(Orthogonal) Partial Least Square ((O)-PLS), Support Vector Machine (SVM) and Artificial Neural Network (ANN), comparing methods for modeling cell culture production processes for predicting the quality of batches

M. Kerbrat[1] C. Lang [1] T. Schmidberger[1]

[1] Sartorius Stedim Biotech, maina.kerbrat@sartorius.com, chloe.lang@sartorius.com, timo.schmidberger@sartorius.com

Keywords: batch modeling, cell culture development, OPLS, SVM, Artificial Neural Network.

1 Introduction

Market demand and regulatory requirements place great emphasis on product quality in industry. Consequently, process monitoring and control have become important aspects of systems engineering. The end of batch and fed-batch processes is typically reached when a cell viability drops below a specified limit. Understanding which conditions are responsible for maintaining high viability is a key for any bioprocess development. In that context, batch modeling is performed to study the evolution of the process of batches with time. This is a way to detect fault diagnosis for cases when continuous data profiles are to be analyzed.

2 Material and methods

In this study, fourteen fed-batch cell culture processes were performed for process development objectives. The data for the model contain of process measurements of 12 process points as well as measurements at the end of the process Regression methods were used to predict the final Titer, final Viable Cell density (VCD) and final viability of the batch using the data from the 12 process days were used. The aim of this study is to compare Multivariate data analysis regression methods such as Partial Least square (PLS) and Orthogonal Partial Least Square (OPLS) regarding machine learning technics such as Support Vector Machine (SVM) and Artificial neural network (ANN).

To apply these techniques of regression, the dataset is unfolded in a way to have one column per day and per variable, to have one row per batch. Due to the limited number of batches, a Leave-one-out cross-validation is performed on the dataset.

SIMCA software 17.0.2 is used to perform PLS and OPLS regressions. R4.1.2 is used to perform SVM and ANN.

3 Results and discussion

The table 1 shows the R^2Y and the RMSECV for each method and each target.

Regardless of the target, the prediction performance of the PLS regression seems lower than for the other prediction methods. For final titer and final viability, the best prediction model is obtained by ANN with the highest R^2Y and the lowest RMSECV. For Final VCD, the highest R^2Y is obtained by OPLS and the best RMSECV is obtained by ANN.

Table 1 –R^2 and RMSECV for PLS, OPS, SVM and ANN for final titer, final VCD and Final Viability

	R^2Y				RMSECV			
MODEL	**PLS**	**OPLS**	**SVM**	**ANN**	**PLS**	**OPLS**	**SVM**	**ANN**
Final titer	0.933	0.994	0.990	**0.994**	0.225	0.150	0.043	**0.033**
Final VCD	0.994	**0.996**	0.991	0.991	3.419	1.005	0.423	**0.414**
Final Viability	0.930	0.967	0.992	**0.996**	1.109	0.678	0.149	**0.100**

Another important aspect in batch modeling is to be able to understand the deviation of the batch during the process and to correct at the right moment. To make some process corrections, drill-down functionalities are obtainable only using multivariate techniques such as the OPLS method. The contribution barplot (Figure 1) is used to highlight which variables are responsible for the deviation of the deviating batch. The Sodium and pCO2 are detected as outside limits, indicating that the pH control is not working. This deviating batch has a strange behavior regarding the trend of other batches (figure 2). Recalibration of the pH sensor might be necessary in order to ensure the process is controlled correctly and potentially correct the trajectory to become inside the limits.

Figure 1 – contribution barplot of a deviating batch

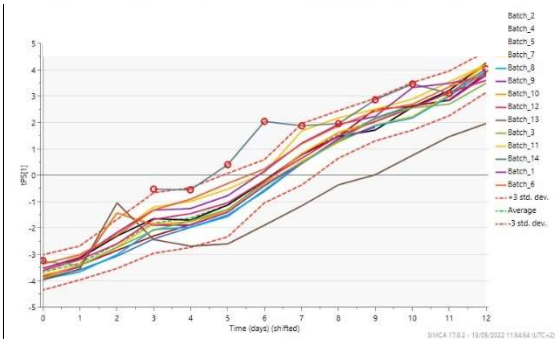

Figure 2 – Multivariate control chart

4 Conclusion

To conclude, we highlighted that ANN, in a batch monitoring context is slightly better performing compared to more traditional methods for regression to predict the final quality of the batch in real-time during a cell culture development. To correct the quality of the batch in real-time, more interpretable methods are preferable. In that way, final quality parameters such as Titer, VCD and Viability can be monitored to keep the best quality as possible and save a deviating batch.

5 References

[1] Anurag S. Rathore et al: *Application of Multivariate Data Analysis for Identification and Successful Resolution of a Root Cause for* a Bioprocessing Application, biotechnology progress, volume24, issue 3

[2] Elif Seyma Bayrak, Tony Wang, Aditya Tulsyan, Myra Coufal, Cenk Undey, *Product Attribute Forecast: Adaptive Model Selection Using Real-Time Machine Learning,* IFAC-PapersOnLine, Volume 51, Issue 18,2018,