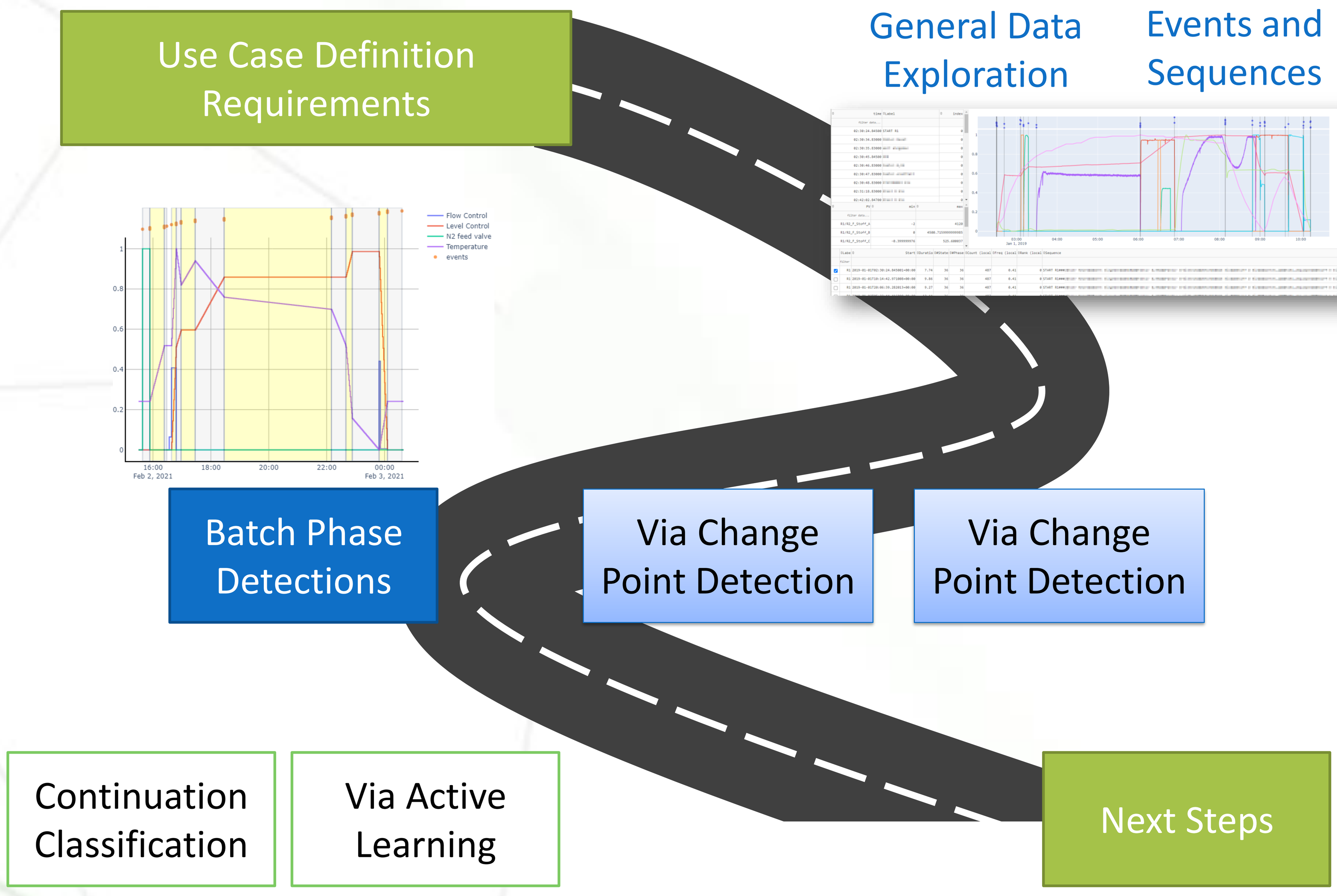
Towards Automatic Batch Phase Detection
First Results using Machine Learning

F. Bähner[1]*, M. Gärtler[2]*, B. Klöpper[2]*, S. Maczey[2]*,
S. Merkelbach[3]*, P. Sander[4]*, C. Song[2]*, C. Sonntag[4]*, R. Tan[2]*

Motivation

- Batch phase statistics make batch process data and analyses accessible
- They have a wide range of use cases
  - Data analysis, process modelling (scheduling, planning, retrofit optimization), but also realtime-hard process monitoring
- Batch phase statistics (start, end, attributes) are not always archived automatically, and not always at high quality, especially in
  - Old plants, systems with islands of automation, manually operated plants
- Timeseries values of measurements are more often archived
- The inherent reproducibility of batch recipes makes automatic posterior batch phase identification based on timeseries values a promising task for ML
  - ML can help to distinguish recurrent & irregular behaviors as well as noise

Roadmap



Detection of Start and End Times of Batch Phases [1]

- How easy or hard is it to identify batch phases?
  - Given labelled information to train on
  - Without labels and looking for candidates
- Simple classification works extremely good on ideal data from simulation
  - Real-World Data need further work
- Performance of Change Point Detection depends on the characteristics of timeseries
  - Overall relevant / interesting parts can be identified

Results Classification

- 2 approaches for the classification of batch phases were tested:
  - Random forest classification on sensor measurements
  - Random forest classification on statistical features [2]
- Experiments were made on:
  - Simulated data with 4 variables and 13 different phases
  - Real data with 8 variables and 43 different phases

Accuracy of prediction	Simulated data	Real plant data
Classification on sensor measurements	99.94%	69.92%
Classification on statistical features	99.94%	58.49%

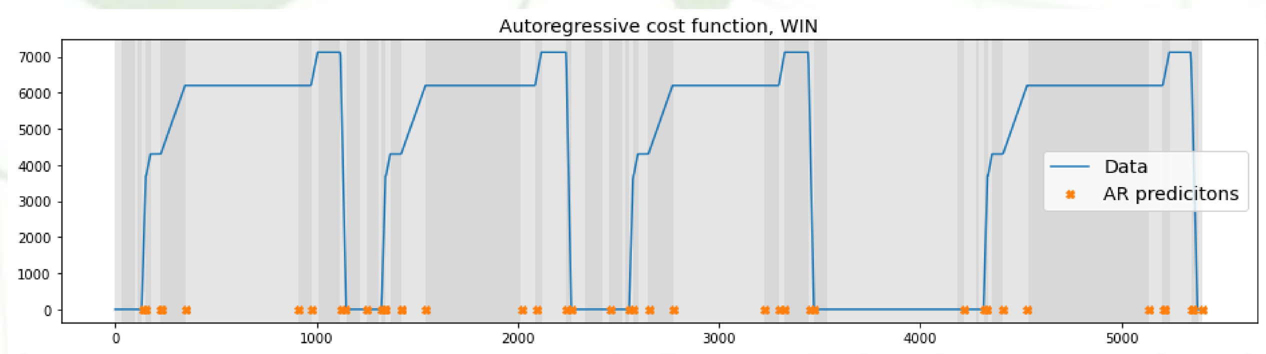
Results - Change Point Detection

- Unsupervised CPD as an optimization problem:

$$\min_{\tau} V(\tau) + pen(\tau) \;=\; \min_{\tau} \sum_{t \in \tau} c(\,y_{t_j:t_{j+1}}\,) + \beta|\tau|$$

τ: change points; y: original signal;
V: criterion function for all segments ; pen: penalty; c: cost function

- Results for automatic batch phase detection



- Findings: cost function influences the performance of CPD and should be chosen according to the characteristic of the data.

Next Steps

- Batch phase classification (5.2)
  - Multivariate optimization to create better classification results
  - Removal of erroneous batches
  - Refine feature creation to achieve better results
  - Include sequences in classification
- Active Learning (5.4)
  - Evaluation of Idea - Provide few exact labels (via Experts)
  - Comparison with classification / Supervised Learning
  - Improve the visualization of the prediction quality during active learning interaction

[1] S. Merkelbach, R. Tan, F. Bähner, M. Gärtler, J. Gatter, R. Gedda, L. Urbas: Automatic Detection of start and end times of batch phases for the process industry, AUTOMATION 2021, accepted (2021)

The authors would like to acknowledge the contribution of master students Josefine Gatter and Rebecca Gedda to the work.

[2] B. Esmael, R. K. Fruhwirth, and G. Thonhauser, "A Statistical Features Based Approach for Operations Recognition in Drilling Time Series," vol. 5, p. 4545, 2013.

[1] Bayer AG
[2] ABB
[3] TU Dresden
[4] INOSIM

* Each author contributed equally to the overall progress

www.keen-plattform.de

Contact

Franz D. Brähner (Use Case Owner)

Bayer AG
www.bayer.com
franz.baehner@bayer.com

Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag