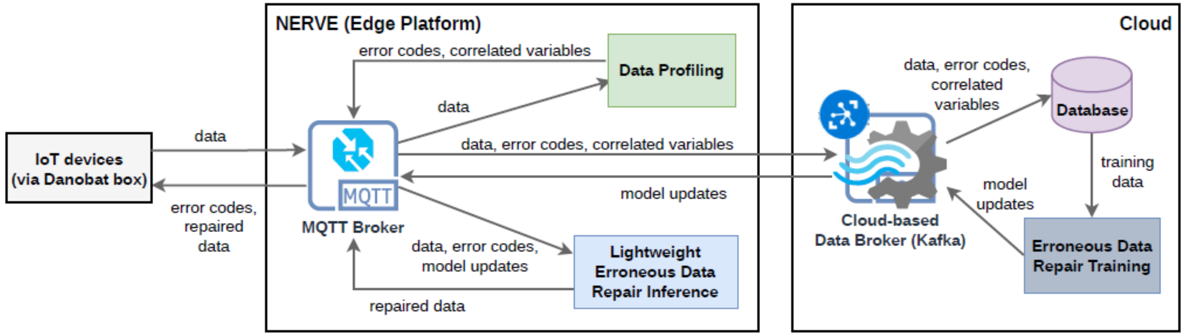
Ensuring Data Quality for the Quality of Product, Process towards Zero-Defect Manufacturing

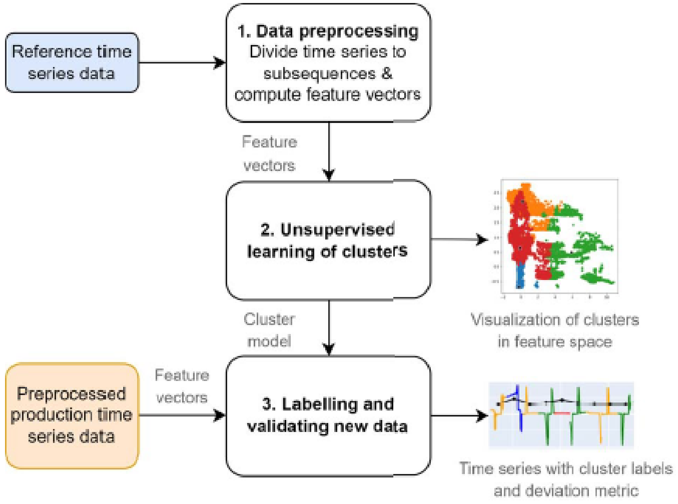
Adela Nedisan Videsjorden, Phu Nguyen
(SINTEF Digital)

DAT4.ZERO

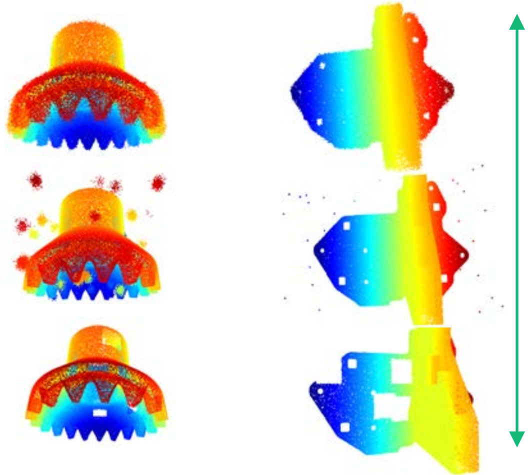
1) Data Quality on Edge



2) UDAVA



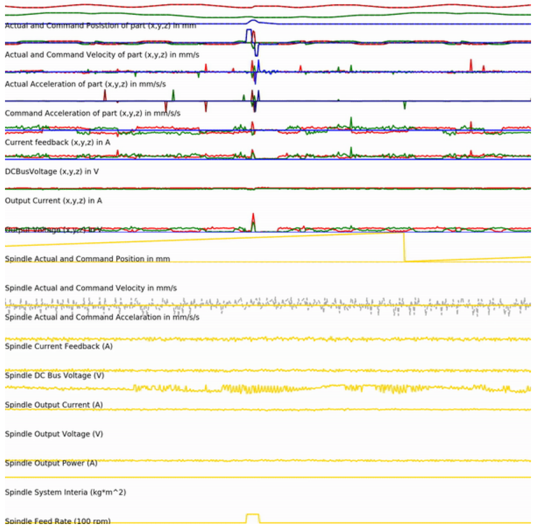
3) 3D-DaVa :
3D Point Cloud
Data Validation



Data are driving the world and data quality must be the driving license!

Many things could go wrong with data being collected …

How is **data acquired** during manufacturing?



High-frequency multivariate sensor data

Several KHz

Summarization

(Frequencies, severity, skislope 1 Hz)

But has data quality been neglected so far?

What are the existing approaches addressing data quality and its management?

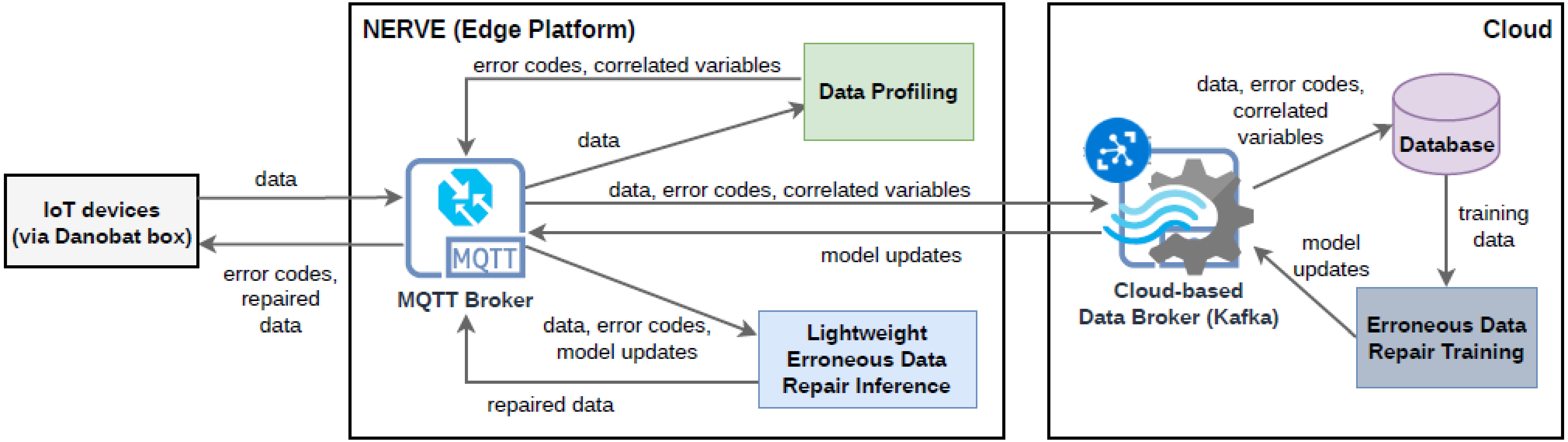
Some highlights from a Systematic Review of Data Quality in CPS and IoT for Industry 4.0

- A variety of data quality issues were addressed: outliers, missing values, and data veracity are the three main ones.
- Not many studies discuss the source of data quality issues and the implications of different computing architectures for data quality issues.
- Existing data quality management techniques do not support deployment on different IoT layers for online and offline scenarios. Future techniques should be able to run on a standalone machine, edge device, or the Cloud, with data access to support online and offline data repair and cleaning on the edge and in the Cloud.

Arda Goknil, Phu Nguyen, Sagar Sen, Dimitra Politaki, Harris Niavis, Karl John Pedersen, Abdillah Suyuthi, Abhilash Anand, and Amina Ziegenbein. 2023. A Systematic Review of Data Quality in CPS and IoT for Industry 4.0. *ACM Comput. Surv.* 55, 14s, Article 327 (July 2023), 38 pages. https://doi.org/10.1145/3593043

1) Data Quality on Edge

- Leveraging the computational abilities of edge devices enables data profiling and repair tasks at the edge, allowing for immediate remediation of erroneous data within the data stream and improved scalability through distributed repair across multiple edge devices.

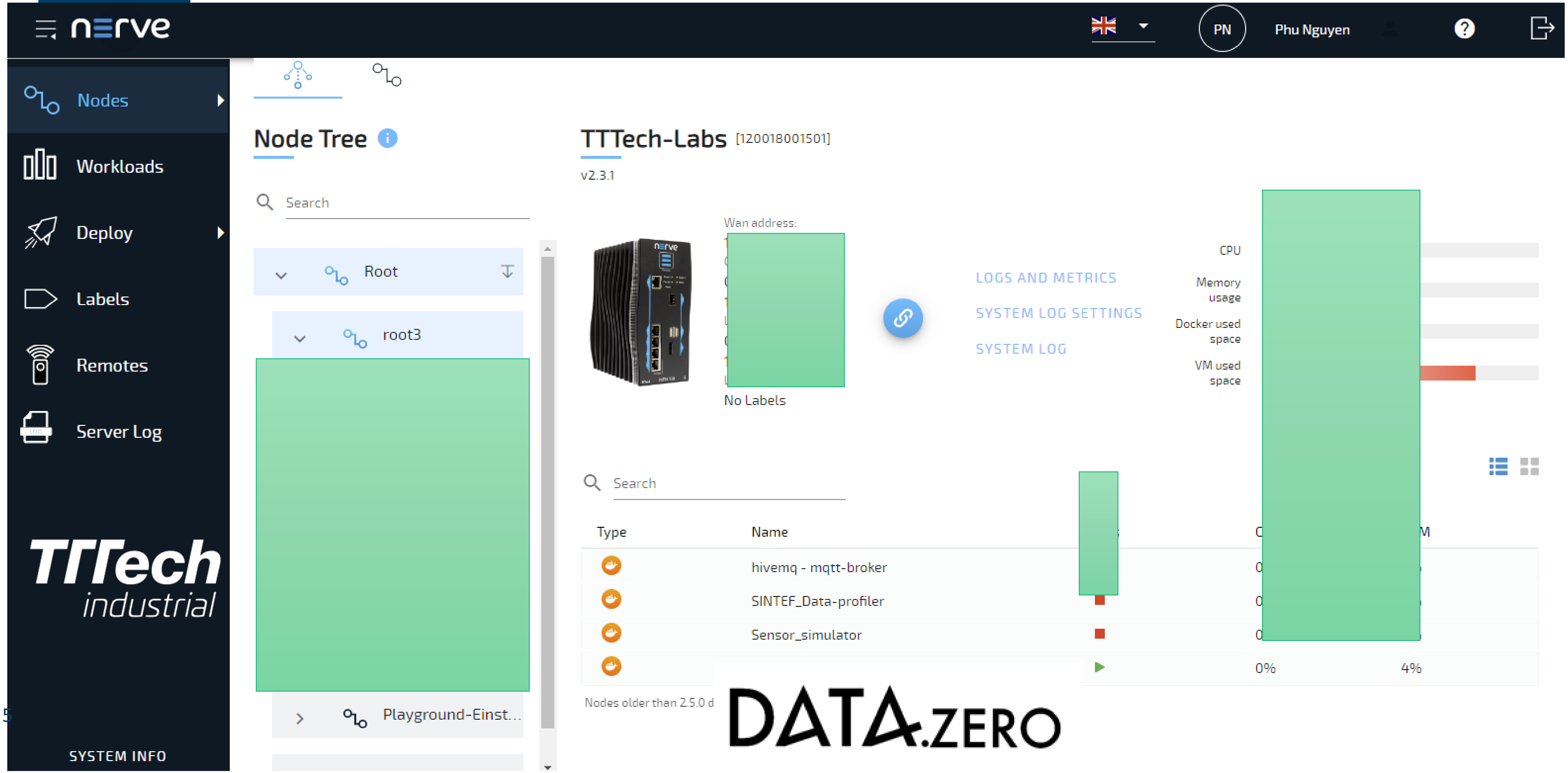


Simeon Tverdal, Arda Goknil, Phu Nguyen, Erik Johannes Husom, Sagar Sen and Jan Ruh, Francesca Flamigni. 2023. Edge-based Data Profiling and Repair as a Service for IoT. In Proceedings of IoT Conference (IoT'23). ACM, New York, NY, USA, Article 4, 9 pages.

SINTEF

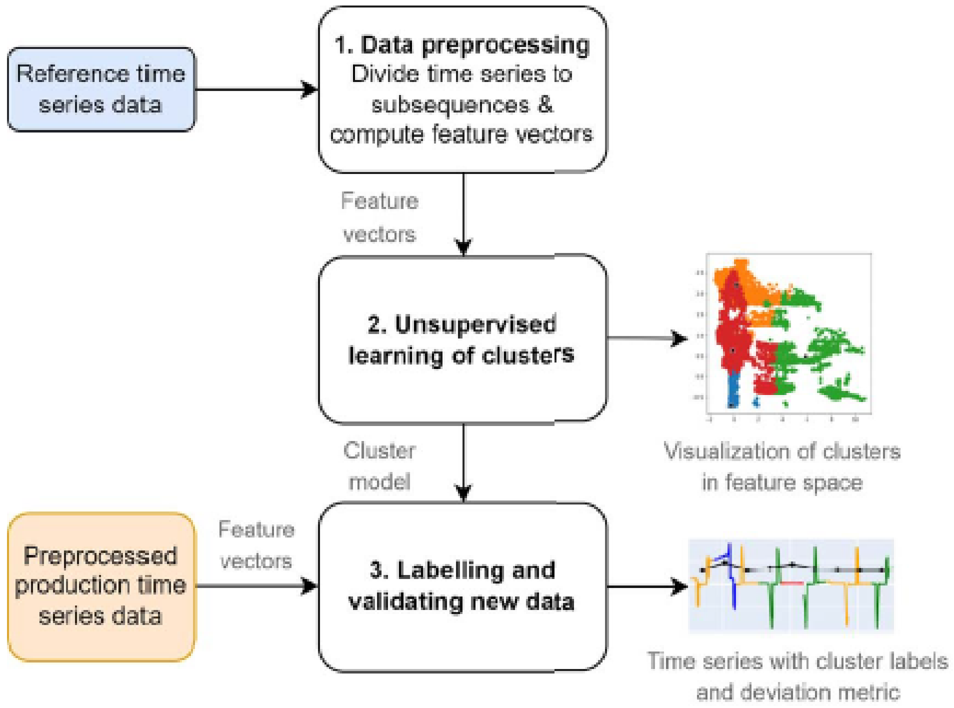
Data Quality on Edge: An example of how data (quality) can be managed at the Edge layer

Simeon Tverdal, Arda Goknil, Phu Nguyen, Erik Johannes Husom, Sagar Sen and Jan Ruh, Francesca Flamigni. 2023. Edge-based Data Profiling and Repair as a Service for IoT. In *Proceedings of IoT Conference (IoT'23)*. ACM, New York, NY, USA, Article 4, 9 pages. https://doi.org/xx.xxx/xxx_x



2) UDAVA: An Unsupervised Learning Pipeline for Sensor Data Validation in Manufacturing

- High-volume and velocity multivariate sensor data acquired during manufacturing introduce a challenging task for human operators to find diverse patterns of interest and track their deviations (e.g., process shifts and drifts) over multiple production cycles.
- UDAVA, an unsupervised machine learning pipeline, automatically discovers process behavior patterns in sensor data for a reference production cycle.



Erik Johannes Husom, Simeon Tverdal, Arda Goknil, and Sagar Sen. 2022. UDAVA: An Unsupervised Learning Pipeline for Sensor Data Validation in Manufacturing. In *1st Conference on AI Engineering - Software Engineering for AI (CAIN'22), May 16–24, 2022, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3522664.3528603

SINTEF

3D-DaVa :
3D Point Cloud Data Validation

Use cases:

Benteler & Dentsply Sirona

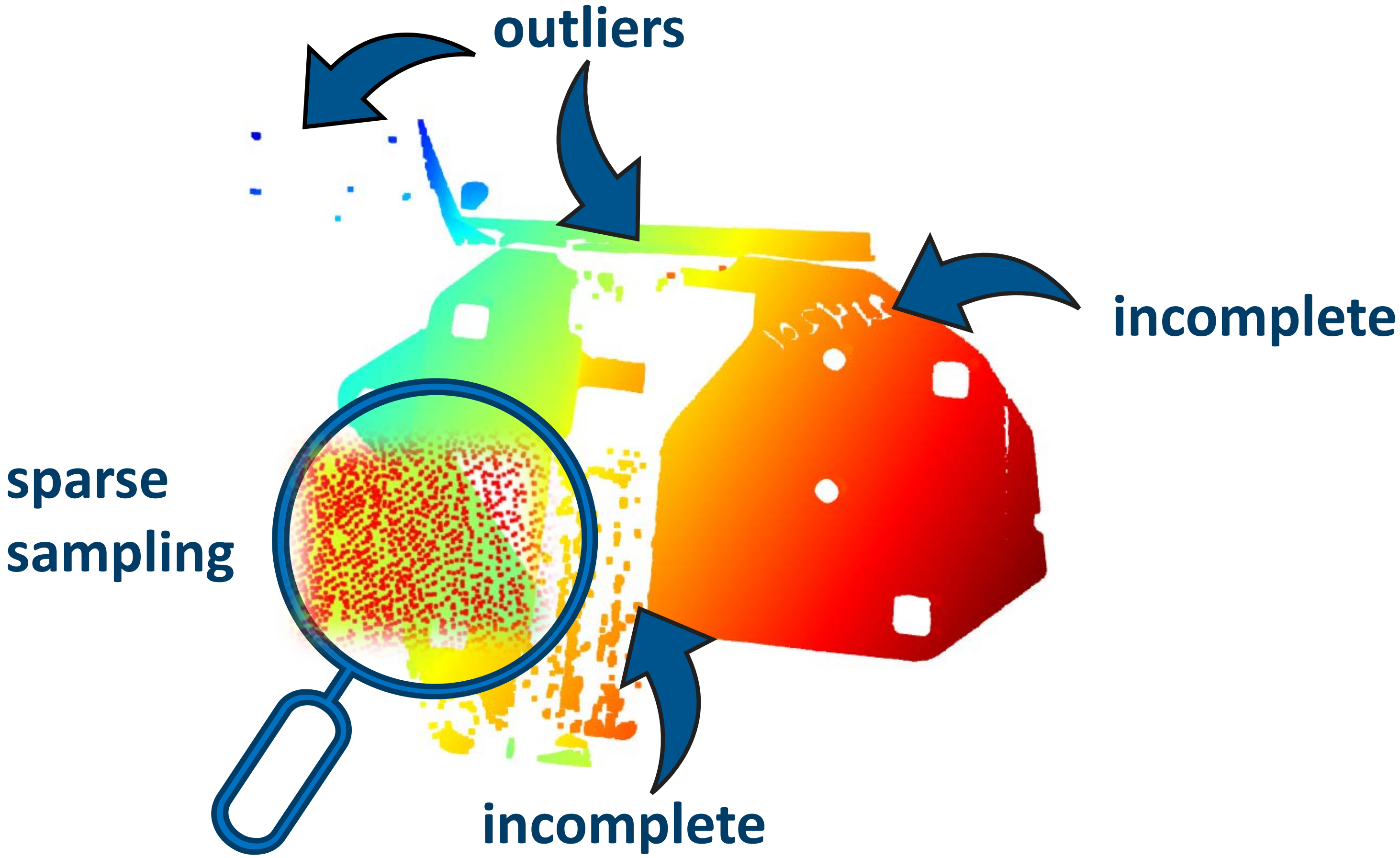
DATA.ZERO

Point clouds

- **Point clouds** are 3D representations of physical objects
- Typically obtained through non-contact **scanning technologies**
- Source of information within **Cyber-Physical Systems** (CPS)
- **Applications** in manufacturing: quality control, defect detection, progress monitoring etc.
- **High quality data leads to better results !**

Problem Statement

Point clouds can contain noise, outliers and/or be incomplete.



Data: Benteler

How can we increase reliability?

1. Obtain point cloud quality indicators

Goal: trigger corrective action (e.g. rescan, calibrate equipment).

2. Filter point cloud

- Goal: use filtered inputs to improve performance of models that use point cloud data.

3D-DaVa: Architecture

Input:

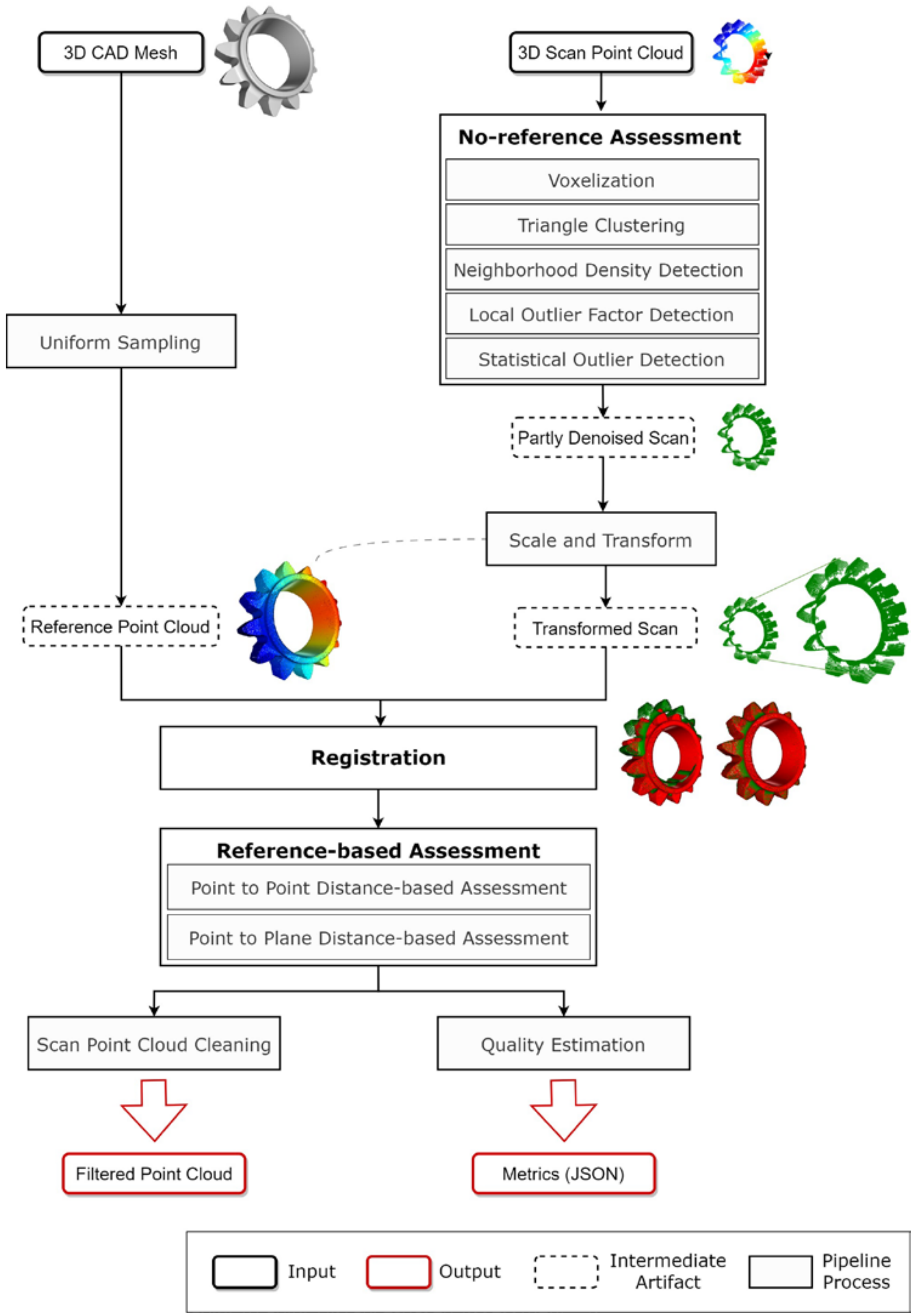
- Scanned point cloud
- Reference (i.e CAD)
- User-specified weights and parameters (optional)

Pipeline contains **two assessment blocks**:

- No-reference (no CAD) – 5 steps
- Reference-based (scan aligned with CAD) – 2 steps

Output:

- Quality metrics (accuracy/validity/completeness)
- Filtered scanned point cloud

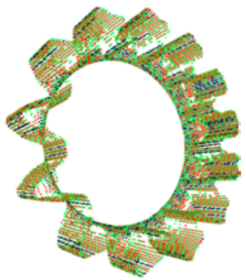


Data: Dentsply Sirona

Flow Overview

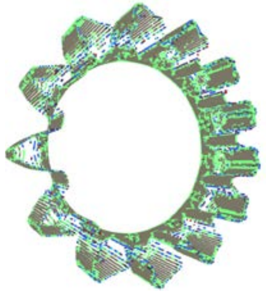
I. No-reference Assessment:

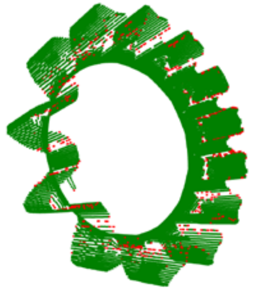

Scan


1.Voxelization

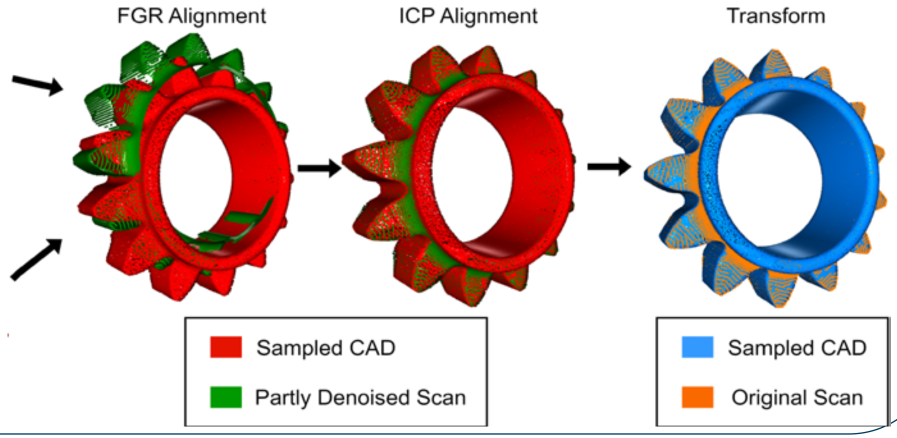

2.Triangle Clustering


3.Neighborhood Density


4. Local Outlier Factor Detection

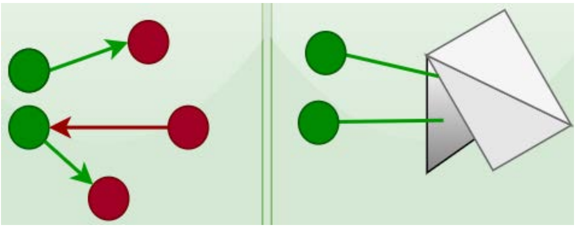

5.Statistical Outlier Detection

II. Registration (partly represented):



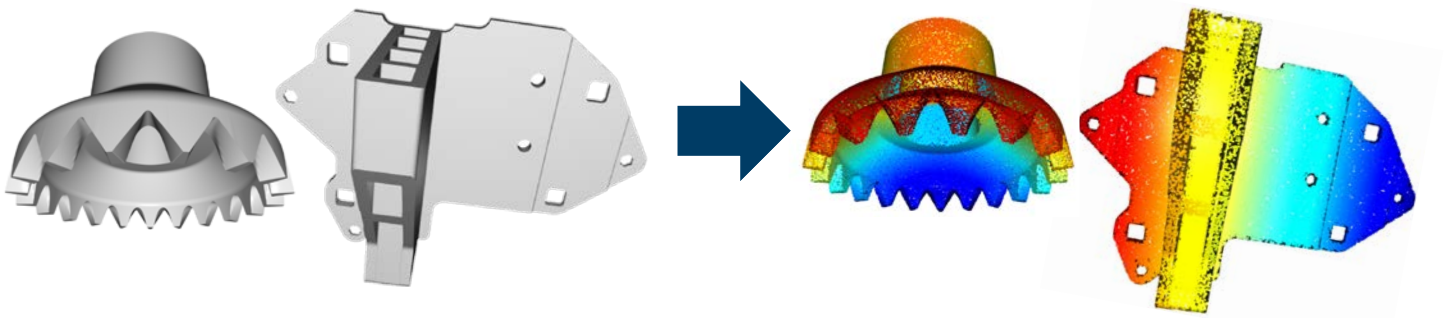
III. Reference-based Assessment:

Distance-based Assessments:
1. Point-to-point
2. Point-to-plane

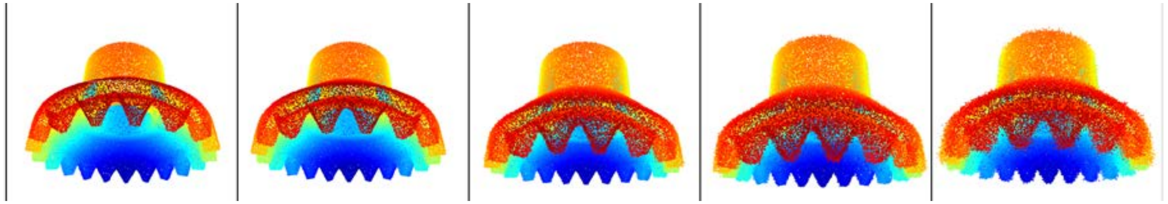


Output

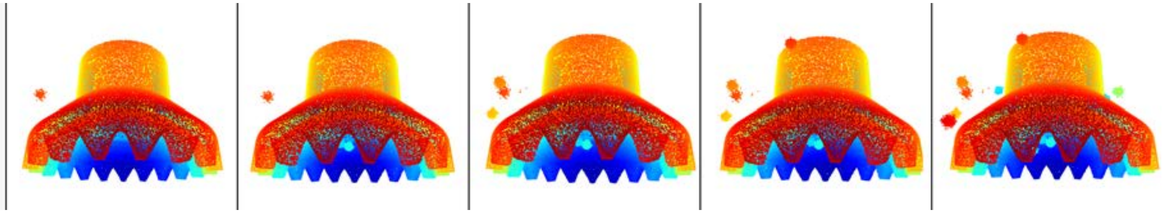
Validation



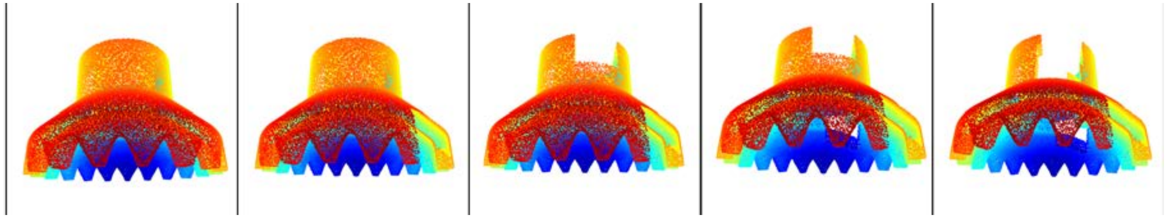
- Facilitated by **simulating distortions** on sampled CADs, for each use case.
- Vary over 5 levels of distortion intensity.
- We simulate:



a) Gaussian noise: threat to accuracy



b) Outlier clusters: threat to validity



c) Local Missing: threat to completeness

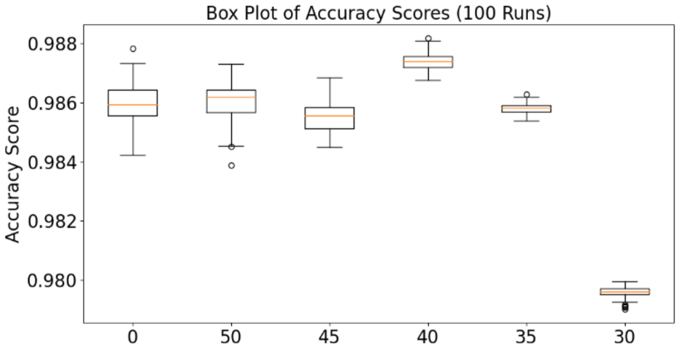
We plot quality for each distortion, and check if metrics drop with an increase of distortion intensity.

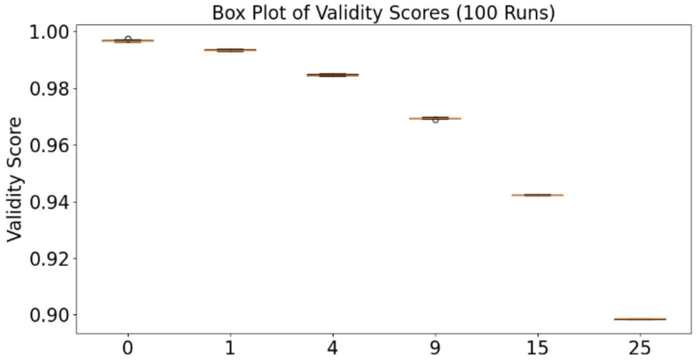
Results

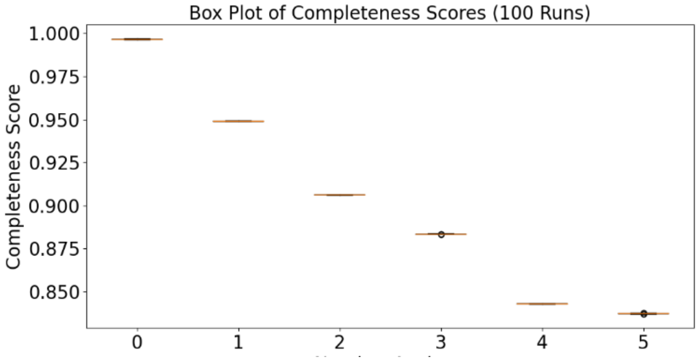
Expected:
Higher distortion intensity
→ loss in quality

Results:
As expected for outliers and missing values.

Dentsply Sirona







Benteler

Addressing limitations

- **Noisy data is more difficult to quantify.**

Best at capturing low intensity noise. Keep blue noise in mind. Allow user-specified weights.

- **Good alignment fitness is necessary.**

Align, remove outliers based on current alignment, then re-align.

- **A distortion might threaten several quality dimensions.**

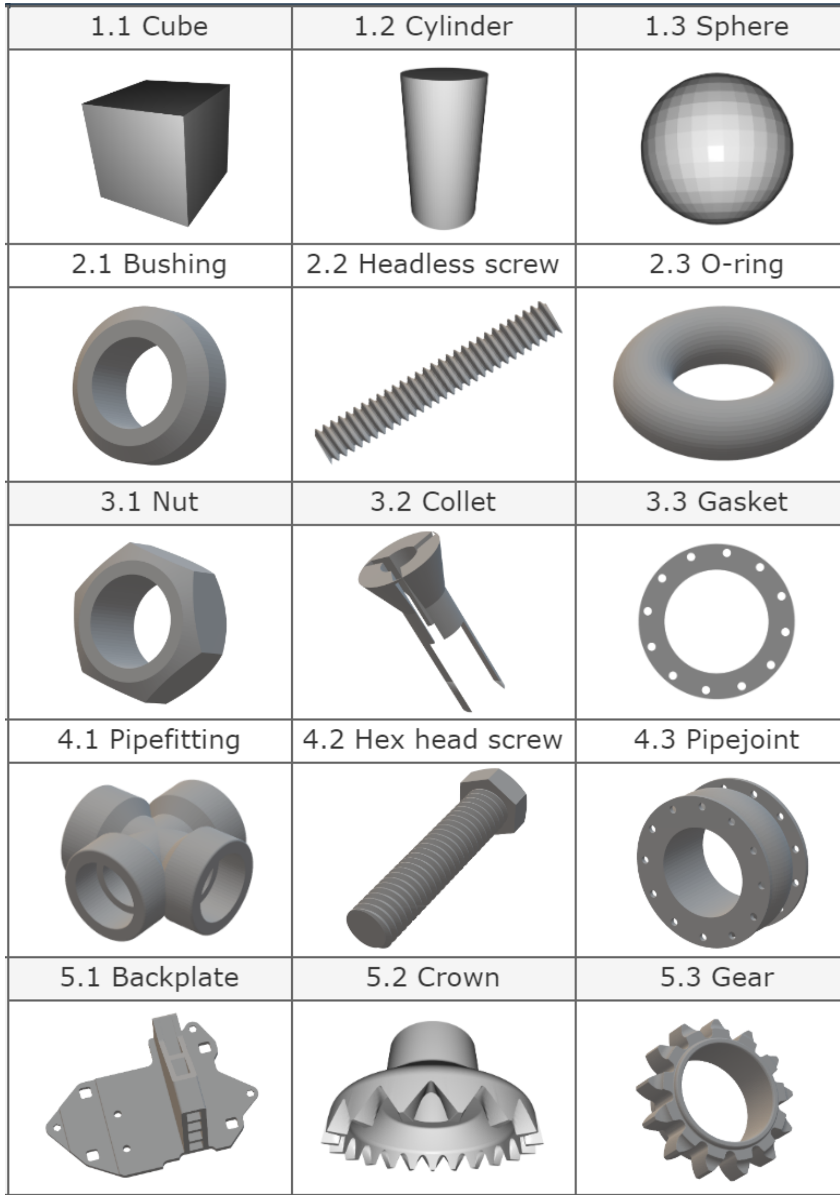
A sensitivity investigation showed good results and explains the noise values.

Dissemination

- Paper submitted to ACM: Journal of Data and Information Quality in January.
- Titled:

  3D-DaVa: Enhancing 3D Point Cloud Data Reliability for Industrial Applications

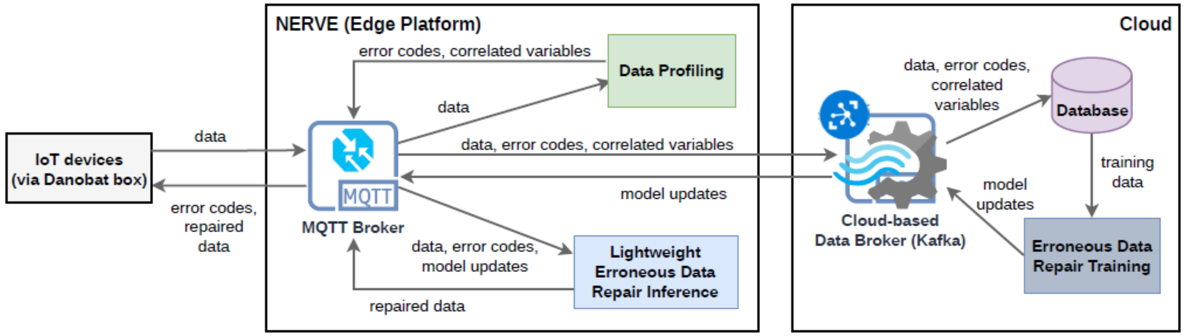
- Same approach, several point cloud examples.
- Outlines future work.



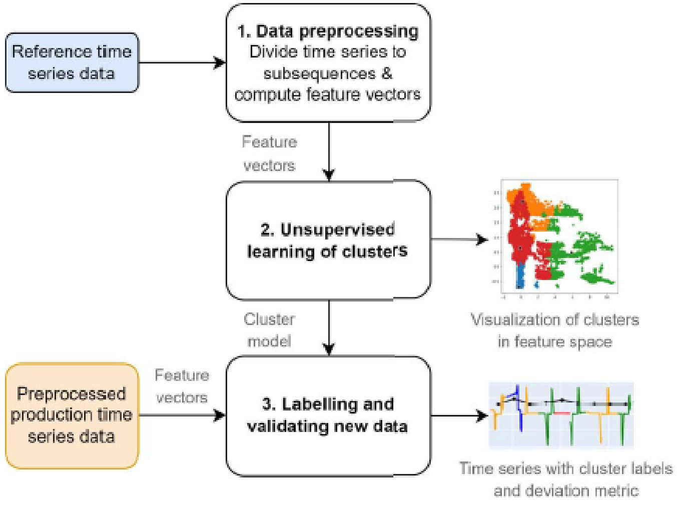
Recap

- We discussed the importance of data quality.
- We briefly give a high-level overview of
  - Erroneous data repair on Edge approach
  - UDAVA approach for discovering process behaviour patterns in sensor data for a reference production cycle.
- We presented the main points of our 3D Point Cloud Data Validation work

1) Data Quality on Edge



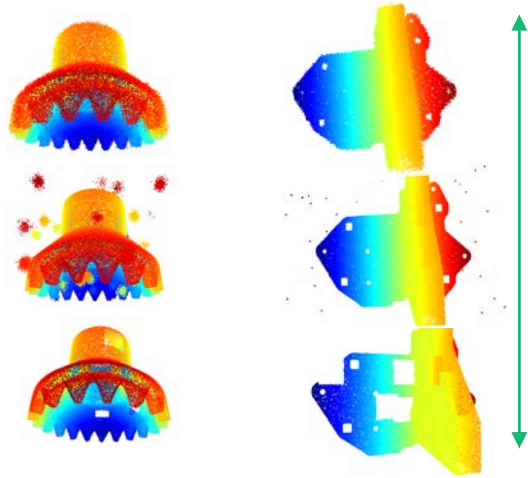
2) UDAVA



3) 3D-DaVa :

3D Point Cloud

Data Validation



References

- Isaja, Mauro, Phu Nguyen, Arda Goknil, Sagar Sen, Erik Johannes Husom, Simeon Tverdal, Abhilash Anand et al. "A blockchain-based framework for trusted quality data sharing towards zero-defect manufacturing." Computers in Industry 146 (2023): 103853.
- Goknil, Arda, Phu Nguyen, Sagar Sen, Dimitra Politaki, Harris Niavis, Karl John Pedersen, Abdillah Suyuthi, Abhilash Anand, and Amina Ziegenbein. "A Systematic Review of Data Quality in CPS and IoT for Industry 4.0." ACM Computing Surveys 55, no. 14s (2023): 1-38.
- Sen, Sagar, Erik Johannes Husom, Arda Goknil, Dimitra Politaki, Simeon Tverdal, Phu Nguyen, and Nicolas Jourdan. "Virtual sensors for erroneous data repair in manufacturing a machine learning pipeline." Computers in Industry 149 (2023): 103917.
- Sen, Sagar, Erik Johannes Husom, Arda Goknil, Simeon Tverdal, Phu Nguyen, and Iker Mancisidor. "Taming data quality in AI-enabled industrial internet of things." IEEE Software 39, no. 6 (2022): 35-42.
- Cassoli, Beatriz Bretones, Nicolas Jourdan, Phu H. Nguyen, Sagar Sen, Enrique Garcia-Ceja, and Joachim Metternich. "Frameworks for data-driven quality management in cyber-physical systems for manufacturing: A systematic review." Procedia CIRP 112 (2022): 567-572.
- Erik Johannes Husom, Simeon Tverdal, Arda Goknil, and Sagar Sen. 2022. UDAVA: An Unsupervised Learning Pipeline for Sensor Data Validation in Manufacturing. In 1st Conference on AI Engineering - Software Engineering for AI (CAIN’22), May 16–24, 2022, Pittsburgh, PA, USA. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3522664.3528603

SINTEF

https://sea4dq.github.io/

Important Dates

- **Paper Submission Deadline:** ~~March 15, 2024~~ March 29, 2024
- **Notification of Acceptance:** April 26, 2024
- **Camera-Ready Submission:** May 17, 2024
- **Workshop Day:** July 15, 2024

Teknologi for et bedre samfunn