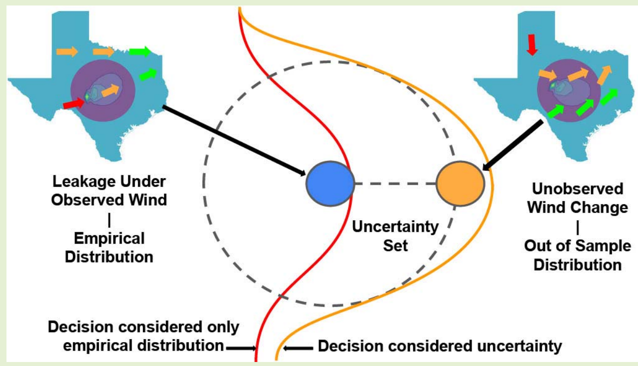
Distributionally Robust Optimal Sensor Placement Method for Site-Scale Methane-Emission Monitoring

Yuan Zi, *Student Member, IEEE*, Lei Fan, *Senior Member, IEEE*, Xuqing Wu, *Member, IEEE*, Jiefu Chen, *Senior Member, IEEE*, and Zhu Han, *Fellow, IEEE*

***Abstract*—Recent research in deterministic sensor placement optimization technologies has improved the capability of monitoring large-scale field environments with a limited budget. In traditional stochastic mixed-integer linear programming formulations, minimizing the expectation of detection time can lead to a detector placement with good average behavior but unexpected worst case behavior. The uncertainty factors in the complex environment and sensor system significantly challenge the effects of the placement strategy provided by stochastic programming (SP). These factors include unknown leakage rate and location, sensor delay, and primary uncertainty of wind conditions. This article introduces a distributionally robust optimization (DRO) formulation of sensor placement under the uncertainty of wind conditions and improves a sensor network's detection robustness. The method is demonstrated using the atmospheric simulation with site-specific methane-emission scenarios that capture partial natural wind conditions and emission characteristics. DRO techniques are employed to determine sensor locations that minimize the detection time expectation of the emission scenarios with a significantly better worst case behavior. Experiment results show that the proposed DRO method outperforms the sensor placement methods based on SP.**



***Index Terms*— Carbon monitoring, distributionally robust optimization (DRO), methane sensor, mixed-integer programming, optimal sensor placement.**

NOMENCLATURE

$e \in \mathcal{E}$	Set of all events.
L	Set of all candidate sensors.
L_e	Set of all sensors that are capable of detecting event e.
p_e	Probability of occurrence for event e.
$d_{e,i}$	Damage coefficient for leak event e at location i.
$d'_{e,i}$	Worst case expectation of $d_{e,i}$ under uncertainty.
$x_{e,i}$	Indicator for location i that first detects event e.
y_l	Binary variable indicating if a sensor is installed at location l.
c_i	Cost of sensor i.
c	Sensors' budget.
κ	Radius of the uncertainty ball.
$\mathbb{B}_{\kappa}^{d_{e,i}}$	Uncertainty set.
Q	Arbitrary distribution within uncertainty set.
T	Empirical distribution of $d_{e,i}$.
G_+	Set of all probability distributions.
$\mathcal{D}_W$	Wasserstein distance.
S	Number of historical data for empirical distribution.
H	Number of bins for empirical distribution.
γ	Confidence level.

Manuscript received 24 August 2022; revised 5 October 2022; accepted 7 October 2022. Date of publication 1 November 2022; date of current version 30 November 2022. This work was supported in part by NSF under Grant CNS-2107216, Grant CNS-2128368, and Grant CMMI-2222810. The work of Lei Fan was supported in part by NSF under Grant ECCS 2045978. The associate editor coordinating the review of this article and approving it for publication was Prof. Subhas Chandra Mukhopadhyay. *(Corresponding author: Zhu Han.)*

Yuan Zi and Jiefu Chen are with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA.

Lei Fan is with the Department of Engineering Technology, University of Houston, Houston, TX 77004 USA.

Xuqing Wu is with the Department of Information and Logistics Technology, University of Houston, Houston, TX 77004 USA.

Zhu Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 02447, South Korea (e-mail: hanzhu22@gmail.com).

Digital Object Identifier 10.1109/JSEN.2022.3214176

I. INTRODUCTION

METHANE is the primary component of natural gas and the second most anthropogenic greenhouse gas emitted into the atmosphere, which has global warming power estimated to be 28–36 times more than CO_2 over the last 100 years from release [1], [2] and 80 times more potent over 10–20 years from release. Roughly, a third of the contemporary anthropogenic methane emissions come from the fossil fuel energy sector worldwide [3], [4], [5]. Curbing methane emissions happening in new and existing oil and gas infrastructure and restricting unexpected leakage and intentional releases of methane at oil and natural gas facilities is considered an effective strategy to slow the rate of near-term climate warming [6]. New sensors and monitoring technologies have been increasingly studied and developed to identify and remove methane leakage in the oil and gas industry. A key challenge of methane-emission monitoring is that of comprehensively, accurately, and robustly measuring large areas at a resolution that allows the separation of different types of local sources with the lowest cost. Therefore, optimal methane sensors' placement serves as one of the essential techniques to identify the leakage of methane.

Sensor placement optimization problems as one category of facility location problem (FLP) [7] have been studied extensively and applied to a wide range of applications, such as energy management [8], wireless sensor placement [9], [10] [11], water quality sensor placement [12], and surveillance cameras placement [13]. Sensor placement optimization is often formulated as coverage formulation or P-median [14] mixed-integer linear programming (MILP) formulation. With a coverage formulation, the optimal sensors are placed to maximize the geographic sensing covered region straightforwardly and the capability to detect the most scenario from numerous leakage scenarios. With a P-median formulation, sensors are placed to minimize specific emission impact metrics, e.g., the cost associated with damage. Usually, the emission damage is proportional to the first detection time after the emission occurred. These two formulations are early proposed to solve the water sensor placement problem [12]. Then, they are applied to the gas sensor placement [15], [16] in petrochemical facilities and site-scale methane-emission monitoring [17] recently.

Klise et al. [17] proposed a sensor placement optimization scheme based on atmospheric dispersion models and an open-source MILP solver and developed an open-source Python package named Chama. In this sensor placement scheme, both maximum-coverage and minimum-impact problems are formulated and solved in two steps. The first step is the simulation of methane-emission dispersion maps based on the possible leak data (leak rate and leak locations) and wind data (wind speed and wind direction). The second step is to solve an MILP problem with given sensor data (sensitivity and budget) using the MILP solver. This scheme uses multiple simulations, called scenarios or leakage events, to represent the system under different conditions to account for uncertainty. The methane sensing system discussed in this article is a point sensor system placed in cubic space with averaged weather data sampled in Fort Worth, Texas, which contains oil&gas facilities as potential methane leakage sites. The current oversimplified weather model would cause performance to deteriorate when the real condition is different from the condition used for sensor placement simulation. The current solution algorithm would need to simulate all possible scenarios with all possible weather conditions on this site to prevent this problem. Again, the methane sensor placement is facing the following challenges.

1) Unexpected wind conditions caused detection time shifts.
2) Sensor bias, noise, and failure caused detection time shift.
3) Methane propagation physical modeling error compares to the real environment.

One straightforward way to attempt to solve these challenges could be given as follows:

1) utilizing more complex wind prediction models or increasing simulation numbers trying to capture the distribution of nature [18], [19];
2) improving the sensor's detection time by using more advanced sensors, denoising technology, or simply increasing the sensor's number [20], [21] [22];
3) improving the physical model of methane propagation [23].

Each method has its technical challenges and can significantly increase the solution's complexity and costs, such as sensor budget and computational cost. Moreover, the facts mentioned above will cause the detection time distribution to shift between simulated and reality.

To solve uncertainty challenges generally, there are three optimization approaches under the uncertainty in the optimization domain: robust optimization (RO), stochastic programming (SP), and distributionally RO (DRO). The RO method considers the worst case scenarios as constraints, with a significantly lower probability of actual occurrences. Although it provides a safe guarantee, the RO method leads to the most conservative strategies with moderate performance. The SP method typically assumes that the decision-maker has complete information on the uncertainty distribution. However, this assumption is too extreme since it usually does not hold. In most problems, decision-makers do not know the true distribution of the uncertainty, especially when the data are limited. The previous P-median MILP formulation is an SP method assuming a uniform distribution of simulated leakage events. Ultimately, DRO [24] bridges the gap between RO and SP, which builds an uncertainty set of the distribution for uncertainty parameters based on the data. The data-driven method, the DRO method, can efficiently utilize the limited dataset and provide a robust solution. Moreover, this solution has a better performance than the RO method and a safer guarantee than traditional methods that do not consider uncertainty in optimization tasks, such as portfolio selection [25], flexible generation resources management [26], and computational offloading [27]. Therefore, this article proposes a distributionally robust methane sensor placement optimization framework that is robust to detection time shifts caused by unexpected wind conditions. The proposed method can resist the unseen weather performance in the experimental site because it had estimated the uncertainty level before the optimization and optimized the sensor placement under the worst case scenario. Instead of improving the specific component of the sensor placement scheme, we propose to estimate the possible worst time-shift distribution based on the historical data. With this uncertainty estimation, the method provides a more robust sensor placement strategy for the sensor placement. This proposed method also adaptively adjusts the uncertainty constraints level of the uncertainty set based on a data-driven way.

The contribution of this article can be summarized as follows.

1) The experiments in this article show that the performance of the traditional sensor placement optimization method will become worse as the environment distribution shift, which is a typical challenge for all methane sensor placement problems.
2) To the authors' best knowledge, this article is the first data-driven DRO method for solving sensor placement problems with risk awareness.
3) In formulation, this article is the first to consider the impact of leakage events like the first detection time as a distribution instead of a deterministic variable. This distribution assumption makes quantification of uncertainty possible.
4) In solution, this article proposes the integrated solution that combines the new proposed risk estimation and previous open-source MILP solution to solve the distributional shift under uncertainty caused by any uncertain facts for methane sensor placement. Moreover, the example and experiments of single-sensor placement, single-sensor detection time analysis, and multiple wind conditions show the proposed method's effectiveness.

The remainder of this article is arranged as follows. Section II describes the problem formulation and distributionally robust approach in detail. Section III shows the optimization results of our method using the synthetic training dataset and testing dataset of leakage constructed based on the historical leak and wind data.

II. METHOD

A. Problem Statement

The methane sensor placement problem is a practical general sensor placement problem for carbon neutrality. Sensor placement is the subset of location planning problems that involves specifying the physical position of facilities that provide demanded services. Examples of facilities include hospitals, restaurants, ambulances, retail and grocery stores, schools, and fire stations. In this article, the facilities are the methane sensors, and the demanded service provided by the methane sensors is methane leakage monitoring. The ultimate goal is to use the least amount of sensor budget to design sensor networks for unknown leakage events with the fastest alert time under unknown environmental conditions. There were various models to solve this problem, and the most common model considered a specific type of discrete location model for numerical simulation purposes. Leaking sources and sensors are in discrete positions in a 3-D space. The P-median formulation is a specific type of discrete problem formulation originally proposed to place p facilities to minimize the (demand-weighted) average distance between a demand node and the location in which a facility was placed. In this article, the modified P-median formulation is proposed for tackling the uncertainty challenges in the sensor methane placement problem.

B. Problem Formulation

The distributionally robust P-median mixed-integer programming formulation used in methane-emission applications to place sensors that minimize detection time under uncertainty is given by

$$\min_{y_l} \sum_{e\in\mathcal{E}_e} p_e \sum_{i\in\mathbb{L}_e} \sup_{Q\in\mathbb{B}_\kappa^{d_{e,i}}} E\left(Q\right) x_{e,i} \tag{1}$$

subject to

$$\sum_{l\in\mathbb{L}} y_l \leq c \tag{2}$$

$$x_{e,i} \leq y_i \quad \forall\, e \in \mathcal{E},\, i \in L_e \tag{3}$$

$$\sum_{i\in\mathbb{L}_e} x_{e,i} = 1 \quad \forall\, e \in \mathcal{E} \tag{4}$$

$$y_l \in \{0, 1\} \quad \forall\, l \in \mathbb{L} \tag{5}$$

$$0 \leq x_{e,i} \leq 1 \quad \forall\, e \in \mathcal{E},\, i \in L_e \tag{6}$$

$$\mathbb{B}_\kappa^{d_{e,i}} := \{Q \in G_+ : \mathcal{D}_W\left(Q,\, T\right) \leq \kappa\}. \tag{7}$$

The notation used in the problem is summarized in the Nomenclature. In the interest of the methane leakage site, there exists a set of potential methane sensor locations, defined as $L = \{1, 2, \ldots, N\}$, where L is the set of all N potential locations, indexed by l. $\mathcal{E} = \{1, 2, \ldots, M\}$ represents the set of leakage events considered, where $\mathcal{E}$ is the set of all M leakage events. A single event is indicated by e. Not all sensor locations are affected by each leakage event. The subsets $L_e \subseteq L, \forall e \in \mathcal{E}$, are defined such that L_e contains all the sensor locations that can detect particular leakage event e. Parameter p_e is the probability associated with leakage event e out of all events in $\mathcal{E}$, and $\sum_{e\in\mathcal{E}} p_e = 1$.

It is important to note that $d_{e,i}$ is the applicable damage coefficient if leak event e is first detected by a sensor at location i. For the methane case study, previous studies [15], [28] pursue to minimize the expected value of the first detection time of all simulated leak events with the assumption of the total volume of leakage from the event e corresponding to first detection time by a methane sensor at location i. Therefore, $d_{e,i}$ is designed to be the time required for a sensor at location i to detect the leak simulated in the event e. In previous studies' [15], [28] formulation, $d_{e,i}$ is a deterministic parameter and neglects the statistical nature of the real environment. In our proposed formulation, $d_{e,i}$ is a distribution considering sensor delay caused by the uncertainty of environment modeling.

The binary decision variable y_l represents the existence of a sensor ($y_l = 1$) or not ($y_l = 0$) at location l. Equation (2) constrains the maximum number of methane detectors placed to be no more than the parameter c, the total sensor budget. Each $x_{e,i}$ is an indicator variable that has a value of 1 if a sensor at location i is the first sensor to detect the leak event e among all sensors active for this event and 0 otherwise. The difference between y and x is that there might be multiple sensors installed and activated by the leak event, but only one is the first sensor to detect this event. Equation (3) ensures that location i can only be the first to detect leak event e if there is a sensor placed at location i. Therefore, if $y_i = 0$, the location i cannot claim to be the first to detect an event.

The objective function equation (8) can be reformulated to a standard MILP problem's minimum objective function [17] by solving the following distributionally optimization problem first:

$$d'_{e,i} = \max_{Q \in \mathbb{B}_{\kappa}^{d_{e,i}}} E(Q) \quad \text{subject to} \tag{8}$$

$$\mathbb{B}_{\kappa}^{d_{e,i}} := \{Q \in G_+ : \mathcal{D}_W(Q, T) \leq \kappa\} \tag{9}$$

$$\kappa = \left(\frac{H}{2S}\right) \log\left(\frac{2H}{1-\gamma}\right) \tag{10}$$

where G_+ represents the set of all possible distributions. In (9), $\mathcal{D}_W$ denotes the probability distance between any arbitrary distribution and the reference empirical distribution, $d_{e,i}$ within uncertainty set $\mathbb{B}_{\kappa}^{d_{e,i}}$, $Q \in \mathbb{B}_{\kappa}^{d_{ei}}$. In this article, we use the L_1 Wasserstein distance as $\mathcal{D}_W$. According to [29, Proposition 1], given a set of historical data, (10) defines the relationship between the size of data and the value of uncertainty level κ under the L_1 norm. Here, S, H, and γ denote the size of historical data, the number of bins, and the confidence level, respectively.

C. Solution

The major innovation of the proposed method is that it provides an uncertainty-resistant robust solution by leveraging the cutting-edge uncertainty-aware optimization theory and historical data of methane leakage and environment features generated from the physical numerical simulation model. Following [17], this article focuses on the remaining open challenge of the persistent nature of methane leaks and uncertainty in atmospheric conditions, which prevent researchers from studying P-median sensor placement formulations for methane leakage sensor placement. We considered the uncertainty of wind speed and constructed the framework based on Fig. 1, composed of five core processes. The first two steps demonstrate similar modeling processes as in [17] and prepare the methane dispersion geospatial characteristic events for optimization. Instead of adopting events' group simulated based on upscaled wind data for MILP directly, we propose a new data-driven component to estimate the uncertainty of the current dataset for optimization of worst case events.

The five core steps of our new proposed solution framework are given as follows.

1) *Historical Data Input*: This step relates to acquiring and preprocessing the data required for the solution. Oil and gas facility locations, including wells, natural gas pipelines, and processing plants, available in the public domain, are identified for the area of interest, which could be inferred from satellite images or civil construction records. Historical emissions distribution data related to source leak rates are a function of the oil and gas facility type. Therefore, the leak rates are selected from empirical data published in the Fort Worth Natural Gas Air Quality Study [30]. Multiple events are simulated to capture methane concentration distribution concerning uncertain leak location, leak rate, leak height, and wind speed and direction. Wind speed and direction are defined using data collected in Fort Worth in 2015 by NOAA [31].
2) *Methane Dispersion Simulation*: This is the physical-based element of the framework that models the movement of methane in the atmosphere over the area of interest using atmospheric-based dispersion models. This article adopts the Gaussian plume model to the atmospheric dispersion model for methane transport schemes.
3) *Estimate Uncertainty of Historical Data*: There are three ways to explore the persistent nature of uncertainty in atmospheric conditions: stochastic optimization (SO), RO, and DRO. Most studies focus on SO: decreasing the uncertain level by increasing the simulation events' number or using more complex atmospheric transport models. In theory, increasing the sample number can result in a better-fit model to the empirical distribution as a known distribution to optimize the objective function. However, the distribution will shift over time due to facts such as climate change [32]. This no-distribution-shift assumption limited these SO methods being applied to the out-of-sample scenarios. Moreover, these methods face significant inaccurate challenges when modeling the atmosphere and simulating with limited computational resources [33]. The RO aims to determine the uncertainty boundary of natural atmospheric distribution and optimize theoretical worst case events, which is over pessimist because the worst case has a very small probability of happening in reality. Therefore, we propose the DRO to optimize over the worst case within an uncertainty set constructed based on the empirical distribution and uncertain level estimated from data. Once the events have been simulated using the physical dispersion model and upscaled wind data in step 2, each potential sensor placement grid point will have an empirical distribution of detection time for each leakage event. Considering that the actual distribution might be different from this empirical distribution, we use statistical inference to define the uncertainty set corresponding to a given radius of the uncertainty set and the Wasserstein L_1 distance. Within this uncertainty set, we can determine the worst Dirac delta distribution of detection time at each sensor grid point.
4) *MILP:* The newly generated worst case events have the same data schema as the original P-media formulation and can be solved by any MILP solver. In this article, we utilize the open-source python-based sensor placement optimization software Chama [28], which constructed optimization using Pyomo [34] and is solved using an open-source GNU linear programming kit (GLPK) [35]. The open-source solver GLPK can be used for many applications. For future work, there are several ways to improve the accuracy of the proposed methods and lower the missing leakage events in the

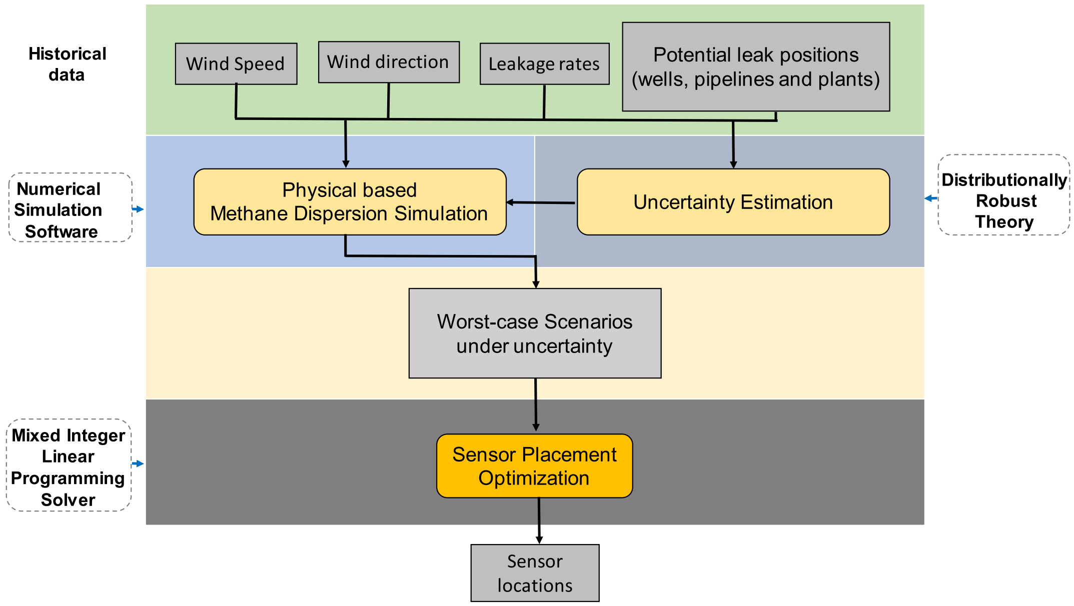


Fig. 1. Workflow of the distributionally robust sensor placement optimization for site-scale methane-emissions monitoring.

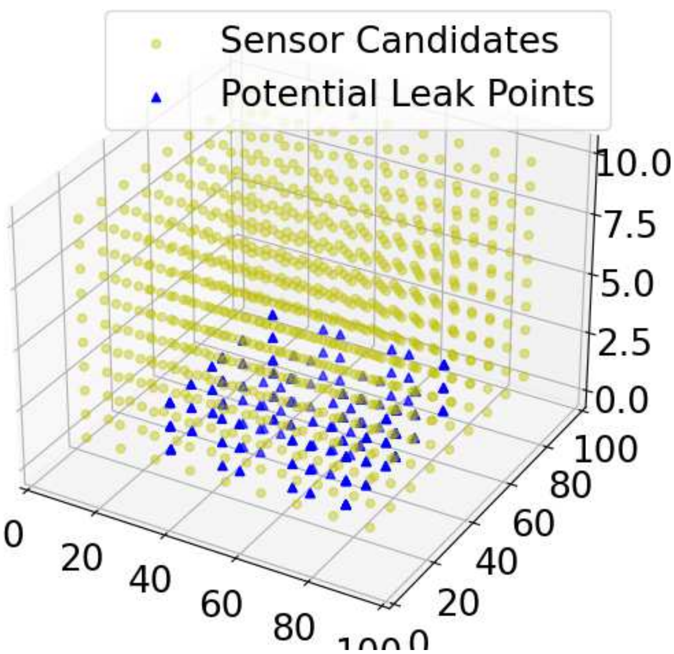


Fig. 2. 3-D view of potential leak locations (blue triangles) and candidate sensor locations (yellow circles).

out-of-sample testing set. First, increasing the sampling number will likely increase the confidence level of avoiding those missing events. Second, use a more reliable and sensible sensor. Third, satellite and air-bone sensor has shown the promising capability to detect methane on a site scale [36], [37]. A DRO multimodel sensor fusion sensor network could also improve the accuracy.

D. Single-Sensor Placement Example

Assuming that there is one leak source and two sensor placement candidates' locations, the total sensor budget can only afford to place one sensor near this methane leak source. Based on the historical data, the minimum detection time in two locations is two distribution, $d_A = [0, 2, 6, 6, 6]$ and $d_B = [3, 3, 3, 3, 8]$. The ultimate goal is to pick one sensor, place location to place sensor, and expect to minimize the detection time in future leak events. If we use the SP method, both sensor candidates have the same expectation detection time $E[d_A] = 4$ and $E[d_B] = 4$, which cannot help distinguish the superior sensor candidates. On the other hand, the RO method will take the worst case scenario of each distribution to make the decision, $\max(d_A) = 6$, $\max(d_B) = 8$, resulting in choosing sensor A as the placement strategy. DRO-based worst case expectations of sensors A and B's minimum detection times are calculated by using the L_1 Wasserstein distance with confidential $\gamma = 0.9$; based on 10, $\kappa = 2$, $E_{\text{worst}}(d_A) = 6$, and $E_{\text{worst}}(d_B) = 14/3$, and choose that the sensor B is the best choice. $E_{\text{worst}}(d_A)$ and $E_{\text{worst}}(d_B)$ are calculated by solving the following equations:

$$\begin{cases} \frac{1}{5}|E_{\text{worst}}(d_A)| + \frac{1}{5}|E_{\text{worst}}(d_A) - 2| \\ +\frac{3}{5}|E_{\text{worst}}(d_A) - 6| = 2 \\ \frac{4}{5}|E_{\text{worst}}(d_B) - 3| + \frac{1}{5}|E_{\text{worst}}(d_B) - 8| = 2. \end{cases} \tag{11}$$

III. Experiment

The case study in Section III-A is similar to [28]. The open-source data [31] are utilized to determine the wind conditions setting. This case study provides a general methane emissions setting that allows methods' development and evaluation. The case study's simulation and optimization are performed using Chama [17], and the open-source solver GLPK [35] with custom-developed uncertainty quantification and detection-time-shift correction tool.

A. Simulations Setting

Fig. 2 shows the 3-D simulation domain. The simulation area is defined as a 100 × 100 × 10 m cubic region. There are 30 potential source points (blue triangles) within this area,

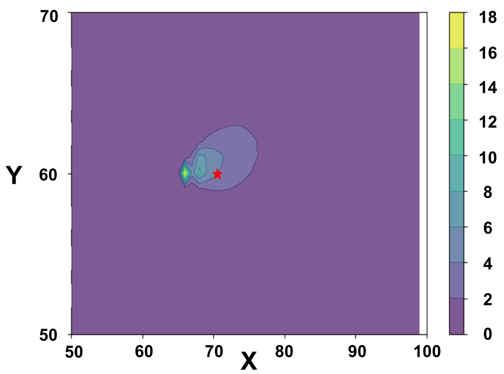


Fig. 3. Methane plume simulation using one wind speed trace, one sensor (red star), and one leak rate in the top-view view.

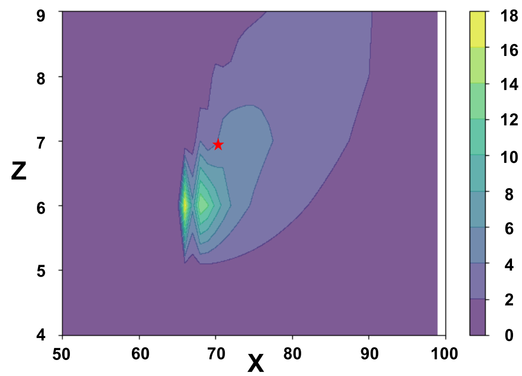


Fig. 4. Methane plume simulation using one wind speed trace, one sensor (red star), and one leak rate in the cross-sectional view.

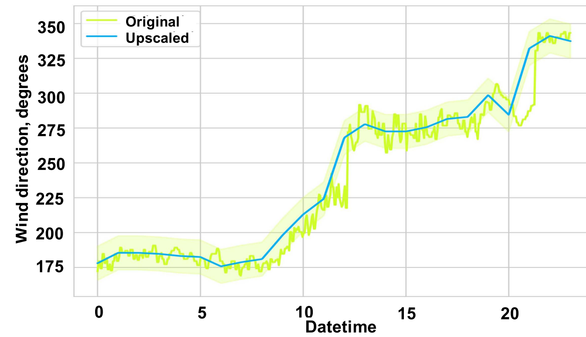


Fig. 5. Wind direction over a single day, including original 1-min original data (green line) and data averaged to 1 h (blue line).

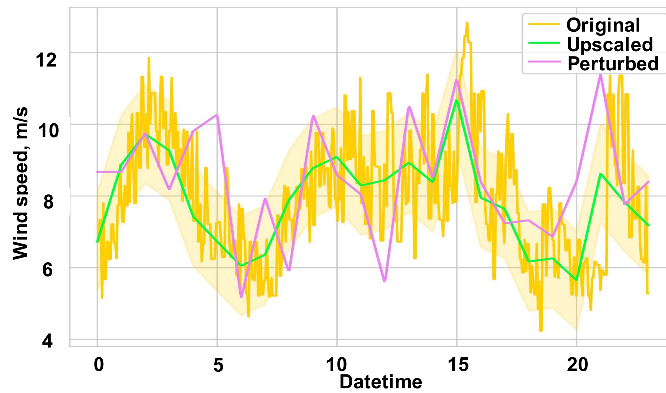


Fig. 6. Wind speed over a single day, including original 1 min original data (orange line), data averaged to 1 h (green line), and one perturbation wind trace (pink line).

indicating ten potential locations near the surface (0 m–2 m). These points represent equipment and devices that have the potential to emit methane, such as well, pipelines, and plants. Moreover, candidate sensor locations (yellow circles) are the grid points that can install sensors for monitoring.

For the physical simulation, the methane leakage propagation processing and concentration are modeled using the Gaussian plume model in Chama [17]. Each leak event simulates a single emission scenario (one leak source point with one leak rate) over one day using a 1-h time step. In addition, we set the impact of this event to 72 h if the event is failed to be detected. In space, the simulation model domain is divided into one m^3 cell. Figs. 3 and 4 show the top view and side view of an example of a leak event extent scenario and one sensor candidate position near the source location, respectively. According to the physical model, the extent of methane plume increases as the leak rate increases. The extent of methane plume and detection time for the candidate sensor is also affected by the wind speed and direction. In the simulation, the steady-state Gaussian plume model's extent is recomputed at each time step as the wind speed and direction change over the day.

Multiple events are simulated as events set and used to capture methane concentration concerning the uncertainty of leak location, leak rate, leak height, wind speed, and direction. However, it is difficult to capture the distribution of all uncertainty by simply increasing the simulation scenarios' number. Although the leak position and leak range might be stable in a particular location and range, the wind data's uncertainty can be hard to capture with limited data. Moreover, the computational costs also increase drastically when the simulation's wind conditions increase.

In the previous study, Klise et al. [17] proposed reducing the amount of wind data used in the simulations by random selection and vector averages. Wind speed and direction are defined using data collected in Fort Worth in 2015 by NOAA [31]. A comparison of the original wind data and averaged wind data, for one example day's wind direction and wind speed, are shown in Figs. 5 and 6, respectively. The averaging result focuses on capturing predominant wind characteristics throughout the year but results in a loss of wind variability. Fig. 6 also shows one perturbation wind speed condition sampled from assumed real wind speed distribution, assuming as a Gaussian distribution with mean as average wind speed and before average variance as the stand deviation. Leak rates are selected from empirical data as in [17], as shown in Fig. 7. Each leak event simulation is defined by one day of wind speed trace, one wind direction trace, one leak location, and one leak rate.

In this case study, the locations for candidate sensors are defined using a regular grid throughout the model domain, same as setting in [17], at 10-m spacing in the x- and y-directions and 1-m spacing (starting at the ground elevation) in the z-direction. This setting results in 810 ($9 \times 9 \times 10$) candidate sensors. The total sensor budget is set to \$100 000, and each sensor has 0.1 active thresholds with \$10 000 cost.

TABLE I
SENSORS

Sensors	Type	Range	Time resolution	Accuracy	Cost
Point detector [17]	Ideal point sensor	0.1-100 ppm	1/3600 Hz	0	$10,000
TGS 2611-E00	MOS type	500–10,000 ppm	1 Hz	1.7 ppm	$15
Los Gatos UGGA	Spectroscopic	0.1-100 ppm	1 Hz	2 ppb	$50,000

B. Sensor's Discussion

The sensor used in our numerical simulation experiment is an ideal point sensor with characters listed in Table I. We also provide both one high cost and low cost, most commonly used sensor as the reference for practical use reference. Methane is presented in the atmosphere at low concentrations, with a global background concentration of around 1.8 ppm [38]. The most used sensor in environments with low methane concentrations is the cavity ring-down spectrometry (CRDS)-type sensor [39]. One example of a CRDS sensor is Los Gatos UGGA that served as a reference instrument for methane [40]. It has a high precision (one standard deviation <2 ppb at 1 Hz) and has been used on long-term measurements [41], [42]. However, the cost and the power consumption remain an issue. Therefore, there is a growing interest in incorporating it into monitoring networks [43]. One example of a low-cost sensor is the Taguchi Gas Sensors, such as TGS 2611(Figaro Engineering Inc., Osaka, Japan), which are designed to measure ambient methane mixing ratios between 500 and 10 000 ppm.

C. Detection Time Distributionally Robust Analysis

Fig. 3 shows one leak event leaked at (65, 60, 5). The one sensor candidate near this source is plotted as a red star located at (70, 50, 7). The leak rate is set to 5, and the detection threshold of this sensor is set to 0.01. With the Gaussian wind speed distribution, randomly sample 25 wind speed traces to generate detection time's empirical distribution, as shown in Fig. 8. The blue bins indicate the empirical distribution of the minimum detection time at this sensor place. The green bin is the expectation of the minimum detection time at this sensor candidate place, and the red bin indicates the distributionally robust expectation of the minimum detection time. From Fig. 8, the confidential level changes from 0.5 to 0.7 and 0.9, and the associated distributionally worst spike distribution is the shift from the low value near the mean to the maximum value of the original distribution. Accordingly, the uncertainty set contains more distribution and the method because more conservative. When the confidentiality level increases, the result obtained from DRO will become closer to that obtained from RO.

D. Dataset's Preparing

With the vector averaged wind conditions as the fixed environment, 37 basic leak events set are generated by sampling different values from the leak rate data and leak height distribution. As shown in Figs. 5 and 6, we assume that the true natural wind speed and direction distribution are the Gaussian distributions with averaged wind speed and direction trace as the mean. The variances of these Gaussian distributions are calculated from the actual measured

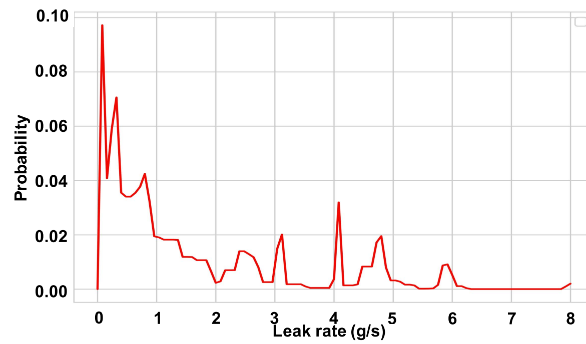


Fig. 7. Probability distribution of leak rate data.

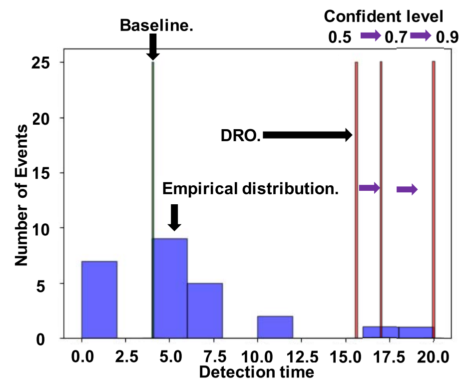


Fig. 8. Probability distribution of minimal detection data (blue bins). The expectation of this distribution (green spike). Distributionally robust expectation spike distribution (red) with different confidence levels.

wind data. For the testing dataset, we sample four disturbed wind conditions from Gaussian. Wind conditions, combined with basic leak events set to generate testing leak events dataset, result in 148 events (four wind conditions × 37 leak scenarios). Like the testing leak events set, we assume that the observation wind condition is different from the true wind distribution. We modeled the observed wind distributions as two Gaussian distributions with perturbated wind speed like Fig. 6 and perturbated wind direction traces as the mean, which can cause up to 30% scale wind speed and wind direction shift. The observed wind distribution also has a half-scaled variance than the true distribution, making it more biased. We sampled two wind conditions from the observation wind distribution and utilized them to generate 74 leak events (two wind conditions × 37 leak scenarios) as the observation leak events set.

E. Results

Optimal sensor placements of different methods are validated using testing leak events set, which has different scenarios than those used in the optimization. This comparison helps determine how optimal sensor placement methods perform under different conditions. Three methods are compared in this

TABLE II
TESTING ACCURACY

Methods	Testing Accuracy of in-sample events	Testing Accuracy of Out-of-sample events	Accuracy Regret Value
MEAN [17]	100%	79.73%	20.27%
SO [19]	95.95%	84.46%	11.49%
DRO	95.95%	87.16%	8.78%

TABLE III
TESTING OBJECTIVE

Methods	Testing Objective of in-sample events	Testing Objective of Out-of-sample events	Objective Regret Value
MEAN [17]	16.97297297	24.79054054	-7.81756757
SO [19]	17.97297297	20.11486486	-2.14189189
DRO	18.55405403	18.86486486	-0.31081083

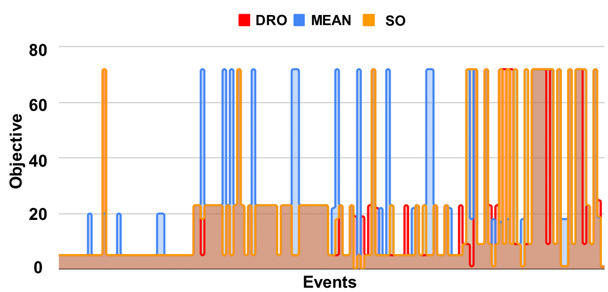


Fig. 9. Minimal detection time on testing leak events set: DRO (red), MEAN (blue), and SO (orange).

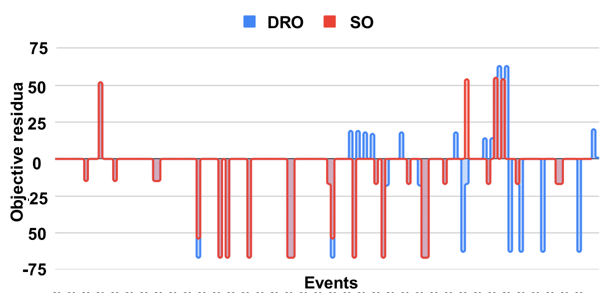


Fig. 10. DRO and SO residual of minimal detection time on testing leak events set with the mean method as the baseline.

section, DRO sensor placement method, SO sensor placement method with averaged wind condition (MEAN), and SO sensor placement with all leak events in observation wind conditions (SO). Also, these methods are validated in leak events used for their optimization. The comparison optimization performance and the testing performance of traditional averaged wind method and traditional stochastic method on the limited observation leak events face the performance dropping when facing out-of-sample performance dropping. In other work, these traditional methods are biased toward observing leak events and cannot be generalized to the out-of-sample scenarios.

Fig. 9 compares minimal detection time distributions from a distributionally robust optimal sensor placement strategy to an average wind optimization sensor placement strategy and stochastic optimal sensor placement strategy using the minimal detection time distributions. Also, with the mean method's distribution result as the baseline, the residual of DRO and SO methods has been shown in Fig. 10. Table II compares DRO, SO with observation wind conditions, and SO with mean wind conditions using the detection accuracy. Moreover, Table III compares these three methods using the

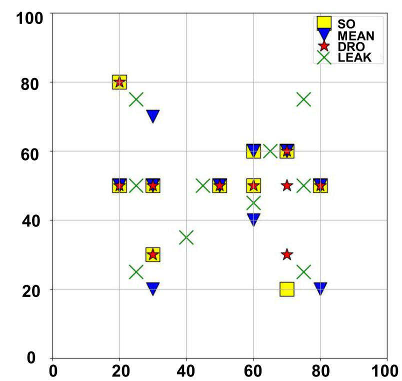


Fig. 11. Optimal sensor placements using SO, MEAN, and DRO.

minimal detection time expectation objective. Results show that the optimal placement consistently outperforms sensors placed using the SO method. As the newest proposed method in the literature, the mean method has high accuracy and objective result in leak events used for its optimization. It can detect all 37 basic leak events (100% accuracy) with 16.97297297-h expected minimal detection time. However, it fails to detect 30 leak events in the testing leak events set with 148 events in total, which results in roughly 20% accuracy dropping (79.73%), and the objective increasing from 16.97297297 to 24.79054054 h. Similar results are observed when we increase the leak events used for the SO. The SO method can achieve 95% accuracy with three events missing and 17.97297297 objectives in the observation leak events set but still has significant performance dropping in the testing leak events set, 84% and 20.11486486, which means that it fails to detect 23 events out of 148 testing events. The proposed DRO method has a relatively robust performance with the accuracy performance of 95% accuracy in observation events set with three missing events and 18.55 objectives and in the testing set with 87% accuracy, 19 missing events, and 18.86 objectives. Fig. 11 shows the sensor placement strategies of the stochastic method by using a single mean wind condition or multiple observation wind conditions, the placement of DRO with the same observation wind conditions, and the leakage source locations.

All methods have performance dropping when validating their placement optimized from their observation dataset to the testing dataset because the observation dataset and testing dataset are sampled from two different distributions with a distribution shift. The observation is limited to the number and observation environment and bias; therefore, the optimization result solely based on these observation data faces the generalization problem. However, because DRO is considered, the worst case events are based not only on the empirical distribution (observation distribution) but also on all possible distributions within the uncertainty set. In the DRO theory, this uncertainty set contains the true distribution of the natural environment at a confident level. The validation performance can remain relatively robust with the DRO optimization result event when validation leak events are sampled from a distribution that differs from the observation leak events distribution.

IV. Conclusion

In this article, we proposed a DRO formulation of the site-scale methane-emission sensor placement problem for oil and gas industry carbon footprint monitoring. To the best of the authors' knowledge, this is the first DRO method that has been proposed to provide a data-driven robust sensor placement solution. Accounting for the big gap between observation and the real-life environment is also known as a distribution shift or generalization problem. When the observation sample is limited, we proposed an uncertainty-level estimation component to provide a safe solution under the uncertain. The experiments demonstrated as follows. First, traditional and proposed methods face performance dropping when validating their placement on testing sets that are different from the leak events used in their optimization. Second, insufficient sample or biased observation of the environment will harm the performance of non-RO and cause significant damage such as missing events or increasing the detection time. Third, the proposed DRO sensor placement method has the lowest performance gap dropping in terms of accuracy and objective. Also, it achieved the highest performance on the out-of-sampled testing leak events set. Finally, in all results presented in this article, the proposed distributionally robust sensor placement algorithm has a relatively better robust performance when facing observation limitations (observation data shortage). It took uncertainty into account and performed risk-aware updates during optimization. For future work, there are several ways to improve the accuracy of the proposed methods and lower the missing leakage events in the out-of-sample testing set. First, increasing the sampling number will likely increase the confidence level of avoiding those missing events. Second, use a more reliable and sensible sensor with more practical characteristics. Third, satellite and air-bone sensor has shown the promising capability to detect methane on a site scale [36], [37]. A DRO multimodel sensor fusion sensor network could also improve the accuracy.

References

[1] S. N. Riddick et al., "Methane emissions from oil and gas platforms in the North Sea," *Atmos. Chem. Phys.*, vol. 19, no. 15, pp. 9787–9796, Jun. 2019.

[2] R. A. Alvarez, S. W. Pacala, J. J. Winebrake, W. L. Chameides, and S. P. Hamburg, "Greater focus needed on methane leakage from natural gas infrastructure," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 17, pp. 6435–6440, Apr. 2012.

[3] J. D. Maasakkers et al., "Global distribution of methane emissions, emission trends, and oh concentrations and trends inferred from an inversion of GOSAT satellite data for 2010–2015," *Atmos. Chem. Phys.*, vol. 19, no. 11, pp. 7859–7881, Jun. 2019.

[4] S. Schwietzke et al., "Upward revision of global fossil fuel methane emissions based on isotope database," *Nature*, vol. 538, no. 7623, pp. 88–91, Oct. 2016.

[5] B. Hmiel et al., "Preindustrial $^{14}CH_4$ indicates greater anthropogenic fossil CH_4 emissions," *Nature*, vol. 578, no. 7795, pp. 409–412, Feb. 2020.

[6] I. Karakurt, G. Aydin, and K. Aydiner, "Sources and mitigation of methane emissions by sectors: A critical review," *Renew. Energy*, vol. 39, no. 1, pp. 40–48, Mar. 2012.

[7] Z. Drezner and H. W. Hamacher, *Facility Location: Applications and Theory*. Cham, Switzerland: Springer, 2004.

[8] Y. Cao et al., "Optimal energy management for multi-microgrid under a transactive energy framework with distributionally robust optimization," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 599–612, Sep. 2022.

[9] Z. Ling, F. Hu, H. Zhang, and Z. Han, "Age-of-information minimization in healthcare IoT using distributionally robust optimization," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 16154–16167, Sep. 2022.

[10] M. Younis and K. Akkaya, "Strategies and techniques for node placement in wireless sensor networks: A survey," *Ad Hoc Netw.*, vol. 6, no. 4, pp. 621–655, Jun. 2008.

[11] Y. Liu, H. Zhang, B. Di, J. Wu, and Z. Han, "Deployment for high altitude platform systems with perturbation: Distributionally robust optimization approach," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1126–1130, May 2022.

[12] J. W. Berry, L. Fleischer, W. E. Hart, C. A. Phillips, and J.-P. Watson, "Sensor placement in municipal water networks," *J. Water Resour. Planning Manage.*, vol. 131, no. 3, pp. 237–243, 2005.

[13] A. Ahmadzadeh, J. Keller, G. Pappas, A. Jadbabaie, and V. Kumar, "An optimization-based approach to time-critical cooperative surveillance and coverage with UAVs," in *Experimental Robotics*. Cham, Switzerland: Springer, 2008, pp. 491–500.

[14] M. Karatas, N. Razi, and H. Tozan, "A comparison of p-median and maximal coverage location models with Q–coverage requirement," *Proc. Eng.*, vol. 149, pp. 169–176, Dec. 2016.

[15] S. W. Legg et al., "A stochastic programming approach for gas detector placement using CFD-based dispersion simulations," *Comput. Chem. Eng.*, vol. 47, pp. 194–201, Dec. 2012.

[16] A. J. Benavides-Serrano, M. S. Mannan, and C. D. Laird, "Optimal placement of gas detectors: A P-median formulation considering dynamic nonuniform unavailabilities," *AIChE J.*, vol. 62, no. 8, pp. 2728–2739, Aug. 2016.

[17] K. A. Klise, B. L. Nicholson, C. D. Laird, A. P. Ravikumar, and A. R. Brandt, "Sensor placement optimization software applied to site-scale methane-emissions monitoring," *J. Environ. Eng.*, vol. 146, no. 7, Jul. 2020, Art. no. 04020054.

[18] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renew. Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, May 2009.

[19] A. Roy, S. Datchanamoorthy, S. Nundy, B. Keely, O. Kim, and G. Chan, "Fugitive emission monitoring system using land-based sensors for industrial applications," in *Proc. Abu Dhabi Int. Petroleum Exhib. Conf.*, Abu Dhabi, UAE, Nov. 2021.

[20] A. P. Ravikumar, J. Wang, and A. R. Brandt, "Are optical gas imaging technologies effective for methane leak detection?" *Environ. Sci. Technol.*, vol. 51, no. 1, pp. 718–724, Jan. 2017.

[21] X. Zhou, V. Amaral, and J. D. Albertson, "Source characterization of airborne emissions using a sensor network: Examining the impact of sensor quality, quantity, and wind climatology," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 4621–4629.

[22] K. Zheng et al., "Near-infrared off-axis integrated cavity output spectroscopic gas sensor for real-time, in situ atmospheric methane monitoring," *IEEE Sensors J.*, vol. 21, no. 5, pp. 6830–6838, Dec. 2020.

[23] A. F. Stein, R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. Cohen, and F. Ngan, "NOAA's HYSPLIT atmospheric transport and dispersion modeling system," *Bull. Amer. Meteorol. Soc.*, vol. 96, no. 2, pp. 2059–2077, Dec. 2016.

[24] F. Lin, X. Fang, and Z. Gao, "Distributionally robust optimization: A review on theory and applications," *Numer. Algebra, Control Optim.*, vol. 12, no. 1, p. 159, Mar. 2022.

[25] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Oper. Res.*, vol. 58, no. 3, pp. 595–612, Jun. 2010.

[26] S. Wang, C. Zhao, L. Fan, and R. Bo, "Distributionally robust unit commitment with flexible generation resources considering renewable energy uncertainty," *IEEE Trans. Power Syst.*, early access, Feb. 8, 2022, doi: 10.1109/TPWRS.2022.3149506.

[27] Z. Ling, F. Hu, Y. Zhang, L. Fan, F. Gao, and Z. Han, "Distributionally robust chance-constrained backscatter communication-assisted computation offloading in WBANs," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3395–3408, May 2021.

[28] K. A. Klise, B. L. Nicholson, and C. D. Laird, "Sensor placement optimization using Chama," Sandia Nat. Lab (SNL-NM), Albuquerque, NM, USA, Tech. Rep., Oct. 2017.

[29] C. Zhao and Y. Guan, "Data-driven stochastic unit commitment for integrating wind generation," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2587–2596, Jul. 2016.

[30] E. R. Group et al., "City of fort worth natural gas air quality study," Eastern Res. Group, Morrisville, NC, USA, Tech. Rep., Jul. 2011. [Online]. Available: https://www.scribd.com/document/60046362/Air-Quality-Study-Final

[31] NOAA. *NOAA Automated Surface Observing System (ASOS)*. Accessed: Apr. 19, 2022. [Online]. Available: https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/automated-surface-observing-system-asos

[32] K. L. McInnes, T. A. Erwin, and J. M. Bathols, "Global climate model projected changes in 10 m wind speed and direction due to anthropogenic climate change," *Atmos. Sci. Lett.*, vol. 12, no. 4, pp. 325–333, Oct. 2011.

[33] W. F. Dabberdt and E. Miller, "Uncertainty, ensembles and air quality dispersion modeling: Applications and challenges," *Atmos. Environ.*, vol. 34, no. 27, pp. 4667–4673, Jan. 2000.

[34] W. E. Hart et al., *Pyomo-Optimization Modeling in Python*, vol. 67. Cham, Switzerland: Springer, 2017.

[35] A. Makhorin. *GLPK (GNU Linear Programming Kit)*. Accessed: Apr. 19, 2022. [Online]. Available: https://www.gnu.org/software/glpk

[36] E. D. Sherwin, Y. Chen, A. P. Ravikumar, and A. R. Brandt, "Single-blind test of airplane-based hyperspectral methane detection via controlled releases," *Elementa, Sci. Anthropocene*, vol. 9, no. 1, Mar. 2021, Art. no. 00063. [Online]. Available: https://doi.org/10.1525/elementa.2021.00063

[37] E. D. Sherwin et al., "Single-blind validation of space-based point-source methane emissions detection and quantification," EarthArXiv, 2022. [Online]. Available: https://eartharxiv.org/repository/view/3465/, doi: 10.31223/X5DH09.

[38] T. Stocker, *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

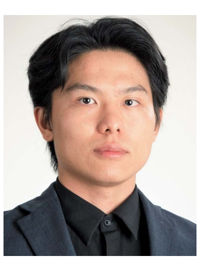
[39] H. Chen et al., "High-accuracy continuous airborne measurements of greenhouse gases (CO_2 and CH_4) using the cavity ring-down spectroscopy (CRDS) technique," *Atmos. Meas. Techn.*, vol. 3, no. 2, pp. 375–386, 2010.

[40] S. N. Riddick et al., "The calibration and deployment of a low-cost methane sensor," *Atmos. Environ.*, vol. 230, Jun. 2020, Art. no. 117440.

[41] S. Connors et al., "Estimates of sub-national methane emissions from inversion modelling," *Atmos. Chem. Phys. Discuss.*, vol. 2018, pp. 1–19, Dec. 2018.

[42] S. N. Riddick et al., "Development of a low-maintenance measurement approach to continuously estimate methane emissions: A case study," *Waste Manage.*, vol. 73, pp. 210–219, May 2018.

[43] A. Somov, A. Baranov, D. Spirjakin, A. Spirjakin, V. Sleptsov, and R. Passerone, "Deployment and evaluation of a wireless sensor network for methane leak detection," *Sens. Actuators A, Phys.*, vol. 202, pp. 217–225, Nov. 2013.



Yuan Zi (Student Member, IEEE) received the B.S. degree in exploration technology and engineering (geophysical exploration) from the China University of Petroleum (East China), Qingdao, China, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA, co-advised by Dr. Jiefu Chen and Dr. Zhu Han.

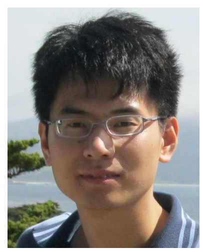
He was a Machine Learning Summer Intern with Siemens Sustainable Automation Solution (SAS), Princeton, NJ, USA, in 2021, and Shell AI, Houston, in 2022. His research interests include sustainability, machine learning, anomaly detection, inverse problem, optimization, and signal processing.

Lei Fan (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Hefei University of Technology, Hefei, China, in 2009, and the Ph.D. degree in industrial and systems engineering from the University of Florida, Gainesville, FL, USA, in 2015.

He was an Application Engineer with General Electric, Schenectady, NY, USA, and a Software Engineer with Siemens Industry, Inc., Minnetonka, MN, USA. He is currently an Assistant Professor with the University of Houston, Houston, TX, USA. His research interests include the optimal operations and planning of complex systems (e.g., power energy systems and wireless communication systems).

Xuqing Wu (Member, IEEE) received the bachelor's degree in automation from the University of Science and Technology Beijing, Beijing, China, in 1995, and the Ph.D. degree in computer science from the University of Houston, Houston, TX, USA, in 2011.

He is currently an Associate Professor at the University of Houston. Prior to joining the University of Houston, in 2015, he was a Data Scientist and a Software Engineer in the energy and IT industry. His research interests include machine learning, probabilistic modeling, and their applications in subsurface sensing.



Jiefu Chen (Senior Member, IEEE) received the B.S. degree in engineering mechanics and the M.S. degree in dynamics and control from the Dalian University of Technology, Dalian, China, in 2003 and 2006, respectively, and the Ph.D. degree in electrical engineering from Duke University, Durham, NC, USA, in 2010.

From 2011 to 2015, he was a Staff Scientist with the Advantage Research and Development Center, Weatherford International Ltd., Houston, TX, USA. He joined the University of Houston, Houston, in 2015, where he is currently an Associate Professor of Electrical and Computer Engineering. His research interests include computational electromagnetics, inverse problems, machine learning for scientific computing, oilfield data analytics, seismic data processing, underground and underwater wireless communication, and well logging.

Zhu Han (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2003, respectively.

From 2000 to 2002, he was a Research and Development Engineer with JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, Boise, ID, USA. He is currently a John and Rebecca Moores Professor with the Department of Electrical and Computer Engineering and the Department of Computer Science Department, University of Houston, Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid.

Dr. Han has been a Fellow of American Association for the Advancement of Science (AAAS) since 2019 and a Distinguished Member of Association for Computing Machinery (ACM) since 2019. He received the NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the *Journal on Advances in Signal Processing* in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (Best Paper Award in IEEE JSAC) in 2016, and several best paper awards in IEEE conferences. He has been a 1% Highly Cited Researcher since 2017 according to Web of Science. He is the winner of the 2021 IEEE Kiyo Tomiyasu Award for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks." He was a Distinguished Lecturer of the IEEE Communications Society from 2015 to 2018.