UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DEVELOPMENT OF A MACHINE-LEARNING ENHANCED HIGH PERFORMANCE METHANE SENSING INSTRUMENT FOR FIELD APPLICATIONS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

BY

YANG YAN
Norman, Oklahoma
2025

DEVELOPMENT OF A MACHINE-LEARNING ENHANCED HIGH PERFORMANCE METHANE SENSING INSTRUMENT FOR FIELD APPLICATIONS

A THESIS APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Binbin Weng, Chair

Dr. Samuel Cheng, Member

Dr. Chao Lan, Outside Member

Acknowledgments

I would like to express my sincere gratitude to my friends, professors, and classmates who have supported me throughout my undergraduate and graduate studies at the University of Oklahoma. I am especially thankful to my graduate advisor, Dr. Binbin Weng, for his continued guidance and support in both my research and lab work in the School of Electrical and Computer Engineering.

I also extend my appreciation to my committee members, Dr. Chao Lan and Dr. Samuel Cheng, for their valuable advice on my machine learning research, and to our former lab post-doctoral researcher, Dr. Jim Jeffers, for his assistance during the development of the device.

Special thanks go to my lab members, Lkhanaa Mijiddorj, Bek Mijiddorj, and Mark Luke Banquite, for their contributions in building, testing, and field-deploying the device. I also appreciate Tyler Beringer for helping improve the circuit design and enclosure setup, making the final system more robust and professional.

In addition, I am grateful to Lkhanaa Mijiddorj once again, as well as our PSU partners Dr. Bin Li, Zifan Zhou and Xuan Wang, for their contributions to software updates and server related development.

Finally, I sincerely thank my entire family for their continuous understanding and for providing both financial and emotional support throughout my pursuit of a master's degree.

Table of Contents

1 Introduction . . . 1

1.1 Research Background . . . 1

1.2 Research Motivation . . . 2

1.3 Research Objectives . . . 3

2 Technical Review of Relevant Sensing Solutions . . . 6

2.1 Calorimetric sensors . . . 6

2.2 Pyroelectric Sensor . . . 8

2.3 Semiconducting metal oxide sensors . . . 10

2.4 Electrochemical sensors . . . 11

2.5 Optical Sensors . . . 13

2.6 Machine Learning Implementation . . . 15

3 Development of the NDIR-based Sensing Instrument . . . 17

3.1 Overview of the Hardware Design . . . 17

3.2 Operating Hardware Development of the NDIR Sensor . . . 18

3.2.1 The Choice of the NDIR-based Methane Sensing Module . . . 18

3.2.2 Design and Integration of the Microcontroller Control and Data Handling Module . . . 19

3.2.3 Wireless Communication Interface . . . 20

3.2.4 Active Sampling via Gas Pump . . . 21

3.2.5 Power System and Solar Integration . . . 22

3.3 Electric Circuit Design & Breadboard Test . . . 23

3.4 Device Box Set-up . . . 25

4 Machine Learning Enhanced Data Analysis for the NDIR based Sensing Instrument . . . 28

4.1 ML Model Selection . . . 30

4.1.1 Multiple Linear Regression . . . 31

4.1.2 Support Vector Regression . . . 31

4.1.3 Gaussian Process Regression 32

4.1.4 Artificial Neural Networks 32

4.2 Machine Learning Data Training 33

4.2.1 Environment Setup & Data Collection 33

4.2.2 Training Result and Model Selection 36

4.3 In-lab calibration and validation 38

4.4 Outdoor calibration and validation 41

5 Preliminary Demonstration of NDIR-based Sensing Network . . . 46

5.1 Field Deployment of Distributed Sensor Instruments 46

5.2 Field test and validation 48

5.3 Mobile test and validation 49

6 Conclusion and Future Works 56

6.1 Research Summary 56

6.2 Future Work Roadmap 58

List of Abbreviations 61

List of Key Symbols 62

References 62

List of Figures

1.1 Primary Source of Methane Release [1] 2

2.1 Calorimetric gas sensor principle 7
2.2 Schematic of pyroelectric sensor based on infrared heating 8
2.3 Gas sensing mechanism of metal oxides 10
2.4 Schematic of an electrochemical sensor 12
2.5 Principle of Optical Sensor 13

3.1 The Design Architecture for the NDIR-based Methane Sensing Instrument 17
3.2 Diagrams related to the Senseair K96 module 19
3.3 Schematic of the Feather M0 Microcontroller 20
3.4 Sim 7070G modem 21
3.5 Power Consumption 22
3.6 Circuit Design Schematic for the device 24
3.7 Internal structure setup of the two different version device 26
3.8 Real-Time Monitoring Interface for Sensor Network 27

4.1 A field installation of the NDIR sensor instrument 28
4.2 Methane Readings from the Field-Deployed Device 29
4.3 Principle of NDIR Methane Sensor 30
4.4 Block diagram of Training Data Collection Setup 33
4.5 Block diagram of Training Data Collection Setup 34
4.6 Calibration Data 35
4.8 Root Mean Square Error (RMSE) for Training, Validation, and Test Datasets 38
4.9 In-Lab Validation Results Using Original Data Order to Show Methane Peaks 40
4.10 R-squared validation result for the in-lab test 41
4.11 Model's Accuracy on Test Data Set 43
4.12 Balcony test with the AIMNet device and Licor 7810 43
4.13 Result and comparison for balcony test 44

5.1 Field Installation of Field Monitoring System 47
5.2 Device Installation Location 48
5.3 Field Validation Test 48
5.4 Methane Response from Device F030 in Field 49
5.5 Contribution of greenhouse gas emissions by sector 51
5.6 Comparison between Fs002 and Licor-7700 methane measurement . . . 53
5.7 Heat Map for Licor-7700 54
5.8 Heat Map for FS-002 55

Abstract

Methane - a well-known potent greenhouse gas - has significantly contributed to global warming through their emissions from both natural sources, like wetlands, and human activities, including particularly the fossil fuel production sectors. The earlier to detect the emission sources, would allow a fast response to deploy the mitigation strategies to contain methane from releasing the environment. Therefore, making accurate and fast monitoring of methane emission is vastly crucial, particularly in the oil and gas industries where large-scale emissions are prevalent. However, the existing challenges on the field deployable sensing solutions still exist. They are either too costly involving human participation or with unstable or low sensitivity performance to provide an accurate and low false alarm sensing outcomes.

Having this motivation, this thesis presents the development of a new cost effective, high performance and environmentally robust methane sensing instrument based on the NDIR (Nondispersive Infrared) sensing method and machine learning enhancement algorithm, which could be largely distributed to form a real-time methane monitoring network that can be strategically positioned for scalable and comprehensive area coverage, such as from facility-level to production basin or even regional level.

Specifically, the thesis includes six chapters, starting from the background introduction to provide an brief overview of the existing methane emission issues and monitoring studies. Chapter two will provide a systematically review of the point sensor technologies that can be used for creating real-time sensing network in the field to detect methane emission dynamics. Chapter Three will primarily describe the design of the circuit board and the overall setup of the device. After finalizing the design of a single device, Chapter Four will focus on the selection and training of the machine learning models used to process the data and mitigate environmental influences. Chapter Five will present the three validation methods we employed to verify the accuracy of the device's readings: in-lab validation, open-area validation, and field validation.

CHAPTER 1

Introduction

1.1 Research Background

Greenhouse gases trap heat in the Earth's atmosphere, leading to a gradual rise in global temperatures, which has profound implications for ecosystems, weather patterns, sea levels, and human societies, including agriculture, health, and infrastructure. Over the past century, global warming has emerged as one of the most pressing environmental issues worldwide. The primary driver behind this escalating problem is the significant increase in greenhouse gas emissions resulting from human activities from various sources. [2]

The main components of greenhouse gases include carbon dioxide (CO2), methane (CH4), water vapor (H2O), and nitrous oxide (N2O). Among these, CH4 stands out as the second most critical contributor after CO2 to global warming due to its high global warming potential. Methane is approximately 28 times more potent than carbon dioxide over a 100-year period and has an even greater impact—up to 84 times—over a shorter, 20-year timescale. [2] This potency underscores the urgent need to prioritize controlling and mitigating methane emissions to slow the rate of climate change effectively.

Currently, about 30 percent of global methane emissions originate from activities associated with oil, gas, and coal extraction industries. [3] These emissions arise primarily due to leaks, venting, and inefficient combustion processes during extraction, processing, and transportation. Methane leakage not only exacerbates climate change but also represents significant economic losses for the energy industry due to wasted natural resources. Additionally, uncontrolled methane emissions contribute to air pollution, posing health risks to nearby populations and ecosystems.

1.2 Research Motivation

The continued growth and expansion of the oil, gas, and coal industries, coupled with inadequate emission control and outdated monitoring strategies, further aggravate global warming challenges. Particularly for the conventional monitoring methods, they are either too costly involving human participation, having very low discrete scanning frequencies, or with unreliable or low sensitivity performance to provide an timely, accurate and low false alarm sensing results. Consequently, to advance the technical solution with an accurate and continuous capability to monitor methane emissions from these field sectors is obviously essential for developing effective climate management policies, which could lead to effective regulatory compliance, improved resource efficiency, reduced environmental damage, and significant cost savings by identifying and addressing leaks quickly.

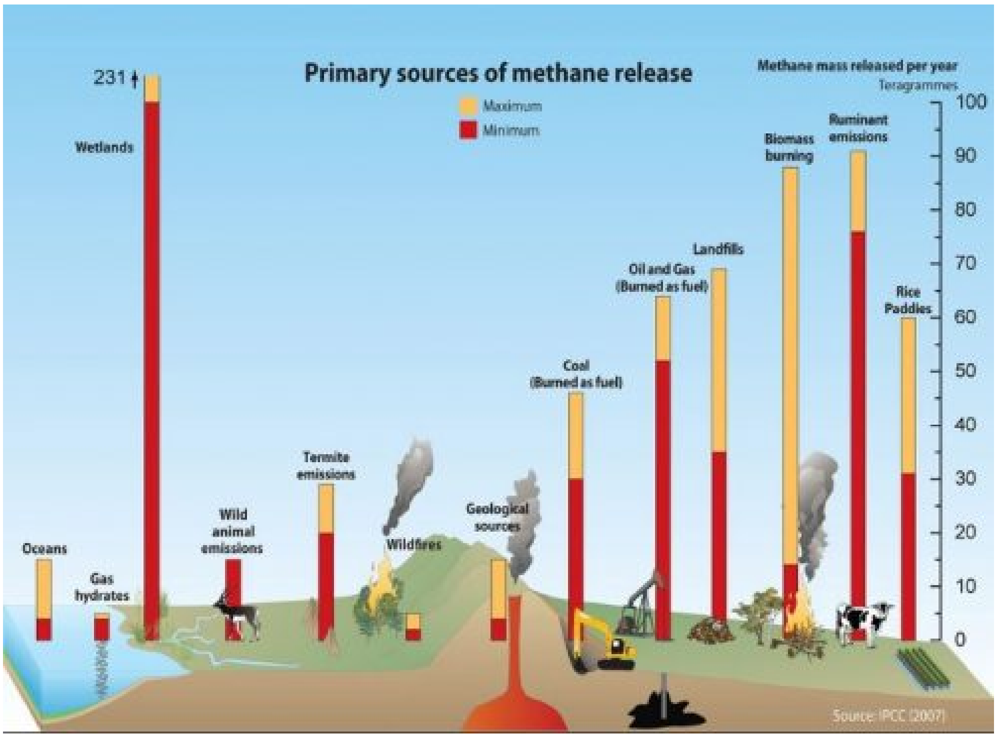


Fig. 1.1: Primary Source of Methane Release [1]

As shown in figure 1.1, most of the the oil and gas extraction operations are located in remote regions, far from urban centers. This geographic isolation poses significant challenges for emissions monitoring and regulation enforcement, making conventional monitoring methods difficult and costly. [3] For example, sending leak inspectors to those locations to detect the potential leaks from valves or pipe connectors can only be done in monthly routine, or even a several-month cycle. However, studies have found that leaks particularly large leaks in the fields are not continuous and mostly are random in time and space domain, thus such a low-frequent visiting style inspection method is likely missing the leak events or largely delay the awareness of those events. Hence, there is a clear need for reliable, durable, and autonomous methane monitoring devices capable of operating independently in harsh and remote environments. To ensure the performance, such devices must also provide precise, real-time methane measurements over extended periods, enabling accurate identification, quantification, and rapid response to emission events. Again, as mentioned previously, addressing this technological need will greatly enhance our capability to manage methane emissions effectively, mitigate their environmental impact, and contribute positively to global climate stabilization efforts.

1.3 Research Objectives

The ultimate goal of the study is to develop a distributive real-time gas monitoring network capable of continuously and precisely monitoring methane emissions over extensive infrastructure areas such oil and gas production fields, landfills or waste water treatment plants, etc. To achieve this, a non-dispersive infrared (NDIR) methane sensor was selected due to its proven sensitivity and reliability. In order to enable its sensing functionality and allowing the real-time data collection over a remote field, the first objective of this thesis work focuses on the design and development of a complete methane sensing hardware and its corresponding software operation system, which include a detailed circuit board design tailored specifically for the NDIR methane sensor, microcontroller programming, and complete system integration, encompassing power supply management and gas handling systems.

Given the need for high precision and reliability in methane detection, the second objective of this work focuses on the implementation of several signal processing techniques and machine learning algorithms to enhance the sensing data accuracy. These methods aim to filter noise and improve measurement accuracy under varying environmental conditions. This thesis will also detail the collection, preprocessing, training, and validation of data for machine learning models. A robust dataset comprising 10,000 data points, incorporating a wide range of temperature, humidity, and methane concentration conditions, will be utilized to train and validate these predictive models, ensuring their accuracy and reliability in operational scenarios. After completing the hardware and machine learning algorithm, the main objective was to apply the design in real-world testing. The first concept involved mounting the device on a vehicle to enable mobile testing. The device should be installed at the front or on top of the vehicle. With a high-precision GPS setup, the device can provide real-time methane readings along with positional data, which can later be used to generate a heatmap of the area. This enables analysis of potential leak points and regions with higher methane concentration. After that, to effectively monitor methane emissions across an entire oil and gas field, an array of 20 individual methane sensing devices are deployed in the objective 3 throughout the study site in the oil and gas production basin in Anadarko, oklahoma for proving the concept. Each device will receive the standard pre-calibration before installation while with the ability to function independently in the field providing real-time methane concentration readings. Data from each device will be collected through an integrated network and transmitted to a central monitoring station via MQTT protocol for remote, continuous data logging and analysis using the machine learning algorithms developed in the prior objective.

Through the completion of these objectives in this thesis, we have developed an robust field deployable methane sensing instrument that can be distributed to form a small scale sensing network to function for capturing methane emission events. The research will continue via scaling up the network's node density and the location coverage, and the future study will focus on improving the sensor's durability and performance in extreme environmental conditions by addressing issues such as signal degradation, transmission limitations, and sensor reliability outside controlled environments. Further enhancements in machine learning algorithms and signal processing filters will be explored to refine data accuracy, ensuring the sensor's performance closely matches that observed under stable laboratory conditions, and to further minimize environmental influences to estimate the emission target points and the volumes.

CHAPTER 2

Technical Review of Relevant Sensing Solutions

As mentioned, our ultimate goal is to develop a point-sensor based distributive methane emission monitoring system. The performance of such a system largely depends on the choice of the point sensor, as it directly influences the accuracy, responsiveness, and consistency of the methane measurements in field test situation. In other words, it is crucial to select a proper sensing technology that is of high performance, cost effective and robust for the real-world implementation. Thus, to make an informed right selection, we need to understand the strengths and limitations of the commonly available sensor types and evaluate how they align with the specific requirements.

At present, four main types of methane sensors are widely used for the point sensing instrument applications [4]: calorimetric sensors, pyroelectric sensors, electrochemical sensors and optical sensors. Each of these technologies has been applied in various gas detection scenarios and offers different levels of performance, cost, and durability. Therefore, in this chapter, I will provide a general overview of these four sensor types, discussing their typical advantages and limitations. The purpose is to establish a clear rationale for selecting the most appropriate sensing method for our following hardware design and development. In addition, we will briefly introduce potential machine learning strategies that can enhance the overall instrument performance for better field operation and sensing reliability.

2.1 Calorimetric sensors

The calorimetric thermoelectric gas sensor (TGS) integrates catalytic combustion and thermoelectric conversion to detect combustible gases such as methane (CH_4). Such a device typically adopts a dual-catalyst structure[5]: Pd/θ-Al_2O_3 (θ: metastable phase [6]) is coated on the hot side and Pt/α-Al_2O_3 (α: stable phase [6]) on the cold side. As shown in figure 2.1 Methane and hmmydrogen combust exothermically on the hot side, generating a local temperature increase, while hydrogen combustion also occurs on the cold side, reducing the temperature difference. This thermal gradient is then converted into a voltage output by the Seebeck effect [7] across the thermoelectric thin film.

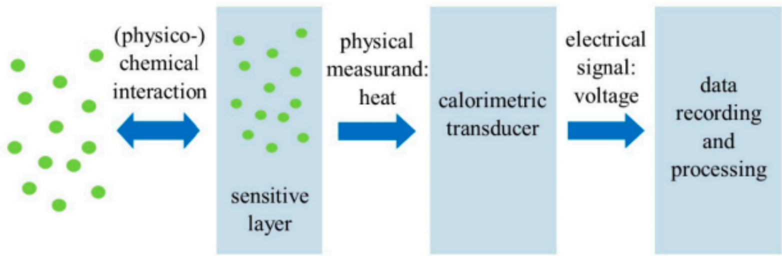


Fig. 2.1: Calorimetric gas sensor principle

Mathematically speaking, the voltage response ΔV is proportional to the temperature difference ΔT between the two sides [8], governed by:

$$\Delta V = \alpha \cdot \Delta T \tag{2.1}$$

where the α here is the Seebeck coefficient of the thermoelectric material - different from the one in the prior paragraph used to describe the stable phase of sapphire. And the response voltage is linearly correlated with the combustion heat Q:

$$Q = K \cdot \Delta V \tag{2.2}$$

where K is the device-specific conversion factor in J/mV.

This design typically can enable the detection of CH_4 in the range of 200–2000 ppm at operating temperatures below 400°C. If a dual-catalyst strategy is applied, it can significantly suppress hydrogen's dominance in total heat release, improving methane detection accuracy. Further, by employing additional wt% Pt/α-Al_2O_3, could make the device achieve up to 50% reduction in H_2 response while maintaining linearity for

CH_4, highlighting its balance and resolution capability [5].

Calorimetric sensors offer advantages such as simple design, low manufacturing costs, and mechanical robustness. These sensors are widely used in industrial applications due to their portability and reliability. However, they typically require high operational temperatures, resulting in significant power consumption. Moreover, calorimetric sensors lack the sensitivity needed to detect small variations in methane concentration (e.g., changes as small as a single digit ppm level or even lower), which is a critical requirement for our system. Furthermore, their performance may degrade over time due to catalyst poisoning or saturation, leading to reduced sensitivity and accuracy. This degradation necessitates frequent calibration and maintenance, which further increases the operational complexity and cost, as well as reduces its long-term reliability [4]. Given our goal to develop a low-cost, durable, and sensitive methane detection system capable of detecting changes as small as 1 ppm and performing long-term outdoor monitoring, calorimetric sensors is not a suitable selection for this project.

2.2 Pyroelectric Sensor

The pyroelectric methane sensor 2.2 operates based on the principle that certain dielectric materials, such as lithium tantalate ($LiTaO_3$), generate an electric charge in response to a change in temperature.

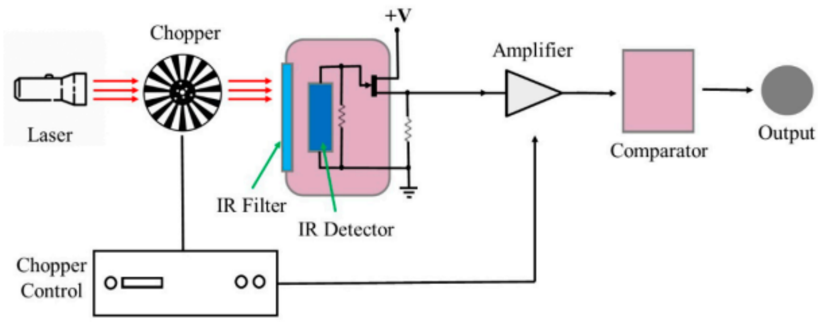


Fig. 2.2: Schematic of pyroelectric sensor based on infrared heating

As shown in figure 2.2, when infrared radiation, modulated at a certain frequency, passes through a gas chamber containing methane, the gas absorbs specific wavelengths (notably around 3.4 μm), following the Beer–Lambert law:

$$I = I_0 \cdot \exp(-\alpha \cdot C \cdot L) \tag{2.3}$$

where I_0 and I are the incident and transmitted light intensities, α is the absorption coefficient, C is the gas concentration, and L is the optical path length.

The pyroelectric element absorbs the transmitted radiation and converts the resultant temperature change ΔT into an electric current according to the equation:

$$I_p = p \cdot A \cdot \frac{dT}{dt} \tag{2.4}$$

where p is the pyroelectric coefficient, A is the sensitive area, and $\frac{dT}{dt}$ is the rate of temperature change. The output voltage is expressed as:

$$V = \frac{I_p}{\omega C} \tag{2.5}$$

with ω being the modulation frequency and C the capacitance of the element.

In a 2024 study *ACS Omega*, a MEMS-based pyroelectric sensor integrated with a Fabry–Pérot filter and signal processing via wavelet algorithms achieved a detection limit as low as 5.6 ppmv, with a fairly quick response time ($\mathrm{t}_{90} < 3\,\mathrm{s}$) and a regression accuracy of $R^2 = 0.982$. This demonstrates the sensor's good sensitivity, relatively fast dynamics, and robustness for real-time methane monitoring applications.[9] However, there are still critical limitations for the field distributive network focusing application. Firstly, pyroelectric methane sensors are typically very expensive, which makes them hard to be considered for large-scale or cost-sensitive deployments. Secondly, due to their structural complexity—including delicate optical components, infrared filters, and MEMS elements — they are highly susceptible to damage when exposed to harsh environmental conditions in the field, such as high humidity, dust, mechanical shock, or temperature fluctuations. These vulnerabilities significantly reduce their durability and long-term stability, making them a suboptimal choice for robust outdoor sensing platforms.

2.3 Semiconducting metal oxide sensors

Semiconducting metal oxide (SMOx) methane sensors operate based on the chemiresistive effect, where the presence of methane induces a change in the sensor's electrical resistance. Typically fabricated from n-type metal oxides such as SnO_2, ZnO, or WO_3, the sensor surface initially adsorbs oxygen molecules from ambient air, as shown in figure 2.3, which capture free electrons from the conduction band and form negatively charged species such as O, O^2. This process leads to the formation of a surface depletion layer, resulting in an increased baseline resistance.[10]

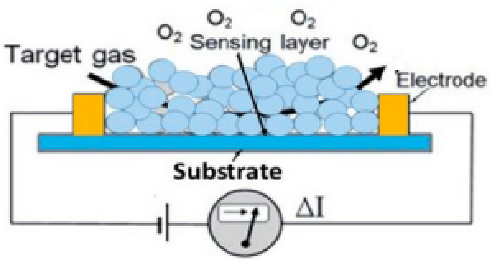


Fig. 2.3: Gas sensing mechanism of metal oxides

When exposed to reducing gases like methane (CH_4), the gas molecules react with the adsorbed oxygen species on the sensor surface, releasing trapped electrons back to the conduction band. This reduces the resistance of the sensing material, which can be quantitatively expressed by the sensitivity equation:

$$S = \frac{R_g}{R_a}, \tag{2.6}$$

where R_g is the resistance in the presence of methane, and R_a is the resistance in clean air.

The surface redox reaction can be summarized as:

$$\mathrm{CH_4} + 4O^{-}(\mathrm{ads}) \rightarrow \mathrm{CO_2} + 2\mathrm{H_2O} + 4e^{-}. \tag{2.7}$$

This type of sensors are attractive due to their low cost, fast response time, and ease of fabrication. However, they generally require elevated operating temperatures (200–450°C), leading to high power consumption. In addition, these sensors exhibit strong cross-sensitivity to humidity and other reducing gases, which can compromise their long-term stability and selectivity. To improve performance, enhancements such as noble metal doping (e.g., Pt, Pd) and nanostructuring (e.g., nanowires or porous films) are often employed.

Despite these improvements, such sensors still struggle from their unpredictible highly sensitive challenges to the environmental fluctuations. In our application, the field conditions involve significant changes in both humidity and temperature, making it difficult for SMOx sensors to maintain consistent accuracy and stability. Therefore, due to their vulnerability to such environmental factors, SMOx sensors are not a suitable choice for our system.

2.4 Electrochemical sensors

Electrochemical methane sensors operate based on gas-phase oxidation reactions occurring at the working electrode surface, which are facilitated in the presence of a suitable electrolyte and reference electrode. In the configuration described by Sekhar et al.[11], the sensor utilizes a mixed-potential type mechanism involving a platinum/indium tin oxide (Pt/ITO) electrode coupled with a yttria-stabilized zirconia (YSZ) solid electrolyte, as shown in figure 2.4.

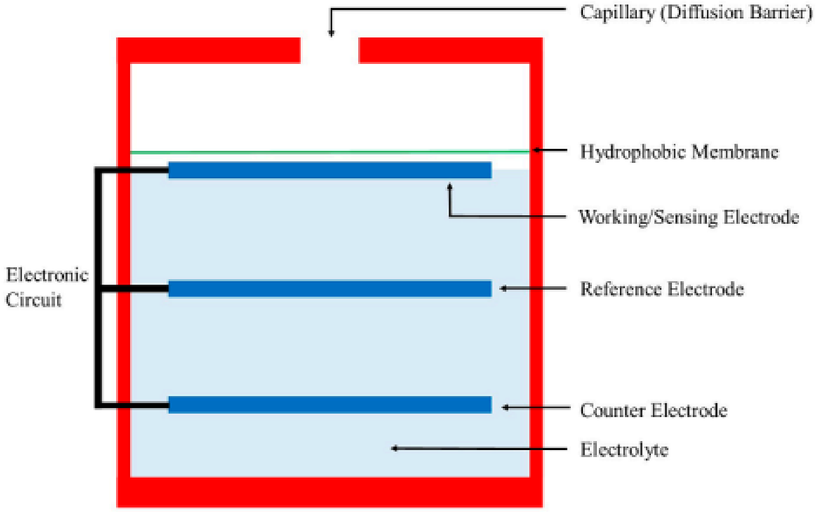


Fig. 2.4: Schematic of an electrochemical sensor

When methane diffuses through the gas-permeable membrane and reaches the electrode-electrolyte interface, it undergoes electrochemical oxidation, producing CO_2 and H_2O. This reaction generates a measurable voltage potential due to the imbalance of reaction rates at the working and reference electrodes. The output voltage, V_{out}, exhibits a logarithmic relationship with methane concentration C_{CH_4} as follows:

$$V_{\text{out}} = K \cdot \ln(C_{\text{CH}_4}) + B \tag{2.8}$$

where K and B are empirical constants dependent on the electrode materials and system configuration.

This sensor type offers low power consumption, a compact form factor, and a rapid response time ($t_{90} < 15$ seconds). In tests with 100 ppm of CH_4, it produced a voltage response of approximately 44 mV, indicating high sensitivity. These characteristics make it one of the most commonly selected options for methane monitoring applications. However, its performance is prone to drift over time and is significantly affected by environmental factors such as humidity and temperature. In outdoor conditions, these influences can accelerate sensor drift and lead to unstable readings that are difficult to recover or recalibrate. Therefore, despite its advantages, this sensor is not suitable either for our sensing network development goal.

2.5 Optical Sensors

The optical sensor specifically, the non-dispersive infrared (NDIR) methane sensors, similar to pyroelectric sensors, also operate under the Beer–Lambert gas absorption law, but with a different signal collection mechanism to detect the transmitted infrared intensity as shown in Figure 2.5, which can be directly correlated gas concentration without the intermediate light to heat conversion process that pyroelectric sensors require.

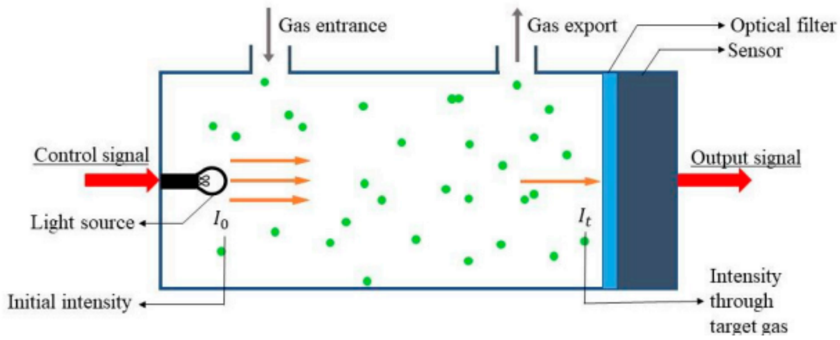


Fig. 2.5: Principle of Optical Sensor

The governing equation is given by:

$$I = I_0 \exp(-\alpha c L) \tag{2.9}$$

where I_0 is the incident light intensity, I is the transmitted light intensity, α is the absorption coefficient of methane at a specific wavelength, c is the gas concentration, and L is the optical path length.

In the work by Zhu et al. [12], an enhanced NDIR methane sensor was developed using a modulated IR source combined with a single-frequency Fast Fourier Transform (FFT) filter to improve the signal-to-noise ratio (SNR). The proposed system achieved a detection limit as low as 1 ppm, significantly outperforming conventional NDIR designs. The optical path was optimized to enhance absorption while maintaining compactness. Additionally, real-time digital filtering at the modulation frequency helped isolate the methane absorption signal, effectively suppressing environmental noise and drift. This design demonstrates that integrating spectral filtering with advanced signal processing can substantially improve the detection resolution and stability of low-concentration methane monitoring systems.

NDIR sensors offer several advantages, including high selectivity, low cost, ease of maintenance, minimal long-term drift, and reliable performance under a wide range of environmental conditions. These sensors are capable of accurately measuring methane concentrations from sub parts per million (ppm) up to percentage levels, making them suitable for both low-level detection and higher-range monitoring.

However, as a trade-off for their low-cost nature, NDIR sensors can be affected by interference from other gases that share overlapping infrared absorption bands. This necessitates careful calibration and often the use of reference channels to enhance measurement accuracy. Additionally, the performance of the sensor's photodiode and infrared light source is temperature-dependent, which becomes a critical factor when operating in outdoor environments where temperature fluctuations can be rapid and significant.

Fortunately, many of these fluctuations can be modeled using physical principles, providing a data processing pathway for compensating the resulting noise and false signals through algorithmic correction, particularly the fast growing machine learning techniques. Given its balance of affordability, accuracy, and adaptability—especially when paired with environmental compensation techniques — NDIR technology presents the most promising solution for formulating our methane network system.

2.6 Machine Learning Implementation

Based on the discussion above, we have chosen the optical methane sensor—specifically, the NDIR type — as the most suitable option. But at the same time, given the need for deployment in outdoor environments, it is essential to address the limitations of NDIR sensors, which are often influenced by environmental factors such as temperature, humidity, and pressure. Instead of relying solely on traditional physical equations or conventional signal processing techniques which are difficult to be comprehended in such complex situation, recent advancements in machine learning (ML) and neural networks (NNs) have demonstrated strong potential in mitigating these environmental impacts and enhancing the performance of outdoor methane sensing systems.

Unlike fixed calibration methods, ML models can dynamically adapt to changing environmental conditions and correct for sensor drift over time. For example, Furuta et al. (2024) integrated linear regression into a low-cost sensor system and significantly improved methane concentration estimation, achieving a root-mean-square error (RMSE) below 0.6 ppm under ambient air conditions [13]. This level of accuracy is crucial for outdoor monitoring, where environmental noise is often unavoidable. In a complementary study, Lin et al. (2023) demonstrated how data-driven approaches—such as random forests and support vector machines—can effectively identify and compensate for cross-sensitivities, particularly those caused by coexisting gases like carbon monoxide (CO) or fluctuations in humidity [14]. These supervised learning models utilize co-measured parameters (e.g., temperature, relative humidity, and CO_2) to reconstruct stable and accurate methane readings, even when raw sensor outputs are unstable [13].

Clearly, leveraging the use of ML and NNs as effective tools for improving the sensor's robustness, adaptability, and long-term reliability, will enpower our capability to maximize the eventual sensing network performance in our project. Detailed work will be elaborated in chapter 4. In which, we will explore and compare several machine learning models—including support vector machines (SVM) and random forests (RF), as highlighted in related literature—to determine the most suitable approach for enhancing the outdoor performance of our NDIR-based system. But before that, in the following chapter, my focus will be on the sensing instrument hardware development.

CHAPTER 3

Development of the NDIR-based Sensing Instrument

3.1 Overview of the Hardware Design

In this chapter, I will present the hardware development effort to create a robust NDIR-based gas sensing instrument that can detect the gas concentration at a single digit ppm level. As shown in figure 3.1, the initial overall design architecture of this sensing instrument comes with several critical components to enable functions including the data collection, data storage, data transmission.

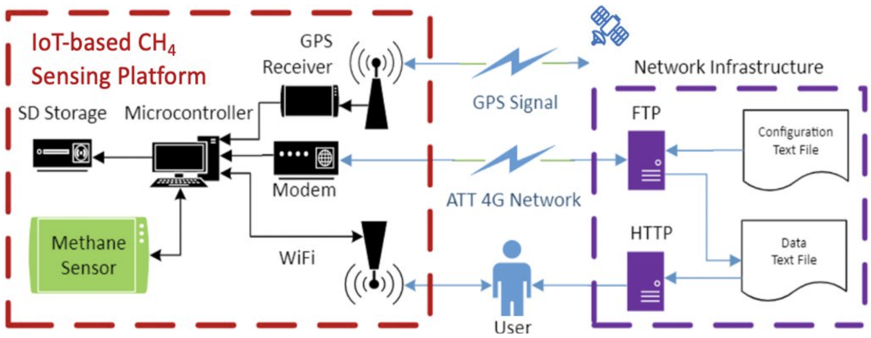


Fig. 3.1: The Design Architecture for the NDIR-based Methane Sensing Instrument

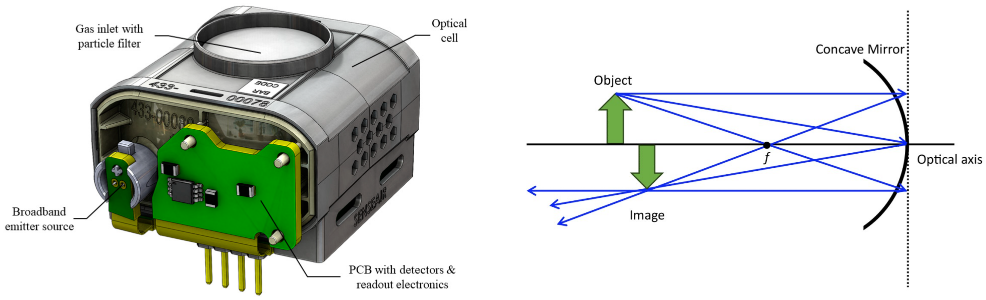
To enable the mentioned functions, the instrument integrates a commercial-ready NDIR sensing module connected to a low-power-consuming microcontroller, which handles data acquisition, processing, and logic control. A GPS module provides the location information for each instrument, and all sensor readings are stored locally on an SD card to ensure data retention even when network connectivity is unavailable. For communication, the instrument includes both a 4G LTE modem and a Wi-Fi module, allowing for flexible data transmission and remote access. Data can be sent to a central server using either FTP or HTTP protocols, where it is stored and analyzed. It is noted that through our development, we identify a better protocol to securely tranmit the data which uses MQTT instead of FTP and HTTP. It will be elaborated later. Beyond the functionality aspect, to support long-term autonomous operation, the system also includes a solar-powered charging mechanism, which will be presented later in this chapter. Lastly, designed with modularity and power efficiency in mind, this architecture will also be housed in a weather-proof case to ensure reliable performance in harsh field conditions.

3.2 Operating Hardware Development of the NDIR Sensor

3.2.1 The Choice of the NDIR-based Methane Sensing Module

At the core of the system lies the methane detection unit. As discussed in the previous chapter, the NDIR sensor was identified as the most suitable solution for our application. However, the sensitivity of most of the market avaiable NDIR sensor can barely reach to the level below 10 ppm when sensing methane gas. The major limitation comes with the limited optical path these devices can offer. To address this issue, our team identified a to-be-commercialized NDIR module manufactured by the company SenseAir in Sweden. They introduced an unique intra-cavity optical mirror reflector design [15] into their NDIR gas sensor which elongates the effective optical path by one order of magnitude (i.e., around 28cm) without physically increase the device size as shown in Fig. 3.2. Through which, this sensor was able to reach its high laboratory-demonstrated precision of 0.5 ppm. Such high sensitivity enables the detection of fairly small variations of ambient methane concentration particularly in the oil and gas land, which could meet — and in many cases exceed — the accuracy requirements for our demanded to detect large leak of methane from production sites.

Furthermore, the K96 sensor supports robust digital communication protocols, facilitating fast and seamless integration with microcontrollers. It also provides the sensing signals of water and CO2 which allow us to correlate the methane sensing signal with these channels of data to decouple the signal fluctuations related to the humidity and temperature out and thus improve its accuracy to detect methane. In other words, through our system engineering, the instrument running this sensor module could operate reliably across a wide range of environmental temperatures and humidity levels which is suitable for outdoor deployment in real-world conditions [16].



(a) A diagram of the Senseair's K96 methane sensing module

(b) Principle of concave mirror inside the sensor

Fig. 3.2: Diagrams related to the Senseair K96 module

3.2.2 Design and Integration of the Microcontroller Control and Data Handling Module

To support sensor operation and enable autonomous field data acquisition, a capable and energy-efficient microcontroller unit (MCU) is required. The Adafruit Feather M0 was selected due to its robust features, including an integrated SD card slot for data storage, multiple I/O channels, and low-power operation capabilities. The microcontroller is tasked with managing the power cycles of the NDIR sensor, collecting sensor readings at scheduled intervals, applying basic preprocessing routines (e.g., averaging, timestamping), and storing the resulting data to the onboard SD card.

In addition, the microcontroller facilitates serial communication with the methane sensor. This connection follows the standard digital communication protocol supported by the Senseair sensor, ensuring compatibility and low-latency data transmission. The microcontroller also supports over-the-air (OTA) configuration updates, which is advantageous for remote firmware management in field deployments.

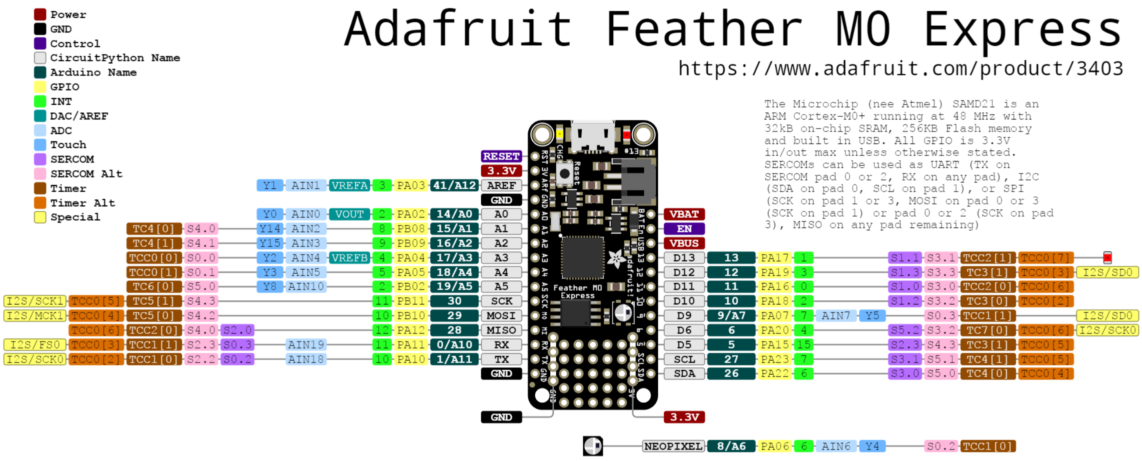


Fig. 3.3: Schematic of the Feather M0 Microcontroller

The inclusion of an SD card-based storage solution allows the system to function independently of real-time network connectivity. This design ensures data is not lost during transmission outages and provides a reliable backup for post-deployment data analysis.

3.2.3 Wireless Communication Interface

Real-time data transmission is often essential for continuous monitoring applications, especially when methane leak detection or emissions reporting is required. To meet this need, a cellular communication module was integrated into the system, enabling over-the-air data transfer to a central server. Given the wide geographical spread and limited infrastructure expected in deployment zones, cellular IoT solutions were prioritized over short-range or Wi-Fi-based alternatives.

Among the modules evaluated, the SIM7070G was selected for integration. This multi-mode module supports LTE Cat-M1, NB-IoT, and EGPRS, which makes it suitable for low-bandwidth IoT applications in rural or underdeveloped areas. Following a comparison of available cellular modules, the SIM7070G demonstrated superior stability and coverage relative to its price point, while maintaining low power consumption during idle and active transmission states.

This communication interface enables the system to either push data periodically to a remote server or respond to polling requests, depending on deployment configuration. The module is managed via AT commands issued by the microcontroller, and supports secure data transmission protocols.

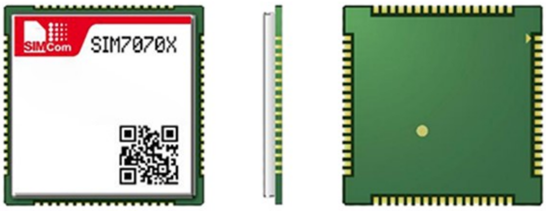


Fig. 3.4: Sim 7070G modem

3.2.4 Active Sampling via Gas Pump

To ensure consistent sensor performance and representative gas sampling, an active airflow system was incorporated into the design. Passive diffusion may be inadequate in windy or stagnant environments; therefore, a miniature gas pump is used to draw air into the sensor chamber at a fixed flow rate.

Based on a comparative evaluation of available pumps in the market, we selected a model that offers low flow rate with high power efficiency, aligning well with the requirements of our device. The chosen pump was selected for its optimal balance between flow stability, noise level, power consumption, and mechanical durability. It operates at 1170 mW and maintains a consistent flow rate of 1.5 L/min, which complies with manufacturer recommendations for typical gas sensor chamber throughput.

This design decision is particularly critical for sensors such as NDIR units, which depend on a steady and continuous gas flow to ensure accurate and repeatable measurements. A stable flow minimizes signal fluctuations and enhances the sensor's ability to detect subtle changes in methane concentration.

3.2.5 Power System and Solar Integration

A fundamental requirement for long-term field deployment is the ability to operate autonomously without frequent maintenance or recharging. To fulfill this requirement, the system includes an integrated solar power supply combined with a high-capacity rechargeable lithium polymer (Li-Po) battery. The solar panel is responsible for recharging the battery during daylight hours, while the battery supplies power to the system continuously.

Each system component has distinct power requirements. The methane sensor typically operates at 5V with low current draw, while the microcontroller and communication module may have peaks of higher current during transmission cycles. The gas pump adds additional load during active sampling phases. According to the estimated consumption values compiled in form 3.5, the system's continuous power draw is calculated to be approximately 1404mW.

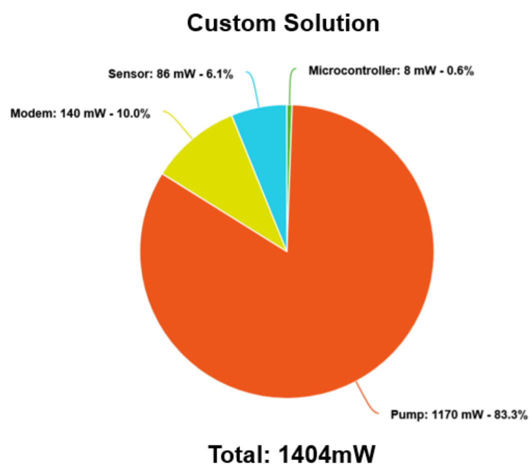


Fig. 3.5: Power Consumption

After evaluating a range of batteries for energy density, discharge rate, and environmental durability, the configuration selected includes a 153.6 Wh battery paired with a 50 W solar panel. Based on energy balance calculations, this system configuration is expected to support up to 4 days of operation without sunlight, meeting the redundancy requirements for off-grid reliability. A charge controller is used to regulate the solar charging cycle and prevent battery over-discharge.

3.3 Electric Circuit Design & Breadboard Test

To support the integration of all electrical components within a compact and robust system, we designed a custom circuit layout tailored to the specific requirements of the methane monitoring device. The process began with the creation of a detailed circuit diagram, illustrated in Graph 1, which guided the functional distribution and power delivery across the various modules.

A key feature in the circuit is the implementation of a buck converter to regulate voltage from a single 12V power supply. This converter is responsible for stepping down the voltage to 7V, 5V, and 3.3V, which are the respective operating voltages required for the methane sensor, SIM communication module, and microcontroller. Each voltage rail is isolated and routed through appropriate paths to ensure component safety and to minimize power loss due to mismatched voltage levels.

Communication between components is established through standard protocol connections. The SCL (Serial Clock Line) and SDA (Serial Data Line) from the microcontroller are connected directly to the Sensirion methane sensor using an I^2C interface, enabling digital communication for sensor readings and configuration. In addition, an analog output pin on the microcontroller is linked to the SIM communication module to manage its power and signal control states during operation.

All modules are powered via their respective voltage inputs and share a common ground. Special care was taken to prevent ground loops and signal interference by carefully segmenting analog and digital sections of the circuit. Furthermore, a potentiometer is added to regulate the voltage supplied to both the system fan and the

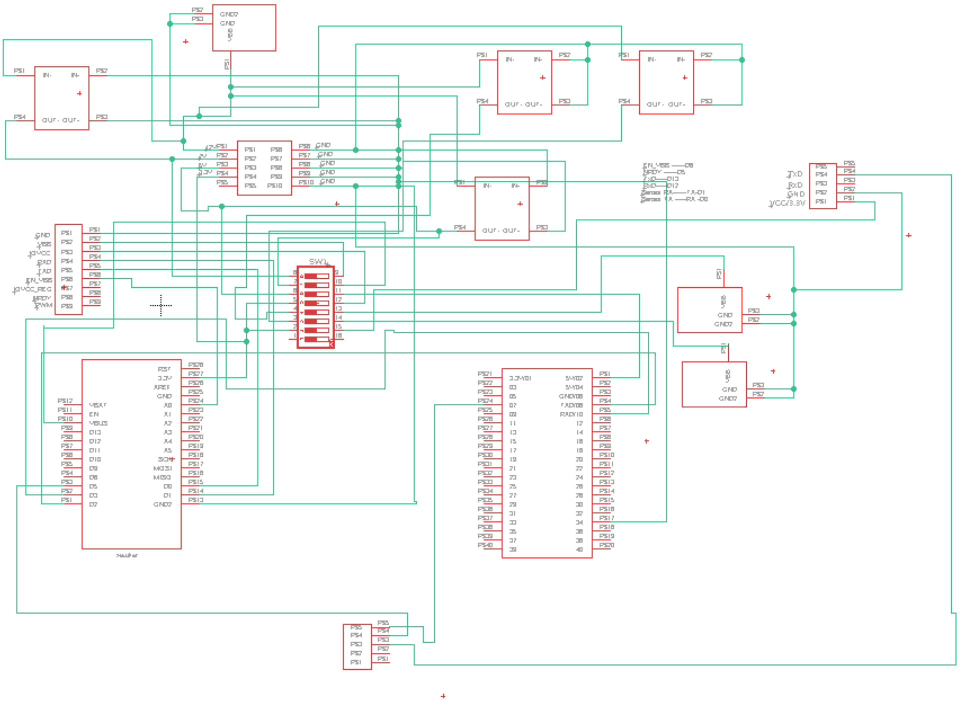


Fig. 3.6: Circuit Design Schematic for the device

gas pump. This allows manual adjustment of the air system's flow rate and supports dynamic cooling control, particularly important for maintaining stable environmental conditions inside the enclosure.

Once the circuit design was finalized and tested on a breadboard, we transitioned the layout into a formal block diagram, followed by the development of a dedicated printed circuit board (PCB) layout, as shown in Graph 4.4. This initial PCB design faithfully replicates the tested connections and module placements from the breadboard prototype. The first version of the PCB prioritized modularity and verification, with some components still mounted externally for flexibility during early testing.

In subsequent development phases, we refined and optimized the board design through several updated versions. The latest iteration integrates the buck converter directly onto the PCB, thereby eliminating the need for an external power regulation module. Additionally, we removed the independent control boards previously used for the SIM module and microcontroller, opting instead to embed these components directly into the PCB design. This integration significantly reduces system size and simplifies the overall assembly process.

Our overarching goal in this iterative hardware design process is to produce a fully self-contained control board that houses all critical components in a single structure. This design reduces the complexity of manual wiring, minimizes connection errors during assembly, and enhances the mechanical stability and durability of the sensing device when deployed in the field.

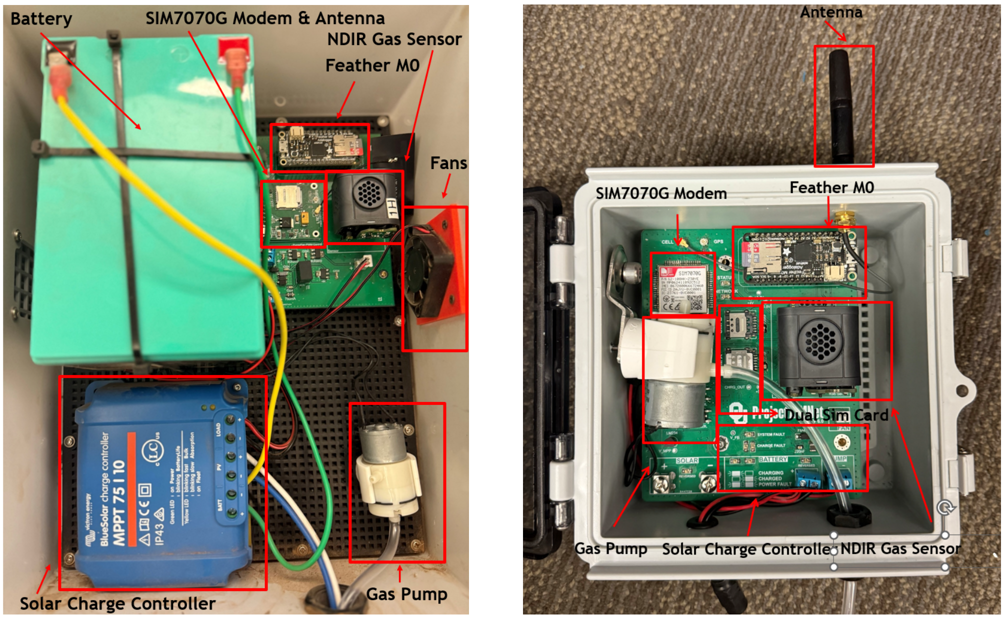
3.4 Device Box Set-up

In the first version of our enclosure setup, we mounted all core components—including the soldered circuit board, rechargeable battery, solar charging controller, cooling fan, and gas pump—directly inside a plastic box using zip ties and screws for mechanical support. A small 3D-printed case was added to house the NDIR sensor, along with a dust filter positioned at the inlet to prevent particulate contamination. For ease of assembly, all electronic components were either plug-and-play or designed to be easily soldered, allowing rapid prototyping and on-site replacements.

Following field deployment of the initial version, several limitations became apparent, prompting design improvements in the second iteration. To reduce overall system cost, weight, and assembly complexity, we integrated the solar charge controller directly onto the main PCB, eliminating the need for a separate breakout board. Additionally, a buck converter was embedded onto the PCB to handle voltage step-down regulation.

The Adafruit breakout modules used in the first version, particularly for communication modules and sensors, were prone to mechanical instability due to their plug-in form factor. To improve reliability, these components were replaced with surface-mount equivalents. Specifically, the SIM card module was transitioned to a surface-mount version, as the plug-in design frequently experienced pin damage during field use.

Moreover, field testing revealed that an unsealed enclosure was highly vulnerable to environmental intrusion—particularly dust and insects—which led to potential sensor malfunction. As a result, we adopted a fully sealed enclosure design and implemented corresponding circuit board modifications, as illustrated in Fig. ??. These upgrades significantly enhanced durability and reliability for extended outdoor deployment.



(a) Unsealed Version **(b)** Sealed Version

Fig. 3.7: Internal structure setup of the two different version device

After confirming that the sensor design was complete, we utilized the built-in MQTT communication protocol to enable real-time data transmission from field-deployed devices. In collaboration with our partner teams at the University of Oklahoma (OU) and Portland State University (PSU), a web-based visualization platform was developed. This platform allows real-time monitoring of methane concentration and environmental data collected from the field, significantly improving our ability to analyze and interpret system performance under real-world conditions.

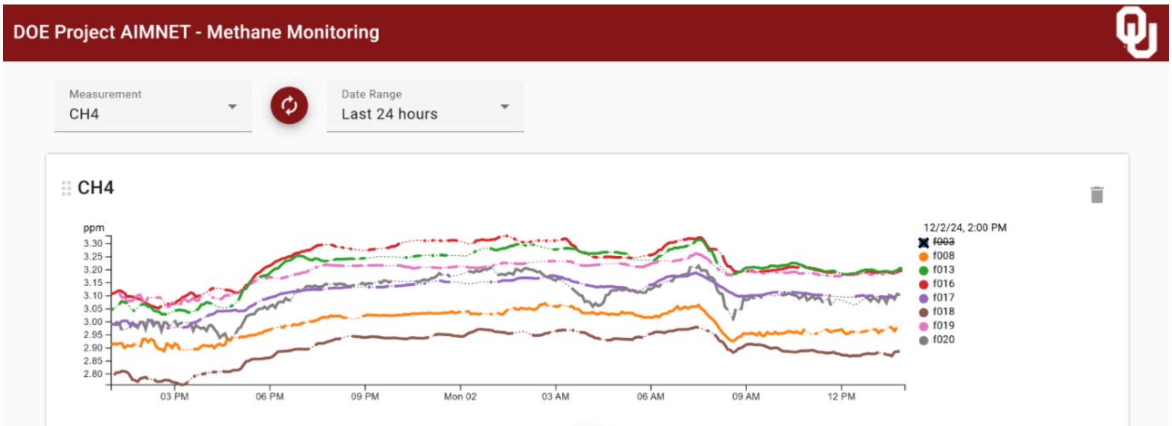


Fig. 3.8: Real-Time Monitoring Interface for Sensor Network

In addition to real-time monitoring, the system also supports direct data export, which allows us to collect historical datasets for training and validating future signal processing or machine learning models. This continuous data pipeline is critical for long-term system optimization and adaptive calibration strategies.

CHAPTER 4

Machine Learning Enhanced Data Analysis for the NDIR based Sensing Instrument

After completing the development of the sensor hardware, we began installing the sensor in open-field environments under real-world outdoor conditions to conduct preliminary performance assessments. Figure 4.1 shows my groupmate Dr. Jim Jeffers and me are installing a set of complete sensor instrument which includes the solar panel power supply together with the sensor hardware detailed in the previous chapter. This is the first outdoor NDIR sensing system we installed back in 2024, and it is located in the oil and gas facility that our industrial partner Devon Energy owns over the Andarko basin. It is worth noting that after the installation, this device had been running continuously for months with no disruption until got removed when a new system upgrade was needed.

Fig. 4.1: A field installation of the NDIR sensor instrument

Nevertheless, through the data we collected, we unfortunately observed that much of the raw data from the sensors was hard to be understood or in other words, not immediately meaningful. As illustrated in Figure 4.2, the recorded methane readings were significantly affected by environmental factors. The sensor readings fluctuated in tandem with variations in temperature, humidity, and other ambient conditions, rather than reflecting actual changes in methane concentration. According to the sensor configuration, a 1-unit change in the raw output corresponds to approximately

0.5 ppm of methane.

However, as shown in the graph, abnormal fluctuations were observed, which appeared to strongly correlate with temperature changes. Further data analysis revealed that humidity and pressure also contributed to these deviations. The influence of environmental factors on the sensor was substantial, effectively masking the true methane signal and significantly reducing the reliability of the raw measurements. This finding highlights the limitation of using NDIR sensors in outdoor environments. Although NDIR technology is known for its robustness, the results demonstrate that its readings can still be heavily influenced by environmental variability—especially when methane concentrations are below 30 ppm. This level of interference significantly reduces the reliability of the raw sensor output and falls short of the precision required for our monitoring objectives. Therefore, further data processing or correction mechanisms are necessary to extract meaningful insights from the sensor data.

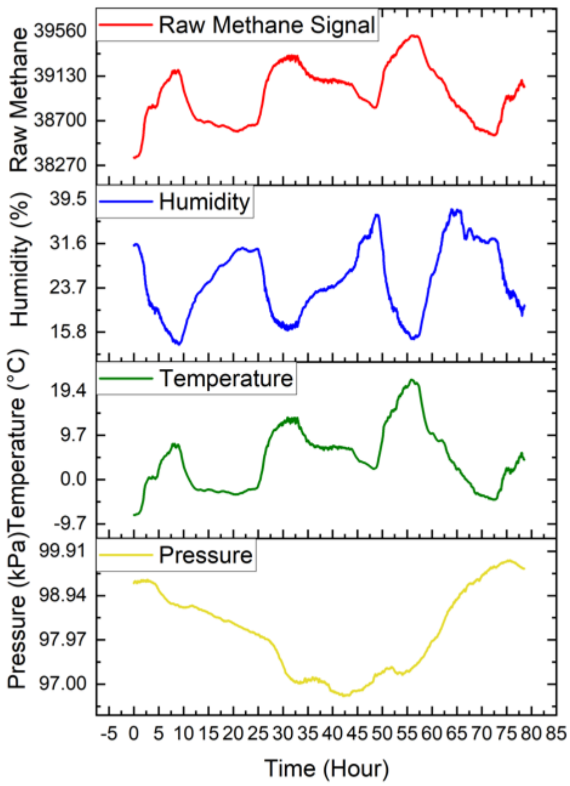


Fig. 4.2: Methane Readings from the Field-Deployed Device

From the figure 4.2, One of the most critical issues observed is the effect of humidity and water vapor on the sensor's performance. As illustrated in figure 4.3, the presence of water particles inside the sensor chamber can interfere with the infrared measurement mechanism. Specifically, when water vapor enters the sensing region, it may partially block or scatter the infrared light beam. This phenomenon can cause the sensor to falsely interpret the absorption signature as methane, leading to erroneous concentration readings. Such interference is particularly problematic in high-humidity environments or during rapid temperature changes that cause condensation within the device.

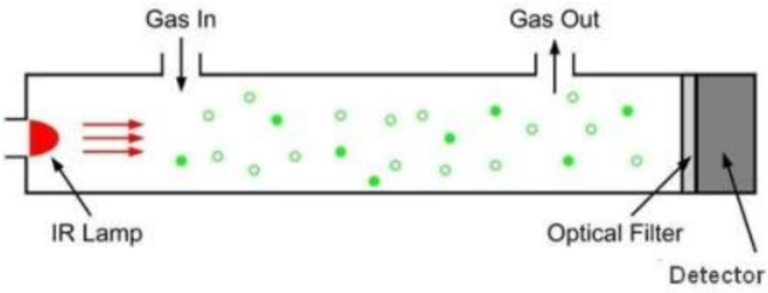


Fig. 4.3: Principle of NDIR Methane Sensor

In addition to humidity, ambient pressure and temperature also exert a strong influence on the sensor output. Variations in atmospheric pressure can alter the density of the air sample inside the sensor chamber, which in turn affects the optical path length and infrared absorption. Similarly, temperature fluctuations can shift the calibration curve of the sensor, especially since NDIR sensors are sensitive to thermal drift. These factors combined result in considerable noise and bias in the raw sensor data, which must be compensated for through robust correction methods.

To overcome these limitations, we integrated machine learning (ML) models into the data processing pipeline to enhance the accuracy of methane concentration readings. The use of ML techniques allows us to capture nonlinear relationships between the sensor output and environmental variables, enabling data-driven corrections that traditional calibration techniques may overlook.

4.1 ML Model Selection

To ensure optimal performance of machine learning in processing the sensor data, it is crucial to select a suitable model capable of minimizing the impact of environmental factors. Therefore, we selected several commonly used models that have been widely applied in other gas sensing applications for comparison and evaluation.

4.1.1 Multiple Linear Regression

Multiple Linear Regression (MLR) models the relationship between one dependent variable and multiple independent variables using a linear approach. It efficiently estimates variable coefficients by minimizing the residuals' squared differences.[17]

MLR is ideal for data with linear relationships, excelling in fields like economics and biological sciences where interactions are straightforward. Its simplicity and computational efficiency make it a go-to method for predictive modeling when dealing with clear, additive effects among variables. Overall, MLR provides a direct and effective solution for linear data analysis across various applications.[18]

4.1.2 Support Vector Regression

Support Vector Regression (SVR) is effective for predicting complex and non-linear relationships between variables. It extends the principles of Support Vector Machines to regression, focusing on fitting data within a specified tolerance margin while minimizing error. SVR is particularly robust against overfitting in high-dimensional spaces due to its emphasis on structural risk minimization.

The flexibility of SVR comes from its use of various kernel functions, enabling it to adeptly handle non-linear data transformations. This makes SVR well-suited for environmental sensor data, where complex interactions between variables are common. Overall, SVR provides a powerful and efficient approach for sophisticated predictive modeling tasks. The support Vector Regression is a kernal function to find out the gaussian basis function [19]

4.1.3 Gaussian Process Regression

Gaussian Process Regression (GPR) is highly effective for predictions involving complex and non-linear relationships between variables, making it suitable for environmental sensor data. GPR uses a Gaussian process prior characterized by a mean and a covariance function, allowing flexibility in adapting to data intricacies. This adaptability is crucial for modeling behaviors in scenarios where environmental factors significantly impact sensor readings.

Unlike traditional models, GPR quantifies prediction uncertainty, enhancing decision-making in environmental monitoring. Its probabilistic nature and independence from structured model forms offer significant advantages, particularly in applications with poorly understood physical dynamics. Thus, GPR stands out for its robust, flexible, and accurate modeling capabilities in complex data environments. [20]

4.1.4 Artificial Neural Networks

Artificial Neural Networks (ANNs) have become a popular tool for predicting emissions where the relationship between input and output variables is non-linear. Compared to the model previously discussed, ANNs generally exhibit superior performance in managing the complex interactions between sensor readings and environmental data. Existing research indicates that ANNs have been successfully employed in various predictive applications involving gas sensors, often yielding high accuracy.

In the specific ANN model in question, the architecture comprises three layers. The activation function for the first two layers is ReLU (Rectified Linear Unit). ReLU is favored for its simplicity and efficiency in promoting non-linear properties without significantly impacting computational requirements. This choice of activation function helps in handling non-linear data effectively, enhancing the model's ability to learn complex patterns without falling into the pitfalls of overfitting or gradient vanishing problems commonly encountered in deep neural networks. This structure allows the ANN to not only capture intricate relationships within the data but also maintain robustness, making it highly suitable for environmental sensing applications where accuracy and reliability are paramount. [21]

4.2 Machine Learning Data Training

4.2.1 Environment Setup & Data Collection

To build an environmental simulation setup, as illustrated in Fig. 4.4, we designed a controlled testing system capable of adjusting methane concentration, humidity, temperature, and pressure. The sensor was directly connected to calibration gas cylinders containing both methane and nitrogen. These gases were routed through a mass flow controller, allowing precise control over the flow rates and mixing ratios to achieve the desired methane concentrations.

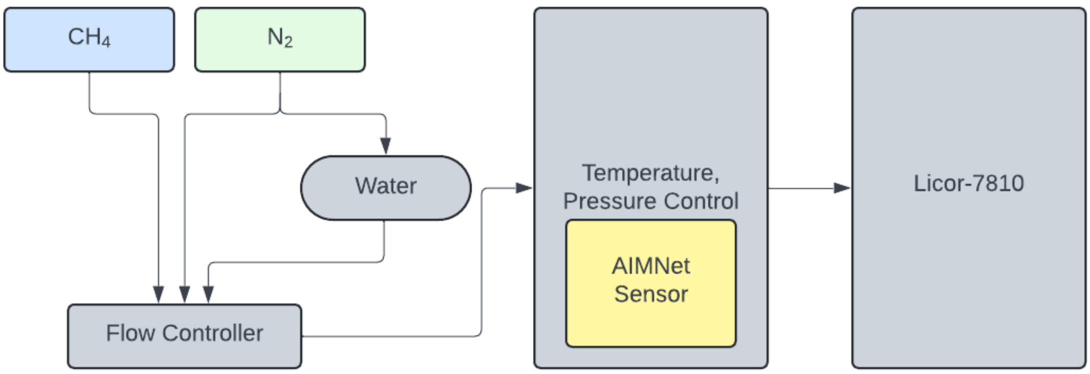


Fig. 4.4: Block diagram of Training Data Collection Setup

To simulate varying humidity conditions, a portion of the nitrogen gas was bubbled through a water reservoir, producing humidified air. By adjusting the mixing ratio between dry and humid gas streams, we were able to regulate the relative humidity reaching the sensor. Like the Fig. 4.5 shown, the entire system was controlled by the

Alicat Flow control device, it could control the flow of input , and between the control and device is the place we modify the humidity, we can use descant or water humity to remove or add water to the air then make the combination of three way of the gas.

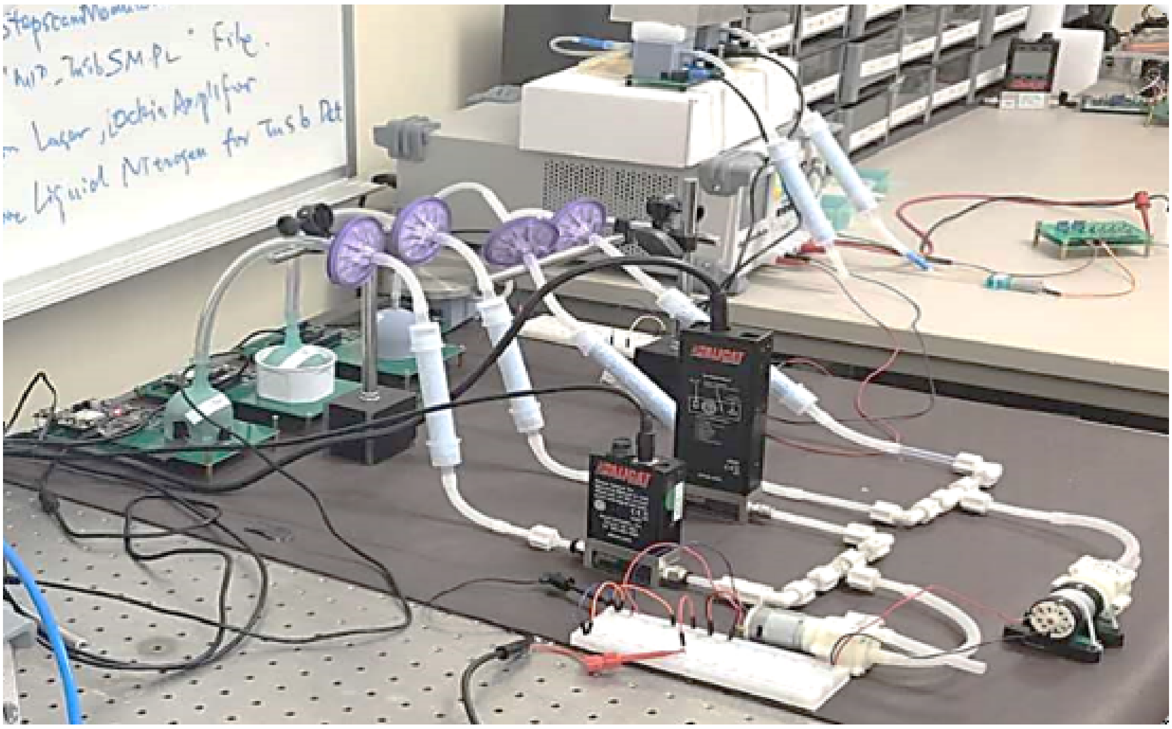


Fig. 4.5: Block diagram of Training Data Collection Setup

Then the sensor unit was then placed inside a sealed metal oven chamber to simulate elevated temperature conditions, replicating the high surface temperatures (up to 50°C) typically encountered during summer in Oklahoma. The sealed enclosure also allowed us to manipulate internal pressure conditions. By modifying the flow rate and outlet restrictions, we were able to induce slight pressure variations within the sensor chamber.

This comprehensive setup enabled precise control over environmental variables—methane concentration, humidity, temperature, and pressure—allowing us to collect data under realistic and repeatable test conditions for model training and validation.

To simulate real-world environmental variability, we developed two distinct datasets under controlled yet representative conditions. The first dataset encompassed a wider methane concentration range of 0 to 50 ppm, reflecting scenarios with substantial methane emissions commonly observed in field deployments. The second dataset focused on a narrower concentration range of 0 to 20 ppm, specifically designed to evaluate the model's prediction accuracy in low-emission scenarios where detection sensitivity is critical.

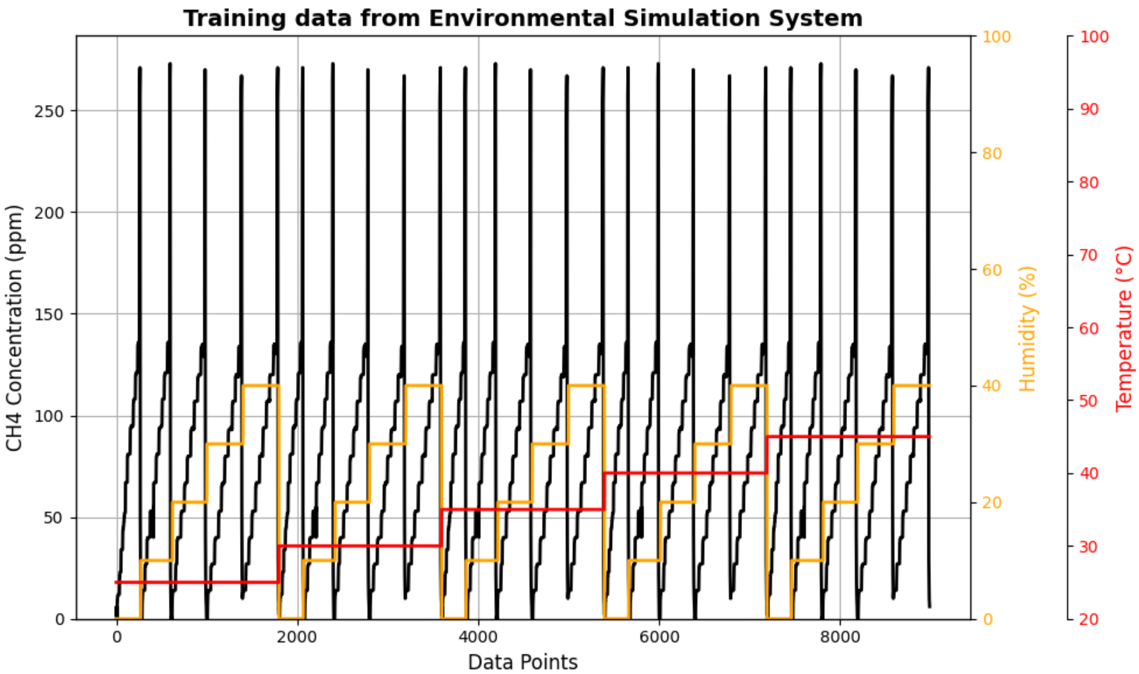


Fig. 4.6: Calibration Data

Both datasets were tested under a range of environmental parameters to emulate typical outdoor conditions. Temperature settings ranged from 25°C to 50°C, corresponding to the peak surface temperatures measured on sensor enclosures during summer operations in Oklahoma. To account for varying atmospheric moisture levels, three distinct humidity conditions—0 percentage, 25 percentage, and 50 percentage —were applied during testing. These values were chosen based on observed field humidity ranges and were integrated systematically across all test cases.

Comprehensive data collection was conducted across the full matrix of temperature and humidity combinations, as shown in figure 5.5, which ensures that the resulting datasets capture the environmental diversity during actual field deployments.

4.2.2 Training Result and Model Selection

For the training process, we planned to evaluate several different models to identify those most suitable for our system design. To ensure robust model development, it was crucial to implement steps that would validate the training performance and avoid overfitting. As described earlier, we collected approximately 20,000 data points from two datasets generated using our environmental simulation system. These data were divided into three parts.

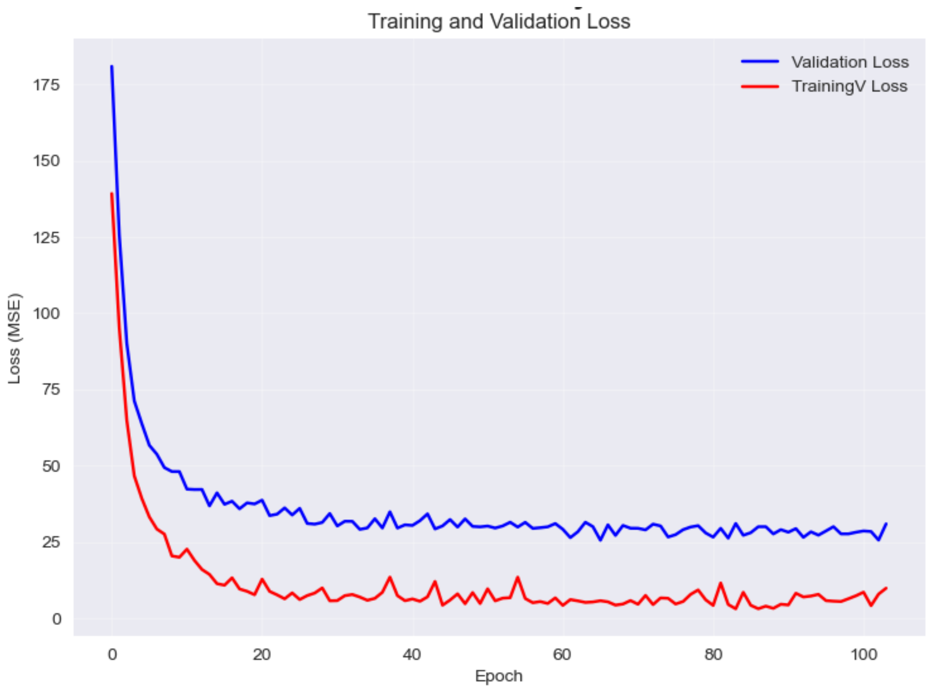


Fig. 4.7: Training and validation loss over epochs

To establish a reliable training dataset and prevent overfitting, we first removed duplicate entries, unclear data, and manually inserted anomalies. We then randomly shuffled the remaining data to break any inherent sequential patterns that might bias the model. From this cleaned dataset, 80 percentage (around 10,000 data points) were used for training, while the remaining 20 percentage (approximately 2,500 data points) were set aside as a validation dataset to monitor model performance during training.

From Fig.4.7, we can find out the model was trained when which the Loss drop from 200 to at about 3 to 4, and the comparision of training loss and validation loss mean the model is not overfitting and it still provide good enough loss.

For the training results, we employed baseline model structures without fine-tuning the sensor hardware itself. Instead, we explored a wide range of model combinations to identify the optimal configuration. Most models exhibited low error in both the training and validation datasets, with RMSE values ranging from approximately 0.5 to 2. Given that our prediction target spans 0 to 100 ppm, this level of error indicates that the models effectively captured the underlying data relationships. Among all evaluated models, the multilayer perceptron (MLP) demonstrated the best performance. As shown in Fig. 4.8, the RMSE values for both training and validation sets remained consistently low, confirming the model's ability to predict methane concentration with minimal deviation from the actual values.

Additionally, to further assess the generalization capability of each model, we created a third dataset consisting of new methane concentration values ranging from 1 to 100 ppm at uniform intervals. This dataset also included several values outside the original training range to evaluate how well the models could handle unseen environmental conditions. This third set included approximately 2,500 data points and followed a similar structure to the previous datasets.

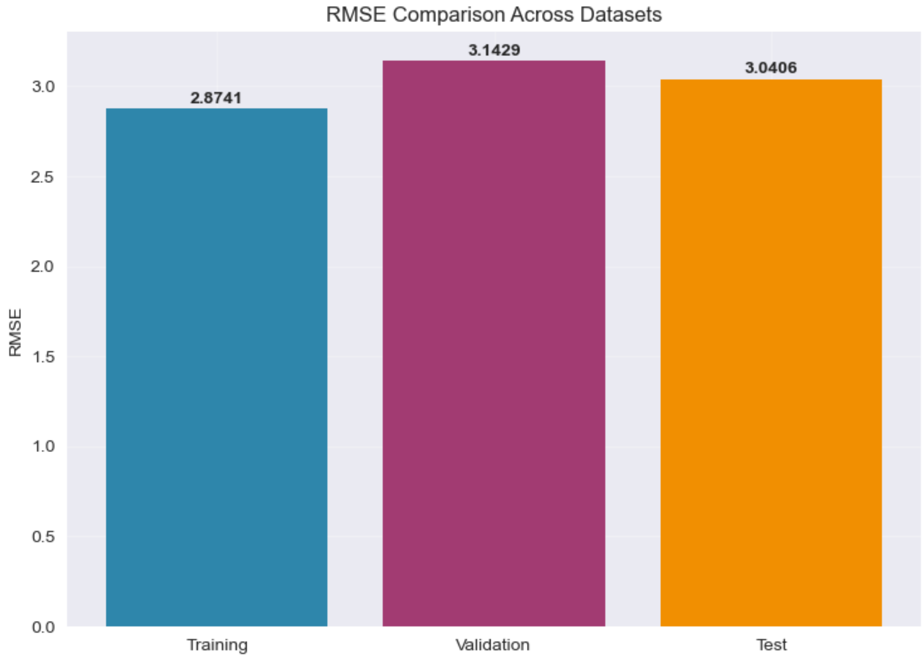


Fig. 4.8: Root Mean Square Error (RMSE) for Training, Validation, and Test Datasets
Blue: Training dataset, Pink: Validation dataset, Yellow: Test dataset

4.3 In-lab calibration and validation

After establishing the model for our NDIR methane sensor network system, it is essential to evaluate its performance through a structured validation process. The primary objective is to ensure high accuracy and stable data quality across a range of environmental conditions, especially considering that methane sensing in outdoor environments can be significantly influenced by temperature, humidity, and atmospheric pressure. Although the NDIR sensor was originally designed for laboratory settings, our research focuses on modifying and deploying it for robust, long-term outdoor applications.

To thoroughly validate the sensor's performance, we implemented a three-phase validation procedure. This process is designed to progressively transition from controlled conditions to increasingly complex and realistic environments. The first phase is an in-lab test, conducted in a controlled indoor environment. This step ensures that the sensor's baseline functionality aligns with the manufacturer's specifications and provides a reference point for subsequent outdoor experiments.

Following the laboratory validation, the second phase consists of two distinct outdoor tests. The first is a static balcony test, which simulates a stable outdoor environment while still offering some environmental protection. This scenario enables us to examine the sensor's response under real sunlight, airflow, and ambient variations without the unpredictable disturbances found in mobile or field settings.

Since the ultimate objective of our research is to develop a methane sensing system suitable for long-term outdoor deployment, it is essential to conduct rigorous validation that reflects the dynamic environmental conditions typical of real-world applications. To simulate such variability while maintaining control over the experimental setup, we continued using our environmental simulation system. Unlike the data collection process used for training our machine learning models, which involved labeled and structured data acquisition, this validation phase was conducted as a blind test. In this context, the system was tested without prior knowledge of the emission timeline, allowing us to evaluate its responsiveness and accuracy under uncertain and changing conditions.

The validation experiment was carried out over a continuous 12-hour period inside a controlled laboratory environment. During this time, temperature and humidity were intentionally varied to simulate realistic diurnal patterns. The temperature ranged from 25°C to 55°C, while relative humidity was adjusted between 0 percentage and 60 percentage. These variations were achieved using heating, ventilation, and humidification equipment integrated into the simulation chamber.

Methane emissions were introduced at five different time intervals throughout the test. Each emission phase was manually triggered and included plume concentrations ranging from 0 ppm to 100 ppm. These emissions were delivered in the form of localized methane plumes with varying intensity and spatial coverage to mimic real-world scenarios such as fugitive leaks or burst emissions. Each stage of the test—including both emission and non-emission periods—was maintained for a duration of 5 minutes before transitioning to the next condition.

To evaluate the sensor's ability to distinguish between environmental fluctuations and actual methane presence, we divided the validation test into two main segments. In the first half, both environmental changes and controlled methane emissions were introduced simultaneously. In the second half, no methane was emitted, but temperature and humidity were continuously varied. This portion assessed the sensor's environmental stability and its potential for false positives under changing ambient conditions.This helped evaluate the sensor's sensitivity and selectivity in distinguishing true methane signals from background noise induced by external variables.

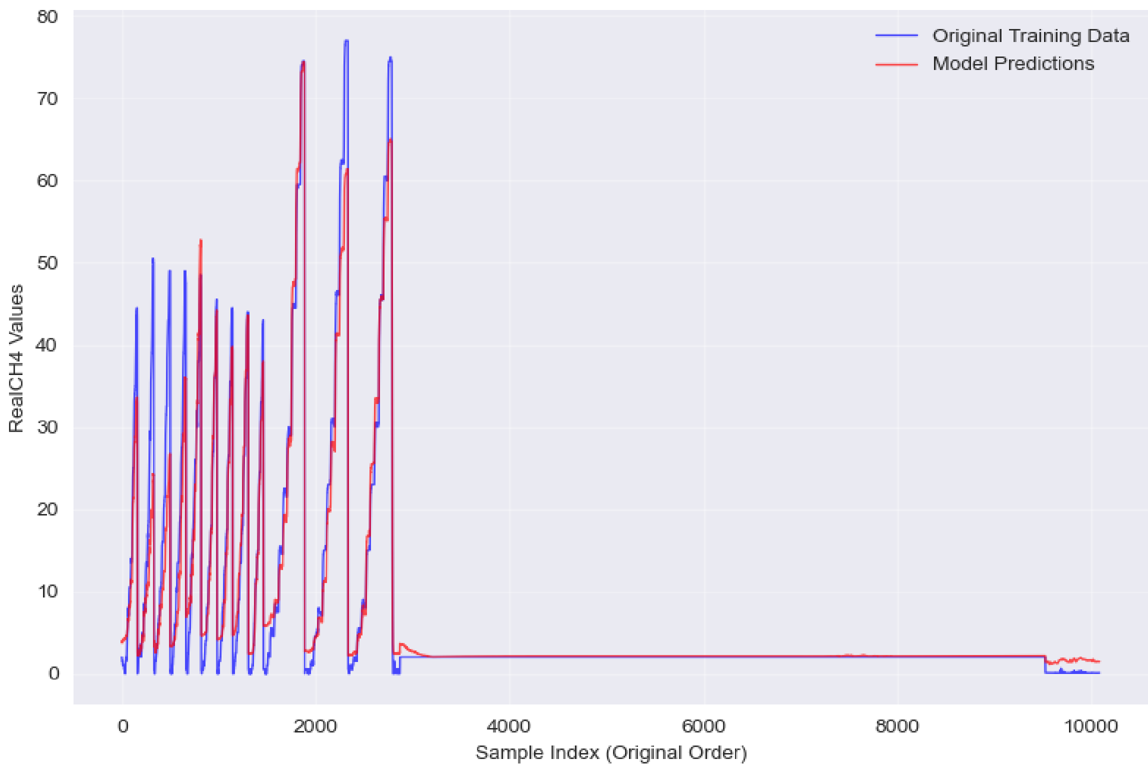


Fig. 4.9: In-Lab Validation Results Using Original Data Order to Show Methane Peaks

As shown in Fig. 4.10, the model maintains good performance on the new dataset, achieving an R-squared value of approximately 0.8688. Compared to the training and validation datasets, this represents a drop of around 0.07, which remains within an acceptable range. The scatter plot further illustrates that the model performs more accurately at lower methane concentrations, whereas higher concentrations tend to result in increased prediction error. This indicates that while the model generalizes well overall, its accuracy diminishes slightly when methane levels are elevated.

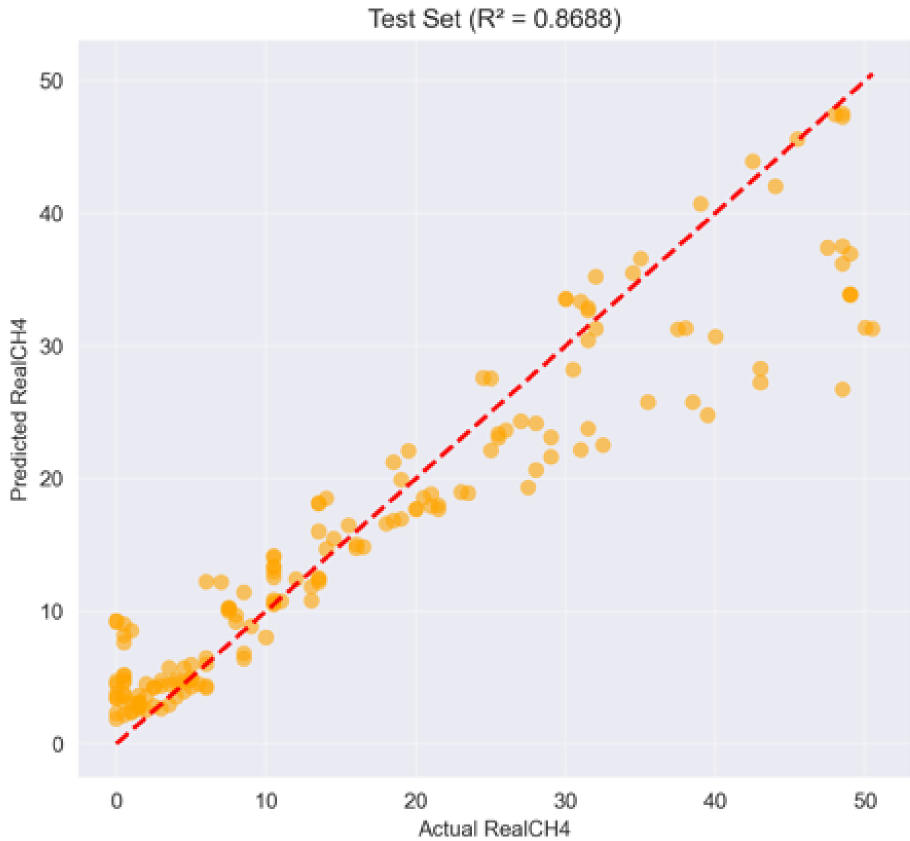


Fig. 4.10: R-squared validation result for the in-lab test

4.4 Outdoor calibration and validation

After completing the in-lab validation phase, we proceeded to the next step of our testing protocol by relocating the sensor system to an outdoor environment. The goal was to evaluate the sensor's performance under more realistic atmospheric conditions before committing it to long-term field deployment. As an intermediate stage between lab and full-field testing, we selected a stable outdoor setting—the balcony of Devon

Energy Hall on the University of Oklahoma campus. This location provides controlled exposure to environmental factors while maintaining easy access to power and safety monitoring, making it ideal for semi-controlled outdoor trials.

.Unlike the in-lab tests, where methane gas was released in close proximity to the sensor inlet, the outdoor tests introduced spatial separation between the emission source and the detection unit. Due to limited access to large volumes of methane gas, we conducted tests using certified calibration gas cylinders containing methane concentrations of 10 ppm, 50 ppm, and 100 ppm. The gas source (a calibration tank connected to a release nozzle) was positioned at various distances from the methane sensor, ranging from direct contact (0 feet) to a maximum distance of 5 feet. This setup allowed us to assess how detection accuracy and response time were affected by distance and diffusion in an open-air environment.

We designed five separate test scenarios to account for different environmental conditions, including daytime versus nighttime, high temperature, low temperature, and wet conditions such as light rainfall. For each methane concentration, gas was released continuously for 2 minutes. A mandatory 5-minute interval was maintained between each gas emission to allow sufficient time for the previous methane concentration to dissipate and for the sensor readings to return to baseline levels. This procedure ensured that each test remained independent and unaffected by residual gas in the surrounding environment.

To validate the results, we used a high-precision reference instrument: the LI-COR 7810, a state-of-the-art methane analyzer with a detection accuracy of approximately 1 part per billion (ppb)—roughly 1,000 times more sensitive than our NDIR sensor. For comparison, we connected the inlet of the LI-COR 7810 directly to the output path of our methane sampling chamber. This ensured that both sensors sampled the same gas mixture under identical environmental conditions, enabling accurate side-by-side performance assessment.

Method	RMSE (Large Concentration Range Data Set)	RMSE(Small Concentration Range Data Set)
Linear Regression	9.72	4.26
Polynomial Regression	20.71	6.11
Random Forest	7.83	5.92
SVR (Support vector Regression)	5.34	3.18
Gaussian Process Regression	---------	---------
BNN	5.79	3.55
ANN	6.98	3.91
MLP	6.98	4.17
XGBoost	----------	--------

Fig. 4.11: Model's Accuracy on Test Data Set

Two list is included, left is the RMSE when model applied to the large methane concentration range data set, and right is the RMSE for the small methane concentration range data set, – means the value is too bad too show

The goal of this intermediate outdoor validation was to identify potential limitations in detection sensitivity, evaluate environmental interference effects, and test the consistency of our sensor under semi-controlled but variable outdoor conditions. The outcomes from this phase guided further adjustments before full-scale field deployment.

(a) Test Distance 0.5 ft (b) Test Distance 1 ft (c) Test Distance 1.5 ft

Fig. 4.12: Balcony test with the AIMNet device and Licor 7810

[Three different distance applied for the test for the data validation]

The results of the three outdoor tests are presented in Fig. 4.13. As shown in the graph, the raw signal data obtained from the sensor exhibits significant fluctuations due to environmental variation, making it difficult to clearly identify actual methane emission events. The influence of temperature, humidity, and other ambient factors introduces substantial background noise, which masks the true emission signature.

However, after applying our machine learning model to the raw data, it becomes evident that much of the environmental background has been effectively removed. The processed data more clearly reveals the timing and magnitude of methane emissions. Despite this improvement, the presence of systematic distortion remains observable. Although the calibration gases used were 100 ppm, 50 ppm, and 10 ppm, the corresponding processed signals consistently peaked at approximately 80 ppm. This indicates that while the model successfully suppresses environmental noise, it does not fully preserve the true amplitude of the methane concentration, resulting in a reduced concentration ratio in the final output.

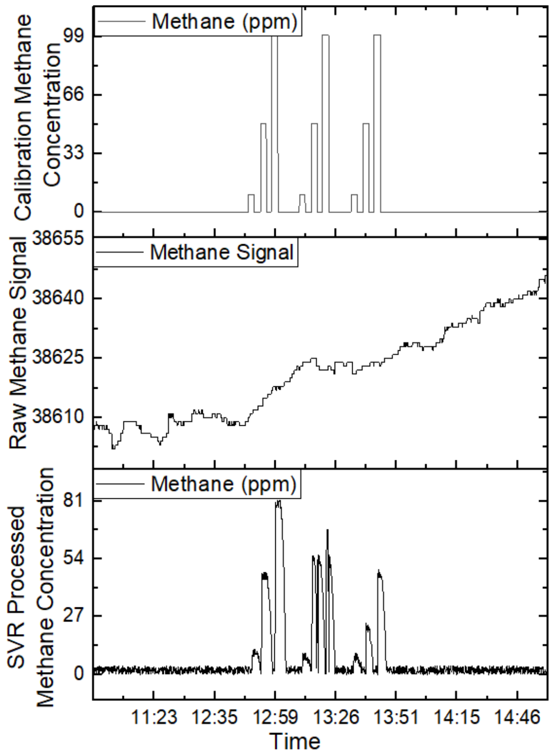


Fig. 4.13: Result and comparison for balcony test

Upper : Calibration Gas Concentration

Middle: Raw Methane Signal

Lower: Data after SVR predicted

Additionally, the effect of distance between the emission source and the sensor is clearly reflected in the results. As the test distance increases, the measured methane concentration decreases significantly, and the processed signal begins to exhibit more noticeable noise artifacts. This suggests that when the sensor is positioned farther from the emission source—especially under low concentration conditions—the signal-to-noise ratio deteriorates, even after machine learning correction.

From the result of the graph, we can make the comparison. I made 3 peaks of emission from the calibration gas tank I prepared and as plan I set. However, althought the algrithm successfully detect the

Nonetheless, if appropriate post-processing filters are applied, such as ratio correction and additional denoising techniques, the final results can achieve up to approximately 80 percentages accuracy compared to the true concentration values. This confirms that while the sensor performance degrades with distance and low signal strength, meaningful and reasonably accurate results can still be extracted through advanced data correction methods.

CHAPTER 5

Preliminary Demonstration of NDIR-based Sensing Network

With the completion of the development of NDIR point sensor instrument having enhanced data analysis algorithm, in this chapter, the effort has been focused on the preliminary demonstration of the NDIR-based sensing network. And starting from the first section, I will present the work on deploying the distributed type sensing network over the oil and gas field in Anadarko Basin testbed. It is noted that due to the supply-chain challenge, at the moment, we only built 20 instruments, thus the deployed network has a limited density and spatial coverage to sufficiently demonstrate its capability to capture the methane emission plume. Because of that, in the following section, I will also present our alternative demonstration solution by placing the instrument on a mobile vehicle. In this way, we can obtain almost a spatial continuous sensing data set along the driving path around a methane emission facility. We think this method can be used to mimic a high density distributed type network and thus prove the concept of our eventual goal to create an effective methane sensing network using the cost-effective NDIR sensing instruments.

5.1 Field Deployment of Distributed Sensor Instruments

After completing the validation phase, we built 20 sensor instruments, and began the installation of all of these devices onto the power posts along the road over the Anadarko basin, as shown in figure 5.1. Each enclosure was mounted beneath a solar panel to provide shade and reduce the risk of overheating due to intense summer sunlight. And the gas sampling port is connect to a PVC tube whose height is extended about 10-15 feet above the ground attached next to the instrument.

Fig. 5.1: Field Installation of Field Monitoring System

Figure 5.2 illustrated all deployed 20 devices over a 3 mile by 3 mile land. As shown, the bottom-left corner of the map has an oil and gas test field, where we deployed five sensors to provide dense coverage. For the larger surrounding area, we spaced the remaining sensors approximately 0.5 miles apart, arranging them in a rectangular formation to ensure comprehensive coverage of the entire oil and gas facility. When any sensor detects a methane concentration peak, the corresponding circle on the map visualization increases in size and displays the measured value. This real-time mapping interface helps to quickly identify the location and intensity of methane emissions across the monitored region.

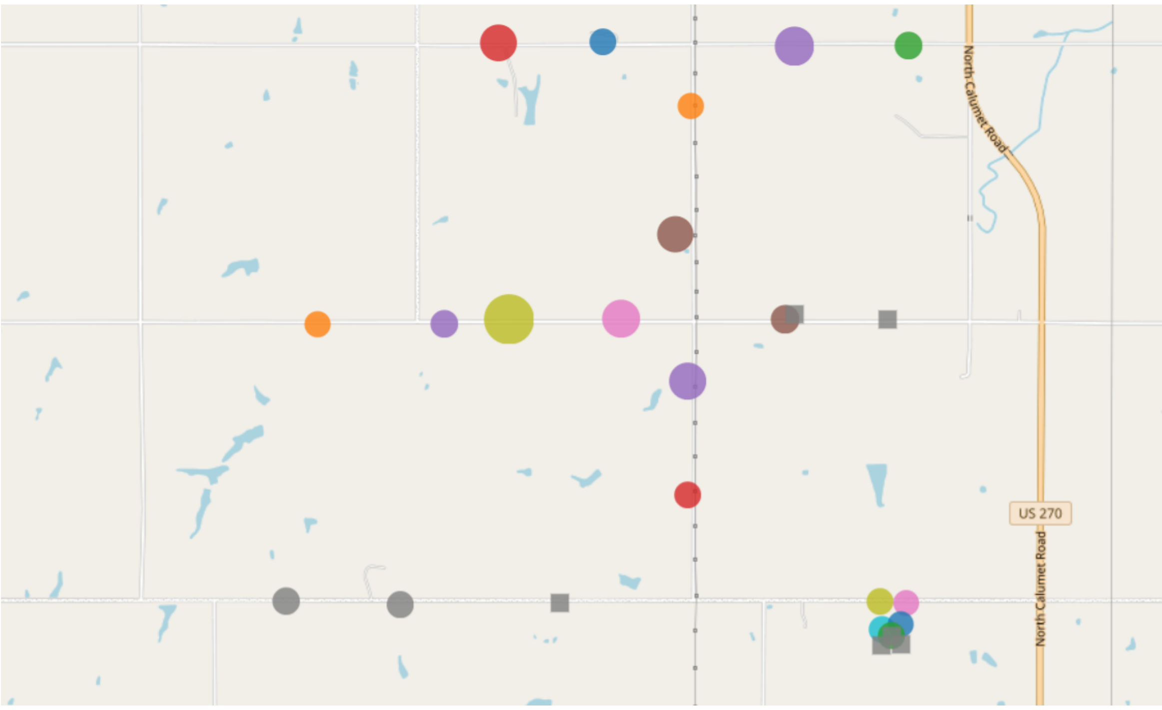


Fig. 5.2: Device Installation Location

5.2 Field test and validation

To ensure the device operating correctly under field conditions, we repeated a validation procedure similar to the one conducted during the earlier balcony test, as shown in figure 5.3. After installation, we selected one device #F030 for performance verification. To make the results more distinguishable, we extended the calibration period. For this test, we applied methane concentrations of 10 ppm, 50 ppm, and 100 ppm, with each stage lasting approximately 3 minutes. A second round of the same test was conducted after a 15-minute interval.

Fig. 5.3: Field Validation Test

As shown in Figure 5.4, the initial peak corresponds to the manual calibration injection. Following that, the graph shows a clear pattern of four distinct peaks—including the test injections—that align well with the designed concentration steps. These peaks exhibit a predictable rise and gradual decline in signal, confirming the device's responsiveness to varying methane levels.

However, one noticeable issue is that the measured values were consistently around 1.2 times higher than the known calibration concentrations. This suggests that, although the machine learning model successfully mitigates most environmental interferences, some residual influence remains. As a result, the final readings may still deviate by approximately 10–20 percentage, indicating room for further model refinement to improve accuracy in real-world outdoor applications.

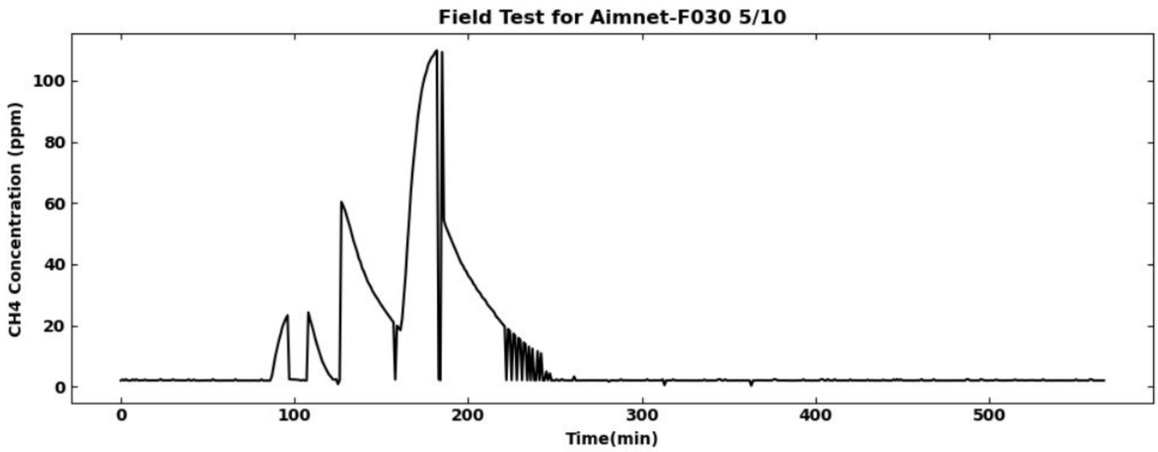


Fig. 5.4: Methane Response from Device F030 in Field

4 methane peaks detected, first one is test with 10 ppm , and last three is the calibration gas with 10, 50, 100 ppm

5.3 Mobile test and validation

Following the stable open-environment validation, we conducted further experiments to evaluate the feasibility of deploying our NDIR methane sensor system in remote outdoor installations. The goal was to test sensor performance under real-world environmental conditions and compare it against high-precision reference instruments. For this validation, we incorporated the LI-COR 7700, an open-path, laser-based methane sensor known for its high accuracy and suitability for outdoor field applications. Alongside it, we continued using the LI-COR 7810, a closed-path methane analyzer that had been employed in our previous experiments. Both LI-COR instruments served as ground-truth references for assessing the accuracy, responsiveness, and overall reliability of our sensor units.

To ensure mobility and flexibility during field testing, all equipment was mounted on a laboratory vehicle. This mobile platform included the NDIR sensors, reference instruments, data acquisition units, and a high-precision GPS module capable of achieving sub-meter accuracy (approximately 0.5 meters). The GPS was embedded into the system architecture to record precise geographic coordinates of all measurements in real time. This positional data was essential for post-processing and interpreting sensor output in a spatial context, especially under conditions where methane concentrations vary over short distances.

We selected a wastewater treatment site located to the south of the University of Oklahoma (OU) campus as the primary testing area. This site is known to emit a relatively consistent amount of methane due to ongoing biological processes. The expected concentration range at the site varies from 5 ppm to 30 ppm, making it an ideal location for sensor validation. Furthermore, natural wind patterns and environmental variability introduce spatial and temporal changes in the methane plume distribution. These dynamics allow for comprehensive testing of the sensor system's ability to detect real-time concentration changes and its adaptability to complex outdoor scenarios.

At the site, we deployed three of our NDIR sensor nodes, labeled FS002, FS003, and FS004. All three sensors were co-located with the LI-COR 7700 to ensure accurate comparative measurement under identical environmental conditions. FS002 and FS003 were configured identically with a pump flow rate set to 30 percentage, which is representative of standard low-power operation. FS004, on the other hand, was configured with a 50 percentage pump rate. This configuration was intentionally varied to investigate the influence of internal airflow and sampling rate on sensor accuracy and response time. By comparing these different settings, we aim to determine whether pump rate optimization can improve the sensor's ability to capture fast-changing methane concentrations.

Through this extended validation campaign, we not only assessed the consistency and accuracy of our sensor nodes but also evaluated their robustness in mobile outdoor environments. The results from this test provide valuable insight into the operational limits and configuration options necessary for deploying our sensor system in remote methane monitoring applications.

Fig. 5.5: Contribution of greenhouse gas emissions by sector

It is evident that all four sensors successfully detected two significant methane plumes emitted from the wastewater source during the mobile test. The Licor-7700, a high-precision reference instrument, was used as the ground truth to evaluate the accuracy of our custom NDIR sensors. Among the four devices, sensor FS003 demonstrated the most accurate and stable performance like the graph shown in Fig. 5.6. It not only captured the first major methane peak promptly but also successfully detected a secondary, smaller peak that appeared later during the test route. This consistency highlights FS003's sensitivity and suitability for mobile environmental monitoring applications.

For the result, Sensor FSO02 performed well overall; however, it still recorded one big false positive—an additional peak that did not correspond to a true methane event as verified by the reference device. This suggests that while FS002 is generally reliable, it is more susceptible to transient disturbances, possibly due to environmental fluctuations caused by vehicle movement.

Sensor FS004, on the other hand, delivered unsatisfactory results. It failed to detect the key emission peaks correctly, and the overall signal was highly unstable. This can be attributed to the high pump duty cycle of 50 percentage, which likely introduced excessive airflow noise and reduced the sensor's response time. These findings indicate that an overly aggressive sampling rate can degrade sensor performance under mobile conditions.

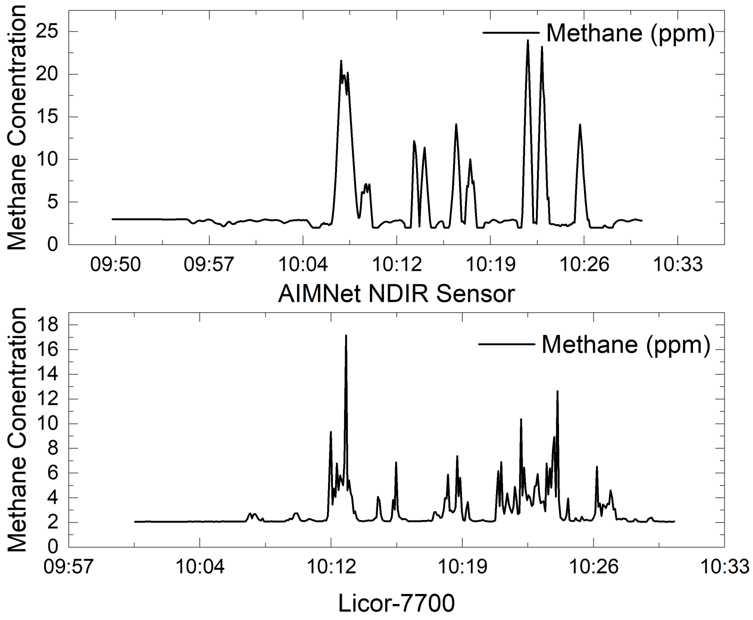


Fig. 5.6: Comparison between Fs002 and Licor-7700 methane measurement

In contrast, sensors operating at a 30 precentage pump duty cycle exhibited better stability. At this flow rate, the sensors were more capable of capturing high methane concentrations with minimal signal distortion. However, at lower methane concentrations, the signal tended to show increased noise, suggesting a trade-off between sensitivity and airflow rate.

Moreover, high vehicle speed during mobile operation appears to negatively affect detection performance, potentially leading to misalignment in plume capture. Future designs should consider optimized pump rates and vehicle speed to improve mobile sensing accuracy.

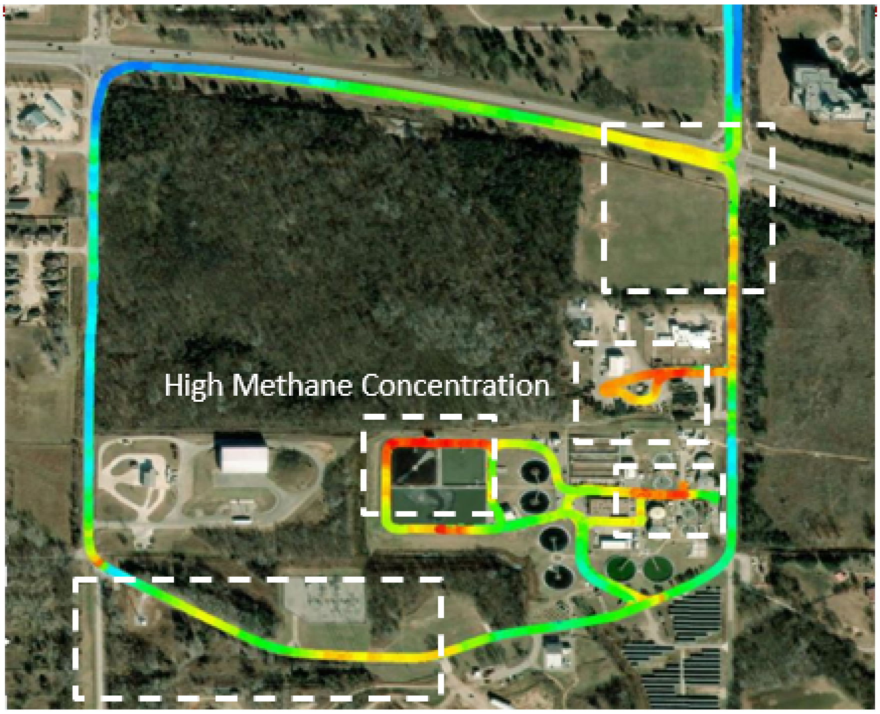


Fig. 5.7: Heat Map for Licor-7700

5 methane peaks were detected and 3 of them have high methane concentration, high methane concentration inside white area

From the heat map of the Licor-7700, as shown in figure 5.7, we can observe that the red rectangular outline highlights three major methane emission points located near the wastewater treatment facility and the trash station. At these points, methane concentrations exceed 15 ppm, which indicates a significantly high emission level. Additionally, moderate methane readings—ranging from 3 to 8 ppm—were detected along Highway 9 and the small trail adjacent to the wastewater station. These values suggest minor methane leaks in those areas, although not at a critical level.

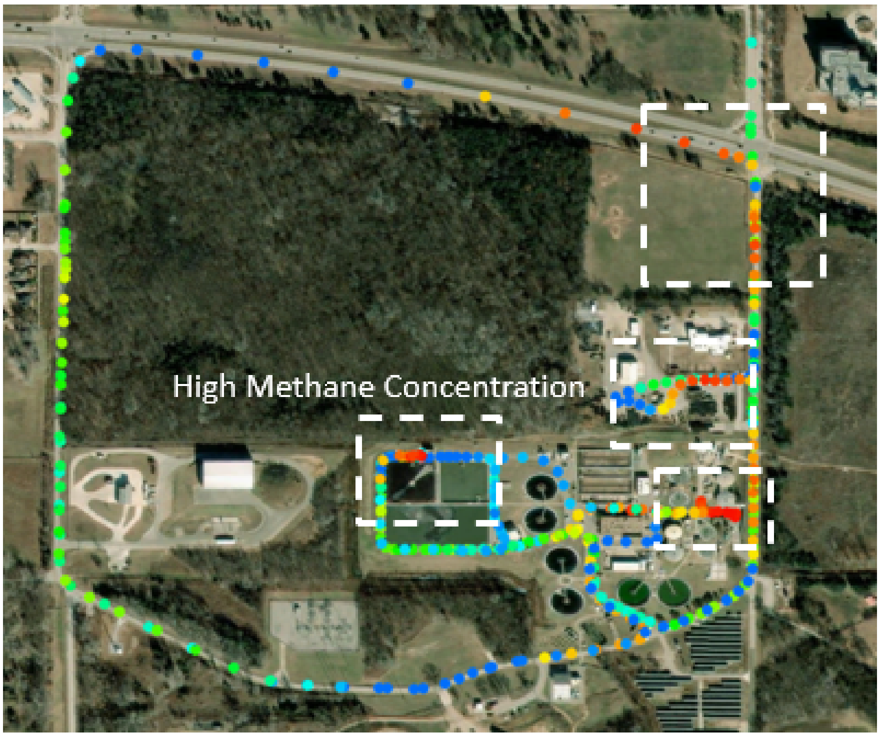


Fig. 5.8: Heat Map for FS-002

Left lower corner's methane peak is missing

When comparing this with the heat map from the FS-003 sensor as shown in figure 5.8, it is evident that FS-003 successfully captured the three major high-concentration methane emission points. It also detected elevated methane levels along Highway 9. The only emission source missed was the moderate-level reading along the small trail to the south of the wastewater station.

In summary, the FS-003 sensor identified 5 out of the 6 primary methane emission locations during the mobile testing. Notably, it demonstrated 100 percentage detection accuracy for high-concentration emission points and only missed one moderate-level source. This suggests that FS-002 provides strong performance in detecting significant methane emissions in real-world scenarios.

CHAPTER 6

Conclusion and Future Works

6.1 Research Summary

In this thesis, I present my work to develop and validate a machine learning–enhanced, NDIR-based methane sensing system designed for low-cost, scalable deployment in outdoor environments. This effort is in response to the growing demand for continuous, high-resolution methane monitoring in oil and gas fields, where traditional sensing approaches often suffer from high costs, low frequency, or poor reliability under variable environmental conditions.

The work I have completed can be divided into two major components. The first part involved the design and development of a custom methane sensor hardware system.

Fist of all, the system is built around a commercial-ready methane sensor, specifically the Senseair K96 module, provided from our industrial parter SenseAir. This sensor employs an intra-cavity reflective optical design to significantly extend the effective optical path (28 cm). Unlike conventional NDIR sensors with limited path lengths and lower detection precision, this module achieves a factory-calibrated resolution of approximately 0.5 ppm. This technology and specifically this K96 device was selected after a thorough evaluation of alternative sensing mechanisms—including calorimetric, electrochemical, pyroelectric, and metal oxide–based sensors—which were ultimately rejected due to limitations in sensitivity, environmental robustness, or scalability.

Having the selected sensor module, this system was then built entirely in-house and includes a fully integrated command module, MQTT-based wireless data transmission, onboard SD card storage, and a complete power management system. Specifically, the sensor module is embedded within a custom-designed hardware system that includes a microcontroller (Adafruit Feather M0), a real-time clock, data logging via onboard SD card, wireless communication through a SIM7070G LTE module, and an active gas sampling system driven by a miniature pump to ensure consistent airflow. Besides, the power supply is provided by a solar-powered energy subsystem, incorporating a 50 W solar panel and a 153.6 Wh lithium polymer battery pack, along with a high-efficiency power management system utilizing buck converters and a solar charge controller. This design enables the system to operate continuously for 4+ days without sunlight, ensuring reliable off-grid deployment.

One of the primary challenges in deploying NDIR sensors outdoors is the influence of environmental variables—including temperature, humidity, and barometric pressure—which distort the optical signal and reduce the accuracy of raw methane measurements. To overcome this, In the second part of the project, we implemented a data-driven correction layer using machine learning (ML). An environmental simulation testbed was constructed to generate controlled training datasets. This setup allowed us to precisely manipulate methane concentrations (0–100 ppm), temperature (25–55°C), and relative humidity (0–60%), producing over 20,000 labeled data points across realistic operational ranges.

In order to choose the most suitable ML models, I investigated and compared several ML models for sensor correction, including Multiple Linear Regression (MLR), Support Vector Regression (SVR), Gaussian Process Regression (GPR), and Multi-Layer Perceptron (MLP) neural networks. Each model was trained on a large dataset and evaluated using both in-distribution and out-of-distribution testing sets. The SVR and MLP models consistently outperformed others, demonstrating strong generalization capabilities even under novel environmental conditions and unseen gas concentrations. These models were subsequently embedded into the system's software stack for real-time correction of field data.

To validate the full system, I adopted a three-phase experimental strategy:

1) In-lab validation using the environmental chamber confirmed the system's ability to track dynamic methane concentrations accurately while compensating for temperature and humidity effects.

2) Semi-controlled outdoor validation on the balcony of Devon Energy Hall allowed us to test response time, accuracy, and environmental resilience under real weather conditions. Data from these tests were benchmarked against a high-end LI-COR 7810 analyzer (ppb-level precision).

3) Full-scale field deployment was carried out over a 3×3 mile area in the Anadarko Basin, using 20 independently operating sensing nodes. Each unit transmitted data via MQTT protocol to a central monitoring dashboard. Field data demonstrated the system's ability to detect elevated methane concentrations and localize leak sources spatially.

In addition, mobile testing was performed by mounting a sensing unit on a vehicle, paired with GPS logging. This approach allowed the creation of spatially continuous concentration maps—mimicking the effect of a dense sensor network—and proved particularly useful for rapidly surveying large areas.

Taken together, this work establishes a complete, field-validated methane sensing platform that balances cost, sensitivity, power efficiency, and scalability. The integration of machine learning into sensor calibration workflows further enhances the data quality, ensuring robust performance even in the face of environmental interference.

6.2 Future Work Roadmap

Looking ahead, several research directions can further enhance the performance, scalability, and practical deployment of the developed methane sensing system.

First of all, from the hardware aspect, it needs to be emphasized that scaling the sensor network for larger field coverage requires rethinking communication strategies.

While the current design relies on LTE cellular modules, this limits functionality in areas with weak or no signal. Future iterations may benefit from adopting low-power wide-area network (LPWAN) protocols such as LoRaWAN or incorporating mesh networking capabilities, which allow nodes to relay data across greater distances without relying on centralized cellular infrastructure. This would enhance scalability, reduce communication costs, and improve overall energy efficiency. [22] [23]

Another promising direction related to the data analysis is the integration of embedded machine learning models for real-time, on-device data correction. The machine learning model that I chose to implement for environmental compensation in my thesis operates through scripts or external post-processing. Transitioning this model to efficient, lightweight inference engines such as TensorFlow Lite or TinyML would allow the system to run corrections locally on the microcontroller, eliminating reliance on cloud connectivity and reducing latency. This edge-AI approach is particularly important for deployment in remote areas with limited communication infrastructure.[24]

In addition, the sensor drift and long-term calibration stability should warrant further attention. Despite environmental corrections through machine learning, gradual drift in sensor output over weeks or months can affect accuracy. Research into self-calibration strategies, including periodic exposure to reference gases or the use of co-located sensor redundancy, could help maintain consistent performance. Alternatively, adaptive algorithms or transfer learning methods could be implemented to update the correction model over time based on new field data, improving resilience without requiring full retraining.

On the other hand, since the ultimate goal for creating the distributive sensor network is to enable real-time methane leak localization and emission rate estimation, it must involve combining methane concentration data from multiple sensors with wind speed and direction measurements to reconstruct emission plumes and identify source locations. Therefore, advanced modeling approaches—such as Gaussian plume modeling, Bayesian inference frameworks, Kalman filtering or GHG-CHEM could be employed to estimate both the geographic origin and intensity of emission events. Developing these capabilities would elevate the system from a passive monitoring tool to an active diagnostic and response platform. It is worth noting that our group has been working with Dr. Xiao-Ming Hu, Dr. Chenghao Wang and Dr. Ming Xue - the experts on weather-coupled gas molecule transportation modeling - in the national weather center, aiming at the development of a high resolution accurate network observation data integrated emission source retrieval and plume dispersion forecasting model. [25]

Lastly, in terms of usability and user engagement, the development of a centralized monitoring platform with a cloud-based dashboard would significantly enhance the value of the system. Such a platform should support real-time visualization of methane maps, system health diagnostics, historical data trends, and customizable alert settings. Compatibility with industry reporting formats and regulatory frameworks, such as the EPA's emission standards or the OGMP 2.0 protocol, would further promote adoption in industrial and environmental applications. This is another ongoing effort our team has been working with researchers from OU data institute to pursue. A few web-based visualized data figures presented in the previous chapters are coming from the preliminary web dashboard that the team recently created. [26]

Overall, the continuation of this research offers numerous exciting opportunities for enhancing the precision, adaptability, and application scope of low-cost methane sensing technologies. With targeted improvements in system robustness, intelligence, and autonomy, we anticipate that the developed platform holds strong potential to become a cornerstone in next-generation environmental monitoring and methane mitigation strategies.

List of Abbreviations

NDIR	Non-Dispersive Infrared Sensor
SVR	Support Vector Regression
GPR	Gaussian Process Regression
RF	Random Forest
XGBoost	Extreme Gradient Boosting
ML	Machine Learning
DL	Deep Learning
AI	Artificial Intelligence
NN	Neural Network
ANN	Artificial Neural Network
GIS	Geographic Information System
MSE	Mean Squared Error
IEEE	Institute of Electrical and Electronics Engineers

Bibliography

[1] T. Stein, "U.s. methane emissions flat since 2006 despite increased oil and gas activity," https://phys.org/news/2019-05-methane-emissions-flat-oil-gas.html, May 2019, accessed: 2025-06-29.

[2] M. Filonchyk, M. P. Peterson, L. Zhang, V. Hurynovich, and Y. He, "Greenhouse gases emissions and global climate change: Examining the influence of co2, ch4, and n2o," *Science of The Total Environment*, p. 173359, 2024.

[3] M. Omara, D. Zavala-Araiza, D. R. Lyon, B. Hmiel, K. A. Roberts, and S. P. Hamburg, "Methane emissions from us low production oil and natural gas well sites," *Nature Communications*, vol. 13, no. 1, p. 2085, 2022.

[4] T. Aldhafeeri, M.-K. Tran, R. Vrolyk, M. Pope, and M. Fowler, "A review of methane gas detection sensors: Recent developments and future perspectives," *Inventions*, vol. 5, no. 3, p. 28, 2020.

[5] N.-H. Park, T. Akamatsu, T. Itoh, N. Izu, and W. Shin, "Calorimetric thermoelectric gas sensor for the detection of hydrogen, methane and mixed gases," *Sensors*, vol. 14, pp. 8350–8362, 2014.

[6] Y. Huang, X. Peng, and X.-Q. Chen, "The mechanism of θ-to α-al2o3 phase transformation," *Journal of Alloys and Compounds*, vol. 863, p. 158666, 2021.

[7] K.-i. Uchida, H. Adachi, T. Kikkawa, A. Kirihara, M. Ishida, S. Yorozu, S. Maekawa, and E. Saitoh, "Thermoelectric generation based on spin seebeck effects," *Proceedings of the IEEE*, vol. 104, no. 10, pp. 1946–1973, 2016.

[8] N.-H. Park, T. Akamatsu, T. Itoh, N. Izu, and W. Shin, "Calorimetric thermoelectric gas sensor for the detection of hydrogen, methane and mixed gases," *Sensors*, vol. 14, no. 5, pp. 8350–8362, 2014.

[9] W. Dong, Y. Sugai, Y. Wang, H. Zhang, X. Zhang, and K. Sasaki, "Experimental study on enhanced methane detection using an mems-pyroelectric sensor integrated with a wavelet algorithm," *ACS omega*, vol. 9, no. 18, pp. 19 956–19 967, 2024.

[10] J. Li, X. Han, Y. Jiang, and D. Ba, "Theoretical study of the gas sensitivity and

response time of metal oxide thin films," *Thin Solid Films*, vol. 520, no. 2, pp. 881–886, 2011.

[11] P. K. Sekhar, J. Kysar, E. L. Brosha, and C. R. Kreller, "Development and testing of an electrochemical methane sensor," *Sensors and Actuators B: Chemical*, vol. 228, pp. 162–167, 2016.

[12] Z. Zhu, Y. Xu, and B. Jiang, "A one ppm ndir methane gas sensor with single frequency filter denoising algorithm," *Sensors*, vol. 12, no. 9, pp. 12 729–12 740, 2012.

[13] D. Furuta, B. Wilson, A. A. Presto, and J. Li, "Design and evaluation of a low-cost sensor node for near-background methane measurement," *Atmospheric Measurement Techniques Discussions*, vol. 2024, pp. 1–31, 2024.

[14] J. J. Lin, C. Buehler, A. Datta, D. R. Gentner, K. Koehler, and M. L. Zamora, "Laboratory and field evaluation of a low-cost methane sensor and key environmental factors for sensor calibration," *Environmental Science: Atmospheres*, vol. 3, no. 4, pp. 683–694, 2023.

[15] M. Berry, "Inflection reflection: images in mirrors whose curvature changes sign," *European Journal of Physics*, vol. 42, no. 6, p. 065301, 2021.

[16] B. Wastine, C. Hummelgård, M. Bryzgalov, H. Rödjegård, H. Martin, and S. Schröder, "Compact non-dispersive infrared multi-gas sensing platform for large scale deployment with sub-ppm resolution," *Atmosphere*, vol. 13, no. 11, p. 1789, 2022.

[17] D. Maulud and A. M. Abdulazeez, "A review on linear regression comprehensive in machine learning," *Journal of applied science and technology trends*, vol. 1, no. 2, pp. 140–147, 2020.

[18] K. F. Nimon and F. L. Oswald, "Understanding the results of multiple linear regression: Beyond standardized regression coefficients," *Organizational research methods*, vol. 16, no. 4, pp. 650–674, 2013.

[19] M. Awad, R. Khanna, M. Awad, and R. Khanna, "Support vector regression," *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*, pp. 67–80, 2015.

[20] C. Williams and C. Rasmussen, "Gaussian processes for regression," *Advances in neural information processing systems*, vol. 8, 1995.

[21] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research," *Journal of pharmaceutical and biomedical analysis*, vol. 22, no. 5, pp. 717–727, 2000.

[22] U. Raza, P. Kulkarni, and M. Sooriyabandara, "Low power wide area networks: An overview," *ieee communications surveys & tutorials*, vol. 19, no. 2, pp. 855–873, 2017.

[23] J. Haxhibeqiri, E. De Poorter, I. Moerman, and J. Hoebeke, "A survey of lorawan for iot: From technology to application," *Sensors*, vol. 18, no. 11, p. 3995, 2018.

[24] P. P. Ray, "A review on tinyml: State-of-the-art and prospects," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 4, pp. 1595–1623, 2022.

[25] X.-M. Hu, W. T. Honeycutt, C. Wang, B. Weng, B. Zhou, and M. Xue, "Observation and simulation of methane plumes during the morning boundary layer transition," *Journal of Geophysical Research: Atmospheres*, vol. 130, no. 7, p. e2024JD042317, 2025.

[26] E. Di Martino, M. Gasperi, A. Lo Nigro, and C. Valentini, "Fugitive emissions estimation and quantification according ogmp 2.0 standard," in *Offshore Technology Conference*. OTC, 2025, p. D021S028R002.