Comparison of equations describing band broadening in high-performance liquid chromatography

L. Kirkup [a,*], M. Foot [b], M. Mulholland [b]

a *Department of Applied Physics, University of Technology, P.O. Box 123, Broadway, Sydney, NSW 2007, Australia*
b *Department of Chemistry, Forensic and Materials Science, University of Technology, Sydney, NSW 2007, Australia*

Abstract

Several models are well established that describe band broadening in gas and liquid chromatography, including those due to Van Deemter and Knox. Comparison of competing models is complicated if raw data are noisy or if the equations to be fitted to data contain many adjustable parameters. This paper describes a comparison of fitting the Van Deemter, Knox and other equations to low noise data gathered during the separation of propyl- and methylparaben by HPLC. Equations are compared using established statistical methods, including analysis of residuals, inference of parameter estimates and Akaikes Information Criterion for model identification. This work indicates that equations that account for non-linear band broadening at elevated mobile phase velocities are more successful at describing the relationship between height equivalent to a theoretical plate, H, and the velocity of the mobile phase, u.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Band broadening; Dispersion; Mathematical modelling

1. Introduction

Models of band broadening in LC and GC can be traced back to work carried out in the middle of the 20th century. An understanding of the mechanisms underlying band broadening permits improved columnar design as well a predicting what conditions must be contrived in order to minimise the height equivalent to a theoretical plate, H, for any given column. Several relationships between H and the linear flow velocity, u, have been proposed [1]. Work of Van Deemter et al. [2] led to a model describing the relationship between H and u which is widely cited in the literature. This model has three terms:

(a) Path dependent term: solute molecules take different paths through the stationary phase. The path lengths of individual molecules, therefore, differ from one another in a random fashion, leading to band broadening. This term in the Van Deemter model depends on the diameter of particles packed into the column, but is independent of u.

(b) Analyte diffusion term: any analyte injected into a column will tend to diffuse, leading to band broadening of the chromatographic zone. The less the velocity of the mobile phase, the greater will be effect of the diffusion. In the Van Deemter model, this term has a $1/u$ dependence.

(c) Mass transfer term: the analyte must equilibrate over time between the stationary and mobile phases. When the mobile phase velocity is large and the analyte has a strong affinity with the stationery phase, the analyte in the mobile phase moves ahead of analyte in the stationary phase, causing band broadening. This term is proportional to u.

Several other models which account for band broadening have been proposed and the functional relationship between band broadening and flow velocity remains a matter of debate [3]. While major contributions to band broadening are attributed to columnar effects, there is evidence that in many experimental situations, extra columnar effects are important. For example, broadening may be due to finite time constants of the electronic system used to detect and record changes in the response variable in an HPLC system. Studies in GC have shown that such broadening is non linear in u. [4]. The work reported here includes an investigation of the applicability of the model in [4] to HPLC data.

---

* Corresponding author. Tel.: +61-2-95142206; fax: +61-2-95142219.
E-mail address: les.kirkup@uts.edu.au (L. Kirkup).

This paper offers an inter-comparison of several established models of band broadening applied to data obtained in the separation of propyl- and methylparaben by HPLC. The comparison is extended to a novel empirical equation that attempts to quantitatively describe non-linear band broadening that occurs at high mobile phase velocities.

2. Theory

Popular equations fitted to H versus u data include those due to Van Deemter and Knox, given by Eqs. (1) and (2), respectively [1,5].

$$H = a + \frac{b}{u} + cu \tag{1}$$

$$H = au^{1/3} + \frac{b}{u} + cu \tag{2}$$

a, b and c are constants.

More recently, a study on band broadening in gas–liquid chromatography has indicated that for a wide range of experimental data the so called Golay–Guiochon equation given by:

$$H = \frac{b}{u} + cu + du^2 \tag{3}$$

better describes the relationship between H and u. The origin of the term in u^2 was attributed to extra columnar broadening caused by time constants in the measurement system [6].

As curvature in H versus u data at large values of u is quite common, an empirical equation to account for the curvature is proposed. The equation is:

$$H = \frac{b}{u} + f \exp(gu) \tag{4}$$

where b, f and g are constants.

While Eq. (4) appears to be quite different to Eqs. (1)–(3), it does reduce to forms similar to Eqs. (1) and (3) in certain circumstances. A series expansion of the second term in Eq. (4) gives:

$$H = \frac{b}{u} + f\left[1 + gu + \frac{(gu)^2}{2!} + \text{higher order terms}\right] \tag{5}$$

$$H = \frac{b}{u} + f + (fg)u + \left(\frac{fg^2}{2!}\right)u^2 + \text{higher order terms} \tag{6}$$

In situations in which the higher order terms in u^2 and above can be neglected, Eq. (6) can be approximated by,

$$H \approx \frac{b}{u} + f + fgu \tag{7}$$

Eq. (7) can be compared with Eq. (1), suggesting:

$$a \equiv f, \tag{8}$$

and

$$c \equiv fg \tag{9}$$

For larger values of u, it may be required to retain the term in u^2 appearing in Eq. (6), thus creating an equation similar to Eq. (3) (but including an additional term independent of u).

3. Materials and methods

3.1. Chemicals and samples

Samples for analysis were methyl- and propylparaben both of which were of above 99% purity. The mobile phase used was 60% Omnisolv HPLC gradient analysis grade acetonitrile (EM Science Gibbstown, NJ, USA) and 40% Milli-Q water (Millipore, WI, USA). Samples were prepared by dissolving 100 mg of the methyl- and propylparaben in 100 ml of mobile phase. The samples were filtered through 0.45 μm syringe filters (Bonnet Equipment, Taren Point, Australia).

3.2. HPLC analysis

HPLC analysis was carried out on a Waters 2690 separations module linked to a Waters 996 photodiode array detector (Waters Australia, Rydalmere, Australia). The experiments were carried out on a Phenomenex Phenossphere-NEXT C_8 column (150 mm × 4.6 mm, 5 μm; Phenomenex, USA). The instrument was controlled using Waters Millennium version 3.20 software (Waters, Milford, MA, USA). Column efficiencies (N) and other chromatographic information were calculated by the system suitability add-on to Millennium.

Flow rates were chosen in the range 0.1–2.0 ml/min with intervals of 0.1 ml/min. Three replicate injections at each flow rate were performed and the injection volume was 10 μl. The diode array detector was set to scan 220–260 nm and the chromatogram from 254 nm was extracted for this work. The resolution of the detector was 1.2 nm. The sampling period was 15 ms. All experiments were performed at 40 °C. Linear flow velocity was calculated by dividing column length by the dead volume time.

4. Results and discussion

H versus u data for propylparaben are shown in Fig. 1. The maximum standard error in H values, expressed as a percentage of mean H, did not exceed 3% for any flow velocity. Examination of the data indicates that at large u, the relationship between H and u is non-linear, suggesting that Eqs. (3) and (4) may be more appropriate than either Eq. (1) or Eq. (2).

Inspection of Fig. 1 indicates that the scatter of data increases as H increases, i.e., the data display heteroscedasticity. In order to properly account for the scatter, fitting by weighted least squares should be adopted. Eqs. (1)–(4) were fitted to data by minimising the weighted sum of squares of
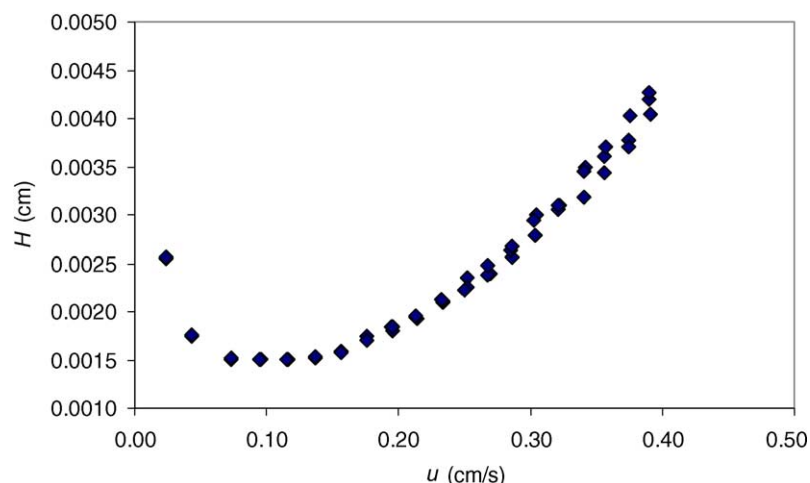
Fig. 1. H vs. u data for propylparaben separated by HPLC.

residuals (WSSR), given by [7]:

$$\text{WSSR} = \sum \left(\frac{H_i - \hat{H}_i}{\sigma_i} \right)^2 \tag{10}$$

where H_i is the ith experimental value, $\hat{H}_i$ is the ith predicted value based on the equation fitted to the data, and σ_i is the ith standard deviation. It is assumed that the scatter can be well described by:

$$\sigma_i \propto H_i \tag{11}$$

Weighted fitting was carried out on a personal computer using GraphPad Prism [8].

As Eqs. (1)–(3) are linear in the fitting parameters, these parameters may be estimated using linear least squares [9]. By contrast, Eq. (4) is non-linear in the fitting parameters and so fitting by non-linear least squares is required [10].

To establish the effect of weighting in the manner described by Eq. (10), residuals are shown in Fig. 2 for Eq. (4) fitted to the data in Fig. 1 using unweighted and weighted last squares.

Residuals in Fig. 2a exhibit a characteristic typical of heteroscedastic data, namely the residuals becoming more scattered as u increases [11]. By contrast, the weighted residuals in Fig. 2b demonstrate minimal trend after fitting using weighted least squares. To allow for comparison of parameter estimates and standard errors in estimates, unweighted fitted using least squares was also carried out and is reported here.

Fitting was carried out on methyl- and propylparaben data obtained by HPLC, and all necessary statistics are reported. As results were similar for both methyl- and propylparaben, graphical plots shown are for propylparaben only. Fig. 3 shows the weighted fits for Eqs. (1)–(4) fitted to H versus u data obtained for propylparaben.

Inspection of the fitted lines in Fig. 3 indicates that those generated by fitting Eqs. (1) and (2) deviate progressively

from the data above flow velocities of about 0.3 cm/s. While lines based Eqs. (3) and (4) appear to follow the data more closely, it is difficult to establish clearly the quality of the fit. In order to make any model violation more apparent, the weighted residuals, given by:

$$i\text{th weighted residual} = \frac{H_i - \hat{H}_i}{H_i} \tag{12}$$

are plotted against u. The weighted residuals determined when Eqs. (1), (3) and (4) are fitted to data in Fig. 1 are shown in Fig. 4. As Eq. (2) is very similar to Eq. (1) with respect to the fitted line it generates, the weighted residuals for Eq. (2) are omitted from Fig. 4.

As a guide, lines are shown connecting the middle of the three replicate points at each flow velocity in Fig. 4. The
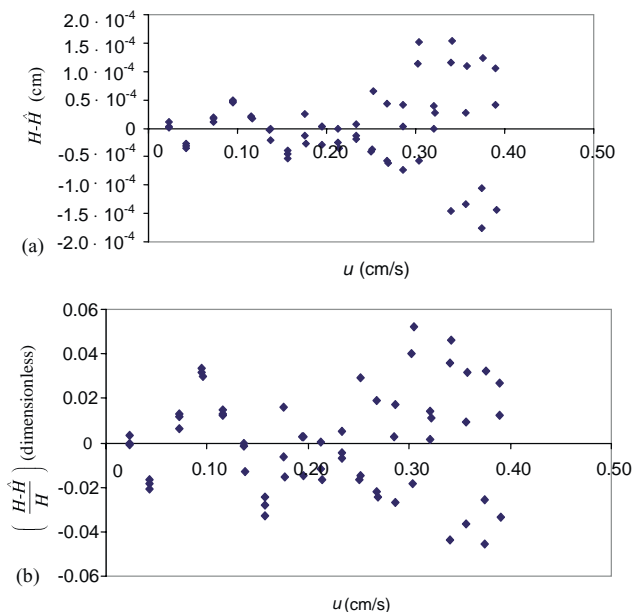

Fig. 2. Residuals for: (a) unweighted and (b) weighted fitting of Eq. (4) to data in Fig. 1.
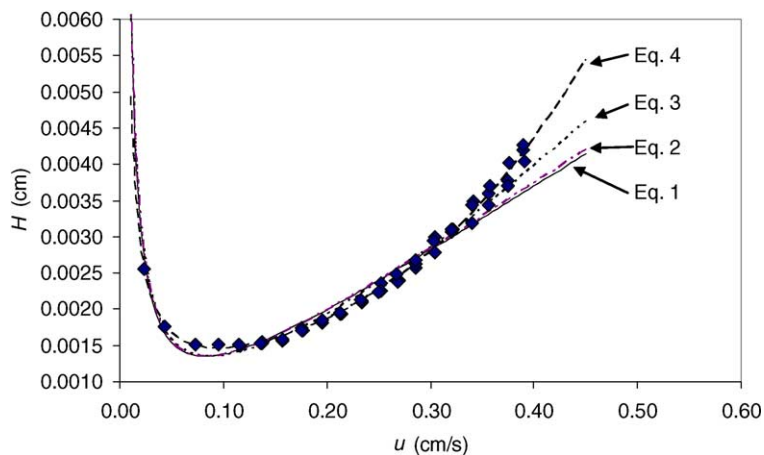
Fig. 3. Eqs. (1)–(4) fitted propylparaben to data.

weighted residuals for Eqs. (1) and (3) fitted to data exhibit the shape recognised as characteristic of model violations. The weighted residuals determined when Eq. (4) is fitted to data are smaller for Eqs. (1) and (3) and exhibit no clear trend.

Table 1 contains the best estimates of parameters appearing in Eqs. (1)–(4) along with their respective standard errors, for both unweighted (UW) and weighted (W) fitting to data in Fig. 1. The statistical significance of the parameter estimates can be established using a *t*-test [12]. Applying such a test reveals that all parameter estimates for both unweighted and weighted fitting of Eqs. (1)–(4) are significant at the 0.05 level of significance.

The standard error in each parameter may be expressed as a percentage of the parameter estimate. The only term common to Eqs. (1)–(4) is that incorporating the *b* parameter estimate. Table 2 shows $(\sigma_b/b) \times 100\%$ for all equations fitted to methyl- and propylparaben by weighted least squares.

Table 2 contains the weighted sum of squares of residuals, WSSR, the coefficient of determination R^2 and the Akaikes Information Criterion (AIC). AIC allows for goodness of fit comparisons when fitting equations have differ-

ing numbers of adjustable parameters [13]. This criterion takes into account the WSSR, but also includes a term proportional to the number of parameters used. AIC may be written,

$$\text{AIC} = n \ln \text{WSSR} + 2M \tag{13}$$

where n is the number of data and M the number of parameters in the equation. The best equation fitted to data is that which gives the lowest value for the AIC. In the work described here all fitted equations have three adjustable parameters, allowing WSSR to act as an adequate goodness of fit criterion to establish the 'best' equation, therefore, it is not required that the AIC be used. AIC has been included as there are other equations which describe band broadening in LC that have more than three adjustable parameters (for example in [14,15]), therefore, a criterion such as the AIC would be required to assist in the determination of the goodness of fit. Examination of Table 2 indicates that the WSSR (and hence the AIC) is much less for Eq. (4) fitted to data than for Eqs. (1)–(3). R^2 is also rather better for Eq. (4) than Eqs. (1)–(3).
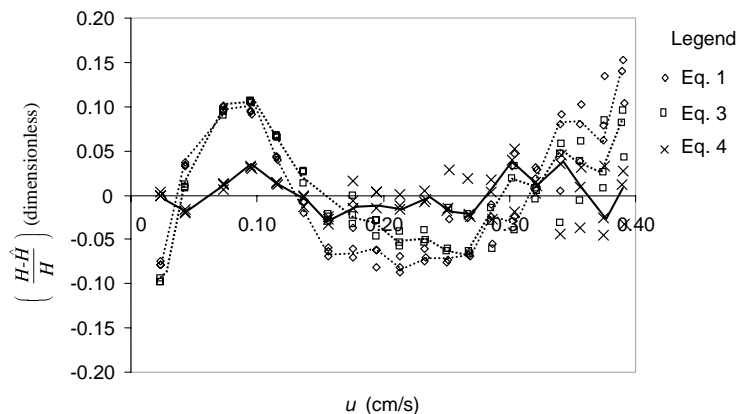

Fig. 4. Weighted residuals for Eqs. (1), (3) and (4) fitted to propylparaben data.

L. Kirkup et al. / J. Chromatogr. A 1030 (2004) 25–31

29

Table 1
Parameter estimates and standard errors in parameter estimates

	Equation	Weighting	Parameter estimates						Standard errors in estimates					
			a	b	c	d	f	g	σ_a	σ_b	σ_c	σ_d	σ_f	σ_g
Methylparaben	(1)	UW	−0.00123	8.39×10^{-5}	0.0178	N/A	N/A	N/A	0.000170	6.60×10^{-6}	0.000559	N/A	N/A	N/A
		W	−0.00061	6.98×10^{-5}	0.0150	N/A	N/A	N/A	0.000128	4.94×10^{-6}	0.000541	N/A	N/A	N/A
	(2)	UW	−0.00306	7.40×10^{-5}	0.0207	N/A	N/A	N/A	0.000365	4.91×10^{-6}	0.008097	N/A	N/A	N/A
		W	−0.00165	6.58×10^{-5}	0.0168	N/A	N/A	N/A	0.000293	3.75×10^{-6}	0.000767	N/A	N/A	N/A
	(3)	UW	N/A	5.89×10^{-5}	0.0065	0.0238	N/A	N/A	N/A	2.33×10^{-6}	0.000479	0.00149	N/A	N/A
		W	N/A	5.90×10^{-5}	0.0080	0.0183	N/A	N/A	N/A	1.80×10^{-6}	0.000433	0.00161	N/A	N/A
	(4)	UW	N/A	3.56×10^{-5}	N/A	N/A	0.000804	5.36	N/A	2.00×10^{-6}	N/A	N/A	2.40×10^{-5}	0.0880
		W	N/A	3.67×10^{-5}	N/A	N/A	0.000775	5.47	N/A	1.10×10^{-6}	N/A	N/A	1.45×10^{-5}	0.0637
Propylparaben	(1)	UW	−0.00042	6.74×10^{-5}	0.0104	N/A	N/A	N/A	0.000095	3.68×10^{-6}	0.000312	N/A	N/A	N/A
		W	−0.00019	6.34×10^{-5}	0.0093	N/A	N/A	N/A	0.000080	3.46×10^{-6}	0.000308	N/A	N/A	N/A
	(2)	UW	−0.00109	6.46×10^{-5}	0.0115	N/A	N/A	N/A	0.000210	2.82×10^{-6}	0.000466	N/A	N/A	N/A
		W	−0.00054	6.28×10^{-5}	0.0100	N/A	N/A	N/A	0.000184	2.74×10^{-6}	0.000454	N/A	N/A	N/A
	(3)	UW	N/A	6.00×10^{-5}	0.0060	0.00976	N/A	N/A	N/A	1.55×10^{-6}	0.000318	0.00099	N/A	N/A
		W	N/A	6.16×10^{-5}	0.0067	0.00711	N/A	N/A	N/A	1.59×10^{-6}	0.000310	0.00109	N/A	N/A
	(4)	UW	N/A	4.25×10^{-5}	N/A	N/A	0.000648	4.70	N/A	9.98×10^{-7}	N/A	N/A	1.35×10^{-5}	0.0619
		W	N/A	4.26×10^{-5}	N/A	N/A	0.000647	4.70	N/A	8.05×10^{-7}	N/A	N/A	1.00×10^{-5}	0.0515

Table 2
Standard error in b expressed as a percentage, WSSR, R^2 and AIC obtained by fitting Eqs. (1)–(4) to methyl- and propylparaben data

	Equation	$(\sigma_b/b) \times 100\%$	WSSR	R^2	AIC
Methylparaben	(1)	7.1	0.5380	0.929	−31
	(2)	5.7	0.4851	0.940	−37
	(3)	3.1	0.2306	0.979	−82
	(4)	3.0	0.05921	0.991	−164
Propylparaben	(1)	5.5	0.3015	0.939	−66
	(2)	4.4	0.2862	0.944	−69
	(3)	2.6	0.1887	0.971	−94
	(4)	1.9	0.03215	0.993	−200

5. Discussion

The least successful equation, with respect to fitting data obtained in this study, is the Van Deemter equation (Eq. (1)). In fact, the fit based on Eq. (1) is only marginally worse than that obtained with Eq. (2) (the Knox equation). The main source of difficulty is the curvature in the H versus u data at large u, which cannot be qualitatively or quantitatively accounted for by either Eq. (1) or Eq. (2).

The predictive ability of each of the four equations is shown in Table 3. The predicted optimum linear flow rate and H are closer to the measured values for Eq. (4) with all other methods overestimating the column performance at optimum flow rate. The ability of each equation to predict the rapid observed decrease in performance at large flow rates is also shown in Table 3. Eq. (4) estimated the H at the maximum flow rate closest to the measured H, all other equations excepting the weighted model for Eq. (3) under-estimated the H.

The constant term in Eq. (1) can be related to the average diameter, d_p, of the particles in the column by Eq. (2), as follows:

$$a = 2\lambda d_p \tag{14}$$

where λ is a positive constant, the value close to 0.6 [16].

As the quantities on the right-hand side of Eq. (14) are positive, a must be positive. Inspection of Table 1 shows that for both Eqs. (1) and (2), a is negative when weighted and unweighted fitting is carried out on both propyl- and methyl-paraben data. A negative value for a cannot be defended on physical grounds. The reason for the negative sign for a can be traced to the curvature in the data appearing in Fig. 1. Fitting a straight line to the data at large u, as effectively occurs when fitting the Van Deemter and Knox equations, yields a line with a large slope. When that line is extrapolated to the H axis, an intercept (and hence a value) is obtained which is negative.

Eq. (8) indicates the equivalence of the term a in Eq. (1) with f in Eq. (4). Table 1 shows that a weighted fit of Eq. (4) to propylparaben data gives $f = 6.47 \times 10^{-4}$ cm. Using Eq. (14) leads to a prediction for the value for d_p (assuming $\lambda = 0.6$) of $d_p = 5.4 \times 10^{-4}$ cm, i.e. 5.4 µm. A similar calculation for methylparaben data gives $d_p = 6.5$ µm. These diameters are of the right order of magnitude for the particle diameter in the packed column, which was 5 µm. Apparent consistency between the pre-exponential factor in Eq. (4) and the column particle size is encouraging, though more work is required if Eq. (4) is to be regarded as anything other than empirical.

As Eq. (3) allows for curvature in the H versus u data, through the u^2 term, it is more successful than either Eq. (1) or Eq. (2) at accounting for the shape of the curve. This can be seen quantitatively through examination of Table 2. The value of WSSR for Eq. (3) is less than half that calculated when Eqs. (1) and (2) were fitted to the propylparaben data. Nevertheless, at large u there remain systematic deviations of the line of best fit based on Eq. (3) from the data, as revealed by the pattern of weighted residuals in Fig. 4.

On all statistical measures, Eq. (4), is superior to the other equations considered in this study. Perhaps the most convincing indicator of goodness of fit is the random scatter of weighted residuals as shown in Fig. 4. It is not possible to say at this stage whether the non-linearity in H versus u in the data reported here is due to intra or extra columnar broadening and work is in progress to clarify the dominant broadening mechanisms.

There are many other equations that describe band broadening in HPLC that could be fitted to data. It is important

Table 3
Predicted H and optimum linear flow rates for each equation

	Sample	Actual H (cm)		Eq. (1)		Eq. (2)		Eq. (3)		Eq. (4)	
				u (cm/s)	H (cm)	u (cm/s)	H (cm)	u (cm/s)	H (cm)	u (cm/s)	H (cm)
Optimum	Propylparaben	0.00150	Unweighted	0.080	0.001256	0.082	0.001254	0.088	0.001289	0.094	0.001460
			Weighted	0.082	0.001351	0.084	0.001345	0.086	0.001345	0.094	0.001460
	Methylparaben	0.00165	Unweighted	0.068	0.001215	0.070	0.001241	0.076	0.001410	0.076	0.001676
			Weighted	0.068	0.001439	0.070	0.001439	0.076	0.001487	0.076	0.001658
Maximum	Propylparaben	0.00430	Unweighted		0.003814		0.003845		0.003996		0.004160
			Weighted		0.003611		0.003645		0.003855		0.004151
	Methylparaben	0.00630	Unweighted		0.005941		0.006008		0.006320		0.006595
			Weighted		0.005435		0.005523		0.006040		0.006640

that any attempt to compare models (empirical or otherwise), be based on data gathered over a wide range of mobile phases velocities and that the increment of velocity is small, such the number of mobile phase velocities is large (say ≥ 20). Data must be of low noise [17] and sufficient replicates made, so that when fitting by linear or non-linear least squares, account may be taken of any heteroscedasticity observed. Though equations containing three, four or more parameters may be fitted to the data, it is vital to report the size of the standard errors in those parameters in order to establish whether the parameters are significant to the fit. The AIC is another valuable statistic that should be included in any comparative study so that over-parameterised models may be rejected.

References

[1] W.-C. Lee, G.-J. Tsai, G.T. Tsao, Sep. Technol. 3 (1993) 178.

[2] J.J. Van Deemter, F.J. Zuiderweg, A. Klinkenberg, J. Chem. Eng. Sci. 5 (1956) 272.

[3] J.H. Knox, J. Chromatogr. A 960 (2002) 7.

[4] G. Gaspar, R. Annino, C. Vidal-Magar, G. Guiochon, Anal. Chem. 50 (1978) 1512.

[5] G.J. Kennedy, J.H. Knox, J. Chromatogr. Sci. 10 (1972) 549.

[6] V.G. Berezkin, V.I. Malyukova, D.S. Avoce, J. Chromatogr. A 872 (2000) 111.

[7] J. Neter, M.J. Kutner, C.J. Nachtsheim, W. Wasserman, Applied Linear Regression Models, Times Mirror Higher Education Group Inc., Chicago, 1996.

[8] GraphPad Software, San Diego, CA, USA.

[9] S. Weisberg, Applied Linear Regression, Wiley, New York, 1985.

[10] D.M. Bates, D.G. Watts, Nonlinear Regression Analysis and Its Applications, Wiley, New York, 1988.

[11] L. Kirkup, Data Analysis with Excel: An Introduction for Physical Scientists, Cambridge University Press, Cambridge, 2002.

[12] J.L. Devore, Probability and Statistics for Engineering and the Sciences, third ed., Brookes/Cole, CA, 1991.

[13] H. Akaike, IEEE Trans. Autom. Control 19 (1974) 716.

[14] J.C. Giddings, J. Chromatogr. 5 (1961) 61.

[15] C. Horvath, H.-J. Lin, J. Chromatogr. 149 (1978) 43.

[16] R.P. Scott, J. Chromatogr. 468 (1989) 99.

[17] E. Katz, K.L. Ogan, R.P.W. Scott, J. Chromatogr. 270 (1983) 51.