Independent of local properties mathematical models for the calculation of retention indices in programmed temperature gas chromatography

L. Lourici, M.-L. Souici, N. Rebbani and D. Messadi*

Pollution Laboratory, Mokhtar Badji University, BP. 12, 23200 Annaba, Algeria

Abstract. Retention indices of some phthalates separated in temperature programming on SE-30 packed column were calculated by smoothing calibration data with Bezier curves, and from 2 to 6 order B-splines. The values thus obtained were compared to the corresponding ones calculated in a classical way. Whatever the standard n-alkane mixtures used (homologous series, alternate members with even or odd carbon atoms, any mixture with consecutive members not exceeding 4 carbon atoms between each two) the B-splines interpolations lead to retention indices values in better agreement with these ones, although Bezier curve smoothing still leads to values more consistent with the scheme of retention indices. Referring the phthalates to n-alkane standard mixtures, with consecutive members not exceeding 3 or 4 carbon atoms between each 2, connecting the calibration data by B-splines, with orders from 2 to 5, and selecting the set of retention indices corresponding to the smallest value of the sum of squared second divided differences one may recover, with a good accuracy, classical programmed retention indices.

Key words. Programmed temperature gas chromatography – retention indices – non-linear smoothing.

Van den Dool and Kratz [1] have shown, empirically, that identification scheme based on logarithmic relationships among isothermal net retention times, developed by Kováts [2], may be transferred to programmed temperature gas chromatography (PTGC) in terms of linear relationships among retention temperatures.

Sources of error in determining retention indices are due primarily to instrumental variations in temperature and...
carrier gas flow rates, inaccuracies in the measurement of elution temperatures (particularly with packed column on account of some parameters such as thermal capacity of column material and packing, temperature gradient existing between the column wall and its axis), and effects caused by support activity and impurities in the stationary phase [3].

The calculation of the linear PTGC retention index is generally based on the linear interpolation between retention temperatures of neighbouring straight chain homologs (in general n-alkanes) and therefore requires the presence of a complete homologue series of compounds in the sample. In view of the complexity of many samples it is often impossible to add the complete nor-alkane series to the sample without creating interferences with the sample itself.

On the other hand the relationship between the homologues' carbon number (which by definition is proportional to the index) and their elution temperatures is not strictly linear as supposed by linear interpolation [4,5].

To avoid these drawbacks and to gain in precision non-linear interpolations (natural cubic splines; polynomials of different orders) have been used to compute programmed retention indices [4-10].

The easy manipulation of polynomials functions explain the method simplicity which leads to slightly different results (less than 5 i.u.) than those obtained by connecting reference data with natural cubic splines. Furthermore, standard deviations associated to retention indices calculated using these two interpolations don’t present significant differences, which means that these two methods lead to comparable methodological errors [11].

Otherwise, these two methods don’t require the presence of a complete series of homologue straight chain alkyl compounds in the sample and may use also non-alkane sample compounds as references for the calculation of retention indices [12]. Nevertheless, notice that contrary to the polynomial interpolations, natural cubic splines don’t introduce extraneous oscillations [12].

Finally, we must emphasise that the two methods are sensitive to the choice of interpolation nodes and they don’t allow foreseeing the variation of retention indices when the set of standards is modified.

To avoid this global dependence from local properties, we suggested the use of other techniques, such as the B-splines or the Bezier curves [13].

In this work, we applied these two techniques to the calculation of retention indices (B-splines: B<sub>6</sub>, order i = 1 to 6; Bezier curves B<sub>6</sub>) of 7 phthalates of a test mixture, eluted in temperature programming on a slightly polar packed column (SE-30). The retention indices thus obtained, by referring the phthalates to different mixtures of n-alkanes, will be compared to the retention indices determined with a classical way (t<sub>i</sub> using the Van den Dool and Kratz’s relation and a complete set of standard hydrocarbons. We will deduct from it a method of calculation of retention index in PTGC, which eliminates this last constraint.

### Theoretical

#### Bezier Curves [14,15]

The Bernstein’s polynomials permit, by the definition of a polygonal line, to construct a curve joining the first point to the last point of the line, the intermediate tops used to pattern the shape of the curve. Such curves are called Bezier curves.

Let (x<sub>j</sub>, y<sub>j</sub>); j = 0, …, m a given set of points. The Bezier curve of m order is definite by the parametric equation:

\[
\begin{align*}
  x(t) &= \sum_{j=0}^{m} B_{j,m}(t) x_j \\
  y(t) &= \sum_{j=0}^{m} B_{j,m}(t) y_j
\end{align*}
\]

\( t \) varying from 0 to 1, and \( B_{j,m}(t) \) is a polynomial of Bernstein:

\[
B_{j,m}(t) = C_j^m t^j (1-t)^{m-j}.
\]

The coefficients of the binomial:

\[
C_j^m = \frac{m!}{j!(m-j)!}
\]

being calculated by the recurrence relation:

\[
C_j^m = C_j^{m-1} + C_{j+1}^{m-1}.
\]

#### Approximation by B-splines functions [14-16]

Suppose we wish to approach a \( g \) function defined on the interval [0,1], in such a way that the interpolated \( \pi g \):

1- coincide with \( g \) in a finished number of points \( \tau_i \), \( j = 1, \ldots \), \( n \) such as:

\[
0 \leq \tau_1 < \tau_2 < \ldots < \tau_j < \tau_{j+1} < \ldots < \tau_n \leq 1
\]

i.e.,

\[
\pi g(\tau) = g(\tau), \; j = 1, \ldots, n
\]

2- be polynomial by pieces of degree \( k - 1 \), with integer \( k \geq 1 \). To do this, one supposes that both integers \( k \) and \( n \) are fixed, \( k \leq n \), and one introduces the partition of the interval [0,1] with the help of the series of points:

\[
t = \{t_1; i = 1, \ldots, n+k\}
\]

verifying:

\[
t_1 = \ldots t_k = 0; \; t_{i+1} = \ldots = t_{i+k}; \; t_i < t_{i+k}, \; \forall i = 1, \ldots, n.
\]

Then the interpolated \( \pi g \) is like:

\[
\pi g|_{[t_i,t_{i+1}]} \in P_{k-1}(\xi);
\]
3.- be of class $C^{k-2}([0,1])$, i.e.,

$$\pi g \in C^{k-2}([0,1])$$

(10)

these interpolated are linear combinations of basis B-splines functions of order $k$, i.e., $B_{j,k,t}^r, j = 1, \ldots, n$; these basis functions, to the number of $n$ (parameter introduced in (5)), depend on the degree $k-1$ chosen (see (9)) and on the series of $t$ points (see (7) to (9)). They are defined by recurrence relations:

$$B_{j,k,t}(\xi) = \begin{cases} 1 & \text{if } \xi \in [t_j, t_{j+1} [; j = 1, \ldots, n-1 \text{ or if } \xi \in [t_n, t_{n+1}), \{j < t_{n+1}\} \\ 0 \text{ elsewhere .} \end{cases}$$

Then, for $k \geq 2$, we have:

$$B_{j,k,t}(\xi) = \begin{cases} \frac{\xi - t_j}{t_{j+k-1} - t_j} B_{j,k-1,t}(\xi) + \frac{t_{j+k} - \xi}{t_{j+k} - t_{j+1}} B_{j+1,k-1,t}(\xi), & j = 1, \ldots, n-1 \\ \frac{\xi - t_j}{t_{j+k-1} - t_j} B_{j+k-1,t}(\xi), & j = n. \end{cases}$$

The approximation $\pi g$, of the $g$ function, will be then defined by:

$$\pi g(\xi) = \sum_{j=1}^{n} \alpha_i B_{j,k,t}(\xi), \forall \xi \in [0,1].$$

(11)

The explicit determination of the interpolated $\pi g$ of the $g$ function returns to find the $j$ coefficients $\alpha_i = \alpha_i(g)$ which appear in the relation (11). According to the relation (6), one is thus conduct to solve the linear system:

$$\sum_{j=1}^{n} \alpha_i B_{j,k,t}(\xi_j) = g(\xi_j), j = 1, \ldots, n$$

(12)

whose matrix is band of width $2k-1$.

The calculations were done using a program, compatible on PC, written in Turbo-Pascal. In order to facilitate the exploitation of the results we printed the values taken by the non-linear interpolation of the whole of calibration data is maladjusted because of the important errors that it could generate. The simplest calibration curve is obtained by joining the reference data by a polygonal line (polygonal interpolation $I_p$). Notes that the irregularities in the shape thus delineated are attenuated by using the two suggested techniques [17].

In these conditions any linear interpolation of the whole of calibration data is maladjusted because of the important errors that it could generate. The simplest calibration curve is obtained by joining the reference data by a polygonal line (polygonal interpolation $I_p$). Notes that the irregularities in the shape thus delineated are attenuated by using the two suggested techniques [17].

The interpolation by B-spline of first order, which is reduced to a polygonal interpolation, leads to the values of the second column of table I (Part A) which reproduces the values of the retention index of the phthalates referred to the successive nor-alkanes C12–C26. One points out that the difference ($I_{k} - I_{\pi_0}$) increases practically, in absolute value, with the order ($i = 2$ to 6) of the B-splines, but remaining enclosed between 1 and 3 i.u., which is of the order of the magnitude of the scattering about the average values of $I_k$ (Col. 3-Part A); the difference ($I_{L} - I_{\pi_0}$) is more important with 2 to 12 i.u.

According to chromatographic theory, there are linear relationships between the retention temperature of higher members of homologous series and their carbon number, provided that linear temperature programs are applied. The retention indices of the other compounds are defined on the

Flow-rate of Nitrogen U (ENGI-Algeria) carrier gas was 100 mL/min at 25 °C.

The chromatograph was an IGC-16 (Delsi-France) working with pressure regulation, and supplied with a two-flame differential F.I.D. The temperature is elevated from 100 to 290 °C at the rate of 8 °C/min. The chromatogram and the temperature were registered simultaneously on the same graph with a Graphispot (Sefram) two channel recorder.

The peaks recorded at the column outlet being Gaussian in shape, retention temperature corresponded to column temperature when the peak maximum has been reached.

Results and discussion

The retention temperatures (in °C) of the standard nor-alkanes with $z$ carbon atoms (C2), and the test phthalates are reproduced here after: C12 (137.5) – C13 (153) – C14 (163) – C15 (174) – C16 (183.5) – C17 (194) – C18 (203) – C19 (214) – C20 (225) – C21 (233) – C22 (243) – C23 (250) – C24 (257) – C25 (266.5) – C26 (275): dimethyl phthalate (DMP: 165)- diethyl phthalate (DEP: 181.5)- diallyl phthalate (DAP: 195)- dibutyl phthalate (DBP: 218)- bis (methyl glycol) phthalate (BMGP: 219.5)- bis (butyl glycol) phthalate (BBGP: 265)- di (cyclohexyl) phthalate (DCHP: 268.5).

It is known that the programs of temperature of commercial devices could deviate considerably from the supposed linearity, particularly with older equipment. These deviations which are important in beginning and in end of programming are also due to the variations of the ambient temperature. The dependence: flow-temperature increases the non-linearity of retention times, or retention temperatures, of the compounds of a homologue set.

Material and method

For chromatographic purposes, solutions were made by dissolving studied compounds in benzene, and pre-set volumes of 1 μL (Hamilton syringe) were injected manually in the chromatograph.

A stainless steel column (2 m length; 3 mm internal diameter), packed with 10.328 g of stationary phase constituted of chromosorb-Q (60–80 mesh) impregnated with 2.5% (with regard to the naked support) of silicone SE-30, was used for the analyses.
basis of these linear relationships. In order to determine which of the three methods considered yields results more consistent with this definition, the sums of squared second divided differences \[15\] were calculated for the consecutive phthalates:

\[
\sum_M = \sum_j \frac{1}{T_j-T_{j+1}} \left( \frac{r_j^{(M)} - r_j^{(M+1)}}{T_j-T_{j+1}} - \frac{r_j^{(M)} - r_j^{(M+1)}}{T_j-T_{j+1}} \right)^2
\]

(13)

where \( M = L, \text{Be}, \text{Bs}. \) \( T_j \) designates the retention temperature of the phthalate and \( r_j^{(L)}, r_j^{(\text{Be})}, r_j^{(\text{Bs})} \), its retention indices computed by the linear, the Bezier, and the B-spline method respectively.

The interpolation by the B-splines \((i \geq 2)\) and the smoothing by Bezier curve lead to more consistent values with the scheme of the retention index, since the sum of squared second divided differences \( (\sum_M \text{ last line-Part A}) \) is then systematically inferior to the one obtained for \( I_L \). This sum, which is the smallest value corresponds to the smoothing by Bezier curve, decreases when the order of the B-splines grows, to stabilise starting from order 5.

We compared the retention indices thus obtained to the retention indices calculated by the classical way. The regressions \( I(L), I(\text{Be}) \) found by the method of least squares \[18\] were done for every case. The standard deviations \( \sigma \) measure the precision with which were obtained the parameters \( a \) and \( b \) if one takes into account the number of degrees of freedom \((N-2)\); \( N \) is the number of paired values \( I, I_L \).

\[ F_{\text{obs}} = \frac{\sigma^2}{s^2} \]

\( s^2 \) is the variance due to the regression (variation of \( I \) with \( I_L \)), and \( s^2 \) the residual variance about the regression line (measures the scattering of the values of \( I \) about the regression line). The values of calculated \( F \) (Tab. I-Part B) are very higher than the value given by the tables for a level of confidence of 99.9%. The test is therefore highly meaningful, and thus there is a correlation between \( I \) and \( I_L \).

Note that the best correlation is obtained for the interpolation by B-spline of order 2.

Some convenient considerations could impose the utilisation of alternating n-alkanes. We referred the phthalates to n-alkanes with even number (Tab. II) and with odd number of carbon atoms. In this last case the retention indices increase systematically with the order of the B-splines of interpolation, and the values calculated by smoothing with Bezier curve are higher than the values reported in columns 2 and 9 of table I (Part A).

The sums \( \sum_M \) are smaller than the corresponding assembled in tables I and II, the smallest value being again observed for the smoothing by Bezier curve; it stabilises around the value \( 1.69 \times 10^{-3} \) starting from order 3, for the B-spline interpolation (table III, last line).

The best correlations with the \( I_L \) (Tab. I-Part A; Col. 2) are obtained with the B-splines when one uses \( n \)-alkanes with even number of atoms of carbon. Practically the B-splines of orders 2, 3 and 4 give the best possible compromise between the values of Fobs, \( s^2 \), and \( \sum_M \). The highest scattering about regression lines are observed with Bezier

\[ \text{Table I. A) Retention indices of some phthalates eluted in temperature programming} \ (T_i = 100^\circ \text{C}, k_1 = 8^\circ \text{C/min}) \text{ using: - polygonal interpolation} \ (I_L, \text{ average of 5 measurements}); \text{ the B-spline of order 1 to 6} \ (I_{BS} \rightarrow I_{BS}); \text{ the smoothing with Bezier curves. B) Coefficients} \ a \text{ and } b \text{ of the least square straight line: } L = a + bI_L. \]

<table>
<thead>
<tr>
<th>Solute</th>
<th>( I_L )</th>
<th>( \sigma_L )</th>
<th>( I_{BS}^1 )</th>
<th>( I_{BS}^2 )</th>
<th>( I_{BS}^3 )</th>
<th>( I_{BS}^4 )</th>
<th>( I_{BS}^5 )</th>
<th>( I_{BS}^6 )</th>
<th>( I_{BS}^{10} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMP</td>
<td>1418.2</td>
<td>1.5</td>
<td>1417.7</td>
<td>1417.3</td>
<td>1417.2</td>
<td>1417.1</td>
<td>1417.3</td>
<td>1426.2</td>
<td></td>
</tr>
<tr>
<td>DEP</td>
<td>1578.9</td>
<td>2.2</td>
<td>1578.5</td>
<td>1578.1</td>
<td>1577.9</td>
<td>1577.9</td>
<td>1577.9</td>
<td>1581.8</td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>1711.1</td>
<td>2.7</td>
<td>1712.3</td>
<td>1712.9</td>
<td>1713.4</td>
<td>1713.5</td>
<td>1713.5</td>
<td>1713.6</td>
<td></td>
</tr>
<tr>
<td>DBP</td>
<td>1936.4</td>
<td>2.9</td>
<td>1936.4</td>
<td>1936.6</td>
<td>1937</td>
<td>1937.4</td>
<td>1937.9</td>
<td>1946.6</td>
<td></td>
</tr>
<tr>
<td>BMGP</td>
<td>1950</td>
<td>2.5</td>
<td>1950</td>
<td>1950.6</td>
<td>1951.2</td>
<td>1951.8</td>
<td>1952.4</td>
<td>1962.2</td>
<td></td>
</tr>
<tr>
<td>BBGP</td>
<td>2484.1</td>
<td>1.9</td>
<td>2484.8</td>
<td>2485.3</td>
<td>2485.8</td>
<td>2485.5</td>
<td>2485.4</td>
<td>2488.4</td>
<td></td>
</tr>
<tr>
<td>DCHP</td>
<td>2523.5</td>
<td>2.6</td>
<td>2524</td>
<td>2524.5</td>
<td>2525.8</td>
<td>2526.4</td>
<td>2526.7</td>
<td>2533</td>
<td></td>
</tr>
<tr>
<td>10^3 ( \sum_M )</td>
<td>4.24</td>
<td>–</td>
<td>4.21</td>
<td>3.14</td>
<td>2.68</td>
<td>2.25</td>
<td>2.10</td>
<td>1.54</td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{a} & = -1.16 \quad -2.41 \quad -3.85 \quad -4.04 \quad -3.68 \quad 10.82 \\
\text{b} & = 1.007 \quad 1.0014 \quad 1.0024 \quad 1.0026 \quad 1.0026 \quad 0.9981 \\
\text{SE (a)} & = 1.11 \quad 1.63 \quad 2.03 \quad 2.32 \quad 2.65 \quad 7.83 \\
\text{SE (b)} & = 0.0006 \quad 0.0008 \quad 0.0010 \quad 0.0012 \quad 0.0013 \quad 0.0039 \\
10^5 F_{\text{obs}} & = 31.85 \quad 14.79 \quad 9.55 \quad 7.35 \quad 5.64 \quad 6.51 \\
\text{s}_2^2 & = 0.3432 \quad 0.7400 \quad 1.1482 \quad 1.4921 \quad 1.9439 \quad 17.0180 \\
\end{align*}
\]

252
curve which lead to better correlations with \( n \)-alkanes with odd number (1) and even number (2) of carbon atoms.

As a testing we used the reference mixture constituted of \( n \)-paraffines with even number of carbon atoms (C8 to C28) as standards.

The values of \( \Sigma_M \) are higher than the corresponding values reproduced in the table I, and the lowest value is again observed for the smoothing by Bezier curve. This parameter \( \Sigma_M \) is minimal for the interpolation by B-spline of order 4, which corresponds (Tab. IV) to the best values of \( F_{obs} \) and \( s^2_{x} \); therefore, in this case, to the best correlation between the retention indices \( I_{BS} \) and \( I_{L} \).

We attempted to improve our results by trying to estimate the retention temperatures of the missing standard \( n \)-alkanes: C17 – C19 – C21 – C23 – C24 – C25. By treating, by the two suggested techniques, the available results for the \( n \)-alkanes, we determine (iteration of order 0) a first set of values. The predicted and experimental results are treated again as many times as necessary until obtaining, for every missing \( n \)-alkane, successive temperatures which don’t differ from each other more than 0.1 °C. Whatever the technique used, 5 iterations to the most are sufficient to ensure such a result. The predicted retention temperatures agree with those measured to less than 1 °C, the only overtaking (2 °C) is noted for the temperature of C24 determined by interpolation with the B-splines of different orders.

The calculated retention indices are in general stable starting from the third iteration for the interpolation by B-splines functions of order inferior or equal to 3. For the superior orders and for the smoothing by Bezier curve, retention indices of BBGP and DCHP decrease when the order of iteration varies, \( \Sigma_M \) go through a minimum in the case of interpolation by B-splines (orders 3, 4, and 5), and they are the lowest in the case of the smoothing by Bezier curve without for all that presenting a minimum.
The statistical analysis show that for the B-splines the minimum of the sums of squared second divided differences corresponds to the best correlation between the retention indices $I_{BS}$ and $I_z$, and that for the smoothing by Bezier curve the iterations don’t improve this correlation.

The differences between the retention indices calculated by a classical way ($I_z$) and those obtained by B-splines interpolation are in general of the order of magnitude of the standard deviation observed for the first retention index, except with the BBGP for which this difference reaches 7 i.u. In restricting the interval of interpolation (C22 – C26) to (C23 – C26), this difference reduces to 2.6 i.u (and to 0.3 i.u for the DCHP).

**Conclusion**

Beyond the parameters which influence the determination of Kováts’ index, the precision of the retention index in programmed elution depends on the determination of the retention temperatures, and it is conventional to add the standards to the sample, previously to any analysis. Some convenient considerations (complexity of the sample; availability of the successive homologues set) could render this precaution null and void. The interpolating functions used in the literature (polygonal, natural cubic spline, polynomial) depending on the nodes of interpolation, it is not possible to foresee the variation of the retention index when one changes the set of standards.

To avoid this inconvenience we treated the calibration data with two functions, which possess the property of local control, that is, which graphs are modified only around the point that one displaces.

The application was made on a test mixture of seven phthalates separated on a column packed with SE-30, and referred to different mixtures of $n$-alkanes.

If the smoothing by Bezier curve provides retention index that present the highest standard deviations in comparison with the corresponding retention index calculated by classical way (standard error of the order of 7 i.u.), these two types of index are still in correlation for a probability higher than 99%.

The smoothing by Bezier curve always leads to the retention index “the most linear”: it is, apparently, its sole interest.

The results obtained with the B-splines interpolation become stable starting from order 5. The sum of squared second divided differences of the retention indices calculated by this technique varies with the order of the B-splines; the minimum corresponding to the best set of computed retention indices.

We took advantage of these results in order to suggest a method of calculation of retention index in temperature programming which doesn’t necessitate the utilization of a complete set of homologues, and which allows to recover the classical retention index (the standard error being inferior to 2 i.u.). This method based on B-spline interpolation allows to bring back the data of literature, often obtained in heterogeneous conditions, to the norms fixed by the IUPAC (compounds related to a set of contiguous standards hydrocarbons treated by the classical Van den Dool and Kratz method).

The described method must be confirmed with other classes of separated compounds, in different conditions (initial temperature, heating rate), on columns (packed or capillary) with various retention mechanisms. The stability of the computed retention index according to a method or another must be studied, and the standard deviations compared.

**References**