0. FERMENTATION PROCESS

Spectral data is collected from the process in real time, and stored as a `.txt` file every minute.

1. DATA DRIVEN MODEL

⟳ Updated every 15 min

>> Input: a matrix $M \sim (n \times m)$ containing the spectral data. n are the samples collected (spectrum at eachtime point). m are the different variables (wavenumbers in the spectrum).

Step 1.1. Data pre-processing: preprocessing the matrix M:

Step 1.1.1. First derivative Savitzky-Golay filter.

Step 1.1.2. Mean center the data.

Step 1.2. PLS modelling: PLS regression is used to estimate the vector

$$\hat{y} \sim (n \times 1)$$

containing the glucose concentration at each time n point from the matrix M.

>> Output: a vector $\hat{y} \sim (n \times 1)$ with the estimated concentration of glucose at each time point.

2. DYNAMIC MECHANISTIC MODEL

⟳ Updated every 15 min

>> Input: a vector $\hat{y} \sim (n \times 1)$ with the estimated glucose concentration at each time point.

>> Input: a dynamic mechanistic model describing the fermentation process with the form:

$$dz/dt = f(z,\theta,t)$$

where z are the state variables of the model, θ the model parameters and t the time.

>> Input: a vector $z0$ with the initial conditions of the state variables.

Step 2.1. Filter the data: the vector $\hat{y}$ is smoothed using a median filter.

2.2. Recursive parameter estimation:

Step 2.2.1. Parameter update: the parameters

$$[Ki_{FA,g}, Ki_{HAc,g}]$$

(inhibition constants on glucose uptake by furfuryl alcohol and acetic acid) are estimated using maximum likelihood estimation (MLE), by fitting the model to the vector $\hat{y}$.

Step 2.2.2. Uncertainty in the estimated parameters: using the bootstrap approach, the parameters $[Ki_{FA,g}, Ki_{HAc,g}]$ are re-estimated 50 times allowing to calculate their mean values and standard deviations.

Step 2.3. Model prediction:

Step 2.3.1. The unceratinty in input parameters is defined from the Step 2.2 and the remaining parameters are assumed to have an uncertainty of 1 %.

Step 2.3.2. Perform 100 Monte Carlo simulations of the model sampling from the parameter space defined by their uncertainties.

Step 2.3.3. Extract statistical data (mean and standard deviation) of the prediction of each state variable.

>> Output: a matrix $\hat{Y}$ containing the predictions for all the state variables (measured and unmeasured) and their 95 % confidence interval.