Motor Speed Overload Detection and Analysis Using PCA Models

This presentation covers the development of a data-driven model to monitor and predict overload conditions in an industrial equipment train. The goal is to detect overloads early to prevent equipment failure and minimize downtime, which can cause major production losses.

We will explore data preprocessing, the use of Principal Component Analysis (PCA) to understand data structure, and how the model identifies abnormal operating conditions effectively.

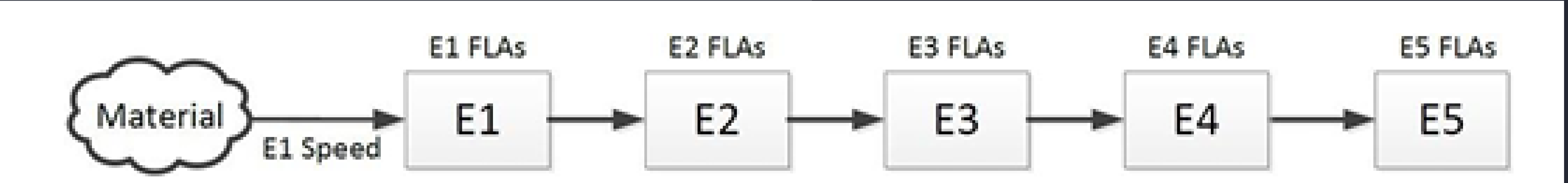
Objectives

- Develop a data-driven model to monitor and predict overload conditions in a production train consisting of equipment E1 to E5
- Create a calculated variable to indicate when the train is overloaded
- Analyze and clean the data to account for system downtimes and no-load conditions
- Use Principal Component Analysis (PCA) to understand the data structure and detect abnormal operating conditions

Early detection of overload states is essential

Because

- Equipment failure leads to total system shutdown
- Resumption of operations after failure can take up to many hours
- These shutdowns cause major production losses



Why Preprocessing is Essential

Filtering invalid entries:

Rows where 'E1 FLAs' values were less than 20 were removed as these often indicated system downtime or no-load conditions, which could distort the analysis

Handling categorical variables:

Categorical columns like 'bad inputs' were excluded after filtering as they predominantly contained zeros

Centering and scaling

All variables were standardized to have a mean of zero and standard deviation of one, ensuring all measurements contributed equally to the analysis regardless of their original scales

Removal of non-informative columns:

The 'time' column was excluded as it didn't contribute meaningful information for the PCA

Special Pre-treatment to the Data

K-Nearest Neighbors (KNN) imputation:

Used to handle missing or zero values resulting from the filtering steps, preserving the dataset's structure and relationships.

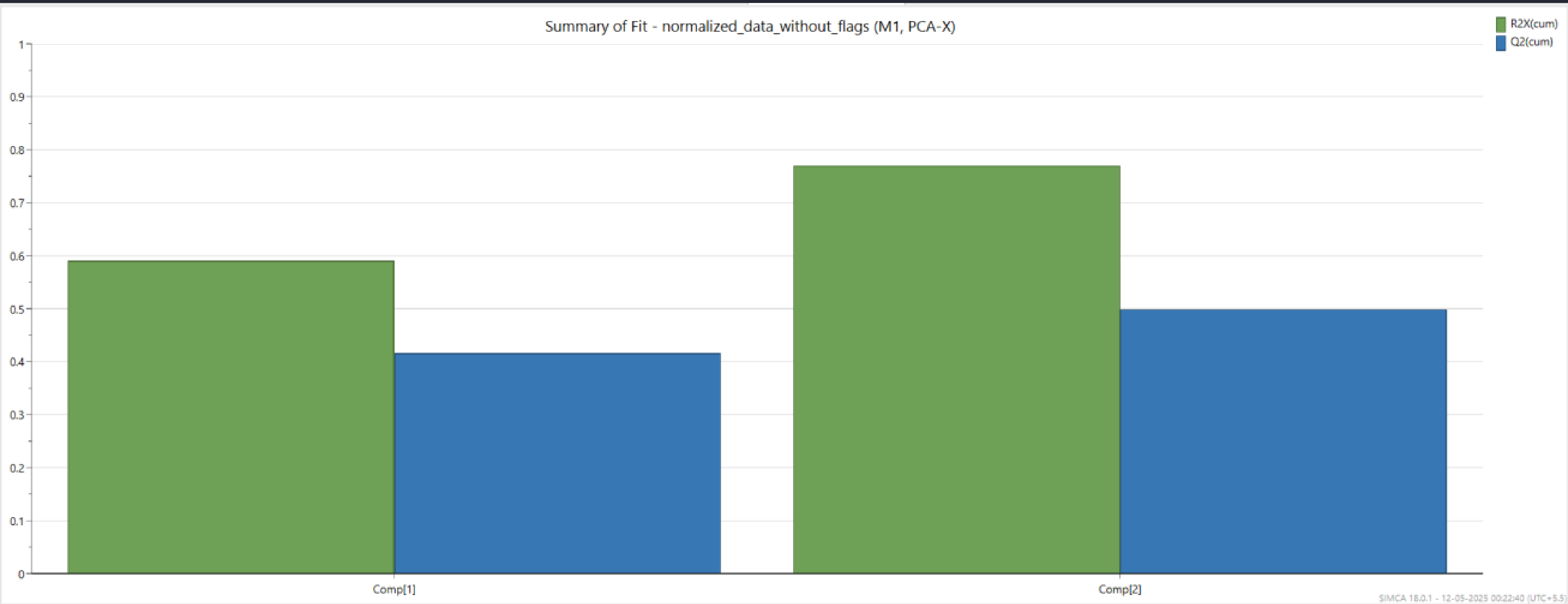
WHY?

Rather than simply removing incomplete data or using simplistic methods like mean imputation, KNN maintains the multivariate relationships that exist in the data

Why PCA is Appropriate for This Objective

- It successfully captured 77% of the total data variance with just two principal components (59% in PC1, 18% in PC2)
- It allowed visualization of relationships between variables through loadings plots, showing which variables contributed most to system variability
- It provided a way to detect outliers and abnormal conditions through Hotelling's T^2 and DModX plots
- It helped identify correlations between different equipment units (e.g., E4 FLAs and E5 FLAs were strongly correlated)

PCA-X Model Performance



1 Progressive Model Improvement:

The explanatory power (R^2X) consistently increases with each additional component, starting at ~0.6 for Comp[1] and reaching nearly 0.77 by Comp[2], showing systematic improvement in capturing data variance.

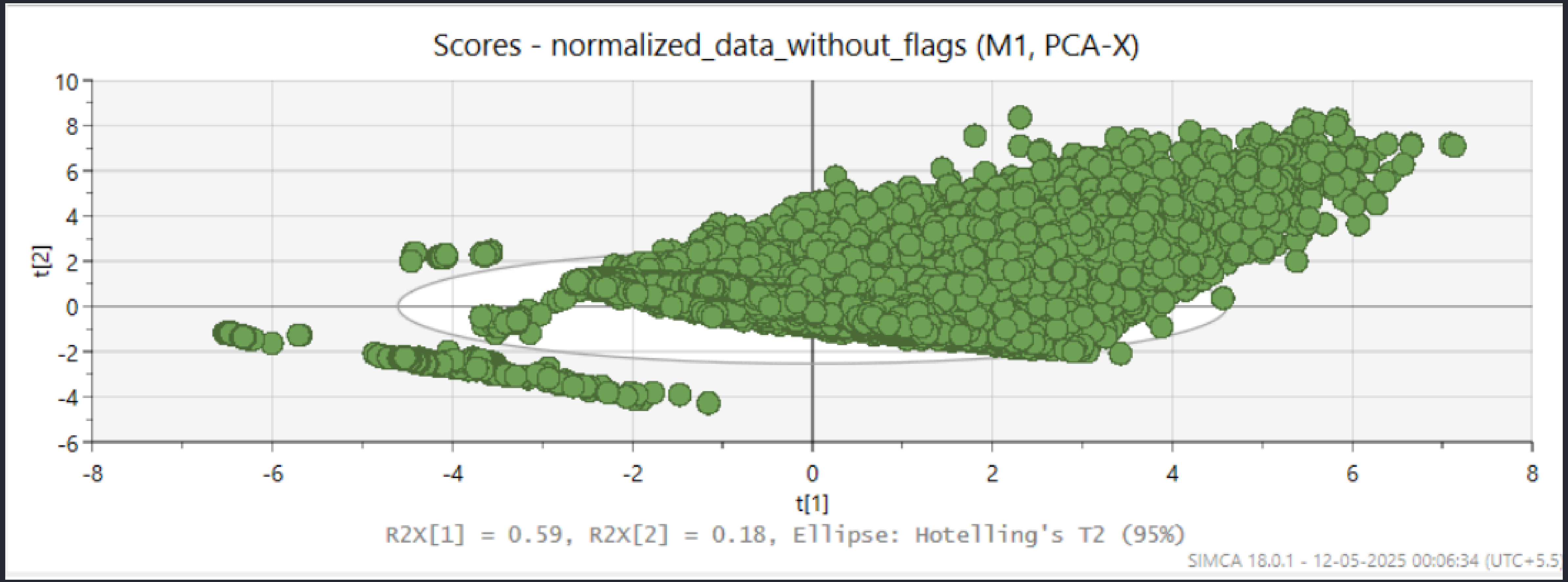
2 Predictive Power Enhancement:

The Q^2 values show a moderate increase between components, with component [2] providing only a limited improvement in predictive capability."

3 Excellent Final Model Quality

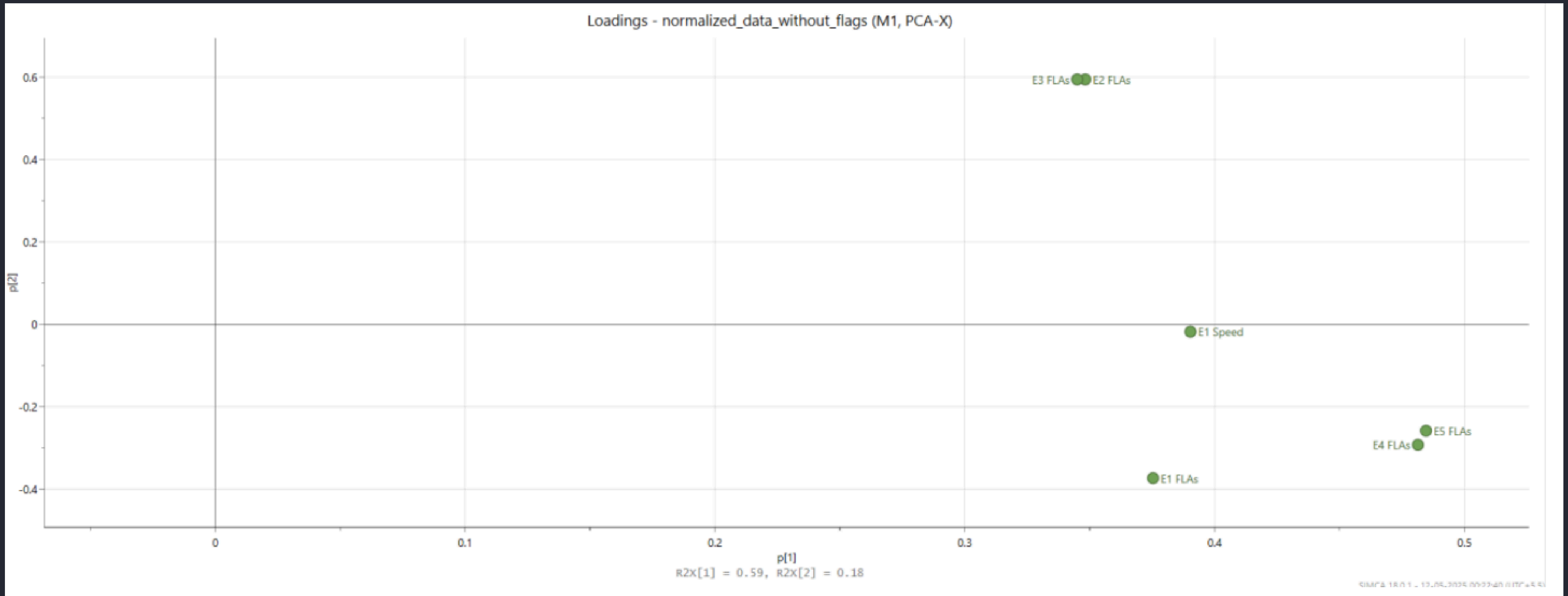
By Comp[2], both R^2X(cum) and Q^2(cum) show moderate values, indicating that the model explains a reasonable portion of the variance but does not yet demonstrate high predictive reliability

PCA Scores Plot:

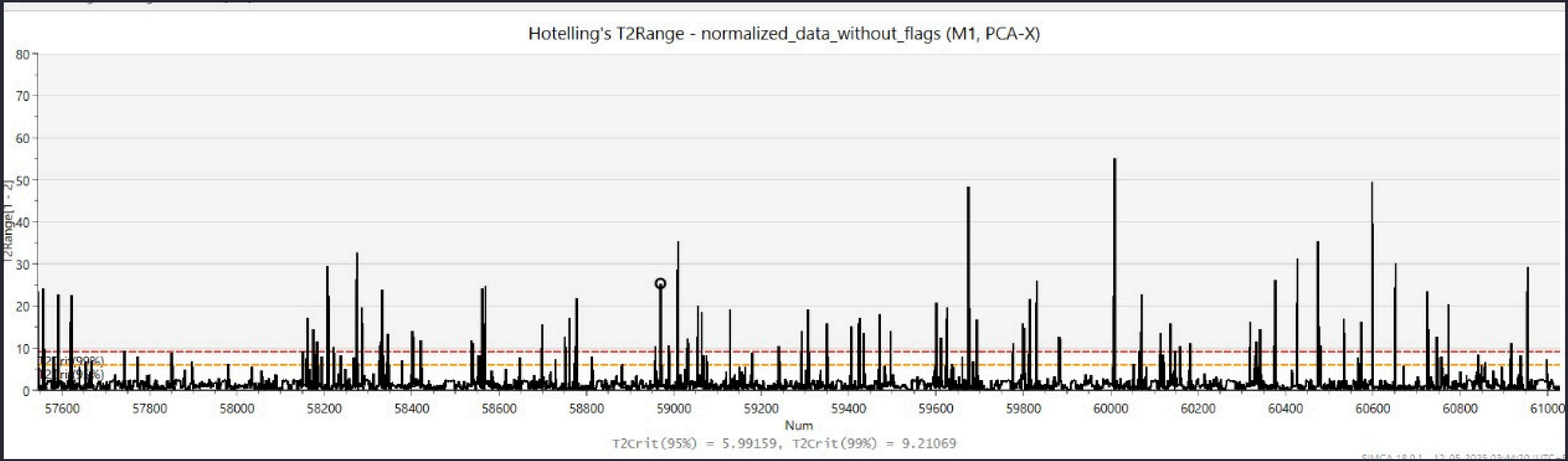


- Variance Explained: PC1 captures 59% and PC2 captures 18% of the data variability (total 77%).
- Data Pattern: Most samples cluster on the positive side of t[1], showing a dominant trend along PC1.
- Model Fit: Most data falls within the confidence ellipse, suggesting a good PCA model
- Cluster Shape: The wedge-like spread along PC1 suggests increasing sample variability in the positive direction.

PCA Loadings Plot



- Strong Influence Variables: Variables like E4 FLAs, E5 FLAs, and E1-FLAs have high positive loadings on PC1 (p[1]), meaning they strongly influence the first principal component.
- E2-FLAs Behavior: E2-FLAs is positioned high on the p[2] axis, meaning it influences the second principal component (PC2) more than PC1.
- Direction Matters: Variables grouped closely (like E4 FLAs and E5 FLAs) suggest they behave similarly across the samples.
- Variance Explained: Together, PC1 (59%) and PC2 (18%) explain 77% of the total variance, meaning the loadings shown capture most important patterns.



Hotelling's T² Range Analysis

Extreme Variability:

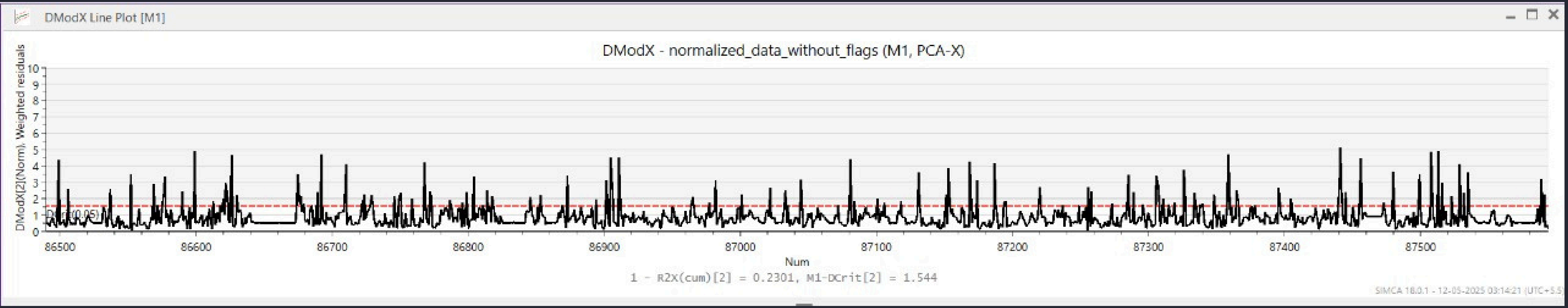
Most samples exceed the 95% and 99% T^2 limits, showing very high variability and multiple outliers.

Outliers and Causes:

Many extreme outliers are present, possibly due to measurement errors, process shifts, equipment issues, or changes in data patterns.

Out of the total samples, 12,380 points fall below the 95% Hotelling's T^2 critical limit, indicating that the vast majority of observations lie within the normal variability range of the PCA model.

DModX Line Plot



A total of 12287 observations are below the D-Crit(2) limit, indicating potential moderate outliers based on model residuals.

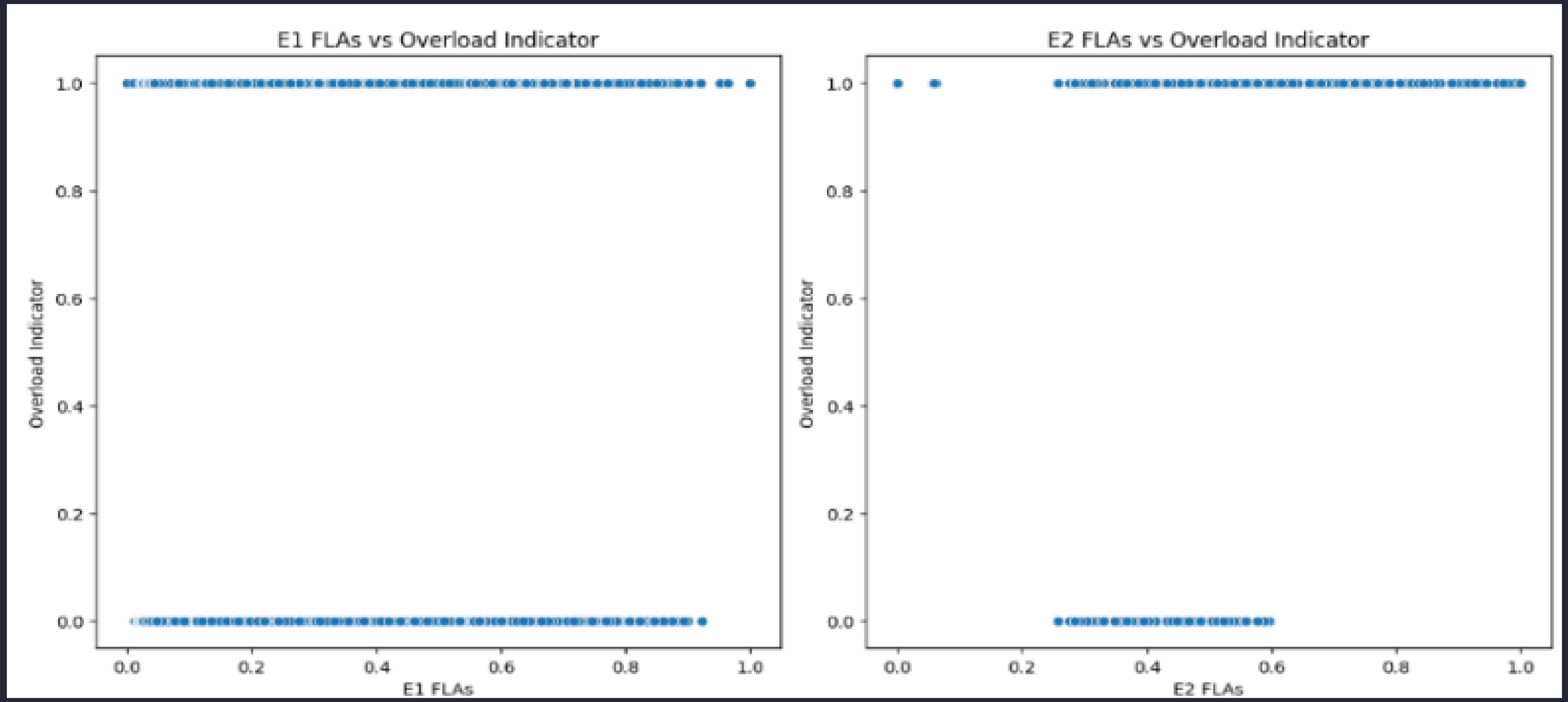
Normal Regions:

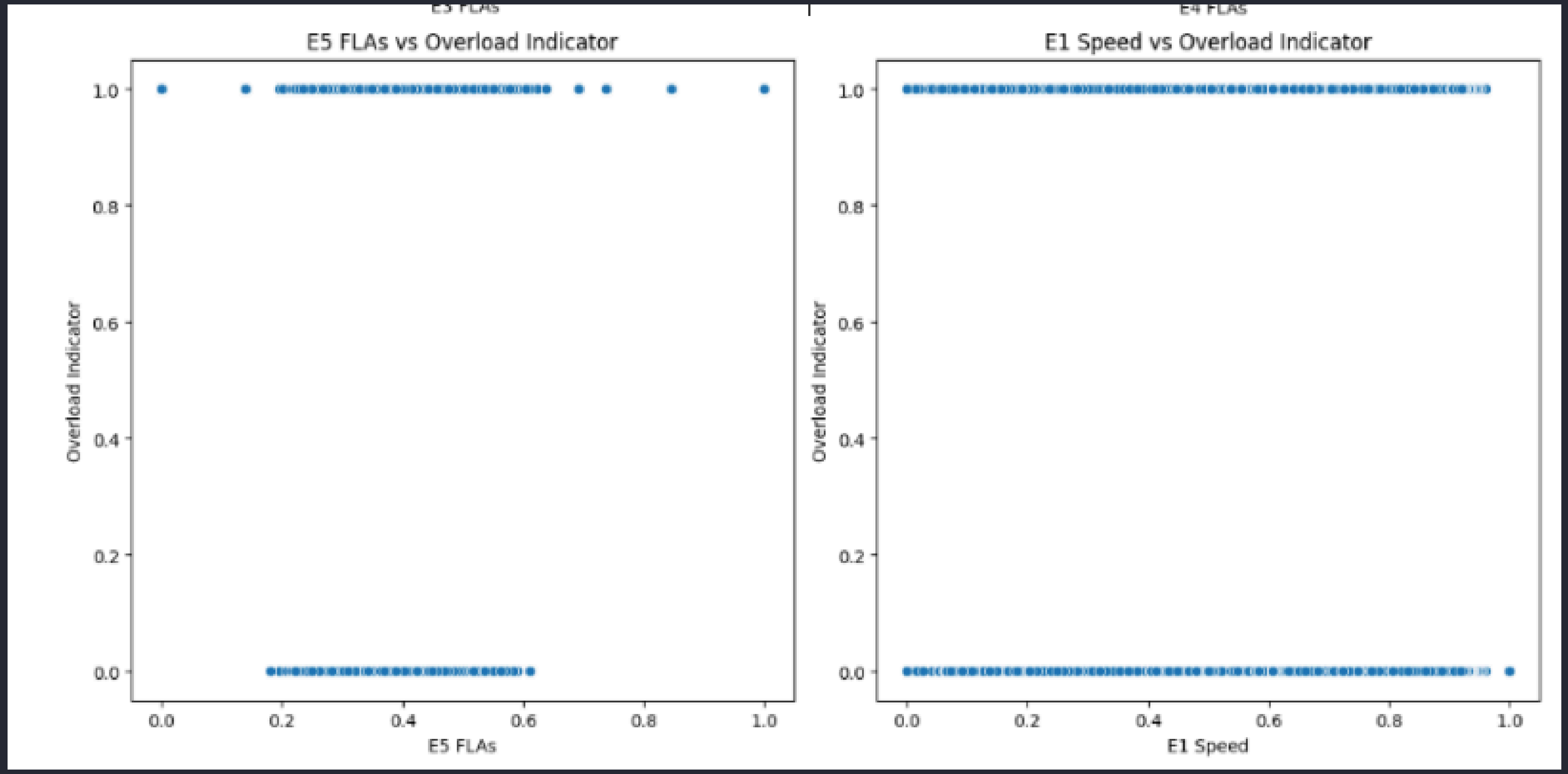
- Most samples are stable and lie below the critical limit, showing good overall process modeling.

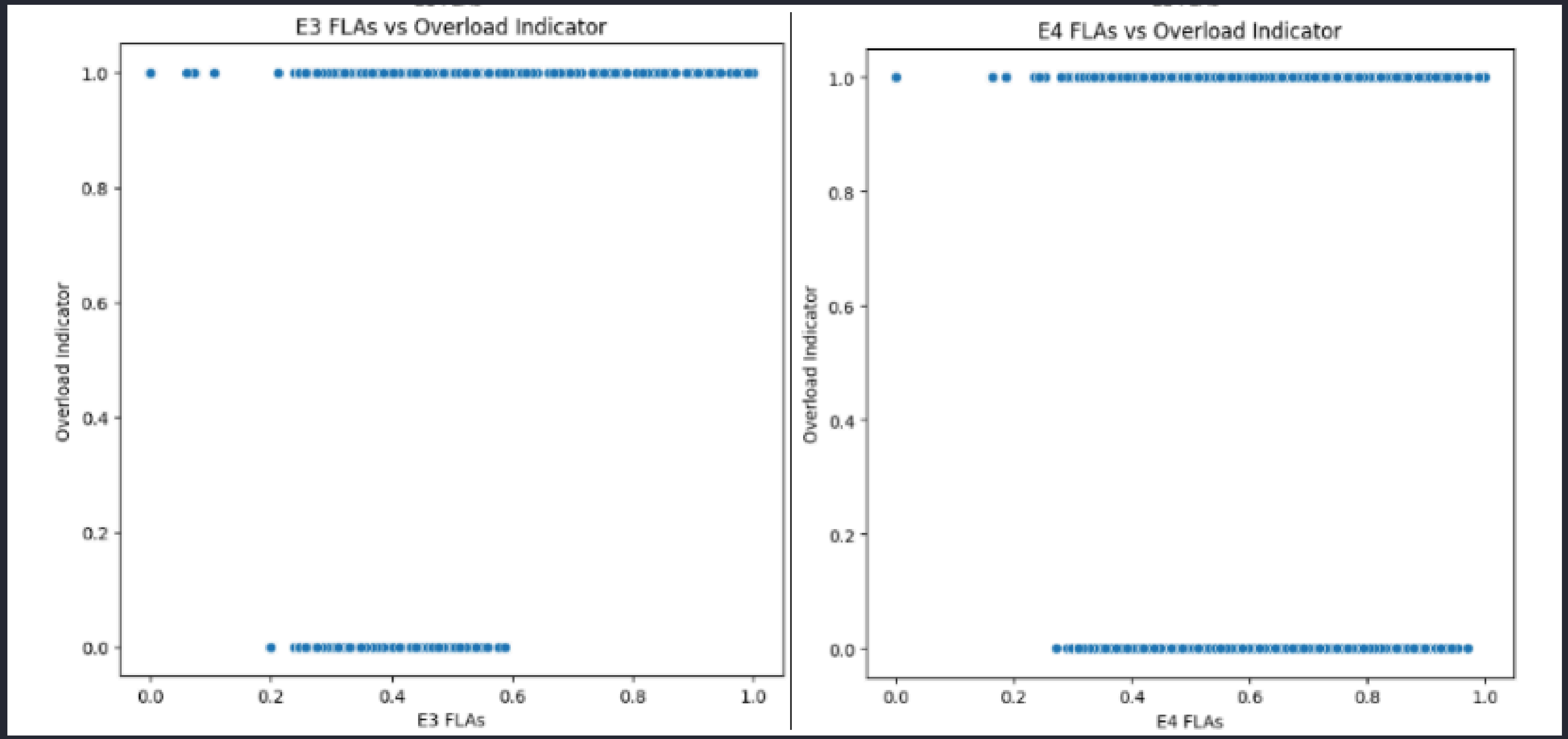
Outliers and Red Marks:

- Red points highlight extreme DModX outliers, needing further root cause analysis for rare or abnormal behavior.

Overload Indicator Correlation with Motor Parameters







Key Observations

Distinct Separation:

- In all plots, there is a clear separation between points where the Overload indicator is 0 and where it is 1, suggesting that FLAs and Speed are strongly associated with the overload condition.

Overload Correlation:

- For each entity (E1 to E5), higher FLA values are generally linked with an Overload indicator of 1, while lower FLA values correspond to an indicator of 0. This pattern is consistent across all five entities.

Speed vs. Overload:

- In the E1 Speed vs. Overload indicator plot, higher speeds are more often associated with overload (indicator = 1), whereas lower speeds relate to normal operation (indicator = 0).

Binary Nature of Target:

- The Overload indicator is strictly binary (0 or 1), with scatter plots showing two distinct horizontal lines, reinforcing the binary classification nature of the problem.

What Was Learned That Was New?

- Equipment correlations: E4 and E5 FLAs are closely grouped in the loadings plot, indicating they share similar behavior and are major drivers of the first principal component
- Distinct behavior: E2 FLAs shows a strong contribution to the second principal component, suggesting it captures variance different from the other variables
- Process instability: The Hotelling's T^2 Range Analysis shows that many observations exceed the 95% and 99% confidence limits, indicating significant process instability
- Clear binary separation: The scatter plots revealed a clear separation between normal operation and overload conditions based on FLA values and E1 Speed
- Predictive indicators: All FLA measurements and E1 Speed showed strong correlation with overload conditions, making them valuable predictors

Thank You