Introduction and motivation

There is intense interest in reducing methane emissions from hydrocarbon production infrastructure. Methane is an extremely potent greenhouse gas: vents and leaks have a large near-term climate impact (IPCC, 2013). For example, U.S. oil and gas supply chain methane emissions were estimated to be 2.3% of natural gas production, a loss rate at which the 20-year climate impact of supply chain methane emissions nearly equals the CO$_2$ climate impact from the total U.S. natural gas combustion (Alvarez et al., 2018). Some of these emissions can be easily mitigated at a relatively low cost (ICF International, 2015), and in cases the methane saved can be sold to offset mitigation costs. In addition to the climate impact, emissions can contain other hazardous species (Garcia-Gonzales et al., 2019).

Although there are many methods to decrease emissions that are technically simple (e.g., changing pneumatic valves to models that emit less); reducing emissions from fugitive sources presents a more serious challenge. Fugitive sources can occur randomly across a vast supply chain, in both remote and populated areas (Yacovitch et al., 2018; Zavala-Araiza et al., 2017). Fugitive sources can exist near other emitting equipment or can represent abnormal operation of normally emitting infrastructure where emissions are above design specification. Some fugitive plumes are mixed with plumes from allowable sources.

A major challenge in mitigating fugitive emissions is finding the sources – and finding them quickly. Searching every site and component for leaks with traditional techniques such as optical gas imaging or the U.S. Environmental Protection Agency Method 21 is labor intensive and expensive (ICF International, 2015). However, research has suggested that significant emissions come from a class of sources known as 'super-emitters', which are infrequent, high magnitude sources caused by abnormal process conditions (Brandt et al., 2016; Zavala-Araiza et al., 2017). These super-emitters, if not repaired quickly, can make up most of the sum of emissions from a field (Brandt et al., 2016). Thus, the optimum strategy for reducing emissions is not clear, ranging from: (i) focusing on inexpensive screening methods implemented frequently to find super-emitters soon after they begin (Fox et al., 2019a), or (ii) using more expensive, but more thorough methods to mitigate a wider range of source rates. Simulations suggest that the optimum solution is likely between these two extrema, but depends sensitively on the detection probabilities of methods and leak distribution (Kemp et al., 2016).

There are a variety of mobile screening methods in active development, targeting a diversity of scales, with...
an assortment of application models (e.g., Albertson et al., 2016; Atherton et al., 2017; Feitz et al., 2018; Ravikumar et al., 2019). Understanding the emissions reduction potential of each is essential to support regulatory approval, base industry adoption, and understand the role of each in a leak detection and repair (LDAR) program. There is a broad need to estimate the leaks that a screening technology will detect to model the relation between deployment cost and mitigation potential (Figure 1).

To effectively feed leak mitigation models and predict emissions reduction, detection must be understood. A common term used to describe detection efficacy is the ‘minimum detection limit’ (MDL). However, detection is better modeled as a series of probabilities (Kemp et al., 2016). Previous research has shown small differences in the detection probabilities can have large impacts on the emissions reduction potential of the method (Ravikumar and Brandt, 2017). Unfortunately, there remains little guidance for evaluating detection probabilities of screening technologies, and there is no well-understood approach for comparing different technologies.

Leak rate is commonly seen as a main control on detection probability (e.g., Kemp et al., 2016). However, detection probability is also affected by other factors affecting plume behaviour and the ability of methane-sensing technologies to distinguish those plumes. For example, sensors that must physically be placed in the plume require a prediction of the wind direction and plume characteristics that will always have error. To better explore these variables, here we evaluate the detection probabilities of a drone-based laser spectrometer system (described by Barchyn et al., 2018) (Figure 2) using controlled releases. This system consists of a long range, fixed wing drone, mounted with an integrated laser spectrometer to measure in-situ methane concentrations. We discuss the process of constructing a detection model with controlled release testing and examine factors that modulate detection probability. We also discuss practicalities of controlled release testing to better inform protocol development. This paper builds on recent work by Ravikumar et al., (2019).

**Drone and plume detection algorithms**

**Drone description**

We use a fixed wing long-range drone with integrated laser spectrometer (Boreal Laser GasFinder 2) that measures methane concentrations in the air the drone passes through. This system is optimized for long-range leak detection missions in remote terrain. The drone is flown downwind of infrastructure and the measured methane concentrations are analyzed for anomalies, which may represent plumes, and trigger follow-up in operational settings (see Barchyn et al., 2018 for full details).

**Detection approaches and experimental considerations**

Detection algorithms must explicitly balance false positive rate to detection effectiveness. Too many false positives will be costly as follow-up crews will be searching for non-existent leaks. Conversely, an algorithm that isn’t sensitive enough will have low detection probabilities. As such, there is an economic optimization exercise involving both survey and repair costs. We expect this optimization to be done operationally as both survey and repair costs are variable (ICF International, 2015). In this study we choose one set of parameters, but other sets may be chosen.

To clarify the terminology used here: we define a plume transect as a discrete traverse from one side of a predicted plume location to another, where the start and end of the plume transect are reliably in background methane concentrations. We discuss the process of constructing a detection model with controlled release testing and examine factors that modulate detection probability. We also discuss practicalities of controlled release testing to better inform protocol development. This paper builds on recent work by Ravikumar et al., (2019).

---

**Figure 1: Linking detection probabilities to emissions reductions.** Prediction of the mitigation potential of LDAR technologies requires simulating the interplay between at least two curves: (i) leak size distribution and (ii) the detection probability of a given LDAR technology. Typically, the curve of present leaks (red) is skewed such that there are a few large sources and many smaller leaks. High leak rate sources are important as they emit more methane per time interval. Hypothetical detection probability curves of two technologies are shown. LDAR Tech A (yellow) is a more sensitive, but more expensive technology that can detect a greater proportion of leaks. LDAR Tech B (blue) is a less expensive but less sensitive technology that could also be applied, likely more frequently, detecting only the upper tail of the curve. Both technologies could reduce the same volume of methane. DOI: [https://doi.org/10.1525/elementa.379.f1](https://doi.org/10.1525/elementa.379.f1)
conditions far from the influence of the target plume. We define a plume as a region of classified methane anomaly (further defined below). Detected plumes can then be classified as true positives (likely a real plume), or false positives (unlikely a real plume).

Close-range plumes, when sampled in sub-minute transects are often not Gaussian shaped or diffuse (Nathan et al., 2015; von Fischer et al., 2017; Yacovitch et al., 2018; Weller et al., 2018). They tend to have small, high concentration ‘tongues’ of methane. Typically, over time-averaging of minutes to hours, the spatial pattern of concentrations can more closely approximate conventional concentration models (e.g., Brantley et al., 2014). Most mobile screening techniques are designed to produce plume detections at sub-minute timescales (Yacovitch et al., 2018), and generally the economic promise of these techniques lies in an ability to survey quickly. A survey can pass through where a plume is expected to be (as predicted with a time-averaged model) but not find a methane anomaly as the instantaneous plume at the time of the transect could have been too low or high, or off to one side. As gusts, lulls, and wind shifts at this timescale are common, this effect is real and introduces a vital caveat to close range, fast timescale detection: detection is not strictly a signal processing or anomaly detection exercise, it also includes elements of natural close-range plume variability (Nathan et al., 2015). Thus, purely on these theoretical grounds, it seems unlikely that any mobile screening technologies can offer a detection probability of 1.0.

Although it is not fully understood, stability classes likely help denote persistence in the spatial positioning of plumes such that more stable atmospheres result in plumes that more frequently hug ground level, and unstable conditions could lead to plumes that loft more frequently. Atmospheric stability is frequently represented by Pasquill-Gifford stability classes, but these classes only partially represent the effect of different atmospheric conditions on plume positioning (Caulton et al., 2018).

Leak detection systems must be evaluated together (platform, sensor, algorithms, method, and operator) (Ravikumar et al., 2018). For example, two different platforms will have different detection probability models as the probability of intersecting a plume relates to the positioning of the sensor and vehicle. Many plumes do not loft high above the surface if released at ambient temperature (evidenced by reliable ground level detection: Brantley et al., 2014; Yacovitch et al., 2018). Further, the vertical pattern of typical concentration enhancements in plumes as evidenced by models such as the Gaussian plume model suggests that detection at higher elevations is less probable. As a second example, the speed of the vehicle and temporal speed of the sensor control the spatial resolution of measurements, controlling detection probability by affecting the spatial sample size. Insufficient sampling speed or excessively fast vehicle travel can miss plumes (Nathan et al., 2015). As such, laboratory tests can easily be misrepresented – field testing is required.

As a final experimental consideration, it is important to emphasize that this system (and other leak detection tools) are industrial tools designed to be applied at very large scales with explicit consideration of cost. For example, one may be able to predict the location of the plume with an array of anemometers and ground sensors to better guide the path of the drone. This is not economical in real application. Similarly, deploying repair or follow up crews to find leaks that don’t exist (false positives) can be very expensive. The scientific challenge in developing these types of systems is not optimizing detection probability – it is optimizing detection probability as a function of application cost.
**Plume detection algorithms**

Before discretizing plumes, we apply a series of data preprocessing steps to address (i) high frequency noise and (ii) low frequency drift in sensor response. To suppress high frequency noise in the sensor response, we apply a running mean across the concentration. The high frequency noise in laboratory conditions is approximately 0.05 ppmv (1 standard deviation) (Barchyn et al., 2018). We don’t make an *a priori* assumption about the source of the high frequency variability: we analyze running means ranging from 0 (no smoothing) to a diameter of 11 adjacent points. For context, data from the drone are produced at approximately 3 Hz – a running mean of 11 adjacent points corresponds to a time window of approximately 3.3 s, or approximately 66 m (the drone travels ~15–20 m/s). From these convoluted data, we create an anomaly series that shows the residual from a symmetrical running median filter of 15 s radius. This procedure is designed to model the ambient conditions and address low frequency sensor drift (Barchyn et al., 2018). A median is preferable to a running mean as a representation of ambient conditions because a running median is not affected by a plume anomaly in the window that is less than 15 s wide. A running median also has fewer step artifacts and sensitivities to non-representative extreme minima (likely caused by sensor noise). Results are not particularly sensitive to anomaly filter size as the low frequency drift in the sensor is relatively slow.

We define a plume as a segment of the plume transect with elevated concentration. A plume transect may have multiple plumes (please note we refer to both true positive and false positive anomalies as plumes). Thus, we require a method to discretize individual plumes, mapping their extent. To do this, we use a region-growing algorithm that begins with a global maxima, and ‘grows’ to fill in the full plume extent.

A plume discretization begins with an anomaly measurement exceeding a defined initial plume detection threshold. From this point, we test adjacent data points forward and backwards in time and continue defining the plume if anomalies remain positive. Plume definitions are extended 1 s past the last positive anomaly to span negative anomalies from sensor noise and extend the plume, similar to manual interpretation. The following pseudocode describes how the region growing algorithm expands from an anomaly maxima with index $i$.

If the concentration anomaly ($C_{\text{anomaly}}(i)$) is positive, the plume discretization continues by assigning the time associated with the present record to the record of last positive anomaly ($t_{\text{last_pos_anomaly}}(i)$) and the plume ID to the record of plume IDs ($P(i)$). If the concentration anomaly is negative, the plume continues expanding so long as the absolute value time difference is less than 1 s (1d, 1e). The same code runs to grow the plume discretization backwards in time from the anomaly maxima (1f is replaced with $i = i - 1$). Generally, we use lenient amalgamation parameters to avoid production of large numbers of false positive plume detections that are unrealistic.

Plumes are defined in order of descending anomaly, starting from the largest anomaly, until there are no anomalies present that exceed the initial detection threshold.

For every plume discretized, we calculate a series of basic extent and positioning metrics, including distance downwind, time span, flight elevation, spatial span, concentration and anomaly metrics, and metrics of the release such as leak rate and local wind speed. We also calculate the predicted concentration enhancement from a Gaussian plume model as a predictor of probability detection (see Supplemental methods S1). We capped the number of detected plumes to 10 per transect, beyond which the density of plumes is unrealistic for the use case of this drone.

**Controlled release experiments**

**Methodology**

Controlled release experiments were performed to evaluate the ability of the drone system to detect the presence of plumes. We set the release rate to a constant value for a series of transects at various distances, testing the ability of the drone to detect the presence of the plume. Typically, each plume transect took 75 s. We used compressed line gas with 95% CH$_4$ metered with a NuFlow Scanner 2000 flow meter to a 2.29 m high release stack. This is analogous to a leak in a pipeline riser or other ground infrastructure. The gas was released at below ambient temperatures due to depressurization and was released through a wide orifice with minimal velocity. Optical gas imaging of the plume as it exited the stack showed horizontal or minor (<0.5 m) sinking adjacent to the stack. All tests were performed near Brooks, Alberta, Canada (50.451°N, −112.120°W).

**Ancillary data**

To better understand the nature of the plume and ensure the drone transects were being conducted in an area where there would be a plume, we used a series of ancillary data sources. To quantify the initial dilution of the methane, we used an RM Young 81000 sonic anemometer to measure the wind ~10 m upwind from the source at 10 Hz (Figure 4a). We produced average wind speed values for each transect. During the flights, the drone operator used the onboard wind direction data to position the drone downwind of the source.

We also used a mobile ground lab (MGL) to locate the plume (e.g., Caulton et al., 2018). The operators of the MGL conducted transects through the plume and were...
in contact with the drone operator to ensure the drone flight path was intersecting the planimetric location of the plume. The plume was always detectable at surface level throughout all experiments with the MGL, thus giving confidence that drone transects covered the area where there was a real plume. In many cases the drone likely transected above the plume, but these situations are real and part of the results.

Figure 3: Plume detection examples. The raw anomaly is shown in blue. The plume detection algorithm located a plume where the red line equals 1. (a) Plume detection where the required initial anomaly was 0.8 ppmv, with no running mean. (b) Plume detection where the required initial anomaly was 0.3 ppmv and the raw data was subject to 7 point running mean to address high frequency noise. Both panels are from the same transect, demonstrating how different detection algorithm parameters can affect the results and present different interpretations of the same raw data. DOI: https://doi.org/10.1525/elementa.379.f3

Figure 4: Controlled release experiments. (a) Controlled release stack. We used an anemometer upwind of the release stack to provide ancillary data to support plume modeling. Plume detections were classified as true or false positives (b) by whether the detection occurred within a true positive wedge, defined by wind direction and ancillary information. Subsequent detections and those outside the true positive wedge were considered false positives. DOI: https://doi.org/10.1525/elementa.379.f4
Classification of detections

After completing the plume detection algorithm to discretize the plumes found in a transect, we classified whether the plumes were false positives or true positives. For each plume transect, whether the drone detected the plume (a true positive exists) or did not detect the plume (no true positives exist) was logged. To denote a true positive, we first evaluated whether there were any plumes within a window of 10–20° surrounding the true wind direction as measured from the release anemometer. We tightened or moved these bounds if we had a coeval MGL measurement of the plume location. The first detection within these bounds was denoted as a true positive. Every subsequent detection was denoted as a false positive, even if within the azimuth window as we assume there is only one contiguous real plume. Most transects had a real plume present, so we assume that the first detection was indeed the real plume. Although there is a possibility of branched plumes, we did not measure any discretely branched plumes with the MGL, supporting our assumption of plume contiguity. As plumes were defined in descending order of anomaly, the real plume in the azimuth window corresponded to the plume with highest maximum anomaly.

Test results

Controlled release coverage

The controlled releases covered a wide range of conditions (Figure 5), but did not cover all possible conditions due to limited testing resources (data available in Barchyn et al., 2019). Most wind speeds tested were between 1–5 m s⁻¹. We tested a range of leak rates that are high, generally within the ‘super-emitter’ range: 0.0, 5.35, 10.71, 16.06, 21.42, and 32.13 g s⁻¹. The experiments with no emissions were used in false positive calculations only. We did not test lower release rates as we knew from previous tests that the system would not detect the plume. The range of distances tested (Figure 5c) is wider than one may use in practice as this system would fly a pre-defined flight path at an optimized distance downwind from the infrastructure (Barchyn et al., 2018). The test transects were generally crosswind, with cumulative times near 1–2 minutes and distances less than 3500 m. All transects were conducted with approximately neutral stability (stability class D). The test site had no trees and flat terrain. Emissions were detected from nearby production sites – but MGL data suggest these emissions rates were far lower than our release.

Detection sensitivity and running mean size

To understand the most appropriate detection algorithm parameters we test a series of data running means and initial detection thresholds. We evaluated running means of 0, 3, 5, 7, 9, and 11 points and the following detection sensitivities: 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.2, 2.4, 2.6, 2.8, and 3.0 ppmv. The combination of these options gives 138 scenarios for evaluation. We compare the number of false positives per km against the detection probability for plume transects where the Gaussian plume model enhancement was greater than 0.0001 g m⁻³ (~0.142 ppmv) (Figure 6). This removes some transects that were poorly positioned and for the purposes of optimizing detection algorithm parameters. We include these transects in subsequent analyses.

From Figure 6, we select a scenario for subsequent analyses. Different scenarios may be more attractive depending on tolerance for false positives and desired sensitivity. We use a running mean of 5 points and detection threshold of 1.0 ppmv for subsequent analyses. This yields a mean single transect detection probability of 0.205 and false positive rate of 0.019 false positives per km. The drone travels approximately 1 km/minute and flies...
>1.5 hrs during a leak detection mission. This would produce ~1.72 false positives per mission. Choosing a scenario with a much lower false positive rate also results in a lower capacity to detect real plumes, and vice versa (Figure 6b). As noted previously, balancing sensitivity to false positive rates is beyond the scope of this article.

**Detection probabilities**

With the chosen scenario, we evaluate the detection probabilities of two populations: (i) all of the transects, considered together \( n = 198 \), and (ii) a subset where the distance downwind was from 750–2000 m \( n = 104 \), representing transects where the drone was well-positioned for a neutral atmosphere and typical minimum flight heights (~40–50 m). Generally, the second subset would be more representative of normal operations where the drone is positioned downwind to increase the probability of detecting the plume at ~40–50 m.

First, results show there is only a slight increase in detection probability with leak rate, and with the series of transects that had a leak rate of 21.42 g s\(^{-1}\) there was 0 positive detections (Figure 7a, b). There were only 10 transects at this rate (Figure 7c), but with no detections these results confirm that parameters other than leak rate must be considered to model detection probability reliably. These data points were all collected on the same day, we are not certain why the system did not detect the plume on this day. Similarly, there is no clear monotonic or sigmoid increase in skill with leak rate, even when the data are subset to locations positioned 750–2000 m downwind from source. This contrasts with detection probability modeling of close-range techniques such as optical gas imaging (Ravikumar et al., 2018) where one can fit data to a type curve and assume near perfect detection at high leak rates.

Second, there is a peak in detection probability with wind speeds ~2.5 m s\(^{-1}\). With these wind speeds flow is defined, and the plume may more closely match a theoretical description of plume diffusion. Stronger wind speeds may show lower detection probabilities as the plume may be less likely to mix upwards in daytime conditions. Third, distance downwind from source also relates to detection probability (Figure 7g, h). This supports the need for precise positioning in surveys as detection probability is low close to source and is not optimum further downwind. Finally, the predicted concentration enhancement provides a noisy, but possibly useful predictor of detection probability. Concentration enhancement from the Gaussian plume model encapsulates many of the previous variables: leak rate, wind speed, diffusion, flight elevation, and distance downwind. There is little detection probability at low concentration enhancements, but a clear increase in probability with higher enhancements. The predicted concentration enhancement is not a measurement or useful prediction of the real concentration enhancement, it is a proxy predictor for the positioning and extent of the plume.

**Discussion**

**Detection probability**

These results demonstrate that (i) detection probability is a function of many variables, and (ii) detection probabilities of this system are on average 0.205 for large sources. Certain situations are better than others: e.g., positioning the drone at least 750 m downwind improves detection. In the conditions tested, results suggest the plume did not reliably diffuse upwards to flight elevation (40–50 m) until at least 750 m downwind. This distance could be less in unstable conditions. However, a requirement
Figure 7: Detection probabilities. Left column shows all the transects, right column shows transects subset to distances downwind from 750–2000 m. The detection probability as a function of leak rate (a, b) shows general increase in skill with leak rate but the relation is not strong. The number of transects in each leak rate bin (c, d), is shown to qualify (a, b). Detection probability as a function of wind speed shows a peak in skill near 2–3 m s⁻¹ (e, f). Detection probability as a function of distance shows a peak >800 m (g, h). Detection probability as a function of Gaussian plume model enhancement shows a more complex relation (i, j). Detection probability as a function of drone elevation shows little relation (above ground level, AGL) (k, l). Plots e–j show points representing an equal number of transects, where divisible evenly. Plots k, l are split evenly across the range of values. DOI: https://doi.org/10.1525/elementa.379.f7
to survey so far downwind means that localization skill using wind direction suffers as the impact of small wind direction errors is magnified at further distances downwind. The practical implications of this are dependent on the situation.

The need for ~2.5 m s$^{-1}$ of wind is expectable. Low wind speeds do result in less initial dilution and greater concentrations in the atmosphere, but at these low wind speeds the plume is less well formed (observed from the MGL), and we hypothesize that it diffuses upwards less reliably. Lower detection probabilities at strong wind speeds do fit classical understanding of plume dynamics.

Considered as a mean, detection probabilities are ~0.205, even with some releases reaching ~6000 scf (greater than production at many wells, Omara et al., 2018). Detection at these high leak rates is not guaranteed. Detection probabilities appear to level off with increase in predicted concentration enhancement from the Gaussian plume model. This supports a hypothesis that a combination of instrument noise and flight elevation is a limiting factor for this system. Such limitations would also apply to similar systems.

The detection probabilities can be increased if more false positives are acceptable (Figure 6). However, operationally, a false positive is likely costlier than a true positive as follow-up crews may search longer to find a nonexistent leak. The reputational risk of reporting false positives suggests that operators of mobile screening methods may strive to avoid false positives, implicitly or explicitly biasing sensitivity down. The parameters chosen here, which result in >1 false positive per mission, may be too sensitive operationally, suggesting that in practice, results shown here for detection probabilities are likely practical maxima.

**Outlook for improving detection probabilities**

We are unable to separate missed detections that were missed because the sensor on this system is not sensitive enough, from missed detections that were due to the plume being incorrectly positioned during the transect. This noted, if the sensor on this system was improved, it could fly further downwind where the plume more reliably mixes upwards. Detections at distances further downwind are more difficult to attribute as attribution skill becomes sensitive to wind measurements.

A major limitation of this system is an inability to fly very low. Risk of collision with trees or power lines is a real risk when flying less than 40 m above ground level. Drone operators are understandably collision adverse, even though many collisions will only harm the drone. At present, drone collision avoidance systems are not available to reliably avoid collisions at fixed wing flight speeds in unmapped terrain. We expect future iterations of the drone used here will be able to fly lower at reduced risk. A possible solution is the use of hybrid drones that fly primarily as a fixed wing but also have capabilities to hover and autonomously navigate close to the source, flying at speeds where optical collision avoidance systems are more reliable.

Less stable atmospheres may result in plumes rising above the surface and increase detection probabilities for this system. However, better mixed plumes may be more dilute, reducing detections. Due to limited testing resources, we were unable to test the system in all conditions. This noted, relatively stable atmospheres are quite common in winter and spring/fall months in Canada and other high latitude regions – locations with considerable quantities of oil and gas infrastructure.

Additional data could be collected on the ground to better constrain wind flow – but these data come at a considerable cost, likely rendering the drone uneconomic for operations. This noted, the system does offer some probability of detection at potentially an extremely low cost, particularly compared to manual methods and in places where other screening technology are less favourable. The economics of this system for leak detection is tightly related to the leak distribution of a given target field and the spatial configuration of surveyed assets. Consequently, we cannot say whether these results are favourable or not as the metrics of favourability must incorporate deployment cost.

**Generalizing detection probabilities**

The lack of a predictable relationship between leak rate and detection probability requires emphasis. These results suggest that other data must also be considered, and it may not be possible to reliably use only leak rate to discriminate detection probability in mobile screening technologies (Kemp et al., 2016; Ravikumar et al., 2018) (Figure 1).

This raises a question surrounding the use of detection probabilities and the models that use them (e.g., Kemp et al., 2016). For example, if using predicted concentration enhancement from the Gaussian plume model to inform detection probability, one must model the position of the drone, the position of the infrastructure, the weather, and then return a probability of detection. If the models use leak rate as a sole discriminator of detection probability – one could easily misrepresent practical skill. For example, if a predicted detection probability from a leak of >30 g s$^{-1}$ is taken from Figure 7b as ~0.36, but wind speeds are >5 m s$^{-1}$, the real detection probability is much lower. Two strategies for resolving this could be defining an operational envelope and using the envelope to limit application or modeling the detection probability across a broader range of conditions using a more granular simulation environment, perhaps considering the spatial positioning of assets. Further, the applicability of detection probabilities to evaluate screening skill may also be limited in steps beyond detection. If there are many closely spaced sources, it is not clear that this system would be able to disambiguate the sources, as plumes would mix.

Presently, testing coverage is not adequate to generalize across all situations (Figure 5). We have no tests in treed or hilly terrain, nor do we have evaluations of performance in more unstable or stable atmospheres. Although a predictor such as concentration enhancement from the Gaussian plume model can be evaluated for any situation, we lack data to suggest results can be generalized outside of our test conditions. Although predicted enhancement is an imperfect proxy – the metric does incorporate sufficient data to make a reasonable first-order guess on the location of a plume from a site.
Testing protocol
The testing protocols used here are labour and cost intensive but have some important advantages. First, the use of controlled releases gives an absolute indicator of performance. Although it is easier to use an existing source (e.g., compressor station exhaust, etc., e.g., Nathan et al., 2015), comparing the response of multiple systems directly with an uncontrolled source to elucidate detection probabilities is unlikely to yield generalizable data. This is further emphasized by the complexity of the detection probabilities seen here.

Testing the system in a manner that emulates operations is vital. For example, we hypothesize that a major limitation of this system is its inability to fly low within the plume. If testing was performed in a synthetic manner on the ground, the results would be less reliable to generalize.

Regulatory and policy implications
In jurisdictions where LDAR is regulated, these results have policy implications. In some jurisdictions such as Canada, screening methods are being considered for ‘equivalency’ with more conventional component level surveys (Government of Canada, 2018; Fox et al., 2019b). Elsewhere, the broad motivation for implementation of these technologies is driven by a desire to optimize the ratio of cost to emissions reductions. The results from this study demonstrate detection probabilities of mobile systems are likely not simple nor predictable, possibly warranting a certification approach. With certification, including both a test suite coverage and probability models for that coverage, a technology can be used in an optimization model with other technologies, and a prediction of emissions reductions can be made and adjusted to meet the target standard.

As an example, the drone tested here could be certified for use in conditions equivalent to the test suite here, and detection probabilities could be taken from Figure 7. The drone could be integrated into a LDAR program where it is applied relatively frequently to search for super-emitters. Simple quality control protocols such as requiring multiple detections for a given source to trigger follow-up could statistically improve the efficacy of the system and facilitate operating the system with more sensitive detection settings. Expanding these quality controls to require detections on different days would help increase certainty. The total emissions reductions can be predicted (Kemp et al., 2016), and presumably the program would be implemented with a lower operator cost.

Although regulatory certification following detection probability testing seems attractive, there are risks that outcomes will not match predictions. Mobile screening technologies, like the drone evaluated here, are quite sensitive to seemingly innocuous hardware configuration and detection algorithm changes. The variability and specificity of our results suggest that certification of technology must be specific, and auditing should be used to ensure compliance.

Data Accessibility Statement
All underlying data are available at: https://doi.org/10.7910/DVN/BR37R3 (Barchyn et al., 2019).

Supplemental file
The supplemental file for this article can be found as follows:

• Supplemental Methods S1. Description of Gaussian plume model concentration enhancement calculations. DOI: https://doi.org/10.1525/elementa.379.s1

Acknowledgements
We thank Kirk Osadetz, Clay Wearmouth, and Marshall Staples for field assistance. We thank Stephen Myshak and Owen Brown for drone operations. We thank the editors and anonymous reviewers for comprehensive and helpful reviews that improved the manuscript considerably.

Funding information
Funding for this project was provided by Natural Resources Canada’s Energy Innovation Program and the Petroleum Technology Alliance of Canada (TEB, CHH, TAF). We acknowledge in-kind support from Carbon Management Canada Research Institutes (TEB, CHH, TAF).

Competing interests
The authors have no competing interests to declare.

Author contributions
• Contributed to conception and design: TEB, CHH, TAF
• Contributed to acquisition of data: TEB, CHH, TAF
• Contributed to analysis and interpretation of data: TEB, CHH, TAF
• Drafted and/or revised the article: TEB, CHH, TAF
• Approved the submitted version for publication: TEB, CHH, TAF

References

Supplemental Methods S1.
Description of Gaussian plume model concentration enhancement calculations. DOI: https://doi.org/10.1525/elementa.379.s1