



Prediction of the Physical Properties of Pure Chemical Compounds
through Different Computational Methods

by

Farhad Gharagheizi

MSc. Chemical Engineering

University of KwaZulu-Natal

Supervisor: Prof. Deresh Ramjugernath

Co-supervisor: Prof. Amir H Mohammadi

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

*I dedicate this thesis to
my wife, Poorandokht, and my mom and dad
for their constant support and unconditional love.*

I love you all cordially.

اسب خرد از گنبد گردون بجماند

بیدارش نماید که بس نخته نماید

لحمان خرک خویش به منزل برساند

در جهل مرکب ابد الدهر بماند

آنگس که بداند و بداند که بداند

آنگس که بداند و نداند که بداند

آنگس که نداند و بداند که نداند

آنگس که نداند و نداند که نداند

Abstract

Liquid thermal conductivities, viscosities, thermal decomposition temperatures, electrical conductivities, normal boiling point temperatures, sublimation and vaporization enthalpies, saturated liquid speeds of sound, standard molar chemical exergies, refractive indices, and freezing point temperatures of pure organic compounds and ionic liquids are important thermophysical properties needed for the design and optimization of products and chemical processes. Since sufficiently purification of pure compounds as well as experimentally measuring their thermophysical properties are costly and time consuming, predictive models are of great importance in engineering.

The liquid thermal conductivity of pure organic compounds was the first investigated property, in this study, for which, a general model, a quantitative structure property relationship, and a group contribution method were developed.

The novel gene expression programming mathematical strategy [1, 2], firstly introduced by our group, for development of non-linear models for thermophysical properties, was successfully implemented to develop an explicit model for determination of the thermal conductivity of approximately 1600 liquids at different temperatures but atmospheric pressure. The statistical parameters of the obtained correlation show about 9% absolute average relative deviation of the results from the corresponding DIPPR 801 data [3]. It should be mentioned that the gene expression programming technique is a complicated mathematical algorithm and needs a significant computer power and this is the largest databases of thermophysical properties that have been successfully managed by this strategy.

The quantitative structure property relationship was developed using the sequential search algorithm and the same database used in previous step. The model shows the average absolute relative deviation (AARD %), standard deviation error, and root mean square error of 7.4%, 0.01, and 0.01 over the training, validation and test sets, respectively.

The database used in previous sections was used to develop a group contribution model for liquid thermal conductivity. The statistical analysis of the performance of the obtained model shows approximately a 7.1% absolute average relative deviation of the results from the corresponding DIPPR 801 [4] data.

In the next stage, an extensive database of viscosities of 443 ionic liquids was initially compiled from literature (more than 200 articles). Then, it was employed to develop a group contribution model. Using this model, a training set composed of 1336 experimental data was correlated with a low AARD% of about 6.3. A test set consists of 336 data point was used to validate this model. It shows an AARD% of 6.8 for the test set.

In the next part of this study, an extensive database of thermal decomposition temperature of 586 ionic liquids was compiled from literature. Then, it was used to develop a quantitative structure property relationship. The proposed quantitative structure property relationship produces an acceptable average absolute relative deviation (AARD) of less than 5.2 % taking into consideration all 586 experimental data values.

The updated database of thermal decomposition temperature including 613 ionic liquids was subsequently used to develop a group contribution model. Using this model, a training set comprised of 489 data points was correlated with a low AARD of 4.5 %. A test set consisting of

124 data points was employed to test its capability. The model shows an AARD of 4.3 % for the test set.

Electrical conductivity of ionic liquids was the next property investigated in this study. Initially, a database of electrical conductivities of 54 ionic liquids was collected from literature. Then, it was used to develop two models; a quantitative structure property relationship and a group contribution model. Since the electrical conductivities of ionic liquids has a complicated temperature- and chemical structure- dependency, the least square support vector machines strategy was used as a non-linear regression tool to correlate the electrical conductivity of ionic liquids. The deviation of the quantitative structure property relationship from the 783 experimental data used in its development (training set) is 1.8%. The validity of the model was then evaluated using another experimental data set comprising 97 experimental data (deviation: 2.5%). Finally, the reproducibility and reliability of the model was successfully assessed using the last experimental dataset of 97 experimental data (deviation: 2.7%).

Using the group contribution model, a training set composed of 863 experimental data was correlated with a low AARD of about 3.1% from the corresponding experimental data. Then, the model was validated using a data set composed of 107 experimental data points with a low AARD of 3.6%. Finally, a test set consists of 107 data points was used for its validation. It shows an AARD of 4.9% for the test set.

In the next stage, the most comprehensive database of normal boiling point temperatures of approximately 18000 pure organic compounds was provided and used to develop a quantitative structure property relationship. In order to develop the model, the sequential search algorithm was initially used to select the best subset of molecular descriptors. In the next step, a three-layer feed

forward artificial neural network was used as a regression tool to develop the final model. It seems that this is the first time that the quantitative structure property relationship technique has successfully been used to handle a large database as large as the one used for normal boiling point temperatures of pure organic compounds. Generally, handling large databases of compounds has always been a challenge in quantitative structure property relationship world due to the handling large number of chemical structures (particularly, the optimization of the chemical structures), the high demand of computational power and very high percentage of failures of the software packages. As a result, this study is regarded as a long step forward in quantitative structure property relationship world.

A comprehensive database of sublimation enthalpies of 1269 pure organic compounds at 298.15 K was successfully compiled from literature and used to develop an accurate group contribution. The model is capable of predicting the sublimation enthalpies of organic compounds at 298.15 K with an acceptable average absolute relative deviation between predicted and experimental values of 6.4%.

Vaporization enthalpies of organic compounds at 298.15 K were also studied in this study. An extensive database of 2530 pure organic compounds was used to develop a comprehensive group contribution model. It demonstrates an acceptable %AARD of 3.7% from experimental data.

Speeds of sound in saturated liquid phase was the next property investigated in this study. Initially, A collection of 1667 experimental data for 74 pure chemical compounds were extracted from the ThermoData Engine of National Institute of Standards and Technology [5]. Then, a least square support vector machines-group contribution model was developed. The model shows a low AARD% of 0.5% from the corresponding experimental data.

In the next part of this study, a simple group contribution model was presented for the prediction of the standard molar chemical exergy of pure organic compounds. It is capable of predicting the standard chemical exergy of pure organic compounds with an acceptable average absolute relative deviation of 1.6% from the literature data of 133 organic compounds.

The largest ever reported databank for refractive indices of approximately 12 000 pure organic compounds was initially provided. A novel computational scheme based on coupling the sequential search strategy with the genetic function approximation (GFA) strategy was used to develop a model for refractive indices of pure organic compounds. It was determined that the strategy can have both the capabilities of handling large databases (the advantage of sequential search algorithm over other subset variable selection methods) and choosing most accurate subset of variables (the advantages of genetic algorithm-based subset variable selection methods such as GFA). The model shows a promising average absolute relative deviation of 0.9 % from the corresponding literature values.

Subsequently, a group contribution model was developed based on the same database. The model shows an average absolute relative deviation of 0.83% from corresponding literature values.

Freezing Point temperature of organic compounds was the last property investigated. Initially, the largest ever reported databank in open literature for freezing points of more than 16 500 pure organic compounds was provided. Then, the sequential search algorithm was successfully applied to derive a model. The model shows an average absolute relative deviations of 12.6% from the corresponding literature values.

The same database was used to develop a group contribution model. The model demonstrated an average absolute relative deviation of 10.76%, which is of adequate accuracy for many practical applications.

Declaration

I, Farhad Gharagheizi, declare that:

- (i) The research reported in this dissertation, except where otherwise indicated and is my original work.
- (ii) This dissertation has not been submitted for any degree or examination at any university.
- (iii) This dissertation does not contain other person's data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
- (iv) This dissertation does not contain other person's 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.
- (v) Where I have reproduced a publication of which I am an author or co-author I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications:
- (vi) This dissertation does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source are detailed in the dissertation and in the References sections.

F Gharagheizi (candidate)

As the candidate's supervisor/co-supervisor I agree/ do not agree to the submission of this dissertation.

Prof. D Ramjugernath (supervisor)

Prof. A.H Mohammadi (co-supervisor)

Acknowledgements

First of all, I thank God who gave me life and thinking ability, who helped me to learn what I didn't know, who gave me diligence and patience to repeatedly try to solve problems, and who has accompanied me in my entire life.

I would like to sincerely thank my beloved family for all their supports in all steps of my life. I cordially thank my wife, Poorandokht, whose patience inspired me and encouraged me. Her contribution in the majority parts of this work are acknowledged.

I would like to express my greatest acknowledgments to my dear thesis supervisors, Prof. Deresh Ramjugernath and Prof. Amir H. Mohammadi. I will never forget their help and support during my PhD studies. I would like to say that I have gained much during their supervisions, which would be of utmost significance in my future career. Also, I would like to thank the staff members of the University of KwaZulu-Natal.

I wish to gratefully thank Prof. William E. Acree (University North Texas) for his significant contribution in providing the experimental database of sublimation and vaporization enthalpies.

I would like to express my gratitude to Dr. Michael Frenkel, the director of National Institute of Standards and Technology (NIST), for providing the latest version of NIST ThermoData Engine for this work.

The financial assistance of the National Research Foundation (NRF) and Pelchem under the Fluorochemical Expansion Initiative (FEI) toward my PhD thesis is highly acknowledged.

Finally, I would like to thank the examiners for taking their valuable time to enhance the level of this thesis.

Durban, July 2014

Farhad Gharagheizi

List of Publications

- 1- **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 Journal](#). 59, 1702-1708, **2013**.
- 2- **F. Gharagheizi**, P. Ilani-Kashkouli, M. Sattari, A. H. Mohammadi, D. Ramjugernath, D. Richon, *Quantitative Structure - Liquid Thermal Conductivity Relationship for Chemical Compounds*, [Fluid Phase Equilibria](#). 355, 52-80, **2013**.
- 3- **F. Gharagheizi**, P. Ilani-Kashkouli, M. Sattari, A. H. Mohammadi, D. Ramjugernath. D. Richon, *A Group Contribution Method for Determination of Thermal Conductivity of Liquid Chemicals at Atmospheric Pressure*, [Journal of Molecular Liquids](#), 190, 223–230, **2013**.
- 4- **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*, [Chemical Engineering Science](#). 80, 326-333, **2012**.
- 5- **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*, [Chemical Engineering Science](#). 84, 557-563, **2012**.
- 6- **F. Gharagheizi**, P. Ilani-Kashkouli, M. Sattari, A. H. Mohammadi, D. Ramjugernath, *Development of a Group Contribution Method for Estimating the Thermal Decomposition Temperature of Ionic Liquids*, [Fluid Phase Equilibria](#). 355, 81-86, **2013**.
- 7- **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*, submitted to [Chemical Engineering Science](#). 101, 478-485, **2013**.

- 8- **F. Gharagheizi**, M. Sattari, P. Ilani-Kashkouli, A. H. Mohammadi, D. Ramjugernath, D. Richon, *Development of a LSSVM-GC Model for Estimating the Electrical Conductivity of Ionic Liquids*, [Chemical Engineering Research and Design](#). 92, 66-79, **2014**.
- 9- **F. Gharagheizi**, S.A. Mirkhani, P. Ilani-Kashkouli, A. H. Mohammadi, D. Ramjugernath, D. Richon, *Determination of the Normal Boiling Point of Chemical Compounds Using a Quantitative Structure–Property Relationship Strategy: Application to a very Large Dataset*, [Fluid Phase Equilibria](#). 354, 250-258, **2013**.
- 10- **F. Gharagheizi**, P. Ilani-Kashkouli, W.E. Acree, A. H. Mohammadi, D. Ramjugernath, A *Group Contribution Model for Determining the Sublimation Enthalpy of Organic Compounds at the Standard Reference Temperature of 298 K*, [Fluid Phase Equilibria](#). 354, 265-285, **2013**.
- 11- **F. Gharagheizi**, P. Ilani-Kashkouli, W.E. Acree, A. H. Mohammadi, D. Ramjugernath, A *Group Contribution Model for Determining the Vaporization Enthalpy of Organic Compounds at the Standard Reference Temperature of 298 K*, [Fluid Phase Equilibria](#). 360, 279–292, **2013**.
- 12- **F. Gharagheizi**, P. Ilani-Kashkouli, A. H. Mohammadi, D. Ramjugernath, *Toward a Group Contribution Method for Determination of Speed of Sound in Saturated Liquids*, submitted to [Journal of Molecular Liquids](#), 194, 159-165, **2014**.
- 13- **F. Gharagheizi**, P. Ilani-Kashkouli, A. H. Mohammadi, D. Ramjugernath, *Toward a Group Contribution Method for Determination of Speed of Sound in Saturated Liquids*, submitted to [Journal of Molecular Liquids](#), 194, 159-165, **2014**.
- 14- **F. Gharagheizi**, P. Ilani-Kashkouli, A. H. Mohammadi, D. Ramjugernath, *A Group Contribution Method for Determination of the Standard Molar Chemical Exergy of Organic Compounds*, [Energy](#) 70, 288-297, **2014**.

15- P. Ilani-Kashkouli, Hashemi, H., **F. Gharagheizi**, Babaei, S., A. H. Mohammadi, D. Ramjugernath, *Gas Hydrate phase Equilibrium in Porous Media: An Assessment Test for Experimental Data*. [Fluid Phase Equilibria](#). 360, 161–168, 2013.

16- P. Ilani-Kashkouli, Babaei, S., **F. Gharagheizi**, Hashemi, H., A. H. Mohammadi, D. Ramjugernath, *Assessment Test of Phase Equilibrium Data of Water Soluble and Insoluble Clathrate Hydrate Formers*. [Fluid Phase Equilibria](#). 360, 68-76, 2013.

17- M. Sattari, **F. Gharagheizi**, P. Ilani-Kashkouli, A. H. Mohammadi, D. Ramjugernath. D. Richon, *Toward a Group Contribution Method for Estimation of Heat Capacity of Ionic Liquids*. [Journal of Thermal Analysis and Calorimetry](#), 363, 27–31, 2013.

18- M. Sattari, **F. Gharagheizi**, P. Ilani-Kashkouli, A. H. Mohammadi, D. Ramjugernath. D. Richon, *Estimation of Heat Capacity of Ionic Liquids: A QSPR Approach*. [Industrial & Engineering Chemistry Research](#). 52, 13217–13221, 2013.

19- M. Sattari, **F. Gharagheizi**, P. Ilani-Kashkouli, A. H. Mohammadi, D. Ramjugernath, *Toward a Group Contribution Method for Determination of Speed of Sound in Saturated Liquids*, [Journal of Molecular Liquids](#), 196, 7-13, 2014.

20- **F. Gharagheizi**, P. Ilani-Kashkouli, A. Kamari, A. H. Mohammadi, D. Ramjugernath, *A Group Contribution Model for the Prediction of Refractive Indices of Organic Compounds*, submitted to [Journal of Chemical and Engineering Data](#), 59, 1930-1943, 2014..

Submitted Manuscripts

21- **F. Gharagheizi**, P. Ilani-Kashkouli, A. Kamari, A. H. Mohammadi, D. Ramjugernath, *Chemical Structure based Model for the Estimation of Refractive Indices of Organic Compounds*, submitted to [Fluid Phase Equilibria](#).

22- F. Gharagheizi, P. Ilani-Kashkouli, A. Kamari, A. H. Mohammadi, D. Ramjugernath, *Freezing Point of Organic Compounds: A Quantitative Structure-Property Relationship Approach*, submitted to [Journal of Molecular Liquids](#).

23- F. Gharagheizi, P. Ilani-Kashkouli, A. Kamari, A. H. Mohammadi, D. Ramjugernath, *A Group Contribution Model for the Prediction of Freezing Point of Organic Compounds*, submitted to [Fluid Phase Equilibria](#).

24- P. Ilani-Kashkouli, Babae, S., **F. Gharagheizi**, Hashemi, H., A. H. Mohammadi, D. Ramjugernath, *Evaluation of Experimental Data for Gas Solubility in Liquid Water in Equilibrium with Gas Hydrates*, submitted to [Neural Computing and Applications](#).

Congresses and Seminars

25- F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, *A Group Contribution Model for Determining the Vaporization Enthalpy of Organic Compounds at the Standard Reference Temperature of 298 K*, South African Chemical Institute (SACI) Convention, 1st -6th December, East London, South Africa, **2013**.

26- P. Ilani-Kashkouli, **F. Gharagheizi**, A. H. Mohammadi, D. Ramjugernath, *A Quantitative Structure-Property Relationship for the Prediction of the Enthalpy of Vaporization of the Pure Organic Fluorochemicals*, South African Chemical Institute (SACI) Convention, 1st -6th December, East London, South Africa, **2013**.

Table of Contents

Abstract	i
Declaration.....	viii
Acknowledgements.....	ix
List of Publications	xi
List of Tables	xix
Chapter 1 Introduction	1
Chapter 2 Literature Survey.....	4
2.1 Liquid Thermal Conductivity of Pure Chemical Compounds at Atmospheric Pressure	6
2.2 Viscosity of Ionic Liquids	8
2.3 Thermal Decomposition Temperature of Ionic Liquids	13
2.4 Electrical Conductivity of Ionic Liquids	14
2.5 Normal Boiling Point Temperature of Pure Organic Compounds.....	17
2.6 Sublimation Enthalpy at the Standard Reference Temperature of 298 K	23
2.7 Vaporization Enthalpy of Organic Compounds at the Standard Reference Temperature of 298 K	27
2.8 Speed of Sound in Saturated Liquids	29
2.9 Standard Molar Chemical Exergy of Organic Compounds	31
2.10 Refractive Indices of Organic Compounds.....	35
2.11 Freezing Point Temperature of Organic Compounds	37
Chapter 3 Mathematical Methods & Techniques.....	39
3.1 Subset variable Selection Techniques.....	39
3.1.1 The Genetic Function Approximation Technique (GFA)	40
3.1.2 The Sequential Search Method (SS).....	42
3.1.3 Artificial Neural Networks.....	43
3.1.4 Least Square Support Vector Machines	45
3.2 The Gene Expression Programming Technique	49
Chapter 4 Databases.....	56
4.1 Liquid Thermal Conductivity of Pure Chemical Compounds at Atmospheric Pressure	56
4.2 Viscosity of Ionic Liquids	56
4.3 Thermal Decomposition Temperature of Ionic Liquids	57
4.4 Electrical Conductivity of Ionic Liquids	58

4.5	Normal Boiling Point Temperature of Pure Organic Compounds.....	59
4.6	Sublimation Enthalpy at the Standard Reference Temperature of 298 K	59
4.7	Vaporization Enthalpy at the Standard Reference Temperature of 298 K	60
4.8	Speed of Sound in Saturated Liquids	61
4.9	Standard Molar Chemical Exergy of Organic Compounds	61
4.10	Refractive Indices of Organic Compounds	62
4.11	Freezing Point Temperature of Organic Compounds	65
Chapter 5	Adaption & Modifications of the Algorithms	67
5.1	Liquid Thermal Conductivity of Pure Chemical Compounds at Atmospheric Pressure	67
5.1.1	Developing a General Model for Liquid Thermal Conductivity.....	67
5.1.2	Developing a Quantitative Structure-Property relationship	69
5.1.3	Developing a Group Contribution Model for Liquid Thermal Conductivity.....	70
5.2	Viscosity of Ionic Liquids	71
5.3	Thermal Decomposition Temperature of Ionic Liquids.....	71
5.3.1	Developing a Quantitative Structure- Property Relationship.....	71
5.3.2	Developing the Group Contribution method.....	72
5.4	Electrical Conductivity of Ionic Liquids	72
5.4.1	Developing a Non-Linear QSPR.....	72
5.4.2	Developing a Non-Linear GC Model.....	73
5.5	Normal Boiling Temperature of Pure Organic Compounds	74
5.5.1	Developing a non-linear QSPR.....	74
5.6	Sublimation Enthalpy at the Standard Reference Temperature of 298 K	75
5.6.1	Developing a GC model.....	75
5.7	Vaporization Enthalpy at the Standard Reference Temperature of 298 K.....	76
5.7.1	Developing a GC model.....	76
5.8	Speed of Sound in Saturated Liquids	76
5.9	The GC model for Determination of the Standard Molar Chemical Exergy of Organic Compounds	78
5.10	Refractive Indices of Organic Compounds	79
5.10.1	The QSPR Model	79
5.10.2	The GC model.....	80
5.11	Freezing Point Temperature of Organic Compounds	82

5.11.1	The QSPR Model	82
5.11.2	The GC Model.....	83
Chapter 6	Results.....	86
6.1	Liquid Thermal Conductivity of Pure Chemical Compounds at Atmospheric Pressure	86
6.1.1	The General Model	86
6.1.2	The QSPR Model	89
6.1.3	The GC Model.....	92
6.2	Viscosity of Ionic Liquids	94
6.3	Thermal Decomposition Temperature of Ionic Liquids	98
6.3.1	The QSPR Model	98
6.3.2	The GC Model.....	102
6.4	Electrical Conductivity of Ionic Liquids.....	104
6.4.1	The Non-Linear QSPR Model	104
6.4.2	The Non-Linear GC model	106
6.5	The Non-Linear QSPR for Normal Boiling Point Temperature	108
6.6	Sublimation Enthalpy at the Standard Reference Temperature of 298 K	112
6.7	Vaporization Enthalpy of Organic Compounds at the Standard Reference Temperature of 298 K 114	
6.8	Speed of Sound in Saturated Liquids	120
6.9	The Standard Molar Chemical Exergy of Organic Compounds	122
6.10	Refractive Indices of Organic Compounds	135
6.10.1	The QSPR model	135
6.10.2	The GC Model.....	138
6.11	Freezing Point Temperature of Organic compounds	139
6.11.1	The QSPR Model	139
6.11.2	The GC Model.....	141
Chapter 7	Conclusions	144
Chapter 8	Recommendations for Future works	149
References	150
Appendix	202
Statistical Parameters: Mathematical Definition	202

List of Tables

<i>Table 2.1. Comparison of the previous models proposed for liquid thermal conductivity of pure compounds.</i>	<i>8</i>
<i>Table 2.2. Comparison of the previous models proposed for viscosity of ILs.</i>	<i>10</i>
<i>Table 2.3. Comparison of the previous models proposed for thermal decomposition temperature of ionic liquids.....</i>	<i>14</i>
<i>Table 2.4. Comparison of the previous models proposed for electrical conductivity of ionic liquids.....</i>	<i>15</i>
<i>Table 2.5. Comparison of the previous models proposed for the normal boiling point temperature of pure organic compounds.</i>	<i>19</i>
<i>Table 2.6. Comparison of the previous models proposed for sublimation enthalpy of pure organic compounds.</i>	<i>24</i>
<i>Table 6.1 The statistical parameters of the general model proposed for the liquid thermal conductivity of pure organic compounds.</i>	<i>88</i>
<i>Table 6.2 The molecular descriptors selected by sequential search algorithm to describe the thermal conductivity of pure organic compounds.</i>	<i>90</i>
<i>Table 6.3 The coefficients of the eq.6.4</i>	<i>91</i>
<i>Table 6.4 The statistical parameters of the model proposed for viscosity of ionic liquids.....</i>	<i>96</i>
<i>Table 6.5. The average absolute relative deviation of the proposed model for viscosity of ionic liquids from the corresponding experimental based on different classes of ionic liquids studied (In the table the units of T and η are respectively in K and) cP.</i>	<i>97</i>

<i>Table 6.6 The deviation of the predicted (the QSPR) thermal decomposition temperatures of ionic liquids from the corresponding experimental data based on different chemical families of ionic liquids.....</i>	<i>101</i>
<i>Table 6.7 The deviation of the predicted (the group contribution model) thermal decomposition temperatures of ionic liquids from the corresponding experimental data based on different chemical families of ionic liquids.</i>	<i>103</i>
<i>Table 6.8 The optimal subset of molecular descriptors describing electrical conductivity of ionic liquids along with their definition.....</i>	<i>105</i>
<i>Table 6.9 The statistical parameters of the proposed QSPR for electrical conductivity of ionic liquids.....</i>	<i>106</i>
<i>Table 6.10. Comparison of the sublimation enthalpy model and the model proposed by Ouvrard and Mitchell [106]</i>	<i>114</i>
<i>Table 6.11. Comparison between the presented vaporization enthalpy model and the previous models suggested by Ducros et al. [141, 142], Guthrie and Taylor [145], Chickos et al. [137], Domalski and Hearing [439] , Constantinou and Gani [138], and Kolská et al. [146], using the data set comprised of 83 compounds used in Santos and Leal [154] studies.....</i>	<i>117</i>
<i>Table 6.12 The contribution of each chemical substructure to the standard molar chemical exergy pure organic compounds (parameters of equation 6.14).....</i>	<i>124</i>
<i>Table 6.13. Comparison of the presented model and the previous model proposed by Gharagheizi for the estimation of standard molar chemical exergy of pure organic compounds.</i>	<i>130</i>

List of Figures

<i>Figure 2.1. The schematic stages of QSPR development.....</i>	<i>5</i>
<i>Figure 3.1 The schematics structure of the 3FFNNs used in this study.....</i>	<i>43</i>
<i>Figure 3.2 A typical computer LISP program in the genetic programming algorithm represented as a parse tree (expression tree), which represents the algebraic expression $[a+(a/b)] \times [b-(a \times b)]$ by a two-gene chromosome.....</i>	<i>52</i>
<i>Figure 3.3. A typical Karva language program in the gene expression programming strategy, which represents the algebraic expression $[(a / b)] + [(\sqrt{a \times c})]$ by a two-gene chromosome....</i>	<i>54</i>
<i>Figure 3.4. The gene expression programming algorithm main steps.....</i>	<i>55</i>
<i>Figure 4.1 Distribution of refractive indices in the databank.....</i>	<i>63</i>
<i>Figure 4.2 Distribution of molecular weights in the refractive index databank.....</i>	<i>64</i>
<i>Figure 4.3 Distribution of atom numbers in the refractive index databank.....</i>	<i>64</i>
<i>Figure 6.1 (Left) Predicted viscosities of ionic liquids by eq.1 versus the corresponding experimental values. (Right) Relative deviation of predicted viscosities of ionic liquids versus the corresponding experimental ones. (*) and (o) denote the training set and the test set.....</i>	<i>95</i>
<i>Figure 6.2 (Left) Comparison of the predicted viscosities of ionic liquids and their corresponding experimental values: approximately 54 % of the viscosities are predicted within 0-5 %, 26 % within 5-10 %, 10 % within 10-15 %, 6 % within 15-20 %, 2 % within 20-25 %, and the remaining 2 % within 25-31 %. (Right) Viscosity as function of temperature (The unit for η is cP).....</i>	<i>97</i>
<i>Figure 6.3 The normal boiling point temperatures distribution of error based on the training, the validation and the test sets.....</i>	<i>110</i>

Chapter 1 Introduction

One of the main obstacles of both the design and optimization of processes and products in chemical industries is that how the properties of the chemical compounds at the process conditions can be accurately estimated. As stated by Mackay et al. [6], there are approximately 50,000 to 100,000 chemical compounds being manufactured commercially worldwide. Approximately, 1000 new compounds are added each year. Despite of many significant efforts that have been undertaken to experimentally measure different properties of chemical compounds, few sets of the properties have been thus far measured and published in open literature. According to the most extensive databases like Dortmund Data Bank [2], NIST ThermoData Engine [7], and DIPPR 801 [4], the number of pure compounds, for which at least a single property value has been measured is around 43,000 [2]. The issue rises more intensely when we deal with the properties of the mixtures. As a result, property estimation methods are of great importance.

Recently, virtual screening of large databases of chemical compounds, mostly not synthesized, has been one of the main steps toward designing new chemical compounds majorly in pharmaceutical related industries. We can refer to the National Center for Biotechnology Information (NCBI) database which contains more than 10 million chemical compounds. As a result, development of new predictive models is critically required to expedite the process of development of new chemical compounds.

As a consequence of the efforts on designing more advanced experimental techniques/apparatuses, the quality and accuracy of the recently measured data have been increased. Additionally, many extensive databases including almost all the published and most of confidential (unpublished) experimental data for various properties have been emerged, in which the uncertainties of

experimental data are reported. This latter issue makes it possible to rank the experimental datasets with respect to their quality. Thus, we may claim that the amount of available experimental data is much larger than those used to develop the existing models for determination of various properties of chemicals.

Furthermore, some novel computational methods (such as Gene Expression Programming) have been recently appeared. On the other hand, some of the currently mathematical algorithms (such as quantitative structure-property relationships) have been significantly improved employing the data mining algorithms (such as support vector machines, and artificial neural networks).

As a result, a comprehensive attempt is required to develop some new accurate models and, indeed, renovate some of the currently available semi-empirical or theoretical models (such as corresponding states methods) using these advanced mathematical tools.

Almost all of the existing methods for the estimation of different properties of pure compounds can be categorized into three main categories as follows:

1. The models, which use other physico-chemical properties to correlate a property of interest (Type 1).
2. The models, which apply chemical structure-based parameters to estimate a desired property (Type 2).
3. The models, which apply both chemical structure-based parameters and other physical properties to estimate a desired property (Type 3).

The main scope of this study has been to develop more comprehensive and more accurate models of first two categories particularly the second one which uses the chemical structure-based

parameters. Consequently, the following steps have been pursued in this thesis to develop new predictive models for several important properties of pure compounds;

- 1- Choosing the appropriate properties of pure compounds to focus on; it should be mentioned that a literature survey was performed to enlist most important thermophysical properties for which there was no sufficiently accurate and comprehensive estimation method.
- 2- Collecting the most comprehensive possible experimental dataset for each property
- 3- Development of the new property prediction methods
- 4- Validation of the models using several statistical methods
- 5- Evaluation of the predictive power of the obtained models in each case
- 6- Comparison between the obtained models and the previous methods in terms of accuracy, reliability, predictive power, and comprehensiveness of the model.

A detailed literature review of the previous models for each property is presented in chapter 2. Chapter 3 devoted to the mathematical techniques used implemented in this study. The modifications of the algorithms are discussed in the chapter 3. Chapter 4 includes a detailed information about the databases used to develop the models. The adaptation/modification of the mathematical techniques is discussed in chapter 5. Chapter 6 dedicated to the obtained models and their results. A brief conclusion is presented in chapter 7. General recommendations for future studies in chapters 8.

Chapter 2 Literature Survey

Nowadays, most of pure compounds primary thermophysical properties such as critical properties, pitzer acentric factors, and normal boiling points have been tabulated in chemical engineering thermodynamics text books which can be used to correlate other properties. In this study, we present a novel computational technique that can be used to develop general models of this type. This method is called Gene Expression Programing and will be extensively discussed in section 3.2. This kind of models are much more popular because they usually have simple shapes and can be easily used in many practical applications.

As mentioned in previous chapter, the number of compounds for which at least one experimental property reported in literature is far less than the nowadays science and engineering requirements. Thus, it is of great interests to develop chemical structure-based models which have predictive capabilities.

Quantitative structure-property relationships strategy (QSPR) is one of the most widely used techniques to develop chemical structure-based models for the estimation of the thermophysical properties of pure compounds and their mixtures. The QSPR assumes that chemical structure of a molecule contains all the required information to describe its thermophysical properties.

The major steps of the QSPR technique is schematically shown in Figure 2.1. Collection of a comprehensive database of experimental data for the desired property is regarded as the first step in QSPR. It significantly affects the comprehensiveness of the final model; the larger database, the more comprehensive model. The next step is to transform the chemical knowledge encoded within a schematic representation of a molecule into useful parameters called “molecular

descriptors”. In other words, the term “molecular descriptor” is used for any chemical structure-based parameter. A detailed explanations about the types of molecular descriptors are presented by Todeschini and Consoni [8]. The values of a significant types of molecular descriptors strongly depend on bond lengths and angles. Thus, the molecular structures geometries should initially be optimized so that all atoms lie in their right places.

According to Todeschini and Consoni [8], a few thousands of molecular descriptors have so far been proposed [8]. Therefore, a computational approach should be implemented in the next step to select the optimal subset of molecular descriptors and develop the model. The final stage of a QSPR study will be the evaluation of the predictive power of the obtained model. A detailed information about the most commonly-used validation techniques are presented by Todeschini and Consoni [8].

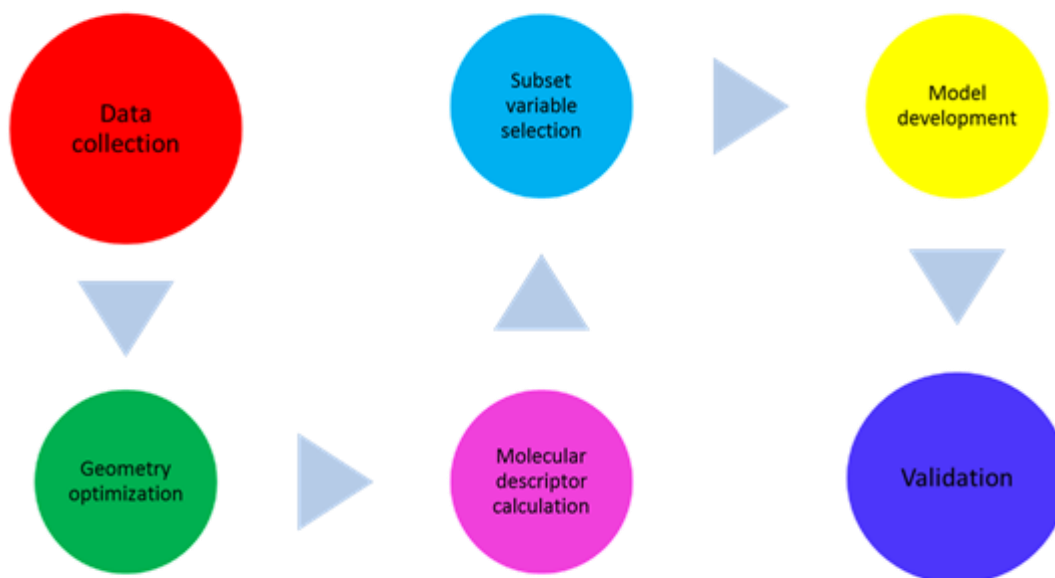


Figure 2.1. The schematic stages of QSPR development

As mentioned earlier, molecular descriptors have different types. One of the most widely used and perhaps simplest types is “functional group counts”. Functional groups can be simply counted from chemical structure without any need to additional tools. These counts can then be used to correlate various properties.

In this chapter, a literature review of the thermophysical properties investigated in this thesis is presented.

2.1 Liquid Thermal Conductivity of Pure Chemical Compounds at Atmospheric Pressure

In most chemical processes, it is needed to perform an energy balance on units, in order to calculate how much heat is either given off or absorbed in various equipment such as evaporators, furnaces, dryers, distillation units, reaction vessels, *etc.* [9]. Heat transfer can take place by three mechanisms, one of which is thermal conduction. In the mechanism, heat transfer takes place due to a temperature gradient [9] which is attributed to vibrational movements of molecules and expressed as Fourier's Law [10].

One of the important transport properties is liquid thermal conductivity which is required for calculation of the thermal conduction. Accurate measurements of thermal conductivity is not straightforward and special attention should be made in experimental measurements due to the possible heat losses and presence of convective currents. Thus, the uncertainty in the reported experimental data is relatively large compared to other thermophysical properties reported [11-15].

The aspects of kinetic theories of thermal conductivity for mono-atomic liquids were initially presented in 1950 [16]. Since then, many researchers have attempted to develop models for the estimation of liquid thermal conductivities of pure compounds. Most of them are empirical in

nature and have been obtained for particular chemical families of compounds and cannot be used for other chemical groups. According to Sastri and Rao [17, 18], in most of the existing methods, thermal conductivity is correlated at a reference temperature such as the normal boiling point (for instance in the models proposed by Sato-Riedel [15, 19], and Sastri and Rao [17, 18]) or at 293.15 K (for instance in the models by Missenard [20] and API Technical Data Book [21]). Thus, the temperature dependency of the model is investigated.

A detailed review of the existing models for the estimation of liquid thermal conductivity of pure compounds has been presented by Poling *et al.* [22] which is illustrated in Table 2.1. In this review [22], it has been mentioned that although the methods presented by Latini *et al.* and Baroncini *et al.* [23-29] and Sastri *et al.* [30] are generally better than the others below the normal boiling temperature, the deviation of the models vary widely, typically less than 15%. Additionally, although the [23-29] methods have been successfully applied for refrigerants up to reduced temperatures equal 0.9, it has been mentioned that there are few experimental liquid thermal conductivity data for reduced temperatures greater than 0.65. Therefore, it can be concluded that the models currently available may not accurately predict the liquid thermal conductivity for reduced temperatures greater than 0.65. Furthermore, the model suggested by Baroncini *et al.* and Latini *et al.* [23-29] is presented for several particular chemical families of compounds, viz. olefins, saturated hydrocarbons, cycloparafins, esters, aromatics, organic acids, alcohols, ketones, and refrigerants. These chemical families do not involve a significant number of important chemical families such as mercaptanes, amines, silanes/siloxanes, sulfides/thiophenes, inorganic compounds, epoxides, aldehydes, nitriles, elements, and peroxides. This issue significantly affects its applicability when using the model. As a result, the model cannot be regarded as a general model.

Table 2.1. Comparison of the previous models proposed for liquid thermal conductivity of pure compounds.

Model	Advantage	Shortcomming
Baroncini <i>et al.</i> and Latini <i>et al.</i> [23-29]	<ul style="list-style-type: none"> - Generally good results when $T < NBP$ typically less than 15% deviation from experimental data - Acceptable results for refrigerants up to $T_r = 0.9$ 	<ul style="list-style-type: none"> - They are a series of models (It is not a general model). - They have not been sufficiently evaluated for $T_r > 0.65$ - They have been developed for particular chemical families of compounds which do not cover a significant number of important chemical families such as sulfides/thiophenes, amines, silanes/siloxanes, mercaptanes, inorganic compounds, nitriles, aldehydes, peroxides, elements, and epoxides
Sastri <i>et al.</i> [30]	-It is a single model	<ul style="list-style-type: none"> - The temperature dependency of the model has not been well evaluated (The model has been evaluated just for 748 experimental data for 208 compounds). -Just 23 of 748 data points have $T > NBP$

**NBP* means normal boiling point temperature

+ T_r means reduced temperature.

Moreover, the model proposed by Sastri *et al.* [30] is based on 748 data points for 208 compounds below the normal boiling point, and 186 data points for only 23 pure compounds above the normal boiling point. Therefore, it affirms that the model cannot be regarded as a general model. Moreover, Poling *et al.* [22] stated that none of the existing models can estimate the large changes of thermal conductivity in the critical point region.

2.2 Viscosity of Ionic Liquids

Ionic liquids are a class of salts composed of ions which are usually liquids at ambient conditions with insignificant vapor pressure [31, 32]. There are a number of features which make them appropriate options for superseding the conventional solvents. Their chemical, physical, and biological features have the potential to be adjusted by the combination of appropriate anion-cation pairs. However, this is not a simple task due to the vast possible number of ionic liquids possible via the permutations of anion-cation pairs [33]. As a result, it is required to develop predictive tools capable of contributing to enable tailoring new ionic liquids with desired properties.

Viscosity is the internal resistance to flow or friction resulted from intermolecular interactions and is consequently very important in all chemical processes involving fluid movement or components dissolved in them. For that reason, any pharmaceutical or chemical unit operations dealing with the fluid or energy transfer needs knowledge of the viscosity. It is also regarded as the most important property when considering any scale-up of ILs applications [34, 35].

Prediction of viscosity is often difficult, and comprehensive predictive models will require further experimental data points. There are several models presented for the estimation of viscosity of ILs. The main features of the models are presented in Table 2.2.

Abbott [36] modified the “free volume theory” originally suggested by Fürth [37, 38] and developed a new model for the estimation of viscosity of ionic liquids. He used 11 imidazolium-based ILs viscosity data at three different temperatures (298, 303 and 364 K) to evaluate his model. In spite of good theoretical interpretation, the model could not estimate the viscosity with acceptable accuracy (average absolute relative deviation (AARD) = 122%).

Table 2.2. Comparison of the previous models proposed for viscosity of ILs.

Model	R2	AARD%	Comments	Shortcoming
Abbott [36]		122%	<ul style="list-style-type: none"> - It is a theoretical model. -It is based on 11 imidazolium-based ILs. -Just three temperatures 298, 303, and 364 were studied 	<ul style="list-style-type: none"> - Very primitive - Primitive in temperature decency
Matsuda <i>et al.</i> [39]	0.8971		<ul style="list-style-type: none"> - It is a polynomial expansion-QSPR model. - 300 experimental data points were studied -Just ILs based on pyrrole, alkylamine, pyridine, piperidine and imidazole were studied - Temperature range was 263-353 K - It is a polynomial expansion-QSPR. 	<ul style="list-style-type: none"> - Limited to a few chemical families of ILs - Insufficiently accurate
Bini <i>et al.</i> [40]	0.8755 at 293K 0.9460 at 353 K		<ul style="list-style-type: none"> -It is based on 33 ILs based on imidazolium, pyridinium, piperidinium and morpholinium cations, bearing linear alkyl or oxyalkyl chains - Just two temperatures 293 and 353 K were studied - It is a Group Contribution model. 	<ul style="list-style-type: none"> - Limited to a few chemical families of ILs - Primitive in temperature decency
Gardas and Coutinho [41]		8%	<ul style="list-style-type: none"> -It is based on 498 experimental data for 29 ILs based on imidazolium, pyridinium, and pyrrolidinium - Temperature range was 293-393 K 	<ul style="list-style-type: none"> - Limited to a few chemical families of ILs - The model needs the density of ILs as well.
Ghatee <i>et al.</i> [42, 43]			<ul style="list-style-type: none"> - It is power law empirical model - It is based on 403 experimental data for 49 ILs. 	<ul style="list-style-type: none"> - Lack of predictive power - It is purely regression-based model.
Han <i>et al.</i> [44]			<ul style="list-style-type: none"> - It is a QSPR. - It is based on 84 experimental data for Imidazolium-based ILs at 298 K and 1 atm. 	<ul style="list-style-type: none"> -Limited to a few chemical families of ILs - Without temperature dependency.
Mirkhani <i>et al.</i> [45]		9%	<ul style="list-style-type: none"> - It is a QSPR. - it is based on 435 experimental data for 293 ILs. 	<ul style="list-style-type: none"> - Linear temperature dependency assumption.

Matsuda *et al.* [39] presented a polynomial expansion-QSPR model for the prediction of viscosity of ionic liquids. It used terms related to temperature, cation, the side alkyl branch attached to cation and the other side chain except for the anion and alkyl chain. The model proposed was based on only 300 experimental viscosity data belonging to pyrrole, piperidine alkylamine, pyridine, and imidazole chemical families. Their data points include the temperature range 263 to 353 K. The model shows inferior estimations for pyridine cation and ethylsulfate. The model also shows a squared correlation coefficient (R^2) of 0.8971 compared with experimental data. As mentioned earlier, the model is based on a small number of anions and cations, and therefore the estimation of viscosity of ILs for which either their anion or cation were not part of the data set should be undertaken with caution. Regardless of this, relating the viscosity which has an intermolecular origin merely to the graph structure of the molecules is exceptional. Additionally, for a new anion-cation pair, all calculations need to be repeated to calculate the proper correlation parameters.

Bini *et al.*[40] proposed a QSPR for prediction of viscosity of ILs. They studied 33 ionic liquids based on pyridinium, imidazolium, morpholinium and piperidinium cations, bearing oxyalkyl or linear alkyl chains at two different temperatures (293 and 353 K). After eliminating all data points pertaining to nitrile-functionalized ILs as outliers, they developed a four-parameter model for the prediction of viscosity at 293 and 353 K with the R^2 values of 0.876 and 0.946, respectively. Based on their results, the model could not well describe low-viscosity data. Furthermore, the absence of other ILs chemical families in addition to very limited temperature range, significantly confine its predictive capability.

Gardas and Coutinho [41] proposed an Orrick–Erbar-based group contribution model for the estimation of viscosity of ILs. They studied a large data set of 498 experimental data related to 29

imidazolium-, pyridinium- and pyrrolidinium-based ILs. The data set covers a wide range of temperature from 293 to 393 K and viscosity from 4 to 21000 cP. The model estimates the viscosity of ionic liquids with an AARD of less than 8%. However, the suggested model needs the density of ILs as model parameter, which should be estimated prior to the prediction. On the other hand, they just investigated three classes of ionic liquids which confine the applicability domain of their model. As stated in previous studies, the cation and anion chemical structures profoundly affect the viscosity of ILs. Therefore, the predictive power of the model merely limits its applicability to just the three ILs chemical classes and any extrapolation for other classes of ionic liquids may result in erroneous predictions. Gardas and Coutinho [46] presented a new model based on the Vogel-Tammann-Fulcher equation without any density term. However, the model application still limited just top three classes of ILs.

Ghatee *et al.* [42, 43] used the well-known power law equation of viscosity and suggested a three-parameter model with two terms; characteristic exponent parameters and temperature dependent. They studied 403 experimental viscosity data for 49 ionic liquids. The major drawback of their model is that its parameters are solely obtained from statistical and regression computations. In other words, for ionic liquids not present in their study, experimental data are required to determine the model parameters.

Han *et al.* [44] developed a QSPR for estimation of the viscosity of imidazolium-based ILs. First, they compiled a data set of 1731 experimental data for 255 ILs (comprising 79 cations and 71 anions) at various pressures and temperatures from literature. Next, they derived some QSPRs for viscosity of ionic liquids at 298.15 K and 1 atm. For this purpose, they extracted a sub-data set containing 84 viscosity data from their data set. They divided the data set into 4 subsets and derived a correlation for each one: $[\text{BF}_4]^-$, $[\text{Tf}_2\text{N}]^-$, $[\text{C4mim}]^+$ and $[\text{C2mim}]^+$. There are several

shortcomings associated with their models. Firstly, the obtained models are not temperature dependent, which limits their application to estimate viscosity merely at 298.15 K. Secondly, their study aimed just imidazolium-based ionic liquids; however, they did not develop a unified model for all ionic liquids present in their study.

Recently, Mirkhani and Gharagheizi [45] developed a QSPR for the prediction of viscosity of ionic liquids. The model shows an AARD of less than 9% from a data set comprised of 435 experimental data for 293 ILs.

2.3 Thermal Decomposition Temperature of Ionic Liquids

Thermal decomposition temperature of ionic liquids (T_d) refers to the maximum temperature at which ILs can exist before undergoing chemical decomposing. In other words, it is a measure of the maximum processing temperature of ionic liquids.

Very recently, two correlations have been proposed for prediction of T_d of ionic liquids. Implementing a data set comprised of 198 experimental data to develop and validate his model (120 data points for developing the model and the remaining 78 for its validation), Lazuss [47] developed a group contribution model for the estimation of T_d of ionic liquids. The model uses 58 functional group counts (27 cation-based and 31 anion-based) for the prediction of T_d of ionic liquids. In order to find the contribution of every functional group on T_d , he used a genetic algorithm optimization technique to minimize the sum of least square errors as an objective function. The AARD of the model results from experimental data for the model and validations sets are 4.3 % and 4.2 %, respectively.

Recently, Yan et al. [48] studied T_d of ionic liquids. Having collected 158 experimental T_d data, they developed a 25-parameter QSPR (they used 126 data points for developing and the remaining

32 ones for validating their model). The AARD of the model estimations from the corresponding experimental data were reported as 3.1 % (for the training set) and 3.5 % (for test set), respectively.

Table 2.3. Comparison of the previous models proposed for thermal decomposition temperature of ionic liquids.

Model	Number of ILs	AARD%	Comment
Lazuss [47]	198	4.3 for the training 4.2 for the validation set	<ul style="list-style-type: none"> - It is a Group Contribution model. - The model has 58 parameters; 27-cation based and 31-anion based - The training and validations sets contain 120 and 78 experimental data, respectively.
Yan et al. [48]	158	3.1 for the traing set 3.5 for the test set	<ul style="list-style-type: none"> - It is a QSPR. - The mode has 25 molecular descriptors. - The traingin and test sets contain 126 and 32 experimental data, respectively.

2.4 Electrical Conductivity of Ionic Liquids

The electrical conductivity of ionic liquids (σ) is one of the important thermophysical properties, which is difficult to estimate due to its non-linear chemical structure and temperature dependency. According to Coutinho et al. [49], four approaches have so far been suggested to correlate the electrical conductivity of ionic liquids. Shows a quick review of the models. Slattery et al. [50] employed the molecular volume (sum of the ionic volume of the ionic counterparts) to correlate the electrical conductivities of 20 ionic liquids based on $[\text{Tf}_2\text{N}]$, $[\text{MF}_n]$, and $[\text{N}(\text{CN})_2]$. The major shortcoming of the model is that it is anion-independent. In other words, different set of parameters is required for each anion. Furthermore, the model cannot be used for functionalized cations [49]. In another attempt, Bogdanov et al. [51] proposed a model based on “residual volume” approach to estimate the electrical conductivity of ionic liquids. As respects, their temperature independent

model has the same shortcoming of the model presented by Slattery et al. In order to overcome the drawbacks of the previous methods, Eiden et al. [52] tried to develop different correlations to predict the electrical conductivity of ionic liquids. Their models showed root mean square error and squared correlation coefficient of 0.22 and 0.91 from the corresponding experimental data, respectively.

Table 2.4. Comparison of the previous models proposed for electrical conductivity of ionic liquids.

Model	Statistical Parameters	Comment	Shortcoming
Slattery et al. [50]		- 20 ILs based on [Tf2N], [MFn], and [N(CN)2] were studied.	-It is anion independent. -Limited just to ILs based on [Tf2N], [MFn], and [N(CN)2]. - Several parameters are presented for each of ILs homologues series.
Bogdanov et al. [51]		-61 ILs based on imidazolium, ammonium, pyrrolidinium, piperidinium, sulfonium, and phosphonium were evaluated.	- It is temperature independent. - Several parameters are presented for each of ILs homologues series.
Eiden et al. [52]	RMSE= 0.22 R ² =0.91	- 69 ILs (596 experimental data) were studied.	- The correlation need viscosity of ILs.
Gardas and Coutinho [46]	AARD=5.4%	- 15 ILs (300 experimental data) were studied.	- Very primitive
Matsuda <i>et al.</i> [39]	R ² = 0.91	- 206 experimental data were studied.	- It cannot predict low electrical conductivity values.
Tochigi and Yamamoto [53]	R ² =0.9745	- 15 ILs (139 experimental data) were studied.	- Very Primitive
Abbott [54]	AARD=27.5%	- 30 ILs at room temperature were studied.	- Temperature independent
Zhao et al. [55]	AARD=2.2%	- 24 ILs at room temperature were studied.	- Temperature independent

The second approach is to relate molar conductivity to viscosity through Walden rule [56]. Galinski et al. [57] reported that Walden rule is almost correct for a wide range of ionic liquids. Gardas and Coutinho [46] proposed a simple linear correlation between logarithms of electrical conductivity and viscosity of ILs. They used a data set comprised of 300 experimental data related to 15 ionic liquids.

The third approach that has been used to correlate the electrical conductivity of ILs is the QSPR. Matsuda *et al.* [39] implemented the “polynomial expansion approach” to develop a model for the estimation of the electrical conductivity of ionic liquids. They used a data set including 206 experimental data values for fitting of their correlation parameters. The model has several terms for cation, temperature, the alkyl chain of the side branch attached on the cation, the other side chain except for the alkyl chain and anion. The squared correlation coefficient of their model compared with the corresponding experimental data is reported as (R^2) 0.91. Unfortunately, they have not reported the deviation of the model predictions from the corresponding experimental data. It appears that the deviation of their model predictions from experimental data is significantly high (several hundred percent) in low electrical conductivity regions.

Tochigi and Yamamoto [53] proposed a complicated QSPR for the prediction of electrical conductivity of ILs. To develop their model, they used a data set comprised of 139 experimental electrical conductivity data points of 15 different anion families and 5 cation. They have stated that the model can predict the electrical conductivity of ILs with squared correlation coefficient, absolute average error, and standard deviation of 0.9745, 0.457, and 0.63, respectively.

The “hole theory” is the fourth approach which have been used to estimate the electrical conductivity of ionic liquids. Abbott [54] was initially tried this approach on a data set of electrical conductivities for 30 ionic liquids at room temperature. He reported an average absolute relative deviation of 27.5% from the corresponding experimental data. Zhao et al. [55] modified the Abbott’s approach and developed another model. It shows 2.2% deviation from the corresponding experimental data.

2.5 Normal Boiling Point Temperature of Pure Organic Compounds

The normal boiling point temperature (NBP) is defined as the temperature at which the vapor pressure of a saturated liquid equals to vapor pressure of 1 atm. Taking into consideration the zero changes of Gibbs free energy of a pure compound at vapor-liquid equilibrium condition, the Clausius-Clapeyron equation can be derived for the NBP:

2.1

$$NBP = \frac{\Delta H_{\text{vap}}}{\Delta S_{\text{vap}}}$$

where ΔS_{vap} and ΔH_{vap} are the entropy and enthalpy of vaporization, respectively. NBP is one of the few thermophysical properties available for most organic compounds. In addition, NBP is an indicator of the physical state of a pure organic compounds (liquid or solid). Having an accurate knowledge of the NBP is of great interest for design of chemical processes and equipment dealing with fluid phase equilibria; evaluation of environmental impacts; and hazard studies (HAZOP). The NBP has also been commonly implemented to estimate a number of important thermophysical properties for instance critical temperature [58], enthalpy of vaporization [59] and chromatographic retention indices [60].

Besides, the NBP is regarded as a measure of the strength of the intermolecular forces. The stronger the intermolecular forces, the more tightly packed the atoms and, as a result, the higher the normal boiling point. The NBP can therefore be directly correlated to the molecular structure of the molecule.

Chemical structure-based models are of great importance since synthesis and experimental measuring the NBP of new compounds are costly, and even dangerous if the compound under study is radioactive or has other hazardous characteristics.

An exhaustive literature review of prior models are presented in Table 2.5. A careful consideration of the models reveals that they can be divided into three main categories from model point of view. The first category of models are originated from the well-known Clausius-Clapeyron equation. The models are based on either enthalpy or entropy of vaporization. The NBP is not directly predicted in this type of methods and therefore their validities are disputable.

The second category of methods for the prediction of NBP is the group contribution models. The molecular basis of this approach root in the assumption that majority of intermolecular forces in the liquid phase are short-range [61]. In this methods, chemical structure are decomposed into predefined chemical substructures/segments, each of which adds a constant contribution to the value of the desired property of a compound. This method provides exceptional results for small and non-polar molecules.

Third category of the correlations are QSPRs in which the NBP is correlated with molecular descriptors.

Table 2.5. Comparison of the previous models proposed for the normal boiling point temperature of pure organic compounds.

Proposed by	Chemical Family	Number of data points	R^2	RMSE	Comment	Reference
Hine and Ehrenson	Halomethanes	46	-	2.4 K	The model has correlates NBP to normal boiling point of methane. It has four correction parameters for halogens (F, Br, Cl, I)	[62]
Ogilvie and Abu-Elgheit	Alkane+N-heterocyclic compounds	50+40	0.99	2.3 K	1. The model has 8 parameters. They account number of hydrogen types as well as the number of bonds between carbon atoms. The square and Cube of abovementioned factors are used in the model 2. The 10-parameter model using fragment and Bond features	[63]
White	Polycyclic Aromatics	40	0.994	8.59 K	The model correlates NBP to First-Order Valence Molecular Connectivity	[60]
Hansen and Jurs	Olefins	123	0.999	1.78 K	The model correlates the NBP to 7 molecular-based parameters: count of methyl groups, square root of the molecular weight, degree of alkene substitution, count of paths of length 2, 3xp, count of clusters of size 3, count of ring atoms path 3 molecular connectivity, count of ethyl groups	[64]
Lai	Diverse	1169	-	(1.29 %) average absolute percent error	The model is a group contribution type. IT has several correction factors for branched side chains, functional group positions, cyclic rings and Aromatic rings for both hydrogen and non-hydrogen bonding compounds	[65]
Rordorf	halogenated dibenzo-p-dioxins	29	-	15 K	2 nd order polynomial of degree of chlorine substitution	[66]
Smeeks and Jurs	$C_5 - C_8$ aliphatic alcohol	120	0.993	2.24 K	The model has 6 parameters; square root of molecular weight, absolute total sigma charge on molecule, count of CH_3 groups, path-two molecular connectivity, path-three valence molecular connectivity, cluster-three molecular connectivity, count of size five path-clusters	[67]

Balaban et al.	Halogenated Alkanes $C_1 - C_4$	532	0.97	11.59 K	The model has 6 p parameters involving the first molecular connectivity index and the number of H,F,Br,Cl,I	[68]
Balaban et al.	Acyclic Ethers, Peroxides, Acetals, and Their Sulfur Analogues	185	0.971	8.2 K	The model has 3 parameters involving the number of methyl groups and electrotopological state of the O or S heteroatom, the valence molecular connectivity index order 0	[69]
Satou et al.	hydrocarbons in a heavy oil	238	0.993	13.6 K	The model is based on the group contribution approach involving several parameters such as boiling point of an n-paraffin with the same total carbon number, the increment of the atomic group and the number of the particular atomic group. Atomic group are considered as follows: Aromatic rings, Naphthenic rings, Aromatic conjunction carbons, Aromatic inner carbons	[70]
Stanton et al.	Pyrans and Pyrroles	752	0.954	13.1 K	The model has 11 parameters involving: number of of single bonds, valence corrected 1st order molecular connectivity, valence corrected 3rd order molecular connectivity, molecular ID/no. of atoms in molecule, average distance sum molecule connectivity, radius of gyration, max positive partial atomic charge dipole moment and 3 charged partial surface area to account for hydrogen bonding	[71]
Simamora et al.	Rigid Aromatic Compounds	241	0.9994	13.66	A 22- parameter group-contribution model.	[72]
Egolf et al.	Diverse (industrially important compounds)	268	0.988	11.85	The model has 8 parameters involving CPSA descriptors, 3 topological descriptors, 2 descriptors accounting for hydrogen bonding and dipole-dipole interactions	[73]
Stein and Brown	Diverse	4426 (model derivation)+6584 (Test)	-	15.5 (model derivation)+20.4 (Test set)	A 85-parameter group contribution model The authors the model proposed Jobak et al. They categorized the original functional groups to smaller one for the sake of higher accuracy, introducing new groups by combining two or three original functional groups and introducing new groups not available in the original model proposed by Jobak et al.	[74]
Wang et al.	Diverse	541	0.96	17.07 K	A 49-parameter group contribution model	[75]
Krzyzaniak et al.	Diverse (non-hydrogen bonding)	870	0.999	14.4 K	A 24-parameter group contribution model. The authors also use the thermodynamic definition of normal boiling point to	[76]

					correlate enthalpy of boiling and entropy of boiling, as well.	
Tsibanogiannis et al.	Diverse (Medium/High Molecular weight compounds)	126	-	Average Error (1.7%)	The model is based on molecular weight and density at 20 °C.	[77]
Wessel and Jurs	Diverse	296	0.997	6.3 K	The model has 6 parameters including charge on most negative atom partial positive minus partial negative surface area, fractional negative surface area, total paths/total no. of atoms, path cluster 3 molecular connectivity and square root of molecular weight	[78]
Katritzky et al	O, N, Cl, and Br containing compounds	298	0.9732	12.41	The model has 4 parameters involving: AM1 most negative atomic charge, bulk cohesiveness descriptor, area-weighted surface charge of the hydrogen-bonding donor atom and number of Cl atoms	[79]
Katritzky et al	Diverse	584	0.965	15.5	The model has 8 parameters involving: the cubic root of the gravitation index, the charged surface area of the hydrogen-donor atoms, (relative number of F atoms, Number of CN group, the surface area of hydrogen acceptors in the molecule, topographic electronic index, charged surface area of hydrogen and Cl atom	[80]
Balaban et al.	Acyclic saturated and unsaturated carbonyl compounds	200	0.964	6.93	The model has 5 parameters involving: information content indices order 1 and 2, sum of square roots of vertex degrees, average distance based molecular connectivities index as well its modification for the presence of heteroatom and multiple bonds	[81]
Cholakov et al.	Diverse hydrocarbons	235	0.999	4.95	The model has 8 parameters including: Total Energy, Van der Waals volume, Number of C atoms in aliphatic C groups, Number of C atoms in aliphatic CH=CH groups, Unsaturated Van der Waals surface, Bond energy, Number of C atoms in aliphatic CH ₂ groups, total number of C atoms.	[82]
Chalk et al.	Diverse	6000+629	0.947	10.8 (Standard Deviation)	The model is a QSPR based on artificial Neural Networks.	[83]
Cordes and Rarey	Diverse	2550	-	8.9 K	The model is a 96-parameter group contribution.	[84]
Ericksen et al.	Diverse	1141(test)+384 (Train)	-	7.75 K-13 K	The model parameters include group contribution+Van der Waals volume+Molecular weight	[85]

Toropov and Toropova	Acyclic carbonyl compounds	100 Test+100 (train)	0.972-0.975	6.12,6	The model correlates NBP to Nearest Neighboring Code	[86]
Ren	Oxo compounds	184	0.9977	3.99 (Standard Error)	The model correlates NBP to Atomic Index and Xu index	[87]
Nannoolal	Diverse	2850	-	6.37	A 212-parameter group contribution model.	[88]
Tomas Öber	Halogenated aliphatic compounds	240	-	4.9	The model has 6 parameters.	[89]
Sanghvi and Yalkowsky	Diverse	1322	0.98	9.3	The model is based on the definition of the boiling point according to entropy and enthalpy of vaporization.	[90]

2.6 Sublimation Enthalpy at the Standard Reference Temperature of 298 K

The phase change enthalpy is of great interest in chemical and environmental engineering, chemistry, and physics since it represents interactions at molecular level [91].

The sublimation enthalpy is a phase change enthalpy which is a measure of molecular interactions in solid phase [92]. It has many applications in crystal science since it is directly use to define the crystal lattice energy [93]. Consequently, the sublimation enthalpy is implemented to determine the solubility of molecules in solvents, particularly drug molecules [94-99].

The sublimation enthalpy is also employed to assess the contaminants transport in the atmosphere; in modeling environmental fate; to calculate materials discolorations; and represent dispersion of dyes [100]. Moreover, crystalline compounds gas phase formation enthalpy can be determined from the sublimation and combustion enthalpies. Furthermore, it can be employed to predict other thermophysical properties such as vapor pressure through the well-known Clausius-Clapeyron equation [101].

Several methods have so far been suggested for the prediction of sublimation enthalpies of pure organic compounds which are compared in Table 2.6. Rice et al. [102] employed the quantum mechanically computed electrostatic potentials of isolated molecules to develop a model for the estimation of the sublimation enthalpy. They used a data set of 35 pure organic compounds to validate their model.

Table 2.6. Comparison of the pervious models proposed for sublimation enthalpy of pure organic compounds.

Model	Accuracy	N*	Shortcoming
<i>Sublimation enthalpy at 289.15 K</i>			
Rice et al. [102]	RMSE=15 kJ.mole ⁻¹	35	- Very primitive - Complexity of its parameters; Density Functional Theory-based model.
Politzer et al. [103]	R ² =0.95	34	- Very primitive - Complexity of its parameters; Density Functional Theory-based model.
Matheieu and Simonetti [104]	RMSE=0.21 kJ.gr ⁻¹	34	- Very primitive - Complexity of its parameters; Density Functional Theory-based model.
Kim et al. [105]	R ² =0.9	34	- Very primitive - Complexity of its parameters; Density Functional Theory-based model.
Ouvrard and Mitchell [106]	R ² =0.925 (training set) R ² =0.937 (test set)	260 (236 for training set and 35 for test set).	- Not sufficiently accurate - Not sufficiently comprehensive
Politzer et al. [107]	AAE=11.7 kJ.mole ⁻¹	105	- Primitive - Not sufficiently comprehensive
Byrd and Rice [108]	RMSE=12.5 kJ.mole ⁻¹	35	- Very primitive - Complexity of its parameters; Density Functional Theory-based model.
<i>Sublimation enthalpy at triple point</i>			
Gharagheizi [109]	R ² =0.9746 RMSE=27.56 kJ.mole ⁻¹	1348	
Gharagheizi et al. [110]	R ² =0.9746 AARD=3.54% RMSE=4.21 kJ.mole ⁻¹	1384	
Mathieu [111]	R ² =0.986 AARD=3.1%	1300	
Salahinejd et al. [112]	R ² =0.96 RMSE=7.9 kJ.mole ⁻¹	1304	

*N denotes number of compounds studied.

The model shows a root mean square error (*RMSE*) and a maximum deviation 15 and 52 kJ.mole⁻¹ from the corresponding experimental data, respectively. Politzer et al. [103], Matheieu and Simonetti [104], and Kim et al. [105] independently developed several models based on the van der Waals electrostatic surface potentials. They independently developed several parameters to determine the sublimation enthalpy. Although, their models shows low deviation from the corresponding experimental data for a small data set of 34 organic compounds, they are so primitive to be practically used for estimation of sublimation enthalpies of pure organic compounds. The major drawback of their model is that they have been developed just based on very small data sets. Ouvrard and Mitchell [106] developed for the prediction of the sublimation enthalpy. They used a training set of 226 compounds for generating their model and another dataset of 35 compounds as a test set for evaluating the model's predictive power. A comparison between the models results and the corresponding experimental data reveals the squared correlation coefficients (R^2) of 0.925 and 0.937, respectively. Politzer et al. [107] developed a QSPR for the prediction of the sublimation enthalpies. They used the molecular surface area and surface electrostatic potential-based information as input parameters to correlate the sublimation enthalpy. The model predictions shows an average absolute deviation of 11.7 kJ mol⁻¹ from the corresponding experimental data for 105 amino acids and small organic compounds. Byrd and Rice [108] employed quantum mechanical information to correlate the sublimation enthalpy. They mentioned that their model is capable of predicting the sublimation enthalpies of 35 organic compounds with a *RMSE* and maximum error of 12.5 and 217.7 kJ.mole⁻¹, respectively.

So far, a few models have been developed for the prediction of sublimation enthalpy at the triple point. Gaharagheizi [109] suggested a 5-parameter QSPR for the prediction of the sublimation

enthalpies of 1348 pure mostly organic compounds. He compared his model results with DIPPR 801 data and reported the R^2 , RMSE, and maximum absolute relative deviation of 0.9746, 5.46, and 27.56 kJ.mole⁻¹, respectively. Gharagheizi et al. [110] proposed an artificial neural network-group contribution model to predict the sublimation enthalpies of 1384 pure mostly organic chemical compounds. The model shows promising R^2 , average absolute relative deviation (AARD%), and root-mean square error of 0.986, 3.54% and 4.21 kJ.mole⁻¹, respectively.

Employing a subset of the data set used by Gharagheizi [109] (1300 out of 1348 data), A model based on the fragment contributions was developed by Mathieu [111]. The R^2 , RMSE, and AARD% of the model compared with DIPPR 801 are 0.986, 4 kJ.mole⁻¹, and 3.1%, respectively. More recently, Salahinejd et al. [112] used another subset of the data set used by Gharagheizi [109] (1304 out of 1348 data) to develop a QSPR model for the estimation of the sublimation enthalpy. They compared their model results with the corresponding experimental data and reported values of 0.96 and 7.9 kJ.mole⁻¹ for the R^2 and the average absolute error, respectively. As simply found, the model proposed by Salahinejd et al. predicts the sublimation enthalpy with a lower accuracy than the one suggested by Gharagheizi [109]. Neither Mathieu [111], nor Salahinejd et al. [112] mentioned why they removed 48 and 44 compounds, respectively, from the complete data set used by Gharagheizi ²⁰.

A comprehensive comparison among the previous models suggested for the prediction of the sublimation enthalpy of pure compounds shows that:

- 1- Majority of the correlations presented for the prediction of the sublimation enthalpy at 298.15 K have been developed/evaluated for small chemical families of compounds.

Moreover, the most comprehensive data set was used by Ouvrard [106] (261 compounds).

The model developed by Gharagheizi et al. [110] shows better performance among the other methods proposed for the sublimation enthalpy at triple point.

2.7 Vaporization Enthalpy of Organic Compounds at the Standard Reference Temperature of 298 K

The vaporization enthalpy is defined as the difference between the enthalpies of a unit mole of the saturated vapor and the saturated liquid of a compound [22]. The vaporization enthalpy represent the strength of intermolecular interactions existed in vapor/liquid phase.

The vaporization enthalpy is regarded as one of the important properties required in the design and the optimization of chemical processes in which vapor-liquid phase change takes place.

So far, several models have been proposed for the estimation of the vaporization enthalpies of pure organic compounds at the standard temperature of 298.15 K. All the previous models, the can be divided into three major classes based on the type of parameters they use.

The first class are those correlations which correlate the vaporization enthalpy to the other most widely used thermophysical properties, i.e. critical properties, normal boiling temperature, acentric factor, and vapor pressure. The models proposed by Fishtine [113-115], Wadsö [116], and Zhao et al. [117] originate from Trouton's rule [118] which employs the normal boiling temperature to predict the vaporization enthalpy. The Clapeyron [22], Antoine [22], Reidel [119], Wagner [120, 121], and Ambrose–Walton [122] equations implement vapor pressure and critical properties to correlate the vaporization enthalpy. The methods suggested by Pitzer [123], Wang–Shi [124], and

Morgan–Kobayashi [125] are corresponding states correlations which employs critical properties and the acentric factor to estimate the vaporization enthalpy. The models proposed by Giacalone [22], Kistiakowsky [22], Chen [126], Vetere [127, 128], and Liu [129] correlate the vaporization enthalpy to vapor pressure and the normal boiling temperature.

The second class are chemical structure-based models. Several QSPRs have so far been developed by Ren [130], Zhokhova et al. [131], and Krasnykh et al. [132]. Lin et al. [133] and Wang et al. [134] developed a model based on the solvation free energy. To date, several group contribution models have been proposed for the prediction of the vaporization enthalpy [135-154]. Chen [155] developed an approach based on the quasi-chemical lattice theory to estimate the vaporization enthalpy. Stefanis et al. [156] developed a group contribution method based on the theory of non-random hydrogen bonding equation of state.

The third class involve those models in which both the chemical structure-based parameters and the thermophysical properties are employed to correlate the vaporization enthalpy. Screttas and Micha-Screttas [157] developed a model based on the normal boiling temperature and carbon number to correlate the vaporization enthalpy of homologous series of compounds. Greenshields and Rossini [158] proposed a model based on the normal boiling temperature and some parameters taking into account chain branching. So far, several group contribution-based models has been proposed which use other thermophysical properties to correlate the vaporization enthalpy [159-167].

To date several models have also been proposed for the prediction of the normal boiling vaporization enthalpy [168-171] . In this study, we do not review them because it is out of scope of this study

The models within the first and third group suffer from a general drawback. As stated above, they correlate the vaporization enthalpy to other thermophysical properties, namely, they are not applicable when one of the required thermophysical properties are not available. From this point of view, the second class, namely, chemical structure-based models, surpasses the others. Furthermore, the majority of the models presented so far for the correlation of the vaporization enthalpy have been proposed and assessed using relatively small number of experimental data sets. Therefore, there are some limitations in their applications.

2.8 Speed of Sound in Saturated Liquids

Speed of sound is an important property in thermodynamics. Like density, it can be measured with a high level of experimental accuracy, typically at least one order of magnitude greater than other thermodynamic quantities. Density has been conventionally used for the development of models for saturated pure liquids, however recently greater attention has been given to speed of sound measurements for the development of rigorous models capable of describing fluid properties over a wide range of temperatures and pressures. Quick and highly accurate measurement capabilities for the speed of sound make it a reliable quantity to measure in order to represent other thermodynamic properties with high precision. Almost all observable thermodynamic properties of a fluid phase can be directly obtained from the speed of sound by integration of partial differential equations which relate it to the other thermodynamic properties. This procedure of indirect determination of thermodynamic properties via speed of sound offers promising predictions over conventional direct approaches.

An important application of speed of sound is to determine the heat capacities of liquids along with $p - \rho - T$ data. This method may be a good alternative approach instead of the conventional calorimetric method. Furthermore, the heat-capacity ratio γ and the isentropic compressibility

κ_s of pure liquids can be related to speed of sound. The latter offers a promising experimental route to measure these thermodynamic properties.

2.2

$$u^2 = \frac{1}{\rho\kappa_s}$$

2.3

$$u^2 = \frac{\gamma}{\rho\kappa_T}$$

Where:

2.4

$$\kappa_s = \frac{1}{\rho} \left(\frac{\partial \rho}{\partial P} \right)_s$$

2.5

$$\kappa_T = \frac{1}{\rho} \left(\frac{\partial \rho}{\partial P} \right)_T$$

2.6

$$\gamma = \frac{C_p}{C_v}$$

At high pressures where measuring $p - \rho - T$ data can generally be difficult, speed of sound measurements in liquids are probably of the greatest value [172].

Although there are some models that can be used just to estimate the speed of sound of single compounds [173-181], there is no general method for the estimation of the speed of sound of saturated liquids for a wide range of temperatures. A few available estimation approaches for speed of sound for a wide range of temperatures are restricted to mixtures and are based on equation of states [182].

2.9 Standard Molar Chemical Exergy of Organic Compounds

In many engineering systems, all the substances involved may be in stable equilibrium with each other; however, they may not be in mutual equilibrium with their surroundings. This lack of a mutual stable equilibrium may result in the production of shaft work. The second law of thermodynamics (SLT) provides information about the spontaneous direction in which a process proceeds. Thus, it can be used to evaluate how close a system is to stable equilibrium with its surrounding when all the system's species are in equilibrium with each other. In other words, thermodynamic analysis methods based on the SLT are much more comprehensive than the analysis methods that are just based on the first law of thermodynamics, due to the fact that they consider the equilibrium state of a system both internally (between substances involved) and externally (between system and its surroundings).

One of the SLT-based analysis techniques that has attracted the attention of many researchers in various fields is the method of exergy analysis. Exergy is a very useful method for the evaluation of the thermodynamic and energy performance of various chemical processes such as study of auto-cascade Rankine cycles [183], evaluation of polygeneration energy systems [184], environmental impact assessment of turboprop engines [185], investigation of steam methane pre-reforming systems [186], analysis of refrigeration shaft power in industrial clusters [187], analysis of thermosolar and heat pumps [188], utilizing of the criteria for decision making in energy systems

[189], improving the performance of natural gas fired combined cycle power plants by coupling Kalina cycles [190], assessment of silicon production processes [191], modifying pulp and paper mills [192], analysis of oil and gas processing platforms [193], analysis of coal systems [194], and bottoming rankine cycle for engine exhaust heat recovery [195]. In chemical processes, the components of process streams are changed as a result of different unit operations which occur in different unit operations such as distillation towers, reactors, etc. The application of exergy analysis in these systems is however in general limited due to the lack of chemical exergy data for components of interest.

Chemical Exergy (CE) refers to the maximum work obtainable from a system at the temperature and pressure of the reference environment [196]. It is also defined as the minimum work required to synthesize a pure compound at the reference state of the environment from its constituents at the same state. In other words, chemical exergy has two parts; one that is related to the reaction needed to produce species that are not in their stable forms in the reference state, and another that is related to translating the species from the system state to the reference state. Therefore, the reference state should be introduced in terms of temperature and pressure so that the chemical exergy can be calculated. Generally, the reference environment is considered as 298.15 K and 1 atm in terms of temperature and pressure, respectively.

Naturally, there are several compounds that are in their most stable states at the reference environment conditions (e.g. CO₂, O₂, N₂, H₂O, noble gases). Therefore, one is not able to obtain more useful work from these kinds of compounds. Thus, their chemical exergy is equivalent to 0 kJ·mol⁻¹ [197, 198].

From a thermodynamic point of view, the standard molar chemical exergy (SMCE) of a compound can be calculated from the appropriate Gibbs energy of formation and the values of the chemical exergy of its constituent chemical elements. Taking into account the reversible reaction for chemical formation of a compound, the SMCE may be written as follows:

2.7

$$\varepsilon^\circ = \Delta G_f^\circ + \sum n_j \varepsilon_j^\circ$$

where ΔG_f° , n_j , and ε_j represent the standard Gibbs free energy of formation; the number of atoms of the constituent element j; and the SMCE of the constituent element j, respectively.

Several models have, to date, been proposed for the estimation of chemical exergy. Probably the first attempt to correlate the chemical exergy is Rant's work [199] in which a constant ratio of chemical exergy to calorific value for solid, and then for liquid fuels was proposed. Szargut and Styrylska [200] showed that Rant's correlation constant depends on the composition of fuels and they developed a new correlation taking this fact into account.

In another model developed by Shieh and Fan [201] a simple correlation for the estimation of the chemical exergy of materials was introduced that had complicated chemical structures. While developing their model, they assumed that the entropy of a fuel is equal to the entropies of its constituent elements. This assumption is far from the reality in many cases and therefore their correlation is not entirely correct.

A modification of Shieh and Fan's model [201] proposed by Stepanov [202] used the entropy model proposed by Ikumi et al. [203]. Using the same idea, Bilgen [204] implemented the entropy correlation proposed by Eisermann et al. [205] to modify Shieh and Fan's model [201]. The final

models proposed by Stepanov [202] and Bilgen [204] are very complicated to use and have a limited applicability domain because both the models have been proposed for coal or coal liquids.

Song et al. [206] developed a model for the estimation of entropy of major constituents of biomass and then extended Shieh and Fan's model [201] for major components of biomass. According to Szargut and Styrylska's investigations [200], the model proposed by Song et al. [206] demonstrated a high level of accuracy for the estimation of biomass, however, its applicability domain was very narrow and limited to a few substances, as in the case of Stepanov's [202] and Bilgen's [204] models.

Using a database of pure organic compounds, Gharagheizi and Mehrpooya [207] developed a quantitative structure-standard molar chemical exergy relationship. Their model was simple and has just three parameters; viz. molecular weight, sum of atomic polarizabilities (scaled on carbon atoms), and number of atoms in any compound. Although their model was simple, the secondary parameter of the model needed computation using a computer program and could not be simply calculated.

Recently, Song et al. [208] developed a model for liquid and solid fuels on a dry basis. They considered the specific exergy of a dry fuel in two separate parts; one for organic matter and the other for inorganic matter. They subsequently developed a model for the estimation of entropy of organic matter for solid and liquid fuels. Implementing the ash analysis data, they also developed a model for the estimation of chemical exergy of inorganic compounds. They showed that the chemical exergy of both ash and inorganic compounds can be neglected when compared with the chemical exergy of dry solid and liquid fuels.

2.10 Refractive Indices of Organic Compounds

The refractive index (RI), also called the index of refraction, is a measure of the change in velocity of a light wave as it travels from one medium to another [209]. It is also equal to the velocity c of light of a given wavelength in empty space divided by its velocity v in a substance, or $n=c/v$ [210]. The refractive index (n) is frequently employed to characterize organic compounds [211] and it is also one of the most important properties in light scattering measurements of dilute polymer solutions which are applied for the estimation of molecular weight, size, and shape [212]. Values of refractive index can be measured experimentally and are normally used to correlate density and/or other physical properties of chemicals [213]. Information obtained from the RI measurements is therefore valuable in various chemical engineering calculations due to its application in the design of new optical materials. Moreover, RI measurements in combination with density, melting point, boiling point and other analytical data are very useful industrially for the specification and characterization of substances like oils, waxes, sugar syrups, *etc* [214].

As mentioned above, the refractive index is a measure of the change in velocity of a light wave and it is directly related to the molecular state of the material it traverses. RI can thus be related to the density of the material, and it is also directly related to the dielectric constant of the material [209]. Most theoretical treatments for the estimation of RI have been proposed in terms of molar refraction, which quantifies the intrinsic refractive power of the basic structural units of a material [215]. Alternative definitions of molar refraction (R) have been developed by Lorentz and Lorentz [216] (eq 1), Gladstone and Dale [217] (eq 2), and Vogel [218] (eq3) as follows [215]:

2.8

$$R_{LL} = \frac{n^2 - 1}{n^2 + 2} V$$

2.9

$$R_{GD} = (n - 1)V$$

2.10

$$R_V = nM$$

The calculation of the refractive index is based on the above equations, where the molecular refraction (R) is calculated as a sum of corresponding atom and bond contributions, and volume (V) is estimated as a van der Waals volume of the compound divided by the average coefficient of molecular packing [215]. Hence, it is relatively easy to calculate the refractive index using the molar refraction, the molecular volume (V), and molar weight (M). Two major approaches which include the group contribution (GC) strategy and the quantitative structure-property relationship (QSPR) approach have been applied for the estimation of the molar refraction [215]. Agrawal and Jenekhe [219] indicated that the refractive index of π -conjugated polymers estimated by existing group contribution methods can have deviations from experimental values as high as 22%. The source of these discrepancies is believed to be large optical dispersion and π -electron delocalization impacts in conjugated polymers. In order to solve this problem, Yang and Jenekhe [220] proposed new Lorentz and Lorentz [216] molar refraction group contributions for 24 functional groups commonly found in conjugated polymers. They successfully applied these new R_{LL} data to calculate the refractive indices of 33 conjugated polymers (with an average absolute relative deviation average error of 0.9%) [220].

Kier and Hall [221] presented a good QSPR model ($R^2=0.998$) for a dataset of 55 alkanes using four topological indices. The same authors proposed another good QSPR model ($R^2=0.998$) for 24 alkyl substituted benzenes containing six to ten carbon atoms with two topological indices. Brekke et al. [222] used the partial least squares (PLS) technique to relate physical properties, including the RF of 12 component mixtures of n-alkanes, iso-alkanes, cyclo-alkanes, and aromatics to the intensities and the chemical shifts of C NMR spectra. The dataset contained a total of 24 samples, and the average absolute relative deviation of the estimation was 2.5%.

A thorough comparison of the previous models proposed for the estimation of the refractive index of pure chemical compounds reveals that most of the models have been developed/evaluated for small chemical groups/families of compounds. Therefore, it was decided to use a very large database in an attempt to develop a general group contribution relationship for the prediction of the refractive index.

2.11 Freezing Point Temperature of Organic Compounds

Freezing point temperature is regarded as a temperature at which solid phase is in equilibrium with liquids phase. The freezing point is always lower than the melting point for mixtures and for organic compounds such as fats [1].

Freezing point and/or melting point (depending on some considerations in their descriptions) are fundamental physical property specifying the transition temperature between liquid and solid phases [2]. Furthermore, they have been used for the prediction of other physical properties such as aqueous solubility [3-5]. Hence, accurate prediction of this fundamental thermo-physical property seems an essential necessity. To date, there have been a few quantitative structure-property relationships (QSPR) methods, such as the property-property relationships (PPR) [6], and

group contribution methods [7-9] applied in attempt to estimate freezing/melting point. There are some successful estimations of melting points, e.g. for 24 normal alkanes ($R^2=0.998$) using topological indices like the carbon number, Wiener index, and the Balaban distance sum connectivity index [10]. Nevertheless, some models such as the QSPR models proposed by Needham et al. [11] indicate poor predictability ($R^2=0.570$) for their use of 56 normal and branched alkanes. A QSPR model [12] for melting point using a dataset containing 443 mono- and di-substituted benzenes which was correlated with a set of structural parameters, and a nine-parameter model showed a R^2 of 0.837. Descriptors related to hydrogen bonding ability, molecular packing in crystals, and other intermolecular interactions such as charge transfer and dipole-dipole interactions contributed to the prediction of melting point.

Burch et al. [13] recently proposed multi parameters models to estimate melting points of alkanes having 10-20 carbon atoms and only one methyl group, which are of special interest to petroleum engineers manufacturing synthetic diesel fuel. A nonlinear regression model with satisfactory predictability was acquired based on the Wiener path numbers, the number of carbon atoms, the methyl locant index, and the mean Wiener index.

A comprehensive review on the previous methods developed for the prediction of the freezing point of chemical compounds demonstrates that most of them have been developed for small chemical groups/families of compounds using small databanks.

Chapter 3 Mathematical Methods & Techniques

In this thesis, a collection of most advanced computational techniques was used to develop the models. As mentioned in chapter 1, the main aim of this thesis has been developing more accurate and more comprehensive models of Type 1, and Type 2.

In order to develop the correlations of Type 1 (the models in which the property is correlated with other widely used physical properties), the Gene Expression Programming is successfully used. The computational details of the algorithm are discussed in this chapter.

In order to develop the chemical structure-based models, the QSPR is used in which various computational algorithms such as Genetic Function Approximation (GFA), and Sequential Search (SS) are implemented. Also, several non-linear techniques such as Least Square Support Vector Machines (LSSVM) and Artificial Neural Networks (ANN) are used to develop more accurate models. The details of computational algorithms are discussed in this chapter.

It should be mentioned that the Group Contribution method (GC) was also used to develop several models in this thesis. The GC approach is considered as one of the most widely used branches of QSPRs due to its simplicity. The computational algorithms used in this study are identical to those used in general QSPR.

3.1 Subset variable Selection Techniques

In many computational applications, it is required to select an optimal subset of variables from a larger set of variables. There are several mathematical methods for this purpose. A review of the subset variable selection methods are presented elsewhere [8]. In majority of the suggested techniques, a multi-linear equation between the subset of variables and the desired property is

preliminary assumed. Next, the optimal subset of variables is chosen using appropriate mathematical strategies.

In this work, the Genetic Function Approximation (GFA) technique which uses the genetic algorithm optimization technique to find the best subset of variables from a larger set of variables was used.

It should be stated that the GFA method cannot be used when studying large databases of compounds due to two main obstacles as follows;

1. The run-time dramatically goes up when the size of the dataset under consideration becomes larger.
2. The amount of installed memory required on a PC rapidly increases when the size of the dataset under consideration increases.

The sequential search (SS) algorithm was found the only algorithm which is capable of handling large databases of compounds.

It should be mentioned that linear models may not be able to well describe many properties therefore; they can be used in couple with some network-based methods such as Artificial Neural Networks and Support Vector Machines. This idea lead to developing much more accurate model compared with linear models obtained using GFA and SS.

3.1.1 The Genetic Function Approximation Technique (GFA)

The Genetic Function Approximation (GFA) is used to select the optimal subset of variables. GFA is a genetic algorithm-based subset variable selection technique. It involves a multivariate adaptive

regression splines (MARS) algorithm [223] coupled with with a genetic algorithm [224] to develop a series of multi-linear equations that best fit the training set data. The strategy was originally suggested by Rogers and Hopfinger [225].

Generally, a QSPR model is presented as follows:

3.1

$$F(X) = a_0 + \sum_{k=1}^M a_k X_k$$

where a_0 is the intercept, a_k is the model coefficient and X_k s are input molecular descriptors (model variables). The initial equations are generated by random selection of the number of molecular descriptors. In the next stage, the derived models are assessed using Friedman's LOF scoring function as follows:

3.2

$$LOF(\text{model}) = \frac{1}{N} \frac{LSE(\text{model})}{\left(1 - \frac{(c + 1 + (d \times p))}{N}\right)^2}$$

where c is the number of non-constant basis functions, N is the number of compounds in the data set, d is a model smoothing factor to be adjusted by the user, and p is the total number of parameters in the equation and the LSE denotes the least square error of the model. LOF directs the computations to equations with better predictive capability without over-fitting.

The *genetic recombination* or *crossover* operation should be repeatedly performed as follows:

- Two good model equations are chosen based on their fitness as “parents”.
- Each parent is randomly 'cut' into two parts and finally two new model equations emerges (children).

- The parents with the worst fitness is replaced by their children.
- The process stops when no more improvement in the fitness functions is observed.

The method has been successfully employed in our previous [21, 45, 226-233].

In this study, the number of population and the number of maximum generations are set to 100 and 5000, respectively. The Mutation probability parameter is set 1.5 in this study.

3.1.2 The Sequential Search Method (SS)

The basic idea of the SS method is to replace each variable one at a time with all the remaining ones and see whether an improved model is obtained. The major steps of the algorithm are as follows:

step 1- Introducing all the variables

step 2- Considering all the one-variable linear correlations between the property under consideration and the variables

step 3- The first optimal variable is the one which predict the property under consideration with lowest possible AARD%

3.3

$$AARD\% = \frac{100}{N} \sum_i^N \frac{|pred - exp|}{exp}$$

step 4- Define number of variable(s) in the first step i=1

step 5- Considering all the linear correlations between the desired property and i+1 molecular descriptors including the i optimal variable(s) selected in the previous step(s)

step 6- The optimal subset of $i+1$ variables is the one whose linear correlation predicts the desired property the lowest AARD%

step 7- Check if the difference between AARD% of the model obtained using the optimal subset of $i+1$ variables and the AARD% the model obtained using the optimal subset of i input variable is less than the stopping criteria. If not $i=i+1$ then go to step 5. If yes the model is obtained.

3.1.3 Artificial Neural Networks

Artificial Neural Networks (*ANNs*) are composed of simple elements working in a parallel computational strategy. These elements are called neurons and inspired by biological nervous systems [234].

Several types of *ANN* strategies have been proposed in the literature. One of the most-widely used is called “Three-Layer Feed Forward Neural Networks (*3FFNN*) with Sigmoid (Hyperbolic Tangent) Transfer Function”. This type is used to generate a non-linear correlation between input and output parameters. In other word, this strategy is used as a regression tool. The schematic structure of the *3FFNNs* is shown in Figure 3.1 [234].

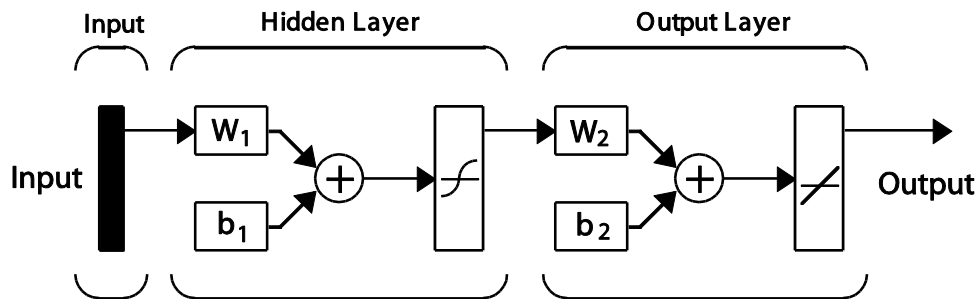


Figure 3.1 The schematics structure of the 3FFNNs used in this study

This type of *ANN* contains three layers. The first one is the input layer and the second one is the hidden layer of the *3FFNN*. It consists of n neurons. The parameter n is one of the important parameters of the *3FFNN* method and needs to be obtained by optimization just after generating the main structure of the *3FFNN*. The mathematic formulation of a *3FFNN* is as follows:

3.4

$$Output(i) = W_2 \times \tanh(W_1 \times Input(i) + b_1) + b_2$$

where i denotes the data point, the parameter W_1 represents the first layer weight, W_2 denotes the second layer weight, b_1 is the second layer bias, and b_2 stands for third or output layer bias. W_1 , W_2 , b_1 , and b_2 needs to be obtained by a process called “Training”.

Generally, in many scientific and engineering problems, input and output parameters should be defined for solving the problem. To generate a *3FFNN* model, some parameters should be defined regarding the proposed problem. The parameters include the number of inputs (nip), number of outputs (nop), and the number of objects ($nobj$) *i.e.* number of data points with known input parameter (ip) and corresponding output parameter (op).

The dimensions of each part in equation 3.4 can be determined using n neurons. The input (i) is a row containing nip input parameters of the i^{th} object, so it is of dimension $nip \times 1$ and W_1 is of dimension $n \times nip$. Therefore, the $W_1 \times Input(i)$ should be of dimension $n \times nip$. The b_1 is of dimension $n \times nip$. Thus, the results of the second layer output are of dimension $n \times 1$. It is easy to figure out the dimensions of W_2 and B_2 . W_2 should be of dimension $nop \times n$, and b_2 should be of dimension $nop \times 1$. As a result, the $output(i)$ is of dimension $nop \times 1$.

There are unknown parameters for a desired number of neurons n , which in a *3FFNN* model are W_1 , W_2 , b_1 , and b_2 . These parameters should be obtained using the “Training” process. This process

applies an optimization method to minimize an objective function defined between output and the experimental values.

The parameters W_1 , W_2 , b_1 , b_2 are the most significant factors for selecting the optimization method. Firstly and perhaps most importantly, the method should be accurate enough to deal with the specified problem. Another point to consider is that the applied method should be robust when we have several tens or hundreds of parameters in a simple *3FFNN*. The Levenberg - Marquart method is the most widely used technique for optimization of the ANNs which was implemented in this thesis.

3.1.4 Least Square Support Vector Machines

Artificial Neural Networks (ANN) approach have shown high accuracy [235]. One of their main drawbacks may be the non-reproducibility of their results, partly as a result of random initialization of their structure and variation of the stopping criteria during optimization [236-238]. The support vector machines technique (SVM) is a popular strategy developed from the machine-learning community. The advantages of the SVM methods in comparison with the traditional ANNs are generally as follows [235-238]:

1. More probability for convergence to the global optimum;
2. Normally find a solution that can be quickly obtained by a standard algorithm (quadratic programming);
3. No need to determine the network topology in advance; which can be automatically determined as the training process ends;

4. There is generally a lower probability of the SVM strategy to be faced with the problem of over-fitting or under-fitting due to a lower number of parameters required for its development than the ANN methods.

The SVM has superior performance over the traditional empirical risk minimization principles. Furthermore, as a result of their specific formulation, sparse solutions can be found and both linear and nonlinear regression can be performed [235-238].

Suykens et al. [235] have considered a modification to the original SVM [235] to overcome the weakness of the previous algorithm in finding the final solution because it requires the solution of a set of nonlinear equations (quadratic programming). Their method, named as Least-Squares Support Vector Machines (LSSVM), encompasses advantages similar to those of SVM, but it requires the solving of a set of only linear equations (linear programming), which is much easier and more rapid compared to the traditional SVM method [235-238].

The cost function (penalized cost function) of the applied [235] method has been defined as follows [235-238]:

3.5

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

subject to the following constraint:

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

In the two abovementioned equations, x is the input vector of parameters of the model (molecular descriptor), y denotes the outputs (the desired property), b is the intercept of the linear regression in LSSVM method, w is the regression weight (the linear regression slope), e_k denotes the regression error for N training objects (the least-squares error approach), γ represents the relative weight of the regression errors summation compared to the regression weight (first right hand side of equation 3.5), and superscript T denotes the transpose matrix.

Employing the Lagrange function [235-238], the weight coefficient (w) can be written as follows [235-238]:

3.7

$$w = \sum_{k=1}^N \alpha_k x_k$$

where,

3.8

$$\alpha_k = 2\gamma e_k$$

Generally in LSSVM algorithm, a linear regression between the independent and dependent variables is assumed. Therefore, equation 3.6 can be re-written as [235-238]:

3.9

$$y = \sum_{k=1}^N \alpha_k x_k^T x + b$$

Therefore, the Lagrange multipliers are computed as [235-238]:

3.10

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

The abovementioned linear regression can be well developed to a non-linear one using the Kernel function as follows [235-238]:

3.11

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

where $K(x, x_k)$ is the Kernel function obtained from the inner product of the two vectors x and x_k in the feasible space built by the inner product of the vectors $\Phi(x)$ and $\Phi(x_k)$ as follows [235-238]:

3.12

$$K(x, x_k) = \Phi(x)^T \cdot \Phi(x_k)$$

In this work, the radial basis function (RBF) Kernel has been implemented as below [235-238]:

3.13

$$K(x, x_k) = \exp(-\|x_k - x\|^2 / \bar{\sigma}^2)$$

where $\bar{\sigma}^2$ is regarded to be a decision variable, which is optimized by an external optimization algorithm when developing the model. The mean square error (MSE) of the LSSVM predictions [235] is defined as:

$$MSE = \frac{\sum_{i=1}^n (A_{\text{pred}_i} - A_{\text{exp}_i})^2}{N}$$

where A_{pred} and A_{exp} stand for the predicted, and experimental values of the desired property.

3.2 The Gene Expression Programming Technique

To date, several theoretical approaches have been used to develop models for the estimation of various thermophysical properties [22]. These models can then be used to predict the corresponding properties other systems at specific conditions within their applicability domain. Nevertheless, the stated thermophysical properties may not be estimated merely using the basic thermodynamics. Although the promising advances that have been achieved in recent years, the thermodynamic basic concepts bring about reduction in the complexity of the molecular-based theories. This can be done by correlating one physical property to another one. As consequence of that two main categories of models emerge: the empirical models and the semi-empirical ones.

The empirical models can estimate many thermophysical properties within the range of the conditions and the compounds and/or mixtures, which have been implemented for their development; however, any extra-polation cannot be recommended. The semi-empirical models employ theoretical concepts introducing some parameters to modify the empirical models. As results, coupling empirical models with theoretical relations generally leads to powerful reliable and predictive correlations.

In any case, certain parameters of the abovementioned models should be regressed over selected experimental data sets. Many mathematical linear and nonlinear regression strategies have been so

far suggested for this purpose. A careful consideration of the techniques reveals that they should satisfy the following criteria [239-247]:

1. Have the capability of converging to the global optimum;
2. High probability of finding a mathematically-correct solution;
3. No requirement for determination of the mathematical functionality in advance
4. Low probability to be encountered with over-fitting/under-fitting problems;
5. Satisfactory generalization performance;
6. Lower number of adjustable parameters;
7. Relying on the population-based initialization;
8. Use of the stochastic evolutionary principles basis;
9. Ability to handle non-differentiable, nonlinear, and multimodal cost functions;
10. Few, robust, and easy to choose control parameters to steer the minimization of the objective function;
11. No sensitivity to starting points i.e. starting decision variables or objective function values
12. Consistent and consecutive improvement of the solutions in each generation.

The abovementioned criteria would eventually contribute to developing reliable predictive models that yield satisfactory results in acceptable computational time.

The Genetic Algorithm (GA), initially introduced by Holland [248], is regarded as a heuristic optimization technique (among the evolutionary algorithms) that inspired by the process of natural evolution. Generally in biology, an organism involves a set of rules, explaining the way that the organism is constructed from the very small building blocks of life. These rules are generally preserved in the genes that are connected together in the shape of strings called chromosomes. The genetic algorithm generally develops mathematical chromosomes (population of strings), which encode solutions to optimization problems through specific operators like selection, mutation, and crossover [240].

The final solutions of a genetic algorithm technique are encoded in fixed length binary (0 and 1) strings. The improvements of this algorithm mainly include manipulation of the mentioned operators. The Genetic Programming (GP) [249, 250] is an effective modification of the genetic algorithm, in which the solutions are represented as nonlinear structures of parse trees (behaved as functions) instead of fixed length binary solutions. This improvement brings about searching among variety of possible functions for finding the final solution [249, 250]. Considering the drawbacks of the GP [249, 250] (which will be discussed later), Ferreira [251] introduced a new modification to the original Genetic Programming algorithm [249, 250]. In the new technique, called “Gene Expression Programming (GEP)” [251], ramified structures of different sizes and shapes (parse trees) are completely encoded in the linear solutions of fixed length that finally result in more effective models [240, 251]. The details of the GEP algorithm are described in the next sections.

3.2.1.1 Genetic Programming

As stated above, the genetic programming [249, 250] is an extension of the genetic algorithms. The defined problem (the forms of the functions, number of parameters etc.) has no effect on the

genetic programming algorithm [249, 250]. The main difference between the GP [249, 250] and the GA [248] is that in the first one, the chromosomes contain of nonlinear structures similar to parse trees though they are similar to the GA [248] linear structures, which are naked replicators working as genotype and phenotype. These parse trees, adopted like the protein molecules, involve diverse forms of functionality. Thus, the final solution of a particular problem can be found among more various types of functions. As a matter of fact, the genetic operators (such as recombination, crossover, and mutation) also operate during the computational steps of the genetic programming [249, 250] exactly similar to those of the original genetic algorithm [248] however they resemble to pruning and grafting of trees. As mentioned by Ferreira [240], the main shortcomings of the genetic programming is that the complex replicators (parse trees structures) can be only amended in limited ranges because their reproduction should be done just on the parse trees.

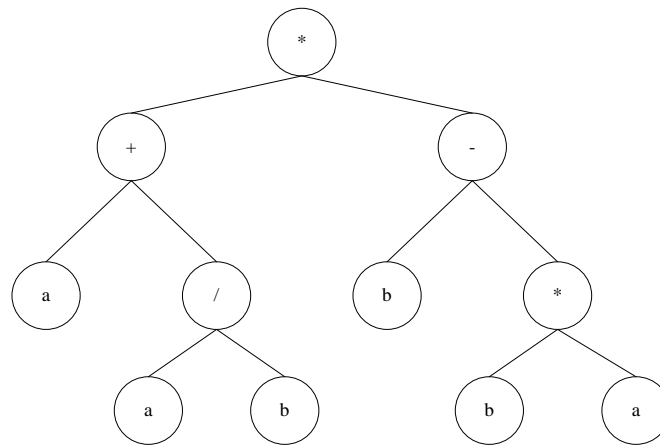


Figure 3.2 A typical computer LISP program in the genetic programming algorithm represented as a parse tree (expression tree), which represents the algebraic expression $[a+(a/b)] \times [b-(a \times b)]$ by a two-gene chromosome

These modifications involve modifying or exchanging definite branches of the corresponding parse trees [240], that may bring about invalid (unacceptable) trees structures. A typical computer LISP program based on the genetic programming [249, 250] algorithm is shown in Figure 3.2.

3.2.1.2 Gene Expression Programming

The gene expression programming is a modification of the genetic programming. In the gene expression programming, the population individuals are symbolic expression trees unlike those of genetic programming, in which the individuals are encoded as linear chromosomes, which are later translated into the expression parse trees i.e. the genotype and phenotype are eventually separated. Therefore, the gene expression programming algorithm has many of the advantages of the evolutionary algorithms [252]. Another element of the gene expression programming is that the chromosomes are such designed that can permit the creation of multiple genes. Consequently, the novel structures of the genes in the gene expression programming algorithm, allows encoding of any program for efficient evolution of the solutions. The organized structure of the genes also permits powerful and efficient genetic operators searching for the solutions in the entire feasible space of the problem. These operators are directly recombined on the linear encoding (before it is translated into a tree). Recombination, as a matter of fact, is sharing the information from the genes of the parents to the gene of the offspring. As a result, the improved parts of the resulting expression trees generally experience little resemblance to their previous ancestors. On the other hand, the GEP special codes use Karva language to make it possible to infer exactly the phenotype given the sequence of a gene and vice versa. For instance, the algebraic expression $(a/b) + (\sqrt{a \times c})$ can be easily shown as a diagram or Expression Tree (ET) like Figure 3.3, with the Karva language representation of $* / Q + a b a c$ (Q stands for the squared root function). Each characters places in one position from 0 to 7 and can be shown as 0 1 2 3 4 5 6 7. To conclude, the operators in this strategy bring about generation of the valid parse trees that may be complex mathematical structure or even artificial neural networks.

3.2.1.3 The gene expression programming general computational steps

Ferreira [240] presented the general computational steps of the gene expression programming strategy as follows:

1. Initialization of the population comprising the random generated chromosomes of a certain number of individuals;
2. Fitness of the individuals based on fitness functions (cases);
3. Choosing the individuals based on their fitness to regenerate with modification.
4. The new individuals are treated using the same procedure including expression of the genomes, confrontation of the selection environment, selection, and reproduction with modification.

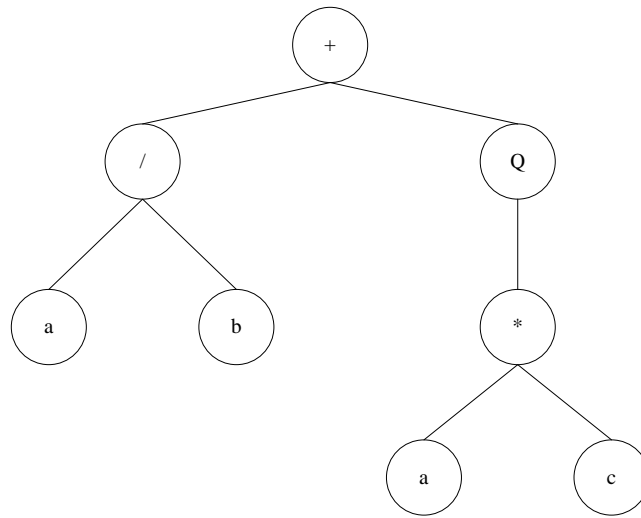


Figure 3.3. A typical Karva language program in the gene expression programming strategy, which represents the algebraic expression $[(a / b)] + [(\sqrt{a \times c})]$ by a two-gene chromosome

5. Initialization of the population comprising the random generated chromosomes of a certain number of individuals;
6. Fitness of the individuals according to fitness functions (cases);
7. Choosing the individuals according to their fitness to reproduce with modification.

8. The new individuals are treated using the same procedure including expression of the genomes, confrontation of the selection environment, selection, and reproduction with modification.

9. Repeating the aforementioned steps for a certain number of generations or until a good solution has been found (convergence of the algorithm according to the defined criteria).

The same algorithm has been followed in this section to develop the model.

The above mentioned steps are schematically shown in Figure 3.4.

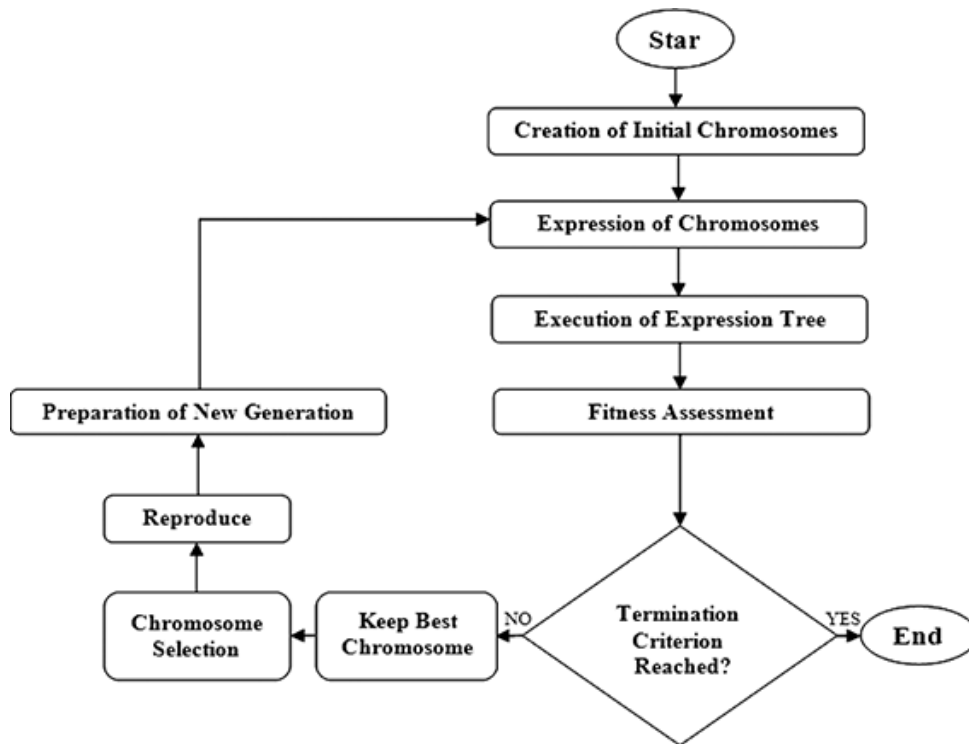


Figure 3.4. The gene expression programming algorithm main steps.

Chapter 4 Databases

4.1 Liquid Thermal Conductivity of Pure Chemical Compounds at Atmospheric Pressure

In this study, approximately 19000 liquid thermal conductivities of more than 1600 pure mostly organic compounds at a pressure of 1 atm for temperatures below normal boiling point and at saturation pressure for temperatures above the normal boiling point were extracted from DIPPR 801 database [4] and used to develop and validate the model.

The database [3] has been randomly split into three sub-data sets; the “training”, the “validation”, and the “prediction” sets. The shares of the training, the validation and the test sets from the database have been 80 %, 10 % , and 10 %, respectively [252]. The database are presented as supplementary file in the supplementary DVD.

4.2 Viscosity of Ionic Liquids

An extensive literature survey was conducted to collect a comprehensive experimental viscosity data set for ionic liquids at atmospheric pressure. Consequently, 1672 data points related 443 ILs comprised of 191 cations and 76 anions were collected from 204 different references.

The most recent corresponding data point was included in case of multiple reported data points. The ionic liquids studied were categorized into 17 different chemical groups containing 1,3-dialkyl imidazolium, 1-alkyl imidazolium, amino acids, ammonium, double imidazolium, guanidinium, isoquinolinium, morpholinium, oxazolidinium, phosphonium, piperidinium, pyridinium, pyrrolidinium, pyrroline, sulfonium, tetra-alkyl imidazolium and tri-alkyl imidazolium. The database are presented as supplementary file in the supplementary DVD.

To develop a valid and reliable model, the data set were split into the two sections: the “training” and the “test” sets. The ”training” set was implemented to obtain the model and the ”test” set was

employed to assess the predictive power of the obtained model. The K-means clustering technique was employed to choose the training and the test sets. This method is used to split n data points into k partitions in which each data point belongs to the partition with the closest mean. As discussed in a previous article the shares of the training and the test sets from the main data set should be selected with caution [253]. As a result, approximately 20 % of the data including 336 data points was used for the test set. The remaining 80% data was employed for developing the model.

4.3 Thermal Decomposition Temperature of Ionic Liquids

A comprehensive literature survey was conducted to gather an extensive experimental T_d data set for ionic liquids. Consequently, the experimental T_d data for 586 ionic liquids composed of 305 cations and 52 anions and were extracted from 71 peer-reviewed reference sources. It should be mentioned that in case of multiple reported values for a unique ionic liquid, the most recent one was kept in the data set for further studies.

The ionic liquids within the data set were classified into 21 chemical groups containing 1,3-Dialkyl imidazolium, 1-Alkyl imidazolium, 4,4-Dimethylimidazolium, Amino acids, Ammonium, Double imidazolium, Guanidinium, Morpholinium, Oxazolidinium, Phosphonium, Piperidinium, Pyridazinium, Pyridinium, Pyrrolidinium, Quinary alkyl imidazolium, Sulfonium, Tetra-alkyl imidazolium, Tetrazolium, Tri-alkyl imidazolium, Triazolium, and Uronium.

The database are presented as supplementary file in the supplementary DVD.

Employing the K-means clustering technique similar to the previous section, the T_d data set was also divided into two sub-data sets including the training and the test sets, respectively. 20 % of

the data set including 119 data points was assigned to the test set. The remaining data points was allocated to the training set for model development purpose.

4.4 Electrical Conductivity of Ionic Liquids

An extensive experimental electrical conductivity data set for ionic liquids was collected from 21 different references. The database includes 977 data points for 54 ionic liquids composed of 18 cations and 17 anions. In case of multiple published experimental data point at a fixed temperature, the most recent ones (from chronological point of view) were kept in the data set. The data set encompass a wide range of electrical conductivity values from 0.00009 to 20 S.m⁻¹ and a wide range of temperatures from 238 to 480 K. Before doing any further processing on the data set, all the experimental data have been compared to remove doubtful data. Finally, 100 experimental data were eliminated. As result, the database used to next computations contains 977 experimental data.

The database are presented as supplementary file in the supplementary DVD.

Employing the K-means clustering technique similar to the previous section, the electrical conductivity data set was also divided into three sub-data sets including the training, the validation and the test sets, respectively. 10 % of the data set including 97 data points was assigned to the test set. Similar to the test set, 10 % of the data set including 97 data points was allocated to the validation set. The remaining data points (783 data points) were employed as the training set for model development purpose.

4.5 Normal Boiling Point Temperature of Pure Organic Compounds

An enormous database of normal boiling point temperature data for 17768 chemically diverse organic compounds extracted from Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds [254] was employed to develop and then to validate a QSPR model. The database was randomly divided into three sub-data sets including the training, the validation and the test sets, respectively. 10 % of the data set including 1776 normal boiling point temperatures was assigned to the test set. Similar to the test set, 10 % of the data set including 1776 normal boiling point temperatures was allocated to the validation set. The remaining 80 % of the database including 14216 normal boiling point temperatures (863 data points) was employed as the training set for model development purpose.

The database are presented as supplementary file in the supplementary DVD.

4.6 Sublimation Enthalpy at the Standard Reference Temperature of 298 K

Inj this study a databased proposed by Acree and Chickos [91] together with the most recent published sublimation enthalpy data during the past three years [255-327] was employed to develop the model and then to validate it.

The data is comprised of 1645 experimental sublimation enthalpies of 1269 compounds. A single experimental sublimation enthalpy was reported for 1018 compounds. Multiple values were found for 251, in which case the arithmetic averages were included in the data set. No attempt was performed to compare the independently determined values, which mostly differed by less than 6 kJ mol⁻¹. A careful analysis of the compounds within the dataset shows that the sublimation enthalpies range between 34 and 240 kJ.mole⁻¹.

Implementing the K-means clustering technique explained earlier, the data set was divided into three sub-data sets including the training, the validation and the test sets, respectively. 10 % of the data set including 127 data points was assigned to the test set. Similar to the test set, 10 % of the data set including 127 data points was allocated to the validation set. The remaining data points (1015 data points) were employed as the training set for model development purpose.

The database are presented as supplementary file in the supplementary DVD.

4.7 Vaporization Enthalpy at the Standard Reference Temperature of 298 K

The collection of vaporization enthalpies at the standard temperature of 298.15 K collected by Acree and Chickos [91] was employed to model development and then to model validation. The recently published vaporization enthalpies were also included in the database [257-259, 264, 266, 273, 277, 280, 282, 286, 290, 296-298, 305, 317, 323-325, 328-389].

The data set includes 4320 experimental vaporization enthalpies of 2811 pure organic compounds. There are just 2049 pure compounds for which just a single experimental data point has been reported in literature. In other words, multiple vaporization enthalpies were published for the 762 compounds. In this study, the arithmetic average of multiple reported values was included in the data set. It should be mentioned that no attempt was performed to compare independently reported vaporization enthalpies, which mostly part differed by less than 9 kJ.mol⁻¹. A quick analysis of the data set demonstrates that the vaporization enthalpies range between 15.6 and 424.5 kJ.mole⁻¹.

Using the K-means clustering technique mentioned earlier, the data set was divided into three sub-data sets including the training, the validation and the test sets, respectively. 10 % of the data set

including 281 data points was assigned to the test set. Similar to the test set, 10 % of the data set including 281 data points was allocated to the validation set. The remaining data points (2249 data points) were employed as the training set for model development purpose.

The database are presented as supplementary file in the supplementary DVD.

4.8 Speed of Sound in Saturated Liquids

In order to develop a reliable and accurate model, application a high quality data set is essential. For this purpose, the ThermoData Engine of the National Institute of Standard and Technology [5] was used in the present work to provide 1667 saturated liquid speed of sound data for 74 pure organic compounds. Information about the name of the compounds and the original reference for each data point are presented as supplementary file in the supplementary DVD [390].

Implementing the K-means clustering technique mentioned earlier, the data set was divided into three sub-data sets including the training, the validation and the test sets, respectively. 10 % of the data set including 166 data points was assigned to the test set. Similar to the test set, 10 % of the data set including 166 data points was allocated to the validation set. The remaining data points (1335 data points) were employed as the training set for model development purpose.

The database are presented as supplementary file in the supplementary DVD.

4.9 Standard Molar Chemical Exergy of Organic Compounds

The dataset used in this study is that of Kotas [197]. The dataset comprises standard molar chemical exergy values for 133 pure organic compounds. The data values range from 303580

$\text{kJ}\cdot\text{mole}^{-1}$ (for formic acid) to $13254570 \text{ kJ}\cdot\text{mole}^{-1}$ (for tetraphenylmethane). The organic compounds within the database are composed of elemental carbon, hydrogen, nitrogen, oxygen, and sulfur. The minimum and maximum number of carbon, and hydrogen atoms of the compounds are for methane (1 carbon atom) and tetraphenylmethane (25 carbon atoms), formaldehyde (2 hydrogen atoms) and octadecane (38 hydrogen atoms), respectively. In the dataset, there are 22 nitrogen, 63 oxygen, and 18 sulfur containing compounds. The maximum occurrence of nitrogen, oxygen, and sulfur atoms is related to melamine (6 nitrogen atoms), β -maltose (11 oxygen atoms), and cyctine (2 sulfur atoms), respectively. Therefore, the dataset covers a fairly wide range of organic compounds.

The *K*-means clustering technique was used to split the data set to the training, validation, and test sets [391, 392]. In order to avoid these issues, approximately 80% of the data was assigned to the training set (107 data points) and the remaining was divided into two equal sub-datasets and were allocated to the validation and test sets, respectively (13 data points each).

The database are presented as supplementary file in the supplementary DVD.

4.10 Refractive Indices of Organic Compounds

The superiority of a proposed model is dependent on the extensiveness of the dataset used for both the development and testing of the model. In other words, the applicability, reliability and accuracy of the model for estimation of physical properties depend on the comprehensiveness of the dataset employed in its development [393-399]. There are relatively few studies in literature which used of a very large dataset for model derivation with the aid of the group contribution method. This makes the dataset of 11918 diverse mostly organic compounds drawn from Yaws' Handbook of

Thermodynamic and Physical Properties of Chemical Compounds [400] the most comprehensive and extensive dataset for the develop of a GC model for refractive index of pure compounds.

A quick analysis of the compounds within the dataset shows that the refractive indices range between 1 and 1.872 and molecular weights between 16.042 and 891.497. It should be that there are a few cadmium, tin, antimony, lead, tellurium, and bismuth compounds in the database. Figure 4.1,

Figure 4.2, and Figure 4.3 indicate the distribution of refractive indices, molecular weights, and atom numbers, respectively.

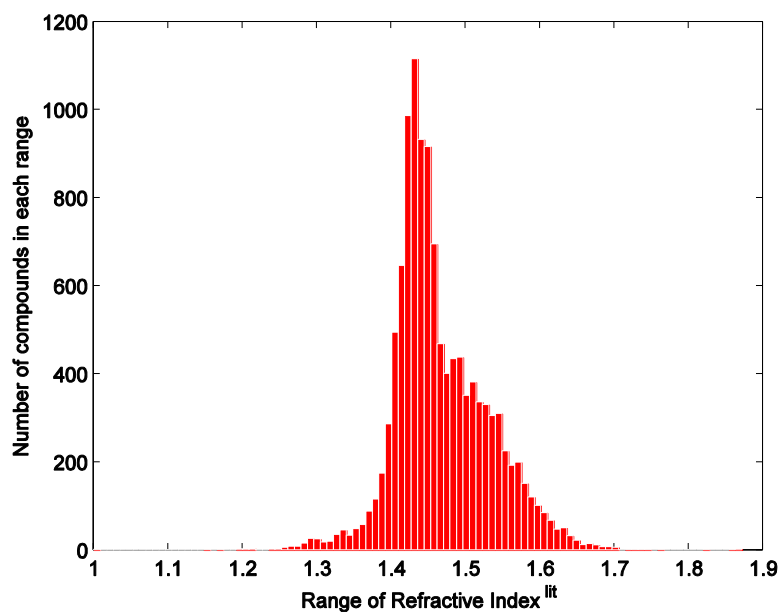


Figure 4.1 Distribution of refractive indices in the databank

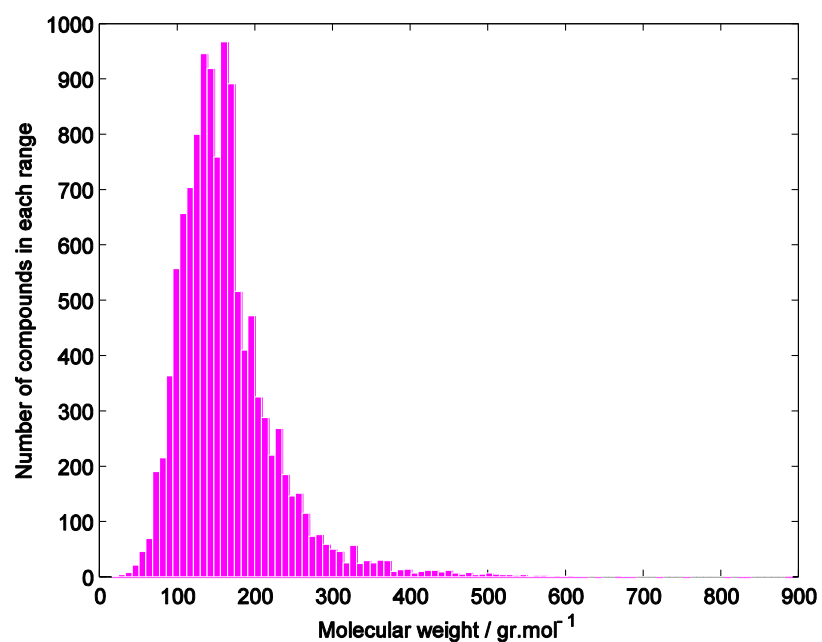


Figure 4.2 Distribution of molecular weights in the refractive index databank

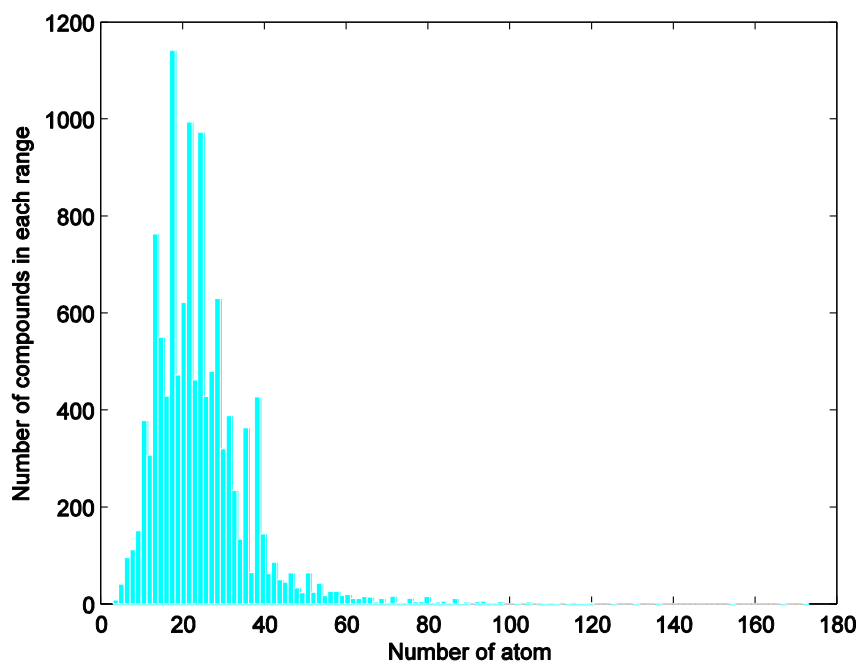


Figure 4.3 Distribution of atom numbers in the refractive index databank

Using the K-means clustering technique mentioned earlier, the data set was divided into three sub-data sets including the training, the validation and the test sets, respectively. 10 % of the data set including 1191 data points was assigned to the test set. Similar to the test set, 10 % of the data set including 1191 data points was allocated to the validation set. The remaining data points (9536 data points) were employed as the training set for model development purpose.

The database are presented as supplementary file in the supplementary DVD.

4.11 Freezing Point Temperature of Organic Compounds

The important step in developing comprehensive predictive models is to apply an informative, inclusive and representative dataset [398, 399, 401]. The essential criteria for a sufficiently predictive model are the availability of a set of data of adequate size, diversity and measured under the same (or similar) conditions with sufficiently reproducibility and accuracy [402]. Consequently, there are almost few previous studies (mostly reported by the author) in literature which report the handling of large datasets to derive a group contribution method. Hence, a dataset of freezing point values for more than 16,500 diverse mostly organic compounds extracted from Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds [400] was used in this study. A quick analysis of the compounds within the dataset indicates that the freezing points range between 54.26 and 914.05 K.

Using the K-means clustering technique mentioned earlier, the data set was divided into three sub-data sets including the training, the validation and the test sets, respectively. 10 % of the data set was assigned to the test set. Similar to the test set, 10 % of the data set including was allocated to

the validation set. The remaining data points were employed as the training set for model development purpose.

The database are presented as supplementary file in the supplementary DVD.

Chapter 5 Adaption & Modifications of the Algorithms

As mentioned in previous chapters, several thermo physical properties were investigated in this thesis. The mathematical algorithms used in this study were extensively discussed in chapter 3. In this chapter, the modifications/adaptions of the mathematical algorithms for each thermophysical properties are discussed in details.

5.1 Liquid Thermal Conductivity of Pure Chemical Compounds at Atmospheric Pressure

5.1.1 Developing a General Model for Liquid Thermal Conductivity

The gene expression programming as explained in section 3.2 has been pursued to obtain a reliable model. The corresponding states principle parameters, namely, critical temperature (T_c), critical pressure (P_c), critical volume (V_c), acentric factor (ω), along with normal boiling temperature (T_b), temperature (T), and the molecular weight (M_w) have been considered as input parameters of the model and then introduced to the gene expression programming for the sake of a non-linear model development. At the beginning, the liquid thermal conductivities of pure organic compounds was assumed to be correlated to the abovementioned properties as follows:

Error! Reference source not found.**Error! Reference source not found.**

$$k = f(T_c, P_c, V_c, \omega, T_b, T, M_w)$$

The following steps were pursued:

1. Setting an initial population *i.e.* In this step. The chromosomal structures were randomly generated using a set of mathematical operators such as (-,+,*,/,^) presented as pars trees and input parameters as terminals.

2. Computation of the fitness value for every individual of the generated population by the following objective function (OF):

Error! Reference source not found.

$$OF(i) = \frac{100}{N} \sum_i^N \frac{|k(i)^{pred} - k(i)^{exp}|}{k(i)^{exp}}$$

where N stands for the number of the liquid thermal conductivities studied, n represents the number of the model parameters, and superscripts *pred* and *exp* are the predicted and experimental liquid thermal conductivities. As mentioned earlier the DIPPR 801 [3] data set was used to provide the main data set of liquid thermal conductivities.

3. Selection of the best model trees to make proper parents for replacement. In this step, each model is assessed using the fitness function represented in above equation. In this study, the tournament technique has been employed to produce an sufficiently diverse population in each generation.

4. Employing the genetic algorithm operators such as replication, mutation, and inversion for gene reproduction with modification computational steps.

a. Replication operator: It copies exactly the chromosomes of the individuals chosen in the selection step (step 3).

b. Mutation operator: It brings about efficient adaption of individuals populations. The point mutation has been implemented, in which a random node (in the structures of the chromosomes) is selected and the stored information is replaced with a different random primitive of the same arity taken from the initial (old) set [1, 2]. Having defined the mutation rate (p_m), the mutation can happen everywhere in the structural organization of chromosomes; however, with preservation of

the original structure. The mutation can be normally performed through changing the heads of genes symbols and terminals of the tails.

c. Inversion operator: This operator is applied to create new individuals through modification of the heads of randomly selected genes. It has already been argued that all the new individuals created by inversion are considered as correct programs. The performance of this operator can be adapted choosing a value for the inversion rate (p_i).

5. Transposition and insertion sequence elements: The transposable elements of gene expression programming are a part of the genome that can be activated and jumped to another place in the chromosome, which include three types as implemented by Ferreira [1, 2]: “Short fragments with either a function or terminal in the first position transpose to the head of genes, short fragments with a function in the first position that transpose to the root of genes (root IS elements or RIS elements), and entire genes that transpose to the beginning of chromosomes.”

6. Recombination: This step, which is conducted in three manners including one-point recombination, two-point recombination, and gene recombination, randomly chooses two chromosomes to exchange specific material with each other, leading to appearance of two new chromosomes. Consequently, new generation is created. The preceding procedure is repeated until the defined stopping criteria (can be user-defined convergence criteria or maximum number of generations) is satisfied. The details of this procedure along with comprehensive examples are provided by Ferreira [1, 2].

5.1.2 Developing a Quantitative Structure-Property relationship

Initially, the chemical structures of all the compounds within the data set were prepared for geometry optimization. This step was done employing the Dreiding Force field in Chemaxon's

JChem [403]. In the next step, Dragon software [404] was used to calculate the molecular descriptors. Approximately 3000 molecular descriptors were calculated for each chemical structure. Then, those molecular descriptor which have not been calculated for all the chemical structures were eliminated. Also, a pair-correlation analysis was performed to remove highly correlated molecular descriptors. It was assumed that any pair of molecular descriptors whose squared correlation coefficient of their linear regression is higher than 0.9 behave the same and one should be eliminated to decrease the solution space. It should be noted that temperature was also considered as an experimental descriptor. As a result of the abovementioned steps, the molecular descriptors and the electrical conductivities are introduced to the sequential search algorithm (SS) [405] for the sake of choosing the proper subset of molecular descriptors.

5.1.3 Developing a Group Contribution Model for Liquid Thermal Conductivity

The chemical structures of all the pure compounds were analyzed with much attention and approximately 650 functional groups/segments/substructures were identified useful to describe the liquid thermal conductivities of pure mostly organic compounds. Subsequently, the frequency of each one was counted and used for next steps. The pair correlation analysis as explained in previous section was performed to eliminate the irrelevant functional groups/segments/substructures. Consequently, 321 functional groups/segments/substructures were remained and introduced to the sequential search algorithm (as explained in section 3.1.2) for the sake of calculation of their contribution on the liquid thermal conductivities. It is required to prevent entering irrelevant functional groups/segments/substructures in to the final group contribution model. Irrelevant functional groups/segments/substructures means those parameters which has no effect on the accuracy of the model.

5.2 Viscosity of Ionic Liquids

In order to develop a reliable group contribution model for the prediction of the viscosity of ionic liquids, the chemical structures of all counterparts of ionic liquids within the data set were carefully investigated to identify all the possible functional groups/segments/substructures. Finally, nearly 310 functional groups/segments/substructures were identified. Subsequently, the frequency of each one was counted and used for next steps. It should be noted that temperature was included in the molecular descriptors pool as an experimental descriptor. Performing the pair correlation as mentioned in previous section, the number of functional groups/segments/substructures was decreased to 200. Consequently, the sequential search (as explained in section 3.1.2) was used to determine the optimal subset of functional groups/segments/substructures and their corresponding contribution on viscosity of ionic liquids. Finally, 46 functional groups/segments/substructures (24 for anions and 22 for cations) in addition to temperature chosen by sequential search algorithm. In order to increase the quality of the model, the dependency of the viscosity of ionic liquids to temperature was further evaluated by trying several non-linear terms.

5.3 Thermal Decomposition Temperature of Ionic Liquids

5.3.1 Developing a Quantitative Structure- Property Relationship

In order to develop a QSPR for the thermal decomposition temperature of ionic liquids, the geometry optimization was performed on the chemical structures of all the ionic counterparts of all the ionic liquids within the data set. The optimization was performed using the Dreiding Force Field algorithm [406] in Chemaxon's JChem. In the next step, the optimized chemical structures were introduced to Dragon software [404] for molecular descriptors calculation. Consequently, approximately, 3000 molecular descriptors were calculated for each ionic counterparts. To develop the model the molecular descriptors of cations and anions of each ionic liquids were used. The

same steps explained in previous sections was done to remove molecular descriptors which have not been computed for all the ionic counterparts. Furthermore, the pair correlation analysis as explained in previous section was performed to omit the irrelevant molecular descriptors. Finally, the genetic function approximation technique as explained in section 3.1.1 was used to select the proper subset of molecular descriptors from the remaining set.

5.3.2 Developing the Group Contribution method

The thermal decomposition temperature of ionic liquids was also studied to develop a group contribution model. Consequently, the chemical structures of the ionic counterparts of all the ionic liquids were considered to determine the all the possible functional groups/segments/substructures. As a result, 425 unique functional groups/segments/substructures were identified. Subsequently, the frequency of each one was counted and used for next steps. The number of functional groups/segments/substructures was dropped to 261 using pair correlations analysis as explained in previous sections. Employing the sequential search (as explained in section 3.1.2), 30 functional groups/segments/substructures (10 for anions and 20 for cations) was selected to optimally describe the thermal decomposition temperature of ionic liquids.

5.4 Electrical Conductivity of Ionic Liquids

5.4.1 Developing a Non-Linear QSPR

The electrical conductivities of ionic liquids was investigated to develop a QSPR. Therefore, the ionic counterparts of all the ionic liquids were investigated for geometry optimization and then molecular descriptor calculation. It should be noted that the Dreiding Force field in Chemaxon's JChem was used for geometry optimization of the chemical structures [403]. Subsequently, the

optimized chemical structures were introduced to Sarchitect software (Sarchitect) [23] to compute the molecular descriptors. Approximately 1000 molecular descriptors were calculated for each ionic species. The number of molecular descriptors was dropped as a consequent of elimination of the incompletely calculated molecular descriptors and performing the pair correlations as explained in previous sections. It should be noted that temperature was also considered as an experimental molecular descriptor. Implementing the sequential search algorithm as explained (as explained in section 3.1.2) coupled with the least square support vector machines approach (as illustrated in section 3.1.4), a new non-linear LSSVM-QSPR was developed. It should be mentioned that the application sequential search algorithm which results in linear models is not appropriate for those properties which have non-linear dependency to either temperature or molecular descriptors.

5.4.2 Developing a Non-Linear GC Model

The ionic species (cations and anions) of all the ionic liquids within the data set were carefully analyzed to identify all the possible functional groups/segments/substructures. Consequently, 146 unique functional groups/segments/substructures were identified. Subsequently, the frequency of each one was counted and used for next stages. Implementing the pair correlation analysis the number of functional groups/segments/substructures was dropped to 78. Employing the sequential search (as explained in section 3.1.2) coupled with least square support vector machines (22 functional groups/segments/substructures (11 for anions and 11 for cations) were selected to optimally predict the electrical conductivity of ionic liquids.

5.5 Normal Boiling Temperature of Pure Organic Compounds

5.5.1 Developing a non-linear QSPR

The geometry optimization was performed on all of the compounds within the database using the Dreiding force fields in Chemaxon's JChem [403]. The optimized molecular structures were then introduced to Dragon software [407] for the molecular descriptor computation. Nearly 3000 molecular descriptors were computed for each compound. As result of elimination of the incompletely calculated molecular descriptors for all the compounds and performing the pair correlations as explained in previous sections, the number of molecular descriptors was dropped to 1143 molecular descriptors. Using the sequential search algorithm as explained (as explained in section 3.1.2) coupled with the three-layer feed forward neural networks (3FFNN) with Sigmoid (hyperbolic tangent) transfer function (as illustrated in section 3.1.3), a new non-linear 3FFNN-QSPR was developed.

Handling a very large database was one of the major achievement of this study. This challenge was overcome by employing a parallel computing technique. In order to decrease the computational round-off errors, the literature values of the NBPs are normalized to values between -1 and 1 using minimum and maximum values of input data. This procedure, done in the optimization section, is performed to obtain the neural network parameters (W_1, W_2, b_1, b_2) and has no effects on the results of the model.

In the next step the database is randomly divided into three subsets:

- i. A Training Test set to generate the structure of the ANN (80% of the studied compounds, 14216 compounds);
- ii. A Validation (Optimization) set for optimizing the model parameters (10% of studied compounds, 1776 compounds); and

iii. A Test (Prediction) set to assess the capability of the model to estimate the desired property for compounds not used in the model development (10% of studied compounds, 1776 compounds)

Finally, the optimal values of the weight matrices (W_1, W_2) as well as the bias vectors (b_1, b_2) are obtained by minimization of the objective function as explained in section 3.1.3.

5.6 Sublimation Enthalpy at the Standard Reference Temperature of 298 K

5.6.1 Developing a GC model

To develop a reliable correlation model, one must use parameters which enable one to distinguish each compound from the others. In other words, one needs a unique set of parameters for each compound that can adequately describe the sublimation enthalpy. Based on past experience [20, 21] it was decided to generate the parameters from the molecular structures. As a result, a collection of 294 chemical substructures were gathered which have previously been implemented by the authors to correlate other important physical properties [110, 408-410]. In the next step, the frequency of appearance of each of these 294 chemical substructures was counted in each compound. The pair correlation between each pair of the 294 chemical substructures was then evaluated to avoid entering irrelevant parameters into the final model. In the next step, if the pair correlation of a pair of chemical substructures was more than the threshold value of 0.95, one of them was eliminated and the other kept for the next step. Performing this procedure, the collection of the chemical substructures was reduced to 251 chemical substructures. In order to determine the final model and to choose the optimal subset of chemical substructures affecting the sublimation enthalpy, the SS algorithm explained in section 3.1.2 was applied [409].

5.7 Vaporization Enthalpy at the Standard Reference Temperature of 298 K

5.7.1 Developing a GC model

The physio-chemical properties of chemical compounds may be regarded as the intermolecular interactions originating from their chemical sub-structures. Therefore, it seems to be rational that one relates the properties to an efficient subset of these chemical sub-structures. The term “efficient” means a subset which enables one to distinguish each chemical compound from the others. In order to find an efficient subset of chemical sub-structures to predict the vaporization enthalpy, 1600 chemical sub-structures which have been previously used by the authors to correlate other important properties [110, 408-410] were generated. Next, the number of occurrences of each of these 1600 chemical sub-structures in chemical structures of the chemical compounds were collected in a table. Then, the pair correlation between each pair of these 1600 chemical substructures was considered as the major criteria to prevent entering irrelevant chemical sub-structures. In this step, if the pair correlation of a pair was more than the threshold value of 0.9, one of them was eliminated and the other kept for the subsequent computations. Following these preliminary calculations, the collection of chemical structures was reduced to 323 chemical sub-structures.

In order to find the optimal subset of chemical sub-structures affecting the vaporization enthalpy and also to develop the final model, the SS algorithm as explained in section 3.1.2 was implemented [409].

5.8 Speed of Sound in Saturated Liquids

The first step in developing a group contribution model is to determine which set of chemical substructures can be most efficient in describing the desired property under consideration.

In the next step, it is required to select the most efficient subset of substructures which can describe the desired property with highest accuracy. It should be noted that great attention should be paid when choosing the most efficient subset of chemical substructures to avoid selecting those substructures that have no or low contribution to the desired property. The latter makes the problem of developing a group contribution model difficult. Therefore, there is a requirement to develop new strategies. Another obstacle in developing a group contribution model is to determine how the desired property can be related to the most efficient subset of chemical substructures. In most cases the process of introducing the most efficient subset of chemical substructures and the developing the model are considered as two separate steps:

1. Introducing the most efficient chemical substructures according to some rule of thumb, some basic findings about how to decompose a chemical structure to its basic units, or their effectiveness on other physical properties.
2. Determining the contribution of each chemical substructure on the desired property using a multivariate linear regression method (In some cases, it is possible to develop some nonlinear correlations based on the linear contribution of each chemical substructures.)

In order to avoid choosing irrelevant chemical substructures, a subset variable selection strategy was employed. Also, nonlinear regression methods were implemented to develop the group contribution model for cases in which multivariate linear regression methods give poor results.

As a result, a combination of SS algorithm discussed in section 3.1.2 as a subset selection technique and least squares support vector machine (LSSVM) as a non-linear regression method is implemented in this study.

5.9 The GC model for Determination of the Standard Molar Chemical Exergy of Organic Compounds

The chemical structures for all 133 compounds within the dataset were carefully investigated to find the most effective chemical substructures that can be used to distinguish compounds from each other. To achieve this goal, a collection of the most important chemical substructures that had previously been implemented for the estimation of the other physical properties of pure organic compounds [110, 409, 410] was utilized. Thus, a collection of 187 chemical substructures was prepared. In the next step, the number of occurrences of each chemical substructure in each compound was counted. Finally, the results were imported into a table.

The final model was developed in two steps as follows;

- 1- Firstly, the pair correlation between each pair of 187 chemical substructures was evaluated to prevent entering irrelevant parameters into the final model. As a result, if the squared correlation coefficient (R^2) of a linear correlation between a pair of chemical substructures was more than the threshold value of 0.95, one of them was eliminated and the other one was kept for the next step. Performing this step, the collection of the chemical substructures was reduced to 140 chemical substructures. This step is used to pre-evaluate the collection of 187 chemical substructures. In this step, some of the irrelevant parameters are eliminated.
- 2- In the next step, the SS algorithm as explained in section 3.1.2 was used.

5.10 Refractive Indices of Organic Compounds

5.10.1 The QSPR Model

In developing a QSPR model, molecular descriptors are one of the most important ingredients. Furthermore, the molecular descriptors are the final result of a logical and mathematical procedure in technical terms, which transform chemical structure information, encoded within a symbolic representation of a molecule, into a useful number or the result of some standardized experiment [411]. The optimized molecular structures are a necessity to calculate molecular descriptors. The molecular structures are optimized with accurate Dreiding force fields as defined by Chemaxon's JChem[403]. In order to calculate the molecular descriptors, the optimized molecular structures must be loaded into Dragon software[21]. It is capable of calculating over 3000 descriptors from several diverse classes. These classes consist of Topological indices, Burden Eigen values, Constitutional descriptors, Connectivity indices, Information indices, 2D autocorrelations, Walk and path counts, Functional group counts, Atom-centered fragments, Molecular properties, Edge-adjacency indices, topological charge indices, Eigenvalue-based indices, geometrical descriptors, Randic molecular profiles, 3D-MORSE descriptors, RDF descriptors, WHIM descriptors, GETAWAY descriptors, charge descriptors, 2D binary fingerprint, and 2D frequency finger print. The descriptors obtained were analyzed carefully and those which were not able to be calculated for some compounds were neglected completely.

After calculation of the descriptors, the next step is to gather the subset from the descriptor pool which can correlate the refractive index well. One of the crucial issues in this study is the handling of a large number of compounds as their associated descriptors in model development. From experience gained in previous studies[21, 412], it can be concluded that the sequential search strategy [413] is the correct choice for the subset variable selection/reduction, in terms of its

capability of handling the large number of data, as well as acceptable computational run-times. However, the algorithm uses a simple scheme to select the best features and develop the final model. As a result, a two-step strategy was used in this section as follows;

1-Reduction of the number of molecular descriptors to several tens so that the application of a more sophisticated method such as Genetic Function Approximation (GFA) is feasible.

2- Selection of the best subset of molecular descriptors from the output of the previous step to develop the final model.

In the first step, a SS mathematical strategy explained in section 3.1.2 is applied to reduce the number of molecular descriptors to several tens of descriptors.

In the next step, the Genetic Function Approximation (GFA) as explained in section 3.1.1 is applied for selection of the most efficient subset of variables from the small subset of variables selected by the sequential search algorithm in the previous step.

5.10.2 The GC model

The chemical structures of all the components used in the database were examined thoroughly to find out the most efficient sub-structures for the estimation of the refractive index. In other words, having defined the compounds present in the database, the chemical structures of all of the studied compounds have been analyzed to recognize the functional groups. These functional groups are generally selected from a series of groups containing approximately 500 different chemical groups as follows [21, 92, 395, 397, 408, 410, 414-421]:

a) Functional groups are partitioned in different categories, each one including two pairs from all of the groups.

b) A mathematical strategy (algorithm) is applied to establish a linear relationship between the two groups in a pair:

5.1

$$GC_i = a \times GC_j + b$$

where GC denotes the functional groups, a and b are the parameters of the linear regression, and subscripts i and j refer to i th and j th functional groups.

c) In the case where the squared correlation coefficient of equation 5.1 is greater than a selected value (0.9 in this study), one of the groups is omitted from the investigated pair because it has no significant effects on the model finally developed and results in an increase in the model parameters (final functional groups).

The preceding procedure is pursued until the most efficient contributions for evaluation of the corresponding property (refractive index) have been determined. In order to propose an accurate and reliable correlation model, one must utilize parameters which enable one to distinguish one compound from the other. In other words, one needs a unique set of parameters for each compound that can sufficiently describe the refractive index. In the present study, it was decided to generate the parameters from the molecular structures. As a result, a collection of nearly 500 chemical substructures were collected. In the next step, the frequency of appearance of each of the chemical substructures was counted in each compound. The pair correlation between each pair of the chemical substructures was then evaluated to avoid entering irrelevant parameters into the final model. Then, if the pair correlation of a pair of chemical substructures was more than the threshold value of 0.9, one of them was eliminated and the other kept for the next step. Performing this

procedure, the collection of the chemical substructures was decreased to nearly 200 chemical substructures.

In order to develop the final model and select the optimal subset of chemical substructures affecting the refractive index, the SS algorithm explained in section 3.1.2 was applied [417].

5.11 Freezing Point Temperature of Organic Compounds

5.11.1 The QSPR Model

The optimized molecular structures are a necessity to calculate molecular descriptors. The molecular structures are optimized with accurate Dreiding force fields as defined by Chemaxon's JChem [403]. In order to calculate the molecular descriptors, the optimized molecular structures must be loaded into Dragon software [26]. It is capable of calculating over 3000 descriptors from several diverse classes. These classes consist of Topological indices, Burden Eigen values, Constitutional descriptors, Connectivity indices, Information indices, 2D autocorrelations, Walk and path counts, Functional group counts, Atom-centered fragments, Molecular properties, Edge-adjacency indices, topological charge indices, Eigenvalue-based indices, geometrical descriptors, Randic molecular profiles, 3D-MORSE descriptors, RDF descriptors, WHIM descriptors, GETAWAY descriptors, charge descriptors, 2D binary fingerprint, and 2D frequency finger print. The descriptors obtained were carefully analyzed and those which were not able to be calculated for some compounds were completely neglected.

The *K*-means clustering technique is employed to partition the main dataset into the training, the validation, and the test sets. Consequently, 80% (13,313 points), 10% (1,664 points), and 10%

(1,664 points) of the main dataset were assigned to each one of the training, the validation and the test sets. It should be noted that the “Training” set is applied to generate the model structure and the “Validating” set as well as the “Test (prediction)” set are employed to investigate its validity and prediction capability.

One of the most important aspects of this study is the handling of a large number of compounds as their associated descriptors in model development. For this purpose, powerful mathematical tools are required. From experience gained from previous studies [412, 422], it can be concluded that the sequential search mathematical strategy [413] is appropriate for the subset variable selection in terms of its capability of handling the large number of data, as well as acceptable computational run-times. Therefore, a sequential search mathematical strategy is applied to reduce the number of molecular descriptors to several tens of descriptors. The sequential search strategy implements an extensive search throughout the feature subsets [423]. The disadvantage of sequential forward selection (SS) is that if several features collectively are good predictors, but alone each is a poor prediction, none of the features may be chosen [423]. However, it should be mentioned that the SS with percentage of average absolute relative deviation as an objective function is successfully used for selection of variables.

5.11.2 The GC Model

For the prediction of pure component properties, group-contribution models such as those developed by Lyman et al. [424], Lydersen [425], Joback and Reid [426], Horvath [427], Ambrose [428], and Klinecicz and Reid [429] have been widely utilized. In these models, the property of a compound is a function of structurally-dependent parameters, which are determined by summing the frequency of each group occurring in the molecule and multiplying by its contribution. These techniques provide the advantage of quick prediction without requiring substantial computational

resources [430]. In proposing an efficient and reliable group-contribution model for the prediction of freezing point, the chemical structures of all the compounds were tested thoroughly to find out the most efficient sub-structures. Hence, having defined the compounds present in our database, the chemical structures of all of the studied compounds were analyzed to recognize the chemical substructures. These functional groups are normally selected from a series involving approximately 500 varying chemical groups [417].

In the next step, the frequency of appearance of each of the chemical substructures was counted in each compound. The pair correlation between each pair of the chemical substructures was then evaluated to avoid entering irrelevant parameters into the final model. Next, if the pair correlation of a pair of chemical substructures was more than the threshold value of 0.9, one of them was removed while the other was kept for the next step. Conducting these steps, the collection of the chemical substructures was decreased to nearly 300 chemical substructures.

In order to select the optimal subset of chemical substructures which affect the freezing point and finally proposing the final group-contribution model, the SS algorithm was implemented [417]. The method is suitable for the subset variable selection in terms of its capability of handling the large number of data, as well as for an acceptable computational run-time.

The “Training” set is applied to generate the model structure, while the “Validation” set as well as the “Test” set are employed to investigate its prediction validity and capability. In other words, the first set is for developing the model, the second set is for evaluation of the internal validity of the group-contribution model, and the final set is for assessing the predictive capability. In splitting the dataset into sub-data sets, the *K*-means clustering technique is implemented to partition the main dataset into the training, the validation, and the test sets. As a consequence, 80% (13,533

points), 10% (1,694 points), and 10% (1,694 points) of the main databank were assigned to each of the training, the validation, and the test sets, respectively.

Chapter 6 Results

6.1 Liquid Thermal Conductivity of Pure Chemical Compounds at Atmospheric Pressure

6.1.1 The General Model*

The Gene expression programming [1, 2] technique allowed determining the convenient parameters for the most accurate model from the following parameters: $T_c, P_c, V_c, \omega, T_b, T, M_w$.

The final correlation can be reported as follows:

$$k = 1 \times 10^{-4} \left(10\omega + 2P_c - 2T + 4 + 1.908 \left(T_b + \frac{1.009B^2}{M_w^2} \right) + \frac{3.9287M_w^4}{B^4} + \frac{A}{B^8} \right) \quad 6.1$$

where

$$A = 3.8588M_w^8(1.0045B + 6.5152M_w - 8.9756) \quad 6.2$$

and

$$B = 16.0407M_w + 2T_b - 27.9074 \quad 6.3$$

In 6.1 **Error! Reference source not found.**, 6.2, and 6.3 units for k , P_c , and M_w are $\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$, bar, and $\text{g}\cdot\text{mol}^{-1}$, respectively. T and T_c are in K.

The statistical parameters of the obtained results presented in Another important drawback of the model proposed by Latini *et al.* and Baroncini *et al.* [23-29] is that it cannot be used for any pure compounds. The parameters of the model are presented just for several chemical families of compounds comprised of saturated hydrocarbons, olefins, cycloparafins, aromatics, alcohols, organic acids, ketones, esters, and refrigerants. These chemical families do not cover a significant number of widely used chemical families such as amines, silanes/siloxanes, inorganic compounds,

* The model has been published in *AIChE Journal*. 59, 1702-1708, 2013.

sulfides/thiophenes, mercaptanes, epoxides, peroxides, nitriles, elements, and aldehydes. This point reveals that the model is not a general model.

Table 6.1 indicate that the average absolute relative deviation of the determined liquid thermal conductivity values from the experimental ones [3] is about 8.7 %.

The detailed results using the developed model for all of the studied compounds along with the values of the input parameters of the model are presented as supplementary file in the supplementary DVD.

Furthermore, the capability of the proposed method for determination of the liquid thermal conductivity values for different chemicals/chemical families are presented in the published article [252].

Reasonable deviations from the DIPPR 801 [3] dataset values can be observed. This issue asserts the acceptable applicability of the proposed model for many of the chemical families. This point is confirmed when taking a look at the uncertainty of the DIPPR 801 data [3]. According to the results, it can be concluded that the deviation of the model results are mostly within or very close to the average uncertainty of DIPPR 801 data [3]. This latter demonstrates the capability of the model.

It is worth it to know that the computational steps in each generation of the GEP [1, 2] strategy needs parallel computing and consequently high amounts of CPU time. For development of the method, we have defined a convergence criterion for the algorithm, which is the difference between the accuracy of the obtained results from the current generation and the previous one in each step. Consequently, it is possible to develop a more accurate model through continuation of the calculation steps producing more generations from the subsequent populations. However, careful investigation using more powerful computers should be made to verify this concept.

In comparison with the models proposed by Latini *et al.* and Baroncini *et al.* [23-29] and Sastri *et al.* [30], the presented model here is simpler and has smaller number of parameters. The previous models have a major drawback. Since, both the models have several chemical family- dependent parameters; many compounds may be fitted into more than one chemical family of compounds.

As a result some difficulties may happen when prediction of the liquid thermal conductivity of multifunctional compounds.

Another important drawback of the model proposed by Latini *et al.* and Baroncini *et al.* [23-29] is that it cannot be used for any pure compounds. The parameters of the model are presented just for several chemical families of compounds comprised of saturated hydrocarbons, olefins, cycloparafins, aromatics, alcohols, organic acids, ketones, esters, and refrigerants. These chemical families do not cover a significant number of widely used chemical families such as amines, silanes/siloxanes, inorganic compounds, sulfides/thiophenes, mercaptanes, epoxides, peroxides, nitriles, elements, and aldehydes. This point reveals that the model is not a general model.

Table 6.1 The statistical parameters of the general model proposed for the liquid thermal conductivity of pure organic compounds.

Statistical parameters	
<i>training set</i>	
R ²	0.8835
Average absolute relative deviation	8.96
Standard deviation error	0.0146
Root mean square error	0.0148
N	15221
<i>optimization set</i>	
R ²	0.8715
Average absolute relative deviation	8.95
Standard deviation error	0.0150
Root mean square error	0.0152
N	1902
<i>prediction set</i>	
R ²	0.9105
Average absolute relative deviation	8.96
Standard deviation error	0.0144
Root mean square error	0.0147
N	1902
<i>total set</i>	
R ²	0.8856
Average absolute relative deviation	8.68

Standard deviation error	0.0146
Root mean square error	0.0149
N	19025

Another important drawback of the model proposed by Latini *et al.* and Baroncini *et al.* [23-29] is that it cannot be used for any pure compounds. The parameters of the model are presented just for several chemical families of compounds comprised of saturated hydrocarbons, olefins, cycloparafins, aromatics, alcohols, organic acids, ketones, esters, and refrigerants. These chemical families do not cover a significant number of widely used chemical families such as amines, silanes/siloxanes, inorganic compounds, sulfides/thiophenes, mercaptanes, epoxides, peroxides, nitriles, elements, and aldehydes. This point reveals that the model is not a general model.

Furthermore, the major advantage of the presented model over previous models is that it has been evaluated using nearly 19000 experimental data for nearly 1600 compounds.

Therefore, we can conclude that the developed model is a valid one from a statistical point of view and we may not be able to doubt about the applied data using the characteristics of the model.

6.1.2 The QSPR Model*

According to method explained in section 5.1.2, the optimum model was obtained as follows:

6.4

$$k = 0.228925489(\pm 0.000706464) + \sum_{i=1}^{20} k_i C_i$$

As can be found from eq. 6.4, the optimum number of descriptors is 20. As mentioned above, the optimum number of molecular descriptors should be obtained by tracking the accuracy of the model.

* The model has been published in *Fluid Phase Equilibria*. 355, 52-80, 2013..

The optimum subset of molecular descriptors selected by SS algorithm along with their definition is presented in Table 6.2. Furthermore, the coefficient of each molecular descriptor in eq. 6.4 is presented in Table 6.3.

The average absolute relative deviation (AARD %), standard deviation error, and root mean square error of the model from experimental data are respectively 7.4%, 0.01, 0.01. It should be noted that the model shows the same parameters for the training set, validation set and test set.

The model shows reasonable deviation from DIPPR 801 data [4]. Furthermore, it can be concluded that the deviation of the model results are mostly within or very close to the average uncertainty of DIPPR 801 [4] data. This latter demonstrates the capability of the model.

Table 6.2 The molecular descriptors selected by sequential search algorithm to describe the thermal conductivity of pure organic compounds.

ID	Descriptor	Definition
k ₁	T	Temperature
k ₂	SIC0	structural information content (neighborhood symmetry of 0-order)
k ₃	ATS2m	Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic masses
k ₄	MATS1v	Moran autocorrelation - lag 1 / weighted by atomic van der Waals volumes
k ₅	BEHm1	highest eigenvalue n. 1 of Burden matrix / weighted by atomic masses
k ₆	JGI1	mean topological charge index of order1
k ₇	DP17	molecular profile no. 17
k ₈	RDF130e	Radial Distribution Function - 13.0 / weighted by atomic Sanderson electronegativities
k ₉	Mor30e	3D-MoRSE - signal 30 / weighted by atomic Sanderson electronegativities
k ₁₀	E1p	1st component accessibility directional WHIM index / weighted by atomic polarizabilities
k ₁₁	R5p	R autocorrelation of lag 5 / weighted by atomic polarizabilities
k ₁₂	nRCN	number of nitriles (aliphatic)
k ₁₃	nHDon	Sum of the hydrogens linked to all of the Os and Ns in the molecule
k ₁₄	N-067	number of aliphatic Al ₂ -NH
k ₁₅	TPSA(Tot)	topological polar surface area using N, O, S, P polar contributions

k ₁₆	BLTF96	Verhaar model of Fish base-line toxicity for Fish (96h) from MLOGP (mmol/l)
k ₁₇	B02[F-F]	presence/absence of F-F at topological distance 2
k ₁₈	F03[C-C]	frequency of C-C at topological distance 3
k ₁₉	F05[C-N]	frequency of C-N at topological distance 5
k ₂₀	F07[C-O]	frequency of C-O at topological distance 7

Another point that should be considered is that the proposed model uses just chemical structure-based parameters and doesn't require any other physical properties like normal boiling point and critical properties as required in the models proposed by Latini *et al.*, Baroncini *et al.* [23, 24, 26-29, 431, 432] and Sastri *et al.* [30].

Therefore, this model is superior to the previous models proposed by Latini *et al.*, Baroncini *et al.* [23, 24, 26-29, 431, 432] and Sastri *et al.* [30].

Table 6.3 The coefficients of the eq.6.4

Coefficient ID	Coefficient	Coefficient Error
C ₁	-0.000200805	1.02874E-06
C ₂	0.025053989	0.001129233
C ₃	-0.004309863	0.000226684
C ₄	-0.00388204	0.000231799
C ₅	-0.002597065	0.000144797
C ₆	-0.029900996	0.000762129
C ₇	0.000968139	4.73535E-05
C ₈	0.000157524	1.52522E-05
C ₉	-0.009813803	0.000330394
C ₁₀	-0.022988099	0.000766659
C ₁₁	-0.016816946	0.001009798
C ₁₂	0.019440866	0.000471029
C ₁₃	0.010927854	0.000139057
C ₁₄	-0.019446752	0.000536146
C ₁₅	0.000259336	6.47995E-06
C ₁₆	0.003567252	0.000153197
C ₁₇	-0.0310118	0.000549327

C ₁₈	0.001290887	2.7123E-05
C ₁₉	-0.005157446	0.000226558
C ₂₀	-0.002039102	9.16509E-05

As a matter of fact, the previous models have a major drawback. Since, both the models have several chemical family- dependent parameters; many compounds may be categorized into more than one chemical family of compounds. As a result some difficulties may take place when prediction of the liquid thermal conductivity of multifunctional compounds.

Another important drawback of the model proposed by Latini *et al.* and Baroncini *et al.* [23, 24, 26-29, 431, 432] is that it cannot be used for any pure compounds. The parameters of the model are presented just for several chemical families of compounds comprised of saturated hydrocarbons, olefins, cycloparafins, aromatics, alcohols, organic acids, ketones, esters, and refrigerants. These chemical families do not cover a significant number of widely used chemical families such as amines, silanes/siloxanes, inorganic compounds, sulfides/thiophenes, mercaptanes, epoxides, peroxides, nitriles, elements, and aldehydes. This point reveals that the model is not a general model.

Moreover, the main advantage of the proposed QSPR model over previous models is that it has been developed and validated using nearly 19000 evaluated thermal conductivity data for 1635 chemical compounds provided from DIPPR 801 data [4]. As a result, the developed model is a valid one from a statistical point of view and we may not be able to doubt about the applied data using the characteristics of the model.

6.1.3 The GC Model*

The proposed GC model for representation/prediction of liquid thermal conductivity is as follows:

6.5

$$k = \sum_{i=1}^{49} N_i k_i + a + bT$$

where N_i , k_i , T , a , and b are the number of occurrences of i th chemical substructure, the contribution of the i th chemical substructure, temperature, the intercept of equation 6.5, and the temperature contribution, respectively. The parameters of the model are presented in the published article [433].

The DIPPR 801 [4] data can be described successfully using the group contribution method. The average absolute relative deviation (AARD %), standard deviation error, and root mean square error of the model from experimental data are 7.2%, 7.2%, 7.1%, 0.01, 0.01, 0.01, 0.01, 0.01, and 0.01 for the training set, validation set and test set respectively.

Reasonable deviations from the DIPPR 801 [4] data can be observed. This observation supports the acceptable applicability of the presented model for many of the chemical families. This point is confirmed when taking into consideration the uncertainty of the DIPPR 801 [4] data. According to the results, it can be concluded that the deviation of the model results are mostly within or very close to the average uncertainty of DIPPR 801 [4] data. The latter demonstrates the capability of the model.

There are some points for which the model does not produce very accurate thermal conductivity values (generally in the higher than 13 AARD % range) and a possible explanation is given below. There are three chemical families for which the model results show AARD% more than 13%; INORGANIC GASES (24.8%), OTHER INORGANICS (21.8%) and POLYFUNCTIONAL ACIDS (13.4%).

According to earlier discussions, experimental measurement of accurate liquid thermal conductivity data is very difficult. Thus, the first possibility which may prevent the model in

* The model has been published in *Journal of Molecular Liquids*, 190, 223–230, 2013.

providing good representation/prediction for the data points may be due to the high experimental uncertainties. A quick consideration of the supplementary table demonstrates that the uncertainties of the data generally lie between 1 to 50 % (available in the supplementary DVD).

In comparison with the models proposed by Latini *et al.* and Baroncini *et al.* [23-29] and Sastri *et al.* [30], the model proposed herein is simpler and has a smaller number of parameters. The previous models have a major drawback. Since, both the models have several chemical family-dependent parameters; many compounds may be fitted into more than one chemical family of compounds. As a result some problems may occur when predicting the liquid thermal conductivity of multifunctional compounds.

Furthermore, the model proposed by Latini *et al.* and Baroncini *et al.* [23-29] cannot predict the liquid thermal conductivity of any random choice of pure compounds due to the fact that the parameters of the model do not exist for several important chemical families. Therefore, the model cannot be considered as a general model.

Moreover, the main advantage of the proposed GC model over previous models is that it has been developed and validated using nearly 19000 experimental data for nearly 1600 compounds. As a result, the developed model is from a statistical point of view more robust and has greater validity. There is probably greater confidence, as a result, in the application of the model due to this characteristic.

6.2 Viscosity of Ionic Liquids^{*}

The group contribution technique has been pursued to develop an accurate model for estimation of the viscosity of ionic liquids.

According to the aforementioned procedure, the contributions of each of 46 parameters were determined. The proposed model for estimation of viscosity of ILs is found as follows:

6.6

$$\log(\eta) = \sum_{i=1}^{N_a} N_{ai} \eta_{ai} + \sum_{i=1}^{N_c} N_{ci} \eta_{ci} + AT + \frac{B}{T} + CT^2 + \frac{D}{T^2} + \eta_0$$

^{*} The model has been published in *Chemical Engineering Science* 80 (2012) 326–333.

where N_{ai} , N_{ci} , η_{ai} , η_{ci} , and η_0 are, respectively, the number of occurrence of i th sub-structure of anions and cations, the contribution of the i th sub-structure of anions and cations, and the intercept of eq.6.6. The obtained group contributions, namely η_{ai} , η_{ci} , η_0 , as well as the temperature coefficients A, B, C, and D are presented in the published article [228].

The predicted viscosities and their absolute relative deviations from experimental values and the statistical parameters of the model are presented as supplementary file in the supplementary DVD [228]. The results demonstrate that this GC model can reliably estimate the viscosity of ILs. The model results, respectively, shows an average absolute relative deviation (%AARD), standard deviation, and root mean square error of 6.4 %, 0.21, and 0.21 from the 1672 experimental data.

The AARD% of the model for each class of ILs used in this study are presented in Table 6.5.

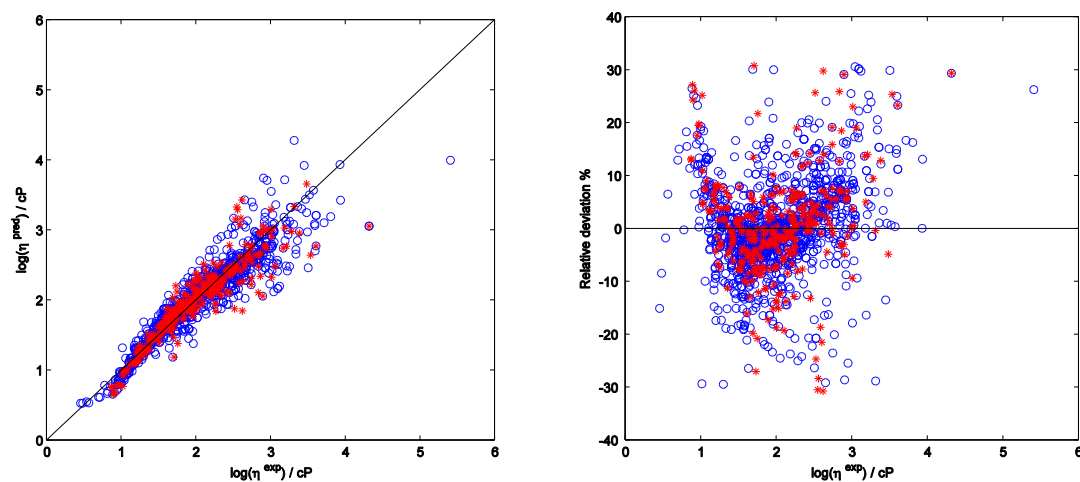


Figure 6.1 (Left) Predicted viscosities of ionic liquids by eq.1 versus the corresponding experimental values. (Right) Relative deviation of predicted viscosities of ionic liquids versus the corresponding experimental ones. (*) and (o) denote the training set and the test set.

The predicted viscosities in comparison with the experimental values are presented in Figure 6.1 (Left). In addition, the deviation of the model in comparison with the experimental data is presented as Figure 6.1 (Right).

These comparisons clearly show that this model can be used as a predictive tool for the prediction of the viscosity of various ILs. A comparison between the model predictions and their corresponding experimental values is depicted in Figure 6.2 (Left).

The models results for [C2mim][DCA], [C10mim][NTf2], [C1mim][NTf2], and [C2mim][C2SO4] as a function of temperature are presented in Figure 6.2 (Right). As depicted, the model can accurately predict the viscosity of ILs as a function of temperature.

One of the major causes of error in experimental data that may bring about large deviations when developing models is the presence of contaminations [45] in the ILs or ILs which are not of extremely high purity being used.

Table 6.4 The statistical parameters of the model proposed for viscosity of ionic liquids

Statistical Parameter	
training set	
R ²	0.880
Absolute average relative deviation	6.21
Standard deviation error	0.20
Root mean square error	0.20
N	1336
test set	
R ²	0.854
Absolute average relative deviation	6.79
Standard deviation error	0.22
Root mean square error	0.22
N	336
total	
R ²	0.874
Absolute average relative deviation	6.32
Standard deviation error	0.21
Root mean square error	0.21
N	1672

They can significantly affect various properties of ILs particularly their viscosities. The contaminants are mostly composed of water, halides, and other organic solvents. For instance existence of traces of Cl⁻ ion in [C₄mim][BF₄] will result in a significant deviation in its viscosity. Moreover, with increasing the concentration of such ions, the viscosity dramatically increases. For

instance when the concentration of Cl^- increases from 0.01 to 0.5 m, the viscosity of $[\text{C}_4\text{mim}][\text{BF}_4]$ increases from 154 to 201 cP [434].

In comparison with the previous models that have been proposed, this model is more applicable across a wider domain. It has been developed and evaluated using 443 ILs composed of 76 and 191 unique anions and cations. Additionally, the temperature range of experimental data in this study is 50-433 K which is wider than that of previous models.

Table 6.5. The average absolute relative deviation of the proposed model for viscosity of ionic liquids from the corresponding experimental based on different classes of ionic liquids studied (In the table the units of T and η are respectively in K and) cP.

No	Class	AARD%	T range		$\log(\eta^{\text{exp}})$ range			$\log(\eta^{\text{pred}})$ range			N	
1	1,3-Dialkyl imidazolium	5.86	50.00	-	433.15	0.53	-	3.99	0.46	-	5.41	1381
2	1-Alkyl imidazolium	10.06	293.15	-	298.15	1.56	-	2.91	1.56	-	2.74	26
3	Amino acids	9.87	353.15	-	353.15	1.95	-	3.71	2.01	-	3.19	2
4	Ammonium	8.01	289.15	-	343.15	0.55	-	3.92	0.54	-	3.45	86
5	Double imidazolium	3.58	298.15	-	358.15	2.09	-	3.22	2.06	-	3.05	19
6	Guanidinium	5.81	293.15	-	333.15	2.05	-	3.36	2.09	-	3.07	9
7	Isoquinolinium	6.69	329.15	-	329.15	2.56	-	2.56	2.75	-	2.75	1
8	Morpholinium	10.85	298.15	-	303.15	2.28	-	2.58	2.35	-	3.01	9
9	Oxazolidinium	8.47	298.15	-	303.15	2.15	-	2.63	1.95	-	2.86	12
10	Phosphonium	16.00	293.15	-	303.15	1.53	-	4.28	1.54	-	3.32	19
11	Piperidinium	8.56	298.15	-	298.15	1.92	-	2.41	1.83	-	3.09	14
12	Pyridinium	3.86	293.15	-	353.15	1.11	-	2.30	1.13	-	2.38	7
13	Pyrrolidinium	7.51	293.15	-	301.15	1.55	-	2.41	1.57	-	3.23	21
14	Pyrroline	3.62	298.15	-	298.15	1.62	-	1.82	1.76	-	1.76	3
15	Sulfonium	9.57	253.15	-	353.15	0.66	-	2.68	0.81	-	2.71	50
16	Tetra-alkyl imidazolium	0.00	333.15	-	333.15	2.99	-	2.99	2.99	-	2.99	1
17	Tri-alkyl imidazolium	7.45	293.15	-	373.15	1.47	-	2.23	1.53	-	2.99	12

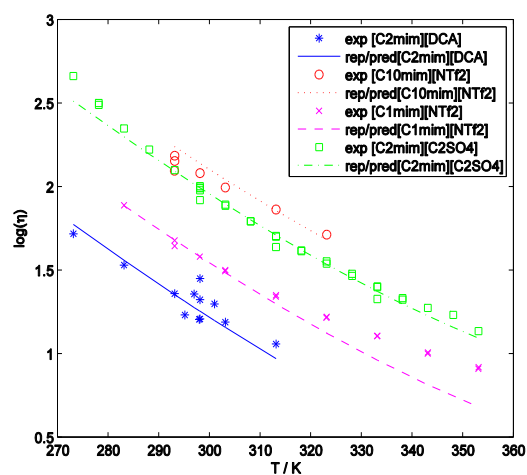
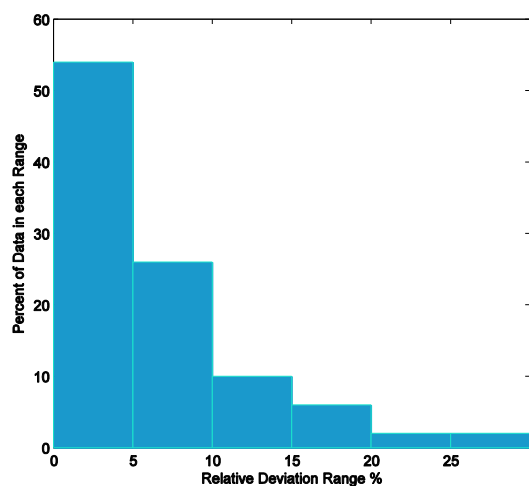


Figure 6.2 (Left) Comparison of the predicted viscosities of ionic liquids and their corresponding experimental values: approximately 54 % of the viscosities are predicted within 0-5 %, 26 % within 5-10 %, 10 % within 10-15 %, 6 % within 15-20 %, 2 % within 20-25 %, and the remaining 2 % within 25-31 %. (Right) Viscosity as function of temperature (The unit for η is cP).

The model is even superior to the previous QSPR model proposed by one of the authors [45] due to its comprehensiveness (76 and 191 unique anions and cations vs. 36 and 146 unique anions and cations in the previous study) and accuracy (6.4 % vs. 9 % AARD in the previous study).

6.3 Thermal Decomposition Temperature of Ionic Liquids

6.3.1 The QSPR Model*

The final QSPR model obtained contains 12 descriptors. In order to investigate the contribution of anions and cations on T_d of ILs, their corresponding descriptors in the model are presented as T_d^{Anion} and T_d^{Cation} , respectively:

6.7

$$T_d = 447.17 + T_d^{Anion} + T_d^{Cation}$$

6.8

$$T_d^{Anion} = -44.75 nCL_{Anion} + 35.97 GATS1m_{Anion} + 236.01 E3v_{Anion} + 209.82 R4e_{Anion} - 21.08 nCp_{Anion} - 40.88 B01[C - O]_{Anion}$$

6.9

$$T_d^{Cation} = -186.39 R1m_{Cation} + 11.58 nImidazoles_{Cation} - 71.17 S - 107_{Cation} - 35.41 B03[N - N]_{Cation} - 24.90 F03[C - O]_{Cation} + 48.62 F04[N - O]_{Cation}$$

In eq. 6.7, 6.8, and 6.9, **Error! Reference source not found.** the subscripts “Anion” and “Cation” denote the descriptors of anions and cations respectively.

Following Todeschini and Consonni [435], we may say:

* The model has been published in *Chemical Engineering Science*, 84, 557-563, 2012.

- “nCL” is the number of chlorine atoms. It seems that the presence of chlorine atoms in anions decrease the T_d of ILs [435].
- “GATS1m” is the Geary auto correlation - lag 1 / weighted by atomic masses. This descriptor shows how T_d is distributed along the anion structure. It has a positive effect on T_d [435].
- “E3v” is the 3rd component accessibility directional WHIM index / weighted by atomic van der Waals volumes. Directional WHIM descriptors are calculated as some univariate statistical indices on the projections of the atoms along each individual principal axis, while the global WHIMs are directly calculated as a combination of the former, thus simultaneously accounting for the variation of molecular properties along the three principal directions in the molecule. In this case, any information individually related to each principal axis disappears and the description is related only to a global view of the molecule. This descriptor has a positive effect on T_d [435].
- “R4e+” is the R maximal autocorrelation of lag 4 / weighted by atomic Sanderson electro negativities. It is a measure of polarity in the anion. When the polarity is boosted, T_d increases [435].
- “nCp” is the number of terminal primary C(sp³). It has negative effect on T_d [435].
- “B01[C-O]” is the presence or absence of C-O at topological distance 1. Its presence results in value “1” and its absence will bring about a value of “0”. It has a negative effect on T_d .
- “R1m+” is the R maximal autocorrelation of lag 1 / weighted by atomic masses. It has a negative effect on T_d [435].

- “nImidazoles” is the number of Imidazoles in the chemical structure of the cation. Presence of such substructure has a positive effect on T_d [435]. “S-107” is the number of R_2S or $RS-SR$ in which R represents any group linked through carbon. It appears that the presence of such substructures has a negative effect on T_d [435].
- “B03[N-N]” is the presence or absence of N-N at topological distance 3. Its presence results in a value of “1” and its absence will result in a value of “0”. It has a negative effect on T_d [435].
- “F03[C-O]” is the number of C-O at topological distance 3. It has a negative effect on T_d [435].
- “F04[N-O]” is the number of N-O at topological distance 4. It has a positive effect on T_d [435].

The average AARD% of the model results from experimental values for the training set and the test are close to 5.2 % and 5.1 %, respectively.

The AARD% of the model results from 586 experimental T_d data is approximately equal to 5.2%. The deviation of the model results from experimental data for different chemical classes of ILs are presented in Table 6.6. As can be seen, the highest AARD% values are related to Quinary alkyl imidazolium and Uronium ILs. It may be due to this fact that the model cannot predict their T_d well, because they are outnumbered in the database. This latter issue prevents the model to modify itself to predict their T_d well. A comparison between the presented model and those recently proposed demonstrates that the database used here (586 ILs) is much more comprehensive than those previously employed by Lazuss [47] (198 ILs) and Yan et al. [48] (158 ILs).

Table 6.6 The deviation of the predicted (the QSPR) thermal decomposition temperatures of ionic liquids from the corresponding experimental data based on different chemical families of ionic liquids.

No.	Compound	ARD%	AARD% range		$T_d^{\text{exp}} / \text{N.m}^{-1}$		$T_d^{\text{pred}} / \text{N.m}^{-1}$		N
1	1,3-Dialkyl imidazolium	5.4	0.3	- 15.9	423.15	- 725.15	469.84	- 667.22	83
2	1-Alkyl imidazolium	4.1	0.2	- 12.2	461.15	- 686.15	478.31	- 661.04	43
3	4,4-Dimethylimidazolium	3.2	3.2	- 3.2	488.15	- 488.15	472.63	- 472.63	1
4	Amino acids	6.2	0.5	- 14.4	392.15	- 514.15	375.74	- 550.17	29
5	Ammonium	5.3	0.0	- 15.3	379.15	- 693.15	389.78	- 665.88	121
6	Double imidazolium	2.9	0.1	- 9.6	577.95	- 715.15	609.38	- 712.3	26
7	Guanidinium	5.0	0.1	- 13.0	463.15	- 587.15	462.53	- 552.24	16
8	Morpholinium	5.5	0.9	- 13.8	397.15	- 685.15	366.7	- 653.53	34
9	Oxazolidinium	4.9	0.1	- 15.6	446.15	- 620.15	452.84	- 601.78	21
10	Phosphonium	5.0	0.7	- 12.8	457.15	- 693.15	456.45	- 663.73	23
11	Piperidinium	5.4	0.0	- 15.1	440.15	- 696.15	429.14	- 658.75	21
12	Pyridazinium	5.1	4.2	- 6.0	573.15	- 573.15	538.97	- 548.93	2
13	Pyridinium	3.0	0.4	- 6.5	500.35	- 677	488.73	- 652.84	6
14	Pyrrolidinium	6.3	1.1	- 13.8	423.15	- 690.15	431.73	- 658.83	25
15	Quinary alkyl imidazolium	8.5	8.0	- 8.9	674.15	- 739.15	620.17	- 673.46	2
16	Sulfonium	3.1	0.1	- 5.3	449.15	- 456.15	455.6	- 472.78	8
17	Tetra-alkyl imidazolium	5.7	0.5	- 14.3	439.15	- 580.15	502.08	- 571.96	6
18	Tetrazolium	6.8	3.5	- 10.3	443.15	- 588.15	458.84	- 549.3	3
19	Tri-alkyl imidazolium	5.6	0.7	- 13.8	528.75	- 730.19	495.84	- 667.97	27
20	Triazolium	5.4	0.1	- 15.3	379.15	- 698.15	363.67	- 685.39	88
21	Uronium	7.4	7.4	- 7.4	493.15	- 493.15	456.52	- 456.52	1

Regardless of that, the number of parameters used in the model (12 molecular descriptors) is much lower than those of Lazuss [47] (58 GCs) and Yan et al. [48] (25 molecular descriptors). Therefore, the model may be regarded as superior to the previous models.

6.3.2 The GC Model*

According to the procedure explained in section 5.3.2, the contribution of each of the 30 sub-structures was determined using the training set. The proposed model for estimation of the T_d of ILs is as follows:

6.10

$$T_d = \sum_{i=1}^{10} N_{ai} T_{dai} + \sum_{i=1}^{20} N_{ci} T_{dci} + T_{d0}$$

where N_{ai} , N_{ci} , T_{dai} , T_{dci} , and T_{d0} are, respectively, the number of occurrence of i th sub-structure of anions and cations, the contribution of the i th sub-structure of anions and cations, and the intercept of eq. 6.10. The computed group contributions, namely T_{dai} , T_{dci} , and T_{d0} , are presented in the published article [436].

The average AARD% of the model results from experimental values for the training set and the test are 4.5% and 4.3%, respectively. The model results show an AARD% of 4.4 % from experimental T_d values for the 613 ILs.

The deviation of the model results from experimental data for different chemical classes of ILs are presented in Table 6.7 .

* The model has been published in *Fluid Phase Equilibria*. 355, 81-86, 2013.

Table 6.7 The deviation of the predicted (the group contribution model) thermal decomposition temperatures of ionic liquids from the corresponding experimental data based on different chemical families of ionic liquids.

No.	Family	%AARD	AARD% range		T_d^{exp} range		T_d^{pred} range		N
1	1,3-Dialkyl imidazolium	4.8	0.0	- 14.8	436.15	- 725.15	389.35	- 735.58	96
2	1-Alkyl imidazolium	2.9	0.1	- 7.7	461.15	- 686.15	478.63	- 664.69	42
3	Amino acids	4.1	0.0	- 12.7	392.15	- 514.15	399.53	- 541.68	32
4	Ammonium	4.8	0.0	- 14.9	374.15	- 693.15	387.22	- 724.66	139
5	Double imidazolium	3.8	0.3	- 8.4	508.15	- 715.15	512.00	- 693.38	28
6	Guanidinium	5.1	1.0	- 10.7	463.15	- 587.15	469.41	- 560.07	16
7	Morpholinium	4.3	0.2	- 14.5	413.15	- 685.15	417.61	- 657.94	32
8	Oxazolidinium	4.0	0.3	- 11.3	446.15	- 620.15	451.91	- 641.29	18
9	Phosphonium	3.5	0.5	- 13.0	457.15	- 693.15	464.56	- 657.94	22
10	Piperidinium	4.5	0.0	- 13.9	473.15	- 696.15	469.41	- 657.94	18
11	Pyridazinium	1.2	1.2	- 1.2	573.15	- 573.15	579.86	- 579.86	2
12	Pyridinium	1.8	0.3	- 4.3	500.35	- 677.00	494.36	- 664.93	5
13	Pyrrolidinium	5.5	0.2	- 12.3	523.15	- 690.15	550.85	- 657.94	20
14	Quinary alkyl imidazolium	1.1	0.4	- 2.3	606.15	- 739.15	603.50	- 742.80	3
15	Sulfonium	1.4	0.8	- 2.3	449.15	- 456.15	445.72	- 445.72	8
16	Tetra-alkyl imidazolium	6.6	3.0	- 11.5	439.15	- 580.15	489.73	- 599.23	6
17	Tetrazolium	8.0	2.8	- 13.7	455.15	- 588.15	452.88	- 507.54	3
18	Tri-alkyl imidazolium	5.4	0.7	- 14.5	526.15	- 730.19	479.72	- 684.73	31
19	Triazolium	4.5	0.0	- 13.5	382.15	- 698.15	393.36	- 647.91	91
20	Uronium	1.8	1.8	- 1.8	493.15	- 493.15	484.43	- 484.43	1

A comparison between the proposed model and those recently presented in literature indicates that the database used in this study (613 ILs) is much more comprehensive than those previously implemented by Lazuss [47] (198 ILs) and Yan et al. [48] (158 ILs). Furthermore, the number of parameters used in the model (30 molecular descriptors) is much lower than those of Lazuss [47] (58 GCs) and very close to that of Yan et al. [48] (25 molecular descriptors). Of course, counting GCs is much simpler than computation of molecular descriptors that needs at least elementary knowledge of molecular simulation. Therefore, the model may be superior to the previous models. Thus, the model is predictive (within 4.4%) and may be conveniently used to predict the T_d of ILs.

6.4 Electrical Conductivity of Ionic Liquids

6.4.1 The Non-Linear QSPR Model^{*}

Implementing the SS algorithm explained in section 5.4.1, the optimal subset of 10 molecular descriptors was selected. Table 6.8 reports the selected molecular descriptors. As mentioned earlier, in all computational steps in this study, temperature is considered as a molecular descriptor.

The LSSVM computational steps as explained have been performed following the described procedure in section 3.1.4. Thus, the optimal values of σ^{-2} and γ are 8.5048 and 7884361327.30237.

The AARD of the model results from experimental data for each of ILs used in this study are presented in Table 6.9.

^{*} The model has been published in *Chemical Engineering Science*, 101, 478-485, 2013.

Table 6.8 The optimal subset of molecular descriptors describing electrical conductivity of ionic liquids along with their definition.

Molecular descriptor	Definition
T / K	Temperature
nP ^a	number of phosphorous atoms
MATS2m ^a	Moran autocorrelation - lag 2 weighted by atomic masses
GATS3e ^a	Geary autocorrelation - lag 3 weighted by atomic Sanderson electronegativities
Mor19p ^a	MoRSE - signal 19 weighted by atomic polarizabilities
CH2X2 ^c	number of CH2X2 groups
X3Av ^c	average valence connectivity index chi-3
R3m ^c	R autocorrelation of lag 3 weighted by atomic masses
R4p ^c	R autocorrelation of lag 4 weighted by atomic polarizabilities
f-veVSA ^c	fractional van der Waals surface area of atoms with negative charge

*Superscripts a and c denotes the anion- or cation-based descriptors.

An important point that needs careful attention is that for some ILs the experimental data reported in various references significantly differs from each other. Thus, it is very difficult to decide about which reported data are correct and which should be eliminated. In addition, there are some fluctuations in the experimental data that may be due to some difficulties that are associated with measuring the experimental data. These issues are major causes of error in most of the ILs studied.

A comparison between the proposed model and the only other model available for estimation of the electrical conductivity of ILs (Matsuda *et al.* [39]) shows that the proposed model demonstrates significantly better results and shows an AARD of 1.9% in a wide range of temperature and for a data set significantly larger than that of Matsuda *et al.*

Table 6.9 The statistical parameters of the proposed QSPR for electrical conductivity of ionic liquids

Statistical Parameter	
training set	
R ²	0.999
Average absolute relative deviation ^a	1.8
Standard deviation error ^b	0.06
Root mean square error ^c	0.06
N ^d	783
validation set	
R ²	0.999
Average absolute relative deviation	2.5
Standard deviation error	0.09
Root mean square error	0.09
N	97
test set	
R ²	0.999
Average absolute relative deviation	2.7
Standard deviation error	0.07
Root mean square error	0.07
N	97
total	
R ²	0.999
Average absolute relative deviation	1.9
Standard deviation error	0.06
Root mean square error	0.06
N	977

6.4.2 The Non-Linear GC model^{*}

The group contributions have been determined using the LSSVM [235] mathematical method explained in section 5.4.2. The optimized values of the LSSVM algorithm [235] have been calculated as follows: $\gamma = 610524.4217$ and $\sigma^{-2} = 14.4637$.

^{*} The model has been published in *Chemical Engineering Research and Design*. 92, 66-79, 2014.

It has been found that the squared correlation coefficients, absolute average deviations, standard deviation errors, and root mean square errors of the developed model for the “training” set, the “validation” set, the “test” set, and the main dataset for the LSSVM-GC model are 0.997, 0.997, 0.994, 0.997, 3.1 %, 3.6 %, 4.9 %, 3.3 %, 0.16, 0.17, 0.20, 0.16, 0.16, 0.17, 0.20, and 0.16, respectively. The AARD of the model results from experimental data for each of ionic liquids used in this study are presented as supplementary file in the supplementary DVD [437].

In comparison with the previous models it should be stated that the model is not comparable with majority of previous models due to this fact that majority of them have been developed just for small number of ILs or just for a few limited temperatures. Furthermore, some of them have been proposed to introduce new theories and they have not been intended to be used new models. As a result, in this section, we compare our model with the models proposed by Matsuda *et al.* [39], Tochigi and Yamamoto [53], and Gardas and Coutinho [46].

In comparison, with the model proposed by Matsuda *et al.* [39], the present model is more accurate and more comprehensive. The model developed here can represent/predict the electrical conductivity of ILs with a lower AARD of 5% over a wide range of temperatures (maximum deviation 40% compared to the several hundred percent with the model of Matsuda *et al.* [39]. Furthermore, the model is more comprehensive (1077 experimental data vs. 139 experimental data) than the model proposed by Tochigi and Yamamoto [53]. Furthermore, in terms of the parameters of the models, the model presented here just uses chemical substructures to represent/predict the electrical conductivities of ILs, however, the parameters of the model proposed by Tochigi and Yamamoto [53] need quantum chemical calculations. Also, the model presented here is better than that of Gardas and Coutinho [46] because it has been developed for a significant larger data set and this makes it more comprehensive. Furthermore, our model

showed lower deviation from experimental data compared to the model proposed by Gardas and Coutinho (3.3% vs. 4.6%) [46].

6.5 The Non-Linear QSPR for Normal Boiling Point Temperature [★]

Pursuing the steps discussed in sections 3.1.3 and 5.5.1, the developed has the structure of 44-40-1 (44 molecular descriptors are regarded as the inputs of the algorithm, the second layer, viz. hidden layer is composed of 40 neurons, and finally one neuron was assigned for the output layer). Parametric description of the proposed ANN model is presented as eqn. 6.11:

6.11

$$NBP_{\text{Calc}}(i) = W_2 \times (\tanh((W_1 \times T_i) + b_1)) + b_2$$

In this equation T_i denotes i -th column of the input matrix (T). The input matrix elements are molecular descriptors (row) of the training set compounds (column). W_1 and b_1 are the weight and bias matrices of second layer, respectively. Similarly, W_2 and b_2 represent the weight and bias matrices of third layer (output layer), respectively. The dimension of the neural network parameters are as follows:

$$W_1 (\text{Weight 1}) = 40 \times 44, W_2 (\text{Weight 2}) = 1 \times 40,$$

$$b_1 (\text{Bias 1}) = 40 \times 1, b_2 (\text{Bias 2}) = 1 \times 1$$

Therefore, the ratio of the data/ (parameters of the ANN module including the weights and biases) is equal to 7.72 (14216 /1841). Therefore, the model is valid and is not over-fitted.

The model derived has forty four descriptors to predict normal melting temperatures. Descriptors

[★] The model has been published in *Fluid Phase Equilibria*. 354, 250-258, 2013.

as well as their definitions and their classes are reported in the published article [438]. Statistical parameters of the derived model and a comparison between predicted versus literature NBP data are presented in the published article [438]. By analyzing the absolute relative deviations, the large portion of the investigated NBP values were successfully predicted within a promising range of 0-3 % (62.5 %). 21.4% of investigated values were predicted within the range of 3.1%-6%; 8.33% of data within the range of 6.1%-9% and merely 7.77% of investigated data had an error greater than 9 %. As was discussed earlier, 84 % of predicted values have deviations of less than 6%. The distribution of the deviations was also studied and is presented as **Error! Reference source not found.** As is shown in this figure, the distributions of deviations are identical for the training, validation and test subsets. The list of compounds studied as well as their predicted normal melting temperatures is presented as supplementary file in the supplementary DVD. [438].

An advantage of the proposed model is the use of the largest datasets for both development and testing of the model. There are a relatively few previous studies in literature which tackle the handling of very large datasets for the sake of model derivation with the aid of QSPR methodology. Difficulties in selecting the appropriate mathematical method and optimization strategy for treating the molecular descriptors to evaluate the normal boiling point may be one of the main reasons for a scarcity of methods for handling large datasets. Apart from considerations with regard to the computation times associated with the model development, there are several

Another drawbacks associated with the model development as well. The procedure of selecting convenient parameters from a descriptor pool is very complex owing to the large number of compounds as well as their associated descriptors.

Contrary to group contribution models which merely contain functional fragments, determination of some of the model parameters (descriptors) are not easy and computational software is required.

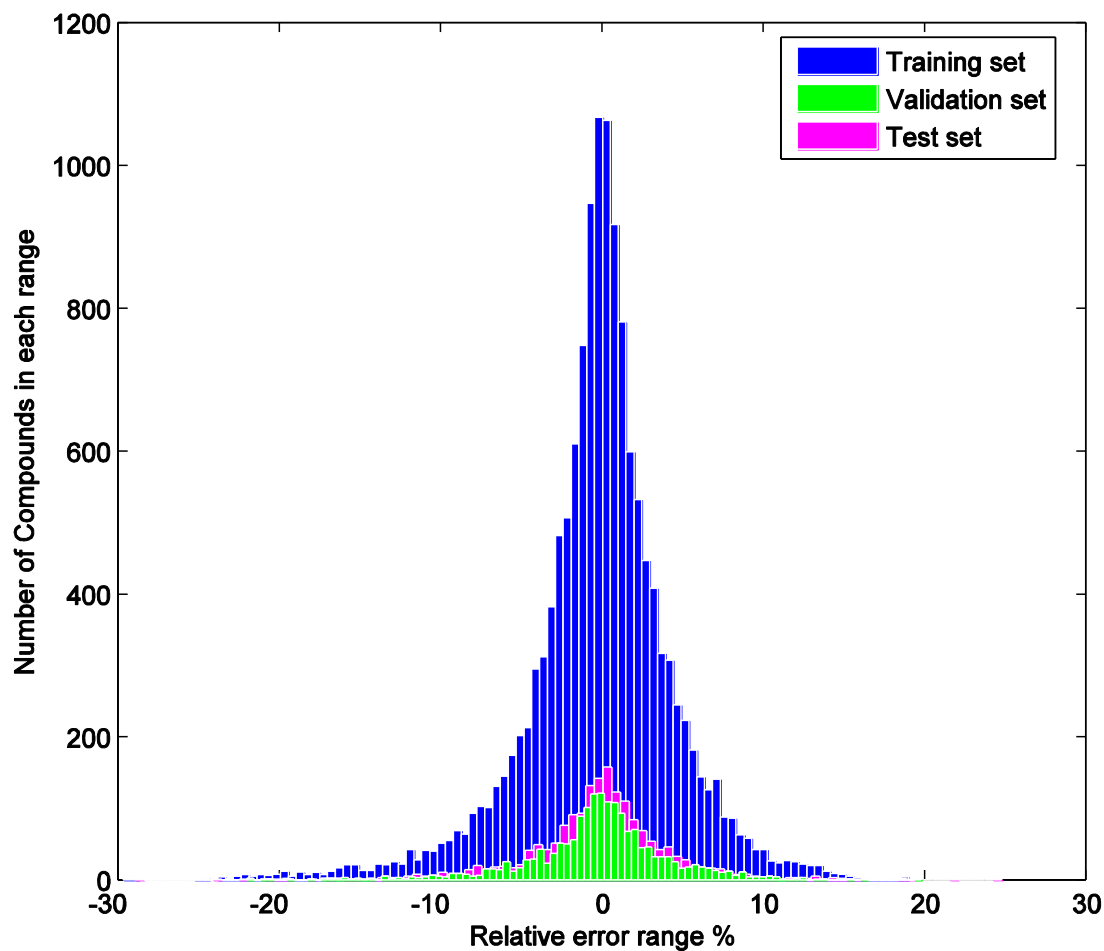


Figure 6.3 The normal boiling point temperatures distribution of error based on the training, the validation and the test sets

A closer look at prior models for the estimation of the normal boiling point of organic compounds revealed that the majority of them employed small datasets to establish predictive correlations. Exploitation of a large database which covers a significant range of chemical compounds is of great importance to develop a widely applicable and accurate model.

Among the previous models, some [74, 83] merit further discussion with regard to the databases implemented.

By employing database of 6000 experimental boiling point data, extracted from Registry of physiochemical Data, Chalk et al. [83] introduced a QSPR model which had satisfactory performance: $R^2 = 0.96$ (Training set) and Error of Standard deviation equal to 16.5 K for 6000 training set values. From a modeling point of view, Chalk et al. did not reveal the detailed procedure for their descriptor selection. This step is the critical bottleneck for handling very large databases. The feature selection procedure implemented by Chalk et al. was formal inference-based recursive modeling (FIRM). Similarly, as 3 layer feed forward artificial neural network with sigmoid transfer function was implemented in Chalk et al.'s work with a network architecture of 18:10:1. The ratio of the number of training molecules to the number of model descriptors in the present model is $(14216/44 \approx 323)$, which is close to the associated value in Chalk et al.'s work $(6000/18 \approx 333)$. The high ratio can be interpreted as the immunity of the proposed models to overtraining. The performance of Chalk et al.'s model is slightly better than the model proposed in this study in terms of the model statistical parameter ($R^2(\text{Chalk model}) = 0.96$, $R^2(\text{this study}) = 0.943$), at expense of using a database which approximately half the size of that used in this study $(14216/6000 \approx 2.369)$. In the Chalk et al. model, 56.0% of the training set estimations have deviations below 10 K, 84.9% below 20 K, and 93.4% below 30 K. In the model proposed in this study, 95.2% of the training set estimations have deviations below 10 K, 99.66% below 20 K, and all of estimated values have a deviation below 30 K. In the light of aforementioned values, the model proposed in the present work can be considered as being superior to that of Chalk et al.

Another example of a large dataset used in literature for prediction of normal boiling points is the GC based model proposed by Stein and Brown [74]. By introducing more functional groups, as well as correction for temperature dependent bias, they developed a modified version of the Jobak and Reid correlation which can predict the boiling points of 6584 compounds with an average absolute error of 20.4 K. No statistical parameter other than average absolute error was reported in their work which makes it difficult for detailed comparison between models. As stated in their work, a large number of boiling points were measured at pressures less than the 1 atm which necessitates extrapolation to obtain corresponding values at 1 atm. Employment of extrapolated data instead of "real" experimental data would resultantly decrease the reliability of the model.

6.6 Sublimation Enthalpy at the Standard Reference Temperature of 298 K[☆]

In order to find the optimal model in terms of both the number of chemical substructures and accuracy, a threshold value of 0.01 was considered for the decrease in *AARD%* as a stopping criterion. It means that when the improvement of the model *AARD%* was less than 0.01, the SS algorithm was automatically stopped and reported the final model. The optimal model was obtained using 147 chemical substructures. The model obtained is as follows:

6.12

$$\Delta_{sublimation}H_m(kJ.mol^{-1}) = \sum_{i=1}^{147} n_i \times \Delta_{sublimation}H_m^i + \Delta_{sublimation}H_m^0$$

where $\Delta_{sublimation}H_m^0$, $\Delta_{sublimation}H_m^i$ and n_i are the intercept of the equation, the contribution of the *i*th chemical substructure to the sublimation enthalpy, and the number of occurrences of the

[☆] The model has been published in *Fluid Phase Equilibria*. 354, 265-285, 2013.

ith chemical substructure in every chemical structure of pure compounds, respectively. The subset of 147 chemical substructures and their contribution to the sublimation enthalpy are tabulated in the in supplementary file in the supplementary DVD. [414].

The model results [414] show that it can successfully predict the standard molar enthalpies of sublimation of pure organic compounds at 298 K. The average absolute relative deviation, standard deviation error, and root mean square error of the model are 6.3%, 10.5, and 10.5 for the training set; 6.3%, 10.7, and 10.7 for the validation set; and 6.3%, 12.7, and 10.8 for the test set, respectively.

Unfortunately, a comprehensive comparison between the presented model and the previous models is not possible because they have mostly developed for small groups/classes of compounds. Even the largest dataset used by Ouvrard and Mitchell [106] which comprised of sublimation enthalpies of 261 organic compounds, when compared with the data used in this study is very small.

In order to compare the performance of the presented model developed in this study with that proposed by Ouvrard and Mitchell [106], a comparison was made based on the chemical families of compounds that were used by Ouvrard and Mitchell [106] in their studies. They categorized the compounds within their data set as aliphatic hydrocarbons, aromatic hydrocarbons, and non-hydrogen bonding compounds. We used the same classification for our main dataset in order to make a comparison. The results are presented in **Error! Reference source not found.** As can be seen, the model presented by Ouvrard and Mitchell [106] predicts the sublimation enthalpies of the aliphatic and aromatic hydrocarbons slightly better than the model presented in our study. A similar behavior can be observed for non-hydrogen bonding compounds. It should be noted that the number of hydrocarbons in our dataset is significantly larger than that of Ouvrard and Mitchell

[106]. Another point to consider is that majority of the compounds for which the experimental sublimation enthalpies have been reported are capable of forming hydrogen bonding. However, most of the compounds used by Ouvrard and Mitchell [106] to develop their model are non-hydrogen bonding. This latter detail may be considered as a drawback of their model.

Table 6.10. Comparison of the sublimation enthalpy model and the model proposed by Ouvrard and Mitchell [106]

Chemical family	Statistical parameter	Ouvrard and Mitchell [106]	The presented model
Aliphatic hydrocarbons	n	33	38
	R^2	0.968	0.932
	$RMSE$	7.42	8.61
Aromatic hydrocarbons	n	50	79
	R^2	0.965	0.836
	$RMSE$	7	10.85
non-hydrogen bonding compounds	n	156	164
	R^2	0.896	0.885
	$RMSE$	9.98	10.56
Various compounds	n	226	1269
	R^2	0.925	0.826
	$RMSE$	9.58	10.79

6.7 Vaporization Enthalpy of Organic Compounds at the Standard Reference

Temperature of 298 K[★]

The procedure outlined in section 5.7.1 was pursued by introducing the collection of 323 chemical sub-structures into the SS algorithm.

According to the algorithm, the optimal vaporization enthalpy model includes 150 chemical sub-structures. The model obtained is as follows:

[★] The model has been published in *Fluid Phase Equilibria*. 360, 279–292, 2013.

$$\Delta_{\text{vaporization}}H_m^\circ(kJ.mol^{-1}) = \sum_{i=1}^{150} n_i \times \Delta_{\text{vaporization}}H_m^i + \Delta_{\text{vaporization}}H_m^0$$

where $\Delta_{\text{vaporization}}H_m^0$, $\Delta_{\text{vaporization}}H_m^i$ and n_i are the intercept of the equation, the contribution of the i th chemical sub-structure to the vaporization enthalpy and the number of occurrences of the i th chemical sub-structure in every chemical structure of pure compounds, respectively. The subset of 150 chemical sub-structures and their contribution to the vaporization enthalpy are tabulated and presented as supplementary file in the supplementary DVD [436].

The model results demonstrate that it can successfully predict the standard molar vaporization enthalpies of pure organic compounds at 298 K. The %AARD, standard deviation error, and root mean square error of calculated values based on the model from the experimentally measured vaporization enthalpies are 3.7%, 4.45, and 4.45 for the training set, 3.7%, 4.51, and 4.52 for the validation set, and 3.7%, 3.78, and 3.78 for the test set, respectively.

According to the results, the model predicts the enthalpies of 689 hydrocarbons with an %AARD of 2.4%. There are 14 hydrocarbons for which the model shows an %AARD of more than 10%. The compounds are trans 2,2,4,6,6-pentamethyl-3-heptene (24%), cis 2,2,4,6,6-pentamethyl-3-heptene (20.7%), cis 1,2-diphenylethylene (20.3%), 4,4-dimethyl-1-hexene (19.5%), 1-cyclopropyl-1,3-pentadiene (19.2%), 1-trans-5-trans-9-cis-cyclododecatriene (13.8%), 1,1,4-trimethylcyclohexane (12%), bicyclo[2.2.1]hepta-2,5-diene (11.7%), dibenz[a,c]anthracene (10.9%), 5,5-bis(3,3'-dimethylbutyl)-2,2,8,8-tetramethylnonane (10.5%), 5-ethylidene-2-norbornene (10.4%), perylene (10.3%), and 2-methylbicyclo[2.2.2]oct-2-ene (10%).

The model predicts the vaporization enthalpy of nitrogen compounds within the AARD of 4.7% deviation compared with experimental values. The %AARD of the model results from experimental vaporization enthalpy for oxygenated compounds is 4.5%. The majority of the outliers are for oxygenated compounds (55 out of 71). According to the results, the model can predict the vaporization enthalpies of the phosphorous-containing compounds with a reasonable AARD of 2.7%. Sulfur-containing compounds are a class of compounds for which the model shows an AARD of 3.6%. Fluorine-containing compounds are one of the important classes of compounds whose vaporization enthalpies are estimated by the model with an AARD of 4.4%. The model predicts the vaporization enthalpies of chlorine-containing compounds better than fluorine-containing compounds in terms of AARD (3.7% vs. 4.4%). Based on the model analysis, the vaporization enthalpies of bromine-containing compounds are successfully predicted by the model. The model AARD for this class of compounds is 2.5%, which is less than those of previously mentioned halogenated compounds.

Unfortunately, a comprehensive comparison between the model proposed in this study and previously proposed models is not possible because they have mostly been developed for small groups/classes of compounds and it is not possible to use most of the previous models for the compounds used in this study.

Very recently, Santos and Leal [154] reviewed various models for the estimation of the vaporization enthalpy at the standard temperature of 298.15. They used a dataset of 83 hydrocarbons including alkanes, alkenes, polyenes, diynes, cycloalkanes, alkylidenecycloalkenes, cycloalkenes, and benzene related compounds to compare their model with the models proposed by Ducros et al. [141, 142], Guthrie and Taylor [145], Chickos et al. [137], Domalski and Hearing [439], Constantinou and Gani [138], and Kolská et al. [146]. The same dataset used here to

Table 6.11. Comparison between the presented vaporization enthalpy model and the previous models suggested by Ducros et al. [141, 142], Guthrie and Taylor [145], Chickos et al. [137], Domalski and Hearing [439], Constantinou and Gani [138], and Kolská et al. [146], using the data set comprised of 83 compounds used in Santos and Leal [154] studies.

No.	Compound	exp.	This work		Santos and Leal		Ducrosa et al.		Guthrie and Taylor		Chickos et al.		Domalski et al.		Constantinou and Gani		Kolská et al.		
			This work	AE	%ARD	AE	%AARD	AE	%ARD	AE	%ARD	AE	%ARD	AE	%ARD	AE	%ARD	AE	%ARD
1	3,3-Diethylpentane	42.6	43.5	0.9	2.1	1.1	2.6	0.1	0.2	0.7	1.6	2.9	6.8	0.4	0.9	0.6	1.4	0.2	0.5
2	2,2,4-Trimethylhexane	40.7	40.3	0.4	0.9	0.1	0.3	0.5	1.2	1.3	3.2	3.1	7.6	0.9	2.2	0.7	1.7	0.7	1.7
3	2,2,5-Trimethylhexane	40.2	40.7	0.5	1.2	0.1	0.3	1.0	2.5	1.8	4.5	2.6	6.5	1.4	3.5	0.1	0.3	0.3	0.8
4	2,3,5-Trimethylhexane	41.4	41.5	0.1	0.3	0.5	1.2	0.9	2.2	1.3	3.1	2.5	6.0	0.9	2.2	0.6	1.5	0.3	0.7
5	Nonane	46.4	46.1	0.3	0.7	0.3	0.7	0.3	0.7	0.4	0.9	1.3	2.8	0.0	0.0	1.2	2.6	0.9	1.9
6	2,4-Dimethyloctane	47.1	47.3	0.2	0.4	0.6	1.3	1.4	3.0	2.0	4.3	1.5	3.2	1.7	3.6	0.6	1.3	0.1	0.2
7	2-Methylnonane	49.6	49.2	0.4	0.8	0.3	0.6	0.2	0.4	0.9	1.8	1.9	3.8	0.5	1.0	0.5	1.0	0.1	0.2
8	Decane	51.4	51.1	0.3	0.6	0.4	0.8	0.3	0.6	0.5	1.0	1.6	3.1	0.1	0.2	0.9	1.8	0.6	1.2
9	Dodecane	61.5	61.1	0.4	0.7	0.9	1.5	0.4	0.7	0.5	0.8	2.3	3.7	0.2	0.3	0.1	0.2	0.1	0.2
10	Tri-tert-butylmethane	55.3	48.5	6.8	12.4	3.5	6.3	1.5	2.7	0.0	0.0	9.9	17.9	1.4	2.5	7.0	12.7	6.1	11.0
11	5-Ethyl-5-methyldecane	60.5	61.5	1.0	1.7	2.0	3.3	1.9	3.1	3.1	5.1	2.1	3.5	2.7	4.5	1.3	2.2	0.8	1.3
12	3,3,6,6-Tetraethyloctane	74.3	72.8	1.5	2.1	1.6	2.2	0.6	0.8	1.0	1.4	7.3	9.8	1.0	1.4	3.0	4.0	3.6	4.9
13	Hexadecane	81.4	81.0	0.4	0.4	1.4	1.7	0.4	0.5	0.9	1.1	3.5	4.3	0.7	0.9	1.2	1.5	1.1	1.4
14	2,4,4-Trimethyl-2-pentene	37.5	35.7	1.8	4.9	0.2	0.5	0.7	1.9	0.3	0.8	2.5	6.7	0.5	1.3	1.1	2.9	1.0	2.7
15	(Z)-2,2-Dimethyl-3-hexene	37.2	36.7	0.5	1.4	0.1	0.3	0.7	1.9	0.2	0.5	2.2	5.9	0.2	0.5	1.3	3.5	0.4	1.1
16	(E)-2,2-Dimethyl-3-hexene	37.2	36.7	0.5	1.4	0.3	0.8	0.7	1.9	0.2	0.5	2.2	5.9	0.4	1.1	1.3	3.5	0.4	1.1
17	1-Octene	40.4	40.3	0.1	0.2	0.0	0.0	0.1	0.3	0.5	1.2	0.1	0.3	0.0	0.0	0.1	0.3	0.6	1.5
18	(Z)-2-Octene	40.2	40.5	0.3	0.7	1.0	2.5	1.3	3.2	2.3	5.7	0.3	0.8	0.8	2.0	0.4	1.0	1.2	3.0
19	(E)-2-Octene	40.2	40.5	0.3	0.7	1.3	3.2	1.3	3.2	2.3	5.7	0.3	0.8	1.2	3.0	0.4	1.0	1.2	3.0
20	(Z)-3-Octene	39.7	40.5	0.8	1.9	1.2	3.0	1.9	4.8	3.3	8.3	0.8	2.0	1.0	2.5	0.6	1.5	1.5	3.8
21	(E)-3-Octene	40.2	40.5	0.3	0.7	1.0	2.5	1.4	3.5	2.8	7.0	0.3	0.8	0.9	2.2	0.1	0.3	1.0	2.5
22	(Z)-4-Octene	39.7	40.5	0.8	1.9	1.2	3.0	1.9	4.8	3.3	8.3	0.8	2.0	1.0	2.5	0.6	1.5	1.5	3.8
23	(E)-4-Octene	42.9	40.5	2.4	5.7	1.7	4.0	1.3	3.0	0.1	0.2	2.4	5.6	1.8	4.2	2.6	6.1	1.7	4.0
24	2,6-Dimethyl-1-heptene	45.9	43.1	2.8	6.1	2.3	5.0	1.3	2.8	1.0	2.2	2.8	6.1	1.9	4.1	3.4	7.4	3.0	6.5
25	1-Nonene	44.7	45.1	0.4	0.8	0.5	1.1	0.8	1.8	1.2	2.7	0.4	0.9	0.8	1.8	0.5	1.1	1.0	2.2

26	1-Decene	50.4	50.1	0.3	0.7	0.3	0.6	0.1	0.2	0.6	1.2	0.6	1.2	0.2	0.4	0.6	1.2	0.0	0.0
27	3-Methyl-3-propyl-1-heptene	50.9	50.8	0.1	0.2	1.0	2.0	0.9	1.8	0.4	0.8	1.9	3.7	1.7	3.3	0.1	0.2	0.2	0.4
28	1-Undecene	54.3	55.1	0.8	1.4	0.6	1.1	1.2	2.2	1.8	3.3	0.2	0.4	1.4	2.6	0.2	0.4	0.8	1.5
29	1-Dodecene	60.8	60.0	0.8	1.2	1.1	1.8	0.3	0.5	0.3	0.5	1.6	2.6	0.0	0.0	1.7	2.8	0.9	1.5
30	1-Hexadecene	80.2	80.0	0.2	0.2	1.1	1.4	0.2	0.3	1.2	1.5	2.3	2.9	1.0	1.3	2.5	3.1	1.4	1.8
31	1,5-Hexadiyne	31.9	35.4	3.5	11.1	1.4	4.4	3.5	11.0	31.9	100.0	0.8	2.5	6.9	21.6	0.3	0.9	0.7	2.2
32	(Z)-1,3,5-Hexatriene	34.1	31.7	2.4	7.1	1.1	3.2	0.5	1.5	25.7	75.4	3.0	8.8	1.0	2.9	2.4	7.0	3.8	11.1
33	1,7-Octadiyne	41.7	44.9	3.2	7.6	1.3	3.1	3.7	8.9	31.6	75.8	1.2	2.9	7.3	17.5	0.8	1.9	0.3	0.7
34	Bicyclopropyl	31.7	33.9	2.2	6.9	3.3	10.4	5.8	18.3	5.3	16.7	0.6	1.9	14.6	46.1	0.7	2.2	2.8	8.8
35	1-Cyclopropylpenta-1,3-diyne	51.9	50.0	1.9	3.7	1.2	2.3	41.0	79.0	41.8	80.5	11.4	22.0	32.3	62.2	5.9	11.4	5.7	11.0
36	1,1,2-Trimethylcyclopentane	38.1	36.6	1.5	4.0	0.5	1.3	0.4	1.1	2.5	6.6	3.1	8.1	6.1	16.0	0.5	1.3	2.5	6.6
37	1,1,3-Trimethylcyclopentane	36.6	35.4	1.2	3.1	0.1	0.3	1.1	3.0	1.0	2.7	0.5	1.4	4.6	12.6	0.0	0.0	2.9	7.9
38	Butylcyclopentane	45.9	45.6	0.3	0.7	0.3	0.7	0.4	0.9	1.7	3.7	0.8	1.7	5.3	11.6	0.3	0.7	0.2	0.4
39	1,1,3-Trimethylcyclohexane	41.9	39.5	2.4	5.7	0.6	1.4	0.1	0.2	1.2	2.9	0.1	0.2	0.9	2.2	0.9	2.2	4.2	10.0
40	1,1,4-Trimethylcyclohexane	45.6	40.1	5.5	12.0	4.3	9.4	3.8	8.3	4.9	10.8	3.8	8.3	2.8	6.1	1.0	2.2	7.9	17.3
41	Isopropylcyclohexane	44.0	44.2	0.2	0.4	0.2	0.5	0.1	0.2	1.2	2.7	0.9	2.1	0.9	2.1	0.3	0.7	0.3	0.7
42	Bicyclopentyl	50.4	49.4	1.0	2.0	1.4	2.8	1.0	2.0	3.8	7.5	0.6	1.2	10.5	20.8	2.0	4.0	0.1	0.2
43	tert-Butylcyclohexane	47.0	46.7	0.3	0.7	0.6	1.3	0.3	0.6	1.2	2.6	2.6	5.5	0.5	1.1	2.3	4.9	0.0	0.0
44	Bicyclohexyl	58.0	58.8	0.8	1.4	1.4	2.4	1.6	2.8	1.3	2.2	1.2	2.1	3.5	6.0	0.9	1.6	0.5	0.9
45	Decylcyclohexane	78.7	80.5	1.8	2.3	0.3	0.4	1.5	1.9	0.9	1.1	0.8	1.0	3.3	4.2	0.9	1.1	0.6	0.8
46	Dodecylcyclohexane	88.9	90.5	1.6	1.8	0.3	0.3	1.3	1.5	0.8	0.9	1.6	1.8	3.3	3.7	1.8	2.0	1.4	1.6
47	3-Isopropyl-6-methylene cyclohexene	49.2	50.3	1.1	2.2	1.0	2.0	2.6	5.3	10.0	20.3	1.5	3.1	4.2	8.5	3.8	7.7	1.1	2.2
48	Cyclohexene, 1-methyl-4-(1-methylethylidene)-	53.2	52.0	1.2	2.3	1.0	1.9	3.2	6.0	5.2	9.8	3.4	6.4	1.6	3.0	5.3	10.0	4.0	7.5
49	1,3-Dimethylcyclopentene	35.0	36.5	1.5	4.4	0.4	1.1	1.8	5.1	1.8	5.1	0.8	2.3	3.6	10.3	0.4	1.1	4.7	13.4
50	1,4-Dimethylcyclopentene	34.5	36.5	2.0	5.9	0.1	0.3	2.8	8.1	1.1	3.2	1.3	3.8	4.1	11.9	0.6	1.7	5.2	15.1
51	1,5-Dimethylcyclopentene	36.5	37.1	0.6	1.6	1.4	3.8	0.3	0.8	0.3	0.8	2.8	7.7	2.1	5.8	1.0	2.7	4.3	11.8
52	4-Ethylcyclopentene	36.8	36.6	0.2	0.5	0.1	0.3	0.1	0.3	1.5	4.1	1.0	2.7	2.0	5.4	3.2	8.7	1.1	3.0
53	3-Methylcyclohexene	38.8	36.6	2.2	5.6	3.1	8.0	2.0	5.2	2.3	5.9	3.0	7.7	1.0	2.6	4.7	12.1	1.4	3.6
54	1-Methyl-4-isopropyl-1,4-cyclohexadiene	51.4	51.3	0.1	0.1	0.6	1.2	8.7	16.9	9.6	18.7	3.7	7.2	2.3	4.5	5.7	11.1	1.1	2.1
55	1-Ethenyl-2-methylbenzene	49.3	48.0	1.3	2.7	0.0	0.0	10.7	21.7	12.4	25.2	6.2	12.6	3.1	6.3	3.3	6.7	0.1	0.2
56	1-Ethenyl-3-methylbenzene	49.5	47.7	1.8	3.6	0.8	1.6	11.8	23.8	12.6	25.5	4.4	8.9	2.0	4.0	3.1	6.3	0.5	1.0
57	1-Ethenyl-4-methylbenzene	48.9	48.0	0.9	1.8	0.2	0.4	11.7	23.9	12.0	24.5	3.8	7.8	0.7	1.4	3.7	7.6	1.7	3.5

58	(Z)-1-Propenylbenzene	49.1	47.4	1.7	3.4	0.4	0.8	10.1	20.6	10.2	20.8	4.0	8.2	0.5	1.0	2.5	5.1	2.4	4.9
59	(E)-1-Propenylbenzene	46.9	47.4	0.5	1.1	2.9	6.2	7.9	16.8	8.0	17.1	1.8	3.8	3.1	6.6	4.7	10.0	4.6	9.8
60	Allylbenzene	46.2	46.4	0.2	0.5	0.0	0.0	3.9	8.4	4.2	9.1	1.1	2.4	3.6	7.8	2.4	5.2	0.2	0.4
61	1-Ethyl-2-ethenylbenzene	52.1	52.3	0.2	0.5	2.0	3.8	9.3	17.9	11.8	22.7	4.4	8.5	2.5	4.8	5.0	9.6	1.1	2.1
62	1-Ethyl-3-ethenylbenzene	53.9	52.4	1.5	2.8	0.4	0.7	12.1	22.5	13.6	25.2	4.1	7.6	2.9	5.4	3.2	5.9	0.1	0.2
63	1-Ethyl-4-ethenylbenzene	52.6	52.7	0.1	0.2	0.9	1.7	11.3	21.5	12.3	23.4	2.8	5.3	1.0	1.9	4.5	8.6	1.9	3.6
64	1-Buten-2-ylbenzene	51.8	51.8	0.0	0.1	2.1	4.1	8.0	15.5	8.3	16.0	2.0	3.9	4.4	8.5	3.6	7.0	2.9	5.6
65	(3-Methyl-1-buten-2-yl) benzene	53.2	54.9	1.7	3.2	3.4	6.4	6.3	11.8	4.8	9.0	0.8	1.5	2.0	3.8	5.5	10.3	4.2	7.9
66	3-Methyl-1-tert-butylbenzene	51.1	53.4	2.3	4.5	1.5	2.9	2.8	5.5	1.0	2.0	2.1	4.1	0.3	0.6	7.4	14.5	0.3	0.6
67	4-Methyl-1-tert-butylbenzene	52.2	53.7	1.5	2.9	0.4	0.8	1.2	2.3	0.1	0.2	3.2	6.1	0.2	0.4	6.3	12.1	0.3	0.6
68	(3,3-Dimethyl-1-buten-2-yl) benzene	53.2	57.4	4.2	7.9	7.0	13.2	3.2	6.0	5.1	9.6	0.5	0.9	0.9	1.7	7.1	13.3	6.6	12.4
69	Cyclohexylbenzene	60.8	60.7	0.1	0.2	1.6	2.6	0.7	1.2	1.8	3.0	1.6	2.6	0.4	0.7	2.3	3.8	1.6	2.6
70	3,5-Dimethyl-1-tert-butylbenzene	56.6	57.9	1.3	2.4	0.8	1.4	2.4	4.2	0.2	0.4	2.9	5.1	2.3	4.1	7.5	13.2	1.1	1.9
71	1,3,5-Triethylbenzene	59.2	61.7	2.5	4.3	2.3	3.9	1.0	1.7	1.5	2.5	0.0	0.0	4.2	7.1	2.8	4.7	0.3	0.5
72	1,3-Di-isopropylbenzene	56.2	58.7	2.5	4.5	0.7	1.3	0.6	1.1	0.3	0.5	1.2	2.1	1.6	2.9	1.7	3.0	0.9	1.6
73	1,4-Di-isopropylbenzene	56.5	59.0	2.5	4.4	1.0	1.8	0.2	0.4	0.0	0.0	1.5	2.7	1.3	2.3	1.4	2.5	0.7	1.2
74	1,3-Di-tert-butylbenzene	59.6	63.6	4.0	6.8	3.3	5.5	5.6	9.4	1.5	2.5	2.0	3.4	1.6	2.7	14.5	24.3	0.1	0.2
75	1,4-Di-tert-butylbenzene	63.0	63.9	0.9	1.5	0.1	0.2	1.7	2.7	1.9	3.0	5.4	8.6	1.2	1.9	11.1	17.6	2.7	4.3
76	1,3,5-Tri-isopropylbenzene	64.6	71.0	6.4	9.9	2.3	3.6	4.4	6.8	3.3	5.1	2.4	3.7	0.6	0.9	6.6	10.2	0.9	1.4
77	1,2,4,5-Tetra-isopropylbenzene	75.7	82.3	6.6	8.7	3.9	5.2	7.0	9.3	3.6	4.8	0.9	1.2	4.4	5.8	8.8	11.6	2.1	2.8
78	1,1-Diphenylethylene	71.2	71.4	0.2	0.3	2.9	4.1	9.9	13.9	11.0	15.5	2.6	3.7	2.0	2.8	9.2	12.9	7.1	10.0
79	(Z)-1,2-Diphenylethylene	70.5	72.7	2.2	3.1	3.9	5.5	14.0	19.9	14.4	20.4	1.9	2.7	7.9	11.2	10.5	14.9	9.6	13.6
80	(E)-1,2-Diphenylethylene	79.6	72.7	6.9	8.7	2.5	3.1	23.1	29.0	23.5	29.5	11.0	13.8	0.8	1.0	1.4	1.8	0.5	0.6
81	1,1-Diphenylethane	68.9	70.3	1.4	2.1	3.8	5.5	1.4	2.0	2.1	3.1	2.4	3.5	23.3	33.8	4.4	6.4	4.3	6.2
82	2-Methyl-1,1-diphenylpropane	73.2	76.9	3.7	5.0	0.7	1.0	5.8	7.9	6.6	9.0	2.7	3.7	27.9	38.1	7.3	10.0	7.0	9.6
83	Tri-phenylethene	91.8	100.5	8.7	9.5	13.0	14.2	7.1	7.7	7.7	8.4	4.9	5.3	15.8	17.2	23.4	25.5	20.1	21.9
	Total			1.6	3.1	1.4	2.6	3.7	7.0	5.1	10.6	2.4	4.6	3.3	6.4	3.1	5.5	2.1	4.1

compare the obtained model with the previous ones, even the one proposed by Santos and Leal [154]. The results are presented in Table 6.11 [436].

As demonstrated, the model proposed predicts the vaporization enthalpies of the compounds with an %AARD of 3.1%. However, the models proposed by Ducros et al. [141, 142], Guthrie and Taylor [145], Chickos et al. [137], Domalski and Hearing [439], Constantinou and Gani [138], and Kolská et al. [146] predicts the dataset with an %AARD of 7%, 10.6%, 4.6%, 6.4%, 5.5%, and 4.1%, respectively. Therefore, the model predicts the vaporization enthalpies of the selected 83 compounds better than these previous models. It should be noted that the model proposed by Santos and Leal [154] estimates the vaporization enthalpies of the selected 83 hydrocarbons with an %AARD of 2.6% which shows that their model is slightly better than the present model (2.6% vs. 3.1%). However, the model proposed by Santos and Leal [154] is capable of use for just for hydrocarbons, i.e. it cannot be considered as a general model. As result, the model proposed here can be considered better than previous models.

6.8 Speed of Sound in Saturated Liquids [☆]

Nearly 600 chemical substructures were gathered from previous articles published by the authors and considered to be useful for the representation/prediction of saturated liquid speed of sound.

Since the saturated liquid speed of sound is a temperature dependent property, temperature is included in the database of chemical substructures as a separate column and considered as a chemical substructure.

[☆] The model has been published in *Journal of Molecular Liquids*, 194, 159-165, 2014.

The combination of sequential search algorithm and LSSVM was used to select the most efficient subset of chemical substructures as well as developing a non-linear LSSVM model [235].

The most efficient subset of chemical substructures comprised of 44 chemical substructures including temperature. It is presented as supplementary file in the supplementary DVD [390].

The optimized values of the LSSVM algorithm [235] were calculated by means of the simulated annealing optimization technique [235] as follows: $\gamma = 6737788.288$ and $\sigma^2 = 0.993333622$. The numbers of significant digits for the two aforementioned parameters were obtained by sensitivity analysis of the overall errors of the optimization procedure.

The determined saturated liquid speed of sound values and their absolute relative deviations using this method and compared with the experimental values are presented as supplementary file in the supplementary DVD. [390] It has been found that the squared correlation coefficients, average absolute deviations (AARD%), standard deviation errors, and root mean square errors of the developed model for the “training” set, the “validation” set, the “test” set, and the main dataset for the LSSVM-GC model are 0.998, 0.999, 0.998, 0.998, 0.5 %, 0.6 %, 0.7 %, 0.6 %, 17.31, 9.43, 21.15, 17.12, 17.30, 9.41, 21.13, and 17.12, respectively.

According to the results, there are 9 data points, for which the model shows more than a 10% absolute relative deviation (ARD%); 4 points of which are for propane, 2 points for methane, and one point for each one of carbon dioxide, ethylbenzene, and 3-bromohexane.

A careful consideration of the model results shows that it represents/predicts the saturated liquid speed of sound of propane, methane, carbon dioxide, ethylbenzene, and 3-bromohexane with AARD% of 3.73%, 1.93%, 2.39%, 0.92%, and 1.93%, respectively. Furthermore, the highest deviation of the model results from experimental data is related to propane and carbon dioxide

with deviations of 3.73% and 2.39, respectively. As a result, the model cannot predict the saturated liquid speed of sound of propane and carbon dioxide as accurately as for the other compounds. Moreover, since the model predicts the saturated liquid speed of sound of methane, ethylbenzene, and 3-bromohexane very well, there may be issues with regard to the accuracy of measurement of the data points.

6.9 The Standard Molar Chemical Exergy of Organic Compounds[☆]

The sequential search was initially started with a collection of 140 chemical substructures prepared in the previous step (section 5.9).

The optimal model is obtained using a subset of 47 chemical substructures. The model is presented as follows:

6.14

$$\varepsilon^{\circ} = \varepsilon^{\circ}_0 + \sum_{i=1}^{47} n_i \varepsilon^{\circ}_i$$

where ε°_0 and ε°_i are the intercept of the equation, and the contribution of *i*th chemical substructure to the standard chemical exergy of pure organic compounds, respectively (The units for both ε°_0 and ε°_i are kJ mole⁻¹). The subset of 47 chemical substructures and their contributions to standard chemical exergy are presented in Table 1.

The model can successfully be described using the group contribution method. The average absolute relative deviation, standard deviation error, and root mean square error of the model from

[☆] The model has been published in *Energy* 70, 288-297, 2014.

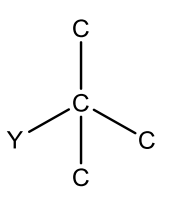
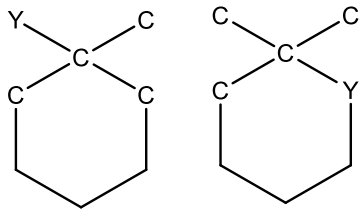
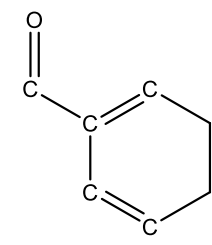
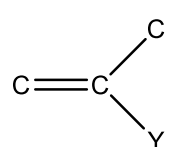
experimental data are (1.6%, 158601, 158393) (1.6%, 260872, 254247) and (1.6%, 145868, 140166) for the training set, validation set, and test set respectively.

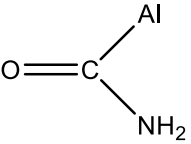
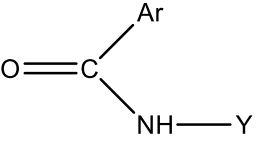
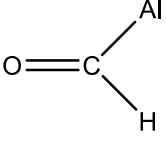
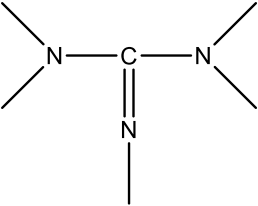
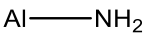
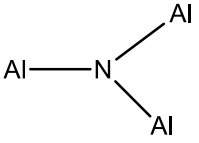
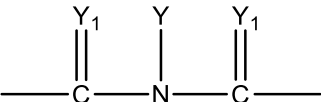
The *AARD%* of the model from literature values for hydrocarbons is equivalent to 2.1%. The maximum absolute relative deviation of the model from literature hydrocarbon data is for ethylene with a value of 11.9%.

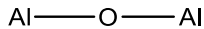
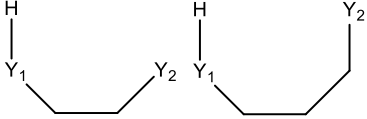
According to the dataset, there are 21 nitrogen containing compounds within the database for which the model gives a promising *AARD%* of 0.3%. The maximum absolute relative deviation among nitrogen compounds is for urea for which the value is 1.5%.

An analysis of the compounds in the dataset shows the presence of 62 oxygenated compounds. The *AARD%* of the model from literature values for the oxygenated compounds is a promising value of 0.9%. The maximum relative deviation for the model among the oxygenated compounds is for phthalic anhydride which has a value of 9.3%.

Table 6.12 The contribution of each chemical substructure to the standard molar chemical exergy pure organic compounds (parameters of equation 6.14)

No.	ID	Chemical substructure	Comment	Substructure Contribution
	ϵ°_0		intercept of eq. (2)	626,757.2
1	ϵ°_1		number of total tertiary C(sp ³) Y = H or any heteroatom	171,171.8
2	ϵ°_2		number of ring tertiary C(sp ³) Y = H or any heteroatom	-209,682.5
3	ϵ°_3		number of non-aromatic conjugated C(sp ²)	-89,442.6
4	ϵ°_4		number of aliphatic secondary C(sp ²) Y = H or any heteroatom	115,290.6

5	ϵ°_5		number of primary amides (aliphatic) Al = H or aliphatic group linked through C	-253,798.9
6	ϵ°_6		number of secondary amides (aromatic) Y = Ar or Al (not H, not C = O)	-615,532.7
7	ϵ°_7		number of aldehydes (aliphatic)	-61,177.5
8	ϵ°_8		number of guanidine derivatives	183,077.7
9	ϵ°_9		number of primary amines (aliphatic) Al = aliphatic group linked through C (not C = O)	154,555.3
10	ϵ°_{10}		number of tertiary amines (aliphatic) Al = aliphatic group linked through C (not C = O)	-13,275.5
11	ϵ°_{11}		number of imides (thio-) Y = H or C Y1 = O or S	-361,685.5

12	ϵ°_{12}		number of ethers (aliphatic) Al = aliphatic group linked through C (not C = O, not C # N)	48,097.1
13	ϵ°_{13}		number of intramolecular H-bonds Y1 = B, N, O, Al, P, S Y2 = N, O, F	-100,695.5
14	ϵ°_{14}	CH ₃ X		253,982.7
15	ϵ°_{15}	R--CR--R		31,298.4
16	ϵ°_{16}	R-C(=X)-X or R-C#X or X=C=X		-135,027.2
17	ϵ°_{17}	H ^a attached to C ⁰ (sp ³) no X attached to next C		57,481.2
18	ϵ°_{18}	H ^a attached to C ³ (sp ³) or C ² (sp ²) or C ³ (sp ²) or C ³ (sp)		-26,030.7
19	ϵ°_{19}	H ^a attached to C ⁰ (sp ³) with 1X attached to next C		41,406.9
20	ϵ°_{20}	phenol or enol or carboxyl OH		-200,332.3
21	ϵ°_{21}	Al-O-Ar or Ar-O-Ar or R..O..R or R-O-C=X		69,949.7
22	ϵ°_{22}	Ar ₂ NH or Ar ₃ N or Ar ₂ N-Al or R..N..R ^b		-146,741.0
23	ϵ°_{23}	absence/presence of C-C	0 or 1	258,074.9
24	ϵ°_{24}	absence/presence of C-O	0 or 1	39,315.2
25	ϵ°_{25}	absence/presence of C-S	0 or 1	803,943.3
26	ϵ°_{26}	absence/presence of C-A-C	0 or 1 (A means any atom except hydrogen)	201,602.5
27	ϵ°_{27}	absence/presence of C-(A) ₂ -C	0 or 1 (A means any atom except hydrogen)	211,229.7
28	ϵ°_{28}	absence/presence of N-(A) ₂ -O	0 or 1 (A means any atom except hydrogen)	-68,150.9
29	ϵ°_{29}	absence/presence of C-(A) ₃ -C	0 or 1 (A means any atom except hydrogen)	178,596.3
30	ϵ°_{30}	absence/presence of N-(A) ₃ -O	0 or 1 (A means any atom except hydrogen)	-179,334.0
31	ϵ°_{31}	absence/presence of O-(A) ₃ -O	0 or 1 (A means any atom except hydrogen)	-45,758.6
32	ϵ°_{32}	absence/presence of C-(A) ₄ -C	0 or 1 (A means any atom except hydrogen)	165,809.6

33	ϵ°_{33}	absence/presence of C-(A) ₅ -C	0 or 1 (A means any atom except hydrogen)	144,194.3
34	ϵ°_{34}	absence/presence of C-(A) ₆ -C	0 or 1 (A means any atom except hydrogen)	131,264.1
35	ϵ°_{35}	absence/presence of C-(A) ₇ -O	0 or 1 (A means any atom except hydrogen)	717,611.9
36	ϵ°_{36}	absence/presence of C-(A) ₈ -C	0 or 1 (A means any atom except hydrogen)	399,063.2
37	ϵ°_{37}	absence/presence of C-(A) ₉ -C	0 or 1 (A means any atom except hydrogen)	543,611.6
38	ϵ°_{38}	absence/presence of C-(A) ₉ -O	0 or 1 (A means any atom except hydrogen)	-204,606.1
39	ϵ°_{39}	C-C		319,737.5
40	ϵ°_{40}	N-A-N	number of N-A-N substructure (A means any atom except hydrogen)	16,773.2
41	ϵ°_{41}	C-(A) ₂ -N	number of C-(A) ₂ -N substructure (A means any atom except hydrogen)	127,623.3
42	ϵ°_{42}	N-(A) ₂ -O	number of N-(A) ₂ -O substructure (A means any atom except hydrogen)	26,865.4
43	ϵ°_{43}	O-(A) ₂ -O	number of O-(A) ₂ -O substructure (A means any atom except hydrogen)	-5,498.2
44	ϵ°_{44}	C-(A) ₃ -C	number of C-(A) ₃ -C substructure (A means any atom except hydrogen)	55,195.2
45	ϵ°_{45}	C-(A) ₃ -N	number of C-(A) ₃ -N substructure (A means any atom except hydrogen)	4,884.9
46	ϵ°_{46}	N-(A) ₃ -O	number of N-(A) ₃ -O substructure (A means any atom except hydrogen)	160,371.9
47	ϵ°_{47}	C-(A) ₅ -C	number of C-(A) ₅ -C substructure (A means any atom except hydrogen)	-10,059.0

R represents any group linked through carbon;

X represents any electronegative atom (O, N, S, P, Se, halogens);

Al and Ar represent aliphatic and aromatic groups, respectively;

= represents a double bond;

represents a triple bond;

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

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

^b Pyrrole-type structure

The model can successfully predict the standard chemical exergy of all the oxygenated compounds within an AARD% of 2.5%, except for three compounds for which the model shows a significant deviation (butanol with 7.8%, L-Sorbose with 8.9% and phthalic anhydride with 9.3%).

There are 3 sulfur containing compounds within the dataset for which the model results in a promising *AARD% value* of 0.3%. The maximum deviation of the model among sulfur containing compounds is for cysteine which has a value of 0.8%.

In order to compare the model proposed in this study with previous models reported in literature [198, 200-202, 204, 206-208], the following points have to be considered. First of all, the proposed model predicts the standard chemical exergy of pure organic compounds at their most stable state without any information about what is this state at the reference state [208]. This point is very important because the knowledge of the physical state of the compound is needed in the majority of previous models to select the correct correlation for the estimation of the chemical exergy [208]. In some cases the physical state of a compound at the reference state may be controversial, as has been reported by Song et al. [208].

In most of the previous models the elemental compositions are used to estimate the chemical exergy of pure organic compounds [198-202, 204, 206, 208] instead of the chemical structure. Although this may be useful for some of the fuels studied, which are composed of various organic and inorganic compounds, the application of elemental based models for pure organic compounds would be limited. This is the most likely reason why the elemental based models cannot take into account the interactions among various chemical groups in pure organic chemical compounds.

The last point that should be considered when comparing the proposed model with previous reported models is that the model presented in this study proposes a direct procedure to estimate

the chemical exergy of pure organic compounds. As has been mentioned earlier, most of the currently available models are based on the estimation of entropy [201-206, 208]. Therefore, most of the available models are developed in two steps, the first step which involves the estimation of entropy of pure organic compounds. This two-step approach may generate some errors in the estimation of the chemical exergy of pure organic compounds. Thus, direct methods for the estimation of chemical exergy of pure organic compounds are better than indirect methods. As a result, a comparison of the proposed model is made with just one of the recently proposed direct methods for the estimation of the standard chemical exergy of pure organic compounds, viz. the model proposed by Gharagheizi and Mehrpooya [207] using the same dataset as used in this study.

As shown in Table 6.13, A comparison between the model proposed by Gharagheizi and Mehrpooya and the model proposed here shows that, in terms of the *AARD%*, the model proposed in this study is more accurate and gives a lower value (1.6% vs. 3%) than the model proposed by Gharagheizi and Mehrpooya [207]. Although the model proposed by Gharagheizi and Mehrpooya [207] gives slightly better results for hydrocarbons (1.5% versus 2.1%), it gives significantly higher deviations for nitrogen containing compounds (5.5% vs. 0.3%), oxygenated compounds (4.8% vs. 0.9%), and sulfur containing compounds (4.3% vs. 0.3%). Also, the proposed model is simpler to apply compared to the model proposed by Gharagheizi and Mehrpooya [207]. This simplicity is mainly due to the fact that the parameters of the proposed model are several chemical substructures that are used to constitute any chemical structure and can be easily counted without any need for additional computational facilities.

Table 6.13. Comparison of the presented model and the previous model proposed by Gharagheizi for the estimation of standard molar chemical exergy of pure organic compounds.

No.	Name	ε° [197]	ε° [207]	ε° (this model)	%ARD	Status
1	Methane	836510	837479	856682	2.4	training
2	Ethane	1504360	1493749	1549457	3.0	training
3	Propane	2163190	2159909	2185759	1.0	training
4	Butane	2818930	2816466	2831689	0.5	training
5	Pentane	3477050	3472976	3500180	0.7	training
6	Hexane	4134590	4129492	4155884	0.5	training
7	Heptane	4786300	4786117	4779915	0.1	training
8	Octane	5440030	5442718	5391015	0.9	training
9	Nonane	6093550	6099346	5870851	3.7	training
10	Decane	6749750	6765669	6749750	0.0	training
11	Undecane	7404520	7422373	7773198	5.0	training
12	Dodecane	8059340	8079121	8253034	2.4	training
13	Tridecane	8714200	8735920	8732870	0.2	training
14	Tetradecane	9368970	9392787	9212706	1.7	test
15	Pentadecane	10023870	10049717	9692542	3.3	test
16	Hexadecane	10678810	10706722	10172378	4.7	training
17	Cyclopropane	2052490	1935966	2188932	6.6	training
18	Cyclobutane	2707730	2602430	2825234	4.3	training
19	Cyclohexane	3928100	3917013	3905863	0.6	training
20	Methylcyclo-hexane	4573030	4573714	4535844	0.8	training

21	Ethylcyclohexane	5246900	5229963	5246744	0.0	training
22	Propyl cyclohexane	5878430	5886942	5925969	0.8	training
23	Butylcyclo-hexane	6534510	6543524	6527010	0.1	training
24	Ethylene	1366610	1279652	1204570	11.9	training
25	Propene	2010840	1936747	2013644	0.1	training
26	1-Butene	2668920	2603054	2659573	0.4	training
27	1-Hexene	3984330	3916220	3983769	0.0	training
28	1-Heptene	4641570	4572698	4607799	0.7	training
29	Acetylene	1269310	1064274	1204570	5.1	validation
30	Propyne	1904070	1722194	1898353	0.3	training
31	1-Butyne	2561190	2379366	2544283	0.7	validation
32	1-Hexyne	3876600	3702643	3868478	0.2	training
33	1-Heptyne	4534300	4359199	4492509	0.9	training
34	Benzene	3310540	3269841	3216089	2.9	training
35	Toluene	3952550	3936233	3942062	0.3	validation
36	Ethylbenzene	4610250	4592769	4652961	0.9	training
37	Propyl benzene	5262930	5249352	5332187	1.3	training
38	Butyl benzene	5908120	5906035	5933228	0.4	training
39	Decyl benzene	9730670	9860632	9744860	0.1	training
40	Naphthalene	5264190	5268941	5490353	4.3	training
41	1,2,4,5-Tetramethyl benzene	5896060	5906209	5860390	0.6	training
42	2-Methylnaphthalene	5892920	5926221	6227060	5.7	validation
43	Pentamethylbenzene	6534420	6563080	6518156	0.2	training
44	Hexamethylbenzene	7191670	7219706	7175923	0.2	training

45	Anthracene	7229600	7259330	7745706	7.1	training
46	Phenanthrene	7213270	7260270	7876215	9.2	training
47	1,1-Diphenylethane	7682020	7682565	7806947	1.6	test
48	Octadecane	11981110	12032547	11132050	7.1	validation
49	Triphenyl methane	10127620	10114360	10109522	0.2	training
50	1,3,5-Triphenylbenzene	12510990	12556476	12952432	3.5	training
51	Tetraphenylmethane	13254570	13194206	13257495	0.0	training
52	Formaldehyde	541650	467365	552833	2.1	training
53	Formic acid	303580	303694	304682	0.4	training
54	Ethanol	1370800	1337401	1368106	0.2	training
55	Dimethylether	1426440	1336314	1423737	0.2	training
56	Acetaldehyde	1167860	1124829	1156677	1.0	training
57	Ethylene oxide	1288990	1122547	1291982	0.2	validation
58	Ethylene glycol	1214210	1171872	1238387	2.0	training
59	Acetic acid	923570	960880	908525	1.6	training
60	Propan-2-ol	2007820	1994356	2013666	0.3	training
61	Acetone	1798440	1781915	1765225	1.8	training
62	Butanol	2472470	2653804	2666412	7.8	training
63	Butan-2-one	2441780	2438659	2468635	1.1	training
64	Furan	2123420	2012248	2154912	1.5	test
65	Butyric acid	2224950	2273868	2248238	1.0	test
66	Ethyl acetate	2278750	2273237	2269389	0.4	test
67	Pentan-1-ol	3325530	3307392	3334903	0.3	training
68	2-Methylbutan-2-ol	3289530	3307914	3290800	0.0	test

69	Cyclopentanol	3121220	3094945	3119990	0.0	training
70	Furfuryl alcohol	2694580	3546876	2680381	0.5	training
71	Hexan-1-ol	3977170	3963991	3990608	0.3	training
72	Cyclohexanol	3764560	3751633	3765919	0.0	training
73	Heptan-1-ol	4637550	4630048	4614638	0.5	training
74	Benzyl alcohol	3804960	3770707	3808933	0.1	training
75	(2R,3S)-butane-1,2,3,4-tetraol	2240470	2172446	2233939	0.3	training
76	Succinic acid	1616310	1741722	1625473	0.6	training
77	Maleic acid	1500300	1527654	1498284	0.1	validation
78	Fumaric acid	1476100	1528134	1498284	1.5	test
79	Phenol	3135370	3114266	3086370	1.6	training
80	Galactitol	3212360	3163713	3216518	0.1	training
81	Mannitol	3220860	3163380	3216518	0.1	training
82	alpha-D-Galactose	2942570	2953493	2937417	0.2	validation
83	L-Sorbose	2952820	2953141	3216518	8.9	training
84	Benzoic acid	3350440	3393838	3384131	1.0	validation
85	Orthohydroxybenzoic acid	3158140	3238540	3169339	0.4	training
86	Phthalic anhydride	3439410	3469838	3759117	9.3	test
87	Phthalic acid	3419530	3518391	3334262	2.5	training
88	Diphenyl ether	6293850	6201233	6287329	0.1	training
89	Beta-lactose	6013420	5848889	6054024	0.7	validation
90	Sucrose	6033090	5845699	6048526	0.3	training
91	alpha-Lactose	6070860	5839574	6054024	0.3	training
92	beta -Maltose	6091000	5836308	6054024	0.6	training

93	1-Hexadecanol	10532980	10549810	10520107	0.1	training
94	Palmitic acid	10089060	10176611	10101933	0.1	training
95	Oxalic acid	370950	427871	372288	0.4	training
96	Cyanoguanidine	1481960	1460350	1480928	0.1	training
97	Melamine	2127400	2207973	2127928	0.0	validation
98	Adenine	2946840	2948763	2947239	0.0	test
99	2-Cyanopyridine	3251510	3145629	3234102	0.5	training
100	Diphenyl amine	6553430	6514554	6551586	0.0	training
101	Urea	693580	772555	682846	1.5	training
102	Ammonium carbonate	673570	881063	666072	1.1	test
103	Aminoethanoic acid	1055280	1116979	1048660	0.6	training
104	Alanine	1697540	1773538	1694221	0.2	training
105	L-Asparagine	1751920	1898256	1751920	0.0	training
106	Alloxan	1056120	1416361	1056956	0.1	validation
107	Creatinine	2448730	2462233	2446040	0.1	training
108	Creatine	2453170	2519567	2455860	0.1	training
109	Allantoin	1916760	2088837	1917986	0.1	training
110	D-glutamic acid	2403600	2554679	2433764	1.3	test
111	Hypoxanthine	2606950	2637700	2609441	0.1	training
112	Xanthine	2366340	2474488	2364262	0.1	training
113	Uric acid	2294870	2316469	2294051	0.0	training
114	Guanine	2697010	2785235	2698042	0.0	training
115	Hippuric acid	4398530	4485927	4397498	0.0	training
116	Ethanethiol	2139620	2099638	2132734	0.3	training

117	Dimethyl sulphide	2150970	2099466	2140268	0.5	training
118	Propane-1-thiol	2802640	2756220	2785110	0.6	test
119	Methyl ethyl sulphide	2802940	2756216	2799549	0.1	training
120	Butane-1-thiol	3448920	3412950	3431040	0.5	training
121	2-Methyl propane-1-thiol	3444480	3412995	3407056	1.1	training
122	Diethyl sulphide	3456540	3412874	3223316	6.7	training
123	Methyl propyl sulfide	3452810	3412911	3484114	0.9	training
124	Thiophene	2850290	2774243	2849590	0.0	training
125	Ethyl butyl disulphide	4063330	4018875	4585135	12.8	training
126	Pentane-1-thiol	4101510	4079088	4099531	0.0	training
127	2-Methyl thiopene	3402110	3433230	3380558	0.6	training
128	3-Methyl thiopene	3403780	3433195	3553001	4.4	training
129	Thiophenol	3921690	3875538	4051331	3.3	training
130	Dipropyl disulphide	5370870	5341656	4601209	14.3	training
131	3-3-Dithiodipropionic acid	4176460	4269027	4176902	0.0	training
132	Cysteine	2292610	2379938	2273248	0.8	validation
133	Cystine	4425660	4120550	4424628	0.0	training

6.10 Refractive Indices of Organic Compounds

6.10.1 The QSPR model*

A summary of the results obtained using the procedure is as follows:

- Database collection from literature.

* The manuscript has been published in *Journal of Chemical and Engineering Data* 59, 1930-1943, 2014.

- Drawing and optimization of chemical structures according to the method mentioned.
- Molecular descriptor calculation using Dragon software [21].
- After preliminary studies 2328 molecular descriptors are retained.
- Training, validation, test sets allocations using the K-means clustering technique.
- Selection of most effective 58 molecular descriptors using the SS algorithm as explained in section 3.1.2.
- Selection of the most efficient molecular descriptors and development of the final model using the GFA method as explained in section 3.1.1.
- Final correlation is a multi-linear model with 20 parameters.
- The Hat approach is applied to check for outliers, and also to determine whether the newly developed correlation is statistically correct and valid.
- Elemental analysis of the results.

The model which is derived to estimate the refractive index data has a best twenty-parameter correlation equation with a total $R^2 = 0.892$ is as follows:

6.15

$$RI = \sum_{i=1}^{20} a_i X_i + b$$

where $b=0.690850214$ and a_i and X_i values are presented in the published article. The R^2 and average absolute relative deviation of the new correlation in the testing phase are 0.897 and 0.9 %, respectively.

The results obtained demonstrate that good agreement exists between the prediction of equation 6.15 and refractive index literature data. As previously mentioned, in proposing a predictive model or correlation, outlier detection plays a significant role to assess group or groups of data which may differ from the bulk of the data present in a dataset [440-442]. Therefore, to check whether the model is statistically correct and valid; the Williams plot has been sketched for the results obtained. The results of the refractive index predictive correlation illustrate that only a few of the data points may be considered as outlier. It may be possible to eliminate these probable outliers from the correlation results and propose more accurate ones; nevertheless, our aim, herein, has been to study the ability of all of the investigated models to estimate the entire range of refractive index values from a dataset in literature.

An elemental analysis was undertaken to determine the validity and accuracy of the model for various atomic elements contained in the molecules in the databank. The results indicate that there is an acceptable agreement between the literature data and the predicted values.

An advantage of the correlation proposed is the application of the most extensive dataset for both development and testing of the refractive index model. There are a relatively few previous studies in the literature which tackle the handling of very large datasets for the sake of model derivation with the aid of a QSPR strategy. Difficulties in selecting the proper mathematical-based technique and optimization strategy for treating the molecular descriptors to evaluate the refractive index may be one of the main reasons for a scarcity of methods for handling large datasets. Apart from considerations with regard to the computational times associated with the model development, there are several other drawbacks associated with model development. The procedure of selecting convenient parameters from a descriptor pool is complex owing to the large number of compounds as well as their associated descriptors. In contrast to group contribution models which merely

contain functional fragments, determination of some of the model parameters (descriptors) are not easy and specialized computational software is required.

6.10.2 The GC Model*

In order to obtain an accurate and reliable correlation, the collection of nearly 200 chemical substructures prepared as previously discussed, were introduced into the sequential search mathematical algorithm. Furthermore, in order to obtain the optimal correlation in terms of both the number of chemical substructures and accuracy, a threshold value of 0.01 was considered for the reduction in the average absolute relative deviations (AARD) as a stopping criterion. It means that when the improvement of the model AARD% was less than 0.01, the SS algorithm was automatically stopped and reported the final model. The optimal model was obtained using 80 chemical substructures. The best model derived to predict refractive index data for an eighty-chemical sub-structure correlation equation, with a total $R^2 = 0.888$ is as follows:

6.16

$$RI = \sum_{i=1}^{80} n_i RI_i + RI_0$$

where RI_0 , RI_i and n_i are the intercept of the equation, the contribution of the i th chemical substructure to the refractive index, and the number of occurrences of the i th chemical substructure in every chemical structure of pure compounds, respectively. The subset of 80 chemical substructures and their contribution to the refractive index are presented as supplementary file in the supplementary DVD.

* The manuscript has been *submitted to Fluid Phase Equilibria*.

The R^2 and average absolute relative deviation of the new correlation in the testing phase are reported as 0.898 and 0.83, respectively. The results demonstrate that good agreement exists between the prediction of equation 6.16 and refractive index literature data. The model results indicate that it can successfully predict the refractive index of pure organic compounds.

As previously mentioned, in proposing a predictive model or correlation, outlier detection plays a significant role to assess a group or groups of data which may differ from the bulk of the data present in a dataset [440-442]. Therefore, to check whether the model is statistically correct and valid; the Williams plot has been sketched for the results obtained. The results of the refractive index predictive correlation illustrate that only 197 of the data points may be considered as outlier.

6.11 Freezing Point Temperature of Organic compounds

6.11.1 The QSPR Model*

According to the procedure followed in this study, thirty descriptors are selected using the sequential search algorithm. The descriptors, as well as their descriptions are reported as supplementary file in the supplementary DVD. The model derived for the prediction of freezing point data has a best thirty-parameter correlation equation, with total $R^2 = 0.710$, as follows:

6.17

$$FP = \sum_{i=1}^{30} a_i X_i + b$$

where $b = 203.12942119$, a_i and X_i are presented in the published article (The manuscript has been *submitted to Journal of Thermal Analysis and Calorimetry*). To evaluate the accuracy of the proposed correlation, statistical error analysis, in which squared correlation coefficients (R^2),

* The manuscript has been *submitted to Journal of Thermal Analysis and Calorimetry*.

average absolute relative deviations (AARDs), standard deviation errors STD, and root mean square errors (RMSEs), and graphical error analysis, in which crossplot and error distribution is sketched, have been utilized that are presented in the published article. The R^2 and average absolute relative deviation of the new correlation in the testing phase are 0.723 and 12.5%, respectively.

The results obtained demonstrate that good agreement exists between the prediction of equation 6.17 and the literature data of freezing point.

As previously mentioned, in developing a predictive model or correlation, outlier detection plays a significant role to assess group or groups of data which may differ from the bulk of the data present in a dataset [441-443]. The results of the freezing point predictive correlation illustrate that there are only 220 of data points whose literature values may be erroneous.

An elemental analysis on the databank was performed. According to the results, the model can predict the freezing points of the 2,491 hydrocarbons with an AARD % of 13.53 %. There are 18 hydrocarbons which are outliers in the applicability domain of the implemented model.

As demonstrated, the AARD% of the model from literature data for nitrogen-containing compounds is 12.79 % and there are 136 outlier data points for nitrogen-containing compounds. The AARD% of the model results from literature data for oxygen-containing compounds is 12.11 %. There are 129 oxygen compounds which are outliers in the applicability domain of the implemented model. There are just 6 phosphorous-containing compounds within the dataset for which the model gives an AARD% of 17.05%. Sulfur-containing compounds are another class of compounds for which the model shows an AARD% of 13.74%. It should be noted that there are 28 outlier points for sulfur-containing compounds. Fluorine-containing compounds are one of the important classes of compounds whose freezing points are predicted by the model with an AARD% of 14.94 % and there are 12 outlier data points. The model predicts the freezing points

of chlorine-containing compounds better than fluorine-containing compounds in terms of AARD% (12.96 % vs. 14.94 %) and worse in term of outlier data (16 vs. 12). Based on our fragmental analysis, the freezing points of bromine-containing compounds are successfully estimated by the model. The model AARD% for this class of compounds is 12.55 % which is less than those of halogen-containing compounds mentioned above. Moreover, there are 7 bromine-containing compounds which are outliers in the applicability domain of the model implemented, which is less than for fluorine- and chlorine-containing compounds. Iodine-containing compounds are another class of halogen compounds whose freezing points are successfully predicted by the model. The model shows an AARD% of 12.7 % with only 4 outlier data points which is the minimum outlier number among all the lighter halogen compounds. According to the obtained results, many of the outliers for the model are for nitrogen- and oxygen-containing compounds. This may be the major cause for the high deviation in the predictions of the freezing points for these nitrogen- and oxygen-containing compounds. There is therefore high probability that the literature data maybe in error.

6.11.2 The GC Model*

In order to develop an efficient and reliable group-contribution model, the collection of nearly 300 chemical substructures prepared as described in the previous section, was introduced into the SS algorithm. Moreover, in developing the optimal model, in terms of both accuracy and the number of chemical substructures, a threshold value of 0.01 in terms of improvement was considered for the average absolute relative deviations (*AARD*) as a stopping criterion. According to the aforementioned procedure, 112 chemical substructures were finally selected using the sequential

* The manuscript has been *submitted to Fluid Phase Equilibria*.

search algorithm. The chemical substructures as well as their contribution values are reported as supplementary file in the supplementary DVD.

As mentioned above, the group-contribution model derived to predict freezing point data has a correlation equation (best 112 chemical structures) with total $R^2 = 0.735$ as follows:

6.18

$$FP = \sum_{i=1}^{112} g_i FP_i + FP_0$$

where FP_0 , FP_i and g_i are the intercept of the equation, the contribution of the i th chemical substructure to the freezing point, and the number of occurrences of the i th chemical substructure in every chemical structure of pure compounds, respectively.

The R^2 of the newly developed group-contribution model for the training, validation and testing phase are reported as 0.736, 0.705 and 0.756, respectively. In addition, the total average absolute relative deviation is 10.76 %. The results demonstrate that a good agreement exists between the predictions of equation 6.18 and the literature data for freezing points. The model results indicate that it can successfully predict the freezing point of pure organic compounds.

Outlier detection (or diagnostics) is of importance when developing and proposing mathematical models [440, 441]. Detection of outliers easily indicates individual datum (or groups of data) that may differ from the bulk of the data present in a databank [440-442, 444]. Consequently, there is indeed a necessity to evaluate the available experimental/literature data for freezing point, since uncertainties in the experimental data used in the model development would affect the estimation capability of the proposed group-contribution model. Hence, the Leverage Value Statistics technique [444, 445] was applied to determine the reliability, accuracy, and consistency of

experimental data. The graphical detection of suspect data or outliers is undertaken through sketching the Williams plot on the basis of the calculated H values [440, 441]. A detailed discussion of computational procedure and equations for this technique can be found elsewhere [440, 441]. As already mentioned, the compounds in our database are composed of Carbon, Nitrogen, Oxygen, Phosphorus, Sulfur, Fluorine, Chlorine, Bromine, Iodine, Boron, Aluminum, Silicon, Iron, Germanium, Arsenic, Selenium, Cadmium, Tin, Antimony, Tellurium, Mercury, Lead and Bismuth. In order to evaluate the performance of the model with regard to groups and elemental constituents of the molecules, an elemental analysis on the database was conducted. Consequently, the model predicts the freezing points of 2,460 hydrocarbons with an *AARD* of 10.27 %. There are 26 hydrocarbons which are outliers in the applicability domain of the model implemented. According to the results, the accuracy of the predicted freezing points for higher atom numbers is better than for lower values. The *AARD* of the model from 5700 literature data is 10.89% and there are 140 outlier data points for nitrogen-containing compounds. The *AARD* of the model results from literature data for 9777 oxygen-containing compounds is 10.52 %. There are 146 oxygen-containing compound data values which are outliers in the applicability domain of the model implemented. Sulfur-containing compounds are another class of compounds for which the model shows an *AARD* of 12.78 %. It should be noted that there are 38 outlier points for sulfur-containing compounds.

Fluorine-containing compounds have freezing points which are predicted by the model with an *AARD* of 11.05 % and there are 10 outlier data points. For chlorine-containing compounds, the *AARD* of the model from literature data is 10.14% and there are 23 data point outliers.

Chapter 7 Conclusions

The liquid thermal conductivity of pure organic compounds was the first property investigated in this study and three models were developed for its prediction. Three correlations based on the gene expression programming, the QSPR, and the group contribution were developed. All the models were evaluated using a comprehensive database of experimental liquid thermal conductivities for approximately 1600 pure organic compounds. According to Table 2.1 all the three models are more comprehensive and more accurate than the previous models. Among the three models proposed, the first one, namely the general model which correlate the liquid thermal conductivity to other thermophysical properties is simpler in shape and less accurate than the other two (AARD: 9% vs. 7.4% for QSPR and 7.1% for group contribution). The QSPR is more predictive than the other two and has moderate accuracy among them. Several statistical techniques were used to evaluate its predictive capability. However, the most accurate one is the third model, namely, the group contribution model which shows an AARD of 7.1% from experimental data. It is worth it mentioning that the applicability domain of all the models in terms of temperature and liquid thermal conductivity is satisfactory wide.

In the next step, an extensive database for viscosity of ionic liquids at standard pressure of 1 atm including 1672 experimental data points for 443 ionic liquids was collected from more than literature sources. Using the database, a group contribution model was developed for the estimation of viscosity of ionic liquids. The model shows an AARD% of 6.8. According to Table 2.2, the model is more comprehensive and more accurate than the previous model.

As next part of this study, a comprehensive database of thermal decomposition temperature data for 586 ionic liquids was initially collected from literature sources. Implementing the database, a

QSPR was developed. The models shows 5.2% AARD from the corresponding experimental data. The database was updated to 613 ionic liquids by collecting more experimental data from literature sources and then was used to develop a group contribution model. The model shows an AARD% of 4.5% from the corresponding experimental data. According to Table 2.3, both the QSPR and GC models are more comprehensive and more accurate than previous models.

Electrical conductivity of ionic liquids was the next property investigated in this study. An extensive data set of 1077 electrical conductivity data for 54 ionic liquids was initially collected from literature. Then, two independent non-linear models were developed; a QSPR and a group contribution model. In terms of accuracy, the QSPR surpass the group contribution model (AARD: 1.9% vs. 3.3%). However, in terms of simplicity of input parameters, the group contribution model surpass the QSPR.

In the next part of study, an enormous database of normal boiling point temperature of pure organic compounds was successfully handled and employed to develop a very comprehensive QSPR. One of most novel part of this study is the handling more than 17500 chemical structures which is the largest ever reported database in open literature for which a QSPR has been developed. The model shows promising AARD% of 3.6% from the corresponding literature data. According to Table 2.5, the model is significantly more accurate and more comprehensive than the previous models and can be regarded as a step forward in the field of study.

In the next stage of this thesis, the sublimation and vaporization enthalpies at the standard temperature of 298.15 K were investigated. First, a database of sublimation enthalpies at the standard temperature of 298.15 K for 1269 pure organic compounds was collected from literature. Employing the database, a group contribution model was developed. The model shows an AARD

of 6.4% from the corresponding experimental data. According to Table 2.6, the model is the more comprehensive and more accurate than the previous models.

Also, an extensive database of vaporization enthalpies of 2249 pure organic compounds at the standard temperature of 298.15 K was collected from literature sources. Employing the database, a group contribution model was developed which shows an AARD of 3.7% from the corresponding experimental data. As shown in Table 6.11, the model is more accurate than the previous models. In terms of the comprehensiveness, the database used in this study is the largest ever reported database for vaporization enthalpy at the standard temperature of 298.15 K.

Speed of sound of saturated liquids was the next property studied in this thesis. A collection of 1667 experimental data for 74 pure chemical compounds were extracted from the ThermoData Engine of National Institute of Standards and Technology [5]. The database was used to develop a group contribution model. The model shows an AARD% of 0.6 from the corresponding experimental data. It should be noted that there was no previous model which directly estimate the saturated liquids speed of sound.

One of the important properties for which a group contribution model was developed in this study is the standard molar chemical exergy of pure organic compounds. The model shows an AARD% of 1.6% from the corresponding experimental data for 133 pure organic compounds. It should be mentioned that the database used to evaluate the model is the most comprehensive database for the standard molar chemical exergy of pure organic compounds. The performance of the model was compared with one of the previous models. As shown in Table 6.13, the model shows significantly better AARD than one of the previous accurate models (1.6% vs. 3%). It is worth it

mentioning that in terms of the simplicity of calculation of input parameters, the proposed model surpasses the previous models.

Again another large database of refractive indices of approximately 12000 pure organic compounds was initially used to develop a QSPR with slightly new strategy than the one used for normal boiling point temperatures. A combination of sequential search algorithm and genetic function approximation was successfully used to develop an accurate model for the estimation of refractive index. The model shows an AARD of 0.9% from the corresponding literature values. In the next, step, a group contribution model was developed. The model shows an AARD of 0.83%. As concluded, the group contribution model shows better results in terms of accuracy. As mentioned earlier in section 6.10, both the QSPR and group contribution are more accurate and more comprehensive compared with previous models.

Freezing point temperatures of pure organic compounds was the most recent studied property in this thesis. One more time, a large database of nearly 16500 pure organic compounds was used to develop a QSPR and a group contribution model. A comparison between two models demonstrate that the group contribution model performance is better than the QSPR model (AARD: 10.78% vs. 12.6%).

In this thesis several computational techniques were successfully used to model various thermophysical properties particularly large databases. The gene expression programming is a promising technique in developing empirical and semi-empirical models in which a thermophysical property is correlated with other thermophysical properties. This technique is in its initial stages of development and it should be improved in terms of speed.

In terms of the computational methods used in developing QSPRs, it is worth it pointing that the genetic algorithm-based subset variable selection and model development methods such as genetic function approximation show the best results among various algorithms. The algorithms are available in several commercial software packages such as Materials Studio, Sarchitect, and QSARINS which are capable of handling just small to moderate size databases (less than for instance 2000 molecules with less than 2000 molecular descriptors). The software packages mainly suffers from their programming bases which lets use of the maximum 3.2 GB of RAM (32-bit programming base). This issue may be resolved by development of more commercial software packages which enable unlimited RAM usage (64-bit programing base). Another issue which seriously confines the computational speeds of the currently available commercial software packages is that they are capable of using just a single processor. Nowadays, most of desktops and notebooks have multi-core processors. Multi-thread-based programming's can significantly improve the speed this type of computations.

Chapter 8 Recommendations for Future works

- 1- The QSPR technique, the group contribution approach, and the gene expression programming technique can be applied for all thermophysical properties.
- 2- The abovementioned methods needs to be extended for binary and ternary mixtures. Since there is no basic mixing rule for molecular descriptors components of binary and ternary mixtures, novel planes should be devised.
- 3- The much descriptive chemical structure-based parameters such as new chemical substructures/segments in the GC method or new molecular descriptors in QSPR method can be suggested to enhance the accuracy of the currently available methods for the estimation of various physical properties of compounds.
- 4- Almost the majority of subset variable selection techniques that are currently in use in QSPR/GC approach are based on linear correlations. Non-linear models bring about more accurate models.
- 5- Outlier detection is a crucial step in developing QSPR/GC methods when particularly dealing with large databases of compounds. Although the leverage approach was used in this study for outlier detection; more accurate outlier detection techniques are needed.

References

- [1] C. Ferreira, Gene expression programming: a new adaptive algorithm for solving problems, *Complex Syst.*, 13 (2001) 87–129.
- [2] C. Ferreira, *Gene Expression Programming*, 2nd ed ed., Springer-Verlag, The Netherlands, 2006.
- [3] DIPPR-801, Project 801, Evaluated Process Design Data, Public Release Documentation, Design Institute for Physical Properties (DIPPR), American Institute of Chemical Engineers (AIChE), 2006
- [4] DIPPR-801, Evaluated process design data, public release documentation, American Institute of Chemical Engineers (AIChE), 2006.
- [5] M. Frenkel, R.D. Chirico, V. Diky, C. Muzny, E.W. Lemmon, X. Yan, Q. Dong, NIST Standard Reference Database 103b: ThermoData Engine (TDE), Version 7.0, National Institute of Standards and Technology, Gaithersburg, <http://www.nist.gov/srd/nist103b.cfm>, (2012).
- [6] D. Mackay, W. Ying Shiu, K.-C. Ma, S.C. Lee, *Handbook of physical-chemical properties and environmental fate for organic chemicals*, 2nd ed., CRC/Taylor & Francis, Boca Raton, FL, 2006.
- [7] M. Frenkel, R.D. Chirico, V. Diky, C.D. Muzny, A.F. Kazakov, J.W. Magee, I.M. Abdulagatov, J.W. Kang, NIST ThermoData Engine, NIST Standard Reference Database 103b, Version 7.3 (<http://trc.nist.gov/tde.html>), (2012).
- [8] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing / Volume II: Appendices, References*, 2nd, Revised and Enlarged Edition ed., Wiley-VCH; Chichester: JohnWiley, Weinheim, Chichester, 2009.

- [9] J.M. Coulson, J.F. Richardson, Coulson & Richardson's chemical engineering, Butterworth-Heinemann, Oxford ; Boston, 1996.
- [10] J.H. Lienhard, A heat transfer textbook, 4th ed., Phlogiston Press, Cambridge, Mass., 2011.
- [11] D.K.H. Briggs, Thermal conductivity of liquids, *Ind. Eng. Chem.*, 49 (1957) 418-421.
- [12] J. David Raal, R.L. Rijdsijk, Measurement of alcohol thermal conductivities using a relative strain-compensated hot-wire method, *Journal of Chemical and Engineering Data*, 26 (1981) 351-359.
- [13] G.M. Mallan, M.S. Michaelian, F.J. Lockhart, Liquid thermal conductivities of organic compounds and petroleum fractions, *Journal of Chemical and Engineering Data*, 17 (1972) 412-415.
- [14] R.W. Powell, H. Groot, Use of thermal comparator method for thermal conductivity measurements on liquids: Values for three organic series: normal alcohols, acids and saturated hydrocarbons, *International Journal of Heat and Mass Transfer*, 15 (1972) 360-366.
- [15] R.C. Reid, J.M. Prausnitz, B.E. Poling, *The Properties of Gases and Liquids*, (1987).
- [16] J.H. Irving, J.G. Kirkwood, The Statistical Mechanical Theory of Transport Processes. IV. The Equations of Hydrodynamics, *The Journal of Chemical Physics*, 18 (1950) 817-829.
- [17] S.R.S. Sastri, K.K. Rao, A new temperature–thermal conductivity relationship for predicting saturated liquid thermal conductivity, *Chemical Engineering Journal*, 74 (1999) 161-169.
- [18] S.R.S. Sastri, K.K. Rao, Quick estimating for thermal conductivity, *Chem. Eng.*, 100 (1993) 106-107.
- [19] J.D. Birkett, Thermal conductivity, *Handbook of Chemical Property Estimation Methods*, (1982).

- [20] A. Missenard, *Conductivité Thermique des Solides, Liquides, Gaz et de Leurs Melanges*, Eyrolles Editions, (1965).
- [21] F. Gharagheizi, A. Eslamimanesh, A.H. Mohammadi, D. Richon, QSPR approach for determination of parachor of non-electrolyte organic compounds, *Chemical Engineering Science*, 66 (2011) 2959-2967.
- [22] B.E. Poling, J.M. Prausnitz, J.P. O'Connell, *The properties of gases and liquids*, 5th ed., McGraw-Hill, New York, 2001.
- [23] C. Baroncini, P. Di Filippo, G. Latini, M. Pacetti, THERMAL CONDUCTIVITY OF LIQUIDS: COMPARISON OF PREDICTED VALUES WITH EXPERIMENTAL RESULTS AT DIFFERENT TEMPERATURES, *High Temperatures - High Pressures*, 11 (1978) 581-586.
- [24] C. Baroncini, P. Di Filippo, G. Latini, M. Pacetti, An improved correlation for the calculation of liquid thermal conductivity, *International Journal of Thermophysics*, 1 (1980) 159-175.
- [25] C. Baroncini, P. Di Filippo, G. Latini, M. Pacetti, Correlations for Thermal Conductivity Evaluation of Paraffins, Cycloparaffins and Ketones between the Melting Point and the Critical Temperature, *CORRELAZIONI PER IL CALCOLO DELLA CONDUCIBILITA TERMICA DI PARAFFINE, CICLOPARAFFINE E CHETONI TRA LA TEMPERATURA DI FUSIONE E LA TEMPERATURA CRITICA.*, 35 (1981) 633-638.
- [26] C. Baroncini, P. Di. Filippo, M. Pacetti, G. Latini, Thermal conductivity estimation of refrigerant fluids, IN: *PROC. 15TH INT. CONGRESS OF REFRIGERATION ON PROGRESS IN REFRIGERATION SCIENCE AND TECHNOLOGY*, (VENICE, ITALY: SEP. 23-29, 2 , Venice, Italy, 15th Int. Congress of Refrigeration Organizing & Sci. Committee, 1980, p.279-284. (1980).

- [27] C. Baroncini, P.D. Filippo, G. Latini, Thermal conductivity estimation of the organic and inorganic refrigerants in the saturated liquid state, *International Journal of Refrigeration*, 6 (1983) 60-62.
- [28] G. Latini, C. Baroncini, EFFECT OF TEMPERATURE AND PRESSURE ON THERMAL CONDUCTIVITY OF ORGANIC LIQUIDS, *High Temperatures - High Pressures*, 15 (1982) 407-411.
- [29] G. Latini, C. Baroncini, P. Pierpaoli, LIQUIDS UNDER PRESSURE: AN ANALYSIS OF METHODS FOR THERMAL CONDUCTIVITY PREDICTION AND A GENERAL CORRELATION, *High Temperatures - High Pressures*, 19 (1987) 43-50.
- [30] S.R.S. Sastri, personal communication, Regional Research Laboratory, Bhubaneswar,, 1998.
- [31] M. Freemantle, *An Introduction to ionic liquids*, Royal Society of Chemistry, Cambridge, 2010.
- [32] G.W. Meindersma, M. Maase, A.B. De Haan, Ionic Liquids, in: *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, 2000.
- [33] R.D. Rogers, K.R. Seddon, Chemistry. Ionic liquids--solvents of the future?, *Science*, 302 (2003) 792-793.
- [34] D. Rooney, J. Jacquemin, R. Gardas, *Thermophysical Properties of Ionic Liquids*, 290 (2009) 185-212.
- [35] C. Reichardt, T. Welton, *Solvents and solvent effects in organic chemistry*, 4th, updated and enlarged ed. ed., Wiley-VCH, Weinheim, 2011.
- [36] A.P. Abbott, Application of Hole Theory to the Viscosity of Ionic and Molecular Liquids, *ChemPhysChem*, 5 (2004) 1242-1246.

- [37] R. Fürth, On the theory of the liquid state, *Mathematical Proceedings of the Cambridge Philosophical Society*, 37 (1941) 281-290.
- [38] R. Fürth, On the theory of the liquid state, *Mathematical Proceedings of the Cambridge Philosophical Society*, 37 (1941) 252-275.
- [39] 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 Equilibria*, 261 (2007) 434-443.
- [40] R. Bini, M. Malvaldi, W.R. Pitner, C. Chiappe, QSPR correlation for conductivities and viscosities of low-temperature melting ionic liquids, *Journal of Physical Organic Chemistry*, 21 (2008) 622-629.
- [41] R.L. Gardas, J.A.P. Coutinho, A group contribution method for viscosity estimation of ionic liquids, *Fluid Phase Equilibria*, 266 (2008) 195-201.
- [42] M.H. Ghatee, M. Zare, A.R. Zolghadr, F. Moosavi, Temperature dependence of viscosity and relation with the surface tension of ionic liquids, *Fluid Phase Equilibria*, 291 (2010) 188-194.
- [43] M.H. Ghatee, M. Zare, Power-law behavior in the viscosity of ionic liquids: Existing a similarity in the power law and a new proposed viscosity equation, *Fluid Phase Equilibria*, 311 (2011) 76-82.
- [44] C. Han, G. Yu, L. Wen, D. Zhao, C. Asumana, X. Chen, Data and QSPR study for viscosity of imidazolium-based ionic liquids, *Fluid Phase Equilibria*, 300 (2011) 95-104.
- [45] S.A. Mirkhani, F. Gharagheizi, Predictive Quantitative Structure–Property Relationship Model for the Estimation of Ionic Liquid Viscosity, *Industrial & Engineering Chemistry Research*, 51 (2012) 2470-2477.

- [46] R.L. Gardas, J.A.P. Coutinho, Group contribution methods for the prediction of thermophysical and transport properties of ionic liquids, *AIChE Journal*, 55 (2009) 1274-1290.
- [47] J.A. Lazzús, A group contribution method to predict the thermal decomposition temperature of ionic liquids, *Journal of Molecular Liquids*, 168 (2012) 87-93.
- [48] F. Yan, S. Xia, Q. Wang, P. Ma, Predicting the Decomposition Temperature of Ionic Liquids by the Quantitative Structure–Property Relationship Method Using a New Topological Index, *Journal of Chemical & Engineering Data*, 57 (2012) 805-810.
- [49] J.A.P. Coutinho, P.J. Carvalho, N.M.C. Oliveira, Predictive methods for the estimation of thermophysical properties of ionic liquids, *RSC Advances*, 2 (2012) 7322-7346.
- [50] J.M. Slattery, C. Daguenet, P.J. Dyson, T.J.S. Schubert, I. Krossing, How to predict the physical properties of ionic liquids: A volume-based approach, *Angewandte Chemie - International Edition*, 46 (2007) 5384-5388.
- [51] M.G. Bogdanov, B. Iliev, W. Kantlehner, The residual volume approach II: Simple prediction of ionic conductivity of ionic liquids, *Zeitschrift für Naturforschung - Section B Journal of Chemical Sciences*, 64 (2009) 756-764.
- [52] P. Eiden, S. Bulut, T. Köchner, C. Friedrich, T. Schubert, I. Krossing, In silico predictions of the temperature-dependent viscosities and electrical conductivities of functionalized and nonfunctionalized ionic liquids, *Journal of Physical Chemistry B*, 115 (2011) 300-309.
- [53] K. Tochigi, H. Yamamoto, Estimation of ionic conductivity and viscosity of ionic liquids using a QSPR model, *Journal of Physical Chemistry C*, 111 (2007) 15989-15994.
- [54] A.P. Abbott, Model for the conductivity of ionic liquids based on an infinite dilution of holes, *ChemPhysChem*, 6 (2005) 2502-2505.

- [55] H. Zhao, Z.C. Liang, F. Li, An improved model for the conductivity of room-temperature ionic liquids based on hole theory, *Journal of Molecular Liquids*, 149 (2009) 55-59.
- [56] P. Walden, Organic solvents and ionization media. III. Interior friction and its relation to conductivity, *Z. Phys. Chem.*, 55 (1906).
- [57] M. Galiński, A. Lewandowski, I. Stepniak, Ionic liquids as electrolytes, *Electrochimica Acta*, 51 (2006) 5567-5580.
- [58] C.H. Fisher, Boiling Point Gives Critical Temperatures, *Chem. Eng.*, 96 (1989) 157-158.
- [59] J. Rechsteiner, Boiling point, in: W.J. Lyman, W.F. Reehl, D.H. Rosenblatt (Eds.) *Handbook of chemical property estimation methods*, McGraw-Hill, New York, 1982.
- [60] C.M. White, Prediction of the boiling point, heat of vaporization, and vapor pressure at various temperatures for polycyclic aromatic hydrocarbons, *Journal of Chemical and Engineering Data*, 31 (1986) 198-203.
- [61] S.W. Benson, J.H. Buss, Additivity rules for the estimation of molecular properties. Thermodynamic properties, *Journal of Chemical Physics*, 29 (1958) 546-572.
- [62] J. Hine, S.J. Ehrenson, An empirical method for prediction of the boiling points of halomethanes, *Journal of Organic Chemistry*, 21 (1956) 819-820.
- [63] J.F. Ogilvie, M.A. Abu-Elgheit, Statistical correlation of molecular structure with boiling points of N-heterocyclic compounds: Multiple linear regression analysis, *Computers and Chemistry*, 5 (1981) 33-39.
- [64] P.J. Hansen, P.C. Jurs, Prediction of olefin boiling points from molecular structure, *Analytical Chemistry*, 59 (1987) 2322-2327.

- [65] W.Y. Lai, D.H. Chen, R.N. Maddox, Application of a nonlinear group-contribution model to the prediction of physical constants. 1. Predicting normal boiling points with molecular structure, *Industrial and Engineering Chemistry Research*, 26 (1987) 1072-1079.
- [66] B.F. Rordorf, Prediction of vapor pressures, boiling points and enthalpies of fusion for twenty-nine halogenated dibenzo-p-dioxins, *Thermochimica Acta*, 112 (1987) 117-122.
- [67] F.C. Smeeks, P.C. Jurs, Prediction of boiling points of alcohols from molecular structure, *Analytica Chimica Acta*, 233 (1990) 111-119.
- [68] A.T. Balaban, N. Joshi, L.B. Kier, L.H. Hall, Correlations between chemical structure and normal boiling points of halogenated alkanes C1-C4, *Journal of Chemical Information and Computer Science*, 32 (1992) 233-237.
- [69] A.T. Balaban, L.B. Kier, N. Joshi, Correlations between chemical structure and normal boiling points of acyclic ethers, peroxides, acetals, and their sulfur analogues, *Journal of Chemical Information and Computer Science*, 32 (1992) 237-244.
- [70] M. Satou, S. Yokoyama, Y. Sanada, Structural analysis and estimation of boiling point of hydrocarbons in a coal-derived liquid by a group contribution method, *Fuel*, 71 (1992) 565-574.
- [71] D.T. Stanton, L.M. Egolf, P.C. Jurs, M.G. Hicks, Computer-assisted prediction of normal boiling points of pyrans and pyrroles, *Journal of Chemical Information and Computer Science*, 32 (1992) 306-316.
- [72] P. Simamora, A.H. Miller, S.H. Yalkowsky, Melting point and normal boiling point correlations: Applications to rigid aromatic compounds, *Journal of Chemical Information and Computer Science*, 33 (1993) 437-440.

- [73] L.M. Egolf, M.D. Wessel, P.C. Jurs, Prediction of boiling points and critical temperatures of industrially important organic compounds from molecular structure, *Journal of Chemical Information and Computer Science*, 34 (1994) 947-956.
- [74] S.E. Stein, R.L. Brown, Estimation of normal boiling points from group contributions, *Journal of Chemical Information and Computer Science*, 34 (1994) 581-587.
- [75] S. Wang, G.W.A. Milne, G. Klopman, Graph theory and group contributions in the estimation of boiling points, *Journal of Chemical Information and Computer Science*, 34 (1994) 1242-1250.
- [76] J.F. Krzyzaniak, P.B. Myrdal, P. Simamora, S.H. Yalkowsky, Boiling point and melting point prediction for aliphatic, non-hydrogen-bonding compounds, *Industrial and Engineering Chemistry Research*, 34 (1995) 2530-2535.
- [77] I.N. Tsibanogiannis, N.S. Kalospiros, D.P. Tassios, Prediction of normal boiling point temperature of medium/high molecular weight compounds, *Industrial and Engineering Chemistry Research*, 34 (1995) 997-1002.
- [78] M.D. Wessel, P.C. Jurs, Prediction of normal boiling points for a diverse set of industrially important organic compounds from molecular structure, *Journal of Chemical Information and Computer Science*, 35 (1995) 841-850.
- [79] A.R. Katritzky, L. Mu, V.S. Lobanov, M. Karelson, Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics, *Journal of Physical Chemistry*, 100 (1996) 10400-10407.
- [80] A.R. Katritzky, V.S. Lobanov, M. Karelson, Normal boiling points for organic compounds: Correlation and prediction by a quantitative structure - property relationship, *Journal of Chemical Information and Computer Science*, 38 (1998) 28-41.

- [81] A.T. Balaban, D. Mills, S.C. Basak, Correlation between structure and normal boiling points of acyclic carbonyl compounds, *Journal of Chemical Information and Computer Science*, 39 (1999) 758-764.
- [82] G. St. Cholakov, W.A. Wakeham, R.P. Stateva, Estimation of normal boiling points of hydrocarbons from descriptors of molecular structure, *Fluid Phase Equilibria*, 163 (1999) 21-42.
- [83] A.J. Chalk, B. Beck, T. Clark, A quantum mechanical/neural net model for boiling points with error estimation, *Journal of Chemical Information and Computer Science*, 41 (2001) 457-462.
- [84] W. Cordes, J. Rarey, A new method for the estimation of the normal boiling point of non-electrolyte organic compounds, *Fluid Phase Equilibria*, 201 (2002) 409-433.
- [85] D. Ericksen, W.V. Wilding, J.L. Oscarson, R.L. Rowley, Use of the DIPPR database for development of QSPR correlations: Normal boiling point, *Journal of Chemical and Engineering Data*, 47 (2002) 1293-1302.
- [86] A.A. Toropov, A.P. Toropova, Modeling of acyclic carbonyl compounds normal boiling points by correlation weighting of nearest neighboring codes, *Journal of Molecular Structure: THEOCHEM*, 581 (2002) 11-15.
- [87] B. Ren, Atom-type-based AI topological descriptors: Application in structure-boiling point correlations of oxo organic compounds, *Journal of Chemical Information and Computer Science*, 43 (2003) 1121-1131.
- [88] Y. Nannoolal, J. Rarey, D. Ramjugernath, W. Cordes, Estimation of pure component properties: Part 1. Estimation of the normal boiling point of non-electrolyte organic compounds via group contributions and group interactions, *Fluid Phase Equilibria*, 226 (2004) 45-63.

- [89] T. Öberg, Boiling Points of Halogenated Aliphatic Compounds: A Quantitative Structure - Property Relationship for Prediction and Validation, *Journal of Chemical Information and Computer Science*, 44 (2004) 187-192.
- [90] R. Sanghvi, S.H. Yalkowsky, Estimation of the normal boiling point of organic compounds, *Industrial and Engineering Chemistry Research*, 45 (2006) 2856-2861.
- [91] W. Acree, J.S. Chickos, Phase Transition Enthalpy Measurements of Organic and Organometallic Compounds. Sublimation, Vaporization and Fusion Enthalpies From 1880 to 2010, *Journal of Physical and Chemical Reference Data*, 39 (2010) 043101.
- [92] F. Gharagheizi, M. Sattari, B. Tirandazi, Prediction of Crystal Lattice Energy Using Enthalpy of Sublimation: A Group Contribution-Based Model, *Industrial & Engineering Chemistry Research*, 50 (2011) 2482-2486.
- [93] A. Gavezzotti, Molecular packing and other structural properties of crystalline oxohydrocarbons, *Journal of Physical Chemistry*, 95 (1991) 8948-8955.
- [94] G.L. Perlovich, S.V. Rodionov, A. Bauer-Brandl, Thermodynamics of solubility, sublimation and solvation processes of parabens, *European Journal of Pharmaceutical Sciences*, 24 (2005) 25-33.
- [95] G.L. Perlovich, T.V. Volkova, A. Bauer-Brandl, Towards an understanding of the molecular mechanism of solvation of drug molecules: A thermodynamic approach by crystal lattice energy, sublimation, and solubility exemplified by paracetamol, acetanilide, and phenacetin, *Journal of Pharmaceutical Sciences*, 95 (2006) 2158-2169.
- [96] G.L. Perlovich, S.V. Blokhina, N.G. Manin, T.V. Volkova, V.V. Tkachev, Polymorphs and solvates of felodipine: Analysis of crystal structures and thermodynamic aspects of sublimation and solubility processes, *CrystEngComm*, 14 (2012) 8577-8588.

- [97] G.L. Perlovich, N.N. Strakhova, V.P. Kazachenko, T.V. Volkova, V.V. Tkachev, K.J. Schaper, O.A. Raevsky, Studying thermodynamic aspects of sublimation, solubility and solvation processes and crystal structure analysis of some sulfonamides, *International Journal of Pharmaceutics*, 334 (2007) 115-124.
- [98] G.L. Perlovich, N.N. Strakhova, V.P. Kazachenko, T.V. Volkova, V.V. Tkachev, K.J. Schaper, O.A. Raevsky, Sulfonamides as a subject to study molecular interactions in crystals and solutions: Sublimation, solubility, solvation, distribution and crystal structure, *International Journal of Pharmaceutics*, 349 (2008) 300-313.
- [99] G.L. Perlovich, T.V. Volkova, A. Bauer-Brandl, Thermodynamic study of sublimation, solubility, solvation, and distribution processes of atenolol and pindolol, *Molecular Pharmaceutics*, 4 (2007) 929-935.
- [100] K. Nakajoh, E. Shibata, T. Todoroki, A. Ohara, K. Nishizawa, T. Nakamura, Measurement of temperature dependence for the vapor pressures of twenty-six polychlorinated biphenyl congeners in commercial Kanechlor mixtures by the knudsen effusion method, *Environmental Toxicology and Chemistry*, 25 (2006) 327-336.
- [101] J.S. Chickos, A protocol for correcting experimental fusion enthalpies to 298.15 K and its application in indirect measurements of sublimation enthalpy at 298.15 K, *Thermochimica Acta*, 313 (1998) 19-26.
- [102] B.M. Rice, S.V. Pai, J. Hare, Predicting heats of formation of energetic materials using quantum mechanical calculations, *Combustion and Flame*, 118 (1999) 445-458.
- [103] P. Politzer, J.S. Murray, M.E. Grice, M. Desalvo, E. Miller, Calculation of heats of sublimation and solid phase heats of formation, *Molecular Physics*, 91 (1997) 923-928.

- [104] D. Mathieu, P. Simonetti, Evaluation of solid-state formation enthalpies for energetic materials and related compounds, *Thermochimica Acta*, 384 (2002) 369-375.
- [105] C.K. Kim, K.A. Lee, K.H. Hyun, H.J. Park, I.Y. Kwack, C.K. Kim, H.W. Lee, B.S.U. Lee, Prediction of physicochemical properties of organic molecules using van der waals surface electrostatic potentials, *Journal of Computational Chemistry*, 25 (2004) 2073-2079.
- [106] C. Ouvrard, J.B.O. Mitchell, Can we predict lattice energy from molecular structure?, *Acta Crystallographica Section B: Structural Science*, 59 (2003) 676-685.
- [107] P. Politzer, Y. Ma, P. Lane, M.C. Concha, Computational prediction of standard gas, liquid, and solid-phase heats of formation and heats of vaporization and sublimation, *International Journal of Quantum Chemistry*, 105 (2005) 341-347.
- [108] E.F.C. Byrd, B.M. Rice, Improved prediction of heats of formation of energetic materials using quantum mechanical calculations, *Journal of Physical Chemistry A*, 110 (2006) 1005-1013.
- [109] F. Gharagheizi, A new molecular-based model for prediction of enthalpy of sublimation of pure components, *Thermochimica Acta*, 469 (2008) 8-11.
- [110] 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.
- [111] D. Mathieu, Simple alternative to neural networks for predicting sublimation enthalpies from fragment contributions, *Industrial and Engineering Chemistry Research*, 51 (2012) 2814-2819.
- [112] M. Salahinejad, T.C. Le, D.A. Winkler, Capturing the Crystal: Prediction of Enthalpy of Sublimation, Crystal Lattice Energy, and Melting Points of Organic Compounds, *Journal of Chemical Information and Modeling*, 53 (2013) 223-229.
- [113] S.H. Fishtine, Reliable latent heats of vaporization, *Ind. Eng. Chem.*, 55 (1963).

- [114] S.H. Fishtine, Reliable latent heats of vaporization, *Ind. Eng. Chem.*, 55 (1963) 49-55.
- [115] S.H. Fishtine, Reliable latent heats of vaporization, *Ind. Eng. Chem.*, 55 (1963) 47-56.
- [116] I. Wadso, Heats of vaporization for a number of organic compounds at 25 °C, *Acta Chem. Scand.*, 20 (1966).
- [117] L. Zhao, N. Ni, S.H. Yalkowsky, A modification of Trouton's rule by simple molecular parameters for hydrocarbon compounds, *Industrial and Engineering Chemistry Research*, 38 (1999) 324-327.
- [118] J.D. Cox, G. Pilcher, *Thermochemistry of Organic and Organometallic Compounds*, (1970).
- [119] L. Riedel, Kritischer Koeffizient, Dichte des gesättigten Dampfes und Verdampfungswärme, *Chem. Ing. Tech.*, 26 (1954).
- [120] W. Wagner, New vapour pressure measurements for argon and nitrogen and a new method for establishing rational vapour pressure equations, *Cryogenics*, 13 (1973) 470-482.
- [121] J.T.R. Watson, Imperial College, London, 1977.
- [122] D. Ambrose, J. Walton, Vapor pressures up to their critical temperatures of normal alkanes and 1-alkanols, *Pure Appl. Chem.*, 61 (1989) 1395-1403.
- [123] K.S. Pitzer, D.Z. Lippmann, R.F. Curl Jr, C.M. Huggins, D.E. Petersen, The volumetric and thermodynamic properties of fluids. II. Compressibility factor, vapor pressure and entropy of vaporization, *Journal of the American Chemical Society*, 77 (1955) 3433-3440.
- [124] R.Y. Wang, J. Shi, A reference value method of the corresponding states principle for the prediction of latent heats of vaporization of pure liquids, *Thermochimica Acta*, 169 (1990) 239-246.
- [125] D.L. Morgan, R. Kobayashi, Extension of Pitzer CSP models for vapor pressures and heats of vaporization to long-chain hydrocarbons, *Fluid Phase Equilibria*, 94 (1994) 51-87.

- [126] N.H. Chen, Generalized correlation for latent heat of vaporization, *Journal of Chemical and Engineering Data*, 10 (1965) 207-210.
- [127] A. Vetere, New correlations for predicting vaporization enthalpies of pure compounds, *The Chemical Engineering Journal*, 17 (1979) 157-162.
- [128] A. Vetere, Methods to predict the vaporization enthalpies at the normal boiling temperature of pure compounds revisited, *Fluid Phase Equilibria*, 106 (1995) 1-10.
- [129] Z.Y. Liu, Estimation of heat vaporization of pure liquid at its normal boiling temperature, *Chemical Engineering Communications*, 184 (2001) 221-228.
- [130] B. Ren, Application of novel atom-type AI topological indices to QSPR studies of alkanes, *Computers and Chemistry*, 26 (2002) 357-369.
- [131] N.I. Zhokhova, V.A. Palyulin, I.I. Baskin, A.N. Zefirov, N.S. Zefirov, Fragment descriptors in the QSPR method: Their use for calculating the enthalpies of vaporization of organic substances, *Russian Journal of Physical Chemistry A*, 81 (2007) 9-12.
- [132] E.L. Krasnykh, Prediction of vaporization enthalpy based on modified Randić indices. Alkanes, *Journal of Structural Chemistry*, 49 (2008) 986-993.
- [133] S.T. Lin, J. Chang, S. Wang, W.A. Goddard Iii, S.I. Sandler, Prediction of vapor pressures and enthalpies of vaporization using a COSMO solvation model, *Journal of Physical Chemistry A*, 108 (2004) 7429-7439.
- [134] S. Wang, S.T. Lin, J. Chang, W.A. Goddard Iii, S.I. Sandler, Application of the COSMO-SAC-BP solvation model to predictions of normal boiling temperatures for environmentally significant substances, *Industrial and Engineering Chemistry Research*, 45 (2006) 5426-5434.
- [135] S.W. Benson, New methods for estimating the heats of formation, heat capacities, and entropies of liquids and gases, *Journal of Physical Chemistry A*, 103 (1999) 11481-11485.

- [136] J.S. Chickos, W.E. Acree, J.F. Liebman, Computational Thermochemistry: Prediction Estimation of Molecular Thermodynamics, (1998).
- [137] J.S. Chickos, D.G. Hesse, J.F. Liebman, Estimating vaporization enthalpies of organic compounds with single and multiple substitution, *Journal of Organic Chemistry*, 54 (1989) 5250-5256.
- [138] L. Constantinou, R. Gani, New group contribution method for estimating properties of pure compounds, *AIChE Journal*, 40 (1994) 1697-1709.
- [139] L. Constantinou, S.E. Prickett, M.L. Mavrovouniotis, Estimation of thermodynamic and physical properties of acyclic hydrocarbons using the ABC approach and conjugation operators, *Industrial and Engineering Chemistry Research*, 32 (1993) 1734-1746.
- [140] L. Constantinou, S.E. Prickett, M.L. Mavrovouniotis, Estimation of properties of acyclic organic compounds using conjugation operators, *Industrial and Engineering Chemistry Research*, 33 (1994) 395-402.
- [141] M. Ducros, J.F. Gruson, H. Sannier, Estimation des enthalpies de vaporisation des composés organiques liquides. Partie 1. Applications aux alcanes, cycloalcanes, alcènes, hydrocarbures benzeniques, alcools, alcanes thiols, chloro et bromoalcanes, nitriles, esters, acides et aldehydes, *Thermochimica Acta*, 36 (1980) 39-65.
- [142] M. Ducros, J.F. Gruson, H. Sannier, I. Velasco, Estimation des enthalpies de vaporisation des composés organiques liquides. Partie 2. Applications aux éthersoxydes, thioalcanes, cétones et amines, *Thermochimica Acta*, 44 (1981) 131-140.
- [143] M. Ducros, H. Sannier, Estimation des enthalpies de vaporisation des composés organiques liquides, partie 3. Application aux hydrocarbures insaturés, *Thermochimica Acta*, 54 (1982) 153-157.

- [144] R.F. Fedors, METHOD FOR ESTIMATING BOTH THE SOLUBILITY PARAMETERS AND MOLAR VOLUMES OF LIQUIDS, *Polymer Engineering and Science*, 14 (1974) 147-154.
- [145] J.P. Guthrie, K.F. Taylor, *Can. J. Chem.*, 61 (1983).
- [146] Z. Kolská, V. Růžička, R. Gani, Estimation of the enthalpy of vaporization and the entropy of vaporization for pure organic compounds at 298.15 K and at normal boiling temperature by a group contribution method, *Industrial and Engineering Chemistry Research*, 44 (2005) 8436-8454.
- [147] K.J. Laidler, A system of molecular thermochemistry for organic gases and liquids, *Can. J. Chem.*, 34 (1956).
- [148] E.G. Lovering, K.J. Laidler, *Can. J. Chem.*, 38 (1960).
- [149] E.G. Lovering, O.M. Nor, *Can. J. Chem.*, 40 (1962) 199-204.
- [150] J. Marrero, R. Gani, Group-contribution based estimation of pure component properties, *Fluid Phase Equilibria*, 183-184 (2001) 183-208.
- [151] J.F. Pankow, W.E. Asher, SIMPOL.1: A simple group contribution method for predicting vapor pressures and enthalpies of vaporization of multifunctional organic compounds, *Atmospheric Chemistry and Physics*, 8 (2008) 2773-2796.
- [152] G.N. Roganov, P.N. Pisarev, V.N. Emel'yanenko, S.P. Verevkin, Measurement and prediction of thermochemical properties. Improved benson-type increments for the estimation of enthalpies of vaporization and Standard Enthalpies of Formation of Aliphatic Alcohols, *Journal of Chemical and Engineering Data*, 50 (2005) 1114-1124.
- [153] S.P. Verevkin, Improved Benson increments for the estimation of standard enthalpies of formation and enthalpies of vaporization of alkyl ethers, acetals, ketals, and ortho esters, *Journal of Chemical and Engineering Data*, 47 (2002) 1071-1097.

- [154] R.C. Santos, J.o.P. Leal, A Review on Prediction Methods for Molar Enthalpies of Vaporization of Hydrocarbons: The ELBA Method as the Best Answer, *Journal of Physical and Chemical Reference Data*, 41 (2012) 043101.
- [155] F. Chen, A group contribution method for estimating heat of vaporization at 298 K, *Chemical Engineering Science*, 46 (1991) 1063-1068.
- [156] E. Stefanis, L. Constantinou, I. Tsvintzelis, C. Panayiotou, New group-contribution method for predicting temperature-dependent properties of pure organic compounds, *International Journal of Thermophysics*, 26 (2005) 1369-1388.
- [157] C.G. Screttas, M. Micha-Screttas, Some properties and trends of enthalpies of vaporization and of Trouton's ratios of organic compounds. Correlation of enthalpies of vaporization and of enthalpies of formation with normal boiling points, *Journal of Organic Chemistry*, 56 (1991) 1615-1622.
- [158] J.B. Greenshields, F.D. Rossini, Molecular structure and properties of hydrocarbons and related compounds, *Journal of Physical Chemistry*, 62 (1958) 271-280.
- [159] D. Hoshino, K. Nagahama, M. Hirata, Prediction of the entropy of vaporization at the normal boiling point by the group contribution method, *Industrial and Engineering Chemistry Fundamentals*, 22 (1983) 430-433.
- [160] V. Svoboda, P. Dočkalová, Extension of the group contribution method for the calculation of the heat of vaporization, *Fluid Phase Equilibria*, 54 (1990) 293-299.
- [161] P. Ma, X. Zhao, Modified group contribution method for predicting the entropy of vaporization at the normal boiling point, *Industrial and Engineering Chemistry Research*, 32 (1993) 3180-3183.

- [162] M. Klüppel, S. Schulz, P. Ulbig, UNIVAP - A group contribution method for the prediction of enthalpies of vaporization of pure substances, *Fluid Phase Equilibria*, 102 (1994) 1-15.
- [163] P. Basařová, V. Svoboda, Prediction of the enthalpy of vaporization by the group contribution method, *Fluid Phase Equilibria*, 105 (1995) 27-47.
- [164] D. Dalmazzone, A. Salmon, S. Guella, A second order group contribution method for the prediction of critical temperatures and enthalpies of vaporization of organic compounds, *Fluid Phase Equilibria*, 242 (2006) 29-42.
- [165] P. Li, Y.H. Liang, P.S. Ma, C. Zhu, Estimations of enthalpies of vaporization of pure compounds at different temperatures by a corresponding-states group-contribution method, *Fluid Phase Equilibria*, 137 (1997) 63-74.
- [166] Y. Nannoolal, J. Rarey, D. Ramjugernath, Estimation of pure component properties part 3. Estimation of the vapor pressure of non-electrolyte organic compounds via group contribution and group interactions, *Fluid Phase Equilibria*, 269 (2008) 117-133.
- [167] C.H. Tu, C.P. Liu, Group-contribution estimation of the enthalpy of vaporization of organic compounds, *Fluid Phase Equilibria*, 121 (1996) 45-65.
- [168] F. Gharagheizi, Determination of normal boiling vaporization enthalpy using a new molecular-based model, *Fluid Phase Equilibria*, 317 (2012) 43-51.
- [169] F. Gharagheizi, O. Babaie, S. Mazdeyasna, Prediction of Vaporization Enthalpy of Pure Compounds using a Group Contribution-Based Method, *Industrial & Engineering Chemistry Research*, 50 (2011) 6503-6507.
- [170] Q. Jia, Q. Wang, P. Ma, Prediction of the enthalpy of vaporization of organic compounds at their normal boiling point with the positional distributive contribution method, *Journal of Chemical and Engineering Data*, 55 (2010) 5614-5620.

- [171] W. Wenying, H. Jinyu, X. Wen, Group vector space method for estimating enthalpy of vaporization of organic compounds at the normal boiling point, *Journal of Chemical Information and Computer Sciences*, 44 (2004) 1436-1439.
- [172] A.R.H. Goodwin, J.P.M. Trusler, Speed of sound, in: K.N.M. A.R.H. Goodwin, W.A. Wakeham (Eds.) *Experimental Thermodynamics*, Elsevier, 2003, pp. 237-323.
- [173] M. Bobik, Thermodynamic quantities for liquid benzene 1. Sound velocities between 283 and 463 K and up to 62 MPa, *The Journal of Chemical Thermodynamics*, 10 (1978) 1137-1146.
- [174] M. Bobik, R. Niepmann, W. Marius, Thermodynamic quantities for liquid carbon tetrachloride 1. Sound velocities between 265 and 435 K and up to 62 MPa, *The Journal of Chemical Thermodynamics*, 11 (1979) 351-357.
- [175] A. Lainez, P. Gopal, J.A. Zollweg, W.B. Streett, Speed-of-sound measurements for liquid trichlorofluoromethane under pressure, *The Journal of Chemical Thermodynamics*, 21 (1989) 773-777.
- [176] A. Lainez, J.F. Miller, J.A. Zollweg, W.B. Streett, Volumetric and speed-of-sound measurements for liquid tetrachloromethane under pressure, *The Journal of Chemical Thermodynamics*, 19 (1987) 1251-1260.
- [177] R. Niepmann, Thermodynamic properties of propane and n-butane 2. Speeds of sound in the liquid up to 60 MPa, *The Journal of Chemical Thermodynamics*, 16 (1984) 851-860.
- [178] R. Niepmann, G.J. Esper, K.A. Riemann, Thermodynamic properties of chlorodifluoromethane and dichloromethane Speeds of sound in the liquid up to 60 MPa, *The Journal of Chemical Thermodynamics*, 19 (1987) 741-749.

- [179] T. Takagi, T. Sakura, H.J.R. Guedes, Speed of sound in liquid cyclic alkanes at temperatures between (283 and 343) K and pressures up to 20 MPa, *Journal of Chemical Thermodynamics*, 34 (2002) 1943-1957.
- [180] T. Takagi, K. Sawada, H. Urakawa, M. Ueda, I. Cibulka, Speeds of sound in dense liquid and vapor pressures for 1,1-difluoroethane, *Journal of Chemical and Engineering Data*, 49 (2004) 1652-1656.
- [181] R. Vedam, G. Holton, Specific volumes of water at high pressures, obtained from ultrasonic-propagation measurements, *J. Acoust. Soc. Am.*, 43 (1968) 108-116.
- [182] M. Khammar, J.M. Shaw, Speed of sound prediction in 1 - n-alcohol + n-alkane mixtures using a translated SAFT-VR-Mie equation of state, *Fluid Phase Equilibria*, 288 (2010) 145-154.
- [183] J. Bao, L. Zhao, Exergy analysis and parameter study on a novel auto-cascade Rankine cycle, *Energy*, 48 (2012) 539-547.
- [184] P. Ahmadi, M.A. Rosen, I. Dincer, Multi-objective exergy-based optimization of a polygeneration energy system using an evolutionary algorithm, *Energy*, 46 (2012) 21-31.
- [185] R. Atilgan, O. Turan, T. Altuntaş, H. Aydin, K. Synylo, Environmental impact assessment of a turboprop engine with the aid of exergy, *Energy*, 58 (2013) 664-671.
- [186] G.G. Dimopoulos, I.C. Stefanatos, N.M.P. Kakalis, Exergy analysis and optimisation of a steam methane pre-reforming system, *Energy*, 58 (2013) 17-27.
- [187] R. Hackl, S. Harvey, Applying exergy and total site analysis for targeting refrigeration shaft power in industrial clusters, *Energy*, 55 (2013) 5-14.
- [188] J.A. Quijera, A. García, M.G. Alriols, J. Labidi, Heat integration options based on pinch and exergy analyses of a thermosolar and heat pump in a fish tinning industrial process, *Energy*, 55 (2013) 23-37.

- [189] A.P. Simpson, C.F. Edwards, The utility of environmental exergy analysis for decision making in energy, *Energy*, 55 (2013) 742-751.
- [190] O.K. Singh, S.C. Kaushik, Reducing CO₂ emission and improving exergy based performance of natural gas fired combined cycle power plants by coupling Kalina cycle, *Energy*, 55 (2013) 1002-1013.
- [191] M. Takla, N.E. Kamfjord, H. Tveit, S. Kjelstrup, Energy and exergy analysis of the silicon production process, *Energy*, 58 (2013) 138-146.
- [192] Z. Utlu, O. Kincay, An assessment of a pulp and paper mill through energy and exergy analyses, *Energy*, 57 (2013) 565-573.
- [193] M. Voldsund, I.S. Ertesvåg, W. He, S. Kjelstrup, Exergy analysis of the oil and gas processing on a north sea oil platform a real production day, *Energy*, 55 (2013) 716-727.
- [194] L. Yan, B. He, X. Pei, X. Li, C. Wang, Energy and exergy analyses of a Zero emission coal system, *Energy*, 55 (2013) 1094-1103.
- [195] S. Zhu, K. Deng, S. Qu, Energy and exergy analyses of a bottoming rankine cycle for engine exhaust heat recovery, *Energy*, 58 (2013) 448-457.
- [196] I.b. Dinçer, M. Rosen, *Exergy : energy, environment, and sustainable development*, 2nd ed., Elsevier, Amsterdam ; Boston, 2013.
- [197] T.J. Kotas, *The exergy method of thermal plant analysis*, Reprint ed., Krieger Pub., Malabar, Fla., 1995.
- [198] J. Szargut, D.R. Morris, F.R. Steward, *Exergy analysis of thermal, chemical, and metallurgical processes*, Hemisphere, New York, 1988.
- [199] Z. Rant, Zur Bestimmung der spezifischen exergie von brennstoffen, *Allg. Wärmetechn.*, 10 (1961) 172-176.

- [200] J. Szargut, T. Styrylska, Approximate evaluation of the exergy of fuels, *Brennstoff Wärme Kraft*, 16 (1964) 589-596.
- [201] J.H. Shieh, L.T. Fan, Estimation of energy (enthalpy) and exergy (availability) contents in structurally complicated materials, *Energy Sources*, 6 (1982) 1-46.
- [202] V.S. Stepanov, Chemical energies and exergies of fuels, *Energy*, 20 (1995) 235-242.
- [203] S. Ikumi, C.D. Luo, C.Y. Wen, Methods of estimating entropies of coals and coal liquids, *Can J Chem Eng*, 60 (1982) 551-555.
- [204] S. Bilgen, K. Kaygusuz, The calculation of the chemical exergies of coal-based fuels by using the higher heating values, *Applied Energy*, 85 (2008) 776-785.
- [205] W. Eisermann, P. Johnson, W.L. Conger, Estimating thermodynamic properties of coal, char, tar and ash, *Fuel Processing Technology*, 3 (1980) 39-53.
- [206] G. Song, L. Shen, J. Xiao, Estimating specific chemical exergy of biomass from basic analysis data, *Industrial and Engineering Chemistry Research*, 50 (2011) 9758-9766.
- [207] F. Gharagheizi, M. Mehrpooya, Prediction of standard chemical exergy by a three descriptors QSPR model, *Energy Conversion and Management*, 48 (2007) 2453-2460.
- [208] G. Song, J. Xiao, H. Zhao, L. Shen, A unified correlation for estimating specific chemical exergy of solid and liquid fuels, *Energy*, 40 (2012) 164-173.
- [209] R.J. Samuels, Application of refractive index measurements to polymer analysis, *Journal of Applied Polymer Science*, 26 (1981) 1383-1412.
- [210] R. García-Domenech, J. de Julián-Ortiz, Prediction of indices of refraction and glass transition temperatures of linear polymers by using graph theoretical indices, *The Journal of Physical Chemistry B*, 106 (2002) 1501-1507.
- [211] I. Finar, *Organic Chemistry* (6th edn.), Longman, in, London, 1973.

- [212] D.W. Van Krevelen, K. Te Nijenhuis, Properties of polymers: their correlation with chemical structure; their numerical estimation and prediction from additive group contributions, Access Online via Elsevier, 2009.
- [213] H. Touba, G. Mansoori, A. Sarem, M, New analytic techniques for petroleum fluid characterization using molar refraction, in: SPE Western Regional Meeting, 1997.
- [214] R. Mehra, Application of refractive index mixing rules in binary systems of hexadecane and heptadecane withn-alkanols at different temperatures, Journal of Chemical Sciences, 115 (2003) 147-154.
- [215] A.R. Katritzky, S. Sild, M. Karelson, Correlation and prediction of the refractive indices of polymers by QSPR, Journal of chemical information and computer sciences, 38 (1998) 1171-1176.
- [216] H. Lorentz, Ueber die Beziehung zwischen der Fortpflanzungsgeschwindigkeit des Lichtes und der Körperdichte, Annalen der Physik, 245 (1880) 641-665.
- [217] T.P. Dale, J. Gladstone, On the influence of temperature on the refraction of light, Philosophical Transactions of the Royal Society of London, 148 (1858) 887-894.
- [218] A.I. Vogel, 369. Physical properties and chemical constitution. Part XXIII. Miscellaneous compounds. Investigation of the so-called co-ordinate or dative link in esters of oxy-acids and in nitro-paraffins by molecular refractivity determinations. atomic, structural, and group parachors and refractivities, Journal of the Chemical Society (Resumed), (1948) 1833-1855.
- [219] A.K. Agrawal, S.A. Jenekhe, Thin-film processing and optical properties of conjugated rigid-rod polyquinolines for nonlinear optical applications, Chemistry of materials, 4 (1992) 95-104.
- [220] C.-J. Yang, S.A. Jenekhe, Group contribution to molar refraction and refractive index of conjugated polymers, Chemistry of materials, 7 (1995) 1276-1285.

- [221] L.B. Kier, L.H. Hall, Molecular connectivity in structure-activity analysis, Research Studies Press Letchworth, 1986.
- [222] T. Brekke, O.M. Kvalheim, E. Sletten, Prediction of physical properties of hydrocarbon mixtures by partial-least-squares calibration of carbon-13 nuclear magnetic resonance data, *Analytica chimica acta*, 223 (1989) 123-134.
- [223] J.H. Friedman, Multivariate Adaptive Regression Splines, *The Annals of Statistics*, 19 (1991) 1-67.
- [224] 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., MIT Press, 1992.
- [225] D. Rogers, A.J. Hopfinger, Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships, *Journal of Chemical Information and Computer Science*, 34 (1994) 854-866.
- [226] S.M. Mousavisafavi, F. Gharagheizi, S.A. Mirkhani, J. Akbari, A predictive quantitative structure-property relationship for glass transition temperature of 1,3-dialkyl imidazolium ionic liquids - Part 2. The nonlinear approach, *Journal of Thermal Analysis and Calorimetry*, (2012) 1-10.
- [227] S.M. Mousavisafavi, F. Gharagheizi, S.A. Mirkhani, J. Akbari, A predictive quantitative structure-property relationship for glass transition temperature of 1,3-dialkyl imidazolium ionic liquids, *Journal of Thermal Analysis and Calorimetry*, (2012).
- [228] S.A. mirkhani, F. Gharagheizi, P. Ilani-Kashkouli, N. Farahani, Determination of the glass transition temperature of ionic liquids: A molecular approach, *Thermochimica Acta*, (2012).
- [229] F. Gharagheizi, A. Eslamimanesh, F. Farjood, A.H. Mohammadi, D. Richon, Solubility Parameters of Nonelectrolyte Organic Compounds: Determination Using Quantitative Structure-

Property Relationship Strategy, *Industrial & Engineering Chemistry Research*, 50 (2011) 11382-11395.

[230] F. Gharagheizi, Chemical Structure-Based Model for Estimation of the Upper Flammability Limit of Pure Compounds, *Energy & Fuels*, 24 (2010) 3867-3871.

[231] 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.

[232] F. Gharagheizi, Prediction of upper flammability limit percent of pure compounds from their molecular structures, *Journal of hazardous materials*, 167 (2009) 507-510.

[233] F. Gharagheizi, A QSPR model for estimation of lower flammability limit temperature of pure compounds based on molecular structure, *Journal of hazardous materials*, 169 (2009) 217-220.

[234] 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, *Chemical Engineering Science*, 66 (2011) 3039-3044.

[235] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least squares support vector machines, World Scientific, River Edge, NJ, 2002.

[236] 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, *Journal of Physical Chemistry B*, 109 (2005) 20565-20571.

[237] 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.

- [238] 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.
- [239] A. Eslamimanesh, F. Gharagheizi, A.H. Mohammadi, D. Richon, Phase equilibrium modeling of structure H clathrate hydrates of methane + water "insoluble" hydrocarbon promoter Using QSPR molecular approach, *Journal of Chemical and Engineering Data*, 56 (2011) 3775-3793.
- [240] C. Ferreira, *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*, (2006).
- [241] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, (2001).
- [242] J.P. Chiou, F.S. Wang, Hybrid method of evolutionary algorithms for static and dynamic optimization problems with application to a fed-batch fermentation process, *Computers and Chemical Engineering*, 23 (1999) 1277-1291.
- [243] R. Storn, K. Price, Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, *Journal of Global Optimization*, 11 (1997) 341-359.
- [244] K. Price, R. Storn, Differential evolution, *Dr. Dobb's Journal*, (1997) 18-24.
- [245] L. Davis, *Handbook of Genetic Algorithms*, (1991).
- [246] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, (1989).
- [247] H.P. Schwefel, *Numerical Optimization of Computer Models*, (1981).
- [248] J.H. Holland, *Adaptation in Natural and Artificial Systems*, (1975).
- [249] N.L. Cramer, A representation for the adaptive generation of simple sequential programs, *Int. Conf. Genetic Algorithms and their Applications*, (1985) 183-187.

- [250] J.R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection, (1992).
- [251] C. Ferreira, Gene expression programming: A new adaptive algorithm for solving problems, *Complex Systems*, 13 (2001) 87-129.
- [252] F. Gharagheizi, A. Eslamimanesh, M. Sattari, A.H. Mohammadi, D. Richon, Development of corresponding states model for estimation of the surface tension of chemical compounds, *AIChE Journal*, 59 (2013) 613-621.
- [253] F. Gharagheizi, QSPR analysis for intrinsic viscosity of polymer solutions by means of GA-MLR and RBFNN, *Computational Materials Science*, 40 (2007) 159-167.
- [254] C.L. Yaws, *The Yaws Handbook of Physical Properties for Hydrocarbons and Chemicals: Physical Properties for More Than 41,000 Organic and Inorganic Chemical Compounds : Coverage for C1 to C100 Organics and Ac to Zr Inorganics*, Gulf Publishing Company, 2005.
- [255] A.R.R.P. Almeida, M.A.R. Matos, M.J.S. Monte, V.M.F. Morais, Experimental and computational thermodynamic study of ortho-, meta-, and para-methylbenzamide, *Journal of Chemical Thermodynamics*, 47 (2012) 81-89.
- [256] A.R.R.P. Almeida, M.J.S. Monte, Thermodynamic study of benzamide, N-methylbenzamide, and N, N -dimethylbenzamide: Vapor pressures, phase diagrams, and hydrogen bond enthalpy, *Journal of Chemical and Engineering Data*, 55 (2010) 3507-3512.
- [257] A.R.R.P. Almeida, M.J.S. Monte, Vapor Pressures and Phase Diagrams of Two Methyl Esters of Substituted Benzoic Acids, *Journal of Chemical and Engineering Data*, 56 (2011) 4862-4867.
- [258] A.R.R.P. Almeida, M.J.S. Monte, Thermodynamic study of phase transitions of imidazoles and 1-methylimidazoles, *Journal of Chemical Thermodynamics*, 44 (2012) 163-168.

- [259] A.R.R.P. Almeida, M.J.S. Monte, The influence of the halogen size in the volatility and melting of methyl p-halobenzoic esters and of their parent acids, *Journal of Chemical Thermodynamics*, 57 (2013) 160-168.
- [260] R.D. Chirico, A.F. Kazakov, W.V. Steele, Thermodynamic properties of three-ring azaromatics. 1. Experimental results for phenazine and acridine, and mutual validation of experiments and computational methods, *Journal of Chemical Thermodynamics*, 42 (2010) 571-580.
- [261] J.C.S. Costa, C.F.R.A.C. Lima, M.A.A. Rocha, L.R. Gomes, L.M.N.B.F. Santos, Phase transition equilibrium of terthiophene isomers, *Journal of Chemical Thermodynamics*, 43 (2011) 133-139.
- [262] J.Z. Dávalos, A. Guerrero, R. Herrero, P. Jimenez, A. Chana, J.L.M. Abboud, C.F.R.A.C. Lima, L.M.N.B.F. Santos, A.F. Lago, Neutral, ion gas-phase energetics and structural properties of hydroxybenzophenones, *Journal of Organic Chemistry*, 75 (2010) 2564-2571.
- [263] J.Z. Dávalos, R. Herrero, A. Chana, A. Guerrero, P. Jiménez, J.M. Santiuste, Energetics and structural properties, in the gas phase, of trans -hydroxycinnamic acids, *Journal of Physical Chemistry A*, 116 (2012) 2261-2267.
- [264] V.N. Emel'Yanenko, E.N. Stepurko, S.P. Verevkin, G.N. Roganov, The thermodynamic properties of 1,4-dioxane-2,6-dione, *Russian Journal of Physical Chemistry A*, 85 (2011) 179-185.
- [265] V.N. Emel'yanenko, S.P. Verevkin, A.A. Pimerzin, The thermodynamic properties of DL- and L-lactides, *Russian Journal of Physical Chemistry A*, 83 (2009) 2013-2021.
- [266] V.N. Emel'Yanenko, S.P. Verevkin, R.V. Ralys, V.V. Turovtsev, V.Y. Orlov, Enthalpy of phase transitions of lactams, *Russian Journal of Physical Chemistry A*, 86 (2012) 1493-1499.

- [267] V.N. Emel'Yanenko, S.P. Verevkin, E.N. Stepurko, G.N. Roganov, M.K. Georgieva, Thermodynamic properties of glycolic acid and glycolide, *Russian Journal of Physical Chemistry A*, 84 (2010) 1301-1308.
- [268] A.I.M.C.L. Ferreira, M.A.V. Ribeiro Da Silva, Thermochemical study of three dibromophenol isomers, *Journal of Chemical Thermodynamics*, 43 (2011) 227-234.
- [269] J.M.S. Fonseca, O. Pfohl, R. Dohrn, Development and test of a new Knudsen effusion apparatus for the measurement of low vapour pressures, *Journal of Chemical Thermodynamics*, 43 (2011) 1942-1949.
- [270] J.M.S. Fonseca, L.M.N.B.F. Santos, M.J.S. Monte, Thermodynamic study of 4- n - alkyloxybenzoic acids, *Journal of Chemical and Engineering Data*, 55 (2010) 2238-2245.
- [271] V.L.S. Freitas, J.R.B. Gomes, L. Gales, A.M. Damas, M.D.M.C. Ribeiro Da Silva, Experimental and computational studies on the structural and thermodynamic properties of two sulfur heterocyclic keto compounds, *Journal of Chemical and Engineering Data*, 55 (2010) 5009-5017.
- [272] V.L.S. Freitas, J.R.B. Gomes, M.D.M.C. Ribeiro Da Silva, Experimental and computational thermochemical studies of 9-R-xanthene derivatives (ROH, COOH, CONH 2), *Journal of Chemical Thermodynamics*, 54 (2012) 108-117.
- [273] I.V. Garist, S.P. Verevkin, J.E. Bara, M.S. Hindman, S.P.O. Danielsen, Building Blocks for Ionic Liquids: Vapor Pressures and Vaporization Enthalpies of 1-(n-Alkyl)-benzimidazoles, *Journal of Chemical and Engineering Data*, 57 (2012) 1803-1809.
- [274] E.M. Gonçalves, C.E.S. Bernardes, H.P. Diogo, M.E. Minas Da Piedade, Energetics and structure of nicotinic acid (Niacin), *Journal of Physical Chemistry B*, 114 (2010) 5475-5485.

- [275] D. Hasty, T. Subramanian, T.C. Winter, J.S. Chickos, A.A. Samarov, A.V. Yermalayeu, S.P. Verevkin, Applications of correlation gas chromatography and transpiration studies for the evaluation of the vaporization and sublimation enthalpies of some perfluorinated hydrocarbons, *Journal of Chemical and Engineering Data*, 57 (2012) 2350-2359.
- [276] C.F.R.A.C. Lima, J.C.S. Costa, L.M.N.B.F. Santos, Thermodynamic insights on the structure and energetics of s-triphenyltriazine, *Journal of Physical Chemistry A*, 115 (2011) 9249-9258.
- [277] C.F.R.A.C. Lima, C.A.D. Sousa, J.E. Rodriguez-Borges, A. Melo, L.R. Gomes, J.N. Low, L.M.N.B.F. Santos, The role of aromatic interactions in the structure and energetics of benzyl ketones, *Physical Chemistry Chemical Physics*, 12 (2010) 11228-11237.
- [278] D. Lipkind, N. Rath, J.S. Chickos, V.A. Pozdeev, S.P. Verevkin, The vaporization enthalpies of 2- and 4-(N,N-dimethylamino)pyridine, 1,5-diazabicyclo[4.3.0]non-5-ene, 1,8-diazabicyclo[5.4.0]undec-7-ene, imidazo[1,2-a]pyridine and 1,2,4-triazolo[1,5-a]pyrimidine by correlation-gas chromatography, *Journal of Physical Chemistry B*, 115 (2011) 8785-8796.
- [279] A.I.M.C. Lobo Ferreira, M.A.V. Ribeiro Da Silva, Experimental and computational thermochemical study of the three monoiodophenol isomers, *Journal of Chemical and Engineering Data*, 56 (2011) 4881-4890.
- [280] A.I.M.C. Lobo Ferreira, M.A.V. Ribeiro da Silva, Experimental and computational study of the molecular energetics of the monoiodoanisole isomers, *Journal of Chemical Thermodynamics*, 48 (2012) 84-92.
- [281] M.A.R. Matos, C.C.S. Sousa, V.M.F. Morais, Thermochemistry of chromone- and coumarin-3-carboxylic acid, *Journal of Thermal Analysis and Calorimetry*, 100 (2010) 519-526.

- [282] R. Maxwell, J. Chickos, An examination of the thermodynamics of fusion, vaporization, and sublimation of ibuprofen and naproxen by correlation gas chromatography, *Journal of Pharmaceutical Sciences*, 101 (2012) 805-814.
- [283] M.S. Miranda, M.A.R. Matos, V.M.F. Morais, J.F. Liebman, Study of energetics and structure of 1,2,3-benzotriazin-4(3H)-one and its 1H and Enol tautomers, *Journal of Physical Chemistry B*, 115 (2011) 6616-6622.
- [284] M.S. Miranda, M.A.R. Matos, V.M.F. Morais, J.F. Liebman, 2,1,3-Benzothiadiazole: Study of its structure, energetics and aromaticity, *Journal of Chemical Thermodynamics*, 50 (2012) 30-36.
- [285] E.A. Miroshnichenko, T.S. Kon'Kova, Y.N. Matyushin, Y.O. Inozemtsev, Bond dissociation energies in nitramines, *Russian Chemical Bulletin*, 58 (2009) 2015-2019.
- [286] M.J.S. Monte, R. Notario, S.P. Pinto, A.I.M.C. Lobo Ferreira, M.D.M.C. Ribeiro Da Silva, Thermodynamic properties of fluoranthene: An experimental and computational study, *Journal of Chemical Thermodynamics*, 49 (2012) 159-164.
- [287] M.J.S. Monte, L.M.N.B.F. Santos, J.M.S. Fonseca, C.A.D. Sousa, Vapour pressures, enthalpies and entropies of sublimation of para substituted benzoic acids, *Journal of Thermal Analysis and Calorimetry*, 100 (2010) 465-474.
- [288] M.J.S. Monte, C.A.D. Sousa, Thermodynamic study on the sublimation of diphenyl and triphenyl substituted acetic and propanoic acids, *Journal of Thermal Analysis and Calorimetry*, 106 (2011) 913-920.
- [289] V.M.F. Morais, C.C.S. Sousa, M.A.R. Matos, Experimental and computational study of the energetics of methoxycoumarins, *Journal of Molecular Structure: THEOCHEM*, 946 (2010) 13-19.

- [290] A.G. Nazmutdinov, I.A. Nesterov, T.A. Nazmutdinov, T.N. Nesterova, S.V. Tarazanov, S.V. Vostrikov, L.L. Pashchenko, E.A. Miroshnichenko, S.P. Verevkin, Vapour pressures and enthalpies of vaporization of a series of the alkylbiphenyls, *Fluid Phase Equilibria*, 335 (2012) 88-98.
- [291] R. Notario, M.V. Roux, C. Foces-Foces, M.A.V. Ribeiro Da Silva, M.D.D.M.C. Ribeiro Da Silva, A.F.L.O.M. Santos, R. Guzmán-Mejía, E. Juaristi, Experimental and computational thermochemical study of N-benzylalanines, *Journal of Physical Chemistry B*, 115 (2011) 9401-9409.
- [292] G.L. Perlovich, A.M. Ryzhakov, N.N. Strakhova, V.P. Kazachenko, K.J. Schaper, O.A. Raevsky, Thermodynamic aspects of solubility, solvation and partitioning processes of some sulfonamides, *Journal of Chemical Thermodynamics*, 43 (2011) 683-689.
- [293] G.L. Perlovich, T.V. Volkova, A.N. Proshin, D.Y. Sergeev, C.T. Bui, L.N. Petrova, S.O. Bachurin, Synthesis, pharmacology, crystal properties, and quantitative solvation studies from a drug transport perspective for three new 1,2,4-thiadiazoles, *Journal of Pharmaceutical Sciences*, 99 (2010) 3754-3768.
- [294] R. Picciochi, H.P. Diogo, M.E. Minas Da Piedade, Thermochemistry of paracetamol, *Journal of Thermal Analysis and Calorimetry*, 100 (2010) 391-401.
- [295] R. Picciochi, H.P. Diogo, M.E. Minas Da Piedade, Thermodynamic characterization of three polymorphic forms of piracetam, *Journal of Pharmaceutical Sciences*, 100 (2011) 594-603.
- [296] S.V. Portnova, E.L. Krasnykh, S.P. Verevkin, Vapour pressure and enthalpy of vaporization of di-iso-propyl and di-tert-butyl esters of dicarboxylic acids, *Fluid Phase Equilibria*, 309 (2011) 114-120.

- [297] M.A.V. Ribeiro da Silva, L.M.P.F. Amaral, Standard molar enthalpies of formation of monochloroacetophenone isomers, *Journal of Chemical Thermodynamics*, 42 (2010) 1473-1477.
- [298] M.A.V. Ribeiro Da Silva, L.M.P.F. Amaral, Thermochemical study of some dichloroacetophenone isomers, *Journal of Chemical Thermodynamics*, 43 (2011) 255-261.
- [299] M.A.V. Ribeiro Da Silva, L.M.P.F. Amaral, Standard molar enthalpies of formation of 3'- and 4'-nitroacetophenones, *Journal of Chemical Thermodynamics*, 43 (2011) 876-881.
- [300] M.A.V. Ribeiro Da Silva, L.M.P.F. Amaral, P. Szterner, Experimental study on the thermochemistry of 2-thiouracil, 5-methyl-2-thiouracil and 6-methyl-2-thiouracil, *Journal of Chemical Thermodynamics*, 57 (2013) 380-386.
- [301] M.A.V. Ribeiro Da Silva, J.I.T.A. Cabral, Standard molar enthalpies of formation of 5- and 6-nitroindazole, *Journal of Thermal Analysis and Calorimetry*, 100 (2010) 457-464.
- [302] M.A.V. Ribeiro Da Silva, J.I.T.A. Cabral, Standard molar enthalpies of formation of three methyl-pyrazole derivatives, *Journal of Chemical Thermodynamics*, 47 (2012) 138-143.
- [303] A.F.L.O.M. Santos, M.A.V. Ribeiro da Silva, Experimental and computational thermochemistry of 1-phenylpyrrole and 1-(4-methylphenyl)pyrrole, *Journal of Chemical Thermodynamics*, 42 (2010) 734-741.
- [304] A.F.L.O.M. Santos, M.A.V. Ribeiro Da Silva, Diaminobenzenes: An experimental and computational study, *Journal of Physical Chemistry B*, 115 (2011) 4939-4948.
- [305] M.A.V. Ribeiro da Silva, M.J.S. Monte, A.I.M.C. Lobo Ferreira, J.A.S.A. Oliveira, A. Cimas, Experimental and Computational Thermodynamic Study of Three Monofluoronitrobenzene Isomers, *Journal of Physical Chemistry B*, 114 (2010) 7909-7919.

- [306] M.A.V. Ribeiro Da Silva, M.J.S. Monte, I.M. Rocha, A. Cimas, Energetic study applied to the knowledge of the structural and electronic properties of monofluorobenzonitriles, *Journal of Organic Chemistry*, 77 (2012) 4312-4322.
- [307] M.A.V. Ribeiro Da Silva, A.F.L.O.M. Santos, Thermochemical properties of two nitrothiophene derivatives : 222-acetyl-5-nitrothiophene and 5-nitro-2-thiophenecarboxaldehyde, *Journal of Thermal Analysis and Calorimetry*, 100 (2010) 403-411.
- [308] M.D.M.C. Ribeiro Da Silva, V.L.S. Freitas, M.A.A. Vieira, M.J. Sottomayor, W.E. Acree Jr, Energetic and structural properties of 4-nitro-2,1,3-benzothiadiazole, *Journal of Chemical Thermodynamics*, 49 (2012) 146-153.
- [309] M.D.M.C. Ribeiro da Silva, M.S. Miranda, C.M.V. Vaz, M.A.R. Matos, W.E. Acree Jr, Experimental thermochemical study of three monosubstituted pyrazines, *The Journal of Chemical Thermodynamics*, 37 (2005) 49-53.
- [310] M.V. Roux, C. Foces-Foces, R. Notario, M.A.V. Ribeiro Da Silva, M.D.D.M.C. Ribeiro Da Silva, A.F.L.O.M. Santos, E. Juaristi, Experimental and computational thermochemical study of sulfur-containing amino acids: L -cysteine, l -cystine, and l -cysteine-derived radicals. S-S, S-H, and C-S bond dissociation enthalpies, *Journal of Physical Chemistry B*, 114 (2010) 10530-10540.
- [311] M.V. Roux, R. Notario, C. Foces-Foces, M. Temprado, F. Ros, V.N. Emel'Yanenko, S.P. Verevkin, Experimental and computational thermochemical study of barbituric acids: Structure-energy relationship in 1,3-dimethylbarbituric acid, *Journal of Physical Chemistry A*, 115 (2011) 3167-3173.
- [312] M.V. Roux, R. Notario, M. Segura, J.S. Chickos, Thermophysical study of 2-thiobarbituric acids by differential scanning calorimetry, *Journal of Chemical and Engineering Data*, 57 (2012) 249-255.

- [313] M.V. Roux, M. Temprado, P. Jiménez, C. Foces-Foces, R. Notario, A.R. Parameswar, A.V. Demchenko, J.S. Chickos, C.A. Deakyne, J.F. Liebman, Experimental and theoretical study of the structures and enthalpies of formation of 3 H -1,3-benzoxazole-2-thione, 3 H -1,3-benzothiazole-2-thione, and their tautomers, *Journal of Physical Chemistry A*, 114 (2010) 6336-6341.
- [314] M.V. Roux, M. Temprado, P. Jiménez, R. Notario, A.R. Parameswar, A.V. Demchenko, J.S. Chickos, C.A. Deakyne, J.F. Liebman, Knowledge of a molecule: An experimental and theoretical study of the structure and enthalpy of formation of tetrahydro-2 H -1,3-oxazine-2-thione, *Journal of Chemical and Engineering Data*, 56 (2011) 4725-4732.
- [315] A.F.L.O.M. Santos, M.A.V.R. Da Silva, Energetics of 1-(aminophenyl)pyrroles: A joint calorimetric and computational study, *Journal of Chemical Thermodynamics*, 43 (2011) 1480-1487.
- [316] A.F.L.O.M. Santos, M.A.V. Ribeiro Da Silva, Experimental and computational energetic study of two halogenated 2-acetylpyrrole derivatives: 2-Trichloroacetylpyrrole and 2-trifluoroacetylpyrrole, *Journal of Chemical Thermodynamics*, 42 (2010) 1079-1086.
- [317] A.F.L.O.M. Santos, M.A.V. Ribeiro da Silva, Molecular energetics of pyrrolecarbonitriles and derivatives: A combined calorimetric and computational study, *Journal of Chemical Thermodynamics*, 48 (2012) 194-200.
- [318] A.F.L.O.M. Santos, M.A.V. Ribeiro Da Silva, The enthalpies of formation of alkyl carbamates: Experimental and computational redetermination, *Journal of Chemical Thermodynamics*, 57 (2013) 454-460.
- [319] L.M.N.B.F. Santos, L.M.S.S. Lima, C.F.R.A.C. Lima, F.D. Magalhães, M.C. Torres, B. Schröder, M.A.V. Ribeiro Da Silva, New Knudsen effusion apparatus with simultaneous

gravimetric and quartz crystal microbalance mass loss detection, *Journal of Chemical Thermodynamics*, 43 (2011) 834-843.

[320] C.C.S. Sousa, M.A.R. Matos, V.M.F. Morais, When theory and experiment hold hands: The thermochemistry of γ -pyrone derivatives, *Journal of Chemical Thermodynamics*, 43 (2011) 1159-1163.

[321] C.C.S. Sousa, V.M.F. Morais, M.A.R. Matos, Energetics of the isomers: 3- and 4-hydroxycoumarin, *Journal of Chemical Thermodynamics*, 42 (2010) 1372-1378.

[322] P. Umnahanant, D. Hasty, J. Chickos, An examination of the thermodynamics of fusion, vaporization, and sublimation of (R,S)- and (R)-flurbiprofen by correlation gas chromatography, *Journal of Pharmaceutical Sciences*, 101 (2012) 2045-2054.

[323] M.A. Varfolomeev, D.I. Abaidullina, B.N. Solomonov, S.P. Verevkin, V.N. Emel'Yanenko, Pairwise substitution effects, inter- and intramolecular hydrogen bonds in methoxyphenols and dimethoxybenzenes. Thermochemistry, calorimetry, and first-principles calculations, *Journal of Physical Chemistry B*, 114 (2010) 16503-16516.

[324] R.M. Varushchenko, A.A. Efimova, A.I. Druzhinina, E.S. Tkachenko, I.A. Nesterov, T.N. Nesterova, S.P. Verevkin, The heat capacities and thermodynamic functions of 4-methylbiphenyl and 4-tert-butylbiphenyl, *Journal of Chemical Thermodynamics*, 42 (2010) 1265-1272.

[325] V.N. Emel'yanenko, S.V. Portnova, S.P. Verevkin, A. Skrzypczak, T. Schubert, Building blocks for ionic liquids: Vapor pressures and vaporization enthalpies of 1-(n-alkyl)-imidazoles, *Journal of Chemical Thermodynamics*, 43 (2011) 1500-1505.

[326] M.L.F. Viveiros, V.L.S. Freitas, N. Vale, J.R.B. Gomes, P. Gomes, M.D.M.C.R. Da Silva, Synthesis and thermochemical study of quinoxaline-N-oxides: Enthalpies of dissociation of the N-O bond, *Journal of Physical Organic Chemistry*, 25 (2012) 420-426.

- [327] J.A. Widegrenand, T.J. Bruno, Vapor pressure measurements on low-volatility terpenoid compounds by the concatenated gas saturation method, *Environmental Science and Technology*, 44 (2010) 388-393.
- [328] L.E. Agafonova, R.M. Varushchenko, A.I. Druzhinina, O.V. Polyakova, Phase equilibria and the thermodynamic properties of methyl and ethyl esters of carboxylic acids. 1. Methyl n-butanoate and ethyl propanoate, *Journal of Chemical Thermodynamics*, 47 (2012) 120-129.
- [329] L.E. Agafonova, R.M. Varushchenko, A.I. Druzhinina, O.V. Polyakova, Y.S. Kolesov, Heat capacity, saturation vapor pressure, and thermodynamic functions of ethyl esters of C-3-C-5 and C-18 carboxylic acids, *Russian Journal of Physical Chemistry A*, 85 (2011) 1516-1527.
- [330] N.C.-B. Ahmed, L. Negadi, I. Mokbel, A.A. Kaci, J. Jose, Experimental determination of the isothermal (vapour plus liquid) equilibria of binary aqueous solutions of sec-butylamine and cyclohexylamine at several temperatures, *Journal of Chemical Thermodynamics*, 44 (2012) 116-120.
- [331] A.R.R.P. Almeida, M.J.S. Monte, Thermodynamic Study of Benzamide, N-Methylbenzamide, and N,N-Dimethylbenzamide: Vapor Pressures, Phase Diagrams, and Hydrogen Bond Enthalpy, *Journal of Chemical and Engineering Data*, 55 (2010) 3507-3512.
- [332] A.R.R.P. Almeida, M.J.S. Monte, Thermodynamic Study of the Three Fluorobenzamides: Vapor Pressures, Phase Diagrams, and Hydrogen Bonds, *Journal of Chemical and Engineering Data*, 55 (2010) 5230-5236.
- [333] A.R.R.P. Almeida, M.J.S. Monte, Thermodynamic study of phase transitions in methyl esters of ortho- meta- and para-aminobenzoic acids, *Journal of Chemical Thermodynamics*, 53 (2012) 100-107.

- [334] L.M.P.F. Amaral, M.A.V. Ribeiro da Silva, Thermochemistry of some methoxypyridines, *Journal of Chemical Thermodynamics*, 48 (2012) 65-69.
- [335] L.M.P.F. Amaral, M.A.V. Ribeiro da Silva, Calorimetric study of 2'-methylacetophenone and 4'-methylacetophenone, *Journal of Chemical Thermodynamics*, 57 (2013) 301-305.
- [336] A. Belabbaci, N.C.-B. Ahmed, I. Mokbel, L. Negadi, Investigation of the isothermal (vapour plus liquid) equilibria of aqueous 2-amino-2-methyl-1-propanol (AMP), N-benzylethanolamine, or 3-dimethylamino-1-propanol solutions at several temperatures, *Journal of Chemical Thermodynamics*, 42 (2010) 1158-1162.
- [337] M. Benziane, K. Khimeche, I. Mokbel, A. Dahmani, J. Jose, Isothermal Vapor-Liquid Equilibria of n-Tetradecane plus Ethyl Hexanoate, Ethyl Decanoate, and Ethyl Tetradecanoate, *Journal of Chemical and Engineering Data*, 58 (2013) 492-498.
- [338] M. Benziane, K. Khimeche, I. Mokbel, T. Sawaya, A. Dahmani, J. Jose, Experimental Vapor Pressures of Five Saturated Fatty Acid Ethyl Ester (FAEE) Components of Biodiesel, *Journal of Chemical and Engineering Data*, 56 (2011) 4736-4740.
- [339] E. Bogdani, R. Daoussi, S. Vessot, J. Jose, J. Andrieu, Implementation and validation of the thermogravimetric method for the determination of equilibrium vapour pressure values and sublimation enthalpies of frozen organic formulations used in drug freeze-drying processes, *Chemical Engineering Research & Design*, 89 (2011) 2606-2612.
- [340] R.D. Chirico, A.F. Kazakov, W.V. Steele, Thermodynamic properties of 9-fluorenone: Mutual validation of experimental and computational results, *Journal of Chemical Thermodynamics*, 54 (2012) 278-287.
- [341] P.S. Davis, J.E. Kilpatrick, Entropy, related thermodynamic properties, and structure of methylisocyanate, *Journal of Chemical Thermodynamics*, 58 (2013) 134-141.

- [342] V.N. Emel'yanenko, S.P. Verevkin, E.N. Stepurko, G.N. Roganov, The thermodynamic characteristics of 15-pentadecanolide and 16-hexadecanolide, *Russian Journal of Physical Chemistry A*, 85 (2011) 348-356.
- [343] V.N. Emel'yanenko, S.P. Verevkin, E.N. Stepurko, G.N. Roganov, M.K. Georgieva, Thermodynamical properties of ϵ -caprolactone, *Russian Journal of Physical Chemistry A*, 84 (2010) 356-363.
- [344] V.N. Emel'yanenko, S.P. Verevkin, M.A. Varfolomeev, V.V. Turovtsev, Y.D. Orlov, Thermochemical Properties of Formamide Revisited: New Experiment and Quantum Mechanical Calculations, *Journal of Chemical and Engineering Data*, 56 (2011) 4183-4187.
- [345] J. Fu, J.W. Rice, E.M. Suuberg, Phase behavior and vapor pressures of the pyrene+9,10-dibromoanthracene system, *Fluid Phase Equilibria*, 298 (2010) 219-224.
- [346] M. Fulem, K. Ruzicka, M. Ruzicka, Recommended vapor pressures for thiophene, sulfolane, and dimethyl sulfoxide, *Fluid Phase Equilibria*, 303 (2011) 205-216.
- [347] B. Gimeno, S. Martinez, J.S. Urieta, P. Perez, Vapor Pressures and Activity Coefficients of (1-Propanol plus 1,8-Cineole) at 10 Temperatures between 278.15 K and 323.15 K, *Journal of Chemical and Engineering Data*, 57 (2012) 3026-3031.
- [348] B. Gimeno, M. Torcal, A.M. Mainar, J.S. Urieta, P. Perez, Total vapour pressure and excess Gibbs energy of ethanol with 1,8-cineole at temperatures between 278.15 K and 323.15 K, *Fluid Phase Equilibria*, 309 (2011) 83-88.
- [349] D. Hasty, J. Drapekin, T. Subramanian, T.C. Winter, J.S. Chickos, A.A. Samarov, A.V. Yermalayeu, S.P. Verevkin, Applications of Correlation Gas Chromatography and Transpiration Studies for the Evaluation of the Vaporization and Sublimation Enthalpies of Some Perfluorinated Hydrocarbons, *Journal of Chemical and Engineering Data*, 57 (2012) 2350-2359.

- [350] K. Khimeche, A. Dahmani, I. Mokbel, Excess properties and vapour pressure of {3-diethylaminopropylamine plus cyclohexane}, *Journal of Chemical Thermodynamics*, 42 (2010) 829-833.
- [351] K. Khimeche, F. Djellouli, A. Dahmani, I. Mokbel, Excess Properties and Vapor Pressure of 2-Diethylaminoethylamine plus n-Heptane, *Journal of Chemical and Engineering Data*, 56 (2011) 4972-4977.
- [352] D. Lipkind, J.S. Chickos, J.F. Liebman, Study of the Anomalous Thermochemical Behavior of 1,2-Diazines by Correlation-Gas Chromatography, *Journal of Chemical and Engineering Data*, 55 (2010) 1628-1635.
- [353] D. Lipkind, N. Rath, J.S. Chickos, V.A. Pozdeev, S.P. Vereykin, The Vaporization Enthalpies of 2-and 4-(N,N-Dimethylamino)pyridine, 1,5-Diazabicyclo 4.3.0 non-5-ene, 1,8-Diazabicyclo 5.4.0 undec-7-ene, Imidazo 1,2-a pyridine and 1,2,4-Triazolo 1,5-a pyrimidine by Correlation-Gas Chromatography, *Journal of Physical Chemistry B*, 115 (2011) 8785-8796.
- [354] S.V. Lipp, E.L. Krasnykh, S.P. Verevkin, Vapor pressures and enthalpies of vaporization of a series of the symmetric linear n -alkyl esters of dicarboxylic acids, *Journal of Chemical and Engineering Data*, 56 (2011) 800-810.
- [355] A.S. Maslakova, E.L. Krasnykh, S.V. Levanova, Saturation vapor pressures and vaporization enthalpies of esters of ethylene glycol and lower carboxylic acids, *Russian Journal of Physical Chemistry A*, 85 (2011) 1695-1700.
- [356] M.S. Miranda, V.M.F. Morais, M.A.R. Matos, J.F. Liebman, Standard molar enthalpy of formation of 1-benzosuberone: An experimental and computational study, *Journal of Chemical Thermodynamics*, 42 (2010) 1094-1100.

- [357] E.A. Miroshnichenko, T.S. Kon'Kova, Y.O. Inozemtsev, Y.N. Matyushin, Bond energies and the enthalpies of formation of mono- and polyradicals in nitroalkanes 3. Nitroalkanes C4-C7, *Russian Chemical Bulletin*, 60 (2011) 36-41.
- [358] I. Mokbel, T. Sawaya, M.-L. Zanota, R. Abou Naccoul, J. Jose, C. de Bellefon, Vapor-Liquid Equilibria of Glycerol, 1,3-Propanediol, Glycerol plus Water, and Glycerol+1,3-Propanediol, *Journal of Chemical and Engineering Data*, 57 (2012) 284-289.
- [359] M.J.S. Monte, R. Notario, M.M.G. Calvino, A.R.R.P. Almeida, L.M.P.F. Amaral, A.I.M.C. Lobo Ferreira, M.D.M.C. Ribeiro da Silva, Experimental and Computational Study of the Thermodynamic Properties of 9-Fluorenone and 9-Fluorenone, *Journal of Chemical and Engineering Data*, 57 (2012) 2486-2496.
- [360] M.J.S. Monte, S.P. Pinto, A.I.M.C. Lobo Ferreira, L.M.P.F. Amaral, V.L.S. Freitas, M.D.M.C. Ribeiro da Silva, Fluorene: An extended experimental thermodynamic study, *Journal of Chemical Thermodynamics*, 45 (2012) 53-58.
- [361] L. Negadi, A.A. Kaci, A. Negadi, J. Jose, Isothermal vapor-liquid equilibria and excess enthalpies of binary mixtures of propyl ethanoate+1-hexyne,+2-hexyne,+3-hexyne, *Fluid Phase Equilibria*, 300 (2011) 135-144.
- [362] V.A. Pozdeev, S.P. Verevkin, Vapor pressure and enthalpy of vaporization of linear aliphatic alkanediamines, *Journal of Chemical Thermodynamics*, 43 (2011) 1791-1799.
- [363] K.R. Reddy, D.B.K. Kumar, G.S. Rao, P.B.S. Sri, C. Rambabu, Vapor-liquid equilibria and excess molar volumes of N-methyl-2-pyrrolidone with 2-alkoxyethanols, *Fluid Phase Equilibria*, 336 (2012) 52-58.
- [364] M.A.V. Ribeiro da Silva, L.M.P.F. Amaral, Standard molar enthalpies of formation of some methylfuran derivatives, *Journal of Thermal Analysis and Calorimetry*, 100 (2010) 375-380.

- [365] M.A.V. Ribeiro da Silva, A.I.M.C. Lobo Ferreira, A. Cimas, Calorimetric and Computational Study of the Thermochemistry of Phenoxyphenols, *Journal of Organic Chemistry*, 76 (2011) 3754-3764.
- [366] M.A.V. Ribeiro da Silva, M.J.S. Monte, I.M. Rocha, Á. Cimas, Energetic Study Applied to the Knowledge of the Structural and Electronic Properties of Monofluorobenzonitriles, *The Journal of Organic Chemistry*, 77 (2012) 4312-4322.
- [367] M.A.V. Ribeiro da Silva, A.F.L.O.M. Santos, L.M.P.F. Amaral, A calorimetric and computational study on the thermochemistry of 2-(5H)-furanone and 2-(5H)-thiophenone, *Journal of Chemical Thermodynamics*, 42 (2010) 564-570.
- [368] M.A.V. Riberio da Silva, L.M.P.F. Amaral, Thermochemical study of 2,5-dimethyl-3-furancarboxylic acid, 4,5-dimethyl-2-furaldehyde, and 3-acetyl-2,5-dimethylfuran, *Journal of Chemical Thermodynamics*, 43 (2011) 1-8.
- [369] R. Rios, J. Ortega, L. Fernandez, Measurements and Correlations of the Isobaric Vapor-Liquid Equilibria of Binary Mixtures and Excess Properties for Mixtures Containing an Alkyl (Methyl, Ethyl) Butanoate with an Alkane (Heptane, Nonane) at 101.3 kPa, *Journal of Chemical and Engineering Data*, 57 (2012) 3210-3224.
- [370] M.V. Roux, G. Martin-Valcarcel, R. Notario, S. Kini, J.S. Chickos, J.F. Liebman, The Joining of Measurement and Prediction: The Enthalpy of Formation of 1,4-Cubanedicarboxylic Acid, *Journal of Chemical and Engineering Data*, 56 (2011) 1220-1228.
- [371] A.F.L.O.M. Santos, M.A.V. Ribeiro da Silva, A Combined Experimental and Computational Thermodynamic Study of the Isomers of Pyrrolecarboxaldehyde and 1-Methyl-pyrrolecarboxaldehyde, *Journal of Physical Chemistry B*, 115 (2011) 12549-12557.

- [372] A.F.L.O.M. Santos, M.A.V. Ribeiro da Silva, Experimental redetermination of the gas-phase enthalpy of formation of ethyl 2-thiophenecarboxylate, *Journal of Chemical Thermodynamics*, 58 (2013) 476-478.
- [373] V.M.T.M. Silva, C.S.M. Pereira, A.E. Rodrigues, S.P. Verevkin, V.N. Emel'yanenko, I.V. Garist, J. Gmehling, Experimental and Theoretical Study of Chemical Equilibria in the Reactive Systems of Acetals Synthesis, *Industrial & Engineering Chemistry Research*, 51 (2012) 12723-12729.
- [374] C.C.S. Sousa, V.M.F. Morais, M.A.R. Matos, Experimental and computational thermochemistry of 6,7-dihydro-4(5H)-benzofuranone, *Journal of Chemical Thermodynamics*, 56 (2013) 83-88.
- [375] V. Stejfa, M. Fulem, K. Ruzicka, C. Cervinka, M.A.A. Rocha, L.M.N.B.F. Santos, B. Schroeder, Thermodynamic study of selected monoterpenes, *Journal of Chemical Thermodynamics*, 60 (2013) 117-125.
- [376] M. Suceska, S.M. Musanic, I.F. Houra, Kinetics and enthalpy of nitroglycerin evaporation from double base propellants by isothermal thermogravimetry, *Thermochimica Acta*, 510 (2010) 9-16.
- [377] A.O. Surov, G.L. Perlovich, V.N. Emel'yanenko, S.P. Verevkin, Thermochemistry of Drugs. Experimental and First-Principles Study of Fenamates, *Journal of Chemical and Engineering Data*, 56 (2011) 4325-4332.
- [378] P. Umnahanant, J. Chickos, An Examination of the Thermodynamics of Fusion, Vaporization, and Sublimation of Several Parabens by Correlation Gas Chromatography, *Journal of Pharmaceutical Sciences*, 100 (2011) 1847-1855.

- [379] P. Umnahanant, J. Chickos, Vaporization and Sublimation Enthalpies of Acetanilide and Several Derivatives by Correlation Gas Chromatography, *Journal of Chemical and Engineering Data*, 57 (2012) 1331-1337.
- [380] S.P. Verevkin, V.N. Emel'yanenko, R. Notario, M. Victoria Roux, J.S. Chickos, J.F. Liebman, Rediscovering the Wheel. Thermochemical Analysis of Energetics of the Aromatic Diazines, *Journal of Physical Chemistry Letters*, 3 (2012) 3454-3459.
- [381] S.P. Verevkin, B. Tong, U. Welz-Biermann, Y. Chernyak, Vapor Pressures and Enthalpies of Vaporization of a Series of Low-Volatile Alkanolamines, *Journal of Chemical and Engineering Data*, 56 (2011) 4400-4406.
- [382] S.P. Verevkin, V.N. Emel'yanenko, A. Heintz, Liquid Organic Hydrogen Carriers: An Upcoming Alternative to Conventional Technologies. *Thermochemical Studies, Industrial & Engineering Chemistry Research*, 51 (2012) 12150-12153.
- [383] S.P. Verevkin, V.N. Emel'yanenko, A.A. Pimerzin, E.E. Vishnevskaya, Thermodynamic Analysis of Strain in the Five-Membered Oxygen and Nitrogen Heterocyclic Compounds, *Journal of Physical Chemistry A*, 115 (2011) 1992-2004.
- [384] S.P. Verevkin, V.N. Emel'yanenko, A.A. Pimerzin, E.E. Vishnevskaya, Thermodynamic Analysis of Strain in Heteroatom Derivatives of Indene, *Journal of Physical Chemistry A*, 115 (2011) 12271-12279.
- [385] J.A. Widegren, T.J. Bruno, Vapor Pressure Measurements on Low-Volatility Terpenoid Compounds by the Concatenated Gas Saturation Method, *Environmental Science & Technology*, 44 (2010) 388-393.

- [386] J.A. Wilson, J. Chickos, Vaporization Enthalpy and Vapor Pressure of Valproic Acid by Correlation Gas Chromatography, *Journal of Chemical and Engineering Data*, 57 (2012) 2281-2285.
- [387] J.A. Wilson, J.S. Chickos, Vapor Pressures and Vaporization, Sublimation, and Fusion Enthalpies of Some Fatty Acids, *Journal of Chemical and Engineering Data*, 58 (2013) 322-333.
- [388] L.K. Zaitri, L. Negadi, I. Mokbel, N. Msakni, J. Jose, Liquid-vapor equilibria of binary systems containing alcohols (1-butanol, or 2-butanol or 1-hexanol) present in the production by chemical process of 2,5-dimethyl furan from biomass, *Fuel*, 95 (2012) 438-445.
- [389] Z.-X. Zeng, J. Chen, W.-L. Xue, P. Zhang, Y. Xie, Vapour pressure and vapourisation enthalpy of pyridine N-oxide, *Canadian Journal of Chemical Engineering*, 90 (2012) 570-575.
- [390] F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, Toward a group contribution method for determination of speed of sound in saturated liquids, *Journal of Molecular Liquids*, 194 (2014) 159-165.
- [391] G.A.F. Seber, *Multivariate Observations*, John Wiley & Sons, Inc., Hoboken, NJ, 1984.
- [392] H. Spath, *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples*. (Translated by J. Goldschmidt.), Halsted Press, New York, 1985.
- [393] S.R. Taghanaki, M. Arabloo, A. Chamkalani, M. Amani, M.H. Zargari, M.R. Adelzadeh, Implementation of SVM framework to estimate PVT properties of reservoir oil, *Fluid Phase Equilibria*, (2013).
- [394] G. Scalabrin, P. Marchi, L. Bettio, D. Richon, Enhancement of the extended corresponding states techniques for thermodynamic modeling. II. Mixtures, *International journal of refrigeration*, 29 (2006) 1195-1207.

- [395] F. Gharagheizi, R.F. Alamdari, M.T. Angaji, A new neural network– group contribution method for estimation of flash point temperature of pure components, *Energy & Fuels*, 22 (2008) 1628-1635.
- [396] A.H. Mohammadi, D. Richon, A mathematical model based on artificial neural network technique for estimating liquid water– hydrate equilibrium of water– hydrocarbon System, *Industrial & Engineering Chemistry Research*, 47 (2008) 4966-4970.
- [397] F. Gharagheizi, Prediction of the standard enthalpy of formation of pure compounds using molecular structure, *Australian Journal of Chemistry*, 62 (2009) 376-381.
- [398] A. Kamari, A. Hemmati-Sarapardeh, S.-M. Mirabbasi, M. Nikookar, A.H. Mohammadi, Prediction of sour gas compressibility factor using an intelligent approach, *Fuel Processing Technology*, 116 (2013) 209-216.
- [399] A. Kamari, A. Khaksar-Manshad, F. Gharagheizi, A.H. Mohammadi, S. Ashoori, Robust Model for the Determination of Wax Deposition in Oil Systems, *Industrial & Engineering Chemistry Research*, 52 (2013) 15664–15672.
- [400] C.L. YAWS, *Yaws handbook of physical properties for hydrocarbons and chemicals*, (2005).
- [401] L. Eriksson, E. Johansson, M. Müller, S. Wold, On the selection of the training set in environmental QSAR analysis when compounds are clustered, *Journal of Chemometrics*, 14 (2000) 599-616.
- [402] A.R. Katritzky, M. Kuanar, S. Slavov, C.D. Hall, M. Karelson, I. Kahn, D.A. Dobchev, Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction, *Chemical reviews*, 110 (2010) 5714-5789.
- [403] S.L. Mayo, B.D. Olafson, W.A. Goddard, DREIDING: a generic force field for molecular simulations, *Journal of Physical Chemistry*, 94 (1990) 8897-8909.

- [404] DRAGON for Windows (Software for Molecular Descriptor Calculations) Version 5.5, in.
- [405] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, Vol. I, Alphabetical Listing, Volume II, Appendices, References, (2009).
- [406] 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.
- [407] Dragon, (Software for molecular Descriptor Calculations) Talete srl, in.
- [408] F. Gharagheizi, G.R. Salehi, Prediction of enthalpy of fusion of pure compounds using an artificial neural network-group contribution method, *Thermochimica Acta*, 521 (2011) 37-40.
- [409] F. Gharagheizi, A. Eslamimanesh, A.H. Mohammadi, D. Richon, Group contribution model for determination of molecular diffusivity of non-electrolyte organic compounds in air at ambient conditions, *Chemical Engineering Science*, 68 (2012) 290-304.
- [410] F. Gharagheizi, A. Eslamimanesh, A.H. Mohammadi, D. Richon, Representation/Prediction of Solubilities of Pure Compounds in Water Using Artificial Neural Network-Group Contribution Method, *Journal of Chemical and Engineering Data*, 56 (2011) 720-726.
- [411] F. Gharagheizi, S.A. Mirkhani, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, D. Richon, Determination of the Normal Boiling point of chemical compounds using a Quantitative Structure-Property Relationship strategy: Application to a very large dataset, *Fluid Phase Equilibria*, (2013).
- [412] F. Gharagheizi, A. Eslamimanesh, P. Ilani-Kashkouli, A.H. Mohammadi, D. Richon, QSPR molecular approach for representation/prediction of very large vapor pressure dataset, *Chemical Engineering Science*, 76 (2012) 99-107.
- [413] R. Todeschini, V. Consonni, *Molecular descriptors for chemoinformatics*, Wiley. com, 2009.

- [414] F. Gharagheizi, P. Ilani-Kashkouli, W.E. Acree Jr, A.H. Mohammadi, D. Ramjugernath, A group contribution model for determining the sublimation enthalpy of organic compounds at the standard reference temperature of 298K, *Fluid Phase Equilibria*, 354 (2013) 265-285.
- [415] F. Gharagheizi, A. Eslamimanesh, A.H. Mohammadi, D. Richon, Determination of Critical Properties and Acentric Factors of Pure Compounds Using the Artificial Neural Network Group Contribution Algorithm, *Journal of Chemical and Engineering Data*, 56 (2011) 2460-2476.
- [416] F. Gharagheizi, A. Eslamimanesh, A.H. Mohammadi, D. Richon, Representation/prediction of solubilities of pure compounds in water using artificial neural network– group contribution method, *Journal of Chemical & Engineering Data*, 56 (2011) 720-726.
- [417] F. Gharagheizi, A. Eslamimanesh, P. Ilani-Kashkouli, A.H. Mohammadi, D. Richon, Determination of Vapor Pressure of Chemical Compounds: A Group Contribution Model for an Extremely Large Database, *Industrial & Engineering Chemistry Research*, 51 (2012) 7119-7125.
- [418] F. Gharagheizi, O. Babaie, S. Mazdeyasna, Prediction of vaporization enthalpy of pure compounds using a group contribution-based method, *Industrial and Engineering Chemistry Research*, 50 (2011) 6503-6507.
- [419] F. Gharagheizi, R. Abbasi, A New Neural Network Group Contribution Method for Estimation of Upper Flash Point of Pure Chemicals, *Industrial & Engineering Chemistry Research*, 49 (2010) 12685-12695.
- [420] F. Gharagheizi, An accurate model for prediction of autoignition temperature of pure compounds, *Journal of Hazardous Materials*, 189 (2011) 211-221.
- [421] F. Gharagheizi, A new group contribution-based model for estimation of lower flammability limit of pure compounds, *Journal of hazardous materials*, 170 (2009) 595-604.

- [422] F. Gharagheizi, A. Eslamimanesh, F. Farjood, A.H. Mohammadi, D. Richon, Solubility parameters of nonelectrolyte organic compounds: determination using quantitative structure–property relationship strategy, *Ind. Eng. Chem. Res.*, 50 (2011) 11382-11395.
- [423] A.Z. Dudek, T. Arodz, J. Galvez, Computational methods in developing quantitative structure-activity relationships (QSAR): a review, *Combinatorial chemistry & high throughput screening*, 9 (2006) 213-228.
- [424] W.J. Lyman, W.F. Reehl, D.H. Rosenblatt, *Handbook of chemical property estimation methods*, (1990).
- [425] A. Lydersen, R.A. Greenkorn, O.A. Hougen, *Estimation of Critical Properties of Organic Compounds by the Method of Group Contributions*, University of Wisconsin, 1955.
- [426] K.G. Joback, R.C. Reid, Estimation of pure-component properties from group-contributions, *Chemical Engineering Communications*, 57 (1987) 233-243.
- [427] A.L. Horvath, *Molecular design: chemical structure generation from the properties of pure organic compounds*, Access Online via Elsevier, 1992.
- [428] D. Ambrose, Correlation and estimation of vapour-liquid critical properties. Part 1: Critical temperatures of organic compounds, (1978).
- [429] K. Klincewicz, R. Reid, Estimation of critical properties with group contribution methods, *AIChE journal*, 30 (1984) 137-142.
- [430] J. Marrero, R. Gani, Group-contribution based estimation of pure component properties, *Fluid Phase Equilibria*, 183 (2001) 183-208.
- [431] C. Baroncini, P. Di Filippo, G. Latini, M. Pacetti, Correlations for Thermal Conductivity Evaluation of Paraffins, Cycloparaffins and Ketones between the Melting Point and the Critical

Temperature, Correlazioni Per Il Calcolo Della Conducibilita Termica Di Paraffine, Cicloparaffine E Chetoni Tra La Temperatura Di Fusione E La Temperatura Critica, 35 (1981) 633-638.

[432] C. Baroncini, P. Di Filippo, G. Latini, M. Pacetti, Organic liquid thermal conductivity: A prediction method in the reduced temperature range 0.3 to 0.8, International Journal of Thermophysics, 2 (1981) 21-38.

[433] 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, Journal of Molecular Liquids, 190 (2014) 223-230.

[434] K.R. Seddon, A. Stark, M.-J. Torres, Influence of chloride, water, and organic solvents on the physical properties of ionic liquids, Pure and Applied Chemistry, 72 (2000) 2275-2287.

[435] R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics, (2009) 2.

[436] F. Gharagheizi, P. Ilani-Kashkouli, W.E. Acree, A.H. Mohammadi, D. Ramjugernath, A group contribution model for determining the vaporization enthalpy of organic compounds at the standard reference temperature of 298K, Fluid Phase Equilibria, 360 (2013) 279-292.

[437] 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, Chemical Engineering Research and Design, 92 (2014) 66-79.

[438] F. Gharagheizi, S.A. Mirkhani, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, D. Richon, Determination of the Normal Boiling Point of Chemical Compounds using a Quantitative Structure–Property Relationship strategy: Application to a very Large Dataset, Fluid Phase Equilibria, 354 (2013) 250-258.

[439] E.S. Domalski, E.D. Hearing, J. Phys. Chem. Ref. Data, 22 (1993) 805-1159.

- [440] 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 and Engineering Chemistry Research*, 51 (2012) 3844-3849.
- [441] A.H. Mohammadi, A. Eslamimanesh, F. Gharagheizi, D. Richon, A novel method for evaluation of asphaltene precipitation titration data, *Chemical Engineering Science*, 78 (2012) 181-185.
- [442] P.J. Rousseeuw, A.M. Leroy, *Robust regression and outlier detection*, Wiley. com, 2005.
- [443] A.H. Mohammadi, F. Gharagheizi, A. Eslamimanesh, D. Richon, Evaluation of experimental data for wax and diamondoids solubility in gaseous systems, *Chemical Engineering Science*, 81 (2012) 1-7.
- [444] P. Gramatica, Principles of QSAR models validation: internal and external, *QSAR & combinatorial science*, 26 (2007) 694-701.
- [445] C.R. Goodall, 13 Computation using the QR decomposition, *Handbook of Statistics*, 9 (1993) 467-508.

Appendix

Statistical Parameters: Mathematical Definition

The mathematical definition of the relative deviation (RD%), average absolute relative deviation (AARD%), root mean square error (RMSE), standard deviation error (Std), and squared correlation coefficient (R^2) are presented as follows:

(A1)

$$RD\% = 100 \times \frac{pred - lit}{lit}$$

(A2)

$$AARD\% = \frac{100}{N} \sum_i^N \frac{|pred(i) - lit(i)|}{lit(i)}$$

(A3)

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (pred(i) - lit(i))^2}{N}}$$

(A4)

$$Std = \frac{1}{N} \sum_i^N \sqrt{(pred(i) - \overline{pred})^2}$$

(A5)

$$R^2 = 1 - \frac{\sum_i^N (pred(i) - lit(i))^2}{\sum_i^N (pred(i) - \overline{pred})^2}$$

where $pred$ and lit denote the predicted value by model and its corresponding value reported by literature, respectively. The terms \overline{pred} and \overline{lit} refer to the mean values over the predicted values by the model and the mean value over the literature reported data. N is the number of data point in each data set or subset.

