Korean J. Chem. Eng., 34(10), 2715-2724 (2017)
DOI: 10.1007/s11814-017-0173-3

pISSN: 0256-1115
eISSN: 1975-7220

A modified scaled variable reduced coordinate (SVRC)-quantitative structure property relationship (QSPR) model for predicting liquid viscosity of pure organic compounds

Seongmin Lee[*,‡], Kiho Park[*,‡], Yunkyung Kwon[**], Tae-Yun Park[**], and Dae Ryook Yang[*,†]

*Department of Chemical and Biological Engineering, Korea University, Seoul 02841, Korea
**ChemEssen Inc., 812, 8th Floor. AceHighTechCity 2-Cha, 25 Seonyu-ro 13-gil, Yeongdeungpo-gu, Seoul 07282, Korea

(*Received 28 March 2017 • accepted 24 June 2017*)

Abstract−Liquid viscosity is an important physical property utilized in engineering designs for transportation and processing of fluids. However, the measurement of liquid viscosity is not always easy when the materials have toxicity and instability. In this study, a modified scaled variable reduced coordinate (SVRC)-quantitative structure property relationship (QSPR) model is suggested and analyzed in terms of its performance of prediction for liquid viscosity compared to the conventional SVRC-QSPR model and the other methods. The modification was conducted by changing the initial point from triple point to ambient temperature (293 K), and assuming that the liquid viscosity at critical temperature is 0 cP. The results reveal that the prediction performance of the modified SVRC-QSPR model is comparable to the other methods as showing 7.90% of mean absolute percentage error (MAPE) and 0.9838 of R^2. In terms of both the number of components and the performance of prediction, the modified SVRC-QSPR model is superior to the conventional SVRC-QSPR model. Also, the applicability of the model is improved since the condition of the end points of the modified model is not so restrictive as the conventional SVRC-QSPR model.

Keywords: Liquid Viscosity, Molecular Descriptor, QSPR, SVRC, Property Estimation, Modeling

INTRODUCTION

Liquid viscosity is one of the important physical properties in chemical processes, especially in modeling and optimization of the processes [1]. Physical properties of each material are required to estimate the performance in a chemical process or to optimize the operating conditions in model-based simulation approaches. However, huge efforts are required to measure the physical properties by an experimental approach, especially in some materials which are toxic, expensive, and unstable. If the physical properties can be estimated from the model which is based on the molecular structure of each material, the efforts for obtaining the properties could be significantly reduced.

Liquid viscosities of pure substances show a non-linear correlation with temperature changes (both under isobaric conditions and as saturated liquids). Within the temperature range from freezing point to nearby boiling point, the natural logarithm of liquid viscosity shows almost linear to the reciprocal of absolute temperature. Thus, usually at low temperature, the liquid viscosity can be predicted using the Andrade equation [2].

For a temperature range beyond the boiling point, the saturated liquid viscosities cannot be predicted by this kind of simple method. To overcome this problem, corresponding state theory (CST) was developed [3]. CST represents a generalization that equilibrium properties which depend on certain inter-molecular forces are related to the critical properties in a universal way. It is valid for liquids containing simple molecular structures that are not strongly polar or hydrogen-bonded. Some spherically-symmetric molecules (for example, CH_4) are well fitted by a two-constant law of corresponding states. However, non-spherical or strongly polar molecules are not fitted adequately by this method. Also, the CST based model requires additional information for modeling, such as normal boiling point, critical point, triple point, and acentric factor. Both the logarithm-based models and the CST based models show relatively poor predictions. Thus, these models are not enough to be utilized in developing accurate and widely applicable liquid viscosity prediction model.

For more accurate prediction, many researchers have suggested models which utilize the detailed chemical structural information of target compounds [4-7]. The group contribution (GC) method is a representative method of utilizing the chemical structure information for the prediction of physical properties. However, since the GC method uses only 2-dimensional structure information, the prediction performance of the method is not sufficient to be employed to the high quality prediction model [8]. Also, the GC method cannot distinguish isomer structures, and is not available if a compound contains a missing fragment which is not already defined in the GC method [9].

One of the proposed models as an alternative to the GC method is the quantitative structure property relationship (QSPR) model [8-15]. In the QSPR model, it is assumed that the physical proper-

[†] To whom correspondence should be addressed.
E-mail: dryang@korea.ac.kr

[‡] These authors contributed equally to this work.

[‡] This article is dedicated to Prof. Ki-Pung Yoo on his honorable retirement from Sogang University.

ties of each material are highly correlated with its characteristics of molecular structure. The molecular characteristics can be described as a form of molecular descriptors which are the quantified molecular structure information. The QSPR model is defined as a function of molecular descriptors to predict the physical properties. Generally, multi-linear regression or artificial neural networks (ANN) are used as regression models in the QSPR [16-19]. Within a narrow temperature range, the QSPR model can usually describe the physical property with high accuracy. Thus, although many QSPR models have shown sufficient estimation performance in predicting for diverse properties, the application of QSPR models has been restricted within limited temperature ranges [20,21]. As the temperature range becomes broader, the estimation performance is degraded significantly. For estimation of temperature-dependent properties, a different approach which can predict the properties with high accuracy should be devised. To develop the liquid viscosity model with high accuracy and over wide prediction range, an approach combining scaled variable reduced coordinate (SVRC) method with the QSPR model, which is called as SVRC-QSPR model, has been developed and utilized [11,22,23]. The SVRC method, which combines the corresponding state theory and the scaling law, is able to predict the liquid viscosity from triple point to critical point with high accuracy. However, the previous SVRC-QSPR model could not predict the liquid viscosity for the materials which do not have triple or critical point information.

We modified the SVRC-QSPR model for predicting liquid viscosity to extend the applicability of the model by replacing the initial and end points of the SVRC framework. The initial point was changed from triple point to 293 K, and the liquid viscosity at the final temperature (critical point) was assumed as 0 cP. With this approach, the limitation of the conventional SVRC model, which is a requirement of the information at the triple and critical points, can be overcome. The SVRC parameters were estimated by utilizing the QSPR methodology. From the modified SVRC-QSPR model, the prediction model for liquid viscosity with higher accuracy and wider applicability can be constructed.

METHODS

1. Data Set

The experimental data of liquid viscosity points were obtained from ThermoData Engine (TDE), which was developed by NIST for standard reference database. In the database, two selection criteria were designed to develop a high performance and robust prediction model for liquid viscosity. First, the liquid viscosity data set should range from 293 K to critical temperature. Second, at least five data points of a certain substance should exist within the temperature range. With these criteria, the database was constructed with including 2,450 of liquid viscosity data in 250 substances.

2. Model Development

2-1. Conventional SVRC Model Framework

In many previous studies, the SVRC was used to correlate the saturation properties of organic molecules from triple point to critical point (for example, vapor pressure, vapor and liquid densities and liquid viscosity) [11,24,25]. The model was developed based on the corresponding state theory and scaling law. By using this model, complex and nonlinear correlations between temperature and properties are represented on a universal line regardless of the chemical species. The generalized SVRC model is described as

$$\frac{Y_\infty^\alpha - Y^\alpha}{Y_\infty^\alpha - Y_0^\alpha} = \Phi(\varepsilon) \tag{1}$$

$$Y^\alpha = Y_0^\alpha \cdot \Phi(\varepsilon) + [1 - \Phi(\varepsilon)] \cdot Y_\infty^\alpha \tag{2}$$

$$\varepsilon = \frac{X_\infty - X}{X_\infty - X_0} \tag{3}$$

where Φ is the correlating function, X is the correlating variable (in this study, temperature), Y the saturation properties (in this study, liquid viscosity) at given X, Y_∞ is the asymptotic value of the saturation property at given X_∞, Y_0 is the initial value of the property at X_0, and α is the scaling exponent. In case of correlating properties between triple and critical point temperatures, Eq. (2) can be converted to

$$Y^\alpha = Y_T^\alpha \cdot \Phi(\varepsilon) + [1 - \Phi(\varepsilon)] \cdot Y_C^\alpha \tag{4}$$

If it is applied to liquid viscosity correlation, the equation is expressed as

$$\eta^\alpha = \eta_T^\alpha \cdot \Phi(\varepsilon) + [1 - \Phi(\varepsilon)] \cdot \eta_C^\alpha \tag{5}$$

where, η_C and η_T are the liquid viscosity at critical and triple points, respectively. The correlating function and scaling exponent value are defined as

$$\Phi(\varepsilon) = \frac{1 - A^{\varepsilon^B}}{1 - A} \tag{6}$$

$$\alpha = \alpha_C - \Delta\alpha \frac{\varepsilon(1 + C \cdot \varepsilon)}{1 + C} \tag{7}$$

$$\varepsilon = \frac{T_C - T}{T_C - T_T} \tag{8}$$

$$\Delta\alpha = \alpha_C - \alpha_T \tag{9}$$

where, A, B and C are the universal correlation constants, and α_C, α_T are the scaling exponent value at critical and triple temperature, respectively. The SVRC model explains the effect of chemical structure and temperature through the correlating function and scaling value.

2-2. Modified SVRC-QSPR Model

For estimation of physical properties by using the conventional SVRC model, the information of the properties at both triple and critical points is required as mentioned above. However, for certain materials, the information of the properties at triple or critical points is not available due to the absence of experimental data. To improve the applicability of the conventional SVRC model, the triple point is replaced as 293 K, which is an ambient temperature. Also, the value of liquid viscosity at critical point is assumed as 0 cP. Even though the liquid viscosity at the critical point is not actually 0, the supercritical fluid exhibits gas-like viscosity [26]. It implies that the liquid viscosity at the critical point is much lower than the liquid viscosity. Also, the measurement data of supercritical fluid have many uncertainties. Therefore, 0 cP at the critical point is assumed to simplify the model development, and not to violate the physical phenomenon at the same time. Then, the conventional SVRC model (Eq. (5)) can be reformulated as

$$\eta^{\alpha} = \eta_{am}^{\alpha} \cdot \frac{1-A^{\varepsilon^B}}{1-A} \tag{10}$$

$$\alpha = \alpha_C - \Delta\alpha \frac{\varepsilon(1+C\cdot\varepsilon)}{1+C} \tag{11}$$

$$\Delta\alpha = \alpha_C - \alpha_{am} \tag{12}$$

$$\varepsilon = \frac{T_C - T}{T_C - T_{am}} \tag{13}$$

where η_{am} is the liquid viscosity at 293 K, α_{am} is the scaling exponent at 293 K, T_{am} is the ambient temperature (293 K), and η_c is the liquid viscosity at critical point, which becomes zero in this modified model.

To develop the modified SVRC model, the scaling exponents at 293 K and critical temperature should be identified. Also, the initial liquid viscosity (at 293 K) should be characterized in all of the substances to predict the liquid viscosity of desired temperature. In this study, the SVRC parameters which include the liquid viscosities at 293 K (η_{am}), scaling exponents at 293 K (α_{am}) and at critical temperature (α_C), were predicted by utilizing QSPR model. Using this approach, the modified SVRC-QSPR model can be applied even at some materials which do not have the information of properties at both end points.

To construct the QSPR approach, molecular descriptors which convert the molecular structure information to a numerical value should be calculated and carefully selected. In this study, more than 1,900 descriptors were calculated by utilizing DragonX software. The descriptor selection and the number of descriptor optimization were conducted in the following section. The correlating function

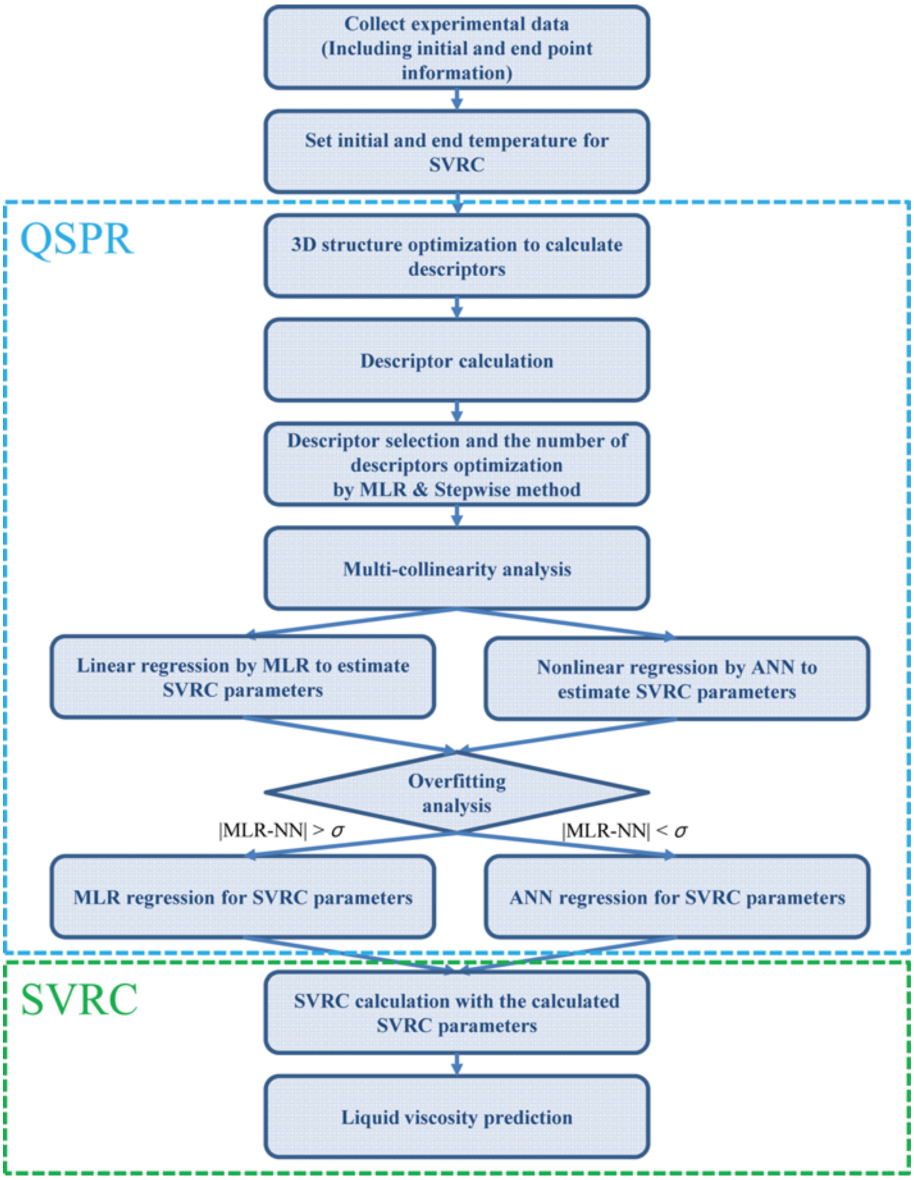
Collect experimental data (Including initial and end point information)
Set initial and end temperature for SVRC
QSPR
3D structure optimization to calculate descriptors
Descriptor calculation
Descriptor selection and the number of descriptors optimization by MLR & Stepwise method
Multi-collinearity analysis
Linear regression by MLR to estimate SVRC parameters
Nonlinear regression by ANN to estimate SVRC parameters
Overfitting analysis
|MLR-NN| > σ
|MLR-NN| < σ
MLR regression for SVRC parameters
ANN regression for SVRC parameters
SVRC
SVRC calculation with the calculated SVRC parameters
Liquid viscosity prediction

Fig. 1. Algorithm for developing the modified SVRC-QSPR model.

between the molecular descriptors and the SVRC parameters was designed. Multi-linear regression (MLR) for describing linear correlation between the molecular descriptors and the SVRC parameters and the ANN for describing non-linear correlation were selected as the correlating functions. By comparing the results of these function, the degree of nonlinearity of the correlating function can be identified, and an overfitting problem which can occur in the ANN model can be characterized and prevented.

3. Model Construction and Improvement

Since the modified SVRC-QSPR approach is based on the data mining analysis, relevant selection of descriptors, covariance characterization among the selected descriptors, and results validation by proper statistical testing should be performed. The detailed algorithm for developing the modified SVRC-QSPR model is presented in Fig. 1. In this section, it is introduced how the SVRC-QSPR model is constructed and improved by statistical methods.

3-1. Descriptor Selection and Optimization of the Number of Descriptors

To develop the QSPR model, relevant descriptors should be chosen as the input variables of the QSPR model. Generally, if more descriptors are included in the input variables, the prediction performance will be enhanced. However, too many descriptors can result in large computation load and overfitting. Thus, the number of descriptors should be optimized to reduce the computational time and undesirable noise fitting in the SVRC-QSPR model without loss of the performance of prediction at the same time. Since the determination coefficient (R^2) cannot consider the adverse effect of the number of input variables, adjusted R^2 was employed in this study to determine the optimal number of descriptors. The adjusted R^2 is defined as

$$R^2_{adj}=1-(1-R^2)\frac{(n-1)}{(n-k-1)} \tag{14}$$

where n is the number of liquid viscosity data in the training set, and k is the number of elements in the descriptor set.

In this study, descriptor selection and optimization of the number of descriptors were carried out by utilizing stepwise regression with forward selection. The MLR was used as a regression model in this step. At first, there were no descriptors in the input variable set. As adding new descriptor whose inclusion shows the most significant improvement in the performance of regression in the MLR model, the important descriptors were selected. The selection criterion was the highest partial correlation coefficient among all of the descriptors. This process was repeated by increasing the number of descriptors until there was no more increment in adjusted R^2. In each step, even if some descriptors were already selected in the previous step, these descriptors could be rejected if the partial correlation coefficients of the descriptors would decrease. Thus, all of the descriptors were newly checked at each step by comparing the partial correlation coefficient of each descriptor, and the top k most relevant descriptors were selected.

3-2. Multi-collinearity Analysis

Multi-collinearity is a phenomenon in which two or more explanatory variables in multiple regression models are highly correlated. Since one of the main issues with stepwise regression is that it is prone to selecting the descriptors with high covariance [27,28], multicollinearity analysis should be performed to remove the highly correlated descriptors with each other.

The multi-collinearity can be analyzed by variance inflation factor (VIF), which is defined as

$$\mathrm{VIF}=\frac{1}{1-R_j^2} \tag{15}$$

where R_j^2 is the coefficient of determination of a regression of variable j by all the other variables. This indicates that if some variables are highly correlated with the variable j, the R_j^2 of the variables would show high value, and it results in the high VIF. In this study, the selected descriptors after the previous step were analyzed by calculating VIF of each descriptor to estimate the degree of multi-collinearity. Generally, a VIF value showing over 10 indicates a multicollinearity problem [29]. Thus, if there were some descriptors showing VIF over 10, the descriptor with the highest VIF was rejected from the selected descriptor set. The VIF calculation was carried out one more time to check whether the multi-collinearity between the descriptors would be enhanced or not. Generally, if one or more descriptors with high VIF value are rejected, the VIF values of the other descriptors are reduced. This procedure was repeated until all of the VIF values of the remained descriptors in the input variable set were shown below 10.

3-3. ANN Model Training

In this study, ANN model was adopted to estimate the nonlinear correlation between the descriptors and the SVRC parameters. Since the ANN model should be trained appropriately to describe the correlation between input and output variables, the training set should be designed by random extraction from the whole data set to avoid skewness of data points in the training set. In the training step of the ANN model, the ratio of training, test for determining the number of hidden nodes, and validation sets is 6 : 2 : 2 for each prediction model.

The ANN model was constructed by utilizing commercial software (MATLAB R2016a and SPSS 23.0), and the Levenberg-Marquardt algorithm with error back propagation method was used to train and optimize the weighting factors in each node of the ANN model. To generate the best prediction performance of the ANN

Table 1. Detailed ANN structure to estimate the SVRC parameters in this study

Model	Number of input variables	Hidden layer			Output layer
		Number of hidden layers	Number of nodes	Activation function	Activation function
η_{am}	13	1	20	Sigmoid	Linear
α_{am}	9	1	13	Sigmoid	Linear
α_C	8	1	12	Sigmoid	Linear

model, each ANN training was repeated more than 100 times for estimating the most reliable SVRC parameters. The number of hidden nodes was decided based on the prediction performance with the test set and also considering calculation time. The detailed structure of each ANN model in this study is shown in Table 1.

3-4. Criteria to Prevent Misuse of Overfitted ANN Model

The training of the ANN model and the descriptor selection by stepwise regression can lead to the overfitting problem, which indicates that the model is too much trained in the training set to apply into the other data set. It should be avoided since the generality of the model could be disrupted. In this study, the overfitting parameter σ is defined as the maximum difference of mean error between the results of MLR model and ANN model in the training step. Thus, if the liquid viscosities of any new substances would be predicted by utilizing the developed model in this study, the results calculated by the MLR and ANN models are compared at first. Then, if the mean error difference between the MLR and ANN models is larger than the overfitting parameter (σ), it can be interpreted that the high prediction performance of the ANN model compared to the MLR is attributed to the overfitting problem. In this case, the MLR model is adopted to estimate the SVRC parameters and the liquid viscosities to avoid the loss of robust-

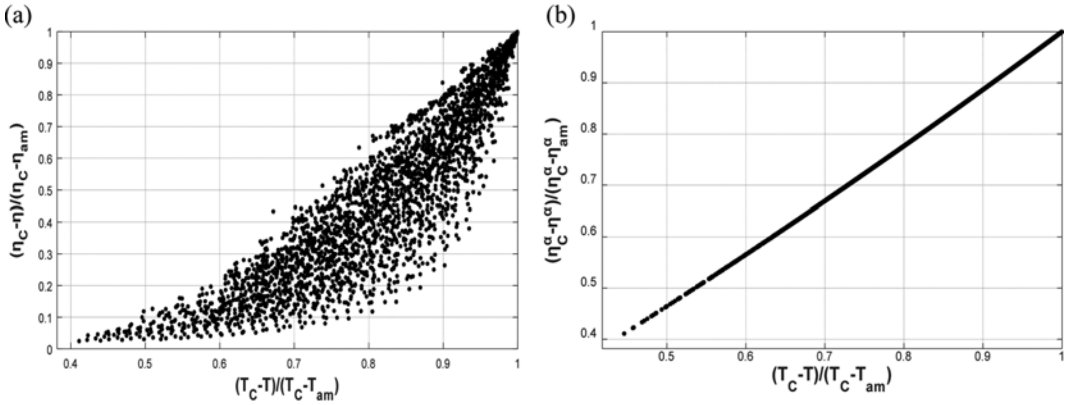

Fig. 2. Correlations between the reduced temperature and the reduced liquid viscosity (a) without the SVRC model and (b) with the SVRC model.

Table 2. Selected descriptor sets after MLR & stepwise method and multi-collinearity analysis to estimate the SVRC parameters

Liquid viscosity at 293 K (η_{am})	Scaling exponent at 293 K (α_{am})	Scaling exponent at critical temperature (α_C)
Wiener W index	Total structure connectivity index	Molecular walk count of order 02
Schultz MTI by valence vertex degrees	Lowest eigenvalue n. 1 of Burden matrix/ weighted by atomic masses	Avg electroph. react. index for a C atom
3D-MoRSE - signal 02/weighted by atomic masses	Balaban V index	Average valence connectivity index chi-4
Radial Distribution Function - 3.0/weighted by atomic van der Waals volumes	Relative number of double bonds	Maximal electrotopological negative variation
Molecular path count of order 10	Number of non-aromatic conjugated C (sp2)	Min 1-electron react. index for a C atom
Number of ring tertiary C (sp3)	Mean topological charge index of order 4	Min net atomic charge
Molecular path count of order 09	Number of double bonds	Molecular path count of order 08
Leverage-weighted autocorrelation of lag 5/ Weighted by atomic van der Waals volumes		H autocorrelation of lag 2/Weighted by atomic polarizabilities
Qyy COMMA2 value/weighted by atomic masses		
2nd component accessibility directional WHIM index/weighted by atomic Sanderson electronegativities		
WNSA-2 Weighted PNSA (PNSA2*TMSA/1000)		
Molecular profile no. 15		
Ghose-Viswanadhan-Wendoloski antiinfective-like index at 80%		

ness in the modified SVRC-QSPR model. Otherwise, the ANN model is employed for high performance of prediction.

The criterion was designed to prevent misuse of overfitted ANN model when the estimation model is used for any other components. The overfitting parameter implies the maximum degree of non-linearity in the liquid viscosity prediction model. If the training set contains a sufficiently large number of data points, the overfitting parameter will show the most non-linear correlation between the descriptors and experimental data. Thus, if the estimation results difference between MLR and ANN models were larger than the overfitting parameter, it could be concluded that the result of ANN model is not an accurate prediction, but due to the overfitting problem.

RESULTS AND DISCUSSION

The effect of the SVRC framework to the liquid viscosity is shown in Fig. 2. From the ambient temperature to the critical temperature, the correlation between the reduced temperature and the reduced liquid viscosity in Fig. 2(a) shows increasing tendencies by increasing temperature regardless of the kind of materials. However, the degree of increment in liquid viscosity as increasing the temperature is different at each material. Thus, Fig. 2(a) looks like a scattered plot. If appropriate scaling exponents could be applied in this correlation, all of the liquid viscosity data would be represented on the universal line regardless of the kind of materials as shown in Fig. 2(b). From the graph, it can be concluded that the liquid viscosity data from 293 K to critical point can be described by the SVRC model framework, and it revealed that the SVRC model can be applied in all of the materials if the appropriate scaling exponents would be found.

The selected descriptor sets for the SVRC parameters are listed in Table 2. The number of selected descriptors is 13, 8, and 9 for the liquid viscosity at 293 K, scaling exponent at 293 K, and scaling exponent at critical temperature, respectively. The selected descriptors were obtained by the MLR & stepwise method, and the descriptors which have high multi-collinearity with the other descriptors were already removed.

With the selected descriptors, the QSPR models for estimating the SVRC parameters were constructed. As shown in Fig. 3, the results of each QSPR model estimated the target data (experimental data) with great agreement of over 0.97 of R^2 in the case of η_{am} and α_{am}, and with a relatively good agreement of over 0.85 of R^2 in the case of α_C. Various statistical indexes such as sum of squared

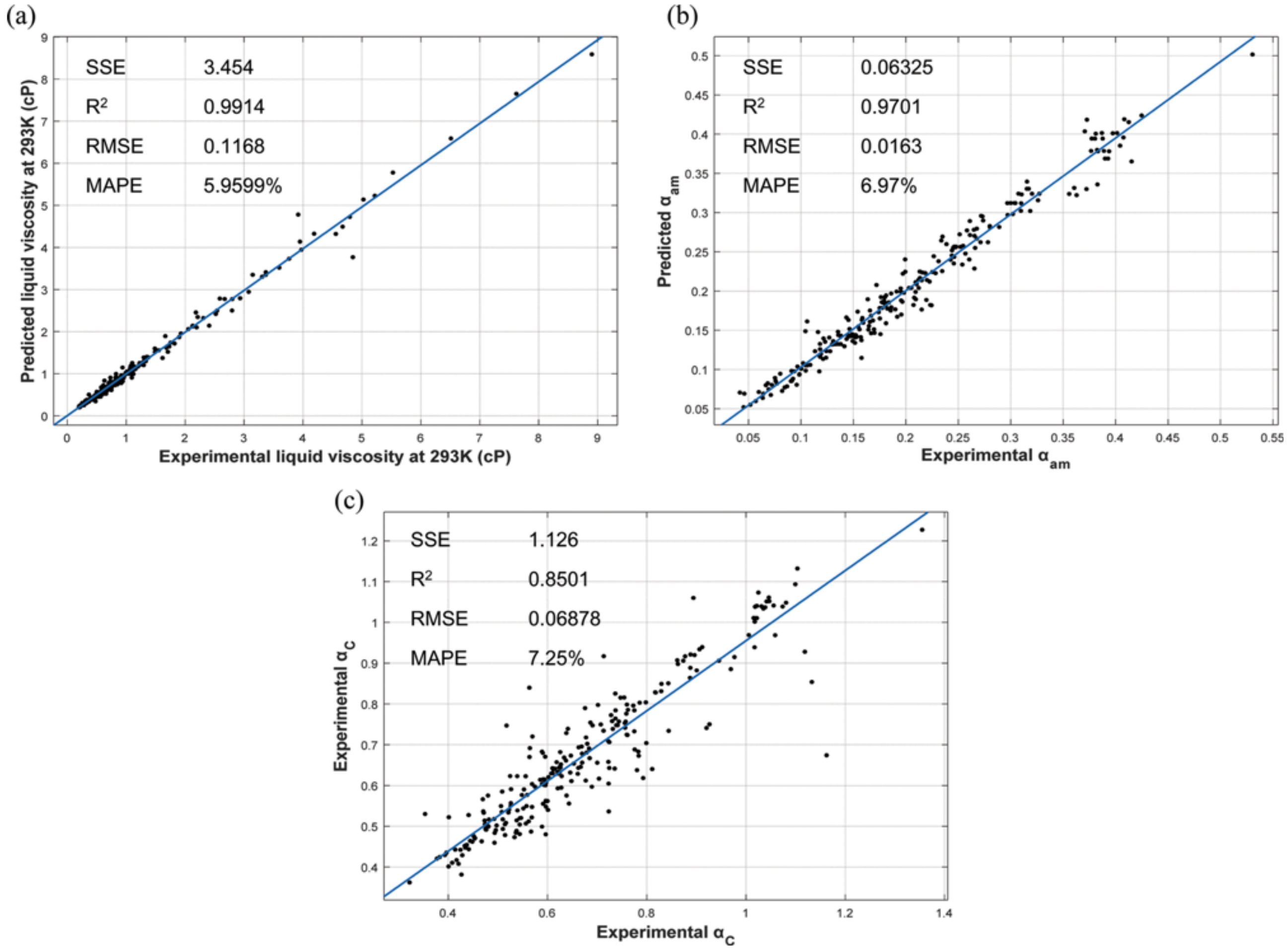

Fig. 3. Results of the QSPR models for estimating each SVRC parameter. (a) liquid viscosity at 293 K, (b) scaling exponent at 293 K, and (c) scaling exponent at critical temperature.

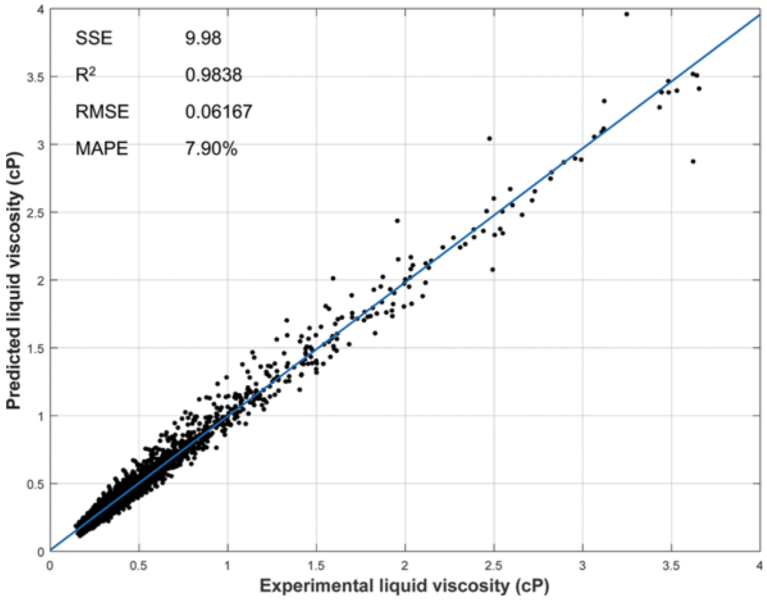

Fig. 4. Liquid viscosity prediction results from the modified SVRC-QSPR model with the experimental liquid viscosity data.

Table 3. Liquid viscosity prediction results of some materials as an example

CAS	Compound	Experimental viscosity (cP)	Predicted viscosity (cP)	Temperature (K)	Error (%)
1002-43-3	3-Mthylundecane	0.710581	0.723566	320.9869	1.827
1002-43-3	3-Mthylundecane	0.185642	0.191392	468.5107	3.097
1002-43-3	3-Mthylundecane	0.204133	0.212593	453.7583	4.144
1002-43-3	3-Mthylundecane	0.282781	0.296079	409.5012	4.703
1002-43-3	3-Mthylundecane	0.251762	0.264087	424.2535	4.896
110-83-8	Cyclohexane	0.589471	0.608815	301.6555	3.282
110-83-8	Cyclohexane	0.487384	0.514995	317.7216	5.665
110-83-8	Cyclohexane	0.410803	0.44219	333.7876	7.640
110-83-8	Cyclohexane	0.352004	0.384089	349.8537	9.115
110-83-8	Cyclohexane	0.32762	0.359211	357.8867	9.642
1120-36-1	1-Tetradecane	0.597224	0.578645	380.394	3.111
1120-36-1	1-Tetradecane	0.462679	0.437811	409.911	5.375
1120-36-1	1-Tetradecane	0.376077	0.341749	439.4281	9.128
1120-36-1	1-Tetradecane	0.317055	0.271632	468.9451	14.327
1120-36-1	1-Tetradecane	0.274965	0.217724	498.4621	20.818
13151-06-9	1-Methyl-1-octene	0.189227	0.210412	399.2722	11.195
13151-06-9	1-Methyl-1-octene	0.22703	0.255364	373.9543	12.480
13151-06-9	1-Methyl-1-octene	0.279686	0.315914	348.6364	12.953
13151-06-9	1-Methyl-1-octene	0.355996	0.402584	323.3186	13.087
13151-06-9	1-Methyl-1-octene	0.472089	0.536666	298.0007	13.679
13389-42-9	Trans-2-octene	0.227185	0.214923	372.6558	5.397
13389-42-9	Trans-2-octene	0.249717	0.235583	360.2922	5.660
13389-42-9	Trans-2-octene	0.308098	0.287735	335.5651	6.609
13389-42-9	Trans-2-octene	0.393084	0.362573	310.8379	7.762
13389-42-9	Trans-2-octene	0.450792	0.413752	298.4744	8.217
3875-51-2	Isopropylcyclopentane	0.330861	0.325128	364.8716	1.733
3875-51-2	Isopropylcyclopentane	0.30091	0.294998	376.6008	1.965
3875-51-2	Isopropylcyclopentane	0.275233	0.268586	388.33	2.415
3875-51-2	Isopropylcyclopentane	0.253059	0.245223	400.0593	3.097
3875-51-2	Isopropylcyclopentane	0.684318	0.658865	294.4963	3.720

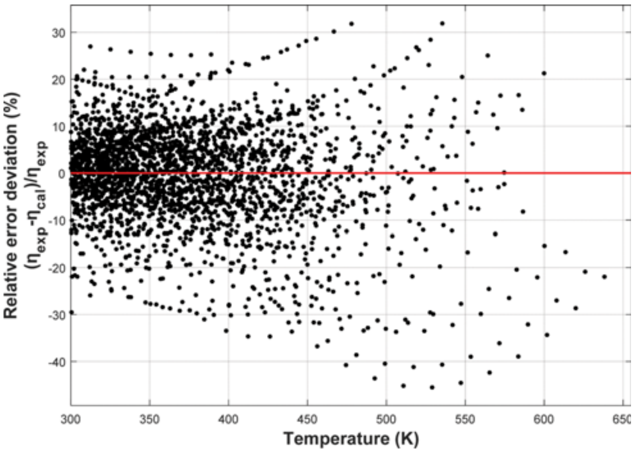

Fig. 5. Relative error deviation between the experimental data and the calculated results from the modified SVRC-QSPR model.

error (SSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) are displayed in Fig. 3. In terms of the MAPE, all of the QSPR models showed less than 8% error for estimating the SVRC parameters. Thus, the performance of estimating the SVRC parameters was sufficient to predict the liquid viscosity through the modified SVRC model framework.

The results of liquid viscosity prediction by the developed modified SVRC-QSPR model are presented in Fig. 4. The modified SVRC-QSPR model can predict more than 2,400 of liquid viscosity data in 250 materials with the performance of 7.90% of MAPE and 0.9838 of R^2. This result revealed that the liquid viscosity can be successfully predicted from the modified SVRC-QSPR model framework. The liquid viscosities of some materials and the prediction results by using the modified SVRC-QSPR model are displayed in Table 3 as an example.

The relative error deviation is shown in Fig. 5. Most of the data are distributed within ±10%. Even if some data points deviate considerably more than 20% of absolute percentage error, the number of these data are quite small compared to the overall data set. Thus, the overall MAPE is shown as 7.9%. However, the data points with high error percentages were concentrated on specific materials. In Fig. 6, the materials showing high prediction error are displayed with the results of the whole data set for comparison. Even if the comparison was conducted in terms of MAPE, the average errors are larger than the whole set even more than three-times in case of pentadecylbenzene. It reveals that a significant portion of the overall error was attributed to these limited number of materials. Also, as described in Table 4, the materials showing high prediction error contain a benzene ring in their chemical structure. Thus, it can be concluded that the liquid viscosity of some chemical compounds containing a benzene ring cannot be easily predicted from the modified SVRC-QSPR model. To improve the prediction performance, descriptors which can distinguish the detailed structure of functional groups attached in the benzene ring should be developed and contained in the modified SVRC-QSPR model.

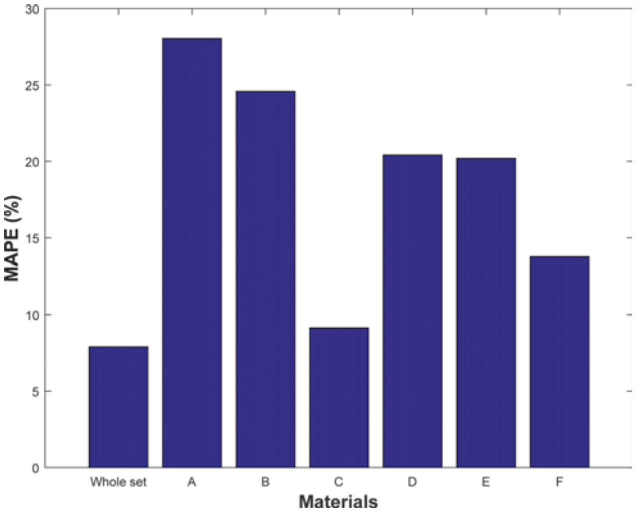

Fig. 6. MAPE comparison between the whole data set and the materials showing high prediction error percentages. The name of materials are A: pentadecylbenzene, B: tetradecylbenzene, C: beta-butylnaphthalene, D: n-tridecylbenzene, E: 1,2,3,5-tetraethylbenzene, and F: 1-phenylundecane.

Although the materials containing a benzene ring show relatively poor prediction results, the overall prediction performance of the modified SVRC-QSPR model is comparable to the results in the references, as shown in Table 5. The number of components is considerably larger than the other methods except the conventional SVRC-QSPR model. Considering both the number of components and the performance of prediction at the same time, the modified SVRC-QSPR model is quite adequate for liquid viscosity prediction compared to the other methods. Also, the modification in the SVRC model framework not only can extend the applicability of the model, but also improve the accuracy of the model at the same time. Since the experimental data at triple and critical tem-

Table 4. Detailed information about the materials showing high prediction error percentages

Substance	CAS	Temperature range	Number of data points	MAPE (%)	Remark
Pentadecylbenzene	2131-18-2	346-638 K	19	28.04	Benzene ring
Tetradecylbenzene	1459-10-5	335-626 K	19	24.58	Benzene ring
Beta-butylnaphthalene	1134-62-9	336-564 K	19	9.14	Naphthalene
N-tridecylbenzene	123-02-4	345-613 K	18	20.42	Benzene ring
1,2,3,5-Tetraethylbenzene	38842-05-6	359-527 K	17	20.19	Benzene ring
1-Phenylundecane	6742-54-7	341-586 K	17	13.18	Benzene ring
Whole set			2450	7.90	

Table 5. Comparison of the performance of prediction in liquid viscosity with the references

	Number of components	Temperature range	MAPE (%)
SVRC-QSPR model [22]	598	Triple-critical points	20.22
ANN model [30]	81	283 K-393 K	6.36
MLR-support vector machine (SVR) model [31]	45	280 K-380 K	3.95
QSPR model [32]	27	273 K-353 K	10
GC model [33]	29	293 K-393 K	8
The modified SVRC-QSPR model (in this study)	250	293 K-critical point	7.90

perature contain high uncertainties due to its extreme condition, the modification of the model in this study is able to utilize the experimental data with low uncertainties at 293 K instead of the triple point. Therefore, the modified SVRC-QSPR model shows superior performance of prediction compared to the conventional SVRC-QSPR model.

The model can be utilized by the following procedure. If someone wants to know the liquid viscosities of some materials, the selected descriptors of the materials should be calculated as shown in Table 2. From the descriptors, η_{am}, α_{am}, and α_C are calculated by using MLR and ANN models. If the difference between the calculation results by MLR and ANN models would not be larger than the overfitting parameter, the results and ANN model would be selected. Then, the liquid viscosity prediction could be carried out from Eqs. (10)-(14).

CONCLUSIONS

A modification of the conventional SVRC-QSPR model for liquid viscosity prediction has been proposed. Since there is little experimental data at triple and critical points, the conventional SVRC-QSPR model has difficulties in training the model parameters from the experimental data. By changing the initial point from the triple point to 293 K, and assuming the liquid viscosity at the critical point as 0 cP, the applicability of the SVRC-QSPR model can be improved. The liquid viscosity prediction results of the modified SVRC-QSPR model showed 7.90% of MAPE and 0.9838 of R^2. Most of the significant errors are attributed to the substances which contain a benzene ring in their chemical structure. Even if these substances showing high prediction error are contained in the data set, the overall performance of the modified SVRC-QSPR model is comparable with the other methods. In terms of both the number of components and the performance of prediction at the same time, the modified SVRC-QSPR model is relatively superior to the other methods.

ACKNOWLEDGEMENTS

This research was respectfully supported by Engineering Development Research Center (EDRC) funded by the Ministry of Trade, Industry & Energy (MOTIE) (No. N0000990), and Korea University.

REFERENCES

1. B. E. Poling, J. M. Prausnitz and J. P. O'Connell, *The properties of gases and liquids*, Mcgraw-hill, NY (2001).
2. T. Ghosh, D. Prasad, N. Dutt and K. Rani, *Viscosity of liquids: Theory, estimation, experiment, and data*, Springer, NY (2007).
3. M. Hobson and J. H. Weber, *AIChE J.*, **2**, 354 (1956).
4. L. Constantinou, R. Gani and J. P. O'Connell, *Fluid Phase Equilib.*, **103**, 11 (1995).
5. H. S. Elbro, A. Fredenslund and P. Rasmussen, *Ind. Eng. Chem. Res.*, **30**, 2576 (1991).
6. B. H. Park, M. S. Yeom, K.-P. Yoo and C. S. Lee, *Korean J. Chem. Eng.*, **15**, 246 (1998).
7. J. Park and D. Paul, *J. Membr. Sci.*, **125**, 23 (1997).
8. D. Sola, A. Ferri, M. Banchero, L. Manna and S. Sicardi, *Fluid Phase Equilib.*, **263**, 33 (2008).
9. P. R. Duchowicz, A. Talevi, L. E. Bruno-Blanch and E. A. Castro, *Biorg. Med. Chem.*, **16**, 7944 (2008).
10. Y. Dadmohammadi, S. Gebreyohannes, B. J. Neely and K. A. Gasem, *Fluid Phase Equilib.*, **409**, 318 (2016).
11. S. S. Godavarthy, R. L. Robinson and K. A. Gasem, *Fluid Phase Equilib.*, **246**, 39 (2006).
12. A. R. Katritzky, V. S. Lobanov and M. Karelson, *Chem. Soc. Rev.*, **24**, 279 (1995).
13. H. Maadani, M. Salahinejad and J. Ghasemi, *SAR QSAR Environ. Res.*, **26**, 1033 (2015).
14. B. Wang, L. Zhou, K. Xu and Q. Wang, *Ind. Eng. Chem. Res.*, **56**, 47 (2017).
15. L. C. Yee and Y. C. Wei, *Current modeling methods used in qsar/qspr*, Wiley-VCH: Weinheim, Germany (2012).
16. L. S. Aiken, S. G. West and S. C. Pitts, *Multiple linear regression: Testing and interpreting interactions*, Sage, CA (1991).
17. Y. Ammi, L. Khaouane and S. Hanini, *Korean J. Chem. Eng.*, **32**, 2300 (2015).
18. A. A. Babaei, A. Khataee, E. Ahmadpour, M. Sheydaei, B. Kakavandi and Z. Alaee, *Korean J. Chem. Eng.*, **33**, 1352 (2016).
19. E. Mohagheghian, H. Zafarian-Rigaki, Y. Motamedi-Ghahfarrokhi and A. Hemmati-Sarapardeh, *Korean J. Chem. Eng.*, **32**, 2087 (2015).
20. M. Luckas and K. Lucas, *AIChE J.*, **32**, 139 (1986).
21. W. D. Monnery, W. Y. Svrcek and A. K. Mehrotra, *Can. J. Chem. Eng.*, **73**, 3 (1995).
22. A. Jegadeesan, *Structure-based generalized models for selected pure-fluid saturation properties*, Oklahoma State University, M.S. Thesis (2006).
23. R. D. Shaver, *New scaled-variable-reduced-coordinate framework for correlation of pure fluid saturation properties*, Oklahoma State University, M.S. Thesis (1990).

24. R. Shaver, R. Robinson and K. Gasem, *Fluid Phase Equilib.*, **64**, 141 (1991).
25. R. Shaver, R. Robinson and K. Gasem, *Fluid Phase Equilib.*, **78**, 81 (1992).
26. M. McHugh and V. Krukonis, *Supercritical fluid extraction: Principles and practice*, Elsevier (2013).
27. W. Sauerbrei and M. Schumacher, *Stat. Med.*, **11**, 2093 (1992).
28. E. W. Steyerberg, M. J. Eijkemans and J. D. F. Habbema, *J. Clin. Epidemiol.*, **52**, 935 (1999).
29. R. M. O'brien, *Quality & Quantity*, **41**, 673 (2007).
30. N. Dutt, Y. Ravikumar and K. Y. Rani, *Chem. Eng. Commun.*, **200**, 1600 (2013).
31. Y. Zhao, X. Zhang, L. Deng and S. Zhang, *Comput. Chem. Eng.*, **92**, 37 (2016).
32. B.-K. Chen, M.-J. Liang, T.-Y. Wu and H. P. Wang, *Fluid Phase Equilib.*, **350**, 37 (2013).
33. R. L. Gardas and J. A. Coutinho, *Fluid Phase Equilib.*, **266**, 195 (2008).