# **18th International Conference on**

# **Polysaccharides-Glycoscience 2022**

**DEVELOPMENT OF SOFTWARE TOOL FOR ESTIMATION OF CHROMATOGRAPHY EQUILIBRIUM DISPERSIVE MODEL PARAMETERS FOR SEPARATION OF SACCHARIDES AND MANNITOL**

**TOMAS SVOBODA1, ADAM SVOBODA2, SVATOPLUK HENKE1, SIMONA GILLAROVÁ1, DANIEL RŮŽIČKA1 and PAVLA KRČOVÁ1**

*1University of Chemistry and Technology, Prague, Faculty of Food and Biochemical Technology, Department of Carbohydrates and Cereals, Technická 5, 166 28 Prague 6 – Dejvice*

[*tomas3.svoboda@vscht.cz*](mailto:tomas3.svoboda@vscht.cz)

*2Czech Technical University in Prague, Faculty of Information Technology, Thakurova 9, 160 00 Prague 6, Czech Republic*

**Abstract**

Preparative chromatography using ion exchange resin is an efficient method for the separation of monosaccharides. Continuous chromatography with a simulated moving bed is a popular technology for industrial production of pure chemicals because of its efficiency in terms of sorbent utilization and eluent consumption. To find the optimal operating parameters for the SMB chromatography system, a mathematical description of the chromatography process is necessary.[7],[8] Therefore, the software tool utilizing functionalities needed for the determination of chromatography model parameters based on various and messy experimental data has been developed. The work proposes data preprocessing steps, partial differential equation solver and formulation of parameters estimation task as a bilevel optimization problem. This leads to an efficient workflow which accelerates scaleup of new preparative chromatography applications.

**Mathematical Model**

The Equilibrium Dispersive Model (EDM) extends the simple equilibrium model published by Don De Vault in 1943.[2] He described chromatography as a concentration wave propagation in a fixed bed column with assumptions of diluted systems, plug flow, negligible pressure-drop, and isothermal operation. The addition of a second-order derivative apparent diffusion (dispersive) term is useful to model nonlinear behavior which manifests itself, for example, in peak asymmetry. The mass balance of the component *i* in a volume element of the column is

|  |  |  |
| --- | --- | --- |
|  |  | *equation 1* |

Combining with the linear adsorption equilibrium isotherm.[8]

|  |  |  |
| --- | --- | --- |
|  |  | *equation 2* |

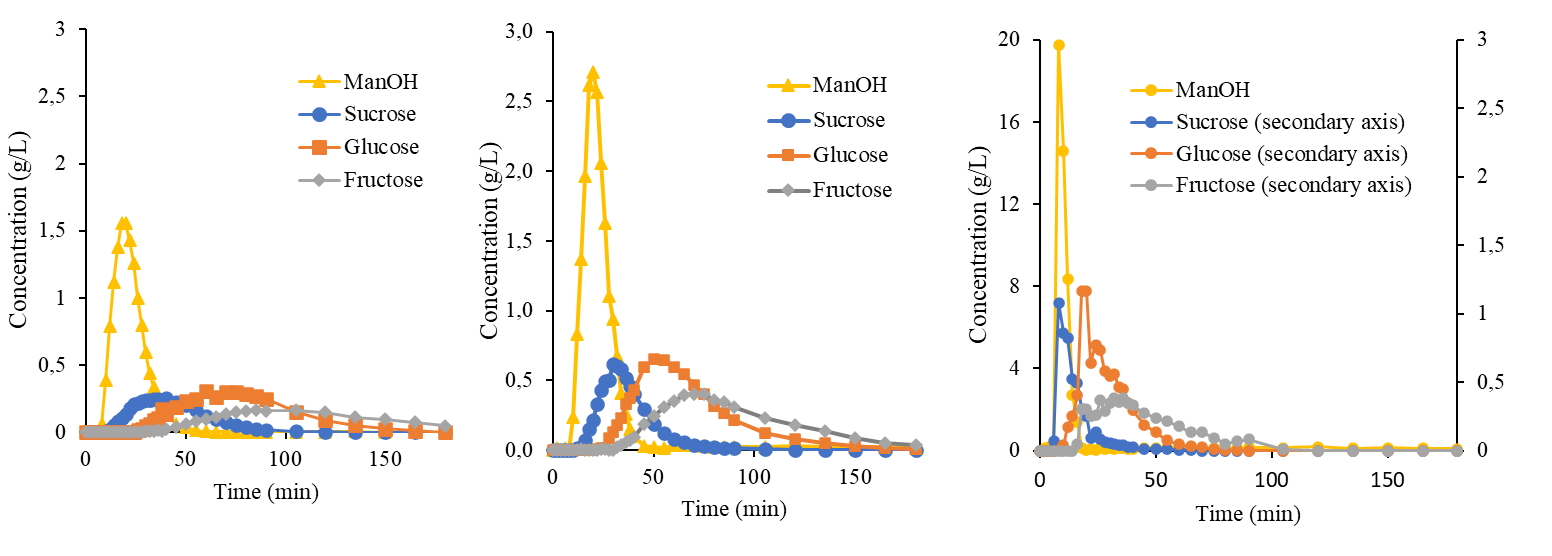
This study presents the original measured data from discontinuous preparative chromatography pulse feed experiments and introduces the methods of preprocessing of the data and way to utilize them to determine the parameters of the EDM with non-competitive linear isotherm. [7]

**Experimental Methods**

The mixture of mannitol, glucose, fructose, and sucrose was separated during the experiments. The experimental setup consisted of the column XK16/40 (GE, USA) with a commercially accessible Dowex 550A anion exchange resin (Dow, USA) with a layer height of 23.5 cm and a diameter of 1.6 cm. The discontinuous station also consisted of a diaphragm dosing pump and a three-way electromagnetically controlled valve used for pulse injection of the mixture. The collected fraction samples were analyzed on HPAEC (High-Performance Anion Exchange Chromatography) Dionex ion-exchange liquid chromatograph with pulsed amperometry detection. The concentrations of the components in the individual fractions were determined by integration of the peaks and comparison with the standards.

**Experimental data**

This work utilizes data from 3 experiments with different conditions. The model solution of sucrose, galactose, fructose, and mannitol has been separated. The concentrations in the injection pulse feed have been 9.09 g/L of sucrose, 18.78 g/L glucose, 20.48 g/L of fructose, and 24.38 g/L of mannitol. [5]

****

*Form 1 – Data from chromatography separation experiments: a) Feed volume 0.85 mL, flowrate 150 mL/h;*

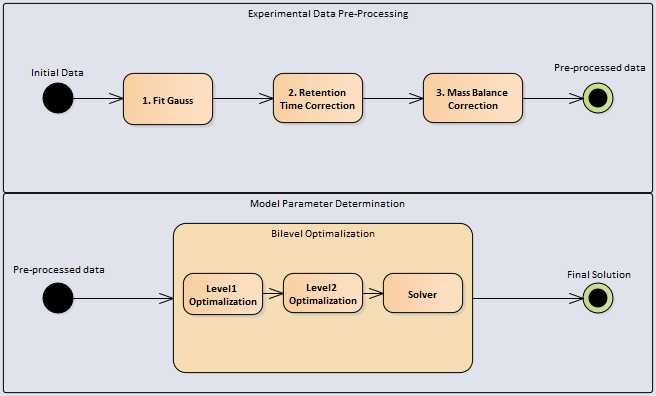
*b) Feed volume 2.52 mL, flowrate 150 mL/h; c) Feed volume 2.40 mL, flowrate 300 mL/h [5]*

**Software tool concept**

The workflow consists of two main parts, preprocessing the experimental data and determining the model parameters based on optimalization. The pre-processing is done in three steps:

1. Fitting an asymmetric Gaussian curve to the measured data of each component in each experiment.
2. Correcting retention time by shifting data points of the same components in similar experiments so the peak time matches. Shifting is done so that the times of the peaks are on average of the peak times.
3. Correcting the mass balance by integrating the time series and calculating the total eluted mass of the component, comparing it with the mass injected during feed and adjusting the feed time to match the values.

The determination of model parameters utilizes 2 levels of optimization and the self-developed PDE solver. Optimization is based on defined loss functions that evaluated the difference between experimental and model concentrations in the column outlet.



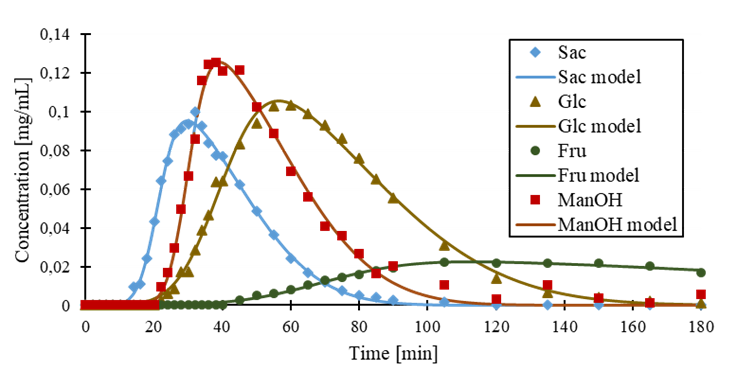
*Form 2 – Measured data point and fitted asymmetric Gaussian model curves*

**Data Preprocessing**

Experimental data points are loaded with errors arising from sample pipetting, integration of analytical peaks, etc. Therefore, preprocessing algorithms have been proposed. The first step is to replace the original measured datapoints with the fitted asymmetric Gauss curve [6].

The second step executes the retention time alignment. This algorithm is feasible when considering two or more experiments with the same conditions (flowrate, column dimensions, etc.). In that case, it calculates the average time of the peaks maxima and shifts all the data points to reach the lowest difference between the average maxima values and the actual maxima values of all

corresponding components.



*Form 3 – Measured data point and fitted asymmetric Gaussian model curves*

Optimization problem arises where is time of peak maximum for component *i* in experiment *j*. is average time of maxima of the peaks from all experiment for component *i*.

|  |  |  |
| --- | --- | --- |
|  |  | *equation 3* |
|  |  | *equation 4* |

The third step of the pre-processing workflow is the correction of feed duration according to total mass balance in the feed and in the end of the column. The total mass balance in the feed is expressed as

|  |  |  |
| --- | --- | --- |
|  |  | *equation 5* |

Where is concentration of component i in the feed and *d* is column diameter. The total mass at the end of the column is expressed as

|  |  |  |
| --- | --- | --- |
|  |  | *equation 6* |

Where is duration of the experiment. Subsequently, the optimization problem arises.

|  |  |  |
| --- | --- | --- |
|  |  | *equation 7* |

**Numerical solution**

After substituting linear isotherm (equation 2) into equation 1 it leads to the following expression:

|  |  |  |
| --- | --- | --- |
|  |  | *equation 8* |

The resulting partial differential equation can be qualified as semi-linear and parabolic. Boundary conditions are Dankwerts’ type and expressed as [8]:

|  |  |  |
| --- | --- | --- |
|  | Left boundary:  Right boundary: | *equation 9*  *equation 10*  *equation 11* |

The initial conditions represent an empty column.

|  |  |  |
| --- | --- | --- |
|  |  | *equation 12* |

The problem can be solved numerically, for example, with implicit finite difference method. In this work, the Crank-Nicolson method was used. Discretization in time and special domains has been done as follows [1]:

Interval in dimension *x* is divided into *n* equal parts: ,

Interval in dimension *t* is divided into *r* equal parts: ,

The Crank-Nicolson scheme was used with centered differential formula and forward/backward formula for boundaries. The substitution of derivatives with formulas leads to the following system of linear equations, which is obtained in one time step:

|  |  |  |
| --- | --- | --- |
|  |  | *equation 13* |

Where are known and is solution vector of next time step.

|  |  |  |
| --- | --- | --- |
|  |  | *equation 14* |

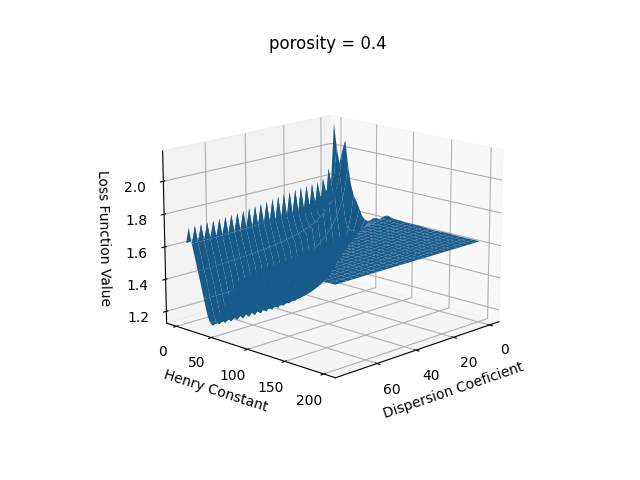
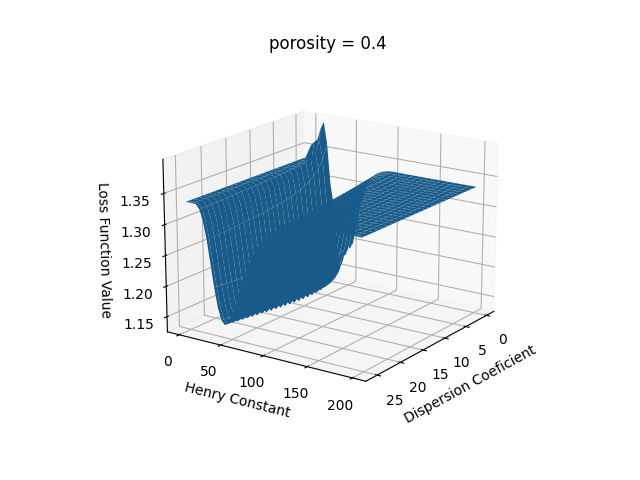
As part of the work, the numerical solver has been designed and implemented using Python script. The solver is specified and optimized for the purposes of EDM with linear isotherm. Matrix A is strictly diagonally dominant (condition for ensured convergence of iterative method) and sparse, allowing us to use its band form and *scipy.linalg.solve\_banded* function, which significantly improved the code performance compared to the traditional Jacobi’s method. Since the method is purely implicit the solution is unconditionally stable [1].

**Definition of Error functions and Sensitivity analysis**

To quantify how well the model solutions fit into the experimental data it was necessary to define loss value which is calculated by loss function. In this work, standard loss value quantification by sum of squared errors has been used, which seems to be a better fitting for parameter estimation (Form 2). The total loss function has been defined as the sum of squared errors for each experiment of each component. To ensure that loss function values are comparable between experiments with different numbers of data points and the sum of loss values is not highly correlated just to one component with the most datapoints, the value is divided by the absolute number of datapoints and *k* refers to a specific point in the experimental data set.

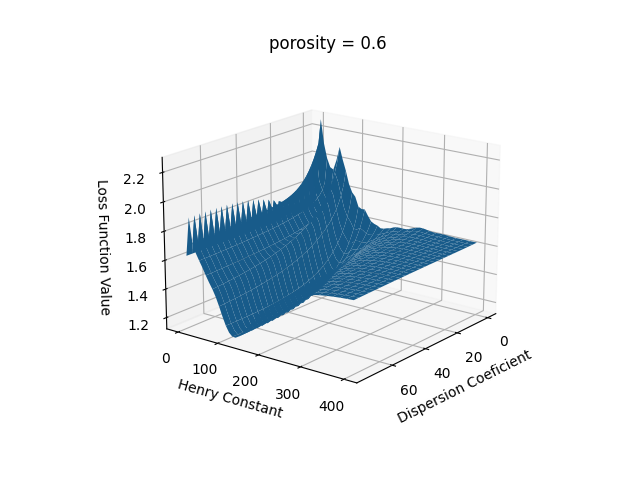
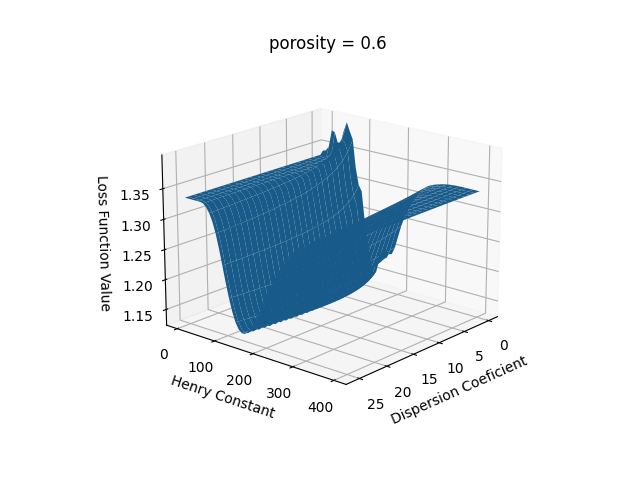
|  |  |  |
| --- | --- | --- |
|  |  | *equation 15* |

As can be seen in Form 2, the isotherm constant is much more correlated with the loss function value than the dispersion coefficient. When we are dealing with data from just a single experiment, it seems that the loss function has just one global minimum and if the initial guess is not on the plateau (area where the retention time is far after the defined timespan), we can use rapidly converging gradient methods such as Newton’s [1]. However, when we add loss values from additional experiments to the total value of loss function, complexity of an optimization problem increases. We can observe that there are some local minima starting to form.



b)

a)



d)

c)

*Form 4 – Loss function value surface (logarithmic scale) in different constant porosities and experimental data: a) and c) are for single experiment dataset, b) and d) are for complete dataset (all 3 experiments)*

**Parameters Estimator**

Estimation of parameters is an optimization problem, where we try to global find minimum of the sum of loss function values of all the components. Naturally, it is of high importance for minima formation how the loss function is defined. Since the model assumes no competition for active groups of the resin between the components, each of them has its own isotherm constant and dispersion coefficient. Porosity is, on the other hand, a common property of the resin, which is assumed to be the same for all the components. Based on this fact, the parameter estimation has been defined as a bilevel optimization problem:

|  |  |  |
| --- | --- | --- |
|  |  | *equation 16* |

To ensure that loss function values are comparable between components with different concentrations and that the resulting sum of loss values is not highly correlated just with component with highest concentrations, we have introduced the relativization step of the loss values based on maximum concentration in the column outlet and number of datapoints.

To find a global minimum in the lower variable space (level 2), the general-purpose algorithm SHGO (simplicial homology global optimization) has been utilized. For the find of minima in upper variable space (level 1), the simplex Nelder-Mead method has been called.[3],[4]

**Results and Discussion**

The work proves that the developed software tool can be used for estimation of the parameters of the linear equilibrium dispersive model. Data preprocessing workflow utilizing fitting of asymmetric Gauss curves, correction of retention times, and mass balance rapidly improves reliability of the software tool and enables utilization of data from numerous experiments in one optimization step. Sensitivity analysis has proven that defined loss function is continuous and has only one global minimum for single dataset, but when data from more experiments are combined several local minima are formed. Therefore, the SHGO optimization algorithm has been proposed as a suitable tool for the solution of the problem. The results of the parameter estimation based on 3 experiments are shown in the following table:

|  |  |  |  |
| --- | --- | --- | --- |
| **Component** |  | **[mm2·s-1]** | **[1]** |
| Sucrose | 11.65 | 13.84 | 0.34 |
| Glucose | 20.95 | 14.61 |
| Fructose | 32.25 | 28.39 |
| Mannitol | 4.89 | 111.52 |

*Table 1 – Results of parameter estimation with developed software tool*

Mathematical modelling of preparative chromatography is a necessary step in development of new industrial applications. The reliable, although relatively simple, model is a crucial prerequisite for successful transition from laboratory to industrial scale. Developed software significantly accelerates the scale-up efforts for the cases where linear equilibrium dispersive model of chromatography is about to be used.

*Acknowledgement*

*„This work was supported from the grant of Specific university research – grant No A1\_FPBT\_2022\_008“*

**References**

1. Chapra S. C. & Canale R. P. (2006). Numerical methods for engineers (5th ed.). McGraw-Hill Higher Education.
2. Devault, D. (1943). The Theory of Chromatography. Journal of the American Chemical Society, 65, 532-540.
3. Endres, SC, Sandrock, C, Focke, WW (2018) “A simplicial homology algorithm for Lipschitz optimisation”, Journal of Global Optimization.
4. Gao, F. and Han, L. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. 2012. Computational Optimization and Applications. 51:1, pp. 259-277
5. Gillarová S., Henke S., Svoboda T., Kadlec P., Hinková A., Bubník Z., Pour V., Sluková M. (2021): Chromatographic separation of mannitol from mixtures of other carbohydrates in aqueous solutions. Czech J. Food Sci., ISSN: 1805-9317 (Online) ISSN: 1212-1800 (Print).
6. Green, M. and X. Chen (2020): Data Functionalization for Gas Chromatography in Python. Journal of Chemical Education 97(4): 1172-1175
7. Nicoud R.M. (2015): Chromatographic Processes: Modeling, Simulation, and Design. Cambridge University Press.
8. Rodrigues A.E., Pereira C., Minceva M. et al. (2015): Simulated Moving Bed Technology. Butterworth-Heinemann, Oxford.