ATMOSPHERIC ENVIRONMENT: X 17 (2023) 100201

Contents lists available at ScienceDirect

Atmospheric Environment: X

journal homepage: www.journals.elsevier.com/atmospheric-environment-x

Measurement-based emissions assessment and reduction through accelerated detection and repair of large leaks in a gas distribution network

Sean MacMullin [a,*], François-Xavier Rongère [b,1]

[a] Picarro, Inc, Santa Clara, CA, USA
[b] Pacific Gas and Electric Company (PG&E), San Francisco, CA, USA

ARTICLE INFO

Keywords:
Methane emissions
Natural gas
Leak detection
Monte Carlo simulation
Bayesian inference

ABSTRACT

A common practice within the current regulatory framework for gas distribution uses an approach where operators report their network emissions by applying an average emission factor for all leaks, sometimes sorted by pipe material and type of assets. Such an approach has two drawbacks: first, it does not account for the specificities of the gas systems and the maintenance processes of the operators; and second, it does not enable the prioritization of large leaks that is key for an effective emissions abatement program. This article describes a method using a mobile leak detection and quantification system to assess methane emissions from a gas distribution network and to reduce them by accelerating the detection and repair of larger leaks. The approach allows for data-driven system-wide emissions quantification that is specific to the network and not subject to operator's leak detection practices that may affect their traditional emission factor-based reporting. Furthermore, we show that for a sensor with a sufficiently low detection limit, the calculated emissions are independent of the precision of the measurement if the uncertainties are correctly addressed. Such a result is important because it assures that methane emissions estimates are not biased and can be used to assess the performance of abatement programs. Finally, we illustrate how the approach can be practically implemented through a program where the largest leaks are rapidly identified and repaired to abate methane emissions while minimizing costs.

1. Introduction

Numerous studies have been performed to update methane emission estimates initially established by the Gas Research Institute (GRI) for the US Environmental Protection Agency (EPA) in 1992-1996 (Harrison et al., 1996) and to provide a baseline for abatement efforts. They all demonstrated the same pattern; most of the leaks are small and generally only contribute marginally to the overall emissions while a relatively small number of leaks can be three to four orders of magnitude larger than the median leak and therefore account for most of the emissions.

For example, Fig. 1 displays the results of measurements performed by GRI in 1996 and by Washington State University (WSU) in 2015 on the pipelines of gas distribution systems (Lamb et al., 2015). It shows that leak sizes vary from less than 10^{-2} ft^3/h up to more than 10^2 ft^3/h[2](a range of more than four orders of magnitude), with a small number of large leaks dominating the total emissions. In fact, this phenomenon is even more patent now; GRI reported that 20% of leaks, greater than 10 ft^3/h, accounted for about 80% of methane emissions while WSU observed that only 2.2% of leaks were greater than 10 ft^3/h but they still represented more than 50% of total emissions.

These results were recently summarized by Brandt et al. (2016) (see Fig. 2), who demonstrated that across assets of the natural gas supply chain as well as across component types, most emissions were owing to a small fraction of leaks, typically 5% or less, that are generally called *Super Emitters*. The term *Super Emitter* is sometimes used in absolute to designate the largest sources of a territory such as in Duren et al. (2019) or greater than a specific threshold such as in Collins et al. (2022), sometimes as a qualification of specific facilities that release more gas than others such as in Zavala-Araiza et al. (2017) or a larger proportion of the gas they produce (Zavala-Araiza et al., 2015). We use it here, in the spirit of Brandt et al.'s article, as a relative designation representing the largest leaks of a distribution system. In that definition, *Super Emitters* in distribution systems may not be considered as large for other facilities or across a territory. This skewed leak size distribution

* Corresponding author.
E-mail address: smacmullin@picarro.com (S. MacMullin).
[1] Current address: Picarro, Inc. Santa Clara, CA, USA.
[2] 1 ft^3/h = 19.2 g/h at Normal Temperature and Pressure (NTP).

https://doi.org/10.1016/j.aeaoa.2023.100201
Received 11 September 2022; Received in revised form 29 November 2022; Accepted 5 January 2023
Available online 12 January 2023

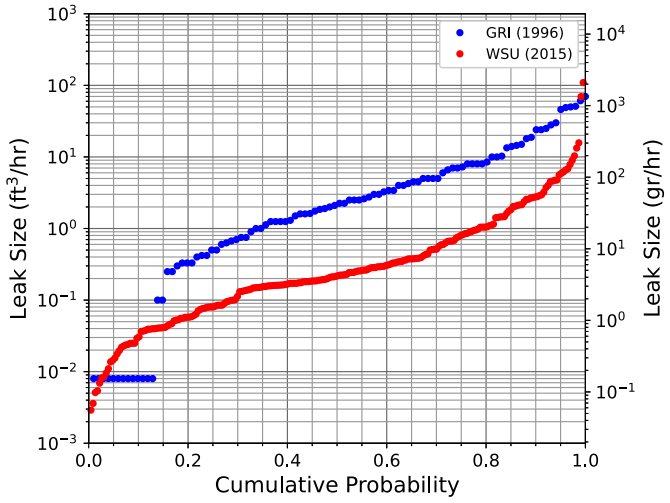

Fig. 1. Cumulative distribution of leak sizes for leaks in a gas distribution network reported by Harrison et al., GRI (1996) and Lamb et al., WSU (2015).

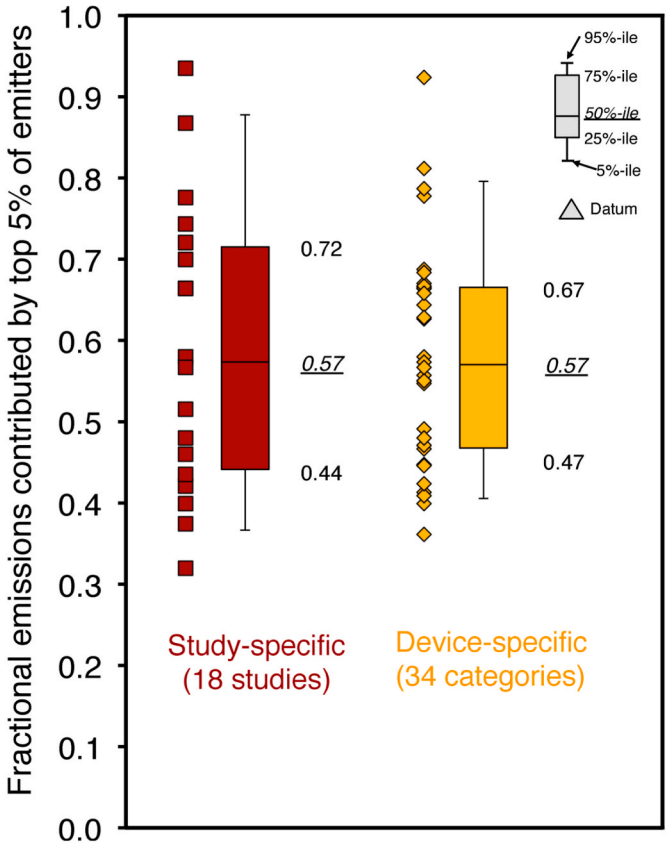

Fig. 2. Share of methane emissions contributed by top 5% of emitters. Reprinted with permission from Brandt et al. (2016). Copyright 2016 American Chemical Society.

represents both a challenge and an opportunity: a challenge because any attempt to characterize the methane emissions of a system must collect large samples to correctly capture the portion of large leaks; an opportunity because substantial reduction can be accomplished with an optimal survey and repair effort if these large leaks are identified rapidly.

2. Methods

2.1. Gas leak detection and emissions quantification

This study used a Picarro vehicle-based mobile platform that identifies the natural gas plumes induced by pipeline leaks as they propagate in the atmosphere and intersect the path of the vehicle. The concentrations of the two main compounds of natural gas: methane and ethane, are established by a parts-per-billion sensitivity gas analyzer based on Cavity Ring Down Spectroscopy (CRDS) (Crosson, 2008) four times per second providing a good spatial resolution of the concentration signal. The system also measures GPS position, wind speed and direction, atmospheric conditions (pressure, humidity and temperature), and uses algorithms to infer the location and flow rates of detected sources. The ratio of ethane and methane concentrations, specific to the utility's natural gas, is used to identify and discard biogenic methane sources that do not contain ethane. Each area of the distribution system is driven four to six times, typically at night when the atmosphere is more stable, and over at least two nights to take advantage of wind shifts to maximize the probability of detecting leaks downwind of the gas infrastructure. The small proportion of pipeline segments still not covered by the measurement system, generally less than 20%, are flagged and accounted for in the emission calculation by extrapolating results obtained on covered areas.

As the vehicle transects the gas plume the emission (flow) rate of a source is calculated by a control volume approach (Conley et al., 2017). The methane flow rate Q is derived from the volumetric flux equation:

$$Q = \iint [C(y,z) - C_0] \cdot u(y) dy dz, \quad (1)$$

where C is the concentration at each measurement point of the cross-sectional area of the plume and C_0 is the background methane concentration. The vehicle as it drives downwind of the leak samples the concentration along a line through the plume in the y direction. The height of the plume, z, is inferred from the measured width in the y direction and the atmospheric conditions. The quantity u is the component of wind speed, measured by the anemometer, normal to the path of the vehicle.

The method relies on the vehicle to make multiple passages through the network to increase the probability to detect leaks. Multiple detections of a plume originating from a source are aggregated using a geospatial clustering algorithm such as DBSCAN (Ester et al., 1996; Birant and Kut, 2007), HDBSCAN (Campello et al., 2013), or OPTICS (Ankerst et al., 1999; Agrawal et al., 2016). The parameters of the clustering algorithm indicate the separation distance scale at which the same plume may be detected over multiple passages. The optimal scale that achieves as close as possible to a 1-to-1 relationship between a cluster of detections and a gas leak, was determined through fitting procedure using hundreds of thousands of detections linked leaks confirmed by gas operators worldwide and is typically 25–30 m. An example of the geospatial clustering is shown in Fig. 3. Detecting a plume multiple times from a single source also serves to improve the precision of the emission estimate. Using the associations determined by the clustering algorithm, multiple independent measurements of plumes from a given source are averaged to report an estimate for the leak size.

2.2. Validation of the measurement

A study performed by NYSEARCH between 2015 and 2017 on three mobile leak detection systems provided a solid validation data set (D'Zurko and Mallia, 2017). The test covered three orders of magnitude in leak size consistent with observations in the field. A similar validation study using the Picarro system has been performed annually since 2018 at the Pacific Gas and Electric Company (PG&E) Gas Safety Academy in Winters, CA. Controlled leaks were setup in various above-ground and below-ground configurations representing leaks typically found in a

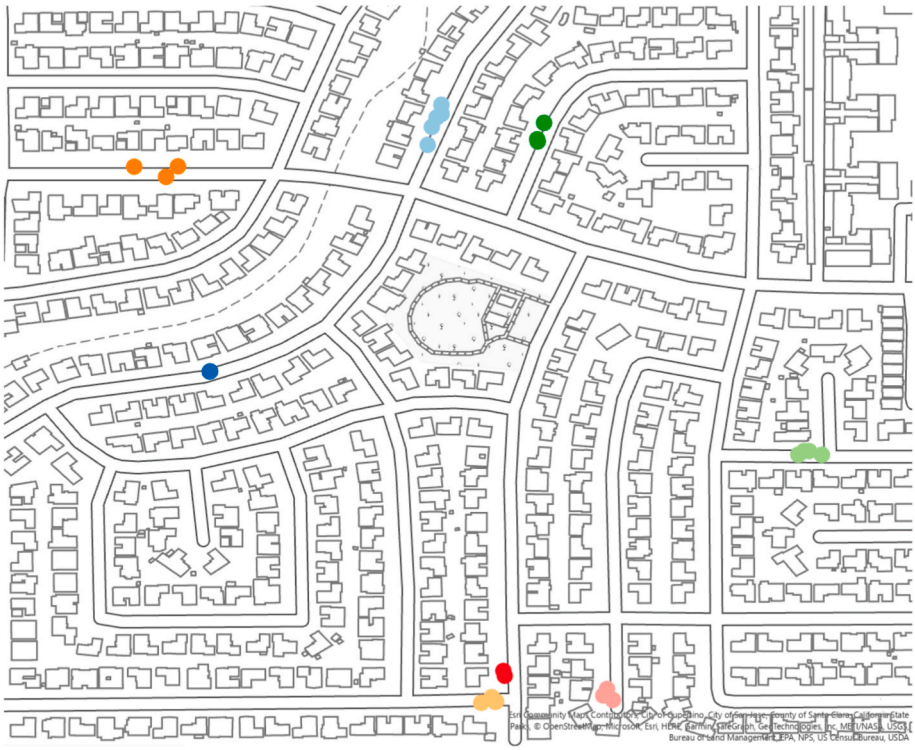

Fig. 3. Example of geospatial clustering to associate one or more detections with a source. Each point represents the location where a single plume was detected. Measurements that are clustered together are shown by like colors. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

distribution network. An inline mass-flow controller provided a precise measurement of the actual flow rate of the leak. Data were collected by the Picarro system according to the standard driving protocol which includes six passes by each leak with three in each direction. In addition to the controlled testing, 53 below-ground leaks that were identified by the Picarro system on PG&E's network were validated using using Bacharach's High Flow sampler following a protocol assuring that all gas was captured. The field validation exercise focused on leaks that were measured by the Picarro system as more than 5 ft^3/h. The purpose of these field tests was to verify that the accuracy observed in a broad range of conditions was consistent with the measurement performed during the validation tests in application of the method developed by NYSEARCH in 2018 with the support of PHMSA (PHMSA, 2019). Fig. 4 summarizes the leak flow rate as estimated by the mobile systems compared to the actual values as measured with a High Flow sampler or as set in the case of controlled leaks. It was observed that the mobile quantification systems were able to estimate the order of magnitude of the leak flow rates: 78% of the data points were within a one order of magnitude band – a factor of $\sqrt{10}$ times greater and $\sqrt{10}$ times less than actual values.

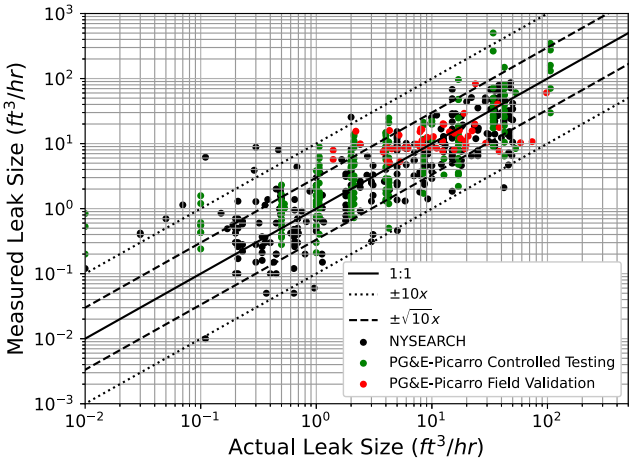

Fig. 4. Validation testing unity plot. Quantification system measurements versus actual leak sizes measured with a High Flow sampler that has an uncertainty of less than 30% (PHMSA, 2019), or as set using a mass-flow controller in the case of controlled leaks with a precision much better of the quantification system to be validated.

3. Theory and calculations

3.1. Emissions measurements in a distribution network

Since 2014, PG&E has used the Picarro vehicle-based methane detection system for its leak survey program (Redding and Glaze, 2015). Any portion of PG&E's distribution system is inspected at the minimum every three years, some areas being surveyed as often as every year. In 2018, to further reduce methane emissions, PG&E introduced a *Super-Emitter Program*, which implemented an additional survey focused only on large leaks. Performed every year on the entire distribution system, it aims to detect only leaks greater than 10 ft^3/h and prioritize their repair to take advantage of their disproportionally large contribution to methane emissions. Based on the leak size distribution observed by WSU, rapidly detecting and repairing these leaks could lead to up to 50% emission abatement. However, in order to correctly assess the impact of the program, special attention must be paid to the representation of uncertainties because, even if the quantification system is evenly calibrated, i.e. it has a symmetrical probability to overestimate or underestimate the flow rate of a given leak, the skewed leak size of the gas distribution system means that, for a measured value of a larger flow rate especially greater than 10 ft^3/h, the actual leak size has a much

higher probability to be overestimated than underestimated.

Fig. 5 illustrates this phenomenon for a leak measured by the mobile system as 1 ft^3/h. The measurement uncertainty spans one order of magnitude (A), consistent with the validation testing. Because of the heavy-tailed leak size distribution, the positive interval, Δ_+, is much smaller than the negative interval Δ_-. This corresponds to approximately 70% chance to overestimate the leak compared to 30% to underestimate the leak. Therefore, the most probable value corresponding to the measurement will be less than the measured value.

The mobile system covers a large range of leak sizes from less than 0.1 to more than 100 ft^3/h. The low Minimum Detection Limit (MDL) associated with a consistent quantification precision across the full range of leak sizes in a distribution network is key to correctly capture the uneven proportion of small and large emitters. In addition, the quantification capability is used to point out larger leaks avoiding the operator the cost of investigating all methane indications. The MDL is not a fixed threshold, but varies as a function of measurement conditions such as terrain, wind speed, and atmospheric stability (Conrad et al., 2022). It must therefore be much lower than the leak sizes that substantially contribute to the total emissions to avoid cases of missed detections and mischaracterization around the MDL that will affect the overall emissions assessment. Using a detection system with an MDL close to the size of the large leaks is not adequate to correctly identify them and evaluate the overall emissions.

3.2. Measurement based emission factors

To evaluate the impact to the system-wide emissions where only the largest leaks are prioritized for repair, leaks and detections are classified in four decade bins from 10^{-2} to 10^2 ft^3/h. These bins can be adjusted as a function of the threshold used to define a large leak. The probability for actual leaks to belong in each bin can be calculated using Bayesian inference:

$$P\langle A_i|B_j\rangle = \frac{P\langle B_j|A_i\rangle \cdot P\langle A_i\rangle}{P\langle B_j\rangle}, \tag{2}$$

where A_i is the statement: *the actual size of the leak is in the bin defined by* $[10^i, 10^{i+1}[$ and B_i is the statement: *the leak size as estimated by the mobile system is in the bin defined by* $[10^j, 10^{j+1}[$. The indices i and j vary from -2 to 1; if i or $j = 1$ the interval is $[10, \infty[$, if i or $j = -2$ the interval is $[0, 0.1[$.

$P\langle A_i|B_j\rangle$ represents the probability for the statement A_i to be true if the statement B_i is true.

$P\langle B_j|A_i\rangle$ represents the probability for the statement B_j to be true if the statement A_i is true.

$P\langle A_i\rangle$ represents the probability of A_i to be true.

$P\langle B_j\rangle$ represents the probability of B_j to be true.

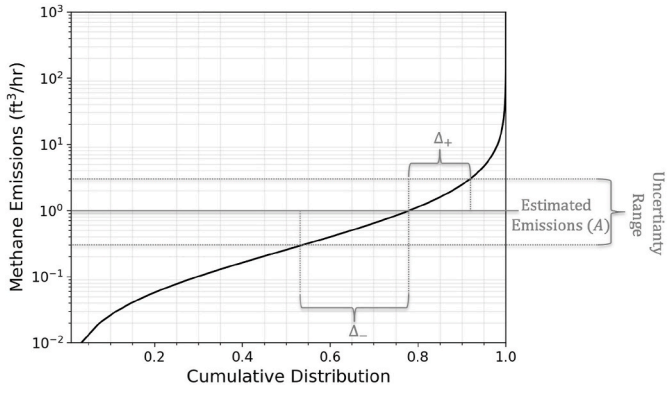

Fig. 5. A single leak measured at $A = 1$ ft^3/h with a $\pm\sqrt{10}$x uncertainty from a skewed distribution will have a larger probability to be overestimated (Δ_-) than underestimated (Δ_+).

$P\langle B_j\rangle$ is calculated using the formula:

$$P\langle B_j\rangle = \sum_i P\langle B_j|A_i\rangle \cdot P\langle A_i\rangle. \tag{3}$$

For this analysis, the experimental data of WSU for distribution mains and services have been fit to a lognormal distribution, shown in Fig. 6. The average flow rate assigned to each bin noted here as $EF(A_i)$ is obtained from the fit to the WSU distribution and is summarized in Table 1.

The validation data described above were used to represent the precision of the mobile quantification system. The distribution of errors observed during the tests was modeled as a lognormal function with a width, $\sigma = 0.95$. The matrix $[v_{ij}]$ defined as $v_{ij} = P\langle B_j|A_j\rangle$ is represented in Table 2. From here, we calculate the inverse matrix $[v_{ij}^{-1}]$ shown in Table 3.

In addition, the flow rate to be assigned to a leak measured as in a bin B_j is obtained by the formula:

$$EF(B_j) = \sum_i P\langle A_i|B_j\rangle \cdot EF(A_i). \tag{4}$$

Table 4 shows the percent of measurements and corresponding average actual flow rate in each order-of-magnitude bin based on the direct measurement. The impact of the precision of the measurement system can be seen here; the number of leaks detected as large leaks (> 10 ft^3/h) is 5% compared to 2% of actual large leaks. Correspondingly, the average flow rate of leaks detected as large leaks is 10.0 ft^3/h compared to 25.2 ft^3/h if each leak could be perfectly classified in its respective bin.

3.3. Estimating total emissions

The emissions associated with each bin is given by:

$$\text{Emissions}(B_j) = N \cdot P\langle B_j\rangle \cdot EF(B_j), \tag{5}$$

where N is the total number of leaks found and $N \cdot P\langle B_j\rangle$ is the number of leaks measured in the bin B_j.

The total emission of the network may then be calculated as:

$$\text{Emissions} = \sum_j \text{Emissions}(B_j) = N \cdot \sum_j P\langle B_j\rangle \cdot EF(B_j). \tag{6}$$

When replacing $EF(B_j)$ from Equation (4) into Equation (6), we obtain:

$$\text{Emissions} = N \cdot \sum_j P\langle B_j\rangle \cdot \left[\sum_i P\langle A_i|B_j\rangle \cdot EF(A_i)\right], \tag{7}$$

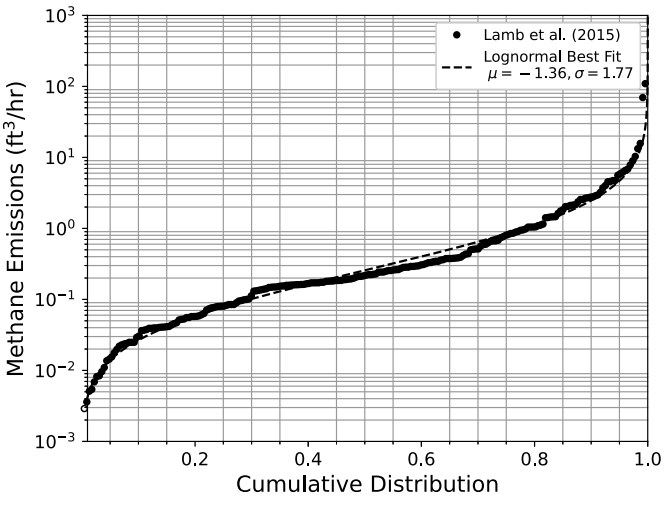

Fig. 6. Experimental data from WSU fit to a lognormal distribution.

Table 1
Fraction of leaks and average emissions in four order-of-magnitude bins based on a distribution of leak sizes fit to the WSU dataset.

Leak Bin	Leak Size (ft^3/h)	% Leaks $P\langle A_i\rangle$	Average Flow Rate in Bin (ft^3/h) $EF(A_i)$
A_1	[10, ∞[	2%	25.2
A_0	[1, 10[	20%	2.8
A_{-1}	[0.1, 1[	48%	0.4
A_{-2}	[0, 0.1[	30%	0.04

Table 2
Probability a leak of size A_i is measured as size B_j.

Actual Leak Size (ft^3/hr)						
	A_1	[10, ∞[	0%	0%	16%	84%
	A_0	[1, 10[	0%	17%	66%	17%
	A_{-1}	[0.1, 1[	17%	66%	17%	0%
	A_{-2}	[0, 0.1[	84%	16%	0%	0%
			[0, 0.1[B_{-2}	[0.1, 1[B_{-1}	[1, 10[B_0	[10, ∞[B_1
			Measured Leak Size (ft^3/hr)			

then replacing $P\langle A_i|B_j\rangle$ from Equation (2) into Equation (7), we obtain:

$$\text{Emissions} = N \cdot \sum_j \sum_i P\langle B_j|A_i\rangle \cdot P\langle A_i\rangle \cdot EF(A_i), \tag{8}$$

with $\forall i, \sum_j P\langle B_j|A_i\rangle = 1$. Then:

$$\text{Emissions} = N \cdot \sum_i P\langle A_i\rangle \cdot EF(A_i) = N \cdot \sum_j P\langle B_j\rangle \cdot EF(B_j). \tag{9}$$

We observe from Equation (9) that the total emissions estimated through direct measurement are equal to the actual emissions independently of the precision of the measurement if the MDL is low enough to cover the full range of emissions and if the prior (estimated distribution of leak size) and uncertainties are correctly considered. On the other hand, ignoring the impact of uncertainty on predicted emissions would lead to a very different result. For the example presented here, the estimated emissions would be 60% greater than the actual emissions as shown in the example presented below. Finally, the four-bin approach implemented for the purpose of capturing methane abatement related to the early detection and repair of large leaks can be expanded to any number of bins towards a continuous approach as presented below using Monte Carlo simulations.

3.4. Monte Carlo simulations

The simulation process starts by sampling events from the prior distribution, modeled from a lognormal fit to the WSU data with parameters $\mu = -1.36$ and $\sigma = 1.77$. Each sample was then converted to a mobile system measurement by multiplying by a sample drawn from its precision as established earlier and modeled as a lognormal distribution with parameters $\mu = 0$ and $\sigma = 0.95$. This process leads to the simulation of the mobile system measurement dataset for leaks following the prior leak size distribution. Fig. 7 shows the result of 10^5 simulated leaks and the resulting distribution of actual leak rates within each of four order-of-magnitude measurement bins. The 1-to-1 line provides a visual cue to help interpret the results. The higher point density on the left side of the 1-to-1 line, especially for actual leak sizes above 1 ft^3/h, illustrates the higher probability for the system to overestimate the leak than underestimate it. The simulation confirms the Bayesian method described in the previous sections when the results are binned according to the measured leak size, $[10^j, 10^{j+1}[$.

This approach may be extended to include any number of bins provided a statistically significant number samples are generated in each bin. The result may then be represented as a continuous function which may be used to evaluate the most likely flow rate and uncertainty range based on any measurement provided by the mobile system. Fig. 8 shows a continuous function, which is a power law-fit to the simulation result separated into 50 log-uniform bins from 10^{-2} to 10^3 ft^3/h.

Table 3
The probability a leak measured as size B_i is actually size A_j for a leak size distribution that follows WSU.

Measured Leak Size (ft^3/hr)						
	B_1	[10, ∞[	0%	0%	68.0%	32.0%
	B_0	[1, 10[	0%	37.5%	61.1%	1.4%
	B_{-1}	[0.1, 1[	11.9%	79.5%	8.6%	0%
	B_{-2}	[0, 0.1[	75.3%	24.7%	0%	0%
			[0, 0.1[A_{-2}	[0.1, 1[A_{-1}	[1, 10[A_0	[10, ∞[A_1
			Actual Leak Size (ft^3/hr)			

Table 4
Fraction of measurements and average emissions in four order-of-magnitude bins based on a distribution of leak sizes fit to WSU.

Measured Leak Bin	Leak Size (ft^3/h)	% Measurements $P\langle B_j\rangle$	Average Flow Rate in Bin (ft^3/h) $EF(B_j)$
B_1	[10, ∞[	5%	10.0
B_0	[1, 10[	22%	2.2
B_{-1}	[0.1, 1[	40%	0.5
B_{-2}	[0, 0.1[	33%	0.09

3.5. Estimating emissions from unknown leaks

In general, the mobile system will report more indications than there are leaks in the network. Although these indications usually represent real sources of methane, they may not be sources of interest for reporting of emissions - e.g. 3rd-party sources, natural/biogenic sources, or natural-gas vehicles. Furthermore, at PG&E, only below-ground (BG) leaks are of interest to the program as meter-set assembly (MSA) emissions are currently reported separately and characterized through static emission factors.

A model for the probability that an indication was generated from a below-ground leak was developed and validated using a set of leaks that were confirmed through field investigation during the routine leak survey process. The model implemented a decision tree algorithm to determine a below-ground probability index based on properties of the detections such as methane and ethane concentration enhancement, calculated emissions, spatial profile of the concentration signal, and number of detections. Using this relationship, the total number of below-ground leaks in the network may be estimated and the reported emissions may be adjusted accordingly. Each detection is assigned a probability to be related to a below-ground leak $P\langle BG\rangle$ and its contribution is calculated as a function of its measured flow rate as:

$$\text{Flow}(BG) = P\langle BG\rangle \cdot AdjFlow[\text{Measured Flow}], \tag{10}$$

and the flow rate of all open below-ground leaks as detected with the vehicle is:

$$\text{TotalFlow}(BG) = \sum_{BG} \text{Flow}(BG) \tag{11}$$

Large gas distribution networks require a significant amount of time to be fully driven by the mobile system. At PG&E the 42,000 miles of mains and 3.6 million services are covered every year from January to December. These data do not provide a picture of the emissions at a point of time but rather a progressive scanning over the year. Detected leaks may have been open since the beginning of the year or for a shorter period. On the other hand, leaks may occur after the survey and produce

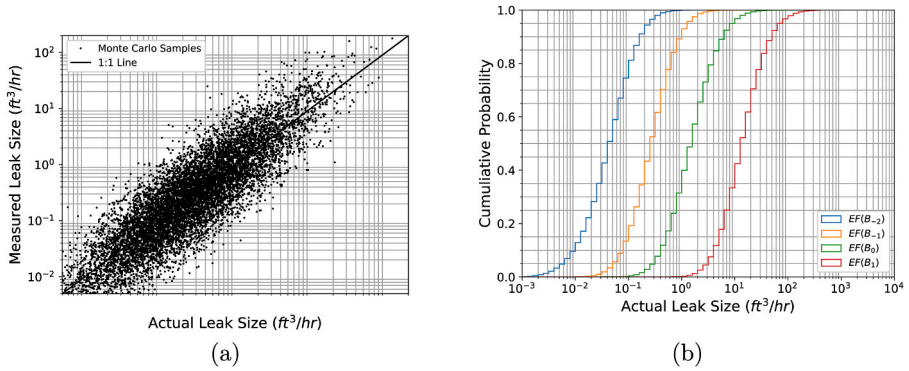

Fig. 7. (a) The result of a Monte Carlo simulation of 10^5 leaks in a gas distribution network. (b) The simulation result represented as a cumulative distribution of the actual leak sizes split into four order-of-magnitude bins based on the measured value.

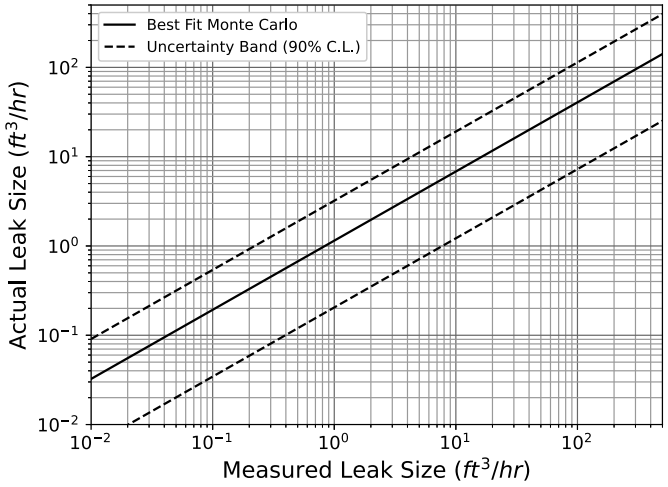

Fig. 8. A relationship between actual leak size and measured leak size for leaks in a gas distribution network. This relationship is called *AdjFlow*.

methane that must be accounted for. If the leaks appear linearly with time and the survey is performed equally along the year, methane emissions from leaks that opened after the measurement equals the overestimate of emission assigned to detected leaks when considering that they are open since the beginning of the year. Therefore, the annual emission of the system can be calculated from the number of leaks detected through the surveys as:

$$Emissions = \sum_{BG} \text{Flow}(BG) \cdot [min(\text{end of year, time of repair}) - \text{beginning of year}]. \quad (12)$$

We have shown therefore that the methane emissions of a gas network can be obtained directly by using the flow rate estimates measured by the survey vehicles without the assumption of any emission factors. In addition, the measurement-based approach assigns a flow rate to every leak detected by the vehicle survey. This unlocks an opportunity for gas operators to prioritize the repair of larger leaks and reduce methane emissions in an effective manner by leveraging the broad range and the skewed distribution of leak sizes highlighted in Section 1.

4. Results

PG&E in its Super Emitter Program uses the leak size estimates of the mobile methane detection system to identify large leaks in areas not scheduled for leak survey. Each leak that was measured by the mobile system as greater than 10 ft^3/h is assigned with the emission factor $EF(B_1) = 10.0$ ft^3/h and leaks measured by the system as less than 10 ft^3/h are assigned with the emission factor $EF(\overline{B_1}) = 0.73$ ft^3/h. The large leaks are repaired in priority independently of their grade. Smaller leaks are repaired in application of the safety standard of the company.

Table 5 illustrates an example of the implementation of such a program and the result of a theoretical case for a network that follows a WSU leak size distribution where 1,000 leaks are found during a 5-year survey (20% of the territory is surveyed every year). It assumes that all leaks are repaired immediately after they are identified. Leaks that are found through survey are therefore assigned with an average lifetime of six months and leaks on non-surveyed areas are assigned with a lifetime of the full year.

In this example the first year is considered as a baseline. The total number of leaks in the non-surveyed areas are estimated by assuming a linear leak appearance over five years. The total emissions are estimated using 1.23 ft^3/h from the average of the WSU leak size distribution. In the first year of a Super-Emitter Program, the non-leak survey areas are driven with the mobile system and the large leaks prioritized for immediate repair. This process results in a 16% emissions reduction compared to the baseline with only 10% increase in the number of repaired leaks. For the second year and later, the number of large leaks found is reduced because the annual detection leaves less time for these leaks to develop. The total emissions reduced is then 39% compared to the baseline with no additional repairs since the program accelerated the detection and repair of large leaks. The large leaks would eventually been detected through routine survey but would have stayed open for a longer time. Additional emissions reduction in the third year and beyond may be realized by a combination of lowering the threshold defining large leak or increasing the measurement frequency in order to reduce the time these large leaks stay undetected.

5. Discussion

The method presented in this article allows for the most probable estimate of all of the leaks for a given prior leak size distribution, independent of the precision of the measurement method. In practice however, a specific network may have an actual distribution of leak size that is similar or different compared to literature. If network-specific measurements are available (ex. a utility that uses mobile data for routine leak survey), the actual distribution may be estimated from the measurements themselves. The actual leak size distribution $P\langle A_i\rangle$ can be adjusted in such a way that $P\langle B_j\rangle$ coincides with the measured leak size distribution. In addition to having a model of the actual distribution that is specific to the network, this approach also offers the convenience to modify the distribution over time to reflect changes in the network owing to emissions abatement efforts. However, it must be noted that the correction of the measured distribution, $P\langle B_j\rangle$, for both measurement uncertainties and for attribution of indications to below-ground leaks is

Table 5
Example emissions reduction scenario by implementing an annual Super-Emitter program in areas not scheduled for leak survey.

		Leak Survey Areas (20% of the network)	Non-Leak Survey Areas (80% of the network)	Total	Change from Baseline
Baseline	Total Number of Leaks	1,000	2,000	3,000	
	Total Number of Repaired Leaks	1,000	0	1,000	
	Annual Emissions (MMcf/y)	5.4	21.5	26.9	
Year 1	Number of Leaks Detected ≥10 ft^3/h	50	100	150	
	Number of Leaks Detected <10 ft^3/h	950	1,900	2,850	
	Total Number of Leaks	1,000	2,000	3,000	
	Number of Repaired Leaks	1,000	100	1,100	+10%
	Annual Emissions (MMcf/y)	5.4	17.2	22.6	−16%
Year 2	Number of Leaks Detected ≥10 ft^3/h	35	78	113	
	Number of Leaks Detected <10 ft^3/h	925	1,862	2,786	
	Total Number of Leaks	960	1,939	2,899	
	Number of Repaired Leaks	960	78	1,037	+4%
	Annual Emissions (MMcf/y)	3.7	13.9	17.5	−35%
Year >2	Number of Leaks Detected ≥10 ft^3/h	28	73	101	
	Number of Leaks Detected < ft^3/hr	899	1,843	2,742	
	Total Number of Leaks	927	1,916	2,843	
	Number of Repaired Leaks	925	73	1000	0%
	Annual Emissions (MMcf/y)	3.1	13.2	16.3	−39%

challenging because of the field validation dataset that it required and the small signal to noise ratio (i.e. below-ground leaks to MSA leaks and other false positives), especially for indications with small concentration enhancement (< 100 ppb).

With their FEAST model Kemp, Ravikumar, and Brandt (Kemp et al., 2016) have argued that it can be more cost effective to accelerate surveys with low-sensitivity tools detecting only large leaks than it is to perform extensive surveys aiming at detecting all leaks such as inspections performed for safety. A key limitation of their approach, however, was the assumption that a low-sensitivity tool would only detect large leaks while, in reality, large uncertainties affect detection and quantification. Low-sensitivity tools may characterize a smaller leak as large or miss a large leak. Uncertainties would therefore substantially impact the effectiveness of their use for an accelerated repair program. The method presented here using an ultra-low MDL system, including a rigorous accounting for uncertainties, circumvents this limitation and supports the use of fast and sensitive detection systems for the estimate of total methane emissions and prioritization of repairs.

CRediT authorship contribution statement

Sean MacMullin: Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **François-Xavier Rongère:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sean MacMullin is an employee of Picarro Inc and François-Xavier Rongére was an employee of Pacific Gas and Electric Company during the development and writing of this manuscript. François-Xavier Rongére is currently an employee of Picarro Inc.

Data availability

Data will be made available on request.

References

Agrawal, K., Garg, S., Sharma, S., Patel, P., 2016. Development and validation of optics based spatio-temporal clustering technique. Inf. Sci. 369, 388–401. https://doi.org/10.1016/j.ins.2016.06.048.

Ankerst, M., Breunig, M., Kriegel, H., 1999. Optics: ordering points to identify the clustering structure. ACM Sigmod record 28, 49–60.

Birant, D., Kut, A., 2007. St-dbscan: an algorithm for clustering spatial–temporal data. Data Knowl. Eng. 60, 208–221. https://doi.org/10.1016/j.datak.2006.01.013.

Brandt, A.R., Heath, G.A., Cooley, D., 2016. Methane leaks from natural gas systems follow extreme distributions. Environ. Sci. Technol. 50, 12512–12520. https://doi.org/10.1021/acs.est.6b04303.

Campello, R., Moulavi, D., Sander, J., 2013. Density-based clustering based on hierarchical density estimates. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 160–172.

Collins, W., Orbach, R., Bailey, M., Biraud, S., Coddington, I., DiCarlo, D., Peischl, J., Radhakrishnan, A., Schimel, D., 2022. Monitoring methane emissions from oil and gas operations. Opt Express 30, 24326–24351. https://doi.org/10.1364/OE.464421.

Conley, S., Faloona, I., Mehrotra, S., Suard, M., Lenschow, D.H., Sweeney, C., Herndon, S., Schwietzke, S., Pétron, G., Pifer, J., Kort, E.A., Schnell, R., 2017. Application of gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases. Atmos. Meas. Tech. 10, 3345–3358. https://doi.org/10.5194/amt-10-3345-2017.

Conrad, B., Tyner, D., Johnson, M., 2022. Robust Probabilities of Detection and Quantification Uncertainty for Aerial Methane Detection: Examples for Three Airborne Technologies. https://doi.org/10.31223/X5S05F.

Crosson, E.R., 2008. A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon dioxide, and water vapor. Appl. Phys. B 92, 403–408. https://doi.org/10.1007/s00340-008-3135-y.

Duren, R., Thorpe, A., Foster, K., et al., 2019. California's methane super-emitters. Nature 575, 180–184. https://doi.org/10.1038/s41586-019-1720-3.

D'Zurko, D., Mallia, J., 2017. Nysearch Methane Emissions Technology Evaluation & Test Program, EPA 2017 Natural Gas STAR and Methane Challenge Annual Implementation Workshop, pp. 24–26.

Ester, M., Kriegel, H., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd 96, 226–231.

Harrison, M., et al., 1996. Methane Emissions from the Natural Gas Industry, vol. 2. Environmental Protection Agency. Technical Report EPA-600/R-96- 080b, Technical Report.

Kemp, C.E., Ravikumar, A.P., Brandt, A.R., 2016. Comparing natural gas leakage detection technologies using an open-source "virtual gas field" simulator. Environ. Sci. Technol. 50, 4546–4553. https://doi.org/10.1021/acs.est.5b06068.

Lamb, B.K., Edburg, S.L., Ferrara, T.W., Howard, T., Harrison, M.R., Kolb, C.E., Townsend-Small, A., Dyck, W., Possolo, A., Whetstone, J.R., 2015. Direct measurements show decreasing methane emissions from natural gas local distribution systems in the United States. Environ. Sci. Technol. 49, 5161–5169. https://doi.org/10.1021/es505116p.

PHMSA, 2019. Nysearch Report – Methane Emissions Quantification Process. URL: https://primis.phmsa.dot.gov/matrix/FilGet.rdm?fil=13229 https://primis.phmsa.dot.gov/matrix/FilGet.rdm?fil=13229.

Redding, S., Glaze, B., 2015. Revolutionizing Leak Management, 26th World Gas Conference. Technical Report, Paris, France.

Zavala-Araiza, D., Lyon, D., Alvarez, R.A., Palacios, V., Harriss, R., Lan, X., Talbot, R., Hamburg, S.P., 2015. Toward a functional definition of methane super-emitters: application to natural gas production sites, 10.1021/acs.est.5b00133. arXiv Environ. Sci. Technol. 49, 8167–8174. https://doi.org/10.1021/acs.est.5b00133. pMID: 26148555.

Zavala-Araiza, D., Alvarez, R.A., Lyon, D.R., Allen, D.T., Marchese, A.J., Zimmerle, D.J., Hamburg, S.P., 2017. Super-emitters in natural gas infrastructure are caused by abnormal process conditions. Nat. Commun. 8, 14012 https://doi.org/10.1038/ncomms14012.