Importance of oxygen signal shape matching for robust parameter estimation in bioprocess development

Judit Aizpuru[1], Annina Kemmer[1], Niels Krausch[1], Peter Neubauer[1], M. Nicolas Cruz Bornazou[1,2], Tilman Barz[3]

[1] Technische Universität Berlin, Chair of Bioprocess Engineering, KIWI-biolab
Straße des 17. Juni 135, 10623 Berlin, Germany
[2] DataHow AG, c/o ETH Zurich, HCI, F137 ,Vladimir-Prelog-Weg, 8093 Zurich, Switzerland
[3] AIT - Austrian Institute of Technology, Center for Energy, Giefingasse 4, 1210 Vienna, Austria

www.kiwi-biolab.de www.bioprocess.tu-berlin.de

Motivation

The characterization of macro kinetic growth models is one of the main tasks in high throughput bioprocess development, where accurate parameter estimations can enable a better understanding of the system and improve the performance of model predictive control actions, among others [1]. The experiments performed for this characterization typically gather different measurements at varying sampling rate, e.g. at-line measurements for biomass and metabolites, versus online monitoring for gases and pH. The usual procedure to fit these data to the systems model, is to minimize the sum of the squared residuals known as Least Square Estimation (LSE), where the measurement errors are assumed to be normally distributed [2]. However, oxygen signals manifest very rapid dynamics, and the optical sensors typically used for its measurement have shown a delayed response. Even if this latter drawback can be solved by calibrating the sensor with a first order step response [3], the variance of the response time remains still high in a system of parallel bioreactors, leading to a poor measurement accuracy and therefore a wrong estimation of the model parameters using LSE. The hypothesis of this work is that using a different estimator which considers the signals shape and not just amplitude, as some estimator derived from Dynamic Time Warping, the overall quality of the fitting problem improves. The comparison study and validation are made via an in-silico data generator with assumed error models and known parameter values.

Concept

Online oxygen measurements show very fast dynamics during bacterial cultivations where bolus feeding is applied [4]. These steep changes combined with the high sampling frequency of the data, are beneficial for model identification.

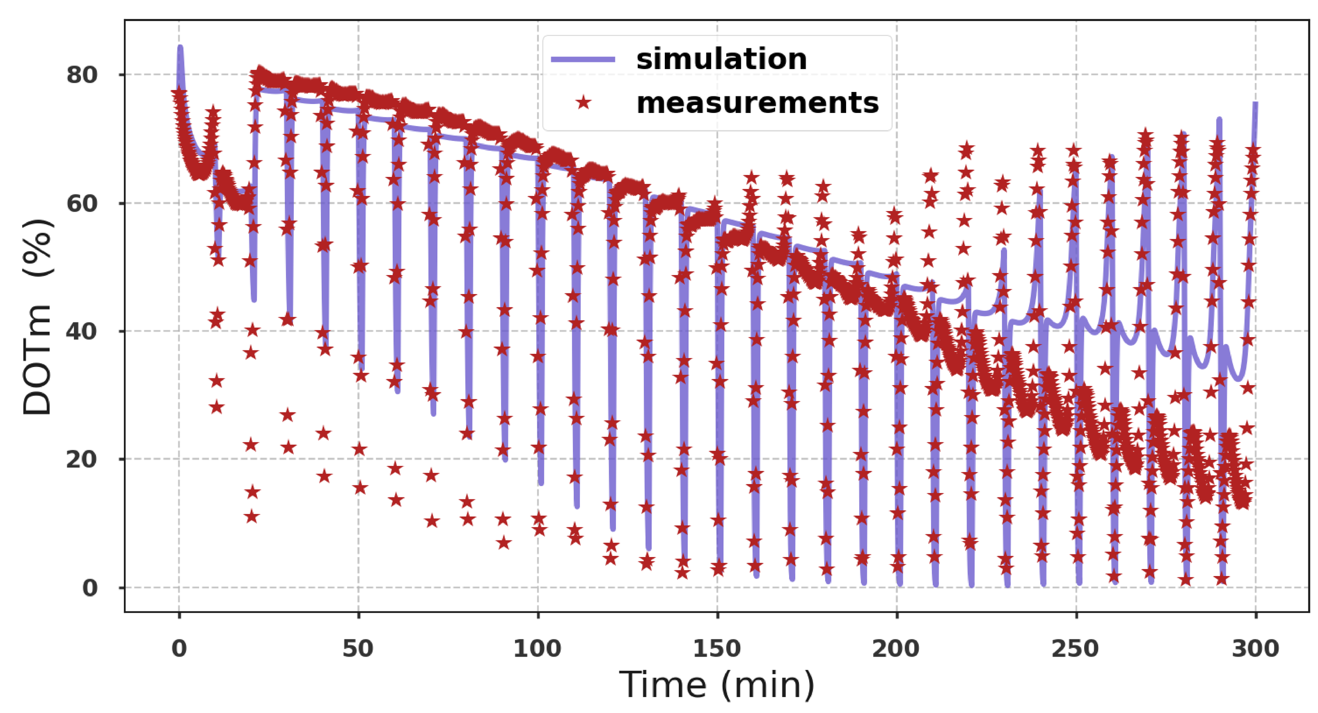

Figure 1: Simulated Dissolved Oxygen Tension (DOT) and experimental data taken from literature [4].

Parameter sensitivities are very useful for analysing the relationship between model outputs and its parameters. Sensitivities can be efficiently computed together with the solution of the original mathematical model. This information can be used for feature selection in an underdetermined problem [2] or for improving the model characterization via optimal experimental design [5].

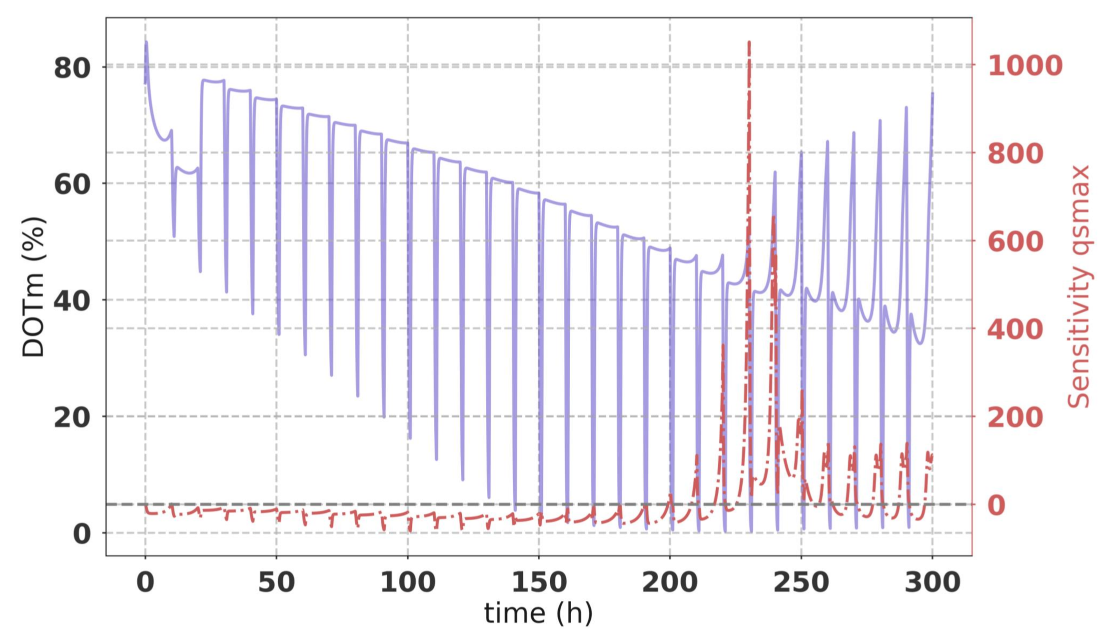

Figure 2: Sensitivity of DOT with respect to the Monod parameter for the maximum substrate uptake rate, *qsmax*.

Optical sensors for online monitoring of dissolved oxygen (Fig. 3) typically show a first order step response behaviour [6]; moreover, the variability in the mixing and aeration in parallel operated mini-bioreactors, and the dissimilarities between their sensor spots, can lead to significant differences in the measured oxygen values and curve shapes even for replicates (Fig. 4).

Figure 3: Integrated optical sensor spots manufactured by PreSens (Regensburg, Germany) for online oxygen and pH measurements

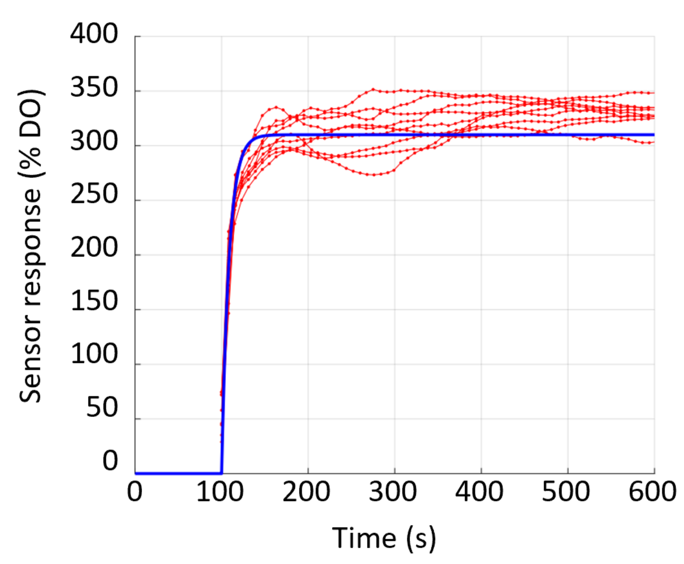

Figure 4: Experimental data provided by Annina Kemmer showing the delayed response of the optical sensor in parallel replicate reactor experiments

Fitting errors, also called residual values, are used to measure the accuracy of the model predictions. Standard least squares algorithms minimize a point-wise squared difference between the measured and predicted values. While these algorithms consider measurement errors in the signal's amplitude, they ignore errors in the time domain. This can lead to a bad fitting and corresponding poor parameter estimates. Figure 5 illustrates the difficulties in aligning steep signals using standard LSE where the data is most informative for determining certain parameters.

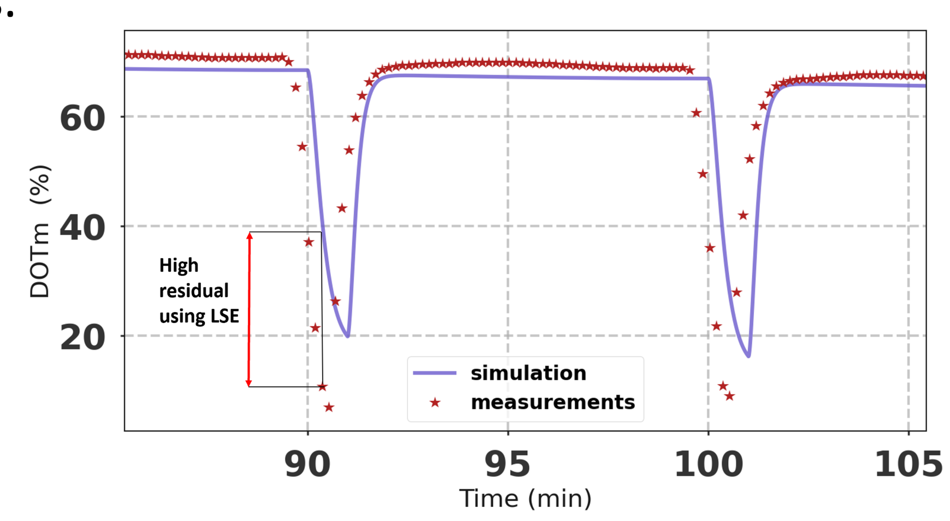

Figure 5: Zoomed portion of figure 1 to highlight the large residuals computed as difference between the measured and predicted DOT for fast changes in oxygen concentration

Conclusions

The fast dynamics that bolus feeding provoke in oxygen signals are highly informative for monitoring of cultivations and calibration of mechanistic growth models. However, if the collected data is not used in an appropriate way, it can hurt more than help the parameter estimation process. Using an optimization criterion which is independent of the delay time of the sensor, as some goodness score that considers shape match between simulated and experimental data, can significantly improve the parameter estimation. The selection and validation of appropriate measurement error models, and corresponding best fitting criteria, are addressed in the research project.

Acknowledgements / References

This work was supported by the German Federal Ministry of Education and Research through the Program "International Future Labs for Artificial Intelligence" (Grant number 01DD20002A)

1] Hans, S. *et al.* (2020) "Automated Conditional Screening of Multiple *Escherichia coli* Strains in Parallel Adaptive Fed-Batch Cultivations", *Bioengineering (Basel)*, *7*(4), 145. DOI: 10.3390/bioengineering7040145
[2] López, D. *et al.* (2015) "Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design", *Computers & Chemical Engineering*, 77, pp. 24–42. DOI: 10.1016/j.compchemeng.2015.03.002
[3] Anane, E. *et al.* (2017) "Modelling overflow metabolism in Escherichia coli by acetate cycling", *Biochemical Engineering Journal*, 125, pp. 23–30. DOI: 10.1016/j.bej.2017.05.013
[4] Anane, E., *et al.*(2019), "Modelling concentration gradients in fed-batch cultivations of *E. coli* – towards the flexible design of scale-down experiments", *J. Chem. Technol. Biotechnol.*, 94: 516-526. DOI: 10.1002/jctb.5798
[5] Cruz Bournazou, M., *et al.* (2017), "Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities", *Biotechnol. Bioeng.*, 114: 610-619. DOI: 10.1002/bit.26192
[6] Puskeiler, R. *et al.* (2005), "Development, parallelization, and automation of a gas-inducing millilitre-scale bioreactor for high-throughput bioprocess design (HTBD)". *Biotechnology and Bioengineering*, 89: 512-523. DOI: 10.1002/bit.20352

Institute of Biotechnology
Berlin Institute of Technology
Bioprocess Engineering

KIWI biolab

Sponsored by the Federal Ministry of Education and Research