**Hybrid Models: A new generation of predictive models for process design, development and optimisation**

Harini Narayanan1 \*, Michael Sokolov1,2, Alessandro Butté1,2, Massimo Morbidelli1,2

*1 Institute of Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland; 2 DataHow AG, Zurich, Switzerland*

*\*Corresponding author: nharini@chem.ethz.ch*

**Highlights**

* Hybrid models have better accuracy in prediction compared to classical PLS models
* Hybrid models have better interpolation and extrapolation capabilities
* Ideal for Experimental Design and process optimization

**1. Introduction**

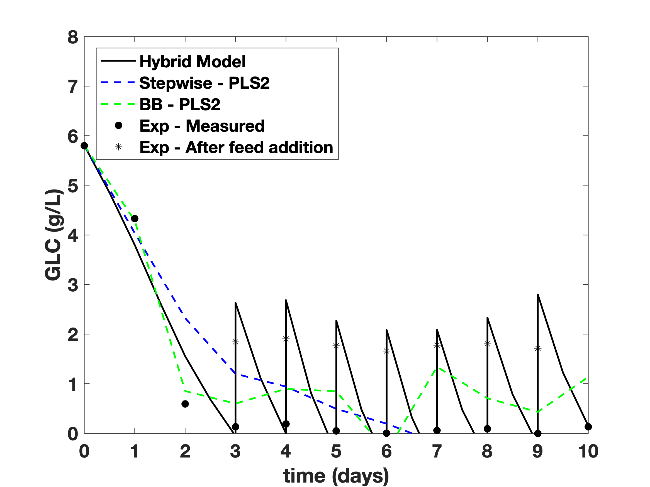
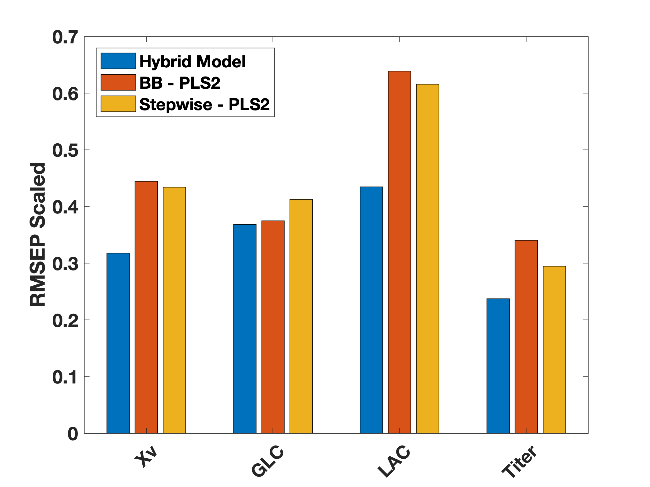
Over the last few years, the biopharmaceuticals market has witnessed tremendous growth reaching a value of worth $186,470 million in 2017. Now, the biggest challenge of the industry is to produce high-quality products in a cost–effective manner. To meet the market demands, there is a need for rapid and efficient process development that calls for the use of model-based methods to assist at several stages of process development, monitoring and control. The lack of complete understanding about the metabolic network and reaction pathways observed in the cells, especially mammalian cells hinders the development of First Principle Models (FPMs). Thus, increased efforts have been devoted to the use of statistical models (such as Partial Least Squares – PLS models) using historical, online spectral and sensor data [1]. However, one of the major limitations of statistical models is its invalidity or poor performance outside of the regions not explored in the underlying experiments. As a result, statistical models are not suitable for process design, development and optimization [2]. We propose and quantitatively show the advantages of using Hybrid modeling strategy which explores a synergy between the classical First Principle models and statistical models.

**2. Methods**

The FPMs provide a mixed system of differential - algebraic equation which often contains specific unknown parameters. The hybrid models establish a synergy by using statistical models to deduce these unknown parameters. Particularly, we pose mass-balance or pseudo mass-balance equations for bioreactor systems and use artificial neural networks to estimate the unknown specific rates. Through different bioprocess case studies, we showcase the added value of hybrid models compared to the state-of-the-art PLS models, in terms of: Predictive accuracy, interpolation and extrapolation capabilities and application in design of experiments for optimization of titer (concentration of monoclonal antibody produced).

**3. Results and discussion**

Hybrid models have the lowest errors in prediction for the key state variables (Figure 1, Left). Lower prediction errors ensure reliable models that can be used as soft sensors to replace expensive analytics. Secondly, given a fed – batch system with glucose feed from day 3, hybrid models correctly predict a daily continuous decrease in glucose concentration whereas the PLS models are not capable of incorporating the feeding strategies in a physically representative manner. Additionally, Hybrid models is seen to predict process variables accurately outside of the trained or observed operating conditions highlighting its potential application in *in-silico* design of experiments and process optimization (results not shown here).



**Figure 1.** (Left) Comparison of the accuracy of the hybrid model against two statistical models (both based on PLS) in terms of standardized errors in prediction for viable cell density (Xv), Glucose (GLC), Lactate (LAC) and titer. (Right) The concentration of glucose predicted by the three models as a function of time along with the experimental observation.

**4. Conclusions**

A success in efficient data management and analytics is expected to have a tremendous effect on improving process development strategies and reducing manufacturing costs, thus making these therapies more available to society. We show that a modest modeling effort is sufficient to generate a hybrid model capable of predicting the complete evolution of the process variables and titer solely based on the initial conditions and the process operating conditions. Hybrid modeling also shows good interpolation and extrapolation capabilities in comparison to statistical models, which face intrinsic limitations to predict outside their training process condition space. This extrapolation capability is not only central for model–based process optimization but also for process scale–up, technology transfer and decision-making in process development. Thus, we strongly believe in investing efforts in the direction of hybrid modeling for process optimization and further extend its application to process monitoring and control.

**References [Calibri 10]**

1. ﻿T. Kourti, “Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions,” J. Chemom., vol. 17, no. 1, pp. 93–109, 2003.
2. ﻿D. Solle et al., “Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation,” Chemie-Ingenieur-Technik, vol. 89, no. 5, pp. 542–561, 2017