Journal of Chromatography A, 1260 (2012) 193–199

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

A method of calculating the second dimension hold-up time for comprehensive two-dimensional gas chromatography

Imhoi Koo[a], Yaping Zhao[a,b], Jun Zhang[a,c], Seongho Kim[d], Xiang Zhang[a,*]

[a] Department of Chemistry, University of Louisville, Louisville, KY 40292, USA
[b] Department of Radiochemistry, China Institution of Atomic Energy, Beijing 102413, China
[c] School of Electronic Engineering & Automation, Anhui University, Hefei, Anhui 230601, China
[d] Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY 40292, USA

ARTICLE INFO

Article history:
Received 16 June 2012
Received in revised form 17 August 2012
Accepted 20 August 2012
Available online 28 August 2012

Keywords:
Second dimension hold-up time
Second dimension retention time
Comprehensive two-dimensional gas chromatography
n-Alkanes

ABSTRACT

A method of calculating the second dimension hold-up time for comprehensive two-dimensional gas chromatographic (GC × GC) data was developed by incorporating the temperature information of the second dimension column into the calculation model. The model was developed by investigating the relationship between the coefficients in each of six literature reported nonlinear models and the relationship between each coefficient and the second dimension column temperature. The most robust nonlinear function was selected and further used to construct the new model for calculation of the second dimension retention time, in which the coefficients that have significant correlation with the column temperature are replaced with expressions of column temperature. An advantage of the proposed equation is that eight parameters could explain the second dimension hold-up time as well as retention time corresponding to *n*-alkanes and column temperature in the entire chromatographic region, including the chromatographic region not bounded by the retention times of *n*-alkanes. To optimize the experimental design for collecting the isothermal data of *n*-alkanes to create the second dimension hold-up time model, the column temperature difference and the number of isothermal experiments should be considered simultaneously. It was concluded that a total of 5 or 6 isothermal experiments with temperature difference of 40 or 50 °C are enough to generate an accurate model. The test mean squared error (MSE) of those conditions ranges from 0.0428 to 0.0532 for calculation of the second dimension hold-up time for GC × GC data.

1. Introduction

The hold-up time t_M in gas chromatography (GC) is a fundamental chromatographic parameter to calculate other important chromatographic parameters such as retention factor k, adjusted retention time t'_R, Kovats' retention index I, and other thermodynamic parameters. Dominguez et al. [1] reviewed the methods of estimating the hold-up time for one-dimensional gas chromatography and categorized them as direct methods (the value of t_M is experimentally generated) and indirect methods (the value of t_M is obtained indirectly from experimental data via data mining). The absence of a true unretained substance for direct calculation of the hold-up time has forced the development of indirect methods [2].

While the comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC × GC-TOF MS) has been widely used to analyze the chemical compounds such as drug, food, oil and biological samples, a small number of literature reports focused on the calculation methods of the second dimensional hold-up time ${}^2t_M$; Klee and Blumberg developed direct calculation with methane-doped carrier gas [3]. The drawback of this approach is that methane retains in GC columns even at high temperatures [4]. Another drawback is the need of accurately controlling extra instrumental module of pump to continuously add methane to the secondary column. Zhao et al. proposed a method to calculate the second dimension hold-up time by fitting a set of hold-up times calculated using an indirect linearity approach at each column temperature [5], where experimental retention times of three *n*-alkane homologues acquired under the same column temperature were used to calculate each hold-up time [6]. The hold-up times calculated by different sets of the three *n*-alkane homologues have large variation and therefore, the multiple noisy hold-up times calculated for the same column temperature can significantly reduce the accuracy of estimating the mean value of hold-up time at each column temperature and also causes large variation across all column temperatures.

In this study, a novel hold-up time calculation method for the second dimension GC was developed by extending an indirect nonlinear model selected from 6 literature reported methods [1], where

---

* Corresponding author. Tel.: +1 502 852 8878; fax: +1 502 852 8149.
E-mail address: xiang.zhang@louisville.edu (X. Zhang).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chroma.2012.08.068

the temperature information of the second dimension column was incorporated into the calculation model because the hold-up time in the second dimension GC is column temperature dependent. A mixture of *n*-alkanes was analyzed on GC × GC-TOF MS in isothermal mode at different temperatures. The hold-up time of the second dimension GC was first calculated at a given temperature to select the most robust model from literature reported indirect nonlinear models. After investigating the relationship between the estimated coefficients in the most robust indirect nonlinear model and the temperature of the second dimension column, the coefficients that have significant correlation with the column temperature are replaced with expressions of the second dimension column temperature. The performance of the developed method was further compared with literature reported approaches.

2. Experimental and methods

2.1. Chemicals

n-Alkanes (C7-C40) were acquired from Sigma–Aldrich Corp., St. Louis, MO. All of the compounds were present at a purity of between 95 and 99%. All the *n*-alkanes were mixed and prepared in dichloromethane at a concentration of 2.5 μg/mL per compound.

2.2. GC × GC-TOF MS analysis

Experiments were performed on a LECO Pegasus® 4D GC × GC-TOF MS instrument equipped with an Agilent 6890 gas chromatography and a Gerstel MPS2 auto-sampler (GERSTEL Inc., Linthicum, MD), featuring a LECO two-stage cryogenic modulator and a secondary oven. The first dimension column was a non-polar DB-5ms (phenyl arylene polymer virtually equivalent to a (5%-phenyl)-methylpolysiloxane) 30 m × 0.25 mm ^1d_c × 0.25 μm ^1d_f column, and the second dimension column was a medium polar BD-17 ((50%-phenyl)-methylpolysiloxane) 1.79 m × 0.1 mm ^2d_c × 0.1 μm ^2d_f column. Both columns were obtained from Agilent Technologies (Agilent Technologies, Santa Clara, CA), and were connected by means of a press-fit connector before the modulator. The flow rate of ultra-high purity helium carrier gas (99.999%) was set to 1.0 mL/min in a constant flow.

A 2 μL liquid sample was injected into the liner using the splitless mode with the injection port temperature set at 260 °C. The ion source chamber was set to 230 °C, the detector voltage was 1700 V, and the electron energy was 70 eV. The thermal modulator was set to +20 °C relative to the primary oven. The second oven was set to + 5 °C with respect to the primary oven. The mass spectra were acquired at a rate of 200 spectra per second with mass range set to *m*/*z* = 45–800. The modulation period was set at 9 s. The first dimension column was set to isothermal at 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 265, 270 and 280 °C, respectively, with a 60 min analysis time.

2.3. Raw data reduction

The LECO ChromaTOF software (version 4.21), equipped with the National Institute of Standards and Technology (NIST) MS database (NIST MS Search 2.0, NIST/EPA/NIH Mass Spectral Library; NIST 2002), was used for instrument control, spectrum deconvolution, and compound identification. By manually visualizing the raw data, we estimated the peak widths on the first dimension column ranges from 2 to 3 modulation periods. Manufacturer recommended parameters for ChromaTOF were used to reduce the raw instrument data into a compound peak list. These parameters are: baseline offset = 0.5; smoothing = auto; peak width in first dimension = 6 s; peak width in the second dimension = 0.1 s; signal-to-noise ratio = 100.0; match required to combine peaks = 500; R.T. shift = 0.08 s; minimum similarity match = 600. The true peak spectrum was also exported as part of the information for each peak in absolute format of intensity values.

2.4. Hold-up time of the second dimension column at a given column temperature

The hold-up time in the second dimension column ^2t_M varies with column temperature. To a fixed temperature ^2T_e of the second dimension column, ^2t_M can be calculated from the following literature reported approaches by setting the carbon number *n* to zero [7–13]:

$$f_1(n) = p_1 + \exp(p_2 + p_3 n) \tag{1}$$

$$f_2(n) = p_1 + \exp(p_2 + p_3 n + p_4 n^2) \tag{2}$$

$$f_2(n) = p_1 + \exp(p_2 + p_3 n + p_4 n^2) \tag{3}$$

$$f_4(n) = p_1 + \exp(p_2 + p_3 n^{p_4}) \tag{4}$$

$$f_5(n) = p_1 + n^{p_2} \cdot \exp(p_3 + p_4 n) \tag{5}$$

$$f_6(n) = p_1 + p_2 \cdot n^{(p_3 + p_4 n)} \tag{6}$$

where p_i is a coefficient of each model and *n* is the carbon number of a *n*-alkane.

To find the best calculation method from these literature reported approaches, a global search approach [14] was applied to estimate coefficients $(p_1, \ldots, p_{m_i})$ for each of the second dimension retention function f_i, $i = 1, \ldots, 6$ using the experimental dataset of a mixture of *n*-alkanes analyzed on 24 isothermal experiments, which has 220 pairs of second dimension retention time ^2t_R and corresponding alkane represented by its carbon number *n* at a given column temperature ^2T_e (Table S-1). The global search approach, designed to find global optima for nonlinear problems with multi-start framework for initial points, is employed to find the optimal values for the coefficients of each retention function f_i. In this study mean squared error (MSE) of prediction at a given temperature $^2T_e = t$ was used as an object function as follows:

$$\mathrm{MSE}(f_i|^2T_e = t) = \mathrm{E}[(^2t_R(n_k) - f_i(n_k))^2] \tag{7}$$

where $^2t_R(n_k)$ and $f_i(n_k)$ are experimental and estimated second dimension retention time for an alkane with carbon number n_k at given column temperature $^2T_e = t$, respectively. To estimate hold-up time for the 24 temperature conditions, a total of 24 coefficient vectors $(p_{1,t_j}, \ldots, p_{m_i,t_j})$, $t_j = 65, \ldots, 285$ °C, are generated since the equations $(f_1, \ldots, f_6)$ are independent of column temperature ^2T_e. Therefore, the total number of coefficients of each retention function f_i is $24 \times m_i$ for all temperature points, m_i is the number of parameters in retention function f_i.

Although the estimated MSE on training data was used to select the best regression model, this approach is prone to over-fit the training data and the prediction accuracy is unknown for new data when the size of the training data is small. Cross-validation for regression can be used to overcome this problem [15]. However, the cross-validation approach cannot be utilized in this study because the sample size at some column temperatures is smaller than the number of parameters in the retention function f_i. The robustness against artificial noise is another alternative criterion to select the best literature reported retention function, since noise is believed to be the fundamental cause of the over-fitting behavior. Therefore, a model with the smallest MSE on analysis of experimental data with additive artificial noise is considered as the most robust model. A new artificial second dimension retention time $^2t_R^{Noise}$ with white Gaussian noise is generated as follows:

$$^2t_R^{Noise} = {}^2t_R + \varepsilon \tag{8}$$

where the noise ε is generated by Gaussian normal distribution $N(0, \sigma^2)$. Three levels of noise with standard deviation $\sigma = 0.1$, 0.2 and 0.3 (their SNRs are 31.3, 25.3 and 21.8 dB) are considered in this simulation.

2.5. Construction of second dimension hold-up time function ^2t_M

Based on the simulation results in Section 3.1, the function $\hat{f}_1$ is considered as the most robust model against the three noise conditions and therefore, it is used to construct a new model for calculation of the second dimension retention time. An estimated coefficient matrix P of the function $\hat{f}_1$ is obtained as follows:

$$P = (\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2, \hat{\boldsymbol{p}}_3) = \begin{pmatrix} \hat{p}_{11} & \hat{p}_{12} & \hat{p}_{13} \\ \vdots & \ddots & \vdots \\ \hat{p}_{31} & \hat{p}_{32} & \hat{p}_{33} \end{pmatrix} \tag{9}$$

where row vectors correspond to each temperature point from 65 to 285 °C and column vectors $\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2, \hat{\boldsymbol{p}}_3$ denote estimated coefficients of the retention function $\hat{f}_1$ through all temperature points. Each of these column vectors is then used as a new data matrix to construct a set of functions that relate the corresponding coefficients $\{P_i\}$ in $\hat{f}_1$ and the column temperature ^2T_e, by calculating their Pearson's correlation coefficients.

The Pearson's correlation coefficients between the three parameters P_1, P_2, and P_3 in function $\hat{f}_1$ Cor(P_1, P_2), Cor(P_1, P_3), and Cor(P_2, P_3) are −0.9411, 0.9496, and −0.9443, respectively. In addition, the Pearson's correlation coefficients Cor($P_1, {}^2T_e$), Cor($P_2, {}^2T_e$), and Cor($P_3, {}^2T_e$) are −0.8876, 0.9304, and −0.9808, respectively.

Due to the strong relationships between the pairs of coefficients (P_3, P_1), (P_3, P_2), and ($P_1, {}^2T_e$) we consider three regression models to represent these relationships as follows:

$$p_1 = X_{P_3}\beta_{P_1} + \varepsilon, \tag{10}$$

$$p_2 = X_{P_3}\beta_{P_2} + \varepsilon, \tag{11}$$

and

$$p_3 = X_T\beta_{P_3} + \varepsilon, \tag{12}$$

where $X_{P_3} = (1, p_3)$ and $X_T = (1,\ t,\ t^2)$ are predictors including one constant term, and $\beta_{P_2} = (a_0, a_1)^\top$, $\beta_{P_1} = (b_0, b_1)^\top$ and $\beta_{P_3} = (c_0, c_1, c_2)^\top$ are vectors of coefficients to be estimated.

Then a function of calculating the second dimension retention time $^2t_{R,1}(n, t)$ can be reformulated using these regression models of P_1, P_2, P_3 derived from $f_1(n)$ as follows:

$$\begin{aligned} {}^2t_{R,1}(n,t) &= \hat{p}_1 + \exp(\hat{p}_2 + \hat{p}_3 n) \\ &= X_{P_3}\hat{\beta}_{P_1} + \exp(X_{P_3}\hat{\beta}_{P_2} + \hat{p}_3 n) \\ &= \hat{a}_1\hat{p}_3 + \hat{a}_0 + \exp(\hat{b}_0 + \hat{p}_3(n + \hat{b}_1)) \\ &= \hat{a}_1(\hat{c}_2t^2 + \hat{c}_1t + \hat{c}_0) + \hat{a}_0 + \exp(\hat{b}_0 + (\hat{c}_2t^2 \\ &\quad + \hat{c}_1t + \hat{c}_0)(n + \hat{b}_1)). \end{aligned} \tag{13}$$

To study the accuracy of the regression model, the MSE and R^2 are calculated. Although the MSE and R^2 statistic of regression model $\hat{P}_1^{Poly}(t) = \hat{a}_1(\hat{c}_2t^2 + \hat{c}_1t + \hat{c}_0) + \hat{a}_0$ is quite good, the accuracy of approximation can be further improved by using an exponential function. A nonlinear regression model is considered using exponential function of column temperature ^2T_e as follows:

$$P_1^{Exp}(t) = a_{2,0} + \exp(a_{2,1} + a_{2,2}t) + \varepsilon. \tag{14}$$

The coefficients $a_{2,0}$, $a_{2,1}$ and $a_{2,2}$ are calculated from the global search algorithm [14]. $P_1^{Exp}(t)$ is named as exponential model and $P_1^{Poly}(t)$ is called polynomial model. Using the exponential model, Eq. (13) can be reformulated as follows:

$$\begin{aligned} {}^2t_{R,2}(n,t) = {} & \hat{a}_{2,0} + \exp(\hat{a}_{2,1} + \hat{a}_{2,2}t) \\ & + \exp(\hat{b}_0 + (\hat{c}_2t^2 + \hat{c}_1t + \hat{c}_0)(n + \hat{b}_1)) \end{aligned} \tag{15}$$

Based on Eqs. (13) and (15), the hold-up time equations for the second dimension GC can be calculated as follows by setting the carbon number of n-alkanes to zero ($n = 0$),

$${}^2t_{M,1}(t) = {}^2t_{R,1}(0,t) = \hat{a}_1(\hat{c}_2t^2 + \hat{c}_1t + \hat{c}_0) + \hat{a}_0 + g(t,0) \tag{16}$$

or

$${}^2t_{M,2}(t) = {}^2t_{R,2}(0,t) = \hat{a}_{2,0} + \exp(\hat{a}_{2,1} + \hat{a}_{2,2}t) + g(t,0) \tag{17}$$

where $g(t, 0) = \exp(\hat{b}_0 + \hat{b}_1(\hat{c}_2t^2 + \hat{c}_1t + \hat{c}_0))$. Therefore, the proposed two retention models use 7 coefficients in Eq. (16) and 8 coefficients in Eq. (17) to calculate the second dimension hold-up time ^2t_M, respectively.

3. Results and discussion

Based on the experimental design, the second dimension GC column was operated under isothermal mode at 65, 75, 85, 95, 105, 115, 125, 135, 145, 155, 165, 175, 185, 195, 205, 215, 225, 235, 245, 255, 265, 270, 275 and 285 °C, respectively. A total number of 220 data points, i.e., the retention time of each alkane and the corresponding second dimension column temperature, were collected with the carbon number (n) of alkanes ranged from 7 to 31. The number of n-alkanes measured at each temperature is 4, 5, 6, 6, 6, 6, 7, 7, 8, 7, 9, 9, 11, 9, 9, 11, 10, 21, 11, 13, 12, 13, 10, and 13, respectively (Table S-1). The final data consist of a vector (^2t_R, t, n), where ^2t_R is the second dimension retention time for an alkane with n carbons at the second dimension column temperature $^2T_e = t$.

3.1. Construction of the second dimension hold-up time function ^2t_M

In order to find the best literature reported hold-up time model, three levels of white Gaussian noise with standard deviation of $\sigma = 0.1$, 0.2 and 0.3 were added to the experimental data, respectively. A total of 20 simulations were performed for each noise level to avoid selection of a wrong literature model by chance. Table 1 shows the mean and standard deviation of MSE of the estimated functions $\hat{f}_1, \ldots, \hat{f}_6$ using data sets of the experimental data and the experimental data with additive noise. The MSE was calculated for the original retention time (experimentally measured value) and the corresponding retention time values calculated from the estimated function. Although the function $\hat{f}_3(n)$ had the best MSE in the original experimental data, the function $\hat{f}_1(n)$ is the most robust one against three noise conditions. For this reason the

Table 1
Mean squared error (MSE) of prediction for the experimental data and experimental data with different levels of noise variances $\sigma_\varepsilon = 0.1$, 0.2, and 0.3. Each MSE is the average of MSE values through all temperatures.

	Clean data	Noise (σ_ε)		
		0.1	0.2	0.3
$f_1(n)$	3.77E−06	0.0032 ± 0.0005	0.0125 ± 0.0019	0.0298 ± 0.0054
$f_2(n)$	2.43E−06	0.0038 ± 0.0005	0.0248 ± 0.0404	0.0438 ± 0.0333
$f_3(n)$	2.18E−06	0.1819 ± 0.1377	0.2922 ± 0.1837	0.3238 ± 0.2371
$f_4(n)$	2.71E−03	0.6280 ± 0.1903	0.7452 ± 0.2932	0.8860 ± 0.2699
$f_5(n)$	2.52E−06	0.0733 ± 0.0500	0.0905 ± 0.0531	0.1260 ± 0.0682
$f_6(n)$	8.76E−06	0.1716 ± 0.0556	0.2138 ± 0.0907	0.3055 ± 0.1095

Table 2
Coefficients of regression models of P_1, P_2, and P_3. In addition, 95% confidence interval, MSE, and R^2 are also tabulated.

		Coefficient	95% confidence interval	MSE	R^2
$\hat{\beta}_{P_1}$	$\hat{a}_0$	1.1736	(1.1490, 1.1981)	4.0793E−04	0.9017
	$\hat{a}_1$	0.3863	(0.3299, 0.4426)		
$\hat{\beta}_{P_2}$	$\hat{b}_0$	−5.0545	(−5.1649, −4.9442)	0.0082	0.8918
	$\hat{b}_1$	−1.6457	(−1.8992, −1.3922)		
$\hat{\beta}_{P_3}$	$\hat{c}_0$	1.0212	(0.9998, 1.0425)	4.0964E−05	0.9984
	$\hat{c}_1$	−0.0050	(−0.0053, −0.0047)		
	$\hat{c}_2$	7.66E−06	(6.93E−06, 8.40E−06)		

function $\hat{f}_1(n)$ was selected to construct a new model for the second dimension retention time.

Table 2 lists the estimated coefficient $\hat{\beta}_{P_2}$, $\hat{\beta}_{P_1}$ and $\hat{\beta}_{P_3}$ calculated by ordinary least square method. The small values of 95% confidence interval and MSE together with high confidence of determination R^2 demonstrate strong correlation between the model and the experimental data. Most importantly, the coefficient P_1 highly correlates column temperature ${}^2T_e$ with a R^2 value of 0.902. Fig. 1 depicts regression curves of dependent variables P_3, P_2 and P_1 using coefficients $\hat{\beta}_{P_3}$, $\hat{\beta}_{P_2}$ and $\hat{\beta}_{P_1}$, respectively. After getting the relationship of parameter pairs $({}^2T_e, P_1)$, (P_3, P_1) and (P_3, P_2) we rewrote the relationship of temperature ${}^2T_e$ and coefficients P_1, P_2 as shown in Fig. 2. In Fig. 2(b), two regression models were depicted for relationship between temperature ${}^2T_e$ and $P1$. The MSE of the exponential model P_1^{Exp} and polynomial model P_1^{Poly} are 3.01×10^{-4} and 4.46×10^{-4}, respectively. The small value of MSE indicates that the exponential model P_1^{Exp} is better than the polynomial model P_1^{Poly}.

Fig. 3 shows the second dimension hold-up times calculated using models ${}^2t_{M,1}$, ${}^2t_{M,2}$, and $\hat{f}_1(n)$. The large variation of calculated hold-up times by function $\hat{f}_1(n)$ was caused by the lack of correlation between the estimated coefficients and column temperature ${}^2T_e$ in function $\hat{f}_1(n)$. The MSE of the calculated second dimension retention time using ${}^2t_{R,1}(n, t)$ and ${}^2t_{R,2}(n, t)$ are 4.53×10^{-2} and 4.42×10^{-2}, respectively, while the MSE of $\hat{f}_1(n)$ is 3.77×10^{-6}. It is natural that the MSE of $\hat{f}_1(n)$ is very small since the estimation procedure for the literature equation $\hat{f}_1(n)$ is independent of column temperature ${}^2T_e$, i.e., a total of 72 (3 × 24) coefficients are used in retention function $\hat{f}_1(n)$ to estimate hold-up time corresponding to all 24 isothermal experiments, which is much larger than 7 and 8 coefficients of ${}^2t_{R,1}(n, t)$ and ${}^2t_{R,2}(n, t)$, respectively.

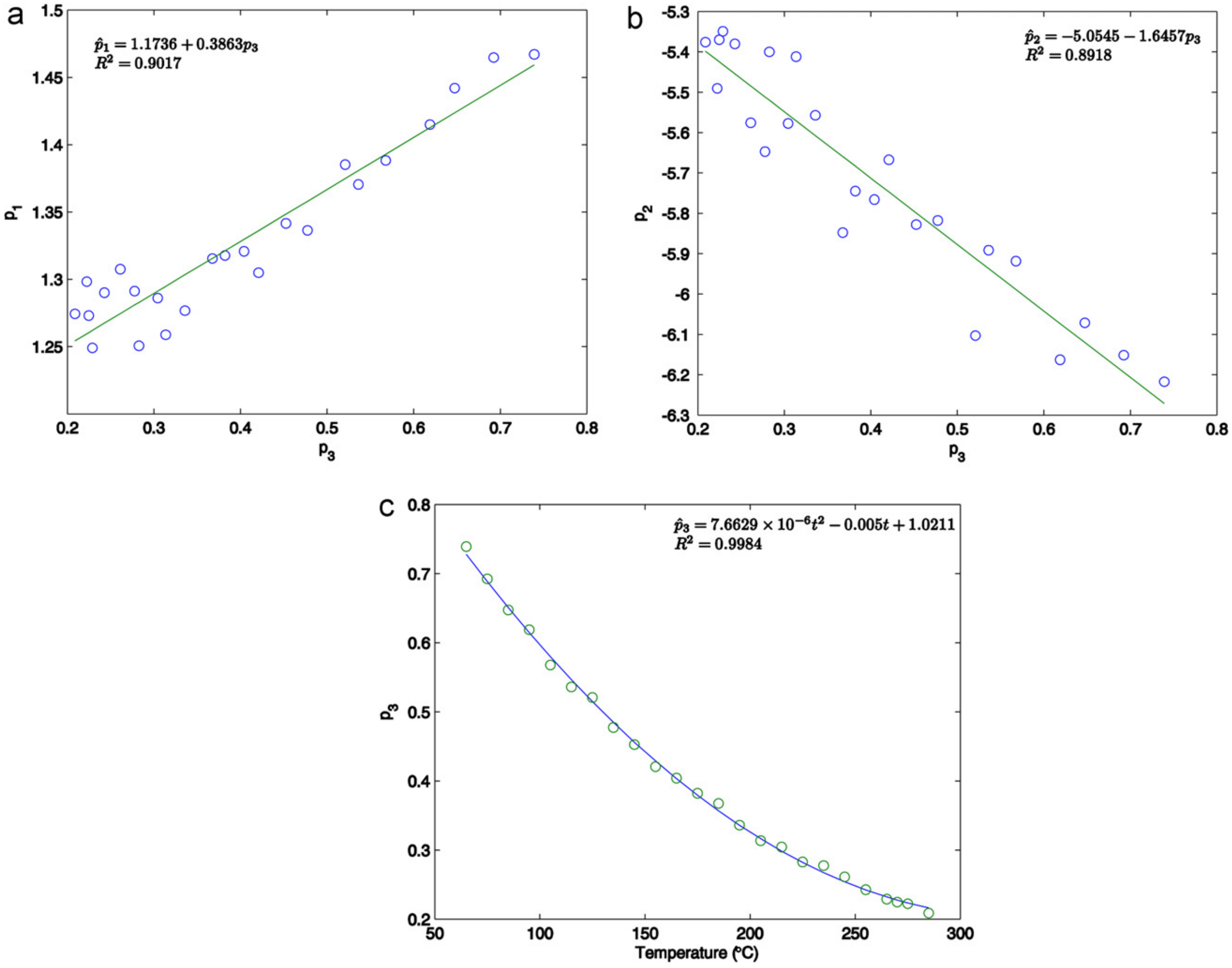

Fig. 1. Plots of regression curves for estimated coefficients (a) P_1 versus P_3, (b) P_2 versus P_3, and (c) P_3 versus column temperature ${}^2T_e$.

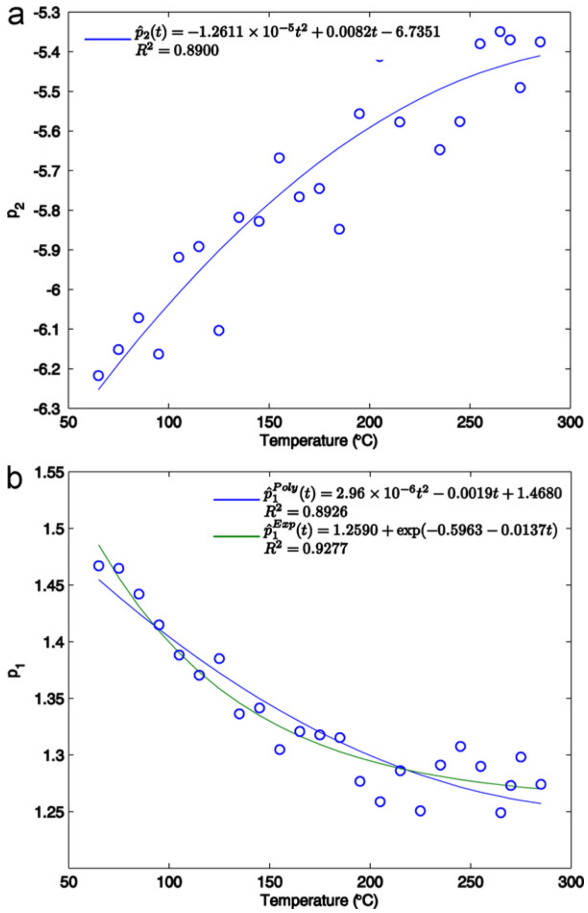

Fig. 2. Relationship between column temperature ${}^2T_e$ and p_1, p_2. (a) Regression of p_2 on temperature ${}^2T_e$ and (b) Regressions of p_1 on column temperature ${}^2T_e$ using the polynomial model (blue line) and the exponential model (green line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Considering relative absolute deviation between the experimentally measured retention time and the retention time calculated using Eq. (17), Supplementary information Fig. S-1 shows boxplot of relative absolute deviation for each *n*-alkane. The relative absolute deviation is defined as

$$\frac{abs({}^2\hat{t}_R(n,t) - {}^2t_R(n,t))}{{}^2t_R(n,t)} \tag{18}$$

where ${}^2\hat{t}_R(n,t)$ is the retention time calculated using Eq. (17) for an alkane with n carbon numbers at column temperature ${}^2T_e = t$, ${}^2t_{R,1}(n,t)$ is experimentally measured primary retention time for the same alkane. Multiple values of relative absolute deviation were obtained for each alkane because of multiple isothermal experiments. The median of relative absolute deviations increases as the increase of carbon number of alkanes. The variation of relative absolute deviation is also increasing. The maximum median of relative absolute deviation is 6.95% at C31-alkane that elutes off the second dimension column at high column temperature.

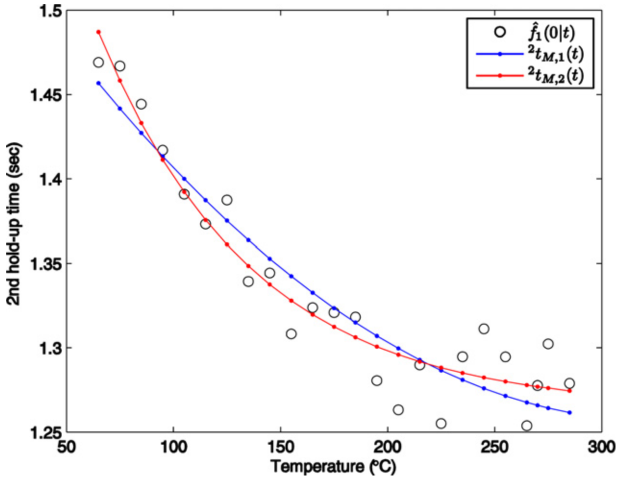

Fig. 3. Calculated second dimension hold-up times using models ${}^2t_{M,1}$, ${}^2t_{M,2}$ and $\hat{f}_1(0)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3.2. Performance analysis of the second dimension hold-up time function ${}^2t_M$

3.2.1. Random temperature selection approach

To evaluate the accuracy of the developed hold-up time function ${}^2t_{R,1}(n,t)$ and ${}^2t_{R,2}(n,t)$, a certain fraction of experimental data were randomly selected from the entire experimental data, i.e., the 220 data points (Table S-1). To simulate real experiments, we first randomly selected a fraction of column temperature from the 24 isothermal experiments. Then, all *n*-alkane data acquired in each of the selected column temperature were used as training data while the remaining data acquired under the unselected column temperatures were used as validation data to study the accuracy of the retention models. As for the training data, 10%, 15%, 20%, 30%, 50%, 70% and 90% of column temperatures were randomly selected, respectively. Even though we can construct the hold-up time function in Eq. (17) by simply setting $n = 0$ in Eq. (15) after nonlinear fitting of all coefficients, the true hold-up time for each experimental condition remains unknown. Therefore, we alternatively evaluate the accuracy of retention time calculation on the validation data using the model described in Eq. (15). The analysis, including data selection, model fitting and validation process, was repeated 100 times for each percentage (sampling ratio) of the column temperature values. The average MSE of predicted retention time and the true value was used to estimate the prediction accuracy.

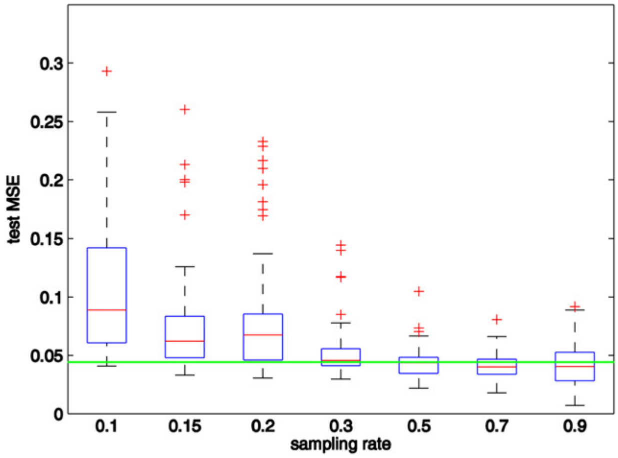

Fig. 4. Boxplot of test mean squared error (MSE) of prediction of ${}^2t_{R,2}(n,t)$ corresponding to random column temperature selection with sampling ratios of 0.1, 0.15, 0.2, 0.3, 0.5, 0.7, and 0.9, respectively. Green line stands for a MSE value of 0.0442 generated by fitting the retention model of Eq. (17) using all 220 experimental data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 4 is the boxplot of prediction results corresponding to sampling ratio. As the number of training temperature data increases, the average of MSE of the validation data decreases and levels off at sampling ratio of 30%. This indicates that 30% of sampling ratio on the 24 isothermal experiments can generate a stable prediction of the second dimension hold-up time, without optimization of isothermal experiments on the distribution of the second dimension column temperature ^2T_e. A 30% of sampling ratio corresponds to 7 isothermal experiments for analysis *n*-alkanes.

3.2.2. Equivalent temperature difference selection approach

We also estimated the effects of hold-up time values calculated using Eq. (17) by selecting a set of training data with equivalent temperature differences. Here, the equivalent temperature difference for a selected temperature set $\{t_1, \ldots, t_i, t_{i+1}, \ldots, t_n\}$ means that all the differences $t_{i+1} - t_i$ of two adjacent column temperatures in the set are constant. All *n*-alkane data acquired under the selected column temperatures are used as training data to fit the model of Eq. (15) while the unselected data are used as validation data to study the model accuracy. MSE of the validation data for predicting model $^2t_{R,2}(n, t)$ is used to evaluate the performance, which is similar to the one used in the previous random selection approach. As looking into all possible temperature differences our goal is to find the optimal temperature difference to fit the model, while the previous random temperature selection indicates the efficient number of temperature size, i.e., the number of isothermal experiments. The possible temperature differences ΔT generated from the 24 isothermal experiments are 10, 20, 30, 40, 50, 60, 70 and 80 °C. For each ΔT, all possible combinations of data acquired during the isothermal experiments were first determined with a requirement of having at least three minimum temperature sets. For example, all data in each of the temperature sets {65, 75, 85 °C}, {75, 85, 95 °C},. . .,{265, 275, 285 °C}, which have adjacent temperature difference $\Delta T = 10\,°\mathrm{C}$, were selected and used to estimate Eq. (15). The temperature sets with more than three, i.e., 4, 5, . . ., 24 with $\Delta T = 10\,°\mathrm{C}$ were also generated and used to fit Eq. (15), as a part the data of $\Delta T = 10\,°\mathrm{C}$. Like the simulation of random temperature selection, the remaining data that were not used to fit the model were used as validation data.

Fig. 5 depicts the boxplot of MSE values of the validation data with respect to each equivalent temperature difference ΔT. The smallest value of MSE at a certain ΔT is the MSE value of validation data that has the smallest number of *n*-alkane data. Although the smallest MSE of each temperature difference increases with the increase of ΔT, the median of MSE is a convex function of ΔT. In detail, the smallest MSE through all temperature sets is 0.0175 in $\Delta T = 10\,°\mathrm{C}$ and the smallest median of test MSE is 0.0560 in $\Delta T = 40\,°\mathrm{C}$. Due to the different number of validation data set, the small variation and also small median of test MSE are more important than the minimum test MSE. As increasing the temperature difference, the variance of test MSE is decreasing although the validation data size of large temperature difference is likely smaller than the one of small temperature difference.

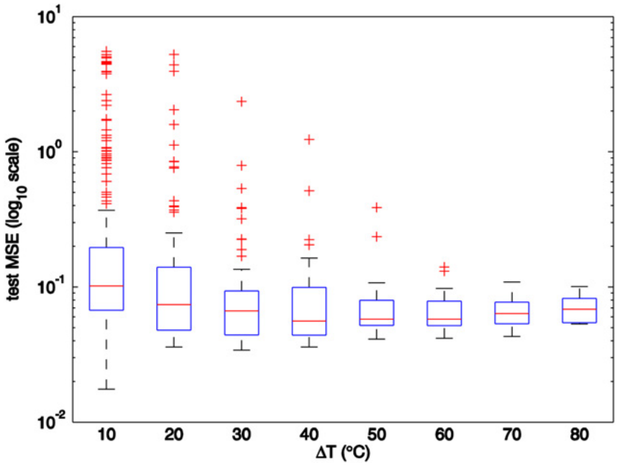

Fig. 5. Boxplot of test mean squared error (MSE) of prediction of $^2t_{R,2}(n, t)$ corresponding to equivalent temperature difference $\Delta T = 10$, 20, 30, 40, 50, 60, 70 and 80 °C.

Supplementary information Figs. S-2 and S-3 show the variation of test MSE corresponding to equivalent temperature difference and the number of temperature sets, respectively. For a fixed equivalent temperature difference, data set with a small number of isothermal experiments (training set) generally has larger variation than the one with a large number of isothermal experiments. This is consistent with our intuition that a large number of samples generate a stable model. On the other hand, the datasets with a large equivalent temperature difference outperform these with a small equivalent temperature difference under a fixed number of isothermal datasets (Fig. S-3). This indicates that the model established using large equivalent temperature differences is more interpolative than extrapolative.

A major concern of the proposed approach is that it requires performing multiple isothermal experiments for analysis of *n*-alkanes. For efficient and economic experiment design to fit the retention model (15) for calculation of the second dimension hold-up time using Eq. (17), the question is what the minimum number of isothermal experiments is needed and what the optimal design of isothermal experiment is for equal column temperature difference. Based on the simulation of random temperature selection, a sampling ratio of 30% on the 24 sets of isothermal experimental data has small variation and its MSE is close to the best median of test MSE. The equal temperature difference simulation shows that the large equivalent temperature difference outperforms small value of equivalent temperature difference under a fixed number of isothermal datasets. In detail, the median of test MSE of the isothermal experiments with column temperature difference of 40 °C monotonically decreases from 0.0796 to 0.0428 at training sample size of three temperature points and six temperature points, respectively, while the median of test MSE of the isothermal experiments with fixed number of column temperature values temperature point for training monotonically decreases from 0.118 to 0.0428 at temperature differences 10 °C and 50 °C, respectively, as shown in Figs. S-2 and S-3. The cases that temperature difference is 40 °C or 50 °C and temperature size (the number of isothermal experiments) is 5 or 6 should to be proper trade-off between the number of isothermal experiments and the accuracy of retention model.

3.3. Comparison with literature reported methods

Zhao et al. developed a comprehensive $^2t_R - {}^2T_e$ function to illustrate the dependency of the second dimension retention time on the carbon number n of alkanes and column temperature of the second dimension GC [5]. By setting $n = 0$ in $\hat{f}_1(n)$, the second dimension retention time actually represents the hold-up time of the second dimension column. Fig. 6 depicts the results of the second dimension hold-up time ^2t_M calculated using three approaches including the selected literature reported model $\hat{f}_1(n)$ in Eq. (1), Zhao's work, and the approach developed in this work. Even though the true hold-up time is unknown, the estimated curve from our proposed approach seems to be proper. Basically, the two approaches proposed by this work and Zhao's work are based on the value of $\hat{f}_1(0)$ and Ettre's model [6], respectively. However, the Ettre's method made very large variation than that of $\hat{f}_1(0)$ within one column temperature and between column temperatures, especially at the high column temperature, even though the average retention time

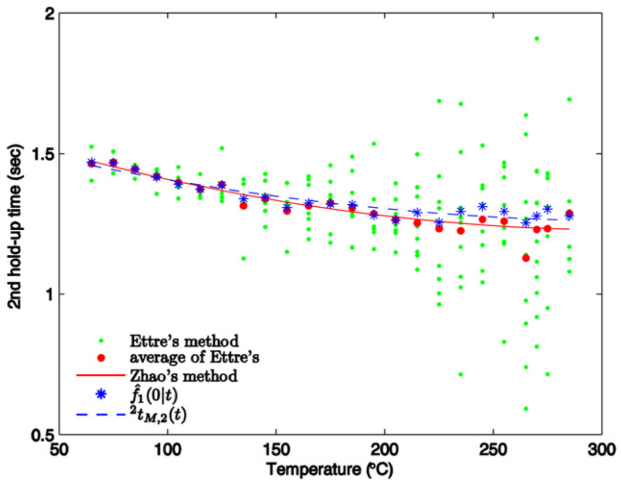

Fig. 6. Comparison of methods for calculation of hold-up time. Green dot and red circle represent the hold-up time calculated using Ettre's method [6] and its average at given temperatures, respectively. Red line is the estimated curve using the method suggested by Zhao et al. [5]. Blue star stands for the calculated hold-up time using the function $f_1(n)$ described in Ref. [7] for at each fixed column temperature. The blue line is the curve generated by the model proposed in this work (Eq. (17)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

estimated by the Ettre's method and $\hat{f}_1(0)$ are similar at the low end column temperatures such as less than 200 °C. Due to the inherent limitation of Ettre's work, Zhao's work generates unstable hold-up time and therefore large variations in prediction of hold-up time at the high temperature end, if limited experimental temperatures and n-alkane are available.

Even though the main purpose of this work is to develop an accurate method to calculate the second dimension hold-up time ^2t_M for GC × GC data, the proposed Eq. (15) could actually be used to generate a n-alkane retention grid at any column temperature ^2T_e. To this end, Eq. (15) is similar to the comprehensive ^2t_R – ^2T_e function proposed by Zhao et al. [5] but with significantly improved accuracy in calculating the second dimension hold-up time. Therefore, with the proposed Eq. (15), knowledge of the temperature program of a GC × GC separation, and knowledge of the primary retention times of the n-alkanes in that GC × GC separation one could create a two-dimensional n-alkane retention grid. This grid could be used to convert the second dimension retention times of any compounds to their corresponding Kovats retention indices. The calculated Kovats retention indices in the second dimension GC and the retention indices (Kovats or Linear retention indices depending on experimental design) calculated from the n-alkane retention times in the first dimension GC could be used together to aid the compound identification, by comparing the experimental retention index of each compound with its retention index value(s) recorded in a reference library such as NIST retention index database.

4. Conclusion

A method of calculating the second dimension hold-up time for GC × GC data was developed, where the temperature information of the second dimension column is incorporated into the calculation model. The correlations among each coefficient of 6 literature reported indirect nonlinear functions and the relationship between each coefficient and the second dimension column temperature was investigated. The most robust nonlinear function (i.e., $f_1(n)$) was determined from additive noise data and further used to construct a new model for calculation of the second dimension retention time, in which the coefficients that have significant correlation with the column temperature are replaced with expressions of column temperature.

Two retention models are proposed in this work, polynomial model (Eq. (16)) and exponential model (Eq. (17)). The small value of MSE indicates that the exponential model is better than the polynomial model. An advantage of the proposed exponential retention model is that 8 parameters could explain the second dimension hold-up time at any column temperature as well as retention time corresponding to column temperature and carbon numbers in n-alkanes, while a total of 72 coefficients are estimated in order for the literature reported nonlinear equation $f_1(n)$ to calculate the hold-up time from the data acquired in 24 isothermal experiments. Furthermore, our equation can predict both hold-up time for the regions in chromatogram where experimental data of n-alkanes are not experimental observed, and can create a two-dimensional n-alkane retention grid in the entire two-dimensional chromatogram. To optimize the experiment design for calculation of the second dimension hold-up time, the column temperature difference and the number of isothermal experiments should be considered simultaneously. In our simulation temperature difference of 40 or 50 °C and temperature size of 5 or 6 are recommended.

Acknowledgements

This work was supported by National Institute of Health (NIH) grant 1RO1GM087735 through the National Institute of General Medical Sciences (NIGMS).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chroma.2012.08.068.

References

[1] J.A.G. Dominguez, J.C. Diez-Masa, V.A. Davankov, Pure Appl. Chem. 73 (2001) 969.
[2] J.E. QuintanillaLopez, R. LebronAguilar, J.A. GarciaDominguez, J. Chromatogr. A 767 (1997) 127.
[3] M.S. Klee, L.M. Blumberg, J. Chromatogr. A 1217 (2010) 1830.
[4] J.A. Garciadominguez, J. Garciamunoz, E. Fernandezsanchez, M.J. Molera, J. Chromatogr. Sci. 15 (1977) 520.
[5] Y.P. Zhao, J. Zhang, B. Wang, S.H. Kim, A.Q. Fang, B. Bogdanov, Z.X. Zhou, C. McClain, X. Zhang, J. Chromatogr. A 1218 (2011) 2577.
[6] L.S. Ettre, Chromatographia 13 (1980) 73.
[7] R. LebronAguilar, J.E. QuintanillaLopez, J.A. GarciaDominguez, J. Chromatogr. A 760 (1997) 219.
[8] K.S. Reddy, J.C. Dutoit, E.S. Kovats, J. Chromatogr. 609 (1992) 229.
[9] W.O. McReynolds, J. Chromatogr. Sci. 8 (1970) 685.
[10] J.A. García Domínguez, J.G. Muñoz, E.F. Sánchez, M.J. Molera, J. Chromatogr. Sci. 15 (1977) 520.
[11] F. Riedo, D. Fritz, G. Tarján, E.s. Kováts, J. Chromatogr. 126 (1976) 63.
[12] V. Ezrets, M. Vigdergauz, Chromatographia 9 (1976) 205.
[13] K.S. Reddy, R. Cloux, E.s. Kováts, J. Chromatogr. A 673 (1994) 181.
[14] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, R. Marti, Informs J. Comput. 19 (2007) 328.
[15] S. Arlot, A. Celisse, Stat. Surv. 4 (2010) 40.