DATAHOW

Understand & Predict Product Quality with 3x Fewer Experiments using Hybrid Modeling Technology

In process development, scientists must decode the complex relationships between critical process parameters (CPPs) and critical quality attributes (CQAs) using advanced analytics and modeling. However, are the commonly used analytical approaches optimal? A study performed with Bristol Myers Squibb compared industry "black box" models against DataHow's hybrid.

Predicting CQA's from process data: Hybrid Model vs. Industry Black Box

Forty-eight 5L scale experiments were designed and executed by BMS to assess the impact of 12 CPPs on 18 CQAs.

The resulting dataset was first analyzed using a multiple linear regression (MLR) approach. Such statistical, data-driven models are widely used across the industry and are commonly referred to as "black-box" models.

The same dataset was subsequently evaluated using DataHow's hybrid models within DataHowLab.

Assessing the Impact DataHowLab's Hybrid Models to understand complex processes

In a direct head-to-head comparison, the performance of DataHowLab's hybrid models was contrasted against industry "black-box" models:

OBJ. 1: Evaluate ability of hybrid models to predict CQAs
How do hybrid models compare with industry "black-box" models in their ability to understand and predict the relationships between CPPs and CQAs?

OBJ. 2: Assess how much experimental data is needed to accurately predict CQA's
Determine how many experiments are needed for each modeling approach to be able to reliably predict CQAs.

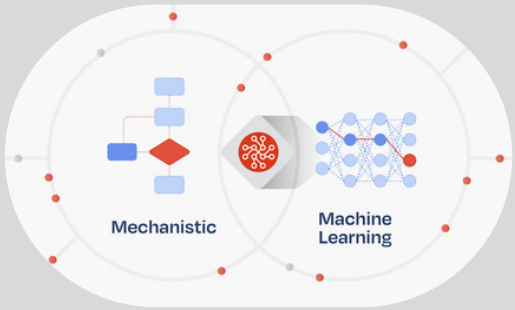
OBJ. 3: Assess other benefits of Hybrid Models
Determine whether hybrid models offer any additional benefits over "black-box" models.

What is a Hybrid Model

DataHowLabs hybrid models are a balance of structured knowledge and flexible learning that are perfectly adapted to the complex, yet data-scarce environment of bioprocessing.

The models have a **mechanistic backbone** which provide:

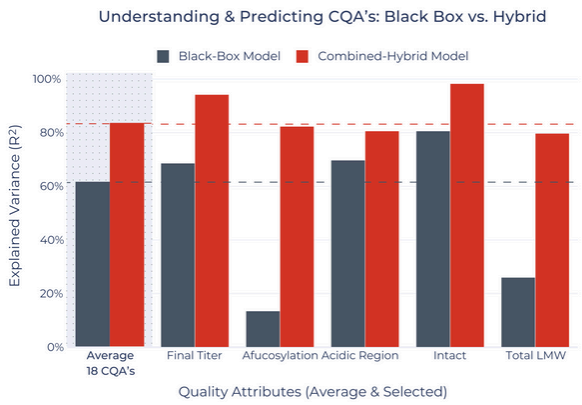
- **Interpretability**, as they describe known process dynamics (e.g. mass balances).
- **Efficiency**, by incorporating prior knowledge they require less data to describe process behaviour.
- **Extrapolation**, allowing you to predict outside of the observed range.



The **machine-learning** component provides data-driven insight where we lack scientific understanding, providing:

- **Accuracy**, learning from data to understand dynamics which are unknown, or complex.
- **Flexibility**, adapting to the specific nature of the process by generating insights from data.
- **Acceleration**, providing deeper insights making extensive experimentation redundant.

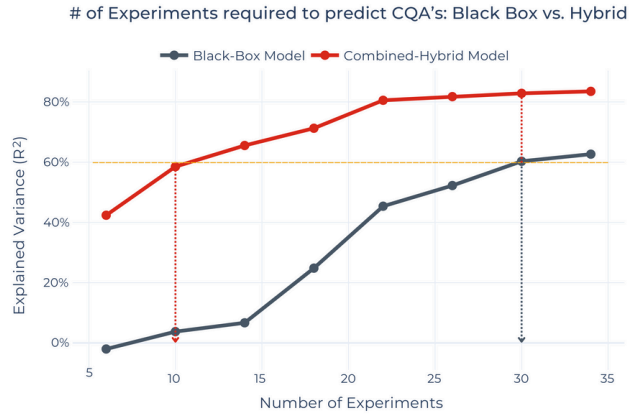
OBJ. 1 - Evaluate the ability of hybrid models to predict CQAs



Hybrid Models are able to more accurately predict CQAs than the industry "black-box" models

- On average, hybrid models improved CQA prediction by 35% (Explained variability of 83% vs 62% for black-box models)
- 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)

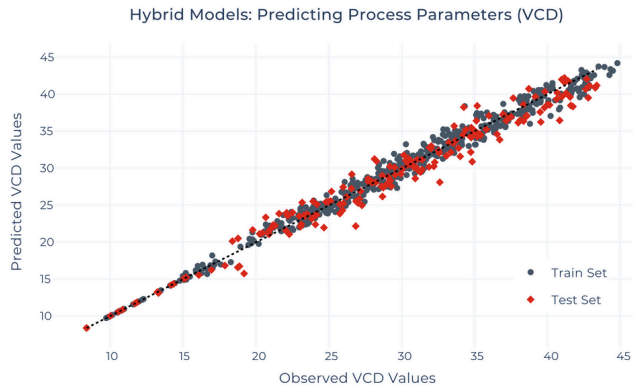
OBJ. 2 - Assess how much data is needed to accurately predict CQA's for each approach



Hybrid models can reliably predict CQAs with 3x fewer experiments

- Black box models needed 30 experiments before they could sufficiently understand the CPP/ CQA interrelationships and reliably predict CQA values
- Hybrid models only required 10 experiments to reach the same level of predictive accuracy
- Hybrid offers development teams greater development efficiency, process performance, and product quality

OBJ. 3 - Assess other benefits of Hybrid Models



Hybrid Models return more process knowledge and insight to support greater process control and reduce risk

- Black-box models are able to predict CQAs but can not support in predicting process parameters. Additional analysis is required, with no industry standard approach.
- DataHow's hybrid models can accurately predict both CQAs and process parameters, offering greater process insight and control
- Highlighted example (VCD): Hybrid models were able to predict 70% of VCD variability

+35% Improved Prediction Accuracy

3x Fewer Experiments

Process Knowledge & Insight

DATAHOW

This was performed in collaboration with BMS as part of a published work

www.datahow.ch