**Hybrid Semi-Parametric Modeling of Preparative Protein Chromatography for Online Monitoring and Real-Time Process Control.**

Anna Christler1, Moritz von Stosch1, Theresa Scharl-Hirsch1,2, Michael Melcher1,2, Friedrich Leisch1,2, Astrid Dürauer1,3, Alois Jungbauer1,3

*1 Austrian Centre of Industrial Biotechnology (ACIB), Muthgasse 11, 1190 Vienna, Austria; 2 Institute of Applied Statistics and Computing, University of Natural Resources and Life Sciences (BOKU), Peter-Jordan-Strasse 82, 1190 Vienna, Austria; 3 Department of Biotechnology,University of Natural Resources and Life Sciences (BOKU), Muthgasse 18, 1190 Vienna, Austria*

*\*Corresponding author: anna.christler@acib.at*

**Highlights**

* A PAT application for preparative protein chromatography was established.
* Non-parametric statistical models were complemented with parametric mechanistic models.
* Hybrid semi-parametric models show enhanced prediction quality of critical quality attributes.

**1. Introduction**

Regulatory agencies encourage pharmaceutical industry to implement Quality-by-Design approaches into manufacturing processes. We integrated additional online sensors, namely infrared (ATR-FTIR), fluorescence, refractive index and static light scattering detectors, into a conventional chromatography workstation. Obtained online signals were statistically correlated to offline measured parameters of quantity, purity and potency. Thereby, predictive models were generated. Hybrid semi-parametric models combine such statistical correlations with first-principles expert knowledge [1]. Hybrid models are able to balance strengths and weaknesses of different sources of knowledge considering for example extrapolation power and data availability. This has been shown among others for monitoring of bacterial fermentation processes [2]. We applied this approach to optimize loading and elution of chromatographic steps where monitoring is currently mostly done via UV/Vis absorption only.

**2. Methods**

Multivariate data analysis enables extraction of unknown or unidentified correlations from large amounts of data sets. Online data obtained from implemented spectroscopic sensors were pre-processed to maximize the information content and reduce unspecific noise. Boosting algorithms allowed automated variable selection. CADET was used for efficient computing of mechanistic models [3].

**3. Results and discussion**

Different combinations of parametric and non-parametric model parts, i.e. parallel and serial hybrid structures, were compared and the best performing identified for monitoring of product attributes. Prediction accuracy and robustness were compared to solely data-based models and differences analyzed. A major advantage of the hybrid models is that predictions can be provided based on the feed composition while non-parametric model estimates are only available retrospectively with a time delay of up to 10 seconds.

**4. Conclusions**

The generated hybrid models enable controlled loading and peak cutting based on defined quality settings. Time and labor intense offline analytics can therefore be reduced to a minimum for model training and final product release. The developed approach is fundamental for the implementation of real-time batch release and continuous manufacturing in biopharmaceutical industry.

**References**

1. M. von Stosch, J. Glassey (Eds.), Hybrid Modeling in Process Industries, CRC Press, Taylor & Francis Group, Boca Raton, Florida, 2018.
2. M. von Stosch, R. Oliveira, J. Peres, S. Feyo de Azevedo, Comput. Chem. Eng. 60 (2014) 86–101.
3. S. Leweke, E. von Lieres, Comput. Chem. Eng. 113 (2018) 274–294.