# Predictive Power: Accelerating Bioreactor Process Development with Hybrid Models

**June 26th, 2025**

Eric Hodgman, Bristol Myers Squibb
Jakub Polak, Datahow

Bristol Myers Squibb™ DATAHOW

# Hybrid model definition

- Hybrid models combine 2 or more models together (i.e. Mechanistic + Machine Learning)

## Mechanistic

- Understanding of mass balances
- Monod equation (cell growth)
- Flux balance / metabolic flux analysis (metabolites, product)

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

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

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

$$\frac{dGlu}{dt} = -q_{glu} \cdot VCD - (F_b + F_p) \cdot Glu$$

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

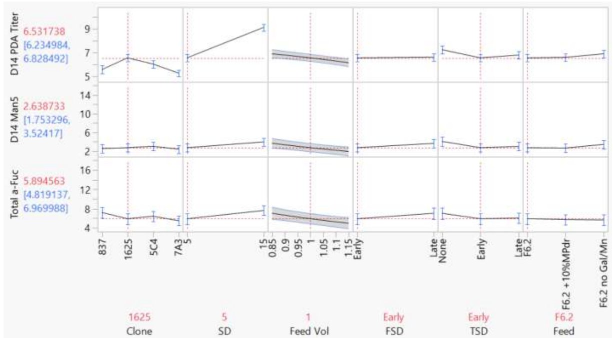
$$\frac{dNH4}{dt} = q_{NH4} \cdot VCD - (F_b + F_p) \cdot NH4$$

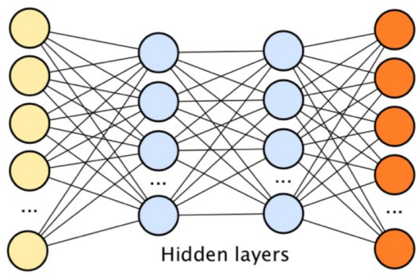
$$\frac{dTiter}{dt} = q_{titer} \cdot VCD - (F_b + F_p) \cdot Titer$$

+

## Machine Learning

- PLS regression
- DoE / LHD design

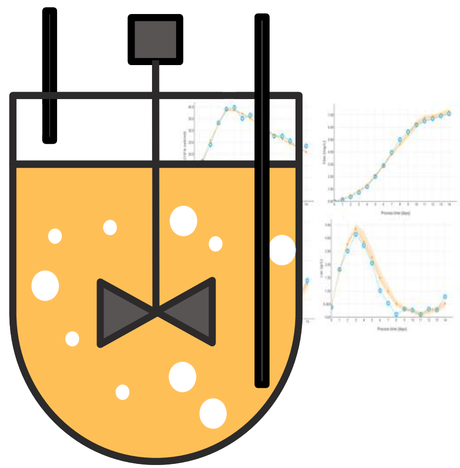



=

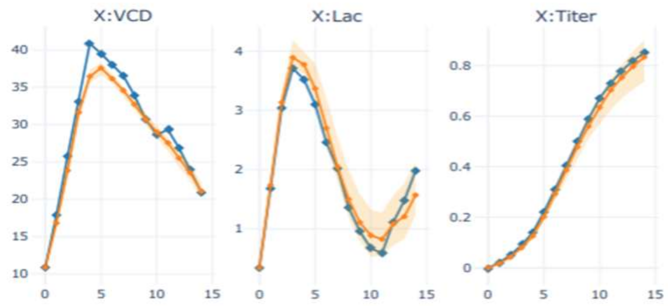
## Hybrid

- Digital representation of the production bioreactor
- Simulate bioreactor performance

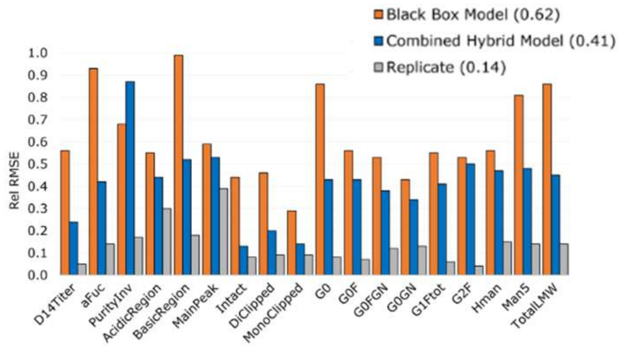


# Collaboration History

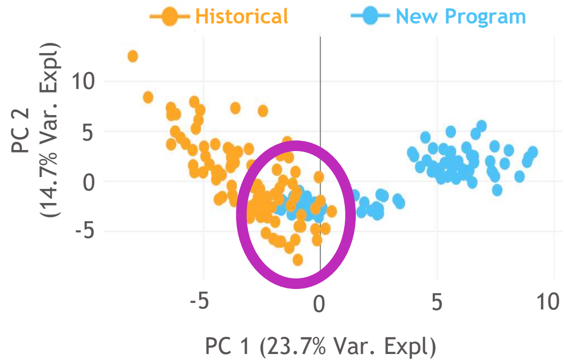
- **Project 1 – Establish Model Capabilities**
  - Goal: Evaluate the Bioreactor Propagation Model
  - Outcome: Propagation models can replicate physical systems



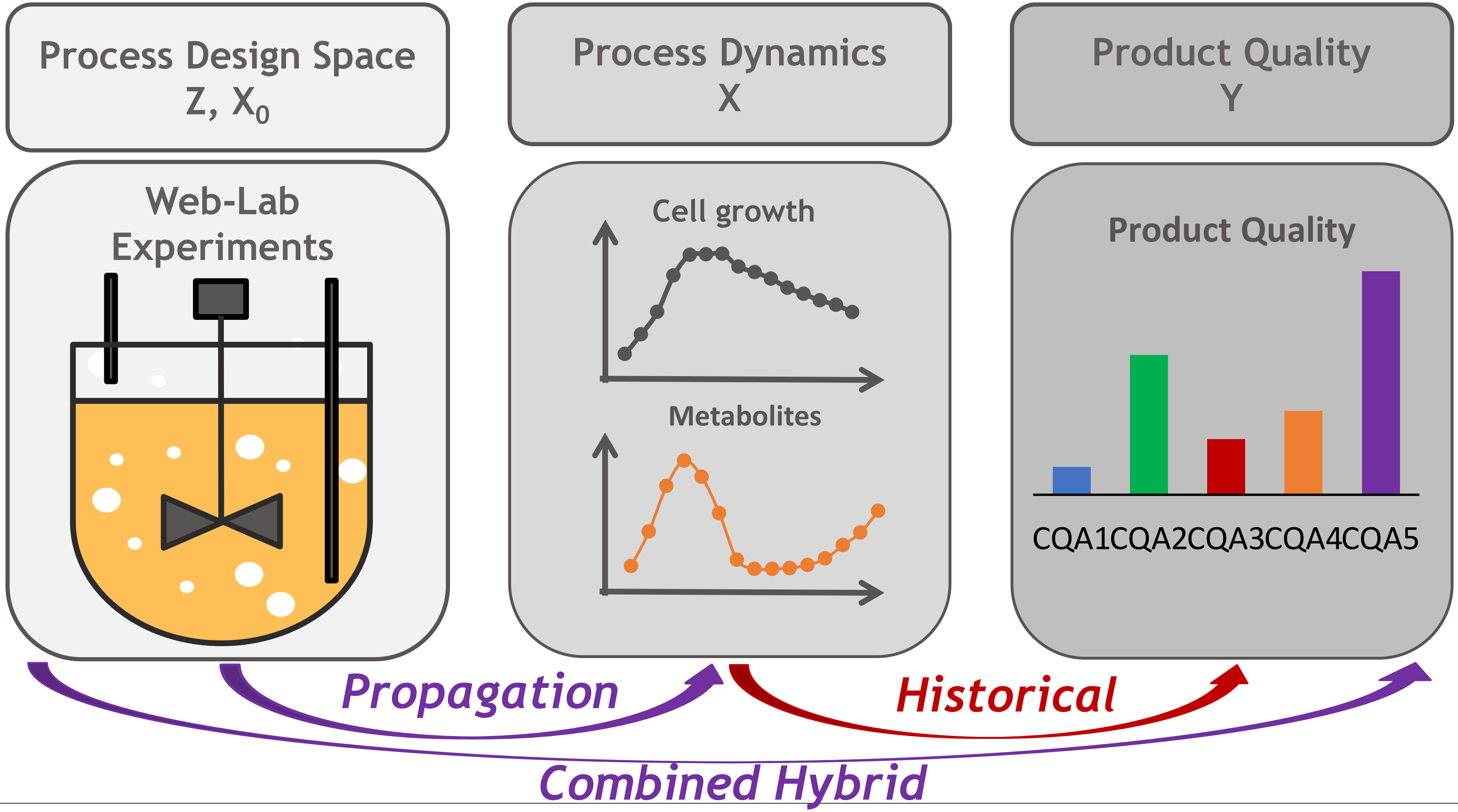

- **Project 2 - Product Quality Prediction**
  - Goal: Evaluate specific objectives including product quality prediction and compare against hybrid modeling alternatives
  - Outcome: Yes, quality can be predicted with hybrid model and are superior to black box models (focus area of this talk)




- **Project 3 – Knowledge Transfer Across Clones, Scales, and Programs**
  - Goal: Evaluate how training data from one source can be used to support another application
  - Outcome: Yes, knowledge can be transferred between applications (focus area of this talk)



# Production bioreactor hybrid models can predict process dynamics *and* product quality from initial conditions






An innovative hybrid modeling approach for simultaneous prediction of cell culture process dynamics and production quality. Polak, Huang, et al, Biotechnology Journal, Sep 2023

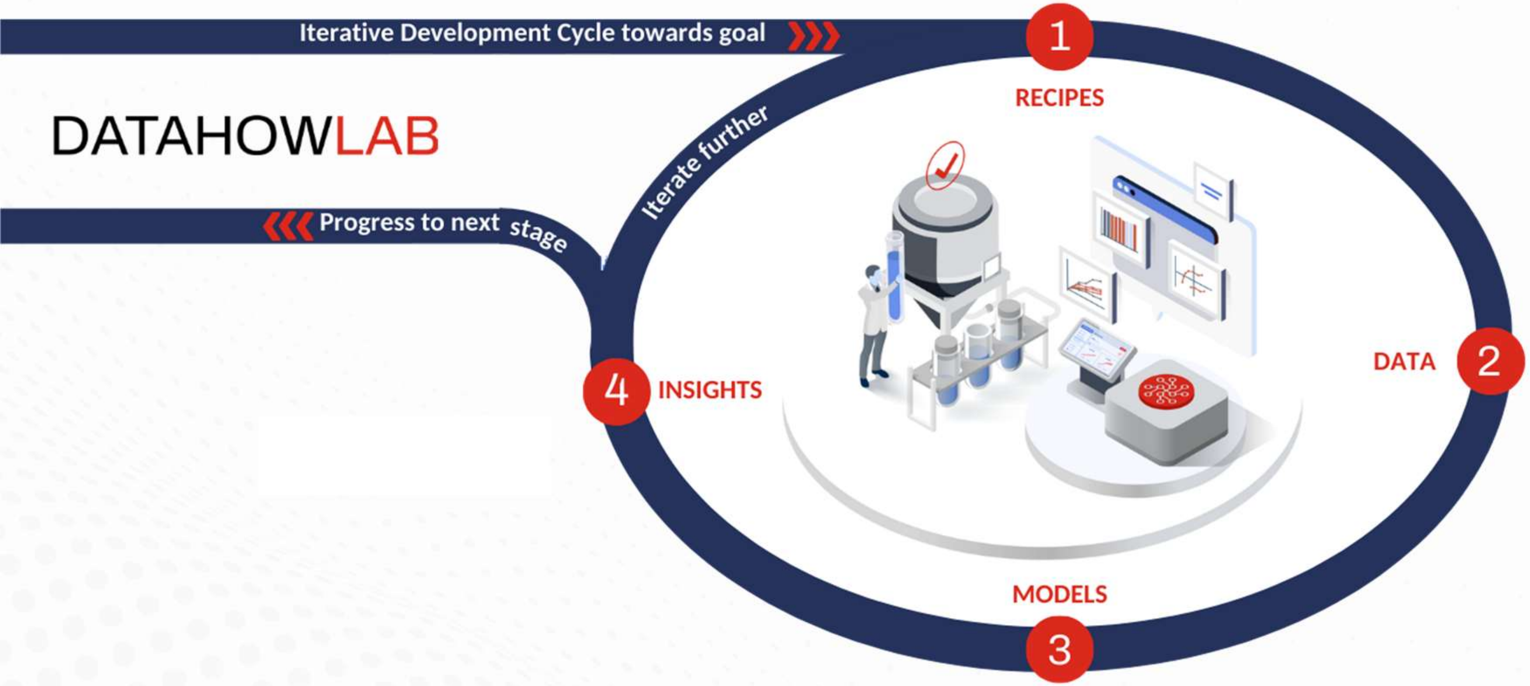
# Case Study:
# Hybrid Modelling vs Industry Standard

*An innovative hybrid modelling approach for simultaneous prediction of cell culture process dynamics and production quality.*

26 June 2025, Jakub Polak

# DataHow Methodology for Model-Based Process Design Workflow

## Hybrid Modelling vs Industry Standard




**This Publication Investigates**

- RECIPES
- DATA
- MODELS
- INSIGHTS

# Hybrid Models vs. Industry Standard

## Understanding & Predicting CQAs

DATAHOW

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.**

**The Objectives:**

1. Evaluate ability of hybrid models to predict CQAs
2. Assess how much experimental data is needed to accurately predict CQA's for each approach
3. Assess other benefits of Hybrid Models for Process characterization

# Hybrid Models vs Industry Standard

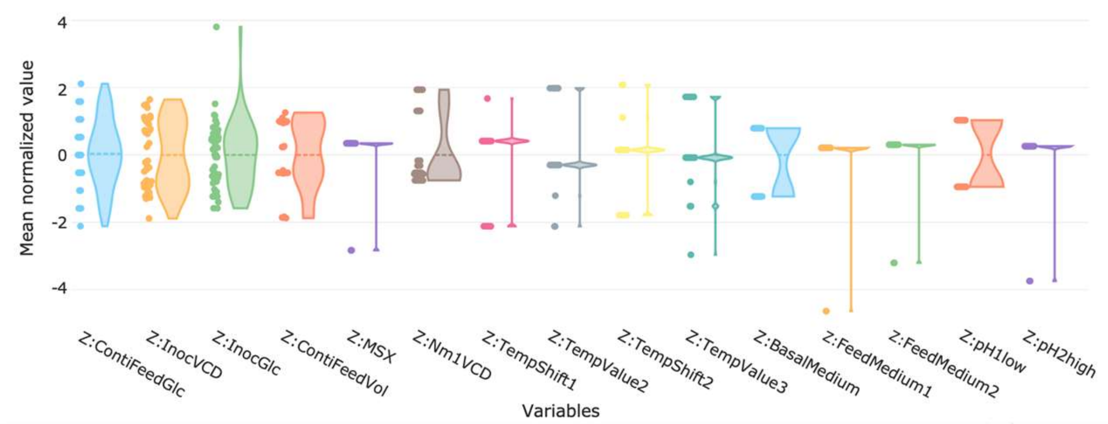
## Closer Look at Experimental Data

2 Data

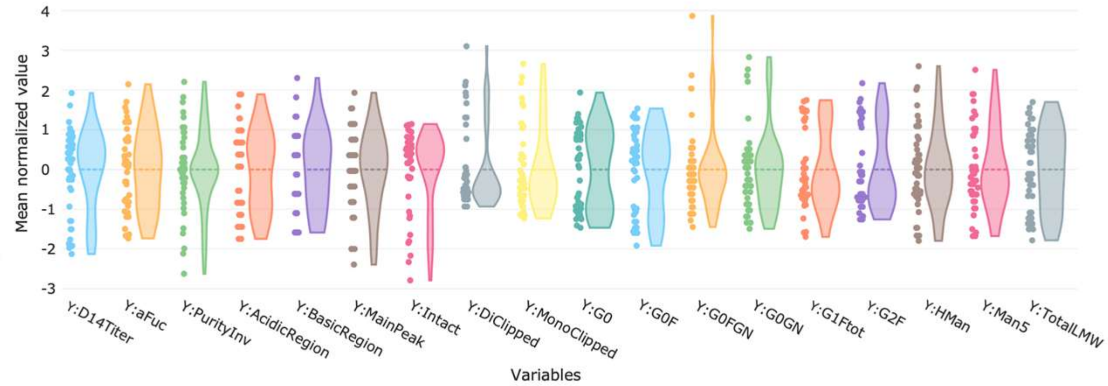
### Desgin of Experiments

| DoE | Batches | Studied parameters |
|---|---|---|
| DoE 1 | 12 × 5L | MSX levels in seed train, Temperature shift, Feed medium, Feed volume |
| DoE 2 | 12 × 5L | Inoculation density, Temperature shift, pH setting, Feed volume |
| DoE 3 | 12 × 5L | MSX levels in seed train, Basal medium, Feed medium, Temperature shifts |
| DoE 4 | 12 × 5L | Inoculation density, Basal medium, Feed medium, Feed volume |

### Inputs Variability




### Outputs Variability



# Hybrid Models vs Industry Standard

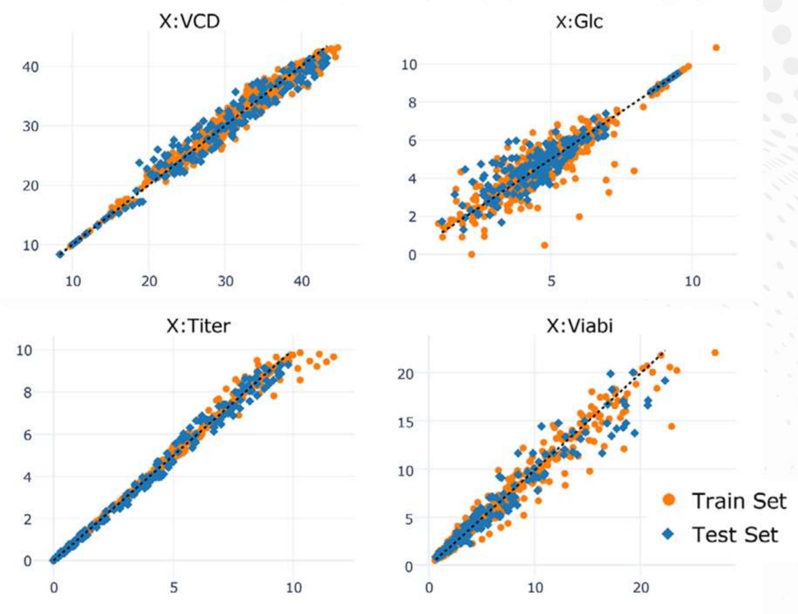
## The Performance of Hybrid Model

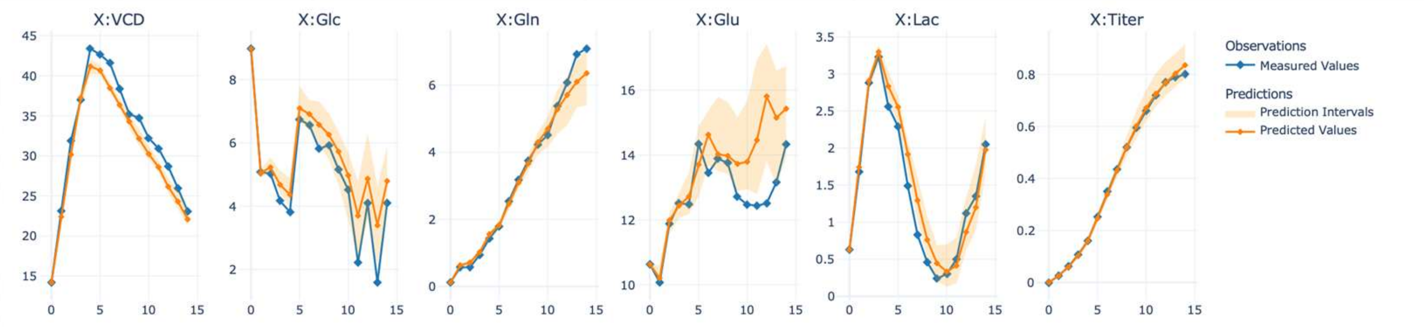
**Propagation Model**

- A discrete hybrid model with a Gaussian Processes was used to characterize the time evolution of the X variables.
- The model propagates the state of the bioreactor adjusted for mass balance equations by predicting the rate of change of the measured metabolites

$$c(t_{i+1}) \approx c(t_i) + \left( GP(s) \cdot V + u_f - c(t_i) \cdot \frac{dV}{dt} \right) \cdot \frac{t_{i+1} - t_i}{V}$$

3 Models



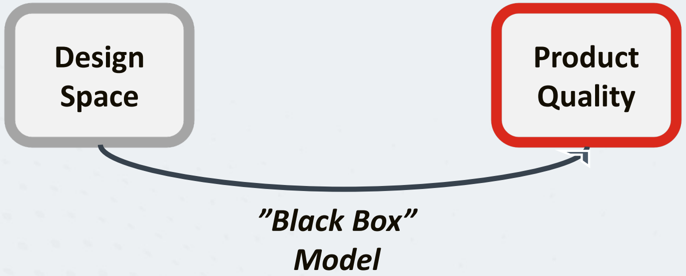
# Hybrid Models vs Industry Standard

## The difference in utilized information between the Approaches

3 Models

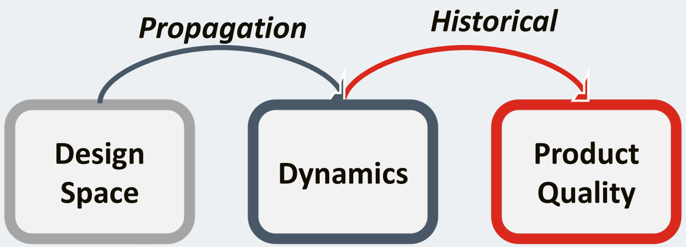
### Industry Standard Response Surface Model

- Utilizing simple linear regression approach to model directly the product quality attributes by using only the designed conditions. It doesn't consider process dynamics.




### Combined Hybrid Model

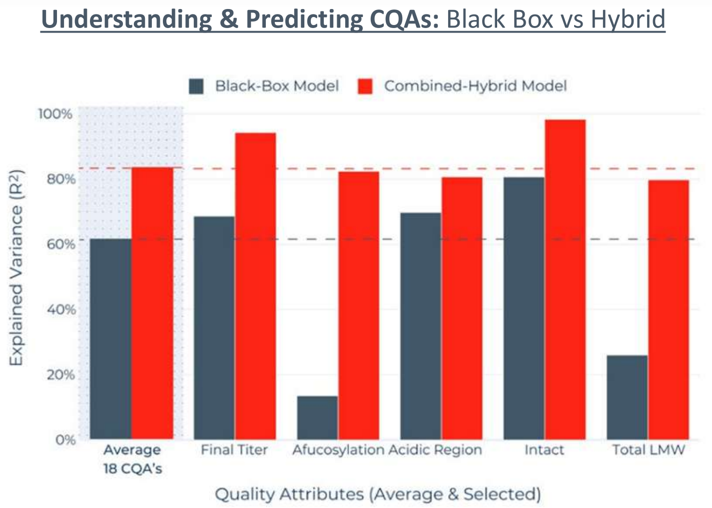
- The combination of Propagation Model and Historical Model, allows to directly link the final properties of the process and the product CQAs to the manipulated process parameters



# Hybrid Models vs Industry Standard

## The Performance of Combined Hybrid Model Compared to ”Black Box” Response Surface Model

3 **Models**





- 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)

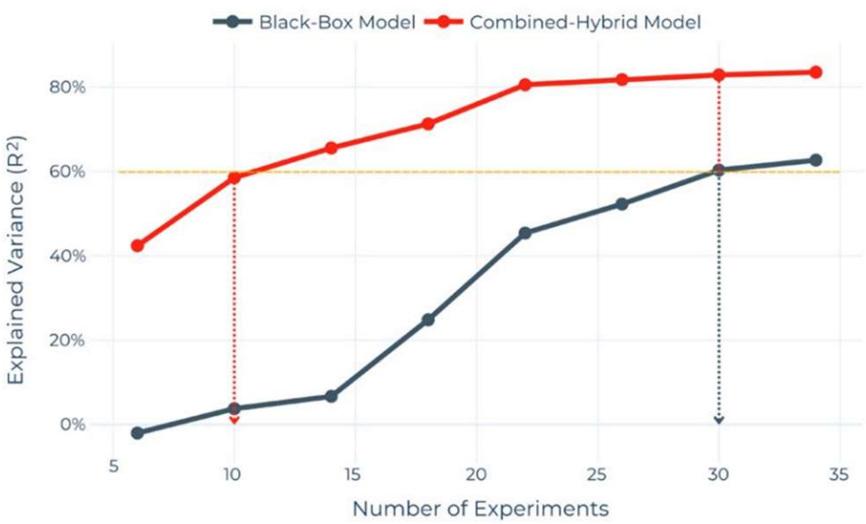


# Hybrid Models vs Industry Standard

## Model Insights with fewer performed experiments

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




- 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

3x

Fewer experiments

4 **Insights**

# Hybrid Models vs Industry Standard

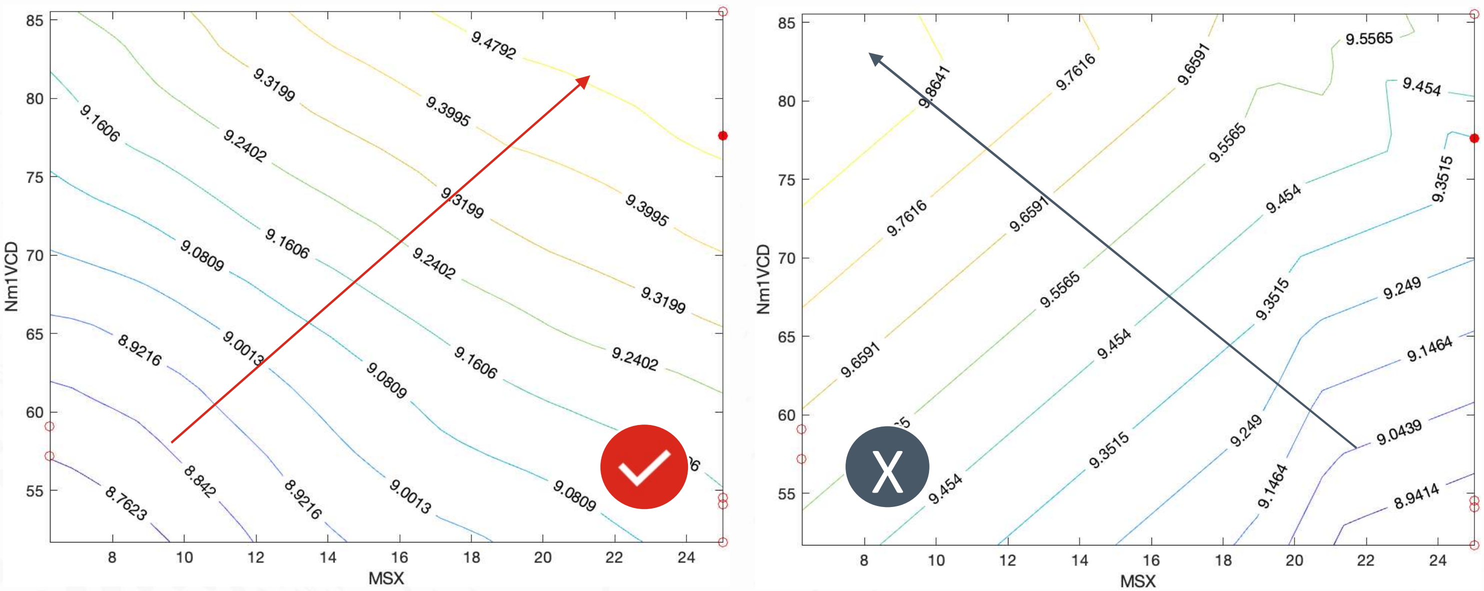
## Models and analytics supporting development objectives towards process characterization

Exploring the design space with:

- **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

Hybrid Models

Black Box Models




4 Insights

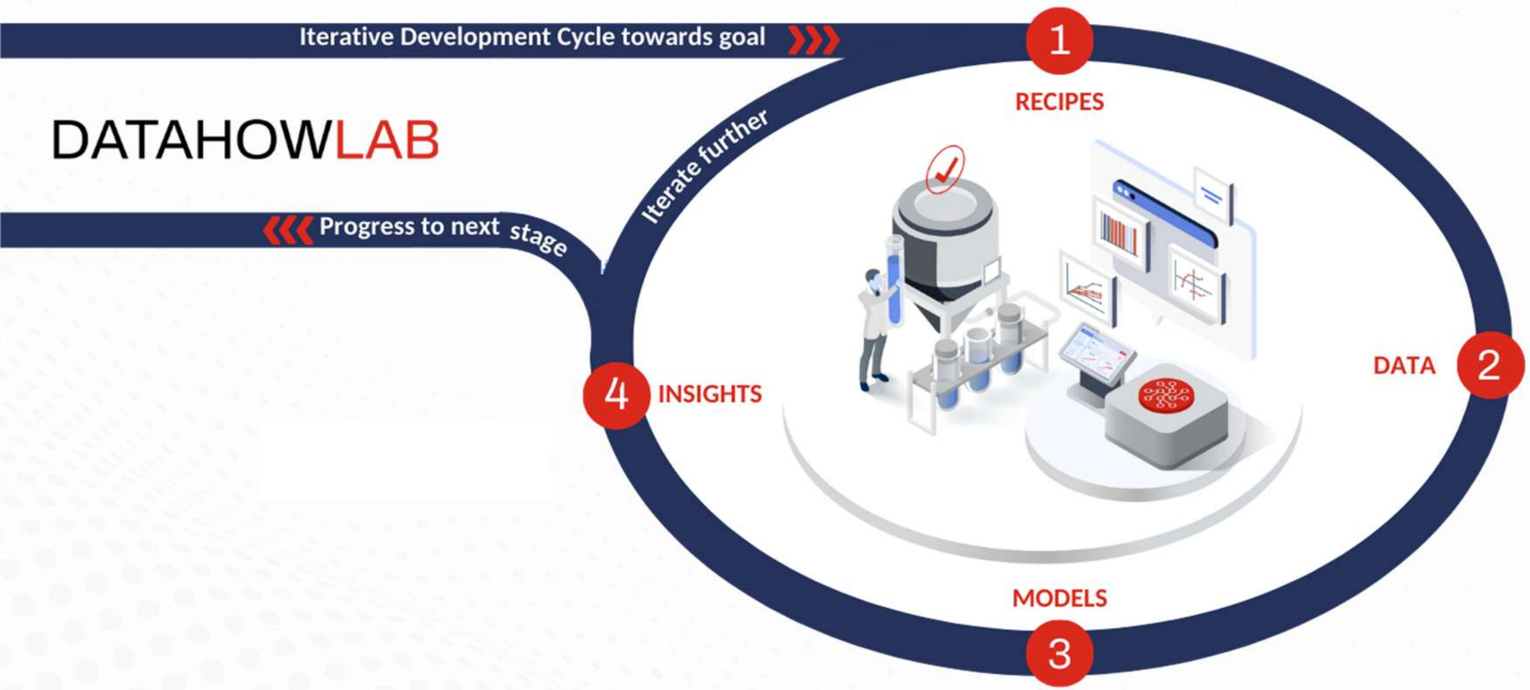
DATAHOW

# Case Study: Influence of Design of Experiments on Modelling Approaches

26 June 2025, Jakub Polak

# DataHow Methodology for Model-Based Process Design Workflow

## Influence of Design of Experiments on Modelling Approaches




**This Case Study Investigates**

- RECIPES
- DATA
- MODELS
- INSIGHTS

# Influence of Design of Experiments on Modelling

## Understanding & Predicting CQAs

Bristol Myers Squibb

**The Project:**

Compare classical **Full factorial design** (FFD) to **latin hypercube design** (LHD) for different modelling approaches such as **response surface model** (BBM) and **combined hybrid model** (CHM)

**The Challenge:**

112 (Tecan scale) experiments were designed and conducted by BMS to evaluate the impact of **10 process parameters** on 18 product CQAs**.** (with focus on Final Titer and Afucosylation)

**The Objectives:**

1. Which design allows to learn this behaviour most efficiently?
2. For which CQAs do we see the largest differences between different designs and model approaches?

# Design of Experiments Options

1 Recipes

### Full Factorial Design (FFD)

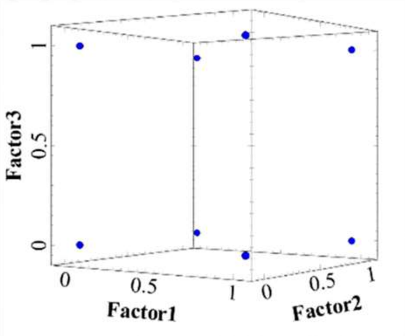
**FFD is optimal for linear models and requires a large number of experiments as a function of the number of variables (factors)**

A 2-level FFD with 9 factors has been designed. This corresponds to a resolution 4 design (all main factors are not confounded) with 32 experiments. Each design on these two levels is also a resolution 4 design.

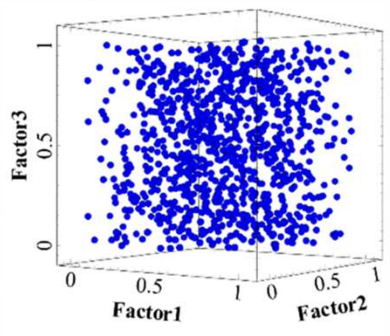
### Space Filling Design (LHD)

**LHD is optimal for ML algorithms, as it uniformly maps the space of the parameters, thus allowing ML to learn higher-order dependencies.**

Two nested LHD's were designed, each with 24 experiments using 10 factors. This will allow us to simulate the effect of a second round of experiments to refine the original model, where the entire space is re-mapped (or the LHD is augmented).



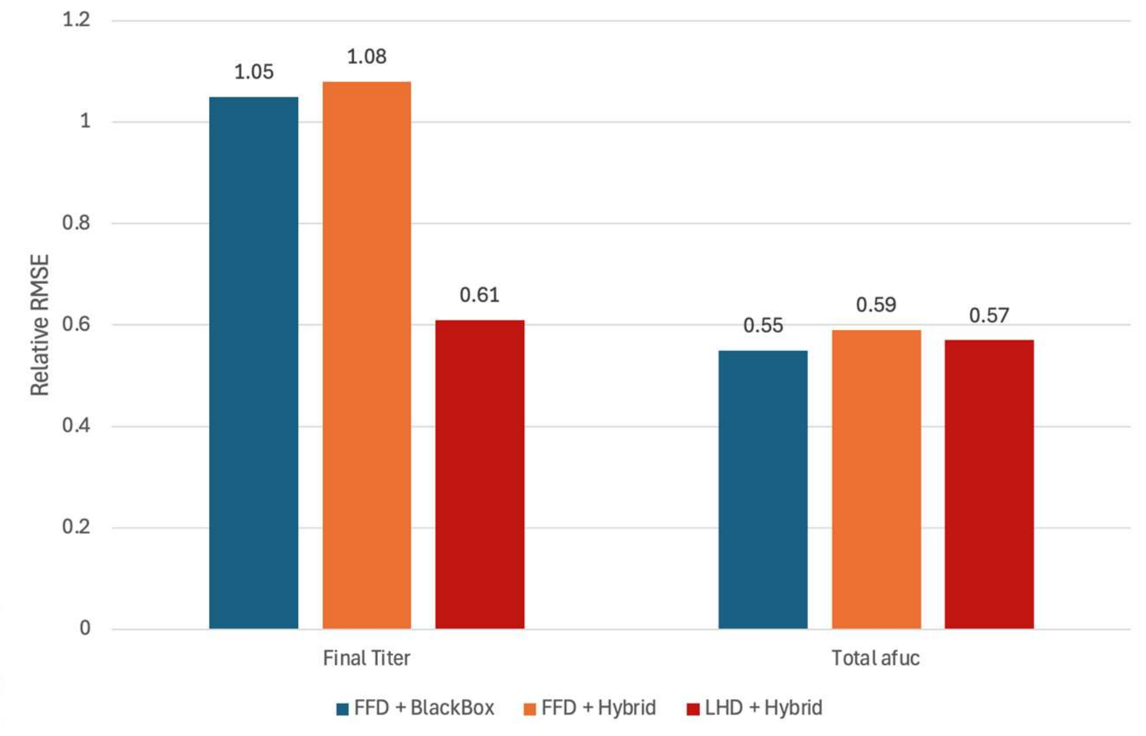

FFD
32 exp




LHD
24 exp

# Influence of Design on Modelling Approaches

**A stark comparison between different values.**



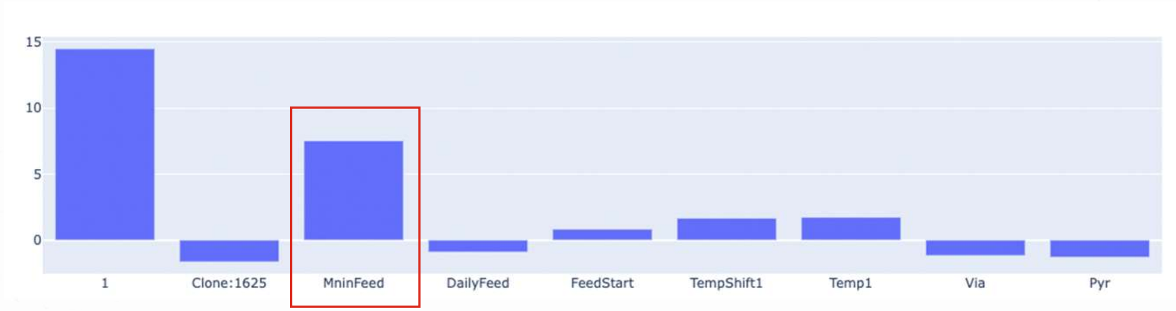

- For Final Titer, the. **Combined Hybrid model trained on LHD experiments delivers the best results** while the FFD design is not sufficient to learn the behavior
- For Total afuc, the model performances **are similar across the different models and designs** used.

3 Models

# Insights on the predictions from the model

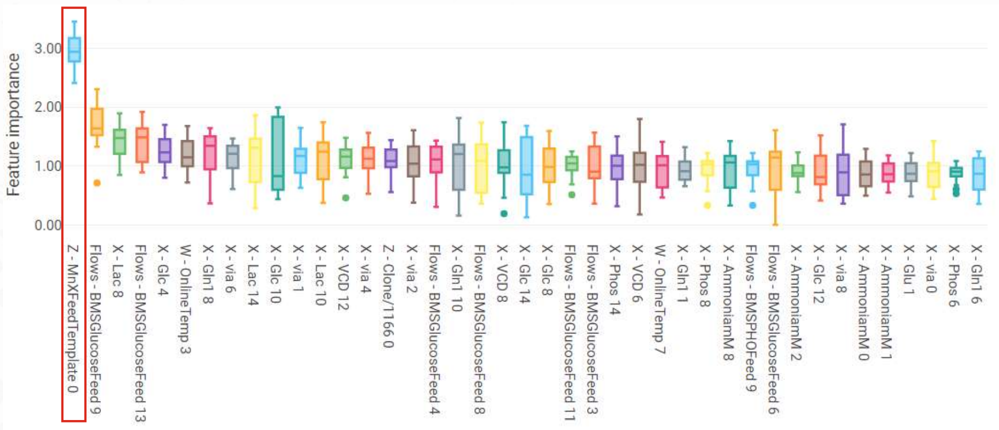
**Variable importances from the model of quality attributes. What does the model sees?**

**RSM Feature Coefficients | Total-afuc**




**Black Box Model**

**PLS Feature Importance | Total-afuc**




**Combined Hybrid**

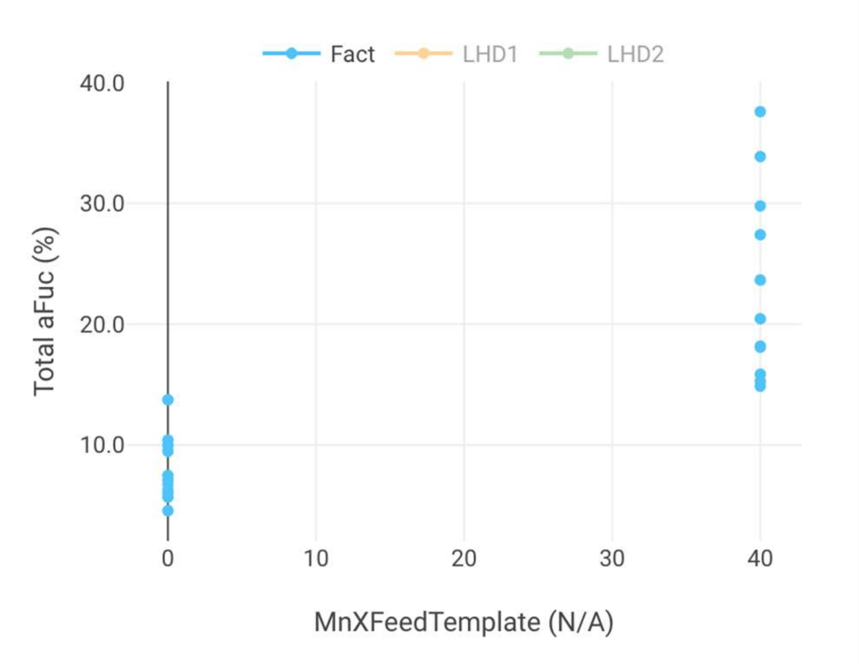
4 Insights

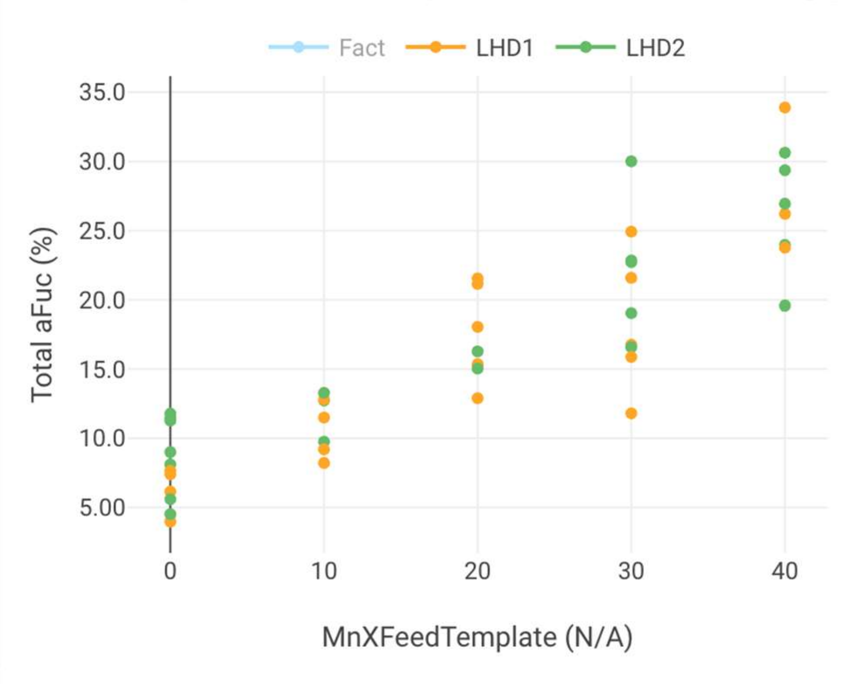
DATAHOW

# Why FFD Design and BlackBox Model is Sufficient

**When the response variable has clear linear dependence.**

2 Data

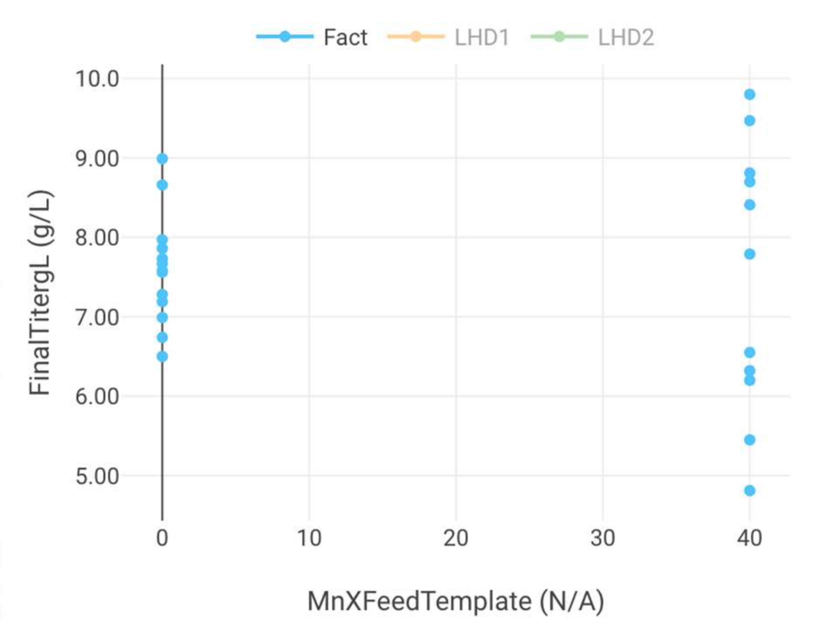


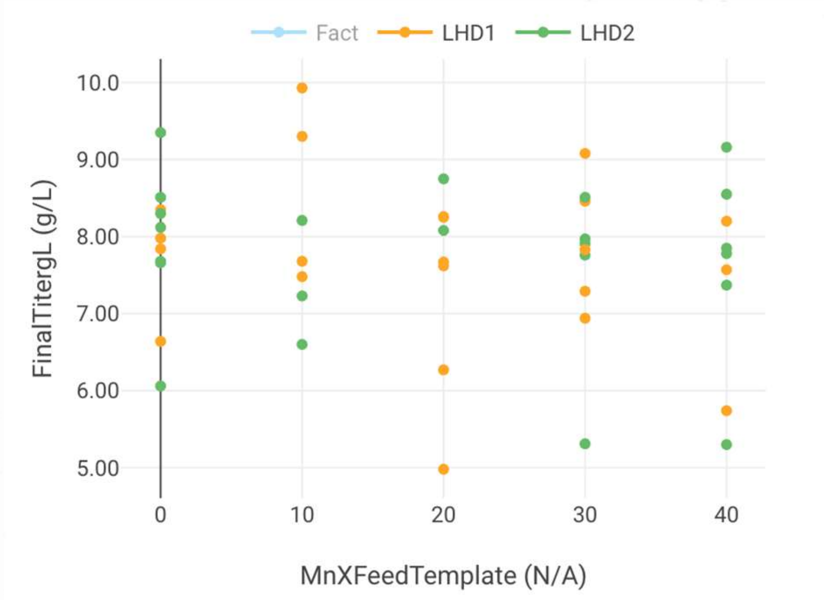



# Why FFD Design and BlackBox Model is NOT Sufficient

**When the response variable doesn't have linear dependence.**

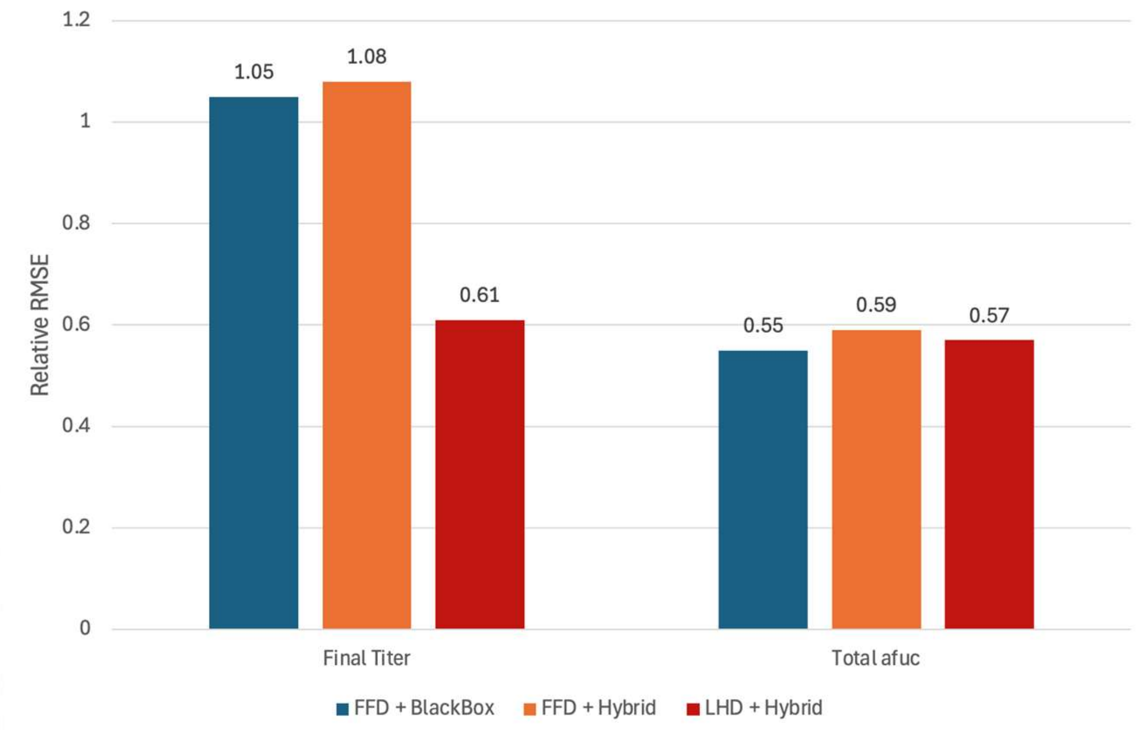
2 **Data**

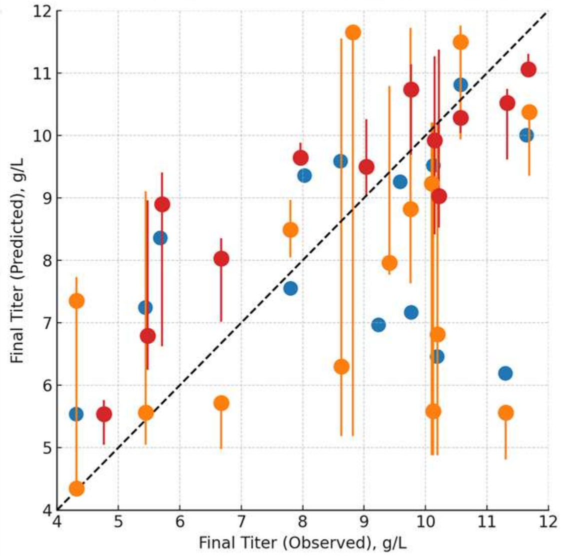


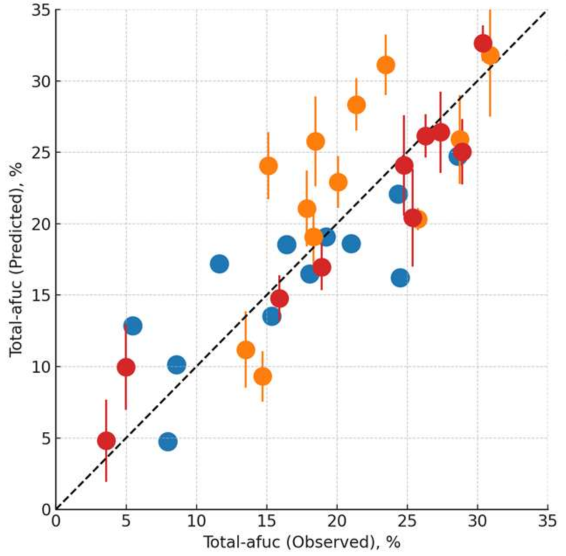



# Influence of Design on Modelling Approaches

**A closer look into model performance in parity plots.**



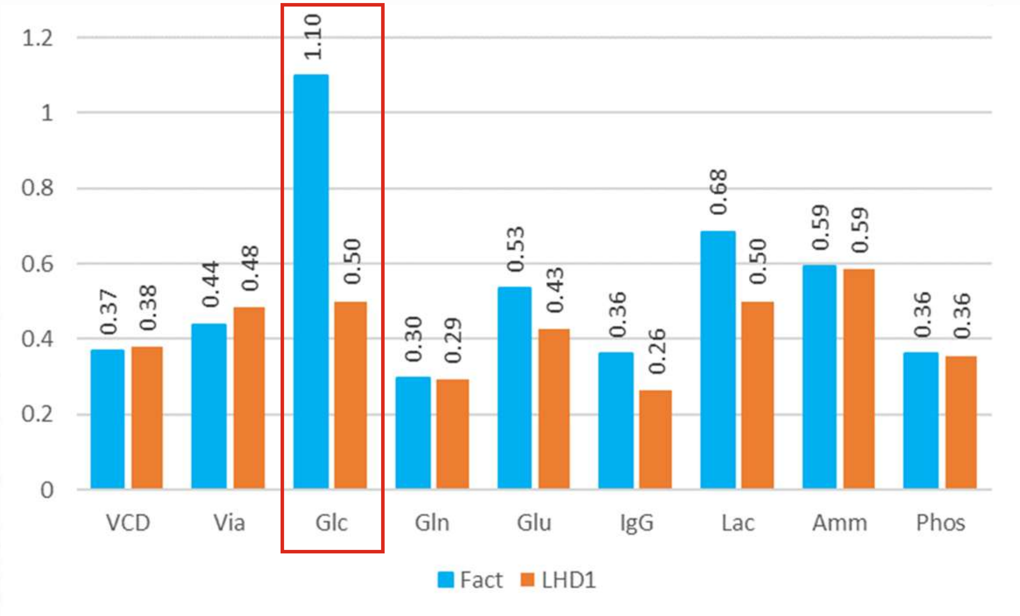




3 Models

DATAHOW

# Insights on the predictions from the model

**Variable importances from the model of quality attributes. What does the model sees?**



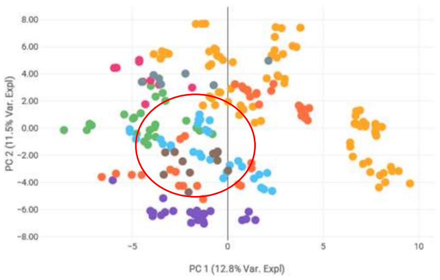

- The propagation models are integral to the Combined Hybrid Model. LHD design demonstrates superior performance.
- Specifically, for Glc, when the model is trained on the FFD, certain experiments are inaccurately predicted, including the VCD.

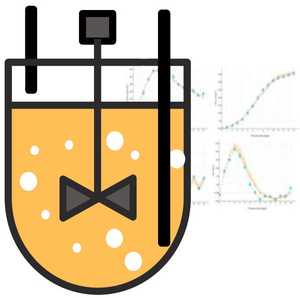
4 **Insights**

# North Star Vision for Hybrid Modeling in Cell Culture

Model-First
Commercial PD

**Biologics Development**



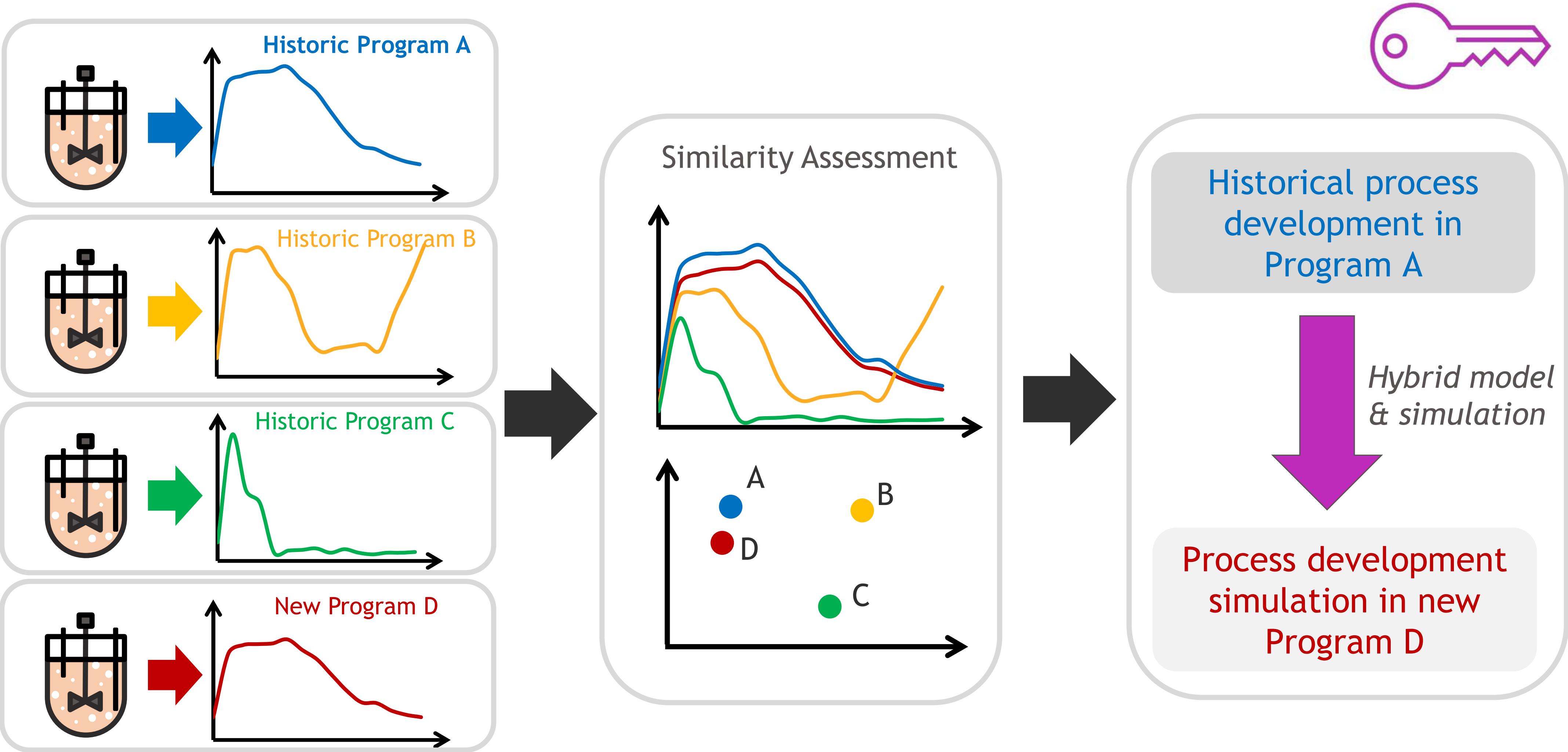



**Goal**: Reduce timelines of highly productive
and robust process development

# Case Study 1:
# Model-First Approach to Process Development
## (Knowledge Transfer Across Programs)

Case Study Credit: Zhuangrong Huang

# Knowledge transfer is key to enable model-first in process development

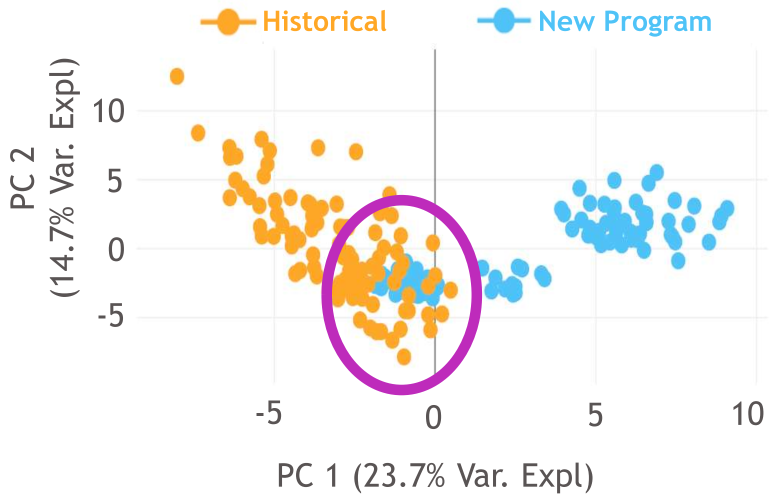





Hutter, von Stosch, et al. *"Knowledge transfer across cell lines using hybrid Gaussian process models with entity embedding vectors."* Biotechnol & Bioeng. 2021. DOE: 10.1002/bit.27907

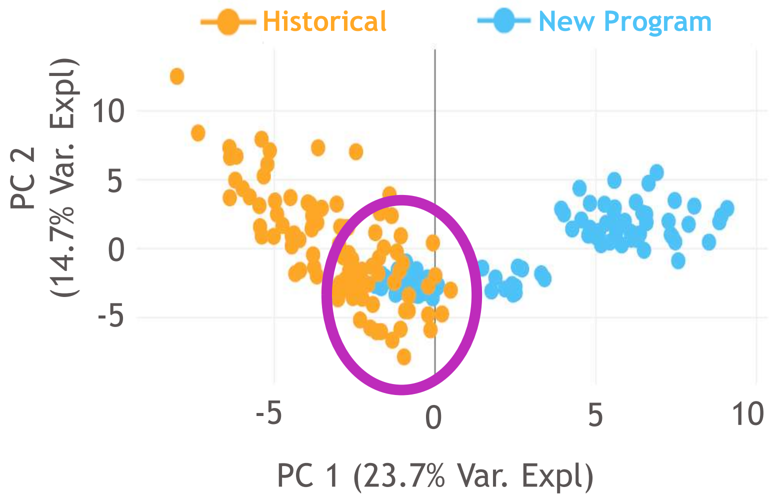
# Knowledge transfer is key to enable model-first in process development

Similarity Assessment

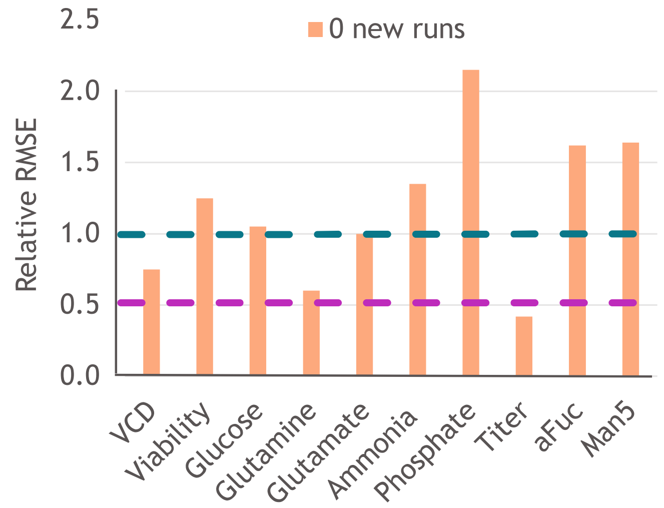


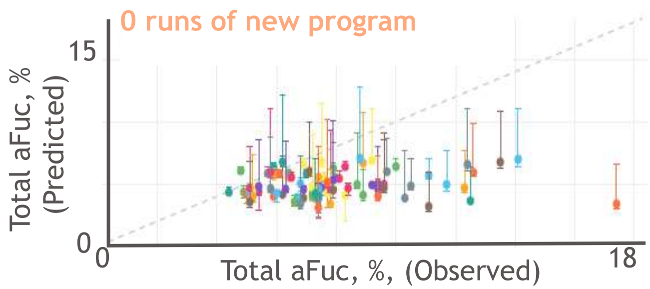
# Knowledge transfer is key to enable model-first in process development

## Similarity Assessment



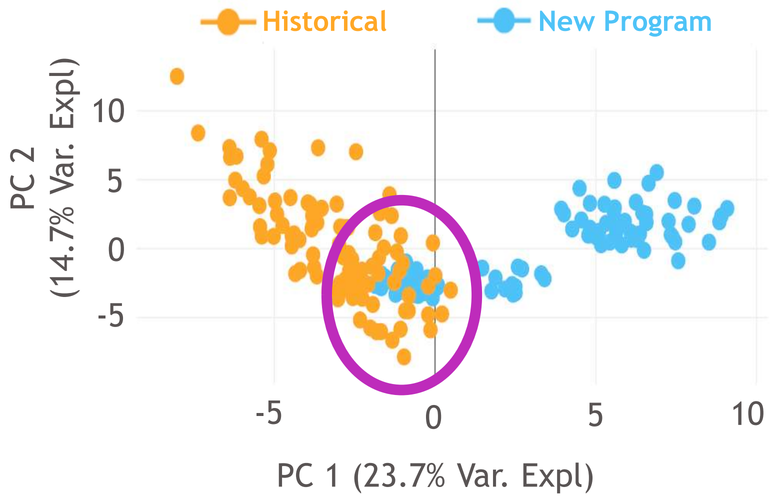

## Model Error



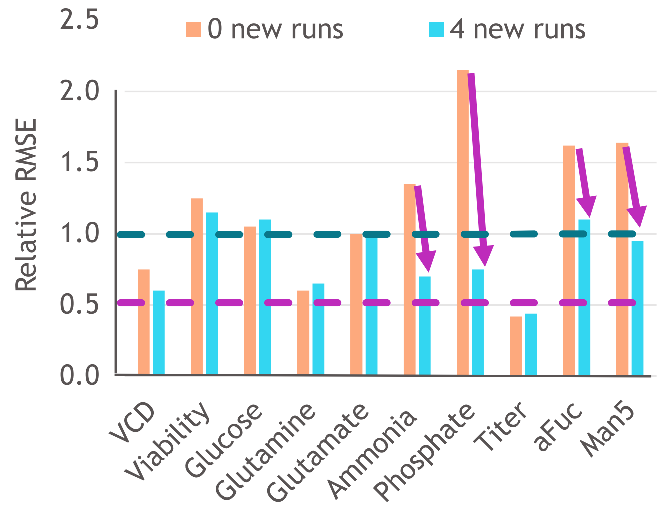



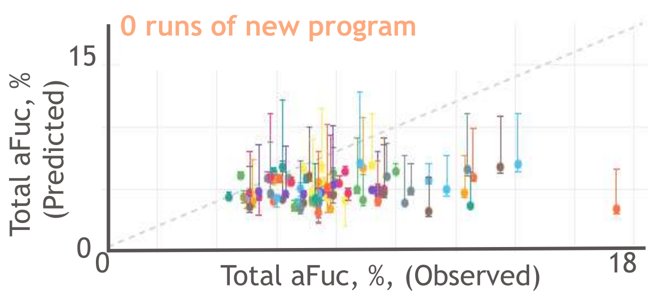
# Knowledge transfer is key to enable model-first in process development

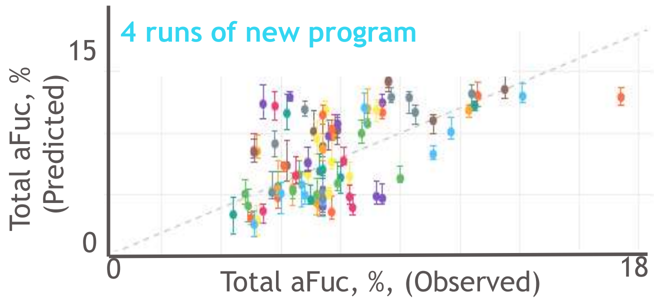
## Similarity Assessment




## Model Error

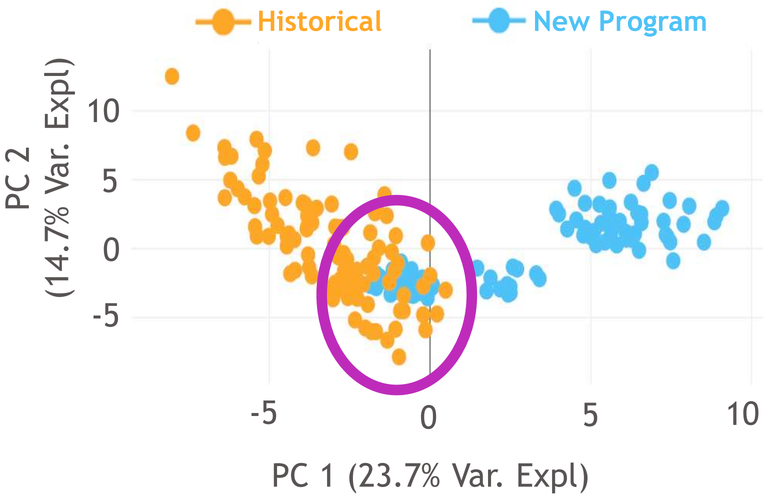


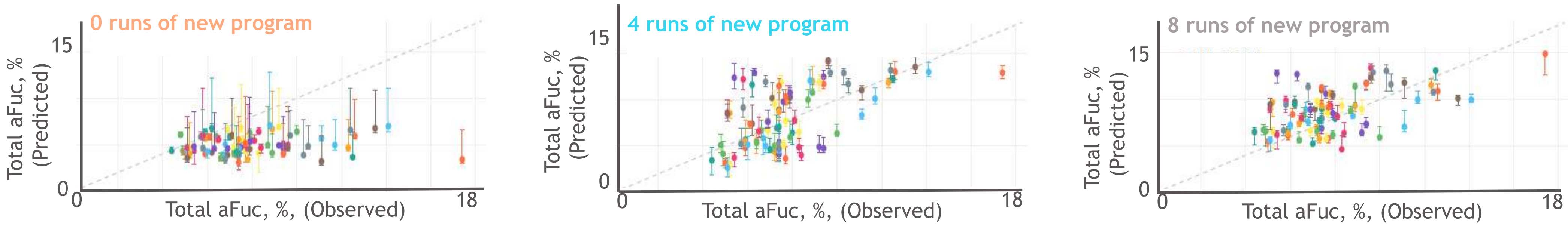



# Knowledge transfer is key to enable model-first in process development

## Similarity Assessment



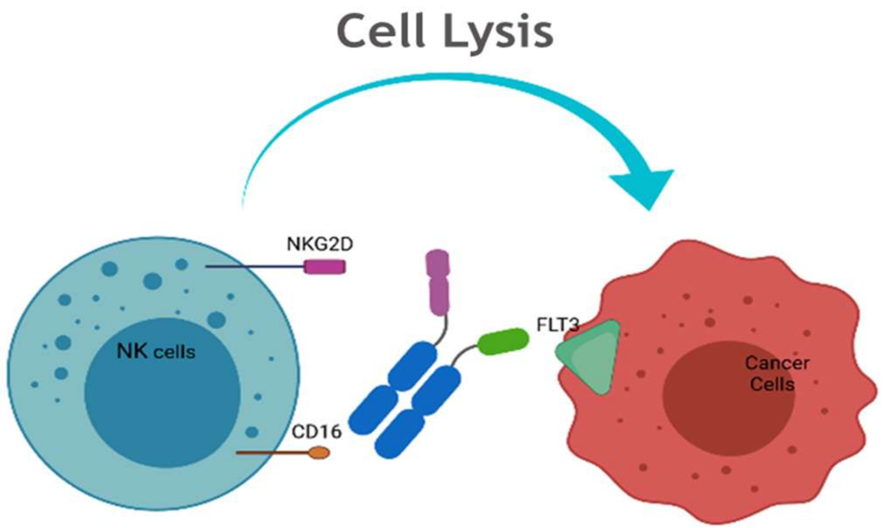

## Model Error

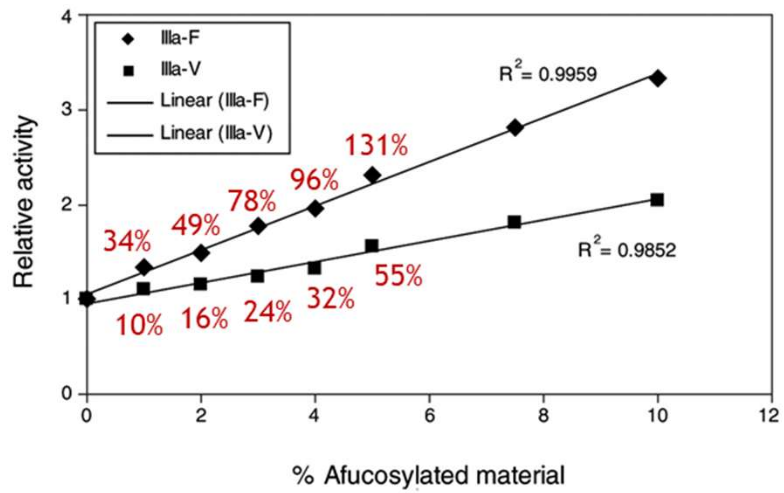



# Case Study 2:
# Recipe-Driven Process Optimization of aFucosylation
## (Knowledge Transfer Across Scales)

Case Study Credit: Khandaker Siddiquee and Yikun Huang

# Why is aFucosylation important?




Chung, Shan, et al. "*Quantitative evaluation of fucose reducing effects in a humanized antibody on Fcγ receptor binding and antibody-dependent cell-mediated cytotoxicity activities.*" MAbs. Vol. 4. No. 3. Taylor & Francis, 2012.

- aFucosylation affects CD16a/FcγRIII binding (efficacy) and antibody dependent cell cytotoxicity (ADCC) (safety)
- Important to control aFucosylation levels

# aFucosylation hybrid modeling strategy

**Dataset Build-up**

- DOE (Tecan (HTS))
- Lab-scale bioreactor (Ambr250)
- Project specific process development

**aFucosylation hybrid model development**

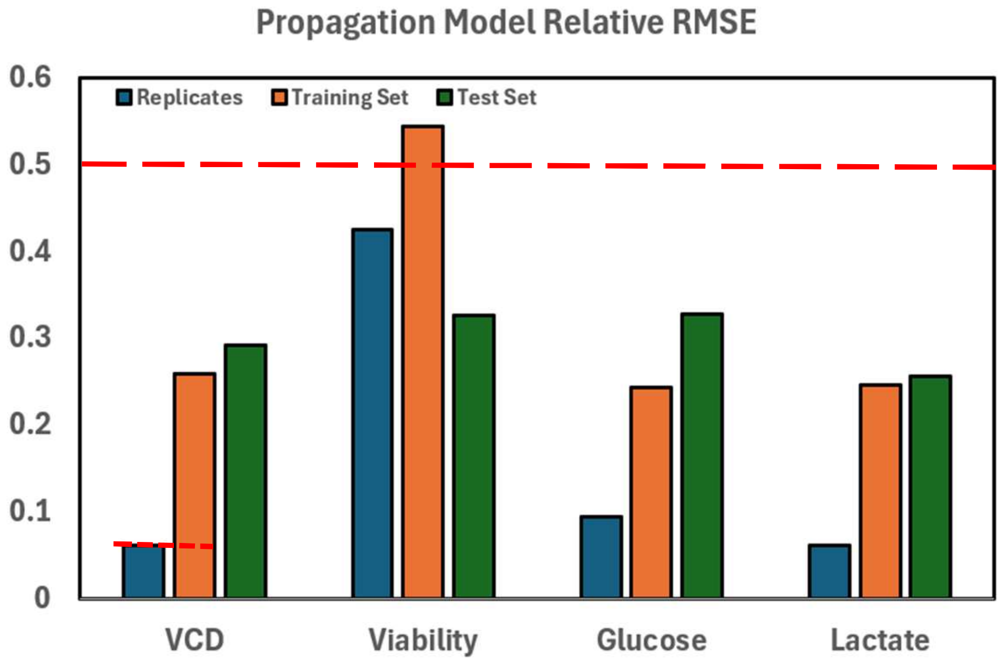
- Model training
- Model evaluation

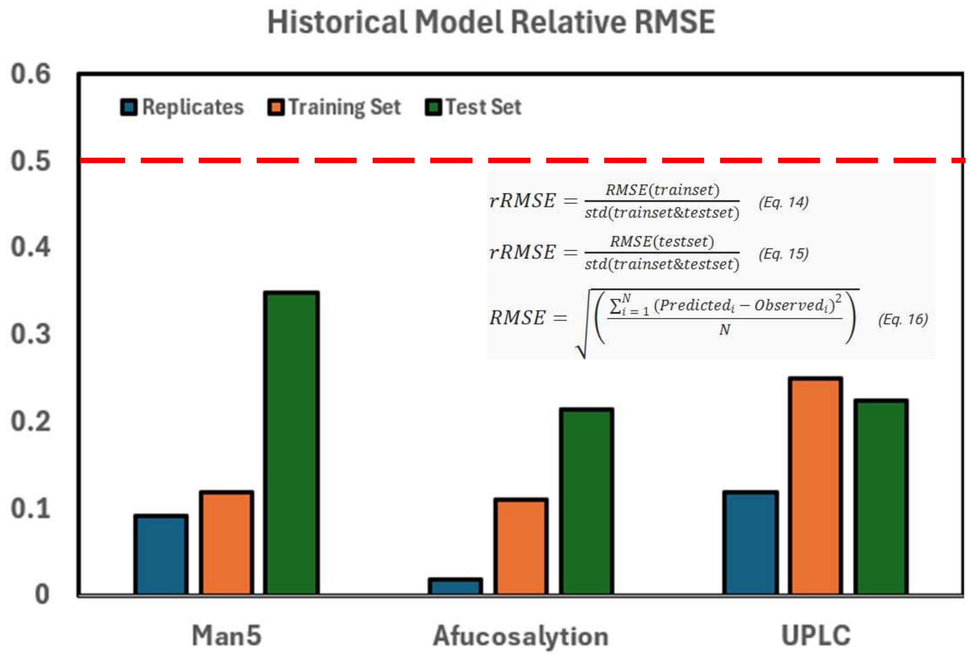
**Simulation & Robustness analysis**

- Predict aFucosylation across programs and scales
- Sensitivity analysis

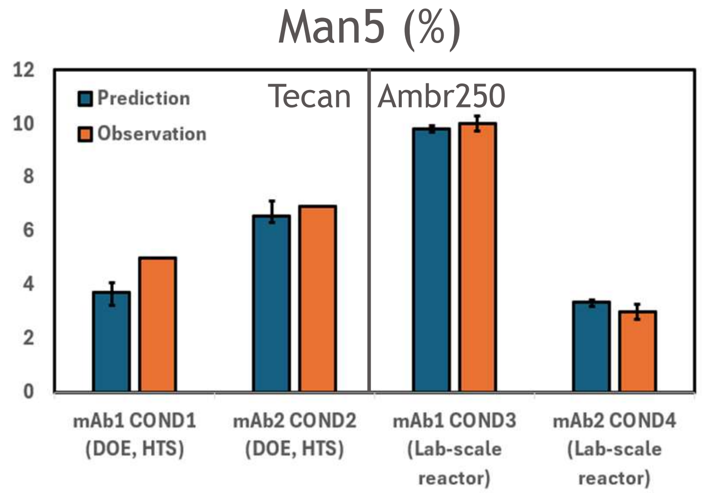
# Propagation and historical models show good performance

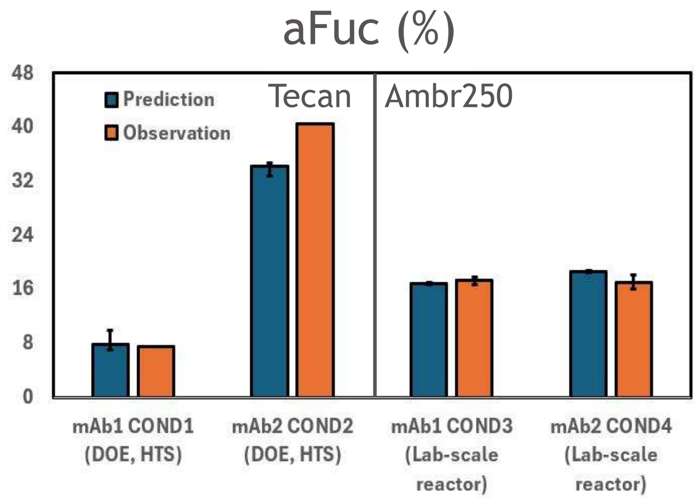
- Training Set: 63 conditions (54 train, 9 test) from Tecan 50-mL conical high-throughput system (HTS)
- Model outcome:
  - **Good performance** with relative RMSE (rRMSE) < 0.5
  - Except with viability (error prone training data)
  - **No overfitting** (training set rRMSEs are greater than replicate rRMSE)



# Strong correlation observed between predictions and experimental data for individual conditions for Tecan (HTS) and Ambr250 (Lab) models



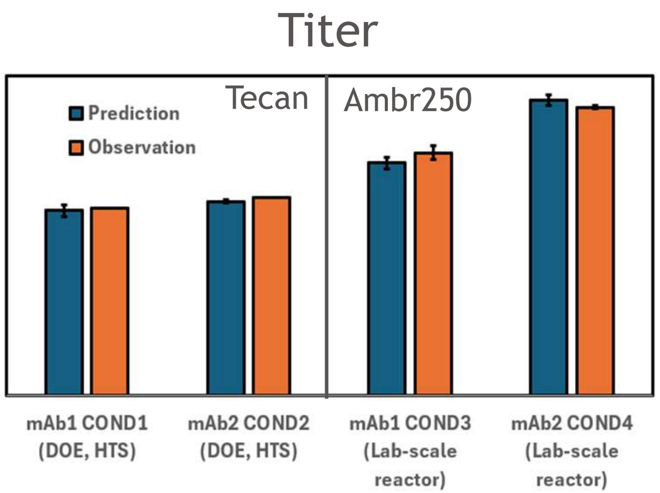

2 PROGRAMS, 1 MODEL PER SCALE

**Model Training 1**

- Tecan: 54 conditions in train set
- AMBR250: 40 conditions in train set

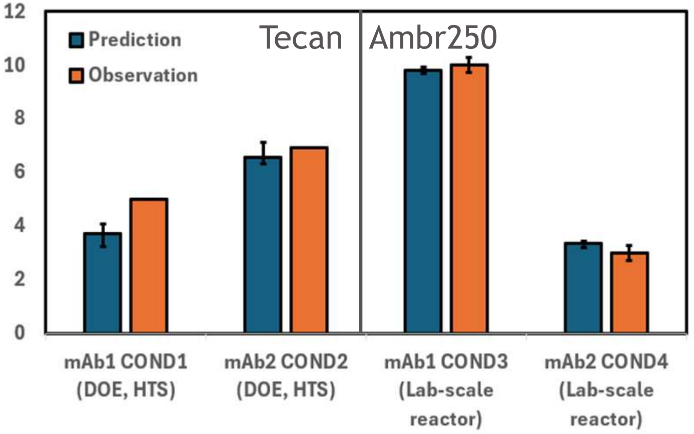
**Takeaway 1**

- Accurate prediction of titer and CQAs across programs within scale

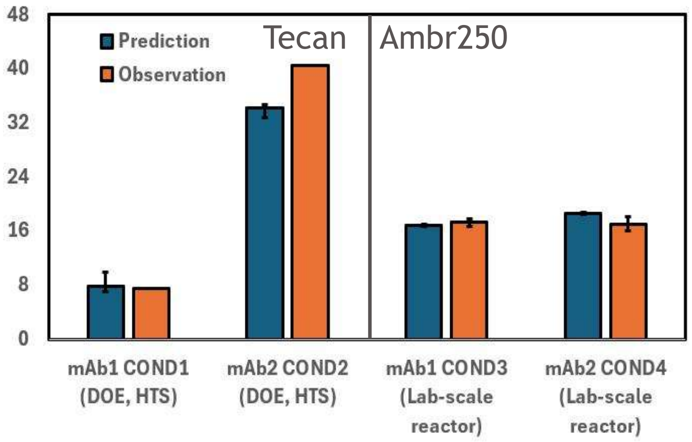


# Strong correlation observed between predictions and experimental data for individual conditions for Tecan (HTS) and Ambr250 (Lab) models

Man5 (%)




aFuc (%)




2 PROGRAMS, 1 MODEL PER SCALE

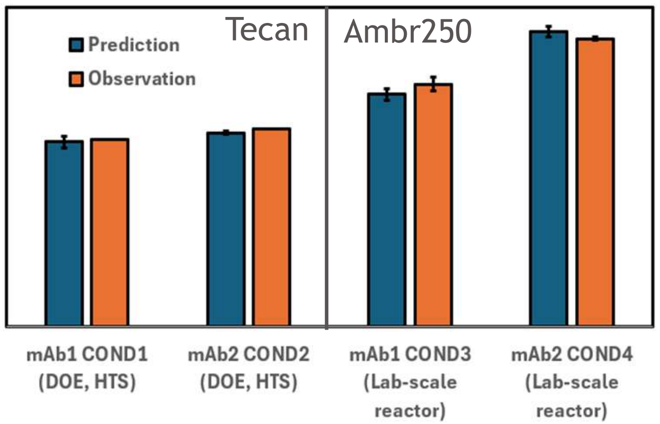
**Model Training 1**

- Tecan: 54 conditions in train set
- AMBR250: 40 conditions in train set

**Takeaway 1**

- Accurate prediction of titer and CQAs across programs within scale

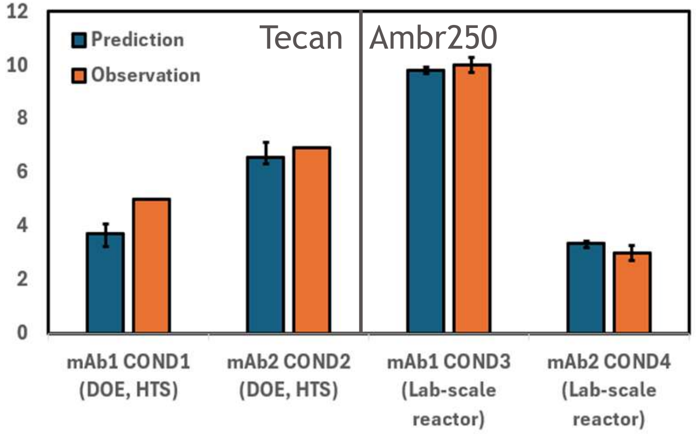
Titer



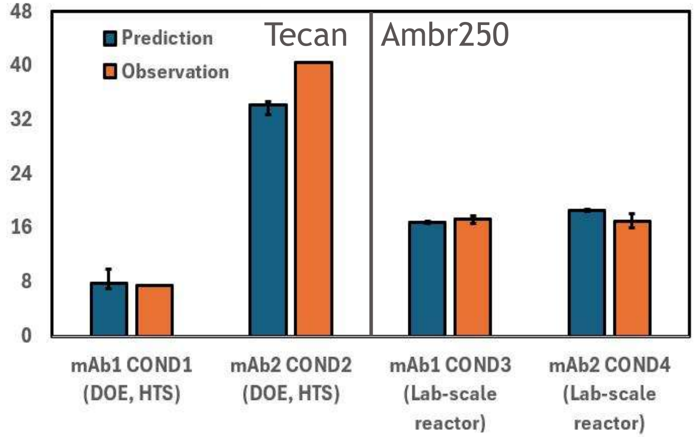

What about across scales?

# Strong correlation observed between predictions and experimental data for individual conditions for Tecan (HTS) and Ambr250 (Lab) models

Man5 (%)




aFuc (%)




## 2 PROGRAMS, 1 MODEL PER SCALE

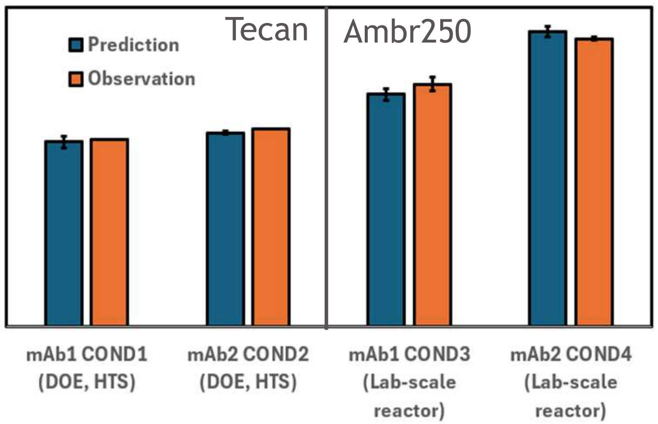
**Model Training 1**

- Tecan: 54 conditions in train set
- AMBR250: 40 conditions in train set

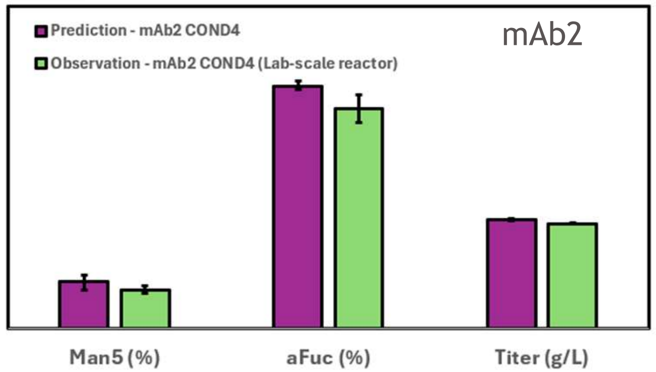
**Takeaway 1**

- Accurate prediction of titer and CQAs across programs within scale

Titer




Prediction Across Scales




## 2 SCALES, 1 MODEL PER PROGRAM

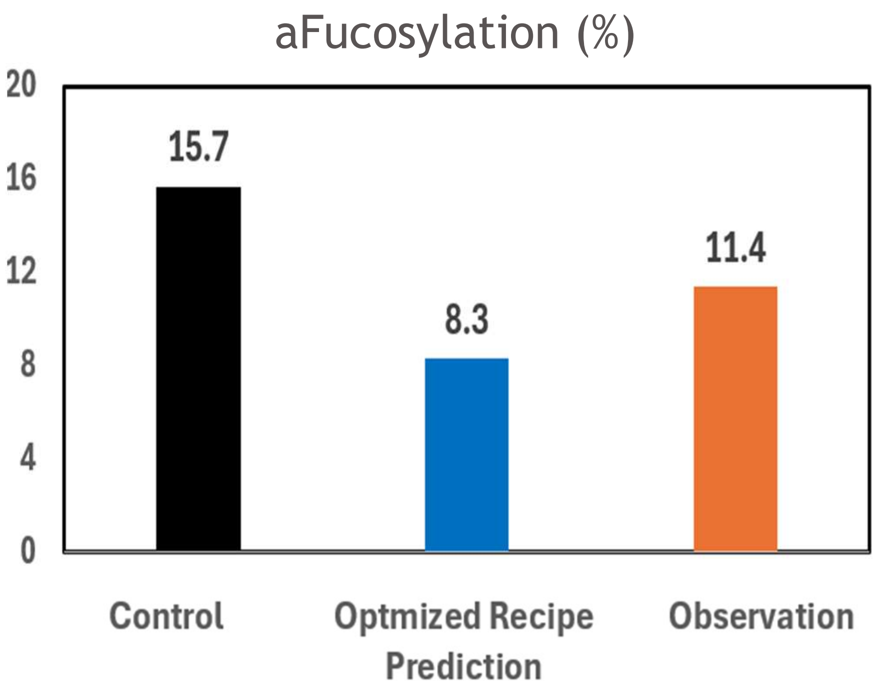
**Model Training 2**

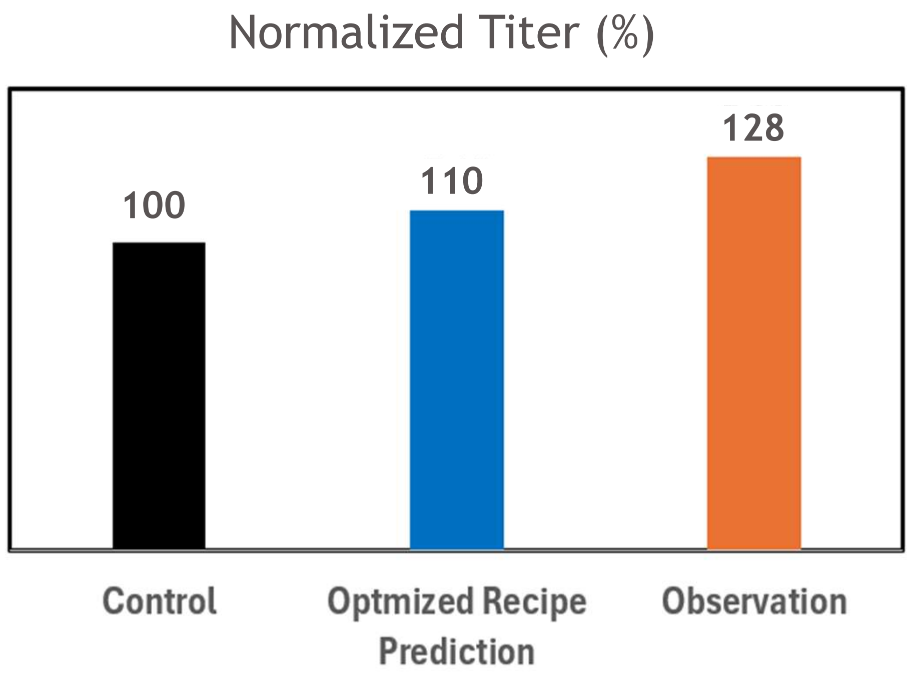
- 28 conditions from Tecan and 4 conditions from AMBR250 in train set

**Takeaway 2**

- Accurate prediction of titer and CQAs across scales

# Recipe driven optimization of aFucosylation and titer




**Takeaway**

✓ Optimization recipes predicted reduced aFucosylation while increasing titer

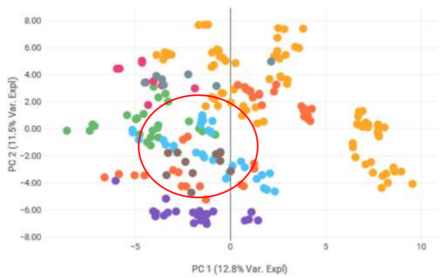
✓ Experimental results align directionally with recipe prediction

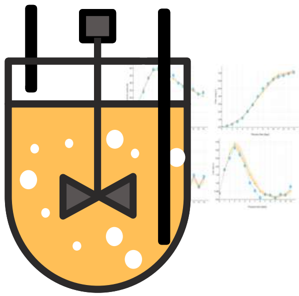
# North Star Vision for Hybrid Modeling in Cell Culture

Improved FIH (Single phase development) → Model-First Commercial PD → Model-Aided Process Characterization → GMP Model Predictive Control

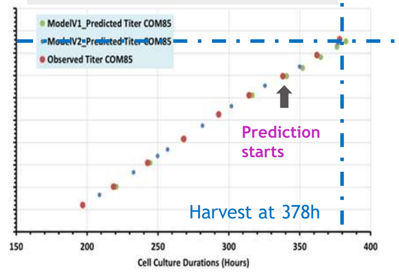
Biologics Development

MS&T

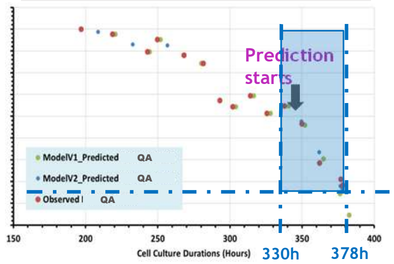


Predicted Harvest Titer




Predicted Harvest Quality




**Goal**: Reduce timelines of highly productive and robust process development

**Result**: 6-week reduction in first case

**Goal**: Predict **optimal timing** of process shifts

Maximize **Titer** and **Quality**

# Acknowledgements


BMS

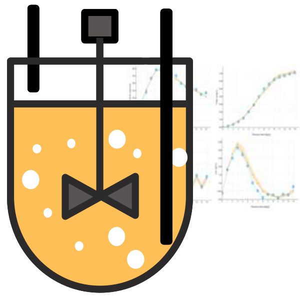
- Anurag Khetan
- Michael Borys
- Zhuangrong Huang (former BMS)
- Khandaker Siddiquee
- Yikun Huang
- Steve Traylor

DataHow

- Michael Sokolov
- Moritz von Stosch
- Alessandro Butté
- Tommaso Sardelli
- Elena Lietta
- Gui Ramos


# Thank you

Biologics Development

**Goal**: Reduce timelines of highly productive and robust process development

**Result**: 6-week reduction in first case

Bristol Myers Squibb™