Journal of Chromatography, 485 (1989) 453–460
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 22 028

MULTI-DIMENSIONAL INTERPOLATION BY THE MOVING LEAST-SQUARES APPROACH FOR MODELLING OF CHROMATOGRAPHIC RETENTION DATA

MATTHIAS OTTO* and UWE HÖRCHNER
Department of Chemistry, Bergakademie Freiberg, Akademiestr. 6, 9200 Freiberg (G.D.R.)
and
WOLFHARD WEGSCHEIDER
Institute for Analytical Chemistry, Micro- and Radiochemistry, Graz University of Technology, Technikerstr. 4, A-8010 Graz (Austria)

SUMMARY

Shepard's method of metric interpolation (moving least-squares) was explored to model retention data as a function of several experimental variables. It is shown that the previously applied mechanistic retention model can be replaced by the empirical method of metric interpolation without loss of information with respect to modelling of the retention data and to evaluating optimum conditions for separating a seven-component mixture of amino acids and dipeptides. The moving least-squares method can be used even in case of irregularly spaced data.

INTRODUCTION

Unattended optimization of chromatographic separations is carried out nowadays by retention modelling of individual solutes followed by evaluating the most appropriate conditions computationally. Mechanistic, semi-empirical or empirical retention models are used to tackle a specific separation problem[1–3].

For solving practical chromatographic optimization problems, most benefit is gained from retention modelling that (i) is very adaptable to different retention mechanisms, (ii) enables several variables to be handled simultaneously and (iii) can treat irregularly spaced variable data. Therefore, a multi-dimensional, empirical modelling technique based on data interpolation will be the method of choice. Multi-dimensional interpolation is known to be based on polynomial or spline interpolation.

Although both methods have been used in chemometrics[4,5], there are several drawbacks with these methods in the present context. Polynomial interpolation is not adaptable to all shapes of hypersurfaces. Spline interpolation lacks practicable solutions for handling dependences with more than two variables and with irregularly spaced data. In a more recent paper[6], metric multi-dimensional interpolation based on the moving least-squares (MLS) approach[7,8] was succesfully applied to the optimization of high-performance liquid chromatographic (HPLC) separations. The MLS method is essentially based on an inverse distance equation. It can deal with as many

0021-9673/89/$03.50

experimental variables as necessary, and also there are no problems with the treatment of irregularly spaced variable data.

In order to judge the general performance of metric interpolation, in this paper retention modelling of diprotic species by the MLS method is compared with earlier results obtained with a semi-empirical model[1] based on distribution and protolysis equilibria.

THEORY

Let $F(P)$ be a function of the point $P=(x_1, x_2, \ldots x_m)$ defined for all P in the m-dimensional real space R^m. Then an interpolant based on Euclidian metric is defined by

$$U(P) = \sum_{i=1}^{n} F_i \prod_{j \neq 1} r^{\beta j}(P,P_j) \Big/ \sum_{i=1}^{n} r^{\beta j}(P,P_j) \tag{1}$$

where

$$r(P,P_j) = [(x_1 - x_{1j})^2 + (x_2 - x_{2j})^2 + \cdots + (x_m - x_{mj})^2]^{1/2} \tag{2}$$

$U(P)$ = interpolated functional value at point P;
F_i = value of F at the knot i;
$r(P,P_j)$ = Euclidian distance between P and P_j;
P = point of interpolation;
P_j = knot (here the measuring point);
β_j = weighting factor of the Euclidian distance;
x_j = coordinates of interpolation points;
x_{ij} = coordinates of knots;
n = number of knots;
m = number of variables.

The interpolation method is essentially a weighted averaging procedure that can be applied to one ($m=1$) or many independent variables likewise.

Some special properties of the method can be discussed:

(a) All interpolated values will be smaller than the maximum knot value used and greater than the minimum knot value.

(b) No interpolation will be feasible at the knots as the Euclidian distance will be 0. At these positions the actual value of the knot is to be used.

(c) The shape of the hypersurface can be tuned by the weighting factor β in eqn. 1. This will be demonstrated under Results and Discussion.

(d) For computing the interpolation points, as many knots as required can be used. The best results are obtained by using the knots closest to the interpolation point.

(e) Interpolation can be based on both regularly and irregularly spaced data (see below).

In some instances it was found with the interpolant of eqn. 1 that the curves undulate severely between the knots. This will be incompatible with the mechanistically expected functional dependence in many situations as here for chromatographic retention modelling. To reduce such undulations inherent in the interpolant of

eqn. 1, Shepard[8] proposed for the special case of all $\beta_j = 2$ a technique for interpolating to given first partial derivatives according to the following equation (bivariate case considered with $P(x,y)$):

$$U(P) = \sum_{i=1}^{n} [F_i + F_{xi}(x - x_i) + F_{yi}(y - y_i)]/r_i^2 \Big/ \sum_{i=1}^{n} 1/r_i^2 \quad (3)$$

with

$$F_i = F(P_i)$$

$$F_{xi} = \left.\frac{\partial F}{\partial x}\right|_{P=P_i}$$

$$F_{yi} = \left.\frac{\partial F}{\partial y}\right|_{P=P_i}$$

The partial derivatives can be estimated from local planar approximations of the general model

$$M(x,y) = a + bx + cy \quad (4)$$

where a, b and c are regression parameters.

RESULTS AND DISCUSSION

Mechanistic retention modelling

As described in an earlier paper[1], the separation of diprotic substances, such as amino acids and dipeptides, can be described by a mechanistic (perhaps more semi-empirical) model that relates relative retention of the solutes (dependent variable) expressed by the capacity factor, k', to the mobile phase variables pH, elution strength (methanol content of the aqueous mobile phase) and ionic strength as the independent variables. This model is based on protolysis and distribution equilibria constants and for the k' dependence on the most important variable, pH, is

$$k' = \frac{k_0 + k_1 \frac{[H^+]}{K_{a1}} + k_{-1} \frac{K_{a2}}{[H^+]}}{1 + \frac{[H^+]}{K_{a1}} + \frac{K_{a2}}{[H^+]}} \quad (5)$$

where k_0, k_1 and k_{-1} are the distribution coefficients for the species HS (partially undissociated), H_2S (protonated) and S (deprotonated), respectively, and K_{a1} and K_{a2} are the consecutive protolysis constants.

As both the distribution coefficients and the protolysis constants depend additionally on the content of the organic modifier and on the ionic strength of the mobile phase, a complicated model is necessary to fit the experimentally obtained

retention data. In the complete model, six linear regression parameters and seven non-linear parameter are to be estimated.

Examples of applying the mechanistic model according to eqn. 5 are given in Fig. 1 for the retention of anthranilic acid (1A) and of L-leucyl-L-tyrosine (1B) as a function of pH and the methanol content of the mobile phase. As the bivariate case is most illustrative, the third variable, ionic strength, was fixed at 0.1 *M* in this study.

Computation of these hypersurfaces is based on eighteen experimental values of k' measured at pH 2, 3, 4, 5, 6 and 7 and at methanol contents of 10, 20 and 30% (v/v) (Fig. 2).

Moving least-squares modelling

In order to apply the MLS approach to the data, first the different scales of the variables pH and elution strength have to be normalized somehow to make calculations of the Euclidian distances in the space of variables comparable. In the present instance, normalization of the experimental variables is carried out linearly by transforing them to the interval [0,1].

For interpolation with eqn. 1 (without considering the partial derivatives), the four knots closest to the interpolation point have been chosen (Euclidian distance measure). The influence of the weighting factor (*cf.*, eqn. 1) was studied for β-values of 0.5, 1, 2 and 3. Fig. 3 shows the results when using β-values of 0.5 (A) and 2 (B) for modelling the retention of anthranilic acid. As can be seen, the shape of the interpolating curves approaches the mechanistic model more closely (Fig. 1A) if the

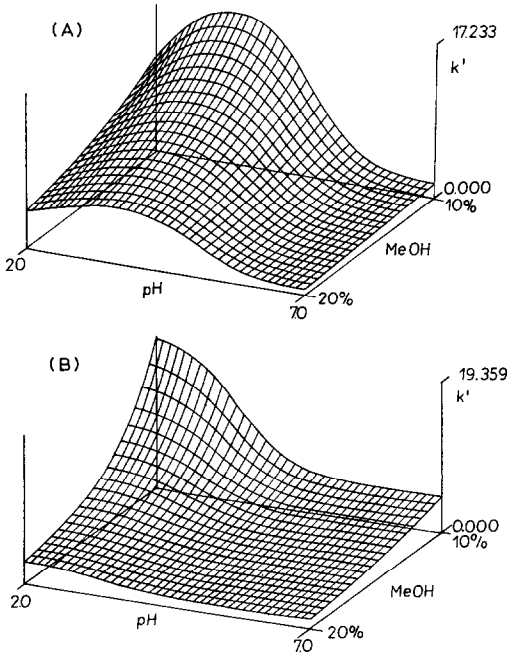

Fig. 1. Retention of (A) amino acids (anthranilic acid) and (B) dipeptides (L-leucyl-L-tyrosine) as a function of pH and methanol content of the mobile phase at an ionic strength of 0.1 *M*. MeOH = methanol.

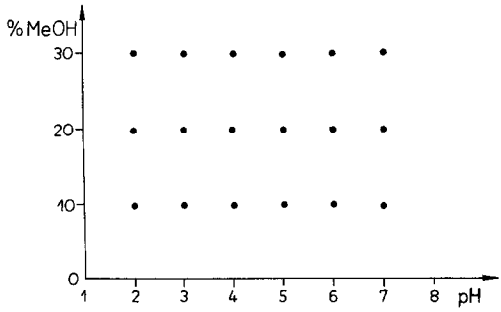

Fig. 2. Experimental design for modelling the retention behaviour as a function of pH and methanol content.

β-values increase. At still higher β-values (>2), the interpolating function behaves like a step function.

A model that compares better with the mechanistic model can be obtained by using eqn. 3 considering the partial derivatives in the interpolant. The plot obtained this way is shown in Fig. 4A for anthranilic acid and as a second example also for the dipeptide L-leucyl-L-tyrosine (Fig. 4B). The similarity of the hypersurfaces modelled with the complicated mechanistic model (Fig. 1) and the MLS method (Fig. 4) is satisfactory.

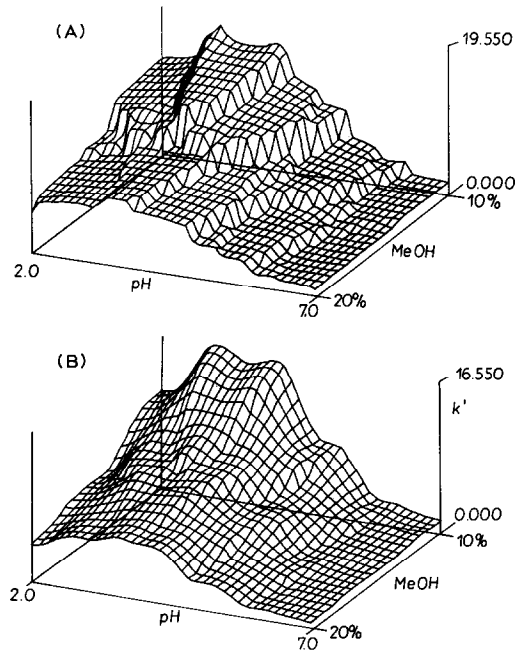

Fig. 3. Modelling of the retention behaviour of anthranilic acid with the MLS method (eqn. 1) at weighting factors β of (A) 0.5 and (B) 2.

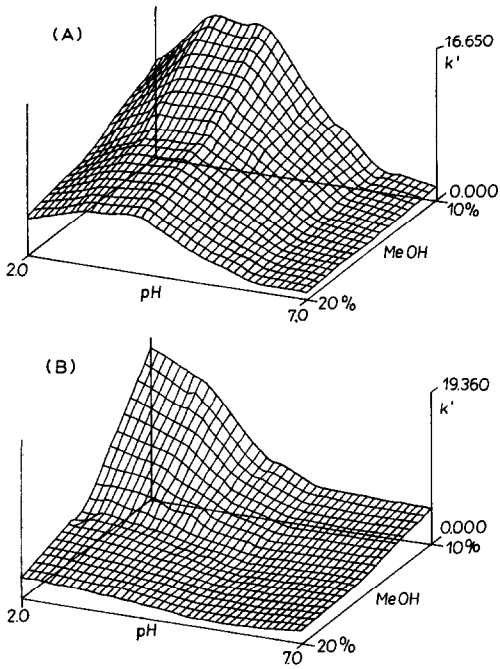

Fig. 4. MLS modelling of (A) anthranilic acid and (B) L-leucyl-L-tyrosine retention by considering the partial derivatives according to eqn. 3.

Irregularly spaced data

For demonstrating the situation where irregularly spaced data occur, the anthranilic acid data were recalculated by omitting the knot at pH 4.0 and 20% methanol. In this instance interpolation is performed in the same way as before, *i.e.*, the four closest knots are selected for applying eqn. 3.

Fig. 5 represents the hypersurface obtained from this case. Although the influence of the missing point is clearly evident, the general shape of the surface has been retained.

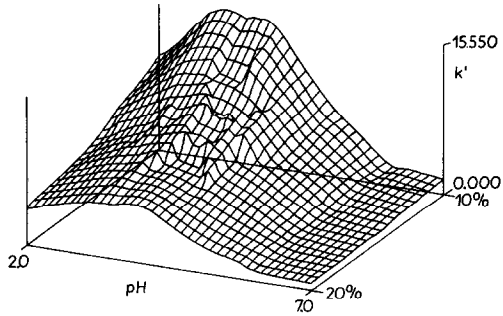

Fig. 5. MLS model of anthranilic acid by omitting experimental point at pH 4 and 20% methanol.

TABLE I

OPTIMUM CONDITIONS FOR SEPARATION OF A SEVEN-COMPONENT MIXTURE EVALUATED BY DIFFERENT MODELLING TECHNIQUES AND THE MINIMUM RELATIVE RETENTION AS THE OBJECTIVE CRITERION ($I=0.1$)

Technique	*pH*	*Methanol (%)*	*Optimum relative retention of the worst separated pair*
MLS (without partial derivatives, eqn. 1)	4.82	11.05	1.46
MLS (with partial derivatives, eqn. 3)	4.95	10.00	1.45
Mechanistic model	5.08	11.05	1.37

Evaluation of optimum separation conditions

As a means of further quantifying the performance of the MLS method in comparison with the mechanistic model, the optimum chromatographic conditions for separating a seven-component mixture[1] were evaluated with respect to the pH and methanol content at a constant ionic strength of 0.1 *M*.

From the retention values, k', of all seven components measured at the experimental points given in Fig. 2, the relative retentions are computed for all pairs of solutes:

$$\alpha = k'_i / k'_j \qquad \text{with } i \neq j$$

and the minimum relative retentions, α_{min}, are then found.

As in the present instance several local optima exist[1], the gobal optimum is found as the maximum value of all the minimum relative retention values. The gobal optima evaluated with different modelling techniques are given in Table I together with their maximum relative retention values. The agreement between the mechanistic model and the MLS approach with or without using partial derivatives is within 0.26 pH unit and 1.05% methanol content. This indicates an excellent agreement, which is also reflected by a difference plot of the minimum α values between the mechanistic model and the MLS modelling with consideration of partial derivatives (Fig. 6).

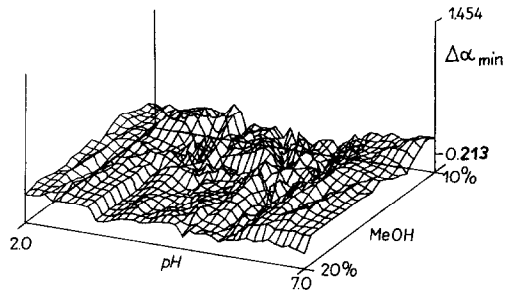

Fig. 6. Difference plot of the minimum relative retention values, α_{min}, of the mechanistic and the MLS approaches.

CONCLUSION

The method of moving least-squares can be used as a very general and simple alternative method to mechanistic modeling of the retention behaviour of solutes in dependence on chromatographic variables. One advantage of MLS is the empirical nature of this modeling technique, *i.e.* any retention mechanism can be described with the same algorithm. The second advantage derives from the fact that regularly as well as irregularly spaced data can be handled enabling simultaneous and sequential optimization strategies to be followed likewise. Applications of the MLS method within the frame of optimization software can be found in ref. 6.

REFERENCES

1 M. Otto and W. Wegscheider, *J. Chromatogr.*, 258 (1983) 11.
2 P. J. Schoenmakers, *Optimization of Chromatographic Selectivity, a Guide to Method Development (Journal Chromatography Library*, Vol. 35), Elsevier, Amsterdam, 1984.
3 J. C. Berridge, *Techniques for the Automated Optimization of HPLC Separations*, Wiley, Chichester, 1985.
4 W. Wegscheider, in D. A. Kurz (Editor), *Cubic Spline Functions for Solving Calibration Problems (ACS Symposium Series*, Vol. 284), American Chemical Society, Washington, DC, 1985, p. 167.
5 M. Otto, P. M. May, K. Murray and J. D. R. Thomas, *Anal. Chem.*, 57 (1985) 1511.
6 E. P. Lankmayr, W. Wetscheider and K. W. Budna, *J. Liq. Chromatogr.*, 12 (1989) 35.
7 W. J. Gordon and J. A. Wixon, *Math. Comput.*, 32 (1978) 253.
8 D. Shepard, in *Proceedings of 1968 ACM National Conference*, 1969, pp. 517–524.