

Predictive Maintenance using MATLAB: Pattern Matching for Time Series Data

Dr. Irina Ostapenko (Dr. Türck Ingenieurbüro GmbH),
Jessica Fisch (Daimler AG)

18.04.2018



Dr. Türck Ingenieurbüro

Zwischen den Zahlen lesen
www.tuerck-optik.de



Mercedes-Benz

Das Beste oder nichts.
www.mercedes-benz.com

Collaboration partners



Mercedes-Benz

Jessica Fisch

Mercedes-Benz Werk Mettingen

Digitale Fabrik Powertrain und Projekt Industrie 4.0

PT/TSD

010 - HPC M427

70546 Stuttgart

jessica.fisch@daimler.com

DR. TÜRCK  **DATA SCIENCE**
INGENIEURBÜRO

The logo for Dr. Türck Data Science features a stylized 'D' shape composed of a blue semi-circle on the left and a yellow semi-circle on the right.

Irina Ostapenko

Senior Data Scientist

Solutions and Algorithms

Kreuzbergstr. 37 Berlin 10965

io@tuerck-optik.de

Collaboration partners



Mercedes-Benz

Jessica Fisch

Focus:

- Digital Transformation
- Big Data
- IIoT

DR. TÜRCK  **DATA SCIENCE**
INGENIEURBÜRO

Irina Ostapenko

io@tuerck-optik.de

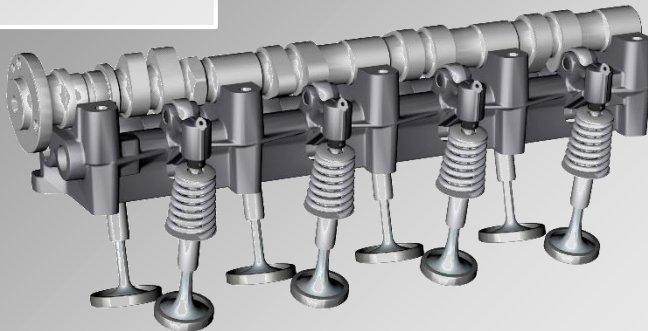
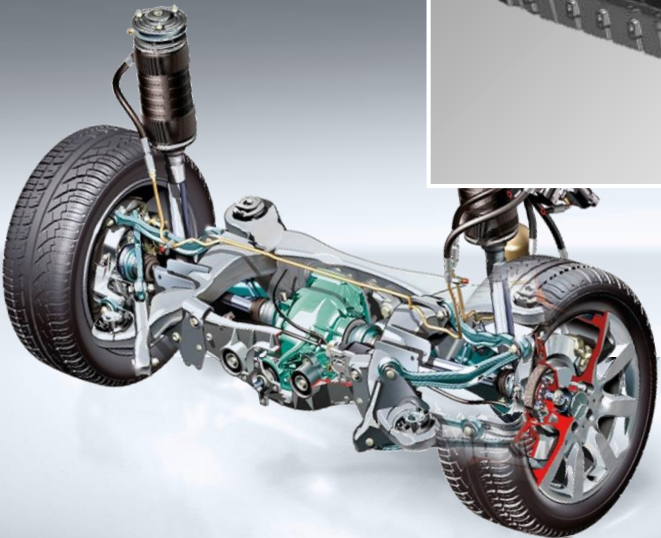
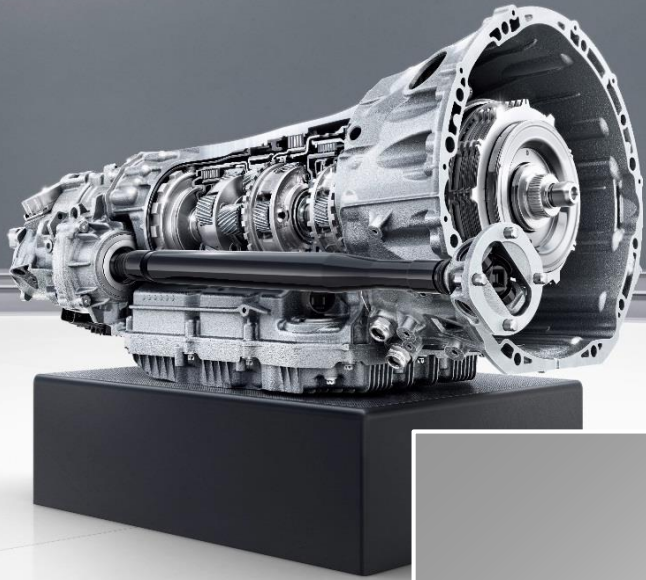
We provide

- Algorithms
- Signal Processing
- Measurement Systems Developing

Optical System Design

Outline

1. Project introduction
2. Task description
3. Solution/Algorithm
4. Summary



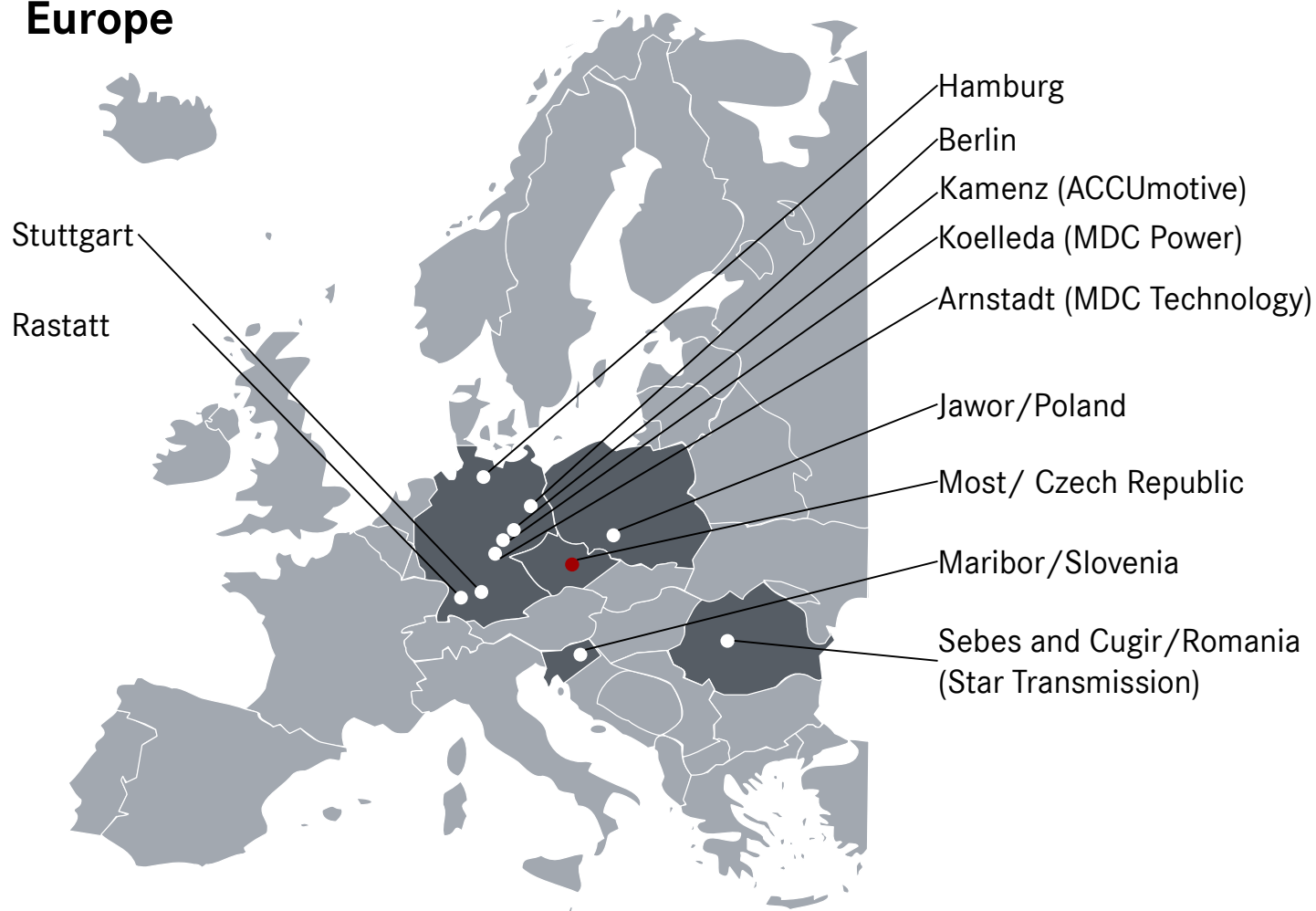
P POWERTRAIN

Five modules form
the core of our cars

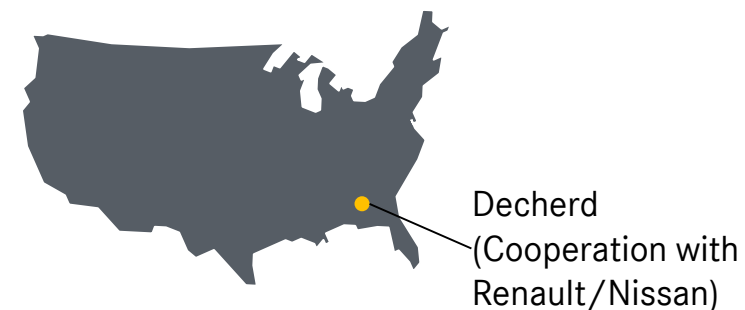


The Powertrain production network is set up globally with lead plant in Germany

Europe



USA



China



Legend

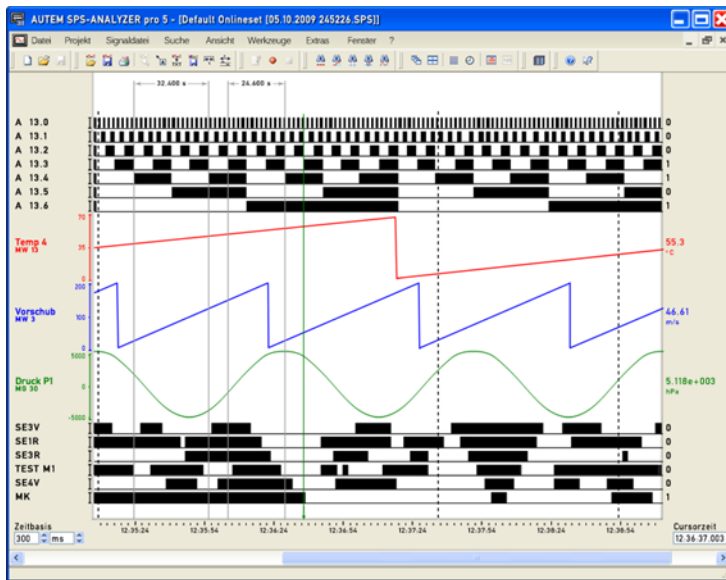
- 100% Daimler
- Majority Holding
- Joint Venture
- Cooperation

Motivation for Anomaly Detection in the Projekt „iLL“



The goal is to detect anomalies in data

PLC-Data today:

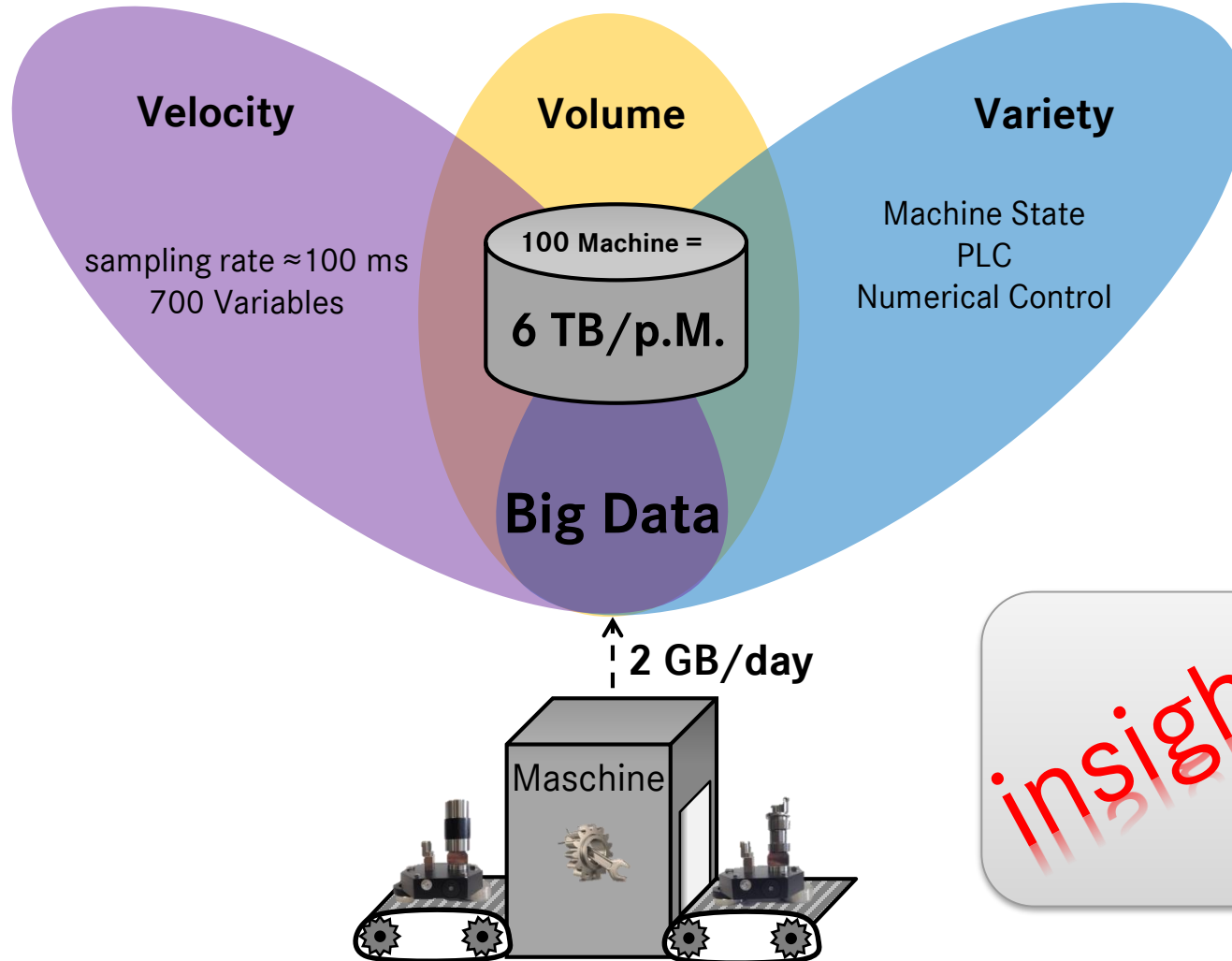


Source: www.autem.de

Automatic Notification of the Deviation:



Data properties in the context of Big Data



The 3 basic V's of Big Data:

- **Velocity:** Speed with which data is generated and analyzed
- **Volume:** Amount of data that traditionally can not be analyzed
- **Variety:** Data diversity refers to unstructured data without a recognizable context

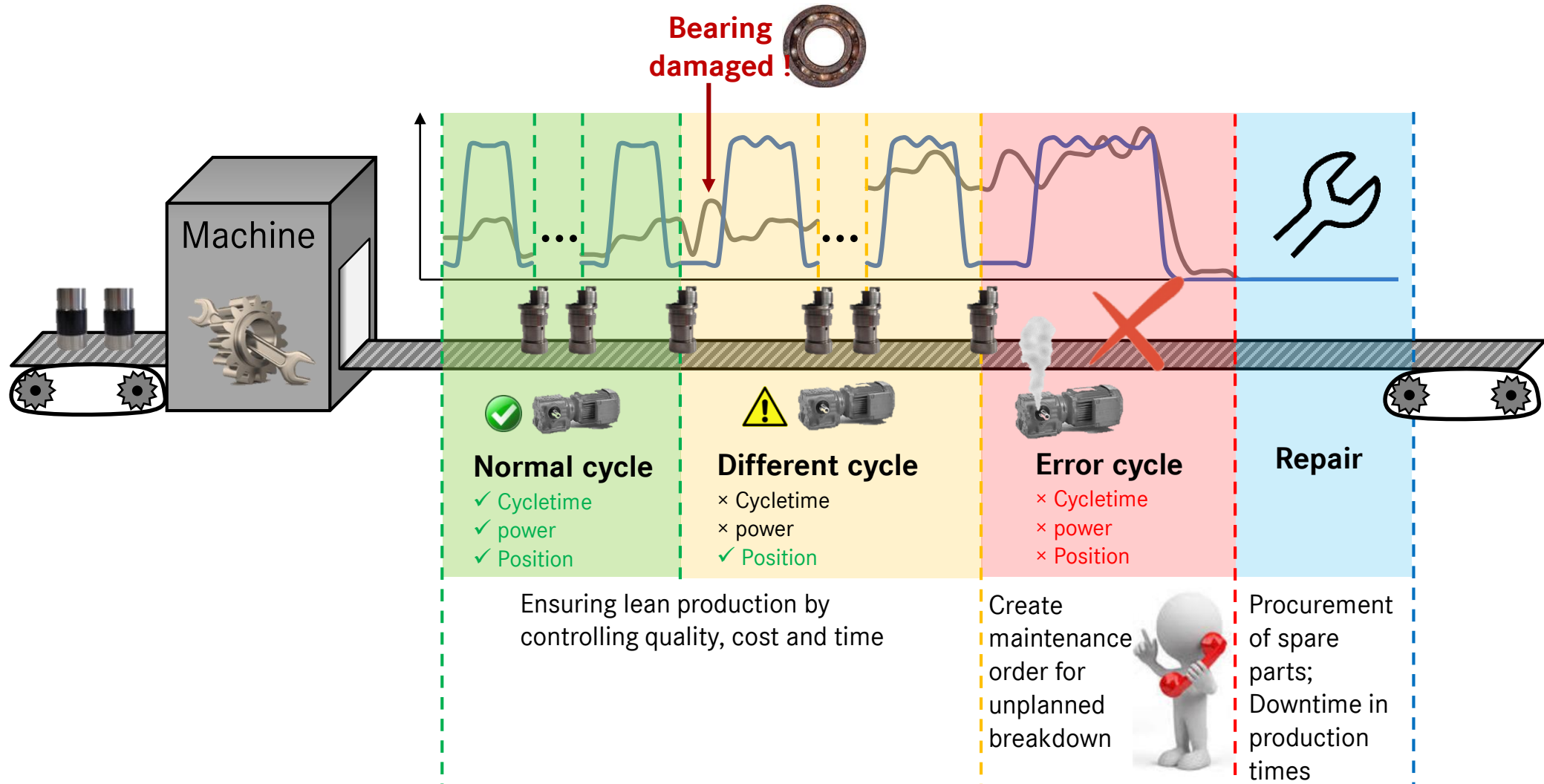
The 2 additional V's:

- **Validity:** Ensuring data quality
- **Value:** measurable benefits from the data

insight

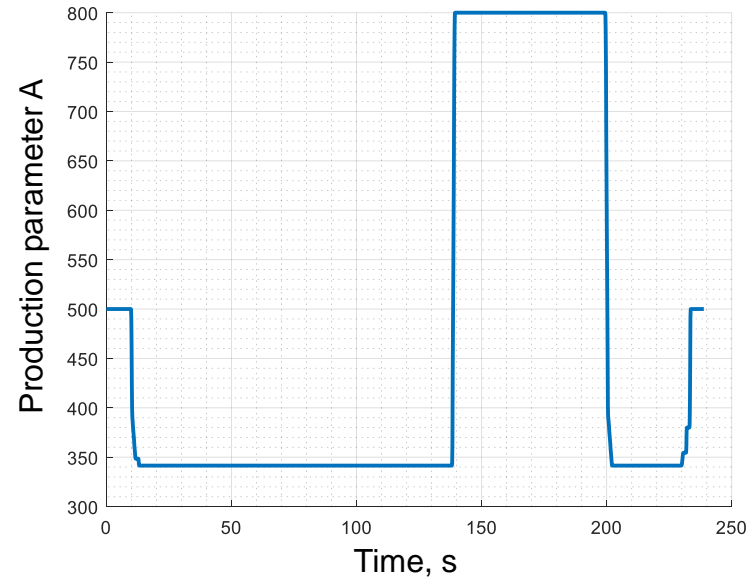
Benefits of „Intelligent Level-Learning“

— Active power engine axis 1
 — Position axis 1



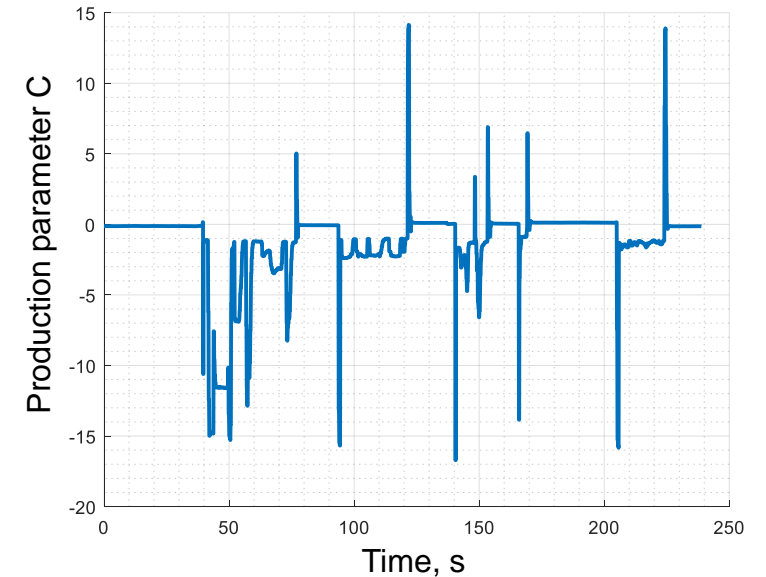
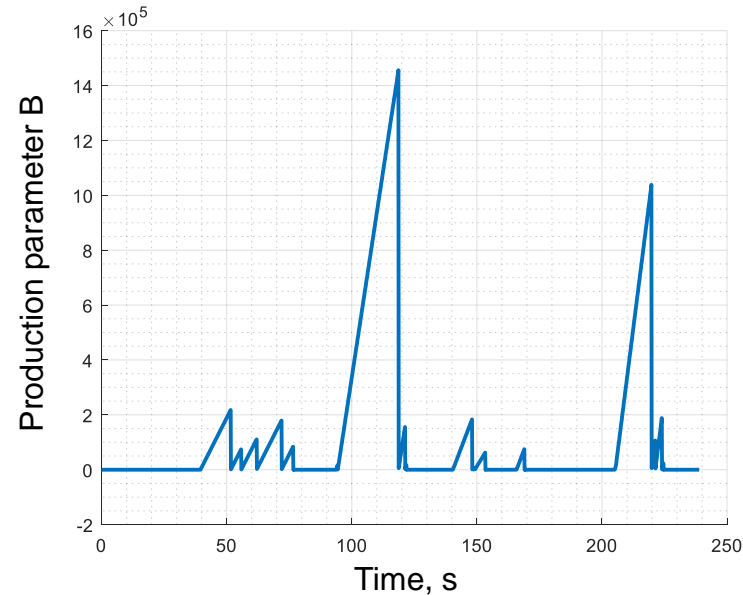
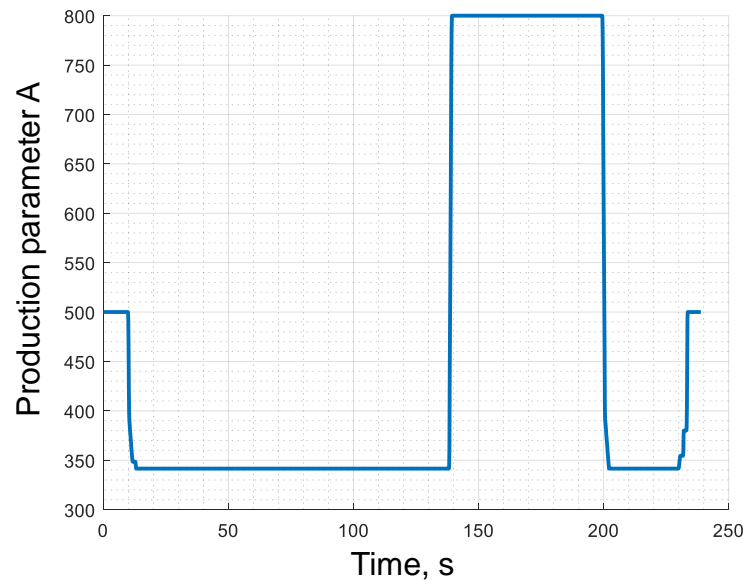
Challenges

- About 700 parameters are continuously monitored in every production cycle yielding 700 individual time-series of about 2500 samples each



Challenges

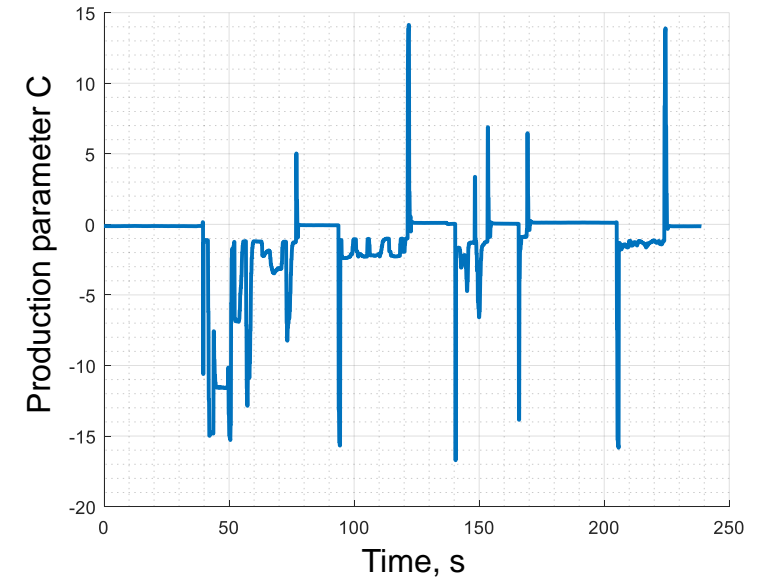
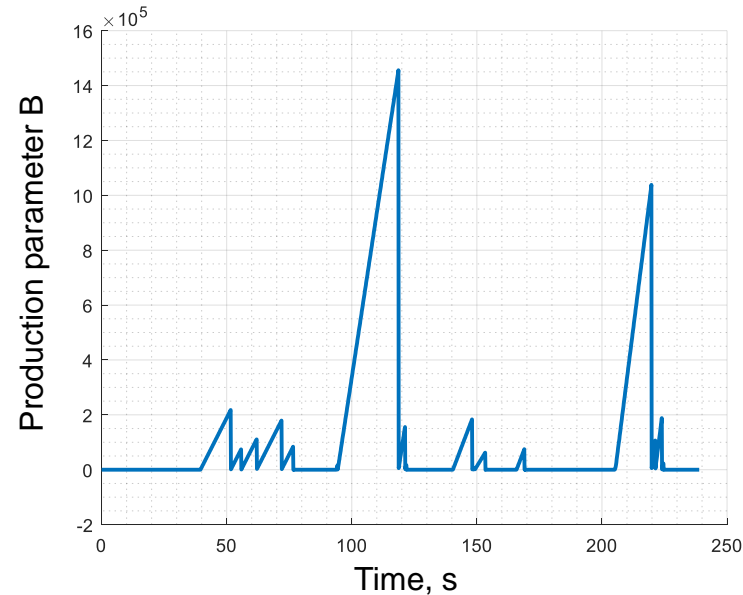
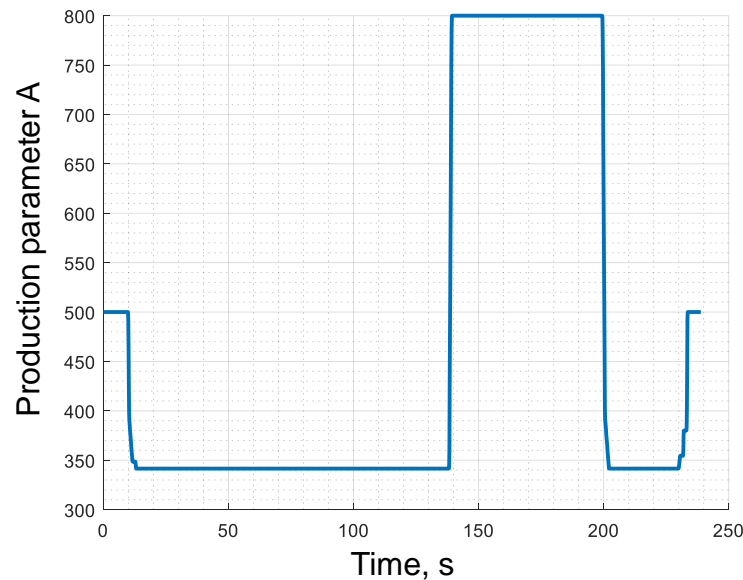
- About 700 parameters are continuously monitored in every production cycle yielding 700 individual time-series of about 2500 samples each



- Different parameters show very different and elaborate features

Challenges

- About 700 parameters are continuously monitored in every production cycle yielding 700 individual time-series of about 2500 samples each

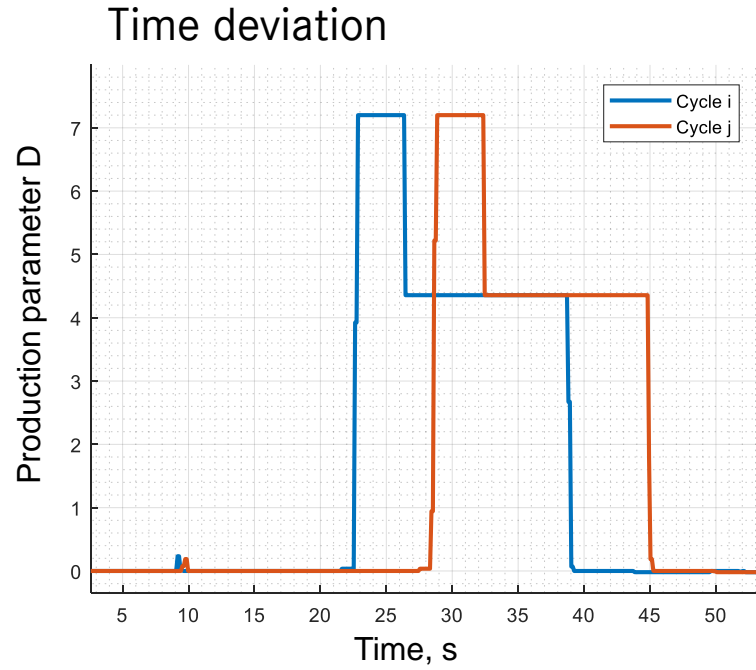


- Different parameters show very different and elaborate features



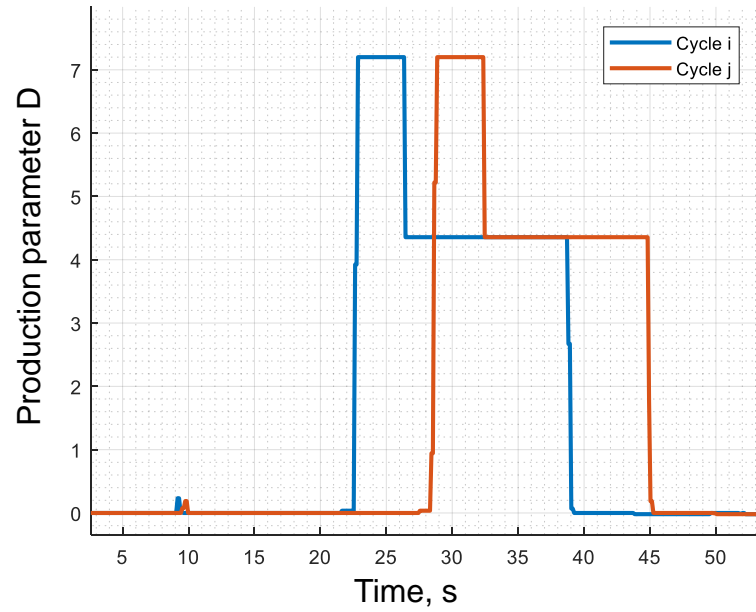
Task: Analyse these 700 time-series and find specific kinds of deviations

Requirements for algorithm

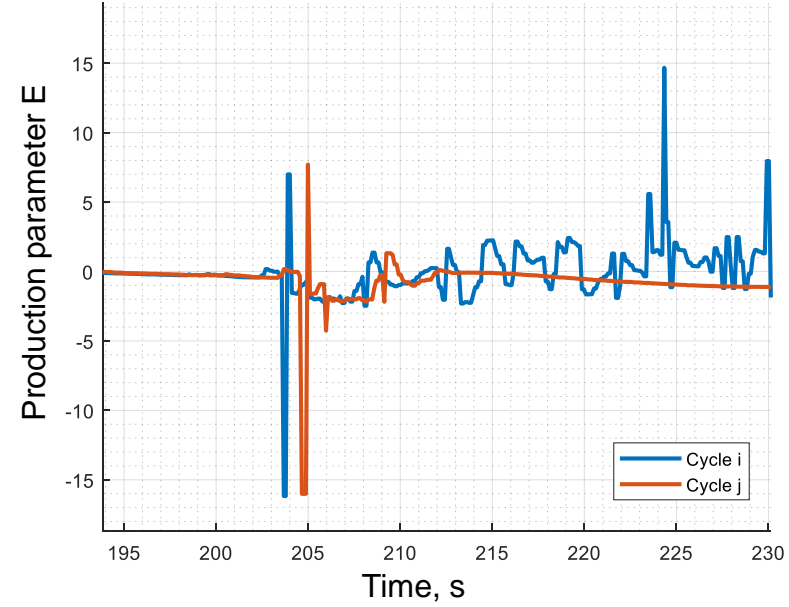


Requirements for algorithm

Time deviation

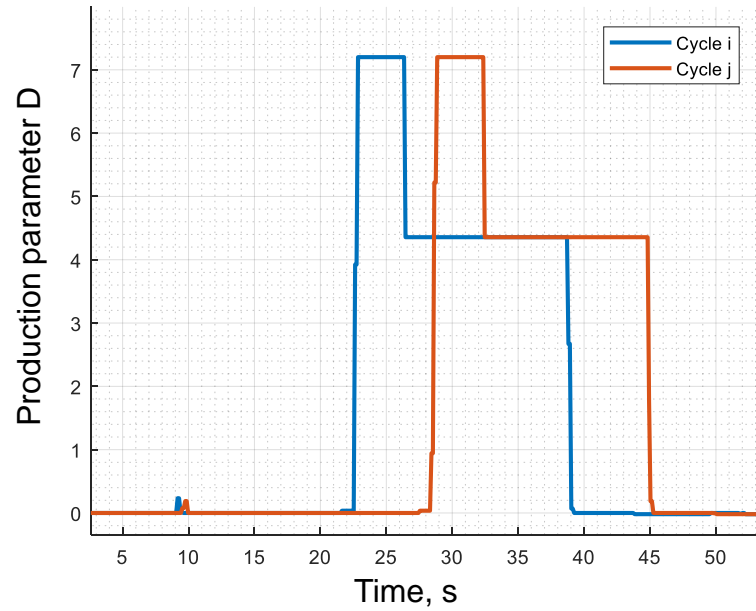


Pattern deviation

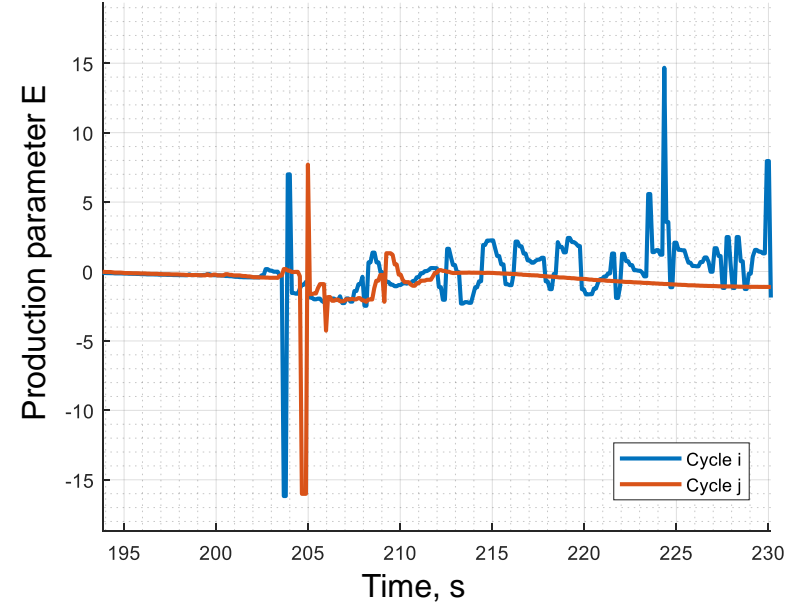


Requirements for algorithm

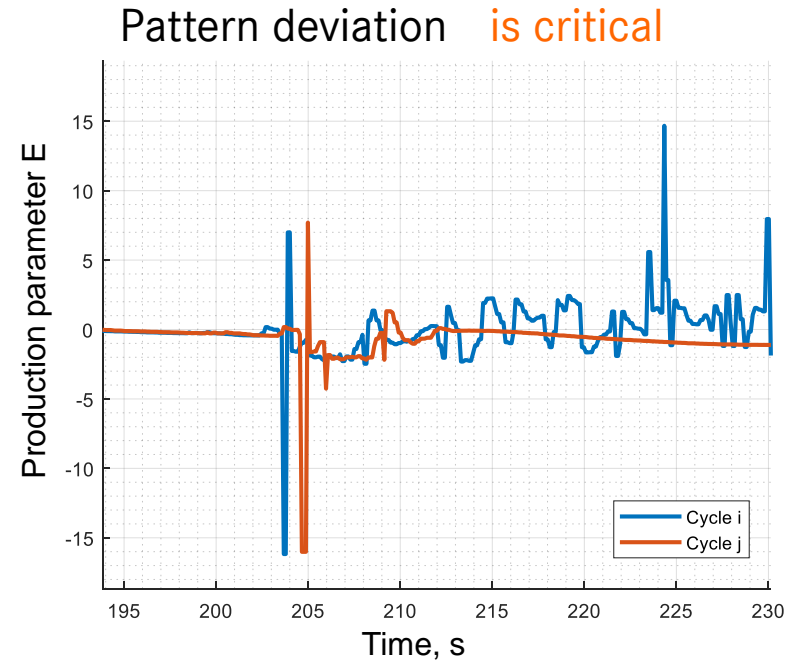
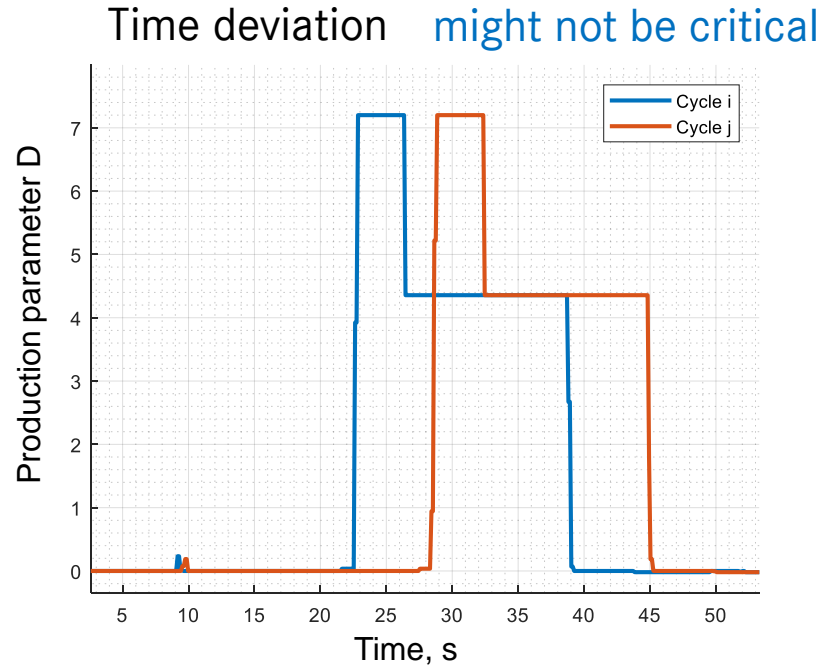
Time deviation might not be critical



Pattern deviation is critical



Requirements for algorithm



What the algorithm should do

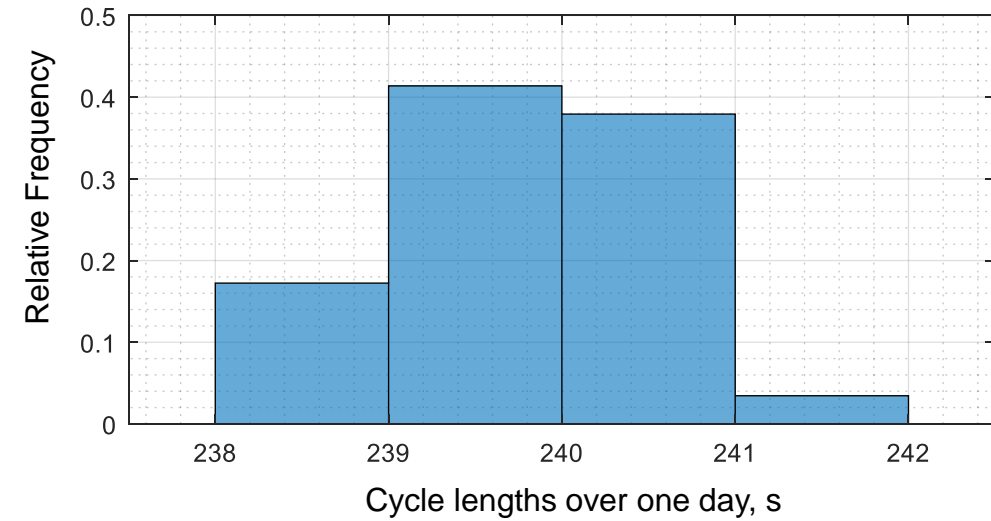
- Time series analysis
- Find deviations from normal cycle and
- Distinguishing between time and pattern deviation

What is *normal*?



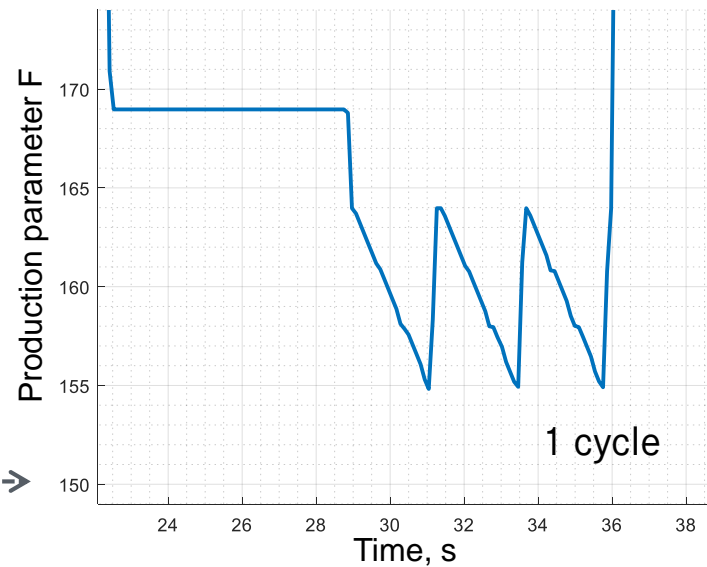
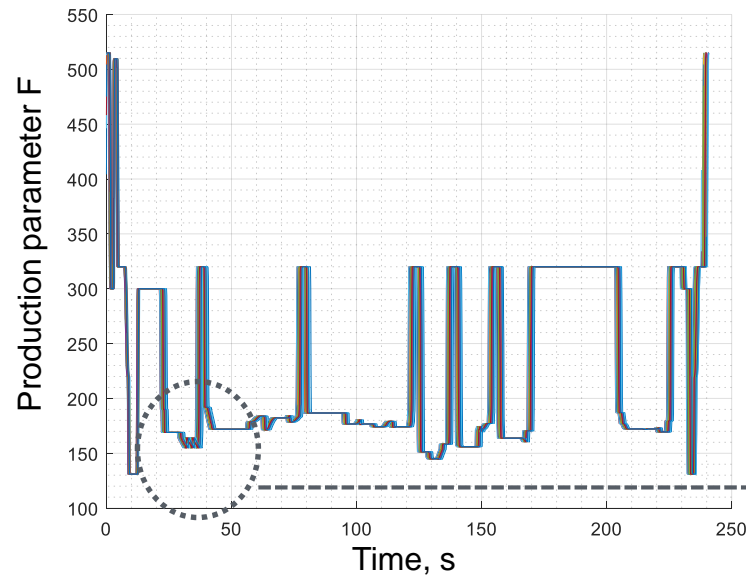
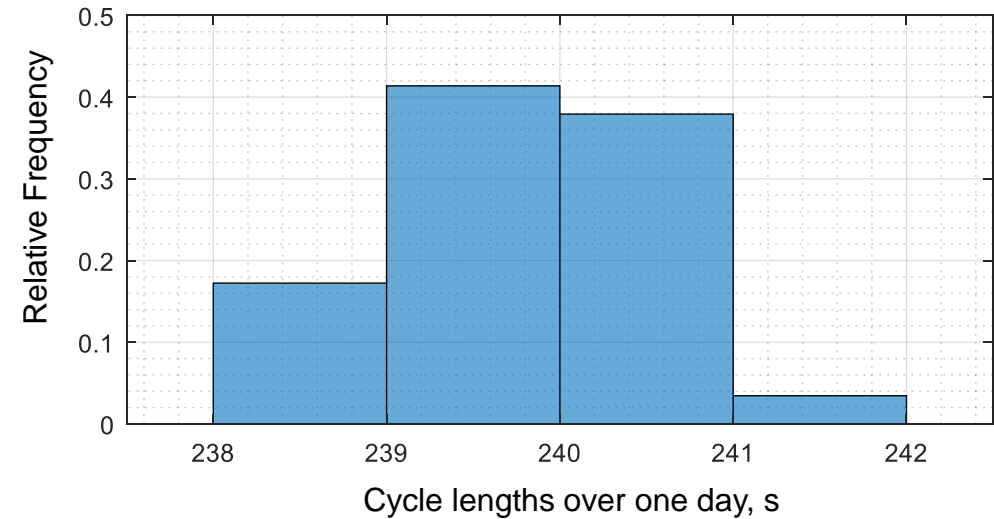
Delays in production cycles

- Length of time-series varies from cycle to cycle even for normal production



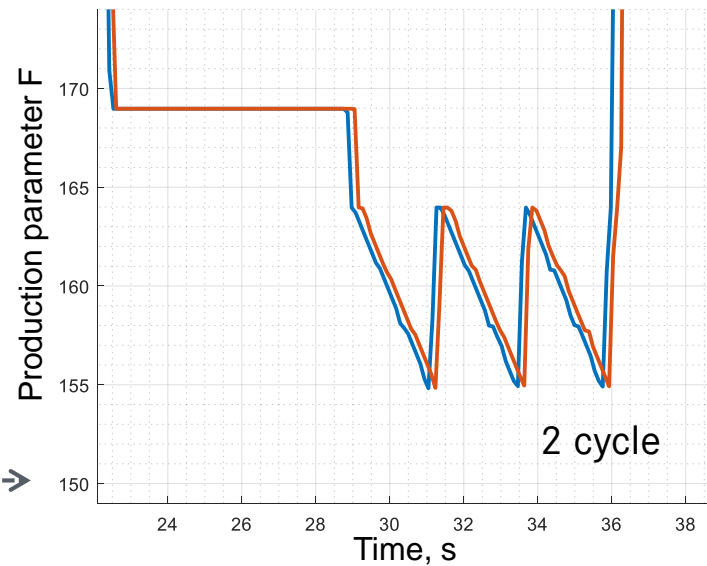
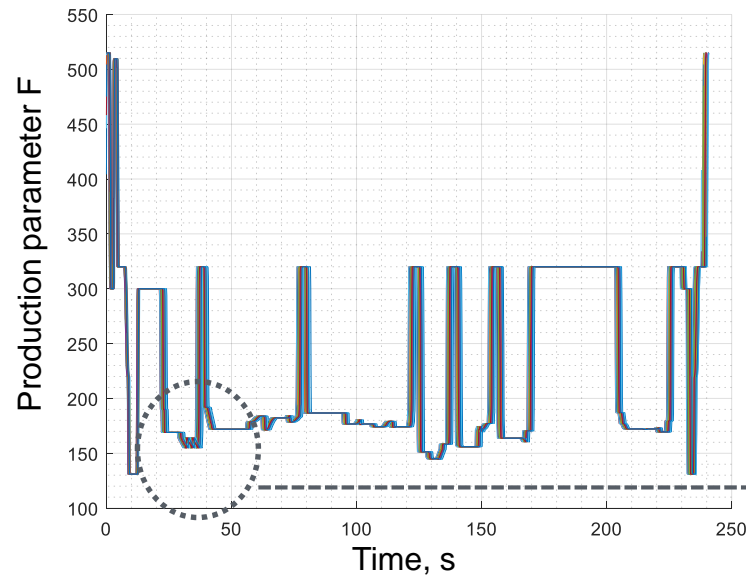
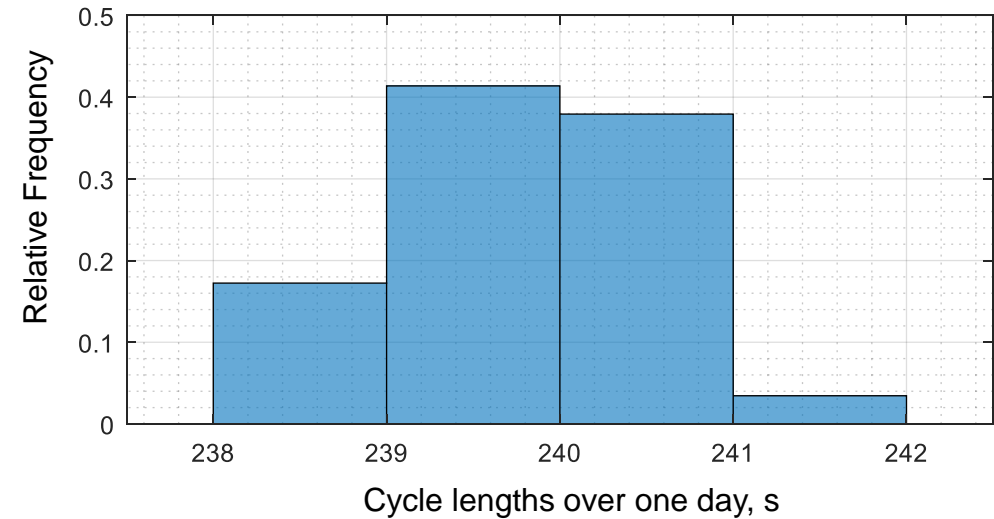
Delays in production cycles

- Length of time-series varies from cycle to cycle even for normal production



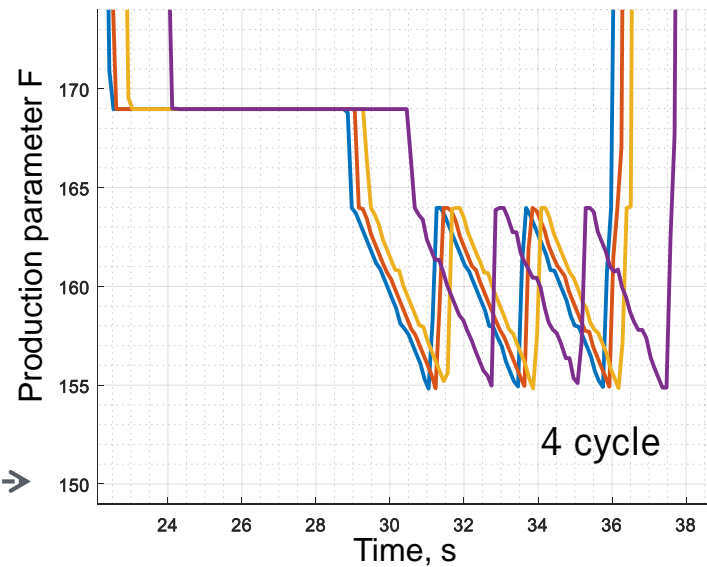
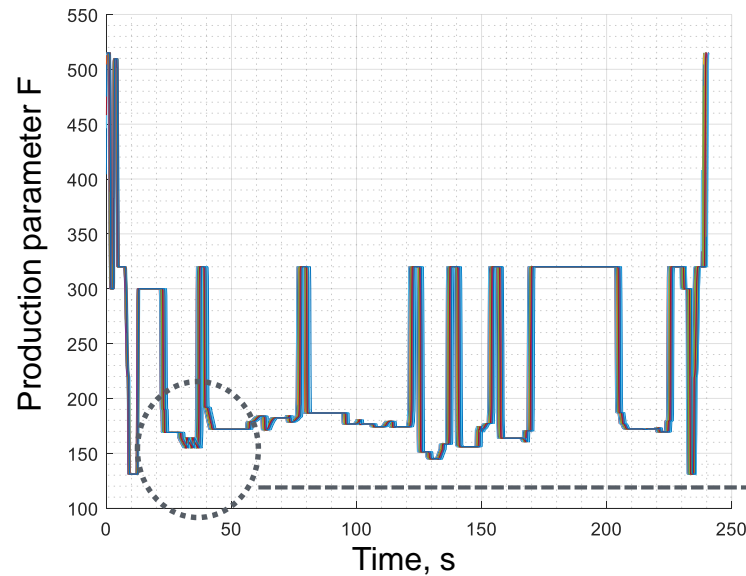
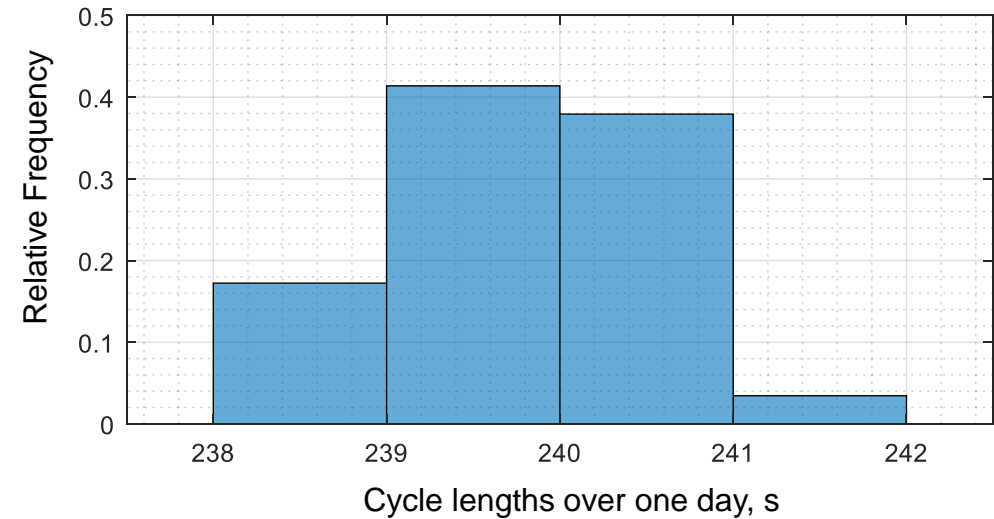
Delays in production cycles

- Length of time-series varies from cycle to cycle even for normal production



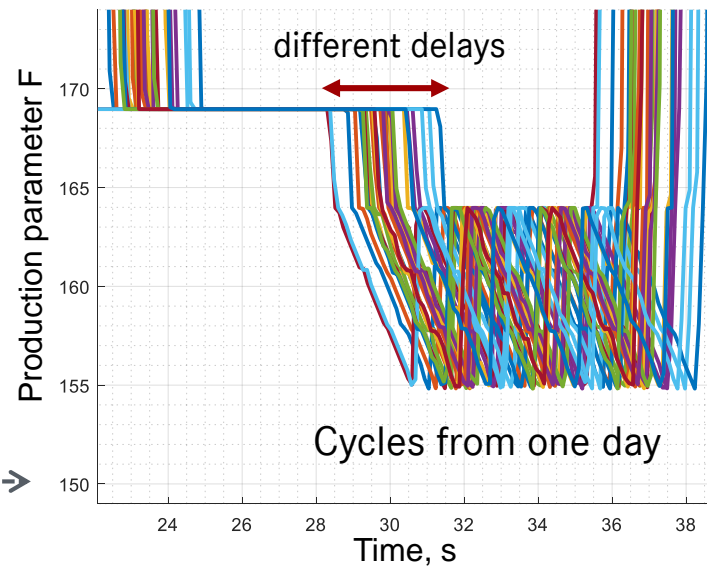
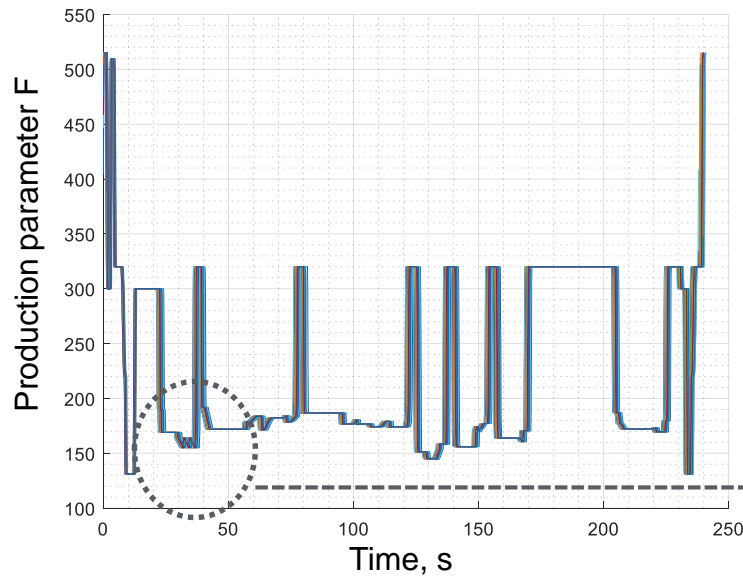
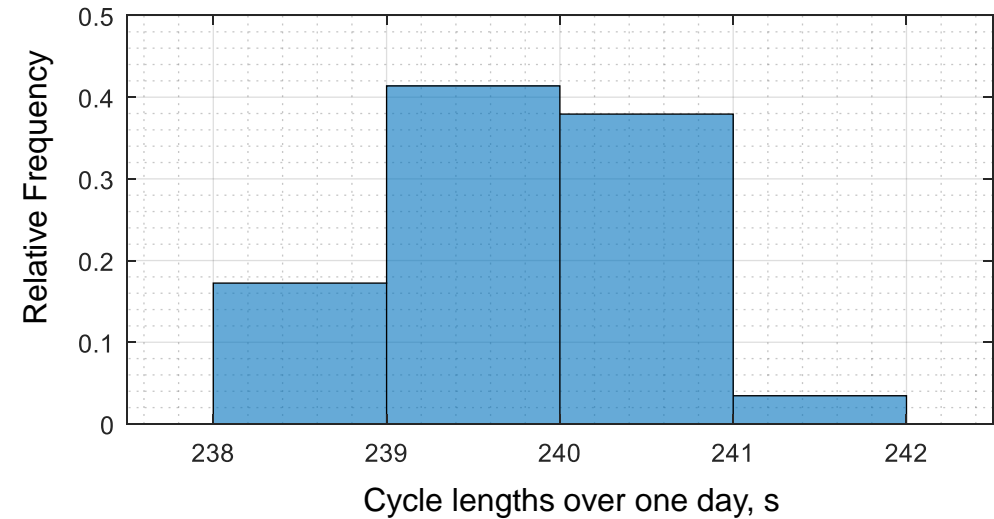
Delays in production cycles

- Length of time-series varies from cycle to cycle even for normal production



Delays in production cycles

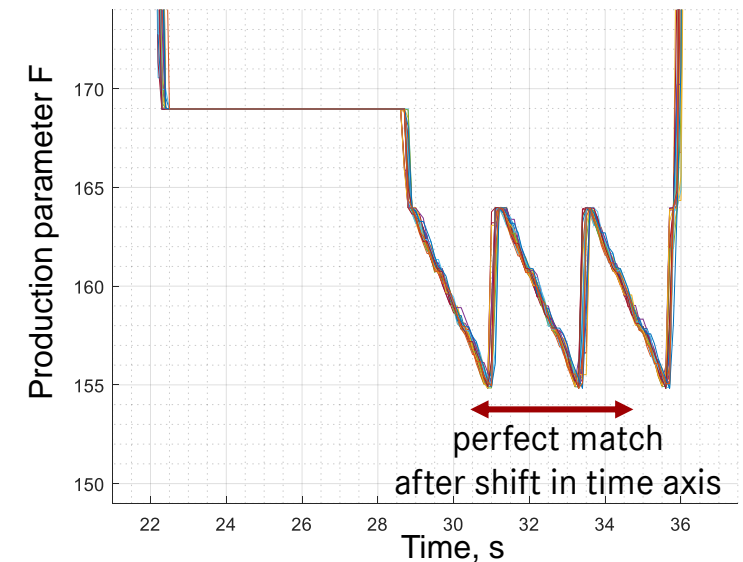
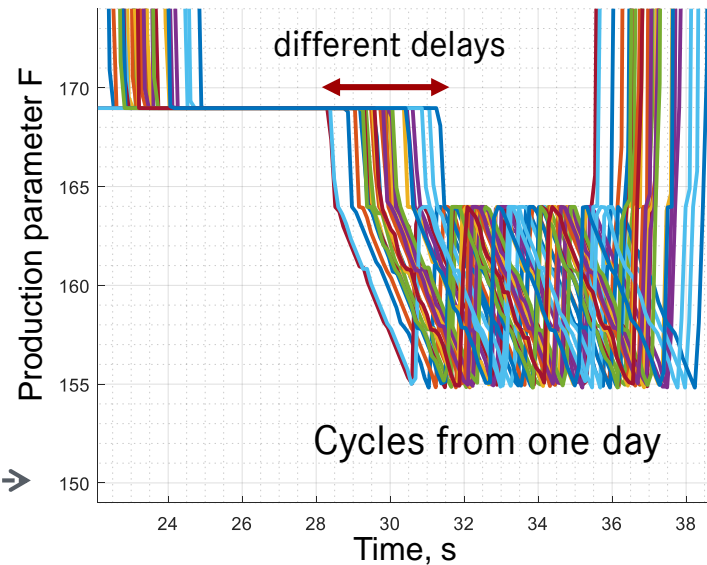
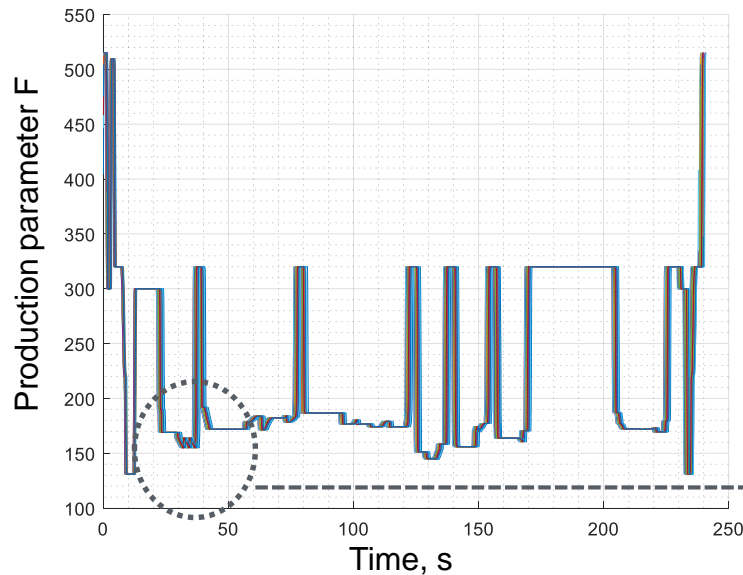
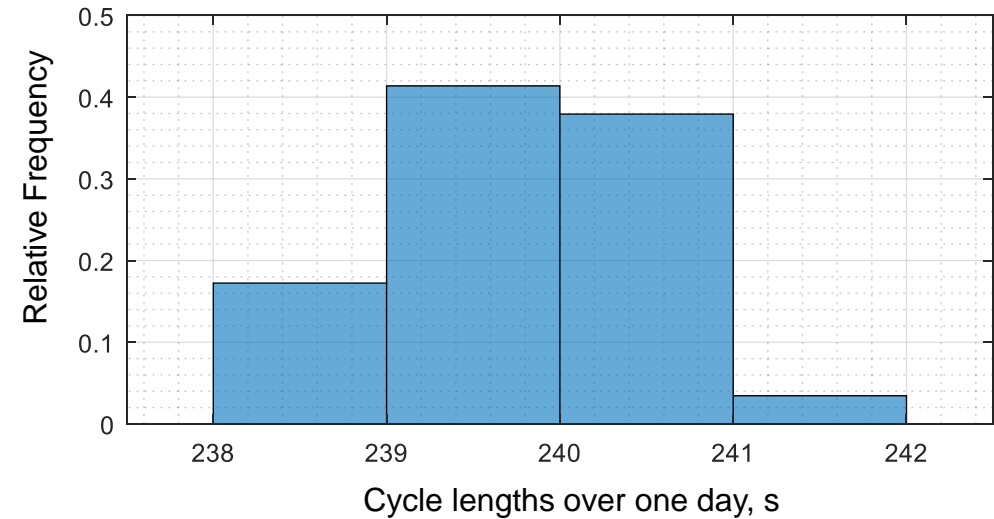
- Length of time-series varies from cycle to cycle even for normal production



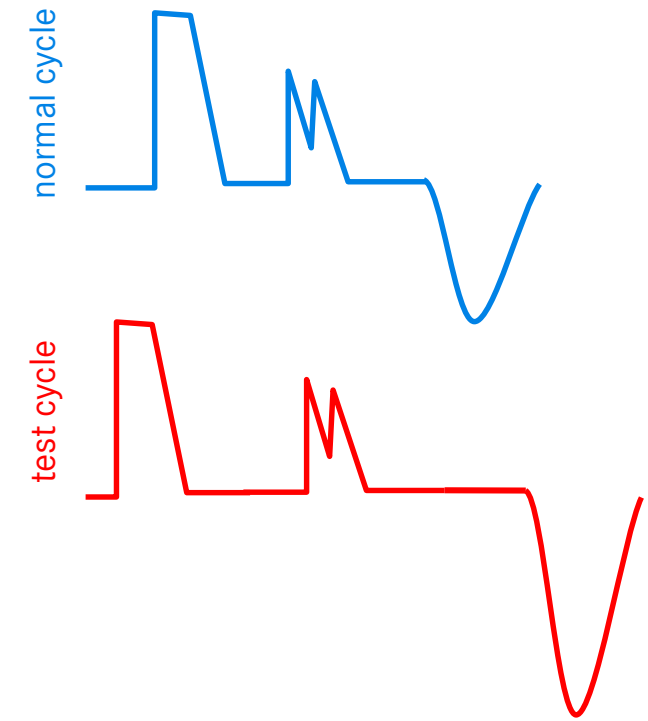
Delays in production cycles

- Length of time-series varies from cycle to cycle even for normal production

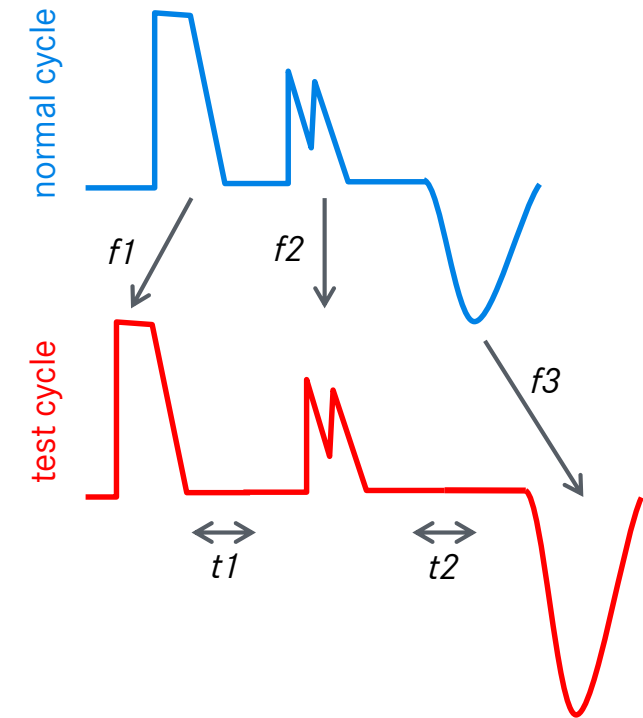
⇒ Normal cycles can be matched to one another through shifting in time axis!



Algorithm principle

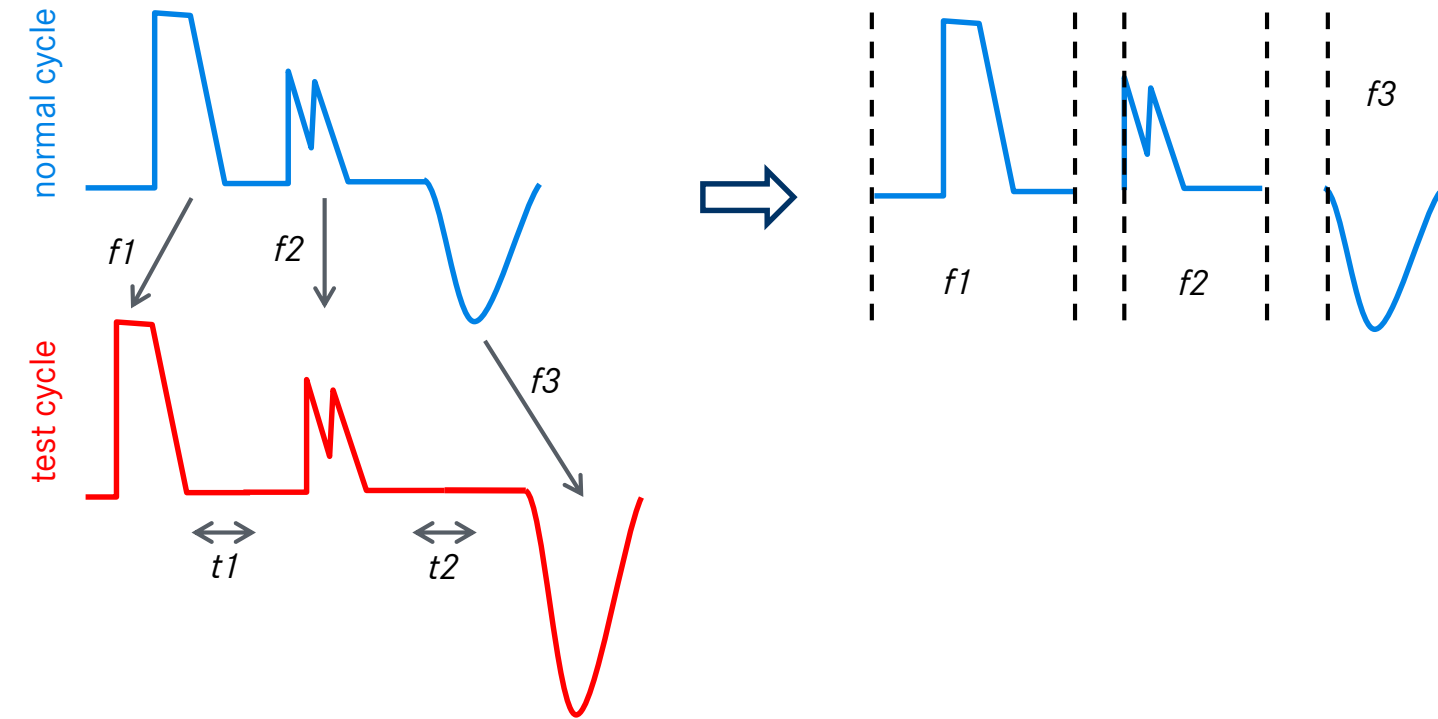


Algorithm principle



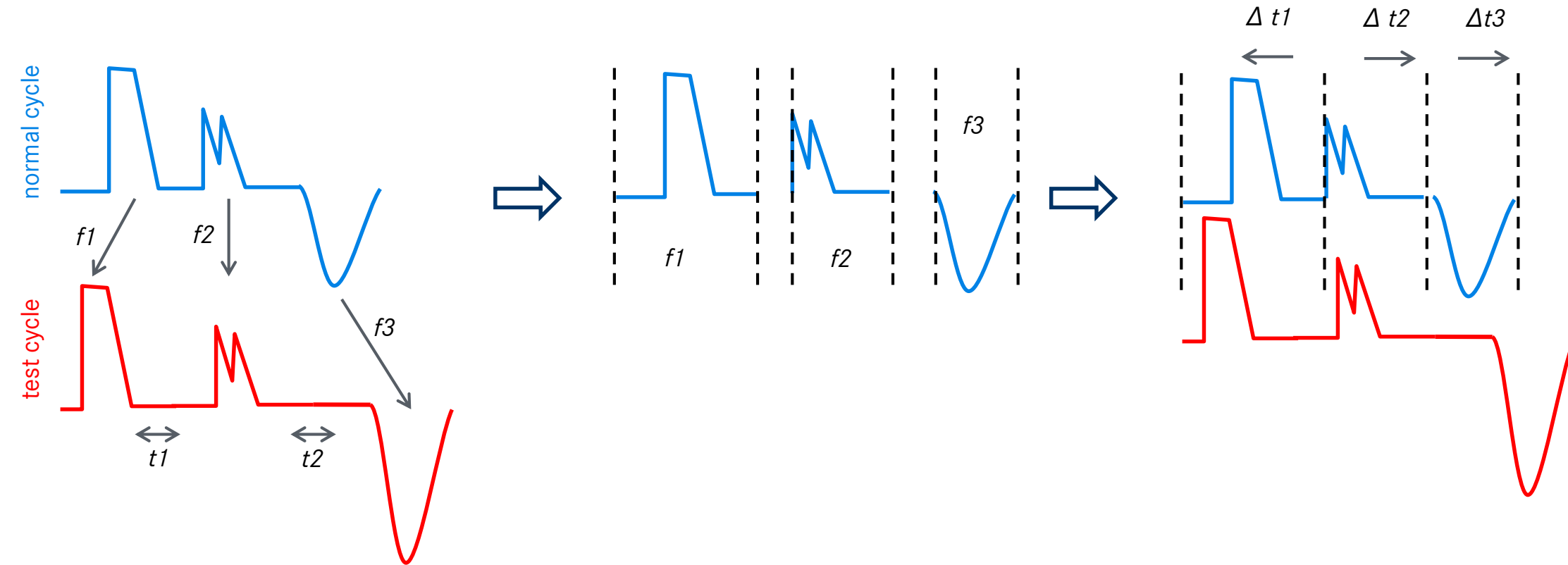
- Cycle can be described as sequence of features $f1$, $f2$, $f3$
- Each cycle can show some delays in time $t1$, $t2$

Algorithm principle



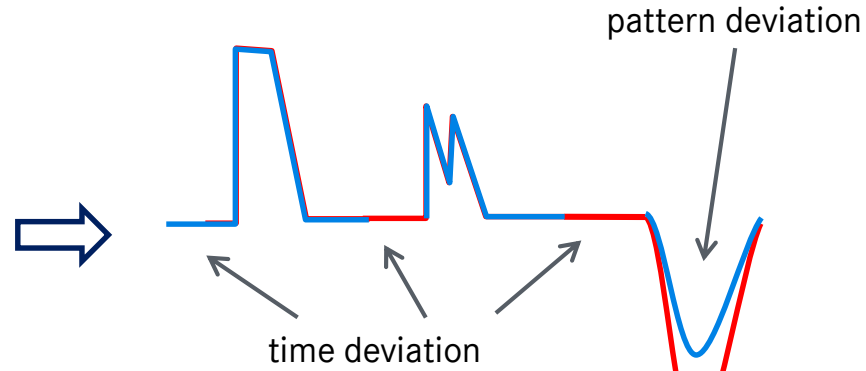
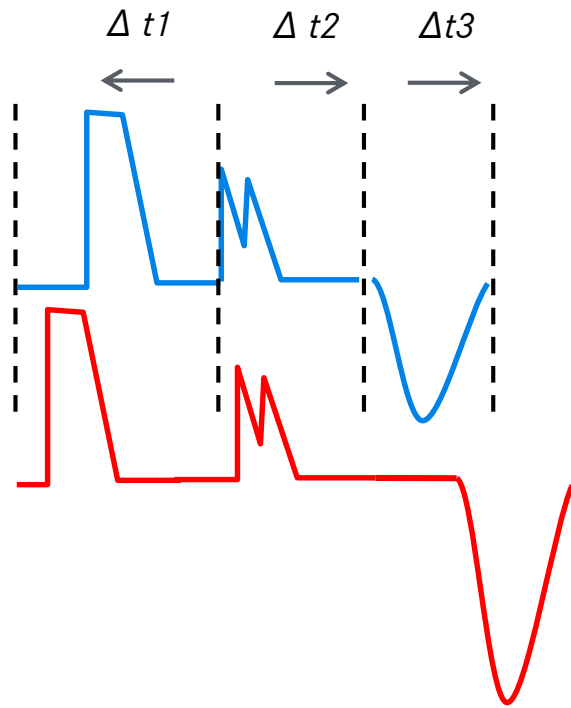
- Cycle can be described as sequence of features $f1, f2, f3$
- Each cycle can show some delays in time $t1, t2$
- Automatic feature detection $f1, f2, f3$

Algorithm principle



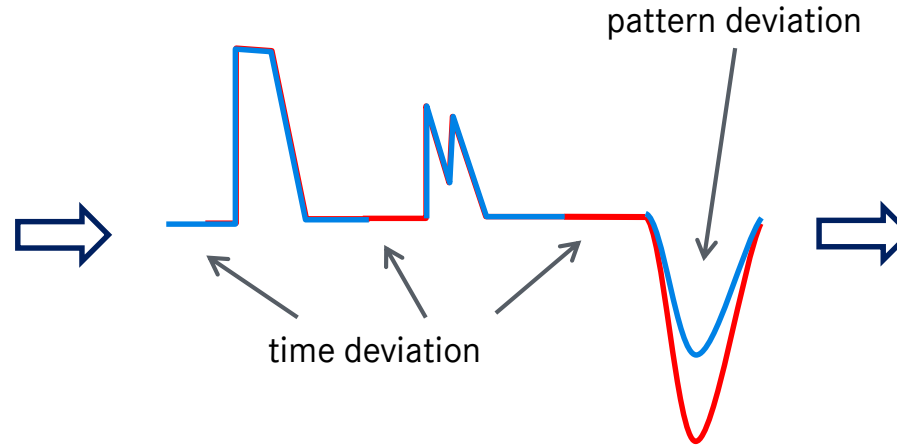
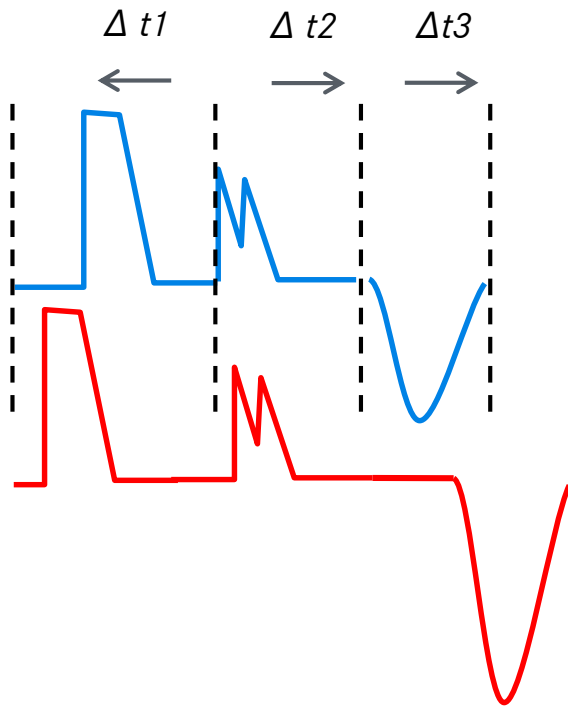
- Cycle can be described as sequence of features f_1, f_2, f_3
- Each cycle can show some delays in time t_1, t_2
- Automatic feature detection f_1, f_2, f_3
- Pattern matching through shift of feature along time axis ($\Delta t_1, \Delta t_2, \Delta t_3$): least square fit (t_{shift} to minimize the Sum of Residual Squares of two signals)

Algorithm principle



- Pattern matching through shift of feature along time axis ($\Delta t_2, \Delta t_2, \Delta t_3$): minimization of SRS

Algorithm principle



$f1$	$f2$	$f3$	
$\Delta t1$	$\Delta t2$	$\Delta t3$	Time deviation
No	No	Yes	Pattern deviation

- Pattern matching through shift of feature along time axis ($\Delta t1$, $\Delta t2$, $\Delta t3$): minimization of SRS

- Description of a cycle as feature sequence
- For each feature time and pattern deviation can be calculated

- Time and pattern deviation for each feature are used as characteristic numbers for test cycle

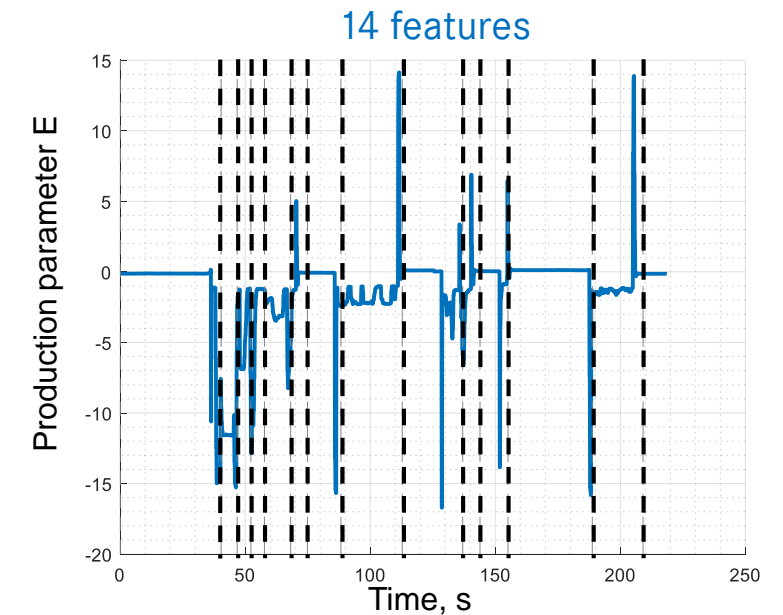
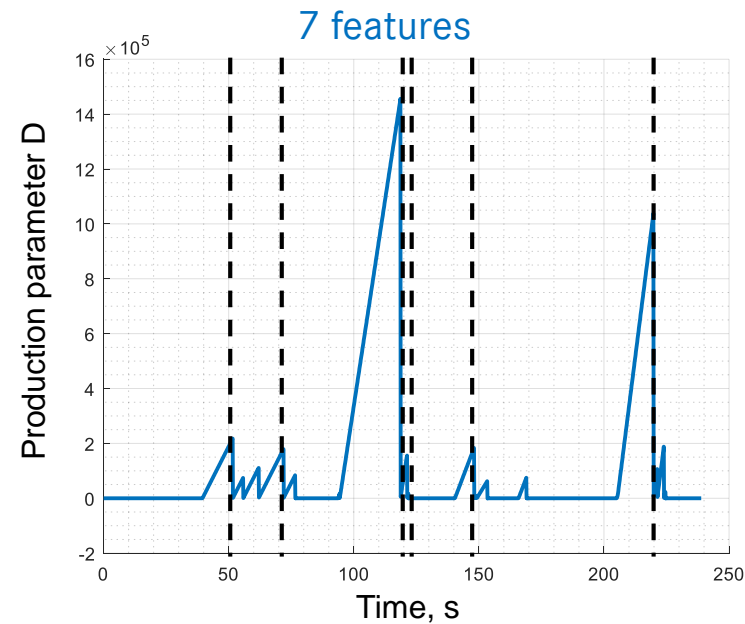
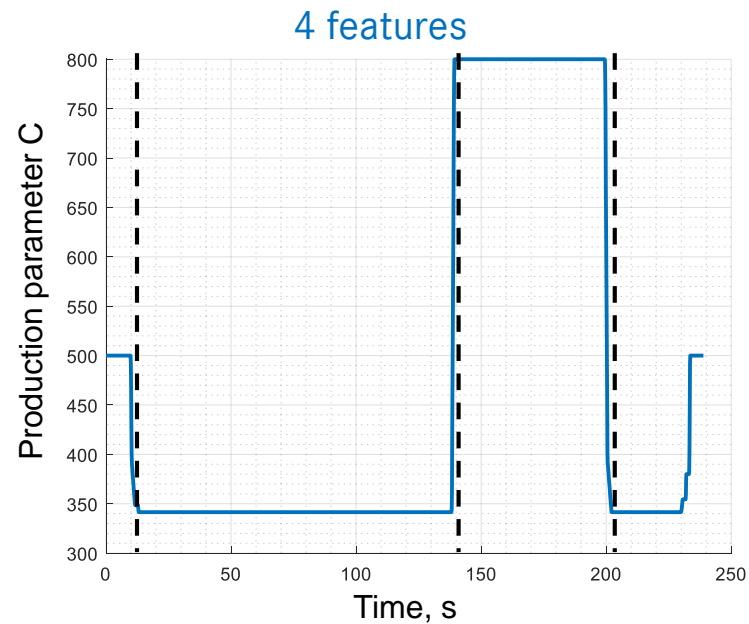


Data reduction!

Automatic feature detection

Time series is split

- After a local extremum (maximum or minimum) or on a plateau
- After a given relative change



➔ Data reduction of time series from 2500 datapoints to sequence of max. 60 features (typically 10)!

Algorithm implementation: machine learning approach in MATLAB



Reference cycles ->

- Build „reference signal“ for each feature
- Limits for time and pattern deviation



Test cycles ->

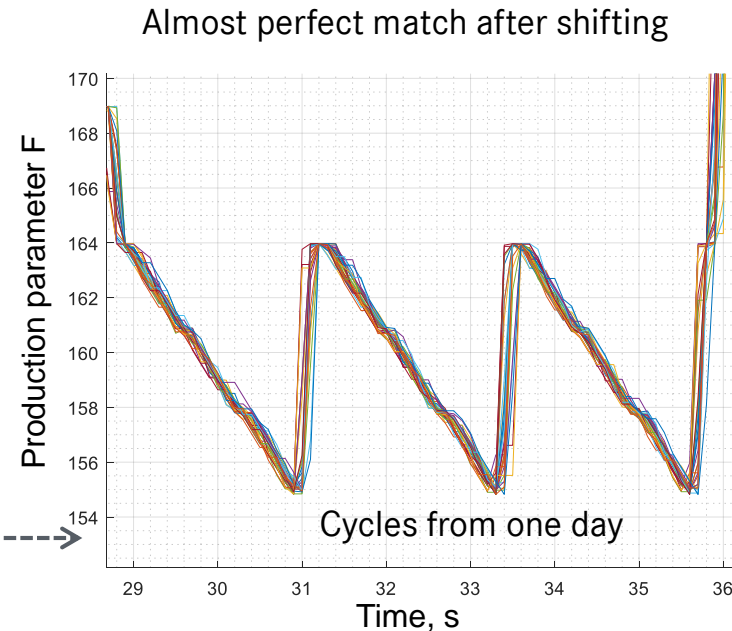
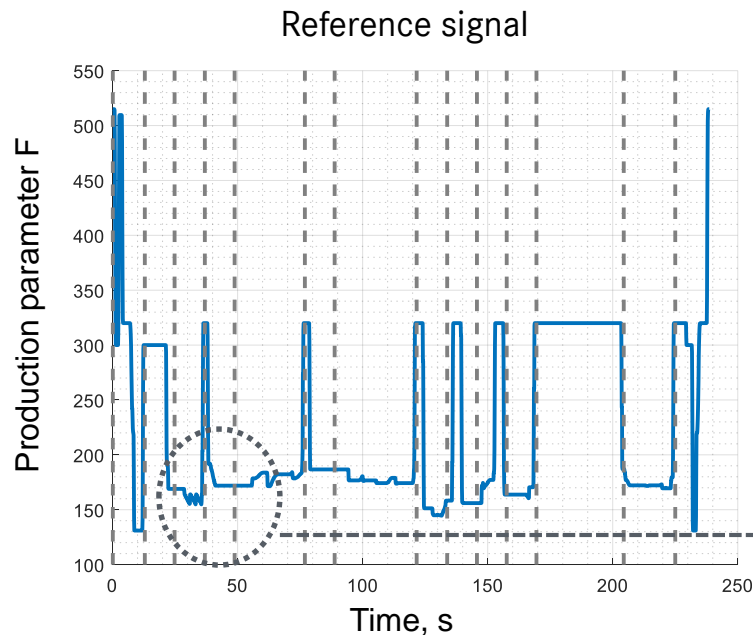
- Comparison of each feature in reference signal
- Is time and pattern deviation within the limits?

Create „reference signal“ for each production parameter



Training

1. For all training cycles - matching to shortest cycle
2. Create „reference signal“ – mean over all matched reference cycles

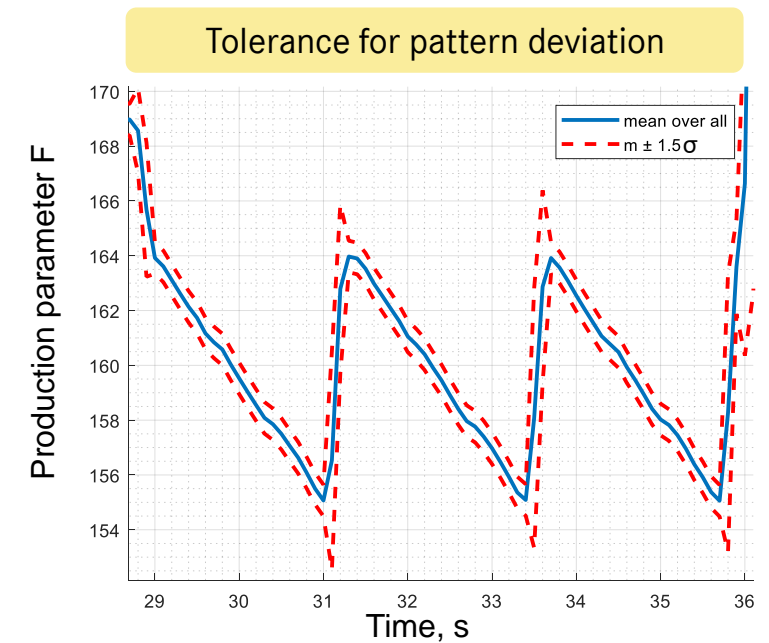
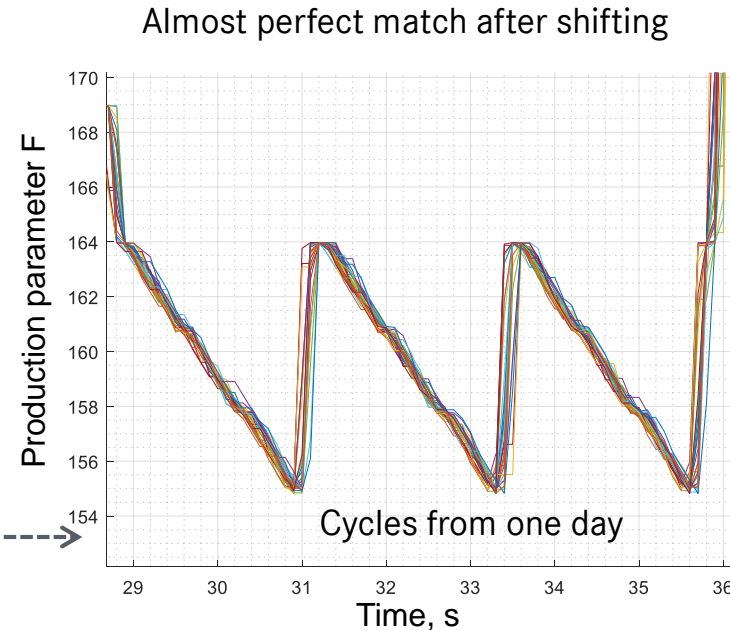
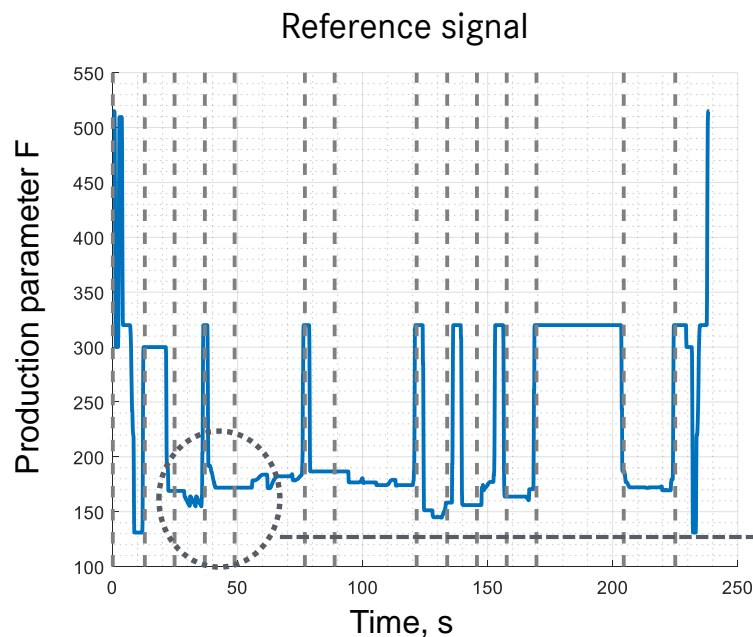


Create „reference signal“ for each production parameter



Training

1. For all training cycles - matching to shortest cycle
2. Create „reference signal“ – mean over all matched reference cycles
3. Possible pattern deviation - standard deviation over all matched reference cycles



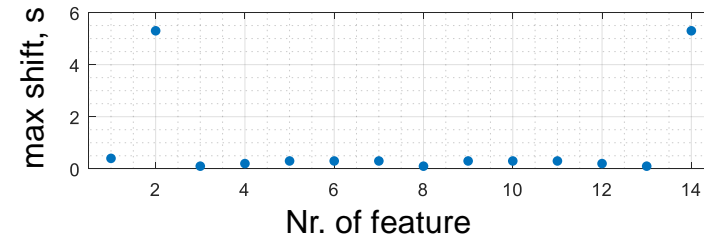
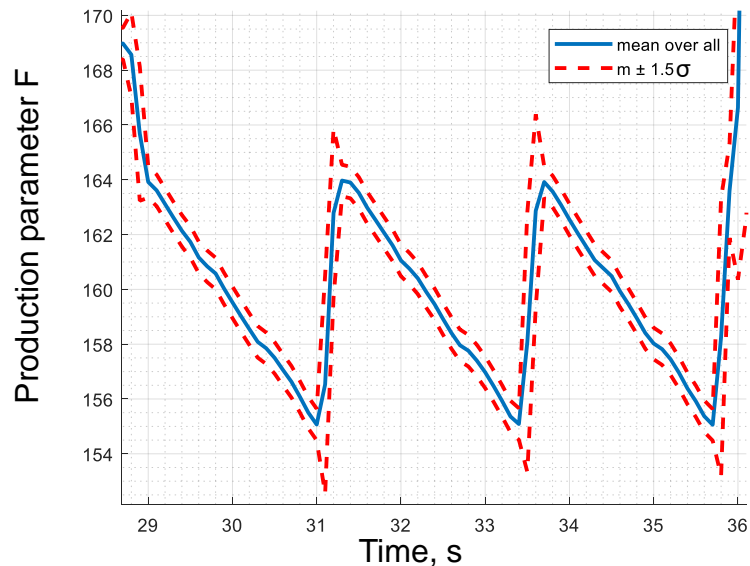
Create „reference signal“ for each production parameter



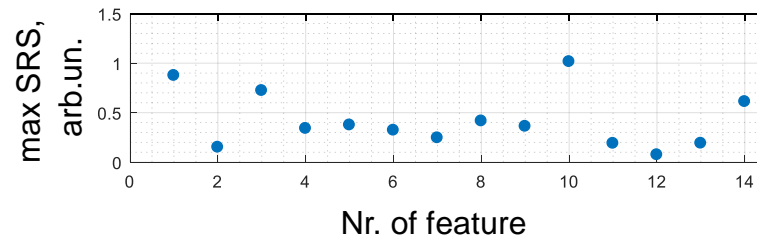
Training

1. Create „reference signal“ – mean over all matched reference cycles
2. Possible pattern deviation - standard deviation over all matched reference cycles, limits for SRS
3. Possible time deviation – maximal absolute shift from matched reference cycles

Tolerance for pattern deviation



Possible time deviation

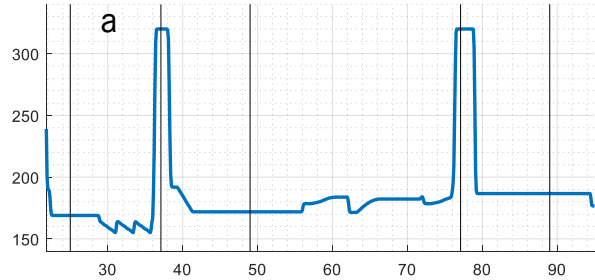


Possible pattern deviation

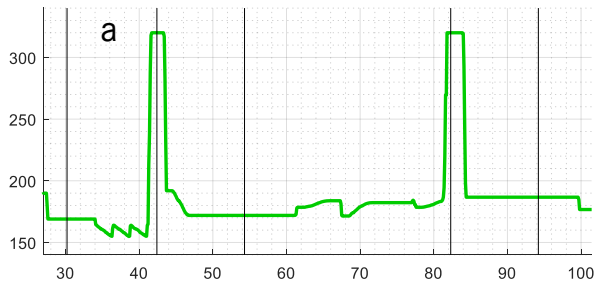
Testing: time and pattern deviation evaluation



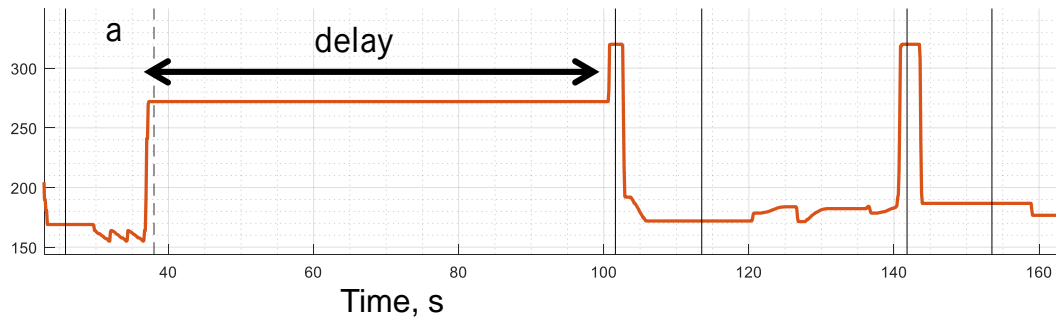
reference signal



normal cycle



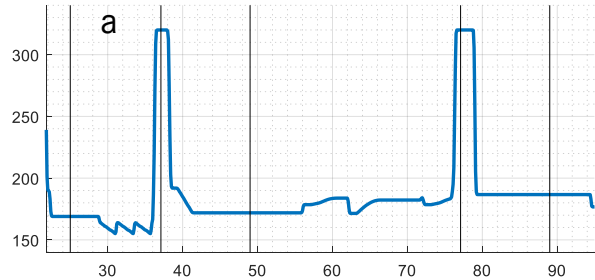
error cycle



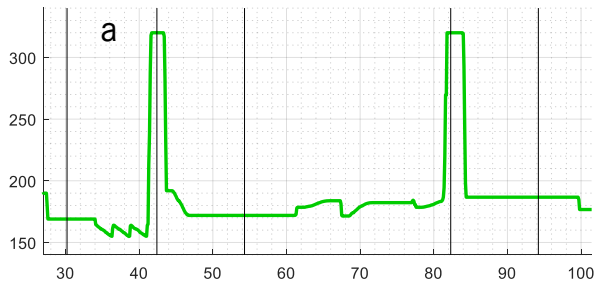
Testing: time and pattern deviation evaluation



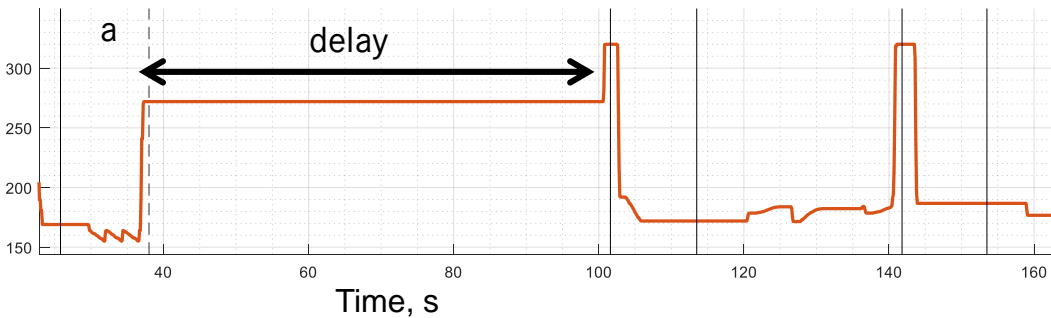
reference signal



normal cycle

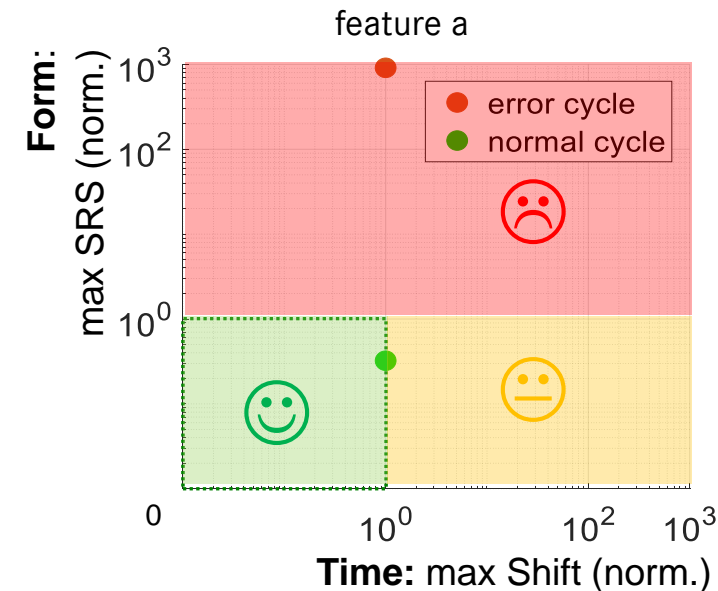


error cycle



Is time and pattern deviation for this feature within the limits?

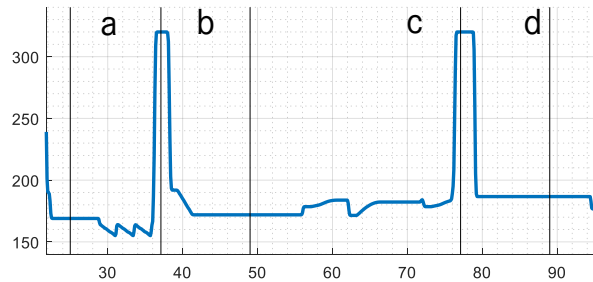
- Tolerance window (Δt , *SRS*)
- Easy to spot a critical deviation



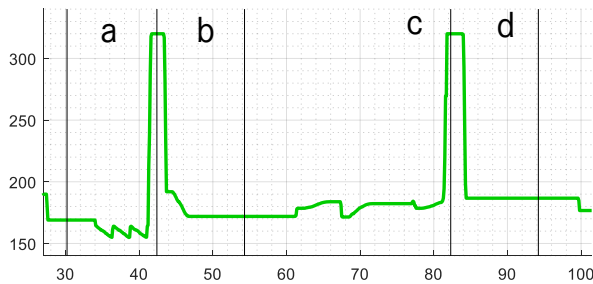
Testing: time and pattern deviation evaluation



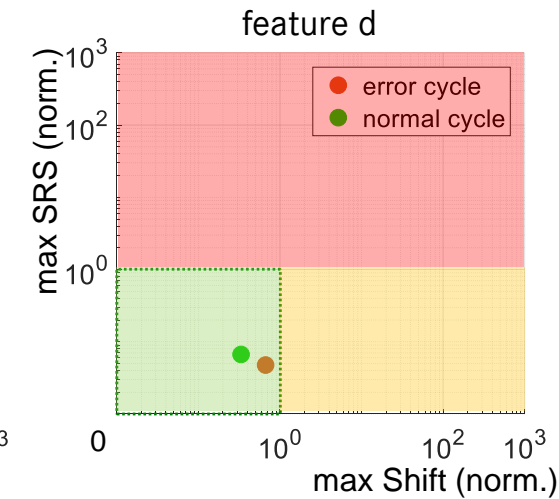
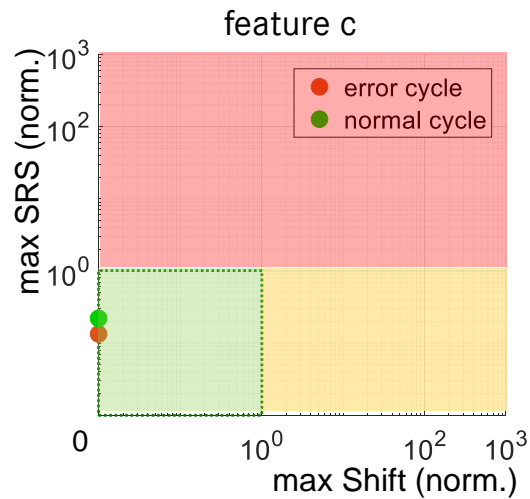
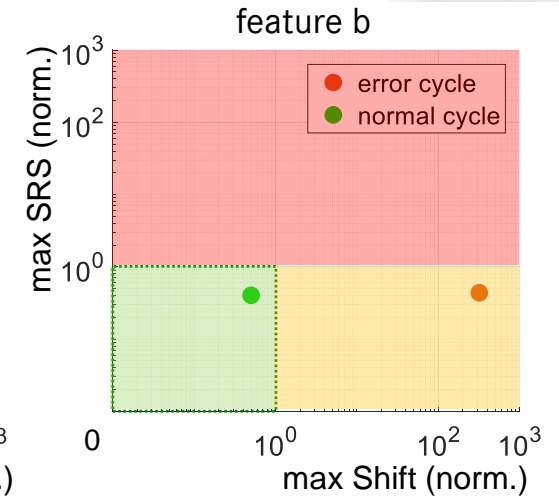
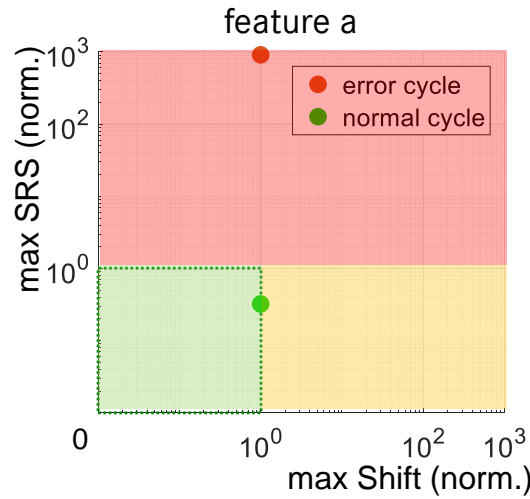
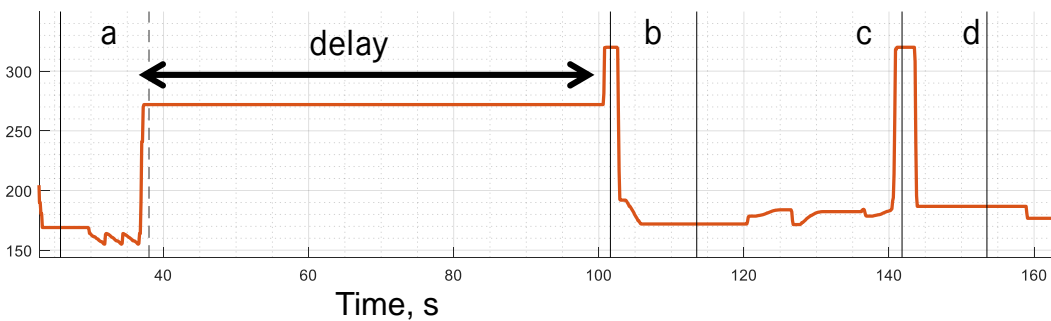
reference signal



normal cycle



error cycle



Algorithm: Summary

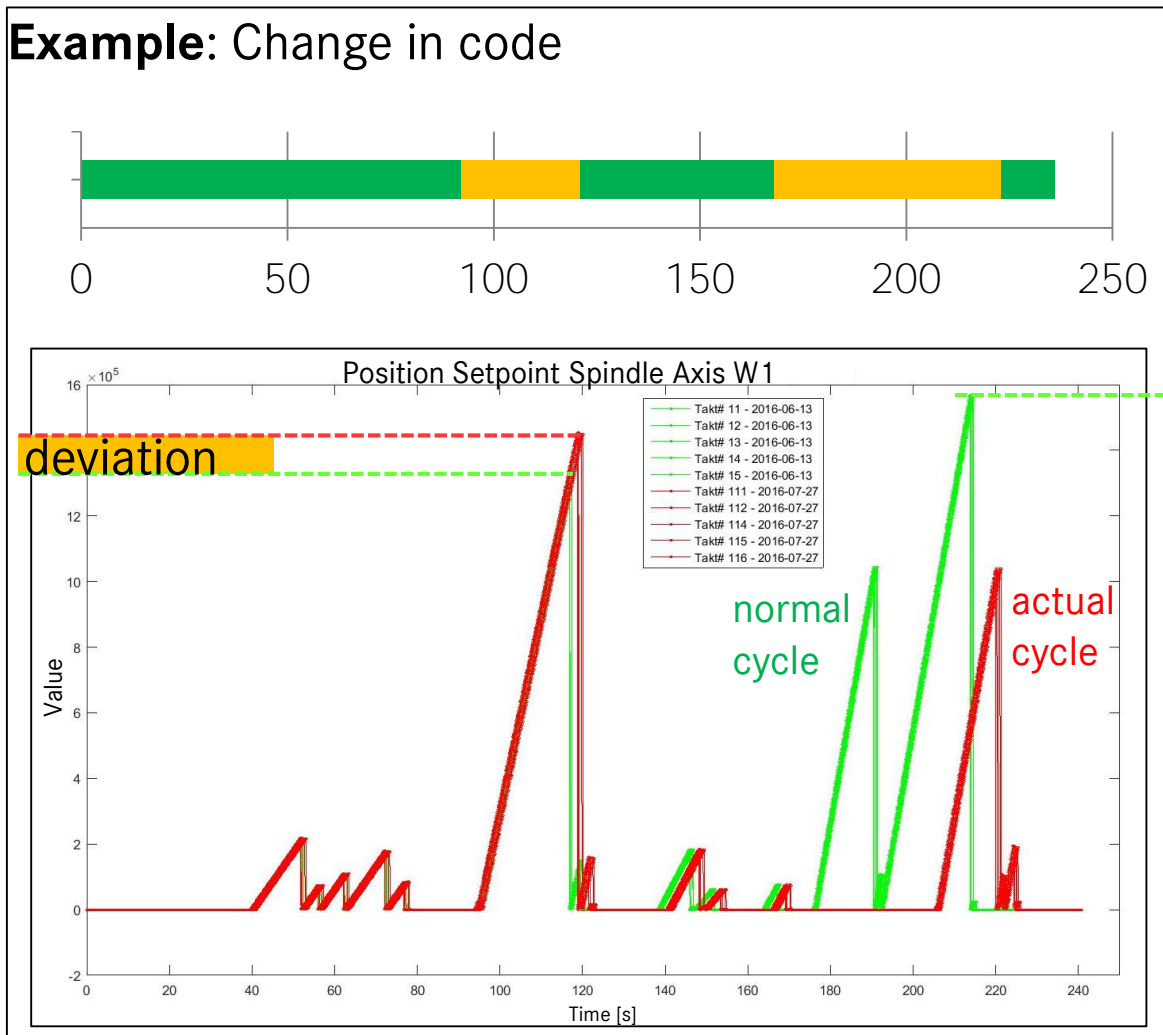
1. Quantitative and qualitative description of production failure
2. Independent of signal form -> universally applicable to other applications or machines
3. Signal description with characteristic numbers, which are easy to interpret
4. Data reduction with a factor **250** without significant loss of information!
5. Easy control of production: recognition of critical errors and non-critical delays online

DR. TÜRCK
INGENIEURBÜRO  DATA SCIENCE
io@tuerck-optik.de

Example of the added value



Example: Change in code



Results:

- Transparency of the process
 - Deviation for each Signal
 - Reason of Cycletime increase found
- ➔ Time and Pattern deviation are recognized

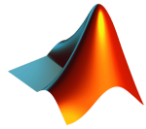
Summary



Algorithm using pattern matching for time series developed and implemented for production data

Why MATLAB?

- easy algorithm implementation
- existing solution for data import
- very good support and broad use in universities



MATLAB Products used:

- Signal Processing Toolbox
- Statistics and Machine Learning Toolbox

Outlook:

- Parallel Computing Toolbox for performance improvement



Prototyp intelligent Level-Learning (iLL)  has a new function for anomaly detection

- Troubleshooting in case of failure (maintenance), Parts Planning, Influences on the quality
→ Optimization of repair time, spreaders amount, ...

Thank you for your attention!



Mercedes-Benz

Jessica Fisch

jessica.fisch@daimler.com

DR. TÜRCK
INGENIEURBÜRO  DATA SCIENCE

The logo for Dr. Türck Ingenieurbüro Data Science features a stylized 'D' shape composed of a blue semi-circle on the left and a yellow semi-circle on the right.

Irina Ostapenko

io@tuerck-optik.de

Picture Credits: © Can Stock Photo / AnatolyM, abluecup, maxxystas