

# Change point detection of particulate matter data.

Degree programme : BSc in Industrial Engineering and Management Science | Specialisation : Business Engineering  
Thesis advisors : Patrik Marti, Prof. Dr. Angela Meyer

Change points indicate when a system moves from one state to a different state. Being able to detect those movements can save money, time as well as lives. The problem this Bachelor thesis tries to solve is to write a python code who automatically detects change points for hourly univariant PM10 time series data sets.

## Introduction

Univariant timeseries, also known as one dimensional timeseries, can be found in all industries. Well known examples are the depiction of share prices at the stock market, the depiction of the heart frequency in hospital monitors or the depiction of temperature development predictions for the next days in the weather forecast. The main characteristic of univariant time series is that only one variable, which is dependent of the time, is listed within the time series. Understanding the data within a univariant time series dataset can have a significant impact on the decision-making process of businesses and individuals. One important aspect which is to be understood are the changepoints of the univariant time series. Change points divide a time series into multiple segments, each segment has its statistical characteristics. The change points are located at the beginning and end of each segment. There are different forms of change points, such as the change in the mean, the change in variance, the change in periodicity and the change in patterns. The impact changepoints can have on the decision-making processes of businesses and individuals can be seen by the decision to buying or selling of shares in a certain stock, the decision to administrate a certain medicine to a patient or the decision to cancel an excursion due to a decline in the temperature. Therefore, the problem this bachelor thesis solves is to write a python code which automatically detects change points in the univariant time series data sets.

## Concept

To achieve the goal of a python code who automatically detects change points in the univariant time series data sets, the project follows the data science life cycle.

- Problem understanding
- Data gathering
- Data cleaning
- Data exploration
- Feature engineering
- Predictive modeling

## Results

The final python code contains 5 million data points from 37 air quality measuring stations all over Switzerland. Those data points are transformed into new features with the help of such methods as the slope, the kurtosis, the skewness, the interquartile range, the power spectral density, the change in the mean value and the change in the variance value. For the final model a random forest classifier is used to predict the change points.



Nico Balz Jakob

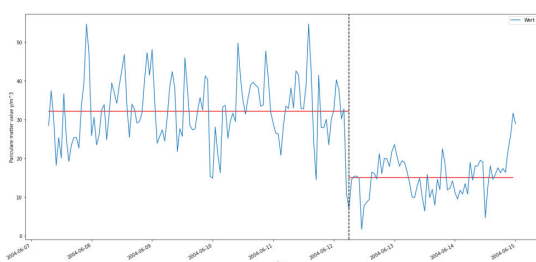


Figure 1: Visualization of a mean change point in particulate matter data.