Supplementary Information

Hybrid models based on machine learning and an increasing degree of process knowledge: Application to cell culture processes

Authors: Harini Narayanan[1], Martin Luna[1], Michael Sokolov[2], Alessandro Butté[2,*], Massimo Morbidelli[2,3,*]

[1]Institute of Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zurich, 8093 Zurich, Switzerland

[2]DataHow AG, Zürichstrasse 137, 8600 Dübendorf, Switzerland

[3]Dipartimento di Chimica, Materiali e Ingegneria Chimica, Giulio Natta, Politecnico di Milano, 20131 Milano, Italy

Corresponding Authors: Alessandro Butté, Zürichstrasse 137, 8600 Dübendorf, Switzerland

E-mail: a.butte@datahow.ch

Massimo Morbidelli, Dipartimento di Chimica, Materiali e Ingegneria Chimica, Giulio Natta, Politecnico di Milano, 20131 Milano, Italy

E-mail: massimo.morbidelli@polimi.it

1. Choice of data-driven modeling method

The different data-driven (DD) modeling approaches considered here are as follows:

1.1. Black Box – Partial Least Square Model (PLS1)

BB-PLS1 are PLS1 models connecting the process conditions, Z and the initial conditions, X(t=0) to the concentration of different process variables at different times, $X_i(t = t_{model})$. Thus, one model per timepoint per process variable is developed, denoted as $PLS1_{i,\,t}$.

$$[\mathrm{Z}, \mathrm{X}(\mathrm{t} = 0)] \xrightarrow[PLS1_{i,t}]{} X_i(t = t_{model}) \qquad (1)$$

where Z, X are the information matrix indicated in section 2.1.1, *i* is the *i-th* process variable and t_{model} is the timepoint for which the model is developed. For the *in-silico* dataset, t_{model} varies from day 0 to day 14 while for the experimental dataset it varies from day 0 to day 10.

1.2. Batch Wise Unfolded - PLS1

In BWU-PLS1, a different PLS1 model is developed per time point per process variable, ($PLS1_{i,\,t}$), using the process condition and the historical information available until a given time, $t = t_{model}$, to predict the process variable at that time. The mathematical representation is as follows:

$$\textit{Training}\text{:}\ \ [\mathrm{Z}, \mathrm{X}(\mathrm{t} < t_{model})] \xrightarrow[PLS1_{i,t}]{} X_i(t) \qquad (2)$$

where Z, X are the information matrix indicated in section 2.1.1, i is the *i-th* process variable and $t = t_{model}$, the timepoint for which the model is developed. For the *in-silico* dataset, t_{model} varies from day 1 to day 14 while for the experimental dataset it varies from day 1 to day 10. It is worth noting that during training, the actual measurements are used as input for the model (Equation (2)), but during testing, predictions from the model at the previous time points are used, as shown in the following equation:

$$\textit{Testing}\text{:}\ \ [\mathrm{Z},\ \mathrm{X}\ (\mathrm{t} = 0),\ X^{predicted}(\mathrm{t} < t_{model})] \xrightarrow[PLS1_{i,t}]{} X_i^{predicted}(t) \qquad (3)$$

where $X^{predicted}$ indicate the prediction of the process variables made by the respective PLS1 models of different time points $PLS1_{i,\,t\,<\,t_{model}}$. A detailed discussion of the BWU-PLS1 models can be found in [38,46].

1.3. Artificial Neural Network

ANN models per variable per time point ($ANN_{i,t}$) are used to predict the process variable at given time, $t = t_{model}$, using the process condition and process variables of the immediate previous time, i.e., $t = t_{model} - 1$.

$$\textit{Training}\text{:}\ \ [\mathrm{Z}, \mathrm{X}(\mathrm{t} = t_{model} - 1)] \xrightarrow[ANN_{i,t}]{} X_i(t = t_{model}) \qquad (4)$$

$$\textit{Testing}\text{:}\ \ [\mathrm{Z}, \hat{\mathrm{X}}(\mathrm{t} = t_{model} - 1)] \xrightarrow[ANN_{i,t}]{} X_i(t = t_{model}) \qquad (5)$$

Here again, the different quantities hold the same meaning as before. Additionally, similar to the BWU-PLS1, actual measurements are used during the training phase, while predicted ones are used during the test phase.

1.4. Variable Wise Unfolded - ANN

A similar modeling strategy as ANN is used for VWU-ANN with the only difference that for VWU-ANN, a single model is used for all time points (ANN_i).

$$Training\text{:}\ \ [\mathrm{Z}, \mathrm{X}(\mathrm{t} = t_{model} - 1)] \xrightarrow[ANN_i]{} X_i(t = t_{model}) \qquad (6)$$

$$Testing\text{:}\ \ [\mathrm{Z}, \hat{\mathrm{X}}(\mathrm{t} = t_{model} - 1)] \xrightarrow[ANN_i]{} X_i(t = t_{model}) \qquad (7)$$

2. Comparison among the Data Driven models

Figure S1 shows the comparison of predictive accuracy among different data-driven (DD) models trained with different number of runs. The models are organized in the order of increasing complexity with the BB-PLS1 model being purely black-box, while the BWU-PLS1 being a linear model but accounting for the historical information. Subsequently, the two NNs are complex nonlinear models where ANN model builds one model per time and the VWU-ANN builds one model accounting for all time points. In this regard, the VWU-ANN resembles closely the other dynamic models, that is, the hybrid and mechanistic models.

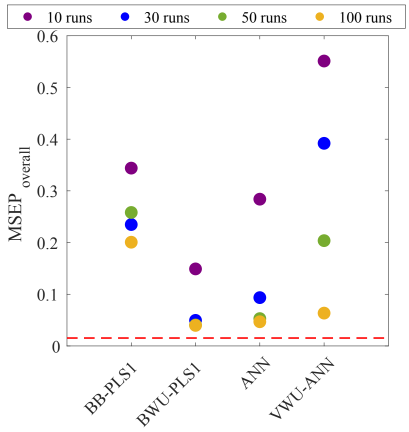

Figure S1: Comparison of the different DD model namely BB-PLS1, BWU-PLS1, ANN and VWU-ANN trained with different number of runs: 10, 30, 50 and 100 runs, based on the predictive accuracy in interpolation.

It can be observed that BB-PLS1, a pure black box model that does not account for any historical process information, performs extremely poor even when trained using 100 runs. This quantitatively validates the fact that BB-PLS1 is too simple to capture the extremely non-linear dynamic behavior of cell culture systems. On the other end, the VWU-ANN is a complex model attempting to capture all the dynamic effects of a variable using a single NN. When the number of training runs are insufficient the model is likely to overfit in training, thus, performing poor on the test set. As a result, with increasing number of experiments the predictive accuracy of the model increases. However, even with 100 runs, the VWU-ANN model ($\mathrm{MSEP}_{\mathrm{overall}}$ = 0.063) does not match the performance of the BWU-PLS1 model

($\text{MSEP}_{\text{overall}}$ = 0.039), indicating that more experiments might be required to train a robust VWU-ANN model.

Amongst the BWU-PLS1 and ANN models, the BWU-PLS1 model performs better when trained with 10 runs. Subsequently, the performance of both the models improves with increasing training data. The BWU-PLS1 model required 50 runs to achieve its best predictive capability after which the model had a constant prediction metric $\text{MSEP}_{\text{overall}}$ = 0.039. However, the ANN required 100 experiments to achieve a comparable prediction metric. The $\text{MSEP}_{\text{overall,analytical}}$ metric for the measurement error ($MSEP_{overall,\, analytical}$) is 0.0163 which marks the predictive limit of the models. Thus, the predictive capability of the BWU-PLS1 model is away from the best predictive performance possible and this can be attributed to the linear nature of the model. On the other hand, use of purely data driven complex non-linear model such as NN requires a considerable number of experiments (e.g., more than 100 experiments for VWU-ANN and ANN model). This problem can be alleviated by supporting the data-driven models with process knowledge.

(1) Narayanan, H.; Behle, L.; Luna, M. F.; Sokolov, M.; Guillén-Gosálbez, G.; Morbidelli, M.; Butté, A. Hybrid-EKF: Hybrid Model Coupled with Extended Kalman Filter for Real-Time Monitoring and Control of Mammalian Cell Culture. *Biotechnol. Bioeng.* **2020**, *117* (9), 2703–2714. https://doi.org/10.1002/bit.27437.

(2) Narayanan, H.; Sokolov, M.; Morbidelli, M.; Butté, A. A New Generation of Predictive Models: The Added Value of Hybrid Models for Manufacturing Processes of Therapeutic Proteins. *Biotechnol. Bioeng.* **2019**, *116* (10), 2540–2549. https://doi.org/10.1002/bit.27097.

(3) Craven, S.; Shirsat, N.; Whelan, J.; Glennon, B. Process Model Comparison and Transferability across Bioreactor Scales and Modes of Operation for a Mammalian Cell Bioprocess. *Biotechnol. Prog.* **2013**, *29* (1), 186–196. https://doi.org/10.1002/btpr.1664.

(4) Xing, Z.; Bishop, N.; Leister, K.; Li, Z. J. Modeling Kinetics of a Large-Scale Fed-Batch CHO Cell Culture by Markov Chain Monte Carlo Method. *Biotechnol. Prog.* **2010**, *26* (1), 208–219. https://doi.org/10.1002/btpr.284.

(5) Narayanan, H.; Sokolov, M.; Butté, A.; Morbidelli, M. Decision Tree-PLS (DT-PLS) Algorithm for the Development of Process: Specific Local Prediction Models. *Biotechnol. Prog.* **2019**, *35* (4), 1–11. https://doi.org/10.1002/btpr.2818.