Selection and optimization of historical data for the training of artificial intelligence in power plant engineering

Dipl.-Ing. Frank Gebhardt
NeuronalNetWorks! GmbH (NNW)

NEURONALNETWORKS!

Content of today's presentation

- **Introduction** to the lecture / Focus
- **Challenges** of "Big Data"
- **Theory** of predictions of process variables with NN
- Integral **prediction**
- **Selection** and **evaluation** of training and test data
- **Practical example**
- 10 practical steps for **deterministic AI**
- AI **interactions** with the **control system**

Introduction to the lecture / Focus

- Application of artificial intelligence (AI) in power plant process engineering

  ►Encounter with "Big Data"!
- Procedures for the selection and optimization of historical data for the training of AI
- Applications for predicting process variables:
  - Steamgeneration
  - NOx, NH3
  - Temperature boiler ceiling
  - etc…
- High expenditure of time for selection of learning patterns

  ► > 64,0000 learning patterns corresponds to > 1.5 Gbyte of data

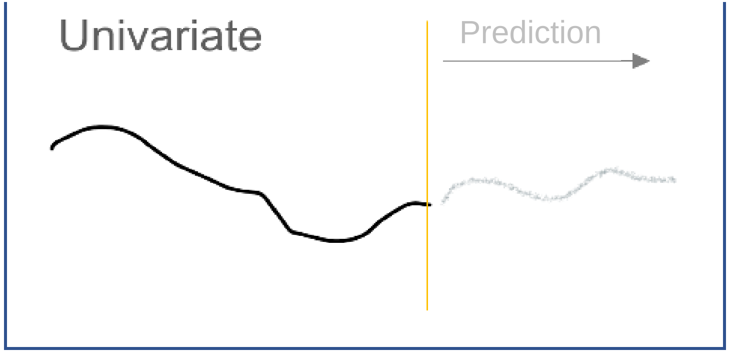
Challenges of "Big Data"

- Data amount
  - cannot be processed because, for example, there are upper limits on the number of rows in Excel
- Performance Issues
  - is available with e.g. Excel, if you want to open a file > 1 GByte
  - may mean several minutes of waiting time
- Search for alternatives is required
- Algorithms for preprocessing the training data very helpful!
  ► Professional processing required for cost-effectiveness!

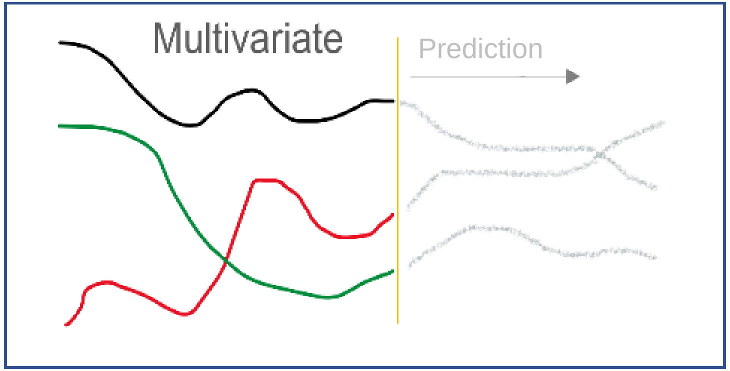
Theory of predictions of process variables with NN 1v2

Different approaches and possibilities for predicting process variables:

- Dependence on only **one** variable:
  - Only one signal is used to predict future behavior, e.g. only steam
  - This dependence is called "univariate" in mathematics



- Dependence on **several** variables:
  - Several signals are used to predict future behavior, e.g. steam, temperature, CO, air, ...
  - This dependence is called "multivariate" in mathematics

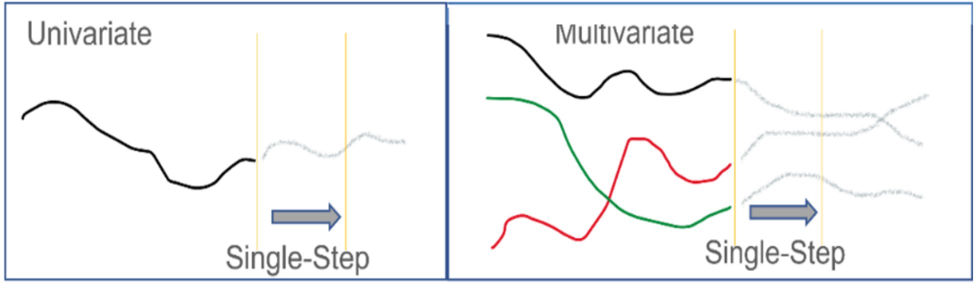


Theory of predictions of process variables with NN 2v2

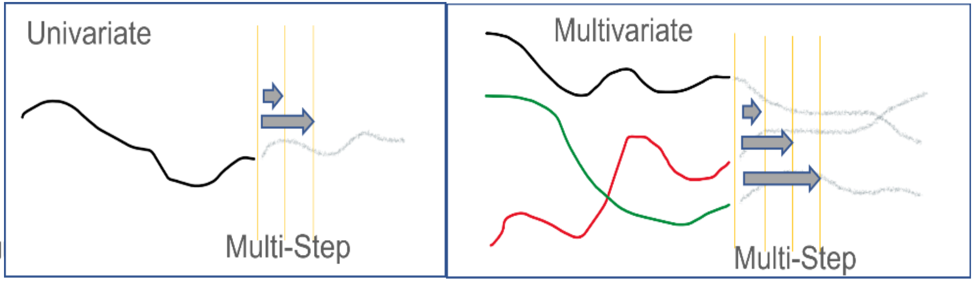
Forecast horizons

There are different forecast horizons for predicting process variables:

- Single-Step: only one forecast horizon, e.g. only 5 minutes:



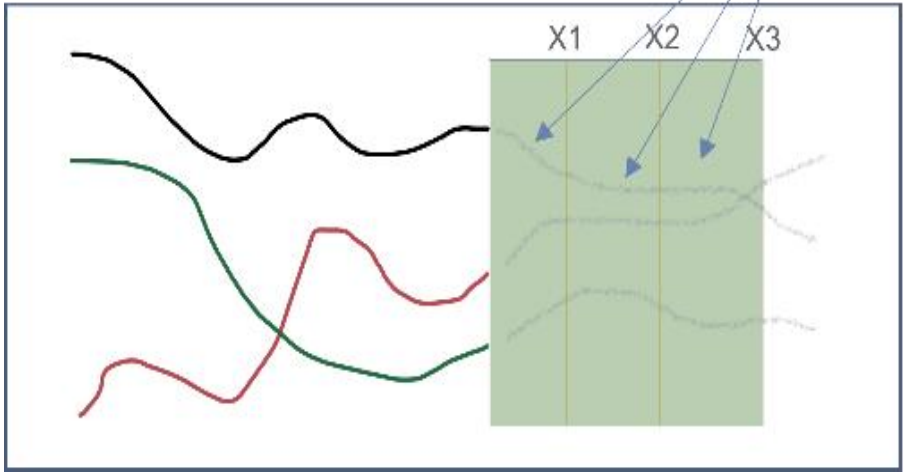
- Multi-Step: Predict multiple forecast horizons, e.g. 5 minutes, 10 minutes, ...



Integral prediction

- Multivariate and multi-level forecasting method
- Averaging over forecast horizon
- **Averages:**
  - easier to predict
  - contain the important process engineering information
- Confirmation by practice
  - ► Prediction of steam mass flows

Averaging



Selection and evaluation of training and test data 1v5

- Training dataset
  - 4 to 6 months for predictive model, e.g. steam prediction
  - Elimination of plant downtimes and disruptive process events
  - Storage of as much data as possible
- Test dataset
  - Selection of a small test set of a few weeks

    ► These are the data that the neural network has not learned, i.e. does not know
- Overfitting
  - Typical strategies to avoid overfitting don't seem to work here
  - Large datasets are the key!
  - Practical confirmation of steam prediction and temperature prediction boiler ceiling

    ► Use of very, very large data sets required!

Selection and evaluation of training and test data 2v5

Complexity optimization (with the same NN topology)

- **Large** data set
  - In case of underfitting (poor learning), the complexity of the NN is too low to model the complexity of the data set

  ► Solution: Reduce the number of trainable outputs to increase the available complexity per output or reduce variarity
- **Small** data set
  - With **Overfitting** the complexity of the NN is too great
  - The NN remembers each sample of the training data set, the generalization of test data will be poor in most of these cases

  ► Solution: Increase the number of trainable outputs or increase training dataset or variarity

Selection and evaluation of training and test data 3v5

Uneven distribution of data in a data set

- "Unbalanced regression"
- Large 6-month data sets reflect typical distribution of actual plant operation

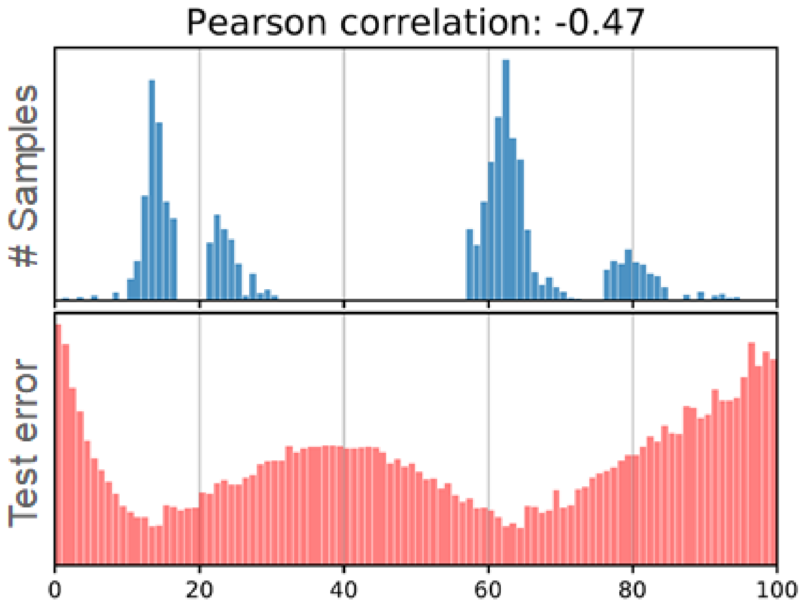
Number of input/output pairs	Spatial density	Probability of alternative similar input/output pairs
Accumulation	high	high
Shortage	low	low

Selection and evaluation of training and test data 4v5

Example of a data set distribution

► Negative correlation (-0.47) between the number of samples for a given operating behaviour and a test error

- **Frequent** data in a particular operating behavior results in a **small test error** for that behavior
- **Rare** data in a certain operating behavior leads to a **high test error** for this operating behavior

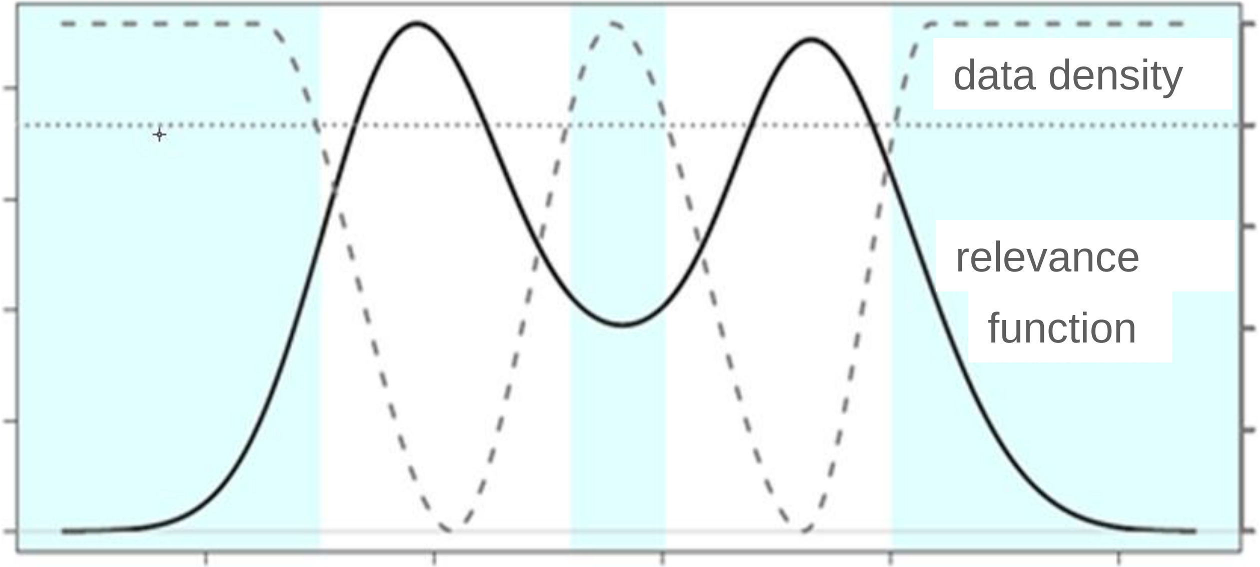


Selection and evaluation of training and test data 5v5

Relevance function

To improve the learning success of the NN:

- **Inverted** data density is used as a relevance function



There are many theoretical methods from the literature:

- SMOTER (2013)
- SMOGN (2017)
- **WERCS (2018)**
- Dense Loss (2021)

Steam prediction:

- Method WERCS
- Specific weighting of synapse weights
- Strong improvement
- Practical!

Practical example - Procedure 1v4

Standard procedure for balancing algorithms for balancing learning patterns according to the previous mentioned methods:

1. **Analysis** of the complete data set for common and rare data

2. **Balancing** the common and rare cases by using a combination of:
   - **Oversampling <u>rare</u>** data by creating synthetic data, such as adding random noise to existing data
   - **Undersampling for data clusters by finding data that:**
     - are close to each other and therefore approximately the same
     - are staggered in time and similar

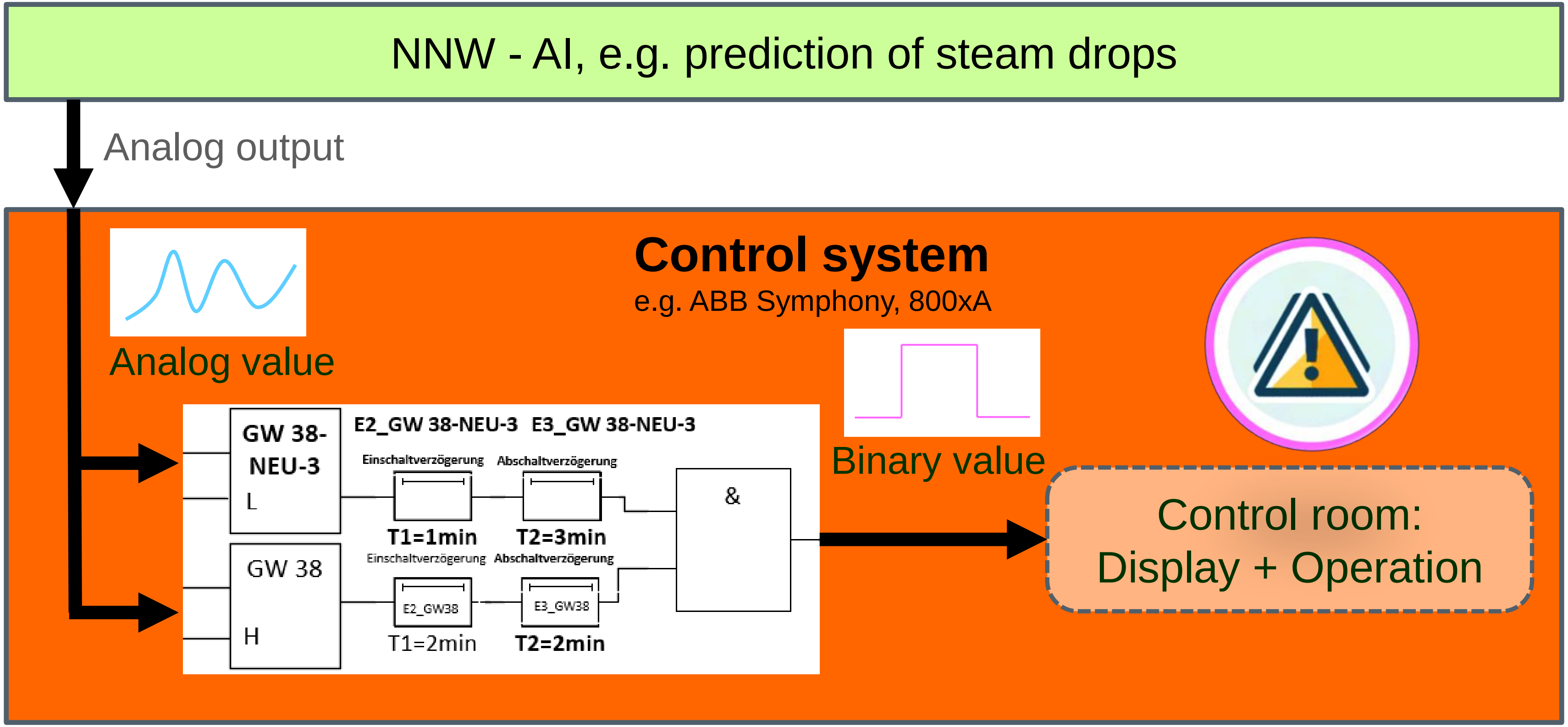
     and can therefore be deleted.

Practical example - steam forecast WIP (waste incineration plant) 2v4

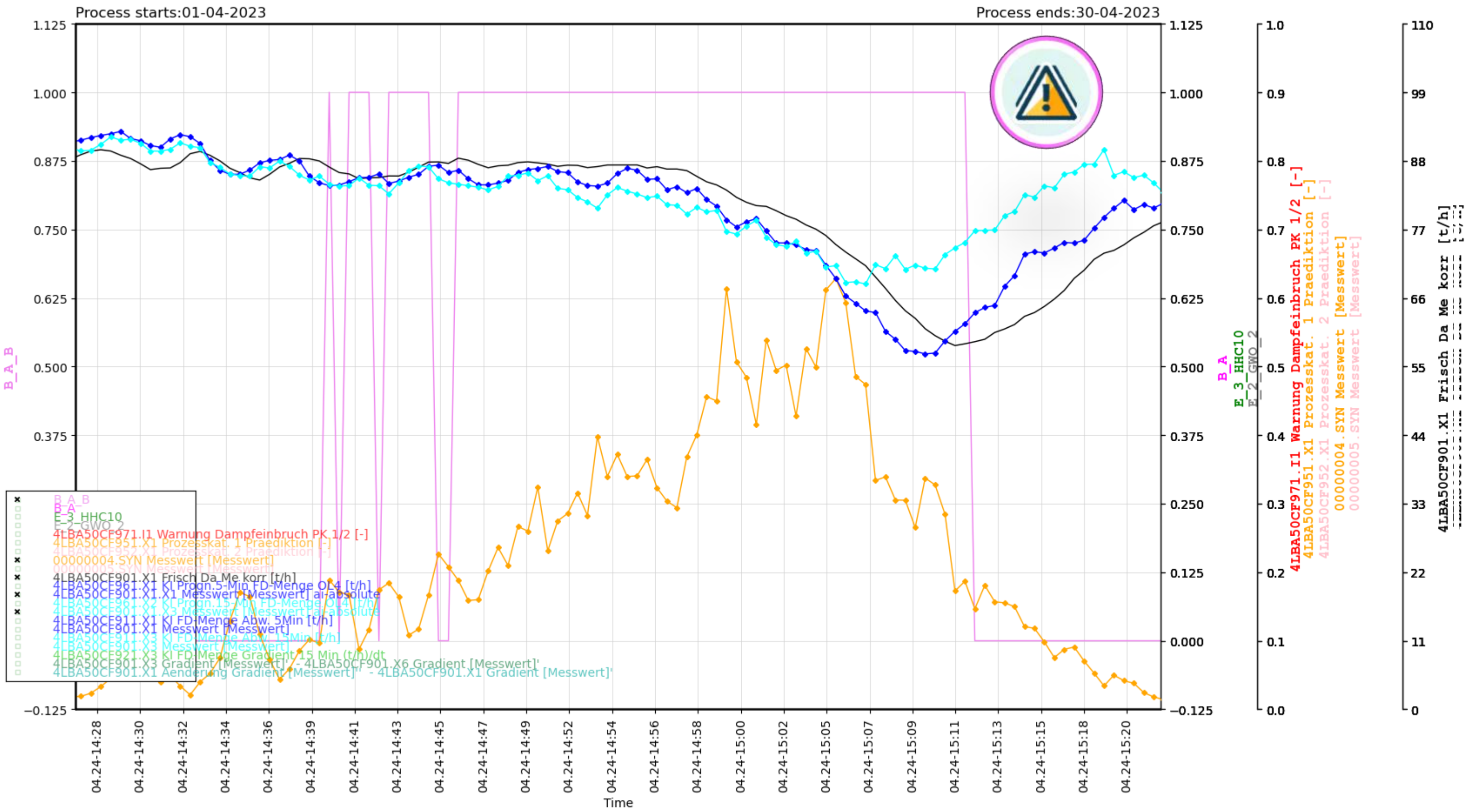
Relevance / Process Engineering Indicator – Definition

- The relevance was also used as a process engineering indicator, which was called here "Process category 1: Waste grate overfill" (double function!)
- Trends in measurements:
  - Primary air pressures: trend upwards
  - Flue gas O2 content: trend upwards
  - Fire rate control grate feed: trend upwards
  - Combustion chamber temperature: trend downwards
  - Moisture measurement: trend upwards
- Future values of steam production
  - 5 minutes: Trend downward
  - 15 minutes: Trend downward
  - 30 minutes: Trend downward
- Calculation of relevance / process engineering indicator
  - Mathematical linking of:
    - Trends in measurements
    - Future values of steam production

Evaluation of outputs with binary logic 3v4



Practical example – Analog and binary Result 4v4



NEURONALNETWORKS!

10 practical steps for deterministic AI

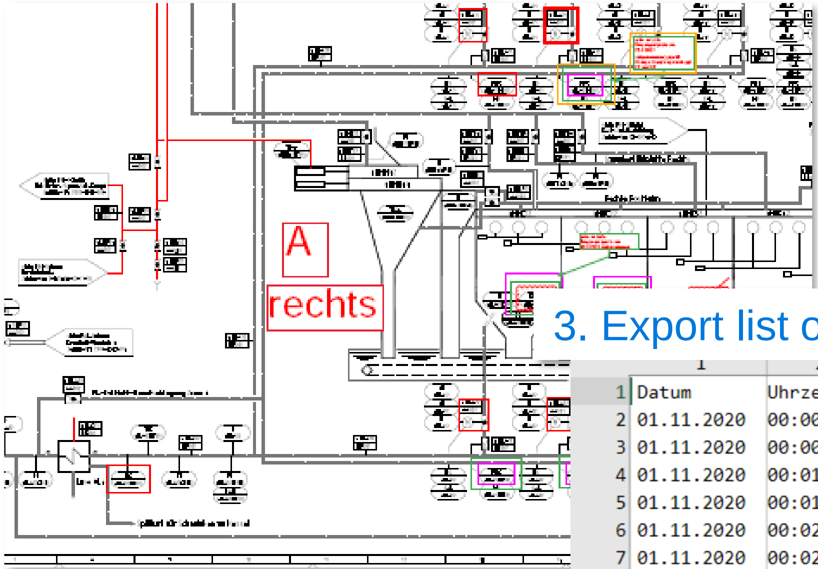
→ The previous presentation includes steps 1 to 4

1. **Export of approx. 50 relevant measurements from process engineering**
2. **Correlation analysis* from measurements to predictions**
3. **Selection of the measurements with the highest correlations**
4. **Use of the NNW AI tools © to prepare the learning patterns**
5. *Training the neural network with NNW Cuda-C learning algorithm ©*
6. *Evaluation of outputs with binary logic © in the control system*
7. *„Offline-Commissioning“* mit training- und test pattern*
8. *Optimizing the training algorithm with NNW Cuda-C Focus-Learning ©*
9. *Loading the scale file and knowledge base into the automation*
10. *Launch of deterministic AI*

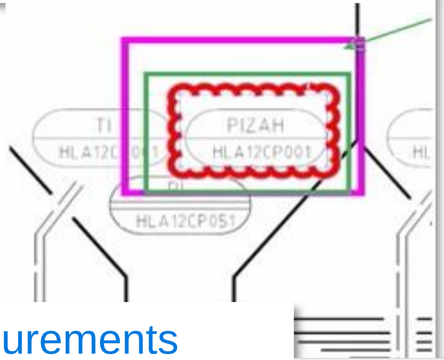
* © NNW-Toolbox: Python, Cuda-C

Regarding 1. Export of approx. relevant 50 measurements

1. Marking of the measurements in the P&ID



2. Measurement details



3. Export list of measurements

	1	2	3	4	5	6
1	Datum	Uhrzeit	xHLA21CP001.X1	xHLA11AA001.A1	xHLA11CP001.X1	xHLA21AA00
2	01.11.2020	00:00:00	15,29	104,03	18,44	99,8
3	01.11.2020	00:00:30	14,94	104,03	18	99,8
4	01.11.2020	00:01:00	15,34	104,03	18	99,8
5	01.11.2020	00:01:30	14,73	104,03	17,56	99,8
6	01.11.2020	00:02:00	16,53	104,03	19,27	99,8
7	01.11.2020	00:02:30	16,24	104,03	18,78	99,8
8	01.11.2020	00:03:00	16,89	104,03	19,3	99,98
9	01.11.2020	00:03:30	16,68	104,03	19,25	100
10	01.11.2020	00:04:00	16,37	104,03	18,88	100
11	01.11.2020	00:04:30	15,98	104,03	18,51	99,86
12	01.11.2020	00:05:00	15,54	104,03	18,06	99,93
13	01.11.2020	00:05:30	15,31	104,03	17,75	100

Regarding 2nd and 3rd correlation analysis (data science)

	A	B	C	D	E	F	
1	Stamm-KKS steht für 21 Derivate!	Text	12HTA10CQ954Q01#4#Nox ende kessel > 0,3	max	12HTA10CQ954Q01#4#Nox ende kessel < -0,3	min	12H Q00 5#N end kess 0,3
2	10LBA50CF041Q02	Dampf Regeler ausgang	WAHR	0,32	FALSCH		FA
3	12HTA10CQ002Q01	O2 ende kessel	WAHR	0,64	WAHR	-0,46	FA
4	12HTA10CQ003Q01	CO ende kessel	FALSCH		WAHR	-0,53	FA
5	12HTA10CQ007Q01	SO2 messung nach e-Filter	WAHR	0,41	WAHR	-0,49	FA
6	10HBK01CT043Q01	Temperatur rost 3	WAHR	0,63	FALSCH		FA
7	10HBK01CT044Q01	Temperatur T4 Feurraum 12 meter	WAHR	0,62	FALSCH		FA
8	10HBK01CT045Q01	Temperatur T5 (dach temperatur kessel)	WAHR	0,32	WAHR	-0,39	FA
9	10HHL14CP001Q01	Druck OW zone 4	WAHR	0,38	FALSCH		FA
10	10HHQ22CP001Q01	Druck SW zone 2	WAHR	0,35	WAHR	-0,32	FA
11	10HFB10DS001C01	Rost 1 %	WAHR	0,35	FALSCH		FA

Regarding 4. Preparation of learning patterns

1. Data in defined format with derivatives calculation

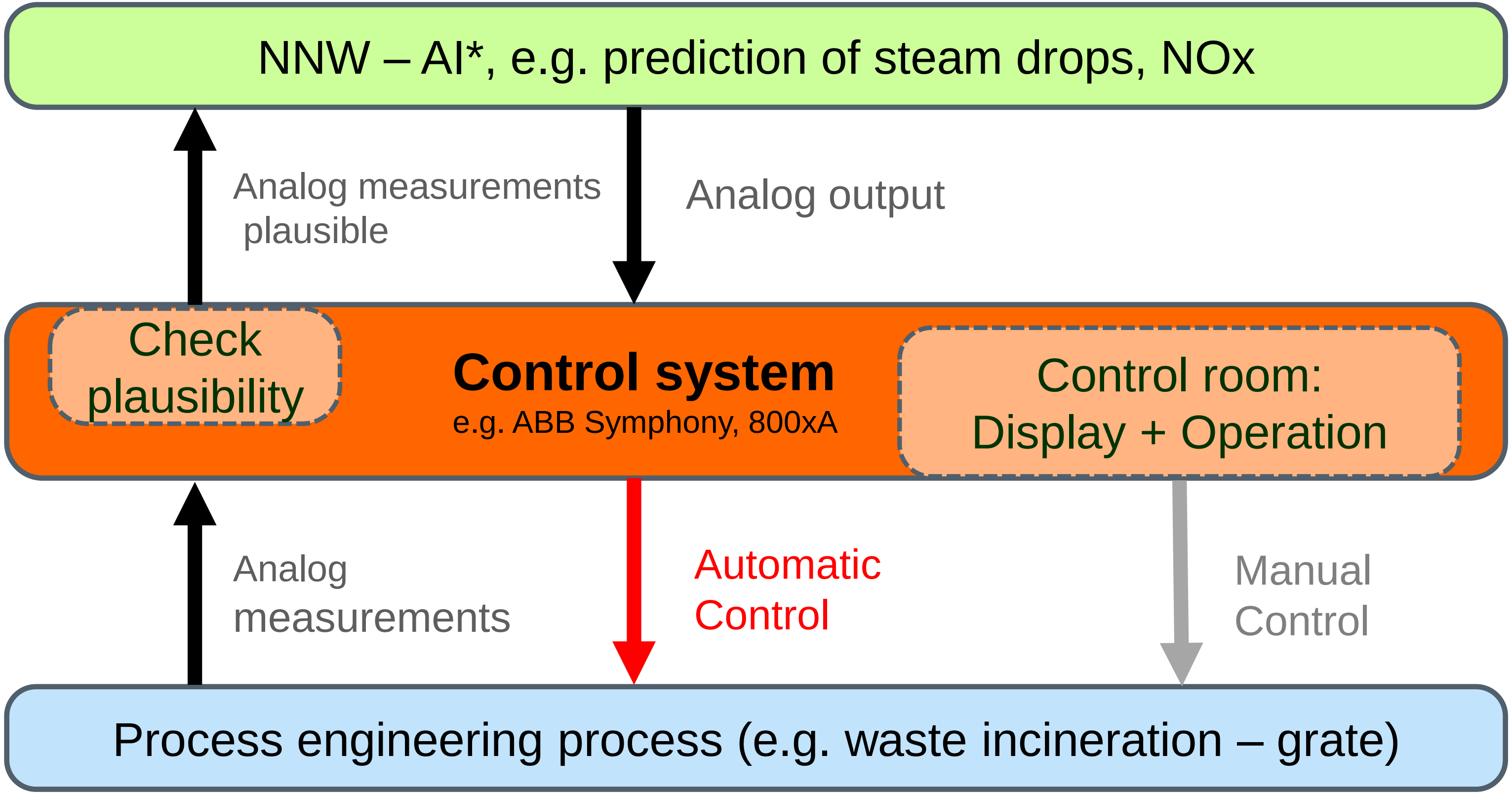
1	2	3	4	5	6
KKS	Eingang	Ausgang	Timestamp	Messwert [Messwert]	Gradient [Messwert]
12HTA10CQ954Q01	1	0	01.11.2023-03:12:36	98,19725800	2,81074524
12HTA10CQ009Q01	1	0	01.11.2023-03:12:36	1,24984741	-0,03890991
10LBA50CT041Q01	1	0	01.11.2023-03:12:36	400,00781250	1,00000000
10LBA50CF041Q02	1	0			
12HTA10CQ002Q01	1	0			
12HTA10CQ003Q01	1	0			
12HTA10CQ007Q01	1	0			
12HTA10CQ954Q01.X1	0	1			
12HTA10CQ954Q01.X2	0	1			
12HTA10CQ954Q01.X3	0	1			

2. Transformation of data into the number range [0; 1]

01-PG_23022024-153653_pattern.pat

1	0.15117043	0.49952223	0.66505734	0.23002498	0.20409163
2	0.47094495	0.48392035	0.49769980	0.48928447	0.37511619
3	0.15117043	0.49946687	0.66496932	0.23010526	0.20413817
4	0.47091336	0.48388655	0.49766705	0.48926124	0.37508642
5	0.15139004	0.49960015	0.66505795	0.23029857	0.20444890
6	0.47082159	0.48372643	0.49752189	0.48924739	0.37503003
7	0.15139004	0.49961478	0.66512000	0.23027736	0.20451720

AI interactions with the control system



Hardware: Siemens S7 system integrator (product available) / ABB Optimax**

** Product development for customer projects

NEURONALNETWORKS!

Frank Gebhardt

CEO (Geschäftsführer)

Frank.Gebhardt@neuronalnetworks.de

+49 177 611 5983

www.neuronalnetworks.de