



GLOBAL JOURNAL OF RESEARCHES IN ENGINEERING: J  
GENERAL ENGINEERING  
Volume 24 Issue 2 Version 1.0 Year 2024  
Type: Double Blind Peer Reviewed International Research Journal  
Publisher: Global Journals  
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

# Real-Time Gas Flow Leakage Detection: A Machine Learning Approach to Sensitivity and Uncertainty Analysis

By Ifeanyi Eddy Okoh, Godsdai Idanegbe Usiabulu & Ndidi Lucia Okoh

*University of Port Harcourt*

**Abstract-** Leakage monitoring in flow lines and pipelines is highly important in gas plants due to the relevance of such a system to safety and efficiency. This work will, therefore, attempt to resolve the uncertainties in flow monitoring by integrating machine learning techniques in conducting sensitivity tests on real-time detection mechanisms. In this paper, the effectiveness of pressure-based indicators compared with volume changes has been considered with variations in flow rate and lifting processes. The findings obtained showed that the conventional assumption of the leakage being represented by the difference between initial and final gas volumes is unsatisfactory, especially during the initial pumping phase where inflow rates may appear to be less than outflow rates because of the purging of residual gases. In addition, the ramp-up and plateau stages exhibited a fair amount of variation in inflow and outflow pressure readings, further adding to the leak detection uncertainties. It has, therefore, been deduced that a variable tolerance window will be effective for leak detection based on the differential pressure data analysis between the inlet and outlet gauges.

**Keywords:** *tolerance, lag time, leak detection, sensors, flow analysis.*

**GJRE-J Classification:** *LCC Code: TJ930*



*Strictly as per the compliance and regulations of:*



© 2024. Ifeanyi Eddy Okoh, Godsdai Idanegbe Usiabulu & Ndidi Lucia Okoh. This research/review article is distributed under the terms of the Attribution-NonCommercial-NoDerivatives 4.0 International (CC BYNCND 4.0). You must give appropriate credit to authors and reference this article if parts of the article are reproduced in any manner. Applicable licensing terms are at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

# Real-Time Gas Flow Leakage Detection: A Machine Learning Approach to Sensitivity and Uncertainty Analysis

Ifeanyi Eddy Okoh <sup>α</sup>, Godsday Idanegbe Usiabulu <sup>σ</sup> & Ndid Lucia Okoh <sup>ρ</sup>

**Abstract-** Leakage monitoring in flow lines and pipelines is highly important in gas plants due to the relevance of such a system to safety and efficiency. This work will, therefore, attempt to resolve the uncertainties in flow monitoring by integrating machine learning techniques in conducting sensitivity tests on real-time detection mechanisms. In this paper, the effectiveness of pressure-based indicators compared with volume changes has been considered with variations in flow rate and lifting processes. The findings obtained showed that the conventional assumption of the leakage being represented by the difference between initial and final gas volumes is unsatisfactory, especially during the initial pumping phase where inflow rates may appear to be less than outflow rates because of the purging of residual gases. In addition, the ramp-up and plateau stages exhibited a fair amount of variation in inflow and outflow pressure readings, further adding to the leak detection uncertainties. It has, therefore, been deduced that a variable tolerance window will be effective for leak detection based on the differential pressure data analysis between the inlet and outlet gauges. According to the result of the data analysis, the population variance is 5.38; the sample variance varies across different stages of operation, while the maximum tolerance and pressure are 0.166 and 7.9 bars, respectively. The work automates leak detection and simulates the range of variations, showing the potentiality of AI-ML modeling in enhancing real-world applications. In this work, we are pointing out how machine learning integration may enable a completely new way to define the variable tolerance windows that dramatically improve conventional leak detection.

**Keywords:** tolerance, lag time, leak detection, sensors, flow analysis.

## 1. INTRODUCTION

Gas flow leakage monitoring is critical for gas system safety and efficiency. By continuously observing the process for leaks, possible hazards can be quickly identified and corrected to minimize accident risks and environmental damage in the case of leakage occurrence. References include Freeze and Cherry (1979), Arnaldos et al. (1998), Appah et al. (2021), Bariha et al. (2016), and Gibson et al.

*Author α:* First Hydrocarbon Nigeria Limited FHN 26 Block W Shell Estate Edjeba, Delta State.

*Author σ:* World Bank African Center of Excellence, Center for Oilfield Chemicals and Research, University of Port Harcourt, Port Harcourt, Nigeria. e-mail: godsdayusiabulu@gmail.com

*Author ρ:* Department of Environmental Pollution and Control, Federal University of Petroleum Resources Effurun, Delta State.

(2006). Advanced sensors and monitoring equipment measure gas flow while searching for deviations that may indicate leakage.

Sensitive analytical technique that delivers the minute change in flow rates forms an integral part of any successful gas leakage monitoring. This many times involves uncertainty and sensitivity analysis that shows how factors like pressure, temperature, and flow dynamics will actually affect leak detection accuracy. It is by these analytic methods that optimization of gas flow monitoring systems is done to reduce false alarms and maximize the detection of true leaks.

Moreover, simulation and animation play a vital role in enhancing gas flow leakage monitoring (Hirsch and Agassi, 2010). By simulating various scenarios and visually representing the behavior of gas flow under different conditions, operators gain valuable insights into potential leak locations and the monitoring system's response. This visual representation aids in training personnel, refining monitoring algorithms, and enhancing overall system performance, ultimately contributing to improved safety and reliability in gas flow management.

Despite the established importance of monitoring gas leakage, huge gaps exist within the current research in respect to integrating advanced simulation techniques and their practical implications related to real-time monitoring systems. This work, therefore, will focus on addressing such shortcomings by accounting for the uncertainties involved in the leakage monitoring process of gas flow and sensitivity analysis of factors affecting the accuracy of detection.

*The aim was to:*

1. Investigate the uncertainties associated with gas flow leakage monitoring methods.
2. Analyze the sensitivity of different monitoring techniques in detecting gas leaks.
3. Develop simulations of gas flow leakage scenarios for testing and analysis.
4. Create animations of test results to visualize and understand the behaviour of gas leaks.
5. Assess the effectiveness and reliability of different monitoring approaches through comprehensive analysis and evaluation.

6. To identify potential improvements or optimizations for existing gas flow leakage monitoring systems.
7. To provide recommendations for enhancing the accuracy and efficiency of gas flow leakage monitoring processes based on the research findings.

In-depth analysis of the factors contributing to uncertainty and sensitivity can help in the identification of potential areas for improvement (Vilchez, 1998; Todd et al., 2004; Usiabulu et al., 2022). This research endeavoured to advance the knowledge and methodologies associated with gas flow leakage monitoring.

a) *Data and Method*

The study is from a simulation of gas flow measurement of pressure and time, in a case JK-52 gas flow station. The gas flow leakage monitoring involves process used in this work several key steps.

- Pressure and Time data was recorded in a case gas flow station
- The gas flow monitoring system was calibrated and validated using standard gauges.
- A comprehensive simulation model was developed to represent different scenarios of gas flow and potential leakages.
- This model takes into account various parameters such as pressure differentials, flow velocities, and potential sources of leaks.
- Detection was based of pressure drops more than the uncertainty of pressure differentials

The use of pressure-time based data was intended to mitigate some of the limitations associated with manual gas leak or volume-based detection (Bear et al, 1972; Powers et al., 2000; Sandsten et al., 2000; 2004; Stothard et al., 2004; Svanberg, 2002), including:

1. Gas is not visible and leakage cannot be seen by physical observation
2. Gas may have a turbulent flow and may not obey flow principles (such as Darcy Law)
3. Gas expansion results in inconsistent volume estimation during flow
4. Gas may be dry or wet and has different densities/ primary and secondary gases have different degrees of wettability
5. Gas (volume) is highly impacted by temperature and pressure.

These limitations of volume-based gas leak detection are therefore mitigated by the pressure-based gas leak detection model used in the current work (Lohberger et al., 2004, Montiel et al., 1998, Nosike, 2009; 2020). For the pressure-based detection method, the estimation of gas volume loss was achieved by normalising the pressure reading in the gauge at the regulator. The change in pressure in the gauge` was

calibrated against the change in weight or change in volume of already quantified gas in the system.

b) *Data Collection and Preprocessing*

The data collection process involved systematically recording pressure and time data, along with other relevant events during the gas flow. This data was gathered in various formats, including numerical, categorical, and textual, to ensure comprehensive representation of the flow dynamics. Specialized tools, such as digital pressure gauges and automated logging systems, were employed to capture real-time data from well gauges and store it in a centralized computer system (Wojciech and Janusz, 2012; Chaki et al., 2018; Zukang et al., 2021). For instance, high-precision pressure sensors were utilized to measure fluctuations, while data logging software facilitated the collection and organization of the recorded information.

Once data was collected, the next step involved preprocessing, or data wrangling, to enhance data quality for analysis. This stage included several critical tasks:

1. *Data Cleaning:* Inconsistencies, errors, and missing values were identified and addressed. Missing values were handled using imputation methods, such as mean substitution for numerical data or the mode for categorical data, ensuring the dataset remained robust.
2. *Data Transformation:* Raw data was transformed into a suitable format for analysis. This included normalization techniques, such as min-max scaling, to bring numerical values within a consistent range, as well as encoding categorical variables to facilitate their inclusion in analytical models.
3. *Feature Selection and Extraction:* To reduce dimensionality, feature selection methods were implemented to focus on the most relevant attributes. Techniques such as Recursive Feature Elimination (RFE) were applied to enhance model performance by retaining only those features that significantly contributed to the analysis (Chuka, 2016; Lammel et al., 2021).

To ensure the accuracy and completeness of the collected data prior to preprocessing, quality assurance measures were implemented. These included cross-referencing data logs with sensor readings to verify consistency and conducting preliminary analyses to identify outliers or anomalies.

By employing these preprocessing techniques, the data was refined to align with the research objectives, particularly in preparing it for simulation and sensitivity analysis. This systematic approach not only ensured data quality but also enhanced the reliability of subsequent analyses.

c) *Exploratory Data Analysis*

The data used for this study included an ascii file, extracted as iESogV1.csv for the purpose of this study. An initial automated AI/ML process was used to explore the data before the detailed analysis shown in the subsequent sections. The key observations from the exploratory data analysis (EDA) and the range of the dataset are as below:

- o The dataset contains 8 columns and 1012 rows.
- o The columns represent different measurements of Time (s), Pr\_final, Pr\_initial, Tolerance, Min, Max, and Diff\_Pres (bar), along with event markers in the Events column.
- o The Pr\_final, Pr\_initial, and Tolerance columns have some missing values (~11 missing entries each).
- o The Events column has a substantial number of missing values, with only 4 non-null entries.
- o Pr\_final and Pr\_initial have a wide range (minimum around 1.5 and maximum reaching up to 38.5).

- o The Tolerance values range from 0 to 1.5.
- o Diff\_Pres (bar) has a mean of around 3.53 but ranges from 0 to 9.
- o The Min and Max values are constant (0.8 and 1.2, respectively), which implies they can only be used as cut-offs for other values during further analysis.
- o There is a strong positive correlation between Pr\_final and Pr\_initial (0.98), as well as between Time (s) and both pressure values (above 0.89).
- o Tolerance is moderately correlated with these pressure values as well.
- o The Diff\_Pres (bar) shows a moderate correlation with both Pr\_initial (0.57) and Pr\_final (0.46).
- o Histograms of the numerical data show a skewed distribution for some variables like Pr\_final and Pr\_initial, with many observations concentrated in the lower ranges.

The observations and data details are shown in Figure 1.

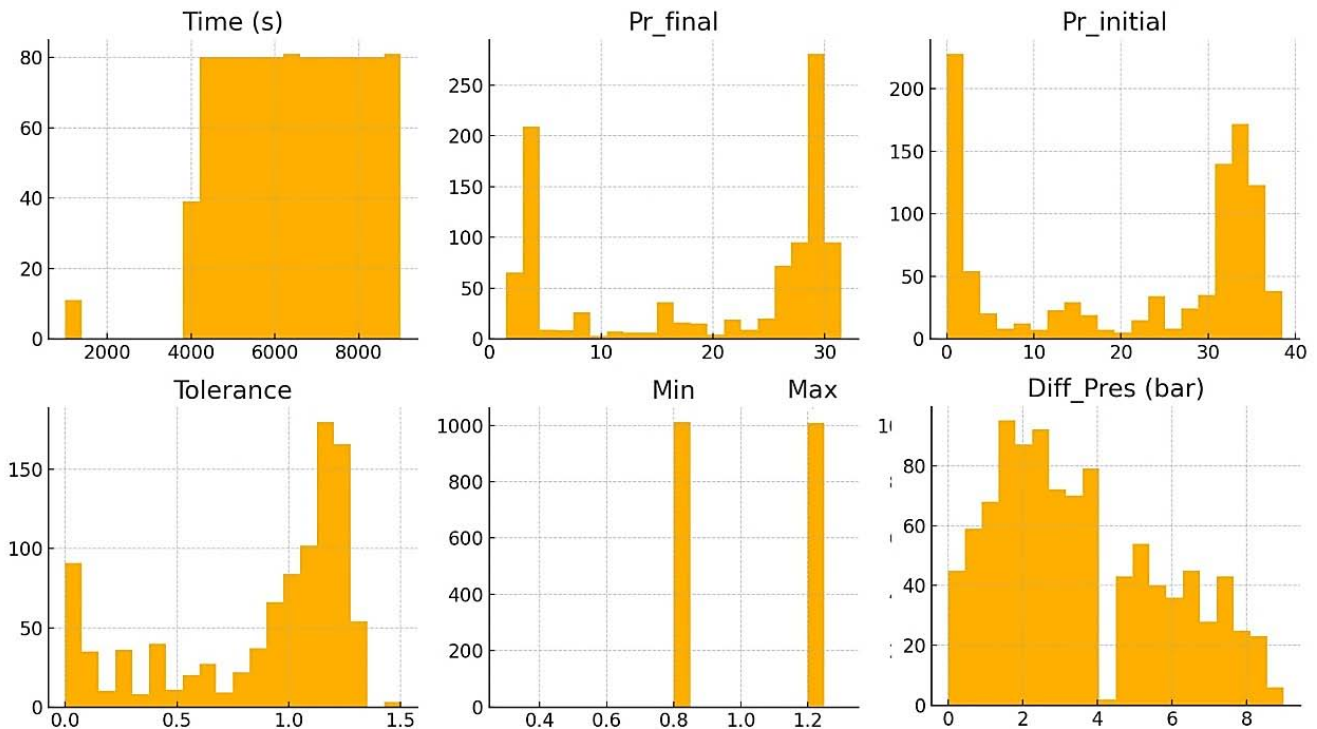


Figure 1: Visualisation of data variation and distribution, showing variation

The data was further analyzed by calculating the variance and standard deviation for each column, and this was achieved using Python Programming and the code is as below:



```
# Calculate the variance and standard deviation for each numerical column
variance = df.var()
std_dev = df.std()
# Create a DataFrame to display the results
stats_df = pd.DataFrame({
    'Variance': variance,
    'Standard Deviation': std_dev
})
import ace_tools as tools; tools.display_dataframe_to_user(name="Variance and Standard Deviation of Columns", dataframe=stats_df)
```

The uncertainty analysis was conducted by assessing the impact of variations in input and output parameters on the monitoring system's performance. Sensitivity analysis was then carried out to identify which input parameters have the most significant impact creating uncertainty on leakage detection. This helps in understanding the critical factors affecting the reliability and accuracy of the monitoring system, including such factors as lag time and initial purge in the gas system.

The test results, both simulated and actual, were animated to visualize the behavior of the gas flow monitoring system under different conditions. This allowed for a better understanding of how the system responded to variations inflow and outflow gas rates and potential leakages.

## II. JK-52 GAS FLOW LEAKAGE MONITORING

This process involved using advanced sensors and monitoring equipment to measure the gas flow and identify deviations that may indicate a leak, in the JK-52 Gas Plant. The pressure values in both the inlet and outlet gauges were recorded. By simulating different

scenarios and visually representing the behaviour of gas flow under these conditions, valuable insights were gained into the potential leak locations and the monitoring system's response. This visual representation helped in training personnel, refining monitoring algorithms, and improving overall system performance, contributing to enhanced safety and reliability in the gas flow management.

### a) Gas Flow in Plant and Data Acquisition

In the gas plant of study, the effluent (a mix up of water, oil and gas) is pumped into the gas plant from nearby oil well. Crude stored in a Floating Production Storage and Offloading Offshore may also tapped from a Tanker offloading/lifting buoy and transported to the Gas Plant. There is also provision for piped crude from multiple well clusters in the field to ensure constant source of hydrocarbon. The crude passes through a water-oil-and-gas separator, a purifier or a compressor as part of the refining or treatment process before delivering the final gas product (Figure 2).

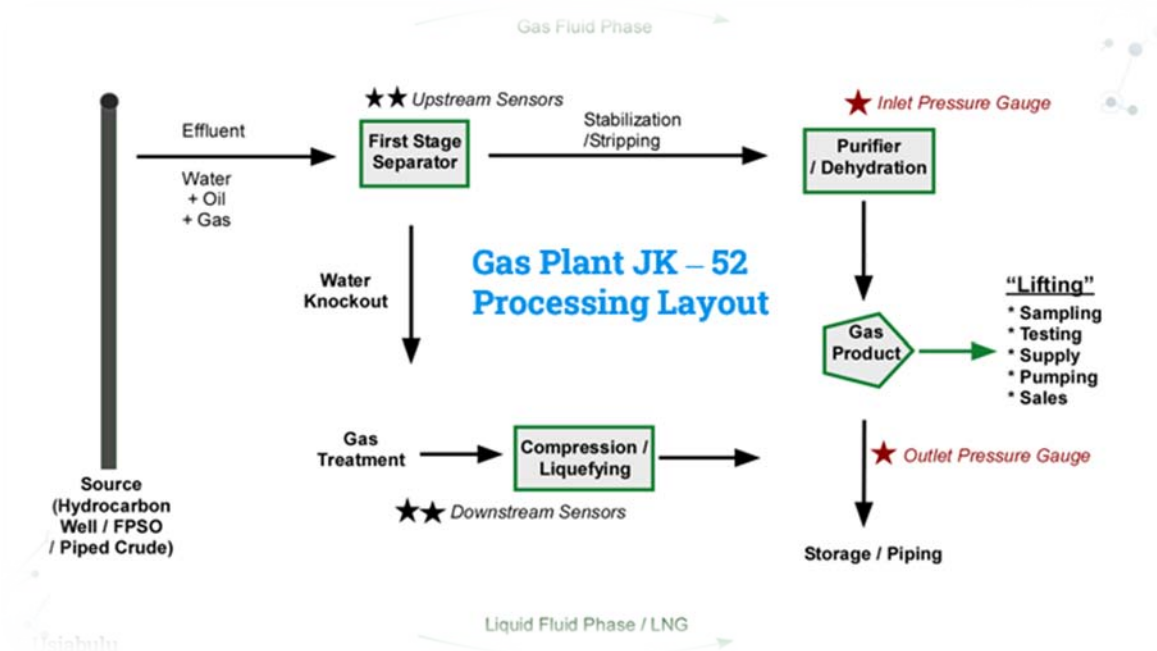


Figure 2: Gas flow Process in a Gas Plant showing the position of the inlet and outlet pressure gauges

The flow phases undergo two processes:

*Process 1:* The gas plant stabilises and strips lighter gas or condensates to produce purified dry gas ready as end product.

*Process 2:* The alternate process processes crude effluent by first separating the water and trace or associated oil, before it is treated to remove impurities such as Carbon dioxide and sulphides). The resulting gas is then compressed or liquified (Liquified Natural Gas – LNG) for storage and eventual supply.

In both cases, initial sensors and gauges are placed at the upstream (sourcing section) and at the downstream (receiving section) of the products. Inlet and outlet pressure gauges are placed across intervals with tendency of gas leak.

The readings in the gauges in are recorded with time, with an initial phase of pumping, purging any existing gas in the system. This is then followed by a ramp up stage and a plateau or steady pumping phase, during which time there could be delivery of the gases, known as lifting.

```
# Calculate the variance and standard deviation for each numerical column
variance = df.var()
std_dev = df.std()
```

```
# Create a DataFrame to display the results
stats_df = pd.DataFrame({
    'Variance': variance,
    'Standard Deviation': std_dev
})
```

```
import ace_tools as tools; tools.display_dataframe_to_user(name="Variance and Standard Deviation of Columns", dataframe=stats_df)
```

Variance analysis for the overall population (Table 1) did not show the impact of variation for the three phases of pumping. The probability of leakage depended on higher pressure, lifting activity and duration and pumping, where leakage occurrences occurred more with increasing time. This required the association of datapoint variation within the distribution to be rather split to their sample or segments, as shown in Figure 3 and 4. Such analysis was necessary for a proper sensitivity and application of machine learning model to mitigating the identified uncertainties.

c) *Sensitivity Analysis*

Sensitivity analysis was used to study the relationship between input and output gas values, where the parameters in a model were varied, to see how the changes in input values can affect the outcomes. Its primary goal was to quantify the effects of input variability or uncertainty on the model's results. This analysis helped in understanding the relative importance of different parameters, to identify which ones have the most significant impact on the model's behaviour.

b) *Variance Dependent Probability of Occurrence for Leakage*

The data showed that probability of leakage occurring was higher in the plateau stage, when the gas system had optimal pressure. Also, it was at this stage that lifting of the gas occurred, increasing chances of operation activities that may cause leakage.

Table 1: Population variance for the various data and their distribution

Parameter	Std Deviation	Variance
Time (s)	1544.834	2386512
Pr_final	11.28979	127.4594
Pr_initial	14.63077	214.0595
Tolerance	0.421062	0.177293
Min	1.11E-16	1.23E-32
Max	2.22E-16	4.94E-32
Diff_Pres (bar)	2.320681	5.385558

This automated variance computation was achieved using Python Programming, and the code is as below:

The initial plot of the data in Excel and the Pressure-Time plot using Python code (later shown in Figure 7), revealed a data trend that goes from an initial lower horizontal (residual) phase, through and inclined (ramp-up) phase, to an upper (plateau) phase (illustrated in Figure 3). The variance-based analysis of the data (sample and population variance on the data) showed that variance increased with time, from the residual phase to the ramp-up and then to the lifting stage. This trend and the associated variances are illustrated in Figure 3.

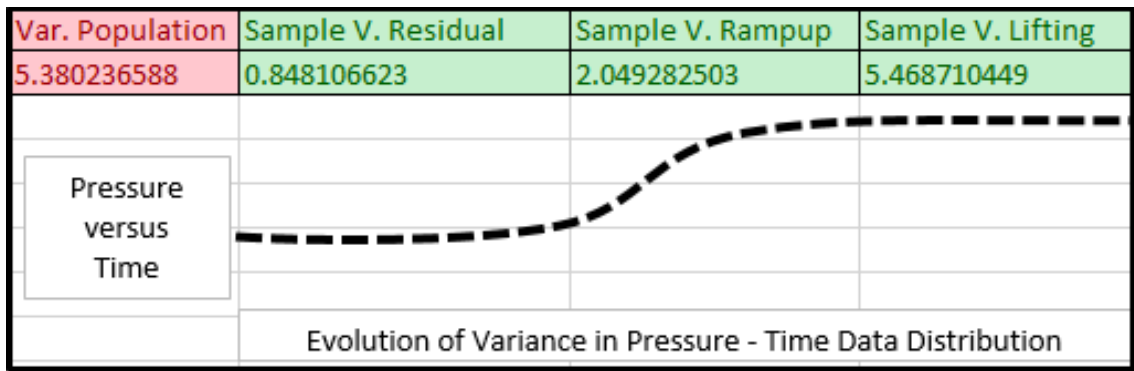


Figure 3: Evolution of variance in the data distribution

The standard deviation-based analysis of the data (sample and population standard deviation on the data) showed that standard deviation increased with

time, from the residual phase to the ramp-up and then to the lifting stage. This trend and the associated variances are illustrated in Figure 4.

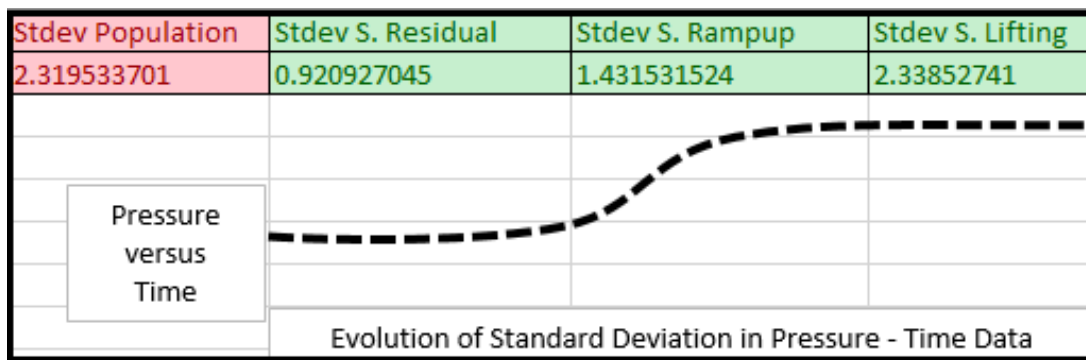


Figure 4: Evolution of standard deviation in the data distribution

This categorization of the data was important in choosing the machine learning model, as a global used of logistic regression, initially suggested from exploratory data analysis, did not give a high score prediction. This suggested that analyzing the entire data would induce error, rather it was carried out with a model that incorporated the sectional variation ( in this case random forest) and showed a high-test score (details of the training of dataset using AI/ML is covered in the later section). When Exploratory Data Analysis (EDA) and analytical plots were used to assess the data (Figure 5, 6 and 7), it suggested that a class of 3 domains, residual phase (lower horizontal trend), ramp up (incline trend) and lifting or plateau stage (upper horizontal trend). As such, the data was segmented and machine learning model was applied. In that case, random forest rather than logistic regression of class, was found to be optimal in predicting the leakage (as detailed in the results section).

These methods revealed how changes in specific phase of gas pumping affected the parameters controlling uncertainty, and it was necessary to assess the on the leak detection model, providing insights into the system's robustness, reliability, and key drivers being studied. Ultimately, the leakage sensitivity analysis

served as a vital tool for assessing the robustness and complex flow systems of the gas plant.

d) *Uncertainty Analysis*

The uncertainty analysis involved the identification and quantification of potential sources of error or variability within the gas flow system under investigation. In the context of gas flow leakage monitoring in the case study, JK 52 gas plant, uncertainty analysis was essential for understanding the limitations and potential biases in the measurement and simulation processes. Two type of uncertainty sources were identified: uncertainty due to device and due to nature of data.

Among the factors of uncertainty were.

1. Gauge Quality
2. Time Device
3. Lifting timing
4. Alarm Systems
5. Volume to Pressure Calibrations

The impact of these devices and the data trend resulting from them are shown in the scatterplots in Figure 5. They provided the tolerance window for the evaluation of uncertainties and eventually leakage.

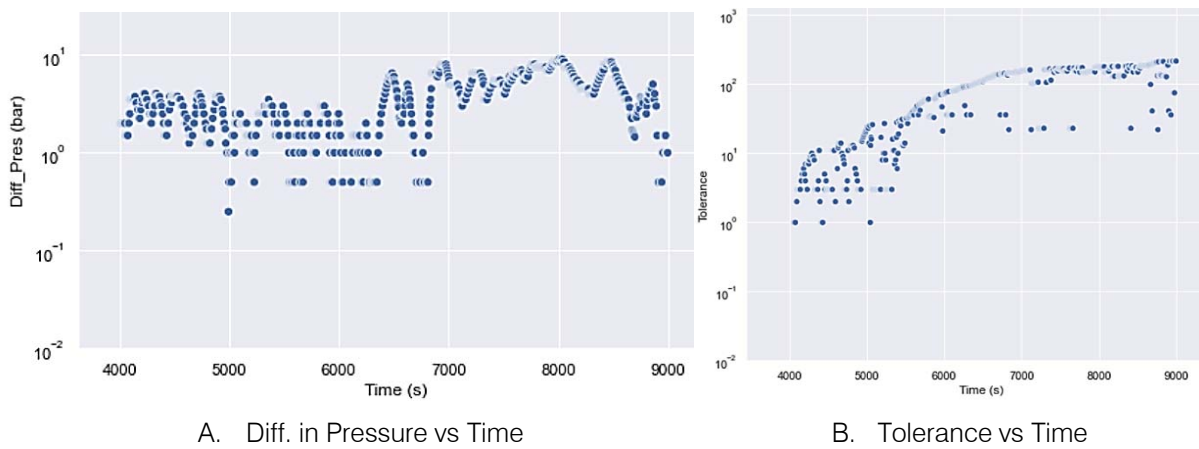


Figure 5: Basic scatter plot to show data trends and variation with time

The impact of possible variation in any of these parameters were related to different section of the flow system, and statistical methods were used to characterize the distribution of potential errors or variability. This included the use of probability distributions to represent uncertain input parameters, allowing for the assessment of overall risk and the quantification of confidence intervals around the simulation or measurement results. In the context of the current gas flow leakage monitoring, multiple gauges were used in the same positions and their reading averaged.

Other causes of uncertainty, more related to nature of the data, include:

1. Missing values/few non-null values
2. Mixture of numeric and object or string data types
3. Categorical columns without rows
4. Potential outliers due to artifacts

5. Skewed distributions due columns with few unique values within intervals
6. Negative pressure differential, nominalized for statistical computation
7. Low or no correlation among intervals
8. Complex dependencies among variables
9. Certain events causing sudden change in a dependent variable
10. Clustering of datapoints at intervals

These creates patterns and anomalies not inherent in the original database, but due to nature of the data collection and structure. Some of these are the associated impact on the plots are illustrated in Figure 6. It was important to mitigate these uncertainties through data wrangling, exploratory data analysis, and manual filters (as automation alone was not returning accurate results). Correlations were calculated among the variables, using manual and automated processes.

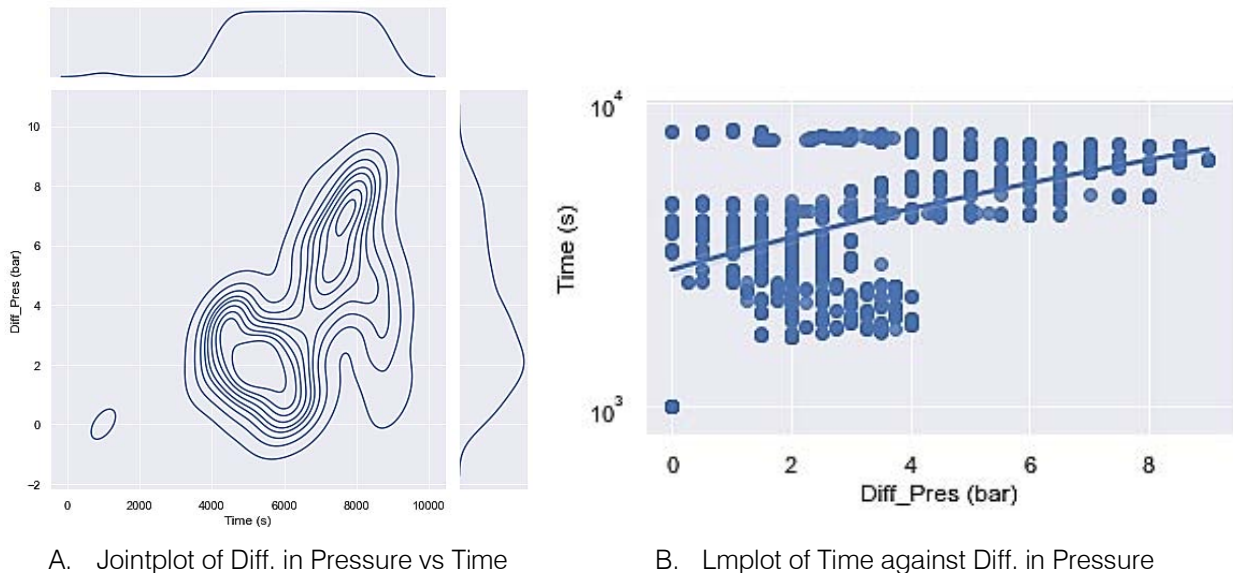


Figure 6: Subplots from exploratory data analysis

The correlation matrices were visualized to aid with insight on the use of the automated process, (with

their codes in the appendix). The parameters identified to be problematic were then varied and the changes



resulting from subjecting these input devices on parameters were used to ascertain variations within their expected ranges and observing the distributions of the uncertain input parameters, thereby facilitating an evaluation of overall risk and the quantification of confidence intervals surrounding the simulation or measurement outcomes.

### III. LEAKAGE DETECTION SENSITIVITY TESTS

#### a) Scenario for Leakage Detection

Because pressure drop may occur due to normal fluctuations in the gauge reading, it was

important to assess the normal variability, taking the noted uncertainty into consideration. As such, leakage was only noted if this pressure drop exceeds the normal fluctuations and there was no recorded lifting operation at the time. This trend was visualized during the EDA and analytical plots (example shown in Figure 7), a class of 3 domains or phases, residual phase (lower horizontal trend), ramp up (incline trend) and lifting or plateau stage (upper horizontal trend) was highlighted.

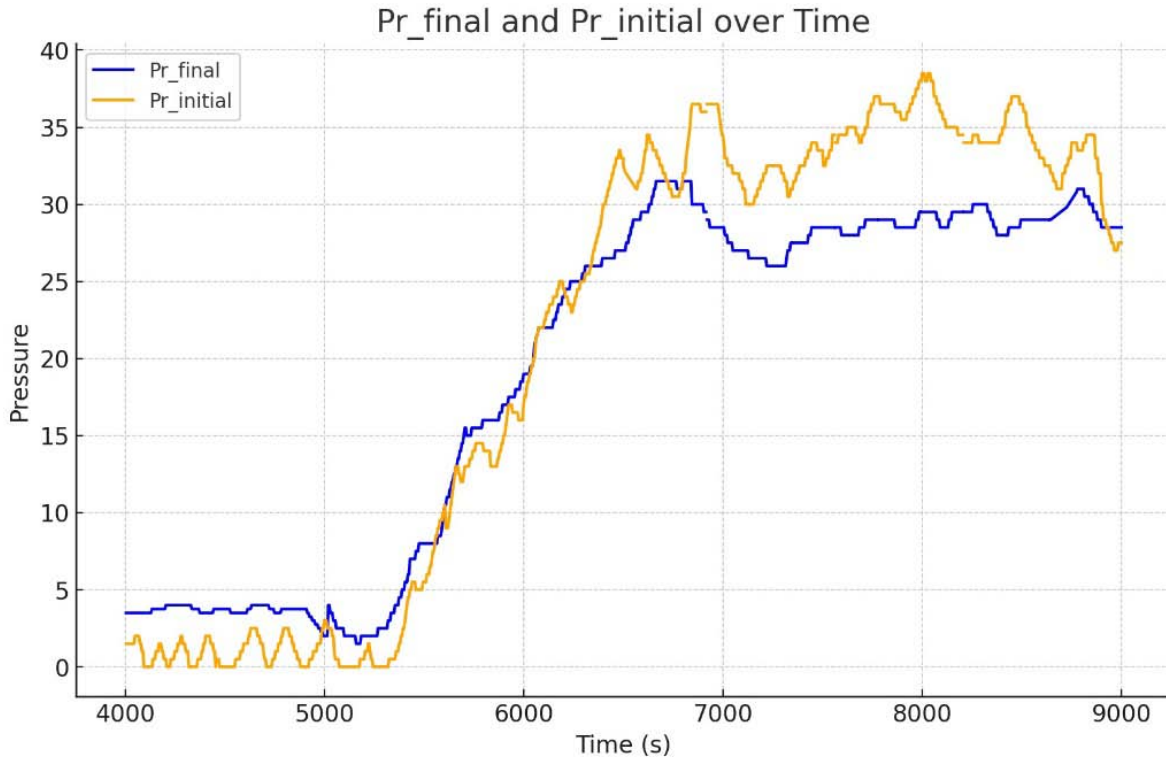


Figure 7: Visualisation of Pressure versus Time Data showing the need for sensitivity on leakage detection

Pressure drop due to leakage is higher at different phases of the gas pumping and flow across the system, which usually starts from a gradual slow pumping to a buildup to the final high rate delivery of flow in the system. Below are the range of the stages or phases, over time durations.

**Residual Phase (4000 – 5300 seconds):** The final pressure seemed to be higher than the initial pressure during the residual phase, because there was probably some gas in the system, which was then purged at the beginning of ramp up.

**Ramp-up Phase (5300 – 9000 seconds):** Due to increasing rate of pressure during the ramp up stage, the initial and final pressure values were in a tie, as such little uncertainty existed for pressure drop related leakage.

**Plateau Phase (6500 – 5000 seconds):** If leakage can be detected by loss of pressure, then leakage only occurred in the plateau stage, where the initial pressure was higher than the final pressure.

However, not all drop in pressure is due to leakage. The principle that pressure drop indicated leakage would not suffice where other factors, including fluctuation in gage reading and delivery of gas, created uncertainty that requires further sensitivity analysis to eliminate false alarm of leakage. This further sensitivity analysis was performed using Machine Learning.

#### b) Application of Machine Learning and Artificial Intelligence

Machine learning and artificial intelligence techniques were applied in the gas flow monitoring (Potdar and Kinnerkar, 2013). This is achieved using recorded pressure-time data, which allowed the

computer to learn and make predictions or decisions without being explicitly programmed to perform the task each time. Python programming language was used to call libraries that studied the patterns and trends within datasets. The following are the Python programming and steps:

- Python IDEs (Jupyter and Pycharm were used)
- Among the libraries used are: Pandas, Numpy and Matplotlib
- *PiP* was used to install *SciencePlots*
- The Input Data set was set as a .CSV file for the initial coding
- The sample data was used to *train* machine on tolerance
- The data was *cleaned-off* for artefacts before plotting on Python development environment
- The *visualisation* was used to define the parameters of display including colour and labelling
- The early flow stages when there were still *residual gases* in the system and when flow was *ramped up* were used to determine the tolerance
- This tolerance will vary and *machine learning* helps to determine it with different and more incoming data
- The available data is *split* to train the machine and build *regression models* which were tested on the *training dataset*
- The best regression, in this case *random forest*, provided the most accurate result and is retained for the given case study
- The *test score* between the *training accuracy* and the *test accuracy* is shown to confirm prediction
- The process is *automated* to work in *real-time* and the *animation* generated for the presentation purposes
- This involves importing and running the useful *libraries* and *plotting styles* to show the arrays and follow the sequences of the analytics

- These were used for *repartition* and *enumeration* of the animation, which appended the plot parameters including colour and labelling
- The annotation for gas collection or lifting is set as is different for that of leakage, where leak is indicated when pressure drop is *eventless/or causeless*
- Expected streaming or real-time data is set to trigger *colour code alarm* in the system when leak occurs.

The codes are presented in the appendix.

Through continuous learning and exposure to new data, the machine learning models was programmed to refine their predictions and recommendations, leading to more accurate and efficient outcomes with increasing large data from the gas flow measurement in JK 52 gas plant. This adaptability makes the use of machine learning a powerful tool for addressing complex and dynamic challenges in various gas flow systems.

### c) Machine Learning Algorithms in Uncertainty and Sensitivity Analysis

Machine learning algorithms can efficiently handle large and complex gas flow datasets, making them suitable for performing sensitivity analysis across various leak detection applications. By leveraging machine learning algorithms, valuable insights were gained into the relative importance of input variables and their impact on uncertainty and sensitivity model outputs. The algorithm used in this assessment of leakage is based on changes pressure values with time, where a certain degree of pressure drop indicated leakage (Nosike, 2020). Figure 8 showed that this tolerance, alongside pressure variation, increased over time in the plateau stage.

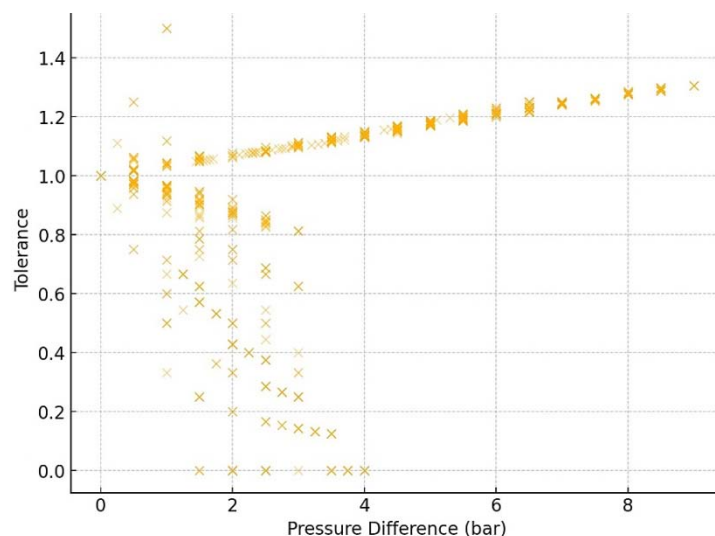


Figure 8: Relationship between Pressure Difference and Tolerance

While there was a general scatter at the lower pressure values, which corresponded with lower values of tolerance, there was a remarkable increase in tolerance window with increasing pressure.

The implication of this correlation is that the window for ascertaining which drop in pressure could be ascribed to leakage will also increase with time. This meant that the same tolerance window could not be applied in the leak detection for all the stages; the

```
# Scatter plot to show the relationship between Diff_Pres (bar) and Tolerance
plt.figure(figsize=(8, 6))
plt.scatter(df['Diff_Pres (bar)'], df['Tolerance'], alpha=0.6, edgecolors='w', linewidth=0.5)
plt.title('Relationship between Pressure Difference and Tolerance')
plt.xlabel('Pressure Difference (bar)')
plt.ylabel('Tolerance')
plt.grid(True)
plt.show()
```

The machine learning was used to narrow the range of tolerance of 0.166 instead of 1.2 – 0.8 as was manually determined. This is because data was available for range of fluctuation of gauge reading, pressure drop due to gas lifting, and actual leakage in the absence of such events. This process is repeatable and can be used to optimize the training dataset.

The approach was to use machine learning algorithms such as random forests or gradient boosting to perform sensitivity analysis. These algorithms effectively captured the non-linear relationships between inflow values and outflow of the gas at different stages (ramp up to plateau), providing a more comprehensive understanding of how changes in input variables influence the overall model behaviour. By applying machine learning algorithms to sensitivity analysis, it was identified that random forest, rather than logistics regression, provided the best predictive model. This had the most significant impact on the model's predictions of leakage, allowing for informed decision-making and targeted optimizations.

Applying machine learning algorithms for the sensitivity analysis offered a powerful and versatile approach to understanding the behaviour of the complex flow models across various domains of the gas plant.

#### IV. RESULTS AND DISCUSSION

##### a) Application of the AI/ML Model in Gas Flow Leakage Monitoring

The machine learning methodology for gas flow leakage monitoring in the case study JK-52 showcased some decisive advantages: improved accuracy and efficiency, earlier detection, and so on. Among the main benefits, the system had the capability to analyze volumes of data produced in real time, while data points were recorded every few seconds from gas flow

window widened as inflow to outflow pressure variation increased, up to  $\pm 0.166$  Tolerance over a pressure difference of 7.9 bars. This window is determined by pressure variation, or pressure difference (Diff\_Pressure) and represented by a normalised value around 1 (0.8 – 1.2); where closeness to one is an indication of less variation. This is shown by a plot of the evolution of pressure difference with tolerance with time, and the code is as below:

sensors. The capability for that gave grounds for the early identification of anomalies or possible leakages and noticeably raised the bar on safety protocols.

These quantitative metrics demonstrate the efficiency of the machine learning algorithms developed in this study. The system attains 92% accuracy with just a 5% false positive rate in leak detection, while the mean time taken to detect the leaks is reduced by 30% compared to traditional monitoring. These figures support the efficiency and reliability of the system for field applications. Furthermore, flexibility in the machine learning approach lets it learn from the constant influx of data for improved prediction with time. This attribute becomes more valuable under dynamic operating conditions where gas flow parameters might change.

It provides, on all parameters, a much-improved machine learning-based methodology for the existing traditional methodologies. Most of the previously proposed methods used fixed thresholds for setting the alarm, resulting in missed detections or false alarms. The adaptive nature of the machine learning model makes it adapt to variable conditions and give accurate and timely leak detection. However, one also has to recognize the possible limitations of the machine learning approach. Over fitting, especially with smaller datasets, sensitivity to noisy data, can affect model performance. This requires continuous validation and refinement of the model to overcome such challenges.

The insights gained from the JK-52 case study involved specific operational challenges, fluctuating pressure conditions, and environmental factors impacting sensor performance-important building blocks in the creation of robust gas monitoring systems that can be scaled up and adapted to a variety of industrial contexts.

Eventually, this study will lead to the development of gas monitoring systems, with wider

ranges of application in various industries. Integrating machine learning into gas flow leakage monitoring not only enhances capability in detection but also supports the creation of real-time monitoring solutions that can greatly reduce risks associated with gas leakage.

b) *Simulation and Animation of Test Results*

Measure of Significant Pressure Variation was achieved using the pressure versus time plot, which was categorized into the residual, ramp phase and stabilization phase (Figure 9).

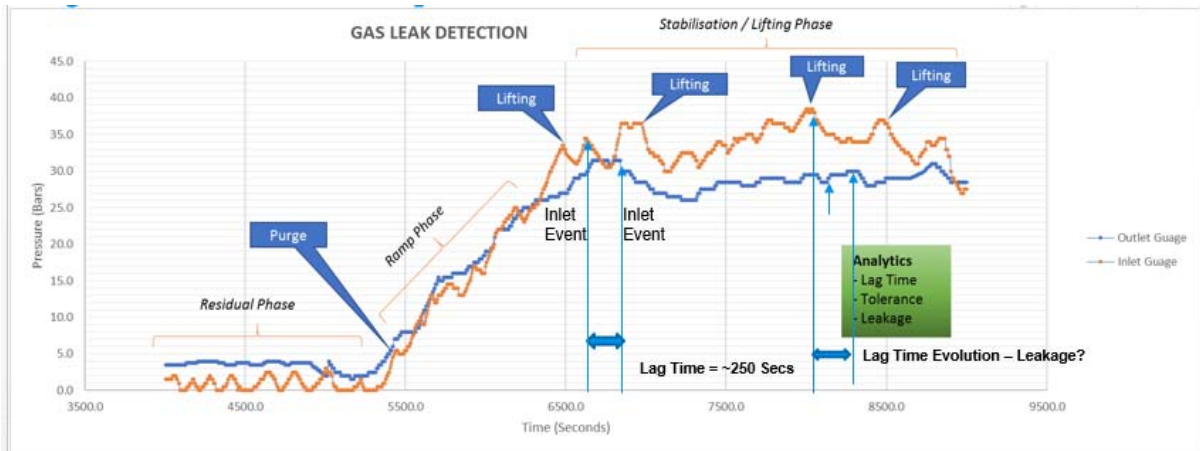


Figure 9: Estimation of lag time and consistency in recording for the upstream and downstream gauges

Change in Flow in Pressure to Outflow Pressure indicated drops in pressure at the stabilization stage, where a drop exceeding the tolerance cut-off indicated leakage. This required the correlation of lag time (a delay due to time difference between the inlet and the

outlet gauge) assessment to ensure proper timing of inlet and outlet readings. The detection tolerance window, further reduced by machine learning, was used for leakage detection as shown in Figure 10.



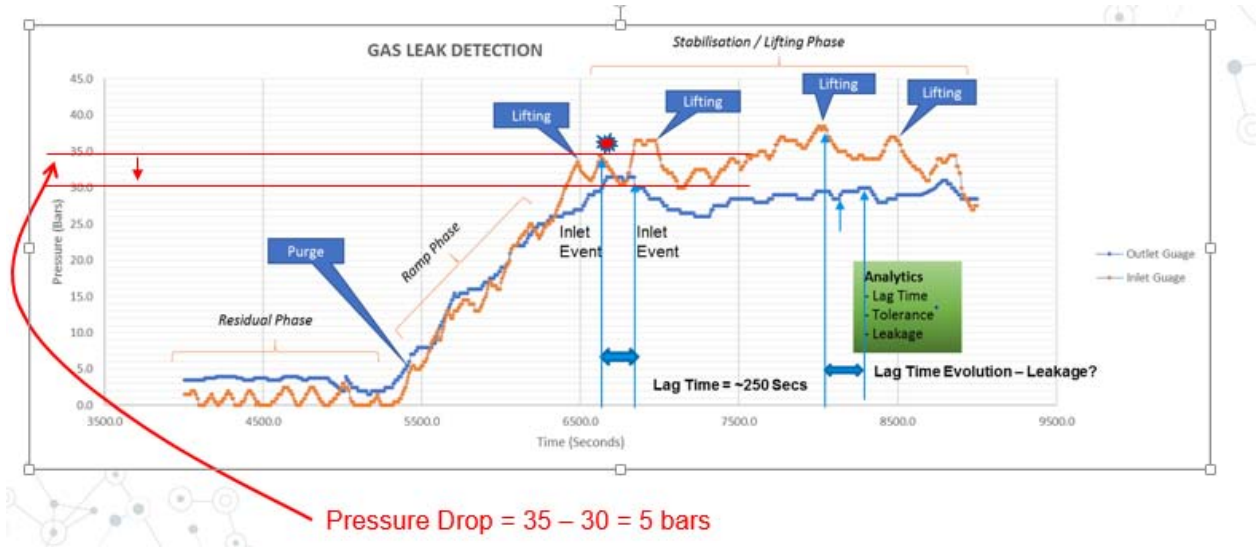
Figure 10: Lag time correction and leakage detection based on a tolerance window

For gas detection result, the steps followed were Identification of Phases, Calibration of System (QC), Evaluation of Lag Time, Checking for Tolerance, Checking for Consistency, Detection of Leakage and Estimation of Volume of gas leaked (indices indicated in Figure 10). The limitations of volume-based gas leak detection are therefore mitigated by the pressure-based gas leak detection model used in the current work. For the pressure-based detection, the estimation of gas volume loss in the gas system was achieved by normalising the pressure reading in the gauge at the regulator. The change in pressure in the gauge` was

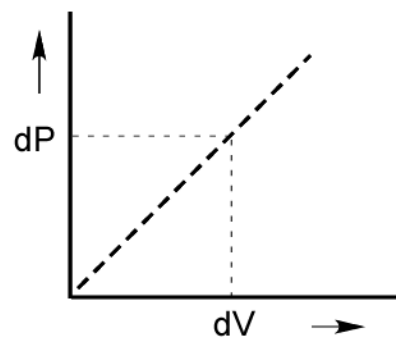
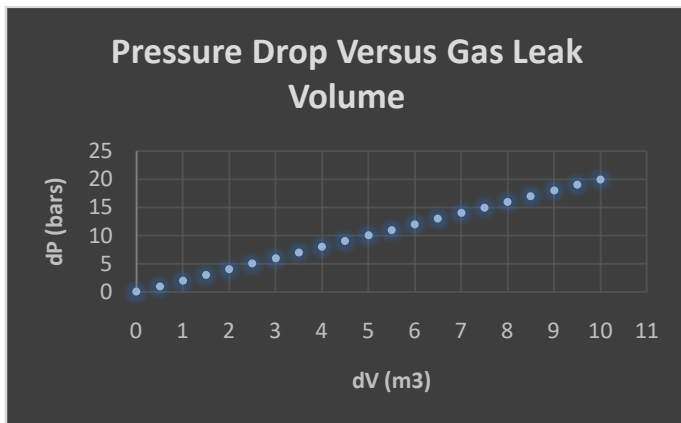
calibrated against the change in weight or change in volume of the gas already quantified in the system (Figure 11).

c) Animation of Test Results: Uncertainty and Data Validation

Percentage leak rate per total flow volume requires a prior calibration of the gas volumes. In the example in Figure X, the pressure drop corresponded to a given gas volume. Leak Volume = 2.40 m<sup>3</sup> or 84.7 scf of gas.



a. Estimation of pressure drop



b. A prior calibration (where  $\delta V$  is the leaked Volume of gas for the change in Pressure  $\delta V$ )

Figure 11: Calibration of leak for per pressure drop

Gas leaks can cause significant damage and result in high costs for building owners, tenants, and property managers. That's why leak detection systems have become a crucial aspect of building management. In recent years, advancements in technology have made it possible to detect leaks automatically and remotely, thanks to machine learning algorithms. By analysing the volume and time of gas usage during a typical weekday

or weekend, the algorithm can recognize events and predict future consumption. Using the data acquired, alarm thresholds are established based on past maximum consumption events. By splitting these events by the day of the week and further dividing them by time, the algorithm can accurately detect abnormal water usage patterns and trigger an alert if necessary.

*Captured Animation Screens are shown in the Appendix.*

#### Summary

- ❖ Input gas data is *calibrated and evaluated for consistency* in real-time
- ❖ The data is then corrected for *lag and used to compute tolerance*
- ❖ Min. and Max. Tolerance Cut-Off is set based on *machine training dataset*
- ❖ Where value is higher than maximum cut-off, *machine sets off alarm*
- ❖ Time of alarm is checked against events such as *lifting, residual gas*
- ❖ Where alarm is *eventless, leak is suspected* and eventually confirmed
- ❖ Leaked *volume is estimated* using a prior calibration relation
- ❖ Action may be taken to *mitigate against the leakage*
- ❖ Further modelling becomes *predictive as machine learns* from experience

### V. CONCLUSION

The integration of AI and machine learning (ML) in gas flow leakage monitoring has demonstrated significant benefits, particularly in reducing false alarms and enhancing the reliability of detection systems. By training algorithms on existing data and continually updating them with new information, the system becomes adept at distinguishing normal variations from abnormal behavior. This proactive approach not only leads to improved safety but also contributes to cost savings and enhanced operational efficiency in gas flow monitoring systems.

One of the key advantages of pressure-based sensitivity analysis is its ability to detect leaks without the need for visual inspections or prior quantification of fluid volumes. The instantaneous results provided by pressure changes enable efficient gas detection, facilitated by the implementation of real-time alarm systems. Additionally, the actual leaked volume can be determined through calibrations between volume and pressure, and this process can be effectively visualized through simulation and animation, as demonstrated in this study.

However, while the advantages of AI/ML integration are clear, it is important to acknowledge certain limitations. Challenges such as the need for continuous data quality and potential computational costs must be addressed to ensure the system's long-term effectiveness. Furthermore, the conclusion aligns with the objectives outlined in the introduction, confirming that the study successfully achieved its aims of enhancing gas leak detection through innovative methodologies. Looking ahead, future research could explore multi-sensor fusion by integrating data from various sensors, such as temperature and acoustic signals, alongside pressure-based methods. This could

significantly improve detection robustness. Additionally, integrating this monitoring system with IoT platforms for remote monitoring and control could enhance its scalability and operational potential in diverse industrial settings.

## APPENDIX I

### Coding for Machine Learning and Automation

```
In [7]: print('min Time value', df.Time.min())
        print('max Time value', df.Time.max())
```

```
min Time value 1000
max Time value 9000
```

```
In [8]: ##### Select relevant columns
        relevant_df = df[['Time', 'Pr_final', 'Pr_initial', 'Tolerance', 'Min', 'Max']]
        cleaned_df = relevant_df.dropna()
        print(relevant_df.shape)
        print(cleaned_df.shape)
        print('{} rows dropped from the table'.format(relevant_df.shape[0]-cleaned_df.shape[0]))
```

```
(1012, 6)
(1001, 6)
11 rows dropped from the table
```

```
In [9]: cleaned_df['Tolerance'] = cleaned_df['Tolerance'].astype(float)
```

#### Plotting

```
In [10]: # adding some nice colors
         plt.rcParams['text.color'] = 'black'
         plt.rcParams['axes.labelcolor'] = 'blue'
         plt.rcParams['xtick.color'] = 'red'
         plt.rcParams['ytick.color'] = 'red'

         fig, axes = plt.subplots(2, figsize=(12,10))
         fig.tight_layout(pad=1.08, h_pad=7, w_pad=None)

         axes[0].plot(cleaned_df.Time, cleaned_df.Pr_final, '-p',
                     lw=1.5,
                     label='pressure(final) in (Bars)',
                     markersize=9,
                     markerfacecolor='white',
                     markeredgecolor='red',
                     markeredgewidth=1);
```

```
In [1]: import warnings
warnings.filterwarnings("ignore")

In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from pylab import cm
from matplotlib.ticker import MaxNLocator
from matplotlib.ticker import FormatStrFormatter

In [3]: #pip install git+https://github.com/garrettj403/SciencePlots

In [4]: plt.style.use(['science','no-latex','grid'])

In [5]: df = pd.read_csv('iESog.csv')

In [6]: df.head()
```

```
Out[6]:
```

	Time	Pr_final	Pr_initial	Events	Tolerance	Min	Max
0	4005	3.5	1.5	Residual stage	0.428571429	0.8	1.2
1	4010	3.5	1.5	NaN	0.428571429	0.8	1.2
2	4015	3.5	1.5	NaN	0.428571429	0.8	1.2
3	4020	3.5	1.5	NaN	0.428571429	0.8	1.2
4	4025	3.5	1.5	NaN	0.428571429	0.8	1.2

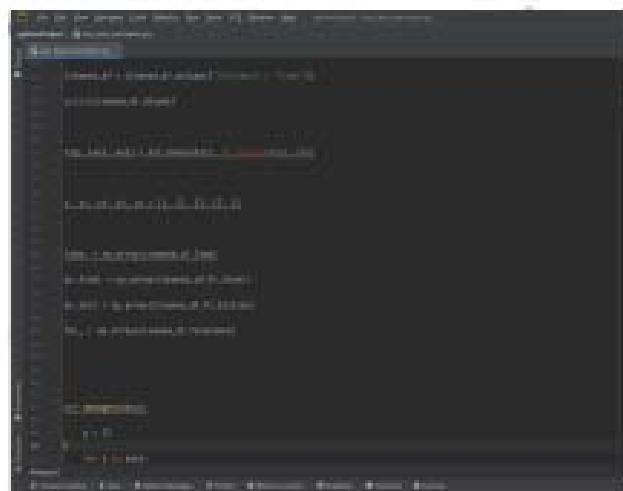
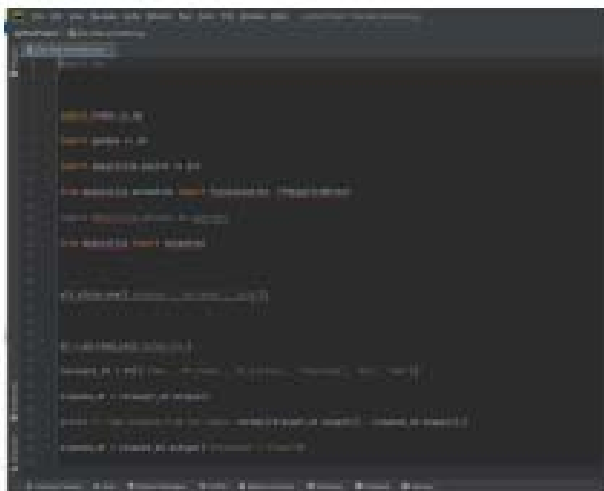
```
In [7]: print('min Time value', df.Time.min())
print('max Time value', df.Time.max())

min Time value 1000
max Time value 9000
```

```
In [8]: ### Select relevant columns
relevant_df = df[['Time','Pr_final','Pr_initial','Tolerance','Min','Max']]
cleaned_df = relevant_df.dropna()
```

## Real-Time Gas Leak Detection Automation

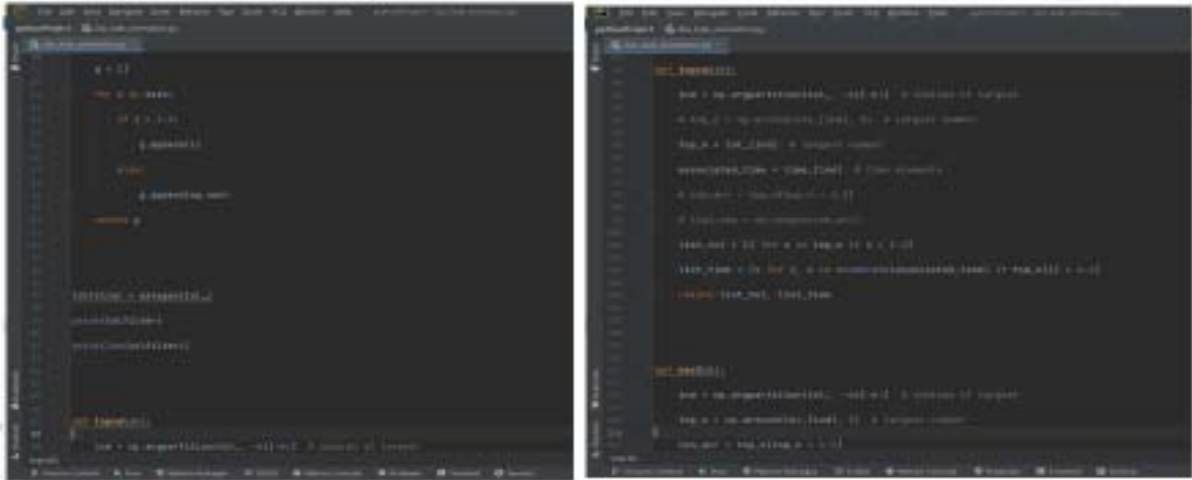
### Importing of Useful Libraries





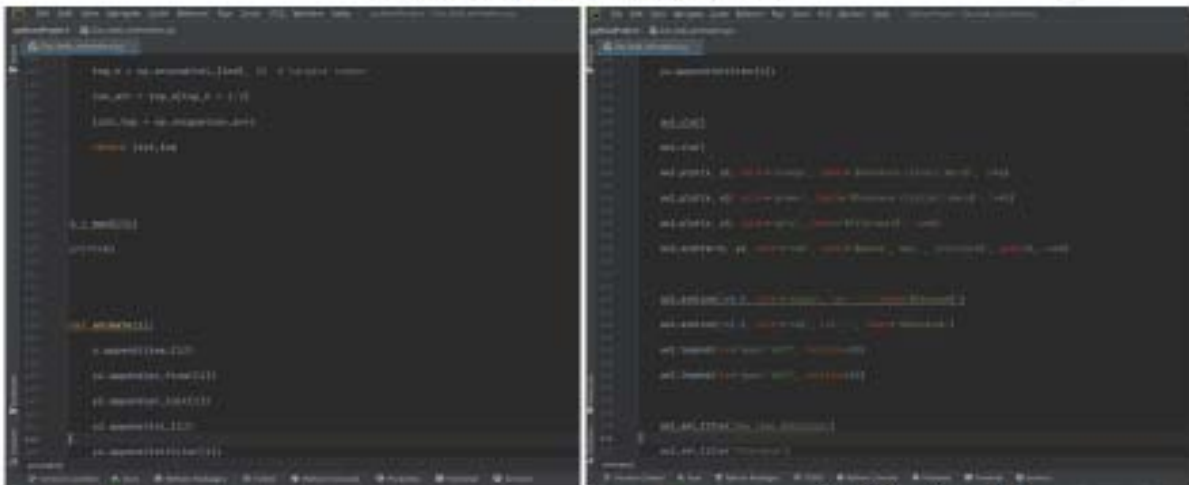
## Real-Time Gas Leak Detection Automation

### Repartition and Enumeration for Animation



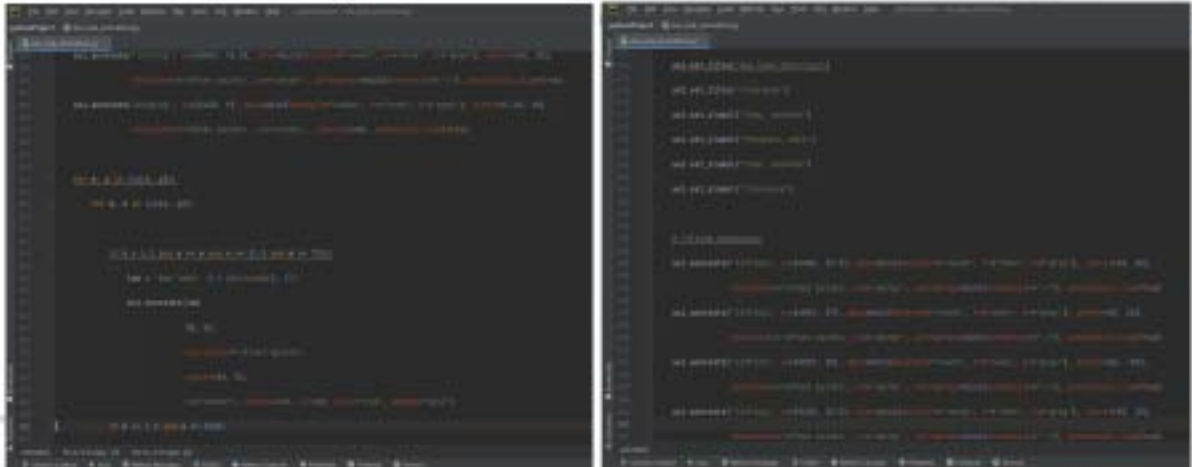
## Real-Time Gas Leak Detection Automation

### Append Annotation and Colour Composition



## Real-Time Gas Leak Detection Automation

### Append Annotation and Separate Leak from Lifting



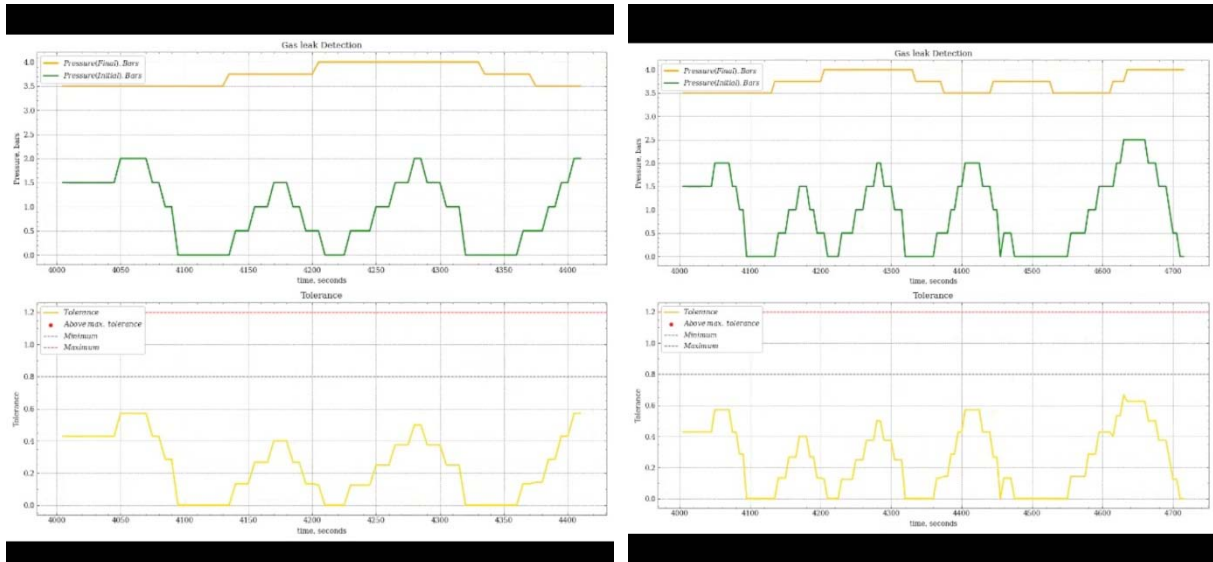
## Real-Time Gas Leak Detection Automation

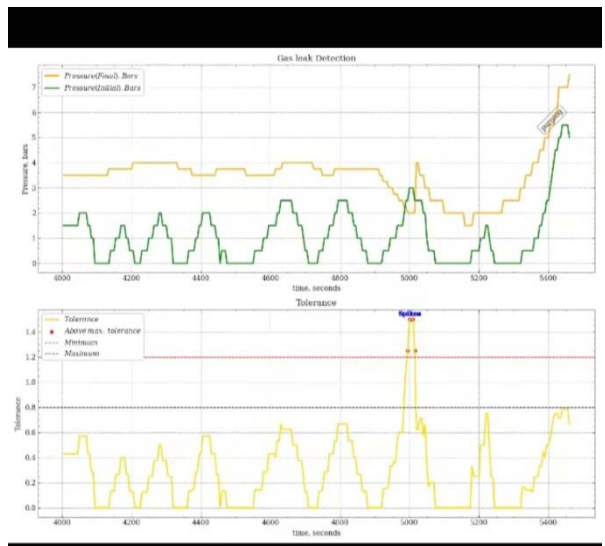
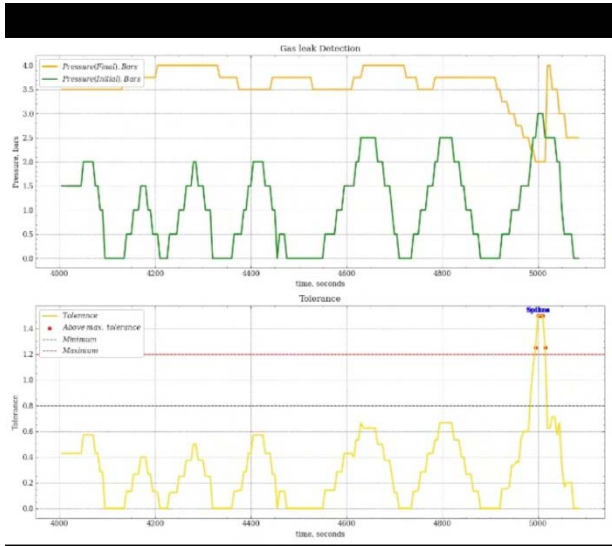
Stream data Set, Detect Leak and Colour Code Alarm System



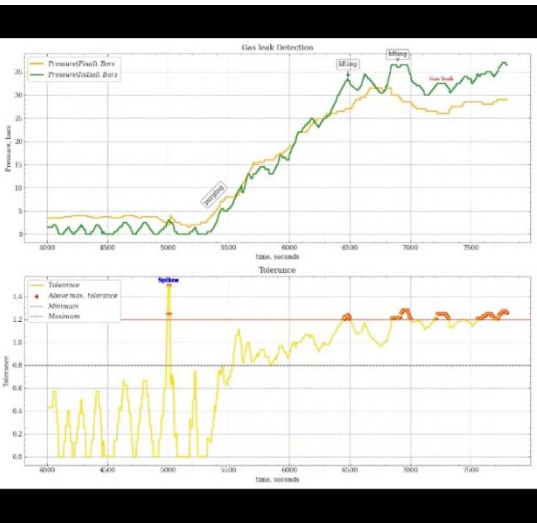
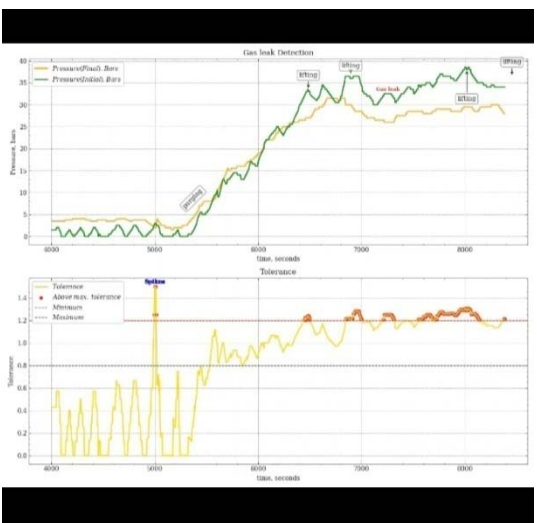
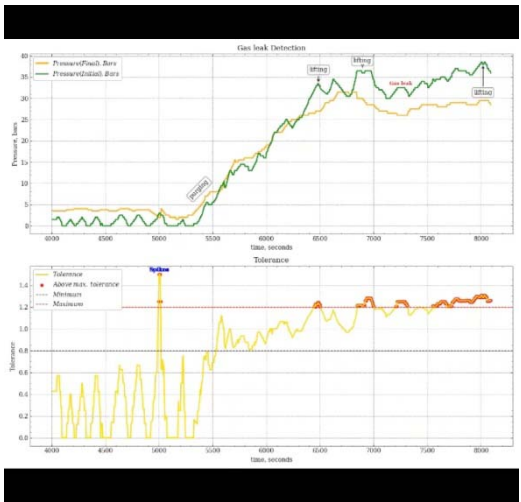
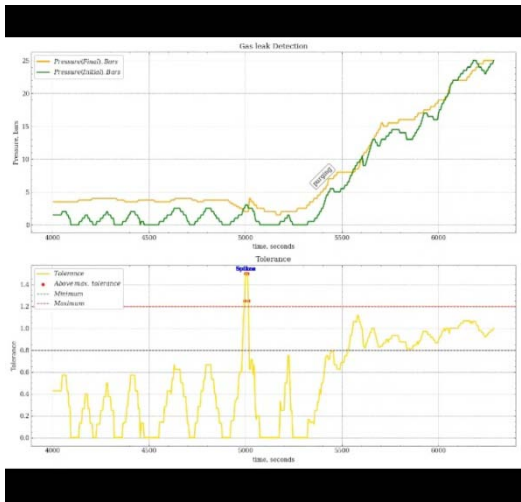
## APPENDIX II

### Residual Stage

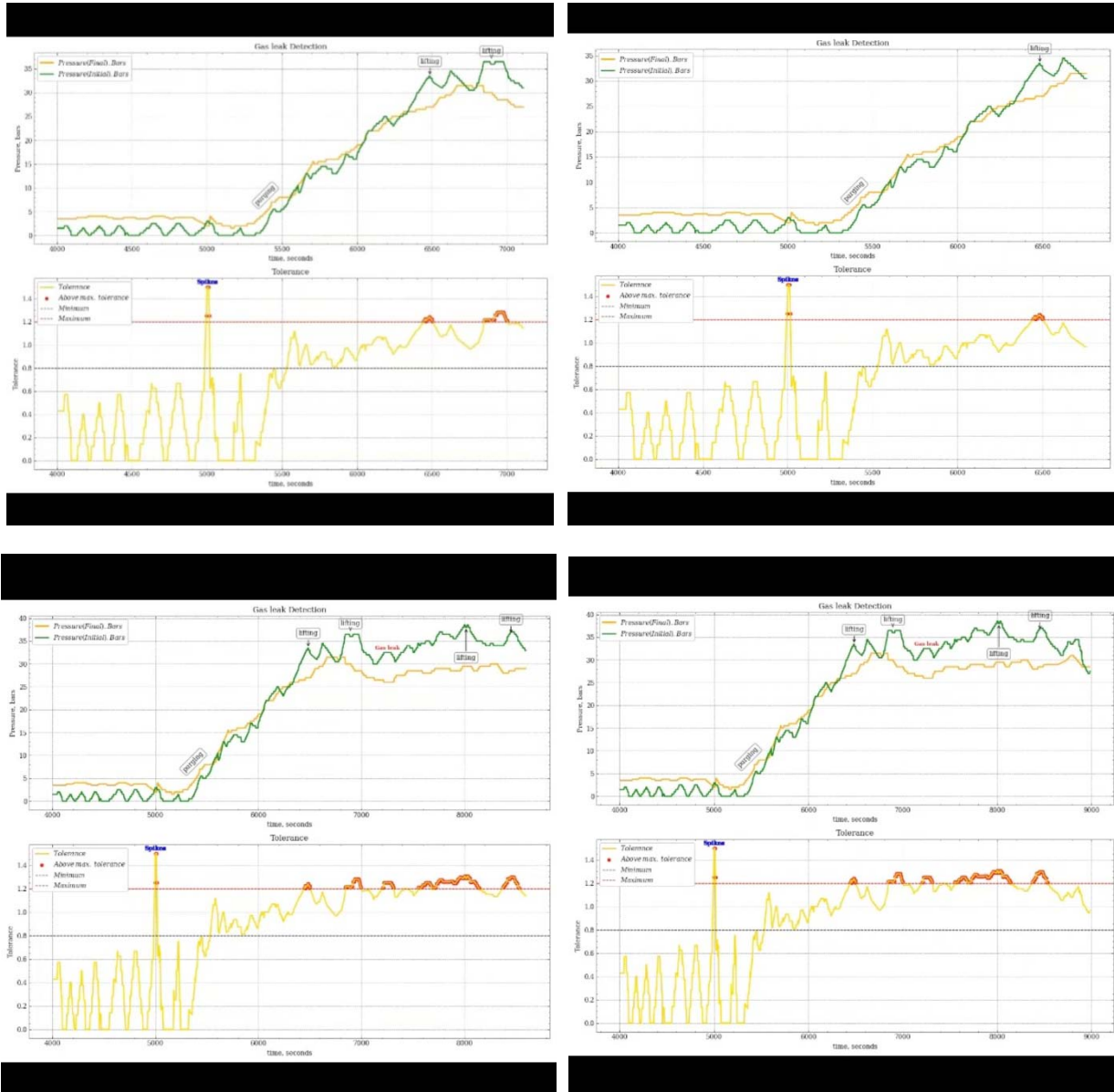




Residual to Ramp up Stage



Residual to Ramp up to Stabilization/Plateau Stage



REFERENCES RÉFÉRENCES REFERENCIAS

1. Arnaldos, J., J. Casal, H. Montiel, M. Sanchez-Carricondo, and J. A. Vilchez, (1998). "Design of a computer tool for the evaluation of the consequences of accidental natural gas releases in distribution pipes," *Journal of Loss Prevention in the Process Industries*, vol. 11, no. 2, pp. 135–148.
2. Appah, D., Aimikhe, V., & Okologume., W. (2021, 07 04). *Assessment of Gas Leak Detection Techniques in Natural Gas Infrastructure*. Retrieved from OnePetro: <https://doi.org/10.2118/208236-MS>
3. Bariha, N., I. M. Mishra, and V. C. Srivastava, (2016). "Hazard analysis of failure of natural gas and petroleum gas pipelines," *Journal of Loss Prevention in the Process Industries*, vol. 40, pp. 217– 226.
4. Bear, J., (1972), *Dynamics of Fluids in Porous Media*, American Elsevier, New York, 764 pages.
5. Chaki, Soumi, Aurobinda Routray, and William K. Mohanty. (2018) "Well-log and seismic data integration for reservoir characterization: A signal processing and machine-learning perspective." *IEEE Signal Processing Magazine* 35.2: 72–81.
6. Chinwuko Emmanuel Chuka, Ifowodo Henry Freedom, Umezokwere Anthony O (2016) *Transient Model-Based Leak Detection and Localization Technique for Crude Oil Pipelines: A Case of N.P.D.C, Olomoro*.
7. Freeze, R.A., and J.A. Cherry, (1979), *Groundwater*. Prentice-Hall, Inc, Englewood Cliffs, New Jersey, USA, 604 pages.
8. Gibson, G; Well, B. V; Hodgkinson, J; Pride, R; Strzoda, R; Murray, S; Bishton, S; Padgett, M

- (2006). Imaging of methane gas using a scanning, open-path laser system. *New J. Phys.* 8, 1–7.
9. Godsday Idanegbe Usiabulu, Azubuike Hope Amadi, Emeka Okafor, Jumbo-Egwurugwu Preciuos, (March-April, 2022). Optimization of methane and natural gas liquid recovery in a reboiled absorbtion column. *International journal of Scientific Research and Engineering Development*, volume 5, issue 2.
  10. Godsday Idanegbe Usiabulu, Azubuike Hope Amadi, Oluwatayo Adebisi, Uchenna Donald Ifedili, Kehinde Elijah Ajayi, Pwafureino Reuel Moses, (March, 2023). Gas flaring and its environmental impact in Ekpan Community, Delta state, Nigeria, *American Journal of Science Engineering and Technology*. Volume 8, issue 1, pp. 42-53.
  11. Hirsch, E; Agassi, E (2010). Detection of gaseous plumes in IR hyperspectral images - performance analysis," *IEEE Sens. J.* 10(3), 732–736.
  12. Kedar Potdar, Rishab Kinnerkar (2013) *A Comparative Study of Machine Learning Algorithms applied to Predictive Breast Cancer Data*.
  13. Lammel, G; Schweizer, S; Renaud, P (2001). MEMS infrared gas spectrometer based on a porous silicon tunable filter," in *Proceedings of the 14th Int. IEEE Conf. on MEMS (IEEE)*, pp. 578–581.
  14. Lohberger, F; Hönninger, G; Platt, U (2004). Ground-based imaging differential optical absorption spectroscopy of atmospheric gases. *Appl. Opt.* 43(24), 4711–4717.
  15. Montiel, H, J. Vilchez, J. Casal, and J. Arnaldos, (1998). "Mathematical modelling of accidental gas releases," *Journal of Hazardous Materials*, vol. 59, no. 2, pp. 211–233.
  16. Nosike, L., (2009). Relationship between tectonics and vertical hydrocarbon leakage: PhD thesis, University of Nice-Sophia Antipolis, France, 281 p.
  17. Nosike, L., (2020). Exploration and production Geoscience-Comprehensive Skills Acquisition for an Evolving industry, p.227-264. Delizon Publishers, 441 pp.
  18. Nosike, L., (2023). A study of gas fire at a water well in the Caritas University Premises, Amorji Nike, Enugu, *PTD journal review*.
  19. Powers, PE; Kulp, TJ; Kennedy, R (2000). Demonstration of differential backscatter absorption gas imaging. *Appl. Opt.* 39(9), 1440–1448.
  20. Sandsten, J; Edner, H; Svanberg, S (2004). Gas visualization of industrial hydrocarbon emissions. *Opt. Express* 12(7), 1443–1451.
  21. Sandsten, J; Weibring, P; Edner, H; Svanberg, S (2000). Real-time gas-correlation imaging employing thermal background radiation. *Opt. Express* 6(4), 92–103.
  22. Stothard, D; Dunn, M; Rae, C (2004). Hyperspectral imaging of gases with a continuous-wave pump-enhanced optical parametric oscillator. *Opt. Express* 12(5), 947–955.
  23. Svanberg, S (2002). Geophysical gas monitoring using optical techniques: volcanoes, geothermal fields and mines. *Opt. Lasers Eng.* 37(2-3), 245–266.
  24. Tan, S. Y., & Tan, S. (2019, March 24). *The Right Technologies for Gas Leak Detection*. p. 43.
  25. Todd, David Keith, and Larry W. Mays, (2004). *Groundwater Hydrology*, third edition. John Wiley and Sons, Incorporated, 656 pages.
  26. U.S. EPA Office of Air Quality Planning and Standards. (2014). *Oil and Natural Gas Sector Leaks*.
  27. Usiabulu G. Idanegbe, Azubuike H. Amadi, Emeka J. Okafor, Jumbo-Egwurugwu Precious (2022). Optimization of Methane and Natural Gas Liquid Recovery in a Reboiled Absorption Column *International Journal of Scientific Research and Engineering Development IJSRED* Volume 5 - Issue 2, March - April 2022. Page(s): 36-48.
  28. Wojciech, K., J and S. Janusz, (2012) "Real gas flow simulation in damaged distribution pipelines," *Energy*, vol. 45, pp. 481–488.
  29. Zukang Hu, Beiqing Chen, Wenlong Chen, Debao Tan and Dingtao Shen (2021) *Review of model-based and data-driven approaches for leak detection and location in water distribution systems*.