Review

Retention models for isocratic and gradient elution in reversed-phase liquid chromatography

P. Nikitas*, A. Pappa-Louisi

Laboratory of Physical Chemistry, Department of Chemistry, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

A R T I C L E   I N F O

Article history:
Available online 19 September 2008

Keywords:
Liquid chromatography
Retention models
Exo-thermodynamic relationships

A B S T R A C T

One- and multi-variable retention models proposed for isocratic and/or gradient elution in reversed-phase liquid chromatography are critically reviewed. The thermodynamic, exo-thermodynamic or empirical arguments adopted for their derivation are presented and discussed. Their connection to the retention mechanism is also indicated and the assumptions and approximations involved in their derivation are stressed. Special attention is devoted to the fitting performance of the various models and its impact on the final predicted error between experimental and calculated retention times. The possibility of using exo-thermodynamic retention models for prediction under gradient elution is considered from a practical point of view. Finally, the use of statistical weights in the fitting procedure of a retention model and its effect on the calculated elution times as well as the transferability of retention data among isocratic and gradient elution modes are also examined and discussed.

© 2008 Elsevier B.V. All rights reserved.

0021-9673/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2008.09.051
1. Introduction

Reversed-phase liquid chromatography (RPLC) is one of the most popular and versatile analytical techniques [1–7]. It may operate under various experimental conditions; from simple isocratic, isothermal and isochoric conditions to gradient elution conditions, where one (single mode) or more (multi-mode) experimental variables change simultaneously with time. Due to the importance and popularity of RPLC, there exists a vast amount of literature concerning the RPLC separation process at a macroscopic/thermodynamic or molecular level [8–20]. The principal aim of most research is to find optimal chromatographic parameters for a given separation problem and to develop new RPLC columns with improved separation characteristics through the understanding of the retention mechanism.

The first step for an optimization process is the prediction of the elution times, $t_R$, of the analytes under examination. Usually the elution time $t_R$ is determined by means of the retention factor, $k$, provided that $k$ is known as a function of the column temperature $T$, the $pH$ of the mobile phase and the volume fractions $\phi_1, \phi_2, \ldots$ of the solvents of this phase. When $k = g(T, pH, \phi_1, \phi_2, \ldots)$ is known, $t_R$ is calculated from

$$t_R = t_0(1 + k)$$  

in isocratic elution mode, and from the solution of the following integral equation [21–23]

$$\int_0^{t_R} \frac{F \, dt_c}{t_0(1 + k)} = 1$$  

in gradient elution mode. Here, $t_0$ is the column hold-up time and the time variable $t_c$ expresses the effective $\phi_1, \phi_2, \ldots$ vs. $t_c$ profile that the analytes feel inside the column. Note that Eq. (2) is not the conventional fundamental equation for gradient elution proposed first by Snyder [3–5]. The drawback of Eq. (2) is that it is not always easy to be solved and recursive numerical approaches have been proposed [21–23]. Alternatively, in gradient elution studies $k$ may be expressed as $k = h(\phi_1, \phi_2, \ldots, \phi_m)$, where $\phi_1, \phi_2, \ldots, \phi_m$ are variables representing the gradient elution conditions, that is the gradient steepness in linear gradients, the initial and the final composition of the mobile phase, etc. In this case $t_R$ is again calculated from Eq. (1).

It is seen that the expression of $k$ either as

$$k = g(T, pH, \phi_1, \phi_2, \ldots)$$  

or

$$k = h(\phi_1, \phi_2, \ldots, \phi_m)$$

is the very basic equation in optimization studies. This equation is called retention model since it may be directly or not the mathematical expression of the model that expresses the retention mechanism. Thus, the term retention model is used either for the model adopted to describe the retention mechanism or for Eqs. (3), which are, in fact, the mathematical expressions of this model. In this work our aim is not to critically review the efforts made to clarify the retention mechanism and its modeling, but the critical evaluation of the various simple models that have been used in isocratic and/or gradient studies for the prediction of the analytes elution times. The selection of the proper retention model is a crucial step in optimization procedures because the quality of such a procedure depends on the reliability and accuracy of the predicted retention times. This review concerns retention models for relatively small molecules. Retention models for macromolecules, ion-pair or micellar RPLC will not be considered in this paper.

2. Basic models for retention mechanisms

Retention in RPLC is a very complex process involving a great variety of interactions that are difficult to describe exactly. The most important of them are analyte–analyte interactions in both the mobile and stationary phases, analyte–solvent and solvent–solvent interactions in the mobile and possibly on the stationary phase, and interactions of the alkyl-chains of the stationary phase among themselves and with the analyte and solvent molecules. We should add that conventional RPLC silica-based columns have about 30–50% of their silanol groups accessible to analytes and solvent, even when coated and end-capped. These groups are weakly acidic and are expected to affect the retention mechanism. An excellent review on this topic is [19].

The complexity of these interactions results in a rather obscure picture about the retention mechanism. The solvophobic theory was one of the first attempts to describe chromatographic retention by means of classical thermodynamics. It is based on the work presented by Sinanoğlu in 1967 for biological binding [24]. This work was further developed by Horváth et al. to explain chromatographic retention [25–28]. The theory adopts a thermodynamic cycle that decomposes the chromatographic retention process into two processes and calculates the retention free energy. In particular, the standard free energy change of the retention process is expressed as the sum of the standard free energy change due to solvation and the standard free energy change for the association of an analyte with an alkyl chain. The solvation process for each species consists of three steps: creation of a cavity in the mobile phase to accommodate the analyte molecule, reduction in the free volume, and subsequent interaction of the analyte molecule placed in the cavity with the surrounding solvent molecules. All these contributions to the free energy change of the retention process are taken into account explicitly yielding rather complicated expressions for $ln k$.

The solvophobic theory has been strongly criticised by Dill [29], who developed the partition model, i.e. a retention model that considers the partitioning of the analyte molecules from the mobile phase into the stationary phase. In particular, this theory [8,30–35] assumes that an analyte molecule is fully embedded within the stationary phase alkyl chains and this process involves three main steps: the creation of a cavity within the stationary phase, the transfer of the analyte into this empty cavity from the mobile phase, and the closing of the cavity in the mobile phase, which was left behind by the analyte. Thus, the partitioning process is controlled by the differences in the molecular interactions of analyte molecules in both the stationary phase and the mobile phase.

An alternative approach proposed to explain the retention process is the adsorption model, according to which the analyte molecules are adsorbed at a surface solution formed on the tips of the hydrocarbon chains of the stationary phase [8,9,29,36–47]. The surface solution has the same constituents with those of the mobile
phase but with different concentrations. In the majority of the studies, the adsorption model assumes a displacement process; analyze molecules and molecules of the organic modifier co-adsorbed at the interface displace solvent molecules.

A combination of adsorption and partition models has been proposed by Jaroniec and Martire [9,48,49]. The retention process is modelled as a process of formation of an adsorbed layer via a displacement mechanism and the subsequent distribution of the analyte between the adsorbed layer and the stationary phase via a partition mechanism.

The adsorption and partition models and their combination have been extensively used in RPLC studies and they have been treated using many variations of classical or statistical thermodynamics or even combinations of them. In this respect, the following theories have been adopted: The regular solution theory [11,45–47,50–53], the Flory–Huggins mean field lattice theory for polymers in solution [11,43,54,55], the UNIFAC theory [56], the Hildebrand’s solubility parameters and the extended solubility parameter model [50–52,57–64], and statistical treatments that involve mainly lattice approaches based on Bragg–Williams approximation [29,38,54,55,65]. The self-consistent field theory for adsorption has also been employed [66].

To sum up, at present we believe that the various interactions in a RPLC column lead to two principal mechanisms for retention, partition and adsorption, although their combination cannot be ruled out. However, due to the complexity of the various interactions, it is still unclear whether retention is better described by partition or adsorption. Moreover, a number of critical issues are still open. For example, it is not clear whether the thermodynamic driving forces for retention lie primarily in the hydrophobic interactions with the aqueous mobile phase or in lipophilic interactions with the stationary phase. For the partition mechanism, it is questioned whether the thermodynamics of partition into the grafted chains of the stationary phase is the same or not with the partition at bulk oil–water system. In what concerns the adsorption mechanism, it is not clear whether it is a displacement process or not. The role of the various size effects, i.e. whether one solute molecule displaces one solvent molecule or a cluster of solvent molecules from the adsorbed layer, remains also unanswered. Finally, it is debated whether the retention mechanism remains the same with the variation of the mobile phase composition.

Molecular studies, like Monte Carlo simulations [67–74] and molecular dynamics [75–78], have been employed to shed light to the above problems. Although significant progress has been made in this direction, all thermodynamic problems are unlikely to be answered from molecular studies. For this reason and taking into account the complexity of the various interactions, the models for the retention mechanisms in RPLC may be divided into two categories [1]. Simple models that result in simple and applicable expressions for \( k \) but provide a poor and usually erroneous description of the retention process. Rigorous models that attempt to describe the reality of the retention process but with a very limited practical value. Here, we critically review the first category. The models of this category may be subdivided into thermodynamic (classical or statistical) models, exo-thermodynamic models and empirical models. In addition, we further subdivide the retention models considering the number of separation variables that affect the expression of the retention factor \( k \).

3. One variable retention models

The most commonly used separation variable is the volume fraction of the organic modifier in the mobile phase, \( \varphi \). Other variables also used in RPLC are the mobile phase \( \text{pH} \), the flow rate, \( F \), and the column temperature, \( T \). The flow rate does not affect \( k \) and for this reason the discussion below concerns the variables \( \varphi, \text{pH}, \) and \( T \).

3.1. Thermodynamic models

3.1.1. Principal thermodynamic processes

The principal thermodynamic processes of retention in a RPLC column, as explained previously, are the partition, the adsorption and their combination. Consider a mobile phase composed of water \( S \), the organic modifier \( B \) and the analyte \( A \). If the retention process is due to partition, it may be described by the simple equilibrium

\[
A^m \rightleftharpoons A^s
\]  

where superscripts \( m \) and \( s \) denote the mobile and the stationary phase, respectively.

For the adsorption mechanism the situation becomes more complicated since the adsorption process depends on whether it is a displacement process or not and whether size effects affect it or not. If it is a displacement process and size effects play a dominant role, the adsorption takes place through the following processes

\[
A^m + rS^m \rightleftharpoons A^s + rS^m
\]  

where superscript \( a \) denotes the adsorbed layer formed on the alkyl chains of the stationary phase and \( r \), \( n \) are the number of water molecules displaced by one analyte and one modifier molecule, respectively. However, process (5) is valid only if the adsorbed layer exhibits a constant width; it is a monolayer or a bilayer. If this prerequisite is not valid, process (5) may be replaced by

\[
A^m \rightleftharpoons A^s
\]  

In addition, there are rather strong indications that water molecules, and very likely the molecules of all polar solvents, form large clusters at the adsorbed layer. These clusters have as a consequence the minimization of size effects in the mechanism of adsorption. In particular, thermodynamic and adsorption studies have shown that when a solute molecule displaces a cluster of solvent molecules from the adsorbed layer, the size parameter \( n \) in Eq. (6) takes a value close to 1 [79–82]. Therefore, at least as a first approximation, we may completely ignore size effects by putting \( n = 1 \) in Eq. (6). Thus, this process may also be replaced by the simpler process

\[
B^m + S^a \rightleftharpoons B^s + S^m
\]  

If both partitioning and adsorption contribute to the retention mechanism, process (8) is still valid but (7) should be replaced by the two steps process

\[
A^m \rightleftharpoons A^a, \quad A^a \rightleftharpoons A^s
\]  

where superscript \( s \) means that the analyte molecules embed into the stationary phase.

3.1.2. Retention models for variations in mobile phase composition

Since the composition of the mobile phase and in particular the concentration of the organic solvent in it is the most frequently used factor for optimization in RPLC, several models have been proposed and tested for accurate predictions of the elution time at varying concentrations of organic solvent [45,46,51,57,83–93]. Here, we show that the most important of them may arise from the proper thermodynamic treatment of the retention process.
If both partition and adsorption contribute to the retention mechanism, then the capacity factor is given by

\[
k = \lim_{n_A^r \rightarrow 0} \frac{n_A^r + n_A^f}{n_A^r}
\]

(10)

where \(n_A^r\), \(n_A^f\), \(n_A^m\) are the number of moles of the analyte in the mobile phase, the adsorbed layer and within the alkyl chains of the stationary phase, respectively. However, if \(\psi_A^r\) is the volume fraction of \(A\) in the \(r\) phase (\(r = a, s, m\) or \(m\)), we may write \(\psi_A^r = V_A^r/V = n_A^r V_{m_A}^r/(V m_{m_A}^r)\), where \(V_A^r\) is the volume of \(A\) in phase \(r\), \(V_{m_A}^r\) the molar volume of \(A\) in the same phase and \(V^s\) is the total volume of phase \(r\). Therefore, \(k\) is rewritten as

\[
k = \lim_{\psi_A^r \rightarrow 0} \frac{k^c \psi_A^r + k^t \psi_A^t}{\psi_A^r}
\]

(11)

where \(k^c = V^s V_{m_A}^r/(V m_{m_A}^r)\) and \(k^t = V^s V_{m_A}^r/(V m_{m_A}^r)\). In order to find \(k^c\) and \(k^t\) to be constant, the molar volumes \(V_{m_A}^r\) are assumed to be independent of the composition of phase \(r\) \((=a, s, m\) or \(m\)). However, we do know that in aqueous solutions of organic solvents the molar volumes depend on the composition and this dependence is associated with the volume contraction of the solution. Therefore, the above assumption is in fact an approximation and all retention models derived below are subject to this approximation.

From the two-step process, i.e. Eq. (9), we obtain

\[
\frac{\psi_A^f}{\psi_A^m} = \beta^m a, \quad \frac{\psi_A^s}{\psi_A^m} = \beta^m
\]

(12)

where \(f_A^m, f_A^s, f_A^f\) are the activity coefficients of the analyte in the mobile phase, the adsorbed layer and within the alkyl chains of the stationary phase, respectively, and \(\beta^m\) are the equilibrium constants of the processes represented by Eq. (9). Substitution of Eq. (12) into Eq. (11) yields

\[
k = \left( k^c + k^t \beta^m f_A^m / f_A^f \right) \frac{\beta^m}{f_A^f} = \left( k^c + k^t \beta^m f_A^m / f_A^f \right) \frac{f_A^m}{f_A^f}
\]

(13)

which may be written as

\[
\ln k = \ln k^c + \ln \left( \frac{f_A^m}{f_A^f} \right) + \ln \frac{f_A^m}{f_A^f}
\]

(14)

where \(k^c = k^t \beta^m a + \psi k^t \beta^s a/k^t\).

It is seen that an analytical expression of \(\ln k\) in terms of the volume fraction of the modifier in the mobile phase, \(\varphi = \varphi^m a\), may be obtained from Eq. (14) if the activity coefficients \(f_A^m, f_A^s, f_A^f\) are expressed as a function of \(\varphi\). If the stationary phase consists of alkyl chains grafted onto silica or polymer surfaces and molecules of \(A\) within the alkyl chains at infinite dilution, then the activity coefficient \(f_A^f\) may be assumed constant and therefore we may write

\[
f_A^m = D_A^f
\]

(15)

where \(D_A^f\) is a constant depending on the stationary phase properties. In what concerns \(f_A^m, f_A^s\) their expression depends on the properties of the mobile phase and the adsorbed layer. In general, their calculation in terms of \(\varphi\) requires drastic approximations. Such an approximation is to assume that both the mobile phase and the adsorbed layer behave as regular ternary mixtures of solvents \(S\), organic modifier \(B\) and analyte \(A\) molecules but with different compositions. In this case and under the condition that \(\varphi_A^m \rightarrow 0, \varphi_A^f \rightarrow 0\) we obtain [45, 46, 94]

\[
\ln f_A^m = D_A^f + D_A^2 \varphi_A^b + C_{\psi}^d (\varphi_A^b)^2
\]

(16)

\[
\ln f_A^f = C_0^b (1 - \varphi_A^b)^2
\]

(17)

\[
\varphi_A^m = \varphi_A^m(\varphi^s a)^2\]

(18)

where \(D_A^f, D_A^2, C_{\psi}\) are constants. We observe that the final expression of \(\ln k\) in terms of \(\varphi\) requires the dependence of \(\psi_A^m a\) upon \(\varphi\). This dependence may arise from the adsorption equilibrium process (8). Since \(\varphi_A^m a \rightarrow 0, \varphi_A^f a \rightarrow 0\), we obtain

\[
\psi_A^m a (1 - \psi_A^m a) = \beta^m
\]

(19)

However, in order to obtain a simple dependence of \(\psi_A^m a\) upon \(\varphi = \varphi^m a\) we are obliged to proceed to a more drastic approximation. In particular, we assume an ideal behaviour for the modifier and the water molecules, i.e. \(f_A^m = f_A^s = f_A^f = 1\), which, according to Eqs. (16) and (18) means that

\[
C_0^b \approx C_0^m \approx 0
\]

(20)

Then Eq. (19) yields

\[
\varphi_A^m a = \frac{\beta^m \varphi}{1 + \beta^m - 
\]

(21)

where \(\beta = \beta^m - 1\). Substitution of Eqs. (15), (16), (20) and (21) into Eq. (14) results in

\[
\ln k = c_0 - \frac{c_1 \varphi}{1 + c_2 \varphi} + c_3 \varphi + \ln(c_4 \ e^{c_4 \varphi / (1 + c_4 \varphi)} + 1)
\]

(22)

where \(c_0 = \ln k^c + D_A^2, c_1 = (\beta + 1) D_A^2, c_2 = \beta, c_3 = D_A^2\) and \(c_4 = (k^e / k^t) e^{D_A^2 - D_A^1} \beta^m\).

In Eq. (22) parameter \(c_4\) is a measure of the retention mechanism. If adsorption predominates, \(\beta^m \rightarrow 0\), which yields \(c_4 \rightarrow 0\), whereas if retention is due to partition, we have \(c_4 \rightarrow \infty\) because \(\beta^m \rightarrow \infty\). Fig. 1 depicts the properties of Eq. (22). When \(c_4 = 0\), we obtain curved \(\ln k\) vs. \(\varphi\) plots, which are concave upwards provided that \(c_1 \leq 0\). Parameter \(c_2\) should be negative or at least equal to zero otherwise at great \(c_4\) values \(\ln k\) becomes an increasing function of \(\varphi\), which does not hold in aqueous mobile phases modified by organic solvents. The curvature of these plots is determined by the value of \(c_2\). The greater this parameter, the greater the curvature. As \(c_4\) increases the \(\ln k\) vs. \(\varphi\) plots are shifted towards higher \(\ln k\) values, their curvature decreases and eventually they become linear.

Eq. (22) could be an interesting retention model, which could give information on the retention process, because it combines the adsorption with the partition mechanism. However, it exhibits serious drawbacks. It is still a rather complicated five-parameter
equation, although it is based on radical approximations. Thus, it assumes first a negligible volume contraction during the mixing of water with the organic modifier and second that the activity coefficients of the organic modifier and the water are equal to one, which mean that both the mobile phase and the surface solution exhibit an ideal behaviour. However, aqueous solutions modified with methanol, acetonitrile and isopropanol, which are the usual organic modifiers in RPLC, exhibit significant positive deviations from Raoult’s law, i.e. they show strong deviations from the ideal behaviour [45]. Additionally, even as a retention model, Eq. (22) has its own limitations. An attempt to apply Eq. (22) to experimental data revealed serious convergence problems. In particular, we attempted to apply Eq. (22) to the retention data of [95]. That is, to the retention data of six non-polar analytes, benzene, toluene, ethylbenzene, propylenebenzene, isopropylbenzene and tert-butylbenzene obtained in aqueous mobile phases modified with methanol, isopropanol, acetonitrile and tetrahydrofuran using C18 and C2 reversed-phase columns. We used these data because in the C2 column the retention is plausibly due to adsorption, whereas the partition is expected to play some role in the C18 column.

In addition, these data cover the whole ϕ range from ϕ = 0 to 1. Unfortunately, in the majority of cases we faced serious convergence problems, which prevented us from taking any useful, even qualitative, information about the retention mechanism.

Despite the above drawbacks, Eq. (22) can be used to generate several useful retention models. Thus, as pointed out above, if adsorption predominates, c4 → 0 and Eq. (22) is reduced to [46,92]

\[ \ln k = c_0 - \frac{c_1 \psi}{1 + c_2 \psi} + c_3 \psi \]  

(23)

which is further simplified to [46,92]

\[ \ln k = c_0 - \frac{c_1 \psi}{1 + c_2 \psi} \]  

(24)

under the assumption that c3 = D^\alpha_2 ≈ 0. Eq. (24) was first proposed empirically by Neue et al. [96].

Eqs. (23) and (24) are four- and three-parameter equations, respectively. However, in these equations, as well as in Eq. (22), c2 = β + β' − 1 is, in fact, the equilibrium constant of the adsorption process of the modifier on the alkyl chains. This means that c2 = β is expected to have a constant value for each modifier. Therefore, Eqs. (23) and (24) may be written as

\[ \ln k = c_0 - \frac{c_1 \psi}{1 + \beta \psi} + c_2 \psi \]  

(25)

\[ \ln k = c_0 - \frac{c_1 \psi}{1 + \beta \psi} \]  

(26)

and they can be treated as three- and two-parameter equations, respectively, provided that the values of β for the various modifiers are known. Common β values have been determined in Ref. [92] and they are shown in Table 1.

The fitting performance of Eqs. (24)–(26) has been tested in Ref. [92] using 293 different experimental systems. It was found that when the two-parameter Eq. (26) is used with a common β value for each modifier, its applicability is rather poor. The results are improved if β takes a common value per group of structurally similar analytes for each modifier. The performance of the three-parameter Eq. (24) is in general very satisfactorily comparable to that of Eq. (25) when the latter uses β values from Table 1. A similar comparison of the performance of the most important retention models discussed in this paper is presented in Section 3.4.

If retention is ruled by partition, c4 → ∞, Eq. (22) yields the linear expression [1,51]

\[ \ln k = c_0 + c_1 \psi \]  

(27)

in which case c1 = D^\alpha_2. However, if we treat the partition process alone by putting \( \psi^a_A \) in Eq. (11) and make use of

\[ \beta^m = \frac{\psi_A^m f_m}{(\psi^m_A f_m^m)} \]  

(28)

where \( \beta^m \) is the equilibrium constant of process (4), then we readily obtain [46]

\[ \ln k = \ln(k^\alpha \beta^m) - D^\alpha_1 + \ln f_m^m \]  

(29)

Now \( f_m^m \) may be expressed as a power series of ψ since it is a continuous function of ψ. Therefore, Eq. (29) results in

\[ \ln k = c_0 + c_1 \psi + c_2 \psi^2 \ldots \]  

(30)

The linear Eq. (27) is valid only if we use very narrow ranges of ψ values. In contrast, the quadratic Eq. (30) exhibits a very satisfactory applicability and for this reason is the most popular retention model. However, it may lead to over-fitting errors or to physically meaningless \( \ln k \) vs. ψ plots, as discussed in Section 3.4.

Eqs. (23)–(26) express the adsorption mechanism under the approximations discussed above. However, in the majority of the treatments presented up to now the adsorption is considered as a displacement process, like process (5). If we adopt that the retention is due to this process, then the retention equation is described from

\[ \beta^m = \frac{\psi^p_A f_a}{(\psi^a_A f_a^p)} \]  

(29)

where we have adopted the approximation of Eqs. (20), which yield \( f_B^\alpha = f_B^p = f_B^m = f_B = 1 \). Now from Eq. (11) with \( k^\alpha = 0 \) we obtain

\[ \ln k = \ln k^\alpha + \ln \frac{\psi^a_A f_a}{\psi^a_A f_a^p} = (\ln(k^\alpha \beta^m) + \ln f_m^m = \ln r + \ln \frac{1 - \psi_B^p}{1 - \psi_B} \]  

(32)

which in combination with Eqs. (16) and (21) results in

\[ \ln k = c_0 - r \ln(1 + c_2 \psi) - \frac{c_1 \psi}{1 + c_2 \psi} + c_3 \psi \]  

(33)

in place of Eq. (23). The physical meaning of \( c_0, c_1, c_2 \) and \( c_3 \) is the same with that of Eq. (22). A similar to this equation but with \( c_1 = 0 \) and the factor +2 in place of = −r has been used in Ref. [97]. When \( c_1 = 0 \), Eq. (33) yields the logarithmic expression of \( \ln k \) [46]

\[ \ln k = c_0 - r \ln(1 + c_2 \psi) \]  

(34)

A similar expression for \( \ln k \) has been derived by Peichang and Xiaoming [38] using statistical thermodynamics but the derivation used has been questioned [46]. This equation is a four-parameter retention model if \( r \) is treated as adjustable variable, and it is reduced to the three-parameter equation [41,46]

\[ \ln k = c_0 - c_1 \ln(1 + c_2 \psi) \]  

(35)

under the assumption that \( c_3 \approx 0 \).

Finally, the reciprocal expression [46,98]

\[ \frac{1}{k} = c_0 + c_1 \psi + c_2 \psi^2 \]  

(36)

Table 1 - Common values of β at various mobile phases

<table>
<thead>
<tr>
<th>Mobile phase</th>
<th>Eq. (25)</th>
<th>Eq. (26)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methanol–water</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Acetonitrile–water</td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>Isopropanol–water</td>
<td>2</td>
<td>3.0</td>
</tr>
<tr>
<td>Tetrahydrofuran–water</td>
<td>2</td>
<td>1.0</td>
</tr>
</tbody>
</table>

From Ref. [92] with permission.
may also arise by means of thermodynamic arguments. In particular, in Ref. [46] we have shown that if we consider only analyte–modifier interactions in the mobile phase and treat these interactions as a pseudo-reaction between analyte and modifier molecules, the adsorption model results in Eq. (36).

To sum up, the thermodynamic approach may lead to the most important retention models used in optimization techniques. However, the necessity for the retention models to be as simple as possible is attained by the use of radical approximations. Thus, all expressions of ln\(k\) in terms of \(\varphi\) are based on the assumption that the molar volumes of the analyte, the modifier and the solvent are constant and independent of the composition of the mobile phase or the adsorbed layer. This assumption means that the contraction of volume observed during the mixing of water with any organic modifier is ignored. Another radical approximation involved in the derivation of Eqs. (22)–(26), (33)–(35) is the assumption that the mixture of the modifier and the solvent (usually water) in the mobile phase exhibits an ideal behaviour, which means that activity effects can be ignored. This is contrary to the fact that aqueous mobile phases modified with methanol, acetonitrile and isopropanol exhibit significant positive deviations from Raoult’s law [45]. Due to these approximations it is dangerous to acquire molecular information from the application of a simple model. Their usefulness is just to express as accurately as possible \(\ln k\) in terms of \(\varphi\), since only this information is needed in optimization procedures.

### 3.1.3. Retention models for variations in mobile phase pH

The mobile phase pH affects the retention of ionizable analytes [26,99–104]. Here, we generalize the treatment of ionizable analytes and examine the case of a n-protic acid, \(HA\), to derive the dependence of \(k\) upon \(pH\).

The acid \(HA\) in the mobile phase yields \(n\) equilibrium processes, which may be formally represented by the following relationships

\[
H_{n-j}A^{j-} \leftrightarrow H^{+} + H_{n-j-1}A^{j-1}, \quad i = 0, 1, \ldots, n - 1
\]

(37)

where \(H_{n-j}A^{j-}\) means \(H^{+}\) and \(H_{n-j}A^{0-}\) stands for \(HA\). Therefore, there are \(n + 1\) species, \(H_{n-j}A^{j-}\), where \(j\) ranges from 0 to \(n\). For simplicity we denote the species \(H_{n-j}A^{j-}\) by 0, 1, 2, \ldots, \(n\), i.e. by the value of \(j\). Then

\[
k = \lim_{n \to 0} \frac{\sum_{i=0}^{n} x_i^{m}}{\sum_{i=0}^{n} x_i^{m}} = \lim_{n \to 0} \frac{\sum_{i=0}^{n} \phi_i x_i^{m}}{\sum_{j=0}^{n} \phi_j x_j^{m}}
\]

(38)

where

\[
k^* = (n_i^* + n_j^*)/(n_i^* + n_j^*)
\]

and \(x_j^m\) is the mole fraction of \(j\). Note that \(n_j^m, n_j \to 0\). From Eq. (38) we obtain

\[
k = \sum_{j=0}^{n} k_j x_j^m
\]

(39)

where

\[
k_j = k_j x_j^m
\]

(40)

is the retention factor of the \(j\) species.

If we ignore activity effects, which have a constant contribution when the volume fraction \(\varphi\) is constant, the equilibrium constant of Eq. (37) may be written

\[
K_i = \frac{[H^+][x_i^m]}{x_i^{m+1}}, \quad i = 0, 1, \ldots, n - 1
\]

(41)

where \([H^+]\) is the concentration of \(H^+\) in the mobile phase. Now, we proceed as follows. We divide the numerator and denominator of Eq. (39) by \(x_0^m\), write the ratios \(x_i^m/x_0^m\) as \((x_i^m/x_{i-1}^m)\), \((x_i^m/x_{i+1}^m)\) and substitute these ratios from Eq. (41), since \(x_i^m/x_{i-1}^m = K_{i-1}/[H^+]\). Then we obtain the final expression for \(k\)

\[
k = \sum_{j=0}^{n} k_j a_j
\]

(42)

where

\[
a_j = \frac{[H^+][x_i^m]}{[H^+] + [H^+]x_i^m - K_{i-1}x_i^m + \ldots + K_{n-1}x_i^m}
\]

(43)

with \(K_j = 1\).

Activity effects may be easily embodied following the same approach. The thermodynamic equilibrium constant of Eq. (37) may be written as

\[
k_j = k_j [H^+]^n \gamma_i^n / \gamma_j^n
\]

(44)

where \(\gamma_i^n\) is the activity of \(H^+\) in the mobile phase and \(\gamma_i^n\) is the activity coefficient of species \(i\) also in the mobile phase. From this equation we obtain again Eq. (42) but with \(a_j\) given by

\[
a_j = \frac{(\alpha x_i^m - \alpha x_j^m + x_i^m K_{i-1}y_i/\gamma_j)}{(\alpha x_i^m + \alpha x_j^m + x_j^m K_{i-1}y_i/\gamma_j)}
\]

(45)

For a monoprotic acid Eqs. (42) and (43) yield the familiar expression

\[
k = k_0 + k_1 \frac{[H^+] - pK}{1 + 10^{[H^+] - pK}}
\]

(46)

whereas from Eqs. (42) and (45) we obtain

\[
k = k_0 + k_1 \frac{[H^+] - pK}{1 + 10^{[H^+] - pK}}
\]

(47)

The above treatment is general enough and covers also bases and ampholytes since their ionization may be formally described by relationships like those of Eq. (37). We also observe that the final expression for \(k\), Eq. (42), is independent on the retention mechanism. However, the retention mechanism affects the values of the retention factors \(k_j\) and therefore it affects the expression of these parameters in terms of \(\varphi\). The effect of \(\varphi\) on \(pH, pK\) and eventually on the retention factor \(k\) is discussed in Section 4.1.2.

### 3.1.4. Retention models for variations in temperature

Up to now the majority of studies on the effect of temperature in RPLC is mainly focused on elucidating the retention mechanism by obtaining thermodynamic parameters from van’t Hoff plots [105–112]. In contrast, there are relatively few studies where temperature has been used as a parameter to improve separations [113–123], although nearly all the physical parameters that play a role in liquid chromatographic separation are a function of temperature. Nevertheless, nowadays column temperature is recognized as an important variable in controlling retention and/or selectivity in RPLC [113,114].

The effect of temperature on the various equilibrium processes taking place inside a chromatographic column may be taken into consideration through the Gibbs–Helmholtz relationship [94,124]
the retention process, $T$ the column temperature, $P$ the pressure and $x$ stands for the composition. By integration Eq. (48) results in

$$\frac{\Delta G}{T} = \frac{\Delta H_0}{T} + \Delta H_1 g(T) + \text{const}$$

(49)

where the function $g(T)$ is given by

$$g(T) = \begin{cases} 
0 & \text{when } \Delta H^f = \Delta H_0 \\
-\ln T & \text{when } \Delta H^f = \Delta H_0 + \Delta H_1 T \\
-\frac{T}{\Delta H^f} & \text{when } \Delta H^f = \Delta H_0 + \Delta H_1 T^2 \\
2/\Delta H^f & \text{when } \Delta H^f = \Delta H_0 + \Delta H_1 / T
\end{cases}$$

(50)

where $\Delta H_0$ and $\Delta H_1$ are constants, coefficients of $\Delta H^f$. If we take into account that $\ln k = -\Delta G/RT + \ln \Phi$, where $\Phi$ is the phase ratio, we readily obtain the following general expression for $\ln k$ provided that $\Phi$ is temperature invariant [47]

$$\ln k = c_0 + \frac{c_1}{T} + c_2 g(T)$$

(51)

In Eq. (51) coefficients $c_0, c_1$, and $c_2$ are independent of $T$, whereas they depend upon the composition of the mobile phase. Note also that Eq. (51) with $c_2 = 0$ has been used extensively in the literature [115–120]. Another equation, which has been adopted to study the temperature effect on the retention properties of analytes, is that with $g(T) = 2/T^2$ [121,122].

We should note that, as in the case of $pH$, the final expression for $\ln k$, Eq. (51), seems to be independent of the retention mechanism. The retention mechanism affects $\ln k$ through the dependence of $c_0$, $c_1$, $c_2$ coefficients on $\Phi$. The retention mechanism might also affect the expression of $\Delta H^f$ that determines the expression of $g(T)$ in Eq. (51).

At this point we should also clarify that the temperature which should be used in Eq. (51) is not the oven temperature, $T_{ov}$, but the effective temperature, $T_{ef}$, that the analyte feels inside the column [23,123]. This effective temperature is identical to the oven temperature under isothermal conditions but it becomes quite different under temperature programming conditions when a column with conventional dimensions is used. In this case the effective temperature may be estimated from Newton’s law [23,123]

$$\frac{dT}{dt} = -h(T - T_{ov})$$

(52)

where $h$ is a hysteresis constant characteristic of the chromatographic system. Thus, this parameter should become less significant in case a capillary column is used instead of a column with conventional dimensions. Under temperature programming conditions the oven temperature $T_{ov}$ varies with time $t$ but since it is a continuous function of $t$ it can always be expressed as a power series of $t$

$$T_{ov} = a_0 + a_1 t + \cdots + a_n t^n$$

(53)

Then Eq. (52) yields

$$T_{ef} = \left( a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \right) - \frac{1}{h} \left( a_1 + 2a_2 t + \cdots + na_n t^{n-1} \right) + \frac{1}{h^2} \left( 2a_2 + \cdots + n(n-1)a_n t^{n-2} \right) + \cdots + \left( -\frac{1}{n^2} \right)^n (n!a_n) + c e^{-ht}$$

(54)

where $c$ is an integration constant calculated from the initial condition $T_{ef} (t=0) = T_{ov} (t=0)$. The hysteresis constant $h$ is estimated experimentally from the comparison of experimental and calculated retention times using two or more T-gradients [23,123]. Fig. 2 depicts the oven temperature profiles of four T-gradients (----) and the corresponding effective temperature (---) that the analyte experiences. They have been obtained for a Zorbax SB-C18 column (3.5 μm, 150 mm × 4.6 mm) thermostatted by a CTO-10AS Shimadzu column oven. The mobile phase was binary mixtures of acetonitrile–water moved with a flow rate equal to 1.0 mL/min.

3.2. Exo-thermodynamic models

Initially the term exo-thermodynamic or extra-thermodynamic relationships was used for every linear relationship between $\ln k$ and the analyte parameters or experimental variables provided that it falls outside the formal structure of thermodynamics. Now it includes every empirical correlation of thermodynamic parameters. They have been used extensively in RPLC to clarify the role of molecular structural parameters in the retention process and thus to shed light on the underlying mechanism. For excellent reviews and comprehensive articles on this topic, see [14,16,19,125,126].

The exo-thermodynamic relationships related to retention models are the linear free energy relationships (LFERs) with main subcategories the linear solvation energy relationships (LSERs) [17,125,127–143] and the quantitative structure–activity relationships (QSARs) [144–157]. Note that one variable exo-thermodynamic models exist only for variation in the mobile phase composition. The retention models for $pH$ and $T$, Eqs. (42) and (51), are strict thermodynamics relationships and leave no room for exo-thermodynamic models.

The LSER expression of the logarithmic capacity factor using recent Abraham’s notation is given by [112,135]

$$\ln k = c + eE + sS + aA + bB + vV$$

(55)

where the letters $E$, $S$, $A$, $B$, $V$ are analyte parameters (descriptors) representing its polarizability, dipolarity, hydrogen bond donating ability, hydrogen bond accepting ability, and molecular size, respectively. A similar to Eq. (55) expression has been also used by Snyder et al. [158–165] to estimate solute retention and column selectivity from $\ln a = \ln(k/k_{ref})$, where $k_{ref}$ is the retention factor of a reference solute. Descriptors for more than 4000 compounds are available [166] plus a software program to estimate analyte descriptors from structure [167]. The coefficients $c, s, a, b, v$ and the constant $c$ reflect properties of the mobile phase and the stationary phase of the chromatographic column and they are determined by multi-parameter linear least squares fit to experimental data. Note that in literature and especially in exo-thermodynamic studies $\log k$ is used instead of $\ln k$. However, in this article for reasons of coherence we use the natural logarithm of $k$ for all expressions of the retention models.
When the column/mobile phase system changes, e, s, a, b, m, and c change as well. Thus, in the same column but at different mobile phase compositions the above coefficients become functions of \( \phi \). Wang et al. proposed a linear dependence of these parameters upon \( \phi \) [125]. For wider ranges of solvent compositions Torres-Lapasió et al. proposed, among others, the following relationship with quadratic terms [143]

\[
\ln k = (c_0 + c_1 \phi + c_2 \phi^2) + (c_0 + c_1 \phi + c_2 \phi^2)E + (s_0 + s_1 \phi + s_2 \phi^2)S + (a_0 + a_1 \phi + a_2 \phi^2)A + (b_0 + b_1 \phi + b_2 \phi^2)B + (v_0 + v_1 \phi + v_2 \phi^2)\V 
\]

The above relationships have been used for the prediction of the elution time under isocratic conditions. Similar expressions have been proposed for gradient elution [168–170].

Other useful correlations can be also established between \( \ln k \) or \( \log k \) and the octanol–water partition coefficients (\( \log P_{O/W} \)) [144,154,171]

\[
\ln k = c_0 + c_1 \log P_{O/W} 
\]

However, it has been shown [150,172] that the correlation of the two scales, \( \ln k \) (or \( \log k \)) and \( \log P_{O/W} \), needs also the hydrogen bond acidity, that is

\[
\ln k = c_0 + c_1 \log P_{O/W} + c_2 A 
\]

It is evident that the above relationships could be used as retention models for isocratic or gradient conditions of variable \( \phi \) provided that coefficients \( c_0, c_1, c_2 \) are expressed in terms of \( \phi \), like in Eq. (56). Thus, Eq. (58) is written as

\[
\ln k = c_0 + c_1 \phi + c_2 \phi^2 + (p_0 + p_1 \phi + p_2 \phi^2) \log P_{O/W} + (a_0 + a_1 \phi + a_2 \phi^2)A + (b_0 + b_1 \phi + b_2 \phi^2)B + (v_0 + v_1 \phi + v_2 \phi^2)\V 
\]

and a corresponding expression holds for Eq. (57).

At this point it is interesting to note that the above formulations of the LSER equations focus on understanding the analyte behaviour. In contrast, the earliest LSER formulations focused on the solvent properties. Thus, for the same analyte and the same column but with different mobile phase compositions Eq. (55) had the expression [173]

\[
\ln k = c_0 + m_m \sigma_m^2 + s_m \pi_m + a_m \beta_m + b_m \alpha_m 
\]

where subscripts \( m \) and \( A \) denote the mobile phase and the analyte, respectively, \( \sigma \) the square of the Hildebrand solubility parameter, \( \pi \) a measure of the dipolarity/polarizability of the molecules, \( \alpha \) an empirical measure of hydrogen bond donating acidity, and \( \beta \) is a similar scale of hydrogen bond acceptor basicity. The coefficients \( m_m, s_m, a_m, \alpha_m \) and \( b_m \) are fitting parameters. For applications, Eq. (60) usually was simplified to [174]

\[
\ln k = c_0 + s_m \pi_m + a_m \beta_m + b_m \alpha_m 
\]

by assuming a linear relationship between \( \sigma_m^2 \) and \( \pi_m, \beta_m, \alpha_m \). Eqs. (60) and (61) can be used as model equations if the dependence of \( \pi_m, \beta_m, \alpha_m \) upon \( \phi \) is known. For this reason \( \pi_m(\phi) \) expressions may be calculated from the corresponding data of Cheong and Carr [175], \( \alpha \) values may be computed form \( E^{N_n}_\phi \) and \( \pi_m \) values [174,176], and finally, \( \beta \) values may be obtained by interpolation of the corresponding values given in Ref. [177].

A useful correlation used extensively as model equation is the two-scales, log scale [173,178–182], which later was replaced by the normalised \( E^{N_n}_\phi \) parameter [174,183]. Eq. (62) has been extensively used in RPLC studies under isocratic conditions. In order to be able to be used in gradient studies, parameter \( E^{N_n}_\phi \) should be expressed in terms of \( \phi \). It has been found that the experimental values of \( E^{N_n}_\phi \) can be fitted to the following equation

\[
E^{N_n}_\phi = \frac{1 + p_1 \phi}{1 + b_1 \phi} 
\]

where \( p_1 \) and \( b_1 \) are constants that take certain values for each modifier [92]. Table 2 depicts the values of \( p_1 \) and \( b_1 \) for four commonly used modifiers and the range of \( \phi \) values where Eq. (63) is valid. Note that almost the same results have been suggested by Rosés and Bosch for the calculation of \( E^{N_n}_\phi \) values in methanol–water and acetonitrile–water solutions [174].

If Eq. (63) is combined with Eq. (62), we obtain the two-parameter equation

\[
\ln k = c_0 + c_1 \frac{1}{1 + b_1 \phi} 
\]

where \( c_0 = m + n p_1/b_1 \) and \( c_1 = n – np_1/b_1 \). This equation is in fact a two-parameter relationship, where \( b_1 \) necessarily takes the values of Table 2. Similar to Eq. (64) expressions have been used by Bosch et al. [183–186]. However, extensive tests of this equation have shown a rather limited applicability [92].

Eq. (64) may be extended by taking into account the analyte interactions at the stationary phase [92]. These interactions may be accounted for by adding to the expression of \( \ln k \) a term like \( qE^{n}_T \), where \( qE^{n}_T \) is the corresponding \( E^{n}_T \) factor of the stationary phase, i.e. a measure of the polarity of this phase. It is evident that the expression of \( qE^{n}_T \) and therefore the contribution of the analyte interactions at the stationary phase depend on the retention mechanism. If the retention is due to partition, \( qE^{n}_T \) is a constant, because the stationary phase is just the hydrocarbon chains. A constant \( qE^{n}_T \) has been first proposed by Bosch et al. [183]. In this case, the term \( qE^{n}_T \) is also a constant, which can be included in parameter \( c_0 \) of Eq. (64). In contrast, if the retention is due to adsorption, a surface solution is formed on the chains and \( qE^{n}_T \) is the measure of its polarity, i.e. a measure of the solvent (water + organic modifier) polarity at the surface solution of the stationary phase. Therefore, we may assume that \( qE^{n}_T \) is given by an expression like Eq. (63) but with different \( b_1 \) and \( p_1 \) values, say \( b_2 \) and \( p_2 \), and with the surface volume fraction \( \phi^2_{\phi} \) of the modifier in place of \( \phi \). That is

\[
qE^{n}_T = \frac{1 + p_2 \phi^2_{\phi}}{1 + b_2 \phi^2_{\phi}} 
\]

and if we use Eq. (21) to interrelate \( \phi^2_{\phi} \) and \( \phi \), we obtain

\[
qE^{n}_T = \frac{1 + p' \phi}{1 + b' \phi} 
\]

where \( p' = \beta + p_1(\beta + 1) \) and \( b' = \beta + b_1(\beta + 1) \). Now the addition of the term \( qE^{n}_T \) to Eq. (62) yields [92]

\[
\ln k = c_0 + c_1 \frac{1}{1 + b' \phi} + \frac{c_2}{1 + b' \phi} 
\]

### Table 2

<table>
<thead>
<tr>
<th>Mobile phase</th>
<th>( p_1 )</th>
<th>( b_1 )</th>
<th>( \phi ) range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methanol–water</td>
<td>0.40</td>
<td>0.83</td>
<td>( 0 \leq \phi \leq 1 )</td>
</tr>
<tr>
<td>Acetonitrile–water</td>
<td>0.51</td>
<td>1.1</td>
<td>( 0 \leq \phi \leq 0.8 )</td>
</tr>
<tr>
<td>Isopropanol–water</td>
<td>0.51</td>
<td>1.6</td>
<td>( 0 \leq \phi \leq 0.9 )</td>
</tr>
<tr>
<td>Tetrahydrofuran–water</td>
<td>0.54</td>
<td>1.9</td>
<td>( 0 \leq \phi \leq 0.8 )</td>
</tr>
</tbody>
</table>

From Ref. [92] with permission.
Kowalska’s model is not the same as that in Table 2 used for Eq. (64) [92].

Models: with more or less empirical bases have been proposed as retention models and the simple rational Eq. (24) from the adsorption model. We have already seen that the quadratic Eq. (30) has been obtained from the partition model. Thermodynamic arguments have been used to support this model. The highest applicability describing satisfactorily more than 90% of the systems studied in Ref. [92], provided that we use the values of \( b_k \) and \( b'_k \) depicted in Table 3. Note that the value of \( b_k \) is not the same as that in Table 2 used for Eq. (64) [92].

### 3.3. Empirical models

In this category belong mainly polynomial and rational function models. The polynomials are popular because they have a simple form, well known and understood properties and they are computationally easy to use. Their major limitation is that they easily create over-fitting problems. This limitation is rare when we use rational functions. A rational function is simply the ratio of two polynomial functions, where the constant term of the polynomial of the denominator is set equal to 1. The rational functions usually exhibit very good fitting performance without over-fitting problems but these functions have a moderately simple form and may present convergence problems.

We should clarify that although polynomials and rational functions are in general empirical models, some of them may be obtained from thermodynamic arguments. We have already seen that the quadratic Eq. (30) has been obtained from the partition model and the simple rational Eq. (24) from the adsorption model.

Apart from the above two categories, the following relationships are in general empirical models, some of them may be obtained from thermodynamic arguments, Eqs. (56), (57) and (59), from fitting the rest ln \( k \) of the main conventional retention models we used the following model:

\[
\ln k = c_0 + c_1\phi + c_2\phi^2 + c_3\sqrt{\phi} 
\]

(Eq. (68))

(b) **Kowalska’s model.** Kowalska [85, 187, 188] has proposed the following model:

\[
k = \frac{1}{c_0 + c_1\sqrt{\phi} + c_2\sqrt{1-\phi}} - 1
\]

(Eq. (69))

This model was originally derived for reversed-phase planar chromatography [85, 187] and later applied to RPLC [188].

### 3.4. A comparison of the retention models

The retention models that express \( \ln k \) in terms of \( \phi \) or \( T \) are based on strict thermodynamic arguments and on a minimum of assumptions. For example, the \( \ln k = f(\phi) \) models may assume negligible or non-negligible activity effects, whereas the final expression of the \( \ln k = f(T) \) models depends upon the dependence of the standard enthalpy of the retention process on \( T \). In contrast, the \( \ln k = f(\phi) \) models are very sensitive on the assumptions made about not only the retention mechanism but also the various intermolecular interactions. For this reason, there exists a great variety of \( \ln k = f(\phi) \) retention models and the question arises which of them are the most appropriate for retention prediction and optimization separation conditions based on a relatively small number of experiments.

From the \( \ln k = f(\phi) \) expressions presented above that with the lowest performance is the linear model of Eq. (27). It is valid only if the range of \( \phi \) values is narrow enough. Despite this, it has been used extensively in gradient runs and in particular in combination to a linear gradient profile. This combination is called linear solvent strength gradient [191–197] and constitutes the base of DryLab, the most widely used HPLC simulation package [198, 199]. The results are improved if we use Eq. (26) and \( \beta \) takes either a common value for each modifier or even better a common value per group of structurally similar analytes at each modifier. The use of two-parameter retention models has the advantage that relatively few experiments are needed to obtain their adjustable parameters. However, they exhibit relatively low fitting performance and should be handled cautiously when used for prediction and optimization.

Extensive comparisons of various retention models have shown that many three-parameter models are a good compromise between simplicity, accuracy and good numerical behaviour [92]. However, all the three-parameter models presented above have not been tested using the same database. In addition, there is not a detailed comparison between exo-thermodynamic relationships and the simple three-parameter models in what concerns their performance for describing the retention behaviour of analytes. For this reason, we have proceeded here to such a comparison.

However, at this point we should stress that there is a distinct difference in fitting the exo-thermodynamic relationships, Eqs. (56), (57) and (59), from fitting the rest \( \ln k = f(\phi) \) retention models. These exo-thermodynamic relationships are applied to the whole data set of all analytes obtained with a particular modifier and column, whereas all the other retention models are applied to the data set of each analyte. In order to compare the fitting performance of the various exo-thermodynamic relationships and that of the main conventional retention models we used the following data sets. Six non-polar analytes, benzene (B), toluene (T), ethylbenzene (EB), propylbenzene (PB), isopropylbenzene (iPB), and tert-butylbenzene (tBB) plus eight phenol derivatives, phenol (P), 4-nitrophenol (4-NP), 2-methylphenol (2-MP), 2-chlorophenol (2-CP), 4-chlorophenol (4-CP), 4-bromophenol (4-BP), 2,4-dimethylphenol (2,4-DMP) and 3-bromophenol (3-BP). The retention data of all analytes were obtained in aqueous mobile phases modified with methanol, isopropanol, acetonitrile and tetrahydrofuran using C18 and C2 reversed-phase columns. Data of benzene derivatives were taken from Tables 1–4 of [95], whereas those of phenol derivatives were presented in the ICCECS 4, Conference (Belgrade, July 18–21, 2004). For the comparisons we used the data in the \( \phi \) range from 0.2 to 0.9 in testing the exo-thermodynamic relationships, Eqs. (56), (57) and (59), and all the three-parameter Eqs. (24), (25), (30), (35), (36), (67), (69) and (71). The compounds used for the test are very simple analytes. They have been selected as model solutes because their retention data are available in a wide \( \phi \) range for four organic modifiers and for
Table 4
Average percentage error between experimental and calculated retention times over all solutes and mobile phase concentrations at a certain column and organic modifier

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MeOH/C_{2}</td>
<td>2.0</td>
<td>2.1</td>
<td>1.9</td>
<td>2.1</td>
<td>109.9</td>
<td>2.7</td>
<td>4.9</td>
<td>4.9</td>
<td>1.6</td>
<td>275.7</td>
<td>23.8</td>
</tr>
<tr>
<td>MeOH/C_{18}</td>
<td>1.8</td>
<td>1.7</td>
<td>1.8</td>
<td>2.3</td>
<td>103.2</td>
<td>3.5</td>
<td>6.5</td>
<td>6.1</td>
<td>2.3</td>
<td>49.4</td>
<td>31.3</td>
</tr>
<tr>
<td>MeCN/C_{2}</td>
<td>1.0</td>
<td>1.0</td>
<td>1.6</td>
<td>1.0</td>
<td>352.6</td>
<td>2.1</td>
<td>3.8</td>
<td>3.5</td>
<td>2.4</td>
<td>206.7</td>
<td>11.0</td>
</tr>
<tr>
<td>MeCN/C_{18}</td>
<td>1.5</td>
<td>1.4</td>
<td>2.5</td>
<td>1.4</td>
<td>90.3</td>
<td>2.9</td>
<td>5.6</td>
<td>5.2</td>
<td>2.0</td>
<td>43.9</td>
<td>16.9</td>
</tr>
<tr>
<td>iPrOH/C_{2}</td>
<td>3.4</td>
<td>2.8</td>
<td>3.0</td>
<td>5.3</td>
<td>34.5</td>
<td>3.0</td>
<td>4.0</td>
<td>4.0</td>
<td>3.7</td>
<td>23.8</td>
<td>48.2</td>
</tr>
<tr>
<td>iPrOH/C_{18}</td>
<td>3.8</td>
<td>3.7</td>
<td>4.0</td>
<td>4.6</td>
<td>72.3</td>
<td>4.2</td>
<td>5.7</td>
<td>5.7</td>
<td>3.7</td>
<td>54.9</td>
<td>47.0</td>
</tr>
<tr>
<td>THF/C_{2}</td>
<td>2.1</td>
<td>2.0</td>
<td>1.9</td>
<td>2.1</td>
<td>106.4</td>
<td>3.3</td>
<td>4.4</td>
<td>4.4</td>
<td>1.7</td>
<td>81.7</td>
<td>37.5</td>
</tr>
<tr>
<td>THF/C_{18}</td>
<td>2.0</td>
<td>2.1</td>
<td>3.1</td>
<td>1.8</td>
<td>75.5</td>
<td>4.2</td>
<td>4.9</td>
<td>4.9</td>
<td>2.1</td>
<td>60.1</td>
<td>32.3</td>
</tr>
<tr>
<td>Average</td>
<td>2.1</td>
<td>2.1</td>
<td>2.6</td>
<td>2.6</td>
<td>118.1</td>
<td>3.2</td>
<td>5.0</td>
<td>4.8</td>
<td>2.4</td>
<td>99.5</td>
<td>31.0</td>
</tr>
</tbody>
</table>

two columns, C_{18} and C_{2} reversed-phase columns. In the first column the partition mechanism is expected to predominate whereas the retention should be due to adsorption in the C_{2} column.

Table 4 depicts the average percentage error between experimental and calculated retention times over all analytes and mobile phase concentrations at a certain column and organic modifier. Note that retention data of polar solutes and non-polar solutes were fitted to Eqs. (56), (57) and (59) separately, because the mobile phases used for polar solutes were aqueous buffers with 0.07 acetic acid and not pure aqueous mobile phases such as those used for the retention study of the non-polar solutes. It is seen that Eqs. (36), (69) and (71) give extremely high errors and therefore cannot be used in fitting procedures. The rest of the three-parameter equations seem to exhibit a good fitting performance. The best results are obtained from Eqs. (24) and (25), at least for the data set used in the present test. Eqs. (67), (30) and (35) also give good results. Among the exo-thermodynamic retention models better results are obtained from Eq. (56).

Fig. 3 shows the average percentage error between experimental and calculated retention times of all analytes over all mobile phase concentrations in MeCN for a C_{18} column. It is seen again that Eqs. (24) and (25) exhibit the best fitting performance and for this system an equally good performance is shown from Eq. (35). Eq. (56) seems to give acceptable results. However, an acceptable average error may screen high errors at certain data. Fig. 4 shows the variation of the percentage error between experimental and calculated retention times of benzene derivatives at the various mobile phase concentrations in THF (A) and MeCN (B) for C_{18} columns. The calculated t_{R} values were obtained by means of Eq. (24) (○) and Eq. (56) (▲), respectively. We observe that despite the relatively low average errors in Table 4, the error in the retention times at certain elution conditions may be quite high for Eq. (56).

We should also point out that the plots ln k_{exp} vs. ln k_{calc} do not give a safe picture about the fitting performance of a certain retention model. In Fig. 5 the ln k_{calc} vs. ln k_{exp} plot in iPrOH and C_{18} column for benzene derivatives using Eq. (59) is linear with R^2 = 0.9969 but the model fails to give acceptable fittings (Fig. 6). In general, from the exo-thermodynamic retention models fitted to the retention data of all solutes as a whole, only Eq.
(56) seems to give acceptable fittings. However, its use as retention model does not give any advantage over the conventional models, like those of the three-parameter Eqs. (24), (25), (30) and (35). The average error between predicted and experimental retention times is higher for Eq. (56) and from a practical point of view this model is more complicated when it is fitted to gradient data. In this case we have to compute 18 adjustable parameters using a non-linear fitting procedure.

Finally, we observe that Eq. (30) exhibits a rather good performance. However, we should stress again that it may create over-fitting problems or give physically meaningless ln $k$ vs. $\phi$ plots \cite{46,104,200,201}. Such a case is depicted in Fig. 7, which shows the ln $k$ vs. $\phi$ plot of L-tryptophan in isopropanol–water mobile phases. Data were taken from \cite{45}. The continuous curves have been calculated from Eqs. (24), solid line, (30), dotted line, and (35), dashed line. It is seen that Eqs. (24) and (35) describe satisfactorily the experimental data, whereas Eq. (30) fails to give an acceptable description of the system.

4. Multi-variable retention models

As the complexity of separations increases, the need of optimized procedures based on the existence of theoretical or empirical models describing the dependence of retention on more than one separation variables also increases. Today is generally accepted that RPLC analyses poorly performed by using isocratic elutions in binary solvent mixtures may be enhanced by using isocratic elution in ternary eluents or multi-mode gradient conditions. However, it is difficult to handle two or more variables simultaneously; there is need of proper optimization algorithms and in the case of gradient elution there is need of an appropriate theory for prediction of the analyte elution times. In this section we review the retention models used to describe the variation of retention with composition in ternary mobile phases or with composition in binary mobile phases simultaneously with mobile phase pH and/or column temperature.

4.1. Thermodynamic models

4.1.1. Retention in ternary systems

In ternary systems the mobile phase consists of two organic modifiers, $B$ and $C$. Therefore, if we again ignore size effects, the formation of the adsorption layer on the tips and stems of the hydrocarbon layer will be the result of processes (8) and (9) plus the following process describing the adsorption of $C$

$$C^m + S^a \rightleftharpoons C^a + S^m$$

(72)
The thermodynamic Eqs. (10)–(15) are still valid but Eqs. (16) and (18) should be replaced by [202]
\[
\ln f_i^j = D_i^j + D_2^j \phi_B + D_3^j \phi_C + C_i^j (\psi_B^j)^2 + C_2^j (\psi_C^j)^2 + C_3^j \psi_B^j \psi_C^j
\] (73)
\[
\ln f_i^j = C_i^j (\psi_B^j)^2 + C_2^j (\psi_C^j)^2 + C_3^j \psi_B^j \psi_C^j, \quad j = m \text{ or } a
\] (74)

In addition, instead of Eq. (21) we now have
\[
\psi_i^j = \frac{(\beta_j + 1) \psi_j}{1 + \beta_j \psi_B + \beta_j \psi_C}, \quad i = B \text{ or } C
\] (75)
valid under the approximation
\[
C_i^j \approx C_i^k \approx 0
\] (76)
which assumes an ideal behaviour of the mixture of B, C and S. All the above equations readily extend the yield of expression of Eq. (22) to
\[
\ln k = c_0 - c_{1B} \psi_B + c_{1C} \psi_C + c_{2B} \psi_B^2 + c_{2C} \psi_C^2
+ \ln[C_i^j (c_{1B} \psi_B + c_{1C} \psi_C) (1 + c_{2B} \psi_B + c_{2C} \psi_C) + 1]
\] (77)
where
\[
c_0 = \ln k^+ + D_0^+ + D_1^0, \quad c_{1B} = (\beta_j + 1) D_2^0, \quad c_{1C} = (\beta_j + 1) D_3^0, \quad
c_{2B} = \beta_j C_{2B}, \quad c_{2C} = \beta_j C_{2C}, \quad c_{2B}^2 = D_2^0, \quad c_{2C} = D_3^0, \quad \text{and} \quad c_4 = (k/r^E) e^{D_1 - D_1^0}.
\]
Again, if adsorption predominates, \(c_4 \rightarrow 0\) and Eq. (77) is reduced to [202]
\[
\ln k = c_0 - c_{1B} \psi_B + c_{1C} \psi_C + c_{2B} \psi_B^2 + c_{2C} \psi_C^2
\] (78)
which is further simplified to
\[
\ln k = c_0 - c_{1B} \psi_B + c_{1C} \psi_C + c_{2B} \psi_B^2 + c_{2C} \psi_C^2
\] (79)
under the assumption that \(D_0^+ \approx 0\) and \(D_0^j \approx 0\). Note that \(c_0, c_{1B}, c_{1C}, c_{2B}, c_{2C}\) are constants calculated by studying the isocratic retention of the sample analyte in the two binary mobile phases composed of aqueous solutions of the organic solvents B and C, respectively. Therefore, Eq. (79), and the same holds for Eq. (78), can predict the retention behaviour of solutes under ternary solvent mixture conditions from known retention characteristics in binary mobile phases. However, experimental tests have shown that this is possible only for dilute solutions, i.e., when \(\psi_B, \psi_C \leq 0.2\) [202]. If this prerequisite does not hold, Eq. (79) fails completely to predict retention times using binary data [203].

If partition is the predominant retention mechanism, then \(c_4 \rightarrow \infty\) and Eq. (77) yields the linear expression
\[
\ln k = c_0 + c_{1B} \psi_B + c_{1C} \psi_C
\] (80)
This expression extends to the following quadratic expression [57,202]
\[
\ln k = c_0 + c_{1B} \psi_B + c_{2B} \psi_C + c_{3B} \psi_B^2 + c_{4B} \psi_C^2 + c_{5B} \psi_B \psi_C
\] (81)
under the assumption that \(\ln f_i^j\) is represented by a similar expression.

4.1.2. Retention in systems of variable pH and mobile phase composition

If we ignore activity coefficients, the final expression for \(k\), Eq. (42), shows that it is a combination of the retention factors, \(k_j\), of all species \(j\) arising from the dissociation of the ionic analyte, the corresponding ionization constants, \(K_j\), and the pH of the mobile phase. Therefore, Eq. (42) may be extended to include changes in the mobile phase composition if \(k_j\) and \(K_j\) are expressed in terms of \(\psi\), \(k_j(\psi)\) and \(K_j(\psi)\). In Section 3 we developed a great number of \(k_j(\psi)\) expressions. In what concerns \(K_j(\psi)\), we can make use of Born’s equation [204–206]
\[
pK = pK_r + \frac{D}{\varepsilon}
\] (82)
where \(pK_r\) and \(D\) are constants and \(\varepsilon\) is the dielectric constant of the mobile phase. The dielectric constant may be measured by a dipolometer and therefore its \(\varepsilon(\psi)\) expression may be known. Otherwise we are obliged to assume a certain \(\varepsilon(\psi)\) expression. Usually it is assumed that either \(\varepsilon\) or \(1/\varepsilon\) varies quadratically with \(\psi\). Thus, we obtain

(a) If \(\varepsilon = \varepsilon_0 + \varepsilon_1 \psi + \varepsilon_2 \psi^2\), then
\[
pK = pK_r + 1/\varepsilon_0 + \varepsilon_1 \psi + \varepsilon_2 \psi^2
\] (83)

(b) If \(1/\varepsilon = \varepsilon_0^+ + \varepsilon_1^+ \psi + \varepsilon_2^+ \psi^2\), then
\[
pK = pK_0^+ + r_1 \psi + r_2 \psi^2
\] (84)

This procedure can be applied to any ionizable analyte. For example, for a monoprotic acid if we make use of Eqs. (24), (46) and (84), we obtain
\[
k = \frac{b_0}{1 + 10^{pK - pK_0^+ - r_1 \psi - r_2 \psi^2}}
\] (85)
where the adjustable parameters are eight: \(b_0, K_0^+, pK_0^+, c_0, c_1, c_2, r_1, r_2\). This retention model has been first proposed by Neue et al. [96]. The same expression is obtained if we include activity coefficients. In this case we have [104]
\[
\frac{\gamma_i^m}{\gamma_i^1} = e^{b_0 + q_i \psi}
\] (86)
where \(q_0, q_1\) are constants, and therefore Eqs. (24), (47), (84) and (86) result again in Eq. (85), where now coefficients \(pK_0^+\) and \(r_1\) include also contributions from \(q_0\) and \(q_1\).

If we use the quadratic Eq. (30) in place of Eq. (24), we obtain the retention model
\[
k = \frac{b_0^0 \psi^2 + c_0 \psi + c_1 \psi^2}{1 + 10^{pK - pK_0^+ - r_1 \psi - r_2 \psi^2}}
\] (87)
with nine adjustable parameters: \(b_0^0, K_0^+, pK_0^+, c_0, c_1, c_2, r_1, r_2\). It is seen that this retention model has one adjustable parameter more than Eq. (85). However, the problem with Eq. (87) is that, depending on the system, serious over-fitting problems may be detected. A typical case is shown in Fig. 8.

Note that usually the pH of the mobile phase is measured in the aqueous buffer before mixing it with the organic modifier. This pH scale denoted by \(\text{pH}^B\) has been adopted in the vast majority of studies because it is very simple, it does not present technical difficulties in automated systems or routine optimization procedures and reduces the required number of pH measurements [99]. However, the rigorous thermodynamic pH scale in mobile phases modified by an organic solvent is defined from \(pK_r = \log a_{H^{+},r}\). This pH scale has been suggested by Bosch et al. [206–210]. It is related to the \(\text{pH}^B\) scale, i.e., when the pH is measured after mixing the aqueous buffer with the organic modifier using a pH electrode system calibrated with common aqueous reference electrode by the following relationship [206–210]:
\[
\text{pH}^B = \text{pH} - \delta
\] (88)
where the quantity \(\delta\) depends only on the composition of the mobile phase. Values of this quantity can be found in [206–210]. For modifier contents less than 15% the term \(\delta\) has a negligible contribution and this contribution is still small but detectable when \(\psi\) is above 30% (\(\psi/\psi^0\)).
Finally, we should point out that a comparison of various retention models for ionisable analytes using both the simple $\ln pH$ and the strict thermodynamic $\ln pH$ scale has shown that the scale does not affect the performance of the fitting equation [104]. Only the $pK$ values calculated in organic-water buffers are strongly affected by the $pH$ scale and reliable values are obtained if the $\ln pH$ scale is used.

4.1.3. Retention in systems of variable $T$ and mobile phase composition

The general $T$ dependence of the activity coefficients is the following [47]

$$\ln f_j = b_0 + b_1 x + b_2 x^2 + \cdots + (b_0' + b_1' x + b_2' x^2 + \cdots)/T$$

$$+ \left[ (b_0'' + b_1'' x + b_2'' x^2 + \cdots) g(T) \right]$$

Thus, Eq. (15) is written as

$$\ln f_j^\prime = \frac{b_0' + b_0'' g(T)}{T}$$

and Eq. (16) as

$$\ln f_j^\prime = \frac{b_0' + b_1' \psi_{j\phi} + b_2' \psi_{j\phi}^2 + b_3' \psi_{j\phi}^3}{T}$$

$$+ \left[ (b_0'' + b_1'' \psi_{j\phi} + b_2'' \psi_{j\phi}^2) g(T) \right], \quad j = m \text{ or } a$$

where $g(T)$ is given by Eq. (50). If we use these expressions for $\ln f_j$, we readily obtain for the partition model [47,122]

$$\ln k = c_0 + c_1 \psi + c_2 \psi^2 + \frac{c_3 + c_4 \psi + c_5 \psi^2}{T} + (c_6 + c_7 \psi + c_8 \psi^2) g(T)$$

where the coefficients $c_0, c_1, \ldots$ are independent of $T$ and $\psi$. Eq. (92) is further reduced to

$$\ln k = c_0 + c_1 \psi + \frac{c_2 + c_3 \psi}{T} + (c_4 + c_5 \psi) g(T)$$

(93)

when the $\psi$ range is relatively narrow. Similarly the adsorption model yields [47]

$$\ln k = c_0 + \frac{c_1}{T} + c_2 g(T)$$

$$- \frac{[c_3 + (c_4/T) + c_5 g(T)] [\exp(c_6 + [c_7/T] + c_8 g(T))] \psi}{1 + [\exp(c_6 + [c_7/T] + c_8 g(T)) - 1] \psi}$$

(94)

Comparisons of these equations with $g(T) = 0$ has shown a better performance of Eq. (94) over Eq. (92), although the latter is more easily fitted to experimental data [47]. Such a comparison of Eqs. (92)–(94) as well as of the empirical Eq. (112) presented below is shown in Fig. 9. This figure illustrates the effectiveness of these equations for describing the dependence of experimental values of $\ln k$ vs. $1/T$ for tert-butylbenzene using a mobile phase with $\varphi = 0.5$ in acetonitrile. Data have been taken from [47]. It is seen that Eq. (94) can describe curved $\ln k$ vs. $1/T$ plots at constant $\psi$ using the least number of adjustable parameters, $c_1, c_2, \ldots, c_8$. However, this has its own cost. In contrast to the other equations, Eq. (94) is not reduced to a simple equation when $\varphi = constant$. It is readily shown that under constant $\varphi$, Eqs. (92) and (93) are reduced to Eq. (51), whereas Eq. (94) is reduced to Eq. (51) only in the limited cases when $\varphi = 0$ or 1. In all other cases it remains a six-parameter equation.

Finally, we should point out that the application of the above retention models to gradients with variable temperature requires the use of the effective temperature, which may be calculated from Eq. (54) [23,123].

4.1.4. Retention in systems of variable $T$, $pH$ and mobile phase composition

The combined effect of solvent content, temperature and $pH$ on the chromatographic behaviour of ionisable compounds has been studied by Heinisch et al. [121] and Pous-Torres et al. [114]. The retention factor in terms of $\varphi$, $T$ and $pH$, $k(\varphi, T, pH)$, arises from the general Eq. (42) if $k_j$ are replaced by expressions like those of Eqs. (92)–(94). In addition, the various equilibrium constants $K_j$ in the form of $pK_j$ may be replaced by an expression of the form

$$pK = c_0 + c_1 \varphi + c_2 \varphi^2 + \frac{c_3 + c_4 \varphi + c_5 \varphi^2}{T} + (c_6 + c_7 \varphi + c_8 \varphi^2) g(T)$$

(95)

which, in fact, combines the quadratic expression of Eq. (92) with a temperature dependence of $pK(T)$ like that of Eq. (51). As expected, the final expressions of $k(\varphi, T, pH)$ are very complicated.
4.2. Exo-thermodynamic models

Although the extension of the one-variable exo-thermodynamic models to multi-variable models is straightforward, the relative work in this area is rare. The first attempt has been made by Espinosa et al. [140] who wrote Eq. (46) as

$$\ln k = \ln k_0 + \ln[1 - D(1 - f)]$$

(96)

where \(f = k_1/k_0\) and \(D\) is the degree of ionization of the analyte

$$D = 10^{pH-pK} / (1 + 10^{pH-pK})$$

(97)

Then they combined Eqs. (55) and (96) to obtain

$$\ln k = c + eE + s + aA + bB + vV + \ln[1 - D(1 - f)]$$

(98)

The combined effect of \(pH\) and the mobile phase composition is taken into account similarly [140]

$$\ln k = \ln k_{0,0} + \varphi \ln k_{0,\varphi} + (e_w - \varphi e_s)E + (s_w - \varphi s_s)S + (\alpha_w - \varphi \alpha_s)A + (b_w - \varphi b_s)B + (v_w - \varphi v_s)V + \ln[1 - D(1 - f)]$$

(99)

using a linear dependence of coefficients \(c, e, s, a, b, v\) upon \(\varphi\).

An alternative approach based on the polarities has been proposed by Ruiz-Angel et al. [211], who (a) replaced \(\ln k_0\) in Eq. (96) by \(p(E_{EN}^N - E_{EN}^B)\) and (b) assumed a linear relationship between \(pK\) and \(\varphi\) to obtain

$$\ln k = c_0 + c_1(E_{EN}^N - E_{EN}^B) + \ln[1 - (1 - f)]$$

(100)

Here, \(E_{EN}^N\) is considered to be constant and \(E_{EN}^B\) is given by expressions similar to Eq. (63).

Finally, Pou-Torres et al. [114] have extended Eq. (100) to include \(pH\), \(\varphi\), and \(T\) changes

$$\ln k = c_0 + \frac{c_1}{T} + c_2E_{EN}^N$$

$$+ \ln\left\{ c_3 + (1 - c_3) \frac{10^{pH+c_4+c_5+c_6c_7+c_8+c_9/c_7+c_9/c_8+c_9/c_8+c_9/c_8}}{1 + 10^{pH+c_4+c_5+c_6c_7+c_9/c_8+c_9/c_8+c_9/c_8}}\right\}$$

(101)

4.3. Empirical models

There is a great variety of empirical models for multi-variable expressions of \(\ln k\). A simple relationship that can be used to obtain some of these expressions is the following [47, 212]

$$\ln k(\varphi_1, \varphi_2, T, p\varphi) = \ln k(\varphi_1) \ln k(\varphi_2) \ln k(T) \ln k(p\varphi)$$

(102)

At this point we should stress that it is not the same to fit Eq. (102) to the experimental data before or after the multiplication. Consider the mathematical function

$$f(x, y) = (a_1 + a_2x)(b_1 + b_2y)$$

(103)

If we fit this equation to experimental data each of the coefficients \(a_1\) and \(a_2\) is, in fact, multiplied by the same quantity \(a_1 + a_2x\), since

$$f(x, y) = b_1(a_1 + a_2x) + b_2(a_1 + a_2x)y$$

(104)

However, if Eq. (103) is written as

$$f(x, y) = c_1 + c_2x + (c_3 + c_4y)$$

(105)

and then is fitted to experimental data, the above restriction does not hold and the applicability of Eq. (105) is higher than that of Eq. (103).

Below we apply Eq. (102) to ternary systems, and systems with variable \(\varphi\) and \(p\varphi\), \(\varphi\) and \(T\) or \(p\varphi\), \(p\varphi\) and \(T\).

4.3.1. Retention in ternary systems

For ternary systems Eq. (102) is written as

$$\ln k(\varphi_1, \varphi_2) = \ln k(\varphi_1) \ln k(\varphi_2)$$

(106)

We can distinguish at least two cases:

(a) if \(\ln k(\varphi)\) is given by a polynomial of order 2 and we want the final expression of \(\ln k(\varphi_1, \varphi_2)\) not to have terms of order higher than 2, then we obtain the familiar expression [57, 202]

$$\ln k(\varphi_1, \varphi_2) = c_0 + c_1\varphi_1 + c_2\varphi_2 + c_3\varphi_1\varphi_2 + c_4\varphi_1^2 + c_5\varphi_1\varphi_2^2 + c_6\varphi_1^2\varphi_2$$

(107)

(b) if \(\ln k(\varphi)\) is given by a rational function, say Eq. (24), then we have

$$\ln k(\varphi_1, \varphi_2) = c_0 - \frac{c_1\varphi_1}{1 + c_2\varphi_1} - \frac{c_3\varphi_2}{1 + c_4\varphi_2} + \frac{c_5\varphi_1\varphi_2}{(1 + c_2\varphi_1)(1 + c_4\varphi_2)}$$

(108)

4.3.2. Retention in systems of variable \(pH\) and mobile phase composition

Here the retention model for \(pH\) is expressed as \(k(p\varphi)\), see Eqs. (42) or (46), (47), whereas the retention model for \(\varphi\) as \(k(\varphi)\). For this reason we may write

$$k(\varphi, p\varphi) = k(\varphi)k(p\varphi) = e^{\ln k(\varphi)p\varphi}$$

(109)

Thus, if we combine the rational function of Eq. (24) with Eq. (46), we obtain

$$k = k_0 e^{-c_1\varphi/(1 + c_2\varphi)} + k_0 e^{-c_1\varphi/(1 + c_2\varphi)} 10^{pH-pK}$$

(110)

In this expression \(p\varphi\) may be either constant or variable with \(p\varphi\) through Eq. (83) or (84). If we use Eq. (84), we recover Eq. (85).

Several other expressions for \(k(\varphi, p\varphi)\) arise if we combine the various expressions of \(\ln k(\varphi)\) presented in this paper with Eq. (46) or (47).

4.3.3. Retention in systems of variable \(T\) and mobile phase composition

As in the previous cases, the final expression of \(\ln k(\varphi, T)\) depends on the choice of the functions \(\ln k(\varphi)\) and \(\ln k(T)\). Thus, from Eqs. (24) and (51) we readily obtain

$$\ln k(\varphi, T) = \left[ a_1 - \frac{c_1\varphi}{1 + b_1\varphi} \right] \left[ a_2 + \frac{b_2}{T} + c_2g(T) \right]$$

(111)

which yields

$$\ln k(\varphi, T) = c_0 - \frac{c_1\varphi}{1 + c_2\varphi} + \left( c_3 - \frac{c_4\varphi}{1 + c_2\varphi} \right) \frac{1}{T} + \left( c_5 - \frac{c_6\varphi}{1 + c_2\varphi} \right) g(T)$$

(112)

It is straightforward to show that this approach yields Eqs. (92) and (93) by combining the appropriate 1D equations.

4.3.4. Retention in systems of variable \(T\), \(p\varphi\) and mobile phase composition

We may start from

$$k(\varphi, p\varphi, T) = k(\varphi, T) k(p\varphi) = e^{\ln k(\varphi, T)p\varphi}$$

(113)

By substitution of \(\ln k(\varphi, T)\) from Eqs. (92), (93) or (94) into Eq. (46) and replacement of the various \(p\varphi\) from Eq. (95), we obtain expressions identical to those in Section 4.1.4.
4.3.5. Models used in factorial designs

Factorial designs are a class of experimental designs that are used to study multivariate response surfaces, \( y = f(x_1, x_2, \ldots, x_m) \) [213–222]. In liquid chromatography \( y = t_0 \) or \( \ln k \) and the variables \( x_1, x_2, \ldots, x_m \), which are called factors, may be any variable that affects \( t_0 \) or \( \ln k \), like mobile phase composition, \( pH \), \( T \) or gradient variables, like the steepness of a linear gradient, the initial value of \( \psi \), etc. The great advantage of factorial designs is that they offer a large amount of information using the minimum number of experiments.

The different values that a factor takes are known as levels. In two-levels factorial designs, each variable (factor) takes only two values. The lower value is indicated by \( -1 \) and the higher value by \( +1 \). Thus, all variables are scaled in the region \([−1, 1] \). Factorial designs are usually used to estimate the effect of all factors and their interactions on the response surface.

Typical retention models (response functions) when the response surface represents either \( t_0 \) or \( \ln k \) for two factors are the following

\[
y = c_0 + c_1 x_1 + c_2 x_2 + c_12 x_1 x_2
\]

or

\[
y = c_0 + c_1 x_1 + c_2 x_2 + c_11 x_1^2 + c_22 x_2^2 + c_12 x_1 x_2
\]

For three factors we similarly have

\[
y = c_0 + c_1 x_1 + c_2 x_2 + c_3 x_3 + c_{12} x_1 x_2 + c_{13} x_1 x_3 + c_{23} x_2 x_3 + c_{123} x_1 x_2 x_3
\]

or

\[
y = c_0 + c_1 x_1 + c_2 x_2 + c_3 x_3 + c_{11} x_1^2 + c_{22} x_2^2 + c_{33} x_3^2 + c_{12} x_1 x_2 + c_{13} x_1 x_3 + c_{23} x_2 x_3 + c_{123} x_1 x_2 x_3
\]

When Eqs. (114)–(117) are fitted to experimental data, we usually start with a general model, Eq. (115) or (117), and leave the software to select the statistically significant terms. In this way we can estimate whether a certain factor or interaction has a significant or not contribution to \( t_0 \) (or \( \ln k \)).

4.3.6. Models exclusively for gradient elution

All the above models can be used for retention prediction and separation optimization under either isocratic or gradient elution conditions. If \( t_0 \) or \( \ln k \) is expressed in terms of the variables representing the gradient elution conditions, like gradient steepness, initial mobile phase composition or temperature, the total gradient time, etc., then the retention model is applicable only to gradient prediction. In this case the retention model is usually empirical, arising from the experimental properties of the system.

For example, let us consider the following two cases:

I. Gradients of \( \psi \) at constant \( T \)

In this case we collect a series of 2D data, which are the elution times of an analyte at various \( \psi \)-gradients at constant \( T \) values. All \( \psi \)-gradients should exhibit the same initial and final value of \( \psi \) but different gradient steepness \( b \). These data are likely to be fitted to the following retention models

\[
t_R = c_0 + c_1 T + (c_2 + c_3 T)b + (c_4 + c_5 T)b^2
\]

or

\[
t_R = c_0 + c_1 T + (c_2 + c_3 T)b
\]

where the \( \psi \)-gradients are of the same type with those used in the fitting procedure and the values of \( b \) and \( T \) lie within those used for fitting.

II. Gradients of \( T \) at constant \( \psi \)

To study these gradients we should have a series of 2D data of the elution times of an analyte at various \( T \) and \( b \) values, \( b \) being the slope of \( T \)-gradients with the same initial and final temperature. The retention models used may be the following

\[
\ln k = c_0 + c_1 \psi + c_2 \psi^2 + (c_3 + c_4 \psi + c_5 \psi^2) b + (c_6 + c_7 \psi + c_8 \psi^2) b^2
\]

or

\[
\ln k = c_0 + c_1 \psi + c_2 \psi^2 + (c_3 + c_4 \psi + c_5 \psi^2) b
\]

or

\[
\ln k = c_0 + c_1 \psi + (c_2 + c_3 \psi) b
\]

It is evident that these equations can be used for prediction under prerequisites similar to those for Eqs. (118) and (119). An application of all these equations is given in Ref. [123].

5. Fitting data to retention models

The determination of the adjustable parameters of a retention model is always attained by fitting to either isocratic or gradient data. In both cases the cost function for fitting may be either

\[
CF = \sum_{j=1}^{N} (t_{R,j, \text{exp}} - t_{R,j, \text{calc}})^2
\]

or

\[
CF = \sum_{j=1}^{N} (\ln k_{j, \text{exp}} - \ln k_{j, \text{calc}})^2
\]

where \( t_{R,j, \text{exp}} \) and \( \ln k_{j, \text{exp}} \) are the experimental retention time and its retention factor of a certain analyte under the \( j \)-th measurement and \( t_{R,j, \text{calc}} \) and \( \ln k_{j, \text{calc}} \) are the corresponding values calculated from the retention model. Note that Torres–Lapasió et al. [224] suggested the use of weighted least squares when we fit \( \ln k \) data with weights calculated from

\[
w = \frac{1}{(\partial \ln k / \partial k)^2} = k^2
\]

The theoretical basis of weighted regression is that the use of weights in Eq. (124) makes it equivalent to Eq. (123).

The two above expressions of the cost function do not give always practical converging results unless the \( \psi \) range is narrow. When this prerequisite is not fulfilled, then the use of Eq. (123) reduces the error in the high \( t_R \) values but increases it in the small values of \( t_R \). Quite the opposite is the behaviour of Eq. (124). This is clearly shown in Fig. 10. This figure shows the percentage absolute error between experimental and calculated retention times of Benzene in \( \text{iPrOH} \) (A) and \( \text{MeCN} \) (B) as a function of the mobile phase composition in \( \psi \). The calculated retention times obtained by means of Eq. (30) using the objective function of Eq. (123) (○), and Eq. (124) with (+) and without (•) weighting factors from Eq. (125). We observe that the use of weighting factors is indeed equivalent to using Eq. (123) and that the use of Eq. (123) or the use of weighting factors has as a result small errors at low \( \psi \) values, which means great retention times, and great errors at high \( \psi \) values, i.e. at low retention times. The use of Eq. (124) seems to have a more balanced behaviour. Finally, it is interesting to note that the error between experimental and calculated retention times may increase
considerably if we use weighting factors to fit exo-thermodynamic relationships. In this case the error may be unacceptably high (Fig. 11).

The fitting to isocratic data is usually easy and can be performed using commercial software, like the Regression and Solver programs of Excel. The fitting to gradient data may require homemade software for non-linear fitting [225]. Previous studies [226] have shown that the fitting problem can be solved using algorithms based on the Levenberg–Marquardt method [227–229]. However, the direct application of the Levenberg–Marquardt (LM) algorithm exhibits several problems, the most important of which is the trapping of the algorithm into local minima other than the global one. For this reason, in recent papers [225,230] we proposed two variations of this algorithm denoted by RND_LM and R_LM, respectively.

The RND_LM modification of the LM algorithm arises if we combine LM with an initial random search of the domain adopted for the adjustable parameters. Thus, a number of vectors \((c_0, c_1, c_2, \ldots)\) are randomly selected from the search domain and the vector that yields the minimum value of the objective function, Eq. (123) or (124), is selected and its coordinates \(c_0, c_1, c_2, \ldots\) are used as initial estimates in the LM algorithm, which determines the minimum of the cost function. In the second modification of the LM algorithm denoted by RLM, the LM algorithm is repeated several times. In particular, the algorithm starts with an initial vector \((c_0, c_1, c_2, \ldots)\) randomly selected from the search domain and the LM method using a small number of iterations (usually 100) determines the local minimum, which is stored. Then a new initial vector \((c_0, c_1, c_2, \ldots)\) is randomly selected and the whole algorithm is repeated for a preset number of iterations. The minimum of the stored local minima is determined and it presumably corresponds to the global minimum of the cost function. The iterative use of the LM method in the above algorithms is expected to reduce convergence problems as well as problems associated with trapping into local minima. Apart from the LM based algorithms, genetic algorithms (GA) [231,232] as well as algorithms based on the descent method (D) [229,233] have also been used to solve fitting problems [234,235]. However, the LM based algorithms are still a good tool for solving fitting problems since they are computationally simple and fast.

In general, although the fitting procedure of isocratic data to a retention model is much easier than the fitting of gradient data, initial isocratic experiments required for fitting isocratic data are more laborious and time-consuming than gradient experiments in case gradient data are used for fitting. However, isocratic experiments give accurate information about retention in an extended retention range and, thus, the determination of the adjustable parameters of a retention model from isocratic data can be used to predict reliably retention in isocratic runs as well as in gradient runs. In contrast, isocratic predictions from gradient data are almost impossible by using most of the retention models since gradient experiments give accurate information about retention only in narrow solvent concentration ranges [236], whereas gradient prediction from gradient data exhibits the following advantages: (1) Less experimental effort. (2) Reduction of predictive errors related to non-equilibration of column [237] due to the use of similar experimental conditions. (3) Reduction of predictive errors related to the retention model. The distribution of the retention times of a sample solute under the gradients used for the calculation of the coefficients of a retention model is much narrower than that if we used isocratic data, and this property reduces errors introduced by the retention model itself. However, a basic prerequisite to obtain high quality gradient predictions from gradient data is the following. The predicted value of the retention time of each solute should lie within the corresponding values of the gradient data used to determine the retention model. If the above prerequisite is not valid, we should add more initial gradient data or replace some gradient data by proper isocratic ones in order to determine the retention model [226].

![Fig. 10. Percentage absolute error between experimental and calculated retention times of benzene derivatives in MeOH at C18 column as a function of the entire data set. Retention times were calculated from Eq. (30) using the objective function of Eq. (123) (○), and Eq. (124) with (+) and without (●) weighting factors from Eq. (125).](Image 352x581 to 587x773)

![Fig. 11. Percentage absolute error between experimental and calculated retention times of benzene derivatives in MeOH at C18 column as a function of the entire data set. Retention times were calculated from Eq. (56) using the objective function of Eq. (124) with (○) and without (●) weighting factors from Eq. (125).](Image 86x423 to 315x773)
Nomenclature

\( a \) as superscript it denotes the adsorbed layer
\( a, b, c, e, s, v \) coefficients of Eq. (55)
\( a_0, a_1, a_{2,2} \) coefficients of \( T_{ov} \) and \( T_{ef} \), Eqs. (53) and (54)
\( a_j, b_j, c_j, e_j, s_j, v_j \) \( (j = 0, 1, 2) \) coefficients of Eq. (56)
\( a_A, b_A, m_A, s_A \) coefficients of Eq. (60)
\( a_A', b_A', s_A' \) coefficients of Eq. (61)
\( A \) analyte molecule
\( A, B, E, S, V \) analyte descriptors in Eq. (55)
\( b \) slope of a linear gradient
\( b_E \) coefficient of Eqs. (63), (64) and (67)
\( b_s \) coefficient of Eq. (65)
\( b_{0,0}, b_{1,0}, \ldots \{ \) coefficients of Eq. (89)
\( b_{0,0}' \) \( \rightarrow \) coefficients of Eqs. (65) and (67)
\( B, C \) organic modifier molecules
\( c_{0}, c_{1}, c_{2, \ldots} \) coefficients of \( \ln \) \( K \) expressions
\( c_{B}, c_{C} \) (i = 1, 2, 3) coefficients of Eqs. (77)–(79)
\( C_j \) \( (j = a, m, s) \) coefficient of Eqs. (17) and (18)
\( D \) constant of Born’s Eq. (82) as well as the degree of ionization of the analyte
\( D_j, D_j' \) \( (j = a, m, s) \) coefficients of Eqs. (15), (16) and (73)
\( C_j, C_{2, \ldots} (j = a, m, s) \) coefficients of Eqs. (15)–(18), (73) and (74)
\( e_0, e_1, e_2 \) coefficients of \( pK \) expression
\( E_N^S \) solvatochromatic parameter
\( E_N^W \) solvatochromatic parameter of the stationary phase
\( E_r(30) \) non-normalised solvatochromatic parameter
\( f \) \( k_1/k_0 \) in Eqs. (96), (98)–(100)
\( f_A^{m, f_A'} \) activity coefficients of \( A \) in the mobile phase, the adsorbed layer and within the alkyl chains of the stationary phase, respectively
\( g \) \( =k_0^2 \beta^\theta |k^a| \) a function of \( T \) defined from Eq. (50)
\( g_1, g_2, \ldots \) variables representing gradient elution conditions
\( \Delta G^\theta \) standard free energy of transfer of an analyte from the mobile to the stationary phase
\( h \) hysteresis constant
\( \Delta H^\theta \) standard enthalpy of the retention process
\( \Delta H_{0}, \Delta H_{1} \) constants, coefficients of \( \Delta H^\theta \)
\( k \) retention factor of the sample solute
\( k_j \) retention factor of species \( j \)
\( k^a, k^t \) retention factor due to adsorption (\( k^a \)) or to partition (\( k^t \)) retention process
\( k_{0}^0, k_{1}^0 \) coefficients of Eqs. (85), (87) and (110)
\( k_{t}^0 \) equilibrium constant of Eq. (37)
\( \log P_{0/NW} \) octanol–water partition coefficients
\( m \) as superscript it denotes the mobile phase
\( n \) the number of water molecules displaced by one modifier molecule
\( n_A^m, n_A^s, n_A^r \) number of the analyte moles in the mobile phase, the adsorbed layer and within the alkyl chains of the stationary phase, respectively
\( p_0, p_1, p_2 \) coefficients of Eq. (59)
\( p_E \) coefficient of Eq. (63)
\( p_s \) coefficient of Eq. (65)
\( p_{Kp} \) constant of Born’s Eq. (82)
\( pK_0 \) value of \( pK \) when \( \psi = 0 \)

References
