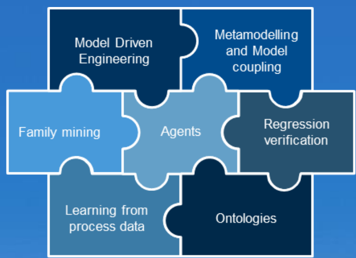
SUMMER SCHOOL
Smart Data Enabled Learning During Operation



Univ.-Prof. Dr.-Ing. Birgit Vogel-Heuser

Ordinaria
Automation and Information Systems (AIS)
Mechanical engineering,
Technische Universität München
www.ais.mw.tum.de; vogel-heuser@tum.de

TUM ASIA SUMMER SCHOOL
24TH – 30TH August 2017

1

Automation
and Information Systems
Technical University of Munich

Agenda – TUM Asia Summer School 29th August

	29th August	30th August
9:00AM – 10:30AM	**Comparison of Industry 4.0, IoT, Smart Factory, Smart Data**	**Case Studies & Successful Demostrators: Applying Enabling Technologies**
10:30AM – 11:00AM	MORNING TEA BREAK	
11:00AM – 12:30PM	**PART I: Enabling Technologies (Agents, Modelling Notations for Automation)**	**Smart Data Enabled Learning During Operation**
12:30PM – 01:30PM	LUNCH BREAK	
01:30PM – 03:00PM	**PART II: Enabling Technologies (Agents, Modelling Notations for Automation)**	**Security and Human in the Loop**

Complete Agenda: https://tum-asia.edu.sg/i4ss/

Automation
and Information Systems
Technical University of Munich

Outline of part smart data as component of Industry 4.0

1. Why is smart data essential for Industry 4.0?
2. Selected case studies
3. Timber production
4. Construction machines- bauflott
5. Devices in process industry-SIDAP
6. Architectures as a basis to access data efficiently
7. Human in the loop – Improve
8. Semiconductor industries
9. Alarm analysis

Automation
and Information Systems
Technical University of Munich

Cyber-Physical Production Systems (CPPS) – Industrie 4.0

Data processing for humans

Assistance systems for Engineering

Data processing and integration for humans

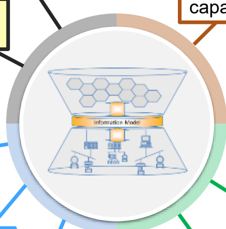
Communication and data consistency

Appropriation of necessary data for configuration, production, negotiation

World wide distribution of data, **high availability, access protection**

Data consistency about different „stakeholders" in different engineering phases and crafts

Digital networks and interfaces for communication (between machine, human and plant, plant and plant)



Architecture models (reference architecture) for a category of aggregation/modules related to properties, capabilities, interfaces…

Intelligent products and production units

Production units with **inherent capabilities**

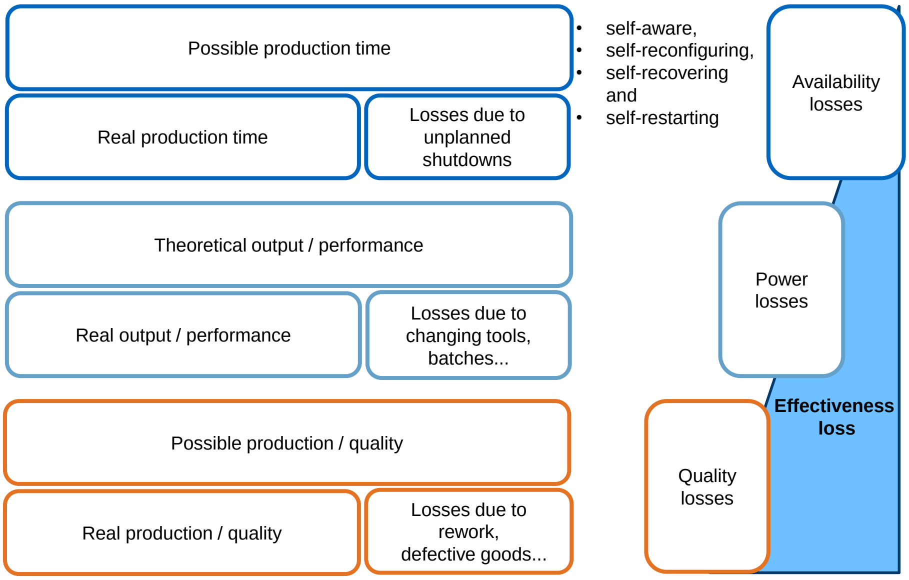
Data analysis of process and alarm data and connection with engineering data

Flexible production units, **adaptable** to modified product requirements, allow also structural changes

Description of product and operating resources, e.g. ontology, for independent analysis, presentation, organisation and execution of a production process

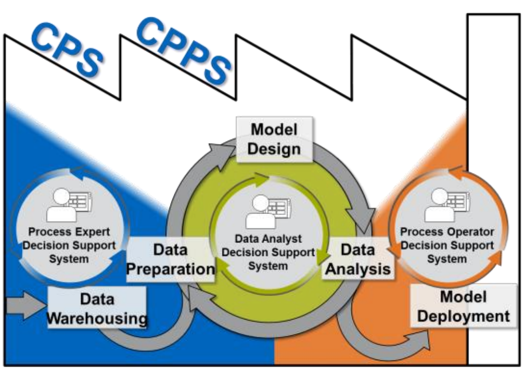
Source: B. Vogel-Heuser, G. Bayrak, U. Frank: Forschungsfragen in "Produktautomatisierung der Zukunft". acatech Materialien. 2012.

Overall equipment effectiveness (OEE) by self-aware, self-reconfiguring, self-recovering and self-restarting



Smart Data in Automated Production Systems

Concept/Workflow of Smart Data in CPS

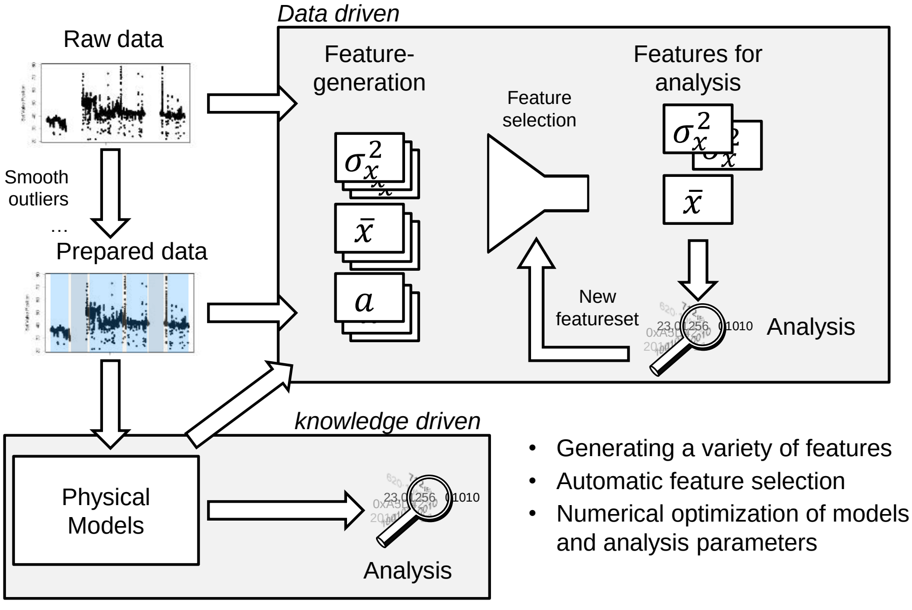


Research Concept:

- **Objective:** Improvement of OEE (Overall Equipment Effectiveness)
- **Challenges**: Unstructured information, heterogeneous data sources, missing meta-models to represent data and inefficient system architecture
- **Methods:** Big Data algorithms, inquiry and integration of expert knowledge
- **Results:** Decision models, data and system architectures, decision support systems

Automation
and Information Systems
Technical University of Munich

Data preprocessing / analysis



- Generating a variety of features
- Automatic feature selection
- Numerical optimization of models and analysis parameters

Automation
and Information Systems
Technical University of Munich

Outline of part smart data as component of Industry 4.0

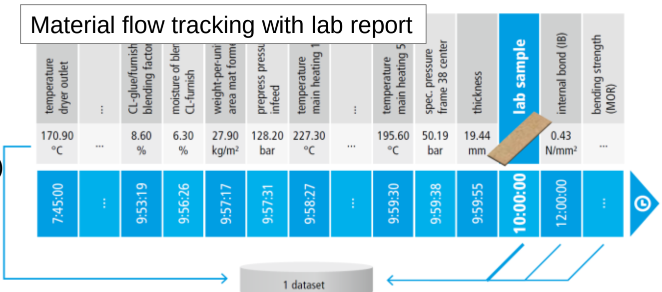
1. Why is smart data essential for Industry 4.0?
2. Selected case studies
3. Timber production
4. Construction machines- bauflott
5. Devices in process industry-SIDAP
6. Architectures as a basis to access data efficiently
7. Human in the loop – Improve
8. Semiconductor industries
9. Alarm analysis

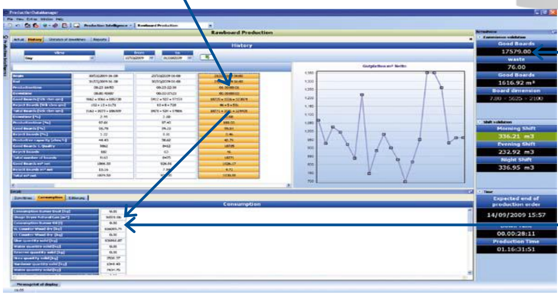
Automation
and Information Systems
Technical University of Munich

OEE: Quality loss-
Quality forecast in the fibreboard manufacture

Historical Information:
- Commision
- Shift
- Day
- Month

(exportable to Excel)
+ generate reports





Current information<

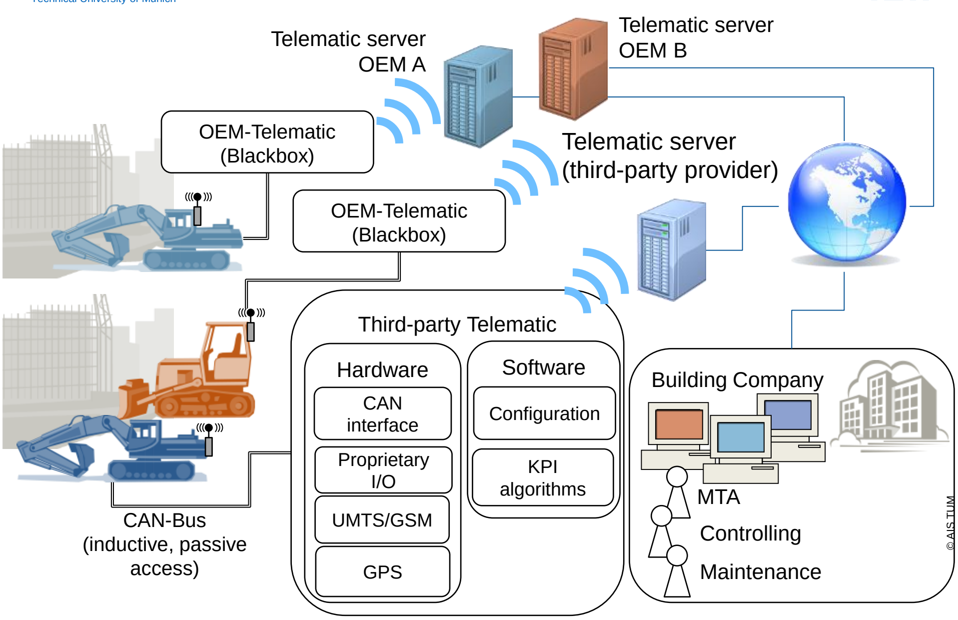
Details:
- Downtimes
- Consumptions
- …

Editor for corrections

Source: Siempelkamp Maschinen- und Anlagenbau GmbH & Co. KG, Prod-IQ

Automation
and Information Systems
Technical University of Munich

Motivation: current state of telematics in civil engineering



AEMP and ISO standardisation of current KPI set

Schedule

International Organization for Standardization

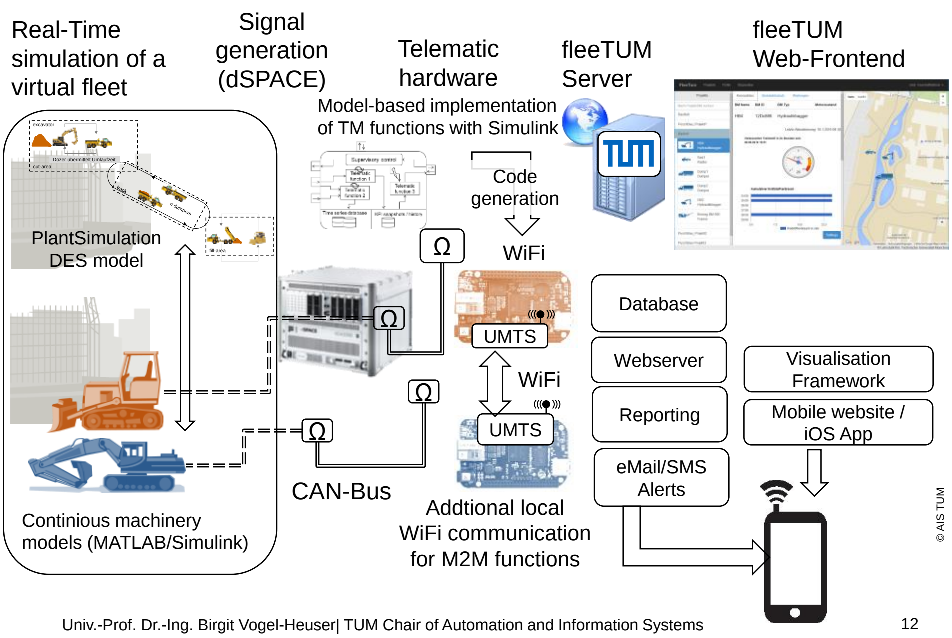
- **09/15**
  Changed draft sent to ISO TC 127 working group 5
- **10/15**
  Meeting of ISO ISO TC 127 working group 5;
  Application of changes and forwarding to ISO TC 127/SC2 Committee
- **11/15**
  Representing countries in ISO TC 127/SC2 Comitee have 90 days for acceptence or revision call
- **03/16**
  After acceptance final approval and certification for ISO standard

Englisch	AEMP v1.2	ISO 15143-3
Equipment information	x	x
Last know location	x	x
Cumulative operating hours	x	x
Cumulative fuel used	x	x
Fuel used in the preceding 24 hours	x	x
Cumulative distance travelled	x	x
Cumulative idle operating hours		x
Fuel remaining ratio		x
Is engine running		x
Digital input state		x
Cumulative power take-off hours		x
Average daily engine load factor		x
Peak Daily Speed for past 24 hours		x
Cumulative Load Count		x
Cumulative Payload Totals		x
Cumulative nonproductive regeneration hours		x
Diagnostic trouble codes		x
Caution code		x
DEF remaining ration		x
Cumulative idle nonoperating hours		x

Italic: Change of Description

Bold: New KPI

Onboard-/Hardware-in-the-Loop evaluation based on a virtual machinery fleet



Outline of part smart data as component of Industry 4.0

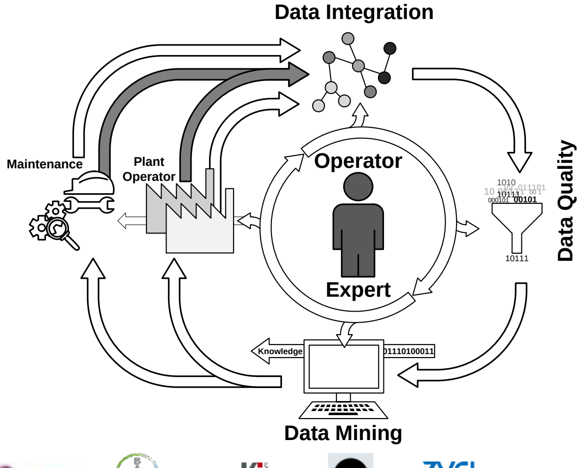
1. Why is smart data essential for Industry 4.0?
2. Selected case studies
3. Timber production
4. Construction machines- bauflott
5. Devices in process industry-SIDAP
6. Architectures as a basis to access data efficiently
7. Human in the loop – Improve
8. Semiconductor industries
9. Alarm analysis

Automation
and Information Systems
Technical University of Munich

Overview

Gefördert durch:
Bundesministerium für Wirtschaft und Energie
aufgrund eines Beschlusses des Deutschen Bundestages

Cross-company data integration and analysis:



Research Concept

- **Objective:** Improvement of OEE (Overall Equipment Effectiveness)
- **Challenges:** "Learning" from different data sources (development, operation, maintenance), different plants and between manufacturer, engineering, operation and maintenance
- **Methods:** Data mining algorithms, data and system architectures, cloud technologies
- **Results:** Prediction of faulty conditions by signal- and model-based methods of diagnosis

www.ais.mw.tum.de/en/research/current-research-projects/sidap/

Project: #SmartData2015 / Data Mining in process industry

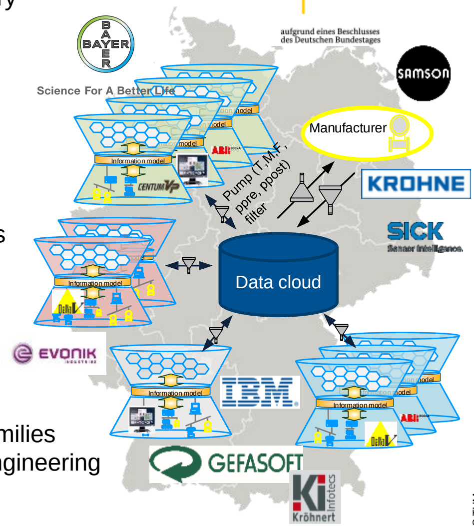
Gefördert durch:
Bundesministerium für Wirtschaft und Energie
aufgrund eines Beschlusses des Deutschen Bundestages

OEE availability loss targeted

Data logistics

- Secure provision and transport
- Secure storage
- Data model

- **Aggregation and analysis of data**
  - Identification of unknown correlations in data
  - Integration of field device manufacturers



- **Data use**
  - Application of the findings to plant families
  - Supporting operating personnel in engineering and maintenance

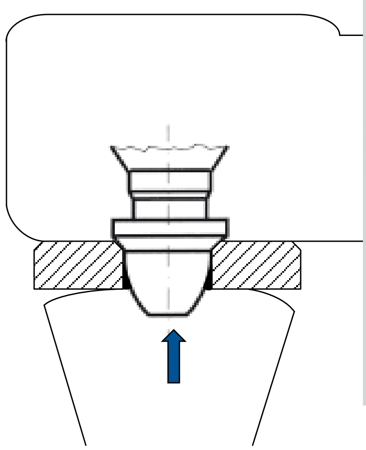
https://www.ais.mw.tum.de/en/research/current-research-projects/sidap/

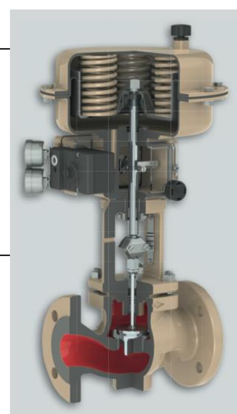
Definition of the task

Error identification of the fault images –
Cone wear and valve closure of control valves in historical usage data

Error images to be identified:

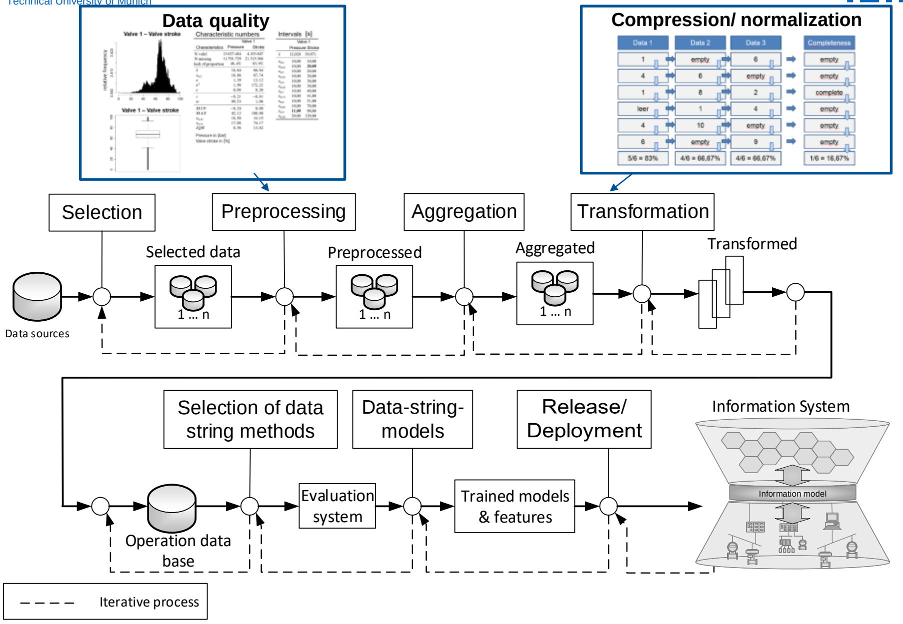
- Cone wear (KV)
- Valve Closure (VV)





Automation
and Information Systems
Technical University of Munich

Workflow for string data in process industry

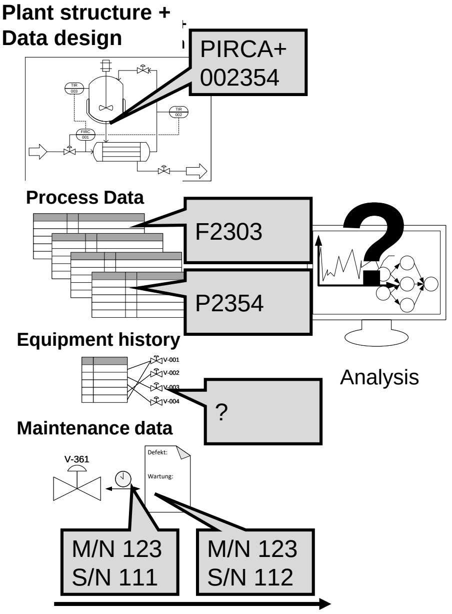


© AIS TUM

Automation
and Information Systems
Technical University of Munich

Data model - Initial Situation

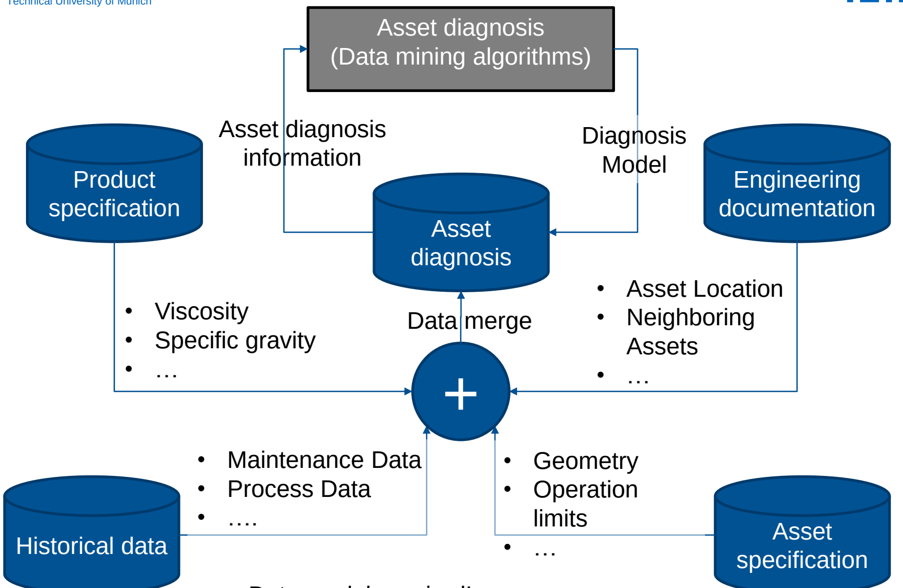
- Variety of different data sources
  - Different locations *(several databases, different sites)*
  - Heterogeneous data formats *(relational databases with different conventions, csv, XML…)*
  - Different names for the same equipment *(semantic gap)*
  - Partially not digital *(e.g. maintenance data on paper)*
  - **Very heterogeneous data base!**
- Hands-on, complex integration of the necessary data
- Data analysis is currently limited
  - No automatic analysis possible *(Manual, specific configuration of analysis per valve)*
  - Always expert knowledge necessary *(Identification of relevant sources, data integration, interpretation of data)*
  - Missing data concerning the representation of the full asset lifecycle



© AIS TUM

Automation
and Information Systems
Technical University of Munich

Data logistics and data model

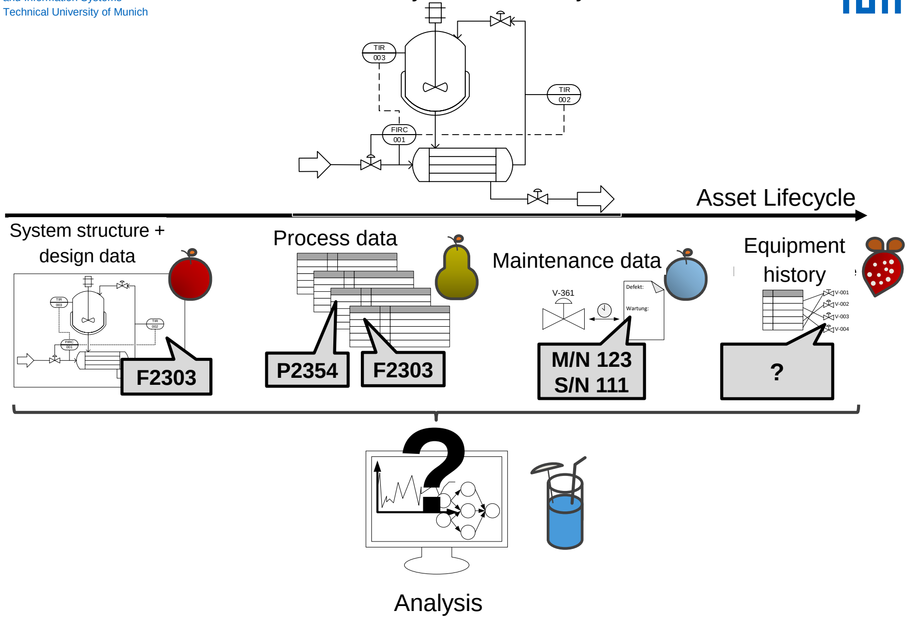


Data model required!

© AIS TUM

Automation
and Information Systems
Technical University of Munich

Data diversity – use for analysis?

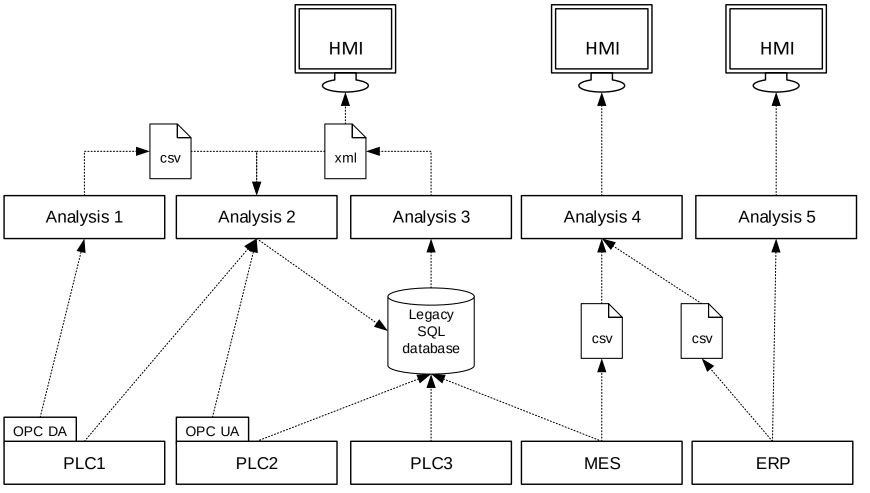


© AIS TUM

Current System Layout

Characteristics of current systems

- Large number of legacy applications and interfaces
- Large number of **Point-to-Point** interfaces (low cross-connectivity, high maintenance costs)
- Reuse of analysis results often only via file export (slow, non-automated)

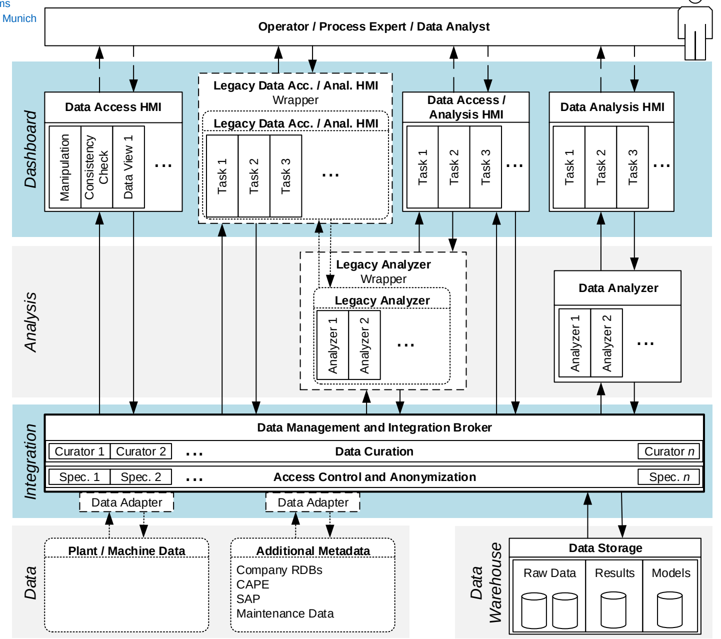


Outline of part smart data as component of Industry 4.0

1. Why is smart data essential for Industry 4.0?
2. Selected case studies
3. Timber production
4. Construction machines- bauflott
5. Devices in process industry-SIDAP
6. Architectures as a basis to access data efficiently
7. Human in the loop – Improve
8. Semiconductor industries
9. Alarm analysis

Automation
and Information Systems
Technical University of Munich

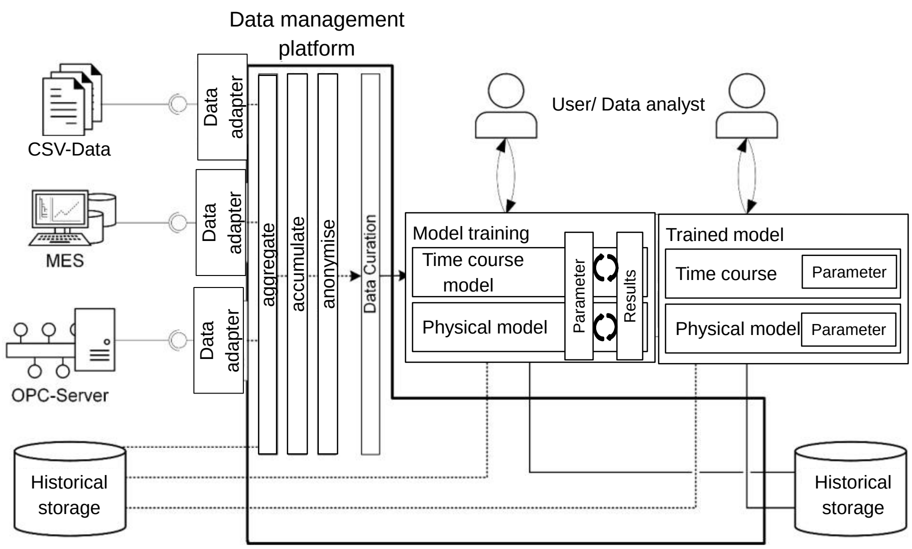
Cloud concept - architectural proposal



© AIS TUM

Automation
and Information Systems
Technical University of Munich

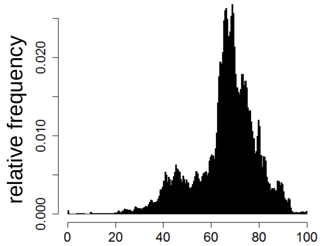
Cloud concept - architectural proposal



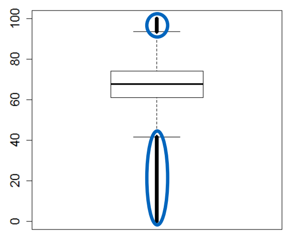
© AIS TUM

A real issue: Data quality

Valve 1 – Valve stroke



Valve 1 – Valve stroke



Characteristic numbers

	Valve 1	
Characteristics	Pressure	Stroke
N valid	13.627.484	4.103.847
N missing	11.791.729	21.315.366
lack of proportion	46,4%	83,9%
$\bar{x}$	16,64	66,04
$x_{0.5}$	16,86	67,74
σ	1,39	13,12
σ^2	1,98	172,21
v	0,08	0,20
s	−9,21	−0,81
w	99,53	1,06
MIN	−0,24	0,00
MAX	42,12	100,00
$x_{0.25}$	16,50	61.15
$x_{0.75}$	17,08	74,17
IQR	0,56	13,02

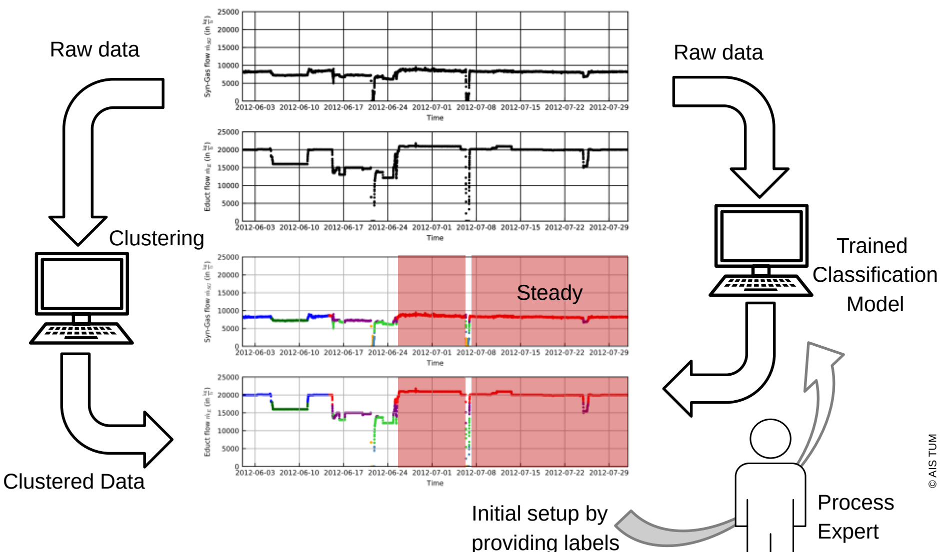
Pressure in [bar] Valve stroke in [%]

Intervals [s]

	Valve 1	
	Pressure	Stroke
$\bar{x}$	12,028	39,971
$x_{0.4}$	10,00	10,00
$x_{0.45}$	10,00	**20,00**
$x_{0.5}$	10,00	20,00
$x_{0.55}$	10,00	20,00
$x_{0.6}$	10,00	30,00
$x_{0.65}$	10,00	30,00
$x_{0.7}$	10,00	40,00
$x_{0.75}$	10,00	41,00
$x_{0.8}$	10,00	51,00
$x_{0.85}$	10,00	70,00
$x_{0.9}$	**11,00**	90,00
$x_{0.95}$	20,00	120,00

Automatic classification of operational states

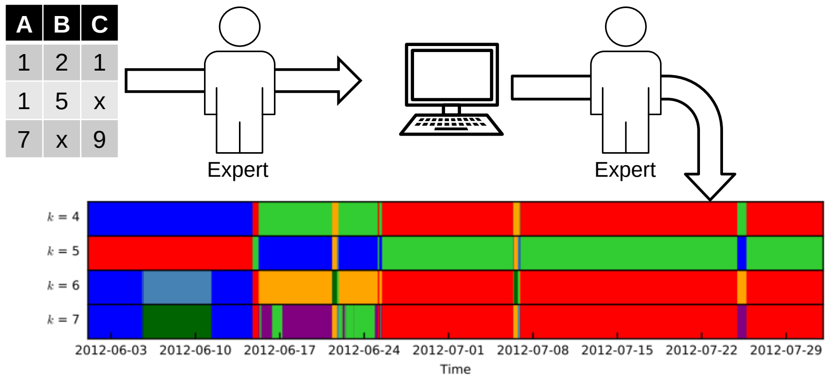
Determination of operational states to increase process understanding and filter data.



Automation
and Information Systems
Technical University of Munich

Operational Phase Classification
K-Means / k-NearestNeighbours

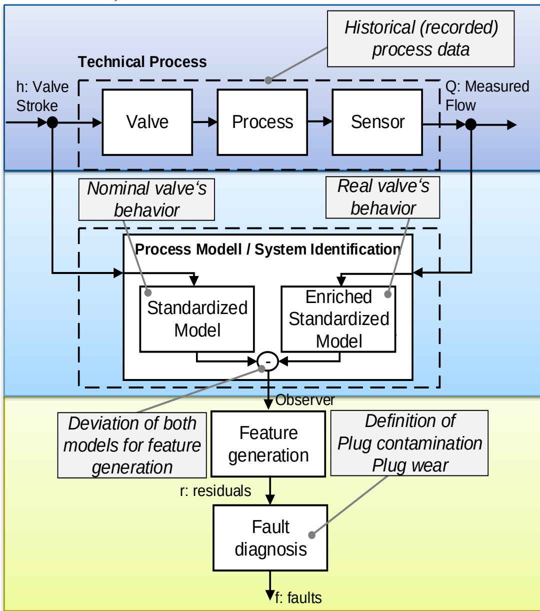
Include data series identified by expert to be relevant for analyis



Identified, distinct operational phases of the underlying process
Process knowledge is necessary:
- Filter raw data to include relevant fields and derive variables (ratios, trends, standard deviations)
- Suitable number of clusters?

Automation
and Information Systems
Technical University of Munich

Valve diagnosis based on model-based and signal-based approaches – comparative study in progress



Data Acquisition / Curation
- Relevant Process Data
- Valve / Plant Specification
- Relations of Field Devices

Process Modell
- Model based valve Diagnosis (including technical data and process data)

System Identification
- Signal based valve Diagnosis (including process data)

Valve Diagnosis
- Detection of plug contamination and wear based on different diagnosis methods
- *In progress: comparative study*

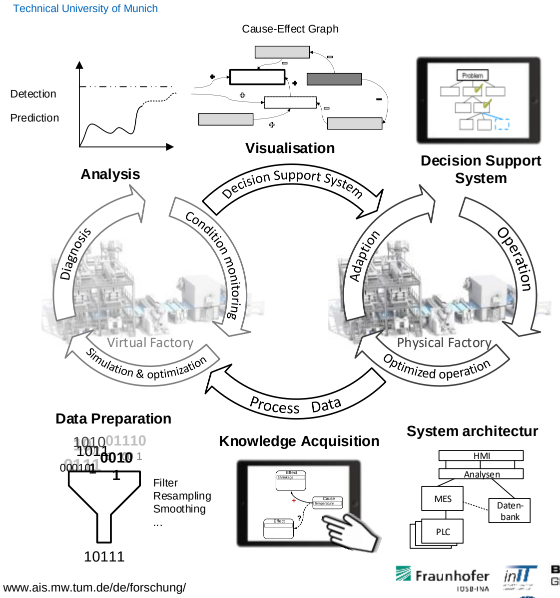
Automation
and Information Systems
Technical University of Munich

Outline of part smart data as component of Industry 4.0

1. Why is smart data essential for Industry 4.0?
2. Selected case studies
3. Timber production
4. Construction machines- bauflott
5. Devices in process industry-SIDAP
6. Architectures as a basis to access data efficiently
7. Human in the loop – Improve
8. Semiconductor industries
9. Alarm analysis

Automation
and Information Systems
Technical University of Munich

Quality prediction and Fault detection



IMPROVE
www.improve-vfof.eu

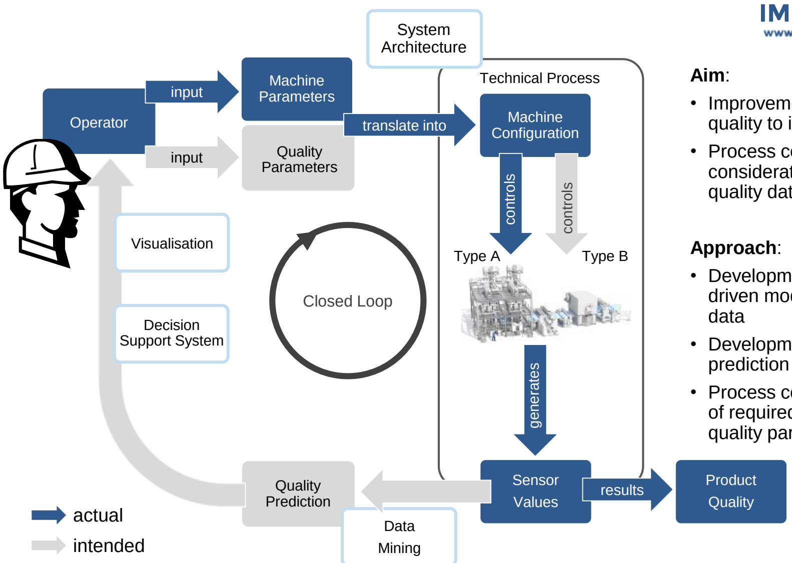
Research concept

- **Objectives:** Identification of data-driven diagnosis and optimization + development of decision support system
- **Challenges**: Development of system architecture, which enables intelligent data management and analytics
- **Approach:** Virtual factory for diagnosis and optimization + development of decision support system for back coupling to the physical factory
- **Results:** Data management platform, algorithms and models, human machine interface

www.ais.mw.tum.de/de/forschung/
aktuelle-forschungsprojekte/improve-eu-project

Automation
and Information Systems
Technical University of Munich

Improving Models of Production Systems by Data Mining of Production Parameters



Aim:
- Improvement of product quality to increase OEE
- Process control under consideration of product quality data (laboratory)

Approach:
- Development of data-driven models by historic data
- Development of prediction models
- Process control by input of required product quality parameters

© AIS TUM

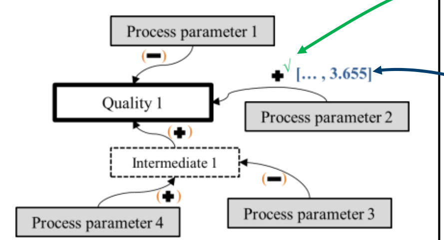
Automation
and Information Systems
Technical University of Munich

Human in the Loop and Data Mining

Expert Knowledge Acquisition through Interviews

=> Mental Models representing the cause and effect relation of process and quality parameters of aPS

Cause-Effect Graph (Excerpt)



+ positive relation √ Confirmed by data analysis
− negative relation ( ) Not (yet) confirmed by data analysis

Data Mining

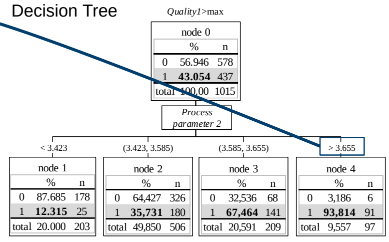
Logistic Regression

Estimator for Quality1>max

	β_j	S.E.	Sig.
absolute term	-44.379	3.225	0.000
Process parameter 1	0.041	0.405	0.920
Process parameter 2	12.285	0.987	0.000
Intermediate 1	0.256	0.350	0.507

S.E. = Standard Error, Sig. = Significance

Decision Tree



© AIS TUM

Source: B. Vogel-Heuser, V. Karaseva, J. Folmer, I. Kirchen. "Operator Knowledge Inclusion in Data-Mining Approaches for Product Quality Assurance using Cause-Effect Graphs," in 20th IFAC World Congress, July 2017.

Automation
and Information Systems
Technical University of Munich

Outline of part smart data as component of Industry 4.0

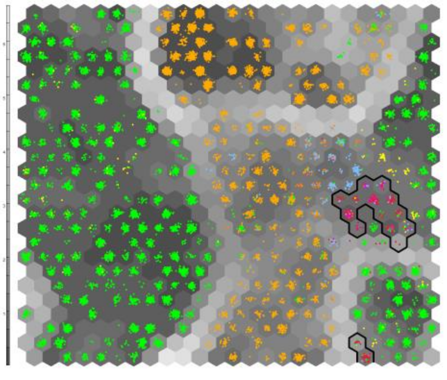
1. Why is smart data essential for Industry 4.0?
2. Selected case studies
3. Timber production
4. Construction machines- bauflott
5. Devices in process industry-SIDAP
6. Architectures as a basis to access data efficiently
7. Human in the loop – Improve
8. Semiconductor industries
9. Alarm analysis

Automation
and Information Systems
Technical University of Munich

detection of process disturbances

Self Organizing Map for fault detection

U-Matrix incl. state assignment



- Normal State
- Shutdown
- Startup/Cutback
- Stabilizing
- Purging
- Fault Type 1
- Fault Type 2

Detection of faults by clustering based on process values

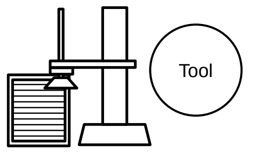
Alarm, if process values similar to a state with faults

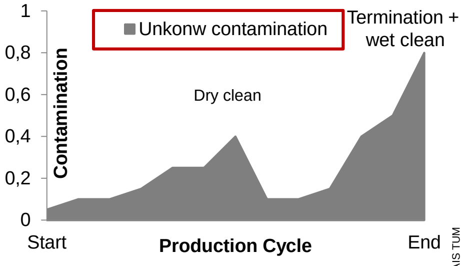
Introduction - Plant and Problem Description

Analyzed system
- Hard masking deposition tool for semiconductor manufacturing

Current maintenance
- **Tool contaminates with every produced wafer**
- Chamber cleaned after fixed number of wafers, maximum amount of time, or based on laboratory results
- **Wet cleans** (after production cycle) and **dry cleans** to maintain chamber

Problem
- Lack of information to **estimate current contamination**
- Lack of information about **reasons for production cycle termination** and wet cleans
- Lack of information about **dry cleans**





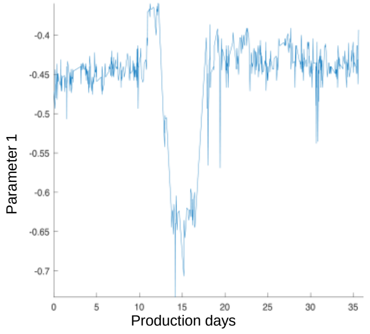
Automation
and Information Systems
Technical University of Munich

Introduction – Research Goal

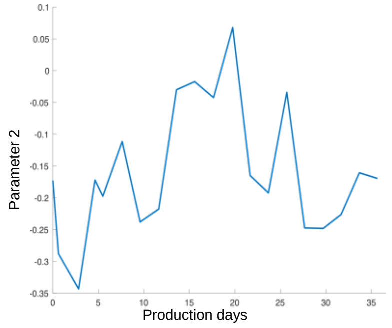
Research Goal
- Cluster sensor data to estimate contamination
- Unsupervised learning due to missing contamination information

Requirements
- Lack of process knowledge
- Dynamic sensor behavior (no recognizable trend in sensor data over production cycle)
- Dynamic sampling rate (varying over time and between sensors)

Dynamic parameter behavior, varying sampling rates



Parameter 1: high sampling rate



Parameter 2: low sampling rate

Automation
and Information Systems
Technical University of Munich

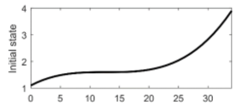
Developed Method

1. Data Preprocessing

- Feature suitability for different Production Cycles with varying value ranges has to be considered when choosing statistical values
- Only **statistical** values **not depending** on a production cycle's value range are used
- Ideal Sliding Window length for a variable cannot be chosen manually because data variables are recorded in a very dynamic manner

2. Data Mining

- **Step 1:** Parameter Initialization
  - **Non-random initialization**

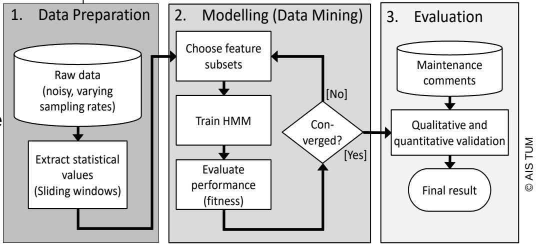


Initial, continuous state deterioration

- **Step 2:** calculate probability of the contamination state p_{fc}
  - If p_{fc} is **high**, a contaminated state requiring a cleaning maintenance is assumed
  - If p_{fc} is **low**, the tool is assumed to be in a non-contaminated condition

3. Feature Selection

- Genetic Algorithm evaluates a feature subset's performance w.r.t. its maintenance prediction capability
- suitable feature subset for maintenance prediction should allow Hidden Markov Models (HMM) to estimate
  - **a low contamination probability** for the beginning and
  - a **higher contamination probability** for the end of all considered production cycles



Automation
and Information Systems
Technical University of Munich

Results and Validation- Estimated Contamination Probabilities

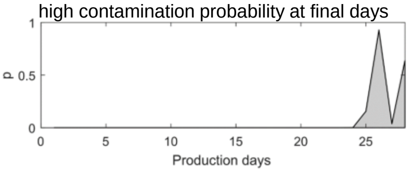
Contamination probability estimated for 8 production cycles

- HMM trained with **6 cycles**
- Cross validation

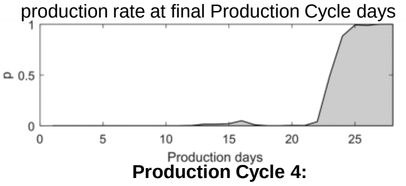
Similar contamination estimation for **cycles without additional cleaning** (only cleaned after termination, first 3 plots)

Varying contamination probability explainable with **additional maintenances** (comments) (last plot, terminated due to amount of produced wafers)

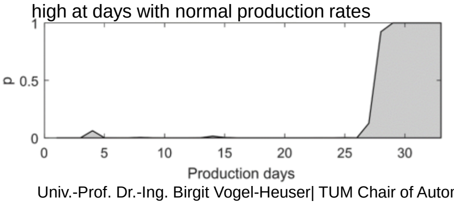
Production Cycle 1:
expected contamination probability distribution, high contamination probability at final days



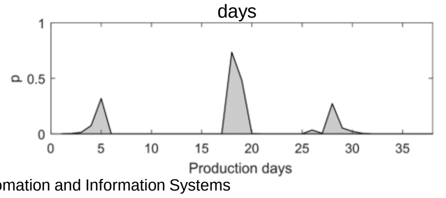
Production Cycle 2:
expected contamination probability distribution, low production rate at final Production Cycle days



Production Cycle 3:
Unexpected contamination probability distribution, very high at days with normal production rates

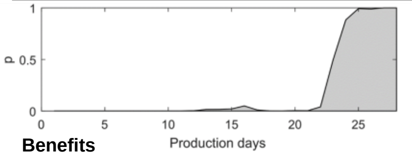


Production Cycle 4:
bad contamination probability distribution, not explainable with expected contamination deterioration or production days

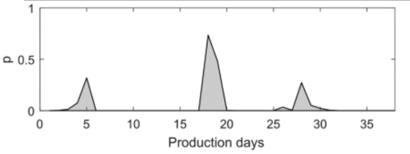


Conclusion

Production Cycle 2:
expected contamination probability distribution, low production rate at final Production Cycle days



Production Cycle 4:
bad contamination probability distribution, not explainable with expected contamination deterioration or production days



Benefits
- Contamination probability estimated based on **dynamic data patterns**
- Results show high contamination probabilities before maintenances
- suitable for industrial data with asynchronous sampling rates
- does not require extensive system or process knowledge

Shortcomings
- Contamination sometimes already high days before cycle termination
- **Maintenance comments** cannot be used for learning
  (manually entered comments, not standardized, hard to interpret)

Conclusion
- sliding windows are introduced transforming the raw data into features suitable for data-mining
- genetic algorithm chooses a feature subset suitable for maintenance prediction, which is clustered by a hidden Markov model
- Results show, that **data patterns suitable to estimate contamination probability** exist

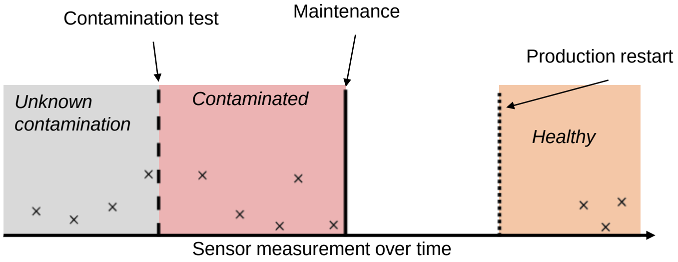
Outlook: Current Work on New Dataset

New dataset of semiconductor production plant available
- Additional information about maintenance time points
- maintenance reasons
- quality parameters

Currently developed approach
- Most data points can be labeled as healthy / not contaminated
- Few contaminated data points

→New supervised classification approach for unbalanced dataset

Known contamination classes can improve the datamining results
→ classification approach for unbalanced data currently tested



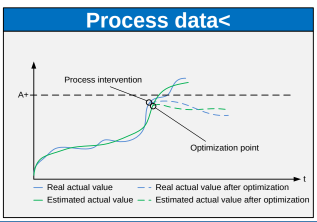
Automation
and Information Systems
Technical University of Munich

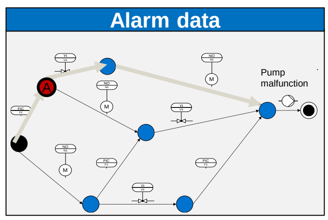
Outline of part smart data as component of Industry 4.0

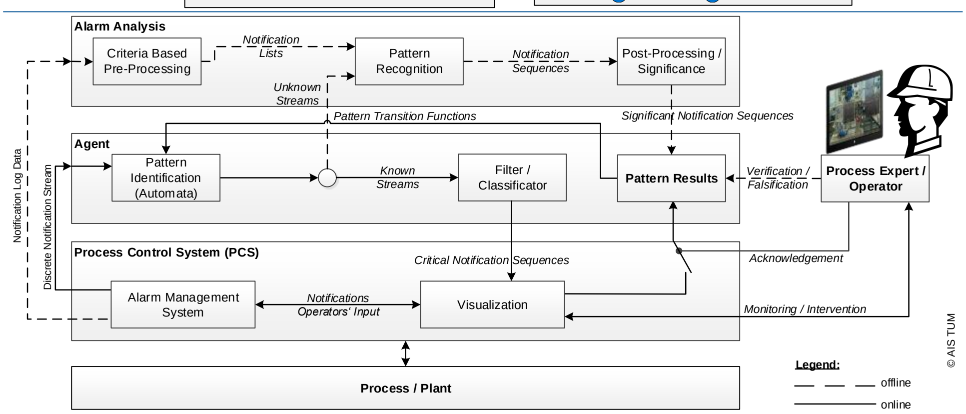
1. Why is smart data essential for Industry 4.0?
2. Selected case studies
3. Timber production
4. Construction machines- bauflott
5. Devices in process industry-SIDAP
6. Architectures as a basis to access data efficiently
7. Human in the loop – Improve
8. Semiconductor industries
9. Alarm analysis

Automation
and Information Systems
Technical University of Munich

Data analysis – Identification von causal message patterns





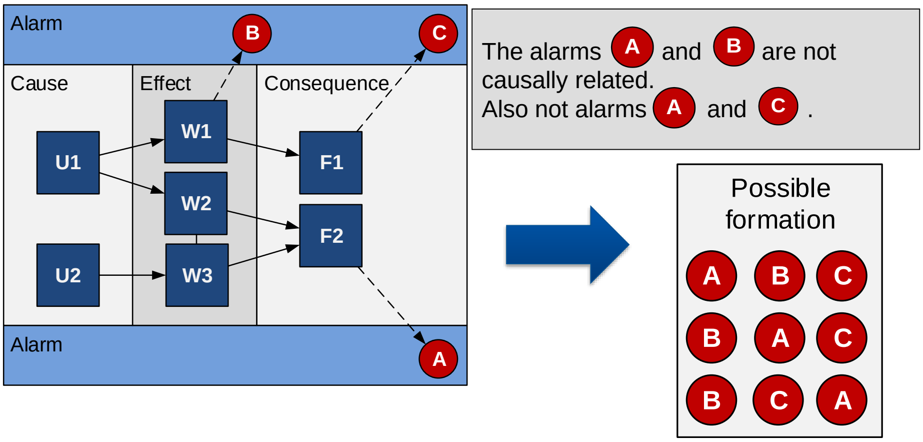


Source: Vogel-Heuser, B. et al.: *Criteria-based Alarm Flood Pattern Recognition using Historical Data from Automated Production Systems (aPS)*. In: Journal Mechatronics, 1-12, 2015. – in press

Automation
and Information Systems
Technical University of Munich

Dependent/ independent reports

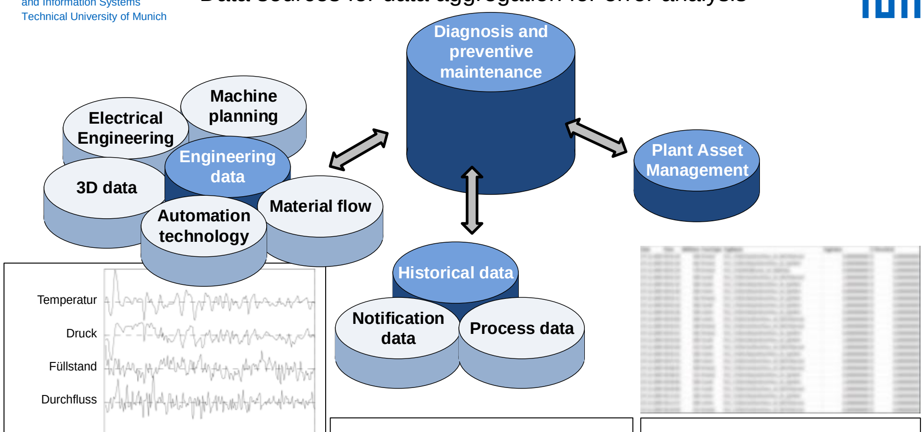
Alarms are located at different positions in an action chain



- The messages are displayed sequentially to the operator or Logfile-System
- The messages may occur in different sequential arrangements

Automation
and Information Systems
Technical University of Munich

Data sources for data aggregation for error analysis



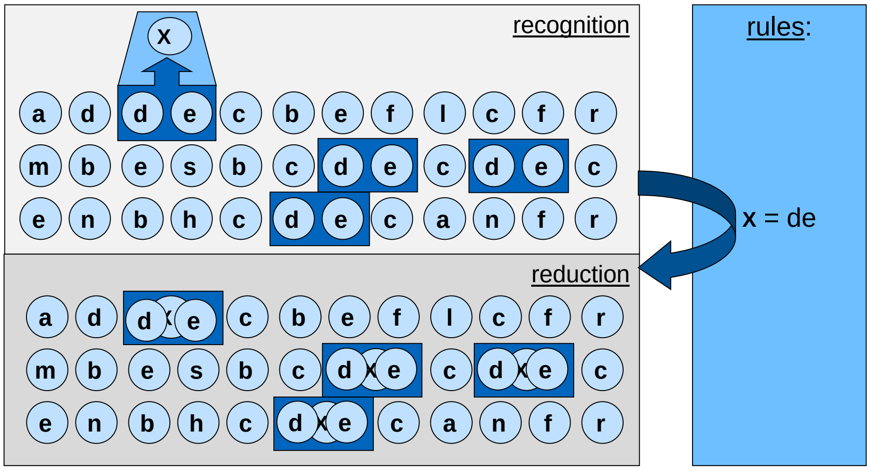
Process data
Analysis of process data to generate process data vectors to predict errors (classification)

Combined consideration:
Finding causal dependencies by combined analysis of alarm and process data (How is the process data flow before / after a specific alarm

Alarm data:
Analysis of alarm data for finding causal dependent message chains

Automation
and Information Systems
Technical University of Munich

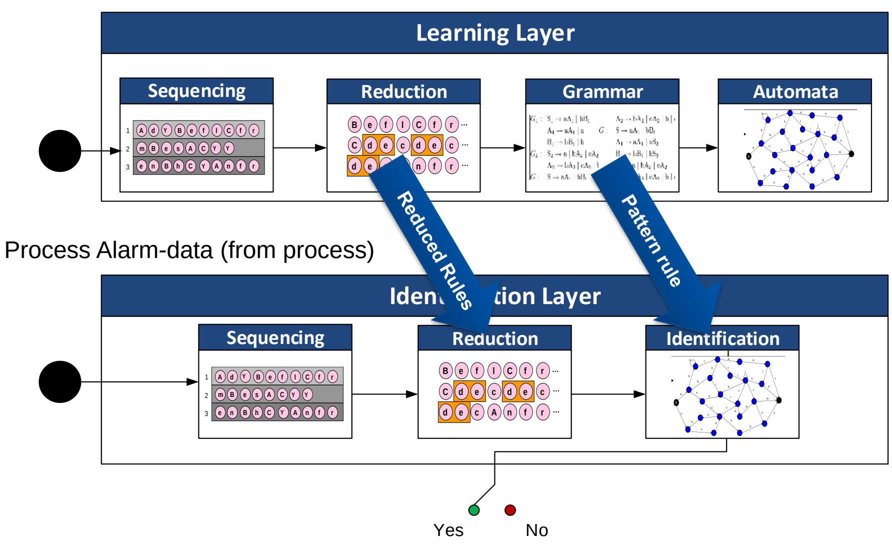
Reduction (1)



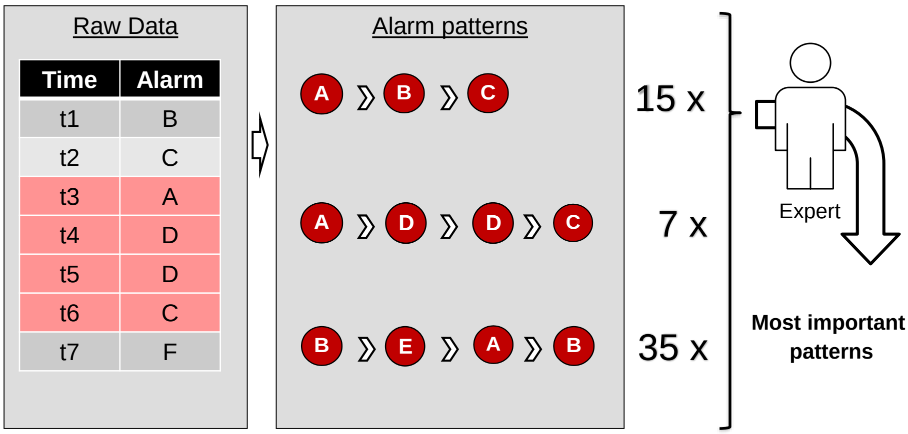
Automation
and Information Systems
Technical University of Munich

Application in operation

Alarm-data (e.g. archive)



Concept evaluation



Summary for alarm analysis

Reduction of alarm shiver / report shiver
Finding causal connections
Mapping the causal relationships in the formal model
Reduction of the alarm screens by hiding (redundant) messages

Calculation of a time-dependent failure rate during operating state
Time-dependent failure rate should correspond better to reality than the constant failure rate
Should support the operator in the decision-making process
- Early replacement of component
- Support in decision-making in case of error

Visualization to relieve the operator
Reduction of Operator Load by Display:
- The actual cause
- Of the MTBF-value for individual devices

Further References

B. Vogel-Heuser, D. Schütz and J. Folmer. "Criteria-based Alarm Flood Pattern Recognition using Historical Data from Automated Production Systems (aPS)," Mechatronics, vol. 31, pp. 89-100, Mar. 2015. (available online free of charge)

E. Trunzer, I. Kirchen, J. Folmer, G. Koltun and B. Vogel-Heuser. "A Flexible Architecture for Data Mining from Heterogeneous Data Sources in Automated Production Systems," in *IEEE International Conference on Industrial Technology (ICIT) - 2017*, Mar. 2017, pp. 1106-1111.

B. Vogel-Heuser, V. Karaseva, J. Folmer and I. Kirchen. "Operator Knowledge Inclusion in Data-Mining Approaches for Product Quality Assurance using Cause-Effect Graphs," in *20th IFAC World Congress (IFAC)*, Jul. 2017.

J. Folmer, C. Schrüfer, J. Fuchs, C. Vermum and B. Vogel-Heuser. "Data-Driven Valve Diagnosis to Increase the Overall Equipment Effectiveness in Process Industry," in *14th IEEE International Conference on Industrial Informatics (INDIN)*, IEEE, Jul. 2016, p. 6

J. Kinghorst, O. Geramifard, M. Luo, H. Chan, K. Yong, J. Folmer and B. Vogel-Heuser. "Hidden Markov Model-Based Predictive Maintenance in Semiconductor Manufacturing: A Genetic Algorithm Approach," in *13th Conference on Automation Science and Engineering (CASE 2017)*, Aug. 2017.

Thank you for your attention!

Univ.-Prof. Dr.-Ing. Birgit Vogel-Heuser

Ordinaria
Automation and Information Systems (AIS)
Mechanical engineering,
Technische Universität München
www.ais.mw.tum.de; vogel-heuser@tum.de