Mechmine

data science technologies

Mechmine LLC – predictive maintenance

Reality-Check bei der Vibrations-basierten Maschinen-überwachung mittels KI

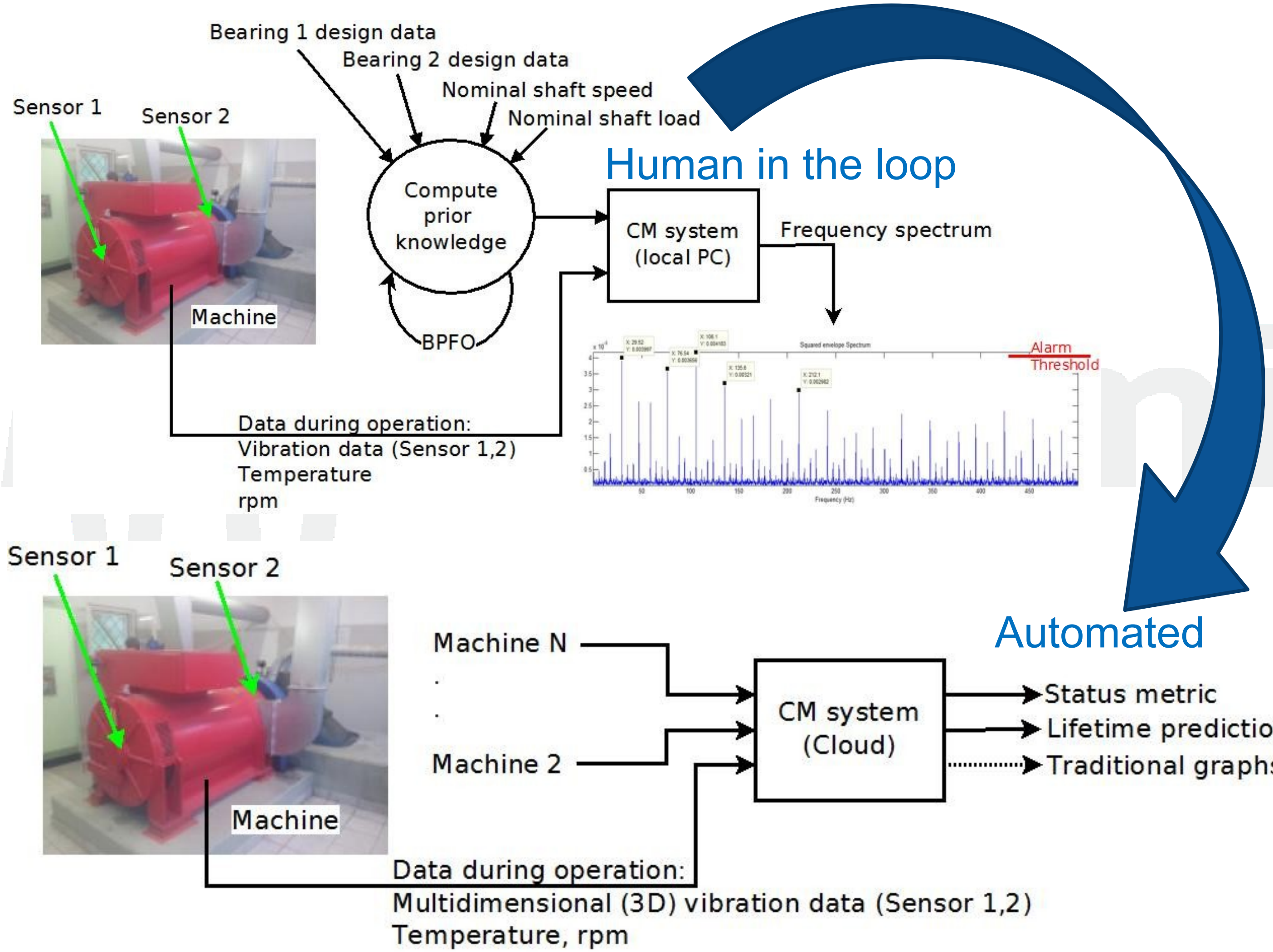
28. May 2019

Presented by Rudolf Tanner

Content

- Context
- Pre-requisites for deploying Machine Learning (ML) in industry
- Availability of data
- State-of-the-art ML in use
- Use of augmented data

Vibration-based monitoring Evolution - from FFT to AI



Applications

Gearbox
Bearings
Pumps
Ventilators
Conveyors
Mills
Escalators
and many more...

Pre-requisites

- Motivation in industry
  - Economical improvement
  - Lowering risk
- Where is Machine Learning (or AI) an option?
  - ML is old: regression, neural networks → operation research
  - Optimize: company processes, maintenance, production
  - Application: energy/resource optimisation, failure prediction, image classification
- Why is ML wanted?
  - Problem has no closed form solution
  - Problem is complex, e.g. highly-dimensional
  - Engineers/staff cannot solve it otherwise
  - Too much data to do it by hand

Data availability

- First, recall some facts
  - RIRO (rubbish in - rubbish out)
  - A fool with a tool is still a fool!
- Data needs
  - Quality – maximise information content
    - Sensor placement, SNR, resolution, bandwidth
  - Quantity – maximise data diversity & representation
    - In «sufficient» numbers: scenarios, context, environment
  - Time – maximise applicability
    - Seasonal effects, business/economic effects

Data availability
Example 1

- Bit resolution
  - Choose the right analg-to-digital converter
  - MEMS sensor: SNR = 40g/5.6mg=7143 $\rightarrow$ 12.8bit
  - Piezo sensor: SNR = 100g/0.1mg=1Mio $\rightarrow$ 19.9 bit
  - Makes little sense to use 16 bit ADC for piezo sensors
  - Higher dynamic range means earlier detection
- Bandwidth
  - Choose the right sampling rate: physical vs. FFT resolution
  - 30 RPM $\rightarrow$ f_c=0.5Hz
  - 10 RPM $\rightarrow$ $f_c \cong$0.16Hz
  - Acceleration-measurement is less suited for slow turning shafts, and velocity or displacement are alternatives, or, some advanced signal processing

Data availability
Example 2

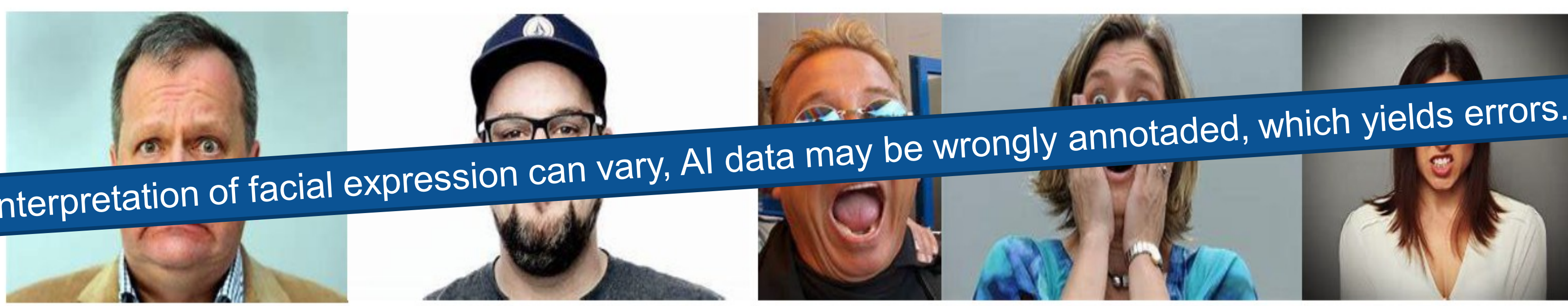
Mechmine
data science technologies

Three tiny data-set examples, taken from the Internet, which show pitfalls and problems with interpretation of data, and consequently AI-based solutions:

M h e n c z e t r i r ... i n a g l

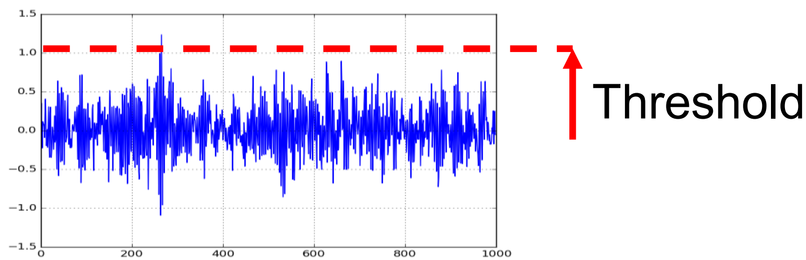
Menschen über 55 in der Schweiz werden l... ...gereien jedes Jahr um 400 Millionen Franken geprellt.

Gaps are too big, AI may not be able to complete the text.

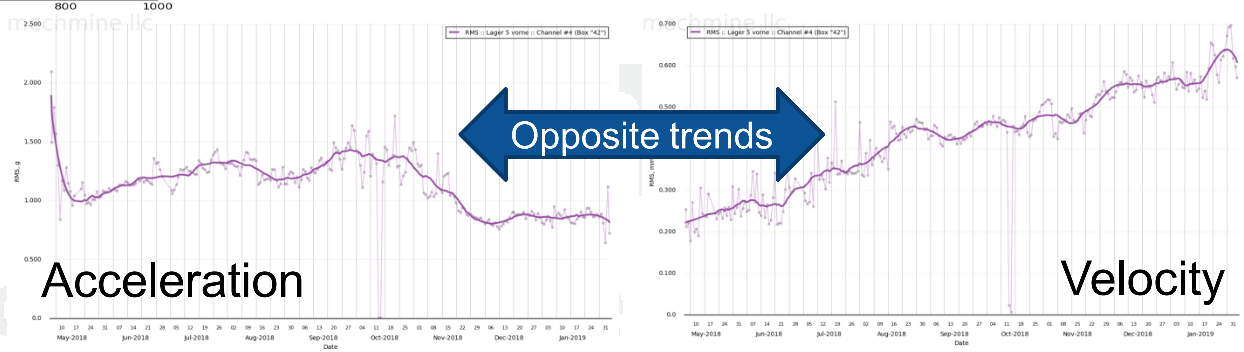


State-of-the-art
Thresholding, still popular today

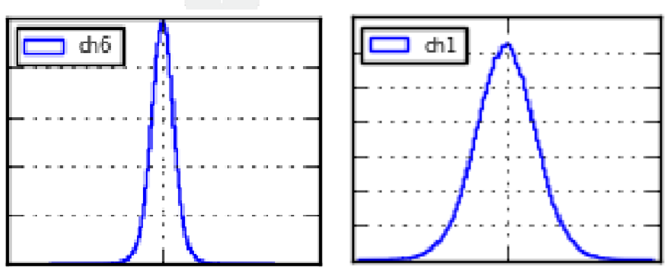
- One does not really need ML in many IoT applications



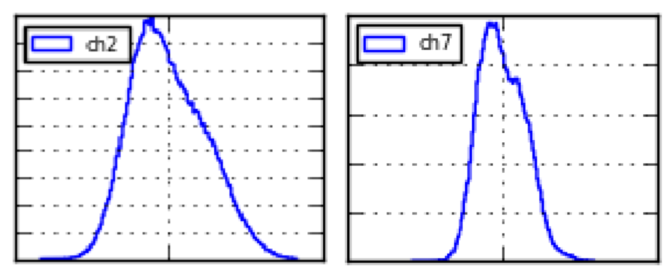
But what to track?



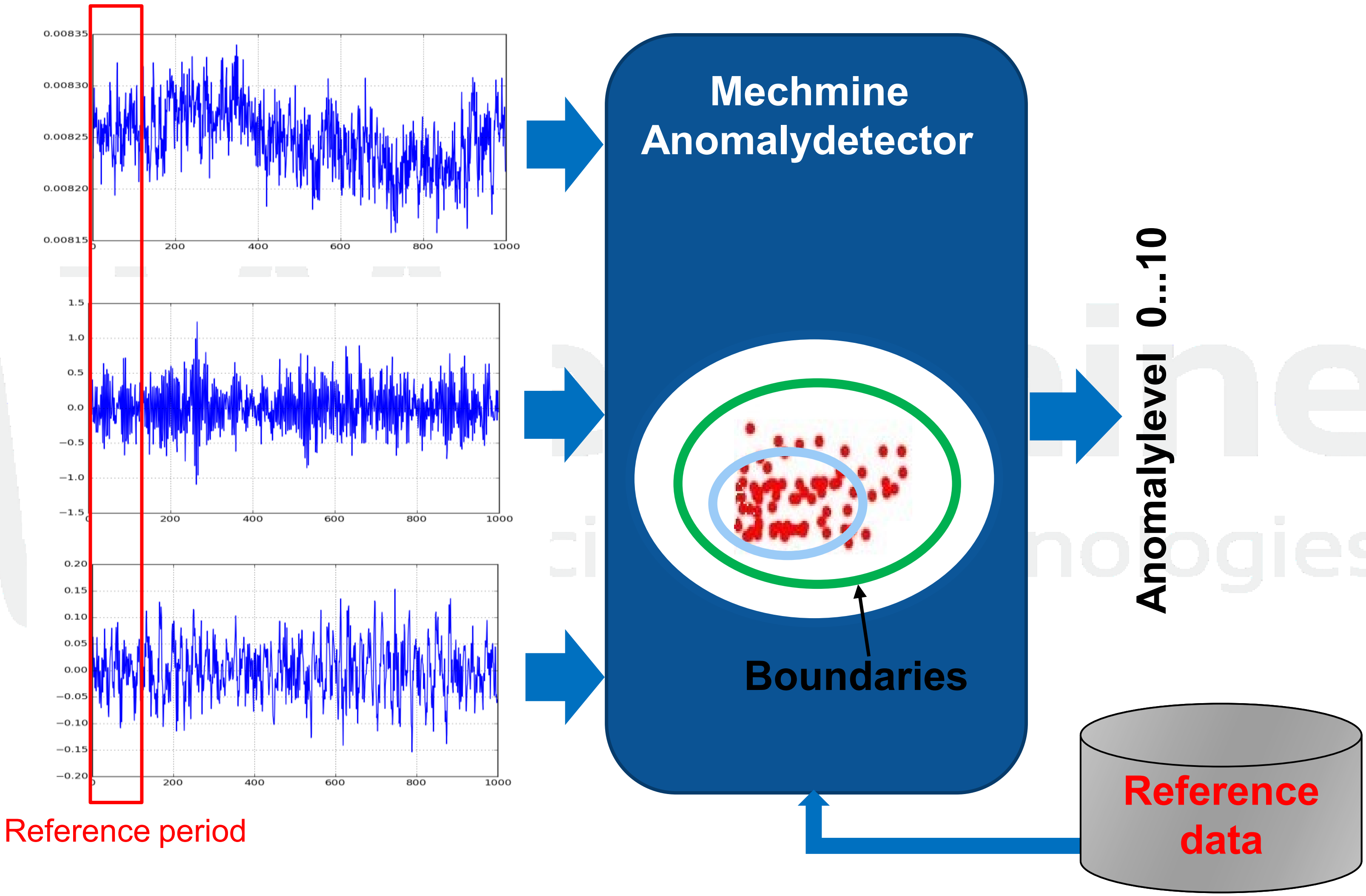
- Common assumption is Normal distribution



But in reality:



State-of-the-art
Outlier detection, a typ. AI solution



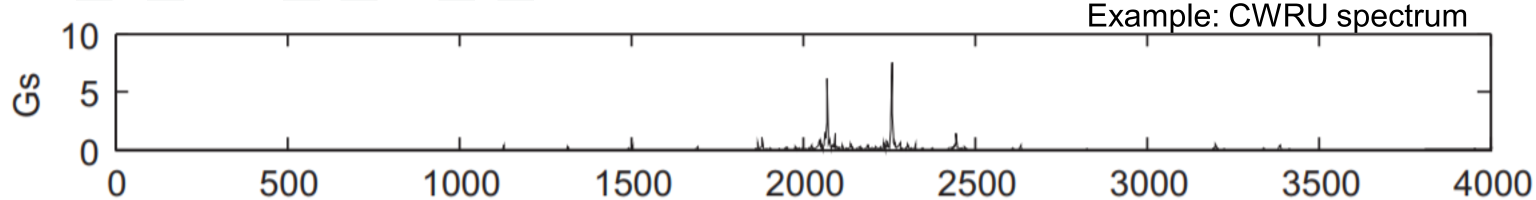
Use of augmented data
Data feeds ML

- Lots of data is needed to train a Deep Neural Network
  - Recall: typ. 1000 images per class for model training
- Assume we get one useful data sequence per machine component, e.g. bearing, per day
  - Then we collect data over 2.5 years
- Machine manufacturer
  - Customers use same machine in very different ways
    - Gearbox on crusher in cement mill or mine
    - Gearbox in cableway
    - Gearbox in sewage treatment plant
- AI prediction applications rely on huge datasets
- Who has data is king

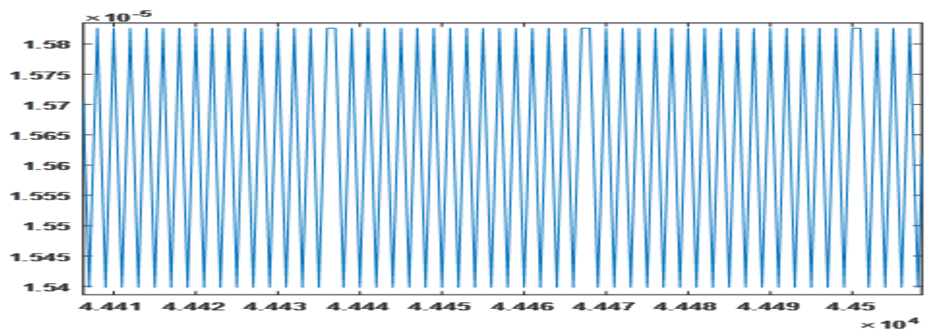
Use of augmented data
Use of „Internet“ data

- Available data in the Internet is generally less suited
  - Used by researchers: CWRU, NASA, PHM, Paderborn
  - Few data, short sequences, bad resolution or even faulty
  - Lab data (too nice to be of practical use)

Example: CWRU spectrum

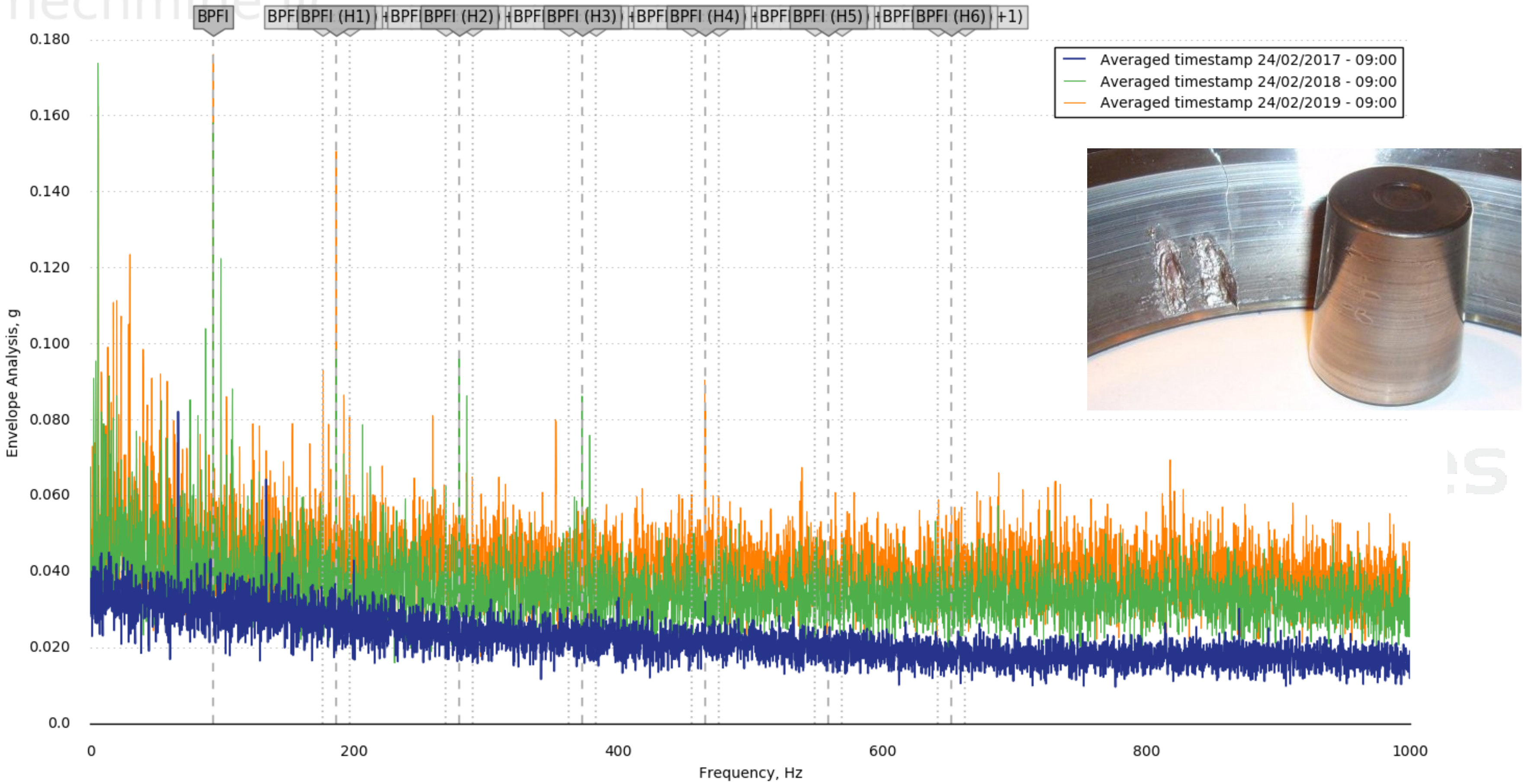


  - Most datasets contain only datasets of one fault class
  - Some are faulty (e.g. sampling jitter)

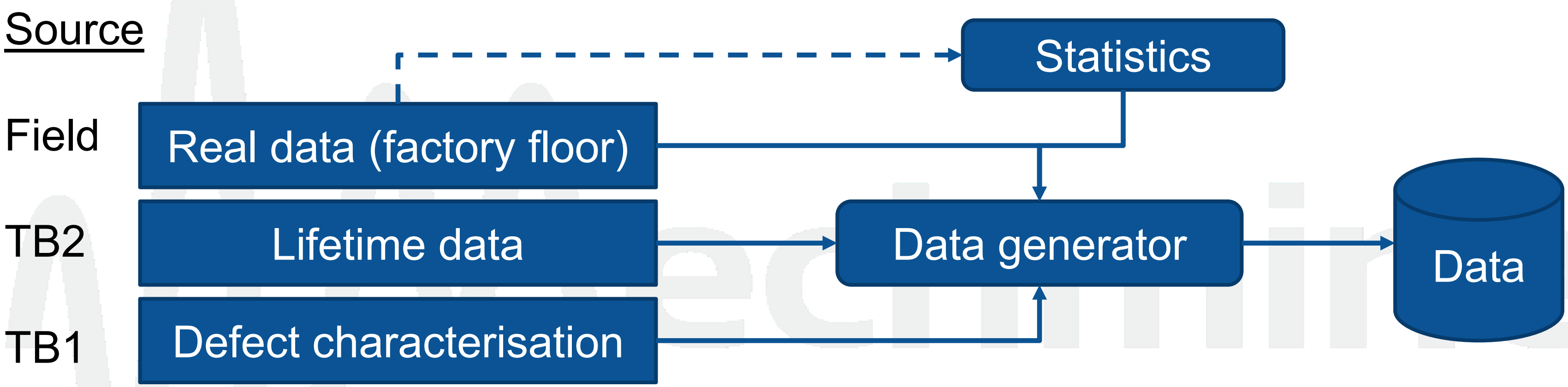


Use of augmented data
Real data

- Mechmine: exemplary BPFI defect



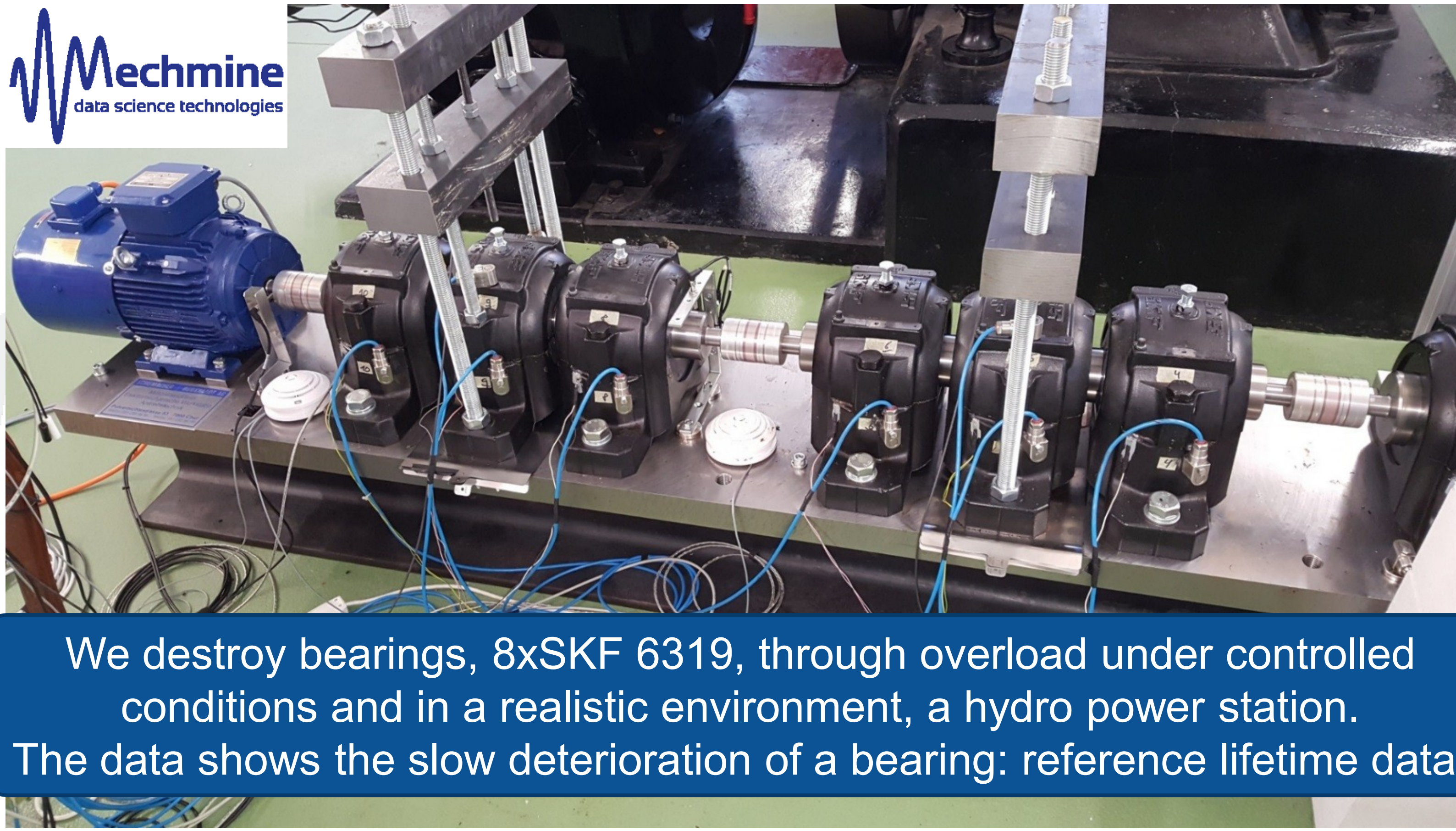
Use of augmented data Concept



We use field data (realistic vibration data) and testbench (TB) data to produce a useful data pool.

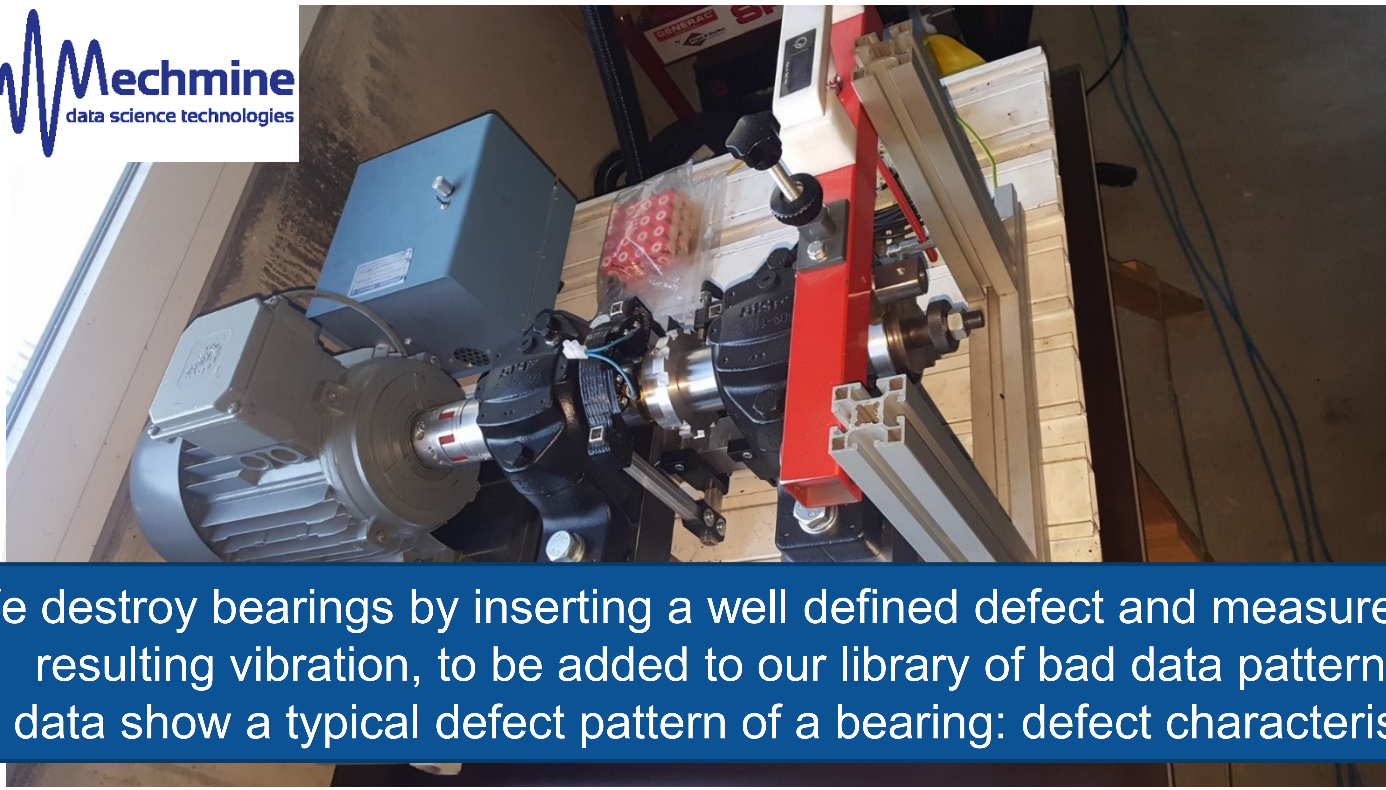
Use of augmented data
Lifetime data TB2

Mechmine
data science technologies



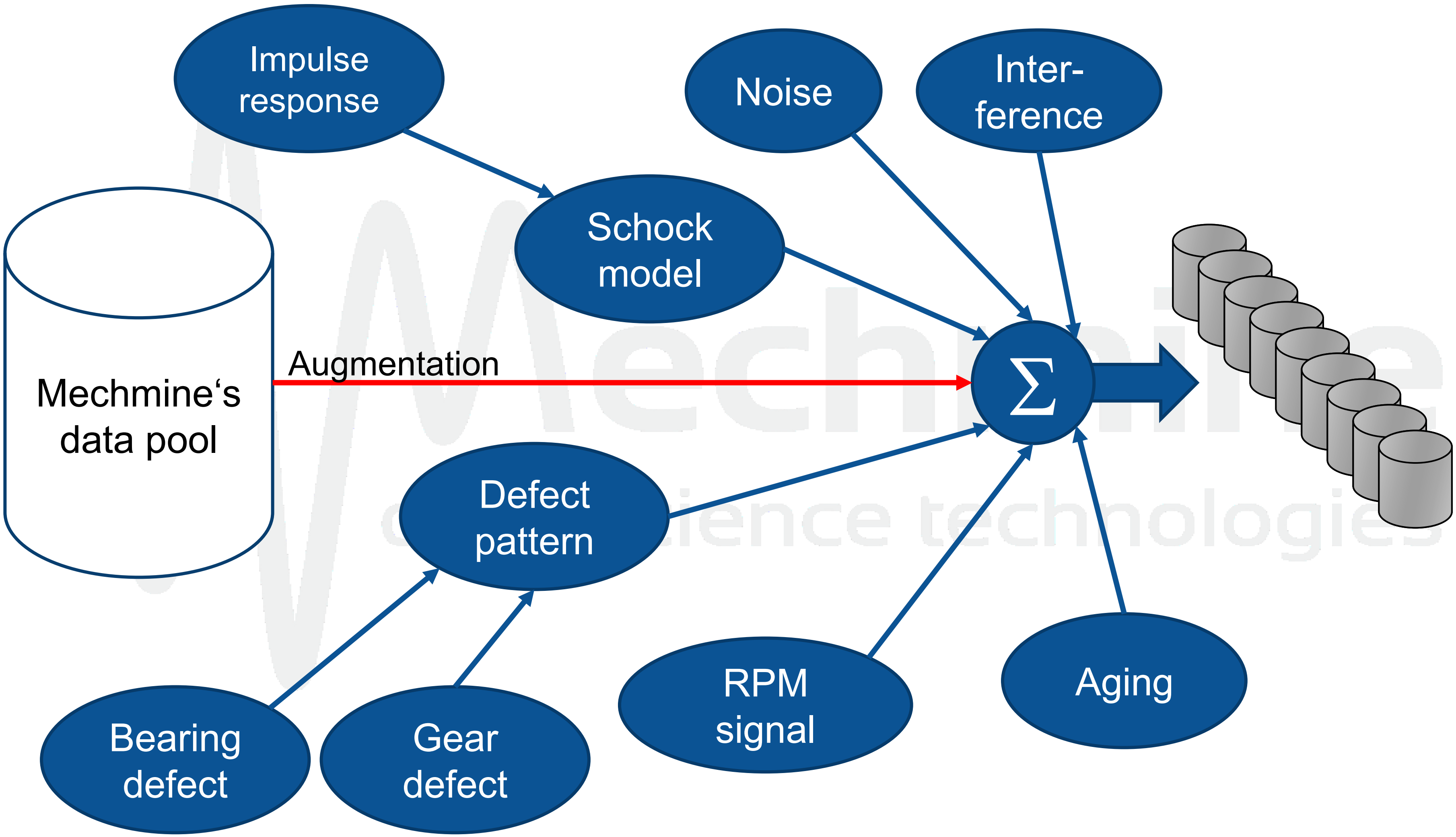
We destroy bearings, 8xSKF 6319, through overload under controlled conditions and in a realistic environment, a hydro power station.
The data shows the slow deterioration of a bearing: reference lifetime data.

Use of augmented data
Defect characterisation TB1



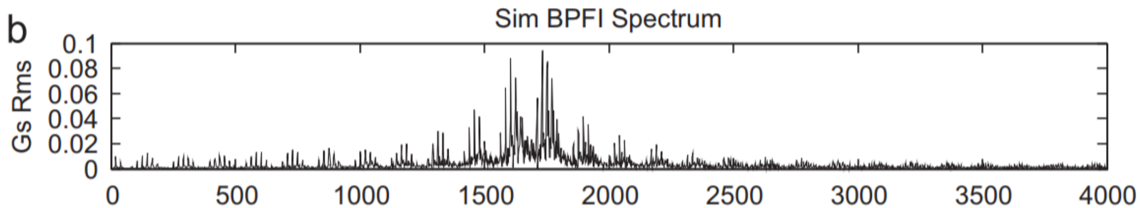
We destroy bearings by inserting a well defined defect and measure the resulting vibration, to be added to our library of bad data pattern.
The data show a typical defect pattern of a bearing: defect characterisation.

Use of augmented data
Data generation approach

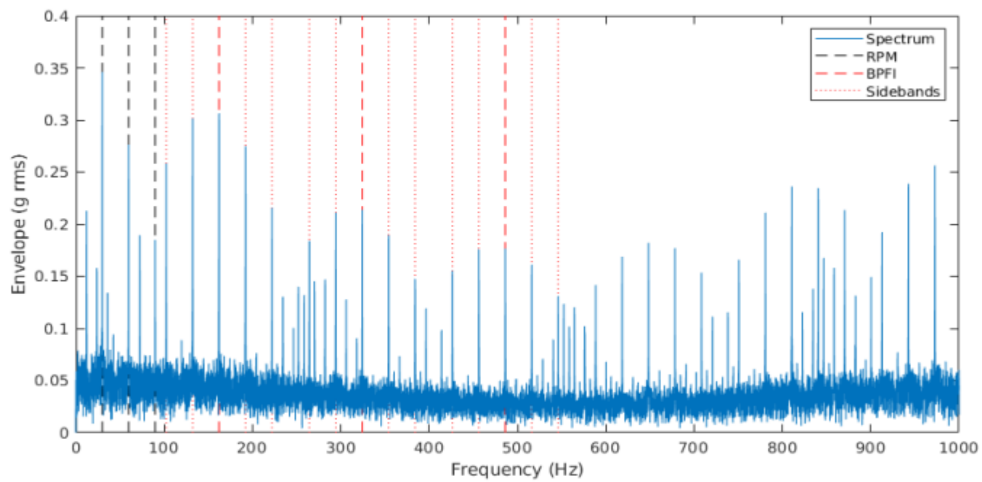


Use of augmented data Testing

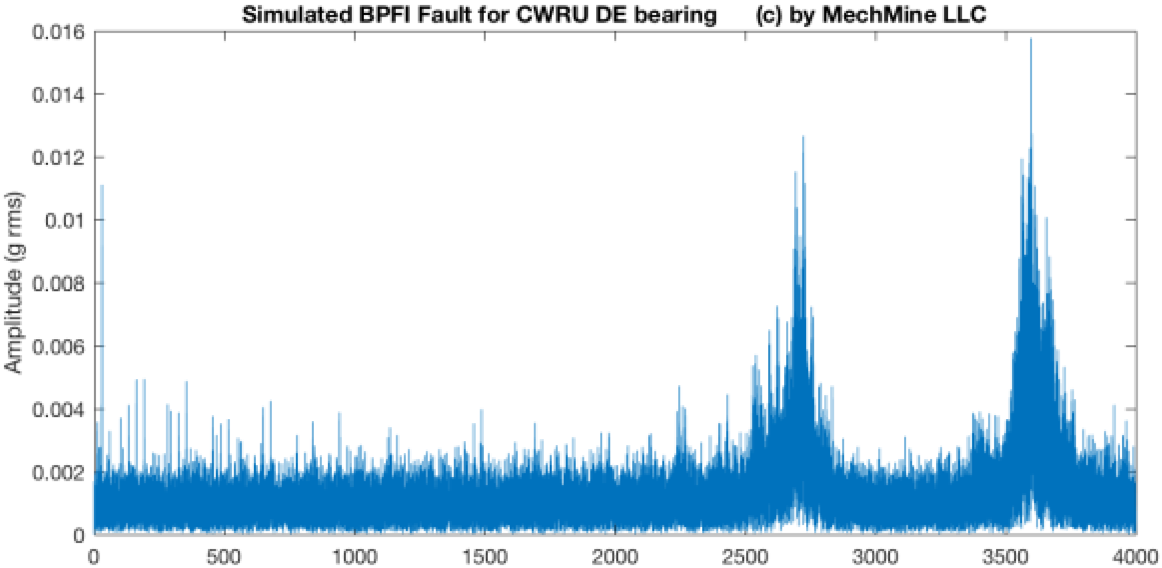
- How do we test the augmented data quality
  - Classifier: trained to discriminate real/synthetic but fails
  - Analysis through many experts
- Example of synthetic data (Gryllias 2012)



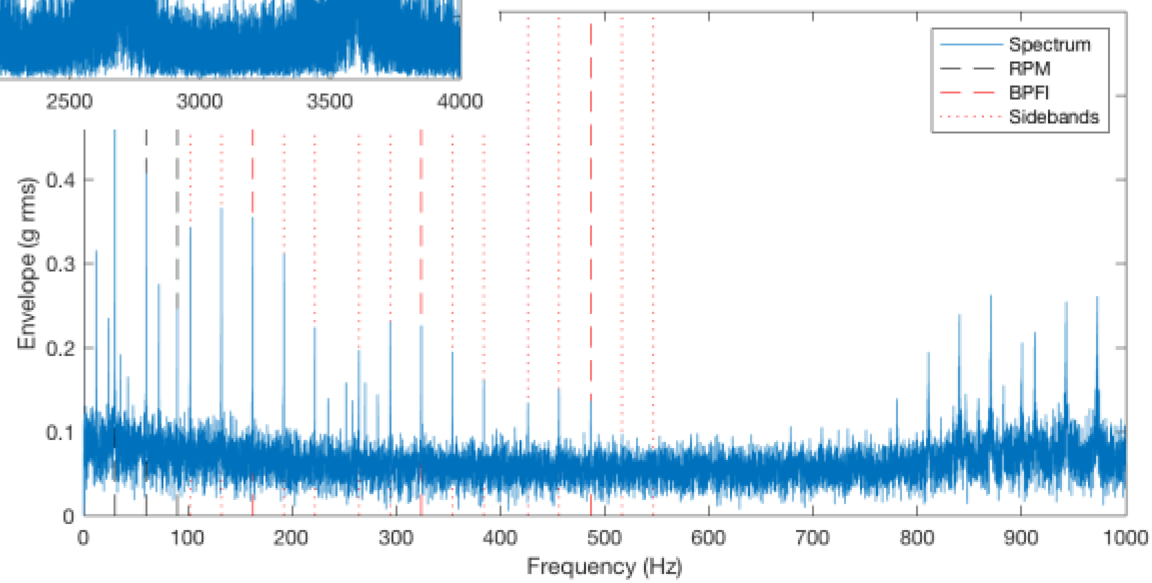
- Exemplary augmented data by Mechmine



Use of augmented data
Data example CWRU



Augmented data by Mechmine: for bearing-type (i.e. system) as used by CWRU experiment, rawdata and its spectrum.



Summary & conclusion

- ML relies on large and applicable data collections
- Collecting data is costly and time consuming
- Never trust the annotator

- Data can be generated by synthetic means
- Established models are too trivial for reality
- Data augmentation is a way out of the data gap
- The data generation model is very critical
- Produce synthetic data derived from real data