



Representation/Prediction of Physico-Chemical Properties of Ionic Liquids through Different Computational Methods

Mehdi Sattari

MSc. Chemical Engineering

Tarbiat Modares University

Supervisors: Prof. Deresh Ramugernath

Co-supervisor: Prof. Amir H. Mohammadi

December 2014

This thesis is dedicated to my beloved wife, Samira, whose love and unlimited patience gave me unceasing strength, motivation, and encouragement during the challenging moments of my life. She always stood behind me all the time, no matter what happened. Without her support, I would never be able to accomplish this work.

This thesis is also dedicated to my dearest father and mother, Firooz and Mahvash, who brought me up with their love. They supported me and have always believed in me throughout my life. I owe them for all achievements that I have got to date.

I love you all.

تا بدانی که کیست عقل و خدا
دل به دانش فرشته باید کرد
راه جوید به آفریننده
دانش ذات خویش میاید

اوحدی

علم خود را مکن ز عقل جدا
تن به دانش سرشته باید کرد
دل شود که به علم بیننده
آنچه در علم بیش میاید

ABSTRACT

The “green” industrial chemical processes are of great interest to scientists and engineers due to elimination of environmental pollution, especially air pollution. One of the most important air pollutants is class of materials called volatile organic compounds (VOCs) which are widely used in different industrial chemical processes. The recent research has revealed that ionic liquids (ILs) are generally the best possible alternative to the conventional solvents; because in general, the ILs have interesting properties such as very low vapor pressure, nonflammability, and high physical and chemical stability.

Ionic liquids are constituted of ions, typically a cation and an anion, and their thermophysical properties are strongly dependent on the type and chemical structure of the cation and anion. As a result, in theory, they can be designed for specific applications with certain properties by choosing the appropriate combination of anion/cation pair. For this purpose, a predictive model is required to estimate the target property based on the chemical structure of ions.

At the initial step of this study, the NIST Standard Reference Database #103b as well as the published papers in the literature was chosen as the source of experimental data of ionic liquids. As a result, a large database was collected covering several thermophysical properties of ILs. Thereafter, the collected data were examined carefully and the duplicated and erroneous data were screened.

Speed of sound, heat capacity, refractive index, viscosity, infinite dilution activity coefficient (γ^∞), and critical temperature of various ionic liquids were modeled by means of two well-known property estimation methods, Group Contribution (GC) and Quantitative Structure-Property Relationship (QSPR) methods. These methods were combined with different computational and regression techniques such as genetic function approximation (GFA) and least square support vector machine (LS-SVM). The combined routines then were applied to select reasonable number of parameters from thousands of variables and to develop the predictive models for representation/prediction of chosen temperature-dependent thermophysical properties of ionic liquids.

Speed of sound in ionic liquids was modeled successfully and two models were developed, one GC and one QSPR model. These models were the first GC and QSPR models developed

for this property in the literature. Both models had better accuracy in terms of average absolute relative deviation (the AARD% of 0.36 for the GC and 0.92% for the QSPR models over 41 ILs) and covered a wider range of ionic liquids compared with the previous models published (AARD% of 1.96% over 14 ILs) and consequently, they were more applicable.

Liquid heat capacity of ionic liquids was studied and one GC and one QSPR model were developed. Both models covered 82 ILs which was a larger number of ionic liquids compared with the best available model in the literature (32 ILs with an AARD% of 0.34%) and had relatively low AARD%. The AARD% of the models was 1.68% and 1.70% for the GC and QSPR models, respectively. In addition, the QSPR model was the first model developed for this property through the QSPR approach.

For the refractive index of ionic liquids, little attention had been given to modeling and consequently, one new GC (AARD% = 0.34%) and the first QSPR (AARD% = 0.51%) models were developed to predict this property using the experimental data for 97 ionic liquids. Both models covered a wider range of ionic liquids and showed very good prediction ability compared with the best available model (an AARD% of 0.18% for 24 ILs).

Viscosity of Fluorine-containing ionic liquids was studied because the insertion of fluorinated moieties in the molecular structure of ionic liquids could result in reduction of viscosity. As a result, one QSPR (AARD% = 2.91%) and two GC models were developed using two different databases, one with fewer number of ionic liquids but with more reliable data (AARD% = 3.23%), the one with larger number of ionic liquids but with lower reliability (AARD% = 4.85%). All of the models developed had better prediction ability compared with the previous models and covered a wider range of fluorinated ionic liquids.

Infinite dilution activity coefficient (γ^∞) of organic solutes was modeled by developing six different models for different types of solutes (alkane, alkene, aromatic, etc.). The model developed were the first GC models for the prediction of γ^∞ of solutes in ionic liquids. They were much easier to use, more comprehensive, and much more accurate compared with the UNIFAC model.

Ultimately, the theoretical critical temperature (T_c) of ionic liquids was tried to model using the GC and QSPR approaches. The experimental data of surface tension of 106 ionic liquids

were used to calculate the critical temperature and then, these values were used to develop the models. It was found that the only available model in the literature was not accurate and predictive enough when its output was compared with the abovementioned T_c values. In addition, it was found that both of the models developed were not predictive enough to calculate the T_c of various types of ionic liquids as the models were developed using a few number of ionic liquids; however both models were accurate enough to fit the used values of T_c . The GC model has an AARD% of 5.17% and the QSPR model showed the AARD% of 4.69%.

In this thesis, much larger databases were used to develop the models compared with the models published previously in the literature. It was found that thermophysical properties of ionic liquids can be modeled fairly well by combination of the GC or QSPR methods with an appropriate regression technique. In addition, the developed models improved significantly the quality of fit and predictions for a wider range of ionic liquids compared with the previous models. Consequently, the models proposed are more predictive and can be used to design the ionic liquids with desired property for specific applications.

DECLARATION

I, Mehdi Sattari, declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original work.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
 - a) their words have been re-written but the general information attributed to them has been referenced;
 - b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.
5. Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.
6. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Mehdi Sattari (Candidate)

Date

As the candidate's Supervisor I agree/do not agree to the submission of this thesis.

Prof. Deresh Ramjugernath (Supervisor)

Date

Prof. Amir H. Mohammadi (Co-Supervisor)

Date

ACKNOWLEDGEMENT

In the Name of GOD, the Most Merciful, the Most Compassionate, all praise is due to Allah, the Lord of the worlds.

First and the foremost, I would like to thank GOD, the Almighty, who created me and gave me an opportunity to live. Without His guidance and bless, this study would have never become truth.

I would like to express my sincere and boundless appreciation to both my supervisors, Prof. Deresh Ramjugernath, and Prof. Amir H. Mohammadi, for their precious advices during conducting the present study. I would like to thank them for their unfailing patience during my PhD program and supporting me through all aspects of my study. I was very fortunate to have been able to work with them in a very peaceful and collaborative environment, the Thermodynamics Research Unit in School of Engineering, University of KwaZulu-Natal.

I would like to extend my appreciation to all staff members of University of KwaZulu-Natal, specifically the Chemical Engineering Department, for their warm and affable comportment with me as an international student. I wish to thank all the students of Thermodynamics Research Unit who were very friendly to me and treated me like a member of their family.

I would like to thank Brian J. Satola for automatic calculation of infinite dilution activity coefficient for several solutes in ionic liquids using the UNIFAC model. Without his collaboration, I had to spend several hours to perform the calculations of activity coefficient.

Special thanks should go to Dr. Michael Frenkel, the Director of the Thermodynamics Research Center (TRC) at the U. S. National Institute of Standards and Technology (NIST) for providing the NIST Standard Reference Database and ThermoData Engine for this research study.

I wish to acknowledge the South African Research Chairs Initiative (SARChI) of the Department of Science and Technology and National Research Foundation (DST-NRF) for funding and financial support of present study.

Last but not the least, I owe profound gratitude to my dearest wife, Samira, who is the reason of all my power and strength. Words cannot express how much I owe to her during my PhD

study. She provided me a warm and lovely home in spite of her endless problems during the course of her Masters' program. I also would like to express my wholehearted appreciation to my parents who supported me emotionally and financially, every second of my life.

TABLE OF CONTENT

ABSTRACT	i
DECLARATION	v
ACKNOWLEDGEMENT	vii
TABLE OF CONTENT	ix
LIST OF FIGURES	xiv
LIST OF TABLES	xx
LIST OF PUBLICATIONS.....	xxiv
NOMENCLATURE.....	xxvii
ABBREVIATIONS.....	xxviii
CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: LITERATURE REVIEW	5
2.1 Scope	5
2.2 Speed of sound in ionic liquids	6
2.3 Liquid heat capacity of ionic liquids	7
2.4 Refractive Index of ionic liquids	8
2.5 Viscosity of fluorine-containing ionic liquids (F-ILs)	10
2.6 Infinite dilution activity coefficient of organic solutes in ionic liquids	12
2.7 Critical temperature of ionic liquids.....	15
CHAPTER 3: DATA MANAGEMENT.....	18
3.1 Scope	18
3.2 Data source for the ionic liquids	18
3.3 Data screening.....	18
3.4 Datasets of ionic liquids for different properties.....	19
3.4.1 Speed of sound in ionic liquids	19
3.4.2 Liquid heat capacity of ionic liquids	20

3.4.3 Refractive index of ionic liquids	21
3.4.4 Viscosity of F-ILs	22
3.4.5 Infinite dilution activity coefficient of organic solutes in ionic liquids.....	23
3.4.6 Critical temperature of ionic liquids	29
CHAPTER 4: COMPUTATIONAL METHODS	31
4.1 Scope	31
4.2 Property Estimation Methods.....	32
4.2.1 Group Contribution Methods	32
4.2.2 QSPR Methods	33
4.3 Calculation of descriptors/functional groups	34
4.4 Variable reduction	36
4.5 Dataset partitioning	36
4.6 Feature Selection Methods.....	37
4.6.1 Stepwise Method	38
4.6.2 Genetic Algorithm	38
4.7 Mathematical Methods	39
4.7.1 Genetic Algorithm based Multiple Linear Regression (GA-MLR)	39
4.7.2 Genetic Function Approximation (GFA).....	40
4.7.3 Support Vector Machine (SVM)	40
4.7.4 Forward feature selection-based LSSVM	43
4.7.5 Genetic Programming (GP).....	43
4.7.6 Gene Expression Programming (GEP)	44
4.8 Fitness function	46
CHAPTER 5: MODEL DEVELOPMENT	48
5.1 Scope	48
5.2 Speed of sound in ionic liquids	48
5.2.1 The GC model	48

5.2.2 The QSPR model.....	49
5.3 Liquid heat capacity of ionic liquids	50
5.3.1 The GC model.....	50
5.3.2 The QSPR model.....	51
5.4 Refractive index of ionic liquids	52
5.4.1 The GC model	52
5.4.2 The QSPR model.....	53
5.5 Viscosity of F-ILs.....	53
5.5.1 The GC model	53
5.5.2 The QSPR model.....	54
5.6 γ^∞ of organic solutes in ionic liquids	54
5.6.1 Aromatic solutes.....	54
5.6.2 Alcohol solutes	55
5.6.3 Alkane solutes	55
5.6.4 Alkene solutes	56
5.6.5 Alkyne solutes	56
5.7 Critical temperature of ionic liquids.....	56
CHAPTER 6: RESULTS	57
6.1 Scope	57
6.2 Speed of sound in ionic liquids	57
6.2.1 The GC model	57
6.2.2 The QSPR model.....	65
6.3 Liquid heat capacity of ionic liquids	70
6.3.1 The GC model	70
6.3.2 The QSPR model.....	83
6.4 Refractive index of ionic liquids	93
6.4.1 The GC model	93

6.4.2 The QSPR model.....	102
6.5 Viscosity of F-ILs.....	111
6.5.1 The GC model.....	111
6.5.2 The QSPR model.....	122
6.6 γ^∞ of solutes in ionic liquids.....	127
6.6.1 Aromatic solutes.....	127
6.6.2 Alcohol solutes.....	137
6.6.3 Alkane solutes.....	146
6.6.4 Alkene solutes.....	161
6.6.5 Alkyne solutes.....	169
6.7 Critical temperature of ionic liquids.....	177
CHAPTER 7: CONCLUSION.....	185
CHAPTER 8: RECOMMENDATIONS FOR FURTHER STUDIES.....	188
REFERENCES.....	190

LIST OF FIGURES

Figure 3.1: The number of ionic liquids in different families used for modeling the speed of sound.	20
Figure 3.2: The number of ionic liquids in different families used for modeling the liquid heat capacity.....	21
Figure 3.3. The number of ionic liquids in different families used for modeling the refractive index.	21
Figure 3.4: The number of ionic liquids in different families used to model the viscosity (refined dataset).	23
Figure 3.5: The number of ionic liquids in different families used to model the viscosity (unreliable dataset).	23
Figure 3.6: The γ^∞ data of “benzene” in “1-butyl-3-methylimidazolium nitrate”.....	24
Figure 3.7: The γ^∞ data of “ethylbenzene” in “1-butyl-3-methylimidazolium thiocyanate”. .	25
Figure 3.8: Number of data points of γ^∞ of aromatic solutes in each class of ionic liquids....	25
Figure 3.9: The γ^∞ data of “1-butanol” in “1-butyl-3-methylimidazolium trifluoromethanesulfonate”.....	26
Figure 3.10: The γ^∞ data of “methanol” in “1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide”.....	26
Figure 3.11: Number of data points of γ^∞ of alcohol solutes in each class of ionic liquids.	27
Figure 3.12: Number of data points of γ^∞ of alkane solutes in each class of ionic liquids.	27
Figure 3.13: Number of data points of γ^∞ of alkene solutes in each class of ionic liquids.	28
Figure 3.14: Number of data points of γ^∞ of alkyne solutes in each class of ionic liquids.	29
Figure 3.15: The number of ionic liquids in different families used for modeling the critical temperature.	30
Figure 4.1: The flow diagram of developing a property estimation model.....	31
Figure 4.2: Functional groups in “acetone” molecule.....	32
Figure 4.3: The schematic microstructure of “acetone” molecule.....	34
Figure 4.4: A typical expression tree in the gene expression programming, which represents $x - x \times x + x$ by a two-gene chromosome.....	45
Figure 6.1: Predicted versus experimental values of speed of sound in ILs.	58
Figure 6.2: Relative deviation of predicted speed of sound from experimental data.	58
Figure 6.3: Percentage of predicted values of speed of sound in different relative deviation ranges.	59

Figure 6.4: Predicted versus experimental values of speed of sound in ILs.	66
Figure 6.5: Relative deviation of predicted speed of sound from experimental data.	66
Figure 6.6: Percentage of the predicted values of speed of sound in different relative deviation ranges	67
Figure 6.7: Effect of the number of functional groups on the accuracy of the classical GC and GFA models of C_{pL}	71
Figure 6.8: Changes in the accuracy of C_{pL} model versus the increasing number of functional groups.	71
Figure 6.9: The effect of the number of C_{pL} model parameters on the accuracy of the training and test sets.	72
Figure 6.10: Predicted versus experimental values of C_{pL} (GC model).	75
Figure 6.11: Relative deviation of predicted C_{pL} from experimental data (GC model).	76
Figure 6.12: Comparison of C_{pL} data for 1-hexyl-3-methylimidazolium tetrafluoroborate measured by Crosthwaite <i>et al.</i> [236] and Garcia-Miaja <i>et al.</i> [237]	77
Figure 6.13: Comparison of C_{pL} data for 1-octyl-3-methylimidazolium tetrafluoroborate measured by Crosthwaite <i>et al.</i> and other researchers.	77
Figure 6.14: Effect of the number of descriptors on the accuracy of the linear QSPR and GFA models of C_{pL}	83
Figure 6.15: Changes in the accuracy of C_{pL} model versus the increasing number of descriptors	84
Figure 6.16: Percentage of the predicted values of C_{pL} in different relative deviation ranges.	86
Figure 6.17: Predicted versus experimental values of C_{pL} (QSPR model).	87
Figure 6.18: Relative deviation of predicted C_{pL} from experimental data (QSPR model).	88
Figure 6.19: Predicted versus experimental values of n_D (— diagonal line).	95
Figure 6.20: Relative deviation of predicted n_D from experimental data.	95
Figure 6.21: Percentage of predicted values of n_D in different relative deviation ranges.	96
Figure 6.22: Predicted versus experimental values of n_D (— diagonal line).	104
Figure 6.23: Relative deviation of predicted n_D from experimental data.	104
Figure 6.24: Percentage of calculated/predicted values of n_D in different relative deviation ranges	105
Figure 6.25: Predicted versus experimental values of $\ln(\eta)$ (refined database) (— diagonal line).	118
Figure 6.26: Relative deviation of predicted $\ln(\eta)$ from experimental data (refined database).	119

Figure 6.27: Predicted versus experimental values of η in linear scale (refined database) (— diagonal line).	120
Figure 6.28: Predicted versus experimental values of $\ln(\eta)$ (entire database) (— diagonal line).	121
Figure 6.29: Predicted versus experimental values of η in linear scale (entire database) (— diagonal line).	121
Figure 6.30: Predicted versus experimental values of $\ln(\eta)$ (— diagonal line).	124
Figure 6.31: Relative deviation of predicted $\ln(\eta)$ from experimental data.	125
Figure 6.32: Predicted versus experimental values of η in linear scale by the QSPR model (— diagonal line).	126
Figure 6.33: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of aromatic solutes (— diagonal line).	130
Figure 6.34: Correlated/Predicted versus experimental values of γ^∞ of aromatic solutes (— diagonal line).	131
Figure 6.35: Percentage of calculated/predicted values of γ^∞ of aromatic solutes in different relative deviation ranges.	132
Figure 6.36: Cation-Cation hydrogen bond for 1-(2-hydroxyethyl)-3-methylimidazolium	133
Figure 6.37: Cation-anion hydrogen bonds for 1-(2-hydroxyethyl)-3-methylimidazolium bis(trifluoromethylsulfonyl)imide.	133
Figure 6.38: Cation-solute hydrogen bonds for 1-(2-hydroxyethyl)-3-methylimidazolium and benzene.	133
Figure 6.39: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of alcohol solutes (— diagonal line).	140
Figure 6.40: Correlated/Predicted versus experimental values of γ^∞ of alcohol solutes (— diagonal line).	140
Figure 6.41: Percentage of calculated/predicted values of γ^∞ of alcohol solutes in different relative deviation ranges.	141
Figure 6.42: Cation-Anion hydrogen bond for 1-ethyl-3-methylimidazolium tetrafluoroborate.	142
Figure 6.43: Cation-Solute hydrogen bond for 1-ethyl-3-methylimidazolium and methanol	143
Figure 6.44: Anion-Solute hydrogen bond for tetrafluoroborate and methanol.	143

Figure 6.45: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of alkane solutes ($nC_{sol}<10$) (— diagonal line).	149
Figure 6.46: Correlated/Predicted versus experimental values of γ^∞ of alkane solutes ($nC_{sol}<10$) (— diagonal line).	149
Figure 6.47: Percentage of calculated/predicted values of γ^∞ of alkane solutes ($nC_{sol}<10$) in different relative deviation ranges.	150
Figure 6.48: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of alkane solutes ($nC_{sol}\geq 10$) (— diagonal line).	155
Figure 6.49: Correlated/Predicted versus experimental values of γ^∞ of alkane solutes ($nC_{sol}\geq 10$) (— diagonal line).	155
Figure 6.50: Percentage of calculated/predicted values of γ^∞ of alkane solutes in different relative deviation ranges ($nC_{sol}\geq 10$).	156
Figure 6.51: The γ^∞ of different alkane solutes in 1-hexyl-3-methylimidazolium trifluoromethanesulfonate.	157
Figure 6.52: The γ^∞ of different alkane solutes in 1-ethyl-3-methylimidazolium diethylphosphate.	157
Figure 6.53: Correlated/Predicted versus experimental values of γ^∞ of different alkane solutes in 1-ethyl-3-methylimidazolium diethylphosphate (— diagonal line).	158
Figure 6.54: Correlated/Predicted versus experimental values of γ^∞ of different alkane solutes in 1-ethyl-3-methylimidazolium methanesulfonate (— diagonal line).	159
Figure 6.55: Correlated/Predicted versus experimental values of γ^∞ of tetradecane in different ionic liquids (— diagonal line).	159
Figure 6.56: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of alkene solutes (— diagonal line).	165
Figure 6.57: Correlated/Predicted versus experimental values of γ^∞ of alkene solutes (— diagonal line).	165
Figure 6.58: Percentage of calculated/ predicted values of γ^∞ of alkene solutes in different relative deviation ranges.	166
Figure 6.59: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of alkyne solutes (— diagonal line).	172
Figure 6.60: Correlated/Predicted versus experimental values of γ^∞ of alkyne solutes (— diagonal line).	172

Figure 6.61: Percentage of calculated/predicted values of γ^∞ of alkene solutes in different relative deviation ranges.....	173
Figure 6.62: The γ^∞ of different alkene solutes in 1-butyl-1-methylpyrrolidinium thiocyanate.	173
Figure 6.63: Correlated/Predicted versus experimental values of γ^∞ of 1-heptyne and 1-octyne in 1-ethyl-3-methylimidazolium thiocyanate (— diagonal line).	174
Figure 6.64: The γ^∞ of different alkene solutes in 1-ethyl-3-methylimidazolium thiocyanate.	175
Figure 6.65: The γ^∞ of different alkene solutes in 1-butyl-3-methylimidazolium thiocyanate.	176
Figure 6.66: The γ^∞ of different alkene solutes in 1-hexyl-3-methylimidazolium thiocyanate.	176
Figure 6.67: Estimated T_c of 1-ethyl-3-methylimidazolium ILs by different models.	181
Figure 6.68: Estimated T_c of 1-butyl-3-methylimidazolium ILs by different models.	182
Figure 6.69: T_c of 1-ethyl-3-methylimidazolium ionic liquids with 45 different anions calculated by the QSPR and Valderrama <i>et al.</i> models.	184

LIST OF TABLES

Table 2.1: Summary of available models for the speed of sound in ionic liquids.....	6
Table 2.2. Summary of available models for the of heat capacity of ILs.	8
Table 2.3: Summary of different models for refractive index of ILs.....	9
Table 2.4: Summary of different models for predicting the $\ln(\eta)$ of ILs.	12
Table 2.5: Summary of Valderrama <i>at al.</i> model for T_c of ionic liquids.	17
Table 6.1: The input variables of LS-SVM Model for speed of sounds in ILs.....	59
Table 6.2: Statistical parameter of the model of speed of sound.....	59
Table 6.3: IL abbreviations and AARD% of ionic liquids modelled by LSSVM.....	62
Table 6.4: Comparison between calculated values of speed of sound in ILs using different models	63
Table 6.5: List of ionic liquids and their frequency used to develop equation (6.1).....	67
Table 6.6: Statistical parameters for equation (6.1).....	68
Table 6.7: Summary of available models for the speed of sound in ionic liquids.....	70
Table 6.8: Description of parameters of the equation (6.2).....	73
Table 6.9: Statistical parameters for equation (6.2).....	75
Table 6.10: Name and AARD% of ionic liquids used to develop equation (6.2).	79
Table 6.11: The AARD% of equation (6.2) for different families of ionic liquids.	82
Table 6.12: Definition of descriptors used in equation (6.3).....	85
Table 6.13: Statistical parameters for equation (6.3).....	86
Table 6.14. Summary of available models for the of heat capacity of ILs.	89
Table 6.15: Name and AARD% of studied ionic liquids for developing equation (6.3).....	90
Table 6.16: Group contribution parameters $n_{a,i}$ and a_i in equation (6.4)	93
Table 6.17: Group contribution parameters $n_{b,i}$ and b_i in equation (6.4)	94
Table 6.18: Statistical parameters for equation (6.4).....	94
Table 6.19: Name and AARD% of studied ionic liquids for developing equation (6.4).....	98
Table 6.20: The AARD% of equation (6.4) for different families of ionic liquids.	101
Table 6.21: Definition of the descriptors used in equation (6.5).....	102
Table 6.22: Statistical parameters for equation (6.5).....	103
Table 6.23: Summary of different models for refractive index of ILs.....	106
Table 6.24: Name and AARD% of ionic liquids studied for developing equation (6.5).....	107
Table 6.25:Parameters of the GC model in equation (6.6).....	112
Table 6.26: Name and AARD% of ionic liquids studied.....	114

Table 6.27: The AARD% of equation (6.6) for different families of ionic liquids.	117
Table 6.28: The statistical error parameters for $\ln(\eta)$ in equation (6.6).	118
Table 6.29: Parameters of the QSPR model in equation (6.6).	123
Table 6.30: The statistical error parameters for the $\ln(\eta)$ in equation (6.6).	123
Table 6.31: Summary of different models for predicting the $\ln(\eta)$ of ILs.	126
Table 6.32: Parameters of the equation (6.8) for γ^∞ of aromatic solutes in ILs.	128
Table 6.33: Statistical parameters for the presented model for γ^∞ of aromatic solutes in ILs.	131
Table 6.34: AARD% of equation (6.8) for different class of ionic liquids.	134
Table 6.35: Summary of result of models for γ^∞ of aromatic solutes in ILs.	135
Table 6.36: γ^∞ of some aromatic solutes in ILs calculated using the GC models proposed (equation 6.8), and the original and modified UNIFAC models.	136
Table 6.37: Parameters of equation (6.10)	138
Table 6.38: Definition of parameter of equation (6.10).	139
Table 6.39: Statistical parameters for equation (6.10)	141
Table 6.40: AARD% of equation (6.10) for different class of ionic liquids.	143
Table 6.41: Summary of result of models for γ^∞ of alcohol solutes in ILs.	144
Table 6.42: γ^∞ of some alcohol solutes in ILs calculated using the GC models proposed (equation 6.9), and the original and modified UNIFAC models.	145
Table 6.43: Parameters of equation (6.11)	146
Table 6.44: Definition of parameter of equation (6.11).	148
Table 6.45: Statistical parameters for equation (6.11)	150
Table 6.46: AARD% of equation (6.10) for different class of ionic liquids.	152
Table 6.47: Parameters of equation (6.12)	153
Table 6.48: Definition of parameter of equation (6.12).	154
Table 6.49: Statistical parameters for equation (6.11)	156
Table 6.50: AARD% of equation (6.12) for different class of ionic liquids.	160
Table 6.51: Summary of result of models for γ^∞ of alkane solutes in ILs.	161
Table 6.52: Parameters of equation (6.13)	162
Table 6.53: Definition of parameter of equation (6.13).	163
Table 6.54: Statistical parameters for equation (6.13)	166
Table 6.55: AARD% of equation (6.13) for different class of ionic liquids.	168
Table 6.56: Summary of result of models for γ^∞ of alkene solutes in ILs.	168

Table 6.57: Parameters of equation (6.14)	169
Table 6.58: Definition of parameter of equation (6.14).	170
Table 6.59: Statistical parameters for equation (6.14).	173
Table 6.60: AARD% of equation (6.14) for different class of ionic liquids.	176
Table 6.61: Summary of result of models for γ^∞ of alkyene solutes in ILs.	177
Table 6.62: Comparison of estimated T_c of ionic liquids by different methods.	178

LIST OF PUBLICATIONS

1. **M. Sattari**, A. Kamari, H. Hashemi, A.H. Mohammadi, D. Ramjugernath, *A Group Contribution Model for Determining the Viscosity with Temperature Dependency for Fluorine-Containing Ionic Liquids*, submitted to: Journal of Fluorine Chemistry.
2. **M. Sattari**, A. Kamari, A.H. Mohammadi, D. Ramjugernath, *Prediction of Refractive Indices of Ionic Liquids - A Quantitative Structure-Property Relationship based Model*, Journal of the Taiwan Institute of Chemical Engineers; In Press (2015).
3. **M. Sattari**, A. Kamari, A.H. Mohammadi, D. Ramjugernath, *A group contribution method for estimating the refractive indices of ionic liquids*, J. Mol. Liq., 200 (2014) 410-415.
4. **M. Sattari**, F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, *Determination of the speed of sound in ionic liquids using a least squares support vector machine group contribution method*, Fluid Phase Equilib., 367 (2014) 188-193.
5. **M. Sattari**, F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, *A chemical structure based model for the determination of speed of sound in ionic liquids*, J. Mol. Liq., 196 (2014) 7-13.
6. **M. Sattari**, F. Gharagheizi, P. Ilani-Kashkouli, A. Mohammadi, D. Ramjugernath, *Development of a group contribution method for the estimation of heat capacities of ionic liquids*, J. Therm. Anal. Calorim., 115 (2014) 1863-1882.
7. **M. Sattari**, F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, *Estimation of the Heat Capacity of Ionic Liquids: A Quantitative Structure–Property Relationship Approach*, Industrial & Engineering Chemistry Research, 52 (2013) 13217–13221.
8. F. Gharagheizi, **M. Sattari**, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, D. Richon, *A “non-linear” quantitative structure–property relationship for the prediction of electrical conductivity of ionic liquids*, Chem. Eng. Sci., 101 (2013) 478-485.
9. F. Gharagheizi, **M. Sattari**, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, D. Richon, *Quantitative structure–property relationship for thermal decomposition temperature of ionic liquids*, Chem. Eng. Sci., 84 (2012) 557-563.
10. F. Gharagheizi, P. Ilani-Kashkouli, **M. Sattari**, A.H. Mohammadi, D. Ramjugernath, *A group contribution method for determination of thermal conductivity of liquid chemicals at atmospheric pressure*, J. Mol. Liq., 190 (2014) 223-230.

11. F. Gharagheizi, P. Ilani-Kashkouli, **M. Sattari**, A.H. Mohammadi, D. Ramjugernath, D. Richon, *Development of a LSSVM-GC model for estimating the electrical conductivity of ionic liquids*, Chem. Eng. Res. Des., 92 (2014) 66-79.
12. F. Gharagheizi, P. Ilani-Kashkouli, **M. Sattari**, A.H. Mohammadi, D. Ramjugernath, D. Richon, *Development of a general model for determination of thermal conductivity of liquid chemical compounds at atmospheric pressure*, AIChE J., 59 (2013) 1702-1708.
13. F. Gharagheizi, P. Ilani-Kashkouli, **M. Sattari**, A.H. Mohammadi, D. Ramjugernath, D. Richon, *Development of a quantitative structure–liquid thermal conductivity relationship for pure chemical compounds*, Fluid Phase Equilib., 355 (2013) 52-80.

Conference papers

14. **M. Sattari**, M.A. Mohammadi, D. Ramjugernath, *Prediction of Refractive Indices of Ionic Liquids - A Quantitative Structure-Property Relationship Approach*, 23rd International Conference of Chemical Thermodynamics and the South African Institution of Chemical Engineers Conference (ICCT/SAIChE), 27 June – 1 August, Durban, South Africa, 2014
15. **M. Sattari**, M.A. Mohammadi, D. Ramjugernath, *A Group Contribution Model for Prediction of Infinite Dilution Activity Coefficients of Ionic Liquids with Benzene and n-Heptane*, 23rd International Conference of Chemical Thermodynamics and the South African Institution of Chemical Engineers Conference (ICCT/SAIChE), 27 June – 1 August, Durban, South Africa, 2014
16. **M. Sattari**, M.A. Mohammadi, D. Ramjugernath, *A Group Contribution Model to Predict the Viscosity of Fluorine Containing Ionic Liquids*, 41st National Convention of the South African Chemical Institute (SACI), 1 - 6 December, East London, South Africa, 2013
17. **M. Sattari**, F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, *A group contribution method for estimating the refractive indices of ionic liquids*, 2nd International Conference on Ionic Liquids in Separation and Purification Technology (ILSEPT), June 29 – July 2, Toronto, Canada, 2014.
- 18.

NOMENCLATURE

Symbols

C_{pL}	Liquid heat capacity at constant pressure
n_D	Refractive index
u	speed of sound
T_c	Critical temperature
T_b	Normal boiling point
V_c	Critical volume
M	Molar mass

Greek letters

γ^∞	infinite dilution activity coefficient
η	Viscosity
σ	Surface tension
γ	Surface tension
γ	weight of the regression error (SVM)
ρ	Density
σ^2	RBF kernel parameter (SVM)

ABBREVIATIONS

ANN	Artificial neural network
F-IL	Fluorine-containing ionic liquid
FFS-LSSVM	Forward feature selection-based Least Squares Support Vector Machine
IL	Ionic liquid
GA	Genetic algorithm
GA-MLR	Genetic Algorithm based Multiple Linear Regression
GC	Group contribution
GEP	Gene expression programming
GFA	Genetic function approximation
LSSVM	Least Squares Support Vector Machine
MLR	Multiple Linear Regression
QSPR	Quantitative structure-property relationship
SVM	Support vector machine
UNIFAC	Uniquac Quasi-Chemical Functional Activity Coefficient
VOC	Volatile organic compound

CHAPTER 1: INTRODUCTION

In recent decades, the rapid growth of the world population has led to fast industrial development and higher usage of chemical materials. Among the chemical compounds, the chemical solvents, especially the volatile organic compounds (VOCs) are of great importance in numerous industrial processes and applications. The main benefit of using VOCs is their ease of removal and evaporation in several applications such as separation and extraction processes; but their critical disadvantages are their adverse health effects such as allergic skin reactions, dyspnea, nose and throat discomfort, and their environmental pollutant role destroying the ozone layer through free radical air oxidation processes [1].

Recently, there are lots of demands among several countries to move toward “green” industries to eliminate the environmental pollutions, especially air pollution. In this regard, one of the basic and fundamental steps is to substitute the VOCs from different industrial chemical processes with a “green” alternative. Recent research has revealed that the ionic liquids (ILs) are typically the best possible environmental friendly alternative to the conventional solvents. In general, they show unusual but interesting properties such as extremely low saturation vapor pressure and negligible volatility, wide liquid range, nonflammability, high thermal conductivity, and high physical and chemical stability [2].

Ionic liquids are defined as molten salts which are generally liquid at or near room temperature (typically below 100 °C) due to the poor coordination of ions [3]. ILs are typically composed of a cation and an anion; so their thermophysical properties are strongly dependent on the type and chemical structure of the ions. As a result, they can be designed for specific applications with desired properties by choosing the proper pair of ions. This feature has made them the “tunable” and “designable” materials [4]. Consequently, ionic liquids are potential to be applied in numerous industrial applications, such as extraction and separation processes [5, 6], battery industry [7, 8], fuel cells [9, 10], solar panels [11, 12], polymer and biopolymer processing [13-15], electroplating [16, 17], lubricants [18-21], waste recycling [22-25], gas separation and CO₂ capturing [26-33], catalysis [34-36], and many others [37-40].

The existence of a large number of combinations of organic cations and anions leads to the generation of various groups of ionic liquids. The most popularly researched groups are imidazolium, ammonium, phosphonium, pyrrolidinium, pyridinium, guanidinium, isoquinolinium, piperidinium, morpholinium, and sulphonium. As a result, thousands of ionic liquids can be synthesized in theory by different types of cations and anions. In addition, it is possible to design the ILs for specific applications before synthesis. In this regard, predictive models can play an important role to relate the physico-chemical properties of ILs to their constituent cation and anion combinations or other properties of ILs.

The models used for property estimations of ionic liquids as well as other chemical compounds are classified into three main types:

1. Models based on chemical structure and functions groups of the compounds.
2. Models based on other physico-chemical properties.
3. Models based on both chemical structure and other physico-chemical properties.

Usually the model which uses the other thermophysical properties of the compounds has better accuracy and prediction ability, especially for nonlinear properties; but it strongly depends on the availability of the experimental data for all parameters of the model for desired compound. If one of the parameters does not exist in the literature, the model becomes useless if the missing parameter cannot be estimated by another model.

The first ionic liquid was discovered in 1888 [41]. Since then, more than a thousand ionic liquids have been reported to date. Despite the great interest of introducing the new ionic liquids, less effort have been done to measure the physico-chemical properties of ionic liquids uniformly. For example, experimental density data is available for more than 500 ionic liquids, but surface tension has been measured for less than 150 ILs. At present, the results of experimental investigations for ionic liquids as well as other compounds are reported comprehensively in the form of large databases by Dortmund Data Bank [42], and NIST ThermoData Engine [43]. For non-ionic liquid compounds, there are few more databases such as DIPPR [44].

Due to lack of enough experimental data for the thermophysical properties of ILs, use of other thermophysical properties of ILs as the parameters of a new model is not applicable and/or straightforward in most cases. Consequently, a chemical structure-based model can overcome such limitations and can be applied to develop more comprehensive and general models.

The aim of this study is to develop the best new accurate and predictive models for prediction of the thermophysical properties of ionic liquids which are more comprehensive but have less limitation. For this purpose, the largest possible databases are collected by the use of the NIST standard reference database and searching the literature for data published recently; so the more comprehensive and predictive models can be developed with the use of various types of ionic liquids and more experimental data points, compared with the available models in the literature.

Furthermore, it is targeted to develop some models for the prediction of infinite dilution activity coefficient (γ^∞) of organic solutes in various ionic liquids. This property has not been modeled by Group contribution (GC) or Quantitative Structure-Property Relationship (QSPR) to date. These novel models make the calculation/prediction of γ^∞ much easier compared with the UNIFAC model.

Another aspect of this study is to assess the use of the QSPR method for development of new models for ionic liquids. The common approach in the model development for ionic liquids is the GC method and consequently, there is an opportunity to develop some new and predictive QSPR models for the first time for some thermophysical properties of ionic liquids.

The main approach of the model development in this thesis is the use of chemical structure-based parameters to develop accurate and predictive property estimation models for ionic liquids. The number of these parameters is more than a thousand for ionic liquids and when a temperature dependent property is studied, it may go up to 40,000 parameters (all parameters are multiplied by each other or by a function of the temperature; e.g. $\frac{a}{b}, \frac{a}{b^2}, \frac{a}{b^3}$). As a result, novel and different mathematical approaches and feature selection methods should be applied to choose the most effective parameters on the target property.

To achieve this goal, there are several steps that have to be undertaken to produce a predictive model. In this thesis, the procedure is as follows:

- 1- Each property was selected carefully. The selected property should have reasonable number of data points and the previous models in the literature should have at least one drawback such as complexity, low accuracy, high number of parameters, limited supported compounds, etc.
- 2- It was tried to collect all available data into a dataset using the NIST standard reference database and searching the literature for recent published data.
- 3- The dataset was examined carefully to screen the duplicated and erroneous data.
- 4- A mathematical method was selected and the model was developed. If the model did not have good accuracy, the mathematical method was changed.
- 5- The model was validated through different statistical methods and its prediction power was evaluated.
- 6- Finally, if it is applicable, the model developed was compared with previous models in term of simplicity, accuracy, prediction power, and comprehensiveness.

All the steps will be explained in detail in other chapters of the thesis. Chapter 2 covers the literature review of the properties studied. In chapter 3, the data management procedure is described. Afterward, the methods used for modeling are explained in chapter 4. Thereafter, chapter 5 discusses about the model development procedure. Chapter 6 includes the result of modeling for physico-chemical properties studied. The conclusion of this thesis is presented in Chapter 7. Finally, Chapter 8 offers some recommendations for future works.

CHAPTER 2: LITERATURE REVIEW

2.1 Scope

In previous chapter, it was discussed that there are three types of models used for property estimation of chemical compounds. For non-ionic liquid compounds, it is common to use the other thermophysical properties of the substances to estimate the desired one. The most famous type is corresponding state models which use the critical properties and relate them to the target property. A collection of such models are available in “The properties of gases and liquids” text book [45].

Ionic liquids have strong inter/intra-molecular interactions and consequently, their theoretical boiling point and critical properties are above the decomposition temperature [46]. As a result, corresponding state modeling is not applicable for the ionic liquids. Another option is to use the other thermophysical properties to model the desired one; but as explained earlier, it is not possible mostly due to lack of enough experimental data for various thermophysical properties of ionic liquids. Consequently, the best available option is the application of chemical structure-based models.

The chemical structure-based models are classified as follows and will be explained in detail in section 4.2.

- 1- Group contribution (GC) models.
- 2- Quantitative structure-property relationship (QSPR) models.

This chapter covers the literature review of thermophysical properties of ILs studied in this thesis. These properties are speed of sound, heat capacity, refractive index, viscosity of fluorine-containing ionic liquids, infinite dilution activity coefficient (γ^∞) of organic solutes in ionic liquids, and the theoretical critical temperature of ionic liquids.

2.2 Speed of sound in ionic liquids

In chemistry and physics, the speed of sound (u) is of great importance and can be used in modeling or deriving other thermophysical properties, such as isentropic and isothermal compressibility, thermal pressure coefficient, the reduced isobaric thermal expansion coefficient, isobaric and isochoric heat capacities, bulk modulus, and the Joule–Thomson coefficient. However, the speed of sound has not been extensively used to derive thermodynamic properties of ILs due to the short supply of experimental data [47-50].

The efforts for modeling the speed of sound in ILs are surprisingly very few. Gardas and Coutinho [48] conducted the first study in 2008 and modeled the 133 experimental data for 14 imidazolium-based ILs using the density and surface tension data. Their model was the modified correlation originally proposed by Auerbach [51] and had the average absolute relative deviation (AARD%) of 1.96%. Despite the relatively low deviation, the lack of experimental data for both the density and surface tension for each ionic liquid is the drawback of their model. To overcome this limitation, two separate correlations were developed to predict the density and surface tension of ILs. This brought about another drawback which required a lot of computations for the use of proposed model in the prediction of the speed of sound in ILs.

The next and only other study in the literature was undertaken by Singh and Singh [52]. The same approach as Gardas and Coutinho was used and a new model was developed for 3 imidazolium-based ILs.

The comparison of available models in the literature is summarized in Table 2.1.

Table 2.1: Summary of available models for the speed of sound in ionic liquids.

Model	Model Type and parameters	N_{ILs} *	N_{data} **	AARD%***
Gardas and Coutinho [48]	Correlation, ρ , σ	14	133	1.96
Singh and Singh [52]	Correlation, ρ , σ	3	60	n.a.

* Number of ionic liquids.

** Number of data points.

$$*** \text{AARD\%} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \times 100$$

2.3 Liquid heat capacity of ionic liquids

Heat capacity at constant pressure (C_p) is defined as the partial derivative of the enthalpy with respect to temperature while pressure kept constant. C_p is related to several thermodynamic properties such as: Gibbs free energy, enthalpy, and entropy and knowing its behavior helps for further study of other properties. In addition, C_p of both solid and liquid state of compounds is used to determine the heat transfer of equipment such as reactors and heat exchangers, which is the important parameter in their design [53].

There are few models available in the literature for the estimation of liquid heat capacity (C_{pL}) of ionic liquids. Gardas and Coutinho [54] used the second order group contribution method (GCM) and developed a 12-parameter model. They used a dataset comprising 19 ILs with 2396 data points over a wide temperature range of (196.36-663.10 K). Despite the good claimed prediction accuracy of the model, it must be noted that only 3 cations and 6 anions were used to develop the model, which limits its applicability. The number of model parameters is also relatively high when compared to the number of ionic liquids used (12 parameters for 19 ILs).

Soriano *et al.* [55] applied a similar approach to Gardas and Coutinho by using the entire structure of cations and anions instead of functional groups. Their dataset had 32 ILs and 3149 data points, with 10 cations and 14 anions, over a temperature range of (188.06-663.10 K). Since their model is dependent on certain type of ions used in its development, it cannot predict heat capacity for ILs for cation-anion combinations which were not in their dataset.

Valderrama *et al.* [56] applied a new approach for the prediction of C_{pL} which was called the mass connectivity index (MCI). It was based on the molecular connectivity concept introduced by Randic [57]. The model was developed from a dataset consisting of 541 data points for 15 ILs. The 40-parameter group contribution model was obtained with two specific constants for each ionic liquid. In another study, Valderrama [58] used the same approach as their previous study, but this time with 31 ILs and 477 data points and developed an artificial neural network (ANN) model for the prediction of C_p . The drawback of their models though is the use of a relatively small dataset and calculation of constants for a limited number of ionic liquids which reduces the model applicability for new ionic liquids with difference cation-anion combinations.

In addition to the models listed above, there are several studies [54, 59-61] published in literature for the prediction of C_{pL} with respect to molar volume (V_m). The major problem with this approach is the need in the model for experimental values for V_m for ionic liquids. This property is generally not readily available for most of the ILs.

The summary of the previous models are shown in Table 2.2.

Table 2.2. Summary of available models for the of heat capacity of ILs.

Model	Model Type and parameters	N _{ILs}	N _{data}	AARD%
Gardas and Coutinho [54]	GC, 12 parameters	19	2396	0.36
Gardas and Coutinho [54]	Correlation, V_m	19	2396	1.85
Soriano <i>et al.</i> [55]	GC, 10 cations and 14 anions	32	2414	0.34
Valderrama <i>et al.</i> [56]	MCI, 40 parameters	15	541	0.8
Paulechka <i>et al.</i> [60]	Correlation, V_m	19	653	6.0 (max error)
Preiss <i>et al.</i> [59]	Correlation, V_m	20	n.a	1.2 (max error)

2.4 Refractive Index of ionic liquids

Refractive index (n_D) is defined by IUPAC as “the ratio of the speed of light in vacuum to that in a given medium” [62]. Refractive index is a fundamental physical property, especially for optics related research fields, and it is used to verify a material and its purity, or to measure the concentration of a mixture. It can also provide valuable information when studying the intermolecular forces or the behavior of molecules in solution. Refractive index is also related to other properties such as the density, surface tension, and dielectric constant through thermodynamic equations [63, 64].

Surprisingly, little attention has been given to modeling and also measuring the refractive index of ILs, despite the simplicity of measurement. For other compounds however, several

models have been developed and published [65-70]. In the NIST Standard Reference Database #103b, there are experimental data available for more than 700 ILs, but for only 97 of them has the n_D been reported.

The first method for the prediction of n_D of ILs was proposed by Deetlefs *et al.* [71] whom related the n_D of 9 methylimidazolium based ILs using Equation (2.1) to the surface tension (σ), molar refraction (R_M), and parachor (P) of the molecule.

2.1

$$n_D^2 - 1 = \frac{P}{R_M} + \frac{\sigma}{R_M^2}$$

All of the parameters in this model are required to be experimentally available or correlated by other experimental properties, and in case of new compounds, this model therefore needs other models to predict the parameters. This unfortunately results in the increased errors in the prediction of n_D . The model has been developed based on 9 ILs, and consequently its applicability is limited and its results cannot be generalized. Even with the limited application, the model is not good and its average absolute relative deviation (AARD%) is 6.4%.

Gardas and Coutinho [63] developed a 10-parameter group contribution model using 24 imidazolium based ILs having 7 different anions. The AARD% of their model was 0.18%, but it couldn't predict the n_D of ILs with different cations. The same approach was used by Soriano *et al.* [72] and Freire *et al.* [73] using a few new imidazolium ILs, but their models suffered the same limitations as that of Gardas and Coutinho.

Table 2.3: Summary of different models for refractive index of ILs.

Model	Model Type and parameters	N_{ILs}	N_{data}	AARD%
Deetlefs <i>et al.</i> [71]	Correlation	9	9	6.4
Gardas and Coutinho [63]	GC, 10 parameters	24	245	0.18

2.5 Viscosity of fluorine-containing ionic liquids (F-ILs)

One of the barriers against replacing commonly used solvents with ionic liquids in various applications is the relatively high viscosity (η) of ILs which results in low diffusion coefficients, slow mass transfer, and low electrical conductivity [74, 75]. Therefore research has to be undertaken to synthesize ILs with low viscosity and low melting points. Initial investigations have shown that the insertion of a fluorine atoms (typically CF_3 groups) into the chemical structure of the cation or in most cases the anion can reduce the viscosity as well as the melting point of ILs [74, 76-79].

ILs consist of ions, viz. a cation and an anion. In theory the cation and anion can be selected to have molecules with desired properties for a particular application. As such, ILs have been termed as “tunable” and “designable” materials [80]. Consequently, the values of viscosity can be tuned by selecting appropriate combinations of ions, especially fluorine-containing ions. The process of selection can be sped up by using a predictive model which has the ability to correlate the viscosity data based on chemical structure of ions or other thermophysical properties.

Abbott [81] recommended the application of hole theory for the definition of the viscosity of ILs and molecular liquids. Despite the introduction of this theory for the estimation of viscosity, the model did not produce satisfactory predictions for ILs and had a percentage average absolute relative deviation (AARD%) of 122%. Bandres et al. [82] predicted the viscosity of 8 pyridinium ionic liquids by modifying the hole theory suggested by Abbott [81]. They introduced an efficient IL radius (R^*) to improve the estimations. The R^* was fitted to real/laboratory viscosity data points at a pressure of 0.1 MPa. Consequently, their method yielded an AARD% of 4.5%. The weakness of their model is its lack of predictive capability for ILs without any experimental data, as the model needs the experimental data to evaluate R^* .

Gardas and Coutinho [83] presented a group contribution (GC) model which could predict the viscosity of ionic liquids based on an Orrick–Erbar-type [84] equation. They developed the model by using 500 data points for 30 ionic liquids with 8 anion and 3 cation types. In terms of performance, the AARD% of the model is 7.78%. The model developed however requires the density of the ionic liquids as an input, which may be considered as a

disadvantage of the approach [85]. This drawback was solved by Gardas and Coutinho [86] by developing a new model for 25 ILs based on the Vogel-Tammann-Fulcher (VTF) equation. Despite this improvement, the main disadvantage of the two models of Gardas and Coutino is the limited number of cations and anions used to develop the models which limits the applicability of the model for ionic liquids.

The first nonlinear QSPR based model was developed by Yamomoto [87] for 62 ILs with a R^2 of 0.9464 . The model shows fairly good results, but it is limited to 6 anions. Thereafter Yamomoto and co-workers [88, 89] developed new nonlinear GC and QSPR models for more than 300 data points. Their last model has a R^2 of 0.931 and an average error of 5.04% on a logarithmic scale which translates into a linear scale average error of 21.84%. This model has the same weakness in terms of applicability for a limited number of cations and anions.

Bini *et al.* [90] presented two 4-parameter QSPR models for 33 ILs at two different temperatures (293 K and 353 K). Their model fails to correlate/predict nitrile-functionalized cations and it is also limited to a relatively small number of ionic liquids.

Gharagheizi *et al.* [91] reported a 47-parameter GC model for 443 ILs with 1672 data points. The model has a R^2 of 0.874 and an average error of 6.32% in a logarithmic scale; however the “real” average error in linear scale is nearly 31%. The data used to develop the model contained 638 duplicated experimental values; thus only 1034 of the data points used were unique. Consequently, by removing the duplicated values, the average error in terms of $\log \eta$ and η increases to 7.1% and 35.7% respectively. In addition, the database used contained 724 data points for fluorine-containing ionic liquids (F-ILs) which was less than the dataset used in this study (863 data points). The average error of this model for only F-ILs is 6.7% and 59.7% for logarithmic and linear scales. Obviously, this model fails to predict the viscosity of F-ILs.

There have been other studies by researchers [92, 93] on Neural Network modeling for viscosity, but these have been for limited numbers of ILs at a single temperature.

The summary of the previous models are shown in Table 2.2 .

Table 2.4: Summary of different models for predicting the $\ln(\eta)$ of ILs.

Model	Model Type and parameters	N _{ILs}	N _{data}	AARD%	Comments
Abbott [81]	Correlation	11	n.a.	122	10 ILs contained fluorine atom.
Bandres et al. [82]	Correlation	8	n.a.	4.5	7 ILs contained fluorine atom.
Gardas and Coutinho [83]	GC, 13 parameters, ρ	30	500	7.78*	Only 19 ILs contained fluorine atom. *AARD% is in linear scale.
Gardas and Coutinho [86]	GC, 12 parameters	25	482	7.50*	Only 16 ILs contained fluorine atom. *AARD% is in linear scale.
Tochigi and Yamomoto [89]	QSPR, 24 parameters	161	334	5.04	149 ILs contained fluorine atom. Most of the ILs had just one data point.
Bini <i>et al.</i> [90]	QSPR, 4 parameters.	33	66	n.a.	Authors proposed two models for T = 293 K and T = 353 K.
Gharagheizi <i>et al.</i> [91]	GC, 47 parameters	443	1672	6.32	638 duplicated data points were used. By removing the duplicates, the AD% was risen to 7.1%. 724 data points were for F-ILs
Valderrama <i>et al.</i> [92]	ANN	58	327	n.a.	
Billard <i>et al.</i> [93]	ANN	99	99	10	The data were only at 298 K.

2.6 Infinite dilution activity coefficient of organic solutes in ionic liquids

As discussed in Chapter 1, the unique properties of ILs make them to be looked upon as an alternative for conventional volatile, often toxic and flammable organic solvents in various applications such as extraction and separation processes, waste recycling, electroplating, gas separation and CO₂ capturing, and many more.

Among these, the main application of ILs is extraction and extractive distillation processes in which ionic liquids can be utilized as the solvent. For this purpose the knowledge of activity coefficient at infinite dilution (γ^∞) would be inevitable and numerous studies have been carried out to find this useful property in different mixtures [94-133]. Infinite dilution activity coefficients is also very beneficial in illustrating the behavior of liquid mixtures, estimating the mutual solubilities, predicting the existence of an azeotrope, analytical chromatography, calculating the Henry constant and partition coefficients, fitting the excess molar Gibbs energy (G^E) parameters of the models such as Wilson, NRTL and UNIQUAC, and developing the thermodynamic models based on the group contribution methods such as original and modified UNIFAC [134, 135].

In fact, infinite dilution activity coefficient is a hard property to model, especially for ionic liquids, due to presence of several interactions among three components in the system (cation, anion, solute). In addition, the values of γ^∞ are dependent on the method of measurement, such as inverse gas chromatography and diluter methods, or solvent column packing concentrations [136-138]. So the reported values of γ^∞ vary if it is not measured with the same method and at the same conditions. As a result, it is hard to develop a model to predict the γ^∞ with very low deviation as there is reasonably high scatter in experimental data.

The predictive models of activity coefficient of solutes in ionic liquids are categorized as group contribution and COSMO-based methods. The most well-known group contribution models for the prediction of γ^∞ are the original and modified UNIFAC (Uniquac Quasi-Chemical Functional Activity Coefficient). The modified UNIFAC (Dortmund) model just has a small difference in evaluation of interaction parameters which is shown in equation (2.11). The activity coefficient in UNIFAC model is described as follows [134, 135, 139]:

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \tag{2.2}$$

where, $\ln \gamma_i^C$, is the combinatorial term defined by:

$$\ln \gamma_i^C = \left(\ln \frac{\Phi_i}{z_i} + 1 - \frac{\Phi_i}{z_i} \right) - 5q_i \left(\ln \frac{\Phi_i}{\theta_i} + 1 - \frac{\Phi_i}{\theta_i} \right) \tag{2.3}$$

in which z_i is the mole fraction of component i , θ_i is the area function and Φ_i is the segment function which are described by following equations:

$$\theta_i = \frac{q_i z_i}{\sum_j q_j z_j} \quad 2.4$$

$$\Phi_i = \frac{r_i z_i}{\sum_j r_j z_j} \quad 2.5$$

$$r_i = \sum_k v_{ki} R_k \quad 2.6$$

$$q_i = \sum_k v_{ki} Q_k \quad 2.7$$

where v_{ki} is the number of groups of type k in molecule i , R_k and Q_k are UNIFAC volume and surface area parameters respectively which need to be adjusted. The interaction between the molecules have been taken to account by residual part of equation (2.2), $\ln \gamma_i^R$, which is expressed by following equation [139] :

$$\ln \gamma_i^R = \sum_k v_{ki} (\ln \Gamma_k - \ln \Gamma_k^i) \quad 2.8$$

where,

$$\ln \Gamma_k = Q_k \left[1 - \ln \left(\sum_m \theta_m \Psi_{mk} \right) - \sum_m \left(\frac{\theta_m \Psi_{km}}{\sum_n \theta_n \Psi_{nm}} \right) \right] \quad 2.9$$

The following equations have been used for evaluation of the interaction parameters, Ψ , of the original and mod. UNIFAC activity model respectively [134, 135]:

$$\Psi_{mn} = \exp \left[- \left(\frac{a_{mn}}{T} \right) \right] \quad 2.10$$

$$\Psi_{mn} = \exp\left[-\left(\frac{a_{mn} + b_{mn}T + c_{mn}T^2}{T}\right)\right]$$

in which a and b , are the temperature-independent group interaction parameters between main groups m and n and reliable experimental data are required for their fittings.

As it can be observed from equation (2.11), the number of fitting parameters intensively depends on the number of subgroups of the considered molecules; so the more complicated molecules such as ILs can produce large number of fitting parameters which increases the complexity of the model and difficulty of its application. In addition, the interaction parameters are not available for all pair of subgroups, especially for ionic liquids, which may limit the use of the model for certain compounds [140].

COSMO-RS (COnductor-like Screening MOdel for Real Solvents) was developed by Klamt [141] which uses quantum chemistry calculations to predict various thermophysical properties of chemical equilibrium in liquid/liquid and vapor/liquid systems. The major difference between COSMO-RS and GCMs is that the predictive ability of a GC model depends on the availability of group interaction parameters, but the only limitation of COSMO-RS is the availability of individual component parameters [140]. COSMO-based models have been parameterized using the experimental data of several molecules covering most of the chemical functionality of these 10 elements: hydrogen, carbon, nitrogen, oxygen, fluorine, phosphorus, sulfur, chlorine, bromine, and iodine [142]. So they have few numbers of parameters, but the calculations are very time consuming and need powerful computers, especially for complex and large molecules such as ionic liquids [143]. In addition, the calculations need the use of commercial software which is not freely available.

Consequently, there is not any easy-to-use GC or QSPR model in the literature to predict the γ^∞ of different solutes in ionic liquids.

2.7 Critical temperature of ionic liquids

The critical temperature (T_c) of a substance is the temperature at and above which vapor of the substance cannot be liquefied, no matter how much pressure is applied. The behavior of a

fluid near its critical point is a specific property necessary to develop the thermodynamic models for fluids [144].

Due to presence of high electrostatic interactions, as well as short-range van der Waals interactions in ionic liquids, the critical properties of ionic liquids usually cannot be measured as they are decomposed before their normal boiling temperature (T_b) are reached [46]. As a result, the experimental data of critical properties of ILs as well as their T_b are not available in the literature.

Despite this limitation for the ionic liquids, some researchers tried to use the available correlations and estimation methods of common organic compounds to estimate the critical properties of ILs. The most accepted approach is to use the relationship between critical temperature, density, and surface tension and thereafter, estimate the T_c of ionic liquids. According to the fact that the surface tension becomes zero at critical temperature, two correlations were presented by Eötvös [145] and Guggenheim [146] which are shown in equation (2.12) and (2.13), respectively.

$$\frac{\gamma}{\rho M} = K(T_c - T) \tag{2.12}$$

$$\gamma = K(T_c - T)^{2/3} \tag{2.13}$$

where γ is surface tension, ρ is density, M is molar mass, k and K are two different constants. These equations usually can estimate the T_c of organic compound pretty well; so it is assumed that they are applicable for ionic liquids too. This approach has been followed by the majority of researchers [144, 147-152] to calculate and report the critical temperature of ionic liquids.

Another approach was followed by Valderrama *et al.* [46] where they used the Lydersen group contribution method to estimate the critical pressure and critical volume, and the Joback and Reid group contribution method to calculate the normal boiling temperature and the critical temperature of ionic liquids. They combined the best results of Lydersen's method

with the best results of the Joback-Reid method to propose a “modified Lydersen-Joback-Reid” method and claimed that this modified method had good results for molecules of high molecular weight. Thereafter, Valderrama *at al.* performed a consistency test and used a correlation for density to validate the estimated critical properties of ionic liquids. This correlation was developed earlier by Valderrama and Abu-Sharkh[153] to relate the critical properties of saturated liquids and petroleum fractions to their density. Consequently, they published the results of calculations for 1130 ionic liquids [154].

Despite their interesting approach, there is not a good agreement between the value of T_c estimated by their method and the first approach for most of the ionic liquids. However, these doubtful calculated critical properties are used by several researchers to develop the corresponding state models for prediction of thermophysical properties of ionic liquids [155-162].

As the experimental data of surface tension is limited for ILs, the more realistic value of T_c can be calculated only for less than 140 ionic liquids. As a result, the detailed analysis of Valderrama *at al.* model is not applicable; however, Table 2.5 shows the deviation of this model from T_c calculated from surface tension data for limited number of ionic liquids.

Table 2.5: Summary of Valderrama *at al.* model for T_c of ionic liquids.

Model	Model Type and parameters	N_{ILs}	N_{data}	AARD%	Comments
Valderrama <i>at al.</i> [154]	GC, 44 parameters	1130	1130	30.2	AARD% was calculated over 51 ionic liquids.

CHAPTER 3: DATA MANAGEMENT

3.1 Scope

The first step in development of a model is to find and gather the published data as much as possible. This can be done by searching the literature or using the comprehensive databases such as NIST Standard Data Reference Database [43] and Dortmund Data Bank [42].

Using such databases has some benefits. They help to save a lot of time by eliminating the process of searching for data, extracting them from the context of the papers, checking the correctness of typed data, etc. In addition, they have a software which can read the database and visualize the data by plotting them, or estimate some of the properties using integrated codes for equation of states and some other property estimation correlations.

3.2 Data source for the ionic liquids

In this research, the NIST Standard Data Reference Database was used as the main source of data for each property of ionic liquids. In addition, the literature search was done to gather the most recent published data which have not been included in NIST database. As a result, all of the properties studied in this thesis have much larger data points and ionic liquids compared with the properties modeled previously by other researchers. The information about the dataset of properties modeled will be shown in next pages.

3.3 Data screening

After gathering the data for the desired property, the next step is to analyze and screen the data, and remove the duplicates and erroneous data points. As the chosen properties are temperature dependent, except the critical temperature, precise analysis of data was required.

The data used in the model development was screened as follows:

- 1) Where there were several reported values of the desired property for a single temperature, the values were plotted against temperature to find and remove the outliers.

- 2) If there were only two sources, the values with the lowest uncertainty were incorporated into dataset utilized.
- 3) If the reported values had the equal uncertainties, the latest published values were utilized.
- 4) To verify the potential outliers in final data points, the target values were plotted against temperature and a (non-)linear curve fitting algorithm was applied to fit the data. The data points which were far from the fitted curve were considered as the outliers and omitted from the dataset.

3.4 Datasets of ionic liquids for different properties

In this section, the information about the datasets used for different properties of ionic liquids is listed. The information includes the number of ionic liquids as well the number of different cations and anions. All the studied properties are temperature dependent and they are analyzed and screened carefully.

3.4.1 Speed of sound in ionic liquids

The NIST Standard Reference Database #103b [43] was the source of data for the speed of sounds in ionic liquids. After screening the extracted data, the resultant dataset contained 446 reported experimental data points for 41 ILs at atmospheric pressure. In addition, the ILs were constituted from 29 types of cations and 11 types of anions. These ionic liquids belonged to the imidazolium, phosphonium, pyrrolidinium, pyridinium, and amino acid families. The data covered a wide range of temperatures (278.15-343.15 K) and values (1128.4- 1885.4 m s⁻¹).

The number of ionic liquids per family classification is presented in Figure 3.1. In addition, the database is provided as a supplementary file in the supplementary CD.

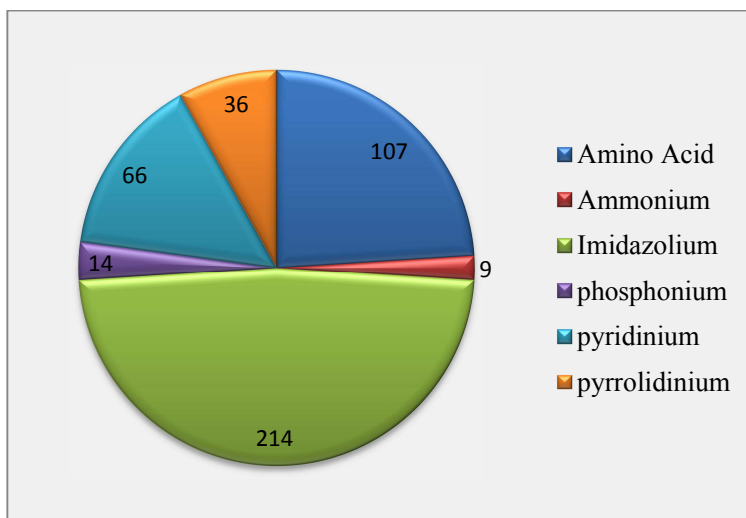


Figure 3.1: The number of ionic liquids in different families used for modeling the speed of sound.

3.4.2 Liquid heat capacity of ionic liquids

Same as the previous property, the reported experimental data of C_{pL} were extracted from the NIST Standard Reference Database #103b. The screening process resulted in a dataset comprising 82 unique ILs with 3726 experimental data points. In addition, the ILs were constituted from 39 types of cations and 24 types of anions.

The data covered a temperature range from (180.06-663 K) and C_{pL} values ranged from (259.09-1805.7 Jmol⁻¹K⁻¹). These ILs belonged to the imidazolium, phosphonium, pyrrolidinium, pyridinium, ammonium, and isoquinolinium class of ionic liquids.

The number of ionic liquids in different families is presented in Figure 3.2. In addition, the database is provided as a supplementary file in the supplementary CD.

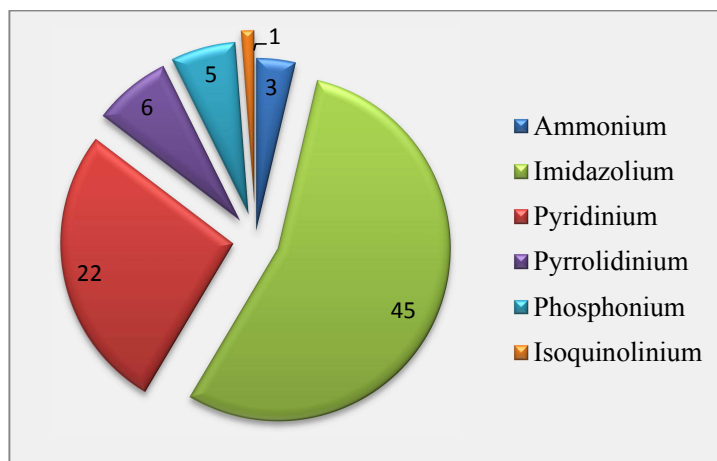


Figure 3.2: The number of ionic liquids in different families used for modeling the liquid heat capacity.

3.4.3 Refractive index of ionic liquids

To prepare a dataset for refractive index ($\lambda = 589 \text{ nm}$) of ionic liquids, the previous procedure was done and consequently, 931 experimental data points for 97 unique ILs constituted from 50 different types of cations and 33 types of anions were obtained. The data covered a temperature range from (283-363.15 K) and n_D values ranged from (1.3551-1.5778). The number of ionic liquids per family classification is presented in Figure 3.3.

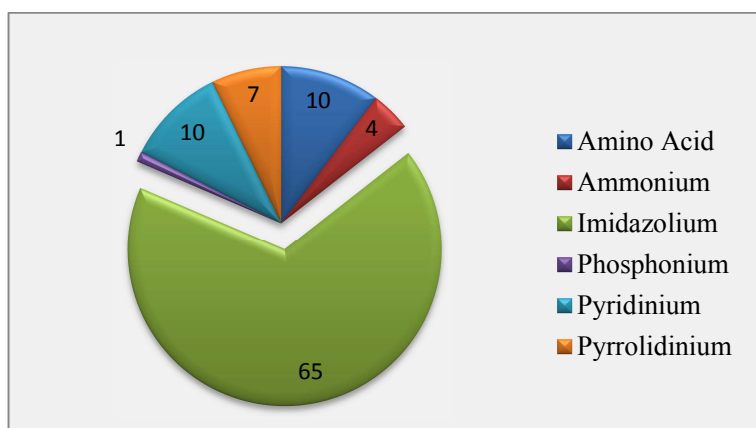


Figure 3.3. The number of ionic liquids in different families used for modeling the refractive index.

3.4.4 Viscosity of F-ILs

Viscosity of ionic liquids is one of the hardest properties to model. Because the presence of small amount of contaminants in ionic liquids changes the viscosity drastically. Seddon *et al.*[163] showed that contamination with water decreases the viscosity, while the presence of chloride ions increases the viscosity. As a result, it was required to select the experimental data of viscosity very carefully and use the reliable sources.

The NIST Standard Reference Database gathered the temperature dependent data of viscosity for 85 fluorine-containing ILs. There were in total 863 data points. The selected ILs were composed of 58 unique types of cations and 9 unique types of anions. The dataset has an extensive range of temperatures from 258.15 to 388.51 K and a wide range of viscosity values from 4.1 to 3067.3 cp. Figure 3.4 shows the number of ionic liquids per family classification including ammonium, imidazolium, phosphonium, pyridinium, pyrrolidinium, and quinolinium.

Zhang *et al.* [164] gathered lots of data from different sources such as patents and papers and published them as a text book entitled “Ionic Liquids:: Physicochemical Properties”, which has been used extensively by Gharagheizi *et al.* [91]. The screening procedure was performed on the dataset reported by Gharagheizi *et al.* and after removing lots of duplicated values, 297 new data points for 247 different ionic liquids were remained. It was observed that most of these new ionic liquids just had one data point at a constant temperature without any reported uncertainties. In addition, some data points seemed to be erroneous. For example, the viscosity of “1-butyl-3-[3-(2-hydroxybenzylamino)propyl]-3H-imidazolium hexafluorophosphate” was reported originally by Quadi *et al.* [165] as 257,000 cP which seems as outlier compared with viscosity of other ionic liquids. Unfortunately, Quadi *et al.* didn’t report the temperature of the measurement; so Gharagheizi *et al.* assumed that it was measured at 298.15 K. Most of the data of these new ionic liquids suffered from unreported value of temperature.

The initial steps of modeling revealed that these new data were not reliable as they had large deviations from the predicted values (mostly more than 70%). As a result, it was decided to ignore these data to avoid negative effects of accuracy of the model developed. However, another model was developed using the entire dataset (1160 data points for 332 ionic liquids) and the output of the model was illustrated in two separate figures in section 6.5. In addition,

the datasets used for two developed models as well as the models parameters is reported in the supplementary CD.

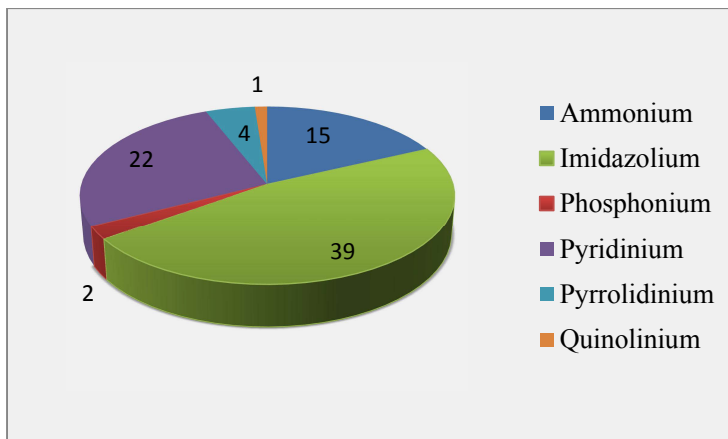


Figure 3.4: The number of ionic liquids in different families used to model the viscosity (refined dataset).

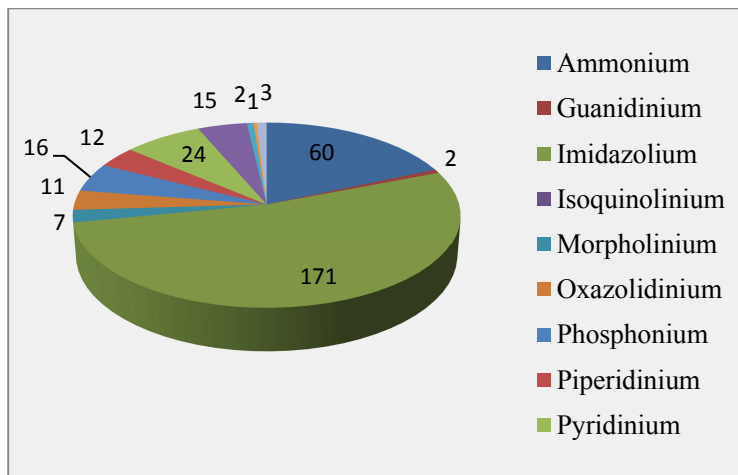


Figure 3.5: The number of ionic liquids in different families used to model the viscosity (unreliable dataset).

3.4.5 Infinite dilution activity coefficient of organic solutes in ionic liquids

To gather the most comprehensive and up-to-date database, the NIST Standard Reference Database #103b and several other papers [94, 96, 97, 99-106, 111, 112, 114, 116, 118, 121, 124,

152, 166-177] were used to extract the γ^∞ data of different solutes in ionic liquids. The entire screened database consisted of 20476 data points for 136 solutes in 126 ionic liquids.

As mentioned in section 2.6, there is not any easy-to-use model for prediction of γ^∞ of solutes in ionic liquids. During the model development, it was observed that it is almost impossible to develop a single model for all kinds of solutes. As a result, it was decided to break the database of γ^∞ of solutes in ILs into several datasets with respect to the chemical families of solutes including alkanes, alkenes, aromatics, alcohols, and so on.

3.4.5.1 γ^∞ of aromatic solutes in ionic liquids

The dataset of γ^∞ of aromatic solutes in ILs was refined as explained in section 3.3. In addition, it was observed that some outliers existed in dataset as presented in Figure 3.6 and Figure 3.7. In such cases, the γ^∞ data of each system was plotted against temperature and the outliers were removed visually. Accordingly, the data of 42 solute-IL systems were refined by this method and the outliers were removed. As a result, the final dataset contained 1653 experimental data points for 10 different solutes in 123 unique ionic liquids which resulted in 354 unique solute-IL systems.

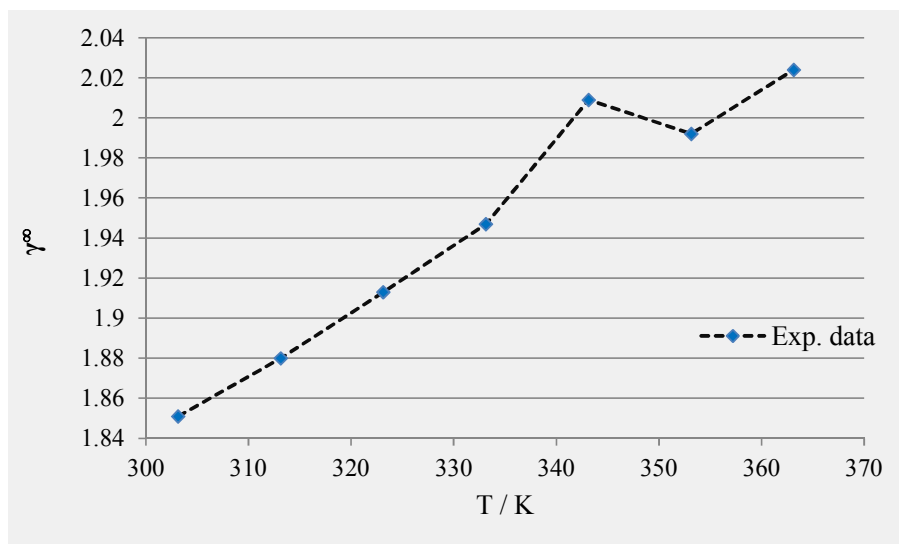


Figure 3.6: The γ^∞ data of “benzene” in “1-butyl-3-methylimidazolium nitrate”.

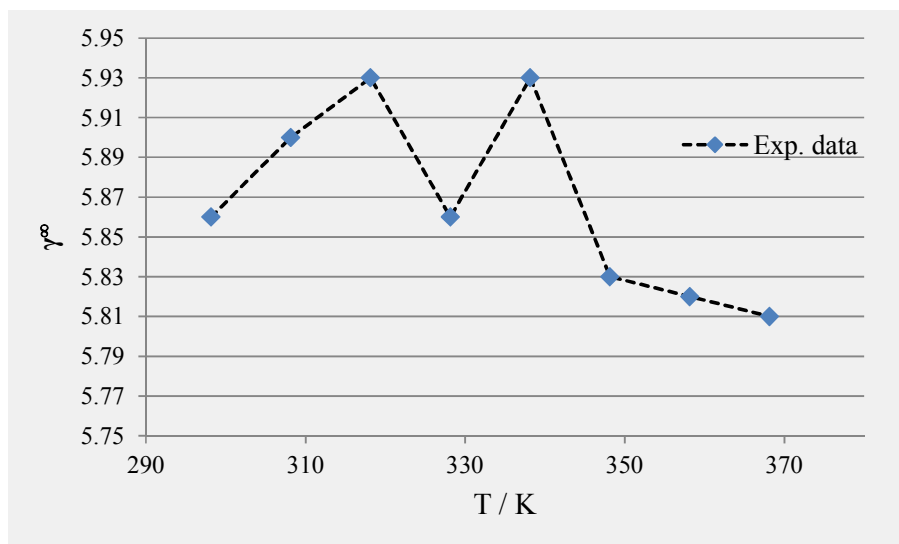


Figure 3.7: The γ^∞ data of “ethylbenzene” in “1-butyl-3-methylimidazolium thiocyanate”.

The ILs studied were constituted of 55 types of cations and 32 types of anions, and belonged to the imidazolium, guanidinium, isoquinolinium, phosphonium, piperidinium, pyridinium, pyrrolidinium, morpholinium, sulphonium, and ammonium class of ILs. The number of data points in each class of ionic liquids is illustrated in Figure 3.8.

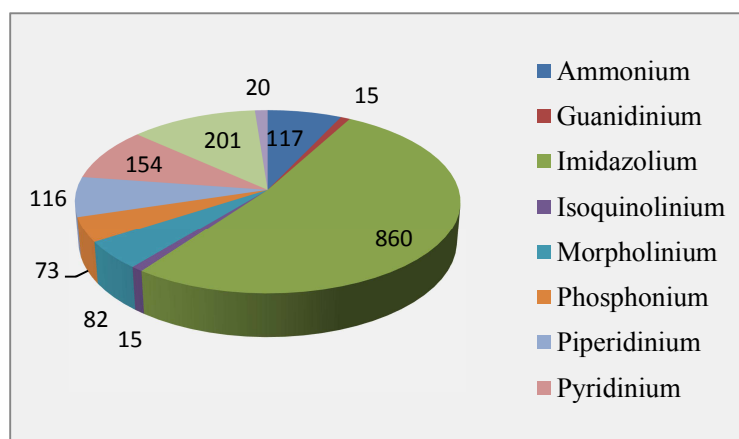


Figure 3.8: Number of data points of γ^∞ of aromatic solutes in each class of ionic liquids.

The complete dataset, including the original source of experimental data, is provided as a supplementary file in the supplementary CD.

3.4.5.2 γ^∞ of alcohol solutes in ionic liquids

The same screening procedure was performed for alcohol solutes and a dataset of γ^∞ data for 17 different alcohols in 126 unique ILs were extracted. According to the previous section, outliers were detected visually and removed for 54 solute-IL systems. Two sample systems are shown in Figure 3.9 and Figure 3.10. Consequently, the dataset had 2785 experimental data points which covered a temperature range from (293.15-413 K) and γ^∞ values ranged from (0.029-12.42) for 615 solute-IL systems.

These ILs were constituted of 54 types of cations and 28 types of anions and belonged to the ammonium, guanidinium, imidazolium, isoquinolinium, morpholinium, phosphonium, piperidinium, pyridinium, pyrrolidinium, and sulphonium class of ionic liquids. The number of data points of each class is shown in Figure 3.11.

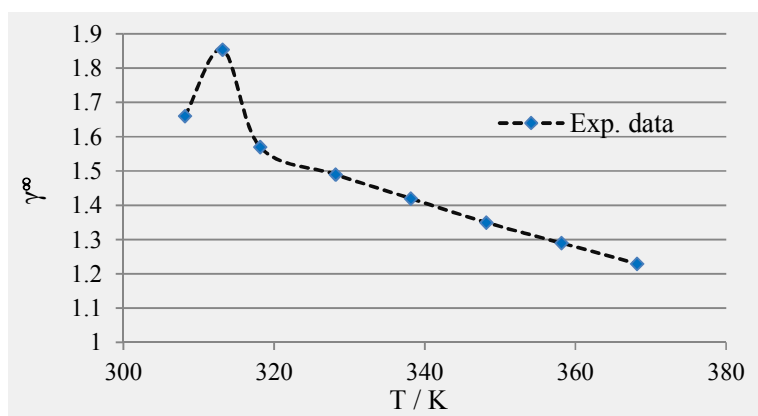


Figure 3.9: The γ^∞ data of “1-butanol” in “1-butyl-3-methylimidazolium trifluoromethanesulfonate”.

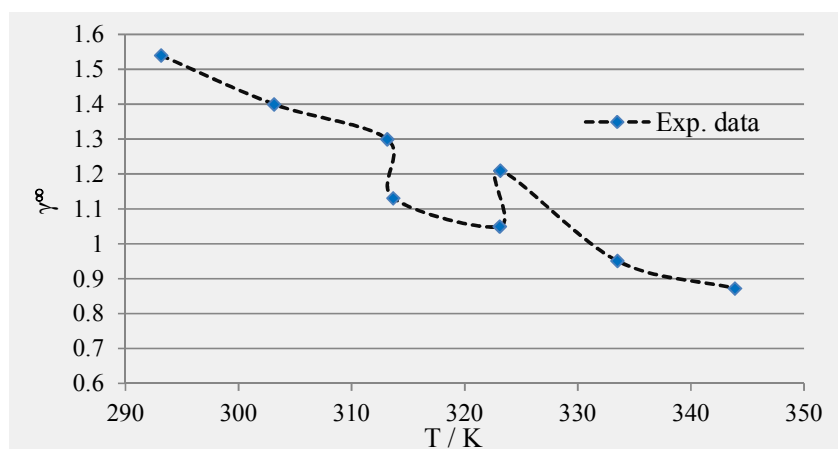


Figure 3.10: The γ^∞ data of “methanol” in “1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide”.

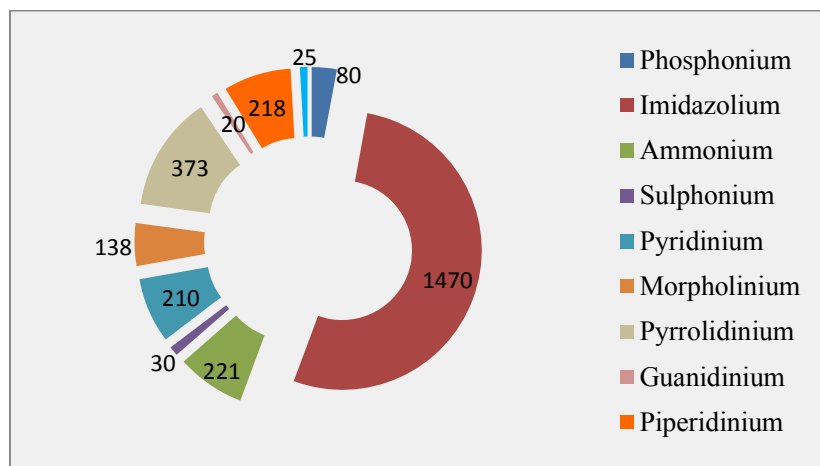


Figure 3.11: Number of data points of γ^∞ of alcohol solutes in each class of ionic liquids.

3.4.5.3 γ^∞ of alkane solutes in ionic liquids

Similar procedure was done for alkanes to extract and refine the data. The outliers were removed for 34 systems and the resultant database had 3935 experimental data points which covered a temperature range from (293.15-374.95 K) and γ^∞ values ranged from (0.104-4065.351) for 882 solute-IL systems. In addition, the database consisted of 18 solutes and 123 unique ionic liquids constituted from 55 types of cations and 36 types of anions. The number of data points per family classification is shown in Figure 3.12.

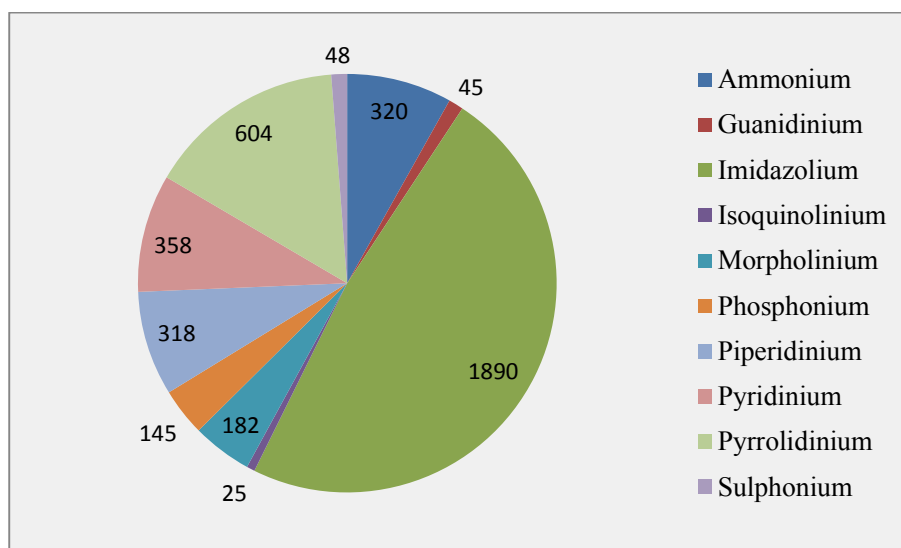


Figure 3.12: Number of data points of γ^∞ of alkane solutes in each class of ionic liquids.

3.4.5.4 γ^∞ of alkene solutes in ionic liquids

For the alkene solutes, the refined database had 2011 experimental data points which covered a temperature range from (293.15-375.05 K) and γ^∞ values ranged from (0.153-374) for 422 solute-IL systems. In addition, the database consisted of 13 solutes and 123 unique ionic liquids constituted from 55 types of cations and 34 types of anions.

The number of data points of each class is shown in Figure 3.13.

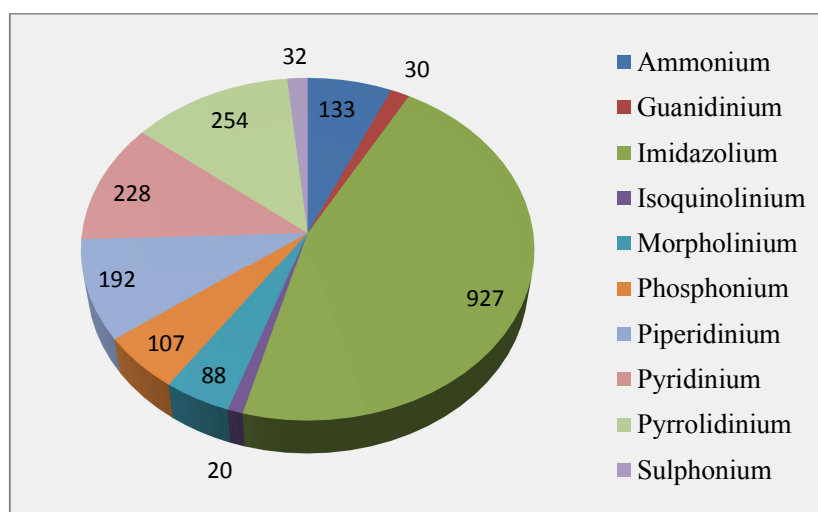


Figure 3.13: Number of data points of γ^∞ of alkene solutes in each class of ionic liquids.

3.4.5.5 γ^∞ of alkyne solutes in ionic liquids

For the alkyne solutes, all available experimental data were screened and the outliers were removed. As a result, 1257 data points were remained which covered a temperature range from (298.15-373.15 K) and γ^∞ values ranged from (0.28-96.61) for 270 systems. The database consisted of 6 solutes and 84 ionic liquids constituted from 44 types of cations and 26 types of anions.

The number of data points of each class is shown in Figure 3.14.

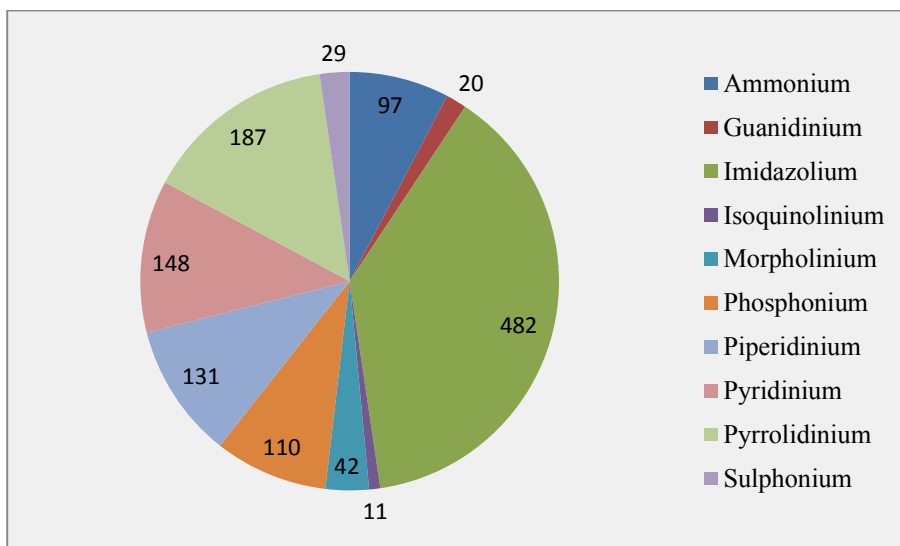


Figure 3.14: Number of data points of γ^∞ of alkyne solutes in each class of ionic liquids.

3.4.6 Critical temperature of ionic liquids

As explained in section 2.7, the real critical temperature of ionic liquids cannot be measured; because they are decomposed before they reach to the boiling point temperature and/or the critical temperature. As a result, the critical temperature can only be estimated in theory. The more realistic value of critical temperature should be calculated by Eötvös and/or Guggenheim equations. The previous studies [144, 147] show that the T_c calculated by these equations are close to each other (less than 100 °C in most cases); however the Eötvös equation is more prone to show deviations due to use of two sources of experimental data as inputs (surface tension and density). In addition, it is hard to find the experimental data of surface tension and density measured by the same laboratories at the same conditions. As a result, the Guggenheim equation was chosen for the estimation of critical temperature of ionic liquids.

The NIST Standard Reference Database and some newly published papers [178-182] were used to extract the experimental data of surface tension of ionic liquids. The result of the screening process was 1513 data points for 139 ionic liquids constituted from 65 types of cations and 42 types of anions. Figure 3.15 shows the number of ionic liquids in each family.

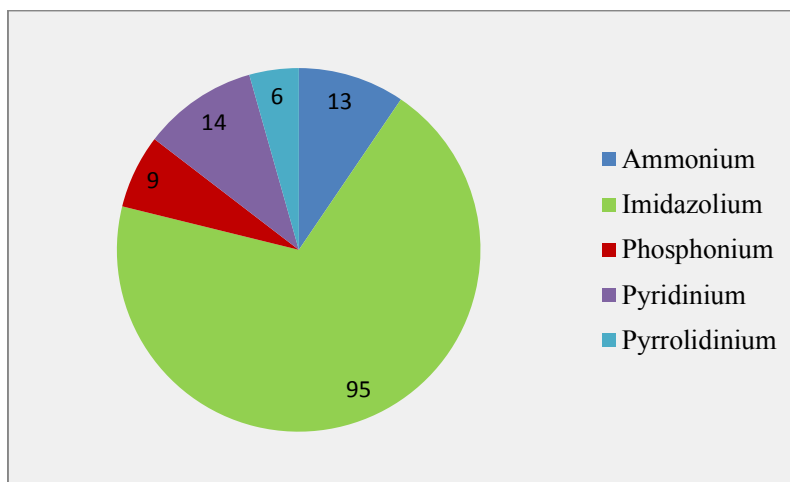


Figure 3.15: The number of ionic liquids in different families used for modeling the critical temperature.

To calculate the critical temperature of ionic liquids, the Guggenheim equation was used to fit the surface tension data. Unfortunately, 33 ILs had only one data point; so the T_c could not be calculated for them. As a result, the critical temperature was calculated for 106 ionic liquids.

In another attempt, it was tried to model the surface tension data and calculate the critical temperature thereafter; but it was observed that a small deviation in the prediction of surface tension, resulted in very large deviations in the calculated values of T_c compared with equation (2.13). As a result, this approach was not successful for the estimation of critical temperature of ionic liquids and only the values calculated by Guggenheim equation was used for model development.

CHAPTER 4: COMPUTATIONAL METHODS

4.1 Scope

This chapter covers all computational and mathematical steps required during the model development process. These steps cover all aspects from the input of data to the model outputs. The procedure and required steps of modeling the physico-chemical properties of chemical compounds are illustrated in Figure 4.1. Accordingly, these steps are discussed in next sections.

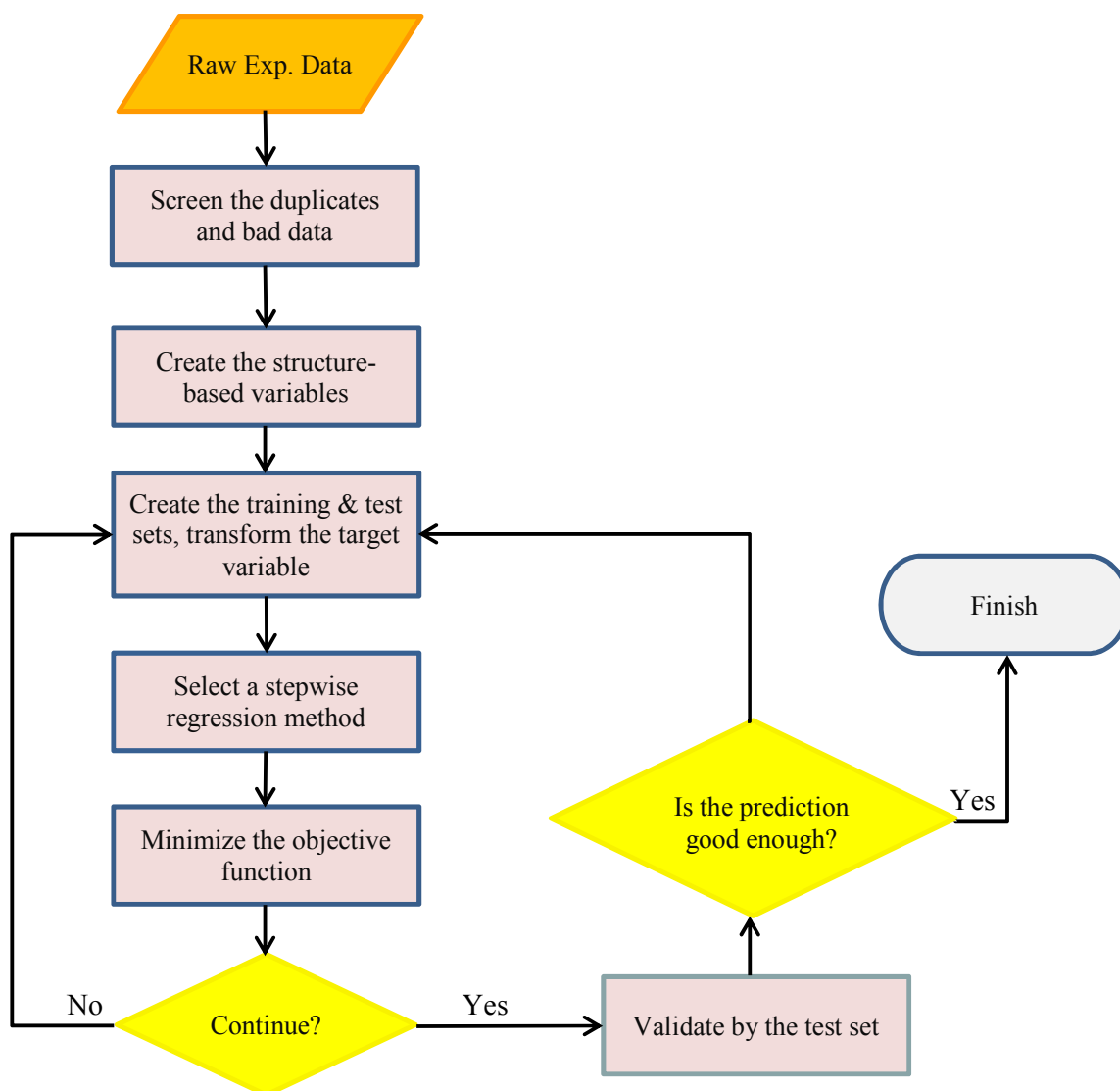


Figure 4.1: The flow diagram of developing a property estimation model.

4.2 Property Estimation Methods

To start modeling the physico-chemical properties of ionic liquids, it is required to select a structure-property estimation method. In the literature, various methods are reported which can be broadly categorized as GC-based methods, QSPR-based methods, and molecular simulations [183].

4.2.1 Group Contribution Methods

The Group Contribution (GC) method was initially developed by Riedel [184] and Lyndersen [185] to estimate the critical properties of pure component from their molecular structure. Since there are an unlimited number of chemical compounds, but the number of functional groups is limited, it is more convenient to use functional groups of existing data to develop a model and then, apply that model to predict the properties of new compounds. Figure 4.2 shows the common functional groups in “acetone” molecule.

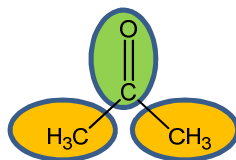


Figure 4.2: Functional groups in “acetone” molecule.

Group contribution methods are based on the so called “additive principle”. That means any compound can be divided into fragments, usually atoms, bonds or group of atoms, etc. All fragments have a partial value called a contribution. The simplest form of GC is the calculation of the physical property by summing up the product of the contributions made by structural groups in the individual molecule and the number of times each group appears. This approach was first used by Joback and Reid [186] to predict the thermo-physical and transport properties of pure compounds [183, 187].

The disadvantage of this method is that it cannot differentiate the isomers and calculated the same result for all of them. In addition, GC method generally cannot deal with proximity effects of groups within the molecules such as inter/intra-molecular interactions, Hydrogen bonding, etc. Moreover, reliable GC models usually can be developed for a limited number of

thermodynamic properties and representing all atomic arrangements is not possible. As a result, it is difficult to have an accurate prediction for complex molecules. Constantinou and Gani [188] tried to apply a correction by introducing the second order groups where they have the first order groups as building blocks. The second order groups are used as follows:

$$X(X) = \sum_i X_i^2 + W \sum_j X_j^2 \quad 4.1$$

where X is the property, X_i is the contribution of the first-order group type- i which occurs M_i times and X_j is the contribution of the second-order group type- j with M_j occurrences in a compound. The constant W is set equal to unity if the second order term is to be used. Similar approach was done later by Marrero and Gani [189] by introducing third-order groups.

Recently, a new GC method introduced by Wang, Ma and Neng [190] for estimation of critical properties, boiling point and melting point for organic compounds which is named the “position group contribution method”. This method differentiates isomers including cis- and trans- structures and takes into account the ortho, meta, and para corrections in benzene ring and pyridines [183].

Despite the benefits of “second”, “third”, and “position” group contribution methods, there is not any software for automating the calculations of groups; so the calculations need lots of time and efforts for large databases, such as properties used in this thesis. In addition, there will be the high risk of errors in manual calculations for large and complex molecules. As a result, the well-known “first” order group contribution method is used in this thesis.

4.2.2 QSPR Methods

The Quantitative Structure-Property Relationship (QSPR) is a well-established and highly respected technique to correlate diverse simple and complex physicochemical properties of a component by its molecular structure and interactions among different molecular groups based on their connectivities [191]. In this approach, it is assumed that the molecular features which are called “molecular descriptors”, affect directly the properties of compounds. Molecular descriptors are based on several different theories, such as quantum-chemistry,

information theory, organic chemistry, graph theory, and so on, and are used to model several different properties of chemicals in scientific fields such as toxicology, analytical chemistry, physical chemistry, and medicinal, pharmaceutical, and environmental chemistry [192].

The molecular descriptors are the ultimate result of a mathematical procedure which transforms chemical information encoded within a molecular formula into an useful number or the result of some standardized experiment [193]. As a result, there are some descriptors which can differentiate isomers and take different values for different isomers. This feature is the most important advantage of QSPR method over GC method. In addition, QSPR models usually have less parameter than GC models at same accuracy, which shows that the descriptors can correlate the target property better than simple functional groups; however manual calculation of descriptors is not possible and it is required to use the special software.

Figure 4.3 shows how the microstructure of the acetone molecule such as functional groups, 3D structure, topological characteristics, bond length, etc. can be used to create several descriptors.

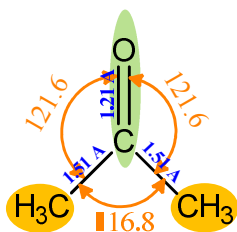


Figure 4.3: The schematic microstructure of “acetone” molecule.

4.3 Calculation of descriptors/functional groups

Ionic liquids consist of cations and anions. In order to associate the desired property with the constituent cation and anion combinations, the functional groups/descriptors of both ions should be calculated for each IL.

In GC method, only the number and types of functional groups/sub-structure are taken into accounts which are just related to the chemical structure of components, not the 3D structure of the molecules. But the majority of descriptors are calculated based on the 3D structure of

the molecules. So QSPR models can deal with isomers and proximity effects (inter/intra-molecular interactions).

In this thesis, the ChemDraw software [194] was used to draw the molecular structure of cations and anions. Thereafter, each chemical structure was saved in an MDL Molfile which holds information about the atoms, bonds, connectivity and coordinates of a molecule. In the next step, the molecular structures were imported separately into ChemAxon JChem Base software [195] to optimize the 3D structure by Dreiding Force field [196]. The optimized structures then saved as new MDL Molfiles.

In the next step, Dragon software [197] was used to calculate over 3000 molecular descriptors/functional groups by importing the optimized structures of all cations and anions, separately. These descriptors belong to 15 classes of descriptors: Constitutional descriptors; Topological indices; Connectivity indices; Walk and path counts; Information indices; Burden Eigen values; Edge-adjacency indices; Molecular properties; Functional group counts; Atom-centered fragments; Eigen value-based indices; topological charge indices; 2D binary finger print; 2D frequency finger print; 2D autocorrelations; and 3D conformational descriptors.

After calculating the structure-based variables for ionic liquids, there are two common approaches to make the final dataset:

- 1- Mixing the cations and anions descriptors together by adding the common descriptors of cations and anions to each other.
- 2- Using the cations and anions descriptors separately.

The second approach is beneficial for two reasons:

- a) It is easier to analyze the model and find out which descriptors of cations or anions have more effect on the model.

- b) It is much easier to calculate the property modeled for the new ionic liquids which their both cations and anions are available in the dataset. So there is no need to calculate the structure-based variables again.

In this thesis, the second approach was followed; so the cations and anions descriptors were collected in one dataset; but two suffixes were added to the name of descriptors to make the cations and anions descriptor distinguishable.

4.4 Variable reduction

After calculating the variables and gathering them together, the number of generated variables should be considered. Usually in GC modeling, the number of variables is not so much. In worst cases, it is less than 500 which is not a big deal in most cases. But in QSPR modeling, the variables are more than 3000 almost in all cases (summation of number of cations and anions descriptors).

To reduce the number of variables, there is a method called “pair correlation” which calculated the correlation coefficient of each pair of variables and allows the exclusion of one of the two descriptors with a correlation coefficient equal or greater than the selected threshold value. In GC modeling, it is common to set the threshold equal to 1.0 to remove the duplicated variables. In QSPR modeling, the common accepted threshold is 0.9 which usually drops the number of variables significantly to 400-800.

The pair correlation can be applied on all variables or on cations and anions variables, separately. In this thesis, second method was used for all properties modeled.

4.5 Dataset partitioning

One of the critical steps during the development of a “predictive” model is to have an external dataset for validating the model in terms of prediction ability. The common accepted method is dividing the entire dataset into two subsets: a “training” and a “test” datasets. The “training” set is utilized to develop the model and should contain different classes of

compounds to have a comprehensive model. On the other hand, the “test” set is implemented to evaluate the reliability and prediction ability of the model and like the “training set”, it should have different classes of compounds. In case of ionic liquids, they are usually classified by their cation type as imidazolium, ammonium, phosphonium, pyrrolidinium, pyridinium, guanidinium, isoquinolinium, piperidinium, morpholinium, and sulphonium ionic liquids. Usually, 80% and 20% of the main dataset is allocated to the “training” and “test” sets, respectively.

In order to distribute similar compounds into both subsets and overcome the problem of unsuitable allocation of subsets, cluster analysis is undertaken to maintain as close as possible a similarity between “training” and “test” subsets. One of the most important methods of cluster analysis is *K*-means clustering which divides *n* observations into *k* clusters in which each observation is counted within the cluster with the nearest mean [4, 198].

After dividing the entire dataset into “training” and “test” sets, the proper mathematical method should be selected and applied on “training” set to fit the data.

4.6 Feature Selection Methods

In QSPR or GC modeling, one of the earliest steps before starting the use of a regression method is to select the most effective variables on target property which is called “feature selection” or “variable subset selection”. Feature selection method can reduce the computational cost by reducing dimensionality of data, improve the prediction performance and the comprehensibility of the models by eliminating redundant and irrelevant (probable noise) features [199]. For example, in QSPR approach, the researcher should select few variables within thousands of available descriptors and then use a mathematical method to find a model. As using all descriptors is not acceptable and a wise decision mainly due to overfitting and large model size, feature selection approach should be combined with a regression method during the model development.

The highly used feature selection methods in GC or QSPR modeling are Stepwise method and Genetic Algorithm.

4.6.1 Stepwise Method

This method was first proposed by Efroymson [200] which consists of two approaches: “forward selection” and “backward elimination”. Forward selection involves starting with no variables in the model, testing the addition of each variable using a selected objective function, adding the variable (if any) which improves the model the most, and repeating these steps until desired number of variables or none improves the model.

Backward elimination involves starting with all candidate variables, testing the deletion of each variable using the defined objective function, removing the variable that enhances the model the most by being deleted, and redoing these steps until desired number of variables in the model or no further improvement is observed. As usually the number of initial variables is high, the backward elimination approach is required more computations and time comparing with forward selection approach; so the former is more applicable and popular.

4.6.2 Genetic Algorithm

Genetic Algorithm (GA) was introduced by Holland [201] which has “ability to exploit accumulating information about an initially unknown search space in order to bias subsequent search into promising subspaces”. GA is a domain independent search method, so its best application is where the theory and domain knowledge is hard or impossible to present [202].

In a genetic algorithm, a population of individuals (candidate solutions to the optimization problem) is evolved in order to find better solutions. Each individual has a set of properties, called chromosomes, which can be mutated and changed.

The development process usually begins from a population of individuals which are generated randomly. The process is iterative and each iteration is called a generation. In each generation, the fitness of all individuals is evaluated. The more fit individuals are selected randomly from the current population, and each individual's genome is modified to form a new generation by recombination and possibly random mutation. Thereafter, the new generation of possible solutions is used in the next iteration. Usually, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population [203].

Genetic Algorithm shows better results compared with the stepwise methods in terms of speed and performance and is the most popular feature selection method in QSPR modeling.

4.7 Mathematical Methods

Use of GC or QSPR approaches needs a mathematical method as a regression tool to create a meaningful relationship between the calculated structure-based variables and the target property. There are several mathematical methods available in the literature which can be combined with GC or QSPR methods such as Multiple Linear Regression (MLR), Genetic Algorithm based Multiple Linear Regression (GA-MLR), Genetic Function Approximation (GFA), Artificial Neural Networks (ANNs), Support Vector Machine (SVM), Gene Expression Programming (GEP), and so on. In this study, GA-MLR, GFA, SVM, and GEP methods are described.

4.7.1 Genetic Algorithm based Multiple Linear Regression (GA-MLR)

Multiple linear regression is an easiest approach to model the relationship between several variables and a target variable by fitting a linear equation to experimental data. Each value of the independent variable x is related to the value of the dependent (target) variable y . So for p explanatory variables and n observation, the MLR model is like following form.

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \dots + \beta_p x_{pj} \quad j = 1, 2, \dots, n \tag{4.2}$$

In QSPR or GC modeling, the application of a feature selection method is compulsory, so the MLR is combined with Genetic Algorithm (GA-MLR) for feature selection and modeling at the same time. This method is very popular in QSPR research and was successfully applied before [204-207].

As MLR finds a linear correlation between selected variables and the response, it will not produce good results for complex and nonlinear responses; so other methods should be used in such cases.

4.7.2 Genetic Function Approximation (GFA)

Genetic Function Approximation (GFA) is a combination of two different algorithms: Multivariate Adaptive Regression Splines (MARS) of Friedman [208] and Genetic Algorithm (GA) introduced by Holland [201]. It was originally propounded in the innovative work of Rogers and Hopfinger [209]. Generally, the target of most GC or QSPR studies is to introduce the linear combination of basic functions $\varphi_{\alpha}(\mathbf{x})$ of the features $X = \{x_1, \dots, x_n\}$ in the training data set of size M:

$$\hat{y}(\mathbf{x}) = \beta_0 + \sum_{\alpha \in \mathcal{A}} \beta_{\alpha} \varphi_{\alpha}(\mathbf{x}) \quad (4.3)$$

This equation in simple form of linear $\varphi_{\alpha}(\mathbf{x})$ is equal to MLR formula, but in complicated form the functions can be splines, step functions, high-order polynomials, etc. In most cases, binary interactions of variables in the forms of simple multiplication or quadratic polynomials are used. This approach can facilitate modeling of some nonlinear responses as successfully applied on some problems which will be explained in the next chapters. For complicated behaviors, the GFA method is not very efficient and it's required to apply the advanced nonlinear techniques such as ANN or SVM.

4.7.3 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a highly accepted algorithm developed from the machine-learning community. SVM methods have outstanding benefits over Artificial Neural Networks (ANNs) which are [210, 211]:

1. Unlike ANN which has a heuristic development path, SVM has a strong theoretical background which provides a high generalization capability so it can avoid local minima.
2. SVM always has a solution that can be quickly acquired by a standard algorithm (quadratic programming).

3. SVMs are less prone to overfitting or underfitting because fewer parameters are required for its development in comparison with ANNs.
4. The SVM need not determine the network topology and complexity in advance, which can be automatically obtained when the training process ends.
5. SVMs use structural risk minimization whilst ANNs use empirical risk minimization. Thus, SVM is usually less vulnerable to the overfitting problem.

As a result of its advantages, the SVM shows outstanding performance and can be used for both linear and nonlinear regression.

Suykens and Vandewalle [212] applied some modifications to the traditional SVM algorithm to simplify the process of finding a model by solving a set of linear equations (linear programming) instead of nonlinear equations (quadratic programming) and named it as Least Squares Support Vector Machine (LSSVM). As a result, LSSVM includes similar advantages of traditional SVM, but it performs faster computationally. The basic concept of SVM is to transform the signal to a higher dimensional feature space and find the optimal hyper-plane in the space that maximizes the margin between the classes [213]. In LSSVM, the target is to fit a linear relation ($y = wx + b$) between the independent variables (x) and the dependent variable (y). To find the best relation, the cost function (penalized cost function) should be minimized.

4.4

$$Q_{LSSVM} = \frac{1}{2}w^T w + \gamma \sum_{k=1}^N e_k^2$$

where

4.5

$$e_k = y_k - (w^T \phi(x_k) + b) \quad k=1, 2, \dots, N$$

The first part of equation (4.4) is the L_2 norm on regression weights which is penalized quadratically. The second term is the summation of regression error (e_k) for all of the N training objects weighted by parameter γ , which has to be optimized by the user. Equation (4.5) as the

definition of the regression error which is the difference between the true and the predicted values, can be seen as a constraint [211]. In both equations, w represents the regression weight, x is the input vector of parameters of the model, y is the independent variable, and b is the intercept of the linear regression in the LSSVM method.

By using the Lagrange function in equation (4.4), the weight coefficient (w) can be written as an expansion of the Lagrange multipliers with the corresponding training objects (x_k):

$$w = \sum_{k=1}^N \alpha_k x_k, \quad \alpha_k = 2\gamma e_k \tag{4.6}$$

By substituting equation (4.6) into the linear relation ($y = wx + b$) between input and output variable, Lagrange multipliers can be written as:

$$\alpha_k = \frac{(y_k - b)}{x_k^T x + (2\gamma)^{-1}} \tag{4.7}$$

Finding these Lagrange multipliers is very simple as comparing with the SVM approach in which a more difficult relation has to be solved to obtain these values [211].

To have a nonlinear regression function, the linear relation can be extended to nonlinear one by introducing Kernel functions, $K(x, x_k)$.

$$f(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b \tag{4.8}$$

There are few types of nonlinear kernel functions, such as polynomial and radial basis function (RBF) kernels which the latter is commonly used. The RBF kernel function is defined as:

$$K(x, x_k) = \exp\left(\frac{-\|x_k - x\|^2}{\sigma^2}\right) \tag{4.9}$$

where σ^2 is a RBF kernel parameter which should be optimized during model development.

4.7.4 Forward feature selection-based LSSVM

As explained earlier, feature selection methods are required to be applied in QSPR or GC modeling and obviously, SVM needs to use such algorithms to produce models with the minimum variables. In fact, LSSVM algorithm is just a regression method and does not originally have the capability of selecting the variables. To solve this issue, there are two approaches:

- Using GA-MLR or GFA to choose the most effective variables and then use them as the inputs of LSSVM afterwards [214, 215].
- Using Genetic Algorithm or stepwise methods with LSSVM algorithm simultaneously.

For the first approach, obviously there is no guaranty that selected variables by a multi-linear or quadratic equations are good candidates and most efficient variables in SVM modeling. So it's better to select variables directly during SVM modeling.

In this study, “forward feature selection” algorithm was combined with LSSVM successfully (FFS-LSSVM) to find the best predictive SVM model with a minimum number of variables. As the calculations are very time consuming for the large number of variables, the variable reduction approach should be considered to decrease the number of variables.

4.7.5 Genetic Programming (GP)

Genetic Programming (GP) is an evolutionary computation technique that automatically solves problems without requiring the user to know or specify the form or structure of the solution in advance [216-218]. GP is a specialization of genetic algorithm which instead of fixed length binary strings, the solutions are recognized as computer programs [219]. Basing on the rules of natural evolution of genetic, the GP solution is a computer program denoted as tree structure. It is presented in a functional programming language so that it automatically creates computer models, while GA generates a string of numbers to denote the solution [220].

Koza [218] explained that the genetic programming algorithm develops the computer programs to solve problems by executing the following three steps:

1. Generate an initial population of computer programs composed of random compositions of the functions and terminals (standard arithmetic operations, standard programming operations, standard mathematical functions, logical functions, etc.).
2. Iteratively perform the following substeps until the termination criterion has been satisfied:
 - a) Execute each program in the population and assign it a fitness value according to how well it solves the problem.
 - b) Create a new population of computer programs by applying the following two primary operations. The operations are applied to computer program(s) in the population chosen with a probability based on fitness.
 - (i) Copy existing computer programs to the new population (crossover).
 - (ii) Create new computer programs by genetically recombining randomly chosen parts of two existing programs (mutation).
3. The best computer program that appeared in any generation (i.e., the best-so-far individual) is designated as the result of genetic programming. This result may be a solution (or an approximate solution) to the problem.

4.7.6 Gene Expression Programming (GEP)

Gene expression programming (GEP) proposed by Ferreira [221] is one of the most important and robust linear-based GP methods that utilizes a fixed length of character strings to denote solutions to the problems, which are subsequently represented as parse trees of different shapes and sizes. As a result, the main components of GEP mathematical algorithm are control parameters, function set, fitness function, terminal set, and termination condition [220]. Those parse trees are known as expression trees (ETs) for GEP algorithm [222].

The GEP nature permits the evolution of more complex programs composed of several substructures or subprograms called as GEP genes. Each GEP gene encompasses a list of symbols with a fixed length that can be any element from the terminal set such as {x, y, z, -2, 1} and a function set such as {×, /, +, -, sin, log} [222]. For instance, a two-gene chromosome can be made of four functions, Q, *, -, and +, (Q expresses the square root function) and four terminals, a, b, c, and d together with its decoded ET and the corresponding mathematical expression (Figure 4.4). The algebraic expression $\sqrt{(a-b)} \times (c+d)$ can be easily translated as a diagram or ET by the Karva language representation. Each character is placed in a position from zero to seven and can be indicated as follows:

0 1 2 3 4 5 6 7
 Q * - + a b c d

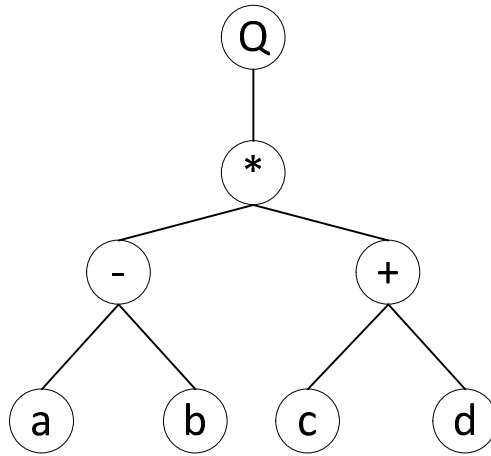


Figure 4.4: A typical expression tree in the gene expression programming, which represents $\sqrt{(a-b)} \times (c+d)$ by a two-gene chromosome.

To reach a termination condition, the GEP mathematical strategy employs some computational steps similar to GP algorithm including reproducing the fixed-length chromosome of each individual randomly for the initial population, assessing fitness of each individual and also representing chromosomes as ETs, choosing the best individuals considering their fitness to generate and repeating the above process until a solution has been found for a certain number of generations [221, 222].

The author has been previously studied the GEP [223-225] and showed that GEP can successfully produce small and accurate correlations for developing the models based on the other physico-chemical properties such as corresponding state models; however the calculations are very time consuming compared with GFA method.

GEP cannot be used directly for developing the GC or QSPR models as it needs to be joint with a feature selection method. GEP has been combined with GA algorithm in GeneXproTools [226] software package to do variable selection and function estimation simultaneously; however it is not efficient enough to obtain a model at a reasonable time regarding this fact that two types of huge and time consuming calculations (variable selection and function estimation) should be done concurrently. Consequently, the best applicability domain of GEP is to develop the models when the number of variables is not too much.

As mentioned in chapter 1, ionic liquids suffer from lack of enough experimental data; therefore, developing the models based on other thermophysical properties of ILs are not applicable and/or straightforward in most cases. Consequently, GEP is not applicable in this thesis as a regression method.

4.8 Fitness function

After selecting the appropriate regression method, it is required to choose a fitness function and minimize it by the regression method. Several types of fitness functions can be defined with respect to the nature of property, data, and desired minimization that the researcher wants to apply.

Common fitness functions and their mathematical representation are shown as follows.

❖ Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

❖ Mean Squared Error (MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^N [y(i) - \hat{y}(i)]^2$$

❖ Root Mean Square Error:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [y(i) - \hat{y}(i)]^2}$$

❖ R-square (Coefficient of determination):

$$R^2 = 1 - \frac{\sum_{i=1}^N [y(i) - \hat{y}(i)]^2}{\sum_{i=1}^N [y(i) - \bar{y}]^2}$$

❖ Average Absolute Relative Deviation (AARD%) which is also known as Mean Absolute Percentage Error (MAPE).:

$$AARD\% = \frac{100}{N} \sum_{i=1}^N \frac{|y(i) - \hat{y}(i)|}{|y(i)|}$$

where N is the number of data points, $y(i)$ is the i th value of target variable (response), and \bar{y} is the average of target variable values.

During the modeling of some properties, it may be required to apply a transformation function on the target property or some variables in order to simplify and speed up the modeling process. The common types of transformation functions are \ln or \log , \exp , and inversion ($1/y$) which can be combined with the fitness function and set it as an objective function to be minimized by the selected regression method.

In this thesis, the AARD% was selected as the fitness function for all properties. In addition, natural logarithm (\ln) was chosen as the transformation function for modeling the viscosity and γ^∞ of ionic liquids.

CHAPTER 5: MODEL DEVELOPMENT

5.1 Scope

This section discusses the steps undertaken in the development of a chemical structure-based model for each properties of ionic liquids which have been studied in this thesis. The theoretical explanation of all steps was discussed in previous chapters.

5.2 Speed of sound in ionic liquids

5.2.1 The GC model

As explained in section 4.3, the chemical structure of cations and anions were drawn by ChemDraw software and saved as separate MDL Molfiles. Thereafter, all the cations or anions files were opened by Dragon Software, simultaneously. Ultimately, the functional groups of both cations and anions were calculated. In the next step, the calculated functional groups of cations and anions were merged with each other to have the functional groups of the ionic liquids. As a result, a dataset comprised of 256 variables (123 for cations, 132 for anions, and T) was built.

During the calculation of functional groups, it is common to have some duplicated variables. For example, the calculations may result in having a variable for number of Chlorine atoms, and another for the number of halogens in the molecule. If there is not any other halogen atom in the molecule, these two variables are equal. As a result, one of them can be removed to reduce the mathematical computations required for choosing the effective variables.

As described in section 4.4, the pair correlation was done and pairs of functional groups with a correlation coefficient equal to 1.0 were removed and remainder kept in the dataset. As a result, the number of variables was reduced to 105 for cations, 68 for anions, and the absolute temperature.

The next step was to divide the dataset into “training” and “test” subset using K -means clustering (section 4.5). As the LSSVM method was selected to model this property, it is common to build three subsets: a “training” set used to develop and train the model, a “validation set” for optimizing the model parameters (internal validation), and a “test” set

which was used to determine the prediction ability of the model for new compounds which have not been used in model development (external validation).

In the group contribution approach, it is assumed that each functional group in the molecule has a contribution to the value of the physical property. This approach is most widely used in the form of linear multivariate models because it is relatively easy to use [191, 227]; but in case of complex and nonlinear relationships between input parameters and desired property, linear modeling fails. In such cases, nonlinear modeling such as Artificial Neural Networks (ANN) and Support Vector Machine (SVM) are used [213, 228].

In primary steps of model development, it was found that GFA model failed to produce the accurate and predictive model. It might be due to complex behavior of molecules and ions or interactions among them. In such cases, functional groups cannot describe the interactions by a linear correlation or completely fail to interpret them. As a result, a nonlinear regression method can overcome this problem.

In recent years, ANNs have been used extensively for modeling in various fields of science; however they may suffer from some disadvantages such as converging at local minima instead of global minima, overfitting if training goes on for too long, and non-reproducibility of results, partly as a result of random initialization of the networks and variation of the stopping criteria during optimization [211, 229]. As explained in section 4.7.3, the SVM overcomes these drawbacks and performs better than the ANN in the development of both linear and nonlinear models.

According to above-mentioned explanations and section 4.7.4, the GFA approach was ignored for choosing the most effective variables. So FFS-LSSVM method was selected to model the speed of sound in ionic liquids using the group contribution approach. Moreover, the AARD% was selected as the objective function.

5.2.2 The QSPR model

According to the method explained in section 4.3, the Dreiding Force field was used to optimize the 3D structure of cations and anions. Thereafter, Dragon software was utilized to

calculate over 3000 molecular descriptors by importing the optimized structures of all cations and anions, separately.

After calculation of descriptors, those which could not be calculated by the software used were completely removed from the list. Then, pair correlation was applied to all descriptors. Accordingly, pair of descriptors with a correlation coefficient greater than 0.9 were removed and the remainder used to develop the model. Consequently, the number of variables was reduced to 230 for cations, 125 for anions, and the absolute temperature.

Thereafter, the “training” and “test” sets were allocated. Finally, the Genetic Function Approximation algorithm was applied on “training” set to select the most effective descriptors based on the AARD% as the objective function.

5.3 Liquid heat capacity of ionic liquids

5.3.1 The GC model

The similar steps were done to calculate the functional groups and eliminate the correlated variables. As a result, the dataset had 56 variables for 3726 data points. Thereafter, the dataset was divided into two subsets: 80% of data points for “training” set and 20% for “test” set.

In typical group contribution models, the summation of functional groups is used to develop a model. The advantage of this method is its simplicity, but its drawback is the weakness in the prediction of properties for compounds which exhibit complex behavior. For such cases, researchers have used more complicated mathematical models like Artificial Neural Networks or Support Vector Machine to find a predictive model using predefined functional groups; but the major problem of these methods is the selection of the most effective functional groups. In order to find an accurate predictive model for certain variables by ANN or SVM, numerous calculations and noticeable amounts of time are required. As a result, it is extremely difficult to determine the most effective variables from the pool available. ANN or SVM models also require specialized mathematical software for their implementation, which further adds to their drawbacks.

Consequently in this part of study, Genetic Function Approximation (GFA) has been used as a tool to select the most effective functional groups and do the regression, simultaneously.

During the initial steps of modeling, it was observed that the relationship between the absolute temperature, functional groups, and the liquid heat capacity of ionic liquids can be formulated as follows.

$$C_{p,liq} = f(\text{functional groups}, T, T^2) \quad 5.1$$

For temperature dependent properties, the above equation can be written in two forms:

$$C_{p,liq} = f(\text{functional groups}) + f(T, T^2) \quad 5.2$$

$$C_{p,liq} = f(\text{functional groups}, T, T^2) \quad 5.3$$

The equation (5.2) is the simplest approach and needs less computations and time, but the final model may have large deviations for some compounds with complex temperature dependence. The second approach results in more accurate models, because it can distinguish between compounds with $f(T)$ or $f(T, T^2)$ as their temperature dependence.

In this study, the absolute temperature was introduced as two variables (T and T^2), and all variables were multiplied by each other; so more than 3700 new variables were generated ($20 \times 20 \times 20$). Ultimately, the AARD% was set as the objective function.

5.3.2 The QSPR model

Dragon software uses the 3D structure of the molecules to calculate more than half of the descriptors. So as a requirement, the Dreiding Force field was utilized to optimize the 3D

chemical structure of ions. Then, the optimized 3D structures were used to calculate the molecular descriptors and thereafter, pair correlation was done to omit the interrelated descriptors with a correlation coefficient greater than 0.9. As a result, 644 variables remained and used to develop the model.

Normally in QSPR modeling, it is assumed that there is a linear relationship between the target property and one or more descriptors. Therefore, the output is a linear function which can be interpreted and understood easily; but this assumption shows weakness in the case of complicated relationships between the target property and the chemical structures of some materials.

Same as the previous section, Genetic Function Approximation was performed to select the most effective descriptors, combine two descriptors by multiplication to create new variables, use them in modeling, and minimizing the AARD% as the objective function.

5.4 Refractive index of ionic liquids

5.4.1 The GC model

After calculating the functional groups by Dragon software, the duplicated variables were eliminated and consequently, 279 variables were remained. Then *K*-mean clustering was used to distribute similar structures in both “training” and “test” sets. As a result, 20% of the data points (172 data points for 15 ILs) were used in the test set and the remainder (759 data points for 82 ILs) used in the training set to develop the model.

By analyzing the data of refractive index, it was observed that n_D has a linear dependency with regard to temperature. Thereafter, it was assumed that the linear summation of functional groups can fit the data of refractive index. So, the GFA method was used and all of the functional groups were multiplied by the absolute temperature and new variables were produced. Finally, the AARD% objective function was minimized.

5.4.2 The QSPR model

Same as the previous sections, the descriptors of cations and anions were calculated and the highly correlated descriptors were removed from the dataset. Thereafter, new variables were produced by multiplication of the absolute temperature and all of the descriptors.

After allocating the “training” and “test” sets, the GFA method was performed to fit the data and minimize the AARD% objective function.

5.5 Viscosity of F-ILs

5.5.1 The GC model

The substructures and their number of events/occurrences in chemical structures of cations and anions plus the absolute temperature were applied as the input parameters of the GC model.

After eliminating the duplicated data by pair correlation, *K*-means clustering was used to divide the main dataset. In this study, approximately 80% (667 data points) and 20% (196 points) of the main dataset were allocated to the “training” and “test” sets, respectively.

Generally, it is assumed that the viscosity can be modeled using the more complex form Arrhenius equation, $\eta = \exp\left(\frac{A}{T} + B\right)$. As a result of greater reliability being required for smaller values of viscosity, taking the logarithm will increase the accuracy for small values of viscosity [89]. In this study, it was assumed that $\ln(\eta)$ had a multi-linear relationship with functional groups, and a nonlinear relationship with absolute temperature. The temperature dependency was assumed as $\ln\left(\frac{\eta}{T}\right)$. Consequently, the “training” set was used to select the effective substructures and fine tuning the temperature nonlinear function in a polynomial form.

It is notable that ILs are glass former and typically one has to use the VFT approach (equation) and with it the Vogel temperature T_0 , which is unfortunately an IL specific quantity.

5.4

$$\ln\left(\frac{\eta}{T}\right) = \ln\left(\frac{\eta_0}{T_0}\right) + \frac{B}{T - T_0}$$

As the η_0 was not available for most of the ionic liquids, unfortunately the VFT approach was not applicable for modeling the viscosity of ionic liquids. As a result, the Arrhenius equation was used in this thesis.

5.5.2 The QSPR model

The molecular descriptors of ionic liquids were calculated using their optimized 3D structure and then, the pair correlation was performed to remove highly correlated descriptors. Thereafter, the main dataset was divided into “training” and “test” sets by *K*-means clustering technique. As a result, 667 data points and 196 data points were allocated to the “training” and “test” sets, respectively.

As mentioned in previous section, it was assumed that $\ln(\eta) = \ln\left(\frac{\eta_0}{T}\right)$ so the “training” set was used to develop and train the model.

5.6 γ^∞ of organic solutes in ionic liquids

5.6.1 Aromatic solutes

From property estimation perspective, the data of γ^∞ consists of three sets of variables which are related to the chemical structure of cations, anions, and solutes.

Similar to previously discussed GC modeling, the functional groups of the cations, anions, and solutes calculated without 3D optimization of the structures. Thereafter, the calculated groups of each component were processed separately to remove any possible unnecessary variables. The result was 497 functional groups of which 137 belonged to cations, 201 to anions and 159 to solutes.

In the next step, the data partitioning was performed and 1278 and 375 data points were chosen as the “training” and “test” sets, respectively. Moreover, the data were analyzed precisely to find the dependency of γ^∞ to the temperature. It was observed that for the majority of solute-IL system, the $\ln(\gamma^\infty)$ was not linear against temperature; but transforming the γ^∞ to $\ln(\ln(\gamma^\infty))$ resulted in a linear curve for most cases; however it was observed that the aforementioned transformation was remained nonlinear for some systems. In addition,

$\ln(\ln(\gamma^\infty))$ was a function of T for some systems, while it was depended to $\frac{1}{T}$ for the others. As a result, the equation (5.5) was used to fit the data; so four forms of the absolute temperature were utilized to generate more than 1000 new variables. Ultimately, the AARD% objective function was minimized.

5.5

$$\ln(\ln(\gamma^\infty))=f(GCs, T, T^2, \frac{1}{T}, \frac{1}{T^2})$$

It should be mentioned that for modeling the γ^∞ of ionic liquids, only the GC approach was performed, because the primary QSPR modeling of this property revealed that an accurate model needed more than 30 descriptors. As large QSPR models are not well accepted, the QSPR modeling of the γ^∞ of ionic liquids was ignored.

5.6.2 Alcohol solutes

The similar procedure was performed to generate new variables. Thereafter, K -means clustering was done and 2048 and 738 data points were allocated in the “training” and “test” sets, respectively. Finally, the equation (5.5) was used to develop the model and minimize the AARD% of calculated data.

5.6.3 Alkane solutes

The functional groups were calculated for cations, anions, and solutes, separately. The initial steps of model development were revealed that all available experimental data cannot be modeled by a single correlation. It was observed that for solutes with lower number of carbon atoms, the γ^∞ was usually less than 100; but as the number of carbon atoms increased, the γ^∞ increased drastically up to 4000. As a result, the main database was divided into two subsets:

- Solute with less than 10 carbon atoms.
- Solute with 10 or more carbon atoms.

The first dataset contained 743 systems with 3368 experimental data points and the second dataset had 140 systems with 567 data points. For both datasets, the “training” and “test” sets were allocated separately by means of K -means clustering. Finally, Genetic Function

Approximation (GFA) was used as a tool to select the most effective functional groups and do the regression, simultaneously. In addition, the equation (5.5) was used to develop the model and minimize the AARD% of calculated data.

5.6.4 Alkene solutes

For alkene solutes, no strange behavior was observed. As a result, the functional groups of solutes, cations, and anions were calculated and thereafter, *K*-means clustering was performed to allocate the “training” and “test” sets. So 1536 experimental data points were selected as the “training” set and 475 data points as the “test” set. Ultimately, GFA method has been used to do the regression and select the effective functional groups, simultaneously. Similar to previous sections, the equation (5.5) was used for development of the model and the AARD% was set as the objective function.

5.6.5 Alkyne solutes

Similar procedures were done for alkyne solutes to screen and refine the database. Thereafter, 1257 experimental data points were divided into two subsets by *K*-means clustering; so 945 data points were considered as the “training set” and 312 data points as the “test” set. Finally, the equation (5.5) was used to develop the model by means of GFA method.

5.7 Critical temperature of ionic liquids

To develop the model, both the group contribution and QSPR approaches were followed up. In GC approach, it was observed that the model failed to predict the data and consequently, only the QSPR modeling were performed; however the results of both methods were explained in section 6.7.

Similar to previous sections, the “training” and “test” sets were allocated by dividing the entire dataset by means of *K*-means clustering.

CHAPTER 6: RESULTS

6.1 Scope

In previous chapters, it was explained how to manage the data and how to do the computational procedures. In this chapter, the models developed and the outputs are discussed.

6.2 Speed of sound in ionic liquids

6.2.1 The GC model*

The forward feature selection method was combined with the LSSVM algorithm to develop a reliable predictive SVM model. The result was an 8-variable model with the absolute temperature as the first variable and chemical substructures, as shown in Table 6.1, as the others. The two parameters of the SVM model are as follows:

$$\gamma = 104.183 \text{ (the weight of the regression error)}$$

$$\sigma^2 = 14.755 \text{ (the parameter of the RBF kernel)}$$

As listed in Table 6.2, the AARD% of the proposed model is 0.21%, 0.68%, and 0.87% for the training, validation, and test sets, respectively. These values indicate that the proposed SVM model can correlate and predict the speed of sound (u) fairly well. The values of the predicted u versus the experimental data are presented in Figure 6.1. In addition, the deviation of the model from the experimental data and the percentage of data points in different AARD ranges are shown in Figure 6.2 and Figure 6.3.

According to Figure 6.3, greater than 99% of calculated data points for the training set, 81% for the validation set, and 64% for the test set are within an AARD% of 0-1%. The percentage of data points with an AARD% of 1-2% is less than 1%, 16%, and 26% for the training, validation, and test sets, respectively. The percentage of data points with an AARD% of 2-3%, is 3% and 10% for the validation and test sets respectively. For the training set, the relative deviation of all calculated data points is less than 2%.

* The results have been published in *Fluid Phase Equilibria* 367 (2014) 188–193.

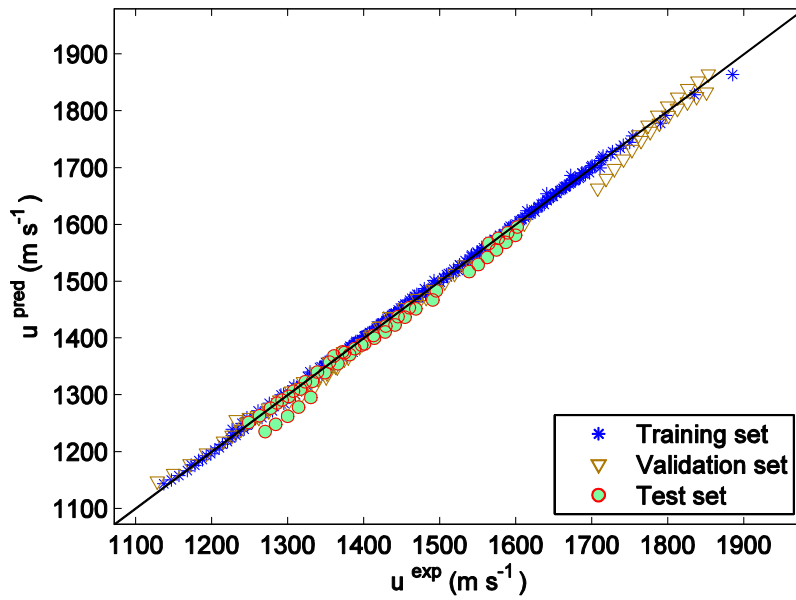


Figure 6.1: Predicted versus experimental values of speed of sound in ILs.

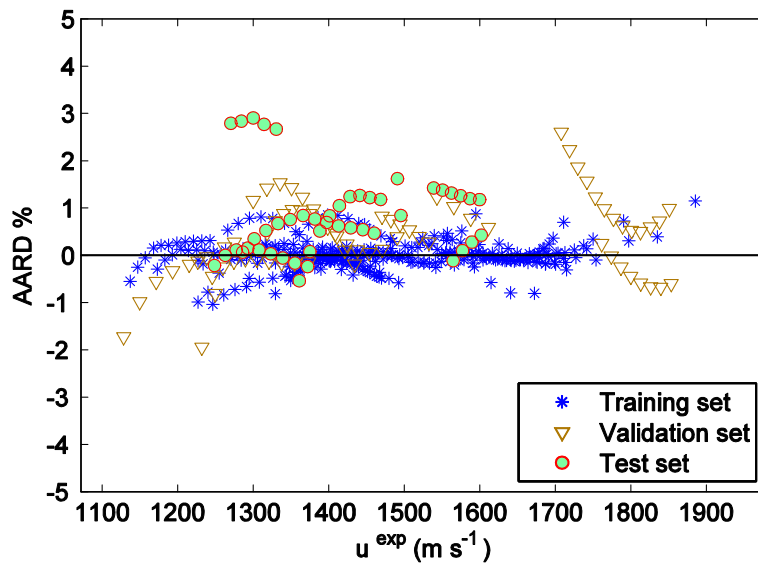


Figure 6.2: Relative deviation of predicted speed of sound from experimental data.

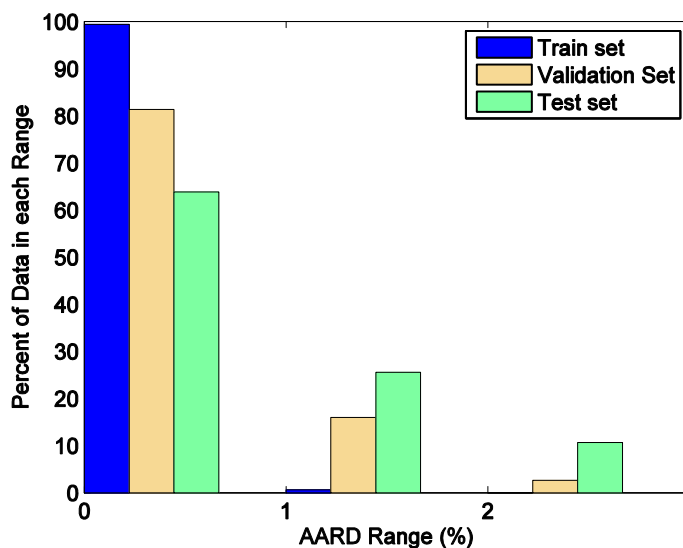


Figure 6.3: Percentage of predicted values of speed of sound in different relative deviation ranges.

Table 6.1: The input variables of LS-SVM Model for speed of sounds in ILs.

No.	Symbol	Description
1	V_T	T/K - Absolute Temperature
2	V_{C1}	presence/absence of $C-(A^*)_3-N$ in cation (value: 1 or 0)
3	V_{C2}	number of $C-(A)-N$ in cation
4	V_{C3}	number of $C-(A)_2-N$ in cation
5	V_{C4}	number of $C-(A)_6-N$ in cation
6	V_{A1}	$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \sum_{m=1}^n \sum_{n=1}^n$ - Number of atoms of anion
7	V_{A2}	Number of F-B in anion
8	V_{A3}	Number of $O-(A)_2-F$ in anion

* A represents any type of atoms

Table 6.2: Statistical parameter of the model of speed of sound

<i>Statistical Parameter</i>	
<i>training set</i>	
R^2	0.999
Absolute average relative deviation	0.21
Standard deviation error	4.34
Root mean square error	4.34
No. of data points	324
<i>validation set</i>	
R^2	0.997
Absolute average relative deviation	0.68
Standard deviation error	11.98
Root mean square error	13.39
No. of data points	75
<i>test set</i>	
R^2	0.988
Absolute average relative deviation	0.87
Standard deviation error	11.54
Root mean square error	16.29
No. of data points	47
<i>total</i>	
R^2	0.997
Absolute average relative deviation	0.36
Standard deviation error	8.18
Root mean square error	8.47
No. of data points	446

As indicated in Table 6.3, the highest average error is observed for [GluC4][DS] (L-glutamic acid, 1,5-bis(2-methylpropyl) ester, dodecyl sulfate) which has a value of 2.79%. This IL was in the test set. In most engineering applications, an error of this magnitude would be acceptable and shows that the proposed model has good prediction capability. However, this IL is the second largest molecule in our database with 86 atoms. As it is a big molecule, it seems that some interactions exist within this molecule and the selected functional groups of the model cannot account for the probable interactions.

The IL with the second highest deviation is [C8MIm][NTf2] (1-octyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]imide) with an AARD% of 1.95%. As this IL has just one reported experimental data point and has been used to tune parameters of the model within the validation set, it has not been able to tune the model parameter for its own benefit due to its weighted contribution.

The model can correlate or predict the remainder ILs with very good accuracy. This claim is justified by considering the [C6MIm] (1-hexyl-3-methylimidazolium) ionic liquids. As indicated in Table 6.3, three [C6MIm] ionic liquids with different anions were used in the validation set. The cation did not exist in the training set and it was not used in the process of variable selection by applied forward feature selection routine. It was explained earlier that in SVM, the training set has the major influence in the process of model development and the validation set is used simply to optimize the model parameters. Thus, if the model is able to predict the test set well, it shows that the model has prediction ability. As a result, the presented SVM model shows the good prediction ability for [C6MIm] ionic liquids for which this cation did not exist in the training and validation sets.

The claim about the prediction ability of the SVM model proposed in this study can also be shown for [bmpy][BF4] (1-butyl-3-methylpyridinium tetrafluoroborate). There was just one IL available in whole dataset with 1-butyl-3-methylpyridinium cation which was in the test set. As a result its structure was not used in either the training set or validation set; however a similar cation (N-octyl-3-methylpyridinium) was used in the model development process. The speed of sound in [bmpy][BF4] was predicted with an AARD% of 1.29% which is fairly low deviation and further indicates the ability of the model to predict values for ILs for which their cations or anions did not exist in the training dataset.

In section 2.2, it was discussed that the available correlations need the experimental data of density and surface tension for calculation of speed of sounds. To compare those models with the SVM model developed, only one experimental data point was used; because for most of the ionic liquids, there was only one reported value for density at certain temperature, mostly at 298.15 K. In addition, there were not any reported values for surface tension of pyrrolidinium, pyridinium, and amino acid ionic liquids. So a model developed by Gharagheizi *et al.* [230] was used to predict the surface tension at desired temperature.

According to Table 6.4, the SVM model had a very low AARD% comparing with other models. It can be found that the model of Gardas and Coutinho [48] had better prediction compared with work of Singh and Singh [52]. The model of Gardas and Coutinho showed that for pyrrolidinium and pyridinium ILs, the predicted values of surface tension were reliable and consequently, the low AARD values were observed. But for amino acid ionic liquids, the AARD values were high because of the error in prediction of surface tension. As a result, the presented model in this study is more applicable as it is not depend on any other physical properties which need to be measured by experiments.

Furthermore, there is no need to calculate the functional groups to estimate the speed of sound in new ionic liquids which their cations and anions are available in this study. It is just required to insert the corresponding groups of cations and anions in the presented model. As a result, the speed of sound in 319 ionic liquids (the number of cations multiplied by number of anions) can be estimated using functional groups calculated previously.

Table 6.3: IL abbreviations and AARD% of ionic liquids modelled by LSSVM.

No.	Compound	T (K) range	u (ms ⁻¹)	Uncertainty (ms ⁻¹)	AARD%	Number of data points	Subset
1	[C2MIm][TfO]	278.15-338.15	1348.51-1482.23	2.08	0.08	10	Train
2	[C2MIm][NTf2]	293.15-293.15	1240	10.00	0.78	1	Train
3	[C2MIm][EtSO4]	288.15-343.15	1566.4-1703.9	7.25	0.06	12	Train
4	[C8MIm][NTf2]	293.15-293.15	1232	11.00	1.95	1	Validation
5	[C8MIm][BF4]	293.15-343.15	1361.1-1495.6	2.09	0.94	11	Test
6	[C8MIm][PF6]	278.15-343.15	1294.6-1481.6	5.00	0.18	14	Train
7	[C5Mim][NTf2]	298.15-298.15	1227	7.00	0.11	1	Train
8	[C4MIm][PF6]	278.15-343.15	1329.4-1492.5	2.31	0.27	14	Train
9	[C4MIm][TfO]	293.15-318.15	1348.1-1403.4	1.30	0.07	6	Train
10	[C4MIm][NTf2]	293.15-293.15	1227	13.00	0.99	1	Train
11	[C4MIm][MeSO4]	278.15-343.15	1552.1-1711.3	2.00	0.21	14	Train
12	[C4MIm][OcSO4]	278.15-343.15	1349.6-1557.2	4.43	0.30	50	Train
13	[C4MIm][BF4]	283.15-343.15	1462.1-1604.5	2.05	0.19	13	Train
14	[C6MIm][BF4]	293.15-318.15	1470.5-1532.7	1.40	0.57	6	Validation
15	[C6MIm][PF6]	278.15-343.15	1318.4-1490.7	5.00	0.54	14	Validation
16	[C6MIm][NTf2]	283.15-343.15	1128.4-1262	1.70	0.52	8	Validation
17	[C1MIm][MeSO4]	283.15-343.15	1708-1851	5.00	1.17	13	Validation
18	[C3MIm][NTf2]	293.15-343.15	1137-1243	2.00	0.26	11	Train
19	[OMIm][Cl]	278.15-343.15	1510.2-1885.4	2.00	0.43	14	Train
20	[C4EPyr][EtSO4]	328.15-343.15	1564.9-1602	5.10	0.23	4	Test
21	[C2MPyr][EtSO4]	308.15-343.15	1665.7-1750.2	5.10	0.10	8	Train
22	[C4MPyr][NTf2]	278.15-343.15	1173-1316	3.00	0.09	14	Train
23	[C4MPyr][MeSO4]	298.15-343.15	1625.5-1741.6	5.10	0.13	10	Train
24	[C2Py][EtSO4]	298.15-343.15	1608-1711	3.00	0.10	10	Train
25	[C8MPyr][BF4]	278.15-328.15	1433.6-1571.9	3.56	0.08	21	Train
26	[bmpy][BF4]	293.15-318.15	1538.9-1599.8	1.40	1.29	6	Test
27	[C3Py][BF4]	278.15-338.15	1549.54-1691.9	2.92	0.07	25	Train
28	[M0-py][BF4]	293.15-323.15	1543.45-1611.47	1.58	0.91	4	Validation
29	[P14,6,6,6][dca]	278.15-343.15	1390-1599	3.00	0.12	14	Train
30	[TEMA][MeSO4]	308.15-343.15	1761.5-1853.5	5.10	0.44	8	Validation
31	[NHHH,(CH2)2OH][ac]	298.15-298.15	1790.73	3.17	0.73	1	Train
32	[GluC3][DS]	288.15-343.15	1233.6-1439.7	5.02	0.17	12	Train
33	[GluC4][DS]	323.15-343.15	1270.5-1330.6	5.00	2.79	5	Test
34	[GlyC3][DS]	303.15-343.15	1245.8-1373.4	5.00	0.18	9	Validation

No.	Compound	T (K) range	u (ms ⁻¹)	Uncertainty (ms ⁻¹)	AARD%	Number of data points	Subset
35	[GlyC4][DS]	303.15-343.15	1248.7-1372.1	5.00	0.12	9	Test
36	[ValC3][DS]	288.15-343.15	1254.9-1438.4	5.02	0.13	12	Train
37	[ValC4][DS]	288.15-343.15	1249.5-1434.4	0.01	0.86	12	Validation
38	[ProC3][DS]	288.15-343.15	1284-1455.5	5.03	0.10	12	Train
39	[AlaC3][DS]	288.15-343.15	1266.3-1432.3	5.01	0.50	12	Train
40	[ProC4][DS]	288.15-343.15	1286.2-1460.2	5.03	0.57	12	Test
41	[AlaC4][DS]	288.15-343.15	1246.6-1447.6	5.02	0.52	12	Train

Table 6.4: Comparison between calculated values of speed of sound in ILs using different models

No.	Abbr	T	This study			Gardas & Coutinho [48]		Singh and Singh [52]	
			u_{exp} (ms ⁻¹)	u_{calc} (ms ⁻¹)	ARD%	u_{calc} (ms ⁻¹)	ARD%	u_{calc} (ms ⁻¹)	ARD%
1	[C2MIm][TfO]	298.15	1435.6	1458.46	1.59	1361.09	5.19	1171.09	18.43
2	[C2MIm][NTf2]	293.15	1240	1249.67	0.78	1193.36	3.76	828.72	33.17
3	[C2MIm][EtSO4]	298.15	1679	1679.86	0.05	1599.37	4.74	1789.85	6.60
4	[C8MIm][NTf2]	293.15	1232	1255.98	1.95	1210.29	1.76	859.99	30.20
5	[C8MIm][BF4]	298.15	1491	1466.86	1.62	1361.98	8.65	1173.09	21.32
6	[C8MIm][PF6]	298.15	1407.8	1409.56	0.13	1310.06	6.94	1059.13	24.77
7	[C5Mim][NTf2]	298.15	1227	1228.35	0.11	1181.18	3.73	806.67	34.26
8	[C4MIm][PF6]	298.15	1442.2	1458.14	1.11	1441.89	0.02	1362.84	5.50
9	[C4MIm][TfO]	298.15	1392.1	1393.18	0.08	1299.87	6.62	1037.62	25.46
10	[C4MIm][NTf2]	293.15	1227	1239.14	0.99	1169.27	4.71	785.45	35.99
11	[C4MIm][MeSO4]	298.15	1658	1659.62	0.10	1541.22	7.04	1623.76	2.07
12	[C4MIm][OcSO4]	298.15	1484.7	1480.43	0.29	1236.19	16.74	909.23	38.76
13	[C4MIm][BF4]	298.15	1576.1	1593.69	1.12	1581.15	0.32	1736.71	10.19
14	[C6MIm][BF4]	298.15	1519.5	1513.47	0.40	1536.07	1.09	1609.52	5.92
15	[C6MIm][PF6]	298.15	1424.2	1420.79	0.24	1371.92	3.67	1195.75	16.04
16	[C6MIm][NTf2]	298.15	1226.8	1228.35	0.13	1173.44	4.35	792.84	35.37
17	[C1MIm][MeSO4]	298.15	1813	1803.98	0.50	1887.14	4.09	2765.37	52.53
18	[C3MIm][NTf2]	298.15	1232	1228.26	0.30	1187.00	3.65	817.16	33.67
19	[OMIm][Cl]	298.15	1715	1720.28	0.31	1428.66	16.70	1330.20	22.44
20	[C4EPyr][EtSO4]*	328.15	1602	1595.18	0.43	1760.51	9.89	2303.74	43.80
21	[C2MPyr][EtSO4]*	308.15	1750.2	1744.24	0.34	1849.91	5.70	2624.24	49.94

No.	Abbr	T	This study			Gardas & Coutinho [48]		Singh and Singh [52]		
			u_{exp} (ms ⁻¹)	u_{calc} (ms ⁻¹)	ARD%	u_{calc} (ms ⁻¹)	ARD%	u_{calc} (ms ⁻¹)	ARD%	
22	[C4MPyr][NTf2]	298.15	1269	1269.15	0.01	1224.24	3.53	886.29	30.16	
23	[C4MPyr][MeSO4]*	298.15	1741.6	1739.29	0.13	1762.56	1.20	2310.80	32.68	
24	[C2Py][EtSO4]*	298.15	1711	1712.08	0.06	1629.49	4.76	1879.85	9.87	
25	[C8MPyr][BF4]	298.15	1511.4	1513.47	0.14	1473.31	2.52	1442.33	4.57	
26	[bmpy][BF4]*	298.15	1587	1567.98	1.20	1567.96	1.20	1698.88	7.05	
27	[C3Py][BF4]	298.15	1641.7	1643.51	0.11	1670.81	1.77	2007.79	22.30	
28	[M0-py][BF4]*	293.15	1611.47	1602.03	0.59	1582.09	1.82	1739.45	7.94	
29	[P14,6,6,6][dca]	298.15	1526	1528.62	0.17	1624.33	6.44	1864.23	22.16	
30	[TEMA][MeSO4]*	318.15	1826.3	1838.54	0.67	1860.92	1.90	2665.51	45.95	
31	[NHHH,(CH2)2OH][ac]*	298.15	1790.73	1777.72	0.73	2021.11	12.86	3311.87	84.95	
32	[GluC3][DS]*	298.15	1389	1391.43	0.17	2520.36	81.45	5917.89	326.05	
33	[GluC4][DS]*	323.15	1330.6	1295.1	2.67	2531.09	90.22	5984.41	349.75	
34	[GlyC3][DS]*	303.15	1373.4	1375.36	0.14	2657.60	93.51	6803.24	395.36	
35	[GlyC4][DS]*	303.15	1372.1	1375.36	0.24	2669.31	94.54	6882.38	401.59	
36	[ValC3][DS]*	298.15	1400.5	1402.15	0.12	2668.72	90.55	6878.37	391.14	
37	[ValC4][DS]*	298.15	1408.6	1402.15	0.46	2689.24	90.92	7018.29	398.25	
38	[ProC3][DS]*	298.15	1419.2	1420.92	0.12	2690.00	89.54	7023.53	394.89	
39	[AlaC3][DS]*	298.15	1401.4	1403.19	0.13	2635.79	88.08	6657.43	375.06	
40	[ProC4][DS]*	298.15	1429.2	1420.92	0.58	2700.78	88.97	7097.74	396.62	
41	[AlaC4][DS]*	298.15	1401.9	1403.19	0.09	2636.98	88.10	6665.35	375.45	
<i>AARD%</i>					<i>0.51</i>			<i>25.69</i>	<i>112.64</i>	

* Surface tension was predicted by the model published by Gharagheizi *et al.* [230]

6.2.2 The QSPR model*

To get the most accurate but simplest predictive model, various models with different numbers of descriptors were examined and the best model in term of R^2 was selected. The final model consisted of absolute temperature, and descriptors cation and anion. To simplify the model, the effect of the cation and anion descriptors is shown as u_{Cation} and u_{Anion} .

6.1

$$T = 3000.68059 + u_{\text{Cation}} + u_{\text{Anion}} - 2.7271 T$$

$$u_{\text{Cation}} = 98.66368 \times nC - 88.27463 \times nF - 638.70715 \times \text{Mor}03u - 49.4333 \times X2A - \text{JGI2}$$

$$u_{\text{Anion}} = -30.37573 \times nC - 47.8716 \times nF - 104.91867 \times \text{Mor}03u - 510.95776 \times X2A$$

In the above equation, T is the absolute temperature and nC and nF are the number of Carbon and Fluorine atoms, respectively. Also $\text{Mor}03u$ is an atom centered fragment descriptor which is the number of Carbon atoms on an aromatic ring that has three carbon neighbors on the same aromatic ring. *Mor03u* is signal 3 / unweighted 3D MoRSE descriptors (3D Molecule Representation of Structures based on Electron diffraction) which are derived from Infrared spectra simulation using a generalized scattering function [231].

X2A is average connectivity index chi-2 which belongs to Kier-Hall Connectivity Indices. A molecular connectivity index is calculated by drawing out a chemical in a hydrogen-suppressed molecular structure and calculating the number of adjacent non-hydrogen atoms for each atom. This descriptor reflects the relative accessibility of each bond to encounter other bonds of the same molecule [232]. *ATS5m* is the Broto-Moreau autocorrelation of a topological structure - lag 5 weighted by atomic masses [233]. *MATS3p* is the Moran autocorrelation - lag 3 / weighted by atomic polarizabilities [234]. *JGI2* is the mean topological charge index of order 2. The Topological Charge Indices evaluate the charge transfer between pairs of atoms and hence the global charge transfer in the molecule [235].

According to equation (6.1) and above-mentioned descriptions, speed of sound is mostly affected by *JGI2* descriptor in anions and *X2A* in cations. This means that the charge

* The results have been published in *Journal of Molecular Liquids* 196 (2014) 7–13

transfers between atoms in anions, and the effects of bonds on each other in cations, are the most important factor on speed of sound in ILs.

The predicted values of u in comparison with the experimental values are presented in Figure 6.4. In addition, the deviation of the model in comparison with the experimental data is shown as Figure 6.5. These figures indicate that the majority of calculated/predicted data points have small deviations from the experimental values. The summary of the statistical parameters of the model, training and test sets can be found in Table 6.6.

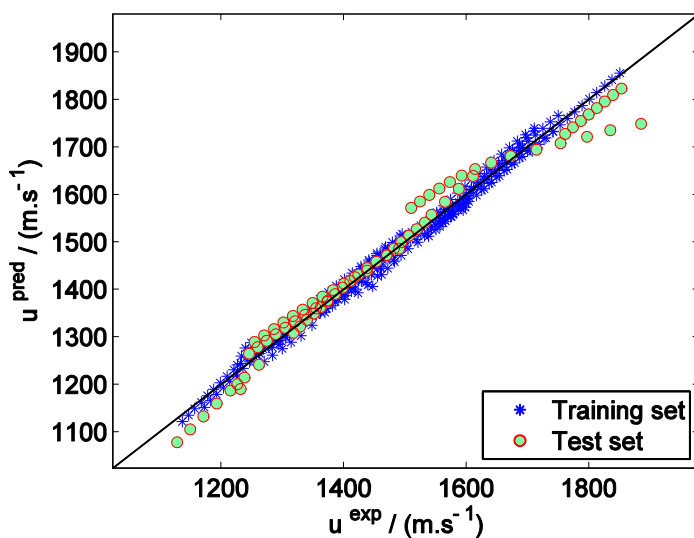


Figure 6.4: Predicted versus experimental values of speed of sound in ILs.

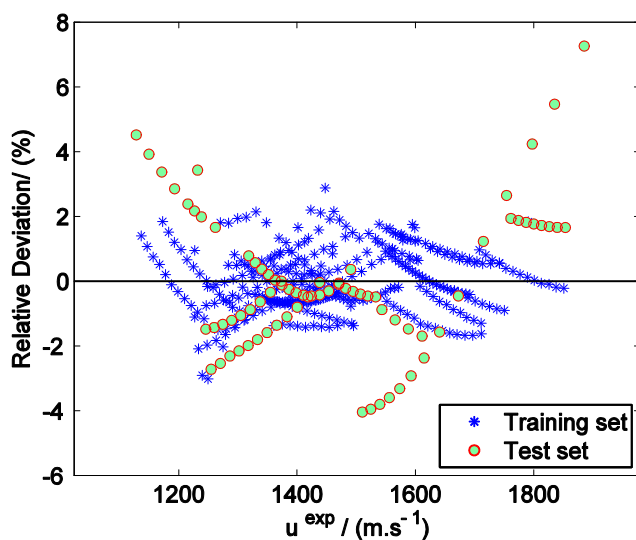


Figure 6.5: Relative deviation of predicted speed of sound from experimental data.

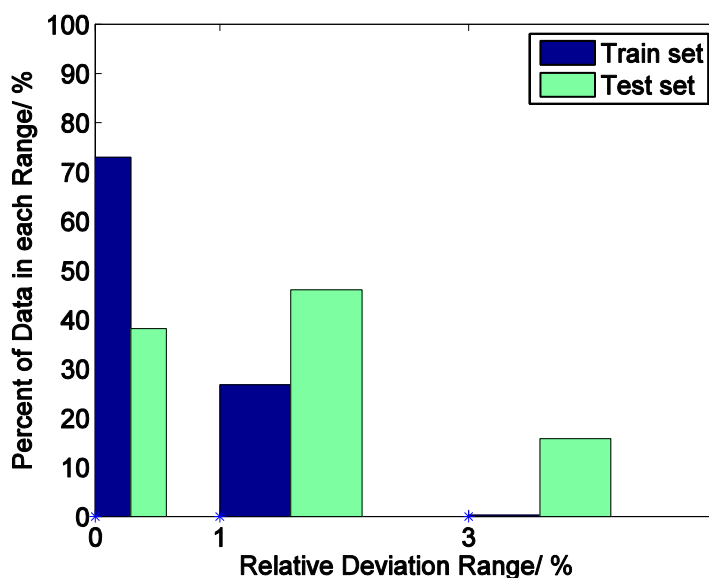


Figure 6.6: Percentage of the predicted values of speed of sound in different relative deviation ranges

Table 6.5: List of ionic liquids and their frequency used to develop equation (6.1)

No.	Compound	T range/K	SS range/m s ⁻¹	AARD/%	Subset	Number of data points
1	[TEMA][MeSO ₄]	308.15-343.15	1761.5-1853.5	1.77	Test	8
2	[NHHH ₂ (CH ₂) ₂ OH][ac]	571.3-571.3	1790.73-1790.73	1.84	Training	1
3	[OMIm][Cl]	278.15-343.15	1510.2-1885.4	3.35	Test	14
4	[C1MIm][MeSO ₄]	283.15-343.15	1708-1851	0.36	Training	13
5	[C2MIm][TfO]	278.15-338.15	1348.51-1482.23	0.63	Training	10
6	[C4MIm][PF ₆]	278.15-343.15	1329.4-1492.5	0.88	Training	14
7	[C4MIm][BF ₄]	283.15-343.15	1462.1-1604.5	1.39	Training	13
8	[C4MIm][TfO]	293.15-318.15	1348.1-1403.4	0.29	Training	6
9	[C2MIm][NTf ₂]	293.15-293.15	1240-1240	2.91	Training	1
10	[C4MIm][NTf ₂]	293.15-293.15	1227-1227	1.11	Training	1
11	[C8MIm][NTf ₂]	293.15-293.15	1232-1232	3.43	Test	1
12	[M0-py][BF ₄]	293.15-323.15	1543.45-1611.47	1.31	Test	4
13	[C3MIm][NTf ₂]	293.15-343.15	1137-1243	0.71	Training	11
14	[C4MPyr][NTf ₂]	278.15-343.15	1173-1316	0.77	Training	14
15	[C3Py][BF ₄]	278.15-338.15	1549.54-1691.9	0.36	Training	25
16	[C6MIm][BF ₄]	293.15-318.15	1470.5-1532.7	0.33	Test	6
17	[C8MIm][BF ₄]	293.15-343.15	1361.1-1495.6	1.32	Training	11
18	[C5Mim][NTf ₂]	298.15-298.15	1227-1227	0.95	Training	1
19	[C6MIm][PF ₆]	278.15-343.15	1318.4-1490.7	0.34	Test	14
20	[C8MIm][PF ₆]	278.15-343.15	1294.6-1481.6	0.40	Training	14
21	[C2MIm][EtSO ₄]	288.15-343.15	1566.4-1703.9	1.04	Training	12
22	[C6MIm][NTf ₂]	283.15-343.15	1128.4-1262	2.86	Test	8
23	[C4MIm][MeSO ₄]	278.15-343.15	1552.1-1711.3	1.31	Training	14

No.	Compound	T range/K	SS range/m s ⁻¹	AARD/%	Subset	Number of data points
24	[C4MIm][OcSO4]	278.15-343.15	1349.6-1557.2	0.57	Training	50
25	[bmpy][BF4]	293.15-318.15	1538.9-1599.8	1.47	Training	6
26	[P14,6,6,6][dca]	278.15-343.15	1390-1599	0.52	Training	14
27	[C8MPyr][BF4]	278.15-328.15	1433.6-1571.9	0.44	Training	21
28	[C4MPyr][MeSO4]	298.15-343.15	1625.5-1741.6	0.70	Training	10
29	[C2MPyr][EtSO4]	308.15-343.15	1665.7-1750.2	0.63	Training	8
30	[C2Py][EtSO4]	298.15-343.15	1608-1711	0.78	Training	10
31	[C4EPyr][EtSO4]	328.15-343.15	1564.9-1602	0.81	Training	4
32	[AlaC3][DS]	288.15-343.15	1266.3-1432.3	0.88	Training	12
33	[GlyC3][DS]	303.15-343.15	1245.8-1373.4	0.93	Test	9
34	[GlyC4][DS]	303.15-343.15	1248.7-1372.1	0.32	Training	9
35	[AlaC4][DS]	288.15-343.15	1246.6-1447.6	1.00	Training	12
36	[GluC3][DS]	288.15-343.15	1233.6-1439.7	1.28	Training	12
37	[GluC4][DS]	323.15-343.15	1270.5-1330.6	1.93	Training	5
38	[ValC3][DS]	288.15-343.15	1254.9-1438.4	1.57	Test	12
39	[ValC4][DS]	288.15-343.15	1249.5-1434.4	0.73	Training	12
40	[ProC3][DS]	288.15-343.15	1284-1455.5	0.93	Training	12
41	[ProC4][DS]	288.15-343.15	1286.2-1460.2	0.54	Training	12

Table 6.6: Statistical parameters for equation (6.1)

<i>Statistical Parameter</i>	
<i>training set</i>	
R ²	0.9919
Average absolute relative deviation	0.76
Standard deviation error	13.53
Root mean square error	13.54
No. of data points	370
<i>test set</i>	
R ²	0.9701
Average absolute relative deviation	1.66
Standard deviation error	34.10
Root mean square error	34.12
No. of data points	76
<i>total</i>	
R ²	0.9862
Average absolute relative deviation	0.92
Standard deviation error	18.72
Root mean square error	18.72
No. of data points	446

According to Table 6.4 and Table 6.5, the overall AARD% of the “training” set is less than 0.8 % which shows that the model fits the data very well. The calculated values of u in “training” set indicates that about 73 % data points have an AARD% of between 0 and 1 %, about 27 % between 1.01 and 3.00 %, and only one data point with deviation over 3 %; i.e. 3.01%. The summary of these results are presented in Figure 6.6 for both “training” set and “test” set.

The AARD% of “test” set is less than 1.7 % which indicates that the model has good predictability. About 38% of predicted values present deviations between 0 to 1.00%, 46 % between 1.01 to 3.00%, and 16% over than 3.0%. The maximum prediction error belongs to “1-octyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]imide” ([C8MIm][NTf2]) which has an AARD % of 3.43 for one data point. Also, “1-methyl-3-octylimidazolium chloride” ([OMIm][Cl]) shows a relatively large deviation because of nonlinear temperature dependence of u (non-predictable by our current linear model) . Other ILs have satisfactory predicted results due to their linear temperature dependence of u .

A comparison between the model developed and the model proposed by Gardas and Coutinho [48] shows that the model developed has greater accuracy (AARD% = 0.92% vs 1.96%) and does not require other thermophysical properties of ILs as the other model needs the data of density and surface tension. Furthermore, there is no need to calculate descriptors or use additional property estimation models to estimate the u of new ionic liquids for which their cations and anions are available in this study. It is just required to insert the corresponding descriptors of cations and anions in the presented model. As a result, the u of 319 ionic liquids (the number of cations multiplied by number of anions) can be estimated using previously calculated descriptors.

To compare the QSPR model with GC model developed in previous section, it is notable that the GC model has better accuracy (AARD% of 0.36% compared to 0.92% for the QSPR model); however it requires a special software package to use the SVM. So the application of QSPR model is easier. The summary of all available models for the speed of sound in ionic liquids is shown in Table 6.7.

All information for the entire dataset as well as the value of descriptors of ILs are available in the supplementary CD.

Table 6.7: Summary of available models for the speed of sound in ionic liquids.

Model	Model Type and parameters	N _{ILs}	N _{data}	AARD%
Gardas and Coutinho [48]	Correlation, ρ , σ	14	133	1.96
Singh and Singh [52]	Correlation, ρ , σ	3	60	n.a.
GC LSSVM Model	LSSVM, 8 parameters (7 GCs)	41	446	0.36
QSPR Model (equation 6.1)	9 parameters (8 descriptors)	41	446	0.92

6.3 Liquid heat capacity of ionic liquids

6.3.1 The GC model*

In order to find the most accurate model with an acceptable number of variables, several possible models with different sizes were investigated. The changes in the accuracy of models in terms of the Average Absolute Relative Deviation with respect to the number of functions groups is shown in Figure 6.7, for models determined by both the GFA and classical linear GC method. It can be easily seen that the GFA models show better accuracy for the same number of parameters after the 6th functional group. As indicated in the figure, improvements in the accuracy of classical linear GC models becomes constant after the addition of the 29th variable; but in GFA models, there is continual improvement by addition of a new functional groups or new combination of previously entered groups in the model.

To determine the least complex, but most accurate model, the changes in accuracy of the model versus the number of functional groups is plotted in Figure 6.8. From this plot it is evident that a 13-parameters model would suffice. The introduction of additional functional groups did not show noticeable changes in model accuracy; however a larger model would have better accuracy. Figure 6.9 shows that after the 16th parameter, the AARD% of the test set becomes lower than that of the training set and this behavior is also observed for larger models. This

* The results have been published in *J Therm Anal Calorim* (2014) 115:1863–1882.

behavior indicates that some new functional groups should be added to the model to get better prediction. As a result, the 16-parameters model was selected as a final model.

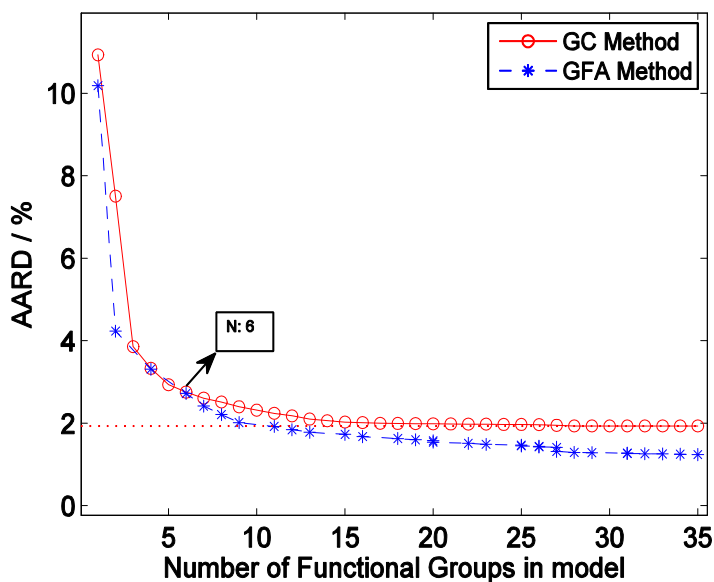


Figure 6.7: Effect of the number of functional groups on the accuracy of the classical GC and GFA models of C_{pL}

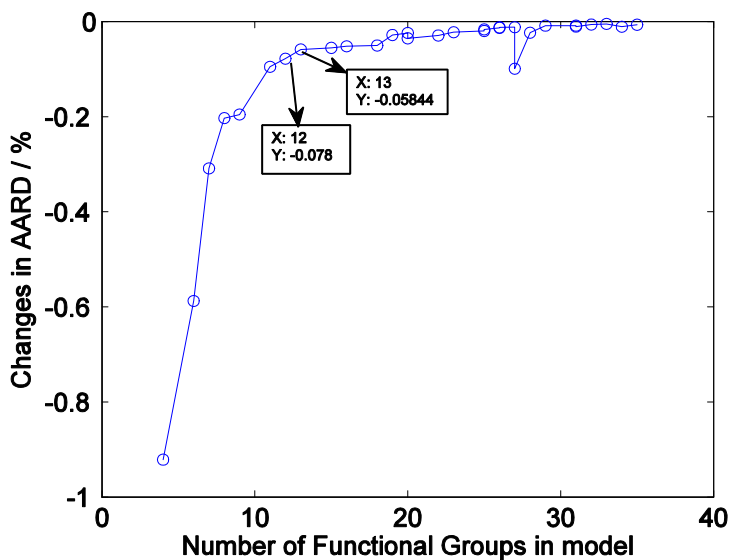


Figure 6.8: Changes in the accuracy of C_{pL} model versus the increasing number of functional groups.

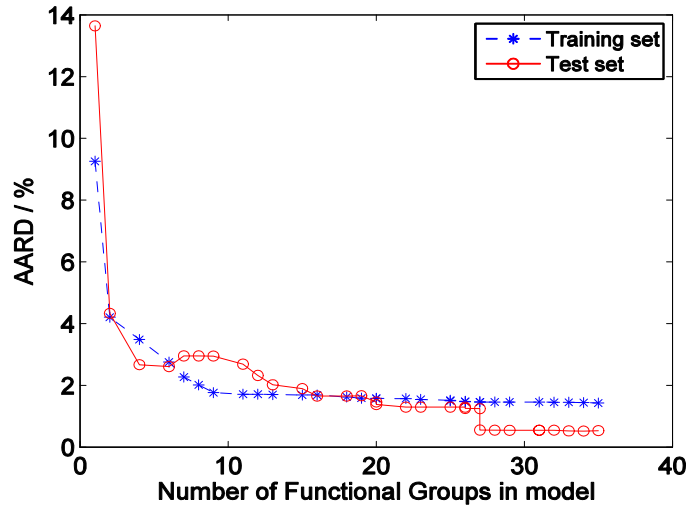


Figure 6.9: The effect of the number of C_{pL} model parameters on the accuracy of the training and test sets.

The result of the proposed modeling approach produces the model presented below where T is absolute temperature. The other model parameters are defined in Table 6.8.

6.2

$$\ln \rho_{pL} = \ln \rho + \ln \rho_{pL} + \ln \rho_{pL}^2$$

$$\ln \rho = 8.292 \ln \rho_{pL} + 11.477 \ln \rho_{pL}^2$$

$$\ln \rho = 0.319$$

$$-7.773 \times 10^{-4} \ln \rho - \ln \rho_{pL} - \ln \rho_{pL}^2 \times \ln \rho - \ln \rho_{pL} - \ln \rho_{pL}^2$$

$$-4.827 \times 10^{-4} \ln \rho_{pL}^3 \times \ln \rho_{pL}^5$$

$$-1.649 \times 10^{-4} \ln \rho_{pL}^2 \times \ln \rho_{pL}$$

$$-1.636 \times 10^{-4} \ln \rho_{pL}^2 \times \ln \rho_{pL}^2$$

$$+2.394 \times 10^{-4} \ln \rho_{pL} \times \ln \rho_{pL}^2$$

$$+5.319 \times 10^{-4} \ln \rho_{pL} \times \ln \rho_{pL}$$

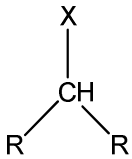
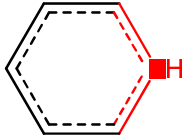
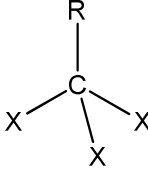
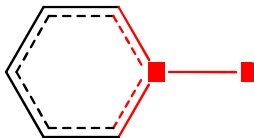
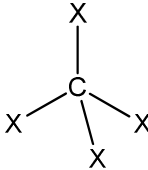
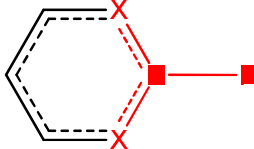
$$-4.838 \times 10^{-4} \ln \rho_{pL} \times \ln \rho_{pL}^3$$

$$+1.786 \times 10^{-4} \ln \rho_{pL} \times \ln \rho_{pL}^4$$

$$-1.355 \times 10^{-4} \ln \rho - \ln \rho_{pL} - \ln \rho_{pL}^2 \times \ln \rho_{pL}^2$$

$$\Delta = +2.473 \times 10^{22} \times \Delta \Delta \Delta \Delta_{\Delta \Delta}$$

Table 6.8: Description of parameters of the equation (6.2)

No.	Symbol	Description	No.	Symbol	Description
1	$\Delta \Delta \Delta \Delta \Delta \Delta$	number of atoms	8	$\Delta \Delta$	number of Nitrogen atoms
2	$\Delta \Delta \Delta$	number of non-h atoms	9	$\Delta \Delta$	number of Sulfur atoms
3	$\Delta \Delta \Delta$	number of triple bonds	10	$\Delta \Delta \Delta$	number of Chlorine atoms
4	$\Delta \Delta 05$	number of 5-membered rings	11	$\Delta \Delta 3 \Delta$	CH ₃ —R
5	$\Delta \Delta \Delta \Delta \Delta$		12	$\Delta - - \Delta \Delta - - \Delta$	
6	$\Delta \Delta \Delta 3$		13	$\Delta - - \Delta \Delta - - \Delta$	
7	$\Delta \Delta 4$		14	$\Delta - - \Delta \Delta - - \Delta$	

The AARD for the proposed model, represented by equation (6.2), is 1.68% for the training set which has 2939 experimental data points for 65 ILs, and 1.65% for the test set which consists of 787 experimental data points for 17 ILs. The coefficient of determination (R^2) for the training and test sets are 0.987 and 0.997, respectively. A summary of the statistical parameters for the model for the training and test sets can be found in Table 6.9.

In order to visualize the results of modeling, the values of the predicted C_{pL} versus the experimental data are presented in Figure 6.10. In addition, the deviation of the model from the experimental data is shown in Figure 6.11. As seen in these figures, there are cases for some ILs for which the proposed model cannot predict the heat capacity well. As indicated in Table 6.10, the largest deviation is observed for “1-hexyl-4-(4-methyl-1-piperidiny)pyridinium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide” which has an AARD% of 33.67%. This ionic

liquid however had only two experimental points with an uncertainty of 7% ($64.50 \text{ J mol}^{-1} \text{ K}^{-1}$).

Table 6.9: Statistical parameters for equation (6.2).

<i>Statistical Parameter</i>	
<i>training set</i>	
R ²	0.987
Absolute average relative deviation	1.68
Standard deviation error	19.71
Root mean square error	19.71
No. of data points	2939
<i>test set</i>	
R ²	0.997
Absolute average relative deviation	1.65
Standard deviation error	12.04
Root mean square error	12.50
No. of data points	787
<i>total</i>	
R ²	0.990
Absolute average relative deviation	1.68
Standard deviation error	18.42
Root mean square error	18.42
No. of data points	3726

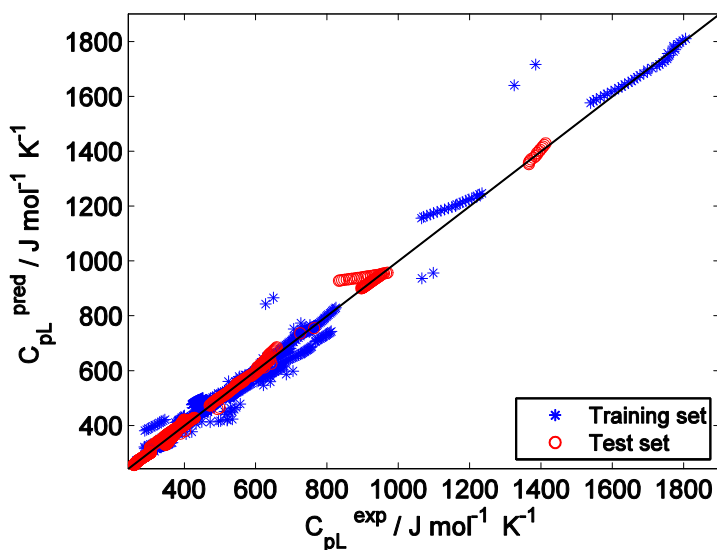


Figure 6.10: Predicted versus experimental values of C_{pL} (GC model).

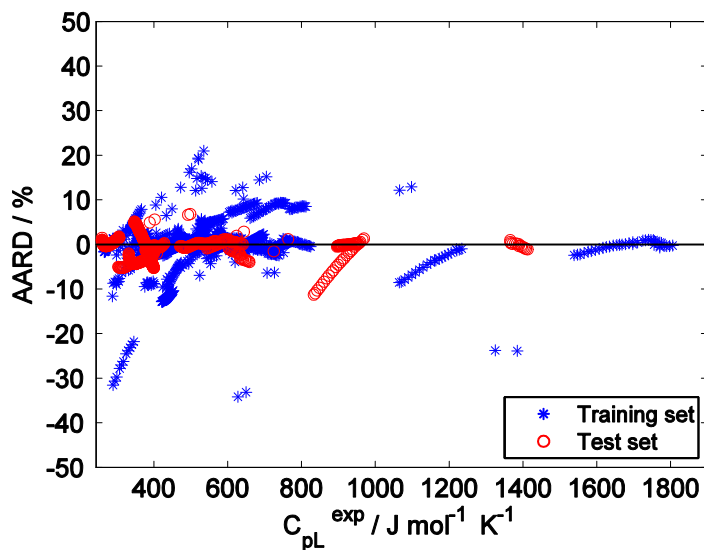


Figure 6.11: Relative deviation of predicted C_{pL} from experimental data (GC model).

There are several other ILs in the dataset used which also have two data points with large uncertainties and predictions for these ILs also show large deviations. Interestingly, all these data points have been measured by Crosthwaite *et al.* [236] whom generally have produced data which are not in good agreement with measurements of other researchers. Figure 6.12 and Figure 6.13 show such deviations for two ionic liquids. It is therefore highly probable that the experimental data for these ILs for which only two data points are available are not reliable. This observation as well as the large uncertainty value of the measured C_{pL} is the most probable explanation for the large observed deviations between the predicted and experimental data.

The second IL with the largest deviation (26.26%) is “1-ethyl-3-methylimidazolium hexafluorophosphate”. There are four IL in the dataset for which the anion is “hexafluorophosphate”. The calculations show that the model presented in this study cannot predict the C_{pL} of the ILs with this anion well. However, the exception is “1-butyl-3-methylimidazolium hexafluorophosphate” which has the AARD% of 0.82%. For “1-ethyl-3-methylimidazolium hexafluorophosphate”, it seems that there is a problem with the experimental data for this, because similar deviation (26.82%) has been reported by Soriano *et al.* with respect to their accurate model [55].

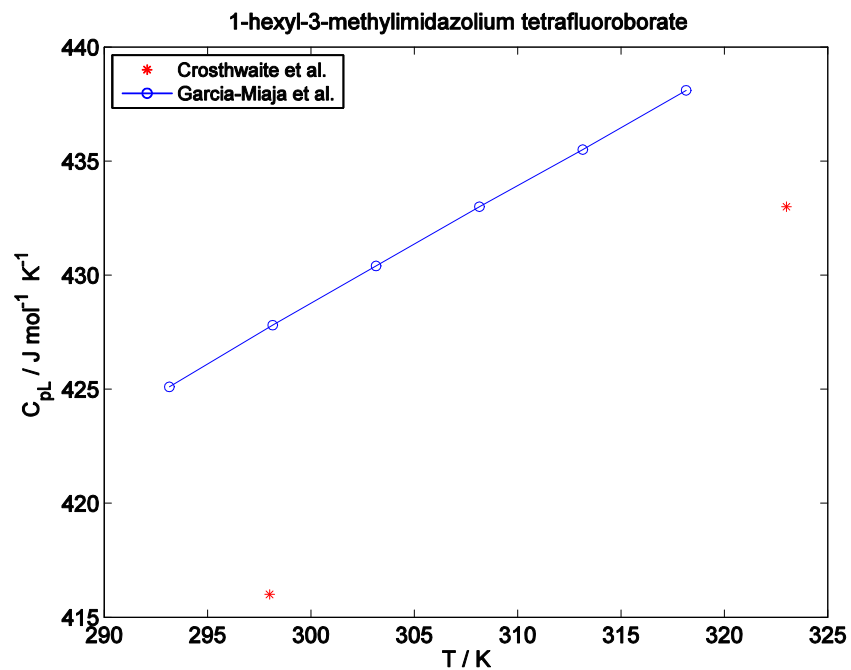


Figure 6.12: Comparison of C_{pL} data for 1-hexyl-3-methylimidazolium tetrafluoroborate measured by Crosthwaite *et al.* [236] and Garcia-Miaja *et al.* [237]

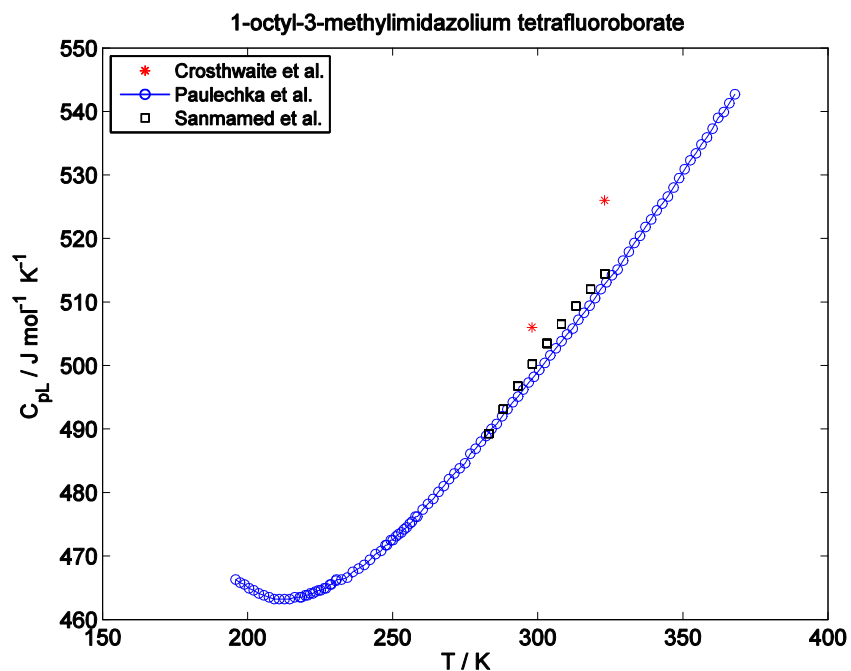


Figure 6.13: Comparison of C_{pL} data for 1-octyl-3-methylimidazolium tetrafluoroborate measured by Crosthwaite *et al.* and other researchers.

Finally, Table 6.11 shows the AARD% of each family of ionic liquids. The results indicate that the model can correlate and predict each family on average with similar accuracy which shows the comprehensiveness of the model. The maximum deviation is observed in Phosphonium (3.65%) liquids and the minimum in Dialkyl Imidazolium ionic liquids (1.51%). It should be noted that the Quinolinium family has the absolute minimum deviation (0.19%); but as there is just one ionic liquid in that family, hence it cannot be considered as a well-predicted class of ILs.

The proposed model performs better than the previous models presented in terms of simplicity, accuracy, and comprehensiveness. According to Table 2.2, the model proposed by Soriano et al. [55] has an AARD% of 0.34%; however it covers only 32 ILs. The model proposed in this study covers 82 ILs resulting in a higher overall AARD% of 1.68%. The presented model can predict the C_{pL} of ionic liquids over a wider range of compounds. The summary of comparison between the models is shown in Table 6.14, at the end of next section.

Table 6.10: Name and AARD% of ionic liquids used to develop equation (6.2).

No.	Compound	T/K range	$C_{pL}/J \text{ mol}^{-1} \text{ K}^{-1}$	Uncertainty/ $J \text{ mol}^{-1} \text{ K}^{-1}$	AARD/%	Number of data points	Subset
1	1-methyl-3-(3,3,4,4,5,5,6,6-nonafluorohexyl)-1H-imidazolium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	298-323	725-752	74.50	1.95	2	Train
2	3-hexyl-1-methyl-1H-imidazolium bromide	298-323	344-357	35.50	3.15	2	Train
3	1-(2-hydroxyethyl)-3-methylimidazolium trifluoroacetate	283.15-343.15	362-394	29.43	0.87	7	Train
4	1-tetradecyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	309.98-368.07	896.3-953.3	7.30	0.18	50	Test
5	1-ethyl-3-methylimidazolium bromide	347.66-367.5	265.34-272.58	1.27	1.43	11	Train
6	1-ethyl-3-methylimidazolium tetrafluoroborate	283.15-358.15	303.4-330.7	7.06	5.09	16	Test
7	1-ethyl-3-methylimidazolium trifluoromethanesulfonate	315.15-425.15	386-425	21.61	0.78	23	Train
8	1-ethyl-3-methylimidazolium methanesulfonate	293.15-343.15	327-350	22.43	0.78	6	Train
9	1-ethyl-3-methylimidazolium hexafluorophosphate	353-453	289.43-345.77	12.53	26.26	11	Train
10	1-ethyl-3-methylimidazolium ethyl sulfate	195-390	346.8-399.9	1.61	2.61	210	Test
11	1-ethyl-3-methylimidazolium hydrogen sulfate	283.15-343.15	290-324	18.59	7.35	7	Train
12	1-ethyl-3-methylimidazolium diethylphosphate	283.15-343.15	517-557	43.17	14.77	6	Train
13	1-propyl-3-methylimidazolium glutamate	244.24-357.68	471-568.06	4.55	0.36	58	Test
14	1-octyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide	281.99-372.62	693.71-742.21	19.32	0.48	10	Train
15	1-octyl-3-methylimidazolium tetrafluoroborate	195.88-367.89	463.2-542.7	3.95	1.03	107	Train
16	1-octyl-3-methylimidazolium trifluoromethanesulfonate	315.15-425.15	604-680	30.51	8.19	23	Train
17	1-octyl-3-methylimidazolium bromide	298-323	392-408	35.50	3.20	2	Test
18	1-butyl-3-methylimidazolium chloride	343-453	298.69-354.58	13.10	2.91	12	Train
19	1-butyl-3-methylimidazolium hexafluorophosphate	300.05-524.87	409.22-510.39	9.35	0.82	1528	Train
20	1-butyl-3-methylimidazolium trifluoromethanesulfonate	290.98-370	423.9-466.4	2.98	1.19	48	Train
21	1-butyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide	190-363.18	515-602.48	10.82	0.99	22	Train
22	1-butyl-3-methylimidazolium trifluoroacetate	190-370	367.4-442.6	5.70	1.20	21	Train
23	1-butyl-3-methylimidazolium nitrate	309.16-370	357.7-383	9.12	2.12	8	Train

No.	Compound	T/K range	$C_{pL}/J\ mol^{-1}\ K^{-1}$	Uncertainty/ $J\ mol^{-1}\ K^{-1}$	AARD/%	Number of data points	Subset
24	1-butyl-3-methylimidazolium acetate	210-300	352.4-384.1	4.40	0.56	11	Train
25	1-butyl-3-methylimidazolium methylsulfate	303.2-358.2	375.47-400.51	7.52	8.81	12	Train
26	1-butyl-3-methylimidazolium tosylate	343.89-380	543.4-569.8	12.71	2.82	5	Train
27	1-butyl-3-methylimidazolium dicyanamide	235.8-367.14	355.9-403.2	4.81	0.81	78	Train
28	1-butyl-3-methylimidazolium bis(oxalato)borate	244.30-292.74	528.69-551.92	12.02	4.55	23	Train
29	1-butyl-3-methylimidazolium octylsulfate	298.15-343.15	635.22-697.64	14.49	0.91	46	Train
30	1-butyl-3-methylimidazolium 2-(2-methoxyethoxy)ethyl sulfate	298-323	643-652	65.50	9.31	2	Train
31	1-n-butyl-3-methylimidazolium bromide	225.62-403.2	289.04-366.4	3.87	3.28	35	Train
32	1-n-butyl-3-methylimidazolium tetrafluoroborate	189.66-367.98	332-400.1	2.94	3.46	79	Test
33	1-hexyl-3-methylimidazolium tetrafluoroborate	293.15-318.15	425.1-438.1	5.82	2.18	6	Train
34	1-hexyl-3-methylimidazolium hexafluorophosphate	293.15-343.15	421.53-452.44	13.95	11.88	51	Train
35	1-hexyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide	188.06-370	572-677	2.73	0.63	191	Train
36	1-hexyl-3-methylimidazolium trifluoromethanesulfonate	315.15-425.15	526-589	29.22	5.03	23	Train
37	1-hexyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	293.15-343.15	725.56-767.81	25.45	0.37	51	Train
38	1-hexyl-3-methylimidazolium bis(oxalato)borate	239.33-397.43	575.78-656.71	3.94	0.96	80	Train
39	1-ethyl-2,3-dimethylimidazolium bis[(trifluoromethyl)sulfonyl]amide	309-323	492.7-498.8	9.95	6.72	2	Test
40	1-butyl-2,3-dimethylimidazolium hexafluorophosphate	298-323	433.6-449.1	8.85	7.23	2	Train
41	1-n-butyl-2,3-dimethylimidazolium tetrafluoroborate	330-372	375.3-406.5	7.80	2.45	2	Train
42	1-hexyl-2,3-dimethylimidazolium bis[(trifluoromethyl)sulfonyl]amide	298-323	686-705	70.00	14.82	2	Train
43	1-methyl-3-propylimidazolium bromide	212.2-368.28	259.0973-306.3	1.64	0.43	203	Test
44	1,2-dimethyl-3-propylimidazolium bis[(trifluoromethyl)sulfonyl]amide	323-663	473.47-631.15	11.04	1.59	35	Train
45	1-ethyl-3-methylimidazolium methylsulfate	283.15-343.15	324-354	20.24	5.10	7	Train
46	1-butyl-1-methylpyrrolidinium bis[(trifluoromethyl)sulfonyl]amide	237.44-368.4	546.8-638	4.85	0.71	72	Train
47	1-butyl-1-methylpyrrolidinium trifluoromethanesulfonate	288.15-308.15	424-441	7.67	8.24	3	Train
48	1-butyl-1-methylpyrrolidinium dicyanamide	288.15-308.15	473-521	4.47	16.27	3	Train
49	1-butyl-1-methylpyrrolidinium tris(pentafluoroethyl)trifluorophosphate	293-358	767-812	36.29	8.29	14	Train

No.	Compound	T/K range	$C_{pL}/J\ mol^{-1}\ K^{-1}$	Uncertainty/ $J\ mol^{-1}\ K^{-1}$	AARD/%	Number of data points	Subset
50	1-Butyl-1-methylpyrrolidinium tetracyanoborate	298.2-323.2	524-554	13.54	5.64	2	Train
51	1-methyl-1-propylpyrrolidinium bis[(trifluoromethyl)sulfonyl]amide	283.15-358.15	544.2-594	8.26	0.44	16	Test
52	1-octyl-3-methylpyridinium bis(trifluoromethylsulfonyl)imide	298-323	669-693	69.00	0.65	2	Train
53	1-hexyl-4-(4-methyl-1-piperidinyl)pyridinium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	298-323	628-650	64.50	33.67	2	Train
54	n-ethyl-4-(n',n'-dimethylammonium)pyridinium bis(trifluoromethylsulfonyl)imide	315.15-425.15	603-659	33.30	2.42	23	Test
55	n-butyl-4-(n',n'-dimethylammonium)pyridinium bis(trifluoromethylsulfonyl)imid	315.15-425.15	672-739	36.35	1.60	23	Train
56	1-hexyl-3-methyl-4-(dimethylamino)pyridinium bis[(trifluoromethyl)sulfonyl]amide	298-323	725-764	75.00	1.36	2	Test
57	1-methylpyridinium methylsulfate	288.15-308.15	288-305	6.00	9.65	3	Train
58	1-ethyl-3-methylpyridinium ethylsulfate	298-323	389-402	40.00	5.24	2	Test
59	N-octyl-3-methylpyridinium tetrafluoroborate	278.15-328.15	433.8-473.7	27.09	6.73	21	Train
60	1-butyl-3-methylpyridinium trifluoromethanesulfonate	288.15-308.15	496-535	30.00	18.83	3	Train
61	1-butyl-3-methylpyridinium bis[(trifluoromethyl)sulfonyl]amide	298-323	622-641	64.00	12.44	2	Train
62	1-butyl-3-methylpyridinium tetrafluoroborate	298-323	405-421	42.00	9.69	2	Train
63	1-hexyl-3-methylpyridinium bromide	298-323	343-358	13.55	1.42	2	Train
64	1-hexyl-3-methylpyridinium bis[(trifluoromethyl)sulfonyl]amide	298-323	624-644	64.00	2.39	2	Test
65	1-hexyl-3,5-dimethylpyridinium bis(trifluoromethylsulfonyl)imide	298-323	620-665	65.00	2.42	2	Train
66	1-propylpyridinium tetrafluoroborate	278.15-338.15	352-385	7.90	1.55	25	Train
67	1-butylpyridinium bis[(trifluoromethyl)sulfonyl]amide	330.15-425.15	587-641	28.86	0.61	20	Test
68	N-butylpyridinium tetrafluoroborate	286.06-390	377.18-428.45	0.88	0.68	62	Test
69	3-methyl-N-butylpyridinium tetracyanoborate	298.2-323.2	495-524	59.98	2.95	2	Train
70	1-hexylpyridinium bis(trifluoromethylsulfonyl)imide	298-323	612-632	63.00	1.95	2	Train
71	4-(dimethylamino)-1-hexyl-pyridinium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	315.15-425.15	750-825	51.37	0.54	23	Train

No.	Compound	T/K range	$C_{pL}/J\ mol^{-1}\ K^{-1}$	Uncertainty/ $J\ mol^{-1}\ K^{-1}$	AARD/%	Number of data points	Subset
72	N-hexylquinolinium bis(trifluoromethylsulfonyl)imide	322.72-370.13	578.47-599.38	58.11	0.19	77	Train
73	trihexyl(tetradecyl)phosphonium bis[(trifluoromethyl)sulfonyl]amide	293-358	1366-1413	65.71	0.52	14	Test
74	trihexyl(tetradecyl)phosphonium tris(pentafluoroethyl)trifluorophosphate	338.15-513.15	1539.1-1805.7	163.40	0.75	36	Train
75	tributyl(methyl)phosphonium methyl sulfate	343.15-463.15	660.8-757.4	72.02	8.16	25	Train
76	trihexyltetradecylphosphonium chloride	338.15-463.15	833.5-969.5	84.57	4.53	26	Test
77	trihexyltetradecylphosphonium dicyanamide	313.15-413.15	1065.1-1234.5	108.29	4.24	21	Train
78	cocosalky pentaethoxi methylammonium methylsulfate	298-323	1066-1098	109.50	12.52	2	Train
79	butyltrimethylammonium bis(trifluoromethylsulfonyl)imide	278.32-367.93	547.5-600.9	8.21	1.16	48	Train
80	tetrabutylammonium docusate	298-323	1325-1385	137.00	23.84	2	Train
81	1-butyl-nicotinic acid butyl ester 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	298-323	707-727	72.00	6.35	2	Train
82	1-ethyl-nicotinic acid ethyl ester ethylsulfate	298-323	513-530	53.00	12.34	2	Train

Table 6.11: The AARD% of equation (6.2) for different families of ionic liquids.

No.	Family	T/K range	$C_{pL}/J\ mol^{-1}\ K^{-1}$	AARD/%	Number of ILs	Number of data points
1	Ammonium	278.32-367.93	547.5-1385	2.47	3	52
2	Dialkyl imidazolium	188.06-524.87	259.0973-953.3	1.51	40	3093
3	Phosphonium	293-513.15	660.8-1805.7	3.65	5	122
4	Pyridinium	278.15-425.15	288-825	2.66	22	229
5	Pyrrolidinium	237.44-368.4	424-812	2.36	6	110
6	Quinolinium	322.72-370.13	578.47-599.38	0.19	1	77
7	Trialkyl imidazolium	298-663	375.3-705	2.74	5	43

6.3.2 The QSPR model*

Same as the GC model, different models were developed to find the most accurate one with less number of parameters. It was observed that enhancement in the accuracy of classical MLR models becomes insignificant after the 30th descriptor; but in the GFA models, continuous improvement was perceived by the addition of new descriptors or new combination of previously entered descriptors in the model (Figure 6.14).

In order to find the most accurate, but least complex model, the variation in the accuracy of the model versus the number of functional groups was plotted in Figure 6.15. Accordingly, it was observed that after the 13th descriptor, there were not any noticeable changes in the accuracy of the model; however larger models would be more accurate.

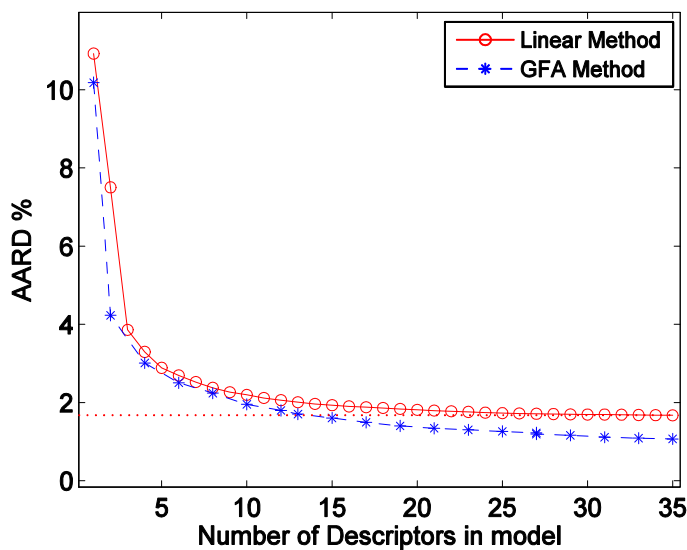


Figure 6.14: Effect of the number of descriptors on the accuracy of the linear QSPR and GFA models of C_{pL}

* The results have been published in *Ind. Eng. Chem. Res.* 2013, 52, 13217–13221

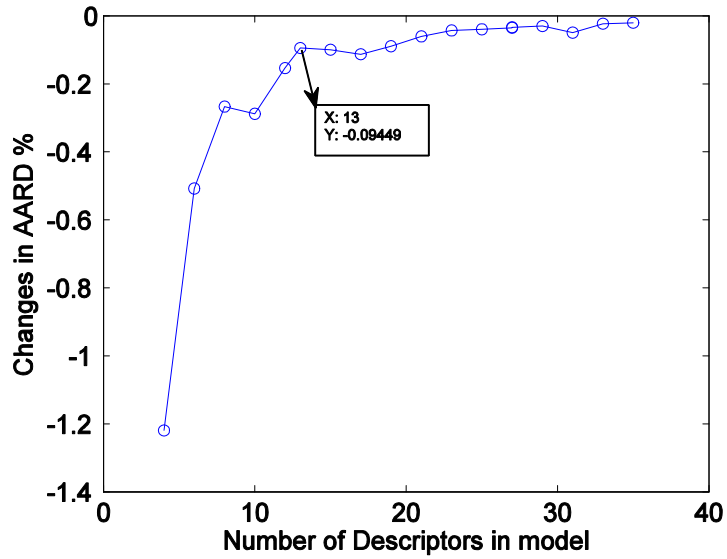
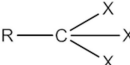


Figure 6.15: Changes in the accuracy of C_{pL} model versus the increasing number of descriptors.

The final model is presented as equation (6.3), where T is absolute temperature. The other parameters of the model are defined in Table 6.12.

$$\begin{aligned}
 \ln \tau_{pL} &= \ln \tau + \ln \tau_{pL} & (6.3) \\
 \ln \tau &= 8.404 \ln \tau_{pL} + 10.246 \ln \tau_{pL} \\
 \ln \tau &= 0.272 \\
 &+ 5.074 \ln \tau_{pL} \times \ln \tau_{pL} + 6.594 \ln \tau_{pL} \times \ln \tau_{pL} \\
 &- 2.235 \times 10^{-2} \ln \tau_{pL} \times \ln \tau_{pL} + 0.448 \ln \tau_{pL} \times \ln \tau_{pL} \\
 &+ 2.790 \times 10^{-2} \ln \tau_{pL} \times \ln \tau_{pL} + 0.108 \ln \tau_{pL} \times \ln \tau_{pL}
 \end{aligned}$$

Table 6.12: Definition of descriptors used in equation (6.3)

No.	Descriptors	Definitions
1	22222 2	Number of atoms
2	222	Number of non-h atoms
3	2232	
4	2223	
5	2 22222	Moran autocorrelation - lag 2 / weighted by atomic masses
6	2 22242	Moran autocorrelation - lag 4 / weighted by atomic masses
7	2 22242	Moran autocorrelation - lag 4 / weighted by atomic Sanderson electronegativities
8	2221302	Radial Distribution Function - 13.0 / unweighted
9	22 22222	Fractional accessible surface area of hydrogen bond acceptors
10	2 222	Eccentric
11	222209	Molecular multiple path count of order 9
12	22222	Eigenvalue sum from Z weighted distance matrix (Barysz matrix)

R represents any group linked through carbon

X represents any heteroatom (O, N, S, P, Se, halogens)

The AARD% for equation (6.3) is 1.55% for the “training” set which has 3001 experimental data points for 61 ILs, and 2.32 % for the “test” set consisting of 725 experimental data points for 21 ILs. The coefficient of determination (R^2) for the “training” and “test” sets is 0.990 and 0.996, respectively. The results of statistical analyses of the model are summarized in Table 6.13.

According to Figure 6.16, 62% of the calculated values of the C_{pL} in the “training” set show deviations between 0 to 1%, 25 % between 1.01 to 3%, 8% between 3.01 to 5, 3% between 5.01 to 10, and 2% over 10%. For the “test” set, the predicted values of the C_{pL} show that 41% of data points are within an AARD range of 0-1%, 33 % within the range of 1-3%, 15% within the range of 3-5%, about 8% in the range of 5-10%, and 3% greater than 10%.

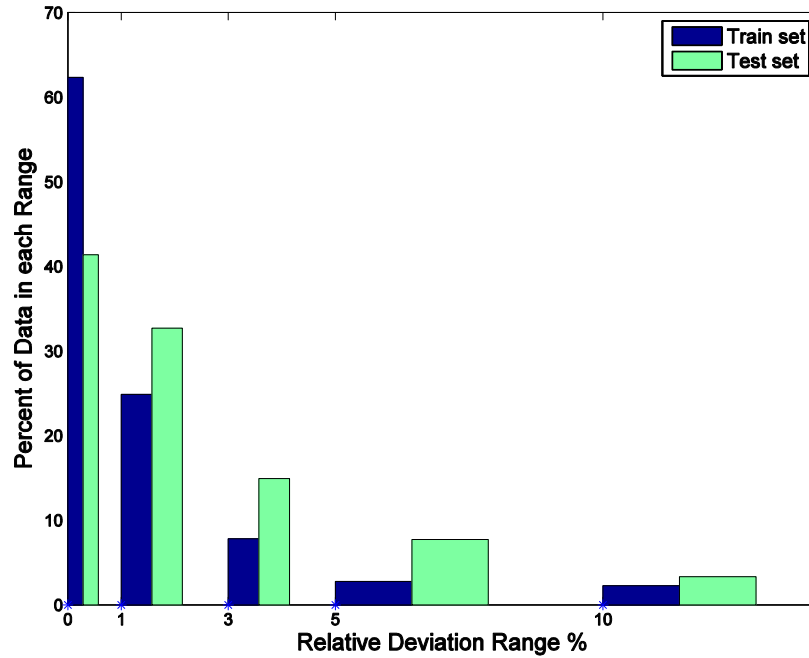


Figure 6.16: Percentage of the predicted values of C_{pL} in different relative deviation ranges.

Table 6.13: Statistical parameters for equation (6.3)

<i>Statistical Parameter</i>	
<i>training set</i>	
R^2	0.990
Absolute average relative deviation	1.55
Standard deviation error	14.19
Root mean square error	14.20
No. of data points	3001
<i>test set</i>	
R^2	0.996
Absolute average relative deviation	2.32
Standard deviation error	18.42
Root mean square error	18.42
No. of data points	725
<i>total</i>	
R^2	0.993
Absolute average relative deviation	1.70
Standard deviation error	15.11
Root mean square error	15.11
No. of data points	3726

Figure 6.17 shows the experimental data for the C_{pL} versus the predicted values. In addition, Figure 6.18 demonstrates the deviation of the model from the experimental data. In these figures, some ILs show a large deviation which means the failure of the model for prediction of these data points. As indicated in Table 6.15, the largest deviation is observed for “1-ethyl-2,3-dimethylimidazolium bis- [(trifluoromethyl)sulfonyl]amide”, with an AARD% of 27.9%. This IL has only two experimental points. Additionally, several other ILs are present in the dataset which have two data points and show large deviations. As mentioned in section 6.3.1, these data points were measured by Crosthwaite *et al.* [236] and it is highly probable that the experimental data for these two-point ILs are not reliable, and consequently the large deviations with respect to the proposed model.

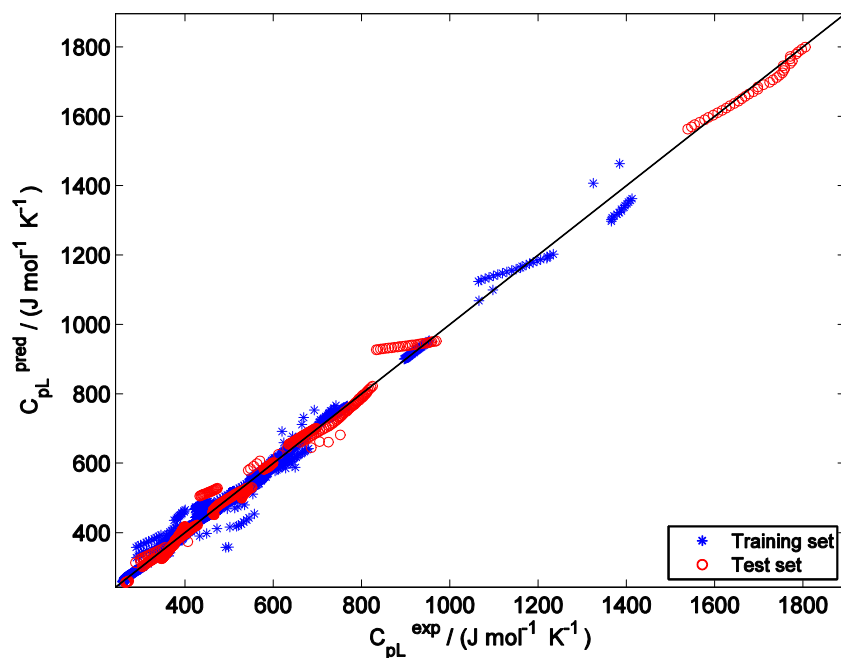


Figure 6.17: Predicted versus experimental values of C_{pL} (QSPR model).

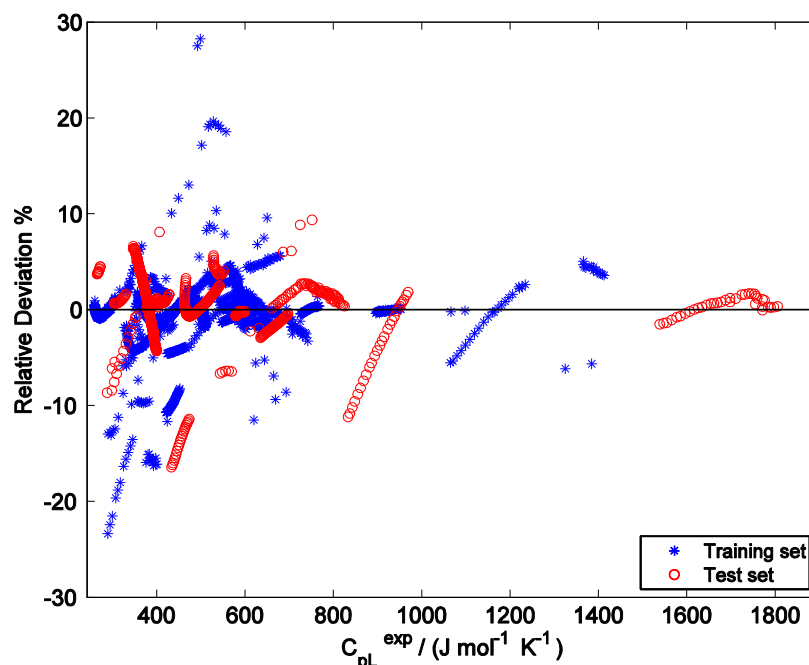


Figure 6.18: Relative deviation of predicted C_{pL} from experimental data (QSPR model).

The IL with the second largest deviation (19.1%) is “1-ethyl-3-methylimidazolium diethylphosphate”. The largest model found during this study, which had 35 descriptors, was also used to predict the C_{pL} of this IL, and it was observed that it can be predicted well with an AARD% of 0.8%. Consequently, more descriptors are required for better prediction of some of the ionic liquids in dataset, but the aim of this study is to produce the simplest model with still keeping fairly acceptable prediction capability for most of the ILs in the dataset. Similar behavior was observed for most of the ILs with large deviations and more than two data points. The only exception was “1-ethyl-3-methylimidazolium hexafluorophosphate”. It seems that the experimental data for this IL is problematic, because similar deviation (26.8%) is reported by Soriano *et al.* with respect to their accurate model [55].

A comparison between equation (6.3), equation (6.2), and previous models (Table 2.2) shows that the proposed QSPR model performs better than other models. Both the QSPR and GC models presented in this study have similar accuracy, 1.70% for the QSPR model and 1.68% for the GC model; but the former has fewer parameters. In section 4.2.2 it was discussed that

QSPR models usually have fewer parameters compared to equivalent GC models; so the QSPR model seems simpler, but the calculation of descriptors cannot be done manually and needs special software. As a result, the GC model is more beneficial and applicable when such kinds of software are not available.

Table 6.14. Summary of available models for the of heat capacity of ILs.

Model	Model Type and parameters	N _{ILs}	N _{data}	AARD%
Gardas and Coutinho [54]	GC, 12 parameters	19	2396	0.36
Gardas and Coutinho [54]	Correlation, V_m	19	2396	1.85
Soriano <i>et al.</i> [55]	GC, 10 cations and 14 anions	32	2414	0.34
Valderrama <i>et al.</i> [56]	MCI, 40 parameters	15	541	0.8
Paulechka <i>et al.</i> [60]	Correlation, V_m	19	653	6.0 (max error)
Preiss <i>et al.</i> [59]	Correlation, V_m	20	n.a	1.2 (max error)
GC Model (equation 6.2)	16 GCs	82	3726	1.68
QSPR Model (equation 6.3)	13 descriptors	82	3726	1.70

Information on the entire dataset and original data sources, as well as the values of the functional groups for ILs are available in the supplementary CD.

Table 6.15: Name and AARD% of studied ionic liquids for developing equation (6.3)

No.	Ionic Liquid	T (K) range	C_{pL}^{exp} (J Mol ⁻¹ K ⁻¹) range	AARD%	N	Subset
1	1-methyl-3-(3,3,4,4,5,5,6,6,6-nonafluorohexyl)-1H-imidazolium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	298-323	725-752	9.09	2	Test
2	3-hexyl-1-methyl-1H-imidazolium bromide	298-323	344-357	3.89	2	Train
3	1-(2-hydroxyethyl)-3-methylimidazolium trifluoroacetate	283.15-343.15	362-394	0.85	7	Train
4	1-tetradecyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	309.98-368.07	896.3-953.3	0.15	50	Train
5	1-ethyl-3-methylimidazolium bromide	347.66-367.5	265.3375-272.5791	4.04	11	Test
6	1-ethyl-3-methylimidazolium tetrafluoroborate	283.15-358.15	303.4-330.7	0.94	16	Test
7	1-ethyl-3-methylimidazolium trifluoromethanesulfonate	315.15-425.15	386-425	2.33	23	Train
8	1-ethyl-3-methylimidazolium methanesulfonate	293.15-343.15	327-350	1.84	6	Train
9	1-ethyl-3-methylimidazolium hexafluorophosphate	353-453	289.43-345.77	18.05	11	Train
10	1-ethyl-3-methylimidazolium ethyl sulfate	195-390	346.8-399.9	2.86	210	Test
11	1-ethyl-3-methylimidazolium hydrogen sulfate	283.15-343.15	290-324	11.98	7	Train
12	1-ethyl-3-methylimidazolium diethylphosphate	283.15-343.15	517-557	19.12	6	Train
13	1-propyl-3-methylimidazolium glutamate	244.243-357.682	471-568.06	3.10	58	Train
14	1-octyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide	281.99-372.62	693.715-742.213	1.94	10	Train
15	1-octyl-3-methylimidazolium tetrafluoroborate	195.88-367.89	463.2-542.7	0.85	107	Test
16	1-octyl-3-methylimidazolium trifluoromethanesulfonate	315.15-425.15	604-680	4.91	23	Train
17	1-octyl-3-methylimidazolium bromide	298-323	392-408	3.89	2	Train
18	1-butyl-3-methylimidazolium chloride	343-453	298.69-354.58	4.14	12	Test
19	1-butyl-3-methylimidazolium hexafluorophosphate	300.05-524.87	409.223-510.392	0.78	1528	Train
20	1-butyl-3-methylimidazolium trifluoromethanesulfonate	290.98-370	423.9-466.4	4.29	48	Train
21	1-butyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide	190-363.18	514.998-602.48	1.45	22	Train
22	1-butyl-3-methylimidazolium trifluoroacetate	190-370	367.4-442.6	0.52	21	Train
23	1-butyl-3-methylimidazolium nitrate	309.16-370	357.7-383	9.64	8	Train
24	1-butyl-3-methylimidazolium acetate	210-300	352.4-384.1	2.38	11	Train
25	1-butyl-3-methylimidazolium methylsulfate	303.2-358.2	375.47-400.51	15.73	12	Train
26	1-butyl-3-methylimidazolium tosylate	343.89-380	543.4-569.8	6.48	5	Test
27	1-butyl-3-methylimidazolium dicyanamide	235.8-367.14	355.9-403.2	0.88	78	Train
28	1-butyl-3-methylimidazolium bis(oxalato)borate	244.303-292.74	528.69-551.92	4.27	23	Test

No.	Ionic Liquid	T (K) range	C_{pL}^{exp} (J Mol ⁻¹ K ⁻¹) range	AARD%	N	Subset
29	1-butyl-3-methylimidazolium octylsulfate	298.15-343.15	635.22-697.64	1.75	46	Test
30	1-butyl-3-methylimidazolium 2-(2-methoxyethoxy)ethyl sulfate	298-323	643-652	6.42	2	Train
31	1-n-butyl-3-methylimidazolium bromide	225.62-403.2	289.0441-366.4	2.11	35	Train
32	1-n-butyl-3-methylimidazolium tetrafluoroborate	189.66-367.98	332-400.1	3.44	79	Train
33	1-hexyl-3-methylimidazolium tetrafluoroborate	293.15-318.15	425.1-438.1	1.87	6	Train
34	1-hexyl-3-methylimidazolium hexafluorophosphate	293.15-343.15	421.527-452.439	9.77	51	Train
35	1-hexyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide	188.06-370	572-677	1.05	191	Train
36	1-hexyl-3-methylimidazolium trifluoromethanesulfonate	315.15-425.15	526-589	2.84	23	Train
37	1-hexyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	293.15-343.15	725.56-767.808	0.24	51	Train
38	1-hexyl-3-methylimidazolium bis(oxalato)borate	239.332-397.436	575.78-656.71	1.27	80	Train
39	1-ethyl-2,3-dimethylimidazolium bis[(trifluoromethyl)sulfonyl]amide	309-323	492.7-498.8	27.87	2	Train
40	1-butyl-2,3-dimethylimidazolium hexafluorophosphate	298-323	433.6-449.1	10.84	2	Train
41	1-n-butyl-2,3-dimethylimidazolium tetrafluoroborate	330-372	375.3-406.5	5.58	2	Test
42	1-hexyl-2,3-dimethylimidazolium bis[(trifluoromethyl)sulfonyl]amide	298-323	686-705	6.07	2	Test
43	1-methyl-3-propylimidazolium bromide	212.2-368.28	259.0973-306.3	0.52	203	Train
44	1,2-dimethyl-3-propylimidazolium bis[(trifluoromethyl)sulfonyl]amide	323-663	473.465-631.146	1.13	35	Train
45	1-ethyl-3-methylimidazolium methylsulfate	283.15-343.15	324-354	5.26	7	Train
46	1-butyl-1-methylpyrrolidinium bis[(trifluoromethyl)sulfonyl]amide	237.44-368.4	546.8-638	0.63	72	Train
47	1-butyl-1-methylpyrrolidinium trifluoromethanesulfonate	288.15-308.15	424-441	10.40	3	Train
48	1-butyl-1-methylpyrrolidinium dicyanamide	288.15-308.15	473-521	16.49	3	Train
49	1-butyl-1-methylpyrrolidinium tris(pentafluoroethyl)trifluorophosphate	293-358	767-812	1.69	14	Test
50	1-Butyl-1-methylpyrrolidinium tetracyanoborate	298.2-323.2	524-554	6.45	2	Train
51	1-methyl-1-propylpyrrolidinium bis[(trifluoromethyl)sulfonyl]amide	283.15-358.15	544.2-594	1.44	16	Train
52	1-octyl-3-methylpyridinium bis(trifluoromethylsulfonyl)imide	298-323	669-693	8.99	2	Train
53	1-hexyl-4-(4-methyl-1-piperidinyl)pyridinium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	298-323	628-650	8.18	2	Train
54	n-ethyl-4-(n',n'-dimethylammonium)pyridinium bis(trifluoromethylsulfonyl)imide	315.15-425.15	603-659	1.02	23	Train
55	n-butyl-4-(n',n'-dimethylammonium)pyridinium bis(trifluoromethylsulfonyl)imide	315.15-425.15	672-739	1.42	23	Train
56	1-hexyl-3-methyl-4-(dimethylamino)pyridinium bis[(trifluoromethyl)sulfonyl]imide	298-323	725-764	1.29	2	Train

No.	Ionic Liquid	T (K) range	C_{pL}^{exp} (J Mol ⁻¹ K ⁻¹) range	AARD%	N	Subset
57	1-methylpyridinium methylsulfate	288.15-308.15	288-305	6.76	3	Test
58	1-ethyl-3-methylpyridinium ethylsulfate	298-323	389-402	0.85	2	Train
59	N-octyl-3-methylpyridinium tetrafluoroborate	278.15-328.15	433.8-473.7	13.70	21	Test
60	1-butyl-3-methylpyridinium trifluoromethanesulfonate	288.15-308.15	496-535	8.21	3	Train
61	1-butyl-3-methylpyridinium bis[(trifluoromethyl)sulfonyl]amide	298-323	622-641	4.59	2	Train
62	1-butyl-3-methylpyridinium tetrafluoroborate	298-323	405-421	2.55	2	Train
63	1-hexyl-3-methylpyridinium bromide	298-323	343-358	8.61	2	Train
64	1-hexyl-3-methylpyridinium bis[(trifluoromethyl)sulfonyl]amide	298-323	624-644	5.41	2	Train
65	1-hexyl-3,5-dimethylpyridinium bis(trifluoromethylsulfonyl)imide	298-323	620-665	9.23	2	Train
66	1-propylpyridinium tetrafluoroborate	278.15-338.15	352-385	1.27	25	Train
67	1-butylpyridinium bis[(trifluoromethyl)sulfonyl]amide	330.15-425.15	587-641	1.14	20	Train
68	N-butylpyridinium tetrafluoroborate	286.06-390	377.18-428.45	0.70	62	Test
69	3-methyl-N-butylpyridinium tetracyanoborate	298.2-323.2	495-524	1.34	2	Train
70	1-hexylpyridinium bis(trifluoromethylsulfonyl)imide	298-323	612-632	2.09	2	Test
71	4-(dimethylamino)-1-hexyl-pyridinium	315.15-425.15	750-825	1.39	23	Test
72	1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide					
72	N-hexylquinolinium bis(trifluoromethylsulfonyl)imide	322.72-370.13	578.47-599.38	0.30	77	Test
73	trihexyl(tetradecyl)phosphonium bis[(trifluoromethyl)sulfonyl]amide	293-358	1366-1413	4.20	14	Train
74	trihexyl(tetradecyl)phosphonium tris(pentafluoroethyl)trifluorophosphate	338.15-513.15	1539.1-1805.7	0.84	36	Test
75	tributyl(methyl)phosphonium methyl sulfate	343.15-463.15	660.8-757.4	1.83	25	Test
76	trihexyltetradecylphosphonium chloride	338.15-463.15	833.5-969.5	4.43	26	Test
77	trihexyltetradecylphosphonium dicyanamide	313.15-413.15	1065.1-1234.5	2.27	21	Train
78	cocosalky pentaethoxi methylammonium methylsulfate	298-323	1066-1098	0.17	2	Train
79	butyltrimethylammonium bis(trifluoromethylsulfonyl)imide	278.32-367.93	547.5-600.9	0.12	48	Train
80	tetrabutylammonium docusate	298-323	1325-1385	5.90	2	Train
81	1-butyl-nicotinic acid butyl ester	298-323	707-727	2.51	2	Train
82	1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide					
82	1-ethyl-nicotinic acid ethyl ester ethylsulfate	298-323	513-530	8.36	2	Train

6.4 Refractive index of ionic liquids

6.4.1 The GC model*

During the development of the GC model for n_D , it was observed that a linear summation of functional groups could produce an accurate and easy-to-use model. It was also observed that the n_D of ILs has a linear dependency with regard to temperature. Hence, the resultant model was a 17-parameter linear model as shown in equation (6.4).

6.4

$$n_D = n_0 + \sum_{i=1}^k a_i n_i$$

$$n_0 = 1.5082 + \sum_{i=1}^k a_i n_i \quad n_0 = -1.4207 \times 10^{-4} + \sum_{i=1}^k a_i n_i$$

where n_i is the number of occurrences of the i th functional group of anions and cations, k is the total number of different functional groups of the anions and cations, and a_i and n_i are the relevant coefficient of the i th functional group according to equation (6.4). The values of a_i and n_i are presented in Tables 2 and 3 respectively.

Table 6.16: Group contribution parameters $n_{a,i}$ and a_i in equation (6.4)

Symbol	Comments	a_i
Cations		
$n_{a,1}$	number of N (Nitrogen)	3.7482×10^{-2}
$n_{a,2}$	number of 6-membered rings	6.2856×10^{-2}
$n_{a,3}$	number of C-(A [†]) ₈ -N	-2.1155×10^{-2}
Anions		
$n_{a,4}$	number of halogen atoms	-3.6817×10^{-3}
$n_{a,5}$	number of H attached to C ¹ (sp ³) / C ⁰ (sp ²)	-4.1550×10^{-3}
$n_{a,6}$	presence/absence of F-A-F (value: 1 or 0)	-7.3159×10^{-2}
$n_{a,7}$	number of C-N	2.0372×10^{-2}
$n_{a,8}$	number of C-S	2.2844×10^{-2}
$n_{a,9}$	number of N-(A) ₃ -N	-2.3814×10^{-2}
$n_{a,10}$	number of C-(A) ₄ -S	6.2458×10^{-2}
$n_{a,11}$	number of C-(A) ₈ -S	-4.7861×10^{-2}

[†] A represents any type of atoms.

* The results have been published in *Journal of Molecular Liquids* 200 (2014) 410–415.

Table 6.17: Group contribution parameters $n_{b,i}$ and b_i in equation (6.4)

Symbol	Comments	b_i
Cations		
$n_{b,1}$	H attached to C ³ (sp3) / C ² (sp2) / C ³ (sp2) / C ³ (sp)	-5.3363×10^{-5}
$n_{b,2}$	presence/absence of C-(A) ₈ -N (value: 1 or 0)	7.8234×10^{-5}
Anions		
$n_{b,3}$	number of H (Hydrogen)	-6.2136×10^{-6}
$n_{b,4}$	number of O (Oxygen)	2.1976×10^{-5}
$n_{b,5}$	number of =O	-6.1456×10^{-5}
$n_{b,6}$	presence/absence of S-A-F (value: 1 or 0)	4.8459×10^{-5}

The AARD% of the proposed model is 0.32% for training set and 0.45% for the test set. The coefficient of determination (R^2) for the “training” and “test” sets is 0.971 and 0.919, respectively. A summary of the statistical parameters of the model for the “training” and “test sets” are listed in Table 6.18.

Table 6.18: Statistical parameters for equation (6.4).

Statistical Parameter	
training set	
R^2	0.971
Absolute average relative deviation	0.32
Standard deviation error	7.43×10^{-3}
Root mean square error	7.43×10^{-3}
No. of data points	759
test set	
R^2	0.919
Absolute average relative deviation	0.45
Standard deviation error	9.94×10^{-3}
Root mean square error	9.99×10^{-3}
No. of data points	172
total	
R^2	0.964
Absolute average relative deviation	0.34
Standard deviation error	7.96×10^{-3}
Root mean square error	7.97×10^{-3}
No. of data points	931

The visual output of the model is illustrated in Figure 6.19, i.e.: the values of the predicted n_D versus the experimental data. In addition, the deviation of the model from the experimental data is shown in Figure 6.20. It can be seen that almost all of the ILs can be predicted fairly well by the proposed model. Figure 6.21 illustrates that the AARD for all of the ILs are less than 3%. For “training” set which has been used to develop the model, 97% of data points are within an AARD% of 0-1%; 2% of the data points are within the range of 1-2%, and 1% of data points which belong to only one ionic liquid shows an AARD% of 2-3%. Similarly according to Figure 6.21, for “test” set, 93% of data points are within an AARD% of 0-1%; 3.5% within the range of 1-2%; and 3.5% within the range of 2-3%.

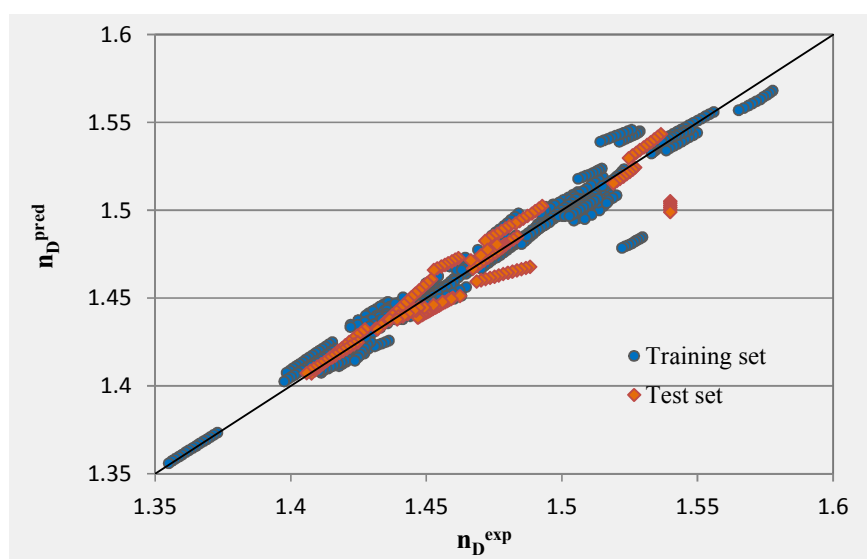


Figure 6.19: Predicted versus experimental values of n_D (— diagonal line).

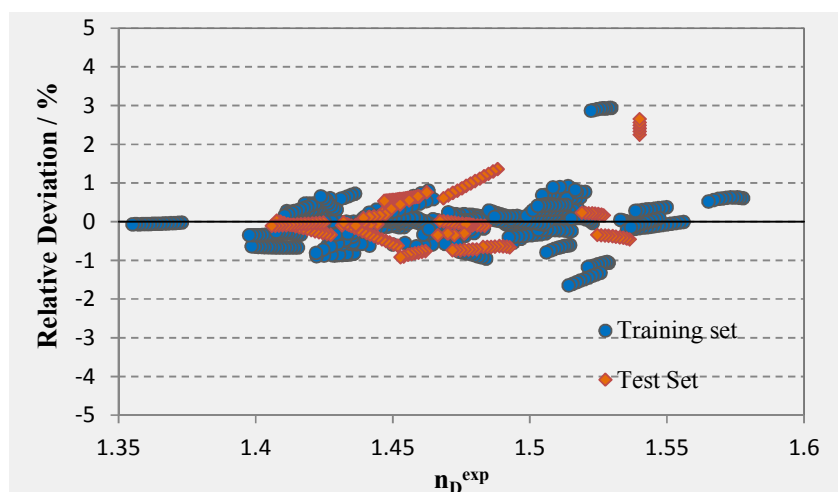


Figure 6.20: Relative deviation of predicted n_D from experimental data.

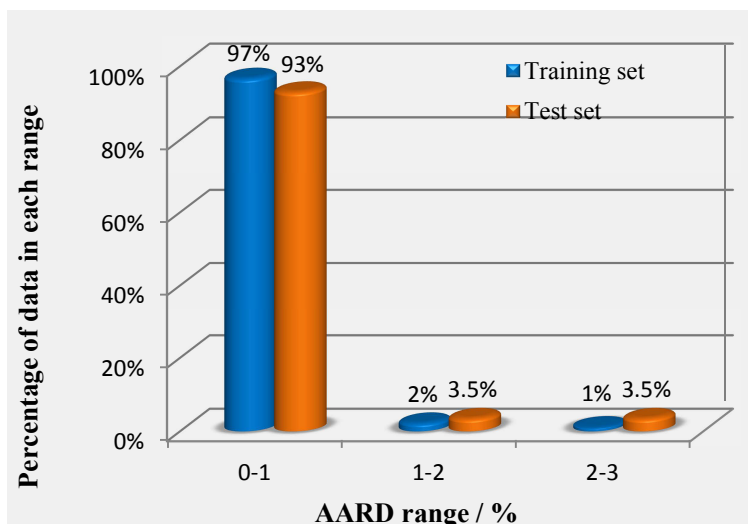


Figure 6.21: Percentage of predicted values of n_D in different relative deviation ranges

As indicated in Table 6.19, the largest deviation is observed for “3-(2-cyanoethyl)-1-hexylimidazolium trifluoromethanesulfonate” which is 2.92%; however this is still a relatively low deviation. The data for this IL was published by Ziyada *et al.* [238] with a reported uncertainty in the measurement of 0.00081 (0.05%). It is probably unlikely that there is error with the data measurement. The possible reason for the deviation is the purity of synthesized ionic liquid, with the reported value being 96.4%. The effect of impurity can be seen in other ionic liquids containing “2-cyanoethyl” side chains in their cations, such as “1-butyl-3-(2-cyanoethyl)-imidazolium chloride” and “1-propyronitrile-3-butylimidazolium bromide”, which have the third and fourth largest deviation (1.48% and 1.10%), respectively. The second largest deviation (2.45%) belongs to 1-butyl-3-methylimidazolium bromide and surprisingly, the reported n_D over the temperature range of 298.3-323 K is constant at 1.54. This data have been reported by Kim *et al.* [239] with an uncertainty of 0.01, which is the largest uncertainty in comparison with the reported values of other ILs. As a result, these data are not precise enough and that is the probable reason of the observed deviation in the test set.

The summary of the results can be found in Table 6.20 which shows the AARD% of proposed model for different families of ionic liquids. It is obvious that the AARD% of the model is less than 1% for all families of ILs which shows that the prediction capability of the model is fairly well.

As mentioned in section 2.4, the previous published models [62, 72, 73, 86] are applicable just for imidazolium-based ionic liquids; so they are not comparable with presented model which covers the various families of ionic liquids and it's not limited to certain type of ionic liquids; however the output of model proposed by Gardas and Coutinho [63] is shown in Table 6.19. It is clearly obvious that its application range is too limited. According to the provided data in the supplementary CD, this model is applicable only for 21 imidazolium based ionic liquids and the AARD% is 0.28%. The model presented in this study has the AARD% of 0.34% over these ionic liquids. So it can be concluded that the GC model developed has the same performance on average compared with the Gardas and Coutinho model; however it had more application range and it is not limited to certain type of ionic liquids.

Information on the entire dataset and original data sources, as well as the values of the functional groups for ILs are available in the supplementary CD.

Table 6.19: Name and AARD% of studied ionic liquids for developing equation (6.4).

No.	Compound	T / K range	Refractive Index	Average Uncertainty	AARD / % GC Model	AARD / % Gardas and Coutinho [63]	Data Points	Subset
1	L-glutamic acid, 1,5-bis(1-methylethyl) ester, dodecyl sulfate	288.15-343.15	1.4392-1.4575	1.10E-03	0.10	n.a.	12	Train
2	L-glutamic acid, 1,5-bis(2-methylpropyl) ester, dodecyl sulfate	323.15-343.15	1.4393-1.4472	1.10E-03	0.14	n.a.	5	Test
3	glycine, 1-methylethyl ester, dodecyl sulfate	303.15-343.15	1.4414-1.463	1.10E-03	0.53	n.a.	9	Train
4	glycine, 2-methylpropyl ester, dodecyl sulfate	303.15-343.15	1.4365-1.4624	1.10E-03	0.36	n.a.	9	Test
5	L-valine, 1-methylethyl ester, dodecyl sulfate	288.15-343.15	1.4388-1.4555	1.10E-03	0.03	n.a.	12	Train
6	L-valine, 2-methylpropyl ester, dodecyl sulfate	288.15-343.15	1.439-1.4568	1.10E-03	0.06	n.a.	12	Train
7	L-proline, 1-methylethyl ester, dodecyl sulfate	288.15-343.15	1.4398-1.4575	1.10E-03	0.10	n.a.	12	Train
8	L-proline, 2-methylpropyl ester, dodecyl sulfate	288.15-343.15	1.4378-1.4648	1.10E-03	0.29	n.a.	12	Train
9	L-alanine, 1-methylethyl ester, dodecyl sulfate	288.15-343.15	1.4371-1.4542	1.10E-03	0.09	n.a.	12	Train
10	L-alanine, 2-methylpropyl ester, dodecyl sulfate	288.15-343.15	1.4336-1.456	1.10E-03	0.16	n.a.	12	Train
11	1-(2-cyanoethyl)-3-(2-propen-1-yl)-imidazolium chloride	293.15-333.15	1.5368-1.5473	4.30E-04	0.07	n.a.	9	Train
12	1-butyl-3-(2-cyanoethyl)-imidazolium chloride	293.15-333.15	1.5142-1.5257	6.20E-04	1.48	n.a.	9	Train
13	1-(2-cyanoethyl)-3-(2-hydroxyethyl)-imidazolium chloride	293.15-333.15	1.5385-1.55	5.20E-04	0.33	n.a.	9	Train
14	3-(6-hydroxyhexyl)-imidazolium chloride	293.15-333.15	1.4989-1.5093	5.00E-04	0.17	n.a.	9	Train
15	1-ethyl-3-(6-hydroxyhexyl)-imidazolium chloride	293.15-333.15	1.5011-1.5115	5.00E-04	0.31	n.a.	9	Train
16	1-butyl-3-(6-hydroxyhexyl)-imidazolium chloride	293.15-333.15	1.5028-1.5132	5.00E-04	0.42	n.a.	9	Train
17	1-propyronitrile-3-decylimidazolium bromide	308.15-333.15	1.5028-1.5079	3.70E-04	0.44	n.a.	6	Train
18	1-butyl-3-ethylimidazolium trifluoromethanesulfonate	288.15-338.15	1.4279-1.4424	5.10E-04	0.43	n.a.	6	Train
19	3-(6-hydroxyhexyl)-1-methylimidazolium chloride	293.15-333.15	1.5002-1.5106	5.00E-04	0.25	n.a.	9	Train
20	1-ethyl-3-methylimidazolium 1,1,2,2-tetrafluoroethanesulfonate	293.15-353.15	1.4251-1.441	4.70E-04	0.00	n.a.	7	Train
21	1-ethyl-3-methylimidazolium acetate	283.15-353.15	1.4848-1.504	6.80E-04	0.13	n.a.	15	Train
22	1-ethyl-3-methylimidazolium ammonioacetate	293.15-333.15	1.4932-1.5063	6.80E-04	0.28	n.a.	9	Train
23	1-ethyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide	283-353	1.4076-1.4274	5.80E-04	0.17	0.28	15	Train
24	1-ethyl-3-methylimidazolium dicyanamide	283-353	1.4956-1.5177	5.97E-04	0.16	n.a.	15	Train
25	1-ethyl-3-methylimidazolium diethylphosphate	298.2-313.2	1.4691-1.4733	6.00E-04	0.57	n.a.	2	Train
26	1-ethyl-3-methylimidazolium dimethylphosphate	298.19-327.64	1.4738-1.4817	3.70E-04	0.05	n.a.	7	Train
27	1-ethyl-3-methylimidazolium ethyl sulfate	283-353	1.4647-1.4832	5.71E-04	0.08	0.79	15	Train

No.	Compound	T / K range	Refractive Index	Average Uncertainty	AARD / % GC Model	AARD / % Gardas and Coutinho [63]	Data Points	Subset
28	1-ethyl-3-methylimidazolium hydrogen sulfate	283-353	1.4851-1.5004	5.50E-04	0.13	n.a.	15	Test
29	1-ethyl-3-methylimidazolium imidodisulfuryl fluoride	283-353	1.4321-1.4518	5.79E-04	0.32	n.a.	15	Test
30	1-ethyl-3-methylimidazolium L-alanine	293.15-333.15	1.4923-1.5059	6.80E-04	0.35	n.a.	9	Train
31	1-ethyl-3-methylimidazolium L-proline	293.15-333.15	1.5031-1.5152	6.80E-04	0.22	n.a.	9	Train
32	1-ethyl-3-methylimidazolium L-serine	293.15-333.15	1.5008-1.512	6.80E-04	0.24	n.a.	9	Train
33	1-ethyl-3-methylimidazolium methanesulfonate	283-353	1.481-1.4999	5.15E-04	0.07	n.a.	21	Train
34	1-ethyl-3-methylimidazolium methyl phosphonate	288.37-317.82	1.4871-1.4953	3.70E-04	0.14	n.a.	7	Train
35	1-ethyl-3-methylimidazolium tetracyanoborate	283-353	1.4292-1.4528	6.08E-04	0.15	n.a.	15	Train
36	1-ethyl-3-methylimidazolium tetrafluoroborate	283-353	1.3986-1.4155	5.60E-04	0.66	0.09	15	Train
37	1-ethyl-3-methylimidazolium thiocyanate	283-353	1.5355-1.556	5.85E-04	0.11	n.a.	15	Train
38	1-ethyl-3-methylimidazolium tosylate	322.73-342.37	1.533-1.5384	3.70E-04	0.03	n.a.	10	Train
39	1-ethyl-3-methylimidazolium tricyanomethide	283.15-313.15	1.5084-1.5154	1.10E-03	0.89	n.a.	4	Train
40	1-ethyl-3-methylimidazolium trifluoromethanesulfonate	288.15-338.15	1.4222-1.436	4.40E-04	0.85	0.13	12	Train
41	1-ethyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	283-353	1.3551-1.3731	5.69E-04	0.05	n.a.	15	Train
42	1-propyl-3-methylimidazolium methylsulfate	298.15-328.15	1.4664-1.4761	2.50E-03	0.32	0.48	4	Test
43	1-propyl-3-methylimidazolium tetrafluoroborate	298.15-338.15	1.4075-1.4165	4.00E-04	0.29	0.06	9	Train
44	1-octyl-3-methylimidazolium hexafluorophosphate	283.15-343.15	1.4106-1.4272	1.01E-03	0.39	0.05	13	Train
45	1-octyl-3-methylimidazolium tetrafluoroborate	283.15-363.15	1.4142-1.4363	8.00E-04	0.41	0.02	16	Train
46	1-butyl-3-methylimidazolium acetate	283.15-353.15	1.4718-1.4927	3.70E-04	0.68	n.a.	15	Test
47	1-butyl-3-methylimidazolium dicyanamide	288.15-308.15	1.5058-1.5121	5.20E-04	0.03	n.a.	3	Train
48	1-butyl-3-methylimidazolium glycine	298.15-313.15	1.5166-1.5202	2.00E-03	0.78	n.a.	4	Train
49	1-butyl-3-methylimidazolium hexafluorophosphate	283.15-343.15	1.3977-1.4133	1.01E-03	0.33	0.03	13	Train
50	1-butyl-3-methylimidazolium L-alanine acid salt	298.15-313.15	1.5135-1.5184	4.00E-04	0.62	n.a.	4	Train
51	1-butyl-3-methylimidazolium methylsulfate	283.15-343.15	1.4672-1.4835	9.52E-04	0.06	0.43	24	Train
52	1-butyl-3-methylimidazolium octylsulfate	288.15-343.15	1.4577-1.4725	5.00E-04	0.14	n.a.	12	Train
53	1-butyl-3-methylimidazolium perchlorate	283.15-353.15	1.4577-1.4763	1.20E-03	0.18	n.a.	15	Train
54	1-butyl-3-methylimidazolium trifluoromethanesulfonate	288.15-338.15	1.4261-1.4401	6.98E-04	0.57	0.04	17	Train

No.	Compound	T / K range	Refractive Index	Average Uncertainty	AARD / % GC Model	AARD / % Gardas and Coutinho [63]	Data Points	Subset
55	1-butyl-3-methylimidazolium bromide	298.3-323	1.54	1.00E-02	2.45	n.a.	6	Test
56	1-butyl-3-methylimidazolium tetrafluoroborate	293.15-353.15	1.4114-1.4259	6.50E-04	0.27	0.41	8	Train
57	1-hexyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide	302.95-332.95	1.4238-1.4296	1.26E-03	0.48	0.10	7	Train
58	1-hexyl-3-methylimidazolium chloride	298.15-343.15	1.5045-1.5172	1.05E-03	0.73	0.61	10	Train
59	1-hexyl-3-methylimidazolium hexafluorophosphate	288.15-318.15	1.4122-1.4206	1.09E-03	0.27	0.11	7	Train
60	1-hexyl-3-methylimidazolium tetrafluoroborate	298.15-338.15	1.4179-1.427	4.00E-04	0.45	0.16	9	Train
61	1-Benzyl-3-methylimidazolium chloride	298.15-343.15	1.5652-1.5778	5.10E-04	0.60	n.a.	10	Train
62	1-benzyl-3-methylimidazolium methylsulfate	298.15-343.15	1.5246-1.5365	5.10E-04	0.39	n.a.	10	Test
63	1,3-dimethylimidazolium methyl sulfate	283.15-343.15	1.4703-1.4866	9.96E-04	0.17	0.43	13	Train
64	1-n-butyl-2,3-dimethylimidazolium tetrafluoroborate	298.15-298.15	1.4330	2.00E-03	0.28	0.93	1	Test
65	1-methyl-3-propylimidazolium bis[(trifluoromethyl)sulfonyl]amide	293.15-343.15	1.4119-1.4267	3.70E-04	0.06	0.20	11	Train
66	1-propyronitrile-3-octylimidazolium bromide	298.15-333.15	1.506-1.5147	3.70E-04	0.68	n.a.	8	Train
67	1-methyl-3-octylimidazolium chloride	288.15-343.15	1.4936-1.5089	5.00E-04	0.21	0.52	12	Train
68	1-methyl-3-pentylimidazolium tetrafluoroborate	298.15-338.15	1.4153-1.4238	4.00E-04	0.24	0.80	9	Train
69	1-propyronitrile-3-butylimidazolium bromide	298.15-333.15	1.5375-1.5454	3.70E-04	0.04	n.a.	8	Train
70	3-(2-cyanoethyl)-1-hexylimidazolium trifluoromethanesulfonate	298.15-333.15	1.5223-1.5297	8.11E-04	2.92	n.a.	8	Train
71	3-(2-cyanoethyl)-1-hexylimidazolium 1,4-bis(2-ethylhexyl) sulfobutanedioate	298.15-333.15	1.4703-1.4797	7.61E-04	0.15	n.a.	8	Train
72	3-(2-cyanoethyl)-1-hexylimidazolium 3-sulfobenzoate	298.15-333.15	1.515-1.523	7.65E-04	0.07	n.a.	8	Train
73	3-(2-cyanoethyl)-1-hexylimidazolium benzenesulfonate	298.15-333.15	1.519-1.527	6.79E-04	0.21	n.a.	8	Test
74	3-(2-cyanoethyl)-1-hexylimidazolium dodecyl sulfate	298.15-333.15	1.4742-1.4841	7.30E-04	0.86	n.a.	8	Train
75	1-propyronitrile-3-hexylimidazolium bromide	298.15-333.15	1.5212-1.5287	3.70E-04	1.10	n.a.	8	Train
76	1-butyl-1-ethylpyrrolidinium ethylsulfate	328.15-343.15	1.4632-1.4671	1.08E-03	0.16	n.a.	4	Train
77	1-ethyl-1-methylpyrrolidinium ethylsulfate	308.15-343.15	1.4612-1.4702	1.08E-03	0.27	n.a.	8	Train
78	1-butyl-1-methylpyrrolidinium bis[(trifluoromethyl)sulfonyl]amide	283.15-343.15	1.4102-1.4272	3.70E-04	0.09	n.a.	13	Train
79	1-butyl-1-methylpyrrolidinium dicyanamide	288.15-308.15	1.4939-1.4997	5.20E-04	0.44	n.a.	3	Train
80	1-butyl-1-methylpyrrolidinium methyl sulfate	298.15-343.15	1.4614-1.4731	1.08E-03	0.22	n.a.	10	Train
81	1-butyl-1-methylpyrrolidinium trifluoromethanesulfonate	288.15-338.15	1.422-1.4354	3.70E-04	0.60	n.a.	6	Train

No.	Compound	T / K range	Refractive Index	Average Uncertainty	AARD / % GC Model	AARD / % Gardas and Coutinho [63]	Data Points	Subset
82	1-methyl-1-propylpyrrolidinium bis[(trifluoromethyl)sulfonyl]amide	283-353	1.4059-1.4247	5.71E-04	0.07	n.a.	15	Test
83	1-ethylpyridinium ethylsulfate	298.15-343.15	1.4931-1.5053	4.57E-04	0.04	n.a.	10	Train
84	1-methylpyridinium methylsulfate	288.15-308.15	1.5108-1.516	3.80E-04	0.49	n.a.	3	Train
85	1-butyl-2-methylpyridinium tetrafluoroborate	298.15-318.15	1.4500-1.4545	5.00E-04	0.57	n.a.	2	Train
86	1-ethyl-3-methylpyridinium ethylsulfate	298.15-343.15	1.4936-1.5067	7.33E-04	0.07	n.a.	10	Train
87	N-octyl-3-methylpyridinium tetrafluoroborate	283.15-328.15	1.4469-1.4598	1.30E-03	0.60	n.a.	19	Test
88	1-butyl-3-methylpyridinium trifluoromethanesulfonate	288.15-308.15	1.4587-1.4645	3.80E-04	0.61	n.a.	3	Train
89	N-butyl-4-methylpyridinium tetrafluoroborate	298.15-318.15	1.447-1.4517	5.00E-04	0.37	n.a.	2	Train
90	1-propylpyridinium tetrafluoroborate	283.15-338.15	1.434-1.4486	5.07E-04	0.15	n.a.	23	Train
91	1-butylpyridinium bis[(trifluoromethyl)sulfonyl]amide	303.15-353.15	1.4269-1.4415	1.10E-03	0.51	n.a.	6	Train
92	1-butylpyridinium tetrafluoroborate	303.15-353.15	1.4315-1.4444	1.10E-03	0.08	n.a.	6	Test
93	trihexyltetradecylphosphonium dicyanamide	283.15-343.15	1.4685-1.4883	3.70E-04	0.99	n.a.	13	Test
94	triethylmethylammonium methylsulfate	308.15-343.15	1.4528-1.4619	1.08E-03	0.83	n.a.	8	Test
95	triethyloctylammonium bis[(trifluoromethyl)sulfonyl]amide	298.15-298.15	1.4287	7.00E-04	0.53	n.a.	1	Train
96	triethylheptylammonium bis[(trifluoromethyl)sulfonyl]amide	298.15-298.15	1.4271	7.00E-04	0.41	n.a.	1	Train
97	triethylhexylammonium bis[(trifluoromethyl)sulfonyl]amide	298.15-298.15	1.4260	7.00E-04	0.34	n.a.	1	Train

Table 6.20: The AARD% of equation (6.4) for different families of ionic liquids.

No.	Family	T (K) range	n_D range	AARD%	No. of ILs	No. of Data Points
1	Amino Acid	288.15-343.15	1.4336-1.4648	0.17	10	107
2	Ammonium	298.15-343.15	1.426-1.4619	0.72	4	11
3	Imidazolium	283-363.15	1.3551-1.5778	0.37	65	657
4	Phosphonium	283.15-343.15	1.4685-1.4883	0.99	1	13
5	Pyridinium	283.15-353.15	1.4269-1.516	0.29	10	84
6	Pyrrolidinium	283-353	1.4059-1.4997	0.20	7	59

6.4.2 The QSPR model*

Same as the GC model, it was found that a linear summation of functional groups could produce a precise and easy-to-use model. In addition, the n_D of ILs showed a linear dependency with regard to the temperature. Hence, the resultant model was a 8-parameter linear model as shown in equations (6.5).

6.5

$$\ln k_p = \ln A + \ln \left(\frac{1}{T} \right)^n$$

$$\ln A = 1.3278 + 4.3161 \times 10^{-4} R_{2P+} + 3.4416 \times 10^{-4} H_{6m} - 2.4714 \times 10^{-4} HATS_m - 2.6988 \times 10^{-4} Mor_{13m}$$

$$\ln \left(\frac{1}{T} \right)^n = 7.0342 \times 10^{-4} - 1.1987 \times 10^{-4} C_{IC1} - 1.5328 \times 10^{-4} B_{03[C-C]} - 6.0483 \times 10^{-4} + 1.6109 \times 10^{-4} [\ln T - \ln T_0]$$

where T is the absolute temperature. The other parameters are described in Table 6.21. The full description of all parameters are also available in Molecular Descriptors for Chemoinformatics [240].

Table 6.21: Definition of the descriptors used in equation (6.5).

No.	Symbol	Definition
1	AMW	average molecular weight
2	R_{2P+}	R maximal autocorrelation of lag 2 / weighted by atomic polarizabilities
3	H_{6m}	H autocorrelation of lag 6 / weighted by atomic masses
4	$HATS_m$	leverage-weighted total index / weighted by atomic masses
5	SIC_0	structural information content (neighborhood symmetry of 0-order)
6	Mor_{13m}	3D-MoRSE - signal 13 / weighted by atomic masses
7	C_{IC1}	complementary information content (neighborhood symmetry of 1-order)
8	$B_{03[C-C]}$	presence/absence of C-(A) ₂ -C substructure which its value is 0 or 1 and A represents any type of atoms.

* The results have been published in *Journal of the Taiwan Institute of Chemical Engineers, In Press (2014)*

The AARD% of the proposed model is 0.47% for the training set and 0.60% for the test set. The coefficient of determination (R^2) for the training and test sets are 0.934 and 0.938, respectively. A summary of the statistical parameters of the model for the training and test sets is listed in Table 6.22.

Table 6.22: Statistical parameters for equation (6.5).

Statistical Parameter	
training set	
R^2	0.934
Absolute average relative deviation	0.47
Standard deviation error	1.02E-02
Root mean square error	1.02E-02
No. of data points	678
test set	
R^2	0.938
Absolute average relative deviation	0.60
Standard deviation error	1.19E-02
Root mean square error	1.19E-02
No. of data points	253
total	
R^2	0.935
Absolute average relative deviation	0.51
Standard deviation error	1.07E-02
Root mean square error	1.07E-02
No. of data points	931

The values of the predicted n_D versus the experimental data are shown in Figure 6.22. In addition, the deviation of the model from the experimental data is presented in Figure 6.23. It can be seen that almost all of the ILs can be predicted fairly well by the proposed model. Figure 6.24 illustrates that the AARD% of all ILs are less than 3%, except for one ionic liquid in the training set. For the training set, which has been used to develop the model, 87% of data points are within an AARD% of 0-1%; 12% of the data points are within the range of 1-2%, and 1% of data points, which belong to only one ionic liquid, shows an AARD% of greater than 2%. Similarly according to Figure 6.24, for test set, 80% of data points are within an AARD% of 0-1%; nearly 18% within the range of 1-2%; and nearly 2% of data points greater than 2% which is the predicted data of only one ionic liquid.

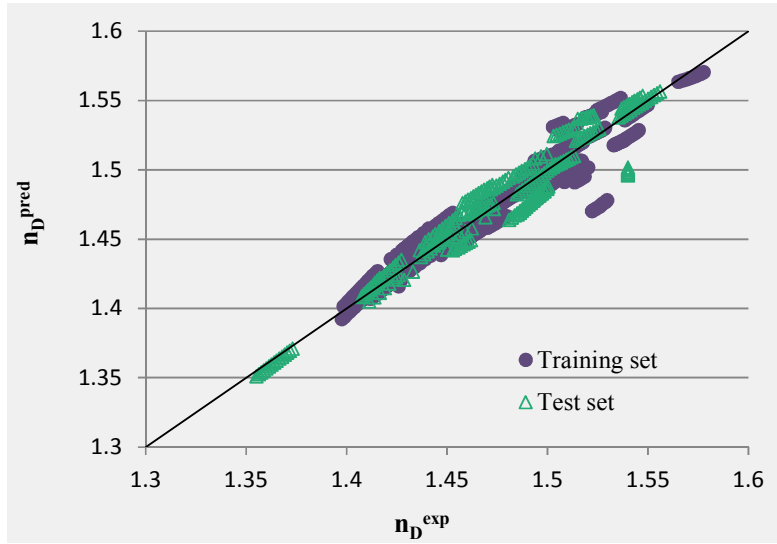


Figure 6.22: Predicted versus experimental values of n_D (— diagonal line).

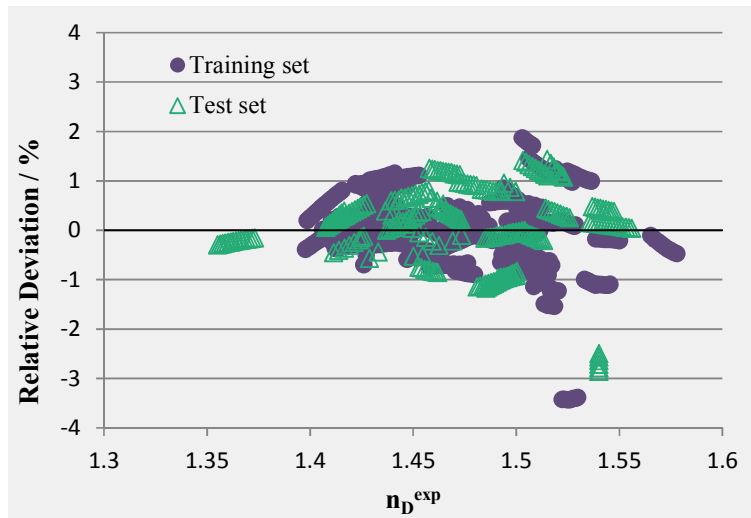


Figure 6.23: Relative deviation of predicted n_D from experimental data.

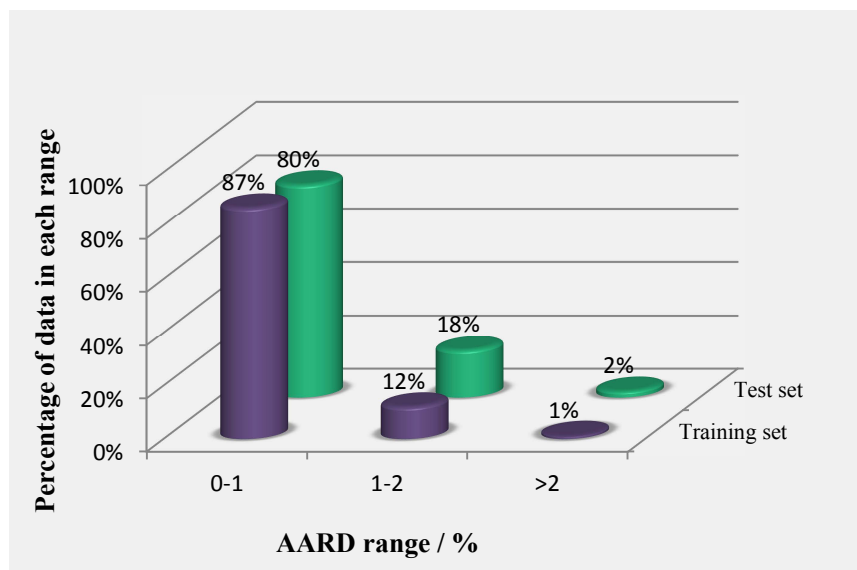


Figure 6.24: Percentage of calculated/predicted values of n_D in different relative deviation ranges

As indicated in Table 6.24, the largest deviation is observed for “3-(2-cyanoethyl)-1-hexylimidazolium trifluoromethanesulfonate” which is 3.41%. In previous section, it was explained that the data for this IL was published by Ziyada *et al.* [238] with a reported uncertainty in the measurement of 0.00081 (0.05%). The possible reason for the deviation is the purity of synthesized ionic liquid which the reported value is 96.4%. The effect of impurity can be seen in other ionic liquids containing “2-cyanoethyl” side chains in their cations, such as “1-propyronitrile-3-decylimidazolium bromide”, and “1-propyronitrile-3-octylimidazolium bromide” which have the third and fifth largest deviations (1.79% and 1.31%) respectively.

The second largest deviation (2.68%) was observed for 1-butyl-3-methylimidazolium bromide and surprisingly, the reported n_D over the temperature range of 298.3-323 K is constant at 1.54. These data have been published by Kim *et al.*[239] with an uncertainty of 0.01, which is the largest uncertainty in comparison with the reported values of other ILs. As a result, these data are not precise enough and that is the probable reason of the observed deviation in the test set.

For the purpose of comparison of this model with previous ones, it was mentioned in the last section that previous models are only applicable for imidazolium-based ionic liquids, but the presented model covers various families of ionic liquids. In previous section it has been

explained that the model proposed by Gardas and Coutinho [63] can calculate the refractive index for only 21 ionic liquids used in this study and the AARD% is 0.28%. The QSPR model presented in this study has the AARD% of 0.35% over these ILs which is not a large difference. As a result, the average accuracy of both models is close to each other, but the QSPR model is more comprehensive.

A comparison between GC and QSPR models developed in this study shows that the QSPR model (equation 6.5) has 8 parameters less than the GC one (equation 6.4) at similar accuracy. It is however a smaller and simpler model, but it requires special software for calculation of descriptors which may limit its usage if one does not have access to such software. The summary of different models for refractive index of ILs is shown in Table 6.23.

Table 6.23: Summary of different models for refractive index of ILs.

Model	Model Type and parameters	N_{ILs}	N_{data}	AARD%
Deetlefs <i>et al.</i> [71]	Correlation	9	9	6.4
Gardas and Coutinho [63]	GC, 10 parameters	24	245	0.18
GC Model (equation 6.4)	17 GCs	97	931	0.34
QSPR Model (equation 6.5)	8 descriptors	97	931	0.51

Table 6.24: Name and AARD% of ionic liquids studied for developing equation (6.5)

No.	Compound	T / K range	Refractive Index	Average Uncertainty	AARD / % QSPR Model	AARD / % Gardas and Coutinho [63]	No. of Data Points	Subset
1	L-glutamic acid, 1,5-bis(1-methylethyl) ester, dodecyl sulfate	288.15-343.15	1.4392-1.4575	1.10E-03	0.69	n.a.	12	Train
2	L-glutamic acid, 1,5-bis(2-methylpropyl) ester, dodecyl sulfate	323.15-343.15	1.4393-1.4472	1.10E-03	0.94	n.a.	5	Test
3	glycine, 1-methylethyl ester, dodecyl sulfate	303.15-343.15	1.4414-1.463	1.10E-03	0.37	n.a.	9	Train
4	glycine, 2-methylpropyl ester, dodecyl sulfate	303.15-343.15	1.4365-1.4624	1.10E-03	0.20	n.a.	9	Test
5	L-valine, 1-methylethyl ester, dodecyl sulfate	288.15-343.15	1.4388-1.4555	1.10E-03	0.06	n.a.	12	Train
6	L-valine, 2-methylpropyl ester, dodecyl sulfate	288.15-343.15	1.439-1.4568	1.10E-03	0.27	n.a.	12	Train
7	L-proline, 1-methylethyl ester, dodecyl sulfate	288.15-343.15	1.4398-1.4575	1.10E-03	0.08	n.a.	12	Train
8	L-proline, 2-methylpropyl ester, dodecyl sulfate	288.15-343.15	1.4378-1.4648	1.10E-03	0.19	n.a.	12	Train
9	L-alanine, 1-methylethyl ester, dodecyl sulfate	288.15-343.15	1.4371-1.4542	1.10E-03	0.19	n.a.	12	Train
10	L-alanine, 2-methylpropyl ester, dodecyl sulfate	288.15-343.15	1.4336-1.456	1.10E-03	0.09	n.a.	12	Train
11	1-(2-cyanoethyl)-3-(2-propen-1-yl)-imidazolium chloride	293.15-333.15	1.5368-1.5473	4.30E-04	0.43	n.a.	9	Train
12	1-butyl-3-(2-cyanoethyl)-imidazolium chloride	293.15-333.15	1.5142-1.5257	6.20E-04	0.35	n.a.	9	Train
13	1-(2-cyanoethyl)-3-(2-hydroxyethyl)-imidazolium chloride	293.15-333.15	1.5385-1.55	5.20E-04	0.19	n.a.	9	Train
14	3-(6-hydroxyhexyl)-imidazolium chloride	293.15-333.15	1.4989-1.5093	5.00E-04	0.06	n.a.	9	Train
15	1-ethyl-3-(6-hydroxyhexyl)-imidazolium chloride	293.15-333.15	1.5011-1.5115	5.00E-04	0.07	n.a.	9	Train
16	1-butyl-3-(6-hydroxyhexyl)-imidazolium chloride	293.15-333.15	1.5028-1.5132	5.00E-04	0.11	n.a.	9	Train
17	1-propyronitrile-3-decylimidazolium bromide	308.15-333.15	1.5028-1.5079	3.70E-04	1.79	n.a.	6	Train
18	1-butyl-3-ethylimidazolium trifluoromethanesulfonate	288.15-338.15	1.4279-1.4424	5.10E-04	0.47	n.a.	6	Train
19	3-(6-hydroxyhexyl)-1-methylimidazolium chloride	293.15-333.15	1.5002-1.5106	5.00E-04	0.36	n.a.	9	Train
20	1-ethyl-3-methylimidazolium 1,1,2,2-tetrafluoroethanesulfonate	293.15-353.15	1.4251-1.441	4.70E-04	1.04	n.a.	7	Train
21	1-ethyl-3-methylimidazolium acetate	283.15-353.15	1.4848-1.504	6.80E-04	0.08	n.a.	15	Train
22	1-ethyl-3-methylimidazolium ammonioacetate	293.15-333.15	1.4932-1.5063	6.80E-04	0.26	n.a.	9	Train
23	1-ethyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide	283-353	1.4076-1.4274	5.80E-04	0.31	0.28	15	Train
24	1-ethyl-3-methylimidazolium dicyanamide	283-353	1.4956-1.5177	5.97E-04	0.17	n.a.	15	Train
25	1-ethyl-3-methylimidazolium diethylphosphate	298.2-313.2	1.4691-1.4733	6.00E-04	0.16	n.a.	2	Train
26	1-ethyl-3-methylimidazolium dimethylphosphate	298.19-327.64	1.4738-1.4817	3.70E-04	0.25	n.a.	7	Train

No.	Compound	T / K range	Refractive Index	Average Uncertainty	AARD / % QSPR Model	AARD / % Gardas and Coutinho [63]	No. of Data Points	Subset
27	1-ethyl-3-methylimidazolium ethyl sulfate	283-353	1.4647-1.4832	5.71E-04	0.19	0.79	15	Train
28	1-ethyl-3-methylimidazolium hydrogen sulfate	283-353	1.4851-1.5004	5.50E-04	1.00	n.a.	15	Test
29	1-ethyl-3-methylimidazolium imidodisulfuryl fluoride	283-353	1.4321-1.4518	5.79E-04	0.10	n.a.	15	Test
30	1-ethyl-3-methylimidazolium L-alanine	293.15-333.15	1.4923-1.5059	6.80E-04	0.67	n.a.	9	Train
31	1-ethyl-3-methylimidazolium L-proline	293.15-333.15	1.5031-1.5152	6.80E-04	1.25	n.a.	9	Train
32	1-ethyl-3-methylimidazolium L-serine	293.15-333.15	1.5008-1.512	6.80E-04	0.82	n.a.	9	Train
33	1-ethyl-3-methylimidazolium methanesulfonate	283-353	1.481-1.4999	5.15E-04	1.00	n.a.	21	Train
34	1-ethyl-3-methylimidazolium methyl phosphonate	288.37-317.82	1.4871-1.4953	3.70E-04	0.58	n.a.	7	Train
35	1-ethyl-3-methylimidazolium tetracyanoborate	283-353	1.4292-1.4528	6.08E-04	1.03	n.a.	15	Train
36	1-ethyl-3-methylimidazolium tetrafluoroborate	283-353	1.3986-1.4155	5.60E-04	0.51	0.09	15	Train
37	1-ethyl-3-methylimidazolium thiocyanate	283-353	1.5355-1.556	5.85E-04	0.10	n.a.	15	Train
38	1-ethyl-3-methylimidazolium tosylate	322.73-342.37	1.533-1.5384	3.70E-04	1.05	n.a.	10	Train
39	1-ethyl-3-methylimidazolium tricyanomethide	283.15-313.15	1.5084-1.5154	1.10E-03	1.03	n.a.	4	Train
40	1-ethyl-3-methylimidazolium trifluoromethanesulfonate	288.15-338.15	1.4222-1.436	4.40E-04	1.02	0.13	12	Train
41	1-ethyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	283-353	1.3551-1.3731	5.69E-04	0.22	n.a.	15	Train
42	1-propyl-3-methylimidazolium methylsulfate	298.15-328.15	1.4664-1.4761	2.50E-03	0.45	0.48	4	Test
43	1-propyl-3-methylimidazolium tetrafluoroborate	298.15-338.15	1.4075-1.4165	4.00E-04	0.22	0.06	9	Train
44	1-octyl-3-methylimidazolium hexafluorophosphate	283.15-343.15	1.4106-1.4272	1.01E-03	0.24	0.05	13	Train
45	1-octyl-3-methylimidazolium tetrafluoroborate	283.15-363.15	1.4142-1.4363	8.00E-04	0.20	0.02	16	Train
46	1-butyl-3-methylimidazolium acetate	283.15-353.15	1.4718-1.4927	3.70E-04	0.87	n.a.	15	Test
47	1-butyl-3-methylimidazolium dicyanamide	288.15-308.15	1.5058-1.5121	5.20E-04	0.50	n.a.	3	Train
48	1-butyl-3-methylimidazolium glycine	298.15-313.15	1.5166-1.5202	2.00E-03	1.22	n.a.	4	Train
49	1-butyl-3-methylimidazolium hexafluorophosphate	283.15-343.15	1.3977-1.4133	1.01E-03	0.18	0.03	13	Train
50	1-butyl-3-methylimidazolium L-alanine acid salt	298.15-313.15	1.5135-1.5184	4.00E-04	1.52	n.a.	4	Train
51	1-butyl-3-methylimidazolium methylsulfate	283.15-343.15	1.4672-1.4835	9.52E-04	0.25	0.43	24	Train
52	1-butyl-3-methylimidazolium octylsulfate	288.15-343.15	1.4577-1.4725	5.00E-04	1.19	n.a.	12	Train
53	1-butyl-3-methylimidazolium perchlorate	283.15-353.15	1.4577-1.4763	1.20E-03	0.70	n.a.	15	Train

No.	Compound	T / K range	Refractive Index	Average Uncertainty	AARD / % QSPR Model	AARD / % Gardas and Coutinho [63]	No. of Data Points	Subset
54	1-butyl-3-methylimidazolium trifluoromethanesulfonate	288.15-338.15	1.4261-1.4401	6.98E-04	0.85	0.04	17	Train
55	1-butyl-3-methylimidazolium bromide	298.3-323	1.54	1.00E-02	2.68	n.a.	6	Test
56	1-butyl-3-methylimidazolium tetrafluoroborate	293.15-353.15	1.4114-1.4259	6.50E-04	0.28	0.41	8	Train
57	1-hexyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide	302.95-332.95	1.4238-1.4296	1.26E-03	0.21	0.10	7	Train
58	1-hexyl-3-methylimidazolium chloride	298.15-343.15	1.5045-1.5172	1.05E-03	0.56	0.61	10	Train
59	1-hexyl-3-methylimidazolium hexafluorophosphate	288.15-318.15	1.4122-1.4206	1.09E-03	0.33	0.11	7	Train
60	1-hexyl-3-methylimidazolium tetrafluoroborate	298.15-338.15	1.4179-1.427	4.00E-04	0.06	0.16	9	Train
61	1-Benzyl-3-methylimidazolium chloride	298.15-343.15	1.5652-1.5778	5.10E-04	0.32	n.a.	10	Train
62	1-benzyl-3-methylimidazolium methylsulfate	298.15-343.15	1.5246-1.5365	5.10E-04	1.09	n.a.	10	Test
63	1,3-dimethylimidazolium methyl sulfate	283.15-343.15	1.4703-1.4866	9.96E-04	0.09	0.43	13	Train
64	1-n-butyl-2,3-dimethylimidazolium tetrafluoroborate	298.15-298.15	1.4330	2.00E-03	0.44	0.93	1	Test
65	1-methyl-3-propylimidazolium bis[(trifluoromethyl)sulfonyl]amide	293.15-343.15	1.4119-1.4267	3.70E-04	0.25	0.20	11	Train
66	1-propyronitrile-3-octylimidazolium bromide	298.15-333.15	1.506-1.5147	3.70E-04	1.31	n.a.	8	Train
67	1-methyl-3-octylimidazolium chloride	288.15-343.15	1.4936-1.5089	5.00E-04	0.60	0.52	12	Train
68	1-methyl-3-pentylimidazolium tetrafluoroborate	298.15-338.15	1.4153-1.4238	4.00E-04	0.09	0.80	9	Train
69	1-propyronitrile-3-butylimidazolium bromide	298.15-333.15	1.5375-1.5454	3.70E-04	1.10	n.a.	8	Train
70	3-(2-cyanoethyl)-1-hexylimidazolium trifluoromethanesulfonate	298.15-333.15	1.5223-1.5297	8.11E-04	3.41	n.a.	8	Train
71	3-(2-cyanoethyl)-1-hexylimidazolium 1,4-bis(2-ethylhexyl) 2-sulfobutanedioate	298.15-333.15	1.4703-1.4797	7.61E-04	0.86	n.a.	8	Train
72	3-(2-cyanoethyl)-1-hexylimidazolium 3-sulfobenzoate	298.15-333.15	1.515-1.523	7.65E-04	1.22	n.a.	8	Train
73	3-(2-cyanoethyl)-1-hexylimidazolium benzenesulfonate	298.15-333.15	1.519-1.527	6.79E-04	1.10	n.a.	8	Test
74	3-(2-cyanoethyl)-1-hexylimidazolium dodecyl sulfate	298.15-333.15	1.4742-1.4841	7.30E-04	0.11	n.a.	8	Train
75	1-propyronitrile-3-hexylimidazolium bromide	298.15-333.15	1.5212-1.5287	3.70E-04	0.13	n.a.	8	Train
76	1-butyl-1-ethylpyrrolidinium ethylsulfate	328.15-343.15	1.4632-1.4671	1.08E-03	0.04	n.a.	4	Train
77	1-ethyl-1-methylpyrrolidinium ethylsulfate	308.15-343.15	1.4612-1.4702	1.08E-03	0.71	n.a.	8	Train
78	1-butyl-1-methylpyrrolidinium bis[(trifluoromethyl)sulfonyl]amide	283.15-343.15	1.4102-1.4272	3.70E-04	0.11	n.a.	13	Train
79	1-butyl-1-methylpyrrolidinium dicyanamide	288.15-308.15	1.4939-1.4997	5.20E-04	0.87	n.a.	3	Train

No.	Compound	T / K range	Refractive Index	Average Uncertainty	AARD / % QSPR Model	AARD / % Gardas and Coutinho [63]	No. of Data Points	Subset
80	1-butyl-1-methylpyrrolidinium methyl sulfate	298.15-343.15	1.4614-1.4731	1.08E-03	0.35	n.a.	10	Train
81	1-butyl-1-methylpyrrolidinium trifluoromethanesulfonate	288.15-338.15	1.422-1.4354	3.70E-04	0.79	n.a.	6	Train
82	1-methyl-1-propylpyrrolidinium bis[(trifluoromethyl)sulfonyl]amide	283-353	1.4059-1.4247	5.71E-04	0.03	n.a.	15	Test
83	1-ethylpyridinium ethylsulfate	298.15-343.15	1.4931-1.5053	4.57E-04	0.46	n.a.	10	Train
84	1-methylpyridinium methylsulfate	288.15-308.15	1.5108-1.516	3.80E-04	0.64	n.a.	3	Train
85	1-butyl-2-methylpyridinium tetrafluoroborate	298.15-318.15	1.4500-1.4545	5.00E-04	0.52	n.a.	2	Train
86	1-ethyl-3-methylpyridinium ethylsulfate	298.15-343.15	1.4936-1.5067	7.33E-04	0.37	n.a.	10	Train
87	N-octyl-3-methylpyridinium tetrafluoroborate	283.15-328.15	1.4469-1.4598	1.30E-03	0.45	n.a.	19	Test
88	1-butyl-3-methylpyridinium trifluoromethanesulfonate	288.15-308.15	1.4587-1.4645	3.80E-04	0.59	n.a.	3	Train
89	N-butyl-4-methylpyridinium tetrafluoroborate	298.15-318.15	1.447-1.4517	5.00E-04	0.57	n.a.	2	Train
90	1-propylpyridinium tetrafluoroborate	283.15-338.15	1.434-1.4486	5.07E-04	0.17	n.a.	23	Train
91	1-butylpyridinium bis[(trifluoromethyl)sulfonyl]amide	303.15-353.15	1.4269-1.4415	1.10E-03	0.52	n.a.	6	Train
92	1-butylpyridinium tetrafluoroborate	303.15-353.15	1.4315-1.4444	1.10E-03	0.27	n.a.	6	Test
93	triethyltetradecylphosphoniumdicyanamide	283.15-343.15	1.4685-1.4883	3.70E-04	0.09	n.a.	13	Test
94	triethylmethylammoniummethylsulfate	308.15-343.15	1.4528-1.4619	1.08E-03	0.81	n.a.	8	Test
95	triethyloctylammoniumbis[(trifluoromethyl)sulfonyl]amide	298.15-298.15	1.4287	7.00E-04	0.57	n.a.	1	Train
96	triethylheptylammoniumbis[(trifluoromethyl)sulfonyl]amide	298.15-298.15	1.4271	7.00E-04	0.50	n.a.	1	Train
97	triethylhexylammoniumbis[(trifluoromethyl)sulfonyl]amide	298.15-298.15	1.4260	7.00E-04	0.70	n.a.	1	Train

6.5 Viscosity of F-ILs

6.5.1 The GC model*

As discussed in section 5.5, the natural logarithm (\ln) was chosen to transform the data of the viscosity to a simpler form to increase the quality of regression. Moreover, the refined dataset of viscosity data with total 863 data points were used to develop the model, as discussed in section 3.4.4.

The modeling results revealed that the best model with the lowest possible number of variables was a 35-parameter GC model with a total $R^2 = 0.977$, as is shown in equation (6.6).

6.6

$$\ln(\eta) = A + B \left(\frac{1000}{T} \right) + C \left(\frac{1000}{T} \right)^2$$
$$A = \sum_{i=1}^k n_i a_i, B = \sum_{i=1}^k n_i b_i, C = \sum_{i=1}^k n_i c_i$$

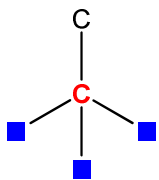
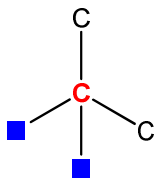
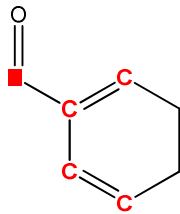
where T is the absolute temperature, n_i is the number of groups of type i , k is the total number of different types of groups, and the parameters a_i , b_i , and c_i are for different substructures. The model parameters are available in Table 6.25.

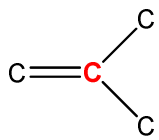
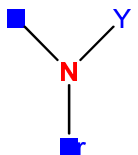
The calculated/predicted viscosities of fluorine-containing ILs used in this study and their families are available in Table 6.26 and Table 6.27, respectively. In addition, the cross-plot of experimental data versus calculated/predicted data and the relative deviation of $\frac{\eta_{\text{calc}} - \eta_{\text{exp}}}{\eta_{\text{exp}}}$ from experimental data are presented in Figure 6.25 and Figure 6.26, respectively.

To assess the performance and the accuracy of the GC model developed, statistical error analyses was undertaken, the results of which are summarized in Table 6.28. The results show that the total average absolute relative deviation (AARD%) of the model is 3.23% (2.91% for the “training” set and 4.31% for the “test” set). In addition, R^2 is 0.977 for all data points, 0.981 for the “training” set and 0.966 for the “test” set.

* The results have been submitted to *Journal of Fluorine Chemistry*.

Table 6.25: Parameters of the GC model in equation (6.6).

No.	Chemical structure	Descriptions	a_i	b_i	c_i
	Intercept		5.05776		
1	$nR06_{Cat}$	number of 6-membered rings	0.15706		
2	nCp_{Cat}	number of terminal primary C(sp ³)	-0.34962		
		Y = any terminal atom or heteroatom			
3	$nCrS_{Cat}$	number of total secondary C(sp ³) Y = H or any heteroatom	0.26051		
					
4	$nCconj_{Cat}$	number of non-aromatic conjugated C(sp ²)	-1.81670		
					
5	$(R-C(=X)-X / R-C\equiv X / X=C=X)_{Cat}$	X represents any heteroatom (O, N, S, P, Se, halogens)	-3.58447	0.69161	
6	$[C-(A)_3-C]_{Cat}$	A represents any atom	-0.08890		
7	$[C-(A)_3-N]_{Cat}$		-0.21752		
8	$[C-(A)_5-C]_{Cat}$		0.02554		
9	$[C-(A)_5-N]_{Cat}$		-0.15649		
10	$[C-(A)_8-P]_{Cat}$		-5.34117		
11	$(R-C(=X)-X / R-C\equiv X / X=C=X)_{An}$		-0.39713		

No.	Chemical structure	Descriptions	a_i	b_i	c_i
12	$nR=Ct_{Cat}$	number of aliphatic tertiary C(sp ²)		2.11721	
					
13	nN^+_{Cat}	number of positive charged N		-4.57101	
14	$(CH3R / CH4)_{Cat}$			0.32222	
15	$CH3X_{Cat}$			0.16892	
16	$[C-(A)_4-N]_{Cat}$			0.04905	
17	$[C-(A)_4-O]_{Cat}$			-0.14875	
18	$[C-(A)_5-O]_{Cat}$			0.14203	
19	nP_{An}	number of Phosphorous atoms		-0.42847	
20	$[C-(A)_2-O]_{An}$			0.03122	
21	nO_{Cat}	number of Oxygen atoms			0.17796
22	$nArNR2_{Cat}$	number of tertiary amines (aromatic) Y = Aromatic or Aliphatic (not C = O)			-0.94346
					
23	$CH2R2_{Cat}$				0.02528
24	$R-CH-R_{Cat}$				0.05543
25	$(R-N-R / R-N-X)_{Cat}$				0.20778
26	$[C-N]_{Cat}$				0.28852
27	$[N-A-N]_{Cat}$				-0.75415
28	$[C-(A)_2-N]_{Cat}$				0.01301

No.	Chemical structure	Descriptions	a_i	b_i	c_i
29	[N-(A) ₂ -O] _{Cat}				0.57373
30	[C-(A) ₃ -O] _{Cat}				-0.21222
31	[C-(A) ₄ -C] _{Cat}				0.00218
32	[C-(A) ₈ -C] _{Cat}				-0.00526
33	[C-(A) ₈ -N] _{Cat}				0.02693
34	[P-F] _{An}				0.03452
35	[S-(A) ₃ -F] _{An}				-0.01262

R represents any group linked through carbon.

X represents any heteroatom (O, N, S, P, Se, halogens).

-- represents an aromatic bond as in benzene or delocalized bonds such as the N-O bond in a nitro group.

Table 6.26: Name and AARD% of ionic liquids studied.

No.	Compound	T (K) range	ln(η_{exp}) (cP)	AARD%	N Data	Subset
1	2-[2-hydroxyethyl(methyl)amino]ethanol tetrafluoroborate	303.15-343.15	3.477-5.341	0.29	5	Train
2	3-hexyl-1-methylimidazolium 1,1,2,2,2-pentafluoro-N-[(1,1,2,2,2-pentafluoroethyl)sulfonyl]ethanesulfonamide	293	5.124	2.97	1	Train
3	1-(2-hydroxyethyl)-3-methylimidazolium tetrafluoroborate	293.15-328.15	5.05-7.349	1.65	5	Train
4	1-decyl-3-methylimidazolium tetrafluoroborate	283-363	3.311-7.636	1.75	9	Train
5	1-decyl-3-methylimidazolium hexafluorophosphate	313-363	3.78-6.111	3.24	6	Train
6	1-decyl-3-methylimidazolium trifluoromethanesulfate	283-363	3.572-7.63	1.52	9	Test
7	1-decyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]imide	298.15-343.15	2.901-4.684	21.06	3	Test
8	1-decyl-3-methylimidazolium bis(perfluoroethylsulfonyl)imide	293	5.342	25.12	1	Train
9	1-undecyl-3-methylimidazolium tetrafluoroborate	293-363	3.818-7.641	6.33	8	Test
10	1-dodecyl-3-methylimidazolium hexafluorophosphate	333-363	3.932-5.366	2.56	4	Train
11	1-(2-ethylsulfonyl)ethyl-3-methyl-imidazolium trifluoromethanesulfonate	303	6.317	6.59	1	Train
12	1-ethyl-3-methylimidazolium tetrafluoroborate	269-363	1.916-4.977	19.88	21	Train
13	1-ethyl-3-methylimidazolium trifluoromethanesulfonate	278.15-363.15	2.1-4.603	11.53	18	Test
14	1-ethyl-3-methylimidazolium hexafluorophosphate	343-363	2.588-3.153	3.14	3	Train

No.	Compound	T (K) range	ln(η_{exp}) (cP)	AARD%	N Data	Subset
15	1-ethyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]imide	263.1-388.19	1.411-5.165	5.32	23	Train
16	1-ethyl-3-methylimidazolium 1,1,2,2-tetrafluoroethanesulfonate	293.15-353.15	2.691-4.766	1.86	7	Train
17	1-isobutenyl-3-methylimidazolium tetrafluoroborate	288.15-313.15	3.978-5.367	0.65	10	Train
18	1-propyl-3-methylimidazolium tetrafluoroborate	298	4.635	2.46	1	Test
19	1-octyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]imide	283.15-353.15	2.555-5.384	7.95	8	Train
20	1-octyl-3-methylimidazolium tetrafluoroborate	283.15-363.15	3.071-6.944	1.79	17	Train
21	1-octyl-3-methylimidazolium hexafluorophosphate	283-363	3.434-7.522	2.16	9	Train
22	1-octyl-3-methylimidazolium bis(perfluoroethylsulfonyl)imide	293	5.293	11.82	1	Train
23	1-octyl-3-methylimidazolium trifluoromethanesulfonate	283-363	3.27-6.862	3.83	9	Train
24	1-octyl-3-methylimidazolium 1,1,2,2-tetrafluoroethanesulfonate	298.15-373.15	2.728-5.633	8.62	4	Test
25	1-butyl-3-methylimidazolium hexafluorophosphate	283.15-353.15	3.249-6.687	1.26	13	Train
26	1-butyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]imide	293.4-387.51	1.435-4.091	3.05	11	Train
27	1-butyl-3-methylimidazolium trifluoroacetate	283-353	2.332-5.256	0.70	9	Train
28	1-butyl-3-methylimidazolium bis(perfluoroethylsulfonyl)imide	283-353	2.646-5.675	2.31	9	Train
29	1-butyl-3-methylimidazolium tetrafluoroborate	273-388.04	1.758-6.236	2.91	23	Train
30	1-hexyl-3-methylimidazolium tetrafluoroborate	288.15-338.15	3.615-5.755	2.12	11	Train
31	1-hexyl-3-methylimidazolium hexafluorophosphate	283-363	3.288-7.268	4.55	9	Train
32	1-hexyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]imide	258.15-373.15	2.01-6.848	2.32	17	Test
33	1-hexyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	293.15-343.15	2.695-4.739	7.60	11	Test
34	1,3-dimethylimidazolium bis[(trifluoromethyl)sulfonyl]imide	283.15-353.15	2.092-4.345	2.16	15	Train
35	1-butyl-2,3-dimethylimidazolium bis[(trifluoromethyl)sulfonyl]imide	273.15-373.15	2.176-6.251	1.48	21	Train
36	1-butyl-2,3-dimethylimidazolium tris(pentafluoroethyl)trifluorophosphate	283.15-373.15	2.309-6.078	1.55	19	Train
37	1-hexyl-2,3-dimethylimidazolium bis[(trifluoromethyl)sulfonyl]imide	283-343	3.091-5.759	2.22	8	Train
38	1-methyl-3-propylimidazolium bis[(trifluoromethyl)sulfonyl]imide	298.15-343.15	2.366-3.777	2.57	10	Train
39	1,2-dimethyl-3-propylimidazolium bis[(trifluoromethyl)sulfonyl]imide	290-365	2.944-4.875	6.19	16	Train
40	3-(2-cyanoethyl)-1-hexylimidazolium trifluoromethanesulfonate	293.15-353.15	3.381-7.716	1.27	13	Train
41	1-(2-methoxyethyl)-1-methyl-pyrrolidinium bis[(trifluoromethyl)sulfonyl]imide	258.15-373.15	1.983-6.43	2.36	24	Train
42	1-(2-methoxyethyl)-1-methyl-pyrrolidinium tris(pentafluoroethyl)trifluorophosphate	258.15-373.15	2.286-7.733	2.06	24	Test

No.	Compound	T (K) range	ln(η_{exp}) (cP)	AARD%	N Data	Subset
43	1-butyl-1-methylpyrrolidinium bis[(trifluoromethyl)sulfonyl]imide	278.15-373.15	2.135-5.439	4.13	20	Train
44	1-butyl-1-methylpyrrolidinium tris(pentafluoroethyl)trifluorophosphate	283.15-373.15	2.568-6.386	5.11	19	Test
45	1-methyl-1-propylpyrrolidinium bis[(trifluoromethyl)sulfonyl]imide	298	4.143	3.47	1	Test
46	1-(2-ethoxy-2-oxoethyl)-pyridinium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	293.15-343.15	3.627-6.667	1.49	6	Train
47	1-hexyl-4-(4-methyl-1-piperidiny)pyridinium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	283-343	2.708-5.652	1.09	8	Test
48	1-(3-hydroxypropyl)pyridiniumbis[(trifluoromethyl)sulfonyl]imide	293.15-343.15	3.127-5.075	1.61	6	Train
49	1-hexyl-3-methyl-4-(dimethylamino)pyridiniumbis[(trifluoromethyl)sulfonyl]imide	283-343	2.944-5.628	0.44	8	Train
50	1-butyl-2-methylpyridinium tetrafluoroborate	288-328	4.041-6.893	5.56	7	Train
51	1-octyl-3-methylpyridinium bis[(trifluoromethyl)sulfonyl]imide	283-343	2.944-5.591	3.74	8	Test
52	1-butyl-3-methylpyridinium bis[(trifluoromethyl)sulfonyl]imide	278.15-363.15	2.138-5.275	1.13	18	Train
53	1-butyl-3-methylpyridinium tetrafluoroborate	283-343	3.135-6.248	2.37	8	Test
54	1-hexyl-3-methylpyridinium bis[(trifluoromethyl)sulfonyl]imide	283-343	2.773-5.283	3.01	8	Train
55	1-butyl-4-methylpyridinium bis[(trifluoromethyl)sulfonyl]imide	278.15-363.15	2.038-5.135	0.58	18	Train
56	N-butyl-4-methylpyridinium tetrafluoroborate	298.15-373.15	2.351-5.325	3.99	4	Train
57	1-hexyl-2-ethyl-3,5-dimethylpyridinium bis[(trifluoromethyl)sulfonyl]imide	283-343	3.367-6.562	2.99	8	Train
58	1-hexyl-3,5-dimethylpyridinium bis[(trifluoromethyl)sulfonyl]imide	283-343	2.89-5.525	1.92	8	Train
59	4-methyl-N-ethylpyridinium bis[(trifluoromethyl)sulfonyl]imide	298.15-363.15	1.823-3.489	1.17	14	Test
60	1-propylpyridinium tetrafluoroborate	278.15-338.15	3.09-6.134	2.20	25	Train
61	1-propylpyridinium bis[(trifluoromethyl)sulfonyl]imide	308.15-338.15	2.603-3.497	0.55	7	Test
62	3-methyl-1-propylpyridinium bis[(trifluoromethyl)sulfonyl]imide	278.15-363.15	2.072-5.074	1.20	18	Train
63	3,5-diethyl-1-hexyl-2-propylpyridinium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	283-343	3.296-6.361	1.84	8	Train
64	1-octylpyridinium bis[(trifluoromethyl)sulfonyl]imide	278.15-363.15	2.463-5.988	0.76	18	Train
65	1-butylpyridinium bis[(trifluoromethyl)sulfonyl]imide	278.15-363.15	2.151-5.203	1.58	18	Train
66	1-hexylpyridinium bis[(trifluoromethyl)sulfonyl]imide	278.15-363.15	2.367-5.775	1.94	18	Train
67	N-hexylquinoliniumbis[(trifluoromethyl)sulfonyl]imide	323.15-348.15	3.533-4.606	0.29	6	Train
68	2-(acetyloxy)-N,N,N-trimethylethanaminium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	293.15-343.15	3.314-5.733	0.77	6	Train
69	N-(2-hydroxyethyl)-N,N-dimethyl-1-propanaminium 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	293.15-343.15	3.131-5.075	1.18	6	Train
70	triethyl(tetradecyl)phosphoniumbis[(trifluoromethyl)sulfonyl]imide	278.15-358.15	3.293-7.128	6.05	17	Test

No.	Compound	T (K) range	ln(η_{exp}) (cP)	AARD%	N Data	Subset
71	triethyl(tetradecyl)phosphoniumtris(pentafluoroethyl)trifluorophosphate	268.15-373.15	2.866-8.029	2.24	23	Train
72	ethylheptyl-di-(1-methylethyl)ammonium bis[(trifluoromethyl)sulfonyl]imide	298	5.892	3.58	1	Train
73	heptyltrimethylammoniumbis[(trifluoromethyl)sulfonyl]imide	298	5.03	6.28	1	Train
74	butyltrimethylammoniumbis[(trifluoromethyl)sulfonyl]imide	283-388.51	1.887-5.567	4.71	12	Train
75	hexyltrimethylammoniumbis[(trifluoromethyl)sulfonyl]imide	298	5.03	1.39	1	Test
76	methyltrioctylammoniumtrifluoroacetate	298	7.443	0.54	1	Train
77	triethyloctylammoniumbis[(trifluoromethyl)sulfonyl]imide	293.05-368.15	2.61-5.664	0.98	17	Test
78	trimethyloctylammoniumbis[(trifluoromethyl)sulfonyl]imide	298	5.198	4.15	1	Train
79	tributylloctylammoniumbis[(trifluoromethyl)sulfonyl]imide	298	6.353	2.51	1	Train
80	tributylloctylammoniumtrifluoromethanesulfonate	298	7.616	3.29	1	Train
81	triethylheptylammoniumbis[(trifluoromethyl)sulfonyl]imide	293.25-368.25	2.54-5.536	1.31	16	Train
82	tributylheptylammoniumbis[(trifluoromethyl)sulfonyl]imide	298	6.407	0.58	1	Test
83	triethylhexylammoniumbis[(trifluoromethyl)sulfonyl]imide	293.25-368.15	2.548-5.457	0.58	16	Train
84	tributylhexylammoniumbis[(trifluoromethyl)sulfonyl]imide	298	6.389	0.19	1	Train
85	1-butyl-nicotinic acid butyl ester 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	283-343	3.871-7.512	1.28	8	Train

Table 6.27: The AARD% of equation (6.6) for different families of ionic liquids.

No.	Family	T (K) range	ln(η) (cP)	AARD%	N Compounds	N Data Points
1	Ammonium	283-388.51	1.887-7.616	1.60	15	87
2	Imidazolium	258.15-388.19	1.411-7.716	4.51	39	393
3	Phosphonium	268.15-373.15	2.866-8.027	3.86	2	40
4	Pyridinium	278.15-373.15	1.823-7.512	1.72	22	249
5	Pyrrolidinium	258.15-373.15	1.983-7.733	3.29	4	88
6	Quinolinium	323.15-348.15	3.533-4.606	0.29	1	6

Table 6.28: The statistical error parameters for $\ln(\eta)$ in equation (6.6).

<i>Statistical Parameter</i>	
<i>training set</i>	
R ²	0.981
Average absolute relative deviation	2.91
Standard deviation error	0.17
Root mean square error	0.18
No. of data points	667
<i>test set</i>	
R ²	0.966
Average absolute relative deviation	4.31
Standard deviation error	0.25
Root mean square error	0.25
No. of data points	196
<i>total</i>	
R ²	0.977
Average absolute relative deviation	3.23
Standard deviation error	0.19
Root mean square error	0.19
No. of data points	863

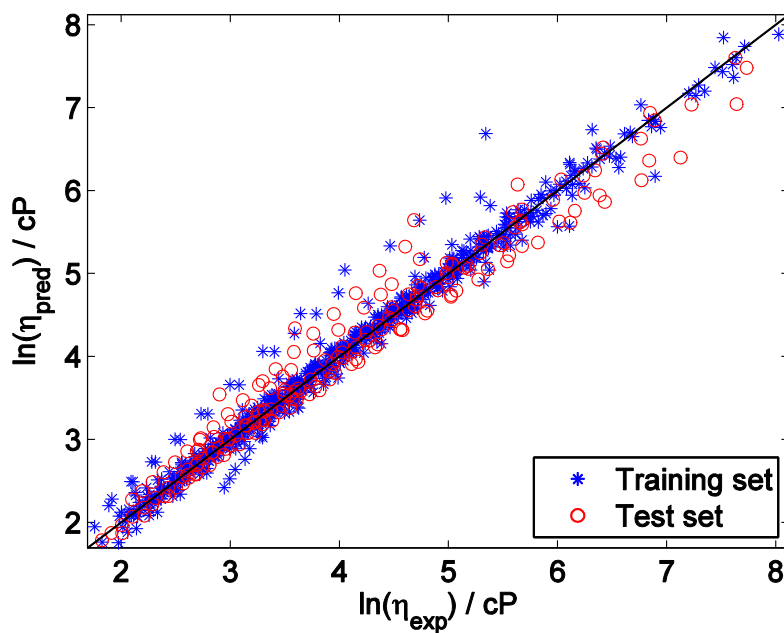


Figure 6.25: Predicted versus experimental values of $\ln(\eta)$ (refined database) (— diagonal line).

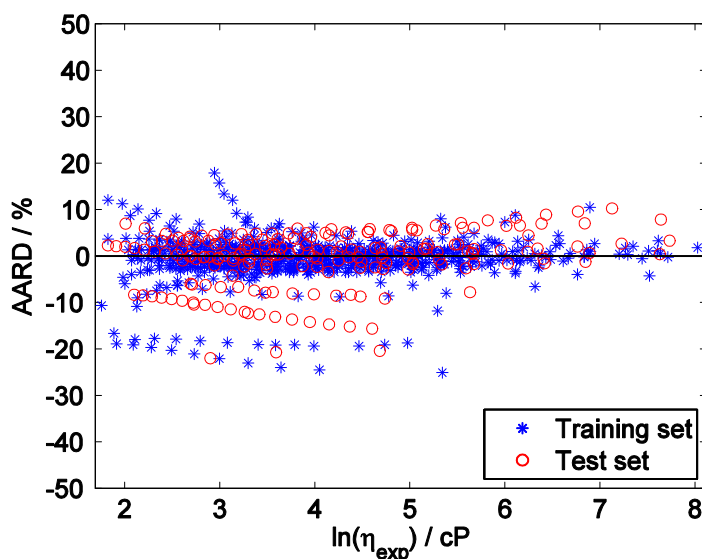


Figure 6.26: Relative deviation of predicted $\ln(\eta)$ from experimental data (refined database).

According to Figure 6.25, a tight cluster of points near the diagonal for the training and testing datasets illustrates the robustness of the proposed GC model for the prediction of the viscosity of fluorine-containing ILs. As indicated in Table 6.26, most of the data points with large deviations are for temperatures over 330 K or for very low viscosity values, for which a small deviation in predicted viscosity produces large AARD values.

Another possible reason for the observed deviations for the model prediction/correlation is due to unreliability in viscosity values of ILs which are not as pure as claimed. Seddon *et al.* [163] showed that the presence of contaminants in ionic liquids changes the viscosity drastically; for example contamination with water decreases the viscosity, while the presence of chloride ions increases the viscosity. That is one of the main reasons for the existence of multiple reported values of viscosity at a single temperature in the literature.

Table 6.28 shows the AARD% of each family of ionic liquids. The results indicate that the model can correlate and predict each family with a similar averaged accuracy and demonstrates the comprehensiveness of the model. The maximum deviation is observed in imidazolium (4.51%) ILs and the minimum in ammonium ILs (1.60%). It should be noted that the quinolinium family has the absolute minimum deviation (0.29%); but as there is just one ionic liquid in that family, it cannot be considered as a well-predicted class of ILs.

To have a better judgment about the model proposed, Figure 6.27 shows the “real” linear scale values of viscosity; however it is common to report just the accuracy of the model in logarithmic scale. By reversing the transformation function and applying the *exp* on the data, the “real” AARD% of the model became 13.31% which is very low compared with the previous models, as shown in Table 2.4.

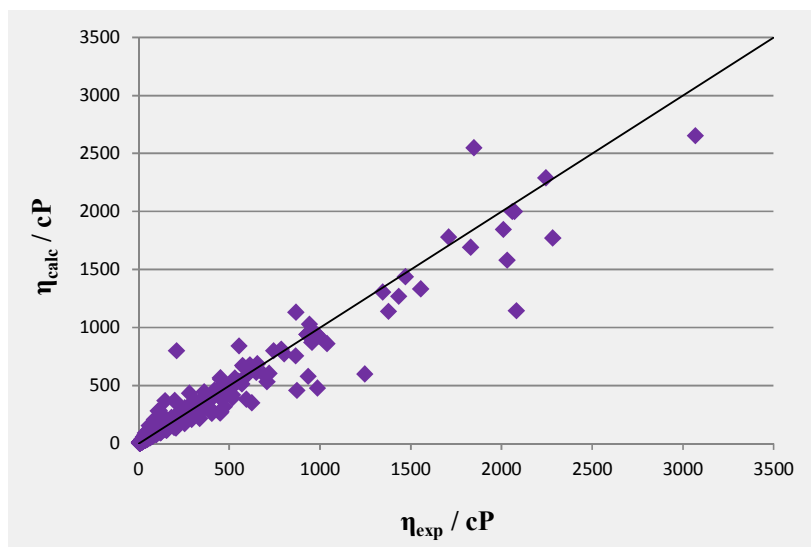


Figure 6.27: Predicted versus experimental values of η in linear scale (refined database) (— diagonal line).

As discussed in section 3.4.4, the viscosity data of 247 ILs with 297 data points were excluded from the development of the model proposed, because there were not any reported uncertainty values for those data points. To have a comprehensive but less accurate model, all available data points (1160 data points) were used to develop a new model. This new model had 36 parameters with an AARD% of 4.85% in logarithmic scale and 21.89% in linear scale. Figure 6.28 and Figure 6.29 illustrate the result of model developed in logarithmic and linear scale.

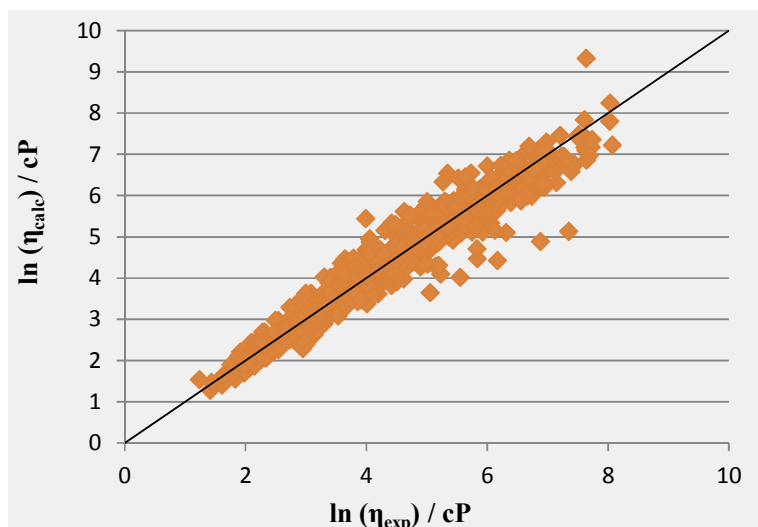


Figure 6.28: Predicted versus experimental values of $\ln(\eta)$ (entire database)
(— diagonal line).

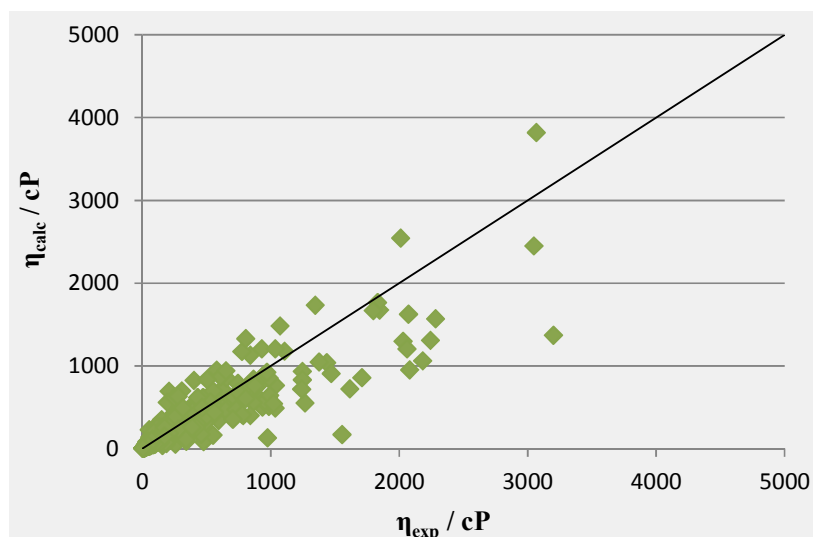


Figure 6.29: Predicted versus experimental values of η in linear scale (entire database)
(— diagonal line).

In comparison with the previous published models, these two models presented in this thesis have more accuracy and applicability for various ionic liquids. The first model which was developed based on less but refined data points, had an AARD% of 3.23% and 13.31% for $\ln(\eta)$ and η , respectively. Similarly, the second model, which was developed by some unreliable data points,

had an AARD% of 4.85% and 21.89% for $\ln(\eta)$ and η , respectively. According to the results, both can predict the viscosity of F-ILs more accurately compared with previous models.

All previous models were developed for few types of ionic liquids, except the model proposed by Gharagheizi *et al.* [91]. They used a dataset consisted of 1034 unique data points of which 724 were for F-ILs. By using only the data of F-ILs, that model showed an AARD% of 6.7% and 59.7% for $\ln(\eta)$ and η which the latter is a very large deviation and calls the applicability of this model for F-ILs into question. As a result, their model has lower accuracy compared with models presented in this thesis.

The comparison between models proposed and previous models is shown in Table 6.31, at the end of next section. In addition, information for the entire dataset, as well as the values of the functional groups of the ILs for both developed models are available in the supplementary CD.

6.5.2 The QSPR model

The data of the viscosity was transformed using the natural logarithm (\ln) function and modeled thereafter. Moreover, the refined dataset with total 863 data points were modeled. Similar to the GC model, the same equation (6.6) was used to develop a model. In order to have a similar accuracy compared with the GC model, a 22-parameter QSPR model with a total $R^2 = 0.983$ was developed.

6.7

$$\ln(\eta) = A + B \left(\frac{1000}{T} \right) + C \left(\frac{1000}{T} \right)^2$$

$$A = \sum_{i=1}^k n_i a_i, \quad B = \sum_{i=1}^k n_i b_i, \quad C = \sum_{i=1}^k n_i c_i$$

where T is the absolute temperature, n_i is the number of groups of type i , k is the total number of different types of groups, and the parameters a_i , b_i , and c_i are for different substructures. The model parameters are available in Table 6.29.

The cross-plot of experimental data versus calculated/predicted data and the relative deviation of $\ln(\eta)$ from experimental data are presented in Figure 6.30 and Figure 6.31, respectively. In addition, the statistical error analyses for the model are summarized in

Table 6.30. The results show that average absolute relative deviation (AARD%) of the model is 3.07% overall data points, 2.91% for the “training” set, and 3.61% for the “test” set. In addition, R^2 is 0.983 for all data points, 0.985 for the “training” set and 0.980 for the “test” set.

Table 6.29: Parameters of the QSPR model in equation (6.6).

No.	Chemical structure	Descriptions	a_i	b_i	c_i
		Intercept	1.241		
1	GGI5 _{Cat}	topological charge index of order 5	-2.728		
2	RDF035m _{Cat}	Radial Distribution Function - 3.5 / weighted by atomic masses	0.0648		
3	R1v _{Cat}	R maximal autocorrelation of lag 1 / weighted by atomic van der Waals volumes	-48.288		6.168
4	R3v _{Cat}	R maximal autocorrelation of lag 3 / weighted by atomic van der Waals volumes	162.761	-58.990	
5	B05[N-O] _{Cat}	presence of [N-(A) ₄ -O] (value 0 or 1)	1.539		
6	Mor32u _{An}	3D-MoRSE - signal 32 / unweighted	-0.522		
7	E3u _{An}	3rd component accessibility directional WHIM index / unweighted	2.368		
8	MSD _{Cat}	mean square distance index (Balaban)		-5.676	
9	EEig02x _{Cat}	Eigenvalue 02 from edge adj. matrix weighted by edge degrees		0.227	
10	EEig03x _{Cat}	Eigenvalue 03 from edge adj. matrix weighted by edge degrees		-0.191	
11	ESpm01d _{Cat}	Spectral moment 01 from edge adj. matrix weighted by dipole moments		-0.614	0.3123
12	E3u _{Cat}	3rd component accessibility directional WHIM index / unweighted		0.996	
13	E3v _{Cat}	3rd component accessibility directional WHIM index / weighted by atomic van der Waals volumes		-3.220	
14	RBF _{Cat}	rotatable bond fraction			1.694
15	nR06 _{Cat}	number of 6-membered rings			0.0389
16	PW2 _{Cat}	path/walk 2 - Randic shape index			1.532
17	piPC07 _{Cat}	molecular multiple path count of order 07			0.0189
18	EEig02d _{Cat}	Eigenvalue 02 from edge adj. matrix weighted by dipole moments			-0.1547
19	L/Bw _{Cat}	length-to-breadth ratio by WHIM			0.0026
20	HATS2m _{Cat}	leverage-weighted autocorrelation of lag 2 / weighted by atomic masses			-0.5904
21	nBM _{An}	number of multiple bonds			-0.0178
22	Mor17p _{An}	3D-MoRSE - signal 17 / weighted by atomic polarizabilities			0.6370

Table 6.30: The statistical error parameters for the $\ln(\eta)$ in equation (6.6).

Statistical Parameter

<i>training set</i>	
R ²	0.985
Average absolute relative deviation	2.91
Standard deviation error	0.15
Root mean square error	0.15
No. of data points	667

<i>test set</i>	
R ²	0.980
Average absolute relative deviation	3.61
Standard deviation error	0.20
Root mean square error	0.20
No. of data points	196

<i>total</i>	
R ²	0.983
Average absolute relative deviation	3.07
Standard deviation error	0.17
Root mean square error	0.17
No. of data points	863

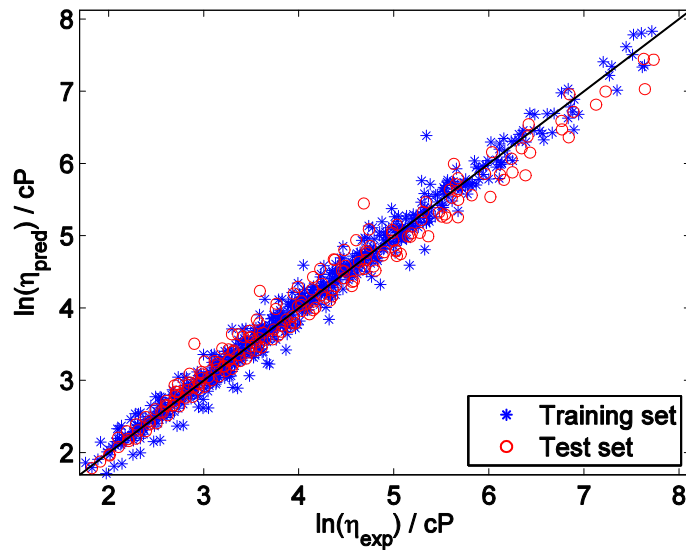


Figure 6.30: Predicted versus experimental values of $\ln(\eta)$ (— diagonal line).

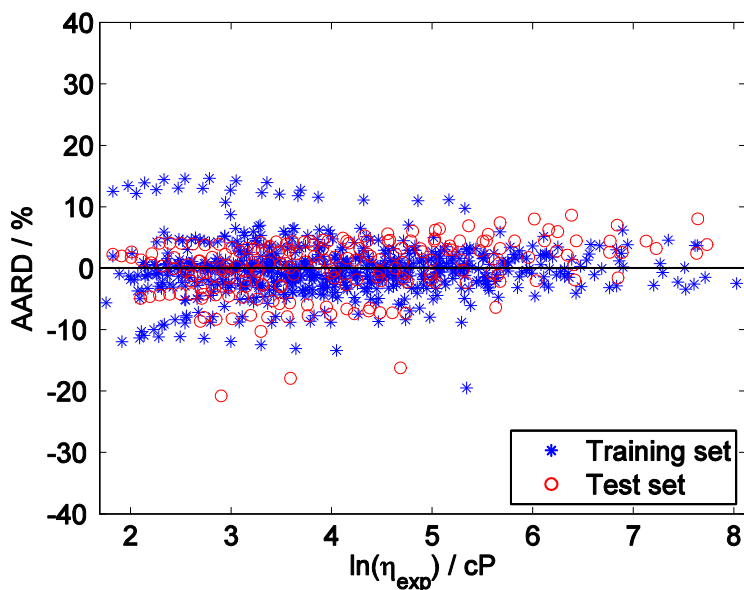


Figure 6.31: Relative deviation of predicted $\ln(\eta)$ from experimental data.

As illustrated in Figure 6.33, almost all data points are near the diagonal which indicates the robustness of the QSPR model for the calculation/prediction of the viscosity of F-ILs. In previous section it was explained that most of the data points with large deviations are for very low viscosity values or for temperatures over 330K, or possible impurity of ionic liquids [163].

To compare this model with the previous GC model more precisely, the linear scale values of viscosity are shown in Figure 6.32. In addition, the AARD% of linear scale data is 12.1% which is close to AARD% of GC model (13.31)%. As discussed in section 4.2.2, the QSPR models have fewer parameters than the GC model of similar accuracy and it can be seen obviously for this model; but the number of parameters is relatively high compared with the QSPR models of other physico-chemical properties of compounds. The common QSPR models have less than 15 parameters and consequently, the larger models are not well accepted by cheminformatics researchers. As a result, further QSPR modeling of this property was ignored; because the model developed had 22 parameters for the refined dataset. So the model would be larger if the complete dataset (1160 data points including unreliable data points) was used.

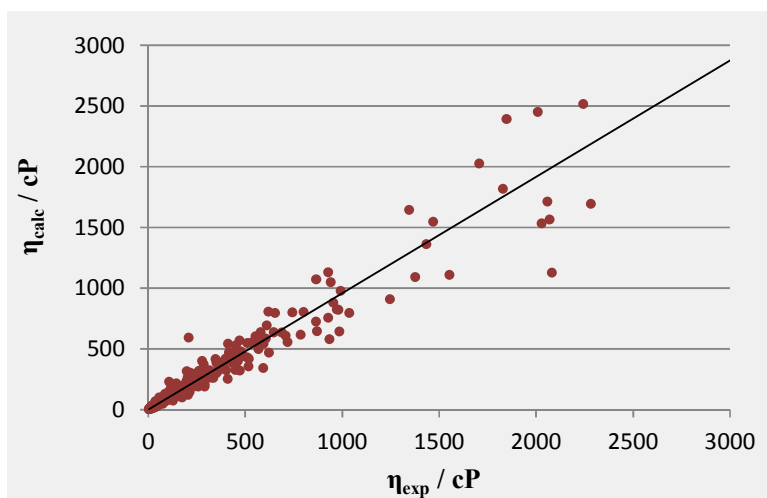


Figure 6.32: Predicted versus experimental values of η in linear scale by the QSPR model (— diagonal line).

Table 6.31: Summary of different models for predicting the $\ln(\eta)$ of ILs.

Model	Model Type and parameters	N_{ILs}	N_{data}	AARD%	Comments
Abbott [81]	Correlation	11	n.a.	122	10 ILs contained fluorine atom.
Bandres et al. [82]	Correlation	8	n.a.	4.5	7 ILs contained fluorine atom.
Gardas and Coutinho [83]	GC, 13 parameters, ρ	30	500	7.78*	Only 19 ILs contained fluorine atom. *AARD% is in linear scale.
Gardas and Coutinho [86]	GC, 12 parameters	25	482	7.50*	Only 16 ILs contained fluorine atom. *AARD% is in linear scale.
Tochigi and Yamamoto [89]	QSPR, 24 parameters	161	334	5.04	149 ILs contained fluorine atom. Most of the ILs had just one data point.
Bini <i>et al.</i> [90]	QSPR, 4 parameters.	33	66	n.a.	Authors proposed two models for $T = 293$ K and $T = 353$ K.
Gharagheizi <i>et al.</i> [91]	GC, 47 parameters	443	1672	6.32	638 duplicated data points were used. By removing the duplicates, the AD% was risen to 7.1%.

Model	Model Type and parameters	N _{ILs}	N _{data}	AARD%	Comments
					724 data points were for F-ILs
Valderrama <i>et al.</i> [92]	ANN	58	327	n.a.	
Billard <i>et al.</i> [93]	ANN	99	99	10	The data were only at 298 K.
GC Model (1)	GC, 35 parameters	85	863	3.23	The reliable data sources were used.
GC Model (2)	GC, 36 parameters	332	1160	4.85	The entire database including unreliable data sources was used.
QSPR Model	QSPR, 22 parameters	85	863	2.91	The reliable data sources were used.

6.6 γ^∞ of solutes in ionic liquids

6.6.1 Aromatic solutes

After examining different computational methods and transformation functions, it was found that fitting the $\ln(\ln(\gamma^\infty))$ produced the more accurate correlation. But the γ^∞ of some system was equal or less than 1 (more than 400 data points) which caused the computational problem in logarithmic scale. To eliminate the negative values after taking the first logarithm, the values of γ^∞ were multiplied by 1000 and then used as the target variable in modeling process. In addition, some systems showed an increase in γ^∞ by temperature increment which was convincing to introduce T and T^2 as the input variables, same as the interaction parameters of mod. UNIFAC (Dortmund) (equation 2.11). Hence, the resultant model was a 42-parameter linear model as shown in equation (6.8).

$$\ln(10^3 \times \gamma^\infty) = \exp\left(\beta_0 + \beta_1 T + \beta_2 T^2 + \frac{\beta_3}{T} + \frac{\beta_4}{T^2}\right) \times \left[1 + \sum_{i=1}^4 \beta_{5+i} \left(\frac{\beta_{5+i}}{\beta_5}\right)^{\beta_{6+i}}\right] \quad (6.8)$$

$$\beta_0 = 2.07 + \sum_{i=1}^4 \beta_{10+i} \left(\frac{\beta_{10+i}}{\beta_{10}}\right)^{\beta_{11+i}}$$

$$\beta_1 = \sum_{i=1}^4 \beta_{12+i} \left(\frac{\beta_{12+i}}{\beta_{12}}\right)^{\beta_{13+i}}$$

$$\beta_2 = \sum_{i=1}^4 \beta_{14+i} \left(\frac{\beta_{14+i}}{\beta_{14}}\right)^{\beta_{15+i}}$$

$$\beta_3 = \sum_{i=1}^4 \beta_{16+i} \left(\frac{\beta_{16+i}}{\beta_{16}}\right)^{\beta_{17+i}}$$

$$\beta_4 = \sum_{i=1}^4 \beta_{18+i} \left(\frac{\beta_{18+i}}{\beta_{18}}\right)^{\beta_{19+i}}$$

where T is absolute temperature, n_i is the number of occurrences of the i th functional group of anions and cations, k is the total number of different functional groups of the anions and cations, and $\alpha_1, \alpha_2, \dots, \alpha_k$ are the relevant coefficient of the i th functional group. In addition, $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ is a correction term for the ionic liquids which have hydroxyalkyl group (2-hydroxyethyl, 3-hydroxypropyl, etc.) in their cation structure. The formula of this correction term is as follows.

6.9

$$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10} = 1.43 + \alpha_1 \alpha_2 + \alpha_3 \alpha_4 + \frac{\alpha_5}{\alpha_6}$$

$$\alpha_1 = \sum_{i=1}^k \alpha_i \alpha_{i+1} \quad \alpha_2 = \sum_{i=1}^k \alpha_i \alpha_{i+2}$$

$$\alpha_3 = \sum_{i=1}^k \alpha_i \alpha_{i+3} h_i$$

It is notable that $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10} = 0$ for ionic liquids which do not have hydroxyalkyl group in their cation structure. The values of $\alpha_1, \alpha_2, \dots, h_i$ are presented in Table 6.32

Table 6.32: Parameters of the equation (6.8) for γ^∞ of aromatic solutes in ILs

β_i	Description	$\beta_{i, \beta}$
β_1	nN _{an}	-1.714E-02
	F-085 _{an}	-5.222E-03
β_2	nAT _{cat}	-0.424
	nCp _{cat}	1.258
	B07[C-O] _{cat}	11.787
	F02[C-C] _{cat}	-0.91
	nSK _{an}	-3.628
	nCp _{an}	4.999
	B03[C-S] _{an}	-8.952
	F03[C-C] _{sol}	8.476
β_3	C-040 _{cat}	7659.61
	B02[N-O] _{cat}	-8420.78
	B05[N-O] _{cat}	6651.02

$\mathbb{Z}_{\bar{e}}$	Description	$\mathbb{Z}_{\bar{e}, \bar{e}}$
B06[C-C] _{cat}	presence/absence of C-(A) ₅ -C	-1108.41
B07[C-C] _{cat}	presence/absence of C-(A) ₆ -C	-901.33
B08[C-O] _{cat}	presence/absence of C-(A) ₇ -O	-16021.12
B10[C-C] _{cat}	presence/absence of C-(A) ₉ -C	-2096.88
F03[C-O] _{cat}	Number of C-(A) ₂ -O	3563.88
F10[C-C] _{cat}	Number of C-(A) ₉ -C	364.26
nRCN _{an}	number of nitriles (aliphatic)	-592.35
B02[S-F] _{an}	presence/absence of S-A-F	-4979.74
F01[O-S] _{an}	Number of O-S	2016.16
F05[O-F] _{an}	Number of O-(A) ₄ -F	361.97
F08[C-C] _{an}	Number of C-(A) ₇ -C	482.61
F06[C-C] _{sol}	Number of C-(A) ₅ -C	1768.79
$\mathbb{Z}_{\bar{e}}$		
F03[C-N] _{an}	Number of C-(A) ₂ -N	-8.172E-06
F05[O-O] _{an}	Number of O-(A) ₄ -O	2.172E-05
$\mathbb{Z}_{\bar{e}}$		
nN ⁺ _{cat}	number of positive charged N	1.064E-07
C-005 _{cat}	Number of CH3X in cation	1.017E-07
C-033 _{cat}	Number of R--CH..X	6.578E-08
H-046 _{cat}	H attached to C ⁰ (sp3), no X attached to next C	-2.060E-08
B09[C-N]	presence/absence of C-(A) ₈ -N	-6.268E-08
C-005 _{an}	Number of CH3X in anion	2.184E-07
B04[C-F] _{an}	presence/absence of C-(A) ₃ -F	4.222E-08
B07[C-C] _{an}	presence/absence of C-(A) ₆ -C	-6.951E-07
F01[C-N] _{an}	Number of C-N	3.422E-08
$\mathbb{Z}_{\bar{e}}$		
nB _{an} × F03[C-C] _{sol}	number of Boron atoms, Number of C-(A) ₂ -C	7.670E-04
nN _{cat} × nCs _{sol}	Number of Nitrogen, number of total secondary C(sp3)	5.563E-07
$\mathbb{Z}_{\bar{e}}$		
nAT _{an} × nCbH _{sol}	Number of atoms, number of unsubstituted benzene C(sp2)	-2.348E-08
$h_{\bar{e}}$		
nSK _{cat} × B04[C-C] _{an}	Number of non-H atoms,	-1.826E+03
nB _{an} × nCs	number of Boron atoms, number of total secondary C(sp3)	3.726E+04

The AARD% of the proposed model for $\ln(\gamma^\infty)$ ($10^2 \times \ln(\gamma^\infty)$) is near 1.33% for both the “training” and “test” sets. Usually the values of $\ln(\gamma^\infty)$ are shown in a scatter plot and it is preferable to calculate the AARD% of $\ln(\gamma^\infty)$ to show the prediction ability of the model, but the AARD% of $\ln(\gamma^\infty)$ is not meaningful for $\ln(\gamma^\infty)$ around 1 which causes division by zero or very large deviations. So the Root Mean Square Error (RMSE) of $\ln(\gamma^\infty)$ of the proposed model is 0.15 for “training” set and 0.20 for “test” set which shows fairly small average deviation between experimental and calculated values of $\ln(\gamma^\infty)$. In addition, the coefficient of determination (R^2) for $\ln(\gamma^\infty)$ is 0.968 and 0.959 for the “training” and “test” sets, respectively. The scattered plot of experimental values versus calculated/predicted values for $\ln(\gamma^\infty)$ is shown in Figure 6.33.

To have the fair comparison for the performance of proposed model, predicted values of γ^∞ in linear scale were plotted in Figure 6.34 against the experimental values of γ^∞ . In addition, the summary of the statistical parameters for the model for γ^∞ for the training and test sets are listed in Table 6.33. Accordingly, the AARD% of model for actual values of γ^∞ is 9.82% and 9.22% for training and test sets, respectively. Furthermore, the RSME of training and test sets are 0.74 and 1.52, respectively which are relatively small values.

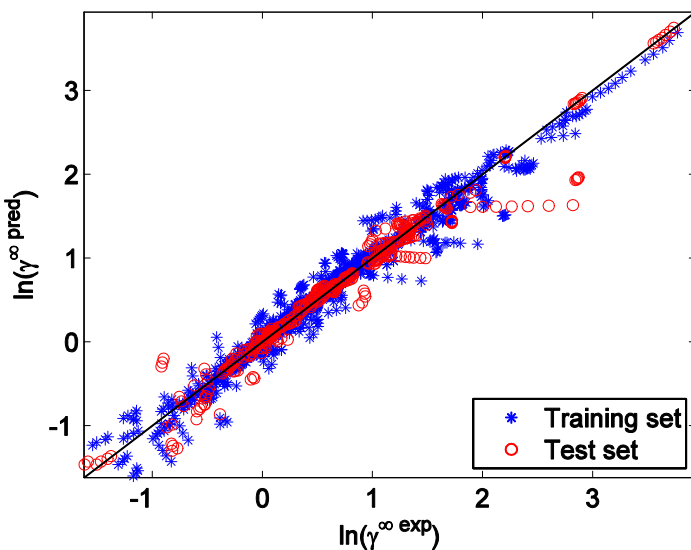


Figure 6.33: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of aromatic solutes (— diagonal line).

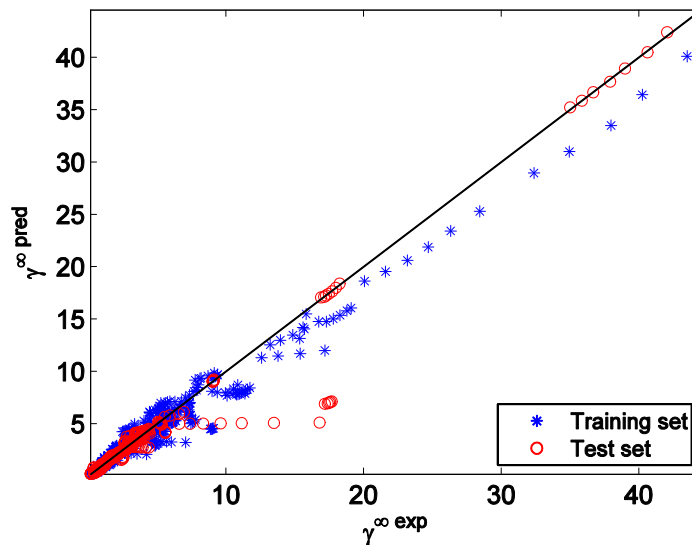


Figure 6.34: Correlated/Predicted versus experimental values of γ^∞ of aromatic solutes (— diagonal line).

Table 6.33: Statistical parameters for the presented model for γ^∞ of aromatic solutes in ILs.

<i>Statistical Parameter</i>	
<i>training set</i>	
R ²	0.970
Average absolute relative deviation	9.82
Standard deviation error	0.73
Root mean square error	0.74
No. of data points	1278
<i>test set</i>	
R ²	0.935
Average absolute relative deviation	9.22
Standard deviation error	1.50
Root mean square error	1.52
No. of data points	375
<i>total</i>	
R ²	0.954
Average absolute relative deviation	9.69
Standard deviation error	0.96
Root mean square error	0.98
No. of data points	1653

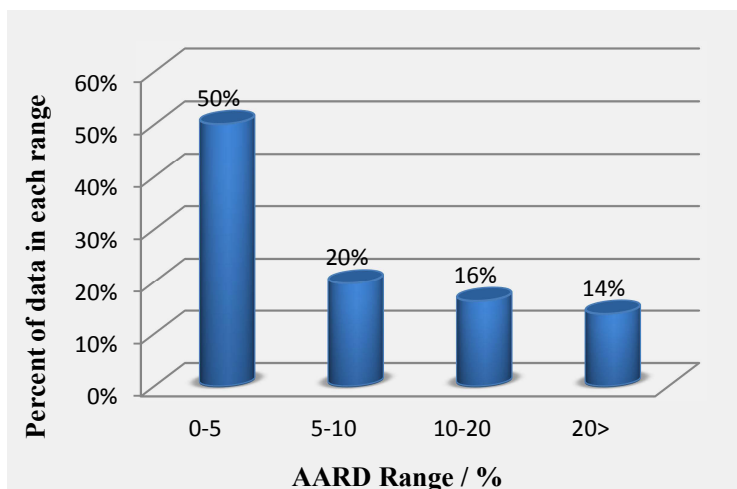


Figure 6.35: Percentage of calculated/predicted values of γ^∞ of aromatic solutes in different relative deviation ranges.

As demonstrated in Figure 6.35, 50% of the calculated/predicted values show deviations between 0 to 5%, 20% between 5 to 10%, 16% between 10 to 20%, 14% over 20 %. According to the data provided in the supplementary CD, the largest deviation is observed for “benzene in trihexyltetradecylphosphonium chloride” which is 93.6%; but after a precise look at the data, it can be found that the range of γ^∞ is 0.4-0.408 and the average error in prediction is -0.378 which is a small deviation; but since the experimental value of γ^∞ for this system is very small, a large deviation is observed. Similar deviations are observed for different systems such as: “benzene and 3-(2-methoxyethyl)-1-methylimidazolium bis(trifluoromethylsulfonyl)imide”, “ethylbenzene and dodecylene-bis(ethylmorpholinium) bis(trifluoromethylsulfonyl)imide”, “ethylbenzene and dodecylene-bis(methylmorpholinium) bis(trifluoromethylsulfonyl)imide”, etc.

There are some other systems with ethyl sulfate or nitrate anion which have large AARD% and relatively large deviations which is caused by presence of strong interactions such as hydrogen bonding between anion and benzene in solute molecule. These systems need further modification in equation (6.8). The typical modification has been done for the systems of which the cations has one hydroxyalkyl group (2-hydroxyethyl, 3-hydroxypropyl, etc.) in their structure. These cations can interact with themselves as well as other anions and solute molecules. Based on the findings of some research groups [165, 241-249], the schematic hydrogen bonding in cation-cation, cation-anion, and cation-solute pair in “benzene” and

“1-(2-hydroxyethyl)-3-methylimidazolium bis(trifluoromethylsulfonyl)imide” are shown in Figure 6.36, Figure 6.37, and Figure 6.38, respectively.

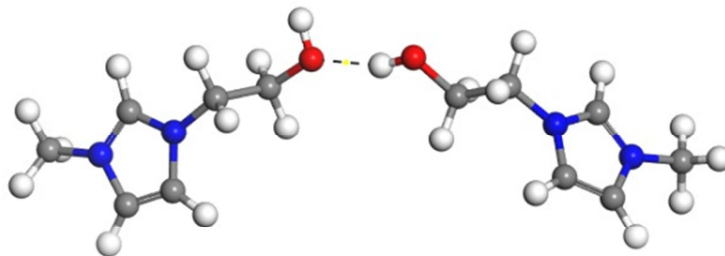


Figure 6.36: Cation-Cation hydrogen bond for 1-(2-hydroxyethyl)-3-methylimidazolium

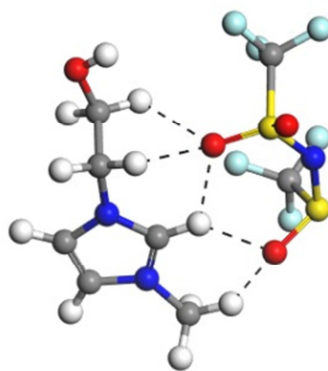


Figure 6.37: Cation-anion hydrogen bonds for 1-(2-hydroxyethyl)-3-methylimidazolium bis(trifluoromethylsulfonyl)imide

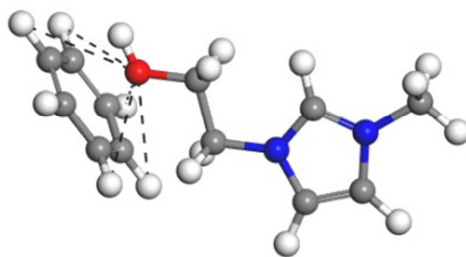


Figure 6.38: Cation-solute hydrogen bonds for 1-(2-hydroxyethyl)-3-methylimidazolium and benzene.

The correction term just for ILs with hydroxyalkyl group has been introduced by term $\frac{V_i^L}{V_i^G} \left(\frac{V_i^L}{V_i^G} \right)^{\frac{V_i^L}{V_i^G}}$ in equation (6.8). These systems could not be correlated well in “training” set and consequently, all of them were located in the “test” set to eliminate their weight on deviation of the model. As this modification improved the prediction results of “test” set, the AARD% of “test” set became less than the “training” set.

Ultimately, the AARD% of calculated/predicted values of γ^∞ for different classes of ionic liquids are presented in Table 6.34. Accordingly, the maximum deviations are observed in Phosphonium and Morpholinium families with AARD% of 16.89% and 14.77% respectively; however the range of γ^∞ is 0.19-2.51 and 0.182-3.27 for these two classes, and as explained before, the small values of γ^∞ are very sensitive to small deviations. The third large deviation class of ionic liquids is Imidazolium with AARD% of 11.58%; but this class consists of 187 different systems with 860 data points and due to the nature of γ^∞ and measurement techniques, it is not a large deviation.

Table 6.34: AARD% of equation (6.8) for different class of ionic liquids.

No.	Family	T /K range	γ^∞ range	AARD%	N Systems	N Data Points
1	Ammonium	303.15-385.45	0.31-7.92	8.53	33	117
2	Guanidinium	308.15-348.15	0.821-1.76	4.218	3	15
3	Imidazolium	293.15-413	0.429-43.51	11.58	187	860
4	Isoquinolinium	328.15-368.15	0.605-1.07	5.72	3	15
5	Morpholinium	308-368.15	0.182-3.27	14.77	18	82
6	Phosphonium	298.15-373.15	0.19-2.51	16.89	16	73
7	Piperidinium	308.15-368.15	0.546-5.18	4.20	21	116
8	Pyridinium	298.15-378.15	0.664-10.87	7.33	29	154
9	Pyrrolidinium	298.15-368.15	0.587-5.35	3.89	41	201
10	Sulphonium	298.15-368.15	1.08-2.45	3.38	3	20

In comparison with the other models, it should be noted that there is not any easy-to-use model available in the literature and the UNIFAC model requires special database and software (DDB) to calculate the γ^∞ of aromatic solutes in ILs. To have a comparison between the output of the proposed GC model and the original and modified UNIFAC models, the γ^∞ is calculated for few systems using the consortium delivery 2014 interaction parameters and

the results are shown in Table 6.36. Accordingly, there are many systems that both original and modified UNIFAC models fail to predict the correct value of γ^∞ ; however there are few systems that modified UNIFAC model performs better than the GC model proposed in this study, such as “benzene” in “1-ethyl-3-methylimidazolium trifluorotris(perfluoroethyl) phosphate”. It seems that such systems have been used to fit the interaction parameters of the modified UNIFAC model. There few other systems that modified UNIFAC model shows the AARD% of over 1000%, such as “ethylbenzene” in “1-butyl-1-methylpyrrolidinium bis(trifluoromethylsulfonyl)imide” (3447.71%); however the original UNIFAC model fails to predict the γ^∞ for them.

The calculated/predicted values of γ^∞ for all of the systems using the GC model proposed, and the original and modified UNIFAC models are available in the supplementary CD. It is notable that the interaction parameters are not available for all of the systems. The summary of results is shown in Table 6.35. Accordingly, the GC model proposed in this study has the best results and it is the first GC model for the calculation/prediction of γ^∞ of aromatic solutes in ILs.

Table 6.35: Summary of result of models for γ^∞ of aromatic solutes in ILs.

Model	Model Type and parameters	N _{Systems}	N _{data}	AARD%	Comments
Original UNIFAC	GC	9	27	64.07	3 solutes in 9 ILs.
Modified UNIFAC	GC	135	389	174.40	4 solutes in 85 ILs.
GC Model (this study)	GC, 42 parameters	354	1653	9.69	10 solutes in 123 ILs.

Table 6.36: γ^∞ of some aromatic solutes in ILs calculated using the GC models proposed (equation 6.8), and the original and modified UNIFAC models.

No.	Solute	IL	T	$\gamma^\infty_{\text{exp}}$	$\gamma^\infty_{\text{GC Model}}$	ARD%	$\gamma^\infty_{\text{org UNIFAC}}$	ARD%	$\gamma^\infty_{\text{mod UNIFAC}}$	ARD%
1	benzene	1-ethyl-3-methylimidazolium trifluoro-tris(perfluoroethyl)phosphate	313.1	1.06	0.727	31.38	n.a.	n.a.	1.017	4.02
2	ethylbenzene	1-ethyl-3-methylimidazolium tetrafluoroborate	313.15	6.3	6.266	0.55	n.a.	n.a.	6.167	2.11
3	ethylbenzene	1-butyl-3-methylimidazolium tetrafluoroborate	313.15	5.537	4.281	22.68	0.755	86.37	1.047	81.09
4	ethylbenzene	1-butyl-1-methylpyrrolidinium bis(trifluoromethylsulfonyl)imide	323.15	1.83	1.833	0.18	n.a.	n.a.	64.923	3447.71
5	toluene	1-butyl-3-methylimidazolium thiocyanate	298.15	3.4	3.469	2.03	n.a.	n.a.	3.391	0.25
6	toluene	1-butyl-1-methylpyrrolidinium tetracyanoborate	338.15	1.14	1.161	1.86	0.765	32.89	0.779	31.68
7	toluene	1-hexyl-3-methylimidazolium trifluoro-tris(perfluoroethyl)phosphate	312.8	0.674	0.670	0.66	n.a.	n.a.	1.521	125.72

6.6.2 Alcohol solutes

To model the data of alcohol solutes, the same procedure was done as explained in section 6.6.1. Hence, the resultant model was a 36-parameter linear model as shown in equation (6.10).

$$\ln(10^{\gamma} \times \gamma^{\infty}) = \exp\left(\beta_0 + \beta_1 T + \beta_2 T^2 + \frac{\beta_3}{T} + \frac{\beta_4}{T^2}\right) \quad 6.10$$

$$\begin{aligned} \beta_0 &= 1.853 + \sum_{i=1}^k \beta_{0i} n_i \gamma_i^{\infty} & \beta_1 &= \sum_{i=1}^k \beta_{1i} n_i \gamma_i^{\infty} \\ \beta_2 &= \sum_{i=1}^k \beta_{2i} n_i \gamma_i^{\infty} & \beta_3 &= \sum_{i=1}^k \beta_{3i} n_i \gamma_i^{\infty} & \beta_4 &= \sum_{i=1}^k \beta_{4i} n_i \gamma_i^{\infty} \end{aligned}$$

where T is absolute temperature, n_i is the number of occurrences of the i th functional group of anions, cations, and solutes, k is the total number of different functional groups of the anions, cations, and solutes, and $\beta_0, \beta_1, \dots, \beta_4$ are the relevant coefficient of the i th functional group. The values of $\beta_0, \beta_1, \dots, \beta_4$ and their description are presented in Table 6.37 and Table 6.38, respectively.

The AARD% of the proposed model for $\ln(10^{\gamma} \times \gamma^{\infty})$ is approximately 1.93% for the “training” set and 2.30% for the “test” set. As mentioned before, usually the values of $\ln(\gamma^{\infty})$ are shown in scatter plots; but the AARD% of $\ln(\gamma^{\infty})$ is not meaningful for γ^{∞} around 1 which causes division by zero or very large deviations. So the RMSE of $\ln(\gamma^{\infty})$ of the proposed model is 0.21 for training set and 0.24 for test set which shows fairly small deviation between experimental and calculated values of $\ln(\gamma^{\infty})$. In addition, the coefficient of determination (R^2) for $\ln(\gamma^{\infty})$ is 0.913 and 0.924 for the “training” and “test” sets, respectively. The scattered plot of experimental values versus calculated/predicted values for $\ln(\gamma^{\infty})$ is shown in Figure 6.39.

To compare the real output and performance of the presented model, Figure 6.40 demonstrates the predicted values of γ^{∞} in linear scale versus the experimental values of γ^{∞} . In addition, Table 6.39 shows the summary of the statistical parameters for the training and test sets of equation (6.9).

Table 6.37: Parameters of equation (6.10)

\mathbb{Z}_F	$\mathbb{Z}_{F,E}$
\mathbb{Z}_F	
$B10[C-N]_{cat} \times F03[C-N]_{cat}$	-1.696E-02
$B08[C-C]_{cat} \times nH_{sol}$	-2.057E-03
$F02[C-C]_{cat} \times nHAcc_{sol}$	-1.123E-04
$B01[C-S]_{an} \times B01[C-C]_{sol}$	5.838E-03
\mathbb{Z}_F	
$B02[C-N]_{an} \times nH_{sol}$	-1.506E-05
\mathbb{Z}_F	
$F01[C-O]_{an}$	-7.880E-07
$B06[C-O]_{cat} \times nB_{an}$	6.403E-07
$F02[C-C]_{cat} \times nSO3_{an}$	-5.762E-08
$nSO4_{an} \times nSK_{sol}$	1.595E-07
$H-051_{an} \times nSK_{sol}$	1.271E-07
$B04[C-O]_{sol} \times H-046_{sol}$	2.375E-08
\mathbb{Z}_F	
$B02[F-F]_{an}$	36.388
$B01[C-C]_{sol} \times nHAcc_{sol}$	-17.994
\mathbb{Z}_F	
$F04[N-O]_{cat} \times F10[C-C]_{cat}$	1810.573
$B10[C-N]_{cat} \times nHAcc_{an}$	678.267
$F02[C-C]_{cat} \times B01[C-S]_{an}$	-431.535
$F09[C-N]_{cat} \times nB_{an}$	3340.265
$B09[C-N]_{cat} \times nH_{sol}$	-713.324
$F06[C-C]_{cat} \times nH_{sol}$	-35.814
$nP_{an} \times nROH_{cat}$	-11439.849
$B02[C-N]_{an} \times F03[C-C]_{cat}$	551.154
$B02[C-N]_{an} \times F08[C-O]_{cat}$	-11746.388
$nN^+_{an} \times nHAcc_{an}$	-2042.971
$nB_{an} \times F03[C-N]_{an}$	252.926
$F02[F-F]_{an} \times F-085_{an}$	50.732
$H-046_{an} \times nS_{an}$	-1230.705
$H-05_{an} \times B01[C-C]_{sol}$	-14227.463
$nOHt_{sol} \times nHAcc_{an}$	-346.107
$nH_{sol} \times B02[F-F]_{an}$	-309.651
$nH_{sol} \times B01[C-C]_{sol}$	2091.650

Table 6.38: Definition of parameter of equation (6.10).

No.	Symbol	Definition
1	B01[C-C]	absence/presence of C-C (0 or 1)
2	B01[C-S]	absence/presence of C-S (0 or 1)
3	B02[C-N]	absence/presence of C-A-N (0 or 1)
4	B02[F-F]	absence/presence of F-A-F (0 or 1)
5	B04[C-O]	absence/presence of C-(A) ₃ -O (0 or 1)
6	B06[C-O]	absence/presence of C-(A) ₅ -O (0 or 1)
7	B08[C-C]	absence/presence of C-(A) ₇ -C (0 or 1)
8	B09[C-N]	absence/presence of C-(A) ₈ -N (0 or 1)
9	B10[C-N]	absence/presence of C-(A) ₉ -N (0 or 1)
10	F01[C-O]	number of C-O
11	F02[C-C]	number of C-A-N
12	F02[F-F]	number of F-A-F
13	F03[C-C]	number of C-(A) ₂ -C
14	F03[C-N]	number of C-(A) ₂ -N
15	F04[N-O]	number of N-(A) ₃ -O
16	F06[C-C]	number of C-(A) ₅ -C
17	F08[C-O]	number of C-(A) ₇ -O
18	F09[C-N]	number of C-(A) ₈ -N
19	F10[C-C]	number of C-(A) ₉ -C
20	F-085	F attached to C ² (sp ²)-C ⁴ (sp ²) / C ¹ (sp) / C ⁴ (sp ³) / X
21	H-046	H attached to C ⁰ (sp ³) no X attached to next C
22	H-051	H attached to alpha-C ^a
23	nB	number of Boron atoms
24	nH	number of Hydrogen atoms
25	nHAcc	Total number of acceptor atoms for H-bonds (N, O, F), excluding N with a formal positive charge, higher oxidation states and pyrrolyl form of N
26	nN+	number of positive charged N
27	nOHt	number of tertiary alcohols
28	nP	number of Phosphorous atoms
29	nROH	number of hydroxyl groups connect to an aliphatic group
30	nS	number of Sulfur atoms
31	nSK	number of non-H atoms
32	nSO3	number of sulfonates (thio- / dithio-)
33	nSO4	number of sulfates (thio- / dithio-)

^a An alpha-C may be defined as a C attached through a single bond with -C=X, -C#X, -C—X

R represents any group linked through carbon

X represents any heteroatom (O, N, S, P, Se, halogens)

A represents any atom

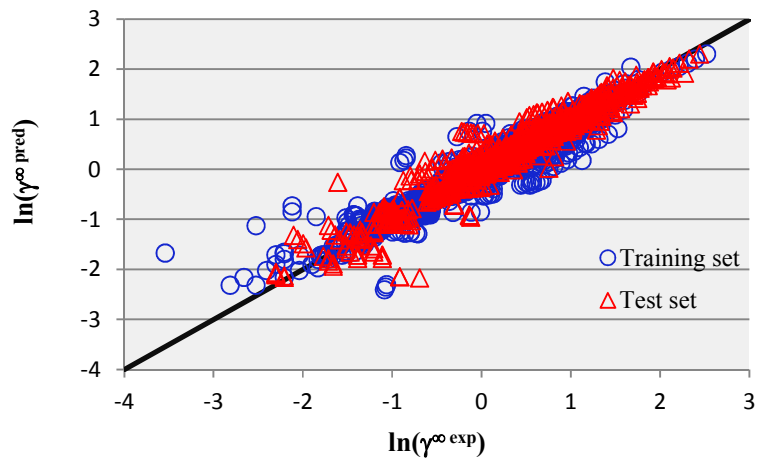


Figure 6.39: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of alcohol solutes (— diagonal line).

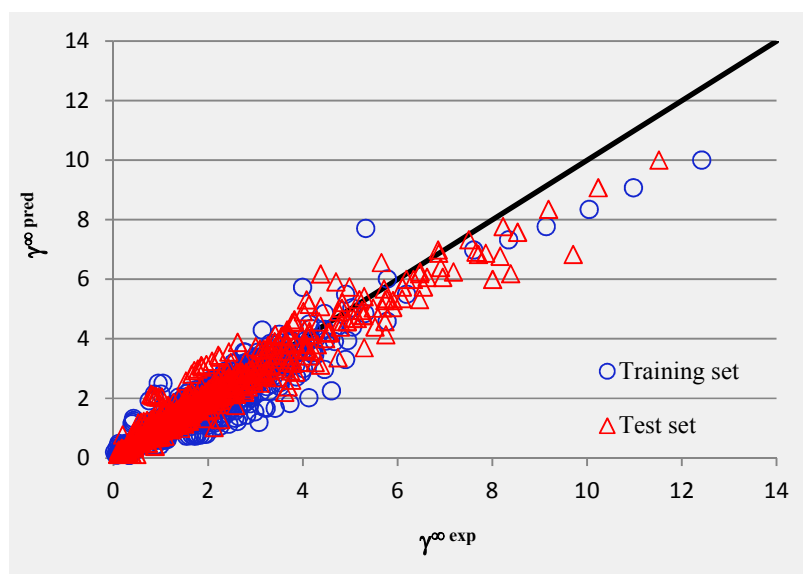


Figure 6.40: Correlated/Predicted versus experimental values of γ^∞ of alcohol solutes (— diagonal line).

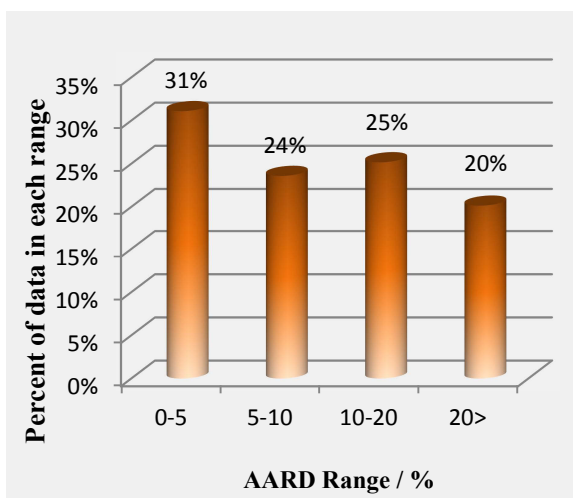


Figure 6.41: Percentage of calculated/predicted values of γ^∞ of alcohol solutes in different relative deviation ranges.

Table 6.39: Statistical parameters for equation (6.10)

<i>Statistical Parameter</i>	
<i>training set</i>	
R^2	0.900
Average absolute relative deviation	13.42
Standard deviation error	0.31
Root mean square error	0.31
No. of data points	2047
<i>test set</i>	
R^2	0.939
Average absolute relative deviation	16.97
Standard deviation error	0.42
Root mean square error	0.42
No. of data points	738
<i>total</i>	
R^2	0.921
Average absolute relative deviation	14.36
Standard deviation error	0.34
Root mean square error	0.34
No. of data points	2785

As indicated in Table 6.39, the AARD% of the model for γ^∞ is 13.42% for the “training” set and 16.97% for the “test” set. Furthermore, the RSME of “training” and “test” sets are 0.31 and 0.42, respectively which are relatively small values.

According to Figure 6.41, 31% of the calculated/predicted values show deviations between 0 to 5%, 24% between 5 to 10%, 25% between 10 to 20%, 20% over 20 %. At the first look, it may seem that the model has large deviation; but as the γ^∞ of more than 900 data points are below 1, a small deviation in predicted values cause a great value of AARD%.

According to the dataset provided in the supplementary CD, the largest deviation belongs to “methanol” in “1,3-dimethylimidazolium dimethylphosphate” which is 547.7%; but it has just one data point and its σ_{AARD}^2 is 0.029. Since the experimental value of γ^∞ for this system is

very small, a large deviation is observed. In addition, as this system only has one data point, it has very low weight in model development process as its deviation does not change noticeably the AARD% of the model. Similar deviations are observed for different systems with just one or two data points such as: “methanol”, “ethanol”, and “1-propanol” in “1-octyl-3-methylimidazolium chloride”; “ethanol” in “1-butyl-3-methylimidazolium chloride”, etc.

There are some other systems with “nitrate” or “tetrafluoroborate” as anion which show large AARD% and relatively large deviations which is caused by presence of strong interactions such as hydrogen bonding between cation and anion. Normally, most of the studied alcohols have H-bond with cations which introduce a deviation in calculations, but the additional H-bond between cation-anion pairs can be the source of large deviations [165, 241-249]. The evidence of such interactions is the presence of “number of acceptor atoms for H-bonds” of anion and solutes ($nHAcc_{an}$ and $nHAcc_{sol}$) in presented model. In addition, anions and solutes can participate in hydrogen bonding too, such as “tetrafluoroborate” as acceptor and hydrogen at –OH group of “methanol” as donator. The proof of this interaction is the presence of $n_{HAcc_{an}} \times n_{HAcc_{sol}}$ in equation (6.10); however in this parameter, the $n_{HAcc_{sol}}$ is the number of tertiary alcohols. The summary of these descriptions are presented in Figure 6.42, Figure 6.43, and Figure 6.44 which show the H-bond in cation-anion, cation-solute, and anion-solute pairs for “methanol” and “1-ethyl-3-methylimidazolium tetrafluoroborate” system. Such systems may be correlated better by further modification of equation (6.10) and inserting a new term or more functional groups to model the systems with above-mentioned anions.

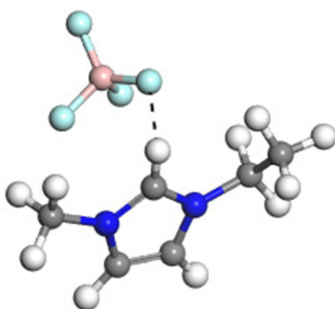


Figure 6.42: Cation-Anion hydrogen bond for 1-ethyl-3-methylimidazolium tetrafluoroborate.

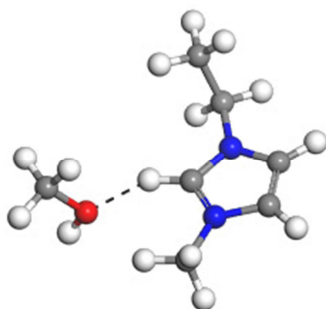


Figure 6.43: Cation-Solute hydrogen bond for 1-ethyl-3-methylimidazolium and methanol

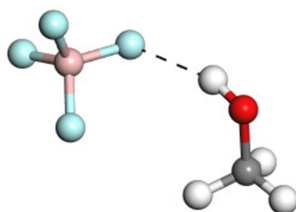


Figure 6.44: Anion-Solute hydrogen bond for tetrafluoroborate and methanol.

Ultimately, the AARD% of calculated/predicted values of γ^{∞} for different classes of ionic liquids are presented in Table 6.40. Accordingly, the maximum deviations are observed in phosphonium and imidazolium families with AARD% of 19.8% and 17.6% respectively; however the range of γ^{∞} for phosphonium class is 0.164-2.27 and as explained before, the small values of γ^{∞} are very sensitive to small deviations. The imidazolium class consists of 337 different systems with 1470 data points and due to the nature of γ^{∞} and measurement techniques; it is not a large deviation.

Table 6.40: AARD% of equation (6.10) for different class of ionic liquids.

No.	Family	T /K range	γ^{∞} range	AARD%	N Systems	N Data Points
1	Ammonium	301.75-386.6	0.1-4.02	12.3	61	221
2	Guanidinium	308.15-348.15	0.907-2.1	9.2	4	20
3	Imidazolium	293.15-413	0.029-12.42	17.6	337	1470
4	Isoquinolinium	328.15-368.15	0.916-1.78	7.9	5	25
5	Morpholinium	308-368.15	0.218-3.73	10.4	28	138
6	Phosphonium	302.45-373.15	0.164-2.27	19.8	20	80
7	Piperidinium	308.15-368.15	0.366-3.75	8.0	38	218
8	Pyridinium	298.15-378.15	0.195-5.759	10.7	40	210
9	Pyrrolidinium	298.15-368.15	0.329-4.48	9.8	78	373
10	Sulphonium	298.15-368.15	0.935-3.32	11.8	4	30

In order to compare this model with previous models, it is notable that there is not any easy-to-use model available in the literature, as discussed in previous section. The only available models are the original and modified UNIFAC models which require special software (DDB) to calculate the γ^∞ of alcohol solutes in ILs. Table 6.42 shows the output of above mentioned models for few alcohol-IL systems. Accordingly, there are some systems that the original and/or modified UNIFAC models predict their γ^∞ better than the GC model.

The calculated/predicted values of γ^∞ for all systems using these three models are available in the supplementary CD. It is notable that the interaction parameters are not available for all ionic liquids. The summary of results is shown in Table 6.35 . Accordingly, the GC model proposed in this study shows the best performance which is the first GC model for the calculation/prediction of γ^∞ of alcohol solutes in ILs.

Table 6.41: Summary of result of models for γ^∞ of alcohol solutes in ILs.

Model	Model Type and parameters	N _{Systems}	N _{data}	AARD%	Comments
Original UNIFAC	GC	47	205	27.34	13 solutes in 7 ILs.
Modified UNIFAC	GC	164	703	32.22	13 solutes in 26 ILs.
GC Model (this study)	GC, 36 parameters	615	2785	14.36	17 solutes in 126 ILs.

Table 6.42: γ^∞ of some alcohol solutes in ILs calculated using the GC models proposed (equation 6.9), and the original and modified UNIFAC models.

No.	Solute	IL	T	$\gamma^\infty_{\text{exp}}$	$\gamma^\infty_{\text{GC Model}}$	ARD%	$\gamma^\infty_{\text{orig UNIFAC}}$	ARD%	$\gamma^\infty_{\text{mod UNIFAC}}$	ARD%
1	1-butanol	1-butyl-3-methylimidazolium trifluoromethanesulfonate	308.15	1.66	2.080	25.30	n.a.	n.a.	1.6603062	0.02
2	1-pentanol	1-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	333.25	2.18	2.114	3.01	1.7951817	17.65	2.1152835	2.97
3	2-butanol	1-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	293.15	2.03	2.597	27.94	1.6116513	20.61	2.0340787	0.20
4	2-butanol	1-(2-hydroxyethyl)-3-methylimidazolium tetrafluoroborate	303.15	6.876	6.869	0.10	n.a.	n.a.	22.200032	222.86
5	2-methyl-1-propanol	1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	313.15	2.85	2.676	6.10	2.8376674	0.43	2.7749396	2.63
6	2-methyl-1-propanol	1-(2-hydroxyethyl)-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	322.55	2.38	2.390	0.44	2.6394176	10.90	3.1033135	30.39
7	2-propanol	1-decyl-3-methylimidazolium tetracyanoborate	328.15	0.938	0.952	1.48	1.2407551	32.28	1.259237	34.25
8	2-propanol	1-ethyl-3-methylimidazolium tetrafluoroborate	323	1.05	2.502	138.26	n.a.	n.a.	5.6600154	439.05
9	methanol	1-decyl-3-methylimidazolium tetracyanoborate	318.15	0.826	0.680	17.63	1.4178283	71.65	1.7590327	112.96

6.6.3 Alkane solutes

As explained in section 5.6.3, the entire database was divided into two subsets based on the number of carbon atoms of the solute; so two separate models were developed.

6.6.3.1 Solutes with less than 10 carbon atoms

For this category, a 45-parameter model was developed of which 20 parameters were the functional groups of cations, 19 for anions, and 6 for solutes.

6.11

$$K_{ow} (10^{\log}) \times \gamma_{ow} = \exp\left(\beta + \sum_{i=1}^n \beta_i \gamma_i + \sum_{j=1}^m \beta_j \gamma_j + \frac{\beta_{cat}}{\gamma_{cat}} + \frac{\beta_{an}}{\gamma_{an}}\right)$$

$$\beta = 2.02 + \sum_{i=1}^n \beta_i \gamma_i + \sum_{j=1}^m \beta_j \gamma_j$$

$$\beta_{cat} = \sum_{i=1}^n \beta_i \gamma_i + \sum_{j=1}^m \beta_j \gamma_j$$

$$\beta_{an} = \sum_{i=1}^n \beta_i \gamma_i + \sum_{j=1}^m \beta_j \gamma_j$$

$$\beta_{sol} = \sum_{i=1}^n \beta_i \gamma_i + \sum_{j=1}^m \beta_j \gamma_j$$

$$\beta_{cat} = \sum_{i=1}^n \beta_i \gamma_i + \sum_{j=1}^m \beta_j \gamma_j$$

The values of β_i , β_j ,... β_n and their description are presented in Table 6.43 and Table 6.44, respectively.

Table 6.43: Parameters of equation (6.11)

β_i	$\beta_{i,E}$
β_i	
B06[C-O] <i>cat</i>	4.354E-02
B09[C-C] <i>cat</i>	-1.532E-02
β_j	
nPyridines <i>cat</i>	-8.743E-03
B10[C-N] <i>cat</i>	-1.063E-02
B05[C-F] <i>an</i>	1.296E-02
B08[C-C] <i>an</i>	-7.521E-03
B03[C-C] <i>sol</i>	2.793E-02
F01[C-C] <i>sol</i>	6.230E-03
β_{cat}	
nR10 <i>cat</i>	1.570E-04
nN+ <i>cat</i>	5.241E-03

\mathbb{Z}_F	$\mathbb{Z}_{F,E}$
CH3X _{cat}	-7.571E-04
CH2X2 _{cat}	-6.395E-04
B06[O-O] _{cat}	3.875E-03
F02[C-C] _{cat}	-4.354E-04
F07[C-O] _{cat}	5.144E-04
F09[C-O] _{cat}	-3.913E-03
nAT _{an}	-4.751E-04
nN _{an}	-6.077E-04
B01[C-S] _{an}	-1.094E-02
B02[S-F] _{an}	1.002E-02
B05[C-C] _{an}	-9.266E-04
F06[C-O] _{an}	9.572E-04
nCL _{sol}	1.858E-03
\mathbb{Z}_F	
R-SH _{an}	1.911E-01
B02[F-F] _{an}	2.539E-03
B05[C-F] _{an}	-2.113E-02
\mathbb{Z}_F	
nCs _{cat}	2.196E-04
nRCN _{cat}	1.859E-02
R--CH--R _{cat}	1.419E-03
B01[C-N] _{cat}	1.425E-03
B07[C-C] _{cat}	-3.501E-03
B08[C-N] _{cat}	-1.388E-03
F02[C-C] _{cat}	-1.079E-03
F07[C-N] _{cat}	-2.538E-03
F08[C-N] _{cat}	1.742E-03
nN _{an}	-1.036E-03
nX _{an}	-1.965E-04
nN+ _{an}	-1.078E-04
CH3R/CH4 _{an}	2.417E-03
R-SH _{an}	-4.012E-02
B02[C-S] _{an}	1.317E-03
B06[C-O] _{an}	1.445E-02
B06[C-S] _{an}	-1.471E-02
B07[C-C] _{an}	-3.362E-02
B08[C-S] _{an}	1.537E-02
F04[O-S] _{an}	-7.609E-03
nCp _{sol}	8.134E-04
nCs _{sol}	1.105E-03
CH2RX _{sol}	1.038E-03
B03[C-C] _{sol}	2.212E-02

Table 6.44: Definition of parameter of equation (6.11).

No.	Symbol	Definition
1	CH3R / CH4	
2	CH3X	
3	CH2RX	
4	CH2X2	
5	R--CH--R	
6	B01[C-N]	absence/presence of C-N (0 or 1)
7	B01[C-S]	absence/presence of C-S (0 or 1)
8	B02[C-S]	absence/presence of C-A-S (0 or 1)
9	B02[F-F]	absence/presence of F-A-F (0 or 1)
10	B02[S-F]	absence/presence of S-A-F (0 or 1)
11	B03[C-C]	absence/presence of C-(A) ₂ -C (0 or 1)
12	B05[C-C]	absence/presence of C-(A) ₄ -C (0 or 1)
13	B05[C-F]	absence/presence of C-(A) ₄ -F (0 or 1)
14	B06[C-O]	absence/presence of C-(A) ₅ -O (0 or 1)
15	B06[C-S]	absence/presence of C-(A) ₅ -S (0 or 1)
16	B06[O-O]	absence/presence of O-(A) ₅ -O (0 or 1)
17	B07[C-C]	absence/presence of C-(A) ₆ -C (0 or 1)
18	B08[C-C]	absence/presence of C-(A) ₇ -C (0 or 1)
19	B08[C-N]	absence/presence of C-(A) ₇ -N (0 or 1)
20	B08[C-S]	absence/presence of C-(A) ₇ -S (0 or 1)
21	B09[C-C]	absence/presence of C-(A) ₈ -C (0 or 1)
22	B10[C-N]	absence/presence of C-(A) ₉ -N (0 or 1)
23	F01[C-C]	number of C-C
24	F02[C-C]	number of C-A-C
25	F04[O-S]	number of O-(A) ₃ -S
26	F06[C-O]	number of C-(A) ₅ -O
27	F07[C-N]	number of C-(A) ₆ -N
28	F07[C-O]	number of C-(A) ₆ -O
29	F08[C-N]	number of C-(A) ₇ -N
30	F09[C-O]	number of C-(A) ₈ -O
31	nAT	number of atoms
32	nCL	number of Chlorine atoms
33	nCp	number of terminal primary Carbon (sp ³)
34	nCs	number of total secondary Carbon (sp ³)
35	nN	number of Nitrogen atoms
36	nN+	number of positive charged N
37	nPyridines	number of Pyridines
38	nR10	number of 10-membered rings
39	nRCN	number of nitriles (aliphatic)
40	nX	number of halogen atoms
41	R-SH	

R represents any group linked through carbon , X represents any heteroatom (O, N, S, P, Se, halogens)

A represents any atom

-- represents an aromatic bond as in benzene or delocalized bonds such as the N-O bond in a nitro group

The value of AARD% is not meaningful for $\ln(\gamma^\infty)$ for values near zero; so the RMSE of $\ln(\gamma^\infty)$ of the proposed model is 0.31 for the “training” set and 0.40 for the “test” set. In addition, R^2 for $\ln(\gamma^\infty)$ is 0.966 and 0.960 for the “training” and “test” sets, respectively. The scattered plot of experimental values versus calculated/predicted values for $\ln(\gamma^\infty)$ is shown in Figure 6.45.

Figure 6.40 demonstrates the predicted values of γ^∞ in linear scale versus the experimental values of γ^∞ . In addition, Table 6.39 shows the summary of the statistical parameters of the model for γ^∞ for the training and test sets.

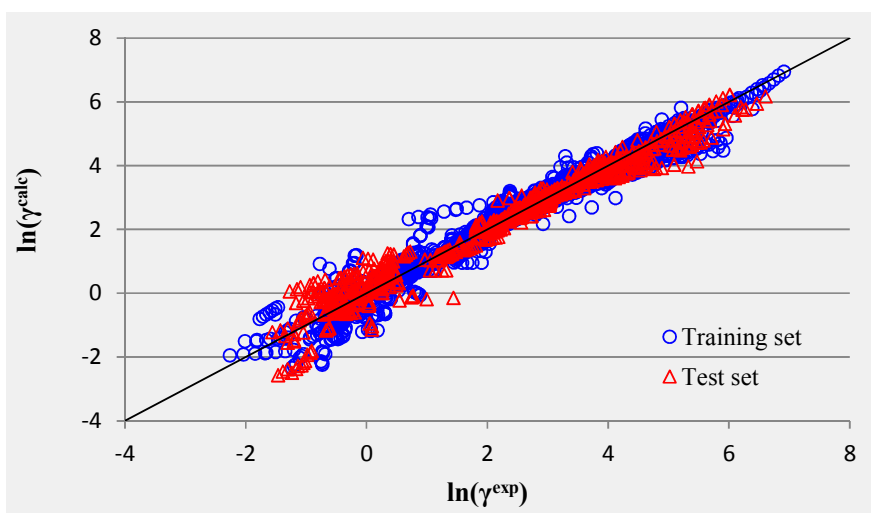


Figure 6.45: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of alkane solutes ($nC_{sol} < 10$) (— diagonal line).

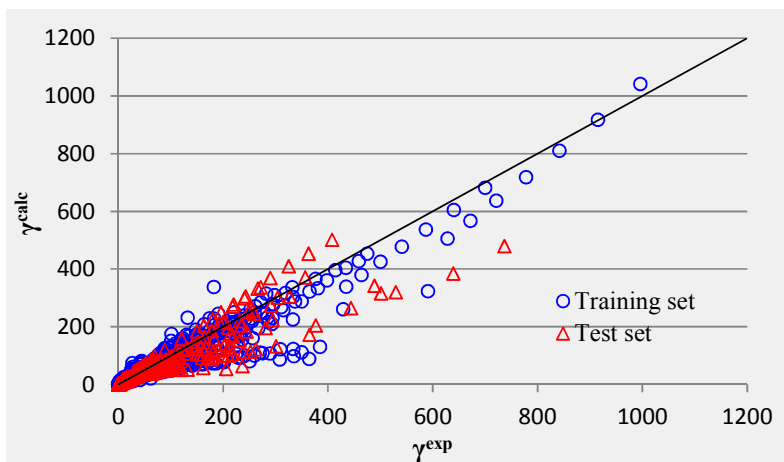


Figure 6.46: Correlated/Predicted versus experimental values of γ^∞ of alkane solutes ($nC_{sol} < 10$) (— diagonal line).

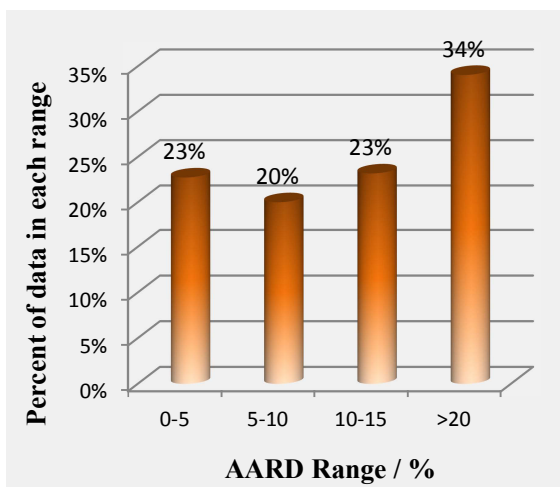


Figure 6.47: Percentage of calculated/predicted values of γ^∞ of alkane solutes ($n_{C_{sol}} < 10$) in different relative deviation ranges.

Table 6.45: Statistical parameters for equation (6.11)

Statistical Parameter	
training set	
R^2	0.920
Average absolute relative deviation	20.56
Standard deviation error	22.13
Root mean square error	22.40
No. of data points	2653
test set	
R^2	0.867
Average absolute relative deviation	28.83
Standard deviation error	31.42
Root mean square error	32.04
No. of data points	715
total	
R^2	0.907
Average absolute relative deviation	22.32
Standard deviation error	24.43
Root mean square error	24.76
No. of data points	3368

According to Table 6.45, the AARD% is 20.56% for the “training” set and 28.83% for the “test” set. Furthermore, the RSME of “training” and “test” sets are 22.40 and 32.04. The value of RMSE is relatively high compared with the models for γ^∞ of aromatic and alcohol solutes. A comparison between γ^∞ values of alkane solutes with aromatic and alcohol solutes (Figure 6.46 versus Figure 6.34 and Figure 6.40) shows that some alkane solutes have large values of γ^∞ . For example, the highest observed γ^∞ for aromatic solutes is 43.51 while there are 685 data points of alkane solutes with γ^∞ of greater than 50.0.

As indicated in Figure 6.46, the model presented has higher deviations in prediction of the γ^∞ values greater than 100.0. It is due to the type of objective function used which minimizes the average absolute “relative” deviation. The minimization of AARD% results in creation of large deviations for larger target values (the AARD% remains relatively low) while it tries to keep the deviation less for smaller target values. As most of the γ^∞ data have small values, the

AARD% objective function is more useful compared with the other types of objective functions such as Mean Squared Error (MSE). MSE objective function is not useful for modeling both the small and large target values which may introduce very high AARD% (over 500%) for values less than 1.0. For example, 10% of deviation for $\gamma^\infty = 1000$ is equal to 100 which is a relatively large deviation; but if the model predicts the $\gamma^\infty = 1.0$ as 2.0, the deviation is low while the AARD% is equal to 100%. As a result, the AARD objective function is more suitable for smaller values.

According to Figure 6.47, 23% of the calculated/predicted values show deviations between 0 to 5%, 20% between 5 to 10%, 23% between 10 to 20%, 34% over 20 %. At the first look, it may seem that the model has large deviations and fails to predict the data; but as the small values of γ^∞ are more sensitive to deviations, a small deviation in prediction results in a great value of AARD%. For the solutes with less than 10 carbon atoms, there are 455 data points with γ^∞ of less than 1.0 and the AARD% over these data points is 47.4%; so the smaller values of γ^∞ are the main reason of the observing large AARD% in calculated/predicted values of γ^∞ .

Regarding to the dataset provided in the supplementary CD, the largest deviations belong to trichloromethane and dichloromethane in 1-ethyl-3-methylimidazolium tosylate which are 446.7% and 300.7%; but these systems have just one data point which have a very low weight in model development process as their deviations do not change noticeably the AARD% of the model. Similar deviations are observed for different systems with just one or two data points.

The other systems with high AARD values have γ^∞ of less than 1.0 which are very sensitive to small deviations. These systems mostly contain chloride and nitrate anions which are highly polar. As a result, the formation of strong H-bond between cation and anion are the reason of such deviations.

According to the dataset provided in the supplementary CD, there are 742 systems for 13 solutes and 122 ionic liquids. Due to presence of strong interactions between components (solute, cation, and anion), it seems that more functional groups should be inserted into the model developed; but as explained previously, large models are not desirable and the aim of this study is to develop the smallest model with reasonable accuracy over the entire database.

Ultimately, Table 6.46 is shown the AARD% of calculated/predicted values of γ^{∞} for different classes of ionic liquids. The maximum deviation belongs to Phosphonium family with AARD% of 46.3%; like the previous class of solutes. As approximately half of the data points of Phosphonium family are below 1.0 and the small values of γ^{∞} are very sensitive to small deviations, such a high AARD% is observed.

Table 6.46: AARD% of equation (6.10) for different class of ionic liquids.

No.	Family	T (K) range	Y range	AARD%	N Systems	N Data Points
1	Ammonium	301.62-374.95	0.29-100.55	24.4	76	256
2	Guanidinium	308.15-348.15	0.786-37.2	9.0	8	40
3	Imidazolium	293.15-373	0.17-996	27.2	367	1612
4	Isoquinolinium	328.15-368.15	3.66-7.45	6.4	4	20
5	Morpholinium	308-368.15	0.104-125	18.3	34	170
6	Phosphonium	298.15-373.15	0.23-61.4	46.3	35	127
7	Piperidinium	308.15-368.15	0.495-333	7.5	47	277
8	Pyridinium	297-368.15	0.577-381	15.0	54	309
9	Pyrrolidinium	298.15-368.15	0.26-408	17.0	112	517
10	Sulphonium	298.15-368.15	12.7-87.2	4.7	5	40

The comparison of model proposed is discussed in next section, over entire database of alkane solutes. Information on the entire dataset and original data sources, as well as the values of the functional groups for ILs are available in the supplementary CD.

6.6.3.2 Solutes with 10 or more carbon atoms

For the second part of γ^{∞} data of alkane solutes, a 28-parameter model was developed of which 18 parameters were the functional groups of cations, 9 for anions, 2 for solutes, and the absolute temperature.

6.12

$$\gamma^{\infty} (10^{\gamma} \times \gamma^{\infty}) = \exp\left(\gamma + \sum_{i=1}^9 \gamma_i \gamma_i^{\infty} + \sum_{j=1}^2 \gamma_j \gamma_j^{\infty} + \frac{\gamma}{T} + \frac{\gamma}{T^2}\right)$$

$$\gamma = 2.505 + \sum_{i=1}^9 \gamma_i \gamma_i^{\infty}$$

$$\gamma = \sum_{i=1}^9 \gamma_i \gamma_i^{\infty}$$

$$\gamma = \sum_{j=1}^2 \gamma_j \gamma_j^{\infty}$$

$$\gamma = \sum_{i=1}^9 \gamma_i \gamma_i^{\infty}$$

$$\gamma = \sum_{i=1}^9 \gamma_i \gamma_i^{\infty}$$

The values of α_{C} , α_{N} , ... α_{O} and their description are presented in Table 6.43 and Table 6.44, respectively.

Table 6.47: Parameters of equation (6.12)

α_{C}	$\alpha_{\text{N},\text{E}}$	
α_{C}	F07[C-C] _{cat}	-2.026E-03
	CH3X	9.973E-02
α_{C}	CH3X _{cat}	-1.861E-03
	F01[C-C] _{cat}	3.490E-04
	F10[C-C] _{cat}	1.355E-03
	F02[C-N] _{an}	-3.382E-03
α_{C}	nCp _{cat}	-1.287E-03
	nRCN _{cat}	8.633E-03
	F02[C-C] _{cat}	-1.087E-03
	B01[C-N] _{an}	3.906E-03
	F03[C-N] _{an}	-9.085E-04
	F06[C-C] _{an}	-1.139E-03
α_{C}	B03[C-N] _{cat}	9.071E-03
	B04[C-N] _{cat}	-2.628E-03
	B08[C-O] _{cat}	-1.042E-02
	nB _{an}	1.284E-02
	F04[C-C] _{an}	-4.123E-02
	F05[C-C] _{sol}	1.235E-02
α_{C}	nCs _{cat}	-7.273E-04
	nN+ _{cat}	2.674E-03
	nROH _{cat}	6.294E-03
	B07[C-C] _{cat}	-2.331E-03
	B08[C-N] _{cat}	-4.246E-03
	B09[C-N] _{cat}	-2.937E-03
	F03[C-N] _{cat}	-3.622E-04
	F03[N-O] _{cat}	2.666E-03
	nAT _{an}	-1.180E-03
	B02[C-S] _{an}	6.688E-03
	F04[C-C] _{an}	1.607E-02
	F08[C-C] _{sol}	-1.170E-03

Table 6.48: Definition of parameter of equation (6.12).

No.	Symbol	Definition
1	CH3X	number of CH3X group
2	B01[C-N]	absence/presence of C-N (0 or 1)
3	B02[C-S]	absence/presence of C-A-S (0 or 1)
4	B03[C-N]	absence/presence of C-(A) ₂ -N (0 or 1)
5	B04[C-N]	absence/presence of C-(A) ₃ -N (0 or 1)
6	B07[C-C]	absence/presence of C-(A) ₆ -C (0 or 1)
7	B08[C-N]	absence/presence of C-(A) ₇ -N (0 or 1)
8	B08[C-O]	absence/presence of C-(A) ₇ -O (0 or 1)
9	B09[C-N]	absence/presence of C-(A) ₈ -N (0 or 1)
10	F01[C-C]	number of C-C
11	F02[C-C]	number of C-A-C
12	F02[C-N]	number of C-A-N
13	F03[C-N]	number of C-(A) ₂ -N
14	F03[N-O]	number of N-(A) ₂ -O
15	F04[C-C]	number of C-(A) ₃ -C
16	F05[C-C]	number of C-(A) ₄ -C
17	F06[C-C]	number of C-(A) ₅ -C
18	F07[C-C]	number of C-(A) ₆ -C
19	F08[C-C]	number of C-(A) ₇ -C
20	F10[C-C]	number of C-(A) ₉ -C
21	nAT	number of atoms
22	nB	number of Boron atoms
23	nCp	number of terminal primary Carbon (sp ³)
24	nCs	number of total secondary Carbon (sp ³)
25	nN+	number of positive charged N
26	nRCN	number of nitriles (aliphatic)
41	nROH	number of hydroxyl groups

R represents any group linked through carbon

X represents any heteroatom (O, N, S, P, Se, halogens)

A represents any atom

-- represents an aromatic bond as in benzene or delocalized bonds such as the N-O bond in a nitro group

Figure 6.48 shows the experimental values versus calculated/predicted values for $\log K_{ow}$ (Eq. 6.12). As mentioned before, the value of AARD% is not meaningful for $\log K_{ow}$ (Eq. 6.12) for values near zero; so the RMSE of $\log K_{ow}$ (Eq. 6.12) of the presented model is 0.24 for the “training” set and 0.29 for the “test” set. In addition, R^2 for $\log K_{ow}$ (Eq. 6.12) is 0.976 and 0.975 for the “training” and “test” sets, respectively.

Figure 6.49 demonstrates the predicted values of γ^{∞} in linear scale versus the experimental values of γ^{∞} . In addition, the statistical parameters for the model are summarized in

Table 6.49. Accordingly, the AARD% of the model is 11.45% for the “training” set and 12.72% for the “test” set. In addition, the RSME of “training” and “test” sets are 184.01 and 110.96. As discussed earlier, this database contains the large values of γ^∞ and AARD% objective function tried to minimize only the “relative” deviation.

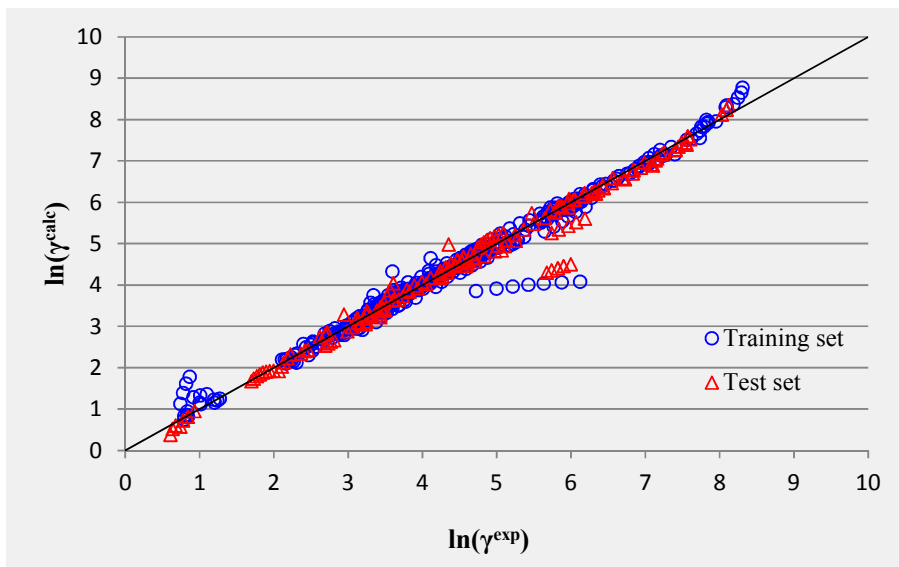


Figure 6.48: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of alkane solutes ($nC_{sol} \geq 10$) (— diagonal line).

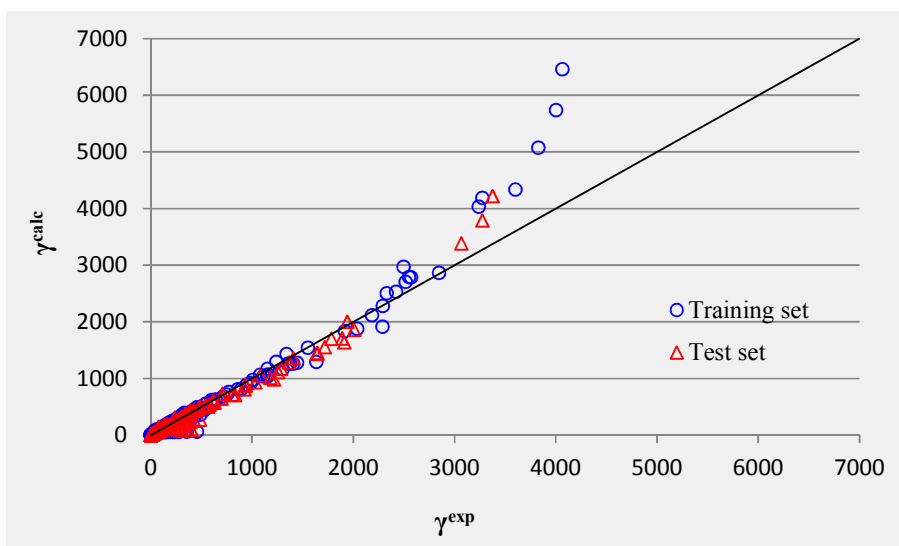


Figure 6.49: Correlated/Predicted versus experimental values of γ^∞ of alkane solutes ($nC_{sol} \geq 10$) (— diagonal line).

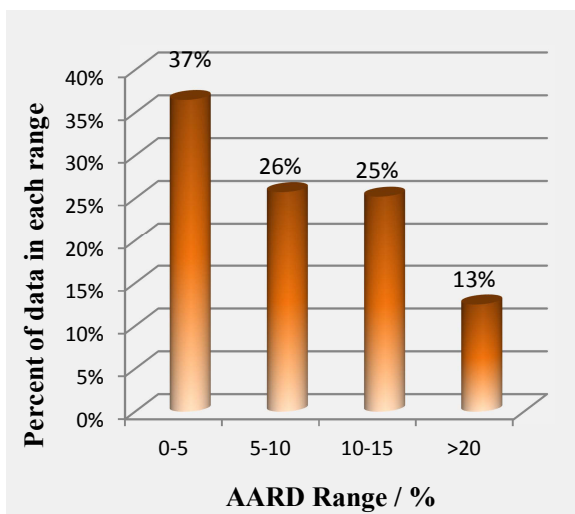


Figure 6.50: Percentage of calculated/predicted values of γ^∞ of alkane solutes in different relative deviation ranges ($nC_{sol} \geq 10$).

Table 6.49: Statistical parameters for equation (6.11)

Statistical Parameter	
training set	
R^2	0.961
Average absolute relative deviation	11.45
Standard deviation error	183.58
Root mean square error	184.01
No. of data points	402
test set	
R^2	0.971
Average absolute relative deviation	12.72
Standard deviation error	109.47
Root mean square error	110.96
No. of data points	165
total	
R^2	0.960
Average absolute relative deviation	11.82
Standard deviation error	166.06
Root mean square error	166.10
No. of data points	567

Figure 6.50 shows that 37% of the calculated/predicted values have deviations between 0 to 5%, 26% between 5 to 10%, 25% between 10 to 20%, 13% over 20 %. Regarding to the dataset provided in the supplementary CD, the largest deviations belong to “decane” in 1-“octyl-3-methylimidazolium tetrafluoroborate” and “trihexyl(tetradecyl)-phosphonium tetrafluoroborate” ionic liquids which are 108.3% and 100.9%. The first system has just one data point which has a very low weight in model development process as its deviation does not change noticeably the AARD% of the model; so the variable selection procedure may ignore the large deviation of this system. Similar deviations are observed for “dodecane” in “1-octyl-3-methylimidazolium tetrafluoroborate” (87.2%) and “undecane” in “1-octyl-3-methylimidazolium tetrafluoroborate” (71.9%) which have just one data point. On the other hand, the average γ^∞ value of above mentioned systems is less than 5.0 while the average γ^∞ value of entire database is 525.5. As this model has been developed for the system with large values of γ^∞ , such deviations are inevitable for small values.

Figure 6.48 indicated that there are two systems with large deviations in logarithmic scale. The first one, which is located in “training” set, is “decane” in “1-hexyl-3-methylimidazolium trifluoromethanesulfonate”. The second system, which is in “test” set, is “decane” in “1-ethyl-3-methylimidazolium diethylphosphate”. For the first system, Figure 6.51 shows the γ^∞ of different solutes in “1-hexyl-3-methylimidazolium trifluoromethanesulfonate”. It is clearly obvious that the data of “decane” is outlier and the measured data are not correct.

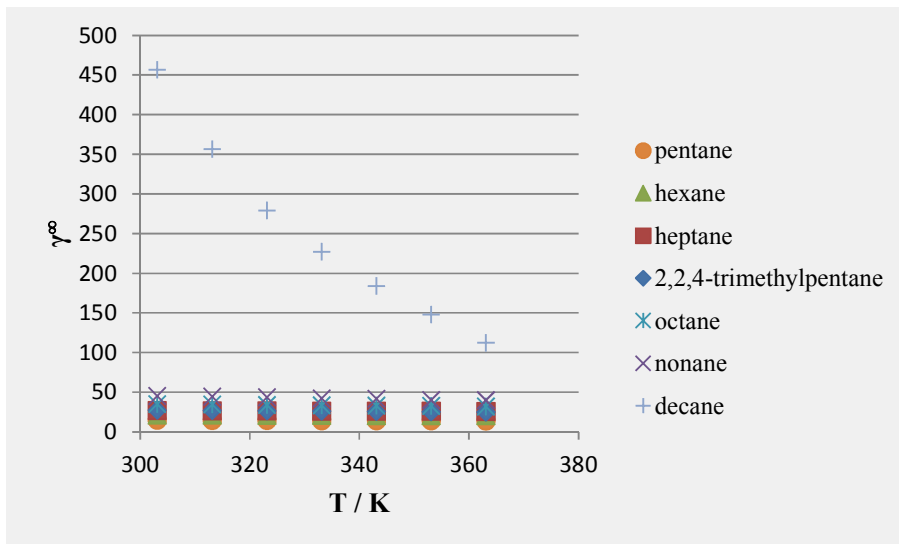


Figure 6.51: The γ^∞ of different alkane solutes in 1-hexyl-3-methylimidazolium trifluoromethanesulfonate.

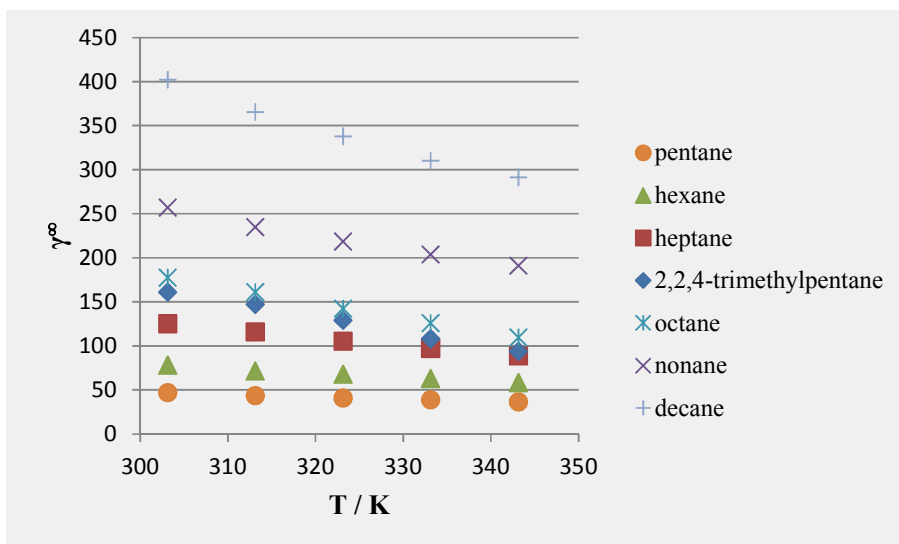


Figure 6.52: The γ^∞ of different alkane solutes in 1-ethyl-3-methylimidazolium diethylphosphate.

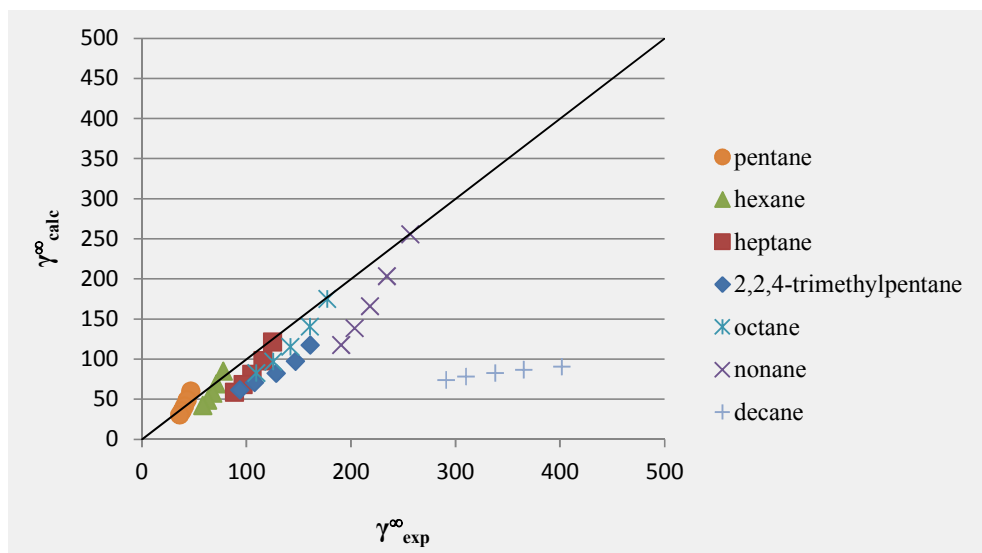


Figure 6.53: Correlated/Predicted versus experimental values of γ^{∞} of different alkane solutes in 1-ethyl-3-methylimidazolium diethylphosphate (— diagonal line).

For the second system, Figure 6.52 shows the γ^{∞} of different solutes in “1-ethyl-3-methylimidazolium diethylphosphate”. Accordingly, there is not any strange behavior while the solutes are changed. For further analysis, the calculated/predicted versus the experimental values of γ^{∞} of different solutes in “1-ethyl-3-methylimidazolium diethylphosphate” is shown in Figure 6.53 (using equation 6.11 and equation 6.12). It is clear that previous model for solutes with $nC_{sol} < 10$ calculates/predicts the γ^{∞} of different solutes in this ionic liquid fairly well; but the model for solutes with $nC_{sol} \geq 10$ fails to predict the γ^{∞} of “decane” in this ionic liquid as it is observed a large deviation from the diagonal line. The analysis of the dataset of solutes with $nC_{sol} \geq 10$ reveals that there is only one ionic liquid with “diethylphosphate” anion in this dataset. As a result, the variable selection procedure has not included a functional group for this anion as it has been located in the “test” set.

Regarding to Figure 6.49, there are few systems with γ^{∞} of higher than 3000 which also show large deviations. These systems belong to different alkane solutes in “1-ethyl-3-methylimidazolium methanesulfonate”. Figure 6.54 shows that the model fails to predict the γ^{∞} of “tetradecane” in “1-ethyl-3-methylimidazolium methanesulfonate”; however the model calculates/predicts the γ^{∞} of tetradecane in other ionic liquids fairly good (Figure 6.55).

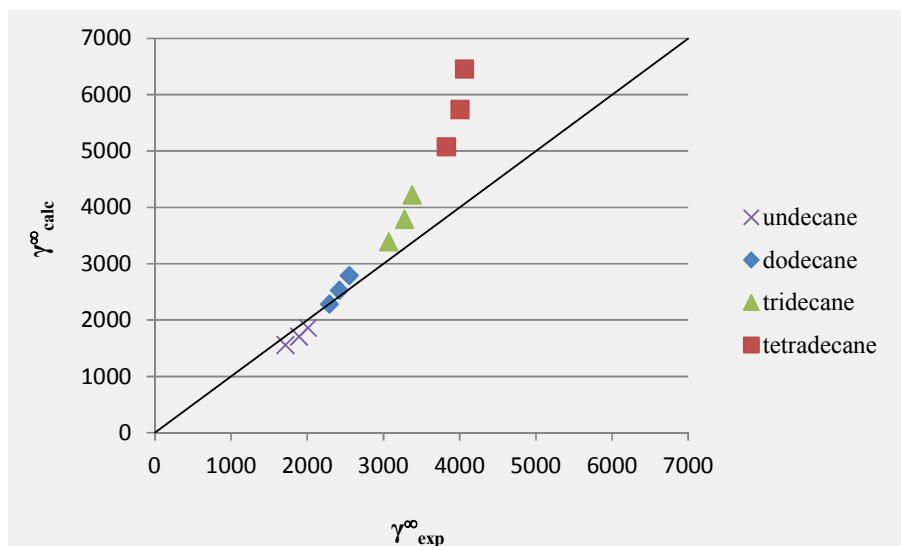


Figure 6.54: Correlated/Predicted versus experimental values of γ^∞ of different alkane solutes in 1-ethyl-3-methylimidazolium methanesulfonate (— diagonal line).

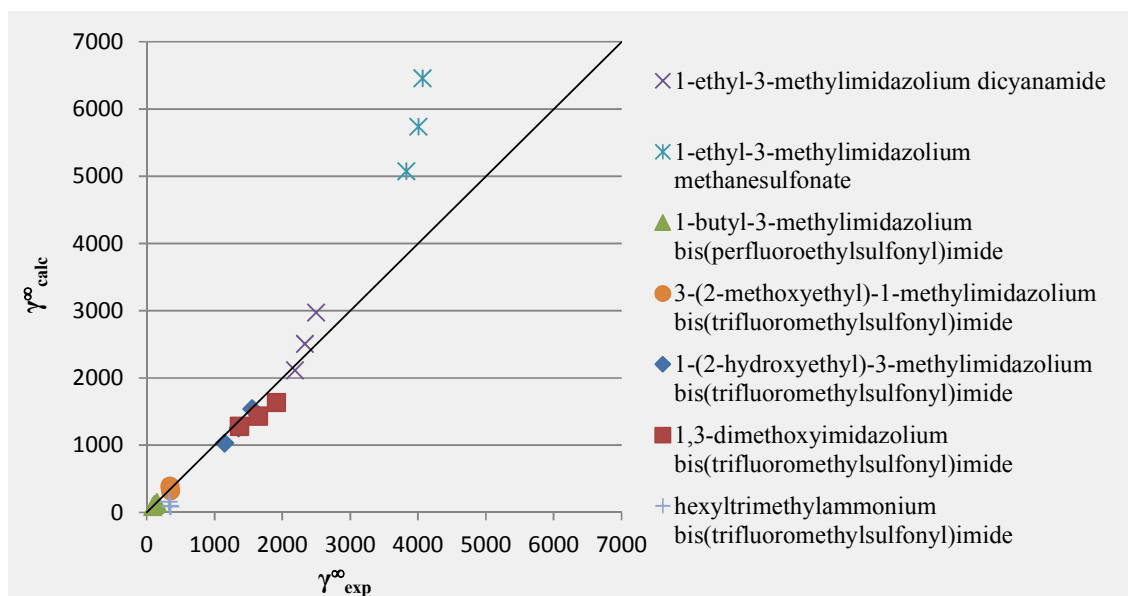


Figure 6.55: Correlated/Predicted versus experimental values of γ^∞ of tetradecane in different ionic liquids (— diagonal line).

Finally, the AARD% of calculated/predicted values of γ^∞ for different classes of ionic liquids are shown in Table 6.52. Accordingly, the maximum deviation belongs to Phosphonium family with AARD% of 29.2%. As discussed before, this class contains several data point with γ^∞ of below 1.0; so they are very sensitive to small deviations. The family with the second highest deviation is imidazolium which is the largest family and consists of 73

different systems with 278 data points. As this family contains the data points with both small and large values of γ^∞ and the presented model has been developed to correlate/predict the large values of γ^∞ , such a relatively large deviation is observed.

Table 6.50: AARD% of equation (6.12) for different class of ionic liquids.

No.	Family	T /K range	γ^∞ range	AARD%	N Systems	N Data Points
1	Ammonium	302.15-374.95	2.21-158.777	6.7	21	64
2	Guanidinium	308.15-348.15	32.8-51.3	4.0	1	5
3	Imidazolium	298-373	3.328-4065.351	13.7	73	278
4	Isoquinolinium	328.15-368.15	7.88-9.25	7.2	1	5
5	Morpholinium	318.15-368.15	55.2-192	9.9	2	12
6	Phosphonium	298.15-373.15	1.84-33.3	29.2	5	18
7	Piperidinium	308.15-368.15	18.3-487	10.4	7	41
8	Pyridinium	297-368.15	24.3-533	11.5	8	49
9	Pyrrolidinium	298.15-368.15	8.7-615	8.0	21	87
10	Sulphonium	298.15-368.15	59.7-131	11.3	1	8

The summary of comparison between the presented models with $n_{C_{sol}} < 10$ and $n_{C_{sol}} \geq 10$, and the original and modified UNIFAC models is shown in Table 6.51. It is obvious that the GC models proposed for alkane solutes has the lowest AARD% and consequently, it is more reliable compared with the original and modified UNIFAC models.

For the modified UNIFAC model, there are 357 data points with an AARD% of greater than 100%. For example, the output of the model for “2,2,4-trimethylpentane” in “1-octyl-3-methylimidazolium nitrate” is 26.739 (AARD% = 100.6%) , while the γ^∞ of “2,2,4-trimethylpentane” in “1-(2-methoxyethyl)-1-methylpiperidinium trifluorotris(perfluoroethyl) phosphate” is predicted as 8144.442 (AARD% = 65055.53%).

Consequently, the GC models proposed in this study have the best results and it is the first GC models for the calculation/prediction of γ^∞ of alkane solutes in ILs.

Information on the entire dataset and original data sources, as well as the values of the functional groups for ILs are available in the supplementary CD.

Table 6.51: Summary of result of models for γ° of alkane solutes in ILs.

Model	Model Type and parameters	N _{Systems}	N _{data}	AARD%	Comments
Original UNIFAC	GC	96	254	331.04	12 solutes in 61 ILs.
Modified UNIFAC	GC	293	849	1927.49	14 solutes in 113 ILs.
GC Model (this study)	GC, 45 parameters for nC _{sol} <10 28 parameters for nC _{sol} ≥10	882	3935	20.81	18 solutes in 123 ILs.

6.6.4 Alkene solutes

As described in section 6.6.1, the same procedure was performed to model the $\ln(\ln(1000 \times \gamma^\circ))$ for alkene solutes. The resultant model was a 44-parameter linear model as shown in equation (6.13).

$$\ln(\ln(10^3 \times \gamma^\circ)) = \exp\left(\frac{A}{T} + \frac{B}{T^2} + \sum_{i=1}^k \frac{C_i}{T} + \sum_{i=1}^k \frac{D_i}{T^2}\right) \quad 6.13$$

$$A = 2.467 + \sum_{i=1}^k \frac{A_i}{T} \quad B = \sum_{i=1}^k \frac{B_i}{T} \quad C_i = \sum_{j=1}^k \frac{C_{ij}}{T} \quad D_i = \sum_{j=1}^k \frac{D_{ij}}{T}$$

where T is absolute temperature, n_i is the number of occurrences of the i th functional group of anions, cations, and solutes, k is the total number of different functional groups of the anions, cations, and solutes, and A_i, B_i, \dots, D_i are the relevant coefficient of the i th functional group. The values of A_i, B_i, \dots, D_i and their description are presented in Table 6.37 and Table 6.38, respectively.

Table 6.52: Parameters of equation (6.13)

\mathbb{Z}_E	$\mathbb{Z}_{E,E}$
\mathbb{Z}_E	
nN ⁺ _{cat}	4.049E-02
nHAcc _{cat}	-2.718E-02
X--CH..X _{cat}	-3.496E-02
F02[C-C] _{cat}	-1.622E-02
F07[C-N] _{cat}	-7.445E-02
F08[C-N] _{cat}	5.492E-02
F10[C-C] _{cat}	7.793E-03
nCs _{an}	-2.967E-02
nO ⁻ _{an}	-4.480E-02
F01[O-P] _{an}	-1.314E-02
nCL _{sol}	-4.191E+00
F05[C-C] _{sol}	-2.641E-01
\mathbb{Z}_E	
CH3R/CH4 _{cat}	-8.292E-03
nO _{an}	7.470E-03
nHAcc _{an}	-9.283E-03
F07[C-O] _{an}	-1.344E-02
F08[C-S] _{an}	4.459E-02
F09[C-C] _{an}	6.119E-03
nR06 _{sol}	-1.455E-02
\mathbb{Z}_E	
F04[C-N] _{cat}	8.795E-04
F10[C-N] _{cat}	1.012E-03
nF _{an}	6.528E-02
nX _{an}	-6.393E-02
CH3R/CH4 _{an}	-2.367E-03
CH3X _{an}	2.317E-03
F03[C-O] _{an}	-4.863E-04
F04[C-C] _{an}	8.818E-04
F04[O-S] _{an}	-1.773E-03
\mathbb{Z}_E	
nRCN _{cat}	5.952E-02
nROH _{cat}	1.611E-02
nPyridines _{cat}	-6.318E-03
F04[N-O] _{cat}	1.204E-02
F06[O-O] _{cat}	2.556E-02
F07[C-O] _{cat}	-1.019E-02
nF _{an}	-1.409E-01
nX _{an}	1.362E-01
F01[C-C] _{an}	-2.905E-03

F02[C-F] _{an}	3.385E-03
nCL _{sol}	2.789E+00
F05[C-C] _{sol}	1.924E-01

□

nCs _{cat}	-1.970E-03
nCrS _{cat}	7.915E-04
F04[C-C] _{cat}	9.973E-04
F05[N-O] _{cat}	5.367E-03
R#N/R=N ⁻ _{an}	-2.457E-03
F01[P-F] _{an}	7.663E-04
F05[C-S] _{an}	3.040E-03
nCL _{sol}	-4.681E-01
F03[C-C] _{sol}	3.171E-03
F05[C-C] _{sol}	-3.420E-02

Table 6.53: Definition of parameter of equation (6.13).

No.	Symbol	Definition
1	CH3R / CH4	
2	CH3X	
3	X--CH..X	
4	F01[C-C]	number of C-C
5	F01[O-P]	number of O-P
6	F01[P-F]	number of P-F
7	F02[C-C]	number of C-A-C
8	F02[C-F]	number of C-A-F
9	F03[C-C]	number of C-(A) ₂ -C
10	F03[C-O]	number of C-(A) ₂ -O
11	F04[C-C]	number of C-(A) ₃ -C
12	F04[C-N]	number of C-(A) ₃ -N
13	F04[N-O]	number of N-(A) ₃ -O
14	F04[O-S]	number of O-(A) ₃ -S
15	F05[C-C]	number of C-(A) ₄ -C
16	F05[C-S]	number of C-(A) ₄ -S
17	F05[N-O]	number of N-(A) ₄ -O
18	F06[O-O]	number of O-(A) ₅ -O
19	F07[C-N]	number of C-(A) ₆ -N
20	F07[C-O]	number of C-(A) ₆ -O
21	F08[C-N]	number of C-(A) ₇ -N
22	F08[C-S]	number of C-(A) ₇ -S
23	F09[C-C]	number of C-(A) ₈ -C

24	F10[C-C]	number of C-(A) ₉ -C
25	F10[C-N]	number of C-(A) ₉ -N
26	R#N / R=N ⁻	
27	nCL	number of Chlorine atoms
28	nCrS	number of ring secondary Carbon (sp ³)
29	nCs	number of total secondary Carbon (sp ³)
30	nF	number of Flourine atoms
31	nHAcc	number of acceptor atoms for H-bonds (N,O,F)
32	nN ⁺	number of positively charged N
33	nO	number of Oxygen atoms
34	nPyridines	number of Pyridines
35	nR06	number of 6-membered rings
36	nRCN	number of nitriles (aliphatic)
37	nROH	number of hydroxyl groups
38	nX	number of halogen atoms
39	nO ⁻	number of negatively charged Oxygen atoms

^a An alpha-C may be defined as a C attached through a single bond with -C=X, -C#X, -C—X

R represents any group linked through carbon

X represents any heteroatom (O, N, S, P, Se, halogens)

A represents any atom

-- represents an aromatic bond as in benzene or delocalized bonds such as the N-O bond in a nitro group

.. represents aromatic single bonds as the C-N bond in pyrrole.

As mentioned before, the value of AARD% for γ^{∞} does not represent the model accuracy very well; because γ^{∞} tend to infinity when γ^{∞} approaches zero. As a result, the RMSE is 0.20 and 0.23 for the “training” and “test” sets, respectively. In addition, the coefficient of determination for γ^{∞} is 0.973 and 0.957 for the training and test sets, respectively. The scattered plot of experimental values versus calculated/predicted values for γ^{∞} is shown in Figure 6.56.

To demonstrate the real output and performance of the presented model, Figure 6.57 shows the predicted values of γ^{∞} in a linear scale versus the experimental values of γ^{∞} . In addition, Table 2.1 shows the summary of the statistical parameters of the model for the training and test sets.

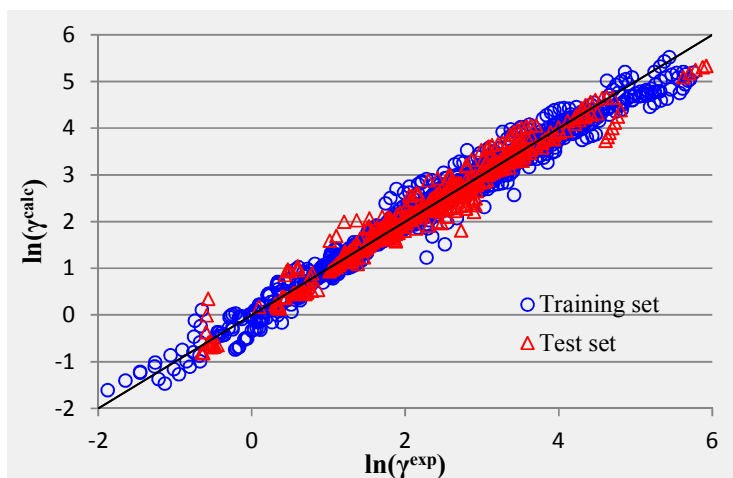


Figure 6.56: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of alkene solutes (— diagonal line).

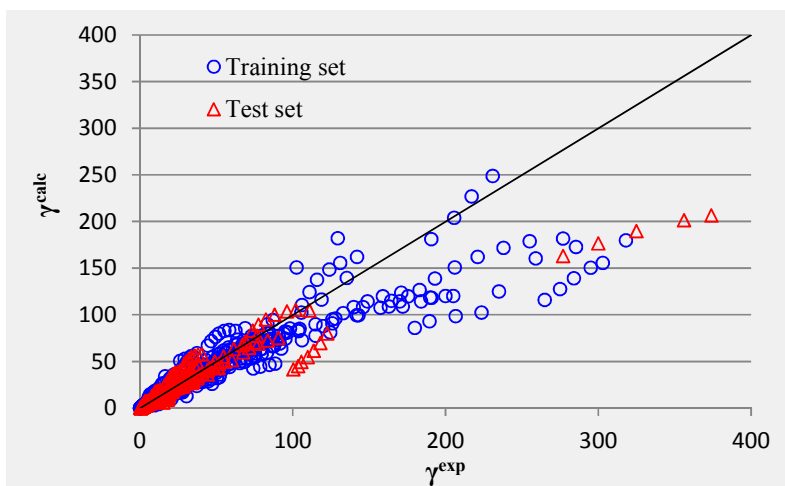


Figure 6.57: Correlated/Predicted versus experimental values of γ^∞ of alkene solutes (— diagonal line).

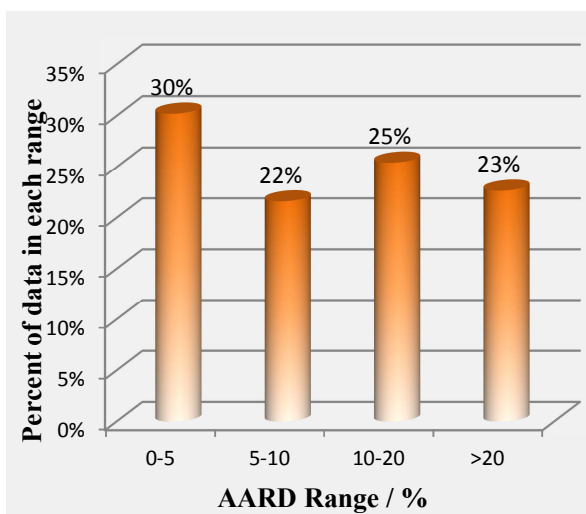


Figure 6.58: Percentage of calculated/ predicted values of γ^∞ of alkene solutes in different relative deviation ranges.

Table 6.54: Statistical parameters for equation (6.13)

Statistical Parameter	
training set	
R ²	0.891
Average absolute relative deviation	13.83
Standard deviation error	14.93
Root mean square error	15.11
No. of data points	1536
test set	
R ²	0.890
Average absolute relative deviation	16.86
Standard deviation error	16.22
Root mean square error	16.31
No. of data points	475
total	
R ²	0.890
Average absolute relative deviation	14.54
Standard deviation error	15.25
Root mean square error	15.40
No. of data points	2011

According to Table 6.54, the AARD% of the model is 13.83% for the “training” set and 16.86% for the “test” set. Furthermore, the RSME of the “training” and “test” sets are 15.11 and 16.31. The value of RMSE is relatively high compared with the models for γ^∞ of aromatic and alcohol solutes. A comparison between γ^∞ values of alkene solutes with aromatic and alcohol solutes (Figure 6.57 versus Figure 6.34 and Figure 6.40) shows that some alkene solutes have very large values of γ^∞ . For example, the highest observed γ^∞ for aromatic solutes is 43.51 while there are more than 200 data points of alkene solutes with γ^∞ of greater than 50.0. According to Figure 6.57, the presented model has more deviation in prediction of the γ^∞ of greater than 100.0. It is due to the type of objective function used which minimizes the average absolute “relative” deviation. The minimization of AARD% results in production of large deviations for the larger target values (the AARD% remains relatively low) while it minimized the deviation for smaller target values. As most of the γ^∞

data have small values, the AARD% objective function is more useful compared with other types of objective functions such as Mean Squared Error (MSE). MSE objective function is not useful for modeling the small target values which may introduce very high AARD% (over 500%) for values less than 1.0.

According to Figure 6.58, 30% of the calculated/predicted values show deviations between 0 to 5%, 22% between 5 to 10%, 25% between 10 to 20%, 23% over 20 %. At the first look, it may seem that the model has a large deviation; but as the small values of γ^∞ are more sensitive to deviations, a small deviation in prediction results in a great value of AARD%.

Regarding to the dataset provided in the supplementary CD, the largest deviation belongs to “1-octene” in “1-octyl-3-methylimidazolium tetrafluoroborate” which is 100.2%; but it has just two data points which has a very low weight in model development process as their deviations do not change noticeably the AARD% of the model. Similar deviations are observed for different systems with just one or two data points such as “1-dodecene” in “1-octyl-3-methylimidazolium tetrafluoroborate”, “1-heptene” in “1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide”, etc.

It is notable that alkene solutes have strong interactions with cations and anions, similar to aromatic and alcohol solutes. In addition, the database used comprised of 422 different systems and it may be required to include more functional groups into the model to have a better prediction of γ^∞ ; however inclusion of more parameters increases the complexity of the model which is not desirable.

Ultimately, the AARD% of calculated/predicted values of γ^∞ for different classes of ionic liquids is presented in Table 6.55. Accordingly, the maximum deviation belongs to Phosphonium family with AARD% of 28.9%; however about half of the data points are below 1.0 and as explained before, the small values of γ^∞ are very sensitive to small deviations. The imidazolium class consists of 337 different systems with 1470 data points and due to the nature of γ^∞ and measurement techniques; it is not a large deviation.

Table 6.55: AARD% of equation (6.13) for different class of ionic liquids.

No.	Family	T /K range	γ^∞ range	AARD%	N Systems	N Data Points
1	Ammonium	301.85-374.95	0.74-76.8	10.9	33	133
2	Guanidinium	308.15-348.15	3.92-26.5	5.4	6	30
3	Imidazolium	293.15-375.05	1.1-374	17.0	213	927
4	Isoquinolinium	328.15-368.15	2.53-4.85	6.3	4	20
5	Morpholinium	308-368.15	0.153-87.8	11.4	16	88
6	Phosphonium	298.15-373.15	0.468-30.5	28.9	28	107
7	Piperidinium	308.15-368.15	2.75-90.6	6.5	33	192
8	Pyridinium	298.15-368.15	3.018-142	12.1	40	228
9	Pyrrolidinium	298.15-368.15	2.57-62.49	13.6	45	254
10	Sulphonium	298.15-368.15	6.69-27	4.2	4	32

To compare the model proposed with the original and modified UNIFAC models, the AARD% and number of systems and data points are shown in Table 6.56 for all models. The entire results are available in the supplementary CD. Accordingly, the model proposed performs better than both the original and modified UNIFAC models in terms of better predictions, number of data points, and systems covered. As a result, the GC model proposed can be used easier as there is no need to use any special software, and it produces more reliable results.

Table 6.56: Summary of result of models for γ^∞ of alkene solutes in ILs.

Model	Model Type and parameters	N _{Systems}	N _{data}	AARD%	Comments
Original UNIFAC	GC	49	212	36.80	11 solutes in 8 ILs.
Modified UNIFAC	GC	111	486	70.74	12 solutes in 29 ILs.
GC Model (this study)	GC, 44 parameters	422	2011	14.54	13 solutes in 123 ILs.

6.6.5 Alkyne solutes

As mentioned previously for other solutes, $\ln(\ln(1000 \times \gamma^{\infty}))$ was modeled for alkyne solutes. The final model has 37 parameters as shown in equation (6.13) of which 21 parameters are the functional groups of cations, 15 for anions, and 2 for solutes.

6.14

$$\ln(10^3 \times \gamma^{\infty}) = \exp\left(\frac{A}{T} + B + C \ln T + D \ln^2 T + \frac{E}{T} + \frac{F}{T^2}\right)$$

$$A = 2.044 + \sum_{i=1}^k n_i a_i \quad B = \sum_{i=1}^k n_i b_i$$

$$C = \sum_{i=1}^k n_i c_i \quad D = \sum_{i=1}^k n_i d_i \quad E = \sum_{i=1}^k n_i e_i$$

where T is absolute temperature, n_i is the number of occurrences of the i th functional group of anions, cations, and solutes, k is the total number of different functional groups of the anions, cations, and solutes, and a_i, b_i, \dots, e_i are the relevant coefficient of the i th functional group. The values of a_i, b_i, \dots, e_i and their description are presented in Table 6.37 and Table 6.38, respectively.

Table 6.57: Parameters of equation (6.14)

a_i	b_i, e_i
a_i	
F06[C-N] _{cat}	-5.724E-03
nCar _{an}	-4.952E-03
b_i	
R--CR--R _{cat}	-7.026E-03
F01[C-N] _{cat}	4.034E-03
CRX3 _{an}	-7.306E-04
R-SH _{an}	2.327E-02
F05[C-F] _{an}	-9.289E-04
c_i	
nCrs _{cat}	4.694E-04
nROR _{cat}	-1.514E-03
nPyridines _{cat}	-3.339E-03
R--CH--R _{cat}	2.290E-03
F02[C-C] _{cat}	-1.910E-04
F02[C-N] _{cat}	1.934E-04

$\mathbb{2}_F$		$\mathbb{2}_{\mathbb{2},F}$
	F03[N-O] _{cat}	1.812E-03
	F05[C-N] _{cat}	-6.661E-04
	F06[O-O] _{cat}	2.720E-03
	F10[C-O] _{cat}	-7.997E-04
	nB _{an}	-3.153E-04
	F02[C-S] _{an}	-1.186E-03
	F02[F-F] _{an}	1.425E-04
	F08[C-O] _{an}	-1.283E-03
$\mathbb{2}_F$		
	nR06 _{cat}	-2.451E-02
	nCp _{cat}	1.220E-02
	nRCN _{cat}	6.668E-02
	F02[C-C] _{cat}	-2.197E-03
	F02[N-N] _{cat}	-6.073E-03
	F04[C-N] _{cat}	-1.711E-03
	nP _{an}	9.496E-03
	CH3X _{an}	2.427E-02
	F05[C-S] _{an}	3.443E-02
	nCs _{sol}	1.162E-02
$\mathbb{2}_F$		
	nR06 _{cat}	8.114E-03
	nCs _{cat}	-8.140E-04
	CH3R/CH4 _{cat}	-3.429E-03
	F10[C-C] _{cat}	4.818E-04
	nSO4 _{an}	-1.184E-02
	F02[C-C] _{an}	-6.743E-04
	F03[N-F] _{an}	-2.208E-04
	F06[F-F] _{an}	-1.015E-04
	F08[C-C] _{sol}	-2.271E-03

Table 6.58: Definition of parameter of equation (6.14).

No.	Symbol	Definition
1	CH3R / CH4	
2	CH3X	
3	CRX3	
4	R--CH--R	
5	R--CR--R	
6	F01[C-N]	number of C-N
7	F02[C-C]	number of C-A-C
8	F02[C-N]	number of C-A-N
9	F02[C-S]	number of C-A-S

No.	Symbol	Definition
10	F02[F-F]	number of F-A-F
11	F02[N-N]	number of C-A-N
12	F03[N-F]	number of N-A-F
13	F03[N-O]	number of N-(A) ₂ -O
14	F04[C-N]	number of C-(A) ₃ -N
15	F05[C-F]	number of C-(A) ₄ -F
16	F05[C-N]	number of C-(A) ₄ -N
17	F05[C-S]	number of C-(A) ₇ -O
18	F06[C-N]	number of C-(A) ₅ -N
19	F06[F-F]	number of F-(A) ₅ -F
20	F06[O-O]	number of O-(A) ₅ -O
21	F08[C-C]	number of C-(A) ₇ -C
22	F08[C-O]	number of C-(A) ₇ -O
23	F10[C-C]	number of C-(A) ₉ -C
24	F10[C-O]	number of C-(A) ₉ -O
25	nB	number of Boron atoms
26	nCar	Sum of all the carbons belonging to any aromatic and heteroaromatic structure
27	nCp	number of terminal primary Carbon (sp ³)
28	nCrs	number of ring secondary Carbon (sp ³)
29	nCs	number of total secondary Carbon (sp ³)
30	nP	number of Phosphorous atoms
31	nPyridines	number of Pyridines
32	nR06	number of 6-membered rings
33	nRCN	number of nitriles (aliphatic)
34	nROR	number of ethers (aliphatic)
35	nSO4	number of sulfates (thio- / dithio-)
36	R-SH	

R represents any group linked through carbon

X represents any heteroatom (O, N, S, P, Se, halogens)

A represents any atom

-- represents an aromatic bond as in benzene or delocalized bonds such as the N-O bond in a nitro group

The scattered plot of experimental values versus calculated/predicted values for γ^{∞} (γ^{∞}) is shown in Figure 6.59. In addition, Figure 6.60 shows the calculated/predicted values of γ^{∞} in linear scale versus the experimental values of γ^{∞} . According to Table 6.59, the AARD% of the “training” set is 10.44% while it is 10.76% for the “test” set.

To analyze the model output precisely, Figure 6.61 represents the percentage of calculated/predicted data points in different AARD% ranges. Accordingly, 41% of the calculated/predicted values show deviations between 0 to 5%, 25% between 5 to 10%, 20%

between 10 to 20%, 14% over 20 %. Same as the previous models, smaller values of γ^∞ are the source of obtaining large deviated outputs.

Regarding to Figure 6.59 and Figure 6.60, there is a system in “test” set which shows a large deviation from the diagonal line. This system is “1-pentyne” in “1-butyl-1-methylpyrrolidinium thiocyanate”. In the initial steps of modeling, this system was in the “training” set, but it was observed that the models with different number of parameters failed to represent this system. As a result, it was suspected to be an outlier and consequently, it was moved to the “test” set to eliminate its weight on the model developed.

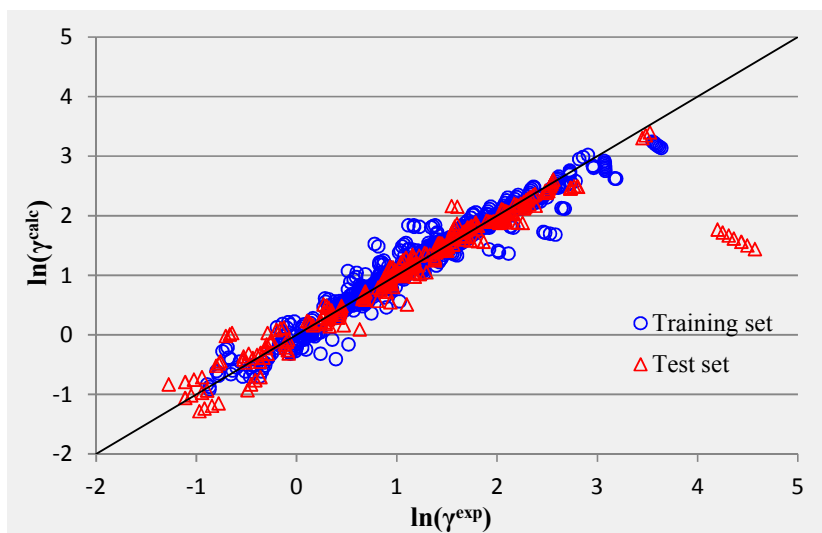


Figure 6.59: Correlated/Predicted versus experimental values of $\ln(\gamma^\infty)$ of alkyne solutes (— diagonal line).

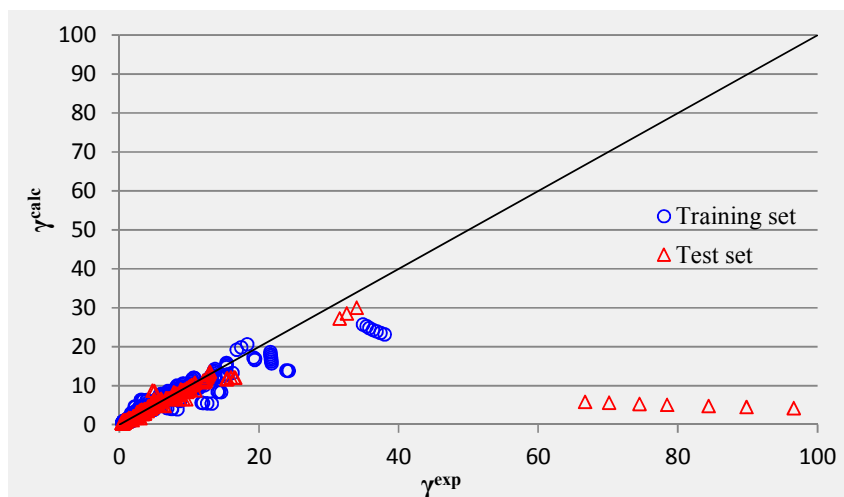


Figure 6.60: Correlated/Predicted versus experimental values of γ^∞ of alkyne solutes (— diagonal line).

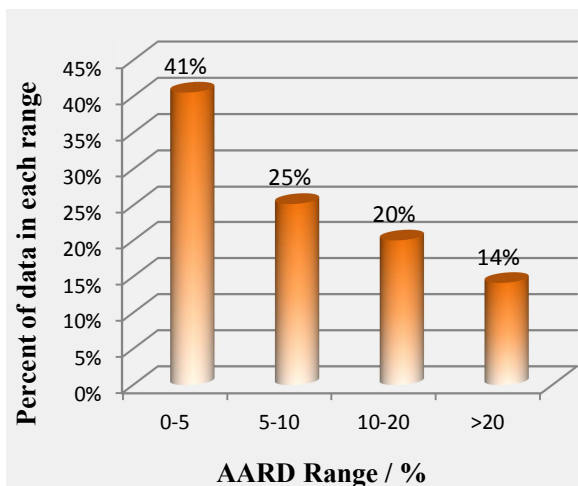


Figure 6.61: Percentage of calculated/predicted values of γ^∞ of alkene solutes in different relative deviation ranges.

Table 6.59: Statistical parameters for equation (6.14)

<i>Statistical Parameter</i>	
<i>training set</i>	
R^2	0.916
Average absolute relative deviation	10.44
Standard deviation error	1.52
Root mean square error	1.53
No. of data points	938
<i>test set</i>	
R^2	0.973
Average absolute relative deviation	10.76
Standard deviation error	0.92
Root mean square error	0.95
No. of data points	305
<i>total</i>	
R^2	0.929
Average absolute relative deviation	10.62
Standard deviation error	1.39
Root mean square error	1.40
No. of data points	1250

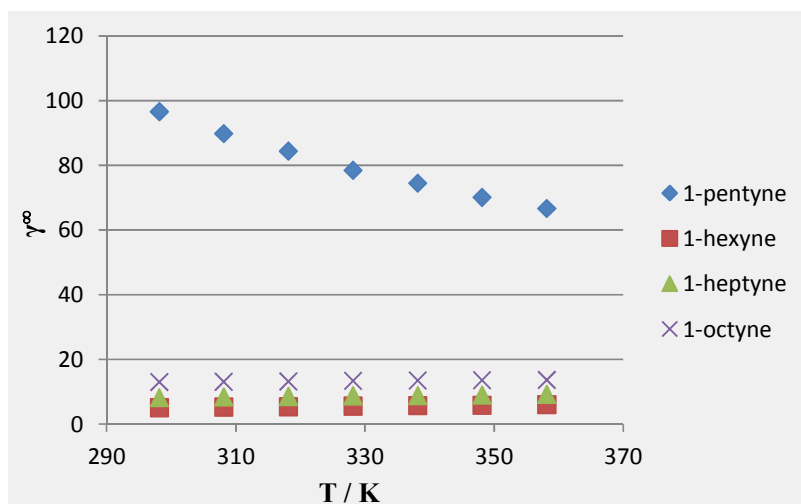


Figure 6.62: The γ^∞ of different alkene solutes in 1-butyl-1-methylpyrrolidinium thiocyanate.

A precise analysis of the γ^∞ data of different solutes in “1-butyl-1-methylpyrrolidinium thiocyanate” is presented in Figure 6.62. This figure shows that the γ^∞ of different alkene solutes in this ionic liquid are less than 20.0 except the abovementioned “1-pentyne” solute. As a result, the γ^∞ data of “1-pentyne” in “1-butyl-1-methylpyrrolidinium thiocyanate” is outlier and the quality of measured data is questionable.

Regarding to Figure 6.60, there are two other systems with “thiocyanate” anion which the presented model fails to predict correctly. These systems are “1-heptyne” and “1-octyne” in “1-ethyl-3-methylimidazolium thiocyanate” which are shown in Figure 6.63. Accordingly, the model fails to predict correctly the change in γ^∞ versus temperature.

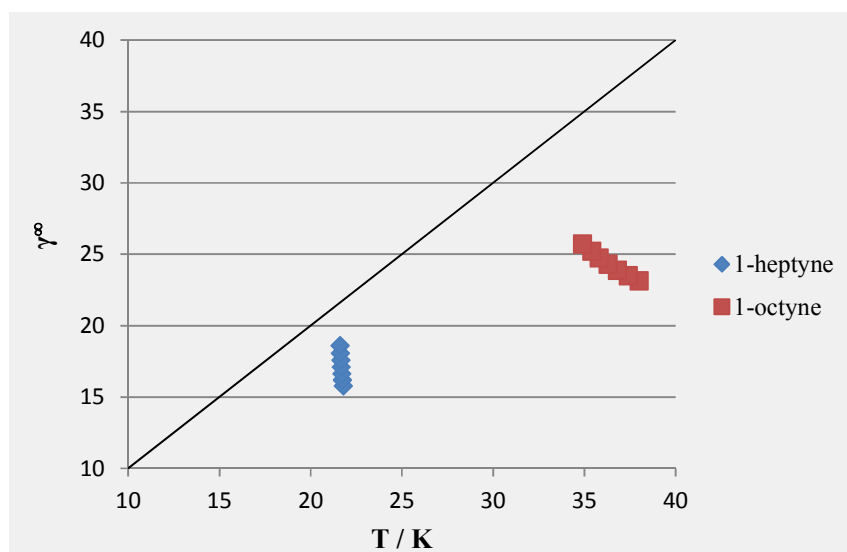


Figure 6.63: Correlated/Predicted versus experimental values of γ^∞ of 1-heptyne and 1-octyne in 1-ethyl-3-methylimidazolium thiocyanate (— diagonal line).

There are two possible reasons for observing such failures:

- Quality of measured data is doubtful.
- There are some unknown interactions among solute, cation, and anion.

Figure 6.64, Figure 6.65, and Figure 6.66 show the correlated/predicted versus experimental values of γ^∞ of different alkene solutes in “1-ethyl-, 1-butyl-, and 1-hexyl-3-methylimidazolium thiocyanate”. These figures demonstrate that the γ^∞ of alkyne solutes is proportional to the number of carbon atoms in the solute molecule and γ^∞ increases as the temperature increases. The only exceptions that the presented model fails to predict are the systems “1-heptyne” and “1-octyne” in “1-ethyl-3-methylimidazolium thiocyanate”.

Despite the similarity in the cation of these three ionic liquids, such behavior is not observed in “1-butyl- and 1-hexyl-3-methylimidazolium thiocyanate” systems. As a result, the quality of reported data is questionable.

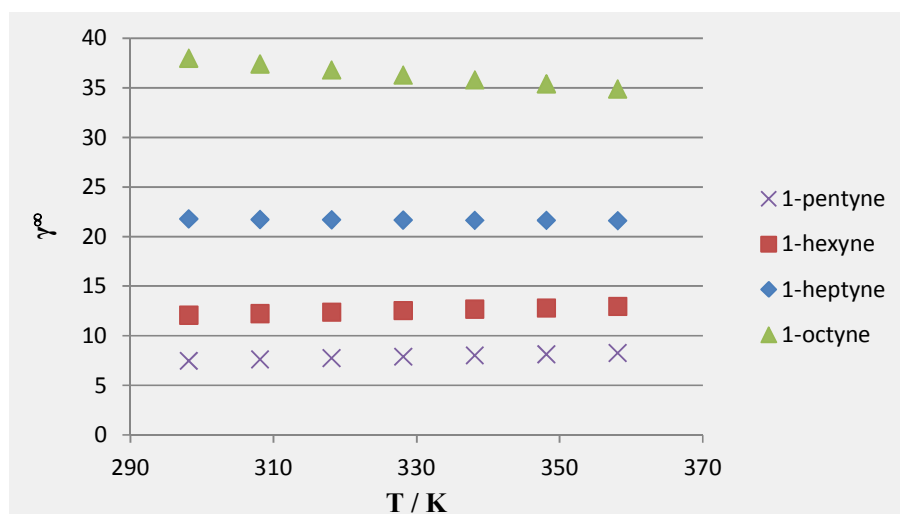


Figure 6.64: The γ^∞ of different alkene solutes in 1-ethyl-3-methylimidazolium thiocyanate.

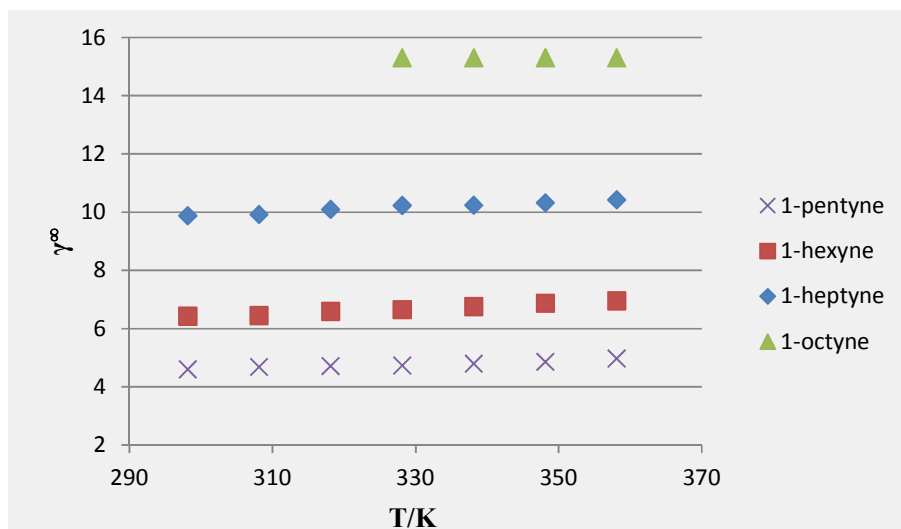


Figure 6.65: The γ^∞ of different alkene solutes in 1-butyl-3-methylimidazolium thiocyanate.

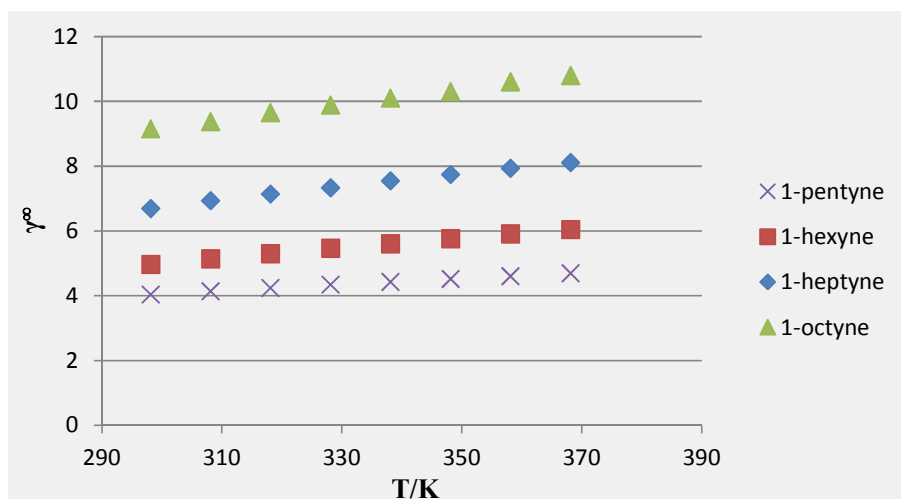


Figure 6.66: The γ^∞ of different alkene solutes in 1-hexyl-3-methylimidazolium thiocyanate.

Ultimately, the AARD% of calculated/predicted values of γ^∞ for different classes of ionic liquids are presented in Table 6.55. The maximum deviation is observed for Phosphonium family with AARD% of 21.2%; however 64% of data points (70 out of 110 data points) are below 1.0 and as explained before, the small values of γ^∞ are very sensitive to small deviations.

Table 6.60: AARD% of equation (6.14) for different class of ionic liquids.

No.	Family	T/K range	γ^∞ range	AARD%	N	N Data
-----	--------	-----------	-----------------------	-------	---	--------

				Compounds	Points	
1	Ammonium	303.15-352.65	0.52-3.567	10.4	29	97
2	Guanidinium	308.15-348.15	1.99-5.24	1.6	4	20
3	Imidazolium	298.15-368.2	0.899-38	13.5	110	482
4	Isoquinolinium	328.15-368.15	1.6-2.44	6.9	3	11
5	Morpholinium	318.15-368.15	2.52-11.4	6.3	7	42
6	Phosphonium	298.15-373.15	0.28-3.19	21.2	29	110
7	Piperidinium	308.15-368.15	1.42-13.7	3.0	24	131
8	Pyridinium	298.15-368.15	1.67-16.2	5.4	24	148
9	Pyrrolidinium	298.15-368.15	1.34-96.61	9.7	36	187
10	Sulphonium	298.15-368.15	2.52-7.87	4.1	4	29

Table 6.61 shows the comparison between output of model proposed and the original and modified UNIFAC models. Accordingly, both forms of the UNIFAC model are limited to 3 systems, but the model presented covers more solute-IL systems. As a result, the alkyne-ILs systems have been modeled successfully and the model developed can be used to calculate/predict the γ^∞ data.

Table 6.61: Summary of result of models for γ^∞ of alkyene solutes in ILs.

Model	Model Type and parameters	N _{Systems}	N _{data}	AARD%	Comments
Original UNIFAC	GC	3	15	68.06	3 solutes in 1 ILs.
Modified UNIFAC	GC	3	15	48.44	3 solutes in 1 ILs.
GC Model (this study)	GC, 37 parameters	270	1257	10.62	6 solutes in 84 ILs.

6.7 Critical temperature of ionic liquids

To assess the possibility of developing a predictive model for estimating the T_c of ionic liquids, a 30-parameter GC and a 25-parameter QSPR models were developed. The GC model has an AARD% of 5.17% over 106 data points, and the QSPR model shows the

AARD% of 4.69%. As a result, the models can calculate the “training” and “test” sets fairly well. The models parameters as well as their outputs are available in the supplementary CD.

To compare the model on other ionic liquids, the T_c for 33 additional ionic liquids were calculated. According to section 3.4.6, these ionic liquids only had one experimental surface tension data point; so it was not possible to use Guggenheim equation (2.13) for estimation of critical temperature. Table 6.62 shows the output of the GC and QSPR models for these ionic liquids. Unfortunately, the coefficients of functional groups were not available in paper published by Valderrama *et al.* [154] and it was not therefore possible to calculate the T_c for these ionic liquids. Only the T_c of some of them was reported by these authors.

According to Table 6.62, for some ionic liquids, there is an agreement between GC, QSPR, and Valderrama *et al.* methods such as “1-hexyl-3-methylimidazolium nitrate”. On the other hand, the T_c of some ionic liquids calculated by the GC and QSPR models is similar, but different from Valderrama’s method (e.g. “1-ethyl-3-methylimidazolium acetate”); and some are different at all (“1-ethyl-3-methylimidazolium bis[(pentafluoroethyl)sulfonyl]amide”). The worst cases are the ionic liquids with “bis[1-methylimidazolium]” cations. The output of the GC and QSPR models are two different values and the trend of changes are not also similar. In addition, the GC model produces the strange negative values of T_c which are obviously meaningless and erroneous. In these cases, the QSPR model has better output compared with the GC model, but the trend of changes are not acceptable.

For ionic liquids with “3,3'-(1,10-decanediyl)bis[1-methylimidazolium]” as the cation and “bis(trifluoromethylsulfonyl)imide”, “tetrafluoroborate”, and “1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide” as the anion, the value of T_c is 915.6 K, 964.9 K, and 908.5 K respectively; but for the “1,1,2,2,2-pentafluoro-N-[(1,1,2,2,2-pentafluoroethyl)sulfonyl]ethanesulfonamide” anion, the T_c is 1961.2 K. It is obvious that 1000 K jump in the value of critical temperature by changing the anion is not correct and acceptable. In this case, the GC model produces negative value, as mentioned before. Unfortunately, Valderrama *et al.* have not calculated the critical temperature for these ionic liquids.

Table 6.62: Comparison of estimated T_c of ionic liquids by different methods.

T_c

No.	Name	Guggenheim	GC Model	QSPR Model	Valderrama et al. [154]
1	1,2-dimethyl-3-propylimidazolium bis[(trifluoromethyl)sulfonyl]amide	n.a.	536.5	543.8	1269.7
2	1-butyl-1-methylpyrrolidinium tris(pentafluoroethyl)trifluorophosphate	n.a.	1058.0	1037.2	n.a.
3	1-butyl-3-methylimidazolium nitrate	n.a.	1102.1	1174.2	954.8
4	1-dodecyl-3-methylimidazolium hexafluorophosphate	n.a.	901.2	829.1	857.6
5	1-dodecyl-3-methylimidazolium tetrafluoroborate	n.a.	901.2	822.5	784.6
6	1-ethyl-3-methylimidazolium acetate	n.a.	1400.7	1358.7	807.1
7	1-ethyl-3-methylimidazolium bis[(pentafluoroethyl)sulfonyl]amide	n.a.	492.3	1732.5	1231.4
8	1-ethyl-3-methylimidazolium hexylsulfate	n.a.	870.9	1025.5	n.a.
9	1-ethyl-3-methylimidazolium octyl sulfate	n.a.	870.9	892.4	n.a.
10	1-hexyl-3-methylimidazolium nitrate	n.a.	991.9	1044.1	991.8
11	1-isobutenyl-3-methylimidazolium tetrafluoroborate	n.a.	561.5	1133.3	n.a.
12	1-methyl-3-pentylimidazolium 1,1,2,2,2-pentafluoro-N-[(1,1,2,2,2-pentafluoroethyl)sulfonyl]ethanesulfonamide	n.a.	371.4	1616.6	n.a.
13	1-methyl-3-pentylimidazolium nitrate	n.a.	990.8	1082.0	n.a.
14	1-methyl-3-propylimidazolium bis(pentafluoroethylsulfonyl)imide	n.a.	466.0	1642.8	n.a.
15	1-methyl-3-propylimidazolium nitrate	n.a.	1085.5	1108.1	n.a.
16	1-octyl-3-methylimidazolium bromide	n.a.	980.9	981.3	912.3
17	3,3'-(1,10-decanediyl)bis[1-methylimidazolium] 1,1,2,2,2-pentafluoro-N-[(1,1,2,2,2-pentafluoroethyl)sulfonyl]ethanesulfonamide	n.a.	-825.9	1961.2	n.a.
18	3,3'-(1,10-decanediyl)bis[1-methylimidazolium] bis(trifluoromethylsulfonyl)imide	n.a.	556.3	915.6	n.a.
19	3,3'-(1,10-decanediyl)bis[1-methylimidazolium] tetrafluoroborate	n.a.	562.9	964.9	n.a.
20	3,3'-(1,12-dodecanediyl)bis[1-methylimidazolium] 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	n.a.	531.0	908.5	n.a.
21	3,3'-(1,12-dodecanediyl)bis[1-methylimidazolium] 1,1,2,2,2-pentafluoro-N-[(1,1,2,2,2-pentafluoroethyl)sulfonyl]ethanesulfonamide	n.a.	-851.2	1955.6	n.a.
22	3,3'-(1,12-dodecanediyl)bis[1-methylimidazolium] tetrafluoroborate	n.a.	537.6	957.8	n.a.
23	3,3'-(1,5-pentanediy)bis[1-methylimidazolium] 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	n.a.	755.0	896.9	n.a.
24	3,3'-(1,5-pentanediy)bis[1-methylimidazolium] 1,1,2,2,2-pentafluoro-N-[(1,1,2,2,2-pentafluoroethyl)sulfonyl]ethanesulfonamide	n.a.	-627.2	1942.5	n.a.
25	3,3'-(1,6-hexanediy)bis[1-methylimidazolium] 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	n.a.	636.0	1133.0	n.a.
26	3,3'-(1,6-hexanediy)bis[1-methylimidazolium] 1,1,2,2,2-pentafluoro-N-[(1,1,2,2,2-pentafluoroethyl)sulfonyl]ethanesulfonamide	n.a.	-746.2	2181.1	n.a.
27	3,3'-(1,8-octanediy)bis[1-methylimidazolium] 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	n.a.	610.7	892.5	n.a.
28	3,3'-(1,9-nonanediy)bis[1-methylimidazolium] 1,1,1-trifluoro-N-[(trifluoromethyl)sulfonyl]methanesulfonamide	n.a.	569.0	907.6	n.a.
29	3,3'-(1,9-nonanediy)bis[1-methylimidazolium] 1,1,2,2,2-pentafluoro-N-[(1,1,2,2,2-	n.a.	-813.2	1955.2	n.a.

No.	Name	T_c			
		Guggenheim	GC Model	QSPR Model	Valderrama et al. [154]
	pentafluoroethyl)sulfonyl]ethanesulfonamide				
30	3,3'-(1,9-nonanediyl)bis[1-methylimidazolium] tetrafluoroborate	n.a.	575.5	956.9	n.a.
31	3-ethyl-1-methylimidazolium butyl sulfate	n.a.	1068.9	1486.6	n.a.
32	3-hexyl-1-methylimidazolium 1,1,2,2,2-pentafluoro-N-[(1,1,2,2,2-pentafluoroethyl)sulfonyl]ethanesulfonamide	n.a.	372.4	1576.8	n.a.
33	butylammonium formate	n.a.	1441.2	1093.4	546.7

As the number of data points used in the model development procedure is not high enough, it is not possible to develop a model with a wide range of applicability; so such deviations are expected and inevitable.

The outputs of the GC model as well as the QSPR one have been compared with the Valderrama's method in the supplementary CD. According to the data provided, the results can be visualized for specific cation and different anions. For example, Figure 6.67 shows the critical temperatures calculated by Guggenheim equation, GC, QSPR, and Valderrama's method. Accordingly, it is obvious that Valderrama's model fails to predict the critical temperature of "1-ethyl-3-methylimidazolium" ionic liquids. Despite the deviations in values of T_c , that model also fails to estimate the T_c in a reasonable trend according to the anions type. Based on the experimental data of surface tension of "1-ethyl-3-methylimidazolium" ionic liquids, the critical temperature is related to the anion as follows; but Valderrama's model fails to represent this relationship.

[DEP] < [NTf2] < [EtSO4] < [DCA] < [BF4] < [TfO]

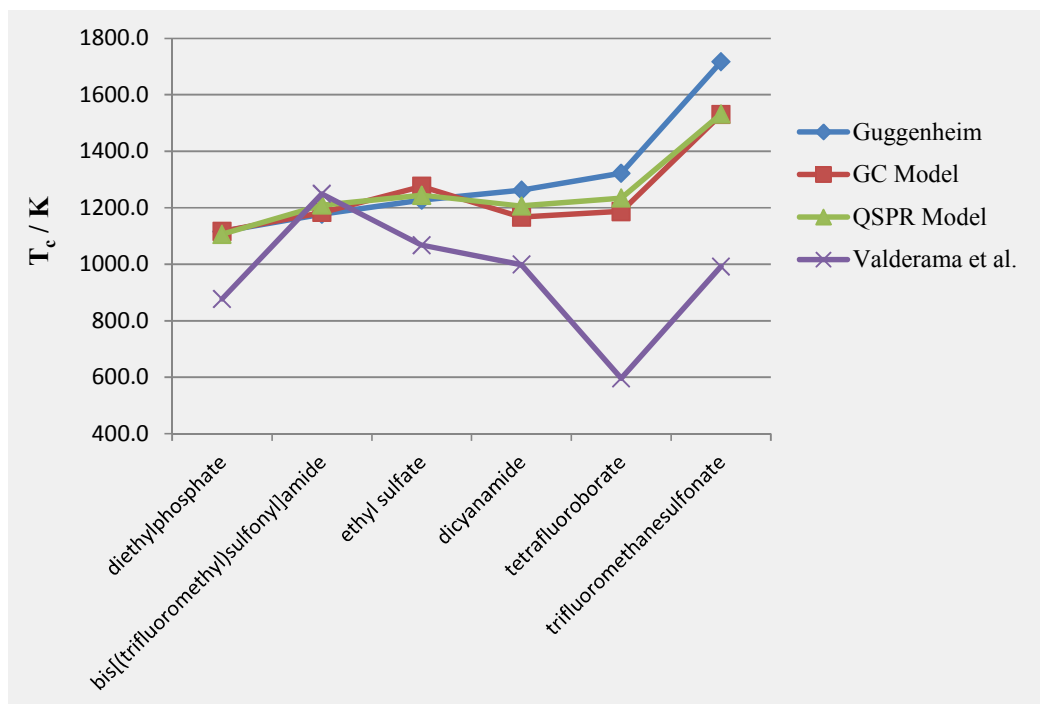


Figure 6.67: Estimated T_c of 1-ethyl-3-methylimidazolium ILs by different models.

In another comparison, the T_c of “1-butyl-3-methylimidazolium” ionic liquids are compared and the results are illustrated in Figure 6.68. Accordingly, similar fluctuations are observed in Valderrama’s model which indicates that the model does not produce reliable results.

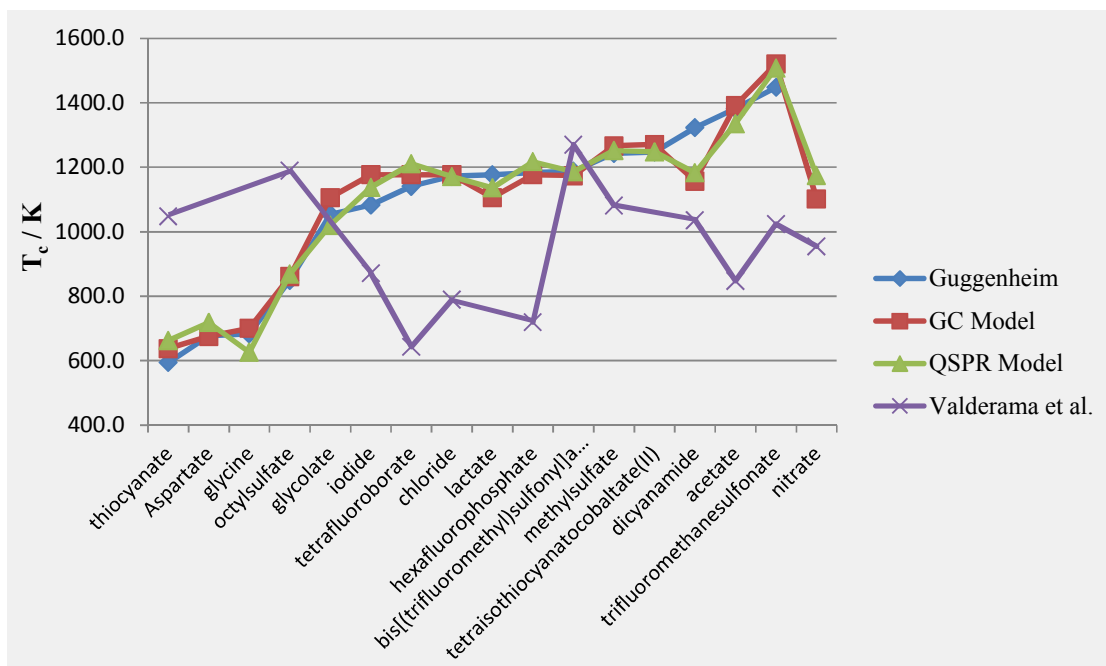


Figure 6.68: Estimated T_c of 1-butyl-3-methylimidazolium ILs by different models.

According to the published papers by Valderrama *et al.* [46, 154, 250], they used a density correlation to validate the calculated critical properties and normal boiling point of ionic liquids which had been developed for saturated liquids and petroleum fractions [153].

6.15

$$\rho = \frac{\rho_c}{\rho_b} + \frac{2}{7} \rho_c \left\{ \frac{(\rho_c - \rho_b)}{\rho_c} \right\} \frac{\rho - \rho_b}{\rho_c - \rho_b}$$

$$\rho = \rho_c + \rho_c^2 / \rho_b$$

$$\rho = \left(\frac{\rho_c}{\rho_b} + \frac{\rho_c}{\rho_b} \right) \rho_b^2$$

$$\rho_c = 0.3411, \quad \rho_b = 2.0443, \quad \rho = 0.5386$$

$$\rho_c = 0.0393, \quad \rho_b = 1.0476$$

It was reported that the density calculated using the above-mentioned correlation had average deviation of less than 19%. As a result, Valderrama *et al.* concluded that the estimated critical properties as well as the normal boiling point were valid and reliable. In this thesis, it was

shown that the critical temperatures calculated by Valderrama's model were not valid and reliable compared with the T_c calculated using the experimental data of surface tension of ionic liquids. Thus, the estimated T_b , V_c , other critical properties, and the validation technique is questionable and these data are not reliable. In recent years, some corresponding state models [155-162] have been published to model the thermophysical properties of ionic liquids using the critical properties estimated by Valderrama *et al.* model. According to the aforementioned results, these models are also questionable and are not reliable. So developing the models for critical properties of ionic liquids and using them for developing the corresponding state models need more research and considerations.

As discussed earlier, the QSPR model developed seems to be more reliable than the GC one as the later calculates the negative T_c values for some ionic liquids. Thus, the QSPR model is chosen to calculate the T_c of 1130 ionic liquids which has been used by Valderrama *et al.* [154]. According to the data provided in the supplementary CD, there are some ionic liquids that the QSPR model produces negative or very high values of T_c . As discussed before, it is not possible to develop a comprehensive model using only few number of data points. In this thesis, the QSPR model has been developed using 41 cations and 34 anions, but the Valderrama's database consists of 484 cations and 113 anions. It's completely expectable that the QSPR model developed fails to calculate proper value for some ionic liquids that their cation and/or anion have not been in the "training" subset.

To have a precise comparison, the calculated T_c of "1-ethyl-3-methylimidazolium" ionic liquids with 45 different anions is shown in Figure 6.69. According to Figure 6.67, the average critical temperature of "1-ethyl-3-methylimidazolium" ionic liquids should be approximately 1200 K which is shown by a red tie line in Figure 6.69. Accordingly, the QSPR model represents the better results and only four strange values are observed; two for around 700 K and two for over 1800 K. The Valderrama's model has more fluctuations between 600 K and 1000 K. Despite the better estimated results of the QSPR model proposed, both models are not reliable to be used for developing the corresponding state models; however the QSPR model seems to have better predictions for the ionic liquids which their cation and anion are present in the "training" set.

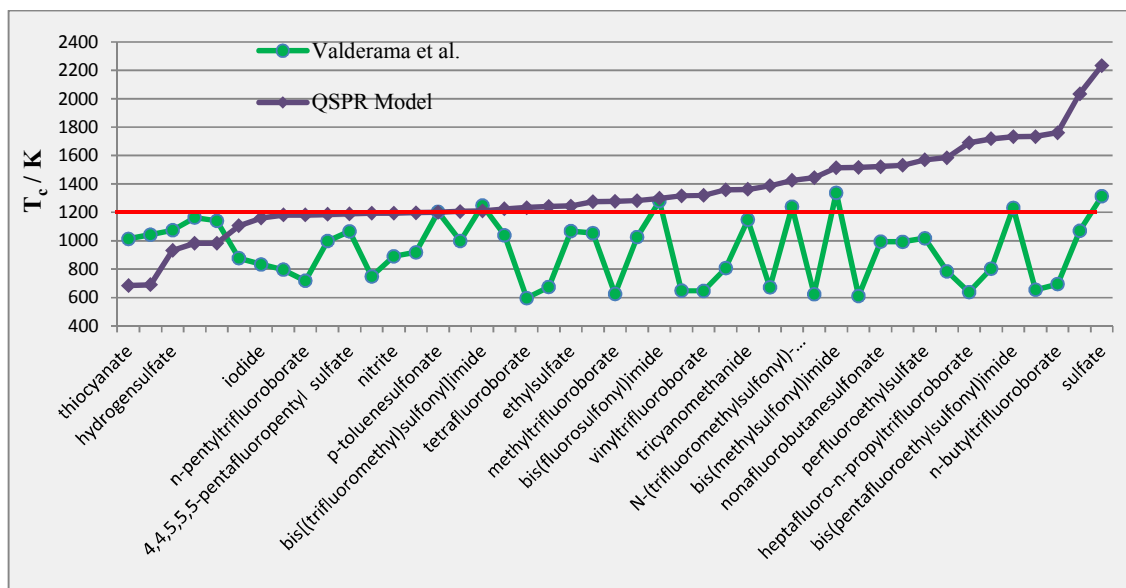


Figure 6.69: T_c of 1-ethyl-3-methylimidazolium ionic liquids with 45 different anions calculated by the QSPR and Valderrama *et al.* models.

At this stage, only the critical temperature of ionic liquids can be estimated which highly depends of the quality of measurements and uncertainty of surface tension data. In addition, further model developments requires larger dataset of experimental data of surface tension and at this time, lack of enough data leads to the questionable models in term of prediction ability.

CHAPTER 7: CONCLUSION

The objective of this thesis was to use the largest possible databases and different computational methods to represent and predict the physico-chemical properties of ionic liquids. To assess the effects of computational methods on quality of models developed, several thermophysical properties of ionic liquids were modeled by combining two well-known property estimation methods (GC and QSPR) with different mathematical regression techniques. In addition, the larger databases were used compared with the models published previously in the literature, and the quality of fit and predictions were improved significantly.

Speed of sounds in ionic liquids was the first property studied. One GC and one QSPR models were developed using the same database comprised of the experimental data for 41 ILs. The GC model was developed using FFS-LSSVM method and its AARD% was 0.36%. Thereafter, the QSPR model was developed using the GFA method with an AARD% of 0.92%. The results showed that both models had better fit and prediction ability compared with the best model published by Gardas and Coutinho [46] which had an AARD% of 1.96% for only 14 ILs.

Liquid heat capacity of ionic liquids was chosen afterwards, and GFA method was used to develop both GC and QSPR models using the experimental data for 82 ionic liquids. It was found that binary multiplication of variables resulted in better prediction for the models. As a result, the GC and QSPR models had the AARD% of 1.68% and 1.70%, respectively. Both models were more comprehensive regarding to the model developed by Soriano *et al.* [53] which had an AARD% of 0.34% using only 32 ionic liquids.

Refractive index of ionic liquids was the third property to examine the applicability of the previous approaches. In this regard, the experimental data for 97 ionic liquids were collected and through the GFA method, one GC and one QSPR model were developed. The models had the AARD% of 0.34% and 0.51%, respectively. In terms of number of covered ionic liquids, both models were more comprehensive regarding to the best previous model with the AARD% of 0.18% for 24 ionic liquids [61].

For the viscosity of fluorine-containing ionic liquids, the available experimental data were used to create two different databases. The first one was screened carefully and unreliable data points were removed; so the refined database contained the experimental data points for

85 F-ILs and it was used to develop one GC and one QSPR model. On the other hand, the second database included the unreliable data points for additional 247 different F-ILs and it was used to develop another GC model in order to have a comprehensive model for a wider range of ILs. All of the models had the AARD% of below 5% for $\ln(\eta)$ which was less than the best model published by Gharagheizi *et al.* [89] (AARD% of 7.1%).

The next reviewed property was infinite dilution activity coefficient (γ^∞) of different organic solutes in ionic liquids. Developing a single model for all of the solutes was impossible and consequently, the data of γ^∞ was split into several parts for aromatic, alcohol, alkane, alkene, and alkyne solutes.

The dataset of aromatic solutes consisted of 1654 data points with 10 solutes and 123 ionic liquids which resulted in 354 different solute-IL systems. The result of the GFA method was a GC model with an AARD% of 9.69%. This model was much more simple and easier to use compared with the UNIFAC model as it required the interaction parameter of chemical subgroups in the molecules. The original UNIFAC model was able to predict only 9 systems with the AARD% of 64.07%. The modified UNIFAC model could predict the γ^∞ of 135 systems; however the prediction error was 174.40%.

For the alcohol solutes, 2785 data points for 615 solute-IL systems (17 solutes and 126 ILs) were used and through a GFA method, a GC model was developed with an AARD% of 14.63%. The original UNIFAC model could calculate the γ^∞ for only 47 systems with an AARD% of 27.34%. The AARD% of the modified UNIFAC model was 32.22% for 164 systems.

The database of alkane solutes was divided into two subsets: one for solutes with less than 10 carbon atoms in their structure, and one for solutes with equal or more than 10 carbon atoms. Using the GFA approach, two different models were developed with the AARD% of 20.81% over the entire database of alkane solutes (882 solute-IL systems). The original UNIFAC model was able to calculate the γ^∞ for only 96 systems with an AARD% of 331.04%. The AARD% of the modified UNIFAC model was 1927.49% for 293 systems.

For the alkene solutes, 422 systems were studied and the GC model was developed with an AARD% of 14.54%. The original UNIFAC had the prediction error of 36.80% for only 49 systems. The AARD% of the modified UNIFAC model was 70.74% for 111 systems.

As the last γ^∞ database, 270 systems of alkyne solute in ionic liquids were studied and the resultant GC model had an AARD% of 10.62%. Unfortunately, both the original and modified UNIFAC models could calculate the γ^∞ for only 3 systems with the AARD% of 68.06% and 48.44%, respectively.

Consequently, the GC models developed for different solutes were more comprehensive, accurate, and applicable compared with the original and modified UNIFAC models.

Finally, the theoretical critical temperature of ionic liquids was calculated using the experimental data of surface tension of ionic liquids by means of the Guggenheim equation. Thereafter, two different GC and QSPR models were developed to assess the prediction ability of the models and compare the results with the pioneering work of Valderrama *et al.* [46, 154, 250]. The study revealed that the QSPR model had better results compared with the GC one in terms of prediction and production of the meaningful results. In addition, it was shown that the critical temperatures calculated by Valderrama's model had the relatively large deviations compared with values driven from surface tension data. Furthermore, it was discussed that the validation method used by Valderrama *et al.* was questionable and thus, the reported critical properties were not reliable enough to be used for developing the corresponding state models for different thermophysical properties of ionic liquids.

As studied in this thesis, the GC method was easier to use in terms of manual calculation of variables. This method was able to model the properties studied with reasonable number of variables and relatively good accuracy. On the other hand, the QSPR method demonstrated its ability to correlate the target property better than the group contribution method with less number of parameters; however it was required the Dragon software for calculating the descriptors.

This thesis results in successful modeling of several thermophysical properties of ionic liquids using different computational methods. The models developed improve the quality of prediction for larger number of ionic liquids in comparison with the currently available models in the literature. In addition, these models, specially the models developed for prediction of γ^∞ for organic solutes in ionic liquids, are easier to use and more comprehensive and consequently, more applicable.

CHAPTER 8: RECOMMENDATIONS FOR FURTHER STUDIES

According to the properties and techniques studied, there are some opportunities for future studies.

1. The QSPR method can be used to develop the smaller and/or more accurate models for the prediction of σ^o of solutes in ionic liquids; however these models may not being accepted by some researchers due to high number of descriptors used.
2. Several new descriptors are developed periodically. Using the new version of Dragon software provides more than 1000 new descriptors which may increase the accuracy and prediction ability of the models developed.
3. The provided techniques can be applied to develop models for other properties of binary mixture of ionic liquids and organic compounds such as binary diffusion coefficient, viscosity, heat capacity, etc.; however at the moment, the number of experimental data points are not high enough for model development.
4. Different nonlinear regression methods such as GEP can be applied on results of linear modeling which may improve the accuracy of the models for complex and nonlinear behavior of ionic liquids; however it need more powerful computers and lots of time to perform the computations and consequently, it was ignored due to having limited amount of time during this PhD course.
5. The calculation of critical temperature as well as other critical properties of ionic liquids needs some revisions. It is required to gather the more comprehensive data set for surface tension data of ionic liquids for calculation of the critical temperature. So one can develop the better model in future when the database of surface tension of ionic liquids becomes large enough.

REFERENCES

- [1] M. Lancaster, *Green Chemistry: An Introductory Text*, Royal Society of Chemistry, 2010.
- [2] J.R. Bell, H. Luo, S. Dai, Superbase-derived protic ionic liquid extractants for metal ion separation, *Sep. Purif. Technol.*
- [3] G. Wypych, *Handbook of Solvents*, ChemTec Publishing, Toronto, 2001.
- [4] M. Sattari, F. Gharagheizi, P. Ilani-Kashkouli, A. Mohammadi, D. Ramjugernath, Development of a group contribution method for the estimation of heat capacities of ionic liquids, *J. Therm. Anal. Calorim.*, 115 (2014) 1863-1882.
- [5] M. Larriba, P. Navarro, J. García, F. Rodríguez, Liquid-Liquid Extraction of Toluene from Heptane Using [emim][DCA], [bmim][DCA], and [emim][TCM] Ionic Liquids, *Industrial & Engineering Chemistry Research*, 52 (2013) 2714-2720.
- [6] T.V. Hoogerstraete, B. Onghena, K. Binnemans, Homogeneous Liquid-Liquid Extraction of Metal Ions with a Functionalized Ionic Liquid, *The Journal of Physical Chemistry Letters*, 4 (2013) 1659-1663.
- [7] A. Lewandowski, A. Świdarska-Mocek, Ionic liquids as electrolytes for Li-ion batteries—An overview of electrochemical studies, *J. Power Sources*, 194 (2009) 601-609.
- [8] S. Menne, J. Pires, M. Anouti, A. Balducci, Protic ionic liquids as electrolytes for lithium-ion batteries, *Electrochem. Commun.*, 31 (2013) 39-41.
- [9] R.F. de Souza, J.C. Padilha, R.S. Gonçalves, J. Dupont, Room temperature dialkylimidazolium ionic liquid-based fuel cells, *Electrochem. Commun.*, 5 (2003) 728-731.
- [10] B. Lin, L. Qiu, J. Lu, F. Yan, Cross-Linked Alkaline Ionic Liquid-Based Polymer Electrolytes for Alkaline Fuel Cell Applications, *Chem. Mater.*, 22 (2010) 6718-6725.
- [11] V. Jovanovski, V. González-Pedro, S. Giménez, E. Azaceta, G. Cabañero, H. Grande, R. Tena-Zaera, I. Mora-Seró, J. Bisquert, A Sulfide/Polysulfide-Based Ionic Liquid Electrolyte for Quantum Dot-Sensitized Solar Cells, *J. Am. Chem. Soc.*, 133 (2011) 20156-20159.
- [12] G.-H. Kim, H.-B. Kim, B. Walker, H. Choi, C. Yang, J. Park, J.Y. Kim, Effects of Ionic Liquid Molecules in Hybrid PbS Quantum Dot–Organic Solar Cells, *ACS Applied Materials & Interfaces*, 5 (2013) 1757-1760.
- [13] D. Holbrey John, J. Chen, B. Turner Megan, P. Swatloski Richard, K. Spear Scott, D. Rogers Robin, Applying Ionic Liquids for Controlled Processing of Polymer Materials, in: *Ionic Liquids in Polymer Systems*, American Chemical Society, 2005, pp. 71-87.
- [14] M.H. Luke, P.F. Matthew, E.K. Brown, M.F. Douglas, C.D.L. Hugh, C.T. Paul, Natural Fiber Welding: Ionic Liquid Facilitated Biopolymer Mobilization and Reorganization, in: *Ionic Liquids: Science and Applications*, American Chemical Society, 2012, pp. 145-166.
- [15] P.K. Singh, B. Bhattacharya, R.K. Nagarale, K.-W. Kim, H.-W. Rhee, Synthesis, characterization and application of biopolymer-ionic liquid composite membranes, *Synth. Met.*, 160 (2010) 139-142.

- [16] Q.X. Liu, S.Z. El Abedin, F. Endres, Electroplating of mild steel by aluminium in a first generation ionic liquid: A green alternative to commercial Al-plating in organic solvents, *Surf. Coat. Technol.*, 201 (2006) 1352-1356.
- [17] A.P. Abbott, G. Frisch, K.S. Ryder, Electroplating Using Ionic Liquids, *Annual Review of Materials Research*, 43 (2013) 335-358.
- [18] M. Uerdingen, Ionic Liquids as Lubricants, in: *Handbook of Green Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, 2010.
- [19] V. Khare, M.-Q. Pham, N. Kumari, H.-S. Yoon, C.-S. Kim, J.-I.L. Park, S.-H. Ahn, Graphene–Ionic Liquid Based Hybrid Nanomaterials as Novel Lubricant for Low Friction and Wear, *ACS Applied Materials & Interfaces*, 5 (2013) 4063-4075.
- [20] B.A. Kheireddin, W. Lu, I.C. Chen, M. Akbulut, Inorganic nanoparticle-based ionic liquid lubricants, *Wear*, 303 (2013) 185-190.
- [21] A. Somers, P. Howlett, D. MacFarlane, M. Forsyth, A Review of Ionic Liquid Lubricants, *Lubricants*, 1 (2013) 3-21.
- [22] M. Han, W. Yi, Q. Wu, Y. Liu, Y. Hong, D. Wang, Preparation of biodiesel from waste oils catalyzed by a Brønsted acidic ionic liquid, *Bioresour. Technol.*, 100 (2009) 2308-2310.
- [23] L. Fischer, T. Falta, G. Koellensperger, A. Stojanovic, D. Kogelnig, M. Galanski, R. Krachler, B.K. Keppler, S. Hann, Ionic liquids for extraction of metals and metal containing compounds from communal and industrial waste water, *Water Res.*, 45 (2011) 4601-4614.
- [24] M.T. Hamed Mosavian, Z. Es'haghi, N. Razavi, S. Banihashemi, Pre-concentration and determination of amitriptyline residues in waste water by ionic liquid based immersed droplet microextraction and HPLC, *Journal of Pharmaceutical Analysis*, 2 (2012) 361-365.
- [25] X. Zeng, J. Li, H. Xie, L. Liu, A novel dismantling process of waste printed circuit boards using water-soluble ionic liquid, *Chemosphere*, In Press (2013).
- [26] D. Camper, J.E. Bara, D.L. Gin, R.D. Noble, Room-Temperature Ionic Liquid–Amine Solutions: Tunable Solvents for Efficient and Reversible Capture of CO₂, *Industrial & Engineering Chemistry Research*, 47 (2008) 8496-8498.
- [27] M. Gonzalez-Miquel, J. Palomar, S. Omar, F. Rodriguez, CO₂/N₂ Selectivity Prediction in Supported Ionic Liquid Membranes (SILMs) by COSMO-RS, *Industrial & Engineering Chemistry Research*, 50 (2011) 5739-5748.
- [28] P.T. Nguyen, B.A. Voss, E.F. Wiesenauer, D.L. Gin, R.D. Noble, Physically Gelled Room-Temperature Ionic Liquid-Based Composite Membranes for CO₂/N₂ Separation: Effect of Composition and Thickness on Membrane Properties and Performance, *Industrial & Engineering Chemistry Research*, 52 (2012) 8812-8821.
- [29] T.K. Carlisle, E.F. Wiesenauer, G.D. Nicodemus, D.L. Gin, R.D. Noble, Ideal CO₂/Light Gas Separation Performance of Poly(vinylimidazolium) Membranes and Poly(vinylimidazolium)-Ionic Liquid Composite Films, *Industrial & Engineering Chemistry Research*, 52 (2012) 1023-1032.

- [30] J. Chau, G. Obuskovic, X. Jie, T. Mulukutla, K.K. Sirkar, Solubilities of CO₂ and Helium in an Ionic Liquid Containing Poly(amidoamine) Dendrimer Gen 0, *Industrial & Engineering Chemistry Research*, 52 (2013) 10484-10494.
- [31] J. Zhang, C. Jia, H. Dong, J. Wang, X. Zhang, S. Zhang, A Novel Dual Amino-Functionalized Cation-Tethered Ionic Liquid for CO₂ Capture, *Industrial & Engineering Chemistry Research*, 52 (2013) 5835-5841.
- [32] J.R. Switzer, A.L. Ethier, K.M. Flack, E.J. Biddinger, L. Gelbaum, P. Pollet, C.A. Eckert, C.L. Liotta, Reversible Ionic Liquid Stabilized Carbamic Acids: A Pathway Toward Enhanced CO₂ Capture, *Industrial & Engineering Chemistry Research*, (2013).
- [33] J.R. Switzer, A.L. Ethier, K.M. Flack, E.J. Biddinger, L. Gelbaum, P. Pollet, C.A. Eckert, C.L. Liotta, Reversible Ionic Liquid Stabilized Carbamic Acids: A Pathway Toward Enhanced CO₂ Capture, *Industrial & Engineering Chemistry Research*, 52 (2013) 13159–13163.
- [34] D. Zhao, M. Wu, Y. Kou, E. Min, Ionic liquids: applications in catalysis, *Catal. Today*, 74 (2002) 157-189.
- [35] P. Wasserscheid, M. Haumann, Catalyst Recycling Using Ionic Liquids, in: D. Cole-Hamilton, R. Tooze (Eds.) *Catalyst Separation, Recovery and Recycling*, Springer Netherlands, 2006, pp. 183-213.
- [36] H. Olivier-Bourbigou, L. Magna, D. Morvan, Ionic liquids and catalysis: Recent progress from knowledge to applications, *Applied Catalysis A: General*, 373 (2010) 1-56.
- [37] J. Stoimenovski, D. MacFarlane, K. Bica, R. Rogers, Crystalline vs. Ionic Liquid Salt Forms of Active Pharmaceutical Ingredients: A Position Paper, *Pharm. Res.*, 27 (2010) 521-526.
- [38] J.P. Gutiérrez, G.W. Meindersma, A.B. de Haan, COSMO-RS-Based Ionic-Liquid Selection for Extractive Distillation Processes, *Industrial & Engineering Chemistry Research*, 51 (2012) 11518-11529.
- [39] M. Liu, S. Jia, Y. Gong, C. Song, X. Guo, Effective Hydrolysis of Cellulose into Glucose over Sulfonated Sugar-Derived Carbon in an Ionic Liquid, *Industrial & Engineering Chemistry Research*, 52 (2013) 8167-8173.
- [40] X. Liang, Novel Efficient Procedure for Biodiesel Synthesis from Waste Oils Using Solid Acidic Ionic Liquid Polymer As the Catalyst, *Industrial & Engineering Chemistry Research*, 52 (2013) 6894-6900.
- [41] S. Gabriel, J. Weiner, Ueber einige Abkömmlinge des Propylamins, *Berichte der deutschen chemischen Gesellschaft*, 21 (1888) 2669-2679.
- [42] U. Onken, J. Rarey-Nies, J. Gmehling, The Dortmund Data Bank: A computerized system for retrieval, correlation, and prediction of thermodynamic properties of mixtures, *Int. J. Thermophys.*, 10 (1989) 739-747.
- [43] M. Frenkel, R.D. Chirico, V. Diky, C.D. Muzny, A.F. Kazakov, J.W. Magee, I.M. Abdulagatov, K. Kroenlein, C.A. Diaz-Tovar, J.W. Kang, R. Gani, ThermoData Engine, NIST Standard Reference Database #103b (Pure Compounds, Binary Mixtures, and Chemical

Reactions), version 7.0, in, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA, <http://www.nist.gov/srd/nist103b.cfm>, 2011.

[44] A.A. Kline, C.R. Szydlak, T.N. Rogers, M.E. Mullins, An overview of compiling, critically evaluating, and delivering reliable physical property data from AIChE DIPPR® Projects 911 and 912, *Fluid Phase Equilib.*, 150–151 (1998) 421-428.

[45] B. Poling, J. Prausnitz, J.O. Connell, *The Properties of Gases and Liquids*, McGraw-Hill Education, 2000.

[46] J.O. Valderrama, P.A. Robles, Critical Properties, Normal Boiling Temperatures, and Acentric Factors of Fifty Ionic Liquids, *Industrial & Engineering Chemistry Research*, 46 (2007) 1338-1344.

[47] C.W. Lin, J.P.M. Trusler, The speed of sound and derived thermodynamic properties of pure water at temperatures between (253 and 473) K and at pressures up to 400 MPa, *The Journal of Chemical Physics*, 136 (2012) 094511-094511.

[48] R.L. Gardas, J.A.P. Coutinho, Estimation of speed of sound of ionic liquids using surface tensions and densities: A volume based approach, *Fluid Phase Equilib.*, 267 (2008) 188-192.

[49] A.F. Estrada-Alexanders, D. Justo, New method for deriving accurate thermodynamic properties from speed-of-sound, *The Journal of Chemical Thermodynamics*, 36 (2004) 419-429.

[50] I.A. Johnston, The Noble-Abel Equation of State: Thermodynamic Derivations for Ballistics Modelling, in, Defence Science And Technology Organisation Edinburgh (Australia), Weapons Systems Division, 2005.

[51] R. Auerbach, Oberflächenspannung und Schallgeschwindigkeit, *Cellular and Molecular Life Sciences*, 4 (1948) 473-474.

[52] M.P. Singh, R.K. Singh, Correlation between ultrasonic velocity, surface tension, density and viscosity of ionic liquids, *Fluid Phase Equilib.*, 304 (2011) 1-6.

[53] M. Borhani zarandi, H. Amrollahi Bioki, Z.A. Mirbagheri, F. Tabbakh, G. Mirjalili, Effect of crystallinity and irradiation on thermal properties and specific heat capacity of LDPE & LDPE/EVA, *Appl. Radiat. Isot.*, 70 (2012) 1-5.

[54] R.L. Gardas, J.o.A.P. Coutinho, A Group Contribution Method for Heat Capacity Estimation of Ionic Liquids, *Industrial & Engineering Chemistry Research*, 47 (2008) 5751-5757.

[55] A.N. Soriano, A.M. Agapito, L.J.L.I. Lagumbay, A.R. Caparanga, M.-H. Li, A simple approach to predict molar heat capacity of ionic liquids using group-additivity method, *Journal of the Taiwan Institute of Chemical Engineers*, 41 (2010) 307-314.

[56] J.O. Valderrama, R.E. Rojas, Mass connectivity index, a new molecular parameter for the estimation of ionic liquid properties, *Fluid Phase Equilib.*, 297 (2010) 107-112.

[57] M. Randic, Characterization of molecular branching, *J. Am. Chem. Soc.*, 97 (1975) 6609-6615.

- [58] J.O. Valderrama, G. Martinez, C. Faúndez, Heat Capacity of Ionic Liquids Using Artificial Neural Networks and the Concept of Mass Connectivity, *Int. J. Thermophys.*, 32 (2011) 942-956.
- [59] U.P.R.M. Preiss, J.M. Slattery, I. Krossing, In Silico Prediction of Molecular Volumes, Heat Capacities, and Temperature-Dependent Densities of Ionic Liquids, *Industrial & Engineering Chemistry Research*, 48 (2009) 2290-2296.
- [60] Y.U. Paulechka, A.G. Kabo, A.V. Blokhin, G.J. Kabo, M.P. Shevelyova, Heat Capacity of Ionic Liquids: Experimental Determination and Correlations with Molar Volume, *Journal of Chemical & Engineering Data*, 55 (2010) 2719-2724.
- [61] L. Glasser, H.D.B. Jenkins, Ambient Isobaric Heat Capacities, $C_{p,m}$, for Ionic Solids and Liquids: An Application of Volume-Based Thermodynamics (VBT), *Inorg. Chem.*, 50 (2011) 8565-8569.
- [62] M. Deetlefs, K.R. Seddon, M. Shara, Neoteric optical media for refractive index determination of gems and minerals, *New J. Chem.*, 30 (2006) 317-326.
- [63] R.L. Gardas, J.A.P. Coutinho, Group contribution methods for the prediction of thermophysical and transport properties of ionic liquids, *AIChE J.*, 55 (2009) 1274-1290.
- [64] F.M. Maia, N. Calvar, E.J. González, A.P. Carneiro, O. Rodriguez, E.A. Macedo, Modeling of Ionic Liquid Systems: Phase Equilibria and Physical Properties, in: J.-i. Kadokawa (Ed.) *Ionic Liquids - New Aspects for the Future*, InTech, 2013.
- [65] G. Astray, A. Cid, O. Moldes, J.A. Ferreira-Lage, J.F. Gálvez, J.C. Mejuto, Prediction of Refractive Index of Polymers Using Artificial Neural Networks, *Journal of Chemical & Engineering Data*, 55 (2010) 5388-5393.
- [66] A.S. Korotkov, V.V. Atuchin, Accurate prediction of refractive index of inorganic oxides by chemical formula, *J. Phys. Chem. Solids*, 71 (2010) 958-964.
- [67] J. Biles, A correlation of the refractive index to the melting point and molecular weight of some organic compounds, *Mikrochimie vereinigt mit Mikrochimica acta*, 39 (1952) 69-72.
- [68] S. Krishnaraj, P. Neelamegam, Prediction of refractive index of organic compounds using structure-property studies, *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, 3 (2012) 597-611.
- [69] H. Redmond, J.E. Thompson, Evaluation of a quantitative structure-property relationship (QSPR) for predicting mid-visible refractive index of secondary organic aerosol (SOA), *PCCP*, 13 (2011) 6872-6882.
- [70] J. Xu, B. Chen, Q. Zhang, B. Guo, Prediction of refractive indices of linear polymers by a four-descriptor QSPR model, *Polymer*, 45 (2004) 8651-8659.
- [71] M. Deetlefs, K.R. Seddon, M. Shara, Predicting physical properties of ionic liquids, *PCCP*, 8 (2006) 642-649.
- [72] A.N. Soriano, B.T. Doma Jr, M.-H. Li, Density and refractive index measurements of 1-ethyl-3-methylimidazolium-based ionic liquids, *Journal of the Taiwan Institute of Chemical Engineers*, 41 (2010) 115-121.

- [73] M.G. Freire, A.R.R. Teles, M.A.A. Rocha, B. Schröder, C.M.S.S. Neves, P.J. Carvalho, D.V. Evtuguin, L.M.N.B.F. Santos, J.A.P. Coutinho, Thermophysical Characterization of Ionic Liquids Able To Dissolve Biomass, *Journal of Chemical & Engineering Data*, 56 (2011) 4813-4822.
- [74] X. Li, Z. Zeng, S. Garg, B. Twamley, J.n.M. Shreeve, Fluorine-Containing Ionic Liquids from N-Alkylpyrrolidine and N-Methylpiperidine and Fluorinated Acetylacetones: Low Melting Points and Low Viscosities, *Eur. J. Inorg. Chem.*, 2008 (2008) 3353-3358.
- [75] S. Zahn, M. Brehm, M. Brüssel, O. Hollóczki, M. Kohagen, S. Lehmann, F. Malberg, A.S. Pensado, M. Schöppke, H. Weber, B. Kirchner, Understanding ionic liquids from theoretical methods, *J. Mol. Liq.*
- [76] N. Khupse, A. Kumar, Dramatic Change in Viscosities of Pure Ionic Liquids upon Addition of Molecular Solvents, *J. Solution Chem.*, 38 (2009) 589-600.
- [77] H. Xue, J.n.M. Shreeve, Ionic Liquids with Fluorine-Containing Cations, *Eur. J. Inorg. Chem.*, 2005 (2005) 2573-2580.
- [78] H. Xue, R. Verma, J.n.M. Shreeve, Review of ionic liquids with fluorine-containing anions, *J. Fluorine Chem.*, 127 (2006) 159-176.
- [79] Z.B. Zhou, H. Matsumoto, K. Tatsumi, Low-melting, low-viscous, hydrophobic ionic liquids: N-alkyl(alkyl ether)-N-methylpyrrolidinium perfluoroethyltrifluoroborate, *Chem. Lett.*, 33 (2004) 1636-1637.
- [80] Y. Chen, Y. Cao, X. Sun, T. Mu, Hydrogen bonding interaction between acetate-based ionic liquid 1-ethyl-3-methylimidazolium acetate and common solvents, *J. Mol. Liq.*, 190 (2014) 151-158.
- [81] A.P. Abbott, Application of hole theory to the viscosity of ionic and molecular liquids, *ChemPhysChem*, 5 (2004) 1242-1246.
- [82] I. Bandrés, R. Alcalde, C. Lafuente, M. Atilhan, S. Aparicio, On the viscosity of pyridinium based ionic liquids: an experimental and computational study, *The Journal of Physical Chemistry B*, 115 (2011) 12499-12513.
- [83] R.L. Gardas, J.A. Coutinho, A group contribution method for viscosity estimation of ionic liquids, *Fluid Phase Equilib.*, 266 (2008) 195-201.
- [84] R.C. Reid, J.M. Prausnitz, B.E. Poling, *The properties of gases and liquids*, (1987).
- [85] S. Aparicio, M. Atilhan, F. Karadas, Thermophysical properties of pure ionic liquids: Review of present situation, *Industrial & Engineering Chemistry Research*, 49 (2010) 9580-9595.
- [86] R.L. Gardas, J.A. Coutinho, Group contribution methods for the prediction of thermophysical and transport properties of ionic liquids, *AIChE J.*, 55 (2009) 1274-1290.
- [87] H. Yamamoto, Structure Properties Relationship of Ionic Liquid, *Journal of Computer Aided Chemistry*, 7 (2006) 18-30.
- [88] H. Matsuda, H. Yamamoto, K. Kurihara, K. Tochigi, Computer-aided reverse design for ionic liquids by QSPR using descriptors of group contribution type for ionic conductivities and viscosities, *Fluid Phase Equilib.*, 261 (2007) 434-443.

- [89] K. Tochigi, H. Yamamoto, Estimation of Ionic Conductivity and Viscosity of Ionic Liquids Using a QSPR Model†, *The Journal of Physical Chemistry C*, 111 (2007) 15989-15994.
- [90] R. Bini, M. Malvaldi, W.R. Pitner, C. Chiappe, QSPR correlation for conductivities and viscosities of low-temperature melting ionic liquids, *J. Phys. Org. Chem.*, 21 (2008) 622-629.
- [91] F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, D. Richon, Development of a group contribution method for determination of viscosity of ionic liquids at atmospheric pressure, *Chem. Eng. Sci.*, 80 (2012) 326-333.
- [92] J. Valderrama, J. Muñoz, R. Rojas, Viscosity of ionic liquids using the concept of mass connectivity and artificial neural networks, *Korean J. Chem. Eng.*, 28 (2011) 1451-1457.
- [93] I. Billard, G. Marcou, A. Ouadi, A. Varnek, In Silico Design of New Ionic Liquids Based on Quantitative Structure–Property Relationship Models of Ionic Liquid Viscosity, *The Journal of Physical Chemistry B*, 115 (2010) 93-98.
- [94] K. Tumba, T. Letcher, P. Naidoo, D. Ramjugernath, Activity coefficients at infinite dilution of organic solutes in the ionic liquid trihexyltetradecylphosphonium hexafluorophosphate using gas–liquid chromatography at T = (313.15, 333.15, 353.15, and 363.15) K, *The Journal of Chemical Thermodynamics*, 49 (2012) 46-53.
- [95] G.M. Foco, S.B. Bottini, N. Quezada, J.C. de la Fuente, C.J. Peters, Activity Coefficients at Infinite Dilution in 1-Alkyl-3-methylimidazolium Tetrafluoroborate Ionic Liquids, *Journal of Chemical & Engineering Data*, 51 (2006) 1088-1091.
- [96] K. Heydar, M. Nazifi, A. Sharifi, M. Mirzaei, H. Gharavi, S. Ahmadi, Determination of Activity Coefficients at Infinite Dilution of Solutes in New Dicationic Ionic Liquids Based on Morpholine Using Gas–Liquid Chromatography, *Chromatographia*, 76 (2013) 165-175.
- [97] P. Reddy, M. Aslam Siddiqi, B. Atakan, M. Diedenhofen, D. Ramjugernath, Activity coefficients at infinite dilution of organic solutes in the ionic liquid PEG-5 cocomonium methylsulfate at T=(313.15, 323.15, 333.15, and 343.15) K: Experimental results and COSMO-RS predictions, *The Journal of Chemical Thermodynamics*, 58 (2013) 322-329.
- [98] R. Kato, J. Gmehling, Activity coefficients at infinite dilution of various solutes in the ionic liquids [MMIM]⁺[CH₃SO₄]⁻, [MMIM]⁺[CH₃OC₂H₄SO₄]⁻, [MMIM]⁺[(CH₃)₂PO₄]⁻, [C₅H₅NC₂H₅]⁺[(CF₃SO₂)₂N]⁻ and [C₅H₅NH]⁺[C₂H₅OC₂H₄OSO₃]⁻, *Fluid Phase Equilib.*, 226 (2004) 37-44.
- [99] K. Tumba, T.M. Letcher, P. Naidoo, D. Ramjugernath, Activity coefficients at infinite dilution of organic solutes in the ionic liquid trihexyltetradecylphosphonium bis (trifluoromethylsulfonyl) imide using gas–liquid chromatography at T = (313.15, 333.15, 353.15 and 373.15) K, *The Journal of Chemical Thermodynamics*, 65 (2013) 159-167.
- [100] A. Marciniak, M. Wlazło, Activity coefficients at infinite dilution and physicochemical properties for organic solutes and water in the ionic liquid 1-(2-methoxyethyl)-1-methylpyrrolidinium bis(trifluoromethylsulfonyl)-amide, *The Journal of Chemical Thermodynamics*, 54 (2012) 90-96.
- [101] A. Marciniak, M. Wlazło, Activity coefficients at infinite dilution and physicochemical properties for organic solutes and water in the ionic liquid 1-(2-methoxyethyl)-1-

methylpiperidinium trifluorotris(perfluoroethyl)phosphate, *The Journal of Chemical Thermodynamics*, 57 (2013) 197-202.

[102] A. Marciniak, M. Wlazło, Activity coefficients at infinite dilution and physicochemical properties for organic solutes and water in the ionic liquid 1-(2-methoxyethyl)-1-methylpyrrolidinium trifluorotris(perfluoroethyl)phosphate, *The Journal of Chemical Thermodynamics*, 60 (2013) 57-62.

[103] P. Reddy, K.J. Chiyen, N. Deenadayalu, D. Ramjugernath, Determination of activity coefficients at infinite dilution of water and organic solutes (polar and non-polar) in the Ammoeng 100 ionic liquid at T = (308.15, 313.5, 323.15, and 333.15) K, *The Journal of Chemical Thermodynamics*, 43 (2011) 1178-1184.

[104] F. Mutelet, E.-S.R.E. Hassan, T.W. Stephens, W.E. Acree, G.A. Baker, Activity Coefficients at Infinite Dilution for Organic Solutes Dissolved in Three 1-Alkyl-1-methylpyrrolidinium Bis(trifluoromethylsulfonyl)imide Ionic Liquids Bearing Short Linear Alkyl Side Chains of Three to Five Carbons, *Journal of Chemical & Engineering Data*, 58 (2013) 2210-2218.

[105] E.F. Órfão, V. Dohnal, A. Blahut, Infinite dilution activity coefficients of volatile organic compounds in two ionic liquids composed of the tris(pentafluoroethyl)trifluorophosphate ([FAP]) anion and a functionalized cation, *The Journal of Chemical Thermodynamics*, 65 (2013) 53-64.

[106] K. Padászyński, U. Domańska, Experimental and theoretical study on infinite dilution activity coefficients of various solutes in piperidinium ionic liquids, *The Journal of Chemical Thermodynamics*, 60 (2013) 169-178.

[107] Ł. Marcinkowski, A. Kloskowski, J. Namieśnik, Measurement of activity coefficients at infinite dilution of organic solutes in the ionic liquid 1-hexyl-1,4-diaza[2.2.2]bicyclooctanium bis(trifluoromethylsulfonyl)imide using gas-liquid chromatography, *The Journal of Chemical Thermodynamics*, 71 (2014) 84-90.

[108] I. Bahadur, B.B. Govender, K. Osman, M.D. Williams-Wynn, W.M. Nelson, P. Naidoo, D. Ramjugernath, Measurement of activity coefficients at infinite dilution of organic solutes in the ionic liquid 1-ethyl-3-methylimidazolium 2-(2-methoxyethoxy) ethylsulfate at T = (308.15, 313.15, 323.15 and 333.15) K using gas + liquid chromatography, *The Journal of Chemical Thermodynamics*, 70 (2014) 245-252.

[109] U. Domańska, M. Królikowski, Measurements of activity coefficients at infinite dilution for organic solutes and water in the ionic liquid 1-ethyl-3-methylimidazolium methanesulfonate, *The Journal of Chemical Thermodynamics*, 54 (2012) 20-27.

[110] U. Domańska, M. Królikowski, W.E. Acree Jr, Thermodynamics and activity coefficients at infinite dilution measurements for organic solutes and water in the ionic liquid 1-butyl-1-methylpyrrolidinium tetracyanoborate, *The Journal of Chemical Thermodynamics*, 43 (2011) 1810-1817.

[111] U. Domańska, M. Królikowski, W.E. Acree Jr, G.A. Baker, Physicochemical properties and activity coefficients at infinite dilution for organic solutes and water in a novel bicyclic guanidinium superbase-derived protic ionic liquid, *The Journal of Chemical Thermodynamics*, 58 (2013) 62-69.

- [112] U. Domańska, E.V. Lukoshko, Measurements of activity coefficients at infinite dilution for organic solutes and water in the ionic liquid 1-butyl-1-methylpyrrolidinium tricyanomethanide, *The Journal of Chemical Thermodynamics*, 66 (2013) 144-150.
- [113] U. Domańska, E.V. Lukoshko, Thermodynamics and activity coefficients at infinite dilution for organic solutes and water in the ionic liquid 1-butyl-1-methylmorpholinium tricyanomethanide, *The Journal of Chemical Thermodynamics*, 68 (2014) 53-59.
- [114] U. Domańska, E.V. Lukoshko, M. Królikowski, Measurements of activity coefficients at infinite dilution for organic solutes and water in the ionic liquid 1-butyl-1-methylpyrrolidinium tris(pentafluoroethyl)trifluorophosphate ([BMPYR][FAP]), *Chem. Eng. J.*, 183 (2012) 261-270.
- [115] U. Domańska, E.V. Lukoshko, M. Wlazło, Measurements of activity coefficients at infinite dilution for organic solutes and water in the ionic liquid 1-hexyl-3-methylimidazolium tetracyanoborate, *The Journal of Chemical Thermodynamics*, 47 (2012) 389-396.
- [116] Z. Jiao, Y. Sun, Q. Yang, X. Wang, Activity coefficients at infinite dilution of organic solutes in the ionic liquid ethyl(2-hydroxyethyl)dimethyl-ammonium diethylphosphate using gas-liquid chromatography, *Fluid Phase Equilib.*, 325 (2012) 15-19.
- [117] Z. Ma, X. Dong, Y. Hu, B. Zhang, C. Xu, Y. Liu, Effect of Ionic Liquids on Organic Reactions Based on Activity Coefficients at Infinite Dilution, *Chin. J. Chem. Eng.*, 21 (2013) 1370-1375.
- [118] A. Marciniak, M. Wlazło, Activity coefficients at infinite dilution and physicochemical properties for organic solutes and water in the ionic liquid 1-(2-methoxyethyl)-1-methylpiperidinium bis(trifluoromethylsulfonyl)-amide, *The Journal of Chemical Thermodynamics*, 49 (2012) 137-145.
- [119] A. Marciniak, M. Wlazło, Activity coefficients at infinite dilution and physicochemical properties for organic solutes and water in the ionic liquid 4-(2-methoxyethyl)-4-methylmorpholinium bis(trifluoromethylsulfonyl)-amide, *The Journal of Chemical Thermodynamics*, 47 (2012) 382-388.
- [120] A. Marciniak, M. Wlazło, Activity coefficients at infinite dilution and physicochemical properties for organic solutes and water in the ionic liquid 1-(2-hydroxyethyl)-3-methylimidazolium trifluorotris(perfluoroethyl)phosphate, *The Journal of Chemical Thermodynamics*, 64 (2013) 114-119.
- [121] M. Wlazło, A. Marciniak, Activity coefficients at infinite dilution and physicochemical properties for organic solutes and water in the ionic liquid 4-(2-methoxyethyl)-4-methylmorpholinium trifluorotris(perfluoroethyl)phosphate, *The Journal of Chemical Thermodynamics*, 54 (2012) 366-372.
- [122] S. Çehreli, J. Gmehling, Phase equilibria for benzene-cyclohexene and activity coefficients at infinite dilution for the ternary systems with ionic liquids, *Fluid Phase Equilib.*, 295 (2010) 125-129.
- [123] N. Deenadayalu, S.H. Thango, T.M. Letcher, D. Ramjugernath, Measurement of activity coefficients at infinite dilution using polar and non-polar solutes in the ionic liquid 1-

methyl-3-octyl-imidazolium diethyleneglycolmonomethylethersulfate at T = (288.15, 298.15, and 313.15) K, *The Journal of Chemical Thermodynamics*, 38 (2006) 542-546.

[124] U. Domańska, A. Marciniak, Activity coefficients at infinite dilution measurements for organic solutes and water in the 1-hexyloxymethyl-3-methyl-imidazolium and 1,3-dihexyloxymethyl-imidazolium bis(trifluoromethylsulfonyl)-imide ionic liquids—The cation influence, *Fluid Phase Equilib.*, 286 (2009) 154-161.

[125] U. Domańska, K. Padiuszyński, Gas–liquid chromatography measurements of activity coefficients at infinite dilution of various organic solutes and water in tri-iso-butylmethylphosphonium tosylate ionic liquid, *The Journal of Chemical Thermodynamics*, 42 (2010) 707-711.

[126] U. Domańska, K. Padiuszyński, Measurements of activity coefficients at infinite dilution of organic solutes and water in 1-propyl-1-methylpiperidinium bis{(trifluoromethyl)sulfonyl}imide ionic liquid using g.l.c, *The Journal of Chemical Thermodynamics*, 42 (2010) 1361-1366.

[127] T.M. Letcher, A. Marciniak, M. Marciniak, U. Domańska, Activity coefficients at infinite dilution measurements for organic solutes in the ionic liquid 1-hexyl-3-methyl-imidazolium bis(trifluoromethylsulfonyl)-imide using g.l.c. at T=(298.15, 313.15, and 333.15) K, *The Journal of Chemical Thermodynamics*, 37 (2005) 1327-1331.

[128] T.M. Letcher, P. Reddy, Determination of activity coefficients at infinite dilution of organic solutes in the ionic liquid, trihexyl(tetradecyl)-phosphonium tris(pentafluoroethyl)trifluorophosphate, by gas–liquid chromatography, *Fluid Phase Equilib.*, 235 (2005) 11-17.

[129] T.M. Letcher, P. Reddy, Determination of activity coefficients at infinite dilution of organic solutes in the ionic liquid, tributylmethylphosphonium methylsulphate by gas–liquid chromatography, *Fluid Phase Equilib.*, 260 (2007) 23-28.

[130] E. Olivier, T.M. Letcher, P. Naidoo, D. Ramjugernath, Activity coefficients at infinite dilution of organic solutes in the ionic liquid 1-octyl-3-methylimidazolium hexafluorophosphate using gas–liquid chromatography at T=(313.15, 323.15, and 333.15) K, *The Journal of Chemical Thermodynamics*, 42 (2010) 646-650.

[131] P. Reddy, N.V. Gwala, N. Deenadayalu, D. Ramjugernath, Activity coefficients at infinite dilution of organic solutes in the ionic liquid, methyl(trioctyl)ammonium thiosalicylate, [N1888][TS] by gas–liquid chromatography at T = (303.15, 313.15, and 323.15) K, *The Journal of Chemical Thermodynamics*, 43 (2011) 754-758.

[132] Y. Shimoyama, K. Ikeda, Y. Iwai, Effect of anion species on infinite dilution activity coefficients of epoxides in imidazolium-based ionic liquids, *Fluid Phase Equilib.*, 294 (2010) 241-245.

[133] K. Tabar Heydar, H.G. Gharavi, M. Nazifi, M. Mirzaei, A. Sharifi, Using binary mixtures of dicationic ionic liquids for determination of activity coefficients at infinite dilution by gas–liquid chromatography, *Fluid Phase Equilib.*, 353 (2013) 93-100.

[134] J. Gmehling, J. Li, M. Schiller, A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties, *Industrial & Engineering Chemistry Research*, 32 (1993) 178-193.

- [135] A. Fredenslund, R.L. Jones, J.M. Prausnitz, Group-contribution estimation of activity coefficients in nonideal liquid mixtures, *AIChE J.*, 21 (1975) 1086-1099.
- [136] S. Nebig, J. Gmehling, Measurements of different thermodynamic properties of systems containing ionic liquids and correlation of these properties using modified UNIFAC (Dortmund), *Fluid Phase Equilib.*, 294 (2010) 206-212.
- [137] T.M. Letcher, B. Soko, D. Ramjugernath, N. Deenadayalu, A. Nevines, P.K. Naicker, Activity Coefficients at Infinite Dilution of Organic Solutes in 1-Hexyl-3-methylimidazolium Hexafluorophosphate from Gas-Liquid Chromatography, *Journal of Chemical & Engineering Data*, 48 (2003) 708-711.
- [138] T.M. Letcher, A. Marciniak, M. Marciniak, U. Domańska, Determination of Activity Coefficients at Infinite Dilution of Solutes in the Ionic Liquid 1-Butyl-3-methylimidazolium Octyl Sulfate Using Gas-Liquid Chromatography at a Temperature of 298.15 K, 313.15 K, or 328.15 K, *Journal of Chemical & Engineering Data*, 50 (2005) 1294-1298.
- [139] T. Magnussen, P. Rasmussen, A. Fredenslund, UNIFAC parameter table for prediction of liquid-liquid equilibria, *Industrial & Engineering Chemistry Process Design and Development*, 20 (1981) 331-339.
- [140] R. Putnam, R. Taylor, A. Klamt, F. Eckert, M. Schiller, Prediction of Infinite Dilution Activity Coefficients Using COSMO-RS, *Industrial & Engineering Chemistry Research*, 42 (2003) 3635-3641.
- [141] A. Klamt, Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena, *The Journal of Physical Chemistry*, 99 (1995) 2224-2235.
- [142] E. Mullins, R. Oldland, Y.A. Liu, S. Wang, S.I. Sandler, C.-C. Chen, M. Zwolak, K.C. Seavey, Sigma-Profile Database for Using COSMO-Based Thermodynamic Methods, *Industrial & Engineering Chemistry Research*, 45 (2006) 4389-4415.
- [143] F. Eckert, A. Klamt, Prediction of halocarbon thermodynamics with COSMO-RS, *Fluid Phase Equilib.*, 210 (2003) 117-141.
- [144] M.H. Ghatee, F. Moosavi, A.R. Zolghadr, R. Jahromi, Critical-Point Temperature of Ionic Liquids from Surface Tension at Liquid-Vapor Equilibrium and the Correlation with the Interaction Energy, *Industrial & Engineering Chemistry Research*, 49 (2010) 12696-12701.
- [145] J.L. Shereshefsky, Surface Tension of Saturated Vapors and the Equation of Eötvös, *The Journal of Physical Chemistry*, 35 (1930) 1712-1720.
- [146] E.A. Guggenheim, The Principle of Corresponding States, *The Journal of Chemical Physics*, 13 (1945) 253-261.
- [147] L.P.N. Rebelo, J.N. Canongia Lopes, J.M.S.S. Esperança, E. Filipe, On the Critical Temperature, Normal Boiling Point, and Vapor Pressure of Ionic Liquids, *The Journal of Physical Chemistry B*, 109 (2005) 6040-6043.
- [148] U. Domańska, K. Skiba, M. Zawadzki, K. Padaszyński, M. Królikowski, Synthesis, physical, and thermodynamic properties of 1-alkyl-cyanopyridinium

bis{(trifluoromethyl)sulfonyl}imide ionic liquids, *The Journal of Chemical Thermodynamics*, 56 (2013) 153-161.

[149] U. Domańska, Z. Żołek-Tryznowska, Effect of Temperature and Composition on the Surface Tension and Thermodynamic Properties of Binary Mixtures of Boltorn U3000 with Alcohols and Ether, *J. Solution Chem.*, 39 (2010) 864-876.

[150] A. Shariati, S.-S. Ashrafmansouri, M. Osbuei, B. Hooshdaran, Critical properties and acentric factors of ionic liquids, *Korean J. Chem. Eng.*, 30 (2013) 187-193.

[151] U. Domańska, M. Zawadzki, A. Lewandowska, Effect of temperature and composition on the density, viscosity, surface tension, and thermodynamic properties of binary mixtures of N-octylisoquinolinium bis{(trifluoromethyl)sulfonyl}imide with alcohols, *The Journal of Chemical Thermodynamics*, 48 (2012) 101-111.

[152] M. Zawadzki, L. Niedzicki, W. Wieczorek, U. Domańska, Estimation of extraction properties of new imidazolidine anion based ionic liquids on the basis of activity coefficient at infinite dilution measurements, *Sep. Purif. Technol.*, 118 (2013) 242-254.

[153] J.O. Valderrama, B.F. Abu-Sharkh, Generalized Rackett-type correlations to predict the density of saturated liquids and petroleum fractions, *Fluid Phase Equilib.*, 51 (1989) 87-100.

[154] J.O. Valderrama, L.A. Forero, R.E. Rojas, Critical Properties and Normal Boiling Temperature of Ionic Liquids. Update and a New Consistency Test, *Industrial & Engineering Chemistry Research*, 51 (2012) 7838-7844.

[155] K.-J. Wu, Q.-L. Chen, C.-H. He, Speed of sound of ionic liquids: Database, estimation, and its application for thermal conductivity prediction, *AIChE J.*, 60 (2014) 1120-1131.

[156] O.M. Basha, Y.J. Heintz, M.J. Keller, D.R. Luebke, K.P. Resnik, B.I. Morsi, Development of a Conceptual Process for Selective Capture of CO₂ from Fuel Gas Streams Using Two TEGO Ionic Liquids as Physical Solvents, *Industrial & Engineering Chemistry Research*, 53 (2014) 3184-3195.

[157] M. Zoubeik, A. Henni, Experimental and thermodynamic study of CO₂ solubility in promising [TF₂N and DCN] ionic liquids, *Fluid Phase Equilib.*, 376 (2014) 22-30.

[158] A. Tagiuri, K.Z. Sumon, A. Henni, Solubility of carbon dioxide in three [Tf₂N] ionic liquids, *Fluid Phase Equilib.*, 380 (2014) 39-47.

[159] M.A. Ahmadi, R. Haghbakhsh, R. Soleimani, M.B. Bajestani, Estimation of H₂S solubility in ionic liquids using a rigorous method, *The Journal of Supercritical Fluids*, 92 (2014) 60-69.

[160] A. Shafiei, M.A. Ahmadi, S.H. Zaheri, A. Baghban, A. Amirfakhrian, R. Soleimani, Estimating hydrogen sulfide solubility in ionic liquids using a machine learning approach, *The Journal of Supercritical Fluids*.

[161] S. Hong, Y. Park, D. Pore, Experimental determination and prediction of phase behavior for 1-butyl-3-methylimidazolium nonafluorobutyl sulfonate and carbon dioxide, *Korean J. Chem. Eng.*, 31 (2014) 1656-1660.

- [162] A. Eslamimanesh, F. Gharagheizi, A.H. Mohammadi, D. Richon, Artificial Neural Network modeling of solubility of supercritical carbon dioxide in 24 commonly used ionic liquids, *Chem. Eng. Sci.*, 66 (2011) 3039-3044.
- [163] R. Seddon Kenneth, A. Stark, M.-J. Torres, Influence of chloride, water, and organic solvents on the physical properties of ionic liquids, in: *Pure Appl. Chem.*, 2000, pp. 2275.
- [164] S. Zhang, X. Lu, Q. Zhou, X. Li, X. Zhang, S. Li, *Ionic Liquids:: Physicochemical Properties*, Elsevier Science, 2009.
- [165] A. Ouadi, B. Gadenne, P. Hesemann, J.J.E. Moreau, I. Billard, C. Gaillard, S. Mekki, G. Moutiers, Task-Specific Ionic Liquids Bearing 2-Hydroxybenzylamine Units: Synthesis and Americium-Extraction Studies, *Chemistry – A European Journal*, 12 (2006) 3074-3081.
- [166] W.E. Acree, G.A. Baker, A.-L. Revelli, J.-C. Moise, F. Mutelet, Activity Coefficients at Infinite Dilution for Organic Compounds Dissolved in 1-Alkyl-1-methylpyrrolidinium Bis(trifluoromethylsulfonyl)imide Ionic Liquids Having Six-, Eight-, and Ten-Carbon Alkyl Chains, *Journal of Chemical & Engineering Data*, 57 (2012) 3510-3518.
- [167] U. Domańska, M. Królikowska, Measurements of Activity Coefficients at Infinite Dilution in Solvent Mixtures with Thiocyanate-Based Ionic Liquids Using GLC Technique, *The Journal of Physical Chemistry B*, 114 (2010) 8460-8466.
- [168] U. Domańska, M. Królikowski, Thermodynamics and Activity Coefficients at Infinite Dilution Measurements for Organic Solutes and Water in the Ionic Liquid N-Hexyl-3-methylpyridinium Tosylate, *The Journal of Physical Chemistry B*, 115 (2011) 7397-7404.
- [169] U. Domańska, A. Marciniak, Activity Coefficients at Infinite Dilution Measurements for Organic Solutes and Water in the Ionic Liquid 1-Ethyl-3-methylimidazolium Trifluoroacetate, *The Journal of Physical Chemistry B*, 111 (2007) 11984-11988.
- [170] U. Domańska, A. Marciniak, Measurements of activity coefficients at infinite dilution of aromatic and aliphatic hydrocarbons, alcohols, and water in the new ionic liquid [EMIM][SCN] using GLC, *The Journal of Chemical Thermodynamics*, 40 (2008) 860-866.
- [171] U. Domańska, A. Marciniak, Activity coefficients at infinite dilution measurements for organic solutes and water in the ionic liquid triethylsulphonium bis(trifluoromethylsulfonyl)imide, *The Journal of Chemical Thermodynamics*, 41 (2009) 754-758.
- [172] U. Domańska, A. Marciniak, Activity coefficients at infinite dilution measurements for organic solutes and water in the ionic liquid 4-methyl-N-butyl-pyridinium bis(trifluoromethylsulfonyl)-imide, *The Journal of Chemical Thermodynamics*, 41 (2009) 1350-1355.
- [173] U. Domańska, A. Marciniak, Physicochemical Properties and Activity Coefficients at Infinite Dilution for Organic Solutes and Water in the Ionic Liquid 1-Decyl-3-methylimidazolium Tetracyanoborate, *The Journal of Physical Chemistry B*, 114 (2010) 16542-16547.
- [174] U. Domańska, M. Zawadzki, M. Królikowska, M. Marc Tshibangu, D. Ramjugernath, T.M. Letcher, Measurements of activity coefficients at infinite dilution of organic compounds and water in isoquinolinium-based ionic liquid [C8iQuin][NTf2] using GLC, *The Journal of Chemical Thermodynamics*, 43 (2011) 499-504.

- [175] Y.-X. Feng, L.-S. Wang, Y. Li, Activity Coefficients at Infinite Dilution of Organic Solutes in 1-Butyl-3-methylimidazolium Nitrate Using Gas–Liquid Chromatography, *Journal of Chemical & Engineering Data*, 56 (2011) 2730-2736.
- [176] A. Marciniak, M. Wlazło, Activity Coefficients at Infinite Dilution Measurements for Organic Solutes and Water in the Ionic Liquid 1-(3-Hydroxypropyl)pyridinium Trifluorotris(perfluoroethyl)phosphate, *The Journal of Physical Chemistry B*, 114 (2010) 6990-6994.
- [177] K. Padaszyński, U. Domańska, Limiting Activity Coefficients and Gas–Liquid Partition Coefficients of Various Solutes in Piperidinium Ionic Liquids: Measurements and LSER Calculations, *The Journal of Physical Chemistry B*, 115 (2011) 8207-8215.
- [178] H. Jiang, Y. Zhao, J. Wang, F. Zhao, R. Liu, Y. Hu, Density and surface tension of pure ionic liquid 1-butyl-3-methyl-imidazolium l-lactate and its binary mixture with alcohol and water, *The Journal of Chemical Thermodynamics*, 64 (2013) 1-13.
- [179] M.H. Ghatee, M. Bahrami, N. Khanjari, Measurement and study of density, surface tension, and viscosity of quaternary ammonium-based ionic liquids ([N222(n)]Tf2N), *The Journal of Chemical Thermodynamics*, 65 (2013) 42-52.
- [180] Y. Wei, Y. Jin, Z.-J. Wu, Y. Yang, Q.-G. Zhang, Z.-H. Kang, Synthesis and Physicochemical Properties of Amino Acid Ionic Liquid 1-Butyl-3-methylimidazolium Aspartate and Binary Mixture with Methanol, *Journal of Chemical & Engineering Data*, 58 (2013) 349-356.
- [181] H.F.D. Almeida, J.A. Lopes-da-Silva, M.G. Freire, J.A.P. Coutinho, Surface tension and refractive index of pure and water-saturated tetradecyltrihexylphosphonium-based ionic liquids, *The Journal of Chemical Thermodynamics*, 57 (2013) 372-379.
- [182] A. Xu, Y. Zhang, Z. Li, J. Wang, Effect of Substituent Groups in Anions on Some Physicochemical Properties of 1-Butyl-3-methylimidazolium Carboxylate Ionic Liquids, *Journal of Chemical & Engineering Data*, 58 (2013) 2496-2501.
- [183] M. Harini, J. Adhikari, K.Y. Rani, A Review on Property Estimation Methods and Computational Schemes for Rational Solvent Design: A Focus on Pharmaceuticals, *Industrial & Engineering Chemistry Research*, 52 (2013) 6869-6893.
- [184] L. Riedel, Ein neues Verfahren zur Abschätzung unbekannter kritischer Drucke von organischen Verbindungen, *Zeitschrift für Elektrochemie und angewandte physikalische Chemie*, 53 (1949) 222-228.
- [185] A.L. Lydersen, E.E. Station, R.A. Greenkorn, O.A. Hougen, Estimation of Critical Properties of Organic Compounds by the Method of Group Contributions, University of Wisconsin, 1955.
- [186] K.G. Joback, R.C. Reid, Estimation Of Pure-Component Properties From Group-Contributions, *Chem. Eng. Commun.*, 57 (1987) 233-243.
- [187] Y. Nannoolal, J. Rarey, D. Ramjugernath, Estimation of pure component properties: Part 2. Estimation of critical property data by group contribution, *Fluid Phase Equilib.*, 252 (2007) 1-27.

- [188] L. Constantinou, R. Gani, New group contribution method for estimating properties of pure compounds, *AIChE J.*, 40 (1994) 1697-1710.
- [189] J. Marrero, R. Gani, Group-contribution based estimation of pure component properties, *Fluid Phase Equilib.*, 183–184 (2001) 183-208.
- [190] Q. Wang, P. Ma, S. Neng, Position Group Contribution Method for Estimation of Melting Point of Organic Compounds, *Chin. J. Chem. Eng.*, 17 (2009) 468-472.
- [191] M. Sattari, F. Gharagheizi, Prediction of molecular diffusivity of pure components into air: a QSPR approach, *Chemosphere*, 72 (2008) 1298-1302.
- [192] R. Todeschini, V. Consonni, R. Mannhold, *Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing / Volume II: Appendices, References*, Wiley-VCH, 2009.
- [193] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, Chichester, 2000.
- [194] N. Mills, ChemDraw Ultra 10.0 CambridgeSoft, 100 CambridgePark Drive, Cambridge, MA 02140. www.cambridgesoft.com. Commercial Price: \$1910 for download, \$2150 for CD-ROM; Academic Price: \$710 for download, \$800 for CD-ROM, *J. Am. Chem. Soc.*, 128 (2006) 13649-13650.
- [195] JChem Base in, ChemAxon.
- [196] S.L. Mayo, B.D. Olafson, W.A. Goddard, DREIDING: a generic force field for molecular simulations, *The Journal of Physical Chemistry*, 94 (1990) 8897-8909.
- [197] Talete srl, DRAGON for Windows (Software for Molecular Descriptor Calculations) Version 5.5, 2007., in.
- [198] M. Sattari, F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, Estimation of the Heat Capacity of Ionic Liquids: A Quantitative Structure–Property Relationship Approach, *Industrial & Engineering Chemistry Research*, 52 (2013) 13217–13221.
- [199] F. Tan, X. Fu, Y. Zhang, A.G. Bourgeois, A genetic algorithm-based method for feature subset selection, *Soft Comput.*, 12 (2007) 111-120.
- [200] M.A. Efroymson, Multiple regression analysis, in: A. Ralston, H.S. Wilf (Eds.) *Mathematical Methods for Digital Computers*, John Wiley & Sons Inc, New York, 1960.
- [201] J.H. Holland, *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*, 1st MIT Press ed. ed., University of Michigan Press, 1975.
- [202] H. Vafaie, K. De Jong, Genetic algorithms as a tool for feature selection in machine learning, in: *Tools with Artificial Intelligence*, 1992. TAI '92, Proceedings., Fourth International Conference on, 1992, pp. 200-203.
- [203] D. Whitley, A genetic algorithm tutorial, *Stat Comput*, 4 (1994) 65-85.

- [204] F. Gharagheizi, M. Sattari, Estimation of molecular diffusivity of pure chemicals in water: a quantitative structure-property relationship study, SAR and QSAR in environmental research, 20 (2009) 267-285.
- [205] F. Gharagheizi, M. Sattari, Prediction of the θ (UCST) of polymer solutions: A quantitative structure-property relationship study, Ind. Eng. Chem. Res., 48 (2009) 9054-9060.
- [206] F. Gharagheizi, M. Sattari, Prediction of Triple-Point Temperature of Pure Components Using their Chemical Structures, Industrial & Engineering Chemistry Research, 49 (2009) 929-932.
- [207] F. Gharagheizi, M. Sattari, Prediction of triple-point temperature of pure components using their chemical structures, Ind. Eng. Chem. Res., 49 (2010) 929-932.
- [208] J.H. Friedman, Multivariate Adaptive Regression Splines, The Annals of Statistics, 19 (1991) 1-67.
- [209] D. Rogers, A.J. Hopfinger, Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships, J. Chem. Inf. Comput. Sci., 34 (1994) 854-866.
- [210] L. Wang, Support Vector Machines: Theory and Applications, Springer, 2005.
- [211] H. Liu, X. Yao, R. Zhang, M. Liu, Z. Hu, B. Fan, Accurate Quantitative Structure-Property Relationship Model To Predict the Solubility of C60 in Various Solvents Based on a Novel Approach Using a Least-Squares Support Vector Machine, The Journal of Physical Chemistry B, 109 (2005) 20565-20571.
- [212] J.A.K. Suykens, J. Vandewalle, Least Squares Support Vector Machine Classifiers, Neural Processing Letters, 9 (1999) 293-300.
- [213] C.J. Lee;, G. Lee;, W. So;, E.S. Yoon, A new estimation algorithm of physical properties based on a group contribution and support vector machine, Korean J. Chem. Eng, 25 (2008) 568-574.
- [214] X. Yao, H. Liu, R. Zhang, M. Liu, Z. Hu, A. Panaye, J.P. Doucet, B. Fan, QSAR and Classification Study of 1,4-Dihydropyridine Calcium Channel Antagonists Based on Least Squares Support Vector Machines, Molecular Pharmaceutics, 2 (2005) 348-356.
- [215] E. Pourbasheer, R. Aalizadeh, M. Ganjali, P. Norouzi, QSAR study of $\alpha 1\beta 4$ integrin inhibitors by GA-MLR and GA-SVM methods, Struct. Chem., 25 (2014) 355-370.
- [216] R. Poli, W.B. Langdon, N.F. McPhee, J.R. Koza, A Field Guide to Genetic Programming, Lulu.com, 2008.
- [217] N.L. Cramer, A representation for the adaptive generation of simple sequential programs, in: Int. Conf. Genetic Algorithms and their Applications, 1985, pp. 183-187.
- [218] J.R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection, Massachusetts Institute of Technology, 1992.
- [219] E.S. Mostafavi, S.S. Ramiyani, R. Sarvar, H.I. Moud, S.M. Mousavi, A hybrid computational approach to estimate solar global radiation: An empirical evidence from Iran, Energy, 49 (2013) 204-210.

- [220] E.S. Mostafavi, S.M. Mousavi, F. Hosseinpour, Gene Expression Programming as a Basis for New Generation of Electricity Demand Prediction Models, *Computers & Industrial Engineering*, (2014).
- [221] C. Ferreira, Gene expression programming: a new adaptive algorithm for solving problems, *Complex Systems*, 13 (2001) 87-129.
- [222] A.H. Alavi, A.H. Gandomi, H.C. Nejad, A. Mollahasani, A. Rashed, Design equations for prediction of pressuremeter soil deformation moduli utilizing expression programming systems, *Neural Computing and Applications*, 23 (2013) 1771-1786.
- [223] F. Gharagheizi, A. Eslamimanesh, M. Sattari, A.H. Mohammadi, D. Richon, Corresponding States Method for Evaluation of the Solubility Parameters of Chemical Compounds, *Industrial & Engineering Chemistry Research*, 51 (2012) 3826-3831.
- [224] F. Gharagheizi, A. Eslamimanesh, M. Sattari, A.H. Mohammadi, D. Richon, Corresponding States Method for Determination of the Viscosity of Gases at Atmospheric Pressure, *Industrial & Engineering Chemistry Research*, 51 (2012) 3179-3185.
- [225] F. Gharagheizi, A. Eslamimanesh, M. Sattari, B. Tirandazi, A.H. Mohammadi, D. Richon, Evaluation of Thermal Conductivity of Gases at Atmospheric Pressure through a Corresponding States Method, *Industrial & Engineering Chemistry Research*, 51 (2012) 3844-3849.
- [226] C. Ferreira, "What is GEP?" From GeneXproTools Tutorials – A Gepsoft Web Resource, in, 2010.
- [227] F. Gharagheizi, M. Sattari, B. Tirandazi, Prediction of crystal lattice energy using enthalpy of sublimation: A group contribution-based model, *Industrial and Engineering Chemistry Research*, 50 (2011) 2482-2486.
- [228] M.-C. Lee, C. To, Comparison of Support Vector Machine and Back Propagation Neural Network in Evaluating the Enterprise Financial Distress, *International Journal of Artificial Intelligence & Applications*, 1 (2010) 31-43.
- [229] D.T. Manallack, D.J. Livingstone, Neural networks in drug discovery: have they lived up to their promise?, *European Journal of Medicinal Chemistry*, 34 (1999) 195-208.
- [230] F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, Group contribution model for estimation of surface tension of ionic liquids, *Chem. Eng. Sci.*, 78 (2012) 204-208.
- [231] J.H. Schuur, P. Selzer, J. Gasteiger, The Coding of the Three-Dimensional Structure of Molecules by Molecular Transforms and Its Application to Structure-Spectra Correlations and Studies of Biological Activity, *Journal of Chemical Information and Computer Sciences*, 36 (1996) 334-344.
- [232] L.B. Kier, L.H. Hall, *Molecular connectivity in structure-activity analysis*, Research Studies Press, 1986.
- [233] P. Broto, G. Moreau, C. Vandycke, Molecular structures: Perception, autocorrelation descriptor and sar studies. Use of the autocorrelation descriptor in the qsar study of two non-narcotic analgesic series, *European Journal of Medicinal Chemistry*, 19 (1984) 79-84.

- [234] P.A.P. Moran, Notes on Continuous Stochastic Phenomena, *Biometrika*, 37 (1950) 17-23.
- [235] J. Galvez, R. Garcia, M.T. Salabert, R. Soler, Charge Indexes. New Topological Descriptors, *Journal of Chemical Information and Computer Sciences*, 34 (1994) 520-525.
- [236] J.M. Crosthwaite, M.J. Muldoon, J.K. Dixon, J.L. Anderson, J.F. Brennecke, Phase transition and decomposition temperatures, heat capacities and viscosities of pyridinium ionic liquids, *The Journal of Chemical Thermodynamics*, 37 (2005) 559-568.
- [237] G. García-Miaja, J. Troncoso, L. Romání, Excess properties for binary systems ionic liquid + ethanol: Experimental results and theoretical description using the ERAS model, *Fluid Phase Equilib.*, 274 (2008) 59-67.
- [238] A.K. Ziyada, M.A. Bustam, C.D. Wilfred, T. Murugesan, Densities, Viscosities, and Refractive Indices of 1-Hexyl-3-propanenitrile Imidazolium Ionic Liquids Incorporated with Sulfonate-Based Anions, *Journal of Chemical & Engineering Data*, 56 (2011) 2343-2348.
- [239] K.-S. Kim, B.-K. Shin, H. Lee, F. Ziegler, Refractive index and heat capacity of 1-butyl-3-methylimidazolium bromide and 1-butyl-3-methylimidazolium tetrafluoroborate, and vapor pressure of binary systems for 1-butyl-3-methylimidazolium bromide + trifluoroethanol and 1-butyl-3-methylimidazolium tetrafluoroborate + trifluoroethanol, *Fluid Phase Equilib.*, 218 (2004) 215-220.
- [240] R. Todeschini, V. Consonni, R. Mannhold, H. Kubinyi, G. Folkers, *Molecular Descriptors for Chemoinformatics*, Wiley, 2009.
- [241] S. Zhang, X. Qi, X. Ma, L. Lu, Q. Zhang, Y. Deng, Investigation of cation–anion interaction in 1-(2-hydroxyethyl)-3-methylimidazolium-based ion pairs by density functional theory calculations and experiments, *J. Phys. Org. Chem.*, 25 (2012) 248-257.
- [242] S. Schneider, G. Drake, L. Hall, T. Hawkins, M. Rosander, Alkene- and Alkyne-substituted Methylimidazolium Bromides: Structural Effects and Physical Properties, *Z. Anorg. Allg. Chem.*, 633 (2007) 1701-1707.
- [243] M. Levitt, M.F. Perutz, Aromatic rings act as hydrogen bond acceptors, *J. Mol. Biol.*, 201 (1988) 751-754.
- [244] S. Suzuki, P.G. Green, R.E. Bumgarner, S. Dasgupta, W.A. Goddard, G.A. Blake, Benzene Forms Hydrogen Bonds with Water, *Science*, 257 (1992) 942-945.
- [245] K. Compaan, R. Vergenz, P. Von Rague Schleyer, I. Arreguin, Carbon-donated hydrogen bonding: Electrostatics, frequency shifts, directionality, and bifurcation, *Int. J. Quantum Chem.*, 108 (2008) 2914-2923.
- [246] J. Komasa, K. Szalewicz, J. Leszczyński, Does the methyl group form a hydrogen bond? Ab initio post-Hartree–Fock study on ethane–hydrogen cyanide complex, *Chem. Phys. Lett.*, 285 (1998) 449-454.
- [247] S.J. Knak Jensen, T.-H. Tang, I.G. Csizmadia, Hydrogen-Bonding Ability of a Methyl Group, *The Journal of Physical Chemistry A*, 107 (2003) 8975-8979.

[248] P. Suarez, A.Z., S. Einloft, J. Dullius, E.L., R. de Souza, F., J. Dupont, Synthesis and physical-chemical properties of ionic liquids based on 1-butyl-3-methylimidazolium cation, *J. Chim. Phys.*, 95 (1998) 1626-1639.

[249] J. Pernak, J. Feder-Kubis, Synthesis and Properties of Chiral Ammonium-Based Ionic Liquids, *Chemistry – A European Journal*, 11 (2005) 4441-4449.

[250] J.O. Valderrama, W.W. Sanga, J.A. Lazzús, Critical properties, normal boiling temperature, and acentric factor of another 200 ionic liquids, *Ind. Eng. Chem. Res.*, 47 (2008) 1318-1330.

