3rd European Forum on New Technologies: Chemical Engineering in the Plant of the Future

UTILIZING BIG DATA ANALYTICS IN THE CHEMICAL INDUSTRY: EXPERIENCES AND CHALLENGES AT THE DOW INC. COMPANY

IVAN CASTILLO, LEO CHIANG, RICARDO RENDALL AND MARK JOSWIAK

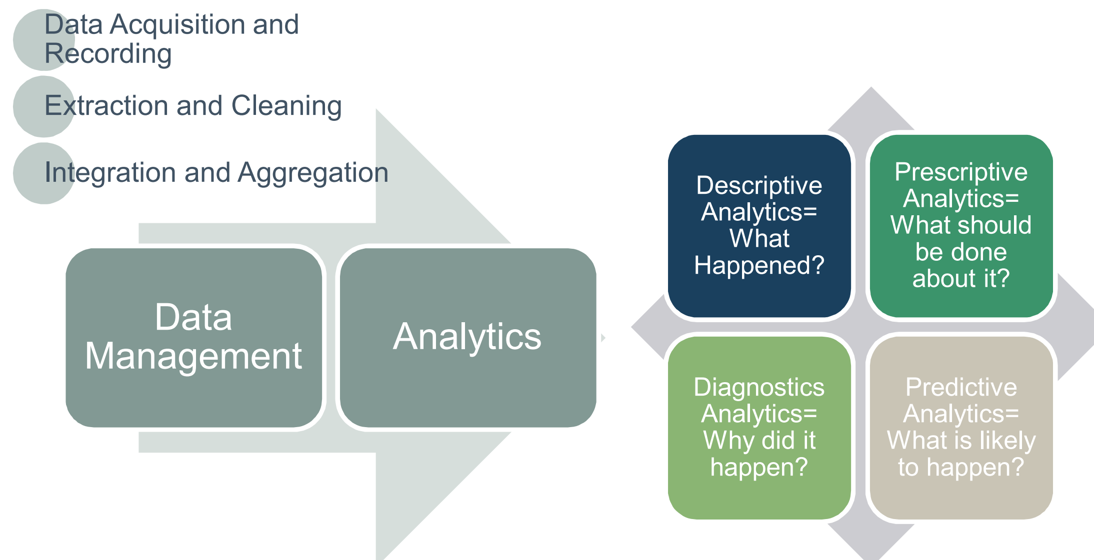
CHEMOMETRICS AND AI GROUP, DATA SERVICES

DOW INC.

9/11/2020

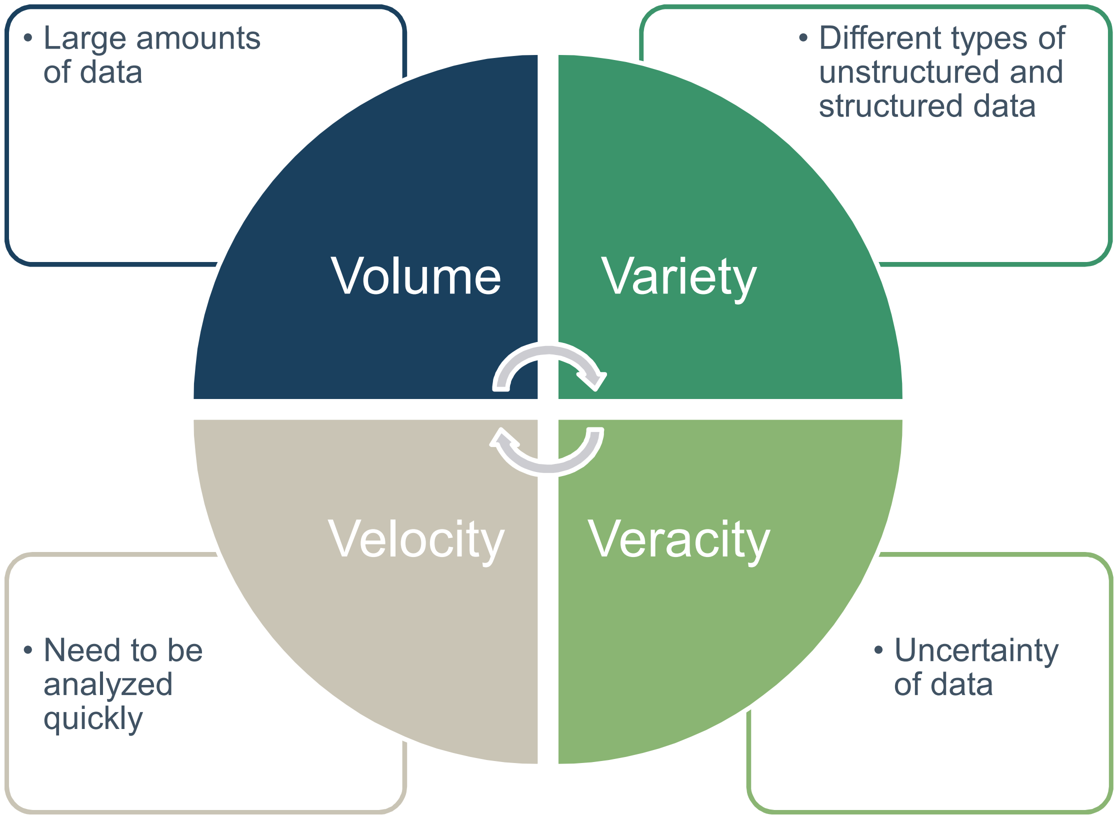
THE FINAL GOAL WHEN APPLYING ANALYTICS: TO DRIVE DECISION MAKING

Large amounts of data



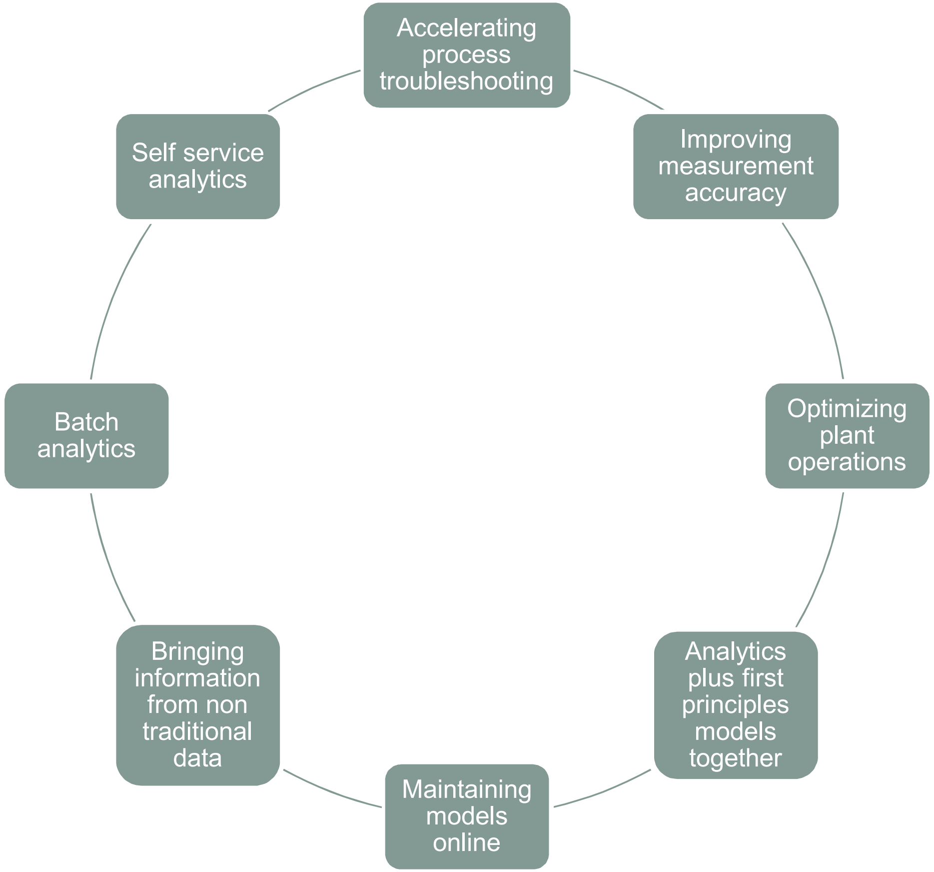
Colegrove, L. F., Seasholtz, M. B., & Khare, C. (2016). Big data: Getting started on the journey. *Chemical Engineering Progress*.

Applications of Data Analytics at Dow Inc.

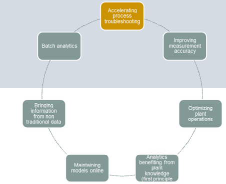


Chiang, L., Lu, B., & Castillo, I. (2017). Big data analytics in chemical engineering. *Annual Review of Chemical and Biomolecular Engineering*

Big Data Analytics for the Plant of the Future



Analytics applied to chemical manufacturing process data



ACCELERATING PROCESS TROUBLESHOOTING

Visualizing High-Dimensional Data

How can the cup be represented in this image?
As points in a 3 dimensional space (three color channels)
or 2 dimensions in black and white (the shadow)?

Advantages of Dimensionality Reduction

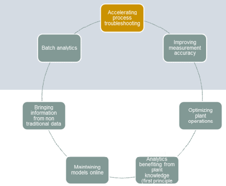
Increases interpretability (clusters, outliers)

Decreases the risk of overfitting and noise

Eliminates irrelevant and redundant features

Decreases computational time

ACCELERATING PROCESS TROUBLESHOOTING

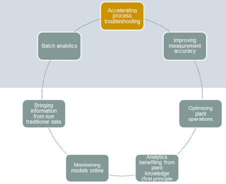


Approaches to Visualize High Dimensional Data

Unsupervised	Supervised	Reinforcement Learning
PCA **tSNE** **UMAP** Autoencoders	PLS PLS-DA UMAP	

Further details on visualization:

- M. Joswiak, Y. Peng, I. Castillo, and L. Chiang, Visualizing Chemical Processes Utilizing Dimensionality Reduction Methods: Survey and Applications, *Control Engineering Practice*, 2019.

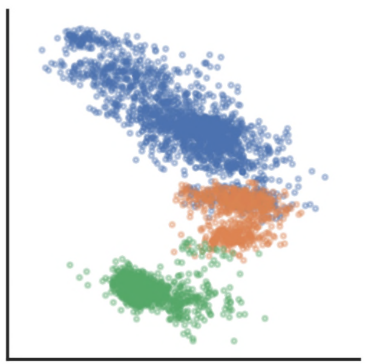


Accelerating process troubleshooting

Case study: What occurs prior to an unplanned event?

UMAP and t-SNE immediately provide new insights

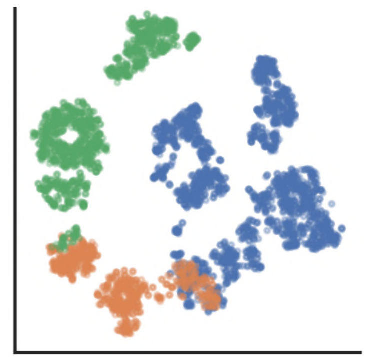
PCA

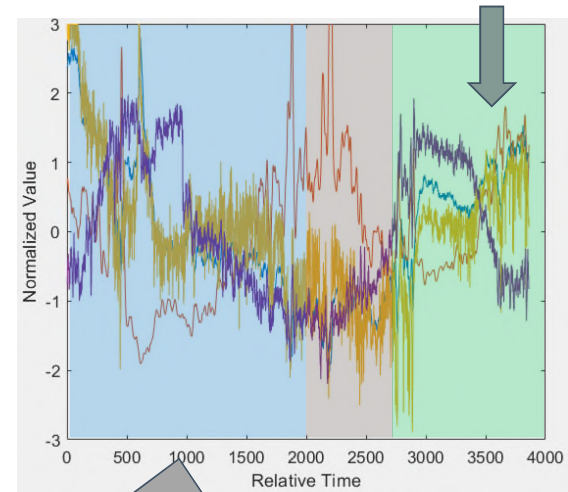


UMAP

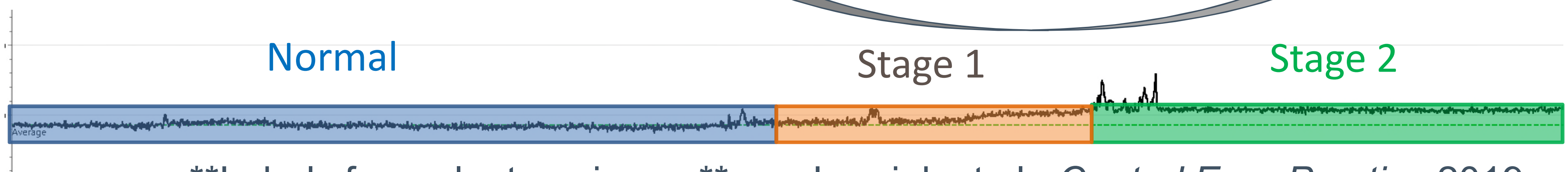
2 clusters!

t-SNE





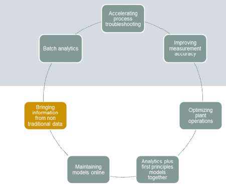
What's different?



! plant down

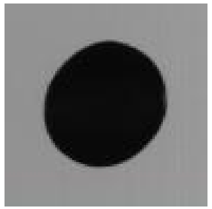
Labels from plant engineers

Joswiak et al., *Control Engr Practice* 2019

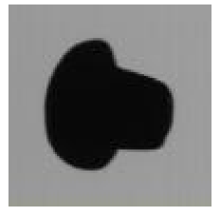


BRINGING INFORMATION FROM NON TRADITIONAL DATA

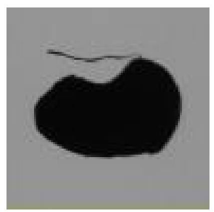
Good Pellet



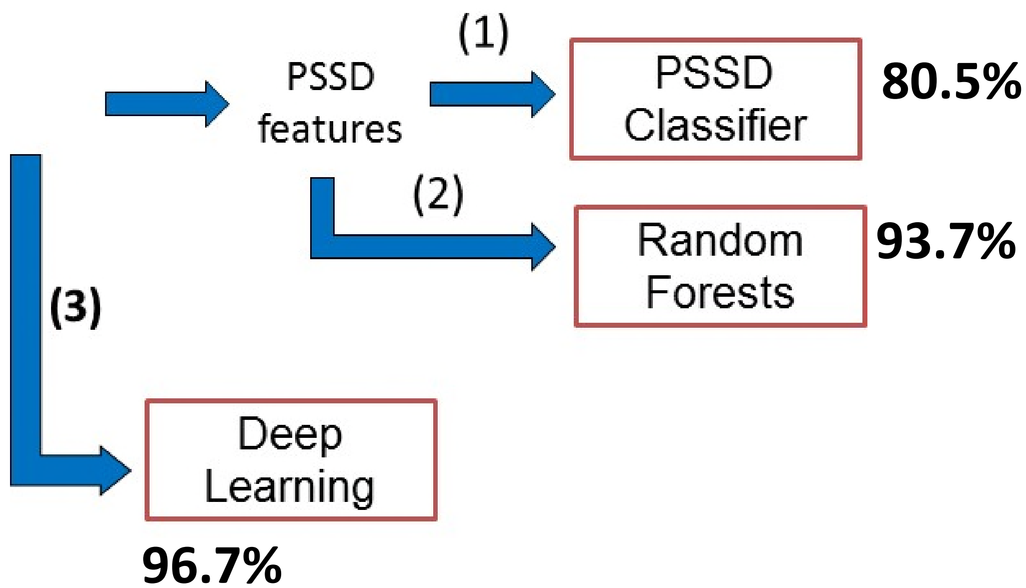
Bad Pellet



Bad Pellet with Tails

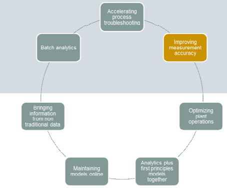


Process experts manually labelled 6000+ images from Particle Shape and Size Distribution (PSSD)



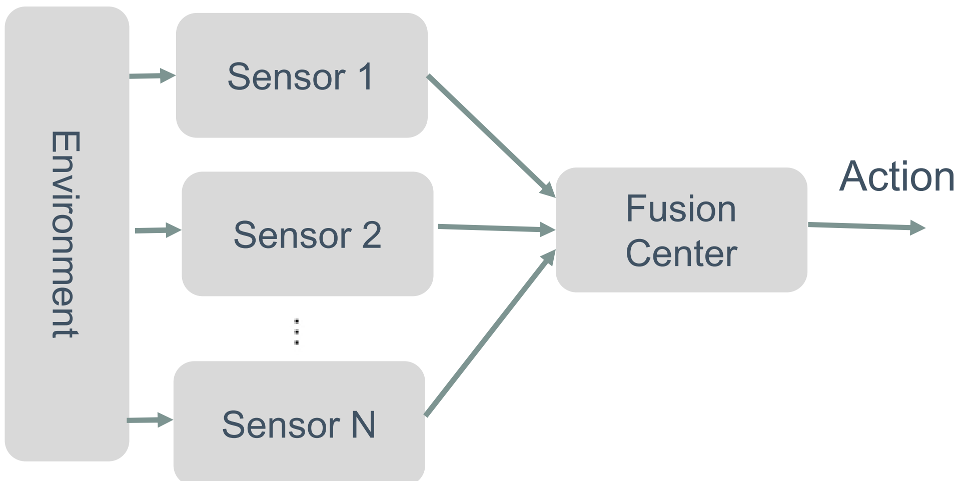
R. Rendall, I. Castillo, B. Lu, M. Broadway, B. Colegrove, L. Chiang, and M. Reis (2018 AIChE spring meeting, Chemo & Intel Lab Systems, 2018)

IMPROVING MEASUREMENT ACCURACY

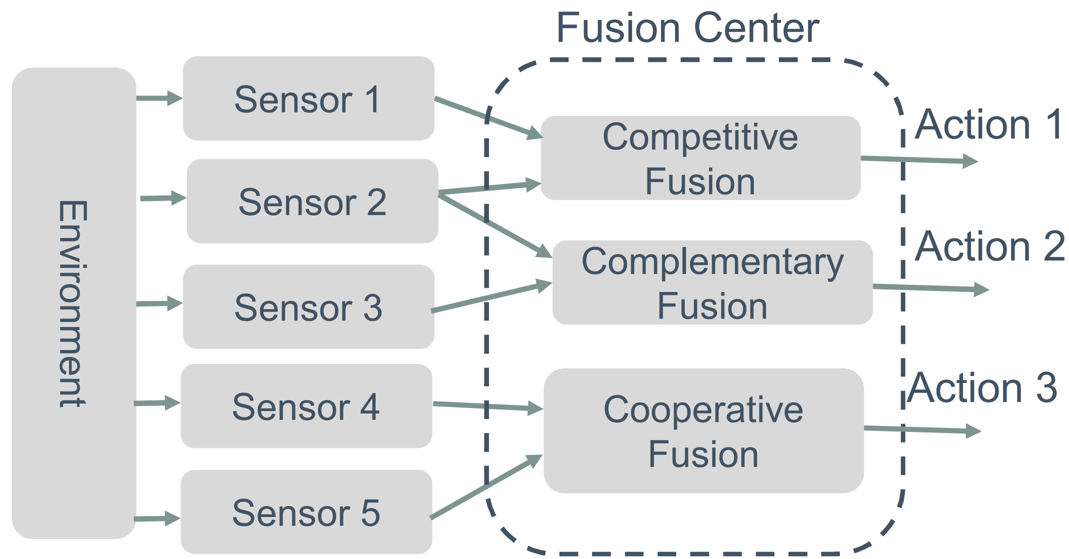


- *Sensor fusion* is the process of merging data from multiple sensors to reduce uncertainty in model predictions

Example of a centralized architecture

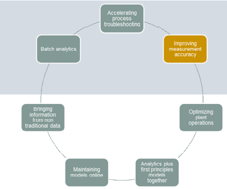


Sensor fusion networks



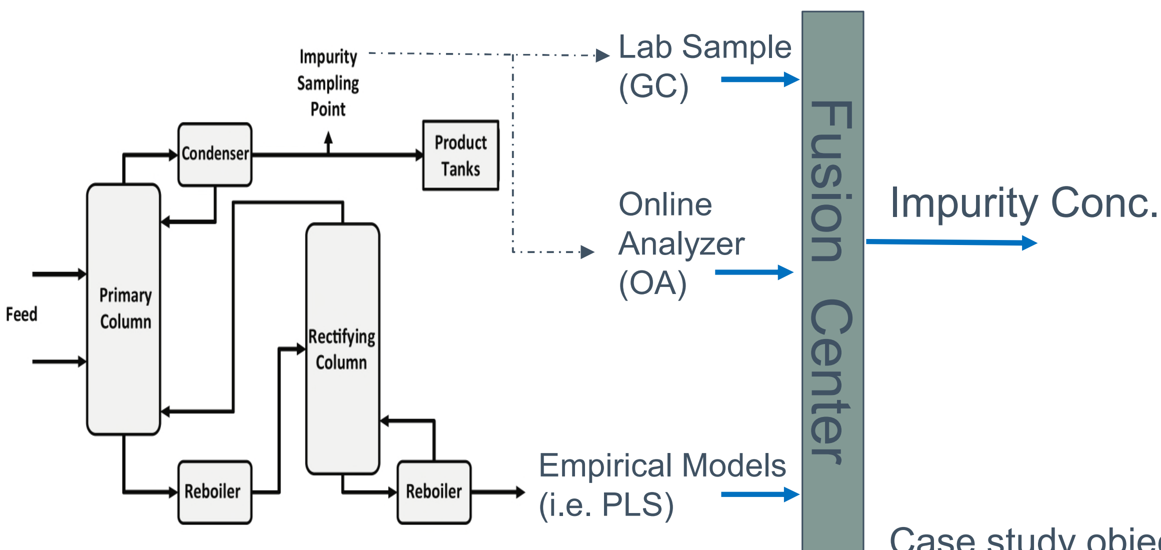
Graphic based on research work of Spyros G. Tzafestas, in Introduction to Mobile Robot Control, 2014

Graphic based on research work of Diego Galar, Uday Kumar, "*Sensors and Data Acquisition*" , in eMaintenance, 2017



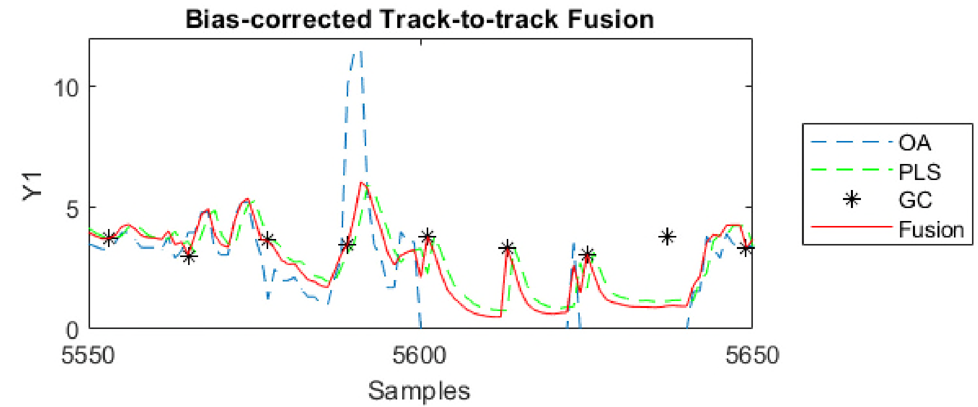
IMPROVING MEASUREMENT ACCURACY

Case Study: Monitoring Impurities of EO



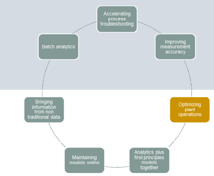
Fusion Results

Online Analyzer False Alarm



Case study objectives:

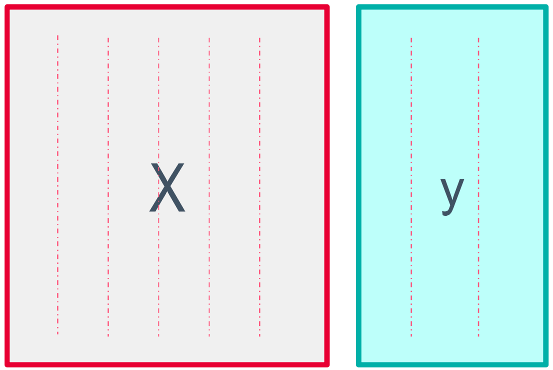
1. Reduce false alarms (type-I error)
2. Reduce misidentification of off-spec product (type-II error)
3. Increase monitoring precision
4. Reduce sampling frequency for lab measurement



Optimizing Plant Operations

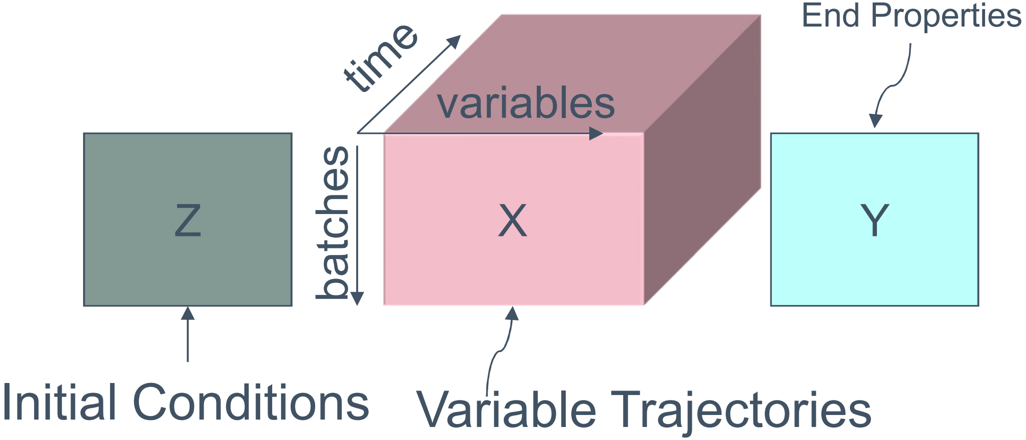
Our Classic Data Problems

Continuous Process



- Steady-state continuous process
- Data size is small
- Usually for prediction of quality or latent variables

Batch Process



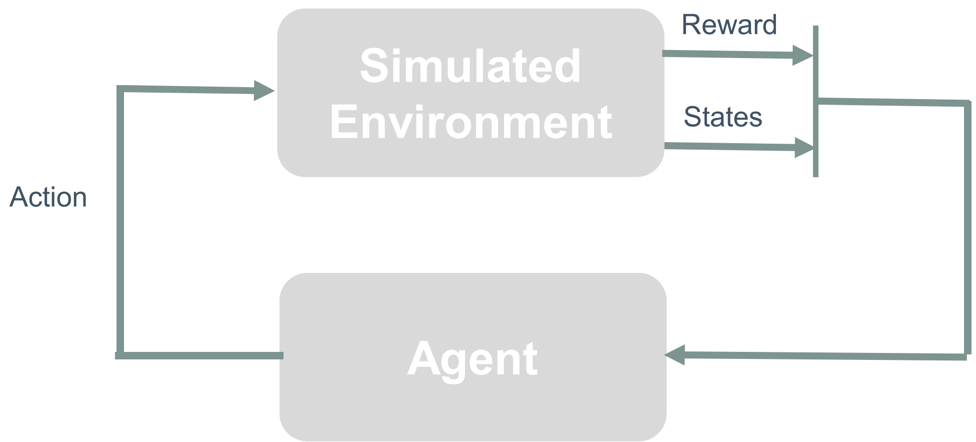
- Usually transient and dynamic
- Data variety / volume challenge, need to match context, multi-dimensional data
- High frequency, large volume datasets

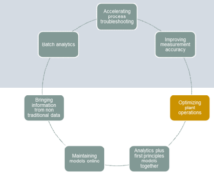
Optimizing Plant Operations

Reinforcement Learning (RL) in Batch Processes

<u>Objective:</u>

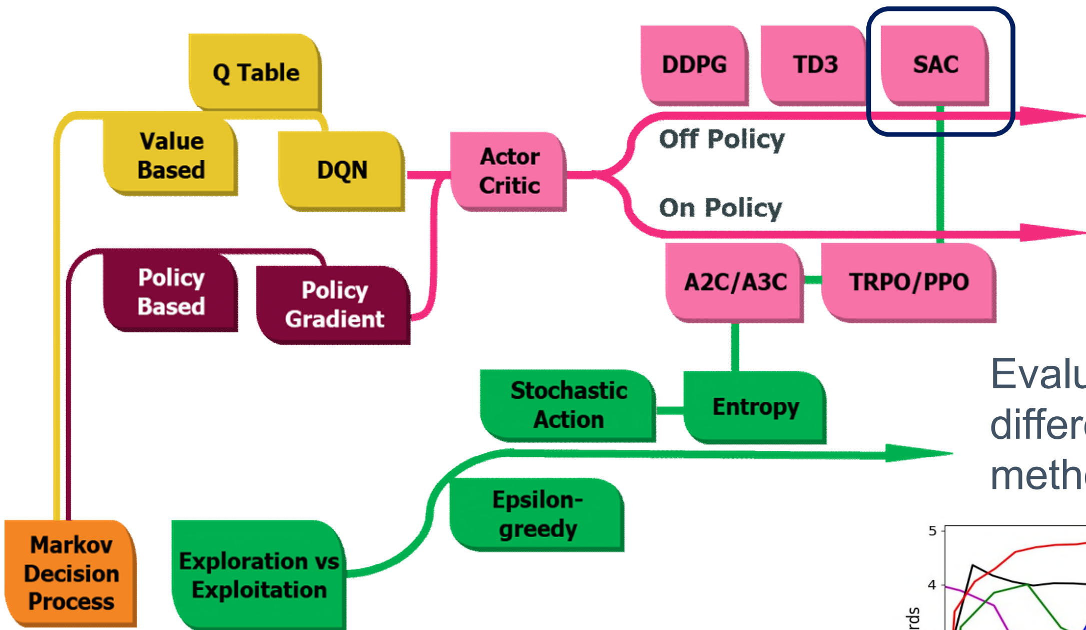
- Determine Optimal trajectory for maximizing batch process production
    - Optimization with states constraints





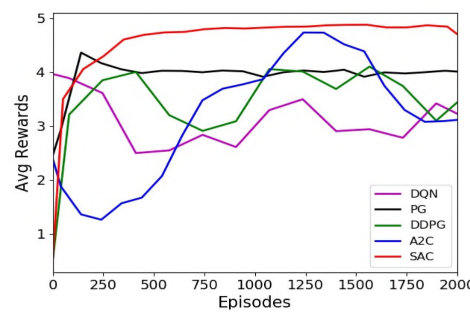
OPTIMIZING PLANT OPERATIONS

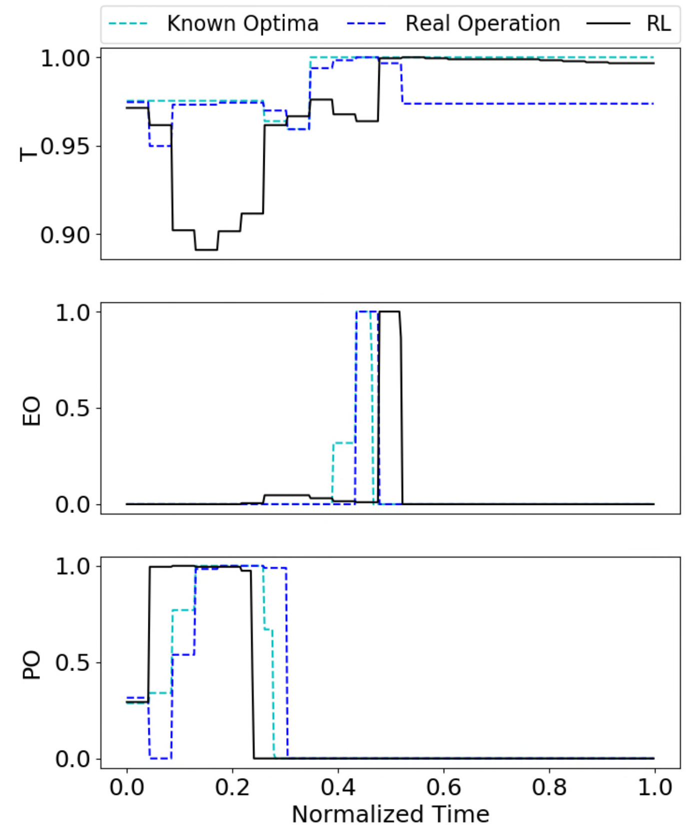
- RL identified optimal trajectories in order to improve batch performance



RL Approaches

Evaluation curve for different DRL methods

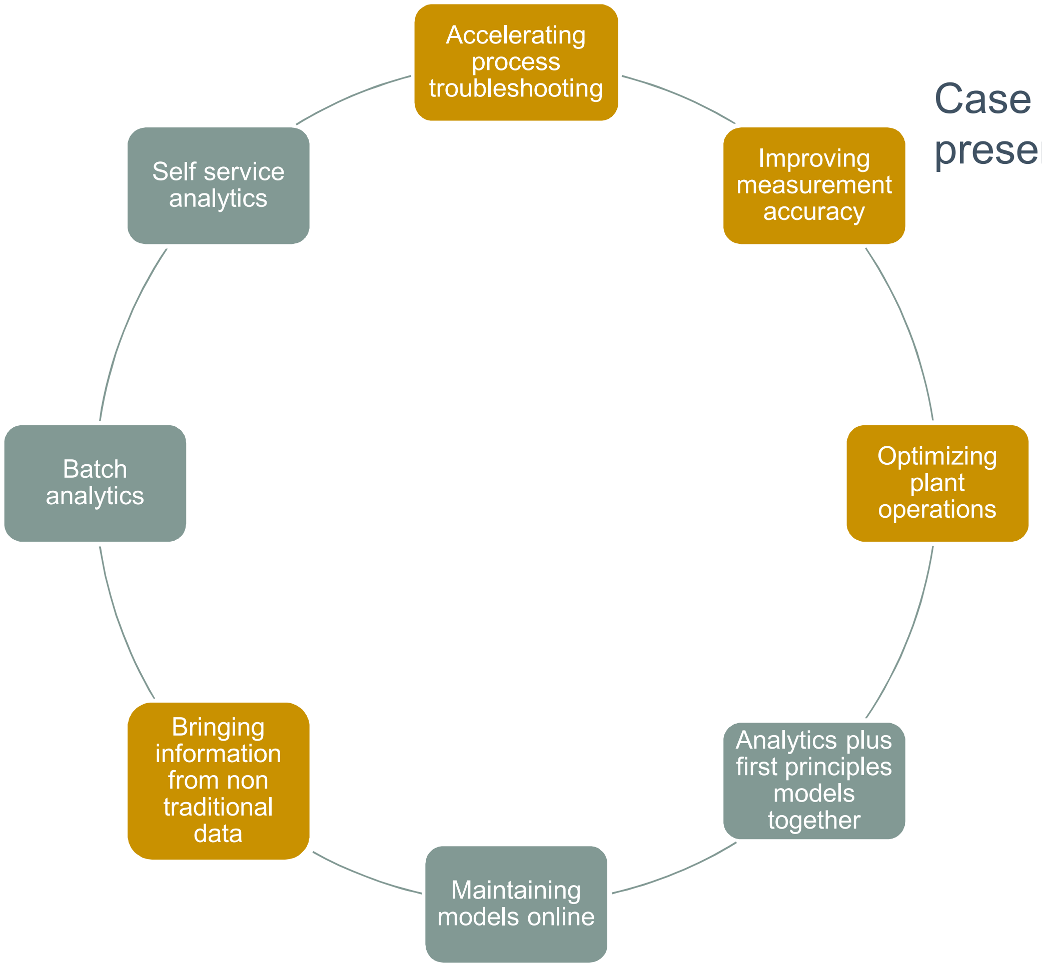




Challenges

- Identify the right fit for new analytics approaches, such as deep learning (DL) and reinforcement learning (RL), for the chemical industry, while maintaining safety and reliability.
- New analytics technologies require multiple successful case studies to speed up adoption within the chemical industry.
- There is a hesitancy to adopt black box models to operate plants. Hybrid approaches and improving interpretability will be helpful for adoption within an industrial context.
- Model maintenance once models are deployed real-time requires further exploration.

Summary: Analytics for the Plant of the Future



Case studies covered in this presentation