Talanta 126 (2014) 151–156

Contents lists available at ScienceDirect

Talanta

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

Determination of boiling point of petrochemicals by gas chromatography–mass spectrometry and multivariate regression analysis of structural activity relationship

Sayo O. Fakayode [a,*], Breanna S. Mitchell [b], David A. Pollard [b]

[a] North Carolina A&T State University, Greensboro, NC 27411, United States
[b] Chemistry Department, Winston-Salem State University, Winston-Salem, NC 27110, United States

ARTICLE INFO

Article history:
Received 24 January 2014
Received in revised form
14 March 2014
Accepted 17 March 2014
Available online 25 March 2014

Keywords:
Gas chromatography
Petrochemicals
Boiling point determination
Multivariate regression analysis
Structural activity relationship
Pattern recognition

ABSTRACT

Accurate understanding of analyte boiling points (BP) is of critical importance in gas chromatographic (GC) separation and crude oil refinery operation in petrochemical industries. This study reported the first combined use of GC separation and partial-least-square (*PLS1*) multivariate regression analysis of petrochemical structural activity relationship (SAR) for accurate BP determination of two commercially available (D3710 and MA VHP) calibration gas mix samples. The results of the BP determination using *PLS1* multivariate regression were further compared with the results of traditional simulated distillation method of BP determination. The developed *PLS1* regression was able to correctly predict analytes BP in D3710 and MA VHP calibration gas mix samples, with a root-mean-square-%-relative-error (*RMS%RE*) of 6.4%, and 10.8% respectively. In contrast, the overall *RMS%RE* of 32.9% and 40.4%, respectively obtained for BP determination in D3710 and MA VHP using a traditional simulated distillation method were approximately four times larger than the corresponding *RMS%RE* of BP prediction using MRA, demonstrating the better predictive ability of MRA. The reported method is rapid, robust, and promising, and can be potentially used routinely for fast analysis, pattern recognition, and analyte BP determination in petrochemical industries.

1. Introduction

Crude petroleum is a complex mixture containing petrochemicals of wide range boiling point (BP) distributions. The utility of petroleum is highly dependent on its BP, necessitating the need for the development of analytical techniques, capable of crude oil separation into its various components. Separation of crude petroleum typically involves a conventional fractional distillation (FD) technique based on differences in petrochemicals BP distribution. Unquestionably, the determination of petrochemical BP distribution using conventional FD strategy is useful and critical. However, FD has significant challenges and drawbacks including, large sample size requirement, lengthy analysis time, and relatively poor accuracy. The drawbacks of FD has resulted in the current use of a simulated distillation using gas chromatography (GC) analysis as a better alternative strategy for petrochemical BP distributions determination in oil and gas industries [1–10].

Determination of BP of unknown sample using GC simulated distillation approach involves a two-step procedure. In the first step, the unknown sample is co-separated with standards calibration mixture of known BP using GC. In the second step, a calibration curve involving the plot of BP of the standard calibration mixture versus analyte retention times is constructed. The constructed calibration curve is then subsequently used to evaluate the BP of the unknown sample. The determination of BPs distribution of petrochemicals using simulated distillation technique is not only simple and effective, but it is also rapid and requires a relatively small sample size. In addition, this strategy has been very effective and robust, with a reasonable BP prediction error. This technique has become routine for simulation of petrochemical's BP in the oil and gas industries.

The simulated distillation approach only focused on the use of ordinary univariate regression analysis (URA) of BP and analyte retention times. However, the similarities and/or differences in petrochemical structural activity relationship (SAR) could preclude the use of URA for accurate determination of BP distribution solely from GC retention times. For instance, analyte retention time is not always linearly related with BP, especially for analytes with extremely low or high boiling points. Besides, analyte SAR, including the shape, size, molecular weight, number of carbons, number of hydrogens, number of single bond, number of double bond, number of benzene ring, and the presence of other functional groups may

---

[*] Corresponding author. Tel.: +1 336 285 2245; fax: +1 336 334 7124.
E-mail address: sofakayo@ncat.edu (S.O. Fakayode).

http://dx.doi.org/10.1016/j.talanta.2014.03.037
0039-9140/

have notable influence, not only on analyte retention times, but also on analyte elution order. In addition, analyte flash point, refractive index, density, and melting point may affect both the analyte retention times and elution order, hindering the use of URA for accurate BP prediction.

In this study, we hypothesized that, the use of simulated distillation from GC data in conjunction with multivariate regression analysis of petrochemicals structural activity relationship may provide a better alternative strategy to improve the accuracy of petrochemicals BP determination. The practical applications of MRA in conjunction with SAR for accurate prediction of analyte retention times of samples of petrochemical, pharmaceutical, and environmental interest have been demonstrated in chromatographic separation [11–15]. In a recent study, the potential utility of molecular weight, specific gravity, and cumulative weight fraction as inputs for artificial neutral networks for the estimation of boiling point distributions of C_{7+} has also been demonstrated [16]. Unquestionably, the use of analyte molecular weight, specific gravity and cumulative weight fraction structural activity relationship input are significant for the estimation of boiling point. However, other analytes structural activity relationship including density, melting point, number of carbon, number of hydrogen, number of single bonds, number of double bonds, number of benzene rings, flash point, and refractive index may have considerable effects on the GC separation and/or petrochemical boiling point. This study therefore reported a comprehensive and the first combined use of a gas chromatography–mass spectrometry separation, analyte structural activity relationship, and multivariate analysis for the determination of petrochemicals BPs. The results of the BP analysis using MRA in this study were further compared with the results of a conventional simulated distillation method using univariate regression analysis of BP determination from GC data. Furthermore, the study explored the potential utility of analyte SAR in conjunction with MRA for petrochemical pattern recognition in GC separations.

2. Experimental

2.1. Material and chemical

Methanol, 2,2 dimethyl butane, hexene, benzene, 2,2,4 trimethyl pentane, n-heptane, toluene, p-xylene, n-nonane, n-decane, dodecane, n-tridecane, and n-tetradecane were obtained from Sigma-Aldrich. A Rtx-XLB column, D3710, and MA VPH calibration mix samples were purchased from Restek, Bellefonte, PA, USA. The chemicals used for the study were ACS certified grade or better purity.

2.2. GC separation and multivariate data analysis

A boiling point calibration mixture sample consisting of a 0.05% v/v of 2,2-dimethyl butane, hexene, benzene, 2,2, 4 trimethyl pentane, n-heptane, toluene, p-xylene, n-nonane, n-decane, dodecane, n-tridecane, and n-tetradecane was prepared in HPLC grade methanol. The analytes in the calibration mixture were chosen for the study because of variation in their structural activities relationship (SAR). A 1 µL aliquot of the calibration mixture sample was directly injected and separated using a GC instrument equipped with a mass-spectrometer detector (GCMS-QP5000, Shimadzu). The GC separation was performed in a Rtx-XLB column (inner diameter: 0.25 mm; film thickness: 0.25 µm; and length 30 m). The GC separation was performed in a temperature programming mode to ensure better analyte resolution. The column temperature was operated in a gradient programming mode, with an initial temperature of 33 °C, held for 3 min at 33 °C and then increased at the rate of 1 °C min^{-1} to 100 °C. The temperature was then held for 4 min at 100 °C, and then increased at a rate of 10 °C min^{-1} to 200 °C. Helium (He) gas was used as the mobile phase, with a column injection pressure of 52.1 kPa, a total flow of 30 mL min^{-1}, column flow of 1.1 mL min^{-1}, linear velocity of 37.6 cm s^{-1}, and a split ratio of 25:1. The GC–MS interface temperature was set at 300 °C. The mass spectrometer detector was operated in a scan acquisition mode, scanning from 40m/z to 400m/z. Each analyte peak in the calibration mixture GC chromatogram was identified and confirmed using the mass spectrometer.

2.3. Analyte structural activity relationship and multivariate regression analysis

A partial-least-square (*PLS1*) multivariate regression analysis was used to correlate changes in analyte structural activity relationships and corresponding retention times from GC separation with analyte BP in the calibration mixture. Table 1 shows the analyte structural activity relationship and GC retention time data set used for *PLS1* regression model development. The *PLS1* model was carefully optimized and subsequently used to predict analyte BPs in two commercially available calibration gas mix samples (D3710 and MA VHP). Multivariate data analysis was performed using chemometric software (9.8, The Unscrambler, CAMO Incorporation, NJ).

3. Result and discussion

3.1. GC separation and multivariate regression analysis

Fig. 1 shows the chromatogram of the GC analysis of the calibration mixture sample. Obviously, all analytes in the mixture

Table 1
Structural activity relationship of calibration petrochemical mixtures data set for multivariate analysis.

Analyte	D (g mL^{-1})	MP(°C)	MW (g mol^{-1})	# SB	# DB	# C	# H	RI	VD	FP (°C)	BP (°C)	RT (min)
2,2 Dimethyl butane	0.6490	−100	86.18	19	0	6	14	1.369	2.97	−29	50	1.377
Hexene	0.6780	0	84.16	16	1	6	12	1.388	3	−25	63	1.652
Benzene	0.8740	5.5	78.11	9	3	6	6	1.5	2.77	−11	80	2.211
2,2,4 Trimethyl pentane	0.6920	−107	114.23	25	0	8	15	1.39	3.9	−12	99	2.522
n-Heptane	0.6840	−91	100.2	22	0	7	16	1.387	3.5	−4	98	2.709
Toluene	0.8650	−93	92.14	12	3	7	8	1.496	3.2	4	111	3.948
p-Xylene	0.8610	13	106.17	11	3	8	10	1.495	3.7	25	139	6.635
n-nonane	0.718	−53	128.26	28	0	9	20	1.405	4.41	31	151	7.617
n-Decane	0.7300	−30	142.28	31	0	10	22	1.411	4.9	46	174.1	10.763
Dodecane	0.7500	−9.6	170.33	37	0	12	26	1.421	5.96	74	216.3	15.357
n-Tridecane	0.7560	−5.3	184.36	40	0	13	28	1.425	6.4	94	234	17.003
n-Tetradecane	0.7620	5.5	198.39	43	0	14	30	1.429	6.83	100	253	18.978

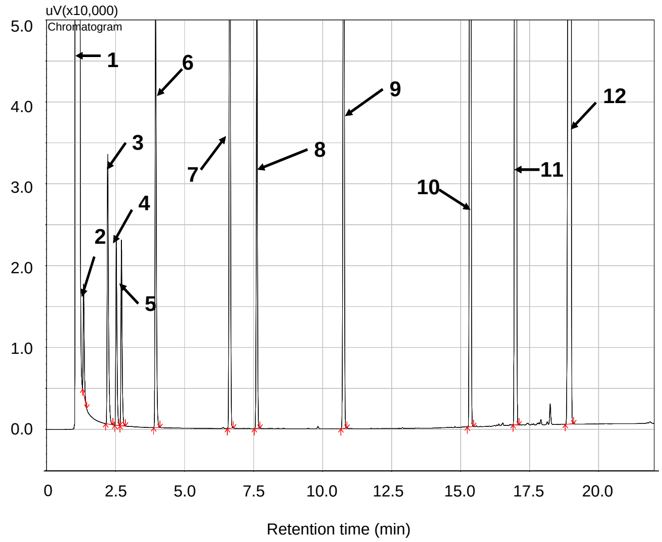

Fig. 1. GC chromatogram of calibration gas mixture separation. (**1**) 2,2 dimethyl butane; (**2**) hexene; (**3**) benzene; (**4**) 2,2,4 trimethyl pentane; (**5**) n-heptane; (**6**) toluene; (**7**) p-xylene; (**8**) n-nonane; (**9**) n-decane; (**10**) dodecane; (**11**) n-tridecane; and (**12**) n-tetradecane.

Table 2
X-loading of structural activity relationship variables to *PC1* and *PC2*.

Variable	*PC1*	*PC2*
Density (*D*)	1.82E – 04	– 8.35E – 04
Melting point	0.441	– 0.937
Molecular weight	0.583	0.299
Number of single bonds	0.142	0.13
Number of double bonds	– 6.41E – 03	– 1.96E – 02
Number of carbon atom	4.04E – 02	1.77E – 02
Number of hydrogen atom	9.72E – 02	8.52E – 02
Refractive index	6.53E – 05	– 5.30E – 04
Vapor density	2.01E – 02	9.54E – 03
Flash point	0.66	0.158
Retention time	9.40E – 02	2.19E – 02

were well resolved and eluted within 25 min. The analytes SAR and the resulting retention time obtained from GC analysis in Table 1 was subsequently subjected to a *PLS1* multivariate regression analysis. Detailed mathematical procedures of multivariate regression analysis for sample calibration have been reported elswhere [17–19] and can also be found in several monographs [20–27]. In brief, a multivariate regression equation can be simplified and represented by the following equation:

$$y=b_0+x_1b_1+x_2b_2+x_3b_3+x_4b_4+x_5b_5+x_6b_6+\ldots\ldots\ldots x_nb_n \tag{1}$$

where *y* is the *dependent variable* (analyte BP in this study), x_1, x_2, ……, x_n are the *independent variables* (analyte density, melting point, molecular weight, number of carbon, number of hydrogen, number of single bonds, number of double bonds, number of benzene rings, flash point, refractive index, and retention time in this study), b_0 is the intercept of the regression equation, and b_1, b_2, ………, b_n are the regression coefficents of *x-variables*. Eq. (1) can be expressed in matrix notation as shown in the following equation:

$$[Y]=[X]b \tag{2}$$

where *[Y]* contains the matrix values of the dependent variables for all samples, *[X]* is a matrix composed of values of independent variables of all samples, and *b* contains the regression vector. In other words, the regression vector relates the independent and dependent variables. The goal in developing any regression model is to first determine the values of the regression vector using a data set of known independent and dependent variables, a process known as *regression or model calibration*. Once the value of the regression vector is established from the regression equation in the calibration phase, the value of regression vector can then be combined with the independent variables of future unknown samples to predict the dependent variables of the future samples. This process is known as *regression or model validation*.

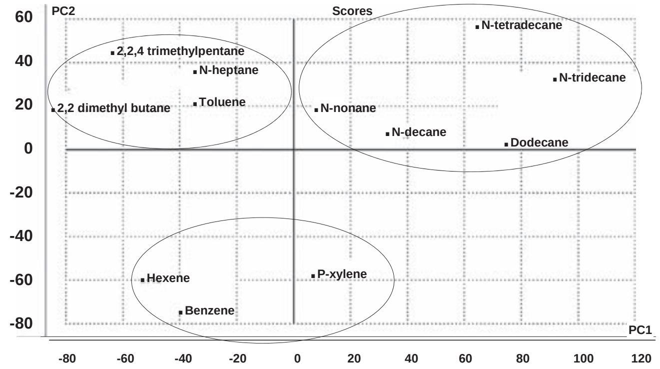

Fig. 2. *Scores* plot of *PC1* versus *PC2*.

The assumption of a multivariate regression model is that there is no co-linearity between *x-variables*. However, in practical terms, there is often significant co-linearity between *x-variables*, hindering effective predictability of a multivariate regression model for future sample. It is therefore critical to first remove any co-linearity between *x-variables* to ensure accurate prediction of future samples. Co-linearity between *x-variables* is usually eliminated using a modern principal component analysis (*PCA*) technique, where the original *x-variable* data set is transformed and transposed to new orthogonal variance scaled data set [20–26]. Since the principal components (*PCs*) in the new variance scale data set are orthogonal to each other, co-linearity in the original *x-variable* data set is eliminated. Besides the removal of co-linearity among *x-variables*, the dimension of a data set is often reduced, allowing the use of relatively fewer *PCs* to re-represent the data, further eliminating the inherent noise in a data set.

The developed *PLS1* model in the calibration phase resulted in a square correlation coefficient (R^2) of 0.996908. A careful analysis of the *PLS1* regression model revealed that, the first two principal components (*PC1* and *PC2*) accounted for 99% in the variability of the independent or *x-variables* data set. In addition, *PC1* and *PC2* also explained approximately 98% in the variability of dependent or *y-variable* (boiling point) data set. Therefore, two *PCs* are adequate to represent the data set. Table 2 shows the *x*-loading value of each structural activity relationship variable to the *PC1* and *PC2*. The *x-loading* is typically used to evaluate the contribution or significance of each independent variable in a data set to the *PLS1* regression model [20–27]. While all the structural activity relationship variables in the data set are significant and contribute to the *PLS1* regression model for the accurate prediction of BP, the contribution of each variable is widely varied. For example, the presence of a double bond negatively contributed to *PC1* while analyte density, melting point, and refractive index negatively contributed to *PC2*. However, other variables positively contributed to the regression model. In general, the flash point, molecular weight, melting point, and number of single bond of petrochemicals were found to be the major contributors to *PC1* and *PC2*.

Fig. 2 is the scores plot of *PC1* versus *PC2*. A scores plot is typically utilized for pattern recognition in a multivariate and/or complex data set. The scores plot is also used to obtain information that is often hidden and obscured from ordinary data examination. The result of the scores plot obtained in this study was remarkable and showed interesting groupings of analytes in the calibration mixture sample. For example, high molecular weight, straight chain alkanes with higher boiling points (n-nonane, n-decane, dodecane, n-tridecane, and n-tetradecane) were grouped in the top quadrant of the scores plot. In contrast, the low molecular weight, branched chain alkanes with relatively low boiling points (2,2 dimethyl butane, 2,2,4 trimethyl pentane, and n-heptane) were grouped on the left quadrant of the scores plot. However, aromatic compounds, or compounds containing at least one double bond, with moderate boiling points (hexene, benzene, and p-xylene) were grouped in the bottom lower quadrant of the scores plot. The scores plot can therefore be potentially used for pattern recognition and for rapid screening of future unknown analyte identities.

3.2. *Validation study*

3.2.1. *Determination of boiling points of D3710 and MA VPH calibration gas mix*

While the R^2 of the developed *PLS1* regression model in the calibration was impressive, the practical application of any regression model is the capability of the model to accurately predict the future unknown samples. To evaluate the accuracy of the developed *PLS1* model for the prediction of BP of future samples, two independent and commercially available calibration gas mix samples (D3710 and MA VPH) were selected for the validation study. The D3710 calibration gas mix consists of: 2-methylbutane, n-pentane, n-hexane, 2,4 dimethyl pentane, n-heptane, toluene, n-octane, p-xylene, n-propylbenzene, n-butyl benzene, n-decane, dodecane, n-tridecane, n-tetradecane, and n-pentadecane. The BP distribution of D3710 calibration gas mix ranged between 36 °C and 270 °C. Also, the physicochemical properties and structural activity relationship of D3710 varied widely. The MA VPH calibration gas mix is composed of: n-pentane, 2,2,4 trimethyl pentane, methyl-tert-butyl ether, n-nonane, benzene, toluene, ethylbenzene, p-xylene, o-xylene, 1,2,4-trimethylbenzene, and naphthalene.

The D3710 and MA VPH gas mix samples were subjected to GC analysis under the same experimental conditions initially used for the separation of the calibration mixture sample used for *PLS1* model development. Each analyte in D3710 and MA VPH samples were also identified and confirmed using the mass spectrometer. Tables 3 and 4 show the analytes, SAR, and retention times of D3710 and MA VHP calibration gas mix used for the validation study. The developed *PLS1* model in the calibration phase was subsequently used to predict the analyte BP of D3710 and MA VPH gas mix samples.

Tables 5 and 6 show the actual and the predicted analyte BP in D3710 and MA VPH calibration gas mix using *PLS1* multivariate regression analysis from the calibration study. The accuracy of the *PLS1* regression model to correctly predict analyte BP was evaluated by a root-mean-square-%-relative-error (*RMS%RE*).where

$$\%\,\text{RE} = \frac{\text{predicted} - \text{actual boiling point}}{\text{actual boiling point}} \times 100 \qquad (3)$$

The validated study conducted for analyte BP prediction in the D3710 gas mix sample resulted in a prediction *RMS%RE* of 6.4%. A similar validation study conducted for MA VHP resulted in a prediction *RMS%RE* of 10.8%.

3.3. *Comparative analysis of boiling point determination using univariate regression and PLS1 multivariate regression analysis*

As noted earlier, the current simulated distillation method of BP determination from GC analysis involves URA of analyte retention time and BP. The major drawback of this method is that, analyte BP may not be linearly related to the retention time, especially, for analytes with extremely low or high boiling points. For example, Fig. 3 is the plot of analyte BP versus the retention times of the calibration mixture sample, showing a curvature in the graph. Similar curvatures in the plots of boiling point against the retention times at a very low or high boiling point have also been reported by American Society for Testing and Material standard test method for boiling point distribution of samples [1,4,5,9,10]. A curvature or non-linearity between the BP and retention time hinders the utility of traditional URA for accurate prediction of analyte BP solely from the retention time. To evaluate the effectiveness of MRA *PLS1* model over the current simulated distillation method, a URA involving direct modeling of retention time and BP of the calibration mixture was developed.

Tables 7 and 8 show the summary of the results of validation studies conducted for D 3710 and MA VHP using URA. The overall prediction *RMS%RE* of 32.9% obtained for D3710 using URA was approximately four times larger than the corresponding *RMS%RE* of 6.4% of analyte BP prediction obtained using multivariate *PLS1* regression. Also, the overall prediction *RMS%RE* of 40.4% obtained for MA VPH using URA was significantly larger than the corresponding *RMS%RE* of 10.8% of analyte BP prediction using multivariate *PLS1* regression. Obviously, the validation results of BP predictions obtained using the multivariate *PLS1* regression model were noticeably better than similar validation studies conducted using traditional URA. This clearly demonstrated that analyte's structural relationship activity play significant role not only in GC analysis, but also has a substantial influence on the capacity of the regression model to accurately predict PB from GC separation.

Table 3

Structural activity relationship of D 3710 gas mix data set used for validation study.

Analyte	D (g mL^{-1})	MP (°C)	MW (g mol^{-1})	# SB	# DB	# C	# H	# O	RI	VD	FP (°C)	BP (°C)	RT (min)
n-Pentane	0.6260	−130	72.15	16	0	5	12	0	1.358	2.48	−49	36	1.875
2 Methylbutane	0.6200	−100	72.15	16	0	5	12	0	1.354	2.60	−29	30	1.792
n-Methyl pentane	0.6530	−154	86.18	18	0	6	14	0	1.371	3	−7	62	2.225
2,4 Dimethyl pentane	0.6730	−123	100.2	22	0	7	16	0	1.381	3.48	−7	80	2.839
n-Heptane	0.6840	−91	100.2	22	0	7	16	0	1.387	3.5	−4	98	4.210
Toluene	0.8650	−93	92.14	12	3	7	8	0	1.496	3.2	4	111	5.975
n-Octane	0.7030	−57	114.23	25	0	8	18	0	1.397	3.9	13	126	6.123
p-Xylene	0.8610	13	106.17	11	3	8	10	0	1.495	3.7	25	138	8.489
n-Propylbenzene	0.8620	−99	120.19	9	3	9	12	0	1.491	4.14	42	159	10.236
n-Butyl benzene	0.8600	−88	134.22	21	3	10	14	0	1.489	1	59	183	12.941
n-Decane	0.7300	−30	142.28	31	0	10	22	0	1.411	4.9	46	174	10.991
Dodecane	0.7500	−9.6	170.33	37	0	12	26	0	1.421	5.96	74	216.3	16.966
n-Tridecane	0.7560	−5.3	184.36	40	0	13	28	0	1.425	6.4	94	234	19.230
n-Tetradecane	0.7620	5.5	198.39	43	0	14	30	0	1.492	6.83	100	253	21.113
n-Pentadecane	0.7690	9	212.41	46	0	15	32	0	1.431	7.4	132	270	22.758

Table 4
Structural activity relationship of MA VPH gas mix data set used for second validation study.

Analyte	D (g mL^{-1})	MP(°C)	MW (g mol^{-1})	# SB	# DB	# C	# H	# O	RI	VD	FP (°C)	BP (°C)	RT (min)
n-Pentane	0.6260	−130	72.15	16	0	0	12	0	1.358	2.48	−49	36	1.829
2,2,4 trimethyl pentane	0.6920	−107	114.23	25	0	8	18	0	1.39	3.9	−12	99	3.801
Methyl-tert-butyl ether (MTBE)	0.74	−100	88.15	17	0	5	12	1	1.369	3.1	−33.0	55.5	2.304
n-Nonane	0.718	−53	128.26	28	0	9	20	0	1.405	4.41	31.0	151	8.741
Benzene	0.8740	5.5	78.11	9	3	6	6	0	1.5	2.77	−11	80	
Toluene	0.8650	−93	92.14	12	3	7	8	0	1.496	3.2	4	111	5.899
Ethylbenzene	0.867	−95	106.17	15	3	8	10	0	1.495	3.7	15.0	136	5.899
p-Xylene	0.8610	13	106.17	15	3	8	10	0	1.495	3.7	25	138	8.419
o-Xylene	0.879	−24	106.17	15	3	8	10	0	1.505	3.7	31	144	8.900
1,2,4 Trimethylbenzene	0.876	−44	120.19	18	3	9	12	0	1.504	4.1	48	168	11.243
Naphthalene	0.888	80	128.17	14	5	10	8	0	1.506	4.4	80	218	17.335

Table 5
Actual and predicted analytes boiling point in D3710 calibration mix by MRA.

Analytes	Predicted BP (°C)	Actual BP (°C)	Absolute error (°C)	% RE
n-Pentane	34.2	36.0	−1.8	−5
n-Hexane	66.0	69.0	−3	−4.4
2,4 Dimethyl pentane	95.4	80.0	15.4	19
n-Heptane	97.8	98.0	−0.2	−0.2
Toluene	110	111	−1	−0.9
n-Octane	121.7	126	−4.3	−3.4
p-Xylene	136	138	−2	−1.5
n-Propylbenzene	167	159	8	5.0
n-Decane	172	174	−2	−1.2
n-Butyl benzene	192	183	9	4.9
n-Decane	171	174	−3	−1.7
Dodecane	213	216	−3	−1.4
n-Tridecane	243	234	9	3.9
n-Tetradecane	253	253	0	0
n-Pentadecane	299	270	29	11
RMS%RE				***6.4***

Table 6
Actual and predicted analytes boiling point in MA VHP calibration mix by multivariate regression analysis.

Analytes	Predicted BP (°C)	Actual BP (°C)	Absolute error (°C)	% RE
n-pentane	44.2	36.0	8.2	23
Methyl-tert-butyl ether	66.1	55.5	11	19
Benzene	93.7	80.0	14	17
2,2,4 trimethyl pentane	93.3	99.0	−6	−5.8
Toluene	113	111	2	12
Ethylbenzene	128	136	−8	−6
p-Xylene	141	138	3	2
N-Nonane	146	151	−5	−3
o-Xylene	147	144	3	2
1,2,4-Trimethylbenzene	169	168	1	0.6
Naphthalene	211	218	−7	−3.2
RMS%RE				***10.8***

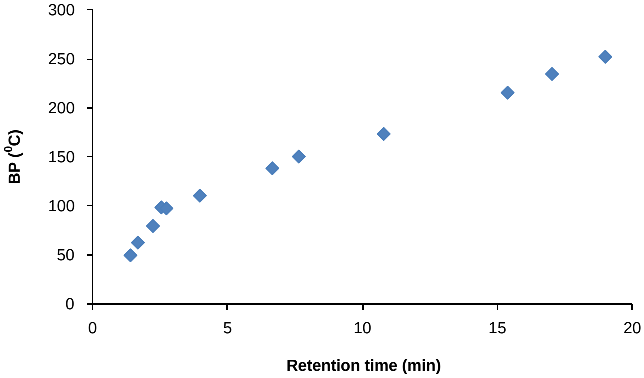

Fig. 3. Plot of analyte boiling point versus retention time of calibration mixture GC separation.

Table 7
Actual and predicted boiling point in D3710 calibration mix sample by univariate regression analysis.

Analytes	Predicted BP (°C)	Actual BP (°C)	Absolute error (°C)	% RE
n-Pentane	79.8	36	44	122
n-Hexane	86.6	69	17.6	26
2,4 Dimethyl pentane	89.8	80	9.8	12
n-Heptane	104	98	6	6
Toluene	122	111	11	9.9
n-Octane	124	126	−2	−2
p-Xylene	149	138	11	8.0
n-Propylbenzene	167	159	8	5
n-Decane	175	174	1	0.6
n-Butyl benzene	195	183	12	6.6
n-Decane	175	174	1	0.6
Dodecane	237	216	21	9.7
n-Tridecane	260	234	26	11
n-Tetradecane	280	253	27	11
n-Pentadecane	297	270	27	10
RMS%RE				***32.9***

Undoubtedly, more study is needed to further reduce the prediction *RMS%RE* of BP obtained using multivariate analysis. A lower *RMS%RE* may further be achieved by increasing the number of analytes in the calibration mixture sample. In addition, incorporation of more analyte structural activity relationship into the model calibration may also further reduce the prediction error of analytes boiling point. Nevertheless, this study demonstrated the superiority of the combined use of analytes GC retention times and analyte structural activity relationships in conjunction with multivariate regression analysis for analyte boiling point determination over the traditional use of univariate regression. The results and the outcome of the separation of crude oil samples is expected to remain the same as the separation of the gas mix described in this study. The use of MRA and principal component analysis (PCA) eliminate the noise or interference in a complex data set such as crude oil samples. The utility of experimental design and MRA for simultaneous optimization of chromatographic separation conditions to achieve efficient chromatographic separations of challenging and complex analytes have been established in our research group [28,29]. Studies from other labs have also demonstrated the combined use of analytical chemistry and MRA for accurate determination of moisture, color and pH in cooked, pre-sliced turkey hams with no interference from complex organic matrix [30]. Also, the use of analytical chemistry and MRA for accurate

Table 8
Actual and predicted analytes boiling point in MA VHP calibration mix by conventional univariate regression analysis.

Analyte	Predicted BP (°C)	Actual BP (°C)	Absolute error (°C)	% RE
n-Pentane	79.3	36	43	120
Methyl-tert-butyl ether	84.3	55.5	28.8	52
Benzene	97	80	17	21
2,2,4 trimethyl pentane	99.8	99	0.8	0.80
Toluene	122	111	11	9.9
Ethylbenzene	144	136	8	6
p-Xylene	148	138	10	7.2
N-Nonane	151	151	0	0
o-Xylene	153	144	9	6
1,2,4-Trimethylbenzene	177	168	9	5
Naphthalene	241	218	23	11
RMS%RE				***40.4***

determination of the quality parameters, including the protein, fat and moisture content and the free acidity of intact olive fruits [31] and other agricultural product with no sample preparation have been well documented [32,33]. Furthermore, the quantitative measurement of muscle oxygen saturation without influence from skin and fat spectroscopic analysis and in-vivo, noninvasive measurement of muscle pH during exercise [34,35] using analytical chemistry and multivariate regression analysis have been established. The reported strategy in this study is therefore promising and with further study, has a potential practical application for accurate petrochemicals BP determination. The method can be used as an alternative method to conventional fractional distillation for the determination of the boiling point distribution of gasoline products. Besides, the method will find practical application for quality control and quality assurance in oil and gas industries for rapid screening, pattern recognition and accurate petrochemicals BP determination.

4. Conclusion

In conclusion, this study explored the potential utility of a combined utility of GC–MS separation, analytes structural activity relationship (SAR), and multivariate regression analysis (MRA) for effective pattern recognition and accurate petrochemicals boiling point (BP) determination. The results of this study indicate that analyte SAR and GC data can be modeled to correctly predict petrochemicals BP, with low prediction error. Besides, while all variables are significant and contributed to the model predictive ability, molecular weight, flash point, melting point, and analyte retention time were found to be the most significant variables, influencing the performance of regression models for accurate BP prediction. The results of the validation studies conducted for D3710 and MA VHP demonstrated the superiority of multivariate regression over the conventional univariate regression analysis for accurate petrochemical BP determination from chromatographic data. The BP determination strategy reported in this study is more appealing and advantageous to traditionally fractional distillation strategy for petrochemicals' BP determination. First, this method is rapid and requires small sample size, reducing both the cost and analysis time. In addition, the method is robust, once the regression model is developed and properly calibrated it can be routinely used in a refinery for petrochemical BP determination without the need for recalibration, further reducing the cost and time of analysis. Furthermore, the technique can be potentially used for fast screening and pattern recognition of new petrochemical products.

Acknowledgments

The study was supported by NSF-HBCU-UP: Award # 0927905 (Start September 1, 2009 to August 31 2014).

References

[1] G. Boczkaj, A. Przyjazny, M. Kamiński, Anal. Bioanal. Chem. 399 (2011) 3253–3260.
[2] ASTM D1160, Test Method for Distillation of Petroleum Products at Reduced Pressure, American Society for Testing and Material, West Conshohocken, 2006.
[3] L.E. Green, L.J. Schumauch, J.C. Worman, Anal. Chem. 36 (1964) 1512–1516.
[4] ASTM D2887, Standard Test Method for Boiling Range Distribution of Petroleum Fractions by Gas Chromatography, American Society for Testing and Material, West Conshohocken, 2008.
[5] ASTM D3710, Test Method for Boiling Range Distribution of Gasoline and Gasoline Fractions by Gas Chromatography, American Society for Testing and Material, West Conshohocken, 1995.
[6] ASTM D5307, Standard Test Method for Determination of Boiling Range Distribution of Crude Petroleum by Gas Chromatography, American Society for Testing and Material, West Conshohocken, 1997.
[7] ASTM D7096, Standard Test Method for Determination of the Boiling Range Distribution of Gasoline by Wide-bore Capillary Gas Chromatography, American Society for Testing and Material, West Conshohocken, 2010.
[8] ASTM D6352, Standard Test Method for Boiling Range Distribution of Petroleum Distillates in Boiling Range from 174 to 700 °C by Gas Chromatography, American Society for Testing and Material, West Conshohocken, 2004.
[9] ASTM D7169, Standard Test Method for Boiling Point Distribution of Samples with Residues such as Crude Oils and Atmospheric and Vacuum Residues by High Temperature Gas Chromatography, American Society for Testing and Material, West Conshohocken, 2005.
[10] ASTM D7213, Standard Test Method for Boiling Range Distribution of Petroleum Distillates in the Boiling Range from 100 to 615 °C by Gas Chromatography, American Society for Testing and Material, West Conshohocken, 2005.
[11] Z. Garkani-Nejad, M. Ahmadvand, Chromatographia 73 (2011) 733–742.
[12] V.K. Guptaa, H. Khani, B. Ahmadi-Roudi, S. Mirakhorli, E. Fereyduni, S. Agarwal, Talanta 83 (2011) 1014–1022.
[13] S. Riahi, M.R. Ganjali, E. Pourbasheer, P. Norouzi, Chromatographia 67 (2008) 917–922.
[14] J. Ghasemi, S. Asadpour, A. Abdolmaleki, Anal. Chim. Acta 588 (2007) 200–206.
[15] S. Li, C. Sun, Y. Wang, S.Y.S. Xu, L. Wang, J. Environ. Sci. 14 (2002) 418–422.
[16] G.R. Moradi, A.A. Khoshmaram, M.R. Riazi, J. Pet. Sci. Eng. 76 (2011) 57–62.
[17] S.O. Fakayode, I.M. Swamidoss, M.A. Busch, K.W. Busch, Talanta 65 (2005) 838–845.
[18] K.W. Busch, I.M. Swamidoss, S.O. Fakayode, M.A. Busch, J. Am. Chem. Soc. 125 (2003) 1690–1691.
[19] K.W. Busch, I.M. Swamidoss, S.O. Fakayode, M.A. Busch, Anal. Chim. Acta 525 (2004) 53–62.
[20] M. Otto, Chemometrics, Wiley-VCH, Weinheim, 1999.
[21] K.R. Beebe, R.J. Pell, M.B. Seasholtz, Chemometrics – A Practical Guide, Wiley-Interscience, New York, 1998.
[22] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewis, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part B, Elsevier, Amsterdam, 1998.
[23] H. Martens, T. Naes, Multivariate Calibration, Wiley, Chicehster, 1989.
[24] E.R. Malinowski, Factor Analysis in Chemistry, Wiley-Interscience, New York, 1991.
[25] R. Kramer, Chemometric Techniques for Quantitative Analysis, Marcel Dekker, New York, 1998.
[26] M.A. Sharaf, D.L. Illman, B.R. Kowalski, Chemometrics, Wiley-Interscience, New York, 1986.
[27] K.W. Busch, M.A. Busch, in: K.W. Busch, M.A. Busch (Eds.), Chiral Analysis by Regression Modeling of Spectral Data in Chiral Analysis, Elsevier, Amsterdam, 2006.
[28] C.A. Luces, S.O. Fakayode, M. Lowry, I.M. Warner, Electrophoresis 29 (2008) 889–900.
[29] A.A. Williams, S.O. Fakayode, X. Huang, I.M. Warner, Electrophoresis 27 (2006) 4127–4140.
[30] I. Iqbal, D.W. Sun, P. Allen, J. Food Eng. 117 (2013) 42–51.
[31] L. Salguero-Chaparro, F. Peña-Rodríguez., LWT – Food Sci. Technol. 56 (2014) 363–369.
[32] D.A. Burns, D.A. Ciurczak (Eds.), Handbook of Near-Infrared Analysis, second ed.,Marcel Dekker, New York, 2001.
[33] M.J. Adams, Chemometrics in Analytical Spectroscopy, Royal Society of Chemistry, Cambridge, 1995.
[34] Y. Yang, O.O. Soyemi, P.J. Scott, M.R. Landry, S.M.C. Lee, L. Stroud, B.R. Soller, Opt. Express 15 (2007) 13715–13730.
[35] Y.E. Yang, O.O. Soyemi, M.R. Landry, B.R. Soller, Appl. Spectrosc. 16 (2007) 223–229.