Case Study

Accelerating Strain Screening through Transfer Learning

Reducing experimental workload by up to 50% while improving prediction accuracy across scales and conditions.

Introduction

A change in strains and yeast extracts can significantly impact the cost of goods for large-scale probiotics manufacturing. To optimize production, different strains and yeast extracts are screened during development to identify the best-performing combination.

To increase throughput and reduce cost and time, these screenings are typically performed at microliter scale. A few top-performing strains are then selected for evaluation at larger scales, typically ranging from 250 milliliters to 20 liters. However, performance at small and large scales often does not align.

Challenge

Can we learn scale dependency from historical data and predict how a new strain or yeast extract will behave at large scale based on small-scale data? Procelys by Lesaffre screens many yeast extracts on hundreds of strains under various conditions and scales, each year, in order to provide the best yeast nutrient to fit needs of different microorganism. To begin this study, we focused on a dataset containing four probiotic strains evaluated with NuCel® yeast extracts at two scales.

Objective

OBJ. 1: Evaluate knowledge transfer potential: Using data from both scales, two NuCel® yeast extracts, and three strains, determine what data is required for the fourth strain to predict its behavior with both yeast extracts at large scale.

OBJ. 2: Increase process capacity: Assess whether the model can replace some experiments, reducing the number of runs per strain and yeast extract at both scales.

Approach

Procelys by Lesaffre shared process data for four strains from different genera. Each strain was cultivated with two NuCel® yeast extract at two scales using a BioLector XT and an ambr250. Four replicates were conducted per condition at BioLector XT scale and three at ambr250 scale.

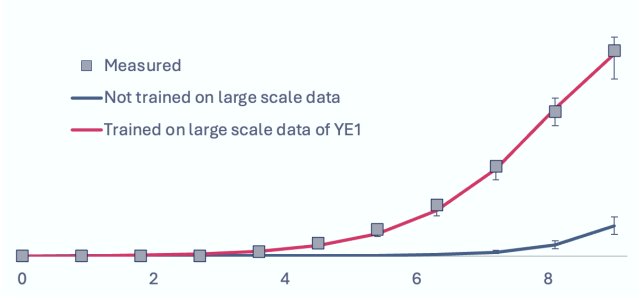
Using data from three strains across both scales and yeast extracts, along with small-scale data from the fourth strain, we evaluated how much and what type of data was needed to predict the fourth strain's process behavior at large scale.

Predictions were compared to experimental data to assess performance. This process was repeated for each strain and yeast extract, ensuring every strain was predicted once at large scale. The focus was on predicting biomass growth, as it determines process performance.

How Can Knowledge Be Transferred?

There are several machine learning approaches to transfer knowledge, including embeddings, meta-learning, transfer learning, one-hot encoding, and scaling. While each has strengths depending on the scenario, all enable the transfer of general process behavior across scales—requiring only small-scale run data to capture strain-specific behavior.

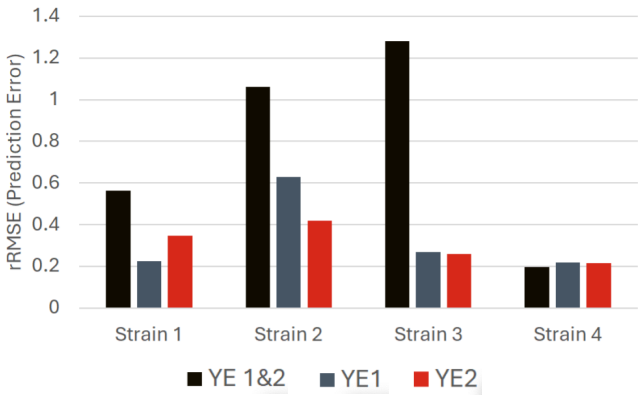
1. Qualitative Assessment of Measured & Predicted Evolution of Biomass Concentration



Using strain 3 (Lactobacillus) as an example the model could not predict its large-scale behavior, as it differed significantly from its small-scale behavior.

Adding data for strain 3 with YE2 at large scale enabled the model to accurately predict strain 3's large-scale behavior with YE1, successfully transferring knowledge from the smaller scale.

2. Knowledge Can Be Transferred Across Scales

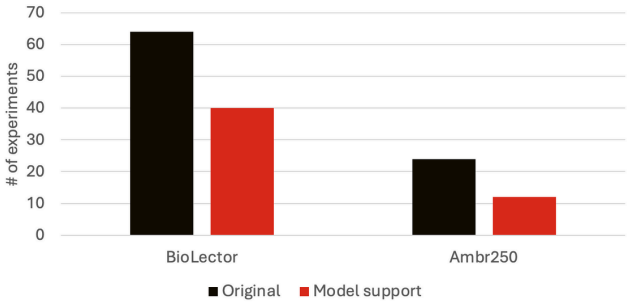


For most strains, prediction error at large scale was high when no large-scale data was available for that strain, as each strain behaved differently.

Adding large-scale data from the same strain with one yeast extract (YE1 or YE2) enabled the model to accurately predict the strain's large-scale behavior for the respective yeast extract.

Transfer learning performance is expected to improve as the amount of training data increases, reducing experimental requirements and increasing process capacity.

3. Reduction in Replicates Possible due to Transfer Learning Capability



While replicates are important for understanding variability, the model can reduce the need for them by transferring knowledge about strain and yeast extract variability.

In this case, the number of experiments could be reduced by 38% at BioLector scale and 50% at ambr250 scale without compromising decision quality.

Further reductions may be possible with additional strain data, especially when strains belong to the same genus.

38%

Reduction in Number of Experiments at BioLector Scale

50%

Reduction in Number of Experiments at ambr250 Scale