ALARM FORECASTING IN NATURAL GAS PIPELINES

by

Colin Quinn, B.S.

A Thesis submitted to the Faculty of the Graduate School,

Marquette University,

in Partial Fulfillment of the Requirements for

the Degree of Master of Science

Milwaukee, Wisconsin

May 2020

ABSTRACT
ALARM FORECASTING IN NATURAL GAS PIPELINES

Colin Quinn, B.S.

Marquette University, 2020

This thesis examines alarm forecasting methods for a natural gas production pipeline to assure the efficient transportation of high-quality natural gas. Natural gas production companies use pipelines to transport natural gas from the extraction well to a distribution point. Forecasting natural gas pipeline pressure alarms helps control room operators maintain a functioning pipeline and avoid costly down time. As gas enters the pipeline and travels to the distribution point, it is expected that the gas meets certain specifications set in place by either state law or the customer receiving the gas. If the gas meets these standards and is accepted at the distribution point, the pipeline is referred to as being in a steady-state. If the gas does not meet these standards, the production company runs the risk of being shut-in, or being unable to flow any more gas through the distribution point until the poor-quality gas is removed.

Sensors are used to collect real-time gas quality information from within the pipe, and alarms are used to alert the control operators when a threshold is exceeded. If operators fail to keep the pipeline's gas quality within an acceptable range, the company risks being shut-in or rupturing the pipeline. Predicting gas quality alarms enables operators to act earlier to avoid being shut-in and is a form of predictive maintenance.

We forecast alarms by using a 10^{th}-order autoregressive model, autoregressive model with exogenous variable, simple exponential smoothing with drift (Theta Method) and an artificial neural network with alarm thresholds. The alarm thresholds are defined by the production company and are occasionally adjusted to meet current environment conditions.

The results of the alarm forecasting method show that we accurately forecast natural gas pipeline alarms up to a 30-minute time horizon. This translates into sensitivity rates that drop from around 100% at one minute to 82.7% at a 30-minute forecast horizon. This means that at 30 minutes, we correctly forecast 82.7% of the alarms. All alarm forecasting models outperform the state-or-the-art forecaster used by the production company, with the artificial neural network performing the best.

ACKNOWLEDGEMENTS

Colin Quinn, B.S.

I would like to thank the friends who were always there for me, the professors who pushed me to become the best version of myself, and my siblings who kept me grounded. This work is a testament of the love you have shown me, for which I could not be more grateful.

I would especially like to thank Drs. Povinelli, Corliss, and Bansal for their patience, motivation, and countless hours of guidance. I cannot express how much you each have helped me to become a better researcher and human.

I could not have accomplished so much were it not for my parents, Tom and Annette. I dedicate this work to you both. Thank you for making this all possible.

TABLE OF CONTENTS

CHAPTER 1

Introduction to Natural Gas Production

This thesis examines alarm forecasting methods for a natural gas production pipeline to assure the efficient transportation of high-quality natural gas. Our goal is to help a natural gas production company transition from maintaining the pipeline reactively to carrying out predictive maintenance. Predictive maintenance is acting based on forewarning to find or mitigate degradation [1]. This thesis explores four real-time alarm prediction methods used to detect the onset of system degradation so that flow assurance is maintained within the pipeline.

Flow assurance is a term used in the hydrocarbon production industry to refer to ensuring a continuous stream of natural gas from the extraction reservoir to the distribution (sales) point [2]. As an infrastructure, natural gas pipelines are vulnerable to damaging conditions that threaten flow assurance and warrant action, resulting in a loss of profit and extra labor. To warn pipeline control operators of these damaging conditions, alarms are used to monitor the health of the pipeline and alert control operators when action is needed. Acting after an alarm has been triggered is often more costly to carry out because damage has already occurred, leading to shutdowns, loss of profit, and dangerous environments. Avoidance of unprofitable consequences can be achieved through this work on early detection of alarms within non-stationary streaming time series data. The alarm forecasting algorithms described in this work aid pipeline controllers in achieving flow assurance and allow them to conduct preventative maintenance to decrease operation cost, unsafe environments, and damage to the environment.

This work is sponsored by a natural gas production company operating in southwest Texas. To protect sensitive information being exposed from this thesis, some data, names, and particular details have been altered to meet the nondisclosure agreement made between the natural gas production company and Marquette University. In March 2018, the sponsoring production company met Marquette University's GasDay lab to discuss the possible development of predictive algorithms for early alarm detection in a natural gas production pipeline. In April 2018, a project proposal was agreed upon by both the production company and Marquette University, and development began on phase one alarm forecasting models. Since April 2018, the GasDay lab has worked closely with pipeline controllers of the production company to understand how the pipeline operates, what their alarm prediction needs are, and to implement real-time forecasting algorithms in their system controls. This research is conducted at the GasDay lab within Marquette University, Milwaukee, Wisconsin. The focus of this research is applicable to both the natural gas industry and other university research labs. The language, industry-related terms, and processes used in this thesis reflect those used at the production company sponsoring this work.

1.1 Chapter Objectives

This chapter introduces the objective of this thesis. Beginning with this project's highest level of abstraction, we provide an overview of the natural gas production process. Then, we give a closer look at production company standards and how they maintain an economical operation. This will lead to the definition of a pipeline alarm and how alarms are used to help pipeline operators reactively service the pipeline. This introduces the forecasted alarm and the benefits of preventative maintenance in the production process. Finally, we give a brief survey of this research project as a whole and a summary of the remaining chapters.

1.2 Introduction to Natural Gas Production Pipelines

Natural gas production companies use pipelines to transport natural gas from point *A* to point *B*. Point *A* is where the gas is extracted from the earth, and point *B* is where the gas is sold to distributors. Figure 1.1 depicts this transfer.

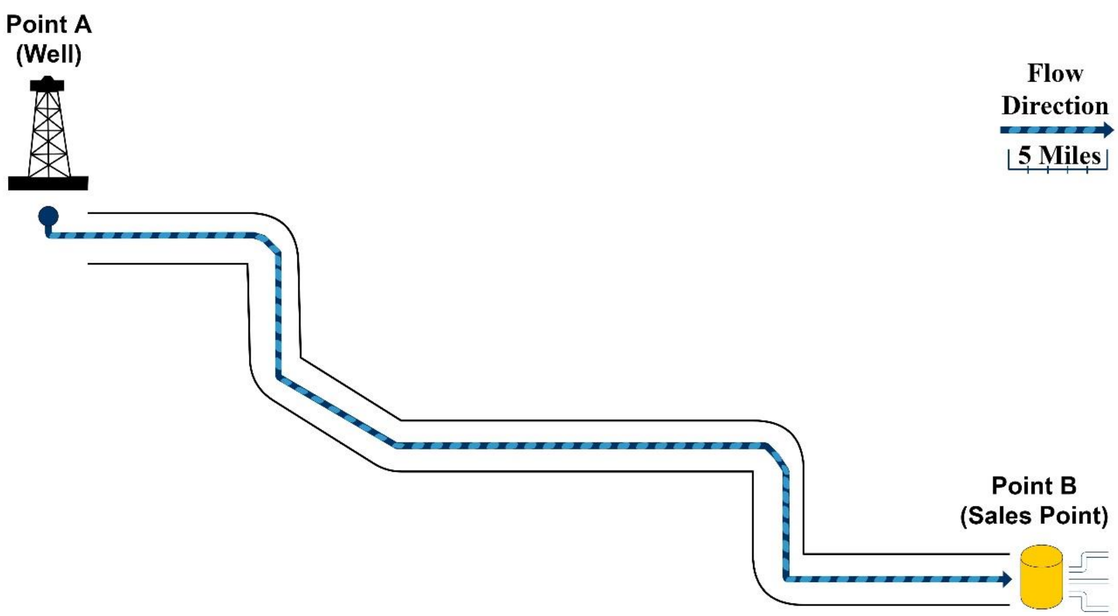

Figure 1.1: A natural gas pipeline transporting gas from point A to point B

Natural gas production companies strive to complete this task as efficiently and cost-effectively as possible. A production company operating at full capacity simultaneously extracts gas from the ground at point A, transports it through the pipeline, and sells it at point B twenty-four hours a day, seven days a week [3]. The pipeline connecting these two points plays a critical role in this operation, as its throughput determines whether the production company's revenue outweighs the cost of operation.

A natural gas production pipeline requires billions of dollars of infrastructure and highly skilled people to operate correctly [4], [5]. There are numerous moving parts in a natural gas production company that are interdependent. The profit margin of a production company depends on the success of transporting gas from A to B, and can vary widely from day to day. Until the last few years, the state-of-the-art solution to ensure reliable production and transportation of natural gas was with human pipeline operators and supervisory control and data acquisition (SCADA) systems [6]. Although pipeline operators are experts in the field of natural gas production, and there have been large technological advancements in SCADA software and pipeline monitoring [7], the growing demand for natural gas as an energy source requires new tools to help automate the production process.

The extraction, processing, and transportation of the natural gas is called the upstream operation of the natural gas industry [5], [8]. This research concentrates on the upstream operation and the flow assurance of a production pipeline (successful transportation of natural gas through a pipeline). The goal of this project is to enhance the current error-prone processes of upstream operations with the modern advancements of data analysis and prediction. Problems can occur in the pipeline that can slow or stop the flow of gas. Alarms are used to notify pipeline control operators that a problem is occurring and that action is needed. Alarm forecasting allows the pipeline operators to act before an alarm is triggered, which minimizes downtime and reduces the number of potential errors in day-to-day operations. If a pipeline operator can detect a problem that will slow production with a forecasted alarm, there is less chance of the operation slowing or halting. Improving this upstream operation returns a larger amount of gas being sold to the distribution vendors, increasing profits and protecting equipment from long-term damage.

There are many opportunities for error in upstream operations of a production company. The most common errors on which this thesis focuses are found in the quality and characteristics of the gas in the pipeline. The quality of gas refers to the chemical makeup of the natural gas, while the gas's pressure, heat content, and flow rates within the pipeline represent the gas's characteristics. The alarms that alert a pipeline controller to a problem correspond to these conditions. To understand what a natural gas pipeline alarm is, how it is used in the production process, and the potential benefit of a forecasted alarm, the next section presents the production process.

1.3 Natural Gas Production

The natural gas production procedure discussed in this section provides a simplified version summarized in three steps. This overview of production sets up the remaining sections in this chapter and represents the highest level of abstraction needed to recognize the contribution of this thesis. Figure 1.2 can be used as a visual representation of each step in the production procedure.

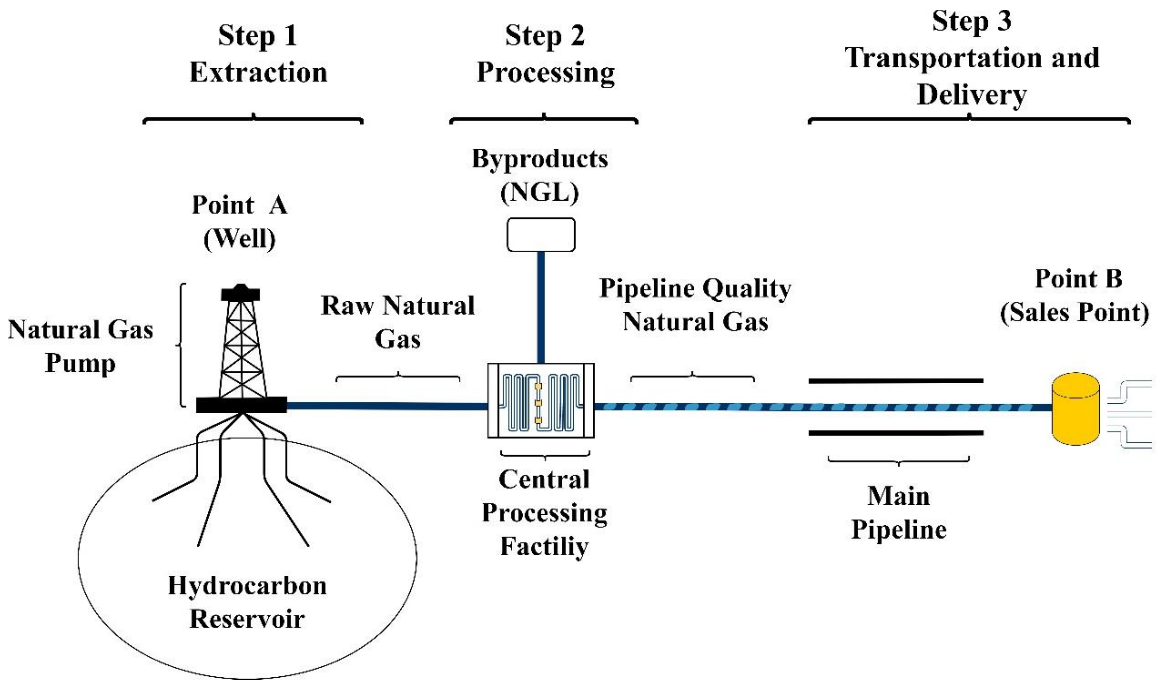

Figure 1.2: Steps in the production procedure

The first step of the upstream operation is to extract natural gas from the earth. Natural gas is accessed using an extraction well, an aperture encased in concrete and steel used to access deposits of natural gas deep within the earth [6]. There are a number of drilling techniques that have made natural gas and other hydrocarbon resource extraction more efficient over the last decade [9], [10], [11], [12]. These techniques will not be covered in this work. However, the advancements in hydraulic fracturing and horizontal drilling have made the U.S. the world's leading natural gas producer at 30 trillion cubic feet in 2018, 31% of the total U.S. primary energy consumption [12]. At the top of these extraction wells, pumps are used to extract the gas slowly from small pockets in rock formations or other hydrocarbon reservoirs [13]. Once extracted, a gathering system made up of several small-diameter lines take the extracted natural gas from the wellhead to a central processing facility [14]. The natural gas in the underground reservoir and in the gathering lines is known as raw natural gas. The chemical makeup of raw natural gas differs from the quality of gas allowed in the main pipeline. Raw natural gas contains impurities that must be removed before being pressurized and injected into the main line. Processing plants are used to collect the gas from the low-pressure gathering lines, process the gas to pipeline quality, and inject it into the pipeline.

The second step in natural gas production is to process the raw natural gas into pipeline quality gas. Natural gas is composed of combustible hydrocarbons, gases, water, and oil [15]. Processing raw natural gas involves separating non-methane hydrocarbons and other impurities from the gas [4]. The plants equipped to do this are known as central processing facilities (CPF) and are located near wells along the pipeline. The raw natural gas drawn from the wellhead consists of both heavy and light hydrocarbons [4]. In general, processing raw natural gas removes water and the heavy hydrocarbons (ethane, propane, butane, and pentane) to achieve a quality considered acceptable to transport in a pipeline [3], [5]. Regulations set by either state law or by the customer receiving the gas require the gas to meet certain specifications [6]. Specifics of what hydrocarbons are removed from the gas specific to this project will be discussed in the following section on natural gas processing.

The byproducts created while processing natural gas are also valuable and are collected for future sale. There are multiple side-operations taking place during this refining process to capture profitable substances [5]. Byproducts such as liquefied natural gas (LNG) can be separated from the hydrocarbon stream and sold. Although out of the scope of this project, such byproducts can be equally valuable as pipeline quality natural gas, and entirely different processes are carried out to retrieve them [16]. Not all production companies process natural gas the same, as equipment varies from pipeline to pipeline and depends on the size of the operation.

The final step of the production process is to flow the gas down the pipeline to be sold at a distribution point. In some instances, a pipeline may have several CPF's operating in parallel, all injecting natural gas into the main pipe simultaneously. This means that the gas arriving at the distribution point is really a combination of gas from several wells from further up the line. The production pipeline operates in a similar way, with Figure 1.3 depicting a version of this coalescent gas stream.

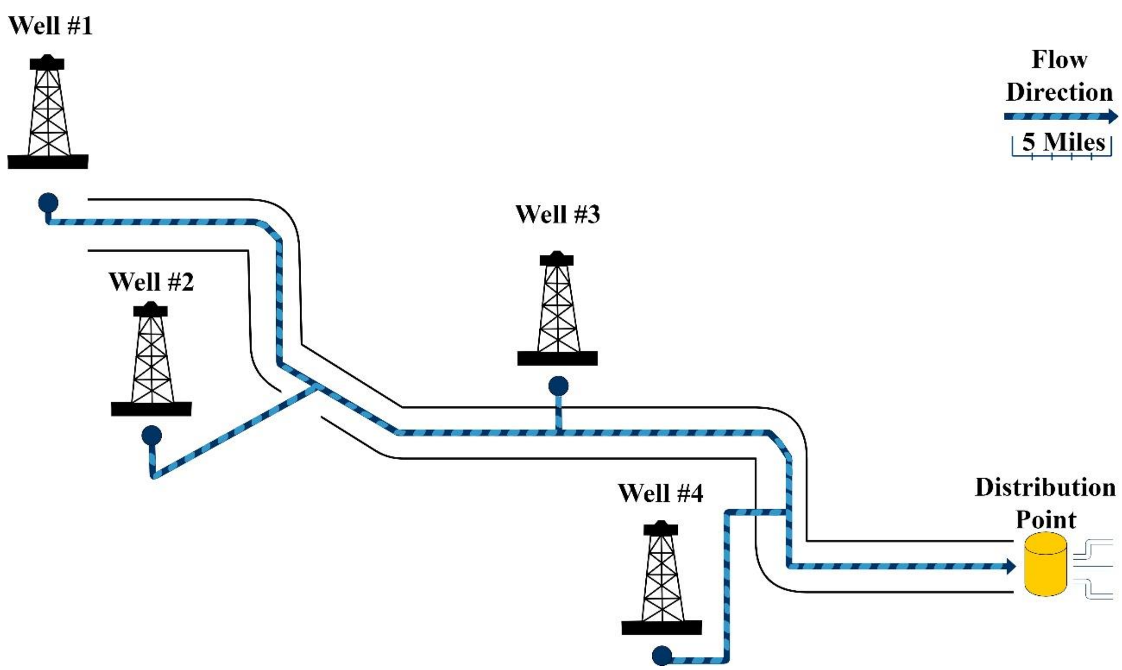

Figure 1.3: Four wells simultaneously injecting gas into the line as the gas flows to the distribution point

No two natural gas wells are identical, and the chemical make-up of each well results in different quality gas being injected post-processing [5]. Even the gas extracted in the morning can be different from the gas taken from the same well the day before [6]. This is an important concept to this work, as the quality of gas being received at the distribution point determines whether the gas will be purchased. The quality of gas can fluctuate, which is why the pipeline control room monitors the condition of the natural gas within the pipeline to coordinate its processing before it reaches the distribution point. After the three steps described in this section, the production company hopes to have a high-quality natural gas that the distributors will purchase.

1.4 Composition of Natural Gas and Production Company Standards

Now that a general overview of the natural gas production process has been presented, this section describes the alarm-triggering situations that arise in daily operations. As previously described, the errors that occur in the production process normally concern the quality of natural gas being received at the distribution point. This section will begin by giving a brief introduction to natural gas found in the U.S.A. and then move into the specifics of the gas being produced from the wells along the sponsoring production company's pipeline. The company standards will be discussed in relation to the distributor's needs, which will transition to alarm forecasting**.**

Extraction wells found in the U.S.A. produce one of two types of natural gas: conventional or nonconventional gas (Figure 1.4). Conventional gas can be extracted with traditional (vertical) drilling techniques and can be found in geological formations that are generally more accessible and straightforward to develop [6], [17]. Conventional natural gas is either associated or non-associated with crude oil. Associated gas is found in oil wells, where the gas can be separate from the oil (free gas) or dissolved into the crude oil (dissolved gas) [6]. If the well is producing dissolved gas, the oil must be separated from the gas at the wellhead and thoroughly processed before transit. The oil and other byproducts from the processing are captured and sold. Non-associated gas wells produce gas that is mixed with little to no crude oil less and requires less post-extraction processing.

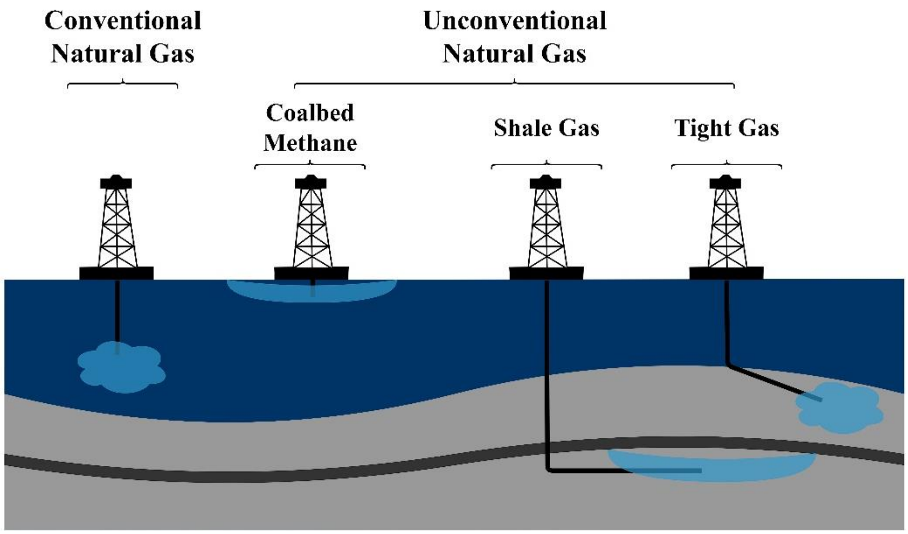

Figure 1.4: Visual representation of conventional and unconventional gas wells accessing natural gas formations

Unconventional gas is held in formations that are accessed with newer drilling techniques and only recently have proved an economically viable alternative to conventional gas wells [18]. Unconventional gas is found in reservoirs with low permeability, meaning the gas is trapped in the formation and is unable to flow through the tight sands that hold it [19]. Coalbed methane, tight gas, and shale gas are non-associated and often extracted from these formations with vertical and horizontal drilling. Horizontal drilling and hydraulic fracking make natural gas one of the most abundant resources in the U.S. The gas measured in this project comes from both conventional and unconventional non-associated gas wells. Operating over 3 million acres, the production company operates different wells, and each well produces a different type of gas.

The natural gas resource for this project is a part of the Permian Basin, located in southwest Texas, primarily in Reeves County (Figure 1.5). There, a pipeline spanning approximately 70 miles across the basin carries gas from extraction wells that generate the pressure and gas quality signals used in this work. This work will concentrate on four of the wells along the pipeline flowing towards a single distribution point, similar as to what Figure 1.3 depicts. Of these four wells, all are producing non-associated natural gas but vary in chemical makeup. To distinguish between the different types of gas at these wells, the gas is further classified into either wet (rich) or dry (lean) gas.

Figure 1.5: The Permian Basin located in Reeves Country, Texas (highlighted in blue)

The difference between wet and dry natural gas is the amount of recoverable hydrocarbons present in the gas [5]. The terms wet and dry natural gas are often used in the production pipeline's control room to describe the quality of gas in the pipeline. If the line is heavy with wet gas, it is more likely that an alarm will be triggered and errors will occur. If dry gas is flowing through the line, the control room is comfortable with current operations and may even try to increase the pipe's throughput. Understanding the differences between these two types of gas provides intuition for the problems that occur in a pipeline, thus a formal definition of natural gas' chemical makeup is provided.

Natural gas is a naturally occurring combustible hydrocarbon gas. The typical chemical composition of natural gas consists of primarily methane (CH_4) and less prominent hydrocarbons. The less prominent hydrocarbons — Ethane, Propane, Butane, etc. – are impurities and processed out before transportation. Table 1.1 shows the typical make-up of natural gas.

The more methane present in the gas, the less processing is needed. Methane-dense gas falls into the dry gas category and is more valuable than its richer counterpart. Gas with high levels of methane already resembles pipeline quality gas and can be produced at a faster rate. By the time natural gas is used for residential or commercial purpose, the composition of the gas is almost pure methane [6]. Refineries are used to achieve this near 100% methane composition in the downstream sector of the industry. Despite raw natural gas consisting of 70-90% methane upon extraction, it must still be processed to be considered pipeline quality dry gas. Pipeline quality gas differs from production company to production company. However, it can be assumed that the gas flowing through the main line is as methane-rich as possible.

Table 1.1: Typical chemical composition of natural gas

Hydrocarbon	**Chemical Formula**	**Percent**
Methane	CH_4	70-90%
Ethane	C_2H_6	0-20%
Propane	C_3H_8	
Butane	C_4H_{10}	
Carbon Dioxide	CO_2	0-8%
Oxygen	O_2	0-0.2%
Nitrogen	N_2	0-5%
Hydrogen Sulfide	H_2S	0-5%
Rare gases	A, He, Ne, Xe	trace

The impurities processed out of raw natural gas include water, ethane, propane, butane, and pentanes. These associated hydrocarbons are the natural gas liquids previously mentioned as byproducts of the processing procedure and can consist of 0-20% of the original chemical makeup. The more liquid content present in raw natural gas, the richer the gas is. Rich gas, synonymous to wet gas, is removed to create a product that has a higher sales value [18]. This removal creates lean gas, or dry gas, consisting of the lighter hydrocarbons. Liquid content is one of the main classifiers of natural gas, with rich gas indicating that a more rigorous processing procedure is needed, and lean gas indicating the gas already has a low liquid content and ultimately less processing is needed. The heavier components of gas, such as ethane, propane, and butane are the main contributors to the liquid content.

Pipeline quality gas is defined by regulations and customer needs. A number of impurities can affect the final product gas being delivered to a distribution point [5]. Although gas being delivered is considered pipeline quality, impurities can be present that effect the final consistency received at the distribution point. At the distribution point, other companies can choose to flow gas from the production company's pipeline into their own. If this exchange takes place, the transaction has been made, and the natural gas has been sold. In this transaction, the quality of gas must meet specific conditions to be allowed to flow into the purchasing company's pipe. As previously stated, regulations are set by either state law or by the customer receiving the gas that the gas meets certain specifications before the customer is allowed to accept the gas [20]. This decision of which gas to accept is made by a careful monitoring of the natural gas's quality in a pipeline control room.

In the production company's control room, the pipeline is monitored and remotely controlled. These controllers are the ones maintaining flow assurance for the pipeline and are the first to act when a problem is present in the system. Different gas qualities are tracked and presented to these operators to ensure high-quality gas is being received at the distribution point. It is because of this that such a detailed explanation of raw natural gas processing has been given thus far. If the buyer at the distribution point sees gas arriving from the pipeline that contains an unacceptable quality, they have the option to reject the gas from flowing into their pipeline and shutting the valve allowing the flow of gas. To counteract the potential problem of being unable to flow gas to the distributor, production companies use alarms to warn pipeline controllers that issues are present.

1.5 Natural Gas Pipeline Alarms

This section defines what it means for the production company's pipeline to be shut in, how alarms are used in the control room, and the potential for forecasted alarms. The actual alarm thresholds specific to this project are presented in Chapter 3 so that they can be visualized with the time series to which they relate. In June 2018 and May 2019, we visited the production company to learn the specifics of their pipeline operation. The information in this section comes from what we learned during these meeting and the remote meetings throughout the duration of this project.

If a distributor chooses to close the valve that allows the flow of gas from the production pipeline into their own, this is called being shut in. Avoiding being shut in is the goal of the alarm forecasting algorithms developed in this thesis. Being shut in triggers a chain of events that is extremely costly and time-consuming for any production company to fix. Once a distributor decides to shut in the pipeline, gas from the extraction wells continue to flow down the pipeline and begin to pack the line with bad gas, gas that contains a quality that exceeds contractual thresholds and is deemed unacceptable to the distributor. While more bad gas builds up near the distribution point, the production control room operators instruct the processing center operators to pull out of the line, or to stop injecting more gas into the pipe. For the pipeline to become functional again, the poor-quality gas must either be diffused with gas further down the line or flared from the system entirely.

Diffusing the low-quality gas is a technique practiced by the production company that is usually the first attempt at resolving the issue of being shut in. Diffusing the line involves slowly mixing the low-quality gas with high-quality gas in an attempt to achieve a quality of gas acceptable to the distributor. Diffusing the gas is preferred over flaring the system, as the gas already in the line does not need to be removed. However, flaring the gas can take a long time to complete. Depending on how packed the line is, it is sometimes more economically sensible to flare the gas instead of diffusing.

Flaring the pipe involves the removal of all gas from a segment of pipe. Depending on how much bad gas in packed into the line, the flared segment of pipe can span back from the distribution point to the majority of the main pipeline. This technique is faster than diffusing the gas; however, it can cost anywhere from $15,000-$25,000 an hour, plus the operation costs to extract, process, and transport that gas in the pipe. Due to these large penalties of being shut in, many precautions are made to avoid being shut in.

1.6 Natural Gas Processing and Transportation

Natural gas often is found in remote places far from a local market [6]. For the gas to be sold, it must be transported from its well of origin to a distributor. For decades, pipelines have been the most secure, reliable, and economical tool for this job [3], [6], [21], [22]. However, because raw natural gas contains impurities that must be removed from transportation, the gas must be processed before it is injected into the pipeline. This section breaks down the process of turning raw natural gas into pipeline quality gas, and how each process effects the signals used to forecast pipeline alarms.

Flow assurance is a term used in the production industry that refers to ensuring the flow of hydrocarbons from the extraction well to the distribution sales point [2]. This section will be concentrating on mid-stream flow assurance issues such as gas hydrate formations, corrosion, erosion, and severe slugging within the pipeline. Each of these flow assurance risks has the potential to slow or stop the flow of gas in the production process. These issues are prevented by processing the impurities out of natural gas before it is injected into the line. Within the pipe, the injected gas is monitored by sensors, which produce the signals used to represent the real-time quality of gas. These signals are used in the alarm prediction algorithm described in Section 4.2, and understanding how the signal reacts to different components of the processing procedure is domain knowledge needed to make accurate forecasts.

Consequences of flowing poor-quality gas through the pipeline fall into two categories. The first category involves the flow assurance risks that affect the design and integrity of the pipeline (hydrate formation, corrosion, etc.). The second consequence stems from marketing/federal law regulations. The production company is held by a contractual agreement to deliver a certain amount of high-quality gas to the distribution point. CPF’s are used to control the quality of gas and the amount flown to the distribution sales point. If this contract is not met, the company could be subject to fines and possibly being shut in. Requirements are placed on the hydrocarbons listed in Table 1.1 as well as internal pipe pressure (measured in pounds per square inch) and the heat content (measured in BTU) of the gas. In conjunction with distributor contracts, the production company must meet federal regulations.

If the poor-quality gas enters the U.S. nation's natural gas transportation network, the company providing the gas can be subject to increased tariffs as the poor quality gas can affect the overall network [23]. While the definition of pipeline quality gas varies from different organizations, the U.S. Energy Information Administration provides general guidelines of the characteristics of pipeline quality gas. The general specifications are:

1) *The gas must be within a specific BTU range (1035 BTU per cubic foot, +/- 50 BTU)*
2) *Be delivered at a specified hydrocarbon dew point temperature level (below which any vaporized gas liquid in the mix will tend to condense at pipeline pressure)*
3) *Contain no more than trace amounts of elements such as hydrogen sulfide, carbon dioxide, nitrogen, water vapor, and processing oxygen.*
4) *Be free of particulate solids and liquid water that could be detrimental to the pipeline or its ancillary operating equipment.*

List 1.1: U.S. Energy Administration's Generalized Pipeline Quality Gas [23]

Depending on the location of the well, these guidelines become more specific to the gas being produced in that area [8]. In this work, the pipeline quality gas specifications are set by the distributor at the sales point, and the processed gas is well within the federal standards. The specific pipeline quality gas is shown in Table 1.2.

Table 1.2: Pipeline Quality Gas Requirements for the Production Pipeline

Quality	**Upper Limit**	**Waiver Dependent**
Moisture (H_2O)	≤ 7 lbs	NO
Carbone Dioxide (CO_2)	≤ 2%	YES
Heat Content (BTU)	≤ 1100 BTU	YES
Hydrogen Sulfide (H_2S)	≤ 5 PPM	NO
Maximum Allowable Operating Pressure (MAOP)	≤ 1400 psi	NO

As an example of this list's generality, the BTU content limit specified for the production pipeline used in this work is higher than the U.S. Energy Administration's limits. This is allowed due to the rating at which the production company operates.

The waiver depended column of Table 1.2 refers to the contract between the production company and distributor. In some instances, the production company or the distributor would like to flow gas outside the limits stated in Table 1.2, so a waiver can be activated. Reasons for activating a waiver usually has to do with a gas quality problem farther down the line. For example, sometimes it is necessary to enrich the gas's heat content, so heavier hydrocarbons may be blended with the gas to offset the low BTU levels [23]. The qualities that may not allow to be altered with a waiver are the qualities that threaten flow assurance and the integrity of the pipeline. Flow assurance for this project is controlled by the central processing facilities located along the pipeline and is referred to as field processing [14]. The role of a CPF is to upgrade poor-quality gas to pipeline quality. Figure 1.6 shows where field processing CPF's typically are located in the production process.

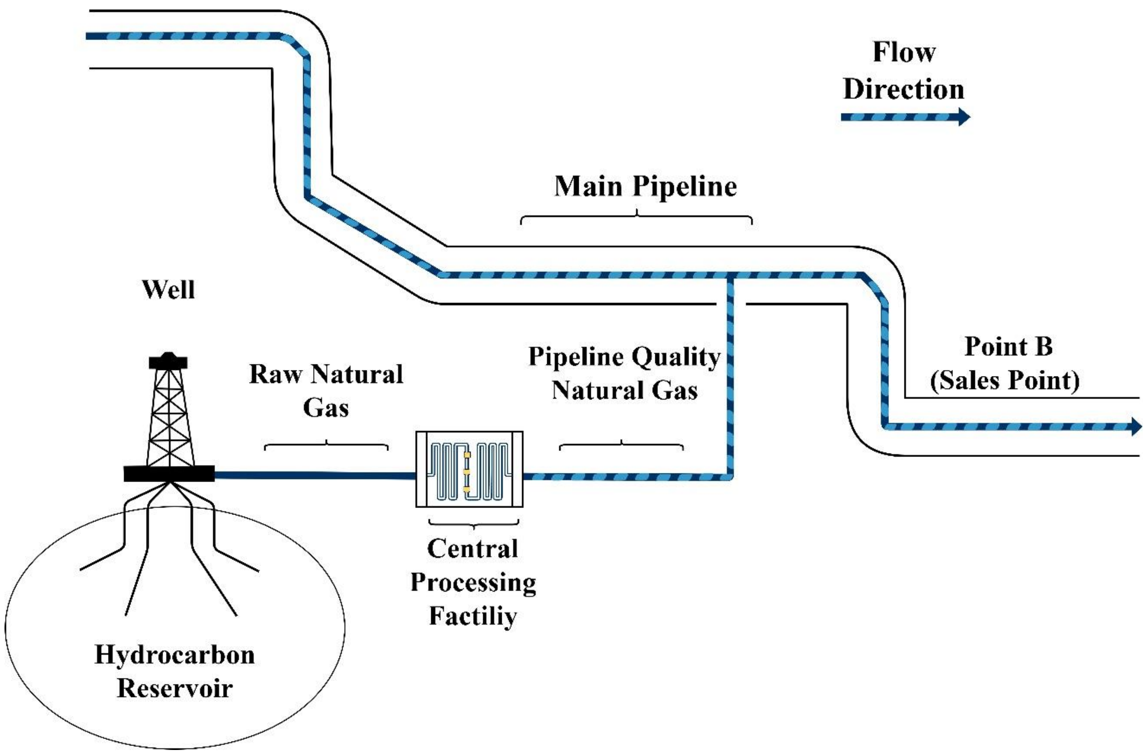

Figure 1.6: The typical location of a field CPF processing raw natural gas into pipeline quality gas

Generally, processing gas involves the separation of non-methane hydrocarbons and other fluids from methane. This is a several step process [5], [6], [14], [23] that begins at the extraction wellhead where the associated or dissolved natural gas is separated from the crude oil. One of the main objectives of a natural gas processing is to remove the high concentration of carbon dioxide from sour gas and other sulfur components to meet stringent emission standards [6]. This process begins with a conventional separator using gravity and compression to heat and cool the gas, which allows the heavier oil and gas to sink below the lighter hydrocarbons. Then refrigeration units are used to dehydrate the gas stream. This removes water and is vital in avoiding the formation of gas hydrates in the main pipe during transportation. Once much of the water has been removed, the gas is subjected to contaminate removal and methane separation. Contaminate removal removes the hydrogen sulfide, carbon dioxide, water vapor, helium, and oxygen from the gas. This is achieved with amine gas treating, where the gas is sweetened using aqueous solutions of alkylamines [24]. To separate the NGL from the methane, absorptive oil is mixed with the gas stream. The absorptive oil soaks up the NGLs (ethane, propane, butane, etc.), while methane stays in gaseous form. The NGL and methane are separated with extreme cold temperatures, and the methane-dense gas rises above the sinking NGLs.

After altering the chemical makeup of the raw natural gas, it is compressed and injected into the pipeline. Compressor stations are critical to the production process and are responsible for the flow of gas through the pipe despite elevation changes, friction, and long distances. As the gas is compressed, heat is generated. With every 100 pounds per square inch (psi) the gas is compressed, the heat content of the gas increases approximately seven to eight BTU per cubic foot [25]. To counteract this, cooling units are used so that by the time the gas is injected into the pipe, the gas is at a temperature that the pipeline operators deem acceptable. Once the raw natural gas has been compressed, the pressure generated by the compressor units forces the gas to flow in the direction of the distribution point.

The signals used to forecast pipeline alarms reflect this process. Depending on how each CPF is operating, the signals will change to reflect the current status of the system. For example, if a refrigeration unit fails, it is likely that the gas being injected into the main line by that CPF is heavy with water. Subsequently, as the wet gas flows towards the distribution point, it is likely that the H_2O signal is increasing towards or exceeding an alarm threshold. By the time the pipeline controller is alerted of the triggered H_2O alarm, the pipeline may already be shut in. Similarly, if pipeline controllers are made aware of a marketing waiver to increase the heat content of the stream, they may instruct a CPF to adjust their processing procedure to output ‘hotter’ gas. This changes the behavior of the BTU and other gas quality signals.

One error in the production process at a single CPF can cause catastrophic failures through the entire pipeline, engendering shutdowns, loss of profit, and dangerous environments. As demand for natural gas as a clean burning fuel continues to grow, the production industry is being pushed to operate at higher pressures [26]. Operating at higher pressures means more pipeline throughput and requires more gas processing. Having access to the latest technologies will provide efficient and resource-saving improvements to production companies.

1.7 Contribution of Thesis

The use of machine learning and artificial intelligence in the energy industry has proven itself to be beneficial and effective. However, many areas of this industry have yet to be explored [27]. There is little work published in the field of natural gas production pipeline alarm predicting. Based on an extensive literature review (Chapter 2), this is the first published algorithm to predict natural gas pipeline alarms. This is due to several reasons: First, this problem is specific to a single natural gas production company, and second, until this point, pipeline operators have been the main source of detecting issues in the system. Although different production companies have different means of production, this work can provide the foundation on which algorithms are developed to aid their systems.

1.8 Outline of Remaining Chapters

The remaining portion of this thesis begins with a review in Chapter 2 of current literature and work done in the field of energy forecasting and time series analysis. In Chapter 3, we will introduce the data used in this work, anomaly detection and imputation, and the framework for real-time alarm forecasting. Chapter 4 continues with the implementation in several forecasting techniques: 10-order autoregressive model, 10-order autoregressive model with exogenous variables, simple exponential smoothing with drift model (Theta method), and an artificial Neural Network. Chapter 5 presents discussion, interpretation, and comparison of the experimental results. Finally, Chapter 6 covers final thoughts such as a project summary, future work, and final words.

CHAPTER 2

Project Relevance and Literature Review

2.1 Chapter Objectives

This chapter presents the history and context of this work, and it analyzes, interprets, and critically evaluates the existing literature on alarm forecasting in natural gas pipelines. Beginning with this project's relevance and incentive, the following sections present the reader with a background of natural gas production, pipeline technology, and the growing need for preventative maintenance. Mathematical work will be discussed involving modern time series, regression, and real-time error detection applications that are reviewed and linked to alarm forecasting in the natural gas field.

2.2 Project Relevance and a Change in Natural Gas Production

To meet the growing demand for fossil fuels, natural gas production companies need to embrace new technologies and develop more capable processes to maintain flow assurance while simultaneously increasing production. This idea of increasing production is not isolated to just the natural gas industry, but all the energy production industries. The harnessing of energy through the use of new technologies has fueled the U.S. economy since the industrial revolution [28]. As humans evolve, more energy is needed to meet our needs. Hence, energy production has transformed over time [29]. The first known practical use of natural gas was in 500 B.C., when the Chinese used naturally occurring gas to boil sea water, producing salt. They achieved this using hollowed bamboo trunks to capture the gas seeping from the earth's surface, unknowingly making the world’s first natural gas pipeline [6], [30]. Although the technology has changed, we still use the same fundamental idea today.

From a few bamboo trunks to the three million miles of carbon-steel pipe that spans the United States today, the goal of a natural gas pipeline is still to transport gas from its extraction point to its place-of-use [31]. Clean burning, abundant, and versatile, natural gas consumption in the United States has doubled since the 1980s and reached an all-time high in 2018 [32], [33]. The U.S. Energy Information Administration reports a predicted 5% rise in natural gas usage by 2050, as well as a 11% decrease in coal and a 7% decrease in nuclear energy [34]. The U.S. is the world’s leader in natural gas production and consumption, making this energy industry a significant part of the economy with 31% of the total U.S. energy consumption being supplied from natural gas [12]. This continuous increase in natural gas use puts pressure on production companies to meet the demand, extracting and flowing more natural gas through their systems than ever before.

This surge in the production industry has come with a cost. Despite natural gas emitting less global warming emissions than coal or oil, carbon dioxide and other heat-trapping gasses are still released when natural gas is combusted [29], [35]. Drilling, extracting, and transporting natural gas introduces the possibility of methane leakage, an even more detrimental occurrence than carbon dioxide contributing to greenhouse gas emissions [35]. The natural gas production industry accounted for about a third of the methane released into the atmosphere in 2018 [6], [32], [36]. Hence, stricter environmental regulations force industry compliance, resulting in the production industry’s adoption of new technologies to reduce production errors. Strict emission standards enforced by the U.S. government keep the increasing demand of North America's natural gas production and transportation in check. Production companies can produce as much gas as they want, but are required to maintain certain standards or otherwise be subject to penalties discussed in Chapter 1 [23]. While the environmental impact of any hydrogen-based energy source's production is an unfortunate trade off, other repercussions from the production process is pressuring the natural gas industry to find alternative solutions for safely keeping up with demand.

Over the last 40 years, pipeline accidents have killed more than 500 people, injured 4000 more, and cost nearly seven billion dollars in property damage across the U.S. [21]. Of these figures, natural gas production was specifically responsible for 24 deaths, 99 injures, and over a billion dollars' worth of damage from 2010 to 2018 alone [37]. Each accident that occurs is a blemish on the natural gas industry, rightly bring up questions of safety and putting pressure on pipeline companies to make changes. Pipeline failure can occur for many reasons. The most common cause is that pipelines are becoming older and may not be maintained over time [21]. With the introduction of the Natural Gas Pipeline Safety Act of 1968 [38], programs such as the Pipeline and Hazardous Material Safety Administration [39] are actively enforcing federal regulations and industry standards to force outdated production pipelines to use the new technologies available in production.

Despite the negative environmental impacts and safety faults of the natural gas industry, natural gas is the most energy efficient and cleanest-burning fossil fuel [29], [40], [41], [42]. Pipelines are the most cost effective and safest ways to move natural gas over long distances [3], [6], [21], [22]. With the affordability of today's digital technologies and research in the field of natural gas production, new ways to maintain, protect, and control pipelines are becoming more accessible than ever before. Smarter production leads to more volume being produced, fewer errors, and responsible care for the earth; all while supplying energy to those who rely on it.

The production company sponsoring this work is aware of the repercussions that result from an inefficient and unsafe production process. Situations that slow or stop the flow a gas cost the production company significant amounts of time and money to correct. Therefore, their incentive for sponsoring this pipeline alarm forecasting project is to remain on the leading edge of natural gas production technology and to reduce the cost of reactively maintaining the pipeline. Searching for ways to maximize the throughput of the pipeline projects such as this are investments for the future of the production company and represent their first steps in predictively maintaining the pipeline.

2.3 Types of Maintenance and Natural Gas Production Technology

Without the appropriate technology to assist in the increasing demand of North America's natural gas consumption, outdated and under-maintained production pipelines can result in serious financial loss for production companies and ecological disasters [43]. Once shut-in, the production company suffers a loss as pipeline workers try to identify and correct the disruption to flow assurance. This reactive process is a fault in the production company's operational efficiency, and new forms of maintenance have been introduced in an attempt to reduce downtime and expended resources. Understanding these forms of maintenance motivates how production companies stand to benefit from alarm forecasting in natural gas pipelines.

Because of entropy, maintenance is required to keep anything in working condition. A human body needs nutrition and exercise, while a gas pipeline needs periodic cleaning and replacement of corroded or weak segments. There is value in different kinds of maintenance to offset cost and labor and to resume common function. According to the Federal Energy Management Program, three types of maintenance are common [1].

Known as the “run it 'till it breaks" model, reactive maintenance is the simplest to adopt. Labor and capital cost is deferred until something breaks. At that point, what is broken is fixed. No other action is taken on the machine while it is running. While rudimentary, reactive maintenance has its advantages in low costs and less staff while nothing is out of service. Preventative maintenance is acting based on a schedule or time to find and mitigate degradation. This is analogous to periodically cleaning the inside of the pipe to flush out any accumulated flow blockage. However, no amount of preventative measures will prevent catastrophic failures, but rather decrease the number of regular deteriorations. Predictive maintenance is based on actual measurements that can detect the onset of system degradation. This is not based on time but rather on condition. For production facilities, the cost and time benefit of conducting predictive maintenance can be appreciable by saving 8% to 12% over a preventative model [1].

Historically, the natural gas production industry performs a form of reactive maintenance [3]. This form of operation is outdated, as the maintenance is required after the problem has occurred, and the damage has been done. Although reactive maintenance is logical and will always be a part of a dynamic system such as natural gas processing, new technologies allow for action to be taken before a reactive process is carried out.

Different from the scheduled preventative maintenance carried out on a pipeline [1], [3], [6], [44], predictive maintenance allows pipeline controllers to act before issues occur. This is made possible through the constant monitoring of the pipeline, a control center, and pipeline operators.

Constant monitoring of the pipeline is achieved through wireless sensor networks (WSN) systematically installed throughout the pipeline. Digital technologies and wireless communications allow the sensors to relay real-time information back to the pipeline control center to help determine machinery health, plan maintenance intervals, and reduce downtime. Especially valuable in the oil and gas industry due to extraction wells often being in remote places, [45] shows how the deployment of wireless sensor networks in pipelines has been a large contributor to safer and more efficient natural gas transmission by connecting offsite pipeline controllers and onsite pipeline personnel. Alakbarov [46] walks through the architecture of a modern WSN system and stresses the importance of reliable communication between the pipeline and the control room. These works point out that these sensor networks are so valuable to the production process, it is not unusual for gas plants to employ a full-time instrument technician to ensure accurate sensor calibration and maintain communication with the control room [6].

Pipelines often have many sensors simultaneously sending a stream of data to the control room. This leads to a huge amount of daily data generation. Similar to the technology used in this thesis, [47] addresses the large-scale data being communicated from the WSN installed on pipelines using big data techniques. Once the data has been recorded, it is communicated to the pipeline control room for immediate analysis. This vital analysis is enabled by software packages that receive and parse the incoming WSN data.

Pipeline control rooms like the one used in this research can be outfitted with supervisory control and data acquisition (SCADA) systems [6]. SCADA systems provide highly configurable industrial hardware/software applications used to manage process control and remote data transmission [7]. The natural gas pipeline technology overview [3] explains the importance of these systems and how with WSN's, SCADA systems give pipeline operators more control over equipment, processes, and communication from remote places. Article [48] discusses how SCADA systems continuing role in gas production has evolved over the last 30 years, increasing recognition and popularity for IT-based automation. Still, the coordination of a natural gas production pipeline involves many complicated processes simultaneously occurring. Uraikul [22] explains how the near-instantaneous information provided by SCADA systems gives pipeline controllers the consistent, fast and reliable decision support they need to ensure safe transportation of the large qualities of gas flowing through the pipeline.

Good pipeline controllers are familiar with the system they are operating, knowledgeable of the tools at their disposal, and quick to recognize immediate threats to the pipeline. They direct, control, and monitor the gas from extraction well to distribution sales point. The importance of a qualified pipeline controller is critical to the success of production. Thus many guides, manuals, and other relevant literature has been published by production companies and the U.S. government to aid these workers [49], [50], [51]. These guides and regulations inform pipeline controllers of the limits within which they can operate the pipeline as well as federal regulations. Operators follow their own set of guidelines, as each production pipeline is rated for different flow rates and performance. They are the decision-makers that keep the pipeline operating and the primary users of the alarm forecasting algorithms and the other tools described in this section. It is with these tools and operation experts that it is possible to perform predictive maintenance on a production pipeline.

Natural gas is being produced at unprecedented rates [52]. The only way the modern production pipeline can ensure the most reliable, productive, and safe operation is through the adoption of new pipeline technologies. Developments such as WSN and SCADA enable new real-time data analysis to help production companies predictively maintain their pipeline. New areas of research have been developing in the field of natural gas production with the intent to manage the safe transportation of this fossil fuel.

2.4 Past Work in the Field of Natural Gas Production Pipelines

The ability to forecast natural gas alarms in production pipelines comes from a foundation of years of research from engineers, mathematicians, and industry experts. This section highlights some recent work leading to this thesis. The application of real-time pipeline data to forecast alarms in natural gas production pipelines is a relatively new area of research. In fact, the definition of an alarm used in this work is not an industry standard, rather a standard of the production company sponsoring this research. Therefore, little work has been published in the natural gas production field that includes the use of alarm thresholds as a form of predictive maintenance. However, there have been many closely related works that strive to achieve the same objective of maintaining a production pipeline using machine learning, artificial intelligence, and big data analytics. In most of the following work, the goal is the same: to protect the pipeline from failure.

Continuing the theme of Chapter 1, if gas being injected into the pipeline is low-quality, numerous problems can threaten flow assurance. Three commonly found problems in production pipelines include hydrate formation, leaks, and corrosion. Work devoted to combating these problems is relevant to this thesis's concentration as they all fall under the umbrella of obstacles that our alarm forecasting is trying to overcome. Many of the variables used to analyze and predict these problems are the same used to forecast alarms. While we are not specifically concentrating on hydrate formation, pipeline leaks, or corrosion, our general-case forecaster can alert controllers to the situations in which these problems can occur or may occur in the future.

One of the three common internal issues pipelines are combating today is the formation of natural gas hydrates. Gas hydrates are clathrate physical compounds of water and natural gas, where the molecules in the gas are trapped in polygonal crystalline structures made of water molecules [53]. These crystalline structures, or simply ice-looking crystals, can accumulate within a pipeline, causing potentially production-halting blockage, damage to pipeline structural integrity, and transport system equipment failure [4]. As shown in work such as [54], hydrates can form anywhere in the pipeline where hydrocarbons and water are present at the right temperature and pressure. Presenting an additional concerning aspect of hydrate formations, [5] points out that they can occur within minutes without prior warning — stressing the importance of real-time detection systems.

Thus, several computational approaches address the issue of hydration formation. Naseer [26] discusses how the formation of gas hydrates can be combated with computational fluid dynamics, locating and predicting hydrate build up in certain sections pipe. Research in [55] describes a method using kinetic inhibition to prevent flow channel blockage of these hydrates. [56] follows a control strategy of using thermodynamic inhibitors to push the hydrate formation phase boundary away from the temperature and pressure conditions at which natural gas hydrates form. While all these approaches are substantially different, they all rely on the data being produced within the pipeline. Specifically, the data used to forecast alarms such as pressure, temperature, H_2S, and H_2O.

Other work in pipeline failure includes leak detection. Like the data sets used in this research, leak detection is heavily reliant on time series and rates-of-change in various signals. Because a leak has serious detrimental effects on both pipeline operations and the environment, a production pipeline will undergo preventative maintenance through periodical inspections conducted by maintenance personnel. This requires intensive human involvement and fails to provide real-time feedback to pipeline operators. The leak detection methods described in [57] help reduce these periodical inspections by incorporating hierarchical leak detection and localization through the use of WSNs. Wan [57] uses the phrase “alerting pipeline operators” and describes false alarms and the reliability of WSNs in natural gas pipelines. Summarizing recent advancements of pipeline monitoring and leak detections, [43] provides an excellent overview of the different types of leak detection systems, including many that involve temporal-based signal processing. Natural gas pipeline leaks are serious problems to encounter. As such, there is an equal concentration on how and where these leaks originate.

Corrosion causes natural gas pipeline leaks [58]. As described in Section 1.4, raw natural gas consists of different compounds. There is dry gas, gas that requires little post-extraction processing, and there is wet gas, which must be thoroughly processed before pressurization and injection into the pipe. One of the reasons why wet gas must be processed significantly more than dry-gas wells is the high level of water, CO_2, and H_2S present in the gas. Gas with high levels of water and these dissolved gasses is referred to as acid gas for its potential to corrode the inside of a pipeline. Several works [58]–[61] have carried out analysis of pipeline corrosion due to the presence of acid gas to reduce leakage accidents and pipeline segment weakening. While these works show the problem of pipeline corrosion is prevalent and makes production companies susceptible to large economic loss, stopping leaks before they begin has caught the attention of many.

The corrosion of pipelines has led to several works being published aiming to predict and combat acid gas corrosion. ObaniJesu [62] focuses on the development of a predictive model for the corrosion rate in natural gas pipelines, specifically with H_2S as the corroding agent in different operating situations. Much like how the alarm forecasting methods in this thesis need to adapt to different operating situations, [62] models situations with varying temperature, pressures, and acidity of the gas within the pipe. Anticipating the other challenges of flowing low-quality gas through the line, [63] ties in gas hydrate formation and its contribution to corrosion rate along subsea pipelines.

Failure to detect and correct corrosive gas damage to a pipeline can result in large scale ruptures or explosions. Bedairi [64] shows how a finite-element method using an elastic-plastic fracture mechanics approach can predict crack-in-corrosion defects, while also bringing to light the lack of assessment methods or current codes for these large-scale incidents. The methods described in each of these works are based in mathematical foundations and are further discussed in Section 2.5.

2.5 Time Series Analysis and the Natural Gas Industry

Forecasting alarms with machine learning is approached as either a classification or a regression problem. The output of a classification-based model is binary: An alarm is either present, or it is not present. A regression-based approach predicts future values which are compared against rules that define an alarm. The benefit of a regression-based model is in its output, since it can be used to diagnose the state of the pipeline rather than just an alarm being imminent. Several models can be trained with multiple time horizons that give control operators more discretion in avoiding unsafe states or unacceptable gas.

One of the first steps in many data analyses applications is performing regression analysis [65]. Autoregressive models have gained popularity over the last few decades due to their simplicity, effectiveness, and practical nature in the time series domain. Such an analysis can provide useful information about correlation and the directionality of the data, how to estimate the model coefficients, and determining the validity and usefulness of the model [66]. The correlation found in data sets indexed by time has led to the significant development of time series work in industrial production [66]. The temporal aspects of the data sets used in this work contain valuable information relating to how the system responds to issues threatening flow assurance. Hence, much work has been put into the fitting and analysis of time series models.

Development of such work is seen in [67] and [68] when the fitting of time series models and autocorrelation analysis was published in the 1960s. The inspiration behind these works and many others was to find efficient ways for parameter estimation while working with data serial dependence. Due to the nature of time data points, it is understandable that one observation is often statistically dependent on another observation recorded at a different time [65]. This property of time series data violates one of the fundamental assumptions of statistical modeling that the data must be statistically independent. Work such as [69] show how to test and avoid the misleading results that can arise from serial dependence in time series forecasting. If the proper steps are taken, there are many examples of successful regression-based time series models.

Linear regression has been proven successful in the energy production industry. Thus, it is the first method explored in this work (Section 4.4). In a similar application to forecasting alarms in natural gas pipelines, Vitullo [70] demonstrates the use of time series to forecast the amount of gas local utilities need to flow to satisfy hourly and daily demand. Similar papers [71] and [72] also provide examples of successful forecasting in the natural gas industry through the use of historical demand and consumption of natural gas. Often, univariate time series forecasting models are augmented by including other data sets measuring similar qualities.

The autoregressive models with exogenous variables (ARX) presented in Section 4.5 exploit relationships one signal has with others. It can be very beneficial to consider a group of time series variables as opposed to concentrating on one single series, thus making the model more dynamic and sensitive to changes elsewhere in the system. Spliid [73] lays out how large multivariate time series can be used with distributed lags for fast estimation forecasting models. Akouemo [74] applies this idea to the natural gas industry and incorporates the important idea of how an ARX anomaly detection can be used to detect and impute anomalous data.

Time series analysis has attracted much attention. To determine some of the most accurate methods, competitions such as the M-Competitions [75] are designed to test extrapolation methods in a variety of scenarios and areas of research. Using three thousand time series, the M-3 competition [76] tested each model entered using real-world objectives with the aim to help forecasters make business decisions. The winner of the M-3 competition was the Theta method [77], the third forecasting method used in this work. Out of the 24 methods submitted in this competition, the Theta method performed the best based on empirical and efficiency-benchmarking assessments.

The Theta method is a specific decomposition technique that uses the projection and combination of individual components [77]. Otherwise known as simple exponential smoothing with drift, as proved in [78], the decomposition of both long-term and short-term components are extracted from the data and are referred to as the ‘theta lines.’ The long-term trend component is the first theta line, which removes the curvature of the time series so that it can be a good estimator for long-term behavior of the series. The short-term trend component of the data doubles the curvatures of the series to gain better approximations of the short-term behavior. Then, components are combined with optimized weightings to produce a forecast value of the original series. Time series work has been continued with the Theta method and has found success in non-competition work such as [79]. More in-depth analysis has been carried out [80] to optimize univariate and multivariate time series forecasting to better fit each application of this method in specific business settings.

The last method described in Chapter 4 is an artificial neural network (ANN) forecasting model. Al-Fattah [81] points out the advantages of ANN models in time series forecasting. In some applications, they outperform traditional time series models. Al-Fattah [82] shows how to predict natural gas production in the U.S. using an ANN similar to the network described in 0. While introducing their technique, [82] describes how the nonlinearity of ANN transfer functions introduces advantages in time series forecasting compared to conventional regression techniques. Nonlinear ANNs have proven successful as huge collections of data have become available over the last few years [47]. These data-driven, self-adaptive models work well with the natural gas production industry's large datasets, and success has been found using dimension reduction techniques seen in [83] to identify production variables that have direct flow assurance implications.

Chapter 2 has described the balance needed in the natural gas production industry between increasing output and maintaining flow assurance. The combination of the above works reflects the inspiration of the research completed in this thesis. A creative aspect of each work is how the author applies the data available to them to best achieve their objective. Without data, these methods would not exist and the internal workings of production pipelines would be less well understood. Chapter 3 presents and analyzes the data use in this thesis to forecast alarms in natural gas pipelines.

CHAPTER 3

Introduction to Signal Data and Alarm Thresholds

3.1 Chapter Objectives

This chapter introduces the gas quality signals used in this project. For each signal, we define alarm thresholds and discuss the behavior of the signal. Then, we explain the data cleaning process and the tools used to conduct this research.

3.2 Natural Gas Signals and Alarm Thresholds

Tens of sensors are within the pipeline providing a constant flow of information to the pipeline operators in the control room. Each sensor measures a pipeline condition, such as gas composition (Table 1.1), internal pressure, flow, etc. These sensors allow operators to monitor changes within the system, determine processing machinery health, and assist the pipeline controllers with the transportation of large amounts of natural gas safely through the pipeline. This section specifies which sensor signals are used to forecast alarms.

While there are many sensors concurrently collecting and sending information to the control room, only a few are used to forecast alarms. The signals chosen in this work are the ones that have the greatest impact on flow assurance. The pipeline conditions these signals measure are deciding factors of whether the pipeline operates as normal or gets shut in. For example, controllers have less interest in trace amounts of rare gases than the internal pressure of the pipe. The production company has provided five signals that they regard as most important to ensure flow assurance: pressure (psi), heat content (BTU), hydrogen sulfide (H_2S), carbon dioxide (CO_2), and moisture (H_2O).

All five pipeline signals are recorded at five locations along the pipeline. Four sets of the signals are generated at each central processing facility (CPF), while the last set is from the distribution point. The set of signals being generated at the distribution point is the target data to be forecast. The distribution point signals are used by the sales point operators (different from the production pipeline operators) to determine if the gas flowing into the sales point is of an acceptable quality. [5] provides a general overview of acceptable quality gas. This work uses more stringent characterizations in the form of alarm thresholds to define what is acceptable.

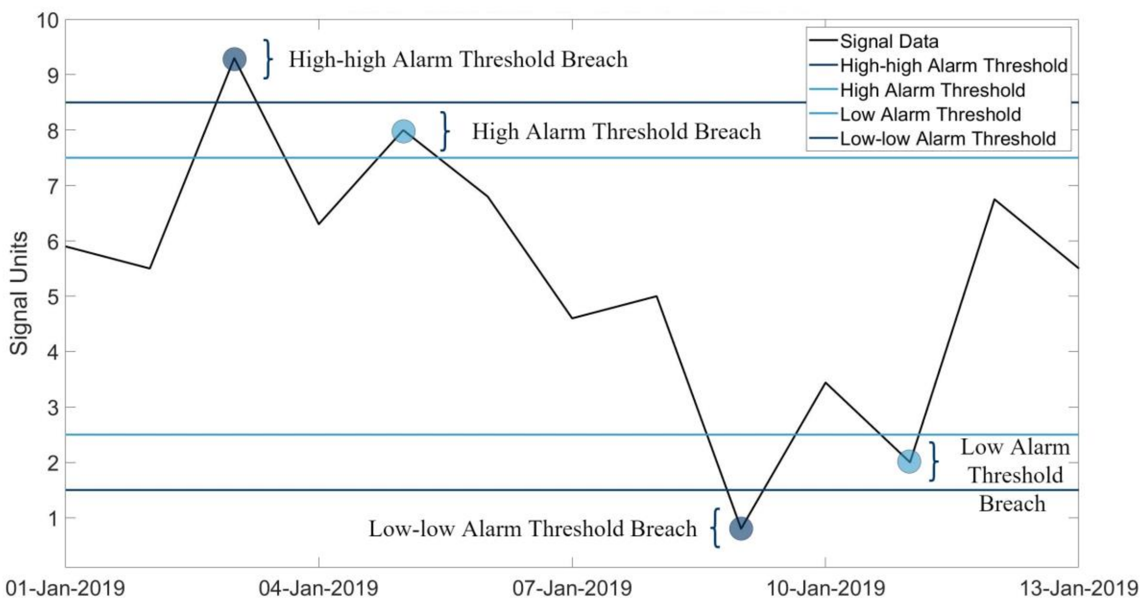

Figure 3.1: Alarm thresholds for a generic time series

Alarm thresholds can be thought of as gas quality limits. If a sensor measurement of a pipeline condition exceeds or falls below an alarm threshold, an alarm may be imminent, and the pipeline may get shut in. The production company defined four types of alarms — high-high, high, low, and low-low. Figure 3.1 shows the generic four alarm thresholds.

The high-high alarm signifies an extreme system lapse, and the pipeline is either already shut in or close to it. If a high-high alarm is triggered, the main concern of the pipeline controller is to protect the production equipment from damage, and to reduce the amount of line pack building at the distribution point. The next alarm threshold is a high alarm. Lower that a high-high alarm, high alarms indicate a serious problem is forming in the system, and action is needed to correct the trajectory of the signal. Conversely, a low alarm indicates that the signal is falling beneath the acceptable level. A low-low alarm alerts controllers of a potential equipment failure or a shut-in worthy problem in the system. Tables 3.1 - 3.5 present each target signal along with its corresponding alarm thresholds.

3.3 Pressure Signal (psi)

Pressure is what moves gas through the pipe, with the gas flowing from high pressure to low pressure [6], [25]. This is a fundamental principle of a natural gas production pipeline and is the main tool used by the pipeline controllers to control the natural gas delivery system [84]. By closely regulating the pressure, the controllers manage how much gas is in the system, how fast it is moving, and coordinate the production of several wells at once. Pressure is a measure of the pounds per square inch (psi) within the pipe and has been deemed the most important condition in the system by the production company involved with this project. Excessive pressure is the most common reason for the pipeline being shut in. Figure 3.2 shows the pressure time series for the distribution point fluctuating between 950 psi and 1250 psi.

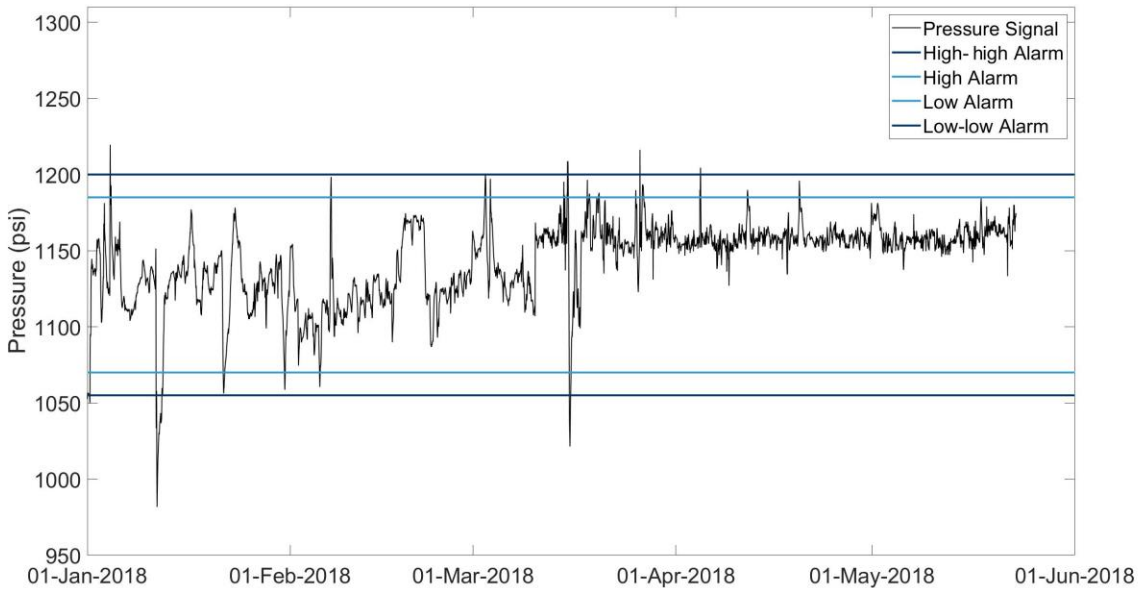

Figure 3.2: Pressure time series recorded at the distribution point from January 2018 – May 2018

The exact thresholds for the pressure time series and their occurrences within our data are summarized in Table 3.1.

Table 3.1: Pressure alarm thresholds and their observed occurrences and frequency percentage (N = 210,000)

PSI			
	Threshold	*Occurrences*	Frequency (%)
High-high	> 1200	294	0.14
High	> 1185	2458	1.20
Low	< 1070	3171	1.55
Low-low	< 1055	1730	0.85

Table 3.1 shows the overall frequency of triggered alarms is quite low. We expand upon this effect in Chapters 4 and 5 when choosing model structures and error

metrics. The next signal discussed, measuring the gas heat content (BTU), shows a similar number of alarms triggered.

3.4 Heat Content Signal (BTU)

The term “heat content” is used in the production industry to help characterize the quality of natural gas. When the gas is sold at the distribution point, its heating value is a main determinant of its sales price. Gas with a lower heating value is not as valuable as a gas with a higher value [85]. This heating value variable depends on the gas consistency and how much energy is released when the gas is burned [40]. It is measured in British Thermal Units (BTU) (the amount of energy needed to increase the temperature of one pound of water by a one degree Fahrenheit [17]). These qualities are important to both the production company and the distributor, as a contractual agreement holds the production company responsible to deliver gas that meets the standards of the distributor. Figure 3.3 shows the heating value signal recorded at the distribution inlet varying between 1000 BTU to 1160 BTU.

Figure 3.3 shows that most of late March is operating under a low alarm. The pipeline operators confirmed this irregularity is authentic and not anomalous data.

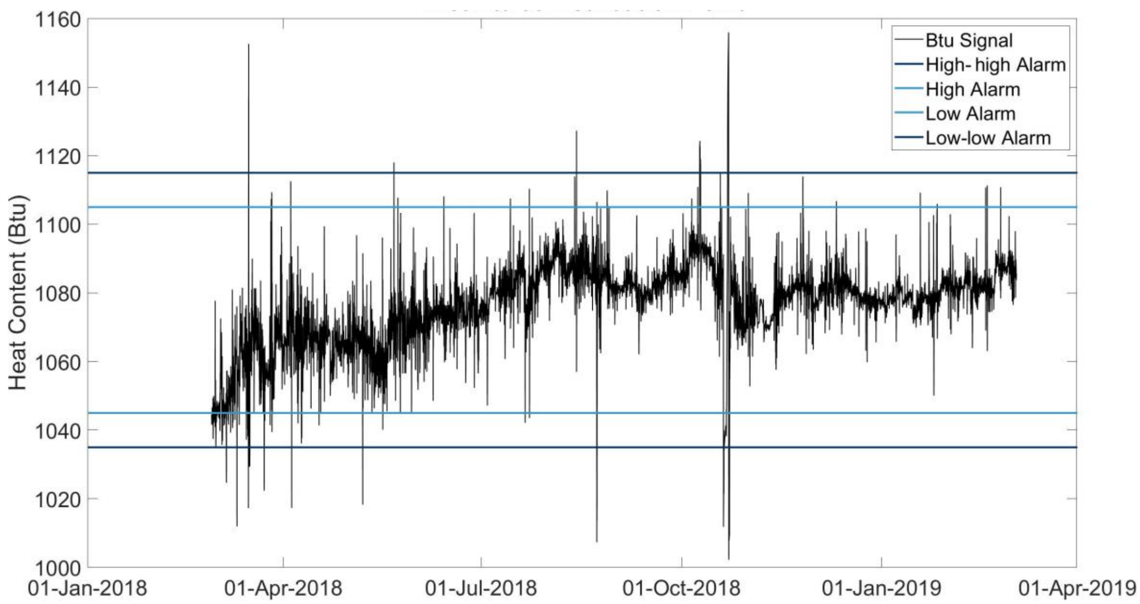

Figure 3.3: Heat content (BTU) time series signal recorded at the distribution point from January 2018 to April 2019

Table 3.2 shows that the number of low alarms is higher than the number of other alarms during this time.

Table 3.2: BTU alarm thresholds and their observed occurrences and frequency percentages (N = 523,600)

BTU			
	Threshold	Occurrences	Frequency (%)
High-high	> 1115	1309	0.25
High	> 1105	1966	0.37
Low	< 1045	6980	1.31
Low-low	< 1035	1419	0.27

Pressure and BTU are the first signals identified due to their importance. While all signals in this work are being monitored constantly in the pipeline control room, the pipeline operators identified Pressure and BTU signals to have triggered the highest number of alarms in recent production. However, looking beyond recent production, the next signal examined is the sulfur content of the gas, which represents an extreme threat to the pipeline's long-term structural health if not closely regulated.

3.5 Hydrogen Sulfide Signal (H_2S)

The sulfur content, or the amount of hydrogen sulfide (H_2S) present in gas, is one of the two components that determines if gas is "sweet" or "sour." Sweeter gas contains a lower sulfur content and less carbon dioxide, while sour gas contains an unacceptable amount of these gases. H_2S is a carefully monitored quality, as sour gas is not accepted at the sales points due to its corrosive nature and potential to damage the pipeline [6]. Sulfur stress cracking has been an issue within the production industry, and considerable research has led to new methods for sweetening of sour natural gas [86]. In this work, gas is labeled sour when the H_2S sensors read values greater than 3 parts-per-million (ppm).

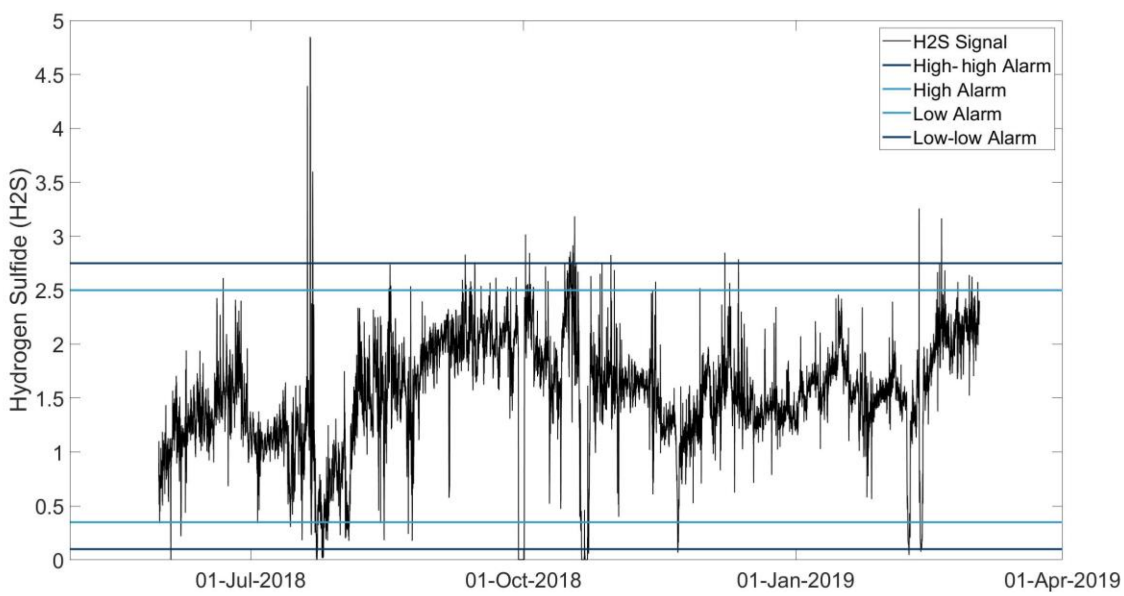

Figure 3.4: Hydrogen sulfide (H_2S) time series signal recorded at the distribution point from May 2018 to April 2019

Figure 3.4 displays the H_2S signal, showing that the controllers of the pipeline do a relatively good job maintaining the level of H_2S under 3ppm. The exact thresholds for the H_2S time series and their occurrences within our data are summarized in Table 3.3.

Table 3.3: H_2S alarm thresholds and their observed occurrences and frequency percentage (N = 388,880)

H_2S			
	Threshold	Occurrences	Frequency (%)
High-high	> 2.75	700	0.18
High	> 2.50	3660	0.92
Low	< 0.35	13209	3.31
Low-low	< 0.10	7251	1.82

Both Table 3.3 and Figure 3.4 show the low and low-low alarm being triggered more often than any high-alarm. This was interesting to us, as having very little hydrogen sulfide in the gas stream indicates very lean, high-quality gas. The need for any low alarms seemed unnecessary, yet the pipeline controllers told us that the low alarms can help manage machinery at the CPF. If a gas stream at a CPF is registering almost zero sulfur content, the pipeline controllers use that information to check on the equipment and possibly reduce the revolutions per minute of the gas sweetening machinery to prevent unnecessary wear. A similar situation can be seen when monitoring the carbon dioxide signal.

3.6 Carbon Dioxide Signal (CO_2)

Carbon dioxide is the second component of sour gas. CO_2-rich gas is not corrosive and has little to do with flow assurance on its own. However, when combined with hydrogen sulfide, acid gas forms which leads to critical problems for production pipelines (Section 2.3). If the pipeline were to get shut-in, flaring any amount of acid gas from the system must be avoided due to the amount of greenhouse gas released. This usually forces production to halt until the gas can either be diffused with sweet gas or back-flown for reinjection into the ground. Figure 3.5 shows the levels of CO_2 from May 2018 to April 2019 fluctuating between 0.25 and 1.75 parts-per-million.

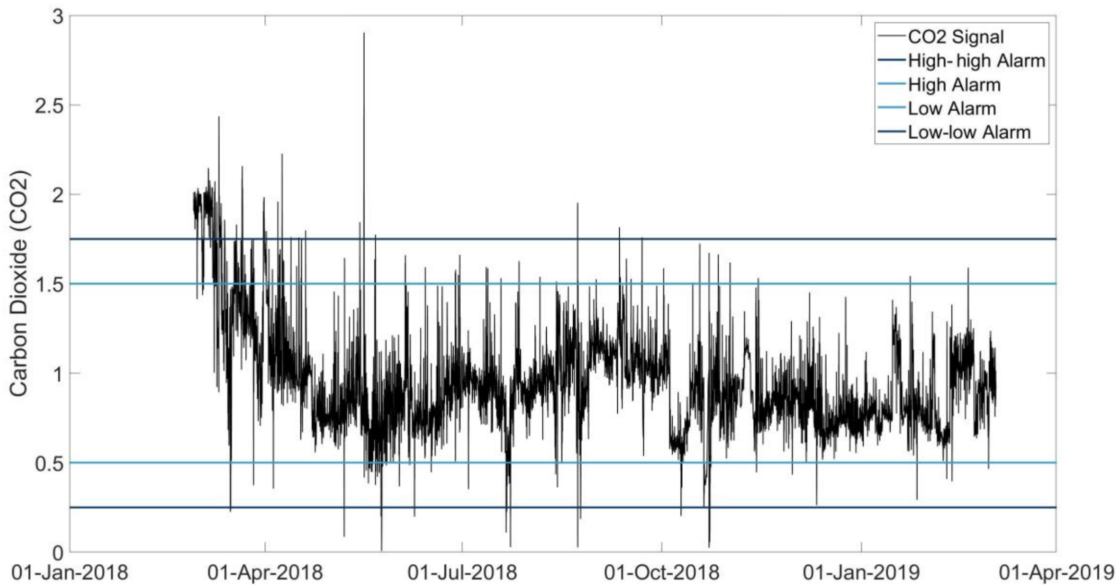

Figure 3.5: Carbon dioxide (CO_2) time series signal recorded at the distribution point from May 2018 to April 2019

The exact thresholds for the CO_2 time series and their occurrences within our data are summarized in Table 3.4.

Table 3.4: CO_2 alarm thresholds and their observed occurrences and frequency percentage (N = 530,842)

CO_2			
	Threshold	Occurrences	Frequency (%)
High-high	> 1.75	15129	2.85
High	> 1.50	24829	4.67
Low	< 0.50	6516	1.23
Low-low	< 0.25	679	0.13

Figure 3.5 shows the carbon dioxide signal triggering high alarms in late March. This correlates with the BTU signal triggering low alarms in late March and is a result of the production pipeline producing higher quality gas during that time. Table 3.4 shows the H_2S alarm thresholds and their frequency of being triggered. Similar to how the H_2S signal triggers low alarms, the low alarms for CO_2 do not indicate a failure; rather they relay information back to the pipeline controllers that they use for adjusting the system.

3.7 Water Content Signal (H_2O)

The moisture content is measured in pounds of water per million standard cubic feet of gas [6]. Figure 3.6 shows the water content measured at the distribution point. Water or water vapor (H_2O) is almost always present in raw natural gas, ranging from trace amounts to saturation [5]. In this production operation, gas is dehydrated as it is pulled from the wellhead to the processing plant via refrigeration units.

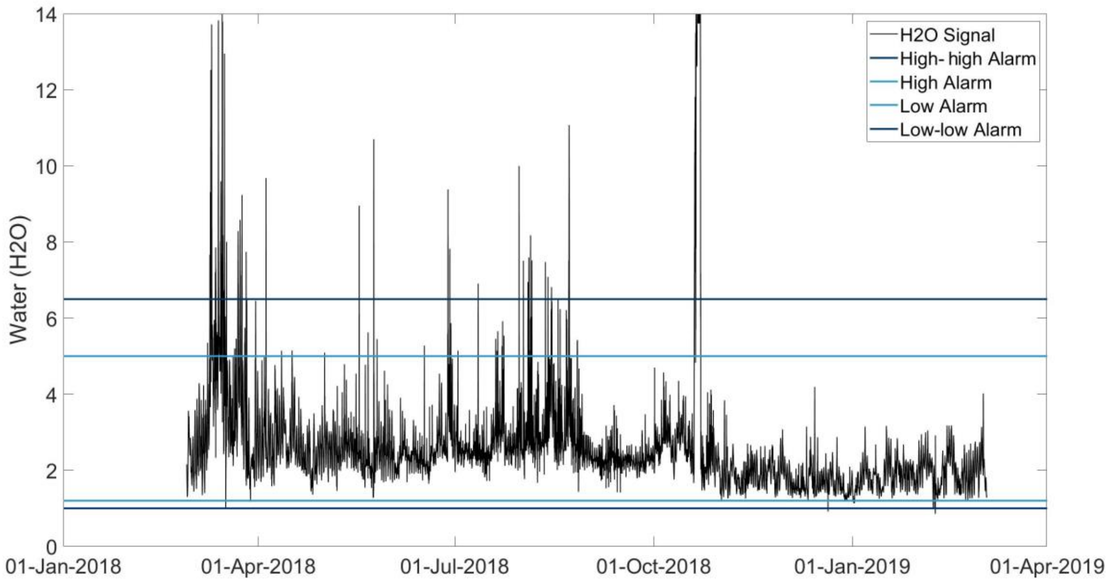

Figure 3.6: Water content (H_2O) time series signal recorded at the distribution point from May 2018 to April 2019

The exact thresholds for the H_2O time series and their occurrences within our data are summarized in Table 3.5.

Table 3.5: H_2O alarm thresholds and their observed occurrences and frequency percentage (N = 533,026)

H_2O			
	Threshold	Occurrences	Frequency (%)
High-high	> 6.5	8102	1.52
High	> 5.0	14809	2.78
Low	< 1.2	1058	0.20
Low-low	< 1.0	87	0.02

The number of low and low-low alarms shown in Table 3.5 are the fewest of all the signals, with low-low alarms being triggered only 0.02 percent of the time. This fact was discussed with the pipeline controllers, who decided to keep the low alarm thresholds as-is but declared the lower H_2O alarms generally less important than the other signals discussed. The effect of this decision is discussed in the Chapter 5 when considering the performance metrics used to evaluate the H_2O forecast models.

3.8 Preparation of Raw Time Series Data

This section describes the cleaning process carried out on the time series signals presented in Sections 3.3-3.7. The process begins with the initial retrieval of the data in raw form and ends with the cleaned time series used to train the alarm forecasting models. The following section begins with a brief overview of the data conversion from comma separated value (csv) files to MATLAB time series. Then, the cleaning of those time series objects leads to a discussion of non-uniform time series and linear modeling. Finally, the anomaly detection and imputation method is examined.

The production company supplies the data used in the project via csv files, with each file containing the historical signals generated from central processing facilities located along the pipeline. As in many cases when using signal data collected in the field, there are many ‘NULL’ entries in each file, either from communication failure between the pipeline sensors and the control room or temporary equipment failure. Unix scripts are used to located and delete any ‘NULL’ time-value pairs. After each signal is separated into its respective csv file, the ‘time’ column vector of each csv is converted from a Microsoft Excel timestamp to a character vector in preparation to turn the csv signals into MATLAB time series objects.

The csv files are read into MATLAB and stored as .mat files. In most time series models, the sampling rate or interval at which observations are recorded is vital information when processing raw data [66]. The time series and signal processing algorithms described in Chapter 4 require a signal with a consistent sample rate. The signals received by the production company are asynchronous [87], meaning that there is no uniform amount of time between each sensor reading. Each signal must be converted to be uniformly sampled while also maintaining the natural behavior of the data. Each gas quality behaves differently in the pipeline, so a consistent sample rate must be chosen for all five time series such that the natural behaviors of each signal remain present in the interpolation.

The non-uniformly sampled pipeline signals are received in the control room and presented to the operators via a supervisory control and data acquisition (SCADA) system. The SCADA system provides our algorithms the data needed to make real-time forecasts, so the sampling rate matters. The production company's SCADA system samples on average every nine seconds; however, the SCADA sample rates can range from 5 to 200 seconds, as shown in Figure 3.7.

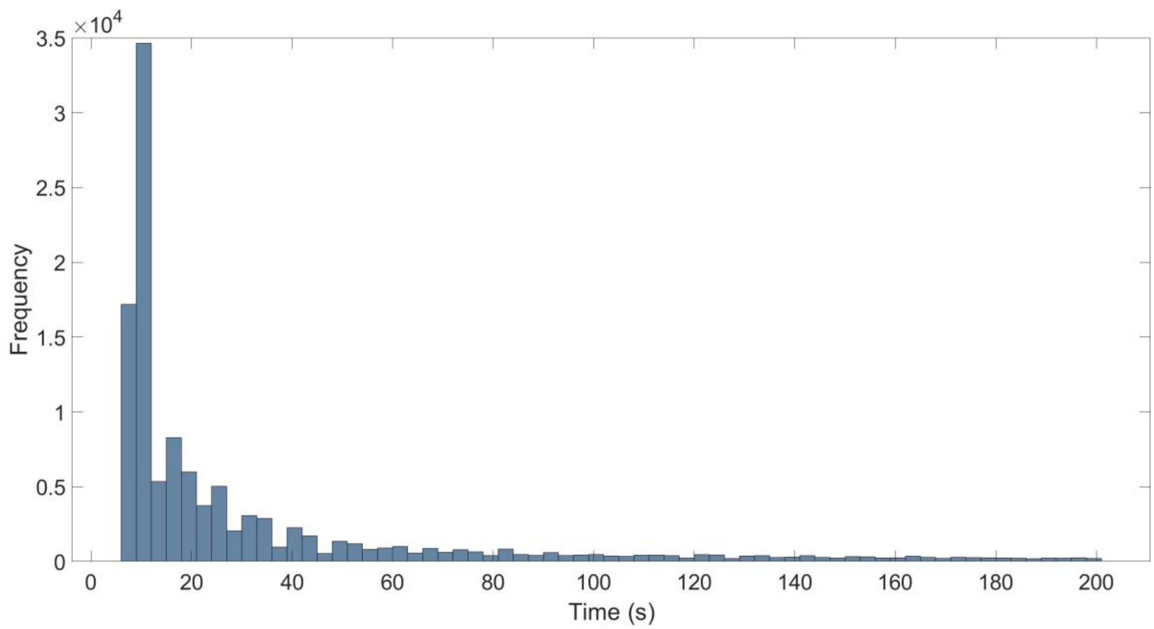

Figure 3.7: Histogram of sampling intervals of the distribution point's pressure signal

Figure 3.7 shows the most frequent sampling rate is 9 seconds. The pipeline controllers determined that a consistent ten-second sampling rate would maintain the integrity of the data and that the SCADA system would be able to provide consistent observations to the alarm forecasting algorithms using a sample-and-hold approach. This allows us to resample the time series objects using ten-second intervals with built-in MATLAB R2019b resample functions [88] and produce a uniform time series. Then we begin the process of detecting and correcting possible inaccurate observations with our time series anomaly detection and imputation algorithm.

3.9 Time Series Cleaning — Anomaly Detection and Imputation

Anomalous data degrades our alarm forecasting model parameters and real-time forecasts. To avoid anomalous data being used in our parameter estimation and forecasting, we implement an anomaly detection and imputation technique. This section defines what an anomalous observation is and how we differentiate between real and erroneous signal observations.

In many real-time monitoring tasks, it is vital to have accurate machinery and skilled workers to identify abnormal behavior quickly. In this setting, we have the sensors within the natural gas production pipeline and the pipeline controllers monitoring the SCADA system. The controllers interpret the data produced from the sensors to determine the state of the system. If an unusual event is occurring, the controllers are the first to identify and categorize what is happening. For example, if pipeline maintenance requires the internal pressure of the pipe to be lowered for a few hours, alarms are triggered, but no action is taken because the controllers are aware of the necessary maintenance. Such human-interactive events create data that does not represent the usual day-to-day signals needed to train the forecasting model parameters and ultimately degrades our forecasting models. Conversely, if a naturally occurring event appears in the data, it is crucial to include that event in the training data so that we can forecast the correct alarms for that situation.

Knowing the difference between naturally occurring events and anomalous data falls into domain knowledge that the production controllers possess. Figure 3.8 shows examples of anomalous observations (circled in red) found within the raw time series signal. The anomaly detection and imputation considers the domain knowledge of the pipeline controllers and statistical likelihoods of each point being a natural observation or an error.

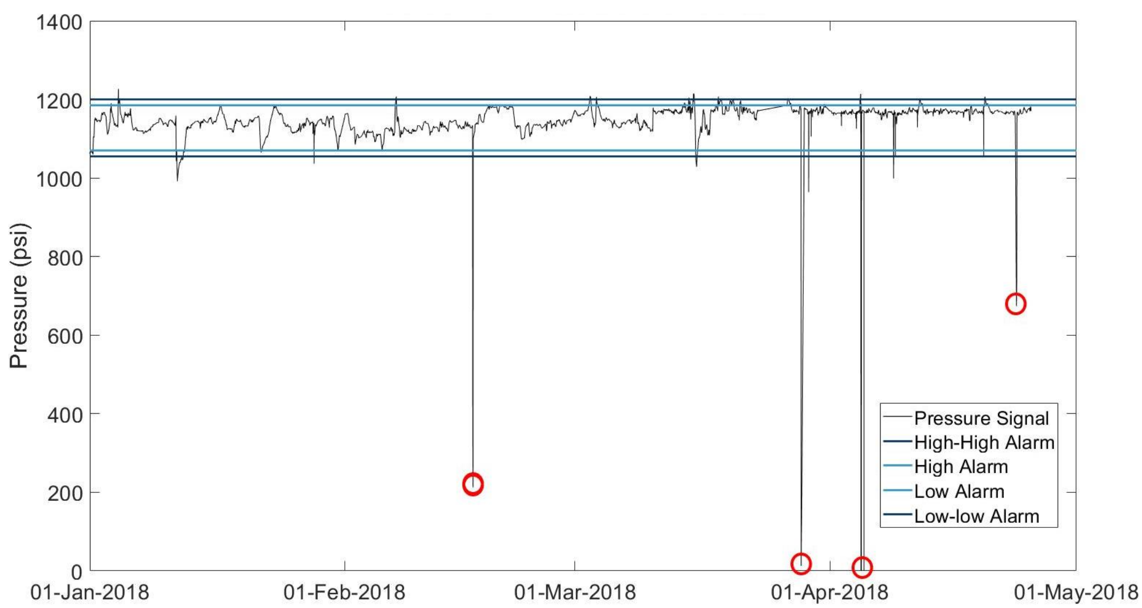

Figure 3.8: Pressure time series with confirmed anomalies circled in red

The anomaly detection and imputation algorithm used in this work is based on [74] by Akouemo, who applied this technique to the similar problem of natural gas energy forecasting. The converted algorithm used in this work stems from her hypothesis-driven outlier detection method but has been modified to fit the problem of this research.

There are two areas in which anomaly detection is used. The first is to determine whether the training data is legitimate when estimating the model coefficients. If a model is trained using anomalous signals, the model may produce erroneous forecasts. This is especially true when applying the model to out-of-sample signals, which is the main intent of the production company controllers.

The second application of anomaly detection is operation of our forecasting algorithm in real-time. As the SCADA system receives new data, it is possible that some observations are anomalous. To detect these occurrences, the anomaly detection algorithm is fed the most recent observation and uses a Bayesian maximum likelihood classifier [66] to label it anomalous or not. If identified as anomalous, the data point is replaced with model estimates.

Chapter 3 introduced the gas quality signals used in this project to forecast pipeline alarm. For each signal, we present alarm thresholds and explain the signal’s behavior. The data cleaning process is described to give an idea of the issues in our datasets, which leads to the anomaly detection and imputation algorithm discussion. Once the data has been cleaned and parsed, it is possible to forecast each signal using the methods presented in Chapter 4.

CHAPTER 4

Forecasting Methods and Framework

4.1 Chapter Objectives

This chapter begins by providing a description of the alarm forecasting framework used in this work. We then define the training and testing data sets and a baseline model to help evaluate each forecasting method. The first method implements a 10^{th}-order autoregressive model. The second method is an extension of the first autoregressive model but incorporates exogenous variables from different central processing facilities (CPF). Then, the Theta method is examined, where simple exponential smoothing with drift is applied to the data sets. Finally, an artificial neural network is used to forecast pipeline signals.

4.2 Framework for Real-Time Alarm Forecasting

This section introduces the notation used in this chapter and describes the framework for real-time alarm forecasting. A time series is a set of data ordered in time [66]. In this work, a distribution point signal, Y, is

$$Y = \{\, y(t), t = 1, \dots, N\}.$$

In this form, $y(t)$ is the value of the distribution point signal y at time t. Time t is uniformly spaced at 10-second intervals (Section 3.8), and the values of $y(t)$ have been tested for anomalies (Section 3.9). The signals at the distribution point are differentiated with signal type subscripts. Thus, Y_{psi}, Y_{Btu}, Y_{H_2S}, Y_{CO_2}, and Y_{H_2O} are the signals for pressure, heat content, hydrogen sulfide, carbon dioxide, and moisture content, respectively.

The notation used to describe the alarm forecasting algorithms is shown in Figure 4.1.

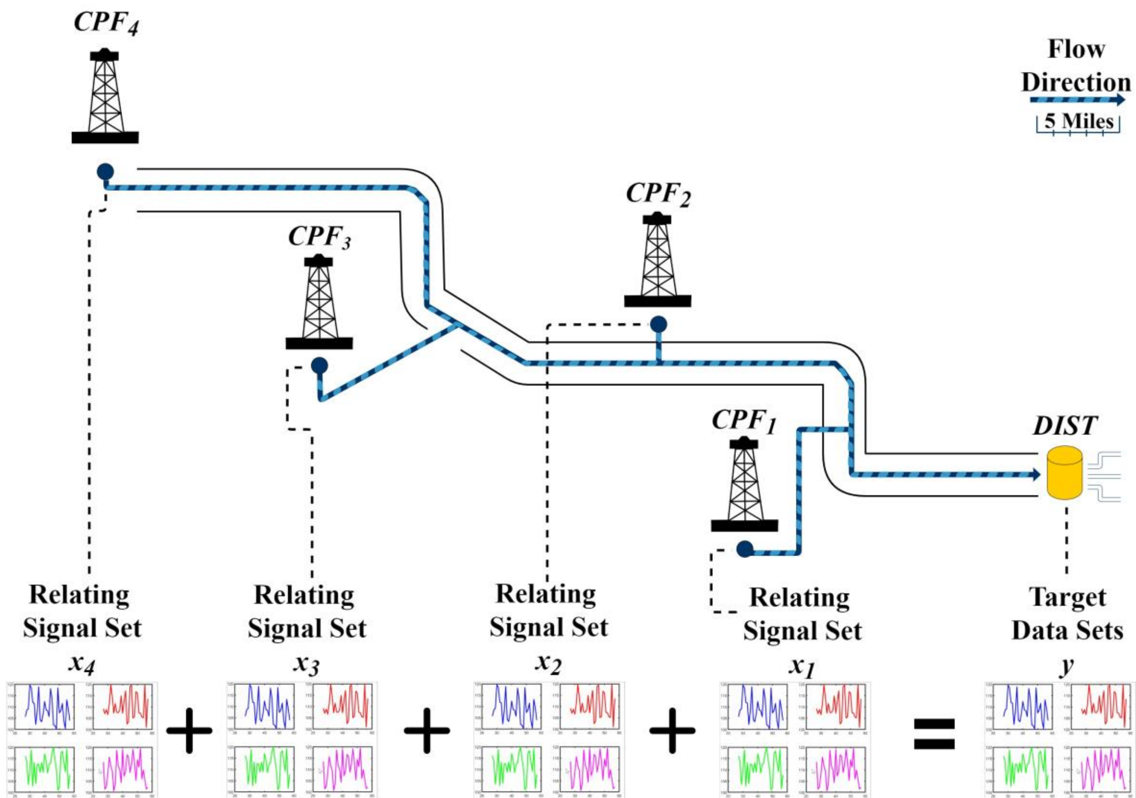

Figure 4.1: Notations of each central processing facility data set showing how each well is referenced in the alarm forecasting equations

The gas received at the distribution point is a function of the signals from the CPFs,

$$Y = F(X_1, X_2, X_3, X_4), \tag{4.1}$$

where X_1, X_2, X_3, and X_4 are the exogenous signals from CPF_1, CPF_2, CPF_3, and CPF_4, respectively. Similar to the notation used for the signal Y from the distribution point, X is subscripted to indicate the CPF from which it came and the type of signal. For example, $X_{2,Btu}$ is the BTU signal from CPF_2.

The target signal produced at the distribution point Y is modeled as lagged versions of the signals from the CPFs. Let l_1, l_2, l_3, and l_4 be the time it takes the natural gas to flow from CPF_1, CPF_2, CPF_3, and CPF_4 to the distribution point, respectively. Equation 4.2 is the model currently used by the control room operators to combine the lagged signals from the CPFs to predict the distribution point signal. Let w_i represent weights for each exogenous signal.

$$\begin{aligned} \hat{y}(t) = \; & w_1 x_1(t - l_1) + \; w_2 x_2(t - l_2) \\ & + w_3 x_2(t - l_3) + w_4 x_4(t - l_4), \end{aligned} \tag{4.2}$$

where $t >$ max (l_1, l_2, l_3, l_4).

Our algorithm forecasts pipeline signals 1 to 30 minutes into the future. After each forecast is made, the 30 estimates are compared against the alarm thresholds (Figure 3.1) for the predicted gas quality. If any of the forecasts cross the low, low-low, high, or high-high thresholds, an alarm is raised for that time horizon. It is possible to have several alarms in different signals being triggered at once. The term time horizon, represented with h, is used to indicate which of the 30 forecasted time horizons is triggering an alarm. Forecasted distribution point values h minutes into the future are denoted by $\hat{y}_{signal}(t + h)$. For example, the H_2O signal at the distribution point forecasted 10 minutes into the future is $\hat{y}_{H_2O}(t + 10)$.

The methods below are tools for the pipeline operator and have been developed with their requests in mind. Each method is implemented in an algorithm that fires every ten seconds to display the alarm forecasts to the pipeline operators via the SCADA system. Algorithm 4.1 shows the general algorithm for each forecasting technique as it operates on one signal.

```
1. Receive new time series values from SCADA system every ten seconds
2. Conduct anomaly detection; Impute if necessary
3. Enter new data with old data into forecaster
4. Compare forecasts against alarm thresholds:

   If(forecast ≥ High-high alarm threshold) : Trigger high-high alarm
   If(forecast ≥ High alarm threshold) : Trigger high alarm
   If(forecast ≤ Low alarm threshold) : Trigger low alarm
   If(forecast ≤ Low-low alarm threshold) : Trigger low-low alarm

5. Alert if necessary
```

Algorithm 4.1 — The alarm prediction algorithm in the SCADA system.

Algorithm 4.1 triggers a single alarm at once. A high-high alarm takes precedence over a high alarm, and a low-low alarm over a low alarm.

Now that the time series used in this work are defined, we split them into training and testing data sets.

4.3 Testing Data and the Naïve Model

A 70/30 percent split of each signal is used for training and testing the four forecasting methods explained below. Throughout the training and testing sets, the low frequency of triggered alarms creates an issue discussed in Chapter 6’s summary. In addition to the error metrics calculated for each model, a performance measurement known as the naïve model is implemented as a basis for comparison.

Equation 4.2 is an example of a simplistic model that is used to forecast signal data. Our first objective with each forecasting model is to outperform a basic, effective model. Hence, a naïve model is used as a basis for comparison. The naïve model for a general distribution point signal y at time horizon h is

$$\hat{y}_{signal}(t+h) = y_{signal}(t). \tag{4.3}$$

The naïve model simply uses whatever the signal's current value is at time t to be the forecasted value at $t+h$ [66].

This elementary forecasting technique is used for two reasons. First, is it is similar to how the control room operators manage the pipeline without the alarm forecasting algorithms. The easiest way to get an idea of the future is to consider the present, so control room operators use the most recent information available to make decisions. Second, despite the naïve model's simplicity, it is accurate enough to be used as a basis for comparison. If any of the models presented below cannot outperform the naïve model, it is considered unacceptable. This is first tested with the 10th-order autoregressive model.

4.4 10th-order Autoregressive Model (AR(10))

One of the first steps in many data analysis applications is performing regression analysis [65]. In time series applications, it is common to see autoregressive terms used in forecasting (Section 2.5). Let $y(t-n)$ represent a lagged value from the distribution point series y, and let $\hat{y}(t+h)$ be the forecasted value. A general autoregressive model is a function of values at previous time steps.

$$\hat{y}(t+h) = \mathrm{F}\big(y(t-1), y(t-2), \dots, y(t-p)\big), \tag{4.4}$$

where p is the number of lagged values. The value $y(t-1)$ is the signal value recorded one minute in the past.

The autoregressive model predicts future distribution point values, $\hat{y}(t+h)$, using the past ten minute's worth of signal data in y and weights $\vec{\beta}$. The order-P equation is

$$\hat{y}(t+h) = \sum_{p=1}^{P} (\beta_m y(t-p)) + \beta_0. \tag{4.5}$$

The equation's order, P, is 10.

The weights, $\vec{\beta}$, used in Equation (4.5) are found using least squares [67], minimizing the mean square of residuals

$$\frac{1}{n}\sum_{i=1}^{n}(y(t) - \hat{y}(t))^2 \tag{4.6}$$

with respect to $\{\beta_m; m = 1, 2, \cdots, M\}$ in Equation (4.6). A 10th-order AR model is trained for each time horizon, resulting in 30 models. Next, we incorporate exogenous variables.

4.5 10th-order Autoregressive Model with Exogenous Variables (ARX)

The next logical approach to modeling this system is to incorporate exogenous variables into the forecasting method. In this case, the exogenous variables are the lagged signals of the central processing facilities (CPF). Because each signal holds important historical information, autoregressive terms are incorporated for each exogenous signal. This combination of autoregressive terms and exogenous variables (ARX) form the second forecasting method.

Referencing Figure 4.1, let X_1, X_2, X_3, and X_4 be the signals of CPF_1, CPF_2, CPF_3, and CPF_4. Let l_1, l_2, l_3, and l_4 be the time it takes the natural gas to flow from

CPF_1, CPF_2, CPF_3, and CPF_4 to the distribution point, respectively. Let the weight matrix β be size five by ten, with the first row holding the weights for the autoregressive terms and the remaining rows the lagged exogenous signals. The autoregressive model with exogenous variables is

$$\begin{aligned} \hat{y}(t+h) \ = \ \sum_{m=1}^{M} (\beta_{1,m} y(t-m) \\ + (\beta_{2,m} x_1(t-l_1-m) \\ + (\beta_{3,m} x_2(t-l_2-m) \\ + (\beta_{4,m} x_3(t-l_3-m) \\ + (\beta_{5,m} x_4(t-l_4-m). \end{aligned} \tag{4.7}$$

Lags $l_1, \ldots, l_4$ are found by comparing a lagged cross-correlation between each signal coming from $CPF_1 - CPF_4$ and the signal at the distribution point. Figures 4.3 and 4.4 show an example of a lagged signal.

Figure 4.2 depicts a lag in the H_2O signal data from CPF_1 (blue line) to the distribution point (black line). The red line is used to help show that when the X_1 H_2O signal increases, a similar increase is seen in the distribution point's signal l_1 minutes later.

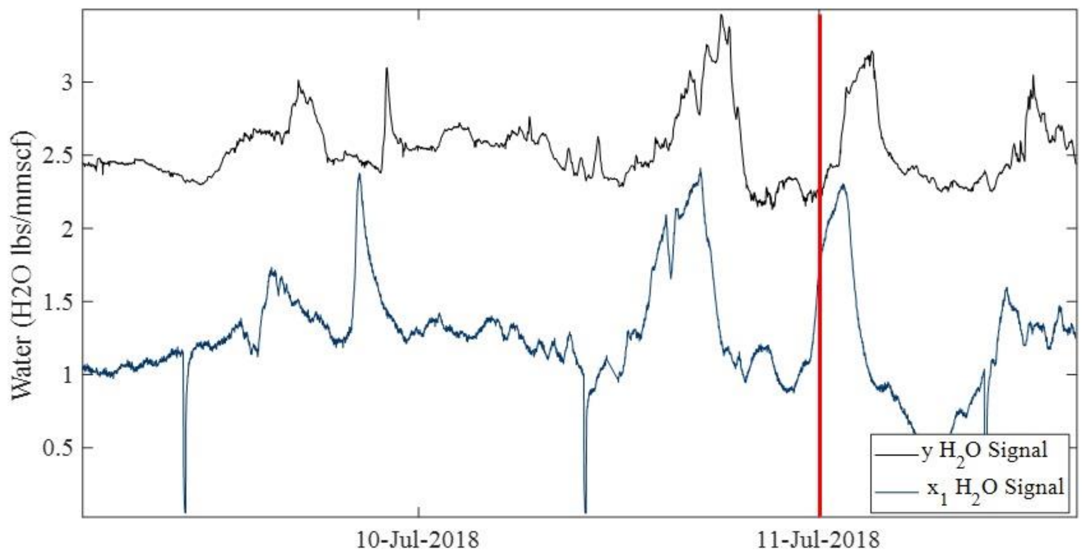

Figure 4.2: Visualizing a lag between CPF_1 and the distribution point's H_2O signals

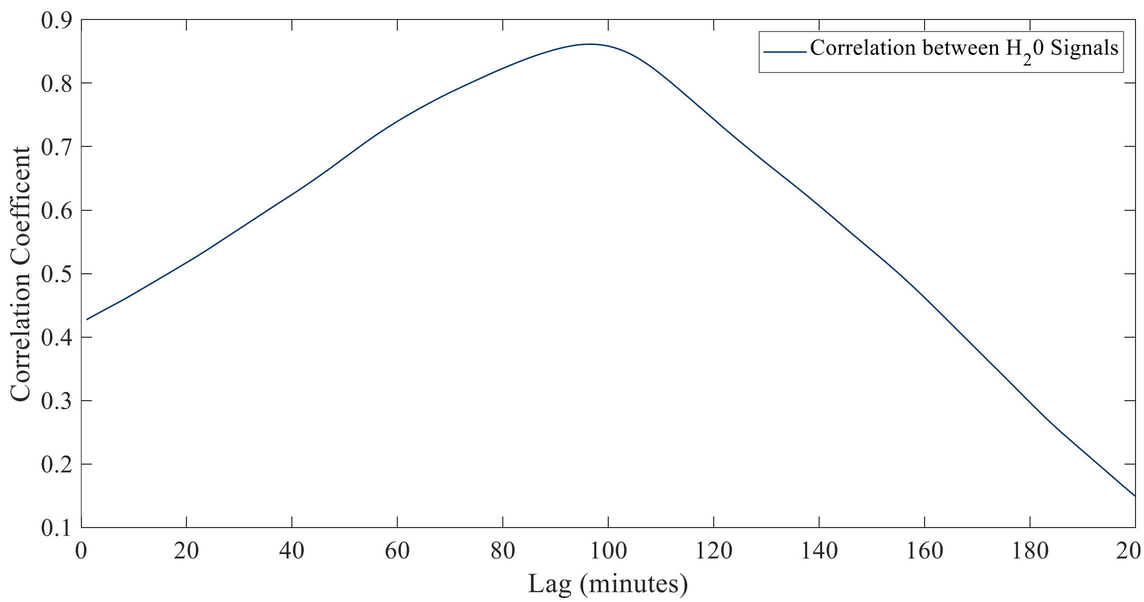

Figure 4.3: Lagged cross-correlation between $X_{1_{H2O}}$ and Y_{H2O}

Figure 4.3 shows the lagged cross-correlation plot between $\mathrm{X}_{1_{\mathrm{H2O}}}$ and $\mathrm{Y}_{\mathrm{H2O}}$. We choose lag l_1 to be the minute with the highest correlation, which is seen at 97 minutes in this example. Table 4.1 summarizes the time it takes for gas entering the pipe at any CPF to reach the distribution point. The l_i's in Equation (4.7) are found under the column 'Lag Time to Distribution' measured in minutes for each respective gas quality.

Weight matrix $\beta_{5,10}$ is structured so that each row of 1x10 weights are used in a linear combination of the 10 lagged values of each CPF time series. The first column vector $\beta_{:,1}$ consists of a bias vector of 1's to consider the intercept of each series.

$$\beta_{5,10} = \begin{bmatrix} 1 & \beta_{1,2} & \cdots & \beta_{1,10} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \beta_{5,2} & \cdots & \beta_{5,10} \end{bmatrix} \quad \begin{matrix} \text{\} Weights used with series } Y \\ \vdots \\ \text{\} Weights used with series } X_4. \end{matrix} \tag{4.8}$$

Equation 4.9 illustrates how to calculate the forecasted BTU at the distribution point by putting together all the notation defined in this section.

$$\begin{aligned} \hat{y}_{Btu}(t+h) \ = \ \sum_{m=1}^{M} (\beta_{1,m} y_{Btu}(t-m) \\ + (\beta_{2,m} x_{1_{Btu}}(t - l_1 - m) \\ + (\beta_{3,m} x_{2_{Btu}}(t - l_2 - m) \\ + (\beta_{4,m} x_{3_{Btu}}(t - l_3 - m) \\ + (\beta_{5,m} x_{4_{Btu}}(t - l_4 - m). \end{aligned} \tag{4.9}$$

The l_i's are gathered from Table 4.1, which illustrates that the further away the CFP, the longer the time the gas takes to travel to the distribution point.

Table 4.1: The amount of time it takes for a signal's behavior to effect the signal being generated at the distribution sales point

CPF	**Average Lag Time to Distribution (L_i - minutes)**	**Average Calculated Correlation**
X_1	92	0.78
X_2	105	0.62
X_3	173	0.41
X_4	209	0.26

The next forecasting method does not use any autoregressive terms as inputs. Instead, the Theta method considers recent trends in the signal data to make forecasts.

4.6 Simple Exponential Smoothing with Drift (Theta Method)

The Theta Method [77], otherwise known as simple exponential smoothing with drift, is used to forecast signals at the distribution point. The unoptimized Theta method is

$$\hat{y}(t+h) = 0.5 * \widehat{L0}(t+h) + 0.5 * \widehat{L2}(t+h) \tag{4.10}$$

where $\hat{y}(t+h)$ is the forecasted distribution sales point series value h time steps ahead of current time t. $\widehat{L0}(t+h)$ represents a forecasted long-term component extracted from the data h time steps ahead of current time t, and $\widehat{L2}(t+h)$ is the forecasted short-term component extracted from the data h time steps ahead of current time t. Forecasting is carried out by an equal weighted combination of these two components.

$L0(t)$, the first Theta line, is calculated with the linear trend estimates b_o and b_1,

$$L0(t) = b_o + b_1(t). \tag{4.11}$$

Once completed, the second Theta line is found using $L0(t)$ in Equation (4.12).

$$L2(t) = 2 * \ y(t) * L0(t). \tag{4.12}$$

With both Theta lines defined, $\widehat{L0}(t+h)$ and $\widehat{L2}(t+h)$ are calculated to get the final forecast $\hat{y}(t+h)$. $\widehat{L0}(t+h)$ is found in the same way as $L0(t)$ with the exception that $\widehat{L0}(t+h)$ is forecasting h steps ahead.

$$\widehat{L0}(t+h) \ = b_o + b_1(t+h). \tag{4.13}$$

$\widehat{L2}(t+h)$ is forecasted using simple exponential smoothing and incorporates an optimized smoothing parameter α,

$$\widehat{L2}(t+h) \ = \alpha(L2(t)) + (1-\alpha)\widehat{L2}(t)\ . \tag{4.14}$$

The Theta method, 10^{th}-order autoregressive model, and 10^{th}-order autoregressive model with exogenous variables are linear models used to forecast the different signals at the distribution point. The next method, an artificial neural network, introduces a nonlinear aspect to this work.

4.7 Artificial Neural Network

An artificial neural network (ANN) is applied to this work in a manner similar to past applications seen in the energy industry (Section 2.4). We use a one hidden-layer, five-node ANN in this work, as suggested in Figure 4.5. The hidden layer neurons are sigmoid. The output layer has a linear transfer function.

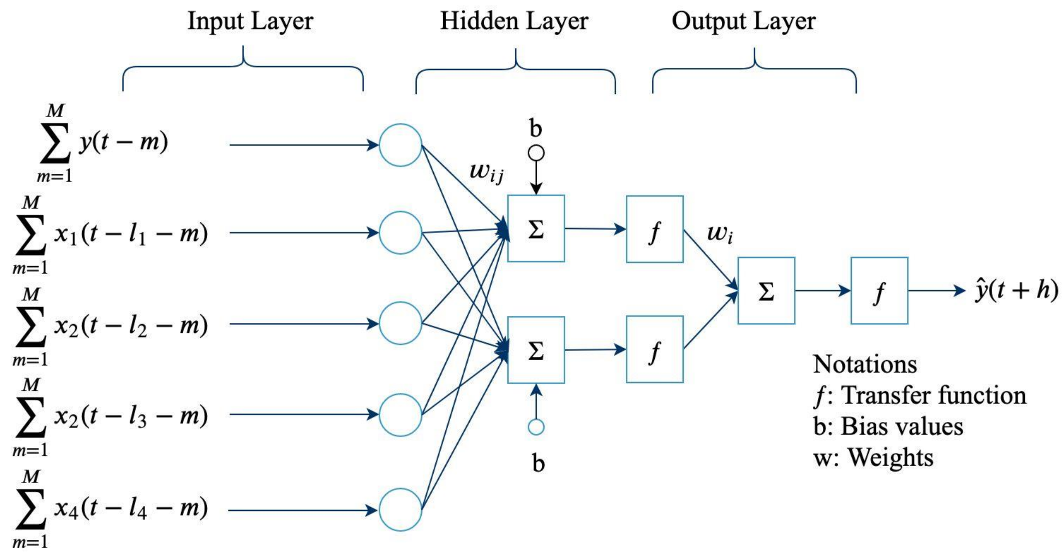

Figure 4.4: ANN feed forward architecture with one hidden layers and 40 nodes

The term neuron follows the traditional naming convention of ANN's and represents the interconnected processing elements grouped in layers. The ANN uses the same inputs as the ARX model, with each input node being ten minutes of lagged signal data from the distribution point, $CPF_1, CPF_2, CPF_3,$ and CPF_4. The ANN outputs a single node, $\hat{y}(t+h)$, which is the forecasted value at the current time plus the time horizon, $t+h$.

Following a feed-forward architecture, each minute's input is modified and summed in the hidden layer. The hidden layer in the ANN consists of a sigmoid transfer function,

$$\mathrm{sigmoid}(x) = \frac{1}{1+e^{-x}}. \tag{4.15}$$

The sigmoid transfer function takes the values provided by the input nodes and produces a scalar that is combined with weights $w_{i,j}$. These combinations are summed, and appropriate bias values, b, are applied. The output layer consists of a linear activation function

$$\mathrm{linear}(x) = x. \tag{4.16}$$

The linear activation function, as in the regression-based models, transforms the weighted observations from the hidden layer to obtain a continuous output that represents the forecasted signal value. Finally, we use a Levenberg-Marquardt (LM) training algorithm to train the ANN [89]. The LM algorithm iteratively optimizes the network weights and bias values using a sum of squares error loss function (Equation 4.6).

Chapter 4 presents the forecasting framework and models used in this work. The forecasting framework is developed to be compatible with the existing SCADA infrastructure, easily expandable, and require little computational resources to operate. The 10th-order autoregressive, autoregressive with exogenous variables, Theta method, and artificial neural network makes forecasts 1 to 30 minutes into the future. With the forecasting methods defined, Chapter 5 presents an empirical analysis to determine the forecasting model that best fits the need of the production company.

CHAPTER 5

Empirical Results and Discussion

5.1 Chapter Objectives

Chapter 5 presents discussion, interpretation, and comparison of the experimental results. We present the error metrics used to determine each model's accuracy along with explanations of why each metric is relevant and an appropriate choice. We compare the forecasting models to the naïve model and state its performance across all time horizons. Numerical and graphical performance reviews are given to compare each forecasting model with the others.

5.2 Error Metrics

Each pipeline alarm forecasting algorithm predicts alarms at time horizons 1 through 30 minutes. We predict alarms by forecasting the value of each gas quality and then detecting if the quality threshold is met by the forecast (Algorithm 4.1). This threshold check is done for each time horizon. Two measures are used to summarize the performance of the algorithm: root-mean-square error (RMSE) and mean-absolute-percentage error (MAPE). These error metrics demonstrate the accuracy of the gas quality forecasts.

RMSE is a measure of the average residual. Let $y(t)$ be the value of signal y at time t, $\hat{y}(t)$ be the predicted quality at time t, and T be the total number of forecasts. The RMSE is

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{T}(\hat{y}(t) - y(t))^2}{T}}. \tag{5.1}$$

The RMSE is a good measure of error but can be overly influenced by outliers. RMSE is measured in the units of signal y.

We also present the MAPE, which measures the average absolute error. It is unitless and represents the percent of error.

$$\mathrm{MAPE} = \frac{1}{T}\sum_{i=1}^{T}\frac{|(\hat{y}(t) - y(t))|}{y(t)}. \tag{5.2}$$

RMSE and MAPE are based on the continuous output of each model, meaning they do not represent how well each model forecasts alarms. To measure binary presence of an alarm, we use sensitivity as a metric. The sensitivity is a measure of how many alarms are correctly predicted out of all possible alarms. Let *TP* (true-positive) be the number of alarms that are correctly predicted, and *FN* (false-negative) be the number missed alarms. Sensitivity is then found with equation 5.3.

$$\mathrm{Sensitivity} = \frac{TP}{TP + FN}, \tag{5.3}$$

We present the sensitivity because the number of non-alarms is far greater than the number of alarms. Summarizing Tables 3.1 – 3.5, the average of each alarm considering each signal is presented in Table 5.1.

Table 5.1: General frequencies of all alarms considering all time series signals

Alarm Frequency (%)						
Alarm	PSI	BTU	H_2S	CO_2	H_20	**Avg. Alarm Frequency (%)**
High-high	0.14	0.25	0.18	2.85	1.52	0.988
High	1.2	0.37	0.92	4.67	2.78	1.988
Low	1.55	1.31	3.31	1.23	0.2	1.52
Low-low	0.85	0.27	1.82	0.13	0.02	0.618

The high alarm is triggered most frequently, but it only occurs approximately 2% of the time. Since all the methods in this work are based on regression analysis, and that an alarm is either triggered or not, sensitivity helps us obtain statistical measures of this binary classification.

The experimental results are shown in two respects. The first considers the numerical output of each forecasting method in comparison with the true values of the series (RMSE, MAPE). The second considers the binary classification of the presence of an alarm (sensitivity). The five natural gas pipeline signals being used in this work are forecasted 1-30 minutes into the future. We select results from time horizons 1, 15, and 30 minutes to compactly compare the performance of each model. This is shown in Figure 5.2. The sensitivity of the forecasted alarms is plotted across all 30 time horizons to show that out of all the alarms that were truly triggered, how many each forecasting method correctly predicted. To provide the production pipeline operators the most amount of warning, the 30-minute time horizon is an important upper-bounding case of the performance of each model.

5.3 Naïve Model Results

Section 4.2 describes the naïve model and how it is used as a baseline performance metric. This section displays the naïve model results and acts as an introduction to understanding the result plots used for each of the forecasting methods (Sections 5.4 – 5.9). We present the MAPE and sensitivity of the naïve model below for the pressure target data set (the signals recorded at the distribution point). In the empirical results section (Section 5.4), we see that the naïve model is a comparable alternative to the purposed forecasting methods.

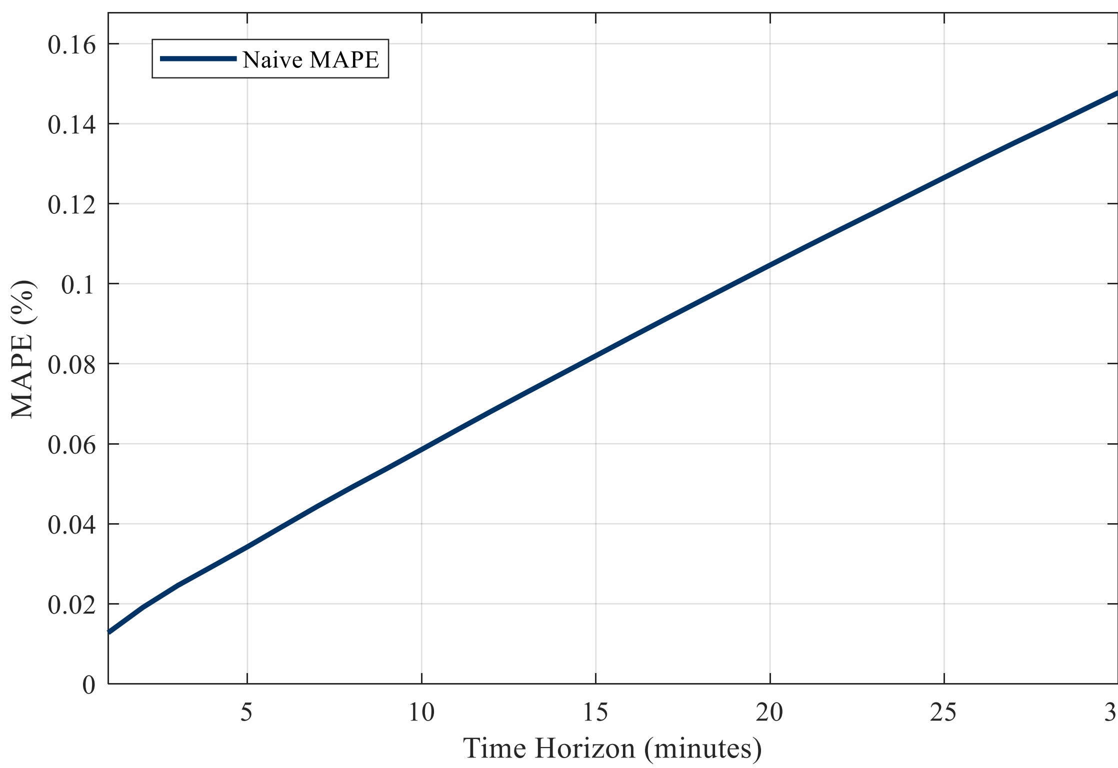

Figure 5.1: Naïve model MAPE for pressure time series over all time horizons

Figure 5.1 shows the MAPE over each time horizon of the naïve model forecasting the pressure at distribution point. The naïve model's MAPE increases almost linearly with the time horizon. This is expected since the naïve model simply uses the current value as the forecasted value at time horizon 30.

The sensitivity plot shows the true-positive rate of the naïve model over time. Figure 5.2 shows how the naïve model predicts the four pressure alarm thresholds over all time horizons.

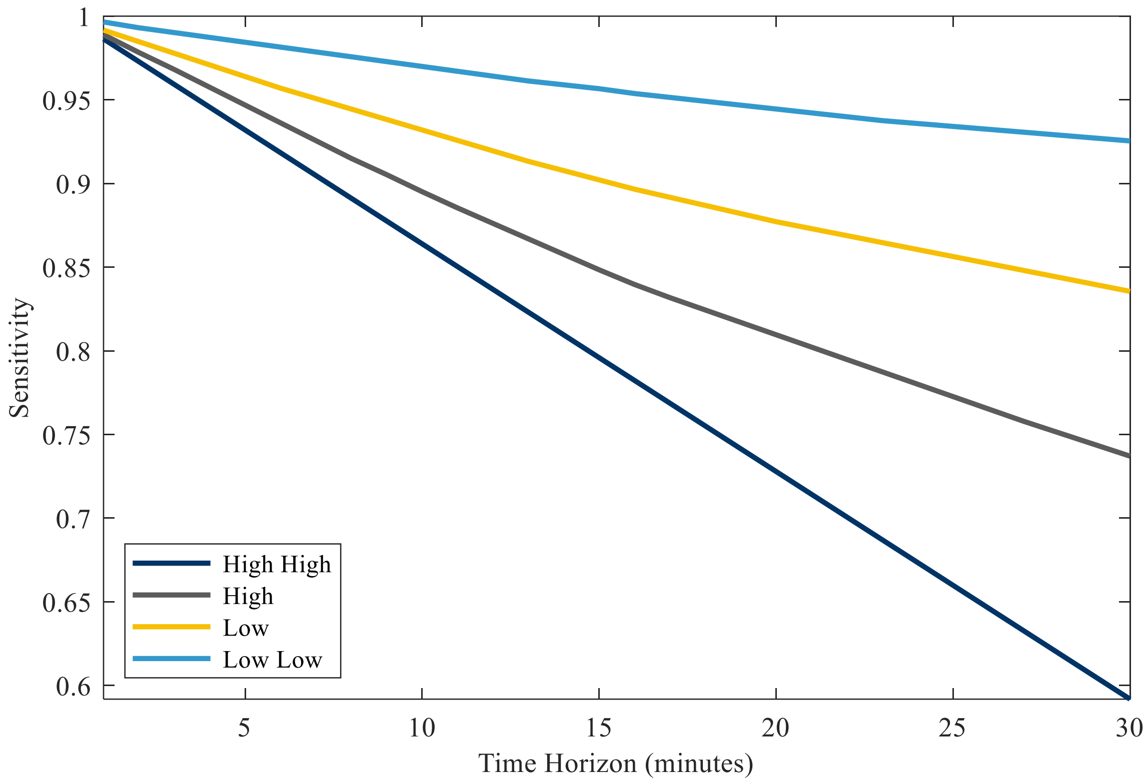

Figure 5.2: Naïve model pressure (psi) sensitivity w.r.t time horizon

The sensitivity shows how many alarms were predicted correctly at each time horizon where 1 indicates 100%. Here we see the alarm prediction accuracy falls approximately linearly from near 100% accuracy at 1 minute to 60-95% accuracy at 30 minutes. The naïve model performed the worst for the high-high alarm forecasts, while the low-low alarms were predicted correctly 95% of the time. When evaluating the five sets of models below, only the high alarm sensitivity rate is considered. This is because the high alarm is the most frequently triggered alarm out of all the datasets (Table 5.1) and pipeline controllers are generally more concerned with high alarms for each signal than the low. Higher pressure, H_2S, CO_2, and H_2O contents indicate serious conditions in the pipeline that require the immediate attention of the controller, while their corresponding low alarms are not as urgent. We begin the analysis of results below.

5.4 Empirical Results of all Models

This section presents the empirical results of the four model types described in Chapter 4: 10^{th}-order autoregressive (AR(10)), autoregressive with exogenous variables (ARX), Theta, and artificial neural network (ANN) forecasting models. Table 5.2 summaries the numerical findings for these models. These results measure the ability of each model to forecast future signal values and do not represent alarm forecasting accuracy. For each signal in Table 5.2, the best performing forecasting model is highlighted in blue.

Two models consistently outperform all others. The ARX and ANN are the best predictors at time horizon 30. For the pressure signal, the ARX performs best at time horizon 30, while the ANN forecasts with higher accuracy at the 15-minute mark. For all other signals, the ANN performs best at time horizon 30.

Table 5.2: The results of each forecasting method at time horizons 1, 15, and 30 minutes

PSI						
	Time Horizon = 1 min.		Time Horizon = 15 min.		Time Horizon = 30 min.	
Model	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
Naïve	0.23	0.01	1.68	0.08	3.06	0.15
AR	0.19	0.01	1.07	0.06	2.07	0.11
ARX	0.23	0.01	0.94	0.05	1.29	0.07
Theta	0.22	0.01	1.55	0.08	2.82	0.14
ANN	0.19	0.01	0.83	0.05	1.45	0.08
BTU						
	Time Horizon = 1 min.		Time Horizon = 15 min.		Time Horizon = 30 min.	
Model	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
Naïve	0.56	0.01	2.77	0.12	3.84	0.17
AR	0.60	0.01	2.33	0.09	3.21	0.13
ARX	0.48	0.01	1.71	0.07	2.94	0.10
Theta	0.52	0.01	2.57	0.11	3.56	0.16
ANN	0.62	0.01	3.72	0.05	2.62	0.06
H_2S						
	Time Horizon = 1 min.		Time Horizon = 15 min.		Time Horizon = 30 min.	
Model	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
Naïve	0.04	0.48	0.14	5.73	0.14	6.18
AR	0.02	0.48	0.09	4.86	0.11	7.01
ARX	0.02	0.47	0.08	4.20	0.10	5.24
Theta	0.02	0.48	0.10	5.28	0.12	5.71
ANN	0.02	0.47	0.03	3.59	0.03	4.02
CO_2						
	Time Horizon = 1 min.		Time Horizon = 15 min.		Time Horizon = 30 min.	
Model	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
Naïve	0.02	0.34	0.06	3.49	0.08	5.08
AR	0.01	0.36	0.05	3.05	0.05	4.11
ARX	0.01	0.40	0.05	3.37	0.06	4.68
Theta	0.01	0.32	0.06	3.24	0.07	4.71
ANN	0.01	0.43	0.02	1.15	0.02	1.67
H_2O						
	Time Horizon = 1 min.		Time Horizon = 15 min.		Time Horizon = 30 min.	
Model	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
Naïve	0.02	1.03	0.06	1.95	0.10	2.87
AR	0.02	0.62	0.05	1.69	0.09	2.60
ARX	0.02	0.82	0.06	1.88	0.09	2.69
Theta	0.02	0.95	0.06	1.82	0.09	2.67
ANN	0.03	1.72	0.04	2.06	0.05	2.43

While the ARX and ANN models are the overall best forecasters at time horizon 30, there are instances where other models are better at shorter time horizons. Sections 5.5-5.10 provide an in-depth analysis of the results for each signal over all time horizons and compares the forecasters visually.

5.5 Pressure (psi) Signal Alarm Forecasting Results

The control room operators labeled the pressure signal to be the most critical pipeline variable. We test our pressure forecasting models and compare their performance. Figure 5.3 shows each mode's MAPE performance across all time horizons.

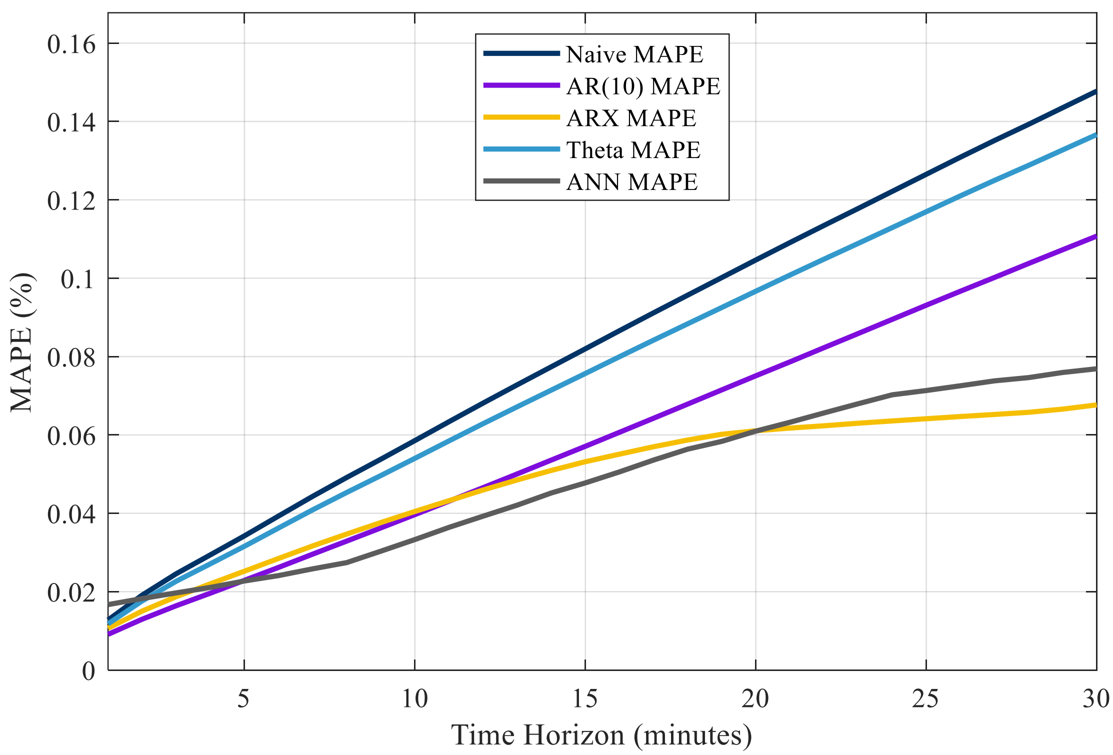

Figure 5.3: MAPE of all pressure forecasting models over the 30 time horizons

All pressure models outperform the naïve model as measured by MAPE. In time horizons 1-15, the simple autoregressive model (AR10), artificial neural network (ANN), and autoregressive model with exogenous variables (ARX) do the best short-term forecasting. In the later time horizons, we see the ARX and ANN models perform the best. Both these models use inputs from exogenous variables. This confirms that the lagged pressure signals coming from the upstream central processing facilities (CPF) have a significant influence on the pressure signals being recorded at the distribution point. It should also be noted that CPF_2 does not have enough pressure data to be included as an exogenous variable, so that signal is excluded from training the pressure models (Equation 4.7). Next, we evaluate each pressure model's high alarm forecasting accuracy.

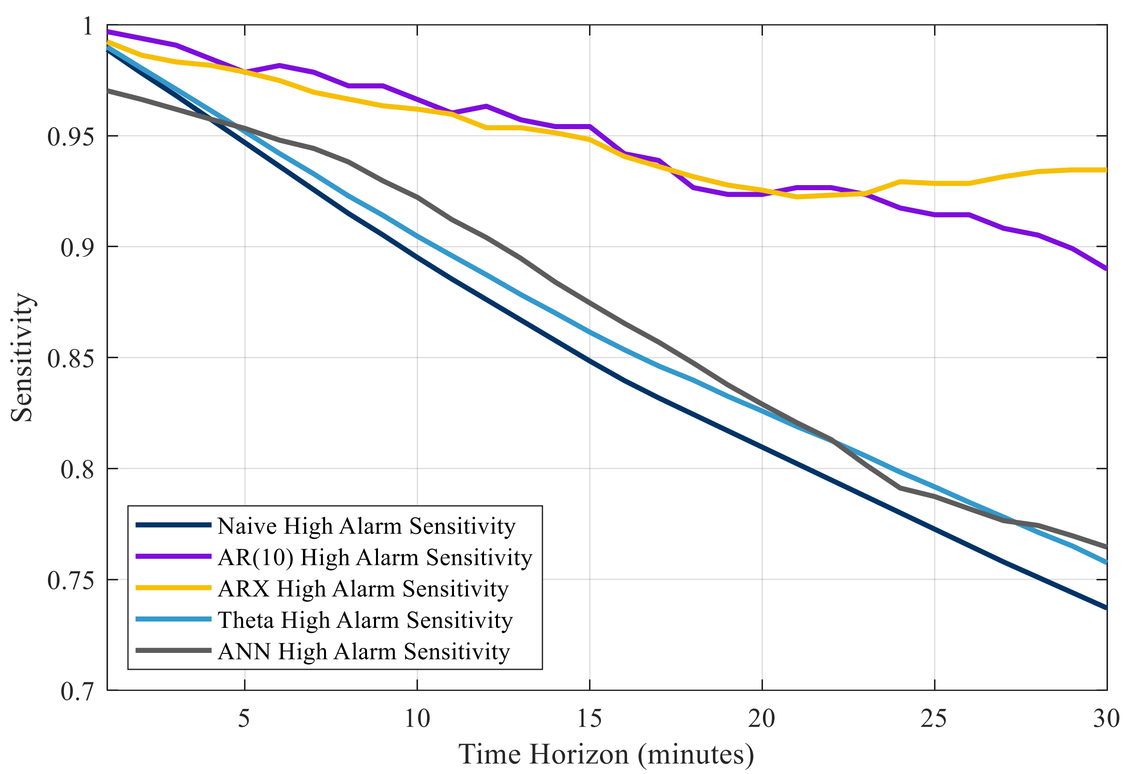

Figure 5.4: High alarm sensitivity of all pressure forecasting models over the 30 time horizons

A higher sensitivity rate translates to more alarms being predicted correctly. In Figure 5.4, the ARX model (yellow line) correctly predicts 93.3% of the alarms at time horizon 30. Because the pipeline control operators need as much time as possible to react to a triggered alarm, the farthest time horizon is the most important. The results from Figure 5.4 indicate that the ARX model is the best overall pressure forecasting model for time horizons 24-30.

The sensitivity plot contradicts the MAPE results seen in Figure 5.3. The ANN has the lowest MAPE scores. However, it does poorly when forecasting high alarms according to the sensitivity plot. The ANN struggles to forecast alarm-triggering pressure events, but does well when forecasting the steady-state signal (when no alarms are present). This quality of the ANN is unhelpful to the pipeline controllers and suggests that the ARX model is the best pressure signal forecaster. Figure 5.4 illustrates an example where the sensitivity plot provides more useful information than the MAPE plot when trying to identify appropriate models.

For each signal in this chapter, it is important to consider the frequency of alarm occurrences per threshold (Table 3.1). The testing dataset for the pressure time series contained the largest number of high alarms, so the sensitivity rate of the high alarm threshold is the best threshold to test for the pressure models. In other testing datasets, a different alarm threshold may be a more logical choice to test. The heat content forecasting model is an example of this. The low alarm threshold is the most important for monitoring the gas heat content, but the lack of these alarms in the testing data set results in the high threshold being used to calculate the model’s sensitivity.

5.6 Heat Content Signal (BTU) Alarm Forecasting Results

The heat content (BTU) signal also is prioritized by the production company as an essential signal when maintaining flow assurance. Figure 5.3 shows how each BTU forecasting model performs at the 30 different forecasting horizons in terms of MAPE.

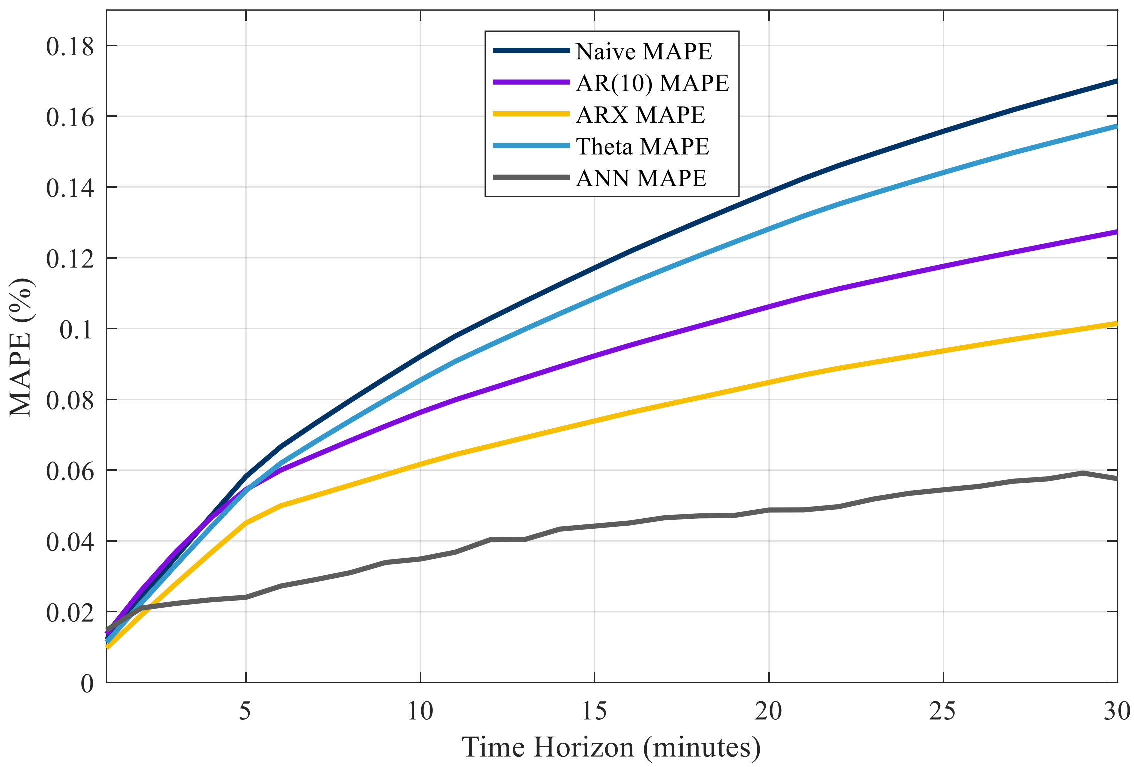

Figure 5.5: MAPE of all BTU forecasting models over the 30 time horizons

Figure 5.5 shows the ANN and ARX are the two best BTU forecasting models when considering MAPE. Of the two, the ANN outperforms the ARX over time horizons 2-30. We verify if the ANN remains the best performer when forecasting the high alarms by examining the sensitivity in Figure 5.6.

Unlike the pressure model ANN, the BTU ANN is a top performer when looking at the sensitivity of high BTU alarms. Considering the ANN's MAPE and sensitivity rate, the BTU ANN is the best model for forecasting pipeline alarms. The AR model outperforms the ARX model, which signifies that the exogenous variables for BTU are not as significant as in the pressure model. The Theta method was one of the worst performers, only correctly identifying 75% of the high alarms at time horizon 30.

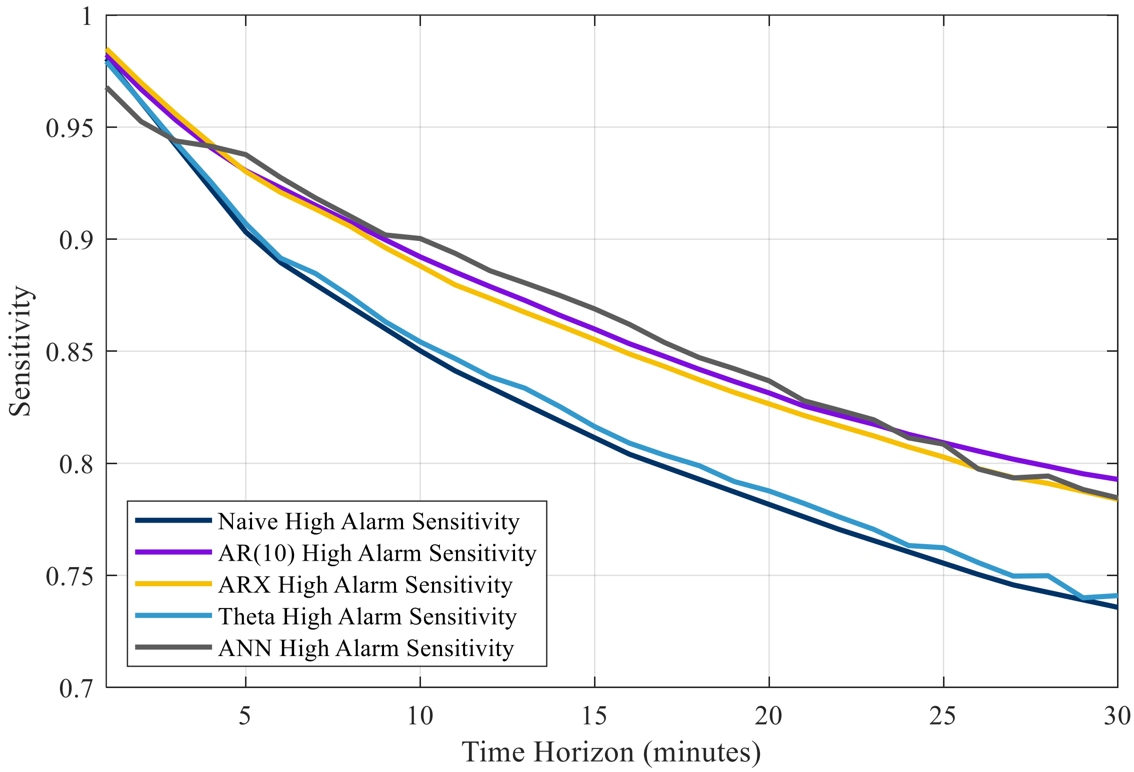

Figure 5.6: High alarm sensitivity of all BTU forecasting models over the 30 time horizons

For the BTU signal, low alarms are more important to pipeline operators than high alarms. As opposed to pressure, where a high alarm requires the immediate attention of the pipeline controllers, high heat content indicates high-quality gas is in the pipe and does not pose an immediate threat. Yet, a low BTU signal can violate the agreement held between the production company and buyer, causing the buyer to shut in the pipeline. Ideally, we want to test the low alarm's sensitivity, but because the testing set does not include any low alarms we use the high alarm threshold to evaluate each model's alarm sensitivity.

Despite the pressure and BTU signals being identified as the most important signals by the pipeline controllers, the remaining signals discussed in this chapter still have the potential to shut in the pipeline. However, the hydrogen sulfide, carbon dioxide, and moisture content signals are more easily influenced by human intervention, and our data cleaning process identified significantly more outliers in these data sets. This is due to the pipeline operators having more control over the processing equipment that controls these signals and the general volatility of each signal. Next, we describe the results of modeling the remaining signals.

5.7 Hydrogen Sulfide (H_2S) Signal Alarm Forecasting Results

Figure 5.7 shows the MAPE of each H_2S model. The H_2S MAPE scores are higher than the pressure or BTU model MAPEs. This higher error rate is also seen in MAPE scores for the CO_2 and the H_2O signals (Figure 5.9 and 5.11). By normalizing each signal and calculating its variance, we find that the H_2S, CO_2, and H_2O signal variability is higher than the pressure or BTU variability, thus explaining the higher MAPE range.

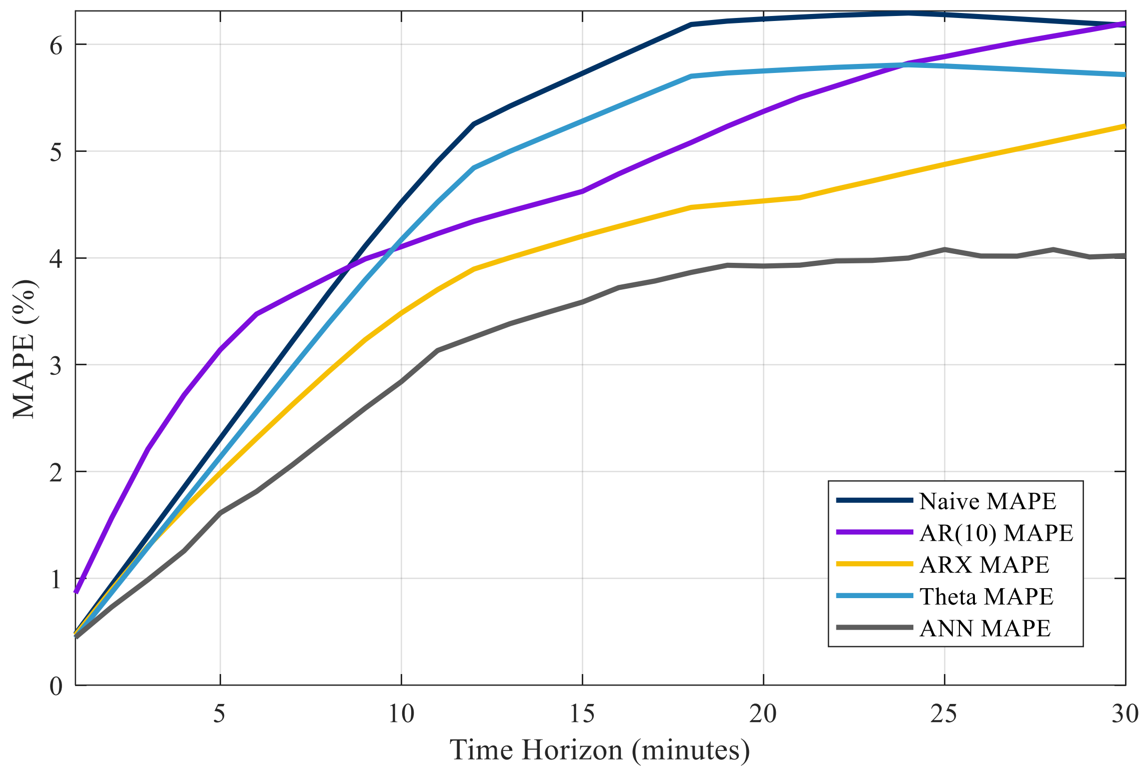

Figure 5.7: MAPE of all H_2S forecasting models over the 30 time horizons

All H_2S alarm forecasting models outperform the naïve model, except that the AR(10) model fails over time horizons one through eight. MAPE errors range from 4.0 - 6.2%, with the ANN performing the best at 4.1%. The AR(10) model performs the worst, exceeding 6% at time horizon 30.

Figure 5.8 shows the sensitivity of high alarms for each H_2S model. The sensitivity plot confirms the ANN to be the best performing H_2S forecasting model. However, all other models fail to predict more than 70% of high alarms correctly at time horizon 30, perhaps because there are few high alarms in the H_2S testing dataset. Table 3.3 shows the frequency of high alarms is less than 1%. With such a low occurrence rate of high alarms, the sensitivity is not the best criteria by which to evaluate the H_2S alarm forecasting models. Therefore, the MAPE plot (Figure 5.7) is considered to be a better representation of the forecasting performance of the H_2S model.

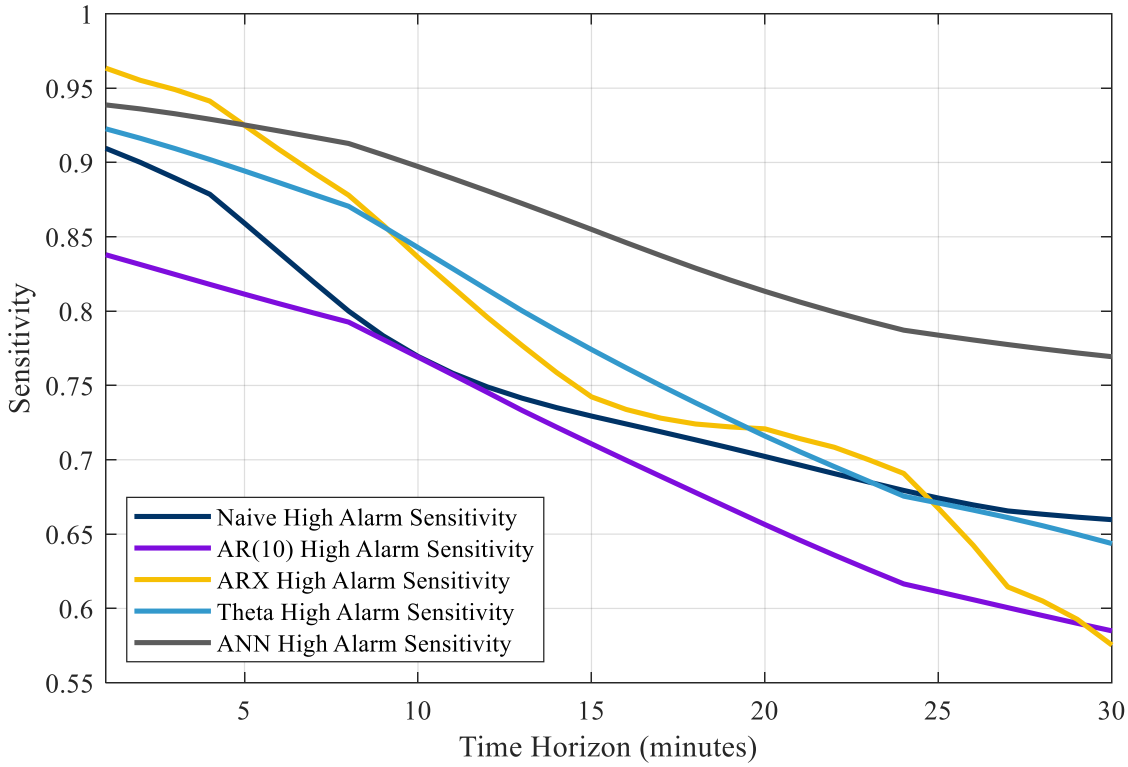

Figure 5.8: High alarm sensitivity of all H_2S forecasting models over the 30 time horizons

While the results for the H_2S models are generally not as good as the pressure or BTU models, the ANN forecasting 76% of the high alarms out of the few occurrences that exist in the testing signal is promising. These results are similar as those seen for the CO_2 models presented next.

5.8 Carbon Dioxide (CO_2) Signal Alarm Forecasting Results

As discussed in Section 3.5, the CO_2 signal is directly related to the H_2S content signal. Figure 5.9 presents the MAPE plot of all CO_2 forecasting models below.

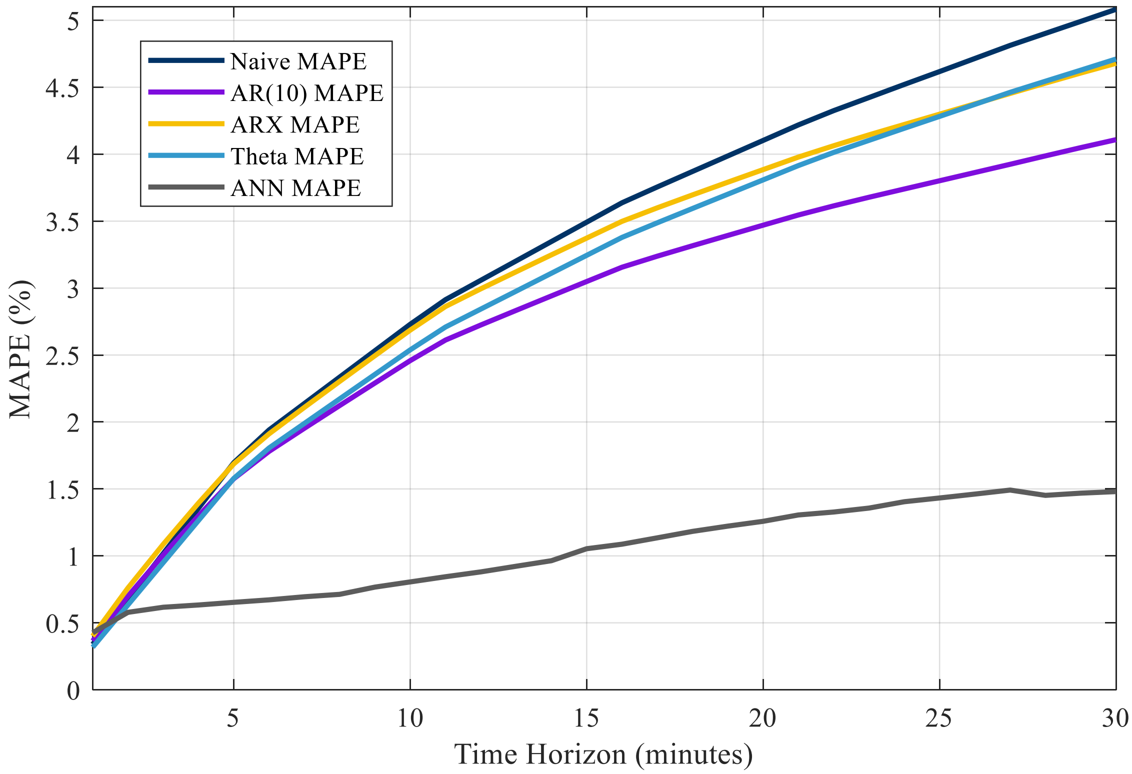

Figure 5.9: MAPE of all CO_2 forecasting models over the 30 time horizons

Figure 5.9 shows that the ANN is the best model to forecast the CO_2 signal. We see that the naïve, AR(10), ARX, and Theta model results are grouped together over the 30 time horizons and forecast a maximum 4.0 – 5.2% error. While all the CO_2 models outperform the naïve basis comparison model, only the ANN performs under 1.5% MAPE error at time horizon 30.

Figure 5.10 shows the CO_2 low alarm sensitivity rate.

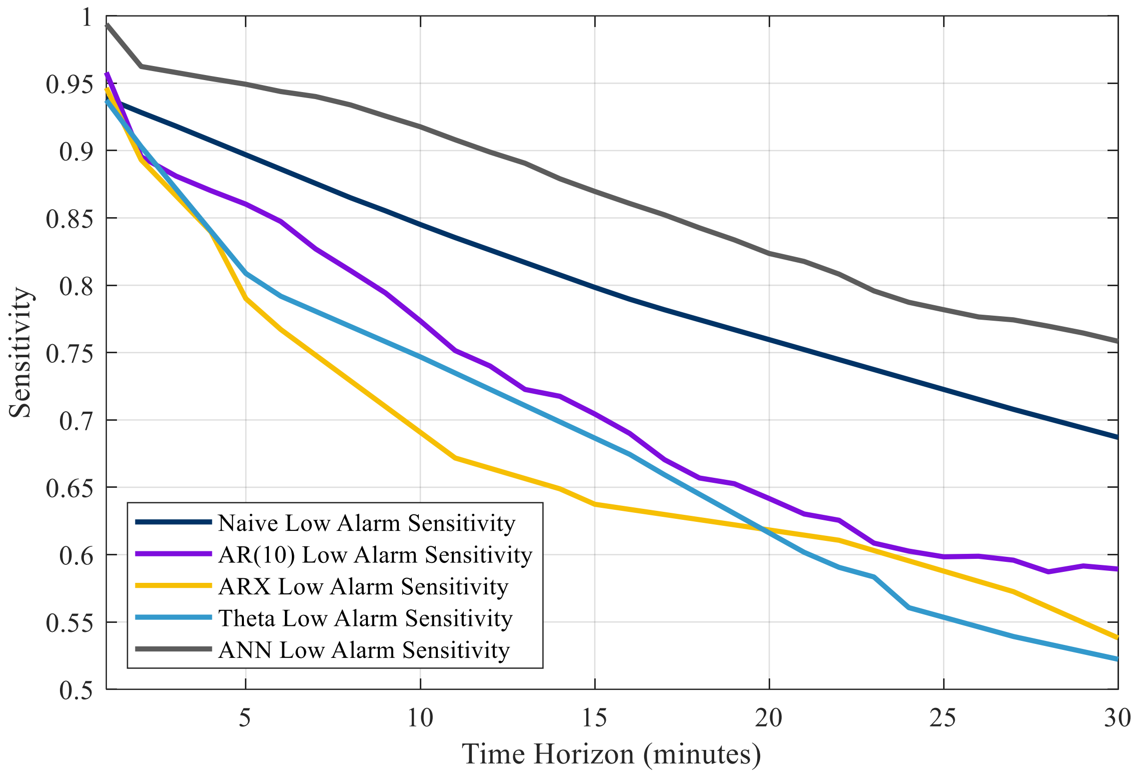

Figure 5.10: Low alarm sensitivity of all CO_2 forecasting models over the 30 time horizons

The naïve model outperforms all but the ANN in the sensitivity plot. The worst performer, the Theta method, forecasts low alarms just above a 50% success rate. The Theta methods forecast is a combination of a local curvature forecast and linear trend. The weighting of these components are of equal parts, but these results indicate that the Theta model would benefit from a heavier weighting on curvature based on the volatility of the CO_2 signal. Figure 5.10 introduces the lowest sensitivity rates seen in any signal thus far. The final signal, H_2O, is examined next.

5.9 Moisture Content Signal (H_2O) Alarm Forecasting Results

The moisture content (H_2O) signal contains periodic characteristics and has a testing set that contains no high or high-high alarms. The MAPE of each H_2O forecasting model is presented below in Figure 5.11.

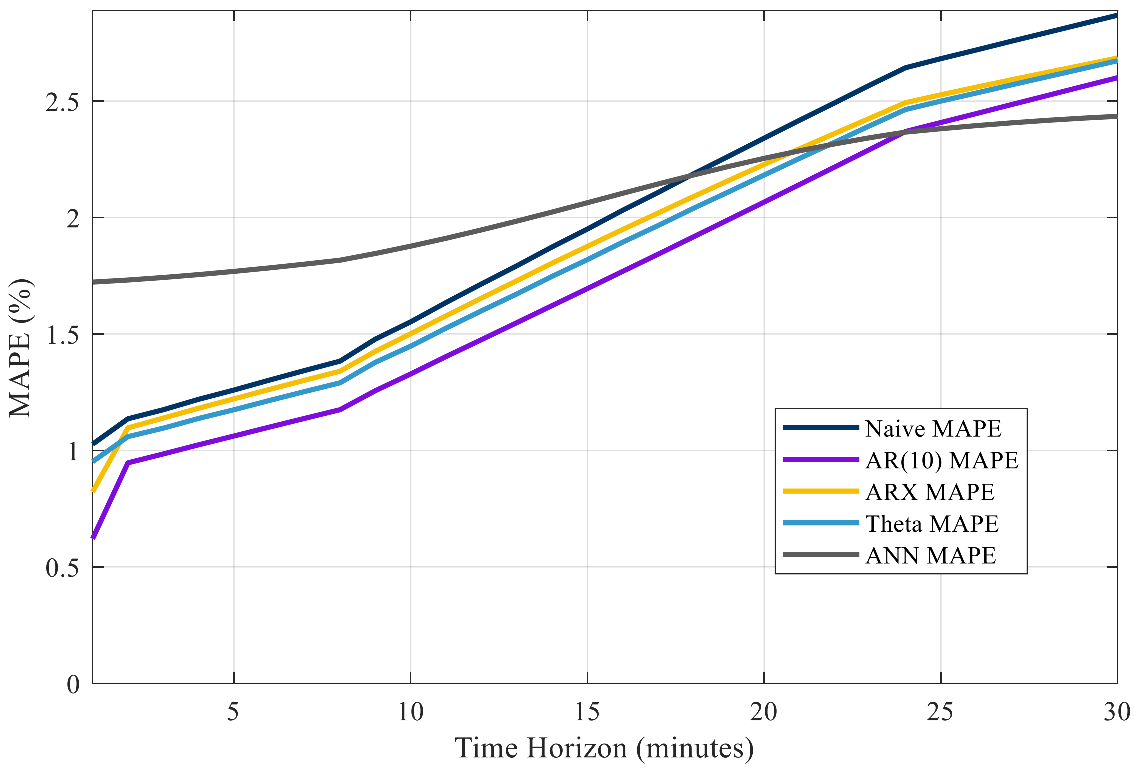

Figure 5.11: MAPE of all H_2O forecasting models over the 30 time horizons

Due to a lack of high or high-high alarms, the low alarm will be analyzed with the sensitivity plot. Despite a high alarm being more concerning to a pipeline operator, the low alarm sensitivity gives an approximation of how well the models forecast general H_2O signal alarms.

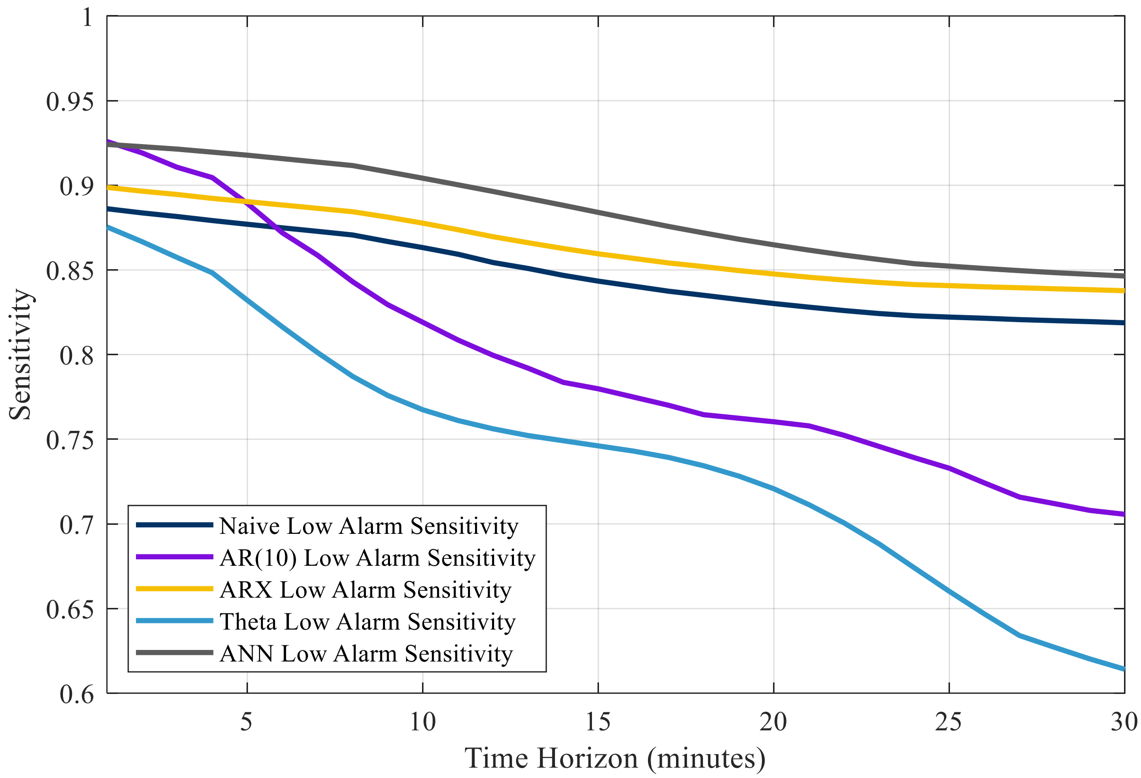

Figure 5.12: Low alarm sensitivity of all H_2O forecasting models over the 30 time horizons

Figure 5.12 shows that not all H_2O forecasting models outperform the naïve model. The models that failed to do this is the AR(10) and Theta, with sensitivity rates below 75% at time horizon 30. The ANN and ARX models outperformed the basis model, but not by an amount that is statistically significant.

5.10 Overall Alarm Forecasting Model Comparison

This chapter presents four methods for alarm forecasting in natural gas production pipelines. The results show all forecasting model performances in terms of RMSE, MAPE, and sensitivity. Some models perform better than others, and other models provide valuable insights in understanding the pipeline system as a whole. To demonstrate the contributions of this research, each model developed to forecast alarms is tested against the current state-of-the-art forecaster used in the pipeline control room at the start of this project.

Chapter 4 introduces the current state-of-the-art forecasting method used in the pipeline control room. A weighted average of each signal recorded at CPF_1- CPF_4 is used to keep the pipeline operators aware of the quality of gas estimated to arrive at the distribution point. This weighted average model (Section 4.3) is replaced by the naïve model (Section 4.3) as the state-of-the-art alarm forecasting model after it was proven the naïve model presented better results when forecasting alarms. The weighted average model struggles to forecasting signal values outside the steady state range, which is understandable as the system is in a steady state more than 95% of the time (Table 5.1). The naïve model is accurate enough to give estimates signals up to 30 minutes in advance (i.e., forecasts at least 50% of alarms correctly at time horizons 30) and forecasts rare signal event with more consistency. The results above show each of our purposed forecasting models outperform the weighted average and naïve model, allowing us to draw final conclusions for the 10th-order autoregressive model (AR(10)), the autoregressive model with exogenous variable (ARX), the Theta method, and the artificial neural network (ANN) models.

The first conclusion is that the artificial neural network performs the best out of all forecasting techniques. The ANN averages an alarm forecasting sensitivity of approximately 94.2% for time horizon 5, 88.6% for time horizon 15, and 82.7% for time horizon 30. This means we are confident the ANN predicts 82.7% of all alarms 30 minutes into the future. 30 minutes’ warning offers pipeline controllers more time to correct system issues that they otherwise would not have of until an alarm is already triggered or the pipeline is shut in.

The autoregressive model with exogenous variables (ARX) performs second best. Although performing poorly when forecasting CO_2 alarms, the ARX model consistently has the second lowest MAPE and RMSE when forecasting all other signals. The ARX averages a sensitivity of 90.3% for time horizon 5, 81.8% for time horizon 15, and 79.1% for time horizon 30. Both the ANN and the ARX use lagged exogenous variables (signals being generated at the upstream central processing facility). We calculate lags in Section 4.4 by looking at the lagged cross-correlation of the distribution points signal and each CPF. The success of these models shows that lagged gas quality signals coming from the upstream CPFs influence the signals being recorded at the distribution point.

The AR(10) model is not as successful as the ARX or ANN, but still provides useful information to understand the system we are forecasting. In cases where the AR(10) outperforms the ARX (BTU forecasting, Figure 5.10), it is reasonable to assume that the ARX's exogenous variables are not significant when forecasting that signal, and could possibly be hurting its performance. The same hypotheses could be formed about the exogenous variables in ANN. The AR(10) averages a sensitivity of 87.5% for time horizon 5, 79.6% for time horizon 15, and 71.2% for time horizon 30. The small amount of information the AR(10) forecasts with can limit its ability to forecast at large time horizons. The lowest performing model, the Theta method, also has higher error rates at larger time horizons.

The Theta method produces an average sensitivity of 88.3% for time horizon 5, 77.0% for time horizon 15, and 66.8% for time horizon 30. The Theta method outperforms the AR(10) in the early time horizons. However, because the Theta method relays on an equal combination of local curvature and long-term trend information to forecast, we see a significant drop in performance in the larger time horizons. Many of the signal's local curvatures at time t are substantially different $t + 30$ minutes into the future. Hence, the Theta methods error rates increase more than usual the farther out it forecasts. We discuss this limitation in the future work section of Chapter 6.

A major error metric used in this work is the sensitivity of a model forecasting a specific alarm threshold. For all forecasting models, we encounter too few triggered alarms to test all four alarm thresholds (high-high, high, low, low-low). Subsequently, we test each model's alarm forecasting accuracy with the alarm threshold that contains the highest occurrence rate. This means if the testing set contains the highest number of low alarms, the low alarm threshold is selected to calculate the sensitivity rates across all the time horizons. The disadvantage of this decision is that we make assumptions on overall model performance based on a sometimes-irrelevant alarm threshold. An example of this is using a low alarm threshold to test the H_2O forecasting models. Less moisture in the pipeline is a good thing; hence, low alarms do not have the same importance as high or high-high alarms. Similar concerns as this is expanded in the research summary in Chapter 6.

CHAPTER 6

Benefits of Alarm Forecasting in Natural Gas Pipelines and Future Work Considerations

6.1 Chapter Objectives

Chapter 6 discusses the contributions of this thesis and suggests opportunities for future work. The contributions of our research include an alarm forecasting framework and four techniques to forecast alarms in natural gas production pipelines. Future work is suggested based on the experience gained from this project and the initial response of the pipeline controllers using the forecasting algorithms. Then, a conclusion is drawn about the work.

6.2 Contributions of Our Work

Natural gas production companies use pipelines to transport natural gas from point the extraction well to distribution point. This work helps a natural gas production company achieve this goal by providing the tools needed to forecast pipeline alarms up to 30 minutes in advance. Pipeline alarms alert control room operators of immediate threats to the system. Delaying action until after an alarm has been triggered is often costly because damage may have already occurred, potentially leading to shutdowns, loss of profit, and dangerous environments. This research shows how production pipelines can avoid unprofitable consequences through the application of our alarm forecasting algorithm. The alarm forecasting algorithms described in this work aid pipeline controllers in achieving flow assurance and allow them to conduct preventative maintenance to decrease operation cost, unsafe working conditions, and damage to the environment.

The top performing alarm forecasting algorithm in this research uses an artificial neural network with exogenous variables to forecasts future pressure, heat content, hydrogen sulfide, carbon dioxide, and moisture content signals. The ANN averages an alarm forecasting sensitivity of 94.2% for time horizon five, 88.6% for time horizon 15, and 82.7% for time horizon 30. This outperforms the current state-of-the-art forecaster (naïve model), which forecasts alarms five minutes into the future with an average accuracy of 89.4%, 15 minutes into the future with an average accuracy of 83.8%, and 30 minutes into the future with an average accuracy of 74.1%. These higher accuracies return a higher number of correctly predicted alarms. More correctly predicted alarms return larger amounts of gas being sold to the distribution vendors, increasing profits, and protecting equipment from long-term damage.

The forecasting framework requires little computational resources to operate, provides real-time anomaly detection, and is developed to work seamlessly with existing control room software. The framework is flexible enough to allow pipeline controllers to change alarm thresholds at their discretion and add additional signals to be forecasted with ease. The framework manages the forecasting algorithms and alerts the pipeline controllers to immediate threats in the system.

A defining quality of this project was the decision to build our alarm forecasting models using regression-based methods. The regression-based approach to predicting alarms is used favoring a classification-based approach because the alarm thresholds can be changed after the algorithm is deployed. In practice, unsafe and alarm-triggering values are avoided by control operators, which makes actual alarm occurrences in reported data scarce. Since alarms are triggered when a threshold is exceeded, a regression model cannot only predict when an alarm will trigger but tell expected values at multiple time horizons to allow operators to perform more appropriate corrective actions.

6.3 Future Work

This work presents several opportunities for future research. Forecasting alarms with machine learning is commonly approached with either classification or regression. Our initial search for alarm forecasting methods included a classification approach [90], but because of the relative infrequency of positive cases of alarms, our classifier did not perform well. The output of a classification-based model is binary: An alarm is either present, or it is not present. Due to this nature of an alarm, this research problem will do well to further explore the classification modeling techniques seen in industry.

For the techniques we did implement in this work, some can be expanded upon. The Theta method is known for being one of the most robust time series forecasting methods available. However, in our work, the Theta forecaster consistently performs worst out of all methods. We believe this to be a penalty of our short-term forecasting, as explained in Section 5.10. However, the optimization of tuning parameters mentioned in [77], [81] as well as multivariate expansions of the Theta method [79] might benefit the alarm forecasting model.

The artificial neural network (ANN) and autoregressive model with exogenous variables (ARX) find success when using the lagged signal values from the central processing facilities upstream. All methods would benefit from further exploration of this idea. In the models discussed in this work, only a one signal type is used as input to forecast the same signal h minutes into the future. That is, the hydrogen sulfide forecasting models only take hydrogen sulfide signals as inputs. We speculate that if we included other signals besides hydrogen sulfide as inputs to the hydrogen sulfide forecaster, the relationship the hydrogen sulfide signal has with other signals might uncover ways increase model performance.

The relationships between hydrogen sulfide and carbon dioxide discussed in Chapter 3 introduces a chance to test this theory. The combination of high hydrogen sulfide and carbon dioxide is a common reason for the pipeline to get shut in. If we were to include both these signals with lagged observations in the training of the ANN and the ARX model, the relationship between the two signals may help when forecasting at time horizons beyond 30 minutes. This is just one example of where we might be able to exploit the relationship of one signal to another to forecast alarms.

Looking to new signals entirely to forecast alarms at further time horizons may prove beneficial. Chapter 2 points out that a pipeline control room is monitoring anywhere from 80 to 100 signals at once. The models in this work use only the pressure, heat content, carbon dioxide, hydrogen sulfide, and moisture content signals. If we can access the other signals the control room is monitoring, a more robust forecaster may be obtainable. For example, the internal pressure of the pipe is directly related to the amount of gas being flown into it. If we acquired the flow-rate time series for each CPF and the distribution point, the flow-rate data may improve our pressure forecasting ability. Other signals, such as machinery revolution-per-minute rates, may also help combat the problem of human intervention in day-to-day operations.

The pipeline control room operators are trained to log unusual occurrences seen in the pipeline. An example of an unusual occurrence would be if one of the central processing facility’s dehydration units fails. If this event occurs, the control room operators log the failure and instruct the pipeline workers to fix the issue. Our models are not aware these events, which hurt our forecasting ability. If we could gain access to the control log that the operators maintain, we may be able to develop an algorithm that detects the failure and adjusts the forecasting models appropriately based on the type of failure.

Natural gas production companies produce massive datasets every day. While the production company sponsoring this work requested this specific alarm forecasting project, there are many opportunities for improvement through all upstream production. From leak detection to flow dynamics, new ideas are being applied to similar datasets like the ones in this work to ensure the safe and reliable transportation of natural gas.

6.4 Conclusion

In conclusion, the alarm forecasting models in this thesis are powerful tools to help natural gas production companies predictively maintain their pipeline. Forecasting alarms help pipeline controllers avoid being shut-in, saving production companies significant resources that otherwise would be spent on reactively maintaining the pipeline. The end user of this work, the pipeline controller, contributed valuable insight to this project and worked closely with us to explain the characteristics of the pipeline signals used in this research. Through our combined efforts, all alarm forecasting models outperform the state-or-the-art forecaster being used by the production company, with the artificial neural network successfully predicting 82.7% of all alarms 30 minutes in advance.

BIBLIOGRAPHY

[1] G. Sullivan, R. Pugh, P. Melendez, and W. Hunt, “Operations & Maintenance Best Practices,” 2010.

[2] J. M. Saleh, *Fluid Flow Handbook*, 1st ed. New York: McGraw - Hill, 2002.

[3] S. M. Folga, “Natural Gas Pipeline Technology Overview,” 2007.

[4] Y. F. Makogon, “Natural Gas hydrates – A Promising Source of Energy,” *J. Nat. Gas Sci. Eng.*, vol. 2, no. 1, pp. 49–59, 2010.

[5] W. Parrish, A. Kidnay, and D. McCartney, *Fundamentals of Natural Gas Processing*, 2nd ed. CRC Press, 2013.

[6] S. Mokhatab, W. Poe, and J. Mak, *Handbook of Natural Gas Transmission and Processing*, no. 2. Gulf Professional Publishing, 2018.

[7] B. Parvez, J. Ali, U. Ahmed, and M. Farhan, “Framework for Implementation of AGA 12 for Secured SCADA Operation in Oil and Gas Industry,” *2015 2nd Int. Conf. Comput. Sustain. Glob. Dev.*, pp. 1281–1284, 2015.

[8] H. D. Beggs, *Gas Production Operations*, 1st ed. Tulsa, Oklahoma: Oil and Gas Consultants International, Inc, 1991.

[9] M. King and T. Bott, *Extraction of Natural Products Using Near-Critical Solvents*, 1st ed. Glasgow: Blackie Academic & Professional, 1993.

[10] J. Golden and H. Wiseman, “The Fracking Revolution: Shale Gas as a Case Study in Innovation Policy,” *Emory Law J.*, vol. 64, no. 4, p. 955, 2015.

[11] A. Shah, G. A. Leeke, S. Rigby, and M. Greaves, “A Review of Novel Techniques for Heavy Oil and Bitumen Extraction and Upgrading,” *Energy Environ. Sci.*, vol. 3, no. June 2010, pp. 700–714, 2010.

[12] U.S. Energy Information Administration, “Total Energy: Monthly Data Collection,

Section 4 Natural Gas," 2018.

[13] D. M. Kargbo, R. G. Wilhelm, and D. J. Campbell, "Natural Gas Plays in the Marcellus Shale: Challenges and Potential Opportunities," *Environ. Sci. Technol.*, vol. 44, no. 15, pp. 5679–5684, 2010.

[14] Natural Gas Supply Association, "Processing Natural Gas," 2013. [Online]. Available: http://naturalgas.org/naturalgas/processing-ng/. [Accessed: 26-Sep-2019].

[15] R. W. Baker and K. Lokhandwala, "Natural Gas Processing with Membranes: An Overview," *Industrial and Engineering Chemistry Research*, vol. 47, no. 7. pp. 2109–2121, 2008.

[16] C. W. Remeljej and A. F. A. Hoadley, "An Exergy Analysis of Small-scale Liquefied Natural Gas (LNG) Liquefaction Processes," *Energy*, vol. 31, no. 2006, 2019.

[17] J. Speight, *Shale Gas Production Processes*, 1st ed. Laramie, Wyoming: Gulf Professional Publishing, 2013.

[18] U.S. Energy Information Administration, "The Geology of Natural Gas Resources," 2011. [Online]. Available: http://www.eia.gov/todayinenergy/detail.php?id=110. [Accessed: 16-Sep-2019].

[19] J. Lee, "Conventional vs Unconventional Natural Gas," *Oil & Gas Blog*, 2015. [Online]. Available: https://www.croftsystems.net/oil-gas-blog/conventional-vs.-unconventional. [Accessed: 17-Mar-2019].

[20] D. Gussow, "Natural Gas Specs Sheet: Florida Power and Light Company," 2004.

[21] L. V. Groeger, "Pipelines Explained: How Safe are America's 2.5 Million Miles of Pipeline?," *ProPublica*, 2015. [Online]. Available: https://www.propublica.org/article/pipelines-explained-how-safe-are-americas-2.5-million-miles-of-pipelines. [Accessed: 02-Jan-2020].

[22] V. Uraikul, C. W. Chan, and P. Tontiwachwuthikul, "Development of an Expert System for Optimizing Natural Gas Pipeline Operations," *Expert Syst. Appl.*, vol. 18, pp. 271–282, 2000.

[23] J. Tobin, P. Shambaugh, and E. Mastrangelo, “Natural Gas Processing: The Crucial Link Between Natural Gas Production and Its Transportation to Market,” 2006.

[24] R. Haws, “Contaminants in Amine Gas Treating,” in *GPA Houston Regional Meeting*, 2001, no. 1, pp. 1–13.

[25] D. Messersmith, D. Brockett, and C. Loveland, “Understanding Natural Gas Compressor Stations,” *Pen State Extension*. The Pennsylvania State University 2015, pp. 1–5, 2011.

[26] M. Naseer and W. Brandstätter, “Hydrate Formation in Natural Gas Pipelines,” *WIT Trans. Eng. Sci.*, vol. 70, pp. 261–269, 2011.

[27] T. Hong, “Energy forecasting: Past, Present and Future,” *Foresight Int. J. Appl. Forecast.*, vol. 32, pp. 43–48, 2013.

[28] D. Price, “Energy and Human Evolution,” *Popul. Environ.*, vol. 16, no. 4, pp. 301–319, 1995.

[29] V. Smil, *Energy and Civilization: A History*. London, England: The MIT Press, 2017.

[30] “History of Natural Gas,” *NaturalGas.org*, 2020. [Online]. Available: http://naturalgas.org/overview/history/. [Accessed: 04-Jan-2020].

[31] U.S. Energy Information Administration, “Natural Gas Explained,” *Natural Gas Pipelines - Independent Statistics and Analysis*. pp. 101–114, 2019.

[32] U.S Energy Information Administration, “Global Natural Gas Consumption Doubled from 1980 to 2010,” 2012.

[33] A. McFarland, “In 2018, the United States Consumed More Energy Than Ever Before,” *U.S. Energy Information Administration - Today in Energy*, 2019. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=39092. [Accessed: 14-Jan-2020].

[34] K. Dyl, “Annual Energy Outlook 2018 with Projections to 2050,” *Annu. Energy*

Outlook, vol. 44, no. 8, pp. 1–64, 2018.

[35] “Environmental Impacts of Natural Gas,” 2014. [Online]. Available: http://www.ucsusa.org/clean_energy/our-energy-choices/coal-and-other-fossil-fuels/environmental-impacts-of-natural-gas.html#.VSsoifnF-So. [Accessed: 09-Dec-2019].

[36] B. Mandler and E. Allison, “Methane Emissions in the Oil and Gas Industry,” *American Geosciences: Petroleum and the Environment*, vol. 19. AAPG Foundation, pp. 1–2, 2018.

[37] M. Kelso, “Pipeline Incidents Continue to Impact Residents,” *FractTracker Alliances*, 2018. [Online]. Available: https://www.fractracker.org/2018/12/pipeline-incidents-impact-residents/.

[38] *Natural Gas Pipeline Safety Act of 1968*. United States of America: Natural Gas Pipeline Safety Act of 1968, 1968, pp. 720–729.

[39] Pipeline and Hazardous Materials Safety Administration, “Field Operations Overview,” 2018. [Online]. Available: https://www.phmsa.dot.gov/regulatory-compliance/pipeline/field-operations-overview. [Accessed: 02-Jan-2020].

[40] V. Chandra, *Fundamentals of Natural Gas: An International Perspective*. Tulsa, Oklahoma: PennWell Corporation, 2006.

[41] J. Chow, R. J. Kopp, and P. R. Portney, “Energy Resources and Global Development,” *Science (80-.).*, vol. 302, no. 5650, pp. 1528–1531, 2003.

[42] R. Hefner, “The Age of Clean Gases,” *American Clean Skies*, pp. 49–53, 2007.

[43] M. A. Adegboye, W. K. Fung, and A. Karnik, “Recent Advances in Pipeline Monitoring and Oil Leakage Detection Technologies: Principles and Approaches,” *Sensors (Switzerland)*, vol. 19, no. 11, 2019.

[44] P. Hopkins, “Pipeline Integrity Reviews,” in *International Pipeline Conference*, 2000, p. 15.

[45] M. R. Akhondi, A. Talevski, S. Carlsen, and S. Petersen, “Applications of

Wireless Sensor Networks in the Oil, Gas and Resources Industries," *Proc. - Int. Conf. Adv. Inf. Netw. Appl. AINA*, pp. 941–948, 2010.

[46] R. G. Alakbarov and M. A. Hashimov, "Application and Security Issues of Internet of Things in Oil-Gas Industry," *Int. J. Educ. Manag. Eng.*, vol. 8, no. 6, pp. 24–36, 2018.

[47] R. Aliguliyev and Y. Imamverdiyev, "Big Data Strategy for the Oil and Gas Industry: General Directions," *Probl. Inf. Technol.*, vol. 08, no. 2, pp. 31–42, 2017.

[48] CFE Media, "What the Oil and Gas Industry Needs from SCADA," *Oil & Gas Engineering*, 2017. [Online]. Available: https://www.oilandgaseng.com/articles/what-the-oil-and-gas-industry-needs-from-scada/. [Accessed: 03-Jan-2020].

[49] Pipeline and Hazardous Materials Safety Administration and Office of Pipeline Safety, "Guidance Manual for Operators of Small Natural Gas Systems," 2017.

[50] D. Havard, *Oil and Gas Production Handbook - An Introduction to Oil and Gas Production, Transport, Refining and Petrochemical Industry*, 3rd ed. ASEA Brown Boveri, 2013.

[51] Pipeline and Hazardous Materials Safety Administration, "Operations & Maintenance Enforcement Guidance," 2017.

[52] J. Deyette, S. Clemmer, R. Cleetus, S. Sattler, A. Bailie, and M. Rising, "The Natural Gas Gamble: A Risky Bet on America's Clean Energy Future." p. 56, 2015.

[53] E. Dendy Sloan and C. A. Koh, *Clathrate Hydrates of Natural Gases*, 3rd ed. Boca Raton, Florida: CRC Press, 2008.

[54] E. Dendy Sloan and B. J. Bloys, *Hydrate Engineering*, 21st ed., vol. 21. Richardson, Tex, USA, 2000.

[55] J. P. Lederhos, J. P. Long, A. Sum, R. L. Christiansen, and E. D. Sloan, "Effective Kinetic Inhibitors for Natural Gas Hydrates," *Chem. Eng. Sci.*, vol. 51, no. 8, pp. 1221–1229, 1996.

[56] C. A. Koh, "Towards a Fundamental Understanding of Natural Gas Hydrates," *Chem. Soc. Rev.*, vol. 31, no. 3, pp. 157–167, 2002.

[57] J. Wan, Y. Yu, Y. Wu, R. Feng, and N. Yu, "Hierarchical Leak Detection and Localization Method in Natural Gas Pipeline Monitoring Sensor Networks," *Sensors*, vol. 12, no. 1, pp. 189–214, 2012.

[58] W. Zhao, T. Zhang, Y. Wang, J. Qiao, and Z. Wang, "Corrosion Failure Mechanism of Associated Gas Transmission Pipeline," *Materials (Basel).*, vol. 11, no. 10, 2018.

[59] H. Pouraria, J. K. Seo, and J. K. Paik, "A Numerical Study on Water Wetting Associated with the Internal Corrosion of Oil Pipelines," *Ocean Eng.*, vol. 122, pp. 105–117, 2016.

[60] A. Elhoud, F. Jewilli, K. Abouswa, and O. Rageai, "Failure Analysis: Internal Corrosion Rupture of a 6-in Gas Line," *Mater. Perform.*, vol. 51, pp. 70–73, May 2012.

[61] M. Javidi and S. Bekhrad, "Failure Analysis of a Wet Gas Pipeline Due to Localized CO2 Corrosion," *Eng. Fail. Anal.*, vol. 89, no. December 2017, pp. 46–56, 2018.

[62] E. O. ObaniJesu, "Modeling the H2S Contribution to Internal Corrosion Rate of Natural Gas Pipeline," *Energy Sources, Part A Recover. Util. Environ. Eff.*, vol. 31, no. 4, pp. 348–363, Jan. 2009.

[63] E. O. Obanijesu, V. Pareek, and M. O. Tade, "Modeling the Contribution of Gas Hydrate to Corrosion Rate Along the Subsea Pipelines," *Pet. Sci. Technol.*, vol. 32, no. 21, pp. 2538–2548, 2014.

[64] B. Bedairi, D. Cronin, A. Hosseini, and A. Plumtree, "Failure Prediction for Crack-in-Corrosion Defects in Natural Gas Transmission Pipelines," *Int. J. Press. Vessel. Pip.*, vol. 96–97, pp. 90–99, 2012.

[65] A. Dotis-Georgiou, "Autocorrelation in Time Series Data," *DZone: A Devada Media Property*, 2019. [Online]. Available: https://dzone.com/articles/autocorrelation-in-time-series-data. [Accessed: 02-Jan-2020].

[66] R. Shumway and D. Stoffer, *Time Series Analysis and Its Applications*, Fourth. Springer, 2017.

[67] H. Akaike, "Fitting Autoregressive Models for Prediction," *Ann. Inst. Stat. Math.*, vol. 21, no. 1, pp. 243–247, 1969.

[68] J. Durbin, "The Fitting of Time-Series Models," *Int. Stat. Inst.*, vol. 28, no. 3, pp. 233–244, 1960.

[69] V. Flovik, "How (not) to use Machine Learning for time series forecasting: Avoiding the pitfalls," *Towards Data Science*, 2018. [Online]. Available: https://towardsdatascience.com/how-not-to-use-machine-learning-for-time-series-forecasting-avoiding-the-pitfalls-19f9d7adf424. [Accessed: 07-Jan-2020].

[70] S. R. Vitullo, R. H. Brown, G. F. Corliss, and B. M. Marx, "Mathematical Models for Natural Gas Forecasting," *Can. Appl. Math Q.*, vol. 17, no. 4, pp. 1–13, 2010.

[71] V. Majazi Dalfard, M. Nazari Asli, S. M. Asadzadeh, S. M. Sajjadi, and A. Nazari-Shirkouhi, "A Mathematical Modeling for Incorporating Energy Price Hikes into Total Natural Gas Consumption Forecasting," *Appl. Math. Model.*, vol. 37, no. 8, pp. 5664–5679, 2013.

[72] R. Hribar, P. Potočnik, J. Šilc, and G. Papa, "A Comparison of Models for Forecasting the Residential Natural Gas Demand of an Urban Area," *Energy*, vol. 167, pp. 511–522, 2018.

[73] H. Spliid, "A Fast Estimation Method for the Vector Autoregressive Moving Average Model with Exogenous Variables," *J. Am. Stat. Assoc.*, vol. 78, no. 384, pp. 843–849, Dec. 1983.

[74] H. N. Akouemo and R. J. Povinelli, "Data Improving in Time Series Using ARX and ANN Models," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3352–3359, 2017.

[75] "Makridakis Open Forecasting Center," 2020. [Online]. Available: https://mofc.unic.ac.cy/. [Accessed: 17-Jan-2020].

[76] S. Makridakis and M. Hibon, "The M3-Competition: Results, Conclusions and Implications," *Int. J. Forecast.*, vol. 16, pp. 451–476, 2000.

[77] F. Petropoulos and K. Nikolopoulos, "The Theta Method," *Foresight Int. J. Appl. Forecast.*, no. 46, pp. 11–17, 2017.

[78] R. J. Hyndman and B. Billah, "Unmasking the Theta method," *Int. J. Forecast.*, vol. 19, no. 2, pp. 287–290, 2003.

[79] P. Deng, Z. Han, and R. D. Reitz, "Forecasting Multivariate Time Series with the Theta Method," *Proc. Inst. Mech. Eng. D J. Automot. Eng.*, vol. 230, pp. 885–898, 2016.

[80] J. A. Fioruci, T. R. Pellegrini, F. Louzada, and F. Petropoulos, "The Optimized Theta Method," pp. 1–18, 2015.

[81] M. Khashei and M. Bijari, "An Artificial Neural Network (p, d, q) Model for Timeseries Forecasting," *Expert Syst. Appl.*, vol. 37, no. 1, pp. 479–489, 2010.

[82] S. M. Al-Fattah and R. A. Startzman, "Predicting Natural Gas Production Using Artificial Neural Network," *Res. Gate*, no. April, 2001.

[83] A. Miguel, "A Review of Dimension Reduction Techniques," *Dep. Comput. Sci. Univ. Sheffield. Tech. Rep. CS-96-09*, pp. 1–69, 1997.

[84] "How Does the Natural Gas Delivery System Work?," *American Gas Association*, 2019. [Online]. Available: https://www.aga.org/how-does-natural-gas-delivery-system-work. [Accessed: 16-Dec-2019].

[85] D. Haeseldonckx and W. D'Haeseleer, "The Use of the Natural-gas Pipeline Infrastructure for Hydrogen Transport in a Changing Market Structure," *Int. J. Hydrogen Energy*, vol. 32, no. 10–11, pp. 1381–1386, 2007.

[86] F. Geuzeborek and T. van der Lei, "Sour Gas," *Sci. Direct - 23rd Eur. Symp. Comput. Aided Process Eng.*, 2013.

[87] A. E. Sjogren and C. J. Myers, "Interfacing Synchronous and Asynchronous Modules within a High-Speed Pipeline," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 8, no. 5, pp. 573–583, 2000.

[88] "MATLAB resample Function: Time Series," *MathWorks*, 2019. [Online].

Available:
https://nl.mathworks.com/help/matlab/ref/timeseries.resample.html?searchHighlight=resample&s_tid=doc_srchtitle. [Accessed: 27-Dec-2019].

[89] H. Yu and B. M. Wilamowski, *The Industrial Electronics Handbook - Intelligent Systems*, 2nd ed. Boca Raton, Florida: CRC Press, 2011.

[90] X. Huang, A. Maier, J. Hornegger, and J. A. K. Suykens, "Indefinite Kernels in Least Squares Support Vector Machines and Principal Component Analysis," *Appl. Comput. Harmon. Anal.*, vol. 43, no. 1, pp. 162–172, 2017.