Recombinant Protein Titer Imputation using Kalman Filters

This is a brief report of the methodology and underlying assumptions employed for the imputation of missing titer values in our dataset of 32 batches, recorded every 2 hours over a total fermentation time of 48 hours. Due to the substantial amount of missing titer data—particularly with only 16 batches having recorded titer values at the end of the experiments—it was necessary to adopt a robust approach for imputing these missing values.

Approach and Assumptions

The method chosen for the imputation was the Kalman Filter, known for its efficacy in handling time series data with missing entries. The specific steps and assumptions made in the imputation process are as follows:

1. Assumption 1: Initial Titer Values:

Titer values are assumed to be zero from the start of the experiment until a significant depletion in phosphate levels (≤ 3 mg/ml). This assumption is based on the understanding that phosphate, being a critical carbon source, must deplete significantly before any substantial protein production occurs.

2. Assumption 2: End-Point Titer Values:

For the 16 batches that had recorded titer values, these values at the 48-hour mark ranged from 0.3 to 2.2. It is assumed that the titer values at the end of the experiment for any batch would fall within this range.

3. Assumption 3: Titer Growth Post-Phosphate Depletion:

After significant phosphate depletion, titer values are likely to increase exponentially. An exponential growth equation was used to model the titer values for 5 out of 32 batches from zero

at the point of significant phosphate depletion to the end values ranging from 0.3 to 2.2 (or the titer value for cases where the dataset had titer).

These assumptions were crucial in setting the initial conditions for the Kalman Filter, enabling us to extrapolate the missing titer data for the remaining batches.

Purpose of these Assumptions

The primary reason for these assumptions is to create a subset of data—batches with complete titer values from start to finish of the experiment. This subset serves as a critical learning set for the Kalman Filter to impute missing titer values accurately across other batches.

Results from the Kalman Filter Method

Estimated Labels vs. Time for Batch exp_220329c2_6

Estimated Labels vs. Time for Batch exp_220329c2_4