Optimizing continuous monitoring sensor placement on oil and gas sites

Meng Jia,* Troy Sorensen, and Dorit Hammerling

Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO, USA

E-mail: mjia@mines.edu

Abstract

We propose a generic, modular framework to optimize the placement of continuous monitoring sensors on oil and gas sites aiming to maximize the methane emissions detection efficiency. Our proposed framework substantially expands the problem scale compared to previous related studies and can be adapted for different objectives in sensor placement. This optimization framework is comprised of five steps: (1) simulate emission scenarios using site-specific wind and emission information; (2) set possible sensor locations under consideration of the site layout and any site-specific constraints; (3) simulate methane concentrations for each pair of emission scenario and possible sensor location; (4) determine emissions detection based on the site-specific simulated concentrations; and (5) select the best subset of sensor locations, under a given sensor budget, using genetic algorithms combined with Pareto optimization. We demonstrate the practicality and effectiveness of our framework through its application to an oil and gas emission testing facility with a large search space of possible sensor locations; a setting which is computationally infeasible to solve with commonly used mixed-integer linear programming formulations. Additionally, a case study illustrates the success-

ful application of our algorithm to a real oil and gas site, showcasing its real-world applicability and effectiveness.

Keywords: methane emission reduction, oil and gas, continuous monitoring systems, sensor placement, genetic algorithms, Pareto optimization

Synopsis: We develop a sensor placement optimization framework for continuous monitoring systems for methane emissions detection on oil and gas sites.

Introduction

Mitigating anthropogenic methane emissions is pivotal for achieving the goal of limiting increase in the global average temperature to below 1.5°C above pre-industrial levels set by the 2015 Paris Climate Agreement.[1,2] Oil and gas emerges as a key sector for reducing emissions, contributing 22% of the global anthropogenic methane emissions[3,4] and 32% in the United States.[5] In June 2024, the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Energy (DOE) announced $850 million in federal funding aimed at reducing methane emissions from the oil and gas sector.[6] Methane emissions from this sector are characterized by significant temporal fluctuations,[7,8] with sporadic, short-lived events of high emissions—known as super emitter events—accounting for a substantial portion of the total emissions.[9,10] In this context, Continuous Monitoring Systems (CMS), comprising multiple point-in-space sensors that continuously measure methane concentrations, are becoming increasingly vital for emission monitoring. They offer near real-time tracking capabilities essential for identifying both highly fluctuating emissions and transient super emitters.[11–13] CMS plays a crucial role in detecting, localizing, and quantifying site emissions.[14–16]

CMS sensor placement, the strategy of positioning multiple fixed point sensors on an oil and gas site, significantly affects the efficiency of methane emissions detection, especially when considering a limited sensor budget.[17] Sensor placement optimization has been extensively studied in environmental monitoring applications such as air pollutant monitoring,[18,19]

water contamination monitoring,[20–24] temperature monitoring,[25,26] and wind monitoring.[27,28] The goal is to optimize the locations of a limited number of sensors to maximize objectives such as coverage, detection capability, and data quality for subsequent analyses. Until recently, sensor placement optimization for methane emission monitoring on oil and gas sites had not been extensively studied due to CMS being a relatively new technology. Other methane emission monitoring technologies, such as satellite,[29] aerial monitoring systems,[30] and handheld optical gas imaging (OGI) cameras,[31] do not require sensor placement. These systems can either cover entire sites (e.g., satellite and aerial systems) or be moved to various locations on-site (e.g., OGI). In contrast, as a ground-based monitoring system, CMS involves sensors that cannot be easily moved, and a limited number of sensors cannot cover an entire site. Therefore, effective sensor placement optimization methods are essential for maximizing the efficiency of methane emissions detection.

In the realm of methane emission monitoring at oil and gas sites, Klise et al.[32] developed a sensor placement optimization framework designed to enhance the detection capabilities of sensor networks. This framework begins by employing site-specific wind data and emission characteristics to create simulated emission scenarios, thereby closely approximating the actual conditions present at the site. Following this, the framework applies mixed-integer linear programming (MILP) to identify optimal sensor placements. Klise et al.[32] showcased their methodology through a case study within a confined space of 100 x 100 x 10 meters, arranging potential sensor locations at intervals of 10 meters horizontally and 1 meter vertically, culminating in a search space comprising 100,000 potential sensor locations. This complexity is further amplified by introducing three sensor categories with different detection threshold and cost, expanding the search space to 300,000 distinct configurations. Moreover, to refine the simulation of real-world conditions, 1,200 emission scenarios were generated, capturing a range of meteorological conditions and emission patterns. Building upon Klise et al.[32]'s foundational work, Zi et al.[33] introduced a novel approach by incorporating wind condition uncertainties and employing a distributionally robust optimization (DRO) strategy to

bolster the sensor network's detection robustness, particularly in worst-case scenarios. This methodology was evaluated using the same spatial domain as in Klise et al. [32]'s study, demonstrating significant improvements in detection capabilities. Both studies leveraged MILP for optimization, showcasing its efficacy in smaller search spaces. However practical applications such as reducing sensor spatial resolution to 1 meter on oil gas sites or extending the analysis to larger regions reveal the limitations of MILP due to its computational demands, which escalate exponentially with the increase in decision variables — in this context, the potential sensor locations. Although advanced computational strategies like branch-and-bound [34] or branch-and-cut [35] can alleviate some computational burdens, the nature of MILP algorithm means that solving them remains NP-hard in the most challenging scenarios.

In the present study, we endeavor to expand the search space for potential sensor locations and advocate for the utilization of genetic algorithms (GAs) [36] to overcome the computational challenges encountered with MILP. GAs, distinguished by their emulation of natural selection processes, demonstrate high efficiency in navigating extensive and intricate solution spaces. Through the application of evolutionary mechanisms such as selection, crossover, and mutation, GAs are posited to achieve near-optimal solutions with greater rapidity and reduced computational burden compared to conventional MILP methods, particularly in instances where exact optimization proves to be computationally prohibitive. GAs have been widely used in sensor placement optimization across various applications, including structural health monitoring [37] and seismic building monitoring. [38] While GAs do not guarantee finding the exact optimal solution, their high efficiency in finding a sub-optimal solution that is very close to the optimum makes them well-suited for applications in methane emissions detection on oil and gas sites, where errors in the modeled emissions make an exact solution superfluous. Within this context, we conceptualize the sensor placement as a multi-objective optimization problem guided by the principles of Pareto optimization [39] which entails the simultaneous optimization of multiple conflicting objectives. Specifically, our objective is twofold: first, to identify a subset of sensor locations that optimizes detection capabilities, and second, to

ensure that the number of selected sensors remains within a predetermined sensor budget. To address this dual-objective optimization problem, we employ the method developed by Qian et al.[40] which leverages GAs to facilitate an effective solution search.

Compared to the work of Klise et al.[32] and Zi et al.[33], we use a much higher spatial resolution to define possible sensor locations resulting in finding a sensor placement with higher detection efficiency. Also, we utilize all available wind data for a site with a resolution of one minute to achieve a more precise approximation of the actual meteorological conditions at the site. Moreover, we simulate methane emissions with the Gaussian puff model rather than the Gaussian plume model owing to its superior accuracy in capturing the dynamics of atmospheric dispersion by allowing for variable wind conditions. As a result of the aforementioned enhancements, our sensor placement optimization algorithm delivers solutions of increased accuracy and granularity, thereby offering substantial benefits for practical applications. As such our algorithm offers an effective solution for optimal sensor placement which can serve as an open-source reference tool for CMS technology vendors and advance the quest towards rapid elimination of methane emissions on oil and gas sites.

Methods

In this study, we introduce a data-driven algorithm designed to optimize the placement of CMS sensors. The algorithm is modular and a visual summary of the algorithm is provided in Section S1 in the Supporting Information (SI) document. The algorithm requires three critical inputs: site-specific wind patterns, emission characteristics, and a predefined sensor budget which dictates the number of sensors that can be deployed. Utilizing these inputs, our algorithm calculates the optimal sensor arrangement within the budget, aiming to maximize emission detection efficiency, defined as the frequency or likelihood with which at least one sensor in the system detects an emission. Detection efficiency, the objective to be optimized in the sensor placement problem, can also be defined using other metrics such as time to

first detection or detection fraction based on total time or amount of emissions. Due to the flexibility of GAs, the objective function can be adjusted accordingly without altering the entire framework. The workflow of the algorithm is structured into five distinct steps:

1. Generation of numerous emission scenarios that reflect the site's actual emission characteristics and wind conditions.

2. Specification of possible sensor locations based on site geometry and operational guidelines if any.

3. Simulation of methane concentrations at all possible sensor locations for each emission scenario.

4. Evaluation of detection for each sensor location and emission scenario pair.

5. Optimization of sensor placements that ensure maximum detection across all scenarios under a given budget.

In the following subsections, we first outline the experimental data utilized to showcase our algorithm. We then detail each step by first providing a general description and then explaining the experimental approach used in this study. It is important to reiterate that our algorithm is highly modular and that the components demonstrated in our experimental study, such as the wind and emission distribution estimators and the atmospheric transport model, can be readily replaced.

Data description

To validate the effectiveness of our proposed algorithm, we employ wind data acquired from the 2022 Advancing Development of Emissions Detection (ADED) campaign[41] conducted at the Methane Emissions Technology Evaluation Center (METEC), which was operational from January 28 to May 16, 2022. Due to data availability and for demonstration purposes, we used a period of less than four months of wind data. In practice, a full year of wind

data is recommended to accurately represent the actual meteorological conditions, including seasonal variations, at the site. This is implemented in the case study of a real oil and gas site as detailed in the final section of this manuscript. The configuration of the METEC site is illustrated in Figure 1(a), where five distinct sets of equipment are outlined, each representing a potential emission source. Wind speed and direction were continuously recorded by three on-site anemometers throughout this period, with data captured at one-minute intervals. For the purposes of our analysis, we assumed wind conditions to be homogeneous across the site, thereby aggregating the data from the three anemometers to compute mean wind speed and direction. Note that the calculation of wind direction used the circular mean, appropriately addressing the circular nature of directional data. Figure 1(b) shows the wind rose that describes the wind distribution throughout the period.

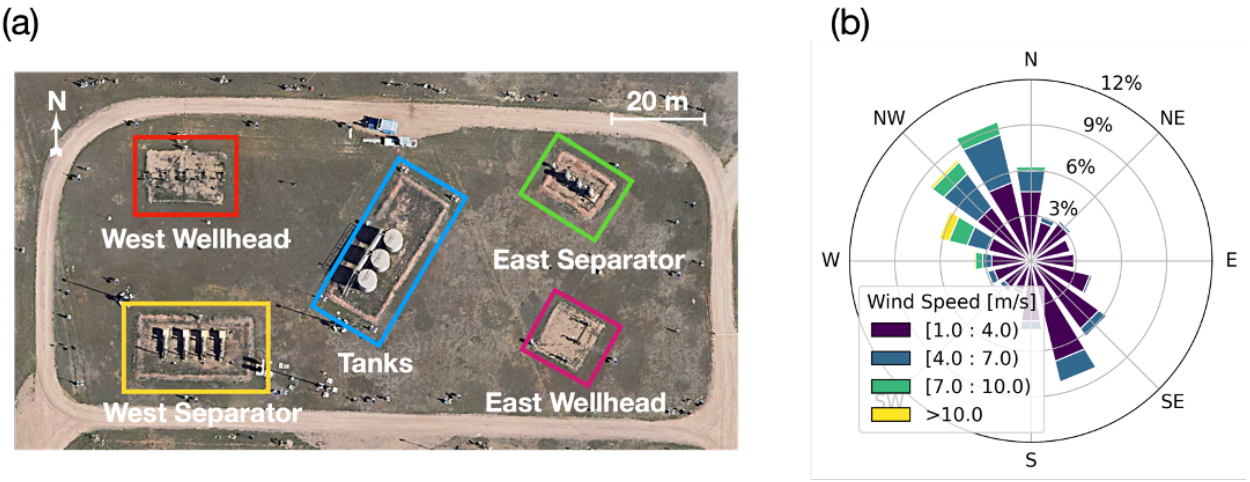

Figure 1: Experimental setup used to demonstrate the sensor placement optimization algorithm. (a) Satellite imagery of the Methane Emissions Technology Evaluation Center (METEC) site. Potential emission source locations are marked with colored boxes. (b) Distribution of wind data during the Advancing Development of Emissions Detection (ADED) experiment period from January 28 to May 16, 2022.

Step 1. Generation of emission scenarios

In accordance with the definition in Klise et al.[32], an emission scenario is determined by a combination of wind conditions and emission characteristics during a period of time. To

generate emission scenarios, two probability distribution functions are constructed: a joint distribution f_e of emission source attributes including locations $\mathbf{x}_0$, durations τ, and rate q, and a joint distribution f_w of wind speed $\mathrm{ws}(t)$ and wind direction $\mathrm{wd}(t), t \in [0, \tau]$. The wind distribution can be inferred from historical wind data recorded on or in the vicinity of the site. The emission distribution can be obtained from operational logs and/or expertise from operators. Practically, we can decompose the joint distribution f_e using conditional distributions, i.e., $f_e(\mathbf{x}_0, \tau, q) = f(\tau, q|\mathbf{x}_0) \times f(\mathbf{x}_0)$. That is, each potential source, based on its functionality, may have its own emission pattern. The two distributions f_w and f_e are deemed independent, which we believe to be a reasonable assumption since the wind conditions should not affect the source attributes of an emission event and vice versa. Nevertheless, it is acknowledged that correlations may arise via confounding temporal factors. For instance, a certain type of emission may tend to occur at a specific hour of the day or in a specific season when the wind might have a specific pattern. In cases where prior knowledge pertaining to such relationships is available, it can be readily incorporated into the distributions facilitating a more nuanced modeling approach. Then, emission scenarios, which are defined by a tuple of $(\mathbf{x}_0, q, \tau, \mathrm{ws(t)}, \mathrm{wd(t)}, t \in [0, \tau])$, are generated by randomly sampling from f_e and f_w.

In this case study, we consider only single-source emissions because our goal is to maximize detection efficiency. Thus, achieving successful detection of single-source emissions should inherently guarantee success in detecting multiple-source emissions. For demonstration purposes, we fix the emission duration for each emission event to be one hour and assume all five equipment have equal likelihood to be an emission source. Additionally, we prescribe three possible emission rates, $[1, 5, 10]$ kg/h to each of the potential emission sources with equal probability. This constructs a simple uniform distribution, i.e., $P(\mathbf{x}_{0,i}, q_j) = 1/15, i = 1, 2, \ldots, 5, j = 1, 2, 3$. It's important to mention that the fixed duration value, emission rates and probabilities we used in this experiment are only for demonstration purpose. In practice, a realistic distribution should be derived based on actual site-specific data and practitioners' insights, and an example of this guidance is provided

in the case study at the end of this paper. For wind data, we approximate the true wind distribution f_w by using one-hour non-overlapping segments from the historical wind data throughout the entire experiment period to reflect the actual meteorological conditions on the site within the experiment period. In practice, a full year of wind data is desired to thoroughly represent the actual meteorological conditions, including seasonal patterns, which is implemented in the case study for a real oil and gas site. Finally, this yields a total number of 38,685 emission scenarios which comprehensively represent the actual emission conditions on the site during the experiment period.

Step 2. Specification of possible sensor locations

The determination of possible sensor locations is contingent upon the desired spatial search grid resolution as well as the geometry of the site. Specifically, areas occupied by equipment or driveways must be excluded from the search space. Furthermore, regions restricted by operational guidelines also necessitate exclusion. In summary, information regarding the search grid resolution sought by the site operator and any pertinent restrictions is essential for the specification of possible sensor locations.

In this case study, we first set possible sensor locations by creating a 3D grid over the entire site. The resolution for the two horizontal directions (Northing and Easting) is 1 m while the resolution for the vertical direction is 0.5 m. The dimensions of the METEC site are approximately 125 meters in Easting and 75 meters in Northing. Considering practical feasibility, the vertical range was designated as spanning from 1 to 10 meters. Subsequently, we excluded all points within the grid that were occluded by any of the five equipment groups present on the site. This exclusion process resulted in a total of 21,774 viable sensor locations, forming the search space for the subsequent optimization analysis. Figure 2 shows the possible sensor locations as well as the potential emission sources on the METEC site. The gap areas correspond to spaces obstructed by equipment and hence devoid of possible sensor locations.

Step 3. Simulation of methane concentrations

In this step, we simulate methane concentrations for each emission scenario generated in Step 1 at all potential sensor locations specified in Step 2, using an atmospheric transport model. This step is summarized by Equation 1:

$$c_i^j(t) = g\left(\mathbf{x}_{0,i}, q_i, \mathrm{ws}_i(t), \mathrm{wd}_i(t); \mathbf{x}_j\right), t \in [0, \tau], i = 1, 2, \ldots, M; j = 1, 2, \ldots, N. \quad (1)$$

Here, c_i^j represents the simulated methane concentration time series for emission scenario i at sensor location j; M and N denote the total numbers of emission scenarios and potential sensor locations, respectively; g is the atmospheric transport model. We will use these simulated methane concentrations to assess detection capabilities at each sensor location for each emission scenario.

In our experiment, the simulations were conducted using the Gaussian puff model which is more accurate than the Gaussian plume model as it accounts for the wind dynamics within the simulation period (see Jia et al.[42] for details). All the parameters used in the Gaussian puff simulations are listed in Section S2 of the SI. We note that other, e.g. more sophisticated, atmospheric transport models can straightforwardly be integrated within the proposed optimization framework by replacing the Gaussian Puff model. Finally, there are a total of $38,685 \times 21,774 = 842,327,190$ one-hour simulated concentration time series. Memory issues arise when dealing with such large data. To address it, we utilize memory-mapping files provided in Numpy[43] to store the data.

Step 4. Evaluation of detection

In this step, we establish detection of every sensor location and emission scenario pair based on the simulated methane concentrations c_i^j obtained in the previous step. We then combine all the detection results into a single detection matrix which serves as the input to the final optimization problem.

We determine detection by applying two thresholds to the simulated methane concentrations: an amplitude threshold, A, to identify significant high values and a temporal threshold, B, to ensure the persistency of the high readings. Therefore, a successful detection is defined to be elevated concentrations above A ppm for more than $B\%$ of the total time steps in a simulated concentration time series as shown in Equation 2.

$$\|c(t) \geq A\|_1 \geq |c(t)| \cdot B\% \tag{2}$$

where $c(t)$ is the simulated methane concentration time series (or vector); $\|\cdot\|_1$ is the L_1 norm of a vector and $|\cdot|$ is the length of a vector. The value of A will typically be set as a function of the sensor quality, with lower quality sensors requiring a higher threshold to distinguish true spikes from the noisy background which consists of ambient atmospheric concentrations and sensor noise. Since we use simulated concentrations in this step, representing concentration enhancements in the observed data, the detection threshold A should be based on the background variations in the observed data to ensure that the signal corresponding to true emissions stands out. Specifically, A is set to be the difference between the upper bound and the lower bound of the background concentration, ensuring that any concentration enhancement above A is guaranteed not to be drowned out by the background. The value of B dependents on the tolerance for false positives. In practice, sporadic spikes may correspond to sensor malfunctions or error from raw data processing rather than actual emissions. We need a metric to evaluate the persistence of the high values defined using the amplitude threshold A. In this study, we use an amplitude threshold of $A = 0.5$ ppm and $B = 20\%$ for the main analysis. An example of a successful detection is illustrated in Section S3 of the SI. We also analyzed how different sensor types may affect the detection efficiency by using $A = 5.0$ ppm and $B = 20\%$ to generate results corresponding to a lower quality sensor. There are various ways to define detection which can be easily accommodated within our framework, and we included a brief analysis of different choices in Section S4 of the SI.

Next, we synthesize all detection results to construct a detection matrix, denoted as D, whose rows correspond to sensor locations and columns to emission scenarios. The entries of the matrix are binary; zero indicates a non-detection for a specific scenario at that sensor location, and one indicates a detection, see Equation 3.

$$\mathbf{D}_{ij} = \begin{cases} 1 & \text{if emission scenario } j \text{ is detected at sensor location } i, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Step 5. Optimization of sensor placement

In the final step, we find the optimal sensor placement under a given sensor budget. Using the detection matrix D constructed in the previous step, we can model the optimization problem as a best subset selection problem. Recall that the rows of the detection matrix correspond to possible sensor locations and columns to emission scenarios. Given the sensor budget k, i.e, the number of sensors to install, selecting k sensors translates to selecting k rows from the detection matrix. In the resulting sub-matrix $D_{\text{sub}}^{(k)}$, we assess the column-wise sums which indicate the number of sensors that detect the corresponding emission scenario. The number of non-zero column-wise sums is the total number of detected emission scenarios, which is called detection coverage. The objective is to find the optimal selection of k rows from this matrix to maximize the detection coverage.

The next step is to apply Pareto optimization to the best subset selection problem we have formulated. Pareto optimization has been developed to handle situations with conflicting objectives. In our case, we want to find a subset of rows from the detection matrix to maximize the detection coverage, while minimizing the size of the subset, that is, using fewer sensors if possible. The Pareto optimization problem is formulated by following the

set up used in Qian et al. [40].

$$\begin{aligned} &\underset{\mathbf{x}\in\{0,1\}^N}{\operatorname{argmax}} \left(g_1(\mathbf{x}), g_2(\mathbf{x})\right) \\ &g_1(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \|\mathbf{x}\|_1 \leq 2k \\ -\infty & \text{otherwise.} \end{cases} \\ &g_2(\mathbf{x}) = -\|\mathbf{x}\|_1 \end{aligned} \tag{4}$$

Here, N is the number of rows of the detection matrix, $\mathbf{x}$ is a binary vector of length N and $\mathbf{x}_i = 1$ if the ith row of the detection matrix is selected; f is the detection coverage and $\|\mathbf{x}\|_1$ is the L_1 norm of the binary vector which is the number of selected rows. Note that for solutions which unduly exceed the budget (by two times in this setting), the corresponding objective value in g_1 is set to be negative infinity to make the subsequent searching process more efficient.

For large-scale optimization problems, such as those involving matrices with dimension in the order of hundreds of thousands, traditional approaches like exhaustive search or standard MILP algorithms are impractical due to their computational cost. Instead, here we use genetic algorithms, a type of randomized algorithms to efficiently search a solution that is close to the optimal one. By conducting selection, mutation, and combination operations inspired by natural selection, the solutions evolve to an optimal or near optimal state in an efficient way. Specifically, we use the Pareto optimization with recombination for subset selection (PORSS) proposed by Qian et al. [40]. The pseudo-code for the algorithm is presented in Section S5 in the SI and the full code is available at the GitHub repository provided at the end of this paper. The algorithm accepts three primary inputs: a detection matrix $\mathbf{D}$, a sensor budget k, and an objective function g_1 which is the detection coverage in this case. The iterative nature of the algorithm is governed by the parameter I, specifying the total number of iterations to execute. The desired output is an optimized subset comprising no more than k rows from the detection matrix $\mathbf{D}$ that maximizes the objective function g_1.

Given the monotonic property of g_1, which ensures that the objective value of the original set is always greater than or equal to that of any of its subsets, the output will consist precisely of k rows.

We start with a null vector $\mathbf{x}$ of length N, the total number of possible sensor locations, initializing a solution population set P with this vector. The iteration counter i is set to zero to track the progression of the algorithm.

The core of the algorithm is an iterative loop that continues until either the iteration counter i reaches the predefined limit I or the prescribed early stop condition is satisfied. Within each iteration, the following steps are undertaken:

1. Selection: A pair of vectors, $\mathbf{x}, \mathbf{y}$, are randomly chosen from the solution population set P.

2. Recombination: The selected vectors undergo a recombination process to produce offspring vectors $\mathbf{x}', \mathbf{y}'$. Specifically, a cut point is randomly selected to split each of $\mathbf{x}, \mathbf{y}$ into two segments: $(\mathbf{x}_1, \mathbf{x}_2), (\mathbf{y}_1, \mathbf{y}_2)$. Then, $\mathbf{x}_1$ is combined with $\mathbf{y}_2$, and $\mathbf{y}_1$ is combined with $\mathbf{x}_2$ to produce the offspring vectors $\mathbf{x}'$ and $\mathbf{y}'$.

3. Mutation: A bit-wise mutation is applied to the offspring $\mathbf{x}'$ and $\mathbf{y}'$, resulting in new vectors $\mathbf{x}'', \mathbf{y}''$. Specifically, a subset of points is randomly selected, and the values at these points are swapped between 0 and 1 to produce the new vectors $\mathbf{x}'', \mathbf{y}''$.

4. Survival of the Fittest: For each of $\mathbf{x}''$ and $\mathbf{y}''$, the algorithm assesses whether it represents an improvement over the existing solutions in P. Only vectors not dominated by any member of P are retained. Here, we define a solution $\mathbf{x}$ dominating another solution $\mathbf{x}'$, i.e, $\mathbf{x} \succ \mathbf{x}'$ if $g_1(\mathbf{x}) \geq g_1(\mathbf{x}')$ and $g_2(\mathbf{x}) \geq g_2(\mathbf{x}')$ **and** either $g_1(\mathbf{x}) > g_1(\mathbf{x}')$ or $g_2(\mathbf{x}) > g_2(\mathbf{x}')$.

5. Early Termination: At the end of each iteration, an early stop condition is evaluated to determine if the optimization can be concluded ahead of schedule.

Upon reaching the termination criterion, the algorithm concludes by selecting from P the solution $\mathbf{x}$ that adheres to the budget constraint k and maximizes the objective function g_1.

The proposed algorithm demonstrates a robust capability for optimizing complex subset selection problems. Through the iterative process of selection, recombination, mutation, and survival of the fittest, the algorithm converges towards an optimal or near-optimal solution. The inclusion of an early termination condition enhances computational efficiency, enabling the algorithm to halt upon sufficient convergence, thereby saving computational resources.

To address the inherent stochastic nature of genetic algorithms which can result in diverse outcomes due to their random selection and mutation processes, we have implemented a strategy that involves running multiple independent genetic algorithms in parallel. By conducting these algorithms separately, we can harness the variability in their search patterns, thereby mitigating the risk of converging to local optima that might be mistaken for global solutions. Upon completion of these independent runs, we aggregate their results, systematically comparing and combining the distinct solutions to identify the most optimal outcomes. This approach not only leverages the exploratory strengths of genetic algorithms but also significantly enhances the robustness of the solution set, ensuring that the best possible solutions are identified amidst the randomness.

To evaluate the performance of our algorithm, we generated test data with known ground truth. These test data were designed to present more challenging scenarios than the original problem in this study, and our algorithm successfully identified the true solution in all cases. Detailed descriptions and results of the tests are provided in Section S6 of the SI.

Results

We first present the optimal placement of sensors across the METEC site, considering various sensor budget constraints ranging from 1 to 12. For each budget level, we conducted 20

independent optimization trials and aggregated the results to identify the most effective sensor configuration. Figure 2 showcases exemplary configurations for the optimal placement of four and eight sensors, with additional configurations for varying budgets detailed in Section S7 of the SI. The sensors' horizontal positions are marked with red dots, while their heights are denoted by corresponding numbers. The top row shows the optimal sensor placements by searching from all possible sensor locations specified in Step 2 across the entire site. The detection coverage achieved by these configurations is 0.52 and 0.75 for the four-sensor and eight-sensor systems, respectively, indicating that the systems could detect 52% and 75% of simulated emission scenarios, respectively. Not surprisingly, the arrangement of the best sensor placement is clearly informed by the predominant wind directions from the northwest and southeast (see Figure 1(b)). Additionally, the sensor heights are highly correlated with the heights of the emission sources. In situations where operational guidelines restrict sensor installation exclusively to the perimeter of a site, we also explore optimal fenceline sensor placements. Adhering to the methodology outlined in the Methods section, we limited potential sensor locations to a narrow zone along the site boundary, specifically within a 2-meter buffer area. The bottom row shows the optimal configurations for placing four and eight sensors, as determined by our algorithm. Again, the arrangement of the optimal sensor placements is informed by the predominant wind directions. Specifically, rather than being uniformly distributed along the perimeter, the sensors are shifted towards the northwest and southeast direction, the predominant wind directions over the experiment period. The four sensor locations in the best four-sensor solution are approximately retained in the best eight-sensor solution, but their heights are adjusted to synergize with the newly added sensors, enhancing detection efficiency.

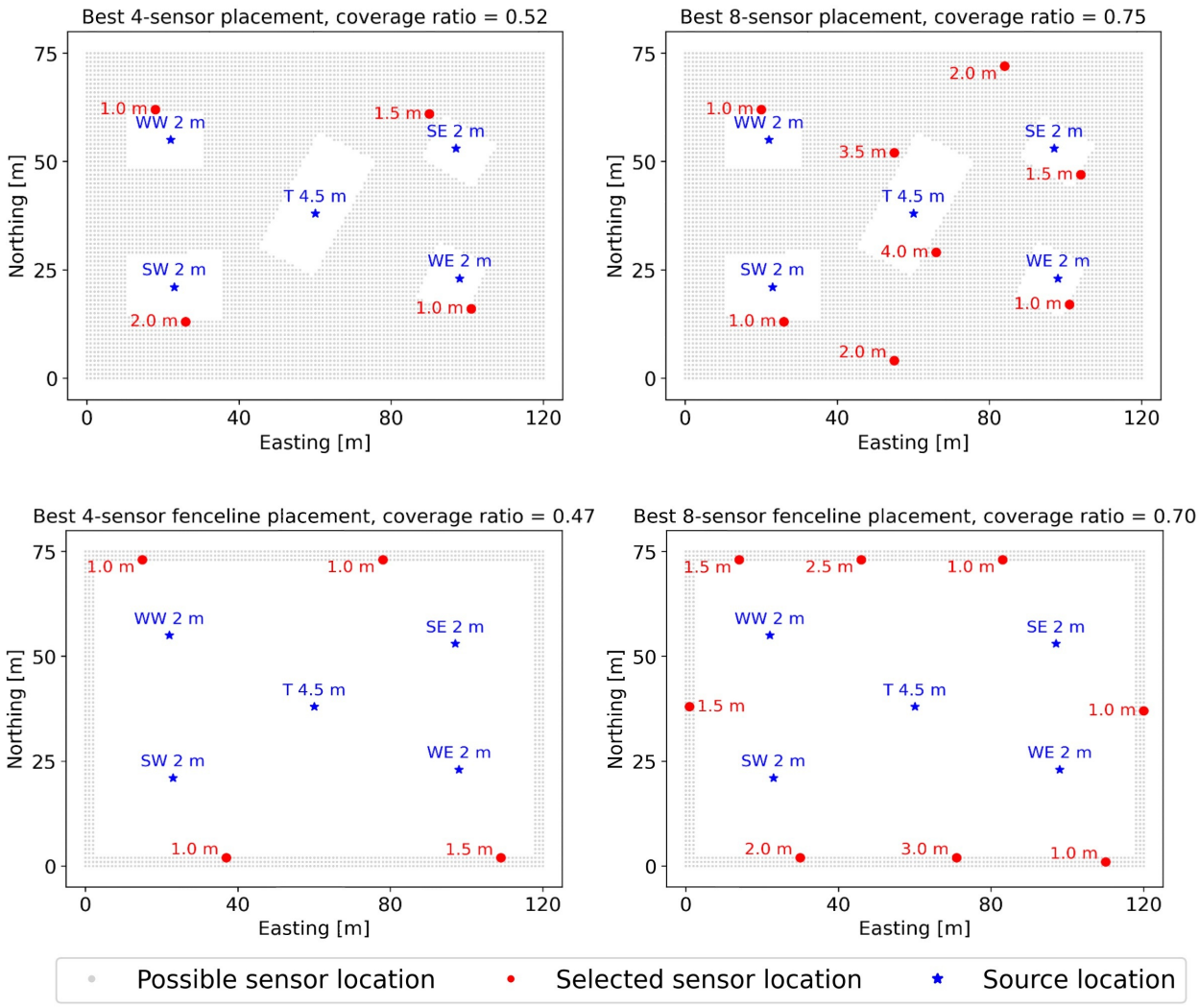

Figure 2: Optimal placements for four and eight sensors as determined by the PORSS algorithm. The top row shows the optimal sensor placements across the entire site, and the bottom row shows the optimal sensor placements constrained along the fenceline of the site. Gray dots indicate possible sensor locations, while blue stars represent potential emission source locations, with their heights indicated by the adjacent numbers. Red dots denote the sensors selected by the algorithm, with their corresponding heights also indicated by the adjacent numbers. The optimal placement of four sensors across the site and along the fenceline achieves detection coverage of 52% and 47%, respectively. The optimal placement of eight sensors across the site and along the fenceline achieves detection coverage of 75% and 70%, respectively.

Figure 3 illustrates the detection coverage ratio, the proportion of detected emission scenarios out of the total emission scenarios generated in Step 1, achieved with the optimal sensor placement for each budget. The blue lines show the coverage ratio associated with the

optimal sensor placements derived by the PORSS algorithm in this setting. The observed trend indicates diminishing returns in terms of detection coverage with increasing sensor budgets. This type of analysis enables us to determine the minimum number of sensors required for a desired detection coverage. Additionally, we compared the results under two different detection amplitude thresholds: 0.5 ppm and 5 ppm representing high and low end sensors, respectively. Low-end sensors generally exhibit larger background variation compared to high-end sensors and hence require higher detection thresholds.[44] The analysis shows that to achieve a specific detection coverage ratio, more low-end sensors are needed. For example, to achieve a detection coverage ratio of 0.50, four high-end sensors or nine low-end sensors are required. In practice, operators can use this analysis to decide on the sensor type and the corresponding number of sensors.

The effectiveness of the PORSS algorithm was further evaluated by benchmarking it against a greedy search method which served as the baseline for comparison. In the greedy search approach, the best current row is iteratively chosen based on a continually updated detection matrix. This matrix evolves by excluding previously selected rows and any columns containing at least a value of one across these rows. As depicted in Figure 3, the solutions derived from the PORSS algorithm represented by the blue lines consistently outperform those obtained through the greedy search method represented by orange lines for both low- and high-end sensors. Notably, for the single-sensor (best-1) configuration, the solution identified by the greedy search is considered the ground truth, which PORSS successfully matched. This outcome underscores the PORSS algorithm's ability to find optimal solutions.

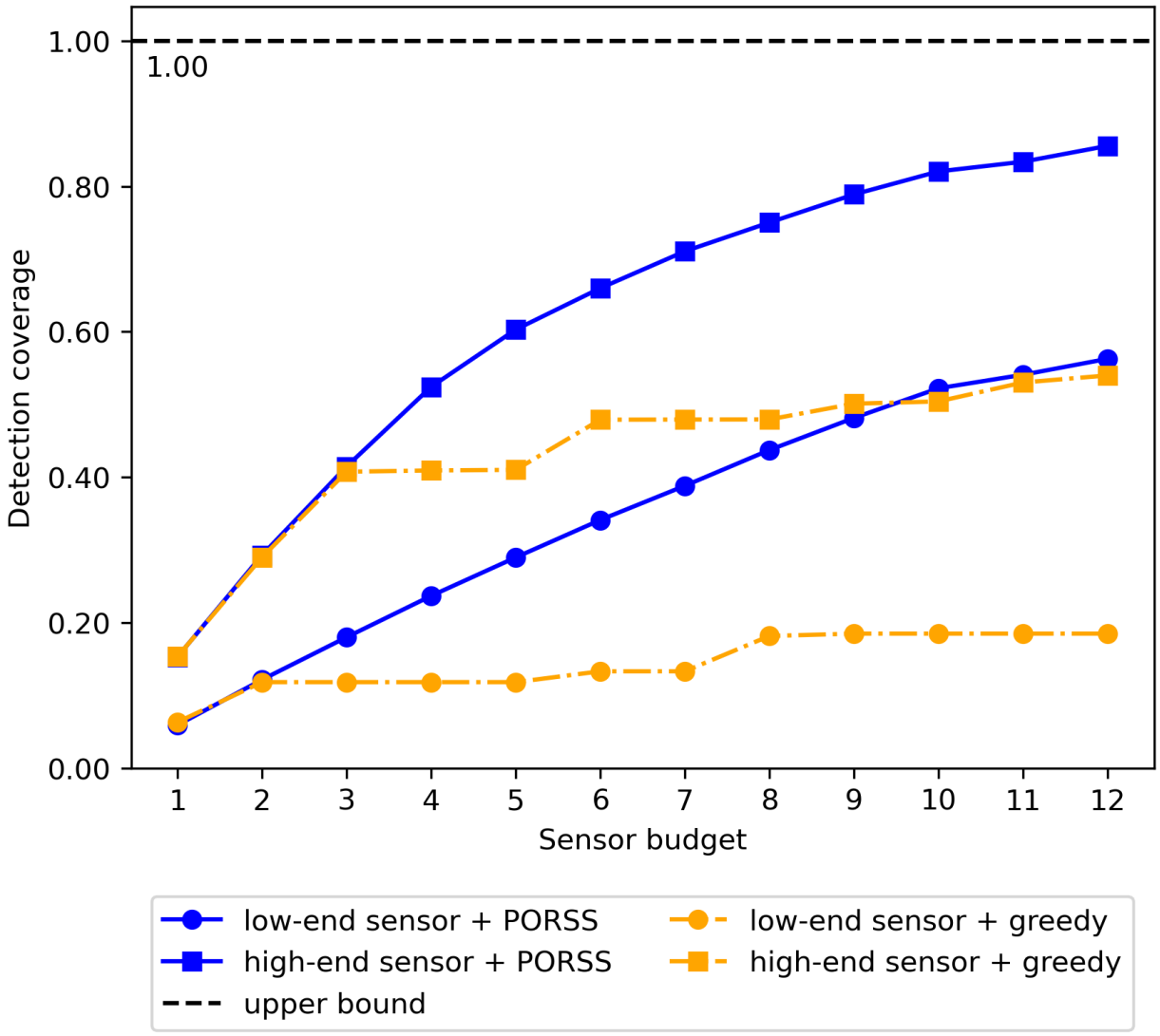

Figure 3: Detection coverage ratios achieved through optimal sensor placement across various sensor budgets. Blue solid lines denote solutions derived from the PORSS algorithm, while orange dash-dot lines represent outcomes from the greedy search method. Circles indicate solutions achieved under a detection amplitude threshold of 5.0 ppm, representing the sensitivity of low-end sensors. Squares denote solutions achieved under a detection amplitude threshold of 0.5 ppm, representing the sensitivity of high-end sensors. The black dashed line illustrates the theoretical maximum coverage ratio, achievable by distributing sensors across all possible locations.

Case Study for a Prototypical Midstream Oil and Gas Site

In this section, we present a case study demonstrating our sensor placement algorithm in a real-world scenario at the request of a site operator. The study illustrates how operator knowledge and site constraints can be incorporated into our optimization scheme for an actual midstream oil and gas site. The site operator and site locations are anonomyzed for confidentiality.

This site initially had nine industry-standard sensors placed on the perimeter fenceline, following the vendors recommendation. The operator wanted to compare this setup to a new configuration using four higher-quality sensors. Given that precise placement becomes more critical with fewer sensors, our framework was used to determine the optimal locations for the new sensors under the operator constraint of restricting placement to the fenceline at a 2-meter height to facilitate installation and minimize operational interference.

Our optimization framework was configured as follows:

Wind Data Wind data was collected over a year at one-minute intervals from two anemometers co-located with the existing sensors. To create a single time series for our framework, we averaged the wind speed and direction (x and y component wise) when both anemometers had data, using the available value when one was NA. Instances with both NA (0.684% of the data) remained NA. We then randomly selected 1,000 non-consecutive hours of gap-free data, with the sampled distribution closely matching the full year's data (see Figure 4).

Emission Scenarios Operator guidance indicated that the 10 site compressors were the primary sources of fugitive emissions, and overflight data had suggested that emissions of 5 kg/hr were typical. We simulated emissions using these values across the 1,000 hourly wind data samples for each compressor, generating 1,000 scenarios per compressor and 10,000 total for all sources combined, allowing us to find the overall optimal placement.

Simulation, Detection, and Optimization All simulations used the Gaussian puff model with parameters listed in Section S2 of the SI. For each simulation, we recorded the candidate locations along the 2 m resolution site fenceline where the simulated methane concentration exceeded 1 ppm for at least 1 minute in accordance with a conservative estimate of sensor sensitivity provided by the sensor vendor. We first determined optimal sensor placements for each compressor individually (see Figure S8 in the SI) then combined the results to create a single detection matrix $\mathbf{D}$ as described in Step 4. The PORSS algorithm (see Step 5) was then used to find the optimal sensor locations. Lastly, we verified our solution's stability by comparing it against 10 million random samples of 4 sensor locations along the fenceline, imposing a 30 m minimum distance between sensors. The PORSS algorithm found the optimal solution in 210 seconds, about 30 times faster than evaluating the random samples which took 6,177 seconds, with nearly identical results but a slightly higher coverage ratio for the PORSS algorithm (see Figure S9 in SI).

The site operators have installed the four new sensors in the recommended locations and are actively recording data as of this paper's publication. While a detailed comparison between these four advanced sensors and the nine existing ones is beyond this paper's scope, this case study illustrates the practical implementation of our framework, integrating operator knowledge effectively.

Conclusion

In this work, we introduced an open-source data-driven algorithm for optimizing sensor placement on oil and gas sites. The demonstration example (METEC experiment) and the real oil and gas site case study outlined in this paper detail the workflow and highlight the efficiency and accuracy of our proposed framework. This algorithm provides site-specific solutions that are more accurate and reliable than traditional sensor placement strategies and

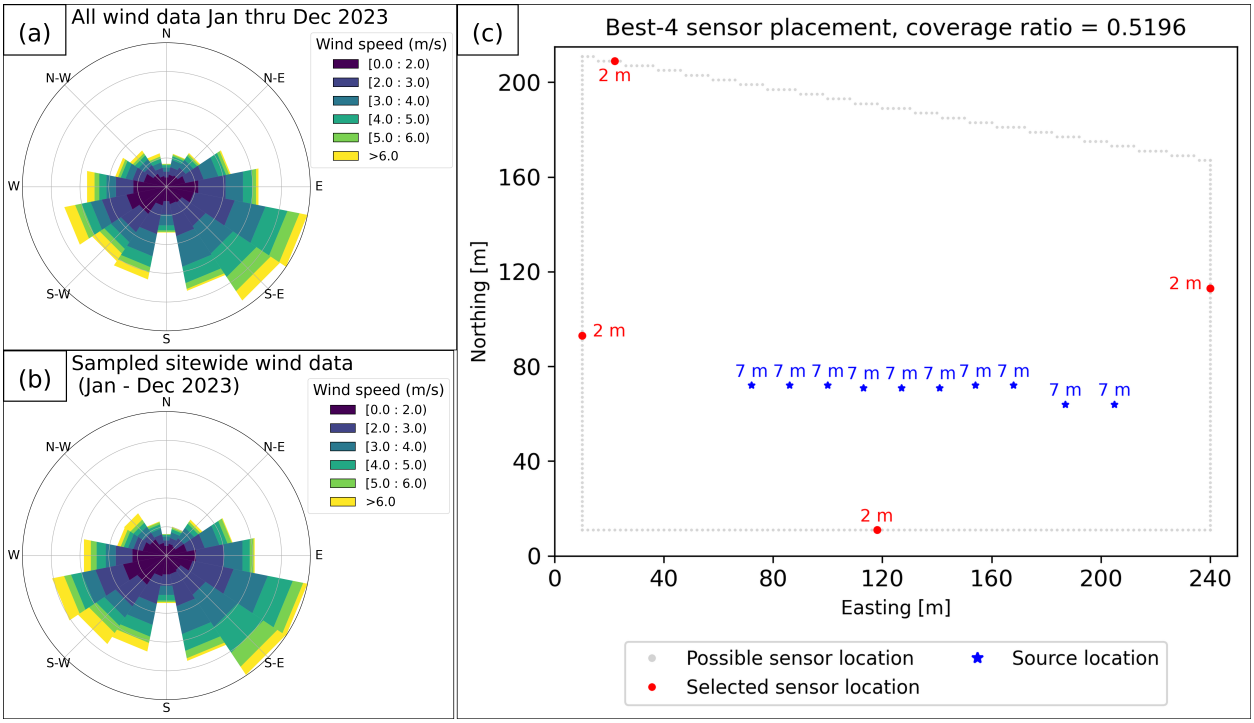

Figure 4: Subplot (a) shows the distribution of wind data recorded over a full year for this site. Subplot (b) shows the distribution of the subset of wind data used within our optimization framework. Subplot (c) shows the optimal fenceline placement for four sensors on this site.

those proposed in prior studies.[32,33] Additionally, our algorithm's analysis of the relationship between sensor budget and detection efficiency is crucial for operators when making decisions during the planning phase.

Our approach significantly expands the problem scale compared to prior research in related studies by efficiently solving large-scale problems often encountered in practice. Further, the high modularity of our algorithm allows users to incorporate their own methods and models, such as wind and emission distribution estimators (in cases with limited or no existing wind data) or different atmospheric transport models. As such, the algorithm holds substantial practical value, particularly as the use of Continuous Monitoring Systems (CMS) becomes increasingly prevalent and critical in the quest to rapidly reduce methane emissions. While this paper focuses on optimizing sensor placement for methane emission detection efficiency, the algorithm can be extended to various other applications, involving emission localization and quantification, by simply changing the objective function in the optimization problem. This flexibility is due to the genetic algorithm's ability to handle non-linear objective functions, a capability not possible with the Mixed-Integer Linear Programming (MILP) approach adopted in prior studies.[32,33]

References

(1) Collins, W. J.; Webber, C. P.; Cox, P. M.; Huntingford, C.; Lowe, J.; Sitch, S.; Chadburn, S. E.; Comyn-Platt, E.; Harper, A. B.; Hayman, G.; Powell, T. Increased importance of methane reduction for a 1.5 degree target. *Environmental Research Letters* **2018**, *13*, 054003.

(2) Schleussner, C. F.; Rogelj, J.; Schaeffer, M.; Lissner, T.; Licker, R.; Fischer, E. M.; Knutti, R.; Levermann, A.; Frieler, K.; Hare, W. Science and policy characteristics of the Paris Agreement temperature goal. *Nature Climate Change 2016 6:9* **2016**, *6*, 827–835.

(3) O'Rourke, P. R.; Smith, S. J.; Mott, A.; Ahsan, H.; McDuffie, E. E.; Crippa, M.; Klimont, Z.; McDonald, B.; Wang, S.; Nicholson, M. B.; Feng, L.; Hoesly, R. M. CEDS v-2021-02-05 Emission Data 1975-2019. `http://doi.org/10.5281/zenodo.4509372`, 2021; Zenodo.

(4) Crippa, M.; Guizzardi, D.; Muntean, M.; Schaaf, E.; Lo Vullo, E.; Solazzo, E.; Monforti-Ferrario, F.; Olivier, J.; Vignati, E. EDGAR v6.0 Greenhouse Gas Emissions. `http://data.europa.eu/89h/97a67d67-c62e-4826-b873-9d972c4f670b`, 2021; European Commission, Joint Research Centre (JRC).

(5) EPA *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2020*; 2022.

(6) US EPA, Office of EPA and DOE EPA, DOE Announce $850 Million to Reduce Methane Pollution from the Oil and Gas Sector. `https://www.epa.gov/newsreleases/epa-doe-announce-850-million-reduce-methane-pollution-oil-and-gas-sector`, Accessed: 2024-07-10.

(7) Allen, D. T.; Cardoso-Saldaña, F. J.; Kimura, Y. Variability in Spatially and Temporally Resolved Emissions and Hydrocarbon Source Fingerprints for Oil and Gas Sources in Shale Gas Production Regions. *Environmental Science and Technology* **2017**, *51*, 12016–12026.

(8) Wang, J. L.; Daniels, W. S.; Hammerling, D. M.; Harrison, M.; Burmaster, K.; George, F. C.; Ravikumar, A. P. Multiscale Methane Measurements at Oil and Gas Facilities Reveal Necessary Frameworks for Improved Emissions Accounting. *Environmental Science & Technology* **2022**, *56*, 14743–14752.

(9) Brandt, A. R.; Heath, G. A.; Cooley, D. Methane Leaks from Natural Gas Systems Follow Extreme Distributions. *Environmental Science and Technology* **2016**, *50*, 12512–12520.

(10) Cusworth, D. H.; Duren, R. M.; Thorpe, A. K.; Olson-Duvall, W.; Heckler, J.; Chapman, J. W.; Eastwood, M. L.; Helmlinger, M. C.; Green, R. O.; Asner, G. P.; Dennison, P. E.; Miller, C. E. Intermittency of Large Methane Emitters in the Permian Basin. *Environmental Science and Technology Letters* **2021**, *8*, 567–573.

(11) Bell, C.; Ilonze, C.; Duggan, A.; Zimmerle, D. Performance of Continuous Emission Monitoring Solutions under a Single-Blind Controlled Testing Protocol. *Environmental Science and Technology* **2023**, *57*, 5794–5805.

(12) Daniels, W. S.; Wang, J. L.; Ravikumar, A. P.; Harrison, M.; Roman-White, S. A.; George, F. C.; Hammerling, D. M. Toward Multiscale Measurement-Informed Methane Inventories: Reconciling Bottom-Up Site-Level Inventories with Top-Down Measurements Using Continuous Monitoring Systems. *Environmental Science & Technology* **2023**,

(13) Barchyn, T. E.; Hugenholtz, C. H.; Gough, T.; Vollrath, C.; Gao, M. Low-cost fixed sensor deployments for leak detection in North American upstream oil and gas: Operational analysis and discussion of a prototypical program. *Elementa: Science of the Anthropocene* **2023**, *11*.

(14) Chen, Q.; Modi, M.; McGaughey, G.; Kimura, Y.; McDonald-Buller, E.; Allen, D. T. Simulated methane emission detection capabilities of continuous monitoring networks in an oil and gas production region. *Atmosphere* **2022**, *13*, 510.

(15) Chen, Q.; Kimura, Y.; Allen, D. T. Defining Detection Limits for Continuous Monitoring Systems for Methane Emissions at Oil and Gas Facilities. *Atmosphere* **2024**, *15*, 383.

(16) Daniels, W. S.; Jia, M.; Hammerling, D. M. Detection, localization, and quantification of single-source methane emissions on oil and gas production sites using point-in-space continuous monitoring systems. *Elementa: Science of the Anthropocene* **2024**, *12*.

(17) Chen, Q.; Schissel, C.; Kimura, Y.; McGaughey, G.; McDonald-Buller, E.; Allen, D. T. Assessing detection efficiencies for continuous methane emission monitoring systems at oil and gas production sites. *Environmental Science & Technology* **2023**, *57*, 1788–1796.

(18) Sun, C.; Yu, Y.; Li, V. O.; Lam, J. C. Multi-type sensor placements in Gaussian spatial fields for environmental monitoring. *Sensors* **2019**, *19*, 189.

(19) Mukherjee, R.; Diwekar, U. M.; Kumar, N. Real-time optimal spatiotemporal sensor placement for monitoring air pollutants. *Clean Technologies and Environmental Policy* **2020**, *22*, 2091–2105.

(20) Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J.; Glance, N. Cost-effective outbreak detection in networks. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. 2007; pp 420–429.

(21) Mukherjee, R.; Diwekar, U. M.; Vaseashta, A. Optimal sensor placement with mitigation strategy for water network systems under uncertainty. *Computers & Chemical Engineering* **2017**, *103*, 91–102.

(22) Kelp, M. M.; Lin, S.; Kutz, J. N.; Mickley, L. J. A new approach for determining optimal placement of PM2. 5 air quality sensors: case study for the contiguous United States. *Environmental Research Letters* **2022**, *17*, 034034.

(23) Liu, Z.; Li, X. Sensor layout strategy for source term estimation of external pollution sources in urban neighbourhoods. *Building and Environment* **2022**, *220*, 109276.

(24) Liu, X.; Yeo, K.; Klein, L.; Hwang, Y.; Phan, D.; Liu, X. Optimal Sensor Placement for Atmospheric Inverse Modelling. 2022 IEEE International Conference on Big Data (Big Data). 2022; pp 4848–4853.

(25) Malings, C.; Pozzi, M.; Klima, K.; Bergés, M.; Bou-Zeid, E.; Ramamurthy, P. Surface

heat assessment for developed environments: Probabilistic urban temperature modeling. *Computers, Environment and Urban Systems* **2017**, *66*, 53–64.

(26) Krause, A.; Singh, A.; Guestrin, C. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research* **2008**, *9*.

(27) Du, W.; Xing, Z.; Li, M.; He, B.; Chua, L. H. C.; Miao, H. Optimal sensor placement and measurement of wind for water quality studies in urban reservoirs. IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks. 2014; pp 167–178.

(28) Gao, H.; Liu, J.; Lin, P.; Hu, G.; Patruno, L.; Xiao, Y.; Tse, K.; Kwok, K. An optimal sensor placement scheme for wind flow and pressure field monitoring. *Building and Environment* **2023**, *244*, 110803.

(29) Cooper, J.; Dubey, L.; Hawkes, A. Methane detection and quantification in the upstream oil and gas sector: the role of satellites in emissions detection, reconciling and reporting. *Environmental Science: Atmospheres* **2022**, *2*, 9–23.

(30) Johnson, M. R.; Conrad, B. M.; Tyner, D. R. Creating measurement-based oil and gas sector methane inventories using source-resolved aerial surveys. *Communications Earth & Environment* **2023**, *4*, 139.

(31) Fox, T. A.; Barchyn, T. E.; Risk, D.; Ravikumar, A. P.; Hugenholtz, C. H. A review of close-range and screening technologies for mitigating fugitive methane emissions in upstream oil and gas. *Environmental Research Letters* **2019**, *14*, 053002.

(32) Klise, K. A.; Nicholson, B. L.; Laird, C. D.; Ravikumar, A. P.; Brandt, A. R. Sensor placement optimization software applied to site-scale methane-emissions monitoring. *Journal of Environmental Engineering* **2020**, *146*, 04020054.

(33) Zi, Y.; Fan, L.; Wu, X.; Chen, J.; Han, Z. Distributionally Robust Optimal Sensor Placement Method for Site-Scale Methane-Emission Monitoring. *IEEE Sensors Journal* **2022**, *22*, 23403–23412.

(34) Land, A. H.; Doig, A. G. *An automatic method for solving discrete programming problems*; Springer, 2010.

(35) Padberg, M.; Rinaldi, G. A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems. *SIAM review* **1991**, *33*, 60–100.

(36) Zbigniew, M. Genetic algorithms+ data structures= evolution programs. *Comput Stat* **1996**, 372–373.

(37) Kim, C.; Oh, H.; Chang Jung, B.; Moon, S. J. Optimal sensor placement to detect ruptures in pipeline systems subject to uncertainty using an Adam-mutated genetic algorithm. *Structural health monitoring* **2022**, *21*, 2354–2369.

(38) Zhang, C.; Chun, Q.; Leng, J.; Lin, Y.; Qian, Y.; Cao, G.; Dong, Q. Optimal placement method of multi-objective and multi-type sensors for courtyard-style timber historical buildings based on Meta-genetic algorithm. *Structural Health Monitoring* **2024**, *23*, 1468–1497.

(39) Deb, K. *Multi-objective evolutionary optimisation for product design and manufacturing*; Springer, 2011; pp 3–34.

(40) Qian, C.; Bian, C.; Feng, C. Subset selection by pareto optimization with recombination. Proceedings of the AAAI Conference on Artificial Intelligence. 2020; pp 2408–2415.

(41) Bell, C.; Ilonze, C.; Duggan, A.; Zimmerle, D. Performance of continuous emission monitoring solutions under a single-blind controlled testing protocol. *Environmental Science & Technology* **2023**, *57*, 5794–5805.

(42) Jia, M.; Daniels, W.; Hammerling, D. Comparison of the Gaussian plume and puff atmospheric dispersion models for methane modeling on oil and gas sites. **2023**,

(43) Harris, C. R.; Millman, K. J.; Van Der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; others Array programming with NumPy. *Nature* **2020**, *585*, 357–362.

(44) Furuta, D.; Sayahi, T.; Li, J.; Wilson, B.; Presto, A. A.; Li, J. Characterization of inexpensive metal oxide sensor performance for trace methane detection. *Atmospheric Measurement Techniques* **2022**, *15*, 5117–5128.

TOC Graphic