23ème Congrès de Chimiométrie, Université de Bretagne Occidentale, Brest, Février 2022.

Digital Twin of Hepatitis B vaccine culture step

T. Cornet[1]
[1] GSK, Rue de l'institut 89 B-1330 Rixensart Belgium, thomas.p.cornet@gsk.com

Keywords: Digital Twin, Raman, FTIR, software sensor, culture, vaccines.

1 Introduction

The process to produce the Hepatitis B vaccine developed by GSK uses recombinant *Saccharomyces cerevisiae* yeast to produce the antigen of interest. During the culture step, the glucose is transformed into CO2 via the respiratory pathway. But, if glucose is added in excess, the fermentative pathways are activated leading to ethanol and acetate productions. These metabolites are by-products that we want to minimize because it is less energetically efficient, and they can induce the stop of the yeast growth at high concentrations.

Therefore, we want to create a digital twin of this culture step developing, in a first step, software sensors to be able to follow these metabolites in-line. In a second step, we will develop control loops so that we can modify the culture process parameters if deviations are observed by our software sensors. This poster describes the software sensors building.

2 Material and methods

To build our software sensors, we used two complementary spectroscopy techniques : Raman and FTIR. We used the Raman system RXN2 from Kaiser and the FTIR system ReactIR 702L from Mettler Toledo.

We performed 8 fermentation lots of 90h using different process parameters and metabolites spikings. We took 1 spectrum (Raman and FTIR, lasting 5 minutes) every 2h and took at the same time an offline sample to calibrate our models. In this offline sample, we measured biomass and our metabolites of interest (glucose, ammonium, acetate and ethanol) using gas chromatography and a photometric analyzer (Gallery from Thermofisher). In total, we analyzed 360 samples.

3 Results and discussion

We obtained good model performances for biomass (measured by wet cell weight) and our four studied metabolites (ethanol, acetate, glucose and ammonium) using a PLS regression and a multiblock approach combining the Raman and the FTIR spectra (figure 1).

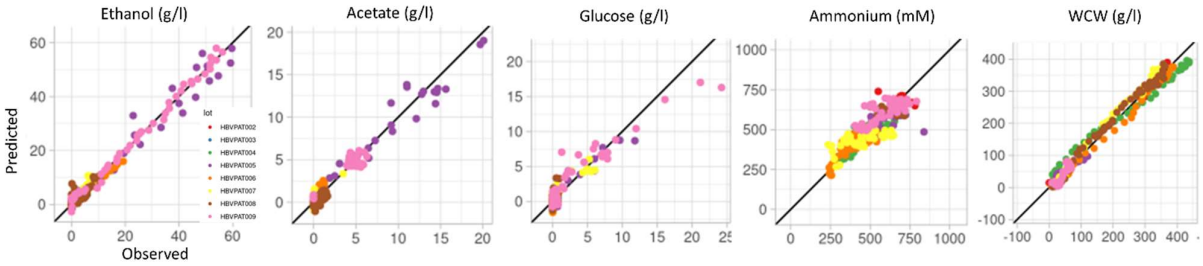

Figure 1 – Observed vs predicted values for ethanol, acetate, glucose, ammonium and biomass (measured by wet cell weight) models using a PLS regression combining the Raman and FTIR spectra. A 10-fold cross validation was performed.

We evaluated other models using only FTIR or Raman spectra to see if only one spectroscopy technique was enough. We can see that FTIR and Raman spectroscopy techniques are complementary because some metabolites/outputs were better predicted with FTIR spectra (ethanol, acetate, glucose and CO2) whereas ammonium and biomass were better predicted with Raman spectra (table 1). However, combining FTIR and Raman data generally leads to better models with lower RMSECV (root mean square error of cross-validation).

We also evaluated other regression methods using machine learning techniques (random forest, support vector machines, regularized linear regression). They generally lead to better models with lower RMSECV compared to models obtained with PLS regression.

Table 1 – Cross-validation results with PLS models (with combined spectra (both), only FTIR or only Raman) or other machine learning models.

RMSECV	PLS models			ML models	
	Both	FTIR	Raman	Both or FTIR	ML method
ethanol (g/L)	1,61	2,23	2,89	1,60	random forest
acetate (g/L)	0,62	0,62	0,89	0,43	regularized linear regression
glucose (g/L)	0,86	1,04	1,07	0,66	random forest
ammonium (mM)	63,95	115,68	84,97	40,82	random forest
OD (650nm)	6,10	11,73	4,99	3,24	random forest
WCW (g/L)	18,63	36,57	20,76	9,09	random forest
DCW (g/L)	4,15	8,80	5,07	2,48	random forest
CO2 (%)	0,49	0,46	0,49	0,26	support vector machine

4 Conclusion

We built fast and efficient software sensors that allows to follow in-line the yeast culture biomass and metabolites of interest such as glucose, ethanol, acetate and ammonium. We will test these software sensors on new lots to have a better view on the model performances.

This is the first step for the building of a digital twin of the culture step of the hepatitis B vaccine. The second step will be to build control loops to modify process parameters based on the software sensors values. This will be done by developing a yeast metabolic model and a multivariable controller.