Journal of Chromatography, 95 (1974) 147–157

CHROM. 7452

NON-LINEAR FITTING METHOD FOR RECORDED CHROMATOGRAPHIC PEAKS

E. CUSÓ, X. GUARDINO, J. M. RIERA and M. GASSIOT*

Departamento de Química Analítica del Instituto Químico de Sarriá (Patronato Alfonso X el Sabio), Barcelona (Spain)

(Received January 23rd, 1974)

SUMMARY

A method for fitting mathematical functions to experimental chromatographic peaks is described, with special reference to those gas–solid chromatographic peaks that have particular fitting difficulties owing to their high degree of skewness. The method was used with different functions and with peaks of diverse characteristics. The computer program used for the calculation and the different results obtained are also discussed.

INTRODUCTION

A knowledge of the mathematical functions that describe chromatographic peaks is fundamental towards the optimization of the calculations and for obtaining better results in gas chromatography, such as the calculation of the peak area (when an integrator is not available)[1], the resolution of overlapping peaks[2,3] and physico-chemical studies based on gas chromatography[4,5].

In gas chromatography, as the elution curves give the variation of concentration with time at the exit of the column, they can be represented by curves of density probability following a determined statistical distribution. We have utilized this approach in most of the cases studied.

It is in gas–solid adsorption chromatography in which it is most difficult to know the exact shape of the chromatographic peaks, as the bands are usually very asymmetric, with a vertical front and a very protracted tail. For this reason, the present study of mathematical functions is centred mainly on those that take asymmetry into account as an important factor. This problem has been dealt with by several workers, including Anderson *et al.*[6], Grushka *et al.*[7], Gladney *et al.*[8] and, more recently, Chesler and Gram[9], who employed for the first time a method of adjustment on the basis of a convolution function, the result of which is expressed in analogue form so as to make possible a selective modification of the adjustment parameters; these workers used as many as eight parameters in order to achieve their aim. In the present work, we have obtained an iterative adjustment of peaks on the basis of an automatic non-linear regression, several functions being successively employed. In the method proposed, no special calculation or measurement is required

---

* To whom correspondence should be addressed.

TABLE I

SOME FUNCTIONS FOR THE DESCRIPTION OF CHROMATOGRAPHIC PEAKS

Function	*Mathematical expression*	*Function characteristics*	*Applications*
Gauss[10]	$y=h_{\max.}\cdot\exp[-(t-t_R)^2/2\sigma^2]$	Continuous distribution. Peak substantially symmetric.	Linear non-ideal chromatography. Theoretical peak and some hydrocarbons.
Poisson[10]	$y=h_{\max.}\cdot\exp(-t_R)\cdot\frac{(t_R)^t}{t!}$	Discontinuous distribution.	Low number of theoretical plates. Liquid–liquid chromatography.
Modified Gauss[11]	$y=h_{\max.}\cdot\left(\frac{t_R}{t}\right)^{\frac{1}{2}}\cdot\exp[-2t_R(t_R^{\frac{1}{2}}-t^{\frac{1}{2}})^2/2\sigma^2]$	Continuous distribution. Slightly skew.	Experimental peaks in GLC.
Convolution Gauss–exponential	(A) Anderson *et al.*[12]: $y=\frac{h_{\max.}}{2}\int_0^{s}\exp(-t'/\tau)\cdot\exp\left[-\left(\frac{t-t_R-t'}{\sigma}\right)^2\right]\cdot\mathrm{d}t'$ (B) Gladney *et al.*[8]: $y=\frac{h_{\max.}}{\tau}\cdot\int_{-\infty}^{t}\left[\exp\left(-\frac{(t'-t_R)^2}{2\sigma^2}\right)\cdot\exp((-t-t+t_R)/\tau)\right]\cdot\mathrm{d}t'$	Continuous distributions. A new parameter (asymmetry function) appears. Asymmetry is related to the τ/σ ratio.	Peaks are fitted, from $\tau=0$ (pure Gauss), to $\sigma=0$ (pure exponential decay). They can be used for GSC peaks with moderate asymmetry.
Composition Gauss–polynomic function	(A) Grushka *et al.*[7]: $y=\frac{1}{\sigma\sqrt{2\pi}}\cdot\exp\left(\frac{-(t-t_R)^2}{2\sigma^2}\right)\cdot\left(1+\sum_{i=3}\frac{C_i}{i!}\cdot H_i\left(\frac{t}{\sigma}\right)\right)$ (B) Grubner[13,14]: $y=\frac{1}{\sqrt{2\pi}}\cdot\exp\left(-\frac{(t-t_R)^2}{2\sigma^2}\right)\cdot\left[1+\frac{S}{6}\cdot\left(\left(\frac{t-t_R}{\sigma}\right)^3-3\cdot\left(\frac{t-t_R}{\sigma}\right)\right)-\frac{E}{24}\cdot\left(\left(\frac{t-t_R}{\sigma}\right)^4-6\cdot\left(\frac{t-t_R}{\sigma}\right)^2+3\right)\right]$	There are based statistical moments of the independent variable (t). H_i = Hermite polynomials. C_3 = S = skewness. C_4 = E = excess.	All type of curves are fitted (symmetrical as well as asymmetric).

for the estimation of the initial data of the iteration, nor is any complementary instrument necessary: the recorded graph and the corresponding computer suffice.

The best known functions for the description of a chromatographic peak that we have found in the literature are summarized in Table I.

METHOD OF FITTING

Once the various functions that we have described briefly have been studied, the main problem is to establish which is the most adequate mathematical function for each case (given an actual peak) and which are the values of its parameters. In this section, we describe the mathematical method employed in order to obtain the parameters of each function, in the most exact form possible, thus obtaining a correct mathematical expression for it. With this aim, we considered it convenient to try a fitting method that would provide an equation that adequately describes an actual chromatogram.

The adjustment method selected is that employed by Law and Bailey[15], based on the known method of least squares, which is made applicable to any type of function through a technique of general, non-linear, regression. This method has been employed previously in problems related to gas chromatography by Campreciós and Gassiot[3].

In order to effect the adjustment, a simultaneous variation is made of all the parameters that occur in the function considered, until there are obtained, through successive iterations, some parameter values such that, given the experimental points, the function presents a minimal mean quadratic error.

Consider a function of the type

$$y_i=\varphi(x_{1i}, x_{2i}, x_{3i}, \ldots, x_{mi}, B_1, B_2, B_3, \ldots, B_{np})\pm e_y \tag{1}$$

where

y_i = dependent variable;
x_{ji} = independent variable;
B_k = function parameters;
i = 1, 2, 3, ..., n=number of points;
j = 1, 2, 3, ..., m=number of independent variables;
k = 1, 2, 3, ..., np=number of parameters;
e_y = experimental error.

If this function is developed into a Taylor series, terminated after the first derivative, and brought in the assimilation $\Delta B_k \simeq \mathrm{d}B_k$, we have:

$$\varphi^{i+1} = \varphi^i + \left(\frac{\partial\varphi}{\partial B_1}\right)^i \cdot \Delta B_1 + \left(\frac{\partial\varphi}{\partial B_2}\right)^i \cdot \Delta B_2 + \ldots + \left(\frac{\partial\varphi}{\partial B_k}\right)^i \Delta B_k + \ldots \tag{2}$$

where i is the iteration order

$$\Delta B_k = B_k^{i+1} - B_k^{\,i} \tag{3}$$

so that, in each iteration there is a linear increment of a non-linear function.

The condition imposed is that the mean quadratic error should be minimal:

$$\frac{\partial \sum_{i=1}^{n} S_i^2}{\partial B_k} = 0 \tag{4}$$

where $S_i = y_i - \varphi_i$ (in each point)

The system of equations to be solved in each iteration is of the type:

$$A\,|b> = |c> ;\ A^T \cdot A\,|b> = A^T\,|c> ;\ |b> = (A^T \cdot A)^{-1} \cdot A^T\,|c> \tag{5}$$

where

A = matrix of the elements: $\left(\frac{\partial \varphi}{\partial B_k}\right)_i$;

$|b>$ = vector of the unknown: ΔB_j;

$|c>$ = vector of the errors: S_i.

In order to solve the system, some initial values are given to B_k and, if the scheme converges for values of B_k that minimize $\sum S_i^2$, the vector $|b>$ represents the ΔB_k values, and therefore eqn. 3 can be solved for B_k^{i+1}. These new values can then be used in the next iteration, and thus successively, until ΔB_k eventually becomes negligible. The limitation of the method lies in the fact that it may lead to divergent values if the initial parameters calculated are not in the region of the correct values. However, if some "restrictive" factors are introduced, the convergence is always obtained.

APPLICATION OF THE METHOD TO CHROMATOGRAPHIC PEAKS

According to the above discussion, a chromatographic peak is of the type

$$y = \varphi(t, B_1, B_2, \ldots, B_{np})$$

where time (t) is the only independent variable and B are the parameters; in all of the functions presented here, their number varies from 3 to 5, according to whether the shape differs more or less from that of a perfect Gaussian peak.

The nomenclature used is as follows:

B_1 = height of the peak;
B_2 = retention time;
B_3 = standard deviation;
B_4 = skewness;
B_5 = excess.

Hence it is correct to apply to a chromatographic band the method of fit described above.

However, when attempting to work with any of the proposed functions, the problem arises of how to estimate correctly the initial value of the parameters involved in each of them. The answer is simple when there are symmetrical (or virtually symmetrical) peaks, but is more difficult when the asymmetry of the peak is consider-able, because in such cases even the parameters of direct measurement, such as retention time and peak height, are modified. For this reason, we believe that the best means of calculating the parameters of the different functions is through the statistical moments of the variable, defined as follows.

Statistical moments of the variable with respect to the origin:

$$m_i = \frac{\int_0^\infty c(t)\cdot t^i \mathrm{d}t}{\int_0^\infty c(t)\cdot \mathrm{d}t}$$

Statistical moments of the variable with respect to the mean:

$$\bar{m}_i = \frac{\int_0^\infty c(t)\cdot (t-m_1)^i \mathrm{d}t}{\int_0^\infty c(t)\cdot \mathrm{d}t}$$

where

$c(t)$ = peak profile.

The computer applies the "method of trapezes" for the evaluation of the integrals in the above expression with the same points used in the adjustment.

In a statistical distribution, it is verified that:

$$\bar{x} = m_1$$

$$\sigma^2 = \bar{m}_2$$

$$S = \frac{\bar{m}_3}{\sigma^3}$$

$$E = \frac{\bar{m}_4}{\sigma^4} - 3$$

We use the following nomenclature

$$B_2 = t_R = m_1$$

$$B_3 = \sigma = \sqrt{\bar{m}_2}$$

$$B_4 = S = \frac{\bar{m}_3}{\sigma^3}$$

$$B_5 = E = \frac{\bar{m}_4}{\sigma^4} - 3$$

We find the height in the middle of the distribution by means of the following equation:

$$y_{\text{max.}} = \varphi(t_R, h_{\text{max.}}, t_R, \sigma, S, E)$$

where $h_{\text{max.}}$ is unknown.

As we have already indicated, the values of both B_2 and B_1 do not correspond exactly to t_R and to the maximum height when the peak is very asymmetric.

PRACTICAL APPLICATION

From the above discussion, one can appreciate the complexity of the calculation involved in this fitting procedure. In order to make it easy to apply, a program has been designed in Fortran IV and operated on an IBM 1130 computer. The program is written in such a way that several alternatives of calculation exist that can be selected by the use of cards or by entering the data by means of a console. The scheme of the program is shown in Fig. 1.

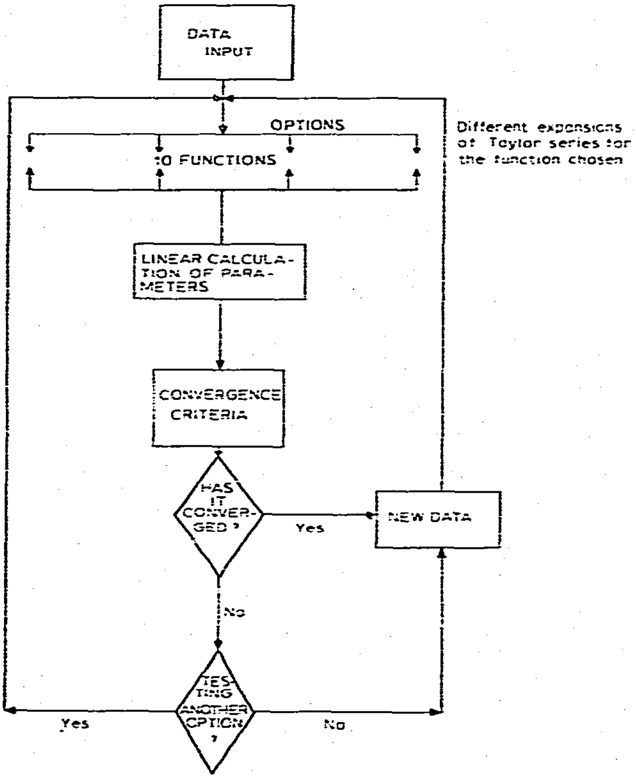

Fig. 1. Scheme of computer program (PAPAT).

The program functions as follows. First, the data are introduced, which consist of peak identification, points on the chromatogram and indicator of the function chosen, followed by the calculation of the statistical moments by means of the above equations. After finding the function which has been indicated to it, the program applies the method of adjustment discussed above, calculates the derivatives, finds the new value of each parameter and compares the result with the maximal error allowed for each function. According to whether the difference between y_i and

φ_i is larger or smaller than this error, it proceeds to work on a new iteration, or else it writes the results and goes on to calculate a new problem.

This program can calculate any type of chromatographic peak, employing as many mathematical functions as may be desired. In fact, it is able to work simultaneously with a maximum of ten different functions. We have worked with only four types of functions, namely the Gauss, Littlewood, Grubner and Gladney types, and, in general, we obtained satisfactory results.

The calculation program is applicable not only to chromatographic peaks, but also to any type of problem in which it may be desired to fit to a set of experimental points a curve following a given statistical distribution.

RESULTS

The results obtained in our work differed according to the type of peak studied. Some individual cases are described briefly below.

For actual peaks obtained in gas–liquid chromatography the result is always coherent, *i.e.*, the curve adjusts itself to the actual points, within a margin of minimal error (Table II). However, with peaks obtained in gas–solid chromatography, the agreement is less satisfactory, as the moments obtained are different from their actual values (Table III).

TABLE II

EXPERIMENTAL PEAK (GAS–LIQUID CHROMATOGRAPHY)

Number of data points: 21. Maximum allowable tolerance of parameters: 0.01.

Parameter	*Initial calculated value*	*Fitted parameters*		
		Gauss: number of iterations = 2	*Littlewood: number of iterations = 2*	*Grubner: number of iterations = 6*
Height	6.030	6.214	6.214	6.139
Retention time	5.648	5.652	5.652	5.666
Standard deviation	0.115	0.106	0.106	0.109
Skewness	0.364	–	–	0.515
Excess	–0.522	–	–	0.051

TABLE III

EXPERIMENTAL PEAK (GAS–SOLID CHROMATOGRAPHY)

Number of data points: 20. Maximum allowable tolerance of parameters: 0.01.

Parameter	*Initial calculated value*	*Fitted parameters*			
		Gauss: number of iterations = 3	*Littlewood: number of iterations = 3*	*Grubner: number of iterations = 9*	*Gladney: number of iterations = 11*
Height	12.327	13.650	13.675	16.632	11.219
Retention time	38.263	37.831	37.849	38.431	35.910
Standard deviation	2.281	2.017	2.015	1.937	0.859
Skewness	1.906	–	–	1.087	1.917
Excess	7.012	–	–	1.325	–

We have also examined theoretical peaks generated by the functions studied, for which fitting by means of other functions presents no problems, so that the correctness of the choice of the functions studied and, at the same time, the acceptability of the method could be confirmed (Tables IV–VI).

TABLE IV

THEORETICAL PEAK (GAUSS)

Number of data points: 15. Maximum allowable tolerance of parameters: 0.01.

Parameter	*Initial calculated value*	*Fitted parameters*			
		Gauss: number of iterations = 1	*Littlewood: number of iterations = 2*	*Grubner: number of iterations = 3*	*Gladney*
Height	20.000	19.998	19.842	19.992	—
Retention time	8.000	7.999	8.064	8.000	—
Standard deviation	1.995	2.000	2.009	1.997	—
Skewness	$0.399 \cdot 10^{-3}$	—	—	$0.683 \cdot 10^{-3}$	—
Excess	−0.035	—	—	$-0.966 \cdot 10^{-3}$	—

TABLE V

THEORETICAL PEAK (GLADNEY)

Number of data points: 28. Maximum allowable tolerance of parameters: 0.05.

Parameter	*Initial calculated value*	*Fitted parameters*			
		Gauss: number of iterations = 2	*Littlewood: number of iterations = 2*	*Grubner: number of iterations = 5*	*Gladney: number of iterations = 5*
Height	22.220	21.750	21.556	21.115	16.342
Retention time	14.719	14.570	14.590	14.902	13.603
Standard deviation	1.398	1.110	1.127	1.219	0.697
Skewness	1.086	—	—	1.054	1.393
Excess	2.029	—	—	0.266	—

TABLE VI

SYMMETRICAL PEAK

Number of data points: 31. Maximum allowable tolerance of parameters: 0.01.

Parameter	*Initial calculated value*	*Fitted parameters*			
		Gauss: number of iterations = 3	*Littlewood: number of iterations = 3*	*Grubner: number of iterations = 4*	*Gladney*
Height	128.297	146.399	145.726	125.973	—
Retention time	30.000	30.000	30.047	30.000	—
Standard deviation	4.190	3.844	3.861	3.309	—
Skewness	$-0.388 \cdot 10^{-3}$	—	—	$-0.105 \cdot 10^{-2}$	—
Excess	−1.482	—	—	−1.521	—

Of the functions studied, that which presents the greatest difficulty of fit is that of Gladney *et al.*[8], even when the peak is almost Gaussian in shape. This is a reasonable result, because the convolution becomes the product of infinite and infinitesimal terms when the skewness is also infinitesimal. From the formal point of view, we can eliminate this indetermination, but this is not possible when purely numerical work is carried out in a computer, where the precision is finite.

Contrary to the function given by Gladney *et al.*[8], the function that shows adjustment, in general, for all peaks with a lower error with regard to the actual values is that of Grubner[13]. This result is logical, as functions of a polynomial type are usually a perfect substitute for any other type of function, if the parameters are adequate and the number of terms in the polynomial is not too small (Table VII and Fig. 2).

For the maximal error allowed, we have established in most cases a value of 0.01 unit. In some cases, however, a convergence for a tolerance of 0.01 unit has not been obtained, while a convergence has been obtained, allowing a larger error of the order of 0.05 unit. Also, in other cases, the contrary occurred, *i.e.*, when the tolerance was reduced to 0.001 unit, the result remained coherent. The discrete nature of the system of adjustment (in the sense of not having a function, but only several points), does not preclude that some local minima with no relation to the chromatographic peaks could be obtained. This effect results in some functional parameters devoid of sense, such as a negative standard deviation.

TABLE VII

COMPARISON OF RESULTS OBTAINED WITH SOME FUNCTIONS

t_R (*experimental*)	y				
	Experimental	*Gauss* (*3 iterations*)	*Littlewood* (*3 iterations*)	*Grubner* (*9 iterations*)	*Gladney* (*11 iterations*)
34.00	0.10	2.17	2.05	0.14	0.29
35.00	2.32	4.96	4.91	3.82	3.12
35.50	6.80	6.83	6.86	6.80	6.80
36.00	11.45	8.86	8.95	10.01	10.78
37.05	14.57	12.52	12.65	14.49	15.21
37.50	14.03	13.37	13.46	14.00	14.55
38.00	12.91	13.58	13.59	13.50	12.96
38.50	11.80	12.97	12.92	11.63	11.19
39.00	10.16	11.65	11.57	9.67	9.59
40.00	7.37	7.83	7.79	7.01	6.90
41.00	5.22	4.12	4.18	5.60	5.06
42.00	3.45	1.69	1.87	3.82	3.68
43.00	2.23	0.54	0.63	1.85	2.68
44.00	1.45	0.13	0.18	0.60	1.95
46.00	0.59	0.00	0.00	0.01	1.03
48.00	0.34	0.00	0.00	0.00	0.54
50.00	0.25	0.00	0.00	0.00	0.28
52.00	0.20	0.00	0.00	0.00	0.15
54.00	0.11	0.00	0.00	0.00	0.08
56.00	0.00	0.00	0.00	0.00	0.04

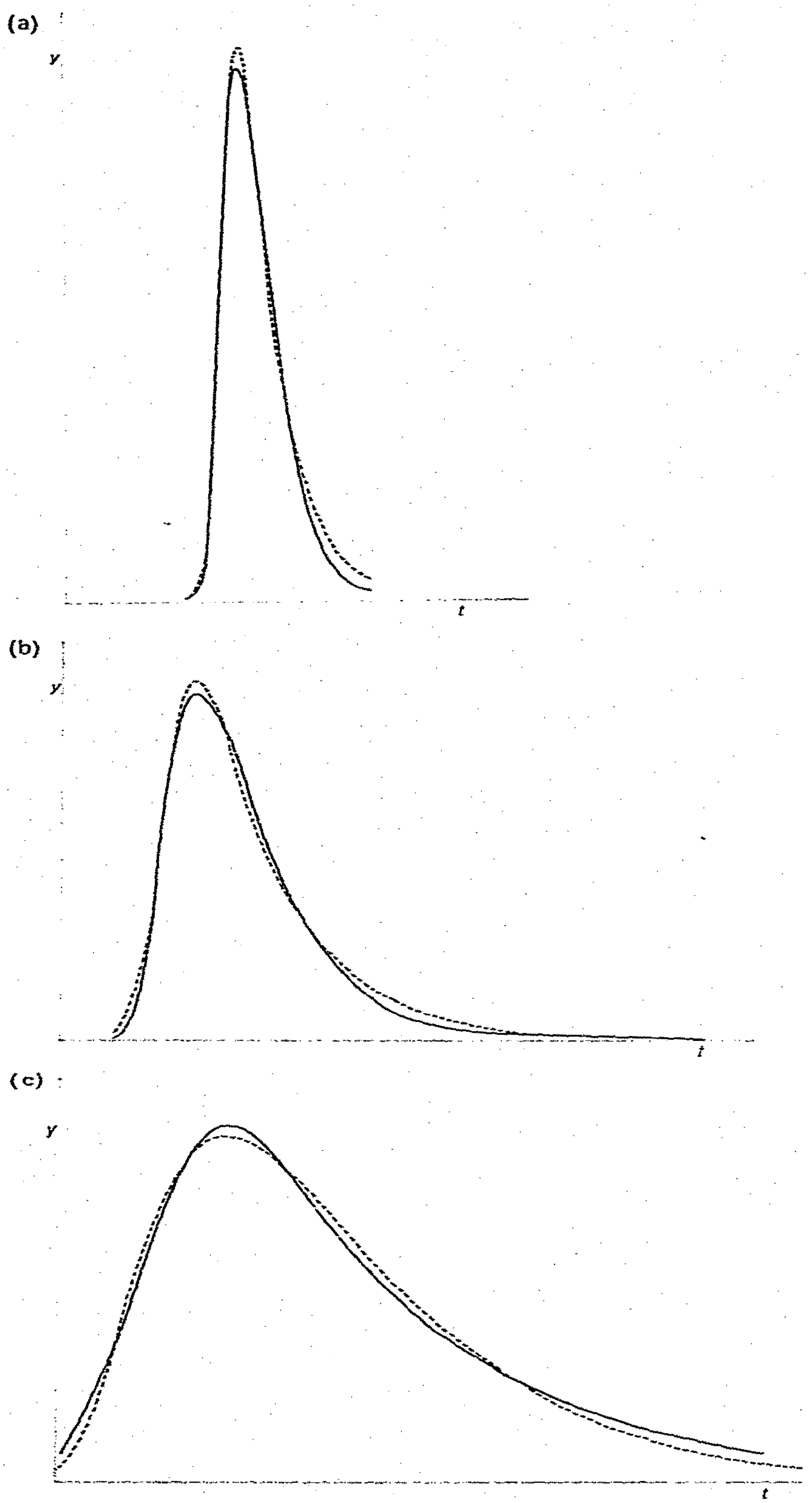

Fig. 2. Shapes of experimental (———) and calculated (– – – –) fitting peaks for acetic acid. Flow-rates of carrier gas: (a), 60 ml/min; (b), 30 ml/min; (c), 15 ml/min.

REFERENCES

1 X. Guardino, E. Cusó and M. Gassiot, *Determinación de Isotermas de Adsorción Sobre Porapak Q, Mediante Cromatografía de Gases*, Trabajo de Fin de Carrera, I.Q.S., Barcelona, 1972, p. 41.
2 S. Polezzo and M. Taramasso, *Chim. Ind. (Milan)*, 44 (1962) 33.
3 A. Campreciós and M. Gassiot, *Estudio de Métodos de Ajuste por Mínimos Cuadrados para Resolución de Picos Solapados en Cromatografía de Gases*, I.Q.S., Barcelona, 1971, p. 25.
4 M. Gassiot and B. Monrabal, *Chromatographia*, 3 (1970) 547.
5 E. Cusó, X. Guardino and M. Gassiot, *Afinidad*, 29 (1972) 965.
6 A. M. Anderson, T. C. Gibb and A. B. Littlewood, *Anal. Chem.*, 42 (1970) 434.
7 E. Grushka, H. N. Myers, P. O. Schettler and J. C. Giddings, *Anal. Chem.*, 41 (1969) 889.
8 H. M. Gladney, B. F. Dowden and J. D. Swalen, *Anal. Chem.*, 41 (1969) 883.
9 S. N. Chesler and S. P. Gram, *Anal. Chem.*, 45 (1973) 1354.
10 A. B. Littlewood, *Gas Chromatography: Principles, Techniques and Applications*, Academic Press, New York, 1962, pp. 125–130.
11 A. H. Anderson, T. C. Gibb and A. B. Littlewood, in C. L. A. Harbourn (Editor), *Gas Chromatography 1968*, Institute of Petroleum, London, 1969, p. 309.
12 A. H. Anderson, T. C. Gibb and A. B. Littlewood, *J. Chromatogr. Sci.*, 8 (1970) 640.
13 O. Grubner, *Anal. Chem.*, 43 (1971) 14.
14 O. Grubner, *Advan. Chromatogr.*, 6 (1968) 204.
15 V. L. Law and R. V. Bailey, *Chem. Eng. Sci.*, 18 (1963) 189.