

PCCP

Accepted Manuscript

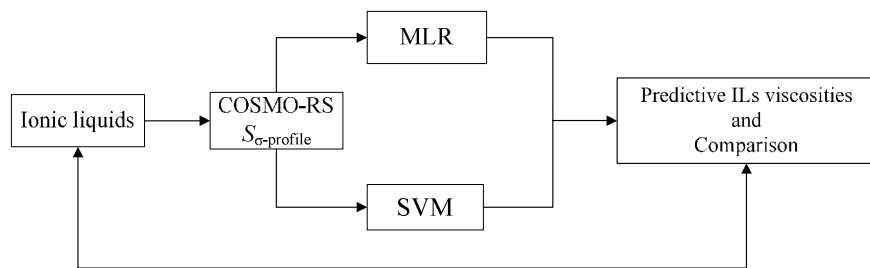


This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



QSPR study for ILs using MLR and SVM algorithms based on COSMO-RS molecular descriptors ($S_{\sigma\text{-profile}}$)

Quantitative prediction of viscosity of ionic liquids using $S_{\sigma\text{-profile}}$ molecular descriptors

Yongsheng Zhao^{a,b}, Ying Huang^{a,b}, Xiangping Zhang^{*,a}, Suojiang Zhang^{*,a}

^a Beijing Key Laboratory of Ionic Liquids Clean Process, State Key Laboratory of Multiphase Complex Systems, Key Laboratory of Green Process and Engineering, Institute of Process Engineering, Chinese Academy of Sciences, 100190, Beijing, China

^b School of Chemistry and Chemical Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: In this study, two novel QSPR models are developed to predict the viscosity of ionic liquids (ILs) using multiple linear regression (MLR) and support vector machine (SVM) algorithms based on Conductor-like Screening Model for Real Solvents (COSMO-RS) molecular descriptors ($S_{\sigma\text{-profile}}$). A total data set of 1502 experimental viscosity data points under a wide range of temperature and pressure for 89 ionic liquids (ILs), is employed to train and verify the models. The Average Absolute Relative Deviation (AARD) of total data set of the MLR and SVM is 10.68% and 6.58%, respectively. The results show that both the MLR and SVM models can predict the viscosity of ILs, and the performance of nonlinear model developed by the SVM is superior to the linear model (MLR). Furthermore, the derived models also can throw some light onto what structural characteristics are related to the viscosity of ILs.

1. Introduction

Ionic liquids (ILs) are a novel class of room temperature molten salts composed of only organic cations and inorganic or organic anions, and have attracted enormous interests as the hot and state-of-the-art topic in green chemistry^{1,2}. ILs have many unique properties, as diverse as thermal and electrochemical stability, negligible vapor pressure, non-inflammability, and tunable property. In this context, ILs are ubiquitously applied in various fields of modern chemistry including organic synthesis, catalytic, extraction separation, and electrochemistry, etc³.

Viscosity is one of the most important properties of ILs, which has great influences on the transfer performance of the IL containing system. However, because the combination of cation and

*Corresponding author. Tel./Fax: +86 01062558174

E-mail address: xpzhang@ipe.ac.cn, sjzhang@ipe.ac.cn

anion can yield 10^{18} IL systems¹, it is experimentally time consuming, expensive and impossible by measuring all possible ILs to screening optimum IL for special purpose. Furthermore, to the best of our knowledge, unlike other properties, such as density, the comparable theoretical basis for the estimation of liquid viscosity has not been elucidated yet⁴. Hence, developing a cost-effective, accurate and reliable method for predicting the viscosity of ILs is important and timely.

Currently, many models⁵⁻¹⁷ have been developed to predict the viscosity of ILs, and some have been good reviewed by Coutinho et al¹⁸. Among these, Group Contribution (GC) and Quantitative Structure-Property Relationship (QSPR) methods are most commonly applied. GC is a very important method to be utilized to predict a variety of physical and thermodynamic properties. The main advantage of GC is its simplicity and capability to give a reasonable accuracy if all the necessary group increments are obtained from the experimental data⁶. In previous study, Gardas and Coutinho¹⁶ proposed a GC method to predict the viscosity of ILs based on the Orrick-Erbar-type equation. The proposed model can be used to predict the viscosity of new ILs in wide ranges of temperature and can be extended to a larger range of ILs. However, the model has a drawback that it needs knowledge of density of ILs. Therefore, they proposed a new GC method using the Vogel-Tammann-Fulcher (VTF) equation to solve this problem¹³. The later proposed model presented an average deviation of 7.5%, and could easily be extended to other ILs. Recently, Padaszyński et al⁵ proposed a new GC model based on feed-forward artificial neural network (FFANN). The overall relative deviation of the whole data set is 11.4%. Although a large set of 1484 ILs covered a wide temperature (253-573 K) and pressure (0.06-350 Mpa) ranges were studied, the FFANN model requires as many as 242 descriptors.

QSPR is an effective method to connect physical or chemical properties to a set of molecular descriptors, which is not only capable of estimating and screening the compound with the desired property, but also is capable of uncovering the underlying relationship between micro-structure and macro-property^{19,20}. QSPR has been profusely employed to investigate the properties of ILs, e.g., melting point²¹⁻²⁵, density^{26,27}, refractive indices²⁸, gas solubilities in ILs²⁹, surface tension^{26,30}, and toxicity^{31,32}, etc. In order to progress in predicting the viscosity of ILs, Tochigi et al³³ proposed a QSPR model to predict IL viscosity, but the correlation coefficient R^2 of the data set was below 0.9, and the deviation between experimental values and calculated values is too large

for some systems. In Han's study³⁴, the CODESSA software was used to build four QSPR models, but the established models have some disadvantages. Firstly, only a small amount of data of imidazolium-based ILs was used to build the model. Secondly, the established models can only be adopted to predict the viscosity at 298.15K. Mirkhani et al³⁵ presented a GFA-based QSPR model for predicting the viscosity of ILs. A large set of 435 data points of 293 ILs based on imidazolium, pyridinium, piperidinium, pyrrolidinium, pyrroline, oxazolidinium, ammonium, phosphonium, sulfonium, guanidinium, amino acids, morpholinium, isoquinolinium were studied. The involved data covered a wide viscosity range (5.7-2824 cP) and temperature range (253-273 K), but the correlation coefficient R^2 is relatively low as 0.8096, and the absolute average deviations (AAD) is 8.77%. Recently, Chen et al²⁰ applied CODESSA software to develop eight QSPR models under different temperatures (283, 293, 298, 303, 313, 323, 333, 343 K). All the correlation coefficients (R^2) of the eight models are above 0.82. However, these models can only be suitable to ILs whose anions are [Tf₂N]⁻, and thus its application is limited.

Klamt and co-workers³⁶⁻³⁹ put forward the Conductor-like Screening Model for Real Solvents (COSMO-RS) approach and developed the COSMOtherm software. In the COSMOtherm, the viscosity is predicted using QSPR model which employs the computed sigma-moments and other parameters as independent variables. Torrecilla et al⁴⁰ pioneered in the use of COSMO-RS for predicting the viscosity of ILs, but the results are unsatisfactory, one possible reason is that the coefficients of the QSPR model are derived from a set of 175 room temperature viscosities of organic liquids instead of ionic liquids. Moreover, the QSPR model can only predict the viscosity under specific temperature due to it does not include a temperature dependency term. All afore-mentioned QSPR investigations are performed only at atmospheric pressure so far. Although the viscosity of ILs at low pressure is influenced mainly by temperature, the influence of pressure at higher ranges becomes crucial for particular purposes⁴¹. Thus, it is very significant to develop and create new models to investigate the viscosity of different ILs at wider temperature and pressure ranges.

Distribution area of σ -profile ($S_{\sigma\text{-profile}}$) is an a priori quantum-chemical descriptor that quantitatively represents the molecule's polar surface screen charge on the polarity scale, and can be achieved from the histogram function σ -profile given by COSMO-RS computation. It has been used effectively as a parameter in QSPR/QSAR models to predict the density⁴² and toxicity³² of

ILs, and the advantages have been proved. Therefore, in this work, $S_{\sigma\text{-profile}}$ is utilized as a parameter to predict the viscosity of ILs.

QSPR can be implemented in two ways: linear and non-linear algorithms. The linear approach is very easy and can intuitively show the impact of each parameter on properties due to the equation parameters have clear meanings. However, it is incapable of processing complex problems. Considering many real-world scenarios cannot be simplified to linear problems, non-linear algorithms are more suitable for accurate prediction. SVM is an increasingly popular non-linear algorithm in various fields of chemistry which can effectively avoid the over-fitting problem of artificial neural networks (ANN), and thus is widely used in QSPR due to its excellent generalization performance⁴³.

In this work, firstly, a large amount of viscosity data was collected from literature under a wide range of temperature and pressure. Secondly, the multiple linear regression (MLR) and SVM algorithms are utilized to construct the linear and nonlinear QSPR models for predicting the viscosity of ILs based on the $S_{\sigma\text{-profile}}$ descriptors. The characteristic parameters that have great impact on the viscosity of ILs were screened, and the performances of the obtained models were investigated and verified.

2. Data and methods

2.1. Viscosity data points and $S_{\sigma\text{-profile}}$ of ILs

In the present work, all the experimental data points are collected from IL Thermo Database⁴⁴. A total of 1502 experimental viscosity data points of 89 ILs mainly based on alkyl-substituted cations, including imidazolium [Im]⁺, piperidinium [Pip]⁺, pyrrolidinium [Pyr]⁺, pyridinium [Py]⁺, morpholinium [Mor]⁺, ammonium [N]⁺, phosphonium [P]⁺, Sulfonium [S]⁺, and different anions, such as hexafluorophosphate [PF₆]⁻, tetrafluoroborate [BF₄]⁻, nitrate [NO₃]⁻, acetate [Ac]⁻, ethylsulfate [EtSO₄]⁻, pentafluoroethyltrifluoroborate [C₂F₅BF₃]⁻, n-butylsulfate [C₄SO₄]⁻, dicyanamide [DCA]⁻, bis(trifluoromethylsulfonyl)imide [BTI]⁻, trifluoromethyltrifluoroborate [CF₃BF₃]⁻, octylsulfate [C₈SO₄]⁻, bis(pentafluoroethylsulfonyl)imide [BETI]⁻, thiocyanate [SCN]⁻, trifluoromethylsulfonate [TfO]⁻, methylsulfate [MeSO₄]⁻, trifluoroacetate [TfA]⁻, halide [X]⁻, are investigated. The collected viscosity data points (8.28-142000 cP) cover a wide range of pressure (1-3000 bar) and temperature (253.15-395.32 K) (see Table 1). To develop a reliable model, all the

selected experimental data points are divided into a training set of 1205 data points to develop the models and a test set of the remained 297 data points to verify its prediction performance. Each category (nearly 80% of the data as training set, 20% as test set) of the total data points is separated at random. The detailed information of ILs investigated in this work are listed in Supporting Information.

COSMO-RS is a new method for quantitative calculation of solvation phenomenon on the basis of continuous medium solvation theory in combination with statistical mechanics method. $S_{\sigma\text{-profile}}$ represents molecule area of surface screening charge density, which is an a priori two-dimensional quantum-chemical parameter to characterize electronic structure and molecular size of ILs. The detailed introduction of the $S_{\sigma\text{-profile}}$ can be found in literature³². All the $S_{\sigma\text{-profile}}$ of ILs in this study are taken from the COSMO IL database, which were calculated in terms of the BP-TZVP quantum chemical level³⁸.

Table 1 Temperature, pressure, viscosity range, and data points for different ILs

No.	Class	Temperature (K)	Pressure (KPa)	Viscosity (cP)	Data points
1	Imidazolium	258.15-395.32	100-300000	7.83-142000	1380
2	Pyridinium	283-353.15	101.325	10-464.49	62
3	Pyrrolidinium	283.1-353.1	101.325	13.1-167.8	20
4	Phosphonium	293.15-303.15	101.325	268.63-2077.91	8
5	Ammonium	283.1-353.1	101.325	14.1-1017	17
6	Morpholinium	298.15	101.325	466-1035	2
7	Piperidinium	298.15	101.325	102-456	5
8	Sulfonium	253.15-313.15	101.325	12.1-120	15

*All the literature sources of data points can be found from Supporting Information

2.2. Multiple linear regression (MLR) algorithm

MLR⁴⁵ is a commonly used algorithm within QSPR to quantify the relationship between several independent x and a dependent variable y . The following information can be achieved from the general MLR form given in equation (1): which independent has great impact on dependent variable; direction of impact for each independent; the model can be used to anticipate the value of dependent variable y when the value of current independent variable is given.

$$\ln(y) = b_1x_1 + b_2x_2 + \dots + b_nx_n + b_0 \quad (1)$$

where b_1 to b_n are the regression coefficients of the descriptors and b_0 is the intercept. The

parameters (x_1 to x_n) comprised in the equation (1) are employed to explain the chemical structure of compounds, and n is the number of the parameters to seek the best model.

2.3. Support Vector Machine (SVM) algorithm

SVM was originally proposed by Cortes and Vapnik et al⁴⁶, which is a relative new and very promising classification and regression algorithm based on the statistical learning theory and structural risk minimization principle. In the SVM, the input data is first mapped to high-dimensional feature space through kernel function, and then the linear regression will happen in the feature space⁴⁷. The nonlinear feature mapping allows the treatment of nonlinear problem in a linear space. More detailed introduction of the SVM theoretical basis can be referred in some excellent books and tutorials^{48, 49}.

2.4. Evaluation of the model performance

Model performance can be measured by different metrics. R^2 , which gives the fraction of explained variance for a data set, is utilized to evaluate the model's fit performance. Average absolute relative deviation (AARD) and mean squared error (MSE) are applied to measure the model's predictive effectiveness, as shown by the following equations:

$$\text{AARD (\%)} = 100 \times \sum_{i=1}^{N_p} \left| \frac{y_i^{\text{cal}} - y_i^{\text{exp}}}{y_i^{\text{exp}}} \right| / N_p \quad (2)$$

$$\text{MSE} = \frac{\sum_{i=1}^{N_p} (y_i^{\text{cal}} - y_i^{\text{exp}})^2}{N_p} \quad (3)$$

Where y_i^{exp} is the experimental property, y_i^{cal} is the actual prediction of the models, and N_p is the number of compounds of the data set.

3. Results and discussion

3.1. Results of MLR model

Based on the collected ILs viscosity data points, the MLR equation (4) is established by using stepwise regression algorithm:

$$\begin{aligned} \text{Log}_{10}(\eta / \text{cp}) = & -0.0177T - 0.047S_{A0.004} + 0.072S_{A0.02} + 3.82 \times 10^{-6}P + 0.036S_{C0.003} \\ & - 0.016S_{A0.012} + 43.091S_{A-0.022} + 7.425 \end{aligned}$$

(4)

$$(n=1205, R^2=0.803, S=0.2616, F=699.199)$$

where η stands for the viscosity of ILs, S represents distribution area with a specific surface screening charge density, subscript A is anion, C is cation, and the numerical value after A or C represents surface screening charge density of ILs. T and P are the temperature and pressure respectively. As shown in Table 2, all of the linear correlation coefficient of any two parameters is less than 0.8, which suggests that there is no strong linear relation between the descriptors⁵⁰. It is observed that all the descriptors contained in the MLR model have physical meaning, and these parameters can account for structural features affecting the viscosity of the ILs studied.

Table 2 Correlation matrix of seven descriptors

	T	$S_{A0.004}$	$S_{A0.02}$	P	$S_{C0.003}$	$S_{A0.012}$	$S_{A-0.022}$
T	1						
$S_{A0.004}$	0.098	1					
$S_{A0.02}$	0.069	0.086	1				
P	0.137	0.086	0.082	1			
$S_{C0.003}$	0.004	0.101	0.079	0.077	1		
$S_{A0.012}$	0.094	0.077	0.236	0.115	0.089	1	
$S_{A-0.022}$	0.079	0.053	0.025	0.036	0.124	0.048	1

In equation (4), positive sign in front of a parameter means positive correlation between the parameter and ILs viscosity, and negative sign stands for negative correlation. The structure descriptors in equation (4) have been arranged in the descending order of t-test values so that the most important one comes first. Thus, T is the most important descriptor, which is in agreement with reference⁵¹ that ILs's viscosities are very sensitive to temperature. The negative sign before T reveals that the temperature contributes negatively to the viscosity of ILs. The second descriptor is $S_{A0.004}$ in the non polar range ($-0.0082 < \sigma < 0.0082 \text{ e \AA}^{-2}$) having a negative sign, and thus indicates that the values of viscosity decrease as $S_{A0.004}$ increase. This might be used to explain some experimental phenomena for e.g., as shown in Figure 1. When ILs have the same cation $[\text{Bmim}]^+$, the volume trend of these anions is as $[\text{BF}_4]^- < [\text{CF}_3\text{BF}_3]^- < [\text{C}_2\text{F}_5\text{BF}_3]^-$, and the viscosity of these ILs shows as follows: $[\text{BF}_4]^- > [\text{CF}_3\text{BF}_3]^- > [\text{C}_2\text{F}_5\text{BF}_3]^-$, which is consistent with the order of $S_{A0.004}$ of these anions. The reason may be that with the decrease of the volume of anion, the

negative charge of anion cannot be effectively distributed, so that the coulomb force increasing, which causes the increase of the viscosity. The next descriptor is $S_{A0.02}$ which is the surface of the 0.02 molecule surface screening charge density of anion. The charge density value (0.02) of anion is higher than the cutoff 0.0082, which implies these anionic fragments may contribute to hydrogen bonds as acceptors. Thus the descriptor $S_{A0.02}$ manifests the ability of accepting hydrogen bonds of anion which has a positive influence on the viscosity of ILs. The fourth descriptor is P , the positive sign of this descriptor discloses that the pressure contributes positively to viscosity. The fifth descriptor is $S_{C0.003}$, and the charge density 0.003 of cation located at non polar range too, thus the surface of non polar charge density of cation has a positive impact of viscosity of ILs. Finally, the last two descriptors, $S_{A0.012}$ and $S_{A-0.022}$, are both at polar range, and thus indicates the surface of polar charge density of anion has also certain influence on viscosity of ILs.

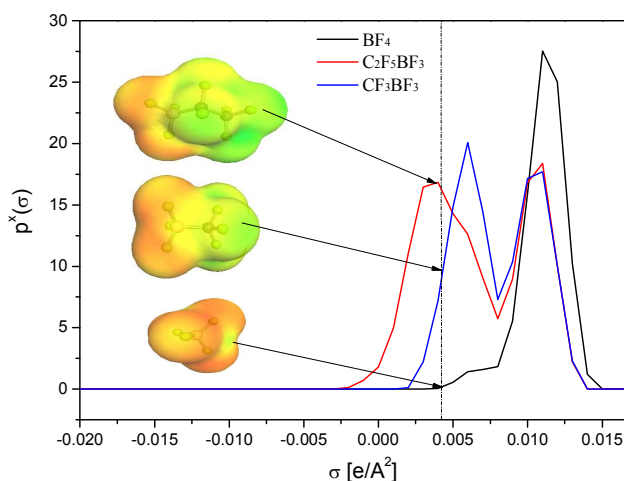


Fig. 1. Surface polarization charge density and σ -profile of representative anions of [Bmim] ionic liquids

From the foregoing, some results are obtained: (1) the interionic electrostatic interaction plays a more important role on the viscosity of ILs than hydrogen bonding interaction, and similar result was obtained by literature³⁴. (2) the surface of non polar charge density of ILs is more critical than the surface of polar charge density.

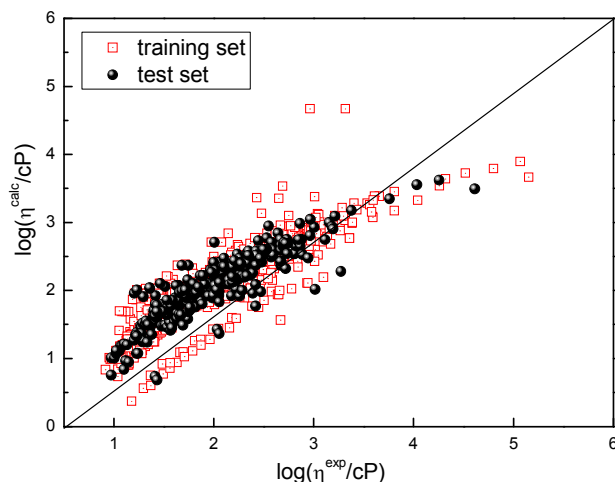


Fig. 2. Calculated versus experimental viscosities using the MLR algorithm for all ionic liquids

Equation (4) with $R^2=0.803$, $\text{MSE}=0.171$ which is established by 1250 data points could be useful for predicting the viscosity of ILs, but the deviation is still higher. As shown in Figure 2, most of the data points of viscosity are close to the diagonal line, and it can be observed from Figure 3 that relative deviation of most substances are within 20%, but quite a number of data points are greater than 20%, and some even reach to 68%. This demonstrates that viscosity prediction of ILs is not a simple linear problem, and QSPR model established by the MLR linear algorithm has certain limitation. Thus, a nonlinear algorithm is required to establish model to predict viscosity of ILs more accurately.

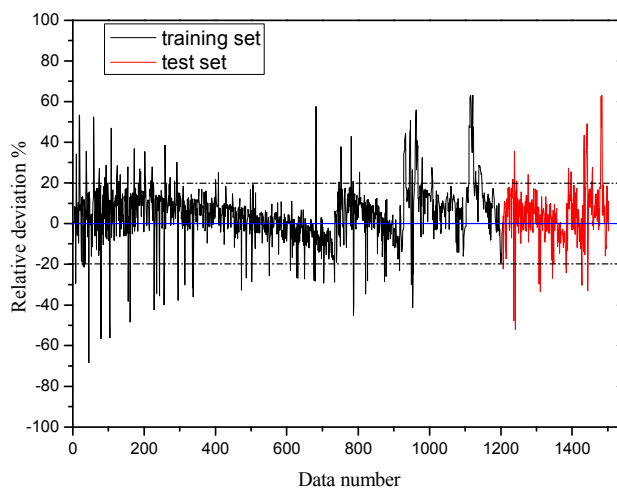


Fig. 3. Relative deviation of calculated using the MLR algorithm versus experimental viscosity values

3.2. Results of SVM model and comparison

The same descriptors applied by the MLR algorithm were utilized as the input parameters to

establish an accurate nonlinear model by the SVM. The performance of the SVM depends on the combination of C , ε , and the corresponding parameter r of the kernel function. Gaussian radial basis function (RBF) is commonly applied in many works due to its good generalization ability. In this study, the RBF was applied too. The optimization of SVM parameters was carried out by systemically varying the value and calculating the MSE of the training set. The lowest MSE is selected as the best condition. When $C=80$, $\varepsilon=0.0213$, $r=5.25$ with 176 support vectors, the best model was obtained. The predicted viscosities, experimental values, and absolute relative deviations are shown in Supporting Information. The plot of predicted in comparison with experimental viscosity logarithm values for each data set is recorded in Fig. 4. The statistical parameter R^2 for the training, test and whole sets are 0.948, 0.930 and 0.944 respectively. Relative deviations of the predicted viscosity values vs. experimental data are depicted in Fig. 5. After taking into account the experimental uncertainties of viscosity data points published in the paper, such errors are acceptable.

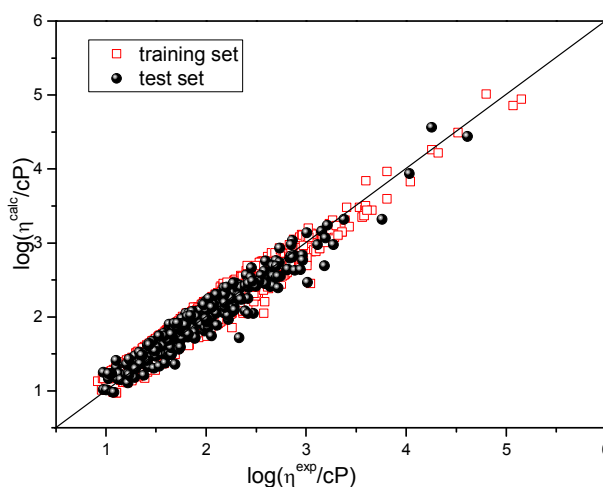


Fig. 4. Calculated versus experimental viscosities using the SVM algorithm for all ionic liquids

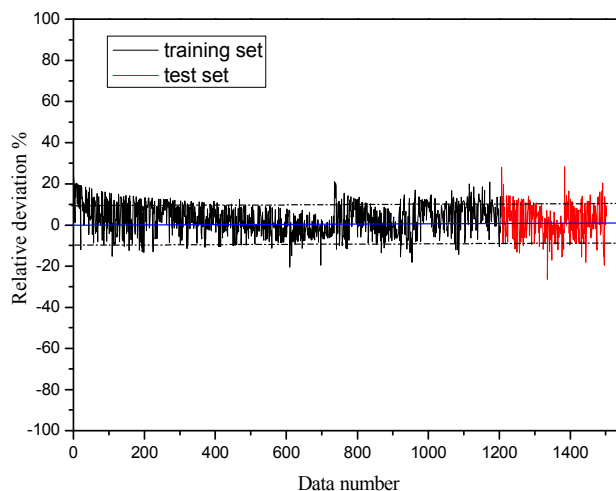


Fig. 5. Relative deviation of calculated using the SVM algorithm versus experimental viscosity values

The relative deviation distribution for all the data set is shown in Fig. 6. It is seen that most of the associated prediction error (77.3%) of the entire data set lies in the range of 0-10%. The maximum calculated absolute deviation from experimental values is 28.2% for [C₂mim][C₄SO₄], which may occur randomly, probably by the source of error for the impurity effect on the viscosity of ILs. The presence of chloride contamination, for example, increasing Cl⁻ concentration from 0.01 to 0.5 m in [C₄mim][BF₄] will increase the viscosity from 154 to 201 cP⁵².

A summary of the performance of the MLR and SVM models for all the data set is given in Table 3. In general, according to R^2 , MSE and AARD, the nonlinear model by SVM performs better (higher R^2 and lower error) than the linear model by the MLR, which indicates the proposed SVM model has better predictive power and general performance. This means that the relationship between viscosity and the $S_{\sigma\text{-profile}}$ descriptors are not just linear, and nonlinear modeling is more rational. Table 4 compares different models for correlation and prediction of the viscosity of ILs. It can be seen that the QSPR and GC are the most frequently employed. Although the models in the Ref. [6] and [53] present lower overall deviations than the models in this work, their models require more parameters, which will increase the complexity of the models. In addition, most of the reported models can only be used at atmospheric pressure. Therefore, the model presented here is suitable to wider application range.

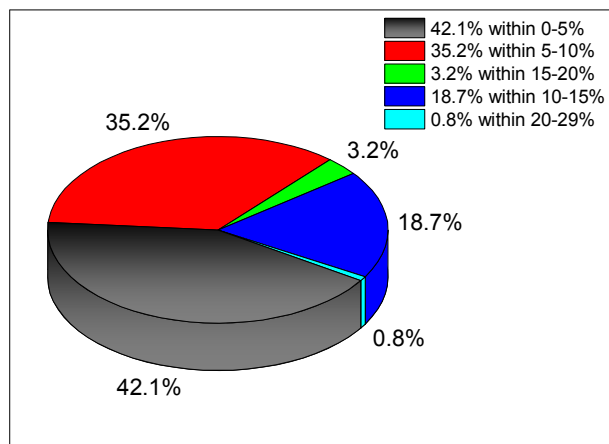


Fig. 6. Percent of value in different deviation range of the SVM model
42.1% of the viscosities are estimated within 0-5% (relative deviation range %); 35.2% within 5-10%, 18.7% within 10-15%, 3.2% within 15-20%, 0.6% within 20-25%, and 0.2% within 25-29%.

Table 3. Comparison of the statistical parameters by the MLR and SVM algorithms

Algorithms	Data set	No.	R^2	MSE	%AARD
MLR	Training	1205	0.803	0.171	10.70
	Test	297	0.800	0.187	10.61
	Total	1502	0.803	0.171	10.68
SVM	Training	1205	0.948	0.021	6.58
	Test	297	0.930	0.025	6.75
	Total	1502	0.944	0.022	6.58

Table 4. The comparison of different models for viscosities of ILs

Method	N_p	N_{IL}	Trange/K	Pressure/MPa	% AARD	Reference
GC	13	29	293–393	0.1	7.7	16
GC	12	25	293–393	0.1	7.7	13
GC	18	146 ^{*a}	293–363	0.1	4.17	53
Correlation		72	253-373	0.1		54
QSPR+GC	27	300 ^{*a}	263–353	0.1		17
QSPR	24	125	298	0.1	a little poor	55
QSPR	7	293	253-373	0.1	8.77	35
GC	242	1484	253-573	0.06-350	11.4	5
QSPR+GC	17	26	258.15-433.15	0.101	2.45	6
QSPR(SVM)	7	89	253.15-395.32	0.1-300	6.58	This work

N_p is the number of parameters, N_{IL} is the number of ILs, and ^{*a} is the data points

4. Conclusions

In this work, two novel QSPR models are established to predict the viscosity based on the $S_{\sigma\text{-profile}}$ descriptor and 1502 experimental data points of 89 ILs. The results show that both the

linear (MLR) and nonlinear (SVM) models can provide acceptable results. The SVM model predicts the viscosity of ILs more effectively than the MLR model over a wide range of temperature (253.15-395.32 K), and pressure (1-3000 bar). These results provide important information for the synthesis and screening of ILs of suitable viscosity. Moreover, since the viscosity is the function of $S_{\sigma\text{-profile}}$ of ILs, the presented derived models can provide some insight into what structural features are connected with the viscosity of ILs.

Acknowledgement

We would like to acknowledge the support from the National Basic Research Program of China (973 Program) (No.2015CB251403, No.2013CB733506), the Beijing Natural Science Foundation (2141003) and the National Natural Science Foundation of China (No.21036007, No.21106146). The LIBSVM program is supported by Chih-Jen Lin team (National Taiwan University, Taiwan).

Reference

- 1 R. D. Rogers and K. R. Seddon, *Science*, 2003, **302**, 792-793.
- 2 S. Chen, S. Zhang, X. Liu, J. Wang, J. Wang, K. Dong, J. Sun and B. Xu, *Phys. Chem. Chem. Phys.*, 2014, **16**, 5893-5906.
- 3 X. Zhang, X. Zhang, H. Dong, Z. Zhao, S. Zhang and Y. Huang, *Energy & Environmental Science*, 2012, **5**, 6668-6681.
- 4 J. M. P. B.E. Poling, J.P. O'Connell, *The Properties of Gases and Liquids*, 5th ed., McGraw-Hill, New York,, 2001.
- 5 K. Padaszynski and U. Domanska, *Journal of Chemical Information and Modeling*, 2014, **54**, 1311-1324.
- 6 B.-K. Chen, M.-J. Liang, T.-Y. Wu and H. P. Wang, *Fluid. Phase. Equilibr.*, 2013, **350**, 37-42.
- 7 F. Llovell, R. M. Marcos and L. F. Vega, *J. Phys. Chem. B*, 2013, **117**, 8159-8171.
- 8 A. C. F. Mendonca, N. Doerr and A. A. H. Padua, *Proceedings of the Institution of Mechanical Engineers Part J-Journal of Engineering Tribology*, 2012, **226**, 965-976.
- 9 I. Polishuk, *Ind. Eng. Chem. Res.*, 2012, **51**, 13527-13537.
- 10 Y. Miao, D. W. Rooney and Q. Gan, *International Journal of Computational Intelligence Systems*, 2012, **5**, 460-471.
- 11 S. N. Butler and F. Mueller-Plathe, *Chemphyschem*, 2012, **13**, 1791-1801.
- 12 J. Omar Valderrama, J. Makarena Munoz and R. Erasmo Rojas, *Korean J. Chem. Eng.*, 2011, **28**, 1451-1457.
- 13 R. L. Gardas and J. A. P. Coutinho, *AIChE J.*, 2009, **55**, 1274-1290.
- 14 O. Borodin, G. D. Smith and H. Kim, *J. Phys. Chem. B*, 2009, **113**, 4771-4774.
- 15 M. G. Bogdanov and W. Kantlehner, *Z. Naturforsch. B*, 2009, **64**, 215-222.
- 16 R. L. Gardas and J. A. P. Coutinho, *Fluid. Phase. Equilibr.*, 2008, **266**, 195-201.

- 17 H. Matsuda, H. Yamamoto, K. Kurihara and K. Tochigi, *Fluid. Phase. Equilibr.*, 2007, **261**, 434-443.
- 18 J. A. P. Coutinho, P. J. Carvalho and N. M. C. Oliveira, *Rsc Advances*, 2012, **2**, 7322-7346.
- 19 Y. Zhao, X. Zhang, J. Zhao, H. Zhang, X. Kang and F. Dong, *Progress in Chemistry*, 2012, **24**, 1236-1244.
- 20 G. Yu, L. Wen, D. Zhao, C. Asumana and X. Chen, *J. Mol. Liq.*, 2013, **184**, 51-59.
- 21 J. S. Torrecilla, F. Rodriguez, J. L. Bravo, G. Rothenberg, K. R. Seddon and I. Lopez-Martin, *Phys. Chem. Chem. Phys.*, 2008, **10**, 5826-5831.
- 22 A. R. Katritzky, A. Lomaka, R. Petrukhin, R. Jain, M. Karelson, A. E. Visser and R. D. Rogers, *J. Chem. Inf. Comp. Sci.*, 2002, **42**, 71-74.
- 23 S. Trohalaki, R. Pachter, G. W. Drake and T. Hawkins, *Energ. Fuel*, 2005, **19**, 279-284.
- 24 R. Bini, C. Chiappe, C. Duce, A. Micheli, R. Solaro, A. Starita and M. R. Tiné, *Green Chem.*, 2008, **10**, 306-309.
- 25 D. M. Eike, J. F. Brennecke and E. J. Maginn, *Green Chem.*, 2003, **5**, 323-328.
- 26 R. L. Gardas and J. A. P. Coutinho, *Fluid. Phase. Equilibr.*, 2008, **267**, 188-192.
- 27 Y. A. Sanmamed, D. Gonzalez-Salgado, J. Troncoso, C. A. Cerdeirina and L. Romani, *Fluid. Phase. Equilibr.*, 2007, **252**, 96-102.
- 28 P. Diaz-Rodriguez, J. C. Cancilla, N. V. Plechkova, G. Matute, K. R. Seddon and J. S. Torrecilla, *Phys. Chem. Chem. Phys.*, 2014, **16**, 128-134.
- 29 A. A. Oliferenko, P. V. Oliferenko, K. R. Seddon and J. S. Torrecilla, *Phys. Chem. Chem. Phys.*, 2011, **13**, 17262-17272.
- 30 R. L. Gardas and J. A. Coutinho, *Fluid. Phase. Equilibr.*, 2008, **265**, 57-65.
- 31 Y. Zhao, J. Zhao, Y. Huang, Q. Zhou, X. Zhang and S. Zhang, *J. Hazard. Mater.*, 2014, **278C**, 320-329.
- 32 J. S. Torrecilla, J. Palomar, J. Lemus and F. Rodríguez, *Green Chem.*, 2010, **12**, 123-134.
- 33 H. Matsuda, H. Yamamoto, K. Kurihara and K. Tochigi, *Fluid. Phase. Equilibr.*, 2007, **261**, 434-443.
- 34 C. Han, G. Yu, L. Wen, D. Zhao, C. Asumana and X. Chen, *Fluid. Phase. Equilibr.*, 2011, **300**, 95-104.
- 35 S. A. Mirkhani and F. Gharagheizi, *Ind. Eng. Chem. Res.*, 2012, **51**, 2470-2477.
- 36 F. Eckert and A. Klamt, *AIChE J.*, 2002, **48**, 369-385.
- 37 A. Klamt and F. Eckert, *Fluid. Phase. Equilibr.*, 2000, **172**, 43-72.
- 38 F. Eckert and A. Klamt, *Available at ww w. cosmologic. de*, 2013.
- 39 A. Klamt, V. Jonas, T. Bürger and J. C. Lohrenz, *The Journal of Physical Chemistry A*, 1998, **102**, 5074-5085.
- 40 J. S. Torrecilla, J. Palomar, J. García and F. Rodríguez, *Journal of Chemical & Engineering Data*, 2009, **54**, 1297-1301.
- 41 N. V. K. Dutt, Y. V. L. Ravikumar and K. Y. Rani, *Chem. Eng. Commun.*, 2013, **200**, 1600-1622.
- 42 J. Palomar, V. R. Ferro, J. S. Torrecilla and F. Rodríguez, *Ind. Eng. Chem. Res.*, 2007, **46**, 6041-6048.
- 43 C.-C. Chang and C.-J. Lin, *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2011, **2**, 27.
- 44 in *NIST Standard Reference Database*, <http://ilthermo.boulder.nist.gov/ILThermo/mainmenu.uix>, 2010.

- 45 R. Draper and H. Smith, Wiley, New York.
- 46 C. Cortes and V. Vapnik, *Mach. Learn.*, 1995, **20**, 273-297.
- 47 Y. Ren, H. Liu, X. Yao and M. Liu, *Anal. Chim. Acta*, 2007, **589**, 150-158.
- 48 N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press, 2000.
- 49 B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*, MIT press, 2002.
- 50 Y. Ren, H. Liu, C. Xue, X. Yao, M. Liu and B. Fan, *Anal. Chim. Acta*, 2006, **572**, 272-282.
- 51 P. Wasserscheid, R. van Hal and A. Bosmann, *Green Chem.*, 2002, **4**, 400-404.
- 52 K. R. Seddon, A. Stark and M. J. Torres, *Pure Appl. Chem.*, 2000, **72**, 2275-2287.
- 53 H. Matsuda, H. Yamamoto, K. Kurihara and K. Tochigi, *J. Comput. Aided Chem*, 2007, **8**, 114-127.
- 54 P. Eiden, S. Bulut, T. Köchner, C. Friedrich, T. Schubert and I. Krossing, *The Journal of Physical Chemistry B*, 2010, **115**, 300-309.
- 55 I. Billard, G. Marcou, A. Ouadi and A. Varnek, *J. Phys. Chem. B*, 2011, **115**, 93-98.