CHROMSYMP. 965

EFFECT OF MODEL INACCURACY ON SELECTIVITY OPTIMIZATION PROCEDURES IN REVERSED-PHASE LIQUID CHROMATOGRAPHY

PETER J. SCHOENMAKERS*

Philips Research Laboratories, P.O. Box 80 000, 5600 JA Eindhoven (The Netherlands)

and

THOMAS BLAFFERT

Philips Research Laboratories, Vogt-Koelln Strasse 30, D-2000 Hamburg 54 (F.R.G.)

SUMMARY

Interpretive methods for the optimization of the selectivity in chromatography require the description of the retention surfaces of all individual solutes in a mixture by some kind of model. The requirements and availability of such models are discussed for the particular case of optimizing ternary and quaternary mobile phase compositions in reversed-phase liquid chromatography.

It is concluded that in order to allow an adequate prediction of the optimal conditions, a model equation should describe the experimental data to within 1% or less (in terms of the capacity factor k). Several suggested models from the literature were tried, but it appears that none of the currently available models provides a description of the data with the required accuracy. It is demonstrated that this situation does not improve when more experimental data become available.

The alternative of using piecewise (linear) interpolation is discussed, and it is demonstrated that this approach may provide a sufficiently accurate description of the data on the basis of a limited number of carefully selected experiments.

INTRODUCTION

Methods for the optimization of chromatographic selectivity can roughly be divided into simultaneous, sequential and interpretive procedures[1]. In simultaneous (or "grid search") procedures, a (large) number of chromatograms are recorded according to a pre-planned experimental design ("grid") and the optimum is identified as the best chromatogram obtained. In sequential procedures (*e.g.*, Simplex optimization), the optimum is approached in a stepwise manner. A minimum number of chromatograms is recorded and the resulting data are used to establish conditions for a subsequent experiment that is expected to be closer to the optimum.

Both categories of optimization procedures suffer from the large number of experiments that are required to locate the optimum (see Table I). Typically, a grid search optimization involving two parameters may require about 100 experiments, and a typical number for Simplex optimization is 25–40[2]. Such a large number of

experiments is unattractive for the optimization of the selectivity in reversed-phase liquid chromatography (RPLC) for two reasons: (1) the time required for equilibrating the system and recording a chromatogram is typically about 30 min; and (2) some of the most relevant parameters (*e.g.*, the mobile phase pH) are difficult to control automatically in small steps over a wide range.

The reason why so many experiments are required for the optimization of the selectivity in chromatography is the complexity of the response surface, which describes the variation of the quality of the chromatogram ("optimization criterion") as a function of the parameters considered. A typical response surface may contain a number of maxima (local optima). Only one of these represents the true (global) optimum. For several reasons[1], the global optimum should be the aim of optimization procedures. Because of the complexity of the response surface, the application of a simultaneous optimization method requires a fine grid and hence a large number of experiments to locate the global optimum. The result of a sequential optimization procedure is likely to be one of the local optima.

In order to locate the global optimum with a small number of experiments (typically between 5 and 15), a number of procedures have been developed specifically for the optimization of chromatographic selectivity. These methods can be classified as interpretive methods. By definition, interpretive methods are those which:

(1) interpret a chromatogram in terms of the retention times of the individual solutes;

(2) describe the retention surface of each individual solute with some kind of model;

(3) use this model for the retention surfaces and a suitable optimization criterion to calculate the response surface; and

(4) locate the optimum on the response surface.

The first of these steps greatly increases the complexity of the optimization procedure, as is indicated in Table I. Not only are the retention times of the peaks in each chromatogram required, but also the order in which the peaks appear is relevant for establishing the retention surfaces of the individual solutes. The peaks obtained for a particular solute need to be assigned the same label or number in each chromatogram, which implies that peaks need to be recognized (but not identified) in each chromatogram. This problem has recently received much attention, especially in connection with RPLC[3], and will not be discussed further in this paper.

TABLE I

SUMMARY OF METHODS FOR OPTIMIZING CHROMATOGRAPHIC SELECTIVITY

Characteristic	Procedure		
	Simultaneous	*Sequential*	*Interpretive*
Required number of experiments	Very large	Large	Small
Optimum found	Global	Local	Global
Complexity of method	Low	Moderate	High
Accuracy of optimum	Low	High	Variable*

  * The accuracy of the predicted optimum is the subject of this paper. A high accuracy may be obtained with an iterative procedure (see below).

The second step in the above series is the most critical. Interpretive methods owe their existence to the observation that retention surfaces of individual solutes are much simpler than the response surface of the entire mixture and may therefore be described adequately by some kind of model. This model may be a (set of) mathematical equations or (for one-parameter optimization problems) a graphical relationship. Alternatively, the retention surface may be approximated by linear interpolation between the experimental data points.

If the model provides a perfect description of the true retention surface, then the calculation of the response surface (step 3 above) and the location of the true (global) optimum (step 4) are almost trivial. Unfortunately, such perfect models do not exist and therefore the reliability and the accuracy of the optimum that results from an interpretive optimization method is determined by the accuracy with which the true retention surfaces are described by the model.

One way to improve the accuracy of the predicted optimum is to use iterative methods. These are interpretive methods, in which the predicted optimum is verified by a new experiment (at or around the location of the predicted optimum) and the resulting data are used to improve the model and to calculate a new optimum.

In this work we investigated the description of the retention surface in RPLC using ternary and quaternary mobile phase mixtures. Several mathematical models were used in attempts to describe the retention surfaces within experimental error. The practical consequences of deviations between the models and the true retention surfaces are demonstrated using simulated chromatograms.

TERNARY MOBILE PHASE MIXTURES

Experimental data for ternary mobile phases were taken from the compilation in ref. 4. Practical optimization processes should focus on mobile phase mixtures that yield roughly optimal capacity factors (*e.g.*, $1 < k < 10$). Hence, interest is limited to solvents within a small range of eluotropic strengths (see Table II), *i.e.*, to a narrow band in the triangle representing all possible ternary mixtures. Fig. 1 illustrates the experimental locations that were considered for mixtures consisting of water, methanol and tetrahydrofuran (THF). The solvent compositions are given in Table II,

TABLE II

COMPOSITION AND ELUOTROPIC STRENGTH OF MOBILE PHASE MIXTURES CONSIDERED FOR THE SYSTEM WATER–METHANOL–THF

Data point No.	Water (%)	Methanol (%)	THF (%)	Eluotropic strength*
1	40	60	0	60
2	30	70	0	70
3	50	25	25	65
4	40	30	30	78
5	60	0	40	65
6	50	0	50	81
7	35	65	0	65
8	40	45	15	69

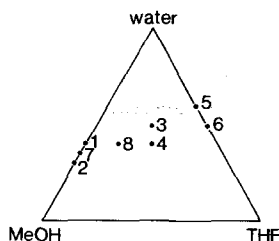* Expressed as the corresponding percentage of methanol.

Fig. 1. Experimental data points considered for the system water–methanol–THF. Experimental data were taken from ref. 4. See also Table II. MeOH = methanol.

together with the eluotropic strength expressed as the corresponding percentage of methanol (see refs. 5 and 6). In Table II 1% of methanol was assumed to correspond to 0.62% of THF[6].

The experimental data for the logarithm of the capacity factor can be described by a quadratic equation[4]:

$$\ln k = A_m \varphi_m^2 + A_t \varphi_t^2 + B_m \varphi_m + B_t \varphi_t + C + D \varphi_m \varphi_t \tag{1}$$

where φ_m is the volume fraction of methanol in the mobile phase and φ_t is the volume fraction of THF. A_m, A_t, B_m, B_t, C and D are the six coefficients of the quadratic expression, the values of which will be dependent on the solute. If six experimental capacity factors are available, then the coefficients can be determined and it is in principle possible to calculate the capacity factors at other compositions. For example, the first six data points in Table II may be used to calculate the coefficients of eqn. 1. This could be done for 30 solutes using the data in ref. 4. Using the coefficients thus obtained, the retention data at points 7 and 8 were calculated and compared with the experimental data. For point 7, which is located between and close to data points 1 and 2 (see Fig. 1), the average deviation between the calculated and experimental values for ln k was 0.031. This corresponds to an average error in the predicted capacity factors of about 3%. However, for point 8, which is not so close to the other experimental data points, the average deviation was 0.193, which corresponds to an average error in k between the model equation and the experimental data of about 20%.

The effect of the large deviations between the model equation and the experimental data at point 8 is illustrated in Fig. 2, which shows the (simulated) experimental chromatogram for five selected solutes (thin line) and the predicted chromatogram from the model (thick line). The simulated chromatogram is reconstructed by a computer program, using the experimentally observed capacity factors. Obviously, the two chromatograms are entirely different.

Fig. 3a shows the response surface for the separation of the five solutes in Fig. 2, calculated using the retention surfaces for the individual solutes, which were calculated from the retention data at the first six data points. The response surface shows the variation of a selected optimization criterion, representing the quality of the separation, as a function of the parameters in the parameter space. The criterion
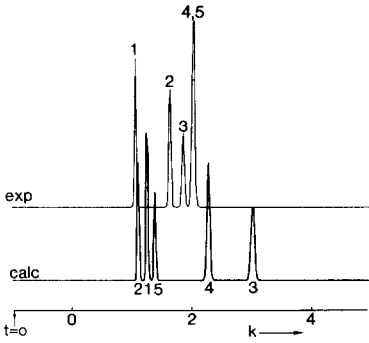
Fig. 2. Calculated (thick line) chromatogram at point 8 using the quadratic model of eqn. 1 fitted through the first six data points. The thin line shows a (simulated) experimental chromatogram. Solutes: 1 = anisole; 2 = *o*-cresol; 3 = N-methylaniline; 4 = *p*-nitroacetophenone; 5 = 3-phenylpropanol. Mobile phase: water–methanol–THF (40:45:15).
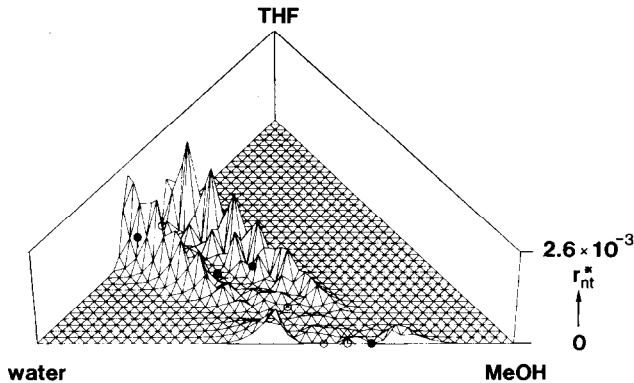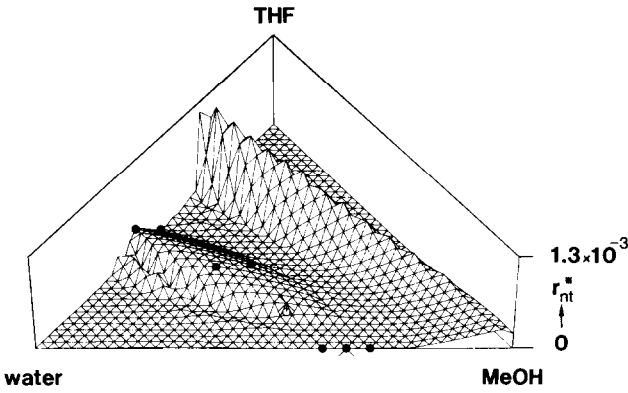

Fig. 3. (a) Response surface for the five solutes in Fig. 2 in the triangle shown in Fig. 1, calculated from the retention surfaces of the individual solutes using the quadratic model of eqn. 1 fitted trough the first six data points. Deviations between the experimental response and the calculated response are indicated for points 7 and 8. Crosses indicate data points not included in establishing the model. (b) As (a), but now with the quadratic model fitted through seven data points (1–6 and 8). Error bars may be observed for data point 7 (not included) and several other data points.

used in Fig. 3 is the time-corrected resolution product[1]. This criterion is based on the calibrated normalized resolution product (r^*) of Drouen et al.[7]:

$$r^* = \prod_{i=0}^{n} (R_s/\bar{R}_s) \tag{2}$$

where R_s is the resolution between two successive peaks in the chromatogram and $\bar{R}_s$ the average resolution for all pairs of successive peaks. Both the product and the average are taken from $i = 0$ (the zeroth peak being a hypothetical peak at $t = t_0$) to $i = n$ (where n is the number of peaks in the chromatogram). The time-corrected resolution product is

$$r^*_{nt} = \exp\left(\frac{\ln r^*}{n}\right)\Big/ t_{ne} \tag{3}$$

where t_{ne} is a measure of the required analysis time and n is again the number of peaks.

   In Fig. 3a, the actual response calculated from the experimental capacity factors is indicated at the locations of the six points in Fig. 1. The bold dots are located above the response surface (i.e., the predicted response is lower than the experimental response). Open circles represent response values below the predicted surface. A cross indicates the points that were not included in fitting the model. The difference between the actual (experimental) response at points 7 and 8 and the predicted response is indicated by vertical error bars. It is seen that a smooth response surface is predicted in the experimental range, but that extrapolation outside this area suggests that the location of the global optimum is at much higher concentrations of organic modifier.

   It can be concluded from Figs. 2 and 3a that an optimization procedure in which (1) experimental data are obtained at the locations of the first six points in Table II (or Fig. 1) and (2) the retention surfaces are fitted by the quadratic model of eqn. 1 is not sufficient accurate to predict the location of the optimum. This conclusion is not only valid for the example involving five solutes, but may be extended to the entire set of 30 solutes of ref. 4 (see Table III). We also fitted the model equations to several other selections of six data points from Fig. 1, but the description of the two remaining data points was always much worse than in the situation described above. Therefore, in order to improve this situation, either the number of experiments needs to be increased or a better model should be used to describe the retention surfaces.

REGRESSION ANALYSIS

   As point 8 appears to be badly described by the model if only six data points are available, it is logical to "obtain" additional experimental data at this location. The average deviation between the quadratic model and the experimental data using seven data points (1–6 and 8) to establish the six coefficients in eqn. 1 is illustrated in Table III (last column).

Table III shows that the large deviation for point 8 is reduced by a factor of about 3. However, the errors are now distributed more evenly over the entire parameter space and the first six data points, which were by definition described exactly by the model through six points, now show average deviations that are typically a few percent, but up to about 8% (in terms of k) for data point 3. Fig. 4 shows examples of (simulated) experimental and calculated chromatograms for the same solutes as in Fig. 2 at the compositions of data points 8, 3, 2 and 4 (Fig. 4a–d, respectively).

Fig. 4a shows that the prediction of the chromatogram at point 8, although better than in Fig. 2, is still unacceptable for optimization purposes. At point 3, where the average deviation between the model and the experimental data is about 8%, we see that the elution order is predicted correctly by the model and that the baseline separation predicted by the model can indeed be obtained for all solutes in practice (see Fig. 4b). Nevertheless, there are considerable differences in retention times (k values) between the predicted and the experimental chromatograms. In fact, Fig. 4b represents a fortuitous example, as is demonstrated in Fig. 4c. This figure, showing the calculated and experimental chromatograms at point 2, yields only minor deviations between the predicted and experimental capacity factors (the average deviation is 2.9% at point 2). However, the model predicts solutes 1 and 5 to be reasonably separated, whereas solutes 2 and 3 are expected to co-elute. In practice, solutes 1 and 5 overlap completely and solutes 2 and 3 are separated to the baseline. Finally, Fig. 4d shows that at point 4, where the average deviation is 0.3% in terms of k, the model yields a perfect prediction of the experimental chromatogram.

It appears from Fig. 4 that in order to use model equations for the retention surfaces for optimization purposes, deviations between the model and the experimental data of a few percent cannot be accepted (Fig. 4c). The accuracy of the model needs to be (much) greater than 1% to yield reliable predictions of the optimal chromatographic conditions (Fig. 4d). This conclusion should be seen in the perspective of possible experimental errors in the determination of capacity factors. A reasonable

TABLE III

AVERAGE DEVIATIONS (× 100) BETWEEN THE CALCULATED AND EXPERIMENTAL VALUES FOR ln k USING THE QUADRATIC MODEL OF EQN. 1 TO DESCRIBE THE RETENTION SURFACES

Thirty solutes for which data are available in ref. 4 were included in the calculations. Data points refer to Table II.

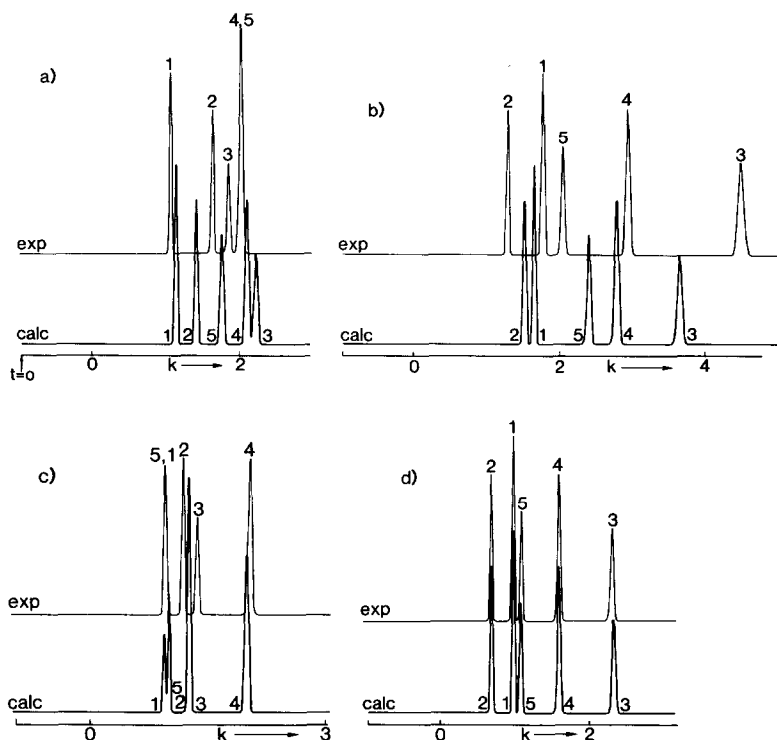| Data point No.	No. of data points included	
	6	7
1	—	1.3
2	—	2.9
3	—	8.2
4	—	0.3
5	—	2.9
6	—	0.6
7	3.1	7.3
8	19.3	7.2

Fig. 4. Calculated (thick line) chromatogram at points 8, 3, 2 and 4 using the quadratic model of eqn. 1 fitted through seven data points (1–6 and 8). The thin lines show (simulated) experimental chromatograms. Solutes as in Fig. 2. Mobile phases: (a) water–methanol–THF (40:45:15) (point 8; average deviation for all solutes 7.2%); (b) water–methanol–THF (50:25:25) (point 3; average deviation 8.2%); (c) water–meth-anol (30:70) (point 2; average deviation 2.9%); (d) water–methanol–THF (40:30:30) (point 4; average deviation 0.3%).

estimate for the experimental error in terms of k is 1% (see also ref. 4). A more accurate determination of k would require considerable extra precautions. Therefore, we should aim to describe the data within experimental error.

Fig. 3b shows the response surface obtained for the five solutes in Fig. 2, now calculated using the retention surfaces fitted through seven data points (data point 7 was not included, as is indicated by a cross in the figure). Error bars may also be observed for other points in this figure, but the most striking feature is the vastly different overall appearance of the response surface. This underlines the sensitivity of interpretive optimization procedures to slight alterations in the model fitted through the retention surfaces. Moreover, the vast differences between Fig. 3a and b are indicative of the fact that whereas an accurate description of the response surface within the experimental parameter space is difficult, extrapolations outside the experimental range are tentative at best.

In order to investigate the effect of additonal data points on the accuracy of the description of the retention surfaces by a quadratic model, we referred to the ternary system water–methanol–acetonitrile. The locations of the available experi-
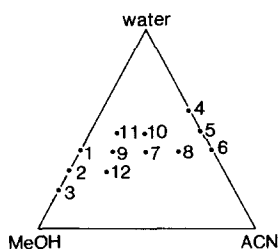
Fig. 5. Experimental data points considered for the system water–methanol–acetonitrile. Experimental data were taken from ref. 4. See also Table IV. ACN = acetonitrile.

mental data in a limited part of the triangle are illustrated in Fig. 5 and Table IV. In Table IV, 1% of methanol was assumed to correspond to 0.78% of acetonitrile[6].

Table V shows the average deviations observed between the quadratic model of eqn. 1 and the experimental data at the 12 data points shown in Fig. 5 and Table IV. The number of data points included in the regression analysis is increased progressively from 7 to 12 in the table. Again, the average deviations could not be reduced by selecting another set of seven data points. It is shown that the average deviations are generally larger for data points that are not included in the regression analysis (typically 10–30%) than for data points that are included (typically 10% or less). The observation in Table III that the errors are distributed more evenly over the entire parameter space when more experimental data points are added to the model is underlined in Table V. It is obvious from Table V that the addition of more data points does not lead to a description of the data that is sufficiently accurate in the light of the conclusions obtained from Fig. 4.

TABLE IV

COMPOSITION AND ELUOTROPIC STRENGTH OF MOBILE PHASE MIXTURES CONSIDERED FOR THE SYSTEM WATER–METHANOL–ACETONITRILE

Data point No.	Water (%)	Methanol (%)	Acetonitrile (%)	Eluotropic strength*
1	40	60	0	60
2	30	70	0	70
3	20	80	0	80
4	60	0	40	51
5	50	0	50	64
6	40	0	60	77
7	40	30	30	68
8	40	15	45	73
9	40	45	15	64
10	50	25	25	57
11	50	37.5	12.5	54
12	30	52.5	17.5	75

* Expressed as the corresponding percentage of methanol.

TABLE V

OBSERVED AVERAGE DEVIATIONS ($\times 100$) IN $\ln k$ BETWEEN THE QUADRATIC MODEL (EQN. 1) AND EXPERIMENTAL DATA FROM REF. 4 FOR THE SYSTEM WATER–METHANOL–ACETONITRILE FOR 49 SOLUTES

The number of data points included in the regression analysis increases in the table.

$n^\star$	Data point No.											
	1	2	3	4	5	6	7	8	9	10	11	12
7	3.7	6.4	2.8	2.0	3.2	1.3	0.0	30.7	7.8	21.2	16.5	21.6
8	5.0	5.2	2.3	1.6	2.6	7.5	13.5	18.0	15.0	29.4	20.9	12.7
9	7.1	4.8	2.1	1.5	2.4	9.2	9.1	20.5	9.7	25.2	16.6	13.2
10	11.0	6.6	1.2	6.2	2.1	10.5	4.5	23.2	6.8	16.2	15.1	17.9
11	11.6	6.2	1.5	6.8	5.2	9.9	4.3	22.5	6.0	16.4	6.9	17.8
12	13.3	5.9	4.0	7.9	6.6	9.6	8.2	20.1	6.1	18.3	6.5	11.0

    $\star$ First n data points in the table were included in the regression analysis.

ALTERNATIVE MODELS

We have concluded above that the prediction of the response surface using the quadratic model cannot be improved sufficiently by adding more experimental data points. The logical conclusion then is that a model is required which describes the experimental retention data more accurately, preferably within experimental error. Several different mathematical expressions have been fitted to the data for the water–methanol–THF system in order to try and find a more accurate description. All of these models are based on some kind of physical picture for the chromatographic process, as described in the original literature. To allow a fair comparison with eqn. 1, the different alternative models were all used in the form of equations with six coeficients.

Reciprocal model
Plotting $1/k$ instead of $\ln k$ against composition in RPLC has been suggested by McCann *et al.*[8]. The resulting plots yield slightly curved lines for binary mixtures. Therefore, a quadratic model analogous to that described by eqn. 1 has been fitted to the data in terms of $1/k$:

$$1/k = A_m \, \varphi_m^2 + A_t \, \varphi_t^2 + B_m \, \varphi_m + B_t \, \varphi_t + C + D \, \varphi_m \, \varphi_t \qquad (4)$$

Extended linear model
The retention *vs.* composition lines in binary mixtures over limited ranges of composition may be accurately described by a straight line[9,10]. Eqn. (1) reduces to a quadratic curve if either φ_m or φ_t is zero (binary mixtures) or if the ratio φ_m/φ_t is held constant (pseudo-binary mixtures). An equation that yields a straight line for both binary and pseudo-binary mixtures is

$$\ln k = A \, R^2 + B \, R + C + (D \, R^2 + E \, R + F)\psi \qquad (5)$$

where ψ is the total volume fraction of solvents ($\psi = \varphi_m + \varphi_t$) and R is a ratio defined as

$$R = \varphi_t/(\varphi_m + \varphi_t) = \varphi_t/\psi \tag{6}$$

Logarithmic model

An extended version of the logarithmic model of Lu and Lu[11] for ternary mixtures may be formulated as

$$\ln k = A + B\,\varphi_m + C\,\varphi_t + D\ln(1 + E\,\varphi_m + F\,\varphi_t) \tag{7}$$

We were not able to fit this model, nor various similar equations involving a logarithmic function, to our data. It appears that over limited ranges of eluotropic strength the linear and logarithmic contributions to $\ln k$ are not sufficiently different for the numerical process to converge.

Piecewise linear interpolation

The retention surfaces can be approximated by piecewise linear interpolation. For example, the retention data at point 7 can be estimated from those at points 1 and 2, and the retention data at point 8 from points 1 and 4 (see Fig. 1). An important aspect of linear interpolation is that the description of the retention data at the experimental locations remains exact if additional experiments become available. We have seen above that this is not the case if a model equation (*e.g.*, the quadratic eqn. 1) is used, as was demonstrated in Table V, We shall return to this aspect later.

Table VI summarizes the results obtained with the different models. It is seen that the average deviations between the model and the experimental data do not vary a great deal between the different models, and that in general none of the above models provides a description of the retention surfaces to within the required accuracy of 1% or better.

TABLE VI

AVERAGE DEVIATIONS ($\times$ 100) BETWEEN THE CALCULATED AND EXPERIMENTAL VALUES FOR $\ln k$ USING VARIOUS MODELS TO DESCRIBE THE DATA

All solutes for which data are available in ref. 4 have been included in the calculations. Data points refer to Table II.

Model	Eqn.	n^*	Data point No.							
			1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Quadratic	1	6	—	—	—	—	—	—	3.1	19.3
Quadratic	1	7	1.3	2.9	8.2	0.3	2.9	0.6	7.3	7.2
Reciprocal	4	6	—	—	—	—	—	—	4.9	17.9
Reciprocal	4	7	2.3	2.2	10.4	0.2	3.7	0.3	4.9	17.9
Extended linear	5	7	5.0	0.0	0.0	9.9	1.6	3.3	2.8	13.3
Linear interpolation**	—	—	—	—	—	—	—	—	2.8	23.7

* Number of data points used to model the retention surfaces.
** Data at point 7 can be found from a linear interpolation between points 1 and 3. Point 8 can be found from points 1 and 4.

QUATERNARY MOBILE PHASE MIXTURES

A convenient way to create quaternary mobile phase mixtures for RPLC is to combine variable proportions of three isoeluotropic binary mixtures[12]. This has the effect that all resulting mixtures may also be expected to be isoelutropic, *i.e.*, give rise to roughly the same capacity factors. Such a strategy has been applied for optimization purposes by Glajch *et al.*[12]. The composition of isoeluotropic quaternary mixtures can again be represented by a triangle, in which the isoeluotropic binary mixtures are located at the corners and identified by bars above the abbreviations for the modifiers (*e.g.*, methanol). This design is illustrated in Fig. 6.
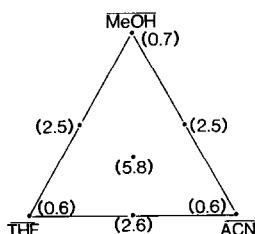

Fig. 6. Experimental design involving isoeluotropic binary (vertices), ternary (edges) and quaternary (centre) mobile phase mixtures for RPLC. The figures indicate the observed average deviation between the experimental data and the model of eqn. 8 in terms of ln k.

The data of Glajch *et al.*[12] may again be fitted to a quadratic expression, similar to eqn. 1:

$$\ln k = A_m \bar{\varphi}_m^2 + A_t \bar{\varphi}_t^2 + B_m \bar{\varphi}_m + B_t \bar{\varphi}_t + C + D \bar{\varphi}_m \bar{\varphi}_t \tag{8}$$

where $\bar{\varphi}_m$ and $\bar{\varphi}_t$ now represent the volume fractions of the isoeluotropic binary mixtures of methanol and water and THF and water, respectively. Data for nine substituted naphthalenes are available from ref. 12. The resulting average deviations are indicated in Fig. 6. The observed deviations are small for the binary mixtures on the vertices of the triangle, a few percent for the ternary mixtures along the sides and up to 6% for quaternary mixtures in the centre. A figure of 6% was also reported by D'Agostino *et al.*[13] for the description of retention data in (non-isoelutropic) quaternary mixtures. This implies that once again the accuracy of the description is insufficient for optimization purposes (see Fig. 4). However, the situation is slightly more favourable than is indicated by the data in Fig. 6. Fig. 7 shows the retention surfaces for two of the solutes. Clearly, these are smooth surfaces. The deviations between the model and the experimental data are indicated by vertical bars. For the two solutes shown in Fig. 7, and also for the other six solutes, the model predicts k values that are slightly too high for all three ternary mixtures and too low for the quaternary mixture. This implies that deviations from the model will be systematic and that small errors in the predicted absolute retention times (k values) will be levelled out to some extent when it comes to relative retentions (α values).

Fig. 8 shows the calculated chromatogram (predicted from the model) and the (simulated) experimental chromatogram for the quaternary mixture in the centre of
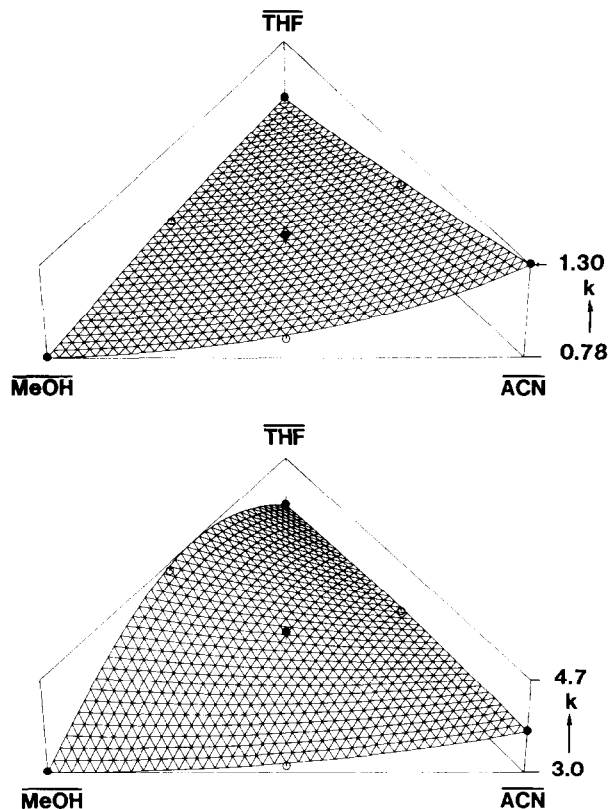
Fig. 7. Retention surfaces for 2-naphthyl methyl sulphone (solute 2) and 1-nitronaphthalene (solute 5) in the parameter space of Fig. 6 calculated using eqn. 8. Deviations between the model equation and the experimental data are indicated. Bold dots are located above the surface; open circles are located underneath.
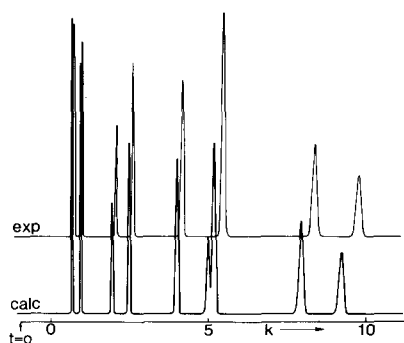

Fig. 8. Calculated (thick line) chromatogram at the quarternary mobile phase composition in the centre of Fig. 6 using the quadratic model of eqn. 8 fitted through all seven data points. The thin line shows a (simulated) experimental chromatogram. Solutes: 1 = 1-acetaminonaphthalene; 2 = 2-naphthyl methyl sulphone; 3 = 2-hydroxynaphthalene; 4 = 1-acetylnaphthalene; 5 = 1-nitronaphthalene; 6 = 2-methoxynaphthalene; 7 = naphthalene; 8 = 1-naphthyl methyl sulphide; 9 = 1-chloronaphthalene. Mobile phase: water–methanol–THF–acetonitrile (49:21:13:17).
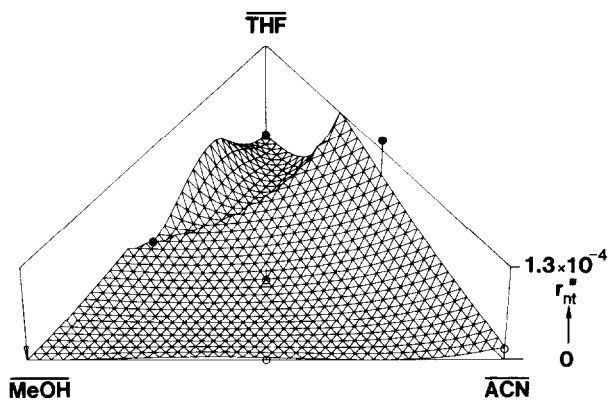
Fig. 9. Response surface calculated from the coefficients in Table VII for the solutes shown in Fig. 8. The deviations between the true (experimental) response and the calculated surface are indicated. Bold dots are located above the surface; open circles are located underneath.

the triangle. It can be seen that all components of the mixture elute slightly later in practice than predicted by the model. Again, these small differences in terms of k values may be significant in terms of resolution, as is evident in Fig. 8 for solutes 6 and 7.

Fig. 9 shows the response surface obtained for the nine solutes in Fig. 8. Considerable error bars are observed for the quaternary mixture and for the ternary mixture of THF, acetonitrile and water. The experimental response in the methanol–water binary mixture is located just below the response surface. It is obvious that the accuracy of the quadratic model is again insufficient for optimization purposes.

Table VII lists the coefficients obtained for the nine solutes in Fig. 8 using eqn. 8 to model the retention surfaces.

TABLE VII

COEFFICIENTS OF EQN. 8 DESCRIBING THE RETENTION SURFACES OF THE SOLUTES IN FIG. 8

Solute No.*	Coefficient						$2A_m + 2A_t + D$	$\Delta\varphi$** (eqn. 11a)
	A_m	A_t	B_m	B_t	C	D		
1	0.84	−0.55	−0.90	0.36	−0.38	1.24	1.82	0.22
2	0.39	0.03	−0.88	−0.29	0.24	0.56	1.40	0.25
3	0.37	−0.55	−0.47	1.15	0.29	1.51	1.15	0.28
4	0.34	0.04	−0.55	−0.16	1.02	0.79	1.93	0.22
5	0.22	−0.10	−0.44	0.12	1.33	1.39	1.63	0.23
6	0.22	−0.22	−0.34	0.23	1.51	1.20	1.20	0.27
7	0.37	−0.22	−0.52	0.31	1.55	1.34	1.64	0.23
8	0.13	−0.24	−0.17	0.21	1.93	1.26	1.04	0.29
9	0.14	−0.54	−0.16	0.38	2.06	1.16	0.36	0.50

* For solute identification see Fig. 8.
** Eqn. 11a.

PIECEWISE LINEAR INTERPOLATION

We have seen above that neither the addition of extra data points nor the use of alternative model equations will lead to a description of the retention surfaces that is sufficiently accurate to allow the application of interpretive optimization procedures with fixed experimental designs. We also mentioned above that when the retention surface is approximated by linear interpolation between data points, the effect of additional data becoming available will be (1) to yield an exact description of the retention surfaces (and hence an exact prediction of the response value) at each experimental location (assuming the absence of experimental error) and (2) to increase the accuracy of interpolation.

It can easily be shown[7] that the error caused by linear interpolation of data points along a retention surface that is more accurately described by a quadratic equation (e.g., eqn. 1) is largest in the middle between two data points and amounts to

$$d = 1/4(A_m \, \Delta\varphi_m^2 + A_t \, \Delta\varphi_t^2 + D \, \Delta\varphi_m \, \Delta\varphi_t) \tag{9}$$

where d is the difference between the value of ln k obtained by linear interpolation and the value obtained using a quadratic equation that is exact for the existing data points. In eqn. 9 $\Delta\varphi_m$ and $\Delta\varphi_t$ are the distances between the existing data points in terms of the two composition parameters φ_m and φ_t.

We can use an equation similar to eqn. 9 to investigate the possibility of using piecewise linear interpolation for the description of the retention surfaces for quaternary mobile phases in Table VII. For example, if we were to approximate the complete triangle with a series of smaller equilateral triangles, the equivalent expression for the deviation between a quadratic expression and a planar triangle through three data points becomes for the centre of the triangle

$$d = 1/9(2A_m \, \Delta\varphi_m^2 + 2A_t \, \Delta\varphi_t^2 + D \, \Delta\varphi_m \, \Delta\varphi_t) \tag{10}$$

where $\Delta\varphi_m$ and $\Delta\varphi_t$ are now the lengths of the sides of the triangle in terms of φ_m and φ_t, and with $\Delta\varphi_m = \Delta\varphi_t = \Delta\varphi$ we find

$$\Delta\varphi = 3\sqrt{d} \, (2A_m + 2A_t + D)^{-1/2} \tag{11}$$

If we tolerate a deviation of 0.01 in terms of ln k (about a 1% variation in k), then

$$\Delta\varphi = 0.3(2A_m + 2A_t + D)^{-1/2} \tag{11a}$$

In Table VII the values for $\Delta\varphi$ calculated from eqn. 11a are listed for the nine retention surfaces. It can be seen that for solute 9 the entire retention surface can be described to within 1% if the triangle is divided into four smaller ones, each half the original size. This may be achieved with only six data points (Fig. 10a). For the eight remaining solutes the required size of the triangle is about one quarter of the initial size. This would require 15 data points for the entire triangle.
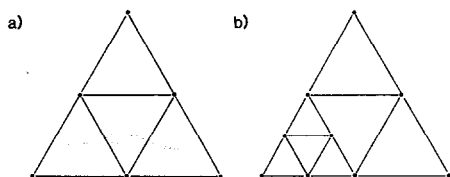
Fig. 10. (a) Possible six-point initial design for the optimization of quaternary mobile phase mixtures for RPLC. The parameter space (triangle) is divided into four parts. (b) As (a), but with three additional experiments to improve the accuracy of prediction.

A final remark concerning piecewise linear interpolation is that we have been discussing maximum errors, for instance the deviation between the model and the quadratic surface in the centre of a triangle. Anywhere else in the triangle the deviations will be smaller, so that the average deviation throughout the entire parameter space is much smaller.

ITERATIVE OPTIMIZATION PROCEDURES

Piecewise linear interpolation has been suggested by Schoenmakers et al.[14] and by Drouen and co-workers[7,15] as part of iterative optimization procedures. In such procedures, a minimum number of experimental data are collected. These data are used to make a first prediction of the location of the optimum. The next experiment is then performed at (or close to) the optimum. The additional data point is used to make a new prediction of the optimum and to establish the location of the next experiment, and so on. The main advantage of such a procedure is that the accuracy of the prediction is verified as part of the procedure and improved during each iteration cycle. Drouen et al.[15] described the application of an iterative procedure for optimization of a quaternary mobile phase composition in RPLC. We shall discuss below an example of a procedure that combines an initial fixed experimental design with an iterative approach to the optimum.

If piecewise linear interpolation is used for the optimization of quaternary mobile phase mixtures for RPLC, then an initial design of six data points may be used to divide the full triangle of Fig. 6 into four smaller ones (see Fig. 10a). According to eqn. 10, the maximum interpolation error is then (with $\Delta\varphi_m = \Delta\varphi_t = 0.5$)

$$d = 1/36(2A_m + 2A_t + D) \tag{10a}$$

Using the coefficients in Table VII, we expect the deviations between the calculated and experimental values of k to be of the order of 3–5% (in the centres of the four triangles). This is not sufficient to predict the optimum composition accurately (see Fig. 4). However, if it can be decided in which of the four triangles the optimum is located, then a further three experiments will allow this triangle in its turn to be divided into four smaller ones, and the accuracy of the prediction to be increased to within 1%. This is illustrated in Fig. 10b. The process may be repeated to increase further the accuracy of the predicted optimum.

As an alternative to piecewise linear interpolation, non-linear models may be used to describe parts of the parameter space with greater accuracy, as was suggested by Lankmayr and Wegscheider[16].

CONCLUSIONS

In order to make reliable predictions of optimum mobile phase compositions in interpretive optimization methods, the retention data for the individual solutes need to be known with an accuracy of greater than 1%.

Quadratic equations that describe the (logarithm of the) capacity factor as a function of composition do not provide the required accuracy. This situation cannot be improved by collecting more experimental data points, nor by using any of various alternative model equations.

A more accurate description of the retention surfaces may be obtained from a piecewise linear (or non-linear) interpolation between the available data points.

Piecewise interpolation in combination with iterative interpretive optimization procedures appears to be the most promising approach.

REFERENCES

1 P. J. Schoenmakers, *Optimization of Chromatographic Selectivity. A Guide to Method Development*, Elsevier, Amsterdam, 1986.
2 J. C. Berridge, *J. Chromatogr.*, 244 (1982) 1.
3 A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 57 (1985) 962.
4 P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 218 (1981) 261.
5 P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 205 (1981) 13.
6 P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *Chromatographia*, 15 (1982) 205.
7 A. C. J. H. Drouen, P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *Chromatographia*, 16 (1982) 48.
8 M. McCann, J. H. Purnell and C. A. Wellington, *Faraday Soc. Symp. Ser.*, 15 (1980) 82.
9 L. R. Snyder, J. W. Dolan and J. R. Gant, *J. Chromatogr.*, 165 (1979) 3.
10 P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 185 (1979) 179.
11 P. Lu and X. Lu, *J. Chromatogr.*, 292 (1984) 169.
12 J. L. Glajch, J. J. Kirkland, K. M. Squire and J. M. Minor, *J. Chromatogr.*, 199 (1980) 57.
13 G. D'Agostino, L. Castagnetta, F. Mitchell and M. J. O'Hare, *J. Chromatogr.*, 338 (1985) 1.
14 P. J. Schoenmakers, A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Chromatographia*, 15 (1982) 688.
15 A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 352 (1986) 127.
16 E. P. Lankmayr and W. Wegscheider, *9th International Symposium on Column LC, Edinburgh, July 1-5, 1985*, paper L3.2.