols is necessary to compare the performance of CM solutions.

Past studies employed study-specific protocols for testing,³³ which are generally difficult to repeat, making it difficult to compare solution performance from multiple test programs. Additionally, previous evaluations of CM solutions encountered limitations related to testing complexity and prevailing meteorological/environmental conditions.^{34–37} Partly in response to these results, a consensus protocol was developed by the advancing development of emissions detection (ADED) project³⁸ and was used by Bell and Zimmerle for the first peer-consensus CM testing with a standardized protocol. The study result showed high variability between solutions, high uncertainty, and some bias in most assessed metrics across all of the CM solutions tested.

The study presented here represents the second implementation of the ADED protocol³⁸ by testing 9 CM solutions, including 4 that also participated in the prior study.³⁹ CM solutions were tested for 11 weeks between February and April, 2023, at the Methane Emissions Technology Evaluation Center (METEC), Colorado State University (CSU), Colorado, USA. This study also divides CM solutions into the same two classes utilized in the prior study: (a) *point sensor network*—solutions that deploy multiple point sensors that sense hydrocarbons and use proprietary algorithms to combine meteorological and concentration readings to infer detections, etc. (b) *Scanning/imaging*—solutions, which use scanning lasers or short/midwave infrared cameras to visualize gas plumes, which are then combined with meteorological data to infer detections, etc. The protocol specifies both testing methods and how performance metrics are calculated. By using the same primary metrics for evaluation, results from the current study can be compared with those from the prior study³⁹ and related studies⁴⁰ to determine if solutions have progressed between test programs.

METHODOLOGY

Test Facility. Testing was conducted between February 8th and April 28th, 2023, at METEC; an 8-acre (3.2 ha) outdoor controlled testing facility primarily designed to simulate methane emissions from North American onshore equipment in a controlled manner. METEC is furnished with inactive surface equipment units (e.g., wellheads, separators, etc.) intentionally fitted with leak points concealed at commonly observed sources, such as valve packing, flanges, and fittings. Units are arranged into 5 wellpads (pads 1–5) of varying size, complexity, and equipment unit layouts. Testing was conducted exclusively on pads 4 and 5 covering ≈ 8450 m², and was made up of 7 separators, 3 condensate tanks, 8 wellheads, and 2 flares (see Zimmerle et al. and [Supporting Information](#), Sections S-1 and S-2). [Table 1](#) includes a brief summary of the *equipment units* and *equipment groups* in pads 4/5 and how their tags are interpreted. This study utilized 53 unique emission points on pads 4/5, each of them used more than once during testing. Over the duration of the study, $\approx 80\%$ of emission points were ≤ 2 m in height, with the rest between 2 and 6 m (see [Supporting Information](#), Figures S-7 and S-7, for the distribution of the heights of emission points used in this study).

Testing Process. The ADED protocol was developed with contributions from multiple stakeholders, including O&G industry players, academic institutions, LDAQ solution

developers, environmental nongovernmental organizations, and regulatory agencies (state and federal).³⁸ The protocol is designed to test an integrated CM solution and does not test individual subsystems, e.g., sensing performance, optimal deployment, and/or algorithm/analytic capability. In all solutions tested here, point or imaging sensors collect raw sensor readings, which are processed by proprietary algorithms to infer actionable data, including the presence and absence of emitters (detections), emission rate estimates, and emitter locations.

According to the protocol, testing involves a series of *experiments* conducted 24 h per day, every day, for an extended period (weeks or months). Each *experiment* consists of 1 or up to 5 simultaneous controlled releases of gas, each emitting at a steady emission rate for a specified duration (hours). Experiments with multiple controlled release points (>1) evaluated solutions' ability to characterize each emitter. Successive experiments were separated from one another by a break period (hours), during which there was no controlled release, signaling solutions of a return to background atmospheric concentration levels. Experiments were designed with the intention to sweep the range of test (e.g., emission rate, release duration, etc.) and meteorological (e.g., wind speed, temperature, etc.) conditions needed to characterize the *POD* curves of solutions tested. The entire test program was single blind—the participating solutions were unaware of the timing, location(s), durations, and emission rates of controlled releases by the test center.

CSU recruited participating CM solutions through an open invitation advertised on METEC's Web site and also leveraged on contacts gathered during the development of the protocol to directly contact CM LDAQ solution developers. The study team required vendors of solutions participating in the testing to install their systems at least 3 days before the start of testing to participate in the mock testing by METEC and to allow both METEC and CM solution vendors to troubleshoot their respective setups. A portable compressed natural gas (CNG) trailer was connected to METEC's gas supply system to support large and long duration controlled releases. In between refills of the CNG trailer, the study team conducted controlled releases from the storage gas cylinders. All controlled gas releases during testing were CNG, with a mean gas composition by volume of 84.8% methane, 13.1% ethane, 1.6% propane, and trace amounts of heavier hydrocarbons and other gases. For each controlled release, METEC logged the timing, location, metered emission rate, and associated uncertainties, gas composition, and prevailing meteorological conditions, which were time averaged over the release duration.

Testing was conducted day and night across all meteorological conditions that supported the operation of METEC for the entire duration of the study. Exceptions included winter conditions with temperatures below the operating specifications of METEC's thermal flow meters (OMEGA FMA-17xx series). For experiments with 2 or more controlled releases flowing through a flow meter, a precalibration was done before releases officially began to correctly meter and log the rate of each controlled release. The emission rate of controlled releases and experiment duration was selected considering METEC's operational constraints, e.g., available gas supply, emission point orifice size, etc. The study team periodically analyzed the performance of solutions during testing to choose 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, etc.) to

map the *POD* curves of solutions. The resulting range of emission rates and release durations in this testing was 6–7100 g CH₄/h and 0.4–10.2 h, respectively. This implied that the study likely excluded a portion of real-world upstream emissions that are intermittent or of much shorter duration. These include routine emissions from the actuation of pneumatic devices, blowdown events, and routine flash tank emissions, which collectively make up a substantial fraction of methane emissions at typical United States onshore production facilities. Similarly, the study excluded larger releases (≥ 10 kg CH₄/h up to the super emitter rate), which is an important emission source category according to several studies.^{19–21} The study team ensured that no two controlled releases within an experiment flowed through the same *equipment unit* and drastically limited scenarios where 2 consecutive experiments had controlled releases flowing to the same *equipment unit*. This gave CM solutions the best opportunity to isolate and estimate the characteristics of each emitter. This represents a substantial simplification of observed emissions behavior in real O&G facilities where emitters may follow random patterns or emit at variable rates. METEC kept a maintenance record, documenting facility downtime and the timing of faulty experiments and controlled release events noncompliant with the study design (e.g., venting gas supply lines, controlled releases on wellpads not used for the study, etc.).

Performance Metrics. The vendor of each solution sent *detection reports* containing data inferred from sensed emissions (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. In some cases, such human interference delayed detection reporting to the test center by days or weeks; likewise, for 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 noncompliant with the protocol's reporting template.³⁸ This contrasts field deployments where operators bear the burden of inferring web-based dashboards (e.g., interpreting time series methane concentrations/emission rates) of data communicated by the solutions installed in O&G facilities. This detection reporting approach 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 specie measured to infer a detection.

Solution vendors were also allowed to report system *downtime*: periods during testing that solutions were offline (e.g., not taking measurements), which should be ignored by the study team during result analysis. Prior to the performance analysis for each solution, the study team excluded detections (1) reported during METEC maintenance and solution

downtime periods, (2) reports with *EmissionStartTime* 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 FP detections. Similarly, the study team excluded controlled releases (1) conducted during METEC maintenance and solution downtime periods, (2) conducted outside the analysis window of the solution, and (3) with durations shorter than required to obtain a stable flow meter reading. These exclusions were done to avoid spurious FN detections.

All detection reports referencing the same *EmissionSourceID* were grouped together 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*). The study team applied a buffer time of 20 min before and after the timing of each controlled release while matching controlled releases to detection reports. The buffer time accounted for emissions during experiment precalibration periods, and the residual emissions detected by solutions after the end of a controlled release. The matching scheme involved the following steps below:

- The study team sorted all controlled releases by equipment unit identifier, then by emission rate (if reported) in descending order. For each controlled release, all detections identifying the emission source on the same equipment unit as the controlled release were selected. All selected detections with *EmissionStartTime* within the controlled release start and end times (including buffer time) were filtered and sorted by emission rate (if reported) in descending order. The topmost filtered detection report was paired with the controlled release as a correct *equipment unit* level true positive (TP) detection, and the pair was removed from further matching.
- The study team resorted the remaining list of controlled releases from the step above by equipment group identifier, then by emission rate (if reported) in descending order. For each controlled release, all detections identifying the emission source on the same equipment group as the controlled release were selected. All selected detections with *EmissionStartTime* within the controlled release start and end times (including buffer time) were filtered and sorted by emission rate (if reported) in descending order. The topmost filtered detection report was paired with the controlled release as a correct *equipment group* level TP detection, and the pair was removed from further matching.
- The study team resorted the remaining list of controlled releases from the step above by emission rate (if reported) in descending order. For each controlled release, all detection reports with reported *EmissionStartTime* within the controlled release start and end times (including buffer time) were filtered and sorted by emission rate (if reported) in descending order. The topmost filtered detection report was paired with the controlled release as a correct *facility* level TP detection, and the pair was removed from further matching.
- Controlled releases and detection reports remaining after the pairings were identified as FN and FP detections, respectively.

All performance metrics stipulated in the protocol utilized these classification results in their analysis. Key metrics are briefly described below, with full details in the protocol.³⁸

POD: the fraction of binned test conditions (i.e., emission rate, release duration, etc.) classified as TP detections (i.e., $\frac{TPs}{TPs + FNs}$).

Localization Precision—(*Equipment Unit*): the fraction of all TP detections at each detection level (*equipment unit*, *equipment group*, and *facility*).

Localization Accuracy (*Equipment Unit*): the fraction of detection reports (FPs and TPs) at each localization precision level (*equipment unit*). For example, localization accuracy at the equipment group or better is the fraction of all detections localized at both the equipment unit and group levels.

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 evaluated by normalizing the 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 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 with which it is paired with.

Operational factor: the fraction of time a CM solution was operational relative to the total deployment time.

Participating Solutions. All participating CM solutions in this study installed their systems at the test site. Solution vendors decided on the number of sensors, positioning of sensors, and the *equipment groups* 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). All but one solution (N) monitored all *equipment groups* on pads 4 and 5 of METEC. Vendors were requested to install as they would at real facilities. This implied that some vendors installed their solutions either along the fence-line of the pads or around the equipment groups monitored ([Supporting Information](#), Figure S-3). In many field applications, sensor locations are likely restricted to the periphery of the facilities, while the number of sensors installed per facility largely depended on the cost of deployment and the size of the facility. In this study, every solution was responsible for the communication systems to connect their on-site hardware to backend servers and algorithms operating offsite; most solutions utilized cellular data for this purpose. After installation and initial testing, solution personnel left the test center and operated their systems remotely, except to fix their hardware failures or other severe failures of their system(s). The test center assessed the performance of solution capabilities supported by the data reported, as shown in [Table 1](#).

Nine CM solutions participated in this study; 4 were also part of the previous study (Bell et al.) approximately a year earlier. The participating solutions, in alphabetical order, are Honeywell, Molex, Project Canary, QLM Technology, Qube Technologies, Sensia solutions, Sensirion, Sensit, and SLB. Testing was performed under confidentiality agreements. Therefore, each solution is identified here by a unique identifier,

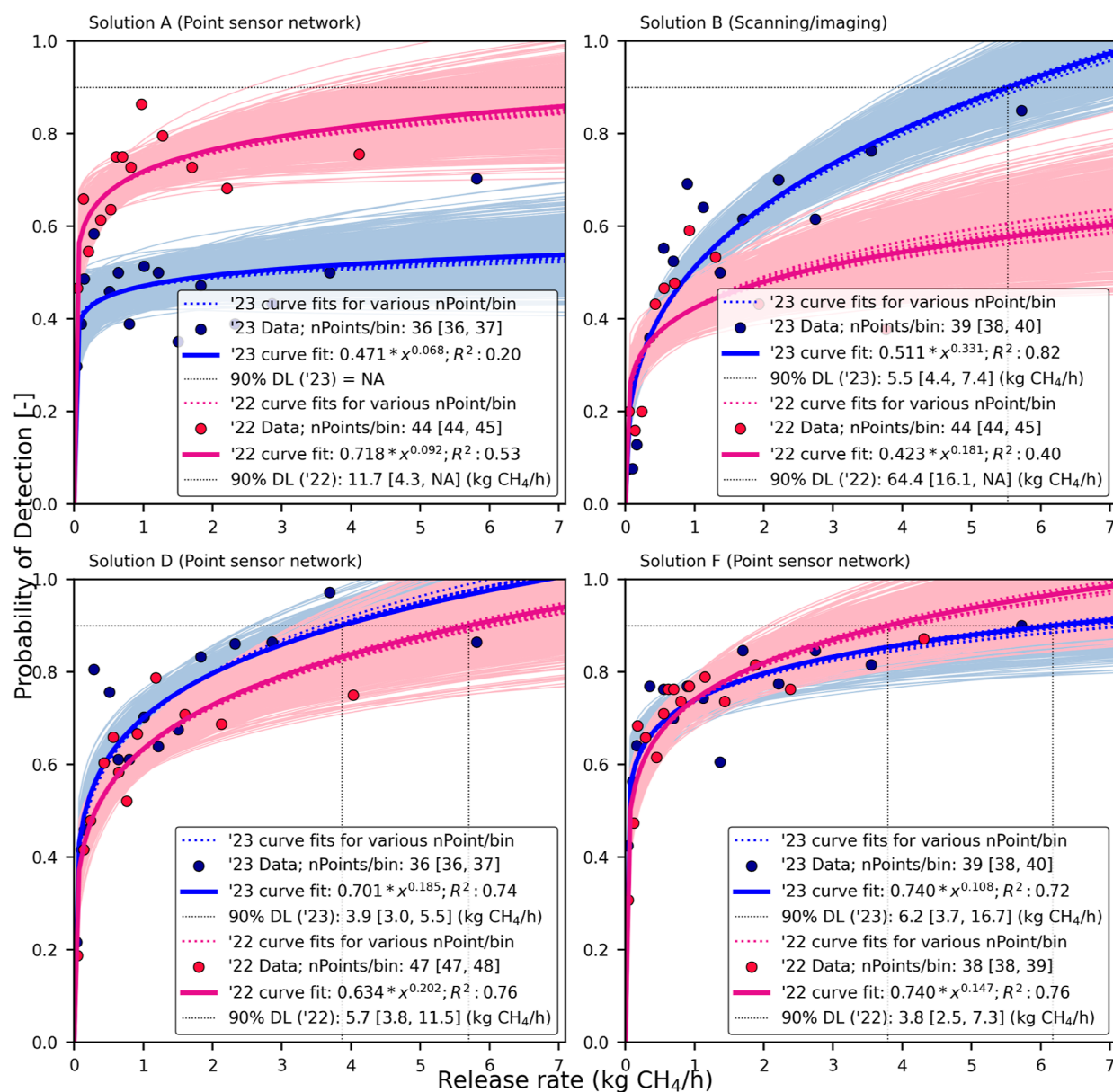


Figure 1. Probability of detection (POD) versus emission rate (kg CH₄/h) for point sensor network solutions (A,D,F) and a scanning/imaging solution (B) fitted using power functions. The *x*-axis is divided into equal-sized bins, with each marker (pod) as the fraction of controlled releases in a bin classified as true positives. Data points from the study by Bell et al. (2022) are overlaid on the current results for comparison. 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 the 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. DL90s of 3 of the 4 solutions (B,D,F) in the current study were within the tested emission rate range. Mean count of points per bin, along with the min and max counts across all bins, is also shown in the figure.

with those that participated in the prior study³⁹ retaining their identifiers. Not all solutions were tested for all metrics. Table 1 summarizes the characteristics of the solutions that participated in the study and which functionality was tested on each.

Data Processing. The study by Bell et al. used binary logistic regression models (*f*) to map the POD curves of solutions over the range of tested conditions (i.e., emission rate, release duration, etc.) and to predict the emission rate at which a solution achieved 90% POD. However, the model in some cases produced curves with nonpractical applications like unrealistic POD at zero emissions: POD at an emission rate of zero was nonzero. To correct for these issues, this study utilized power functions for POD estimation, with the intercept set to zero. The

power curve was fit to a detection fraction from equal-sized sets (quantile-based discretizations) of test conditions. The quantile used for each solution was constrained by the range $30 < N_p < 50$, where N_p is the number of points in each bin. N_p was set by using the quantile-based discretization that produced the highest goodness of fit (R^2) value. See Supporting Information for analysis on picking bin size for all solutions in this study and Table 14 for the recalculated POD curve for solutions from Bell et al.

As described earlier, detection reports were classified as TP or FP, while unreported controlled releases were classified as FN.³⁸ The protocol penalized excess detection reports identifying emission sources already identified earlier or emission sources

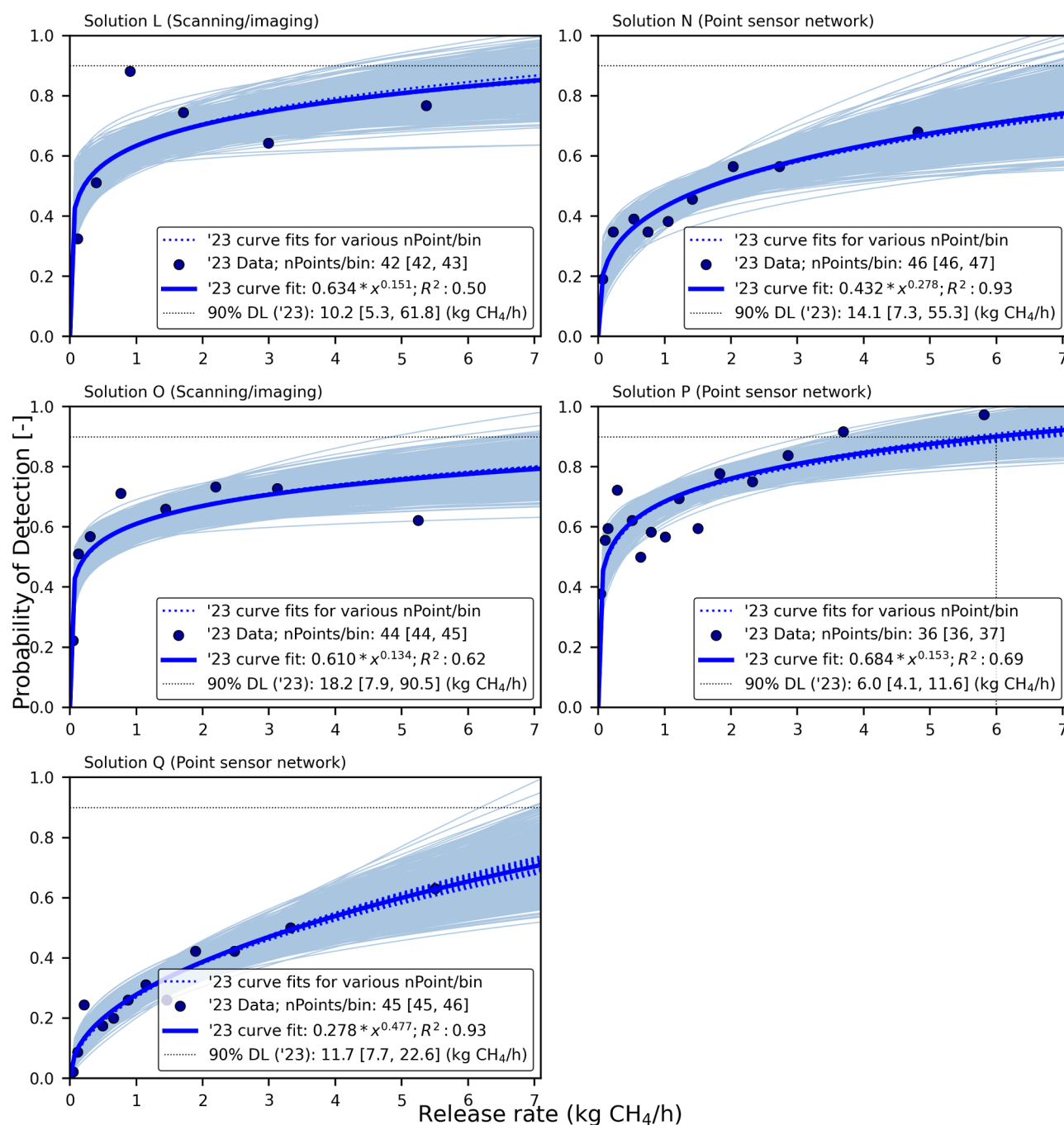


Figure 2. POD versus emission rate (kg CH₄/h) for solutions L, N, O, P, and Q fitted using power function. Solution N, P, and Q are point sensor networks, while solution L and O is a scanning/imaging solution. The x-axis of each plot is divided into equal-sized bins, with each marker (pod) calculated 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 CL 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. Emission rate at which the POD reaches 90% is indicated as the method DL90 for each solution. Best estimate of the DL90 of only solution P is within the tested emission rate range. Mean count of points per bin, along with the min and max counts across all bins, is stated in the plot legend.

not emitting during an experiment as FPs. However, in some field applications of CM solutions, such as facility level monitoring, less priority might be placed on these excess detection notifications if at least one of the alerts correctly identified an emitter. Therefore, in a break with the previous study, this study utilized 2 classifications for FP detection reports:

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 that identified a controlled release already correctly matched to another detection report as a new and/or different emission source. For example, reporting detections with

Table 2. Summary of the Number of Controlled Releases and Detection Reports Considered in the Analysis of Each CM Solution^a

ID	count		FP (%) ^b			FN (%)	DL90 ^c (kg CH ₄ /h)
	controlled release	detection reports	all	no controlled release	excess detections		
Result from the Current Study for All Participating CM Solutions							
D	547	403	6.9	28.6	71.4	31.4	3.9 [3.0, 5.5]
B	547	300	7.7	39.1	60.9	49.4	5.5 [4.4, 7.4]
F	547	444	10.6	8.5	91.5	27.4	6.2 [3.7, 16.7]
P	547	423	13.2	23.2	76.8	32.9	6.0 [4.1, 11.6]
N	417	223	18.4	29.3	70.7	56.4	14.1 [7.3, 55.3]
L	256	254	35.0	95.5	4.5	35.5	10.2 [5.3, 61.8]
O	357	324	34.6	33.0	67.0	40.6	18.2 [7.9, 90.5]
Q	547	260	38.1	21.2	78.8	70.6	11.7 [7.7, 22.6]
A ^d	547	487	47.8	61.8	38.2	53.6	NA
Results from Bell et al. for the 4 CM Solutions That Participated in Both Studies							
D	574	376	10.4	79.5	20.5	41.3	5.7 [3.8, 11.5]
F	574	516	22.5	39.7	60.3	30.3	3.8 [2.5, 7.3]
B	445	250	31.2	61.5	38.5	61.3	64.4 [16.1, NA]
A	574	986	59.8	26.9	73.1	31.0	11.7 [4.3, NA]

^aThe breakdown of the FP rates for all solutions using the ADED protocol is also shown, together with the FN rate, and DL90s predicted by each solution. Solutions are sorted in order of increasing all FP rate. ^bAll is the percentage of all detections classified as FP based on the ADED protocol. **No controlled release** is the fraction of all FPs that is due to detection reports sent when there was no controlled release at the test center. **Excess TP Detections** are the fraction of all FPs that is due to excess detections identifying controlled releases that have been matched already as a new and/or different emitters. ^cWhen the POD curve could not evaluate the DL90, they are given as "NA". Similarly, when the lower and upper empirical 95% confidence interval (CI) on a solution's DL90 could not be evaluated, they are given as 0 and NA, respectively. ^dOne of the sensors installed failed during the study.

different *EmissionSourceIDs* during an experiment with one controlled release.

Limitations and Scope of Study. As extensively discussed in Bell et al., the protocol assumes that while solutions provide near-CM, they issue discrete detection alerts that are source-resolved whenever emissions were detected. This is not always the case, as several solutions provide time-series sensor readings through web-based dashboards for operators to read and infer detection decisions. Additionally, pads 4 and 5 (Supporting Information, Figures S-1 and S-3) at METEC used for this study were designed to mimic simplified on-shore production facilities (see Zimmerle et al. for details on how it differs from a real facility). Hence, the result from this study might not be applicable to more complex or midstream facilities, which likely have different site configurations and emissions behaviors. This study assesses and compares the efficacy of current solutions for emissions detection, localization, and quantification with the findings of Bell et al. A subsequent publication will conduct a detailed performance analysis, focusing on the impact of variables such as sensor count and placement, characteristics of emission points, meteorological conditions, and additional relevant factors.

RESULTS AND DISCUSSION

In this section, we discuss study results based on the following metrics: (1) POD, (2) source localization (precision and accuracy), (3) quantification accuracy, and (4) time to detection. This section further shows the changes in performance of solutions individually and as a group relative to the study by Bell et al.

Primary Results and Analysis. POD. A POD curve relates the probability that a solution will detect an emission of a given rate, as composite performance over all other test conditions like release duration, emission flow rate, wind speed, etc. that could affect the POD of solutions. A multivariable logistic regression

analysis of the impact of these factors on POD over the tested range showed varying statistical significance across all solutions. Results indicate that emission rates significantly ($p < 0.05$) affected the POD of all solutions, with other variables affecting only a subset of solutions (see Table 27 in SI). Figures 1 and 2 show the POD curves for all solutions mapped over the range of emission rates tested. Figure 1 compares curves for the 4 solutions that participated in both the current study and that by Bell et al., while Figure 2 is for the other 6 solutions. Bell et al. defined the method detection limit (DL90) of each solution as the emission rate at which the solution, as deployed (method), detected emitters 90% of the time over a wide range of meteorological conditions. The study team deviated from the acronym MDL used by Bell et al. to avoid it being misinterpreted as "minimum detection limit", which might mean something different. The DL90 metric is an important consideration during the formulation of methane emissions reduction policies/programs²⁷ by regulations and their implementation by O&G operators. Figures 1 and 2 and Table 2 show the DL90s of solutions.

- Performance from the current study (2023): overall, Figures 1 and 2 showed that the POD curves predicted the DL90s of 8 of the 9 solutions ranging from 3.9 [3.0, 5.5] kg CH₄/h to 18.2 [7.9, 90.5] kg CH₄/h. The DL90s of 4 of the 8 solutions fell within the range of emission rates tested in the study. Table 2 shows that the 4 solutions with the lowest FP rates (6.9–13.2%) also had the lowest DL90s (3.9 [3.0, 5.5] kg CH₄/h to 6.2 [3.7, 16.7] kg CH₄/h), while 3 of the 4 solutions had the lowest FN rates (27.4–32.9%) in the study (Supporting Information, Figure S-10). This indicates efforts at balancing method sensitivity (i.e., low DL90) with low FP and FN rates. In contrast, the remaining 6 solutions had relatively higher DL90s (no solution within the tested emission rate range), FP rates (all solutions >20%), and FN rates (5

Table 3. Summary of Emission Source Localization (Equipment Unit) Precision and Accuracy for All Participating Solutions Arranged in Decreasing Localization Precision Equipment Unit Level

ID	sensor density (sensors/m ²)	count of TPs	source localization (equipment unit)					
			precision (%)			accuracy (%)		
			unit level	group level	facility level	unit	group level or better	facility level or better
Result from the Current Study for All Participating CM Solutions								
B	0.000118	277	89.5	9.4	1.1	82.7	91.3	92.3
L	0.000118	165	86.7	10.9	2.4	56.3	63.4	65.0
O	0.000118	212	76.4	12.7	10.9	50.0	58.3	65.4
N	0.00213	182	51.6	41.8	6.6	42.2	76.2	81.6
F	0.00118	397	40.8	53.9	5.3	36.5	84.7	89.4
Q	0.00154	161	28.0	54.0	18.0	17.3	50.8	61.9
D	0.000947	375	27.2	68.8	4.0	25.3	89.3	93.1
P	0.00071	367	27.0	56.9	16.1	23.4	72.8	86.8
A ^a	0.000947	254	26.0	49.6	24.4	13.6	39.4	52.2
Results from Bell et al. for the 4 CM Solutions That Participated in Both Studies								
B	0.000118	172	70.9	15.7	13.4	48.8	59.6	68.8
A	0.000947	396	28.0	39.4	32.6	11.3	27.1	40.2
F	0.000947	400	24.8	50.2	25.0	19.2	58.1	77.5
D	0.000947	337	0.0	52.8	47.2	0.0	47.3	89.6

^aOne of the sensors installed failed during the study.

solutions $\geq 50\%$), which might indicate struggles at emissions detection. At a minimum detection threshold of 0.40 kg CH₄/h (as stipulated in the final rule by the USEPA), results indicate that 5 of the 9 solutions will have $\geq 50\%$ POD.⁴¹

For the *scanning/imaging* solutions, FP rate spanned between 7.7 and 34.6%, with the DL90 of 1 of the 3 solutions within the tested range. While the FP rates of *point sensor network* solutions were between 6.9 and 38.1%, with the DL90 of 3 of the 5 solutions that estimated DL90s were within the tested range. A review of the percentage of FPs due to excess detections (4.5–91.5% with 7 of the 9 solutions having values $\geq 50\%$) suggests that if the intended application of most solutions is to correctly alert operators of ongoing emissions with less priority on what is emitting and the number of emitters, then these solutions would have much lower FP rates than predicted by the protocol. Otherwise, follow-up OGI surveys might take a longer time by investigating misleading alerts, which is costly.

As previously highlighted and detailed in [Supporting Information](#), wind speed significantly influenced the POD of 5 out of the 9 solutions tested ($p < 0.05$), with many solutions, especially point sensors, relying on favorable wind transport for effective detection (i.e., sensors must be situated downwind of an emission plume). The results summarized in [Supporting Information](#) demonstrate that the DL90s for these solutions, calculated using the mean wind speed under which each solution was tested, and emission rate normalized by wind speed (as shown in [Supporting Information](#), Figures S-12 and S-13), were statistically distinct from the DL90s derived solely from emission rates. Thus, the four solutions (D, B, F, and P) with a quoted DL90 range of 3.9 [3.0, 5.5] kg CH₄/h to 6.2 [3.7, 16.7] kg CH₄/h based solely on emission rate over tested wind speed range of 0.7 to 9.9 m/s, alternately have a representative DL90 range of 6.2 [4.7, 9.7] kg CH₄/h to 10.0 [6.2, 21.8] kg CH₄/h if calculated using the mean wind speed under which each solution was tested (see [Supporting Information](#)). In field applications, these solutions are deployed to operate continuously and report emissions data at all times, regardless of the prevailing weather conditions, as was the case in this study.

Given that the test center utilized in this study offers conditions close to ideal operational field conditions (e.g., the absence of highly variable baseline emissions from nonfugitive sources) and that weather conditions may vary greatly across different field locations (e.g., Denver-Julesburg Basin, Permian Basin, and Appalachian Basin), the DL90s assessed in this study, as well as those by Bell et al., likely represent best-case scenario estimates.

A Spearman's rank correlation analysis showed that the count of sensors deployed by solutions did not necessarily affect the method sensitivity of solutions (p value > 0.5), as solutions that deployed more sensors did not always have a lower DL90 compared to solutions that installed fewer sensors. Aside from the difference in the sensor type, quality, and proprietary algorithms, which can vary the performance of solutions, one potential explanation for this observation might be overdeployment of sensors by some solutions. However, given the reporting constraints of the test protocol, solutions did not attribute detections to any sensor(s), hence making the assessment of overdeployment (if any) challenging in this study. In general, TP rate tended to increase with the release rate for all solutions, as shown by the figures above and [Supporting Information](#). See [Supporting Information](#), Figures S-12 to S-17, for POD curves for all solutions based on release duration, wind speed, and release rate normalized by wind speed. See [Supporting Information](#), Figures S-18 to S-19, for POD curves using logistic regression for reference.

- Comparing general performance from Bell et al. to the current study: results from the study by Bell et al. showed that more solutions struggled at balancing low MDL, FP rate, and FN rate when compared to current test results. Two of 11 solutions showed efforts at balancing all 3 metrics relative to other solutions, while others showed mixed performance. For example, solution E had the lowest DL90 (1.3 [0.5, 8.1] kg CH₄/h) and FN rate (12.3%) but had the highest FP rate (82.6%) in the study. While solution J was among the solutions with the lowest DL90 (4.0 [3.4, 5.1] kg CH₄/h) and FP rate (0.0%) but had one of the highest FN rate (76.0%) in the study ([Supporting Information](#)). These results have noted the

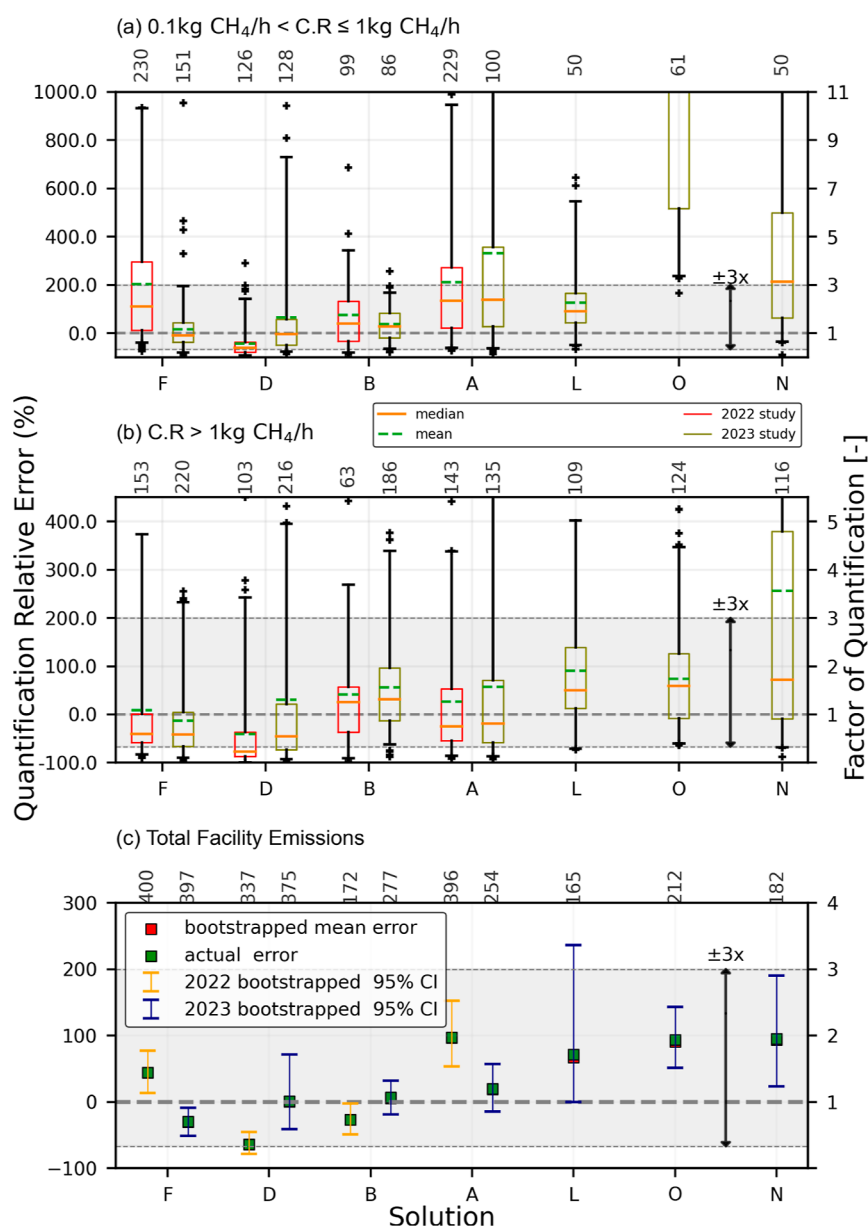


Figure 3. Quantification relative error for solutions categorized by (a) controlled release rate [0.1–1] kg CH₄/h, (b) controlled release rate ≥1 kg CH₄/h, and (c) total facility emissions. Bottom panel (c) summarizes the site-level relative error for each solution arranged in increasing order from left (sol. F) to right (sol. M) based on the current study data. Site-level relative error is bootstrapped to estimate the uncertainty on the actual error. Markers represent bootstrapped site-level mean relative error (red), and the actual site-level relative error (green), respectively. Whiskers represents the 95% CI on the bootstrapped mean relative error. Middle (b) and top panels (a) are boxplots summarizing relative error distribution for each solution over selected range of controlled release rates. Each box represent the interquartile range of data with whiskers including 95% of data. Upper y-axis of (a,b) are arbitrarily trimmed at 400 and 1000%, respectively, with the full 95% CI, as shown in Table 4. Across all panels, results from the study by Bell et al. (2022), is also shown to facilitate comparison. The x-axis of all panels are arranged based on (c), while the shaded zone indicates region within a quantification factor of 3.

tendency for solutions to trade-off detection sensitivity with FP and FN rates: changing solution settings to reduce DL90 tends to increase FP rate. Additionally, at a minimum detection threshold of 0.40 kg CH₄/h, 4 of the 11 solutions had ≥50% POD.⁴¹ In general, setting algorithms to reduce DL90 also makes it more difficult to distinguish smaller fugitive emissions from background concentrations (i.e., sensor or algorithmic noise), leading to background fluctuations being classified as FP emissions detections. Conversely, higher DL90s can imply solutions missing relatively smaller rate emissions,

which typically make up the majority of field measurement studies (by count), resulting in high FN rates. However, generally, solutions from the current study showed more efforts at balancing low DL90 with low FN and FP rates compared to the results by Bell et al.

- Comparing the performance of the four solutions common to both studies: two solutions, B and D, showed reduced DL90 with FN and FP rates relative to the study by Bell et al. The FP and FN rate of solution F—with highly overlapping DL90 uncertainty across both studies—also dropped. These data indicate a general

Table 4. Summary of Single-Estimate Quantification for Solutions along with Their 95% Empirical Confidence Limits^a

ID	estimates within $\pm 3 \times$ (%)			relative quantification error (%)					
	error relative to CR ^c ($-67\% _{\frac{1}{3}} \times$, $+200\% _{3 \times}$)			CR [0.1–1] kg CH ₄ /h			CR > 1 kg CH ₄ /h		
	CR (kg CH ₄ /h)			mean	median	95% CI	mean	median	95% CL
Results from the Current Study for All Participating CM Solutions									
B	90	96	89	37.4	28.1	[−65.0, 168.5]	55.7	31.5	[−62.3, 339.4]
L	81	84	81	126.2	91.2	[−49.5, 546.6]	90.4	50.0	[−70.7, 402.2]
F	78	90	71	15.7	−9.6	[−80.4, 195.5]	−12.8	−41.8	[−89.6, 232.3]
D	67	79	61	64.5	−3.0	[−75.8, 729.0]	30.8	−45.8	[−92.5, 395.7]
A ^b	60	55	72	330.4	138.4	[−62.2, 1803.4]	57.6	−18.3	[−86.3, 612.5]
N	56	48	62	1036.6	212.8	[−36.2, 2900.9]	256.0	72.0	[−68.0, 1671.6]
O	54	1	86	1751.2	1074.9	[235.9, 4071.5]	73.4	58.8	[−60.4, 347.0]
Results from Bell et al. for the 4 CM Solutions That Participated in Both Studies									
B	74	76	80	74.6	39.5	[−81.1, 343.2]	41.9	25.3	[−90.2, 268.8]
F	65	62	75	202.2	110.9	[−39.7, 933.2]	9.2	−40.5	[−82.5, 373.6]
A	64	65	73	211.3	134.2	[−60.9, 946.8]	27.1	−24.2	[−85.6, 338.5]
D	48	60	34	−43.0	−60.1	[−92.6, 141.4]	−40.0	−77.0	[−99.9, 242.4]

^aThe percentage of measurements within a factor of 3 is shown for both the current study and the previous study for comparison. ^bOne of the sensors installed failed during the study. ^cColumns identify the fraction of estimates in the study by Bell et al. and that of the current study, which were within a factor of 3 relative to the controlled release rate.

improvement in efforts to balance method sensitivity with FP and FN rates. Given that these solutions installed the same number of sensors as in Bell et al. except for solution F, which increased from 8 to 10, improved performance could be attributed to improved analytics/algorithms and/or more favorable test conditions, as shown in [Supporting Information](#) (higher emission rates, longer release durations, and lower wind speeds). At higher emission rates, solutions either exceeded or approached their respective DL90s, while testing at calmer wind speeds likely reduced turbulent gas plume dispersion in support of more stable/steady measurements. Longer release durations likely gave *scanning/imaging* solutions multiple opportunities to visualize and identify emissions or longer averaging time of ambient concentration measurement to infer detections by *point sensor network* solutions.

Source Localization. As discussed earlier, the protocol required solutions to report the equipment unit housing any identified emitter. For each solution, sensor density was defined as the ratio of the number of sensors deployed by the solution to the designated test center (pads 4/5 at METEC) surface area (m²). [Table 3](#) and S-30 (in the [Supporting Information](#)) summarizes the sensor densities (sensors/m²) and the emission source localization precision and accuracy results of solutions participating in this study and those in the study by Bell et al. Similar localization metrics were evaluated if solutions reported the GPS coordinates of identified emitters. See the performance report of each solution in the [Supporting Information](#) for those analysis.

- Performance from the current study (2023): at the equipment unit level, all 3 *scanning/imaging* solutions had the highest localization precisions (>70%) and accuracies (>40%), with the smallest sensor densities (0.000118 sensors/m²). For the 6 *point sensor network* solutions, only 1 solution (also with the largest sensor density) had localization precision and accuracy >40%. At the equipment group level or better (equipment group +

unit level), all *scanning/imaging* solutions had >95% localization precision and an accuracy range of 58.3–91.3%, while for the *point sensor network* solutions, 3 solutions had precisions >90% and accuracies >70%, with a sensor density range of 0.000947 sensors/m² to 0.00213 sensors/m². These results illustrate the higher tendency of *scanning/imaging* solutions in this study to correctly narrow down emitters for follow-up OGI surveys than *point sensor network* solutions, despite installing the lowest number of sensors. In general, 6 of the 9 solutions had localization precisions more than 90% at the equipment group level or more, while 5 of the 9 solutions had localization accuracy >70% also at that level. As indicated earlier, for operators deploying CM solutions at multiple (sometimes in 100 s), bigger facilities, narrowing down the source of emitters, if fit for purpose, can have huge time- and cost-saving benefits for operators. However, this functionality might not be an important consideration if the intended application or the inherent capacity of a solution does not support source level localization (i.e., facility level emissions monitoring).

- Comparing the general performance from Bell et al. to the current study: at the equipment unit level, 3 of 5 *scanning/imaging* solutions had the highest localization precisions (>60%) and accuracies (>40%), with the sensor density range of 0.000118 sensors/m² to 0.00416 sensors/m². All *point sensor network* solutions had precisions <50% and accuracies <20% at that level. At the equipment group level or better (equipment group + unit level), one solution (with the largest sensor density) had >90% localization precision and >70% accuracy. As a group, when compared to the current study results, performance generally improved from the study by Bell et al. These improvements could be attributed to the rapid development of the algorithms/analytics of solutions; often the major driver of source localization in CM solutions. Favorable test conditions, as shown in [Supporting Information](#) (higher emission rates, longer release

durations, and lower wind speeds), could also be a factor as solutions had longer and multiple opportunities to see the gas plumes (*scanning/imaging* solutions) or gather ambient measurement data (*point sensor network* solutions) at relatively calmer wind conditions to arrive at better localization estimates relative to prior studies.

- Comparing the performance of the four solutions common to both studies: the localization precisions and accuracies of solutions B, D, and F (with a larger sensor density in the current study) improved at both equipment unit level and equipment group level or better, relative to the study by Bell et al. Solution A had a mixed result, with only localization precision at equipment group level or better improving.

Quantification Accuracy. Seven of the 9 solutions tested the emissions quantification capability. Panels (a,b) of Figure 3 are box and whisker plots showing quantification relative error distribution for each solution for controlled release rate ranges of [0.1–1] kg CH₄/h and >1 kg CH₄/h, respectively. Emission rates in the range of [0.1–1] kg CH₄/h roughly represent equipment component leak rates typically identified through OGI surveys,^{23,42,43} while rates in the range >1 kg CH₄/h represent relatively larger leak rates due to process failures at production facilities.^{42,44} Panel (c) of the figure is an error bar plot showing facility level quantification relative errors (actual and simulated mean) for solutions over the duration tested, along with associated uncertainties obtained through bootstrapping (see Supporting Information, Section S-9.2 for the bootstrapping procedure). Across all panels, the gray shaded area shows an emission rate estimation range within a quantification factor of 3 ($-67\% \left| \frac{1}{3} \times, +200\% \right| 3 \times$) of actual release rates. The results of the 4 solutions that were also tested in the study by Bell et al. are shown in the plots for comparison. Tables 4 and S-31 (in the Supporting Information) summarize for both this study and Bell et al. the percentage of reported estimates within a factor 3 for (1) all controlled releases detected, (2) detected controlled releases within the range of [0.1–1] kg CH₄/h, and (3) detected controlled releases within the range >1 kg CH₄/h.

- Performance from the current study (2023): considering all controlled release rates classified as TP, solutions had 54–90% of their estimates within a factor of 3. For emission rates within the range of [0.1–1] kg CH₄/h, Figure 3 and Table 4 shows that the individual estimate relative errors of all solutions were positively skewed (mean > median). Four of the 7 solutions (including 2 of 3 *scanning/imaging* solutions) in this range had 79–96% of their estimates within a factor of 3, while the remaining solutions had 1–55% of their estimates also within that factor. At a 95% empirical confidence interval, 1 of the 7 solutions (*scanning/imaging*) had both the lower and upper individual estimate relative error limits within a factor of 3, while 4 of the 7 solutions (including 2 of 5 *point sensor network* solutions) had both of their limits within a factor of 10 ($-90\% \left| \frac{1}{10} \times, +900\% \right| 10 \times$). In general, individual estimates ranged from $\approx \frac{1}{5} \times$ to $\approx 42 \times$ the actual rates in this range. Typically, field operations are characterized by a higher background methane concentration than what is obtainable at METEC. Hence, the detection and quantification of some

emissions with rates in this range can be challenging for solutions as emissions are intermittent and can easily blend with background methane concentrations. However, assuming current solution performances are extrapolated to the field, the majority of rate estimates in this range by most solutions may be within a factor of 3 (mostly by overestimation as mean relative errors are skewed high), with individual estimates having wide uncertainty. For emission rates within the range >1 kg CH₄/h, the individual estimate relative errors for all solutions were positively skewed. All the solutions had 61–89% of their estimates of rates in this range within a factor of 3. Five of the 7 solutions (including all *scanning/imaging* solutions) had $\geq 71\%$ of their estimates within a factor of 3, while the remaining solutions having about 62% of their estimates also within the factor. At a 95% empirical confidence interval, 5 of the 7 solutions (including all *scanning/imaging* solutions) had both lower and upper individual estimate relative error limits within a factor of 10. In general, single estimates ranged from $\approx \frac{1}{13} \times$ to $\approx 18 \times$ the actual rates in this range. In field deployments, the wide uncertainty limit on individual estimates for rates in this range can produce grossly misleading results for LDAR programs. For example, overestimating a relatively large emission (e.g., leak rate of 7.1 kg CH₄/h—maximum rate tested in this study) by 18 \times can lead to a bogus alert of emissions at a scale of a super emitter (≥ 100 kg CH₄/h). Generally, in this emission rate range, solutions with a majority of their estimated emissions within a factor of 3 increased, indicating that solutions were likely better at quantifying larger emissions compared to smaller ones (Supporting Information, Figure S-26).

At the facility level, over the study duration, 6 of the 7 solutions estimated emissions within a factor of 2 ($-50\% \left| \frac{1}{2} \times, +100\% \right| 2 \times$) with their respective simulated lower and upper limits within a factor of ≈ 3 . Given the increasing interest in facility-level quantification as inferred from the USEPA final rule, this result indicates that these solutions are likely to provide facility-level emissions estimation with higher accuracy and narrower uncertainty than single estimates, which has important policy implications. See Supporting Information, Figures S-27 and S-28, for the impact of release duration and wind speed on the individual estimates of the relative errors of solutions.

- Comparing general performance from Bell et al.: the quantification performance in the study by Bell et al., as summarized in Tables 4 and S-31 (in the Supporting Information), showed that, as a group, solutions experienced more difficulty at accurately quantifying emissions relative to the current study. For all controlled release rates, 3 solutions in the current study had fraction of estimates within a factor of 3, greater than the highest value obtained in the study by Bell et al. (74%). Similarly, for emission rate ranges [0.1–1] kg CH₄/h and >1 kg CH₄/h, 4 and 3 solutions had fractions of estimates within a factor of 3, greater than the highest values (76 and 80%, respectively) obtained in the previous study. In general, as a group, the quantification results from the current study improved relative to those by Bell et al. although individual estimates still have high uncertainty. There

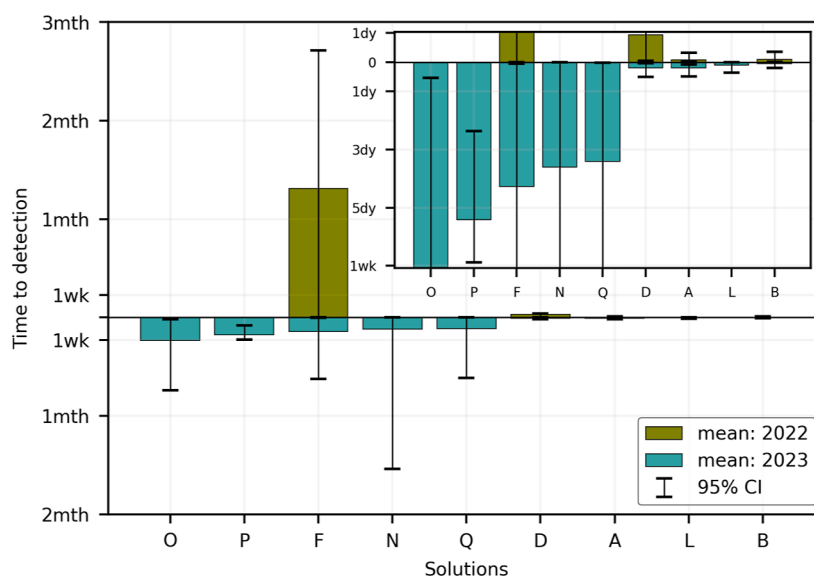


Figure 4. Time to detection for all participating solutions from both the previous (2022) and current studies. Bars representing the mean time to detection are sorted in decreasing order from left (solution O) to right (solution B) using data from the current study. The time to detection of the 4 solutions (A, B, D, and F) from Bell et al. is shown in the upper half of the figure, while that of the current study is shown in the bottom half. Whiskers represent the 2.5 (lower) and 97.5 (upper) percentiles of the data for each solution. The insert is a miniature version of the original plot with the upper y -axis trimmed at a time to detection of 1 day and the lower y -axis trimmed at a time to detection of 1 week.

was no drastic change in facility level quantification performance as all but one solution estimated facility level emissions within a factor of 2 in both studies. As highlighted earlier, this improvement as a group could be attributed to favorable testing conditions and/or improvement in the analytics of solutions, especially for the 4 solutions retesting in the current study.

- Comparing the performance of the four solutions common to both studies: relative to the study by Bell et al., for all controlled releases detected, and those within the range [0.1–1] kg CH₄/h, the percentage of estimates within a factor of 3 increased for solutions B, F, and D, while only that of solutions B and D increased for emission rates in the range >1 kg CH₄/h. At a 95% empirical confidence interval, the individual estimate relative error limits of solution B became narrower for emission rates in the range [0.1–1] kg CH₄/h but had a mixed result for rates in the range >1 kg CH₄/h. Solutions D and F had mixed results for both emission rate ranges, while for solution A, the uncertainty got wider for both emission rate ranges. At the facility level, all 4 solutions improved in quantification accuracy.

Time to Detection. As briefly discussed earlier, the emissions mitigation potential of CM solutions also depends on the fraction of deployment duration during which solutions are operational to collect and transmit data (operational time)²⁷ and how quickly emitters are identified and communicated to operators. Tables S-4 and S-3 (in [Supporting Information](#)) show the operational factors of solutions in this study and in Bell et al. [Figure 4](#) and [Table 32](#) (in [Supporting Information](#)) shows the calculated time to detection for TP detections by solutions. In this study, 2 solutions (P and Q) did not automate their detection reporting processes. Since the study team could not assess the extent of human support (if any) for solutions with automated reporting, especially when there was failure in data transmission, the assessed time to detection also captured the inefficiencies likely introduced by human interference, e.g., time

taken to manually prepare detection reports, as prescribed by the test protocol.³⁸

- Performance from the current study (2023): [Figure 4](#) shows that at a 95% empirical confidence interval, 4 of the 9 solutions had mean times to detection <5 h with upper limits <15 h; 2 solutions had upper limits less than the maximum release duration in this study (10.2 h). Unlike the profile of emissions in this study (steady rates released for hours), several leaks typically found in the field are intermittent; hence, solutions typically have shorter windows than those available in this study to collect and communicate measurement data to operators. Additionally, results show that 6 of the 9 solutions were operational at least 90% of their deployment time, with 5 solutions operational throughout the study (operational factor of 1—[Supporting Information](#)). The USEPA stipulates a rolling 12 month average operational downtime <10% (operational factor >90%) in their final for CM solutions.⁴¹
- Comparing general performance from Bell et al. For the study by Bell et al., at a 95% empirical confidence interval, 3 of the 11 solutions had mean times to detection <5 h with upper limits <15 h and 1 solution had upper limit less than the maximum release duration considered for the solution. Results shows that, as a group, relative to Bell et al., current study results generally improved in this area.
- Comparing the performance of the four solutions common to both studies: at a 95% empirical confidence interval, the mean times to detection, their respective lower and upper limits, and operational factors for solutions B, D, and F improved relative to previous results in Bell et al.

Implications. The growing interest by stakeholders, including operators and regulators, in CM as a faster, temporally resolved approach for methane emissions detection, measurement, and mitigation is driving the rapid development of CM solutions. Therefore, regular and robust testing of solutions is

required to characterize and compare performance levels (intra and inter solutions) using a standardized/consensus testing protocol. This study is the second implementation (first by Bell et al.) of a consensus protocol (ADED CM protocol) to assess the progress of solutions. The results from the study highlight a few key points. First, solutions that were tested before generally exhibited better performance on many performance metrics relative to (1) their previous performance in Bell et al. and (2) other solutions tested for the first time under the protocol. The majority of solutions that were retested in this study had the lowest FP rates and DL90s, the highest localization accuracy at the equipment group, or better performance in the study. They were also among the solutions with the lowest FN rates and highest quantification performance (estimates within a factor of 3) across different emission rate ranges ($[0.1-1]$ kg CH₄/h and >1 kg CH₄/h). Similarly, across all metrics assessed, most of the solutions that were retested improved in performance when compared to their previous results, highlighting the benefits of regular quality testing. Users, however, should be cautious given that these results are likely more representative of non-intermittent emissions from fugitive events, which make up a relatively smaller fraction of reported upstream emissions. Second, single source emission estimates by solutions still have wide uncertainty, which is unsuitable for accurate measurement-based inventory development and reporting programs. On the other hand, solutions had better quantification accuracy with narrower uncertainty at the facility level. This result, if replicable in the field and applied to sites similar to METEC, shows promises of reliable facility-level quantification performance by these solutions, especially when adopted for regulatory programs in the near-future, provided that the observed rapid development of CM solutions was sustained. Overall, solutions need not have excellent performance across all metrics assessed in this study to be useful; i.e., rapid detection of large emissions sources for repairs might not require accurate quantification. As well, higher DL90 at a low FP rate could mitigate larger emissions with minimal cost of follow-up investigations.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.est.3c08511>.

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) (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Daniel Zimmerle – Energy Institute, Colorado State University, Fort Collins, Colorado 80524, United States; orcid.org/0000-0003-2832-048X; Email: Daniel.Zimmerle@colostate.edu

Authors

Chiemezie Ilonze – Department of Mechanical Engineering, Colorado State University, Fort Collins, Colorado 80523, United States; orcid.org/0000-0003-4231-3278

Ethan Emerson – Energy Institute, Colorado State University, Fort Collins, Colorado 80524, United States; orcid.org/0000-0003-0350-3810

Aidan Duggan – Energy Institute, Colorado State University, Fort Collins, Colorado 80524, United States; orcid.org/0000-0003-4344-5977

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.est.3c08511>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

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.

■ REFERENCES

- (1) IEA. Methane and Climate Change—Global Methane Tracker 2022—Analysis. 2023, <https://www.iea.org/reports/global-methane-tracker-2022/methane-and-climate-change> (accessed July 20, 2023).
- (2) NOAA. Greenhouse Gases Continued to Increase Rapidly in 2022. 2023, <https://www.noaa.gov/news-release/greenhouse-gases-continued-to-increase-rapidly-in-2022> (accessed Sept 3, 2023).
- (3) EPA. Overview of Greenhouse Gases. 2023, <https://www.epa.gov/ghgemissions/overview-greenhouse-gases> (accessed July 20, 2023).
- (4) U.S. Energy Information Administration (EIA). Natural Gas Explained. 2023, <https://www.eia.gov/energyexplained/natural-gas/> (accessed July 20, 2023).
- (5) IPCC. Summary for Policymakers—Global Warming of 1.5 °C. 2023, <https://www.ipcc.ch/sr15/Chapter/spm/> (accessed July 20, 2023).
- (6) Brandt, A.; Heath, G.; Kort, E.; O'Sullivan, F.; Petron, G.; Jordaan, S.; Tans, P.; Wilcox, J.; Gopstein, A.; Arent, D.; Wofsy, S.; Brown, N.; Bradley, R.; Stucky, G.; Eardley, D.; Harriss, R. Methane Leaks from North American Natural Gas Systems. *Science* **2014**, *343*, 733–735.
- (7) 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. *Environ. Sci. Technol.* **2022**, *56*, 14743–14752.
- (8) Lavoie, T. N.; Shepson, P. B.; Cambaliza, M. O. L.; Stirn, B. H.; Conley, S.; Mehrotra, S.; Faloon, I. C.; Lyon, D. Spatiotemporal Variability of Methane Emissions at Oil and Natural Gas Operations in the Eagle Ford Basin. *Environ. Sci. Technol.* **2017**, *51*, 8001–8009.
- (9) 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. *Environ. Sci. Technol.* **2017**, *51*, 8832–8840.
- (10) 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. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 11712–11717.
- (11) 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.
- (12) Zavala-Araiza, D.; Lyon, D. R.; Alvarez, R. A.; Davis, K. J.; Harriss, R.; Herndon, S. C.; Karion, A.; Kort, E. A.; Lamb, B. K.; Lan, X.; Marchese, A. J.; Pacala, S. W.; Robinson, A. L.; Shepson, P. B.; Sweeney,

- C.; Talbot, R.; Townsend-Small, A.; Yacovitch, T. I.; Zimmerle, D. J.; Hamburg, S. P. Reconciling Divergent Estimates of Oil and Gas Methane Emissions. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 15597–15602.
- (13) Alvarez, R. A.; Zavala-Araiza, D.; Lyon, D. R.; Allen, D. T.; Barkley, Z. R.; Brandt, A. R.; Davis, K. J.; Herndon, S. C.; Jacob, D. J.; Karion, A.; Kort, E. A.; Lamb, B. K.; Lauvaux, T.; Maasakkers, J. D.; Marchese, A. J.; Omara, M.; Pacala, S. W.; Peischl, J.; Robinson, A. L.; Shepson, P. B.; Sweeney, C.; Townsend-Small, A.; Wofsy, S. C.; Hamburg, S. P. Assessment of Methane Emissions from the U.S. Oil and Gas Supply Chain. *Science* **2018**, *361*, 186–188. eaar7204
- (14) Miller, S. M.; Wofsy, S. C.; Michalak, A. M.; Kort, E. A.; Andrews, A. E.; Biraud, S. C.; Dlugokencky, E. J.; Eluszkiewicz, J.; Fischer, M. L.; Janssens-Maenhout, G.; Miller, B. R.; Miller, J. B.; Montzka, S. A.; Nehrkorn, T.; Sweeney, C. Anthropogenic Emissions of Methane in the United States. *Environ. Sci. Technol.* **2013**, *110*, 20018–20022.
- (15) IEA. The Global Methane Pledge—Global Methane Tracker 2022—Analysis. 2023, <https://www.iea.org/reports/global-methane-tracker-2022/the-global-methane-pledge> (accessed Sept 3, 2023).
- (16) European Commission. Methane Emissions. 2023, https://energy.ec.europa.eu/topics/oil-gas-and-coal/methane-emissions_en (accessed Sept 3, 2023).
- (17) EPA. EMC: Continuous Emission Monitoring Systems. 2023, <https://www.epa.gov/emc/emc-continuous-emission-monitoring-systems> (accessed Sept 3, 2023).
- (18) US EPA. Leak Detection and Repair: A Best Practices Guide. 2023, <https://www.epa.gov/compliance/leak-detection-and-repair-best-practices-guide>, (accessed July 20, 2023).
- (19) 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. *Environ. Sci. Technol.* **2015**, *49*, 8167–8174.
- (20) Zavala-Araiza, D.; Alvarez, R. A.; Lyon, D. R.; Allen, D. T.; Marchese, A. J.; Zimmerle, D. J.; Hamburg, S. P. Super-Emitters in Natural Gas Infrastructure Are Caused by Abnormal Process Conditions. *Nat. Commun.* **2017**, *8*, 14012.
- (21) Chen, Y.; Sherwin, E. D.; Berman, E. S.; Jones, B. B.; Gordon, M. P.; Wetherley, E. B.; Kort, E. A.; Brandt, A. R. Quantifying Regional Methane Emissions in the New Mexico Permian Basin with a Comprehensive Aerial Survey. *Environ. Sci. Technol.* **2022**, *56*, 4317–4323.
- (22) Subramanian, R.; Williams, L. L.; Vaughn, T. L.; Zimmerle, D.; Roscioli, J. R.; Herndon, S. C.; Yacovitch, T. I.; Floerchinger, C.; Tkacik, D. S.; Mitchell, A. L.; Sullivan, M. R.; Dallmann, T. R.; Robinson, A. L. Methane Emissions from Natural Gas Compressor Stations in the Transmission and Storage Sector: Measurements and Comparisons with the EPA Greenhouse Gas Reporting Program Protocol. *Environ. Sci. Technol.* **2015**, *49*, 3252–3261.
- (23) Zimmerle, D. J.; Williams, L. L.; Vaughn, T. L.; Quinn, C.; Subramanian, R.; Duggan, G. P.; Willson, B.; Opsomer, J. D.; Marchese, A. J.; Martinez, D. M.; Robinson, A. L. Methane Emissions from the Natural Gas Transmission and Storage System in the United States. *Environ. Sci. Technol.* **2015**, *49*, 9374–9383.
- (24) Caulton, D. R.; Shepson, P. B.; Santoro, R. L.; Sparks, J. P.; Howarth, R. W.; Ingraffea, A. R.; Cambaliza, M. O. L.; Sweeney, C.; Karion, A.; Davis, K. J.; Stirm, B. H.; Montzka, S. A.; Miller, B. R. Toward a Better Understanding and Quantification of Methane Emissions from Shale Gas Development. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 6237–6242.
- (25) Marchese, A. J.; Vaughn, T. L.; Zimmerle, D. J.; Martinez, D. M.; Williams, L. L.; Robinson, A. L.; Mitchell, A. L.; Subramanian, R.; Tkacik, D. S.; Roscioli, J. R.; Herndon, S. C. Methane Emissions from United States Natural Gas Gathering and Processing. *Environ. Sci. Technol.* **2015**, *49*, 10718–10727.
- (26) EPA. Greenhouse Gas Reporting Rule: Revisions and Confidentiality Determinations for Petroleum and Natural Gas Systems. 2023, [https://www.federalregister.gov/documents/2023/08/01/2023-14338/greenhouse-gas-reporting-rule-revisions-and-](https://www.federalregister.gov/documents/2023/08/01/2023-14338/greenhouse-gas-reporting-rule-revisions-and-confidentiality-determinations-for-petroleum-and-natural)
- [confidentiality-determinations-for-petroleum-and-natural](https://www.federalregister.gov/documents/2023/08/01/2023-14338/greenhouse-gas-reporting-rule-revisions-and-confidentiality-determinations-for-petroleum-and-natural) (accessed Sept 5, 2023).
- (27) EPA. Standards of Performance for New, Reconstructed, and Modified Sources and Emissions Guidelines for Existing Sources: Oil and Natural Gas Sector Climate Review. 2023, <https://www.federalregister.gov/documents/2022/12/06/2022-24675/standards-of-performance-for-new-reconstructed-and-modified-sources-and-emissions-guidelines-for> (accessed Sept 5, 2023).
- (28) Karion, A.; Sweeney, C.; Pétron, G.; Frost, G.; Michael Hardesty, R.; Kofler, J.; Miller, B.; Newberger, T.; Wolter, S.; Banta, R.; Brewer, A.; Dlugokencky, E.; Lang, P.; Montzka, S.; Schnell, R.; Tans, P.; Trainer, M.; Zamora, R.; Conley, S. Methane Emissions Estimate from Airborne Measurements over a Western United States Natural Gas Field. *Geophys. Res. Lett.* **2013**, *40*, 4393–4397.
- (29) Peischl, J.; Ryerson, T.; Aikin, K.; de Gouw, J. A.; Gilman, J.; Holloway, J.; Lerner, B.; Nadkarni, R.; Neuman, J.; Nowak, J.; Trainer, M.; warneke, C.; Parrish, D. Quantifying Atmospheric Methane Emissions from the Haynesville, Fayetteville, and Northeastern Marcellus Shale Gas Production Regions: CH₄ Emissions from Shale Gas Production. *J. Geophys. Res. Atmos.* **2015**, *120*, 2119–2139.
- (30) Schwietzke, S.; Pétron, G.; Conley, S.; Pickering, C.; Mielke-Maday, I.; Dlugokencky, E. J.; Tans, P. P.; Vaughn, T.; Bell, C.; Zimmerle, D.; Wolter, S.; King, C. W.; White, A. B.; Coleman, T.; Bianco, L.; Schnell, R. C. Improved Mechanistic Understanding of Natural Gas Methane Emissions from Spatially Resolved Aircraft Measurements. *Environ. Sci. Technol.* **2017**, *51*, 7286–7294.
- (31) Karion, A.; Sweeney, C.; Kort, E. A.; Shepson, P. B.; Brewer, A.; Cambaliza, M.; Conley, S. A.; Davis, K.; Deng, A.; Hardesty, M.; Herndon, S. C.; Lauvaux, T.; Lavoie, T.; Lyon, D.; Newberger, T.; Pétron, G.; Rella, C.; Smith, M.; Wolter, S.; Yacovitch, T. I.; Tans, P. Aircraft-Based Estimate of Total Methane Emissions from the Barnett Shale Region. *Environ. Sci. Technol.* **2015**, *49*, 8124–8131.
- (32) Dubey, L.; Cooper, J.; Staffell, I.; Hawkes, A.; Balcombe, P. Comparing Satellite Methane Measurements to Inventory Estimates: A Canadian Case Study. *Atmos. Environ.: X* **2023**, *17*, 100198.
- (33) 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. *Elementa: Science of the Anthropocene* **2019**, *7*, 37.
- (34) 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.
- (35) Siebenaler, S. P.; Janka, A. M.; Lyon, D.; Edlebeck, J. P.; Nowlan, A. E. Methane Detectors Challenge: Low-Cost Continuous Emissions Monitoring. *Proceedings of the 2016 11th International Pipeline Conference Vol. 3; Operations, Monitoring and Maintenance; Materials and Joining*; ASME, 2016; p V003T04A013.
- (36) Gardiner, T.; Helmore, J.; Innocenti, F.; Robinson, R. Field Validation of Remote Sensing Methane Emission Measurements. *Rem. Sens.* **2017**, *9*, 956.
- (37) Titchener, J.; Millington-Smith, D.; Goldsack, C.; Harrison, G.; Dunning, A.; Ai, X.; Reed, M. Single Photon Lidar Gas Imagers for Practical and Widespread Continuous Methane Monitoring. *Appl. Energy* **2022**, *306*, 118086.
- (38) Bell, C.; Zimmerle, D. METEC Controlled Test Protocol: Continuous Monitoring Emission Detection and Quantification. *METEC Controlled Test Protocol*; Colorado State University, 2020.
- (39) Bell, C.; Ilonze, C.; Duggan, A.; Zimmerle, D. Performance of Continuous Emission Monitoring Solutions under a Single-Blind Controlled Testing Protocol. *Environ. Sci. Technol.* **2023**, *57*, 5794–5805.
- (40) 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. *Environ. Sci. Technol.* **2020**, *54*, 11506–11514.
- (41) EPA. Standards of Performance for New, Reconstructed, and Modified Sources and Emissions Guidelines for Existing Sources: Oil and Natural Gas Sector Climate Review. 2024, <https://www.epa.gov/>

system/files/documents/2023-12/eo12866_oil-and-gas-nsps-eg-climate-review-2060-av16-final-rule-20231130.pdf (accessed Jan 18, 2024).

(42) 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.

(43) Zimmerle, D.; Vaughn, T.; Luck, B.; Lauderdale, T.; Keen, K.; Harrison, M.; Marchese, A.; Williams, L.; Allen, D. Methane Emissions from Gathering Compressor Stations in the U.S. *Environ. Sci. Technol.* **2020**, *54*, 7552–7561.

(44) Allen, D. T.; Torres, V. M.; Thomas, J.; Sullivan, D. W.; Harrison, M.; Hender, 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. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 17768–17773.