Journal of Chromatography, 462 (1989) 3–12
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 20 998

CHROMATOGRAPHIC DISPERSION CORRECTIONS UTILIZING THE GENERALIZED EXPONENTIAL FUNCTION

R. D. HESTER*, R. A. VAIDYA and J. P. DICKERSON
Department of Polymer Science, University of Southern Mississippi, Southern Station Box 10076, Hattiesburg, MS 39406-0076 (U.S.A.)
(First received March 7th, 1988; revised manuscript received September 13th, 1988)

SUMMARY

Chromatographic dispersion, sometimes referred to as band broadening or zone spreading, reduces instrument resolution especially for long analysis times. Tung has described chromatographic dispersion as resulting from the convolution of a true chromatogram with a spreading function. This convolution produces an observed chromatogram which is broader than the true chromatogram. An algorithm which uses the generalized exponential function has been developed for solving Tung's dispersion equation for the true (non-dispersed) chromatogram. Mathematical constraints and data analysis methods have been developed which enable the generation of realistic dispersion-corrected chromatograms.

INTRODUCTION

Separation of components in chromatography can occur because of retention differences produced by molecular adsorption, component size or charge character. Proper selection of packing materials and mobile phase can maximize separation. Dispersion, also referred to as zone spreading, produces signal band broadening.

Several complex mass transfer or diffusional phenomena are responsible for dispersion[1]. These are related to packing material type, mesh size, column packing efficiency, mobile phase flow-rate and macromolecular size. Therefore, dispersion is unique to the chromatographic system. Dispersion corrections for one size-exclusion chromatography system cannot be applied to a different system or to different operating conditions within the same system. Dispersion is always present in chromatography and works to decrease separation capabilities. However dispersion corrections can be performed by applying mathematical techniques.

An observed chromatogram, h, is the result of a convolution between a true chromatogram, f, and an instrument spreading function, g.

$$h = f * g \tag{1}$$

The integral form of this convolution is given by eqn. 2.

0021-9673/89/$03.50

$$h(x) = \int_{-\infty}^{\infty} f(u)\, g(x - u)\, du \tag{2}$$

As shown by eqn. 2, the value of the observed chromatogram at each elution volume x or $h(x)$ is obtained by an integration involving the true chromatogram and a shifted spreading function.

Tung[2] defined the spreading function g as a normal function of mean zero and standard deviation s.

$$g(x - u) = \frac{1}{s\sqrt{2\pi}} \exp\left[-\frac{(x - u)^2}{2s^2}\right] \tag{3}$$

The standard deviation s, usually called the spreading factor, controls the amount of dispersion produced on each element, $f(u)\, du$, of the true chromatogram. The observed chromatogram at elution volume $h(x)$ is the sum of the dispersion produced on each element of the true chromatogram evaluated at x. In exclusion chromatography, eqn. 2 is generally referred to as Tung's dispersion equation.

Observed chromatograms of standards having known or true distributions can be used to solve eqn. 2 for the unique single spreading factor associated with the instrument. By knowing the spreading factor, Tung's equation can thereafter be inverted (solved) to obtain the dispersion-corrected or true chromatogram for other samples analyzed by the instrument.

Solution of Tung's dispersion equation for function $f(u)$ is difficult and usually requires a computer program involving an algorithm which solves eqn. 1. Several algorithms have been previously developed[3-5]; however, because of the ill-conditioned nature of eqn. 1, the true chromatogram solution produced is usually unsatisfactory, especially if the observed chromatographic data contain random signal noise[6].

In this work we will estimate a solution to eqn. 1 by making the assumption that the true chromatogram $f(u)$ can be described by a modified generalized exponential (GEX) function[7]. In the past, we have used this function extensively to fit noisy chromatographic data[8]. We have shown that the GEX function is very general in nature and can fit negatively or positively skewed data with low or high kurtosis (flatness). The overall generality of the GEX makes the fixed chromatographic shape assumption less limiting. We will use constrained non-linear regression analysis to solve for the GEX function parameters which best fit Tung's dispersion equation. The best fit GEX parameters will then define an estimate of the true chromatogram.

THE GENERALIZED EXPONENTIAL FUNCTION*

The modified GEX function which will be used to define the true or dispersion corrected chromatogram is given by eqn. 4.

* For symbols, see the list at the end of the paper.

For $u > u_o$,

$$f(u) = u_h \left[\frac{u - u_o}{u_m - u_o}\right]^{B-1} \exp\left\{\frac{B-1}{A}\left(1 - \left[\frac{u - u_o}{u_m - u_o}\right]^{A}\right)\right\}$$

with $A > 0$ and $B > 1$.

For $u \leqslant u_o$,

$$f(u) = 0 \tag{4}$$

The GEX has five parameters: u_o, u_m, u_h, A and B. The values of u_o, u_m and h_m can be easily obtained from the chromatogram (see Fig. 1). The A and B parameters are related to the shape of the GEX function and define the inflection points (u_1, u_{1h}; u_2, u_{2h}). It can be shown that:

$$u_1 = u_o + (u_m - u_o)t^+ \tag{5}$$

$$u_2 = u_o + (u_m - u_o)t^- \tag{6}$$

with $t^{\pm} = \left\{\frac{3 - A - 2B \pm \sqrt{A^2 - 6A + 4AB + 1}}{2 - 2B}\right\}^{\frac{1}{A}}$

$$u_{1h} = f(u_1) \tag{7}$$

$$u_{2h} = f(u_2) \tag{8}$$

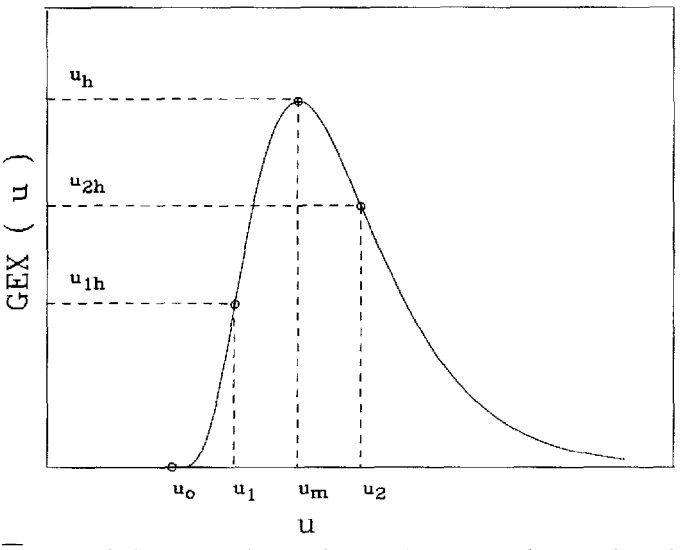

Fig. 1. Typical GEX function starting at u_o having a maximum at (u_m, u_h) and inflection points at (u_1, u_{1h}) and (u_2, u_{2h}).

Also it has been shown that the nth moment of a GEX function about an axis parallel to the ordinate at an elution volume equal to z is given by[9]:

$$\gamma_n^z = \sum_{r=0}^{n} \Theta_r \left[\frac{A}{B-1}\right]^{\frac{B+r}{A}} \Gamma\left[\frac{B+r}{A}\right] \tag{9}$$

where $\Theta_r = \frac{u_h}{A} {}^nC_r \, \beta^{r+1} \, \alpha^{n-r} \exp[(B-1)/A]$

$$\beta = u_m - u_o$$

$${}^nC_r = \frac{n!}{r!\,(n-r)!}$$

$$\alpha = u_o - z$$

Eqn. 9 can be used to determine the area within the GEX function (area = γ_0^0), the mean of the function (mean = γ_1^0/area), variance of the function (variance = γ_2^{mean}/area), and skewness of the function {skewness = γ_3^{mean}/[area (variance)$^{3/2}$]}, and kurtosis of the function {kurtosis = γ_4^{mean}/[area (variance)^2]}.

REGRESSION ANALYSIS

In regression analysis an objective function which has a set of adjustable parameters is optimized. Optimization involves maximizing or minimizing an objective function. In non-linear regression, optimization will usually converge provided that the initial start values of the parameters (first-guess values) are not too distant from the best fit parameter values. A successful regression is crucially dependent upon having an objective function that has a minimum number of parameters and also upon having "good" first guess parameter values.

To solve Tung's equation a least squares objective function of the form

$$\sum_{i=1}^{N} \left[h(x_i) - \int_{-\infty}^{\infty} f(u)\, g(x_i - u)\, du\right]^2 \tag{10}$$

was minimized by using a Levenberg–Marquardt regression algorithm[10]. The objective function is the sum of N terms each of which is the square of the difference between the observed chromatogram and the convolution integral both evaluated at elution volume x_i. x_1 and x_N are the first and last elution volumes respectively which span the total elution volume over which the observed chromatogram signal is detected.

Unfortunately there exists a large number of significantly different parameter sets each of which adequately minimize the objective function. Thus the regression is ill-conditioned and the parameter solution set found may not be realistic. To

compensate for the ill-conditioned nature of the regression, constraints must be used to limit the number of possible parameter solution sets. If the constraints are not adequate, the regression will probably find a solution which is distant from the correct solution. Therefore it is critical that regression constraints be found that are both realistic and as confining as possible.

REGRESSION CONSTRAINTS

Inherent in the GEX function are three intrinsic constraints which realistically confine the regression solution space. These are (1) continuity of signal; (2) non-negativity of signal; and (3) signal appearance over a finite range.

Most single-component chromatographic signals are continuous and smooth. The GEX function satisfies this requirement. We do not normally expect to see both negative and positive appearance of a single component at a detector. Thus the chromatographic signal should always be positive and the GEX function answers this requirement. A chromatographic signal should only deviate from non-zero vlues during the time interval over which material is eluting through a detector. The GEX function meets this condition because it is zero for elution volume values less than u_0 and also because it approaches zero as the elution volume becomes large.

Unfortunately, the above inherent constraints are usually not sufficient to restrict the regression solution space. Additional constraints are necessary to insure reasonable regression convergence to an acceptable solution.

Additional constraints can be imposed by noting that in the convolution h of two functions, g and f that[11] (1) the area of the convolution is the product of the areas of the functions g and f; (2) the mean of the convolution is the sum of the means of functions g and f; and (3) the variance of the convolution is the sum of the variances of functions g and f.

The spreading function g has unit area, zero mean, and variance s^2. Thus the true chromatogram or function f has an area and mean equal to the area and mean of the observed chromatogram or the convolution h. Also, the variance of the function f is the variance of the convolution h less the variance of the spreading function g.

We can easily calculate the area S, mean $\bar{X}$, and variance σ^2 of he observed chromatogram. By using eqn. 9, the area, mean and variance of the GEX function representing the true chromatogram can be expressed in terms of the GEX parameters. We can use this knowledge to eliminate three of the five GEX parameters. The most convenient parameters to eliminate are u_o, u_m and u_h.

$$u_o = \bar{X} - \sqrt{\sigma^2/(C - 1)} \tag{11}$$

with $C = \Gamma[B/A]\ \Gamma\ [(B + 2)/A]/\{\Gamma[(B + 1)/A]\}^2$

$$u_m = u_o + (\bar{X} - u_o)D \tag{12}$$

with $D = [(B - 1)/A]^{1/A}\ \Gamma(B/A)/\Gamma[(B + 1)/A]$

$$u_h = SE/(u_m - u_o) \tag{13}$$

with $E = A \exp [(1 - B)/A]\ [(B - 1)/A]^{B/A}/\Gamma(B/A)$

The true chromatogram function then becomes a GEX function having only two shape parameters. The regression algorithm must find the two shape parameters (A,B) which best minimize the much simplified and more constraining objective function.

Our past experience in fitting chromatograms has shown that the GEX shape parameters, A and B almost always have values that are less than ten. Also the GEX function only exists in real space when A and B have values that are greater than zero and one, respectively. Acceptable first guess values for the shape parameters are the shape parameters found by fitting a GEX function to the observed chromatogram. Table I summarizes the information needed to set up parameter first guess values and their constraints when regression is used to find an estimate of the true chromatogram.

It should be emphasized that the regression solution constraints that have been developed apply only to a Gaussian spreading function in which the spreading factor, s, is constant. If the spreading function was made to vary with the elution volume then some of the constraints would not necessarily apply.

EXPERIMENTAL

A non-linear regression routine was developed which contains the objective function (eqn. 10) and incorporates the parameter constraints previously discussed. A computer program of this regression, named "CDC" for chromatogram dispersion correction, was generated using a Pascal compiler (Turbo Pascal 3.0, Borland International) operating with a Z148 personal computer (Zenith Data Systems) having a 8087 math coprocessor. The math coprocessor enable calculations to be performed on real data with 16 digits accuracy within a range of $4.19 \cdot 10^{-307}$ to $1.67 \cdot 10^{+308}$.

Two dispersed chromatograms were produced using the convolution operator, eqn. 2. GEX functions were used to represent the true (non-dispersed) chromatograms and the normal spreading functions were defined by assigning specific spreading factors. Both dispersed chromatograms, made by the convolutions which represent the observed chromatograms, were each fitted to a GEX function so that first guess values for the shape parameters could be made for the the program CDC. The computer program CDC was then used to reverse the convolution or invert eqn. 2. Thus CDC deconvolutes the obeserved (dispersed) chromatograms and thereby should return the original GEX functions representing the true chromatograms. Using the above technique it was possible to gauge the ability of the program CDC to deconvolute typical chromatogram data.

TABLE I

TRUE CHROMATOGRAM REGRESSION SOLUTION SPACE

Δ = Small positive amount, *i.e.*, 0.01.

Parameter	*Maximum*	*Minimum*	*First guess value*
A	10	Δ	A_x
B	10	$1 + \Delta$	B_x

TABLE II

CASE 1 CHROMATOGRAMS

Information	*Symbol*	*True chromatogram*	*Observed or dispersed chromatogram, $s = 2.5$*	*CDC estimate of the true chromatogram*
Shape parameter	A	3.14	2.99	3.03
Shape parameter	B	3.45	5.62	4.73
Signal start point	u_o	36.0	30.0	34.3
Maximum of signal	u_m	45.0	45.0	45.0
Maximum height	u_h	2.06	1.61	2.07
1st Inflection point	u_1	41.6	40.9	41.7
2nd Inflection point	u_2	48.3	49.1	48.2
Area	S	0.161	0.161	0.161
Mean	$\bar{X}$	45.1	45.05	45.05
Variance	σ^2	9.00	15.4	9.15
Skewness		0.105	0.068	0.079
Kurtosis		2.74	2.85	2.82

RESULTS AND DISCUSSION

Tables II and III give information on the two cases tested. Case 1 was intended to test CDC performance when a true chromatogram is nearly Gaussian or normal in shape and is only slightly dispersed ($s = 2.5$). Case 2 was designed to evaluate CDC when a true chromatogram is skewed right and has been severely dispersed ($s = 10.0$). The three chromatogram plots associated with each test case are shown in Figs. 2 and 3.

Case 1 is a less demanding test of CDC capability than Case 2. The data in Table II and Fig. 2 show that CDC deconvoluted the dispersed chromatogram with a great

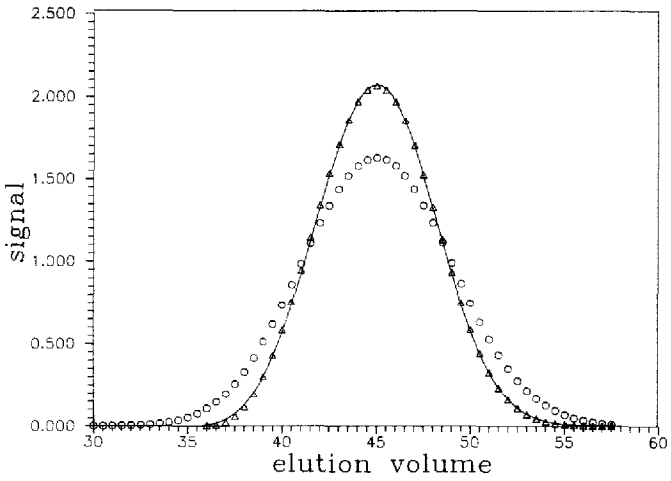

Fig. 2. Test Case 1 Chromatograms. Circles and triangles are data points for dispersed and true chromatograms, respectively. Solid line is the CDC estimate of the true chromatogram.

TABLE III

CASE 2 CHROMATOGRAMS

Information	*Symbol*	*True chromatogram*	*Observed or dispersed chromatogram* $s = 10.0$	*CDC estimate of the true chromatogram*
Shape parameter	A	0.50	2.74	0.663
Shape parameter	B	8.50	4.30	9.11
Signal start point	u_o	36.0	12.0	34.4
Maximum of signal	u_m	45.0	47.5	45.4
Maximum height	u_h	1.00	0.430	0.932
1st Inflection point	u_1	40.5	35.3	40.3
2nd Inflection point	u_2	49.6	59.4	50.5
Area	S	12.5	12.5	12.5
Mean	$\bar{X}$	48.2	48.0	48.0
Variance	σ^2	36.2	135.7	35.7
Skewness		1.26	0.162	0.962
Kurtosis		5.70	2.83	4.52

deal of accuracy. The true chromatogram and the CDC-estimated chromatogram made from inverting the dispersed chromatogram are almost identical.

Although Case 2 is a harsh and rigorous test of CDC, the computer algorithm performed much better than expected. The data in Table III and the chromatogram plots in Fig. 3 show that the estimate of the true chromatogram made by CDC is very close to the true chromatogram.

The time required for CDC to converge to a deconvolution solution was 630 and 760 s for Cases 1 and 2, respectively. Thus computer time requirements are reasonable even for a microcomputer. Computation time would have been much less (probably two orders of magnitude less) on a mainframe computer and/or if a more efficient regression algorithm was used.

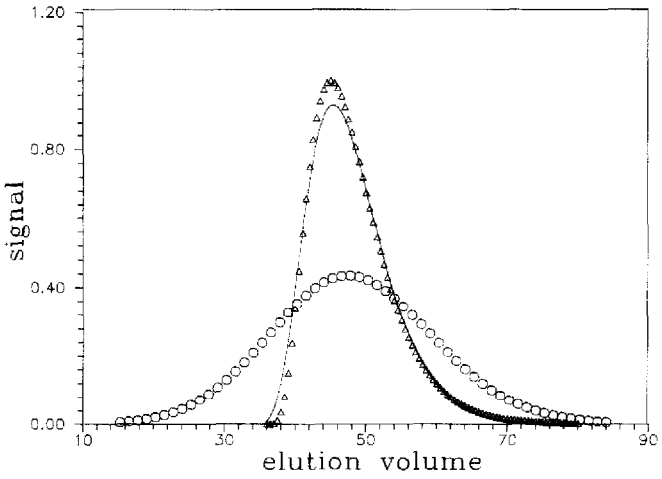

Fig. 3. Test Case 2 Chromatograms. Circles and triangles are data points for dispersed and true chromatograms respectively. Solid line is the CDC estimate of the true chromatogram.

CONCLUSIONS

The use of a GEX function and constrained non-linear regression has been shown to be a very effective in correcting the dispersion produced in chromatography. This is possible because realistic regression solution constraints have been developed which minimize the computational time required to reach a solution and simultaneously reduce the possibility of converging to a false solution. Because of the above features, the regression can be performed on a personal computer.

SYMBOLS

A	first shape parameter of the GEX function
B	second shape parameter of the GEX function
C	factor in eqn. 11
${}^nC_r$	combination of n taken r at a time, see eqn. 9
D	factor in eqn. 12
E	factor in eqn. 13
f	true chromatogram
g	spreading function
h	observed chromatogram
n	order of a moment
N	total number of data points
r	index variable
S	area
s	spreading factor, see eqn. 2
u	true chromatogram elution volume
u_o	GEX starting point
u_1	abscissa value of the first GEX inflection point
u_2	abscissa value of the second GEX inflection point
u_h	GEX function maximum signal value
u_m	abscissa point of GEX maximum signal
u_{1h}	signal value of the first GEX inflection point
u_{2h}	signal value of the second GEX inflection point
t^+, t^-	factor used in eqn. 6
x	observed chromatogram elution volume
$\bar{X}$	mean elution volume value
z	axis for a moment (axis $\equiv u = z$)
α	factor used in eqn. 9
β	factor used in eqn. 9
Γ	gamma function
γ_n^z	nth moment about axis z
Θ_r	factor used in eqn. 9
σ^2	variance

REFERENCES

1 W. W. Yau, J. J. Kirkland and D. D. Bly, *Modern Size-Exclusion Liquid Chromatography*, Wiley, New York, 1979.
2 L. H. Tung, *J. Appl. Polym. Sci.*, 10 (1966) 375.
3 M. Kubin, D. Horak and F. Svec, *Angew. Makromol. Chem.*, 71 (1978) 51.
4 J. A. M. Smit, C. J. P. Hoogervorst and A. J. Staverman, *J. Appl. Polym. Sci.*, 15 (1971) 1479.
5 P. E. Pierce and J. E. Armonas, *J. Polym. Sci.*, 21 (1968) 23.
6 K. S. Chang and R. Y. M. Huang, *J. Appl. Polym. Sci.*, 13 (1969) 1459.
7 R. A. Vaidya and R. D. Hester, *J. Liq. Chromatogr.*, 287 (1984) 231.
8 C. E. Lundy and R. D. Hester, *J. Liq. Chromatogr.*, 7(10) (1985) 1911.
9 R. A. Vaidya and R. D. Hester, *J. Chromatogr.*, 333 (1985) 152.
10 J. L. Kuester and J. H. Mize, *Optimization Techniques with Fortran*, McGraw-Hill, New York, 1973, p. 240.
11 R. N. Bracewell, *The Fourier Transform and Its Application*, McGraw-Hill, New York, 2nd ed., 1986, p. 302.