The 8th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2022
5 – 9 June 2022, Oslo, Norway

HYBRID PROCESS MODELLING COMBINING MECHANISTIC EQUATIONS WITH MACHINE LEARNING

Jazib Hassan[1], Eric von Lieres[1]

[1] Research Center Jülich, Institute of Biotechnology 2, Wilhelm-Johnen-Str., 52425 Jülich, Germany.
j.hassan@fz-juelich.de

Key Words: *Hybrid Process Modelling, Chromatography, Machine Learning.*

Chromatography is commonly applied for separating and purifying target molecules in various biotechnological manufacturing processes. Chromatography modelling is increasingly used to better understand the separation of products and impurities, including their binding mechanisms. Over the past decades, the efficiency of numerical algorithms and computation power have significantly increased. Hence, complex separation processes can be now be modelled with comprehensive mechanistic models [1]. Even though the developed chromatography software can simulate and predict various physical mechanisms, the adsorption mechanisms of proteins in preparative chromatography are still not fully understood. This is because there is a complex interplay of many physical and environmental parameters. In this work, a hybrid process modelling technique is proposed in which single component mechanistic adsorption models are replaced by several data-driven models. Adsorption is modeled with cubic splines, Gaussian Process Regression (GPR), and Artificial Neural Networks (ANN), while the transport through the column is still modelled mechanistically with partial differential equations. CADET software is used to simulate the chromatography process [2]. The training data (adsorption isotherm data) for the data-driven methods used in this study is either synthetically generated, using established mechanistic adsorption models like the Langmuir isotherm, or experimental isotherm data obtained from a published study [3]. For synthetic datasets, various sampling techniques, e.g. uniform, equal-arc length, equal tangent slope increment, and based on smoothing splines with elbow plot, are used to sample a fixed number of training data points. The trained model for each method was then implemented in CADET. For the synthetic dataset, all data-driven methods were able to exactly capture the Langmuir adsorption behaviour and the corresponding breakthrough curve. For experimental adsorption data as training data, the results showed that all the trained data-driven isotherm models can capture the adsorption behaviour better than the Langmuir isotherm fitted to batch experimental data, but the resulting breakthrough curves for both cases are quite similar. Thus, the data-driven models reproduce the mechanistic model very well, without a need to have a closed-form of it. For isotherms with shallow slope, all the sampling techniques can be utilized for all the methods, but for steep slope isotherms, sampling based on smoothing splines with elbow plot is recommended. In terms of computational time, cubic splines have almost similar computational time when compared with the Langmuir model, while GPR and ANN are ca. 2 times and ca. 3.5 times slower, respectively.

References

[1] J. Diedrich, W. Heymann, S. Leweke, C. Kunert, W. Johnson, S. Hunt, B. Todd and E. von Lieres, "Multi-state steric mass-action model and case study on complex high loading behavior of mAb on ion exchange tentacle resin," *Journal of Chromatography A,* pp. 60-70, 2017.

[2] S. Leweke and E. von Lieres, "Chromatography Analysis and Desing Toolkit (CADET)," *Computers and Chemical Engineering,* pp. 274-294, 2018.

[3] T. Pabst, J. Thai and A. Hunter, "Evaluation of recent Protein A stationary phase innovations for capture of biotherapeutics," *Journal of Chromatography A,* vol. 1554, pp. 45-60, 2018.