UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING
Department of Electrical & Computer Engineering

Large-scale Carbon Monitoring with Data-Driven Optimization and Machine Learning-enhanced Signal Processing

Yuan Zi
Advisors: Dr. Jiefu Chen and Dr. Zhu Han

Committee Member: Dr. Jiefu Chen, Dr. Lei Fan, Dr. Zhu Han, Dr. Saurabh Prasad, Dr. Jie Zhang, Dr. Xuqing Wu

Ph.D. Dissertation Defense

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING
Department of Electrical & Computer Engineering

University of Houston
Doctor of Philosophy - PhD, Electrical and Electronics Engineering
2019 - 2024

National mineral resources university (University of mines)
Visiting Student, Exploration Geophysics
2018 - 2019

China University of Petroleum 中国石油大学(华东)
Bachelor's degree, Exploration Technology and Engineering(Exploration Geophysics)
2015 - 2019

Volunteering

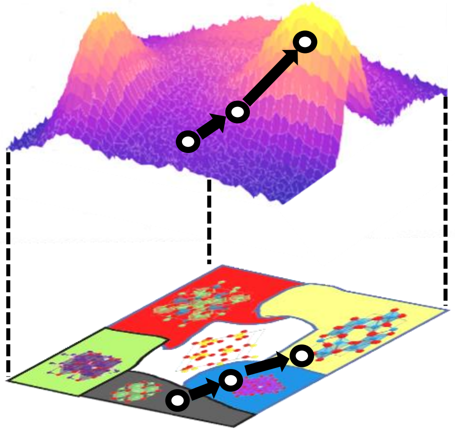
Student Fellow
Martin Trust Center for MIT Entrepreneurship
May 2022 - Present · 1 yr 6 mos
Environment

Anomaly Detection

2021 Summer
AI for industry intern
Sustainable Automation Solution Lab Princeton NJ

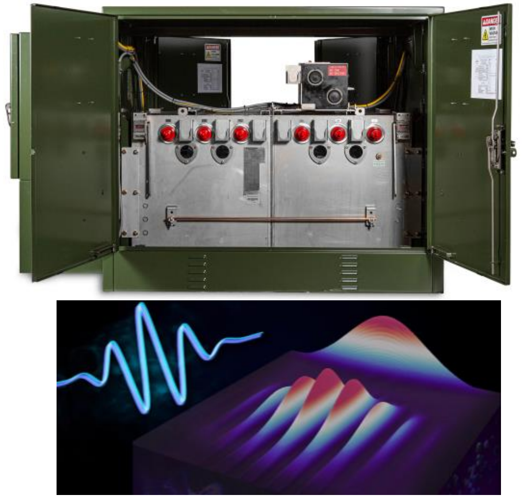
SIEMENS

Catalyst Discovery



2022 Summer
AI intern
Shell AI Houston TX

Predictive Maintenance



2023 Summer
Industrial AI intern
ABB German Research Center

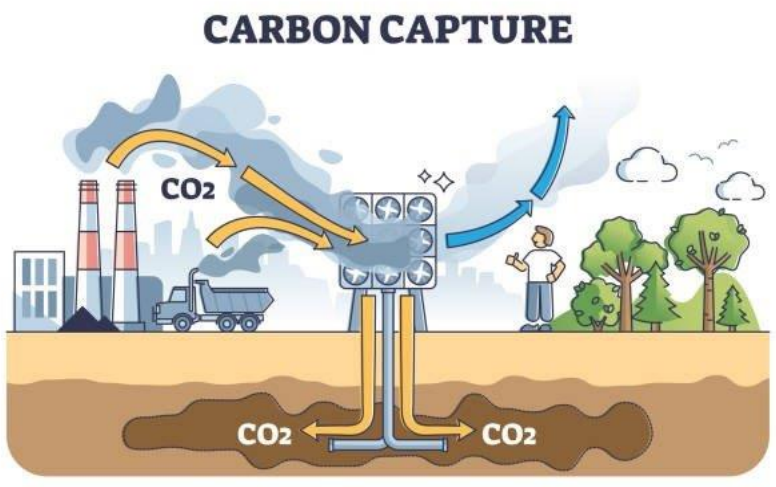
ABB

Outline

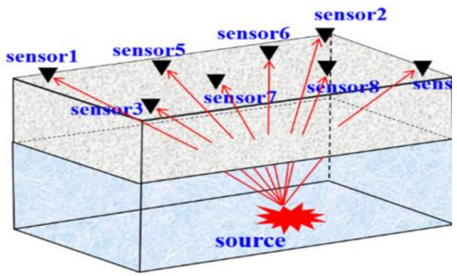
- **Introduction**
  - Carbon sequestration induced seismicity and gas leak monitoring
  - Data-driven optimization methods
  - Dissertation contributions
- **Carbon Monitoring Sensor Placement**
  - Work case 1: Stochastic optimal induced seismicity sensor placement
  - Work case 2: Distributionally robust optimal methane sensor placement
- **Stack of Machine Learning Enhanced Signal Processing Methods**
  - Denoising Work: Swell noise attenuation with self-supervised learning
  - Pattern Matching Work: Well-logging pattern matching with reinforcement learning
- **Remote Methane Monitoring**
  - Segmenting Hyperspectral Images of Methane Plumes with a Large Machine Learning Model
- **Conclusion**

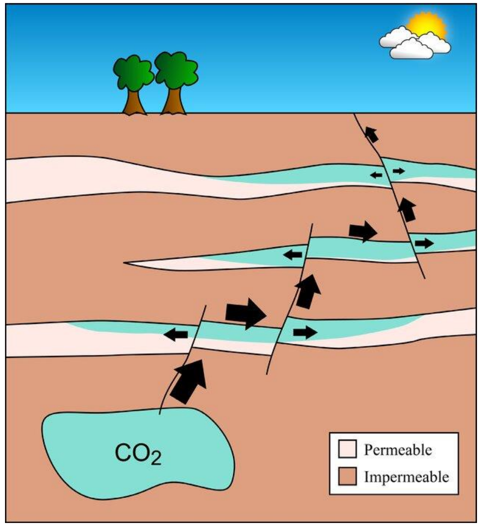
UNIVERSITY of HOUSTON
YOU ARE THE PRIDE
CULLEN COLLEGE of ENGINEERING
Department of Electrical & Computer Engineering

Carbon Sequestration
Injecting Scope

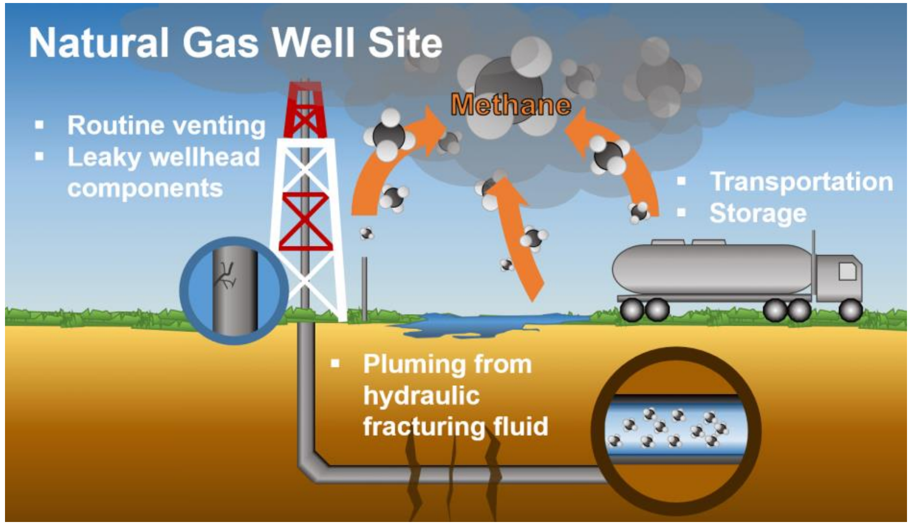


After-injecting monitoring
Induced-seismicity alarm





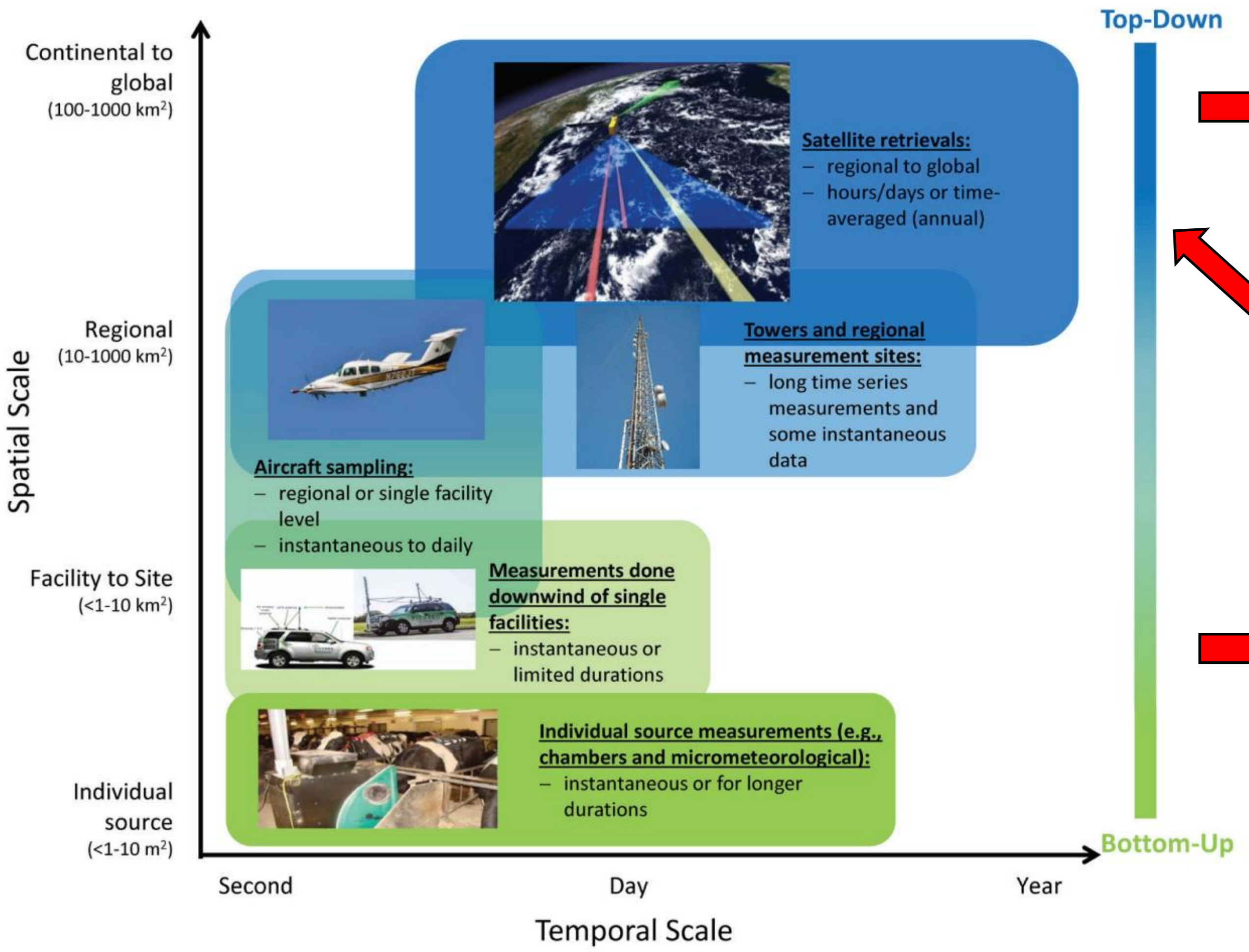
Gas Leak from a facility example
Gas Leak alarm



Ref link @https://github.com/yohanesnuwara/carbon-capture-and-storage.
https://gardaerlangga.wordpress.com/2014/07/06/well-logging-definisi-dan-sejarahnya/
https://www.cgg.com/geoscience/subsurface-imaging
https://www.mdpi.com/1424-8220/21/17/5815
https://www.researchgate.net/publication/254528542_Modeling_Leakage_Through_Faults_of_CO2_Stored_in_an_Aquifer/figures?lo=1

Carbon Monitoring Challenges

Various Sensors of Carbon Monitoring



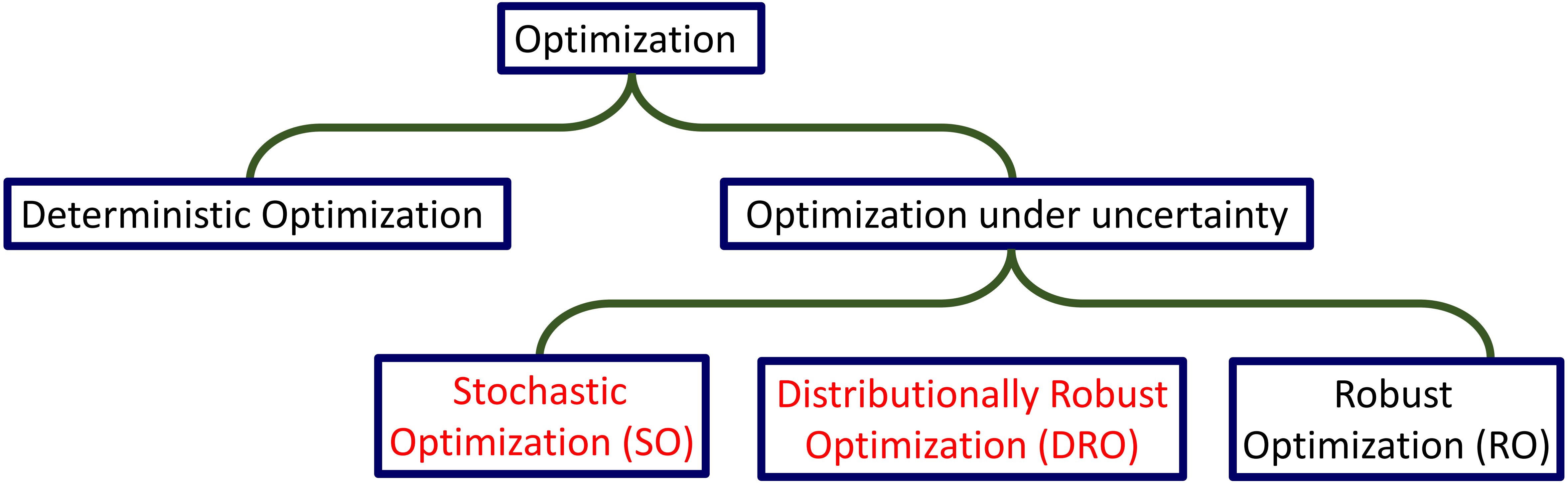
Challenge 3: Processing large-scale remote data with regular methods that need human interaction involved is inefficient and hard to scale up to the global scale

Challenge 1: Design a network to cover the monitoring of field scale with grid sensor network needs massive regular point sensors which cost budget issue.

Challenge 2: Leakage is uncertain in terms of large-scale potential leak locations and environment conditions

Data-driven Optimization Methods

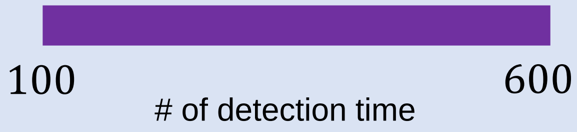
- Optimization Classification



Robust Optimization

$$\min_{x\in\chi}\max_{\xi\in\mathcal{P}} h(x,\xi)$$

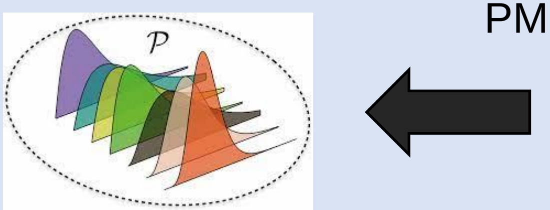
- **Probability distribution** of random parameters is **unknown,** but the **range** is **known**
- Obj. function : find a decision x that minimizes the worst-case cost over an uncertainty set

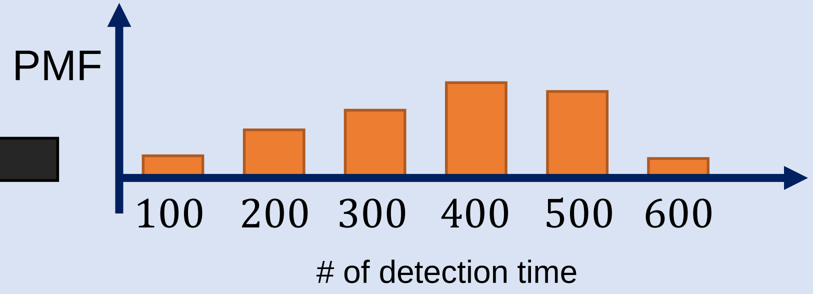


Stochastic Optimization

$$\min_{x\in\chi} E_p[h(x,\xi)]$$

- **Probability distribution** of random parameters is **known**
- Obj. function : find a decision x that minimizes a functional of the expected cost.



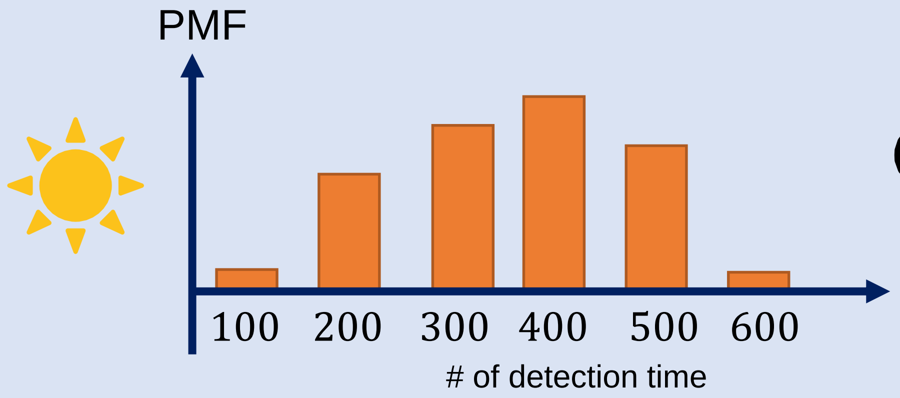


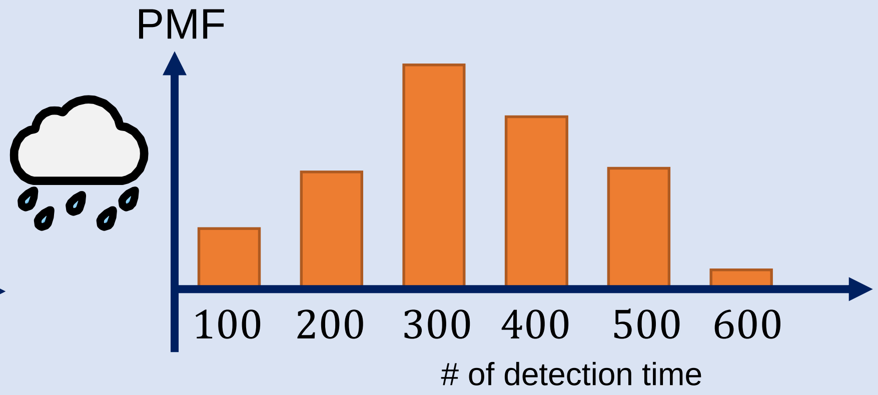
Distributionally Robust Optimization

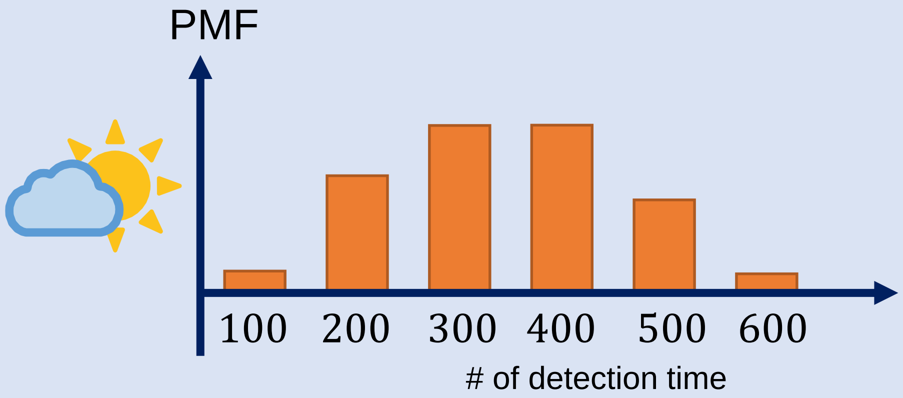
$$\min_{x\in\chi}\max_{P\in\mathcal{P}} E_P[h(x,\xi)]$$

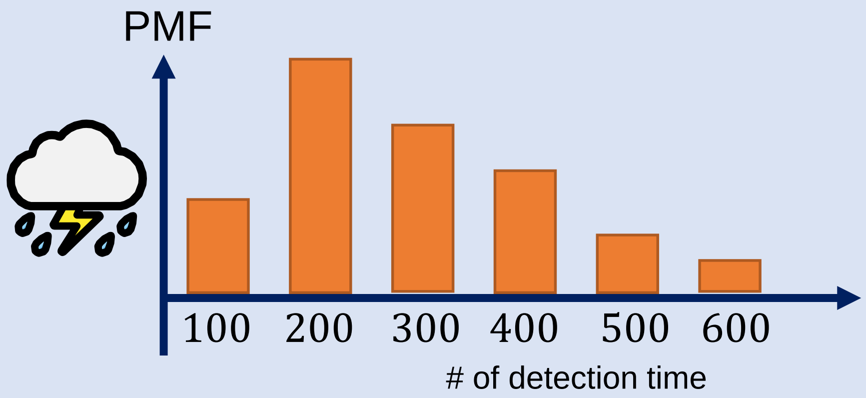
$$\min_{x\in\chi}\max_{P\in\mathcal{P}} E_P[h(x,\xi)]$$

- **Random parameters** are **uncertain**
- **Probability distribution** of random parameters is **uncertain**
- Obj. function : find a decision x that minimizes the worst-case expected cost over an uncertainty set









Contributions

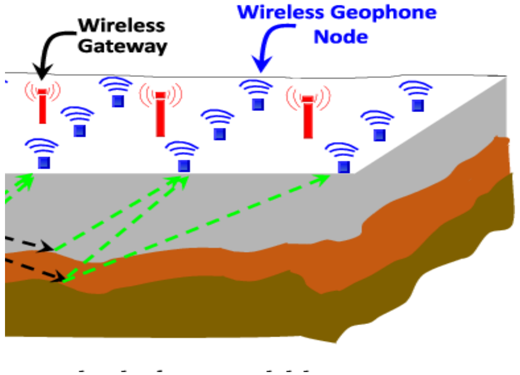
- **Contribution I**: Developed a mathematical framework for optimizing carbon monitoring sensor placement with stochastic programming to address leak source uncertainty.
- **Contribution II**: Enhanced the sensor placement framework using distributionally robust optimization to tackle both leak source and environmental uncertainties.
- **Contribution III**: Advanced carbon monitoring capabilities by integrating machine learning techniques for improved signal processing and data accuracy in variable conditions.

Outline

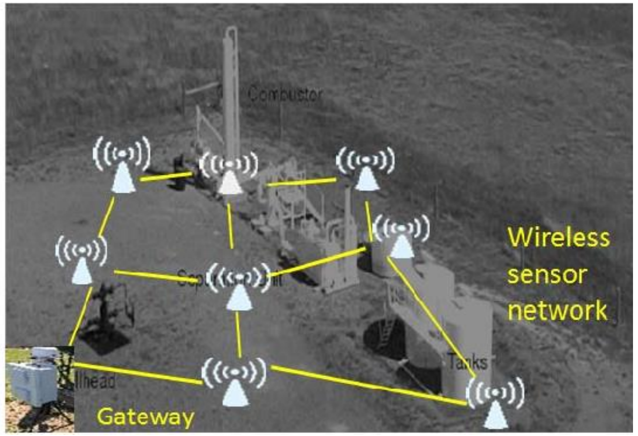
- **Introduction**
  - Carbon sequestration induced seismicity and gas leak monitoring
  - Data-driven optimization methods
  - Dissertation contributions
- **Carbon Monitoring Sensor Placement**
  - Work case 1: Stochastic optimal induced seismicity sensor placement
  - Work case 2: Distributionally robust optimal methane sensor placement
- **Stack of Machine Learning Enhanced Signal Processing Methods**
  - Denoising Work: Swell noise attenuation with self-supervised learning
  - Pattern Matching Work: Well-logging pattern matching with reinforcement learning
- **Remote Methane Monitoring**
  - Segmenting Hyperspectral Images of Methane Plumes with a Large Machine Learning Model
- **Conclusion**

UNIVERSITY of HOUSTON
YOU ARE THE PRIDE
CULLEN COLLEGE of ENGINEERING
Department of Electrical & Computer Engineering

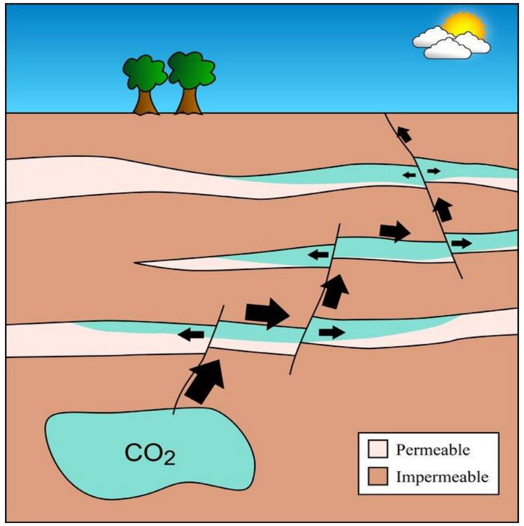
Monitoring Sensor Placement

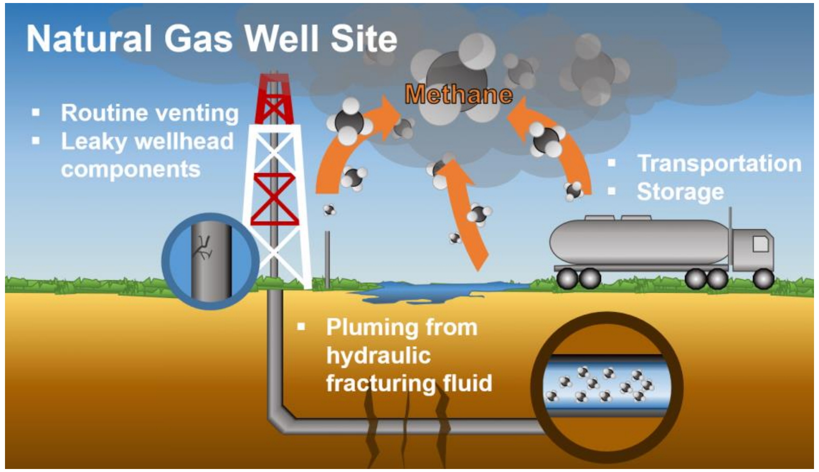


Geophone placement
Induced-seismicity alarm



Methane sensor placement
Gas Leak alarm

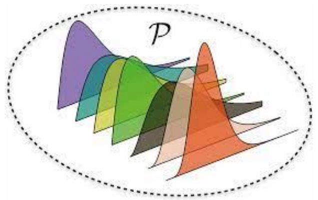




Ref link @https://www.semanticscholar.org/paper/Wireless-Geophone-Sensing-System-for-Real-Time-Data-Attia-Gaya/03eaf9343fda2feed26f3dc31680ad3c7dd537b7
https://chama.readthedocs.io/en/latest/overview.html
https://www.semanticscholar.org/paper/Wireless-Sensor-Networks-for-Fugitive-Methane-in-Klein-Ramachandran/bd5dc788f349d1dc6a20ffb9984c2419a14d391c/figure/1

Robust Optimization (RO)

$$\min_{x\in\chi}\max_{\xi\in U} h(x,\xi)$$



x -- Decision variables

ξ -- Uncertain Parameter

χ -- Convex set of feasible solutions

U -- Uncertain set

$h(x,\xi)$ -- Objective function in x that depends on parameters ξ

Benefits:

Be able to solve optimization problem with uncertainty

Choose an intermediate approach to obtain a robust form of distributed optimization problem ***(DRO):***

Stochastic Optimization (SO) $\min_{x\in\chi} E_P[h(x,\xi)]$ $\Longrightarrow$ $\min_{x\in\chi}\max_{P\in\mathcal{P}} E_P[h(x,\xi)]$ **Distributionally Robust Optimization (DRO)**

$\mathcal{P}$ is an uncertain set of probability distributions constructed from the samples.

Motivation

- In many situations, we have an empirical estimate of the underlying probability distributions.
- A natural way to hedge against the distributional ambiguity is to consider a neighborhood of the empirical probability distribution

Discrepancy

Ambiguity sets based on probability distance:

$$\mathcal{P} = \{P: d(\widehat{P_N}, P) \le \varepsilon\}$$

$\widehat{P_N}$ -- Empirical probability
ε -- Radius
$d(\widehat{P_N}, P)$ -- ***Metric*** of the similarity of two distributions

By selecting a suitable *metric*, certain *infinite-dimensional* convex DRO problems can be transformed into *finite-dimensional* convex optimization problems

Is there a metric that is simple to calculate and suitable for discrete / continuous distributions? → **Wasserstein distance**

Wasserstein distance

used to measure the distance between two distributions.

Definition:

$$d_W(P_1, P_2) = \inf_{\gamma \sim \prod(P_1, P_2)} \mathbb{E}_{(x,y)\sim\gamma}\,[||x - y||]$$

$\prod(P_1, P_2)$: the set of all possible joint distributions of P_1 and P_2.

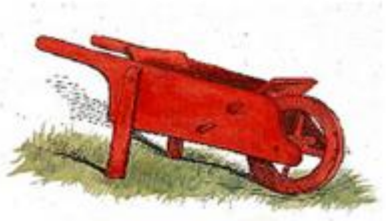
$(x, y)\sim\gamma$: samples under joint distribution γ

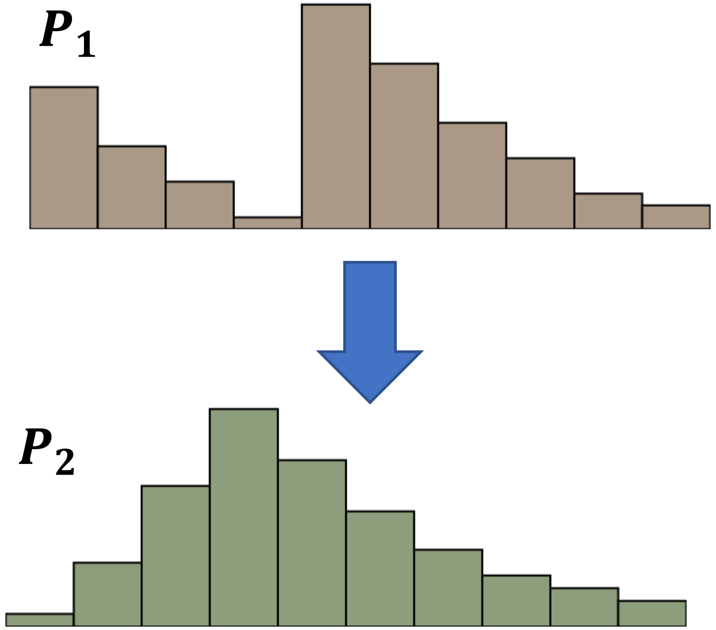
$||x - y||$: sample distance

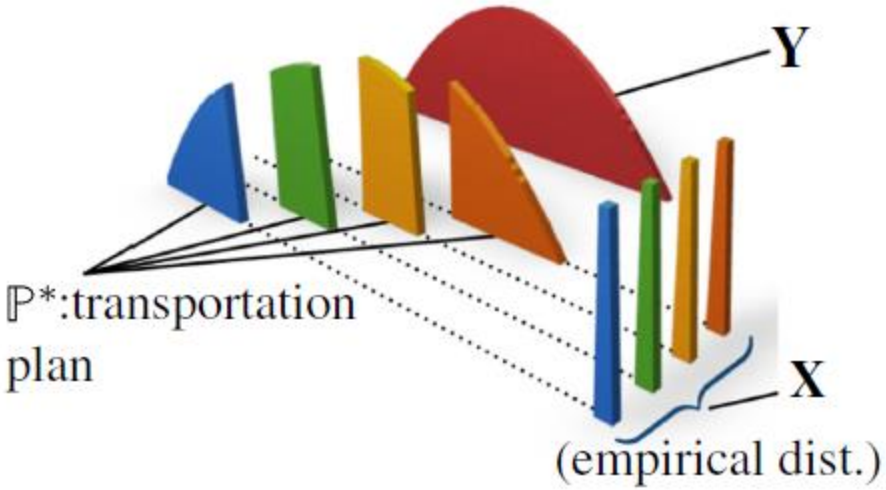
$\mathbb{E}_{(x,y)\sim\gamma}[||x - y||]$: expectation of distance for sample x and y under joint distribution γ

Wasserstein distance of P_1 and P_2: the lower bound of this expectation.

Wasserstein distance







Move mass P_1 into the shape and position of P_2.

$\prod(P_1, P_2)$: transportation plan
$\|x - y\|$: distance the soil moves
$Y(x, y)$: amount of moving soil from x to y
$\mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|]$: bulldozing cost

Bulldozing cost : amount of moving soil multiplied by the distance the soil moves.

Wasserstein distance: the smallest bulldozing cost from P_1 to P_2.

Wasserstein distance

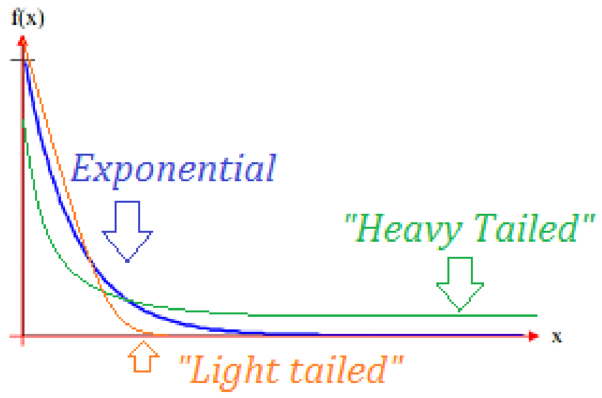
Wasserstein distance-based ambiguity set:

$$\mathbb{B}_{\varepsilon}\left(\widehat{P_N}\right) = \{Q\colon d_W\ (\widehat{P_N}, Q) \leq \varepsilon\}$$

- The ambiguity set Q can be viewed as a Wasserstein ball which contains all probability distributions whose Wasserstein distance to the empirical distribution $\widehat{P_N}$ is less than ε.
- Q will cover the true distribution with a higher probability with a larger value of ε.
    - There exists a trade-off between the accuracy and the complexity
    - It is important to well design the value of ε

Wasserstein distance

How to calculate ε of ambiguity set



Light-tailed distribution assumption: Distribution $\mathbb{P}$ is call light-tailed i $a > 1$ such that

$$\mathbb{E}^{\mathbb{P}}\left[\exp(\|\xi\|^a)\right] = \int_{\Xi} \exp(\|\xi\|^a)\, \mathbb{P}(\mathrm{d}\xi) < \infty.$$

The assumption guarantees that the ambiguity set can cover most of the possible distributions.

Radius selection: Number of samples → Confidence level → Radius selection

$$\varepsilon_N(\beta) := \begin{cases} \left(\frac{\log(c_1\beta^{-1})}{c_2 N}\right)^{1/\max\{m,2\}} & \text{if } N \geq \frac{\log(c_1\beta^{-1})}{c_2}, \\ \left(\frac{\log(c_1\beta^{-1})}{c_2 N}\right)^{1/a} & \text{if } N < \frac{\log(c_1\beta^{-1})}{c_2}. \end{cases}$$

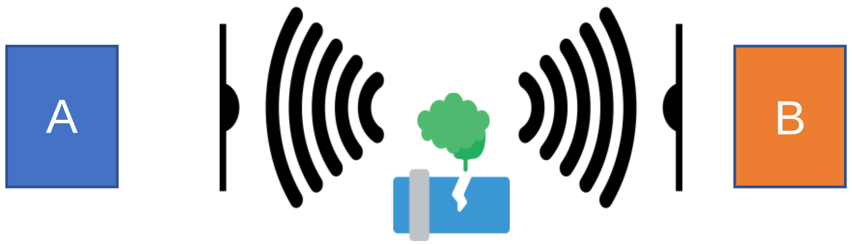
Number of samples

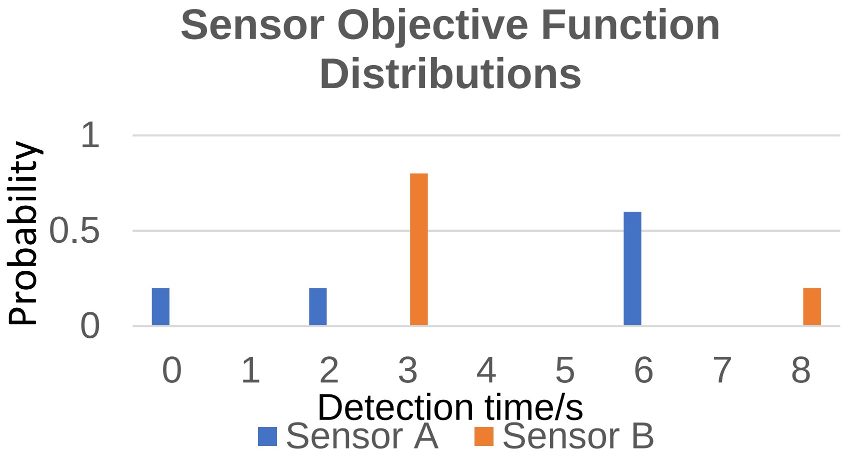
Sensor Placement under uncertainty example

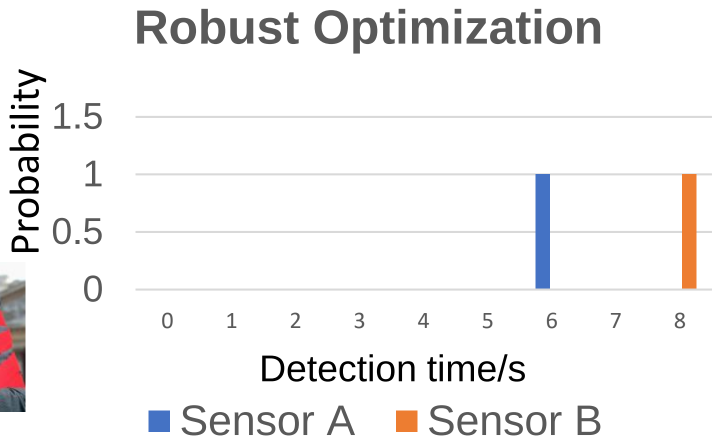
Historical detection time data for sensor A/B

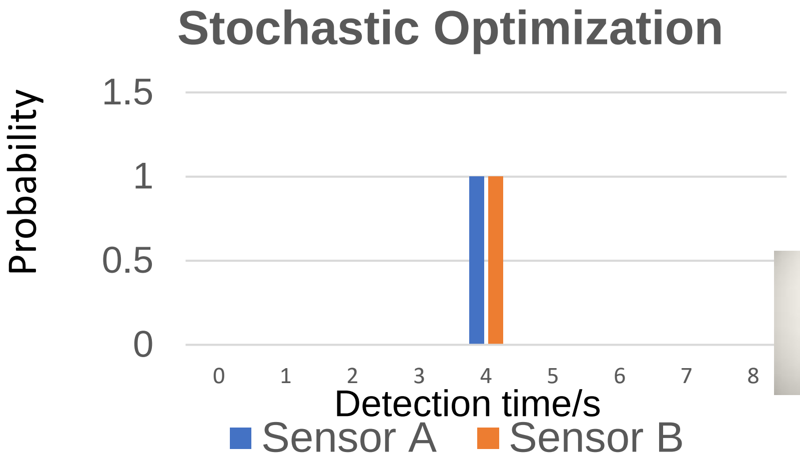
Faster is better

Scenario: 1 leak source, 2 sensor candidates

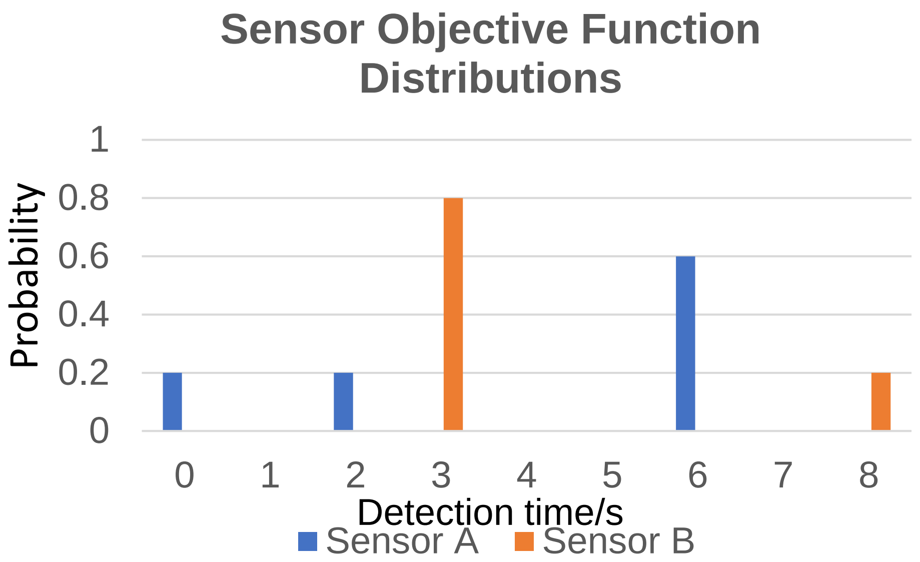


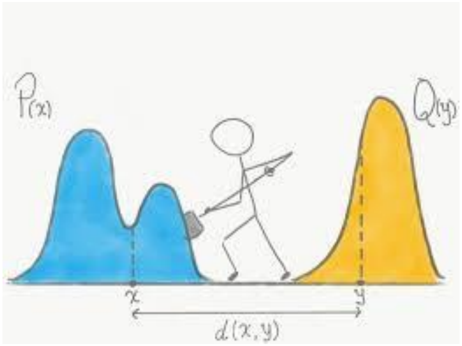


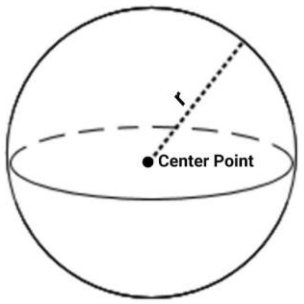


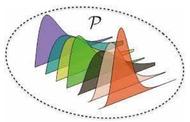


Sensor Placement under Uncertainty Example





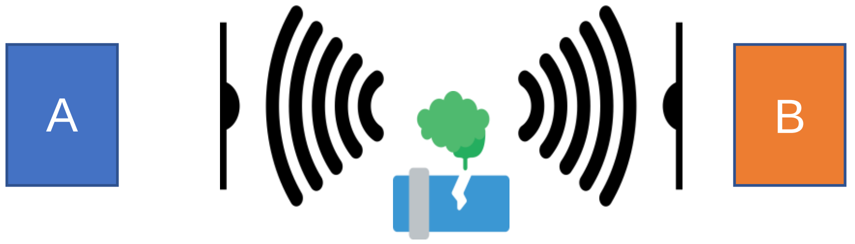


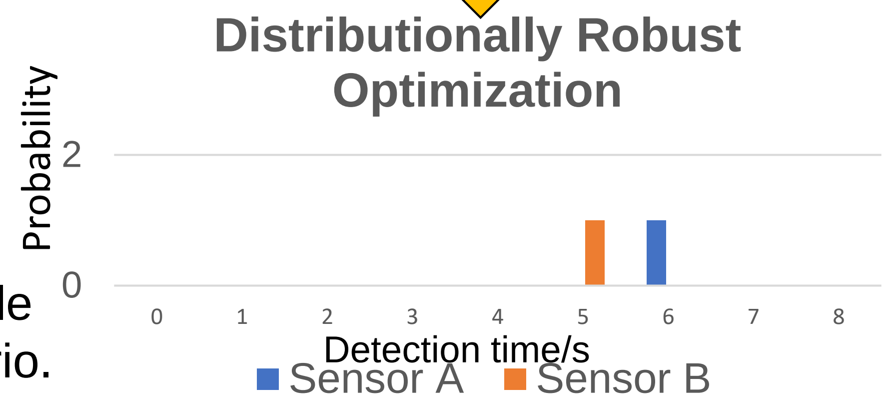


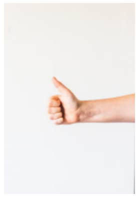
Uncertainty level = 2 (Confidence level = 0.9)

$$2 = \frac{1}{5}|d_{\text{Aworst}-0}| + \frac{1}{5}|d_{\text{Aworst}-2}| + \frac{3}{5}|d_{\text{Aworst}-6}|$$

$$2 = \frac{1}{5}|6-0| + \frac{1}{5}|6-2| + \frac{3}{5}|6-6|$$







With DRO method we can chose the right sensor while other robust optimization method cannot in this scenario.

Problem Formulation 1: Optimal Detection Time

$$\min_{y_l} \sum_{e \in \mathcal{E}_e} p_e \sum_{i \in L_e} d'_{e,i} x_{e,i},$$

Objective: Detection time Expectation

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

Binary Decision Variable

$$\sum_{l \in L} c_i y_l \leq c,$$

Cost Constrain

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

$$\sum_{i \in L_e} x_{e,i} = 1 \quad \forall\, e \in \mathcal{E},$$

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

Supplementary constraints for decision variable define

TABLE I: Tables of Symbols

Symbol	**Meaning**
$e \in \mathcal{E}$	The collection of all events.
L	The group of all potential sensors.
L_e	The set of sensors that can detect event e
p_e	The probability of event e taking place
$d_{e,i}$	The damage coefficient for passive-seismic event e at location i
$x_{e,i}$	Binary variable indicating if location i first detects event e
y_l	Binary variable indicating sensor presence at location l
c_i	The expense associated with sensor i
c	The allocation of funds for the sensors
s_e	passive-seismic source spatial location for event e
$\hat{s}_e$	Predicted passive-seismic source spatial location for event e with heuristic optimization
s'_e	passive-seismic source spatial location ground truth for the event e
t_{obs}	Observation of passive-seismic wave-arrival time.
t	Measurement of passive-seismic wave-arrival time.
$\mathcal{F}$	Ray tracing forward operator.
v	Subsurface velocity model.

Problem Formulation 2: Optimal Sources Localization

$$\min_{x_i} \sum_{e \in \mathcal{E}_e} p_e MSE(s'_e, \hat{s}_e),$$

Objective: Localization Accuracy Expectation

$$\hat{s}_e = \arg\min_{s_e} MSE(t_{obs}, t),$$

Localization Inversion Problem

$$t = \mathcal{F}(y, v, s_e),$$

Forward

$$\sum_{i \in I} c_i y_i \leq c,$$

$$y_i \in \{0, 1\} \quad \forall\, i \in I.$$

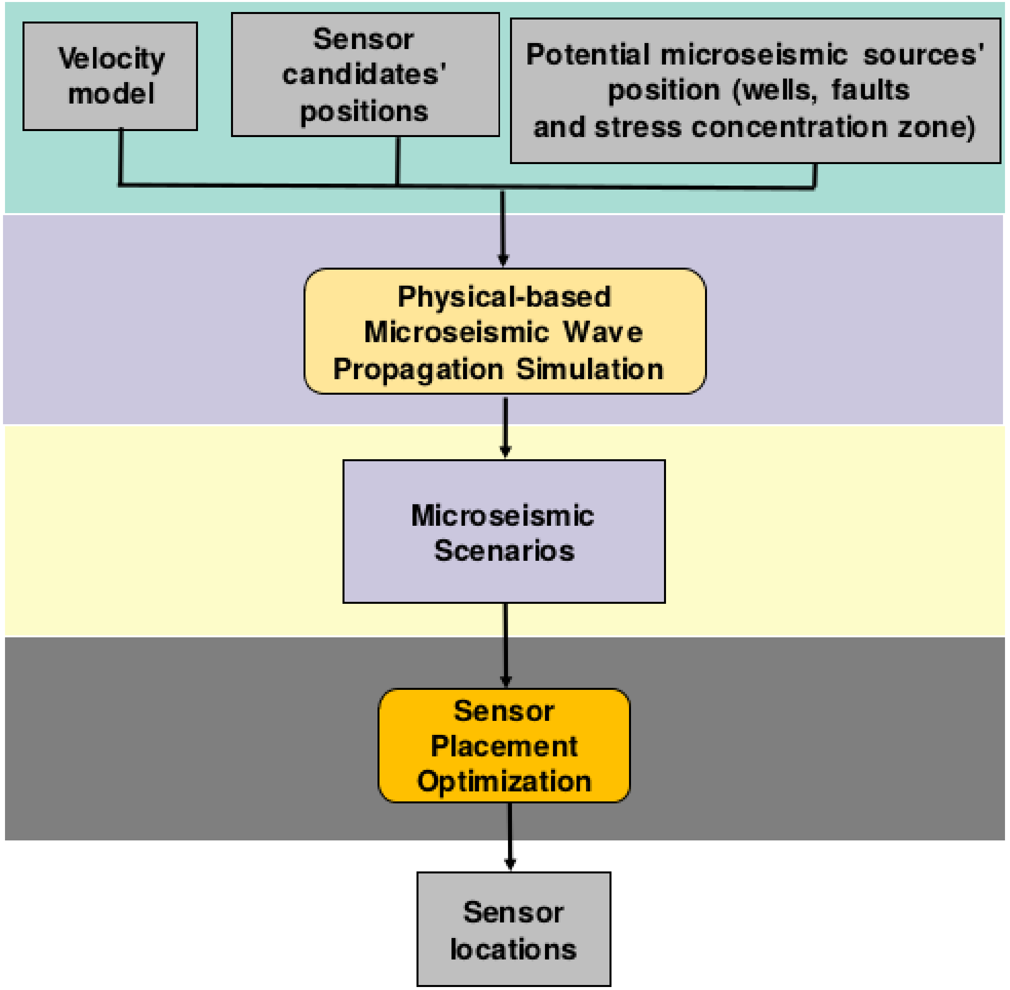
Cost constrain and
decision variable define

TABLE I: Tables of Symbols

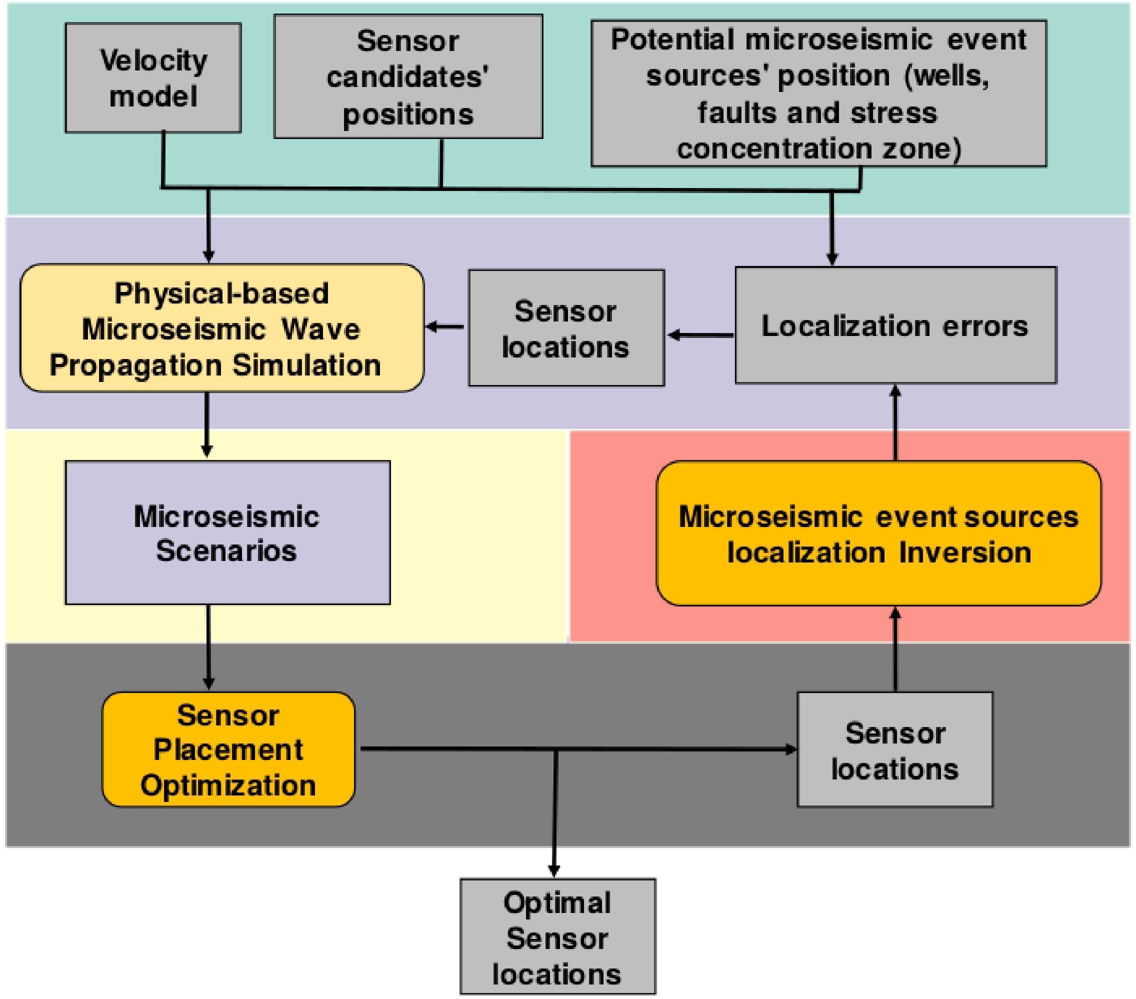
Symbol	**Meaning**
$e \in \mathcal{E}$	The collection of all events.
L	The group of all potential sensors.
L_e	The set of sensors that can detect event e
p_e	The probability of event e taking place
$d_{e,i}$	The damage coefficient for passive-seismic event e at location i
$x_{e,i}$	Binary variable indicating if location i first detects event e
y_l	Binary variable indicating sensor presence at location l
c_i	The expense associated with sensor i
c	The allocation of funds for the sensors
s_e	passive-seismic source spatial location for event e
$\hat{s}_e$	Predicted passive-seismic source spatial location for event e with heuristic optimization
s'_e	passive-seismic source spatial location ground truth for the event e
t_{obs}	Observation of passive-seismic wave-arrival time.
t	Measurement of passive-seismic wave-arrival time.
$\mathcal{F}$	Ray tracing forward operator.
v	Subsurface velocity model.

Passive-seismic Sensor Placement Optimization Workflow

Detection time optimization workflow



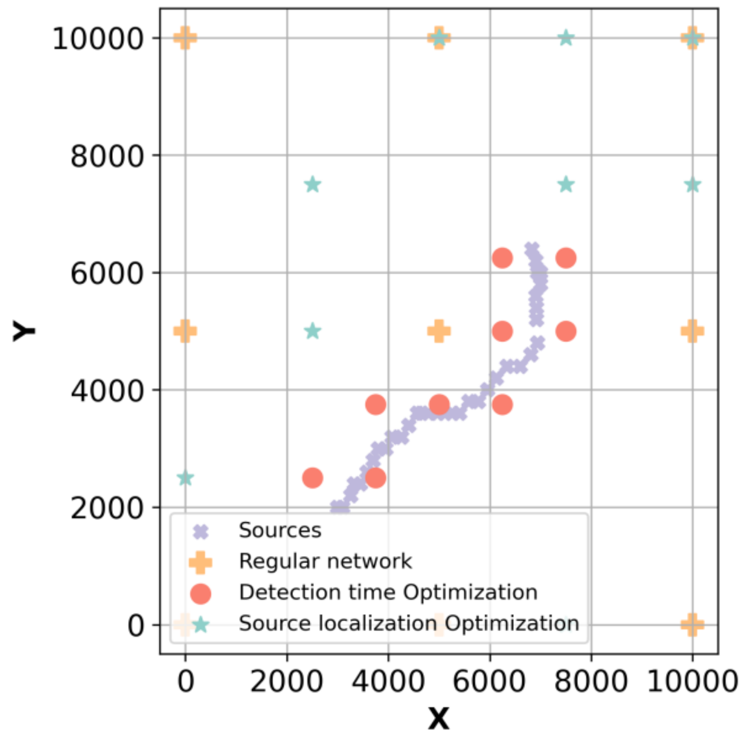
Localization accuracy optimization workflow

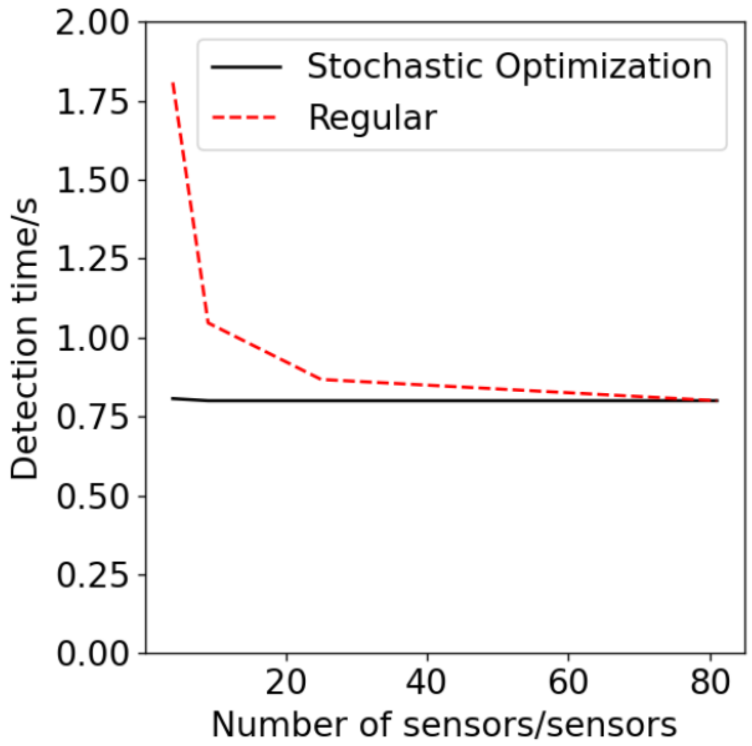


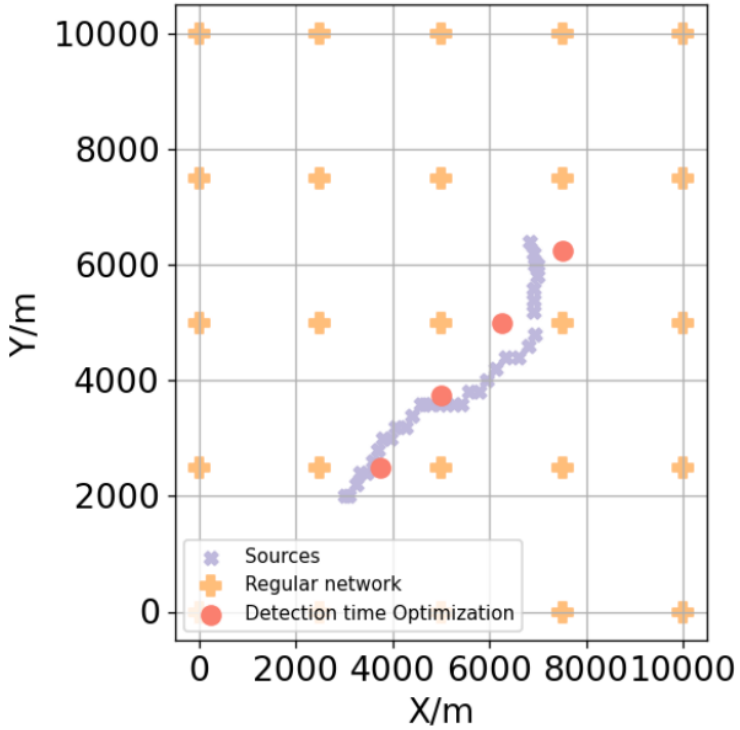
Same Budget Placement strategies for detection time and localizations

When budget increase, both methods performance converge

Same performance, SO method use less sensor budget







Results
1. Achieve 20% higher performance with same budget.
2. Achieving same performance level with less sensors.

Problem Formulation 3: Optimal Worst Detection Time

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

subject to

$$\sum_{l \in \mathbb{L}} y_l \le c, \quad (2)$$ cost

$$x_{e,i} \le y_i \quad \forall\, e \in \mathcal{E},\, i \in L_e, \quad (3)$$ Earliest detection Exist

$$\sum_{i \in \mathbb{L}_e} x_{e,i} = 1 \quad \forall\, e \in \mathcal{E}, \quad (4)$$ Earliest detection Sole

$$y_l \in \{0, 1\} \quad \forall\, l \in \mathbb{L}, \quad (5)$$ Binary

$$0 \le x_{e,i} \le 1 \quad \forall\, e \in \mathcal{E},\, i \in L_e, \quad (6)$$ Earliest detection Range

$$\mathbb{B}_\kappa^{d_{e,i}} := \{Q \in G_+ : \mathcal{D}_W(Q, T) \le \kappa\}. \quad (7)$$ Uncertainty set

Main Problem Formulation

Inner Problem Formulation

$$d'_{e,i} = \max_{Q \in \mathbb{B}_\kappa^{d_{e,i}}} E(Q), \quad (8)$$ Inner Max obj

subject to

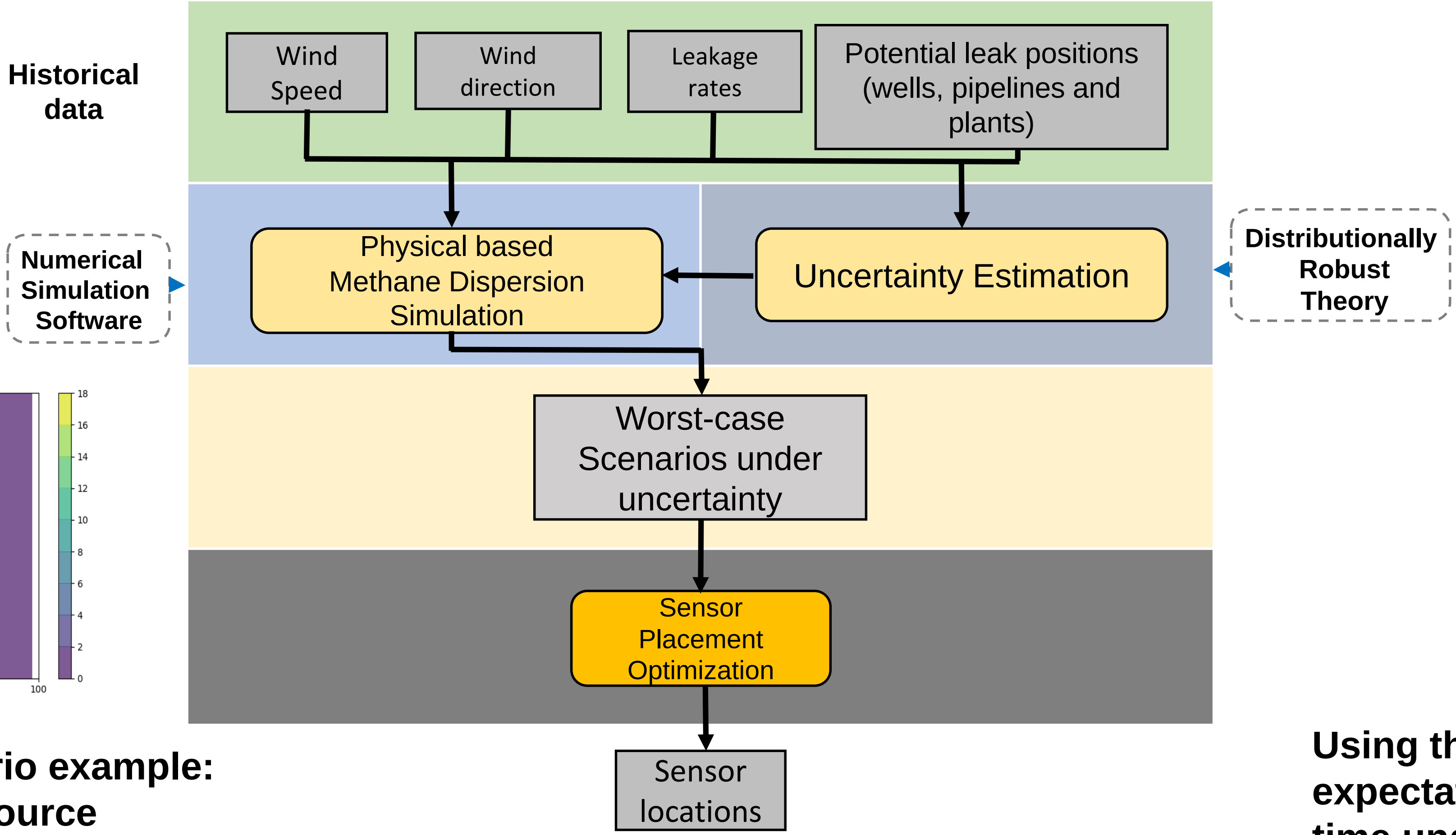
$$\mathbb{B}_\kappa^{d_{e,i}} := \{Q \in G_+ : \mathcal{D}_W(Q, T) \le \kappa\}, \quad (9)$$ Uncertainty set

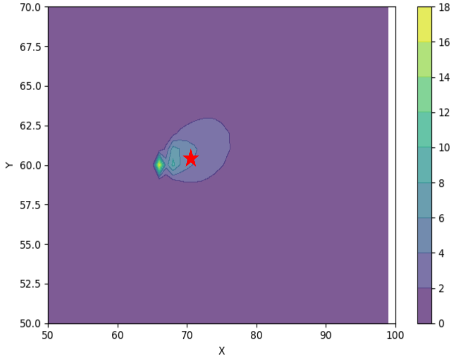
$$\kappa = \left(\frac{H}{2S}\right) \log\left(\frac{2H}{1-\gamma}\right), \quad (10)$$ Uncertainty Level Estimation

Symbol	Meaning
$e \in \mathcal{E}$	The set of all events.
L	The set of all candidate sensors.
L_e	The set of all sensors that are capable of detecting event e
p_e	The 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	The cost of sensor i
c	The sensors' budget
κ	The 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_+	The 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

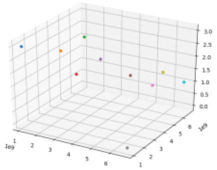
Table of Symbols

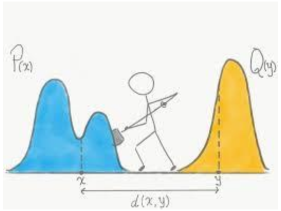
Methane DRO Worst Detection Time Optimization Workflow

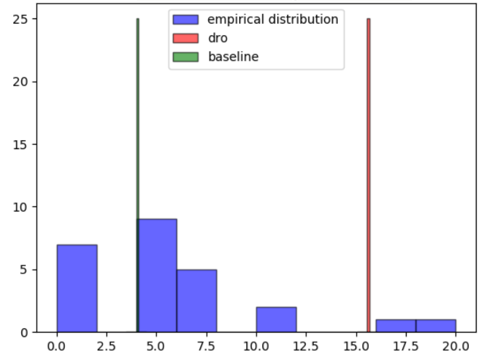




One leak scenario example:
Red star: leak source
Plume: methane propagation after 24 hours



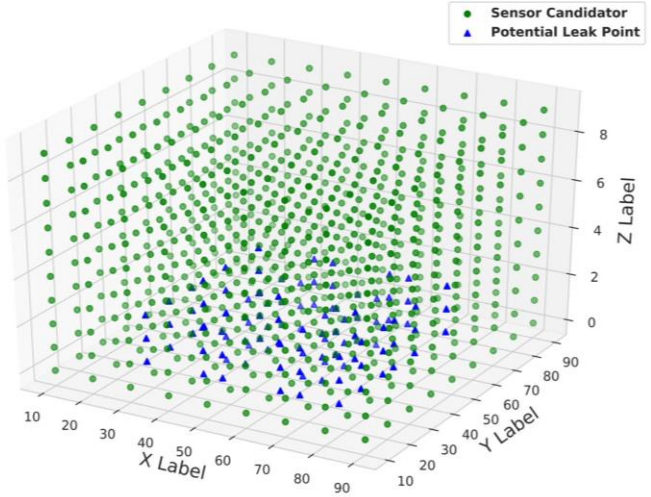




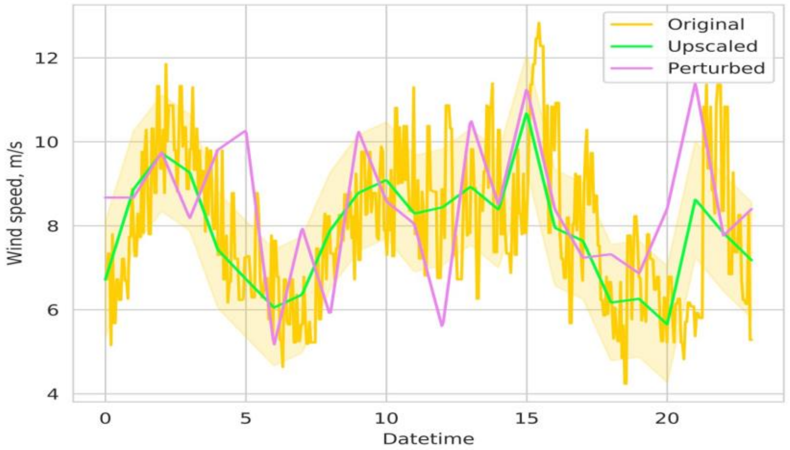
Using the distributionally robust expectation of worst detection time under uncertainty for sensor placement optimization.

Methane DRO Worst Detection Time Optimization Simulation under Wind Perturbation

Simulation cube



Wind Speed



- Simulate 74 leakage events as observed events optimization database
- Test baseline wind speed average method, Stochastic optimization and DRO results on unseen dataset contains 148 leak events.

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

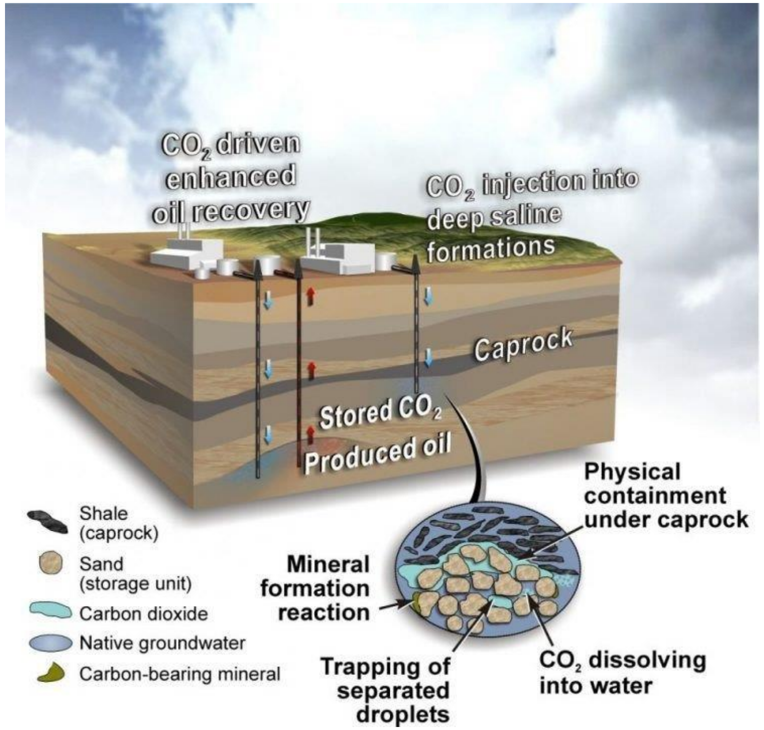
Outline

- **Introduction**
  - Carbon sequestration induced seismicity and gas leak monitoring
  - Data-driven optimization methods
  - Dissertation contributions
- **Carbon Monitoring Sensor Placement**
  - Work case 1: Stochastic optimal induced seismicity sensor placement
  - Work case 2: Distributionally robust optimal methane sensor placement
- **Stack of Machine Learning Enhanced Signal Processing Methods**
  - Denoising Work: Swell noise attenuation with self-supervised learning
  - Pattern Matching Work: Well-logging pattern matching with reinforcement learning
- **Remote Methane Monitoring**
  - Segmenting Hyperspectral Images of Methane Plumes with a Large Machine Learning Model
- **Conclusion**

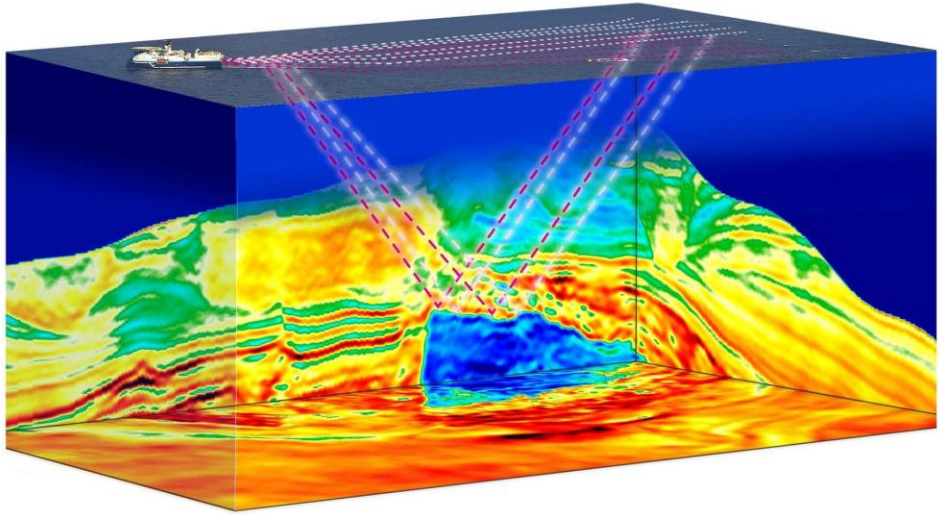
UNIVERSITY of HOUSTON
YOU ARE THE PRIDE
CULLEN COLLEGE of ENGINEERING
Department of Electrical & Computer Engineering

Carbon sequestration requires subsurface surveys (imaging, well-logging) to design injection plans and monitor injection safety.

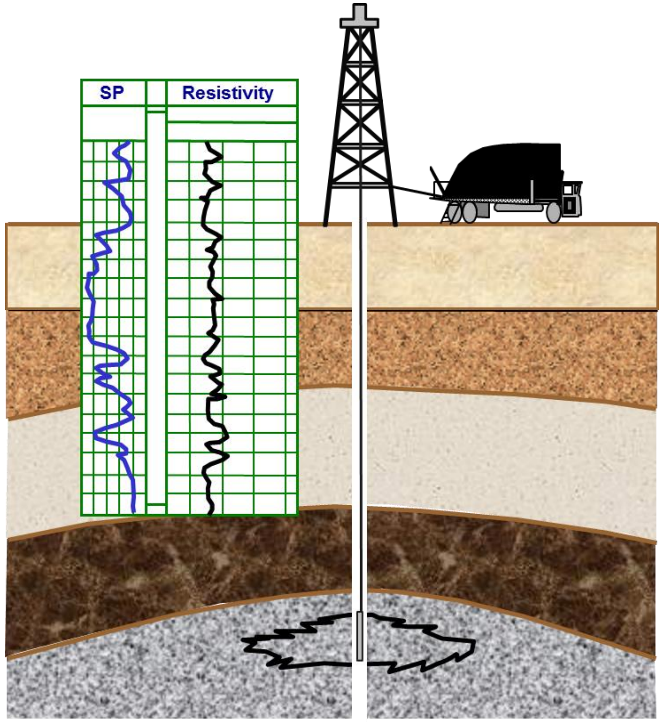
Carbon Sequestration
Injecting Scope



Pre-injecting survey
Imaging



Pre-injecting survey
Well-logging

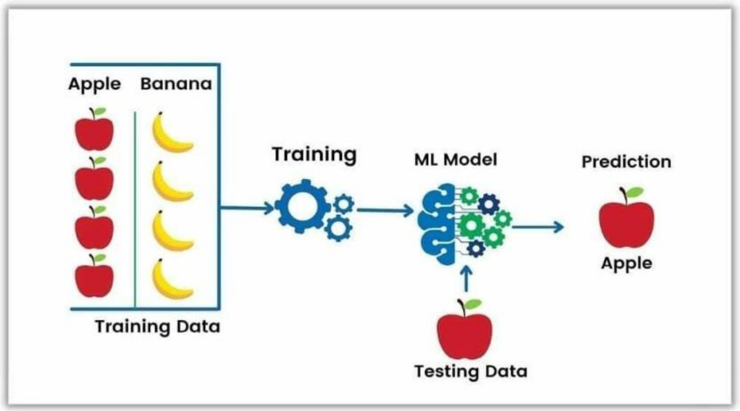


Ref link @https://github.com/yohanesnuwara/carbon-capture-and-storage.
https://gardaerlangga.wordpress.com/2014/07/06/well-logging-definisi-dan-sejarahnya/
https://www.cgg.com/geoscience/subsurface-imaging
https://www.mdpi.com/1424-8220/21/17/5815
https://www.researchgate.net/publication/254528542_Modeling_Leakage_Through_Faults_of_CO2_Stored_in_an_Aquifer/figures?lo=1

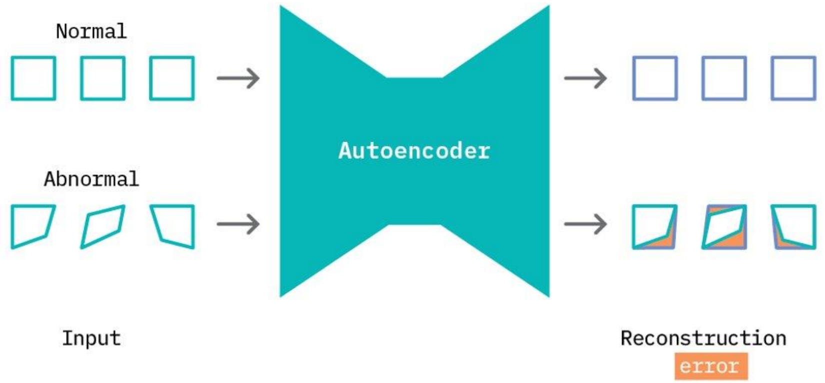
Benefits:

1. **Be able to processing large scale data automatically**
2. **Be able to learn complex high dimensional pattern**

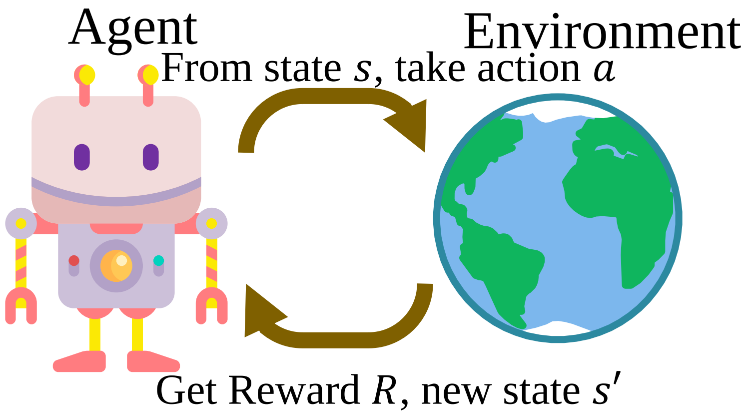
Supervised Learning



Self-supervised Learning



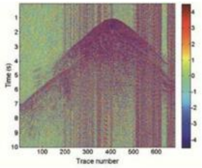
Reinforcement Learning

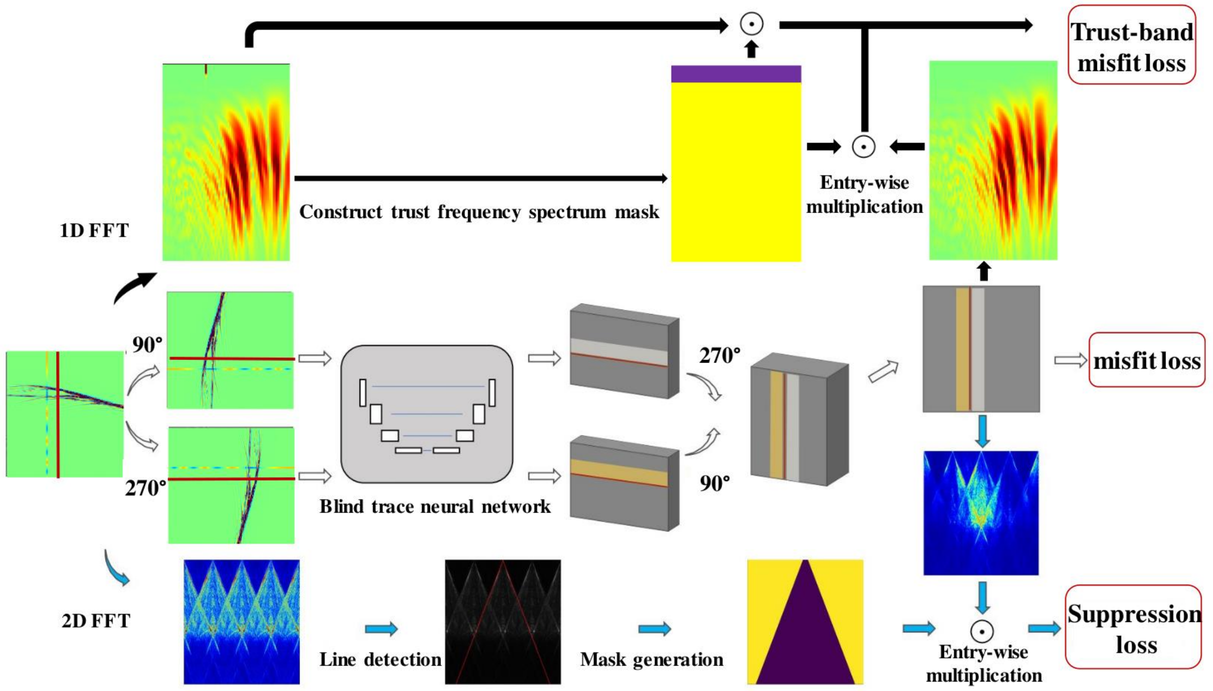


Challenge:

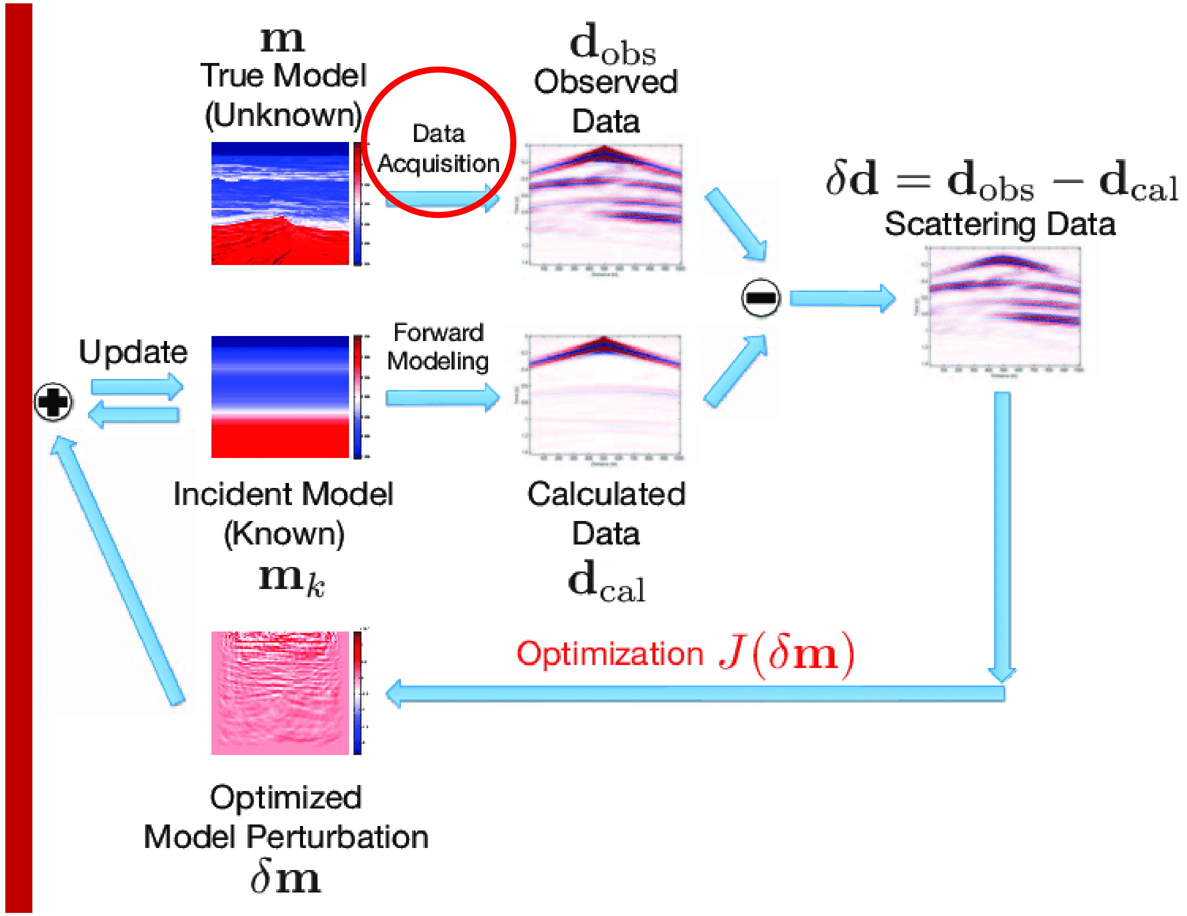
Generalization: the practice data distribution is different from the training dataset

Challenge: The seismic data acquired from the field has noise such as swell and monochromatic noise and the traditional denoising method has limitations which also would attenuate the signal when performing denoising

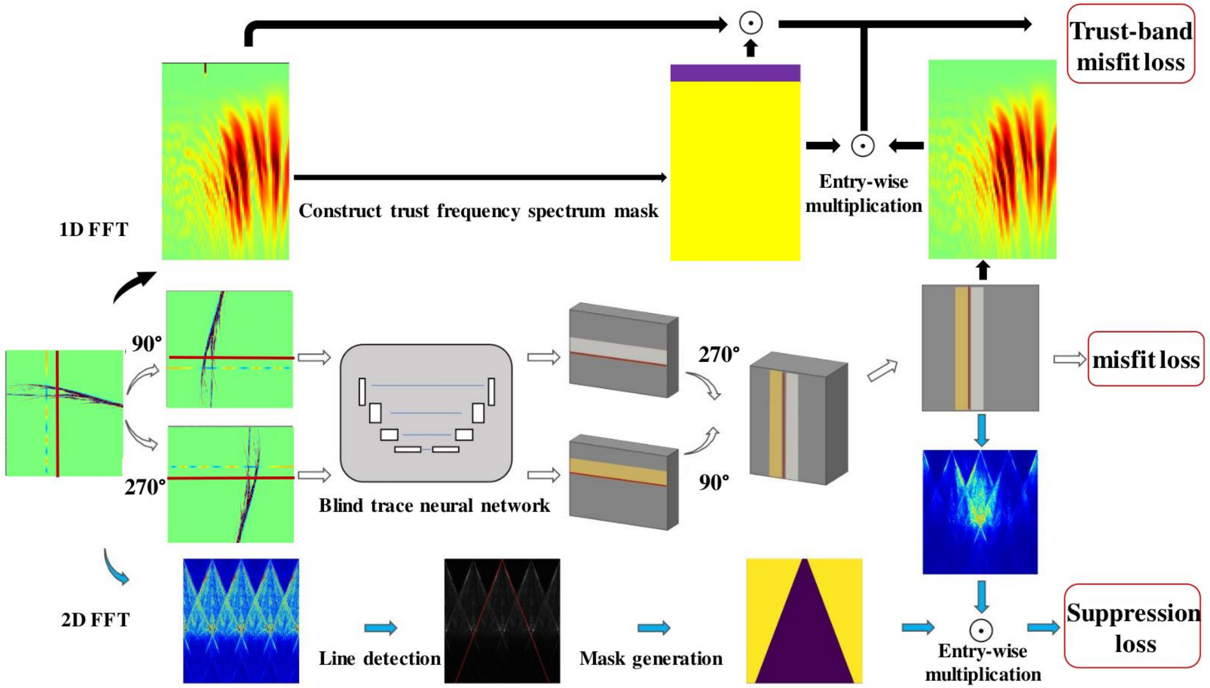


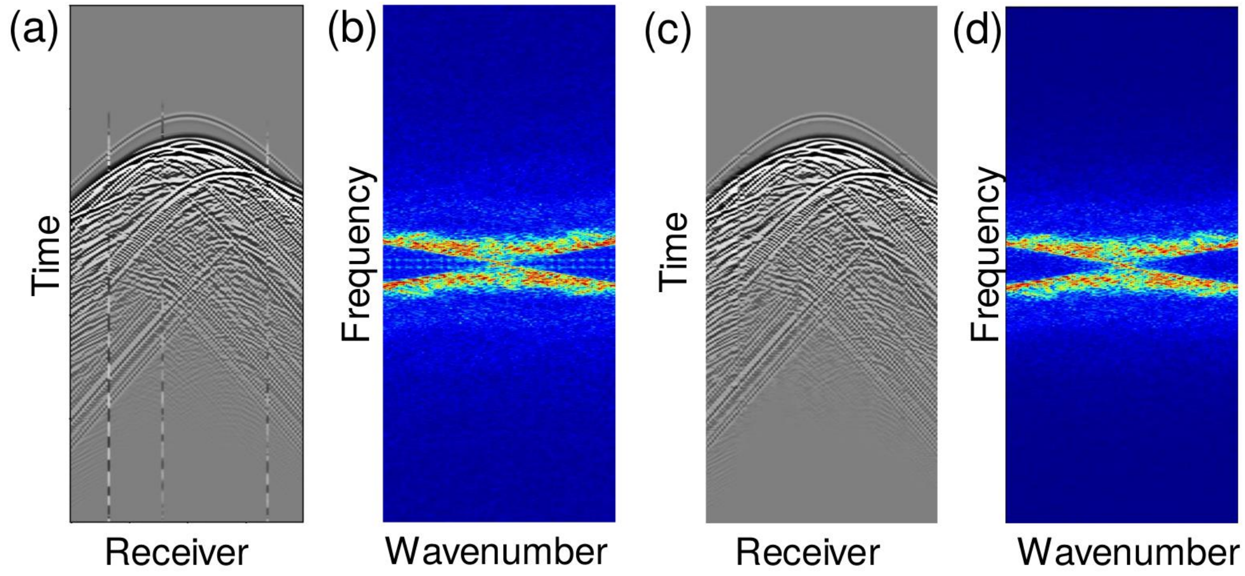


Blind Trace Signal Autoencoder can only learn reconstruct signal remove noise.



Self-Supervised Learning Results





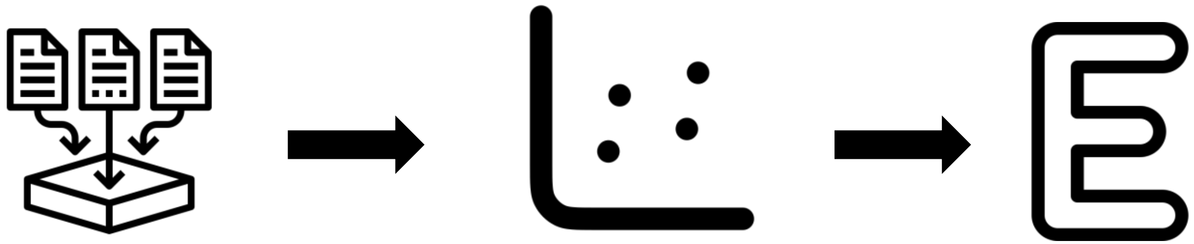
[1]

Physical priors and domain knowledge are keys to formulating self-supervised learning ignoring data distribution shifts because it learns from the testing data itself.

[1] Zi, Yuan, Shirui Wang, Pengyu Yuan, Xuqing Wu, Jiefu Chen, and Zhu Han. "Self-supervised learning for seismic swell noise removal." In Second International Meeting for Applied Geoscience & Energy, pp. 1910-1914. Society of Exploration Geophysicists and American Association of Petroleum Geologists, 2022.

Reinforcement Learning Signal Localization

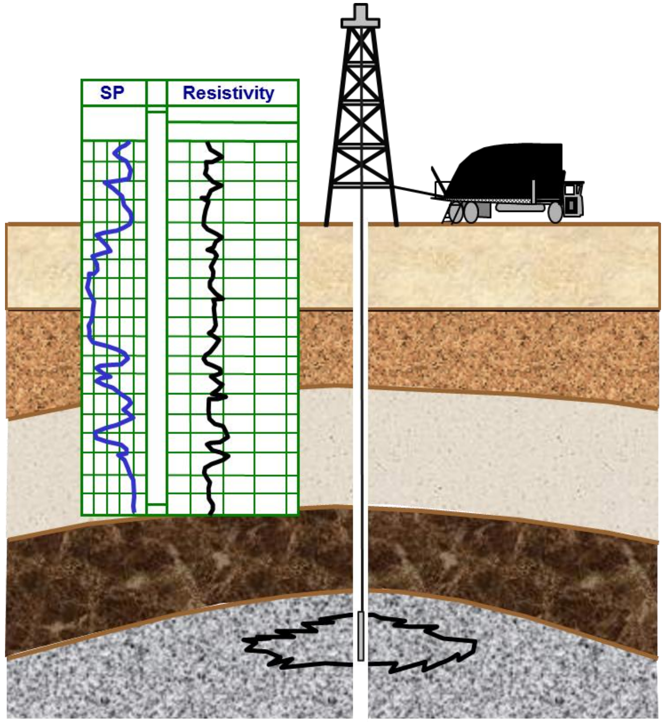
Motivation Using DRRL for Well Logging



data collection → **limited samples** → **estimation errors in Policy Iteration**

Learning to match patterns of this behavior itself constitutes a reinforcement learning task, wherein we anticipate it to achieve:

- 1. Adhering to a conservative policy in an unfamiliar sample.
- 2. Adopting an optimistic policy in a familiar sample.



[4] Zi, Yuan, Fan, Lei, Wu, Xuqing, Chen, Jiefu, Wang, Shirui, and Zhu Han. "Active gamma-ray well logging pattern localization with reinforcement learning." Paper presented at the SEG/AAPG International Meeting for Applied Geoscience & Energy, Houston, Texas, USA, August 2022. doi: https://doi.org/10.1190/image2022-3745281.1

[5] Y. Zi, L. Fan, X. Wu, J. Chen, S. Wang and Z. Han, "Active Gamma-Ray Log Pattern Localization With Distributionally Robust Reinforcement Learning," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-11, 2023, Art no. 5911011, doi: 10.1109/TGRS.2023.3278491.

Traditional method-Tedious:

1. Human vision
2. Prior knowledge about rocks
3. Rough correlation metrics

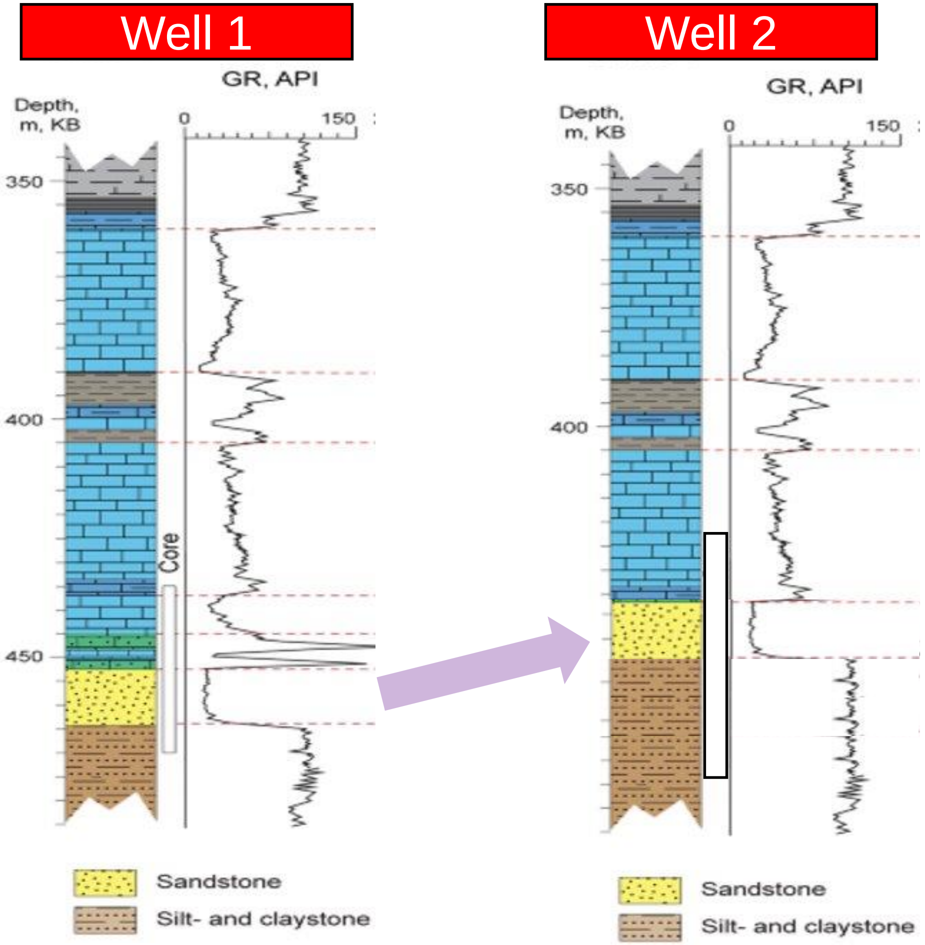
RL-Automatic:

1. Machine vision
2. Pattern recognition
3. Localization Loss

DRRL-Automatic and Robust:

1. conservative in unfamiliar samples
2. optimistic in familiar samples

Gamma-Ray Signal Localization Task

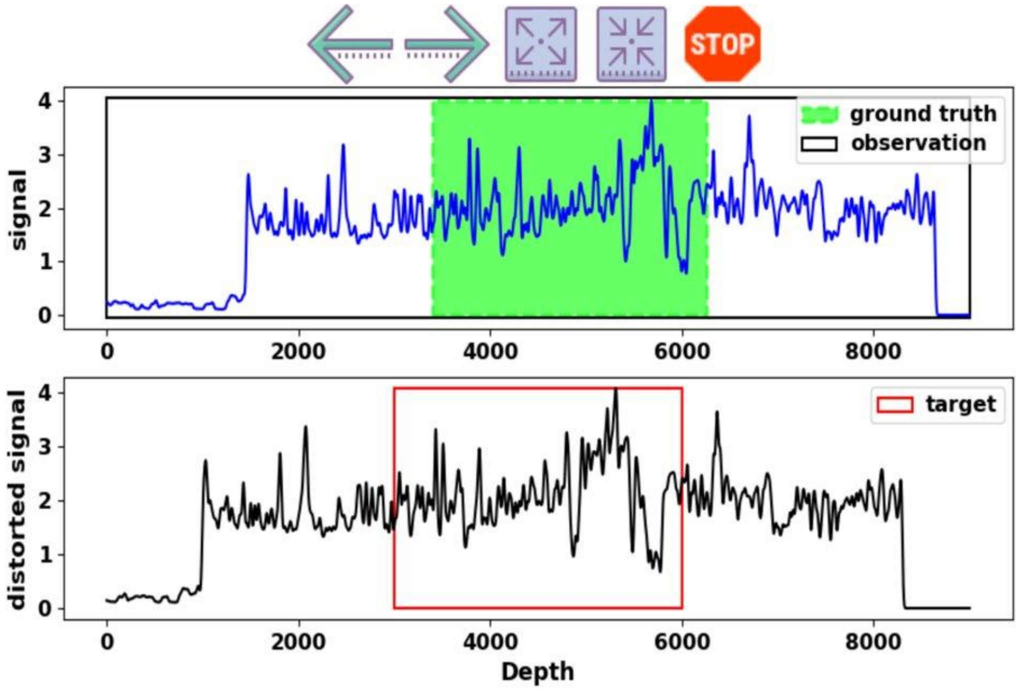


[4] Zi, Yuan, Fan, Lei, Wu, Xuqing, Chen, Jiefu, Wang, Shirui, and Zhu Han. "Active gamma-ray well logging pattern localization with reinforcement learning." Paper presented at the SEG/AAPG International Meeting for Applied Geoscience & Energy, Houston, Texas, USA, August 2022. doi: https://doi.org/10.1190/image2022-3745281.1

[5] Y. Zi, L. Fan, X. Wu, J. Chen, S. Wang and Z. Han, "Active Gamma-Ray Log Pattern Localization With Distributionally Robust Reinforcement Learning," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-11, 2023, Art no. 5911011, doi: 10.1109/TGRS.2023.3278491.

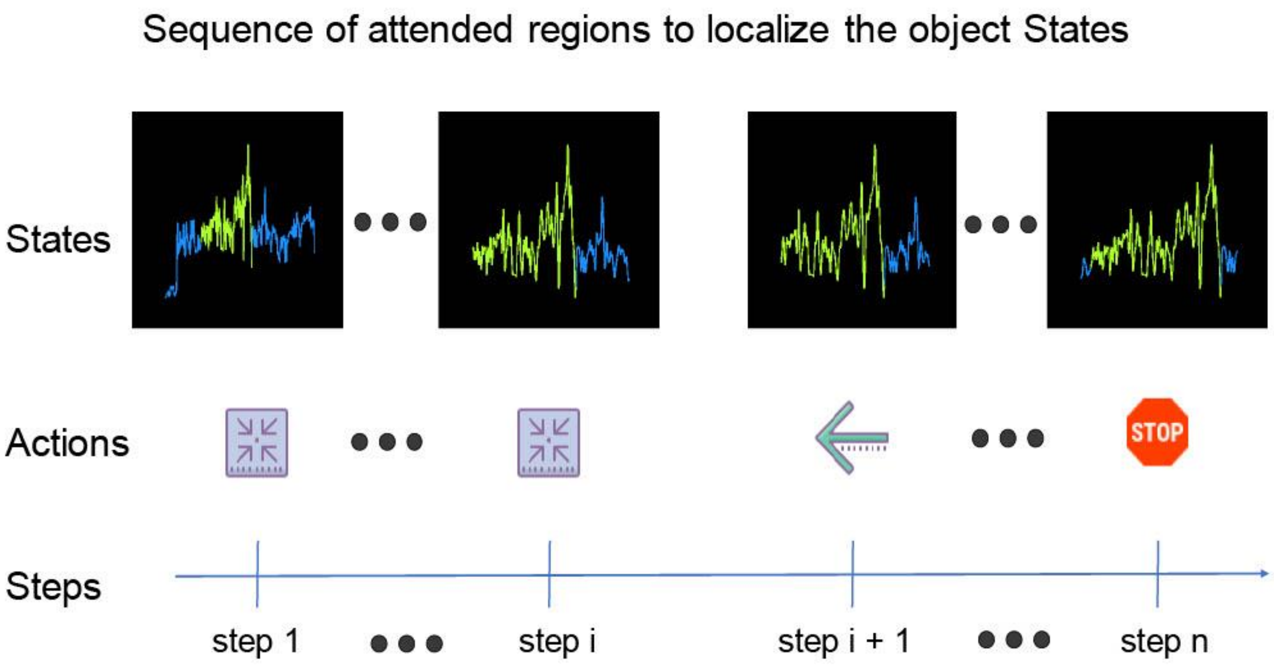
Reinforcement Learning Signal Localization Scheme

How does the problem look like



1. Two series, each containing a signal fragment as a reference/target, separately.
2. Search for target, given reference and new trace contain the target.

How does the solution process look like



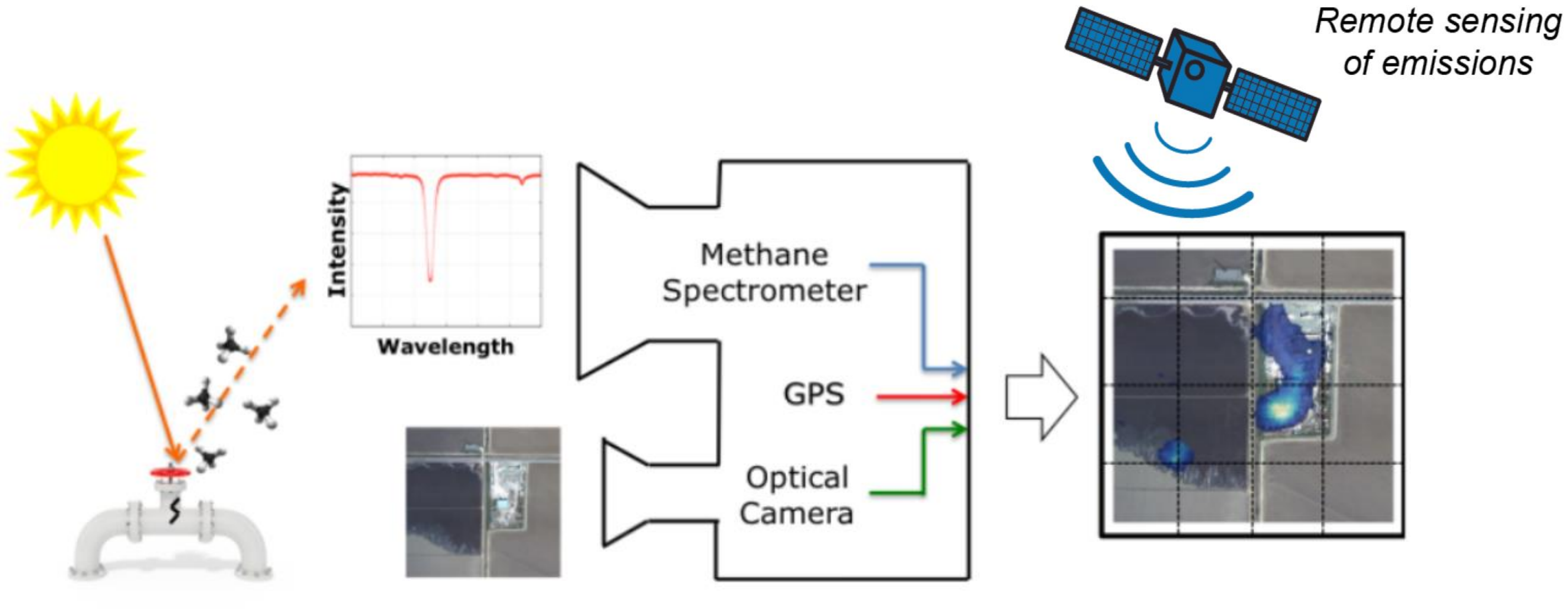
3. Initial the whole new log trace as the agent's observation.
4. Let the agent move (left, right, expand, shrink) to search the reference pattern.

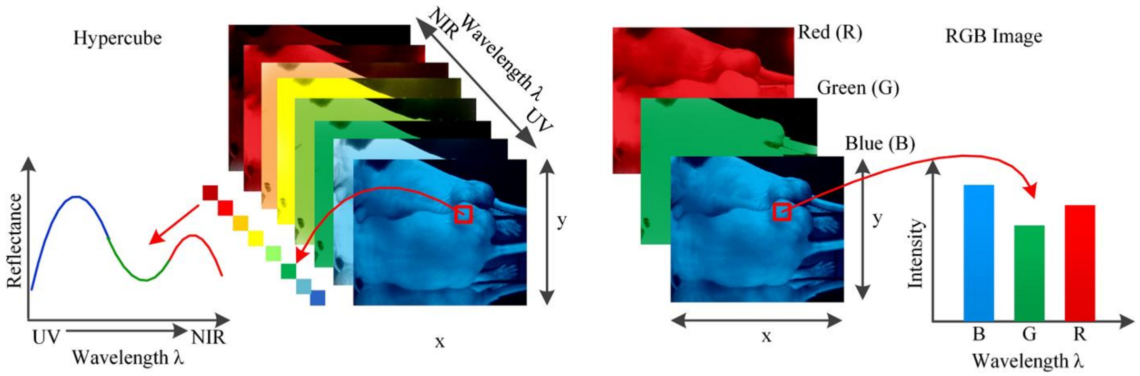
Outline

- **Introduction**
  - Carbon sequestration induced seismicity and gas leak monitoring
  - Data-driven optimization methods
  - Dissertation contributions
- **Carbon Monitoring Sensor Placement**
  - Work case 1: Stochastic optimal induced seismicity sensor placement
  - Work case 2: Distributionally robust optimal methane sensor placement
- **Stack of Machine Learning Enhanced Signal Processing Methods**
  - Denoising Work: Swell noise attenuation with self-supervised learning
  - Pattern Matching Work: Well-logging pattern matching with reinforcement learning
- **Remote Methane Monitoring**
  - Segmenting Hyperspectral Images of Methane Plumes with a Large Machine Learning Model
- **Conclusion**

UNIVERSITY of HOUSTON
YOU ARE THE PRIDE
CULLEN COLLEGE of ENGINEERING
Department of Electrical & Computer Engineering

Remote Methane Detection Physical Principle





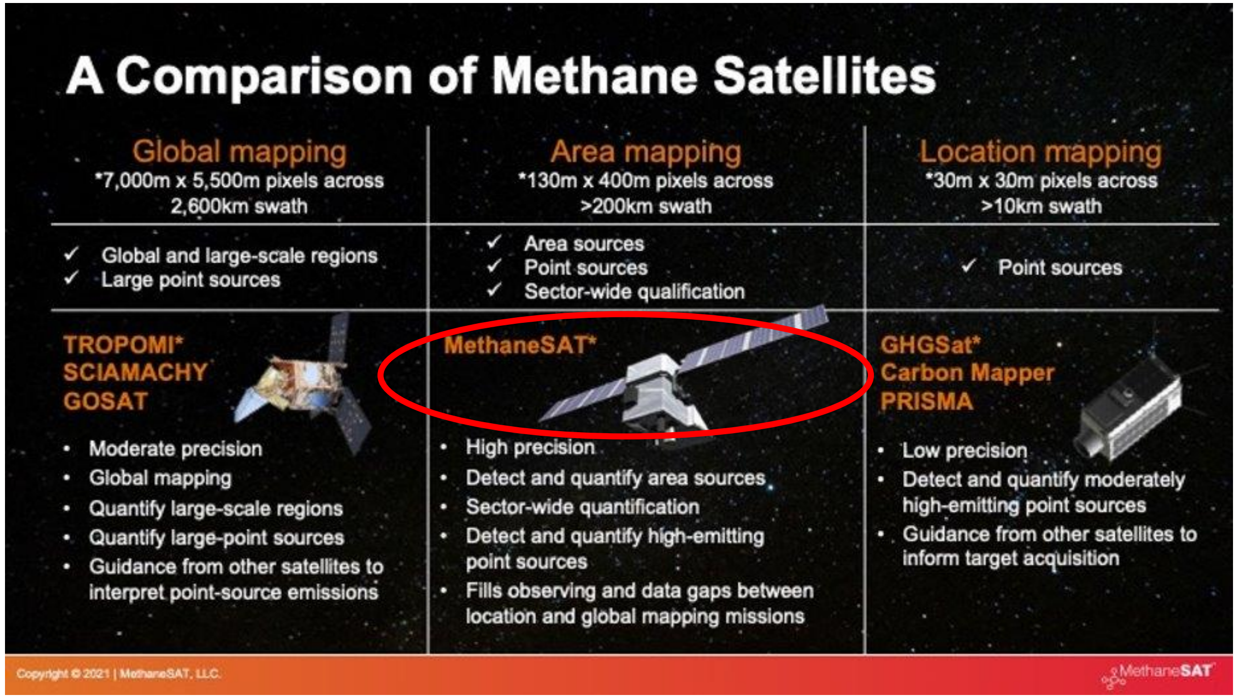
- Reflection light carries carbon element's signature.
- The hyperspectral image have more channels than natural RGB image.

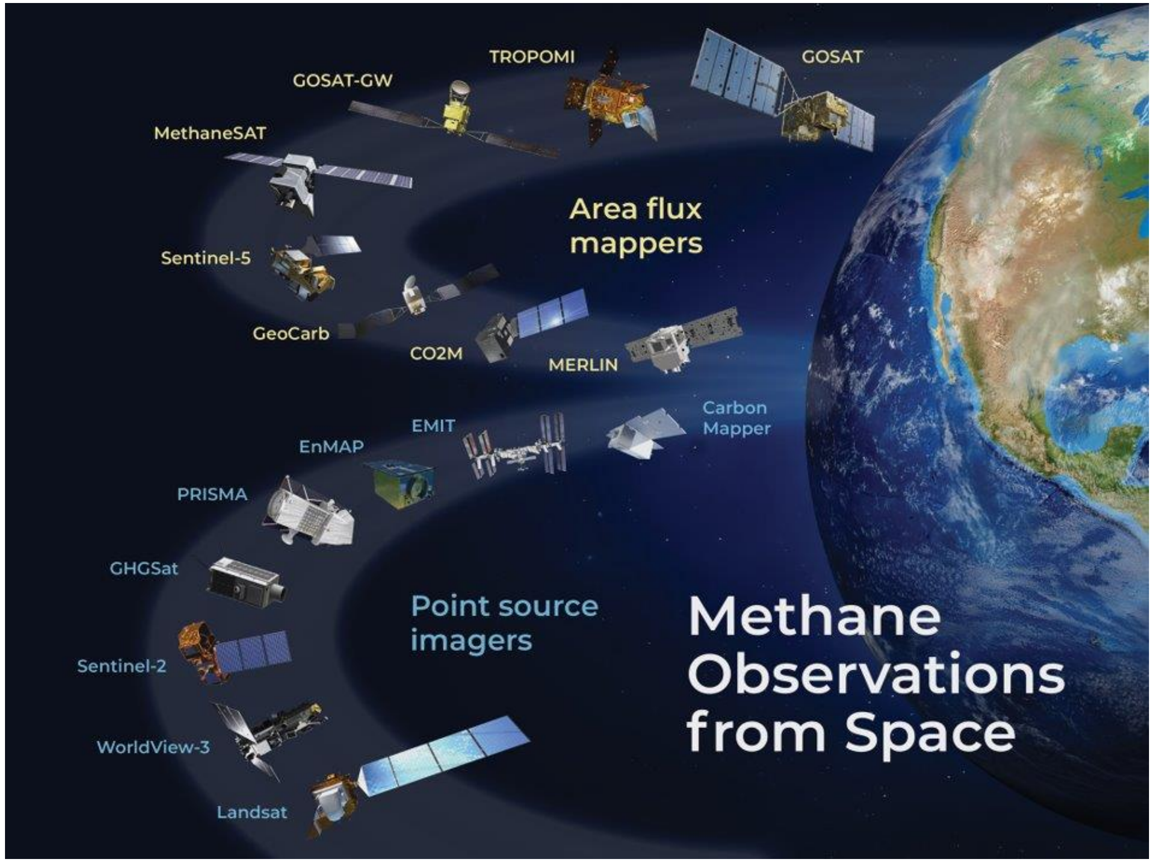
Industry Remote Methane Monitoring Progress

- Majority satellites have low resolution for point source detection.
- Recently some satellites have had the potential to detect point sources but they are commercially owned.
- The public one has detection capability is yet to be launched.

Satellite	Coverage	Constellation size	Swath [km]	~Revisit time (per satellite)	Data availability
GHGSat-C2 [18]	Targeted	5 (C1-C5)*	12	14 days	Commercial
WorldView 3 [20]	Targeted	1	13.1	4.5 days‡	Commercial
PRISMA [21]	Targeted	1	3	7 days	Public
Landsat-8 [22]	Global	1	185	16 days	Public
Sentinel-2 [23]	Global	2	290	10 days	Public

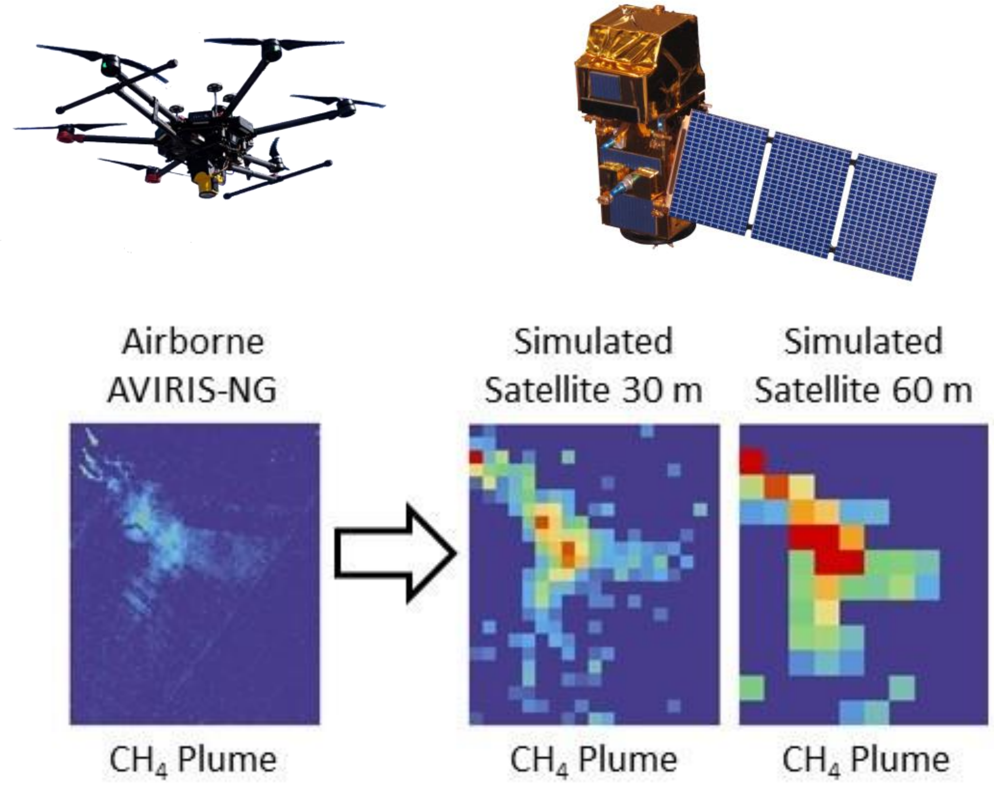
*GHGSat C3-C5 were launched after the conclusion of testing.
‡For best resolution within 20° off nadir. WorldView 3 has 1-day revisit time at lower guaranteed resolution.





Remote Methane Monitoring Researches

Using drone data for satellite research



[6]

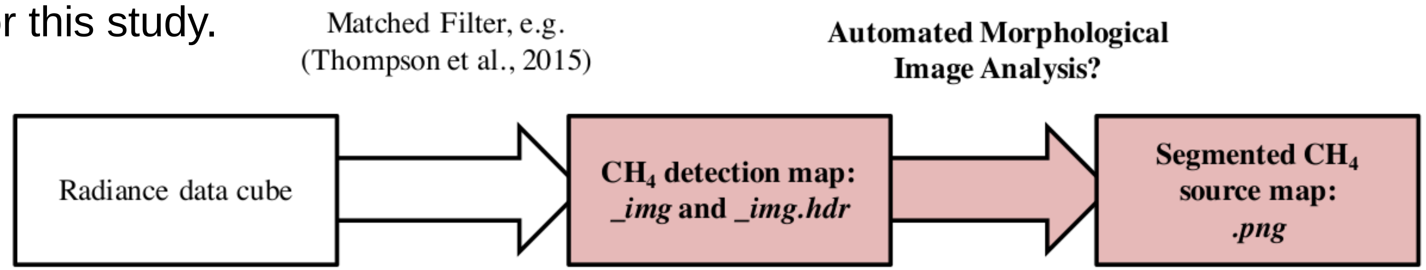
Segmentation of Leakage Plume Task

[7]

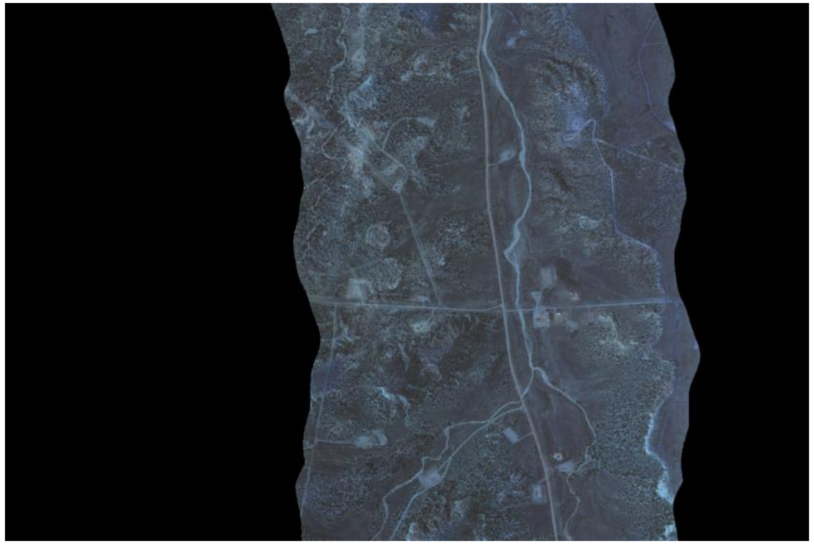
[6] Kumar, Satish, et al. "Deep remote sensing methods for methane detection in overhead hyperspectral imagery." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020.
[7] Jongaramrungruang, Siraput, et al. "MethaNet–An AI-driven approach to quantifying methane point-source emission from high-resolution 2-D plume imagery." Remote Sensing of Environment 269 (2022): 112809.

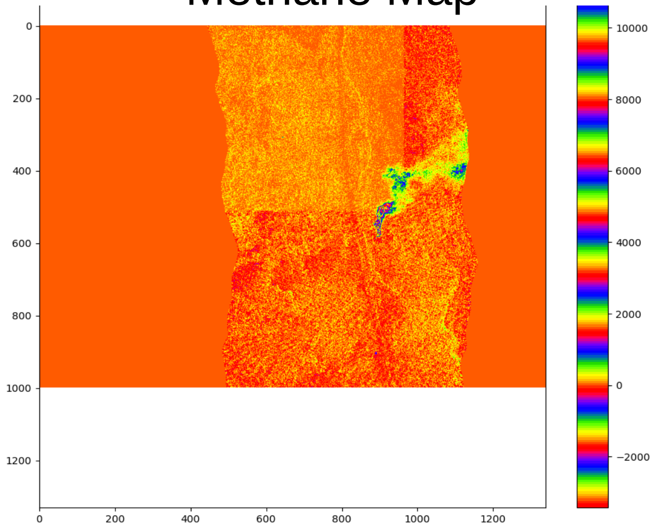
Data Visualization and Analysis

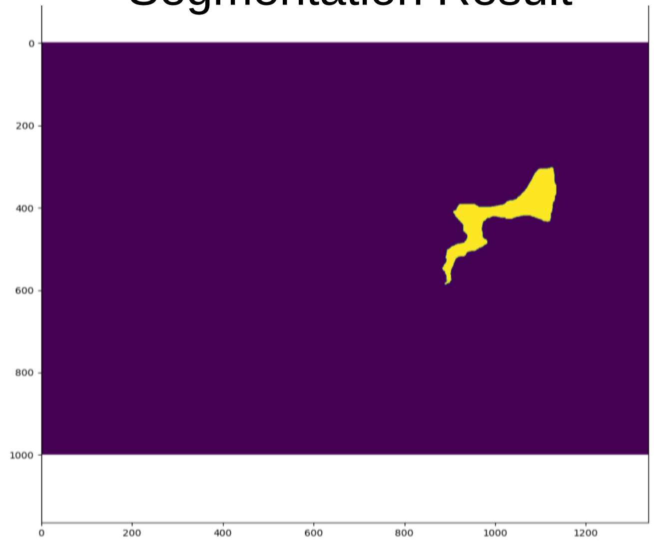
Introduction of the dataset for this study.



RGB Image





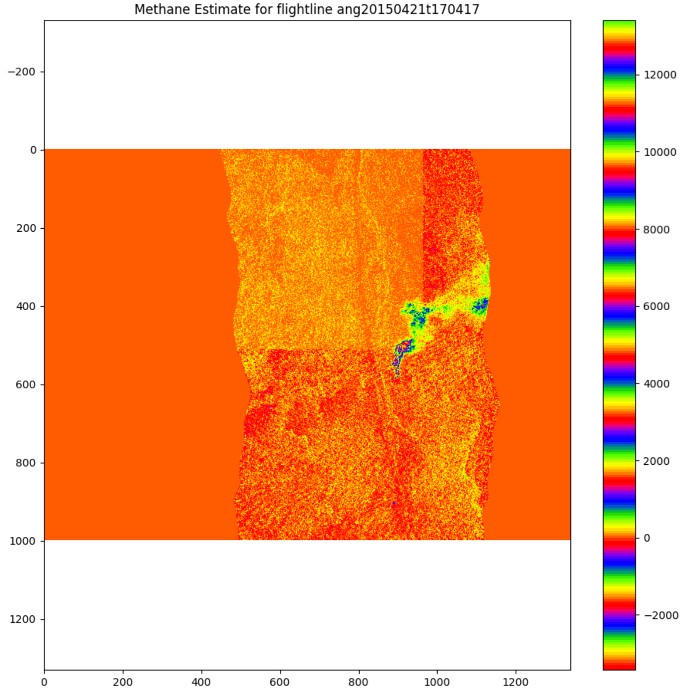


AIVIRIS-NG is an airborne imaging spectrometer that measures radiance in the visible through the (SWIR).
(AVIRIS-NG) = Airborne Visible and Infrared Imaging Spectrometer Next Generation
(SWIR) = Short-wave infrared
AVIRIS-NG dataset is available in https://avirisng.jpl.nasa.gov/benchmark_methane_data.html

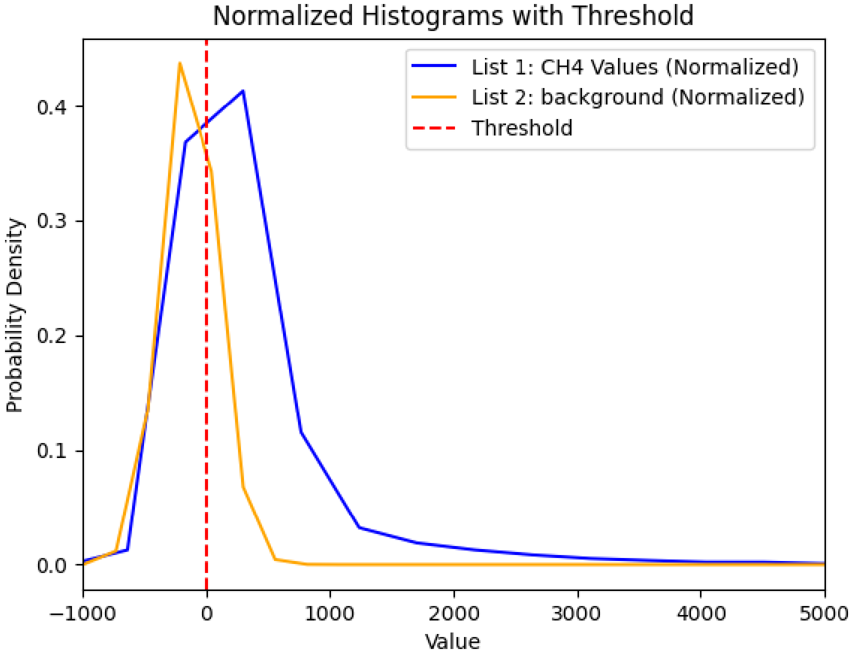
Academic Remote Methane Monitoring Research Tasks

Identify the false alarm problem of the traditional method

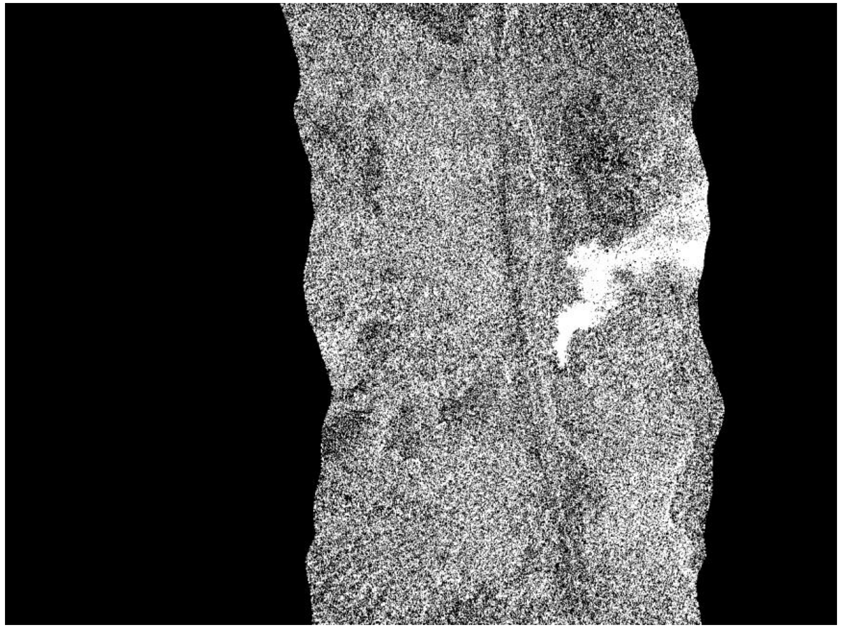
- Background noise mixed with methane signal.
- Filtering false methane light spots is the major focus of this study.



Methane Map



Distributions of Methane signal and Background Noise



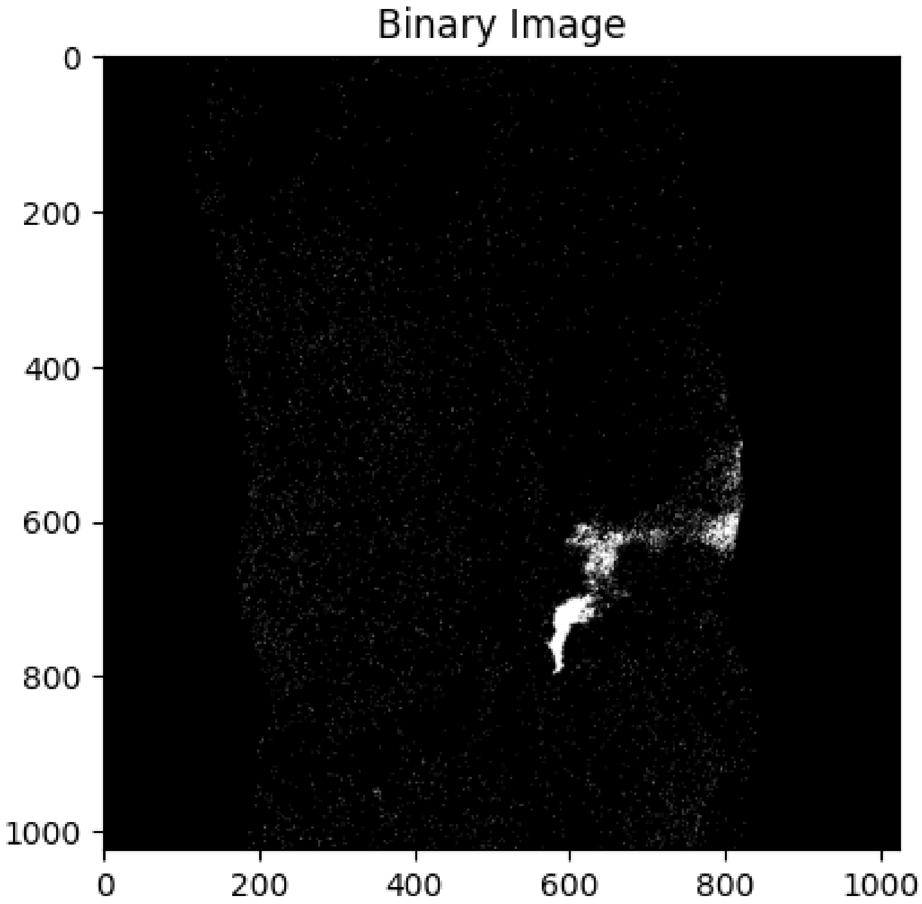
Baseline result
Strong False Alarm

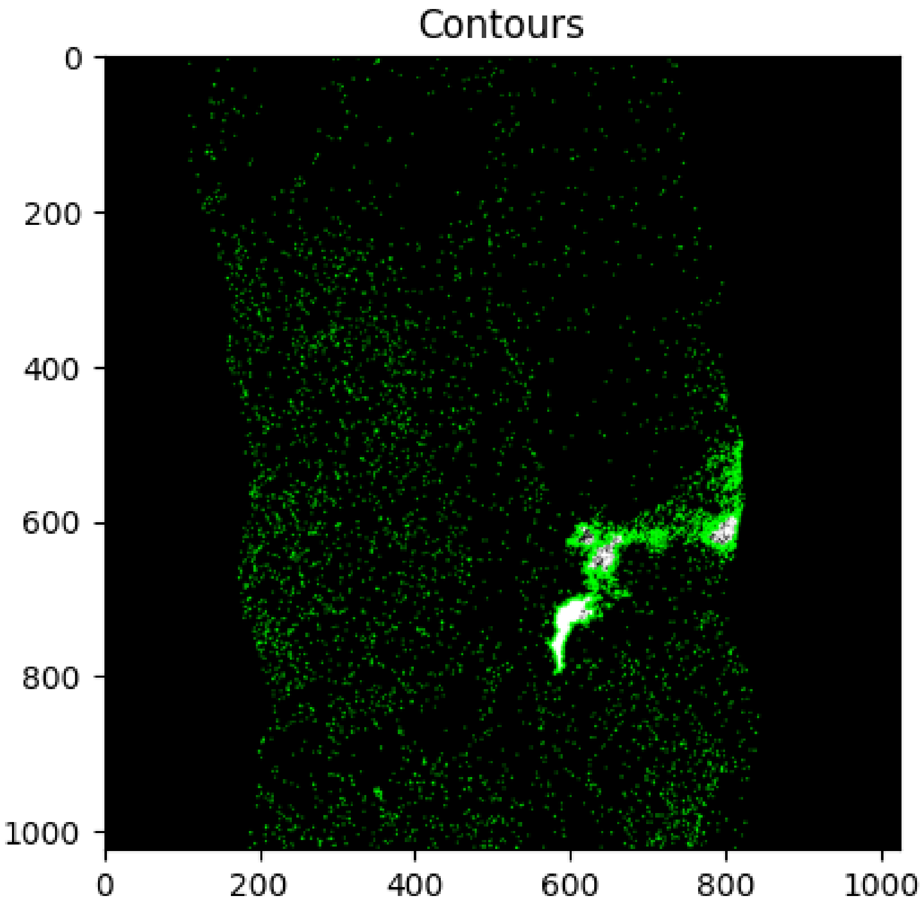
Traditional Method Experiment Result

Visualization of the traditional segmentation result.

Optimal Threshold

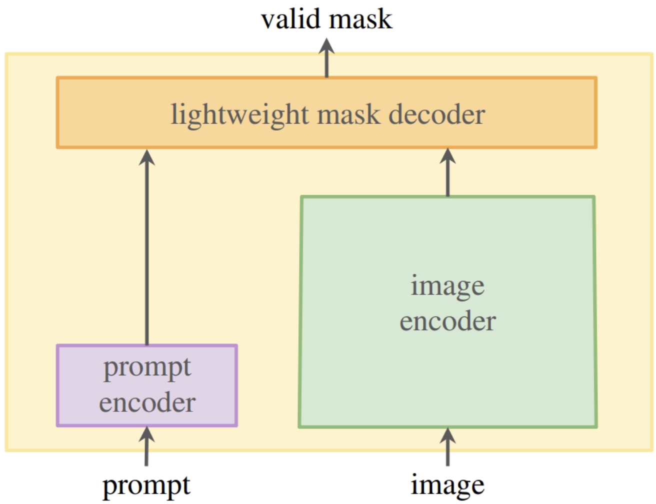
Problem: False alarm contours



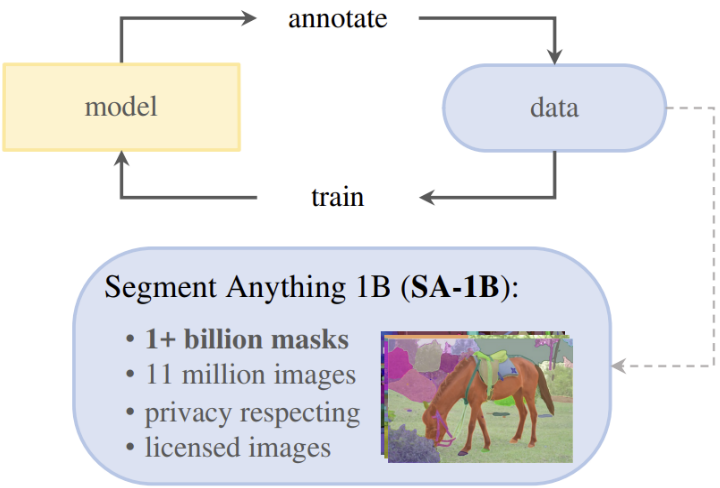


Plume Segmentation Method **Workflow**

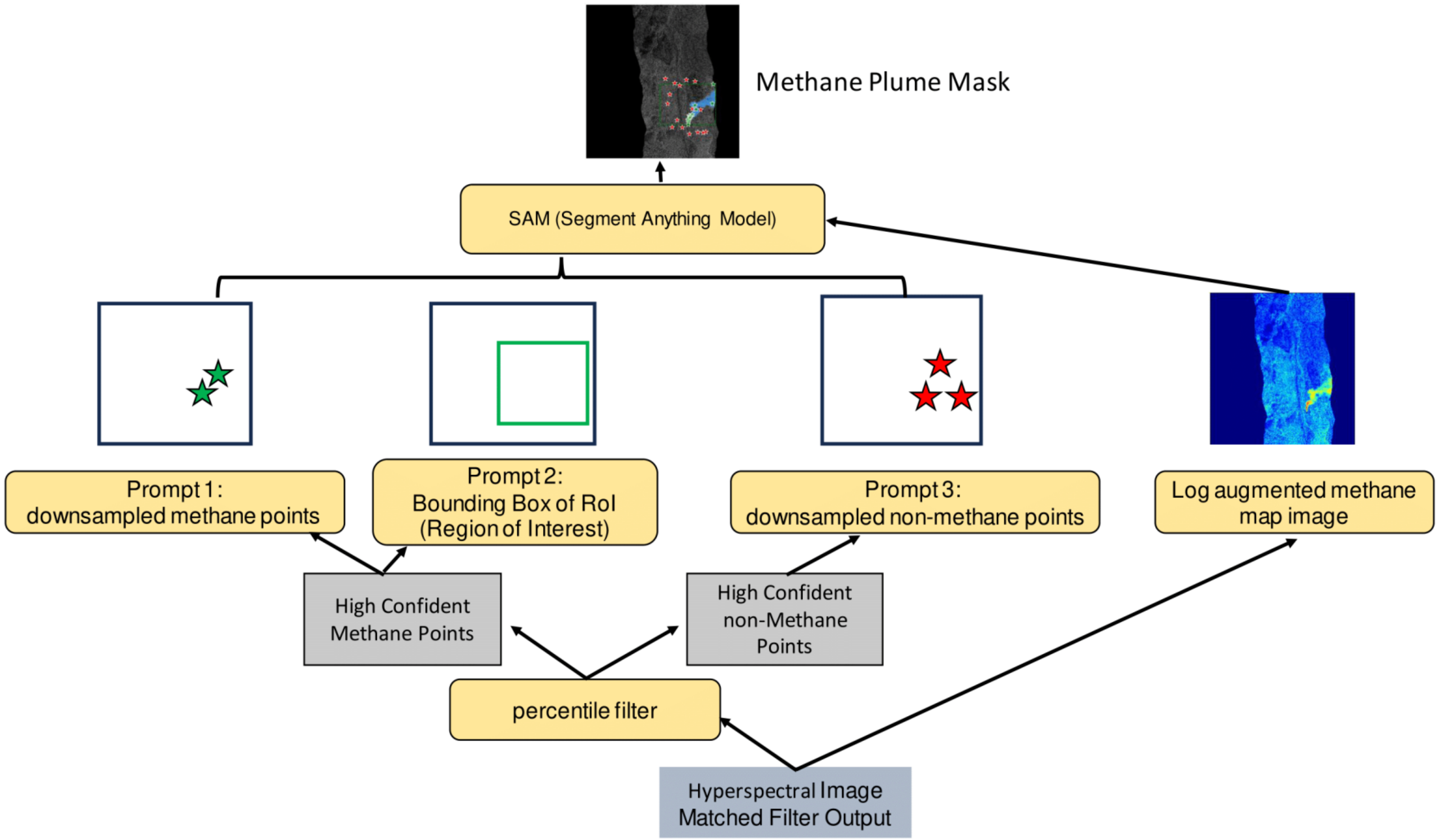
Segment Anything Model



Segment Anything Data



Segment Any Methane Plumes

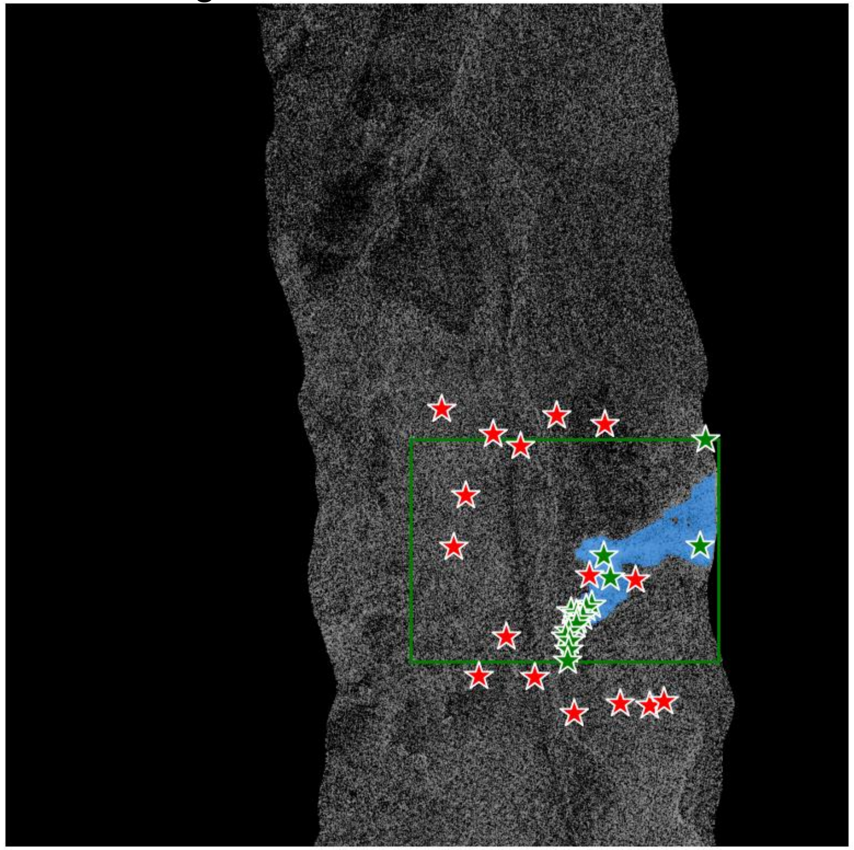


[8] Kirillov, Alexander, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao et al. "Segment anything." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015-4026. 2023.

Segment Anything for Methane Plume Result

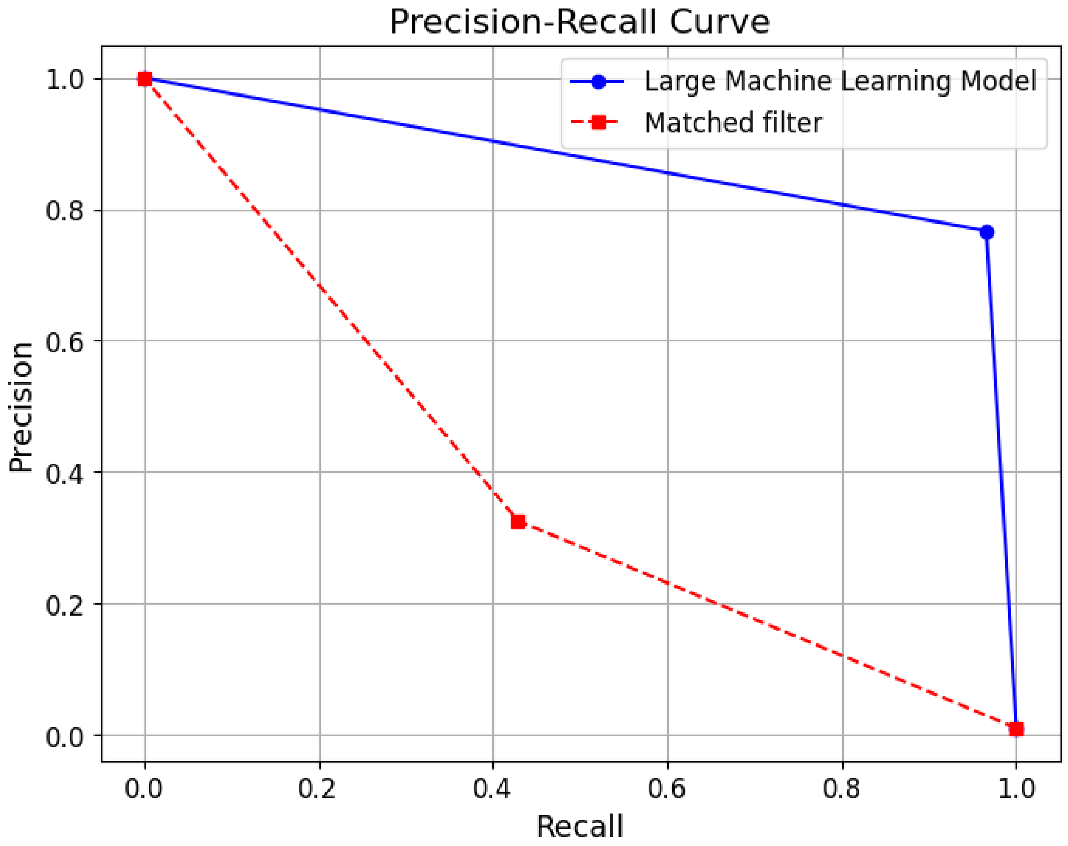
Visualization of the segmentation results in qualitative and quantitative ways.

Segmented Methane Plume Mask



Red stars: high confident background points
Green stars: high confident methane points
Bounding box: high confident methane region

Compare to the Label



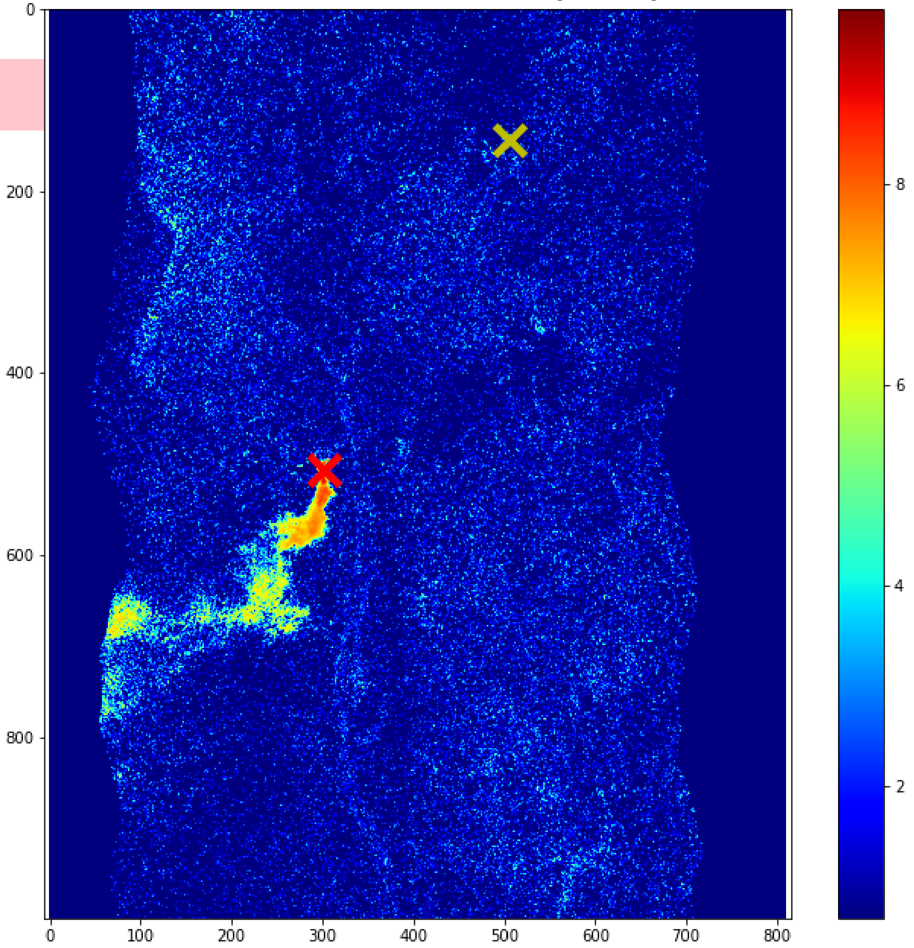
Methane Spectrum Examination

Extract two spectrum traces, one true methane signal and one false alarm for examination

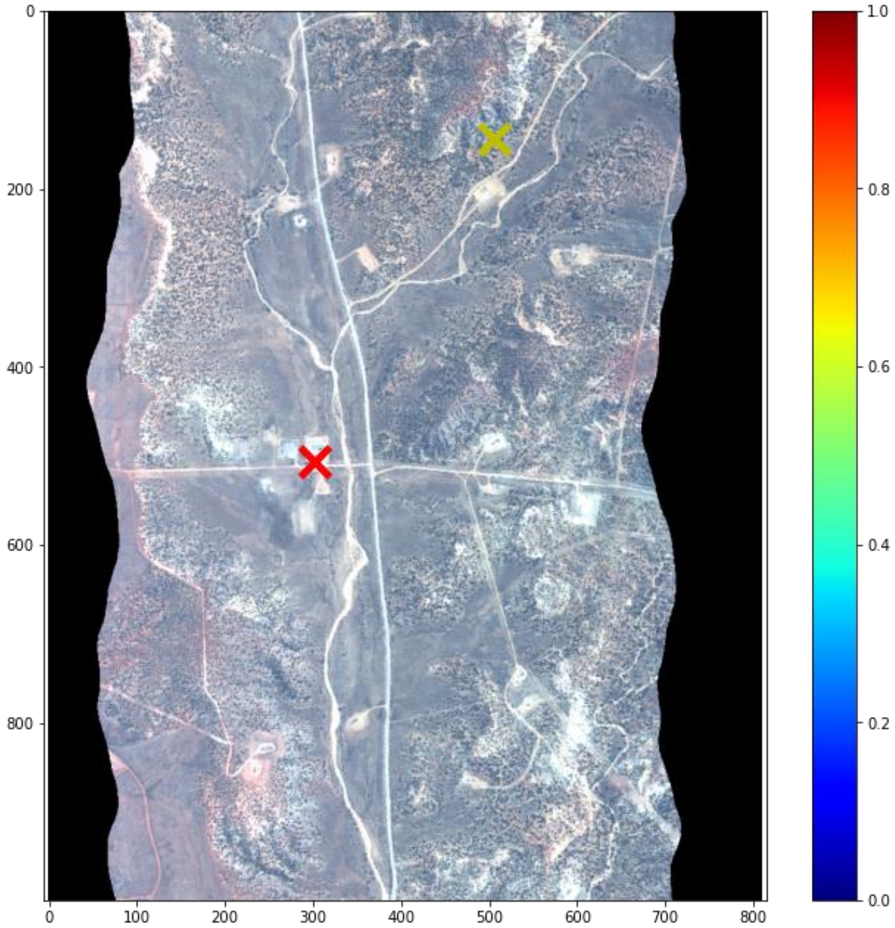
Red cross: true positive; Yellow cross: false alarm

Trace examination

CH4 Intensity map

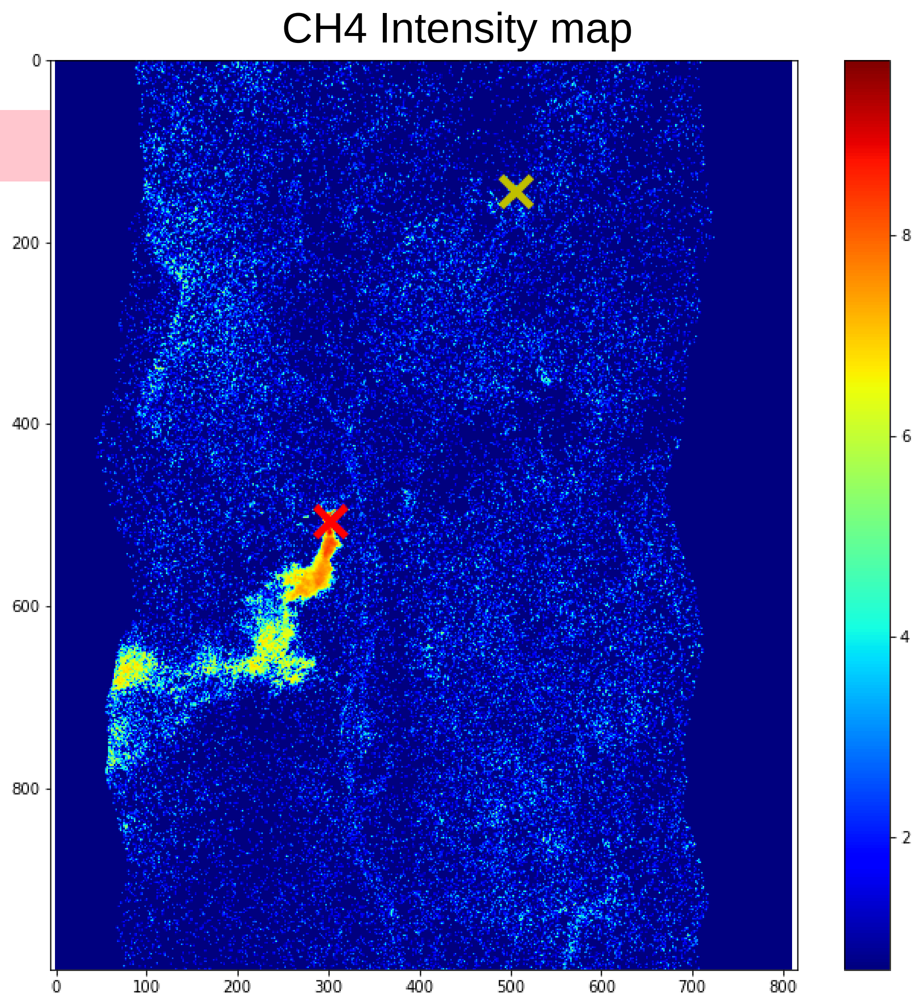


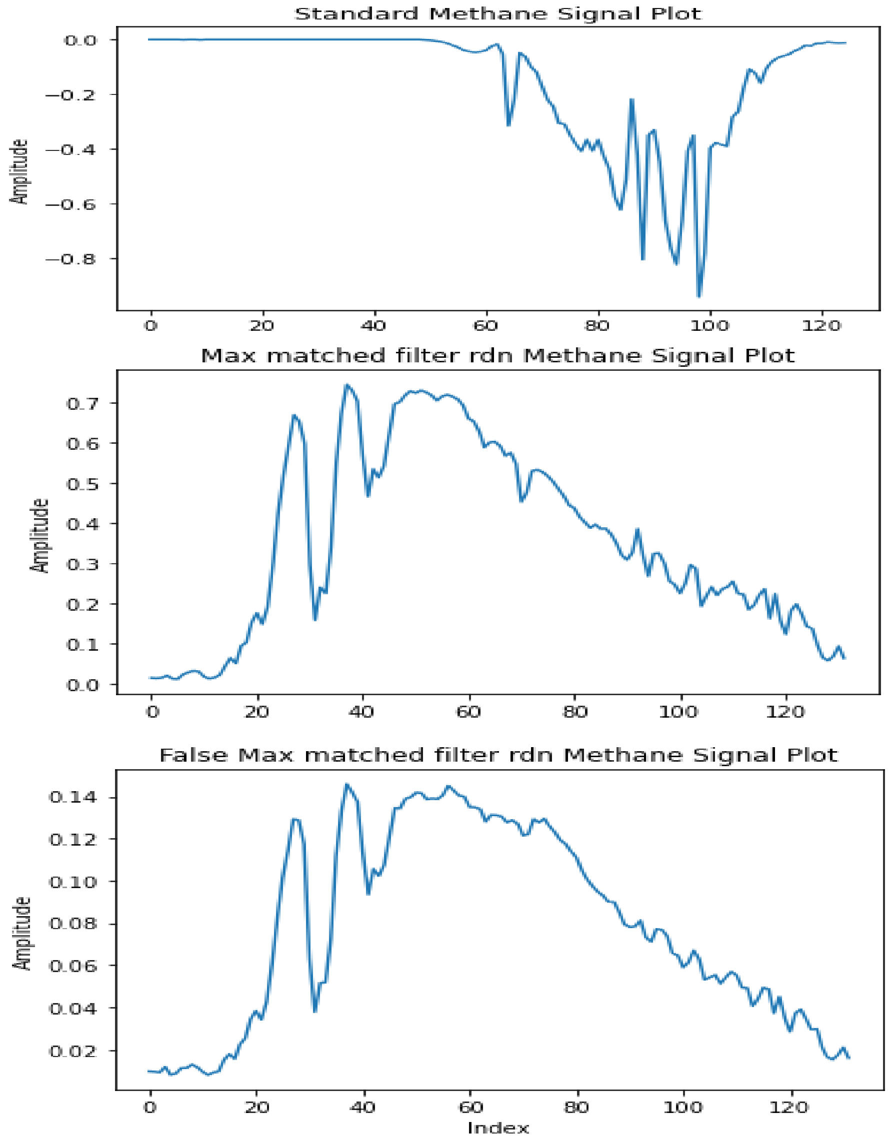
RGB image



Methane Spectrum Examination

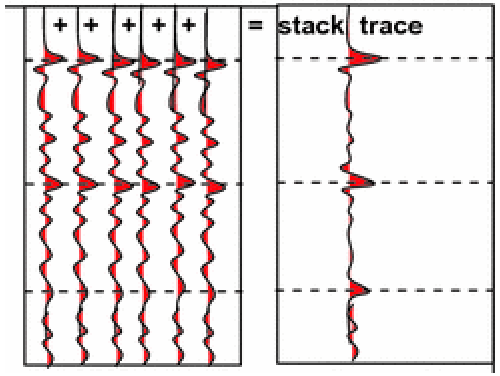
Trace examination



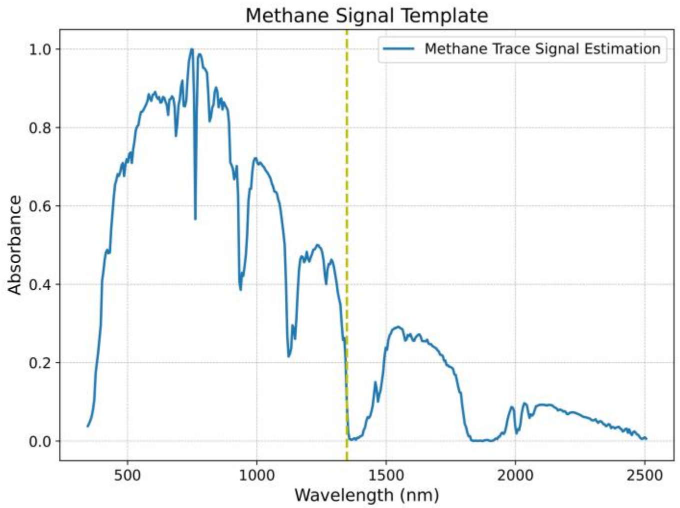


Plume Segmentation Method **Workflow**

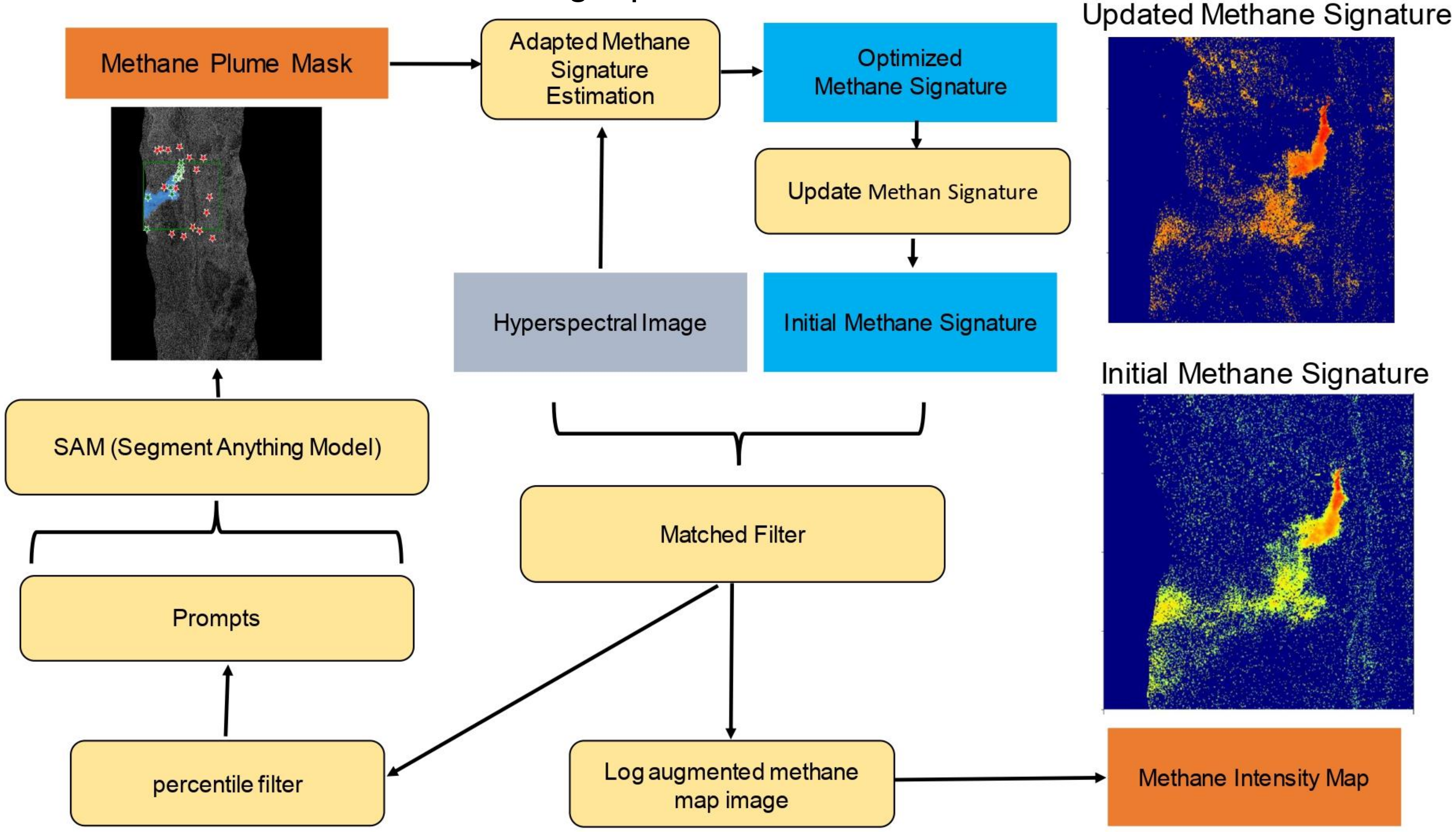
Stacking Operation



Estimated Methane Trace



Stacking Operation



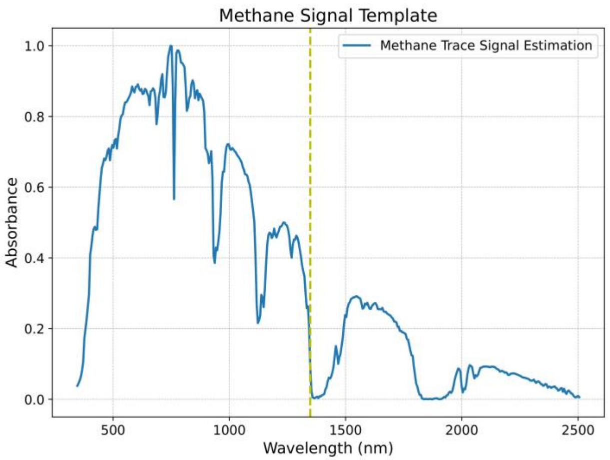
Conclusions

- The proposed sensor placement strategies have enhanced carbon monitoring performance by 20% and demonstrated the ability to meet performance standards with fewer sensors.
- Machine learning techniques have effectively remove noise in subsurface signals, enabling automated processing even when data distribution shifts.
- Advanced machine learning methods applied to remote hyperspectral imaging data have proven effective in segmenting methane plumes and improving filtering accuracy, even without customized training data.

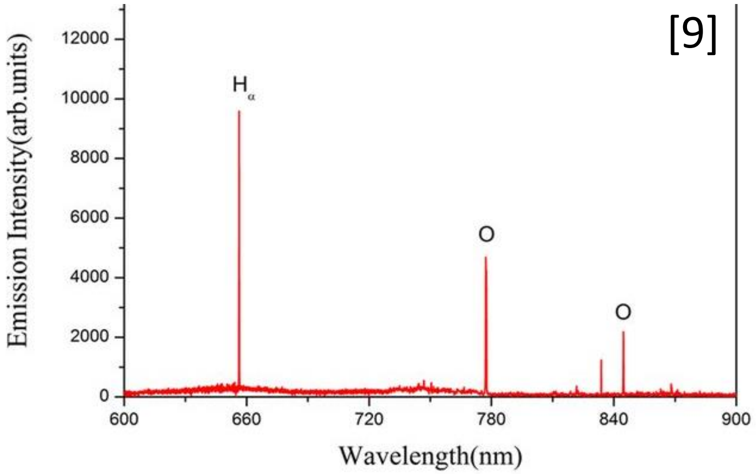
UNIVERSITY of HOUSTON
YOU ARE THE PRIDE
CULLEN COLLEGE of ENGINEERING
Department of Electrical & Computer Engineering

Future Work 1: Remote hydrogen exploration and monitoring

Hyperspectral Trace

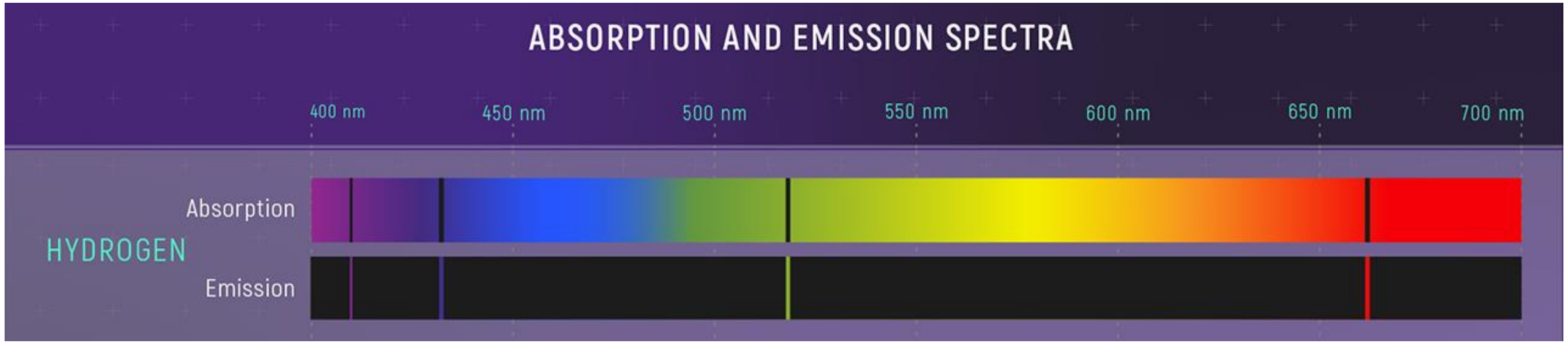


Emission Intensity Spectral Trace



Matched Filter — Segment Anything

Radiance data → H2 detection map → Segmented H2 source map



[10]

[9] ZHANG, J., Dezhi, X.I.A.O., Shidong, F.A.N.G., Xingsheng, S.H.U., Xiao, Z.U.O., Cheng, C., Yuedong, M.E.N.G. and Shouguo, W.A.N.G., 2015. Characteristics of low power CH4/air atmospheric pressure plasma jet. Plasma Science and Technology, 17(3), p.202.
[10] https://webbtelescope.org/contents/media/images/01F8GF9E8WXYS168WRPPK9YHEY

Future Work 2: Hybrid Methane Monitoring System

Hybrid Sensors

Satellite

Drones

Ground Portable gas analyzer

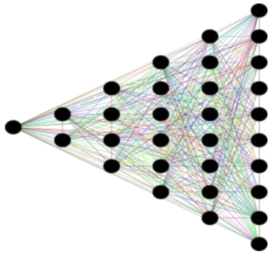
Pointer data

Drones Data

Satellite data

Physical simulation

Neural Network
AI
operator/Optimizer



Optimal sensor placement/ survey scheduling

Human in loop checking

Publications

Journal

1. **Yuan Zi**, Lei Fan, Xuqing Wu, Jiefu Chen, Shirui Wang, and Zhu Han. "Active Gamma-ray Log Pattern Localization with Distributionally Robust Reinforcement Learning." IEEE Transactions on Geoscience and Remote Sensing (2023).
2. Jin, Yuchen, **Yuan Zi**, Wenyi Hu, Yanyan Hu, Xuqing Wu, and Jiefu Chen. "A Robust Learning Method for Low-Frequency Extrapolation in GPR Full Waveform Inversion." IEEE Geoscience and Remote Sensing Letters 19 (2022): 1-5.
3. Jin, Yuchen, Wenyi Hu, Shirui Wang, **Yuan Zi**, Xuqing Wu, and Jiefu Chen. "Efficient progressive transfer learning for full-waveform inversion with extrapolated low-frequency reflection seismic data." IEEE Transactions on Geoscience and Remote Sensing 60 (2021): 1-10.
4. **Yuan Zi**, Lei Fan, Xuqing Wu, Jiefu Chen, and Zhu Han. "Distributionally Robust Optimal Sensor Placement Method for Site-Scale Methane-Emission Monitoring." IEEE Sensors Journal 22, no. 23 (2022): 23403-23412.
5. **Yuan Zi**, Lei Fan, Xuqing Wu, Jiefu Chen, and Zhu Han. "Passive-seismic Sensor Placement Optimization for Geologic Carbon Storage" geoenergy science and engineering 2023

Conference

1. Jin, Yuchen, **Yuan Zi**, Wenyi Hu, Yanyan Hu, Xuqing Wu, and Jiefu Chen. "Solving Full Waveform Inversion Enhanced by Efficient Progressive Transfer Learning." In 2022 United States National Committee of URSI National Radio Science Meeting (USNC-URSI NRSM), pp. 40-41. IEEE, 2022.
2. Jin, Yuchen, **Yuan Zi**, Xuqing Wu, and Jiefu Chen. "An Enhanced GPR FWI Scheme with Low-Frequency Data Extrapolated by Progressive Transfer Learning." In 2022 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium), pp. 120-121. IEEE, 2022.
3. **Yuan Zi**, Lei Fan, Xuqing Wu, Jiefu Chen, Shirui Wang, and Zhu Han. "Active gamma-ray well logging pattern localization with reinforcement learning." In SEG International Exposition and Annual Meeting, p. D011S018R002. SEG, 2022.
4. **Yuan Zi**, Shirui Wang, Pengyu Yuan, Xuqing Wu, Jiefu Chen, and Zhu Han. "Self-supervised learning for seismic swell noise removal." In Second International Meeting for Applied Geoscience & Energy, pp. 1910-1914. Society of Exploration Geophysicists and American Association of Petroleum Geologists, 2022.
5. Jin, Yuchen, **Yuan Zi**, Wenyi Hu, Xuqing Wu, and Jiefu Chen. "A deep learning enhanced full waveform inversion scheme." In 2021 International Applied Computational Electromagnetics Society Symposium (ACES), pp. 1-4. IEEE, 2021.
6. **Yuan Zi**, Wenyi Hu, Jiefu Chen, Xuqing Wu, and Zhu Han. "Physics-guided self-supervised learning for monochromatic noise removal." In SEG International Exposition and Annual Meeting, p. D011S058R002. SEG, 2021.
7. **Yuan Zi**, Lei Fan, Xuqing Wu, Jiefu Chen, and Zhu Han. "Passive-seismic Sensor Placement Optimization for Geologic Carbon Storage" SEG International Exposition and Annual Meeting 2023 accepted
8. **Yuan Zi**, Lei Fan, Xuqing Wu, Jiefu Chen, and Zhu Han. "Segmenting Hyperspectral Images of Methane Plumes with a Large Machine Learning Model" submitted to IMAGE 2024

Thank You

Q&A

UNIVERSITY of HOUSTON
YOU ARE THE PRIDE
CULLEN COLLEGE of ENGINEERING
Department of Electrical & Computer Engineering