

# Uncertainty analysis in mass transfer parameter estimations for chromatographic separation of glucose and fructose

**A. Prieto-Moreno, O. Llanes-Santiago**

Automatic and Computing Department, CUJAE

19390 Havana, Cuba

E-mail: albprieto@electronica.cujae.edu.cu, orestes@electronica.cujae.edu.cu.

**L.D.T Câmara, A.J. Silva Neto, C. Oliveira**

Instituto Politécnico, UERJ - Universidade do Estado do Rio de Janeiro

28601-970 Nova Friburgo, RJ, Brasil

dcamara@iprj.uerj.br, ajsneto@iprj.uerj.br, E-mail: coliveira@uerj.br

**Abstract:** *In this paper, a statistical approach for the analysis of the propagation of uncertainty is shown, in the estimate of the kinetic parameters of mass transference used to model a chromatographic column in Simulated Moving Bed. The modeling of the chromatography column was accomplished intervening the new approach front velocity. The analysis of how it is propagated the operational factors uncertainty involved in the process of chromatography toward the estimated parameters was carried out by the use of response surface methodology. Furthermore, chromatographic regions where factors cause bigger variation in the output and their respective patterns were determined. The analysis was applied to the separation process of glucose and fructose.*

**Keywords:** *uncertainty propagation, parameter estimation, response surface, chromatographic column, front velocity, factorial experimental design*

## Introduction

The characterization of a chromatographic column is an important factor in determining the properties of mass transport, which are used in the design of industrial units such as the process Simulated Moving Bed (SMB). Modeling and simulation of chromatographic systems leads to an understanding of the main mechanisms of mass transfer, and the operating conditions that can be used to improve the separation/purification of the molecules.

The application of inverse problems methodology to chromatography has proven to be important in the optimization studies as well as in the determination of mass transfer parameters with higher accuracy as attested by several research involving the study and analysis of adsorption chromatographic systems [4, 10].

The uncertainty of a measurement or estimate is defined as “a parameter, which is associated with the result of a measurement or estimation and characterizes the dispersion of values”. The result of a measurement or estimation is considered the best estimate of the true value, and all sources of uncertainty influence in its propagation. Therefore, the result can't be interpreted properly without any knowledge about its uncertainty. The quantification of uncertainty will allow to establish the confidence intervals for the estimated parameters, which can be very important from an engineering point of view.

The objective of this work is to analyze how the uncertainty in the values of some variables that are involved in the characterization of a chromatographic column, propagates in the estimation of kinetic parameters of the model front velocity. For this purpose, an alternative

procedure to the the use of the Monte Carlo method [7, 5] is applied, based on the use of response surface [9] for obtaining the probability distribution function (pdf) of the estimated parameters. This proposal has the advantage, over the Monte Carlo method, that require shorter compute time and to offer direct interpretation of how acting factors uncertainty has influence over the variation of the estimated parameters.

## 1 Front velocity modelling approach of chromatography

The chromatographic column model determines the performance in the final separation of the SMB, depending on the number of interconnected columns. In general, different research groups employ dispersion models [6] to represent the chromatographic columns. These models are robust and efficient, but need deep numerical treatment of the partial differential equations which requires high computational time.

In this work it is used a new modeling approach, known as the front velocity convection [2, 3]. In this approach, the convection of the liquid phase is regarded as the main phenomenon in the transport of molecules along the chromatographic column, followed by the mass transfer between the solid adsorbent and the liquid phase.

### 1.1 Front velocity

Due to the fact that the flow rate driven into the column is determined by an external pumping system, the time that the liquid phase needs to go over the length of the chromatographic column can be determined if the volumetric flow rate, the porosity, and the column volume are known experimentally. To model the mass transfer, two transfer models of concentrated mass are assumed, described by the Equations (1) and (2), where  $C$ ,  $q$ ,  $k_1$  and  $k_2$  are respectively, the concentration in the liquid and solid phases, the maximum capacity of adsorption and the global mass transfer kinetic constants of adsorption and desorption.

$$\frac{dC}{dt} = -k_1C + k_2q \tag{1}$$

$$\frac{dq}{dt} = -\frac{dC}{dt} \tag{2}$$

The simulation from the front velocity modeling approach was compared to the experimental data available in [1]. Figure 1 shows a comparison between the experimental results and the data obtained by simulation, for which the following parameters were used: flow-rate = 30ml/min, porosity = 0.4, injec-vol. = 300ml, and injec-conc. = 15mg/ml.

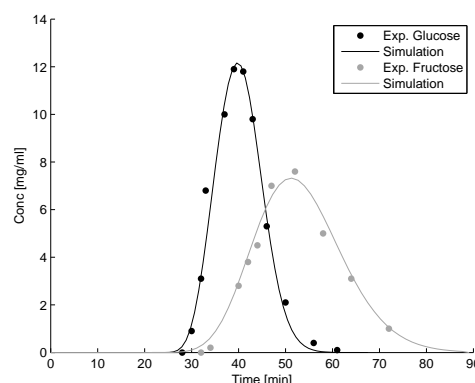


Figure 1: Comparison between simulation results (lines) and experimental adsorption data (points) of glucose and fructose.

## 1.2 Parameter estimation

In order to estimate the unknown parameters of the model, the inverse problem is formulated implicitly like an optimization problem, where it is wanted to minimize the squared function of the residuals

$$S(\vec{K}) = [\vec{C}_{exp} - \vec{C}_{cal}(\vec{K})]' [\vec{C}_{exp} - \vec{C}_{cal}(\vec{K})] = \vec{R}'\vec{R} \quad (3)$$

where  $\vec{C}_{exp}$  is the vector of experimental solute concentration values,  $\vec{C}_{cal}$  is the vector of calculated values,  $\vec{K} = (k_1, k_2)'$  is the vector of unknown parameters to be determined, and  $\vec{R}$  is the vector of residues .

The optimization method used to solve the inverse problem was Simulated Annealing [12]. To avoid a great delay in the estimation of the parameters, an iterative approach was used, in which the algorithm is configured to perform just a few cycles and then restart with the result as an initial condition.

## 2 Uncertainty propagation in a chromatographic column

### 2.1 Uncertainty assessment by means of simulation

The analysis of uncertainty propagation by an analytical method may involve a complicated mathematical manipulation, for which an analytical solution is possible only through approximations. A method that avoids the algebra and calculus, is to use numerical simulation techniques, for which only the system or process model and knowledge of the distribution or range of uncertainties are required. With this approach, given the values of the input variables, the outputs are calculated repeatedly with small changes in the inputs. After sufficient replicates, the distributions of the outputs take the correct form, from which it can determine the mean and standard deviation, e.g. In the design of the simulations it must be taken into account all possible factors of variation.

The input variables that were considered in this work are: flow-rate, porosity, injection volume, and injection concentration; which have also been considered in other papers [8]. For each variable variations up to 10% were defined, as shown in [8].

### 2.2 Response Surface Methodology

As mentioned previously, a widely used method for this type of analysis is the Monte Carlo method. However, for the process under study, this approach is not recommended due to the required number of executions of the parameter estimation method to obtain their pdf. In this work it is used the response surface methodology approach, consisting in a group of mathematical and statistics techniques used to develop a relational function between an answer of concern,  $y$ , and a number of associated control (or input) variables denoted by  $x_1, x_2, \dots, x_k$  [9]. In general, such a relationship is unknown but can be approximated by a low-degree polynomial model of the form

$$y = f'(\mathbf{x})\boldsymbol{\beta} + \epsilon \quad (4)$$

where  $\mathbf{x} = (x_1, x_2, \dots, x_k)$ ,  $f(\mathbf{x})$  is a vector function of  $p$  elements that consist of powers and cross-products of powers of  $x_1, x_2, \dots, x_k$  up to a certain degree denoted by  $d(\geq 1)$ ,  $f'$  means transpose function,  $\boldsymbol{\beta}$  is a vector of  $p$  unknown constant coefficients, and  $\epsilon$  is a random experimental error assumed to have a zero mean. The quantity  $f'(\mathbf{x})\boldsymbol{\beta}$  represent the mean response (expected value of  $y$ ), denoted by  $\mu(\mathbf{x})$ .

In general, two basic models are used [9], which are special cases of Equation (4): the first-degree model ( $d = 1$ )

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \epsilon, \quad (5)$$

and the second-degree model ( $d = 2$ )

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \sum_{i=1}^k \beta_i x_i^2 + \epsilon. \tag{6}$$

In general, at any point,  $\mathbf{x}$ , in a experimental region, denoted by  $\mathcal{R}$ , the predicted response is  $\hat{y}(\mathbf{x}) = f'(\mathbf{x})\hat{\beta}$ ,  $\mathbf{x} \in \mathcal{R}$ .

### 2.3 Factorial design

Obtaining the response surface involves the study of the effect of two or more variables, called factors in this context. In general, the factorial design, used as an approach to obtain the response surface design, is the most efficient for this goal.

The factorial design have several advantages: 1) it allows to analyze the effect of each factor individually and the interaction among them; 2) it allows the effects of a factor to be estimated at several levels of the other factors, yielding conclusions that are valid over a range of experimental conditions; and 3) need less experiment runs [11].

## 3 Outcomes and discussion

To illustrate the uncertainty propagation analysis, the mass transfer kinetic parameters estimation problem in a chromatographic column through front velocity is used. Table 1 shows the values of the variables selected for the analysis and a 10% of deviations from the respective nominal experimentation values.

Variable	Rang	Eng Unit
Flow-rate (A)	27-33	ml/min
Porosity (B)	0.36-0.44	
Injection vol. (C)	270-330	ml
Injection conc. (D)	13.5-16.5	mg/ml

Table 1: Operational factors for the uncertainty analysis in model parameter estimation.

Figure 2 shows how the flow-rate variation in the process output influences. The plots in 2a are obtained solving the direct problem for the flow-rate values used in the factorial design, and show the different concentration profiles over the time obtained. The absolute value concerning the difference among the profiles obtained for the maximum and minimum values, denoted by  $\delta$ , represents the output uncertainty region generated by the flow-rate variation. These uncertainty regions are shown in the graphs in 2b.

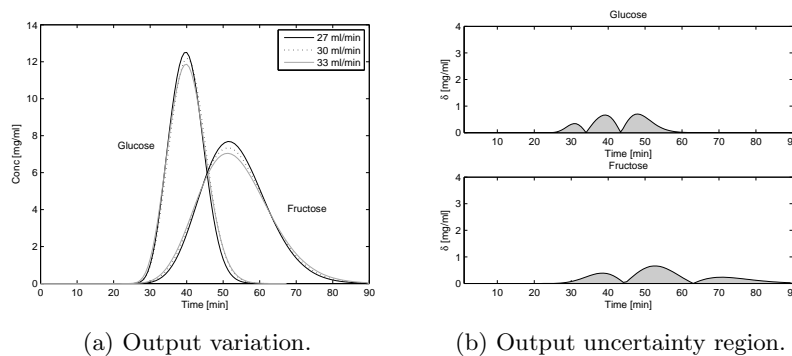
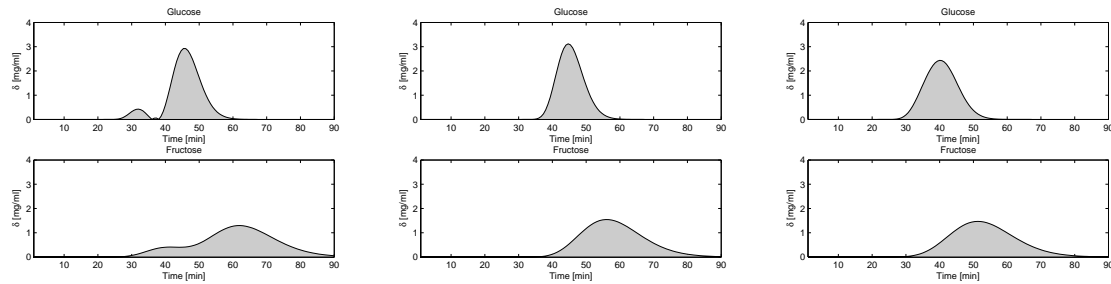


Figure 2: Flow-rate variations influence.

Figure 3 shows the output uncertainty regions generated by the variation at the remainders factors of concern. As can be seen, the concentration uncertainty is variable over the time, and the dispersion due to each factor causes a different uncertainty pattern. This characterization of the uncertainty allows to offer a support to the experimenters, of which are the regions of greater variation caused by changes in the factors involved in the chromatography.



(a) Uncertainty due to the porosity. (b) Uncertainty due to the volume. (c) Uncertainty due to the concentration.

Figure 3: Output uncertainty region generated by the variation at the other factors.

### 3.1 Estimation of the confidence interval

For the response surface design in this work it will be used the 2k factorial analysis. This design leads to the fact that the  $\hat{\beta}$  elements are uncorrelated, and their variance have minimum values. This means that the design provides maximum precision for estimating the unknown parameters in model Equation (5). For each of the experiments indicated by the analyzes two replicas were carried out.

To determine the influence of the factors in the estimation of kinetic mass transfer parameters, an analysis of variance (ANOVA) is performed. The  $F$  statistic indicates that most of the factors and their combinations have a significant effect on the  $k_1$  estimation. However, this test is sensitive to the number of replicates carried out for each experiment, since the degrees of freedom of the statistical test depend on the number of replicates performed. This implies that results that are identified as significant for a number of replicas, may not be for a different number of these. Therefore, the percent of contribution to the behavior of the studied variable is taken as a selection criteria to be included in the response surface generator polynomial. Which presents no major changes when varying the number of experiment replicates.

To select the factors that will be used in the polynomial generator, to incorporate those who contribute more than 2% the behavioral output is taken as a criterion.

For the glucose, the factors incorporated to the polynomial generator determine the 96.92% of the  $k_1$  parameter variation. Once selected the factors of greater effect, it is proceeded to construct the multiple regression model that relates these factors with the results of the  $k_1$  estimation. The regression coefficients of the model are obtained applying the Least Squares method, involving each  $k_1$  estimated value with the respective values of the factors used in the direct problem solution. The polynomial generator obtained was

$$\hat{k}_1 = (18.16 - 1.27A + 2.08C + 3.59D) \cdot 10^{-3} \text{ min}^{-1} \tag{7}$$

The influence of factors of concern on the  $k_1$  estimate variations is more homogeneous in the case of the fructose. This homogeneity implies that it is necessary to incorporate a larger amount of these to the response surface polynomial generator. For this case, it is necessary to incorporate the effect of the four main factors and the combined effect (AB), to explain the

96.97% of the variation in  $k_1$ . The polynomial generator obtained is:

$$\hat{k}_1 = (13.23 - 1.45A - 0.71B + 1.59C + 1.23D + \dots \cdot 0.41AB) \cdot 10^{-3} \text{ min}^{-1} \tag{8}$$

Following the same procedure for  $k_2$ , the polynomial generator obtained for glucose and fructose are, respectively:

$$\hat{k}_2 = (30.13 - 2.03A + 2.41C + \dots \cdot 6.19D) \cdot 10^{-3} \text{ min}^{-1} \tag{9}$$

$$\hat{k}_2 = (11.17 - 1.20A - 0.51B + 1.22C + 1.15D + \dots \cdot 0.31AB) \cdot 10^{-3} \text{ min}^{-1} \tag{10}$$

which explain the 97.35% and 97.13% of the  $k_2$  variations.

For all cases, the indicated statistical test to determine the existence of quadratic curvature on the response surface was carried out, obtaining as result that there is no evidence of quadratic effects; therefore, it can be concluded that a first-degree surface is appropriate.

After obtaining the polynomials generators that describe the estimated parameters behavior to variations in the factors of concern, random values within the defined range for each factor are assigned and the pdf for each one of the mass transfer kinetic parameters is esteemed.

Table 2 shows the pdf that characterize the variability of the estimated parameters of the kinetic model. The specified parameter intervals correspond to a 95% confidence.

Substance	Param	PDF	Interval
Glucose	$k_1$	$\mathcal{N}(0.0183, 0.588 \cdot 10^{-5})$	$[18.298, 18.321] \cdot 10^{-3}$
	$k_2$	$\mathcal{N}(0.0304, 1.494 \cdot 10^{-5})$	$[30.345, 30.403] \cdot 10^{-3}$
Fructose	$k_1$	$\mathcal{N}(0.0133, 0.249 \cdot 10^{-5})$	$[13.250, 13.259] \cdot 10^{-3}$
	$k_2$	$\mathcal{N}(0.0112, 1696 \cdot 10^{-5})$	$[11.187, 11.194] \cdot 10^{-3}$

Table 2: Estimated distributions for the kinetic parameters.

Figure 4 shown the join probability distribution obtained for the  $k_1$  and  $k_2$  model parameters.

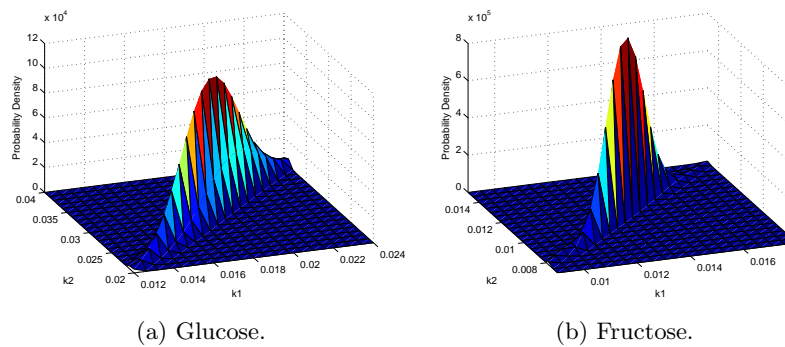


Figure 4: Join probability distribution for the kinetic parameters estimation.

## Conclusions

In this paper an analysis of how the uncertainty of the continuous chromatography operational factor values influences in the mass transfer kinetic parameter estimation, used for the characterization of a chromatographic column modeled on simulated moving bed, is presented.

It was demonstrated how the use of factorial analysis for the construction of the response surface allowed, in a direct way, to establish which are the operational factors of greater influence in the estimated parameters. In addition, regions of the chromatograms where these factors cause the greatest variation were determined. Verifying that each factor creates a region with a different pattern in the output dispersion, which provides a support to the experimenters of which ones are the regions of greater variation. It also showed that the variations are mostly determined by the main factors and that the combined effects of these have a very small contribution. This characterization of the factor's contribution allows to identify the uncertainty sources with the highest incidence, and therefore the ones in which we have to assure bigger exactness for experimentation.

By using the response surface, the probability densities that describe the intervals in which the esteemed parameters can be found were obtained, from defining the range of values that the sources of uncertainty can take. All of this without requiring the execution of a large number of experiments.

## References

- [1] D.C.S. Azevedo, A. Rodrigues, SMB chromatography applied to the separation/purification of fructose from cashew apple juice, *Braz J Chem Eng*, 17 (2000) 507-516.
- [2] A.J. Bihain, A.J. Silva Neto, O. Llanes-Santiago, J.C. Afonso, L. D. Câmara, The front velocity modelling approach in the chromatographic column characterization of glucose and fructose separation in SMB, *TRENDS IN CHROMATOGRAPHY* 2 (2012) 57-73.
- [3] L.D. Câmara, Stepwise Model Evaluation in Simulated Moving-Bed Separation of Ketamine, *Chemical Engineering & Technology*, 37 (2014) 1-10.
- [4] L.D. Câmara, A.J. Silva Neto, Inverse Stochastic Characterization of Adsorption Systems by a Random Restricted Window (R2W) Method, *In: EngOpt 2008 - International Conference on Engineering Optimization*, (2008).
- [5] C. Díez, O. Cabellos, D. Rochman, A. Koning, J. Martínez, Monte Carlo uncertainty propagation approaches in ADS burn-up calculations, *Annals of Nuclear Energy*, 54 (2013) 27-35.
- [6] G. Guiochon, Preparative liquid chromatography, *Journal of Chromatography A*, 965 (2002) 129-161.
- [7] D. Hibbert, J. Jiang, M.I. Mulholland, Propagation of uncertainty in high-performance liquid chromatography with UV/VIS detection, *Analytica Chimica Acta*, 443 (2001) 205-214.
- [8] E. Hund D. Massart, J. Smeyers-Verbeke, Comparison of different approaches to estimate the uncertainty of a liquid chromatographic assay, *Analytica Chimica Acta* 480 (2003) 39-52.
- [9] A.I. Khuri, S. Mukhopadhyay, Response surface methodology, *WIREs Comp Stat*, 2 (2010) 128-149.
- [10] J. Lugon, A.J. Silva Neto, C.C. Santana, A hybrid approach with artificial neural networks, Levenberg/Marquardt and simulated annealing methods for the solution of gas/liquid adsorption inverse problems, *Inverse Problems in Science and Engineering*, 17 (2009) 85-96.
- [11] D.C. Montgomery, "Design and Analysis of Experiments", 8th edn. Wiley, 2012.
- [12] A.J. Silva Neto, J.C. Becceneri, "Técnicas de inteligência computacional inspiradas na natureza: Aplicação em problemas inversos em transferência radiativa", 2nd edn. Sociedade Brasileira de Matemática aplicada e Computacional, 2012.