

RESEARCH ARTICLE

Characterizing detection probabilities of advanced mobile leak surveys: Implications for sampling effort and leak size estimation in natural gas distribution systems

Emily Luetschwager¹, Joseph C. von Fischer², and Zachary D. Weller^{1*}

Advanced leak detection (ALD) to survey local natural gas distribution systems has reached a point in technological maturity where new federal regulations will require its use in compliance surveys. Because most of these deployments are conducted by commercial providers, there has been little publicly available data documenting characteristics of the underlying methane (CH₄) plumes that are the core features measured in ALD surveys. Here, we document key features of CH₄ plumes measured in ALD surveys of 15 U.S. metropolitan areas where we had deployed high-sensitivity CH₄ analyzers on Google Street View cars. Our analysis reveals that CH₄ concentration enhancements from CH₄ sources exhibit high temporal variability, often differing by more than 10-fold among repeated observations. This variability introduces challenges for estimating source emission rates because the same source can appear to be large on one drive-by and small on the next. Additionally, the frequency distribution of CH₄ enhancements from a given source generally has a strong positive skew that can lead to overestimation of leak size. The magnitude of CH₄ enhancements from a source measured with a mobile sensor can also change quickly over time, as indicated by decreasing temporal correlation between mobile measurements longer than an approximately hourly time scale. To manage the uncertainty, we demonstrate how additional survey effort can help overcome this variability and instability to allow discrimination among the wide range of leak sizes. We quantify the probability of source detection, finding that it increases with estimated leak size. Combining these results, we develop a simulation that demonstrates the potential for ALD to detect leaks and quantify emissions as a function of sampling (driving) effort. Our results suggest that five to eight drives of each roadway in a target area would detect >90% of leaks and provide adequate emissions quantification for repair/replacement prioritization.

Keywords: Methane, Advanced leak detection, Natural gas, Mobile survey, CH₄, Greenhouse gas

1. Introduction

In recent years, highly sensitive methane (CH₄) analyzers have been used as a means to detect and monitor CH₄ emissions from the natural gas (NG) supply chain. Within the local distribution segment of the supply chain, these sensors have been deployed on mobile platforms (e.g., on cars, drones) to detect, locate, and quantify NG leaks. These mobile surveys are commonly coupled with data analysis algorithms to produce numerous data products such as maps of leak indications and estimated leak sizes. Together, these surveys, algorithms, and data products are commonly referred to as advanced mobile leak detection (ALD). Numerous studies have examined the results of ALD to understand the health

of local distribution systems (Phillips et al., 2013; Jackson et al., 2014; Gallagher et al., 2015; Chamberlain et al., 2016; Von Fischer et al., 2017; Sanchez et al., 2018).

ALD has the ability to be an integral part of local distribution system management and is being adopted as a best practice for pipeline management (California SB1371, 2016; Cuellar, 2020; Weller et al., 2018b). As the algorithms for processing data from these mobile surveys improve, so will their ability to provide crucial insights for maintaining pipeline safety, prioritizing leak repairs and mitigating CH₄ emissions. Several vendors have developed and are marketing proprietary ALD packages of sensors, algorithms, and software for commercial use (LGR-ABB, 2020; Picarro, 2020).

Understanding the benefits and limitations of ALD surveys is critical for informing survey practice and policy. For example, quantifying the likelihood of detecting elevated CH₄ concentrations at a leak location can inform the level of survey effort required to declare an area as having a low probability of containing a leak as part of a compliance

¹Department of Statistics, Colorado State University, Fort Collins, CO, USA

²Department of Biology, Colorado State University, Fort Collins, CO, USA

* Corresponding author:
Email: zach.weller@colostate.edu

survey. Similarly, understanding temporal variation and persistence in CH₄ concentrations can inform the timing and repeatability of sampling efforts. Quantifying ALD survey properties improves understanding of the cost-benefits and limitations of ALD surveys, which enables comparisons with other technologies (Fox et al., 2019).

Several studies have made progress in documenting the features of ALD data and data products. Many of the aforementioned studies document isotopic composition or ethane to CH₄ ratios of leak indications derived from mobile surveys (Phillips et al., 2013; Jackson et al., 2014). Weller et al. (2018a) estimated detection probabilities in two U.S. metropolitan (metro) areas for a capture-recapture model. Weller et al. (2018b) examined the correspondence between leak indications and leaks and bias in leak size estimates. Zhou et al. (2019) investigate the effects of environmental conditions on CH₄ enhancement levels in plumes and the resulting emissions estimates from mobile sensors for controlled releases. None of these studies examine variation in CH₄ concentrations from point sources over longer time scales (e.g., days), examine the effects of sampling effort on detection and variability in size estimates for the purpose of informing survey policy, or take advantage of ALD survey results across multiple cities.

We present an analysis of several features of ALD surveys. These features include the temporal persistence of CH₄ at a leak location, variation in estimated emission rates, and the probability of leak detection. Combining these results, we develop a simulation demonstrating the efficacy of ALD for a hypothetical NG distribution system. Our analysis uses data from ALD surveys conducted in 15 U.S. metro areas, which have not been previously analyzed together. Our analysis of these features provides insights into the use of ALD for detecting and quantifying CH₄ emissions which are important for understanding the abilities and limitations of ALD and informing survey practices and policies. We also discuss how the results of our analysis illuminate areas of improvement for ALD survey practices and data processing.

2. Materials and methods

We conducted surveys in 15 metro areas with highly sensitive CH₄ analyzers placed on Google Street View vehicles. The instruments, survey protocols, and data processing steps for these surveys have been described in detail elsewhere (Von Fischer et al., 2017; Weller et al., 2018b; Weller et al., 2019). We provide an abbreviated description of the methods and additional details as relevant for this study. The algorithm used for data processing is version 2.0 as described in Weller et al. (2019).

We identified regions within each of the 15 metro areas to target for our mobile surveys. These regions were typically identified using census information in order to obtain a representative sample of the metro area. For example, we surveyed areas that differed in socioeconomic characteristics (e.g., percentage minority, income, median housing age) and building type (e.g., commercial vs. residential). In some cases, the regions were identified in collaboration with the local utility company to target areas known to have leak prone pipes.

We outfitted Google Street View vehicles with highly sensitive CH₄ analyzers (Picarro or Los Gatos Research) and placed the air intake near the front bumper. These instruments recorded CH₄ concentrations down to the part-per-billion (ppb) level at a frequency of 2 Hz as the survey regions were driven. We instructed drivers to drive every roadway within target regions at least two times. We screened the raw data for various quality control concerns. The instrument reports the mole fraction of CH₄ in dry air, and we use the term “concentration” to refer to this mole fraction.

From our survey data, we estimated the background CH₄ concentration and used this to identify areas of elevated CH₄ concentrations. Elevated methane concentrations are defined as those that are at least 110% above the baseline value. A typical value for this threshold is 2.2 ppm, but it varies slightly over space and time to account for ambient changes over these domains. We describe these elevated CH₄ measurements as observed peaks, due to the rise and fall of the CH₄ concentration over the background concentration. We describe the difference between the CH₄ reading and the background concentration as the CH₄ enhancement. Observed peaks are indicators of the presence of a CH₄ emitting source, such as an NG leak or biogenic production site. Observed peaks are defined by a time-location-CH₄ combination, denoting the time stamp that the peak was observed, the location of the peak in latitude and longitude coordinates, and maximum CH₄ (or maximum CH₄ enhancement over background) in the peak.

After identifying locations of elevated CH₄ concentrations (observed peaks), we consolidated observed peaks that are in close spatial proximity (within 30 m) into a single verified peak. That is, we have verified that elevated CH₄ concentrations are present in an area when two or more observed peaks are present. We use this verification protocol in an attempt to screen elevated readings that arise due to nonstationary CH₄ sources (e.g., compressed NG buses). In our public reporting of survey results (Environmental Defense Fund, 2020), we only report verified peaks within the local distribution companies' service territory. At many locations, elevated CH₄ levels are detected only once and never detected on other drives of the area. These single observations of elevated readings are not reported as leak indications. Additionally, some verified peaks are located outside of the collaborating local distribution company's distribution system and are therefore not reported publicly. For the purposes of studying the features of CH₄ concentration properties, we have included those verified peaks in our reporting here, which may lead to different counts and leaks/mile than those given in other publications of our survey results.

We use CH₄ concentrations from the survey to estimate CH₄ emission rates from observed and verified peaks. Our algorithm for estimating these emission rates was calibrated using controlled CH₄ releases (Weller et al., 2019). In these controlled releases, we drove the survey vehicle through the CH₄ plumes created by our controlled CH₄ releases of various magnitudes in order to quantify the relationship between leak size and CH₄ plume characteristics. An analysis of the resulting data revealed an association between the maximum CH₄ enhancement,

defined as the maximum CH₄ observed in the plume minus the CH₄ background, and the known emission rate. We used linear regression to model the relationship between leak size and the maximum CH₄ enhancement on the natural log scale. We use this empirical relationship to estimate the unknown emission rate of leaks detected during surveying. Occasionally, this method produces estimates of emission rate that were beyond the range of emission rates used to calibrate the algorithm. For this reason, we cap all estimated emission rates at 100 L/min.

We calculated the total number of sampling attempts (drive-bys) and detections for each location where elevated CH₄ concentrations were observed on at least one occasion (i.e., for each verified peak and for each observed peak that was not a part of a verified peak). For each of these locations, we used spatial analysis (Environmental Systems Research Institute, 2015; Van Rossum and others, 2015) to determine how many times the survey vehicle drove past the location, which we call a sampling attempt. We define a sampling attempt as any time when the survey vehicle passed within 30 m of the location. Although rare, if the survey vehicle drove past a location more than once during a 30-s time window, those drive-bys were counted as a single sampling attempt. Finally, we computed the number of these sampling attempts that resulted in a detection of elevated CH₄ concentrations. For example, one may drive by a location five times and observe (i.e., detect) elevated CH₄ concentrations three times. In this example, there would be five sampling attempts, and three out of five resulted in a detection.

We were not able to discriminate between sampling attempts that were upwind or downwind of leaks because the actual leak expression locations are not known. Additionally, the anemometer used on our survey vehicles was prone to intermittent failures, and we have not been able to derive reliable wind direction or magnitude from this data stream. Thus, our results that rely on CH₄ signatures alone represent a lower bound on detection probability. We anticipate that commercial algorithms that use wind reliably are able to quantify the spatial coverage of surveys and, in combination with infrastructure or known leak locations, improve upon the detection rates reported from our surveys.

3. Statistical analyses

We analyzed the observed peak, verified peak, and detection data from 15 metro areas where we conducted surveying. Our analyses examined patterns in temporal correlation, variation in leak emission rate estimation, and the probability of detecting leaks with survey vehicles. All analyses were conducted using R software (R Core Team, 2018). **Table 1** provides a description of the number of observed and verified peaks in each city as well as leaks per mile of roadway surveyed. **Table 1** also shows the sampling effort in drive-bys and empirical detection probabilities. The mean and median number of drives per verified peak reflect both the total sampling effort in the city and the typical number of drive-bys of a leak under our targeted sampling protocol of driving each roadway two or more times. The mean is larger than the median because some areas are driven a large number of times (e.g., when

they are located on a major highway or close to the driver's residence).

3.1. Temporal correlation

To understand the repeatability of elevated CH₄ concentrations, we focused our analysis on temporal correlation over day, hour, and minute time scales. To quantify how variable or stable atmospheric CH₄ enhancements from a CH₄ source were on differing time scales, we calculated the correlation between pairs of CH₄ concentration measurements from the same putative source. For each verified peak, we extracted all observed peaks that were joined to create the verified peak. Each of these observed peaks consisted of a specific date and time. For each pair of observed peaks, we computed the difference in time and compiled the pairs of observed maximum excess CH₄ concentrations. Next, we created bins using the time differences (e.g., bin together all pairs from the same day, one day apart, etc.). Finally, we computed the correlation between the pairs of log₁₀ maximum excess CH₄ readings for each time bin and plotted the correlations on a line graph. Hereafter, log represents log₁₀ unless specified otherwise. We repeated this same process over three different time scales: day, hour, and minute. These correlations quantify the typical within-source correlation of log atmospheric CH₄ concentration enhancements measured by a mobile sensor.

3.2. Percentage difference in leak size estimation simulation

Our next analysis examined the effect of the number of detections on the variability in estimated leak emission rates. For this analysis, we subset the data to only include verified peaks with 20 or more observed peaks (i.e., where elevated CH₄ had been observed 20 or more times). For each of these verified peaks, we calculated the estimated emission rate using the average natural log of the maximum excess CH₄ of all the observed peaks within that verified peak. We refer to this estimated emission rate derived from all observations of the leak as the leak indication's reference emission rate for this simulation. We then performed a Monte Carlo simulation to assess the variation in estimated emission rates under a different number of detections relative to this reference. For a given number of detections and each verified peak, we randomly sampled that number of detections (observed peaks) from the entire set of detections of the verified peak. For example, we randomly select three observed peaks from the set of 45 observed peaks that compose the verified peak. We used the randomly selected observed peaks to compute a new, simulated leak emission rate. We repeated this random sampling 200 times for each verified peak and number of detections combination. We compiled these results for each of 2–10 detections. Next, we calculated the percentage difference between the reference emission rate and the simulated emission rates for each verified peak, $([\text{simulated} - \text{reference}] / \text{reference}) \times 100$. We computed an average percentage difference between the simulated emission rate and the reference emission rate over all simulations for each verified peak and number of

Table 1. A summary of several advanced leak detection survey results in 15 U.S. metro areas. DOI: <https://doi.org/10.1525/elementa.2020.00143.t1>

Metropolitan Area	Observed Peaks (OPs) ^a	Verified Peaks (VPs) ^b	Proportion of Nonverified OPs (Count)	Mean Drive-Bys per VP ^c	Median Drive-Bys per VP ^d	Overall VP Detect Prob ^e	VP per Mile Surveyed ^f
Birmingham, AL	498	185	0.63 (313)	9.4	7	0.49	0.54
Boston, MA	5,584	2,348	0.58 (3236)	8.8	6	0.64	0.90
Burlington, MA	88	14	0.84 (74)	12.1	7	0.43	0.12
Chicago, IL	1,374	466	0.66 (908)	7.9	7	0.48	0.45
City A ^g	5,592	1,920	0.66 (3672)	13.2	9	0.40	0.26
Dallas, TX	1,422	537	0.62 (885)	11.4	10	0.42	0.76
Elizabethtown, NJ	1,185	350	0.70 (835)	7.6	5	0.60	1.65
Indianapolis, IN	51	3	0.94 (48)	4	4	0.52	<0.01
Jacksonville, FL	459	94	0.80 (365)	11.0	9.5	0.35	0.11
Los Angeles, CA	1,310	389	0.70 (921)	11.5	8	0.42	0.23
Mesa, AZ	46	7	0.85 (39)	7.1	8	0.39	0.04
Clifton, NJ	1,229	430	0.65 (799)	9.2	8	0.44	0.53
Pittsburgh, PA	611	272	0.55 (339)	12.5	9.5	0.43	0.85
Staten Island, NY	2,155	980	0.55 (1,175)	9.3	8	0.49	0.99
Syracuse, NY	704	220	0.69 (484)	11.4	8	0.45	0.45

^aThe number of CH₄ peaks seen during the survey.

^bThe number of locations where CH₄ peaks were seen at least twice.

^cThe mean number of drive-bys among VPs.

^dThe median number of drive-bys among VPs.

^eThe proportion of drive-bys of VP locations that resulted in detection of elevated CH₄ levels.

^fThe number of VPs per mile of roadway surveyed.

^gCity name is anonymized because data are not released publicly due to ongoing rate case.

detections. Then, we computed an overall average percentage difference across all included verified peaks for each number of detections. The average percentage difference for each verified peak and the average across all verified peaks was plotted to show the variation in estimated emission rates (relative to the reference) as a function of the number of detections. We note that this simulation does not evaluate our estimated emission rates relative to validated measurements as in Weller et al. (2018b). However, this simulation does demonstrate the variability in estimated leak sizes relative to a large number of detections (i.e., the reference emission rate) and how that variability changes with more detections.

3.3. Probability of detection

Next, we investigated the relationship between detection probability and estimated leak size. Detection probability is the likelihood that elevated CH₄ levels will be observed on a single drive-by of a leak. For this analysis, we used only the verified peaks from each city. We did not include data from locations where elevated CH₄ is observed on only one occasion because these can arise due to

nonstationary CH₄ sources (e.g., NG buses). We conducted a sensitivity analysis of detection probabilities, including locations where elevated CH₄ was observed on only one occasion (see SI S3). For each verified peak, we aggregated the number of sampling attempts, total number of detections, and the estimated emission rate. We again capped estimated leak size at 100 L/min. We then fit a mixed-effects logistic regression model, modeling detection probabilities as a function of the estimated log leak size. For this regression model, we assumed a common slope parameter across metro areas but allowed for an area-specific random intercept. We chose to drop Indianapolis, Mesa, and Burlington from the logistic regression analyses because these metro areas each had very few (<15) verified peaks leading to unstable parameter estimates.

3.4. Probability of verification

We utilized the detection probabilities from the fitted logistic regression model to compute verification probabilities as a function of leak size and the number of drive-bys. A leak is verified if it is detected on two or more drive-bys. Intuitively, a leak is more likely to be

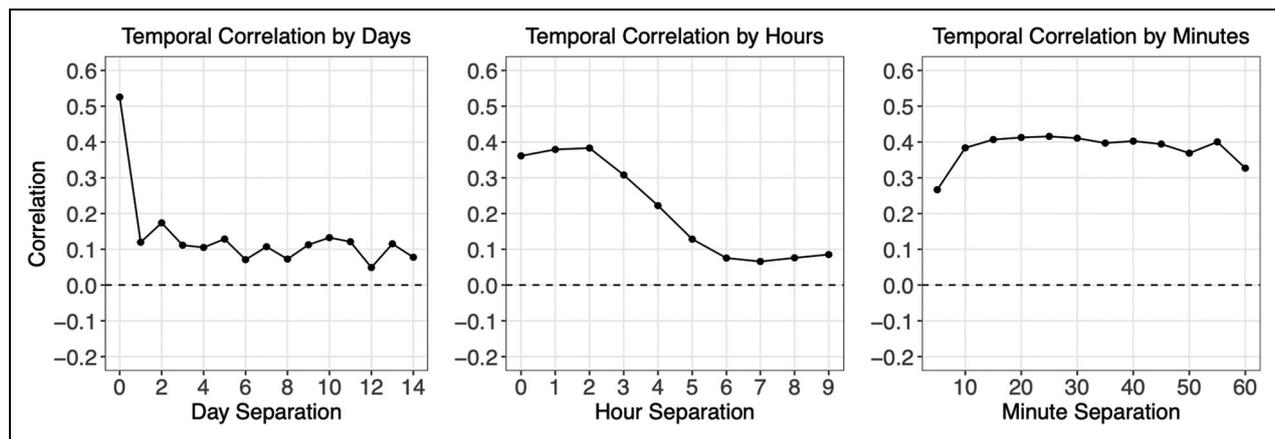


Figure 1. Temporal correlation of CH₄ readings at three different time scales. The log CH₄ enhancements from CH₄ sources observed within the same day are moderately positively correlated; however, this correlation quickly declines when leaks are observed on different days (panel i). Within a day, log CH₄ enhancements are moderately positively correlated when observations of the same source occur within one to 2 h of each other, and then this correlation gradually declines, nearly reaching 0 correlation around the 6-h mark (panel ii). Finally, the correlations between observations of the same source occurring within the same hour remain relatively constant across the minute time scale (panel iii). DOI: <https://doi.org/10.1525/elementa.2020.00143.f1>

verified (i.e., observed at least twice) the more it is driven. We computed verification probabilities assuming a binomial process for leak detection, that is, we assume that each drive-by produces either a detection or nondetection. The probability of detection on a single drive-by is based on the results of our logistic regression model. Given a total number of drive-bys, we compute the probability of verification, which is the probability that at least two of drive-bys result in a detection.

3.5. Monte Carlo synthesis

For our final analysis, we used the results of the percentage difference in size estimation simulation and detection probability analyses to conduct a Monte Carlo simulation that synthesizes our findings and characterizes the ability of ALD to detect leaks and quantify emissions for a hypothetical distribution system under various levels of effort deployed to find leaks. First, we used maximum likelihood to fit a log-normal density to the empirical distribution of emission rates from all 8,191 verified peaks detected during our surveying. Then, using the estimated parameters of the log-normal distribution, we simulated the leak emission rates for a new population of 200 leaks. We treat each of these emission rates as the true emission rate of the leak and use it to compute the system-wide total emissions. We consider these 200 leaks to be representative of (part of) a distribution system where ALD would be used for leak detection and emissions quantification. Next, using the logistic regression results, we computed the detection probability for each of our simulated leaks. We then considered driving by each leak in our sample city a set number of times (e.g., four drive-bys of each leak). Using each leak's detection probability and the number of drive-bys, we simulated the number of times we detected each leak for the given number of drive-bys. Using these results, we computed the number of leaks that were verified (i.e., captured by our mobile survey). For each verified

leak, we randomly adjusted the true leak emission rate to account for variation in size estimation. We made these adjustments using the results from the percentage difference analysis. To calculate the adjusted emission rate, the verified leak from our survey simulation was matched to a representative leak from the percentage difference analysis. We chose the leak with the nearest reference emission rate (in absolute value) to the verified leak from our simulation. Next, we randomly selected a percentage difference from the distribution of percentage differences at the associated number of detections for the representative leak. The true emission rate of the verified leak from the simulation was then either increased or decreased according to this randomly selected percentage difference in order to emulate variation in size estimation. After completing this for each leak that was verified in our simulation, we computed the ALD-based estimate of total emissions from the verified leaks. This process was repeated 500 times for each number of drive-bys, from 2 to 10 drives. For each level of driving effort and each round of the simulation, we computed the proportion of the 200 leaks that were verified, the ratio of the estimated average emission rate of verified leaks to the true average emission rate of those leaks, and the ratio of the estimated total emissions to the true system-wide emissions as a proxy for the proportion of the true system-wide total emissions quantified. For each of these metrics, we also computed the average and empirical 95% confidence intervals (CIs) using the distribution generated over the 500 simulations.

4. Results

4.1. Temporal correlation

Across all time scales, the within-source correlation between log CH₄ concentration enhancement never exceeds moderate levels, with a maximum value just over 0.5. **Figure 1** displays the results of our correlation analysis. When driven multiple times within the same hour,

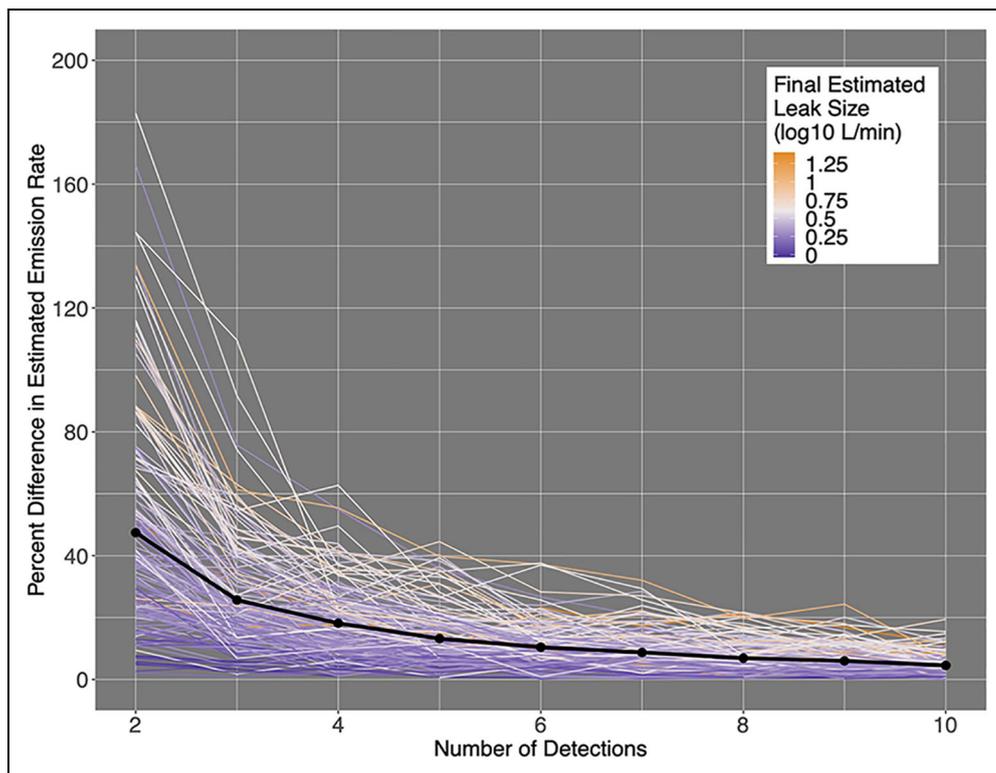


Figure 2. Percentage difference in leak size estimation as a function of the number of detections. The black line shows the average percentage difference for all leaks used in the simulation. Each colored line represents the average percentage difference for a single verified peak relative to its reference emission rate. Across different sizes of leaks, the percentage difference in estimated leak size tends to decrease over the first four detections. The percentage difference varies as a function of leak size with smaller leaks having lower percentage differences and a relatively steady decline over additional detections compared to medium and large leaks. The average percentage difference is positive due to occasional large overestimates of leaks size and the thresholds in place for flagging the detection of a leak. DOI: <https://doi.org/10.1525/elementa.2020.00143.f2>

the correlations across the minute time scale tend to be about 0.4 (**Figure 1**, Panel iii). This correlation persists until it begins to drop at about the 2-h mark (**Figure 1**, Panel ii). The steady decline continues until it asymptotes at about 0.1 for observations separated by more than 6 h.

Our correlation analysis results indicate that there is some degree of CH₄ concentration enhancement measurement repeatability within 2 h but very little after that. This demonstrates the importance of obtaining multiple observations of the same source over time. The CH₄ plumes sampled during ALD surveys are highly variable over time. This finding suggests that it is advantageous to drive past leaks at least 2 h apart in order to maximize the opportunity to observe the full range of CH₄ plumes produced by a CH₄ source. If a single source is observed multiple times over <2 h, the resulting data may not be sufficient to characterize leak size effectively. We expand on this in the next section.

The moderate to weak correlations found in our analysis are likely a result of the numerous factors affecting CH₄ plumes arising from CH₄ emitting sources and therefore the measured CH₄ concentration enhancements. Many of these factors can change dramatically on hourly or daily time scales. These factors include weather, soil conditions, surface type (e.g., pavement vs. soil), traffic,

urban geography, and distance between the leak indication and survey vehicle. We note that our data processing algorithm does not account for wind measurements taken from the survey vehicle because we did not find a reliable signal from this data stream. Similarly, we have explored the use of larger scale prevailing weather conditions (e.g., wind from local weather stations) but also have not found a clear relationship between these weather conditions and measurements taken from the survey vehicle. We note that commercial ALD providers have stated that they are using wind information in their algorithms.

4.2. Percentage difference in leak size estimation simulation

The results of the variation in leak size estimation simulation, shown in **Figure 2**, illustrate the relationship between the number of detections and variation in size estimates, relative to the reference emission rate. We see the largest decrease in variation as detections increase from two to four. As detections increase beyond five and six, we see diminishing reductions of variability. The average percentage difference at two detections is just under +50%. This average percentage difference declines to around 10% after six detections. Uncertainty in individual leak estimates is larger than these average values. For

example, the distribution of an individual leak's percentage differences can range from -40% to $+140\%$, and the skew of the distribution produces an average percentage difference of $+60\%$. Although the uncertainty in estimated size for individual leaks can be large, we note that this percentage uncertainty is still much less than the range of magnitude in leak sizes, which are on the order of $250\times$ different. As a result, more detections improve our ability to discriminate between relative leak sizes.

As shown in **Figure 2**, larger leaks (shown in lighter colors) tend to have a higher percentage difference compared to smaller leaks (shown in dark colors). There are two primary reasons for this phenomenon. First, our algorithm requires that CH_4 levels exceed 110% of the baseline concentration for a detection to be recorded. This requirement creates a lower limit on the excess CH_4 concentrations (e.g., a lower limit of 0.2 ppm under a typical ambient concentration of 2.0 ppm), therefore truncating the distribution of excess CH_4 concentrations. These enhancements are used to estimate leak size, and therefore, leak size estimates are truncated. Second, large leak size estimates typically arise from observing large CH_4 concentration enhancements on just a few drive-bys. These few large observations lead to greater variability in the estimated sizes in our simulation and therefore larger percentage differences. We provide boxplots of CH_4 enhancements from several verified peaks in the SI (see SI S1). These boxplots provide another demonstration of the wide range of variation in CH_4 expressions between and within CH_4 emitting sources.

The CH_4 concentration enhancements observed downwind of a CH_4 emitting source are heavily positively skewed (see SI S1, Figures S1 and S2). In many cases, this positive skew persists even after log transforming the CH_4 enhancements. Because methane concentration is the basis for estimating leak rate, this skew causes average emission rate estimates to be larger than median emission rate estimates, and overestimates of leak size can be unusually large relative to underestimates, which are constrained by our detection threshold. Thus, just two to four large enhancements can have a substantial influence on estimated emission rates, causing overestimates to be large and giving rise to positive differences, on average. In several cases, the variation of within-source CH_4 enhancements seen during our surveys was greater than that seen in our controlled release experiments, suggesting that leak size estimates could be improved through further calibration experiments. In other work, we have implemented a correction for this positive bias (Weller et al., 2020) based on multiple validation studies.

4.3. Probability of detection

The probability of detecting a CH_4 source varied among the different metro areas. **Table 1** displays the overall detection probability for all verified peaks in each city. This overall detection probability is computed by dividing the total number of detections by the total number of drive-bys among all verified peaks. These detection probabilities varied from 35% (Jacksonville) to 64% (Boston). We used our logistic regression analysis to further

investigate how these detection probabilities vary among metro areas and CH_4 sources as a function of estimated emission rate. The fitted logistic regression models for each city had similar slope parameter estimates, prompting our use of a logistic regression model that included a city-specific random-intercept (see SI S2 for further model details).

The results of our logistic regression analysis (**Figure 3**) further demonstrate the variation in detection probabilities among metro areas. A detection probability of 50% corresponds to an emission rate of approximately 0.5 L/min in Boston (the city with the largest intercept), while in City A (the city with the lowest intercept), this detection probability corresponds to an emission rate of about 18 L/min. This variation between metro areas likely arises because of a variety of factors including differences in urban geography (e.g., building density), weather patterns, location of gas pipeline infrastructure, and/or varying traffic levels. This variation indicates that a given level of sampling effort may produce a different proportion of detected leaks in different metro areas. As ALD technology continues to improve quantification of factors that affect leak detection, we anticipate that our understanding of detection probability will also improve.

Our logistic regression model indicated that detection probability increases with increasing emission rates. **Figure 3** shows the positive relationship between estimated leak emission rate and the probability of detection. This result supports the hypothesis that larger leaks are easier to detect than small leaks because they tend to produce larger plumes of CH_4 that are less likely to be diluted below the detection limit. Based on the average regression line, a leak at 1 L/min has an estimated detection probability of 0.36 , while a leak measuring at 10 L/min has an estimated detection probability of 0.56 .

Our detection probability estimates were derived using only the verified peaks from our survey. Because we did not have wind or ethane information, it is not clear how these findings reflect detection probabilities of more sophisticated commercial algorithms, and it is likely that they could be easily improved using this information. For example, we have not included data from locations where elevated CH_4 is observed on only one occasion (singletons). We do not have an estimate of the correspondence between singletons and NG sources for ALD, but the use of ethane to CH_4 ratios could help discriminate singleton (nonrepeated) observed peaks that arise due to NG and non-NG sources. Additional screening or strategic driving practices could also be used to remove or reduce the frequency of non-stationary NG sources. For example, the correspondence of singleton leak indications with bus routes or surveying when buses are not running could be used to reduce indications from NG buses (A van Pelt, personal communication, October 7, 2020). Additionally, there is a trade-off between sensitivity and specificity when defining the threshold for designating a reading as elevated. Lowering the threshold (e.g., to 105%) would likely increase detection probability but would also create more false positives. Among verified peak locations, we

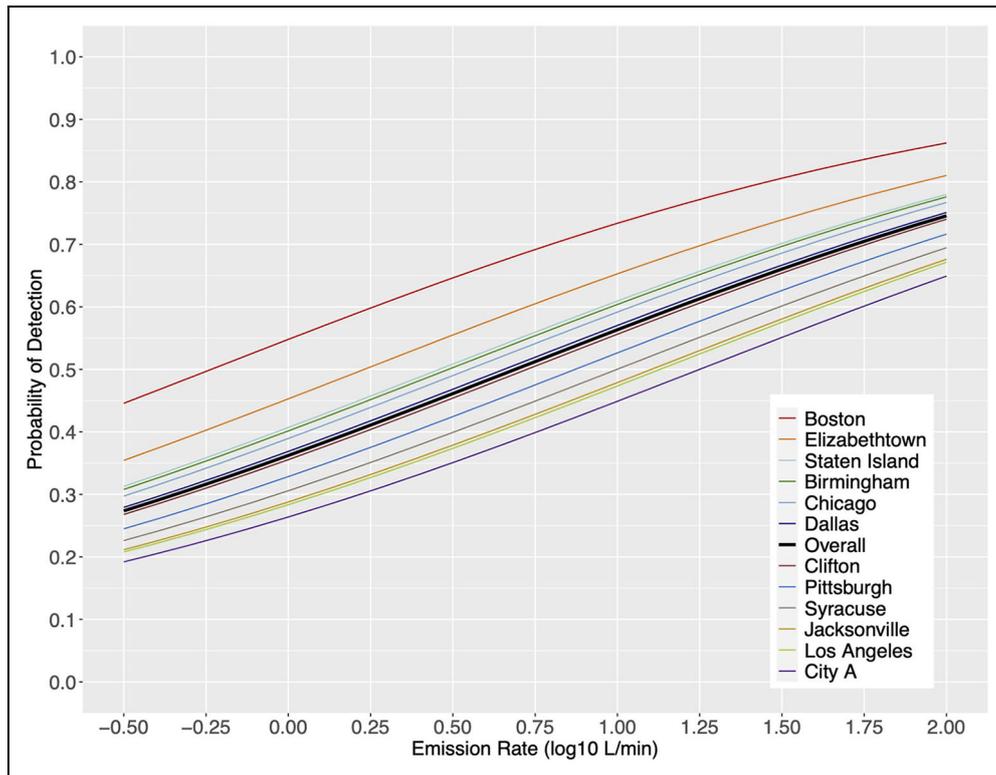


Figure 3. Probability of detecting a leak as a function of leak size. In all metro areas, there is evidence that the probability of detecting a leak increases with emission rate. The black line represents the regression over all the metro areas, and each individual color represents an individual city. The detection probability varies among metro areas with Boston having the highest intercept and therefore the highest detection probabilities across leak size while City A has the lowest. Burlington, Indianapolis, and Mesa are omitted from this analysis because very few (<15) leaks were detected in those metro areas. DOI: <https://doi.org/10.1525/elementa.2020.00143.f3>

previously estimated that 76% (95% CI [60, 88]) of verified peaks correspond to an NG source (Weller et al., 2018b). We performed a sensitivity analysis to examine how the inclusion of singleton observed peaks affects the estimated detection probabilities. When singletons are included in the logistic regression model, estimated detection probabilities decrease across all metro areas and estimated emission rates. See the SI for discussion on limitations of the logistic regression analysis (SI S2) and details of our sensitivity analysis (SI S3).

4.4. Probability of verification

The probability of verifying a leak increases with leak size and sampling effort. There is a rapid increase in verification probability across all sizes of leaks when going from two to five drive-bys. We see diminishing returns in verification probability after five drive-bys, similar to the results of the percentage error simulation. As sampling effort increases, large leaks tend to reach verification probabilities near one. These high verification probabilities suggest that the largest leaks will be detected and reported even under moderate or low sampling effort. This finding demonstrates the efficacy of ALD for detecting the largest leaks that disproportionately contribute to overall emissions (Weller et al., 2019). In contrast, the smallest leaks only reach a verification probability of 0.85 after 12

drives, suggesting that some of these leaks may go unreported even in the presence of substantial sampling effort (Figure 4).

The results from our percentage difference simulation indicated that observation of large leaks has the largest variation in estimated emission rates, especially with a small number of detections. The probability analyses suggest that these large leaks have the highest detection and verification probabilities. In combination, these findings indicate that even under moderate sampling (five to six drive-bys), we can expect moderate uncertainty in size estimates because we are likely to obtain multiple detections from these large leaks.

4.5. Monte Carlo synthesis

Figure 5 displays the results of our Monte Carlo synthesis. The red line displays the proportion of detected leaks as a function of ALD sampling effort. The proportion of detected leaks is based on applying the results of our detection probability analysis. Similarly, the blue and purple lines display the ability of ALD to estimate two population characteristics of leak emission rates: the average emission rate of detected leaks (blue) and the total emissions of all leaks (purple) as a fraction of the system total. These estimates of population characteristics, and their uncertainty bands, are based on which leaks were detected and the

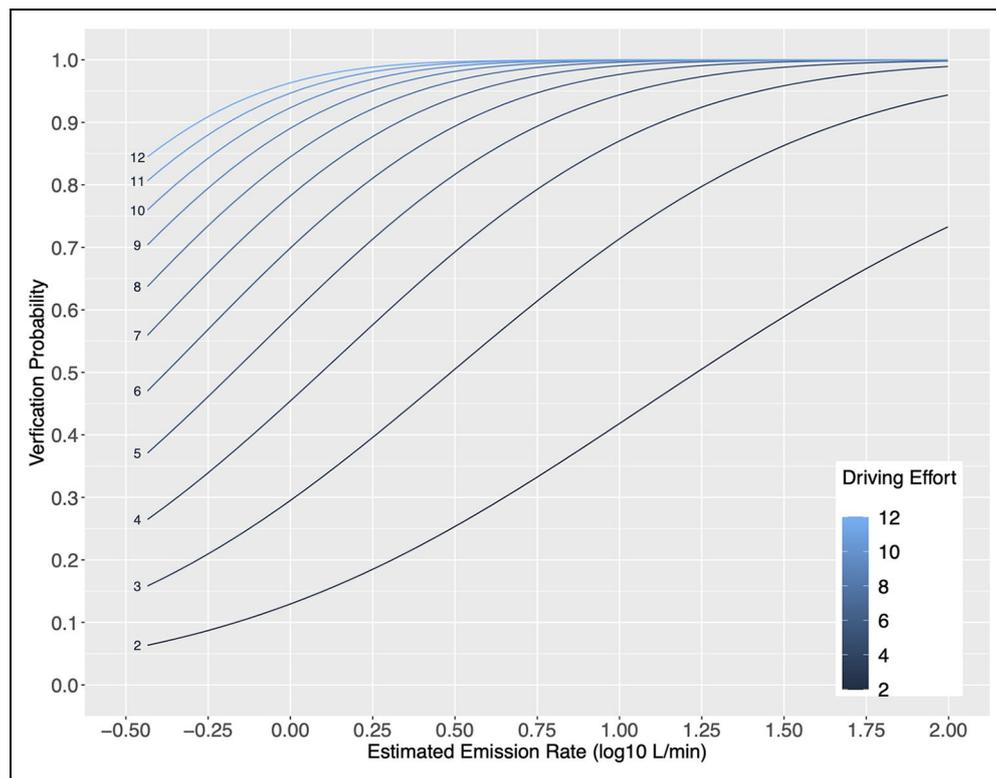


Figure 4. Probability of verifying a leak as a function of leak size over differing sampling efforts. For a leak to be verified and therefore reported in our survey results, our algorithm requires that the instrument detect an elevated CH_4 reading on two or more drive-bys. The estimated probability of verification as a function of survey effort increases more rapidly for large leaks than small leaks and is still less than one for the smallest leaks even after 12 drives, suggesting that a fraction of small leaks may go undetected given limited survey resources. DOI: <https://doi.org/10.1525/elementa.2020.00143.f4>

bias and variance of individual leak size estimates derived from the percentage difference analysis.

As expected, increased sampling effort leads to a greater proportion of leaks being detected in our Monte Carlo synthesis (red line in **Figure 5**). After four drive-bys, we typically verify 85% \pm 5% of the 200 leaks in the population. The average proportion of leaks captured increases quickly up to six drive-bys and then more slowly afterward. Approximately 70% of leaks are verified after three drives and approximately 90% are verified after five drives. If sampling an area with 10 drives or more was economically and practically feasible, and these drives took place over multiple days, it is likely that nearly all of the leaks would be detected and that the remaining undetected leaks would likely be small.

In our Monte Carlo synthesis, the average leak rate tends to be overestimated, but variation in this estimate declines with sampling effort. The overestimation of the average leak rate arises due to overestimation of individual emission rates. This overestimation is largest at low levels of sampling effort (two to four drives) because the largest leaks tend to be detected at this level of sampling effort, and these leaks have the largest percentage difference in size estimation. As sampling effort increases, more small leaks are detected, reducing the overestimation, and leaks are detected more frequently, which reduces the variation in their estimated emission rates.

Similar to the proportion of leaks verified, the proportion of total system emissions quantified also increases with sampling effort in our Monte Carlo synthesis (purple line **Figure 5**). This occurs because of the increasing number of leaks verified, and therefore quantified, with sampling effort. Beyond four drives, the average proportion of total emissions quantified exceeds one due to overestimation in emission rate estimates.

Our Monte Carlo simulation demonstrates how quantification and verification and their uncertainties may vary relatively to one another. The proportion of total system emissions quantified (purple line) increases more quickly than the proportion of leaks detected (red line). The rapid increase in the proportion of total system emissions quantified occurs because the largest leaks, which often account for a majority of the total emissions, are easier to detect. However, the proportion of total system emissions quantified rises above one due to the aforementioned finding of overestimation of emission rates. The uncertainty band around the proportion of total system emissions quantified is much wider than the uncertainty around the proportion of leaks detected. This uncertainty arises due to the variation in emission estimates outlined in our percentage difference simulation and is similar to the uncertainty for estimating average leak emission rate (blue line).

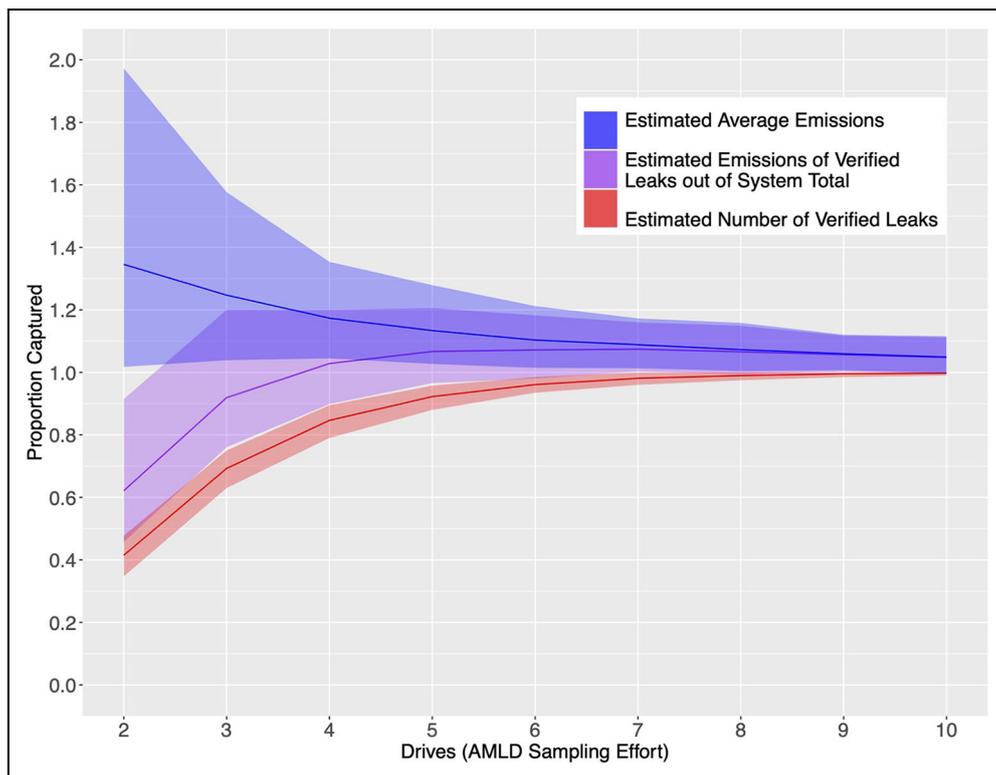


Figure 5. Results of a simulation investigating the ability of advanced leak detection (ALD) to detect leaks and estimate total emissions. This figure displays three features of ALD surveys for detecting leaks and quantifying emissions as a function of sampling effort. The red line displays the average proportion of the 200 leaks that are verified (i.e., reported) for a given level of sampling effort. The blue line displays the estimated average emission rate from verified leaks found during the survey divided by the true average emission rate from the verified leaks found during the survey. The purple line displays the estimated total emissions from verified leaks found during the survey divided by the true system-wide emissions from all 200 leaks. The colored bands display the empirical 95% uncertainty intervals for these quantities from our simulation. DOI: <https://doi.org/10.1525/elementa.2020.00143.f5>

5. Summary and discussion

Based on the results of our analyses, we recommend five to eight drives as an attainable surveillance effort that will capture a large majority (>90%) of leaks and provide reasonable emissions quantification in order to estimate average emissions and discriminate among relative leak sizes. These drives would ideally take place on separate days to improve the chances of observing the leak's range of variability, but if this is not possible, they should be spaced at least 2 h apart. Throughout an ongoing ALD survey, analysis of already collected survey data to assess detection probabilities could be used to increase or decrease survey effort in order to achieve a desired capture percentage. Such efforts have been explored in sophisticated detail for NG production systems using the FEAST model (Fox et al., 2019) and could be readily applied to distribution systems as well.

In combination, our temporal correlation, percentage difference, and detection probability results show the variation in atmospheric CH₄ concentrations arising from CH₄-emitting sources in metro areas. Elevated CH₄ concentrations near a CH₄-emitting source may not be detected with ALD if wind or soil conditions are unfavorable. When sources are detected, there can be large variability among a source's observed CH₄ concentrations

over time. A CH₄ source, such as NG leak, can present plumes with large or small CH₄ enhancements, making the quantification of the emission rate difficult. This is especially true for large leaks, which can produce very small methane enhancements even at relatively far distances from the leak expression location.

The variability between source emission rates is much greater than the apparent within-source variability, which enables the discrimination of relative emission rates. This discrimination improves when sources are detected multiple times. Although turbulent CH₄ plumes affect leak size estimates, our results show how relative size discrimination is still possible through appropriate sampling effort and timing of drive-bys. We note that ALD does not detect every leak every time. Some leaks are likely to go undetected by mobile platforms regardless of sampling effort (Clark et al., 2012; Kerans et al., 2012). Despite these limitations, ALD surveys are becoming an integral part of pipeline management because of their ability to quickly cover large regions, detect leaks, and discriminate relative leak sizes.

The results of our percentage difference Monte Carlo simulation illuminate avenues for adjusting and improving our emissions estimates. More experiments, run under a wide variety of environmental conditions, are needed to better imitate the variation and patterns in excess CH₄

measurements seen in the field. We anticipate that data produced by these experiments would reduce the positive percentage differences and positive bias in size estimates observed in previous validation studies (Weller et al., 2018b). Additionally, estimation methods that are more robust to the influence of these large observations (e.g., through transformations or assuming a skewed distribution) could be used to improve leak size estimates.

In our synthesis simulation, we assumed that all leaks in the city were driven an equal number of times. This assumption is unreasonable in practice because certain leaks, such as those located on arterial roadways, will be driven by more frequently while others, such as those located on collector roads, will be driven less frequently during an ALD survey. Our analysis also does not account for the expenditures associated with ALD surveying, such as the cost of driving or the time required to complete a survey based on the size of the survey region or traffic conditions. These conditions will vary locally and regionally, making ALD more economical in certain areas than others.

There are several avenues of future research that we have not addressed here. We have quantified several features of ALD surveys, but there remain open questions regarding the underlying causes of these features. For example, we anticipate that weather (e.g., rain the previous day, wind) and proximity of the pipeline to the roadway explain some of the variation in leak detectability. As previously alluded to, a cost–benefit analysis would demonstrate the economic effectiveness of ALD. Another area where further research is needed is quantifying the false positive rate of ALD. Understanding the false positive rate could help screen out elevated CH₄ readings that are due to nonstationary sources. Finally, we have outlined ways to improve our leak emission rate estimates.

The quantitative features we have noted here are specific to our instrumentation and data processing algorithms. Although these quantitative features will change with different instruments, inlet heights, and data processing algorithms, we anticipate that the general qualitative features will still hold. For example, not every leak will be detected every time and CH₄ plumes will be highly variable over time. Our results may not be representative of commercial algorithms, especially those that reliably use ethane for source attribution or incorporate the effects of wind and atmospheric stability and their impact on dispersion and dilution.

Data accessibility statement

The data and data descriptions for reproducing the results in this article are available on GitHub: <https://github.com/zdweller/ElementaDetectionProbabilities>.

Supplemental files

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

Tables S1–S2. Figures S1–S3. Text S1–S4. (DOC).

Acknowledgments

The authors thank three anonymous reviewers for feedback that improved the manuscript. They also thank Aaron van Pelt and one anonymous local distribution company employee for comments that improved the manuscript and helped them qualify their findings. They thank the members of the von Fischer laboratory for constructive feedback that improved the manuscript.

Funding

Emily Luetschwager was supported by the CSU Energy Institute's Summer Internship program through a grant from Mr. Alan Cogen. Zachary Weller and Joseph von Fischer were supported through a grant from Environmental Defense Fund.

Competing interests

The authors have declared that no competing interests exist.

Author contributions

Contributed to conception and design: EL, JvF, ZW.

Contributed to acquisition of data: EL, JvF, ZW.

Contributed to analysis and interpretation of data: EL, JvF, ZW.

Drafted and/or revised the article: EL, JvF, ZW.

Approved the submitted version for publication: EL, JvF, ZW.

References

- Chamberlain, SD, Ingraffea, AR, Sparks, JP.** 2016. Sourcing methane and carbon dioxide emissions from a small city: Influence of natural gas leakage and combustion. *Environmental Pollution* **218**: 102–110. DOI: <http://dx.doi.org/10.1016/j.envpol.2016.08.036>.
- Clark, T, Conley, E, Kerans, M, Piazza, M.** 2012. Picarro surveyor leak detection study: Diablo side-by-side study. *Internal Study Report, Pacific Gas and Electric*, in press. Available at <https://edocket.dcpsec.org/apis/api/filing/download?atta>.
- Cuellar, H.** 2020. Text - h.R.133 - 116th congress (2019–2020): Consolidated appropriations act, 2021. Available at <https://www.congress.gov/bill/116th-congress/house-bill/133/text/pl>. Accessed 22 March 2020.
- Environmental Defense Fund.** 2020. Available at <https://www.edf.org/climate/methanemaps>. Accessed 3 August 2020.
- Environmental Systems Research Institute.** 2015. ArcGIS Desktop: Version 10.4.1. Environmental Systems Research Institute. Available at <https://desktop.arcgis.com/en/>.
- Fox, TA, Ravikumar, AP, Hugenholtz, CH, Zimmerle, D, Barchyn, TE, Johnson, MR, Lyon, D, Taylor, T.** 2019. A methane emissions reduction equivalence framework for alternative leak detection and repair programs. *Elementa: Science of the Anthropocene*. DOI: <http://dx.doi.org/10.1525/elementa.369>.
- Gallagher, ME, Down, A, Ackley, RC, Zhao, K, Phillips, N, Jackson, RB.** 2015. Natural gas pipeline

replacement programs reduce methane leaks and improve consumer safety. *Environmental Science & Technology Letters* **2**(10): 286–291. DOI: <http://dx.doi.org/10.1021/acs.estlett.5b00213>.

- Jackson, RB, Down, A, Phillips, NG, Ackley, RC, Cook, CW, Plata, DL, Zhao, K.** 2014. Natural gas pipeline leaks across Washington, DC. *Environmental Science & Technology* **48**(3): 2051–2058. DOI: <http://dx.doi.org/10.1021/es404474x>.
- Kerans, M, Clark, T, Conley, E, Crosson, E, Piazza, M.** 2012. Picarro surveyor leak detection study: Sacramento side-by-side study. *Internal Study Report, Pacific Gas and Electric*, in press. Available at <https://edocket.dcpsc.org/apis/api/filing/download?atta>.
- LGR-ABB.** 2020. Available at <https://new.abb.com/products/measurement-products/analytical/laser-gas-analyzers/abb-ability-mobile-gas-leak-detection-system>. Accessed 3 August 2020.
- Phillips, NG, Ackley, R, Crosson, ER, Down, A, Hutyra, LR, Brondfield, M, Karr, JD, Zhao, K, Jackson, RB.** 2013. Mapping urban pipeline leaks: Methane leaks across Boston. *Environmental Pollution* **173**: 1–4. DOI: <http://dx.doi.org/10.1016/j.envpol.2012.11.003>.
- Picarro.** 2020. Available at https://naturalgas.picarro.com/solutions/leak_survey/advanced_mobile_leak_survey. Accessed 3 August 2020.
- R Core Team.** 2018. Available at <https://www.r-project.org/>.
- Sanchez, NP, Zheng, C, Ye, W, Czader, B, Cohan, DS, Tittel, FK, Griffin, RJ.** 2018. Exploratory study of atmospheric methane enhancements derived from natural gas use in the Houston urban area. *Atmospheric Environment* **176**: 261–273. DOI: <http://dx.doi.org/10.1016/j.atmosenv.2018.01.001>.
- Van Rossum, G.** 2015. *Python 2.7.10 language reference*. Samurai Media Limited.
- Von Fischer, JC, Cooley, D, Chamberlain, S, Gaylord, A, Griebenow, CJ, Hamburg, SP, Salo, J, Schumacher, R, Theobald, D, Ham, J.** 2017. Rapid, vehicle-based identification of location and magnitude of urban natural gas pipeline leaks. *Environmental Science & Technology* **51**(7): 4091–4099. DOI: <http://dx.doi.org/10.1021/acs.est.6b06095>.
- Weller, ZD, Hamburg, SP, Von Fischer, JC.** 2020. A national estimate of methane leakage from pipeline mains in natural gas local distribution systems. *Environmental Science & Technology* **54**(14): 8958–8967. DOI: <http://dx.doi.org/10.1021/acs.est.0c00437>.
- Weller, ZD, Hoeting, JA, von Fischer, JC.** 2018a. A calibration capture-recapture model for inferring natural gas leak population characteristics using data from Google Street View cars. *Environmetrics* **29**(7): e2519. DOI: <http://dx.doi.org/10.1002/env.2519>.
- Weller, ZD, Roscioli, JR, Daube, WC, Lamb, BK, Ferrara, TW, Brewer, PE, Von Fischer, JC.** 2018b. Vehicle-based methane surveys for finding natural gas leaks and estimating their size: Validation and uncertainty. *Environmental Science & Technology* **52**(20): 11922–11930. DOI: <http://dx.doi.org/10.1021/acs.est.8b03135>.
- Weller, ZD, Yang, DK, von Fischer, JC.** 2019. An open source algorithm to detect natural gas leaks from mobile methane survey data. *PLoS One* **14**(2): e0212287. DOI: <http://dx.doi.org/10.1371/journal.pone.0212287>.
- Zhou, X, Montazeri, A, Albertson, JD.** 2019. Mobile sensing of point-source gas emissions using Bayesian inference: An empirical examination of the likelihood function. *Atmospheric Environment* **218**: 116981. DOI: <http://dx.doi.org/10.1016/j.atmosenv.2019.116981>.

How to cite this article: Luetschwager, E, von Fischer, JC, Weller, ZD. 2021. Characterizing detection probabilities of advanced mobile leak surveys: Implications for sampling effort and leak size estimation in natural gas distribution systems. *Elementa: Science of the Anthropocene* 9(1). DOI: <https://doi.org/10.1525/elementa.2020.00143>

Domain Editor-in-Chief: Detlev Helmig, Boulder AIR LLC, Boulder, CO, USA

Associate Editor: Ian C. Faloona, Department of Land, Air and Water Resources, University of California, Davis, CA, USA

Knowledge Domain: Atmospheric Science

Published: June 9, 2021 **Accepted:** April 21, 2021 **Submitted:** September 23, 2020

Copyright: © 2021 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 The Open Access icon, which is a stylized padlock with an open keyhole.