



Robust estimation of thermodynamic parameters (ΔH , ΔS and ΔC_p) for prediction of retention time in gas chromatography – Part II (Application)



Carlos Alberto Claumann^a, André Wüst Zibetti^{c,*}, Ariovaldo Bolzan^a,
Ricardo A.F. Machado^a, Leonel Teixeira Pinto^b

^a Laboratório de Controle de Processos, Departamento de Engenharia Química e de Alimentos, Centro Tecnológico, Universidade Federal de Santa Catarina (UFSC), PO Box 476, Florianópolis, SC 88010-970, Brazil

^b Laboratório de Neuroengenharia Computacional, Departamento de Engenharia Química e de Alimentos, Centro Tecnológico, Universidade Federal de Santa Catarina (UFSC), PO Box 476, Florianópolis, SC 88010-970, Brazil

^c Departamento de Informática e Estatística (INE), Centro Tecnológico, Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC 88010-970, Brazil

ARTICLE INFO

Article history:

Received 11 August 2015

Received in revised form 19 October 2015

Accepted 20 October 2015

Available online 28 October 2015

Keywords:

Parameter estimation

Initialization method

Gas chromatography

Thermodynamic properties

Retention time

Optimization

ABSTRACT

For this work, an analysis of parameter estimation for the retention factor in GC model was performed, considering two different criteria: sum of square error, and maximum error in absolute value; relevant statistics are described for each case. The main contribution of this work is the implementation of an initialization scheme (specialized) for the estimated parameters, which features fast convergence (low computational time) and is based on knowledge of the surface of the error criterion. In an application to a series of alkanes, specialized initialization resulted in significant reduction to the number of evaluations of the objective function (reducing computational time) in the parameter estimation. The obtained reduction happened between one and two orders of magnitude, compared with the simple random initialization.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

For predicting compound retention time within capillary columns in gas chromatography (GC), it is necessary to define a model for retention factor $k(T)$. In the most common approach, the relationship between retention factor and thermodynamic parameters (ΔH , ΔS) [1–3] or (ΔH , ΔS and ΔC_p) [4–6] is considered, as is the temperature program.

In many studies found within the available literature [1,7,8] the estimation of parameters of the retention factor in GC model is performed by way of experimenting with different isothermal conditions, resulting in a linear regression problem that can be solved by least squares.

In other cases [9], although non-isothermal temperature programming is utilized, the method for solving the non-linear

regression is not described in detail, and the optimization method used is not defined, nor is presented any information about convergence and the objective function.

This work, which is Part II of a series, demonstrates performances of different numerical methods in optimizing two different error criteria used for estimating parameters: the sum of square error (SSE) and the maximum error for absolute value (MEAbs).

Optimizations of each error criterion were tested using two types of initialization: one simple, and another one, called *specialized* – which is what is being proposed here, departing from data and information gathered in Part I, which makes use of the format of the surface of the error criterion due to relevant thermodynamic parameters.

According to Part I, from knowledge of the surface of the error function, it is possible to place the points in the area with the highest possibility for the global optimum to be found upon. This result is in line with the heuristic in optimization, that is, that there is a higher probability of convergence toward the global optimum the closer the initial value is to the location of the optimum [10]. Additionally a reduction in the number of objective function evaluations is expected due to the proximity of the initial guess with respect to the optimum point.

* Corresponding author.

E-mail addresses: carlos.claumann@prograd.ufsc.br (C.A. Claumann), azibetti@gmail.com (A. Wüst Zibetti), ariovaldo.bolzan@ufsc.br (A. Bolzan), ricardo.machado@ufsc.br (R.A.F. Machado), leonel.t.pinto@ufsc.br (L.T. Pinto).

The main objective of this series of works (Parts I and II) is to identify an efficient way of initializing parameter estimation; nonetheless, attention is also given to the physically interpretation of the estimated parameters, as well as to how the validation of the model of retention factor in the prediction of retention time, for temperature programs different from those used in estimation, is done.

The experimental data described in Part I and used for parameter estimation was obtained from retention time measurements for a series of alkanes. Additionally, an equation was determined for the variation of dead time of the mobile phase with the temperature. Furthermore, additional data for validation of the method is shown in Section 5.

2. Basic description of the prediction model for time retention

As described in reference [11], there are several equations which may be utilized to represent the influence of the retention factor with temperature. In this study Eqs. (1)–(3) were used as templates for the retention factor, seeing as its parameters (ΔH and ΔS) have thermodynamic meaning and allow correction of these with temperature by the introduction of the parameter ΔC_p .

$$\Delta H(T(t)) = \Delta H(T_0) + \Delta C_p(T(t) - T_0) \quad (1)$$

$$\Delta S(T(t)) = \Delta S(T_0) + \Delta C_p(\ln T(t) - \ln T_0) \quad (2)$$

$$\ln k(T(t)) = -\frac{\Delta H}{RT(t)} + \frac{\Delta S}{R} - \ln \beta \quad (3)$$

where R is the ideal gas constant and β is the ratio between the volume of the column phases.

Parameters ΔH , ΔS and ΔC_p are associated with the solute's thermodynamic transition between the stationary and mobile phases. It is possible to check the consistency of the estimated parameters, especially when comparing different compounds, but from the same functional group.

Once the expressions for dead time and retention factor are defined, retention time can be determined by solving the well-known equation in GC, according to Eq. (4) (see Ref. [8]):

$$1 = \int_0^{t_{r,i}^{(p)}} \frac{dt}{t_M(t) \cdot [1 + k_i(t)]} \quad (p = 1 \dots N_{\text{Exp}}) \quad (4)$$

where N_{Exp} is the number of temperature programs, and $t_{r,i}^{(p)}$ or $t_{r_{\text{mod}},i}^{(p)}$ the retention time as predicted by the model in the case of the p th temperature program $T(t)^{(p)}$, of the i th component.

For solving Eq. (4), the classic procedure described in Ref. [7] was used, where the peak position over time (trajectory) is calculated at small time steps until reaching the end of the column.

3. Description of the methodology for estimating parameters

For the estimation of thermodynamic parameters it is necessary to define a performance criterion, that is, an objective function (error criterion to be minimized) and an optimization method. Depending on the chosen error criterion, a specific method can be selected by the modeler for optimization. In this paper we compared two error criteria (sum of square error and maximum error in absolute value), with two different initialization procedures; the first considers only a random initial guess in the space of the search (called 'simple' procedure), and the other initialization process is based on the division of the error surface in different regions. The error surface was described in Part I of this series. The initial guess,

based off the knowledge of the most likely region where the global optimum can be found, was called 'specialized' initialization.

3.1. Definition of used performance ratios

Some statistical data were established as comparison criteria between the methods for estimating parameters.

In this work, parameters from the retention of coefficient $k_i(T)$ were optimized, considering two different performance rates for an arbitrary i component:

(1) the sum of the square error (SSE), Eq. (5):

$$\text{SSE} = \sum_{p=1}^{N_{\text{Exp}}} \left[t_{r_{\text{exp}},i}^{(p)} - t_{r_{\text{mod}},i}^{(p)}(\Delta H, \Delta S, \Delta C_p, T) \right]^2 \quad (5)$$

(2) the maximum error in absolute value (MEAbs), Eq. (6):

$$\text{MEAbs} = \max \left| t_{r_{\text{exp}},i}^{(p)} - t_{r_{\text{mod}},i}^{(p)}(\Delta H, \Delta S, \Delta C_p, T) \right| (p = 1, \dots, N_{\text{Exp}}) \quad (6)$$

The sum of the square error (SSE) is directly related to the coefficient of determination R^2 , wherein the minimization of the former implies the maximization of the latter [12]. R^2 is a standard approach measure, contrary to what happens with the SSE, which is a discrepancy between the measured data and the estimated model. However, the analyzes described in this paper are shown in terms of R^2 values, even though it was the SSE that which was the minimized criterion. R^2 values of the order of 99.9 or 99.99 usually indicate a good approximation

In the case of the maximum error in absolute value (MEAbs), instead of SSE or R^2 , a well-defined meaning can be found, because it is known, from the value of MEAbs, the worst case of all temperature settings for a given component.

Considering that retention times obtained on Part I of this series were of the order of minutes, prediction errors of t_r around a few seconds are considered acceptable.

Even though the highest relative error [%] in MEP module has not been directly optimized in parameter estimation, it was also calculated according to Eq. (7). MEP values lower than 1% are described in literature as indicators for good prediction performance.

$$\text{MEP} = \max \left| \frac{t_{r_{\text{exp}},i}^{(p)} - t_{r_{\text{mod}},i}^{(p)}(\Delta H, \Delta S, \Delta C_p, T)}{t_{r_{\text{exp}},i}^{(p)}(\Delta H, \Delta S, \Delta C_p, T)} \right| \times 100 \quad (p = 1, \dots, N_{\text{Exp}}) \quad (7)$$

It is noteworthy that for the evaluation of any one of the objective functions described, the model for retention time prediction must be solved N_{Exp} times.

3.2. Computational implementation

The programming of parameter estimation was implemented in *Matlab*[®] and it corresponded to the following data entry steps: initialization, application of optimization methods, and treatment of the results. For optimizers, the functions contained in *Matlab*[®] Optimization Toolbox were used.

Functions *fmincon* and *fminimax* by *Matlab*[®] are used for the optimization of multivariable problems, subject to linear and non-linear restrictions. The parameters of those were kept in their default values. For the present case, solely the restrictions for the search range for parameters ΔH , ΔS and ΔC_p in intervals $[-200$ to 0 kJ/mol], $[-200$ to 0 J/(mol K)] and $[0$ to 200 J/(mol K)], respectively, were used.

3.2.1. Minimization of the SSE

To minimize the sum of square error SSE the *fmincon* function was used. The *fmincon* function is an optimization method based on gradient designed to work with continuous objective function subject to constraints (as is the case with SSE). Further information on how *fmincon* works can be found directly on the online help section of *Matlab*[®] as well as on Refs. [13–18].

3.2.2. Minimization of the MEAbs

The *fminimax* function was used for the minimization of the maximum absolute error. The *fminimax* function transforms the problem of the minimization of the maximum error in absolute value into an equivalent problem (in terms of numerical solution), where sequential quadratic programming [19] can be applied. Further information about *fminimax* can be found on the online help section of *Matlab*[®] and on Refs. [15,19–21].

3.3. Parameter initialization

Considering each of the SSE and MEAbs performance indicators tested for this study, convergence of parameter estimation was compared using two types of initialization: the ‘simple’ and the ‘specialized’ one (the latter was proposed on Part I of this series).

- (a) *Simple*: in this case, a random point in the space of parameters ΔH , ΔS and ΔC_p is allocated, with uniform distribution;
- (b) *Specialized*: ΔS and ΔC_p values are randomly allocated, with uniform distribution in the search interval of these parameters, and a ΔH is determined within the region where it is most likely for the global optimum to be found (S2 region), by way of using the algorithm defined in Part I of this series.

Parameter estimations were compared considering the error criteria SSE and MEAbs for each initialization type. In each combination the parameter estimation was run 100 times, and the proportion of converged solutions was logged.

As defined in Part I of this series, the method to discriminate the optimum region (S2) was used for the specialized initialization. To that effect, five applications of the root search method (departing from different initial conditions) were performed, with 15 set as the maximum number of sign function evaluations for the bisection method; this implies in 75 additional evaluations of the objective function in the case of specialized initialization.

Repeated application of the search method between 3 and 10 times resulted in no significant change in the results. Thus, it was decided to establish such a value by 5 in all parameter estimations effected, in order to maintain robustness.

The number of calls to the objective function requested by the specialized initialization should be added to the number of evaluations of the objective function requested by the optimizer, when calculating the total assessments. As will be observed, the computational time associated with the specialized initialization is small when compared to that demanded by the optimizer, which justifies the implementation of this type of initial guess methodology.

3.4. Definition of statistics for classifying the obtained solutions

With the aim of classifying the solutions obtained, statistics have been defined in order to assist with the comparison between the tested procedures. To that end, approximation intervals were defined in terms of the value of the coefficient of determination, and a classification was adopted for each interval. As described, the parameter estimation problem in the non-isothermal case does not count on an analytical solution; therefore it cannot be asserted with certainty whether a specific obtained solution corresponds or not to the global optimum; however, it was considered, numerically,

that the global optimum – or at least an optimum solution – was obtained under the following conditions: taking into account that the parameter estimation for a component be run a specific number of times, departing from different conditions, the best obtained solution is found, and the frequency with which it shows can be verified (within a specified accuracy) from within all the parameter estimations made ($R_{2,0t}$). In this work, the number of repetitions of parameter estimations was set at 100 for each component.

Aiming at classifying the performances of different solutions, three error levels in relation to the best obtained performance were defined: optimal, satisfactory and poor solutions, defined according to each of the error criteria.

For minimizing SSE ($R_{2,0t}$ in percentage):

- (i) Optimal solution:

$$(R_{2,0t} - R_2) < 0.001 [\%] \quad (8)$$

- (ii) Satisfactory solution:

$$0.001 \leq (R_{2,0t} - R_2) < 1 [\%] \quad (9)$$

- (iii) Poor solution:

$$(R_{2,0t} - R_2) \geq 1 [\%] \quad (10)$$

where $R_{2,0t}$ is the coefficient of determination associated to the global optimum and R_2 is the coefficient of determination for each tested solution.

In the case of minimization of MEAbs, similarly to how solutions were classified, for R_2 , in different error levels in relation to the best MEAbs_{0t} results, amongst all the values obtained from the 100 repetitions. These relationships are shown in Eqs. (11)–(13):

- (i) Optimal solution:

$$(\text{MEAbs} - \text{MEAbs}_{0t}) < 0.01 [s] \quad (11)$$

- (ii) Satisfactory solution:

$$0.01 \leq (\text{MEAbs} - \text{MEAbs}_{0t}) < 10 [s] \quad (12)$$

- (iii) Poor solution:

$$(\text{MEAbs} - \text{MEAbs}_{0t}) \geq 10 [s] \quad (13)$$

where MEAbs_{0t} is the solution with the lowest MEAbs value among all repetitions of parameter estimations for each component.

This way, optimum parameters for a specific component are obtained as a result of analyses of a considerable number of repetitions of parameter estimations, and not only from a single regression.

4. Results

The analysis of parameter estimations was compared for both error criteria tested in this work, SSE and MEAbs.

4.1. Parameter estimation analysis from minimizing SSE

On this session the results of parameter estimations done by way of minimizing SSE, and considering the ‘simple’ and ‘specialized’ initializations are shown.

Displayed in Table 1 are the values of estimated parameters, as well as the performance rates for each component. The values described correspond to the best result obtained from the repetitions (totaling 100 repetitions).

In Table 1, in relation to performance rates, it can be observed that coefficients of determination assumed optimum results. Additionally, good results were obtained in the case of MEPs and MEAbs,

Table 1
Estimated parameters and performance indexes obtained by SSE minimization using simple initialization.

Compound	ΔH (kJ/mol)	ΔS (J/mol K)	ΔC_p (J/mol K)	$R^2\%$	MEP%	MEAbs [s]
Octane	-41.8885	-77.7698	0.0019	99.9653	0.1999	0.5201
Nonane	-41.0474	-67.8453	0.3249	99.9990	0.0707	0.2411
Decane	-49.7331	-87.3977	88.0263	99.9997	0.0854	0.3234
Dodecane	-66.4666	-123.4463	178.3668	99.9991	0.1908	1.0722
Tetradecane	-77.8784	-143.2978	199.9999	99.9995	0.1393	1.3057
Pentadecane	-78.1110	-137.8770	168.0760	99.9999	0.0748	0.4833

Table 2
Statistics of parameters estimation for SSE optimization with simple initialization.

Compound	Evaluations number			Solutions probability [%]		
	Optimal	Sat.	Poor	Optimal	Sat.	Poor
Octane	259	2752	2964	45	6	49
Nonane	244	4456	3204	48	7	45
Decane	233	5000	2611	46	8	46
Dodecane	322	4162	3246	56	16	28
Tetradecane	439	3996	4181	51	9	40
Pentadecane	423	3975	2524	60	9	31

even when taking into account that the aforementioned error criteria were not explicitly optimized.

Estimated thermodynamic parameters ΔH , ΔS and ΔC_p represent changes in enthalpy, entropy and calorific capacity at constant pressure, respectively, associated to the solute movement from the mobile phase through to the stationary one at temperature T . This solute transference from the mobile to the stationary phase represents a process similar to that of condensation. The difference being that, in the latter, the compound presents itself in its purest form both on the liquid and vapor phases. Nevertheless, some characteristics of condensation could also be expected for retaining a solute within a chromatography column. For example, ΔH_c and ΔS_c (enthalpy changes and entropy condensation, respectively) present a negative sign for any compound, for a component in the vapor phase is in a higher energy state than in the liquid phase at a given temperature. Displaying the same behavior in terms of sign is expected of ΔH and ΔS . Additionally, within a functional group, in case there is significant difference in molar mass (M) between two compounds (compounds 1 and 2); where $M_1 < M_2$ it is expected that $\Delta H_{c,1} < \Delta H_{c,2}$ and $\Delta S_{c,1} < \Delta S_{c,2}$ (in absolute values). This way, the same trend is expected for $\Delta H_1 < \Delta H_2$ and $\Delta S_1 < \Delta S_2$.

Consequently, the search for positive values ΔH and ΔS was not allowed, and thus these parameters present, mandatorily, negative signs – see Table 1. With respect to variations of ΔH and ΔS with molar mass, it can be observed that there is an upward trend for the values of these parameters (absolute values) with an increase in the component's molar mass.

In Table 2, SSE optimization statistics on the parameter estimation with simple initialization are shown. The proportion of the ratings obtained (optimal, satisfactory and poor solutions) is displayed, as well as the average number of objective function evaluations for each compound (within the 100 repetitions) with the use of the *fmincon* optimization function. As the optimization method does not guarantee convergence to the global optimum

in all parameter estimations, there is a likelihood of satisfactory and poor solutions. The occurrence ratios of these solutions were obtained by analyzing the results of repeated parameter estimation for each component.

In Table 2 it can be seen that the number of objective function evaluations demanded by the method in order to obtain an excellent solution is much lower than those needed for obtaining satisfactory and poor solutions (by around an order of magnitude), that is, the closer the initial guess is to the highest probability of the optimum to be found implies that fewer evaluations of the objective function are necessary for convergence up to the global optimum.

As it can be seen in Table 2, the simple initialization method does not guarantee that an optimal solution be obtained in 100% of cases. In a practical case this denotes a need for repeating the parameter estimation at least a few times. Still, in some cases the probability of obtaining a poor solution may be higher than what it is to obtain a good or excellent solution.

Next, the performance analysis of repeated parameter estimation was run considering the specialized initialization, and results are shown in Table 3. R^2 values agree with those from Table 1 to the number of decimal places shown. In the case of MEP and MEAbs values similar to those in Table 1 were obtained, but with a noticeable difference. As mentioned, this happens because the SSE was explicitly minimized, whereas the MEP or the MABs were not.

Comparing the values of estimated parameters in Tables 2 and 3, it can be concluded that they present differences, primarily for the pentadecane compound. This occurs for all components, even though almost the same values of R^2 were obtained. Such behavior can be attributed to the simultaneous estimations of several parameters.

In a simultaneous estimation of parameters it is possible that the change in one parameter can, to some extent, be compensated by changing others, when the values predicted by the model are observed. Thus, different sets of parameters can be of substantially the same value as the objective function (and therefore of R^2).

In Table 4 the results of relevant statistics related to estimating SSE optimization parameters with specialized initialization are displayed.

As can be seen in Table 4, excellent solutions were obtained in 100% of the cases, which shows how effective the specialized initialization procedure is when compared to simple initialization.

The number of objective function evaluations with the specialized initialization was always lower than those observed with the simple initialization, even for the better solutions (optimal solutions). This result can be attributed to the fact that there is a

Table 3
Estimated parameters and performance indexes obtained by SSE minimization using specialized initialization.

Compound	ΔH (kJ/mol)	ΔS (J/mol K)	ΔC_p (J/mol K)	$R^2\%$	MEP %	MEAbs [s]
Octane	-41.8923	-77.7816	0.0020	99.9653	0.1996	0.5193
Nonane	-41.0462	-67.8416	0.3168	99.9990	0.0706	0.2407
Decane	-49.7334	-87.3983	88.0263	99.9997	0.0855	0.3236
Dodecane	-66.5389	-123.6580	179.0879	99.9991	0.1904	1.0697
Tetradecane	-77.8787	-143.2986	199.9995	99.9995	0.1393	1.3058
Pentadecane	-81.3553	-147.0259	194.2721	99.9999	0.0741	0.4790

Table 4
Statistics of parameters estimation for SSE optimization with specialized initialization.

Compound	Evaluations number of optimizer	Eval. number of optimizer + Specialized initialization	Optimal solution [%]
Octane	154	229	100
Nonane	159	234	100
Decane	119	194	100
Dodecane	134	209	100
Tetradecane	145	220	100
Pentadecane	108	183	100

tendency of reducing the number of evaluations of the optimized criterion the closer the initial guess is to the global optimum.

Additionally, the total number of evaluations of the objective functions, considering usage of the optimization method, as well as the assessments demanded by the specialized initialization (in this case, equal to 75).

As an estimate, it can be said that the specialized initialization resulted in savings of one to two orders of magnitude, in terms of the objective function evaluations, at the parameter estimation of the retention model factor, considering the minimization of SSE. This statement is based on two observed facts:

The number of evaluations of the objective function demanded for obtaining an excellent solution is much lower than the one needed for satisfactory and poor ones (around one order of magnitude);

The simple initialization method does not warrant the obtainment of an excellent solution in 100% of cases. In a practical case, this implies a need to repeat the parameter estimation at least a few times.

4.2. Analysis of parameter estimation from minimizing MEAbs

In this section the results of the parameter estimations made by way of minimizing MEAbs, and considering the simple and specialized initializations, are shown. Table 5 shows the values of the estimated parameters and performance indexes for each component. The reported values correspond to the best results of the repetitions.

Comparing Table 3 with Table 5, the lower SSE values (consequently, the highest R^2 value) are obtained when this is the performance index itself (Table 3) to be optimized. Something similar happens in the case of the MEAbs, with the lowest MEAbs values in Table 5.

In Table 5, with respect to the variation ΔH and ΔS with molar mass, in the same way as for the minimization of SSE, it can be seen that there is a tendency for these parameters to increase (in absolute terms) with an increase in the molar mass of the component.

Results of statistics referent to MEAbs parameter optimization with simple initialization are shown in Table 6.

As can be observed in Table 6 the simple initialization method does not warrant that excellent solutions be obtained at all times. In the same way as for SSE, in a practical case this will imply in having to repeat parameter estimation at least a few times. The

Table 5
Estimated parameters and performance indexes obtained by MEAbs minimization using simple initialization.

Composto	ΔH (kJ/mol)	ΔS (J/mol K)	ΔC_p (J/mol K)	$R^2\%$	MEP %	MEAbs [s]
Octane	-42.5415	-79.7671	0	99.9542	0.1366	0.3555
Nonane	-41.0083	-67.7241	0	99.9990	0.0638	0.2079
Decane	-50.3154	-89.1665	95.5981	99.9996	0.0572	0.2165
Dodecane	-66.9033	-124.6464	180.3116	99.9988	0.1825	0.8391
Tetradecane	-77.9392	-143.4622	200	99.9993	0.1728	0.9653
Pentadecane	-82.0207	-148.9241	200	99.9999	0.0656	0.4239

Table 6
Statistics of estimated parameters obtained by MEAbs minimization using simple initialization.

Compound	Evaluations number			Solutions probability [%]		
	Optimal	Sat.	Poor	Optimal	Sat.	Poor
Octane	463	–	2200	29	0	71
Nonane	426	–	2247	35	0	65
Decane	602	–	2666	34	0	66
Dodecane	521	–	2588	32	0	68
Tetradecane	643	–	1976	31	0	69
Pentadecane	954	–	2371	32	0	68

Table 7
Statistics of estimated parameters obtained by MEAbs minimization using specialized initialization.

Compound	Evaluations number of optimizer	Eval. number of optimizer + Specialized initialization	Optimal solution [%]
Octane	75	150	100
Nonane	76	151	100
Decane	180	255	100
Dodecane	84	159	100
Tetradecane	79	154	100
Pentadecane	101	176	100

difference is that the likelihood of obtaining an excellent solution was significantly lower in case of minimization of MEAbs.

Next, the analysis based on repeated parameter estimations was performed considering the specialized initialization. The values of the parameters and performance indexes obtained for each component were similar to those shown in Table 5.

The results of statistics referent to MEAbs parameter estimations run by way of the specialized initialization are shown in Table 7.

As it can be observed in Table 7, convergence to the global optimum in 100% of cases was obtained, which again shows the effectiveness of the initialization procedure to minimize the MEAbs.

In the case of excellent solutions, when comparing the specialized initialization with the simple one, the number of evaluations of the objective function was always inferior for the former, even when additional assessments demanded by specialized initializations were considered.

Considering a minimization of MEAbs, in the same way as was observed for SSE, as an estimate, specialized initialization resulted in savings of one to two orders of magnitude, in terms of objective function evaluations, when estimating parameters of the retention factor model.

It is worth noting that the limiting stage in terms of computational time in parameter estimation is the evaluation of the objective function. Thus, to say that the specialized initialization resulted in a savings of one to two orders of magnitude, in terms of evaluations of the objective function, implies that the same will occur with respect to the computational time required.

Departing from results obtained with the minimization of SSE and MEAbs, it can be concluded that the difference between the number of evaluations of the objective function in parameter esti-

Table 8

Comparison between predicted retention time and experimental data (validation program P1).

Compound	$t_{r,Exp,i}$ [min]	SSE minimization			MEAbs minimization		
		$t_{r,mod,i}$ [min]	MEAbs [s]	MEP [%]	$t_{r,mod,i}$ [min]	MEAbs [s]	MEP [%]
Octane	4.2510	4.2256	1.5219	0.3602	4.2206	1.8268	0.4328
Nonane	5.7100	5.7142	0.2515	0.0440	5.7150	0.3025	0.0529
Decane	7.5350	7.5011	2.0367	0.2715	7.4976	2.2433	0.2992
Dodecane	9.8910	9.9462	3.3101	0.3328	9.9449	3.2316	0.3250
Tetradecane	11.2620	11.3442	4.9343	0.4350	11.3410	4.7377	0.4178
Pentadecane	11.7900	11.8680	4.6801	0.3943	11.8740	5.0389	0.4244

Table 9

Comparison between predicted retention time and experimental data (validation program P2).

Compound (i)	$t_{r,Exp,i}$ [min]	SSE minimization			MEAbs minimization		
		$t_{r,mod,i}$ [min]	MEAbs [s]	MEP [%]	$t_{r,mod,i}$ [min]	MEAbs [s]	MEP [%]
Octane	3.9320	3.8964	2.1385	0.5489	3.8904	2.4974	0.6419
Nonane	5.1690	5.1280	2.4571	0.4792	5.1290	2.4025	0.4684
Decane	6.5940	6.5748	1.1548	0.1756	6.5716	1.3452	0.2047
Dodecane	8.8470	8.8678	1.2502	0.1410	8.8650	1.0796	0.1218
Tetradecane	10.5580	10.6081	3.0051	0.2833	10.6036	2.7365	0.2581
Pentadecane	11.3990	11.4673	4.0992	0.3575	11.4712	4.3300	0.3775

mations done with the simple and the specialized initializations can be attributed to two principal features:

- (A) *Need for repeated parameter estimation*: As mentioned, in the case of simple initialization there is a significant likelihood of solutions with satisfactory and poor performances to be found. Since it is not possible to know, in advance, what kind of solution will be obtained from the optimization, repetition is necessary;
- (B) *High number of objective function evaluations in the case of satisfactory and poor solutions*: Solutions that present satisfactory and poor performances imply a sharp increase in the number of objective function evaluations when compared with the excellent solution. The former were observed with significant probability, in the case of the simple initialization, but did not occur in the case of the specialized one.

5. Experimental validation of estimated thermodynamic parameters

Experimental validation of the estimated parameters consisted of resolving the retention time model and comparing it to the respective measured data. For validation, two programs were adopted: P1 and P2, as described below:

P1 Program:

1. Isotherm of 50 °C for 2 min;
2. Ramp at 15 °C/minute up to 100 °C;
3. Isotherm of 100 °C for 2 min;
4. Ramp at 40 °C/minute up to 250 °C;
5. Isotherm of 250 °C for 5 min.

P2 Program:

1. Isotherm of 50 °C for 2 min;
2. Ramp at 30 °C/minute up to 100 °C;
3. Isotherm of 100 °C for 2 min;
4. Ramp at 25 °C/minute up to 200 °C;
5. Isotherm of 200 °C for 5 min.

Tables 8 and 9 display the results from predicting retention time for the series of alkanes in the case of temperature programs P1 and

P2, respectively. Results were calculated from the best solutions that have come from minimizing the SSE and MEAbs indexes.

As it can be observed in Tables 8 and 9, MEAbs in validation is in the order of a few seconds, which is an acceptable error. Additionally, all MEP were lower than 1%, and therefore, it can be considered, in accordance with the literature, that a good prediction performance was obtained for all components.

6. Conclusions

In this work the parameters of the retention factor in GC model were estimated based on two performance criteria (sum of square error and maximum error in absolute value) and two initialization methods were tested: the specialized procedure (proposed in Part I of this series) and the simple one.

The specialized initialization is a rapid convergence procedure (and, consequently, a low computational time one) based on knowledge of the optimization criterion's surface, which is known due to thermodynamic parameters associated to retention.

In an application of the method to a series of alkanes, and comparing the specialized initialization with the simple one, for the former a reduction in computational time for parameter estimations that was between one and two orders of magnitude was noted, considering the approximation error's minimization of a quadratic criterion. The same happened when minimizing the maximum prediction error in absolute value.

It is not guaranteed that, as a rule, such performance gain be obtained in the estimation of parameters in GC, because this is a non-linear problem; however, better performance is expected with the specialized initialization in most cases, according to the heuristic in optimization: The closer an initial guess is the to the global optimum, the more likely the latter is to be found, and the lower the number of evaluations of the objective function necessary for convergence.

Conflict of interest

The authors confirm that there are no conflicts of interest regarding this paper.

Acknowledgments

The authors thank CAPES, CNPq for their financial support. A. W. Zibetti thanks to CAPES-PNPD (2014/2015) for their assistantship.

References

- [1] F. Aldaeus, Y. Thewalim, A. Colmsjö, Prediction of retention times and peak widths in temperature-programmed gas chromatography using the finite element method, *J. Chromatogr. A* 1216 (2009) 134–139.
- [2] H. Snijders, H.-G. Janssen, C. Cramers, Optimization of temperature-programmed gas chromatographic separations II. Off-line Simplex optimization and column selection, *J. Chromatogr. A* 756 (1996) 175–183.
- [3] S. Vezzani, P. Moretti, G. Castello, Classification and comparison of capillary columns by determination of the solution enthalpy of polar and non polar probes, *J. Chromatogr. A* 1101 (2006) 261–267.
- [4] T.M. McGinitie, B.R. Karolat, C. Whale, J.J. Harynuk, Influence of carrier gas on the prediction of gas chromatographic retention times based on thermodynamic parameters, *J. Chromatogr. A* 1218 (2011) 3241–3246.
- [5] T.M. McGinitie, J.J. Harynuk, Prediction of retention times in comprehensive two-dimensional gas chromatography using thermodynamic models, *J. Chromatogr. A* 1255 (2012) 184–189.
- [6] B. Karolat, J. Harynuk, Prediction of gas chromatographic retention time via an additive thermodynamic model, *J. Chromatogr. A* 1217 (2010) 4862–4867.
- [7] F. Aldaeus, Y. Thewalim, A. Colmsjö, Prediction of retention times of polycyclic aromatic hydrocarbons and n-alkanes in temperature-programmed gas chromatography, *Anal. Bioanal. Chem.* 389 (2007) 941–950.
- [8] E.V. Dose, Simulation of gas chromatographic retention and peak width using thermodynamic retention indexes, *Anal. Chem.* 59 (1987) 2414–2419.
- [9] B. Peng, M.-Y. Kuo, P. Yang, J.T. Hewitt, P.G. Boswell, A practical methodology to measure unbiased gas chromatographic retention factor vs. temperature relationships, *J. Chromatogr. A* 1374 (2014) 207–215.
- [10] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons, 2013.
- [11] J. Randon, L. Maret, C. Ferronato, Gas chromatography–mass spectroscopy optimization by computer simulation, application to the analysis of 93 volatile organic compounds in workplace ambient air, *Anal. Chim. Acta* 812 (2014) 258–264.
- [12] C.R. Rao, *Linear Statistical Inference and its Applications*, Wiley, 1965.
- [13] R.H. Byrd, J.C. Gilbert, J. Nocedal, A trust region method based on interior point techniques for nonlinear programming, *Math. Program.* 89 (2000) 149–185.
- [14] T.F. Coleman, Y. Li, On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds, *Math. Program.* 67 (1994) 189–224.
- [15] S.P. Han, A globally convergent method for nonlinear programming, *J. Optim. Theory Appl.* 22 (1977) 297–309.
- [16] R.A. Waltz, J.L. Morales, J. Nocedal, D. Orban, An interior algorithm for nonlinear optimization that combines line search and trust region steps, *Math. Program.* 107 (2005) 391–408.
- [17] T.F. Coleman, Y. Li, An interior trust region approach for nonlinear minimization subject to bounds, *SIAM J. Optim.* (2006).
- [18] M.J.D. Powell, A fast algorithm for nonlinearly constrained optimization calculations, in: G.A. Watson (Ed.), *Numer. Anal. SE*, vol. 13, Springer Berlin Heidelberg, 1978, pp. 144–157.
- [19] R. Brayton, S. Director, G. Hachtel, L. Vidigal, A new algorithm for statistical circuit design based on quasi-Newton methods and function splitting, *IEEE Trans. Circuits Syst.* 26 (1979) 784–794.
- [20] A.C.W. Grace, *Computer-aided control system design using optimization methods*, 1989.
- [21] H. Schjaer-Jacobsen, K. Madsen, Algorithms for worst-case tolerance optimization, *IEEE Trans. Circuits Syst.* 26 (1979) 775–783.