

# Factorial Experiment Design in the Front Velocity Modeling Approach Applied to Chromatographic Separation of Glucose and Fructose

A. Prieto-Moreno<sup>1</sup>, L.D.Tavares Câmara<sup>2</sup> and O. Llanes-Santiago<sup>1</sup>  
and A. J. Silva Neto <sup>2</sup>

**Abstract:** This work deals with a statistical approach to the uncertainty propagation analysis when estimating the kinetic mass transfer parameters used to model a chromatographic column in the Simulated Moving Bed. The chromatographic column modeling was performed using the new front velocity approach. The uncertainty propagation analysis of operational factors intervening in the chromatographic process to estimated parameters was made using the response surface methodology. The application of the factorial experimental design allowed us to establish those operational factors showing a greater influence on continuous chromatography. Besides, the chromatographic regions, where factors cause a greater output variation as well as their respective patterns, were determined. The analysis was applied to the separation of glucose and fructose.

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

## 1 Introduction

The characterization of chromatographic columns constitutes an important step to determine mass transportation properties used in designing industrial units such as the Simulated Moving Bed (SMB) process. The modeling and simulation of chromatographic systems lead to the understanding of the main mass transfer mechanisms and operational conditions that can be used to improve the molecule separation/purification.

The application of the inverse problem methodology to chromatography has provided highly accurate outcomes when determining mass transfer parameters, as

---

<sup>1</sup> CUJAE, Marianao, La Habana, Cuba.

<sup>2</sup> IPRJ-UERJ, Nova Friburgo, RJ, Brazil.

shown by several researches on the study and analysis of absorption chromatographic systems [Vasconcellos, Silva-Neto, and Santana (2003); Câmara and Silva-Neto (2008); Lee, Kasat, Cox, and Wang (2008); Lugon, Silva-Neto, and Santana (2009); Nam and Mun (2012)].

Both in computer-aided prediction and estimation, as well as in experimentation, the accuracy quantification of outcomes is important [Moffat (1988); Roache (1997)]. The error propagation relationship among the physical quantities involved should be known in order to either obtain the actual value and the confidence interval of a physical magnitude of interest, or optimize the experimental conditions or the design from numerical simulations or measurements prone to random errors. A measurement or estimation uncertainty is defined as a parameter associated with the result of a measurement or estimation and it characterizes value dispersion [ISO (2008)]. The result of a measurement or estimation is considered as the best estimation of a real value, and all sources of uncertainty affect its propagation. Therefore, the result cannot be adequately construed without knowing its uncertainty. The uncertainty quantification will allow us to establish the confidence intervals for the estimated parameters, which is very important from the engineering point of view.

The estimation of kinetic mass transfer parameters characterizing the high performance liquid chromatography (HPLC) involves four steps: 1) injection, 2) separation, 3) detection and 4) estimation. This article includes an analysis of the uncertainty present in estimation due to each of the above-mentioned steps. In the injection step, uncertainty is mainly due to random variations in the injected volume or the sample concentration. In the separation step, the main sources of uncertainty are: retention volume, temperature and flow ratio. Detection adds uncertainty to the concentration measurement reaching the detector. The estimation step adds the uncertainty resulting from the intrinsic variability of the estimation algorithms.

Determining the way in which uncertainties of different factors interact with each other and affect the final estimation is a non-linear non-trivial problem. For mathematically simple models, when the number of parameters to be subjected to variation is small and its relationships can be expressed as an algebraic equation, then the uncertainty analysis can be made using analytical methods [Rao (2005)]. However, for more complex processes, such as column chromatography, this approach should be ruled out.

The main purpose of stochastic solutions is to determine the mean (expected) solution of the physical problem and to obtain the solution confidence intervals, for a given uncertainty in some input parameters [Mendes, Ray, Pereira, Pereira, and Trimis (2012)]. There are different stochastic approaches accurately modeling the uncertainty propagation of input parameters in output variables by means of sim-

ulations [Tatang (1995)]. A typical strategy is the use of the Monte Carlo method [Hibbert, Jiang, and Mulholland (2001); Díez, Cabellos, Rochman, Koning, and Martínez (2013)]. This method randomly selects the values of target factors according to their probability distributions. Other methods are based on the spectral representation of parameters uncertainty using the Polynomial Chaos Decomposition [Knio and Maître (2006); Najm (2009)]. There are two procedures to estimate uncertainty of lab findings, namely, the bottom-up and the top-down procedures [Kučera, Bode, and Stvpánek (2000)].

The aim of this work is to analyze the propagation of the uncertainty present in the values of some variables intervening in the characterization of a column chromatography when estimating the kinetic parameters of the model. This allows us to know which factors require a greater experimental accuracy.

For this purpose, an alternative method to the ones above-mentioned should be applied. This alternative method is based on the use of the response surface [Khuri and Mukhopadhyay (2010)] to obtain the probability distribution function (*pdf*) of estimated parameters. Compared with the Monte Carlo method, the advantage of this proposal is that computer time is shorter and the interpretation of how uncertainty affects the factors intervening in the estimated parameters variations is more straightforward.

The work is organized as follows: Section 2 includes the column chromatography modeling technique known as front velocity. Section 3 describes the different uncertainty sources, and the response surface methodology as a way to model uncertainty propagation and the factorial analysis required to study the influence of different uncertainty sources. Findings and their analysis are included in Section 4. And finally, the conclusions.

## 2 Front Velocity for Chromatographic Modeling

The column chromatography model determines the SMB final separation performance depending on the number of interconnected columns. In general, several research groups use dispersion models [Guiochon (2002)] to represent the column chromatography. These are robust and efficient models, though they require a sound numeric treatment of equations in partial derivatives which demand a high computer level.

This work utilize a new modeling approach known as convection front velocity [Bihain, Silva-Neto, Llanes-Santiago, Afonso, and Câmara (2012);Câmara (2014)]. According to this approach, the liquid phase convection is considered the main phenomenon in molecular transportation through column chromatography, followed by a mass transfer between the solid adsorbent and the liquid phase.

This modeling approach can be used as a powerful tool to determine the chromatographic behavior of a sample, since it can be easily implemented in a routine analysis requiring a reduced number of parameters.

### 2.1 Front Velocity

Due to the fact that the flow velocity inside the column is determined by an external pumping system, the time required by the liquid phase to travel along the chromatographic column can be determined if the volumetric flow velocity, the porosity and the column volume are experimentally known.

In the chromatographic column shown in Fig. 1, the  $J^*$ -size control volume travels along the column at the same speed of the eluent flow. In this case, the column longitude is discretized with a  $J^*$ -size control volume.

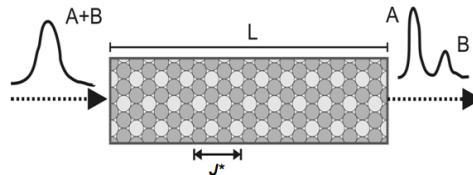


Figure 1:  $L$ -longitude chromatographic column with a discretized volume of  $J^*$  [Câmara (2014)].

The  $(\Delta t)$  time interval in which the liquid phase travels for each control volume is obtained with the following expression

$$\Delta t = \frac{\epsilon V}{nF} \tag{1}$$

where  $\epsilon$ ,  $V$ ,  $n$  and  $F$  correspond to the column bed porosity, the total column volume, the number of control volumes and the liquid flow ratio, respectively.

When modeling the mass transfer, it is necessary to assume two concentrated mass transfer models described by equations (2) and (3), where  $C$ ,  $q$ ,  $k_1$  and  $k_2$  account for the concentration of both liquid and solid phases, as well as the global adsorption and desorption constants of the mass transfer kinetics.

$$\frac{dC}{dt} = -k_1 C + k_2 q \tag{2}$$

$$\frac{dq}{dt} = -\frac{dC}{dt} \tag{3}$$

Modeling simulations using the front velocity are compared with experimental data available in [Azevedo and Rodrigues (2000)]. Fig. 2 shows a comparison between

experimental results and data obtained by simulation, for which the following parameters were used: flow-rate = 30ml/min, porosity = 0.4, injection-vol. = 300ml, and injection-conc. = 15mg/ml.

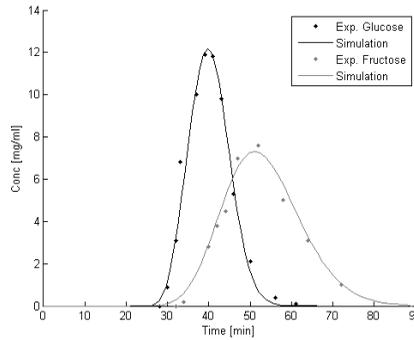


Figure 2: Comparison between the simulation (lines) results and the experimental absorption (dots) for glucose and fructose.

## 2.2 Estimation of Parameters

With the purpose of estimating the unknown model parameters, an implicitly inverse problem is formulated as an optimization problem in order to minimize the residual square function

$$S(\vec{K}) = [\vec{C}_{exp} - \vec{C}_{cal}(\vec{K})]' [\vec{C}_{exp} - \vec{C}_{cal}(\vec{K})] = \vec{R}'\vec{R} \quad (4)$$

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

$$\vec{R} = \vec{C}_{exp} - \vec{C}_{cal}(\vec{K}). \quad (5)$$

The  $\vec{K}^*$  inverse problem solution minimizes the norm given by (4), which is

$$\min_{\vec{K}} S(\vec{K}) = S(\vec{K}^*). \quad (6)$$

The optimization method used for the inverse problem solution was the Simulated Annealing (SA) [Chibante (2010); Silva Neto and . Becceneri (2012)]. However, other methods can be used as shown in [?; Cuco, Silva Neto, Campos Velho, and

de Sousa (2009)]. An SA characteristic, which is common for all global optimization methods, is its high computer effort, as well as the time the direct problem solution requires to obtain the system response. In order to prevent delays in the estimation of parameters, an iterative approach was used, in which the algorithm is configured for barely some cycles and then it is restarted with the outcome achieved as an initial condition. This results in a better use of the initial convergence velocity of this algorithm. The iterative process stops when the difference between two consecutive solutions is below the established value, or when a specific number of iterations is achieved.

Tab.1 includes the global mass transfer parameters obtained.

Table 1: Global mass transfer parameters obtained from the inverse problem.

	Glucose	Fructose
$k_1 [min^{-1}]$	0.01798	0.01302
$k_2 [min^{-1}]$	0.03001	0.01098
$K_{eq} = \frac{k_1}{k_2}$	0.599	1.186
$S [g/L]$	4.59	4.50

### 3 Uncertainty Propagation in a Chromatographic Column

#### 3.1 Uncertainty propagation analysis using simulation

The uncertainty propagation analysis using an analytical method could imply a complex mathematical handling, for which an analytical solution can only be possible through approximations. A method eluding algebra and calculation is the one that applies numerical simulation techniques which only require the system or process model and the distribution knowledge or uncertainty levels. With this approach, and knowing the values of variables, outputs are repeatedly calculated with minor input changes. Once enough repetitions are made, output distributions assume the correct form from which the mean and standard deviations can be determined (or any other appropriate measurement of the distribution).

All potential variation factors should be taken into consideration when designing simulations. The experimenter should decide how many factors would vary significantly and the number of experiments to be conducted. The number of factors depends on the potential uncertainty sources and the number of experiments should be selected in accordance with the required properties, namely: orthogonality, rotativity, uniform precision and optimality. In turn, the samples to be analyzed should be homogeneous and as much representative of future samples as possible [Maroto,

Riu, Bouqué, and Rius (1999)]. If information (historical data) is available, input uncertainty intervals can be quantified as in [Wang and Wang (2011)].

Input variables considered in this work are: flow velocity, porosity, injection volume, and injection concentration, which have been considered in other works too [Hibbert, Jiang, and Mulholland (2001);Hund, Massart, and Smeyers-Verbeke (2003)]. For each variable, variations up to 10% were defined, as indicated in [Hund, Massart, and Smeyers-Verbeke (2003)]. The uncertainty obtained for estimating the parameters will be the outcome from the combination of contributions of every uncertainty sources.

### 3.2 Response Surface Methodology

As previously mentioned, a commonly used method for this kind of analysis is the Monte Carlo Method. However, for the process under study, this approach is not recommended due to the number of executions that must be made in order to obtain the corresponding *pdf*. Fig. 3 shows the Monte Carlo scanning in the uncertainty region, where the estimation of the model parameters should be made for each indicated point. The figure assumes a two-variable process defined in a coded scale from  $-1$  to  $+1$  (the low and high levels of the variable).

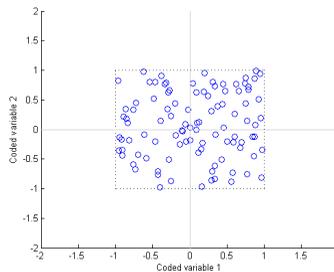


Figure 3: Scanning of the uncertainty region required by the Monte Carlo Method

The response surface methodology approach is used in this work. It involves a set of mathematical and statistical techniques used to develop a relationship function between a  $y$  response of interest and a set of  $x_1, x_2, \dots, x_k$  associated variables (inputs) [Khuri and Mukhopadhyay (2010)]. In general, this relationship is unknown, but can be approximated through a polynomial model

$$y = f'(\mathbf{x})\boldsymbol{\beta} + \varepsilon \quad (7)$$

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  $\varepsilon$  is a random experimental error assumed to have a zero mean. The quantity  $f^l(\mathbf{x})\beta$  represent the mean response (expected value of  $y$ ), denoted by  $\mu(\mathbf{x})$ .

The aim is to consider a model like (7) meeting three objectives:

1. To establish a relationship, though approximated, between  $y$  and  $x_1, x_2, \dots, x_k$ , which might be used to predict the response values for different sets of values in input variables.
2. To determine, through hypothesis tests, the significance of factors whose levels are represented by  $x_1, x_2, \dots, x_k$ .
3. To determine the optimal  $x_1, x_2, \dots, x_k$  combination resulting in a maximum (or minimum) response over a certain region of interest.

Usually, two main models are used [Khuri and Mukhopadhyay (2010); Montgomery (2012)]. These are model special cases (7): the first-order model ( $d = 1$ )

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \varepsilon, \tag{8}$$

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

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \sum_{i=1}^k \beta_i x_i^2 + \varepsilon. \tag{9}$$

In order to achieve the three objectives previously mentioned, a sequence of  $n$  experiments should be conducted. In each experiment, the  $y$  response is measured (or observed) for a specific combination of input variables. All these combinations constitute the so-called response surface design which can be represented in a matrix denoted by  $\mathcal{D}$ , in the  $n \times k$  order

$$\mathcal{D} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

where  $x_{ui}$  denotes the  $u$ th design setting of  $x_i$ . Each row of  $\mathcal{D}$  represents a point, referred to as a design point, in the  $k$ -dimensional Euclidean space. Let  $y_u$  denote the response value obtained as a result of applying the  $u$ th setting of  $\mathbf{x}$ , that is  $\mathbf{x}_u = (x_{u1}, x_{u2}, \dots, x_{uk})'$ . From Eq. (7), it is obtained

$$y_u = f^l(\mathbf{x}_u)\beta + \varepsilon_u, \quad u = 1, 2, \dots, n \tag{10}$$

where  $\varepsilon_u$  denotes error term at the  $u$ th experimental run. Model Eq. (10) can be expressed in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{11}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \mathbf{y} \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

In general,  $\mathbf{y}$  is an  $(n \times 1)$  vector of the observations,  $\mathbf{X}$  is an  $(n \times p)$  matrix of the levels of the independent variables,  $\boldsymbol{\beta}$  is a  $(p \times 1)$  vector of the regression coefficients, and  $\boldsymbol{\varepsilon}$  is an  $(n \times 1)$  vector of random errors.

Assuming that  $\boldsymbol{\varepsilon}$  has a zero mean and a covariance matrix given by  $\sigma^2 \mathbf{I}_n$ , the ordinary least-squares estimator of  $\boldsymbol{\beta}$  is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{12}$$

In general, at any point,  $\mathbf{x}$ , in a experimental region, denoted by  $\mathcal{R}$ , the predicted response is

$$\hat{y}(\mathbf{x}) = f'(\mathbf{x})\hat{\boldsymbol{\beta}}, \quad \mathbf{x} \in \mathcal{R}. \tag{13}$$

### 3.3 Factorial design

Obtaining the response value entails the study of the effect of two or  $p$  more variables known as factors in this context. In general, the factorial design used as an approach to obtain the  $\mathcal{D}$  response surface design is the most efficient for this purpose [Feng, Saal, Salsbury, Ness, and Lin (2007); Hinkelmann and Kempthorne (2008); Hajjaji, Renaudin, Houas, and Pons (2010)]. Besides, it has the advantage of establishing a relationship between the parameters of the regression model (7) and estimating the effects of each factor.

For the simple case of two variables (factors):  $A$  and  $B$ , the observations in a factorial experiment can be described by an effects model [Montgomery (2012)] given by Eq. (14)

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \tag{14}$$

where  $\mu$  is the overall mean effect,  $\tau_i (i = 1, 2, \dots, a)$  is the effect of the  $i$ th level of the factor  $A$ ,  $\beta_j (j = 1, 2, \dots, b)$  is the effect of the  $j$ th level of the factor  $B$ ,

$(\tau\beta)_{ij}$  is the effect of the interaction between  $\tau_i$  and  $\beta_j$ , and  $\varepsilon_{ijk}$  is a random error component. The subindex ( $k = 1, 2, \dots, n$ ) represents the replicate count carried out in each experiment. The effects are defined as deviations from the overall mean, so  $\sum_{i=1}^a \tau_i = 0$ ,  $\sum_{j=1}^b \beta_j = 0$  and  $\sum_{i=1}^a (\tau\beta)_{ij} = \sum_{j=1}^b (\tau\beta)_{ij} = 0$ .

In principle, all factors have the same influence, therefore it is of interest to test the hypothesis about the equality of the effects. In the case of two factors would be:

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_a = 0 \tag{15a}$$

$$H_1 : \text{at least one } \tau_i \neq 0$$

to test the hypotheses about the equality of A's levels effects,

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_b = 0 \tag{15b}$$

$$H_1 : \text{at least one } \beta_j \neq 0$$

to test the hypotheses about the equality of B's levels effects, and also it is of interest determining whether A and B effects interact. Thus, it is also wished to test

$$H_0 : (\tau\beta)_{ij} = 0, \forall i, j \tag{15c}$$

$$H_1 : \text{at least one } (\tau\beta)_{ij} \neq 0$$

These hypotheses are tested using an analysis of variance (ANOVA), in which the sum of squares (SS) of the deviation from the overall mean effect due to each one of the factors intervene, their interactions, and the error (difference between replicates for  $n \geq 2$ ). To test the hypotheses about the effects and their interactions, the value of ratio  $\frac{MS_{Effect}}{MS_E}$  must be analyzed. Large values of this ratio imply that the data do not support the null hypotheses. For the case of two factors,  $MS_{Effect}$  is defined as

$$E(MS_A) = E\left(\frac{SS_A}{a-1}\right) = \sigma^2 + \frac{bn \sum_{i=1}^a \tau_i^2}{a-1} \tag{16}$$

$$E(MS_B) = E\left(\frac{SS_B}{b-1}\right) = \sigma^2 + \frac{an \sum_{j=1}^b \beta_j^2}{b-1} \tag{17}$$

$$E(MS_{AB}) = E\left(\frac{SS_{AB}}{(a-1)(b-1)}\right) = \sigma^2 + \dots + \frac{n \sum_{i=1}^a \sum_{j=1}^b (\tau\beta)_{ij}^2}{(a-1)(b-1)} \tag{18}$$

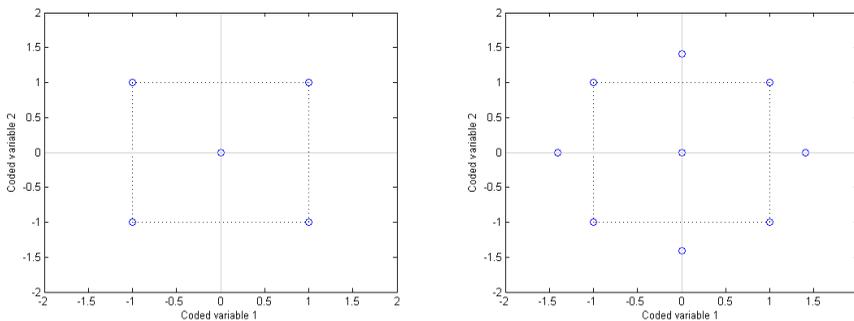
and  $MS_E$  is defined as

$$E(MS_E) = E\left(\frac{SS_E}{ab(n-1)}\right) = \sigma^2 \tag{19}$$

If it is assumed that the model Eq. (14) is adequate and that the error terms  $\epsilon_{ijk}$  are normally and independently distributed with constant covariance  $\sigma^2$ , then each of the ratios of means squares  $\frac{MS_A}{MS_E}$ ,  $\frac{MS_B}{MS_E}$ , and  $\frac{MS_{AB}}{MS_E}$  are distributed as  $F$  with  $a - 1$ ,  $b - 1$ , and  $(a - 1)(b - 1)$  numerator degree of freedom, respectively, and  $ab(n - 1)$  denominator degree of freedom. The effect is significant if the calculated ratio is greater than the threshold defined for the test.

In this work two levels are defined for each one of the factors, that is called  $2^k$  factorial design. This design is widely used in industrial studies and is considered to provide results with adequate accuracy [(Montgomery, 2012)].

The factorial design has several advantages, it: 1) allows the effect analysis of each factor individually and interactions among them; 2) allows the estimation of a factor effects at different levels of other factors, thus leading to conclusions which are valid over a range of experimental conditions; and 3) requires a shorter number of experiments. Concerning the latter advantage, if the number of experiments required by the Monte Carlo Method, Fig. 3, is compared with those required by the  $2^k$  factorial approach, Fig. 4, the factorial approach requires a highly reduced number of experiments. This is due to the fact that results from factorial experiments –also known as factorial points– are used to obtain the response surface generating polynomial by which the output estimation for random values within the uncertainty region is made.



(a) Generating first-order polynomial  $d = 1$     (b) Generating second-order polynomial  $d = 2$

Figure 4: Scanning of the uncertainty region required by the  $2^k$  con  $k = 2$  Factorial Analysis

We must find out if there are significant differences between the mean value of experiments conducted and the experimental findings for the surface center in order to determine the need of incorporating the quadratic terms in the surface response

generating polynomial.

Let's assume that  $\bar{y}_F$  is the mean value of experiments conducted, and  $\bar{y}_C$  the mean value of runs in the center. If the difference  $\bar{y}_F - \bar{y}_C$  is small, then the central point is close to the hyperplane crossing the factorial points, hence there is no quadratic curvature. The square sum for a quadratic curvature is

$$SS_{PQ} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{\bar{y}_F + \bar{y}_C} \quad (20)$$

where  $n_F$  is the number of factorial points and  $n_C$  is the number of runs in the center. This amount can be compared with the error square sum to prove the pure quadrature.

#### 4 Outcomes and discussion

In order to illustrate the uncertainty propagation analysis, we must use the problem to estimate the kinetic mass transport parameters in a chromatographic column modeled by the front velocity. Tab. 2 shows the variables selected for the analysis and deviations of the 10% of the corresponding experimental nominal values.

Table 2: Variables of interest for the uncertainty analysis when estimating the front velocity model parameters.

Variable	Range	Eng Unit
Flow-rate	27-33	ml/min
Porosity	0.36-0.44	
Injection vol.	270-330	ml
Injection conc.	13.5-16.5	mg/ml

Fig. 5 shows how the flow-rate variation affects the output process. Graphs in Fig. 5(a) are obtained by solving the direct problem for the flow-rate values used in the factorial design, showing the different concentration profiles at the time in which they were obtained. The difference module between the profiles obtained for the maximum and minimum values, denoted by  $\delta$ , represent the uncertainty region for the output concentration generated by the flow-rate variations. These uncertainty regions are shown in Fig. 5(b).

Fig. 6 shows the output uncertainty regions generated for variations in the remaining factors of interest. As can be seen, the uncertainty concentration can vary in time and the dispersion resulting from each factor produces a different uncertainty pattern. This uncertainty characterization can provide experimenters with a support to know regions displaying a higher variation due to changes in factors intervening in the chromatography.

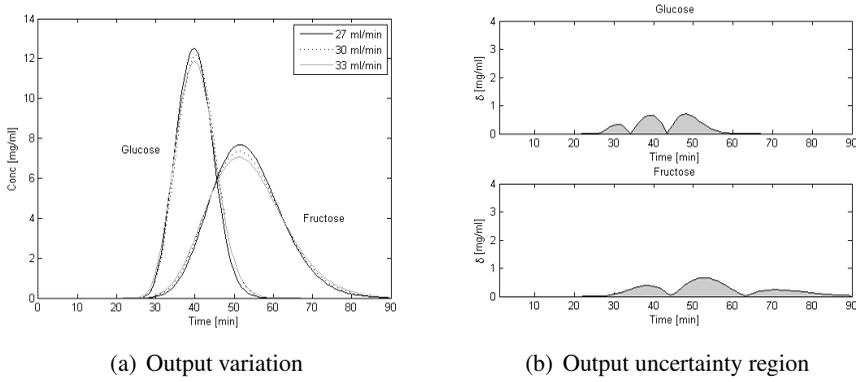


Figure 5: Influence of variations in the flow-rate

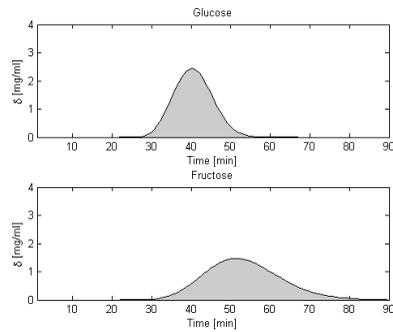
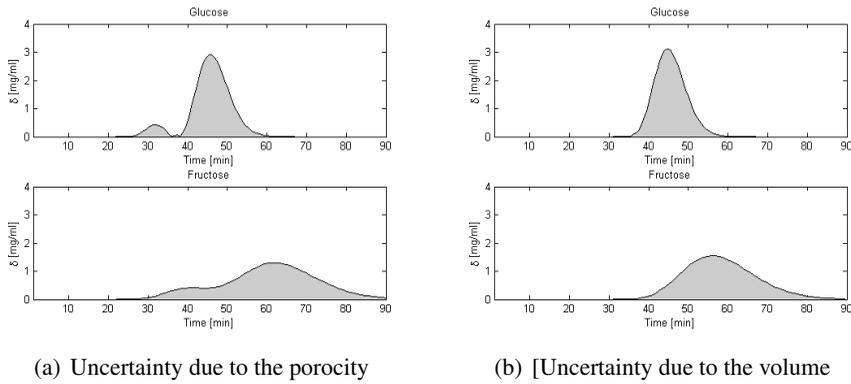


Figure 6: Uncertainty regions in the output concentration generated by variations in the factors of interest.

**4.1 Estimation of the confidence interval**

This work proposes the  $2^k$  [Kandananond (2013)] factorial analysis to design the response surface. According to this design,  $\hat{\beta}$  elements are not correlated and their variances have a minimum value. This implies that the design provides the greatest accuracy when estimating the model parameters (8). Two replicas were made for every experiment indicated by the analysis.

The analysis of variance (ANOVA) was performed in order to determine how factors analyzed affect the estimation of kinetic mass transfer parameters. Tab. 3 shows the ANOVA corresponding to the factorial analysis of the  $k_1$  parameter estimation in glucose. The threshold for the  $F$  statistical test to consider a significant effect is  $t_h = 4.49$ .

Table 3: Analysis of variance to estimate  $k_1$  with glucose.

Source	Value( $10^{-3}$ )	SS( $10^{-5}$ )	DoF	FStat	Sig	%Contrib	Coef( $10^{-3}$ )
flow-rate (A)	-2.54	5.16	1	2992.56	+	8.30	-1.27
porosity (B)	-0.97	0.75	1	434.28	+	1.20	
injection vol. (C)	4.16	13.83	1	8029.09	+	22.27	2.08
injection conc. (D)	7.18	42.21	1	23921.49	+	66.35	3.59
AB	-0.26	0.06	1	35.30	+	0.10	
AC	-0.03	0.003	1	0.55	-	0.00	
AD	0.57	0.26	1	152.15	+	0.42	
BC	0.30	0.07	1	42.18	+	0.12	
BD	-0.24	0.05	1	27.79	+	0.08	
CD	0.06	0.004	1	1.72	-	0.00	
ABC	-0.61	0.29	1	169.98	+	0.47	
ABD	-0.08	0.01	1	3.01	-	0.01	
ACD	-0.60	0.29	1	166.54	+	0.46	
BCD	-0.28	0.06	1	36.14	+	0.10	
ABCD	-0.24	0.04	1	25.68	+	0.07	
Total		62.12	31				18.16

The  $F$  statistics indicates that most factors of interest and their combinations have a significant effect on the  $k_1$  estimation. However, this test is sensitive to the number of replicas made in each experiment, since the freedom degrees of the statistical test depend on the number of replicas made. This means that outcomes considered to be significant outcomes for a number of replicas might not be significant for a different number of them. Therefore, the criterion applied when selecting the effects to be included in the response surface generating polynomial is the percentage contributing with the behavior of the studied variable not having significant changes when the number of replicas of the experiments made does not vary.

The criterion used when selecting factors to be used in the response surface gener-

ating polynomial is the incorporation of those determining the output behavior in more than one 2%. As can be seen in Tab. 3, the variation when estimating the  $k_1$  parameter is mainly determined by the effects of (A), (C) y (D) factors, which determine the 96.92% of the parameter variation.

Once the greatest-effect factors are selected, a multiple regression model is made to relate these factors with outcomes derived from the  $k_1$  estimation. The regression coefficients of the model are obtained by applying the Least Square Method associating each  $k_1$  estimated value with the corresponding values used in the direct problem solution. The generating polynomial obtained was

$$\hat{k}_1 = (18.16 - 1.27A + 2.08C + 3.59D) \cdot 10^{-3} \text{ min}^{-1} \tag{21}$$

Tab. 4 shows the ANOVA corresponding to the factorial analysis of the  $k_1$  parameter estimation for fructose

Table 4: Analysis of variance to estimate  $k_1$  with fructose.

Source	Value( $10^{-3}$ )	SS( $10^{-5}$ )	DoF	FStat	Sig	%Contrib	Coef( $10^{-3}$ )
flow-rate (A)	-2.89	6.69	1	1380.28	+	29.84	-1.45
porosity (B)	-1.41	1.60	1	329.82	+	7.13	-0.71
injection vol. (C)	3.18	8.07	1	1663.81	+	35.97	1.59
injection conc. (D)	2.46	4.85	1	999.18	+	21.60	1.23
AB	0.83	0.54	1	112.31	+	2.43	0.41
AC	0.03	0.001	1	0.16	-	0.00	
AD	-0.07	0.002	1	0.92	-	0.02	
BC	-0.36	0.10	1	20.86	+	0.45	
BD	-0.33	0.09	1	18.38	+	0.40	
CD	-0.07	0.002	1	0.71	-	0.02	
ABC	0.17	0.02	1	4.76	+	0.10	
ABD	0.25	0.05	1	9.93	+	0.21	
ACD	-0.42	0.14	1	29.44	+	0.64	
BCD	-0.48	0.19	1	38.76	+	0.84	
ABCD	0.06	0.001	1	0.64	-	0.01	
Total		22.44	31				13.23

As can be seen, the influence of factors of interest on variations in  $k_1$  estimation is more homogeneous in the case of fructose. This homogeneity implies the incorporation of a greater number of factors to the response surface generating polynomial. In this case, it is necessary to include the effect of the four main factors and the combined effect (AB) to explain the 96.97% of the  $k_1$  variation. The generating polynomial obtained is:

$$\hat{k}_1 = (13.23 - 1.45A - 0.71B + 1.59C + 1.23D + 0.41AB) \cdot 10^{-3} \text{ min}^{-1} \tag{22}$$

Following the same procedure for  $k_2$ , the generating polynomials obtained for both glucose and fructose are, respectively:

$$\hat{k}_2 = (30.13 - 2.03A + 2.41C + 6.19D) \cdot 10^{-3} \text{ min}^{-1} \tag{23}$$

$$\hat{k}_2 = (11.17 - 1.20A - 0.51B + 1.22C + 1.15D + 0.31AB) \cdot 10^{-3} \text{ min}^{-1} \tag{24}$$

with which 97.35% and 97.13% of the  $k_2$  variation can be explained.

For all cases, a statistical test was conducted to determine the existence of a quadratic curvature in the response surface confirming that there is no evidence of quadratic effects. Therefore, we can conclude that a first-order surface is appropriate.

Once the generating polynomials describing the behavior of estimated parameters against the variations of factors of interest are obtained, random values within the range defined for every factor are assigned, and the *pdf* for each kinetic mass transfer parameter is estimated. In this work, the *pdf* estimation was made using the Maximum Likelihood Estimation (MLE) [Myung (2003)] method .

Tab. 5 shows the *pdf* characterizing the variability when estimating the kinetic model parameters. Intervals specified for parameters correspond with the limits for a confidence 95%.

Table 5: Estimated distributions for kinetic parameters.

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	$k_1$	$\mathcal{N}(0.0133, 0.249 \cdot 10^{-5})$	$[13.250, 13.259] \cdot 10^{-3}$
	$k_2$	$\mathcal{N}(0.0112, 1696 \cdot 10^{-5})$	$[11.187, 11.194] \cdot 10^{-3}$

Fig. 7 shows the joint probability distributions obtained for  $k_1$  and  $k_2$  parameters of the model.

#### 4.2 Influence of the adsorption

A very important characteristic of the column chromatography process is the  $K_{eq} = \frac{k_1}{k_2}$  value, which expresses the tendency of the solute to remain in the liquid or solid phase. For  $K_{eq} < 1$  values, the solute tend to remain in the liquid phase. While for  $K_{eq} > 1$  values, the solute has a greater tendency to reach the solid phase. Fig. 8 shows the variability contribution percentages in  $k_1$  estimation for each effect, for different  $K_{eq}$  values. The broken line indicates a 2% contribution.

As can be seen, the estimation variability is mainly determined by the effects of the main factors. Meanwhile, the combined effects have almost no influence. Besides,

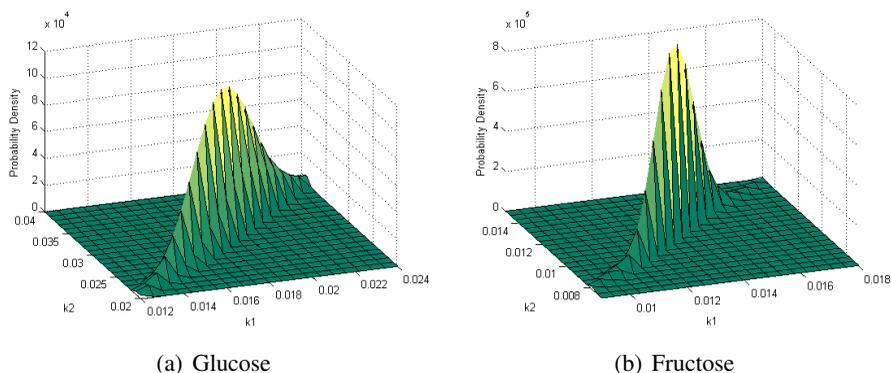


Figure 7: Joint probability distributions to estimated kinetic parameters.

graphs evidenced how the contribution homogeneity of the main effects increases while the  $K_{eq}$  value increases.

In the graph Fig. 8(a) corresponding to glucose, we can see the poor contribution of porosity –less than 1.5%–. This is due to the poor influence of factors affecting the solid phase over substances prone to remain in the fluid phase. Likewise, we can see that the variability in the injected concentration contributes, to a great extent, with the estimation variability. This indicates the need to guarantee a greater accuracy for this parameter in experimentation, since the estimation shows a great sensitivity to changes. In the graph Fig. 8(d) corresponding to fructose, we can see a more uniform contribution of factors and a greater porosity influence.

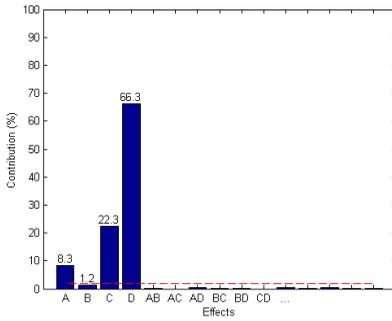
Something similar happens in the behavior of  $k_2$  estimation

Likewise, the  $K_{eq}$  effect is reflected in the estimation of kinetic mass transfer parameters. Fig. 9 shows the mean values of the theoretical quadratic error obtained for each experiment made representing different operational conditions. As can be seen, while de  $K_{eq}$  value increases, it is possible to make estimations closer to the real parameter value and, consequently, to obtain a response which is more adjusted to the real one.

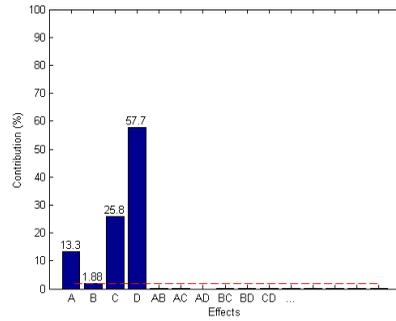
## 5 Conclusions

This work presents an analysis on how uncertainty affects values of operational factors in the continuous chromatography when estimating the kinetic mass transfer parameters used in the characterization of a chromatographic column modeled in the Simulated Moving Bed.

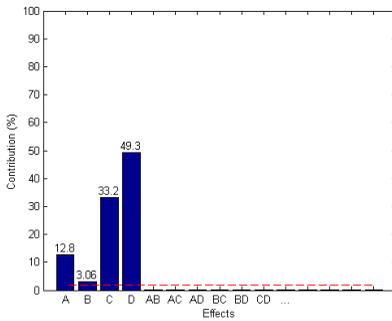
As demonstrated, the use of factorial analysis to develop a response surface allowed



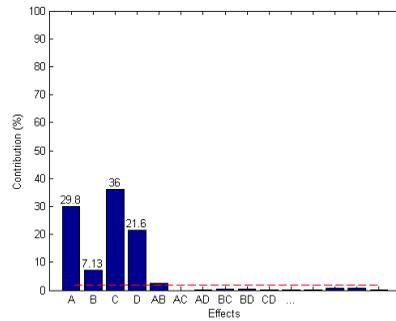
(a)  $K_{eq} = 0.6$



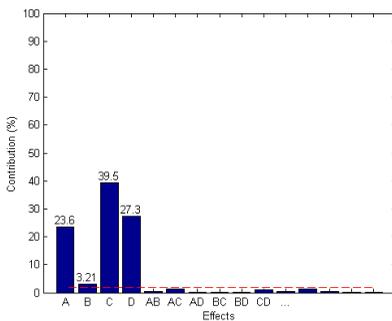
(b)  $K_{eq} = 0.75$



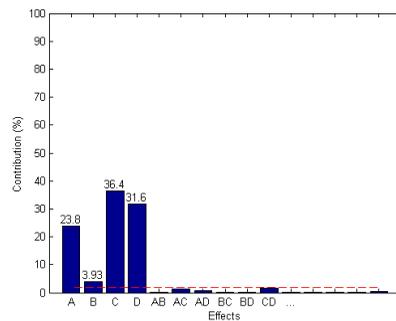
(c)  $K_{eq} = 0.9$



(d)  $K_{eq} = 1.18$



(e)  $K_{eq} = 1.84$



(f)  $K_{eq} = 2.5$

Figure 8: Effects contribution to the estimation variability of  $k_1$ .

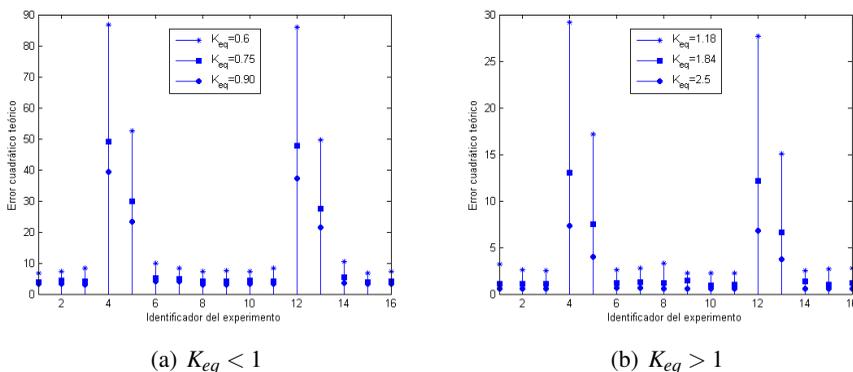


Figure 9: Theoretical mean quadratic error in the estimation of parameters for each experiment.

us to establish, in a direct way, the operational factors having the greatest influence on estimated parameters. Besides, the chromatogram regions where these factors show a greater variation were determined. As was confirmed, every factor creates a region with a different output dispersion pattern, thus supporting experimenters on those regions showing a greater variation. Likewise, as was demonstrated, variations are determined, to a great extent, by the main factors and their combined effects have a very small contribution. This characterization of the contribution of factors allows the identification of uncertainty sources having a greater incidence and, therefore, those requiring a higher accuracy in experimentation.

By using the response surface, we were able to obtain the probability densities describing the intervals in which estimated parameters can be found, by defining the range of values that uncertainty sources can assume. There is no need of implementing a high number of experiments to do all this.

The study evidenced that, when substances involved have a higher absorption rate, higher  $K_{eq}$ , the influence of uncertainty sources on the estimated parameter variations is more homogeneous, and it is possible to make more accurate estimations of parameters, and to obtain model outputs closer to the real ones.

**Acknowledgement:** The authors acknowledge the support provided by FAPER-J, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro, CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, and CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

## References

- Azevedo, D. C. S.; Rodrigues, A.** (2000): SMB chromatography applied to the separation/purification of fructose from cashew apple juice. *Braz. J. Chem. Eng.*, vol. 17, no. 4–7, pp. 507–516.
- Bihain, A. J.; Silva-Neto, A. J.; Llanes-Santiago, O.; Afonso, J. C.; Câmara, L. D.** (2012): The front velocity modelling approach in the chromatographic column characterization of glucose and fructose separation in SMB. *Trends in Chromatography*, vol. 2, pp. 57–73.
- Câmara, L. D.** (2014): Stepwise Model Evaluation in Simulated Moving-Bed Separation of Ketamine. *Chemical Engineering & Technology*, vol. 37, no. 2, pp. 1–10.
- Câmara, L. D.; Silva-Neto, A. J.** (2008): Inverse Stochastic Characterization of Adsorption Systems by a Random Restricted Window (R2W) Method. In *EngOpt 2008 - International Conference on Engineering Optimization*.
- Chibante, R.**(Ed): *Simulated Annealing, Theory with Applications*. Sciyo.
- Cuco, A. P. C.; Silva Neto, A. J.; Campos Velho, H. F.; de Sousa, F. L.** (2009): Solution of an inverse adsorption problem with an epidemic genetic algorithm and the generalized extremal optimization algorithm. *Inverse Problems in Science and Engineering*, vol. 17, no. 1, pp. 85–96.
- Díez, C.; Cabellos, O.; Rochman, D.; Koning, A.; Martínez, J.** (2013): Monte Carlo uncertainty propagation approaches in ADS burn-up calculations. *Annals of Nuclear Energy*, vol. 54, pp. 27–35.
- Feng, C.-X. J.; Saal, A. L.; Salsbury, J. G.; Ness, A. R.; Lin, G. C.** (2007): Design and analysis of experiments in cmm measurement uncertainty study. *Precision Engineering*, vol. 31, no. 2, pp. 94–101.
- Guiochon, G.** (2002): Preparative liquid chromatography. *Journal of Chromatography A*, vol. 965, no. 1–2, pp. 129–161.
- Hajjaji, N.; Renaudin, V.; Houas, A.; Pons, M. N.** (2010): Factorial design of experiment (doe) for parametric exergetic investigation of a steam methane reforming process for hydrogen production. *Chemical Engineering and Processing: Process Intensification*, vol. 49, no. 5, pp. 500–507.
- Hibbert, D.; Jiang, J.; Mulholland, M.-I.** (2001): Propagation of uncertainty in high-performance liquid chromatography with UV-VIS detection. *Analytica Chimica Acta*, vol. 443, no. 2, pp. 205–214.
- Hinkelmann, K.; Kempthorne, O.** (2008): *Design and Analysis of Experiments*. Wiley, 2nd edition.

**Hund, E.; Massart, D.; Smeyers-Verbeke, J.** (2003): Comparison of different approaches to estimate the uncertainty of a liquid chromatographic assay. *Analytica Chimica Acta*, vol. 480, no. 1, pp. 39–52.

**ISO, J. E.** (2008): *International vocabulary of metrology : Basic and general concepts and associated terms (VIM)*. JCGM.

**Kandananond, K.** (2013): Applying 2k factorial design to assess the performance of ann and svm methods for forecasting stationary and non-stationary time series. *Procedia Computer Science*, vol. 22, pp. 60–69.

**Khuri, A. I.; Mukhopadhyay, S.** (2010): Response surface methodology. *WIREs Comp Stat*, vol. 2, no. 2, pp. 128–149.

**Knio, O. M.; Maître, O. P. L.** (2006): Uncertainty propagation in CFD using polynomial chaos decomposition. *Fluid Dynamics Research*, vol. 38, no. 9, pp. 616–640.

**Kučera, J.; Bode, P.; Stvpánek, V.** (2000): The 1993 ISO Guide to the Expression of Uncertainty in Measurement Applied to NAA. *Journal of Radioanalytical and Nuclear Chemistry*, vol. 245, no. 1, pp. 115–122.

**Lee, K. B.; Kasat, R. B.; Cox, G. B.; Wang, N. H. L.** (2008): Simulated moving bed multiobjective optimization using standing wave design and genetic algorithm. *AIChE Journal*, vol. 54, no. 11, pp. 2852–2871.

**Lugon, J.; Silva-Neto, A. J.; Santana, C. C.** (2009): 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*, vol. 17, no. 1, pp. 85–96.

**Maroto, A.; Riu, J.; Bouqué, R.; Rius, F. X.** (1999): Estimating uncertainties of analytical results using information from validation process. *Analytica Chimica Acta*, vol. 391, pp. 173–185.

**Mendes, M. A. A.; Ray, S.; Pereira, J. M. C.; Pereira, J. C. F.; Trimis, D.** (2012): Quantification of uncertainty propagation due to input parameters for simple heat transfer problems. *International Journal of Thermal Sciences*, vol. 60, pp. 94–105.

**Moffat, R. J.** (1988): Describing the uncertainties in experimental results. *Experimental Thermal and Fluid Science*, vol. 1, no. 1, pp. 3–17.

**Montgomery, D. C.** (2012): *Design and Analysis of Experiments*. Wiley.

**Myung, I. J.** (2003): Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, vol. 47, no. 1, pp. 90–100.

**Najm, H. N.** (2009): Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics*, vol. 41, pp. 35–52.

**Nam, H.-G.; Mun, S.** (2012): Optimal design and experimental validation of a three-zone simulated moving bed process based on the Amberchrom-CG161C adsorbent for continuous removal of acetic acid from biomass hydrolyzate. *Process Biochemistry*, vol. 47, no. 5, pp. 725–734.

**Rao, K. S.** (2005): Uncertainty analysis in atmospheric dispersion modeling. *Pure and Applied Geophysics*, vol. 162, no. 10, pp. 1893–1917.

**Roache, P. J.** (1997): Quantification of Uncertainty in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics*, vol. 29, no. 1, pp. 123–160.

**Silva Neto, A. J.; Becceneri, J. C.** (2012): *Técnicas de inteligência computacional inspiradas na natureza: Aplicação em problemas inversos em transferência radiativa*. Sociedade Brasileira de Matemática aplicada e Computacional, 2nd edition.

**Tatang, M. A.** (1995): *Direct incorporation of uncertainty in chemical and environmental engineering systems*. PhD thesis, Massachusetts Institute of Technology. Dept. of Chemical Engineering, 1995.

**Vasconcellos, J.; Silva-Neto, A. J.; Santana, C.** (2003): An Inverse Mass Transfer Problem in Solid-Liquid Adsorption Systems. *Inverse Problems in Engineering*, vol. 11, no. 5, pp. 391–408.

**Wang, X.; Wang, L.** (2011): Uncertainty quantification and propagation analysis of structures based on measurement data. *Mathematical and Computer Modelling*, vol. 54, no. 11–12, pp. 2725–2735.