Applicability of retention modelling in hydrophilic-interaction liquid chromatography for algorithmic optimization programs with gradient-scanning techniques

Pirok, B.W.J.; Molenaar, S.R.A.; van Outersterp, R.E.; Schoenmakers, P.J.

Published in:
Journal of Chromatography A

DOI:
10.1016/j.chroma.2017.11.017

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Applicability of retention modelling in hydrophilic-interaction liquid chromatography for algorithmic optimization programs with gradient-scanning techniques

Bob W.J. Pirok\textsuperscript{a,b,*}, Stef R.A. Molenaar\textsuperscript{a}, Rianne E. van Outersterp\textsuperscript{a}, Peter J. Schoenmakers\textsuperscript{a}

\textsuperscript{a} University of Amsterdam, van ’t Hoff Institute for Molecular Sciences, Analytical-Chemistry Group, Science Park 904, 1098 XH, Amsterdam, The Netherlands
\textsuperscript{b} TI-COAST, Science Park 904, 1098 XH, Amsterdam, The Netherlands

ABSTRACT

Computer-aided method-development programs require accurate models to describe retention and to make predictions based on a limited number of scouting gradients. The applicability of five different retention models for hydrophilic-interaction chromatography (HILIC) is assessed for a wide range of analytes. Gradient-elution equations are presented for each model, using Simpson’s Rule to approximate the integral in case no exact solution exists. For most compound classes the adsorption model, i.e. a linear relation between the logarithm of the retention factor and the logarithm of the composition, is found to provide the most robust performance. Prediction accuracies depended on analyte class, with peptide retention being predicted least accurately, and on the stationary phase, with better results for a diol column than for an amide column. The two-parameter adsorption model is also attractive, because it can be used with good results using only two scanning gradients. This model is recommended as the first-choice model for describing and predicting HILIC retention data, because of its accuracy and linearity. Other models (linear solvent-strength model, mixed-mode model) should only be considered after validating their applicability in specific cases.

© 2017 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Hydrophilic interaction chromatography (HILIC) has become increasingly important for the analysis of highly polar analytes, such as antioxidants [1], sugars (e.g. glycosics [2–4]), (plant) metabolites [5–7], foodstuffs [8], and environmental pollutants [9]. The exact mechanism of retention in HILIC has been intensively investigated and it is thought to be rather complex. The currently accepted mechanism is a combination of (i) partitioning processes of the analytes between a water-poor organic mobile phase and a water-rich layer absorbed on a polar stationary-phase material [10], and (ii) electrostatic interactions between the analytes and the stationary-phase surface [11]. Therefore, HILIC can best be described as a mixed-mode retention mechanism.

To describe retention in HILIC, several retention models have been investigated. The model most commonly used in reversed-phase LC (RPLC) involves a linear relationship between the logarithm of the retention factor ($k$) and the volume fraction of strong solvent ($\varphi$). When a linear gradient is used in RPLC this results in so-called linear-solvent-strength conditions [12]. Already in 1979, LSS conditions have been studied and described in detail by Snyder et al. [12] and equations were also derived for situations in which analytes elute before the gradient commences or after it has been completed [12,13]. However, due to the mixed-mode retention mechanism, this linear model may be less suitable to accurately model retention in HILIC.

To describe retention more accurately across a wider $\varphi$-range, Schoenmakers et al. introduced a quadratic model [13], including relations for the retention factor for analytes eluting within and after a gradient. However, an error function was required to allow partial integration of the gradient equation. This is an impractical aspect of the relationship. Moreover, the model may show deviations from the real values when predicting outside the scanning range. An empirical model proposed by Neu and Kuss circumvented the integration problems, allowing analytical expressions to
be obtained for retention under gradient conditions [14]. A retention model based on surface adsorption has also been proposed for HILIC, predicting retention across narrow ranges of water concentrations in the eluent. To account for the observed mixed-mode behaviour found in HILIC, Jin et al. introduced a three-parameter model [15], which was found to precisely describe retention factors in isocratic mode [16]. However, similar to the quadratic model, integration of the corresponding gradient equation was significantly complicated, involving a gamma function and potentially yielding complex numbers.

For efficient method development, an underlying accurate description of the retention mechanism is crucial. Because gradient elution is an essential tool for analysing or scanning samples, accurate description of gradient-elution patterns is also essential. Computer-aided-optimization tools, such as Drylab [17] for 1D-LC, or PIOTR [18] for 1D and 2D-LC, utilize the concept of so-called “scouting” or “scanning” runs to establish retention parameters [19], from which the optimal conditions and the optimal chromatogram can be predicted. Method development in HILIC following these principles has extensively been studied by Tytėca et al. [20,21]. However, the currently employed retention models for HILIC do not allow accurate prediction of retention times of analytes eluting during or after the completion of gradients based on a very limited number of scouting measurements. This hampers the application of such optimization tools for HILIC.

In this work, we present the results of an evaluation study of each of the five above-listed models for predicting retention times in gradient-elution HILIC based on a limited number of scouting runs for a wide range of applications. First, the equation for each retention model is addressed in the context of gradient-elution chromatography. We use Simpson’s Rule [22] to approximate the integration of resulting gradient equations when an exact solution does not exist, i.e., for the quadratic and mixed-mode models. The performance of each of these models in computer-aided method-development programs is assessed.

2. Theory

In the case that a solute elutes before the start of the gradient, the retention time \( t_{R, \text{before}} \) can be calculated from

\[
t_{R, \text{before}} = t_0 (1 + k_{\text{init}})
\]

where \( k_{\text{init}} \) depicts the retention factor at the start of the gradient and \( t_0 \) the column dead time. The general equation of linear gradients allows calculation of the retention time if a compound elutes during the gradient

\[
\frac{1}{B} \int_{\psi_{\text{init}}}^{\psi_{\text{init}} + B(\psi - \tau)} \frac{d\psi}{k(\psi)} = t_0 - \frac{t_{\text{init}} + t_D}{k_{\text{init}}}
\]

(2)

In this equation \( k(\psi) \) is the retention model, denoting the variation of the retention factor \( k \) with the composition parameter \( \psi \). The change in \( \psi \) as a function of time (i.e. the slope of the gradient) is depicted \( B\psi = \psi_{\text{init}} + B\tau \) and \( \tau \) is the sum of the system dwell time \( t_D \), the waiting time before the gradient is programmed to start \( t_{\text{init}} \), and \( t_0 \) (\( \tau = t_D + t_{\text{init}} + t_0 \)). For useful application of gradient-elution retention prediction models in real cases, it is essential that the retention time cannot only be established if the analyte elutes during the gradient, but also if it elutes after the gradient is completed. In this case, the retention is obtained by integrating the retention model in the following equation

\[
\frac{1}{B} \int_{\psi_{\text{init}}}^{\psi_{\text{final}}} \frac{d\psi}{k(\psi)} = \frac{t_D - \tau - t_0}{k_{\text{final}}} = t_0 - \frac{t_{\text{init}} + t_D}{k_{\text{init}}}
\]

(3)

Here, \( t_c \) represents the gradient time. The application of some of the proposed HILIC retention models is complicated, because the integrals in Eqs. (2) and (3) cannot be analytically solved. The application of each of the HILIC models for gradient-elution separations is the main topic of this paper and this will be described in detail in the following sections.

2.1. Exponential model

In the exponential model (Eq. (4)), \( k_0 \) accounts for the extrapolated retention factor for \( \psi = 0 \) and \( S \) denotes the change in the retention factor with increasing mobile phase strength.

\[ \ln k = \ln k_0 - S\psi \]

(4)

This equation is often referred to as the linear-solvent-strength (LSS) equations, because it corresponds to LSS conditions in combination with linear gradients (\( \psi = \psi_{\text{init}} + B\tau \)). Schoenmakers et al. derived equations for a compound eluting during (\( t_{R, \text{gradient}} \)) and after (\( t_{R, \text{after}} \)) the gradient [13].

\[
t_{R, \text{gradient}} = \frac{1}{SB} \ln \left\{ 1 + SB \cdot k_{\text{init}} \right\} \left( t_0 - \frac{t_{\text{init}} + t_D}{k_{\text{init}}} \right) + \tau
\]

(5)

\[
t_{R, \text{after}} = k_{\text{final}} \left( t_0 - \frac{t_{\text{init}} + t_D}{k_{\text{init}}} \right) - \frac{1}{BS} \left( \frac{1}{t_{\text{final}}} - \frac{1}{t_{\text{init}}} \right) + t_G + \tau
\]

(6)

Here, \( k_{\text{final}} \) represents the retention factor at the end of the gradient and \( t_G \) the duration of the gradient.

2.2. Neue-Kuss empirical model

The empirical model introduced by Neue and Kuss [14] is given by

\[ \ln k = \ln k_0 + 2 \ln(1 + S_2\psi) - \frac{S_1\psi}{1 + S_2\psi} \]

(7)

where the coefficients \( S_1 \) and \( S_2 \) represent the slope and curvature of the equation, respectively. Integration of the gradient equation yields

\[
t_{R, \text{gradient}} = \frac{\ln F}{B(2S_2 - S_1\ln F)} - \frac{\psi_{\text{init}}}{B} + \tau
\]

(8)

with \( F \) defined as

\[ F = S_2Bk_0 \left( t_0 - \frac{t_{\text{init}} + t_D}{k_{\text{init}}} \right) + e^{\frac{S_2\psi_0}{S_1}} \]

(9)

Similarly, introducing Eq. (7) into Eq. (3) and rewriting yields

\[
t_{R, \text{after}} = k_{\text{final}} \left( t_0 - \frac{t_{\text{init}} + t_D}{k_{\text{init}}} \right) + S_2Bk_0 \left( e^{\frac{S_2\psi_{\text{init}}}{S_1}} - e^{\frac{S_2\psi_{\text{final}}}{S_1}} \right) + t_G + \tau
\]

(10)

2.3. Adsorption model

The adsorption model is based on confined surface adsorption as used in normal-phase chromatography and is given by

\[ \ln k = \ln k_0 - n \ln \psi \]

(11)

where \( n \) depicts the ratio of surface areas occupied by a water and a solute molecule [16]. In the events that the compound elutes during or after the gradient retention can be calculated from [23]

\[
t_{R, \text{gradient}} = \left[ k_0 \left( t_0 - \frac{t_{\text{init}} + t_D}{k_{\text{init}}} \right) B(n + 1) + \frac{\psi_{\text{init}}}{B} \right]^\frac{1}{n}
\]

(12)
\[ t_{R,\text{after}} = k_{\text{final}} \left( t_0 - \frac{f_{\text{init}} + f_D}{k_{\text{init}}} \right) \]

\[ - \frac{k_{\text{final}}}{BK_0(n+1)} \left[ \varphi_{\text{final}}^{n+1} - \varphi_{\text{init}}^{n+1} \right] + t_G + \tau \]

(13)

2.4. Mixed-mode model

The mixed-mode aspect of HILIC led Jin et al. to propose a tailored model [15]

\[ \ln k = \ln k_0 + S_1 \varphi + S_2 \ln \varphi \]

(14)

where \( S_1 \) is said to account for the interaction of solutes with the stationary phase and \( S_2 \) for the interaction of solutes with solvents. While seemingly attractive because of its ability to account for the mixed-mode character of HILIC retention, the relation does pose a practical problem upon integration of the gradient equation.

The calculation of the retention time using Eq. (14), requires a gamma function and may possibly result in complex numbers. To circumvent this, we apply Simpsons’ Rule [22] to approximate the integral (Eq. (15)).

\[ \text{Simpsons} \left\{ \Phi_{\text{final}} + \Phi_{\text{gradient} - \varphi} \right\} = BK_0 \left( t_0 - \frac{f_{\text{init}} + f_D}{k_{\text{init}}} \right) \]

(15)

This rule divides the integral over a function, \( f \), in our case \( f(\varphi) = e^{\psi \varphi} \cdot e^{\psi^2} \) in \( m \) segments of width \( \Delta \varphi \). Subsequently, each segment is approximated by a solvable quadratic equation such that, in essence, the complex integral is replaced by \( m \) solvable quadratic equations (Eq. (16)).

\[ \int_{\varphi_{\text{init}}}^{\varphi_{\text{final}}} f(\varphi) d\varphi \approx \frac{\Delta \varphi}{3} \left[ f(\varphi_{\text{init}}) + 4f \left( \frac{1}{m} \Delta \varphi + \varphi_{\text{init}} \right) + 2f \left( \frac{2}{m} \Delta \varphi + \varphi_{\text{init}} \right) + \ldots + 2f \left( \frac{m-2}{m} \Delta \varphi + \varphi_{\text{init}} \right) + f(\varphi_{\text{final}}) \right] \]

(16)

Of course, the approximation is accompanied by an error. With the Simpsons’ Rule, the maximum error depends on \( m \) and can be calculated from

\[ |E| \leq \frac{DX_{\text{final}} - X_{\text{init}}}{180m^4} \]

(17)

where \( D \) represents the maximum value of the fourth derivative of the retention model in the integration range from \( \varphi_{\text{init}} \) to \( \varphi_{\text{final}} \). In this study, we set the acceptable calculation error to 0.001, which is much smaller than the typical experimental error. Rewriting the equation yields

\[ m \geq 4 \sqrt{\frac{X_{\text{final}} - X_{\text{init}}}{0.18}} \]

(18)

In the case of the mixed-mode model, \( \varphi_x \) equals \( \varphi_{\text{init}} \). Likewise, the mixed-mode retention model can be applied in conjunction with Eq. (3), resulting in

\[ t_{R,\text{after}} = k_{\text{final}} \left( t_0 - \frac{f_{\text{init}} + f_D}{k_{\text{init}}} \right) \]

\[ - \frac{k_{\text{final}}}{BK_0(n+1)} \left[ \varphi_{\text{final}}^{n+1} - \varphi_{\text{init}}^{n+1} \right] + t_G + \tau \]

(19)

2.5. Quadratic model

In comparison with the exponential (“LSS”) model, the quadratic model offers a more flexible solution across a broader range of the volume fraction of the strong solvent. Retention is given by

\[ \ln k = \ln k_0 + S_1 \varphi + S_2 \varphi^2 \]

(20)

Where \( S_1 \) and \( S_2 \) represent the influences of the volume fraction of strong solvent. Similar to the mixed-mode model, integration of the gradient-elution equation is complex. Effectively, the gradient equations become

\[ \text{Simpsons} \left\{ \Phi_{\text{final}} + B_{\text{gradient}_1} \varphi \right\} = BK_0 \left( t_0 - \frac{f_{\text{init}} + f_D}{k_{\text{init}}} \right) \]

(21)

\[ t_{R,\text{after}} = k_{\text{final}} \left( t_0 - \frac{f_{\text{init}} + f_D}{k_{\text{init}}} \right) - \frac{1}{BK_0} \]

\[ \text{Simpsons} \left\{ \Phi_{\text{final}} + B_{\text{gradient}_2} \varphi^2 \right\} + t_G + \tau \]

(22)

For calculating the number of partitions \( m \) for the Simpson’s approximation from (Eq. (18)), \( |2\varphi_x| \) now equals \( \varphi_{\text{final}} \), because the fourth derivative is largest at the final solvent strength.

3. Experimental

3.1. Chemicals

Aqueous solutions were prepared using deionized water (Arium 6111UV; Satorius, Germany; R=18.2 MΩ cm). Acetonitrile (ACN, LC–MS grade) was obtained from Avantor Performance Chemicals (Deventer, The Netherlands). Ammonium formate, formic acid (reagent grade, ≥95%) and the peptide mix (HPLC peptide standard mixture, H2016) were obtained from Sigma-Aldrich (Darmstadt, Germany). The earl-grey tea was obtained from Maas International (Eindhoven, The Netherlands). The dyes used in this study were authentic dyes obtained from the reference collection of the Cultural Heritage Agency of the Netherlands (RCE, Amsterdam, The Netherlands).

3.2. Instrumental

All experiments were carried out on an Agilent 1100 LC system equipped with a quaternary pump (G1311A), an autosampler (G1313A), a column oven (G1316A) and a 1290 Infinity diode-array detector (G4212A) (Agilent, Waldbronn, Germany). In front of the column, an Agilent 1290 Infinity In-Line Filter (G5067–4638) was installed to protect the column. The dwell volume was approximately 1.1 mL. The injector needle was set to draw and eject at a speed of 10 μL·min⁻¹ with two seconds equilibration time. Two columns were used, a Waters Acquity BEH Amide (150 × 2.1 mm i.d., 1.7-μm particles, 130-Å pore size; Waters, Milford, MA, USA) and a Phenomenex Luna HILIC column (50 × 3 mm, 3 μm, 200 Å; Phenomenex, Torrance, CA, USA), further referred to as diol column.

3.3. Procedures

3.3.1. Analytical methods

The mobile phase consisted of acetonitrile/buffer [v/v] 97:3 (Mobile phase A) and acetonitrile/buffer [v/v] 1:1 (Mobile phase
The buffer was 10 mM ammonium formate at pH 3. For all experiments recorded on the amide column, five different scouting gradient programs were used. All gradients started from 0.0 until 0.5 min isocratic at 100% A, followed by a linear gradient from 100% A to 100% B in 10 (Gradient A1), 17 (Gradient A2), 30 (Gradient A3), 52 (Gradient A4) and 90 (Gradient A5) minutes. In all cases, 100% B was maintained for 2 min, after which a linear gradient of 1 min brought the mobile phase back to the initial composition of 100% A, which was maintained for 20 min to thoroughly re-equilibrate the column. The flow rate was 0.25 mL min⁻¹.

For the experiments carried out on the diol column, the five different scout gradients each started with 0.0–0.5 min isocratic at 100% A, followed by a linear gradient from 100% A–100% B in 0.67 (Gradient D1), 1.33 (Gradient D2), 2.0 (Gradient D3), 4.0 (Gradient D4) and 6.0 (Gradient D5) minutes. The flow rate was 1.0 mL min⁻¹.

3.3.2. Sample preparation

The peptide mix was used at a concentration of 1000 ppm in deionized water, and was diluted five times with ACN before analysis. The injection volume was 5 μL.

The metabolites were individually prepared as stock solutions in ACN/water with ratios (varying between 9:1 or 8:2 [v/v]) and resulting concentrations depending on the solubility of each compound (see Supplementary Material section S-1). For analysis, mixtures of seven or eight metabolites were prepared with effective metabolite concentrations in these mixtures between 100 and 500 ppm. The injection volume of each mixture was 1 μL. The peaks in the chromatograms were identified using individual injections, clearly distinguishable peak patterns and UV–vis spectra.

The dyes were prepared as individual stock solutions of approximately 5000 ppm in water/MeOH (1:1) [v/v]. Mixtures of 20 dyes were prepared and the resulting solution was diluted four times in ACN. The mixtures were injected at a volume of 3 μL with effective individual dye concentrations of approximately 25 ppm. The peaks in the chromatograms were identified using the UV–vis spectra.

For the preparation of the earl-grey-tea sample, 200 mL of deionized water was heated until the boiling point was reached. The heating was turned off and the earl-grey-tea bag was submersed in the water for 5 min. After cooling, the resulting solution was diluted five times with ACN prior to analysis. The injection volume was 5 μL.

4. Results and discussion

4.1. Goodness-of-fit

To establish a good representation of HILIC behaviour, a set of 57 analytes was compiled comprising organic acids, peptides, synthetic and natural dyes and components found in black tea (Table 1). The set mainly comprises acidic, basic, zwitterionic and neutral compounds with varying polarity and, in most cases, aromatic functionality. The retention times for all analytes on the amide column were recorded with five different gradient programs (see Experimental section, gradients A1–A5). Using the obtained retention times (see Supplementary Material section S-2 for all retention time data), all retention parameters were determined for each retention model by using the nonlinear-programming-solver fminsearch function of MATLAB to fit the retention model to the data. The resulting retention parameters are depicted in Table 1. For most analytes, reasonable values were found. Peptide component 1 exhibited no retention and thus was excluded from this study.

To assess the ability of the different models to describe the retention behaviour, the Akaike Information Criterion (AIC) [24] was used. This criterion provides a measure for the mean-squared error in describing the retention behaviour based on information theory and it can be regarded as a goodness-of-fit indicator. The criterion has been used before to study retention behaviour in HILIC [16] and partition coefficients [25]. One attractive feature of the AIC criterion is that it allows comparison of models with different numbers of terms. For example, a three-parameter model is expected to describe the data better than a two-parameter model. The AIC criterion accounts for this by penalizing the model for each additionally employed parameter, allowing an unbiased comparison. For calculation the AIC, the number of parameters, p, the number of experiments, n, and the sum-of-squares error (SSE) from the fit of the retention model are used.

\[
\text{AIC} = 2p + n \left[ \ln \left( \frac{\text{SSE}}{n} \right) + 1 \right]
\]  
(23)

Table 1 lists the AIC values obtained. Lower AIC values generally indicate a better description of the retention behaviour. Generally, the adsorption and Neue-Kuss models show favourable AIC values. In the case of the Neue-Kuss model, this is in agreement with previous studies [16]. This is also reflected in Fig. 1, where the number of found compounds with a given AIC value are grouped in a histogram per retention model. It can be seen that the LSS model performs poorly in most cases and that both the mixed-mode and quadratic models are not robust across the range of compounds.

Conversely, the plot suggests that the adsorption and Neue-Kuss models perform more reliably, in particular for the metabolites and dyes (see Supplementary Material section S-3 for histograms per class of analytes).

4.2. Prediction errors

With this insight, the ability of the models to reliably predict retention times based on a smaller number of gradients was studied to allow automatic gradient optimization. To simulate such a situation, the retention times from gradients A2 and A4 were excluded, and the retention parameters were determined once more using exclusively the data from gradients A1, A3 and A5. The obtained retention parameters were used to predict the retention times for gradients A2 and A4.

The results are plotted in Fig. 2A for the amide column, where the errors in prediction of the retention times of all analytes in gradient A2 are reflected by the solid bars and the dashed bars reflect the prediction errors for gradient A4. Most models perform very similarly, with an average prediction error of approximately 2% for retention times of gradient A2 and 2.5% for retention times of gradient A4. A significant deviation of this impression is the performance of the Neue-Kuss model, which is surprising because of the low AIC values found for this model. This can be explained by realizing that this is an empirical model now using just three points (A1, A3 and A5) instead of all five such, was the case for Fig. 1. To investigate to which extent the observations obtained thus far also pertain to other stationary phases, the retention times for 17 of the analytes (peptides and metabolites) were also recorded on a cross-linked diol column, using a similar set of scanning gradients (see Supplementary Material section S-4 for the obtained retention times, determined retention parameters and the calculated AIC values; see Section 3.3.1 for gradient programs and conditions). Again, using exclusively data from gradients D1, D3 and D5, the retention times of gradient D2 and D4 were predicted and the prediction errors are shown in Fig. 2B. Interestingly, the prediction errors across the range of models is much better with, with prediction error values of approximately 0.5% for D2 and 1.0% for D4. Noteworthy is, however, the dramatic performance of the Neue-Kuss model for which the box-and-whisker plots are off-scale. The lower prediction errors suggest a more-regular behaviour of the diol column. It is worth to point out that the obtained retention factors for both columns were rather similar. To rule out that this significantly better performance
<table>
<thead>
<tr>
<th>Compound</th>
<th>LSS</th>
<th>Mixed-Mode</th>
<th>Quadratic</th>
<th>Adsorption</th>
<th>Neure-Kuss</th>
</tr>
</thead>
<tbody>
<tr>
<td>tea component 1</td>
<td>3.322</td>
<td>18.069</td>
<td>7.386</td>
<td>0.356</td>
<td>9.400</td>
</tr>
<tr>
<td>tea component 2</td>
<td>3.573</td>
<td>9.612</td>
<td>15.627</td>
<td>9.059</td>
<td>2.000</td>
</tr>
<tr>
<td>tea component 3</td>
<td>3.573</td>
<td>7.508</td>
<td>10.482</td>
<td>7.726</td>
<td>1.980</td>
</tr>
<tr>
<td>tea component 4</td>
<td>4.686</td>
<td>17.599</td>
<td>11.173</td>
<td>8.072</td>
<td>1.982</td>
</tr>
<tr>
<td>tea component 5</td>
<td>5.251</td>
<td>12.932</td>
<td>15.366</td>
<td>0.000</td>
<td>2.390</td>
</tr>
<tr>
<td>tea component 6</td>
<td>5.312</td>
<td>12.854</td>
<td>15.366</td>
<td>0.000</td>
<td>2.060</td>
</tr>
<tr>
<td>tea component 7</td>
<td>4.247</td>
<td>17.496</td>
<td>12.144</td>
<td>4.202</td>
<td>4.048</td>
</tr>
<tr>
<td>tea component 9</td>
<td>3.573</td>
<td>12.069</td>
<td>11.173</td>
<td>8.072</td>
<td>1.980</td>
</tr>
<tr>
<td>tea component 10</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>all</td>
<td>3.573</td>
<td>9.612</td>
<td>15.627</td>
<td>9.059</td>
<td>2.000</td>
</tr>
<tr>
<td>Mixed-Mode</td>
<td>15.627</td>
<td>11.173</td>
<td>20.000</td>
<td>11.173</td>
<td>0.000</td>
</tr>
<tr>
<td>Quadratic</td>
<td>7.386</td>
<td>10.482</td>
<td>7.726</td>
<td>7.726</td>
<td>0.000</td>
</tr>
<tr>
<td>Adsorption</td>
<td>0.356</td>
<td>7.726</td>
<td>0.000</td>
<td>7.726</td>
<td>0.000</td>
</tr>
<tr>
<td>Neure-Kuss</td>
<td>9.400</td>
<td>2.000</td>
<td>1.980</td>
<td>2.000</td>
<td>1.980</td>
</tr>
</tbody>
</table>

**Table of determined retention parameters and calculated AIC values for all 57 analytes and five studied retention models based on five scanning gradients.**
Fig. 1. Quality of fit of 56 retained analytes based on the obtained AIC values for each retention model. The analytes were classified within distinct ranges of AIC values for clarity. See Supplementary Material section S-3 for quality-of-fit histograms per analyte class.

Fig. 2. Box-and-whisker plot showing the errors of prediction for (A) 56 analytes on the amide column, (B) 17 analytes on the diol column (C) 17 analytes on the amide and diol column. For each column/model combination a plot is provided for the prediction errors of both the second and fourth gradient (A2 and A4 for amide and D2 and D4 for the diol). Plot C reflects results exclusively for gradient 2 for both columns. Note the different scales on the y-axes.

of specific models is caused by the absence of the dyes and/or tea components in the data set, the plot shown as Fig. 2C compares exclusively the prediction errors for the 17 selected analytes. The observed trends support those found in Figs. 2A and 2B.

4.3. Repeatability

To study the robustness of fitting the retention parameters and predicting retention times, ten analytes were selected and mea-
sured five times with all five gradients on the amide column. The resulting retention curves are shown in Fig. 3. The error bars signify the spread based on the five determinations. Narrow spreads in the retention curves were observed for the LSS and mixed-mode models, but especially for the adsorption model. The spread was larger for a number of components with the Neue-Kuss model, despite the use of five data points for fitting, and it was dramatic for the quadratic model.

One trend that can be observed is that for specific (mostly early-eluting) analytes the spread is larger towards higher fractions of water. The typical HILIC conditions may no longer be applicable in this range. Earlier-eluting compounds are generally more difficult to fit with the relatively limited gradient range used for scouting. To study the influence of the number of scanning gradients on the error in prediction, retention parameters were determined for each possible combination of gradients where the predicted gradient would fall within the scanning range. For example, gradients A1, A2 and A4 could be combined to predict A3 as the latter falls within the scanning range, whereas this was not the case for predicting gradient A4 from gradients A1, A2 and A3. The results are shown in Fig. 4, with the plot showing the prediction errors for all 56 retained analytes plotted against the number of used scanning gradients. Note that the mixed-mode, quadratic and Neue-Kuss models were not investigated for the use with two scanning gradients as they are three-parameter models. Not surprisingly, a larger number of scanning gradients improves the accuracy of the predictions. However, we can see that for some models this trend is more significant than for other ones. Using four scanning gradients instead of two provides just minor improvements in prediction accuracy for the adsorption model, which appears to perform solidly for the amide column. It should also be noted that the effect of additives, which have been shown to influence the type of interactions in HILIC, have not been evaluated here [26].
5. Concluding remarks

We have investigated the quality-of-fit and prediction accuracies for five retention models and a wide array of compounds from different classes. For most compound classes, the adsorption model provides the most robust performance in terms of its ability to describe and accurately predict HILIC retention based on a limited number of scanning gradients. Prediction accuracies were generally better for a diol column than for an amide column, with the exception of the Neu-Kuss model which performed poorly when using three scouting runs on either column.

While the adsorption model was found to perform robustly for most of the investigated analytes, data suggest a strong effect of the combination of analyte class and stationary-phase chemistry on model performance. The two-parameter model is also attractive, because it can be used with good results using only two scanning gradients. We recommend that the adsorption model be the first-choice model for describing and predicting HILIC retention data, unless data are available to demonstrate that other models ("LSM" model, mixed-mode model) perform adequately in specific situations. Based on the present data, we cannot recommend the quadratic and Neu-Kuss models for use in HILIC.

Acknowledgement

The MANIAC project is funded by the Netherlands Organisation for Scientific Research (NWO) in the framework of the Programmatic Technology Area PTA-COAST3 of the Fund New Chemical Innovations (Project 053.21.113). Agilent Technologies is acknowledged for supporting this work. The Dutch Institute for Cultural Heritage (RCE) is kindly acknowledged for providing the aged synthetic and natural dye samples.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at https://doi.org/10.1016/j.chroma.2017.11.017.

References


