Fault Detection and Sensor Diagnostics Using Control Charts
Week 4: Modelling Plan

Github Repository: Wind Turbines

The goals of this week's task are as follows:
- Feedback incorporation from last week's task
- Document our modelling goal/purpose of our data analysis
- Document our model calibration strategy: tools used, methodology, data used
- Document our model validation strategy: scope of validation, methodology
- Document our model testing strategy: metrics and data used
- Describe the mathematical methods used in our model
- Develop a model diagram/ operations flowchart
- Outline the role of group members in modelling
- Detail document appearance: readability, figures and captures, clarity and completeness

Feedback Incorporation

We have re-centered and re-scaled the data of the faulty turbines using the means and standard deviations of the healthy turbine model. Essentially, this will project the data of the other turbines onto the healthy turbine model and allow for a more accurate comparison and analysis.

Modelling Goal/Purpose of Data Analysis

The primary goal of this modelling project is to develop a fault detection system for wind turbines using control charts and sensor diagnostics. The project entails the following key objectives:
- Model Calibration: Begin by calibrating a robust model for a healthy turbine, optimising the selection of principal components for feature representation.
- Control Chart Development: Create control charts based on multivariate statistical techniques, (T^2 and SPEx control charts), utilising the data from the healthy turbine model. These control charts will serve as the foundation for fault detection.
- Fault Identification: Apply the developed control charts to the data. By investigating the deviations from normal operation, identify and pinpoint the specific faults in the faulty turbines.
- Sensor Contribution Analysis: Analyse the contributions of individual sensors to out-of-control observations. Identify the sensors that effectively capture the faults, providing valuable insights into sensor performance and fault detection capabilities.

Model Calibration Strategy

Wind Turbines are systems susceptible to a range of malfunctions and harsh variable conditions. This can affect their performance as well as their safety. It is crucial to develop methods that are effective in detecting their potential faults in order to reduce the risk that can be caused by their malfunction.

1. **Principal Component Analysis**

We aim to develop a fault detection system for wind turbines using control charts and sensor diagnostics. The first step is to calibrate a robust model for a healthy turbine using Principal Component Analysis(**PCA)** for representing features. PCA will be used for reducing the dimensionality of data while preserving important information. The calibration process will involve the following steps:

- Data Pre-treatment: The data has been normalised and we have already dealt with missing values
- PCA Model Fitting: The PCA model will be fitted to the data from a healthy turbine using a repetitive process of selecting the optimal number of Principal componets. The selection will be based on the percentage of explained variance and the residual analysis.

2. **Data Used:**

   We shall use data from the healthy turbine to calibrate our model.

Model Validation Strategy

The next step is to create control charts based on multivariate statistical techniques, specifically **T2** and **SPEx** control charts. These charts will be used to monitor the turbine performance and detect deviations from normal turbine operation. We shall use them as follows:

- **T2 Control Chart:** It is a control chart based on Hotelling's T2 statistic that measures the Mahalanobis distance of a multivariate observation from the mean of the healthy turbine data. It helps in identifying observations that are significantly defferent from the center of the multivariate distribution.
- **SPEx Control Chart:** It is a control Chart based on the Squared Prediction Error(**SPEx)** which measures the difference between the observed and the predicted values of a multivariate observation. It will be used to detect faults that may not be captured by T2 Chart.

Both Control charts have upper control limits that are computed using statistical formulas and thresholds. The development of the control charts involves:

- **Control Limit Estimation**: The control limits will be estimated using data from healthy turbine model, based on the selected multivariate technics. The

control limits will be adjusted to account for the variability and the correlation of the data.
- **Control chart construction:** The control charts will be constructed using the control limits and the data from the turbines, both healthy and faulty. The control charts will be plotted to visualise the turbine performance and the deviations from normal operation.

These control charts will serve as the foundation for fault detection.

Model Testing Strategy

1. **Metrics**

The primary metric for testing the model's performance is its ability to detect deviations from healthy turbine. The result included identifying and pinpointing specific faults in the faulty turbines. The fault detection will involve applying developed control charts to the data and investigating the deviations from normal operation. This involves evaluating using control charts, observations that exceed control limits which indicates potential faults. The results will show that control charts can detect various types of sensor faults such as sensor bias and drift in different measurements. Additionally, other metrics will be used such as false positive rate to assess the performance of the fault detection system.

2. **Sensor Contribution Analysis**

The sensor analysis will involve analysis of the contributions of individual sensors out of control observations. The analysis will be used to identify the sensors that effectively capture faults, providing insights into sensor performance and fault detection capabilities. The sensor contribution analysis will be done using both **T2** and **SPEx** statistics for the out-of-control observations. The contribution plots will be normalized to have a maximum value of 1. The resulting plot will be interpreted as follows: Sensors with highest contributions to **T2** and **SPEx** are the ones that are affected by the faults.

3. **Data Used**

The testing phase involve applying control charts to both healthy and faulty turbine data. The faulty data having measurments during abnormal conditions or faults in turbines. The plots generated will be analysed to identify any deviations from normal operation and potential fault conditions.

Mathematical Methods Used in Model

The modeling for faulty detection in wind turbines involves principal component analysis (PCA) and control charts based on multivariate statistics (**T2** and **SPEx** control charts). PCA will be used to reduce the dimensionality of sensor data while preserving the most important information. It involves the following steps:

- **Standardization:** First, standardize the data to ensures that all variables have the same scale. By taking the data matrix **X**, the mean vector of all variables **μ,** and standard deviation of σ. The formula for standardization:

$$Z = (X - \mu) / \sigma$$

- **Covariance Matrix:** The second step involves calculating the covariance matrix **Σ** of the dataset **X(nxp)**. Where **n** is the number of observations and **p wis the number of variables.** The covariance matrix characterizes the relationship between variables.
- **Eigenvectors and Eigenvalues:** Next we shall compute eigenvectors(**v1, v2, ….vp)** as well as their corresponding eigenvalues(**λ1, λ2, …., λp).** The eigenvectors represent the directions of maximum variance in the data.
- **Principal Components:** Lastly, we shall calculate the Principal components(PC_i**)** which can be calculated as a linear combination of the original variables and the eigenvectors. They provide a new coordinate system that captures the most important information.

$$PC_i = X.v_i$$

Then, control charts are used to monitor devotions from the expected behavior of a healthy turbine using **T2** and **SPEx**. Hotelling's **T2** is used to observe deviations of from the mean in the data based on the first A PCs. The formula for T2 statistic:

$$T_A^2 = \sum_{i=1}^{A} (t_i^2 / S_{t_i}^2)$$

where $S_{t_i}^2$ *is estimated variance of* t_i^2*.*

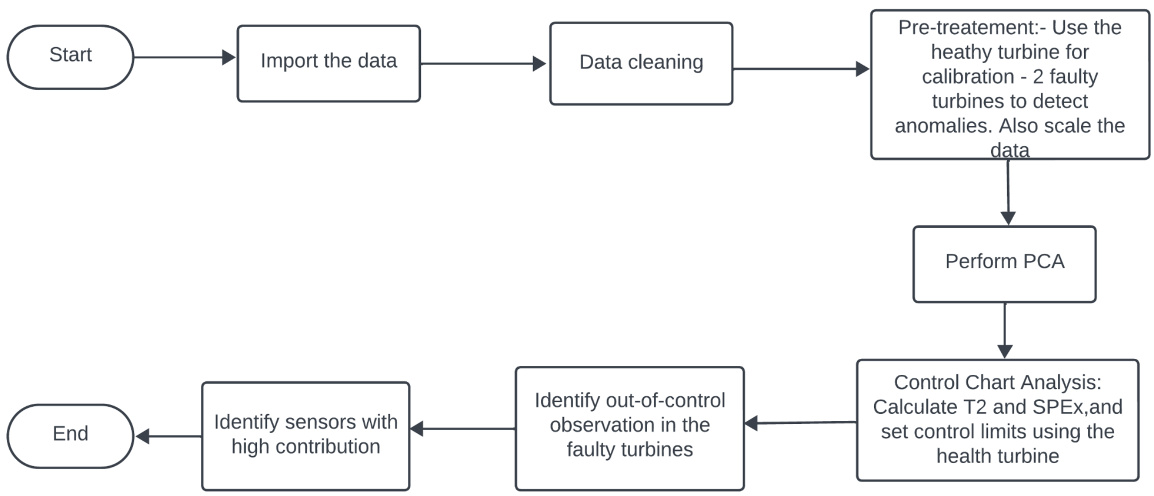
The **SPEx** quantifies the prediction error between observed data and predicted from the PCA model. The formula is given as:

$$SPEx = \sum_{i=1}^{k} (x_i - \widehat{x_i})^2$$

where *k* is number of variables.

By following these steps and formulas, we can build a model for faulty detection in wind turbines and pinpoint which sensors are sensitive to capturing anomalies in the system.

Model Diagram/ Operations Flowchart



Role of Group Members in Modelling

The following division of roles ensures that each member plays a critical part in various aspects of the project while promoting a collaborative and cohesive approach to achieving the project's goals. Whilst the project tasks will mostly be handled jointly, each group member will take lead in different sections of the project. This is outlined as follows:

- Socrates Waka - Model Calibration
- Nadine Cyizere - Control Chart Development and Implementation
- Aime Barema - Fault Identification and Sensor Contribution

In addition, there are general responsibilities for all group members and these are outlined below:

- Collaborate regularly through meetings and communication channels to share progress and insights.
- Contribute to the overall project design, methodology, and documentation.
- Engage in thorough testing and validation of the fault detection system on the selected turbines.
- Assist in the preparation of project reports, presentations, and documentation for the doctoral research.

Detail Document Appearance

The presentation and documentation of our work on this project are critical steps in conveying our results. Effort has already been made in this regard with weekly reports of our work being submitted to the joint Github repository. In addition, to enhance readability and comprehension, we will adhere to the following principles:

Document Structure:

We will largely adhere to the modelling flowchart to structure our document, ensuring a cohesive and organised presentation.

Font and Formatting:
We will use a clear, consistent and legible font style and size for any documentation related to our work, ensuring uniformity in formatting.

Figures and Captions:
Figures, charts and tables will be used to illustrate key findings. Each figure will include a concise and informative caption that depicts its relevance.

Clarity and Completeness:
We will strive to ensure that our writing is clear and jargon-free. Moreover, the writing will follow a logical flow, with each section transitioning smoothly to the next, aiding in comprehension. Lastly, we will strive for completeness by addressing all relevant aspects of the modelling procedures.