RIELAC
Revista de Ingeniería Electrónica,
Automática y Comunicaciones

Application of Genetic Algorithms for Parameter Estimation in Liquid Chromatography

Reynier Hernández Torres

Mirtha Irizar Mesa

Leoncio Diogenes T. Camara

Antonio José da Silva Neto

Orestes Llanes Santiago

RESUMEN / *ABSTRACT*

In chromatography, complex inverse problems related to the parameters estimation and process optimization are presented. Metaheuristics methods are known as general purpose approximated algorithms which seek and hopefully find good solutions at a reasonable computational cost. These methods are iterative process to perform a robust search of a solution space.

Genetic algorithms are optimization techniques based on the principles of genetics and natural selection. They have demonstrated very good performance as global optimizers in many types of applications, including inverse problems.

In this work, the effectiveness of genetic algorithms is investigated to estimate parameters in liquid chromatography.

Keywords: Chromatography, Genetic Algorithms, Inverse problems, Parameter estimation.

En el proceso de cromatografía se presentan problemas inversos relacionados con la estimación de parámetros y la optimización del proceso.

Los métodosmetaheurísticos son conocidos como algoritmos aproximados de propósito general que buscan y encuentran satisfactoriamente buenas soluciones con un costo computacional razonable. Estos métodos son procesos iterativos que realizan una búsqueda robusta en un espacio de solución.

Los algoritmos genéticos son técnicas de optimización basadas en los principios de la genética y la selección natural. Han demostrado un buen rendimiento como optimizadores globales en varios tipos de aplicaciones, incluyendo los problemas inversos.

En este trabajo se analiza la efectividad de los algoritmos genéticos para estimar parámetros en la cromatografía líquida de columna.

Palabras Clave: Algoritmos Genéticos, Estimación de Parámetros, Problemas Inversos, Cromatografía.

Aplicación de Algoritmos Genéticos en la Estimación de Parámetros en Cromatografía Líquida

R. Hernández Torres, M. Irizar Mesa, L. Diogenes T. Camara, A. J. da Silva Neto, O. Llanes Santiago

_____

INTRODUCTION

Liquid chromatography (LC) is a common separation method very important in chemistry, pharmaceutical and biotechnological industries. It is used to isolate one or more compounds in a mixture. In LC, a sample of molecules are injected into a column of adsorbing porous material (stationary phase), frequently silica, and a liquid (mobile phase) is pumped through the column, and the different kinds of molecules are distributed differently between the two phases[1-3]. The structureof a chromatographic column is shown in figure 1.
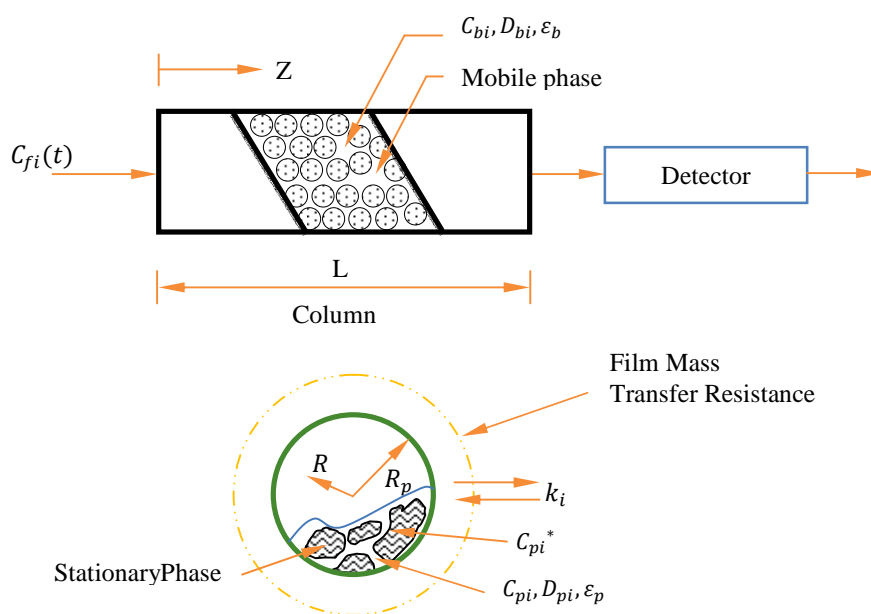

Figure 1 Anatomy of a chromatographic column.[4]

In LC, numerical values of several parameters have to be well estimated. Since geometrical and physical characteristics of the column and of the adsorbent particles are usually known, other parameters like adsorption equilibrium, mass transport and axial mixing parameters have to be estimated, usually through experimental measurements.

Parameters like film mass transfer coefficient, axial dispersion coefficient and effective diffusivity are often not available from literature, or not easily measured by experiments. Nevertheless, they can be estimated with certain accuracy. Opportunely, rate models are not very sensitive to mass transfer parameters. Errors up to a certain degree do not affect the outcome largely.[4]

MATHEMATICAL MODELING OF LIQUID CHROMATOGRAPHY

Consider a fixed-bed adsorption column packed with uniform porous, spherical and solid adsorbents. The process is supposed isothermal and there is no concentration gradient in the radial direction of the column. Another assumption is that there exist local equilibrium for each component between the pore surface and the liquid phase in the macropores inside particles.

Under these suppositions, two differential equations are used to modeling the process of chromatography in this work.[4,6]

Continuity Equation in the Flowing Mobile Phase

The flowing mobile phase equation is:

$$-D_{bi}\frac{\partial^2 C_{bi}}{\partial Z^2} + v\frac{\partial C_{bi}}{\partial Z} + \frac{\partial C_{bi}}{\partial t} + \frac{(1-\varepsilon_b)}{\varepsilon_b}\frac{3k_i}{R_p}\left(C_{bi} - C_{pi,R=R_p}\right) = 0$$

where C_{bi} is the bulk-fluid phase concentration of component i, D_{bi} is the axial dispersion coefficient of component i, Z is the axial coordinate, v is the interstitial velocity, ε_b is the bed void volume fraction, k_i is the film mass transfer coefficient of component i and R_p is the particle radius.

Continuity Equation inside the Macropores

The particle phase equation in spherical coordinates is:

$$\left(1-\varepsilon_p\right)\frac{\partial C_{pi}^*}{\partial t} + \varepsilon_p\frac{\partial C_{pi}}{\partial t} + \varepsilon_p D_{pi}\left[\frac{1}{R^2}\frac{\partial}{\partial R}\left(R^2\frac{\partial C_{pi}}{\partial R}\right)\right] = 0$$

where C_{pi} is the concentration of component i in the stagnant fluid phase inside the particle macropores, C_{pi}^* is the concentration of component i in the solid phase of particle (based on unit volume of particle skeleton), ε_p is the particle porosity, D_{pi} is the effective diffusivity of component i, porosity not included and R is the radial coordinate for particle.

Boundary (B.C.) and initial(I.C.) conditions

The column is initially equilibrated and the initial state may be represented by the following equations:

At $t = 0$,

$$C_{bi} = C_{bi}(0,Z)$$

$$C_{pi} = C_{pi}(0,R,Z)$$

At $Z = 0$,

$$\frac{\partial C_{bi}}{\partial Z} = \frac{v}{D_{bi}}\left(C_{bi} - C_{fi}(t)\right)$$

At $Z = L$,

$$\frac{\partial C_{bi}}{\partial Z} = 0$$

At $R = 0$,

$$\frac{\partial C_{pi}}{\partial R} = 0$$

At $R = R_p$

$$\frac{\partial C_{pi}}{\partial R} = \frac{k_i}{\varepsilon_p D_{pi}}\left(C_{bi} - C_{pi,R=R_p}\right)$$

Dimensionless equations

Let us define the following dimensionless terms:

$$c_{bi} = C_{bi}/C_{oi}$$

$$c_{pi} = C_{pi}/C_{oi}$$

$$c_{pi}{}^* = C_{pi}{}^*/C_{0i}$$

$$\tau = \frac{vt}{L}$$

$$r = \frac{R}{R_p}$$

$$z = \frac{Z}{L}$$

$$Pe_{Li} = \frac{vL}{D_{bi}}$$

$$Bi_i = \frac{k_i R_p}{\varepsilon_p D_{pi}}$$

$$\eta_i = \frac{\varepsilon_p D_{pi} L}{R_p 2v}$$

$$\xi_i = \frac{3 Bi_i \eta_i (1 - \varepsilon_b)}{\varepsilon_b}$$

The Peclet number (Pe_{Li}) reflects the ratio of the convection rate to the dispersion rate while the Biot number (Bi_i) reflects the ratio of the external film mass transfer rate to the intraparticle diffusion rate. Parameters η_i and ξ_i are dimensionless constants.

The equations model becomes:

$$-\frac{1}{Pe_{Li}} \frac{\partial^2 c_{bi}}{\partial z^2} + \frac{\partial c_{bi}}{\partial z} + \frac{\partial c_{bi}}{\partial \tau} + \xi_i \left(c_{bi} - c_{pi,r=1}\right) = 0$$

$$\left(1 - \varepsilon_p\right) \frac{\partial c_{pi}{}^*}{\partial \tau} + \varepsilon_p \frac{\partial c_{pi}}{\partial \tau} - \eta_i \left[\frac{1}{r^2} \frac{\partial}{\partial r}\left(r^2 \frac{\partial c_{pi}}{\partial r}\right)\right] = 0$$

Using the Langmuir isotherm, the model becomes nonlinear. For such a nonlinear multicomponent model, there is no analytical solution. These equations must be solved numerically.

Model Solution

A robust and efficient numerical procedure has been developed byGu[4,6]to solve the system. First, the model is converted to a dimensionless model. Later, the spatial axes, $z$and r, are discretized.

The bulk-fluid phase equation is discretized using the finite elements method (FE) and the particle equation using orthogonal collocation method (OC).

Solution to the ODE system

R. Hernández Torres, M. Irizar Mesa, L. Diogenes T. Camara, A. J. da Silva Neto, O. Llanes Santiago

_____

Finally, if Nz quadratic nodes are used for z-axis in the bulk fluid phase equation and Nr interior OC points are used for the r-axis in the particle phase equation, the discretization procedure gives a total of $NsNz(Nr + 1)$ equations that are solved simultaneously by any of the stiff ordinary differential equations solver (function *ode15s* from MATLAB is used in the simulation).

PARAMETER ESTIMATION

The proposed technique assumes a first principle model structure where some of the parameters are unknown.

The identification objective consists of finding a set of parameters ($\hat{\zeta}$) which minimizes differences between the real response vector of the process ($y(t)$) and the model output vector ($\hat{y}(t)$). A cost function for a minimization in a time interval$[0, t_e)$, known as Sum of Squared Errors, is:

$$J(\zeta) = \sum_{j=1}^{N}\left(y_i(j) - \hat{y}_i(j)\right)^2$$

where$N$ is the number of samples, $y_i(j)$the component i of vector $y(j)$ and $\hat{y}_i(j)$ the component i of vector $\hat{y}(j)$.

Once the cost function is established, it is necessary to choose an optimization technique. For complicated models (nonlinearities, saturation, high order, etc.) the optimization problem can be very complicated, and then a powerful optimization technique is required.Genetic Algorithm is a very good candidate for this role, even more so if there is no restriction on computational cost (it is an off-line identification).

PARAMETER ESTIMATION BASED ON GENETIC ALGORITHMS

Genetic algorithms method (GA) is an optimization technique based on the principles of genetics and natural selection. It was pioneered by J. Holland and his collaborators in the years 1960 and 1970[7]. This technique is based on applying natural selection laws onto a population to achieve individuals that are better adjusted to their environment.

A population is a set of points in the search space. Each individual of the population represents a point in that space by means of his chromosomes. The adaptation degree of the individuals is given by the objective function.

The evolution mechanism of individuals is achieved by genetic operators. The usual operators are:

- Selection: its main goal consists of selecting the chromosomes to integrate the next population (these would depend on the cost function for each individual).
- Crossover: new individuals are generated and integrated by combining the chromosomes of two individuals.
- Mutation: randomly varying of some part of the chromosome of an individual in the population generates new individuals.

GA implementation have some inherent parameters such as size population, number of generations, crossover and mutation probability. There are others parameters that have to be determined before the operation, but definitive general approaches do not exist. These algorithms should work in a wide interval of their parameters, but with differences in the efficiency. Another aspect to consider in genetic algorithms is the fitness function, which offers information about the quality from the possible solutions to a problem. Execution parameters and fitness function define the genetic algorithms completely.

A possible method for parameters estimation based on genetic algorithms, according to the previous algorithm of inverse solution of a partial differential equations system, can be:

- Generate an initial population of individuals in the genetic algorithms, which represent possible values of the parameters.
- Execute the solution algorithm of the partial differential equation for each individual in a sequential way.
- Compare the result of solution algorithm with the simulated concentration in different instants, assigning to each individual a fitness value, according to the grade of correspondence.
- Select the best individuals to create the next generation, considering the metric established previously.
- When the stopping criterion is satisfied, stop the genetic algorithms.

R. Hernández Torres, M. Irizar Mesa, L. Diogenes T. Camara, A. J. da Silva Neto, O. Llanes Santiago

_____

Parameter estimation based on Genetic Algorithms in Liquid Chromatographic model

In chromatography models, the objective is to estimate the parameters related with protein mass transfer as k, D_b and D_p and particle porosity (ε_p). Protein concentration in the liquid phase (c_b) is the variable to be simulated.

In the genetic algorithms, each individual (group of parameters selected) represents a solution to the defined problem. In this case, the codification of the parameters is $[k \quad D_b \quad D_p \quad \varepsilon_p]$.

RESULTS AND DISCUSSION

All calculations were performed on a PC with a 1,67 GHz Intel Core 2 Duo processor and 1 GB of internal memory. The solver and estimation routines were implemented in MATLAB 7.6 (R2008a), over Windows 7.

In the GA implemented, a tournament selection scheme by two individuals was used. A uniform crossover operator is applied with a probability of 0,95 and a uniform mutation operator where each individual has a probability 0,05 of being mutated. Population size is 20.

For simulation, synthetic data were generated running the direct solution of the model. Parameters values were estimated in ten runs of the algorithm with equal operation characteristics. The results are presented in Table 1 and Table 2. In figures 1 and2 is shown breakthrough curves resulting for average estimated values.

Table 1.Best, worst and average estimated values with noiseless data.

	$k(cm/min)$	$D_b(cm^2/min)$	$D_p(cm^2/min)$	ε_p	Fitnessfunction
Exact	_0,01670_	_0,000200_	_0,009600_	_0,400_	_0,000_
Bestparameter set	0,01699	0,000228	0,008819	0,411	$4,339 \cdot 10^{-4}$
Worstparameter set	0,01903	0,000306	0,009656	0,381	$4,884 \cdot 10^{-3}$
Average	0,01798	0,000218	0,009203	0,401	$3,170 \cdot 10^{-4}$

Table 2.Best, worst and average estimated values with 2% noisy data.

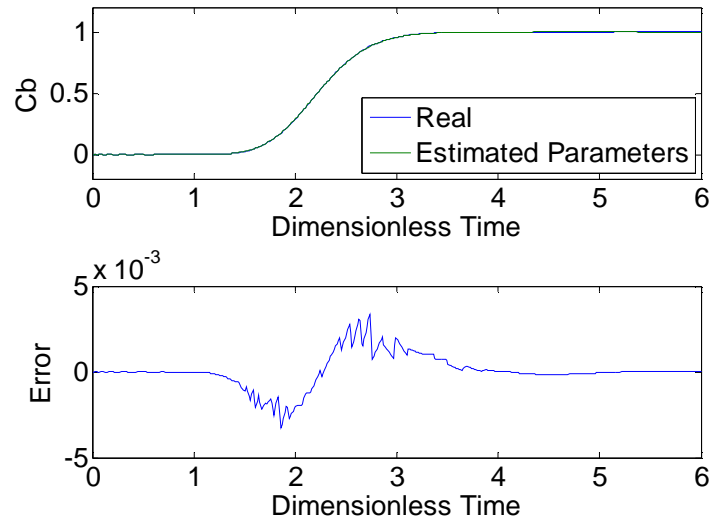|	$k(cm/min)$	$D_b(cm^2/min)$	$D_p(cm^2/min)$	ε_p	Fitnessfunction
2% Noisydata	_0,01670_	_0,000200_	_0,009600_	_0,400_	_$6,286 \cdot 10^{-2}$_
Bestparameter set	0,01971	0,000069	0,007991	0,393	$1,881 \cdot 10^{-3}$
Worstparameter set	0,01316	0,000044	0,006775	0,546	$1,408 \cdot 10^{-1}$
Average	0,01577	0,000132	0,008652	0,424	$3,576 \cdot 10^{-3}$

_____

Figure 2Breakthrough curves (up) and error (down) for real and estimated parameters with GA.
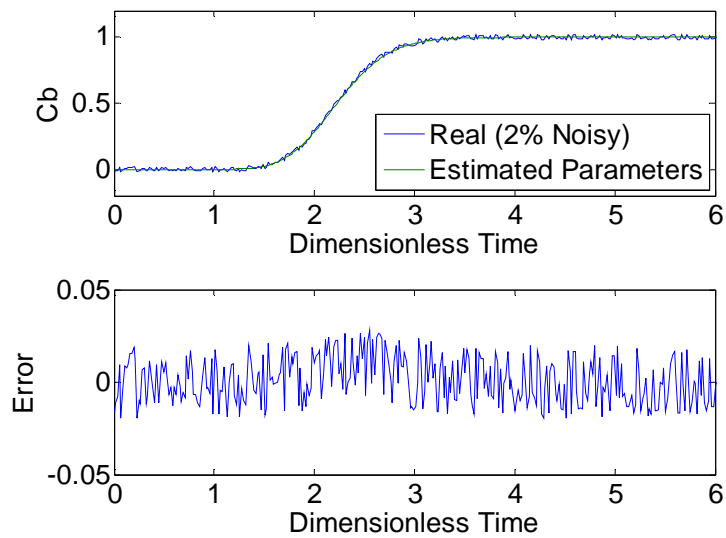

Figure 3 Breakthrough curves (up) and error (down) for real and estimated parameters with GA. (2% noisy data)

The time was approximately 30 minutes per run, with 390 evaluations of the fitness function. The method obtains a relatively accurate.

An improvement of the solution is to use a hybrid method, implying a considerable lower computational effort than simple genetic algorithm, with accurate results.

-19-

R. Hernández Torres, M. Irizar Mesa, L. Diogenes T. Camara, A. J. da Silva Neto, O. Llanes Santiago

_____

CONCLUSIONS

Liquid chromatography is a complex process, where numerical values of several parameters have to be well estimated. Genetic algorithms are efficient global optimizers, and they demonstrate their strength in parameters estimation.

In this work, a simple genetic algorithm was used to estimate four parameters in the liquid chromatography model. Its implementation was made in MATLAB. Ten runs of the method were made, and relatively accurate results were obtained.

In further works, hybrid methods will be applied to decrease the computational effort.

REFERENCES

1. **J.G. DORSEY, ET AL.**, "Liquid Chromatography: Theory and Methodology," Analytical Chemistry, vol. 70, 1998, pp. 591-644.
2. **V. GROSFILS**, "Modelling and parametric estimation of simulated moving bed chromatographic processes (SMB)," 2009.
3. **G. GUIOCHON**, "Preparative liquid chromatography: Review," Journal of Chromatography A, vol. 965, 2002, pp. 129-161.
4. **C. LAZO**, "Simulation of Liquid Chromatography and Simulated Moving Bed (SMB) Systems," 1999.
5. **T. GU, ET AL.**, "Modeling of Nonlinear Multicomponent Chromatography," Advances in Biochemical Engineering Biotechnology, vol. 49, 1993, pp. 45-71.
6. **M. IRIZAR**, "Identificación Basada en Algoritmos Genéticos con Aplicacióna un Proceso de Cromatografía," 2006. Sc. Dr. thesis
7. **T. GU, ET AL.**, "New Approach to a General Nonlinear Multicomponent Chromatography Model," AIChE Journal, vol. 36, 1990, pp. 784-788.

AUTHORS

Reynier Hernández Torres, Engineer in Automatic. Assistant of theDepartment of Automation and Computers. At present, he is working towards the M. Sc. degree.His research interests are parameter estimation in chromatography and metaheuristics methods.
reynier@electrica.cujae.edu.cu

Mirtha Irizar Mesa, Engineer in Electronics, Master in Applied Informatics and Doctor in Technical Sciences. She is Titular Professor of Department of Automation and Computers. Her fundamental research interest is the solution of direct and inverse problems based on computational intelligence.
mirtha@electrica.cujae.edu.cu

LeôncioDiogenes TavaresCâmara,graduated in Chemical Engineering, Master in Chemical Engineering, Doctor in Chemical Engineering, teacher and researcher of the Department of Mechanical Engineering and Energy at Universidade do Estado do Rio de Janeiro.His fundamental research interest is the solution of direct and inverse problems.
dcamara@iprj.uerj.br

Antônio José da Silva Neto,graduated in MechanicalEngineering, Master in Mechanical Engineering, PhilosophyDoctor in Mechanical Engineering, teacher and researcher ofthe Department of Mechanical Engineering and Energy atUniversidade do Estado do Rio de Janeiro. His fundamental research interest is the solution of direct and inverse problems.ajsneto@iprj.uerj.br

Orestes Llanes Santiago, graduated in Electrical Engineering, Master in Automatic Control, Philosophy Doctor in Applied Sciences, teacher and researcher of the Department ofAutomation and Computers. His fundamental research interest is the solution of inverse problems and fault diagnosis.
orestes@electrica.cujae.edu.cu