

PRACTICE BRIDGE

Single-blind inter-comparison of methane detection technologies – results from the Stanford/EDF Mobile Monitoring Challenge

Arvind P. Ravikumar^{*}, Sindhu Sreedhara[†], Jingfan Wang[†], Jacob Englander^{†,‡}, Daniel Roda-Stuart^{†,§}, Clay Bell^{||}, Daniel Zimmerle^{||}, David Lyon[¶], Isabel Mogstad[¶], Ben Ratner[¶] and Adam R. Brandt[†]

Methane leakage regulations in the US and Canada have spurred the development of new technologies that promise faster and cheaper leak detection for the oil and natural gas industry. Here, we report results from the Stanford/EDF Mobile Monitoring Challenge – the first independent assessment of 10 vehicle-, drone-, and plane-based mobile leak detection technologies. Using single-blind controlled release tests at two locations, we analyze the ability of mobile technologies to detect, localize, and quantify methane emissions. We find that the technologies are generally effective at detecting leaks, with 6 of the 10 technologies correctly detecting over 90% of test scenarios (true positive plus true negative rate). All technologies demonstrated pad-level localization of leaks, while 6 of the 10 technologies could assign a leak to the specific piece of equipment in at least 50% of test scenarios. All systems tested here will require secondary inspection to identify leak locations for repair; thus, mobile leak detection technologies can act as a complement, and not a substitute, for currently used optical gas imaging systems. In general, emissions quantification needs improvement as most technologies were only able to generally provide order of magnitude emissions estimates. Improvements to quantification algorithms, reducing false positive detection rates, and identifying early applications will be critical for deployment at scale. Even as this study provides the first independent verification of the performance of mobile technologies, it only represents the first step in the road to demonstrating that these technologies will provide emissions reductions that are equivalent to existing regulatory approaches.

Keywords: Methane emissions; Technology; Oil and gas; Leak Detection and Repair; Policy

1. Introduction

The recent Intergovernmental Panel on Climate Change (IPCC) special report on global warming of 1.5°C highlighted the importance of reducing short-lived greenhouse gases like methane (Intergovernmental Panel on Climate Change, 2018). Methane, a major component of natural gas, has a global warming potential that is 36 times that of carbon dioxide over a 100-year period (Myhre, et al., 2013), and even higher over shorter time periods (Etminan, Myhre, Highwood, & Shine, 2016). Furthermore, methane

emissions contribute to sea-level rise over much longer timescales than their atmospheric lifetimes (Zickfeld, Solomon, & Gilford, 2017). These consequences are troubling given that official methane emissions inventory in the US and Canada have been found to be systematically underestimated (Alvarez, et al., 2018; Johnson, Tyner, Conley, Schwietzke, & Zavala-Araiza, 2017).

Recently, Canada and U.S. states such as Colorado and California implemented regulations to reduce fugitive emissions from the oil and gas industry (Environment and Climate Change Canada, 2018; Colorado Department of Public Health and Environment, 2014). A major component of these regulations is periodic leak detection and repair (LDAR) surveys conducted using established methods like U.S. Environmental Protection Agency (EPA) Method-21 or optical gas imaging (OGI) technologies.

There are two primary challenges to effective methane emissions reductions. First, cost-effectiveness is a critical feature for emission mitigation considering the large spatial extent of oil and gas facilities. OGI-based LDAR surveys, while anecdotally found to be effective (Keating

^{*} Harrisburg University of Science and Technology, Harrisburg, Pennsylvania, US

[†] Stanford University, Stanford, California, US

[‡] California Air Resources Board, Sacramento, California, US

[§] Alphataraxia Management, Los Angeles, California, US

^{||} Colorado State University Energy Institute, Fort Collins, Colorado, US

[¶] Environmental Defense Fund, Washington DC, US

Corresponding author: Arvind P. Ravikumar (aravikumar@harrisburgu.edu)

Research Inc., 2016), are time-consuming – a crew of 2 people can typically visit 4–6 well pads per day, depending on distance between sites. Conducting multiple OGI-based surveys every year at large numbers of facilities or visiting sparsely distributed sites could be costly, especially when low gas prices reduce the economic benefits of increased gas recovery. Furthermore, OGI-based leak surveys are dependent on operator experience and weather conditions (Ravikumar, Wang, McGuire, Bell, Zimmerle, & Brandt, 2018; Ravikumar, Wang, & Brandt, 2017).

Second, methane emissions are highly stochastic. Many recent studies have demonstrated the influence of ‘super-emitters’ on overall methane emissions (Brandt, Heath, & Cooley, 2016). These super-emitters – a small fraction of all emission points (top 5%) that contribute over 50% of total emissions – are caused by abnormal or otherwise unintentional process conditions like equipment malfunction, failure, or operator error (Zavala-Araiza, et al., 2017). Because of the outsize contribution of super-emitters, finding and repairing these anomalous emitters as quickly as possible is key to effective methane reductions.

To address these two challenges, the solution to methane leakage detection must be: (1) faster and more cost-effective on a dollars per site basis than OGI-based leak detection, and (2) performed much more frequently or continuously.

One class of technologies that aims to meet these challenges are mobile methane detectors (Fox, Barchyn, Risk, Ravikumar, & Hugenholtz, 2019). Many new mobile sensor platforms have been developed in recent years that promise faster and more cost-effective methane leak detection. These have been shown to detect methane emissions at various spatial scales and detection thresholds. For example, truck-based measurements in British Columbia have been used to better characterize facility-level and regional methane emissions (Atherton, et al., 2017). Several aircraft- and helicopter-based measurement campaigns in the US and Canada have expanded our understanding of methane emissions and revealed widespread underreporting in official inventories (Englander, Brandt, Conley, Lyon, & Jackson, 2018; Lyon, Alvarez, Zavala-Araiza, Brandt, Jackson, & Hamburg, 2016; Conley, Franco, Faloona, Blake, Peischl, & Ryerson, 2016; Frankenberg, et al., 2016; Yuan, et al., 2015; Johnson, Tyner, Conley, Schwietzke, & Zavala-Araiza, 2017). Recent studies have also demonstrated the use of UAVs to quantify methane emissions (Golston, et al., 2018; Nathan, et al., 2015; Barchyn, et al., 2017). Satellite data are often used to assess regional and global scale methane emissions (Turner, Frankenberg, Wennberg, & Jacob, 2017; Jacob, et al., 2016). Despite promising initial results, there has been no systematic testing of mobile leak detection technologies for applications in LDAR programs. The “methane observation networks with innovative technology to obtain reductions” (MONITOR) program developed by ARPA-E (U.S. Advanced Research Projects Agency (ARPA-E), 2014) has performed the most comprehensive controlled test of new methane detection technologies based on specific cost and performance targets, although these technologies are largely designed for continuous stationary deployment. Similarly, the Methane Detectors Challenge organized by the EDF in partnership with

industry tested continuous monitoring sources for methane leak detection (Southwest Research Institute, 2015).

In this paper, we report results from the Stanford/Environmental Defense Fund (EDF) Mobile Monitoring Challenge (MMC). The MMC was an open study that called for participants to take part in a single-blind, independently administered controlled release study. Section 2 gives an overview of the MMC methods including selection process, participating technologies, and test scenarios. Section 3 describes metrics used to assess the performance of the technologies. Section 4 provides results from each of the teams that participated in the MMC, and section 5 discusses the implications of this work to methane mitigation. Detailed test-related data and further analysis of team performance is provided in the supplementary information (S.I.).

2. Methods

2.1. Team selection

The MMC invited technologists to apply by submitting information on their organization, sensor technical specifications, and commercial characteristics (see S.I. for application form). The project website was advertised widely and remained open for applications for 65 days. The MMC received 25 applications from technologists based in 5 countries. An industry advisory board including members of major oil and gas companies was created to provide industry insights into desirable features of methane detection systems. Scientists and project managers from Stanford and EDF, as well as the industry advisory board, reviewed and scored the applications separately, then gathered in person to discuss the applications and select the final list of participants (see S.I. section 1 and **Table 1**). Selection criteria included scientific soundness, applicability to oil and gas facilities, and path to commercialization. Eleven organizations developing 12 technologies were selected to participate in the MMC – these included 3 truck-, 3 plane-, and 6 drone-based platforms. Due to technical and logistical challenges, two selected teams – Kairos Aerospace and Bluefield Technologies – did not participate in the field trials. After selection, authors (A.P.R. and I.M.) conducted one-on-one phone interviews with the science team of each technology to understand technology features and limitations. Teams were then assigned to one of three testing weeks based on their self-reported methane detection limits. A summary of the technologies selected as part of this study is given in **Table 1** (also see S.I. SM_Table 1 for technical specifications). The tests in this study represents an independent assessment of the performance of methane leak detection technologies as would be observed by a regulator or site operator. As such, the participating teams did not have any interaction with or knowledge of the scientific team’s analysis of their performance after the field tests.

2.2. Test locations and controlled releases

Two test locations were chosen for the MMC. Two weeks of releases were performed at the Methane Emissions Technology Evaluation Center (METEC), a Department of Energy funded controlled release facility in Fort Collins, CO. Release rates of total gas (87% CH₄, see S.I. section 2) at the METEC facility were in the 0–15 standard cubic feet per

Table 1: Summary of the technologies participating in the Stanford/EDF Mobile Monitoring Challenge. DOI: <https://doi.org/10.1525/elementa.373.t1>

Company	Platform Type	Sensor Type/Species Measured	Survey Method	Survey Speed ^a (mph)	Survey Height (m)
ABB/ULC Robotics	Drone	Cavity-enhanced laser absorption spectroscopy – Methane	Modified raster scan (wind responsive)	2–5	5–10 m
Advisian	Helicopter	Laser absorption spectroscopy – methane/ethane	Upwind/Downwind transects with sample tube	2–5	15–20 m
Aeris Technologies	Vehicle	Laser absorption spectroscopy – methane, ethane, water-vapor	Fence-line around equipment groups, facility	~10	1–2 m
Baker Hughes (GE)	Drone	Laser absorption spectroscopy – methane	Concentric circles around equipment	~5	5–10 m
Ball Aerospace	Plane	Airborne differential LIDAR – methane	Fly-overs (multiple passes)	~115	~1000 m
Heath Consultants Inc.	Vehicle	Off-axis integrated cavity output spectroscopy – methane, ethane	Fence-line around equipment groups, facility	~10	~1–2 m
Picarro	Drone and Vehicle	Cavity ringdown spectroscopy – methane, ethane, water-vapor	Upwind/Downwind transects	2–5	5–10 m
Seek Ops Inc.	Drone	Laser absorption spectroscopy – methane	Raster scan, with flux plane mapping	2–5	1–3 m
U Calgary	Vehicle	Open-path laser spectroscopy – methane	Fence-line and highway-based screening	~10 (fence-line) 30–50 (highway)	2–3 m
U Calgary and Ventus Geospatial	Drone (fixed-wing)	Open-path laser spectroscopy – methane	Multiple downwind plume transects	30–40	28–124 m

^a Some technologies were limited in their speed due to speed-limits at the METEC test-site (10 mph).

hour (scfh) range (0–0.25 kg CH₄/h). One week of releases were performed at a facility owned by Rawhide Leasing near Sacramento, CA. Test rates at the Sacramento facility spanned 0–1500 scfh (0–26 kg CH₄/h). Not all releases could be performed at METEC because some teams reported emissions detection limits that were too large for the emissions capability for the equipment and permitting in place at METEC (see S.I. section 2).

Teams were grouped based on self-reported detection thresholds – grouping together teams with similar detection limits ensures that tests are not too facile (for example, only leaks significantly larger than detection limits) nor too difficult (test leaks significantly smaller than detection limits). The final test dates and grouping are shown in S.I. All tests were conducted in a single blind fashion – only authors A.P.R., C.B., and A.R.B. were aware of the actual leak rates and saw leak rates during the test process. All technology teams and other members on the project did not have access to the test scenarios until after the tests were completed. Approximately 3 months after testing was completed, after all teams reported final results to Stanford scientists, the true leak rates were given to the teams for their own use in further technology development.

The blinding of the leakage results could in theory be broken by audible sound or odor from the emission point. Because of the low release volumes, no Stanford staff noted discernable noise of emission while touring sites. For safety reasons, both sites release odorized gas which contains mercaptan compounds. This resulted in frequent odors at both sites, which shifted with the

winds and would be most detectable when the team members were downwind from the release point (either due to the team moving with the vehicle or due to wind shifts). Given the complexity of the release patterns, their frequent temporal changes (every 10 min) and possibility of multiple release points, we do not expect the odors to provide consistent patterns that could be used by teams to break the blind. Furthermore, real oil and gas facilities frequency have odors associated with non-methane compounds in the raw gas, analogous to the test scenario here.

Methane Emissions Technology Evaluation Center (METEC), Fort Collins, CO. METEC is an ARPA-E funded controlled release test site for evaluating new methane emissions detection technologies (see **Figure 1(a)**). The site contains equipment typically found at natural gas production facilities such as wellheads, separators, and tank batteries, organized across 5 clusters of equipment analogous to well-pads (see S.I. section 2). The pads vary in complexity –two of the pads had 1 wellhead, 1 separator, and 1 tank each. Other pads had multiple equipment of the same group, such as 5 wellheads on pad 4. Each team was assigned a pad for initial testing and were rotated across pads periodically to ensure all teams tested on all pads. Each piece of equipment has multiple leak points fashioned out of 0.64 cm (¼ in.) diameter steel tubing – the tubing is well concealed to mimic realistic leak sources such as connectors and flanges. Natural gas (86–88 vol% methane, 8–10 vol% ethane, 2–4 vol% trace gases, with odorant) is sourced from a centrally located tank at 172 bar (2500 psi), with flow controlled by a combination of

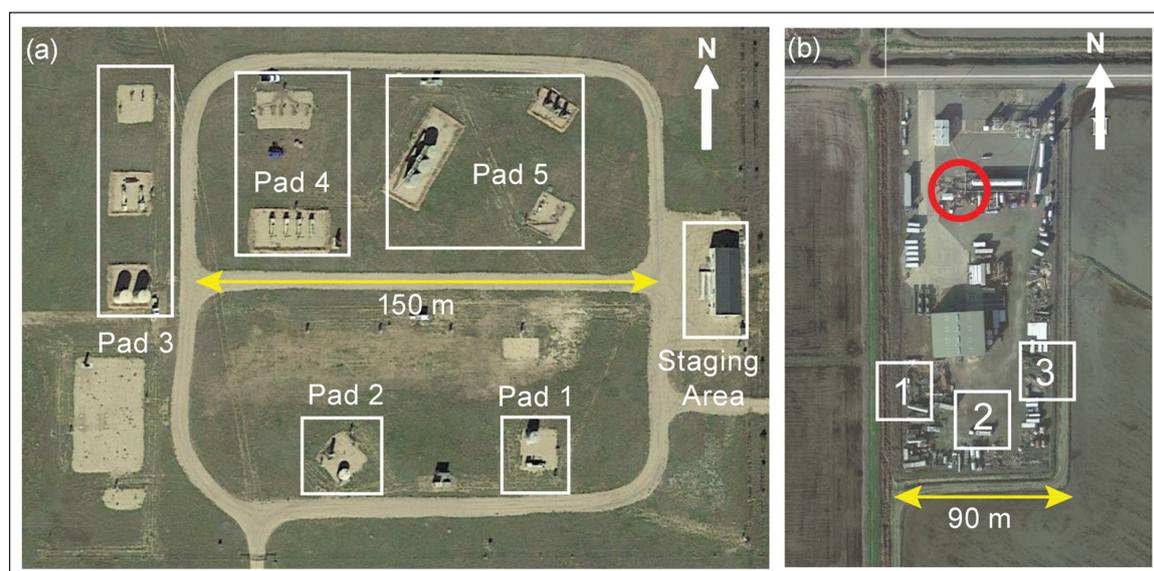


Figure 1: Google Earth image of Stanford/EDF Mobile Monitoring Challenge test locations. (a) METEC facility in Fort Collins, CO showing the pad configurations 1 through 5 and the staging area, **(b)** Rawhide Leasing facility near Sacramento, CA showing the approximate leak locations denoted 1, 2, and 3. The red circle shows the location of an anomalous non-test intermittent methane source. DOI: <https://doi.org/10.1525/elementa.373.f1>

pressure regulators and choked-flow orifices fitted with flow meters. During these tests, flow rates ranged from 0 to about 15 scfh. In addition, the site also included a 3-axis sonic anemometer that collected 1 minute-averaged meteorological data at ~3 m above the ground. The wind data from this instrument is later used to analyze the effect of intra-pad interference during testing (see S.I. section 5 and 6).

Rawhide Leasing Gas Yard, Sacramento, CA.

Controlled release experiments at the Sacramento sites consisted of 3 individual sources separated by 30–60 m (see **Figure 1(b)**, and S.I. section 2). The sources consisted of a 2 m elevated stack of 2.54 cm diameter with test flow rates ranging from 50 scfh (0.87 kg CH₄/h) to about 1500 scfh (26 kg CH₄/h). Each of the sources were individually metered using a Sierra Instruments QuadraTherm 740i thermal mass flow meters with an accuracy of ±0.75% of full-scale reading. Natural gas (91 vol% methane, 6 vol% ethane, 3 vol% trace gases) was sourced from a pressurized tank at 2500 psi and stepped down to 50 psi with a regulator before passing through the flow meters. In addition to flow rates, the mass flow meters also monitored gas temperature along the line. Because over 90% of the flow rates were relatively small, being lower than 400 scfh (<7 kg CH₄/h), we did not experience issues with Joule Thompson cooling effect (Maric, 2005). To allow for effective plume development through the atmosphere for aerial detection, leaks tested at this facility included a 3-minute buffer zone before and after each test period. The pre-test buffer allows the plume to develop while the post-test buffer lets the plume clear the area before the next test to avoid plume-overlap interference. This test site had other methane emissions not part of the controlled release test that were picked up by the technologies tested here (red circle in **Figure 1b**). The teams performed appropriate analysis to remove the effect of the co-located emissions whenever possible.

Test scenarios. We developed a series of test protocols of increasing complexity to assess the performance of mobile leak detection technologies. These tests were designed to assess the ability of technologies to locate and detect leaks, quantify flow rates, resolve multiple leaks that are closely spaced, and do it all within a specified time limit. The test protocols were similar at METEC and Sacramento test locations but varied in complexity (see S.I. section 3 and SM_Table 3). The teams could use as little or as much time as needed within the maximum allotted time for each test. We chose not to time individual teams separately and instead opted for “maximum time allowed” for two reasons: (a) vehicle speeds at both test sites were limited to 10 mph, artificially impacting measurement time for trucks, and (b) test sites are not the same as actively producing well sites and therefore, measurement times here might not be representative of field performance. While some teams stopped after detecting the leak within 2–4 minutes of a timed test, other teams used the entire test duration to improve their localization and quantification precision.

As multiple teams were measuring leaks simultaneously at METEC, study author (A.P.R.) worked in real-time to adjust leak locations across the 5 pads to minimize interference between pads. Leaks were preferentially placed downwind of non-leaks to minimize the amount of methane blowing from leaking sites to non-leaking sites. In addition, participants could drive between leak locations on different pads to sample both upwind and downwind methane data. Real-time monitoring of wind conditions by METEC personnel were used to assign leak configurations across the five pads for each test scenario that would minimize interference. Because teams rotated, and wind conditions changed, each team was given a mix of leak and non-leak observations (generally 50% leaks and 50% non-leaks). In S.I. (section 5 and 6) we present results from cleaning reported data of possible interference, but

present baseline results below. This is done by excluding any test scenarios that had a reasonable possibility of interference from upwind emission sources (see S.I. section 5 for more details on exclusion criteria). To be clear, interference is likely at oil and gas facilities either due to co-located emissions from the same pad or downwind emissions from a different pad. Whether this impacts technology performance is important to understanding the robustness of the algorithms used by the technologies to interpret raw data.

2.3. Performance metrics

A set of common metrics were developed to account for the variety in the sensors used, mobile platforms, survey protocols, analysis algorithms, and reporting parameters. These metrics included – (a) leak detection probability, (b) detection and localization, and (c) quantification accuracy. These are briefly described below.

Leak detection probability: Leak detection probability varies as a function of leak size for each technology. Leak detection probabilities are critical inputs to natural gas field simulators such as the Fugitive Emissions Abatement Simulation Toolkit (FEAST) that can help compare new detection technologies with established methods (Kemp, Ravikumar, & Brandt, 2016). Furthermore, developing estimates of detection threshold will assist in direct comparisons with currently used OGI technologies such as the FLIR GF-320 cameras. In this study, for technologies tested at METEC, we group leak sizes into 5 bins: <1, 1–3, 3–5, 5–8, and >8 scfh, and determine the fraction of test scenarios in each bin that was detected. For technologies tested in Sacramento, CA, the bin sizes were: <150, 150–300, 300–450, 450–600, and >600 scfh. All test scenarios of both leaks and non-leaks (zero tests) are combined into a true/false matrix chart. Four results are possible – a true positive (TP) result is recorded when a team correctly identifies an actual leak; a true negative (TN) occurs when a team correctly identifies a zero-leak test as not containing a leak; a false positive (FP) occurs when a team mis-identifies a zero-leak scenario as a leak; and a false negative (FN) result occurs when a team wrongly characterizes a leak as a zero-leak scenario.

Detection and localization: TP results are grouped into three levels of localization accuracy – level 1, 2, and 3. While some teams reported GPS coordinates that would make exact displacement calculations between actual and measured leak locations possible (i.e., m of offset between expected and actual location), numerous teams specified the equipment type or specific piece of equipment where emissions were detected. We chose this three-level metric to harmonize the different types of location information reported by the teams. All three levels of leak localization will require a secondary inspection to identify the leaking component or the correct leaking equipment for further repair.

Level-1: The team correctly identifies the leaking equipment. In scenarios with multiple equipment of the same group (e.g., 5 wellheads), the teams should also have identified the correct equipment number in that group. This indicates equipment-level attribution ability – for example, a team correctly reporting a leak on wellhead 4 on Pad 4, and corresponds to location accuracy within ~1–4

m. Although the correct equipment has been identified in Level-1 type leaks, a repair crew may still require a method like handheld Method-21, OGI, or bubble test to identify the leaking component.

Level-2: The team correctly identifies the leak equipment group but does not identify (or misidentifies) the equipment number when multiple equipment of the same group is present. For example, a team reporting a leak on wellhead 2 on Pad 4, when wellhead 4 was the actual leak location. Level-2 detection signifies some attributional ability, with effectiveness determined by the spatial density of equipment as well as resolution capabilities of the technology. Level-2 detection corresponds to location accuracy within ~4–10 m. There were no level-2 type leaks at the Sacramento test site because it contained only 3 isolated leak sources and did not have any group sources present. All tests results from the Sacramento site were identified as Level-1 or Level-3 detects. A Level-2 detection requires the operator to first identify the leaking equipment and component using a Method-21 or OGI-based sensor before repairs.

Level-3: The team correctly identifies a leak, but misidentifies the equipment group – for example, a team reporting a leak on separator 2 on Pad 4, when wellhead 4 on Pad 4 was the actual leak location. Teams that did not report any specific location data were automatically assigned Level 3 detection. This level translates to pad-level detection ability (~10+ m) and can be considered as a proxy for screening type technologies. A secondary ground team with a handheld device would be required to identify specific leak location before repairs can occur.

Finally, we also analyze results across equipment type – wellheads, separators, and tanks at METEC, and sources 1, 2, and 3 in California. This will show differences in performance that are affected by the height of the leaking equipment, a critical metric for truck and drone-based systems.

Quantification: Teams were asked to quantify emissions and report estimated flow rates for a subset of the test scenarios. Some teams also quantified emissions in scenarios where it was not required, and these results are scored as well. Quantification performance is shown as a parity chart between actual and estimated leak rates, with error bars if reported by teams. The best-fit linear regression between measured and actual volumes and the 95% confidence interval around the slope is reported.

We choose the charitable interpretation of reported data in the case of ambiguity. For example, consider a scenario where we tested detection and quantification of 2 closely spaced leaks on a separator group, and the team reported one quantification measurement for a separator leak without specifying the number of leaks. We interpreted this result as the team ‘detecting’ both leaks without resolving leak equipment, resulting in 2 level-2 detections. Furthermore, the quantification result would be compared to the combined flux rate of both leaks.

3. Results of the Mobile Monitoring Challenge

This section describes detailed results for participating team. Team performance is presented in alphabetical order. A few caveats should be noted:

- The sample sizes in different tests varies across teams because of the random nature of assigning test scenarios to teams, varying wind directions, robustness of technologies to high winds, and differences in preparation time across the technologies.
- The performance of all technologies is affected by weather conditions to varying degrees. We present data below from all test scenarios, irrespective of weather conditions. S.I. contains detailed analysis of team performance as a function of inter-pad interference.
- The suitability of a given technology for methane leak detection depends not only on the performance of the technologies themselves, but also on parameters such as facility type, and infrastructure density.

3.1. ABB/ULC Robotics

ABB deployed a UAV-mounted methane-only sensor based on cavity enhanced laser absorption spectroscopy. In addition to gas concentration values, the UAV collected GPS coordinates and wind speed using an on-board anemometer.

Figure 2(a) shows the binary detection results of the ABB system. TP rate is 77% ($n = 43$ of 56), all at level-3 localization, indicating detection effectiveness at the pad level. The average leak rate of the 18 (23%) FN indications was 2.4 scfh. FP rate is 22% ($n = 10$ of 45). A majority of these false positive (60%) occurred when multiple leaks were tested, indicating potential issues with leak resolution algorithms.

Figure 2(b) shows the detection probability of the technology as a function of leak-size. Detection probability varies from <30% for leaks <1 scfh, to 100% for leaks >8 scfh. The 53% detection probability for leaks smaller than 3 scfh partially explains the average false negative rate of 2.4 scfh. Re-testing of this technology only at higher leak rates would likely result in improved TP rates.

Figure 2(c) shows the quantification parity chart. The slope of the best-fit line was 0.025, indicating no correlation with the actual leak rate ($R^2 = 0.01$, Pearson's $\rho = 0.02$). The average difference between the actual

and measured leak rate was +2.8 scfh (95% C.I. [1.1, 4.5], $n = 28$). This underestimation was especially severe for leaks larger than 5 scfh, with a mean actual leak rate of 7.4 scfh, and the corresponding average measured leak rate being 3.1 scfh.

3.2. Advisian (Worley Parsons)

Advisian technology employed a Vapor-55 helicopter UAV outfitted with a laser spectroscopy-based methane-ethane sensor. The sample inlet was suspended about 50 ft below the helicopter through an inlet tube pulled behind the helicopter. In addition to gas concentration, the UAV collects GPS coordinates and meteorological data. This team provided two results for each test scenario – one that was immediately available based on 3-dimensional plots of concentration, and the other based on off-site data analysis performed on data uploaded to the cloud. Below we have used the off-site analysis results.

Figure 3(a) shows the TP rate for detection was 94% ($n = 36$), with the level-1, level-2, and level-3 localization at 47%, 25%, and 22%, respectively. The nearly 50% level-1 localization demonstrates equipment-level leak detection capability. However, 10 of the 17 level-2 and level-3 leak detections occurred during the multiple leaks per pad test scenarios, indicating challenges with distinguishing closely-spaced leaks. Across equipment types, the leak detection effectiveness was 90% ($n = 10$) for wellheads, 100% ($n = 24$) for separators, and 50% ($n = 2$) for tanks. The difference between tanks and wellheads/separators was not statistically significant due to the small sample size. The FP rate was 7% ($n = 2$ out of 29).

Figure 3(b) shows that the 100% detection probability cut-off is approximately 3 scfh.

Figure 3(c) shows the quantification parity chart for the sensor, with the slope of best-fit linear regression being 2.7. The error bars shown were directly reported by the team. The average difference between the actual and measured leak rate is -12.7 scfh (95% C.I. [-20.6, -4.8], $n = 33$), representing an average overestimation by approximately 3.5 times the average controlled release rate (3.64 scfh).

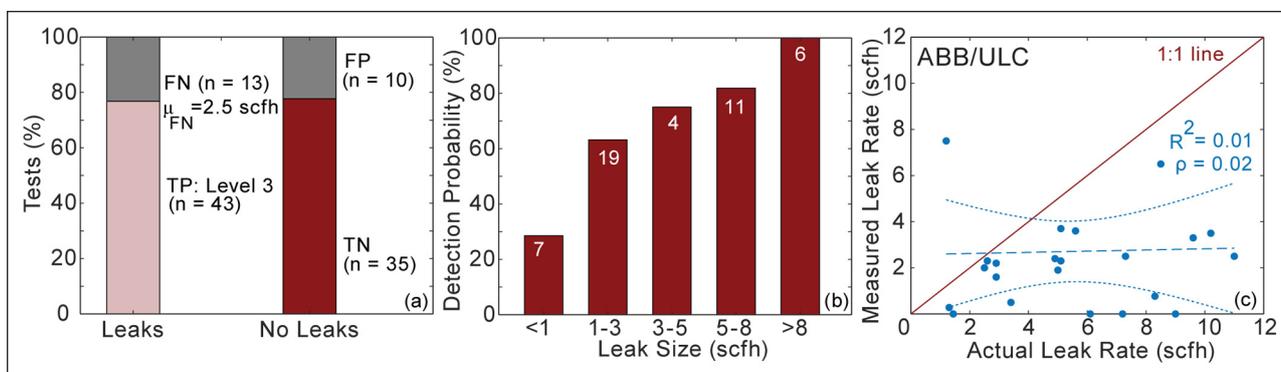


Figure 2: Performance results of ABB (ASEA Brown Boveri)/ULC Robotics in the Stanford/EDF Mobile Monitoring Challenge. (a) Binary detection characteristics of the technology, disaggregated by the level of true positive detection, (b) Leak detection probability across different leak size range, and (c) quantification parity chart between actual and measured leak rates. (TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative). DOI: <https://doi.org/10.1525/elementa.373.f2>

3.3. Aeris Technologies

Aeris Technologies uses a mid-infrared laser spectroscopy-based sensor mounted on a ground vehicle to detect methane, ethane, and water-vapor. In addition to gas concentrations, the system also measures meteorological data and GPS coordinates.

Figure 4(a) shows the detection characteristics for Aeris. Out of 52 total leaks, TP rate was 88%, with 50% at level-1, 15% at level-2, and 23% at level-3 localization. Six leaks were misidentified as zero leaks (FN), with mean FN leak rate of 1.5 scfh. Three of the six FN observations occurred during the multiple leaks per pad test, indicating challenges in spatial resolution of closely located emissions sources. Notably, there is a difference in detection effectiveness between equipment types: wellheads (TP = 87%, n = 15) and separators (TP = 97%, n = 32) had very high success rates, while, tanks had lower success rates (TP = 40%, n = 5). This suggest a possible challenge for measuring from taller equipment from a vehicle-based sensor and would point to the need for a wider sampling path to

allow more time for groundward dispersion of higher leaks.

Out of the 48 zero leaks tested, the FP rate was 15% (n = 7). Of the FP detections quantified (5/7), the average quantified FP leak rate was 0.5 scfh – over 19 times smaller than average measured leak rate of 9.6 scfh for actual leaks. This indicates that false positives were an issue near the detection limits of the technology, as seen in the detection probability curve **Figure 4(b)**.

Figure 4(c) shows the quantification parity chart for Aeris. The slope of the best-fit regression line is 3, indicating overestimation. The average difference between the actual and measured rate was –6.5 scfh, with the 95% C.I. ranging from –10.2 to –2.3 scfh. Five large overestimates (>30 scfh) in quantification are not shown in **Figure 4(c)** for clarity. However, these data points are included in our statistical analysis and are not arbitrarily discarded while calculating the R² and ρ coefficients. Removing these from the statistical analyzes increases R² and ρ coefficients to 0.32 and 0.55, respectively.

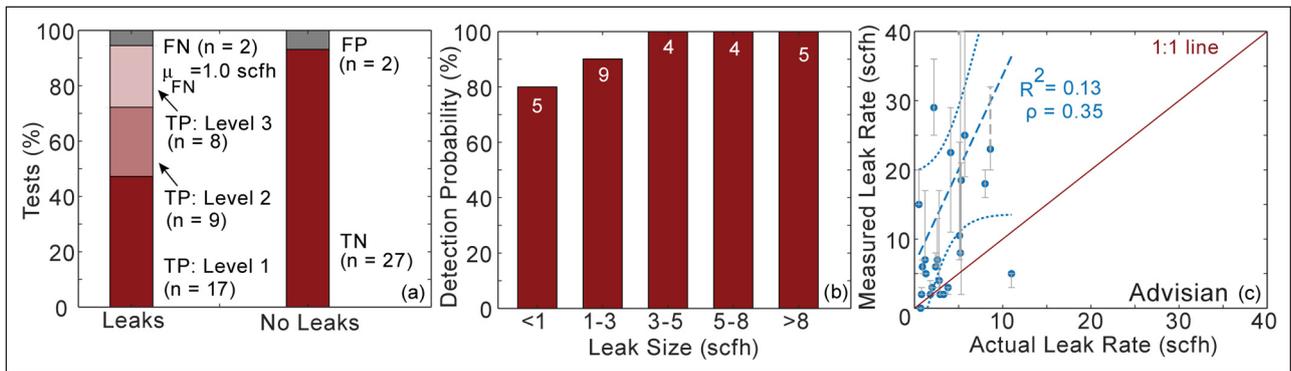


Figure 3: Performance results of Advisian (Worley Parsons Group) in the Stanford/EDF Mobile Monitoring Challenge. (a) Binary detection characteristics of the technology, disaggregated by the level of true positive detection, (b) Leak detection probability across different leak size range, and (c) quantification parity chart between actual and measured leak rates. (TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative). DOI: <https://doi.org/10.1525/elementa.373.f3>

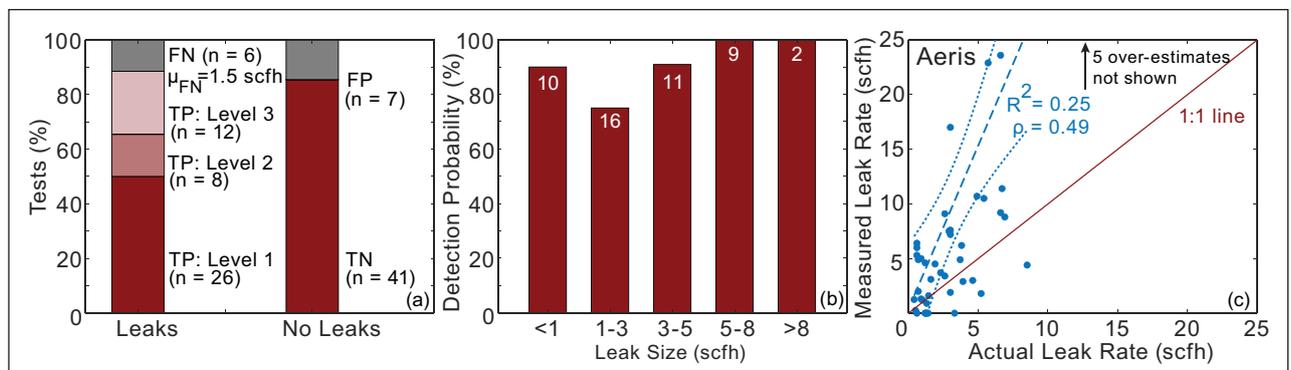


Figure 4: Performance results of Aeris Technologies in the Stanford/EDF Mobile Monitoring Challenge. (a) Binary detection characteristics of the technology, disaggregated by the level of true positive detection, (b) Leak detection probability across different leak size range, and (c) quantification parity chart between actual and measured leak rates. 5 large over-estimates in the data are not shown in (c) for clarity but are included in calculating best-fit lines. (TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative). DOI: <https://doi.org/10.1525/elementa.373.f4>

3.4. Baker Hughes (GE)

BHGE operated an UAV-mounted methane-only sensor based on absorption spectroscopy. The sensor collects single point measurements of methane concentration at 2 Hz frequency along with location information through an on-board GPS. Leaks are analyzed separately by combining with weather parameters from the ground anemometer data made available to the team.

Figure 5(a) shows the detection characteristics of the UAV-mounted sensor. TP rate is 68% (n = 39 of 57). Approximately half the detected leaks – 20 out of 39 – were level-3 localization, indicating pad-level attribution. Mean FN leak rate is 2.5 scfh, which is lower than the 6 scfh detection limit as described by the team prior to testing. FP rate of 71% (32 of 45) is high, indicating a need to improve processing algorithms to reduce false positive detection.

Figure 5(b) shows the detection probability charts for the technology. For leaks below 3 scfh, the detection probability is about 50%, aligning with team reported detection limits. BHGE reliably detected leaks greater than 8 scfh with 100% detection probability.

Figure 5(c) shows the quantification parity chart, Best-fit regression line has slope of 0.05, indicating underestimation and lack of sensitivity to leak size. The mean measured leak rate was 1.2 scfh, corresponding to an average error of +2.2 scfh (95% C.I. [1.4, 3.0], n = 57) – the measured rates were only 35% of the actual leak rates.

3.5. Ball Aerospace

Ball Aerospace tested a methane-only sensor based on airborne differential absorption LIDAR mounted on a single-engine Cessna T206. The sensor samples data at 10 kHz and collects path-integrated methane concentration data in a ‘push-broom’ approach with a spatial resolution of about 2 m on the ground. Meteorological data from nearby ground weather station is integrated with sensor data to develop quantitative flux estimates. The airplane flew at 2800 ft altitude, and the controlled release tests were conducted at the Sacramento, CA site between 21–25 May 2018.

Figure 6(a) shows the detection characteristics of the Ball aerospace team. Out of 50 total leaks that were tested, TP rate is 74% at level-1 localization, demonstrating the

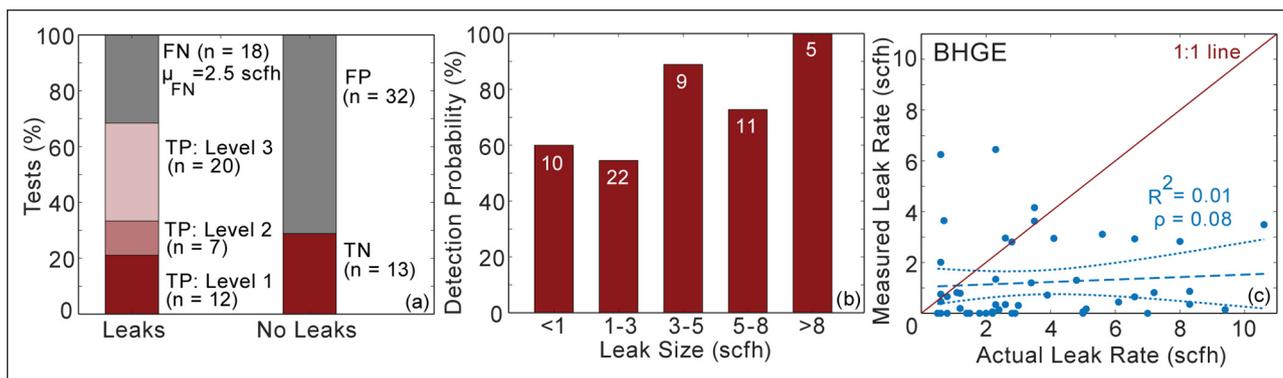


Figure 5: Performance results of Baker Hughes (General Electric) (BHGE) in the Stanford/EDF Mobile Monitoring Challenge. (a) Binary detection characteristics of the technology, disaggregated by the level of true positive detection, (b) Leak detection probability across different leak size range, and (c) quantification parity chart between actual and measured leak rates. (TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative). DOI: <https://doi.org/10.1525/elementa.373.f5>

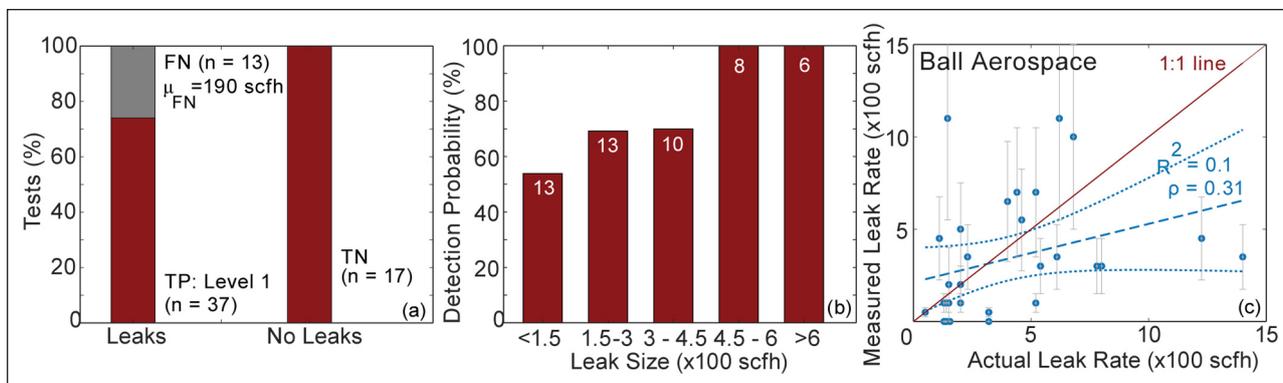


Figure 6: Performance results of Ball Aerospace in the Stanford/EDF Mobile Monitoring Challenge. (a) Binary detection characteristics of the technology, disaggregated by the level of true positive detection, (b) Leak detection probability across different leak size range, and (c) quantification parity chart between actual and measured leak rates. (TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative). DOI: <https://doi.org/10.1525/elementa.373.f6>

source attribution ability of the aircraft-mounted sensor. FN rate is 26% with mean FN rate of 190 scfh. This technology did not detect any FPs ($n = 17$). While the detection effectiveness at source 1 (west) and source 2 (south) were 88% and 80%, respectively, the effectiveness at source 3 (east) was only 46% ($n = 13$). The reason for this discrepancy is not well understood.

The detection probability plot (see **Figure 6(b)**) shows a threshold around 450 scfh. Leaks greater than 450 scfh had 100% probability of detection, while leaks smaller than 450 scfh had an average detection probability of about 64%. The lower detection effectiveness for leaks smaller than 200 scfh also explains the observed mean FN rate (190 scfh, see **Figure 6(a)**).

Figure 6(c) shows the quantification parity chart, with a best-fit linear regression slope of 0.32. The error bars are based on the teams' reports. The average error between actual and measured leak rate was +58 scfh (95% C.I. [-79, 196], $n = 32$), indicating an underestimation of the actual leak rate by ~15%. However, the confidence interval for the average error includes 0.

The effectiveness of airplane-based detection is dependent on the number of passes over the facility. In this study, the Ball Aerospace team averaged 4 passes during the 10-minute tests and 7 passes during the 15-minute tests that required quantification in addition to detection.

3.6. Heath Consultants Inc.

Heath Consultants Inc. tested the Mobile Guard – a vehicle-based leak detection system – that uses off-axis integrated cavity output spectroscopy to detect methane and ethane emissions. In addition to the analyzer, the truck also collected GPS and weather data using an on-board anemometer.

Figure 7(a) shows the detection characteristics of the truck-based measurement system. Out of a total of 92 leaks tested, Heath identified 86 at least partially (levels 1, 2, or 3), resulting in a FN rate of 6.5%. The average leak rate for the false negative tests was 1.8 scfh. 75 of the 86 detected leaks, or 82%, were in the level 1 or level 2 category – the technology identified the correct equipment

group for the leak source the vast majority of the time. In addition to the true positive results, Heath had a false positive rate of 25.6%, with 11 of the 43 zeros incorrectly identified as leaks. This rate was affected by the unusually windy conditions during the week of testing (see S.I. section 5). The mean wind speed during testing was over 13 mph, affecting detection and complicating analysis of raw concentration data. 9 out of 11 false positive detections for Heath occurred during the multiple leaks per pad test scenario, indicating potential challenges in resolving multiple leak sources from spatial concentration data.

Figure 7(b) shows the detection probability curves for Heath as a function of leak size range. This technology has high sensitivity, detecting leaks that are smaller than 1 scfh with approximately 90% success rate. No statistically significant difference in ability to detect leaks across different equipment types exists.

Figure 7(c) shows the quantification parity chart including both single-leak and multi-leak measurements. The slope of the best-fit linear regression line is 0.44 with larger leaks generally underestimated. The overall mis-estimation was skewed negatively (toward underestimation) but not statistically significant from 0 (95% C.I. [-1.4, 0.23], $n = 23$).

3.7. Picarro Inc.

Picarro tested a hybrid drone and vehicle-based methane, ethane, and water-vapor sensor based on optical absorption using cavity ringdown spectroscopy. The sensor was deployed on the ground in a vehicle while the gas inlet for the system was mounted on an unmanned aerial vehicle (UAV). This inlet is tethered to the ground-based sensor using a 150 ft long inlet tube. In addition to pollutant concentrations, the sensor also measured wind speed and GPS coordinates at approximately 1 Hz frequency.

Figure 8(a) shows the detection characteristics of Picarro's drone-based system. A TP rate of 92% (59/64) was achieved at level-2 and level-3 localization, demonstrating detection effectiveness at the pad-level. The average leak rate of the FN measurements – 5 out of the 64 tests – was 3.2 scfh. All tank-related leaks were correctly

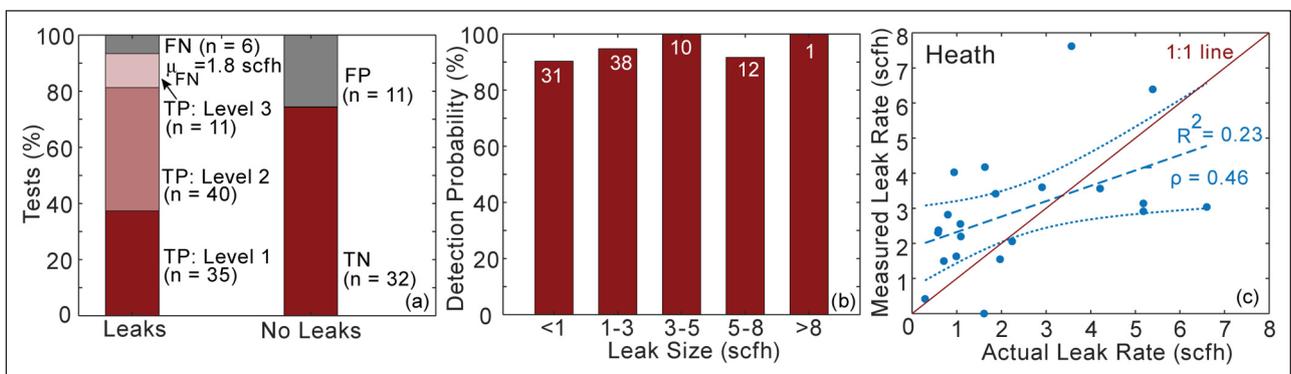


Figure 7: Performance results of Heath Consultants in the Stanford/EDF Mobile Monitoring Challenge. (a) Binary detection characteristics of the technology, disaggregated by the level of true positive detection, (b) Leak detection probability across different leak size range; the numbers on bars refer to sample size in each bin, and (c) quantification parity chart between actual and measured leak rates. (TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative). DOI: <https://doi.org/10.1525/elementa.373.f7>

identified (n = 6), showing success with leaks at height (difference is not statistically significant due to small sample size). A FP rate of 39% was found (9/23).

The level-3 leaks, all identified during the multiple leaks per pad test, point to limited ability to attribute sources at the equipment-group level. However, it was also during the multiple leaks per pad test that this technology tested 8 of the 9 false positive results in this study. This performance indicates suitability at screening pad-level emissions, while also demonstrating the need for improvement in algorithms for source attribution under complex emissions scenarios. The UAV system was not tested on one of the days (April 11th, 2018) because of winds gusting over 23 mph.

Figure 8(b) shows the detection probability curve for Picarro. There is no statistically significant difference in detection between the different leak rates. A high leak detection probability at small leak rates (<1 scfh) points to the underlying sensor's high sensitivity. **Figure 8(c)** shows the quantification parity for a sample size of 86 leaks (all leaks were quantified by Picarro). The error bars in **Figure 8(c)** are 70% confidence intervals as reported by Picarro. The slope of the regression line is 0.36, driven

by underestimation of leaks at larger leak rates (>6 scfh), while smaller leaks are generally overestimated. The average difference between the actual leak rate and the measured leak rate was -0.89 scfh, with a 95% confidence interval between -1.8 scfh and 0.01 scfh.

3.8. Seek Ops Inc.

Seek Ops Inc. tested a methane-only, continuous in-situ monitoring sensor based on laser absorption spectroscopy mounted on a UAV platform. The drone measured methane concentration and GPS coordinates, while wind is measured using a custom ground station on the site erected by the team.

Figure 9(a) shows the detection characteristics of the drone-mounted sensor. This technology had a 100% TP rate (n = 63), with a majority of the leaks (68%) detected at the level-1 scenario. The remaining emissions were equally split (16% each) between level-2 and level-3 detection scenarios. Most level-3 scenario for Seek Ops occurred on pads 1 and 2, where the specific leak location was ambiguous because of the heat map of emission covering more than one equipment. These aggregate statistics also

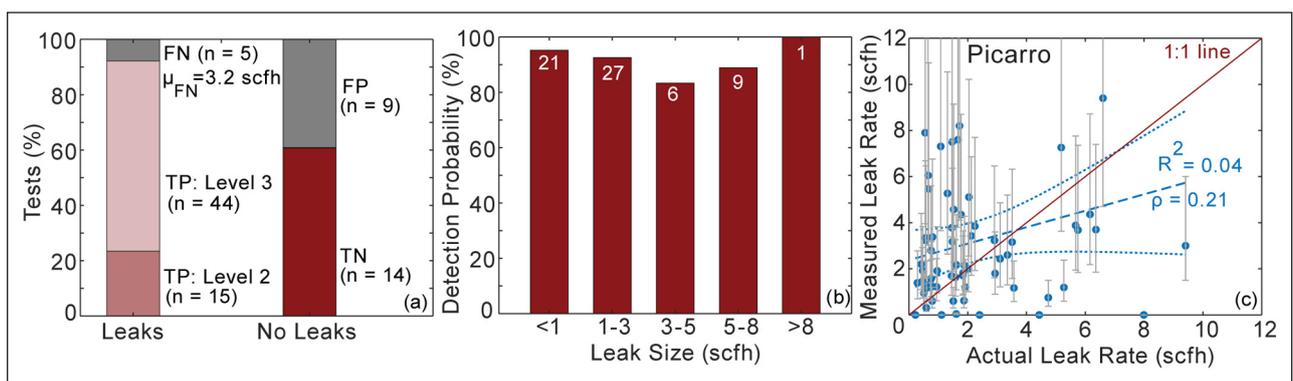


Figure 8: Performance results of Picarro Inc. in the Stanford/EDF Mobile Monitoring Challenge. (a) Binary detection characteristics of the technology, disaggregated by the level of true positive detection, **(b)** Leak detection probability across different leak size range, and **(c)** quantification parity chart between actual and measured leak rates. (TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative). DOI: <https://doi.org/10.1525/elementa.373.f8>

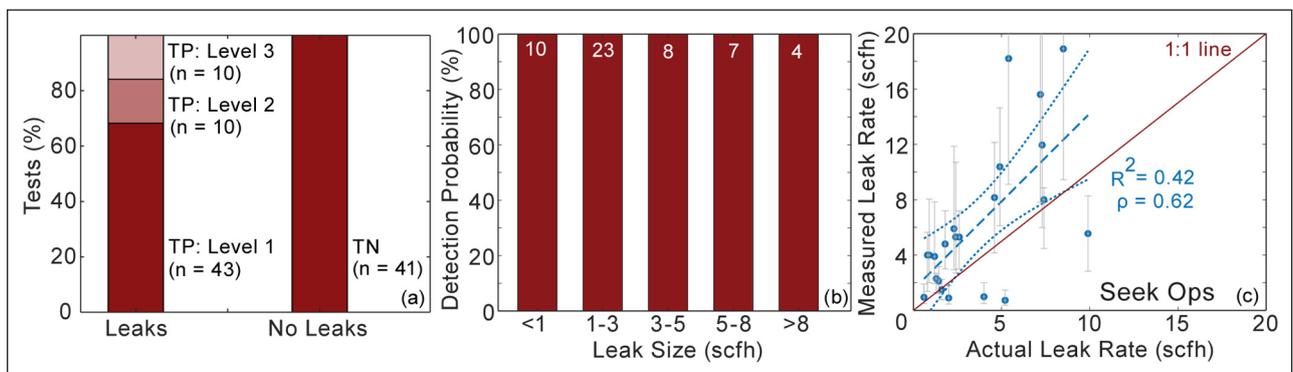


Figure 9: Performance results of Seek Ops Inc. in the Stanford/EDF Mobile Monitoring Challenge. (a) Binary detection characteristics of the technology, disaggregated by the level of true positive detection, **(b)** Leak detection probability across different leak size range, and **(c)** quantification parity chart between actual and measured leak rates. (TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative). DOI: <https://doi.org/10.1525/elementa.373.f9>

include results from the multiple leaks per pad scenarios, demonstrating the ability of Seek Ops algorithms to distinguish multiple closely-spaced emissions sources. The team did not have any FP detection. **Figure 9(b)** shows detection of 100% in all leak classes.

Figure 9(c) shows the quantification parity chart, with the error bars as directly reported by Seek Ops. The slope of the regression line is 1.27, suggesting overestimation of measured flux rates. The average difference between actual and measured leak rates is -2.6 scfh, with a 95% confidence interval between -4.3 and -0.8 scfh ($n = 63$), suggesting intercept (rather than slope) bias towards overestimation of leak rates.

3.9. University of Calgary (UC)

The University of Calgary (UC) team deployed two different technologies – a vehicle-based methane-only sensor, and a fixed-wing drone-based sensor. Both these technologies were tested between May 21–25 2018 near Sacramento, CA. We only include results from the truck-based system here, due to small number of flights with the fixed-wing drone. Results from the drone are presented in S.I. section 4.

The vehicle-based platform is fitted with a roof-mounted, methane-only open-path laser absorption sensor (LICOR LI-7700) that works on the principle of wavelength modulation spectroscopy, a 3D anemometer, and a vehicle position and orientation system. The platform, designed for both fence-line type measurements as well as fast-screening mode from public roads, collects data from all on-board instruments at 10 Hz.

Figure 10(a) shows the detection characteristics of the UC truck-based platform. TP rate is 94%, with $n = 55$ leaks (71%) at level-1 localization, and $n = 18$ leaks (23%) at level-3 localization. 15 of the 18 level-3 detects were from either source 1 (west) or source 2 (south) – interference from the non-test methane emissions from the site under appropriate wind conditions could have contributed to mis-identification. Mean FN flow rate is 121 scfh. A high FP rate (60%) could partly be due to interfering emissions sources from the front of the site.

Figure 10(b) shows the detection probability curve as a function of leak size. Leaks above 450 scfh have a 100% detection probability, even though all leaks are detected at the 80% level or higher. The lowest detection probability (82%) for leaks less than 150 scfh is consistent with the average FN flow rate of 121 scfh. The differences in detection probability across the range of leak sizes considered are not statistically significant.

Figure 10(c) shows the quantification parity chart of the technology, with the slope of the best-fit regression line being 0.4, indicating some underestimation of reported emissions. One reason for the underreporting could be attributed to data processing – the team subtracted the influence of the non-test emission at site by estimating its leak rate. However, the intermittent nature of the non-test leak could have resulted in an overestimation (instantaneous rate > average rate) thereby underestimating test scenario emissions. The average error between the actual and measured leak rate was 185 scfh (95% C.I. [137, 234], $n = 73$), confirming the over-estimation seen in the best-fit regression line.

4. Discussion

Table 2 summarizes the performance of these technologies along parameters chosen to highlight the collective capabilities of mobile systems as well as potential challenges ahead. All technologies are effective at detecting leaks, with 8 of the 9 tested technologies demonstrating a true positive leak rate of at least 75%. More importantly, 5 of 9 technologies show a near perfect true positive detection rate of 90% or higher – this shows the ability of technologies to detect leaks as small as 1 scfh. Despite this, the source attribution capability – denoted by the fraction of leaks detected at level-1 or level-2 (equipment-group level attribution) – varies significantly from 0% to 84%. Technologies such as ABB/ULC Robotics, Picarro, and BHGE largely confine their detection to pad-level attribution – leak repair and mitigation will require a complementary technology to identify emitting equipment and component. For technologies with high level-1 and level-2 detection capabilities, an OGI or similar

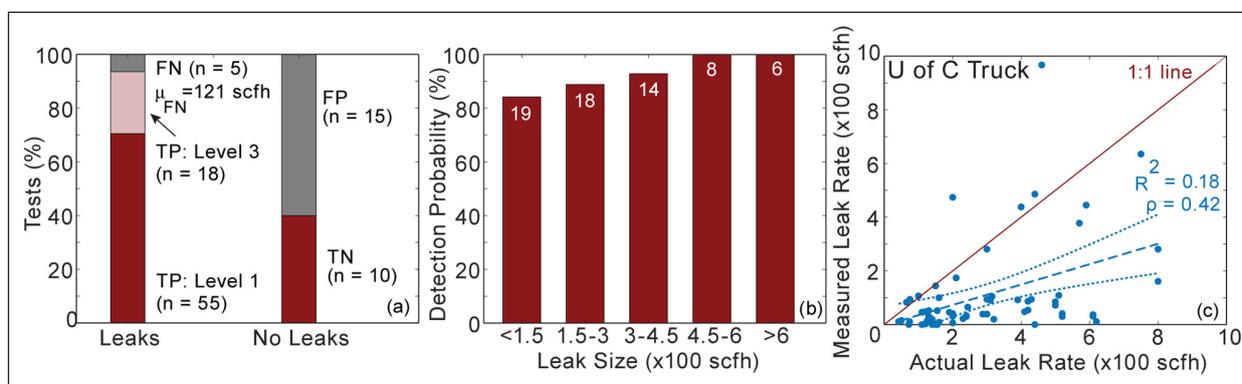


Figure 10: Performance results of University of Calgary (Truck) in the Stanford/EDF Mobile Monitoring Challenge. (a) Binary detection characteristics of the technology, disaggregated by the level of true positive detection, **(b)** Leak detection probability across different leak size range, and **(c)** quantification parity chart between actual and measured leak rates. (TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative). DOI: <https://doi.org/10.1525/elementa.373.f10>

Table 2: Summary of performance of the 9 technologies tested in the Stanford/EDF Mobile Monitoring Challenge. DOI: <https://doi.org/10.1525/elementa.373.t2>

Technology	Technology Type	Detection Effectiveness		False Positive Rate (%)	Detection Limit (leak rate where detection probability is 100%, scfh)	Quantification Accuracy ($\frac{\text{Measured}}{\text{Actual}}$, % tests)	
		True Positive (% , all levels)	True Positive (Levels 1–2(%))			0.5 – 2x*	0.1 – 10x* O(M)
ABB/ULC Robotics	Drone	77	0	22	≥8	30	78
Advisian	Drone	94	72	7	3–5	25	79
Aeris Technologies	Truck	88	65	15	5–8	38	79
Baker Hughes (GE)	Drone	68	33	71	≥8	24	54
Ball Aerospace	Plane	76	76	0	450–600	53	83
Heath Consultants	Truck	93	82	26	≥8	48	95
Picarro	Drone sampling	92	23	39	≥8	45	92
Seek Ops Inc.	Drone	100	84	0	≤1	36	100
U. Calgary (Truck)	Truck	94	71	60	450–600	18	74

* Fraction of tests where the measured emission rates are within (a) 0.5–2 times, and (b) 0.1–10 times of the actual emission rate.

technology may still be required to identify the leaking component and initiate repairs.

The false positive rate is an important indication of a system's ability to differentiate methane signal from noise. Methane is often present at elevated concentrations at oil and gas facilities, and the ability to distinguish natural variability from an emissions source is critical to effective mitigation. This is especially important for technologies that have small leak detection thresholds. Three technologies in this study had false positives rates lower than 10%, four more in the 15–40% range, and two technologies with false positive rates greater 50%. The high false positive rate in some of the technologies occurred despite a high leak detection rate. This indicates that sensor algorithms that process raw concentration data play an important role in the success and failure rate of these technologies. A combination of high sensitivity and ineffective algorithms can lead to high false positive rates because of an inability to clearly distinguish leak signal from background methane noise. Technologists should carefully consider the needs of the application – trade-offs between high sensitivity, high false positives, and quantification may be acceptable in some applications (rapid detection of 'super-emitters'), but unacceptable in others (quantifying mitigation potential, inventory). For technologies tested at the California site, the presence of non-test methane emissions from the site could have contributed to the high false positive rate for the University of Calgary vehicle-based technology.

All the technologies tested at METEC had detection limits lower than 10 scfh – in **Table 2**, we define the detection

limit as the leak rate beyond which the probability of detection is 100% under test conditions. Four of the technologies had a detection limit of at least 8 scfh, while two others were in the 3–8 scfh range. Because SeekOps identified all the leaks, we estimate that their detection limit is lower than 1 scfh. These numbers are comparable to the detection limits of OGI-based leak detection under ideal weather conditions (Ravikumar, Wang, McGuire, Bell, Zimmerle, & Brandt, 2018). Ball Aerospace's aerial system and University of Calgary's truck-based screening system have detection limits in the 450–600 scfh range – these rates are comparable to the 90th percentile of component-level emission rates found at oil and gas facilities (Brandt, Heath, & Cooley, 2016).

In general, quantification performance needs improvement. Most quantification efforts had appreciable errors in average leak rate or slope (or both). This is due to a fundamental issue: quantification of leakage rates from detected concentrations in downwind plumes is a challenging "inverse problem" that is a well-known hurdle in a number of scientific fields. Furthermore, typical plume inversion algorithms may require longer averaging time than the economics of mobile solutions would support. Some quantification results were sufficiently correlated with actual leak sizes that the resulting size estimates might be useful in a simple 3-class binning approach (i.e., small/medium/large to prioritize leak fixes). **Table 2** estimates the accuracy of quantification using two metrics – one, fraction of tests where measured emissions rates are between 0.5x and 2x of the actual emission rate, and two, fraction of tests where measured emission rates are within

an order of magnitude (0.1 – 10x) of the actual emission rate. Only Ball Aerospace estimated leaks within 0.5 – 2x of the actual leak rate in more than 50% of the tests. The overall performance on this metric ranged from a low of 18% to a high of 53%. This performance improves when considering an order of magnitude accuracy level – 8 of the 9 technologies estimated leak sizes to within an order of magnitude of the actual leak rate in at least 74% of test scenarios. In particular, Seek Ops, Heath Technologies Inc., and Picarro Inc. achieved an order of magnitude accuracy in 100%, 95%, and 92% of test scenarios, respectively. In general, the Pearson's coefficient (ρ) was larger than the linear regression coefficient (R^2), indicating that technologies are better at quantifying larger leaks compared to smaller leaks. Finally, the importance of quantification also depends on the application – rapid detection of large emissions sources for effective methane mitigation might not require accurate quantification.

Performance of the technologies are affected not only by inherent sensor capabilities but also factors such as environmental conditions, survey protocol, and facility characteristics. For example, technologies that use a suspended sample inlet (Advisian) or a tethered sample tube (Picarro Inc.) might face additional challenges in the presence of nearby power lines or taller equipment. An important source of error, given our test configuration, is inter-pad interference from wind-borne dispersion of leaks. To account for this, we analyzed the performance of teams tested at METEC under two scenarios – weak and strong interference (see S.I. section 5 and 6). These two analyzes sought to discard test results based on a set of criteria established to identify potential interference issues in leak detection. We found that under both weak and strong interference scenarios, the fraction of tests correctly identified (TPs and TNs) were not statistically different from base-case scenario where all tests were included. This suggests that whatever differences in performance that were observed between the teams did not arise from inter-pad interference.

Some technologies would be well served by re-testing at higher leak rates (>10 scfh). The combined testing format followed here requires supplying a range of leak sizes to satisfy multiple technologies at the same time. More detailed one-on-one testing could allow improved analysis of minimum detection rates and effectiveness. For example, BHGE performed well in the class of leaks >8 scfh and could be re-tested with more samples in that regime. This is especially important considering that a recent study of emissions in the Marcellus shale found that the average emission rate at the pad-level was 5.5 kg/h, corresponding to ~350 scfh (Caulton, et al., 2019). However, these are pad-level estimates, and component-level emissions can be significantly smaller – testing at the METEC facility between 0–15 scfh therefore provides a reasonable test of performance for technologies that detect emissions component-level detection. Conversely, testing at the Sacramento test location with emission rates in the 0–1500 scfh is well suited for technologies that detect aggregated pad-level emissions.

While no single technology can satisfy all the requirements for leak detection and quantification across the

natural gas supply chain, the results demonstrated here provide regulators and the industry with a range of options. There are technologies with strengths in survey speed that are suitable for leak detection along inter-state transmission pipelines, while technologies with high pad-level (but not equipment-level) detection effectiveness indicate potential use as a screening-technology to cover large areas. With potential improvements to algorithms that transform raw concentration data into actionable information, these technologies could become prominent tools to mitigate methane emissions.

A number of practicalities emerged in 3 weeks of testing that are relevant to any attempt to extrapolate these results to field conditions. First: drone technologies tested in this study are still immature, resulting in labor intensity, frequent battery recharge requirements, grounding due to winds, and substantial ground crew effort. Ground-based systems like the truck-mounted Heath and Aeris technologies experienced few of these issues and so have practical advantages that are not represented in above tables. At the same time, drone-based systems can be effective in quantifying emissions from taller equipment and during calm atmospheric conditions where plumes do not disperse but accumulate around the leak source – these conditions pose difficulty for truck-based systems where the plume lofts into the atmosphere and do not intersect the truck-based sensor. Second, drone-based technologies required accommodations that may be difficult to implement in real-world surveys: Advisian and Picarro dangled sample tubes from drones that has the potential to get tangled with equipment or nearby power lines, while SeekOps had a ground technician dedicated to traffic management and avoiding collisions due to the low-flying technique. The employed deployment methods may cause practical difficulties in labor cost and survey time with usage of the technology but will hopefully be solved by technology development.

Even as this study provides the first controlled and independent verification of the performance of mobile leak detection technologies, this is only one step in the road to demonstrating that these technologies will provide emissions reductions that are equivalent to traditional OGI-based methods. Demonstrating equivalence with OGI will require more testing and assessing the performance of these technologies under specific survey protocols (Ravikumar, & Brandt, 2017). Whether the emissions reductions from monthly truck-based screening surveys, for example, are equivalent to emissions reductions from semi-annual OGI-based LDAR survey can be answered through a statistical simulations (for example, using the FEAST simulation platform) (Kemp, Ravikumar, & Brandt, 2016) as well as pilot testing these technologies at oil and gas facilities with co-occurring OGI studies (Fox, et al., 2019). Clearly, the next frontier in mobile methane emissions mitigation is to develop standardized protocols to demonstrate technology equivalence for use across large geographic areas.

It is critical to remember that these results apply to the technologies that are in active development. Many of the systems tested here have undergone changes to both hardware and software since they were tested for this

study. It is a rapidly evolving field and stakeholders should always look for the most recent data to make decisions on deployment and regulatory acceptance.

Supplemental files

The supplemental files for this article can be found as follows:

- **Text S1.** Single-blind inter-comparison of methane detection technologies – Results from the Stanford/EDF mobile monitoring challenge. DOI: <https://doi.org/10.1525/elementa.373.s1>
 - Section 1.** Selection process.
 - Section 2.** Test locations and site configurations.
 - Section 3.** Test protocols.
 - Section 4.** Results from the University of Calgary drone system.
 - Section 5.** Interference analysis: methods.
 - Section 6.** Interference analysis: results.
- **Appendix A.** Stanford/EDF Mobile Monitoring Challenge Application Form. DOI: <https://doi.org/10.1525/elementa.373.s2>
- **Data S1.** MMCResults_Heath. DOI: <https://doi.org/10.1525/elementa.373.s3>
- **Data S2.** MMCResults_Picarro. DOI: <https://doi.org/10.1525/elementa.373.s4>
- **Data S3.** MMCResults_ABB_ULC. DOI: <https://doi.org/10.1525/elementa.373.s5>
- **Data S4.** MMCResults_Advisian. DOI: <https://doi.org/10.1525/elementa.373.s6>
- **Data S5.** MMCResults_Aeris. DOI: <https://doi.org/10.1525/elementa.373.s7>
- **Data S6.** MMCResults_BHGE. DOI: <https://doi.org/10.1525/elementa.373.s8>
- **Data S7.** MMCResults_SeekOps. DOI: <https://doi.org/10.1525/elementa.373.s9>
- **Data S8.** MMCResults_BallAerospace. DOI: <https://doi.org/10.1525/elementa.373.s10>
- **Data S9.** MMCResults_UofCTruck. DOI: <https://doi.org/10.1525/elementa.373.s11>
- **Data S10.** MMCResults_UofCDrone. DOI: <https://doi.org/10.1525/elementa.373.s12>
- **Data S11.** MMC_WeakInterferenceAnalysis. DOI: <https://doi.org/10.1525/elementa.373.s13>
- **Data S12.** MMC_StrongInterferenceAnalysis. DOI: <https://doi.org/10.1525/elementa.373.s14>

Acknowledgements

We appreciate Mike McGuire (formerly with Colorado State University) for his assistance with conducting the field tests at METEC. We also thank the members of the industry advisory board for providing valuable insight on the needs of leak detection at oil and gas facilities, as well as helping the scientific team to understand the commercial viability of the applicants. The industry advisory board included experts from Exxon-Mobil, Schlumberger, Noble Energy, Shell, Equinor, and PG&E. A.P.R. would especially like to thank Nuance Chocolate Inc. (Fort Collins, CO) and the Fat Tire Brewing Company for providing much needed respite after long days in the field.

Funding information

The authors acknowledge funding from the Stanford Natural Gas Initiative and the Environmental Defense Fund.

Competing interests

The authors have no competing interests to declare.

Author contributions

- Substantial contributions to conception and design: APR, ARB
- Acquisition of data: APR, SS, JW, JE, DR, CB
- Analysis and interpretation of data: APR, ARB, SS, DL, DZ, IM, BR
- Drafting the article or revising it critically for important intellectual content: APR, ARB, SS, DZ, IM, BR
- Final approval of the version to be published: All authors

References

- Alvarez, RA, Zavala-Araiza, D, Lyon, DR, Allen, DT, Barkley, ZR, Brandt, AR, Davis, KJ, Herndon, SC, Jacob, DJ, Karion, A, Kort, EA, Lamb, BK, Lauvaux, T, Maasakkers, JD, Marchese, AJ, Omara, M, Pacala, SW, Peischl, J, Robinson, AL, Shepson, PB, Sweeney, C, Townsend-Small, A, Wofsy, SC and Hamburg, SP. 2018. Assessment of methane emissions from the U.S. oil and gas supply chain. *Science* **361**: 186–188. DOI: <https://doi.org/10.1126/science.aar7204>
- Atherton, E, Risk, D, Fougere, C, Lavoie, M, Marshall, A, Werring, J, Williams, JP and Minions, C. 2017. Mobile measurements of methane emissions from natural gas developments in northeastern British Columbia, Canada. *Atmos. Chem. Phys.* **17**: 12405–12420. DOI: <https://doi.org/10.5194/acp-17-12405-2017>
- Barchyn, T, Hugenholtz, C, Myshak, C and Bauer, J. 2017. A UAV-based system for detecting natural gas leaks. *J. Unmanned Vehicle Systems* **6**: 18–30. DOI: <https://doi.org/10.1139/juvs-2017-0018>
- Brandt, A, Heath, G and Cooley, D. 2016. Methane leaks from natural gas systems follow extreme distributions. *Environ. Sci. Technol.* **50**: 12512–12520. DOI: <https://doi.org/10.1021/acs.est.6b04303>
- Caulton, DR, Lu, JM, Lane, HM, Buchholz, B, Fitts, JP, Golston, LM, Guo, X, Li, Q, McSpirtt, J, Pan, D, Wendt, L, Bou-Zeid, E and Zondlo, MA. 2019. Importance of superemitter natural gas well pads in the Marcellus shale. *Environmental Science & Technology* **53**(9): 4747–4754. DOI: <https://doi.org/10.1021/acs.est.8b06965>
- Colorado Department of Public Health and Environment. 2014. Regulation Number 7: Control of Ozone via Ozone Precursors and Control of Hydrocarbons via Oil and Gas Emissions. 5 CCR 1001-9. Denver.
- Conley, S, Franco, G, Faloona, I, Blake, DR, Peischl, J and Ryerson, TB. 2016. Methane emissions from the 2015 Aliso Canyon blowout in Los Angeles, CA. *Science*. DOI: <https://doi.org/10.1126/science.aaf2348>

- Englander, J, Brandt, A, Conley, S, Lyon, D and Jackson, R.** 2018. Aerial Interyear Comparison and Quantification of Methane Emissions Persistence in the Bakken Formation of North Dakota, USA. *Environ. Sci. Technol.* **52**: 8947–8953. DOI: <https://doi.org/10.1021/acs.est.8b01665>
- Environment and Climate Change Canada.** 2018. Regulations Respecting Reduction in the Release of Methane and Certain Volatile Organic Compounds (Upstream Oil and Gas Sector). SOR, 2018-66.
- Etminan, M, Myhre, G, Highwood, E and Shine, K.** 2016. Radiative forcing of carbon dioxide, methane and nitrous oxide: A significant revision of the methane radiative forcing. *Geophys. Res. Lett.* **43**: 12614–12623. DOI: <https://doi.org/10.1002/2016GL071930>
- Fox, T, Barchyn, T, Risk, D, Ravikumar, A and Hugenholtz, C.** 2019. A review of close-range and screening technologies for mitigating fugitive methane emissions in upstream oil and gas. *Environmental Research Letters* **14**(5): 053002. DOI: <https://doi.org/10.1088/1748-9326/ab0cc3>
- Fox, TA, Ravikumar, AP, Hugenholtz, CH, Zimmerle, D, Barchyn, TE, Johnson, MR, Lyon, D and Taylor, T.** 2019. A methane emissions reduction equivalence framework for alternative leak detection and repair programs. *Elem. Sci. Anth.* **7**(1): 30. DOI: <https://doi.org/10.1525/elementa.369>
- Frankenberg, C, Thorpe, AK, Thompson, DR, Hulley, G, Kort, EA, Vance, N, Borchardt, J, Krings, T, Gerilowski, K, Sweeney, C, Conley, S, Bue, BD, Aubrey, AD, Hook, S and Green, RO.** 2016. Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region. *Proc. Natl. Acad. Sciences.* DOI: <https://doi.org/10.1073/pnas.1605617113>
- Golston, LM, Aubut, NF, Frish, MB, Yang, S, Talbot, RW, Gretencord, C, McSpirt, J and Zondlo, MA.** 2018. Natural gas fugitive leak detection using an unmanned aerial vehicle: localization and quantification of emission rate. *Atmosphere* **9**(9): 333. DOI: <https://doi.org/10.3390/atmos9090333>
- Intergovernmental Panel on Climate Change.** 2018. *Global Warming of 1.5 C.* New York, NY: United Nations.
- Jacob, DJ, Turner, AJ, Maasackers, JD, Sheng, J, Sun, K, Liu, X, Chance, K, Aben, I, McKeever, J and Frankenberg, C.** 2016. Satellite observations of atmospheric methane and their value for quantifying methane emissions. *Atmos. Chem. Phys.* **16**: 14371–14396. DOI: <https://doi.org/10.5194/acp-16-14371-2016>
- Johnson, MR, Tyner, DR, Conley, S, Schwietzke, S and Zavala-Araiza, D.** 2017. Comparisons of Airborne Measurements and Inventory Estimates of Methane Emissions in the Alberta Upstream Oil and Gas Sector. *Environ. Sci. Tech.* **51**: 13008–13017. DOI: <https://doi.org/10.1021/acs.est.7b03525>
- Keating Research Inc.** 2016. *The Colorado case study on methane emissions: conversations with the oil and gas industry.* Telluride, CO.
- Kemp, C, Ravikumar, A and Brandt, A.** 2016. Comparing natural gas leakage detection technologies using an open-source virtual gas field simulator. *Environ. Sci. Technol.* **50**: 104–115. DOI: <https://doi.org/10.1021/acs.est.5b06068>
- Lyon, D, Alvarez, R, Zavala-Araiza, D, Brandt, A, Jackson, R and Hamburg, S.** 2016. Aerial surveys of elevated hydrocarbon emissions from oil and gas production sites. *Environ. Sci. Technol.* **50**: 4877–4886. DOI: <https://doi.org/10.1021/acs.est.6b00705>
- Maric, I.** 2005. The Joule-Thomson effect in natural gas flow-rate measurements. *Flow Meas. Instrum.* **16**: 387–395. DOI: <https://doi.org/10.1016/j.flowmeasinst.2005.04.006>
- Myhre, G, Shindell, D, Breon, F, Collins, W, Fuglestedt, J, Huang, J, Koch, D, Lamarque, J, Lee, D, Mendoza, B, Nakajima, T, Robock, A, Stephens, G, Takemura, T, Zhang, H, Aamaas, B, Boucher, O, Dalsoren, SB, Daniel, JS, Forster, P, Granier, C, Haigh, J, Hodnebrog, O, Kaplan, JO, Marston, G, Nielsen, CJ, O'Neill, BC, Peters, GP, Pongratz, J, Prather, M, Ramaswamy, V, Raphael, R, Rotstayn, L, Smith, SJ, Stevenson, D, Vernier, J, Wild, O and Young, P.** 2013. Anthropogenic and Natural Radiative Forcing. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 1–44. New York: Cambridge University Press.
- Nathan, BJ, Golston, LM, O'Brien, AS, Ross, K, Harrison, WA, Tao, L, Lary, DJ, Johnson, DR, Covington, AN, Clark, NN and Zondlo, MA.** 2015. Near-field characterization of methane emission variability from a compressor station using a model aircraft. *Environmental Science & Technology* **49**(13): 7896–7903. DOI: <https://doi.org/10.1021/acs.est.5b00705>
- Ravikumar, A and Brandt, A.** 2017. Designing better methane mitigation policies: the challenge of distributed small sources in the natural gas sector. *Env. Res. Lett.* **12**: 040023. DOI: <https://doi.org/10.1088/1748-9326/aa6791>
- Ravikumar, A, Wang, J and Brandt, A.** 2017. Are optical gas imaging techniques effective for methane leak detection? *Environ. Sci. Technol.* **51**: 718–724. DOI: <https://doi.org/10.1021/acs.est.6b03906>
- Ravikumar, A, Wang, J, McGuire, M, Bell, C, Zimmerle, D and Brandt, A.** 2018. “Good versus Good Enough?” Empirical Tests of Methane Leak Detection Sensitivity of a Commercial Infrared Camera. *Environ. Sci. Technol.* **52**: 2368–2374. DOI: <https://doi.org/10.1021/acs.est.7b04945>
- Southwest Research Institute.** 2015. *Testing of methane detection systems – Phase 2.* Washington, DC: Environmental Defense Fund.
- Turner, A, Frankenberg, C, Wennberg, P and Jacob, D.** 2017. Ambiguity in the causes for decadal trends in atmospheric methane and hydroxyl. *Proc. Natl. Acad. Sciences* **114**: 5367–5372. DOI: <https://doi.org/10.1073/pnas.1616020114>

- U.S. Advanced Research Projects Agency (ARPA-E).** 2014. *Methane Observation Networks with Innovative Technology to Obtain Reductions*. Washington, DC.
- Yuan, B, Kaser, L, Karl, T, Graus, M, Peischl, J, Campos, TL, Shertz, S, Apel, EC, Hornbrook, RS, Hills, A, Gilman, JB, Lerner, BM, Warneke, C, Flocke, FM, Ryerson, TB, Guenther, AB and de Gouw, JA.** 2015. Airborne flux measurements of methane and volatile organic compounds over the Haynesville and Marcellus shale gas production regions. *J. Geophys. Res. Atmos.* **120**: 6271–6289. DOI: <https://doi.org/10.1002/2015JD023242>
- Zavala-Araiza, D, Alvarez, RA, Lyon, DR, Allen, DT, Marchese, AJ, Zimmerle, DJ and Hamburg, SP.** 2017. Super-emitters in natural gas infrastructure are caused by abnormal process conditions. *Nature Communications* **8**: 1–10. DOI: <https://doi.org/10.1038/ncomms14012>
- Zickfeld, K, Solomon, S and Gilford, D.** 2017. Centuries of thermal sea-level rise due to anthropogenic emissions of short-lived greenhouse gases. *Proc. Natl. Acad. Sciences* **114**: 657–662. DOI: <https://doi.org/10.1073/pnas.1612066114>

How to cite this article: Ravikumar, AP, Sreedhara, S, Wang, J, Englander, J, Roda-Stuart, D, Bell, C, Zimmerle, D, Lyon, D, Mogstad, I, Ratner, B and Brandt, AR. 2019. Single-blind inter-comparison of methane detection technologies – results from the Stanford/EDF Mobile Monitoring Challenge. *Elem Sci Anth*, 7: 37. DOI: <https://doi.org/10.1525/elementa.373>

Domain Editor-in-Chief: Detlev Helmig, Institute of Alpine and Arctic Research, University of Colorado Boulder, US

Guest Editor: Brian Lamb, Washington State University, US

Knowledge Domain: Atmospheric Science

Part of an *Elementa* Forum: Oil and Natural Gas Development: Air Quality, Climate Science, and Policy

Submitted: 16 April 2019 **Accepted:** 19 August 2019 **Published:** 10 September 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Elem Sci Anth is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS