TREND ANALYSIS FOR STEADY STATE

The Analysis was performed for two features : **Pressure and Temperature (TB Discussed)**. The subject Pipeline is VVSPL, simulated with noise to match realistic scenarios. However, in order to perform better analysis with refined conclusions, EDA needs to be performed on Real-time Data. This report contains the plots, analysis, explanation and conclusions drawn from the EDA step.

**PART 1 – DATA DESCRIPTION**

The Data was taken from a Scada Data File, which was then converted to a csv. This data is that of the first quarter of the VVASPL pipeline which is similar to a A2B pipeline type. The data contains some structural errors due to simulation of multiple trends for same sensors. However, this error would not be seen again in the further analysis with more simulated data and would be corrected. The Data is induced with noise, to match real-time data and is simulated using flow as hook point with a flow rate of 1255 on both the sides.

The Data consists of Sensor readings of the pipeline over a time period of 43200 seconds or 12 hours of data from 2025-08-12 16:30:53 IST to 2025-08-13 04:30:53 IST. It has 67 columns (including redundant ones). Data consists of Pressure Sensors, Temperature Sensors, Flow Sensors, Valves etc.

The precision of the data is till 9 float points, which is highly sensitive to handle, thus multiple subplots are used instead of multiple readings in a single plot.

**PART 2 – PRESSURE TREND ANALYSIS**

To Understand How Pressure varies across time in a pipeline, I have First conducted a Univariate analysis on a single Pressure Reading over time.

**Subject Sensor : S1-PT-1001A**

**2.1 RAW Pressure over time**

Fig : Pressure Scatter Plot Over Time

A blue and white specks

AI-generated content may be incorrect.

THE STANDARD DEVIATION OF THE PRESSURE TREND :0.041007772108197935

THE MEAN OF THE PRESSURE TREND IS : 81.51618447421274

THE MIN, MAX DIFF OF THE PRESSURE TREND IS : 0.3051014857277039

**2.2 ROLLING / Moving Averages Pressure Over Time**

Rolling Pressure trends are plotted with params as mean and std respectively. Since the Data is highly scattered over a small range, We use Rolling trend (aka moving averages)

Fig : Rolling trend window size = 500

A graph of a stock market

AI-generated content may be incorrect.

Fig : Moving Averages over window size

A graph showing a line

AI-generated content may be incorrect. A graph with blue line

AI-generated content may be incorrect. A graph showing a line

AI-generated content may be incorrect. A blue line graph on a white background

AI-generated content may be incorrect.

From the above figures, we See that the Pressure across fluctuates heavily in a skewerish, unstable manner with no particularly defined upward or downward trend, However, In further smoothing till 7000, shows that after a point of 2000th -ish window, we see a near - steady decrease in the pressure.

This could mean that in real time scenarios, Pressure across the pipeline would decrease, provided the simulations follows real-time conditions closely.

NOTE : Extended study is required on real scenarios

**2.3 Rate of Change**

The Rate of Change is defined as d(param)/d(param2) or how one parameter changes wrt to the other. I have calculated the ROC wrt to time for every two readings for which I got a mean ROC over the trend of -0.000001049996

A blue sound wave with black text

AI-generated content may be incorrect.

Fig : Moving average of ROC

A blue line graph with black text

AI-generated content may be incorrect.

We can conclude that, ROC is not constant, nor does follow any linear trend across time. Thus we can safely conclude that, The Pressure Does not fluctuate linearly, rather in a random manner. Since this is a simulated scenario with pipeline filled with material continuously, we cannot say anything about the changes for sure.

**2.4 Batch Averaging of the Pressure Trend**

Unlike Rolling trend which is continuous, I have plotted Batch Averaging plots that calculate the mean/average for a batch of size n.

This helps us know how mean varies over time per batch. A steady decrease would imply that the pressure is decreasing over time, while a steady increase would imply an increase over time. However, a spiked, non-linear, skewed plot implies that the pressure is unstable and cannot be attributed a particular trend.

A graph of a graph

AI-generated content may be incorrect.

**From the image, we conclude that Pressure does NOT follow a particular trend.**

**Instead for short intervals values are oscillatory with no strict through or crest**

**2.5 BOX plot and Outliers**

Quartiles divide the dataset into four parts, P1, P2, P3, P4 by Q1, Q2, Q3. Using the statistical general of

IQR = Q3 – Q1 && Q1 – 1.5 X IQR < Data < Q3 +1.5 X IQR

We Boxplotted for the Pressure Trend :

A diagram of a box plot

AI-generated content may be incorrect.

Outliers: [81.40546812 81.40564569 81.36806019 81.40570386 81.63008462 81.38632185 81.40150643 81.64336793 81.38486509 81.3999952 81.63255334 81.63523603 81.65050446 81.38430722 81.62898079 81.664347 81.6423993 81.37575388 81.4051306 81.39685469 81.40640422 81.6383005 81.40628823 81.62663302 81.38411943 81.63255505 81.62844157 81.40216709 81.40484131 81.3718027 81.62770926 81.63486776 81.40056052 81.64374334 81.40323706 81.40334767 81.40084448 81.64390668 81.39015175 81.40569847 81.39355794 81.63235896 81.63961404 81.63994102 81.40548555 81.38919288 81.63808146 81.64212403 81.40096141 81.64427555 81.64071791 81.39204029 81.39797717 81.62721107 81.63561595 81.39276037 81.38638739 81.63091664 81.63275351 81.62824858 81.40058146 81.40256137 81.40155731 81.40644657 81.40369359 81.67316167 81.4016244 81.40425074 81.63453658 81.40216983]

A blue and red dots

AI-generated content may be incorrect.

**2.5 SPLINE and Smoothing of the Pressure trend**

Since the above analysis conducted points to the conclusion that the pressure trend is neither decreasing nor increasing in nature, We could try and interpolate to understand if any particular function could be fitted to it.

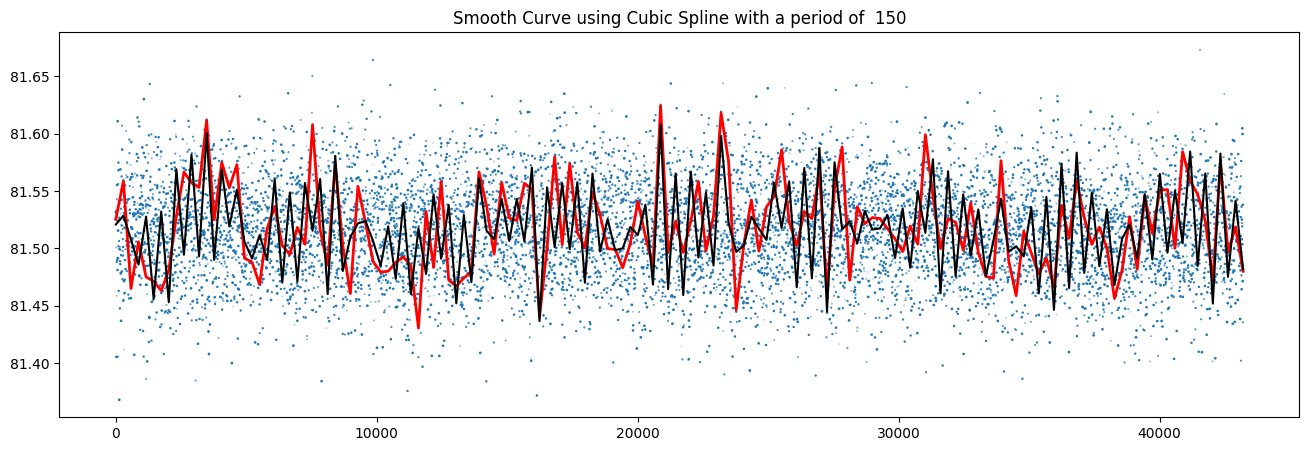
Interpolation is a mathematical method used to estimate **unknown values** that lie **between known data points**.

Spline is a piece-wise interpolation of all the points, A B spline with a degree of 3 was fitted with the data resulting in a graph like this

A red and blue graph

AI-generated content may be incorrect.

However, A spline has no generalization capabilities, hence it cannot be used to detect anomalies as it takes evry point and interpolates / passes through it.



However, I have tried to fit the trend using a custom function called checky\_sin (black in color for the plot) ,this is completely to interpolate which it does in a decent enough manner to conclude that a checky\_sin might generalize the data well, but this is not predictable function and just interpolates based on actual values.

For this, I modified Checky sin, which takes in sinusoidal wave and computes and predicts a new value based on it’s previous until t-1 readings parameters.

Where,

A → frequency factor (trainable)

B → amplitude scaling factor (trainable)

I have trained D2D value for this function and resulting is a an predictor that is parametric, trainable with and RMSE value of 0.070303 which is decent enough for an educated guess that the values are of oscillatory type for a short window.  
  
Even though, the loss is small enough, this is concluded to be not-so-good approach, as it just considers the whole data to be oscialltory as a whole, which defeats the purpose of a controlled thresholded noise induced in it. This generalization, doesn’t work with every sensor reading.

The checky\_sin function has failed, but has provided us with a great insight, i.e capturing temporal dependencies over a small number of previous readings could help to build a regeneration model . A predictive model such as a LSTM predictor or an Auto-Encoder has a scope to capture this while providing us with any anomalies, if they exist.  
  
This study enables us to think about a time series model, that uses a type of regenerative loss to calculate thresholds, and detect anomalies.

CONCLUSION :

The data is highly random and cannot be predicted using linear trend based models. Spline provided us with a highly interpolated model, but failed to generalize the data. Checky\_sin does it to some extent, but fails to understand any other type of oscillatory deviations. This further helps to understand and explore Correlation (inter-sensor type and intra-sensor type) based methods, time series and recurrent, memory-dependent models that shld successfully capture the temporal dependcies of the sensor data.