

Inverse Solution of a Chromatography Model by means of Evolutionary Computation

M. Irizar, L. D. Câmara, A. J. Silva Neto and O. Llanes

Abstract: Modeling of Chromatography allows a better understanding and development of new techniques to be applied at industrial level, although it's relatively complex. The models of this process are represented by systems of partial differential equations with non linear parameters difficult to estimate generally, which constitutes an inverse problem. In general there aren't analytical solutions and therefore numerical methods should be used for their direct solutions. Frequently typical boundary conditions are considered, but it's convenient to study different approaches for those.

Evolutionary Computation has been used successfully in many problems of diverse areas for searching in complex spaces. Considering previous works from the authors, in this article Genetic algorithm and Differential evolution are used for parameters estimation in models of protein chromatography with variants in boundary conditions. In both algorithms each population individual is a supposed condition to the direct solution for the system of partial differential equations, coded in real values, while inverse solution is optimized updating the first one according to a fitness function. A comparative analysis is showed as result.

Keywords: Inverse problems in chromatography, Parameter estimation, Genetic algorithms, Differential Evolution

1 Introduction

Evolutionary Computation, a subfield of Computational Intelligence, is a set of tools for searching in complex spaces which have been used successfully in many problems of diverse areas. These are based on the mechanisms of the genetics (Goldberg, 1989), (Davis, 1991), (Storn and Price, 1995) and constitute efficient techniques of stochastic optimization in cases where the objective functions do not have good mathematical properties (Michalewicz, 1992), (Bäck, 1997). These algorithms carry out their search using a complete population of possible solutions for the problem and implement the survival strategy of the best adapted, as a form

of looking for better solutions. This strategy distinguishes them from the traditional search methods which frequently provide a local optimum, and depend on the initial guess.

In previous works computational methods for inverse problems solution have been applied (Chang et al, 2007), (Liu et al, 2007), (Kanevce et al, 2008), (Marin, 2009), (Amirov et al, 2009). Some of them have developed the parameters estimation in protein chromatography processes starting from experimental data (Fig. 1) which is a complex inverse problem (Tarantola, 2005) for this type of generally nonlinear processes, but it is a very important step for a better understanding of the phenomena involved in this area and can be formulated as a problem of optimization. In those papers, local search methods (Horstmann, 1987), (Gu, 1995), (Altenhöner et al, 1997), (Persson, 2001), (Vasconcellos et al, 2002, 2003) or models constituted by a system of partial differential equations considering typical boundary conditions are considered.

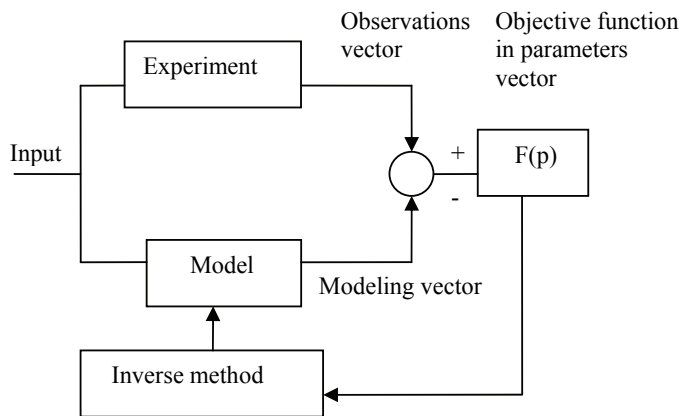


Figure 1: Parameter estimation method scheme.

As it was stated in (Irizar et al, 2009) it's convenient to study different approaches for this last aspect. In that work a Genetic Algorithm was applied to the parameter estimation, but there are some intrinsic parameters of this algorithm implementation such as size population, number of generations, crossover and mutation probability and others to be determined before the operation and definitive general approaches do not exist, therefore a quantity of experiments was required to get the best results. Differential Evolution is a method with a small number of parameters, so in this paper both methods corresponding to the Evolutionary Computation are used for parameter estimation in the same models of protein chromatography with

the variants in boundary conditions already analyzed.

In section 2 the characteristics of adsorption chromatography models are presented. In Section 3 the inverse problem and its solution method by means of Genetic Algorithms and Differential Evolution is reformulated and the developed experiments for parameter estimation in the different models are explained. In section 4 the discussion of results is shown and future works are stated.

2 Models of protein chromatography

Chromatography is a science that studies the separation of molecules based on the differences of its structure and adsorption phenomenon. A mobile phase transports the compounds to be separated and a stationary phase adsorbs those compounds through intermolecular forces.

Mathematical models of chromatography involves a group of parameters whose appropriate estimation can contribute to optimize the production costs.

A model that describes the adsorption of proteins in macro porous solid particles (Blanch, 1997) for chromatography in stirred tank (Fig. 2) includes the transfer of mass mechanisms at the external film and pore diffusion, as well as an expression for the rate of surface reaction.

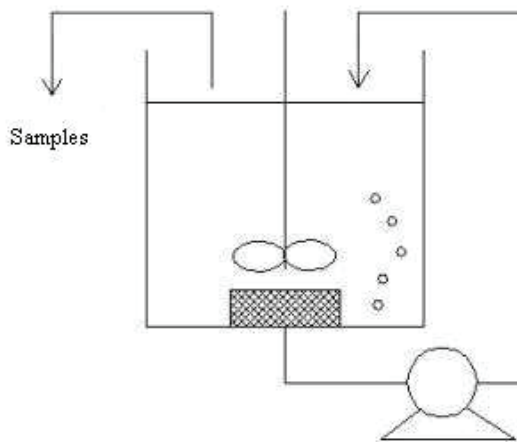


Figure 2: Stirred tank chromatography.

In this model (Eq.1) the left term corresponds to the accumulation of protein inside the pores of the particles and the terms to the right represent the transport by diffusion over radial coordinates and the rate of molecules that have been adsorbed by

the adsorbent phase respectively,

$$\varepsilon_p \frac{\partial C_i}{\partial t} = \varepsilon_p \frac{D_{ef}}{r^2} \frac{\partial}{\partial r} \left[r^2 \frac{\partial C_i}{\partial r} \right] - \rho \frac{\partial q_i}{\partial t} \text{ in } 0 < r < R \text{ for } t > 0 \quad (1)$$

where C_i is the protein concentration in the liquid phase inside the pores of the particles, q_i the protein concentration in the solid phase, D_{ef} the coefficient of effective diffusivity, ρ the density of the adsorbent particles, ε_p the particle porosity, and t and r represent the time and space variables respectively.

The initial condition is

$$C_i(r, t) = 0 \text{ for } t = 0 \text{ in } 0 \leq r \leq R \quad (2)$$

being R is the particles radius.

The boundary conditions are

$$\frac{\partial C_i}{\partial r} = 0 \text{ at } r = 0 \text{ for } t > 0 \quad (3)$$

$$\varepsilon_p D_{ef} \frac{\partial C_i}{\partial r} = k_s (C_b - C_i) \text{ at } r = R \text{ for } t > 0 \quad (4)$$

where C_b is the protein concentration in the liquid phase and k_s is the film mass transfer coefficient.

The mass balance in the bulk liquid phase with regard to the protein concentration can be written as

$$\frac{\partial C_b}{\partial t} = -\frac{3}{R} \frac{1 - \varepsilon_b}{\varepsilon_b} k_s (C_b - C_i)|_{r=R} \quad (5)$$

where ε_b is the bed porosity.

Equation 5 has the following initial condition

$$C_b = C_0 \text{ for } t = 0 \quad (6)$$

This model assumes that q_i follows the Langmuir model, which is determined from equilibrium conditions, i.e.

$$q_i = \frac{q_m C_i}{k_d + C_i} \quad (7)$$

being q_m the maximal adsorption capacity of Langmuir isotherm model and k_d the dissociation constant of Langmuir isotherm model.

Some modification can be made in the model just described. Equation (5) can be substituted for

$$\frac{dC_b}{dt} = -k_s(C_B - \bar{C}_i) \quad (8)$$

and considering the same initial condition (2), but changing boundary conditions according to equations 9 y 10.

$$\frac{\partial C_i}{\partial r} = 0 \text{ para } t > 0 \text{ y } r = 0 \quad (9)$$

$$C_i = C_b \text{ para } t > 0 \text{ y } r = R \quad (10)$$

Assuming Langmuir behavior, equation (1) can be simplified to obtain the equations 11a y 11b.

$$\frac{\partial C_i}{\partial t} = \frac{\psi}{r^2} \frac{\partial}{\partial r} \left[r^2 \frac{\partial C_i}{\partial r} \right] \quad (11a)$$

$$\psi = \frac{D_{ef}}{1 + \frac{\rho q_m k_s}{\varepsilon_p (k_d + C_i)^2}} \quad (11b)$$

Because these equations cannot be solved analytically, the finite difference method (ONeil, 1983) have been implemented for the solution of the direct solid-liquid adsorption problem, that simulate the physical process advancing in time and recalculating the function C_b in each successive instant of time.

3 Inverse Problem Formulation

3.1 Inverse solution for a system of partial differential equations

In most of the scientific disciplines and particularly in engineering there are problems characterized by differential equations with associated initial and boundary conditions. When these problems are solved in a direct way, the result is generally a functional relationship or a system of equations, which can be used to calculate values of the dependent variable for given values of the independent variable.

The inverse solution of systems of partial differential equations constitutes a complex problem, for which there are no universally accepted methods.

Given an applicable direct solution to a system of partial differential equations, it is possible to propose an inverse problem as a problem of optimization. An algorithm to achieve this is (Karr, 2000):

- Suppose a solution to the inverse problem. This can include the supposition of an initial or boundary condition, or a typical parameter for a given problem.
- Feed the supposed condition to the direct solution of the partial differential equation system, calculating in this way values of the dependent variable y . Here the output of the direct solution is a vector of values corresponding to the times in which the values of y are measured. This vector of solutions will be denoted as calculated and it will be represented as \hat{y} .
- Compare the calculated values \hat{y} with the values of the dependent variable y measured in consistent times with those for which \hat{y} was calculated.

The success of this approach is the mechanism for which the supposed condition is improved in the subsequent invocations of the first step. Optimization is the procedure to upgrade the suppositions of the conditions and in this case a genetic algorithm and differential evolution, whose characteristics are explained in the next section, will be used.

The most applied function in the measure of prediction error is the sum of the square error (SSE).

$$SSE(\hat{y}, \hat{\theta}) = \sum_{t_i=1}^{N_T} (y(t_i) - \hat{y}(t_i, \hat{\theta}))^2$$

where θ represents the parameters to be estimated and N_T is the total number of experimental data.

3.2 Evolutionary Computation for parameters estimation in adsorption models

Among the methods of Evolutionary Computation with practical applications are the Genetic Algorithm (GA) and the Differential Evolution Algorithm (DE). Fig. 3 shows the general structure of an Evolutionary Algorithm.

GA is based in three basic operators: selection, crossover or recombination and mutation. These algorithms should work in a wide interval of their parameters, but with differences in the efficiency, what indicates the importance of the designer's approach.

Another of the aspects to consider in a GA is the fitness function, which offers information about the quality from the possible solutions to a problem. Execution parameters and fitness function define the GA completely. Selection, recombination and mutation processes form a generation in the execution of a GA, and are executed until a satisfactory solution or a specified number of generations is reached.

In *DE* the crossover and mutation operators let variations from one generation to the next. Mutation adds a scaled, randomly sampled, vector difference to a third target vector (Price et al, 2005), while crossover combine each population vector with a mutant vector, according to the crossover probability, to generate a trial vector. Selection is used to choose if the parent or the offspring will survive to the next generation. The parameters to be defined in *DE* are the population size, the crossover probability and the scaling factor for the vector difference.

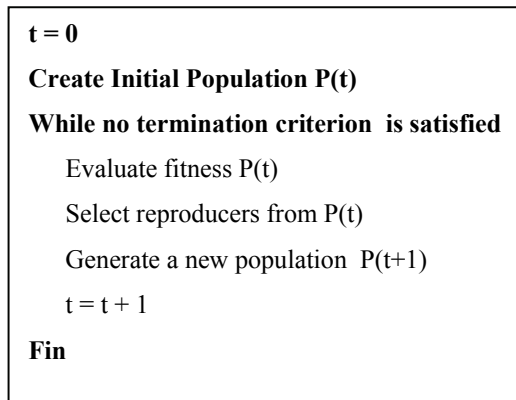


Figure 3: General structure of an Evolutionary Algorithm.

There are different versions for *DE* that differs in the way by which new solutions are generated. It's possible specify how the target vectors are chosen, how many vector differences and the type of crossover.

Considering the different models explained in section 2, the objective is to estimate the parameters related with protein mass transfer as ks and D_{ef} , as well as protein adsorption thermodynamics such as q_m and finally ϵ_p (particle porosity), in the chromatography models. The variable to be simulated is the protein concentration in the liquid phase (C_b). According to the previous algorithm of inverse solution of a partial differential equations system, the developed method for parameters estimation based on a *GA* (Irizar et al, 2008) is extended to *DE* (see Fig. 4).

Applying the finite differences method, the system to be analyzed is divided in discrete points or nodes. This division allows replacing derivatives by approximated expressions in differences.

As for *GA*, in *DE* each individual represents a solution to the outlined problem, that is, a possible group of parameters for model's structure selected previously. In the

```

Generate the synthetic experimental data;
Choose a chromatography model;
Perform a discretization in space and time;
While stopping criterion is not satisfied do
    Solution via finite differences;
    GA, DE
    Parameter Update;
End

```

Figure 4: Developed GA, DE - based parameter estimation method.

analyzed problem the real code is used (Fig. 5), which is the most common code in DE and the most convenient for this type of problem.

D_{ef}	q_m	ε_p	k_s
----------	-------	-----------------	-------

Figure 5: Parameter codification.

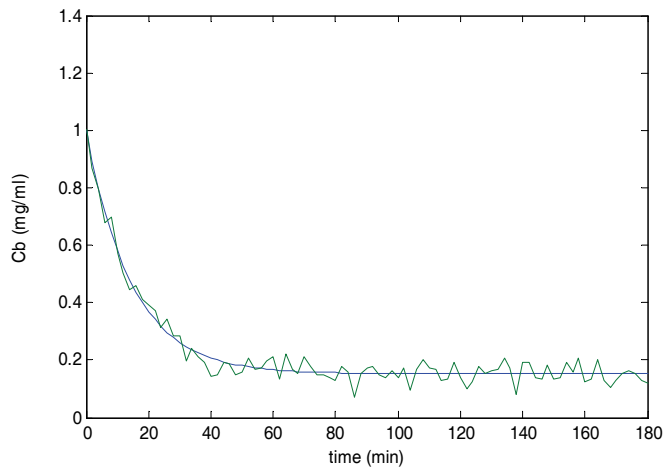


Figure 6: Fitting of noisy data in model with Eq. 3 and 4 for boundary conditions.

Each parameter is coded as a real value included in the intervals shown in Table 1. To put in practice the previous steps a group of functions was programmed with

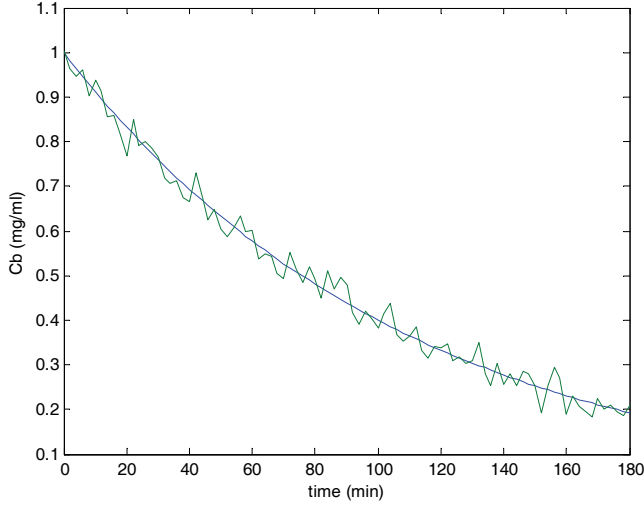


Figure 7: Fitting of noisy data in model with Eq. 9 and 10 for boundary conditions.

satisfactory results. In them were used as the fitness function the sum of the squared error, explained previously.

In the implemented *GA* as selection scheme was used stochastic uniform. As crossover operator was applied a uniform crossover, and a non uniform mutation operator. For the determination of *GA* control parameters like crossover and mutation probabilities, population size and stopping criterion, some amount of experimentation was required, based on some practical criteria.

Table 1: Intervals for parameter codification.

Parameter	Lower limit	Upper limit
$D_{ef}(m^2s^{-1})$	0	1×10^{-6}
$q_m(mg mL^{-1})$	50	100
ϵ_p	0.1	1
$k_s(ms^{-1})$	0	0.5

In the *DE* algorithm the strategy best/1/bin was applied, that is, the target vector was selected as the best individual from the population; one difference vector and the binomial crossover were used.

To simulate real measurements, synthetic data sets were generated running the direct solution of the model with a combination of parameters in the first stage and

adding noise later. Parameters values were identified for ten runs of *GA* and *DE* for each synthetic data set. For both algorithms the population size was 50 and 100 iterations were executed.

In Tables 2 and 3 the results of experiments to assess the performance of *GA* and *DE* are shown. With noisy data both are able to estimate and the model fits the data with accuracy. According to Table 2, mean values for estimate parameters D_{ef} , E_p and k_s are more near to the initial real values applying *DE*, although there is dispersion in them, with exception of parameter k_s . The time of execution is similar in both cases.

In Table 3 similar results can be seen: a better estimation of q_m by the genetic algorithm and almost the same time of execution.

In this nonlinear parameter estimation problem various solutions are obtained, although fitting to the generated curves is adequate. The best approximations to the original parameters are k_s in all cases.

4 Conclusions

A procedure for inverse solution of a system of partial differential equations based on the finite differences method and *GA* previously developed, that allowed parameter estimation of a chromatography process models represented by this system of equations, has been extended to *DE* in the current work.

The estimation of chromatography process parameters were obtained using a *GA* with uniform crossover in a crossover fraction of 0.5 and non uniform mutation operator with mutation probability equal to 0.9 for the model with the different boundary conditions explained, while in *DE* algorithm a scaling factor of 0.5 and a crossover probability of 0.8 were chosen.

The common practice in system identification of simulating models using validation data, and the visual and graphical comparison of correspondence between the real output and the predicted output (Ljung, 1999), (Soderstrom, 1994) was applied. Very small deviations indicate a good quality of the model. For example, Figs. 6 and 7 show *GA* and *DE* capacity to equalize generated concentration values for the studied models.

Another criteria in parameter estimation are statistical properties (Ljung, 1999) such as SSE to measure the total deviation of the fitness, R-square to determine how much successful is the adjustment in the explanation of the variation of data and finally the calculus of mean and variance of the real output, the estimated output and the modeling error. All of them demonstrated that a correspondence exists among predictions given by the model and “observations” for the system.

Table 2: Estimated parameters for *GA* and *DE* in model with Eq. 3 and 4 for boundary conditions.

GA with noisy data (3 %)					
Run	Def 5.37×10^{-7}	qm 70.5	Ep 0.62	ks 0.00892	Time (s)
1	0.19×10^{-6}	70.56	0.22	0.0086	80.91
2	0.50×10^{-6}	66.34	0.98	0.0088	82.38
3	0.56×10^{-6}	67.33	0.68	0.0087	81.71
4	0.16×10^{-6}	70.75	0.12	0.0086	79.66
5	0.43×10^{-6}	70.06	0.28	0.0085	79.66
6	0.51×10^{-6}	69.40	0.33	0.0086	79.40
7	0.20×10^{-6}	71.02	0.10	0.0085	78.87
8	0.45×10^{-6}	68.81	0.52	0.0086	79.27
9	2.74×10^{-15}	72.83	0.10	0.0085	91.91
10	0.82×10^{-6}	70.23	0.10	0.0086	79.60
Mean	3.87×10^{-7}	69.73	0.34	0.0086	81.33
Std. Dev.	2.43×10^{-7}	1.87	0.29	8.46×10^{-5}	3.68
DE with noisy data (3%)					
Run	Def 5.37×10^{-7}	qm 70.5	Ep 0.62	ks 0.00892	Time (s)
1	0.47×10^{-6}	66.31	0.57	0.0089	84.80
2	0.33×10^{-6}	66.30	0.81	0.0089	82.71
3	0.66×10^{-6}	66.31	0.41	0.0089	80.99
4	0.31×10^{-6}	66.29	0.86	0.0089	79.96
5	0.39×10^{-6}	66.30	0.68	0.0089	79.29
6	0.14×10^{-6}	66.32	0.19	0.0089	80.77
7	0.49×10^{-6}	66.31	0.55	0.0089	82.19
8	0.31×10^{-6}	66.29	0.87	0.0089	83.44
9	1.02×10^{-6}	66.32	0.26	0.0089	81.32
10	0.85×10^{-6}	66.32	0.31	0.0089	82.49
Mean	6.26×10^{-7}	66.31	0.55	0.0089	81.80
Std. Dev.	3.63×10^{-7}	0.01	0.25	1.94e-011	1.65

The experiment results demonstrated the feasibility of these techniques for the solution of this problem of parameters estimation. The combination *GA*, *DE* – numeric method can be applied to estimate parameters of other models of complex process, however, it would be convenient to carry out the structural identifiability analysis previously, to determine if it is possible to obtain a unique set of parameters in the analyzed models.

Table 3: Estimated parameters for *GA* and *DE* in model with Eq. 9 and 10 for boundary conditions.

GA with noisy data (3 %)					
Run	Def 5.37×10^{-7}	qm 70.5	Ep 0.62	ks 0.00892	Time (s)
1	0.90×10^{-6}	53.13	0.89	0.0089	94.87
2	0.21×10^{-6}	64.21	0.16	0.0089	93.97
3	0.20×10^{-6}	95.80	0.17	0.0089	93.83
4	0.72×10^{-6}	69.00	0.33	0.0089	95.87
5	0.26×10^{-6}	78.04	0.18	0.0089	94.94
6	0.40×10^{-6}	66.84	0.28	0.0089	95.65
7	0.85×10^{-6}	64.74	0.54	0.0089	94.94
8	0.40×10^{-6}	94.20	0.73	0.0089	94.61
9	0.23×10^{-6}	68.87	0.24	0.0089	95.05
10	0.66×10^{-6}	54.66	0.22	0.0089	94.43
Mean	3.28×10^{-7}	70.95	0.37	0.0089	94.79
Std. Dev.	3.68×10^{-7}	14.52	0.25	1.36×10^{-5}	0.67
DE with noisy data (3%)					
Run	Def 5.37×10^{-7}	qm 70.5	Ep 0.62	ks 0.00892	Time (s)
1	0.94×10^{-6}	95.93	0.32	0.0089	90.32
2	0.40×10^{-6}	113.96	0.13	0.0089	90.77
3	0.25×10^{-6}	122.88	1.56	0.0089	90.80
4	0.10×10^{-6}	57.10	0.90	0.0089	90.43
5	0.46×10^{-6}	88.89	1.00	0.0089	90.21
6	0.82×10^{-6}	48.95	0.24	0.0089	91.08
7	0.11×10^{-6}	75.39	1.25	0.0089	91.29
8	0.43×10^{-6}	84.76	1.10	0.0089	90.93
9	0.46×10^{-6}	63.92	0.03	0.0089	97.01
10	0.96×10^{-6}	66.98	0.26	0.0089	90.69
Mean	6.44×10^{-7}	81.88	0.68	0.0089	91.35
Std. Dev.	3.87×10^{-7}	24.15	0.54	1.18×10^{-11}	2.01

Acknowledgement: The authors acknowledge CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior do Brazil, and MES, Ministério da Educação Superior de Cuba. AJSN and LDTC acknowledge also CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, and FAPERJ, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, both from Brazil.

References

Altenhöner, U.; Meurer, M.; Strube, J.; Schmidt-Traub, H. (1997): Parameter estimation for the simulation of liquid chromatography. *Journal of Chromatography* 769, 5969.

Amirov, A. ; Gölgeleyen, F. ; Rahmanova, A. (2009): An Inverse Problem for the General Kinetic Equation and a Numerical Method. *ICCES*, vol.12, no.4, pp.125-135

Bäck, T. (1997): *Handbook Of Evolutionary Computation*, Oxford: Oxford University Press.

Blanch, H. W; Clark, D.S. (1997): *Biochemical Engineering*, Marcel Dekker Inc.

Chang, C.; Liu, C. ; Chang, J. (2007): The Lie-Group Shooting Method for Quasi-Boundary Regularization of Backward Heat Conduction Problems *ICCES*, vol.3, no.2, pp.69-80, 2007.

De Vasconcellos, J.F.V.; Silva Neto, A.J.; and Santana, C.C. (2003): An inverse mass transfer problem in solid-liquid adsorption systems, *Inverse Problems in Engineering* 11(5), 391-408.

De Vasconcellos, J.F.V.; Silva Neto, A.J.; Santana, C.C.; Soeiro, F.J.C.P. (2002): Parameter estimation in adsorption columns with stochastic global optimization methods, in *4th International Conference on Inverse Problems in Engineering, Rio de Janeiro, Brazil*.

Goldberg, D.E. (1989): *Genetic Algorithms In Search, Optimization, And Machine Learning*, Reading, MA: Addison-Wesley.

Gu, T. (1995): *Mathematical Modelling and Scale-Up of Liquid Chromatography*, Springer Verlag New York.

Horstmann, B. J.; Chase, H. A. (1989): Modelling the affinity adsorption of immunoglobulin g to protein a immobilized to agarose matrices, *Chem. Eng.Res. Des.* 67.

Irizar, M.; Camara, L.D.; Silva Neto, A.J; Llanes, O.; Herrera, F.; Santana, C. (2008): A Hybrid genetic Algorithm for the Inverse Solution of a Chromatography Model with Different Boundary Conditions, *Proceedings of the Iberian Latin American Congress on Computational Methods in Engineering (CILAMCE 2008)*.

Kanevce, G., Kanevce, L. Mitrevski, V., Dulikravich, G. (2008): Inverse Estimation of Moisture Diffusivity by Utilizing Temperature Response of a Drying Body. *ICCES*, vol.8, no.1, pp.1-6.

Karr, C.L.; Yakushin, I.; Nicolosi, K. (2000): Solving inverse initial-value, boundary-value problems via genetic algorithm. *Engineering Applications of Artificial Intel-*

ligence 13, 625-633.

Levenberg, K. (1944): A Method for the Solution of Certain Non-linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2(2):164–168.

Liu, C; Liu, L ; Hong, H. (2007): Highly Accurate Computation of Spatial-Dependent Heat Conductivity and Heat Capacity in Inverse Thermal Problem *CMES*, vol.17, no.1, pp.1-18.

Ljung, L. (1999a): Model validation and model error modeling, *Technical Report LiTH-ISY-R-215, Lund University, Sweden.*

Ljung, L. (1999b): *System identification Theory for the user*, 2nd edn, Prentice Hall ,Upper Saddle River, N.J.

Marin, L. (2009): Boundary Reconstruction in Two-Dimensional Functionally Graded Materials Using a Regularized MFS. *CMES: Computer Modeling in Engineering & Sciences*, vol.46, no.3, pp.221-252.

Marquardt, D. (1963): An Algorithm for the Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal of Applied Mathematics*, 11(2):431–441.

Michalewicz, Z. (1992): *Genetic Algorithms + Data Structures = Evolution Programs*, Berlin: Springer-Verlag.

ONeil, P.V. (1983): *Advanced Engineering Mathematics*, Belmont, CA:Wadsworth Publishing Company.

Persson, P.; Nilsson, B. (2001): Parameter estimation of protein chromatographic processes based on breakthrough curves, in D.Dochain and M.Perrier, eds, *Proceedings of the 8th International Conference on Computer Applications in Biotechnology*.

Price, K. V. ; Storn, R.; Lampinen, J. A. (2005): *Differential Evolution - A Practical Approach to Global Optimization*. Springer – Natural Computing Series.

Soderstrom, T. ; Stoica, P. (1994): *System Identification*, Prentice Hall International, Hemel Hempstead, Paperback Edition.17

Storn, R.; Price, K. (1995): Differential evolution—A simple and efficient adaptive scheme for global optimization over continuous spaces, *Int. Comput. Sci. Inst., Berkeley*, vol. 12, pp 1-16.

Tarantola, A. (2005): *Inverse Problem Theory and Model Parameter Estimation*, SIAM.