

Paris Session 2022



Predicting the Belgian system imbalance better with AI

Study committee D2

PS2: Artificial intelligence

Question 1.1: What experience on improving the performance of machine learning-based systems does exist in terms of addressing the anomalies (rare event with significant consequences), which may pose considerable impact on technological and/or economic performance of the power system? How should we distinct anomalies and data outliers between each other?

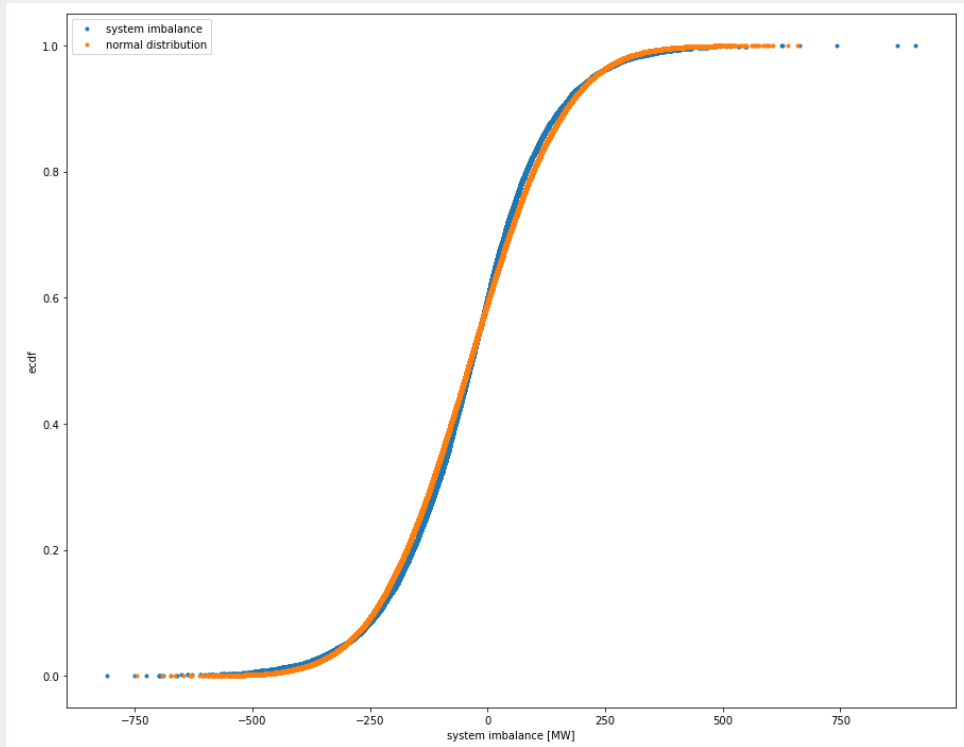
Thomas Craessaerts
Belgium



The Belgian system imbalance

- Difference between production and consumption
- Predicting it better will reduce imbalance costs
- Simple models seem to prefer very well: logistic and linear regression
- But how do we distinguish between outliers and anomalies in the data?

Outlier vs anomaly



The system imbalance is almost normally distributed except for the tails that have too much extremes.

In the plot to the left you can see the empirical cumulative density function of the imbalance in blue compared to a similar normal distribution in orange.

Dropping the extreme points show that the imbalance is normally distributed.

But how do we know which points are outliers from the distribution and which ones are anomalies?

Outlier vs anomaly

- Analysis of the data show the most extreme values are from big outages
 - Losing a big powerplant
 - Big substations tripping
 - Major incidents
- Those events should be excluded from train data. They are clear anomalies and not related to outliers in the normal distribution.