Supplemental Material

Single-blind Inter-comparison of Methane Detection Technologies – Results from the Stanford/EDF Mobile Monitoring Challenge

Arvind P. Ravikumar*[1], Sindhu Sreedhara[2], Jingfan Wang[2], Jacob Englander[2β], Daniel Roda-Stuart[2†], Clay Bell[3], Daniel Zimmerle[3], David Lyon[4], Isabel Mogstad[4], Ben Ratner[4], Adam R. Brandt[2]

[1]Harrisburg University of Science and Technology, Harrisburg, Pennsylvania, US

[2]Stanford University, Stanford, California, US

[3]Colorado State University Energy Institute, Fort Collins, Colorado, US

[4]Environmental Defense Fund, Washington DC, US

*Corresponding author: aravikumar@harrisburgu.edu

August 16, 2019

[β]Current affiliation: California Air Resources Board, Sacramento, California, US

[†]Current affiliation: Alphataraxia Management, Los Angeles, California, US

List of Contents:

SM_Figure 1. Test site configuration at METEC. Site configuration at the Methane Emissions Technology Evaluation Center (METEC) in Fort Collins, CO, indicating typical oil and gas equipment arrange in 5 'pad' configurations.

SM_Figure 2. Test site configuration in California. Site configuration at the Northern California Gas Yard in Knights Landing, CA, showing the approximately locations of the three leak sources. The red circle indicates the location of the anomalous methane emissions from the facility during the test week.

SM_Figure 3. Performance results of University of Calgary (drone) in the Stanford/EDF Mobile Monitoring Challenge. Quantification parity chart between actual and measured leak rates. The drone technology was tested only on one of the days due to airspace restrictions resulting in a small sample size.

SM_Figure 4. Wind speed distribution at METEC during the Stanford/EDF Mobile Monitoring Challenge. Histogram of the 5-minute wind speed collected at METEC during the two weeks of testing. Winds greater than 15 mph were observed 23% of the time during week, and less than 2% the time during week 2.

SM_Figure 5. An example of weak-interference scenario during Week-1 of testing at METEC. The 10-minute binary yes/no detection test scenario had a 2 m/s average wind from 202.6° and leaks on pads 1 and 4. Because pad 4 could experience potential interference from pad 1, the results from pad 4 were discarded from statistical analysis. Here, the 40° cone contained 80% of all wind vectors in the 10-minute test period.

SM_Figure 6. An example of strong-interference scenario during Week-2 of testing at METEC. The 20-minute detection and quantification test scenario had a 1.4 m/s average wind from 58.6° and leaks on pads 3, 4, and 5. Because pad 2 and pad 4 could experience potential interference from pad 5, the results from pad 2 and pad 4 were discarded from statistical analysis. Here, the 40° cone did not contain at least 50% of wind vectors in the 20-minute test interval and was therefore expanded to 49°.

SM_Figure 7. Interference analysis of Heath Consultants' performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

SM_Figure 8. Interference analysis of Picarro Inc.'s performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

SM_Figure 9. Quantification parity chart for Picarro under the weak- and strong-interference scenarios. While there was marginal improvement in R^2, the average error also increased from -0.9 scfh to -1.3 scfh. The parameter values in parenthesis represent base-case scenario.

SM_Figure 10. Interference analysis of SeekOps' performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

SM_Figure 11. Interference analysis of Aeris Technologies' performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

SM_Figure 12. Interference analysis of Advisian's performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

SM_Figure 13. Interference analysis of ABB/ULC Robotics' performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

SM_Figure 14. Interference analysis of Baker Hughes' (GE) performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

SM_Table 1: Specification of the participating technologies. Self-reported performance parameters of the participating technologies. These values correspond to specification as available during the application process for the Stanford/EDF Mobile Monitoring Challenge (October 2017) and can be substantially differently now. Actual field performance may not adhere to these specifications (see main text for results).

SM_Table 2. Stanford/EDF Mobile Monitoring Challenge overview. Test dates, participating teams, test locations, and self-reported detection limits from all the teams participating the Stanford/EDF Mobile Monitoring Challenge

SM_Table 3. Test Protocols in the Stanford/EDF Mobile Monitoring Challenge. General test protocols in the Stanford/EDF Mobile Monitoring Challenge, increasing in complexity from simple yes/no detection tests to complex multi-leak detection and quantification tests.

SM_Table 4. Interference results from Heath Consultants. Performance of Heath Consultants in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

SM_Table 5. Interference results from Picarro Inc. Performance of Picarro Inc. in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

SM_Table 6. Interference results from SeekOps. Performance of Seek Ops Inc. in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

SM_Table 7. Interference results from Aeris Technologies. Performance of Aeris Technologies in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

SM_Table 8. Interference results from Advisian. Performance of Advisian in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

SM_Table 9. Interference results from ABB/ULC Robotics. Performance of ABB/ULC Robotics in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

SM_Table 10. Interference results from Baker Hughes GE. Performance of Baker Hughes (GE) in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

1. Selection process

The selection process for the MMC was undertaken in collaboration with scientists, project managers, and an industrial advisory board set up to advise on the commercial promise of new methane leak detection systems. The 10 technologies that participated in the MMC were chosen in three steps. In the first step, we sought applications from potential participants who submitted written answers to questions on the company, sensor technology, and commercial potential (see Appendix A). We received 28 applications from 5 countries in the 65-day period when applications were accepted. In the second step, scientists at Stanford and EDF, project managers, and the industry advisory board individually evaluated submitted applications on suitability to project goals, scientific capability of the technology, and its commercial potential. In the third step, the evaluators convened in Houston to discuss each application and select the 12 most promising technologies. Out of the 12 initially selected, 2 were unable to participate due to delays in technology development or limited capability to measure under diverse conditions. 10 technologies from 9 institutions participated in the field trials. Detailed, as-reported, specifications of each of technologies are given below in Table T1.

SM_Table 1: Specification of the participating technologies. Self-reported performance parameters of the participating technologies. These values correspond to specification as reported during the application process for the Stanford/EDF Mobile Monitoring Challenge (October 2017) and can be substantially differently now. Actual field performance may not adhere to these specifications (see main text for results).

Technology (Platform)	**Sensor Precision**	**Data Rates**	**Measurement Approach**	**Quantification Approach**	**Quantification Uncertainty**
ABB/ULC Robotics (Drone)	2 ppb (1 Hz) Detection range: 5 ppb – 8000 ppm	Up to 10 Hz	Cavity-enhanced laser absorption spectroscopy – Methane	Modified raster scan (wind responsive)	Not available
Advisian (helicopter)	‘ppb range’ – exact numbers unavailable at time of test		Laser absorption spectroscopy – methane/ethane	Inverse dispersion modeling using concentration maps and in-situ wind data	Depends on wind speed (In this study, average error was -17%/73%)
Aeris Technologies (Vehicle)	1 ppb (1 s) for both methane and ethane	1 – 2 Hz (data log) ~ kHz (intrinsic rate)	Laser absorption spectroscopy in the mid-wave infrared	Emission rates determined using concentration, spatial, and wind information collected in-situ	Not available (see main text)

Baker Hughes (GE) (Drone)	Detection range: 1 – 50,000 ppm-m	2 Hz	Laser absorption spectroscopy produces 2D concentration heat-map	2-D concentration heat maps are combined with wind data to estimate leak size	± 10% (as provided by sensor manufacturer)
Ball Aerospace (Plane)	50 ppm-m above background < 2 m spatial resolution	10,000 Hz	Airborne differential LIDAR around 1650 nm (whisk-broom sensing approach to provide swath image of gas concentration)	Emissions quantified using methane concentration map along with local wind speeds (altitude ~ 3000 ft)	Depends on wind-speed & atmospheric stability (±50% during this study)
Heath Consultants Inc. (Vehicle)	2 ppb (1s) – methane 10 ppb (1s) - ethane	Up to 5 Hz	Off-axis integrated cavity output spectroscopy – methane, ethane	Emissions quantified by combining measurements of gas concentration, local coordinates, and wind conditions,	Not Available
Picarro (Drone and Vehicle)	3 ppb (1s) – methane 10 ppb (1s) - ethane	1 Hz (approx.)	Cavity ringdown spectroscopy – methane, ethane, water-vapor	Flux difference using upwind and downwind transect measurements	70% confidence between 0.5x – 2x
Seek Ops Inc. (drone)	10 ppb (1s) (handheld) 50 ppb (1s) (UAV)	4 Hz (typical), Up to 100 Hz	Tunable diode spectrometer in the mid-wave infrared	Point concentrations & wind data for source localization	±20% (10 – 500 scfh)
U Calgary (Vehicle)	5 ppb (10 Hz)	10 Hz	LICOR LI-7700 open-path wavelength-modulated laser spectroscopy	Data from 3 sensors (sonic anemometer, methane sensor, and vehicle position & orientation system combined to provide localization and quantification.	Not Available (see main text for data)

2. Test locations and site configurations

Two test locations were chosen for field trials – the Methane Emissions Technology Evaluation Center (METEC) in Fort Collins, CO and the Northern California Gas Yard operated by Rawhide Leasing in Knights Landing, CA (40 miles north of Sacramento, CA). Technologies were assigned either of the test location based on their minimum detection limits as described in their application forms and conversations with the teams prior to testing. The final assignments, associated minimum detection limits, and test dates are shown in the table below.

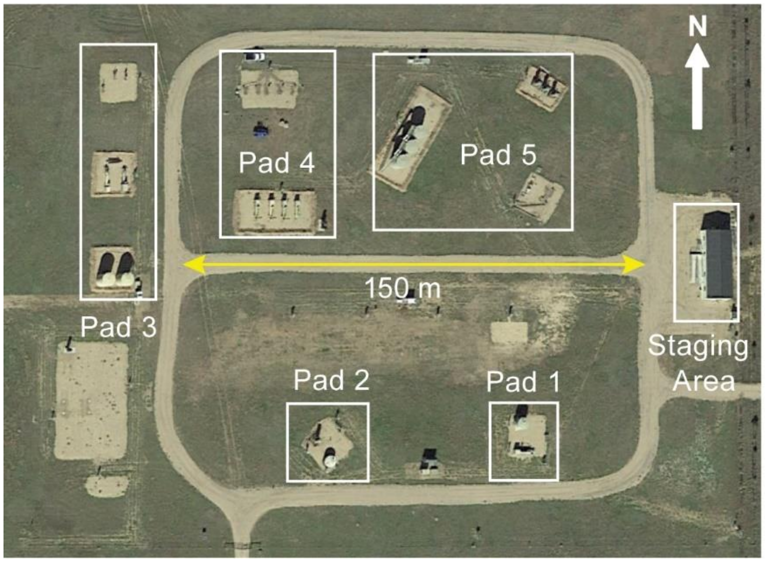
SM_Table 2. Stanford/EDF Mobile Monitoring Challenge overview. Test dates, participating teams, test locations, and self-reported detection limits from all the teams participating the Stanford/EDF Mobile Monitoring Challenge

Test Dates	Participating Teams	Location	Self-reported detection limits
9 – 13 April 2018	Heath Technologies (vehicle) Picarro Inc. (drone/vehicle)	METEC, Fort Collins, CO	1 – 5 scfh
23 – 27 April 2018	Baker Hughes GE (drone), Seek Ops Inc. (drone), Aeris Technologies (vehicle), Advisian (drone), ABB/ULC Robotics (drone)	METEC, Fort Collins, CO	5 – 10 scfh
21 – 25 May 2018	Ball Aerospace (plane), Univ. of Calgary (vehicle), Univ. of Calgary (drone)	Northern CA gas yard, Sacramento, CA	> 100 scfh

2.1. METEC, Fort Collins, CO

METEC is a controlled release test facility funded by the Department of Energy (ARPA-E) MONITOR program and managed by the Colorado State University [1]. The facility consists of typical equipment found at an oil and gas production site including wellheads, separators, and tanks (see Figure S1). These are organized into five pads as shown below. The complexity of the pad varies based on the number of equipment of each type present on the pad – Pads 1 and 2 are the simplest with 1 wellhead, 1 tank, and 1 separator, each. Pad 5 is the most complex with 3 separators, 3 well heads, and 3 tanks. Each equipment had multiple potential emission points made from 1/4” stainless steel tubing concealed to make leaks appear from typical components like flanges and valves.

Flowrates were controlled using the line pressure in the system and monitored using an Omega FMA1700 Series thermal mass flow meter calibrated for use with methane. Gas composition is calculated using a gas chromatograph at the CSU Energy Institute. Composition during these tests were in the following ranges: was 86.7% (± 0.9%) CH_4, 9.9% (± 0.1%) C_2H_6, 0.7% (± 0.2%) C_3H_8, 1.4% (± 0.6%) N_2, and 1.3% (± 0.01%) CO_2 based on 6 repeat gas chromatography samples. The metered rate accurately represented the release rate during single leak tests. For multiple-leak tests, each leak rate (and corresponding inlet pressure and valve position) was individually calibrated before being simultaneously released. The flow rates for individual test scenarios are provided as supplementary excel files.



SM_Figure 1: Test site configuration at METEC. Site configuration at the Methane Emissions Technology Evaluation Center (METEC) in Fort Collins, CO, indicating typical oil and gas equipment arrange in 5 ‘pad’ configurations.

2.2. Northern California Gas Yard – Knights Landing, CA

The Northern California Gas Yard (Rawhide Leasing) test-site location was chosen to accommodate teams whose minimum detection limits were significantly higher than the maximum leak rate available at METEC. The test site configuration, shown in Figure S2, consists of three leak sources, each a 6 feet elevated stack of 1” diameter. The sources were within ~200 ft of each other, with each source individually metered using a Sierra Instruments QuadraTherm 740i thermal mass flow meters with an accuracy of ±0.75% of full-scale reading. Line gas with a composition of 91% CH_4, 6% C_2H_4, 2% N_2, 1% trace gases was sourced from a 2500 psi pressurized tank. The line pressure was reduced to 50 psi using a regulator before being flow through the flow meters and the sources. Because most of the test scenarios had leak rates ranging from 50 scfh to about 400 scfh, we did not experience any substantial Joule-Thompson related cooling effect. In addition to the test-related methane emissions, the California site also had intermittent unintended methane releases from the front of the facility (see Figure S2) from a compressor station and a storage tank – the University of Calgary truck team explicitly accounted for this anomalous emissions source in their analysis by subtracting an estimate of the emissions from the test-scenario emission rate. This compressor station did not run continuously, and the source of the non-test methane emissions was identified to be the vent on the tank though a FLIR GF-320 infrared camera. The location of the vent prevented us from directly quantifying the measurement. The flow rates for individual test scenarios are provided as supplementary excel files.

SM_Figure 2. Test site configuration in California. Site configuration at the Northern California Gas Yard in Knights Landing, CA, showing the approximately locations of the three leak sources. The red circle indicates the location of the anomalous methane emissions from the facility during the test week.

3. Test protocols

Test protocols varied based on the test site, number of teams participating during the test week, and weather conditions. In general, the tests were designed to increase in complexity to progressively test the capabilities of the technologies. Table T2 shows the four major test scenarios used across all the teams and facilities.

SM_Table 3. Test Protocols in the Stanford/EDF Mobile Monitoring Challenge. General test protocols in the Stanford/EDF Mobile Monitoring Challenge, increasing in complexity from simple yes/no detection tests to complex multi-leak detection and quantification tests.

Test Name	Test Description	Number of Leaks per Test	Is Quantification Required?	Maximum Duration (minutes)
Binary Yes/No	Yes/No type detection test	0 – 1	No	10

	Pad known			
Binary Yes/No	Yes/No type detection test Pad known	0 – 1	No	5
Single Leak Quantification	Detection and quantification Pad known	0 – 1	Yes	20
Multi-Leak Quantification	Detection and quantification Pad known	0 – 3	Yes	20

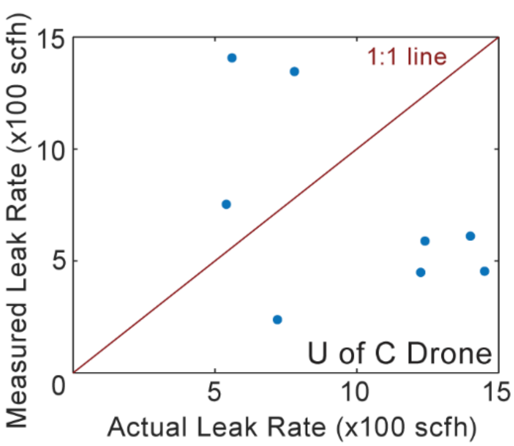
Week 1 test protocols: Two teams tested in week-1 at METEC (Heath Consultants and Picarro Inc.), with each team being assigned a starting pad. For a 10-minute detection only test, both the teams rotate through all four pads in a clockwise direction until each team has tested on all the pads. All pads did not necessarily have leaking components, and the teams were also tested on their ability to detect true negative (and false positive) tests. In addition, we also selectively turned on pads to prevent wind related interference (see analysis below). While the authors A.P.R., M.M., and C.B. controlled the release rates from the staging area during the test, S.S., D.R., J.E., and J.W. assisted with managing the measurement teams on the site. The team managers were not aware of the leak rates or leak locations during the testing.

Week 2 test protocols: Five teams tested in week-2 at METEC – ABB/ULC, Advisian, Aeris, BHGE, and SeekOps. Each team was initially assigned a pad and were rotated clockwise to the next adjacent pad every 30 – 40 minutes. Therefore, under a 10-minute test scenario, each team tested for 3 different leak scenarios on a given pad before moving on to the next pad. This rotation frequency was chosen to minimize the time spend in moving between the pads while also providing for frequent enough rotation such that the weather conditions did not dramatically change before each team tested on all 5 pads (approx. 1 hour). Like week 1, team managers (S.S., D.R., J.E., and J.W.) did not know the test scenarios assigned by the authors A.P.R, M.M, and C.B. In order to account for changing weather conditions, the test scenarios were adjusted in real-time by assigning leaks on pads that minimized inter-pad interference.

Week 3 test protocols: Three teams tested in week-3 at the Northern California Gas Yard in Knights Landing, CA – Ball Aerospace, University of Calgary Truck, and the University of Calgary Drone. Because of airspace conflict with local crop-dusters, the drone system was only able to fly on one of the days of the test and hence statistically significant results could not be obtained. Both the aerial and truck team on this week were tested simultaneously during each test scenario because their survey protocols did not present any logistical difficulty. Furthermore, conducting simultaneous measurements also increased the sample size of test scenarios. Because the teams were measuring the same leak during each test, there was no possibility of interference.

4. Results from the University of Calgary drone system

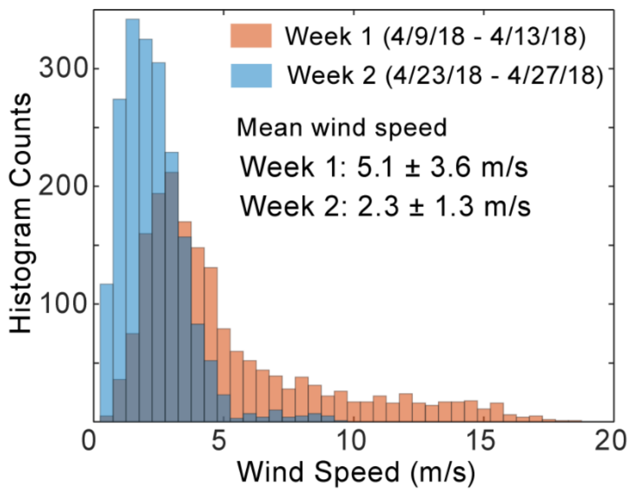
The University of Calgary's UAV-based system, developed in collaboration with Ventus Geospatial, is fitted with a Boreal Laser GasFinder 2 open path laser spectrometer. This sensor is integrated into a C-Astral Bramor UAV that employs a catapult launcher for take-off and a parachute to land. Additional details on the technology can be found in Barchyn et al. [2]. The UAV also collects data on wind direction, speed, and UAV coordinates at 4 Hz frequency, has a flying time of about 2 hours, and requires open fields for launch and landing. We were able to test the performance of the drone on only one of the days (5/9/2018) because of flight restrictions associated with the use of crop dusters in the surrounding rice paddies. While we do report results from this testing, the sample size is too small (n = 8) to draw statistical inferences from the drone-based sensor results. Figure S3 shows the quantification parity chart for the UAV technology (n = 8) – although we observe a leak under-estimation at leak rates > 1000 scfh, small sample size prevents us from drawing any definitive conclusions. Further testing is required to fully characterize this technology.



SM_Figure 3. Performance results of University of Calgary (drone) in the Stanford/EDF Mobile Monitoring Challenge. Quantification parity chart between actual and measured leak rates. The drone technology was tested only on one of the days due to airspace restrictions resulting in a small sample size.

5. Interference Analysis

The orientation of the pads at METEC and the simultaneous testing of technologies can potentially result in inter-pad leak interference. For example, consider that winds are blowing from the north and that there is a leak on pad 5 but not on pad 1 (Figure S1). A technology testing on pad 1 could detect methane from the leak on pad 5 dispersed downwind and result in a false positive identification. This is especially important consideration when wind speeds are high. Figure S4 shows a histogram of the 5-minute average wind speed collected during testing at METEC in both weeks 1 and 2. Tests conducted during week 1 observed an average wind speed of 5.1 m/s, compared to 2.3 m/s in week 2. Winds greater than 15 mph were experienced less than 2% of the time in week 2, compared to 23% of the time in week 1.



SM_Figure 4. Wind speed distribution at METEC during the Stanford/EDF Mobile Monitoring Challenge. Histogram of the 5-minute wind speed collected at METEC during the two weeks of testing. Winds greater than 15 mph were observed 23% of the time during week, and less than 2% the time during week 2.

By analyzing the wind speed and direction across each of the tests, we can classify each pad using a two-digit code. The first digit corresponds to the leak configuration on the pad under consideration (0 for no leak, and 1 for leak), and the second digit corresponds to interference potential (0 for no interference, and 1 for potential interference). Each pad can be classified under one of four options:

L00: No leak on current pad and no leak on any upwind pad
L01: No leak on current pad and leak on at least one upwind pad
L11: Leak on current pad and leak on at least one upwind pad
L10: Leak on current pad and no leak on any upwind pad

As an example, consider a scenario where there are leaks on pads 3 and 5, and no leaks on pads 1, 2, and 4 (Figure S1). When winds are blowing from the west, pads 4 and 5 are downwind of pad

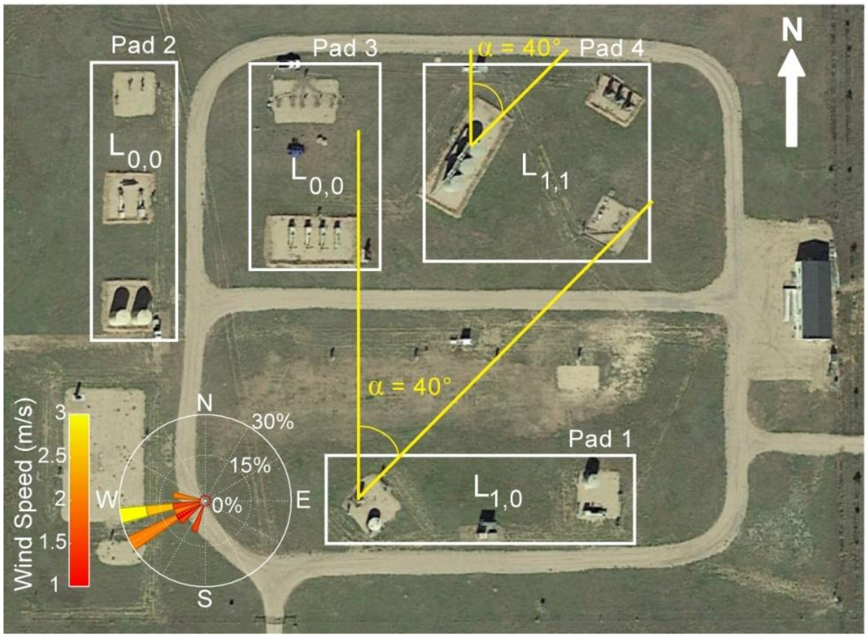
3 that has a leak. Therefore, pad 4 would be designated as L01, and pad 5 as L11. Pad 3, being upwind, will be designated L10. Pads 1 and 2 will be designated as L00 as they do not have leaks and there is no potential for interference from westerly winds.

In our analysis, interference was determined by a combination of vector-averaged wind speed and wind direction during each test duration (5, 10, or 20 minutes). For each leak, we considered a 40° cone centered on the leak along the average wind direction. In cases where the 40° cone did not include at least 50% of the wind vectors in the leak interval, we expanded the cone-angle until 50% of wind vectors is contained in that angle. If there was another leak or pad within this cone, we concluded that there was possible interference on the downwind pad.

Inter-pad interference not only depends on the wind-speed, wind-direction, and leak rate, but also detection limits of the technology and dispersion characteristics. However, whether a technology can detect 0.1 ppm and 1 ppm under specific weather and atmospheric conditions becomes a subjective analysis. To be fair to all technologies and avoid uncertainty around dispersion modeling, we analyzed interferences under three scenarios – no interference, weak interference, and strong interference. These are described below:

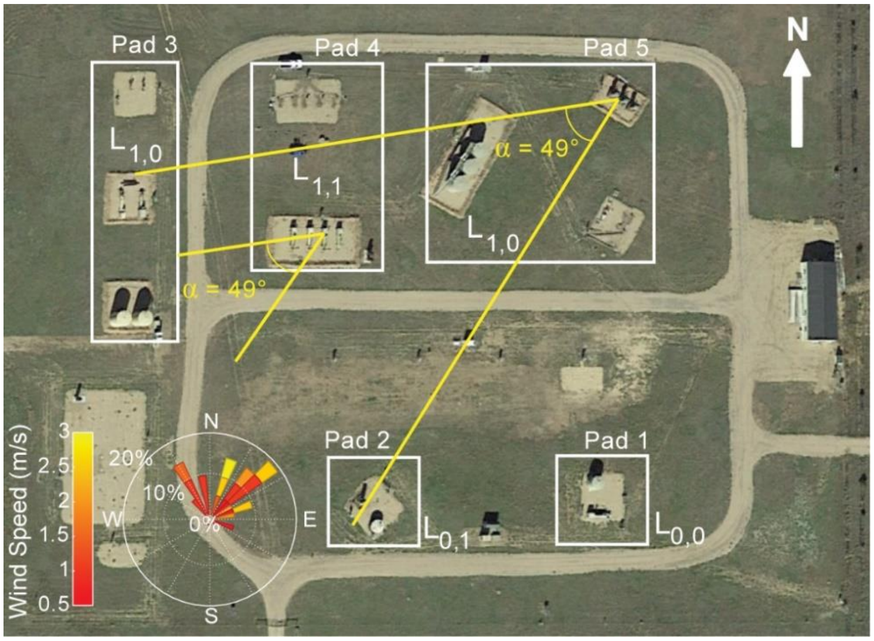
No interference scenario: Interference is negligible. All leaks are either assigned L00 or L10 – this scenario is identified as the 'base-case' scenario in the main text and reproduced here for comparison.

Weak interference scenario: Interference is considered only for those tests where the average wind speed was greater than or equal to 2 m/s. Figure S5 shows an example of a weak-interference scenario with leaks on pads 1 and 4, and a mean wind speed of 2 m/s along 202.6 degrees. The 40° cone contains 80% of all the wind vectors during the 10-minute test duration. In this scenario, data from pads with potential interference issues (pad 4) were discarded before teams' performance was analyzed.



SM_Figure 5. An example of weak-interference scenario during Week-1 of testing at METEC. The 10-minute binary yes/no detection test scenario had a 2 m/s average wind from 202.6° and leaks on pads 1 and 4. Because pad 4 could experience potential interference from pad 1, the results from pad 4 were discarded from statistical analysis. Here, the 40° cone contained 80% of all wind vectors in the 10-minute test period.

Strong interference scenario: Interference is considered for all tests, irrespective of average wind speed. This represents the most conservative analysis where all tests with any possibility of interference are removed from overall statistics. Figure S6 shows an example of a strong interference analysis with leaks on pads 3, 4, and 5. The wind speed averaged 1.4 m/s from 58.6 degrees. Because the 40° did not contain at least 50% of the wind vectors during the 20-minute test interval, we expanded it to 49°. In this scenario, all results from pads with potential interference (Pad 2 and Pad 4) were discarded prior to analyzing teams' performance.



SM_Figure 6. An example of strong-interference scenario during Week-2 of testing at METEC. The 20-minute detection and quantification test scenario had a 1.4 m/s average wind from 58.6° and leaks on pads 3, 4, and 5. Because pad 2 and pad 4 could experience potential interference from pad 5, the results from pad 2 and pad 4 were discarded from statistical analysis. Here, the 40° cone did not contain at least 50% of wind vectors in the 20-minute test interval and was therefore expanded to 49°.

No interference analyzes were performed for teams tested on week 3 in California because of the simplified test set-up (only 3 potential leak sources), large release rates, and that the teams were tested simultaneously on each leak removing the possibility of interfering sources.

6. Results from interference analysis

The main effect of the weak and strong interference analysis across all technologies is a reduction in sample size because of discarding potentially interfering test scenarios. In presenting results from the interference analysis, we consider performance of each team across four possible parameters – true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) arrange in a matrix. In addition, we also consider the fraction of TP leaks across level-1, level-2, and level-3 type detection.

In each of the analysis presented below, we did not find any statistically significant difference in the performance of the teams in the weak- and strong-interference scenarios, compared to the base-case scenario. While the performance of the teams can be affected by many factors including environmental conditions, this analysis shows that inter-pad interference from wind was not one of them. Any difference observed in the teams' performance is likely more impacted by the algorithms that process raw concentration data into useful information such as leak location, flux rate, and the ability to reject noise. The base-case scenario results, presented in the main manuscript, assume zero interference. Results from each of the teams are presented below.

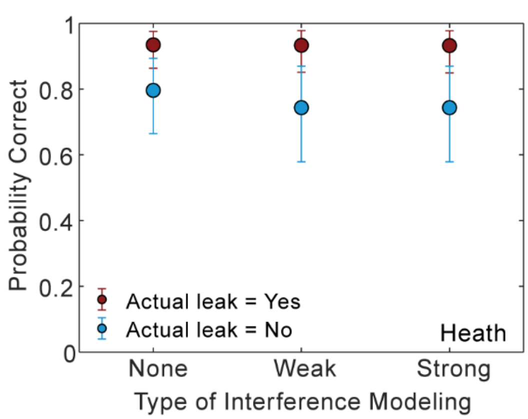
6.1. Heath technologies

No statistically significant change in Heath's performance was observed under mild or strong interference scenarios. This happened because test scenarios during high-wind days were staggered to avoid direct interference downwind of the plumes (e.g., leaks only on Pads 2 and 4 when winds blow from the North).

SM_Table 4. Interference results from Heath Consultants. Performance of Heath Consultants in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

Heath Consultants	Base-Case Scenario		Weak Interference Scenario		Strong Interference Scenario	
	Yes (measured)	No (measured)	Yes (measured)	No (measured)	Yes (measured)	No (measured)
Leak (actual)	0.93 (86)	0.07 (6)	0.93 (70)	0.07 (5)	0.93 (69)	0.07 (5)
No Leak (actual)	0.26 (11)	0.74 (43)	0.26 (10)	0.74 (29)	0.26 (10)	0.74 (29)

True Positives	Base-Case	Weak-Interference	Strong Interference
Level 1	0.41(35)	0.40 (28)	0.39 (27)
Level 2	0.47 (40)	0.47 (33)	0.48 (33)
Level 3	0.12 (11)	0.13 (9)	0.13 (9)



SM_Figure 7. Interference analysis of Heath Consultants' performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

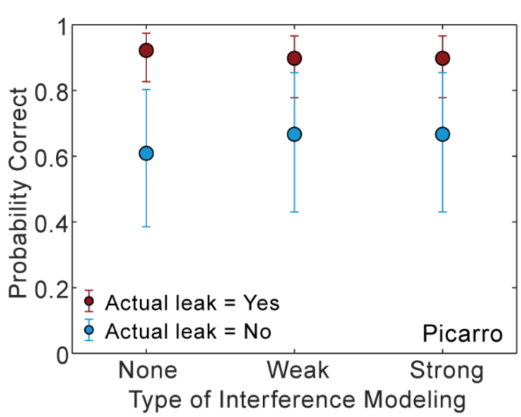
6.2. Picarro Inc.

We observed minor differences in the weak and strong interference scenarios for true negative and false negative rates for Picarro. For example, the fraction of true negative detections increased from about 61% in the base-case scenario to 67% in the weak and strong-interference scenarios, while the correspond false negative detections decreased. However, the error in these two cases overlapped and cannot be assumed to be a significant difference in performance.

SM_Table 5. Interference results from Picarro Inc. Performance of Picarro Inc. in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

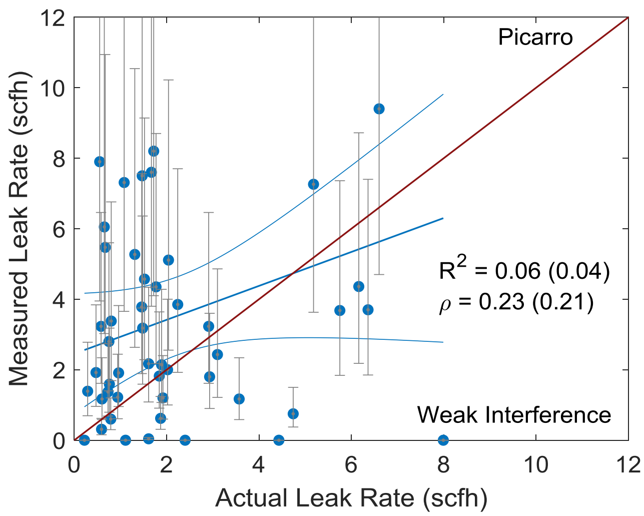
Picarro	Base-Case Scenario		Weak Interference Scenario		Strong Interference Scenario	
	Yes (measured)	No (measured)	Yes (measured)	No (measured)	Yes (measured)	No (measured)
Leak (actual)	0.92 (59)	0.08 (5)	0.90 (44)	0.10 (5)	0.90 (44)	0.10 (5)
No Leak (actual)	0.39 (9)	0.61 (14)	0.33 (7)	0.67 (14)	0.33 (7)	0.67 (14)

True Positives	Base-Case	Weak-Interference	Strong Interference
Level 1	(0)	0.0 (0)	0.0 (0)
Level 2	0.25 (15)	0.30 (13)	0.30 (13)
Level 3	0.75 (44)	0.70 (31)	0.70 (31)



SM_Figure 8. Interference analysis of Picarro Inc.'s performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

Results from quantification are non-trivially different only for Picarro and is therefore shown here. The weak- and strong-interference case shows marginal improvement in R^2 value – the base-case results are shown in parenthesis in Figure S9. The average error increased from -0.89 scfh (95% C.I. [-1.8, 0.01]) in the base-case scenario to -1.29 scfh (95% C.I. [-2.4, -0.19]), demonstrating a clear over-estimation bias.



SM_Figure 9: Quantification parity chart for Picarro under the weak- and strong-interference scenarios. While there was marginal improvement in R^2, the average error also increased from -0.9 scfh to -1.3 scfh. The parameter values in parenthesis represent base-case scenario.

6.3. Seek Ops Inc.

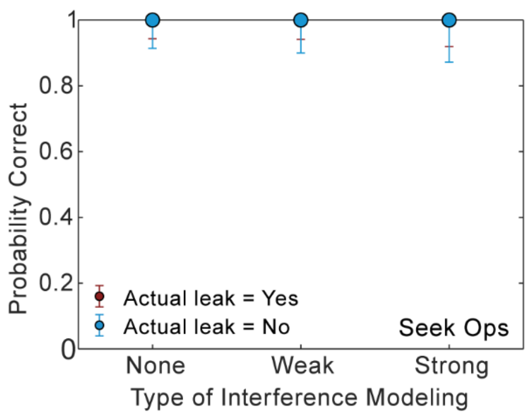
There are no statistically significant changes to Seek Ops' performance in the weak- and strong-interference scenario compared to the base-case.

SM_Table 6. Interference results from SeekOps. Performance of Seek Ops Inc. in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

Seek Ops	Base-Case Scenario	Weak Interference	Strong Interference

Inc.			Scenario		Scenario	
	Yes (measured)	No (measured)	Yes (measured)	No (measured)	Yes (measured)	No (measured)
Leak (actual)	1.0 (63)	0.0 (0)	1.0 (61)	0.0 (0)	1.0 (44)	0.0 (0)
No Leak (actual)	0.0 (0)	1.0 (41)	0.0 (0)	1.0 (35)	0.0 (0)	1.0 (27)

True Positives	Base-Case	Weak-Interference	Strong Interference
Level 1	0.68 (41)	0.64 (39)	0.73 (32)
Level 2	0.16 (10)	0.36 (22)	0.18 (8)
Level 3	0.16 (10)	0.0 (0)	0.09 (4)



SM_Figure 10. Interference analysis of SeekOps' performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

6.4. Aeris Technologies

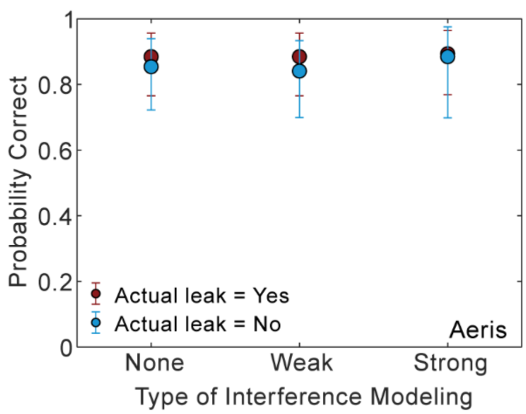
There are no statistically significant changes to Aeris's performance in the weak- and strong-interference scenario compared to the base-case.

SM_Table 7. Interference results from Aeris Technologies. Performance of Aeris Technologies in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

Aeris Technologies	Base-Case Scenario	Weak Interference Scenario	Strong Interference Scenario

	Yes (measured)	No (measured)	Yes (measured)	No (measured)	Yes (measured)	No (measured)
Leak (actual)	0.88 (46)	0.12 (6)	0.88 (46)	0.12 (6)	0.89 (42)	0.11 (5)
No Leak (actual)	0.15 (7)	0.85 (41)	0.16 (7)	0.84 (37)	0.12 (3)	0.88 (23)

True Positives	Base-Case	Weak-Interference	Strong Interference
Level 1	0.57 (26)	0.57 (26)	0.60 (25)
Level 2	0.17 (8)	0.17 (8)	0.17 (7)
Level 3	0.26 (12)	0.26 (12)	0.23 (10)



SM_Figure 11. Interference analysis of Aeris Technologies' performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

6.5. Advisian

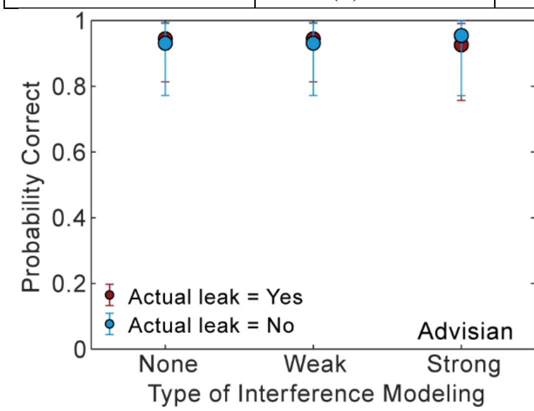
There are no statistically significant changes to Advisian's performance in the weak- and strong-interference scenario compared to the base-case.

SM_Table 8. Interference results from Advisian. Performance of Advisian in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

Advisian	Base-Case Scenario		Weak Interference Scenario		Strong Interference Scenario	
	Yes	No	Yes	No	Yes	No

	(measured)	(measured)	(measured)	(measured)	(measured)	(measured)
Leak (actual)	0.94 (34)	0.06 (2)	0.94 (34)	0.06 (2)	0.93 (25)	0.07 (2)
No Leak (actual)	0.07 (2)	0.93 (7)	0.07 (2)	0.93 (27)	0.05 (1)	0.95 (21)

True Positives	Base-Case	Weak-Interference	Strong Interference
Level 1	0.50 (17)	0.50 (17)	0.44 (11)
Level 2	0.26 (9)	0.26 (9)	0.32 (8)
Level 3	0.24 (8)	0.24 (8)	0.24 (6)



SM_Figure 12. Interference analysis of Advisian's performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

6.6. ABB/ULC Robotics

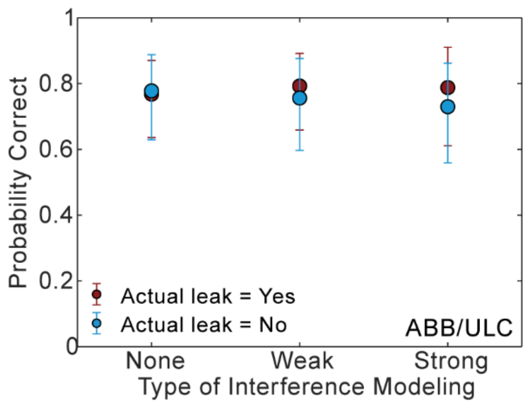
There are no statistically significant changes to ABB's performance in the weak- and strong-interference scenario compared to the base-case.

SM_Table 9. Interference results from ABB/ULC Robotics. Performance of ABB/ULC Robotics in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

ABB/ULC Robotics	Base-Case Scenario		Weak Interference Scenario		Strong Interference Scenario	
	Yes (measured)	No (measured)	Yes (measured)	No (measured)	Yes (measured)	No (measured)

Leak (actual)	0.77 (43)	0.23 (13)	0.79 (42)	0.21 (11)	0.79 (26)	0.21 (7)
No Leak (actual)	0.22 (10)	0.78 (35)	0.24 (10)	0.76 (31)	0.27 (10)	0.73 (27)

True Positives	Base-Case	Weak-Interference	Strong Interference
Level 1	0.0 (0)	0.0 (0)	0.0 (0)
Level 2	0.0 (0)	0.0 (0)	0.0 (0)
Level 3	1.0 (35)	1.0 (31)	1.0 (26)



SM_Figure 13. Interference analysis of ABB/ULC Robotics' performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

6.7. Baker Hughes – GE (BHGE)

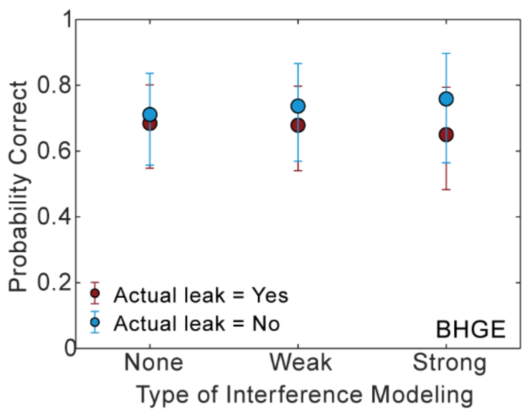
There are no statistically significant changes to Baker Hughes's performance in the weak-interference and strong-interference scenario compared to the base-case.

SM_Table 10. Interference results from Baker Hughes GE. Performance of Baker Hughes (GE) in the weak-interference and strong-interference scenarios, compared to the base-case scenario. The results are presented as percentages, with the sample size in parenthesis.

Baker Hughes GE	Base-Case Scenario		Weak Interference Scenario		Strong Interference Scenario	
	Yes (measured)	No (measured)	Yes (measured)	No (measured)	Yes (measured)	No (measured)
Leak (actual)	0.68 (39)	0.32 (18)	0.68 (38)	0.32 (18)	0.65 (26)	0.35 (14)

No Leak (actual)	0.71 (32)	0.29 (13)	0.74 (28)	0.26 (10)	0.76 (22)	0.24 (7)

True Positives	Base-Case	Weak-Interference	Strong Interference
Level 1	0.31 (12)	0.29 (11)	0.0 (0)
Level 2	0.18 (7)	0.18 (7)	0.0 (0)
Level 3	0.51 (20)	0.53 (20)	1.0 (26)



SM_Figure 14. Interference analysis of Baker Hughes' (GE) performance. The fraction of correctly identified tests for true positives (red), and true negatives (blue) across the three scenarios considered in this analysis. The error bars correspond to 95% confidence intervals associated with finite sample sizes.

References

[1] U.S. Advanced Research Projects Agency (ARPA-E), "Methane Observation Networks with Innovative Technology to Obtain Reductions," Washington D.C., 2014.

[2] T. Barchyn, C. Hugenholtz, C. Myshak and J. Bauer, "A UAV-based system for detecting natural gas leaks," *J. Unmanned Vehicle Systems,* vol. 6, pp. 18--30, 2017.