Predicting Natural Gas Pipeline Alarms

Colin O. Quinn[1], Richard J. Povinelli[1], Jaired R. Collins[2], and Jake Norment

[1]Marquette University, Milwaukee, Wisconsin
[2]Missouri Southern State University, Joplin, Missouri

***Abstract*— Natural gas production companies use pipelines to transport natural gas from the extraction well to a distribution point. The internal pressure of these pipelines is closely regulated to maintain a steady-state system. Sensors are used to collect real-time pressure 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 pressure within an acceptable range, the company risks being shut-in (unable to distribute gas) or rupturing the pipeline. Predicting pressure alarms enables operators to take appropriate action earlier to avoid being shut-in and is a form of predictive maintenance. We forecast alarms by using an autoregressive model (AR) in conjunction with alarm thresholds. The alarm thresholds are defined by the production company and are occasionally adjusted to meet current environment conditions. The forecasting results show that we can accurately predict pressure alarms up to a 30-minute time horizon.**

***Index Terms*— Steady-state System, Predictive Maintenance, Autoregressive Model, Natural Gas Pipelines Alarms, Alarm Thresholds.**

I. Introduction to Natural Gas Production, Pipeline Challenges, and Alarm Forecasting

The transportation of natural gas introduces a variety of challenges as natural gas production companies try to maintain safe, economical, and efficient operations. By the time the 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 [1]. If the gas meets these standards, 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.

Being shut-in is costly and time consuming for the production company. For the pipeline to become functional again, the unexpected gas must either be diffused with gas further down the line or flared from the system entirely. To avoid this, a pipeline control room monitors the condition of the natural gas within the pipeline to coordinate its processing before it reaches the distribution point [2]. With the help of forecasted alarms, it is possible to alert the controllers in advance if the natural gas further down the pipeline moving towards the distribution point will be considered unacceptable by the time it arrives [3]. In our paper, the pressure measured at the distribution point is the variable forecasted to help controllers deliver acceptable gas in a safe and reliable way.

Forecasting alarms with machine learning is commonly approached with either classification or regression [4, 5]. The output of a classification-based model is binary: An alarm is either present or not present. A regression-based approach predicts future values, and these future values are compared against rules that define an alarm and thus are used to forecast alarms. 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.

Our paper is organized as follows: Section 2 presents background of the natural gas industry and description of our autoregressive model. Section 3 describes our methods in predicting alarms. Section 4 shows the results of our regression model, and the ability to forecast alarms. Finally, Section 5 concludes our current work.

We are reporting work sponsored by a natural gas pipeline company in the U.S. The data has been scaled to preserve confidentiality.

II. Project Background and Incentive

The goal of this paper is to develop a method to forecast natural gas pipeline pressure alarms to help control room operators maintain a functioning pipeline.

A. Natural Gas Production Process

Natural gas has become a leading energy source in the United States [6]. Its abundance and energy potential makes it a cost-effective energy source for heating, cooking, and electric power generation [7].

Natural gas production companies extract raw natural gas from the ground, which consists of a variety of combustible hydrocarbons, gases, water, and oil. Processing raw natural gas strips away most components until only methane and ethane remain [8]. Once the correct make-up of the natural gas has been achieved, the gas is considered to be pipeline quality natural gas. The gas is then compressed and transported through a pipeline to its distribution point. Pipelines typically have several sensors to measure pressure, flow, and temperature. The forecast values can be used to predict alarms that are triggered if certain thresholds are breached.

B. Linear Autoregressive Time Series Model

An autoregressive model is used to forecast future pressure values using past observances. An AR model takes the form

$$\hat{y_t}(\vec{y}) = \mathrm{F}(y_{t-1}, y_{t-2} \cdots y_{t-\mathrm{n}}),$$

where $y_{t-\mathrm{n}}$ represents a lagged value from the original series and $\hat{y_t}$ is the forecasted value at time horizon t. Linear regression is often used in forecasting energy demand and is used as a viable technique in [7, 9, 10].

III. Methods

This section presents the data, defines a pipeline alarm, and shows how the alarm forecasts are made.

A. Natural Gas Pipeline Data

The data from a production company was cleaned and resampled. Cleaning the data consisted of imputing data rows where the time series was corrupted. Sampling rates were inconsistent, so the time series was resampled by using zero-order hold every 1 minute. The cleaned time series spans from 1 January 2018 to 22 May 2018, in 204,476 steps. Figure 1 shows the alarm thresholds with the corresponding pressure time series. The first half of the data is used for training, the second half is used for testing.

B. Alarm Definition and Prediction

A regression-based approach to predicting alarms is used in favor of 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 can not only predict when an alarm will trigger but tell expected values at multiple time horizons to allow operators to perform more appropriate corrective action.

The forecast values calculated from our models were compared against four thresholds – high-high (HH), high (H), low (L), and low-low (LL). The exact thresholds and their occurrences within our data is summarized in Table 1.

TABLE I: ALARM THRESHOLDS AND THEIR OBSERVED OCCURANCES AND PERCENTAGE

	Threshold (psi)	Occurrences	Frequency (%)
HH	9.01	1567	0.77
H	8.96	4997	2.44
L	8.54	4016	1.96
LL	8.49	3010	1.47

C. Linear Autoregressive Model Implementation

An autoregressive predictive model was used to forecast pressure values 1 through 30 minutes into the future. The model takes the form

$$\hat{y}(\vec{y};\vec{\beta}) = \beta_0 + \sum_{i=1}^{10} \beta_\mathrm{i} y_{t-i}$$

The predicted future pressure value $\hat{y}$ is calculated using the last ten minutes of pressure data, $\vec{y}$, and weights $\vec{\beta}$. The model is autoregressive. The forecast $\hat{y}$ is obtained from using previous time steps as input in the regression equation. The variable y_t represents the pressure at the current time. Subsequently, y_{t-1} is the pressure recorded one minute in the past, y_{t-2} is the pressure recorded two minutes in the past, etc. The pressure values are extracted and placed in a design matrix A. Because the model uses the last ten minutes of pressure data, A is filled with lagged pressure values spanning from the first timestamp

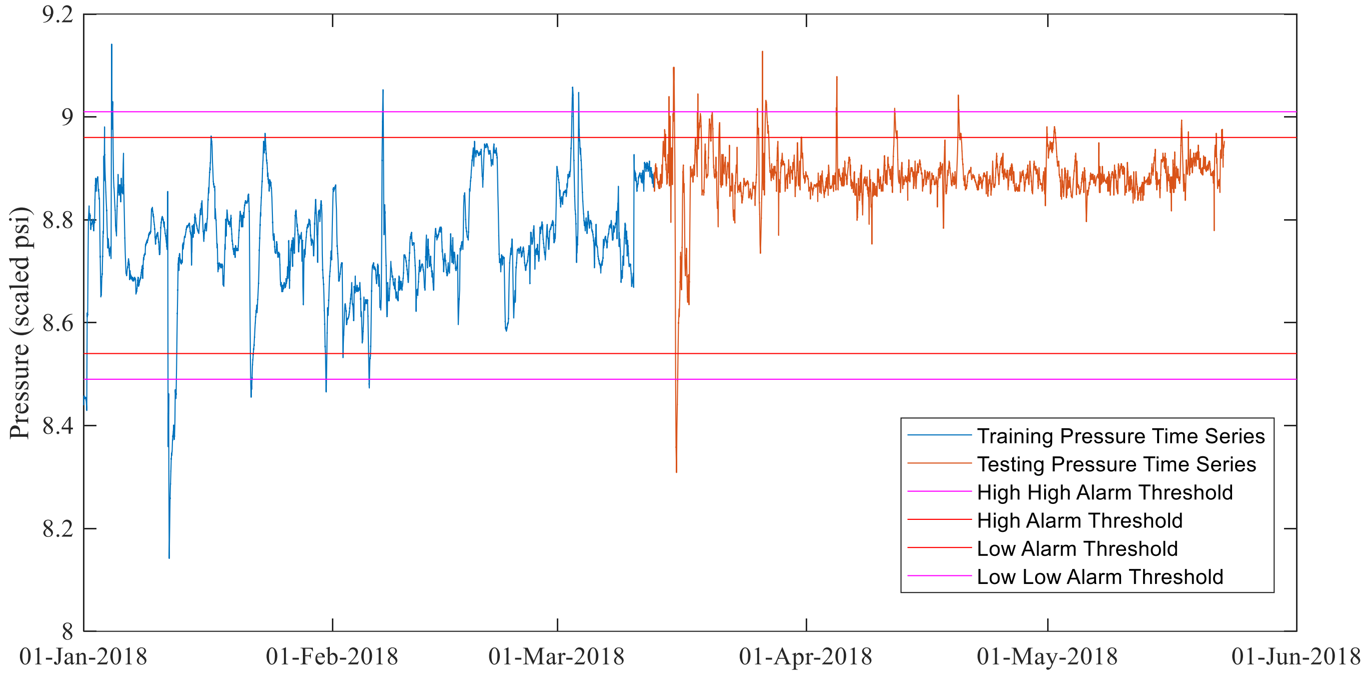

FIGURE 1: SCALED TIME SERIES PRESSURE DATA AND ALARM THRESHOLDS WITH TRAINING SET IN BLUE AND TESTING SET IN ORANGE

pressure value back to ten minutes in the past with a bias coefficient in the first column.

$$A = \begin{matrix} 1 & y_0 & y_1 & \cdots & y_{10} \\ 1 & y_1 & y_2 & \cdots & y_{11} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & y_{n-10} & y_{n-9} & \cdots & y_n \end{matrix}$$

Let $\vec{b}$ be a column vector of the pressure values with the incorporated ten-minute lag.

$$\vec{b} = \begin{matrix} y_{10+TH} \\ y_{11+TH} \\ y_{12+TH} \\ \vdots \\ y_{n+TH}. \end{matrix}$$

TH represents the forecast time horizon. Least squares regression solves the system of equations.

IV. RESULTS OF AR MODEL

Our focus is to predict alarms. However, since we approach this task with regression, we offer metrics to show that our method predicts well. Mean absolute error, mean absolute percent error, and root mean square error are used as regression metrics. Our alarm classification metric is sensitivity, defined as the true positive rate, or in our application, the fraction of gas pipeline alarms that are correctly predicted.

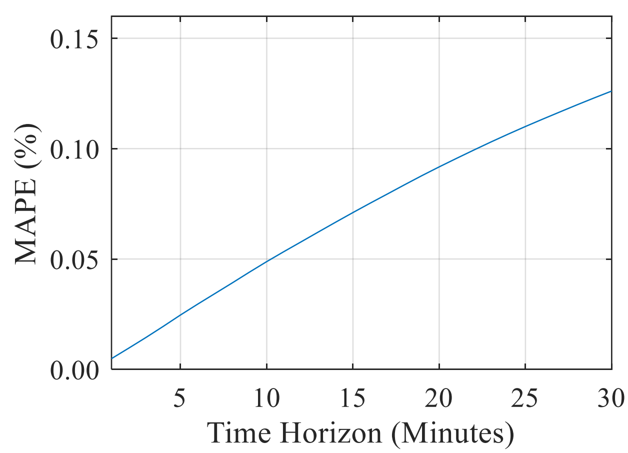

FIGURE 2: MAPE MEASURED ACROSS TIME HORIZONS 1 – 30

A. *Regression Error Metrics:*

The AR model begins forecasting at step 102,238, which is a span of about 71 days. The reason for this is because time steps 1 – 102,238 are used to train the AR model (training dataset). All steps beyond are used as the testing dataset. Time horizons 1-30 minutes are forecasted and considered in these results. Figure 2 shows that the pressure RMSE is less than 3.5 x 10^{-4} scaled psi across all time horizons.Figure 3 shows that the MAPE is less than 0.13% across all time horizons.

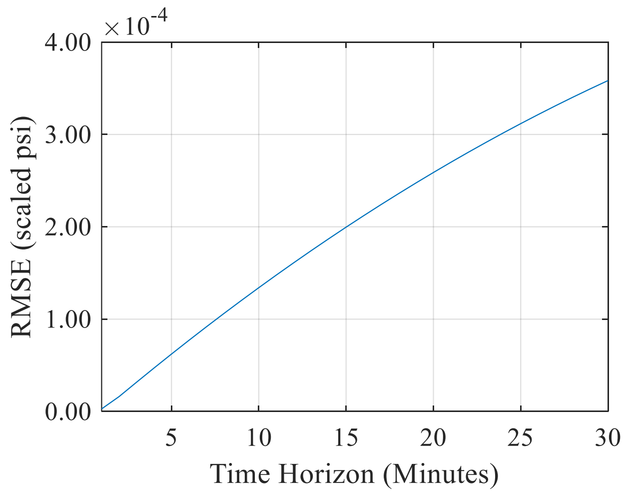

FIGURE 3: RMSE ACROSS TIME HORIZONS 1 – 30

Figure 4 shows the sensitivity (how many times the algorithm correctly predicted at each time horizon). One indicates 100% accuracy. Here we can see that the alarm prediction accuracy falls approximately linearly from near 100% accuracy at 1 minute to 60% accuracy at 30 minutes.

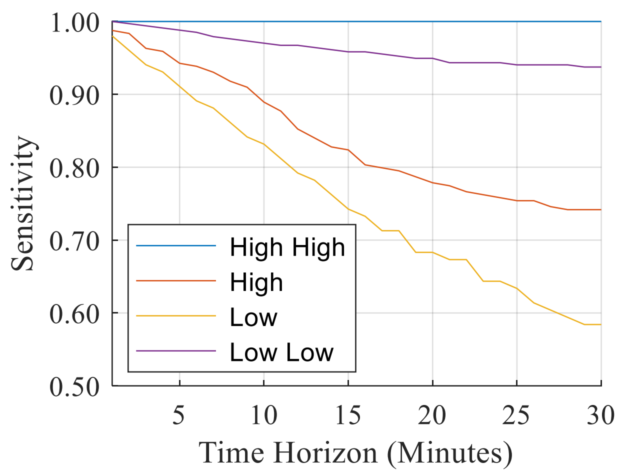

FIGURE 4: SENSITIVITY ACROSS TIME HORIZONS 1 – 30

V. CONCLUSION

This paper presents a method for warning gas production control operators of unacceptable gas with a pipeline. The results show that we can accurately forecast the pressure time series up to a 30-minute time horizon. This translates into true positive rates that drop of linearly from around 100% at one minute to approximately 65% at a 30-minute forecast horizon. This means that at 30 minutes, we correctly forecast 65% of the alarms. We speculate that the sensitivity drops as quickly as it does due to our forecasts fluctuating back and forth over certain alarm thresholds.

REFERENCES

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

[2] S. M. Folga, "Natural Gas Pipeline Technology Overview," 2007.

[3] T. Skripcak and P. Tanuska, "Utilisation of on-line machine learning for SCADA system alarms forecasting," *Sci. Inf. Conf. (…*, pp. 477–484, 2013.

[4] W. Lian, L. Carin, R. Calderbank, J. Reiter, G. Reeves, and K. Heller, "Modeling Time Series and Sequences: Learning Representations and Making Predictions," *Duke Univ. ProQuest Diss. Publ.*, 2015.

[5] D. Jiang, J. Gong, and A. Garg, "Design of early warning model based on time series data for production safety," *Meas. J. Int. Meas. Confed.*, vol. 101, pp. 62–71, 2017.

[6] U.S. EIA, "Annual Energy Outlook 2018 with projections to 2050," *Annu. Energy Outlook 2018 with Proj. to 2050*, vol. 44, no. 8, pp. 1–64, 2018.

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

[8] W. Parrish and A. Kidnay, *Fundamentals of Natural Gas Processing*. 2006.

[9] 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.

[10] 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.