Transforming Process Development with AI

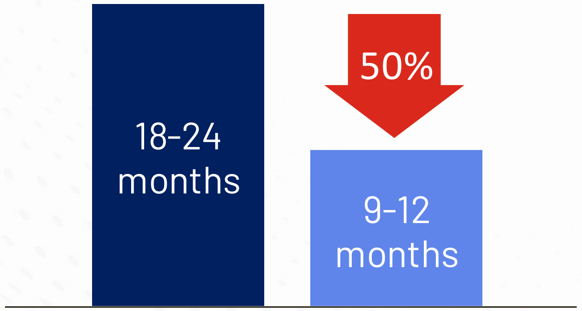
Bayesian hybrid models as enablers for ML-driven digital twins in biopharma productions

13 February 2025, Alessandro Butté, CEO and Co-founder, a.butte@datahow.ch

What is it achievable with AI-Enabled Digital Twins?

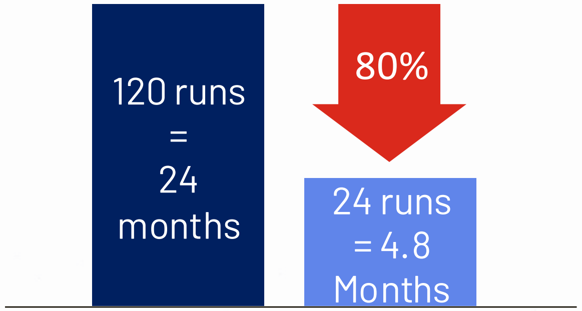
Process Development (Phase I) → **Phase II - III / Tech Transfer** → **Commercial Manufacturing**

Early Phase Development



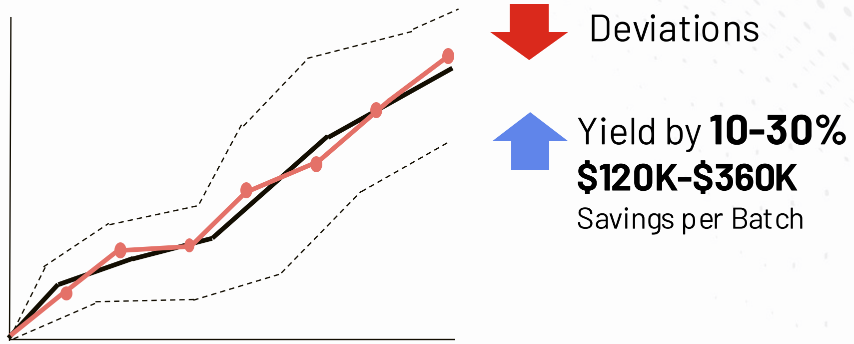
Lengthy Time to Clinical Trial Start

Late Phase Development



High volume of experimental runs required

Continued Process Verification



Variability of product quality and yield

The cost of delays in upstream scale-up can run into hundreds of thousands of dollars a day

DATAHOW

DATAHOW Rockwell Automation

Trusted and deployed by many key industry players

Customers & Partners: A result of more than 10 years of Thought Leadership and First Mover Collaborations

Customers

> **150** industrial process data sets analyzed by DataHowLab
> **40** peer-review publications

> **400** trained DataHowLab users across approx. 35 companies including 14/20 largest pharma and 6/10 largest CDMO

Technology Partners

Academic Partners

Who is DataHow?

Some numbers

2017

Incorporation

2024

Series A investment

ca. 40

Team Members

(Zurich – Milan – Lisbon – Philadelphia)

3

Customer Regions

AI-ENABLED
PROCESS DEVELOPMENT
& MANUFACTURING

WHAT WE DO

DataHow leadership: domain experts & innovators in process modelling

Meet the team

Dr. Alessandro Butté
Co-founder & CEO

M.Sc. Chemical Engineering
PhD Chemical Engineering
International Executive MBA
20 years industrial experience
Co-author of over 100 papers
Member of Italian Academy of Technology and Engineering

Dr. Michael Sokolov
Co-founder & COO

M.Sc. Chemical Engineering
PhD Bioengineering
International Executive MBA
Expert in process development with more than 30 papers

Dr. Fabian Feidl
Co-founder & CTO

M.Sc. Mol. Biotechnology
PhD Bioengineering
International Executive MBA
Expert in process automation

Kevin Healy
Chief Revenue Officer

M.Sc. Chemical Engineering
>25 years industrial experience
Expert in Digital Industry Software

Mark Powell
Chief Commercial Officer

BSc Business Administration
Commercial Finance (GMCA)
International Executive MBA
> 10 years industrial experience

Prof. Massimo Morbidelli
Co-founder

Full professor Politecnico Milano ('86-'95) and ETH Zurich ('96-'19)
Co-author of over 650 papers
Member of the Italian Academy of Science

ACADEMIC LEADERS IN PROCESSES MODELING DEPLOYING DATAHOW'S PRODUCTS & TECHNOLOGY ACROSS 20+ BIG-PHARMA CUSTOMERS & HUNDREDS OF MODELED USP PROCESS DATA SETS

40+ publications on hybrid process modeling

DataHowLab
SpectraHow

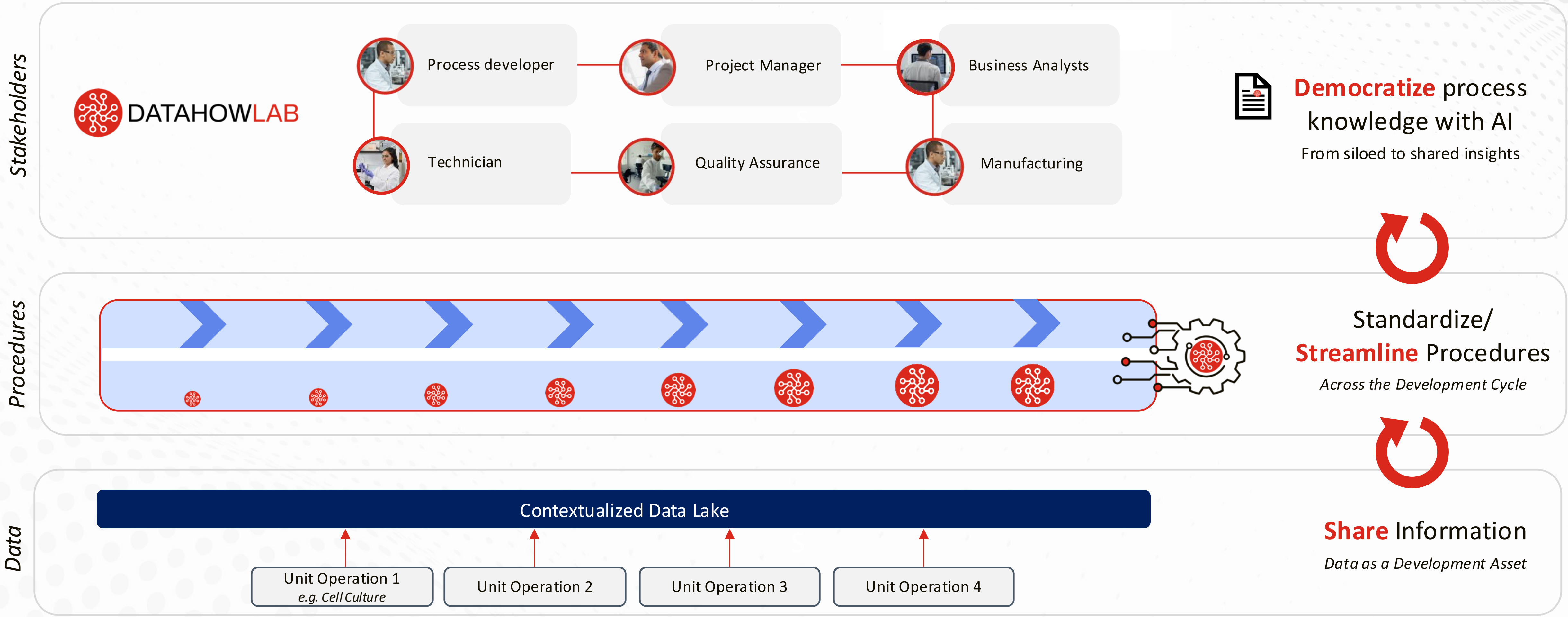
Hybrid Models
Transfer learning
Digital Twins

Enabling process optimization across a range of cases and scenarios

Unparalleled experience and industry process knowledge & best practice

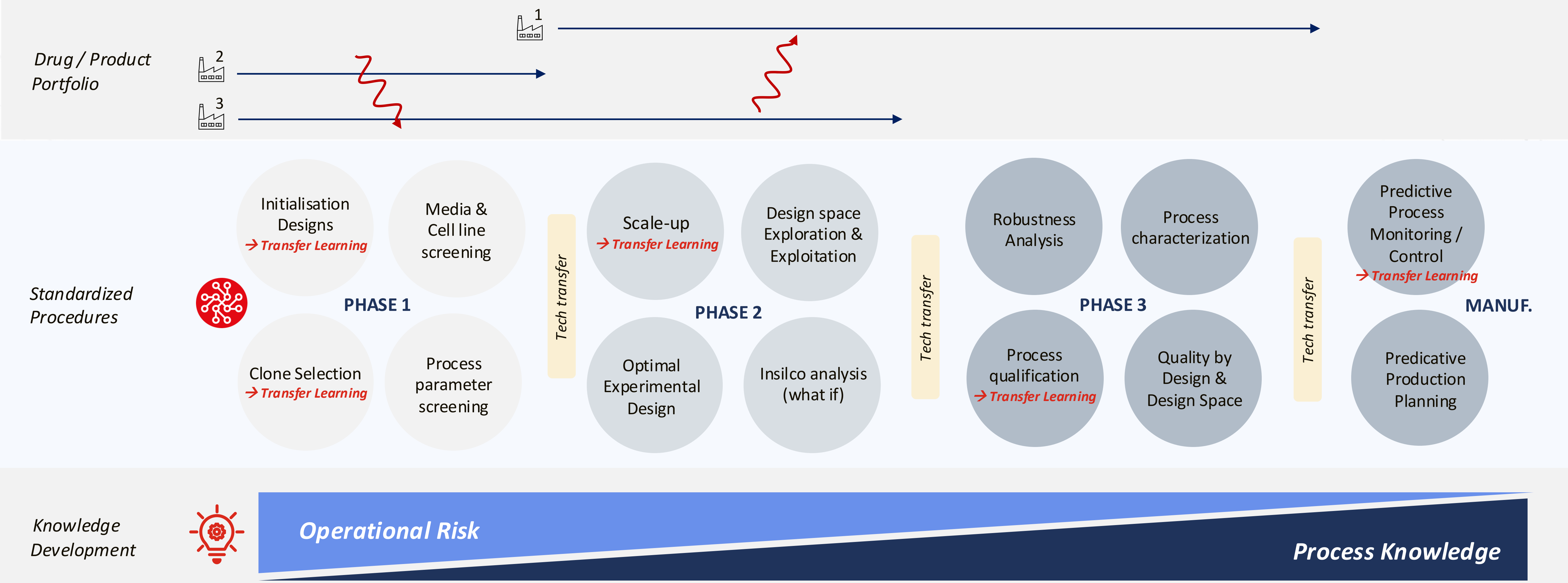
Change the paradigm: from Expert Models to Standardized Solutions

Enabling consistent outcomes, performance & knowledge management



Digital Platform Processes across Full Product Lifecycle

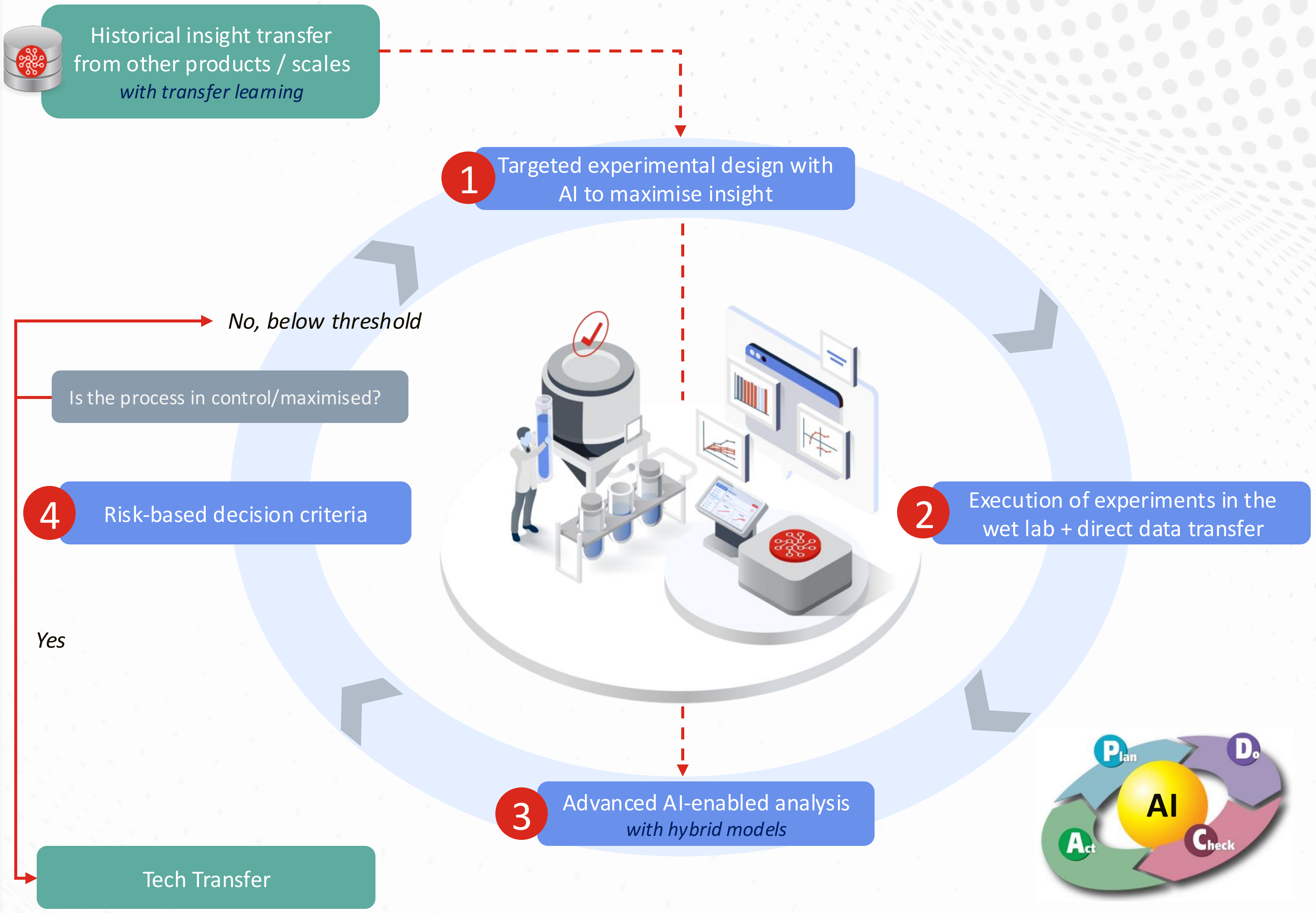
Systematic, efficient development of consistently productive processes



An AI-enabled, systematic approach to development

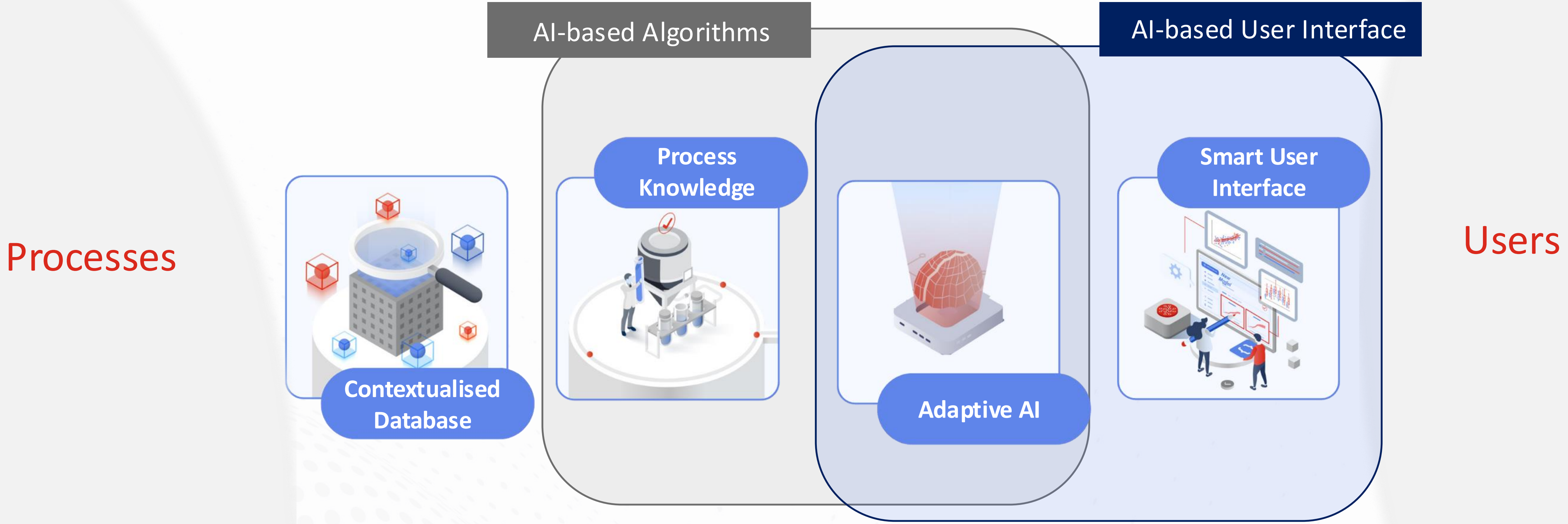
Digital platform processes

Delivering consistent outcomes, fast



Our Methodology Pillars

Adaptive Algorithms and Smart user interface. Hand in Hand.



Core Algorithms

AI-Technologies enabling advanced analytics & digital development

Hybrid Model

AI technology adapted to bioprocessing for ***accelerated learning***

Learn

Transfer Learning

AI-enabled ***transfer of historical process knowledge*** *across projects and scales*

Share

Risk-Based Decision Support

(Bayesian Statistics)

Outcomes & analysis assessed on old and newly available data, based on ***probabilities***

Manage

Bioprocess Digital Twin

Trusted ***digital representation*** *of the bioprocess*

Control

DataHow Hybrid Models

But what are they and how do they help?

Mechanistic: for what we know about the process

Mechanistic models code known engineering and process knowledge. They are not able to fully describe complex biological systems.

Hybrid models: best of both worlds

Explain what we know and use data to understand that which is unknown.

Data-driven (AI): for what we don't know

Machine learning models which determine relationships and patterns from raw process data to help explain complex relationships.

IMPACT: Better insight to develop better processes, faster

- Better prediction and control of quality attributes
- Increased yields and process performance
- Accelerated learning and increased operational efficiency

Hybrid Modeling of Upstream Bioprocesses in a Nutshell

$$\frac{dVCD}{dt} = \mu \cdot VCD - F_b \cdot VCD$$

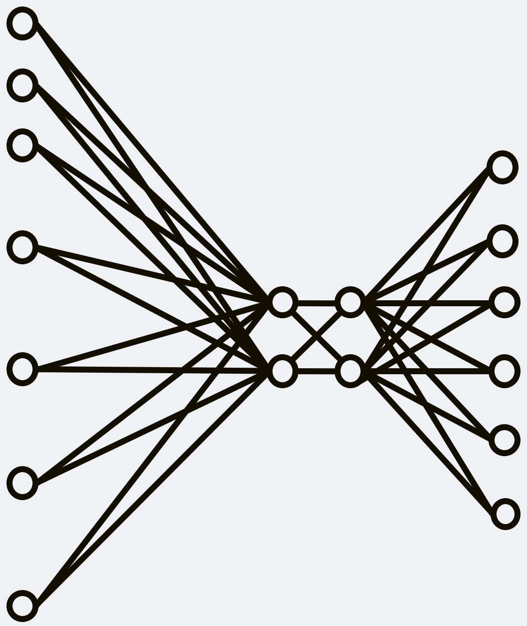
$$\frac{dGlc}{dt} = -q_{Glc} \cdot VCD - (F_b + F_p) \cdot VCD + F \cdot Glc_f$$

$$\frac{dGln}{dt} = -q_{Gln} \cdot VCD - (F_b + F_p) \cdot Gln$$

$$\frac{dAmm}{dt} = q_{Amm} \cdot VCD - (F_b + F_p) \cdot Amm$$

$$\frac{dLac}{dt} = q_{Lac} \cdot VCD - (F_b + F_p) \cdot Lac$$

$$\frac{dTiter}{dt} = q_{titer} \cdot VCD - (F_b + F_p) \cdot Titer$$



$$Process\ Variables \xrightarrow{ML} q_i$$

Process Variables

- Medium type
- Stirring rate
- pH / DO / pCO2
- Temp
- Metabolites
- Cell conc.
- Etc…

Acc. | Rates | VCD | Perfusion / Bleed | Feed | Machine Learning Algorithm | Process State

Bristol Myers Squibb

Hybrid Models vs. Industry State of the Art

Understanding & Predicting CQAs

Bristol Myers Squibb

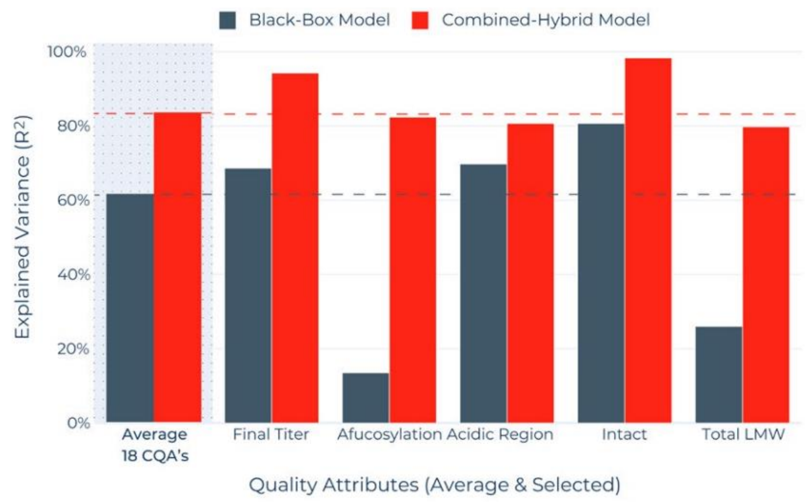
The Project:

Evaluate the ability of hybrid process models to accurately predict CQAs compared to industry state-of the-art "black box" models.

The Challenge:

48 (5 Liter scale) experiments were designed and conducted by BMS to evaluate the impact of **12 process parameters** on **18 product CQAs.**

Understanding & Predicting CQAs: Black Box vs Hybrid



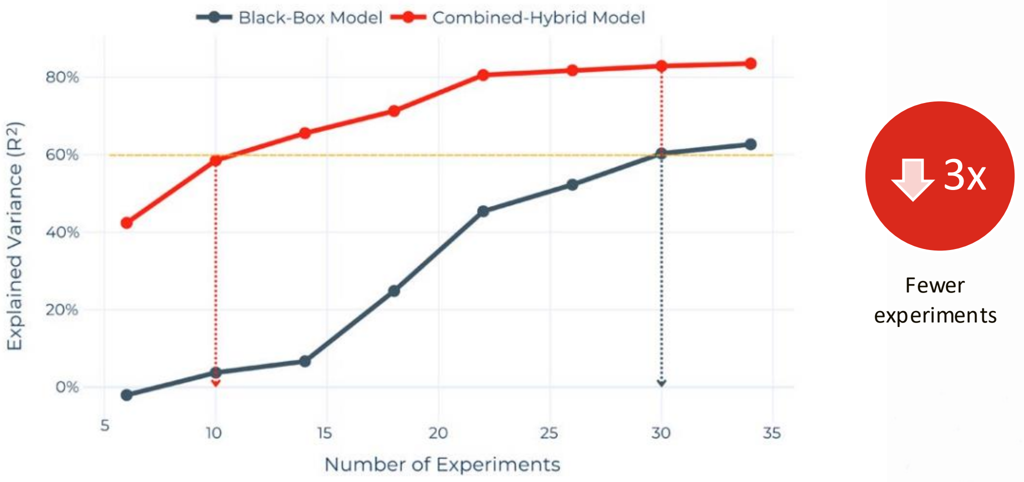
- On average, hybrid models explained **CQA variance +22%** better than black box
- Even after 34 experiments, black-box models were unable to reliably predict 5 of the 18 CQAs (highlighted: Afucosylation / Total LMW)
- For some CQAs, the predictive ability of hybrid models was approaching 100% (highlighted: Titer / Intact)

Bristol Myers Squibb

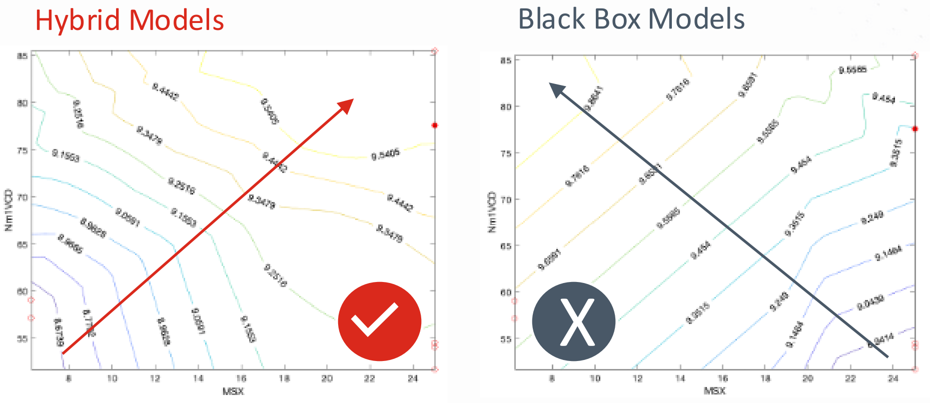
Hybrid Models more effective for development

Models and analytics used today not supporting development objectives

# of Experiments required to predict CQAs: Black Box vs Hybrid



Exploring the design space with:



- Black box models needed **30 experiments** before they could understand the CQA / process parameter interrelationships and reliably predict CQA values
- Hybrid models only required **10 experiments** to reach the same level of predictive accuracy

- **Hybrid models** accurately understood the complex interrelationships to suggest areas of further exploration
- **Black Box** models struggle to understand complex dynamics. They suggest further exploration in the wrong direction

J Polak et al. - Biotechnology Journal, 2024, 19 (3), 2300473

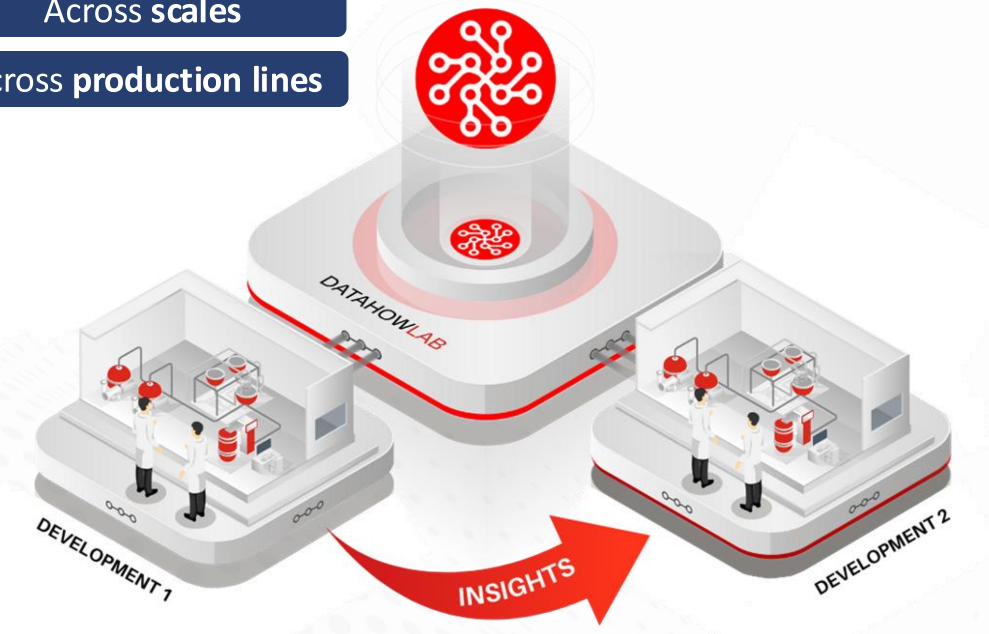
Transfer Learning

Leveraging historical data to inform & accerelate novel developments

Across **products**

Across **scales**

Across **production lines**



What is Transfer Learning?

- For products with a degree of similarity, historical insights can be used to inform novel developments using a meta-learning approach
- These insights can be used to design more targeted "calibration" experiments to accelerate early learning vs. starting from scratc

Reduce experimental effort

Accelerate development timelines

Leverage data as an asset

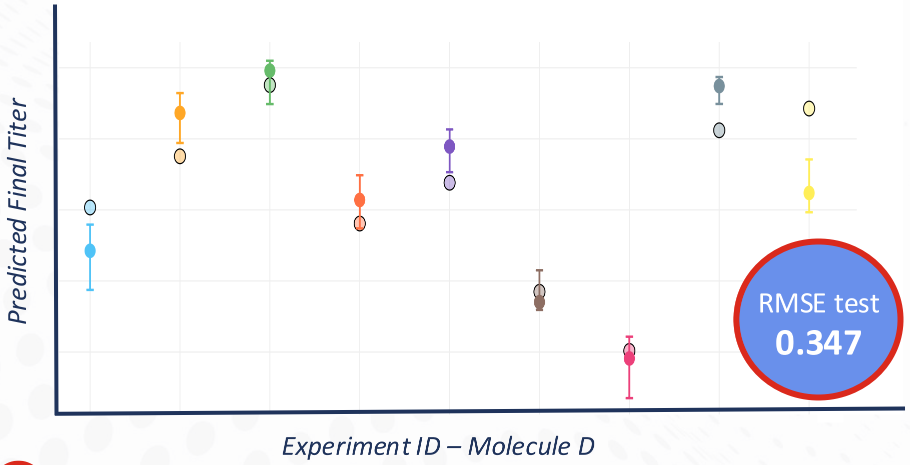
Accelerated Process Design via Transfer Learning

Predicting molecule specific behaviour (Molecule - D) with only 3 runs

Historical Molecules

- A – 15 runs
- B – 15 runs
- C – 15 runs
- D - 3 runs | Predicted values

Predicted vs Test Illustrate **high prediction accuracy**



Model trained on 15 runs of molecules A, B & C + 3 runs of D

- The behaviour of molecule D can be learned effectively from only 3 runs of D and transferred knowledge from molecules A-C.
- Transfer learning can significantly accelerate clone selection process
- The trained model could suggest optimal clone choice & process design

Transfer learning can also be applied across scales

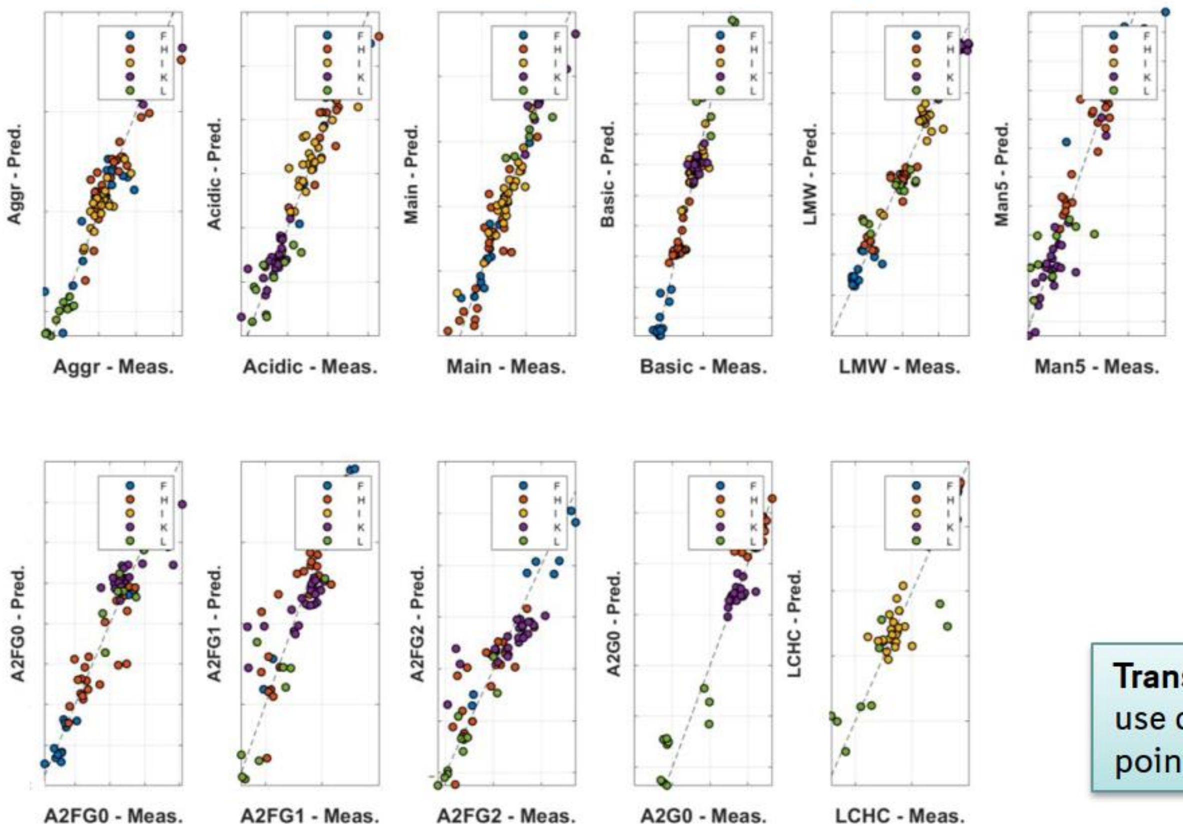
We have been able to learn and **predict large-scale titers** with no or only 1 mid-scale run before continuing to pilot scale.

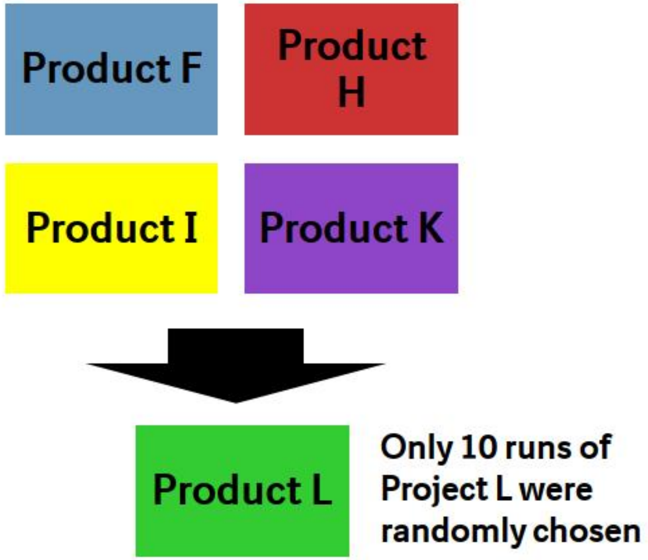
Boehringer Ingelheim

Example: Transfer knowledge for faster process chataterization

Generic model & transfer learning

Observed vs predicted product quality (**train set**)

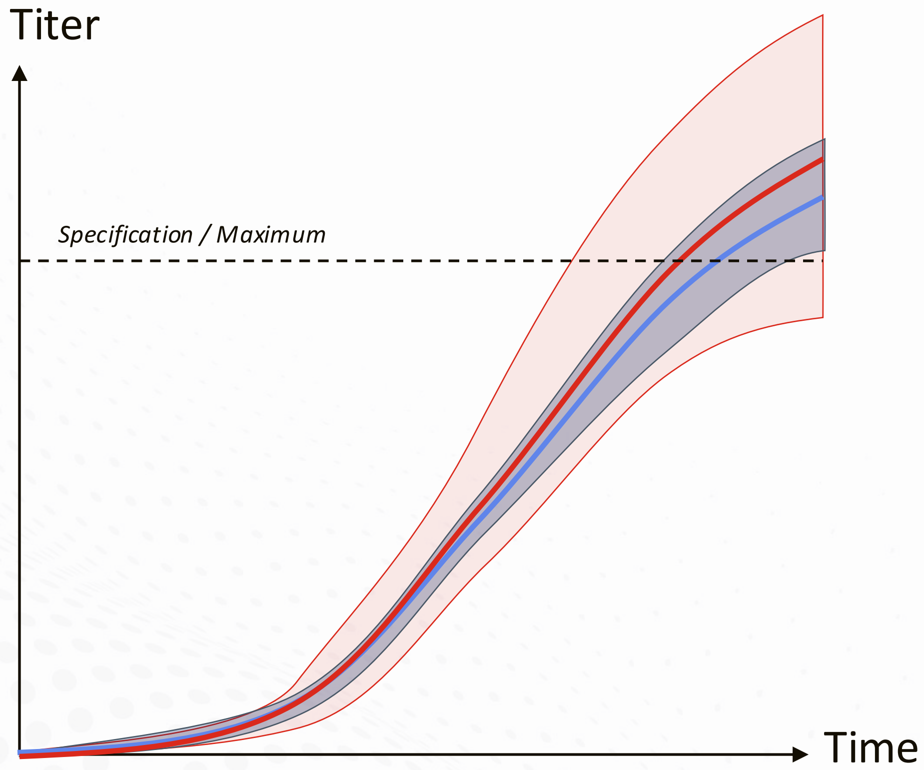




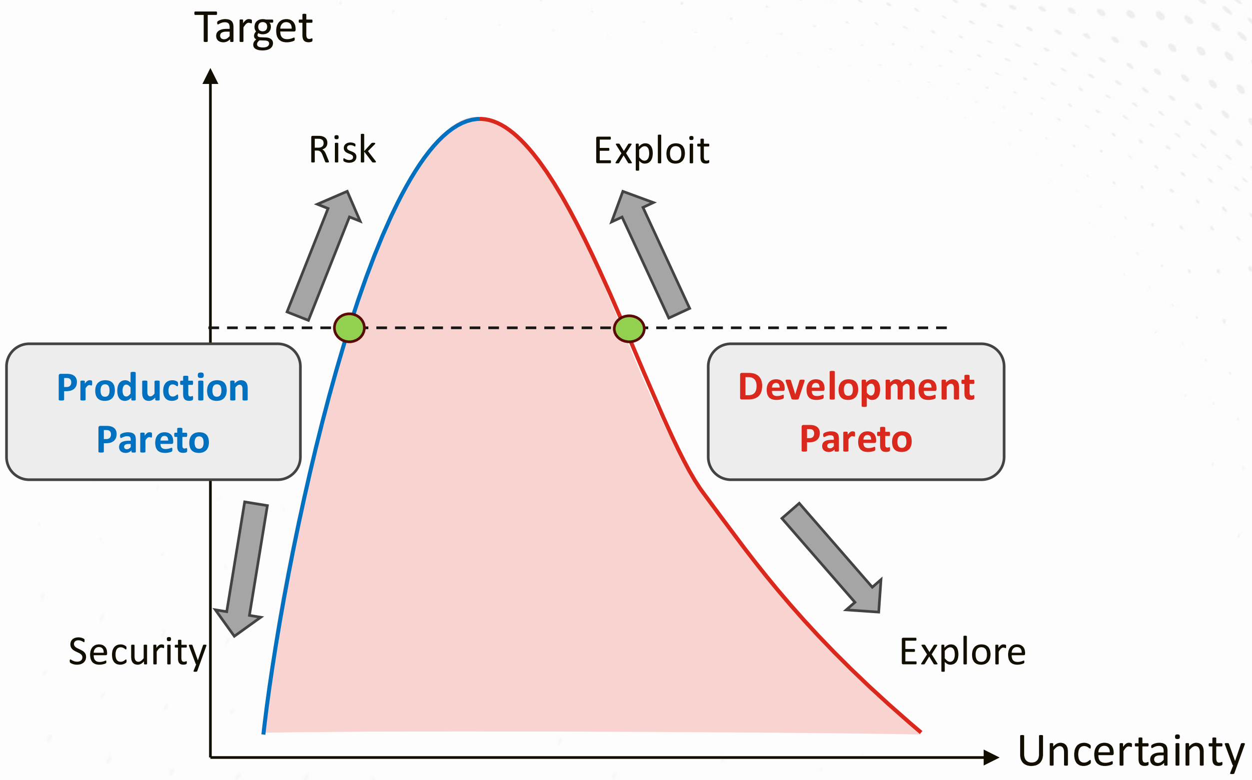
Transfer learning → a generic model allows to use cross-project information as a starting point for project-specific model development

Pareto Bayesian Optimization

How to manage uncertainty to optimize decisions



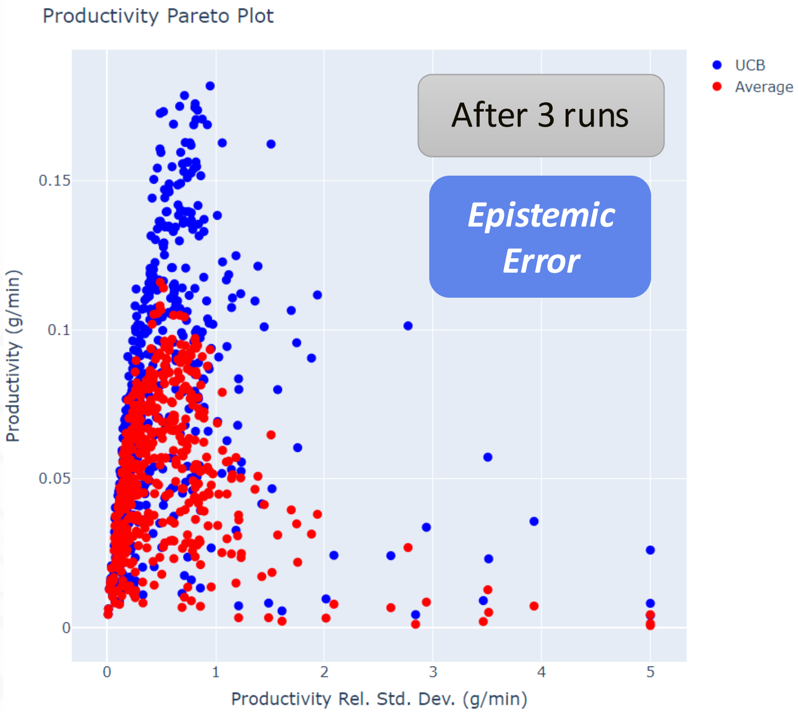
Learning Phase

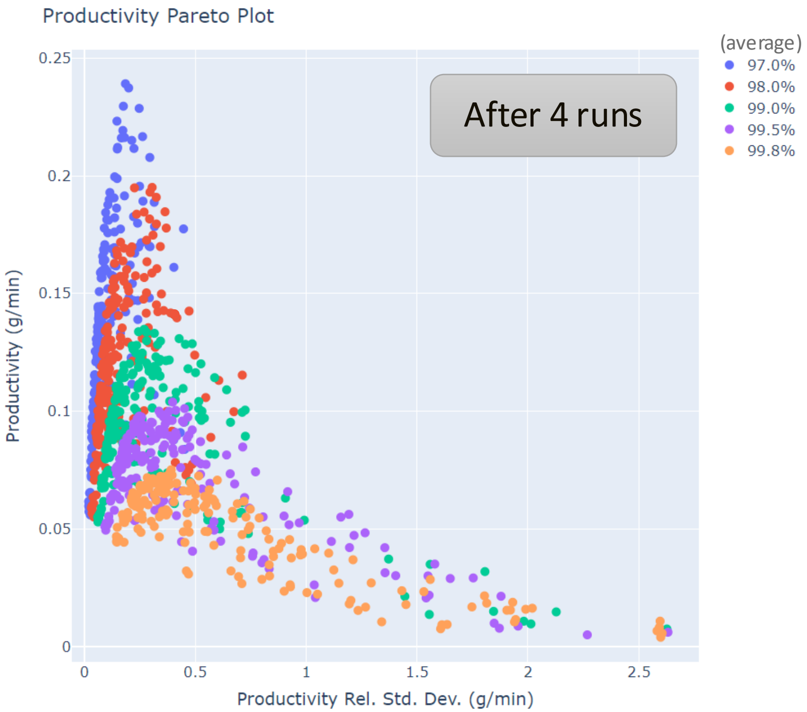


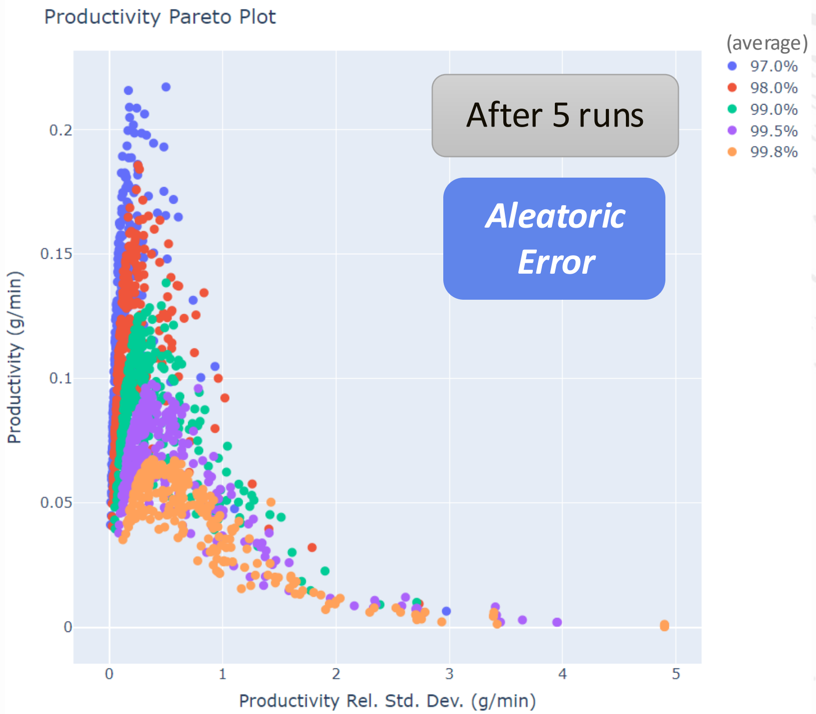
Decision Phase

Hybrid Models for 3-Component Gradient Separation

Design first experiment using the Pareto plot







C1feed	C2feed	C3feed	FlowRate	CVFeed	CVGradient	C1feedTrue	C2feedTrue	C3feedTrue	FlowRateTrue
0.05	1	0.15	1.00	3.00	40.00	0.05	1.08	0.15	1.00
0.1	1	0.05	1.50	1.00	30.00	0.10	0.94	0.05	1.50
0.15	1	0.1	0.50	2.00	20.00	0.15	1.14	0.10	0.50
0.15	1	0.15	1.65	4.26	32.27	0.15	0.97	0.14	1.66
0.15	1	0.15	1.90	4.01	33.32	0.15	0.94	0.16	1.89

Enantiomer Continuous Purification by SMB

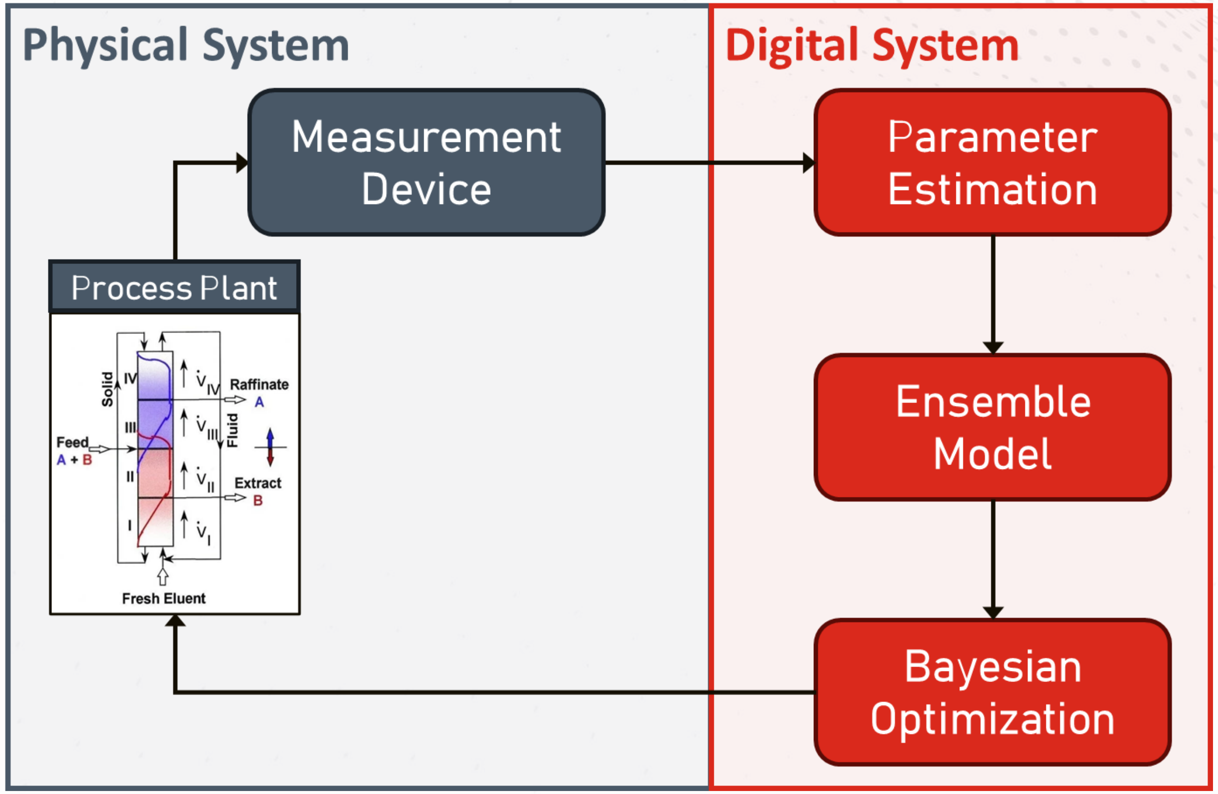
Real time learning and optimization of a 6 column SMB process

The Project:

Create a **robust process controller** for SMB enantiomer separation

The Challenge:

Optimize process conditions in real time in the presence of significant column degradation and variability, while minimizing batch experiments

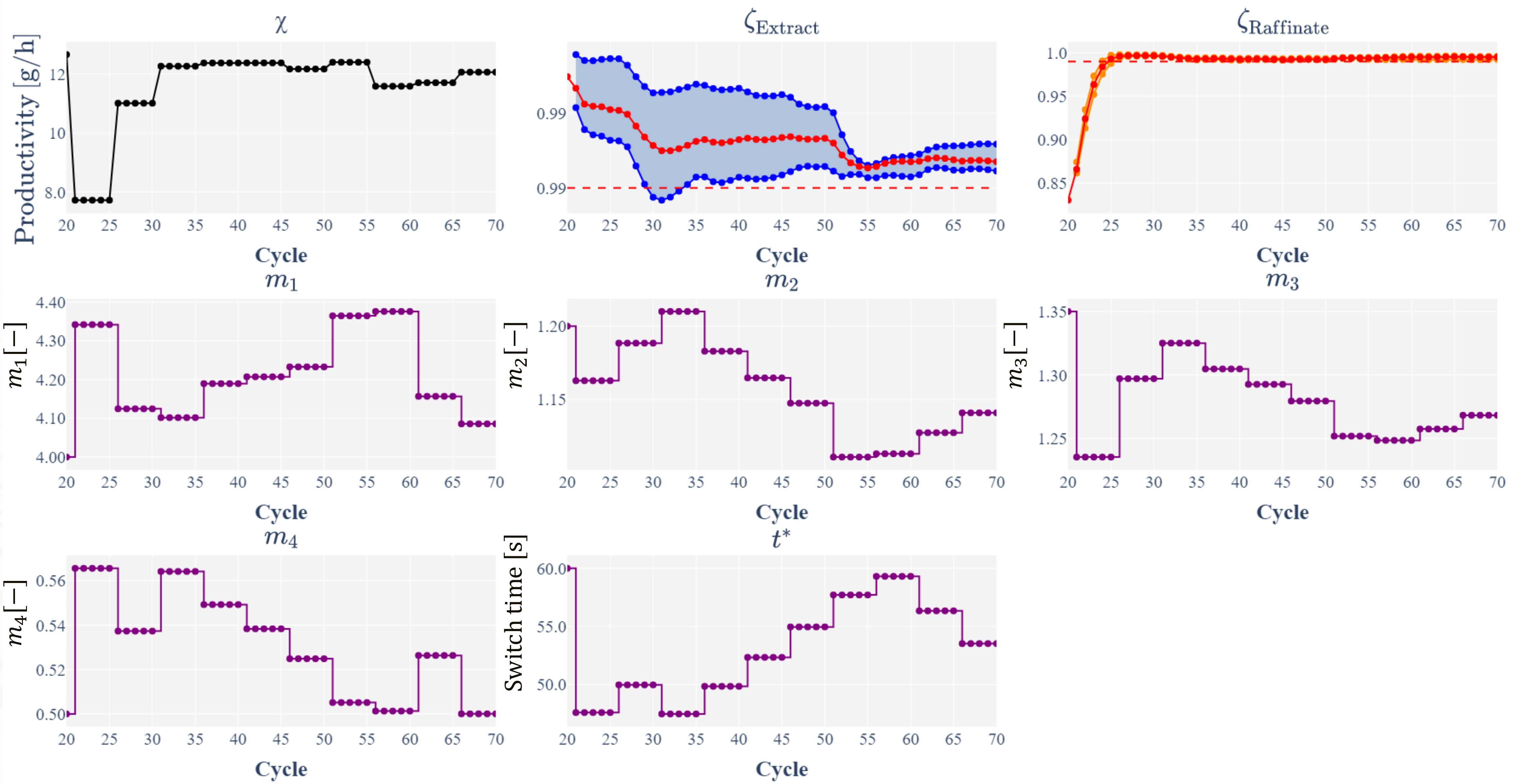


- **Objective**: productivity in 5 cycles
- **Constraint 1**: purity on extract and raffinate > 99%
- **Constraint 2**: yield > 99.5%

Digital Twin for 50 Cycles using UV + Average Purities

Case Study

Transfer Learning

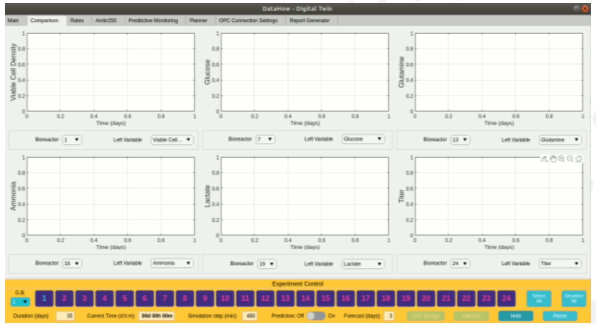


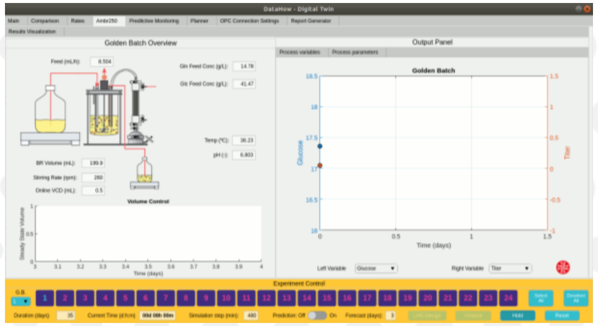
Case Study

Digital Twins

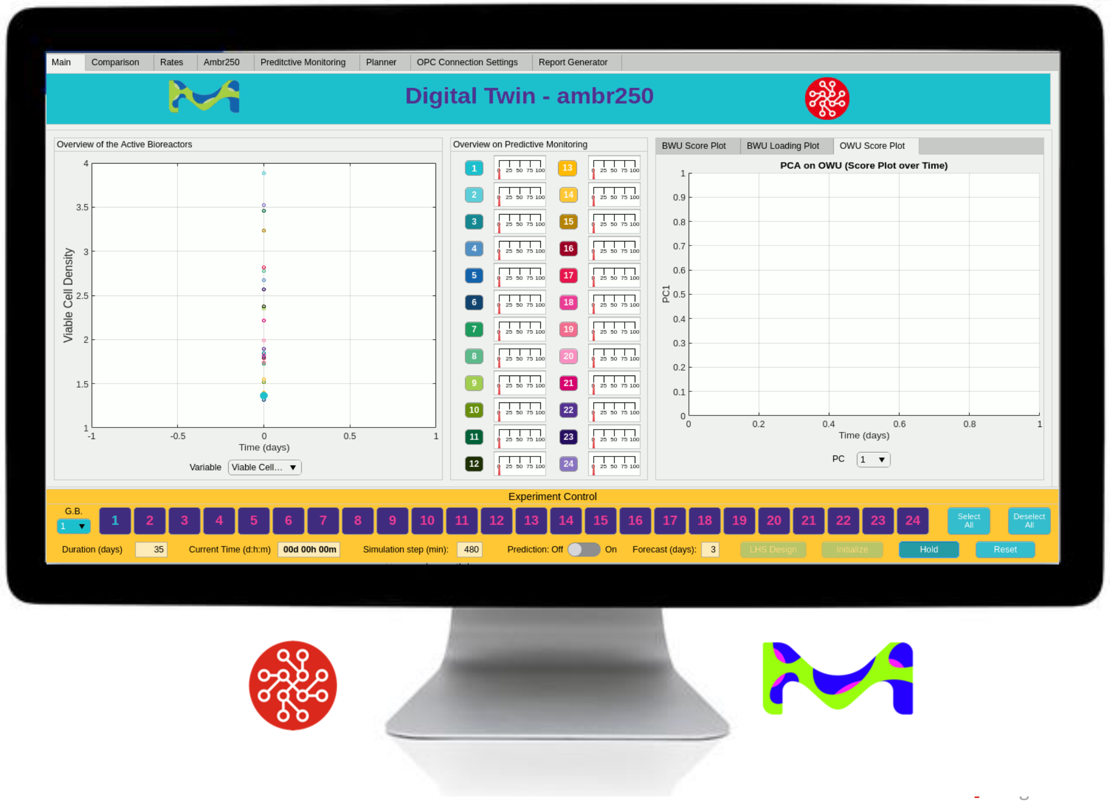
MERCK

Digital Twins for Ambr® Systems





- Real-time control of 24 parallel ambr250 bioreactors in perfusion mode with a digital twin
- Hybrid models from DataHowLab connected to device forecasted future behavior of all runs as basis for feed-forward predictive alarms and controls
- Active learning during run as basis for real-time process optimization and efficient learning in parallel system



A hybrid model is the best abstracted understanding of a process and how it will perform under varying conditions - your bioprocess digital twin

Advanced AI-Analytical Technologies working in the Background

Democratizing access and usability of advanced data-science for non-experts

Advanced Technology Core

- Hybrid Models
- Transfer Learning
- Bayesian Statistics
- Digital Twin

Easy Application

- User Friendly UX
- No-code analysis
- Structured methodology
- Guided workflows
- Automations
- Decision support

Make a model in minutes with DataHowLab's guided workflows

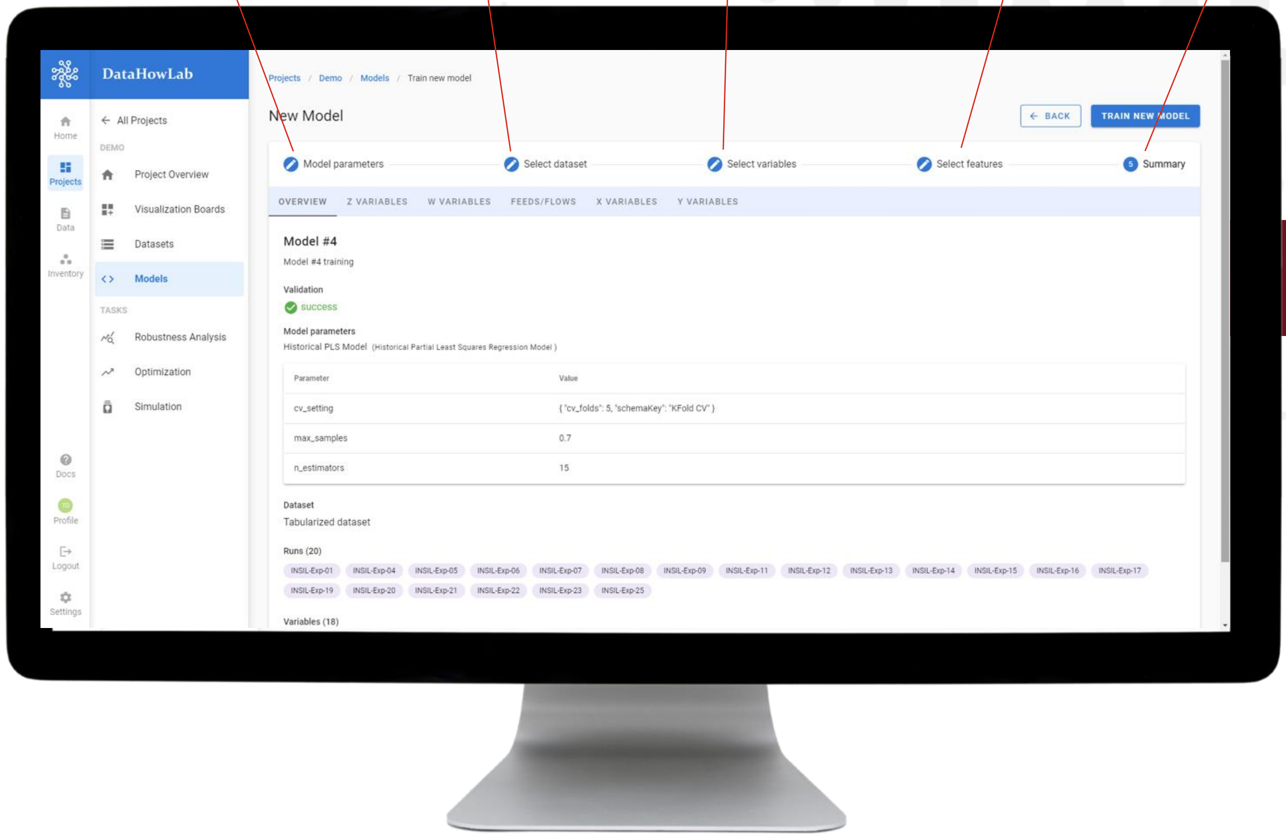
Gateway to model-based tasks

This step has been developed to make model development simple for non-data scientists.

The goal: access insights!

Supported by automations and tutorials

Select Model → Select Dataset → Select Variables → Select Features

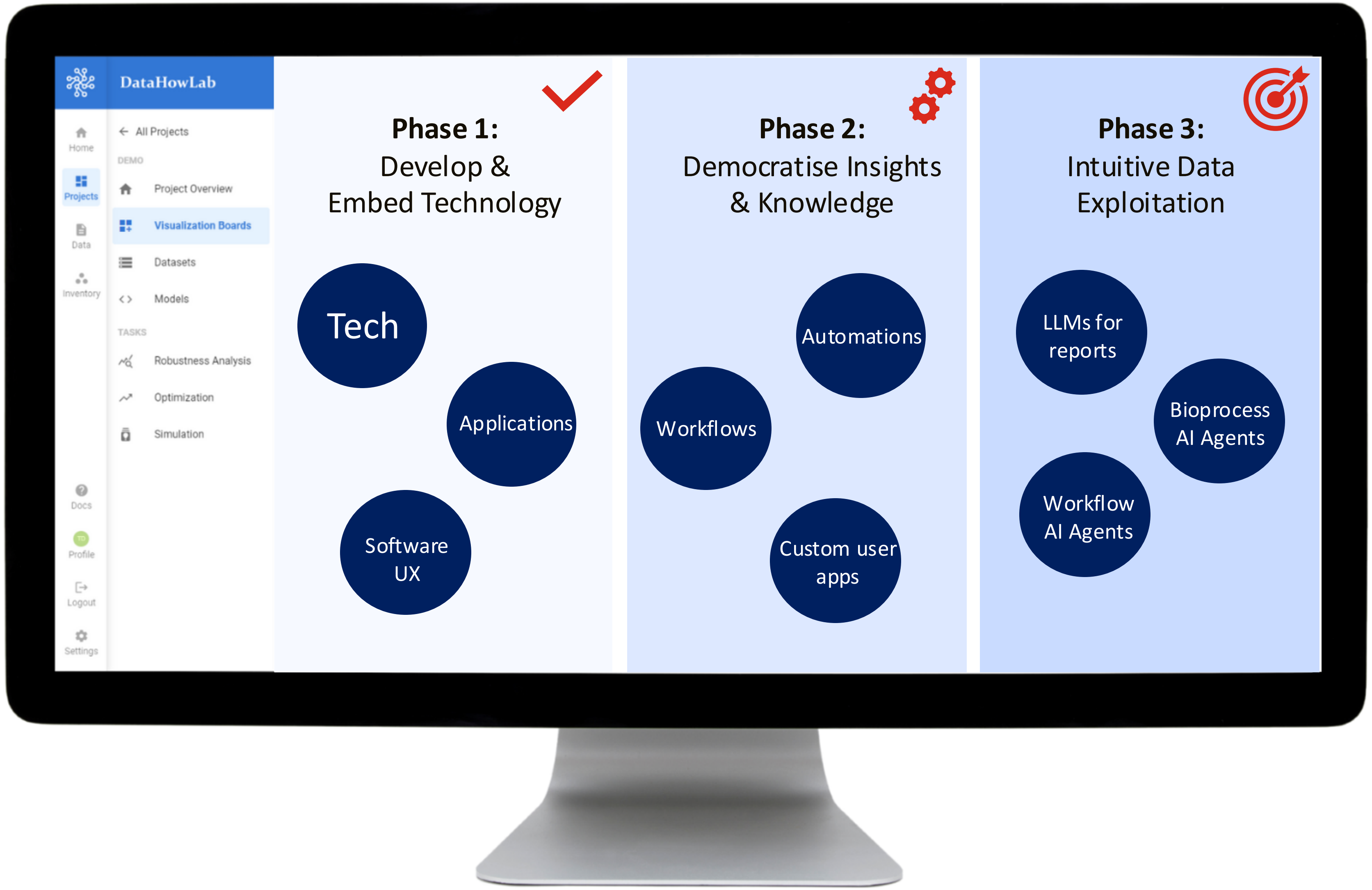


While providing context and traceability to key stakeholders

- **For the process scientist:** model evaluation – compare and select the best model for the project.
- **For the FDA:** model versioning – what has the model been trained on?
- **For the data scientists:** model logbook – what is behind the model?

The bioprocessing software for our AI future

DataHowLab development vision following the AI megatrend



Thank you

Alessandro Butté – CEO and Co-founder of DataHow AG
a.butte@datahow.ch – www.datahow.ch