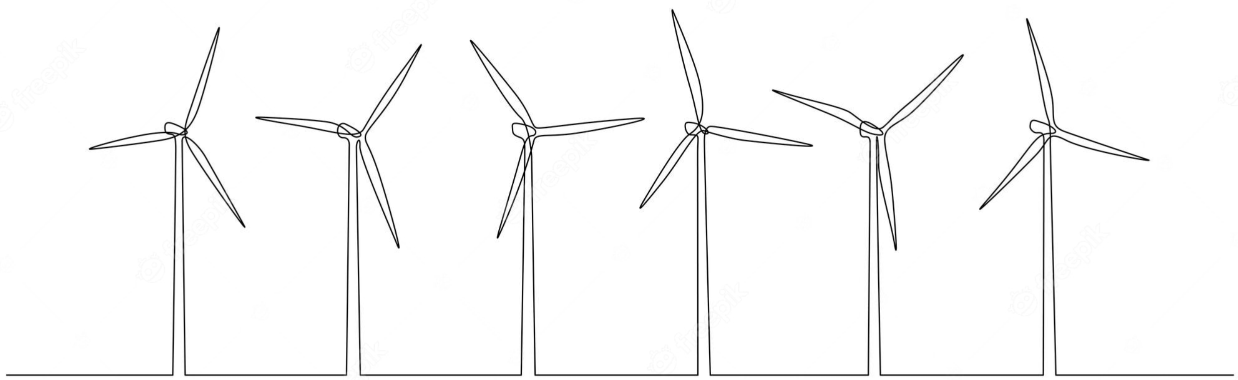

Wind Turbines Failure

Contents:

---

1. Dataset description

The wind turbine failure dataset (SCADA – Supervisory Control and Data Acquisition) has measurements recorded every 10^{th} second, from wind turbine sensors. The variable names are unknown, but they represent sensor measurements that correspond to various physical process properties. The dataset contains four wind turbines: WT2, WT3, WT14 and WT39. The WT2 is the healthy-functioning turbine, whereas the other three present faults (they are developing faults). What we know about the faulty wind turbines is that there is an initial fault, and then they switch at some point to normal operation after the fault has been resolved.

2. Modelling goals

a. **Level B (20p)** Fault detection using biplots.

Faults can be detected using variation captured in the model, if the healthy state of the process is known. PCA biplots are a tool that can distinguish between the healthy and faulty states of a wind turbine. Use various biplots of the healthy model with faulty data

projected into it to identify the faulty observations. Make use of the loadings to interpret how are the values varying in different variables, when the turbine gets the fault.
Note: *Since this is a level B task, you can choose one healthy turbine and one faulty turbine to perform the analysis on.*

b. Level B2 (20p) Fault detection using PLS-DA.

A multivariate regression model coupled with discriminant analysis can distinguish between a healthy and an unhealthy state, if there is prior information on the fault. Create a PLS-DA model in which the healthy turbine represents one class, and the first 20 observations of the faulty turbine represents the second class. If you are predicting the rest of the observations of the faulty turbine, can you evaluate at which point the turbine comes back to normal operation?
Note: *Since this is a level B task, you can choose one healthy turbine and one faulty turbine to perform the analysis on.*

c. Level A (30p) Fault detection using control charts and sensor diagnostics.

Control charts made on the healthy-turbine model can capture the faults of the faulty turbines. Optimize the number of principal components in the healthy model and create control charts based on multivariate statistics (T2 and SPEx control charts). By analyzing the contributions to individual out-of-control observations, identify the sensors that are able to capture the faults.
Note*: Since this is an A task, you need to consider one healthy turbine and two faulty turbines in your analysis.*

d. Level A2 (30p) Sensor importance and multivariate models for filling missing data.

Calibrate a PCA model of the healthy turbine, and separate PCA models for the first 20 observations of the two chosen faulty turbines. Which are the sensors that give most different loadings when compared to the healthy turbine? For the healthy turbine, create PLS models to estimate the values of two important sensors. How well does your model predict? Use, as a test data, one of the faulty turbines. Are you still able to maintain your regression performance in case the turbine has a fault?
Note: *Since this is an A-level task, you need to consider one healthy turbine and two faulty turbines for your analysis.*

e. **Level A3 (30p)** PCA variants and fault identification

Calibrate a PCA model of the healthy turbine, and a robust-PCA model of the healty turbine. What are the differences between the two models in model loadings? Create control charts using PCA and robust-PCA. What are the differences between the two methods? Optimize the number of principal components in the healthy model and create control charts based on multivariate statistics (T2 and SPEx control charts).
Note: *Since this is an A-level task, you need to consider one healthy turbine and two faulty turbines for your analysis.*

a. **Level A4 (30p)** PCA variants and fault identification (2)

Calibrate a PCA model of the healthy turbine, and a k-PCA model of the healty turbine. Create control charts using PCA and robust-PCA. What are the differences between the two methods? Optimize the number of principal components in the healthy model and create

control charts based on multivariate statistics (T2 and SPEx control charts). Identify variable contributions to the residuals.
Note: *Since this is an A-level task, you need to consider one healthy turbine and two faulty turbines for your analysis.*

3. Dataset hints

- One wind turbine (faulty) needs to be dropped from the analysis, because it has an inconsistent number of variables. Since the variables are not guaranteed to be in order, there is no direct correlation to the healthy turbine variables.
- From one of the remaining turbines, one variable needs to be dropped. (The last variable), which represents an extra quality not measured for the other wind turbines.

4. Useful references

[1] - Wen, X., & Xu, Z. (2021). Wind turbine fault diagnosis based on ReliefF-PCA and DNN. Expert Systems With Applications, 178, 115016. doi: 10.1016/j.eswa.2021.115016
[2] - Pozo, F., Vidal, Y. and Salgado, Ó., 2018. Wind Turbine Condition Monitoring Strategy through Multiway PCA and Multivariate Inference. Energies, 11(4), p.749.
[3] – Fu, Y., Liu, Y. and Gao, Z., 2019. Fault Classification in Wind Turbines Using Principal Component Analysis Technique. 2019 IEEE 17th International Conference on Industrial Informatics (INDIN).