UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

TRANSFORMER-BASED MACHINE LEARNING AND MATHEMATICAL MODELS FOR THE PREDICTION OF METHANE EMISSIONS FROM INDUSTRIAL NATURAL GAS COMBUSTION ENGINES

A THESIS
SUBMITTED TO THE GRADUATE FACULTY

in partial fulfilment of the requirements for the

Degree of
MASTER OF SCIENCE

By HUGH NELSON
Norman, Oklahoma
2025

TRANSFORMER-BASED MACHINE LEARNING AND MATHEMATICAL MODELS FOR THE PREDICTION OF METHANE EMISSIONS FROM INDUSTRIAL NATURAL GAS COMBUSTION ENGINES

A THESIS APPROVED FOR THE
SCHOOL OF AEROSPACE AND MECHANICAL ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Pejman Kazempoor, Chair

Dr. Ramkumar Parthasarathy

Dr. Iman Ghamarian

Dr. Wilson Merchan-Merchan

ACKNOWLEDGEMENTS

I would like to take the time to recognize the multitude of influences that have guided me to this point in my life, that have helped me to find the inspiration and motivation to contribute my time and energy as wholeheartedly as possible towards this thesis, my graduate education, my relationships, and my faith.

To start, I would like to extend my thanks and gratitude to Dr. Pejman Kazempoor, Dr. Wilson Merchan-Merchan, and towards Dr. Iman Ghamarian. All three of these excellent examples in my life have shown me kindness, the importance of hard work, and the importance in going for things that seem out of reach. All in their own unique formats, they encouraged and supported me in technical aspects to get this thesis done but also in the great quantities of life uncertainties that exist in many young 20-something year olds. I had the opportunity to both request advice on the work of this thesis consistently throughout its development from each of them and to take my favorite and most educational courses from these professors during my time at the University of Oklahoma. I certainly would not be who I am right now without them, and I am very grateful for the ability to have met, to have learned under, and to have interacted with each of them. I am also thankful to the rest of my professors, advisers, and all those who make the aerospace and mechanical engineering program what it is at the University of Oklahoma.

This work would certainly not have been completed to the degree of quality that I am capable of today without the many phone calls, the many long days and occasional late nights, and great quantity of seemingly pointless though mentally stress-relieving conversations had with Mr. John Christopher Fisher, who is at this time also completing his master's degree in mechanical engineering alongside me. I am eternally grateful that I had the opportunity not only to have a good, helpful, and hardworking research partner during my time in graduate school, but that it was someone as unique and human as Chris. The advantages of having a research partner like Chris is that the work produced is elevated through competition, through encouragement, through support, and through friendship. I recommend anyone who wishes to tackle a task that may take large amounts of time and mental energy find themselves a Chris.

I would like to also make sure that my closest friends are recognized as great supporters in my finishing this project, as they have been a source of peace, of fun, of happiness, and of strength and encouragement during this time and during nearly all of my life before this time. I love you all deeply and I hope you know you are all a source of inspiration to me. You help me to believe in the greater good.

Another huge impact on this work has been the constant, unwavering Toby Lyons, who has been a loving father figure, a wise teacher, and an encouraging supporter long before, during, and I expect after this thesis is completed. I love you greatly and always will.

Finally, I would like to dedicate the work that went into this thesis to my late father and mother, Don Paul Nelson and Judith G. Hart Nelson. Thank you both for being so real, human, so proud, and so loving during the time that I knew you. I know what it means to do those things thanks to you. You continue to teach me and guide me so much every single day and I expect that to continue for the rest of my life. I will never give up my trying to live a good, caring life! I love you both.

ABSTRACT

Aging industrial natural gas engines, such as the AJAX DPC-81 and AJAX DPC-105 models, remain critical components of the natural gas transportation infrastructure. However, cost-effective methods for monitoring methane emissions from these legacy systems are lacking. Traditional mathematical combustion modeling approaches, given the engines' complex mixing and heat transfer behaviors, would require advanced multi-dimensional CFD modeling to achieve accurate predictions—an approach that is often impractical for real-world deployment. This thesis seeks to develop and validate a virtual sensor approach for real-time methane emissions monitoring using machine learning, specifically addressing the questions: Can transformer-based machine learning models, customized for noisy, low-frequency engine sensor data, predict methane emissions accurately? Can these models generalize effectively across varying operating conditions without requiring extensive hardware retrofits or complex physical modeling? Transformer-based machine learning architectures were developed and trained on real-time operating data from the AJAX DPC-81 and DPC-105 engines. The models were tailored to handle noisy input signals and low sampling frequencies. Performance was validated against unseen datasets and benchmarked with outputs from a Chemkin two-zone combustion model to assess predictive capability. The customized transformer models demonstrated promising predictive performance. For the AJAX DPC-105, the model achieved an average Mean Absolute Error (MAE) of 90.43 ppm on the validation dataset, outperforming the training and test MAEs (221.75 and 302.98 ppm, respectively), suggesting strong generalization to certain operating conditions. For the AJAX DPC-81 engine, the model achieved a training MAE of 164.27 ppm and a validation MAE of 337.2 ppm. While the absolute error was higher, the model closely followed the real data trends, achieving high accuracy when averaging predictions over time. The results demonstrate the feasibility of using transformer-based AI models for emissions monitoring in aging natural gas infrastructure. By achieving accurate methane predictions with minimal sensor input and without the need for complex CFD modeling or major hardware upgrades, this work provides a scalable and cost-effective pathway for methane mitigation efforts within the existing industrial fleet.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

CHAPTER 1: INTRODUCTION AND BACKGROUND

To start this work, a context of the relevant global issues will be formed such that the reader may see the importance not only in the specific objectives of this thesis, but the importance of caring about the related issues altogether. After having read this introductory chapter, the reader should feel confident that the work being done is pointed towards a goal justified and focused in its nature – with understanding of how global issues influence every step in the process of the work being conducted.

In general, there is awareness by the American and much of the global public of the fact that we have pushed – and are continuing to push – ourselves into some sort of dilemma concerning the exponential increase in consumption of resources to satisfy an insatiable appetite for convenience, for comfort, and for the sake of progress, all while the very systems that sustain us strains under the building weight. These strained systems are reliant on the beautiful Earth we inhabit. Most are aware of an issue, but there are also many with voices of doubt and ridicule. The reality of the situation is no less real, and the information for understanding our collective problems exists if one simply takes the time to look. What one finds upon looking is not particularly reassuring, and the scale of change required looks to be so vast one might laugh or discard the idea of trying. At some point foundational shifts in the way we live, work, and dream will have to take place. Those who can need to begin making the incremental adjustments to our social, political, and physical infrastructures in the great hope that eventually there is a mass recognition and action taken. Recognition turns a distant hum into a clear voice of reason for many, and the many may be able to make the true push towards sustainability. Failure is true and factual certainty if we do not try.

This thesis will focus in one incremental step towards helping the infrastructures we depend on today adhere to the tightening regulation meant by our regulating bodies to help guide us towards sustainability – continued human progress toward a more utopian world. Particularly, the problem of our energy infrastructures will be addressed, with a focus on the reduction of harmful greenhouse gas emissions from stationary hydrocarbon fueled internal combustion engines. This tiny piece of the pie may be almost comical in comparison to the 'exponential increases in the consumption of finite resources to satisfy an insatiable appetite', but it is something through this researcher's path in graduate school that they have found they can lend a hand to. A small incremental adjustment towards a small incremental adjustment.

SECTION 1.1: EMISSIONS

To begin building the background, a description and history will be given for that of the relevant molecules' interactions and relationship with human progress, which includes a discussion of global warming and of development and modern status of regulation. Regulation is, as stated, our governing bodies (which right now are our best hope at organizing the masses towards a unified goal) way of guiding us to sustainability. Once agreed upon that the regulation and regulating bodies are guiding us in a proper direction, it is helpful to comprehend the regulation such that we can understand what is needed to adhere to it.

SECTION 1.1.1: THE REALITY AND SEVERITY OF GLOBAL WARMING

Currently, the problem of global warming and the validity of the greenhouse effect is an occasional topic of public and political controversy. About one in five Americans believe that global warming is in fact not occurring, and around 29% of those that do believe it to be occurring believe that it is not caused by human activity and is a natural phenomenon [1]. In a study by Kahan of Yale University, 1540 U.S. adults were surveyed for their concern regarding climate change issues. The study found that on the whole, those who were more educated were actually less concerned with these issues than those who were less educated [2]. It was proposed that in the more educated brackets of individuals, it was more common for cultural polarizations on issues to become a bigger factor in the individuals' thoughts on the matter. In light of the still apparent doubt and controversy over the matter, a short discussion as to why the problem of global warming is a true problem that deserves to be addressed will be had.

2024 was the warmest year in temperature-recorded history. Organizations including NASA (National Aeronautics and Space Administration), NOAA (National Oceanic and Atmospheric Administration), ECMWF (European Centre for Medium-Range Weather Forecasts), ESA (European Space Agency), and WMO (World Meteorological Organization) all perform measurements of the global temperature using satellites, land and buoy weather stations, and weather balloons [3], [4], [5], [6], [7].

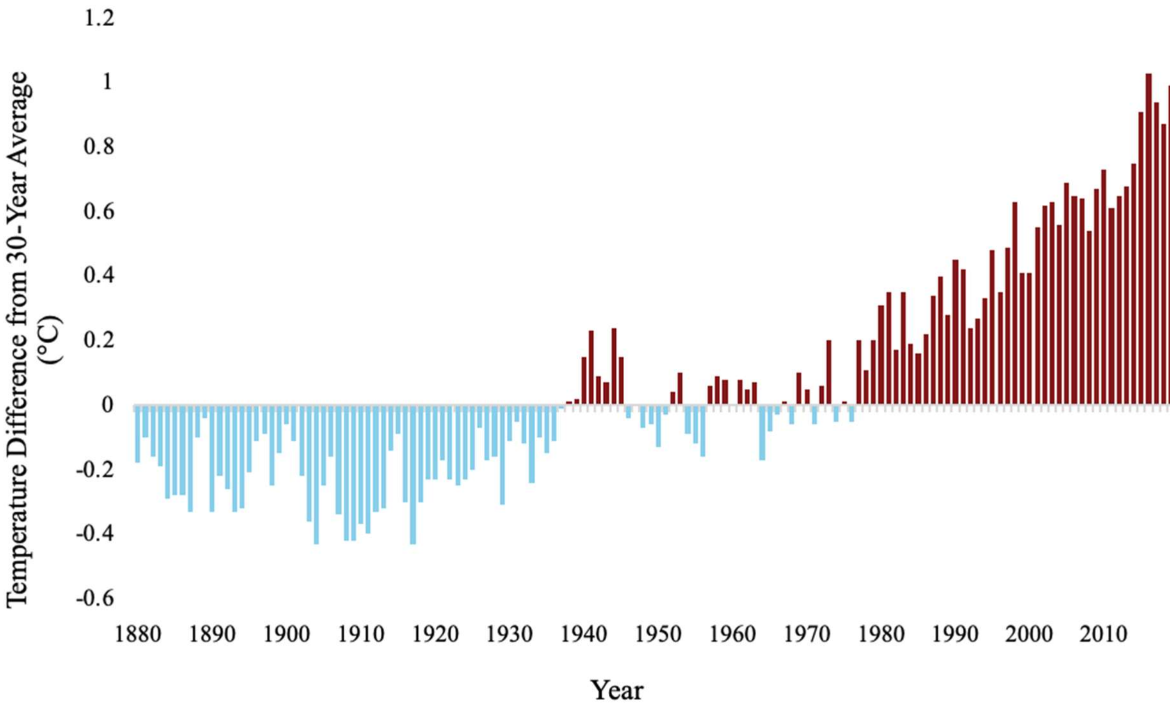


Figure 1: Temperature anomaly compared to the average temperature from 1951-1980 [8]

In Figure 1, the global temperature difference from the average temperature of the 29-year period of 1951-1980 is displayed for every year of recorded data by NASA's Goddard Institute for Space Studies (GISS). All of the other previously mentioned organizations that also track the global temperatures report nearly identical findings, and it is apparent that the global temperature is beginning to increase on a more exponential path. Following the fact that it is clear and provable that the global temperature is rising more and rising more quickly each year, it is also clear and provable that it is due to the exacerbation of the greenhouse effect.

The greenhouse effect is a natural phenomenon that keeps the Earth's surface warmer than it would be without an atmosphere in place, making the planet habitable for us human beings and the rest of the organisms on the planet. The greenhouse effect arises due to the ability of certain gases in the atmosphere to absorb and re-emit infrared radiation. The sun emits electromagnetic radiation across a wide spectrum of wavelengths, with the majority in the visible, the near-infrared, and the ultraviolet ranges. Some percentage of this radiation is reflected back into space, while the rest is absorbed by the surface of the earth and also by the atmosphere, warming the earth. After the absorbed electromagnetic radiation at earth's surface has brought an increase in surface temperature, the earth re-emits the energy in the form of heat, or long-wave infrared radiation. Some of this infrared radiation is then absorbed and re-emitted by the greenhouse gases in the atmosphere. The gases within the atmosphere emit this thermal energy in all directions, which includes back towards the earth's surface, leading to heat retention. The determination of whether or not this absorption and emission maintains the temperature of the earth's surface to a desirable degree depends on the rates of absorption versus emission – there is a delicate balance between absorbed solar radiation, emitted infrared radiation, and the composition of the gases in the atmosphere that determines the climate of earth. Greenhouse gases (GHGs) such as carbon dioxide, methane, or nitrous oxide are particularly effective at absorbing infrared radiation and can remain in the atmosphere for long periods of time. If their concentrations are elevated, the retained heat will be elevated. The following references may be looked into to obtain some background regarding the first scientists to discover and understand this phenomena and a detailed history of the current understanding of the greenhouse effect: [9], [10].

The exponential increase in the release of greenhouse gas emissions can be tracked back in history to the start of the industrial revolution. Starting in some parts of the world as early as the mid-1700s, the Industrial Revolution is the first time technologies built on carbon-based energy sources including coal, steam engines, and then early internal combustion engines in the 1860s are developed and utilized on mass scales [11], [12].

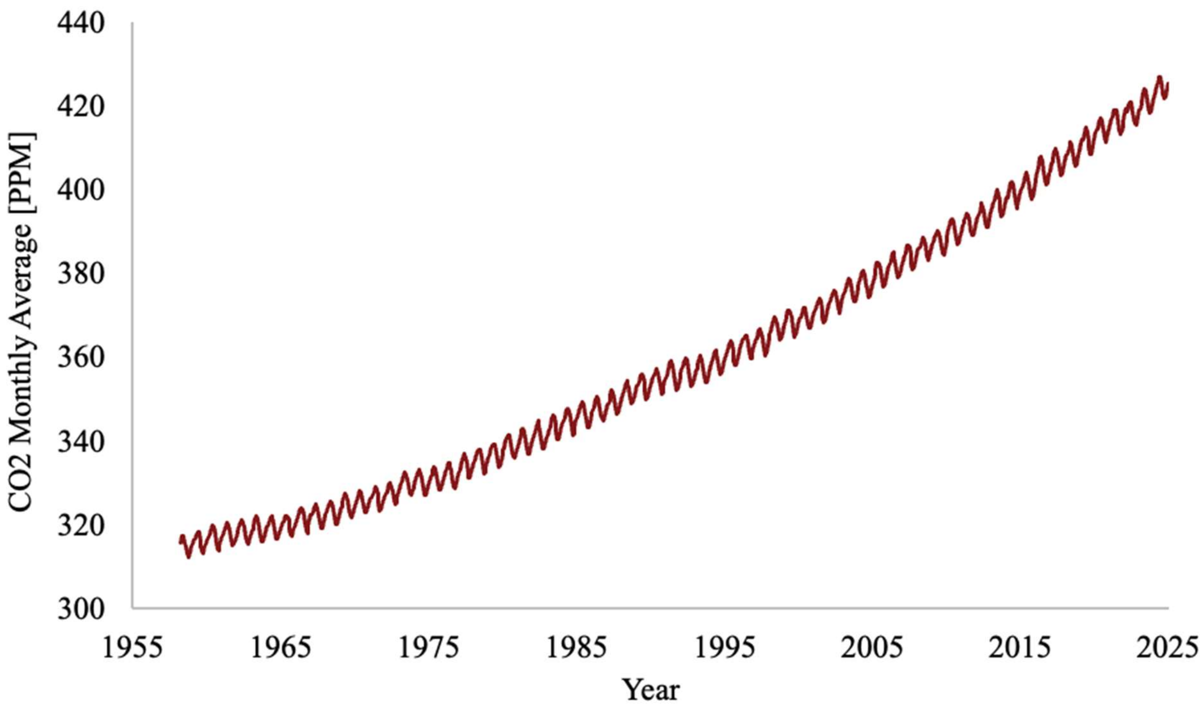


Figure 2: NOAA monthly CO2 averages recorded at the Mauna Loa Observatory [13]

The CO2 PPM measurements at Mauna Loa Observatory may be seen in Figure 2. To provide some intuition for the PPM data prior to the earliest year shown in this figure, 1958, it is important to understand that CO2 PPM counts were closer to the 280 PPM range before the industrial revolution [14]. This is to say, it is apparent and provable that the CO2 PPM counts in our atmosphere are rising in majority part due to our production of the gas for our industrial, commercial, recreational purposes. The other greenhouse gas molecules are oftentimes compared against CO2 in terms of their impact on the greenhouse effect, and the resulting metric that comes out of this comparison is CO2-equivalency. This comparison uses something called Global Warming Potential (GWP), which is a calculation for a given molecule based on the following parameters:

1. Time horizon of interest (commonly 500, 100, or 20 years)
2. The radiative efficiency of the gas of interest, or in other words the ability of the gas to absorb the long-wave infrared radiation
3. How long the gas lasts in the atmosphere

The GWP for each GHG is calculated by the Intergovernmental Panel on Climate Change (IPCC). This value is used to determine the climate forcing of a unit of mass of a particular molecule relative to an identical mass of CO_2. This background should be considered when CO2-equivalence is discussed in future portions of this introductory chapter.

Rise in global temperature leads to greater intensity and frequency of floods, storms, earthquakes, heat waves, droughts, wildfires, and many other natural phenomena that are related to increased hardship and lower quality of life for us humans and the many other beautiful forms of life existing on Planet Earth [15], [16]. The scientists in Table 1 provide more knowledge on topics including global warming, the resilience of earth, tipping elements, planetary boundaries, the severity of these issues, and steps we need to be taking as a species. It is highly recommended to investigate these names and their bodies of work if these topics are of further interest.

Table 1: Climate Change Scientists

Name	**Background**	**Relevant Studies**
Johan Rockström	*Potsdam Institute for Climate Impact Research* *Institute of Environmental Science and Geography, University of Potsdam* *Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden*	- Development of the planetary boundaries' framework - Integrative studies of Earth system sustainability - Research on global environmental limits

Michael E. Mann	*Director, Earth System Science Center of Pennsylvania State University*	- Reconstruction of past climate changes (e.g., the "hockey stick graph") - Paleoclimate reconstructions - Climate modeling and attribution studies
James Hansen	*Former Director of NASA's Goddard Institute for Space Studies* *Professor Emeritus at Columbia University*	- Early and influential climate model simulations - Research on greenhouse gas forcing - Public advocacy and policy-related studies, including notable congressional testimony
Gavin Schmidt	*Director of NASA's Goddard Institute for Space Studies*	- Advances in climate modeling and reanalysis techniques - Studies on global warming attribution - Science communication and public engagement in climate issues
Katharine Hayhoe	*Texas Tech University*	- Regional climate projections and impacts studies - Investigations of climate change effects on communities - Science communication and climate outreach initiatives

Stefan Rahmstorf	*Potsdam Institute for Climate Impact Research, University of Potsdam*	- Research on sea level rise and climate sensitivity - Studies on ocean–atmosphere interactions - Integrated assessments of climate change and its regional/global impacts

SECTION 1.1.2: METHANE, WHAT IT IS AND WHAT IT DOES

Moving forward, the discussion will begin to narrow in scope from the entirety of global warming to one molecule, methane. Methane is of particular importance in the solution of the complex web that is the problem of global warming. Oftentimes labeled the second most important greenhouse gas, methane has a GWP of 27.9 (carbon dioxide representing a 1) on a 100-year time horizon. Though, if the gas is considered on a 20-year time horizon which is closer to its average lifetime in the atmosphere (around 12 years), the calculated GWP is 81.2. Outside of methane's direct ability to absorb and re-emit a large amount of infrared radiation relative to carbon dioxide, it is also converted after its lifetime in the atmosphere to tropospheric ozone, stratospheric water vaper, and carbon dioxide, leading to further radiative forcing and other climate issues post-decay [17], [18]. Due to the shorter lifespan of methane in the atmosphere, if humans were to eliminate all methane emissions the concentration in the atmosphere would drop to pre-industrial levels relatively quickly. This makes methane a unique and well leveraged tool for fighting global warming, as a drastic reduction in methane emissions could reduce warming by as much as 0.5°C in the next two decades [19].

Dropping methane emission to 0 at the drop of a hat is a thought more easily verbalized than executed, however. Methane is an integral part of the infrastructure we have developed today, particularly in heating applications as it is the main constituent of natural gas. The development of our modern infrastructure is not a result of simply ignoring sustainable alternatives. The historical timeline, scientific understanding along that timeline, and economic factors all played roles in getting us to where we are today. The justifications for this development can be provided in a general sense as the following:

1. The infrastructure for natural gas distribution was developed in the 20th century as the gas was used for both lighting and heating prior to the widespread adoption of electricity. The residential demand grew fifty times from 1906 to 1970 [20].

2. Natural gas is oftentimes cheaper than electric heating, especially in regions of abundance and harsh winters [21].
3. Industries that rely on high temperatures achieved easily with combustion would suffer financially if they had to utilize electricity to generate comparable temperatures.
4. Transitioning away from gas to electric heating entirely would require massive electricity grid infrastructure upgrades.

The goal in the writing of this brief background for methane is to provide the reader with some understanding of methane's impact on our current world and also the opportunity methane provides in the battle for sustainability. To provide further understanding of methane's contribution to global warming and the current opportunities to cut down on its emission, an overview of its historical emission and sources will be discussed in the following section.

SECTION 1.1.3: EMISSION OF METHANE

Over the past 250 years, atmospheric methane (CH_4) concentrations have surged by approximately 161.9% [22] due to human activities, making it a major contributor to climate change. The urgency surrounding methane mitigation has grown in recent years as new data indicate that emissions are increasing at a rate aligning with worst-case climate scenarios. Over the past decade, methane concentrations have been tracking the Representative Concentration Pathway (RCP) 8.5, the most extreme warming scenario assessed by the Intergovernmental Panel on Climate Change (IPCC). If this trajectory continues, it could lead to a global temperature increase of approximately 4.3°C by 2100 [17]. Despite efforts to curb emissions, total annual methane output has risen by 61 million tons, about a 20% increase—over the past two decades. This rise is primarily driven by emissions from coal mining, oil and gas extraction and usage, livestock operations, and decomposing organic waste in landfills, as observable in the data shown in Figure 3. While some regions, such as the European Union and potentially Australia, have successfully reduced methane emissions from human activities, the largest regional increases have been observed in China and Southeast Asia [23].

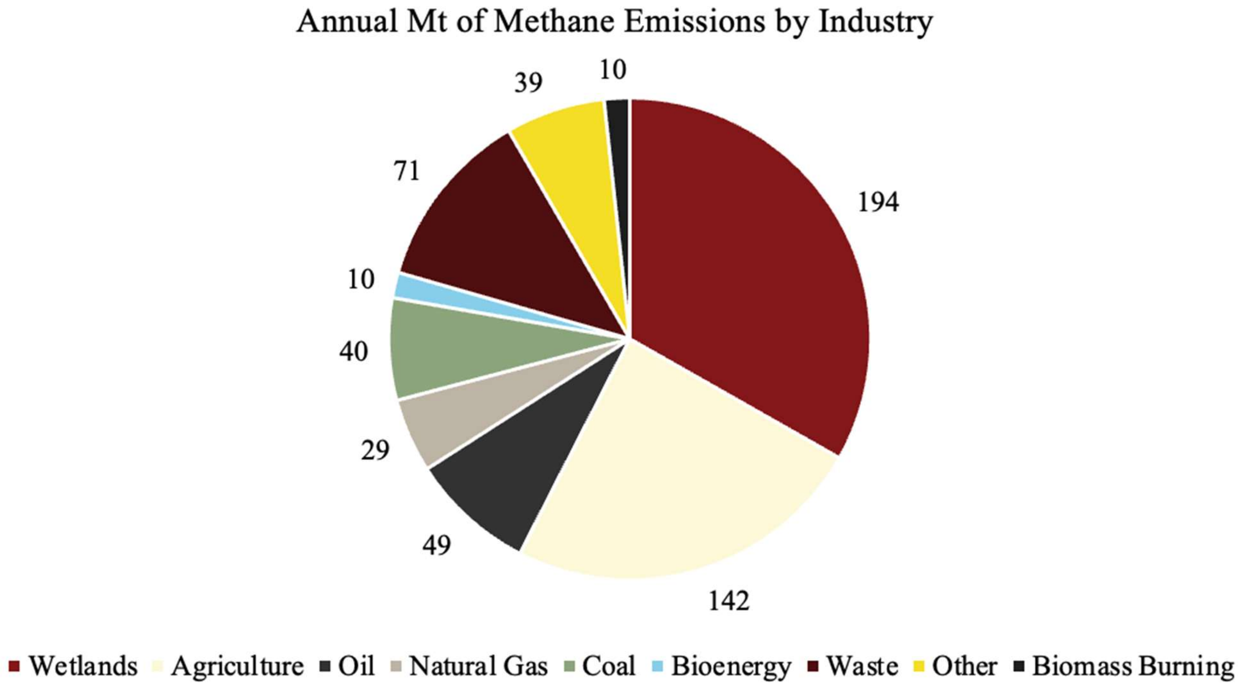


Figure 3: 2023 Breakdown of Annual, Global Emission of Methane by Sector [24]

In the United States, natural gas systems are the second-largest source of anthropogenic methane emissions in 2022, accounting for 24.6% of the country's total methane emissions and 2.7% of total gross greenhouse gas emissions [22]. While methane emissions from natural gas distribution, transmission, and storage have declined by 20.9% since 1990, they still represent a significant portion of overall emissions. Agriculture and waste sectors remain even larger contributors, emitting roughly twice as much methane as the fossil fuel industry.

Despite increased policy attention on methane reduction, new research suggests that existing estimates may severely understate the true scale of emissions. Aerial measurements indicate that U.S. oil and gas producers are releasing methane at over four times the rate reported by the Environmental Protection Agency (EPA). These findings indicate that industry-reported data significantly underestimates emissions, with actual values exceeding corporate emissions targets by a factor of eight. MethaneSAT, a satellite developed by the Environmental Defense Fund and launched in early 2024, is expected to provide further high-resolution insights into global methane emissions as it enters full operation in 2025 [25]. While regulation is improving in focus and stringency, this shows that current monitoring and transparency needs to improve to keep up with the expectations of regulation. As more comprehensive data becomes available, the global community will have a clearer understanding of methane's true contribution to climate change. The challenge ahead lies not only in improving emissions monitoring but also in implementing effective mitigation strategies. As stated, methane's relatively short atmospheric lifespan and heavy short-term impact presents opportunity to make a measurable impact on global temperatures in the coming decades.

SECTION 1.1.4: CURRENT REGULATION – NATIONAL AND ABROAD

Particularly for the U.S., under the Clean Air Act (first established in 1970), the EPA introduced multiple rulemakings for methane emission reduction in 2021, with the final rule having gone into effect May of 2024 [26]. From Section 111(b)(1)(B) of the clean air act, the EPA must identify major sources of air pollution to regulate, establish the standards of performance for the pollutants emitted by new and old sources in the previously identified categories, and establish the guidelines for states to produce standards of performance for pollutants [27]. With these new regulatory actions in the U.S., it is expected that between 2024 and 2038 there will be a reduction in methane emission to the tune of 52.6 million metric tons [26].

Beyond the United States, international regulatory efforts and agreements are also aiming to reduce methane emissions in the short term. The European Union (EU) has implemented stringent regulations targeting methane emissions from the fossil fuel industry. In May 2024, the EU adopted a new regulation requiring fossil fuel companies to monitor, report, and verify their methane emissions while implementing robust leak detection and repair (LDAR) programs [28]. Additionally, strict limitations on routine venting and flaring have been introduced, allowing these practices only in emergency situations [29]. To ensure compliance, the EU plans to establish a maximum methane intensity standard for imported fossil fuels by 2030, compelling global suppliers to meet these requirements [28]. On an even broader scale, the Global Methane Pledge, introduced during COP26 in 2021, represents a key commitment to reducing methane emissions [30]. Led by the United States and the European Union, the initiative aims to cut global methane emissions by at least 30% from 2020 levels by 2030. As of 2024, 159 countries and the European Commission have signed onto the pledge, reinforcing the urgency of mitigating methane's climate impact [31].

These regulatory efforts highlight the growing recognition of methane as a critical driver of climate change. While the U.S. and EU have taken concrete steps to regulate methane emissions, the success of these measures will depend on enforcement, technological advancements in emissions monitoring, and global cooperation. As new technologies such as MethaneSAT provide high-resolution data on methane leaks, policymakers will have more accurate insights to refine and strengthen methane reduction strategies worldwide.

SECTION 1.2: INTERNAL COMBUSTION ENGINES

Now, to further shrink the scope of this introductory chapter towards the particulars of the project of this thesis, a discussion of internal combustion engines will be had. Understanding their role and impact is crucial to address the challenges at hand and to have the capability to analyze and develop effective solutions that help the industries that rely on them adapt to increasingly hefty regulations.

SECTION 1.2.1: IMPACT ON THE EMISSION OF REGULATED MOLECULES

The development of the internal combustion engine gave rise to near limitless possibilities in application and helped to fuel the industrial revolution to the grand scales we have achieved today. The origins of the internal combustion (IC) engine can be traced to the 1860s, in which J. J. E. Lenoir's coal-gas powered engines were produced in the quantity of around 5000 units. During operation of these engines, a coal-gas and air mixture is pulled into the combustion chamber of a piston-cylinder assembly and then combusted (at atmospheric pressures) during the same downward stroke to provide momentum to the piston. Then the combusted gases were exhausted during the pistons travel back toward TDC to complete a cycle. These engines achieved energetic efficiencies of around 5%. From the late 19th century to early 20th century, critical discoveries and engineering from the likes of Nicolaus Otto, Eugen Langen, Rudolf Diesel, Dugald Clerk, James Robson, and Karl Benz caused great forward momentum in internal combustion engine development – greatly increasing power to weight ratios and thermal and mechanical efficiencies [12]. This progress ignited further research and development, leading to the great achievements we have today in IC engine design, manufacture, and performance.

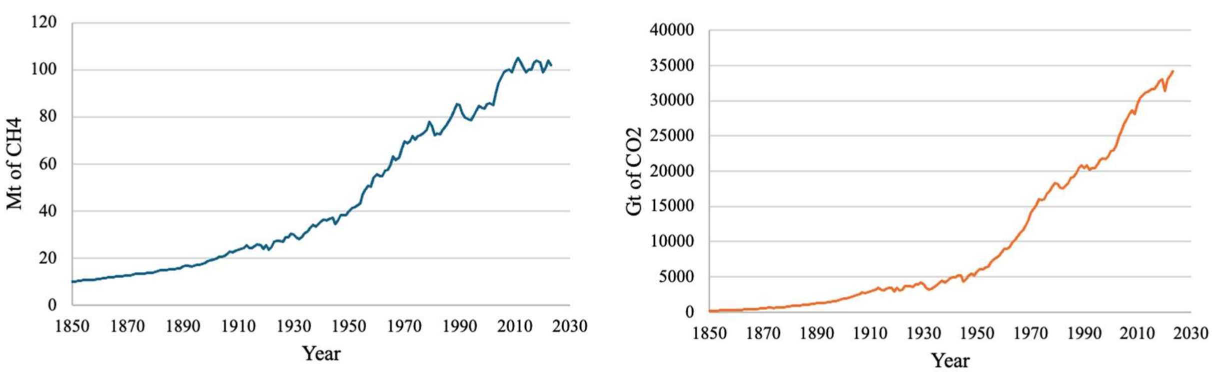


Figure 4: Carbon dioxide and methane production from energy (as defined by IPCC 2006 Guidelines) sectors [32]

As called out in Figure 4's caption, the energy sector's subsectors from the PRIMAP dataset [32] are defined in the IPCC 2006 guidelines [33] and includes a broad range of technologies. These subsectors include electricity and heat, manufacturing and construction, transportation, and other fuel combustion processes. These different processes, activities, and industries that the subsections of the energy segment are made up of are heavily reliant on internal combustion engines, and the energy segment is where IC engines make their biggest contribution in emissions. The correlation in increase in internal combustion engine production and usage is direct with the increase in production of GHG emissions including methane and carbon dioxide.

Modern gasoline and diesel internal combustion engine thermal efficiencies range from 20 to 40%, occasionally achieving close to 50%. Though older examples of these engines (two-stroke engines, carbureted instead of electronically controlled direct injection engines, etc.) often fall short of modern efficiencies, issues of infrastructure, economics, and downtime mean that they are still heavily employed in industries today.

SECTION 1.2.2: NATURAL GAS COMPRESSION AND TRANSPORTATION INFRASTRUCTURE

As mentioned, there are still sectors that are heavily reliant on older IC engine technologies to accomplish their goals. One of these sectors is natural gas compression and transport. The data in the following figures and tables from the Environmental Protection Agency can be used to summarize the impact that this industry has on the emission of methane and carbon dioxide.

Table 2: Natural gas compression infrastructure in the U.S. [34]

Data Element	1990	1995	2000	2005	2010	2015	2018	2019	2020	2021	2022
National station counts	1730	1673	1748	1678	1834	1776	1853	1886	1879	1960	2002
National Compressor Counts	7654	7858	8209	7591	7601	7333	7557	7638	7587	7642	7688
Reciprocating Compressors	6956	6939	6896	6050	5731	5368	5349	5269	5129	5002	4965
Centrifugal Compressors	698	919	1313	1541	1870	1965	2208	2369	2458	2640	2723
Dry Seal Centrifugal Compressors	0	196	545	819	1136	1210	1471	1611	1688	1851	1980
Wet Seal Centrifugal Compressors	698	723	768	722	734	755	737	758	770	789	743

Of the data shown in Table 2, the AJAX compressors used in this study fall into the reciprocating compressors category. These particular models, and others from a similar production timeframe (1940s-1960s) were focused on reliability, durability, and power – meaning fewer moving parts and simple designs. This to say, the focus was not producing the most efficient engine possible with the fewest emissions but producing the engine that got the job done well.

To relate this particular industry's impact back to the emission of harmful emissions including methane, the following data in Table 3 and Figure 5 can be observed to obtain the estimated emission of methane and carbon dioxide from the reciprocating compressors and total emissions associated with the natural gas systems infrastructure in the United States.

Table 3: National Methane and Carbon Dioxide Estimates from U.S. NG Infrastructure [34]

National CH4 Estimates (Kilotons)											
Emission Source	**1990**	**1995**	**2000**	**2005**	**2010**	**2015**	**2018**	**2019**	**2020**	**2021**	**2022**
Reciprocating Compressor	-	-	-	-	-	349	348	342	333	325	323
Total Emissions - 2024 GHGI	2285	2152	1949	1645	1521	1513	1470	1448	1468	1421	1413
National CO2 Estimates (Kilotons)											
Reciprocating Compressor	-	-	-	-	-	10	10	10	10	10	10
Total Emissions - 2024 GHGI	179	174	171	181	227	219	537	1224	2030	874	1185

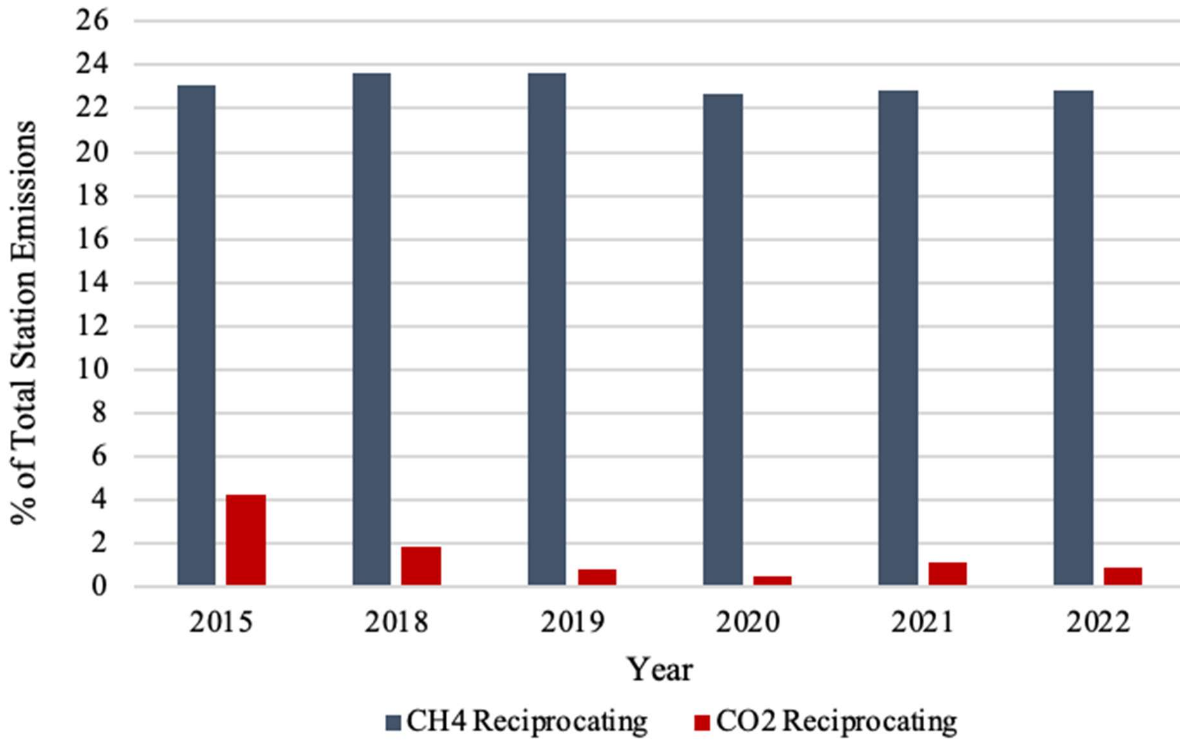


Figure 5: Reciprocating Compressor Contribution to Total Emissions from U.S. Natural Gas Infrastructure [34]

It should be noted that reciprocating compressors make up nearly a quarter of the methane emissions from the natural gas infrastructure of the United States. As noted from Section 1.1.3, the total methane emissions in 2022 (as estimated by the EPA) are close to 27.3 million metric tons [22]. Given this statistic, the natural gas infrastructure accounts for somewhere around 5% of the total methane emissions in the United States. Though this is a small section, it should not be overlooked, and it is not insignificant. We must chip away at the 1%, the 3%, the 5% to get to the target of sustainability.

SECTION 1.3: HOW TO ADDRESS THE CURRENT INFRASTRUCTURE

Given the background discussed to this point – narrowing from the entirety of the problem of global warming down to the contributions (in the form of emission of methane and other products of incomplete combustion) of reciprocal natural gas compressors to global warming – the next logical step in the layout of this thesis is to have some discussion as to what to do about it, and discussion as to the justifications and reasoning for which path the work of this thesis follows to attack the problem. What should be done to reduce the emissions produced by the older technologies used in this sector?

SECTION 1.3.1: TO REPLACE OR TO RETROFIT

The first logical consideration one might make after learning that there are reciprocating natural gas compressors that utilize older, less efficient combustion technology, is to simply replace them with modern technologies. However, this is easier said than done. The main factors that limit this from happening are infrastructure, economics and lack of incentives, malleable emission regulation, and market volatility and uncertainty.

Transitioning from conventional reciprocating natural gas compressors to modern alternatives, such as electric-driven compressors, requires substantial modifications to existing infrastructure. One of the most pressing concerns is the availability and reliability of electrical infrastructure [35]. Electrification necessitates the construction of new substations, upgrades to transformers, and reinforcement of the electrical grid to handle increased demand. Many compressor stations operate in remote locations where grid access is limited or nonexistent, making the expansion of electric infrastructure particularly challenging. The high capital investment required for these upgrades, coupled with long lead times for project completion, poses a significant hurdle to a full-scale replacement strategy.

The financial aspect of compressor replacement is another limiting factor. The cost of acquiring and installing compressors varies significantly, with estimates ranging from $479 to $4,900 per horsepower ($650 to $6,600 per kW), depending on size, technology, and application [36]. Given the thousands of compressors currently in operation, the total capital expenditure for replacement would be immense. Moreover, beyond the initial purchase and installation costs, operators must also account for lost production during transition periods, further increasing the economic burden. As a result, many companies seek solutions that maximize cost efficiency while still achieving compliance with emissions regulations, making retrofitting a more attractive option in many cases.

Regulatory landscapes play a pivotal role in determining the feasibility of compressor replacement versus retrofitting. While environmental policies continue to tighten restrictions on methane emissions and other greenhouse gases, regulatory bodies have, at times, demonstrated flexibility in permitting expansions of natural gas infrastructure even when these expansions appear inconsistent with long-term emissions reduction goals [37], [38]. This regulatory leniency is often motivated by concerns over energy reliability and supply security, as abrupt transitions could disrupt the market. Additionally, existing financial incentives to modernize compressor fleets are limited, with many operators lacking the necessary subsidies or tax credits to justify the substantial cost of full-scale replacement.

Investing in new gas infrastructure carries the risk of becoming a stranded asset, particularly in an era of shifting energy policies and market transitions toward decarbonization [39]. The natural gas industry is experiencing growing uncertainty due to evolving regulatory frameworks, fluctuating energy prices, and the increasing competitiveness of renewable energy sources. If natural gas demand declines more rapidly than anticipated, newly installed compressor systems could become obsolete before reaching the end of their intended service life, resulting in significant financial losses for operators. This uncertainty further strengthens the argument for retrofitting existing infrastructure rather than committing to full replacement.

Outside of these difficulties fighting against the practicality of replacement, retrofitting the existing infrastructure often yields sufficient results in adherence to increasing regulation in addition to solving issues of infrastructure change and economic burden. The retrofit kit currently being developed for integral AJAX natural gas compressors at the University of Oklahoma's Sustainable Energy and Carbon Management Center (SE&CMC) includes a range of sensors for combustion optimization, a bypass valve for more precise control of the air to fuel ratio, and a PLC and cloud server to provide the operator with access to sensor data and performance metrics [40].

SECTION 1.3.2: PROJECT OVERVIEW

With the wholistic background in place, now an introduction to the specific work contained in this thesis will be had. The overarching goal in the completion of the work written out in this thesis is to aid in the progress towards reducing the GHG emissions from older IC engines still largely in use in the natural gas compression and natural gas transportation infrastructures in the United States. The strategy that is pursued in this work – due to advantages of cost, government support, and downtime – is reliant on retrofitting as discussed in Section 1.3.1. With support and collaboration with Cooper Machinery Services, WAGO Automation, Mid Continent Rental, and Kansas State University at the Sustainable Energy & Carbon Management Center (SE&CMC) through the University of Oklahoma, a series of older industrial engines are being tested and evaluated for various solutions to improve combustion and reduce emissions.

More specifically, the end-goal at the completion of the project of this thesis is to have produced, tested, and validated models that are capable of predicting the methane emissions from two AJAX engines, and to make these models of minimal size such that it is possible to incorporate them into the PLC and cloud computing systems used in the retrofitted engines (as defined in Chapter 3, Section 1). The engines being used in this work include a retrofitted AJAX DPC81, and a retrofitted AJAX DPC105. Table 4 can be observed to see the requirements and tasks for the work of this thesis:

Table 4: Requirements and Tasks of the Work of this Thesis

Requirements	Tasks
Develop virtual methane sensors for commonly utilized NG compressors.	Collect quality data for training the machine learning models. This task includes the necessary development and integration of control system networks.
Validate the virtual sensors.	Train models that have acceptable training, testing, and validation accuracy, and minimal computational size for ease of integration into a PLC control system network.
The sensors must be of appropriate size for integration into the PLC system for continuous emission monitoring.	Develop established, modern mathematical physics-based models using the experimental data to evaluate the performance of the virtual sensors.

SECTION 1.3.3: RECAP AND ADDRESSING THE IMPACT OF THIS WORK

The preceding sections have established the urgency required in action taken against the emission of greenhouse gases from industrial natural gas engines, and the impracticality of full-scale replacement of legacy infrastructures. This work approaches the issues with the perspective of improvement of existing technology, using modern technologies such as machine learning and mathematical modeling for emission prediction and control. The work previously conducted at the SE&CMC that this work builds on includes the following references: [41], [42], [43], [44].

At the core of this thesis are two guiding research questions:

1. Can transformer-based machine learning models, customized for noisy, low-frequency engine sensor data, predict methane emissions accurately?
2. Can these models generalize effectively across varying operating conditions without requiring extensive hardware retrofits or complex physical modeling?

To answer these questions, a dual approach is taken: the development of customized transformer architectures trained on real-world engine data, and the benchmarking of their performance against outputs from a simplified two-zone Chemkin combustion model.

The structure of the thesis reflects this approach. Following the introduction, Chapter 2 reviews the technical background and relevant literature. Chapter 3 outlines the methodology, including dataset preparation, model customization, and simulation strategies. Chapter 4 presents and analyzes the engine operation results as well as the machine learning model performance.

Chapter 5 discusses the broader implications and evaluates the performance of the Chemkin model. Chapter 6 concludes by summarizing contributions and identifying future research directions.

Although focused on a specific industrial challenge, the insights gained here extend beyond the immediate application. They demonstrate the potential for artificial intelligence to complement traditional engineering approaches in driving emissions reductions and advancing toward a cleaner, more resilient energy infrastructure. This work, incremental yet impactful, contributes to the ongoing transition toward carbon-neutral operation by ensuring that critical legacy assets can operate more efficiently, intelligently, and sustainably. The ability to monitor, predict, and optimize emissions in real time has significant implications beyond the specific engines studied in this thesis. Insights gained from applying AI to combustion emissions may inform future low-emission engine designs, pushing the transition toward carbon-neutral energy solutions forwards. This thesis, while focused on a specific industrial challenge, represents an incremental yet meaningful step in the transition toward a more sustainable energy infrastructure. By ensuring that aging but critical industrial assets can operate cleaner, smarter, and more efficiently, this research directly addresses the regulatory, economic, and environmental concerns outlined in the introduction.

CHAPTER 2: TECHNOLOGICAL BACKGROUND AND LITERATURE REVIEW

The following sections have been written to provide an overview of the current techniques available for the prediction of emissions and engine performance for internal combustion engines, and a literature review of work similar to that found in this thesis using those techniques. The central topics are mathematical combustion modelling and artificial intelligence modelling. Due to the engines that are being evaluated in this work, the combustion modeling topics will not be holistic to all of combustion modeling for IC engines, but will be tailored toward more specifically spark ignited, gaseous fuel two-stroke and four-stroke engines.

SECTION 2.1: COMBUSTION MODELLING

Combustion modeling is a complex task, with a large variety of scenarios that require unique solutions to achieve accuracy in theoretical results compared to experimental observation. In modern engine development, simulation is used extensively to optimize performance, efficiency, and emissions of engine design prior to physical development. This topic is exceptionally relevant to the work of this thesis, as mathematical modeling of internal combustion engines serves as a way to validate the experimental performance observed, and to predict emissions with a degree of accuracy directly proportional to the accuracy of the model's prediction of the fundamental parameters including combustion chamber pressures, combustion chamber temperatures, fuel and air mixture properties, and flame propagation physics. However, it should be noted that the aim of this work is not to provide the world with a novel mathematical model or new approach towards estimating some physical parameter necessary for modeling combustion processes. The aim of this work is to produce virtual sensors that can accurately, continuously measure methane PPM in industrial natural gas engines using low-cost networks of sensors – and this is kept in mind when selecting the modeling approach for validation of the combustion conditions.

In terms of modeling, combustion engines are broken apart by their fuel and oxidizer mixing method and the combustion ignition method, as these factors are what determine the depth of consideration required of the fluid movement and interaction as well as flame propagation. The following combustion engine concepts (based on those considerations) are the most common concepts used to shape the assumptions in modeling practice.

Table 5: Engine types based on characterization of fuel mixing and method of ignition

Engine Type	**Description**
Homogeneous charge compression (HCCI) engines	This form of engine utilizes a premixed, homogeneous mixture of fuel and oxidizer that is compressed until auto-ignition occurs due to elevated temperature and pressure.

Diesel or charge compression (CI) engines	This form of engine compresses a charge of air, injecting fuel shortly prior to the piston reaching TDC, where the heat of compression enacts the auto-ignition of the mixture. This requires consideration of spray, turbulence, and ignition delay phenomena.
Spark-ignition (SI) engines	Spark-ignition engines can have a variety of fuel and oxidizer mixing methods, requiring different modeling approaches. The combustion ignition is started via a spark at a known location or locations, causing flame propagation to occur from these points. This requires consideration of flame speed, turbulence, and heat release phenomena.

Combustion, and particularly combustion within a moving geometry such as that found in IC engines, can be heavily affected by minute changes in a wide range of physical phenomena. The more complex the scenario becomes, the more complex the modeling approach can become in terms of physical consideration, which generally indicates higher computing costs, longer modeling timeframes, longer solution timeframes, and higher education and skill requirements of the analyst. Generally, the models may be broken up between being thermodynamic or fluid dynamic in their nature; in some scenarios the consideration of the fluid motion is not vital to achieve accuracy in the desired outputs (thermodynamic considerations only) and in some it is vital to consider the geometry and flow of the fluids (fluid dynamic) along with the thermodynamic properties [12]. The engine types listed in Table 5 are the main overarching categorizations, there can be many variations in the forms of these engines. The spark-ignition engine for example, may receive a premixed air and fuel charge via carburetion, or it may receive direct injection of the fuel to mix with the incoming fresh air flow. Liquid fuel direct injection may be done using a variety of nozzles, altering the spray pattern and therefore the mixing. This to say, each engine has its own unique geometry and systems – there are always important considerations as to how much fuel and how much oxidizer is introduced into the system, how well the two are mixed, how the geometry and materials affects these parameters as well as the heat transfer, etc. The following sections will briefly discuss the various modelling approaches currently available and their use-cases, with a narrowing focus towards the engines that are being tested and modeled in this thesis.

SECTION 2.1.1: THE VARYING DEGREES OF COMPLEXITY IN MODELING IC ENGINES

2.1.1.1: ZERO-DIMENSIONAL MODELS

Zero-dimensional models are the simplest, least computationally expensive variant of existing mathematical models utilized to evaluate the physics and chemistry of combustion in internal combustion engines today. This categorization can be further sectioned into single-zone and multi-zone models. Multiple zones can allow for approximation of flame propagation effects, temperature variation in the different zones, and more accurate estimation of the downstream variables dependent on these parameters such as the chemical kinetics.

HCCI Models

Homogenous charge compression ignition models are one of the simplest and most common of IC engine models found in literature. Oftentimes these models rely on a single zone to approximate temperature and pressure in the system. As indicated in the name, the zones, single or multiple, are all comprised of the same homogeneous mixture of fuel and air which is ignited via the heat of compression and not via spark. This description may sound odd, as there are practically no true HCCI-based engines currently used in commercial settings – inducing the question as to why this model is so commonly used in research. The reasoning is this:

In this scenario in which the fuel and oxidizer is well mixed, the entirety of the homogeneous mixture is ignited simultaneously (as in there is no flame propagation) as the heat of compression provides the necessary activation energy for the fuel and oxidizer to react. Due to the lack of complexity, researchers may focus solely on the nature of the chemical kinetics of ignition. All of this considered makes up the reasoning behind HCCI modeling prevalence in modern R&D. HCCI combustion, if reasonably achievable in a commercial setting, would be a very efficient process.

The equations that govern the change in mass and energy in a single zone HCCI model are as follows:

$$\frac{dm}{dt} = 0 \tag{1}$$

$$\frac{dT}{dt} = \frac{1}{mC_v}\left(-P * \frac{dV}{dt} + \dot{Q}_{chem} - \dot{Q}_{loss}\right) \tag{2}$$

To determine the mass into the system, these models usually rely on the ideal gas law and an inlet temperature and pressure input, and only the combustion stroke is considered meaning no mass is transferred into or out of the system (Eq. 1). Then, considering the change in energy associated with the pressure increase with the reducing volume (piston moving to TDC), the chemical kinetics are evaluated as activation energies are reached and additional heat is released from reaction. Chemical kinetics for this model are determined via the reaction rates which should be derived from experimental data for the given fuel being considered. For the case of this thesis, all engines run on natural gas. Fortunately, detailed kinetics of methane combustion have been compiled by UC Berkley [45]. Their dataset, referred to as GRI Mech 3.0, contains thermodynamic data and Arrhenius parameters for 325 reactions and 53 species and is widely used to determine kinetics of methane combustion when paired with any model capable of consideration of the necessary thermodynamics that are used to determine reaction rates. This data is used to determine $\dot{Q}_{chem}$ in the energy equation.

Oftentimes, it is advantageous to move from the simplicity of a single zone model into a multi-zone model as a single zone cannot account for the low temperature regions that occur in the thermal boundary layers and crevices. This generally causes under-prediction of incomplete combustion products such as CO and unburned hydrocarbons as well as over-prediction of peak pressures due to unreasonably complete combustion and higher total heat release [46]. The most common HCCI multi-zone models operate on similar governing equations, resulting in the same homogenous mixtures and total pressure in each zone, but varying temperature.

$$\frac{dY_{\{i,j\}}}{dt} = \frac{\omega_{\{i,j\}}}{\rho_j} \tag{3}$$

$$\frac{dT_j}{dt} = \left(\frac{1}{\left(\rho_j * c_{\{v,j\}}\right)}\right) * \left(-p * \frac{dV_j}{dt} + \dot{q}_{chem,j} - \dot{q}_{ht,j}\right) \tag{4}$$

$$\dot{q}_{chem,j} = -\sum h_{(i,j)} * \omega_{\{i,j\}} \tag{5}$$

$$p = \rho_j * R_j * T_j \quad \& \quad p = \frac{\sum_j m_j * R_j * T_j}{\sum_j V_j} \tag{6}$$

In the species conservation equation (Eq. 3), $Y_{\{i,j\}}$ represents the concentration of species i in zone j, $\omega_{\{i,j\}}$ the production rate per unit volume of species i in zone j, and ρ_j the total mixture density for zone j. In the energy conservation equation (Eq. 4), T_j is the temperature in zone j, $c_{\{v,j\}}$ is the specific heat at constant volume for species i in zone j, p is the common pressure shared across all zones, V_j is the volume in zone j, and $\dot{q}_{chem,j} - \dot{q}_{ht,j}$ is the difference between heat released from chemical reactions and the heat loss from the system in zone j. The heat released from chemical reaction can be described by Eq. 5 as the summation of the product of the temperature-dependent specific enthalpy and production rate of species i in zone j. Eq. 6 represents the ideal gas law, allowing all zones of the model to be tied together, as the shared pressure can be calculated, and then utilize this shared pressure in each zone to calculate the mixture density per zone, ρ_j which is then used to update the other equations for a given time step. Multizone models have been shown to achieve very accurate results in the cases for which combustion is similar to homogenous charge compression ignition [47].

However, the simplicity becomes a hinderance when one wants to obtain detailed kinetics that match that of the real scenario of the engine if it the combustion and fluid dynamic physics do not closely resemble that of a homogenous autoignition scenario. For example, in a spark ignited, carbureted engine, the fuel and oxidizer may be well mixed as in the HCCI scenario, but the flame front begins at the spark location and expands from there. This causes changing gradients in energy (temperature, flow rate, heat transfer, and pressure) through the entirety of the power stroke, which has great impact on the kinetics. In a diesel engine, though it may experience charge compression like the HCCI engine, the diesel fuel is generally injected at high pressure just prior to autoignition which greatly affects mixing, making the assumption of a homogenous mixture a poor choice to achieve accuracy.

Homogeneous SI Models

A common method for modeling spark ignition combustion engines, due to simplicity and generally acceptable results concerning the modeled temperatures and pressures achieved in the combustion chamber if the fuel and air mixture is close to homogenous, is a two-zone, zero-dimensional model. One zone represents the unburned fuel and air mass, and the other the burned or combusted fuel and air mass. The separation of zones allows for approximation of flame propagation and its effect on the chemistry within the system. As explained by Heywood, the main assumptions that are employed in this kind of model are as follows: (a) The fuel and air is uniformly premixed, and (b) the volume of the reaction zone in which the oxidation of the fuel is taking place is small compared to the clearance zone (the flame that is propagating through the uniform mixture of molecules and providing the energy for combustion is thin and localized) [12]. These two assumptions are what support the simplification down to an unburned and burned zone.

The most common method for calculating the transfer of mass between these two zones is to utilize the Wiebe function [12], [48].

$$x_b(\theta) = 1 - exp\left[-a\left(\frac{\theta - \theta_0}{\Delta\theta}\right)^{m+1}\right] \tag{7}$$

In Eq. 7, $x_b(\theta)$ represents the mass fraction of homogenous inlet mixture that has been burned at a given crank angle, $\theta.$ θ_0 represents the starting crank angle, and $\Delta\theta$ is the time step or change in crank angle between calculation steps. This function (Eq. 7) can be used to calculate the profile of the mass fraction burned (x_b), where the parameters a and m are adjustable parameters that should be used to fit the curve to the experimental burn profile, which is calculated from experimental in-cylinder pressure values using methods such as that presented by Rassweiler and Withrow [49]**.** From here, the conservation of mass and energy equations for an open system are applied to calculate the transfer of mass from the unburned to the burned zone. The mass flow rate (negative for the unburned zone and positive for the burned zone) can be calculated by multiplying the total mass of the system by the rate of concentration change which is obtained by differentiating Eq. 7. At a given time step (or crank angle), a two-zone model may calculate detailed chemical kinetics of reactions that take place for the species that enter the burned zone if the necessary activation energy is achieved, and the following heat release and accumulated temperature and pressure are sufficient to promote continued combustion.

This form of model is what is used in this thesis to model the pressure, heat release, and kinetics of the engines used.

2.1.1.3: MULTI-DIMENSIONAL COMPUTATIONAL FLUID DYNAMICS (CFD) MODELS

As computational power and availability has increased, the capability to solve the partial differential equations needed for good representation of fluid systems at an adequate number of discretized cells for a given geometry has become a bigger reality in practical application, such as engine development. Modern software such as Converge or Ansys Fluent make these tasks more accessible than they ever have been prior, but producing a high-quality model still requires a much higher degree of involvement than a relatively simple zero-dimensional model. When a 0-D model becomes insufficient for obtaining necessary results and a CFD model becomes vital for a particular problem, the time, energy, and engineering expertise required becomes much higher, which plays heavily into the motivation behind the work of this thesis: to explore the use of AI modeling as a new approach for achieving accurate predictions of engine performance and emissions for complex problems.

The core components that make up the mathematical model forming the basis of the multidimensional CFD models include the differential equations that describe the conservation of mass, momentum, energy, and species concentrations. Momentum is made up of the set of Navier-Stokes equations in the quantity of the number of dimensions being solved for.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathrm{u}) = 0 \tag{8}$$

$$\rho \left(\frac{\partial u}{\partial t} + \mathrm{u} \cdot \nabla u \right) = -\nabla \mathrm{p} + \nabla \cdot \tau + \rho f \tag{9}$$

$$\frac{\partial(\rho E)}{\partial t} + \nabla \cdot [\mathrm{u}(\rho E + \mathrm{p})] = \nabla(\mathrm{k}\nabla\mathrm{T}) + \Phi \tag{10}$$

$$\frac{\partial(\rho Y_i)}{\partial t} + \nabla \cdot (\rho u \mathrm{Y_i}) = -\nabla \cdot J_i + \dot{\omega}_i \tag{11}$$

Eq. 8 is used to represent the conservation of mass, with ρ representing fluid density, and $\nabla \cdot (\rho u)$ the divergence of the mass flux into or out of the control volume. In the Navier-Stokes equation (Eq. 9), $\frac{\partial u}{\partial t}$ represents the local acceleration (local meaning if you were to observe a given point in space how does the velocity of fluid through this point change with time) while $u \cdot \nabla u$ represents the convective acceleration (convective indicating changes that are dependent on spatial changes such as narrowing passageways). The term $\nabla \rho$ on the righthand side of the equation is used to describe the force from gradients in pressure, $\nabla \cdot \tau$ the viscous stresses (the internal friction from resistance of external force), and $\rho * f$ the external body forces. In the energy equation (Eq. 10), the $\nabla \cdot [u(\rho E + \mathrm{p})]$ is used to represent the divergence of energy from the control volume; $\rho E u \left(\frac{J}{m^2 s}\right)$ is the convection of energy carried by the fluid's velocity and $\mathrm{p} * u \left(\frac{J}{m^2 s}\right)$ the work done by pressure. $\nabla(\mathrm{k}\nabla\mathrm{T})$ is a representation of Fourier's law to account for conduction of heat, and Φ is dissipation function representing energy loss due to viscous effects, radiation, and chemical heat release. Eq. 11 represents the species transport. Y_i is the mass fraction of a given species i, making $\nabla \cdot (\rho u \mathrm{Y_i})$ the divergence of the mass of species i. The diffusive flux, or how a species spreads due to concentration gradients is represented by $\nabla \cdot J_i$, and $\dot{\omega}_i$ is the production rate of a species determined by kinetic reaction mechanisms such as those contained in the previously mentioned GRI-Mech 3.0 datasets.

The complex turbulence of the flow in engine conditions is what presents the first impracticalities for analytical solution of these equations, and several approaches have been developed in literature and then implemented into the popular software choices used today. The standard turbulence models across combustion literature and in modern software are Reynolds-averaged Navier-Stokes (RANS), and large eddy simulation (LES), or a hybrid blend of both models. RANS is less computationally expensive and more widely used in application than that of LES models, where the instantaneous Navier-Stokes equations are deconstructed into mean and fluctuating components. These new governing equations (similar to the original analytical equations) capture the average flow behavior over many cycles. The turbulence in this flow is modeled within the momentum equations through additional equations (these are additional terms added to Eq. 9) that estimate the effect turbulent stresses (known as the Reynolds stresses) have on the average flow [12], [50], [51]. LES modeling directly solves for large eddies using the Navier-Stokes equations, while the effects of the smallest eddies are modeled using simplified equations referred to as sub-grid scale models [52]. These modeling approaches may be more simply differentiated in idea by how they determine what turbulence is. In RANS models, the turbulence is defined as deviation in the flow at any instant from the average flow over many cycles (estimation of the effect eddies have on the flow over multiple cycles), while LES defines turbulence as variation from local averages for a *single* cycle – providing more detail for a given cycle than achievable with RANS models [12].

This description of multi-dimensional CFD models is given such that the reader might gain some new knowledge concerning which scenarios might require you to apply these models. When the complexity of a given problem gives justification for use of one of these models, AI modeling may now become a new tool to use to help achieve the desired results for engine design and analysis.

SECTION 2.1.2: SIMILAR APPLICATIONS AND SOLUTIONS

The coupling of physically accurate combustion modeling with data-driven techniques is not new to the field of engine diagnostics and emissions research; however, the targeted application of this hybrid methodology to older natural gas engines in stationary compression service—like the AJAX engines studied here—presents a novel use-case. The existing literature reveals a progression of modeling complexity, from basic single zone models to chemically detailed two-zone and multi-zone simulations, each with varying capabilities to replicate real-world engine behavior. The focus in this review will be on 0-D, two zone models as that is what is used in this paper to attempt to obtain accurate emission results. Chun and Heywood (1987) offered one of the earliest robust treatments of one-zone thermodynamic models, using the first law of thermodynamics to infer heat release and mass burned from cylinder pressure data. Despite the simplicity of the one-zone structure, their work revealed that accurate results could be obtained by using empirical or fit-based profiles for the ratio of specific heats (γ), as long as careful calibration to two-zone models or experimental benchmarks was performed. This work, along with the compilation of knowledge in Heywood's 'Internal Combustion Engine Fundamentals' was a key influence on the combustion models used in Chemkin-style 0D simulations employed in this work [12], [48], [53].

Building on these ideas, Chindaprasert et al. (2006) incorporated chemical kinetics into a two-zone framework using Chemkin to simulate spark-ignition engine combustion and emission formation. Their method allowed accurate CO emission prediction based on pressure-time data, and they highlighted how wall heat losses and crevice flows affect incomplete combustion species. Their results demonstrate the strengths of using pressure-driven zonal models for emission prediction and validate the kind of emissions modeling used in this thesis for legacy SI natural gas engines [54].

Del Vescovo et al. (2019) and Wang et al. (2020) further advanced the two-zone modeling field, with Del Vescovo's model introducing more depth and consideration for pre-spark heat release (with pre-spark heat release normally only considered by increase in temperature with rising compression pressures in most two-zone models) caused by low temperature chemical reaction between the present species [55]. This pre-spark molecule interaction, depending on the engine geometry and parameters, can release enough energy to affect the thermodynamic, compression ratio, and ignition timing optimization, making it important to consider when trying to limit knocking while also optimizing combustion completeness and work output. Wang's work contributed a novel numerical approach that efficiently simulates inter-zone energy and species exchange, highlighting improvements in computational speed while retaining performance in chemical kinetic calculation [56]. They achieve this by replacing the ODE's that normally govern the unburned zone properties with algebraic equations, making the assumptions that the process occurring in the burned zone is quasi-isentropic – as in entropy is mostly conserved and heat transfer is relatively small over short intervals. This makes the model much less stiff than the normal set of ODE's allows, which makes it easier to use larger time steps and consume less computational power while still maintaining a good level of accuracy for their application.

A more recent study by Yuan et al. (2021) introduced hybrid physics–AI models for predicting emissions in dual-fuel engines, showing strong predictive accuracy while reducing computational needs through the use of AI [57]. Their work validates the use of reduced-order physics as training references for AI models, paralleling the methods used in this thesis to validate transformer-based methane emission predictors and evaluating against Chemkin-generated reference outputs. Yuan et al generates their own two-zone (Wiebe governed) SI engine model to predict burned and unburned zone parameters which are passed to Chemkin Pro for detailed kinetic analysis. Once the model was validated against experimental data, an AI algorithm was trained on the kinetics outputs and mathematical model inputs to make accurate predictions of emissions results. For the case of this thesis, it is apparent with analysis and discussion of the results that this exact route of training an accurate algorithm from the mathematical model would not be possible given the equipment and engines used. The large-bore, two-stroke combustion process present in these engines presents many complexities that make mathematical modeling a tall-task. Instead, AI models were trained from sensing equipment and compared against the mathematical models for further insight. However, inspiration was certainly garnered from this work.

SECTION 2.2: VIRTUAL SENSING FOR COMBUSTION OPTIMIZATION

To understand why artificial intelligence is increasingly integrated into modern industry – and why it is also central to this mechanical engineering thesis – it is important to consider the big picture. In industrial settings, the following topics have some major differences when compared against academia:

1. The nature of research interest – which problems are investigated and why – differs substantially between industry and academia. In short, interest follows the money, the growth and sustainability of a business or product in industry. Whereas in academia an interest might simply be borne due to the curiosity of the researcher, if funding permits.
2. The definition of success – like interest, success is determined by how well the goals of growth, sustainability, and profitability are met in a company. In academia, success is oftentimes simply checking the box of achieving new understanding, in advancing theoretical knowledge for *future* application.
3. Solution robustness against theoretical accuracy – in industry, the bottom line is not to achieve 99% accuracy in the case of development of models for solving difficult problems. The bottom line is to produce a model that gives you results that are accurate enough, reliable enough, and deployable enough that the goals are met to a satisfactory degree. In academia, with the goals of improved understanding, theoretical accuracy and generalization to diverse scenarios is held to a high position on the ladder of priority, which oftentimes leads to reduction in practicality and immediate real-world application.

Artificial intelligence has, especially within the last 5 years, begun to revolutionize nearly every industry through the development of revolutionary capability of approximation of outputs from highly complex input variable relationships. The history, application, theory, and future prospects of artificial intelligence will be discussed in the following sections. Though this work is of academia from an academic institution, it is a work of academia that is focused on bridging the gap between theory and industrial application, and thus the goals outlined in Chapter 1 Section 1.4 must be considered in that context.

SECTION 2.2.1: A BRIEF HISTORY OF MACHINE LEARNING

For the information of the reader and to provide a background and context to the relevancy and growth of machine learning in industry today, a history of the science will be discussed. To begin this short history, it is important to understand the terminology “machine learning.” This phrase captures a huge range of approaches all from the fields of computer science and statistics to approximate patterns recognized in data automatically. The uncovered patterns may be then used to predict future data, or to improve decision making beyond the extent capable via human observation of the data [58]. This definition throws a broad net, but the visions of scientists, visionaries, and creatives since the inception of the computer has included grand thoughts of possible applications of machine learning and artificial intelligence.

Some vital works of literature and media that influenced many scientists in the field of artificial intelligence, computer science, and machine learning might be traced back some time before the first computers were developed in the early to mid-20th century. The first time the word ‘robot’ was introduced to the English language was in the Czech writer Karel Čapek's play 'Rossum's Universal Robots' in the year 1920. The 'robots' in this play were artificial laborers made from synthetic organic matter without original thought, though they eventually rebel against humanity and cause extinction of the human race, presenting perhaps the first concept of artificial intelligence outgrowing and domineering over humanity – which we have seen crop up again and again across literature and film. Some other popular works that influenced both future media and scientists alike include 'Metropolis' (1927) by Fritz Lang, 'Robot' (the series, 1940s-50s) by Isaac Asimov, Alan Turing's 'Computing Machinery and Intelligence' (1950) which is the source of origin for the Turing Test in modern artificial intelligence metrics, 'The Sentinel' (1951) and 'A Space Odyssey' (1968) by Arthur C. Clarke, and Philip K. Dick's 'Do Androids Dream of Electric Sheep?' (1968). The potential originator of the concept that neural network's 'neurons' are based on is derived from a book called The Organization of Thought by Donald Hebbs in 1943 [59], [60]. In this book, the concept known as Hebbian Learning was born, which comes from the biologically-based concept that when neurons in contact with one another repeatedly fire in good timing with each other, the axon of the first cell develops, or enlarges the synaptic knobs in contact with the soma of the second cell, but when they are activated at different times, their physical connection weakens and the size of the synaptic knobs decrease.

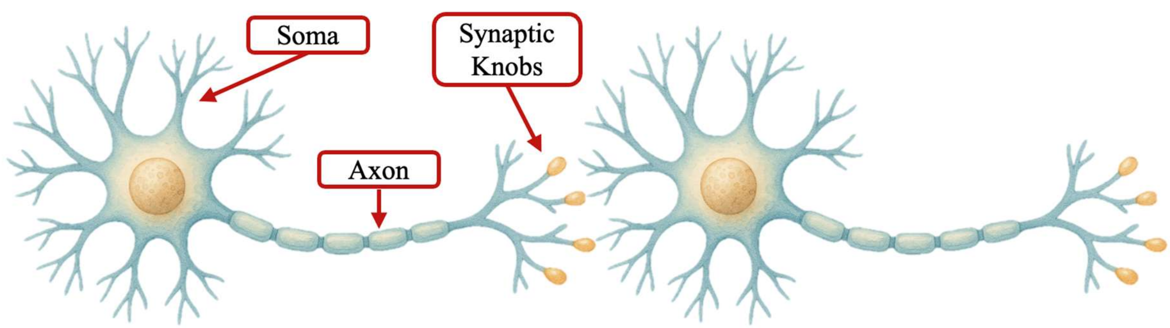


Figure 6: Hebbian Nerve Cell Depiction (Generated through ChatGPT)

Though this concept illustrated in Figure 6 is not entirely truthful to the true biology occurring, Hebbs' ideas that led to the concept of Hebbian Learning has guided the development of many modern ML algorithms, such as the earliest perceptron [61] and later the multilayer perceptron, where a neuron (or node) may be labeled as having a weight (trainable strength of output signal) and bias (allows the neuron to fire or not fire despite external stimuli). The layers of the multilayer perceptron are essentially sets of linear equations that the input variables are passed through to provide some output or outputs, where the slopes (weights) and y-axis intersection points for the case of $x = 0$ (biases) of these linear equations are trainable parameters to match the model's output to the ground truth outputs from the training dataset.

Outside of these early influences, the practical development of algorithms started shortly after the first computers were brought into reality. Early in the 1950s, Arthur Samuel of IBM developed a low-memory (due to the restraints of computing power) computer program to play the game of checkers. In 1952, Samuel actually coined the term ‘Machine Learning’ [60]. From this time in history to now, focus on research and application of models has shifted from models that are more statistically based back to neural networks, as computing power and practical successes have influenced history. Initially, after difficulties with the early perceptron, machine learning was dominated by models such as linear regression, logistic regression, support vector machines, K-nearest neighbors, regression trees, and random forests. These models are well-grounded in statistical mathematics and work well for nicely structured, smaller datasets that fit their approach [62]. After the concept of backpropagation was developed in 1986 by Rumelhart, Hinton, and Williams [63], and consecutive developments from a range of researchers in the field [64], [65], [66], we arrive to the transformer’s development in 2017 [67]. Since this time, the application of models that employ neural networks, especially the transformer, has spread to nearly every industry – and is still on a rapid rate of progression at the time of this thesis. The popularity and grand quantity of use cases for neural network models can be understood by the frequently used name applied to them: “universal function approximators.”

SECTION 2.2.2: SIMILAR APPLICATIONS AND SOLUTIONS

To both validate the approach taken and to show the advantages of AI models in similar applications, a review of some literature in the field will be discussed. The integration of machine learning models for use in combustion optimization and diagnostics, as well as real-time engine control has been explored in a multitude of studies to date. A review of these studies can provide insight into best practices, challenges, and effectiveness. Recent advancements in machine learning (ML) have significantly improved predictive capabilities across various emission-related applications. An emphasis on transformer-based and deep learning models for predicting pollutant emissions in internal combustion engines and related combustion processes is made in this review. Given the timeline of advancement of these models as opposed to the multi-generational development of the combustion modeling techniques discussed in Section 2.1, the papers referenced in this literature review will generally be from more recent years.

To begin, Ricci et al. (2024) introduced a comprehensive deep learning framework to predict pollutant emissions from internal combustion (IC) engines, specifically targeting NOx, CO, and hydrocarbons [68]. In their study, success is found for their particular application using a normal feed-forward neural network, demonstrating the feasibility of employing these models as virtual sensors for real-time emission estimation in gasoline engines if input data is of sufficient quality and type (i.e. captures all necessary physics). Their approach highlights the critical role of advanced ML techniques in improving real-time predictions of emissions, contributing significantly to the potential implementation of these tools in onboard diagnostics and emissions management systems.

Similarly, Li et al. (2021) developed a deep-learning differentiation model for transient NO_x emission prediction from diesel vehicles [69]. They employed advanced noise-reduction techniques, including Singular Spectrum Analysis and Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, to enhance data fidelity prior to applying gated recurrent units and support vector regression models. This multi-step preprocessing and prediction strategy provided enhanced predictive accuracy for transient emissions, showing the necessity of noise-handling and feature engineering techniques when working with dynamic real-world vehicle data like that of NO_x emissions.

Meresht et al. (2023) investigated methane (CH_4) and nitrogen oxide (NO_x) emissions from heavy-duty natural gas engines using various ML regressors, including linear regression, Ridge regression, Support Vector Machines (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) [70]. The Random Forest model notably outperformed the other ML algorithms, achieving an accuracy within ±10% error for 95% of test cases. Again, this paper uses relatively simple models like that of Ricci's work but still achieves highly accurate results due to the quality of their input data. In the case of Meresht et al.'s work, high frequency data (0.1s intervals) from a research engine is collected with input data including the speed, torque, fuel flow, air flow, and exhaust gas temperature all being collected. Their work shows the achievable output quality of statistics-based models such as Random Forests in managing nonlinear relationships and uncertainties inherent in combustion data if input data of high enough quality is collected.

Furthermore, Gomaa et al. (2024) applied multiple ML methods, including Artificial Neural Networks (ANN) and Gradient Boosting, for predicting viscosities of natural gas mixtures under varying conditions [71]. Their study revealed the superior predictive performance of ANN, achieving an R^2 of 0.99, indicating the benefit of neural network models over more statistics-based models for the case of highly complex problems.

Of notable influence on the work performed in this paper, Saez-Perez et al. (2023) introduced a novel transformer-based architecture (CO2ViT) tailored explicitly for CO2 emission prediction in autonomous vehicles [72]. Compared against traditional Long Short-Term Memory (LSTM) networks, CO2ViT significantly improved prediction speed (71.14% faster) and accuracy (achieving an R^2 of 0.9898). The customization techniques for the model and in depth understanding of the transformer architecture helped to guide development of the model employed in this work, which is also an encoder-only transformer-based architecture. This study emphasizes the potential of transformer architectures, originally developed for natural language processing, to effectively capture temporal and spatial dependencies in vehicular emissions data, setting a precedent for their application to stationary combustion systems explored in this thesis.

The studies reviewed in this chapter collectively validate the evolving role of machine learning, particularly deep learning and transformer-based models, as a promising and effective tool for capturing the complex, nonlinear relationships present in combustion phenomena. While traditional statistical models have demonstrated adequacy under ideal conditions, many rely on high-frequency, laboratory-grade sensing networks not commonly found in the field. As discussed, real-world industrial applications, such as those targeted in this thesis, often involve noisier, lower-frequency data streams, and necessitate solutions that can operate reliably within these constraints. In the context of combustion modeling, the literature highlights the usefulness of simplified two-zone models for spark-ignition engines when balancing predictive accuracy against computational feasibility. Though more detailed CFD models can capture fine-scale mixing and flame propagation effects, the computational demands and practical limitations of these approaches make them less suitable for rapid deployment in legacy industrial assets. The reviewed studies show that two-zone modeling, when carefully calibrated, offers a strong foundation for emissions prediction—particularly when augmented by modern AI methods. Building upon this foundation, this thesis applies a transformer-encoder structure to develop virtual methane sensors trained on low-frequency experimental data from aging industrial natural gas engines. By leveraging the strengths of both combustion modeling and modern machine learning, this work stands at the convergence of decades of combustion science and the emergence of AI-assisted diagnostics. In the chapters that follow, the approach, model development, and validation strategies are discussed, leading toward the goal of producing scalable, real-time emissions monitoring solutions suitable for deployment across the existing natural gas compression fleet.

CHAPTER 3: EXPERIMENTAL SETUP AND ALGORITHM DEVELOPMENT

The necessary equipment, software, and models needed to accomplish the goals of this project will be presented – and the reasoning behind their selection, customization, and development will be discussed. Figure 7 displays the workflow of this project that was followed to accomplish the goal of the development of accurate virtual sensors. The following sections will contain all of the information used to accomplish each step of the workflow.

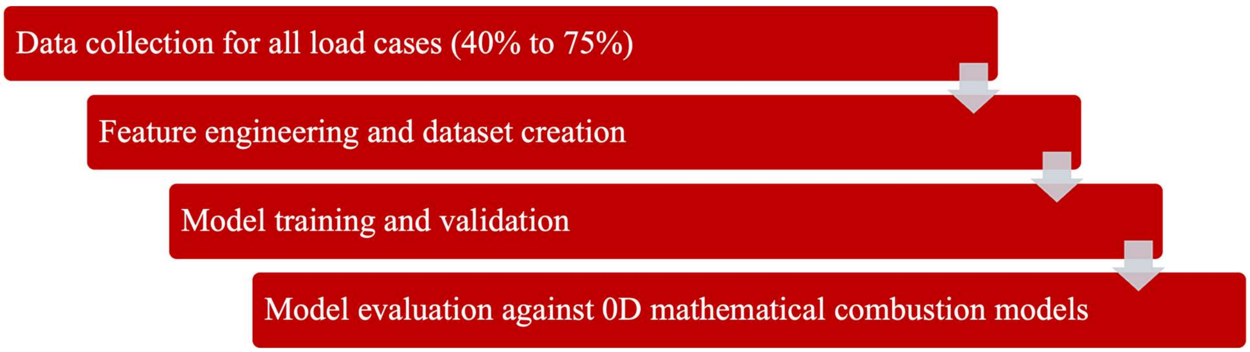


Figure 7: Project Workflow

SECTION 3.1: EXPERIMENTAL SETUP

SECTION 3.1.1: LABORATORY EQUIPMENT

The experimental setup constructed for the achievement of the goals of this thesis includes two industrial natural gas engines, gas analysis equipment, and a range of sensors equipped on the industrial engines that were selected to provide insight into the most important parameters for the combustion process. The laboratory engines include an AJAX DPC-81, and AJAX DPC-105.

Table 6: List of Equipment Used

MODEL	DESCRIPTION
AJAX DPC-81	Two-stroke, single piston, 81 bhp, integral NG compressor
AJAX DPC-105	Two-stroke, single piston, 105 bhp, integral NG compressor
Windrock 6400 Analyzer	Industrial analyzer capable of reading in-cylinder pressure, vibration, and temperature data and calculating performance parameters [73]

MKS FTIR Gas Analyzer (MultiGas 2030 CEM)	Uses Fourier Transform Infrared Spectroscopy to analyze gas composition [74]
DIA-VAC Model # R221-FP-AA1-CJM6	Gas sampling pump used to pull the exhaust emission stream to the MKS through the heated and insulated tubing
WAGO PLC System	The programmable logic controller is used to receive and export sensor data and interact with adjustable electronic components such as the bypass valve
Omega PX359	Pressure transducers used to read compressor-side operating pressures on the laboratory engines
Fox Thermal FT1	Used to measure the volumetric flow rate of the air
Emerson Micro Motion F025S	Used to measure the mass flow rate of the fuel
Dwyer Omega NB1-ICIN-INDUST-TC	Temperature sensor to measure exhaust gas temperatures
Red Lion MP2020	Magnetic pickup sensor to read engine RPM
Bosch Engineering Amperometric NOx Sensor - EGS-NX	Amperometric NOx sensor used to measure nitric oxides as well as oxygen % directly in the exhaust stream

To display the locations and uses of the sensors mentioned in Table 6, the following figure will show a high-level schematic of the sensing network as well as the experimental natural gas loop used to simulate engine load and performance at the Sustainable Energy and Carbon Management Center.

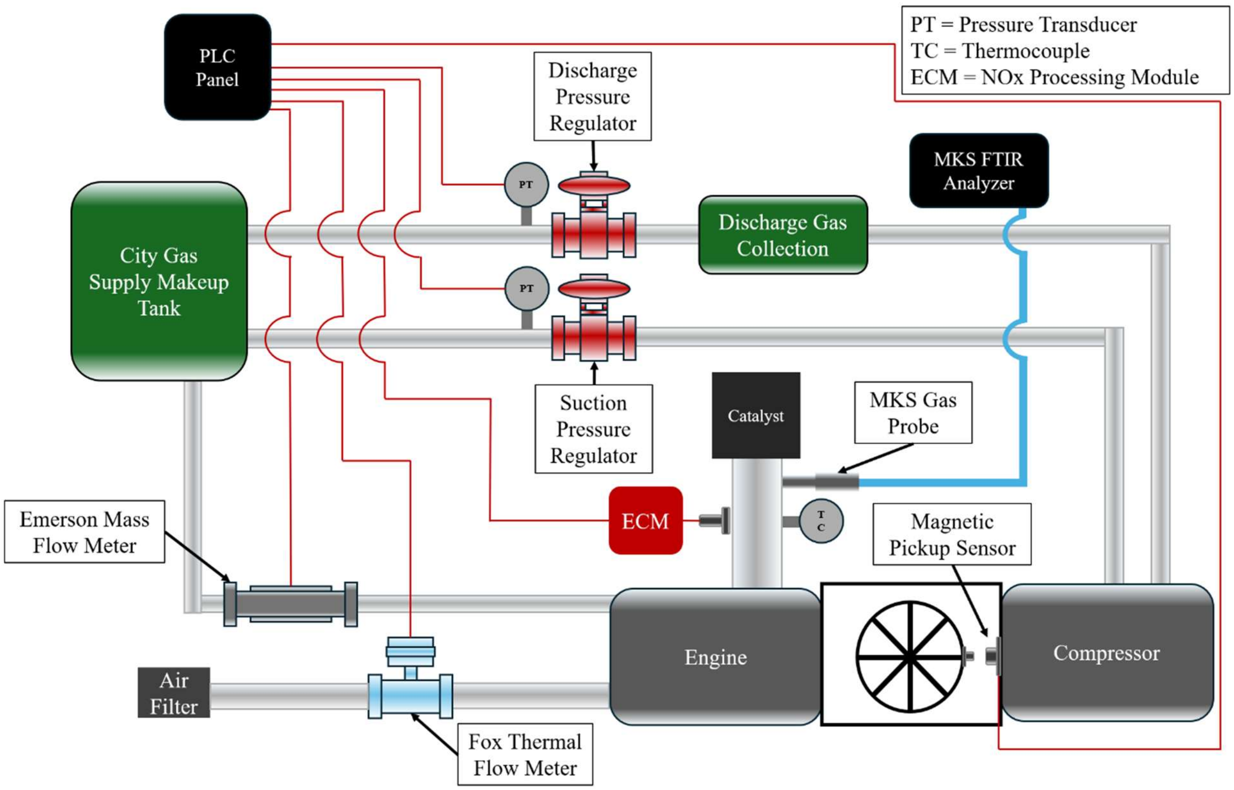


Figure 8: Data Collection Setup

Not included in the depiction of Figure 8 is the bypass valve location for the AJAX engines, discussed in Chapter 1 Section 3.2. Both AJAX models have the retrofitted electronic bypass valves placed between the stock reed valves and the entry location on the top side of the cylinder. The following figures will provide a look at the equipment setup depicted and presented in Table 6 and Figure 8.

Figure 9: AJAX DPC-81 Suction and Discharge Setup

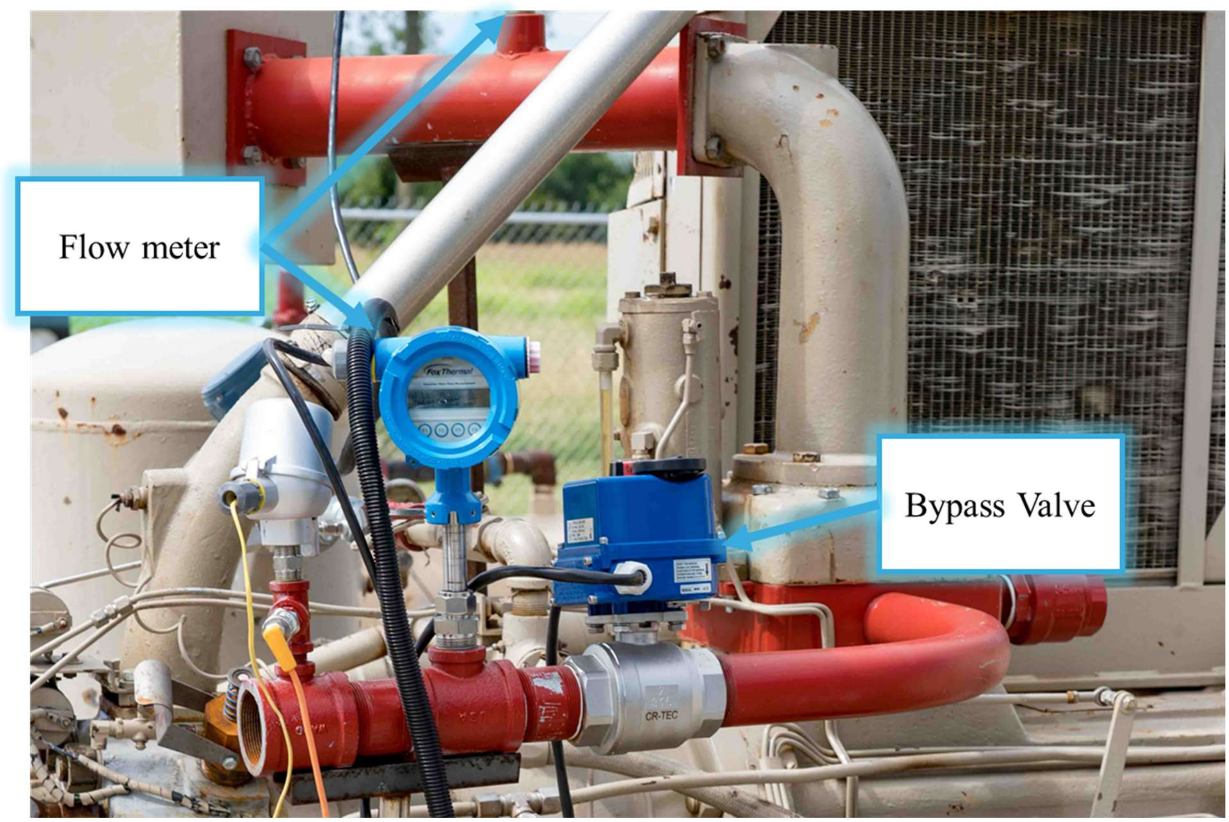


Figure 10: AJAX DPC-81 Bypass Valve and Mass Flow Sensor

Figure 11: Temperature, NOx/O2, and MKS Sampling Line on the AJAX Engines

The same suction and discharge loop is used for the AJAX DPC-81 and DPC-105 (which replaced the DPC-81 in its position in Figure 9 after data collection was concluded on the DPC-81). This loop is depicted in both Figure 8 and Figure 9. The air intake flow measurement and bypass valve setup is the same from the AJAX DPC-81 to the AJAX DPC-105 (Figure 10). The NOx/O2 sensing, exhaust temperature sensing, and MKS gas sampling line is depicted for the engines in Figure 11 for the AJAX models.

SECTION 3.1.2: LABORATORY ENGINES OVERVIEW

Now, a description and important data for each of the laboratory engines will be given to provide the reader with some understanding of general operation and the differences in parameters that affect the combustion process. The same factors discussed in Chapter 2 Section 1 that affect which physical phenomena are vital for proper modeling and must be considered when evaluating the best approaches for choosing the sensor networks for data collection and model building.

Table 7: Laboratory Engine Relevant Details

AJAX DPC-81					
Bore x Stroke (in^2)	*Compression Ratio*	*Fuel & Air Mixing Method*	*Stroke*	*Ignition Process*	*# of Cylinders*
10.5 x 12	**7 - 8**	**Scavenging & mechanical Fuel injection**	**Two-Stroke**	**Spark Ignition**	**1**
AJAX DPC-105					

Bore x Stroke (in^2)	*Compression Ratio*	*Fuel & Air Mixing Method*	*Stroke*	*Ignition Process*	*# of Cylinders*
12 x 14	**7 – 8**	**Scavenging & mechanical Fuel injection**	**Two-Stroke**	**Spark Ignition**	**1**

Table 8: Laboratory Engine Rated Operating Conditions

AJAX DPC-81			
Rated HP	*RPM*	*Suction Pressure*	*Discharge Pressure*
81	**475**	**45**	**450**
AJAX DPC-105			
Rated HP	*RPM*	*Suction Pressure*	*Discharge Pressure*
105	**425**	**3**	**40**

From Table 7, a vital characteristic of the AJAX compressors is that the power cycle is two-stroke. This means that for every revolution of the crank shaft, there is one power stroke as opposed to the two revolutions necessary to achieve a power stroke in a four-stroke engine. The two-stroke spark ignition process compresses a fresh fuel and air charge, the spark ignites some few degrees prior to TDC (11 degrees prior in the case of both the AJAX DPC-81 and DPC-105), and in the process of the combustion event, expansion, and retreating piston, the exhaust and then intake air ports are uncovered in that sequence to allow exhaust emissions to escape and fresh air to be pulled in. Due to the location of these inlet and outlet fluid velocities and the transfer of heat, pressure, and energy, a flow is induced that encourages more fresh air to be pulled in and more exhaust gases to be pushed out. After the piston reaches BDC and begins its acceleration towards TDC again, the air intake port and shortly thereafter the exhaust port is covered again, and the mixture is compressed. In the case of the AJAX models, the fuel injection timing is mechanically governed via the camshaft. At 193 degrees (refer to Figure 12 for timing events for both the DPC-81 and DPC-105) the camshaft begins to actuate a plunger within the injector system that increases hydraulic pressure and opens the poppet valve mechanism – allowing constant pressure (somewhere between 9-15 psig) natural gas to flow into the combustion chamber. In Figure 13, the AJAX scavenging process is depicted with the internal geometry specific to these engines in display.

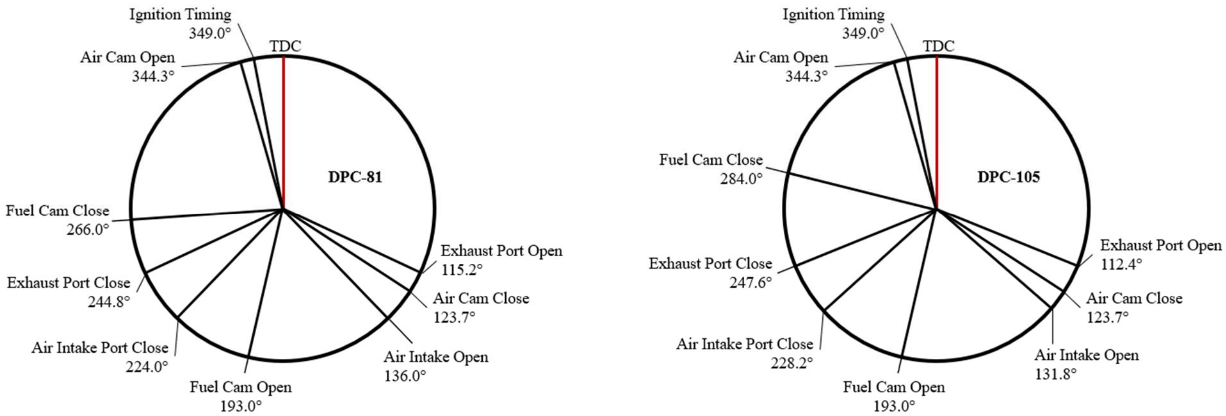


Figure 12: AJAX DPC-81 and DPC-105 Timing Events

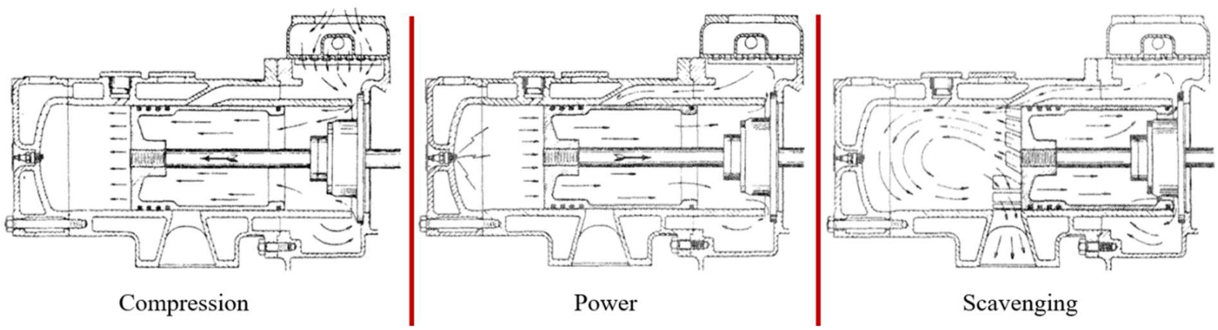


Figure 13: Two-stoke Combustion Process for AJAX Engines [75]

The differences between the two engines are that the DPC105 has a larger swept volume, longer crank radius, larger air intake and higher-pressure fuel injection, and timing that allows for more fuel and air to enter the combustion chamber.

SECTION 3.1.3: EVALUATION OF NOx SENSOR RELIABILITY & DEGRADATION

As referenced in the section title, an experimental study is performed to validate the use of the amperometric NOx sensors in Table 6 for training the machine learning algorithms.

Compared to other molecules that are desirable to measure for reason of government regulation or engine performance, NOx molecules are able to be measured quite easily and reliably. Due to the small family of molecules of similar electronic properties – mainly nitrogen oxide and nitrogen dioxide as opposed to something like the large string of hydrocarbon molecules that might exist between methane and carbon dioxide in the process of oxidation – an electrochemical sensor may be used to measure these molecules. As mentioned in Section 3.2.1 below, NOx is an excellent indicator of peak combustion temperature. This is a vital piece of information for training an AI model for combustion emission prediction, as peak temperatures, along with the peak combustion pressure, may be used to determine the peak activation energies achieved. Modern NOx amperometric electrochemical sensors can provide data at nearly any sampling frequency desired with a high degree of accuracy to most PPM counts.

Given that these sensors are one of the main input sensors for model training in the work of this thesis and are potentially subject to degradation in natural gas exhaust environments, it is desired to test this degradation. If it is excessive, the performance of the sensors will change over the course of data collection used as input training data, potentially degrading model performance upon implementation into field engines. This is the motivation behind testing the degradation.

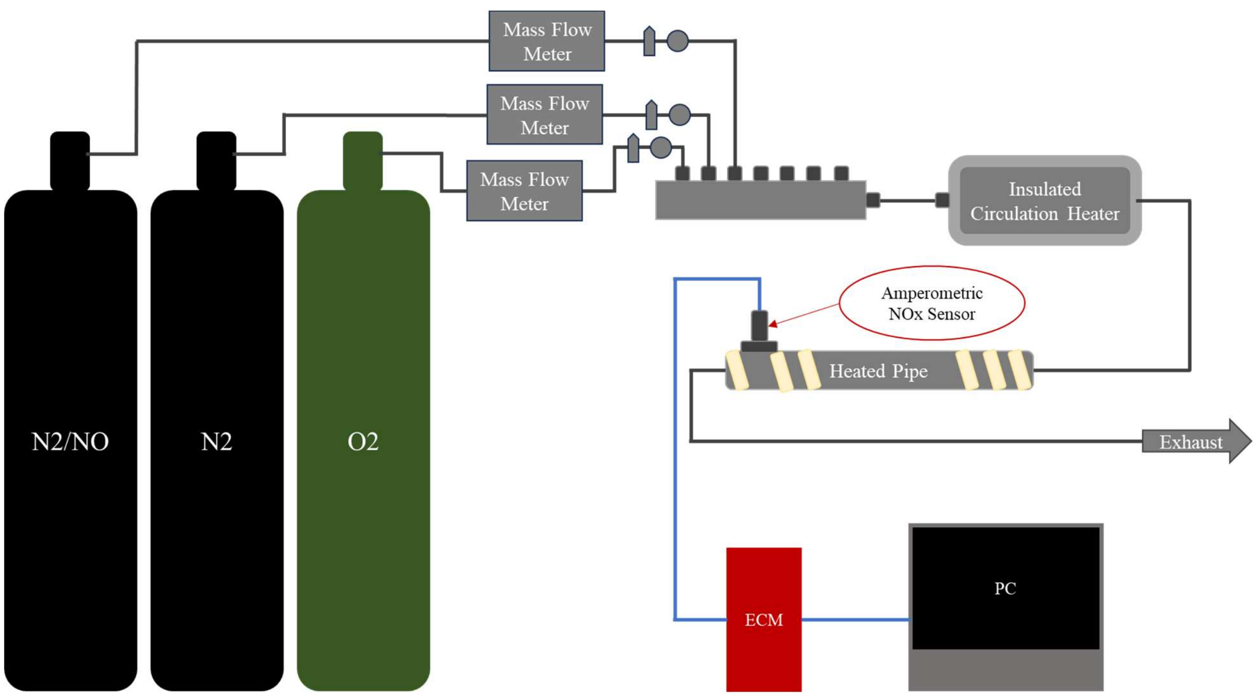


Figure 14: Experimental Testing Apparatus for NOx Sensor Evaluation

Table 9: Experimental NOx Sensor Testing Equipment

Equipment / Software	Notes / Conditions
Compressed 251 Oxygen Cylinder	Standard oxygen cylinder.
Compressed Nitrogen Cylinder	Standard nitrogen cylinder
Compressed Nitrogen / NO Cylinder	Nitrogen, nitric oxide mixture 2536 PPM concentration of nitric oxide
Regulator #1 (oxygen) (Smith Model # 30-20-540)	Regulates the output pressure from the gas cylinder

Regulator #2 (nitrogen) (Smith Model # CGA-580)	Regulates the output pressure from the gas cylinder
Regulator #3 (nitrogen / NO) (Airgas Model # Y11E444D660-A)	Regulates the output pressure from the gas cylinder
Omega Mass Flow Sensors (Model # FMA-1607A)	Utilized to measure the mass flow rate of the three gas cylinder outputs with a ±0.001 g/min uncertainty
CAS Circulating Heat Pump (Model BX8L4M999-0010)	Heats the gas stream mixture to 250°C
Resistance Heating Tape / Wrap	Applied to the experimental testing apparatus piping to keep temperatures elevated and consistent
Twidec Temperature Measurement (Model # MV100)	Utilized to monitor the temperature of the flow that the NO_x/O_2 sensor experiences

Using the equipment referenced in Figure 14 and Table 9, the sensors are tested against known PPM counts of nitrogen oxide in known flow rates of oxygen and nitrogen to determine accuracy. This experimental setup was used to test the sensors between 25-hour stints of use in the AJAX DPC-81 engine. This procedure is outlined in Figure 15.

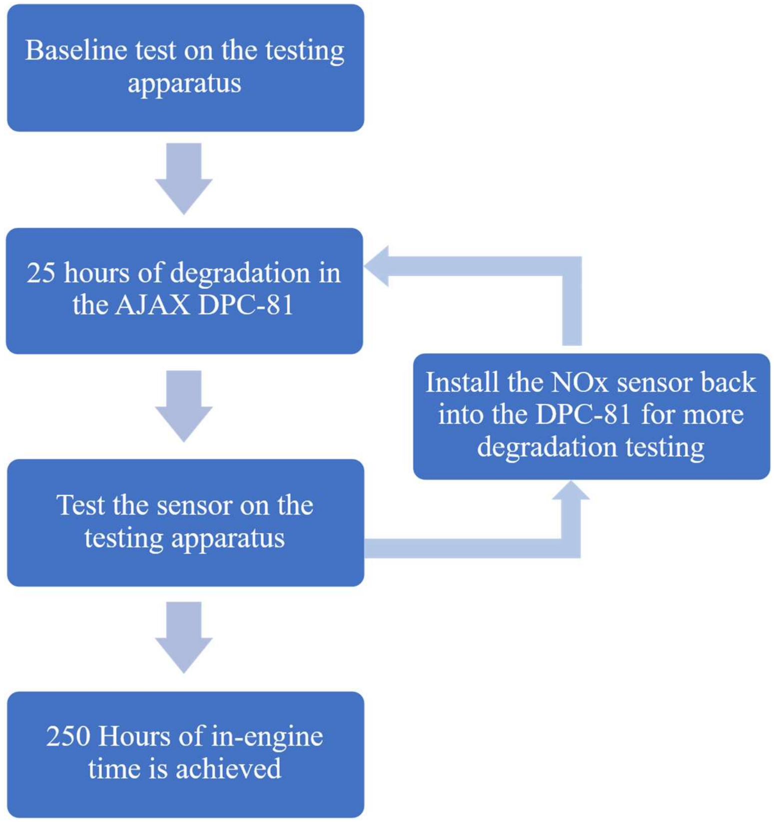
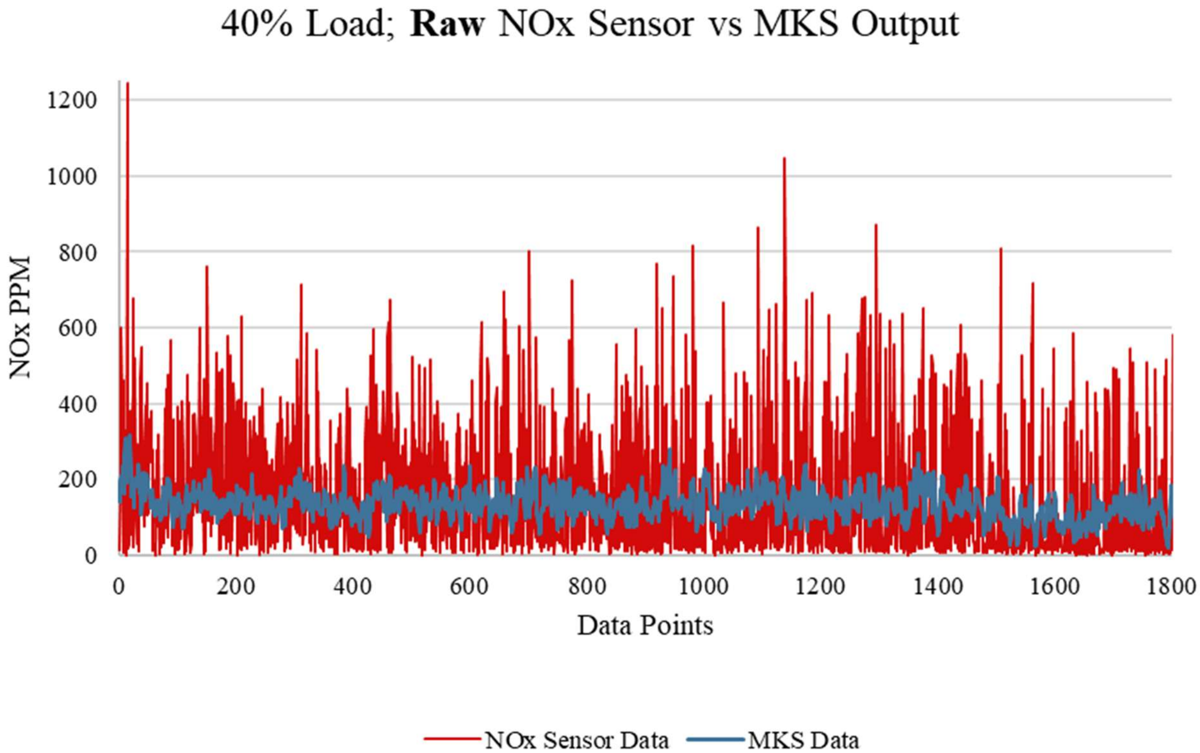


Figure 15: NOx Sensor Evaluation Procedure



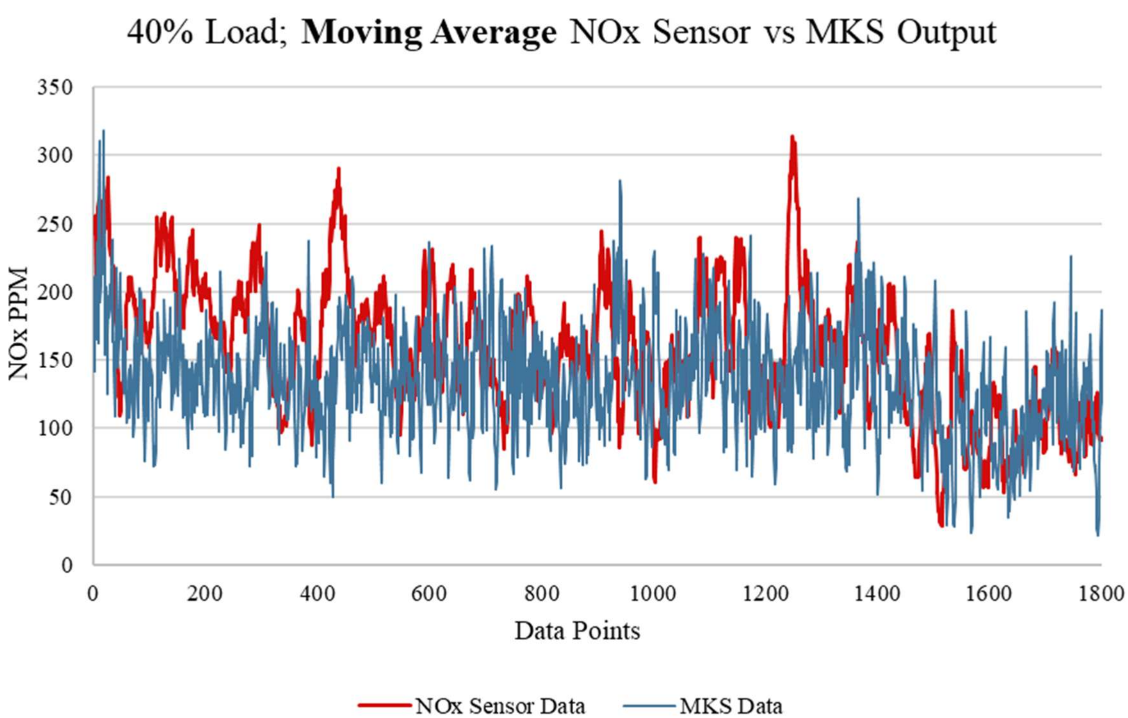


Figure 16: Raw and Moving Average NOx Sensor Data Compared to MKS NOx Data with a New Sensor at 40% Engine Load

Table 10: Averages for comparison; Processed NOx sensor data against MKS 2030 data

Load (%)	100s Moving Average NO_x Sensor Data (PPM)	MKS Data (PPM)
40	159	136
60	66	47
70	91	71

From Figure 16 and Table 10, a new sensor's performance compared to the highly accurate MKS gas analyzer may be observed. The nature of the data collection between the NOx sensor and the MKS are different, as will be discussed again in the feature engineering section of this chapter, Section 3.2.2. The PLC system provides instantaneous 'grabs' from all sensors inputs every 5 seconds, while the MKS device is processing its own data at a 5 second sampling rate – providing an average over these 5 seconds for all emission outputs including NOx. Due to the nature of the PLC data collection system, there is a larger amount of volatility shown in the raw data at the top of Figure 16. To make this more comparable to the MKS as well as more suitable for training a machine learning algorithm, a moving average is applied to smooth the data and show trends more closely related to reality.

From this initial feasibility testing (Figure 16), the sensors were found to closely follow the trends of the MKS device outputs for NOx measurement, with a near constant average difference of 20 PPM over a variety of loads on the AJAX compressor. Over time, the sensors exhibited some detectable degradation: the standard deviation of their experimental testing apparatus readings increased by 138% on average, and the spread of NOx PPM outputs at a given current increased from around ±1 PPM out of the box to around ±2.4 for the worst case after 250 hours of operation (Figure 17 and 18).

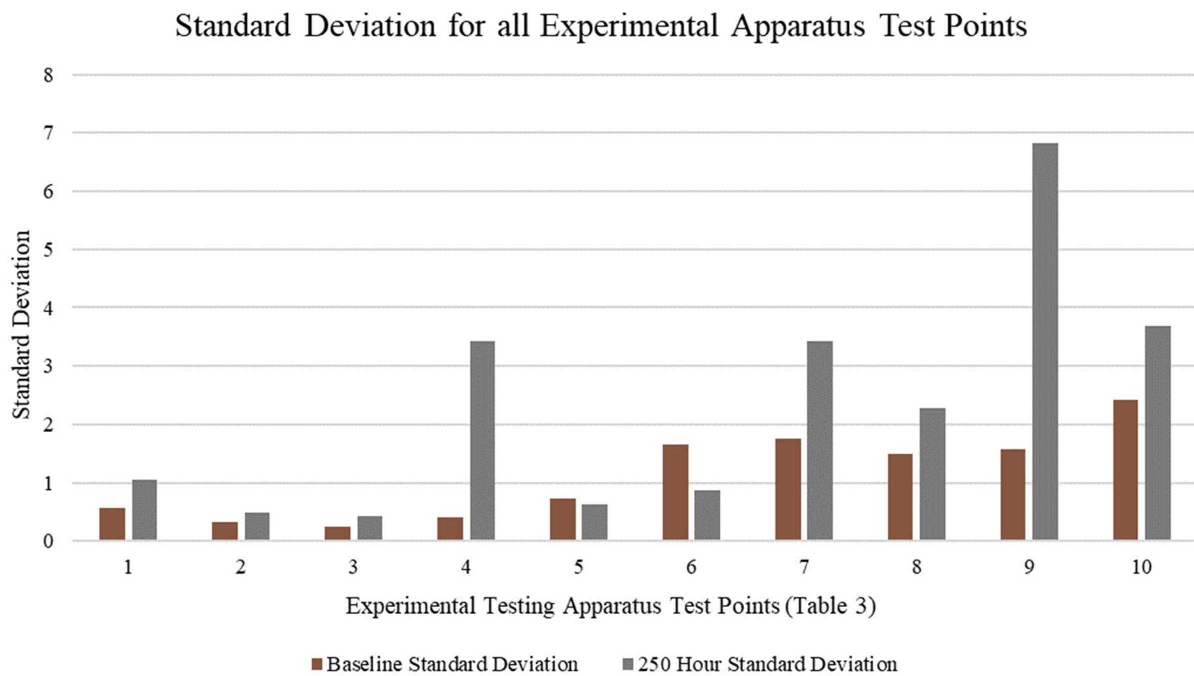


Figure 17: Standard Deviation Analysis Results; Averaged for All Sensors at Baseline and at 250 Operational Hours

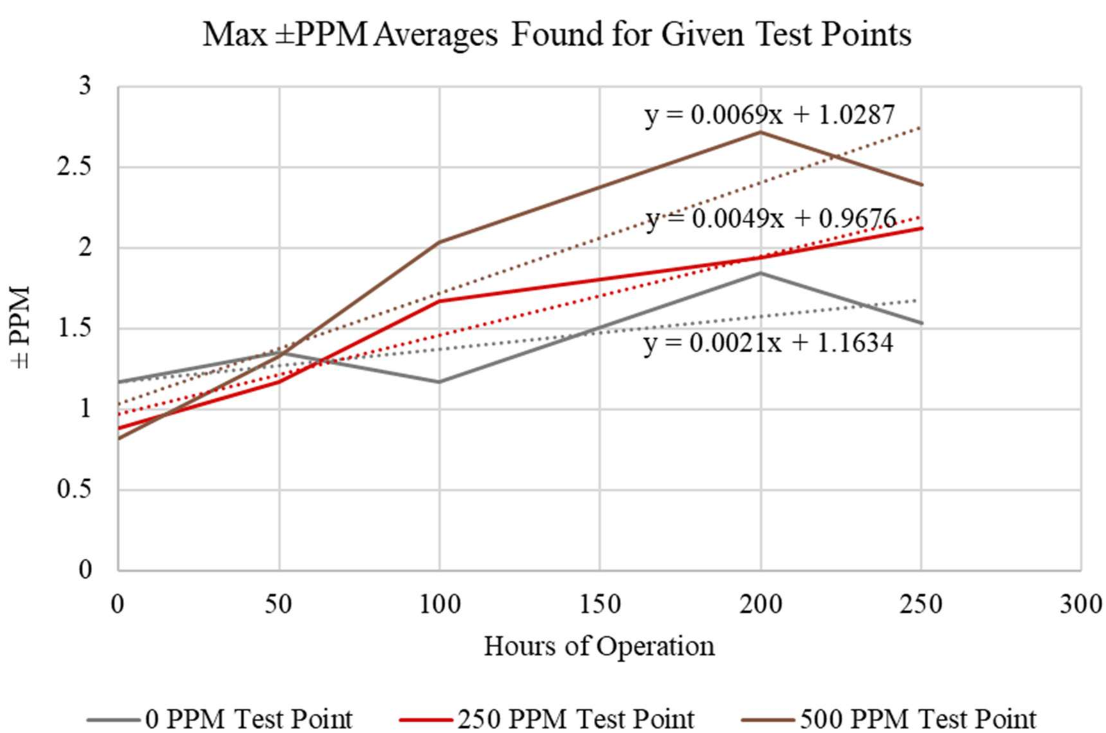


Figure 18: Averaged for all three sensors: Maximum ±PPM readings found at each NO_x test point in the experimental testing apparatus

These metrics indicate increases in spike fault occurrence and scale, particularly at higher NOx concentrations, though the developments were relatively small. The 138% increase in standard deviation correlates to a possibility for outliers that deviate from the mean by 7 PPM in the worst-case scenario on the experimental testing apparatus. Despite this degradation, the sensors maintained a level of performance that is sufficient for short-term monitoring in conjunction with data smoothing techniques. The total amount of hours spent collecting data on all engines for the work of this study is somewhere around 90 hours split between two separate engines with their own amperometric NOx sensors. Given this data analysis along with this timeframe being well under 250 hours for any particular sensor being used, there is confidence in the sensors' ability to maintain performance suitable for training a machine learning model that relies on these sensors' data as input.

Moving forward now from the laboratory equipment, a discussion of the input data, ML model development, and validation approaches will be had.

SECTION 3.2: MODEL DEVELOPMENT

SECTION 3.2.1: INPUT DATA SELECTION AND REASONING

The main parameters chosen (and the sensors chosen to collect these parameters) to evaluate the combustion process are listed in Chapter 3 Section 1. These parameters include the suction pressure, discharge pressure, RPM, fuel mass flow, air mass flow, exhaust temperature, ambient temperature, NOx emissions in the exhaust stream, the oxygen percent in the exhaust stream, and bypass valve positioning. The main reasoning for this selection is the following: It can be costly in time, money, and labor to directly measure the core parameters affecting combustion (namely continuous combustion pressure, temperature, and heat transfer into and out of the system). These parameters require a very high sampling frequency to make use of the information which becomes another factor in cost and computing to handle. Therefore, it is highly desirable to find low-cost sensors that can implicitly inform about these difficult to measure parameters in a sufficient manner to be able to evaluate the combustion to a degree of accuracy that is satisfactory.

As discussed in Chapter 2 Section 1, to quantify the combustion in the engines with minimal error in approximation it is vital to know the chemical composition and quantity of all present molecules within the combustion chamber, the pressure, temperature, energy entering or leaving the system via heat transfer or spark firing, and the effects of fluid motion on the mixing of the fuel and air and also on the heat transfer. To produce a model that can predict or estimate the combustion and these incredibly complex physics accurately, the input variables need to have some correlation or connection to these parameters. In this way, the model can be trained to recognize relationships between the input parameters that indicate the peak combustion temperatures, peak in-cylinder pressures, chemical composition, and heat transfer into and out of the system. Although the model is not explicitly trained on parameters such as peak combustion temperature and chemical composition, these conditions are inherently linked to the chosen input variables. Consequently, a sufficiently sophisticated model can leverage these indirect relationships to accurately estimate emissions.

Table 11: Input Variable Relations to Vital Combustion Parameters

Input Variable	Description of Relation to Necessary Combustion Parameters
Revolutions Per Minute (RPM)	The RPM can be used to calculate the piston speed, which implicitly informs how much volume is being displaced over a unit of time, and if the resistance (suction and discharge pressures) is known, it provides an idea of the load on the engine. The load (power output) is a measurement of power required and gives implicit knowledge of how much fuel is necessary to achieve this power output.
Suction Pressure	As stated for the RPM, if the resistance to the pistons movement is known, the load may be calculated. The suction pressure is what is pulling the fuel in that is used for fuel and to be compressed, so it has an indirect impact on the achievable load through regulating how much discharge pressure is achievable. This measurement of required power may provide implicit knowledge of energy (fuel) requirements and therefore how much heat is being released.
Discharge Pressure	The discharge pressure has direct impact on the resistance the compression piston, and therefore the combustion-side piston, is experiencing. This resistance must be met to keep the RPM constant, which requires a specific amount of fuel which dictates the heat release, combustion pressure, and therefore the chemical kinetics.
Exhaust Temperature	The exhaust temperature provides implicit information regarding how much heat is being lost from the system, what the peak combustion temperatures are, and what peak combustion pressures are. This is directly related to completeness of combustion as well; low exhaust temperatures might indicate incomplete combustion as less energy is released.

Ambient Temperature	The ambient temperature provides further information on heat transfer out of the system, impacting all of the downstream variables: peak combustion temperature, peak combustion pressure, and total fuel input. The fuel (coming in from a city line) is around ambient temperature in the case of the laboratory engine, and the air is being pulled directly from the atmosphere. This alters the density of the fuel and the air, which causes fluctuation in the quantity of molecules and therefore ratios the engine is receiving as the ambient temperature fluctuates.
NOx PPM Exhaust Count	The nitric oxides present in the exhaust are excellent indicators of peak combustion temperatures achieved. This is a connection to peak combustion pressures achieved, heat transfer out of the system, and total heat release.
O2 Exhaust %	The oxygen percent in the exhaust is an indicator of how much excess (or lack of excess) oxygen was present for the combustion process. This is used to understand how complete combustion is or isn't, which is dependent upon the amount of oxygen available for reaction as well as energy that is stolen from reaction by excess oxygen (and also excess nitrogen from the air composition) if it is present in too high of a quantity.
Bypass Valve Opening %	Bypass valve opening provides connection to air to fuel ratio, as it directly impacts how much air is making it into the combustion chamber. This, again, provides knowledge as to what peak values are reached for temperature and pressures in the combustion chamber.

Fuel and air flow sensors are intentionally omitted from Table 11 since their values should be implicitly inferred from other measured variables. Successfully achieving this implicit inference allows for cost savings in a retrofit scenario by reducing sensor count. In the laboratory setup, however, these sensors remain crucial for validating the mathematical model results discussed in Chapter 5. The continuous comparison of mathematical model outputs against measured values helps establish confidence in the implicit relationships utilized, enhancing understanding of each input variable's influence on combustion outcomes.

SECTION 3.2.2: FEATURE ENGINEERING

A vital topic of consideration for any engineer hoping to achieve good results from a ML algorithm is feature engineering. The data must be tailored to the specific goals of the engineer, and to the unique circumstances of the data collection process. For the specific case of this project, the most important considerations can be summarized as the following:

Table 12: Important Data Considerations Prior to Model Development

Point of Interest	Approach for Solution
The input and output data sources are being received by two different systems. The PLC handles all input sensors and the corresponding variables, and the MKS gas analyzer data output is being processed by its own software locally on the laboratory PC. This fact makes it rather important to consider the time stamps between the two sources and to ensure that they line up when creating new, combined datasets.	Take care in insuring that both systems are connected to the internet and running on the same time – pay careful attention when patching the two datasets together in post-processing.
Another point towards the alignment of timestamps: The gas sampling system includes some 30 feet of transport line and a 1/8 hp vacuum pump to pull the fluid to the analyzer. This makes it important to consider the delay it takes for the exhaust gas to reach the MKS analyzer such that the input data and output exhaust emission data may be well aligned.	Determine the delay and then shift the MKS data accordingly such that the delay is accounted for. To determine the delay, any change in engine parameters on a given day of testing is timed to determine how long a change in emissions values takes to be seen on the MKS software output. All recorded time durations are averaged for a given day of testing, and the data is shifted this amount of time to match up with the PLC data output.
The nature of the sampling is inherently different between the MKS analyzer and the PLC system. While both are set to 5 second sampling rates, the PLC is grabbing the instantaneous sensor outputs at those 5 second intervals, while the MKS is providing 5 second averages. In reference to this, the nitric oxides in the exhaust stream may vary at a particular point in the exhaust pipe quite significantly from moment to moment. Due to the instantaneous "grabs", the PLC system outputs quite noisy NOx and oxygen % values that are obtained from the amperometric NOx sensor.	A moving average is applied to the amperometric NOx sensor outputs to reduce noise and increase the correlation to the emissions outputs as well as the other input features like exhaust temperature.

These are the dilemmas associated with the specific equipment and methods used in the laboratory and are a good representation of why feature engineering is such a vital component of building a model that is accurate and capable of accomplishing a given task successfully. If these simple points were neglected, results of acceptable accuracy would be hard to come by.

Outside of these scenario specific considerations, there are other general practices when examining the input dataset for a given machine learning problem. A correlation analysis between all features is a good place to start [62], providing information about how all variables are related to one another. The correlation method used for the datasets in this work is referred to as the Pearson method [76], which is used to find linear correlation coefficients between each pair of variables. Good practice is to search your variable correlation coefficients for values very close to 1, as a high correlation may mean that both variables provide very similar information to your model, which acts as redundancy. Redundancy does a number of things that should be avoided, including increasing dataset size without benefit – which increases the required computing to load and train the model on the data, and an increasing model complexity due to the additional dimensionality which can cause further overfitting to training data and poor performance on validation data.

SECTION 3.2.3: DEVIATIONS IN THE CUSTOM MODEL FROM THE ORIGINAL

This section serves to lay out the thinking behind the model development and answer the questions of how and why certain things were done the way that they are given the problem at hand and the nature of the data. As mentioned in Chapter 2 Section 2.2.3, a transformer-based architecture is customized for the engine datasets due to its ability to pick up on complex nonlinear relationships and to understand the impact previous timeframes can have on a current output. Originally, the transformer architecture was constructed for natural language processing tasks, which follow different procedures for loading and preprocessing the input data than what the data of this thesis requires [67]. In this original construction, the words from the input text file must be tokenized into vector space, embedded into a set of linear equations (layer of neurons), and then finally positionally encoded prior to introduction to the transformer encoder. These encoded outputs are then supplied to the decoder's multi-head attention mechanism, which is attending to the text that is shifted right from what the encoder is attending to. Due to this shift and the decoder's structure, the resulting output is a probability for which words may come next in the sequence. In our case, the input data obtained from the laboratory engines is a consistently structured dataset of a consistent number of columns of numerical data. Since the data is already in numerical format, there is no need to tokenize the data. Since each line of data represents a time step of identically dimensioned data (the data is not sentences of varying length), the positional encoding can be applied directly during the data loading process. The positional encoding functions chosen to apply to our data are as follows:

$$PE_{(pos,\ 2i)} = \sin\left(\frac{pos}{10000 * \frac{2i}{input_{size}}}\right) \tag{12}$$

$$PE_{(pos,\ 2i+1)} = cos\ (\frac{pos}{10000 * \frac{2i}{input_{size}}}) \tag{13}$$

Where pos represents the position or index within the data sequence, and $\mathrm{input\_size}$ represents the number of input columns or features. The sine wave is applied to the even positional indices, and the cosine wave is applied to the odd ones. These positional vectors are then added directly to the input data. To use this newly positioned data in the transformer architecture, which expects sequences of inputs rather than individual timesteps, a custom dataset class was constructed to implement a sliding window approach that constructs overlapping sequences of data that will be provided to the model as simultaneous inputs with the temporal position added thanks to the positional encoding such that temporal relationships may be learned. After the process of positionally encoding and sequencing the data, the inputs are provided to the embedding linear layer. This layer applies a simple linear transformation to the input data using the following equation:

$$x_{linear} = xW + b \tag{14}$$

For a given neuron, $\mathrm{x_{linear}}$ represents the resulting output, W represents the weight, and b represents the bias. This is the first part of the model in which the gradient calculations are required and is therefore a trainable part of the model in which the weights and biases will be adjusted during backpropagation. After the embedding is complete, the data is fed into the multi-head self-attention mechanism. This mechanism helps the model decide which parts of the input are most relevant to focus on when making predictions. Scaled dot-product self-attention is the particular form of attention mechanism used in most transformer architectures and is what is employed for this model is well. The embedded inputs are split between three separate matrices: the query matrix, the key matrix, and the value matrix. Scaled dot-product self-attention may be defined with the following equation:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{15}$$

$$softmax(z_i) = exp(z_i) \,/ \sum_j exp(z_j) \quad (16)$$

In Eq. 15, Q, K, V represents the queries, keys, and value matrices. The term d_k refers to the dimension size of both the queries and keys. In a sense, queries may be considered questions we might ask about the data, and the keys may be considered the connections all inputs have to one another. First, the dot product is performed between the query and key matrices (the exponent T on the key matrix represents the necessary transposition to make a dot product possible between the two matrices) and then scaled by $\frac{1}{\sqrt{d_k}}$. The dot product output is then run through a SoftMax function (Eq. 16), transforming the dot product outputs into a set of fractional weights ranging from 0 to 1 that sum to 1. Once this value is known, the weights, or scores, are multiplied by the values from the value matrix to place importance on certain values (multiply by something close to 1) and drown out other values (multiplying a small number closer to 0). This builds the model's understanding of which input variables (at multiple time stamps) have the biggest impact on the output that the model is attempting to predict. However, not just one attention mechanism is utilized. The model employs multi-head attention like that represented in 'Attention is all you Need.' This means that the query, key, and value matrices are split up to allow for simultaneous, parallel computation of the attention values. Splitting the matrices allows the model to learn different aspects of the overall query matrix's impact on the model's output production. In other words, multi-headed attention allows for both parallel computation as well as better ability to pick up on varying aspects of input data impact on the output values [67], [77].

After exiting the encoder of the Transformer, the data is passed through a small multilayer perceptron (MLP). The MLP consists of the input layer of data that comes from the encoder, one hidden layer of neurons, and one output layer that consists of a single neuron which is the model's final output at a given time step. Between each of these layers are ReLU activation functions. This architecture is depicted as in Figure 19.

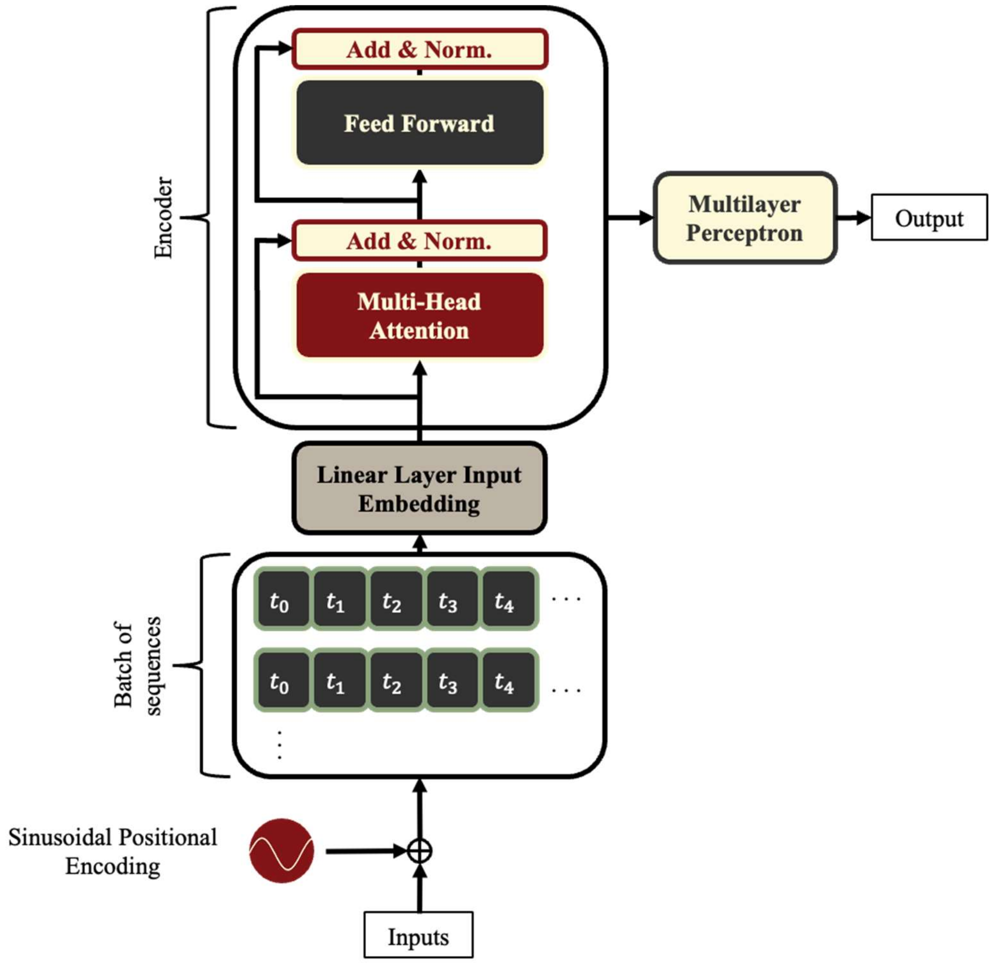


Figure 19: Final Transformer-Encoder Only Architecture

The loss function chosen for this model is Huber Loss, which may be defined with the following equations [78]:

$$\ell(x, y) = L = \{l1, \ldots, lN\}T \tag{17}$$

$$l_n = \begin{cases} 0.5 * (x_n - y_n)^2, & \text{if } |x_n - y_n| < \delta \\ \delta * (|x_n - y_n| - 0.5 * \delta), & \text{otherwise} \end{cases} \tag{18}$$

The reasoning for selection of Huber Loss is directly related to the discussion of feature engineering in Section 3.2.2. From examining Eq. 17 and 18, one who is familiar with mean squared error (MSE = $0.5 * (x_n - y_n)^2$) or mean absolute error (MAE = $|x_n - y_n|$) can see that Huber Loss is a combination of MSE and MAE loss functions, depending on the relationship between the difference between the residual and the δ (delta) values, where δ is a settable parameter to tailor for your particular dataset and architecture. If your difference in prediction vs ground truth (residual) is less than delta, then a MSE loss function is employed. This gives the advantage of the nonlinear response from MSE loss up until the residuals get large enough that it becomes beneficial to employ the linearity of the MAE loss function, as large outliers that cause large residuals will not drastically affect the loss calculation as they would with an MSE loss function. This kind of performance can be observed in Figure 20.

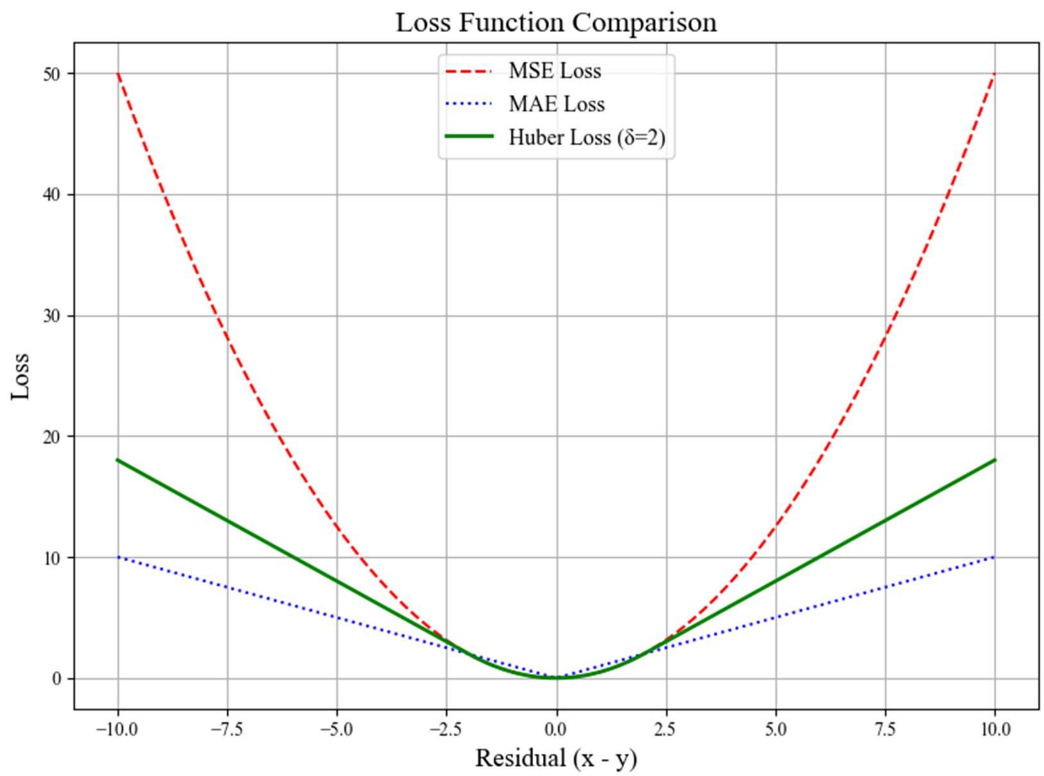


Figure 20: Example and Comparison of Huber Loss against MAE and MSE Loss Functions

The reason this loss function is valuable for our data is that there is some non-negligible noise in the MKS emission data, making it beneficial to have some resilience to outliers. In adjusting for the model performance and dataset values, delta is set to the median absolute deviation [79] of the data and multiplied by a factor greater than one (1.2 was most often used in model training).

To conclude and recap this discussion of the customization of the encoder-only transformer architecture, a brief walk-through of the input data's 'journey' so to speak (pictured in Figure 19) will be had:

1. The sinusoidal positional encoding values are added as additional features to each line of input data using Eq. 12 and 13. This allows the model to know which time stamps are which when learning the temporal relationships using the encoder and then making final output predictions with the final MLP.

2. Then, this new input data set is batched and sequenced – the batching to allow for utilization of GPU-parallelized computation for increased training speed, and the sequencing to break the data up into the desired timespans that one wants the encoder to learn temporal relationships between.
3. From here, the data is passed through a single layer of linear transformations to map the original input feature space into the model's internal dimensional space, preparing it for attention-based processing.
4. After the data is linearly transformed, it is introduced into the encoder layers (multiple layers makes for a more complex model capable of learning more complex relationships and parallel computation). The encoder utilizes the attention mechanism (Eq. 15 and 16), and a feed forward network to learn which time stamps have the biggest impact on the output in focus.
5. After the temporal relationships are learned, the data is passed through a final multi-layer perceptron to attempt to predict the final output based on these learned temporal relationships.

The hyperparameters were tested and tuned for both engines to achieve optimal performance with as small of model size as possible. The results of this testing will be included in Chapter 4, though the following table contains the starting value for all tunable hyperparameters in the developed model:

Table 13: Initial Model Hyperparameters

Hyperparameter Variable Name	Description	Value
d_model	The dimensionality of the input embeddings and all hidden layers in the transformer model. It defines the size of the vector used to represent each timestep and affects the model's representational capacity.	64
dim_feedforward	The size of the hidden layer in the transformer's feedforward network, typically larger than d_model. It determines the capacity of the internal MLP within each encoder block to capture complex patterns	64
mlp_layer_size	The size of hidden layers in custom multilayer perceptron (MLP) heads used for final regression or classification tasks. This controls how much abstraction the	128

	final layers can perform after the transformer.	
nheads	The number of attention heads in multi-head self-attention. Each head learns to focus on different subspaces of the input, enhancing the model's ability to capture varied temporal dependencies.	8
num_encoder_layers	The number of stacked transformer encoder blocks. A greater number of layers allows the model to learn deeper hierarchical representations but increases computational cost and risk of overfitting.	2
dropout_prob	The probability of randomly zeroing out elements during training to prevent overfitting. Applied in attention and feedforward layers, it regularizes the model.	0.1
seq_len	The length of the input sequence window processed by the model. It determines the temporal context available for each prediction and should capture sufficient past information for accurate forecasting.	12
batch_size	The number of sequences processed together during one forward/backward pass. Larger batches improve training stability and GPU utilization but require more memory.	128

These architectural adjustments and parameter selections were developed considering the unique attributes of the collected dataset structure and complexity between input and output relationships. Moving forward, the mathematical model used to evaluate the combustion and further evaluate the ML model performance will be discussed.

SECTION 3.3: THE EVALUATION AND COMPARISON TO MATHEMATICAL MODELS

SECTION 3.3.1: MODEL SELECTION AND CAPABILITIES

In order to evaluate the performance of the model in comparison to mathematical models, Chemkin's 0D SI-engine model is used for the AJAX DPC-81 and the AJAX DPC-105 [48]. In reference to Chapter 2 Section 1.1.1, this model is a two-zone, 0D mathematical solver that transfers the mass of an unburned zone to a burned zone as the time and crank angle position progresses through a power stroke using the Wiebe function. The model is capable of calculating the detailed reaction kinetics at the given energy in the system at each time stamp (crank angle position) using Berkley's GRI Mech 3.0 dataset for methane combustion [45].

SECTION 3.3.2: IMPORTANT CONSIDERATIONS AND JUSTIFICATIONS

This section will cover the necessary equations and theory that are specific to the Chemkin two-zone SI engine model that are vital for proper modeling of engine conditions to achieve the most optimal results. If further information is desired, most necessary equations and explanations of theory and calculation can be found in the Chemkin Theory Manual [46].

The first point of discussion is the method in which this particular model handles intake of fuel and air. As there is no inherent way to couple the Chemkin model with an outside model to handle the air and fuel intake without going to drastic measures, the normal user of this software will have to rely on the following formulation to regulate the intake process.

$$n_{total} = \frac{P_0 V_0}{R T_0} \tag{19}$$

$$\phi = \frac{\left(\frac{A}{F}\right)_{stoich}}{\frac{A}{F}} \tag{20}$$

$$m_{cylinder} = \sum_i n_i M_i \tag{21}$$

$$m_{burned}(\theta) = x_{burned}(\theta) * m_{cylinder} \tag{22}$$

$$m_{unburned}(\theta) = (1 - x_{burned}(\theta)) * m_{cylinder} \tag{23}$$

Eq. 19, in combination with Eq. 20, is used to determine the total number of moles in the system along with the ratio of moles of fuel to moles of air given the user inputs of the equivalence ratio, ϕ, the inlet pressure P_0, the inlet temperature T_0 and the initial volume V_0 given the necessary geometry inputs. Once this is known, the mass is determined via the molar mass for each species, and then the mass of the burned and unburned zones is determined via the Wiebe function (Eq. 7

in Section 2.1.1.1). While this is perfectly acceptable for some engine types, this formulation does not capture the complex physics occurring in the scavenging that occurs in the AJAX two-stroke engines. This requires the user to make some justification and assumption for inlet temperature and pressure. The following are the main factors of complexity that will not agree with the Chemkin model and must be considered carefully:

1. In the case of both of the AJAX engines, a substantial vacuum is created during the exhaust, scavenging, and intake process. The Chemkin model will not allow negative pressures as an initial condition. However, if the inlet temperature and pressure are balanced correctly, one may still achieve similar peak pressures to reality, which should still provide a relatively accurate depiction of the heat release rate and global conversion of fuel. In other words, a model that achieves a similar total balance of energy conversion may still be achieved despite this discrepancy. However, intermediate reactions respond sensitively to changes in temperature and pressure, making the accuracy of these outputs likely to be lower.

2. By the definition of the scavenging affect, the inlet air and fuel are mixed with some remaining gases that were just the cycle prior participating in combustion or interacting with molecules that were participating in combustion, meaning they are at elevated temperatures. With the speed of the piston, high temperatures, and high pressures, some impactful heat transfer will take place into the fresh charge, potentially raising it quite substantially from ambient temperatures. This temperature is not known given the equipment used in this thesis. This is a common part of working with 0-D models - iteration until conditions are representative or as close to representative of reality as possible.

Aside from the inlet fuel and air conditions, another vital consideration is the heat transfer in the system. Referencing the work of Annand's dimensionless heat transfer correlation [80] and Woschni's correlation for average cylinder gas velocity over a combustion cycle [81], the coefficients in Chemkin can be tuned within specified ranges to help achieve higher accuracy in the peak pressure and heat release rate predictions.

$$w = C_1 * S_P + C_2 * \frac{V_d * T_{ref}}{P_{ref} * V_{ref}} * (p - p_m) \quad (24)$$

$$Nu = aRe^b Pr^c \quad (25)$$

In Eq. 24, w represents the fluid velocity, S_P the piston velocity, V_d the total displacement volume or swept volume, and $p - p_m$ the difference between instantaneous true pressure and the 'motored' or compression only (no combustion) pressure. The fluid velocity determined from this correlation is then used to determine the Reynold's number which is used for determination of the Nusselt correlation. As is performed in the work of Woschni and Annand, the values selected for the Chemkin model for C_1 and C_2 in Woschni's correlation (Eq. 24) are 2.28 and 0.00324 respectively. The values of a, b, and c in Annand's correlation (Eq. 25) are in the range of 0.35 – 0.8, 0.7 – 0.8, and 0 respectively. The ranges are tuned to achieve higher accuracy given the other selected conditions (inlet air and fuel ratios, temperature, pressure, burn efficiency, etc.).

Finalizing all discussion of preparation and theory necessary, the following section will contain the results for the ML model concerning the methane prediction on the AJAX DPC-81 and AJAX DPC-105.

CHAPTER 4: MODEL RESULTS AND VALIDATION

SECTION 4.1: AJAX DPC81 – ENGINE OPERATION AND MODEL RESULTS

SECTION 4.1.1: ENGINE EMISSION AND COMBUSTION PARAMETER RESULTS

The following figures and data may be observed to obtain the important emission averages, the air to fuel ratios, and the RPM, suction pressure, and discharge pressure at given load cases for the AJAX DPC-81 Engine. This may be used to build one's understanding of how well the engine's combustion is optimized at varying loads by observing the quality of the air-to-fuel ratio at various engine speeds and resistances, along with the resulting emission balances. All emissions data is reported on a 15% O_2, dry basis.

Table 14: AJAX DPC-81 Engine Conditions at Various Load Cases

40% Load Conditions		
RPM	*Suction Pressure [PSI]*	*Discharge Pressure [PSI]*
350	**30**	**400**
50% Load Conditions		
RPM	*Suction Pressure [PSI]*	*Discharge Pressure [PSI]*
375	**45**	**410**
60% Load Conditions		
RPM	*Suction Pressure [PSI]*	*Discharge Pressure [PSI]*
410	**50**	**410**
70% Load Conditions		
RPM	*Suction Pressure [PSI]*	*Discharge Pressure [PSI]*
430	**50**	**480**

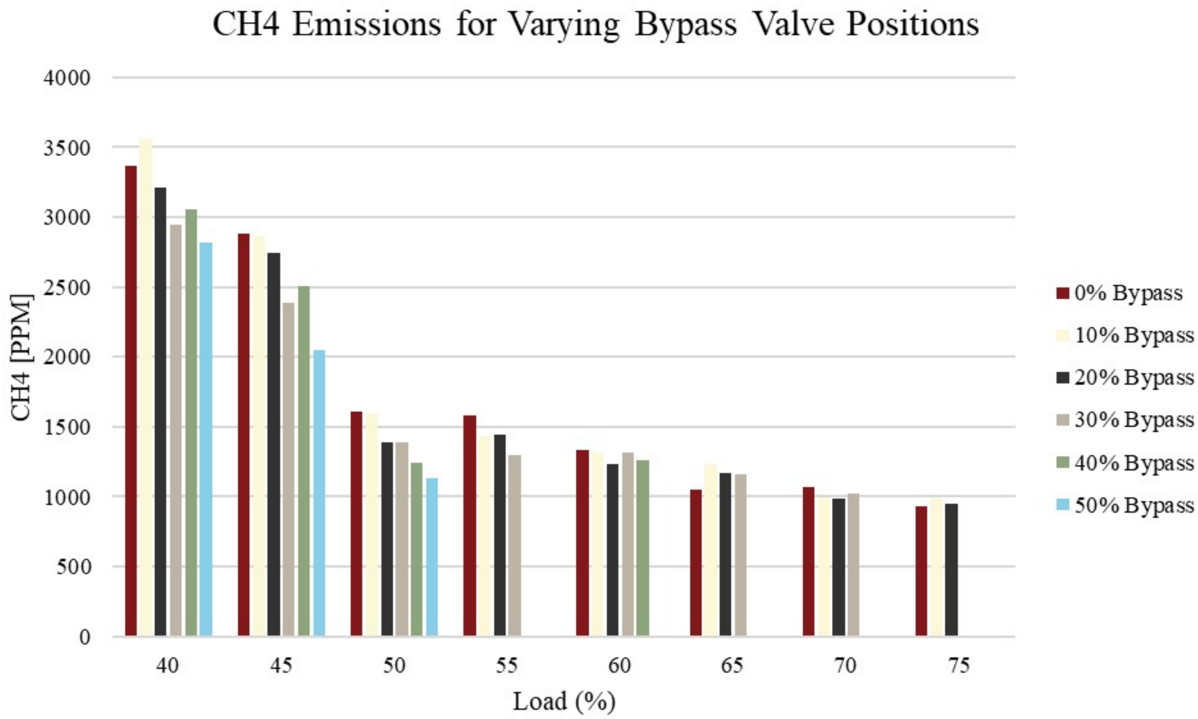


Figure 21: AJAX DPC-81 Methane Emission Averages

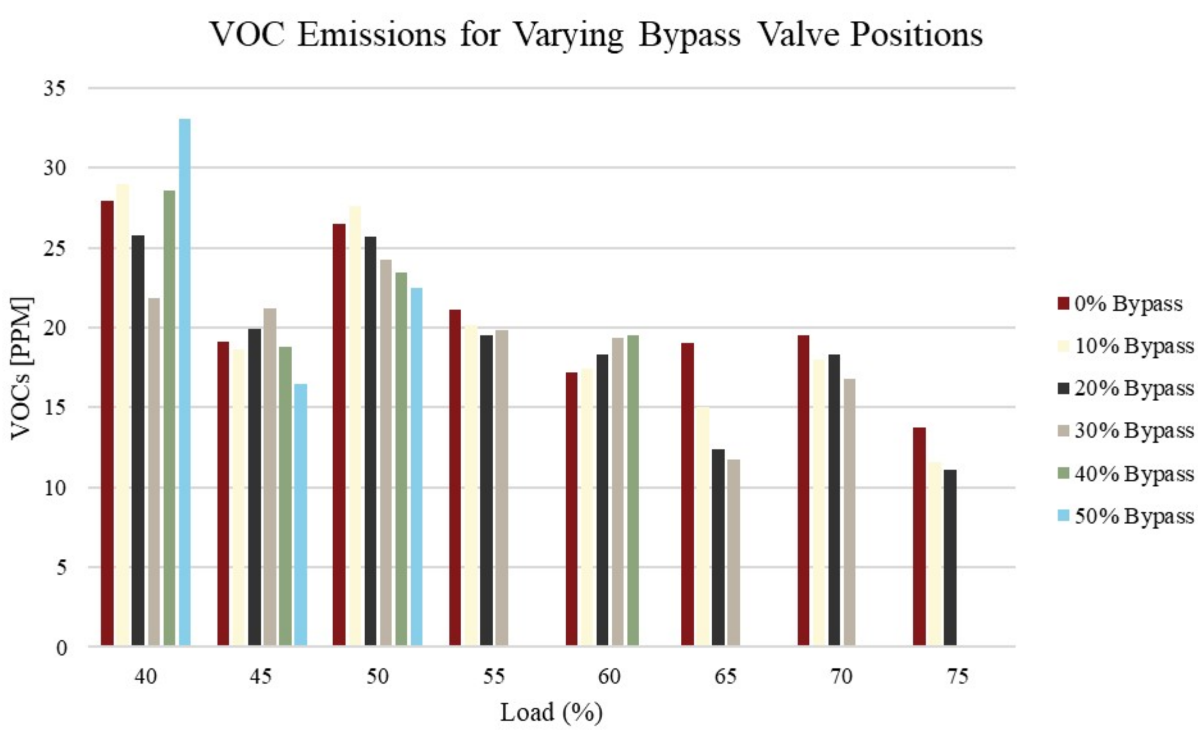


Figure 22: AJAX DPC-81 Volatile Organic Compound Emission Averages

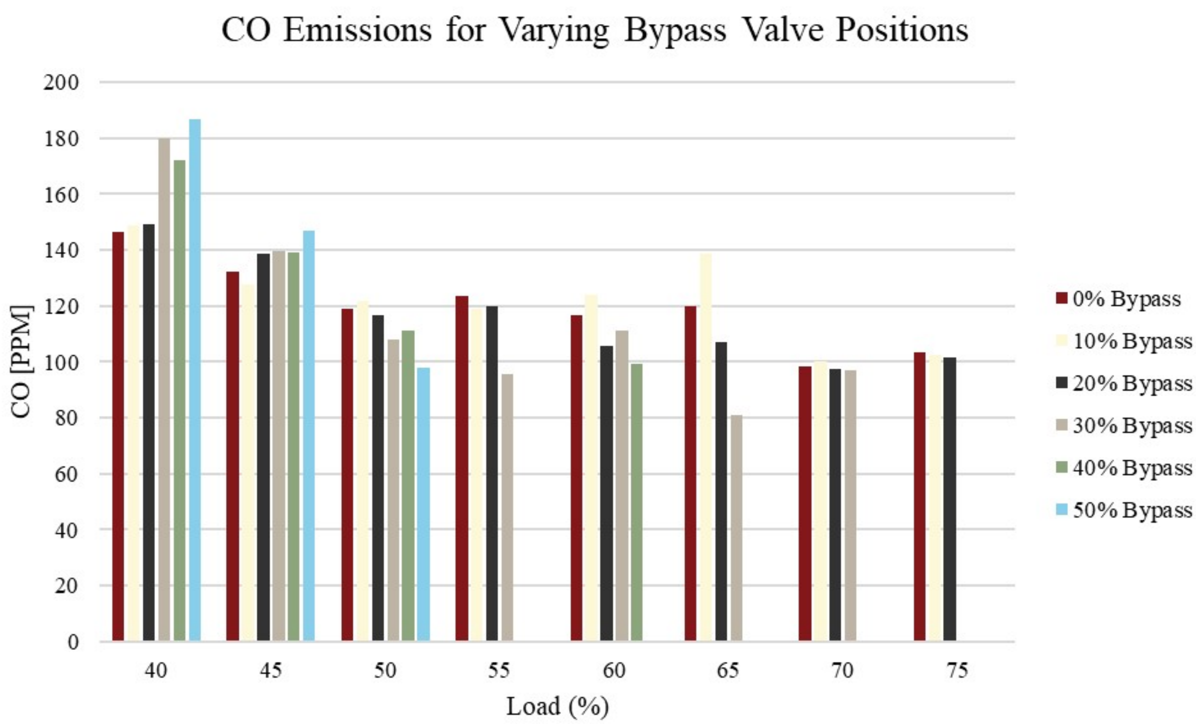


Figure 23: AJAX DPC-81 Carbon Monoxide Emission Averages

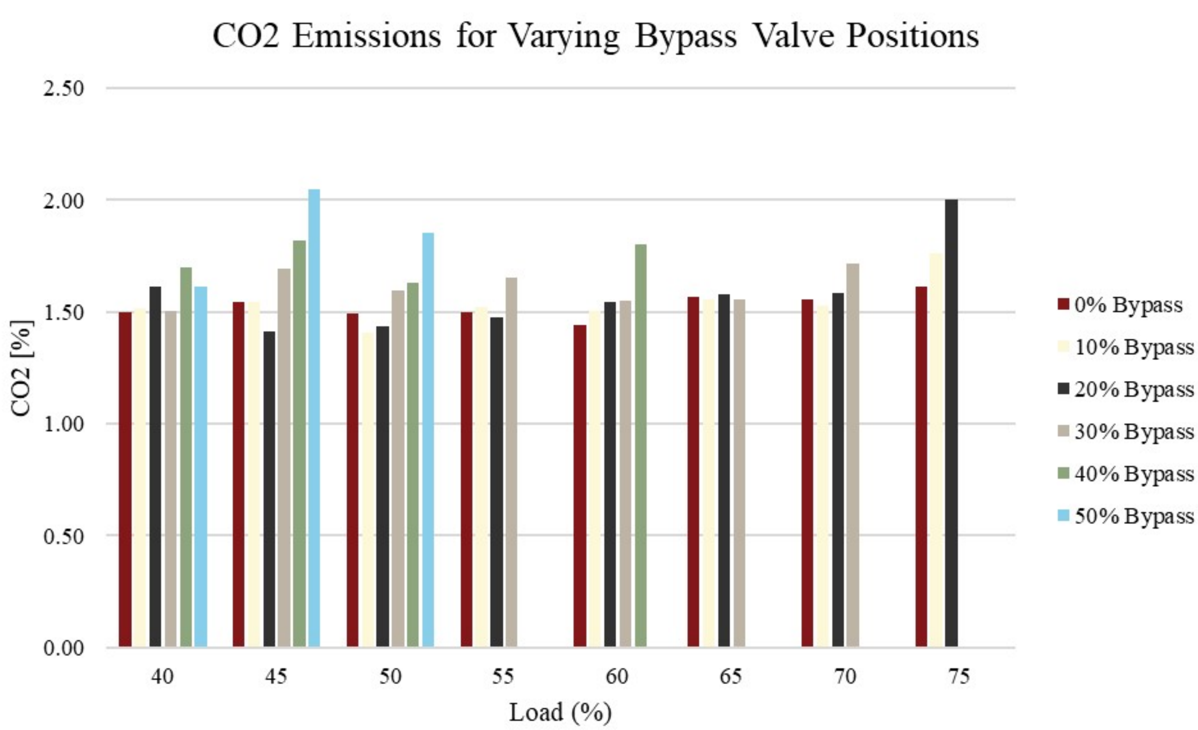


Figure 24: AJAX DPC-81 Carbon Dioxide Emission Averages

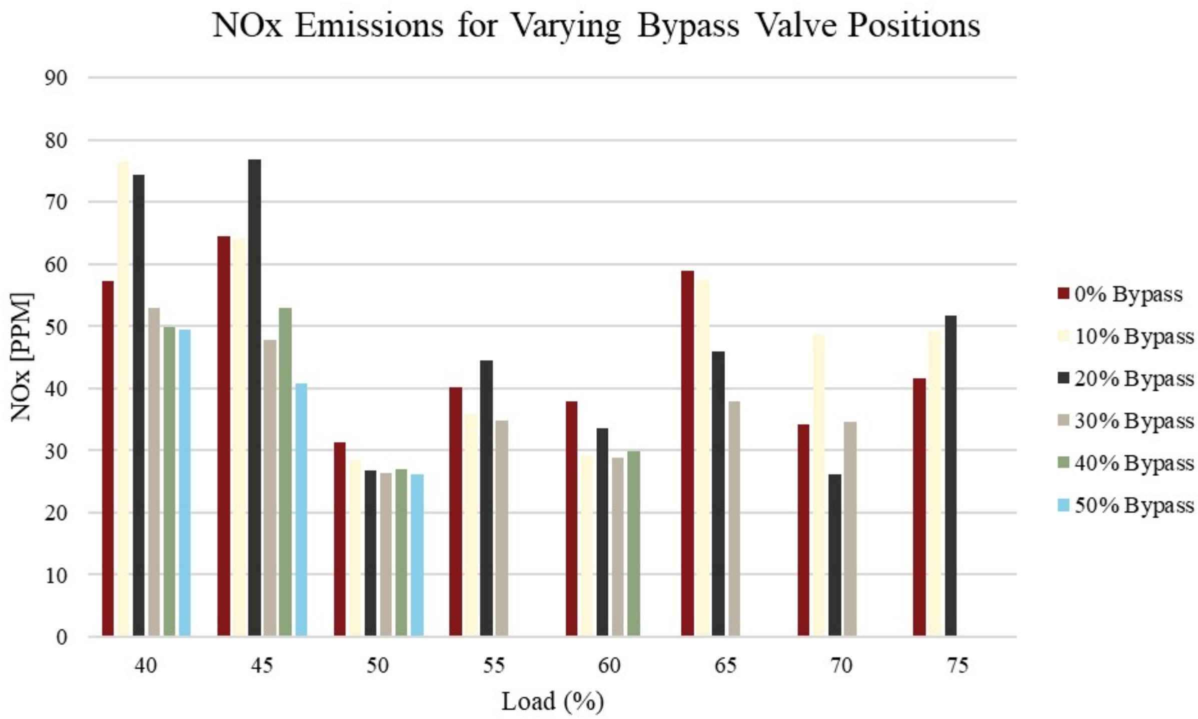


Figure 25: AJAX DPC-81 Nitrogen Oxide Emission Averages

To supplement Figures 21 through 25, Figure 26 can be observed to determine the effect the load and the bypass valve positioning have on the air-to-fuel ratio of the AJAX DPC-81 engine.

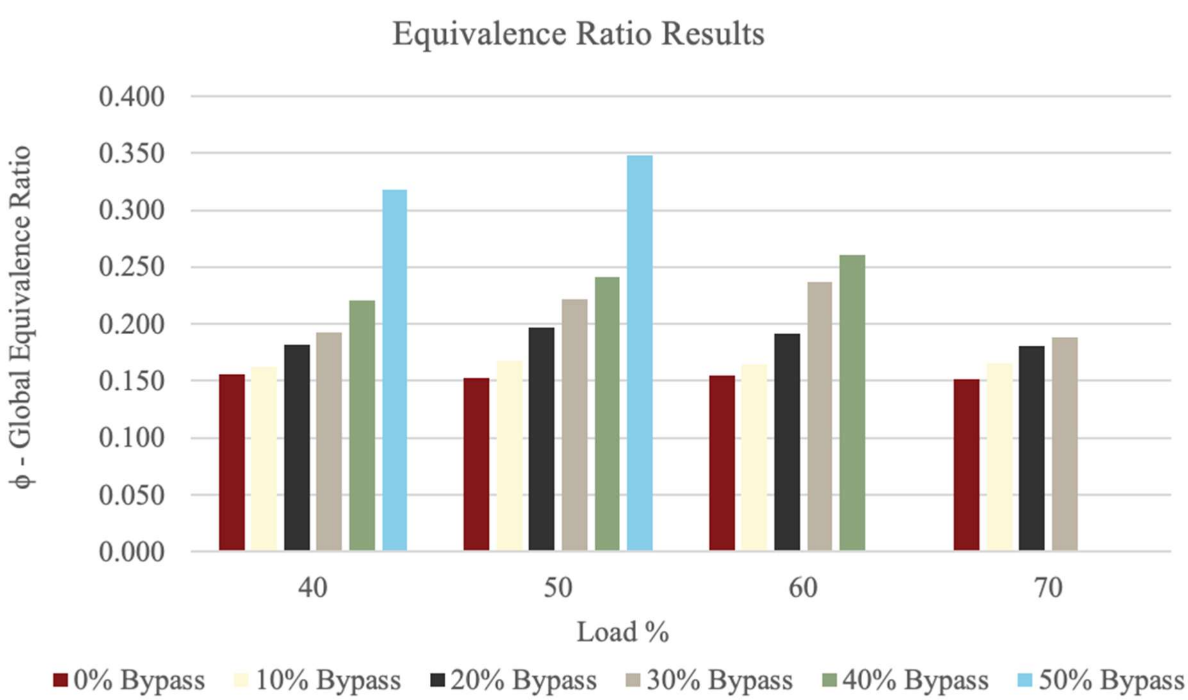


Figure 26: AJAX DPC-81 Global Equivalence Ratio Averages at given Load and Bypass Valve Positionings

From the results shown in Figure 26, some understanding can be gained of the combustion conditions and relate this back to the emissions results observed for the DPC-81. With no air being bypassed, all load conditions for this engine operate at about the same equivalence ratio of around 0.15. From Eq. 20 in Chapter 3 Section 3.3.2, ϕ represents the equivalence ratio and A/F represents the air to fuel ratio. The numerator is the stoichiometric air to fuel ratio for a given reaction, and the denominator the actual ratio. Considering this, an equivalence ratio of less than 1 indicates fuel lean combustion, as in excess air is present, and a ratio of greater than 1 indicates fuel rich combustion where there is less oxidizer present than needed for complete combustion. So, given the quite low equivalence ratios shown for all load and bypass valve cases, the AJAX DPC-81 operates with substantial excess air in the combustion chamber. However, an important consideration is the scavenging process depicted in Figure 13, as the air flow rates calculated are the flow rates coming through the intake piping prior to entry into the combustion chamber. Some of the air that makes it into the combustion chamber is pushed out with the burned fuel through the exhaust port as the scavenging process is conducted, lowering the air to fuel ratio by some extent. This is the reason for the term 'Global' being present in the caption and axis title of Figure 26. Though the impact of this loss is likely not extreme, it is something to consider as it is not directly measurable with the equipment used in this thesis.

Continuing now with the discussion of the results in Figure 26 and their relation to the emissions, a low equivalence ratio can indicate that there may be enough excess air in the combustion chamber to absorb, or steal, energy away from molecules participating in reaction, making complete combustion harder to come by. It can be seen that as the bypass valve is increasingly opened, particularly at lower loads, the equivalence ratio makes its way towards more stoichiometric values. This should result in more energy availability for participating molecules, which should result in further heat release, increased peak combustion pressures, increased peak combustion temperatures, and therefore more complete combustion. This result is indeed the case as can be seen by walking through Figures 21 through 26, focusing attention on 50% load in particular. From no bypass to the case of 50% open, methane PPM is reduced by 29%, volatile organic compound PPMs are reduced by 19%, carbon monoxide PPM is reduced by 18%, and the percentage of carbon dioxide in the exhaust stream is increased by 24%. The incomplete products of combustion (CH_4, CO, VOCs,) are more completely converted to the final product at the end of the reaction chain, CO_2.

For the case of some bypass valve positions, especially at higher load conditions, the exhaust temperature reached a point that exceeded the operating limits provided by AJAX specifications and data could no longer be collected. This is the reason the bypass valve position is not opened greater than 50% and is not opened quite as far the higher the load goes.

To validate the accuracy of the MKS sensor, some chemical reaction equilibrium analysis can be done to the molecules participating in the combustion process to determine appropriate expected ranges of the emissions. This same process should be considered and referenced in the DPC-105 data in Section 4.2.1. To account for methane slip and incomplete reactions leading to hydrogen and carbon monoxide (not water and carbon dioxide) production, a shifted reaction will be considered (Eq. 24) [82].

$$CH_4 + (2)(O_2 + 3.76N_2) \rightarrow CO_2 + 2H_2O + 7.52N_2 \tag{26}$$

$$CH_4 + \left(\frac{2}{\varphi}\right)(O_2 + 3.76N_2) \rightarrow aCH_4 + bO_2 + cCO_2 + dCO + eH_2O + fN_2 \tag{27}$$

Equation 26 represents the stoichiometric combustion of methane, and φ is the equivalence ratio in Equation 27. Performing the mass balance on carbon, hydrogen, oxygen, and nitrogen atoms provides the following results (Table 15):

Table 15: Mass Balance Coefficient Calculations

Molecule	**Balance**
C	$1 = a + c + d$
H	$4 = 4a + 2e$
O	$\frac{4}{\varphi} = 2b + 2c + d + e$
N	$f = 3.76 * \frac{2}{\varphi}$

Given the number of unknowns and the number of equations, some assumptions must be made to calculate these coefficients and determine reasonable ranges for MKS results. At low equivalence ratio, combustion temperatures drop enough for hydrocarbon slip to be high, and for CO production to be elevated [82]. The amount of methane slip depends on a range of factors and is heavily related to the compression ratio at a given equivalence ratio. For our case, methane slip may range between 5 and 10% [83], [84]. Literature also suggests carbon monoxide production may vary significantly depending on residence time, peak temperatures, and compression ratios in engines, but an appropriate range may be somewhere between 1 to 10% [82], [85], [86]. Using some results we achieve in the Chemkin models for SI engine combustion (Chapter 5), we can compare to literature and further educate some appropriate assumptions for methane, carbon monoxide, and hydrogen remnants. As referenced in Chapter 2 and in Chapter 3, this model considers detailed reaction kinetics for methane combustion as well as the geometry and residence time of the specific engines in this thesis, making it a good place to derive some assumptions for this validation. The following table (Table 16) contains data for the DPC-81 at a range of equivalence ratios. Further discussion of this model is contained in Chapter 3 Section 3.3 and in Chapter 5.

Table 16: Chemkin 0D SI Engine Model Results for Methane Combustion at Varying Equivalence Ratio for the DPC-81 Engine

φ	**CH_4 PPM (Dry, 15% O_2)**	**CO PPM (Dry, 15% O_2)**	**CO_2 PPM (Dry, 15% O_2)**	**H_2 PPM (Dry, 15% O_2)**	**Water %**	**Oxygen %**
0.25	2695	0.01	2.78	0.00	4.52	15.9
0.3	2680	0.02	2.76	0.00	5.40	14.9
0.35	2682	0.02	2.77	0.00	6.26	13.9

From the literature cited and Table 16, a range of potentially viable outputs are calculated and tabulated for the range of potential equivalence ratios one might find in the DPC-81 and DPC-105 engines. These results are calculated using Eq. 27 and the assumptions to solve for all unknown coefficients are made for a min-max range of the following: CH_4 slip occurs in the range of 5 to 10%, CO production occurs in the range of 1 to 10%, and there is no hydrogen production.

Table 17: Shifted Equilibrium Reaction Results Considering 5-10% Methane Slip and 1-10% Carbon Monoxide Production

φ	**CH_4 PPM (Dry, 15% O_2)**	**CO PPM (Dry, 15% O_2)**	**CO_2 % (Dry, 15% O_2)**	**H_2 PPM (Dry, 15% O_2)**	**Water %**	**Oxygen %**
0.14	1763 – 3781	353 – 3781	3.32 – 3.02	0	2.75 – 2.61	18.46 – 18.64
0.2	1762 – 3777	352 – 3777	3.31 – 3.02	0	3.91 – 3.7	17.35 – 17.61
0.3	1760 – 3774	352 – 3774	3.31 – 3.02	0	5.8 – 5.49	15.47 – 15.87
0.4	1760 – 3772	352 – 3772	3.31 – 3.02	0	7.66 – 7.24	13.56 – 14.1

Considering the literature, the Chemkin modeling, and the results derived from Eq. 27, the CO_2% found in reality seems to be lower than theoretical results, while other MKS results including the methane and carbon monoxide seem plausible. However, this discrepancy may be refuted considering the oxygen percent as reported by the amperometric NOx sensors which are separate from the MKS gas analyzer system. The oxygen percents found from these sensors are reported for the majority of all measurements, load percentages, and bypass positions for both AJAX engines to be in the range of 9-11% which is lower than that calculated by Chemkin or the shifted equilibrium reaction. This may indicate a higher *local* equivalence ratio (see discussion above after Figure 26), resulting in lower carbon dioxide emissions in reality with similar methane and carbon dioxide results as reported in Table 16 and Table 17. If we consider an oxygen % of 10% as opposed to 18% in the case of an equivalence ratio of 0.14 in Table 17, the correction factor for conversion to a 15% O_2 basis would be $\frac{21-15}{21-10} = 0.545$ as opposed to $\frac{21-15}{21-18} = 2$, resulting in CO_2 measurements more in line with what the MKS results suggest.

SECTION 4.1.2: ANALYSIS OF INPUT FEATURE CORRELATION TO EMISSIONS

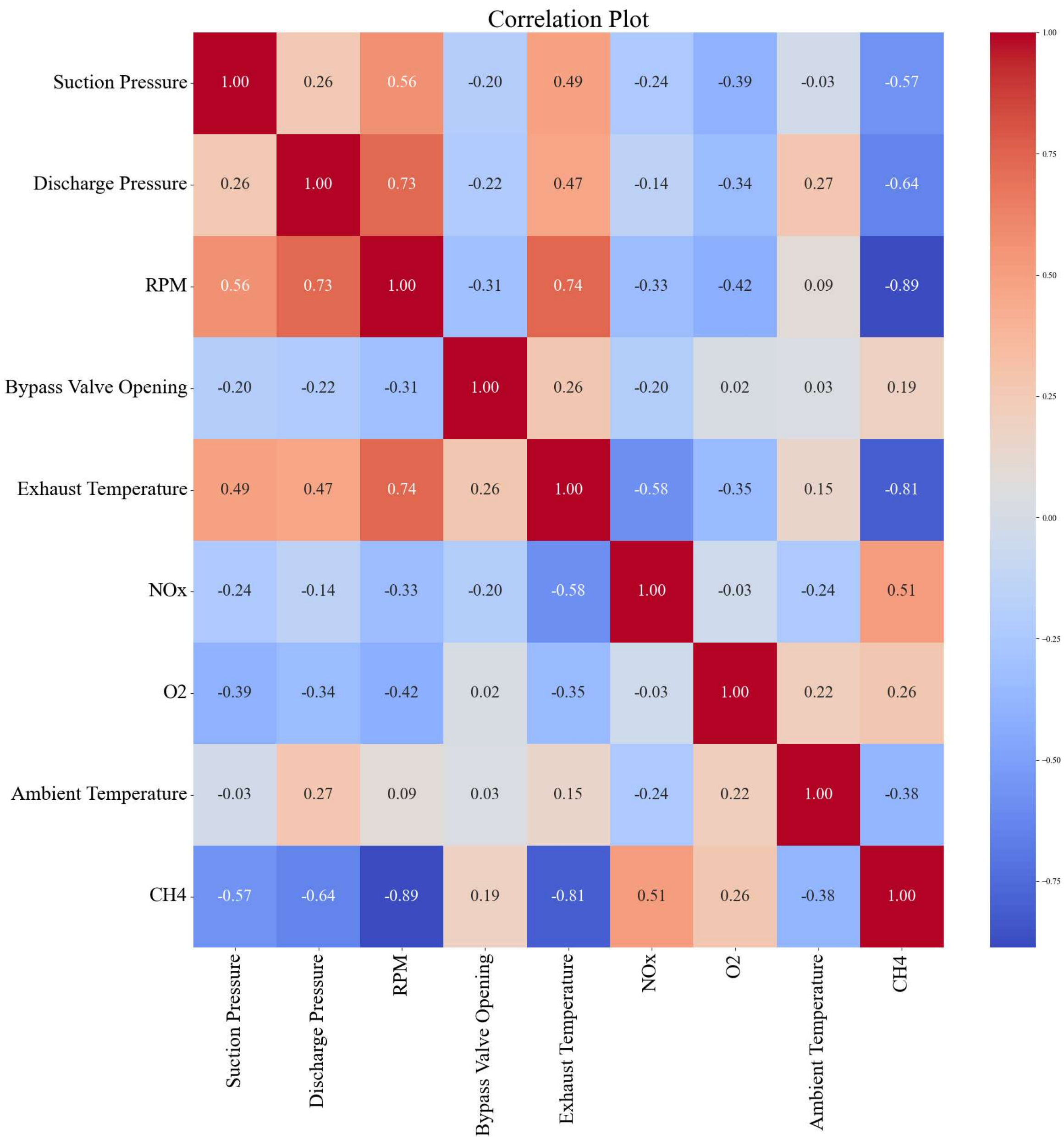


Figure 27: AJAX DPC-81 Training Data Correlation to Emissions

From the correlation analysis, of the final input data selection choices there are no metrics that are linearly correlated to a concerning extent and all sensors have a non-negligible correlation to the methane emissions.

Table 18: Training Dataset Statistics

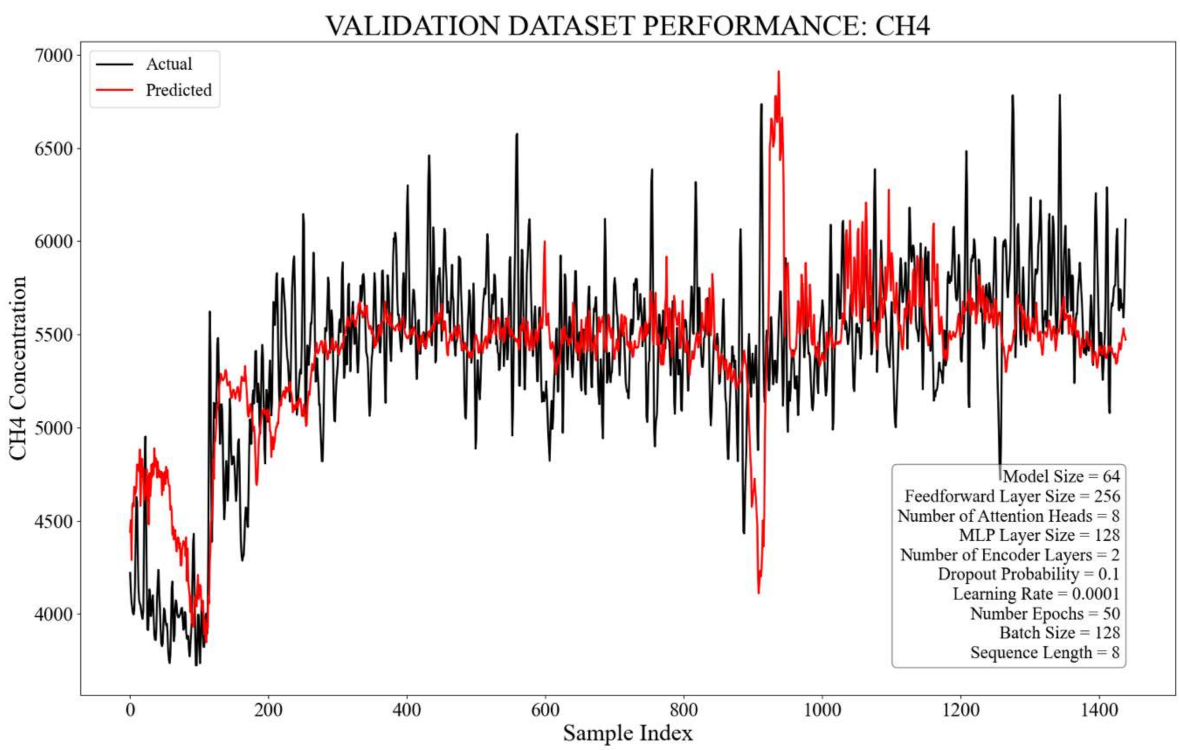
Parameter	Units	Mean	St. Dev.	Min.	Max.	Median	Variability
Suction Pressure	PSI	41.6	14.12	19.63	157.89	39.5	199.37
Discharge Pressure	PSI	411.52	38.4	353.04	533.22	398.94	1474.26
RPM	RPM	394.62	31.86	336.7	494.8	394.6	1014.89

Bypass Valve Opening	%	19.52	15.28	0	99.6	19.7	233.37
Exhaust Temperature	°F	755.94	42	630.5	832.64	766.04	1764.35
NOx	°F	98.33	53.05	1.07	1242.83	87.03	2814.77
O2	PPM	10.38	0.68	7.74	13.16	10.43	0.46
Ambient Temperature	%	69.5	13.32	49.1	97.34	68.72	177.36
CH4 Concentration	PPM	4070.28	1857.15	1114.25	8394.8	4399.49	3449003

SECTION 4.1.3: MACHINE LEARNING MODEL PERFORMANCE

Table 19: Training, Testing, and Validation Average MAE for the AJAX DPC-81 Data

Training MAE	Test MAE	Validate MAE
164.27	200.84	337.2



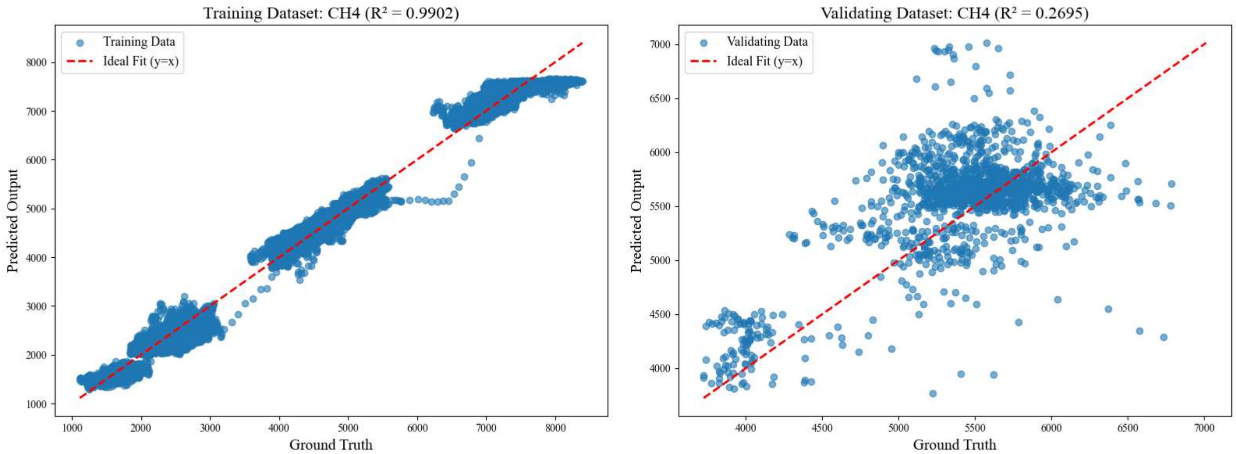


Figure 28: Virtual Sensor Performance on Training and unseen Validation Data for the AJAX DPC-81 Data

In Table 19 and Figure 28, the performance of the model on the unseen validation dataset for the DPC-81 is displayed. While the R^2 plot does not yield the greatest looking results for the validation dataset, if the predictions overlaid on top of the ground truth methane values are observed from the topmost portion of Figure 28, the performance is quite good. The validation dataset is noisier when compared to the training dataset, which may be a result of elevated water content in the MKS gas analyzer on the given day of testing. The trend and average value of the ground truth value is closely followed in this case, and the performance is considered successful. This model would likely perform well across multiple seasons as well due to the long stretch of data collection as the ambient temperature changed through the year.

SECTION 4.2: AJAX DPC105 – *ENGINE OPERATION AND MODEL RESULTS*

SECTION 4.2.1: ENGINE EMISSION AND COMBUSTION PARAMETER RESULTS

The following figures and data may be observed to obtain the important emission averages, the air to fuel ratios, and the RPM, suction pressure, and discharge pressure at given load cases for the AJAX DPC-81 Engine. This may be used to build one's understanding of how well the engine's combustion is optimized at varying loads by observing the quality of the air-to-fuel ratio at various engine speeds and resistances, along with the resulting emission balances. All emissions data is reported on a 15% O_2, dry basis.

Table 20: AJAX DPC-105 Engine Conditions at Various Load Cases

40% Load Conditions		
RPM	*Suction Pressure [PSI]*	*Discharge Pressure [PSI]*
300	**2.6**	**31**
50% Load Conditions		

RPM	*Suction Pressure [PSI]*	*Discharge Pressure [PSI]*
320	**3.5**	**32**
60% Load Conditions		
RPM	*Suction Pressure [PSI]*	*Discharge Pressure [PSI]*
400	**3.0**	**32**
70% Load Conditions		
RPM	*Suction Pressure [PSI]*	*Discharge Pressure [PSI]*
425	**5.0**	**37**

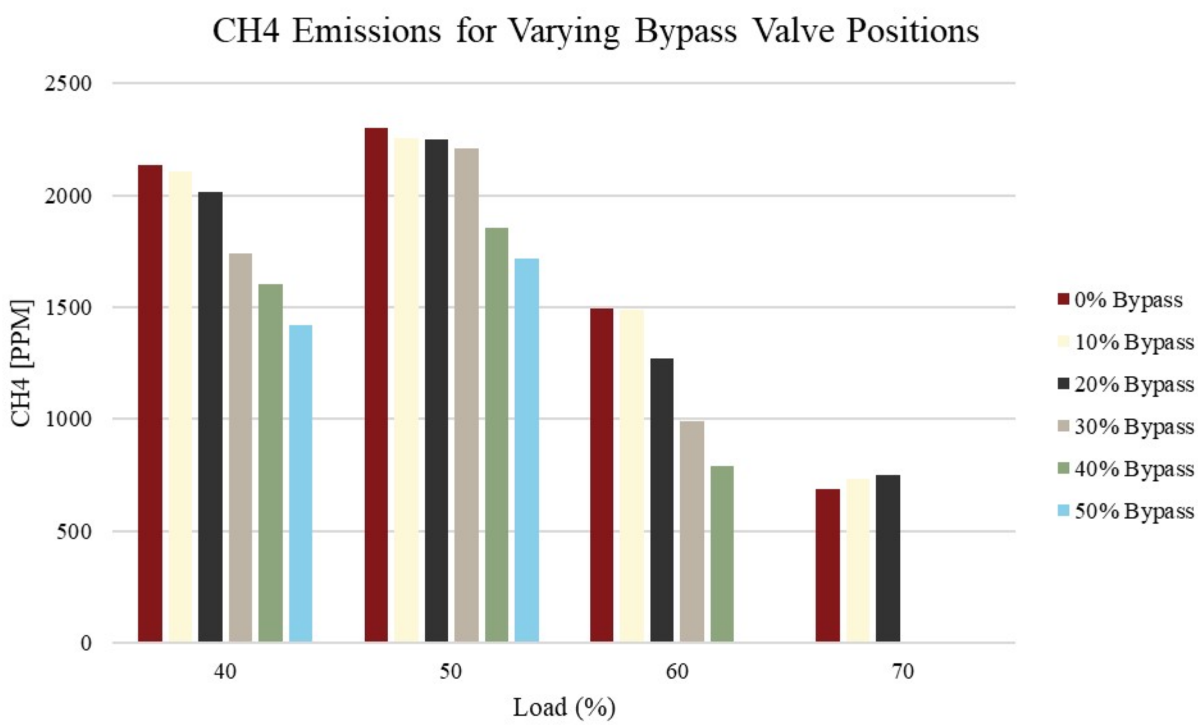


Figure 29: AJAX DPC-105 Methane Emission Averages

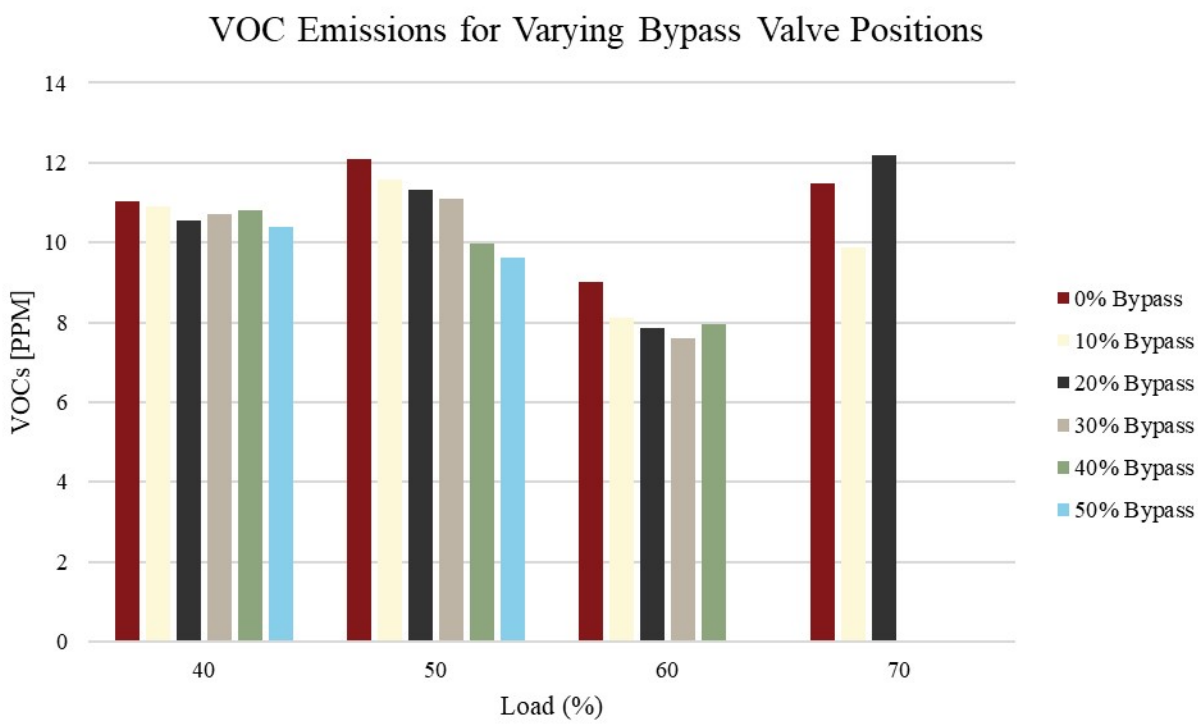


Figure 30: AJAX DPC-105 Volatile Organic Compound Emission Averages

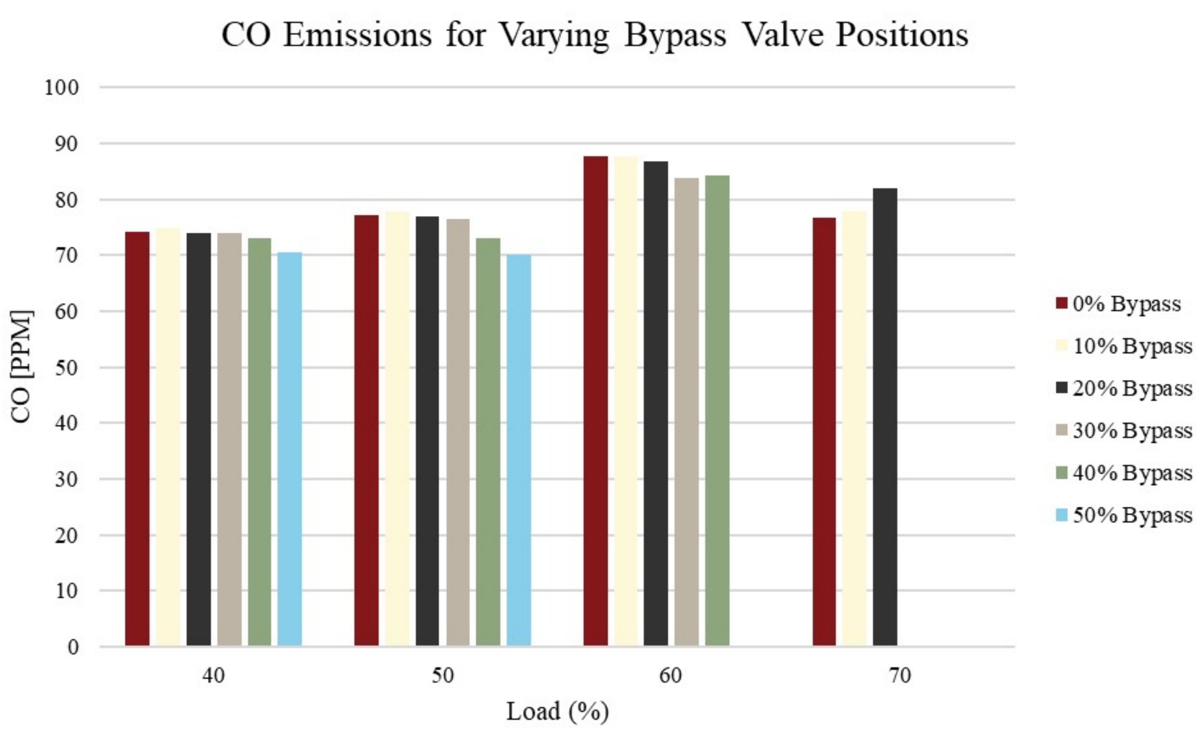


Figure 31: AJAX DPC-105 Carbon Monoxide Emission Averages

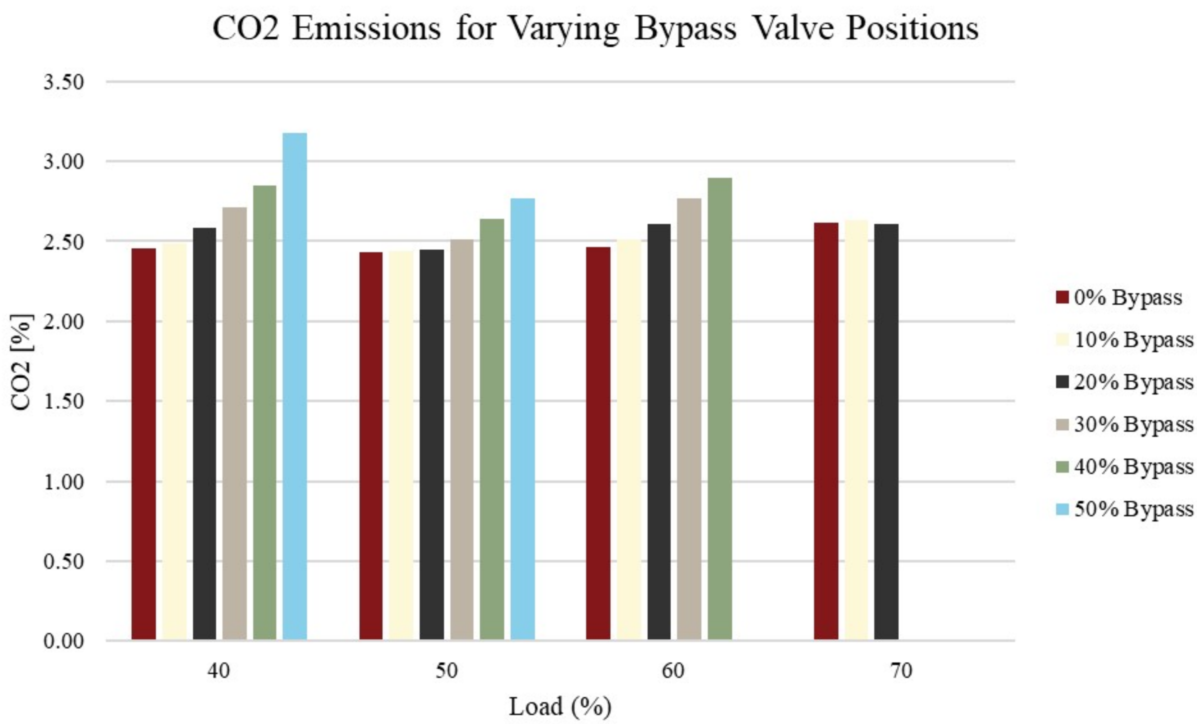


Figure 32: AJAX DPC-105 Carbon Dioxide Emission Averages

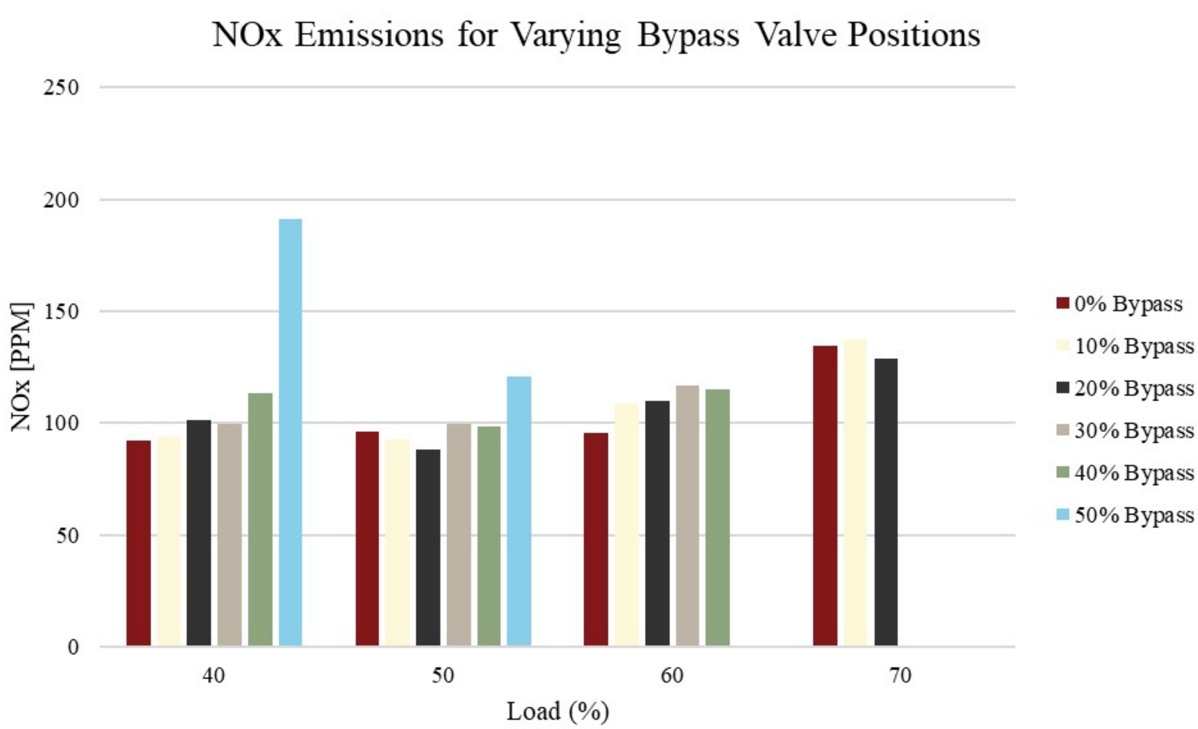


Figure 33: AJAX DPC-105 Nitric Oxide Emission Averages

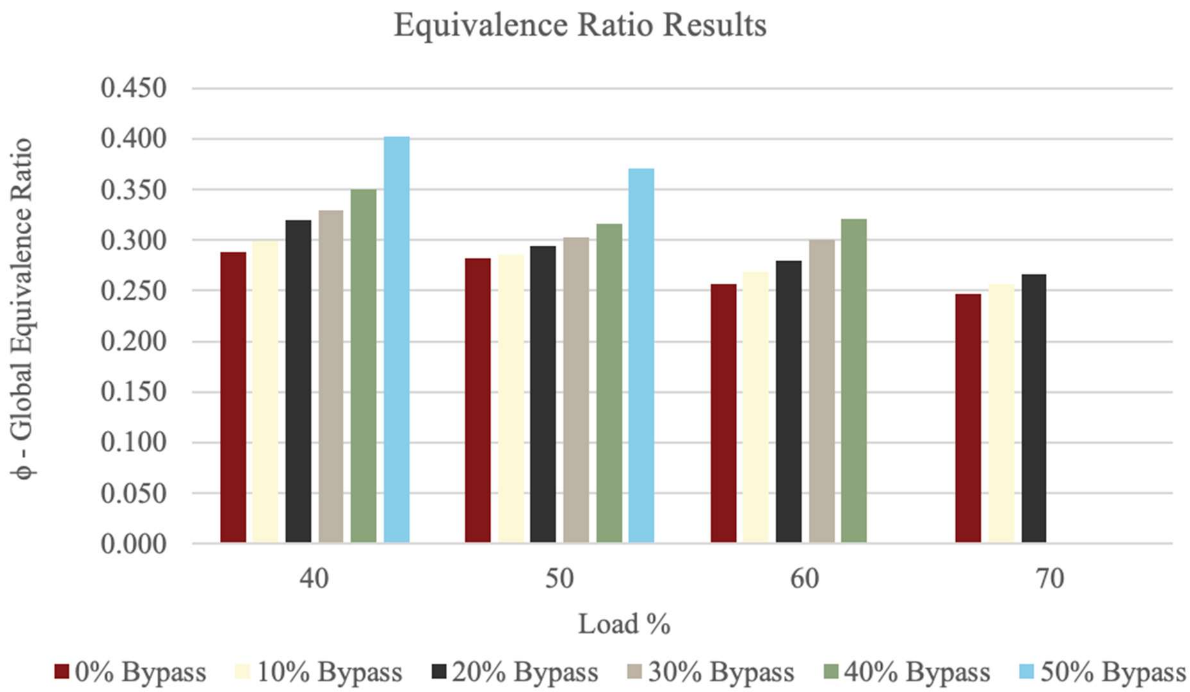


Figure 34: AJAX DPC-105 Equivalence Ratio Averages

The engine emissions data collected (Figures 29-34) across varying load and bypass conditions offer meaningful insights into combustion efficiency, fuel utilization, and pollutant formation. Methane (CH_4) emissions decline consistently with increasing bypass, especially at lower loads. For instance, at 40% load, CH_4 drops from approximately 2132 ppm at 0% bypass to 1418 ppm at 50% bypass. Volatile organic compounds (VOCs) exhibit a similar downward trend with increasing bypass valve opening, though the change is not as substantial. Carbon monoxide (CO) levels show a more complex response. At lower loads, CO exhibits a slight increase with bypass, while at higher loads, it tends to decrease or fluctuate less predictably. These variations may reflect the interplay between richer combustion (which can promote CO formation due to incomplete oxidation) and leaner or cooler combustion (which can also lead to incomplete CO-to-CO_2 conversion). Carbon dioxide (CO_2), the main indicator in this case of combustion completeness, generally increases with both load and bypass, rising from 2.46% to 3.17% across the bypass sweep at 40% load. Nitrogen oxides (NOx) show a nonlinear trend. At lower loads and moderate bypass, NOx levels increase slightly, but at higher bypass and load combinations, the rise becomes more pronounced. For example, at 40% load, NOx climbs from around 92 ppm at 0% bypass to over 190 ppm at 50% bypass. This spike may be attributed to higher combustion temperatures or increased residence times the closer to stoichiometric conditions the combustion phenomena get, both of which favor thermal NOx formation.

The equivalence ratio (ϕ) data from Figure 34 increases with bypass across all loads, particularly at lower loads where it rises from 0.288 to 0.403 at 40% load. At higher loads, equivalence ratios remain leaner, generally staying below 0.30, which contributes to lower CH_4 but sometimes corresponds with elevated NOx due to hotter combustion conditions. In summary, bypassing air in this natural gas engine setup appears to improve combustion completeness as indicated by declining CH_4 and VOCs and rising CO_2 levels. However, the resulting richer mixtures and hotter zones can lead to increased NOx and variable CO levels.

SECTION 4.2.2: ANALYSIS OF INPUT FEATURE CORRELATION TO EMISSIONS

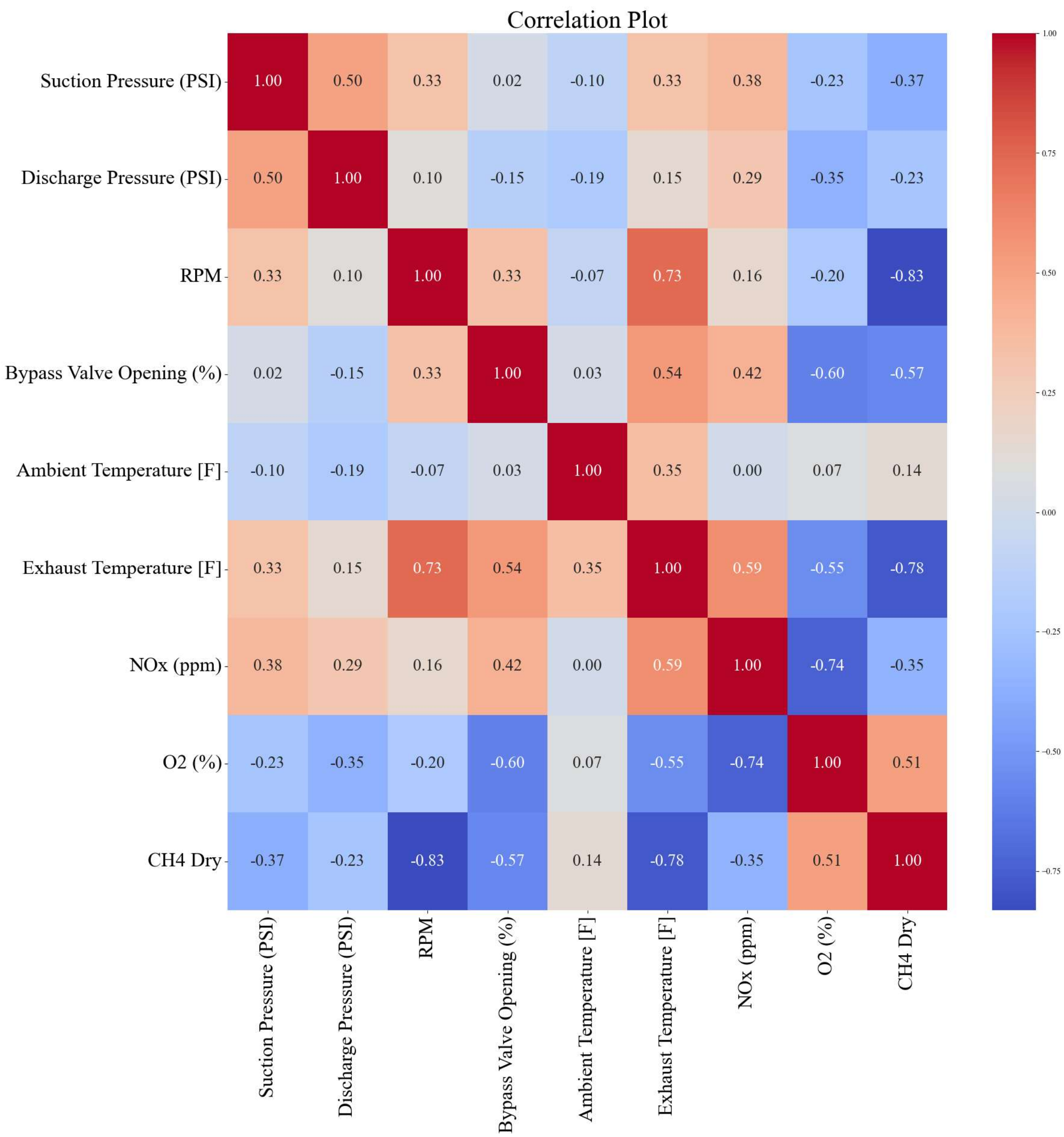


Figure 35: Correlation Analysis for the AJAX DPC-105 Data

Like that of the data correlation analysis performed for the DPC-81, many of the coefficients shown in Figure 35 are close in value to Figure 27. The RPM and exhaust temperature contribute the highest correlations to methane, but do not provide redundancies.

Table 21: Training Dataset Statistics

Parameter	Units	Mean	St. Dev.	Min.	Max.	Median	Variability
Suction Pressure	PSI	3.30	1.12	0.54	10.69	3.08	1.25
Discharge Pressure	PSI	35.55	6.13	23.44	65.43	32.72	37.60
RPM	RPM	355.58	45.07	277.20	429.10	342.90	2030.76
Bypass Valve Opening	%	24.87	21.27	0.00	98.50	19.30	452.57
Ambient Temperature	°F	33.60	18.97	-2.84	374.70	32.40	359.64
Exhaust Temperature	°F	478.04	44.21	404.30	793.94	471.98	1954.47
NOx	PPM	223.48	319.53	1.64	1899.98	121.51	102089.8
O2	%	10.08	1.08	0.53	12.93	10.23	1.16
CH4 Concentration	PPM	1442.37	635.40	363.03	2863.79	1469.13	403698.5

SECTION 4.2.3: MACHINE LEARNING MODEL PERFORMANCE

In similar fashion to the DPC-81 data, the following figures and tables contain the performance data of the transformer-based AI model in the task of predicting methane emissions for the AJAX DPC-105. Some important considerations for this data are the following: Due to time constraints, there are about half of the total datapoints collected when compared to the AJAX DPC-81 dataset. This is still a substantial amount of data (11,301 total datapoints), but the variety of ambient conditions are less diverse due to the shorter seasonal timeframe of testing. This plays heavily into the validation performance discussion, as the temperature outside at the time of the collection of this dataset was around 20°F as opposed to around 60°F during the majority of the training dataset collection.

Table 22: Training, Testing, and Validation Average MAE for the AJAX DPC-105 Data

Training MAE	Test MAE	Validate MAE
221.75	302.98	90.43

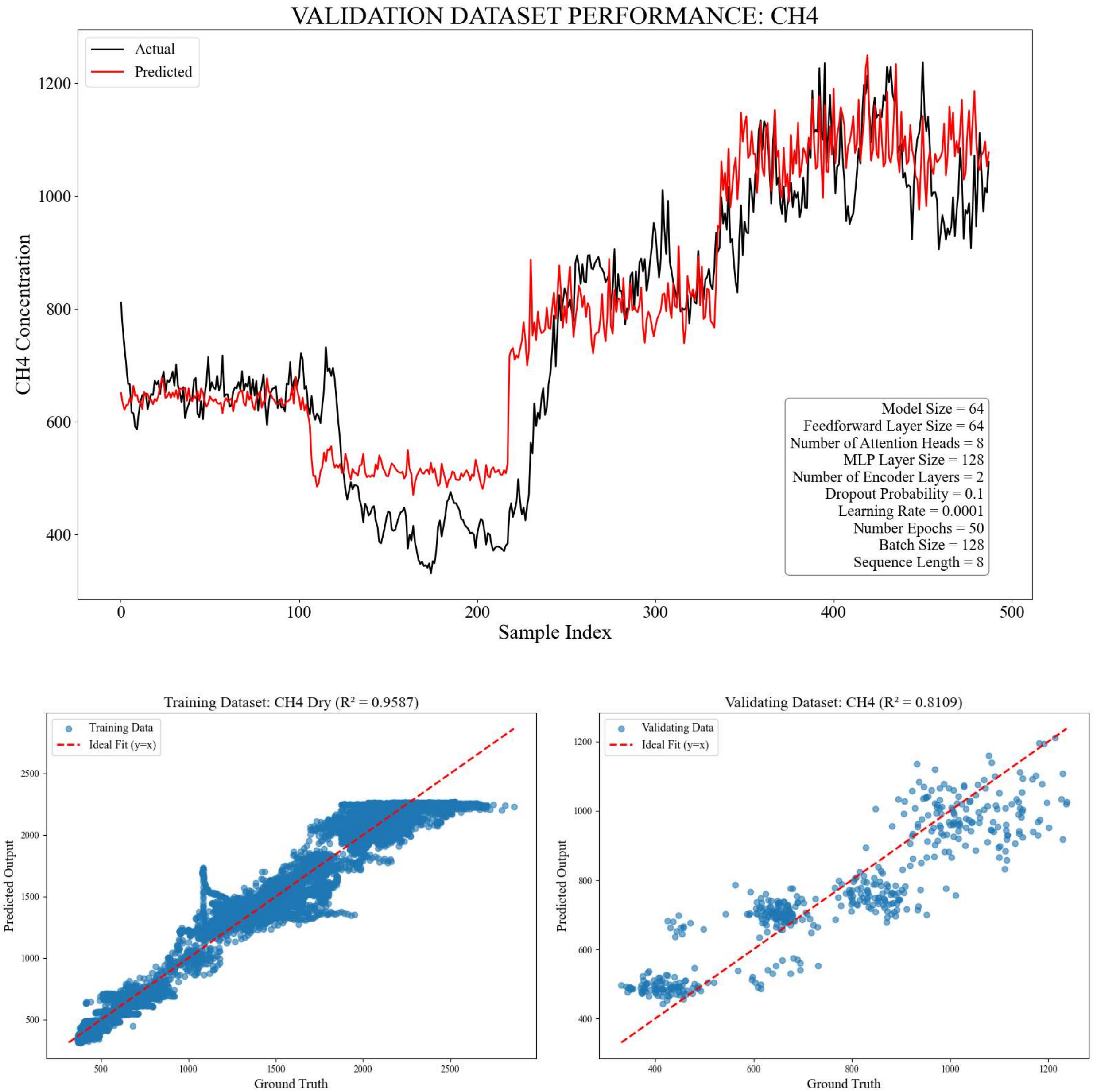


Figure 36: Virtual Sensor Performance on Training and unseen Validation Data for the AJAX DPC-105 Data

From Figure 36 and Table 22, the performance of the transformer-based architecture is shown on the unseen validation dataset for the AJAX DPC-105. While the R^2 score as well as the MAE is substantially improved on the unseen validation dataset in the case of the DPC-105 as compared back to the AJAX DPC-81, the true evaluation of performance is more complex. Due to the reduced amount of data, the model size was able to be reduced (note the hyperparameters listed in the legend), but not all of the minimum and maximum load conditions were captured with as much data as was the case for the DPC-81. Particularly for the dataset, a high RPM (~380) and discharge pressure (~40 psi) paired with a very large bypass valve opening for a short time is what allowed the exceptionally low PPM values below 400 PPM. Due to this state being above the temperature operation limits of the engine if left at steady state, no data at this low of methane concentration was collected for training. The model did a decent job of generalizing from the training data considering this fact. From the training data results in the lower left-hand side of Figure 36, it can also be observed that the training performance falls as methane PPMs get higher than 2000 PPM. Adding more training epochs or using higher complexity models were both adjustments that corrected this discrepancy, but the validation and test dataset performance fell off substantially once this data was learned more closely during training. This is an indicator that more data in this extremely low load, high methane PPM region will likely be necessary to have a model capable of retaining accuracy in these regions.

Moving forward, the mathematical validation models will be presented and the results compared back to the ML model performance shown in these previous sections.

CHAPTER 5: EVALUATION AND COMPARISON

This chapter is written to serve the purpose of comparing the results obtained from the machine learning models to modern internal combustion engine modeling techniques. As mentioned within Chapter 3 Section 3.3, the model employed for this purpose is Chemkin's 0D SI engine model.

SECTION 5.1: AJAX DPC-81 MATHEMATICAL MODEL RESULTS

The results will flow in this order:

1. The Wiebe function parameters are found by fitting the experimental burned mass fraction profile for input into the model.
2. Next a discussion and estimation of an appropriate inlet pressure value is conducted.
3. Following this the results for a single load case at a variety of plausible equivalence ratios are shown. Each set of results is tuned to match the peak pressure and heat release rate plots as closely as possible.
4. Then, the resulting emissions of these tuned models are shown and discussed.

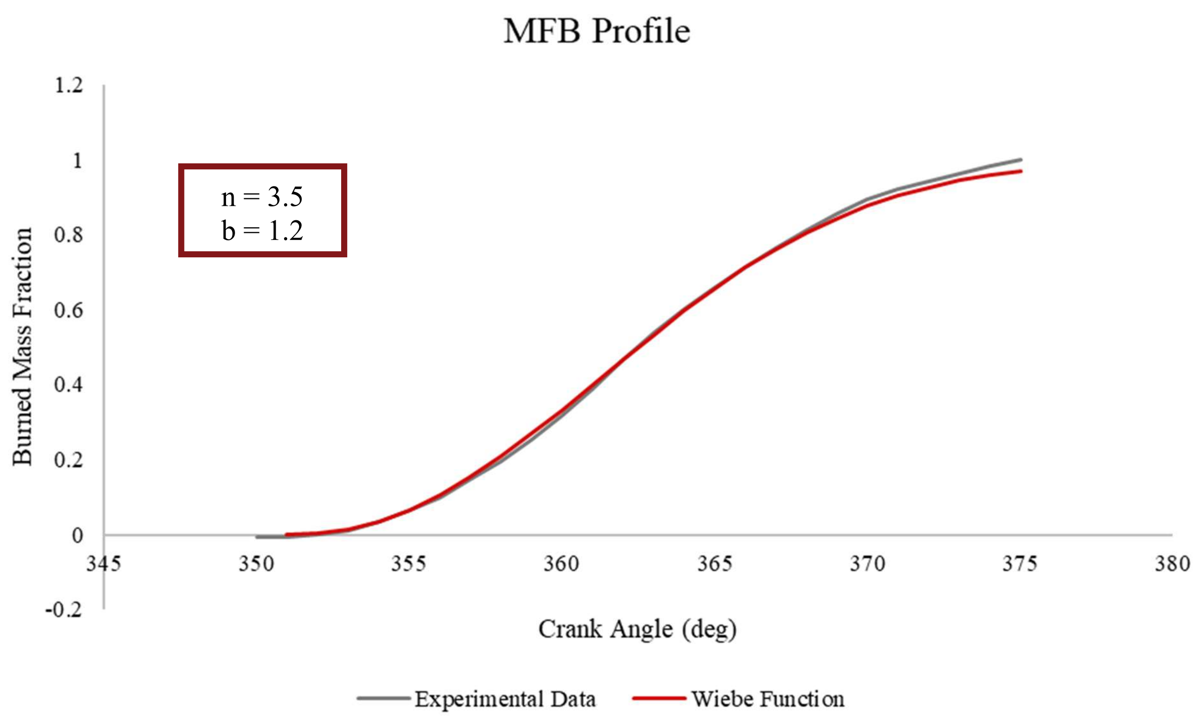


Figure 37: Experimental and Modeled Burned Mass Fraction Profile for the AJAX DPC-81

Based on the fit achieved in Figure 37, the values for n and b were selected as 3.5 and 1.2 respectively for input into Chemkin.

Referring back to Chapter 3 Section 3.3.2, to estimate what pressure the Chemkin model should start at after intake valve closure in order to get as close as possible to experimental results, a quick analysis of adiabatic compression is done to find a potential initial pressure value.

$$\frac{P_{TDC}}{P_{air\ intake\ closure}} = \left(\frac{V_{BDC}}{V_{TDC}}\right)^n \tag{28}$$

In Eq. 25, since the equation is representative of adiabatic conditions (no heat is transferred) n is selected to be slightly lower than the specific heat of air at a value of 1.3 to help account for some heat loss [87]. For the DPC-81, peak compression pressure prior to spark ignition is around 150 psi, and the ratio of volume at bottom dead center to the volume at top dead center is the definition of the compression ratio which is somewhere between a value of 7 and 8. Given this data, the pressure at air intake closure may be estimated to be around 10 psi and is used for the starting pressure value in the Chemkin model.

When using this estimation, due to the nature of the ideal gas law's use in governing the inlet fuel and air quantity, it is possible to bring the inlet temperature down to below 400K, which is closer to what it potentially is in reality considering the ambient status prior to mixing during scavenging. As the temperature comes down, the density goes up and more fuel and air fill the system at a given pressure. If a higher inlet pressure is used, the inlet temperature must be drastically increased to reduce the total amount of fuel and air entering the system if any semblance of reality concerning compression and peak pressures is hoped to be retained. This is true for the AJAX DPC-105 results below as well.

From the data in Figure 26, it can be seen that the AJAX DPC-81 average global equivalence ratio at 70% load for any bypass position is around 0.25. As previously discussed, global indicates the measured value at the fuel and air flow rate sensors, and the scavenging process likely removes additional air from the combustion chamber which should raise the equivalence ratio by some amount. This being said, the 70% load condition is modeled for an increasing range of equivalence ratios and tuned to experimental pressure traces. Final emissions outputs will be shown and described for each case. The most comparable and realistic case will be identified.

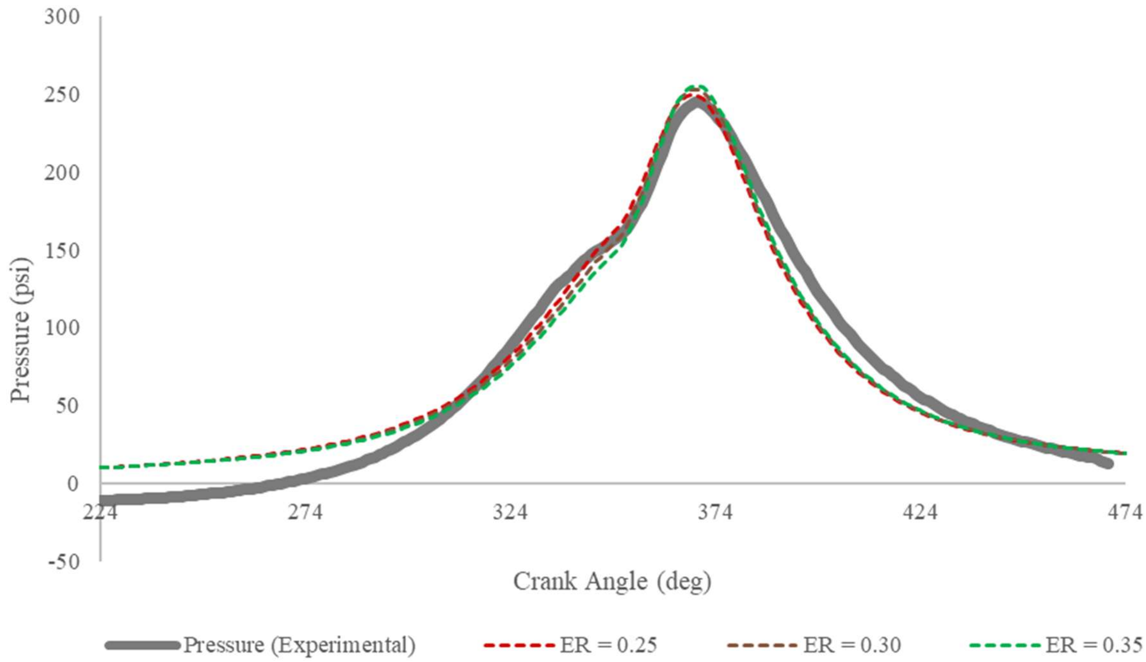


Figure 38: Chemkin SI Engine Modelling In-Cylinder Pressure Results at 70% Load

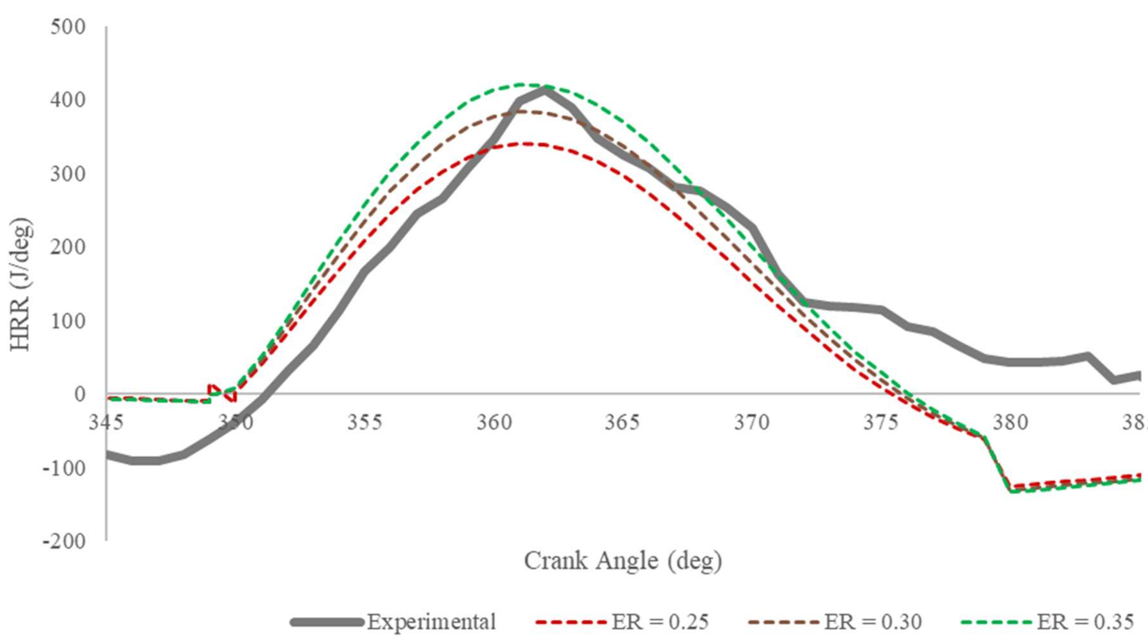


Figure 39: Chemkin SI Engine Modelling Heat Release Rate Results at 70% Load

Table 23: Chemkin SI Engine Model Tabulated Emissions Results for the AJAX DPC-81 (all conditions are on a dry, 15% O_2 basis)

ϕ	CH4 (PPM)	CO (PPM)	CO2 (%)	NOx (PPM)
0.25	2695	0.009	2.78	0.0005
0.30	2680	0.015	2.76	0.0081
0.35	2682	0.020	2.77	0.0839
Reality	1067	98	1.56	34

From Table 23 and Figure 38 and 39, the data can be observed and considered against what is occurring in reality. The simulation results demonstrate a consistent trend of underpredicting several key emissions species relative to experimental measurements. Most notably, NOx levels are drastically underpredicted, with modeled values remaining below 0.1 PPM across all equivalence ratios—orders of magnitude lower than the 34 PPM observed in practice. Given that thermal NOx formation is highly temperature-dependent and primarily governed by the Zeldovich mechanism, this suggests that the model significantly underestimates peak combustion temperatures. This may be due to overly conservative assumptions about wall heat losses, low inlet temperatures (possibly underestimating the impact scavenging-induced heating has on the inlet mixture), or insufficient representation of high-temperature localized zones in the 0D model. These difficulties go hand in hand with the previous discussions of how the model introduces the fuel and air mixture into the system.

CO_2 concentrations are overpredicted in the model, with values consistently around 2.76–2.78%, compared to the measured 1.56%. This discrepancy may indicate an overestimation of combustion completeness in the model, particularly under lean conditions. Supporting this conclusion is the overprediction of CH_4, with model values near 2680 PPM versus an observed 1067 PPM. In real engine operation—featuring more turbulence and likely more stratified ignition—greater oxidation of CH_4 occurs, even at similar global equivalence ratios. CO concentrations are significantly underpredicted in the model. While the experiment measured 98 PPM of CO, the simulations produced values near zero. This underestimation again suggests insufficient peak temperatures: the CO-to-CO_2 conversion is thermally driven and requires elevated post-flame temperatures to proceed to completion. The modeled combustion process may not reach these critical temperature thresholds, again due to limitations of the zero-dimensional approach in representing spatial gradients and localized hotspots.

Overall, while the Chemkin model captures the lean combustion regime's general behavior, the low predicted NOx and CO levels, combined with the mismatched CH_4 and CO_2 values, collectively indicate a consistent underestimation or mischaracterization of combustion intensity and temperature. These findings highlight the limitations of simplified combustion models in reproducing real-world emissions and underscore the importance of proper initial conditions, particularly inlet temperature, pressure, and ignition timing—when simulating older two-stroke engines like the AJAX DPC-81.

SECTION 5.2: AJAX DPC-105 MATHEMATICAL MODEL RESULTS

The flow of this section will follow that of Section 5.1:

1. The Wiebe function parameters are found by fitting the experimental burned mass fraction profile for input into the model.
2. Next a discussion and estimation of an appropriate inlet pressure value is conducted.
3. Following this the results for a single load case at a variety of plausible equivalence ratios are shown. Each set of results is tuned to match the peak pressure and heat release rate plots as closely as possible.
4. Then, the resulting emissions of these tuned models are shown and discussed.

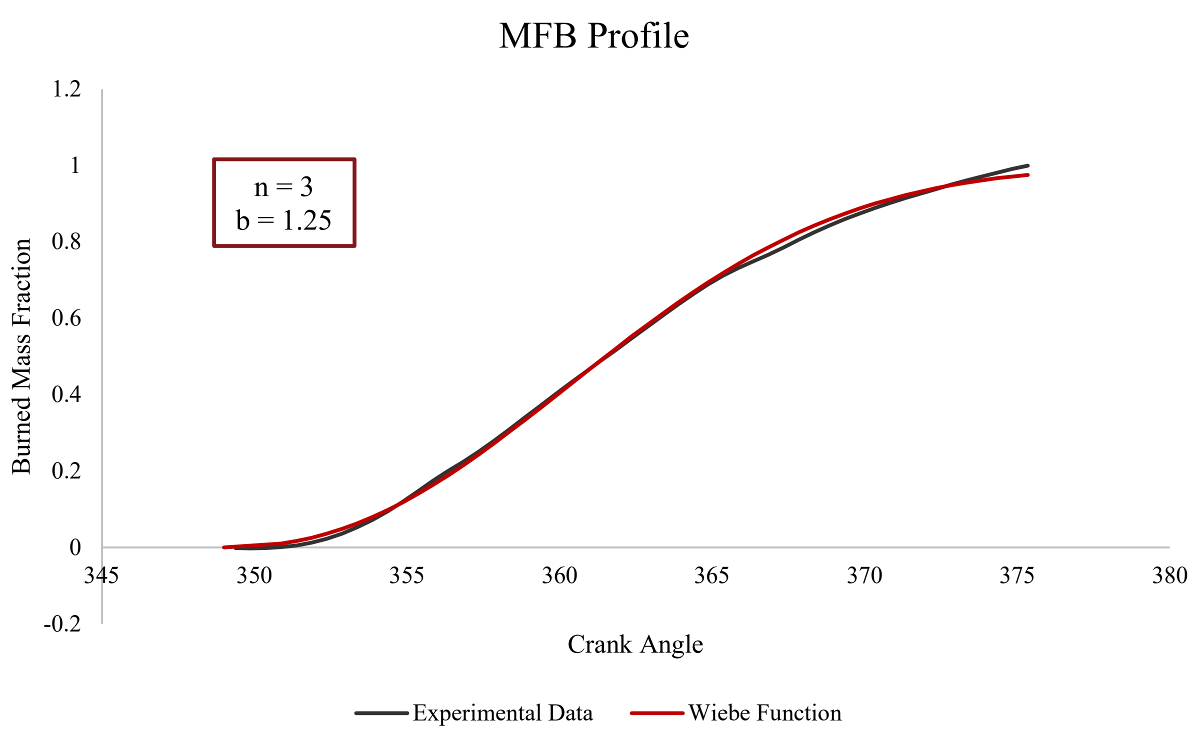


Figure 40: Experimental and Modeled Burned Mass Fraction Profile for the AJAX DPC-105

Based on the fit achieved in Figure 40, the values for n and b were selected as 3 and 1.2 respectively for input into Chemkin.

Following the procedure shown in the previous section for the DPC-81, the DPC-105's peak compression pressure prior to spark ignition is around 100 psi at 60% load, and the compression ratio is a value somewhere between 7 and 8. Given this data, the pressure at air intake closure may be estimated to be around 8 psi and is used for the starting pressure value in the Chemkin model.

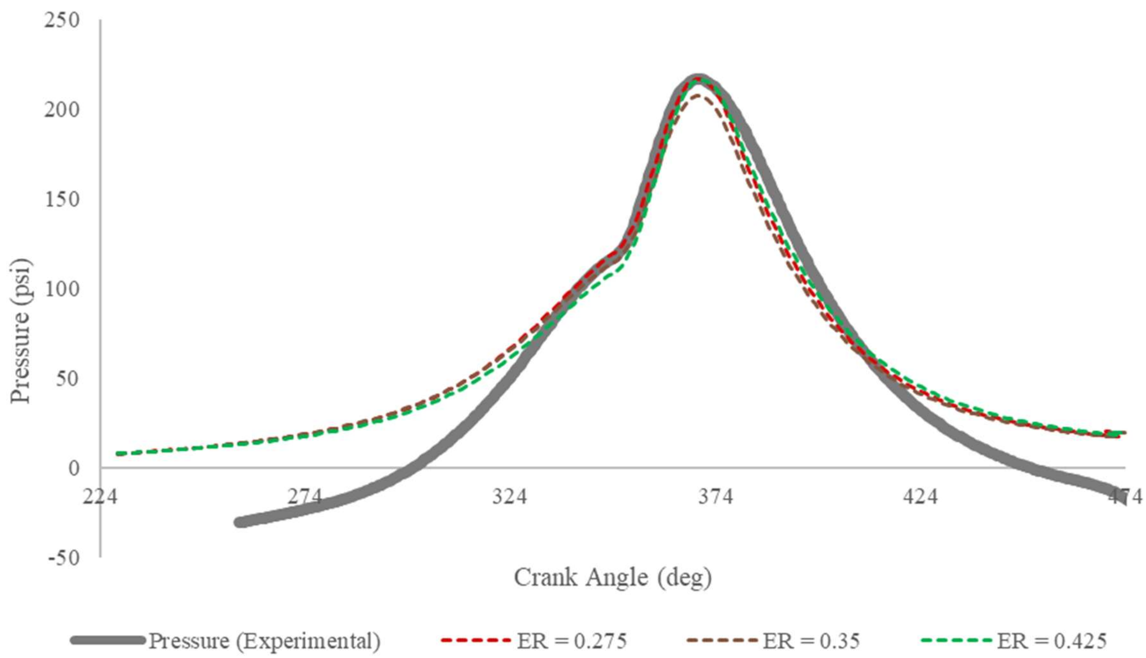


Figure 41: Chemkin SI Engine Modelling In-Cylinder Pressure Results at 60% Load

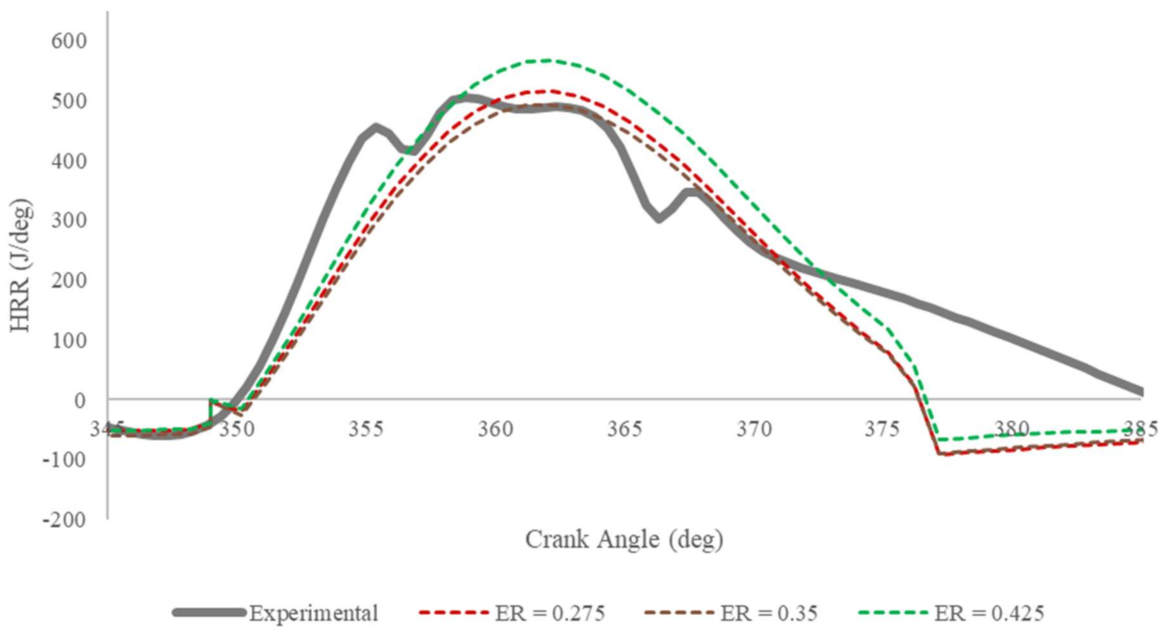


Figure 42: Chemkin SI Engine Modelling Heat Release Rate Results at 60% Load

Table 24: Chemkin SI Engine Model Tabulated Emissions Results for the AJAX DPC-105 (all conditions are on a dry, 15% O_2 basis)

ϕ	CH4 (PPM)	CO (PPM)	CO2 (%)	NOx (PPM)
0.275	2682	0.018	2.77	0.04
0.350	2682	0.019	2.77	0.02
0.425	3231	0.030	3.33	0.01
Reality	1500	85	2.61	100

From Figures 41 and 42, and Table 24, the simulation results for the AJAX DPC-105 can be compared to the case of reality. The analysis points to very similar trends to that of the AJAX DPC-81. Low peak combustion temperatures, not enough heat retained post-flame, and potentially inlet temperatures that are too low all restrict production of NOx and carbon monoxide in the two-zone model. However, there are some interesting differences between these results and the DPC-81's. Despite elevated methane in comparison to reality (indicating less complete combustion), there is elevated carbon dioxide in comparison to reality (indicating more complete combustion). These seemingly conflicting results could have some possible explanations: due to the larger geometry and lower load-case (slower speeds and lower pressures) for the AJAX DPC-105, lower quality mixing may cause localized, stratified regions of lean and rich combustion that could be leading to high amounts of intermediate species being produced that are not reaching full conversion to carbon dioxide. The lack of resolution in a 0-D model is unable to capture spatial differences.

The evaluation of the transformer-based machine learning models against the mathematical Chemkin results highlights the strength of the machine learning approach but also provides valuable insights into the complexity of the combustion process. The insight gained simply from going through the process of conducting the mathematical modeling helped to solidify understanding of which aspects were most important to consider, which helped to guide the selection of input data sources for the experimental data collection. Chapter 6 will now consolidate the key outcomes of this work and propose the future directions.

CHAPTER 6: CONCLUSION

The work presented in this thesis has focused on the development and validation of transformer-based machine learning models and mathematical models for predicting methane emissions from aging industrial natural gas engines. The specific goal of this work was to create low-cost, low-overhead virtual sensors that can be implemented into control systems for older engines still commonly used in the natural gas compression infrastructure. These models aim to provide reliable emission data without requiring expensive upgrades or major changes to existing systems.

Across both the AJAX DPC-81 and DPC-105 engines, virtual sensors were trained using real-time sensor data collected through a series of test campaigns. The predictive performance of these models was evaluated using standard metrics and then compared against a Chemkin-based two-zone combustion model developed for this purpose. The transformer-based models showed strong ability to track methane emissions under different load and operating conditions, even in the presence of noisy sensor data and incomplete system knowledge. These results suggest that machine learning, especially in the form of transformer-based time series models, can serve as a powerful tool for real-time emissions tracking.

The comparison to reduced-order combustion modeling using Chemkin showed the value in ML models. The necessity for a more complex mathematical model to capture the unique geometry and scavenging effects on these engines was apparent, though the process of tuning and evaluating the performance provided great insight into what factors affect combustion. These challenges show the direction that future engine design may start to veer towards. When creating a new engine, where the resources available include the computing power and software to run complex CFD and mathematical models, as well as advanced sensing techniques to inform and to validate these models, machine learning algorithms may be trained based on these model outputs to reduce the amount of experimental data collection required (similar to Yuan et al's paper in the literature review [57]) .

The broader message behind this work is that there are real opportunities to improve the systems we already have. Full engine replacements may not be practical or realistic in every case, especially in remote areas where reliability and uptime are the top priorities. But there is still room for improvement—room to add layers of intelligence and control that reduce emissions and improve efficiency without having to start over from scratch. If this thesis has achieved anything, it is in showing that those small steps are possible, measurable, and worth taking. This work is one small piece of a much larger effort to reduce greenhouse gas emissions and move towards more sustainable energy systems. And while the focus here has been narrow—just a few engines, a few models, the ideas behind it are part of a larger story: that even old machines can adapt, and even small improvements can matter.

REFERENCES

[1] “Climate Change in the American Mind: Beliefs & Attitudes, Spring 2024,” Center for Climate Change Communication, George Mason University, Jul. 2024. [Online]. Available: https://www.climatechangecommunication.org/all/climate-change-in-the-american-mind-beliefs-attitudes-spring-2024/#:~:text=About%20half%20of%20Americans%20(52,1.4.

[2] D. M. Kahan *et al.*, “The Tragedy of the Risk-Perception Commons: Culture Conflict, Rationality Conflict, and Climate Change,” *SSRN Electron. J.*, 2011, doi: 10.2139/ssrn.1871503.

[3] ESA Climate Office, “Land Surface Temperature.” Nov. 23, 2020. [Online]. Available: https://climate.esa.int/en/projects/land-surface-temperature/

[4] NOAA, “NOAAGlobalTemp.” Nov. 21, 2024. [Online]. Available: https://www.ncei.noaa.gov/products/land-based-station/noaa-global-temp

[5] ECMWF, “Open Data.” Nov. 24, 2022. [Online]. Available: https://www.ecmwf.int/en/forecasts/datasets/open-data

[6] J. Marder, “Global Temperature,” NASA, Jan. 2025.

[7] E. Hawkins, “WMO confirms 2024 as warmest year on record at about 1.55°C above pre-industrial level,” World Meteorological Organization, Jan. 2025. [Online]. Available: https://wmo.int/media/news/wmo-confirms-2024-warmest-year-record-about-155degc-above-pre-industrial-level

[8] M. Garrison, “GISS Surface Temperature Analysis.” Jan. 08, 2025. [Online]. Available: https://data.giss.nasa.gov/gistemp/graphs_v4/

[9] R. Jackson, “Who Discovered the Greenhouse Effect,” Royal Insititution. [Online]. Available: https://www.rigb.org/explore-science/explore/blog/who-discovered-greenhouse-effect

[10] S. Weart, *The Discovery of Global Warming*. Harvard University Press, 2008.

[11] The Editors of Encyclopaedia Britannica, “Industrial Revolution,” Banking & Business. [Online]. Available: https://www.britannica.com/event/Industrial-Revolution

[12] J. B. Heywood, *Internal Combustion Engine Fundamentals*. in 87-15251. 1988.

[13] X. Lan and R. Keeling, “NOAA GML DATA.” [Online]. Available: https://gml.noaa.gov/ccgg/trends/mlo.html

[14] The Met Office, “Met Office: Atmospheric CO2 now hitting 50% higher than pre-industrial levels,” Nature and Biodiversity. [Online]. Available: https://www.weforum.org/stories/2021/03/met-office-atmospheric-co2-industrial-levels-environment-climate-change/

[15] NASA, “The Impact of Climate Change on Natural Disasters,” Earth Observatory. [Online]. Available: https://earthobservatory.nasa.gov/features/RisingCost/rising_cost5.php

[16] J. Cohen, L. Agel, M. Barlow, C. Garfinkel, and I. White, “Linking Arctic variability and change with extreme winter weather in the United States,” *Science*, vol. 373, no. 6559, pp. 1116–1121, Sep. 2021, doi: 10.1126/science.abi9167.

[17] K. A. Mar, C. Unger, L. Walderdorff, and T. Butler, "Beyond CO2 equivalence: The impacts of methane on climate, ecosystems, and health," *Environ. Sci. Policy*, vol. 134, pp. 127–136, Aug. 2022, doi: 10.1016/j.envsci.2022.03.027.

[18] R. Cassia, M. Nocioni, N. Correa-Aragunde, and L. Lamattina, "Climate Change and the Impact of Greenhouse Gasses: CO2 and NO, Friends and Foes of Plant Oxidative Stress," *Front. Plant Sci.*, vol. 9, p. 273, Mar. 2018, doi: 10.3389/fpls.2018.00273.

[19] R. Jackson, "'Nobody ever saw anything like this before': how methane emissions are pushing the Amazon towards environmental catastrophe," *The Observer*, Aug. 17, 2024. Accessed: Mar. 02, 2025. [Online]. Available: https://www.theguardian.com/environment/article/2024/aug/17/methane-climate-crisis-amazon-peat-permafrost-vegan-heat-pumps

[20] "Natural gas timeline - Energy Kids: U.S. Energy Information Administration (EIA)." Accessed: Mar. 02, 2025. [Online]. Available: https://www.eia.gov/kids/history-of-energy/timelines/natural-gas.php

[21] "Average energy prices for the United States, regions, census divisions, and selected metropolitan areas : Midwest Information Office : U.S. Bureau of Labor Statistics," Bureau of Labor Statistics. Accessed: Mar. 02, 2025. [Online]. Available: https://www.bls.gov/regions/midwest/data/averageenergyprices_selectedareas_table.htm

[22] E. 430-R.-24-004 U.S. Environmental Protection Agency, "EPA (2024) Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2022." [Online]. Available: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-andsinks-1990-2022.

[23] J. Garthwaite, "Methane emissions are rising faster than ever," Stanford University, Sep. 2024. [Online]. Available: https://news.stanford.edu/stories/2024/09/methane-emissions-are-rising-faster-than-ever

[24] "Sources of methane emissions, 2023 – Charts – Data & Statistics," IEA. Accessed: Mar. 02, 2025. [Online]. Available: https://www.iea.org/data-and-statistics/charts/sources-of-methane-emissions-2023-2

[25] "New Data Show U.S. Oil & Gas Methane Emissions Over Four Times Higher than EPA Estimates, Eight Times Greater than Industry Target." Accessed: Feb. 27, 2025. [Online]. Available: https://www.edf.org/media/new-data-show-us-oil-gas-methane-emissions-over-four-times-higher-epa-estimates-eight-times

[26] "EPA Publishes Rule Significantly Limiting Methane Emissions from Crude Oil and Natural Gas Facilities | Sabin Center for Climate Change Law." Accessed: Mar. 02, 2025. [Online]. Available: https://climate.law.columbia.edu/content/epa-publishes-rule-significantly-limiting-methane-emissions-crude-oil-and-natural-gas

[27] "Regulatory Impact Analysis of the Standards of Performance for New, Reconstructed, and Modified Sources and Emissions Guidelines for Existing Sources: Oil and Natural Gas Sector Climate Review".

[28] "New EU Methane Regulation to reduce harmful emissions from fossil fuels in Europe and abroad." Accessed: Mar. 02, 2025. [Online]. Available: https://energy.ec.europa.eu/news/new-eu-methane-regulation-reduce-harmful-emissions-fossil-fuels-europe-and-abroad-2024-05-27_en
[29] "What Does the New EU Methane Regulation Mean for the Global LNG Supply Chain? - GHGSat." Accessed: Mar. 02, 2025. [Online]. Available: https://www.ghgsat.com/en/newsroom/what-does-the-new-eu-methane-regulation-mean-for-the-global-lng-supply-chain/
[30] "Homepage | Global Methane Pledge." Accessed: Mar. 02, 2025. [Online]. Available: https://www.globalmethanepledge.org/
[31] "Factsheet: 2024 Global Methane Pledge Ministerial." Accessed: Mar. 02, 2025. [Online]. Available: https://www.globalmethanepledge.org/news/factsheet-2024-global-methane-pledge-ministerial
[32] J. Gütschow, D. Busch, and M. Pflüger, "The PRIMAP-hist national historical emissions time series (1750-2023) v2.6." 2024. doi: https://doi.org/10.5281/zenodo.13752654.
[33] Intergovernmental Panel On Climate Change (Ipcc), Ed., "Energy Systems," in *Climate Change 2022 - Mitigation of Climate Change*, 1st ed., Cambridge University Press, 2023, pp. 613–746. doi: 10.1017/9781009157926.008.
[34] "Inventory of U.S. Greenhouse Gas Emissions and Sinks 1990-2022: Updates for Transmission Compressor Station Activity," 2024.
[35] ICF, "Impact of Electrifying Natural Gas Transmission Compression," INGAA Foundation, Inc., Jan. 2024.
[36] J. Apt, A. Newcomer, L. B. Lave, S. Douglas, and L. M. Dunn, "An Engineering-Economic Analysis of Syngas Storage".
[37] C. Campanile, "Kathy Hochul steps on the gas amid Con Edison rate hike furor." Accessed: Feb. 28, 2025. [Online]. Available: https://nypost.com/2025/02/16/us-news/kathy-hochul-steps-on-the-gas-amid-con-edison-rate-hike-furor/
[38] K. Brun, "The US Natural Gas Compression Infrastructure: Opportunities for Efficiency Improvements".
[39] M. Jaffe, "Xcel faces challenges sending natural gas to Colorado mountain towns from its Boulder County station," The Colorado Sun. Accessed: Feb. 28, 2025. [Online]. Available: http://coloradosun.com/2025/02/25/xcel-energy-natural-gas-demand-grand-summit-counties/
[40] H. A. Hassan *et al.*, "Integrated system to reduce emissions from natural gas-fired reciprocating engines," *J. Clean. Prod.*, vol. 396, p. 136544, Apr. 2023, doi: 10.1016/j.jclepro.2023.136544.
[41] A. Llp, "(74) Attorney, Agent, or Firm — Benesch, Friedlander,".
[42] H. A. Hassan, M. Hartless, P. Jha, and P. Kazempoor, "Integrated System to Reduce Emissions from Natural Gas-Fired Reciprocating Engines — Performance Assessment of

Amperometric NOx/O_2 Sensors," SAE Technical Paper, Warrendale, PA, SAE Technical Paper 2022-01–0578, Mar. 2022. doi: 10.4271/2022-01-0578.
[43] H. A. Hassan *et al.*, "Performance and emissions of natural gas/hydrogen blends in large-bore spark-ignition engines," *Int. J. Hydrog. Energy*, vol. 125, pp. 168–180, 2025, doi: https://doi.org/10.1016/j.ijhydene.2025.03.466.
[44] S. R. Gubba, B. Tamma, P. Kazempoor, T. J. Hurley, M. A. Patterson, and G. Hartman, "A novel air management system for a large bore two-stroke naturally aspirated gas engine to reduce emissions," *Int. J. Engine Res.*, vol. 22, no. 2, pp. 364–374, Feb. 2021, doi: 10.1177/1468087419871858.
[45] "GRI-Mech 3.0." Accessed: Apr. 07, 2025. [Online]. Available: http://combustion.berkeley.edu/gri-mech/version30/text30.html
[46] "ANSYS Chemkin Theory Manual 17.0 (15151), Reaction Design: San Diego, 2015."
[47] S. M. Aceves *et al.*, "A Sequential Fluid-Mechanic Chemical-Kinetic Model of Propane HCCI Combustion," presented at the SAE 2001 World Congress, Mar. 2001, pp. 2001-01–1027. doi: 10.4271/2001-01-1027.
[48] "ChemKin_Theory_PaSR.pdf." Accessed: Mar. 26, 2025. [Online]. Available: https://personal.ems.psu.edu/~radovic/ChemKin_Theory_PaSR.pdf
[49] G. M. Rassweiler and L. Withrow, "Motion Pictures of Engine Flames".
[50] A. D. Gosman, "Computer Modeling of Flow and Heat Transfer in Engines, Progress and Prospects," *Proc. Int. Symp. Diagn. Model. Combust. Reciprocating Engines*, vol. COMODIA, no. 85, pp. 529–538, 1985.
[51] M. Baratta, A. E. Catania, E. Spessa, and R. L. Liu, "Multidimensional Predictions of In-Cylinder Turbulent Flows: Contribution to the Assessment of k-ε Turbulence Model Variants for Bowl-in-Piston Engines," *J. Eng. Gas Turbines Power*, vol. 127, no. 4, pp. 883–896, Oct. 2005, doi: 10.1115/1.1852567.
[52] P. Sagaut, Ed., *Large Eddy Simulation for Incompressible Flows: An Introduction*, Third Edition. in Scientific Computation. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. doi: 10.1007/b137536.
[53] K. M. Chun and J. B. Heywood, "Estimating Heat-Release and Mass-of-Mixture Burned from Spark-Ignition Engine Pressure Data," *Combust. Sci. Technol.*, vol. 54, no. 1–6, pp. 133–143, Aug. 1987, doi: 10.1080/00102208708947049.
[54] N. Chindaprasert, E. Hassel, J. Nocke, C. Janssen, M. Schultalbers, and O. Magnor, "Prediction of CO Emissions from a Gasoline Direct Injection Engine Using CHEMKIN®," presented at the Powertrain & Fluid Systems Conference and Exhibition, Oct. 2006, pp. 2006-01–3240. doi: 10.4271/2006-01-3240.
[55] D. A. DelVescovo, D. A. Splitter, J. P. Szybist, and G. S. Jatana, "Modeling pre-spark heat release and low temperature chemistry of iso-octane in a boosted spark-ignition engine," *Combust. Flame*, vol. 212, pp. 39–52, Feb. 2020, doi: 10.1016/j.combustflame.2019.10.009.

[56] Y. Wang, “A Novel Two-Zone Thermodynamic Model for Spark-Ignition Engines Based on an Idealized Thermodynamic Process”.
[57] H. Yuan *et al.*, “Kinetic Modeling of Combustion in a Spark Ignition Engine with Water Injection,” *Fuel*, vol. 283, no. 118814, 2021, doi: https://doi.org/10.1016/j.fuel.2020.118814.
[58] M. Bertolini, D. Mezzogori, M. Neroni, and F. Zammori, “Machine Learning for industrial applications: A comprehensive literature review,” *Expert Syst. Appl.*, vol. 175, p. 114820, Aug. 2021, doi: 10.1016/j.eswa.2021.114820.
[59] R. E. Brown, “Donald O. Hebb and the Organization of Behavior: 17 years in the writing,” *Mol. Brain*, vol. 13, no. 1, p. 55, Dec. 2020, doi: 10.1186/s13041-020-00567-8.
[60] K. D. Foote, “A Brief History of Machine Learning,” DATAVERSITY. Accessed: Mar. 29, 2025. [Online]. Available: https://www.dataversity.net/a-brief-history-of-machine-learning/
[61] F. Rosenblatt, “The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain,” *Psychol. Rev.*, vol. 65, no. 6, 1958.
[62] M. Kuhn and K. Johnson, *Applied Predictive Modeling*.
[63] D. RumelHart, G. Hinton, and R. Williams, “Learning Representations by Back-Propagating Errors,” *Nature*, vol. 323, no. 9, pp. 533–536, Oct. 1986.
[64] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A Fast Learning Algorithm for Deep Belief Nets,” *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: 10.1162/neco.2006.18.7.1527.
[65] R. Collobert, S. Bengio, and J. Mariethoz, “Torch: A Modular Machine Learning Software Library,” IDIAP, Martigny, Switzerland, Research Report 02–46, Oct. 2002.
[66] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, pp. 1735–1780, 1997.
[67] A. Vaswani *et al.*, “Attention Is All You Need,” Aug. 02, 2023, *arXiv*: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.
[68] F. Ricci, M. Avana, and F. Mariani, “A Deep Learning Method for the Prediction of Pollutant Emissions from Internal Combustion Engines,” *Appl. Sci.*, vol. 14, no. 21, p. 9707, Oct. 2024, doi: 10.3390/app14219707.
[69] J. Li, Y. Yu, Y. Wang, L. Zhao, and C. He, “Prediction of Transient NOx Emission from Diesel Vehicles Based on Deep-Learning Differentiation Model with Double Noise Reduction,” *Atmosphere*, vol. 12, no. 12, p. 1702, Dec. 2021, doi: 10.3390/atmos12121702.
[70] N. B. Meresht, S. Munshi, M. Shahbakhti, and G. McTaggart-Cowan, “Developing machine learning models to predict methane and nitrogen oxide engine-out emissions from a heavy-duty natural-gas engine,” in *Progress in Canadian Mechanical Engineering. Volume 6*, Sherbrooke, Canada: Université de Sherbrooke. Faculté de génie, 2023. doi: 10.17118/11143/21169.
[71] S. Gomaa *et al.*, “Machine learning prediction of methane, nitrogen, and natural gas mixture viscosities under normal and harsh conditions,” *Sci. Rep.*, vol. 14, no. 1, p. 15155, Jul. 2024, doi: 10.1038/s41598-024-64752-8.

[72] J. Saez-Perez, P. Benlloch-Caballero, D. Tena-Gago, J. Garcia-Rodriguez, J. Calero, and Q. Wang, "Optimizing AI Transformer Models for CO2 Emission Prediction in Self-Driving Vehicles with Mobile/Multi-Access Edge Computing Support," *IEEE*, vol. 12, pp. 179689–179706, 2024.
[73] "WIN_6400_4 pg Brochure_Rebrand-030923[15].pdf." Accessed: Mar. 22, 2025. [Online]. Available: https://www.championx.com/assets/files/WIN_6400_4%20pg%20Brochure_Rebrand-030923[15].
[74] "MultiGas 2030 FTIR Gas Analyzers," MKS. Accessed: Oct. 25, 2024. [Online]. Available: https://www.mks.com/f/multigas-2030g-ftir-gas-analyzer
[75] "Operation and Maintenance Manual For AJAX Horizontal Gas Engine-Compressors." Cooper Energy Services.
[76] "pandas.DataFrame.corr — pandas 2.2.3 documentation." Accessed: Apr. 10, 2025. [Online]. Available: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html
[77] "The Illustrated Transformer – Jay Alammar – Visualizing machine learning one concept at a time." Accessed: Mar. 31, 2025. [Online]. Available: https://jalammar.github.io/illustrated-transformer/
[78] "HuberLoss — PyTorch 2.6 documentation." Accessed: Apr. 04, 2025. [Online]. Available: https://pytorch.org/docs/stable/generated/torch.nn.HuberLoss.html
[79] P. J. Huber, *Robust statistics*. in Wiley series in probability and mathematical statistics. New York: Wiley, 1981.
[80] Thermodynamics and Fluid Mechanics Group and W. J. D. Annand, "Heat Transfer in the Cylinders of Reciprocating Internal Combustion Engines," *Proc. Inst. Mech. Eng.*, vol. 177, no. 1, pp. 973–996, Jun. 1963, doi: 10.1243/pime_proc_1963_177_069_02.
[81] G. Woschni, "A Universally Applicable Equation for the Instantaneous Heat Transfer Coefficient in the Internal Combustion Engine," in *SAE Technical Paper Series*, 400 Commonwealth Drive, Warrendale, PA, United States: SAE International, Feb. 1967. doi: 10.4271/670931.
[82] S. Turns, *An Intoduction to Combustion*, 2nd ed. McGraw-Hill Higher Education, 2000.
[83] A. Jaworski, H. Kuszewski, K. Balawender, P. Woś, K. Lew, and M. Jaremcio, "Assessment of CH4 Emissions in a Compressed Natural Gas-Adapted Engine in the Context of Changes in the Equivalence Ratio".
[84] M. T. Zarrinkolah, "Methane slip reduction of conventional dual-fuel natural gas diesel engine using direct fuel injection management and alternative combustion modes," 2023.
[85] C. Baukal Jr., *The John Zink Combustion Handbook*, 1st ed. CRC Press, 2001.
[86] R. Kapaku, "Effects of Inlet Pressure and Temperature, and Fuel-Air Equivalence Ratio on Natural Gas Combustion Utilizing the GRI Mech 3.0 Chemical Kinetics Mechanism".
[87] Y. Çengel and M. Boles, *Thermodynamics: An Engineering Approach*, 8th ed. McGraw-Hill Education, 2015.