DATAHOW

Accelerating bioprocess development with digital twins and machine learning

Dr. Moritz von Stosch

1 February 2023

Who is DataHow?

2017

Incorporation

- 4 Co-Founders
- 35yrs R&D experience
- 20yrs Industry experience
- 30yrs Machine Learning

45

Team Members

- 9 EU Countries
- 40% Eng.
- 30% ML
- 30% IT

3

Locations

- HQ in Zürich (CH)
- Subsidiary in Lisbon (PT)
- Subsidiary in Milano (IT)

3

Customer Regions

- 60% Europe
- 30% USA
- 10% Asia

Our Partners

Takeda, gsk, SANOFI, NOVARTIS, MERCK, CSL Behring, Rentschler Biopharma, Bristol-Myers Squibb, Boehringer Ingelheim, ucb, abbvie, novo nordisk®

Customers

> **Hundreds** of industrial process data sets

> **150** trained users on DataHowLab

Technology Partners

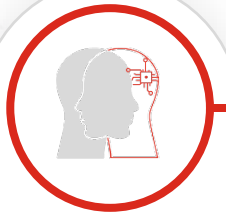
eppendorf, PALL, cytiva, MERCK MILLIPORE, Endress+Hauser, INFORS HT, securecell, PERCEPTIVE ENGINEERING, Genedata solutions in silico

Academic Partners

ETH zürich, POLITECNICO MILANO 1863, TU berlin, KIWI biolab THIS IS THE KIWI-BIOLAB, A BMBF KI FUTURE LAB!

Technologies we believe will change the way process are developed.

Connectivity / Twins



What?
Automatic acquisition of data, data contextualization, in-silico simulations

Where?
DataHowLab

Why?
- Scenario analysis
- Model-driven design of experiment
- Risk management
- Knowledge management

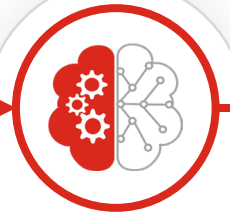
Self-Learning

What?
Direct learning and continuous model refinement from process data

Where?
DataHowLab + SpectraHow

Why?
- Automatically learns from new data
- No need of prior knowledge
- Adapts to new knowledge and improves optimal decision making

Hybrid Models

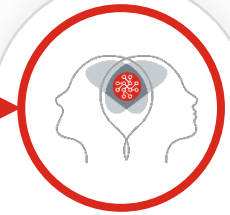


What?
Combine process engineering knowledge with machine-learning

Where?
DataHowLab backend

Why?
- Reduces data requirements
- Extrapolates to multiple products and/or process features
- Increases process understanding

Knowledge Transfer



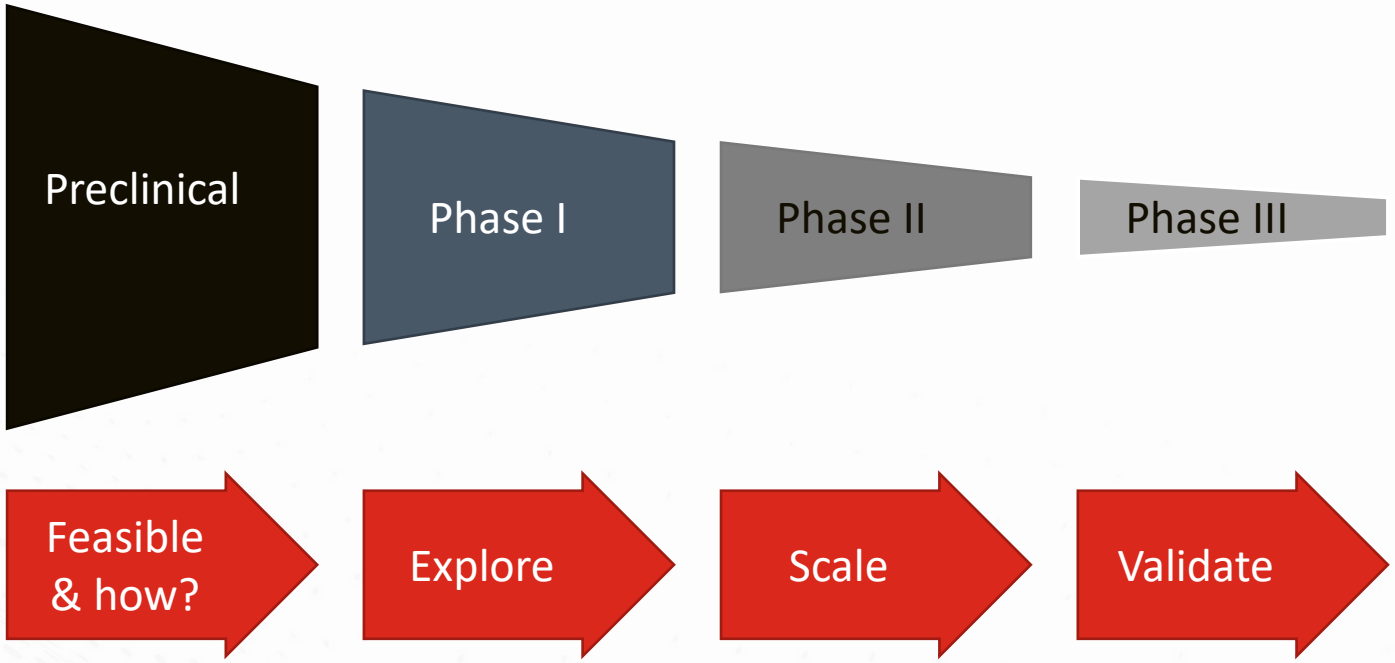
What?
Automatic transfer of knowledge between products & scales

Where?
DataHowLab + SpectraHow

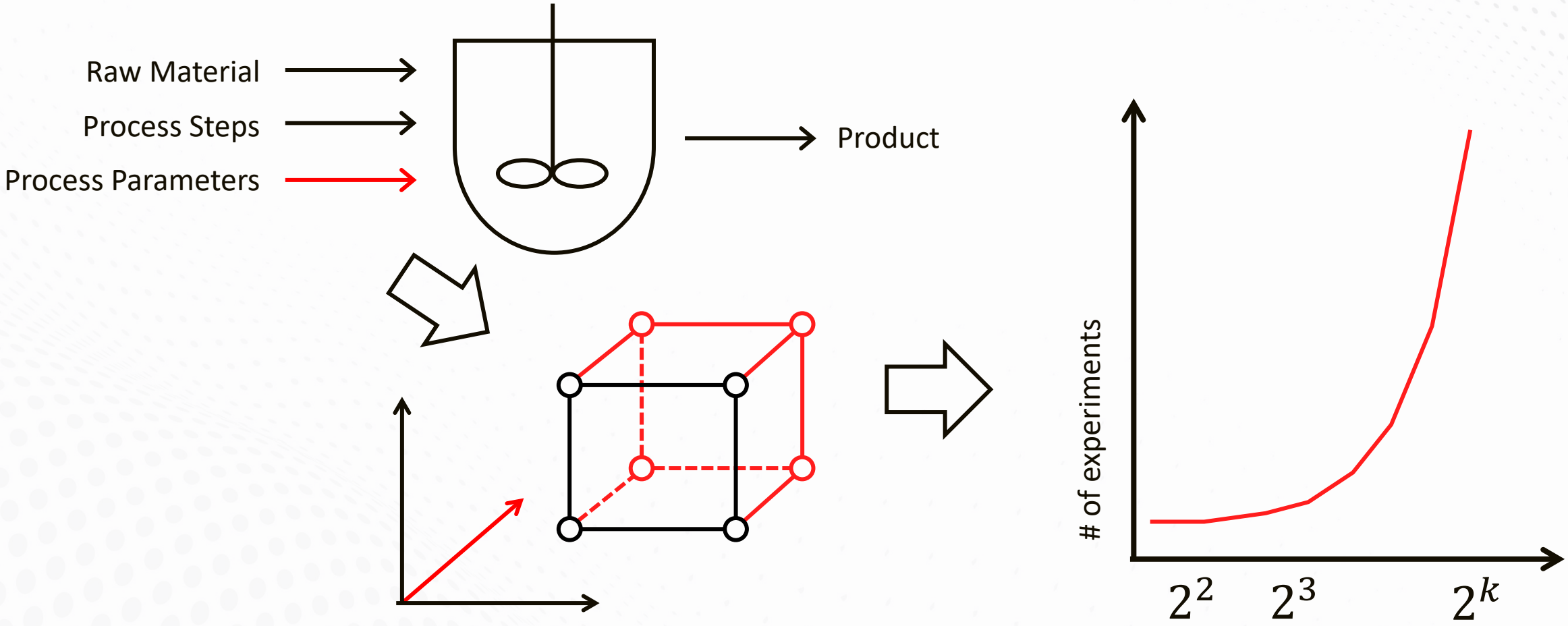
Why?
- Leverages all available data
- Further reduces data requirements
- Improves design of experiments
- Makes Quality by Design economically attractive

Process development in Biopharma

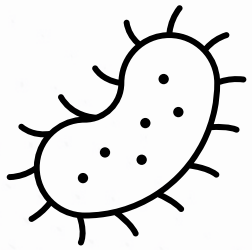
Process development’s role in the time-to-market race



Why does process development take so long?



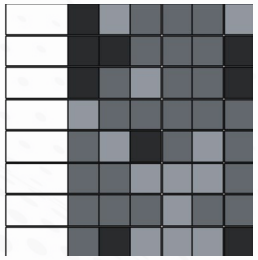
Common approaches to accelerate & de-risk process development

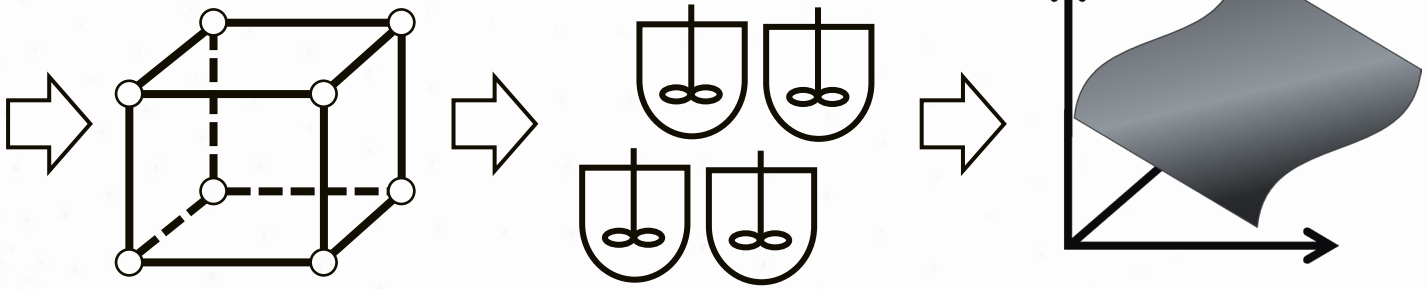


Platform processes
(CHO, Ecoli, Adenovirus,…)

High-throughput
experimentation

Quality by design
(Product centric)
development





Trends shaping the future of CMC development & manufacturing

Preventive Healthcare:

- Reducing margins to 13% or completely eliminating them

AI powered Drug Discovery & Clinical Trials:

- Increase in molecules in the pipeline shifting bottleneck to CMC

Biosimilars & Patent Cliff:

- Will reduce margins for blockbuster drugs, requiring

Cell Therapies & Continuous Processing:

- For these fields modeling no longer is an option but a necessity

Machine-learning:

- Preventive Healthcare
- Drug Discovery
- Clinical Trials
- CMC development & manufacturing

Data:

- Data-infrastructure
- Data standards
- Price of Data generation decreases

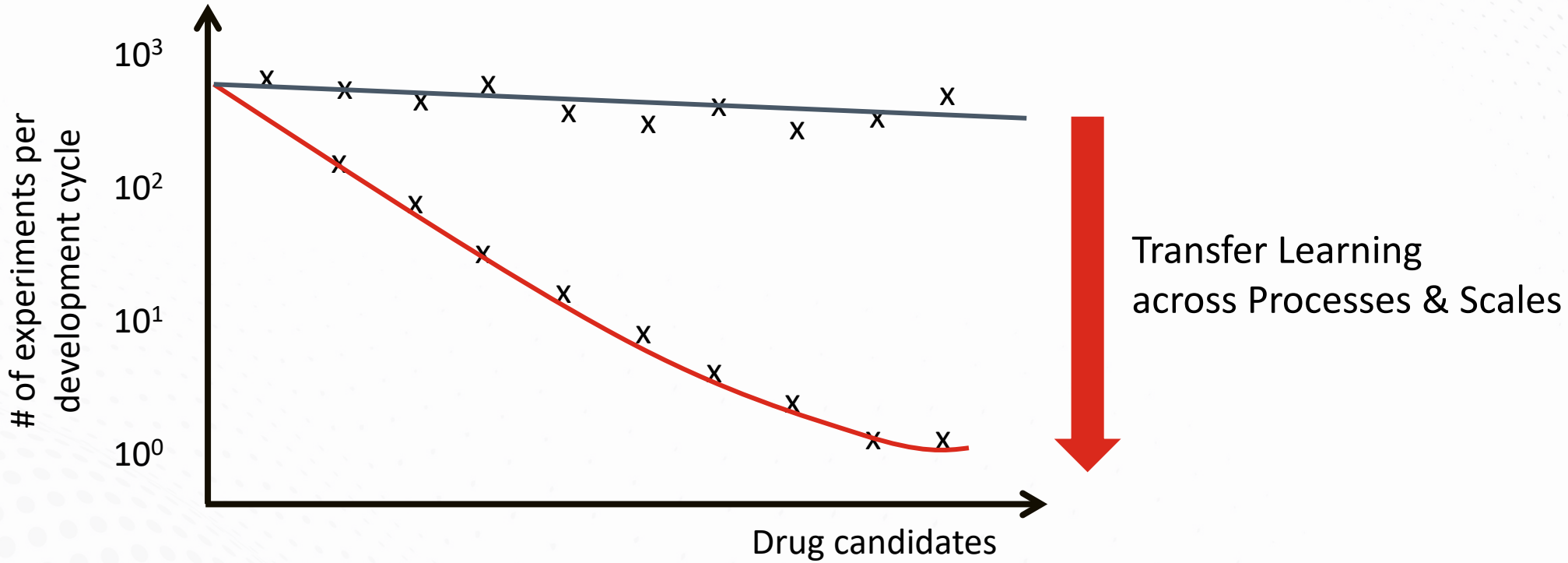
Automation & High-throughput:

- Increase of high-throughput process equipment & analytics will shift the bottleneck to design creation and data analysis

Analytics:

- Increasing availability of online analytics will increase the amount of data
- Increasing availability of off-line analytics (MCMS) will increase information about molecule

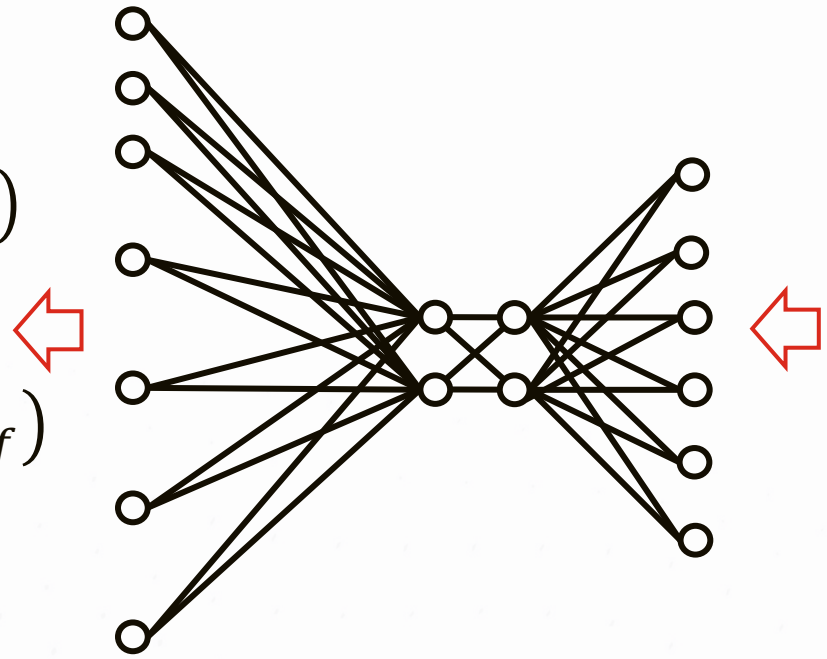
Develop Processes Faster Every Time



Transferring Learning between Processes

The hybrid model for transfer learning

$$\frac{dX_v}{dt} = X_v \cdot (\mu_v - k_d) - D \cdot X_v$$
$$\frac{dGlc}{dt} = -X_v \cdot q_{Glc} - D \cdot (Glc - Glc_f)$$
$$\frac{dLac}{dt} = X_v \cdot q_{Lac} - D \cdot Lac$$
$$\frac{dGln}{dt} = -X_v \cdot q_{Gln} - D \cdot (Gln - Gln_f)$$
$$\frac{dAmm}{dt} = X_v \cdot q_{Amm} - D \cdot Amm$$
$$\frac{dTiter}{dt} = X_v \cdot q_{Titer} - D \cdot Titer$$

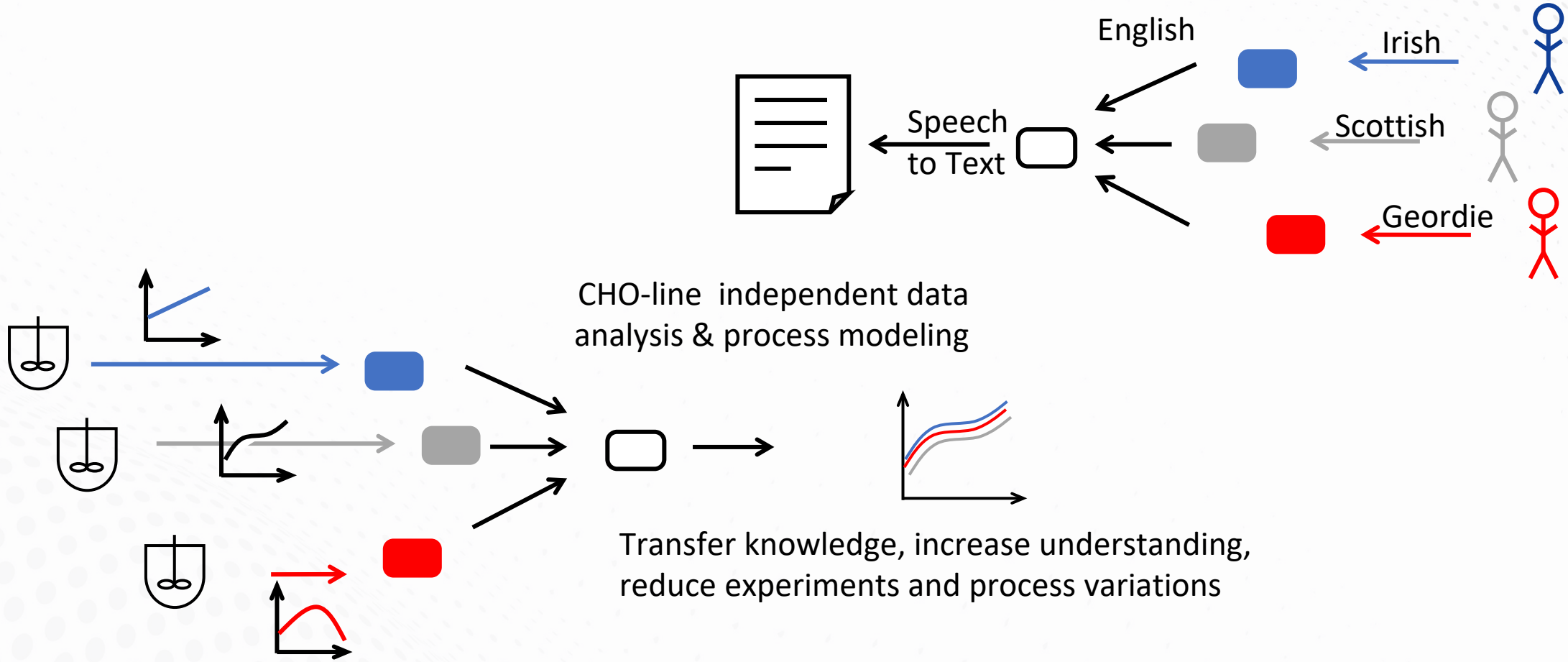


Cell-Line/Product category

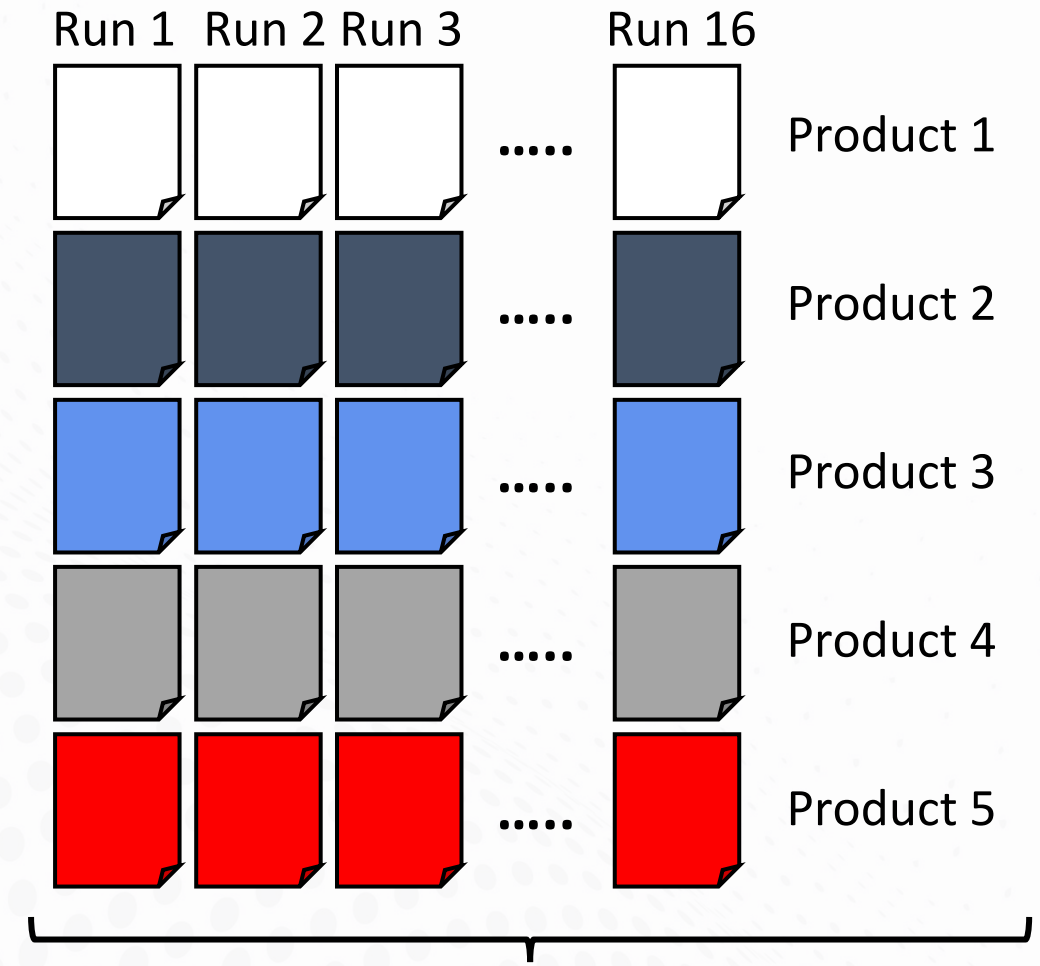
Process Parameter Data
- Medium type
- Stirring rate
- Etc…

Online Process Data
- pH / pCO_2 / …
- Online sensors
- Feed conditions
- Etc…

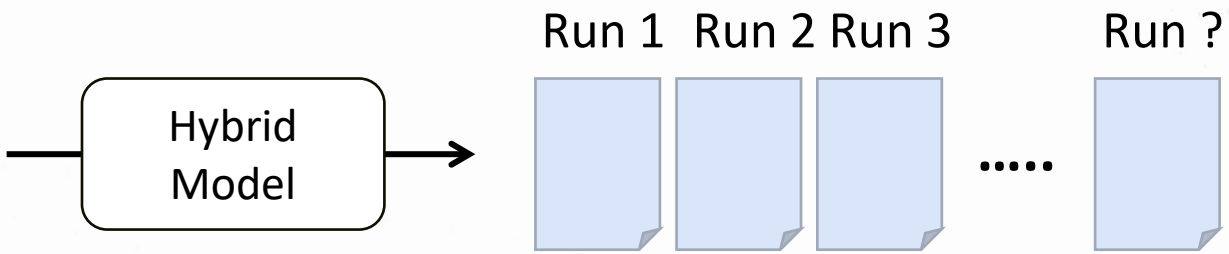
Can we use embedding technology for knowledge transfer to bridge between CHO cell-lines?



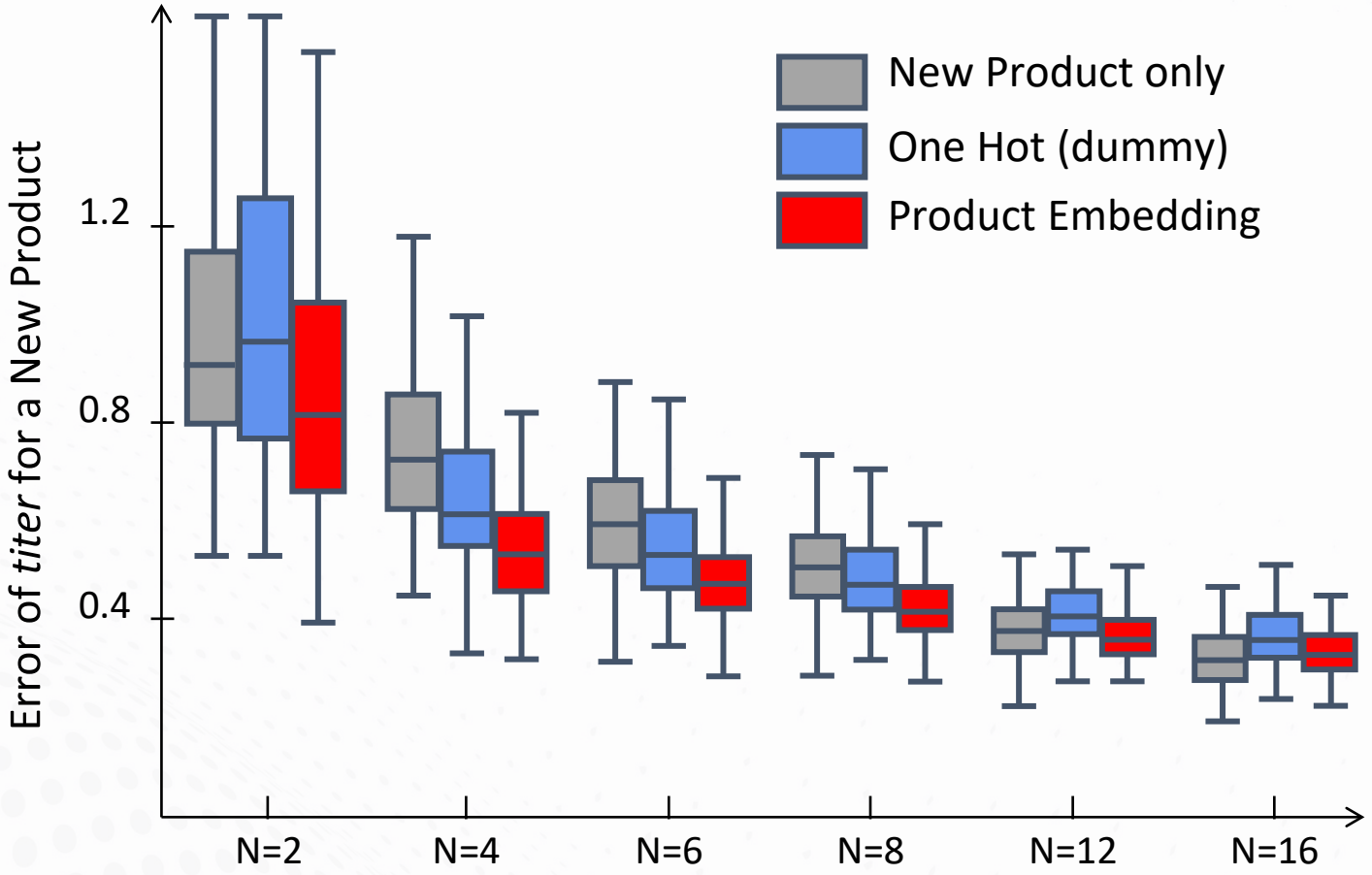
We trained the hybrid embedding model on data of 5 products and then asked ourselves: “How many runs do we need for a new product?”



How many runs do we need to learn the behaviour of the new process?

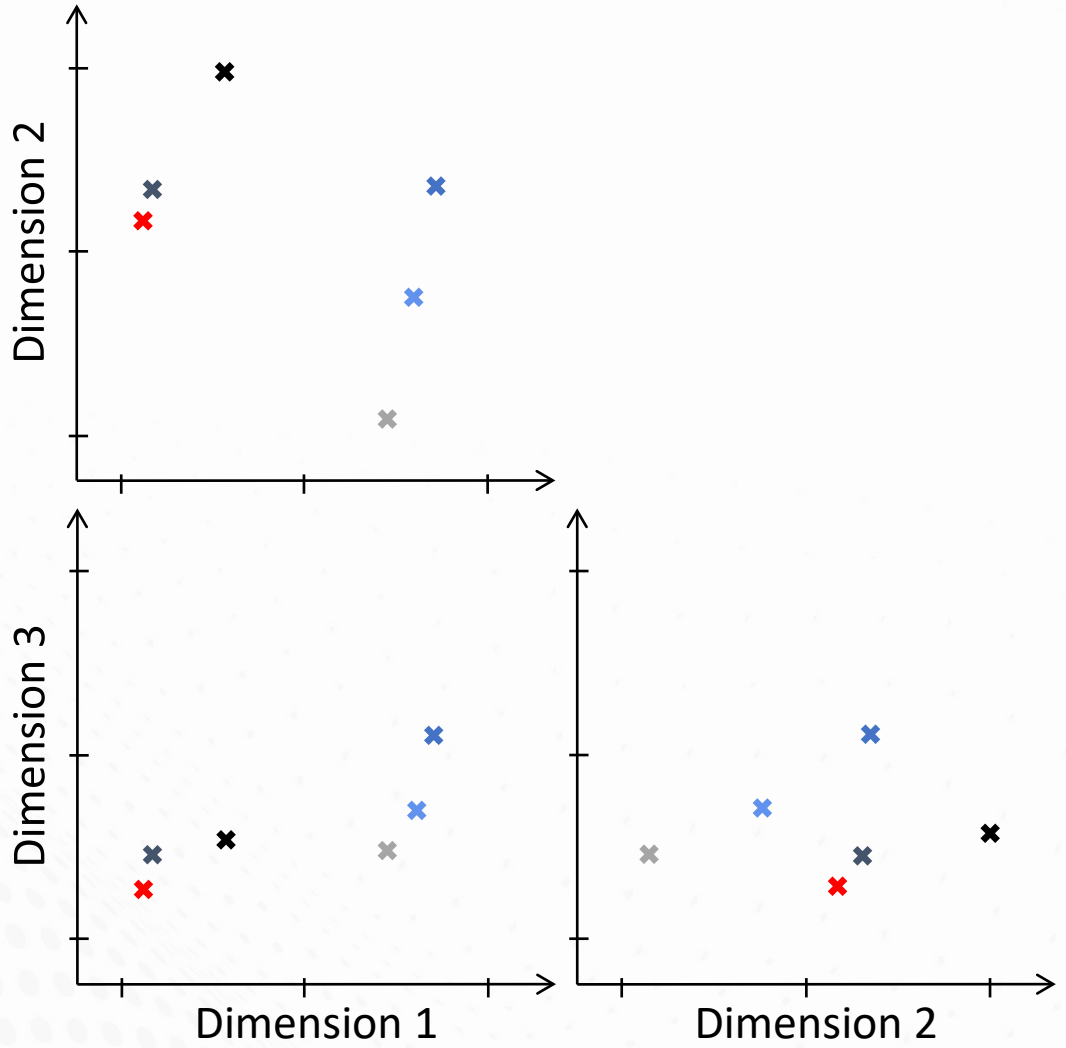


The embedding method outperform classical approaches in describing the systems behaviour with few experiments from novel process.

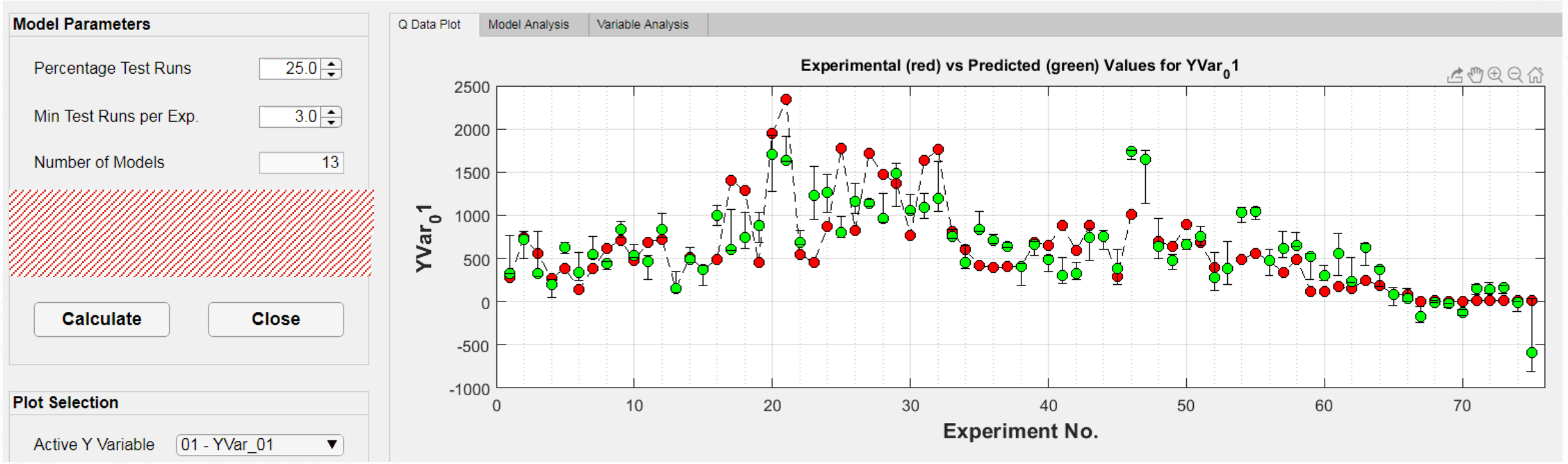


Knowledge transfer across cell lines using hybrid Gaussian process models with entity embedding vectors
Clemens Hutter, Moritz von Stosch, Mariano N. Cruz Bournazou, Alessandro Butté
https://doi.org/10.1002/bit.27907

Embedding technology allows relative comparison of cell lines in reduced dimensional space, gaining insight on their behaviour.



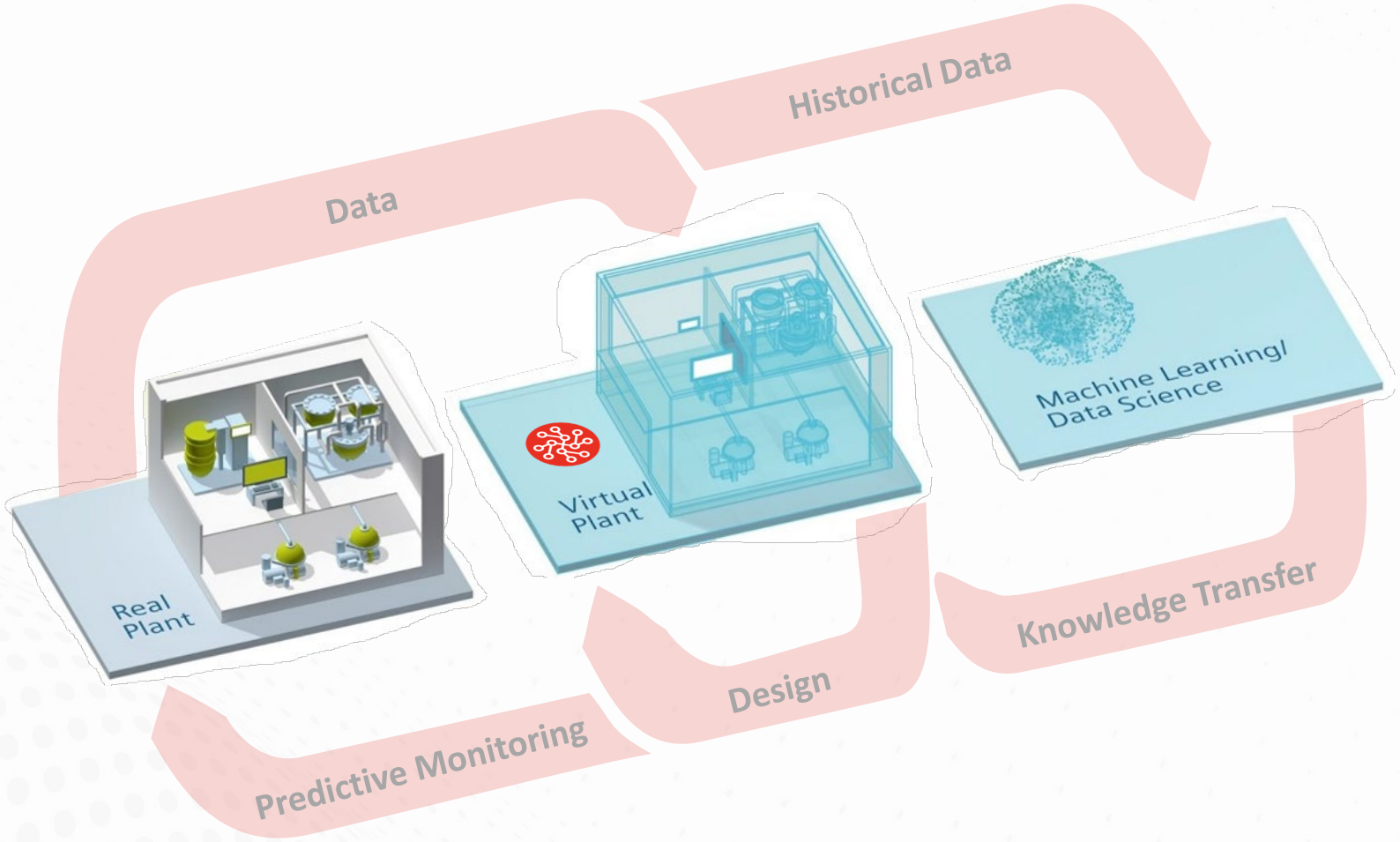
Practical Example: Embedding technology for predicting yields of a platform process in collaboration with a CDMO.



Process A | Pr. B | Pr. C | Process D | Process E | Process F

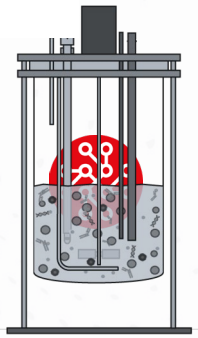
Digital Twins for model-based process development

What is a digital twin?



Start with the end goal in mind, what can the Digital Twin do for you?

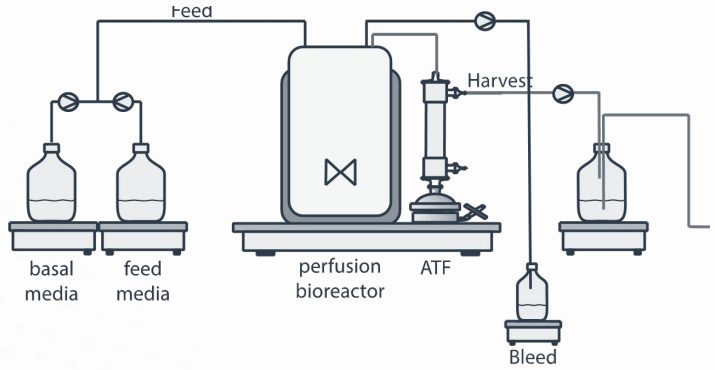
- Knowledge transfer & storage
- Training support
- Process understanding
- Decision support
- Forecasting
- Model based control
- Optimization
- Exp. Planning
- Predictive maintenance
- Automation
- Model driven validation
- Scale consideration
- Economic consideration

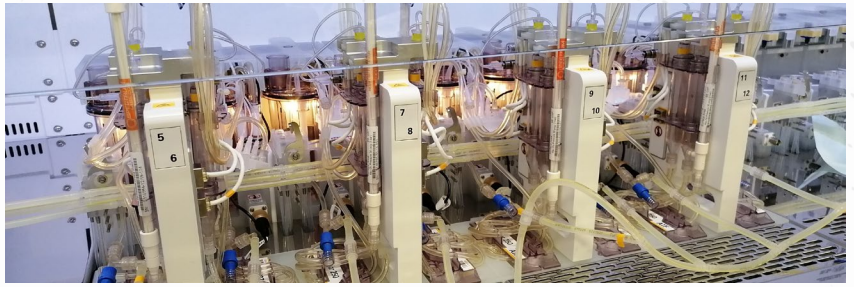


DigitalTwin

Meet the challenger

Why the ambr 250 perfusion system is the perfect challenge





Biological systems are much more complex than cars or turbines

Perfusion processes require a complex control and operation of many inputs in a highly sensitive process

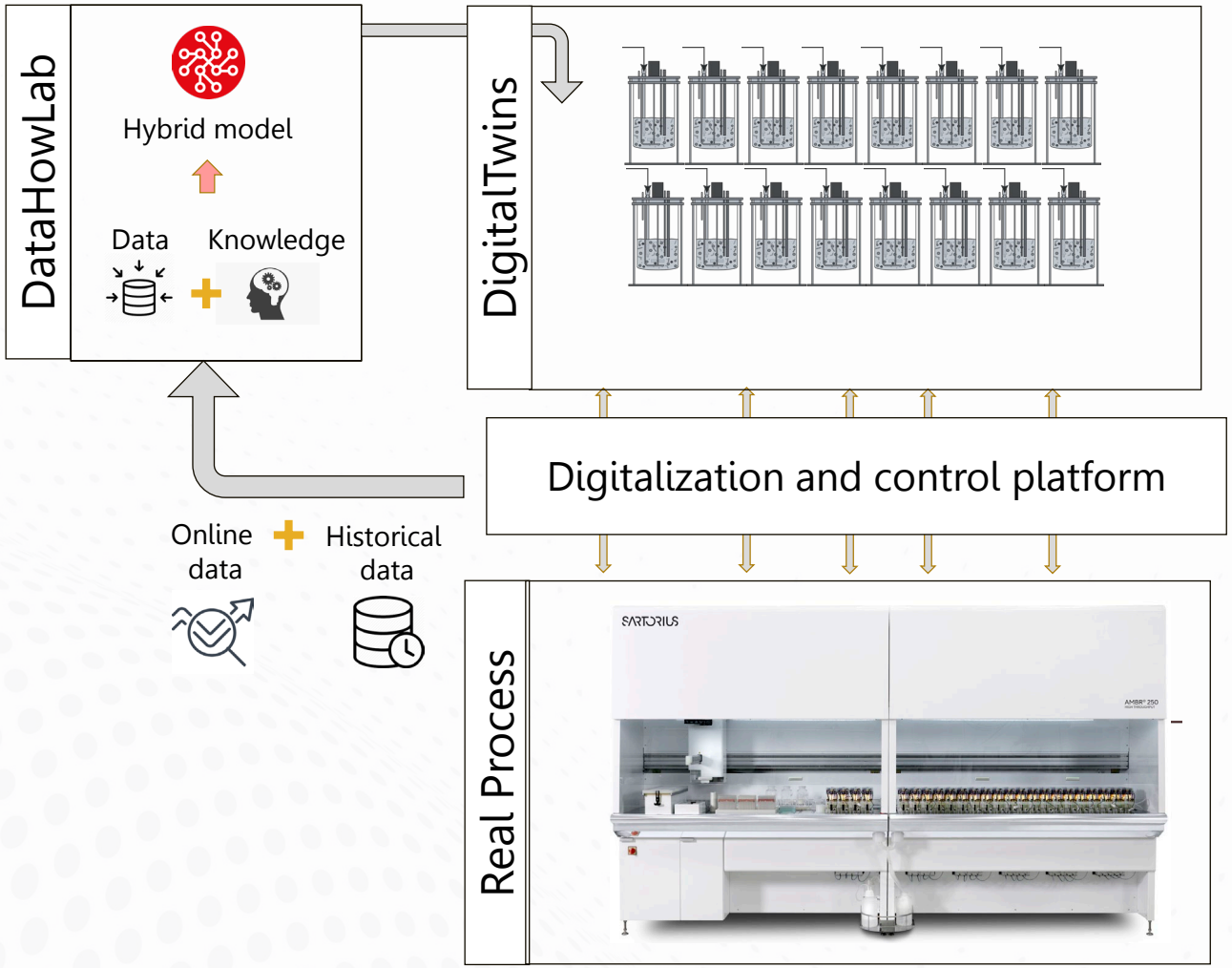
In High Throughput a large number of experiments must be operated at the same time

Perfusion experiments are significantly longer than fed-batch experiments

challenges

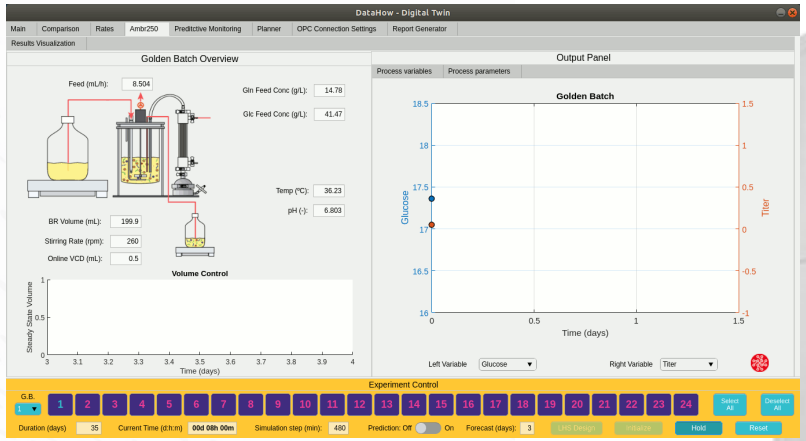
- process complexity
- advanced operation
- high parallelization
- long experimental runs

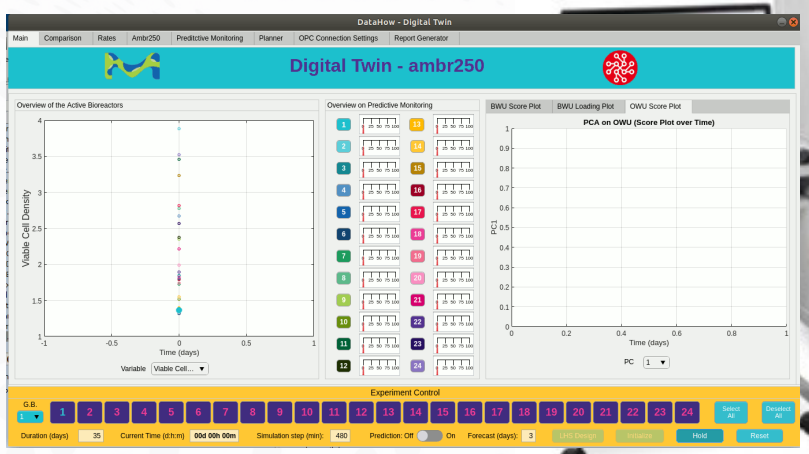
Answering to process development needs with a digital twin



- automation
- forecasting
- model based operation
- predictive maintenance
- exp. planning
- model driven validation

Examples of what could be done with this insilico bioprocess development platform





What-if & Design Space Analysis

Online Forecasting & monitoring

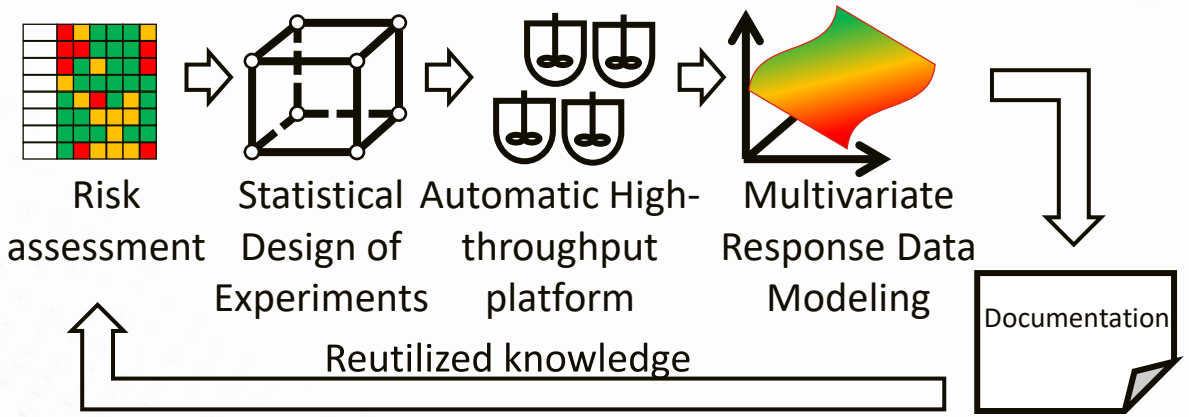
Data Visualization & Analysis

mbDoE & Process Optimization

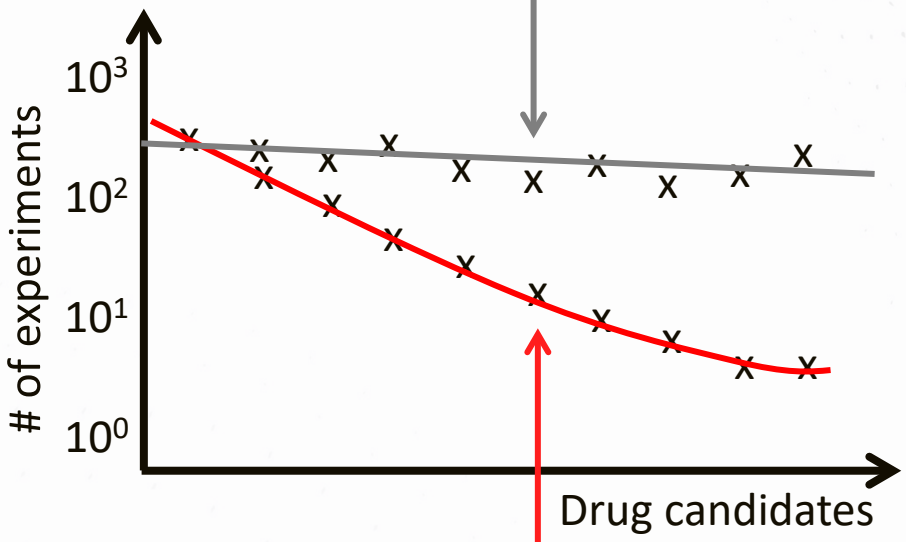
Summary

The traditional approach & the self-learning digital bioprocess twin

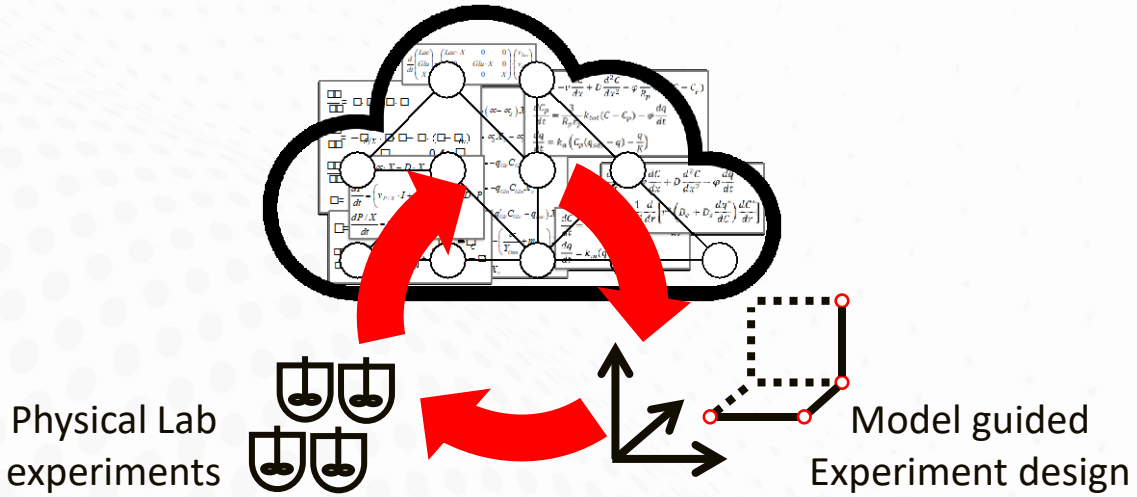
Traditional Quality by Design (QbD) approach



Process quasi developed de novo for every candidate.



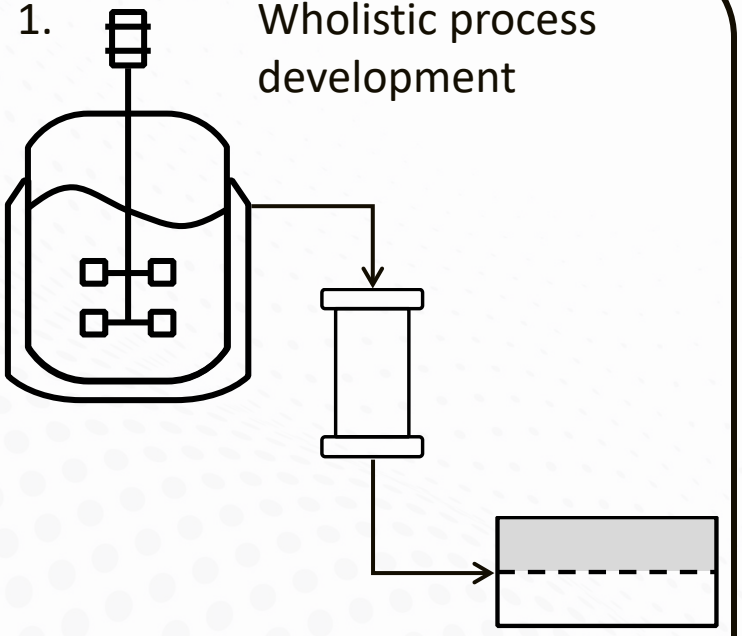
Self-learning digital bioprocess twins QbD2.0



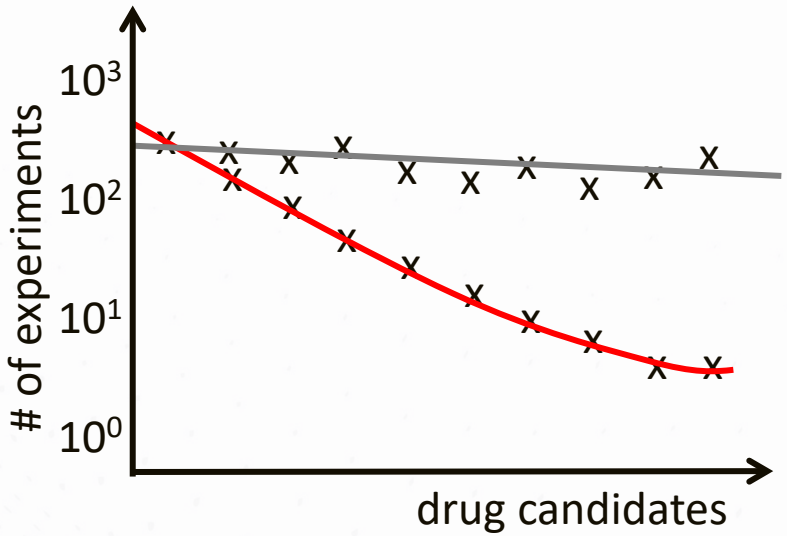
Progressively experiments are run *insilico* on the *Digital* Bioprocess Twins.

Examples of possible future model-based process development scenarios

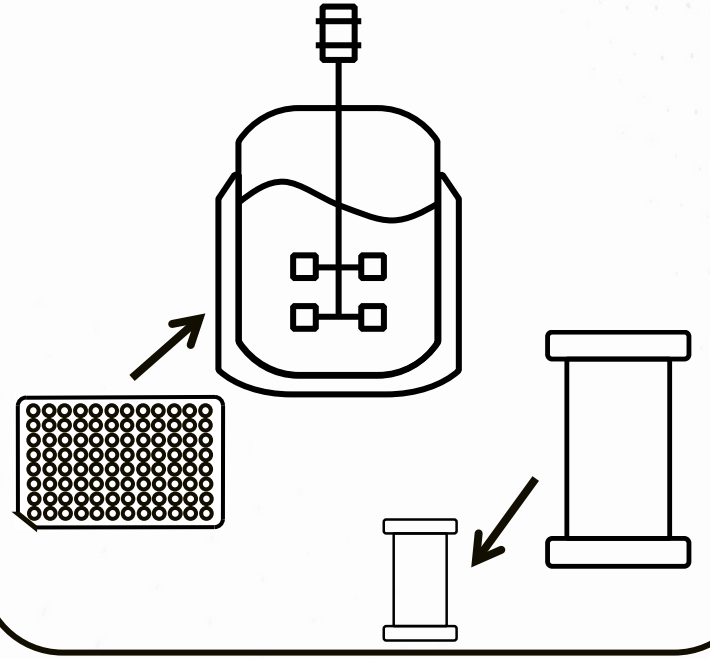
1. Wholistic process development



2. Across candidate learning



3. Across Scale learning



Thank you