Performance of Continuous Emission Monitoring Solutions under a Single-Blind Controlled Testing Protocol

Clay Bell, Chiemezie Ilonze, Aidan Duggan, and Daniel Zimmerle*

ABSTRACT: Continuous emission monitoring (CM) solutions promise to detect large fugitive methane emissions in natural gas infrastructure sooner than traditional leak surveys, and quantification by CM solutions has been proposed as the foundation of measurement-based inventories. This study performed single-blind testing at a controlled release facility (release from 0.4 to 6400 g CH_4/h) replicating conditions that were challenging, but less complex than typical field conditions. Eleven solutions were tested, including point sensor networks and scanning/imaging solutions. Results indicated a 90% probability of detection (POD) of 3−30 kg CH_4/h; 6 of 11 solutions achieved a POD < 6 kg CH_4/h, although uncertainty was high. Four had true positive rates > 50%. False positive rates ranged from 0 to 79%. Six solutions estimated emission rates. For a release rate of 0.1−1 kg/h, the solutions' mean relative errors ranged from −44% to +586% with single estimates between −97% and +2077%, and 4 solutions' upper uncertainty exceeding +900%. Above 1 kg/h, mean relative error was −40% to +93%, with two solutions within ±20%, and single-estimate relative errors were from −82% to +448%. The large variability in performance between CM solutions, coupled with highly uncertain detection, detection limit, and quantification results, indicates that the performance of individual CM solutions should be well understood before relying on results for internal emissions mitigation programs or regulatory reporting.

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

INTRODUCTION

Anthropogenic methane, a potent greenhouse gas (GHG)[1] with an estimated 20 year global warming potential 84−87 times that of CO_2,[2] is emitted from oil and gas facilities, waste management, agricultural processes, and land-use practices. Curbing methane emissions across all sectors is crucial to near-term climate change mitigation. While the importance of mitigating methane emissions is well recognized, measuring these emissions from natural gas facilities remains a challenge, due to the vast extent of natural gas infrastructure and the temporal variability of emission sources.[3−8]

Recent private and public investments have driven improved methane emission measurements. Some efforts have coalesced into multistakeholder initiatives to develop comprehensive and transparent methane emissions and source documentation,[9−12] while others have supported development of new *leak detection and quantification* (LDAQ) solutions.[13] Continuous emissions monitoring (CM) solutions—stationary sensor platforms installed at or near facilities to monitor concentration nearly continuously and process sensed concentrations to detect and quantify emissions—have recently received attention as an improved way of understanding facility emissions. As deployed, each solution consists of one or more sensors installed at the test site in a semipermanent fashion, proprietary algorithms (data analytics) to analyze sensed data, and communications to

transfer sensed data to the analytics platform. For many solutions, data analytics may be assisted by human operators.

Key advantages of CM solutions include providing temporally resolved data, potentially quantifying emissions over extended periods, and detecting large fugitive sources sooner than they otherwise would be detected. However, the performance of CM solutions is poorly understood, including their detection limits, quantification accuracy, and temporal resolution, all of which are impacted by meteorological conditions, sensor placement, and other factors. A better understanding of the performance of CM solutions is required before using data from them to enhance or replace existing emission inventories, particularly if the resulting inventory will be used as part of alternative compliance programs or as the basis for enhanced leak detection and repair (LDAR) programs.[14]

Prior testing of LDAQ solutions has focused on survey techniques, notably the Stanford/EDF Mobile Monitoring Challenge,[15] evaluations of mobile methodologies such as

OTM-33A,[16] and remote sensing using aerial methods.[17,18] Generally, periodic survey methods sense emissions over short time periods with human supervision to adjust the placement and operation of the sensors as needed. In contrast, CM solutions cannot be adequately tested using the same testing methodology as used for periodic survey techniques. CM solutions are expected to operate largely unsupervised from a stationary location in a facility and are therefore more impacted by sensor placement, meteorological conditions, and other factors.

Although some testing of CM solutions has been performed, it is limited in scope, and few peer reviewed publications exist. Testing during the ARPA-E MONITOR program[13] included several CM solutions, but due to small sample sizes, performance of individual solutions was difficult to assess.[19] Other testing, including the Methane Detectors Challenge[20] and tests by the UK National Physical Laboratory,[21,22] suffers from similar limitations: small sample sizes and limited experimental complexity. These studies typically used a single isolated emission source, a limited range of emission rates and environmental conditions, and testing methods that were partially blind (e.g., known emission timing and location, unknown emission rate) instead of fully blind.

This study presents the first testing of CM solutions using a standardized testing protocol specifically developed for CM solutions. The test protocol focuses on evaluating key performance metrics including the probability of detection (POD), localization accuracy and precision, and quantification accuracy. Dependence on meteorological conditions was addressed by testing for an extended period (Testing was performed between March and October in 2021 and between February and May in 2022.), during which emission rates, duration, and source locations were varied. Test conditions therefore mimicked, but did not completely replicate, field conditions. The resulting metrics provide standardized data for stakeholders to assess and compare solutions. Stochastic results such as POD curves are also a key input to LDAR simulation software such as FEAST[23] and LDAR-Sim.[24]

■ METHODOLOGY

Testing Protocol. The Advancing Development of Emissions Detection (ADED) project developed the CM test protocol utilized in this study.[25,26] Over 60 entities (operators, solution developers, regulators, NGOs, etc.) contributed to the protocol development during multiple rounds of presentation, feedback, and revision.

The protocol divides testing into a set of discrete *experiments*, as follows: each experiment ran for a predetermined duration and included one or more controlled releases, each operating at a steady emission rate. Experiments with multiple, simultaneous controlled releases were intended to evaluate a solution's ability to discern and attribute emissions to individual source locations under prevailing atmospheric conditions. Experiments were separated by a period with no emissions to allow solutions to identify the start and end of each controlled release event by recognizing a return to background levels for an extended period (hours). The no-emission period between emitters represents a substantial simplification from field conditions but facilitates data collection and evaluation for testing purposes. All experiments were performed single-blind: solutions were unaware of the timing of experiments or the number, exact location, or emission rate of controlled releases.

Data analytics in CM solutions transform raw sensor measurements (e.g., ambient ppm readings, wind speed, and wind direction) into composite data that are more informative and useful to operators, such as if, where, when, and at what rate, emissions occurred on the facility. Therefore, to evaluate CM solution performance, it is critical to assess the integrated solution rather than the performance of individual solution components. For example, the protocol *does not* assess the performance of the solutions' sensor readings—i.e. the concentration (ppm), path-integrated concentration (ppm-m), camera images, or meteorological data produced by the sensors installed at the test facility. Rather, performance is assessed based upon inferences made by the solution analytics, which may or may not include elements of human review or assistance. Each performance metric is briefly described below; more details are in the protocol.[26]

Probability of Detection. The protocol required solutions to report *detections* based upon data analytics results. The reporting method is detailed in section 5.4.1 of the protocol.[26] In brief, each solution submitted detection reports attributing observed emissions to a unique *EmissionSourceID*. Reports may indicate a new emission source or an update to a previously reported source. A matching procedure (section 6.1 of the protocol) was implemented by the test center to pair reported detection data with controlled release data and to classify each as

- True Positive (TP) - a controlled release and reported detection which were paired.
- False Negative (FN) - a controlled release which remained unpaired
- False Positive (FP) - a reported detection which remained unpaired.

The probability of detection was calculated as the fraction of controlled releases which were classified as TP under a particular set of releases and environmental conditions.

Localization Precision. Detection reports were required to include, at minimum, an estimate of the equipment unit where the emission occurred. Using this data, the localization precision of each TP detection was classified as *equipment unit*, *equipment group*, or *facility* level detection. An *equipment unit* detection correctly attributed an emission source to the major equipment unit (i.e., a specific wellhead, separator, or tank) where a controlled release occurred. An *equipment group* detection correctly attributed an emission to the group of adjacent equipment units where a controlled release occurred but not to the correct unit in that group—for example, reporting the third wellhead in a group, at a time when there was a controlled release on the second wellhead in the same group. A *facility* detection was assigned when a solution correctly reported a source during a controlled release, but the reported location is not within the same equipment group as the controlled release.

Localization Accuracy. Detection reports could include localization data in the form of a GPS coordinate or 2 sets of coordinates which mark the corners of a "bounding box" within which the solution estimates the presence of a controlled release. These data allowed additional localization accuracy and precision metrics to be assessed.

Quantification Accuracy. The protocol required solutions to report the gas species their system measured to perform a detection and, if possible, estimate the emission rate of the source found in g/h of the gas species. This allows the metrics to account for the mix of gas species sensed by each solution's sensors. Some sensors were methane-specific, while some

respond to total hydrocarbons or similar composites of multiple species. The compressed natural gas (CNG) releases used for testing typically contained a mix of methane and ethane. When provided, the test center evaluated the accuracy of the emission estimate for each TP detection and the systematic bias across all TP detections.

Downtime. Since multiple solutions were installed, testing was not stopped if any individual solution was offline. The protocol enabled solutions to report their system's offline periods (section 5.4.2 of the protocol). The test center removed controlled releases during these reported periods from the analysis, avoiding spurious FN detections.

Facility Downtime. Similarly, the test center also removed controlled releases and reported detections that occurred during facility maintenance periods or when a controlled release was noncompliant with the ADED protocol (e.g., when the test center supported other testing at the facility during the period when solutions were installed).

Solutions. A total of 11 CM solutions participated in this study, developed by the following companies, in alphabetical order: Baker Hughes, CleanConnect.ai, Earthview.io, Honeywell, Kuva Systems, Luxmux Technology, Pergam Technical Services, Project Canary, QLM Technology, Qube Technologies, and Sensirion. Due to confidentiality agreements, results are arbitrarily identified in this analysis by an anonymized letter identifier. Tested capabilities and configuration for all participating solutions are listed in Table 1. Solution vendors

Table 1. Characteristics of Participating Solutions

	sensor		reported data[a]		
ID	type	count	detection	quantification	GPS local[b]
A	point sensor network	8	X	X	X
B	scanning/imaging	1	X	X	X
C	point sensor network	6	X	X	X
D	point sensor network	8	X	X	X
E	point sensor network	16	X	X	X
F	point sensor network	8	X	X	NA
G	scanning/imaging	1	X	X	X
H	scanning/imaging	1	X	NA	X
I	scanning/imaging	1	X	NA	NA
J	scanning/imaging	2	X	NA	NA
K	point sensor network	7	X	NA	X

[a]'X' indicates the parameter of interest was reported by the solution. "NA" indicates that it was not reported. [b]The 'GPS local' column indicates if the solution localized emitters by GPS coordinates.

were instructed to deploy their systems in a manner that reflected *typical* field deployment. The number of sensors deployed and locations were selected by the vendors. Some vendors deployed their solution to monitor a portion of the designated testing area, in which case only controlled releases from the specified equipment group(s) were considered in their performance analysis. See SI Section S-3 for more descriptions on solutions' deployment.

The definition of solution types in Table 1 is

- *Point sensor network:* Solution deployed one or more concentration sensors that each sense either methane or hydrocarbons at one point. Analytics combine concentration time series with meteorological data to develop detections.

- *Scanning/imaging:* Solution deployed one or more sensors which produce 2-dimensional images of gas plumes on the surveilled locations. In this study, sensors included scanning lasers (typically LIDAR-type) or short- or midwave infrared cameras that used ambient illumination. Analytics combine images (typically a video sequence) with meteorological data to develop detections.

Testing Process. Testing was conducted during two separate campaigns in 2021 and 2022 at the Methane Emissions Technology Evaluation Center (METEC), an 8-acre outdoor laboratory located at Colorado State University (CSU) in Fort Collins, Colorado, U.S.A. The facility was designed to mimic and simulate a wide-range of emission scenarios associated with upstream and midstream natural gas operations. The facility was built using surface equipment donated from oil and gas operators. A controlled release system allowed metering and control of gas releases at realistic sources such as vents, flanges, fittings, valves, and pressure relief devices found throughout equipment. See Zimmerle et al.[27] and SI Section S-1 for a more comprehensive description of the facility.

The study team designed and scheduled experiments daily during each study period. Each experiment was considered for all solutions installed at that time. Controlled release rates and experiment durations were selected considering facility constraints and the expected detection limits communicated by vendors. The study team reviewed performance of the solutions as testing progressed to inform selection of release rates and durations for subsequent experiments to "fill in" regions where data had a low sample count. For example, if the study team identified that solutions had not yet reached 90% detection rates, then experiments with higher emission rates were integrated into the test schedule.

Testing Constraints. Several operational constraints exist at the test facility. METEC was initially developed to evaluate leak detection systems in the ARPA-E MONITOR program,[13] targeting relatively low emission rates observed from fugitive component leaks in field studies. As noted above, emission rates were varied to sweep, as best possible, across the full range of every solution's probability of detection (POD) curve. In general, this required larger gas releases and gas supply than was available from the installed capacity at METEC. Therefore, while the first test period used the CNG supply at METEC, a CNG trailer with larger capacity was connected to the METEC gas supply system during the 2022 testing period to support longer and larger releases, and more continuous operation.

Additionally, since many CM solutions rely on variability in the emission transport to localize and quantify emissions, each experiment should be of sufficient duration to allow the solutions ample monitoring time and variation in environmental conditions. A large number of experiments were also required to evaluate the repeatability of detection and quantification. These considerations together necessitated a test program lasting several months. In this study, the duration of each experiment was constrained to a maximum of approximately 8 h, which may impact the performance of some solutions which rely on data collected over long time frames (e.g., days) to detect, localize, and quantify emission sources. However, the practicality of the longer time frame approach at operational facilities is questionable since many emission sources are intermittent or unsteady.

One advantage of the long duration test program was the inherent variability in environmental conditions during the

Table 2. Summary of the Localization Precision and Classification of Controlled Releases and Detection Reports of Participating Solutions[b]

	count		number of TP Localization					
ID	controlled releases	detection reports	equipment unit	equipment group	facility	TP (%)[a]	FN (%)	FP (%)
E	567	2382	232	207	58	87.7	12.3	79.1
F	571	469	98	200	100	69.7	30.3	15.1
A	571	834	111	156	129	69.4	30.6	52.5
D	571	346	0	177	158	58.7	41.3	3.2
B	442	213	122	26	23	38.7	61.3	19.7
C	557	214	2	1	191	34.8	65.2	9.3
J	284	68	67	0	1	23.9	76.1	0.0
H	368	37	3	17	14	9.2	90.8	8.1
I	354	31	17	4	7	7.9	92.1	9.7
G	206	12	0	2	6	3.9	96.1	33.3
K	746	2	0	1	1	0.3	99.7	0.0

[a]TP (%) shown in this table is the percentage of controlled releases detected by a solution across all localization levels (equipment unit, group, and facility). [b]Solutions are sorted in the order of the declining true positive (TP) detection rate.

program. Testing was conducted in all weather conditions encountered, and where possible, the influence of wind speed and other meteorological parameters was investigated in the analysis. However, in some cases during the winter season, experiments were either canceled due to limited access to the test facility, or test results were discarded during quality control due to operation at temperatures below the flow meter specifications. The meteorological conditions under which each solution was tested have been summarized in their performance report in the SI.

RESULTS AND DISCUSSION

The first section below discusses each class of major results using illustrative examples for a subset of solutions and comparisons between solutions where appropriate. The following sections provide an overview of all solutions, implications of the results, and a discussion of the challenges uncovered during testing. Refer to the SI for detailed results for each solution.

Primary Results and Analysis. *Classification of Detections.* Detection reports and controlled releases were classified as TP or FP and as TP or FN, respectively, as described earlier. Table 2 summarizes the classified detection reports (FP rates) and controlled releases (TP and FN rates) for all participating solutions, sorted in order of decreasing TP rates. Although many of the solutions participated in testing simultaneously, the number and characteristics of controlled releases included in their performance analysis vary due to solutions installing after the program start, uninstalling prior to the program end, or submitting offline reports during the program. Therefore, it is important to understand the underlying distribution of controlled releases for each solution when interpreting the classification results. A summary of the test periods for each solution is in SI Table S-2, with more details in the performance reports.

Table 2 clearly illustrates the wide range of performance for CM solutions—ranging from near-zero TP to TP rates in an excess of $2/3^{rd}$ of all controlled releases, accompanied by FP rates from zero to over half of all detection alerts. This level of variability clearly indicates the need to set performance standards before qualifying solutions for LDAR deployments or regulatory reporting. The table also illustrates the trade-off between detection sensitivity and false positive rates. Of the 4 solutions with TP rates over 50%, two had FP rates exceeding 50%. In field conditions, a high FP rate may force unacceptably

high follow-up costs. In that case, changing solutions' settings to drop the FP rates may also reduce TP rates. In contrast, the 6 solutions with FP rates below 10% tended to also have lower TP rates—3 had TP rates below 10%, and 3 had rates of 24−59%.

Combining with Table 1 indicates that solutions that deployed more sensors were likely to have higher TP rates: 5 of the 7 solutions with the highest TP rates deployed 6 or more sensors, and 3 of 4 solutions with the lowest TP rates installed only 1 sensor. However, considering one metric individually does not reflect practical performance, individual solutions may be designed for different types of monitoring. For example, one solution may be designed to monitor a specific area of a site for large emitters—a common mode for imaging solutions—while another is designed for full-facility monitoring and quantification. As indicated earlier, a majority of test releases may be above/below the expected sensitivity of any given solution. Assuming the solution is properly deployed for its intended purpose, true performance is typically judged by the balance between TP and FP/FN rates. For example, 2 solutions in the top 4 TP rates had FP rates above 50%, which could be lead to unacceptably frequent follow-up in field deployments.

Additionally, ongoing field work indicates solutions deploy different numbers of sensors—typically fewer—in field deployments than were deployed at the test center and often fewer sensors per unit area. This type of change may have a substantial impact on field performance.

Probability of Detection. A POD curve or surface is a key metric required to model the emission mitigation potential of solutions using tools like FEAST or LDAR-Sim. The POD describes the probability that an emission source will be detected by a solution as a function of many independent parameters including characteristics of the emission source itself (e.g., the emission rate, source type, position, etc.) and environmental conditions (e.g., wind speed and direction, precipitation, etc.).

The POD curves for two point sensor network solutions (A and F - upper panels) and two scanning/imaging solutions (I and J - lower panels) are shown in Figure 1. Logistic regression was used to develop the POD curve, and the regression was bootstrapped to show a cloud of curves providing a visual indication of confidence. In the figure and in summary Table 3, we define a solution's method detection limit (MDL) as the emission rate where the integrated solution—i.e. the *method*—achieves 90% POD. Solutions or other sources may alternatively
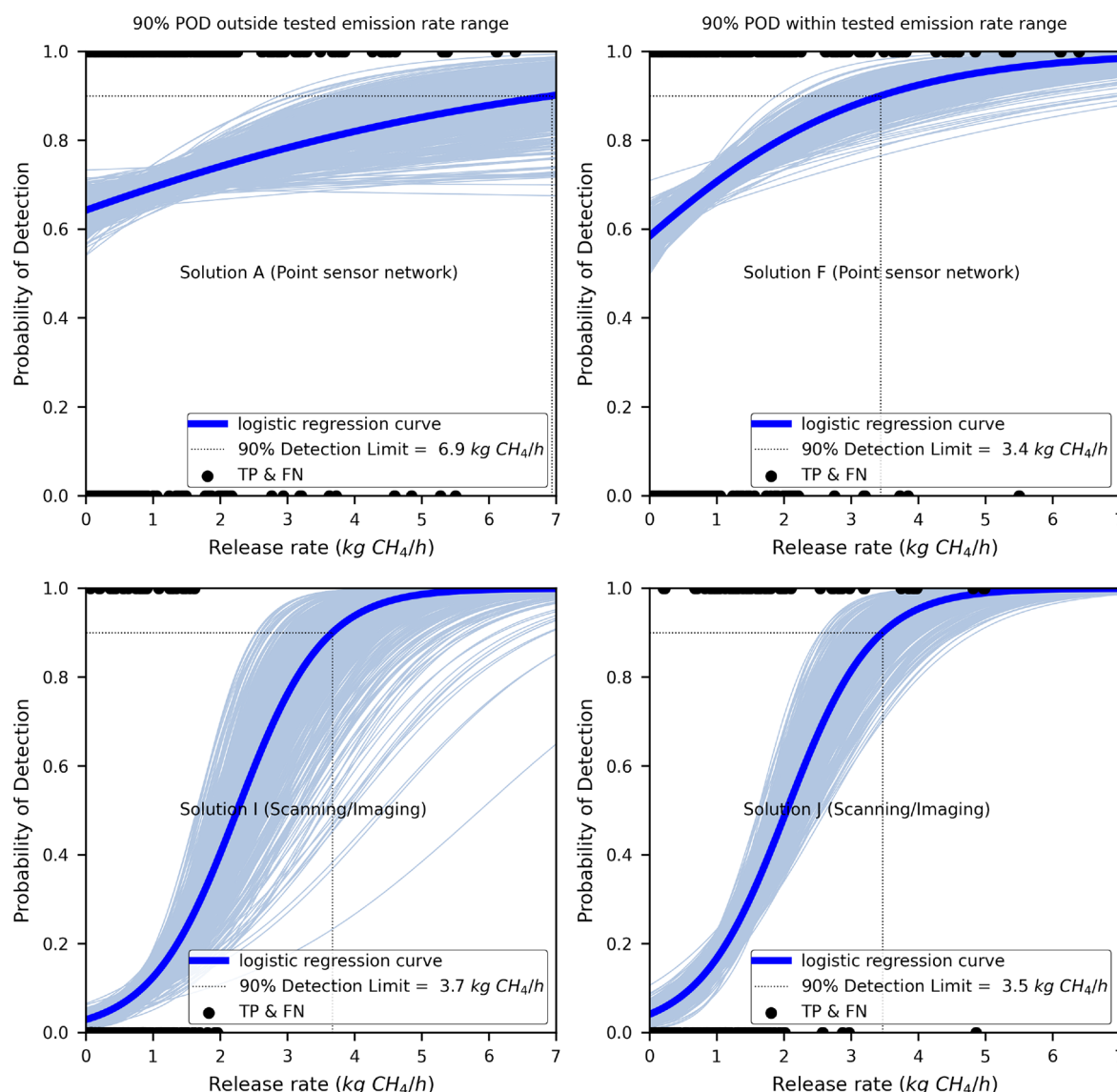
Figure 1. Probability of detection versus emission rate (kg CH_4/h) fit using logistic regression. Solutions A and F (upper two panels) are point sensor networks, while solutions I and J (lower two panels) are scanning/imaging solutions. True positive and false negative controlled releases are shown with markers at y = 1 and y = 0, respectively. The regression is bootstrapped to produce a cloud of curves illustrating uncertainty in the result. The emission rate at which the POD reaches 90% is indicated as the method detection limit for each solution. For solutions A and I, left side, test releases did not exceed the computed 90% detection limit, while for solutions F and J, right side, release did exceed that rate. As a result, POD uncertainty is substantially larger for solutions A and I.

quote a *lower detection limit* (LDL), typically below the MDL, where there is a nonzero probability of detection. Internal reporting or regulatory programs typically require a comprehensive and rigorous MDL metric; here we suggest that metric be defined as the emission rate where the solution will detect nine of ten emission sources across a wide range of meteorological conditions. See SI Section S-8.2 about the details of the logistic regression model.

As indicated by the protocol, the test center should attempt to map the entire POD curve/surface during the testing period. Since many solutions' POD are dependent on the release rate, the test center attempted to span release rates from low rate/low POD to high rate/near 100% POD. This was not always possible and had an impact on the uncertainty of POD results. For point sensor network solution F, where the MDL lies within the range of emission rates, the MDL (3.4 [2.4, 5.4] kg CH_4/h) indicates tighter confidence bounds than for solution A (6.9 [3.6, 29.2] kg

CH_4/h), where the MDL falls outside of the range of controlled release rates tested. The two scanning/imaging solutions (I and J) have comparable MDLs. However, the MDL of solution J (3.5 [2.8, 4.4] kg CH_4/h) falls within the range of tested emission rates, while it falls outside the range for solution I (3.7 [2.7, 5.7] kg CH_4/h), resulting in tighter uncertainty bounds for solution J.

For 3 solutions, the MDLs could not be estimated with logistic regression due to limited distribution of TPs compared to FNs across the range of the independent variable (i.e., controlled emission rate). The remaining 8 solutions had a best estimate MDL range of 2.7−30.1 kg CH_4/h with 6 of the 8 solutions within 2.7−5.9 kg CH_4/h. All but 1 point sensor network solution had an MDL range of 2.7−6.9 CH_4/h. The 3 scanning/imaging solutions considered in the analysis had an MDL range of 3.5−30.1 CH_4/h. SI Section S-6.1 shows similar results when the emission rate is normalized by mean wind speed. See the

Table 3. Summary of Single-Estimate Detection and Quantification for All Participating Solutions Using 95% Empirical Confidence Limits from Bootstrapping to Assess Uncertainty on the MDL Best Estimate

	detection[e]		relative quantification error (%)[d]					
	detection limit (MDL) (kg CH₄/h)		(0.1−1] (kg/h)			>1 (kg/h)		
ID	best est	95% CL	mean	median	95% CL	mean	median	95% CL
A[a]	6.9	[3.6, 29.2]	211.3	134.2	[−60.9, 946.8]	27.1	−24.2	[−85.6, 338.5]
B[a]	30.1	[0.0, NA]	74.6	39.5	[−81.1, 343.2]	41.9	24.4	[−90.2, 268.8]
C	5.9	[4.5, 7.9]	268.9	76.2	[−73.9, 1875.0]	18.2	−29.3	[−88.6, 369.3]
D	5.5	[3.8, 10.2]	−43.6	−60.1	[−92.6, 141.4]	−39.5	−76.8	[−99.9, 242.4]
E[c]	2.7	[0.0, 17.3]	586.2	411.0	[−96.7, 2078.7]	92.2	50.1	[−99.1, 448.3]
F	3.4	[2.4, 5.4]	202.2	110.9	[−39.7, 933.2]	10.2	−40.4	[−82.5, 373.6]
G[b]	NA	[0.0, NA]	NA	NA	NA	NA	NA	NA
H[b]	NA	[0.0, NA]	NA	NA	NA	NA	NA	NA
I[a]	3.7	[2.7, 5.7]	NA	NA	NA	NA	NA	NA
J[c]	3.5	[2.8, 4.4]	NA	NA	NA	NA	NA	NA
K[b]	NA	[0.0, NA]	NA	NA	NA	NA	NA	NA

[a]The MDL best estimate is above the range of controlled releases rates tested. [b]The MDL could not be evaluated by the logistic regression model. [c]Solutions reported whole gas as the gas species measured to perform detection and emissions estimation (if possible), while other solutions reported methane gas. [d]Solutions H to K did not report quantification estimates. Solution G was excluded due to data quality issues. [e]When the logistic regression bootstrapping could not evaluate the lower and upper empirical Confidence Limit (CL) on the best estimate of a solution's MDL, they are given as 0 and NA, respectively.
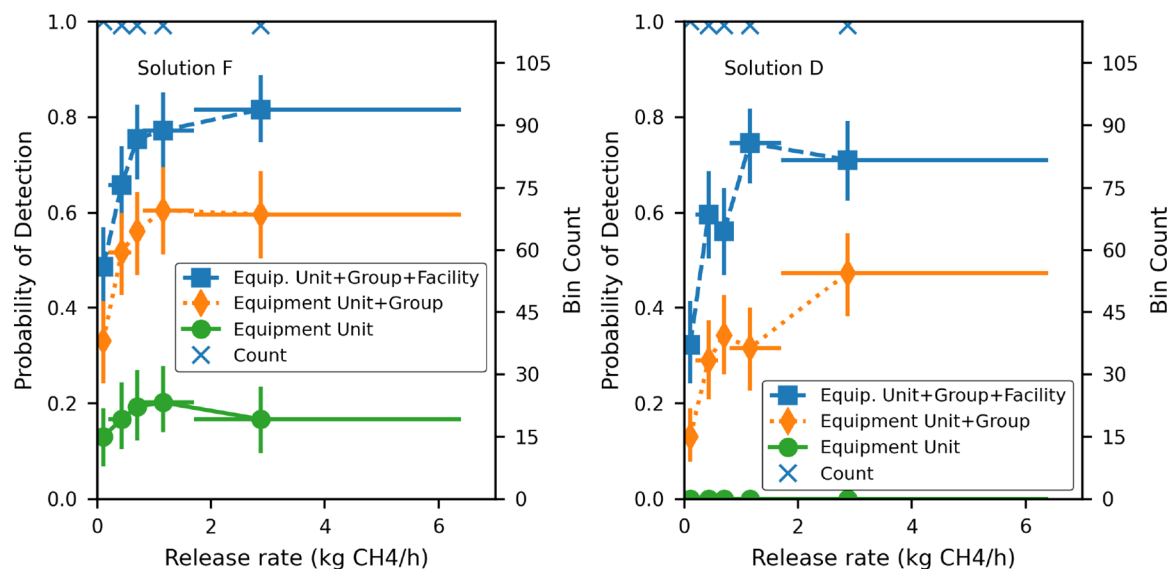

Figure 2. Probability of detection versus release rate (kg CH₄/h) for two example solutions−D and F, binned by reported localization. Separate curves are illustrated for equipment unit, equipment group, and facility-level localization. Categories are cumulative (See text.). Markers represent the mean emission rate and observed probability of detection within each bin. X whiskers indicate maximum and minimum emission rates in each bin. Y whiskers indicate maximum and minimum probability of detection when empirical data is bootstrapped. The number of data points within each bin is plotted using x markers against the right-hand axis.

performance report of each solution to contextualize detection results.

The impact of mean windspeed and release duration were also investigated; see SI Section S-6.1. Considering the 8 solutions with POD curves, MDL may shift upward or downward due to these variables. The impact of release duration on POD over the tested gamut was a POD change within the range of −37% to +20%, while the impact of wind speed was wider, −72% to +4%.

Localization. Table 2 also illustrates the equipment-level localization estimates of solutions. Localization is often dictated by the algorithm implemented by a solution. Some solutions may function only at the "full facility" level and implement no capability to localize within the facility. Others may prioritize localization and can provide specific locations for emission

sources. Four of 11 solutions attributed the majority (46.7% to 98.5%) of TP detections at the equipment unit level, 3 localized a plurality (39.4% to 50.3%) at the equipment group level, while 3 solutions localized most (75% to 100%) TP detections at the facility level. The accuracy and precision of GPS source localization estimates were also assessed if reported by a solution (See SI Section S-6.2.).

Combining POD with Localization. Figure 2 combines equipment-level localization results with POD, utilizing an alternative POD approach where data are binned by the emission rate and the fraction of TP detections in each bin estimates the POD. Vertical error bars are 95% bootstrap confidence intervals (See SI Section S-8.1 on bootstrapping methodology.). Equipment localization was included in the

figure by grouping results by three sets of markers: equipment unit detections, equipment group detections or better, and facility level detections or better.

The logistic regression in Figure 1 and per-solution reports in the SI considered all detections. At any level of localization, the equipment unit and equipment group curves in Figure 2 represent the more precise localization. The example solution F attributed less than ≈20% of TP detections to the correct equipment unit and ≈60% at both the correct equipment unit and group levels combined. Solution D localized sources at the equipment group and facility levels. Although solution D did not provide source attribution below the equipment group level (i.e., no equipment unit), the localization accuracy metric (See performance report in the SI.) shows that 43 of 335 TP detections reported GPS coordinates within 1 m of the actual source. The result for solution F illustrates that, while the POD improved as the release rate increased, relative localization precision did not; approximately the same fraction of TP detections falls in each localization bin.

Localization capabilities may have a substantial impact on the cost of deploying any CM solution. For large facilities—likely larger than the test facility—detections localized to a restricted subset of the facility may substantially reduce follow-up deployment costs.

Quantification. Recent works focusing on the certification of natural gas production have raised interest in the use of CM solutions to provide near-continuous quantification of emissions at facilities.[9−12] For these to be effective, quantification accuracy must be clearly understood. Six solutions reported quantification estimates: 5 reported methane and solution E reported whole gas emissions rate.

Figure 3 shows an example of the reported emission estimate compared to the controlled release rate for TP detections for a solution. The upper panel shows a zero-intercept linear regression to illustrate bias observed across all observations. However, this result should be used with caution for three reasons: First, residuals are strongly heteroskedastic (nonuniform variance of the residual) for most solutions (middle panel), which makes the regression highly uncertain. Second, the emission rates in the controlled releases are unlikely to reflect the mix of emissions at any field location. Finally, many decisions dependent on emission estimates utilize a much smaller number of estimates than used in the fit. In other words, operators only require a small number of observations of emissions events to make decisions on whether to initiate an action.

The dependence of quantification accuracy on the controlled release emission rate is a key variable to understand field performance and is typically the parameter of interest in modeling and regulatory program analysis. However, the emission rate is only one of several independent variables which may have appreciable impact on the quantification performance of a solution. Wind speed and emission duration may have similar impacts; see SI reports for additional analysis. To illustrate another variable, marker color indicates another independent parameter. In the example figure, longer release duration (marker color) did not improve quantification estimates. A similar observation can be made for all other solutions.

The box and whisker plot (lower panel of Figure 3) shows errors when data is binned by the order of magnitude of the controlled release rate. In this example, the mean and median relative errors for all bins were substantially skewed. For instance, the mean error of controlled releases of (0.01−0.1] kg/
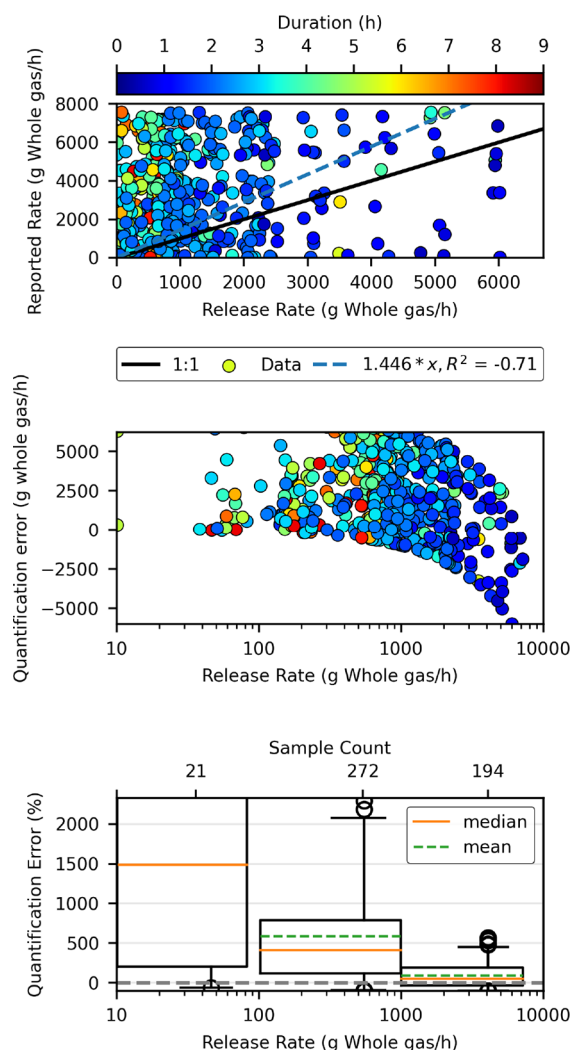

Figure 3. Example of quantification accuracy, using TP detections for solution E as an example. Upper panel: Reported emission rate versus controlled release rate, with markers colored by controlled release duration. This panel indicates bias for a set of the release rates selected for this study but may indicate bias for a different selection of release rates. Center panel: Error in emission rate estimates, i.e. expected error for any individual result reported by the solution. Lower panel: Relative error binned by the order of magnitude of the controlled release rate. Results indicate relative quantification bias over a large number of reported estimates for each emission rate range. Note the logarithmic *x*-axis on the lower two panels. The *y*-axis was restricted to show 95% of the data on each panel; see the SI reports for full data.

h of whole gas was ≈4× the median error. However, as the release rate increased, quantification errors became less skewed (more symmetric) as the mean approached the median for release rates of >1 kg/h of whole gas. While the mean relative error decreased with increasing emission rates, zero mean error was not observed in this testing.

When evaluating solutions for field deployment and emission reporting, the relative error comparable to the lower panel of Figure 3 is of primary interest. Figure 4 shows the results for all solutions that reported quantification estimates, with the *y*-axis constrained to the same values for all plots (Solution G was excluded due to unsatisfactory data quality.). As with Figure 3, each box summarizes data by order of magnitude of the controlled release rate. Results indicate skewness of quantification relative errors for all solutions across all emission rate bins.
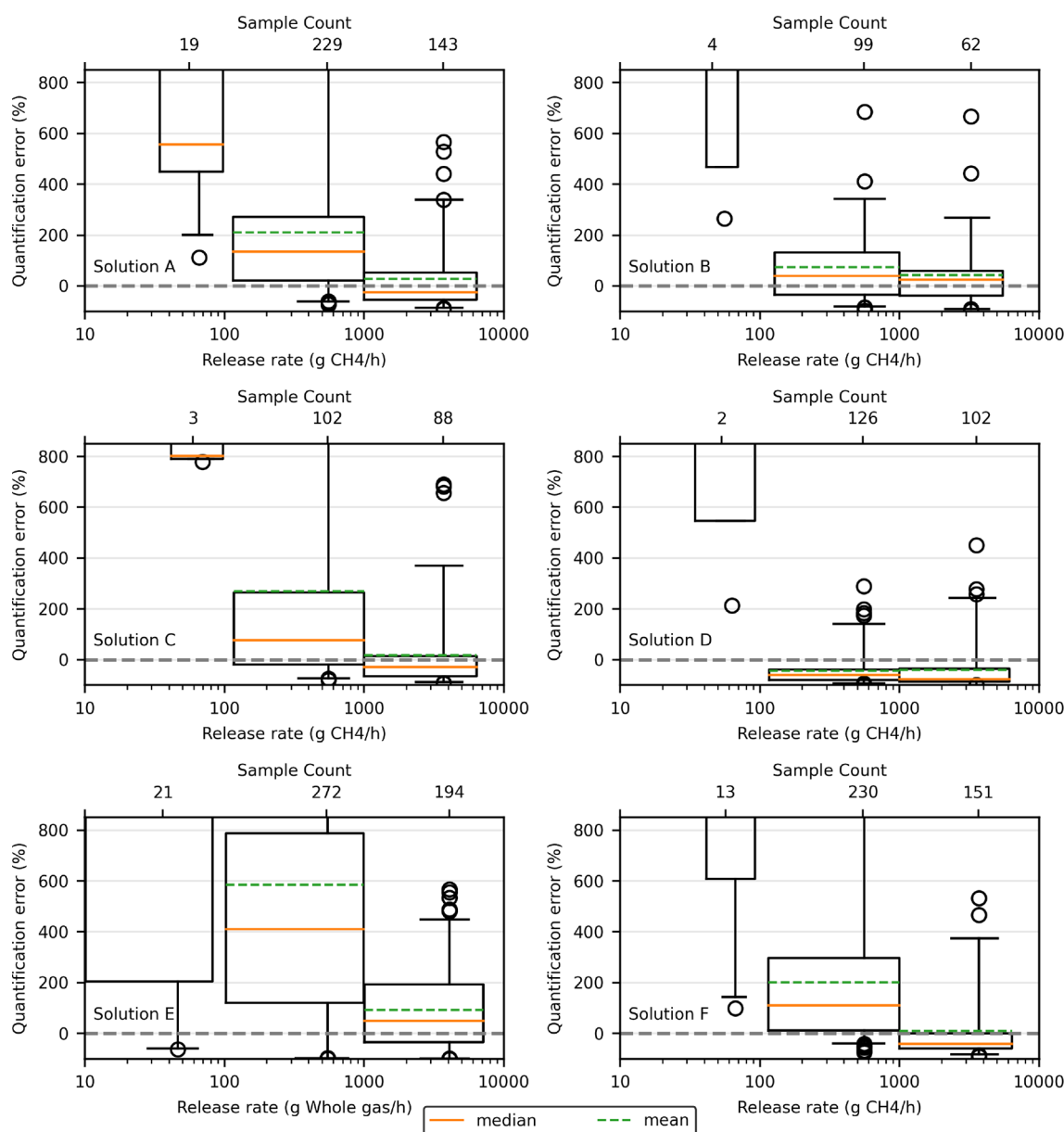
Figure 4. Box plots summarizing the quantification error (%) of emission estimates for all the solutions that reported quantification estimates. Boxes summarize data in each order of magnitude. Difference between mean and median indicate skewness in the data set. The box represents the inner quartile range, whiskers include 95% of the data, and open-circle markers represent outliers. The width of each box represents the minimum to maximum controlled release rate included in each range. The y-axis limit of each panel has been restricted to −100% to 850%, i.e. −1× to +8.5×. The sample count is shown above each box. Solution E reported whole gas emissions. Other solutions reported methane.

When quantifying emissions of (0.01−0.1] kg/h, solutions exhibited mean relative errors of 9× to 60× the release rate (Note the limited sample size in this range.) with individual estimates ranging from −60% to over 130× at the 95% confidence level. Two of the 6 solutions within this range had mean relative errors less than 10× with individual estimates between 5× to 14× the release rate at 95% confidence.

Similarly, for emission rates of (0.1−1] kg/h, the mean relative errors of solutions spanned from −44% to +6× and individual quantification estimates ranging from −97% to +21× (95% CI). Five of the 6 solutions in this range had mean relative errors less than 300% (but greater than −50%) with single estimates ranging from −93% to +19× (95% CI).

At emission rates > 1 kg/h, the mean relative errors of solutions ranged from −40% to +92% with individual estimates

between −1× to +4.5× (95% CI). Two of the 6 solutions in this range had mean relative errors within 20% (±20%). See Table 3 and SI Table S-4 for details.

The two controlled release rate ranges captured in Table 3 roughly correspond to two different types of emitters at upstream facilities. The lower band (0.1−1 kg/h) roughly corresponds to emission rates from component leaks that would be routinely identified and fixed in OGI surveys.[28−30] In this range, no solution reported mean estimates accurate within ±40%. For facilities with no process failures or upsets, or newer facilities with little venting and no atmospheric tanks, nearly all emission sources would likely fall into this emission rate range. Therefore, these data indicate a substantial misestimation of emissions in a critical range of observations.

Individual observations, which are often used to prioritize dispatch after detections, show larger errors than the mean error. Upper 95% confidence exceeds 9 times the emission rate for 4 solutions, and the lower 95% confidence drops below half the true emission rate for all but one solution. Large errors like these could have a substantial impact on the decision to dispatch diagnostic and repair crews, with commensurate impact on cost and emissions mitigation.

The higher band (>1 kg/h) is representative of mid-to-small process upset conditions at upstream facilities.[28,31] While the mean relative errors for solutions in this range are not as tight as expected from typical industrial instrumentation, these uncertainties represent reasonable estimates of larger emitters when observed over many estimates. As shown earlier, individual estimates, however, had a wider confidence interval. The lower confidence limit for all solutions dropped below −80% of the emission rate, calling into question whether a large emitter would be quickly classified as "large" in field conditions. The upper limit exceeded 240% of the emission rate; a large emitter estimated only a few times before mitigation could be substantially overestimated.

In general, quantification estimation by solutions improved with increasing controlled emission rate. Also, with the exception of solution D, which was more likely to underestimate emissions, solutions were more likely to overestimate emissions across all emission rate bins. A figure similar to Figure 3 but for absolute quantification error is in SI Section S-6.3.

Performance in Field Deployments. The controlled test results highlighted the challenge of balancing high sensitivity with a low false positive rate, a key issue faced by CM solutions. Tables 2 and 3 indicate that a solution with the highest TP rate may present issues in field deployments if FP alerts were also considered. For example, the solution with the highest TP rate (solution E, 88% TP rate) also exhibited the lowest MDL of all solutions (2.7 kg CH_4/h). However, 79% of solution E's detection reports were false positives. In field deployment, operators may likely prefer a solution with lower sensitivity but higher confidence and fewer false alarms. For example, solutions C and D had higher MDLs (5.9 kg CH_4/h and 5.5 kg CH_4/h) and lower TP rates (35% and 59%) but much lower false positive rates (9% and 3%). Solutions I and J showed comparable MDLs (3.7 kg CH_4/h and 3.5 kg CH_4/h) with TP rates (8% and 24%) and FP rates (10% and 0%) at the low end of observed performance.

Figure 5 shows the cumulative distributions of direct, component-level measurements performed in three different studies at upstream[28] and gathering[29,32] facilities, overlaid with the range of MDLs observed in this study; ≈5% of the sources measured in these studies exceeds the lowest MDL obtained in this study (2.7 kg CH_4/h).

At operating facilities, emissions consist of nonfugitive emissions, i.e. planned emissions from venting or combusting sources, and unplanned fugitive emissions, caused by process or component failure. The task of the LDAR program is to alert the operator to fugitive emissions while ignoring nonfugitive emissions, which do not require the operator to dispatch personnel to the facility. For the studies in Figure 5 and many similar studies,[31,33−35] the total facility emissions were often dominated by nonfugitive sources, which are not shown in the figure. This type of high rate, highly variable background was not simulated during the controlled testing. Therefore, the detection performance observed in this controlled testing represents a substantially simplified scenario relative to field conditions.
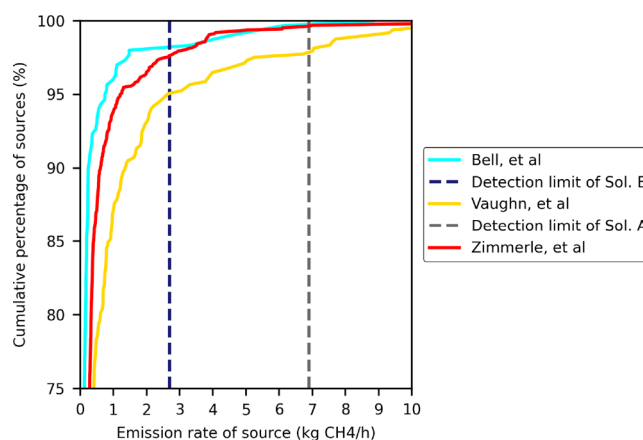

Figure 5. Range of MDLs observed in this study overlaid on a cumulative distribution of direct measurements of component-level fugitive emissions from three different studies across the natural gas supply chain. Note the truncated Y axis. For all three studies, 5% or less of measured sources exceeds the *lowest* 90% probability of detection observed in the study.

During controlled testing, detection of *any* emission source was deemed useful (i.e., treated as if it were a valid fugitive detection). In field conditions, the CM solution (or resulting LDAR program) would need to discriminate between fugitive and nonfugitive sources, which have overlapping emission rates. Therefore, the "true positive" results reported here are only a first-order representation of field performance, as the tested solutions generally cannot, at the time of testing, distinguish between fugitive and nonfugitive sources.

As noted earlier, there is increasing interest in using CM solutions to quantify emissions at facilities. To simulate this type of deployment, the study performed a Monte Carlo (MC) analysis simulating emissions detection and quantification at realistic facilities by using source-level data from Vaughn et al.,[32] the lowest curve in Figure 5. Results are shown in Table 4. Each

Table 4. Summary of Monte Carlo Analysis Results for Solutions That Estimated Emission Rates[a]

	TP detection rate (%)		quantification ratio: [estimates/actual]	
ID	mean	95% CL	mean	95% CL
A	67.1	[62.9, 70.8]	3.6	[1.9, 5.9]
B	37.9	[33.6, 42.3]	7.2	[2.6, 13.0]
C	29.6	[25.9, 33.8]	1.2	[0.9, 1.7]
D	53.9	[49.3, 58.8]	0.8	[0.6, 1.1]
E	86.8	[83.8, 90.1]	14.3	[7.5, 23.1]
F	64.3	[60.1, 68.6]	2.3	[1.7, 3.0]

[a]The Monte Carlo simulation involved 1000 runs, sampling total emissions of 304.0 kg/h from 456 sources from the study by Vaughn et al.

MC iteration applied the POD curve to determine the probability that each source would be detected and then estimated emission rates by multiplying the source rate with a relative quantification error, drawn from the distribution described earlier (See SI Section S-8.3.). Note that this analysis *did not* simulate false positive detections—i.e. reporting an emitter when none existed—to avoid assumptions about the frequency at which FP detections would be reported and presumably included in total emissions.
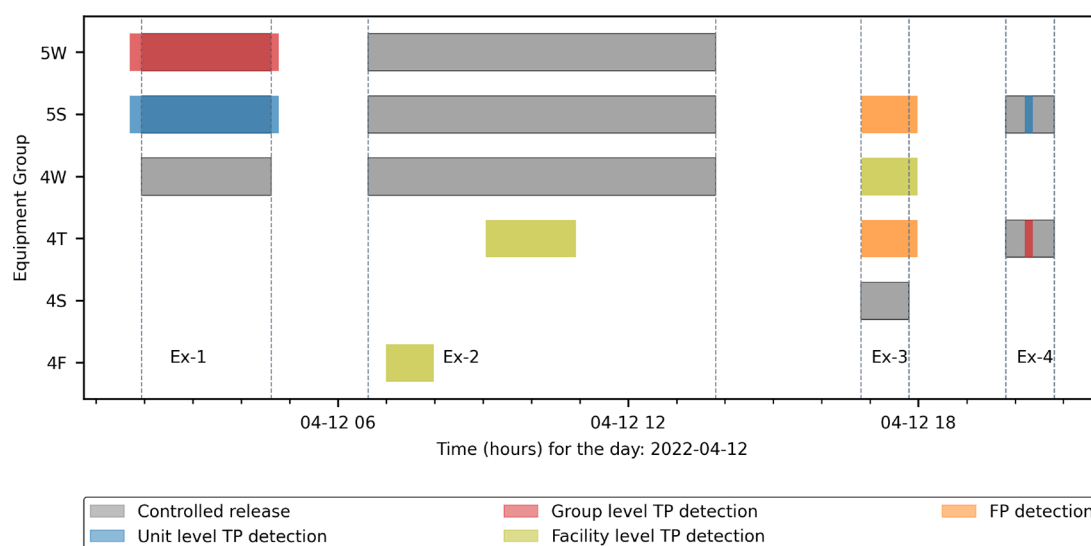
Figure 6. Timing and location of controlled releases conducted by the test center and detection reports received from solution A during a 24-h period that included four experiments. See the text for interpretation of the detections.

Results suggest that all solutions considered in the MC analysis would detect less than 87% of sources (If we exclude solution E, the number drops to about 67%.), and all but one solution would overestimate total emissions for this set of emission sources. For example, solution B would find 37.9% [33.6%, 42.3%] of 456 sources and overestimate the actual emissions by a factor of 7.2 [2.6, 13.0]. In the mean, 4 of 6 solutions estimate more than twice the actual emissions, 2 more than 7×. Simulation results suggest that using quantification estimates from CM solutions for measurement-based inventories may substantially misstate both the number of emitters and emission rate. Therefore, these estimates should be used with caution until both detection limits and quantification accuracy are improved and uncertainties are better characterized.

As discussed earlier, Tables 1 and 2 suggest that a higher sensor count is related to a higher TP rate, likely by increasing the probability that one sensor will be downwind of emissions for any given release or positioned with a favorable field of view of any gas plume. This however has significant cost implications for operators interested in deploying solutions in numerous facilities, especially when aiming to maximize emissions detection (higher TP rate). Therefore, depending on the desired application of a solution by an operator, determining the appropriate and economically feasible number of sensors to install per facility, and where they are installed, remains critical to solutions performance in field conditions.

Limitations of Testing Protocol. This study represents the first large-scale implementation of a consensus testing protocol. Inevitably, challenges arose in areas of test center capabilities, reporting methods, and results analysis. While many of the challenges were procedural and can be readily corrected (e.g., acquiring a large gas supply for the second round of testing), others are foundational and indicate that ongoing discussion of testing methods is required. For this study, key challenges concentrated at the intersection between the maturity of the solutions, the solutions' use cases for field deployments, and variations in data reported by solutions. Two examples are included below:

Use Cases. A key goal of the test protocol is to uniformly assess detection performance, which implies a specified method to report detections, including some level of source localization.

The use case assumed by the protocol was that the leak detection systems would need to issue alerts or notifications to operators when emissions were detected. Several solutions advertise this capability. Others, however, position their solutions as monitors which produce time series of concentration measurements, possibly coupled with periodic emission rate estimates, and do not implement detection reports. Additionally, several solutions provide site-level monitoring and time-resolved quantification estimates, as opposed to detecting and resolving individual sources. This suggests that an additional protocol, or protocol modifications, may be necessary to specify evaluation metrics for different reporting models. Alternatively, purchasers or regulators may settle on a single use case, requiring the protocol and solutions, to adapt to that use case.

Classification of Detections. The protocol, as written, allows each controlled release to be paired with only one *EmissionSourceID* reported by the solution and each *EmissionSourceID* to be paired with only one controlled release. The use case assumed by the protocol was that the CM solution would be capable of identifying that a single source, if detected intermittently, was a single source, at a single location. As of this test program, few solutions appeared to have solved this problem.

Classification may therefore inflate the number of TP, FN, and/or FP notifications relative to other classification methodologies. Figure 6 shows controlled releases performed by the test center and emission sources reported by a solution during 1 day of experiments. The first experiment (Ex-1) provides an illustration of how the classification methodology was intended to work. Three controlled releases located at equipment groups 5W, 5S, and 4W were performed. (SI Section S-1 discusses how to interpret equipment group labels.) The solution reported two emission sources located at 5W and 5S with start and end times closely matching the controlled releases. These were classified as TP detections at the equipment group and equipment unit level, respectively. The third controlled release at 4W was classified as FN.

The second experiment (Ex-2) in Figure 6 illustrates a scenario when the classification methodology potentially inflates TP detections and deflates FP reports. In this experiment, three controlled releases were performed, again located at 5W, 5S, and

4W. Two emission sources were reported by the solution at 4T and 4F. Although these emission sources were not reported at the correct location, the protocol matched these reports with two of the three controlled releases, reporting two TP, facility-level detections and one FN controlled release. An alternative classification procedure would exclude facility-level matching, resulting in three FN controlled releases and two FP detections. Also, note that the reported detections did not match the duration of the release but are still counted as TP.

The third experiment (Ex-3) included one controlled release on 4S. Three emission sources were reported by the solution. The protocol identified one facility-level TP and two FP detections. Looking closely at the timing, it is apparent that all three reported sources overlap the timing of the controlled release, and had an alternative classification methodology been used which allows multiple reported sources to be paired with a single detection, this could have been considered three TP detections; but none would have indicated the correct equipment group or unit.

The last experiment (Ex-4), with two controlled releases and two reported detections, both on correct equipment groups, produces two TP detections: one at the equipment group level and one at the equipment unit level.

Taken across the test program, the classification procedure defined in the protocol may favor the "reporting style" of some solutions, in some conditions, while favoring others in other conditions. In field conditions, similar confusion likely exists: Do staff dispatched to a facility by a detection alert check the entire facility or only the equipment group indicated in the detection? This type of programmatic decision may have significant impact on field performance.

Implications. CM solutions have undergone substantial development, improving deployability, accuracy, and repeatability since the publication of the first single-blind studies.[15,19] In contrast with these first studies, the current study represents the first publication of CM testing completed using a consensus protocol[26] that supports repeatable testing across both time and test centers. While the current testing does not fully simulate field conditions, the duration, scope, and complexity of testing performed here represents substantially more rigorous testing than the same test center performed prior to 2020.[19]

As a result, while the performance of CM solutions has changed substantially, testing has also increased in rigor. Concomitantly, the stakes for next-generation leak detection and quantification solutions have also risen dramatically, as more organizations consider adopting results from these solutions into programs ranging from company-internal emissions mitigation efforts to regulatory programs with financial penalties. These rapid and dramatic changes drive a need for quality testing, critical review of solution performance, and a clear understanding of uncertainties for all result types reported by these methods.

The results presented here indicate that users should utilize CM solutions with caution. Detection limits, probability of detection, localization, and quantification may or may not be fit-for-purpose for any given application. If performance is clearly understood and uncertainties are robustly considered, the solutions tested here, as a group, provide useful information. For example, most will detect large emitters at high probability, and sooner than survey methods, and will quantify those emitters well enough to inform the urgency of a field response. In contrast, relying on quantification estimates from these solutions for emissions reporting is likely premature at this point.

Finally, given the current rate of investment, solution performance changes rapidly. Periodic retesting will be required to ensure that performance metrics are current.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.2c09235.

> Solutions' performance reports, data tables, and data tables' guide, detailed description of the test facility, solutions deployment, additional results, guide to the performance reports, and bootstrapping methodology (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

Clay Bell — *Energy Institute, Colorado State University, Fort Collins, Colorado 80524, United States; BPX Energy, Denver, Colorado 80202, United States;* ⓞ orcid.org/0000-0001-5497-2073

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

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.2c09235

Notes

The authors declare the following competing financial interest(s): Subsequent to manuscript submission, Clay Bell began working for bpx energy, headquartered in Denver, Colorado. bpx energy did not participate in the drafting of this paper, and the views set forth in the paper do not necessarily reflect those of bpx energy.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) *United States Environment Protection Agency, Overview of Greenhouse Gases.* https://www.epa.gov/ghgemissions/overview-greenhouse-gases#methane (accessed 2022-07-18).

(2) *International Energy Agency: Methane and climate change.* https://www.iea.org/reports/methane-tracker-2021/methane-and-climate-change (accessed 2023-02-20).

(3) 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.

(4) 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. *Environ. Sci. Technol.* **2017**, *51*, 8001−8009.

(5) 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.

(6) Zavala-Araiza, D.; et al. Reconciling divergent estimates of oil and gas methane emissions. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 15597−15602.

(7) Alvarez, R. A.; et al. Assessment of methane emissions from the U.S. oil and gas supply chain. *Science* **2018**, *361*, eaar7204.

(8) 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. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 20018−20022.

(9) Oil and Gas Methane Partnership (OGMP). 2020. https://www.ogmpartnership.com/ (accessed 2022-08-08).

(10) Veritas. 2021. https://www.gti.energy/veritas-a-gti-methane-emissions-measurement-and-verification-initiative/ (accessed 2022-11-07).

(11) MIQ. https://miq.org/ (accessed 2022-11-07).

(12) Trustwell Standards. https://www.projectcanary.com/private/trustwell-and-rsg-definitional-document/ (accessed 2022-11-07).

(13) *United States Advanced Research Project Agency Energy (ARPA-E), Methane Observation Networks with Innovative Technology to Obtain Reductions (MONITOR).* 2014. https://www.iea.org/policies/13480-advanced-research-projects-agency-energy-arpa-e-methane-observation-networks-with-innovative-technology-to-obtain-reduction-monitor (accessed 2022-08-08).

(14) United States Environment Protection Agency. *Leak Detection and Repair:A Best Practices Guide*; 2007; pp 31−40.

(15) Ravikumar, A. P.; Sindhu, S.; Jingfan, W.; Jacob, E.; Daniel, R.-S.; Clay, B.; Daniel, Z.; David, L.; Isabel, M.; Ben, R.; Adam, R. B. Single-blind inter-comparison of methane detection technologies − results from the Stanford/EDF Mobile Monitoring Challenge. *Elementa: Science of the Anthropocene* **2019**, *7*, 37.

(16) Edie, R.; Robertson, A. M.; Field, R. A.; Soltis, J.; Snare, D. A.; Zimmerle, D.; Bell, C. S.; Vaughn, T. L.; Murphy, S. M. Constraining the accuracy of flux estimates using OTM 33A. *Atmospheric Measurement Techniques* **2020**, *13*, 341−353.

(17) Sherwin, E.; Chen, Y.; Ravikumar, A.; Brandt, A. Single-blind test of airplane-based hyperspectral methane detection via controlled releases. *Elementa: Science of the Anthropocene* **2021**, *9*, 00063.

(18) Clay, B.; Jeff, R.; Adam, B.; Evan, S.; Timothy, V.; Daniel, Z. Single-blind determination of methane detection limits and quantification accuracy using aircraft-based LiDAR. *Elementa: Science of the Anthropocene* **2022**, *10*, 00080.

(19) Bell, C.; Vaughn, T.; Zimmerle, D. Evaluation of next generation emission measurement technologies under repeatable test protocols. *Elem Sci. Anth* **2020**, *8*, 32.

(20) Siebenaler, S.; Janka, A.; Lyon, D.; Edlebeck, J.; Nowlan, A. *Methane Detectors Challenge: Low-Cost Continuous Emissions Monitoring*; V003T04A013; 2016; DOI: 10.1115/IPC2016-64670.

(21) Gardiner, T.; Helmore, J.; Innocenti, F.; Robinson, R. Field Validation of Remote Sensing Methane Emission Measurements. *Remote Sensing* **2017**, *9*, 956.

(22) 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. *Applied Energy* **2022**, *306*, 118086.

(23) 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. *Environ. Sci. Technol.* **2021**, *55*, 9140−9149.

(24) Fox, T. A.; Gao, M.; Barchyn, T. E.; Jamin, Y. L.; Hugenholtz, C. H. An agent-based model for estimating emissions reduction equivalence among leak detection and repair programs. *Journal of Cleaner Production* **2021**, *282*, 125237.

(25) National Energy Technology Laboratory. https://netl.doe.gov/node/10981 (accessed 2023-02-08).

(26) Bell, C.; Zimmerle, D. *METEC controlled test protocol: continuous monitoring emission detection and quantification*; 2020. https://doi.org/10.25675/10217/235363 (accessed 2023-03-21).

(27) 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.

(28) Bell, C.; Vaughn, T.; Zimmerle, D.; Herndon, S.; Yacovitch, T.; Heath, G.; Pétron, G.; Edie, R.; Field, R.; Murphy, S.; Robertson, A.; Soltis, J. Comparison of methane emission estimates from multiple measurement techniques at natural gas production pads. *Elem Sci. Anth* **2017**, *5*, 79.

(29) 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.

(30) 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.

(31) 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. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17768−17773.

(32) Vaughn, T.; Bell, C.; Yacovitch, T.; Roscioli, J.; Herndon, S.; Conley, S.; Schwietzke, S.; Heath, G.; Pétron, G.; Zimmerle, D. Comparing facility-level methane emission rate estimates at natural gas gathering and boosting stations. *Elem Sci. Anth* **2017**, *5*, 71.

(33) Allen, D.; Sullivan, D.; Zavala-Araiza, D.; Pacsi, A.; Harrison, M.; Keen, K.; Fraser, M.; Hill, A.; Lamb, B.; Sawyer, R.; Seinfeld, J. Methane Emissions from Process Equipment at Natural Gas Production Sites in the United States: Liquid Unloadings. *Environ. Sci. Technol.* **2015**, *49*, 641.

(34) Allen, D. T.; Pacsi, A. P.; Sullivan, D. W.; Zavala-Araiza, D.; Harrison, M.; Keen, K; Fraser, M. P.; Daniel Hill, A.; Sawyer, R. F.; Seinfeld, J. H. Methane Emissions from Process Equipment at Natural Gas Production Sites in the United States: Pneumatic Controllers. *Environ. Sci. Technol.* **2015**, *49*, 633−640.

(35) Lamb, B. K.; Edburg, S. L.; Ferrara, T. W.; Howard, T.; Harrison, M. R.; Kolb, C. E.; Townsend-Small, A.; Dyck, W.; Possolo, A.; Whetstone, J. R. Direct Measurements Show Decreasing Methane Emissions from Natural Gas Local Distribution Systems in the United States. *Environ. Sci. Technol.* **2015**, *49*, 5161−5169.