Model Topology Identification and Parameter Estimation for Purification of BSA and Myoglobin

Guilherme C. Amaral,a,b Ana M. Ribeiro,a,b Idelfonso B.R. Nogueira,c Alexandre F.P. Ferreira,a,b,\*

aLaboratory of Separation and Reaction Engineering−Laboratory of Catalysis and Materials (LSRE-LCM), Department of Chemical Engineering, University of Porto, Porto, 4200-465, Portugal

bALiCE−Associate Laboratory in Chemical Engineering, Faculty of Engineering, University of Porto, Porto, 4200-465, Portugal

*cChemical Engineering Department, Norwegian University of Science and Technology, Sem Sælandsvei 4, Kjemiblokk 5, Trondheim, 793101, Norway*

\*aferreir@fe.up.pt (A.F.P. Ferreira)

Abstract

It is critical to use thorough parameter estimation methodologies to model protein separation processes since these can have very complex mechanisms. The goal of this study was to develop such a methodology and apply it to a particular case study. Hence, this work uses experimental data to present the parameter estimation and model topology identification of myoglobin and BSA separation in a size-exclusion gel SMB equipment. Particle Swarm Optimization and adequate objective functions were used to solve the parameters problem. Additionally, an estimability analysis and uncertainty evaluation were performed to comprehensively characterize the data.

This work was carried out in two steps. In the first step, the parameters were estimated for a fixed bed system. Then, the estimability analysis was carried out, revealing that the mass transfer coefficient parameters had negligible influence on the resulting output. Therefore, the parameters were excluded from further estimations. The second step involved estimating the parameters that generate a simultaneous adjustment of the mathematical model to the SMB and fixed bed experiments. Finally, the uncertainty evaluation was carried out successfully, yielding confidence intervals for each parameter, and a propagated uncertainty, with a value of 0.0522, for the generated model.

Ultimately, the success obtained by using this methodology in this simpler case infers the possibility of its use in the modelling of more complex protein separation processes, where the number of estimable parameters might be higher.

**Keywords**: Adsorption, Simulated Moving Bed, Parameter Estimation, Estimability.

* 1. Introduction

Proteins are present in all biological systems and are extremely important in biological reactions. The three-dimensional structure of the macromolecule and the sequence of amino acids determine the defining properties of each protein and, therefore, their function in the system (Whitford 2013). To study and apply these essential nutrients, protein separation and purification methodologies are critical. Chromatography is a very common choice to perform protein separation due to its versatility and adaptability to the different properties of the proteins (Liu, Li *et al.* 2020). In chromatography, the sample and the mobile phase where it is inserted are introduced into a carrier containing a stationary phase. Due to the varying affinities of the sample proteins to the stationary phase, different retention times will be observed, enabling the separation of the components (Lundanes, Reubsaet *et al.* 2013). Yet, some traditional chromatography techniques are costly. Simulated Moving Bed (SMB) is a prevalent alternative to perform economical chromatography since the countercurrent movement simulated by the equipment results in high productivity and low mobile and stationary phase consumption, reducing operation costs (Houwing 2003).

The development of mathematical models that characterize the adsorption of components is necessary to understand these types of separation better. Therefore, by solving a set of differential-algebraic equations that describe the key adsorption steps, it is possible to gain insight into the adsorption data (Xu, Cai *et al.* 2013). However, these models have parameters that must be identified by solving an optimization problem. For this purpose, deterministic methods guarantee convergence into an optimal solution; however, their effectiveness heavily depends on the initial vector given (Pinto and Schwaab 2007). Meta-heuristic approaches are also recommended due to their flexibility and speed, as they do not rely on initial guesses (Lin, Tsai *et al.* 2012), but do not necessarily reach the optimal solution in a given problem.

In this study, a meta-heuristic technique, Particle Swarm Optimization (PSO) was employed. PSO is described as a technique for continuous nonlinear function optimization, and its algorithm manipulates a defined set of particles to search for the optimal solution using not only the individual memory of each one but also the collective knowledge of the entire particle swarm (Kennedy and Eberhart 1995).

While carrying out a parameters’ estimation procedure, an estimability analysis might be helpful to evaluate the parameters' sloppiness. Local estimability analysis is a technique that employs sensitivity matrices and their orthogonalization to identify possible correlations between parameters and quantify their influence, enabling the streamlining of any given optimization problem (Yao, Shaw *et al.* 2003). This step is critical in parameter estimation since parameter correlation can inhibit successful estimation (Nogueira and Pontes 2017). On the other hand, incorporating uncertainty analysis becomes essential when dealing with parameter and variable measurements prone to errors. However, performing such studies requires a comprehensive knowledge of all the factors contributing to uncertainty (Iso and OIML 1995).

This work aimed to utilize available experimental data collected by Rios and coworkers (Rios, Ribeiro *et al.* 2020) to aid the development of a methodology that allows for the estimation of the mathematical model parameters that optimally describe the performed protein separation for a specific case study, where an SMB equipment, packed with Sephadex G-50, is used to purify these essential components. To achieve this goal, fixed bed experiments were also utilized to facilitate and further corroborate the estimations performed. A more complete description of the presented study can be found in the original document (Amaral 2023).

* 1. Materials, Methodology and Results

The tested proteins were Myoglobin (Mb) and Bovine Serum Albumin (BSA). Mb is a member of the globin protein family and is mainly responsible for binding and transporting oxygen in most vertebrates. Mb has a mass of approximately 17800 Da (Hendgen-Cotta, Kelm *et al.* 2014). BSA is an albumin protein derived from bovine origins and is one of the most available in the blood (Peters 1970). The BSA molecule consists of 563 amino acids, translating to a substantial mass of 66400 Da (Peters Jr 1995). Sephadex G-50, is an organic size exclusion gel with a sieve range spanning from 1500 to 30000 Da, enabling the separation of the aforementioned proteins (Rios, Ribeiro *et al.* 2020).

It was considered that the adsorption isotherm assumed by Rios and coworkers for the Sephadex G-50 separation, a linear isotherm, was the most suitable for the obtained separation data since size exclusion separation has a simpler mechanism, where the retention time of a component in the column is determined solely by its pore accessibility and therefore, no effective adsorption occurs.

* + 1. Particle Swarm Optimization

To ensure the correct application of a PSO algorithm, it is imperative to select an appropriate number of particles and a number of iterations to enable the convergence of the method without requiring excessive time and/or computational power. Furthermore, choosing suitable values for the acceleration coefficients is essential, so that convergence is achieved as efficiently as possible. The number of decision variables and their lower and upper position bounds must also be properly established to construct a feasible region for the optimization problem.

Optimization starts with an initialization step, randomly assigning a position to each particle. The next step involves calculating the objective function (OF) value. Subsequently, the best results, *i.e.*, the lowest OF values yielded by each particle and the overall swarm up to the current iteration, are updated and archived. Then, a value update for the acceleration coefficients follows. After that, velocities are determined while obliging the restriction $v\_{i,k,n}$ ≤ $v\_{max}$. If the constraint is violated, a new velocity that conforms to the rule is given. The particle position is then updated considering region boundaries. Once again, if boundaries are violated, the new positions will assume boundary values so that the rule can be followed. The optimization process is repeated for all particles and iterated until the maximum number of iterations has been achieved.

* + 1. Objective Function (OF)

The OFs to be minimized by the PSO algorithm were designed to guide the PSO algorithm in the search for parameter values that would minimize the residuals computed through a Mean Squared Error (MSE), between the generated mathematical model and the experimental points. Therefore, the lower the OF value, the better the alignment between the experimental data and the model defined by the estimated parameters.

* + 1. Fixed Bed Parameter Estimation

For the fixed bed experiments, the model employed was considered isothermal, and diluted mixtures were assumed, as well as the absence of interactions between solutes. The gel’s particles were considered spherical and uniform. Additionally, plug-flow with axial dispersion was considered for the movement of the fluid phase. The models used were adapted from Rios *et al.*'s studies (2020) and can be observed there. The set of estimable parameters for the fixed bed experiments were $θ\_{1}$ = [$ε\_{b},$ $D\_{ax}$, $Q\_{feed}$, $K\_{SEC, i}$, $k\_{h,i}$], with $ε\_{b}$ pertaining to the bed porosity, $D\_{ax}$ to the axial dispersion, $Q\_{feed}$ to the fixed bed feed flow, $K\_{SEC, i}$, to the size-exclusion constant for protein $i$ and $k\_{h,i}$ to the mass transfer coefficient for protein $i$. Using Particle Swarm Optimization with 100 particles and 100 iterations. This estimation yielded three sets of parameters, one for the myoglobin, one for the BSA, and one for the binary fixed bed experiments, that generated models with lower OF values than those obtained by Rios *et al*., and a close-to-optimal alignment to the experimental data. Figure 1 reveals a comparison for the myoglobin and BSA models.



Figure 1 – Graphical comparison between experimental data (points) and yielded mathematical models for the myoglobin (left) and BSA (right) fixed bed experiments

The estimated parameters yielded in this step would serve as reference parameters in further estimations.

* + 1. Estimability Analysis

The estimability analysis employed in this work is a sensitivity technique based on the Gram-Schmidt principle of orthogonalization. Each parameter has an assigned sensitivity vector. The selection process involves identifying the parameter with the highest sensitivity, after which the remaining vectors are represented orthogonally in the plane perpendicular to the chosen parameter's sensitivity vector. This will be iterated until all parameters are selected or until the sensitivity of the largest parameter falls below a predefined cut-off value (Kravaris, Hahn *et al.* 2013). This analysis enables the removal of correlated parameters or parameters with very low influence in the generated model, simplifying the optimization problem. The mass transfer coefficients, $k\_{h,i}$, consistently fell below the established cut-off value in each iteration in all experiments, meaning that the influence of this parameter in the output is negligible. Fixed values were therefore given to the mass transfer coefficients equal to those obtained by Rios and coworkers.

* + 1. SMB Parameter Estimation

For the SMB experiments, the already available fixed bed models were adapted to represent the SMB columns while the necessary mass balances to the nodes were considered. Furthermore, the model included the necessary global balances and the structure of the additional equipment involved. Dead volumes of the system were also considered. The models can be observed in Rios *et al.* work(2020). The set of chosen parameters for this estimation were $θ\_{2}$ = [$ε\_{b},$ $D\_{ax}$, $Q\_{feed}$, $K\_{SEC, i}$, $ε\_{b}\_{j}^{SMB}$, $Pe\_{j}$, $Q\_{F}$, $Q\_{D}$, $Q\_{X}$, $Q\_{IV}$] with $ε\_{b}\_{j}^{SMB}$concerning the bed porosity of column $j$, $Pe\_{j}$ the Peclet number of column $j$, and $Q\_{F}$, $Q\_{D}$, $Q\_{X}$, $Q\_{IV}$ are the flowrates for the feed, desorbent, extract, and recycling streams, respectively. This study consisted of estimating the parameters set that simultaneously described the SMB experiment and fixed bed experiments, therefore, fixed bed parameters were also part of the estimation procedure. A Particle Swarm Optimization algorithm using 50 particles and 35 iterations was employed. The yielded set of parameters revealed very similar values to the parameters obtained by Rios and coworkers for the SMB experiment, except the axial dispersion coefficient. Nevertheless, a lower objective function value was obtained for the estimated parameter set when compared to Rios’s set of parameters. Figure 2 reveals a comparison between the SMB experimental data and the mathematical models.



Figure 2 - Graphical comparison between SMB experimental data (points) and the models with *Rios et al.* parameters and estimated parameters at the extract (left) and raffinate (right)

An absolute and relative comparison was then performed between the parameters obtained during this task and reference parameters (reported in 2.3). Some parameters revealed deviation from their reference counterpart, particularly the axial dispersion coefficient in all cases and the BSA size exclusion constant for the BSA and binary fixed bed experiments. Aside from a noticeable deviation in the myoglobin case, due to the nature of these parameters, specifically their low absolute value, the resulting mathematical models do not change significantly compared to the ones generated by reference parameters.

* + 1. Uncertainty Analysis

The uncertainty of the parameters was evaluated based on the outcomes yielded by the fixed bed + SMB parameters estimation. However, only the PSO particles that fulfilled the Fisher-Snedecor test were utilized in the uncertainty evaluation. Confidence intervals were obtained for each parameter, assuming Gaussian distribution for the PSO particles. As expected, the values obtained for each parameter during estimation are inserted into these intervals. The propagated uncertainty to the output model was also determined, yielding a value of 0.0522.

* 1. Conclusions

This work developed a methodology to estimate parameters that characterize the equations that describe the separation of BSA and myoglobin in fixed bed and SMB. The methodology consists of creating reference parameters for the fixed bed experiments and obtaining preliminary information about the separation mechanisms, which culminates in a simultaneous SMB and fixed bed parameter estimation.

The parameter estimation for the individual fixed bed experiments, yielded a set of reference parameters for each experiment, with close-to-optimal alignment to the experimental data. The following estimability analysis determined that the parameter mass transfer coefficients for both myoglobin and BSA do not influence the final model in the chosen parameter boundaries. Therefore, their estimation was considered irrelevant and was removed from further studies. The parameter estimation using all the available fixed bed and SMB experiments resulted in a set of parameters that created a model with a simultaneous adjustment to the SMB and the fixed bed experimental data. Using the reference values obtained during the individual fixed bed studies, it was possible to observe slight differences in some parameters, namely the axial dispersion coefficient and size exclusion constant for BSA. Nonetheless, these variations only poorly affected the myoglobin fixed bed experiment, while the BSA and binary experiments do not reveal significant changes in their respective generated models. The uncertainty of each estimated parameter was computed to obtain confidence intervals and characterize the results obtained. The propagated uncertainty into the generated model yielded a value of 0.0522.

The methodology considered and used in this work proved to be a successful course of action to estimate parameters in fixed bed and SMB environments for protein separation processes. Moreover, it revealed to be a reliable technique to validate the results obtained.

References

Amaral, G. (2023). Model topology identification and global parameters estimation for purification of BSA and myoglobin.

Hendgen-Cotta, U. B., M. Kelm and T. Rassaf (2014). "Myoglobin functions in the heart." Free Radical Biology and Medicine **73**: 252-259.

Houwing, J. (2003). "Separation of proteins by simulated moving bed chromatography."

Hunt, B. J. and S. R. Holding (2013). Size exclusion chromatography, Springer Science & Business Media.

Iso, I. and B. J. G. OIML, Switzerland (1995). "Guide to the Expression of Uncertainty in Measurement." **122**: 16-17.

Kennedy, J. and R. Eberhart (1995). Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks.

Kravaris, C., J. Hahn, Y. J. C. Chu and c. engineering (2013). "Advances and selected recent developments in state and parameter estimation." **51**: 111-123.

Lin, M.-H., J.-F. Tsai and C.-S. Yu (2012). "A Review of Deterministic Optimization Methods in Engineering and Management." Mathematical Problems in Engineering **2012**: 756023.

Liu, S., Z. Li, B. Yu, S. Wang, Y. Shen and H. Cong (2020). "Recent advances on protein separation and purification methods." Advances in Colloid and Interface Science **284**: 102254.

Lundanes, E., L. Reubsaet and T. Greibrokk (2013). Chromatography: basic principles, sample preparations and related methods, John Wiley & Sons.

Nogueira, I. B. R. and K. V. Pontes (2017). "Parameter estimation with estimability analysis applied to an industrial scale polymerization process." Computers & Chemical Engineering **96**: 75-86.

Peters Jr, T. (1995). All about albumin: biochemistry, genetics, and medical applications, Academic press.

Peters, T. (1970). Serum Albumin. Advances in Clinical Chemistry. O. Bodansky and C. P. Stewart, Elsevier. **13:** 37-111.

Pinto, J. C. and M. Schwaab (2007). Análise de Dados Experimentais: I. Fundamentos de Estatística e Estimação de Parâmetros, Editora E-papers.

Rios, A. G., A. M. Ribeiro, A. E. Rodrigues and A. F. P. Ferreira (2020). "Bovine serum albumin and myoglobin separation by size exclusion SMB." Journal of Chromatography A **1628**: 461431.

Whitford, D. (2013). Proteins: structure and function, John Wiley & Sons.

Xu, Z., J.-g. Cai and B.-c. Pan (2013). "Mathematically modeling fixed-bed adsorption in aqueous systems." Journal of Zhejiang University SCIENCE A **14**(3): 155-176.

Yao, K. Z., B. M. Shaw, B. Kou, K. B. McAuley and D. W. Bacon (2003). "Modeling Ethylene/Butene Copolymerization with Multi‐site Catalysts: Parameter Estimability and Experimental Design." Polymer Reaction Engineering **11**(3): 563-588.