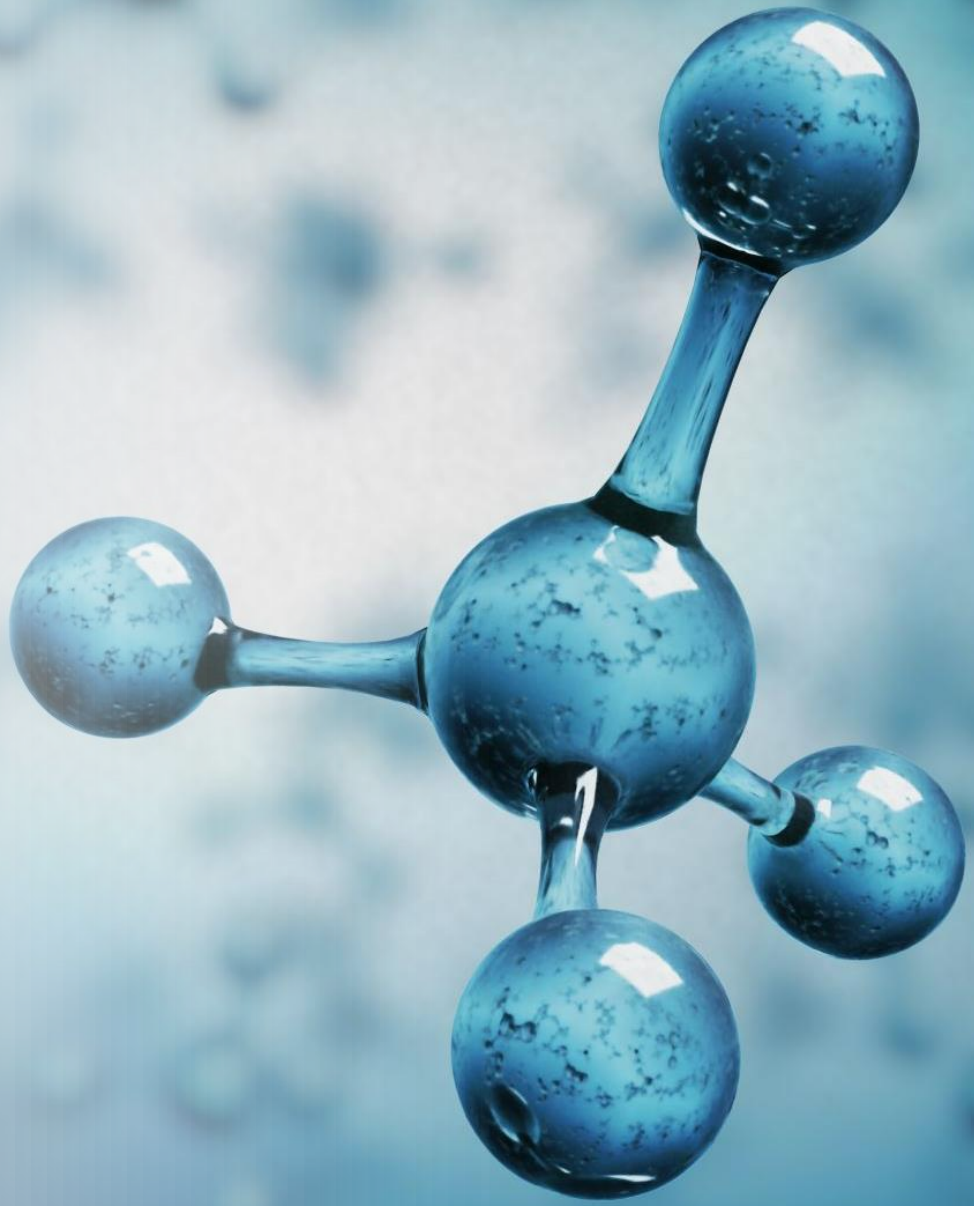
Unsupervised Machine Learning Framework for Sensor Placement Optimization: Analyzing Methane Leaks

Shirui Wang[*], Sara Malvar[*,†], Leonardo O. Nunes, Kim D. Whitehall, Yagna D. Oruganti, Yazeed Alaudah, Anirudh Badam

Microsoft



Methane Leak Detection & Remediation (LDAR) - Motivation

Microsoft

2nd

Most potent GHG.

x 84-87

20-year Global Warming Potential (GWP) than the CO_2. [1]

10.2B USD

Total Available Market (TAM) for O&G in the US

75%

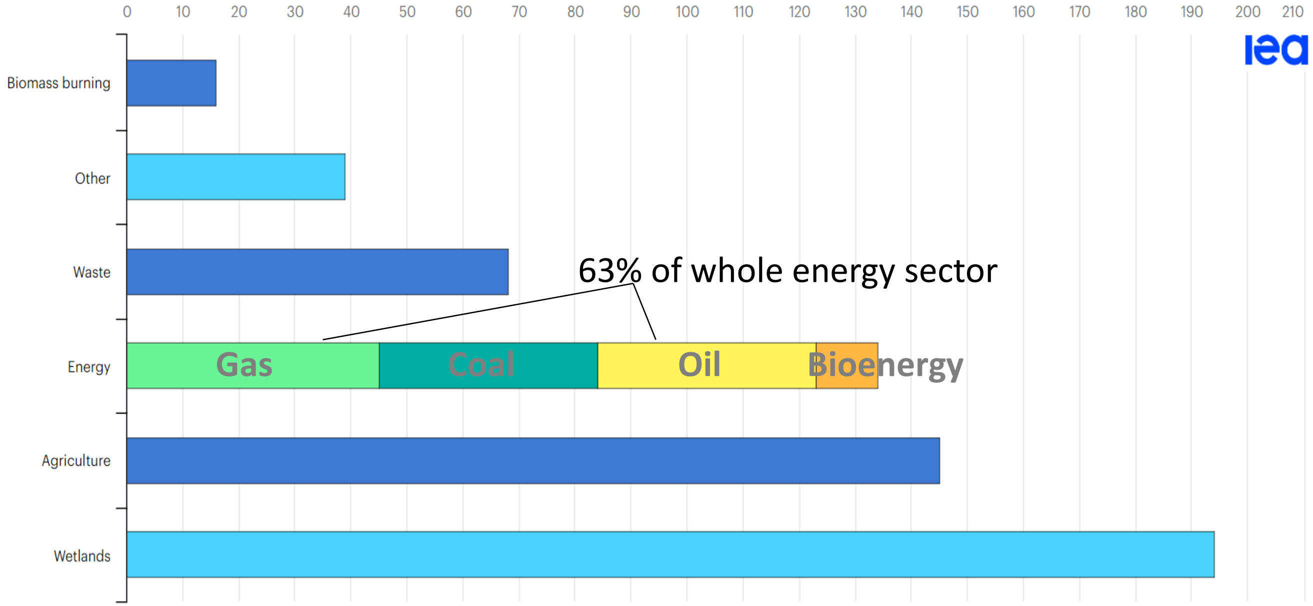
Methane emissions for O&G can be avoided. [1]

40%

With no net cost. [1]

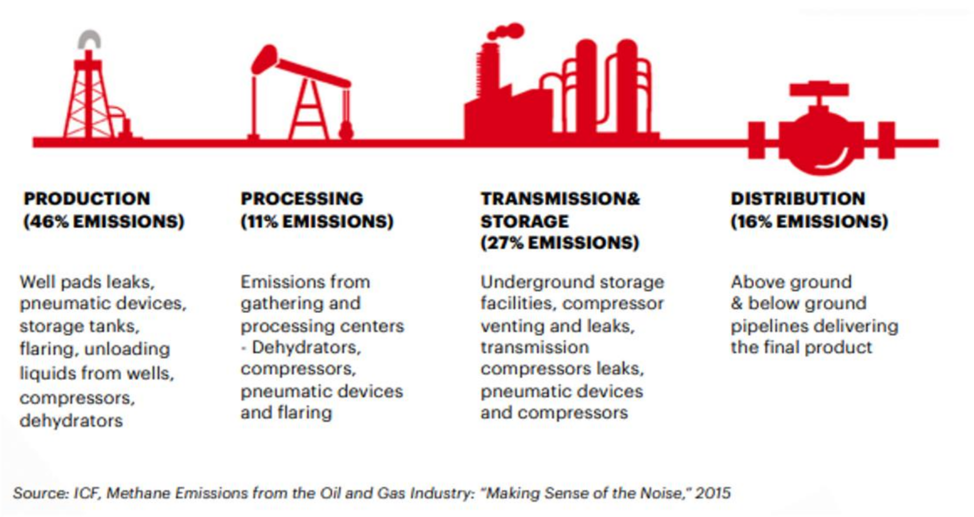
15.4GT of CO_2e

Total emissions abatement opportunity for O&G sector



[2]

Methane Leakage across the O&G Value Chain



[1] IPCC Fourth Assessment Report
[2] IEA, Sources of methane emissions, IEA, Paris https://www.iea.org/data-and-statistics/charts/sources-of-methane-emissions-2

Natural Gas Leak Detection Technologies

Microsoft

Satellite: Surveillance via satellites enables regional & global coverage over a regular period. Analysis of this data can help identify methane hotspots.

Aerial Survey: Aircraft systems conduct on-demand surveys of an area/region of interest and collect high spatial resolution measurements.

Drones: Automated drone flights along a pre-planned path collect

3D near-ground data at a regular cadence. This can be beneficial for remote locations.

Optical Gas Imaging Cameras: EPA uses OGIC evidence for regulatory compliance. Traditionally manual survey to identify leaks & sources.

Fixed Sensors: Fixed sensors provide onsite methane sensing to protect facilities through an early warning system to detect gas leaks.

Ground Sensor Grids: IoT sensor grids with data streams that can be analyzed in near real-time to accurately detect anomalous emissions, perform source attribution and undertake remediation measures.

Problem formulation

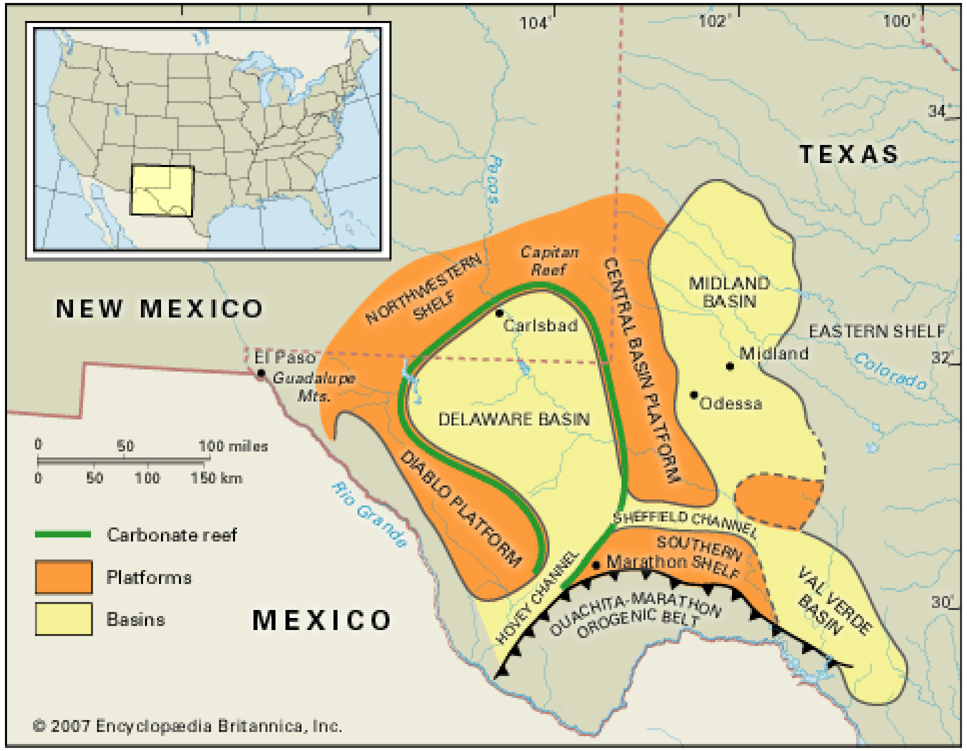
Microsoft

Dense placement (ideal) advantages

- Captures all possible leakages.
- Does not require as much environmental information, such as wind direction, as other techniques;

Dense placement disadvantages

- One of the most prolific Oil and Gas producing regions in the US, the Permian Basin, has over 250,000 km² of area.
- Ensuring sensor coverage over such a vast area can be cost-prohibitive and unrealistic due to budget constraints

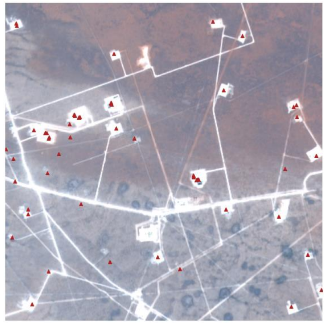


Permian Basin, is a prolific shale play in western Texas and southeastern New Mexico

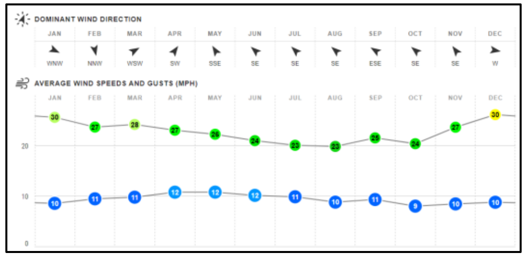
❖ Objective: Propose a sparse sensor placement strategy to capture methane leaks in an Area Of Interest (AOI) timely and accurately.

Methane Sensor Placement Optimization Workflow

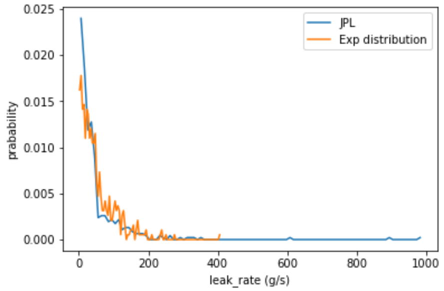
Microsoft



Oil & Gas facility map (well pads, pipelines, processing plants …)



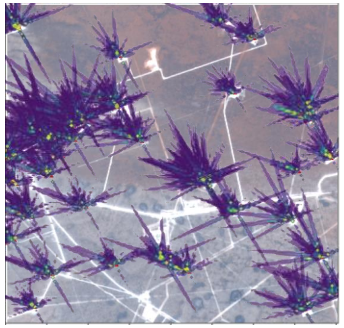
Meteorology variables (wind speed, direction, solar radiation …)



Historical Oil and Gas assets leak rates and distribution

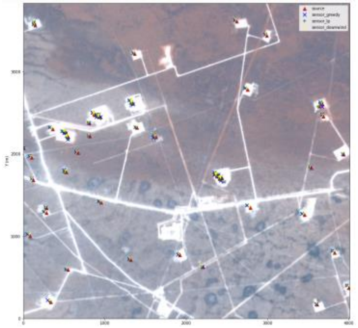
Gaussian plume model for methane dispersion

Consolidated gas presence map



Maximum coverage solver

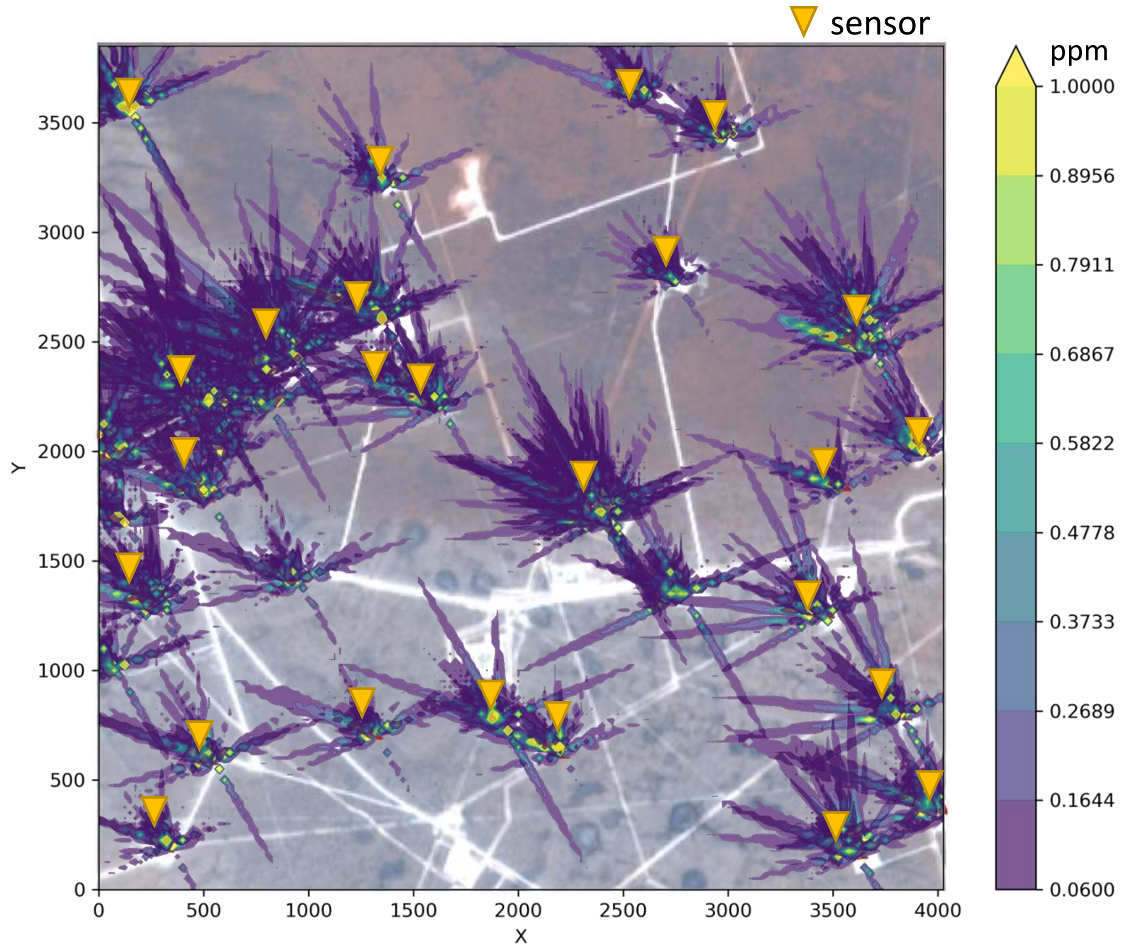
Optimal O&G methane sensor placement



Data Ingestion → Methane Dispersion Modeling → Sensor Placement Optimization

Maximum coverage problem

2D Illustration of Maximum Coverage problem for sensor placement



➢ Given possible sensor locations, find the subset that **maximizes the coverage of possible methane leakage** while **constrained by the number / budget of sensors**.

- Given the simulated methane emission map, the sensor placement optimization is formulated as a **maximum coverage problem**.
  - Given sets $S = \{S_i\}^{i=1,\ldots,N}$ and number k. S_i may contain some entity $e_i \in E$.
  - Find subset $S' = \{S_1, \ldots, S_m\} \subseteq \mathcal{S}$
  - Objective: maximize the covered elements $\left|\bigcup_{S_i \in S'} S_i\right|$, such that $|S'| \leq k$

maximize	$\sum_{e_j \in E} y_j$	(maximizing the sum of covered elements)
subject to	$\sum x_i \leq k$	(no more than k sets are selected)
	$\sum_{e_j \in S_i} x_i \geq y_j$	(if $y_j > 0$ then at least one set $e_j \in S_i$ is selected)
	$y_j \in \{0,1\}$	(if $y_j = 1$ then e_j is covered)
	$x_i \in \{0,1\}$	(if $x_i = 1$ then S_i is selected for the cover)

Current Methodologies

Microsoft

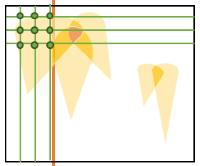
Source locs → Gaussian Plume Model → Plume Ensembles
Weather
Grid

Downwind (Baseline)

Compute the average of the wind direction → Given assigned distance & height → Sensor locations

Linear programming (LP)

Initialize sensor locations on the grid points → Get the sensor-scenario **coverage table** → **Linear programming** to maximize the coverage ratio with given sensor locations → Sensor locations [1]



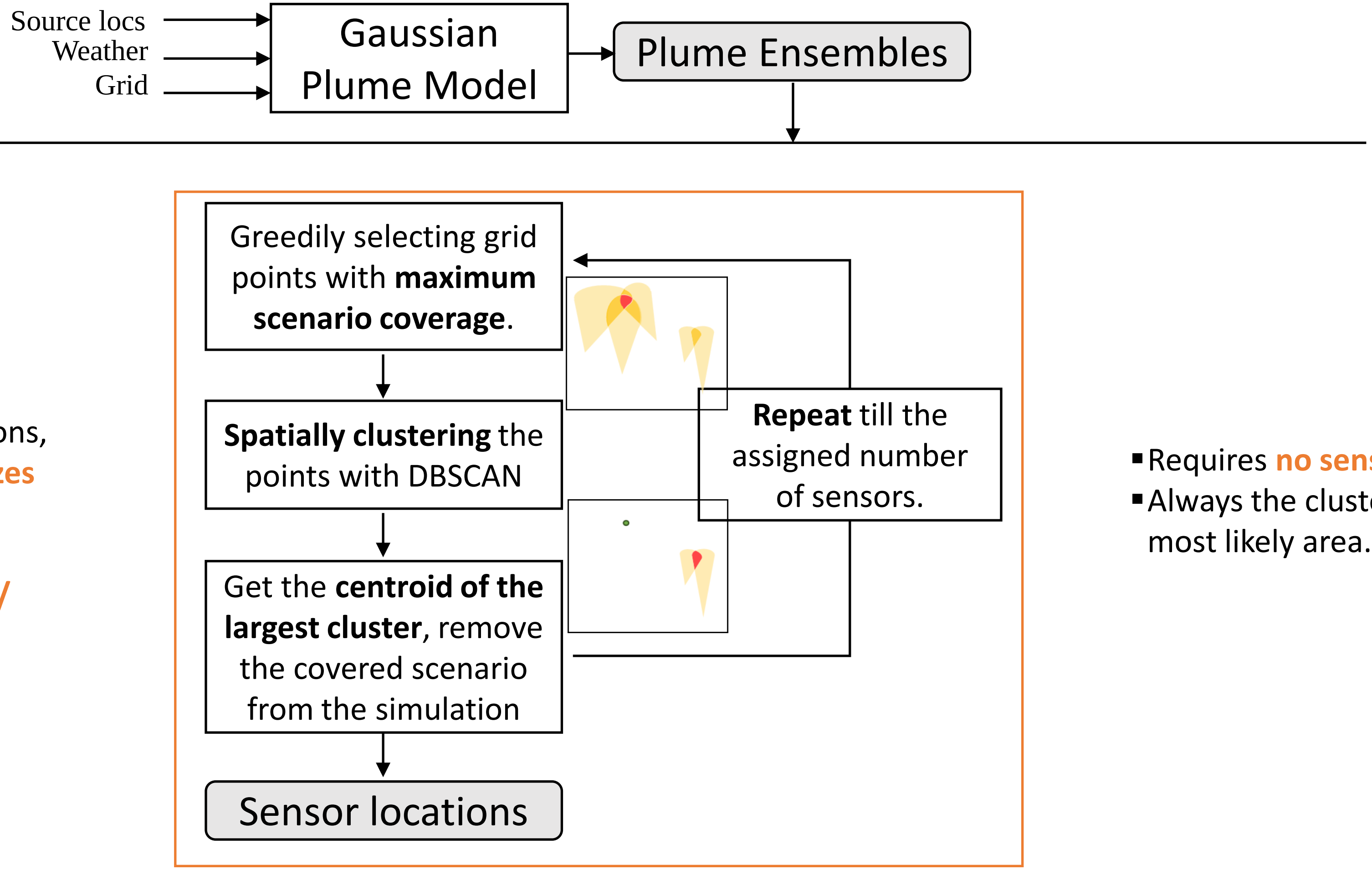
- Compute the dominant / average wind direction and place sensors at a given distance/height near the sources.
- Possible detection height and distance **could vary a lot for different leakages** under different weather conditions.

- Heavily **depending on the initialization** of sensor locations.
  - The sensors can be only put on the subset of initial positions.

[1] Klise, K.A., Nicholson, B.L., Laird, C.D., Ravikumar, A.P. and Brandt, A.R., 2020. Sensor Placement Optimization Software Applied to Site-Scale Methane-Emissions Monitoring. Journal of Environmental Engineering, 146(7), p.04020054.

Proposed Methodology

- Given possible sensor locations, find the subset that **maximizes the coverage of possible methane leakage** while **constrained by the number / budget of sensors**.



Clustering-based Greedy selection

- Requires **no sensor initialization**.
- Always the clustering center of the most likely area.

Results and Conclusion

Microsoft

- **Data ingestion pipeline** incorporating **multi-modal data** (such as organized oil & gas facilities maps, station weather data and historical methane leak rate distributions) has been built for the methane sensor placement optimization problem

- We model **methane dispersion** with the Gaussian Plume Model.

- A **new clustering-based greedy method** is proposed for sensor placement optimization.
  - It explores **spatial diversity** for sensor locations and **captures variance of methane plume dispersion over days**.
  - In one sensor for every three sources (**1:3** sensor-source number ratio) case, the proposed methodology detected **6.8%** more leaks than the baseline
  - The proposed methodology achieves **87.9%** detection rate of the CH_4 leaking sources, as apposed to the **82.8%** of the baseline, with **5.8%** improvement over the detection rate.
  - Our proposed method, in its initial iterations alone, **surpasses** or is **at par** with published literature, with potential for far greater upside.

Future Roadmap

- Improvement of **methane dispersion modeling**.
  - A more complex atmospheric dispersion model / DNN, with the inclusion of more variables such as Digital Elevation Model, gridded weather data etc.
  - Include more types of sources: pipelines, processing plants etc.
- Improve methane leak rate sampling algorithm.
- The **clustering-based greedy algorithm** is more flexible and offers potential for further refinement, such as adding constraints to potential sensor locations and optimizing for source attribution
- **Longer period dataset** is needed to capture the weather of the area for more robust sensor placement.

Thank you!

We would like to acknowledge Mirco Milletari and Fidan Boylu Uz for their contributions to this work.