Introduction

Author:	Paul O'Leary
Date:	6. July 2022
Document number:	CoA2022-MLSC
Presentation type:	Lecture

Chair of Automation
Department Product Engineering, University of Leoben
Peter Tunner Straße 27, 8700 Leoben, Austria

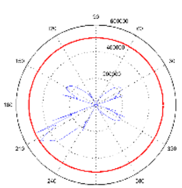
Phone: +43/(0)3842/402-5301
fax: +43/(0)3842/402-5302
email: automation@unileoben.ac.at
web: automation.unileoben.ac.at

Digitalization and CPS at the Chair of Automation

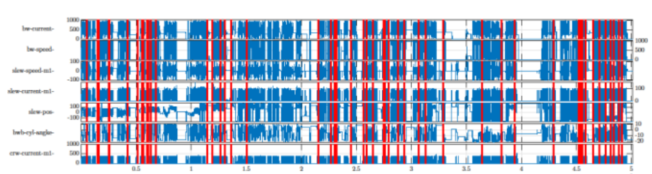
Recent and current projects

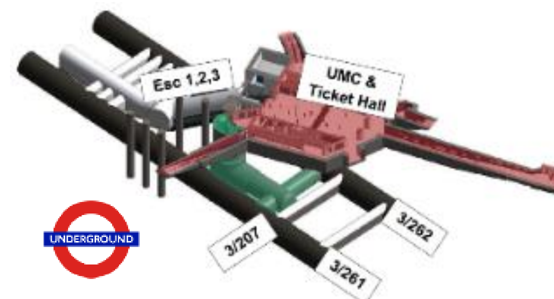
Three selected examples



Bucket wheel reclaimer: life time optimization



Bucket wheel excavator, incident analysis.



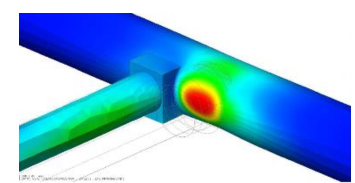
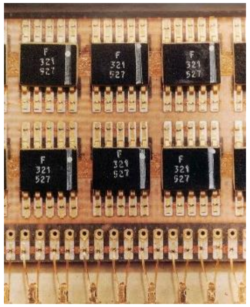


Figure 36: Total displacements of platform tunnel lining after invert excavation

Tunnelling: real-time comparison of observation and simulation.

50 Years of Computing Technology



Apollo Guidance Computer

2,800 ICs

36 K-words memory (16 Bit)

11.72 micro-seconds, (approx. 100 KHz).

16. July 1969

Nexus

CPU: Octa-core, 1.5GHz

GPU: Adreno 430

128 GB Memory

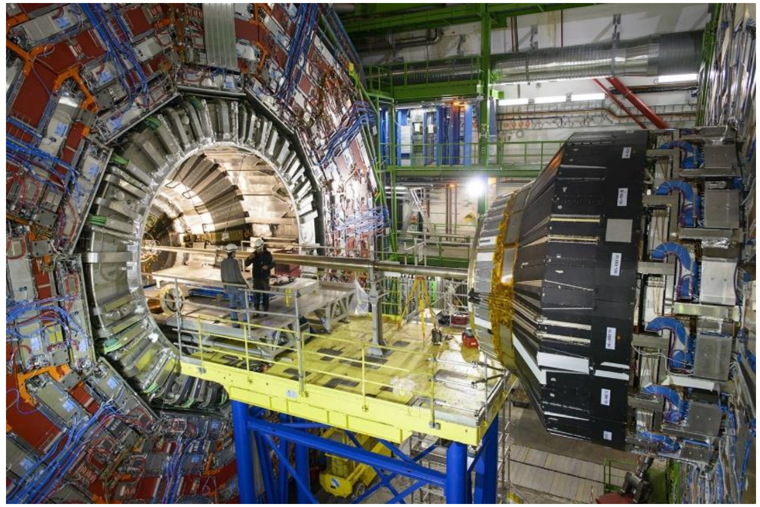
Comparison: 1.000.000 times more computing power

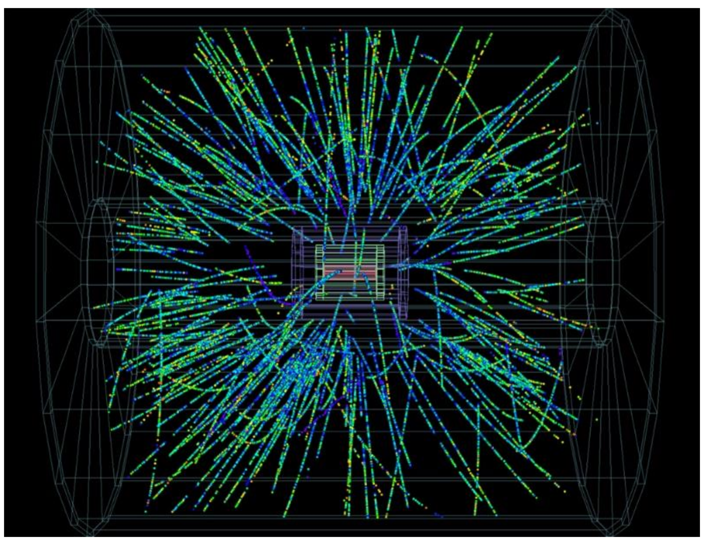
2.000.000 times more memory, etc. etc.

It's not a question of technology
It's what we do with the technology

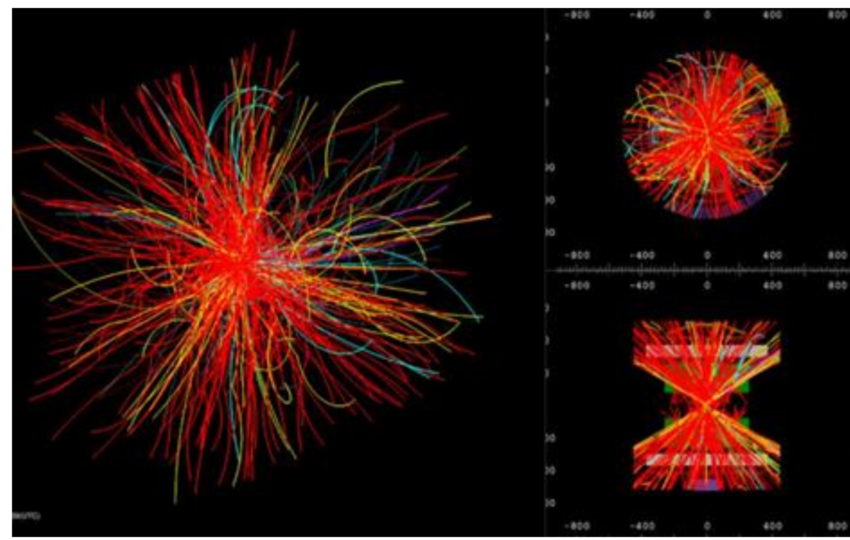
Paul.oleary@unileoben.ac.at

Large Hadron Collider





Atlas



Alice

The most succinct and pertinent definition of a cyber physical system is:

"A CPS is a system with a coupling of the cyber aspects of computing and communications with the physical aspects of dynamics and engineering **that must abide by the laws of physics**. This includes sensor networks, real-time and hybrid systems."

Consequently:

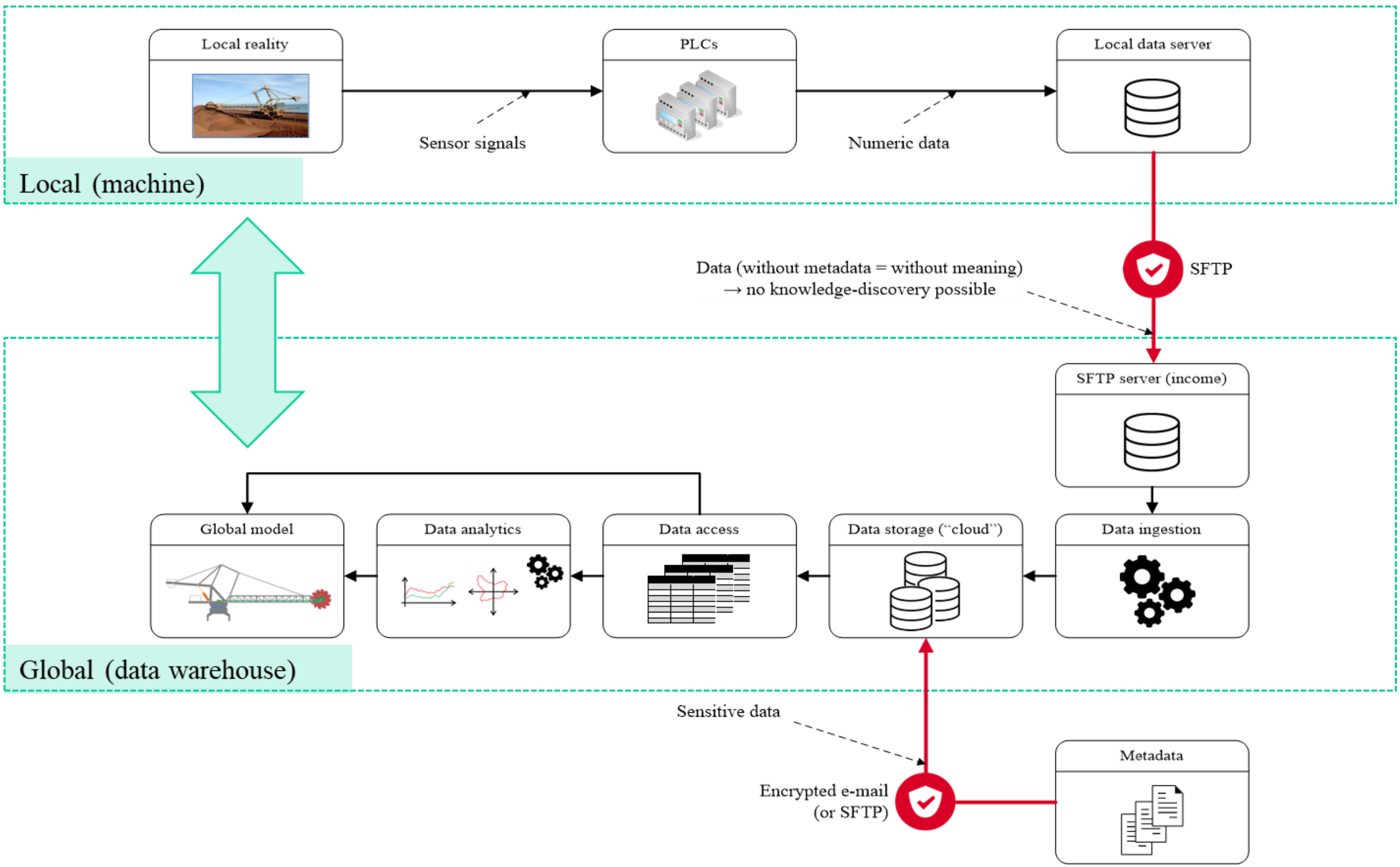
- Mere **correlation** in the sensor data is an inadequate measure of significance.
- System models and their inverse solutions are required if sensor data analytics is to infer knowledge with respect to **causality**.

System Identification, Embedded Simulation, Inverse problem to identify cause.

- No causality → No semantics. No semantics → No knowledge.

© 2015 Chair of Automation
http://automation.unileoben.ac.at

Digital Twin



Economic Justification

- **Incident analysis**

  Given an incident with the plant or machinery we can investigate the cause. This is important for both liability and guarantee reasons. Financially this has proved to be one of the most important issues.

- **Commissioning support:**

  shortening the time to start up complex plant and machinery

- **Automatic operations recognition:**

  Identifying incorrect operation, comparison of operation performance between operators

- **Operational efficiency optimization:**

  Finding invisible lost time

- **Logistics optimization**

  The logistics have an effect on machine life-times

- **Fleet management**

  Comparing the performance of multiple pieces of equipment

- **Plant condition monitoring**

  System identification to determine changes in system behaviour

- **Material condition monitoring**

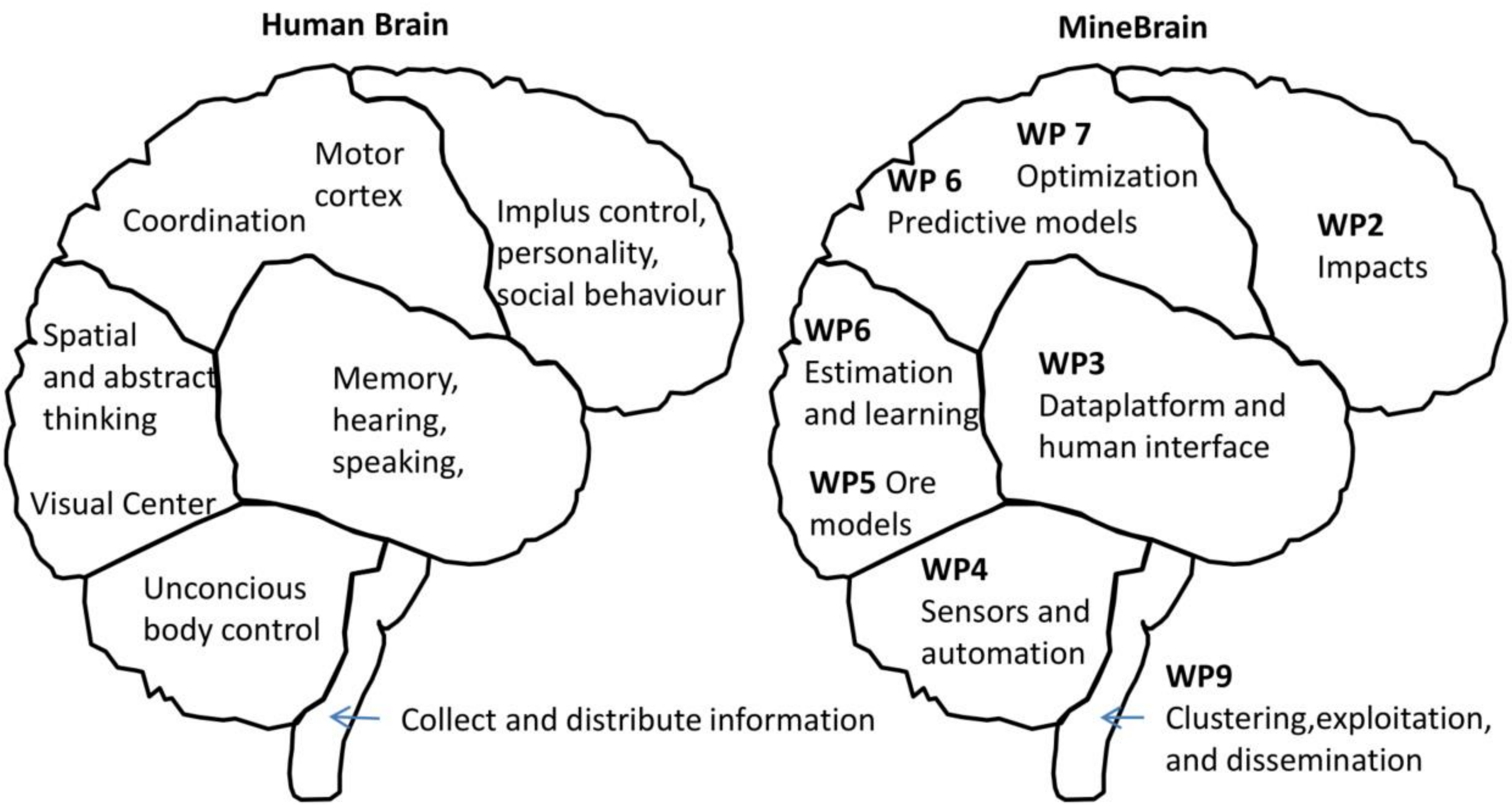
  With established models effects of material condition can be separated from machine condition

- **Engineering feedback**

  Possible improvements in machine design can be identified, important for the next generation.

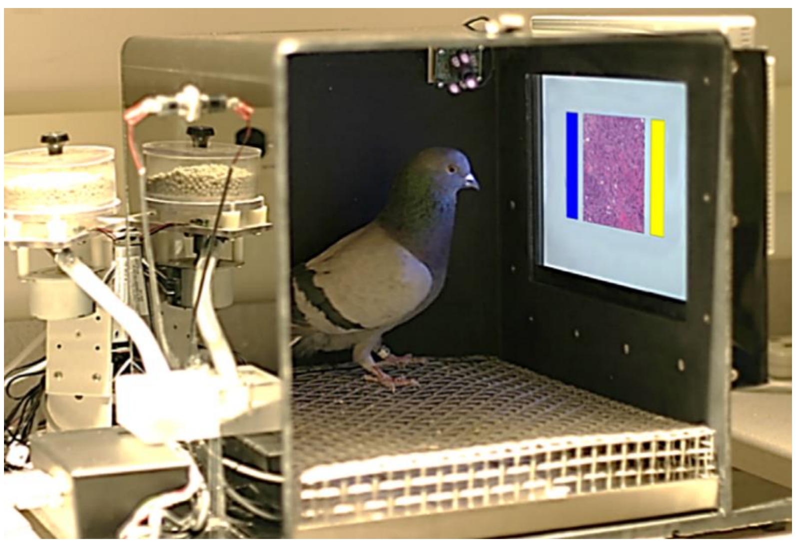
Paul.oleary@unileoben.ac.at

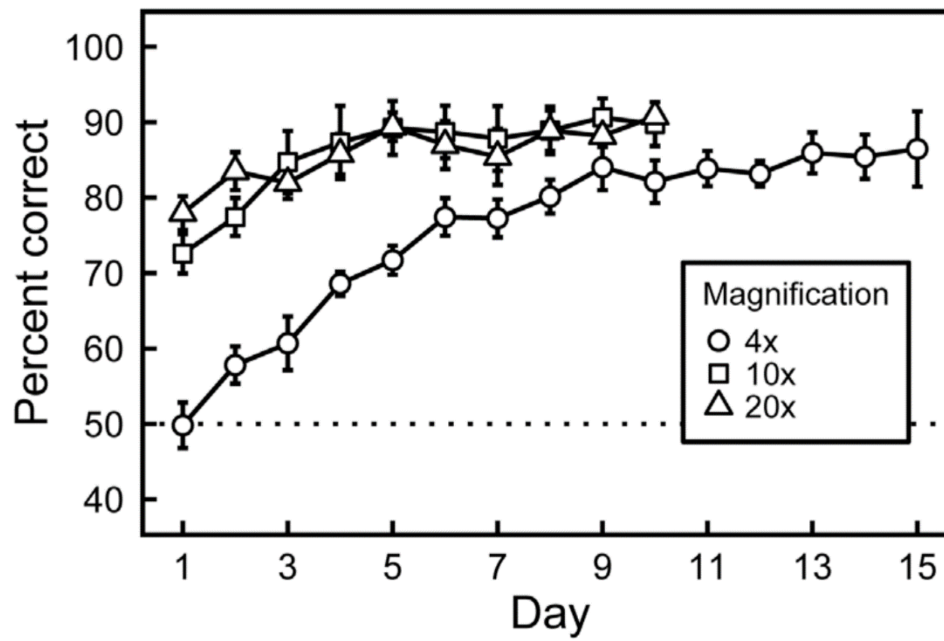
Proposal for the “Mine Brain” project



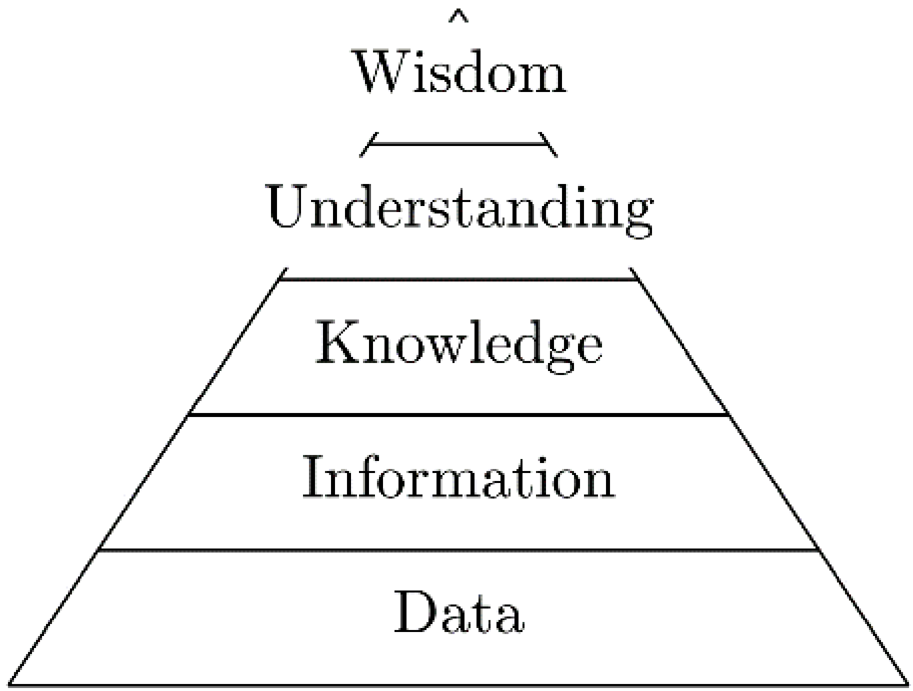
Pigeons (*Columba livia*) as Trainable Observers of Pathology and Radiology Breast Cancer Images

Richard M. Levenson[1]*, Elizabeth A. Krupinski[3], Victor M. Navarro[2], Edward A. Wasserman[2]*





What exists and what do we know?



Ontology
What exists and which relationships are there?

Epistemology
What do we know? how do we know it? And what are the limitations of knowing?

„What is a valid source of knowledge?“

Some ensuing questions:

- What is belief?
- What is justified belief and what justifies justified?
- What is the difference between justified belief, knowledge and understanding?
- How do we know that we know?

Paul.oleary@unileoben.ac.at

„What is a valid source of knowledge?“

European Phenomonology as a possible answer:

- Edmund Husserl (April 8, 1859 – April 27, 1938)
  “… experience is the source of all knowledge…”

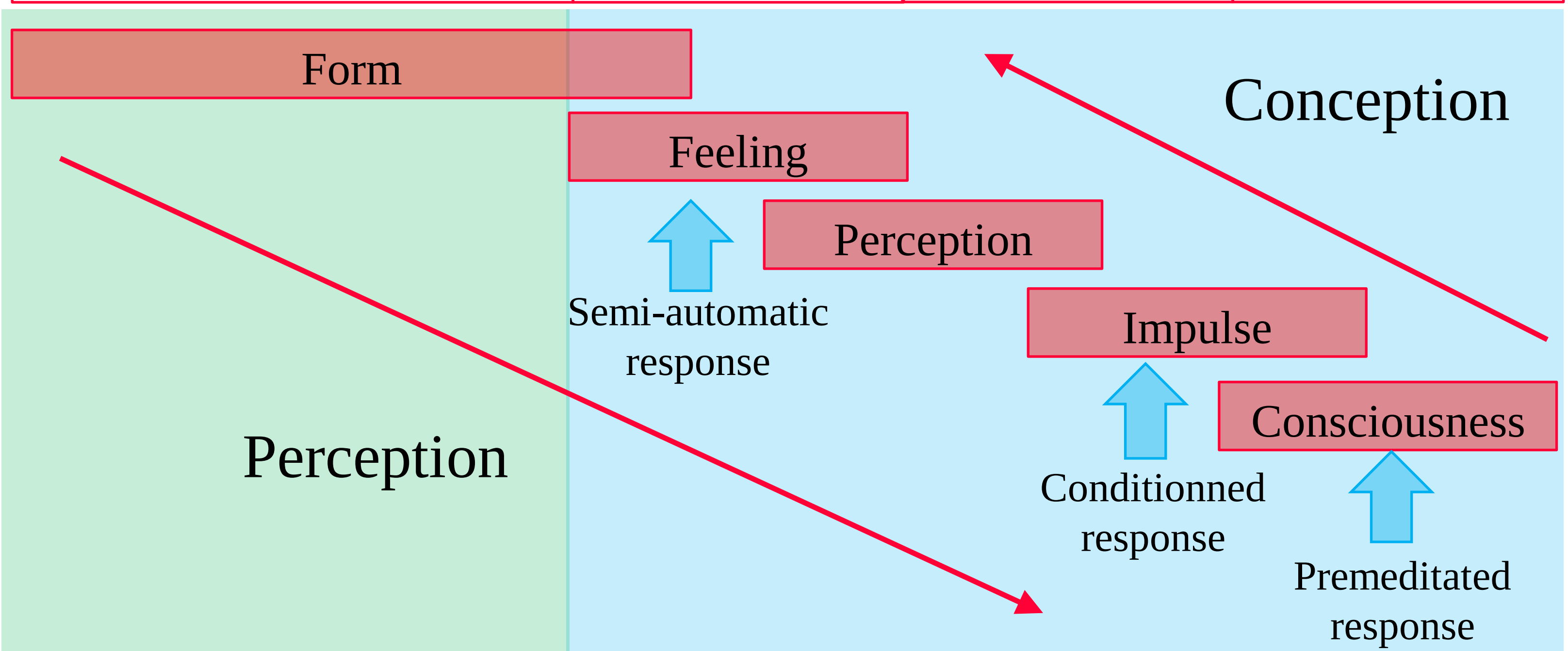
- Martin Heidegger (September 26, 1889 – May 26, 1976)
  “… the things in lived experience always have more to them than what we can see…”.

  (Hidden models)

- Maurice Merleau-Ponty (14 March 1908 – 3 May 1961)
  “…first we experience and then we reflect….”

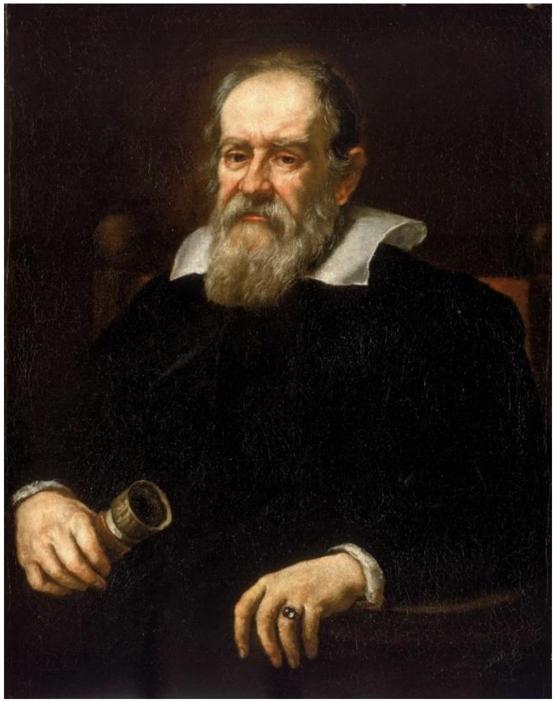
Conception and Perception

Sensory information	Mental models		
Sensory Base	Modelling sensory experience	Self referencing of consequence	Models for past experience
Vijñānas 1..5	Mano-Vijñāna	Manas-Vijñāna	Ālāya Vijñāna

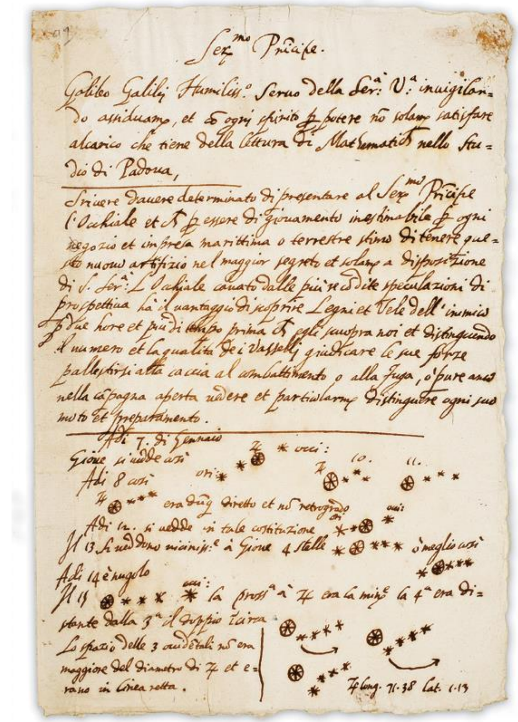


Paul.oleary@unileoben.ac.at

Observational science



Born in 1564, Galileo Galilei's observations of our solar system and the Milky Way revolutionized the understanding the Universe.

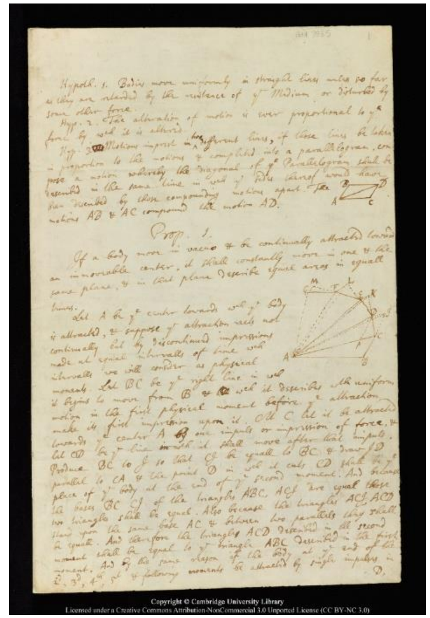

Ser.mo Principe.

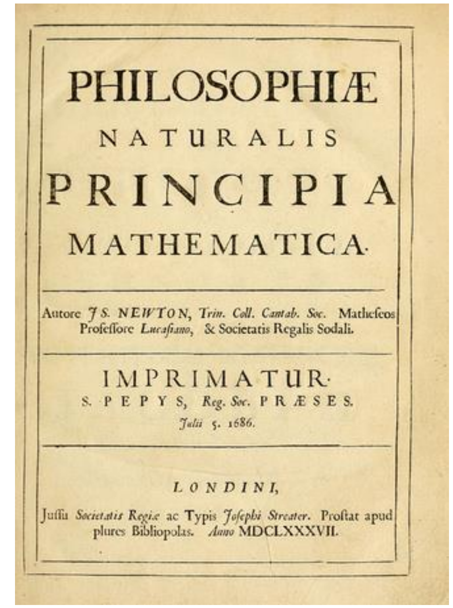
Galileo Galilei Humiliss.o Servo della Ser.tà V.a invigilan=do assiduam.te, et con ogni spirito per potere non solam.te satisfare al carico che tiene della Lettura di Matematiche nello Studio di Padova,

Draft of a letter to Leonardo Donato, Doge of Venice, August, 1609, and Notes on the Moons of Jupiter, January 1610. Credit: University of Michigan Special Collections Library

Isaac Newton (20. Mar. 1726 - 31. Mar. 1727)

Mathematical models of observation (Physical laws)





PHILOSOPHIÆ
NATURALIS
PRINCIPIA
MATHEMATICA.

Autore *J S. NEWTON*, *Trin. Coll. Cantab. Soc.* Matheseos Professore *Lucasiano*, & Societatis Regalis Sodali.

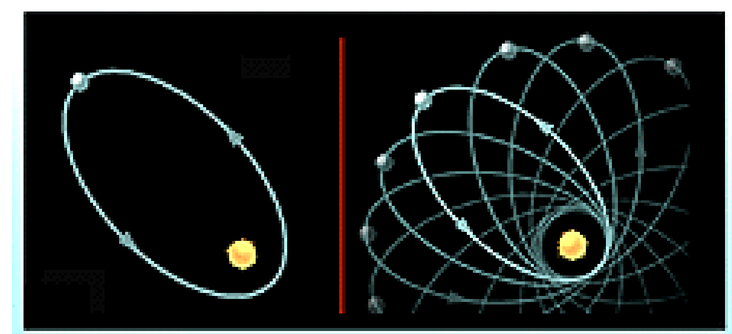
IMPRIMATUR.
S. PEPYS, *Reg. Soc.* PRÆSES.
Julii 5. 1686.

LONDINI,

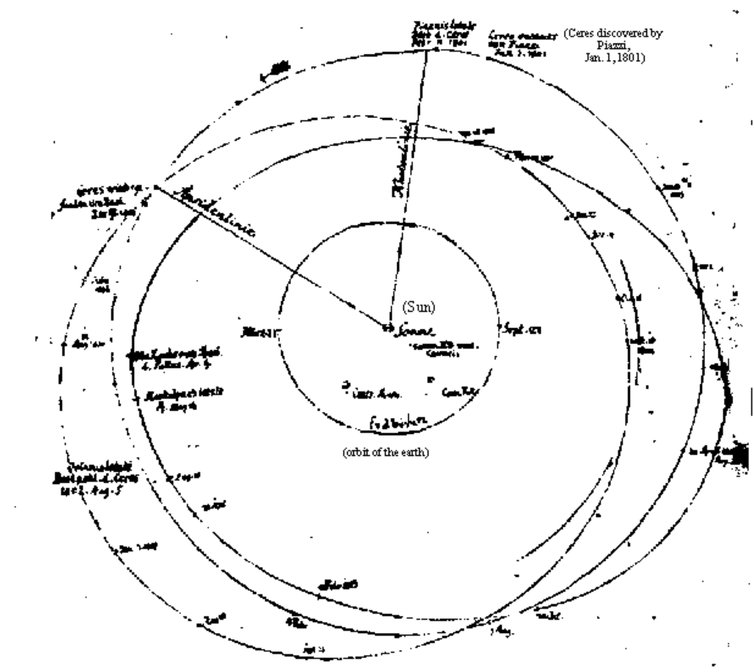
Jussu Societatis Regiæ ac Typis *Josephi Streater*. Prostat apud plures Bibliopolas. *Anno* MDCLXXXVII.

$$F_g = G\frac{m_1 m_2}{r^2}$$

MERCURY'S ORBIT



Mathematical science of systems and uncertainty

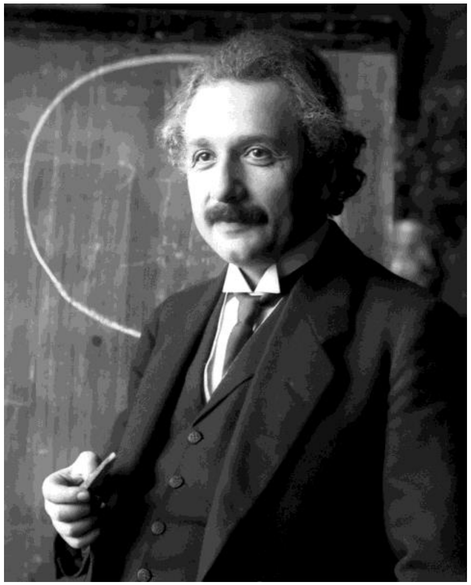


Sketch of the orbits of Ceres and Pallas (nachlaß Gauß, Handb. 4). Courtesy of Universitätsbibliothek Göttingen.

Gauss developed the statistical methods of least squares to determine the orbit of Ceres, which was first observed by Galilei.

Albert Einstein (14. Mar. 1879 - 18. Apr. 1955)

Connectional paradigm change emerging from abstract thinking



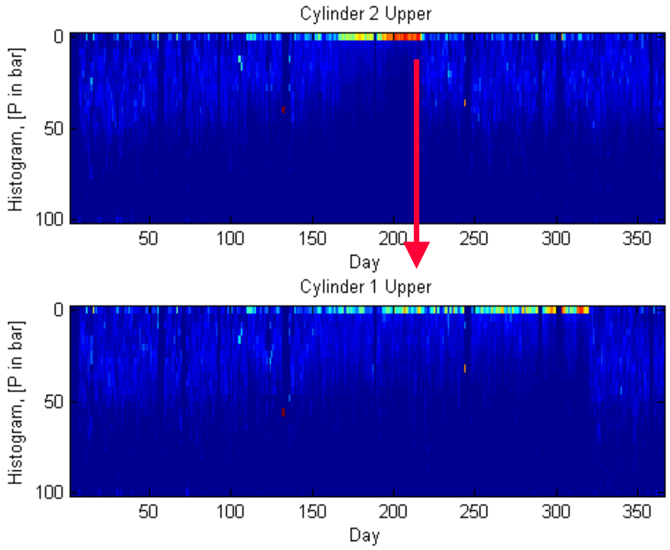
$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

A relationship between gravity and **time**-space.

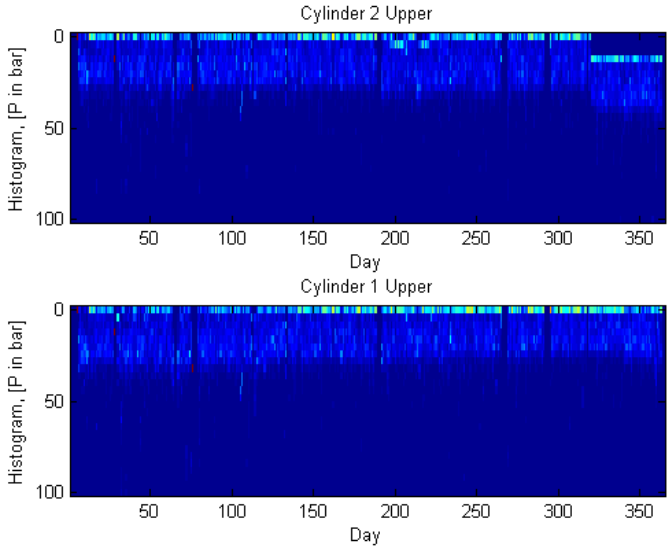
Cape Lambert: an Exemplary Case

One year of operating data for the hydraulics of the two ship loaders

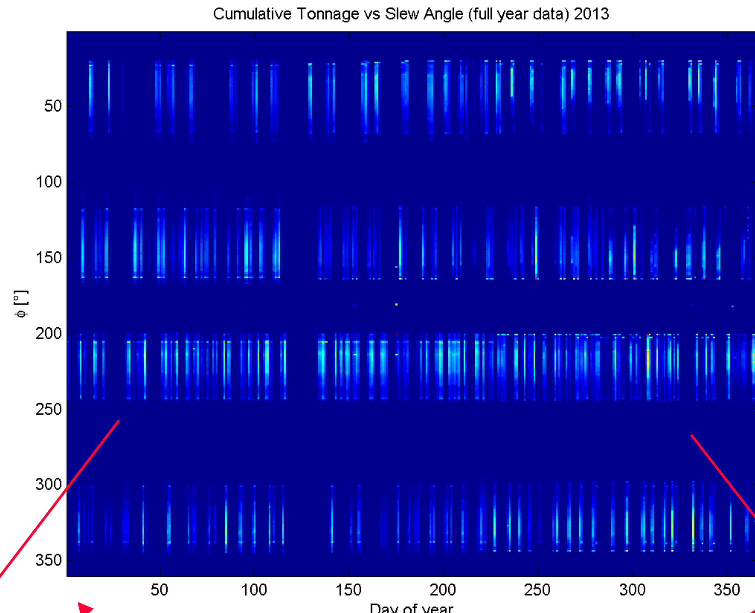
Ship loader 1



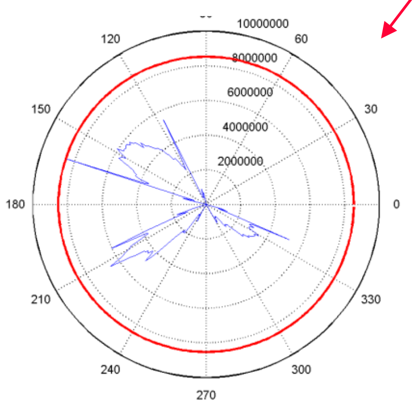
Ship loader 2

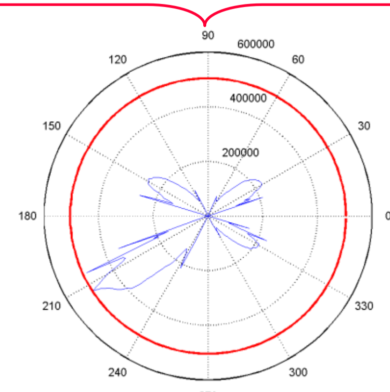


From a Years Overview to Slew-Bearing Loading on a Specific Day



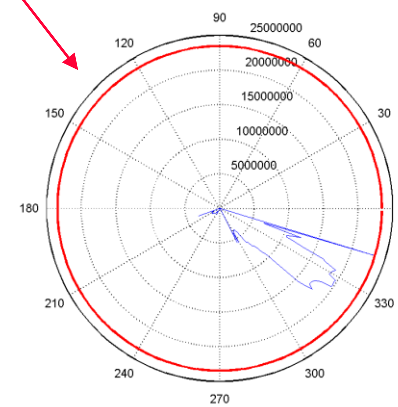
Data for 2013-01-12





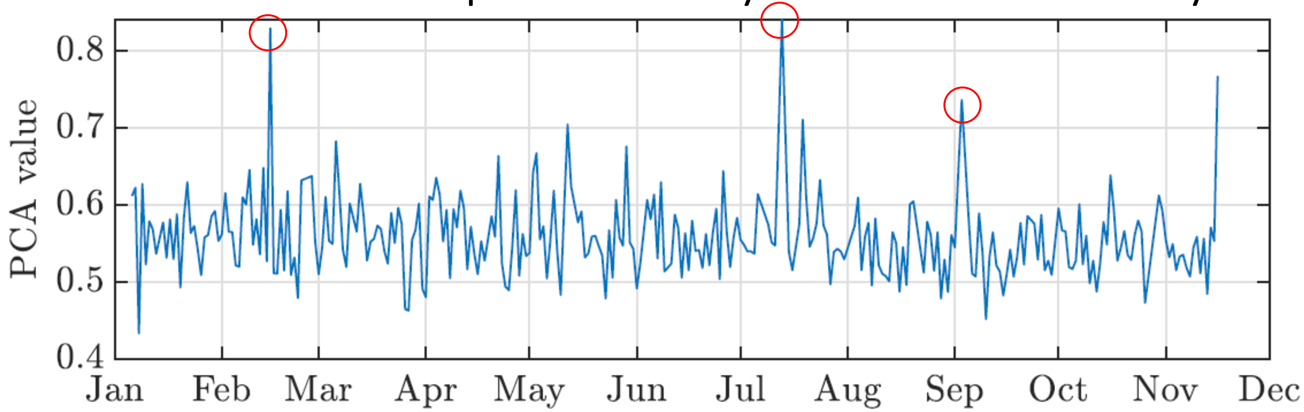
Full Year Summary

Data for 2013-12-11

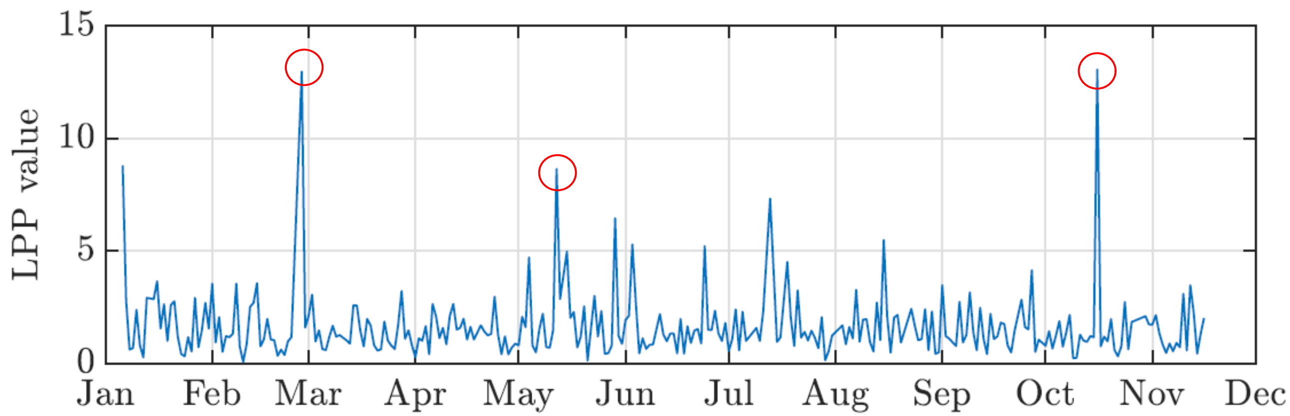


D4.2.3: Example – PCA and LPP

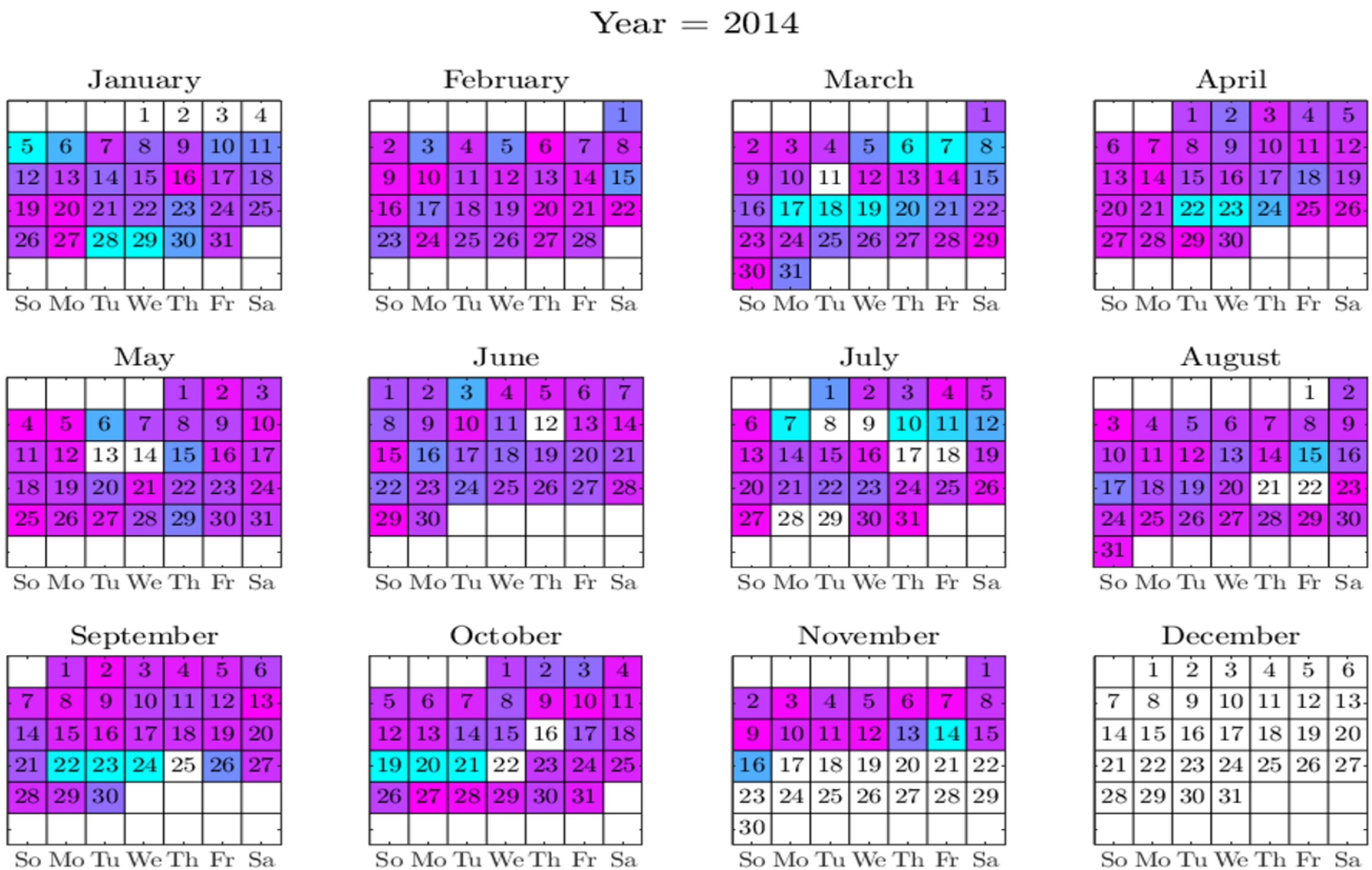
- For the data on the server.
- PCA and LPP were performed to the array of summary tables.
- PCA can be used to represent days of lower activity.



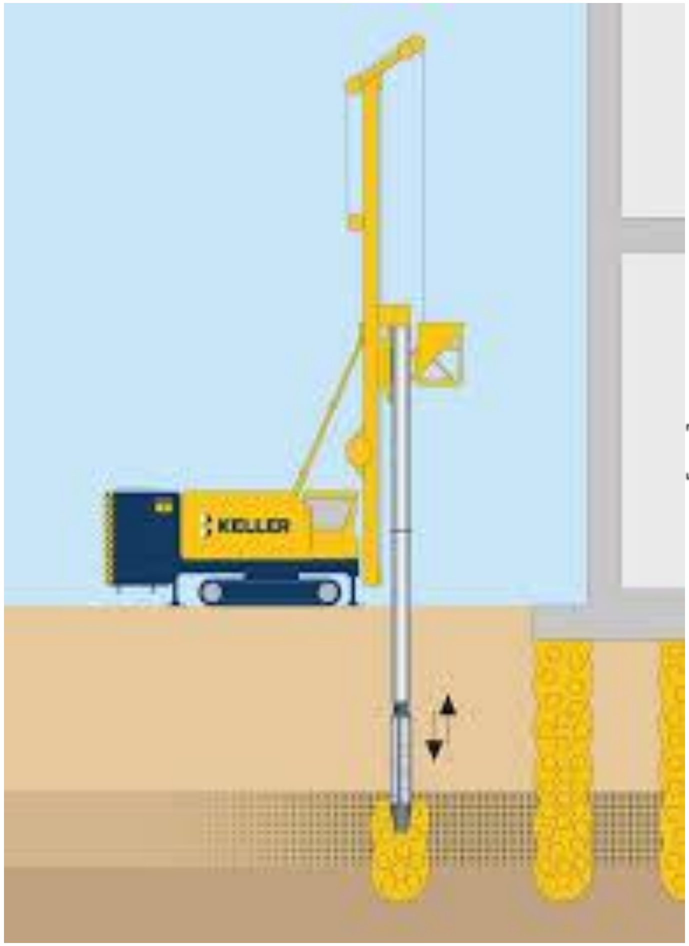
- LPP can be used to represent significant changes in the data.

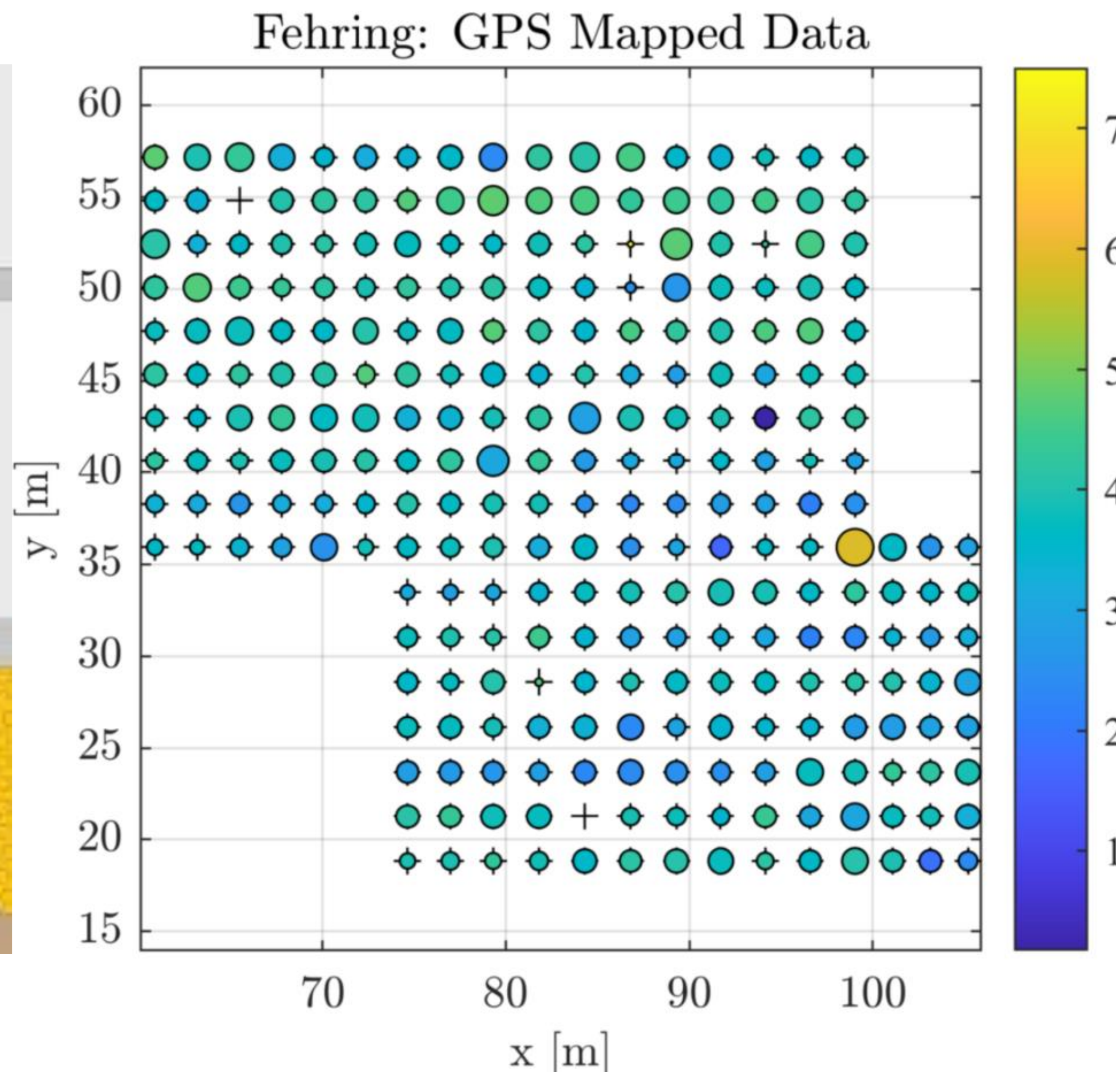


D4.3.1: Visualization of CCA results

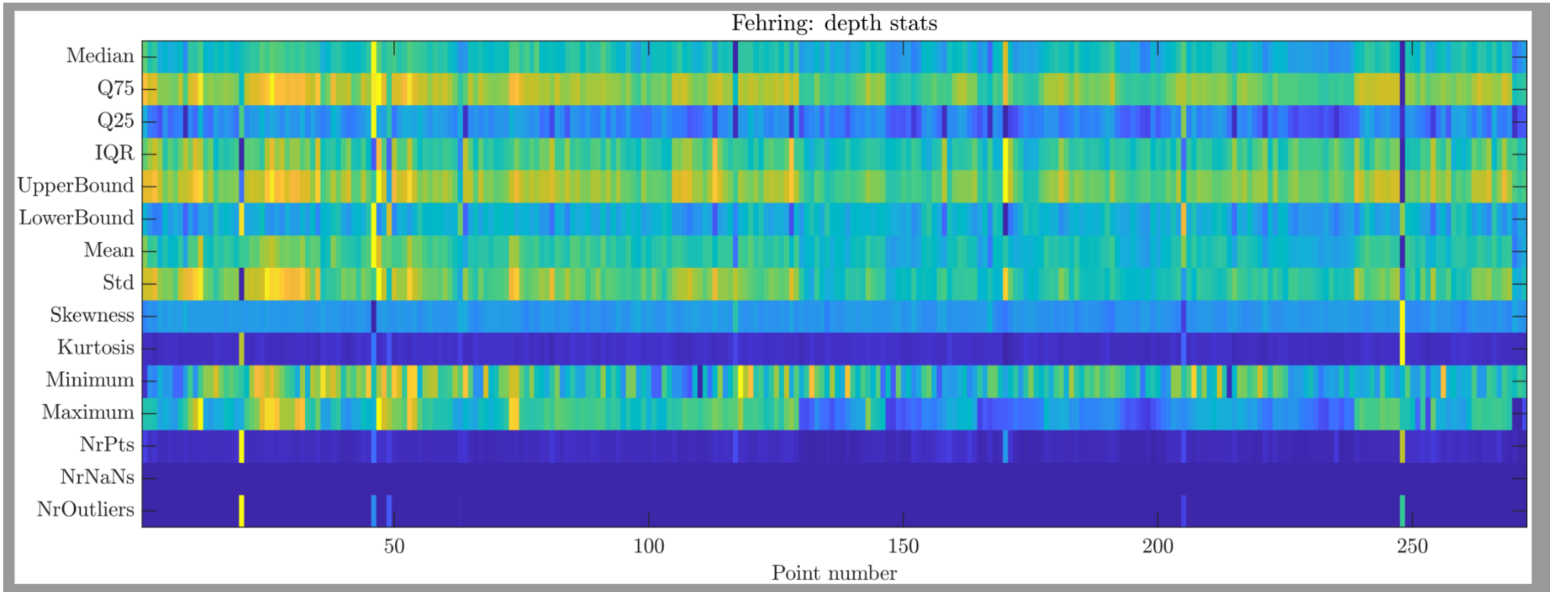


Location Mapped Data





Anomaly Heat Maps



Justified belief or knowledge?

The Universe

- 5% “normal matter”, i.e., the rest - everything on Earth, everything ever observed, everything we claim to know anything about!
- 27%. dark matter.
- 68% dark energy.

Abell 520: observation seems to indicate the current theories of dark matter are not correct.

Hybrid Machine Learning for Anomaly Detection in Industrial Time-Series Measurement Data

Authors:	Anika Terbuch, Paul O‘Leary, Peter Auer
Presenter:	Paul O’Leary

Date:	6 July 2022
Conference:	Summer School – Data Science hub

Chair of Automation
Department Product Engineering, University of Leoben
Peter Tunner Straße 27, 8700 Leoben, Austria

Phone: +43/(0)3842/402-5301
fax: +43/(0)3842/402-5302
email: automation@unileoben.ac.at
web: automation.unileoben.ac.at

Chair of Automation

Motivation

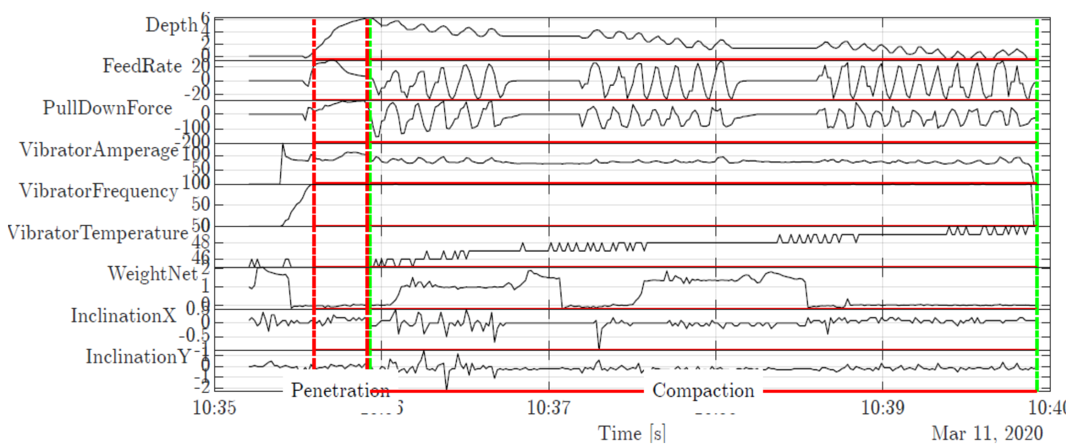
- Improving the detection of anomalies in multivariate time-series (MVTS)
- Evaluating real-time machine data in safety relevant applications
- Developing a generic framework which combines the strengths of KPI-classification and machine learning
  - Machine learning as augmentation

anika.terbuch@unileoben.ac.at

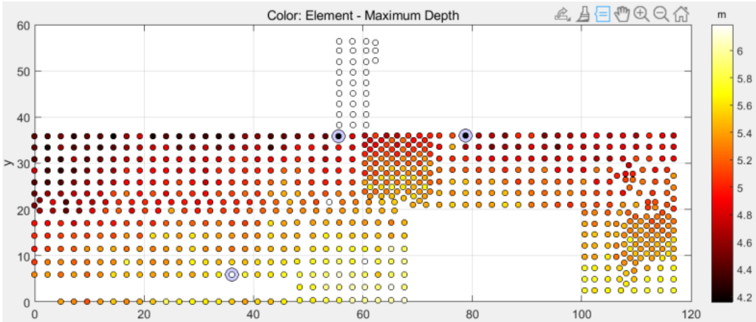
Case Study

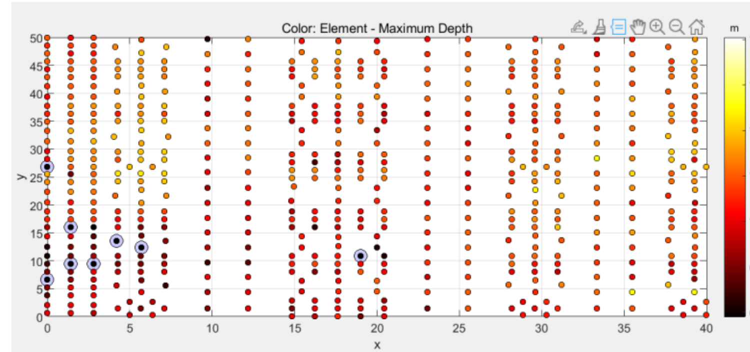
- Vibro replacement ground improvement process
  - Monitored with sensors with n_s= 16 channels
  - Non-uniform sub-surface conditions from site to site
  - Varying creation time for a single column t_c= 6 … 30 min
  - Varying number of MVTS per site m = 400 … 1000 MVTS
  - Manually performed process

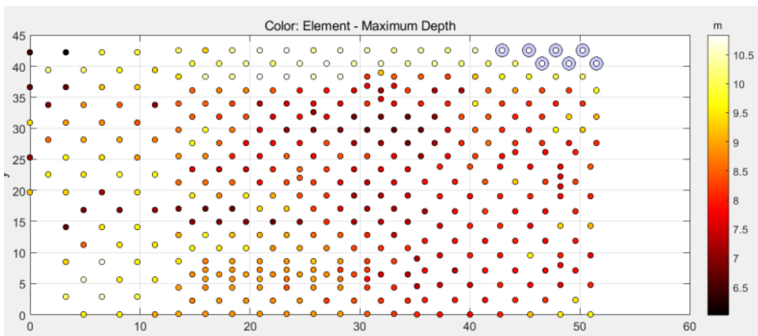
- For each foundation column a MVTS is created

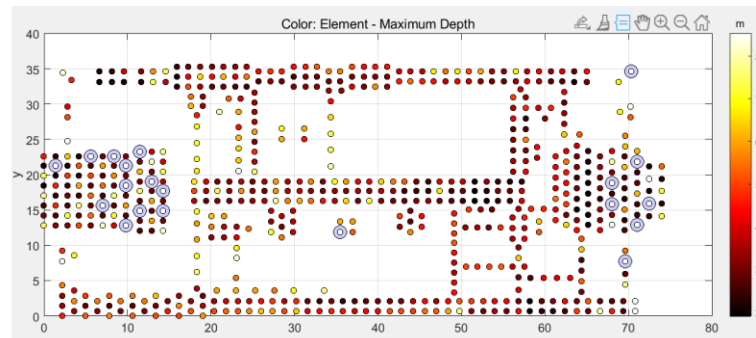


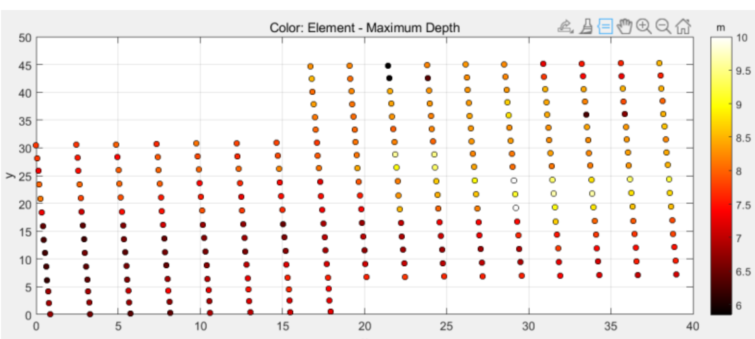
Geo-referenced KPI for multiple sites

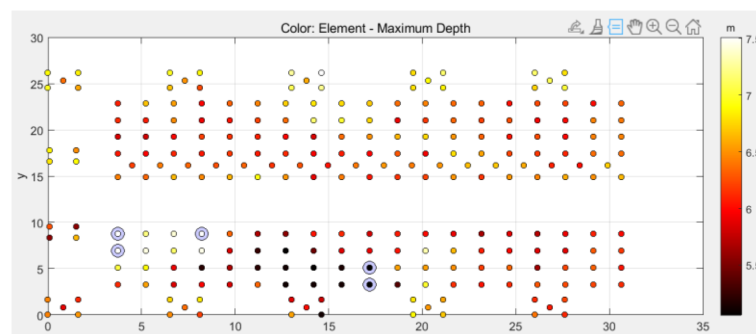






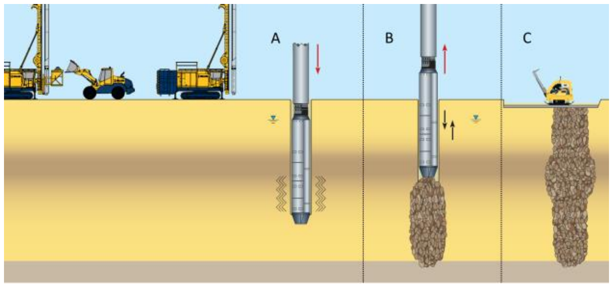




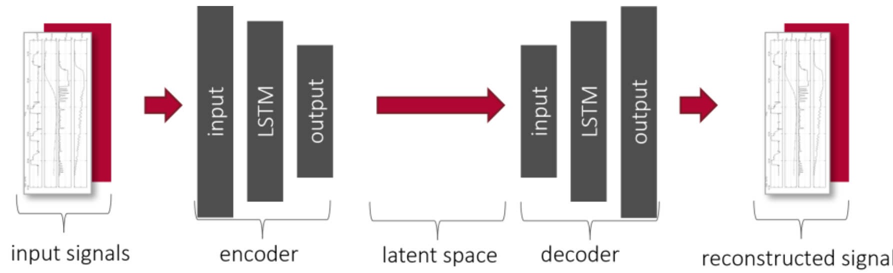


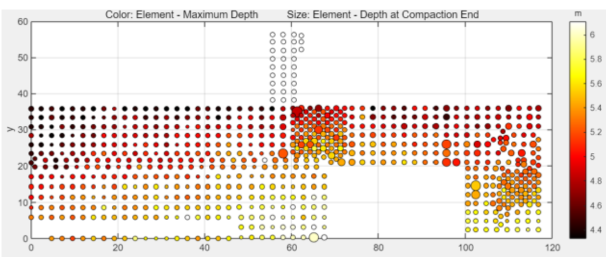
Overview of the Hybrid Classifier

Physical Process



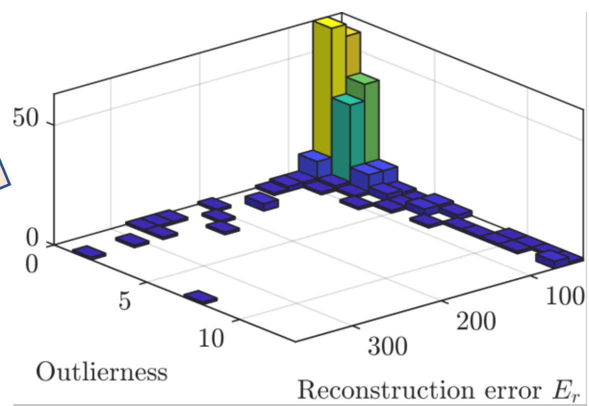
LSTM-VAE

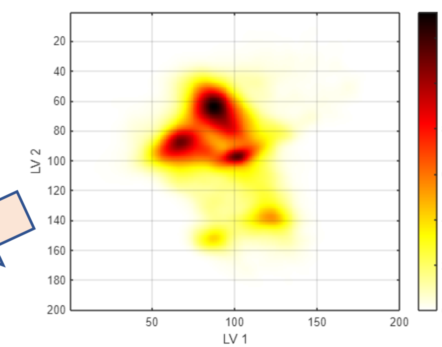




KPI based classification

Hybrid Classification

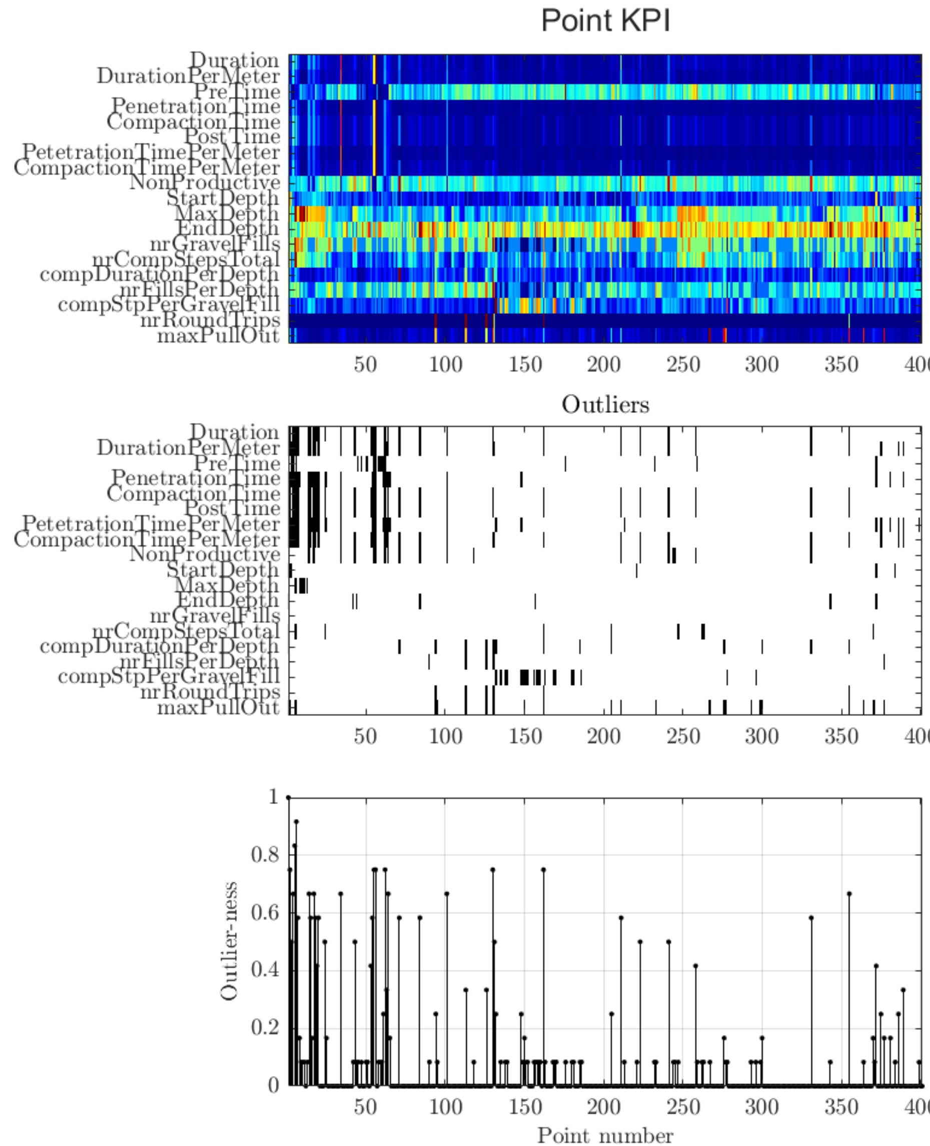




ML based classification

KPI-based classifier

- 49 KPI
- Outlier detection is performed for each sensor channel and each KPI based on IQR
- Outliers - values which are outside the lower bound b_l and the upper bound b_u
  - $b_u = q_{75} + 1.5\ IQR$
  - $b_l = q_{75} - 1.5\ IQR$



- “Outlierness”:
  - Artificial word
  - Quantification of the degree that a MVTS is an outlier

Data pipe line

Classical processing

Raw Real-time machine data

Data acquisition → Data ingestion → KIP Calculation

MVTS

MVTS, KPI, “outlierness” and queryable metadata.

Pre-processing

MVTS

Channel Selection → Rule based segmentation → Data trimming

ML-MVTS

ML Training

ML-MVTS

Deselect outliers → Random MVTS sample → Train bi-LSTM Autoencoder → Validation

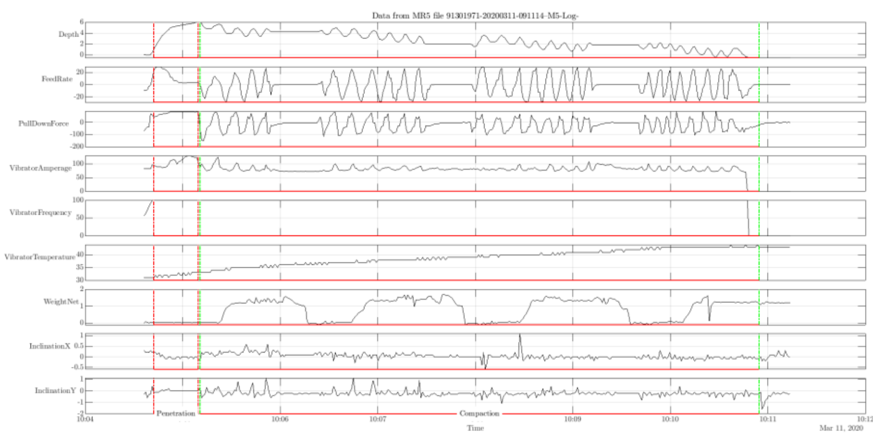
GA Hyperparameter optimization

Hyperparameters, weights, biases

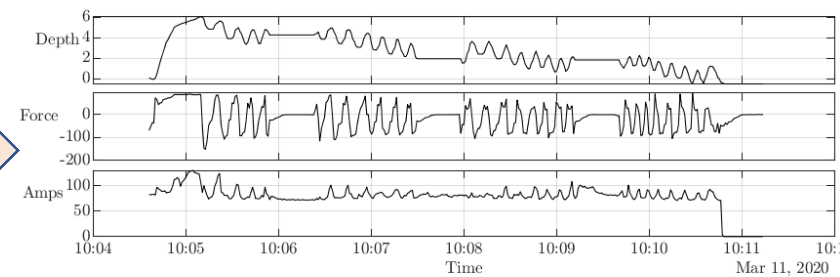
Step 1: Channel Selection for ML

- Different goals may require different selection of channels, e.g., column quality and production efficiency.
- Metadata-based channel selection.

Complete data record



Channels relevant for quality

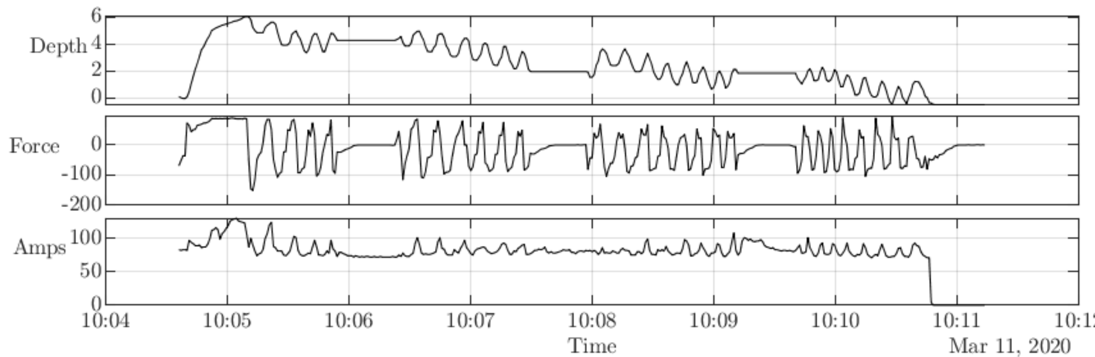


Note: More data channels, does not imply a better model.

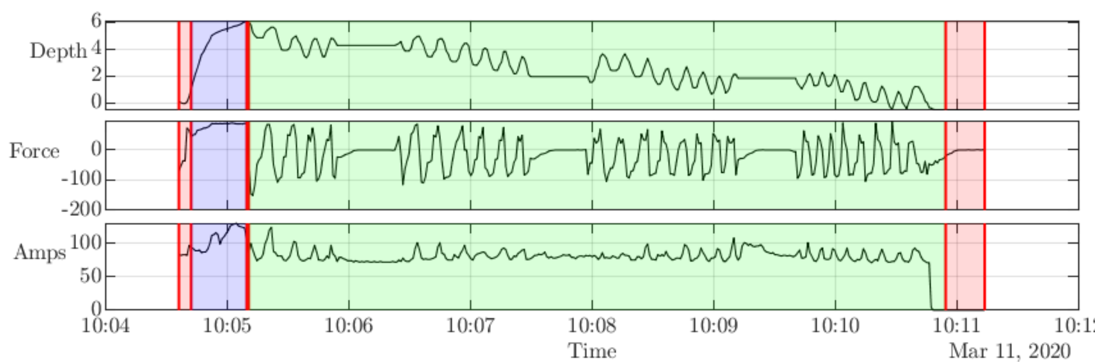
Originally we tested using all the data but the results were not as good as with selected channels.

Step 2: Rule based segmentation

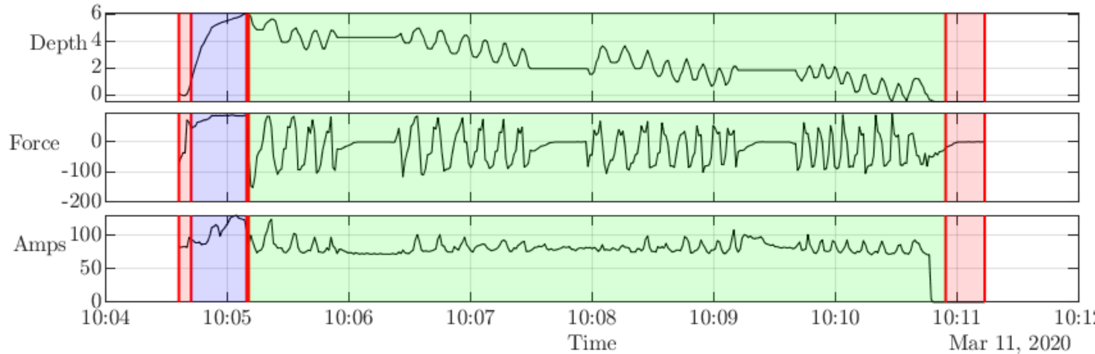
- Expert rules used to segment the data according to subprocesses.



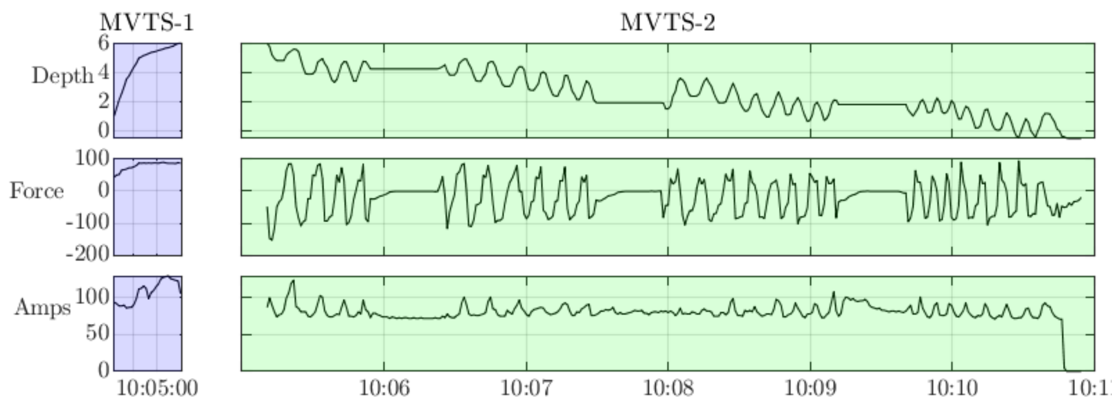
Rule based segmentation



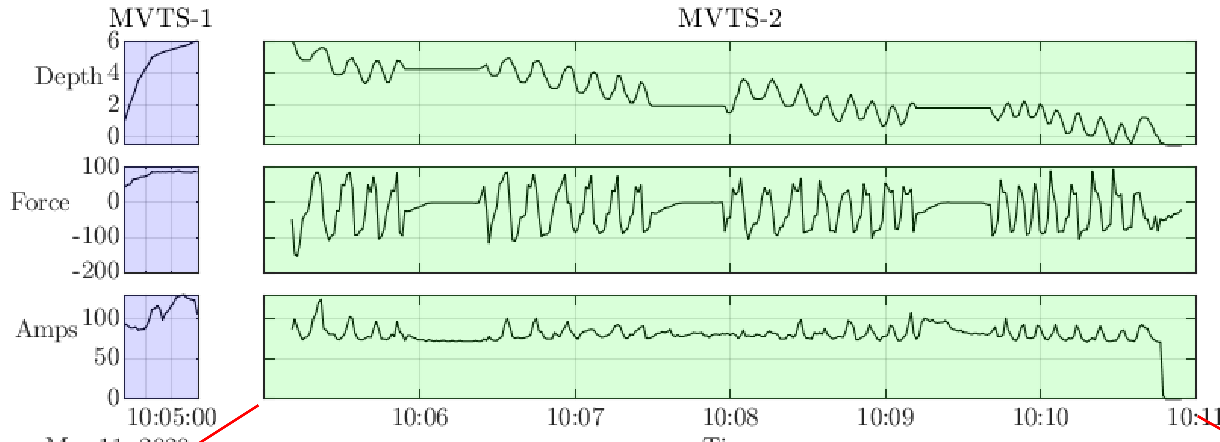
Step 3: Common data structures used for MVTS

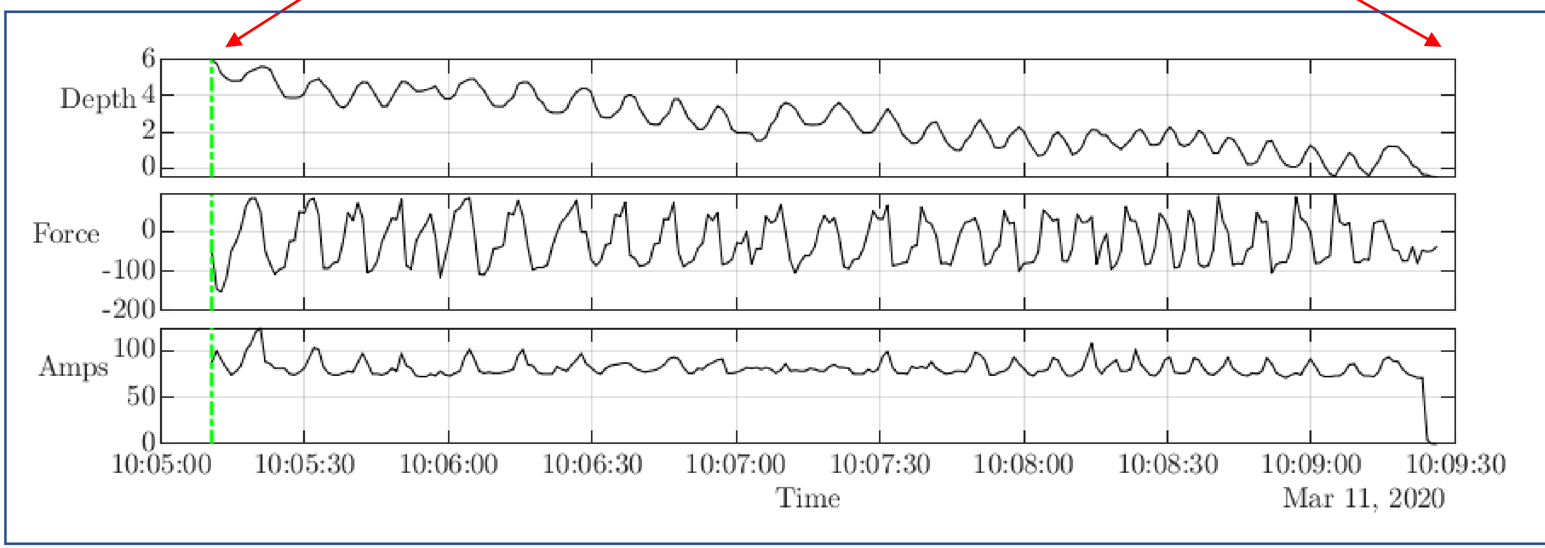


Map segments to standard objects



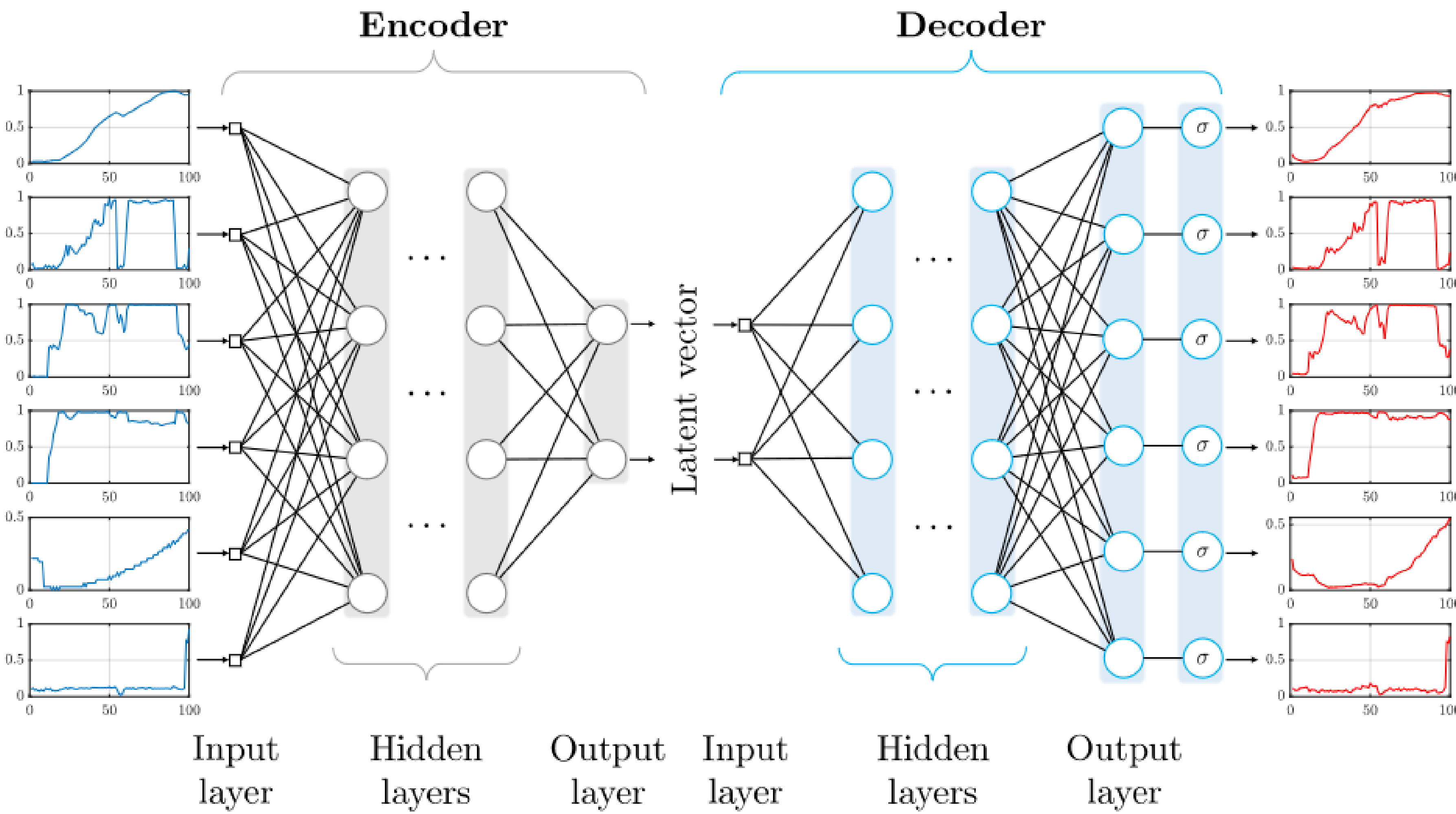
Step 4: Remove no activity (no change of state)





Result: An MVTS binary object for each column and a corresponding “outlierness” value via the KPI.

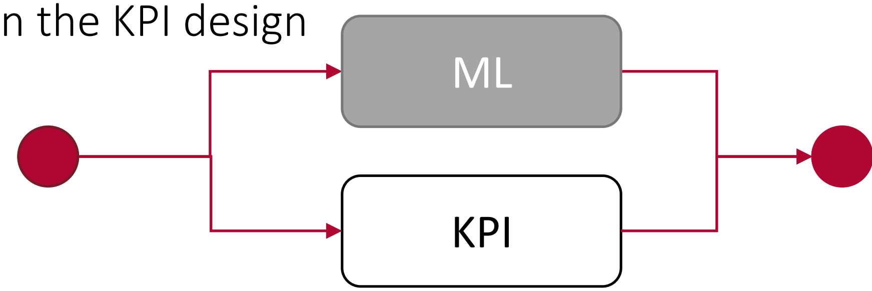
Basic Autoencoder



- Hybrid classifier consists of two parts:

1) KPI classifier
   - Automatic identification of systematic changes in the MVTS

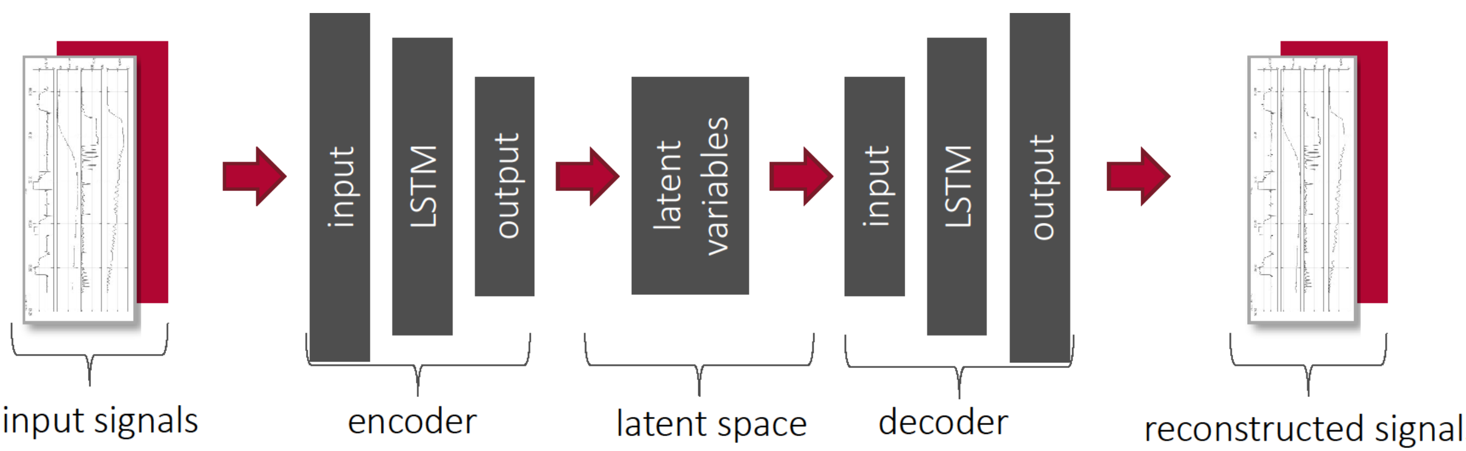
2) Unsupervised machine learning classifier
   - Detecting non-linear relationships between the sensor channels
   - Detecting anomalies which were not addressed by the KPI directly
     - Errors that were not considered in the KPI design



- Serial-Parallel hybrid model
  - Combination of the classifications (logical OR)

ML-based classifier

- LSTM-based variational autoencoder (LSTM-VAE)
  - Performs dimensionality reduction by encoding the input signals into lower dimensional latent variables
  - Variational autoencoder - generative model
    - Latent variables: Gaussian distributions with mean and standard deviation



Cost Function of the LSTM-VAE Classifier

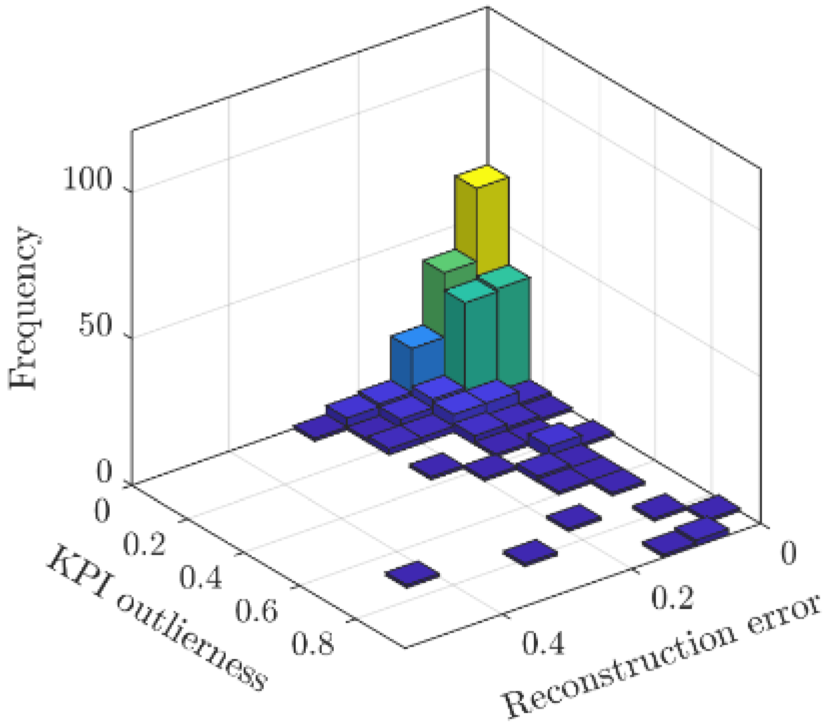
- D_{KL}: Kullback-Leibler divergence
  - Quantification for the similarity of two distributions
  - Assumes independence of variables/ sensor channels
  - Does not take into account the sequence of the data

- E_r : reconstruction error:
  - Difference between initial MVTS (input signal) Y with m signals and the reconstructed MVTS $\widehat{Y}$ (reconstructed signal)
  - $E_r = \sum_{k=1}^{m} ||Y_k - \lambda \hat{Y}_k||_F^2$ with $\lambda = 1$

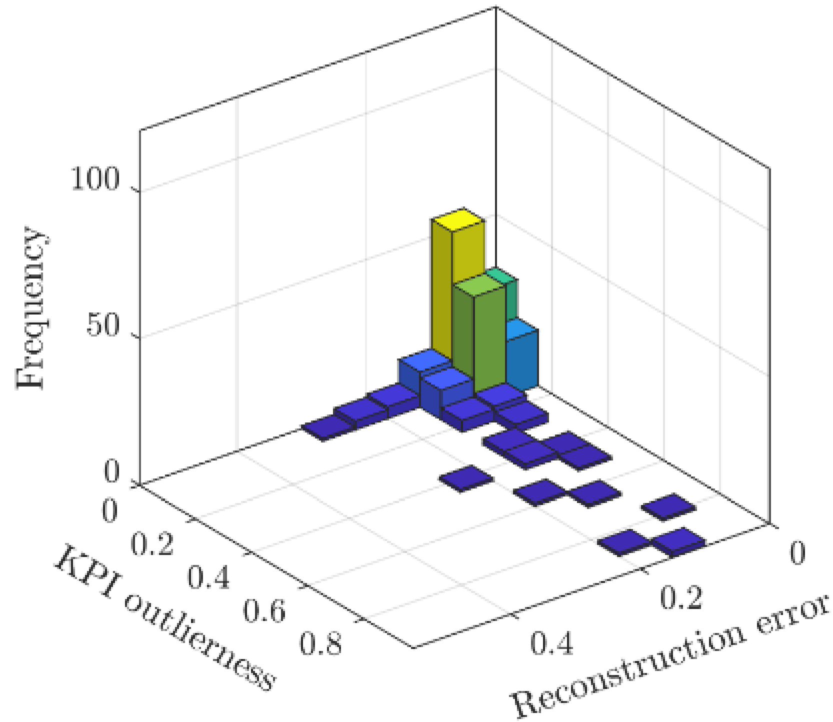
- E_t: cost function VAE-LSTM training
  - $E_t = E_r - D_{KL}\{N(0,1)||N(\mu,\sigma)\}$

anika.terbuch@unileoben.ac.at

Bivariate: “outlierness” vs ML-Reconstion error



(a) Histogram of first run



(b) Histogram of second run

Fig. 7.15: Bivariate histograms of the reconstruction error and the outlierness obtained with KPIs. The training was performed with unlabeled data via 4-fold cross-validation, and the average test error was computed for each sample. For the second run the anomalous samples found in the first run were eliminated from the data set.

Comparison of sub-process performance

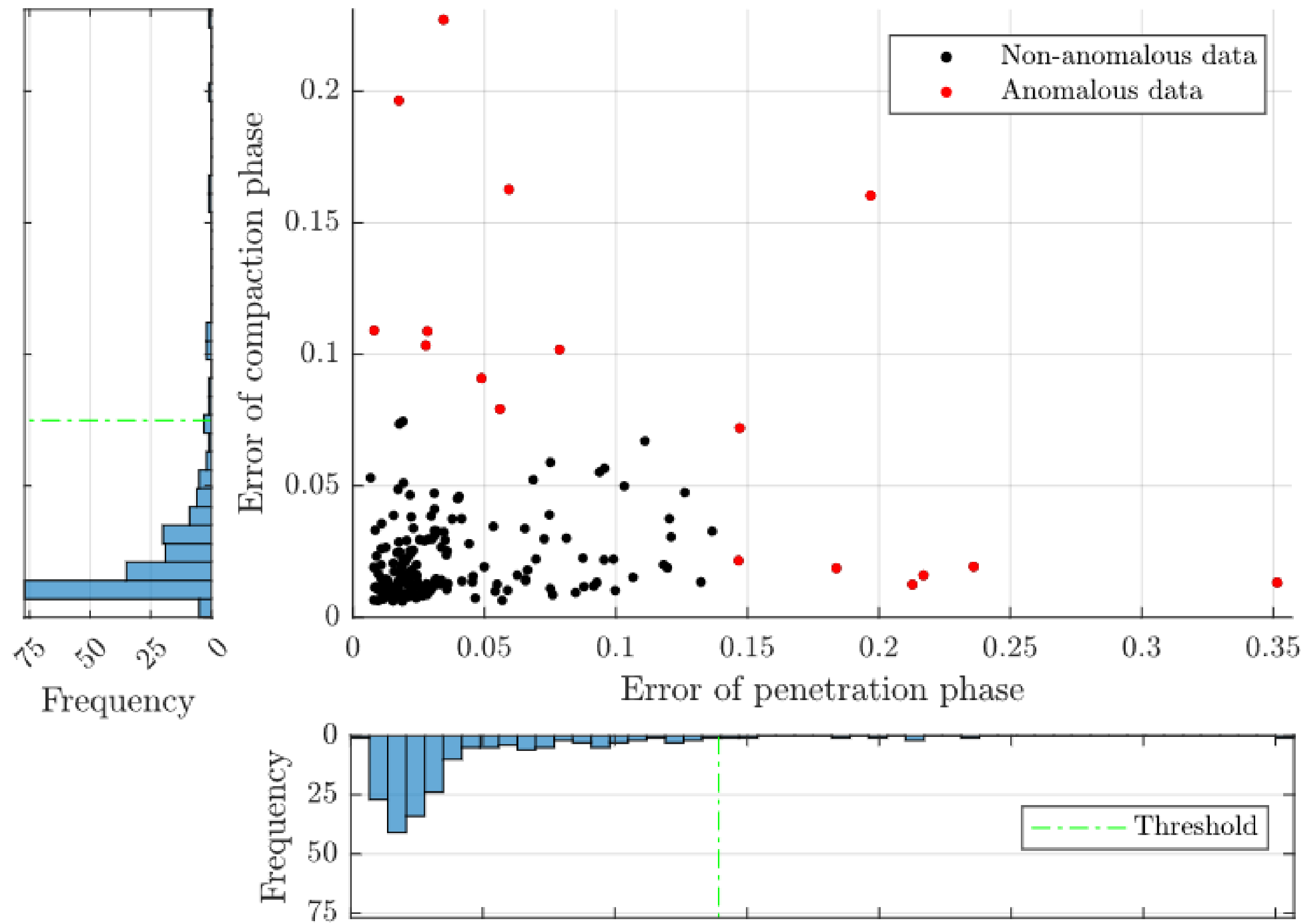
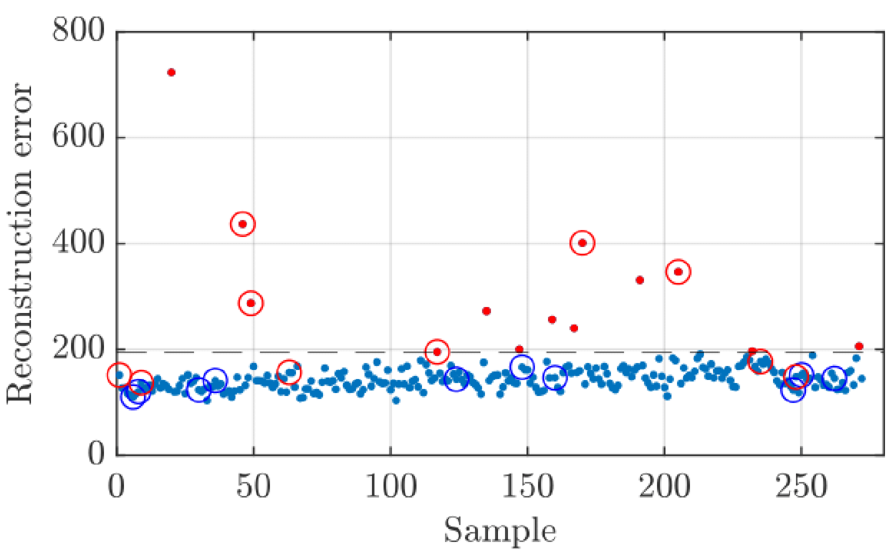


Fig. 7.16: Phasewise error plot of test data samples, with corresponding histograms of the error distributions and threshold values for separating normal data from outliers.

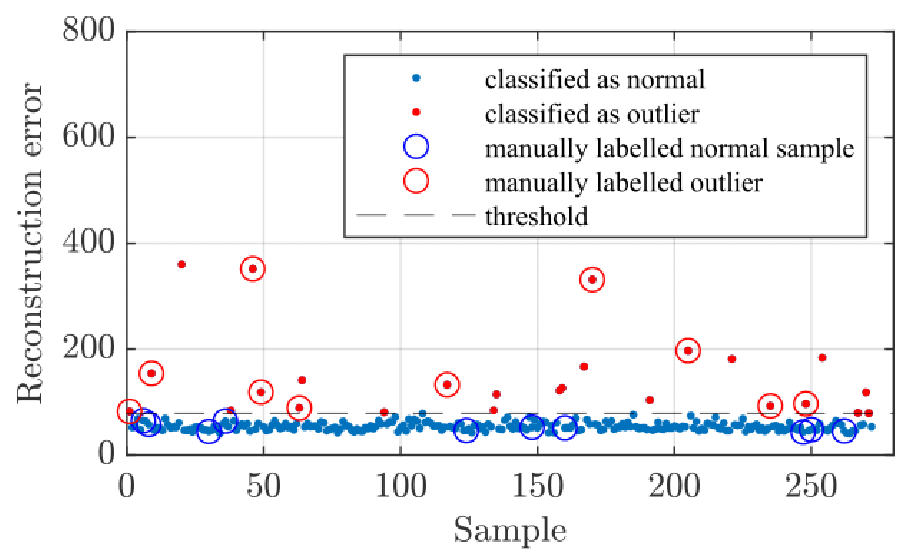
Importance of Hyperparameter Optimization (HPO)

- Improved reconstruction error for classifier trained with optimized hyperparameters
  - Higher distance between the labelled groups of outliers and normal samples
  - Lower variance and IQR within the groups – tighter bounds for the detection of outliers
  - No false classifications of the manually labelled examples after the hyperparameter optimization

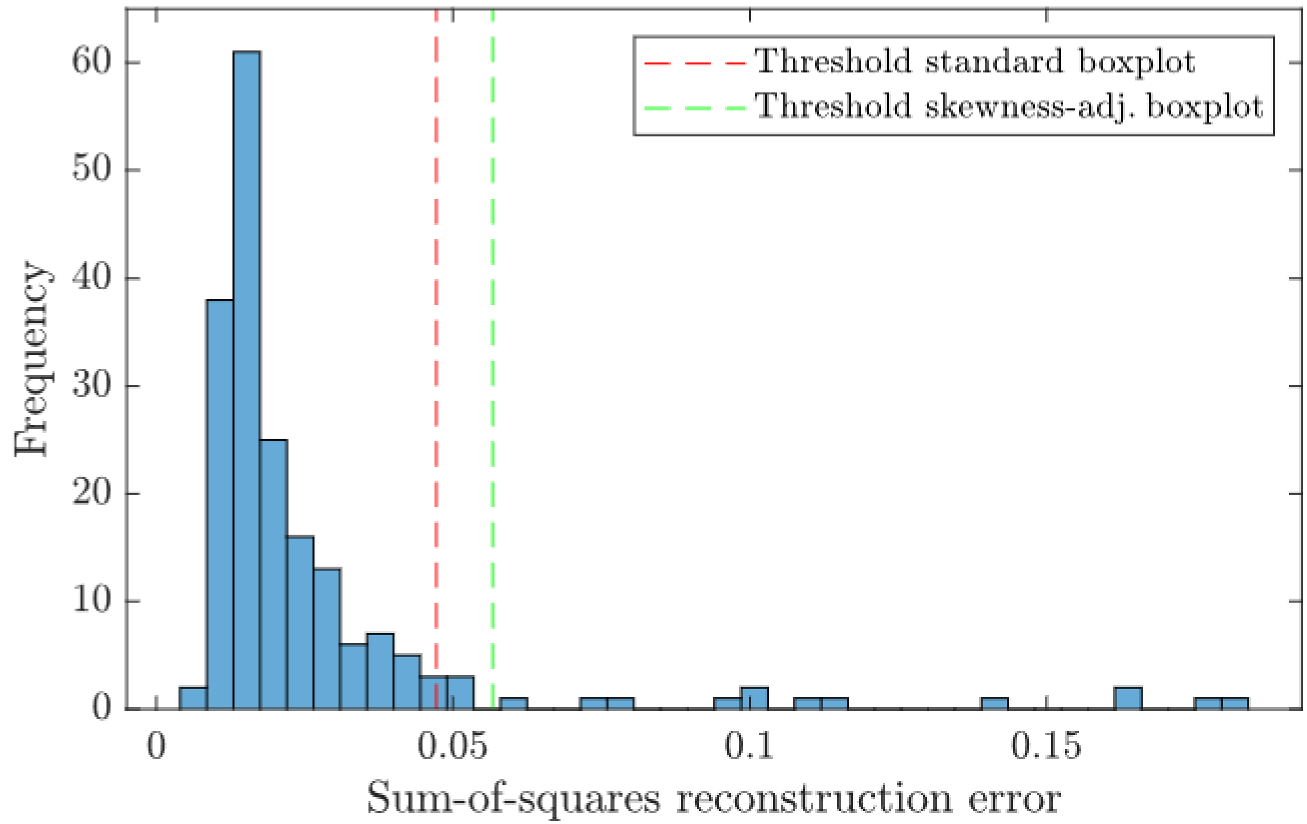
Classifier trained without HPO



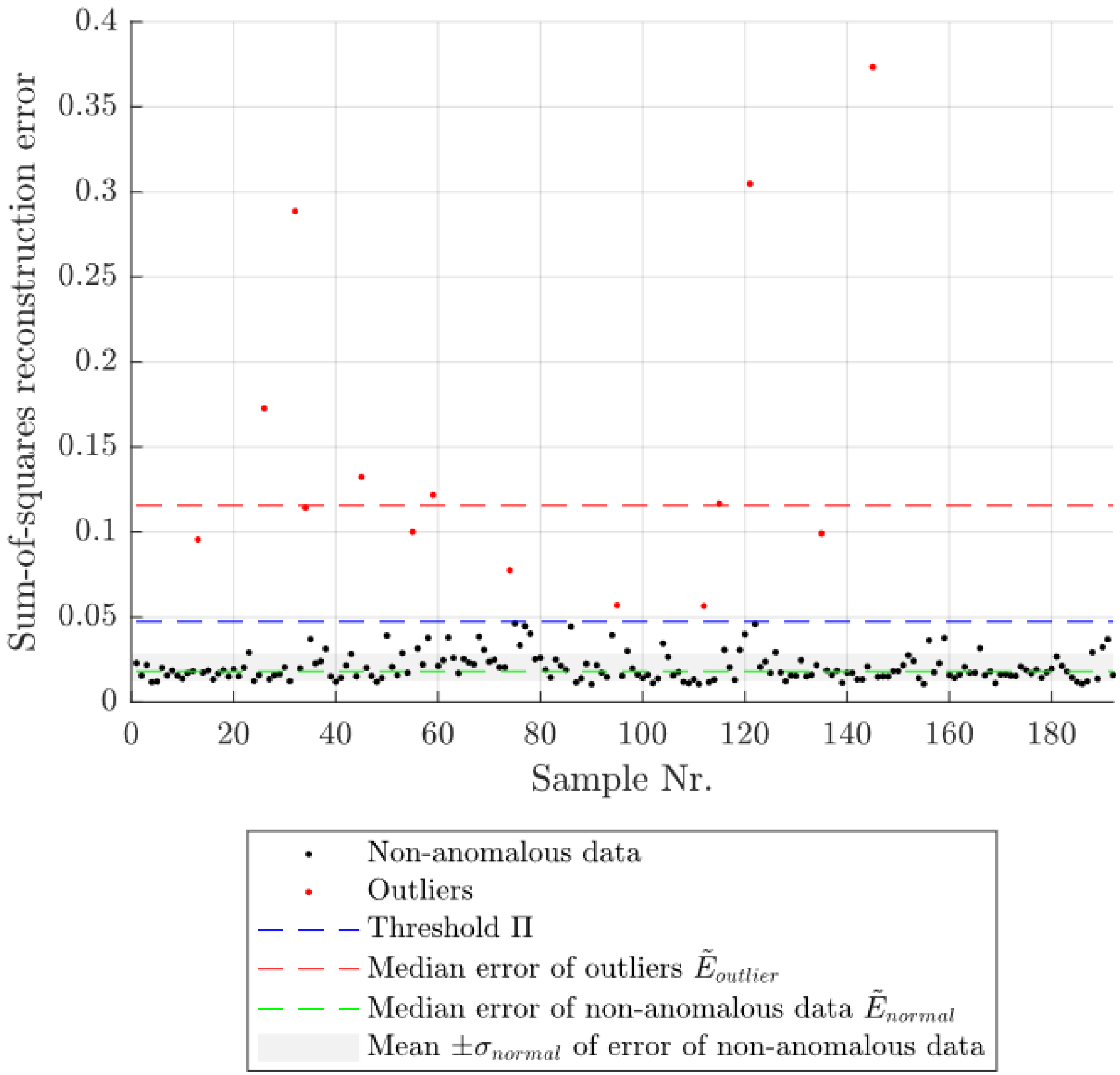
Classifier trained with HPO



Skewness adjustment



ML Classification



anika.terbuch@unileoben.ac.at

Hyperparameter Optimization (HPO)

Settings of the Genetic Algorithm:

Latent dimension: 3

Population size: 20

Maximum number of generations: 7

Mutation probability: 0.05

Random probability: 0.05

Data:

Site: Fehring

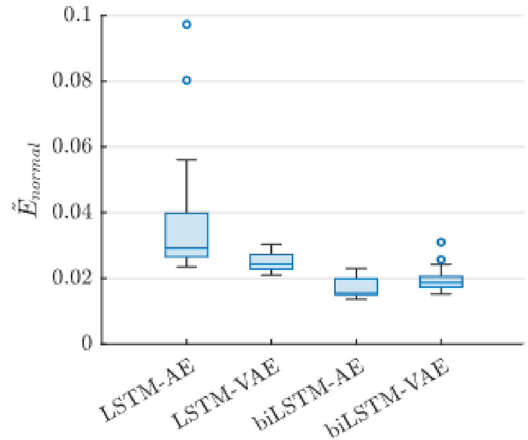
Number of input channels: 6

Number of training samples: 80

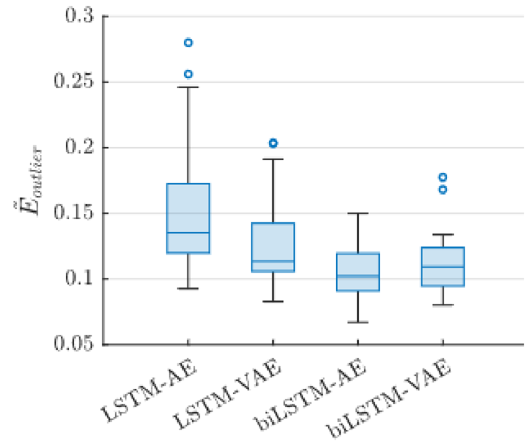
Table 1: Results of HPO with 1-layer networks

Network Structure			Optimized Hyperparameters				
Net Type	Layer Encoder	Layer Decoder	Nr. Epochs	Neurons Encoder	Neurons Decoder	Learning Rate	Batch Size
			[1, 100]	[1, 100]	[1, 100]	[$3\cdot10^{-3}$, $1\cdot10^{-1}$]	[2, 53]
AE	LSTM	LSTM	93	59	19	$3.691\cdot10^{-2}$	12
VAE	LSTM	LSTM	77	55	30	$5.133\cdot10^{-2}$	4
AE	biLSTM	biLSTM	96	70	32	$1.877\cdot10^{-2}$	3
VAE	biLSTM	biLSTM	96	57	47	$2.571\cdot10^{-2}$	6

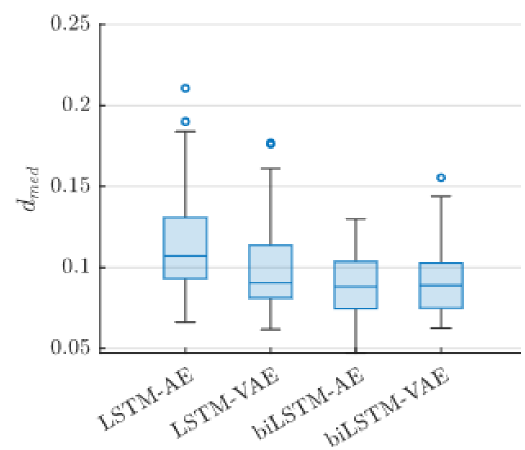
Comparing different architectures



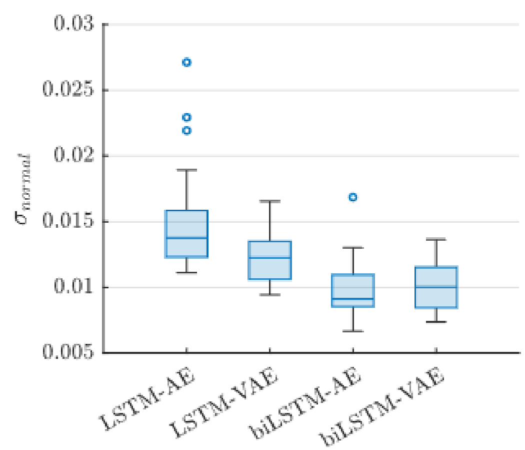
(a) Median error of samples labeled as non-anomalous



(b) Median error of samples labeled as anomalous



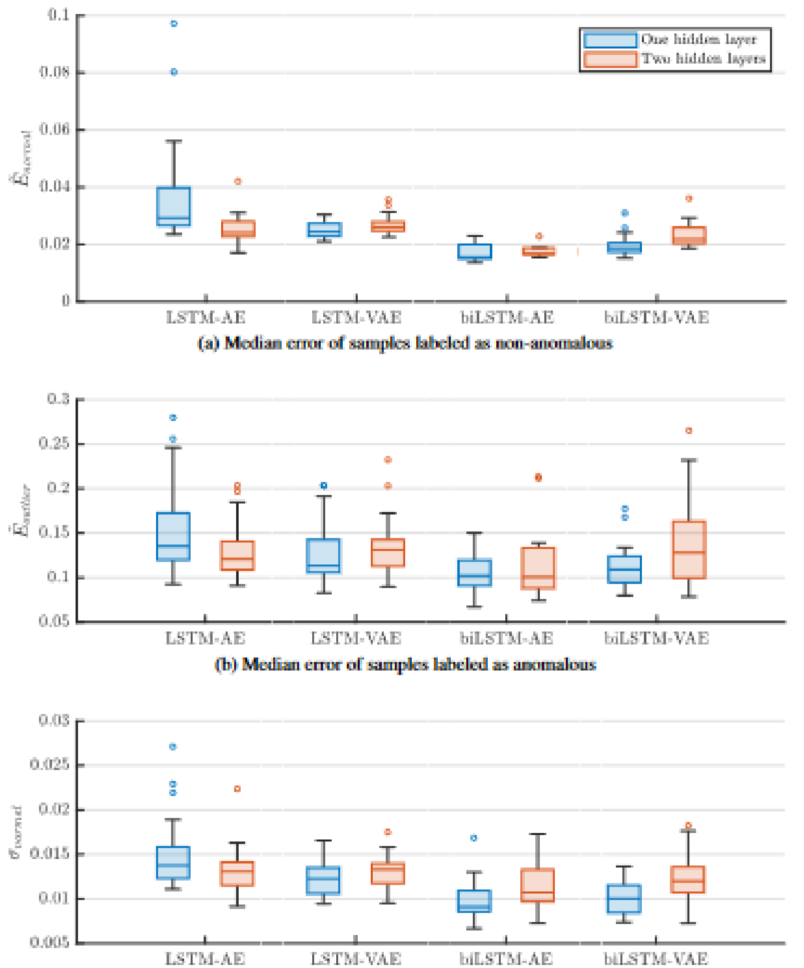
(c) Distance of median errors of non-anomalous samples (a) and anomalous samples (b)



(d) Standard deviation of errors of samples labeled as non-anomalous

Fig. 7.7: Statistical measures of sum-of-squares sample errors obtained with different autoencoder architectures containing one hidden layer in encoder and decoder for $n_{runs} = 25$ test runs.

Comparing LSTM and bi-LSTM



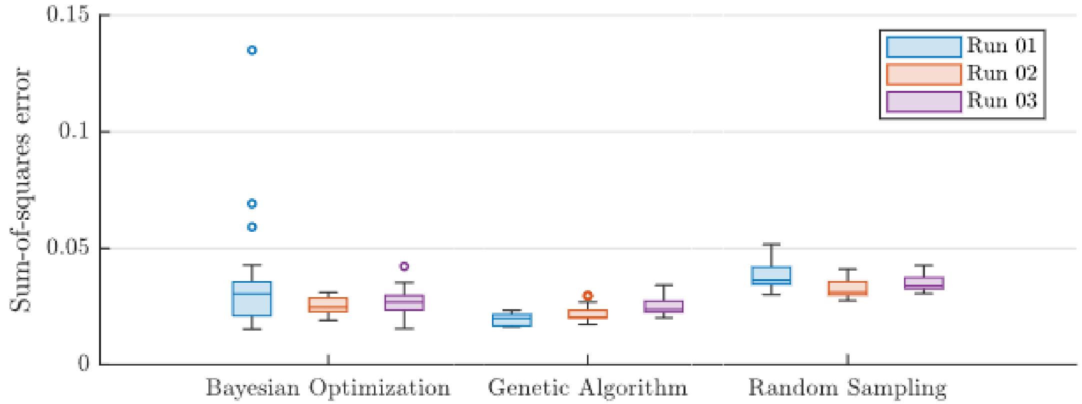
(a) Median error of samples labeled as non-anomalous

(b) Median error of samples labeled as anomalous

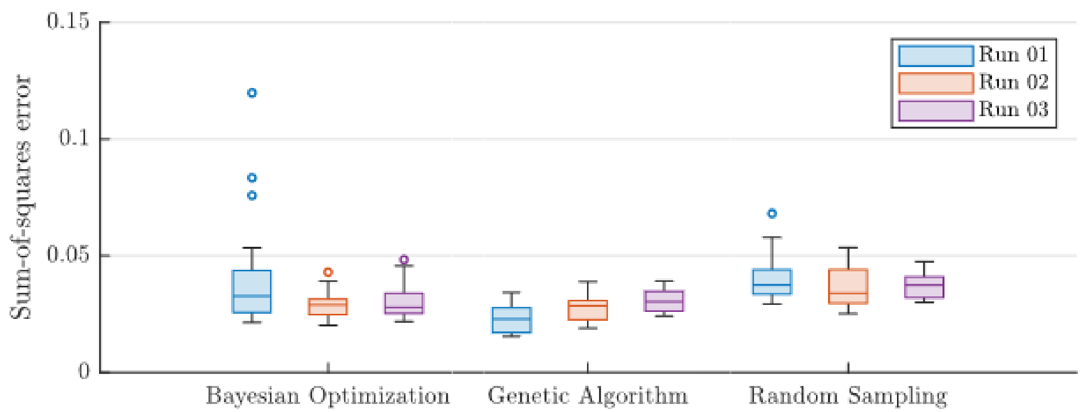
(c) Standard deviation of errors of samples labeled as non-anomalous

Fig. 7.9: Statistical measures of sum-of-squares sample errors obtained with different autoencoder architectures for $n = 25$ test runs. The boxplots in blue are associated with autoencoders, whose encoder and decoder contain one hidden layer. The boxplots in orange correspond to autoencoders, whose encoders and decoders contain two hidden layers.

Run dependent performance vs architecture



(a) Training error per sample



(b) Validation error per sample

Fig. 7.10: Sum-of-squares errors per sample of training samples (a) and validation samples (b) after the last epoch for biLSTM-AEs. The hyperparameters were optimized $n_{HPO} = 3$ times with each of the tested HPO methods, and with each hyperparameter configuration $n_{runs} = 25$ test runs were performed.

Comparison of Weight Initializers

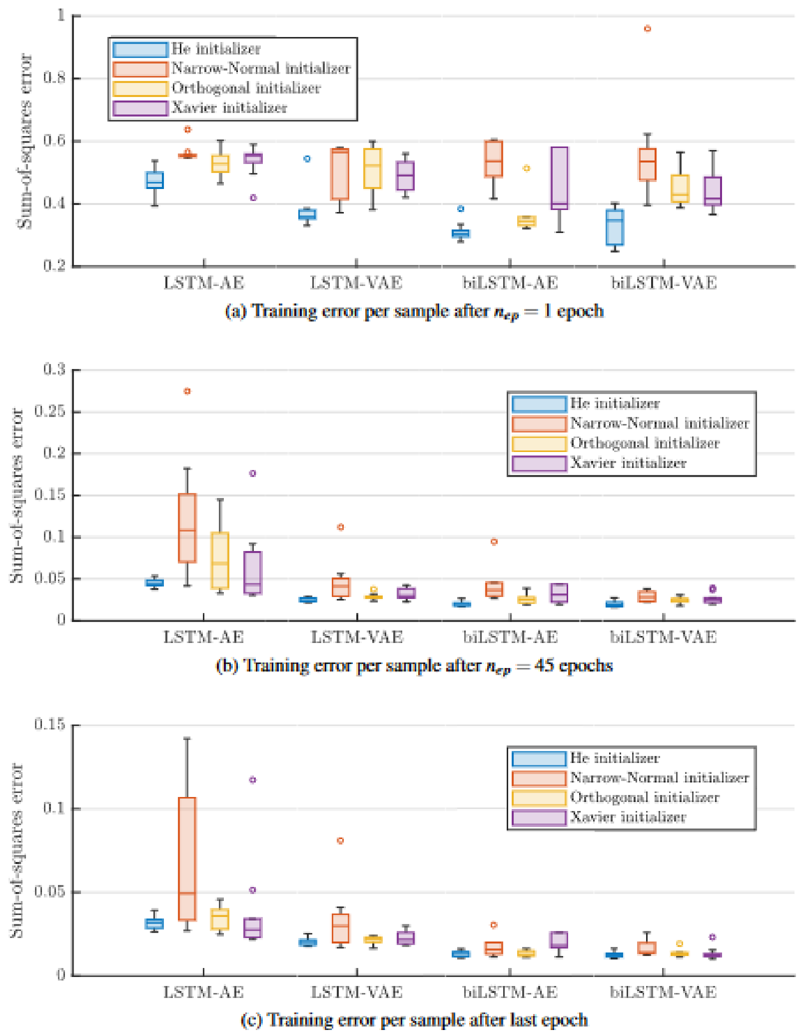


Fig. 7.12: Boxplots of sum-of-squares training error per sample of different autoencoders after a particular number of epochs using four different weight initializing methods. For each method and autoencoder $n_{runs} = 10$ runs were performed.

Comparison of Weight Initializers

Xavier-Initializer (default) [2]: sampling from $\mathcal{U}\left(-\sqrt{\dfrac{6}{n_{Input}+4n_{Neurons}}},\sqrt{\dfrac{6}{n_{Input}+4n_{Neurons}}}\right)$

He-Initializer [2]: sampling from $\mathcal{N}\left(0,\dfrac{2}{n_{Input}}\right)$

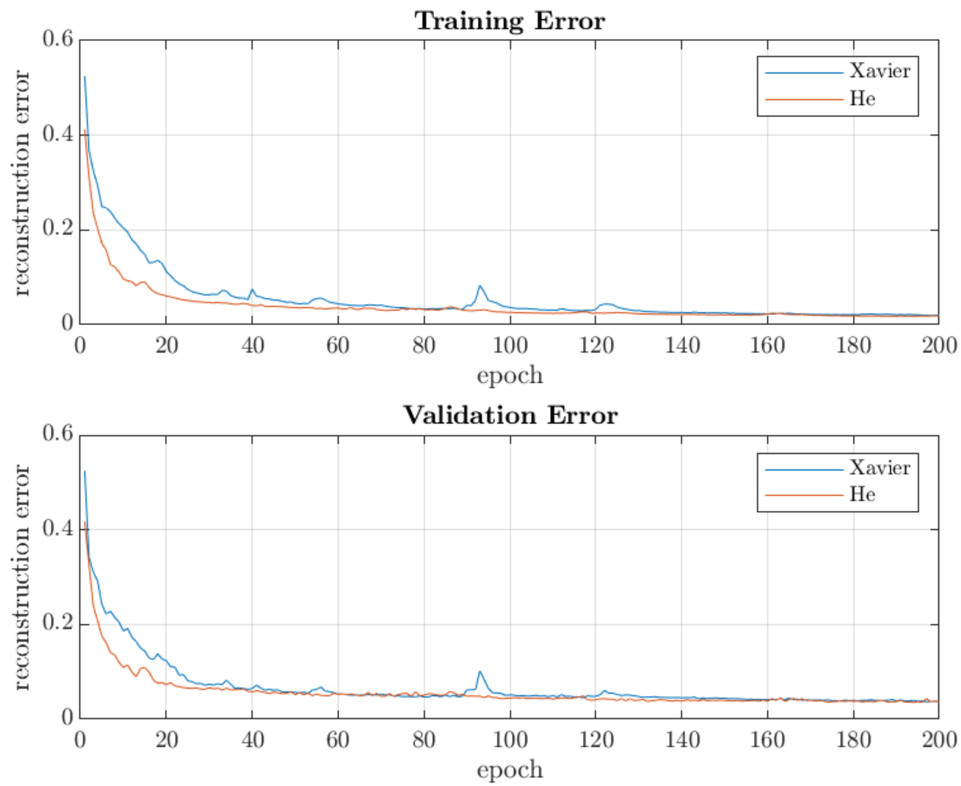


Fig. 9: Plots of mean squared error of training (top) and validation data (bottom) of a biLSTM-biLSTM VAE with different weight initializers

anika.terbuch@unileoben.ac.at

Statistical behaviour of the different architectures

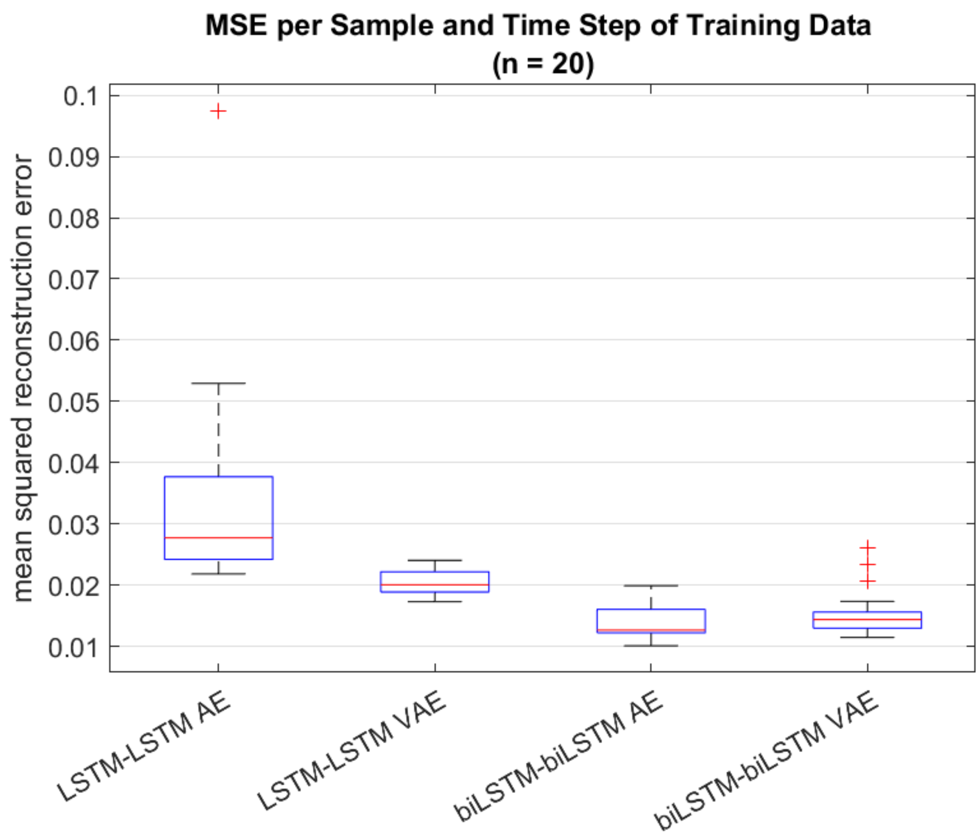


Fig. 10: Boxplot of MSE per sample and time step with training data as input

Architectures tested

Table 2: Results of HPO with 1-layer networks – first and obsolete run

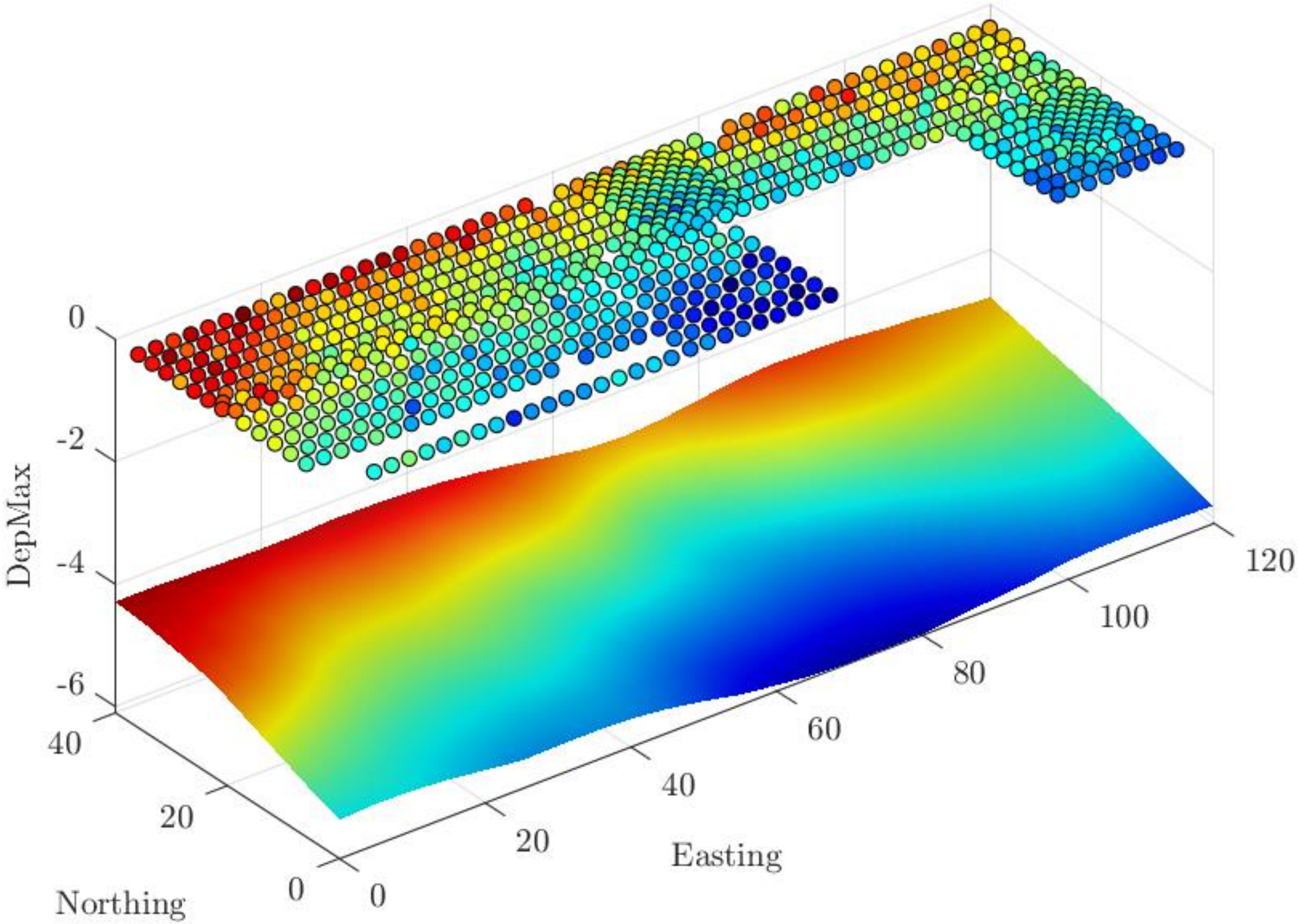
Network Structure			Optimized Hyperparameters				
Net Type	Layer Encoder	Layer Decoder	Nr. Epochs	Neurons Encoder	Neurons Decoder	Learning Rate	Batch Size
Range			[1, 100]	[1, 100]	[1, 100]	[5e-5, 1e-2]	[2, 2/3 x $n_{Samples}$]
AE	LSTM	LSTM	92	75	96	8.22e-3	13
VAE	LSTM	LSTM	88	97	71	9.28e-3	5
AE	biLSTM	biLSTM	83	70	32	9.51e-3	3
VAE	biLSTM	biLSTM	82	62	65	7.83e-3	12

Further work

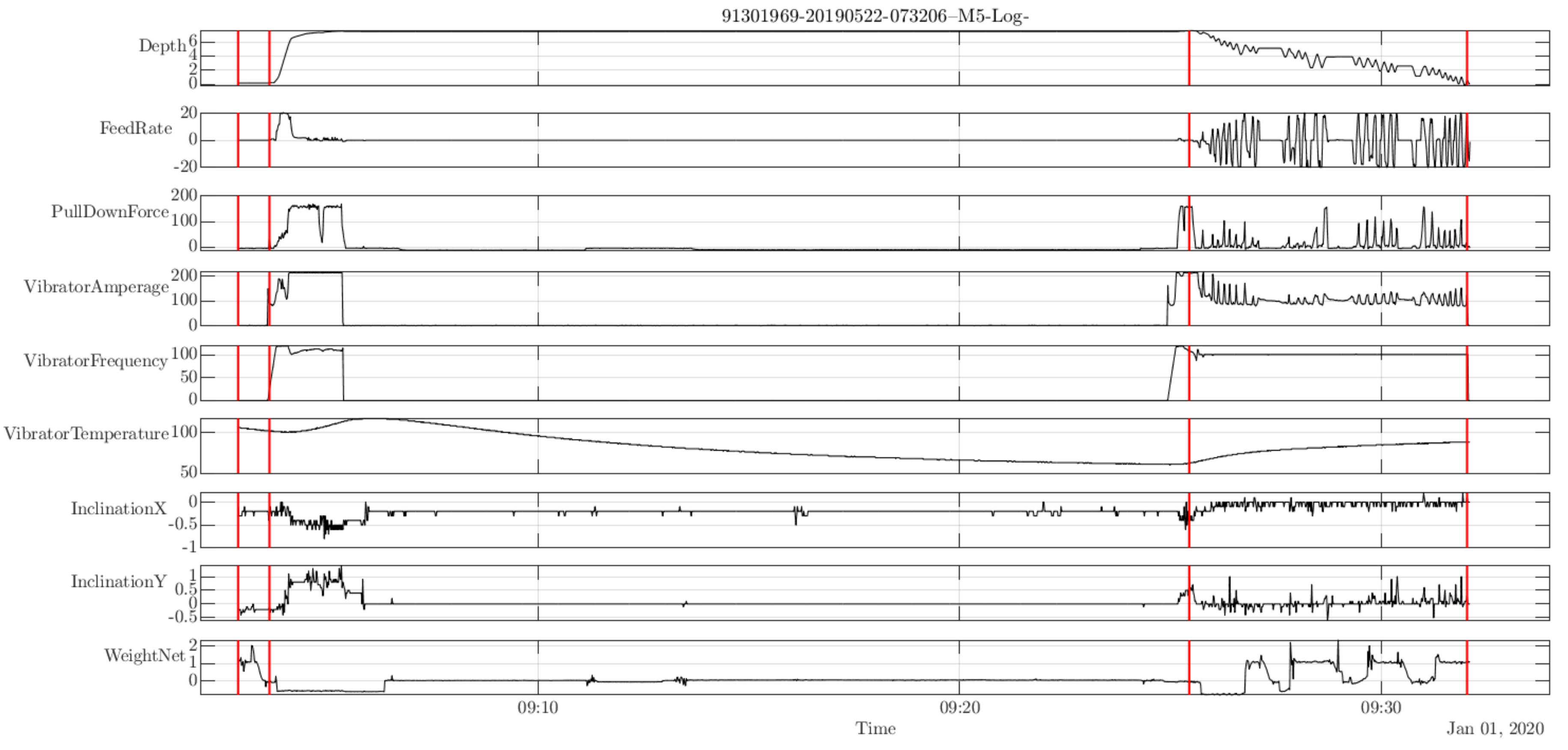
- Alternative cost function / measures in the latent space
- Serial-parallel-hybrid model
  - Pre-selecting the MVTS for training the ML-classifier using KPI-analysis
- Incorporating other normalizations for the sensor signals
- Including other variables (e.g. the latent dimension) in the hyperparameter optimization
- Subsurface modelling to separate systematic from random variations to improve outlier detection via KPI.

anika.terbuch@unileoben.ac.at

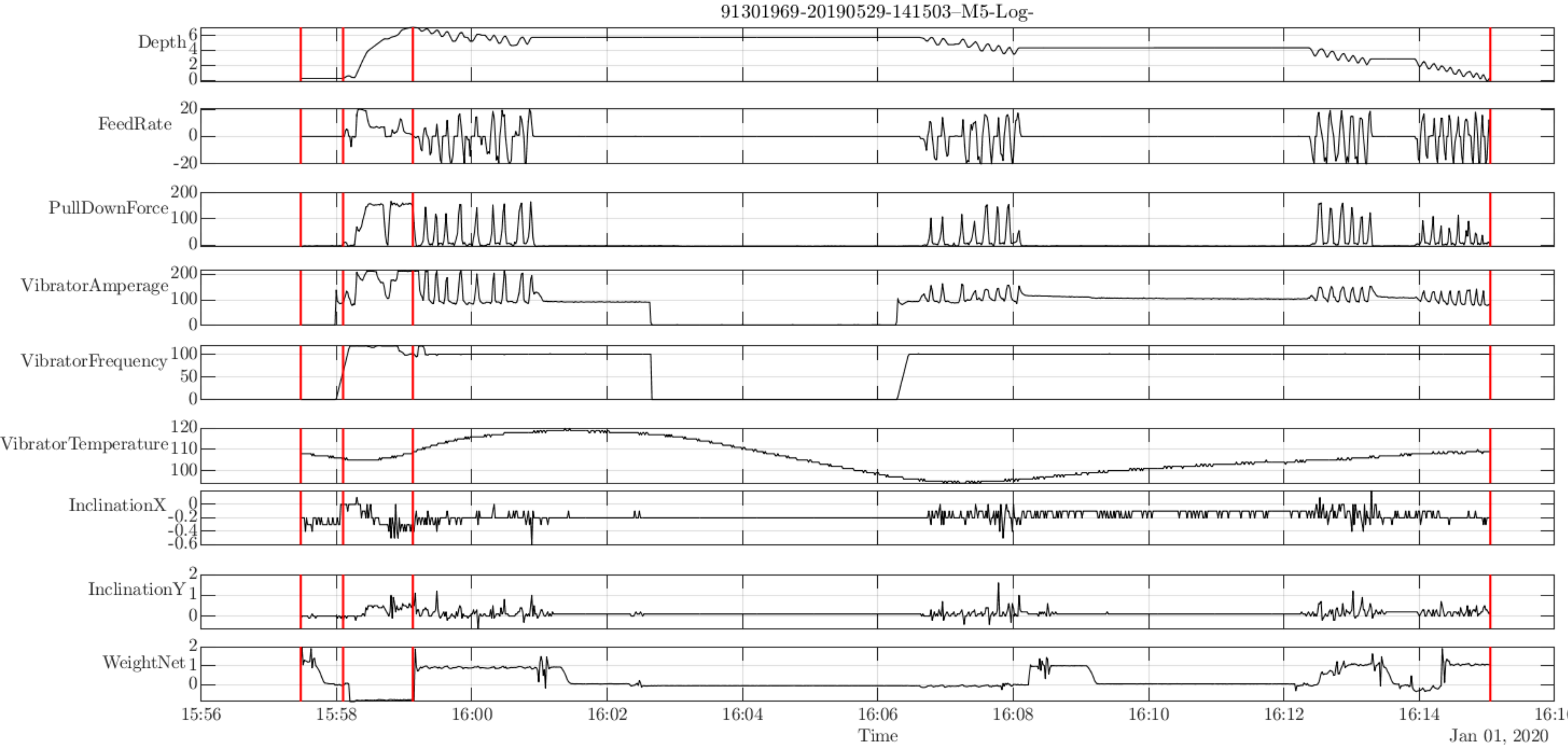
KPI Spatial Modelling



Classified as outlier by both



Classified as outlier by ML



Classified as outlier by KPI