

Reproducibility Report for Updated Results: “Wind Turbine Gearbox Failure Detection Through Cumulative Sum of Multivariate Time Series Data”

1. Updated result

On the published version, we presented our detection result based on an **offline** approach. What it means is that the scoring (i.e. to compute the anomaly score) for all data points was performed in one go (after the test data are available to users). When defining whether a point is an anomaly or not, the algorithm check its k nearest neighbors among **all** points in the entire sets (training + test sets). Furthermore, there is no treatment difference between the training and test data when computing their scores. The role of training data was to set the parameters namely the offset, the accumulation window, and the control limit (H).

This report is to provide the reproducibility of the result using an **online** approach. With the algorithmic procedure remains the same, the first step, which is to compute the anomaly score, is performed differently. We start by computing the anomaly score for the training data first. For the test data, we update the anomaly score by adding one new data point at a time. For example, to compute the anomaly score for data point n in the test set, we use all data points in the training set plus the first $n - 1$ data points in the test set.

Using this online approach, the results in Figure 4, Table 2, and Table 3 will change slightly. Table 2 and Table 3 are given below, but how to reproduce Figure 4 should use the code given in section 3.

Table 2. Gearbox failure detection results summary

| Failure ID | Turbine ID | Failure time | Alarm time | Alarm-to-failure |
|------------|------------|---------------------|---------------------|------------------|
| 1 | T01 | 2016-07-18 02:10:00 | 2016-07-04 02:00:00 | 14 days |
| 2 | T09 | 2016-10-11 08:06:00 | 2016-08-21 01:00:00 | 51 days |
| 3 | T06 | 2017-10-17 08:38:00 | 2017-08-06 06:00:00 | 72 days |
| 4 | T09 | 2017-10-17 08:38:00 | 2017-08-07 00:00:00 | 72 days |

Table 3. Calculated savings from detection using common control limit

| | #TP | #FP | #FN | Calculated saving |
|-----------------------|-----|-----|-----|-------------------|
| Training set | 2 | 9 | 0 | € 41,666.67 |
| Test set | 2 | 4 | 0 | € 140,000.00 |
| Training and test set | 4 | 13 | 0 | € 181,666.67 |

2. Downloading and creating the dataset files

We updated the description on how to obtain the data. The data we used was made available for an open challenge in 2021 called “EDP Wind Turbine Failure Detection Challenge” so there is two ways to retrieve the data from the www.opendata.edp.com website. Either way, user needs to register (free) to get an access.

The data are granted a free use by CC-BY-SA license. **Note that our reproducibility package does not include the datasets**, as we do not own the data. But the instructions below describe precisely where to

download and how to create the dataset files that will be the same as we used for producing the results in the paper. There are two ways to obtain dataset #1 and #2. We use the first way when performing our analysis.

| Dataset | File names | Description |
|---------|----------------------------------|--|
| #1 | wind-farm-1-signals-training.csv | <p>Two ways to obtain the data:</p> <ol style="list-style-type: none"> 1. Download the wind-farm-1-signals-training.csv from this link. 2. Combine data from two files with the following steps: <ol style="list-style-type: none"> (a) Download the wind-farm-1-signals-2016.csv and wind-farm-1-signals-2017.csv from this link. (b) Combine the data from the entire wind-farm-1-signals-2016.csv and that from Jan 1 through August 31 of the wind-farm-1-signals-2017.csv, and save the new data file with the name “wind-farm-1-signals-training.csv”. <p>This file is the recorded SCADA data on wind turbine T01, T06, T07, T09, T11 from January 1, 2016 to August 31, 2017, and will be used as the training set.</p> |
| #2 | wind-farm-1-signals-testing.csv | <p>Two ways to obtain the data:</p> <ol style="list-style-type: none"> 1. Download the wind-farm-1-signals-testing.csv from this link. 2. Download the wind-farm-1-signals-2017.csv from this link. Take the data from Sept 1 through Dec 31 of the wind-farm-1-signals-2017.csv, and save the new data file with the name “wind-farm-1-signals-testing.csv” <p>This file is the recorded SCADA data on wind turbine T01, T06, T07, T09, T11 from September 1, 2017 to December 31, 2017, and will be used as the testing set.</p> |
| #3 | htw-failures-2016.csv | Download the htw-failures-2016.csv directly from EDP Open Data website . This is the wind turbine failures recorded in the year of 2016. |
| #4 | htw-failures-2017.csv | Download the htw-failures-2017.csv directly from EDP Open Data website . This is the wind turbine failures recorded in the year of 2017. |

3. Reproducing the detection results

| Code File | What does it do? | Required Data or Data File | Output |
|----------------------|---|---------------------------------|--|
| LoMSTOnline.R | Provides function to perform LoMST online. It generates the scores for training data first, and then generate new score for each data point added. | n/a | n/a |
| CUSUM_LoMST_Online.R | <ul style="list-style-type: none"> • Arrange the data in hour average and obtain anomaly scores using LoMSTOnline.R • Filters the anomaly scores based on the offset. • Clusters the filtered anomaly scores based on the accumulation window. | LoMSTOnline.R Dataset #1 – 2 | Anomaly scores in <i>LoMST_Gbx.csv</i> and cumulative scores in: <i>cusum_T01.csv</i> <i>cusum_T06.csv</i> <i>cusum_T07.csv</i> |

| Code File | What does it do? | Required Data or Data File | Output |
|-----------|--|---|--|
| | <ul style="list-style-type: none"> Obtains cumulative scores and save them in .csv files. | | <i>cusum_T09.csv</i> <i>cusum_T11.csv</i> |
| Figure4.R | Plots the results of the method on the five wind turbines. | <i>cusum_T01.csv</i> <i>cusum_T06.csv</i> <i>cusum_T07.csv</i> <i>cusum_T09.csv</i> <i>cusum_T11.csv</i> <i>Dataset #3 – 4</i> | Figure 4_Online in .pdf |