



**TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE AGUASCALIENTES**

DOCTOR OF ENGINEERING SCIENCES

***“DEVELOPMENT OF IMPROVED THERMODYNAMIC MODELS FOR THE
PREDICTION OF PROPERTIES FOR PROCESS SYSTEMS ENGINEERING”***

THESIS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

Doctor of Engineering Sciences

By:

M.Sc. Vanessa Villazón León

Thesis Director:

Dr. Juan Carlos Tapia Picazo

Co-Director:

Dr. Adrián Bonilla Petriciolet

Advisors:

Dr. Hilda Elizabeth Reynel Ávila

Dr. José Enrique Jaime Leal

Dr. J. Gabriel Luna Bárcenas

Aguascalientes, Ags., Mexico

February 2025



Instituto **Tecnológico**
de Aguascalientes



Educación
Secretaría de Educación Pública



TECNOLÓGICO
NACIONAL DE MÉXICO



Instituto Tecnológico de Aguascalientes
División de Estudios de Posgrado e Investigación

Aguascalientes, Ags., 17/febrero/2025

No. De Oficio: 019/2025

Asunto: **Autorización de Tesis de Grado Para su Impresión**

M.C. VANESSA VILLAZÓN LEÓN
P R E S E N T E

Los abajo firmantes, Miembros del Jurado para Examen de Grado de Doctorado, hacemos **CONSTAR** que, habiendo revisado el Trabajo de Tesis desarrollado por Usted, bajo el título: **"DEVELOPMENT OF IMPROVED THERMODYNAMIC MODELS FOR THE PREDICTION OF PROPERTIES FOR PROCESS SYSTEMS ENGINEERING"** hemos dictaminado que éste es de aceptarse, por lo que se autoriza su impresión.

A T E N T A M E N T E

Excelencia en Educación Tecnológica®
Ingenio, Cultura y Saber, que conducen a la Excelencia®

PRESIDENTE

DR. JUAN CARLOS TAPIA PICAZO

SECRETARIO

DR. ADRIAN BONILLA PETRICIOLET

VOCAL

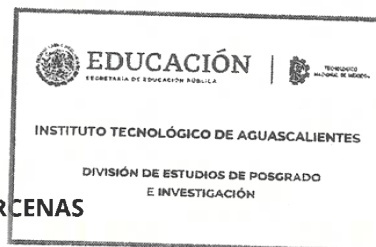
DRA. HILDA ELIZABETH REYNEL ÁVILA

VOCAL

DR. JOSÉ ENRIQUE JAIME LEAL

VOCAL

DR. J. GABRIEL LUNA BÁRCENAS



2025
Año de
La Mujer
Indígena

Av. Adolfo López Mateos #1801. Fracc. Bona Gens
C.P. 20256 Aguascalientes, Ags. Tel. 449 910 5002 ext. 127
e-mail: jefatura.depi@aguascalientes.tecnm.mx
tecnm.mx | www.aguascalientes.tecnm.mx





Instituto Tecnológico de Aguascalientes
División de Estudios de Posgrado e Investigación

Aguascalientes, Ags., **23/enero/2025**
Asunto: Autorización de uso de derechos

VILLAZÓN LEÓN VANESSA, mayor de edad, con domicilio ubicado en **Ultravioleta 104, Jardines del Sol, Aguascalientes, Aguascalientes** en mi calidad de titular y autor de la tesis denominada **"DEVELOPMENT OF IMPROVED THERMODYNAMIC MODELS FOR THE PREDICTION OF PROPERTIES FOR PROCESS SYSTEMS ENGINEERING"** quien para todos los fines del presente documento se denominará **EL AUTOR Y/O TITULAR**, suscribo el presente documento de autorización de uso de derechos patrimoniales de autor a favor del Tecnológico Nacional de México/Instituto Tecnológico de Aguascalientes el cual se registrá por las cláusulas siguientes:

PRIMERA – AUTORIZACIÓN: EL AUTOR Y/O TITULAR, mediante el presente documento autoriza la utilización de los derechos patrimoniales de autor al Tecnológico Nacional de México/Instituto Tecnológico de Aguascalientes, de la tesis denominada **"DEVELOPMENT OF IMPROVED THERMODYNAMIC MODELS FOR THE PREDICTION OF PROPERTIES FOR PROCESS SYSTEMS ENGINEERING"**, a través del Repositorio Institucional del Tecnológico Nacional de México (en lo sucesivo TecNM) y en el Repositorio Nacional, que puede ser consultado en la liga electrónica: (www.repositorionacionalcti.mx/).

SEGUNDA – OBJETO: Por medio del presente escrito, **EL AUTOR Y/O TITULAR** autoriza al Tecnológico Nacional de México/Instituto Tecnológico de Aguascalientes, a través del Repositorio Institucional del TecNM y en el Repositorio Nacional para que de conformidad con la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, use los derechos del documento antes referido, con fines exclusivamente académicos.

TERCERA – TERRITORIO: Los derechos aquí autorizados se dan sin limitación geográfica o territorial alguna.

CUARTA – ALCANCE: La presente autorización se da tanto para formato o soporte material, y se extiende a la utilización en medio óptico, magnético, electrónico, en red, mensajes de datos o similar conocido o por conocer, del ejemplar o número respectivo de la publicación.

QUINTA – EXCLUSIVIDAD: La autorización de uso aquí establecida no implica exclusividad en favor del Tecnológico Nacional de México/Instituto Tecnológico de Aguascalientes. Por lo tanto, **EL AUTOR Y/O TITULAR** en su carácter de autor de la obra objeto del presente documento se reserva el derecho de publicar directamente, u otorgar a cualquier tercero, autorizaciones de uso similares o en los mismos términos aquí acordados.



2025
Año de
La Mujer
Indígena

Av. Adolfo López Mateos #1801. Fracc. Bona Gens
C.P. 20256 Aguascalientes, Ags. Tel. 449 910 5002 ext. 127
e-mail: jefatura.depi@aguascalientes.tecnm.mx
tecnm.mx | www.aguascalientes.tecnm.mx





SEXTA – DERECHOS MORALES (Créditos y mención): La Autorización de los derechos antes mencionados no implica la cesión de los derechos morales sobre los mismos por cuanto en conformidad con lo establecido en los artículos 18, 19, 20, 21, 22 y 23 de la Ley Federal de Derechos de Autor, dada la cuenta que estos derechos son inalienables, imprescriptibles, irrenunciables e inembargables. Por lo tanto, los mencionados derechos seguirán radicados en cabeza de EL AUTOR Y/O TITULAR, y siempre deberá mencionarse su nombre cuando se utilice la obra.

SÉPTIMA – AUTORÍA: EL AUTOR Y/O TITULAR, declara y ratifica que el material objeto de la presente fue realizada por él o por ella sin violar o usurpar derechos de Propiedad Intelectual de terceros.

ATENTAMENTE

Excelencia en Educación Tecnológica®

Ingenio, Cultura y Saber, que conducen a la Excelencia

Vo.Bo.

VILLAZÓN LEÓN VANESSA
AUTOR DE LA TESIS

DR. JUAN CARLOS TAPIA PICAZO
DIRECTOR DE LA TESIS

ccp. Archivo



2025
Año de
La Mujer
Indígena

Av. Adolfo López Mateos #1801. Fracc. Bona Gens
C.P. 20256 Aguascalientes, Ags. Tel. 449 910 5002 ext. 127
e-mail: jefatura.depi@aguascalientes.tecnm.mx
tecnm.mx | www.aguascalientes.tecnm.mx





Instituto **Tecnológico**
de Aguascalientes

Abstract

This thesis addresses the improvement of thermodynamic properties predictions of pure compounds, which are essential for Process Systems Engineering (PSE). The study evaluates the potential of hybrid models based on Group Contribution Models (GCMs) and Machine Learning (ML) algorithms to overcome challenges in property estimation. A comprehensive review of GCMs for the prediction of thermodynamic properties of ionic liquids highlights both their potential and limitations. The integration of GCMs with Artificial Neural Networks (ANNs) demonstrates their ability to improve the prediction of liquid heat capacities over a wide range of operating conditions. An improved thermodynamic model for the prediction of liquid heat capacities of pure compounds, developed using an extensive experimental database, outperformed the GCM implemented in the Aspen Plus[®] simulator in terms of accuracy and versatility, particularly near critical points. Furthermore, comparative analysis and assessment of different approaches based on ML algorithms and GCMs were performed using an AutoML framework, identifying the best model to predict the triple point temperature of pure compounds. Gradient boosting emerged as the best-performing algorithm, achieving low modeling errors for most chemical families. However, challenges persist in estimating the triple point temperature of polar compounds such as halogens and molecules containing NO₂. The findings confirm that combining GCMs with advanced ML algorithms provides a reliable and scalable approach for predicting thermodynamic properties of pure compounds, offering alternative tools for PSE applications, and supporting the design, simulation, and optimization of processes and products.

Dedication

To my beloved grandmother, Dionisia Mercedes del Carmen Capote Chapis, whose love and wisdom continue to guide me every day. She always inspired me to pursue knowledge and persevere in my studies. Although she is no longer with us, I know she would be immensely proud of this achievement.

Acknowledgements

To my parents, for their support throughout every step of my academic journey and for encouraging me to step out of my comfort zone to achieve my dreams. Their decision to move to another country to help me has meant the world to me.

To my brother, for always taking care of me like his little sister and for his invaluable support in helping me reach my goals.

To Coco and Mile, for being by my side in this new chapter and for staying up with me during many late nights in this endeavor.

To my family and friends, whose unconditional support and love, even from a distance, have been a constant source of motivation for me. To those who are physically present, your encouragement and care have made all the difference. And to those who are far away, your understanding, warm messages, and belief in me have kept me strong.

To my religion, my godfather, godmother, and my religious family, thank you for everything.

To my doctoral project advisors, for their guidance and support at every stage of the project. I am grateful for their trust in me to carry out this work.

To TecNM / Instituto Tecnológico de Aguascalientes and CONACYT for their support through funding 786570 and for all the institutional assistance provided in the completion of this doctoral project.

Contents

Chapter 1. Introduction	1
1.1 Contextualization of this research project	1
1.2 Objectives of this project	3
1.3 Hypothesis	4
1.4 Justification	4
1.5 Overview of this thesis.....	5
Chapter 2. A review of group contribution models to calculate thermodynamic properties of ionic liquids for process systems engineering.....	7
2.1 Introduction	7
2.2 Group contribution models for the calculation of thermodynamic properties of ionic liquids	11
2.2.1 Basis and fundamentals of group contribution models	11
2.2.2 Heat capacity.....	12
2.2.3 Density.....	18
2.2.4 Viscosity.....	24
2.2.5 Thermal conductivity	28
2.2.6 Melting point	31
2.2.7 Surface tension	34
2.2.8 Electrical conductivity.....	38
2.2.9 Refractive index	40
2.2.10 Thermal decomposition temperature.....	42
2.2.11 Normal boiling point	45
2.2.12 Critical properties	47
2.2.13 Other thermodynamic properties	48
2.2.14 Estimation of thermodynamic properties using hybrid models from GCM and artificial neural networks	51
2.3 Remarks and perspectives of GCMs for ionic liquids property calculations.....	53
Chapter 3. Calculation of liquid Cp of pure compounds using an improved thermodynamic model based on group contributions and artificial neural networks.....	55

3.1 Introduction	55
3.2 Formulation of an improved GCM–ANNs model for C_p calculation of pure compounds in liquid phase	59
3.2.1 Characteristics of C_p experimental database.....	59
3.2.2 Description of the GCM–ANNs approach used for liquid C_p calculation of pure compounds.....	59
3.2.3 Evaluation of GCM–ANNs model predictions	63
3.3 Results and discussion.....	64
3.3.1 Impact of ANNs structure on the C_p calculation and identification of the best GCM–ANNs model.....	64
3.3.2 Comparative analysis of the best GCM–ANNs model with the GCM model used in the Aspen Plus® simulator.....	80
3.4 Conclusions	83
Chapter 4. Assessment of machine learning algorithms to estimate the triple-point temperature of pure compounds	84
4.1 Introduction	84
4.2 Approach for the comparison and evaluation performance of models based on machine learning algorithms to predict the triple-point temperature of pure compounds.....	86
4.2.1 Experimental database	86
4.2.2 Description of the approach based on machine learning algorithms to estimate the triple-point temperature of pure compounds.....	87
4.2.3 Description of machine learning algorithms.....	89
4.2.4 Performance evaluation metrics.....	91
4.3 Results and discussion.....	92
4.3.1 Comparative analysis of models' performance	92
4.3.2 Capabilities and limitations of the best model based on ML algorithms to estimate the triple-point temperature of pure compounds.....	100
4.4 Conclusions	109
Chapter 5. Final remarks	111
Chapter 6. References	112
Appendix A	135

<i>Appendix B</i>	139
<i>Appendix C</i>	140
<i>Appendix D</i>	141
<i>Appendix E</i>	142

List of Figures

Figure 3.1 General trend of the C_P versus the operating temperature near the critical point.	56
Figure 3.2 C_P estimations obtained for n-nonane using the correlation model proposed by Perry et al. (1997).	57
Figure 3.3 Input variables of the GCM-ANNs model for the calculation of C_P of pure compounds in liquid phase.	63
Figure 3.4 AARD for the C_P calculation of pure compounds in liquid phase using different GCM-ANNs models with symmetric ANNs configuration.	67
Figure 3.5 AARD for the C_P calculation of pure compounds in liquid phase using different GCM-ANNs models with asymmetric ANNs configuration.	68
Figure 3.6 Results of the calculation of pure compounds C_P using different GCM-ANNs models.	70
Figure 3.7 C_P calculation for ethylene and 1,1,1-trifluoroethane using the M_1 and M_2 GCM-ANNs models.	71
Figure 3.8 AARD for the C_P calculation in different T_r ranges using the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.	72
Figure 3.9 ARD_i values as a function of MW and T_r for the C_P calculation of pure compounds in liquid phase using the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.	76
Figure 3.10 C_P calculation for different molecules using the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.	77
Figure 3.11 RD_i values as a function of T_r and MW for the C_P calculation of pure compounds in liquid phase using the Aspen Plus® GCM and M_2 GCM-ANNs models.	81
Figure 4.1 Results of the calculation of T_P of pure compounds using different models based on ML algorithms.	93
Figure 4.2 Histograms of RD_i values obtained with different models based on ML algorithms for the calculation of T_P of pure compounds.	95
Figure 4.3 RD_i values as a function of MW and T_C for the calculation of T_P of pure compounds using different models based on ML algorithms.	98

<i>Figure 4.4 AARD values for the T_p calculation in two different MW ranges using different models based on ML algorithms.....</i>	<i>99</i>
<i>Figure 4.5 RD_i values as a function of T_p for the calculation of T_p of pure compounds using different models based on ML algorithms.....</i>	<i>100</i>
<i>Figure 4.6 Results of the calculation of T_p of pure compounds and histograms of RD values obtained for the training and testing subsets using the M_1 model with the GB algorithm and the random division corresponding to Set_1.....</i>	<i>102</i>
<i>Figure 4.7 AARD values for the T_p calculation in different MW ranges using the M_1 model with the GB algorithm and the random division corresponding to Set_1.....</i>	<i>103</i>
<i>Figure 4.8 Results of the calculation of T_p of pure compounds for different chemical families using the M_1 model with the GB algorithm and the random division corresponding to Set_1.....</i>	<i>106</i>
<i>Figure 4.9 Comparison of modeling errors obtained for the main functional groups in the calculation of T_p of pure compounds using the M_1 model with the GB algorithm and the random division corresponding to Set_1.....</i>	<i>109</i>

List of Tables

Table 2.1 GCMs performance for the heat capacity prediction of ILs.	16
Table 2.2 GCMs performance for the density prediction of ILs.	20
Table 2.3 GCMs performance for the viscosity prediction of ILs.	25
Table 2.4 GCMs performance for the thermal conductivity prediction of ILs.	29
Table 2.5 GCMs performance for the melting point prediction of ILs.	32
Table 2.6 GCMs performance for the surface tension prediction of ILs.	35
Table 2.7 GCMs performance for the electrical conductivity prediction of ILs.	39
Table 2.8 GCMs performance for the refractive index prediction of ILs.	41
Table 2.9 GCMs performance for the thermal decomposition temperature prediction of ILs.	43
Table 2.10 GCMs performance for the critical properties and normal boiling point prediction of ILs.	46
Table 2.11 GCMs performance for the prediction of other thermodynamic properties of ILs.	49
Table 2.12 Different GCM-ANN approaches for the estimation of thermodynamic properties of ILs.	52
Table 3.1 Functional groups considered for the GCM-ANNs model to calculate C_p of pure compounds in liquid phase.	62
Table 3.2 Results of statistical performance metrics of the best GCM-ANNs models for the C_p calculation of pure compounds in liquid phase.	69
Table 3.3 Statistical performance metrics to calculate C_p for the training, testing and validation subsets using the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.	72
Table 3.4 Statistical performance metrics to calculate C_p for different chemical families using the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.	73
Table 3.5 Statistical performance metrics to calculate C_p for the functional groups used by the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.	74
Table 3.6 Comparison of the performance of the Aspen Plus® GCM and the M_2 GCM-ANNs model for the C_p calculation of pure compounds in liquid phase.	81

<i>Table 3.7 Comparison of the performance of the Aspen Plus[®] GCM and the M₂ GCM-ANNs model for the C_P calculation of different chemical families in liquid phase.</i>	<i>82</i>
<i>Table 4.1 Functional groups defined for the development and evaluation of models based on ML algorithms to predict the T_P of pure compounds.</i>	<i>88</i>
<i>Table 4.2 Results of statistical metrics to calculate T_P with the M₁ model for the training, testing and global subsets using different random division sets.</i>	<i>101</i>
<i>Table 4.3 Results of statistical metrics to calculate T_P for different chemical families using the M₁ model with the best performance.</i>	<i>104</i>
<i>Table 4.4 Results of statistical metrics to calculate T_P for the functional groups used by the M₁ model with the best performance.</i>	<i>107</i>

Chapter 1. Introduction

1.1 Contextualization of this research project

Thermodynamic properties are physical or chemical characteristics that describe the macroscopic properties of a system (Honig, 2020). In the context of Process Systems Engineering (PSE), the knowledge of thermodynamic properties of pure compounds and mixtures is crucial for the design, modeling, simulation and optimization of processes and plants (Mann et al., 2022). The availability of reliable and accurate thermodynamic property databases supports the development of innovative and sustainable processes and products. It also facilitates resource optimization in chemical plants through mass and energy integration, contributing to the effective implementation of a circular economy. Furthermore, it enables the reduction of utility and energy consumption, prevents equipment over-sizing and enhances operating conditions (Hosseini et al., 2019; Mowbray et al., 2022).

Overall, the experimental data reported in the literature on thermodynamic properties of pure compounds is scarce and inconsistent due to measurement noise, equipment errors, and failures in equipment and/or laboratory procedures (Jovanović et al., 2011). The availability of reliable databases on thermodynamic properties is limited to well-known chemical families and for restricted operating conditions (Kolská et al., 2008; Gharagheizi and Sattari, 2010). In addition, the experimental procedures for the quantification of thermodynamic properties are usually time-consuming and expensive (Mann et al. 2022). Consequently, developing reliable thermodynamic models for predicting the physicochemical properties of pure compounds is essential in PSE.

Different thermodynamic models have been developed to estimate several thermodynamic properties of pure compounds such as group contribution models (GCMs) (Bouarab et al., 2021; Han et al., 2023; Xia et al., 2023), machine learning (ML) algorithms (Aldosari et al., 2021; Yusuf et al., 2021), quantitative structure property relationships (Hosseini et al., 2019; Sepehri, 2020; Bouarab et al., 2021), equations of state (Hosseini et al., 2019; Zhu et al., 2020), empirical and semi-empirical correlations (Conesa et al., 1998; Dadgostar and Shaw, 2012), corresponding states equations (Gao et al., 2018; Do Carmo et al., 2020), among others (Oster et al., 2018; Hosseini et al., 2019; Bouarab et al., 2021). The features, limitations and performance of these models can differ depending on the molecular structure

of compounds and the thermodynamic property being evaluated. Specifically, GCMs are considered a reliable tool commonly used for estimating a wide spectrum of thermodynamic properties of pure compounds (Hosseini et al., 2019; Chen et al., 2019). These models are distinguished by their versatility, easy implementation and acceptable accuracy for a wide range of compounds, relying on molecular structure as input information (Valderrama and Robles, 2007). On the other hand, ML algorithms have emerged as effective approaches for predicting thermodynamic properties pure compounds, due to their capabilities of handling large databases and capturing nonlinear relationships between variables, providing accurate estimations of the analyzed property (Jirasek and Hasse, 2023).

A common failure of GCMs and ML approaches for the estimation of thermodynamic properties is associated with their high prediction errors for specific chemical families and under some operating conditions. For example, the prediction of the liquid heat capacity of pure compounds (i.e., an essential property for energy transfer calculations and for the design, modeling and optimization of heating, distillation, evaporation, extraction and reaction units) near the phase-transition zones and the critical point is challenging. As a result, the models reported in the literature for the liquid heat capacity estimation may produce unreliable predictions in these regions of the phase diagram. Another basic property for PSE is the triple point temperature of pure compounds, being used in phase equilibria calculations, in the calculation of vapor solid pressure, enthalpy of fusion and solubility of solids in liquids (Gharagheizi and Sattari, 2010). However, only a limited number of studies have focused on the prediction of this thermodynamic property (Gharagheizi and Sattari, 2010). Several GCMs and ML algorithms for the thermodynamic properties calculation have been developed using databases with a limited number of experimental values, which restrict their use in predicting the properties of specific chemicals and emerging compounds. A clear example is ionic liquids, a compound family with a wide range of applications in the PSE tasks, but whose employ could be restricted due to the lack of information about their thermodynamic behavior (Thoms et al., 2017; Bouarab et al., 2021). Despite the strengths of GCMs and ML algorithms in calculating thermodynamic properties, a limited number of studies have utilized hybrid models combining both approaches for these tasks (Wang et al., 2017; Hosseini et al., 2019). Therefore, there are several challenges and opportunity areas for the development of improved thermodynamic models to accurately estimate essential

thermodynamic properties of pure compounds for PSE, especially for chemical families and properties that have not been extensively studied in literature.

In this research, the limitations and capabilities of GCMs for predicting a set of thermodynamic properties of ionic liquids were revised and analyzed. This allows to identify and select the best alternatives for predicting different thermodynamic properties of these compounds. In addition, the main key issues that should be covered to improve the performance of such models were highlighted. The synergy obtained by combining GCMs and artificial neural networks (ANNs) was utilized in the development of an improved thermodynamic model to predict the liquid heat capacity of several chemical families of pure compounds under a wide range of operating conditions. In addition, different approaches based on ML algorithms and the concept of GCMs were evaluated and compared to identify the best hybrid model for estimating the triple point temperature of pure compounds. The development and performance evaluation of these improved models were analyzed in terms of their advantages and gaps for the analyzed properties estimation. This study aims to offer valuable insights into the development of improved and reliable thermodynamic models for the tasks of PSE.

1.2 Objectives of this project

The main objective of this research project was to develop improved thermodynamic models integrating GCMs and ML algorithms approaches to enhance the prediction of thermodynamic properties, supporting innovations in the design, simulation and optimization of processes and products within the framework of PSE. Therefore, the specific objectives were:

- i. To analyze and evaluate the application, performance and accuracy of different GCMs and GCMs-ANNs approaches for estimating the ionic liquids thermodynamic properties, highlighting the strengths and gaps of these models, as well as the main challenges and perspectives for improving their predictive capabilities for PSE.
- ii. To develop an improved GCMs-ANNs model for estimating the liquid heat capacity of pure compounds over a broad temperature range and comparing its performance with other models reported in the literature for the liquid heat capacity estimation.

iii. To develop and assess different models based on ML algorithms and GCMs for predicting the triple point temperature of pure compounds using an AutoML approach, emphasizing the limitations of capabilities of these models for different chemical families and functional groups, while considering the impact of molecular properties on the property estimation.

1.3 Hypothesis

The following hypothesis was tested and proved in this thesis: The enhancement of predictions for essential thermodynamic properties in PSE can be achieved through the development of improved thermodynamic models based on GCMs and ML algorithms approaches.

1.4 Justification

PSE involves design, modeling, simulation and optimization of processes and products. The availability of reliable thermodynamic properties databases is paramount to accomplish these tasks. However, there is a limitation related to the scarcity and uncertainty of experimental data on thermodynamic properties reported in the literature. Different thermodynamic models have been developed to estimate these properties. Specifically, GCMs and ML algorithms offer several advantages to predict thermodynamic properties of pure compounds. However, these models may exhibit high modeling errors for certain chemical families and under specific operating conditions. Several GCMs and ML algorithms for the estimation of thermodynamic properties of pure compounds have been developed using a limited amount of experimental data, restricting their applicability. Therefore, the thermodynamic properties of novel and emerging compounds cannot be predicted using these models. Ionic liquids serve as a prominent example, since the implementation of GCMs and ML algorithms to estimate their thermodynamic properties is affected by the relatively limited amount of experimental data compared to the data contained in databases for traditional chemical compounds. Liquid heat capacity and triple point temperature of pure compounds are essential thermodynamic properties for the design, modeling and optimization of heating, evaporation, distillation and extraction processes. However, it has been identified that these properties are challenging to predict, either due to the limited

number of models reported in the literature for their estimation (e.g., in the case of triple point temperature) or by the high modeling errors under specific operating conditions (e.g., in the case of liquid heat capacity). To overcome the limitations in thermodynamic property estimation it is necessary to develop improved models that can estimate the thermodynamic properties of a wide range of molecules and operating conditions with low predicting errors.

This study is a step forward in that direction since the main objective was to develop improved thermodynamic models for accurate thermodynamic property estimation in order to support innovations in the design, simulation and optimization of processes and products within the framework of PSE. The gaps and potentials of GCMs for estimating thermodynamic properties of ionic liquids were highlighted, emphasizing the main issues to consider when improving such models. Different improved models based on the concept of GCMs and ML algorithms were developed and assessed for the estimation of liquid heat capacities and triple point temperatures of pure compounds, both fundamental thermodynamic properties for PSE. The limitations and capabilities of these models in relation to several chemical families and functional groups were emphasized, considering the impact of other molecular properties on the estimation of the analyzed properties.

1.5 Overview of this thesis

This thesis comprises five additional chapters. Specifically, Chapter 2 contains information about different GCMs reported and applied to estimate the thermodynamic properties of ionic liquids. The application, performance, and accuracy of these models for predicting the ionic liquids properties were analyzed and discussed. Some approaches combining GCMs with ANNs to estimate the thermodynamic properties of ionic liquids are also described and exemplified.

Chapter 3 reports the development, testing, and comparison of an improved thermodynamic model for the calculation of the liquid heat capacity of pure compounds by leveraging the synergy between GCMs and ANNs. An extensive database was used for the model development. Different input variables and ANNs configurations were evaluated to identify the best thermodynamic model for calculating this thermodynamic property. The performance of this improved model was characterized in terms of its capabilities and limitations for C_p calculation of different chemical families with a discussion of the impact

of the molecular weight and reduced temperature. This new model was compared with the Cp model used by the Aspen Plus[®] process simulator and the results showed its superior capabilities and estimations.

The development and assessment of several models based on ML algorithms to estimate the triple-point temperature of pure compounds using the AutoML approach available in the auto-sklearn library in Python were detailed in Chapter 4. Different input variables and ensembles of the ML algorithms were assessed. The limitations and gaps in the models are highlighted for several chemical families and functional groups. The results demonstrated that the algorithm of gradient boosting achieved the best performance in estimating this thermodynamic property. Estimating the triple-point temperature was challenging for certain chemical families, including polar compounds such as halogens and those containing NO₂. This study represents an initial step toward the development of improved models for accurately predicting this thermodynamic property.

Chapter 5 provides the conclusions of this study and Chapter 6 lists all the references used to support the content, analysis and interpretation presented in this thesis. Finally, this thesis includes appendices that provide supplementary information and additional analyses to support the findings and discussions presented in the main chapters.

Chapter 2. A review of group contribution models to calculate thermodynamic properties of ionic liquids for process systems engineering

2.1 Introduction

Ionic liquids (ILs) are compounds with a low melting point (usually $< 100\text{ }^{\circ}\text{C}$) like conventional salts but composed of organic cations and inorganic or organic anions (Brinchi et al., 2003; Hosseini et al., 2019). Common examples of ILs cations include imidazolium, sulfonium, morpholinium, guanidinium, ammonium, pyridazinium, piperazinium, oxazolidinium, tetrazolium, morpholinium, uronium, thiazolium, isoquinolinium, pyrroline, piperidinium, pyrrolidinium, pyridinium and phosphonium (Hosseini et al., 2019; Sepehri, 2020). The anions that frequently conform the ILs molecular structure are BF_4^- , PF_6^- , CF_3SO_3^- , $\text{N}(\text{SO}_2\text{CF}_3)_2^-$, SCN^- , R-PO_4^- , R-SO_3^- and R-SO_4^- where R represents an alkyl substituent (Hosseini et al., 2019).

ILs have attractive thermochemical properties for various applications (Feng et al., 2015; Thoms et al., 2017). They are versatile chemicals due to their low vapor pressure, high solubility for polar and non-polar organic and inorganic substances, thermal stability at high temperature, aggregation liquid state at room temperature, low melting temperature, besides their reusability, solvation, non-flammable and non-corrosive characteristics (Rooney et al., 2010; Ghandi, 2014; Ayan et al., 2020; Chen et al., 2020; Sepehri, 2020; Bouarab et al., 2021). The physicochemical properties of ILs can be tailored for different processes by varying the cations and anions that conform their molecular structures (Bouarab et al., 2021). Theoretically, it is believed that a vast number of ILs could be synthesized with tuned and enhanced properties for the application at hand (Chen et al., 2019).

To date, ILs have been employed to develop novel processes and products thus contributing to the industry sustainability and circular economy, see Figure A.1 and Table A.1 of Appendix A. One of the most relevant applications of these compounds is as green solvents due to their low vapor pressure that allows to reduce the energy consumption and waste generation in purification and separation units (Wang et al., 2009; Zaid et al., 2017). They can be utilized as an alternative to the well-known volatile organic solvents in the chemical industry to develop environmentally friendly technologies and processes. ILs can dissolve organic and inorganic chemicals (Xiong et al., 2013; Liu et al., 2014; Ayan et al.,

2020; Aguilar-Elguezabal et al., 2020), pharmaceuticals (Lee and Lin, 2017), biochemical compounds (Devi et al., 2011; Ortiz-Martínez et al., 2016; Xu et al., 2016; Brogan et al., 2018) and polymeric materials (Valderrama and Robles, 2007; Meksi and Moussa, 2017). For example, Ayan et al. (2020) used ILs for the reactive extraction of propionic acid, while Aguilar-Elguezabal et al. (2020) developed a new process to transform α -pinene into α -terpineol using an IL-based solvent. ILs have been also utilized as extraction agents in distillation processes offering several advantages like a simple recovery, high selectivity and reduction of mass losses (Lei et al., 2014; Qi et al., 2019; Sosa et al., 2020). Qi et al. (2019) evaluated the performance of ethylene glycol and 1,3 – dimethylimidazolium in the extractive distillation of the azeotropic mixture of diisopropyl ether and isopropanol. This IL improved the relative volatility of these compounds and increased the purities obtained from the separation system in comparison to those obtained with ethylene glycol. ILs have also proved to improve the performance of liquid-liquid extraction processes (Marsousi et al., 2019; Sosa et al., 2020; Sommer et al., 2021; Guo et al., 2021). Sosa et al. (2020) evaluated the use of different ILs to separate a mixture of hydrocarbons and ethanol obtaining a better process performance. Guo et al. (2021) carried out a directional solvent extraction with IL for the seawater desalination where a higher water yield was obtained than that generated with the current process using decanoic acid. Another important application of ILs includes catalysis and reaction engineering (Welton, 1999; Adams et al., 2000; Brinchi et al., 2003; Yadav et al., 2003; Irimescu and Kato, 2004; Fukuyama et al., 2007; Arai et al., 2010; Böhm et al., 2014; Lentini et al., 2014; Soares, 2018; Hosseini et al., 2019). Reactive systems with ILs can provide higher activity than their counterparts besides these compounds can facilitate the formation of two phases thus helping the recovery of target compound(s) (e.g., the separation of fatty acid methyl esters during biodiesel production). For instance, Arai et al. (2010) produced biodiesel from soybean oil catalyzed by lipase-producing fungi using four types of imidazolium-based ILs. For interested readers, Vekariya (2017) has reviewed the properties, preparation routes and applications of ILs in catalytic organic transformations. Note that different ILs could show attractive characteristics (e.g., lubricating properties) to be exploited in diverse industrial applications (Ye et al., 2001; Karmakar and Samanta, 2002; Predel et al., 2009).

The design, simulation and optimization of ILs-based processes require a reliable thermodynamic modeling. The experimental and theoretical characterization of the thermodynamic properties of ILs (e.g., heat capacity, density, viscosity, thermal conductivity, melting point, surface tension, electrical conductivity, refractive index, thermal decomposition temperature, normal boiling point, critical properties, freezing point, isobaric expansivity, isothermal compressibility, static dielectric constant) is crucial for the process systems engineering. The availability of a comprehensive database of ILs thermodynamic properties allows to improve the operating conditions, to develop new sustainable processes and products, to reduce the energy requirements, to avoid the equipment oversizing, to enhance other process performance metrics, and to validate theories and models. Consequently, it is fundamental to carry out experimental studies on the quantification of these properties with the aim of supporting the development of a reliable thermodynamic framework for the calculation, as accurately as possible, of ILs physicochemical behavior. Reliable estimations of ILs thermodynamic properties are required to face and handle the different tasks involved in the process design and simulation. However, the experimental data reported for ILs thermodynamic properties are still scarce and some of them could be inconsistent or uncertain due to noise, determination errors and undetected failures in equipment and/or laboratory procedures. Due to the large number of ILs that can be synthesized for a particular application (e.g., it has been estimated that $\sim 10^6$ possible ILs can be synthesized), it could be complex, expensive and time-consuming to quantify experimentally the properties of all these compounds. The development of reliable predictive thermodynamic models to calculate the physicochemical properties of ILs is paramount for PSE.

Thermodynamic models for the estimation of ILs physicochemical properties include GCMs (Aguirre et al., 2012; Lazzús, 2012a, 2012b; Nancarrow et al., 2015; Sattari et al., 2016a; Lazzús et al., 2017b; Shahsavari et al., 2018; Hosseini et al., 2019; Bouarab et al., 2021), quantitative structure property relationships (QSPR) (El-Harbawi et al., 2014; Hosseini et al., 2019; Sepehri, 2020; Bouarab et al., 2021), ANNs (Valderrama et al., 2011a, 2014; Hosseini et al., 2019; Yusuf et al., 2021), least square support vector machine (LSSVM) (Yusuf et al., 2021), equations of state (EoS) (Gadamsetty et al., 2011; Hosseini et al., 2019), molecular dynamics (MD), Monte Carlo (MC) simulations (Vega et al., 2010;

Wang et al., 2014; Weiss, 2015; Hosseini et al., 2019) and their variants and hybrid approaches (Deetlefs et al., 2006; Matsuda et al., 2007; Lazzús, 2010b; Valderrama and Rojas, 2010; Vega et al., 2010; Shen et al., 2011; Valderrama et al., 2011b, 2017; Chen et al., 2013; Wu et al., 2014; Farzi and Esmailzadeh, 2016; Wang et al., 2017; Hosseini et al., 2019; Bouarab et al., 2021). Figure B.1 of Appendix B provides an overview of these modeling tools. The characteristics, limitations and capabilities of these models can vary substantially depending on the ILs molecular structure and thermodynamic property to be analyzed. However, GCMs stand out as effective and widely used approach to estimate the thermodynamic behavior of ILs (Hosseini et al., 2019; Chen et al., 2019).

GCMs offer advantages for the prediction of ILs thermodynamic properties. These models are characterized by their acceptable accuracy without requiring substantial computational efforts and complex calculations, versatility, straightforward and easy implementation for a broad spectrum of compounds where the knowledge of their molecular structures is usually the input information (Valderrama and Robles, 2007). These models assume that the properties of a compound are a function of its molecular structural elements and tested operating conditions. GCMs can be utilized as thermodynamic models for the PSE involving ILs (Valderrama et al., 2012; Hosseini et al., 2019). However, it is convenient to indicate that the main issue that can affect the use and implementation of GCMs for the prediction of ILs thermodynamic properties is the relatively limited amount of experimental data in comparison with the information contained in the thermodynamic database of traditional chemical compounds. Despite this fact, GCMs are still considered as a straightforward (or even the best) alternative to predict the physicochemical properties of ILs.

This chapter covers a diverse set of GCMs that have been reported and applied to predict thermodynamic properties of ILs that are relevant for the design, simulation, optimization and control of separation systems and reaction units. A detailed description of these GCMs is provided thus highlighting their application in the calculation of different ILs thermodynamic properties. This chapter provides to interested readers an easy and comprehensive survey of GCMs for identifying and selecting the best alternatives to estimate several physicochemical properties of ILs.

2.2 Group contribution models for the calculation of thermodynamic properties of ionic liquids

2.2.1 Basis and fundamentals of group contribution models

In GCMs, the thermodynamic property estimation is carried out considering the contributions and frequencies of the elements (e.g., atoms, functional groups, ions) that conform the IL molecular structure (Valderrama et al., 2012; Gani, 2019). They assume that the contribution of each element for the thermodynamic property calculation should be the same in any molecule independently of its structure (Nannoolal et al., 2007). Figure C.1 of Appendix C illustrates the concept of GCMs for the prediction of a given thermodynamic property using an IL molecule as example. The molecular element contributions of the model are calculated from an experimental database of the desired thermodynamic property that contains similar compounds from the target chemical family.

The development and implementation of GCMs consists of three stages: a) the definition of elements (i.e., atoms, functional groups, ions) to obtain the molecular structure of target compounds or chemical family, b) the determination of equation(s) and functionalities to calculate and estimate the thermodynamic property based on the type of structural elements and their frequencies including other additional molecular parameters if any and c) the stage of numerical calculation of structural element contribution values for the thermodynamic property at hand, which usually implies a nonlinear regression process using an experimental database (Su et al., 2017). In the first stage, the target compounds are divided into their structural elements considering their molecular complexity and available experimental information. Then, an equation or model (explicit or implicit) is defined to establish the functionality of the thermodynamic property to be calculated with respect to these structural elements and their frequencies. This equation/functionality can also include other parameters or characteristics available for the compound under analysis (e.g., molecular weight, number of atoms, bond types) and also state variables (e.g., temperature, pressure). Finally, the values of structural element contributions to be used in the equation/functionality are determined by data regression techniques using experimental values of the target thermodynamic property. This last stage implies the resolution of a parameter identification problem where a proper objective function should be minimized via the implementation of a local or global optimization method.

GCMs can be classified into additive, additive-correlative, interaction-based and high order models (Ghasemitabar and Movagharnejad, 2016). Additive GCMs assume that the property depends on the number of structural elements, while the additive-correlative GCMs introduce additional known molecular properties into the model. Interaction-based GCMs calculate the property using additional parameters that describe the interactions between the structural elements that conform the molecule. Higher-order GCMs introduce variables to describe the spatial distribution of structural elements of tested molecules and can also contain first-order groups.

Different GCMs have been reported and applied to calculate the physicochemical properties of ILs and a summary of their applications is given in Tables 2.1–2.11. A detailed description of different GCMs for the estimation of relevant thermodynamic properties of ILs is provided in the next subsections thus discussing its main characteristics, application range and results obtained from their implementation. Two statistical metrics have been considered in the discussion of GCMs performance: the average absolute relative deviation ($AARD$, %) and absolute relative deviation (ARD_i , %), which are given by

$$AARD = \frac{100 \cdot \sum_{i=1}^{N_{dat}} |(X^{cal} - X^{exp})/X^{exp}|_i}{N_{dat}} \quad (2.1)$$

$$ARD_i = 100 \cdot |(X^{cal} - X^{exp})/X^{exp}|_i \quad (2.2)$$

where X^{cal} is the calculated property value, X^{exp} is the experimental property value and N_{dat} is the number of experimental data analyzed in the corresponding study.

2.2.2 Heat capacity

Heat capacity is a basic input data in process simulation software for energy transfer and thermodynamic calculations like entropy and enthalpy changes of reactive and non-reactive systems (Valderrama et al., 2011c; Chen et al., 2019). It is defined as the dependence of enthalpy with respect to the temperature of a given compound (Gardas and Coutinho, 2008a). This property is essential for the design and scale up of distillation, evaporation, extraction and heating systems (Valderrama et al., 2011c; Nancarrow et al., 2015). The variation in heat capacity allows to understand the changes in the compound structure and also provides information on phase changes (Gardas and Coutinho, 2008a). The experimental information of ILs heat capacities is relatively scarce considering the vast amount of these compounds

where the available data have been reported for specific ILs (Gardas and Coutinho, 2008a). Different GCMs for calculating the heat capacity of ILs are summarized in Table 2.1.

Gardas and Coutinho (2008a) developed a second-order GCM to estimate the liquid heat capacity (C_p , J/mol·K) of ILs based on pyrrolidinium, pyridinium and imidazolium cations and trifluoromethane sulfonate, ethyl sulfate, bromide, bis (trifluoromethanesulfonyl) amide, tetrafluoroborate and hexafluorophosphate anions. This GCM was formulated employing the approach proposed by Růžička and Domalski (1993a) (1993b) for the estimation of pure organic liquids heat capacities as a function of temperature. This model is described by the following equations

$$C_p = R \left[A_1 + A_2 \left(\frac{T}{100} \right) + A_3 \left(\frac{T}{100} \right)^2 \right] \quad (2.3)$$

$$A_1 = \sum_{i=1}^k n_i a_{1i} \quad (2.4)$$

$$A_2 = \sum_{i=1}^k n_i a_{2i} \quad (2.5)$$

$$A_3 = \sum_{i=1}^k n_i a_{3i} \quad (2.6)$$

where T is the temperature (K), n_i is the frequency of the structural element i contained in the ILs molecule, k is the total number of different structural elements, a_{1i} , a_{2i} and a_{3i} are the structural element contributions to the model parameters A_i and $R = 8.314472$ J/mol·K is the universal gas constant. This model is capable to predict ILs heat capacities from 264.8 to 825 J/mol·K at 196.36–663.10 K with $AARD = 0.36$ %, see Table 2.1. However, this model was limited to estimate the property for a few ILs families where a total of 2396 experimental data points from 19 ILs including 6 anions and 3 cations were used to develop this GCM. To overcome this drawback, Chen et al. (2019) extended this model to more families of ILs where 3304 data points at 189.66–524.87 K from 61 ILs (based on 6 cations and 14 anions) were utilized. The modeling errors obtained with this GCM were similar to those obtained with the Gardas and Coutinho (2008a) model (i.e., $ARD_i < 5$ %).

GCM proposed by Joback (1984) for the prediction of ideal gas heat capacities was modified to estimate ILs heat capacities using a corresponding states principle (CSP) (Ge et al., 2008). The authors included 3 additional groups (B, P and -SO₂-) commonly found in ILs. The database utilized to develop this model consisted of 961 heat capacity data points

from 53 ILs (e.g., [C_nmim][OTf], [C₄mim][BF₄], [C₄mim][PF₆], [C₄mpyrr][FAP]). The ideal gas heat capacities ($C_{p_{ig}}$, J/mol·K) of ILs as a function of temperature were calculated by

$$C_{p_{ig}} = \left[\sum_i n_i B_{1_{Cpi}} - 37.93 \right] + \left[\sum_i n_i B_{2_{Cpi}} + 0.210 \right] T + \left[\sum_i n_i B_{3_{Cpi}} - 3.91 \cdot 10^{-4} \right] T^2 + \left[\sum_i n_i B_{4_{Cpi}} + 2.06 \cdot 10^{-7} \right] T^3 \quad (2.7)$$

where $B_{i_{Cpi}}$ is the corresponding structural element contribution. The liquid heat capacities C_p were estimated using CSP (Poling et al., 2001) as a function of $C_{p_{ig}}$ and other thermodynamic properties of ILs as follows

$$\frac{C_p - C_{p_{ig}}}{R} = 1.586 + \frac{0.49}{1 - T_r} + \omega \left[4.2775 + \frac{6.3(1 - T_r)^{1/3}}{T_r} + \frac{0.4355}{1 - T_r} \right] \quad (2.8)$$

$$T_r = \frac{T}{T_c} \quad (2.9)$$

where ω is the acentric factor, T_r is the reduced temperature and T_c is the ILs critical temperature. The main drawback of this model is that the values of normal boiling point (T_b) and critical properties (T_c , P_c , V_c) of ILs are needed and, consequently, its application could be restricted due to the lack of experimental information on these properties for several ILs. Therefore, these authors proposed to estimate the missing thermodynamic properties via the GCM proposed by Valderrama and Robles (2007), see Sections 2.2.11 and 2.2.12 of this chapter.

Soriano et al. (2010a) proposed a GCM approach using 3149 heat capacity data from 32 ILs. C_p of ILs was estimated by adding the cation and anion contributions as follows

$$C_p = C_{p,cation} + C_{p,anion} \quad (2.10)$$

$$C_{p,ion} = C_1 + C_2 \cdot \left(\frac{T}{K} \right) + C_3 \cdot \left(\frac{T}{K} \right)^2 \quad (2.11)$$

where $C_{p,ion} = C_{p,cation}$ or $C_{p,anion}$ are the contributions to the molar heat capacity of cation and anion that conformed the tested IL and C_i is an adjustable empirical parameter specific for each ion. The major limitation of this model relies on its applicability to a reduced number of IL molecules that can be obtained from 10 cations and 14 anions that were defined for the database of structural element contributions. Müller and Albert (2014) proposed a similar GCM to calculate the heat capacity of more ILs. This property can be estimated for 1248 pure ILs using this model via the contributions from 39 cations and 32 anions. 2443 heat capacity experimental data from 104 ILs were used to develop this GCM. The modeling

errors obtained in the correlation stage were acceptable with $AARD = 1.4 \%$. The modeling errors increased to predict heat capacities of ILs not used in the data correlation stage (i.e., $AARD = 4.4 \%$), see Table 2.1. The model of these authors was based on the cation and anion contributions (Albert and Müller, 2014), which was defined as

$$C_p = (D_{1cation} + D_{1anion}) - (D_{2cation} + D_{2anion}) \cdot T + (D_{3cation} + D_{3anion}) \cdot T^2 \quad (2.12)$$

$$D_{1ion} = \sum_i (v_i \cdot D_{1i}) + D_{1correction} \quad (2.13)$$

$$D_{2ion} = \sum_i (v_i \cdot D_{2i}) + D_{2correction} \quad (2.14)$$

$$D_{3ion} = \sum_i (v_i \cdot D_{3i}) + D_{3correction} \quad (2.15)$$

where v_i is the frequency of structural element i contained in the ion, D_{1i} , D_{2i} and D_{3i} are the structural element contributions and $D_{icorrection}$ corresponds to additional contributions related to the ion type or ring structure. The heat capacity database included 2419 experimental data from 106 ILs covering a temperature range of 190–663.1 K.

Another GCM for predicting the heat capacities of ILs containing alkyl-ammonium, phosphonium, pyrrolidinium, pyridinium and imidazolium was reported by Valderrama et al. (2011c). This model included an additional structural parameter known as mass connectivity index (φ), which was studied by Valderrama and Rojas (2010). This parameter considers the type of connections between groups and their molecular mass

$$\varphi = \sum \left(\frac{1}{\sqrt{m_i m_j}} \right)_{ij} \quad (2.16)$$

where m_i and m_j are the mass of neighboring groups i and j in a given IL molecule. The functionality of ILs heat capacity with temperature was described as

$$C_p = C_{p0} - 100.70\varphi + \varphi(0.3918T - 1.952 \cdot 10^{-4}T^2) \quad (2.17)$$

$$C_{p0} = 9.0045 + n_i E_{0i} \quad (2.18)$$

where C_{p0} (J/mol·K) is the heat capacity reference value determined by GCM using data of C_p at 298 K only for ILs that are truly liquids at this operating condition, E_{0i} is the contribution value of structural element i and T_0 is the reference T (i.e., 298.15 K). This GCM

Table 2.1 GCMs performance for the heat capacity prediction of ILs.

Modeling performance ¹					
<i>Model</i>	<i>N_{IL}</i>	<i>N_{dat}</i>	<i>AARD, %</i>	<i>R²</i>	<i>Remarks</i>
GCM					
Gardas and Coutinho (2008a)	19	2396	0.36	0.99	Global set
Soriano et al. (2010a)	32	3149	0.69	N/A	Global set
Valderrama et al. (2011c)	32	469	2.80	N/A	Correlation set
	9	65	2.60	N/A	Prediction set
Müller and Albert (2014)	84	2234	1.40	N/A	Correlation set
	20	209	4.40	N/A	Prediction set
Albert and Müller (2014)	86	2210	2.58	0.98	Correlation set
	20	209	5.44	0.94	Prediction set
Ahmadi et al. (2015)	100	4072	5.83	N/A	Correlation set
	28	750	5.61	N/A	Prediction set
Nancarrow et al. (2015)	92	2485	6.76	N/A	Lego approach
	45	1586	5.54	N/A	Lego approach
	45	1586	5.30	N/A	Meccano approach
Chen et al. (2019)	44	2391	0.39	N/A	Correlation set
	17	913	0.62	N/A	Prediction set
GCM-CSP					
Ge et al. (2008)	53	961	2.91	N/A	Global set
GCM-GFA					
Sattari et al. (2014a)	82	3726	1.68	0.99	Global set
	65	2939	1.68	0.99	Correlation set
	17	787	1.65	0.99	Prediction set
AI (GMDH)					
Rostami et al. (2019)	56	2940	1.84	0.98	Global set
¹ Variable description:			<i>R²</i> : Coefficient of determination.		
<i>N_{IL}</i> : Number of ILs used in the study.			N/A: Data not available.		
<i>N_{dat}</i> : Number of experimental data analyzed in the study.					

was formulated via the processing and analysis of 534 heat capacity data (368–1368 J/mol·K) at 250–426 K from 41 ILs. 126 compounds were employed for the correlation stage of this model. It was reported $AARD < 3\%$ for the ILs heat capacity estimation with this approach. However, its main limitation relies on the knowledge of a reference heat capacity value, which is sometimes not available especially for new ILs.

Ahmadi et al. (2015) proposed an atomic-based GCM that can be applied in a wide range of temperatures (190–663.10 K) and heat capacities (93–1764 J/mol·K). This model is given by the next equation

$$C_p = 0.2802T^{1.0854} - 17.5066 \ln(T) + 0.6593W^{1.0793} + 15.9932Nu_{Ca} + 16.1292Nu_{Cc} - 2.8956Nu_N - 11.7667Nu_S - 1.3729Nu_O - 4.1977(Nu_F + Nu_{Br} + Nu_{Cl}) + 6.7438Nu_B - 12.3990 \quad (2.19)$$

where Nu_{Ca} , Nu_{Cc} , Nu_N , Nu_S , Nu_O , Nu_F , Nu_{Br} , Nu_{Cl} and Nu_B are the number of carbon atoms in anion, the carbon atoms in cation, the number of atoms of nitrogen, sulfur, oxygen, fluorine, bromine, chlorine and boron in the IL, respectively. These authors used 4822 heat capacity data points from 128 ILs to develop this GCM. Results showed that this model was outperformed by the model of Müller and Albert (2014). However, the advantage of this GCM relies on its wide application for different ILs since the molecules are represented only as atomic-level contributions.

Nancarrow et al. (2015) implemented and compared the accuracy of two GCMs based on Meccano and Lego approaches. This classification relies on size and nature of the structural building blocks or structural elements employed in each model. Meccano approach uses large functional groups such as ions, e.g.: the model of Gardas and Coutinho (2008a). This type of models is generally more accurate, see Table 2.1, but its application range is limited. On the other hand, the Lego approach is based on the use of simple and small structural elements. The model reported by Ge et al. (2008) falls in this category and is an illustrative example of this approach. These models are less accurate than those based on the Meccano approach. However, they can be applied to construct more IL molecules in comparison to the Meccano-based GCMs. For illustrative purposes, Figure D.1 of Appendix D exemplifies the structural elements considered in the Meccano and Lego approaches. These authors used a refined database containing 2642 experimental data from 96 ILs to evaluate GCMs based on both approaches. This database was divided into two sub-databases: one containing 2584 data

from 92 ILs to implement the Lego approach and another sub-database constituted by 1586 data from 45 ILs that were used with both Meccano and Lego approaches. As expected, the Meccano-based GCM outperformed the predictions obtained with the Lego-based GCM, see results given in Table 2.1.

Other thermodynamic models based on the combination of GCMs with other numerical tools have been reported to estimate C_p of ILs. For instance, Sattari et al. (2014a) used a GCM combined with genetic function approximation (GFA) to estimate C_p of imidazolium-, phosphonium-, pyridinium-, pyrrolidinium- and ammonium-based ILs consisting of 39 cations and 24 anions at atmospheric pressure. The database employed for the model development consisted of 3726 heat capacity experimental data from 82 ILs covering 180.06–663 K and heat capacity values from 259.09 to 1805.7 J/mol·K. GFA is a feature selection method proposed by Rogers and Hopfinger (1994) that allowed to choose the most effective functional groups for the prediction of this thermodynamic property. A hybrid group method of data handling (GMDH), which can be considered as a version of neural network-based system, was proposed to estimate the heat capacity of ILs at 188.06–663.10 K (Rostami et al., 2019). 2940 heat capacity data (226–1413 J/mol·K) from 56 ILs were used to develop this model. The input parameters were T , the number of methyl groups in the IL cation, the number of hydrogen atoms in the anions and the number of atoms in both IL cation and anion. It showed $AARD = 1.84\%$ to predict the ILs heat capacities.

2.2.3 Density

Density is a relevant property of ILs that is required to estimate process equipment sizes, to solve material/energy balance equations and to correlate other thermophysical properties like surface tension and viscosity (Gardas and Coutinho, 2008c; Mendonça et al., 2012; Paduszyński and Domańska, 2012). It has been estimated that the experimental values of density for more than 1000 ILs is available in databases, which have been used to develop models for its estimation including GCMs (Paduszyński and Domańska, 2012; Keshavarz et al., 2016b).

Ye and Shreeve (2007) developed a simple GCM to calculate the densities (ρ , g/cm³) at room-temperature of ILs containing imidazolium, triazolium, tetrazolium and tetraalkylammonium cations and tetrafluoroborate, hexafluorophosphate,

bis(trifluoromethanesulfonyl) amide, nitrate, perchlorate and dinitramide anions. This model is given by the next expression

$$\rho = \frac{W}{NV} \quad (2.20)$$

where W is the ILs molar mass (g/mol), N is the Avogadro constant and V is the IL molar volume (cm³/mol). V was assumed to be a linear function of the volumes of cation (V_+) and anion (V_-) as described by Jenkins et al. (1999) (i.e., for a specific CA_pANI_q IL, $V = pV_+ + qV_-$). This GCM was limited to estimate the densities of ILs at atmospheric pressure and 298.15 K. To overcome this disadvantage, Gardas and Coutinho (2008c) modified this model to expand its application range at 273.15 – 393.15 K and 0.10 – 100 MPa. They assumed that the mechanical coefficients (i.e., isobaric expansivity and isothermal compressibility) of the ILs involved in their study were similar and also constant over a wide range of temperature and pressure. Therefore, they proposed the next expression to calculate ρ (kg/m³) as a function of temperature (T , K) and pressure (P , MPa)

$$\rho = \frac{W}{NV_0(8.005 \cdot 10^{-1} \pm 2.333 \cdot 10^{-4} + 6.652 \cdot 10^{-4} \pm 6.907 \cdot 10^{-7} K^{-1}T - 5.919 \cdot 10^{-4} \pm 2.410 \cdot 10^{-6} MPa^{-1}P)} \quad (2.21)$$

where V_0 is the molar volume (m³/mol) at the reference temperature (T_0 , K) and reference pressure (P_0 , MPa) and W is given in kg/mol. 1086 experimental data from 12 imidazolium-based ILs and 435 experimental data from 8 phosphonium-, pyridinium- and pyrrolidinium-based ILs were used to develop this GCM. It showed *AARD* values from 0.45 % to 1.57 % for the density calculation of pyrrolidinium-, pyridinium-, phosphonium- and imidazolium-based ILs, see Table 2.2. Aguirre and Cisternas (2012) extended this model for the estimation of ammonium-based ILs. 108 experimental data from 20 ammonium-based ILs were used for the correlation of Equation (2.21). *AARD* of this model ranged from 0.07 % to 5.09 % for these ILs.

A simple GCM to predict V_0 (m³/kmol) and a linear relation to calculate of ILs at 258–393 K and 0.09–207 MPa were proposed by Lazzús (2010a), which are defined as

$$V_0 = 0.042 + \sum_i n_i \Delta V_{ci} + \sum_j n_j \Delta V_{aj} \quad (2.22)$$

$$\rho_0 = W \cdot V_0^{-1} \quad (2.23)$$

$$\rho = \rho_0 - 0.7190(T - T_0) + 0.5698(P - P_0) \quad (2.24)$$

Table 2.2 GCMs performance for the density prediction of ILs.Modeling performance ¹

<i>Model</i>	N_{IL}	N_{dat}	<i>AARD, %</i>	R^2	<i>Remarks</i>
GCM					
Gardas and Coutinho (2008c)	12	1086	0.45	N/A	Imidazolium-based ILs.
	8	435	1.43	N/A	Pyridinium-, pyrrolidinium- and phosphonium-based ILs.
Lazzús (2010a)	310	310	1.93	0.97	Global set (V_0 and ρ_0 estimation)
	210	210	1.90	0.97	Correlation set (V_0 and ρ_0 estimation)
	100	100	1.98	0.95	Prediction set (V_0 and ρ_0 estimation)
	76	3530	0.73	0.99	Density as a function of T and P estimation
Qiao et al. (2010)	120	6562	0.88	N/A	Correlation set
	3	188	0.27	N/A	Prediction set
Aguirre and Cisternas (2012)	20	108	1.57	N/A	Ammonium-based ILs
Paduszyński and Domańska (2012)	1028	16830	0.51	N/A	Global set
	828	13135	0.53	N/A	Correlation set
	200	3695	0.45	N/A	Prediction set
Evangelista et al. (2014)	864	21845	0.83	N/A	Global set
Taherifard and Raeissi (2016)	N/A	25850	0.95	N/A	Global set
	N/A	12516	1.02	N/A	Correlation set
	N/A	13334	0.89	N/A	Prediction set
Chen et al. (2019)	143	7360	0.49	N/A	Global set
	90	5039	0.43	N/A	Correlation set
	53	2321	0.67	N/A	Prediction set

¹ Variable description: N_{IL} : Number of ILs used in the study. R^2 : Coefficient of determination. N_{dat} : Number of experimental data analyzed in the study.

N/A: Data not available.

where ρ_0 (kg/m³) is the density at $T_0 = 298.15$ K and $P_0 = 0.101$ MPa, W is given in kg/kmol, ΔV_c and ΔV_a are the structural element contributions for the cation and the anion to the molar volume, respectively. This GCM was formulated with an experimental database that included the molar volume of 310 ILs at 298.15 K and 0.101 MPa. 92 ion groups were defined: 26 cation groups (e.g., pyrazolium, sulfonium, ammonium, phosphonium, guanidinium, oxazolidinium, morpholinium, piperidinium, pyrrolidinium, pyridinium and imidazolium) and 66 anion groups (e.g., metal complexes, trifluoroacetates, triflates, carboxylates, phosphates, borates, amides, imides, tosylates, sulfonates, sulfates, pseudohalides and halides). ρ was calculated as a linear function of T and P via Equation (2.24) where the model parameters were determined with 3530 experimental data (869 – 1641 kg/m³) of 76 ILs at 258 – 393 K and 0.09 – 207 MPa. It showed $AARD = 0.73$ % with $ARD_i < 8$ % for the estimation of density of tested ILs.

Qiao et al. (2010) proposed another GCM to predict ρ (g/cm³) of ILs using T and P as follows

$$\rho = H_1 + H_2P + H_3T \quad (2.25)$$

$$H_1 = \sum_{i=1}^{51} n_i h_{1i} + 0.781297 \quad (2.26)$$

$$H_2 = \sum_{i=1}^{51} n_i h_{2i} + 0.0000976606 \quad (2.27)$$

$$H_3 = \sum_{i=1}^{51} n_i h_{3i} - 0.000742764 \quad (2.28)$$

where h_{1i} , h_{2i} and h_{3i} are the contributions of the structural element i for H_1 , H_2 and H_3 , respectively. 6750 experimental density data from 123 pure ILs including quaternary phosphonium, imidazolium, pyridinium, pyrrolidinium and piperidinium besides compounds of MCl_x family were utilized to develop this model where 51 functional groups were defined. It showed $AARD$ values of 0.88 % and 0.27 % in the data correlation and prediction tests.

Paduszyński and Domańska (2012) reported a GCM based on an extensive experimental density database with 6830 values at 253 – 473 K and 0.1 – 300 MPa from 1028 ILs including amino acid and dicationic and tricationic compounds. Therefore, 177 groups were defined for this GCM. These authors estimated V_0 (m³/mol) at T_0 and P_0 using the next expression

$$V_0 = \sum_i n_i v_i^0 \quad (2.29)$$

where v_i^0 is the contribution of the structural element i . ρ_0 of ILs at T_0 and P_0 was then calculated using Equation (2.23). Then, the density $\rho(T, P_0)$ in kg/m^3 at P_0 can be calculated via the next equation

$$\rho(T, P_0) = \frac{W}{V_0[1+6.439 \cdot 10^{-4} K^{-1}(T-T_0)]} = \frac{\rho_0}{1+6.439 \cdot 10^{-4} K^{-1}(T-T_0)} \quad (2.30)$$

Tait-type equations were proposed for the calculation of ρ (kg/m^3) of compressed ILs

$$\rho = \frac{\rho(T, P_0)}{1-0.081 \ln[1+(P-P_0)J(T)]} \quad (2.31)$$

$$J(T) = \frac{1}{195 \text{ MPa}} [1 + 4.97 \cdot 10^{-3} K^{-1}(T - T_0)] \quad (2.32)$$

This model accurately predicted the densities of a wide diversity of ILs where $AARD = 0.45\%$ for 3695 experimental data from 200 ILs, see Table 2.2.

Evangelista et al. (2014) reported an extension of GCM proposed by Elbro et al. (1991) for the calculation of saturated liquid densities of ILs. This modified model was developed using a database with 21 845 experimental density values from 864 ILs including dicationic and tricationic species at 251.62–473.15 K and 0.1–300.0 MPa. These authors incorporated 102 new IL groups to those reported by Ihmels and Gmehling (2003). The cation groups were classified into “first level cation groups” (i.e., groups directly attached to atoms of the rings) and “second level groups” (i.e., groups directly attached to the “first level cation groups”), while the anion groups were obtained with the same methodology but the largest anions were divided into smaller groups considering the “first level anion groups”. This approach allowed to reduce the number of groups to obtain a wide variety of anions and cations involved in the database. Then, the saturated liquid densities (ρ_{sat} , g/cm^3) were estimated using the next equation

$$\rho_{sat} = \frac{W}{V} = \frac{W}{\sum n_i \Delta v_i} \quad (2.33)$$

$$\Delta v_i = L_{1i} + L_{2i}T + L_{3i}T^2 \quad (2.34)$$

where L_{1i} (cm^3/mol), L_{2i} ($\text{cm}^3/\text{mol}\cdot\text{K}$) and L_{3i} ($\text{cm}^3/\text{mol}\cdot\text{K}^2$) are the model parameters. The next Tait-type equation was employed to predict ρ (g/cm^3)

$$\rho = \frac{\rho(T, P_0)}{1-0.0711 \ln \left[\frac{(M_2+P)}{M_2+P_0} \right]} \quad (2.35)$$

$$M_2 = 287.624 \text{ MPa} - 0.444 \frac{\text{MPa}}{\text{K}} T \quad (2.36)$$

where $\rho(T, P_0)$ is also given in g/cm^3 . *AARD* values of this model were 0.676 % for imidazolium-, 1.21 % for ammonium-, 1.28 % for pyridinium-, 0.81 % for pyrrolidinium-, 0.80 % for phosphonium-, 2.15 % for piperidinium-, 1.51 % for morpholinium-, 1.93 % for azepanium-, 0.45 % for pyrazolium-, 0.42 % for isoquinolinium- and 0.24 % for quinolinium-based ILs.

Taherifard and Raeissi (2016) pointed out that several GCMs to predict ILs densities are based in large functional groups (i.e., cations, anions), which could improve the model accuracy but limits its applicability for different ILs. These authors proposed a simple GCM using only atoms and small functional groups to cover different ILs using the next equations

$$\rho = \frac{W}{V} \quad (2.37)$$

$$V = d_m + \frac{N_{at}}{100} (-20.8279 + 0.3749T - 0.2489P) \quad (2.38)$$

$$d_m = -4.0059 - 17.3357n_r + \sum n_i V_{c_{1i}} + \sum n_j V_{a_{1j}} \quad (2.39)$$

where n_r is the total number of rings existing in the IL, V_{c_1} and V_{a_1} are the contributions of the functional groups for the cation and anion, respectively and N_{at} is the number of atoms in the IL structure. These authors applied an experimental database containing 25 850 density values at 217.58–473.15 K and 0.1–300 MPa. Results showed that *AARD* of this approach was similar to that obtained with other GCMs based on large functional groups for cations and anions. This model was also compared with the models reported by Gardas and Coutinho (2008c) and Paduszyński and Domańska (2012). The performance of the model of Paduszyński and Domańska (2012) (i.e., *AARD* = 0.57 %) was better than that obtained with this new approach (i.e., *AARD* = 0.76 %). However, the model proposed by Gardas and Coutinho (2008c) (i.e., *AARD* = 1.3 %) was outperformed by the model proposed by Taherifard and Raeissi (2016) (i.e., *AARD* = 0.58 %).

Chen et al. (2019) developed a GCM using a database with 7360 experimental values of 143 ILs including 18 anions, 6 cations and 3 side groups thus covering temperatures of 273.15–473.15 K and pressures of 0.1–250.7 MPa. This model is given by

$$\rho = Q_1 + Q_2 T + Q_3 P \quad (2.40)$$

$$Q_1 = \sum_{i=1}^k n_i q_{1i} \quad (2.41)$$

$$Q_2 = \sum_{i=1}^k n_i q_{2i} \quad (2.42)$$

$$Q_3 = \sum_{i=1}^k n_i q_{3i} \quad (2.43)$$

where ρ is reported in kg/m^3 , q_{1i} , q_{2i} and q_{3i} are the structural element contributions. *AARD* values of this model were 0.43 % and 0.67 % for the correlation and prediction tests, respectively. In order to improve the reliability of the model parameters, these authors tried another correlation test using all ILs experimental data for the correlation set with the aim of enhancing the GCM performance. *AARD* = 0.49 % for this improved model with a maximum *ARD_i* value of 4.60 %.

2.2.4 Viscosity

Viscosity is the thermodynamic property utilized to measure the internal resistance of a fluid to flow and it is related to the internal friction within the fluid caused by the intermolecular interactions (Gardas and Coutinho, 2008b; Rooney et al., 2010). The design, simulation and optimization of reactors, crystallization equipment, distillation columns, liquid-liquid extractors, heat transfer equipment, stripping columns and process piping rely on this property for both pure compounds and mixtures (Lazzús and Pulgar-Villarroel, 2015; Chen et al., 2019; Hosseini et al., 2019). ILs are characterized by a high viscosity that could limit their application in industrial processes involving operations like pumping, mixing, heat and mass transfer (Aparicio et al., 2010; Chen et al., 2019; Sattari et al., 2016a). For comparison, the room temperature viscosities of ILs can range from 10 up to more than 105 cP, while the organic solvents have viscosities from 0.2 to 10 cP (Gardas and Coutinho, 2008b). The application of ILs as solvents requires a low viscosity to minimize the pumping costs and to increase the mass transfer rates (Gardas and Coutinho, 2008b). On the other hand, the high viscosities of ILs favor their utilization as lubricants (Lazzús and Pulgar-Villarroel, 2015). This property is highly dependent on the type of anion and cation that conforms the ILs molecule (Rooney et al., 2010). This context opens the possibility of preparing ILs with a tailored viscosity for the desired application (Aparicio et al., 2010). However, different authors (e.g., Aparicio et al., 2010; Chen et al., 2019) have highlighted that the experimental viscosity data of ILs cover a limited number of compounds (mainly well-known ILs) at atmospheric *P* and a limited *T* range. By this reason, a number of GCMs

has been reported and tested to predict IL viscosities using the available data and they are reported in Table 2.3.

Table 2.3 GCMs performance for the viscosity prediction of ILs.

	Modeling performance ¹				
Model	N_{IL}	N_{dat}	AARD, %	R^2	Remarks
GCM					
Gardas and Coutinho (2008b)	29	498	7.78	0.99	Global set
Gardas and Coutinho (2009)	25	482	7.50	0.99	Global set
Gharagheizi et al. (2012a)	443	1672	6.32	0.87	Global set
Daniel et al. (2013)	4	20	7.64	0.98	Global set
Lazzús and Pulgar-Villarroel (2015)	326	1445	4.55	0.94	Global set
	183	1110	4.46	N/A	Correlation set
	143	335	4.84	N/A	Prediction set
Sattari et al. (2016a)	85	863	3.23	0.98	Global set
	N/A	667	2.91	0.98	Correlation set
	N/A	196	4.31	0.97	Prediction set
Chen et al. (2019)	76	1090	3.58	N/A	Global set
	56	778	3.36	N/A	Correlation set
	20	312	5.63	N/A	Prediction set

¹ Variable description:

N_{IL} : Number of ILs used in the study.

R^2 : Coefficient of determination.

N_{dat} : Number of experimental data analyzed in the study.

N/A: Data not available.

For instance, Gardas and Coutinho (2008b) used an Orrick–Erbar-type equation (Reid et al., 1987) to predict the IL viscosity (η , cP) as follows

$$\ln \frac{\eta}{\rho \cdot W} = R_1 + \frac{R_2}{T} \quad (2.44)$$

where ρ (g/cm³) is calculated from Equation (2.21) and the parameters R_1 and R_2 were estimated using the structural element contributions. These authors employed an experimental database with 498 viscosity values (4–21 000 cP) at 293.15–393.15 K from 29

ILs based on pyrrolidinium, pyridinium and imidazolium cations, and trifluoromethanesulfonate, ethyl sulfate, methyl sulfate, acetate, chloride, bis(trifluoromethanesulfonyl) amide, tetrafluoroborate and hexafluorophosphate anions. *AARD* of this model was 7.78 % with a maximum ARD_i value < 28 %. One drawback of this model is that requires ρ data to estimate η of ILs. This drawback was overcome by Gardas and Coutinho (2009) using a new expression based on the Vogel-Tammann-Fulcher (VTF) equation to estimate η (Pa·s).

$$\ln \eta = S_1 + \frac{S_2}{(T - 165.06 \text{ K})} \quad (2.45)$$

where S_1 and S_2 were calculated by the structural element contributions s_{1i} and s_{2i} using

$$S_1 = \sum_{i=1}^k n_i s_{1i} \quad (2.46)$$

$$S_2 = \sum_{i=1}^k n_i s_{2i} \quad (2.47)$$

482 experimental data (0.004–1.065 Pa·s) at 293.15–393.15 K from 25 ILs containing the next ions: trifluoromethanesulfonate, ethyl sulfate, methyl sulfate, acetate, bis[(trifluoromethyl) sulfonyl]imide, tetrafluoroborate, hexafluorophosphate, pyrrolidinium, pyridinium and imidazolium. *AARD* = 7.50 % and a maximum ARD_i value of 23 % were reported for this GCM.

The GCM proposed by Gharagheizi et al. (2012a) considers the independent anion and cation contributions to differentiate their effects on η of ILs. This GCM was formulated using an extensive database of 1672 experimental viscosity values at 250–443 K and 1 atm from 443 ILs where 46 chemical sub-structures were defined (24 for anions and 22 for cations). This database included tri-alkyl imidazolium-, tetra-alkyl imidazolium-, sulfonium-, pyrroline-, pyrrolidinium-, pyridinium-, piperidinium-, phosphonium-, oxazolidinium-, morpholinium-, isoquinolinium-, guanidinium-, double imidazolium-, ammonium-, amino acids- and imidazolium-based ILs. The next expression was used to estimate η (given in cP) of ILs

$$\log(\eta) = \sum_{i=1}^{N_a} N_{ai} U_{ai} + \sum_{i=1}^{N_c} N_{ci} U_{ci} - 0.11968T - \frac{2665.71676}{T} + 0.00013T^2 + \frac{64753.03094}{T^2} + 34.2503 \quad (2.48)$$

where N_{ai} and N_{ci} are the number of repetitions of the i th structural element of IL anions and cations, respectively, while U_{ai} and U_{ci} are the corresponding contributions of these elements. $AARD = 6.32\%$ was obtained for the η estimation.

Daniel et al. (2013) proposed a GCM based on the Orrick-Erbar equation (Reid et al., 1987) to estimate η of magnetic ILs. The database included 20 experimental viscosity values (44–123 500 cP) at 293.15 – 373.15 K where this model showed $AARD = 7.64\%$.

Other GCM based on cation and anion contributions to estimate η of ILs was reported by Lazzús and Pulgar-Villarroel (2015). This model is given by

$$\ln \eta = 6.982 + \left(\sum_{i=1}^{N_c} N_{ci} V_{1ci} + \frac{\sum_{i=1}^{N_c} N_{ci} V_{2ci}}{T} \right) + \left(\sum_{i=1}^{N_a} N_{ai} V_{1ai} + \frac{\sum_{i=1}^{N_a} N_{ai} V_{2ai}}{T} \right) \quad (2.49)$$

where V_{1ci} , V_{2ci} , V_{1ai} and V_{2ai} are the contributions of the ion (cation or anion) i th. 1445 viscosity data from 6 to 20 883 cP at 253 – 395 K corresponding to 326 ILs were employed to define the model contributions. Different ILs based on isoquinolinium, oxazolidinium, morpholinium, ammonium, phosphonium, piperidinium, pyrrolidinium, pyridinium and imidazolium cations, and metal complexes, amino acid, trifluoroacetates, carboxylates, phosphates, borates, amides, imides, tosylates, sulfonates, sulfates, pseudohalides and halides anions were analyzed. This model showed $AARD = 4.55\%$.

The prediction of η of fluorine-containing ILs can be performed with the model of Sattari et al. (2016a) using the next expressions

$$\ln(\eta) = W_1 + W_2 \left(\frac{1000}{T} \right) + W_3 \left(\frac{1000}{T} \right)^2 \quad (2.50)$$

$$W_1 = \sum_{i=1}^k n_i w_{1i} \quad (2.51)$$

$$W_2 = \sum_{i=1}^k n_i w_{2i} \quad (2.52)$$

$$W_3 = \sum_{i=1}^k n_i w_{3i} \quad (2.53)$$

where w_{1i} , w_{2i} and w_{3i} are the structural element parameters. These parameters were obtained from the data processing of 863 viscosity values (4.1–3067.3 cP) at 258.15–388.51 K of 85 fluorine-containing ILs constituted by 58 cations and 9 anions. $AARD$ value of 3.23 % was obtained for the η calculation with this model.

Chen et al. (2019) predicted η (Pa·s) of ILs using the next equations

$$\ln \frac{\eta}{X_0} = X_1 + X_2 \frac{100}{T} + X_3 \left(\frac{100}{T} \right)^2 \quad (2.54)$$

$$X_1 = \sum_{i=1}^k n_i x_{1i} \quad (2.55)$$

$$X_2 = \sum_{i=1}^k n_i x_{2i} \quad (2.56)$$

$$X_3 = \sum_{i=1}^k n_i x_{3i} \quad (2.57)$$

where x_{1i} , x_{2i} and x_{3i} are the structural element contributions and X_0 is a model parameter. 1090 experimental data at 278–408.15 K from 76 ILs involving 8 cations, 16 anions and 3 side groups were used to define these model parameters. The performance of this model is reported in Table 2.3 where $AARD = 3.58 \%$ for the correlation test using all ILs with a maximum ARD_i value of 27.7 %.

Herein, it is convenient to highlight that the available experimental viscosity database of ILs may contain some uncertainties related to the presence of impurities in the samples employed in the quantifications and the accuracy of quantification methodologies to measure this thermodynamic property (Alcalde et al., 2015). Note that the ILs viscosity is larger than other compounds used in the industry. Therefore, these uncertainties can impact the determination of group contribution values and the corresponding model performance.

2.2.5 Thermal conductivity

Thermal conductivity is a transport property that measures the capability of a given material to conduct heat (Gardas and Coutinho, 2009). This property depends on the composition and chemical nature of tested compound besides the corresponding state variables like T (Lazzús, 2015). The design of thermal processes and heat transfer equipment requires the calculation of this property (Lazzús, 2015; Hosseini et al., 2019). For example, the knowledge of thermal conductivity of ILs is essential to analyze the process energy requirements including the potential phase change of ILs due to process operating conditions (Coutinho et al., 2012; Soares et al., 2019). Despite of the relevance of this thermodynamic property for PSE, the available experimental information of ILs thermal conductivities is relatively limited considering the vast number of ILs that have been synthesized (Coutinho et al., 2012; Atashrouz et al., 2016; Hosseini et al., 2019).

In particular, Gardas and Coutinho (2009) developed a GCM to estimate the thermal conductivity of ILs containing phosphonium, pyrrolidinium and imidazolium cations with chloride, ethyl sulfate, trifluoromethanesulfonate, bis[(trifluoromethyl)sulfonyl]imide,

tetrafluoroborate and hexafluorophosphate anions. They used an experimental database with 107 values (0.124–0.199 W/m·K) from 16 ILs at 293–390 K. These authors proposed a linear correlation to estimate the thermal conductivity λ (W/m·K)

$$\lambda = Y_1 - Y_2 T \quad (2.58)$$

$$Y_1 = \sum_{i=1}^k n_i y_{1i} \quad (2.59)$$

$$Y_2 = \sum_{i=1}^k n_i y_{2i} \quad (2.60)$$

where Y_1 and Y_2 were calculated using the contributions (y_{1i} and y_{2i}) of the structural elements conforming the tested IL. This GCM showed $AARD = 1.06$ % with a maximum ARD_i value of 3.5 %, see Table 2.4. It was extended by Gardas et al. (2010) to handle amino acid-based ILs. They used 75 experimental data ($\lambda = 0.146$ – 0.171 , W/m·K) at 298–353 K from 11 amino acid-based ILs. $AARD$ of this model was 0.67 % to predict λ for tested ILs.

Table 2.4 GCMs performance for the thermal conductivity prediction of ILs.

Modeling performance ¹					
Model	N_{IL}	N_{dat}	$AARD$, %	R^2	Remarks
GCM					
Gardas and Coutinho (2009)	16	107	1.06	0.99	Global set
Gardas et al. (2010)	11	75	0.67	0.94	Global set
Wu et al. (2013)	36	286	1.66	N/A	Global set
Albert and Müller (2014)	32	316	2.49	0.95	Correlation set
	7	59	8.09	0.74	Prediction set
Lazzús (2015)	41	400	2.12	0.98	Global set
	30	300	1.90	0.99	Correlation set
	11	100	2.33	0.97	Prediction set
Soares et al. (2019)	33	237	1.16	N/A	Global set

¹ Variable description:

N_{IL} : Number of ILs used in the study.

R^2 : Coefficient of determination.

N_{dat} : Number of experimental data analyzed in the study. N/A: Data not available.

Soares et al. (2019) also utilized the GCM of Gardas and Coutinho (2009) to model ILs containing 13 anion groups (e.g., FAP⁻, VAL⁻, THR⁻, TAU⁻, PRO⁻, CYS⁻, LIS⁻, SER⁻, HOPO₂⁻, CH₃COO⁻, DCA⁻, B(CN)₄⁻, C(CN)₃⁻) and recalculated the contributions for the others groups already reported. They used a database consisting of 237 experimental thermal conductivities from 0.105 to 0.214 W/m·K at atmospheric *P* and 273–390 K from 33 ILs. This model showed *AARD* = 1.16 % with a maximum *ARD_i* of 4.37 %.

Wu et al. (2013) analyzed the thermal conductivity of 36 ILs composed of imidazolium, ammonium and phosphonium cations, and octylsulfate, ethylsulfate, bis(trifluoromethanesulfonyl) amide, hexafluorophosphate, tetrafluoroborate, chloride, trifluoromethanesulfonate, dicyanamide, tris(pentafluoroethyl)trifluorophosphate, acetate, tricyanomethanide, serinate, taurinate, lysinate, threonate, proline and valinate anions. 25 molecular groups were used to develop the GCM via the data processing of 286 experimental values of λ (0.1050–0.2137 W/m·K) at 273.15–353.15 K. This property was calculated with the next expression

$$\lambda = \sum_{j=0}^2 Z_j \left(\sum_{i=1}^k n_i \Delta z_{0i} \right)^j \left[1 + Z_3 (1 - T_r)^{2/3} \right] \quad (2.61)$$

where Z_j is a fixed model parameter, Δz_{0i} corresponds to the structural element contribution, T_r is obtained by Equation (2.9) where T_c is calculated by the GCM proposed by Valderrama et al. (2012) that is described in Section 2.2.12 of this chapter. Overall, this GCM showed low modeling errors (i.e., *AARD* = 1.66 %).

Albert and Müller (2014) applied a database with 372 experimental values from 39 ILs at 273.15–390 K to develop a GCM, which is given by

$$\lambda = (AA_{1cation} + AA_{1anion}) - (AA_{2cation} + AA_{2anion}) \cdot T \quad (2.62)$$

$$AA_{1ion} = \sum_i (v_i \cdot AA_{1i}) + AA_{1correction} \quad (2.63)$$

$$AA_{2ion} = \sum_i (v_i \cdot AA_{2i}) + AA_{2correction} \quad (2.64)$$

where AA_{1i} and AA_{2i} are the structural element contributions, $AA_{1correction}$ and $AA_{2correction}$ are additional contributions related to the type of ion or ring structure. It was reported *AARD* values of 2.49 % and 8.09 % for the correlation and prediction tests, see Table 2.4.

The GCM proposed by Lazzús (2015) estimates λ of ILs as a function of T and P (kPa) as follows

$$\lambda = \left(\sum_{i=1}^{N_c} N_{ci} AB_{1ci} + \sum_{i=1}^{N_a} N_{ai} AB_{1ai} \right) - T \left| \sum_{i=1}^{N_c} N_{ci} AB_{2ci} + \sum_{i=1}^{N_a} N_{ai} AB_{2ai} \right| + P \left| \sum_{i=1}^{N_c} N_{ci} AB_{3ci} + \sum_{i=1}^{N_a} N_{ai} AB_{3ai} \right| \quad (2.65)$$

where AB_{1ci} , AB_{2ci} , AB_{3ci} and AB_{1ai} , AB_{2ai} , AB_{3ai} are the contributions of i th structural element of the cations and anions, respectively. An experimental database of 400 thermal conductivities ranging from 0.1 to 0.22 W/m·K at 100–20 000 kPa and 273–390 K was used. This database included thermodynamic data from 41 ILs containing ammonium, phosphonium, pyrrolidinium, pyridinium and imidazolium cations, and amino acids, acetates, phosphates, borates, imides, tosylates, sulfonates and halides as anions. ARD_i values of this model were $< 10\%$ for all tested ILs. The model performance for correlation and prediction tests is reported in Table 2.4. Results also indicated that this GCM provided better predictions than those obtained with the models of Gardas and Coutinho (2009), Wu et al. (2013) and Albert and Müller (2014).

Finally, Oster et al. (2018) proposed alternative strategies to improve the performance of GCM of Wu et al. (2013). These strategies were based on a more accurate molecular structure characterization of ILs including the impact of cation core atom on the structure and physical properties of these compounds.

2.2.6 Melting point

Melting point (T_m , K) is defined as the temperature to change the physical state of a given compound from solid to liquid at atmospheric pressure (Chen et al., 2019). This thermodynamic property is essential for the calculation of vapor pressure and aqueous solubility (Katritzky et al., 2001) and it can be also employed for the chemical identification and purification (Hosseini et al., 2019). The melting points of ILs are required to characterize their phase behavior in different applications as solvents or heat transfer fluids (Aguirre et al., 2012; Mital et al., 2021). Some authors have indicated that the prediction of T_m for ILs is challenging since it depends on the arrangement of ions inside the IL crystal and their corresponding interactions (Aguirre et al., 2012; Keshavarz et al., 2018). As expected, the

experimental information of ILs melting points is limited and some GCMs have been proposed for their predictions as shown in Table 2.5.

Table 2.5 GCMs performance for the melting point prediction of ILs.

	Modeling performance ¹			
<i>Model</i>	<i>N_{IL}</i>	<i>AARD, %</i>	<i>R²</i>	<i>Remarks</i>
GCM				
Huo et al. (2009)	190	5.86	0.90	Global set
Lazzús (2012b)	400	7.07	0.88	Global set
	200	7.97	N/A	Correlation set
	200	6.16	N/A	Prediction set
Aguirre et al. (2012)	136	7.80	N/A	Global
Gharagheizi et al. (2012c)	799	5.82	0.81	Global set
	598	5.81	0.81	Correlation set
	201	5.85	0.82	Prediction set
Chen et al. (2019)	111	4.70	N/A	Global set
	75	4.70	N/A	Correlation set
	26	7.09	N/A	Prediction set
Valderrama and Cardona (2021)	124	6.47	N/A	Correlation set
	13	6.93	N/A	Prediction set

¹ Variable description:

N_{IL} : Number of ILs used in the study.

R^2 : Coefficient of determination.

N/A: Data not available.

Huo et al. (2009) proposed a GCM to calculate T_m using 190 ILs composed of imidazolium and benzimidazolium. This model was defined as

$$T_m = \frac{\sum_k N_{F_k} AC_{1,k} + 2.355 \sum_j N_{I_j} AC_{2,j}}{-0.4177 + 0.004665\alpha - 0.01859\tau + 0.005428\delta} \quad (2.66)$$

where N_{F_k} and N_{I_j} are the frequencies of group k and ionic group j contained in the IL molecule and $AC_{1,k}$ and $AC_{2,j}$ are the melting point contributions of group k and ionic group

j. Note that τ represents the number of ring groups in the molecule, while the value of α corresponds to different scenarios: $\alpha = 1$ if the groups connected to $> \text{NH}$ and $=\text{N}-$ (ring of imidazolium) are the same, $\alpha = 0$ if different groups are connected to these structural elements. In a similar way, $\delta = 1$ if there is a group connected to the C position in the imidazolium cation and $\delta = 0$ otherwise. This model showed $AARD = 5.86\%$ with a maximum ARD_i value of 32.75% .

Lazzús (2012b) proposed a GCM to calculate this thermodynamic property for pyrazolium-, thiazolium-, morpholinium-, sulfonium-, ammonium-, phosphonium-, piperidinium-, pyrrolidinonium-, triazolium-, pyrrolidinium-, pyridinium- and imidazolium-based ILs with the following anions: metal complexes, trifluoroacetates, triflates, carboxylates, phosphates, borates, amides, imides, tosylates, sulfonates, sulfates, pseudohalides and halides. This model is defined by

$$T_m = 288.7 + \sum_{i=1}^{N_c} N_{ci} AD_{ci} + \sum_{i=1}^{N_a} N_{ai} AD_{ai} \quad (2.67)$$

where AD_{ci} and AD_{ai} are the contributions from the i th structural element of cations and anions, respectively. 400 ILs with T_m from 194 to 567 K were used to define the structural element contributions.

T_m of piperidinium-, phosphonium-, ammonium-, pyrrolidinium-, pyridinium- and imidazolium-based ILs can be calculated using the GCM of Aguirre et al. (2012), which is given by

$$T_m = \frac{\sum_i n_i AE_{1,i}}{0.1 + \alpha_c + 0.012\tau_c} \quad (2.68)$$

$$\tau_c = \sum_k (n(CH_2)_k - 1) \quad (2.69)$$

where $AE_{1,i}$ is the contribution of structural element i , α_c is a symmetry parameter defined for the cation and τ_c is a parameter related to the cation flexibility that is expressed in terms of the number of methylene groups (CH_2) present in the alkyl chain of IL cation minus 1. 136 ILs were analyzed with this model thus obtaining $AARD = 7.8\%$, see Table 2.5.

The following equation was employed by Gharagheizi et al. (2012c) to calculate T_m of ILs.

$$T_m = \sum_{i=1}^{N_a} N_{ai} AD_{ai} + \sum_{i=1}^{N_c} N_{ci} AD_{ci} + 264.2886 \quad (2.70)$$

They used an experimental database with 799 ILs containing tetra-alkyl imidazolium, tetrazolium, thiazolium, trialkyl imidazolium, triazolium, uranium, piperazinium,

piperidinium, pyridazinium, pyridinium, pyrrolidinium, pyrroline, quinary alkyl imidazolium, morpholinium, oxazolidinium, phosphonium, sulfonium, double imidazolium, guanidinium, isoquinolinium, 1,3-dialkyl imidazolium, 1-alkyl imidazolium, amino acids and ammonium.

On the other hand, Chen et al. (2019) proposed a GCM that was formulated in terms of the number of cations (n_c), anions (n_a) and side groups (n_g) in the IL molecule and the corresponding contributions of these cations (t_c), anions (t_a) and substituents (t_g). This GCM was defined as

$$T_m = \sum_{c=1}^{k_c} n_c t_c + \sum_{a=1}^{k_a} n_a t_a + \sum_{g=1}^{k_g} n_g t_g \quad (2.71)$$

111 ILs were analyzed covering 3 side groups, 13 anions and 6 cation cores thus obtaining $AARD = 4.7\%$ where the maximum modeling errors were for the calculation of tetrapropylammonium bis(trifluoromethylsulfonyl)imide IL.

Valderrama and Cardona (2021) proposed a GCM with a theoretical relationship between T_m , the melting heat (ΔH_m , kJ/mol) and the melting entropy (ΔS_m , kJ/mol·K), which was defined as

$$T_m = \frac{\Delta H_m}{\Delta S_m} \quad (2.72)$$

$$\Delta H_m = 3.8315 + \sum [N_{ci} \Delta H_{m,ic}] + \sum [N_{ai} \Delta H_{m,ia}] \quad (2.73)$$

$$\Delta S_m = 9.7736 \cdot 10^{-3} + 2.4599 \cdot 10^{-4} M_c + 1.4582 \cdot 10^{-4} M_a \quad (2.74)$$

$$T_m = \frac{3.8315 + \sum [N_{ci} \Delta H_{m,ic}] + \sum [N_{ai} \Delta H_{m,ia}]}{9.7736 \cdot 10^{-3} + 2.4599 \cdot 10^{-4} M_c + 1.4582 \cdot 10^{-4} M_a} \quad (2.75)$$

where $\Delta H_{m,ic}$ and $\Delta H_{m,ia}$ are the contributions for group i in the cation and anion, respectively, and M_c and M_a are the mass of cation and anion. $AARD$ values of 6.47 % and 6.93 % were obtained for this model in the correlation and prediction tests.

2.2.7 Surface tension

Surface tension is defined as a force acting at the interface of fluid phases and it indicates the cohesive force of liquid molecules in the surface (Gharagheizi et al., 2012b; Lazzús et al., 2017b). This property affects the transfer rates of vapor absorption at the vapor-liquid interface. Therefore, it is important for the study of chemistry and physics at free surfaces and for the process design in flow and mass transfer in porous media, materials science,

reactor engineering, biomedical and biochemical engineering, electronics and electrical applications (Rooney et al., 2010; Chen et al., 2019; Hosseini et al., 2019). ILs can be also used in extraction and multiphase catalytic reactions that take place at the interphase between the IL and an organic or aqueous phase and, consequently, the surface properties of ILs are required in the process analysis (Hosseini et al., 2019). Table 2.6 summarizes GCMs for the calculation of this property.

Table 2.6 GCMs performance for the surface tension prediction of ILs.

Modeling performance ¹					
<i>Model</i>	<i>N_{IL}</i>	<i>N_{dat}</i>	<i>AARD, %</i>	<i>R²</i>	<i>Remarks</i>
GCM					
Gharagheizi et al. (2012b)	51	920	3.61	0.92	Global set
	50	735	3.71	0.92	Correlation set
	45	185	3.20	0.93	Prediction set
Lazzús et al. (2017b)	154	2286	2.81	0.98	Global set
	118	1804	2.69	0.99	Correlation set
	36	482	3.36	0.98	Prediction set
Shahsavari et al. (2018)	61	1246	4.16	0.92	Global set
	N/A	872	4.19	0.92	Correlation set
	N/A	374	4.11	0.91	Prediction set
Chen et al. (2019)	102	1365	3.10	N/A	Global set
	72	952	2.87	N/A	Correlation set
	30	413	4.71	N/A	Prediction set
GCM-CSP					
Wu et al. (2012)	105	1224	4.95	N/A	Global set

¹ Variable description:

N_{IL}: Number of ILs used in the study.

R²: Coefficient of determination.

N_{dat}: Number of experimental data analyzed in the study.

N/A: Data not available.

Gharagheizi et al. (2012b) proposed a GCM to estimate the surface tension (σ , N/m) of ILs considering the contributions of the cation (7 structural elements) and anion (12 structural elements). This model was defined as

$$\sigma = \sum_{i=1}^{N_a} N_{ai} AG_{ai} + \sum_{i=1}^{N_c} N_{ci} AG_{ci} + 0.1362154 \frac{N}{m} - 6.17 \cdot 10^{-5} \frac{N}{m \cdot T} T \quad (2.76)$$

where AG_{ai} and AG_{ci} are the contributions of the i th structural element of anions and cations, respectively. The structural element contributions were obtained from the processing of a database containing 920 surface tension values (0.02499–0.0647 N/m) at 268.3–743.7 K from 51 ILs. Authors concluded that the accuracy of this model was acceptable with $AARD = 3.61$ %.

An atomic-based GCM was proposed by Shahsavari et al. (2018) to estimate σ (mN/m) of ILs at atmospheric P. In this model, the atoms usually found in ILs (e.g., iodine, boron, chlorine, bromine, fluorine, oxygen, phosphorous, sulfur, nitrogen and carbon) were selected as input parameters for the prediction of this thermodynamic property

$$\begin{aligned} \frac{\sigma}{1000} = & 1.138312T^{-0.614173} + 0.007909Nu_{Ca}^{0.150383} + 0.250221Nu_{Cc}^{-0.049454} - \\ & 0.013238Nu_N^{0.501161} - 0.003356(Nu_F + Nu_{Cl} + Nu_{Br} + Nu_I)^{0.092801} - \\ & 0.001367Nu_O^{0.712848} - 0.009602Nu_S^{0.296456} - 0.005233Nu_P^{1.068123} - \\ & 0.002497Nu_B^{0.501312} - 0.192602 \end{aligned} \quad (2.77)$$

where Nu_I and Nu_P are the number of atoms of iodine and phosphorus in the IL, respectively. This GCM was developed using 1246 surface tension data from 0.0155 to 0.0564 N/m at the melting points of 262.89–533.2 K from 61 ILs. It showed $AARD$ values of ~ 4.1 % for the correlation and prediction tests, see Table 2.6.

Lazzús et al. (2017b) reported a GCM to calculate σ of piperidinium-, pyrrolidinium-, pyridinium-, phosphonium-, ammonium- and imidazolium-based ILs. 2286 experimental data from 154 ILs covering a range of σ from 0.015 to 0.062 N/m at 263 – 533 K were used in this study. This GCM is defined by the next expression

$$\sigma = \left| \sum_{i=1}^{N_c} N_{ci} AI_{1ci} + \sum_{i=1}^{N_a} N_{ai} AI_{1ai} \right| - T \left| \sum_{i=1}^{N_c} N_{ci} AI_{2ci} + \sum_{i=1}^{N_a} N_{ai} AI_{2ai} \right| \quad (2.78)$$

where AI_{1ci} , AI_{2ci} and AI_{1ai} , AI_{2ai} are the contributions of the i th structural element of cations and anions, respectively. This model showed $AARD$ values of 0.07–4.03 % with a maximum ARD_i of 14.20 % and $R^2 = 0.91 - 0.99$ to estimate σ for these ILs. It generated lower modeling errors than those obtained with the model of Gharagheizi et al. (2012b) for

small- and large-size ILs, but the performances of both models were similar for medium-size ILs.

Chen et al. (2019) proposed other GCM to calculate this IL thermodynamic property, which was defined as

$$\ln \sigma = AJ_1 + AJ_2 \left(\frac{T}{100} \right) + AJ_3 \left(\frac{T}{100} \right)^2 \quad (2.79)$$

$$AJ_1 = \sum_{i=1}^k n_i aj_{1i} \quad (2.80)$$

$$AJ_2 = \sum_{i=1}^k n_i aj_{2i} \quad (2.81)$$

$$AJ_3 = \sum_{i=1}^k n_i aj_{3i} \quad (2.82)$$

where AJ_i is calculated via the structural element contributions (aj_{1i} , aj_{2i} and aj_{3i}). 1365 surface tension data at 262.89–393 K from 102 ILs involving 6 cations, 18 anions and 2 side groups were compiled and utilized for the model development. Table 2.6 shows the model performance for correlation and prediction tests.

Wu et al. (2012) used an approach that combined a GCM with the corresponding-states theory (GCM-CSP) for the σ prediction of ILs containing ammonium, phosphonium, pyridinium, imidazolium, octylsulfate, methylsulfate, ethylsulfate, propylsulfate, butylsulfate, bis(trifluoromethanesulfonyl) amide, hexafluorophosphate, tetrafluoroborate, chloride, bromide, iodide, triiodide, pentaiodide, heptaoidide, nonaiodide, trifluoromethanesulfonate, dicyanamide, glutamic acid, alanine, thiocyanate, tris(pentafluoroethyl)trifluorophosphate, tetracyanoborate, trifluoroacetate and alkyltrifluoroborate ions. These authors analyzed 1224 surface tension data ranging from 15.50 to 76.20 mN/m at 268.29–533.20 K of 105 ILs. This GCM-CSP utilizes the next expression to estimate σ (mN/m)

$$\sigma = \sum_{j=0}^3 AK_j \left(\sum_{i=1}^k n_i \Delta ak_{0i} \right)^j (1 - T_r)^{11/9} \quad (2.83)$$

where Δak_{0i} are the structural element contributions and AK_j are the model parameters. Note that the value of T_c for the calculation of T_r (see Equation (2.9)) can be obtained by applying the GCM proposed by Valderrama and Rojas (2009), which is described in Section 2.2.12 of this chapter. This model showed $AARD = 4.95$ % with a maximum ARD_i value of 29.92 % for $[C_{10}mim][BF_4]$ at 293.15 K. Another model based on the parachor parameter was reported by Součková et al. (2015) to calculate σ . This approach considers the slight

dependence of the parachor with respect to T and the number of carbon atoms in the cation alkyl chain.

2.2.8 Electrical conductivity

The capability of a given material to transport an electric charge is known as the electrical conductivity and it has been defined as the ratio of the current density to the electric field strength (Gardas and Coutinho, 2009; Chen et al., 2020). Overall, the electrical conductivity of ILs is low at room temperature but some authors have reported a substantial impact of T on this property for different ILs (Aparicio et al., 2010). The electrical conductivity of ILs is required in electrochemical applications like energy storage systems (e.g., supercapacitors, fuel cells and batteries) (Hosseini et al., 2019). Herein, it is convenient to indicate that few known ILs are suitable for electrochemical applications and, consequently, the available experimental data of this property are also limited (Chen et al., 2020). In particular, Gardas and Coutinho (2009) proposed a Vogel-Tammann-Fulcher (VTF) type equation to estimate the electrical conductivities ($\mathcal{E}$, S/m) of ILs using the next expressions

$$\ln \mathcal{E} = \ln AL_1 + \frac{AL_2}{T - 165.06 \text{ K}} \quad (2.84)$$

$$AL_1 = \sum_{i=1}^k n_i al_{1i} \quad (2.85)$$

$$AL_2 = \sum_{i=1}^k n_i al_{2i} \quad (2.86)$$

where al_{1i} and al_{2i} are the structural element contributions. The database of $\mathcal{E}$ used by these authors covered a range of 0.01–12.68 S/m at 258.1–433.15 K with 307 experimental values from 15 ILs containing different cations (e.g., ammonium, pyrrolidinium, pyridinium, imidazolium) and anions (e.g., PF_6^- , BF_4^- , Tf_2N^- , EtSO_4^- , CF_3SO_3^-). This model showed $AARD = 4.57\%$ for the estimation of $\mathcal{E}$ of tested ILs, see Table 2.7.

Wu et al. (2016) developed an alternative approach based also on VTF type equation but defining more complex group contribution equations using a second-order GCM. This model is defined as

$$\ln \mathcal{E} = AM_{1,j} + \frac{AM_{2,j}}{T - T_{0AM,j}} \quad (2.87)$$

$$AM_{1,j} = -2.573 + \sum_{m=1}^2 \sum_{i=1}^I n_i^m (WA_{m,i}) \quad (2.88)$$

$$T_{0AM,j} = 291.166 + \sum_{m=1}^2 \sum_{i=1}^I n_i^m (WT_{0m,i}) \quad (2.89)$$

$$AM_{2,j} = (-21.452 + \sum_{m=1}^2 \sum_{i=1}^I n_i^m (WB_{m,i}))(5.252 + \sum_{m=1}^2 \sum_{i=1}^I n_i^m (WB'_{m,i})) \quad (2.90)$$

where $WA_{m,i}$, $WT_{0m,i}$, $WB_{m,i}$ and $WB'_{m,i}$ are the contributions of structural element i . 2195 electrical conductivity data (0.0001524 – 19.3 S/m) at 238.15–484.1 K from 188 ILs (e.g., imidazolium, phosphonium, ammonium, sulfonium, pyridinium, pyrrolidinium and piperidinium) were used to define the parameters of this model. Table 2.7 indicates that this GCM showed $AARD = 6.02\%$ with a maximum ARD_i value of 39 % for tested ILs. Although the second-order GCMs allow to differentiate the thermodynamic properties of isomers, they require additional information about the IL molecular structures and are also more complex to be implemented due to the multi-order group subdivision.

Table 2.7 GCMs performance for the electrical conductivity prediction of ILs.

Modeling performance ¹					
Model	N_{IL}	N_{dat}	$AARD, \%$	R^2	Remarks
GCM					
Gardas and Coutinho (2009)	15	307	4.57	0.99	Global set
Wu et al. (2016)	188	2195	6.02	N/A	Global set
	177	1978	6.12	N/A	Correlation
	11	217	5.10	N/A	Prediction
Chen et al. (2020)	77	1578	3.62	N/A	Global set
	57	1121	3.30	N/A	Correlation set
	20	457	6.83	N/A	Prediction set
GCM-LSSVM					
Gharagheizi et al. (2014)	54	1077	3.30	0.99	Global set

¹ Variable description:

N_{IL} : Number of ILs used in the study.

R^2 : Coefficient of determination.

N_{dat} : Number of experimental data analyzed in the study.

N/A: Data not available.

It has been indicated that the main drawback of the model of Gardas and Coutinho (2009) relies on its limited number of contribution values that are available for the cations and anions of IL molecules. Chen et al. (2020) tried to overcome this limitation via the implementation

of a GCM that can be applied for a wide spectrum of ILs. This model was developed using 1578 electrical conductivity data (0.0017–9.167 S/m) at 248.05–468.15 K of 77 ILs with 8 cations, 34 anions and 4 side groups. It is given by

$$\ln \frac{\varepsilon}{AN_0} = AN_1 + AN_2 \frac{100}{T} + AN_3 \left(\frac{100}{T} \right)^2 \quad (2.91)$$

$$AN_1 = \sum_{i=1}^k n_i an_{1i} \quad (2.92)$$

$$AN_2 = \sum_{i=1}^k n_i an_{2i} \quad (2.93)$$

$$AN_3 = \sum_{i=1}^k n_i an_{3i} \quad (2.94)$$

where an_{1i} , an_{2i} and an_{3i} are the structural element contributions and AN_0 is a fixed parameter given in S/m. This model showed *AARD* values of 3.30 % and 6.83 % for the correlation and prediction tests, see Table 2.7. These authors indicated that this new model outperformed the calculations with the model of Gardas and Coutinho (2009).

Finally, GCM based on LSSVM was developed by Gharagheizi et al. (2014) to estimate ε of ILs where 1077 experimental data of 54 compounds were analyzed. The performance of this model is also summarized in Table 2.7.

2.2.9 Refractive index

The refractive index (n_D) is defined as the ratio of light speed in vacuum to that obtained for another given medium (Sattari et al., 2014b). This property can be used to test the material purity, to determine a mixture concentration for certain chemical compounds, and to identify or characterize the presence of a particular substance (Gardas and Coutinho, 2009; Sattari et al., 2014b; Wang et al., 2017). n_D is associated with the intermolecular forces and their behavior in solutions and is also related with other thermodynamic properties such as the dielectric constant, surface tension and density (Gardas and Coutinho, 2009). The experimental quantification of this property is simple but the available database covers a limited number of ILs (Sattari et al., 2014b).

To date, there are some GCMs to calculate this property for ILs (Gardas and Coutinho, 2009; Freire et al., 2011; Soriano et al., 2010b; Almeida et al., 2013; Sattari et al., 2014b; Wang et al., 2017), see Table 2.8. For instance, Gardas and Coutinho (2009) proposed the next GCM for the calculation of n_D of ILs composed of imidazolium with PF_6^- , BF_4^- , Tf_2N^- , MeSO_4^- , EtSO_4^- , CF_3SO_3^- and Cl^- anions at 283.15–363.15 K

$$n_D = AO_1 - AO_2T \quad (2.95)$$

$$AO_1 = \sum_{i=1}^k n_i ao_{1i} \quad (2.96)$$

$$AO_2 = \sum_{i=1}^k n_i ao_{2i} \quad (2.97)$$

where ao_{1i} and ao_{2i} are the structural element contributions. 245 experimental data from 24 ILs were used to develop this model, which showed $AARD = 0.18\%$ where a maximum ARD_i of 0.61% .

Table 2.8 GCMs performance for the refractive index prediction of ILs.

Modeling performance ¹					
Model	N_{IL}	N_{dat}	$AARD, \%$	R^2	Remarks
GCM					
Gardas and Coutinho (2009)	24	245	0.18	0.99	Global set
Sattari et al. (2014b)	97	931	0.34	0.96	Global set
	82	759	0.32	0.97	Correlation set
	15	172	0.45	0.92	Prediction set
Wang et al. (2017)	299	2138	0.63	0.89	Global set
	277	1711	0.63	0.88	Correlation set
	190	427	0.60	0.90	Prediction set

¹ Variable description:

N_{IL} : Number of ILs used in the study.

N_{dat} : Number of experimental data analyzed in the study.

R^2 : Coefficient of determination.

This model was also employed by Freire et al. (2011) to calculate n_D of ILs containing 1-ethyl-3-methylimidazolium cation and dimethylphosphate, tosylate, thiocyanate, dicyanamide, trifluoromethanesulfonate, methanesulfonate, methylphosphonate and acetate anions. These authors recalculated the structural element contributions and proposed new values for other elements not included in the original model. Other variants of this GCM were proposed by Soriano et al. (2010b) and Almeida et al. (2013). Sattari et al. (2014b) reported other alternative model to estimate this property, which is given by

$$n_D = AP_1 + AP_2T \quad (2.98)$$

$$AP_1 = 1.5082 + \sum_{i=1}^k n_i ap_{1_i} \quad (2.99)$$

$$AP_2 = -1.4207 \cdot 10^{-4} + \sum_{i=1}^k n_i ap_{2_i} \quad (2.100)$$

where ap_{1_i} and ap_{2_i} are the structural element contributions that were obtained from the processing of a database with 931 experimental values ranging from 1.3551 to 1.5778 at 283–363.15 K. This database included 97 ILs of different families (e.g., amino acid, ammonium, imidazolium, phosphonium, pyridinium, pyrrolidinium) constituted by 50 cations and 33 anions. The model performance is also described in Table 2.8.

Wang et al. (2017) employed 2138 experimental data at 283.15–368.15 K of 299 ILs to obtain the structural element contributions for the prediction of this property in ILs. These contributions covered pyrazolium-, piperidinium-, sulfonium-, guanidinium-, pyrrolidinium-, amino acid-, phosphonium-, ammonium-, pyridinium- and imidazolium-based ILs, which can be represented by 9 cations, 9 anions and 23 substituents attached to the cores. This GCM was established as

$$n_D = (1.4626 + \sum_{i=1}^k n_i aq_{1_i}) + 0.02372 \frac{T}{100} \quad (2.101)$$

where aq_{1_i} is the contribution of structural element i . The performance of this model was also acceptable for correlation and prediction tests, see Table 2.8, with $AARD = 0.628\%$ and $R^2 = 0.89$.

2.2.10 Thermal decomposition temperature

The thermal decomposition temperature is related to the thermodynamic stability of compounds (Keshavarz et al., 2016a). ILs are nonvolatile and, consequently, the thermal decomposition temperature establishes the upper limit of T for their existence as liquid and their corresponding application as solvents in chemical processes (Lazzús, 2012c; Gharagheizi et al., 2013; Mokadem et al., 2016). Therefore, this property and the melting point can be used to define the operation limits of ILs-based processes (Mokadem et al., 2016). Overall, ILs have high thermal stabilities and can operate over a wide T range because these compounds usually decompose at $> 300\text{ }^\circ\text{C}$ (Lazzús, 2012c). In terms of the modeling of this property, Lazzús (2012c) proposed a GCM to calculate the thermal decomposition temperature (T_d , K) of ILs containing thiazolium, pyrazolium, triazolium, oxazolidinium,

morpholinium, ammonium, phosphonium, piperidinium, pyrrolidinium, pyridinium and imidazolium cations, and metal complexes, amino acids, trifluoroacetates, carboxylates, phosphates, borates, amides, imides, tosylates, sulfonates, sulfates, pseudohalides and halides as anions. This model was defined by

$$T_d = 663.85 + \sum_{i=1}^{N_c} N_{ci} AR_{ci} + \sum_{i=1}^{N_a} N_{ai} AR_{ai} \quad (2.102)$$

where AR_{ci} and AR_{ai} are the contributions of the structural elements from cations and anions, respectively. This model was assessed using an experimental database with T_d (403 – 857 K) of 198 ILs. It showed $AARD = 4.28$ % to estimate T_d with a maximum ARD_i value of 25.7 %, see Table 2.9.

Table 2.9 GCMs performance for the thermal decomposition temperature prediction of ILs.

Model	Modeling performance ¹			
	N_{IL}	AARD, %	R^2	Remarks
GCM				
Lazzús (2012c)	198	4.28	0.94	Global set
	120	4.34	N/A	Correlation set
	78	4.18	N/A	Prediction set
Gharagheizi et al. (2013)	613	4.40	0.85	Global set
	489	4.50	0.85	Correlation set
	124	4.30	0.86	Prediction set
Mokadem et al. (2016)	639	4.22	0.87	Global set
	499	4.28	0.86	Correlation set
	140	4.02	0.90	Prediction set

¹ Variable description:

N_{IL} : Number of ILs used in the study.

R^2 : Coefficient of determination.

N/A: Data not available.

GCM of Gharagheizi et al. (2013) can be applied to calculate this property for uranium-, triazolium-, tri-alkyl imidazolium-, tetrazolium-, tetra-alkyl imidazolium-, sulfonium-,

quinary alkyl imidazolium-, pyrrolidinium-, pyridinium-, pyridazinium-, piperidinium-, phosphonium-, oxazolidinium-, morpholinium-, guanidinium-, double imidazolium-, ammonium-, amino acids-, 1-alkyl imidazolium- and 1,3-dialkyl imidazolium-based ILs. These compounds included 58 anions and 313 cations that can be divided into 30 structural elements (i.e., 20 cation structural elements and 10 anion structural elements). This GCM is defined as

$$T_d = \sum_{i=1}^{10} N_{ai} AS_{ai} + \sum_{i=1}^{20} N_{ci} AS_{ci} + 524.5031 \quad (2.103)$$

where AS_{ai} and AS_{ci} are the contributions for T_d of the i th structural element of anions and cations. The contributions of this model were obtained from experimental data of 613 ILs with $T_d = 374 - 740$ K. Table 2.9 summarizes the model performances for the correlation and prediction tests.

Mokadem et al. (2016) proposed a high-order GCM to calculate T_d of ILs using the next expression

$$T_d = 647.492 + \sum_l n_l AT_{2_l} + \sum_k m_k AT_{3_k} - 78.576 AT_0 \quad (2.104)$$

where n_l and m_k are the number of first and second-order groups of the type l and k , respectively and AT_{2_l} and AT_{3_k} are the corresponding group-interaction contributions. The third order contribution is a correction term that incorporates the equivalent number (AT_0) and depends on the IL under analysis. 639 experimental T_d data of ILs were used to define the model parameters. These ILs included the next cations: amino acids, caprolactam, butyrolactam, benzotriazolium, oxazolidinium, ammonium, morpholinium, sulfonium, pyrrolidinium, piperidinium, phosphonium, pyrimidine, guanidinium, tetrazolium, pyridinium, triazolium and imidazolium. The anions included dinitramide, sulfonate, borate, sulfonate, azide, perchlorate, nitrate, 5-nitrotetrazolate, 4,5-dinitroimidazolate, 3-nitro-1,2,4-triazolat, dicyanamide, tricyanomethanide, tris (trifluoromethylsulfonyl) methide, trifluoromethylsulfonate, trifluoroacetate, bis(trifluoromethylsulfonyl)imide, halide, alkylsulfate, carboxylates, dialkylphosphate, tetrafluoroborate and hexafluorophosphate. It was concluded that this model estimated T_d of ILs with low prediction errors, see Table 2.9.

2.2.11 Normal boiling point

Normal boiling point is a relevant property for the application of ILs. This thermodynamic property can be defined as the temperature where the vapor pressure of a liquid is equal to 1 atm (Deng et al., 2016). This property is used as input variable for several correlations and predictive thermodynamic models to estimate the solubility, vapor pressure and critical properties of pure compounds (Valderrama and Robles, 2007; Valderrama et al., 2008, 2012, 2015; Admire et al., 2015). The experimental determination of the normal boiling point of ILs is restricted by their thermal decomposition temperature where several ILs can degrade at temperatures approaching their normal boiling points (Valderrama and Robles, 2007).

The prediction of this property can be performed with GCM proposed by Valderrama and Robles (2007). They modified the model of Valderrama and Venegas (2004), which was based on Lydersen-Joback-Reid approach, to estimate the normal boiling point (T_b , K) of 50 ILs. This model is given by

$$T_b = 198.2 + \sum_i n_i \Delta T_{bi} \quad (2.105)$$

where ΔT_{bi} is the contribution of the structural element i of IL molecule. Note that the estimations obtained with this model were not directly validated because of the lack of experimental data of T_b for several ILs. Therefore, these authors tested the model accuracy via the calculation of ρ of ILs using a generalized correlation proposed by Valderrama and Abu-Sharkh (1989), which only needs the values of T_b , molecular weight and critical properties of tested IL molecules as input variables. The maximum value of ARD_i was 19.3 % using this validation test. The authors stated that the estimations obtained with this GCM were proper to perform engineering calculations using EoS. This model was later extended for 200 ILs (Valderrama et al., 2008) and was also revised to recalculate the group contribution values and incorporating new groups (Valderrama and Rojas, 2009).

Valderrama et al. (2012) updated the studies performed by Valderrama and Robles (2007), Valderrama et al. (2008) and Valderrama and Rojas (2009) where T_b of 1130 ILs were reported. These authors also introduced a new and complementary consistency test based on the calculation of ILs saturation pressure at T_b with a cubic EoS (Valderrama et al., 2012). The model performance for consistency tests based on density and vapor pressure are shown in Table 2.10.

Table 2.10 GCMs performance for the critical properties and normal boiling point prediction of ILs.

Model	Modeling performance ¹		
	N_{IL}	AARD, %	Remarks
GCM			
Valderrama and Robles (2007)	50	5.20	Density test.
Valderrama et al. (2008)	200	5.90	Density test.
Valderrama et al. (2012)	526	15	Density test.
	1130	7	Vapor pressure test.
Valderrama et al. (2015)	667	4.76	Density test.
	1630	2.80	Vapor pressure test.
Valderrama et al. (2021)	637	6.20	Density test.
	4295	3.10	Vapor pressure test.
GC-PT			
Shen et al. (2011)	747	4.40	
FC-CS			
Huang et al. (2013)	490	3.30	
¹ Variable description:			
N_{IL} : Number of ILs used in the study.			

In other study, Valderrama et al. (2015) extended their GCMs to handle ILs with a molar mass > 500 g/mol where 316 compounds were considered. Another extension of these models was proposed by Valderrama et al. (2019) for ILs containing metals where the group contribution values of 29 new metal groups were reported. A computer tool was recently reported by Valderrama et al. (2021) using the Lydersen GCM (Lydersen, 1955) to calculate this IL property. Table 2.10 summarizes the performances of these GCMs. This table also reports the performance of GCMs that have been combined with a fragment contribution approach and the corresponding states method (Huang et al., 2013) and the Patel-Teja EoS (Shen et al., 2011) for the calculation of T_b of ILs.

2.2.12 Critical properties

Critical properties (i.e., volume V_c , pressure P_c , temperature T_c) characterize the vapor-liquid phase behavior of pure components (Tahami et al., 2019). They are used as input variables to identify their chemical nature during process design calculations with thermodynamic models like cubic EoS and to estimate other thermodynamic properties (e.g., density, viscosity, enthalpy, heat capacity, heat of vaporization, vapor pressure, thermal conductivity) (Valderrama et al., 2012). Formally speaking, T_c and P_c of the pure component are the conditions where the properties of its liquid and gas phases are the same because the phase boundaries disappear and, consequently, they cannot be differentiated from each other, while V_c is the corresponding volume of this substance at these thermodynamic conditions (Tahami et al., 2019). There is a limited number of experimental data of critical properties of ILs since several compounds can decompose before reaching the critical point (Valderrama and Robles, 2007). Therefore, the prediction of critical properties of ILs is fundamental to support the design, optimization, modeling and control of different processes. Table 2.10 summarizes GCMs to calculate the critical properties of ILs.

Valderrama and Robles (2007) proposed a modified version of the Lydersen-Joback-Reid GCM to estimate T_c (K), P_c (bar) and V_c (cm³/mol) of ILs as follows

$$T_c = \frac{T_b}{0.5703 + 1.0121 \sum n\Delta T_M - (\sum n\Delta T_M)^2} \quad (2.106)$$

$$P_c = \frac{W}{[0.2573 + \sum n\Delta P_M]^2} \quad (2.107)$$

$$V_c = 6.75 + \sum n\Delta V_M \quad (2.108)$$

where ΔT_M , ΔP_M and ΔV_M are the structural element contributions for the calculation of T_c , P_c and V_c , respectively. These authors highlighted that the estimations of the critical properties obtained with this GCM could not be directly validated for some ILs due to the lack of experimental data. They proposed to perform a density test to assess the model accuracy. This GCM has been utilized to model IL-based processes using different EoS (Carvalho et al., 2009, 2010) and to calculate other properties based on the corresponding states theory (Ge et al., 2008; Valderrama and Zarricueta, 2009). Some updates of this model were reported by Valderrama et al. (2008) and Valderrama and Rojas (2009) where more ILs and structural elements were included and the contribution values were also recalculated.

Valderrama et al. (2012) proposed a new consistency test based on the calculation of the saturation pressure to complement the density test thus including the validation of T_c , P_c , ω and T_b . Other variants of this model were proposed for handling ILs with molecular mass > 500 g/mol (Valderrama et al., 2015) and for ILs containing metals where 29 new metal groups were incorporated (Valderrama et al., 2019).

An approach based on the concepts of the fragment contribution and corresponding states (FC-CS) was proposed by Huang et al. (2013) to estimate the critical properties of ILs. The fragment contribution consists of the division of ILs into fragments: cation fragment, anion fragment and substituent fragment attached to the cation or anion. Other model based on the Patel-Teja EoS and GCM of Valderrama and Robles (2007) (GC-PT) was reported by Shen et al. (2011) to predict the critical properties of ILs. Sattari et al. (2016b) applied GCM and QSPR to estimate T_c of ILs. These authors analyzed the limitations of some GCMs to predict the critical properties of ILs.

Recently, Valderrama et al. (2021) published a simple computer tool based on different approaches (e.g., GCMs, semi-empirical correlations, EoS) to calculate several properties of ILs including critical properties. These authors implemented their GCMs and some improved variants where the consistency tests (density and vapor pressure tests) have been also utilized.

Finally, the results reported in the literature indicated that available GCMs show high modeling errors to estimate the critical properties of some ILs, which fundamentally relies on the lack of experimental information to define proper values of model contributions. However, the accuracy of these models can be improved using alternative numerical tools as will be discussed in this chapter.

2.2.13 Other thermodynamic properties

Different GCMs have been introduced for the calculation of other thermodynamic properties of ILs that are required in PSE. They include the freezing point (Lazzús, 2016a), isobaric expansivity (Gardas and Coutinho, 2009), isothermal compressibility (Gardas and Coutinho, 2009) and static dielectric constant (Zhou et al., 2014). These properties have been estimated using GCMs and some of them are briefly described below where Table 2.11 summarizes their numerical performances.

Table 2.11 GCMs performance for the prediction of other thermodynamic properties of ILs.

		Modeling performance ¹				
Model	Property	N_{IL}	N_{dat}	AARD, %	R^2	Remarks
GCM						
Lazzús (2016a)	Freezing point	63	63	5.30	0.94	Global set
		40	40	5.34	0.95	Correlation set
		23	23	5.25	0.91	Prediction set
Gardas and Coutinho (2009)	Isobaric expansivity	49	109	1.98	0.99	Global set
Gardas and Coutinho (2009)	Isothermal compressibility	22	26	2.53	0.99	Global set
Zhou et al. (2014)	Static dielectric constant	58	128	7.41	N/A	Global set
¹ Variable description:						
N_{IL} : Number of ILs used in the study.			R^2 : Coefficient of determination.			
N_{dat} : Number of experimental data analyzed in the study.			N/A: Data not available.			

For instance, the freezing point is the temperature where a liquid phase changes to a solid state (Lazzús, 2016a). Theoretically, this temperature should be the same as the melting point of a solid but, for the case of ILs, significant differences between these values have been observed (Lazzús, 2016a). A peculiar characteristic of ILs is that they display a substantial supercooling at relatively high T (Hosseini et al., 2019). ILs tend to form crystals during the cooling when they present a freezing transition. For these cases, ILs should be submitted a faster cooling to avoid the crystallization during a freezing transition (Lazzús, 2016a). The experimental determination of freezing point of ILs is difficult due to this type of transition could occur over a wide range of T and also depends on P (Lazzús, 2016a). Lazzús (2016a) proposed a GCM to predict the freezing point of ILs. This model was developed using experimental data of freezing point temperatures from 185 to 466 K of 63 ILs. This model was based on the definition of 10 cation groups (e.g., ammonium, phosphonium, pyrrolidinium, pyridinium and imidazolium) and 21 anion groups (e.g., metal complexes, carboxylates, phosphates, borates, imides, tosylates, sulfonates, pseudohalides and halides). The next equation can be used to estimate the IL freezing point (T_f , K)

$$T_f = 98.599 + \sum_{i=1}^{10} N_{ci}AU_{ci} + \sum_{j=1}^{21} N_{aj}AU_{aj} \quad (2.109)$$

where AU_{ci} and AU_{ai} are the contributions of the i th structural element that conformed the cations and anions, respectively.

The isobaric expansivity is required to perform calculations related to the volumetric behavior of ILs and their changes with respect to state variables (Gardas and Coutinho, 2009). This thermodynamic property is utilized for the analysis of heat transfer fluids and energy storage. Gardas and Coutinho (2009) reported a GCM to calculate this property, which was formulated via the analysis of 109 isobaric expansivity data ($\alpha_p = 4.48 \cdot 10^{-4} - 7.44 \cdot 10^{-4} \text{ K}^{-1}$) at 298.15 K and 0.1 MPa from 49 ILs. This model covered ILs conformed of ammonium, phosphonium, piperidinium, pyrrolidinium, pyridinium and imidazolium. The isobaric expansivity (α_p, K^{-1}) of ILs can be determined with the next expression

$$\alpha_p = AV_1 = \sum_{i=1}^k n_i av_{1i} \quad (2.110)$$

where av_{1i} corresponds to the structural element contributions of this model.

The isothermal compressibility is utilized in fluid dynamics and design like the analysis of compressors and pumps (Hosseini et al., 2019). It has been defined as a measure of the relative volume change of a fluid with respect to P at fixed T (Gardas and Coutinho, 2009). Gardas and Coutinho (2009) introduced a GCM to estimate the isothermal compressibility (k_T, GPa^{-1}) of ILs at 0.33 – 0.73 GPa^{-1} using the next equation

$$k_T = AW_1 = \sum_{i=1}^k n_i aw_{1i} \quad (2.111)$$

where aw_{1i} is the contribution of structural element i . These contributions were determined from the data processing of 26 experimental values at 298.15 K and 0.1 MPa from 22 ILs containing phosphonium, piperidinium, pyrrolidinium, pyridinium and imidazolium.

Finally, the static dielectric constant is required to estimate the electro-viscous effect that occurs when fluids with ions flow in microdevices like microreactors, micropumps and microheat exchangers (Zhou et al., 2014). This property is also a measure of the polarity and solvation capabilities of ILs (Zhou et al., 2014). Few studies have reported static dielectric constant of some ILs (Zhou et al., 2014). The estimation of ILs static dielectric constants (κ) can be performed with the GCM of Zhou et al. (2014), which is given by

$$\kappa = \frac{3}{1 - \frac{AX}{W\rho}} - 2 \quad (2.112)$$

$$AX = \sum_{i=1}^k n_i ax_i \quad (2.113)$$

where ax_i represents the structural element contributions of this model. These authors applied a database of 128 experimental at 278.15 – 343.15 K from 58 ILs to determine the corresponding model parameters.

2.2.14 Estimation of thermodynamic properties using hybrid models from GCM and artificial neural networks

GCMs have some drawbacks for the prediction of physicochemical properties of some ILs with a complex thermodynamic behavior. For example, the viscosity and melting point of ILs are more difficult to predict using only simple structural information of the molecule. As a consequence, the classical GCMs can provide poor estimations for these thermodynamic variables (Hosseini et al., 2019). For these challenging calculations, the ANN can be utilized to improve the model performance and to develop enhanced predictive models.

ANN is a family of statistical learning efficient algorithms that allows to model challenging non-linear behaviors (Lazzús, 2009a; Wang et al., 2017). This numerical tool is based on imitating the complex operation of the biological nervous systems (Fatehi et al., 2014). For interested readers, a detailed description of ANN is provided in the reviews of Abiodun et al. (2018, 2019).

The basis of GCM have been integrated with ANN to develop a new class of thermodynamic models (GCM-ANN) for the estimation of thermodynamic properties of ILs. Some of these models are described in Table 2.12. For example, Valderrama et al. (2009) used a hybrid GCM-ANN model to estimate IL densities. The results indicated that this model predicted the IL density with an acceptable accuracy for engineering calculations. Padaszyński and Domańska (2014) developed a hybrid GCM-ANN model to calculate ILs viscosities. The model performance was substantially better than that of the model of Gardas and Coutinho (2009) to estimate the viscosities of different types of ILs. This model can be applied for predicting the properties of more ILs compared to other models reported in the literature. In these hybrid models, the molecules are generally divided into their structural elements, but instead of using an explicit mathematical expression to define the functionality of these contributions, the ANN is trained to learn the relationship between the input variables (i.e., structural elements that conform the IL molecule including other parameters like the

molar mass, temperature, pressure) and the corresponding output variables that are the desired property for the set of analyzed compounds.

Table 2.12 Different GCM-ANN approaches for the estimation of thermodynamic properties of ILs.

Model	Property	Modeling performance ¹				Remarks
		N_{IL}	N_{dat}	AARD, %	R^2	
Valderrama et al. (2009)	Density	103	399	0.15	N/A	Correlation set
		24	82	0.26	N/A	Prediction set
Lazzús (2009b)	Density	164	2162	0.44	0.99	Global set
		131	1736	0.43	0.99	Correlation set
		33	426	0.49	0.99	Prediction set
Lazzús (2009a)	Density	322	3183	0.55	N/A	Global set
		250	2410	0.57	0.99	Correlation set
		72	773	0.48	0.99	Prediction set
Paduszyński and Domańska (2014)	Viscosity	1484	12405	11.4	0.98	Global set
Lazzús (2016b)	Thermal conductivity	41	400	0.89	0.99	Global set
		30	300	0.91	0.99	Correlation set
		11	100	0.84	0.99	Prediction set
Lazzús et al. (2017a)	Surface tension	162	2307	1.18	0.99	Global set
		108	1728	1.17	0.99	Correlation set
		54	579	1.29	0.99	Prediction set
Wang et al. (2017)	Refractive index	299	2138	0.18	0.96	Global set
		277	1711	0.17	0.97	Correlation set
		190	427	0.22	0.93	Prediction set

¹ Variable description:

N_{IL} : Number of ILs used in the study.

R^2 : Coefficient of determination.

N_{dat} : Number of experimental data analyzed in the study.

N/A: Data not available.

2.3 Remarks and perspectives of GCMs for ionic liquids property calculations

This chapter has covered a wide spectrum of GCMs for the prediction of thermodynamic properties involved in the PSE. The analysis of model performance suggested that some GCMs have been obtained from databases with a limited number of experimental values thus generating that they could not be employed to predict the properties of different ILs for the case of novel molecules. For some operating conditions, the errors of GCMs to predict some ILs thermodynamic properties could be relatively significant to carry out a proper process design and simulation. It appears that the more challenging IL properties to be estimated correspond to the critical temperature and pressure, while the heat capacity and density are the properties more analyzed and studied for ILs where different GCMs can be used for their calculation. Although high order GCMs may offer better results for the IL property calculations, they require more detailed information of IL molecules and additional parameters. Under this scenario, the first-order GCMs should be the first and straightforward option for the prediction of ILs thermodynamic properties but additional efforts are required to reduce their modeling errors especially for those properties that have been not extensively studied in the literature.

The improvement of GCMs for obtaining reliable thermodynamic calculations of ILs should be cover some of the next key issues:

a) As stated, the contributions of structural elements that compose the ILs molecules are obtained from the experimental data processing. For some ILs properties, the available experimental values are limited or can contain significant uncertainties. Consequently, it is convenient to review and recalculate the model contributions of available GCMs via the incorporation of new experimental information that has been reported during the last years. Also, it is necessary to establish experimental working plans to expand the number of ILs and their properties as well as to quantify novel and well-known compounds but at different operating conditions. This aspect is fundamental for creating a ILs database with a vast quantify of experimental values to develop reliable GCMs for the calculations required in PSE.

b) An in-deep analysis of the effect of type of cations and anions conforming the molecular structure of ILs on the prediction of their thermodynamic properties is necessary and should be performed in forthcoming studies. To the best of authors knowledge, this analysis has not

been reported in the literature for the thermodynamic properties of ILs. Since some estimations have indicated that around 10^6 possible ILs can be synthesized, the results of this analysis can contribute to define molecular structure – property relationships with the aim of understanding the thermodynamic behavior of these compounds and to improve the prediction capabilities of GCMs.

c) Several authors have applied local optimization methods to resolve the parameter estimation problems involved in the determination of structural element contributions of CGMs. However, these multivariable parameter identification problems are challenging and imply potential nonconvex objective functions. So, global optimization solvers should be applied to minimize the corresponding objective functions with the aim of identifying the best model parameters. In this direction, the application of novel stochastic optimization methods is an interesting alternative.

d) It seems clear that ANN can improve the performance of GCMs for the calculation of ILs thermodynamic properties. Several advances and improvements of ANN have reported during the last decades thus opening the possibility of applying improved ANN approaches to enhance the capabilities of thermodynamic models based on GCM. Also, other artificial intelligence tools can be tested and categorized to identify the best one for the calculation of a specific thermodynamic property.

GCMs represents an effective approach to support the computer-aided process design of novel processes and operation units based on ILs but it is mandatory to improve their performance for obtaining reliable estimations of performance metrics like separation efficiencies, energy requirements, costs and other relevant metrics utilized in the decision-making stage.

Chapter 3. Calculation of liquid C_p of pure compounds using an improved thermodynamic model based on group contributions and artificial neural networks

3.1 Introduction

Heat capacity (C_p) is an essential property for the PSE in chemical engineering because it is used as input data in several calculations for the design, modeling and optimization of evaporation, reaction, extraction, heating and distillation units (Nancarrow et al., 2015). This thermochemical property is defined in terms of a relationship between the enthalpy and temperature (T) of a thermodynamic system at constant pressure (P) (Gardas and Coutinho, 2008a). It is important for the thermodynamic analysis and energy transfer calculations in reactive and non-reactive systems, as well as the core of energy balance and the evaluation of the effect of T on phase behavior (Zábranský and Růžicka Jr., 2004; Kolská et al., 2008; Chen et al., 2019). Consequently, the process simulation software requires C_p as basic thermodynamic data (Nancarrow et al., 2015).

Overall, C_p of several compounds in liquid phase has been measured at only one T value or, in the best scenario, for a restricted T range such as the saturation conditions (Kolská et al., 2008; Jovanović et al., 2011). This implies that the experimental information on this property, which has been reported in the literature, can present some limitations for specific chemical families, and in the case of new and emerging compounds, experimental C_p measurements are not available (Kolská et al., 2008). A wide spectrum of thermodynamic models has been developed using the available C_p experimental data. Particularly, the C_p calculation of pure compounds can be performed using different approaches such as GCMs (Chen et al., 2019; Taherzadeh et al., 2020; Haghbakhsh et al., 2021; Han et al., 2023; Xia et al., 2023), EoS (Coniglio et al., 1993; Zhu et al., 2020), empirical and semi-empirical correlations (Conesa et al., 1998; Dadgostar and Shaw, 2012), corresponding states equations (Lyman and Danner, 1976; Tarakad and Danner, 1977; Gao et al., 2018; Do Carmo et al., 2020), ML algorithms like ANNs (Aldosari et al., 2021) and models obtained from the combination of these alternatives (Prasad et al., 1993; Diedrichs et al., 2006; Oster et al., 2018). However, a common failure of several C_p models is associated with their high prediction errors in specific regions of the phase diagram mainly for the phase-transition zones and near the critical point. Figure 3.1 illustrates the general trend of C_p versus T where

it can be observed that this property tends to increase exponentially near to the critical point. This thermodynamic behavior has been analyzed and explained in detail by Gopal (1981), and it can be extended to all pure components at operating conditions near phase transitions (e.g., saturation lines for solid-liquid and vapor-liquid equilibria). Therefore, the models reported in the literature can generate unreliable estimations in different regions of the phase diagram. Herein, it is convenient to recall that a common practice of chemical engineers for process design is the estimation of C_p using the well-known polynomial equations based on T , which can be described as follows

$$C_p = C_1 + C_2 \cdot T + C_3 \cdot T^2 + C_4 \cdot T^3 + C_5 \cdot T^4 \quad (3.1)$$

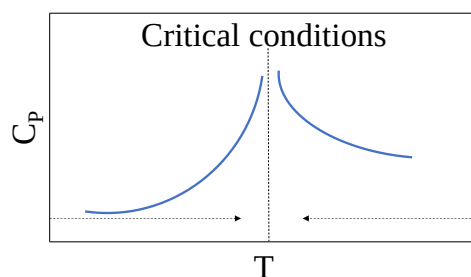


Figure 3.1 General trend of the C_p versus the operating temperature near the critical point.

These correlations are usually unable to estimate the correct trend of C_p , as shown in Figure 3.2, where n-nonane is employed as an example using the parameters reported by Perry et al. (1997). These empirical models can be only applicable at a restricted region of the reduced temperature (T_r), which is 0.37–0.55 for this case. Failure of this model can occur in the C_p calculation of several compounds, even when using the parameters available in commercial process simulators such as Aspen Plus®. On the other hand, some predictive and correlative models (e.g., EoS) have been proposed for specific compounds and they cannot be utilized to estimate this property for other molecules under a wide range of T conditions. These drawbacks generate process design uncertainties and errors due to the inaccurately calculated C_p values. It is clear that the development of reliable C_p models is paramount to build a robust thermodynamic framework for PSE.

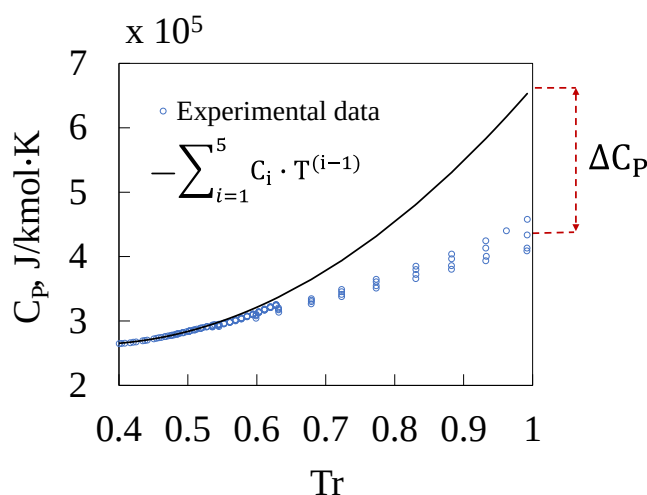


Figure 3.2 C_p estimations obtained for *n*-nonane using the correlation model proposed by Perry et al. (1997).

The thermodynamic models based on the group contribution principle and artificial intelligence offer several advantages to calculate thermophysical properties of pure substances (Taherzadeh et al., 2020; Haghighbakhsh et al., 2021; Han et al., 2023; Xia et al., 2023). For the C_p prediction of liquids, several group contribution-based models have been proposed (Luria and Benson, 1977; Růžička Jr. and Domalski, 1993a, 1993b; Kolská et al., 2008; Ceriani et al., 2009; Jovanović et al., 2011; Chen et al., 2019; Taherzadeh et al., 2020; Haghighbakhsh et al., 2021; Xia et al., 2023). Luria and Benson (1977) reported a GCM for the property calculation of hydrocarbons, while a second-order GCM for organic liquids containing oxygen, halogens, sulfur and nitrogen was introduced by Růžička Jr. and Domalski (1993a, 1993b). Kolská et al. (2008) reported a three-level GCM for C_p calculation of organic liquids, which was based on the studies of Marrero and Gani (2001) and Constantinou and Gani (1994). Ceriani et al. (2009) introduced a GCM for the calculation of C_p of fatty compounds including esters triacylglycerols and alcohols. Another GCM was reported by Jovanović et al. (2011) for C_p calculation of liquids at high P . Recently, several GCM have been developed for the C_p calculation of specific families of compounds of industrial interest, such as deep eutectic solvents (Taherzadeh et al., 2020) and ILs (Gardas and Coutinho, 2008a; Soriano et al., 2010a; Valderrama et al., 2011c; Müller and Albert, 2014; Ahmadi et al., 2015; Nancarrow et al., 2015; Chen et al., 2019).

The thermodynamic models based on ANNs can predict the nonlinear behavior of thermophysical properties using a representative set of input variables, proper network structure and training algorithm (Mondejar et al., 2017). Several authors have used ANNs to estimate the C_p of refrigerants (Neng et al., 2019), deep eutectic solvents (Bunquin and Caparanga, 2021), ILs (Valderrama et al., 2011a), and other chemicals (Santak and Conduit, 2019; Alade et al., 2020). For example, Neng et al. (2019) determined the residual isobaric C_p of pure HFO and HFC liquid refrigerants. They identified the optimal network structure via trial-and-error where the input variables were $1-T_r$, reduced pressure, acentric factor and critical P ratio. Alade et al. (2020) used ANNs for C_p prediction of CuO/water nanofluids. ANNs-based model was trained employing an experimental database of C_p of CuO nanoparticles, their volume fractions and T. C_p calculation of deep eutectic solvents containing phosphonium and ammonium was performed by Bunquin and Caparanga (2021) using ANNs. The input parameters of this model included structure-related parameters, mole ratio, T, and molecular weight. It is convenient to highlight that the available ANNs models for the C_p estimation can be applied to specific chemical families, thus opening the window for developing new models with improved characteristics that can be used on a variety of molecules. Although different studies have addressed the application of ANNs for the C_p calculation, few efforts have focused on exploiting the synergy obtained by combining GCM and ANNs to calculate this thermodynamic property. These hybrid models can utilize the parameters of the molecular structures as ANNs input variables for the property estimation, thus generating a versatile model for thermodynamic calculations of different types of compounds.

This chapter reports the development, performance, characterization and evaluation of a thermodynamic model based on GCM and ANNs for C_p calculation of pure compounds in liquid phase. This improved model was obtained using a first-order approach, in which the functional groups of the molecules are the main input parameters for ANNs. Different structures and configurations for this GCM–ANNs model were evaluated and compared to identify the best one in terms of its performance without data overfitting. A detailed analysis of modeling errors was carried out, in terms of chemical families and functional groups, using a set of statistical metrics with the aim of identifying the limitations and capabilities of this new model. A comparison of the model predictions with those obtained using the process

simulation software Aspen Plus[®] is also reported. The results indicated that this improved GCM–ANNs thermodynamic model is versatile for the C_p prediction of different chemical species and can be used as an alternative tool for PSE tasks.

3.2 Formulation of an improved GCM–ANNs model for C_p calculation of pure compounds in liquid phase

3.2.1 Characteristics of C_p experimental database

An extensive experimental database of liquid-phase C_p at different T conditions was collected from 766 references using NIST. This database included a total of 22715 experimental values of C_p ranging from 6.71E+04 to 1.11E+06 J/kmol·K at 85.59 – 860 K for 441 compounds. The critical temperature (T_c , K) and molecular weight (MW, g/mol) of these compounds ranged from 227.53 to 1149 K and 28.05 to 544.9 g/mol, respectively. The database reported in this study contains more experimental data than those used in other studies for the development of C_p thermodynamic models (e.g., Luria and Benson (1977), Růžicka Jr. and Domalski (1993a, 1993b), Kolská et al. (2008), Ceriani et al. (2009), Jovanović et al. (2011), Valderrama et al. (2011a), Chen et al. (2019), Neng et al. (2019), Santak and Conduit (2019), Alade et al. (2020), Taherzadeh et al. (2020), Bunquin and Caparanga (2021)). This database contained C_p experimental measurements of alkanes, alkenes, alcohols, carboxylic acids, ketones, ethers, esters, compounds containing halogens, sulfur and silicon, among others. Other published C_p models cannot handle several chemical families (Luria and Benson, 1977; Růžicka Jr. and Domalski, 1993a, 1993b; Kolská et al., 2008; Ceriani et al., 2009; Jovanović et al., 2011; Chen et al., 2019; Neng et al., 2019; Taherzadeh et al., 2020; Bunquin and Caparanga, 2021). Finally, isomers were not included in the C_p database.

3.2.2 Description of the GCM–ANNs approach used for liquid C_p calculation of pure compounds

GCM considers that the compound properties depend on the structural elements (i.e., functional groups) that compose the molecules under analysis (Nannoolal et al., 2007). For first-order GCM, the group contribution is assumed to be the same, regardless of the molecule

in which it is present (Valderrama et al., 2012). Therefore, the molecules in the C_p database were divided into 49 functional groups, inside and outside the molecular rings, in addition to oxygen-, nitrogen-, halogenated-, silicon-, sulfur-, phosphorus- and aluminum-based functionalities, see Table 3.1. Appendix E provides an Excel file (Database_compounds.xlsx) containing all the compounds used in this study and the type and frequency of the functional groups that compose the tested molecules. On the other hand, ANNs are surrogate modeling algorithms inspired by the biological nervous system behavior (Fatehi et al., 2014). They consist of a set of interconnected processing units (called neurons), which can represent nonlinear and multivariable spaces (Mondejar et al., 2017). These neurons are arranged in a predefined network structure consisting of hidden, input, and output layers, which also contain different activation functions (Mondejar et al., 2017). Based on these facts, the improved GCM-ANNs model integrates the basis of group contribution with the structure of an artificial intelligence algorithm for the C_p prediction. This improved model uses the functional groups that compose the molecules in the database as the main input parameters of the ANNs, see Figure 3.3. A feed-forward backpropagation ANNs was utilized to establish the implicit nonlinear functionality between these functional groups and C_p of tested molecules. Different input variables and ANNs configurations were evaluated to determine the GCM-ANNs model with the best performance (without overfitting). Four GCM-ANNs models were evaluated and assessed in C_p calculations. The type and frequency (n_i) of the functional groups that compose the molecules of tested compounds and their MW (g/mol) were defined as the ANNs input variables for all the models. These parameters allowed for differentiation between the molecules contained in the database. An additional input variable related to the operating T was included to improve the C_p predictions. The general trend of the C_p versus T near the critical point (see Figure 3.1) suggests that C_p calculation in this phase diagram region is challenging. Therefore, several functionalities involving T_c were tested to determine their impact on the C_p estimations near the critical point. These functionalities were represented by T , T_r , $T - T_c$ or $\frac{1}{(T-T_c)}$. Note that they were included as an additional input variable for the M_1 , M_2 , M_3 and M_4 models, respectively. These models were implemented with 1 and 2 hidden layers and different hidden neurons from 1 to 20 where symmetric and asymmetric distributions of these processing units were also analyzed.

The symmetric configurations implied that the number of hidden neurons in each layer was the same, while the asymmetric configurations operated with hidden layers containing a dissimilar number of hidden neurons. 65 ANNs configurations were evaluated for each model (i.e., M_1 , M_2 , M_3 and M_4) and the best configuration without overfitting was identified and selected for a detailed performance analysis. The output layer used a linear transfer function, while the hidden layers employed a sigmoid transfer function. ANNs parameters (i.e., weights and biases) were optimized using the Levenberg–Marquardt training algorithm, which is recommended for several problems and widely used in the literature (Gharagheizi et al., 2011; Deng et al., 2016). A random procedure was applied to separate the C_p database into three subsets: 70% for training, 15% for testing and the remaining 15% for validation. Note that this random selection was constrained to ensure that different molecules were included in the three subsets (i.e., 406 molecules for training, 380 molecules for testing and 355 molecules for validation). The testing subset included 29 molecules, which were not considered for the training and validation subsets. The identification of the best ANNs biases and weights is a global optimization problem and 100 independent trainings were performed using different initial values of weights and biases. The best solution found from this set of trainings was reported for the GCM–ANNs model. The performance function used in ANNs training was the mean square error (MSE), which was defined by

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (C_{P_i}^{\text{exp}} - C_{P_i}^{\text{cal}})^2 \quad (3.2)$$

where $C_{P_i}^{\text{cal}}$ and $C_{P_i}^{\text{exp}}$ are the calculated and experimental data, and m is the number of analyzed experimental data (i.e., 15904 data for training, 3408 data for testing, 3403 data for validation and 22715 data for the overall database). The calculations were carried out using the model written in Matlab® language, whose acronym is C_p_liquid .

Table 3.1 Functional groups considered for the GCM-ANNs model to calculate C_p of pure compounds in liquid phase.

Groups outside rings					
-CH ₃	-CH ₂ -	>CH-	>C<	=CH ₂	
=CH-	=C<	=C=	≡C-	≡CH	
Groups inside rings					
-CH ₂ -	>CH-	=CH-	>C<	=C<	
Groups with oxygen					
-OH (alcohol)	-OH (phenol)	-O- (outside rings)	-O- (inside rings)	>C=O (outside rings)	
>C=O (inside rings)	-CHO (aldehyde)	-COOH (acid)	-COO- (ester)	=O	
Groups with nitrogen					
>NH (outside rings)	>NH (inside rings)	>N- (outside rings)	>N- (inside rings)	-N= (outside rings)	
-N= (inside rings)	-NH ₂ (outside rings)	-NO ₂	-NO ₃	-CN	
Groups with halogen					
-F	-Cl	-Br	-I		
Groups with sulfur					
-S- (outside rings)	-S- (inside rings)	-SO ₂ - (outside rings)	-SO ₂ - (inside rings)	-SH	=S<
Groups with silicon					
>Si< (outside rings)			>Si< (inside rings)		
Groups with phosphorus					
=P≤					
Groups with aluminum					
>Al-					

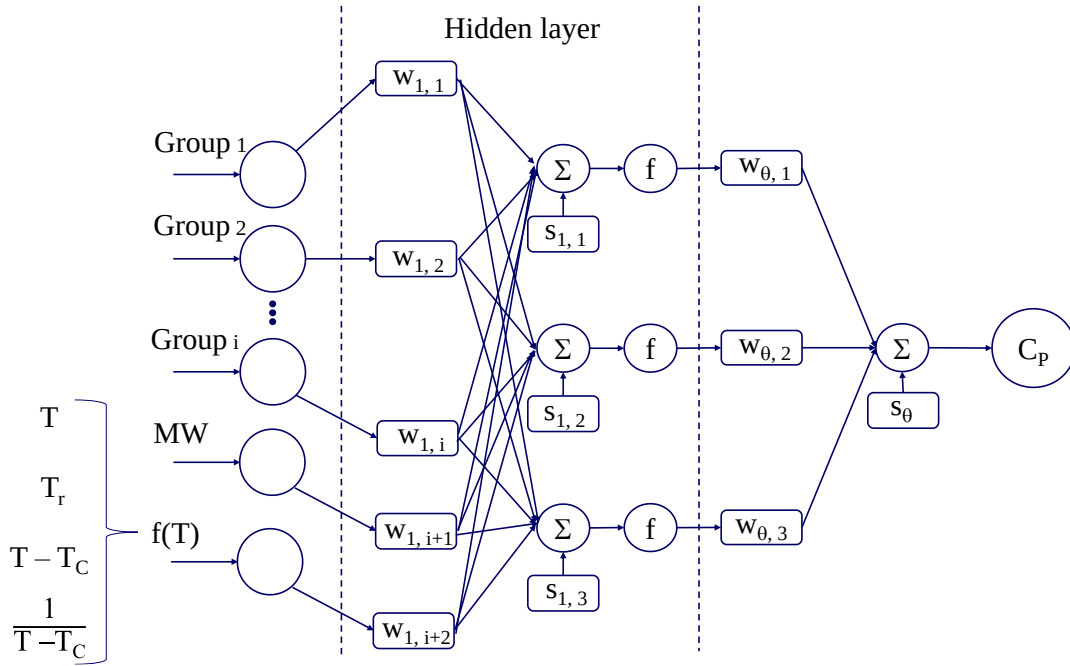


Figure 3.3 Input variables of the GCM-ANNs model for the calculation of C_p of pure compounds in liquid phase.

3.2.3 Evaluation of GCM-ANNs model predictions

Different statistical metrics were employed to evaluate the model performance for the C_p calculation. These statistical metrics included the determination coefficient (R^2), the average absolute relative deviation (AARD), the standard deviation (σ), the relative deviation (RD_i), and the absolute relative deviation (ARD_i), which are given by

$$RD_i = 100 \cdot \frac{C_{P_i}^{\text{exp}} - C_{P_i}^{\text{cal}}}{C_{P_i}^{\text{exp}}} \quad (3.3)$$

$$ARD_i = 100 \cdot \left| \frac{C_{P_i}^{\text{exp}} - C_{P_i}^{\text{cal}}}{C_{P_i}^{\text{exp}}} \right| \quad (3.4)$$

$$AARD = \frac{1}{m} \sum_{i=1}^m 100 \cdot \left| \frac{C_{P_i}^{\text{exp}} - C_{P_i}^{\text{cal}}}{C_{P_i}^{\text{exp}}} \right| \quad (3.5)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^m (ARD_i - AARD)^2}{m}} \quad (3.6)$$

$$R^2 = \left(\frac{\sum_{i=1}^m (C_{P_i}^{\text{exp}} - \overline{C_{P_i}^{\text{exp}}})(C_{P_i}^{\text{cal}} - \overline{C_{P_i}^{\text{cal}}})}{\sqrt{\sum_{i=1}^m (C_{P_i}^{\text{exp}} - \overline{C_{P_i}^{\text{exp}}})^2 \sum_{i=1}^m (C_{P_i}^{\text{cal}} - \overline{C_{P_i}^{\text{cal}}})^2}} \right)^2 \quad (3.7)$$

$$\overline{C_{P_i}^{\text{exp}}} = \frac{\sum_{i=1}^m C_{P_i}^{\text{exp}}}{m} \quad (3.8)$$

$$\overline{C_{P_i}^{\text{cal}}} = \frac{\sum_{i=1}^m C_{P_i}^{\text{cal}}}{m} \quad (3.9)$$

Finally, the best GCM–ANNs thermodynamic model was compared with the calculations obtained using a GCM implemented in Aspen Plus® software. This process simulator has only one GCM for the C_P estimation and corresponded to the model of Růžička Jr. and Domalski (1993a, 1993b). This model can predict this thermodynamic property under specific T conditions, mainly at normal boiling T and 298.15 K. However, it cannot calculate the C_P of some molecules because of the absence of functional group contributions. Therefore, this comparative analysis focused on 168 substances for which the C_P calculations were done at 298.15 K and/or normal boiling conditions. Note that the GCM available in the Aspen Plus® simulator and the best GCM–ANNs model of this chapter can be applied to these conditions and their results can be directly compared without any bias.

3.3 Results and discussion

3.3.1 Impact of ANNs structure on the C_P calculation and identification of the best GCM–ANNs model

Figure 3.4 shows the model performance, based on AARD, of tested GCM–ANNs models using symmetric ANNs configurations for the C_P calculation. In general, an increase in the number of neurons and hidden layers reduced AARD values. For instance, the M_1 model with 1 hidden layer showed AARD values ranging from 6.25 to 1.84% using 1 and 13 neurons, respectively. This model with 2 hidden layers and a symmetric configuration showed AARD values of 6.12 and 1.55% employing 1 and 17 hidden neurons, respectively. Similar results were obtained for M_2 , M_3 and M_4 models, see Figure 3.4. However, the impact of increasing the hidden neurons and layers was directly reflected in the model overfitting. This analysis allowed the identification of the ANNs configuration where the model overfitting occurred. For the symmetric ANNs configuration with 1 hidden layer, the overfitting was present from 3, 7, 8 and 9 neurons for M_4 , M_3 , M_2 , and M_1 models, respectively; while these models with 2 hidden layers overfitted from 2, 4, 5 and 6 neurons, respectively. The AARD results for the thermodynamic models with asymmetric ANNs configurations are shown in Figure 3.5. The

values of this statistical metric ranged from 1.46 to 5.88% for M_1 , 1.52 to 7.15% for M_2 , 1.52 to 7.09% for M_3 and 2.57 to 11.06% for M_4 , respectively. The AARD values also decreased with the increment in hidden neurons, as observed for models with symmetric ANNs configurations. It was also found that all asymmetric configurations with 10, 15, and 20 neurons in the first hidden layer overfitted, regardless of hidden neurons in the second layer. The models with 1 and 5 neurons in the first hidden layer are usually overfitted from 10 neurons in the second layer. These results indicate that asymmetric GCM-ANNs models were outperformed by their symmetric counterparts in terms of model overfitting using a similar number of hidden neurons. The results clearly showed that the model performance for both symmetric and asymmetric configurations was: $M_1 \cong M_2 > M_3 > M_4$.

Table 3.2 summarizes the performance metrics of the best GCM-ANNs models without overfitting. These configurations included ANNs with 1 hidden layer, and both symmetric and asymmetric structures. In general, the symmetric configurations showed better performance than their asymmetric counterparts with AARD values ranging from 2.41 to 11.17% in the C_p calculation. It was found that the ANNs configurations with the best performance corresponded to a symmetric configuration for the M_1 model with 5 neurons in both hidden layers (AARD = 2.41%, $R^2 = 0.99$), the M_2 model with 7 hidden neurons and 1 layer (AARD = 2.61%, $R^2 = 0.99$), the M_3 model with 6 hidden neurons and 1 layer (AARD = 3.02%, $R^2 = 0.99$), and the M_4 model with 2 hidden neurons and 1 layer (AARD = 5.62%, $R^2 = 0.98$). Figure 3.6 compares the experimental and calculated C_p using the best GCM-ANNs models. M_3 and M_4 were outperformed by M_1 and M_2 , which included T and T_r as input parameters to the ANNs, respectively. The deviations between calculated and experimental C_p of the M_2 model were lower than those obtained with other tested GCM-ANNs models. Therefore, the C_p calculation performance of these GCM-ANNs models followed the next trend: $M_4 < M_3 < M_1 < M_2$. As stated, the M_2 model with T_r as an ANNs input parameter improved the predictions obtained near the critical point (i.e., $T_r \rightarrow 1$) for different compounds of the database. For illustration, Figure 3.7 shows the comparison of the C_p estimations obtained for ethylene and 1,1,1-trifluoroethane using the M_1 and M_2 models. These results correspond to data used for training, validation and testing. M_2 model clearly outperformed M_1 model in calculating C_p near to T_c .

Table 3.3 reports the statistical performance metrics of the M_2 model with 1 hidden layer and 7 neurons for the training, testing and validation subsets. The AARD values ranged from 2.32 (for the validation subset) to 4.11% (for the testing subset). The highest error in the C_p estimation was obtained for the validation subset (with an ARD_i of 44.27%). A more detailed analysis of the C_p calculation performance of the M_2 model with 7 neurons in 1 hidden layer is given in Table 3.4 for different chemical families. ARD_i values ranged from 1.48E-05 to 44.27%, and the lowest and highest modeling errors were obtained for 2,4,6-trinitrotoluene and n-pentane, respectively. It was identified that the highest errors for the C_p calculation were obtained for cycloalkanes, ethers, polyfunctional halides, mercaptans and compounds containing NO_2 , F, and I with AARD values of 3.08 – 5.17%. This model behavior could be explained by considering the high uncertainty of the C_p experimental data quantified for some of these chemicals. On the other hand, the M_2 model calculated C_p of different chemical families (e.g., alkenes, polyols, ketones, esters, amines/iminines/amides, silanes/siloxanes, and compounds containing Br) with AARD < 2%. Table 3.5 shows the M_2 model performance in terms of the main functional groups where the AARD values ranged from 1.90 to 3.89%. The highest modeling errors (i.e., AARD > 3%) were obtained for -F, -Cl and >C< (without rings). Figure 3.8 shows the AARD for the C_p calculation as a function of T_r using the best M_2 GCM–ANNs model. These results indicated that the model deviations were the highest (with an AARD of 8.44%) near to the critical point (i.e., T_r from 0.9 to 1). In general, the highest modeling errors were obtained for T_r from 0.8 to 1. In the region of $T_r = 0.3 - 0.8$, this thermodynamic model achieved AARD < 3%.

ARD_i values for the C_p calculation using the best M_2 GCM–ANNs model versus MW and T_r are shown in Figure 3.9. In general, this model showed the highest errors in C_p calculation for low-molecular-weight compounds (i.e., MW < 200 g/mol). This trend has been explained by Nannoolal et al. (2007) to calculate the thermodynamic properties of this type of molecules. Low-molecular-weight compounds generally do not follow the homologous series trend within a chemical family. Therefore, the errors involved in the calculation of their properties using any model usually tend to be higher despite the wide availability of experimental C_p data for these chemicals.

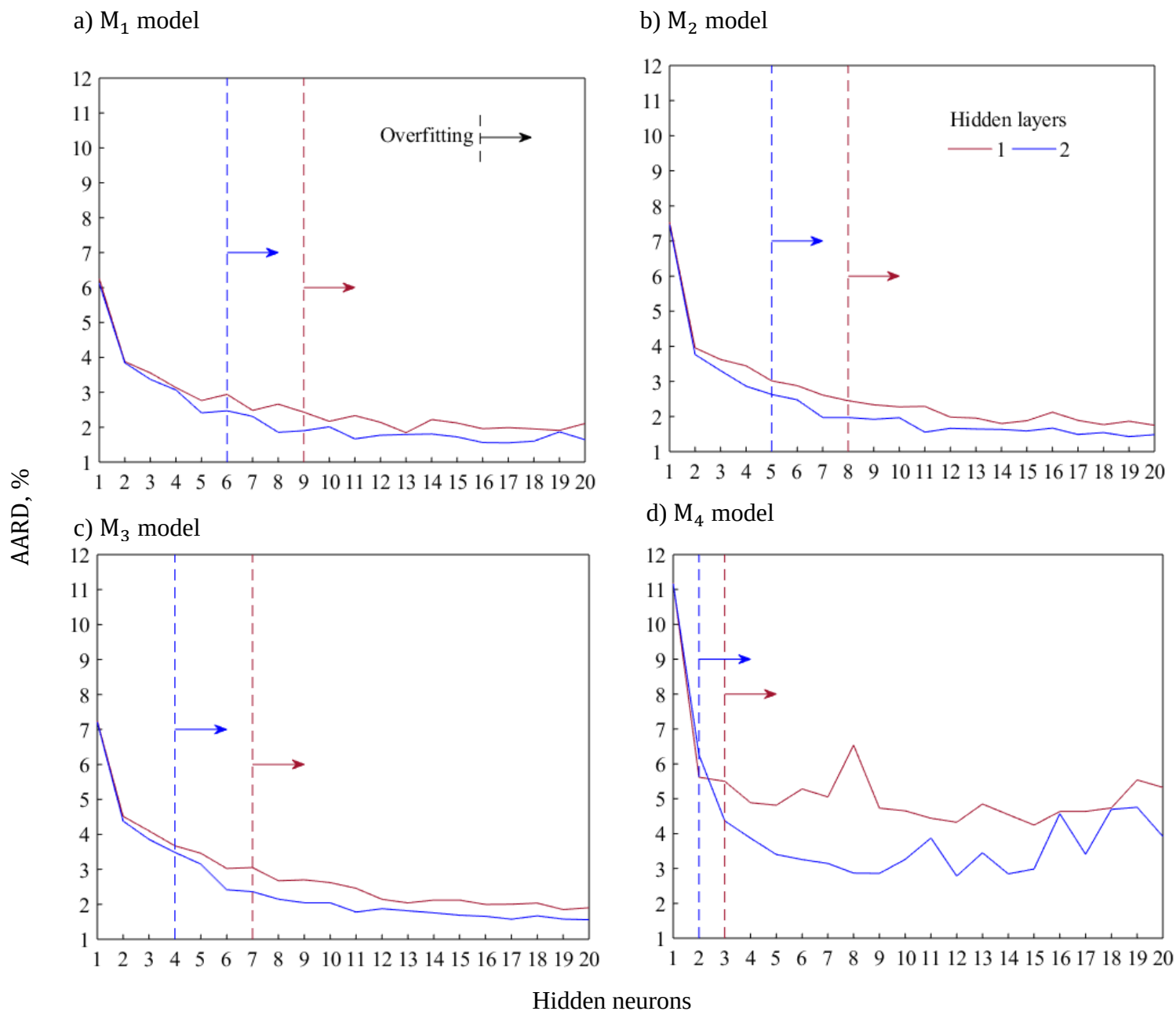


Figure 3.4 AARD for the C_p calculation of pure compounds in liquid phase using different GCM-ANNs models with symmetric ANNs configuration.

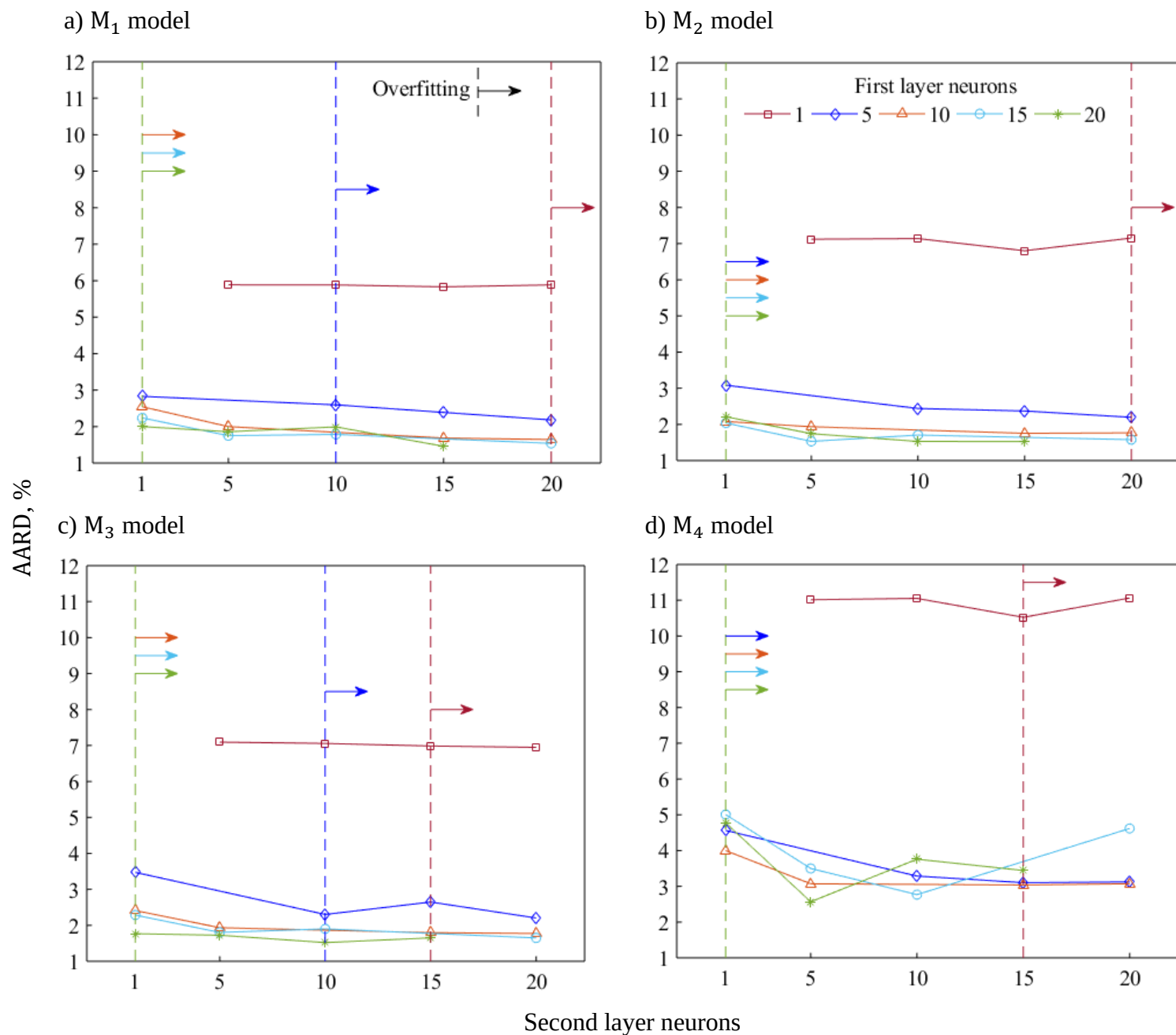


Figure 3.5 AARD for the C_p calculation of pure compounds in liquid phase using different GCM-ANNs models with asymmetric ANNs configuration.

Table 3.2 Results of statistical performance metrics of the best GCM–ANNs models for the C_p calculation of pure compounds in liquid phase.

Model	ANNs inputs	Configuration	Number of neurons		Statistical metric				
			1 st	2 nd	AARD	Min ARD_i	Max ARD_i	σ	R^2
			layer	layer	, %	, %	, %	, %	
M ₁	Functional	1 Hidden layer	7	0	2.48	1.79E-04	70.13	3.56	0.99
	groups, MW,	Symmetric	5	5	2.41	1.30E-04	55.15	3.29	0.99
	T	Asymmetric	5	1	2.83	1.00E-04	62.19	3.77	0.99
M ₂	Functional	1 Hidden layer	7	0	2.61	1.48E-05	44.27	3.68	0.99
	groups, MW,	Symmetric	4	4	2.86	2.23E-05	47.69	3.47	0.99
	T _r	Asymmetric	1	15	6.80	1.06E-04	64.33	7.20	0.98
M ₃	Functional	1 Hidden layer	6	0	3.02	3.26E-05	48.06	4.10	0.99
	groups, MW,	Symmetric	3	3	3.86	4.87E-06	57.08	4.74	0.99
	T – T _C	Asymmetric	5	1	3.47	1.40E-04	80.02	4.70	0.99
M ₄	Functional	1 Hidden layer	2	0	5.62	1.69E-04	52.46	6.01	0.98
	groups, MW,	Symmetric	1	1	11.17	9.95E-05	96.06	9.90	0.94
	$\frac{1}{(T-T_C)}$	Asymmetric	1	5	11.01	1.22E-04	92.17	9.76	0.94

[†] Symbols Min and Max represent the minimum and maximum values of ARD_i .

In summary, the C_p estimations obtained with the M₂ GCM–ANNs model were accurate for several molecules included in the database. For illustration, Figure 3.10 displays the C_p estimations at different T conditions for ethane, ethylene, n-propylcyclohexane, naphthalene, benzaldehyde, oleic acid, benzophenone, isocyanatobenzene, 4-fluorotoluene, 1,1- dichloro-2,2,2-trifluoroethane, benzo[b]thiophene and decamethyltetrasiloxane. The data associated with these compounds were distributed in the training, validation and testing subsets.

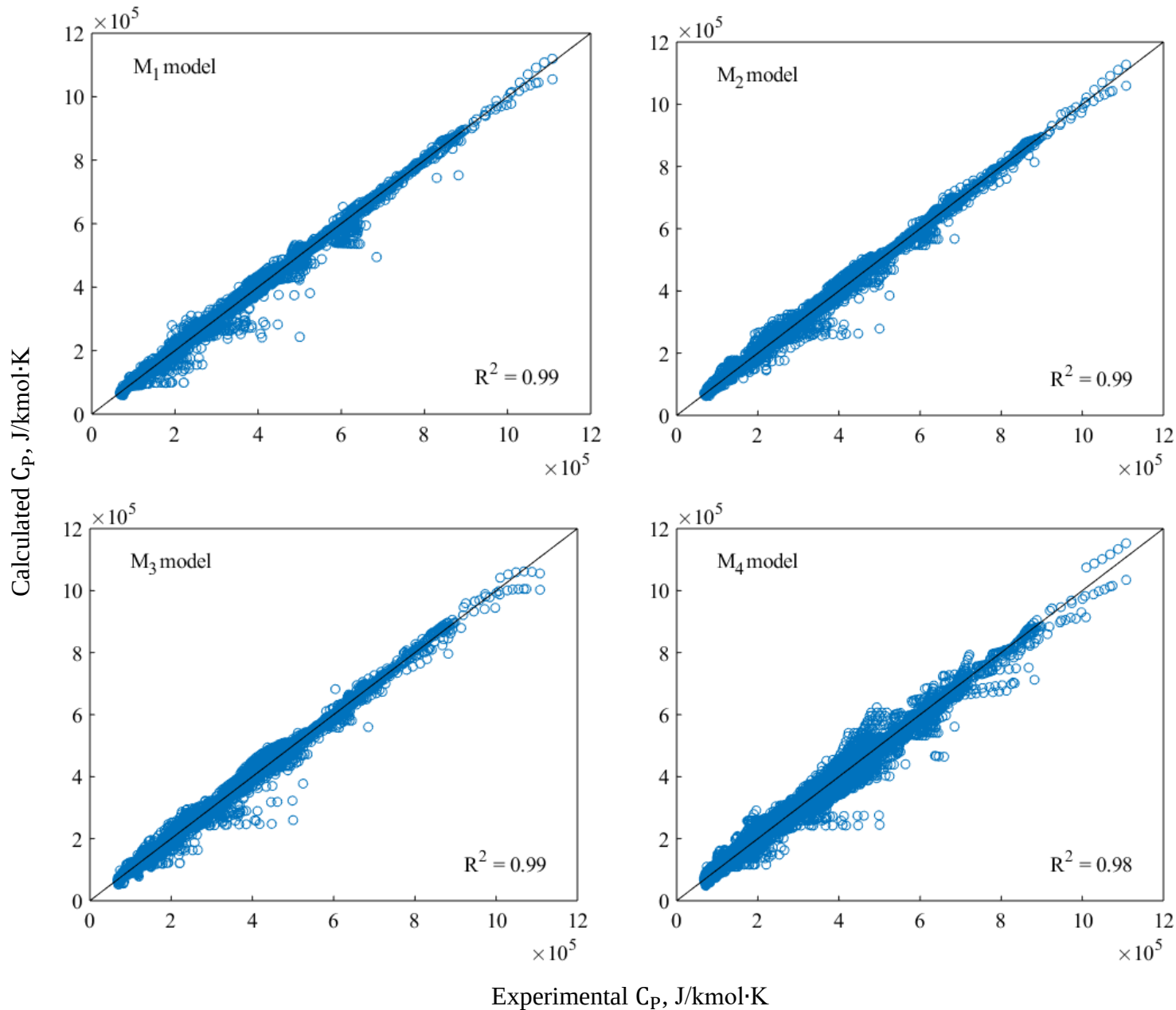


Figure 3.6 Results of the calculation of pure compounds C_p using different GCM-ANNs models.

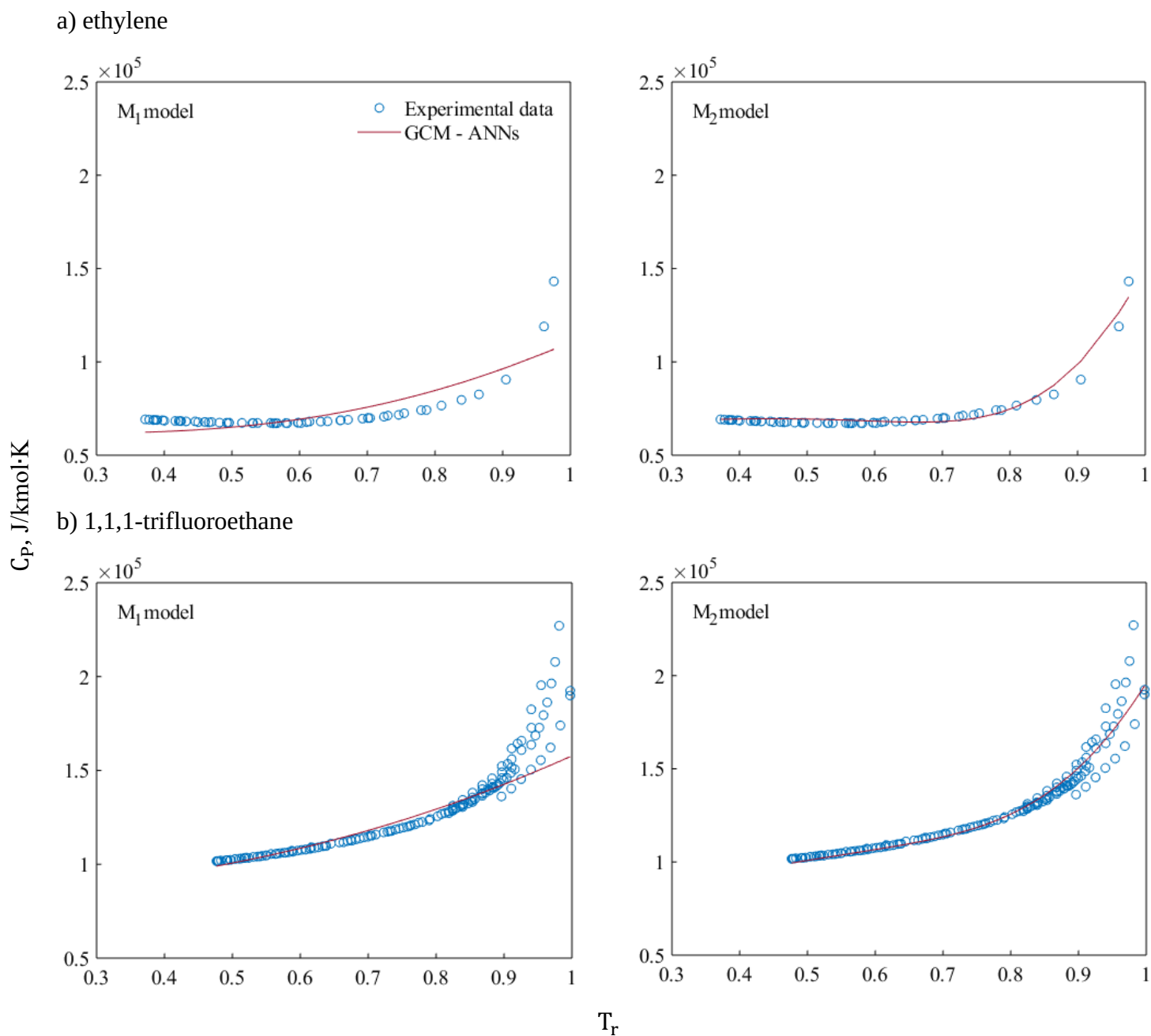


Figure 3.7 C_p calculation for ethylene and 1,1,1-trifluoroethane using the M_1 and M_2 GCM-ANNs models.

Table 3.3 Statistical performance metrics to calculate C_p for the training, testing and validation subsets using the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.

Subset	No. Data	Statistical metric				
		AARD, %	Min ARD_i , %	Max ARD_i , %	σ , %	R^2
Training	15904	2.35	1.48E-05	37.26	3.02	0.99
Testing	3408	4.11	1.10E-03	37.54	6.04	0.99
Validation	3403	2.32	3.24E-04	44.27	2.94	0.99

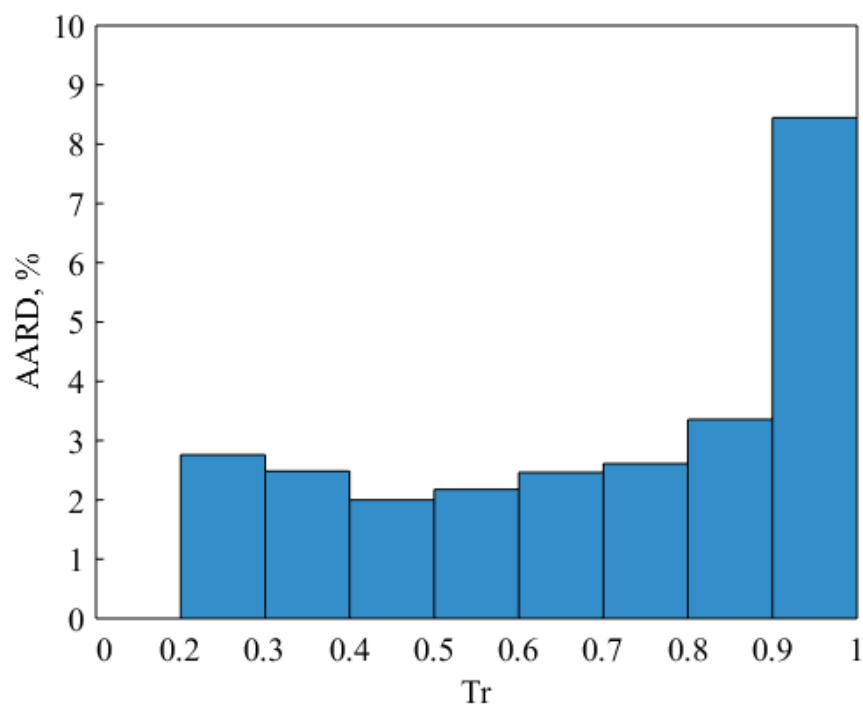


Figure 3.8 AARD for the C_p calculation in different T_r ranges using the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.

Table 3.4 Statistical performance metrics to calculate C_p for different chemical families using the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.

Chemical family	No. Data	Statistical metric				
		AARD, %	Min ARD_i , %	Max ARD_i , %	σ , %	R^2
Alkanes	3634	2.87	8.68E-04	44.27	3.24	0.99
Alkenes	341	1.32	1.00E-02	10.81	1.30	0.99
Alcohols	2419	2.33	2.89E-04	31.13	2.31	0.99
Polyols	399	1.85	1.00E-02	13.08	1.79	0.99
Cycloalkanes	494	3.08	9.43E-04	29.10	4.13	0.96
Carboxylic acids	144	2.39	1.10E-03	10.77	1.87	0.99
Ketones	479	1.30	1.10E-03	8.34	1.20	0.99
Ethers	386	3.35	1.14E-02	15.84	2.80	0.99
Esters	648	1.54	5.03E-04	14.65	1.93	0.99
Nitriles	182	2.30	3.00E-02	28.92	4.42	0.99
Amines/imines/amides	1678	1.72	8.83E-04	19.11	2.01	0.99
Compounds containing NO ₂	181	5.17	1.48E-05	17.68	3.97	0.99
Compounds containing Cl	934	2.74	1.90E-03	26.33	3.70	0.96
Compounds containing F	2649	3.72	6.55E-04	36.50	4.65	0.99
Compounds containing Br	497	1.68	6.67E-04	7.98	1.32	0.99
Compounds containing I	85	3.36	1.00E-02	18.78	4.62	0.99
Multihalogen compounds	1370	2.97	2.90E-03	25.12	3.38	0.96
Polyfunctional halides	466	4.84	6.20E-03	22.69	5.99	0.99
Sulfides/thiophenes	539	2.12	8.06E-04	9.41	2.06	0.99
Mercaptans	315	3.65	3.10E-03	24.73	6.26	0.99
Silanes/siloxanes	838	1.35	1.00E-02	10.18	1.58	0.99

Table 3.5 Statistical performance metrics to calculate C_p for the functional groups used by the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.

Functional group	No. Compounds	No. Data	Statistical metric				
			AARD, %	Min ARD _i , %	Max ARD _i , %	σ , %	R^2
-CH ₃ (without rings)	242	14059	2.51	1.48E-05	44.27	3.17	0.99
-CH ₂ - (without rings)	194	10230	2.43	2.89E-04	44.27	2.64	0.99
=CH- (with rings)	121	5346	1.97	1.48E-05	41.24	2.84	0.99
=C< (with rings)	111	4526	1.92	1.48E-05	27.86	2.53	0.99
-F	66	4279	3.89	6.55E-04	37.54	5.51	0.99
-Cl	59	2462	3.08	1.90E-03	26.33	4.11	0.97
>C< (without rings)	53	2486	3.67	7.71E-04	37.54	5.81	0.99
-OH (alcohol)	44	3346	2.38	2.89E-04	31.13	2.39	0.99
-CH ₂ - (with rings)	43	1986	1.93	8.06E-04	29.10	2.81	0.99
-O- (without rings)	38	1278	2.21	1.80E-03	15.51	2.19	0.99
>CH- (without rings)	32	1880	2.94	1.00E-02	36.08	4.06	0.98
-NH ₂ (without rings)	29	837	2.56	1.00E-02	23.10	2.76	0.99
=CH- (without rings)	24	513	1.90	9.09E-04	12.15	2.27	0.99
-O- (with rings)	23	470	2.08	1.14E-02	20.51	2.89	0.99
-Br	20	853	2.26	6.67E-04	9.99	2.03	0.99
-COO- (ester)	19	652	1.67	5.03E-04	15.51	2.03	0.99
-CN	19	241	2.83	3.00E-02	28.92	4.63	0.99
=CH ₂ (without rings)	18	458	1.48	1.00E-02	10.81	1.55	0.99
>C=O (without rings)	17	592	3.19	1.10E-03	22.68	5.46	0.99
-COOH (acid)	16	212	2.01	9.09E-04	10.77	1.75	0.99
>CH- (with rings)	16	661	1.58	8.06E-04	17.17	1.99	0.99
>Si< (without rings)	14	620	1.62	1.00E-02	10.18	1.74	0.99
-NO ₂	13	260	5.12	1.48E-05	26.07	4.58	0.98
>N- (without rings)	12	353	0.89	1.40E-03	10.31	0.92	0.99
-S- (with rings)	12	417	1.60	2.20E-03	9.17	1.57	0.99
-N= (without rings)	11	178	3.11	1.00E-02	12.15	2.85	0.99

Table 3.5 (continuation).

Functional group	No. Compounds	No. Data	Statistical metric				
			AARD, %	Min ARD _i , %	Max ARD _i , %	σ , %	R ²
-I	11	113	2.61	1.00E-02	18.78	4.21	0.99
-S- (without rings)	11	292	2.68	8.06E-04	9.41	2.32	0.99
>Si< (with rings)	11	218	0.69	1.14E-02	7.74	0.80	0.99
>N- (with rings)	10	346	1.49	7.60E-03	10.09	1.73	0.99
-N= (with rings)	10	330	1.79	1.77E-02	20.51	2.26	0.99
>NH (without rings)	9	164	3.16	6.00E-02	8.67	2.28	0.99
-SH	9	376	8.44	3.10E-03	37.54	12.31	0.98
=C< (without rings)	8	161	2.00	3.00E-02	8.80	2.04	0.99
=O	8	316	1.11	1.80E-03	10.31	1.17	0.99
>C=O (with rings)	8	217	1.06	7.60E-03	5.02	0.94	0.99
>NH (with rings)	7	237	0.87	8.83E-04	5.24	1.01	0.99
>C< (with rings)	7	774	2.14	6.55E-04	36.50	3.01	0.99
=C= (without rings)	6	168	1.42	1.00E-02	6.96	1.18	0.99
-CHO (aldehyde)	6	210	1.09	1.00E-02	9.69	1.49	0.99
-OH (phenol)	4	95	0.98	5.50E-03	5.67	0.82	0.99
=P≤	3	35	0.67	1.80E-03	10.31	1.92	0.99
-NO ₃	3	97	1.09	1.22E-02	8.40	1.41	0.98
≡C- (without rings)	1	31	0.13	1.00E-02	0.61	0.12	0.99
≡CH (without rings)	1	31	0.13	1.00E-02	0.61	0.12	0.99
=S<	1	148	0.74	9.10E-03	2.28	0.54	0.96
-SO ₂ - (without rings)	1	10	0.10	4.24E-04	0.21	0.07	0.97
>Al-	1	22	0.08	1.56E-02	0.26	0.07	0.99
-SO ₂ - (with rings)	1	68	1.05	1.12E-02	4.38	0.97	0.97

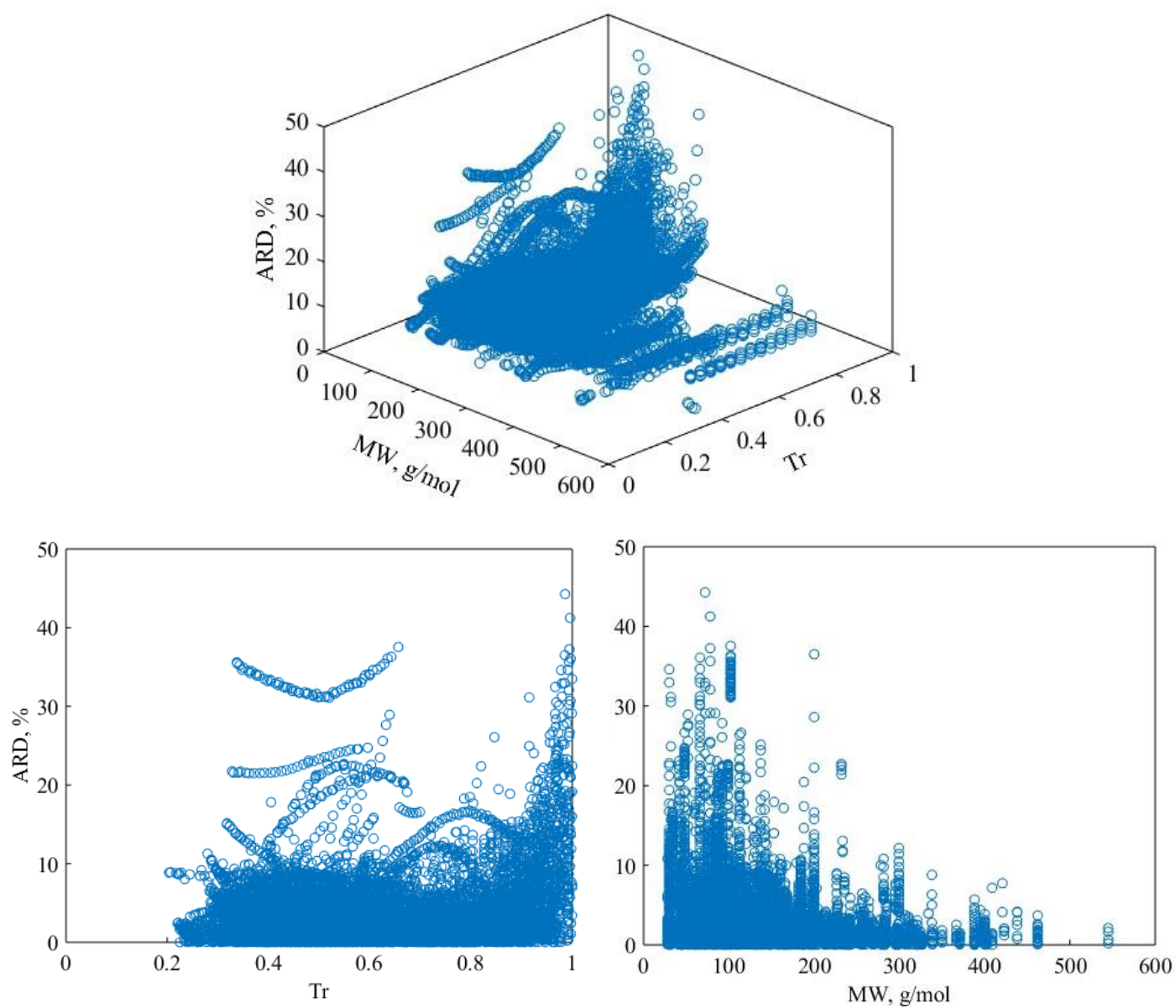


Figure 3.9 ARD_i values as a function of MW and Tr for the C_P calculation of pure compounds in liquid phase using the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.

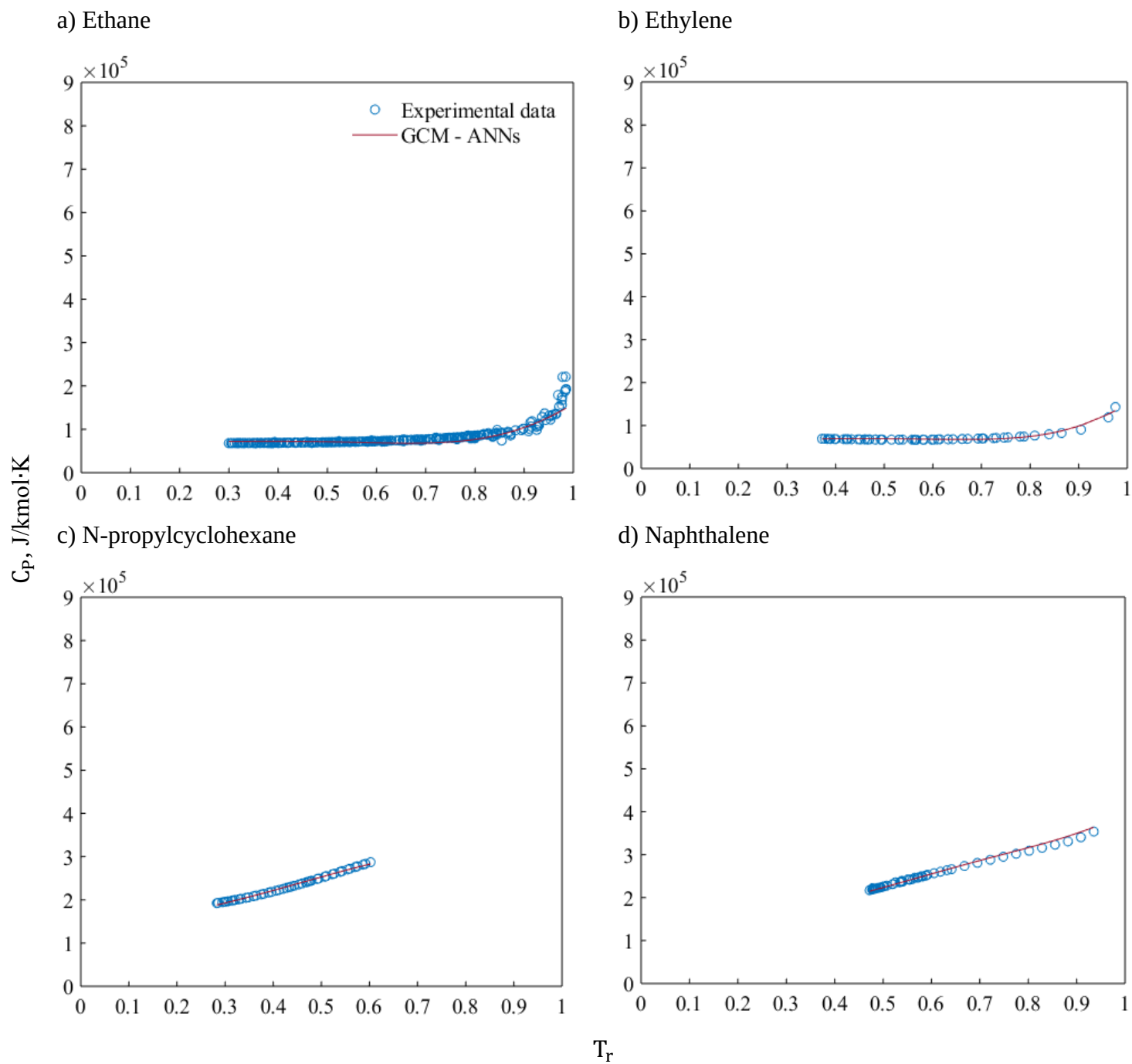
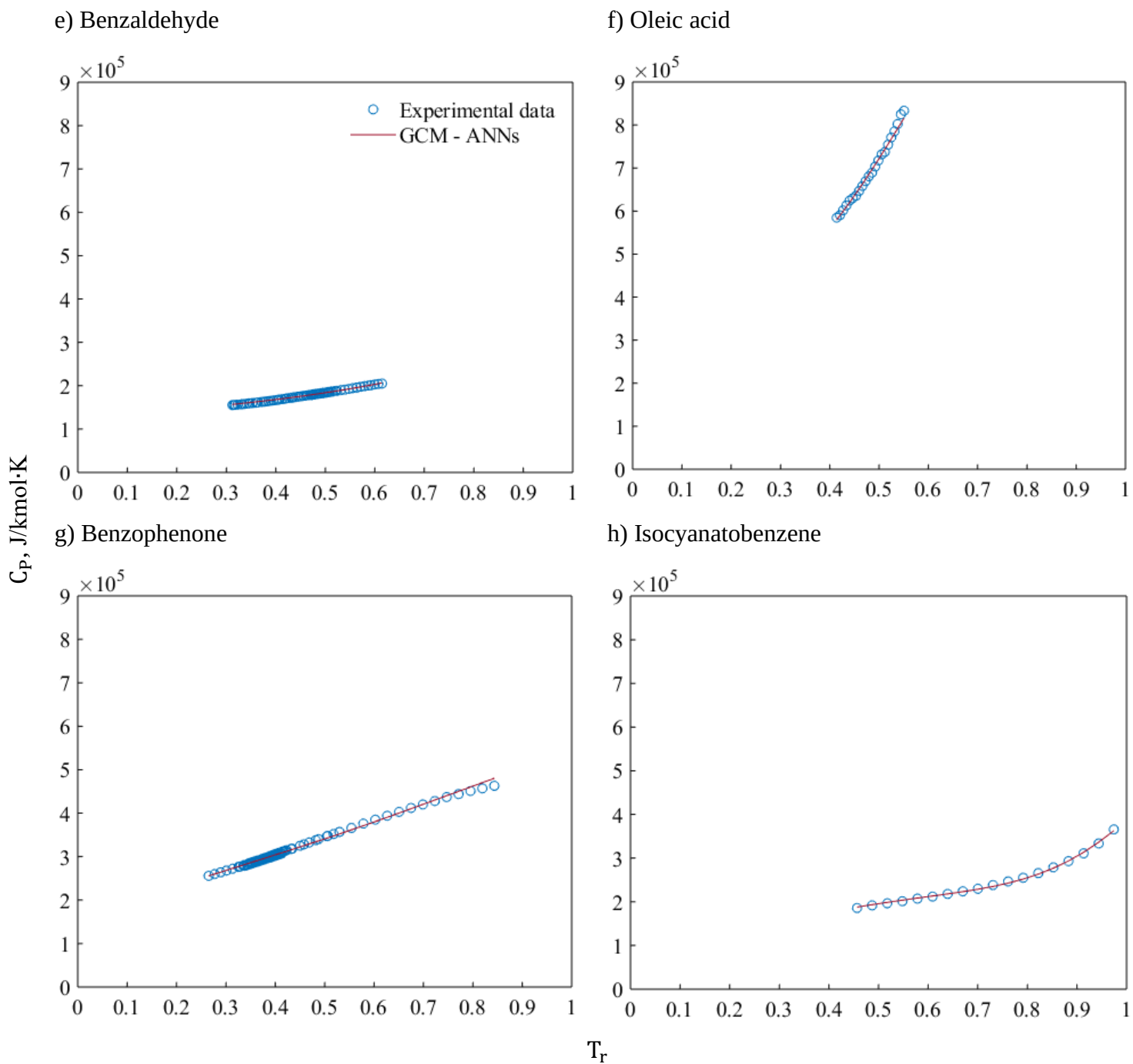


Figure 3.10 C_p calculation for different molecules using the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons.

**Figure 3.10** (continuation).

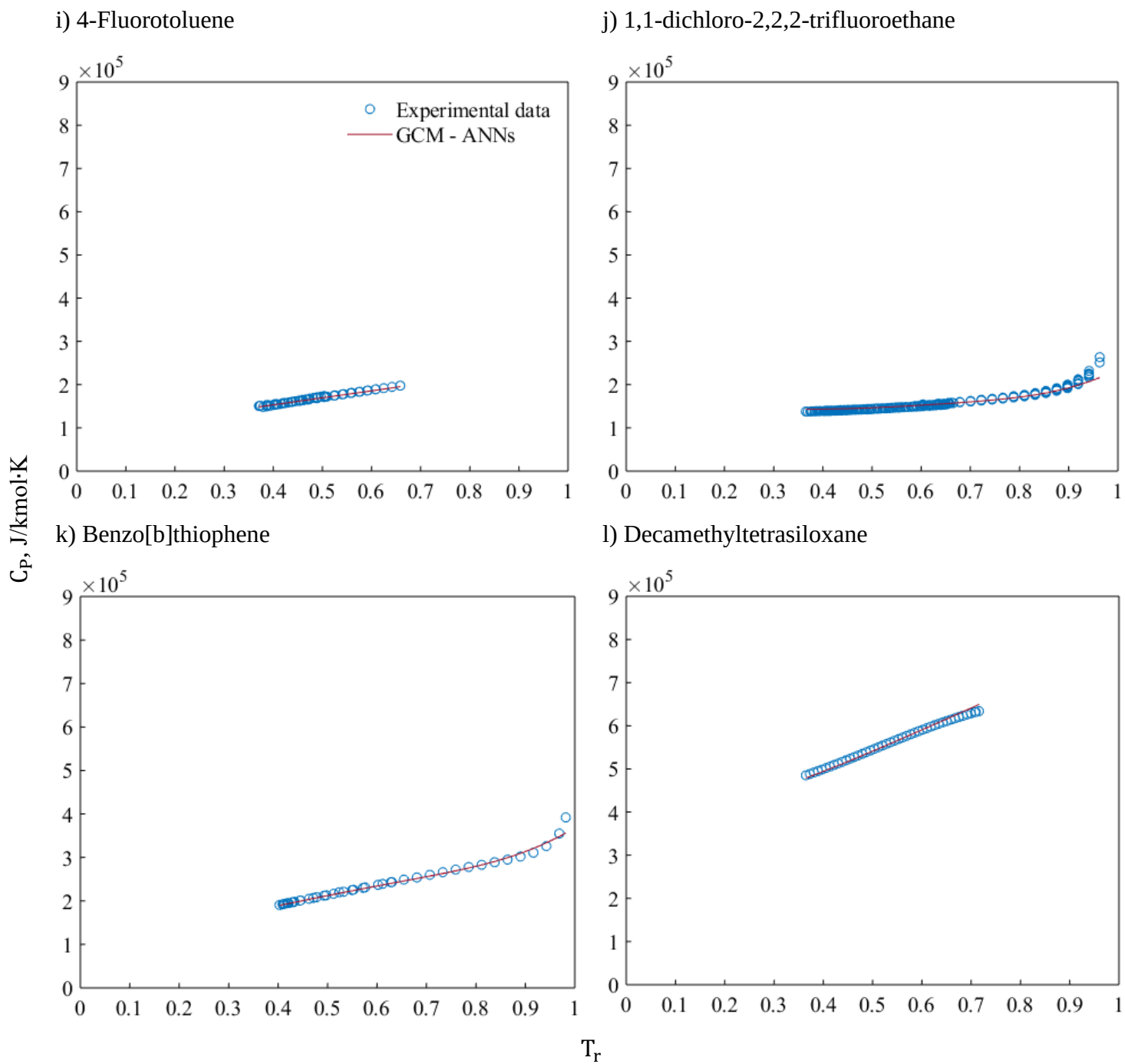


Figure 3.10 (continuation).

3.3.2 Comparative analysis of the best GCM–ANNs model with the GCM model used in the Aspen Plus® simulator

A performance metric analysis of the best M_2 GCM–ANNs model and the GCM implemented in the Aspen Plus® simulator (Růžicka Jr. and Domalski, 1993a, 1993b) is presented in Table 3.6. C_p calculations obtained with the best M_2 GCM–ANNs model were better than those obtained by the model used in this process simulator. AARD values in C_p calculation using the Aspen Plus® GCM and the M_2 GCM–ANNs models were 4.10 and 1.82%, respectively. The highest ARD values were 51.55% for ethane using the Aspen Plus® GCM model and 24.73% for methanethiol using the M_2 GCM–ANNs model. R^2 values were 0.98 and 0.99 for Aspen Plus® GCM and M_2 GCM–ANNs models, respectively. Figure 3.11 shows RD versus T_r and MW for the calculation of C_p using both models. Clearly, the model developed in this study outperformed the Aspen Plus® GCM. It should be noted that the benefits and better calculations of M_2 GCM–ANNs model were more clearly observed near to the critical point and for low-molecular-weight compounds. 78% of the data considered in this analysis showed ARD values lower than 5% using the Aspen Plus® GCM, while the M_2 GCM–ANNs model obtained ARD values lower than 5% for 94% of the data used in this comparison. The Aspen Plus® GCM calculated the C_p of 41 compounds with $ARD > 10\%$; in contrast, the M_2 model showed ARD values higher than 10% for only 3 compounds. Table 3.7 presents a comparison of the performance metrics of the best M_2 GCM–ANNs model and the GCM implemented in the Aspen Plus® simulator for different chemical families. M_2 GCM–ANNs model outperformed Aspen Plus® GCM for most of the chemical families included in this analysis. These results demonstrate the better calculation capabilities of this new GCM–ANNs model, which will help to perform reliable design and PSE calculations. Appendix E contains a Matlab® file (*Cp_liquid.m*) that can be used for the C_p calculation of pure compounds in liquid phase using the best GCMs–ANNs thermodynamic model. The weights and biases of the M_2 GCM–ANNs model were included in a Matlab® file (*Weights_and_biases.mat*) of Appendix E, which allows its standalone use.

Table 3.6 Comparison of the performance of the Aspen Plus® GCM and the M_2 GCM-ANNs model for the C_p calculation of pure compounds in liquid phase.

Model	Statistical performance metric				
	AARD, %	Min ARD_i , %	Max ARD_i , %	σ , %	R^2
Aspen Plus® GCM	4.10	3.38E-05	51.55	6.35	0.98
M_2 GCM-ANNs model	1.82	0.01	24.73	2.36	0.99

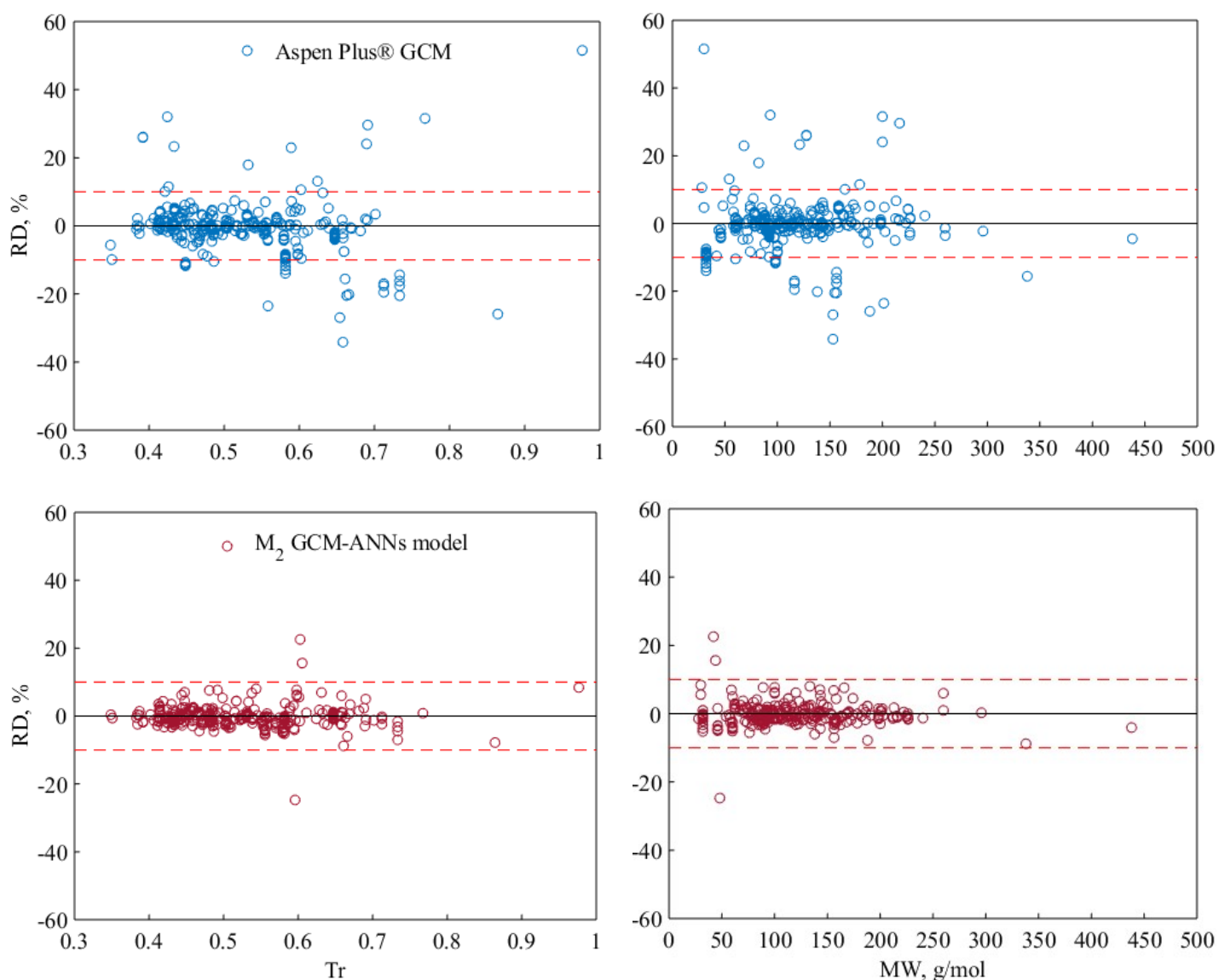


Figure 3.11 RD_i values as a function of T_r and MW for the C_p calculation of pure compounds in liquid phase using the Aspen Plus® GCM and M_2 GCM-ANNs models.

Table 3.7 Comparison of the performance of the Aspen Plus® GCM and the M_2 GCM-ANNs model for the C_p calculation of different chemical families in liquid phase.

Chemical family	Statistical metric									
	Aspen Plus® GCM					M_2 GCM-ANNs model				
	AARD , %	Min ARD _i , %	Max ARD _i , %	σ , %	R^2	AARD , %	Min ARD _i , %	Max ARD _i , %	σ , %	R^2
Alkanes	2.12	3.38E-05	51.55	7.06	0.97	1.27	5.64E-02	8.40	1.43	0.99
Alkenes	6.51	4.09E+00	13.11	3.10	0.99	0.67	5.29E-02	1.49	0.43	0.99
Alcohols	6.62	3.16E-01	19.51	4.33	0.98	2.31	3.44E-02	5.22	1.78	0.99
Polyols	2.14	2.47E-02	9.89	3.81	0.99	1.80	6.04E-01	2.95	0.89	0.99
Cycloalkanes	1.21	3.34E-02	9.57	2.04	0.99	2.24	3.76E-01	22.57	4.97	0.99
Carboxylic acids	1.06	2.86E-01	1.73	0.55	0.99	3.14	2.40E+00	4.36	0.80	0.99
Ketones	8.29	8.04E-01	11.74	4.37	0.78	1.08	2.09E-01	3.26	0.88	0.99
Ethers	1.99	2.01E-01	5.18	1.56	0.99	3.80	1.16E+00	15.60	4.82	0.99
Esters	2.33	1.59E-01	4.41	1.28	0.99	2.09	2.06E-01	4.50	1.13	0.99
Nitriles	0.43	2.00E-03	0.78	0.30	0.99	1.62	3.55E-01	4.12	1.29	0.75
Amines/imines /amides	5.88	5.94E-02	32.03	8.29	0.92	1.73	1.94E-01	6.94	1.66	0.99
Compounds containing NO ₂	1.59	8.99E-01	2.28	0.97	1.00	5.62	3.58E+00	7.67	2.89	1.00
Compounds containing Cl	1.94	2.28E-02	4.32	1.29	0.99	1.83	9.40E-03	8.02	2.21	0.99
Compounds containing F	12.36	1.91E-01	31.57	12.72	0.83	3.02	2.44E-01	8.80	2.83	0.99
Compounds containing Br	2.35	3.23E-02	5.20	1.99	0.99	1.45	6.37E-01	3.08	0.87	0.99

Table 3.7 (continuation).

Chemical family	Statistical metric									
	Aspen Plus® GCM					M ₂ GCM-ANNs model				
	AARD , %	Min ARD _i , %	Max ARD _i , %	σ , %	R ²	AARD , %	Min ARD _i , %	Max ARD _i , %	σ , %	R ²
Compounds containing I	2.72	1.36E-01	5.01	1.59	0.99	1.76	1.99E-02	6.79	2.53	0.99
Multihalogen compounds	13.08	9.09E-02	34.15	13.10	0.55	1.65	8.31E-02	5.98	2.29	0.98
Polyfunctional halides	12.00	1.31E+00	26.10	9.89	0.35	3.90	1.63E+00	7.69	2.51	0.93
Sulfides /thiophenes	6.40	1.35E-01	29.62	11.55	0.79	1.53	3.88E-01	2.60	0.87	0.99
Mercaptans	1.36	4.29E-02	5.23	1.80	0.99	5.17	1.73E-01	24.73	8.81	0.99

3.4 Conclusions

A versatile and reliable thermodynamic model using GCMs and ANNs was developed for the calculation of liquid C_p of pure compounds. Different input variables and artificial neural network configurations were evaluated to identify the best model structure to estimate this thermodynamic property without overfitting. The model that included T_r as input parameter allowed to improve the C_p estimations obtained near to the critical point, and it offered the best performance with low modeling errors. It was found that a symmetric configuration for the distribution of hidden neurons in the ANN layers favored the model performance and helped to reduce the overfitting. This new thermodynamic model showed C_p modeling errors ranging from 0.47 to 7.72% for most of the chemical families tested in this study. This model achieved the highest C_p calculation errors for low-molecular-weight compounds and near the critical point. However, it outperformed the GCM of the Aspen Plus® simulator. Overall, the accuracy and scope of application of this new thermodynamic model are better than the GCM available in the Aspen Plus® process simulator for the C_p estimation of pure compounds in liquid phase. Consequently, this new model is a better option to improve the PSE calculations of different units and equipment.

Chapter 4. Assessment of machine learning algorithms to estimate the triple-point temperature of pure compounds

4.1 Introduction

Machine learning (ML) comprises several computer algorithms that can learn from experience and make decisions without human intervention (Ding et al., 2021; Pence et al., 2023). Because of their capacity to describe complex relationships between variables, they have been employed in numerous real-life applications, including face recognition (Archana et al., 2022), self-driving vehicles (Chopra and Roy, 2020), and image geolocalization (Gu et al., 2022), among others (Sharma et al., 2020; Gao and Wu, 2021; Field et al., 2021; Alanne and Sierla, 2022). Over the past few years, there has been increased interest in the application of ML algorithms to chemical engineering solutions (Kamari et al., 2014; Schmidt et al., 2017; Zendehboudi et al., 2018; Dobbelaere et al., 2021; Trinh et al., 2021; Nakhaei-Kohani et al., 2022). Specifically, ML algorithms have become powerful tools for predicting the thermodynamic properties of pure substances and blends (Mann et al., 2022). Jirasek and Hasse (2023) reported an increase in the quantity of research papers published over time related to fluid thermodynamic modeling using ML algorithms and hybrid models (i.e., models combining ML algorithms with other thermodynamic models).

ML algorithms have notable advantages over traditional thermodynamic models in predicting pure compounds and mixtures properties (Nakhaei-Kohani et al. 2022). These algorithms can provide fast estimations and can handle large databases. The capability of these models to describe nonlinear relationships between variables makes them suitable for predicting complex thermodynamic properties (Ding et al., 2021). These models can discover relationships and patterns between variables that are not as clear via traditional models, allowing accurate estimations of the property under analysis (Soltanian et al., 2024). These algorithms can also handle missing and limited data through different techniques, which allows for a more complete evaluation of the property (Soltanian et al., 2024).

The potential of these models has been exploited for forecasting the thermodynamic properties of various chemicals and advanced materials, such as deep eutectic solvents (Atashrouz et al., 2017; Wang et al., 2021), refrigerants (Chouai et al., 2002; Arcaklıoğlu et al., 2004; Şahin et al., 2012; Mohanraj et al., 2012; Longo et al., 2020; Ahmed et al., 2021;

Noushabadi et al., 2023; Pence et al., 2023), perovskite oxides (Vidyasagar et al., 2022), petroleum derivatives (Jameel et al., 2018), diesel–biodiesel blends (de Oliveira et al., 2017), and others (Zhang et al., 2018; Soltanian et al., 2024). In particular, ML algorithms are considered to be one of the main tools for predicting ILs properties (Shahriari et al., 2018; Mousavi et al., 2021a, b, c; Nakhaei-Kohani et al., 2022). The specific thermodynamic properties of pure substances and blends have been estimated using these algorithms (Ding et al., 2021; Alshehri et al., 2022). These properties include the normal boiling point (Mann et al., 2022), critical properties (Varamesh et al., 2017; Mann et al., 2022), enthalpy of formation (Yalamanchi et al., 2019), density and viscosity (Saldana et al., 2012), melting temperature (Seko et al., 2014), heat capacity (Kauwe et al., 2018), activity coefficients (Behrooz and Boozarjomehry, 2017; Jirasek et al., 2020), aqueous solubility (Palmer et al., 2007), thermodynamic stability of solids (Schmidt et al., 2017), heat of formation (Sun et al., 2014), and various thermal and transport properties (Qian and Yang, 2021).

The knowledge of the thermodynamic properties of pure compounds and mixtures is paramount in the design, modeling and optimization of chemical processes and plants (Mann et al., 2022). One of the most relevant thermodynamic properties of pure compounds is the triple-point temperature (T_p). This property is described as the temperature at which three phases (i.e., liquid, solid and vapor) coexist in equilibrium (Gharagheizi and Sattari, 2010). T_p is a basic property for phase equilibria calculations (e.g., it is essential for the determination of enthalpy of sublimation, since below the triple point it is possible to vaporize the solid without melting). This property is also used for the estimation of other thermodynamic properties, for example the solubility of solids in liquids, enthalpy of fusion, and vapor solid pressure (Gharagheizi and Sattari, 2010).

Despite the relevance of T_p for the effective processes and products design, the availability of experimental data on this property is a constraint. Due to the measurements of T_p are, in many cases, expensive and time consuming, the experimental information published in the literature is scarce and limited for specific chemical families and compounds widely used in industry (Gharagheizi and Sattari, 2010). Furthermore, the number of synthesized compounds increases exponentially, and it is impractical to carry out time-consuming and costly experiments for all of them (Mann et al. 2022). An alternative to address this limitation is to develop thermodynamic and empirical models to estimate this property with acceptable

accuracy. However, only a limited number of studies have focused on this topic. Gharagheizi and Sattari (2010) developed a model based on QSPR for estimating the T_p of pure compounds. The database employed for the development of the model was limited, consisting of 638 pure compounds, and the modeling errors reached 10% absolute percent error.

Although ML algorithms are a great alternative to solve the problem of thermodynamic property estimation, no such model has been reported to predict the T_p of pure compounds. This chapter reports an assessment and comparative analysis of different ML algorithms for estimating this thermodynamic property. The ML algorithms with the greatest potential to estimate T_p were identified using an AutoML approach available in the auto-sklearn library in Python. In addition, different input variables for the ML algorithms were evaluated, including the functional groups that conform to the molecules as the main input, to identify the alternative that displayed the best performance. The limitations and capabilities of the implemented ML approaches were assessed in terms of modeling errors for different chemical families and functional groups. This study allowed to identify the best ML algorithms alternatives for the estimation of T_p of pure compounds and highlights certain gaps to be considered for the estimation of thermodynamic properties using these types of models.

4.2 Approach for the comparison and evaluation performance of models based on machine learning algorithms to predict the triple-point temperature of pure compounds

4.2.1 Experimental database

The experimental database used for the model development and assessment consisted of 3701 T_p data points from 1877 pure compounds ranging from 85.46 to 703.10 K. This information was collected from 1999 references of the National Institute of Standards and Technology (NIST). The molecular weight (MW, g/mol), critical pressure (P_c , Pa) and critical temperature (T_c , K) of the compounds included in the database ranged from 28.05 to 544.90 g/mol, 227.53 to 1981 K and 459314.90 to 10258859.10 Pa, respectively. An initial assessment of the experimental data on T_p was carried out to identify and discard inconsistent values reported by several sources for the identical compound. Furthermore, isomer compounds were not included. The experimental database on T_p contained a wide range of

chemical families including hydrocarbons, aromatic compounds, carboxylic acids, alcohols, aldehydes, esters, ethers, ketones, compounds containing nitrogen, halogens, silicon, sulfur, aluminum and phosphorus, among others. The molecules of the database were segmented into 60 functional groups outside and inside the rings, see Table 4.1.

4.2.2 Description of the approach based on machine learning algorithms to estimate the triple-point temperature of pure compounds

The development of models based on ML algorithms for predicting the T_p of pure compounds requires the evaluation of an ensemble of ML algorithms to determine the one with the most outstanding performance for this task. However, the hyperparameters optimization of each evaluated algorithm and the selection of the optimal one for the T_p estimation is a time-consuming process and heavily relies on human expertise. AutoML has emerged as an innovative solution that enables the automated and rigorous search of several ML algorithms and their hyperparameters (Sinha et al., 2021). This reduces the time and effort required and increases the reliability and robustness of data processing (Conrad et al., 2022). In this research, the AutoML approach employing auto-sklearn library in Python was implemented to assess different ML algorithms ensembles and identify the one with the greatest potential for the T_p estimation.

Furthermore, several input variables to the ML algorithm were evaluated to identify those with the highest impact on T_p estimation. For ease of reference, four models distinguished by a set of different input variables were defined. The frequency and type of the functional groups (n_i) that conform the molecules in the database were included as inputs for all models. Additionally, the variables MW; T_C ; MW and T_C ; and MW, T_C and P_C were selected as input variables of the models M_1 , M_2 , M_3 and M_4 , respectively.

Table 4.1 Functional groups defined for the development and evaluation of models based on ML algorithms to predict the T_p of pure compounds.

Groups outside rings				
-CH ₃	-CH ₂ -	>CH-	>C<	=CH ₂
=CH-	=C<	=C=	≡C-	≡CH
Groups inside rings				
-CH ₂ -	>CH-	=CH-	>C<	=C<
≡C-				
Groups with oxygen				
-OH (alcohol)	-OH (phenol)	-O- (outside rings)	-O- (inside rings)	>C=O (outside rings)
>C=O (inside rings)	-CHO (aldehyde)	-COOH (acid)	-COO- (ester)	=O
Groups with nitrogen				
>NH (outside rings)	>NH (inside rings)	>N- (outside rings)	>N- (inside rings)	-N= (outside rings)
-N= (inside rings)	-NO ₂ (outside rings)	-NO ₂ (inside rings)	>N=O (inside rings)	-NH ₂
-CN	=NH	=N	=N=	-NO ₃
Groups with sulfur				
-S- (outside rings)	-S- (inside rings)	-SO ₂ - (outside rings)	-SO ₂ - (inside rings)	-SO ₄ - (outside rings)
-SO ₄ - (inside rings)	>S=O (outside rings)	>S=O (inside rings)	-SO ₃ -	-SH
=S				
Groups with halogen				
-F	-Cl	-Br		-I
Groups with silicon				
>Si< (outside rings)			>Si< (inside rings)	
Groups with phosphorus				
=P<				
Groups with aluminum				
>Al-				

For model development, the experimental database was randomly separated into a training subset (with 80% of the compounds included in the database) and a testing subset (with the

remaining 20% of compounds). Note that the molecules contained in the test subset were not used for training the model (i.e., 1502 molecules were used for training, and 375 different molecules were used for testing). The ML algorithms were trained by applying a k-fold cross-validation ($k = 5$). The cross-validation resampling strategy uses several random training and testing subsets to evaluate the ML algorithm performance in relation to sampling variation (Zhang et al., 1999). Therefore, the database was randomly divided into five equal subsets or folds (represented by parameter k), and the cross-validation process was performed five times. Each time, a different subset was selected to test the model and the other four subsets were used for training. Thus, all experimental data were used for testing, and most of them were employed in the training process (Zhang et al., 1999).

After identifying the model based on ML algorithms with the highest performance (i.e., the input variables and the ML algorithm and their hyperparameters that displayed the lowest deviations), 30 additional random divisions of the database in subsets of training and testing were performed. The best model was then retrained for all new database splits, and the best results were reported in this chapter. The calculations were carried out using a program written in the Python programming language.

The ML algorithms evaluated for the T_p estimation of pure compounds using Auto-sklearn included Liblinear Support Vector Regression, Adaptive Boosting, Random Forest, Gradient Boosting, Extremely Randomized Trees, K-Nearest Neighbors, Multi-Layer Perceptron, Libsvm Support Vector Regression, Decision Tree and Gaussian Process. The next subsection provides a brief description of ML algorithms.

4.2.3 Description of machine learning algorithms

ANNs are ML algorithms that emulate the behavior of the biological nervous system (Fatehi et al., 2014). In contrast to other sophisticated mathematical and computational approaches, ANNs can establish connections between inputs and outputs without the need for elaborate procedures (Soltanian et al., 2024). The Multi-Layer Perceptron (MLP) is a type of ANNs consisting of three layers (i.e., hidden, input and output). Each layer consists of interconnected processing units called neurons, each characterized by two properties: weights and biases (Fatehi et al., 2014). The nonlinear behavior of the input variables is described by an activation function (Soltanian et al., 2024).

K-Nearest Neighbors is a versatile ML algorithm capable of adapting to diverse datasets. This feature makes this algorithm suitable for scenarios in which the relationships among variables are not established or where the data can exhibit localized patterns (Soltanian et al., 2024). This algorithm assumes that similar data points are located near each other (Vidyasagar et al., 2022). Under this assumption, new values are predicted depending on the level of proximity they resemble the data in the training set (Vidyasagar et al., 2022).

Adaptive Boosting initially trains a weak algorithm on the original dataset, and subsequently constructs additional weak algorithms by adjusting the data distribution for each subsequent training step (Soltanian et al., 2024). In the next step, priority is given to the training data points with the lowest accuracy. At last, the weak algorithms are grouped considering different weights to create a strong algorithm (Soltanian et al., 2024). For further details on this algorithm, the reader can refer to Schmidt et al. (2017).

Support Vector Regression (SVR) is a ML algorithm that extends the Support Vector Machine (SVM) classification algorithm (Yalamanchi et al., 2019). It manages both model complexity and prediction error showing good performance in handling high-dimensional datasets (Zhang and O'Donnell, 2020). This algorithm follows the fundamental concept of SVM, which is a sparse kernel machine conducting classification by utilizing a hyperplane defined by a small number of support vectors (Zhang and O'Donnell, 2020). However, in SVR, the support vectors are represented by a small number of samples from the training data (Zhang and O'Donnell, 2020). The solution will depend on the number of such support vectors, not on the dimension of the input data (Zhang and O'Donnell, 2020).

The Decision Tree algorithm divides intricate decisions or problems into simpler and easily understandable components (Soltanian et al., 2024). In this manner, the data is segmented into smaller subsets to construct a tree (Vidyasagar et al., 2022). In this algorithm, the data features are known as predictor variables, while the class to be mapped represents the target variable (Ahmad et al., 2018). In regression problems, these target variables are continuous (Ahmad et al., 2018). The data points are organized systematically, offering valuable insights into the relationships between different continuous target values and input variables (Soltanian et al., 2024).

Random Forest is an ensemble learning technique that integrates numerous decision trees (Soltanian et al. 2024). This algorithm trains different decision trees on several subsets of a

complete dataset. Although a single decision tree may be susceptible to overfitting, the employ of multiple trees in random forests mitigates variance and enhances model accuracy (Soltanian et al., 2024).

Extra Trees or Extremely Randomized Trees represent a variant of the Random Forest algorithm that is less susceptible to overfitting a dataset (Geurts et al., 2006). It follows the same principle as Random Forest and employs a random subset of characteristics to train each base estimator (John et al., 2016). In contrast to the Random Forest, this algorithm does not utilize the tree bagging process to create the training subset for each tree (John et al., 2016). It uses the global training data to train each regression tree (Ahmad et al., 2018). Extra Trees randomly choose the optimal feature along with its corresponding value to divide the node (John et al., 2016).

Gradient Boosting (GB) is a boost-like algorithm that can be employed in regression problems (Pence et al., 2023). These algorithms combine weak learners to enhance their overall predictive capability (Pence et al., 2023). GB iteratively constructs a regression tree based on the negative gradient of a specified loss function (Pence et al., 2023). The process described enables the creation of a predictive model by combining multiple decision trees (Pence et al., 2023).

The Gaussian Process represents a generalization of the Gaussian probability distribution (Bélisle et al., 2015). Gaussian Process regression calculates the posterior distribution using either the training data or the prior distribution (Bélisle et al., 2015). One of its main advantages is its non-parametric characteristics, making it versatile for diverse situations, particularly for problems involving high-dimensional spaces (Bélisle et al., 2015).

4.2.4 Performance evaluation metrics

For the assessment of models' performance, a variety of statistical metrics were used, including the relative deviation (RD_i), the absolute relative deviation (ARD_i), the average absolute relative deviation (AARD), the determination coefficient (R^2) and the standard deviation (σ), which are determined by

$$RD_i = 100 \cdot \frac{T_{P_i}^{exp} - T_{P_i}^{cal}}{T_{P_i}^{exp}} \quad (4.1)$$

$$ARD_i = 100 \cdot \left| \frac{T_{P_i}^{\text{exp}} - T_{P_i}^{\text{cal}}}{T_{P_i}^{\text{exp}}} \right| \quad (4.2)$$

$$AARD = \frac{1}{m} \sum_{i=1}^m 100 \cdot \left| \frac{T_{P_i}^{\text{exp}} - T_{P_i}^{\text{cal}}}{T_{P_i}^{\text{exp}}} \right| \quad (4.3)$$

$$R^2 = \left(\frac{\sum_{i=1}^m (T_{P_i}^{\text{exp}} - \overline{T_{P_i}^{\text{exp}}})(T_{P_i}^{\text{cal}} - \overline{T_{P_i}^{\text{cal}}})}{\sqrt{\sum_{i=1}^m (T_{P_i}^{\text{exp}} - \overline{T_{P_i}^{\text{exp}}})^2 \sum_{i=1}^m (T_{P_i}^{\text{cal}} - \overline{T_{P_i}^{\text{cal}}})^2}} \right)^2 \quad (4.4)$$

$$\overline{T_{P_i}^{\text{exp}}} = \frac{\sum_{i=1}^m T_{P_i}^{\text{exp}}}{m} \quad (4.5)$$

$$\overline{T_{P_i}^{\text{cal}}} = \frac{\sum_{i=1}^m T_{P_i}^{\text{cal}}}{m} \quad (4.6)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^m (ARD_i - AARD)^2}{m}} \quad (4.7)$$

where $T_{P_i}^{\text{exp}}$ and $T_{P_i}^{\text{cal}}$ are the values of experimental and calculated T_p and m is the number of experimental data analyzed.

4.3 Results and discussion

4.3.1 Comparative analysis of models' performance

The AutoML approach employed in this research allowed to identify the GB as the ML algorithm with the greatest potential for estimating the T_p of pure compounds among the various ensembles of ML algorithms evaluated. These results were consistent for the M_1 , M_2 and M_3 models, while in the case of the M_4 model, a MLP algorithm demonstrated the best performance. Figure 4.1 illustrates the comparison of experimental and calculated T_p for the M_1 , M_2 , M_3 and M_4 models. The results indicated that all models were able to correlate the T_p of pure compounds with an acceptable accuracy displaying R^2 values ranging from 0.96 (for the M_4 model) to 0.99 (for the M_1 model). However, the testing subset exhibited higher modeling errors in predicting the T_p , with R^2 values ranging from 0.75 (for the M_1 and M_4 models) to 0.77 (for the M_3 model). These results suggested that all models might fail to accurately predict the T_p of pure compounds.

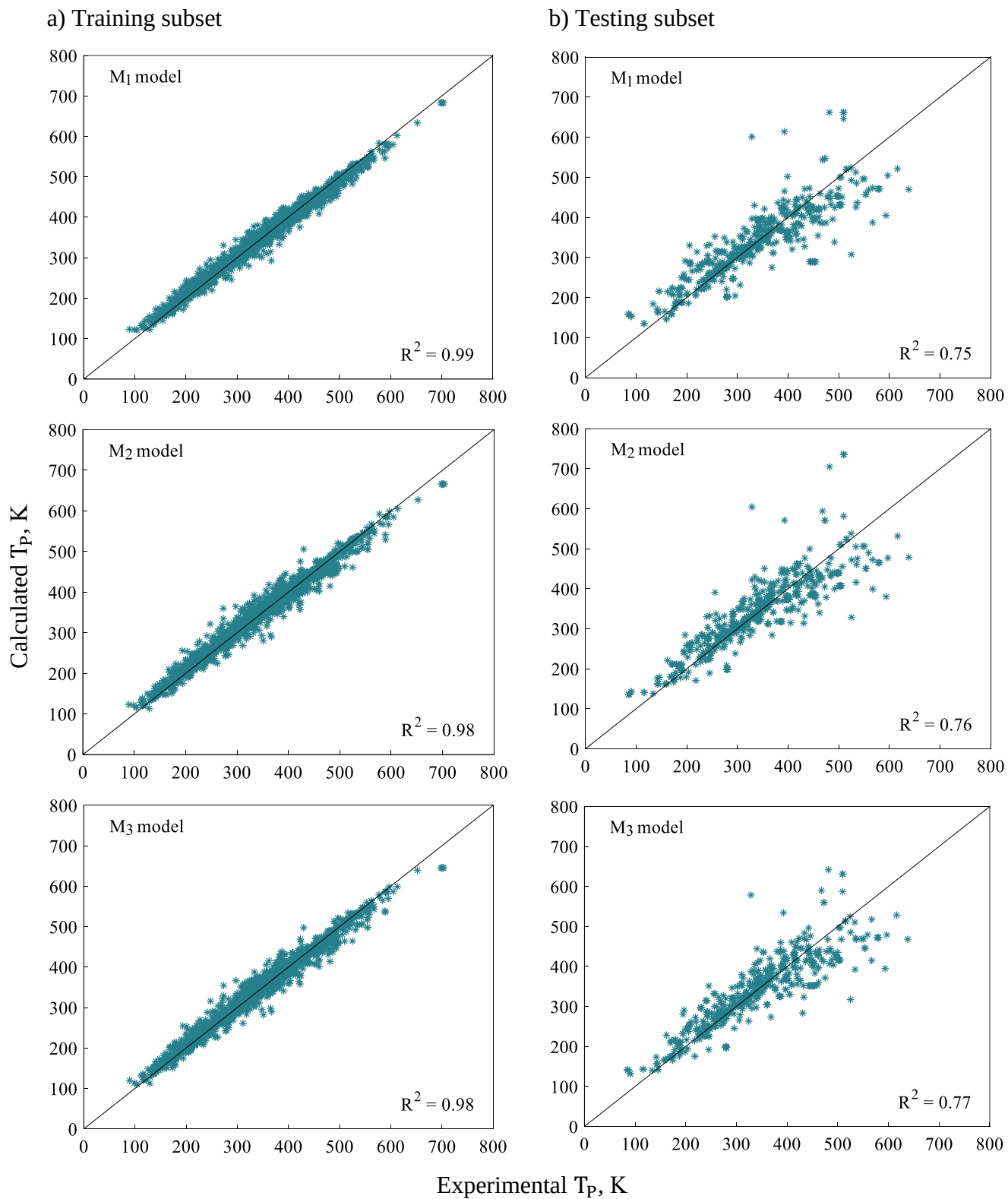


Figure 4.1 Results of the calculation of T_p of pure compounds using different models based on ML algorithms.

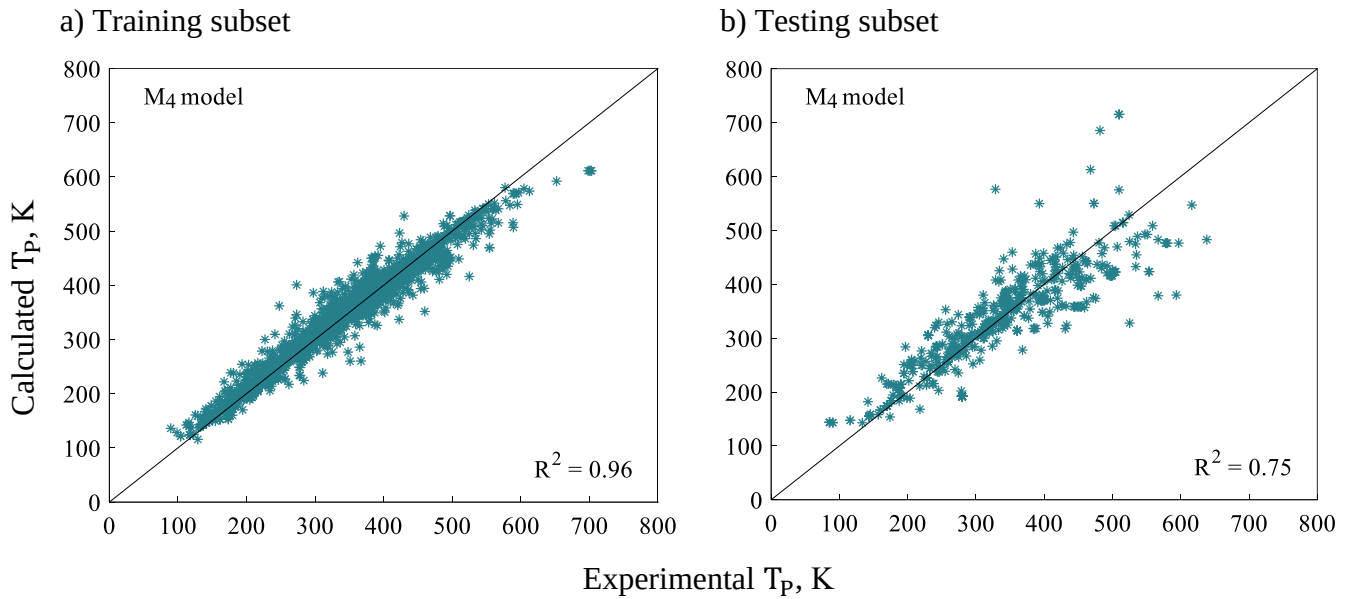


Figure 4.1 (continuation).

The histograms of RD_i values obtained from the evaluated models based on ML algorithms are provided in Figure 4.2. For the M_1 model, modeling errors of ± 5 and $\pm 10\%$ were observed for the 86 and 97% of the data used for training, respectively. For the testing subset, modeling errors of ± 5 and $\pm 10\%$ were displayed for 45 and 66% of the data, respectively. The M_2 and M_3 models demonstrated comparable performances, achieving RD_i values of $\pm 10\%$ for 95% of the training data. For the M_2 and M_3 models, 80 and 81% of the training data exhibited modeling errors of $\pm 5\%$, respectively. RD_i values of $\pm 10\%$ were observed in 62 and 66% of the testing data for the M_2 and M_3 models, respectively. Modeling errors of $\pm 5\%$ were achieved in the testing subset for 45 and 48% of the data, corresponding to the M_2 and M_3 models, respectively. The model with the lowest modeling performance corresponded to M_4 , exhibiting RD_i values of $\pm 10\%$ for 91 and 62% of the training and testing data, respectively. 74% of the training data and 47% of the testing data showed RD_i values of $\pm 5\%$. The performance trend of the models is as follows: $M_1 > M_3 > M_2 > M_4$.

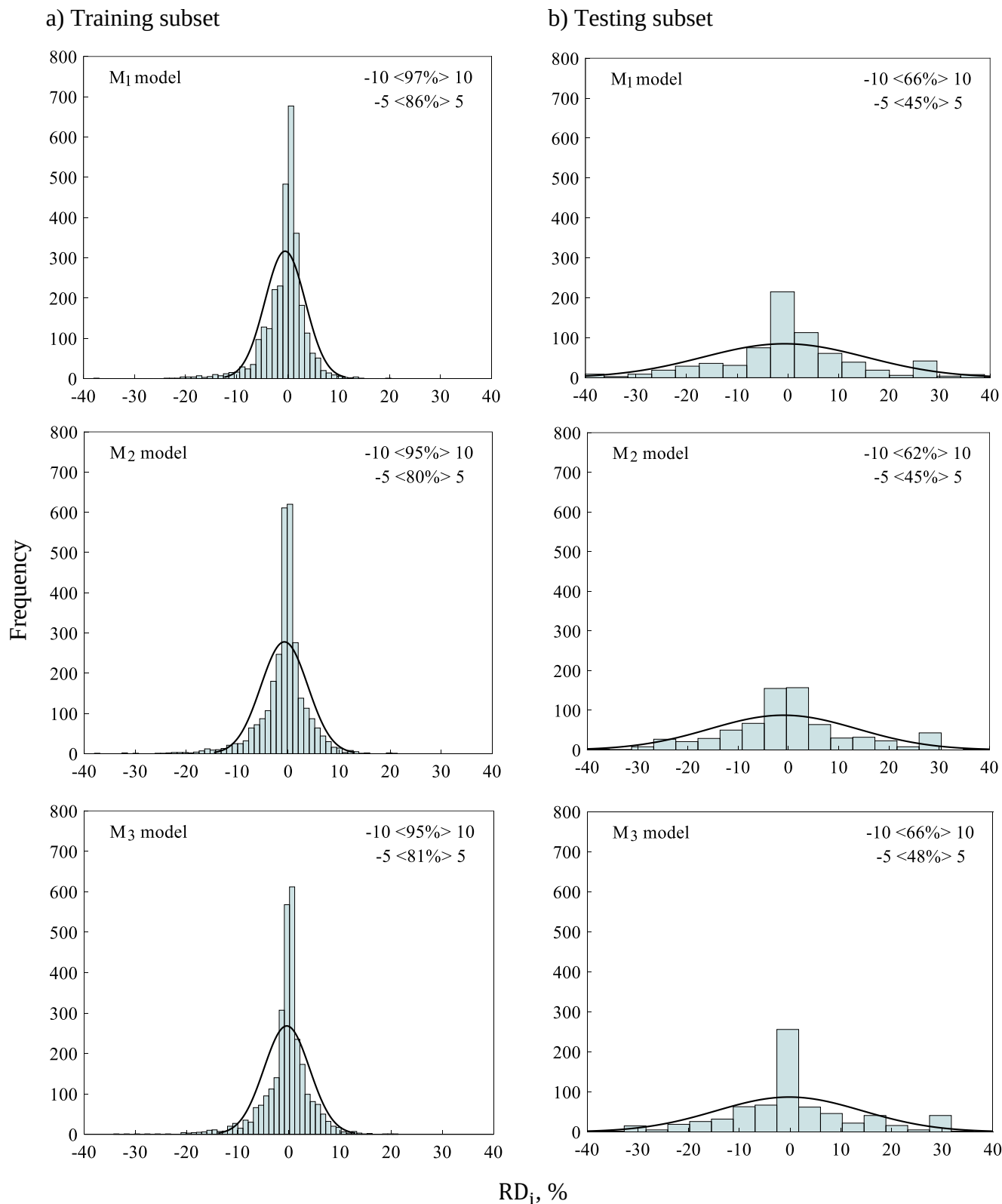


Figure 4.2 Histograms of RD_i values obtained with different models based on ML algorithms for the calculation of T_p of pure compounds.

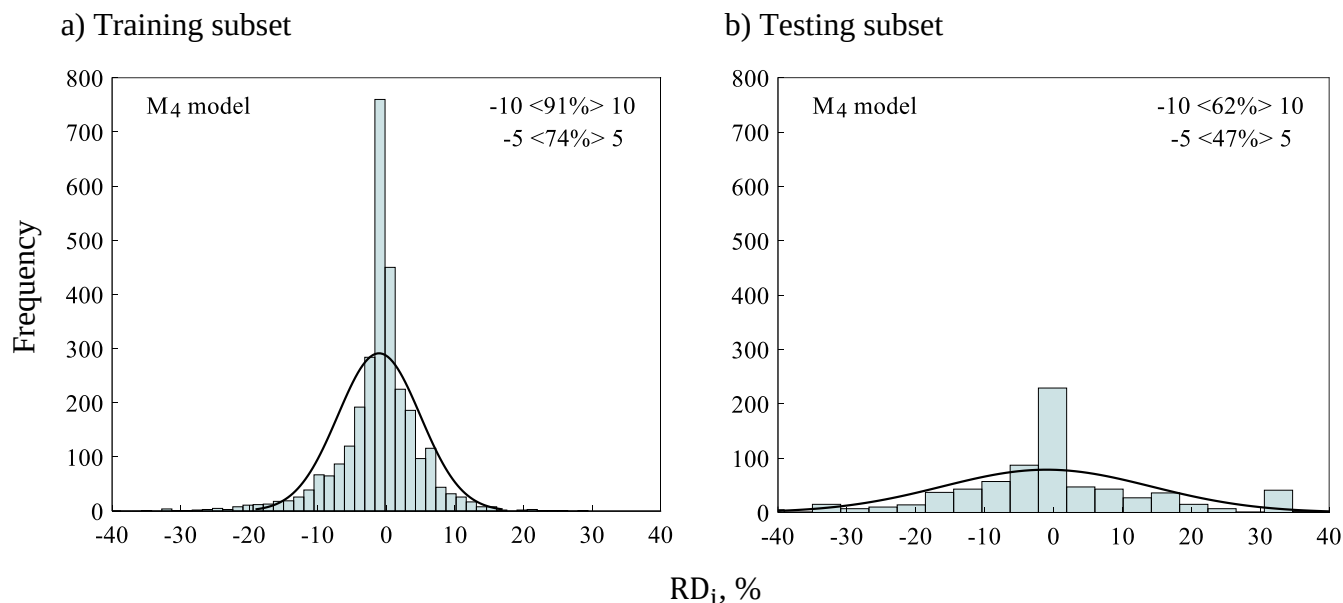


Figure 4.2 (continuation).

These results indicate that the M_1 model, which included the MW and n_i as input variables of the ML algorithm achieved the best performance for estimating the T_p of pure compounds. Therefore, MW, included as an input variable in those models with higher performance (i.e., M_1 and M_3 models) was identified as the variable with the greatest impact on the prediction of T_p . In contrast, the inclusion of T_C and P_C did not result in a positive impact on the models' performance.

Figure 4.3 shows the performance of the model in relation to the MW and T_C for all evaluated models. It was not possible to identify a specific trend between T_C and RD_i values. Although the trend between MW and RD_i is not entirely clear, it is notable that modeling errors were higher for low MW compounds (i.e., $MW < 200$ g/mol). To support this statement, Figure 4.4 displays the AARD values obtained in T_p calculation across different ranges of MW. For all models, the AARD values obtained for low MW compounds (i.e., $MW < 200$ g/mol) are nearly double those obtained for high MW compounds (i.e., $MW > 200$ g/mol). For the M_1 , M_2 , M_3 and M_4 models, the 84, 82, 83 and 78% of high MW compounds exhibited RD_i values of $\pm 5\%$, respectively. For compounds with MW lower than 200 g/mol, 71, 65, 67 and 59% of the data showed modeling errors within $\pm 5\%$ for the M_1 , M_2 , M_3 and M_4 models,

respectively. This trend of low MW compounds exhibiting high modeling errors when calculating their thermodynamic properties was observed by Nannoolal et al. (2007). These types of molecules typically deviate from the homologous series trend within a chemical family. As a result, modeling errors in predicting their properties using thermodynamic models could be higher. However, thermodynamic properties of low MW compounds are usually available in literature, and it is not necessary to estimate them using predictive models. The modeling errors in terms of T_p for the evaluated models are depicted in Figure 4.5. No specific trend could be detected between T_p and RD_i values. From Figures 4.3 and 4.5, it can be inferred that the results were not influenced by specific conditions of T_c and T_p ; however, the models' performances varied across different ranges of MW. The information contained in Figures 4.3, 4.4 and 4.5 corresponds to the data used for training and testing.

The model with the best performance identified in this research (i.e., the M_1 model with the GB algorithm) was retrained for 30 additional random divisions of the database into subsets of training and testing. The modeling errors obtained from those new random divisions were assessed. Table 4.2 presents the performance statistical metrics obtained using the subsets that exhibited the best performance (identified by Set_1), subsets with intermediate performance (Set_2 and Set_3), and the subsets with the highest deviations (Set_4). The values of AARD obtained for the global data (i.e., including training and testing subsets) were 2.67, 3.78, 4.10 and 5.23% for Set_1, Set_2, Set_3 and Set_4, respectively. The highest modeling error obtained by Set_1 and Set_4 corresponded to 64.63 and 83.79%, respectively. These results demonstrate the impact of sampling variation on modeling errors. A more detailed analysis of the performance in estimating the T_p of pure compounds using the best model identified, with the random division corresponding to Set_1 is provided in the next section.

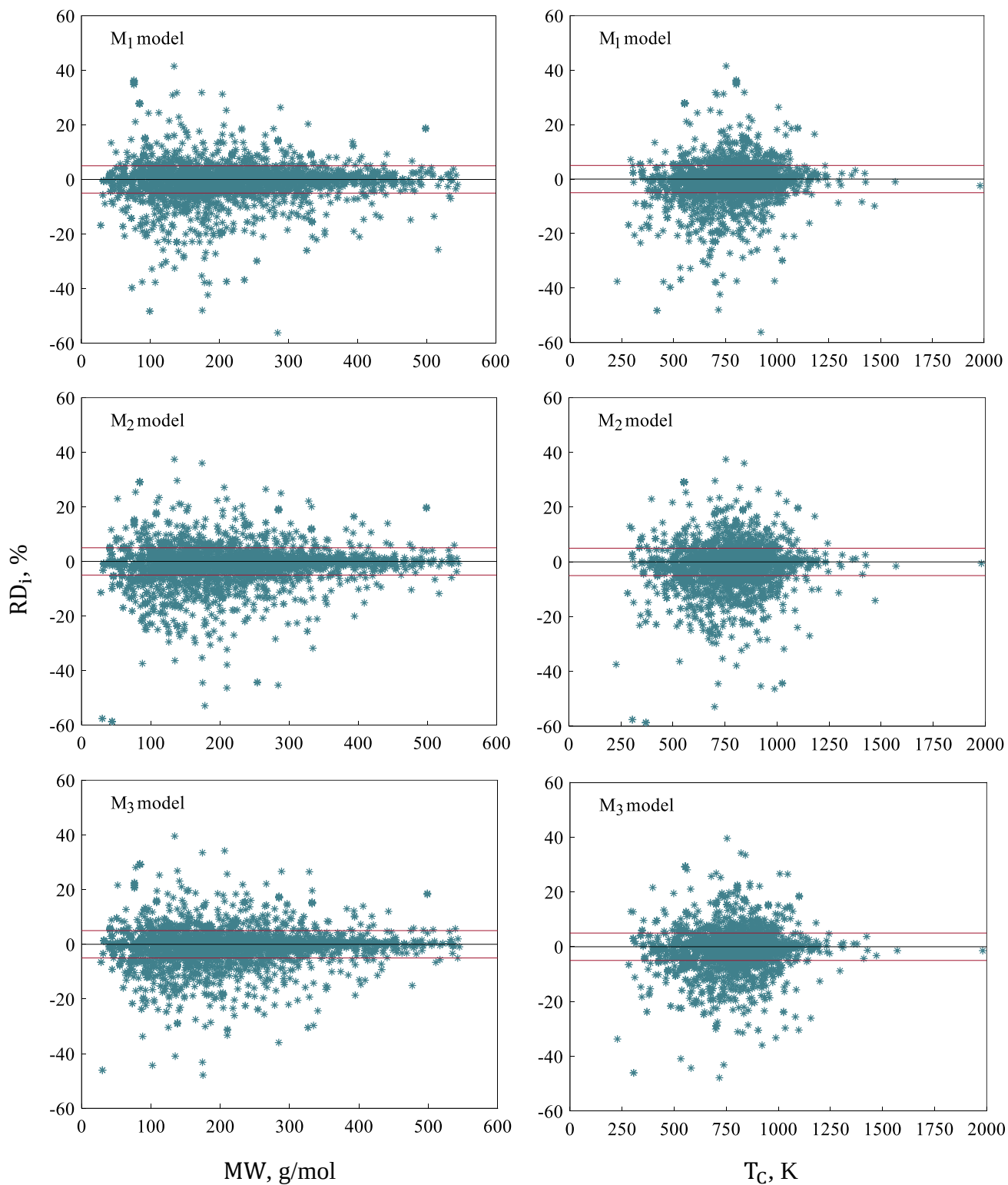


Figure 4.3 RD_i values as a function of MW and T_C for the calculation of T_P of pure compounds using different models based on ML algorithms.

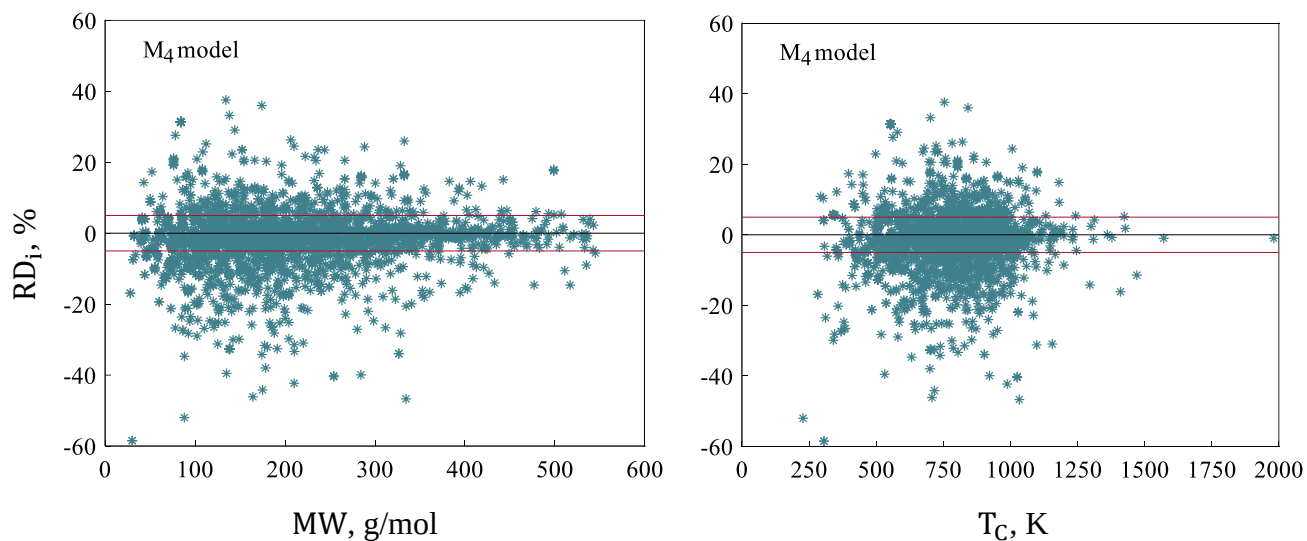


Figure 4.3 (continuation).

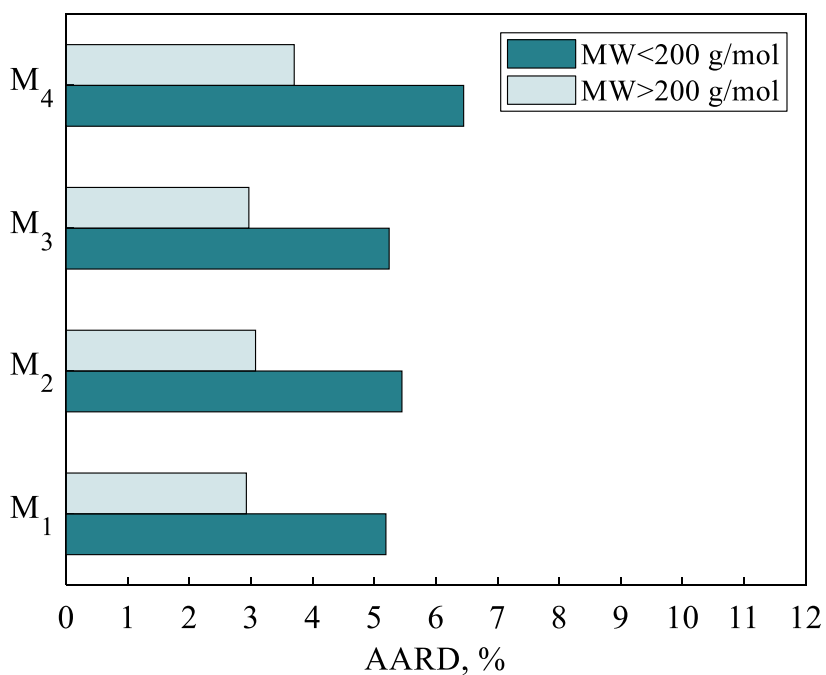


Figure 4.4 AARD values for the T_p calculation in two different MW ranges using different models based on ML algorithms.

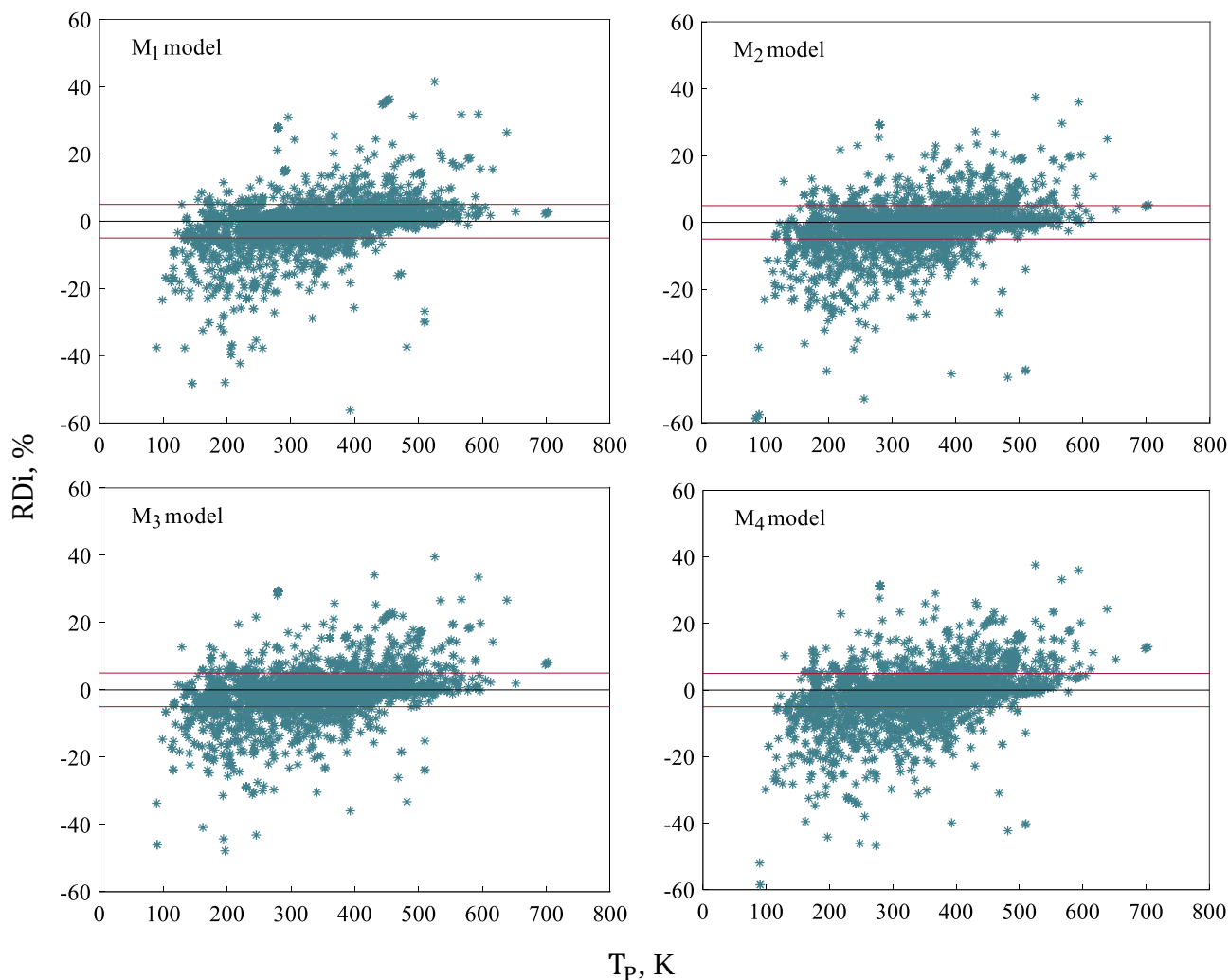


Figure 4.5 RD_i values as a function of T_p for the calculation of T_p of pure compounds using different models based on ML algorithms.

4.3.2 Capabilities and limitations of the best model based on ML algorithms to estimate the triple-point temperature of pure compounds

The comparison between the experimental and calculated values of T_p using the best model identified in this research (i.e., the M₁ model with the GB algorithm and the random division corresponding to Set_1) is depicted in Figure 4.6. The R^2 values were 0.99 and 0.86 for training and testing, respectively. The results indicated that this model has limitations in accurately predicting the T_p of pure compounds. Figure 4.6 also displays the histogram of RD_i values obtained using such model. For the training subset, 99% of the data exhibited RD_i values within $\pm 10\%$, while 97% of the data showed modeling errors within $\pm 5\%$. For the

testing subset, the modeling errors ranged from -10 to 10% and from -5 to 5% for 68 and 46% of the data, respectively. As expected, modeling errors were highest for the testing subset.

Table 4.2 Results of statistical metrics to calculate T_p with the M_1 model for the training, testing and global subsets using different random division sets.

Random division set	Subset	Statistical metric				
		AARD, %	Min ARD_i , %	Max ARD_i , %	σ , %	R^2
Set_1	Training	1.19	2.30E-05	22.13	1.47	0.99
	Testing	8.39	2.60E-03	64.63	8.87	0.86
	Global	2.67	2.30E-05	64.63	5.12	0.96
Set_2	Training	2.52	1.20E-03	40.91	3.19	0.99
	Testing	8.93	1.67E-02	78.20	8.69	0.83
	Global	3.78	1.20E-03	78.20	5.42	0.96
Set_3	Training	2.66	6.56E-04	36.55	3.15	0.99
	Testing	9.00	7.14E-04	65.47	9.89	0.79
	Global	4.10	6.56E-04	65.47	6.08	0.95
Set_4	Training	4.17	7.33E-05	38.03	4.80	0.97
	Testing	9.58	3.80E-03	83.79	9.14	0.84
	Global	5.23	7.33E-05	83.79	6.29	0.95

[†] Symbols Min and Max represent the minimum and maximum values of ARD_i .

Figure 4.7 illustrates the AARD values for the T_p estimation across different ranges of MW using the best model. The highest modeling errors, corresponding to AARD values of 3.52, 3.18 and 3.35% were obtained for compounds with $MW < 100$ g/mol, ranging from 100 to 200 g/mol and $MW > 500$ g/mol, respectively. In regions with intermediate values of MW (i.e., from 200 to 500 g/mol), this model achieved $AARD < 2.15\%$. The significant deviations observed for high MW compounds can be attributed to the limited and inconsistent experimental data on T_p reported in the literature for these types of compounds.

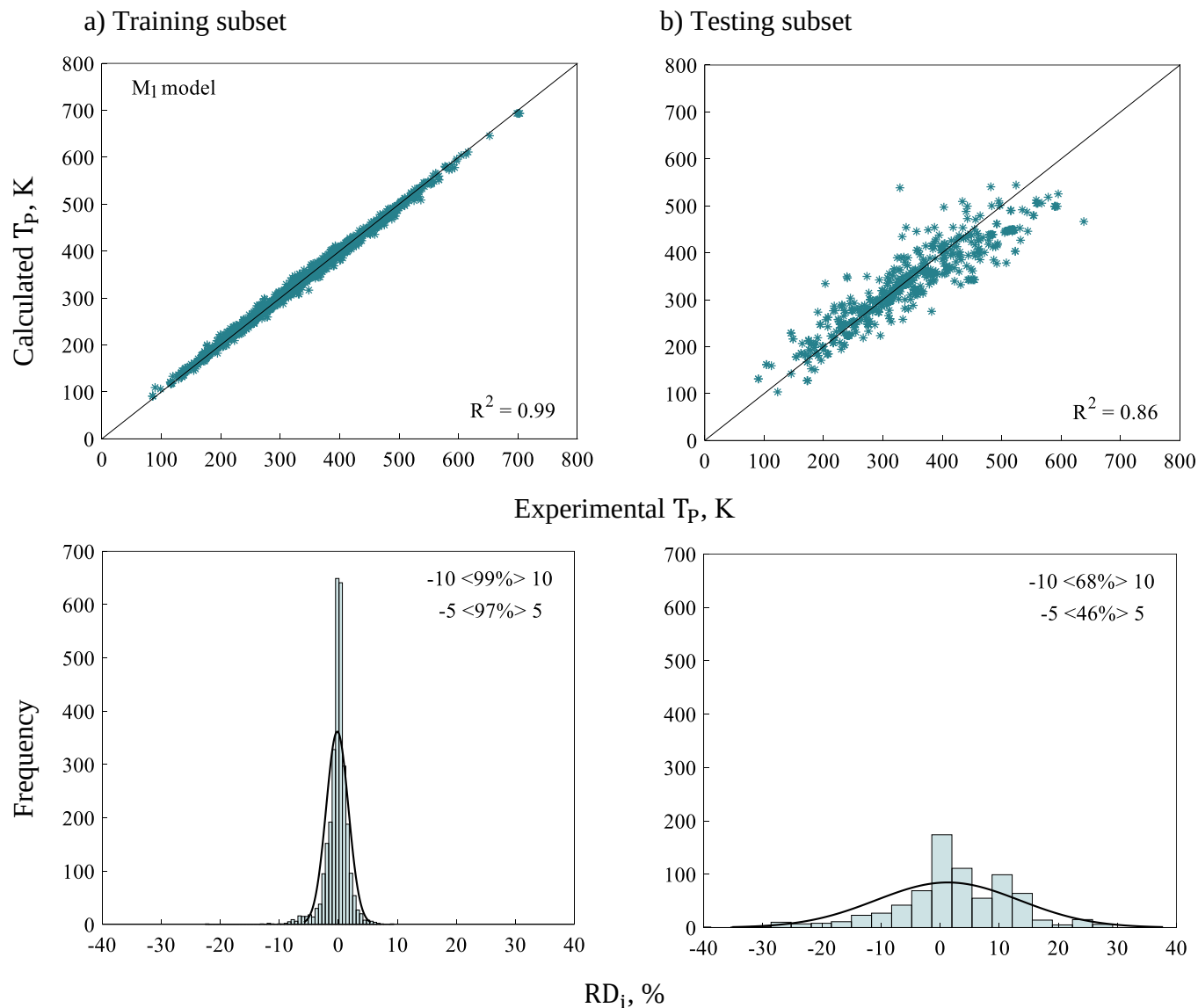


Figure 4.6 Results of the calculation of T_p of pure compounds and histograms of RD values obtained for the training and testing subsets using the M_1 model with the GB algorithm and the random division corresponding to Set_1.

A detailed analysis of the performance of the M_1 model with the GB algorithm and the random division corresponding to Set_1 is presented in Table 4.3 for different chemical families. Overall, the model was able to estimate the T_p of alkanes, alkynes, alcohols, polyols, aromatic alcohols, cycloalkanes, cycloalkenes, nitriles, compounds containing I and anhydrides with acceptable accuracy (i.e., AARD < 2%). The chemical family that exhibited the best performance was polyols, with an AARD value of 0.85%, σ of 0.88%, maximum

value of ARD_i of 3.97% and a R^2 value of 0.99. In contrast, the highest deviations were obtained for alkenes, ethers, esters, amines/imine/amides, compounds containing NO_2 , compounds containing halogens (i.e., Cl, F and Br), polyfunctional halides and sulfide/thiophenes with values of AARD ranging from 3.05 to 5.73%. The chemical family that exhibited the lowest performance in estimating T_p corresponded to polyfunctional halides with an AARD value of 5.73%, σ of 13.81%, maximum value of ARD_i of 57.53% and a R^2 value of 0.89. The poor performance of the model for these chemical families can be attributed to the presence of several polar compounds including halogens and NO_2 . The high electronegativity of these types of compounds makes predicting their thermodynamic properties challenging. Overall, the model accurately estimated the T_p of several chemical families. Figure 4.8 illustrates the results of the calculation of T_p for alkanes, alcohols, cycloalkanes, nitriles, multihalogen compounds and anhydrides achieving R^2 values ranging from 0.92 to 0.99. The information presented in Table 4.3 and Figure 4.8 is associated with the data used for both training and testing.

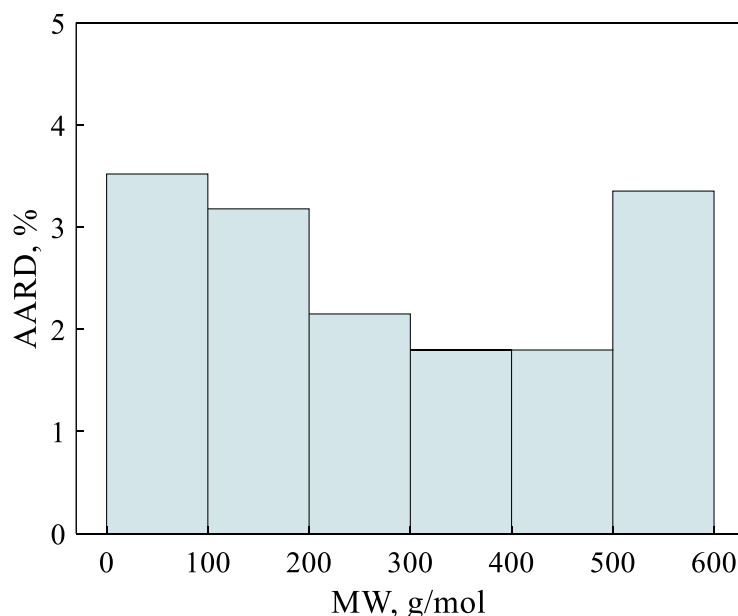


Figure 4.7 AARD values for the T_p calculation in different MW ranges using the M_1 model with the GB algorithm and the random division corresponding to Set_1.

Table 4.3 Results of statistical metrics to calculate T_p for different chemical families using the M_1 model with the best performance.

Chemical family	No.	No.	Statistical metric				
	Compounds	Data	AARD, %	Min ARD _i , %	Max ARD _i , %	σ , %	R^2
Alkanes	37	336	0.93	9.55E-05	45.27	3.58	0.99
Alkenes	20	43	3.90	1.37E-02	55.63	11.69	0.96
Alkynes	7	13	1.24	1.91E-02	4.67	1.55	0.99
Alcohols	23	144	1.13	2.70E-03	45.59	5.31	0.98
Polyols	37	75	0.85	1.42E-02	3.97	0.88	0.99
Cycloaliphatic alcohols	12	18	2.34	5.11E-02	13.51	4.10	0.90
Aromatic alcohols	30	65	1.90	8.90E-03	44.19	5.55	0.92
Cycloalkanes	25	100	1.83	1.22E-01	64.63	6.98	0.92
Cycloalkenes	7	21	1.41	7.16E-02	2.82	0.95	0.99
Carboxylic acids	89	176	2.03	8.50E-03	30.69	4.41	0.95
Aldehydes	21	37	2.77	7.18E-02	18.68	3.78	0.96
Ketones	56	93	2.93	1.96E-02	33.89	5.11	0.95
Ethers	22	30	3.82	1.05E-01	11.45	4.05	0.98
Esters	83	162	4.58	2.90E-03	31.36	5.31	0.95
Nitriles	76	125	1.74	1.19E-02	23.66	3.17	0.97
Amines/imines/amides	300	513	3.05	1.20E-03	43.61	4.78	0.94
Aromatic compounds	128	341	2.94	4.34E-04	45.76	4.64	0.96
Compounds containing NO ₂	73	155	3.12	1.20E-03	19.28	3.88	0.95
Compounds containing Cl	93	196	3.52	7.58E-04	20.94	4.46	0.97
Compounds containing F	57	82	5.53	2.30E-05	41.97	8.82	0.95
Compounds containing Br	75	105	3.78	1.91E-02	19.27	4.72	0.95
Compounds containing I	25	42	1.85	1.50E-02	14.61	3.05	0.99
Multihalogen compounds	18	27	2.64	6.36E-01	7.82	1.91	0.99
Polyfunctional halides	16	32	5.73	2.70E-01	57.53	13.81	0.89
Sulfides/thiophenes	40	53	3.69	4.71E-02	18.05	4.39	0.97
Mercaptans	15	17	2.23	1.79E-01	15.41	3.84	0.98

Table 4.3 (continuation).

Chemical family	No. Compounds	No. Data	Statistical metric				
			AARD, %	Min ARD_i , %	Max ARD_i , %	σ , %	R^2
Silanes/siloxanes	25	32	2.70	7.81E-02	16.47	3.02	0.96
Isocyanates/diisocyanates	3	3	1.27	9.38E-02	2.09	1.04	0.99
Anhydrides	24	46	1.73	1.23E-02	9.49	1.90	0.99
Nitroamines	67	100	2.10	5.80E-03	23.29	3.58	0.93
Peroxides	2	2	1.27	9.43E-01	1.59	0.46	0.99

Table 4.4 lists the global performance metrics (including both training and testing subsets) obtained for various functional groups using the best model (i.e., the M_1 model with the GB algorithm and the random division corresponding to Set_1). The AARD values ranged from 0.76 (for =NH) to 20.98% (for =S). For illustration purposes, Figure 4.9 displays the AARD values obtained for the main functional groups included in the analysis. The model estimates the T_p of compounds containing -CH₂- (outside rings), -OH (alcohol), -CN, -I, -SH, =O, >S=O (outside rings), =C= (outside rings) and =NH with acceptable accuracy (i.e., AARD < 2%), as shown in Table 4.4 and Figure 4.9. The functional groups that exhibited the highest modeling errors (i.e., AARD > 4%) were -F, =C< (outside rings), >NH (inside rings), -S- (inside rings) and =S, as illustrated in Figure 4.9 and Table 4.4. The compound with the highest modeling error was dispiro[2.0.2.1]heptane (with an ARD_i value of 64.63%), while the lowest ARD_i value (of 2.30E-05 %) was achieved for difluoromethane.

The outcomes of this research have facilitated the identification of optimal ML algorithm alternatives for predicting the T_p of pure compounds. Although the model developed in this chapter was able to estimate the T_p of specific compounds with acceptable accuracy, further efforts are needed to develop enhanced models for predicting this thermodynamic property. In that regard, this study represents a starting point for implementing and improving different strategies based on ML algorithms to estimate the T_p of pure compounds for PSE calculations.

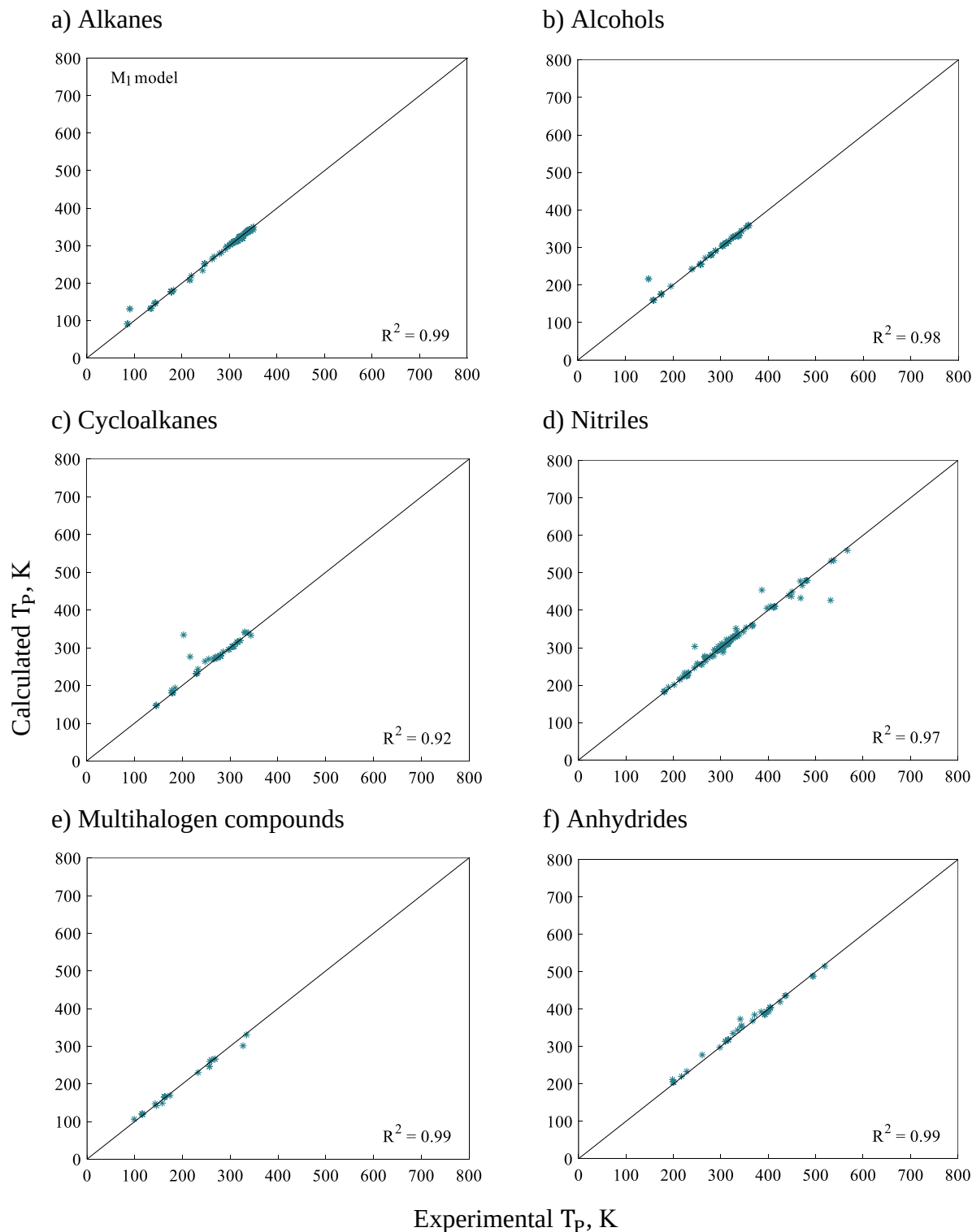


Figure 4.8 Results of the calculation of T_p of pure compounds for different chemical families using the M_1 model with the GB algorithm and the random division corresponding to Set_1.

Table 4.4 Results of statistical metrics to calculate T_p for the functional groups used by the M_1 model with the best performance.

Functional group	No. Compounds	No. Data	Statistical metric				
			AARD, %	Min ARD_i , %	Max ARD_i , %	σ , %	R^2
-CH ₃ (outside rings)	1076	2108	2.30	9.55E-05	63.83	4.44	0.97
=CH- (inside rings)	960	1655	3.03	4.34E-04	63.83	4.79	0.94
=C< (inside rings)	991	1675	2.98	4.34E-04	63.83	4.79	0.94
-CH ₂ - (outside rings)	855	1696	1.94	2.30E-05	45.76	3.73	0.97
-CH ₂ - (inside rings)	321	504	2.79	4.34E-04	64.63	5.36	0.97
>CH- (outside rings)	228	310	2.81	2.90E-03	31.36	5.39	0.96
>CH- (inside rings)	205	264	3.19	2.60E-03	43.61	5.39	0.97
-OH (alcohol)	201	373	1.52	2.60E-03	45.59	4.02	0.98
-NO ₂ (outside rings)	198	356	2.58	1.20E-03	63.83	4.90	0.93
>C< (outside rings)	187	288	3.39	7.58E-04	45.76	6.06	0.96
=CH- (outside rings)	165	217	3.86	2.90E-03	41.97	5.90	0.95
-COOH (acid)	160	269	2.62	8.50E-03	63.83	5.85	0.90
>C=O (inside rings)	151	198	2.89	5.00E-03	33.89	4.63	0.92
-COO- (ester)	145	229	3.43	2.90E-03	31.36	4.58	0.97
>C< (inside rings)	139	161	2.95	4.34E-04	64.63	6.16	0.92
-O- (outside rings)	138	181	2.82	5.70E-03	23.47	3.89	0.96
-NH ₂	125	277	2.40	1.20E-03	43.61	5.41	0.93
-Cl	122	245	3.67	7.58E-04	57.53	6.39	0.97
-N= (inside rings)	118	144	3.26	7.40E-03	20.13	4.48	0.93
>C=O (outside rings)	115	198	3.24	1.20E-03	57.53	7.22	0.93
-O- (inside rings)	102	135	2.16	5.90E-03	16.47	3.12	0.97
>NH (outside rings)	99	132	2.21	5.80E-03	25.47	4.02	0.93
-CN	90	140	1.92	1.19E-02	23.66	3.39	0.97
>N- (outside rings)	83	118	3.44	3.03E-02	24.41	4.78	0.95
-Br	82	115	3.71	1.91E-02	19.27	4.56	0.96

Table 4.4 (continuation).

Functional group	No. Compounds	No. Data	Statistical metric				
			AARD, %	Min ARD _i , %	Max ARD _i , %	σ , %	R ²
-F	81	117	4.70	2.30E-05	41.97	7.58	0.96
=CH ₂ (outside rings)	78	112	3.53	1.37E-02	55.63	8.28	0.97
>N- (inside rings)	76	110	2.53	7.40E-03	13.30	3.74	0.95
=C< (outside rings)	74	105	4.37	1.66E-02	41.97	7.95	0.93
-OH (phenol)	67	143	2.02	8.90E-03	44.19	4.60	0.94
>NH (inside rings)	57	95	5.11	5.00E-03	18.56	5.77	0.90
≡C- (outside rings)	42	49	3.35	1.91E-02	44.19	6.64	0.96
-N= (outside rings)	42	44	3.67	2.60E-03	25.47	5.71	0.85
-S- (inside rings)	41	54	4.22	1.80E-01	22.64	4.94	0.97
-CHO (aldehyde)	34	54	3.55	7.18E-02	22.88	5.28	0.88
-S- (outside rings)	32	40	2.04	1.27E-02	18.05	3.10	0.98
-SO ₂ - (outside rings)	27	41	2.16	3.91E-02	27.95	4.73	0.83
-I	26	43	1.99	1.50E-02	14.61	3.15	0.99
≡CH (outside rings)	26	32	3.91	1.91E-02	44.19	8.07	0.95
-SH	20	25	1.96	1.79E-01	15.41	3.23	0.99
>Si< (outside rings)	15	21	2.31	7.81E-02	7.14	1.80	0.97
=O	14	19	0.88	5.70E-03	3.12	0.85	0.99
>Si< (inside rings)	10	11	3.43	8.97E-02	16.47	4.57	0.84
-SO ₄ - (outside rings)	9	13	2.62	1.36E-01	13.77	4.04	0.93
>S=O (outside rings)	9	18	0.90	9.85E-02	3.39	1.06	0.99
=C= (outside rings)	7	13	0.63	3.91E-02	2.09	0.75	0.99
-NO ₃	7	19	2.32	2.59E-02	15.08	4.01	0.99
-SO ₂ - (inside rings)	5	9	3.43	1.02E+00	10.96	3.02	0.96
=NH	5	8	0.76	5.74E-02	2.90	0.91	0.99
=S	3	10	20.78	3.86E-01	24.63	7.41	0.96
-SO ₄ - (inside rings)	3	3	2.76	1.38E+00	4.88	1.86	0.68
>N=O (inside rings)	3	3	2.92	2.28E+00	3.88	0.84	0.99

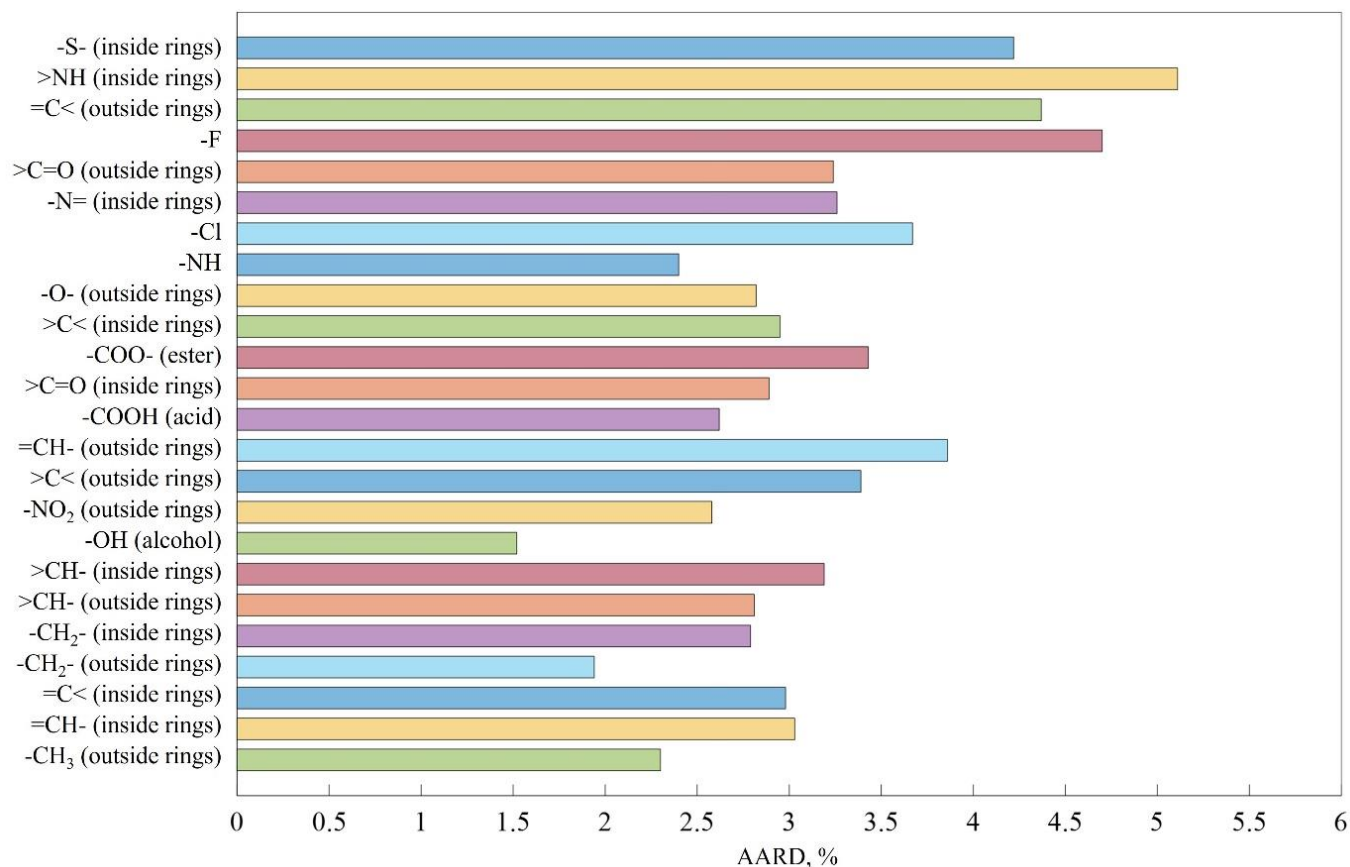


Figure 4.9 Comparison of modeling errors obtained for the main functional groups in the calculation of T_p of pure compounds using the M_1 model with the GB algorithm and the random division corresponding to Set_1.

4.4 Conclusions

An evaluation and comparative analysis of different models based on ML algorithms for estimating the T_p of pure compounds was carried out using an AutoML approach. The results indicated that GB was the ML algorithm with the best potential to estimate this thermodynamic property. Furthermore, the M_1 model, including the MW and n_i as input variables of the ML algorithm accomplished the best performance. The gaps and capabilities of the best model were highlighted for different chemical families and functional groups. The modeling errors for this model varied from 0.85 to 5.73% for the main chemical families included in the analysis. The T_p of alkanes, alkynes, alcohols, polyols, aromatic alcohols,

cycloalkanes, cycloalkenes, nitriles, compounds containing I and anhydrides were estimated with an acceptable accuracy, while for polar compounds (including halogens and those containing NO₂), the deviations were higher. In the training subset, the 99 and 97% of the data exhibited RD_i values within ± 10 and $\pm 5\%$, respectively. For the testing subset, modeling errors of ± 10 and $\pm 5\%$ were obtained for 68 and 46% of the data, respectively. Modeling errors were higher for low and high MW compounds (i.e., MW < 200 g/mol and MW > 500 g/mol). The results indicate that the prediction of T_p of pure compounds using ML algorithms may be challenging. Although this type of approach allows the identification of ML algorithms with better features for predicting this property, better analysis and systematic processes are still required to improve the model performance.

Chapter 5. Final remarks

In this thesis, the prediction of thermodynamic properties of pure compounds was studied and analyzed in detail, focusing on GCMs and ML algorithms approaches. The analysis of GCMs highlighted their potential for predicting thermodynamic properties of ILs. However, their accuracy is constrained by the insufficient experimental data available for these types of molecules. ML algorithms, such as ANNs, can significantly enhance the performance of GCMs for the estimation of thermodynamic properties of ILs. Integrating GCMs with ANNs enhanced the estimations of liquid heat capacities of pure compounds. This versatile and reliable model for the liquid heat capacity calculation outperformed the GCM implemented in the Aspen Plus[®] simulator in accuracy and scope of application. Furthermore, this model demonstrated improved performance near the critical point, where traditional models often fail. A hybrid model based on GCMs and ML algorithms has shown potential for estimating the triple point temperature of pure compounds. An evaluation and comparative analysis using an AutoML approach revealed that gradient boosting was the ML algorithm with the best performance to estimate this thermodynamic property. Despite the promising results, further development is necessary, particularly for polar compounds such as those containing halogens or NO₂ functional groups, where modeling errors were higher.

Overall, this research demonstrates that combining traditional thermodynamic models like GCMs with advanced ML algorithms provides more accurate and versatile approaches for predicting thermodynamic properties, making them alternative tools for PSE calculations of different units and equipment.

Chapter 6. References

Abiodun O.I., Jantan A., Omolara A.E., Dada K.V., Mohamed N.A., Arshad H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* 4 (2018) e00938.

Abiodun O.I., Jantan A., Omolara A.E., Dada K.V., Umar A.M., Linus O.U., Arshad H., Kazaure A.A., Gana U., Kiru M.U. Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition. Institute of Electrical and Electronics Engineers Access 7 (2019) 158820-158846.

Adams C.J., Earle M.J., Seddon K.R. Catalytic cracking reactions of polyethylene to light alkanes in ionic liquids. *Green Chemistry* 2 (2000) 21-24.

Admire B., Lian B., Yalkowsky S.H. Estimating the physicochemical properties of polyhalogenated aromatic and aliphatic compounds using UPPER: Part 1. Boiling point and melting point. *Chemosphere* 119 (2015) 1436-1440.

Aguilar-Elguezabal A., De la Torre-Sáenz L., Román-Aguirre M., Álvarez-Contreras L. Ionic liquid as green solvent for the synthesis of α -terpineol from α -pinene. *Sustainable Chemistry and Pharmacy* 15 (2020) 100207.

Aguirre C.L., Cisternas L.A. Density estimation of ammonium-based ionic liquids. *Chemical Engineering Communications* 199 (2012) 283-289.

Aguirre C.L., Cisternas L.A., Valderrama J.O. Melting-Point Estimation of Ionic Liquids by a Group Contribution Method. *International Journal of Thermophysics* 33 (2012) 34-46.

Ahmad M.W., Reynolds J., Rezgui Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of Cleaner Production* 203 (2018) 810–821.

Ahmadi A., Haghighbakhsh R., Raeissi S., Hemmati V. A simple group contribution correlation for the prediction of ionic liquid heat capacities at different temperatures. *Fluid Phase Equilibria* 403 (2015) 95-103.

Ahmed R., Mahadzir S., Rozali N.E.M., Biswas K., Matovu F., Ahmed K. Artificial intelligence techniques in refrigeration system modelling and optimization: A multi-disciplinary review. *Sustainable Energy Technologies and Assessments* 47 (2021) 101488.

- Alade I.O., Rahman M.A.A., Abbas Z., Yaakob Y., Saleh T.A. Application of support vector regression and artificial neural network for prediction of specific heat capacity of aqueous nanofluids of copper oxide. *Solar Energy* 197 (2020) 485–490.
- Alanne K., Sierla S. An overview of machine learning applications for smart buildings. *Sustainable Cities and Society* 76 (2022) 103445.
- Albert J., Müller K. A Group Contribution Method for the Thermal Properties of Ionic Liquids. *Industrial and Engineering Chemistry Research* 53 (2014) 17522-17526.
- Aldosari M.N., Yalamanchi K.K., Gao X., Sarathy S.M. Predicting entropy and heat capacity of hydrocarbons using machine learning. *Energy and AI* 4 (2021) 100054.
- Almeida H.F.D., Lopes-da-Silva J.A., Freire M.G., Coutinho J.A.P. Surface tension and refractive index of pure and water-saturated tetradecyltriethylphosphonium-based ionic liquids. *The Journal of Chemical Thermodynamics* 57 (2013) 372-379.
- Alshehri A.S., Tula A.K., You F., Gani R. Next generation pure component property estimation models: With and without machine learning techniques. *American Institute of Chemical Engineers Journal* 68 (2022) e17469.
- Aparicio S., Atilhan M., Karadas F. Thermophysical Properties of Pure Ionic Liquids: Review of Present Situation. *Industrial and Engineering Chemistry Research* 49 (2010) 9580-9595.
- Arai S., Nakashima K., Tanino T., Ogino C., Kondo A., Fukuda H. Production of biodiesel fuel from soybean oil catalyzed by fungus whole-cell biocatalysts in ionic liquids. *Enzyme and Microbial Technology* 46 (2010) 51-55.
- Arcaklıoğlu E., Çavuşoğlu A., Erişen A. Thermodynamic analyses of refrigerant mixtures using artificial neural networks. *Applied Energy* 78 (2004) 219–230.
- Archana P., Divyabharathi P., Balaji S.R., Kumaresan N., Veeramanikandan P., Naitik S.T., Rafi S.M., Nandankar P.V., Manikandan G. Face recognition based vehicle starter using machine learning. *Measurement: Sensors* 24 (2022) 100575.
- Atashrouz S., Hemmati-Sarapardeh A., Mirshekar H., Nasernejad B., Moraveji M.K. On the evaluation of thermal conductivity of ionic liquids: Modeling and data assessment. *Journal of Molecular Liquids* 224 (2016) 648-656.

Atashrouz S., Mirshekar H., Hemmati-Sarapardeh A. A soft-computing technique for prediction of water activity in PEG solutions. *Colloid and Polymer Science* 295 (2017) 421–432.

Ayan E., Baylan N., Çehreli S. Optimization of reactive extraction of propionic acid with ionic liquids using central composite design. *Chemical Engineering Research and Design* 153 (2020) 666-676.

Behrooz H.A., Boojarjomehry R.B. Prediction of limiting activity coefficients for binary vapor-liquid equilibrium using neural networks. *Fluid Phase Equilibria* 433 (2017) 174–183.

Bélisle E., Huang Z., Digabel S.L., Gheribi A.E. Evaluation of machine learning interpolation techniques for prediction of physical properties. *Computational Materials Science* 98 (2015) 170–177.

Bottou L. Large-Scale Machine Learning with Stochastic Gradient Descent. *Proceedings of COMPSTAT'2010* (2010) 177–186.

Bouarab A.F., Harvey J.-P., Robelin C. Viscosity models for ionic liquids and their mixtures. *Physical Chemistry Chemical Physics* 23 (2021) 733-752.

Böhm M., Tietze A.A., Heimer P., Chen M., Imhof D. Ionic liquids as reaction media for oxidative folding and native chemical ligation of cysteine-containing peptides. *Journal of Molecular Liquids* 192 (2014) 67-70.

Brinchi L., Germani R., Savelli G. Ionic liquids as reaction media for esterification of carboxylate sodium salts with alkyl halides. *Tetrahedron Letters* 44 (2003) 2027-2029.

Brogan A.P.S., Bui-Le L., Hallett J.P. Non-aqueous homogenous biocatalytic conversion of polysaccharides in ionic liquids using chemically modified glucosidase. *Nature Chemistry* 10 (2018) 859-865.

Bunquin R.M.A., Caparanga A.R. Predicting the heat capacities of ammonium- and phosphonium-based deep eutectic solvents using artificial neural network. *Journal of Physics: Conference Series* 1893 (2021) 012001.

Carvalho P.J., Álvarez V.H., Machado J.J.B., Pauly J., Daridon J.-L., Marrucho I.M., Aznar M., Coutinho J.A.P. High pressure phase behavior of carbon dioxide in 1-alkyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide ionic liquids. *The Journal of Supercritical Fluids* 48 (2009) 99-107.

Carvalho P.J., Álvarez V.H., Marrucho I.M., Aznar M., Coutinho J.A.P. High carbon dioxide solubilities in trihexyltetradecylphosphonium-based ionic liquids. *The Journal of Supercritical Fluids* 52 (2010) 258-265.

Ceriani R., Gani R., Meirelles A.J.A. Prediction of heat capacities and heats of vaporization of organic liquids by group contribution methods. *Fluid Phase Equilibria* 283 (2009) 49–55.

Chen B.-K., Liang M.-J., Wu T.-Y., Wang H.P. A high correlate and simplified QSPR for viscosity of imidazolium-based ionic liquids. *Fluid Phase Equilibria* 350 (2013) 37-42.

Chen Y., Cai Y., Thomsen K., Kontogeorgis G.M., Woodley J.M. A group contribution-based prediction method for the electrical conductivity of ionic liquids. *Fluid Phase Equilibria* 509 (2020) 112462.

Chen Y., Kontogeorgis G.M., Woodley J.M. Group Contribution Based Estimation Method for Properties of Ionic Liquids. *Industrial and Engineering Chemistry Research* 58 (2019) 4277-4292.

Chopra R., Roy S.S. End-to-End Reinforcement Learning for Self-driving Car. *Advanced Computing and Intelligent Engineering* 1082 (2020) 53–61.

Chouai A., Laugier S., Richon D. Modeling of thermodynamic properties using neural networks: Application to refrigerants. *Fluid Phase Equilibria* 199 (2002) 53–62.

Conesa A., Shen S., Coronas A. Liquid densities, kinematic viscosities, and heat capacities of some ethylene glycol dimethyl ethers at temperatures from 283.15 to 423.15K. *International Journal of Thermophysics* 19 (1998) 1343–1358.

Coniglio L., Rauzy E., Berro C. Representation and prediction of thermophysical properties of heavy hydrocarbons. *Fluid Phase Equilibria* 87 (1993) 53–88.

Conrad F., Mälzer M., Schwarzenberger M., Wiemer H., Ihlenfeldt S. Benchmarking AutoML for regression tasks on small tabular data in materials design. *Scientific Reports* 12 (2022) 19350.

Constantinou L., Gani R. New group contribution method for estimating properties of pure compounds. *AIChE Journal* 40 (1994) 1697–1710.

Coutinho J.A.P., Carvalho P.J., Oliveira N.M.C. Predictive methods for the estimation of thermophysical properties of ionic liquids. *Royal Society of Chemistry Advances* 2 (2012) 7322-7346.

Dadgostar N., Shaw J.M. A predictive correlation for the constant-pressure specific heat capacity of pure and ill-defined liquid hydrocarbons. *Fluid Phase Equilibria* 313 (2012) 211–226.

Daniel C.I., Albo J., Santos E., Portugal C.A.M., Crespo J.G., Irabien A. A group contribution method for the influence of the temperature in the viscosity of magnetic ionic liquids. *Fluid Phase Equilibria* 360 (2013) 29-35.

Deetlefs M., Seddon K.R., Shara M. Predicting physical properties of ionic liquids. *Physical Chemistry Chemical Physics* 8 (2006) 642-649.

de Oliveira F.M., de Carvalho L.S., Teixeira L.S.G., Fontes C.H., Lima K.M.G., Câmara A.B.F., Araújo H.O.M., Sales R.V. Predicting Cetane Index, Flash Point, and Content Sulfur of Diesel–Biodiesel Blend Using an Artificial Neural Network Model. *Energy Fuels* 31 (2017) 3913–3920.

Deng S., Su W., Zhao L. A neural network for predicting normal boiling point of pure refrigerants using molecular groups and topological index. *International Journal of Refrigeration* 63 (2016) 63-71.

Devi B.L.A.P., Guo Z., Xu X. Characterization of ionic liquid-based biocatalytic two-phase reaction system for production of biodiesel. *American Institute of Chemical Engineers Journal* 57 (2011) 1628-1637.

Diedrichs A., Rarey J., Gmehling J. Prediction of liquid heat capacities by the group contribution equation of state VTPR. *Fluid Phase Equilibria* 248 (2006) 56–69.

Ding J., Xu N., Nguyen M.T., Qiao Q., Shi Y., He Y., Shao Q. Machine learning for molecular thermodynamics. *Chinese Journal of Chemical Engineering* 31 (2021) 227–239.

Dobbelaere M.R., Plehiers P.P., Van de Vijver R., Stevens C.V., Van Geem K.M. Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats. *Engineering* 7 (2021) 1201-1211.

Do Carmo F.R., da Silva M.R.L., Alves A.A.A., Evangelista N.S. A new method for predicting the isobaric heat capacity of biodiesel-related esters based on the corresponding states principle. *Fluid Phase Equilibria* 521 (2020) 112734.

Elbro H.S., Fredenslund A., Rasmussen P. Group contribution method for the prediction of liquid densities as a function of temperature for solvents, oligomers, and polymers. *Industrial and Engineering Chemistry Research* 30 (1991) 2576–2582.

El-Harbawi M., Samir B.B., Babaa M.-R., Mutalib M.I.A. A New QSPR Model for Predicting the Densities of Ionic Liquids. *Arabian Journal for Science and Engineering* 39 (2014) 6767-6775.

Evangelista N.S., do Carmo F.R., de Santiago-Aguiar R.S., de Sant'Ana H.B. Development of a New Group Contribution Method Based on GCVOL Model for the Estimation of Pure Ionic Liquid Density over a Wide Range of Temperature and Pressure. *Industrial and Engineering Chemistry Research* 53 (2014) 9506–9512.

Farzi R., Esmaeilzadeh F. Prediction of densities of pure ionic liquids using Esmaeilzadeh-Roshanfekr equation of state and critical properties from group contribution method. *Fluid Phase Equilibria* 423 (2016) 101-108.

Fatehi M.-R., Raeissi S., Mowla D. An artificial neural network to calculate pure ionic liquid densities without the need for any experimental data. *The Journal of Supercritical Fluids* 95 (2014) 60-67.

Feng G., Li S., Zhao W., Cummings P.T. Microstructure of room temperature ionic liquids at stepped graphite electrodes. *American Institute of Chemical Engineers Journal* 61 (2015) 3022-3028.

Field M., Hardcastle N., Jameson M., Aherne N., Holloway L. Machine learning applications in radiation oncology. *Physics and Imaging in Radiation Oncology* 19 (2021) 13-24.

Freire M.G., Teles A.R.R., Rocha M.A.A., Schröder B., Neves C.M.S.S., Carvalho P.J., Evtuguin D.V., Santos L.M.N.B.F., Coutinho J.A.P. Thermophysical Characterization of Ionic Liquids Able To Dissolve Biomass. *J. Chem. Eng. Data* 56 (2011) 4813–4822.

Fukuyama T., Inouye T., Ryu I. Atom transfer carbonylation using ionic liquids as reaction media. *Journal of Organometallic Chemistry* 692 (2007) 685-690.

Gadamsetty S., Joshipura M., Saxena P. A review on property prediction of ionic liquids using cubic equation of state. *Institute of Technology, Nirma University, Ahmedabad* 382 (2011) 481.

Gani R. Group contribution-based property estimation methods: advances and perspectives. *Current Opinion in Chemical Engineering* 23 (2019) 184-196.

Gao N., Chen G., Tang L. A corresponding state equation for the prediction of isobaric heat capacity of liquid HFC and HFO refrigerants. *Fluid Phase Equilibria* 456 (2018) 1–6.

Gao Y., Wu L. Efficiently Mastering the Game of NoGo with Deep Reinforcement Learning Supported by Domain Knowledge. *Electronics* 10 (2021) 1533.

Gardas R.L., Coutinho J.A.P. A Group Contribution Method for Heat Capacity Estimation of Ionic Liquids. *Industrial and Engineering Chemistry Research* 47 (2008a) 5751-5757.

Gardas R.L., Coutinho J.A.P. A group contribution method for viscosity estimation of ionic liquids. *Fluid Phase Equilibria* 266 (2008b) 195-201.

Gardas R.L., Coutinho J.A.P. Extension of the Ye and Shreeve group contribution method for density estimation of ionic liquids in a wide range of temperatures and pressures. *Fluid Phase Equilibria* 263 (2008c) 26-32.

Gardas R.L., Coutinho J.A.P. Group contribution methods for the prediction of thermophysical and transport properties of ionic liquids. *American Institute of Chemical Engineers Journal* 55 (2009) 1274-1290.

Gardas R.L., Ge R., Goodrich P., Hardacre C., Hussain A., Rooney D.W. Thermophysical Properties of Amino Acid-Based Ionic Liquids. *Journal of Chemical and Engineering Data* 55 (2010) 1505-1515.

Ge R., Hardacre C., Jacquemin J., Nancarrow P., Rooney D.W. Heat Capacities of Ionic Liquids as a Function of Temperature at 0.1 MPa. Measurement and Prediction. *Journal of Chemical and Engineering Data* 53 (2008) 2148-2153.

Geurts P., Ernst D., Wehenkel L. Extremely randomized trees. *Machine Learning* 63 (2006) 3–42.

Ghandi K. A Review of Ionic Liquids, Their Limits and Applications. *Green and Sustainable Chemistry* 4 (2014) 44-53.

Gharagheizi F., Eslamimanesh A., Mohammadi A.H., Richon D. 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.

Gharagheizi F., Ilani-Kashkouli P., Mohammadi A.H., Ramjugernath D., Richon D. Development of a group contribution method for determination of viscosity of ionic liquids at atmospheric pressure. *Chemical Engineering Science* 80 (2012a) 326-333.

Gharagheizi F., Ilani-Kashkouli P., Mohammadi A.H. Group contribution model for estimation of surface tension of ionic liquids. *Chemical Engineering Science* 78 (2012b) 204-208.

Gharagheizi F., Ilani-Kashkouli P., Mohammadi A.H. Computation of normal melting temperature of ionic liquids using a group contribution method. *Fluid Phase Equilibria* 329 (2012c) 1-7.

Gharagheizi F., Ilani-Kashkouli P., Mohammadi A.H., Ramjugernath D., Richon D. Development of a group contribution method for estimating the thermal decomposition temperature of ionic liquids. *Fluid Phase Equilibria* 355 (2013) 81-86.

Gharagheizi F., Ilani-Kashkouli P., Sattari M., Mohammadi A.H., Ramjugernath D., Richon D. Development of a LSSVM-GC model for estimating the electrical conductivity of ionic liquids. *Chemical Engineering Research and Design* 92 (2014) 66-79.

Gharagheizi F., Sattari M. Prediction of Triple-Point Temperature of Pure Components Using their Chemical Structures. *Industrial and Engineering Chemistry Research* 49 (2010) 929–932.

Ghasemitabar H., Movagharnejad K. Estimation of the normal boiling point of organic compounds via a new group contribution method. *Fluid Phase Equilibria* 411 (2016) 13-23.

Gopal E.S.R. Critical point phenomena, heat capacities and the renormalization group theory of fluctuations. *Bulletin of Materials Science* 3 (1981) 91–99.

Gu F., Jiang K., Hu X., Yang J. Deep Learning-Based Image Geolocation for Travel Recommendation via Multi-Task Learning. *Journal of Circuits, Systems and Computers* 31 (2022) 2250127.

Guo J., Tucker Z.D., Wang Y., Ashfeld B.L., Luo T. Ionic liquid enables highly efficient low temperature desalination by directional solvent extraction. *Nature Communications* 12 (2021) 437.

Haghighbakhsh R., Raeissi S., Duarte A.R.C. Group contribution and atomic contribution models for the prediction of various physical properties of deep eutectic solvents. *Scientific Reports* 11 (2021) 6684.

Han J., Li M., Tian N., Liu C., Zhang Y., Ji Z., Sun X. Prediction of heat capacity of ionic liquids: A simple group contribution method. *Fluid Phase Equilibria* 565 (2023) 113675.

Honig J. M. *Thermodynamics: principles characterizing physical and chemical processes*. Academic Press (2020).

Hosseini S.M., Mulero A., Alavianmehr M.M. Predictive methods and semi-classical Equations of State for pure ionic liquids: A review. *Journal of Chemical Thermodynamics* 130 (2019) 47-94.

Huang Y., Dong H., Zhang X., Li C., Zhang S. A new fragment contribution-corresponding states method for physicochemical properties prediction of ionic liquids. *American Institute of Chemical Engineers Journal* 59 (2013) 1348-1359.

Huo Y., Xia S., Zhang Y., Ma P. Group Contribution Method for Predicting Melting Points of Imidazolium and Benzimidazolium Ionic Liquids. *Industrial and Engineering Chemistry Research* 48 (2009) 2212-2217.

Ihmels E.C., Gmehling J. Extension and Revision of the Group Contribution Method GCVOL for the Prediction of Pure Compound Liquid Densities. *Industrial and Engineering Chemistry Research* 42 (2003) 408–412.

Irimescu R., Kato K. Investigation of ionic liquids as reaction media for enzymatic enantioselective acylation of amines. *Journal of Molecular Catalysis B: Enzymatic* 30 (2004) 189-194.

Jameel A.G.A., Oudenhoven V.V., Emwas A.-H., Sarathy S.M. Predicting Octane Number Using Nuclear Magnetic Resonance Spectroscopy and Artificial Neural Networks. *Energy Fuels* 32 (2018) 6309–6329.

Jenkins H.D.B., Roobottom H.K., Passmore J., Glasser L. Relationships among Ionic Lattice Energies, Molecular (Formula Unit) Volumes, and Thermochemical Radii. *Inorganic Chemistry* 38 (1999) 3609–3620.

Jirasek F., Alves R.A.S., Damay J., Vandermeulen R.A., Bamler R., Bortz M., Mandt S., Kloft M., Hasse H. Machine Learning in Thermodynamics: Prediction of Activity Coefficients by Matrix Completion. *Journal of Physical Chemistry Letters* 11 (2020) 981–985.

Jirasek F., Hasse H. Combining Machine Learning with Physical Knowledge in Thermodynamic Modeling of Fluid Mixtures. *Annual Review of Chemical and Biomolecular Engineering* 14 (2023) 31–51.

Joback K.G. A unified approach to physical property estimation using multivariate statistical techniques. M.C. Thesis in Chemical Engineering, Massachusetts Institute of Technology (1984).

John V., Liu Z., Guo C., Mita S., Kidono K. Real-Time Lane Estimation Using Deep Features and Extra Trees Regression. *Lecture Notes in Computer Science* 9431 (2016) 721–733.

Jovanović J.D., Knežević-Stevanović A.B., Grozdanić D.K. Prediction of high pressure liquid heat capacities of organic compounds by a group contribution method. *Journal of the Serbian Chemical Society* 76 (2011) 417–423.

Kamari A., Bahadori A., Mohammadi A.H., Zendehboudi S. Evaluating the Unloading Gradient Pressure in Continuous Gas-lift Systems During Petroleum Production Operations. *Petroleum Science and Technology* 32 (2014) 2961-2968.

Karmakar R., Samanta A. Solvation Dynamics of Coumarin-153 in a Room-Temperature Ionic Liquid. *Journal of Physical Chemistry A* 106 (2002) 4447-4452.

Katritzky A.R., Jain R., Lomaka A., Petrukhin R., Maran U., Karelson M. Perspective on the Relationship between Melting Points and Chemical Structure. *Crystal Growth and Design* 1 (2001) 261–265.

Kauwe S.K., Graser J., Vazquez A., Sparks T.D. Machine Learning Prediction of Heat Capacity for Solid Inorganics. *Integrating Materials and Manufacturing Innovation* 7 (2018) 43–51.

Keshavarz M.H., Pouretedal H.R., Saberi E. A New Method for Predicting Decomposition Temperature of Imidazolium-based Energetic Ionic Liquids. *Zeitschrift für anorganische und allgemeine Chemie* 643 (2016a) 171-179.

Keshavarz M.H., Pouretedal H.R., Saberi E. A simple method for prediction of density of ionic liquids through their molecular structure. *Journal of Molecular Liquids* 216 (2016b) 732-737.

Keshavarz M.H., Pouretedal H.R., Saberi E. A novel method for predicting melting point of ionic liquids. *Process Safety and Environmental Protection* 116 (2018) 333-339.

Kolská Z., Kukul J., Zábranský M., Růžicka V. Estimation of the heat capacity of organic liquids as a function of temperature by a three-level group contribution method. *Industrial and Engineering Chemistry Research* 47 (2008) 2075–2085.

Lazzús J.A. Estimation of Density as a Function of Temperature and Pressure for Imidazolium-Based Ionic Liquids Using a Multilayer Net with Particle Swarm Optimization. *International Journal of Thermophysics* 30 (2009a) 884-909.

- Lazzús J.A. ρ - T - P prediction for ionic liquids using neural networks. *Journal of the Taiwan Institute of Chemical Engineers* 40 (2009b) 213-232.
- Lazzús J.A. A group contribution method to predict ρ - T - P of ionic liquids. *Chemical Engineering Communications* 197 (2010a) 974-1015.
- Lazzús J.A. Prediction of Flash Point Temperature of Organic Compounds Using a Hybrid Method of Group Contribution + Neural Network + Particle Swarm Optimization. *Chinese Journal of Chemical Engineering* 18 (2010b) 817-823.
- Lazzús J.A. A group contribution method to predict the glass transition temperature of ionic liquids. *Thermochimica Acta* 528 (2012a) 38-44.
- Lazzús J.A. A group contribution method to predict the melting point of ionic liquids. *Fluid Phase Equilibria* 313 (2012b) 1-6.
- Lazzús J.A. A group contribution method to predict the thermal decomposition temperature of ionic liquids. *Journal of Molecular Liquids* 168 (2012c) 87-93.
- Lazzús J.A. A group contribution method to predict the thermal conductivity $\lambda(T,P)$ of ionic liquids. *Fluid Phase Equilibria* 405 (2015) 141-149.
- Lazzús J.A. A Group Contribution Method for Predicting the Freezing Point of Ionic Liquids. *Periodica Polytechnica Chemical Engineering* 60 (2016a) 273-281.
- Lazzús J.A. Estimation of the thermal conductivity $\lambda(T,P)$ of ionic liquids using a neural network optimized with genetic algorithms. *Comptes Rendus Chimie* 19 (2016b) 333-341.
- Lazzús J.A., Pulgar-Villarroel G. A group contribution method to estimate the viscosity of ionic liquids at different temperatures. *Journal of Molecular Liquids* 209 (2015) 161-168.
- Lazzús J.A., Cuturrufo F., Pulgar-Villarroel G., Salfate I., Vega P. Estimating the Temperature-Dependent Surface Tension of Ionic Liquids Using a Neural Network-Based Group Contribution Method. *Industrial and Engineering Chemistry Research* 56 (2017a) 6869-6886.
- Lazzús J.A., Pulgar-Villarroel G., Cuturrufo F., Vega P. Development of a group contribution method for estimating surface tension of ionic liquids over a wide range of temperatures. *Journal of Molecular Liquids* 240 (2017b) 522-531.
- Lee B.-S., Lin S.-T. Prediction and screening of solubility of pharmaceuticals in single- and mixed-ionic liquids using COSMO-SAC model. *American Institute of Chemical Engineers Journal* 63 (2017) 3096-3104.

Lei Z., Dai C., Zhu J., Chen B. Extractive distillation with ionic liquids: A review. *American Institute of Chemical Engineers Journal* 60 (2014) 3312-3329.

Lentini S., Galloni P., Garcia-Bosch I., Costas M., Conte V. Ionic liquids as reaction media in catalytic oxidations with manganese and iron pyridyl triazacyclononane complexes. *Inorganica Chimica Acta* 410 (2014) 60-64.

Liu Z., Meng X., Zhang R., Xu C., Dong H., Hu Y. Reaction performance of isobutane alkylation catalyzed by a composite ionic liquid at a short contact time. *American Institute of Chemical Engineers Journal* 60 (2014) 2244-2253.

Longo G.A., Mancin S., Righetti G., Zilio C., Ceccato R., Salmaso L. Machine learning approach for predicting refrigerant two-phase pressure drop inside Braze Plate Heat Exchangers (BPHE). *International Journal of Heat and Mass Transfer* 163 (2020) 120450.

Luria M., Benson S.W. Heat capacities of liquid hydrocarbons. Estimation of heat capacities at constant pressure as a temperature function, using additivity rules. *Journal of Chemical and Engineering Data* 22 (1977) 90–100.

Lydersen A. L. Estimation of Critical Properties of Organic Compounds. Report 3. University of Wisconsin. College of Engineering. Engineering Experimental Station. Madison (1955).

Lyman T.J., Danner R.P. Correlation of liquid heat capacities with a four-parameter corresponding states method. *AIChE Journal* 22 (1976) 759–765.

Mann V., Brito K., Gani R., Venkatasubramanian V. Hybrid, Interpretable Machine Learning for Thermodynamic Property Estimation using Grammar2vec for Molecular Representation. *Fluid Phase Equilibria* 561 (2022) 113531.

Marrero J., Gani R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria* 183 (2001) 183–208.

Marsousi S., Karimi-Sabet J., Moosavian M.A., Amini Y. Liquid-liquid extraction of calcium using ionic liquids in spiral microfluidics. *Chemical Engineering Journal* 356 (2019) 492-505.

Matsuda H., Yamamoto H., Kurihara K., Tochigi K. 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.

- Meksi N., Moussa A. A review of progress in the ecological application of ionic liquids in textile processes. *Journal of Cleaner Production* 161 (2017) 105-126.
- Mendonça A.C.F., Dörr N., Pádua A.A.H. Predicting thermophysical properties of ionic liquids as a function of temperature and pressure. *Journal of Engineering Tribology* 226 (2012) 965-976.
- Mital D.K., Nancarrow P., Zeinab S., Jabbar N.A., Ibrahim T.H., Khamis M.I., Taha A. Group Contribution Estimation of Ionic Liquid Melting Points: Critical Evaluation and Refinement of Existing Models. *Molecules* 26 (2021) 2454.
- Mohanraj M., Jayaraj S., Muraleedharan C. Applications of artificial neural networks for refrigeration, air-conditioning and heat pump systems—A review. *Renewable and Sustainable Energy Reviews* 16 (2012) 1340–1358.
- Mokadem K., Korichi M., Tumba K. A new group-interaction contribution method to predict the thermal decomposition temperature of ionic liquids. *Chemometrics and Intelligent Laboratory Systems* 157 (2016) 189-195.
- Mondejar M.E., Cignitti S., Abildskov J., Woodley J.M., Haglind F. Prediction of properties of new halogenated olefins using two group contribution approaches. *Fluid Phase Equilibria* 433 (2017) 79–96.
- Mousavi S.P., Atashrouz S., Amar M.N., Hadavimoghaddam F., Mohammadi M.-R., Hemmati-Sarapardeh A., Mohaddespour A. Modeling surface tension of ionic liquids by chemical structure-intelligence based models. *Journal of Molecular Liquids* 342 (2021a) 116961.
- Mousavi S.P., Atashrouz S., Amar M.N., Hemmati-Sarapardeh A., Mohaddespour A., Mosavi A. Viscosity of Ionic Liquids: Application of the Eyring's Theory and a Committee Machine Intelligent System. *Molecules* 26 (2021b) 156.
- Mousavi S.P., Atashrouz S., Rezaei F., Peyvastegan M.-E., Hemmati-Sarapardeh A., Mohaddespour A. Modeling thermal conductivity of ionic liquids: A comparison between chemical structure and thermodynamic properties-based models. *Journal of Molecular Liquids* 322 (2021c) 114911.
- Mowbray M., Vallerio M., Perez-Galvan C., Zhang D., Chanona A. D. R., Navarro-Brull F. J. Industrial data science—a review of machine learning applications for chemical and process industries. *Reaction Chemistry and Engineering* 7 (2022) 1471-1509.

Müller K., Albert J. Contribution of the individual ions to the heat capacity of ionic liquids. *Industrial and Engineering Chemistry Research* 53 (2014) 10343-10346.

Nakhaei-Kohani R., Atashrouz S., Hadavimoghaddam F., Abedi A., Jabbour K., Hemmati-Sarapardeh A., Mohaddespour A. Chemical structure and thermodynamic properties based models for estimating nitrous oxide solubility in ionic Liquids: Equations of state and Machine learning approaches. *Journal of Molecular Liquids* 367 (2022) 120445.

Nancarrow P., Lewis M., AbouChacra L. Group Contribution Methods for Estimation of Ionic Liquid Heat Capacities: Critical Evaluation and Extension. *Chemical Engineering and Technology* 38 (2015) 632-644.

Nannoolal Y., Rarey J., Ramjugernath D. Estimation of pure component properties: Part 2. Estimation of critical property data by group contribution. *Fluid Phase Equilibria* 252 (2007) 1-27.

Neng G., Xuehui W., Yongmei X., Guangming C. An artificial neural network for the residual isobaric heat capacity of liquid HFC and HFO refrigerants. *International Journal of Refrigeration* 98 (2019) 381–387.

Noushabadi A.S., Lay E.N., Dashti A., Mohammadi A.H., Chofreh A.G., Goni F.A., Klemeš J.J. Insights into modelling and evaluation of thermodynamic and transport properties of refrigerants using machine-learning methods. *Energy* 262 (2023) 125099.

Ortiz-Martínez V.M., Salar-García M.J., Hernández-Fernández F.J., Olivares-Carrillo P., de los Ríos A.P., Quesada-Medina J. Ionic liquids in supercritical methanol greatly enhance transesterification reaction for high-yield biodiesel production. *American Institute of Chemical Engineers Journal* 62 (2016) 3842-3846.

Oster K., Jacquemin J., Hardacre C., Ribeiro A.P.C., Elsinawi A. Further development of the predictive models for physical properties of pure ionic liquids: Thermal conductivity and heat capacity. *Journal of Chemical Thermodynamics* 118 (2018) 1-15.

Paduszyński K., Domańska U. A New Group Contribution Method For Prediction of Density of Pure Ionic Liquids over a Wide Range of Temperature and Pressure. *Industrial and Engineering Chemistry Research* 51 (2012) 591-604.

Paduszyński K., Domańska U. Viscosity of Ionic Liquids: An Extensive Database and a New Group Contribution Model Based on a Feed-Forward Artificial Neural Network. *Journal of Chemical Information and Modeling* 54 (2014) 1311-1324.

Palmer D.S., O'Boyle N.M., Glen R.C., Mitchell J.B.O. Random Forest Models To Predict Aqueous Solubility. *Journal of Chemical Information and Modeling* 47 (2007) 150–158.

Pence I., Yıldırım R., Cesmeli M.S., Güngör A., Akyüz A. Evaluation of machine learning approaches for estimating thermodynamic properties of new generation refrigerant R513A. *Sustainable Energy Technologies and Assessments* 55 (2023) 102973.

Perry R.H., Green D.W., Perry's M.J.O. *Chemical Engineers' Handbook*. McGraw- Hill (1997).

Poling B.E., Prausnitz J.M., O'Connell J.P. *The Properties of Gases and Liquids*. McGraw-Hill (2001).

Prasad T.E.V., Rajiah A., Prasad D.H.L. On the dependence of liquid heat capacity on temperature and molecular structure. *Chemical Engineering Journal* 52 (1993) 31–35.

Predel T., Pohrer B., Schlücker E. Ionic Liquids as Alternative Lubricants for Special Applications. *Chemical Engineering and Technology* 33 (2009) 132-136.

Qian X., Yang R. Machine learning for predicting thermal transport properties of solids. *Materials Science and Engineering: R: Reports*. 146 (2021) 100642.

Qiao Y., Ma Y., Huo Y., Ma P., Xia S. A group contribution method to estimate the densities of ionic liquids. *Journal of Chemical Thermodynamics* 42 (2010) 852-855.

Qi J., Zhang Q., Han X., Wu Q., Li Y., Li Q. Vapor-liquid equilibrium experiment and process simulation of extractive distillation for separating diisopropyl ether-isopropyl alcohol using ionic liquid. *Journal of Molecular Liquids* 293 (2019) 111406.

Reid R.C., Prausnitz J.M., Sherwood T.K. *The Properties of Gases and Liquids*. McGraw-Hill. New York (1987).

Rogers D., Hopfinger A.J. Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *Journal of Chemical Information and Computer Sciences* 34 (1994) 854-866.

Rooney D., Jacquemin J., Gardas R. Thermophysical Properties of Ionic Liquids. *Topics in Current Chemistry* 290 (2010) 185-212.

Rostami A., Hemmati-Sarapardeh A., Karkevandi-Talkhooncheh A., Husein M.M., Shamshirband S., Rabczuk T. Modeling heat capacity of ionic liquids using group method of data handling: A hybrid and structure-based approach. *International Journal of Heat and Mass Transfer* 129 (2019) 7-17.

Růžicka Jr.V., Domalski E.S. Estimation of the Heat Capacities of Organic Liquids as a Function of Temperature using Group Additivity. I. Hydrocarbon Compounds. *Journal of Physical and Chemical Reference Data* 22 (1993a) 597-618.

Růžicka Jr.V., Domalski E.S. Estimation of the Heat Capacities of Organic Liquids as a Function of Temperature using Group Additivity. II. Compounds of Carbon, Hydrogen, Halogens, Nitrogen, Oxygen, and Sulfur. *Journal of Physical and Chemical Reference Data* 22 (1993b) 619-657.

Saldana D.A., Starck L., Mougin P., Rousseau B., Ferrando N., Creton B. Prediction of Density and Viscosity of Biofuel Compounds Using Machine Learning Methods. *Energy Fuels* 26 (2012) 2416–2426.

Santak P., Conduit G. Predicting physical properties of alkanes with neural networks. *Fluid Phase Equilibria* 501 (2019) 112259.

Sattari M., Gharagheizi F., Ilani-Kashkouli P., Mohammadi A.H., Ramjugernath D. Development of a group contribution method for the estimation of heat capacities of ionic liquids. *Journal of Thermal Analysis and Calorimetry* 115 (2014a) 1863-1882.

Sattari M., Kamari A., Mohammadi A.H., Ramjugernath D. A group contribution method for estimating the refractive indices of ionic liquids. *Journal of Molecular Liquids* 200 (2014b) 410-415.

Sattari M., Kamari A., Hashemi H., Mohammadi A.H., Ramjugernath D. A group contribution model for prediction of the viscosity with temperature dependency for fluorine-containing ionic liquids. *Journal of Fluorine Chemistry* 186 (2016a) 19-27.

Sattari M., Kamari A., Mohammadi A.H., Ramjugernath D. On the prediction of critical temperatures of ionic liquids: Model development and evaluation. *Fluid Phase Equilibria* 411 (2016b) 24-32.

Schmidt J., Shi J., Borlido P., Chen L., Botti S., Marques M.A.L. Predicting the Thermodynamic Stability of Solids Combining Density Functional Theory and Machine Learning. *Chemistry of Materials* 29 (2017) 5090–5103.

Seko A., Maekawa T., Tsuda K., Tanaka I. Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids. *Physical Review B* 89 (2014) 054303.

Sepehri B. A review on created QSPR models for predicting ionic liquids properties and their reliability from chemometric point of view. *Journal of Molecular Liquids* 297 (2020) 112013.

Shahriari S., Atashrouz S., Pazuki G. Mathematical Model of the Phase Diagrams of Ionic Liquids-Based Aqueous Two-Phase Systems Using the Group Method of Data Handling and Artificial Neural Networks. *Theoretical Foundations of Chemical Engineering* 52 (2018) 146–155.

Shahsavari S., Mesbah M., Soroush E., Farhangian H., Alizadeh S., Soltanali S. A simple group contribution correlation for modeling the surface tension of pure ionic liquids. *Journal of Molecular Liquids* 265 (2018) 292-298.

Sharma A., Jain A., Gupta P., Chowdary V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *Institute of Electrical and Electronics Engineers Access* 9 (2020) 4843–4873.

Shen C., Li C., Li X., Lu Y., Muhammad Y. Estimation of densities of ionic liquids using Patel–Teja equation of state and critical properties determined from group contribution method. *Chemical Engineering Science* 66 (2011) 2690-2698.

Sinha V., Laan J.J., Pidko E.A. Accurate and rapid prediction of pKa of transition metal complexes: semiempirical quantum chemistry with a data-augmented approach. *Physical Chemistry Chemical Physics* 23 (2021) 2557-2567.

Soares B.G. Ionic liquid: A smart approach for developing conducting polymer composites: A review. *Journal of Molecular Liquids* 262 (2018) 8-18.

Soares L.H., Guirardello R., Rolemberg M. A Simple Group Contribution Model to Predict Thermal Conductivity of Pure Ionic Liquids. *Chemical Engineering Transactions* 74 (2019) 1195-1200.

Soltanian M.R., Bemani A., Moeini F., Ershadnia R., Yang Z., Du Z., Yin H., Dai Z. Data driven simulations for accurately predicting thermodynamic properties of H₂ during geological storage. *Fuel* 362 (2024) 130768.

Sommer J., Bromberger B., Robben C., Kalb R., Rossmanith P., Mester P.-J. Liquid-liquid extraction of viral particles with ionic liquids. *Separation and Purification Technology* 254 (2021) 117591.

Soriano A.N., Agapito A.M., Lagumbay L.J.L.I., Caparanga A.R., Li M.-H. A simple approach to predict molar heat capacity of ionic liquids using group-additivity method. *Journal of the Taiwan Institute of Chemical Engineers* 41 (2010a) 307-314.

Soriano A.N., Doma Jr. B.T., Li M.-H. Density and refractive index measurements of 1-ethyl-3-methylimidazolium-based ionic liquids. *Journal of the Taiwan Institute of Chemical Engineers* 41 (2010b) 115-121.

Sosa J.E., Araújo J.M.M., Amado-González E., Pereiro A.B. Separation of azeotropic mixtures using protic ionic liquids as extraction solvents. *Journal of Molecular Liquids* 297 (2020) 111733.

Součková M., Klomfar J., Pátek J. Surface tension and 0.1 MPa density data for 1-C_n-3-methylimidazolium iodides with n = 3, 4, and 6, validated using a parachor and group contribution model. *Journal of Chemical Thermodynamics* 83 (2015) 52-60.

Sun J., Wu J., Song T., Hu L., Shan K., Chen G. Alternative Approach to Chemical Accuracy: A Neural Networks-Based First-Principles Method for Heat of Formation of Molecules Made of H, C, N, O, F, S, and Cl. *The Journal of Physical Chemistry A* 118 (2014) 9120–9131.

Su W., Zhao L., Deng S. Group contribution methods in thermodynamic cycles: Physical properties estimation of pure working fluids. *Renewable and Sustainable Energy Reviews* 79 (2017) 984-1001.

Şahin A.Ş., Köse İ.İ., Selbaş R. Comparative analysis of neural network and neuro-fuzzy system for thermodynamic properties of refrigerants. *Applied Artificial Intelligence* 26 (2012) 662–672.

Tahami S., Movagharnejad K., Ghasemitabar H. Estimation of the critical constants of organic compounds via a new group contribution method. *Fluid Phase Equilibria* 494 (2019) 45-60.

Taherifard H., Raeissi S. Estimation of the Densities of Ionic Liquids Using a Group Contribution Method. *Journal of Chemical and Engineering Data* 61 (2016) 4031-4038.

Taherzadeh M., Haghbakhsh R., Duarte A.R.C., Raeissi S. Estimation of the heat capacities of deep eutectic solvents. *Journal of Molecular Liquids* 307 (2020) 112940.

Tarakad R.R., Danner R.P. An improved corresponding states method for liquid heat capacities. *AIChE Journal* 23 (1977) 944–946.

Thoms E., Sippel P., Reuter D., Weiß M., Loidl A., Krohns S. Dielectric study on mixtures of ionic liquids. *Scientific Reports* 7 (2017) 7463.

Trinh C., Meimaroglou D., Hoppe S. Machine Learning in Chemical Product Engineering: The State of the Art and a Guide for Newcomers. *Processes* 9 (2021) 1456.

Valderrama J.O., Abu-Sharkh B.F. Generalized rackett-type correlations to predict the density of saturated liquids and petroleum fractions. *Fluid Phase Equilibria* 51 (1989) 87-100.

Valderrama J.O., Cardona L.F. Predicting the melting temperature and the heat of melting of ionic liquids. *Journal of Ionic Liquids* 1 (2021) 100002.

Valderrama J.O., Robles P.A. Critical Properties, Normal Boiling Temperatures, and Acentric Factors of Fifty Ionic Liquids. *Industrial and Engineering Chemistry Research* 46 (2007) 1338-1344.

Valderrama J.O., Rojas R.E. Critical Properties of Ionic Liquids. Revisited. *Industrial and Engineering Chemistry Research* 48 (2009) 6890-6900.

Valderrama J.O., Rojas R.E. Mass connectivity index, a new molecular parameter for the estimation of ionic liquid properties. *Fluid Phase Equilibria* 297 (2010) 107-112.

Valderrama J.O., Venegas V.H.Á. Método modificado Lydersen-Joback-Reid para estimar propiedades críticas de biomoléculas. *Alimentaria: Revista de tecnología e higiene de los alimentos* 354 (2004) 55-66.

Valderrama J.O., Zarricueta K. A simple and generalized model for predicting the density of ionic liquids. *Fluid Phase Equilibria* 275 (2009) 145-151.

Valderrama J.O., Sanga W.W., Lazzús J.A. Critical Properties, Normal Boiling Temperature, and Acentric Factor of Another 200 Ionic Liquids. *Industrial and Engineering Chemistry Research* 47 (2008) 1318-1330.

Valderrama J.O., Reátegui A., Rojas R.E. Density of Ionic Liquids Using Group Contribution and Artificial Neural Networks. *Industrial and Engineering Chemistry Research* 48 (2009) 3254-3259.

Valderrama J.O., Martinez G., Faúndez C.A. Heat Capacity of Ionic Liquids Using Artificial Neural Networks and the Concept of Mass Connectivity. *International Journal of Thermophysics* 32 (2011a) 942-956.

Valderrama J.O., Martinez G., Rojas R.E. Predictive model for the heat capacity of ionic liquids using the mass connectivity index. *Thermochimica Acta* 513 (2011b) 83-87.

Valderrama J.O., Toro A., Rojas R.E. Prediction of the heat capacity of ionic liquids using the mass connectivity index and a group contribution method. *Journal of Chemical Thermodynamics* 43 (2011c) 1068-1073.

Valderrama J.O., Forero L.A., Rojas R.E. Critical Properties and Normal Boiling Temperature of Ionic Liquids. Update and a New Consistency Test. *Industrial and Engineering Chemistry Research* 51 (2012) 7838-7844.

Valderrama J.O., Faúndez C.A., Vicencio V.J. Artificial Neural Networks and the Melting Temperature of Ionic Liquids. *Industrial and Engineering Chemistry Research* 53 (2014) 10504-10511.

Valderrama J.O., Forero L.A., Rojas R.E. Extension of a Group Contribution Method To Estimate the Critical Properties of Ionic Liquids of High Molecular Mass. *Industrial and Engineering Chemistry Research* 54 (2015) 3480-3487.

Valderrama J.O., Campusano R.A., Rojas R.E. Glass transition temperature of ionic liquids using molecular descriptors and artificial neural networks. *Comptes Rendus Chimie* 20 (2017) 573-584.

Valderrama J.O., Forero L.A., Rojas R.E. Critical Properties of Metal-Containing Ionic Liquids. *Industrial and Engineering Chemistry Research* 58 (2019) 7332-7340.

Valderrama J.O., Cardona L.F., Rojas R.E. A Simple Computer Tool for Simultaneously Estimating Critical, Transport, Physicochemical, and Phase Change Properties of Ionic Liquids. *Industrial and Engineering Chemistry Research* 60 (2021) 16143–16151.

Varamesh A., Hemmati-Sarapardeh A., Moraveji M.K., Mohammadi A.H. Generalized models for predicting the critical properties of pure chemical compounds. *Journal of Molecular Liquids* 240 (2017) 777–793.

Vega L.F., Vilaseca O., Llovel F., Andreu J.S. Modeling ionic liquids and the solubility of gases in them: Recent advances and perspectives. *Fluid Phase Equilibria* 294 (2010) 15-30.

Vekariya R.L. A review of ionic liquids: Applications towards catalytic organic transformations. *Journal of Molecular Liquids* 227 (2017) 44-60.

Vidyasagar S., Shabari S.B., Kumaraswamy J. Predicting the thermodynamic stability of perovskite oxides using multiple machine learning techniques. *Materials Today: Proceedings* 52 (2022) 457–461.

Wang H., Li Z., Liu Y., Zhang X., Zhang S. Degradation of poly(ethylene terephthalate) using ionic liquids. *Green Chemistry* 11 (2009) 1568-1575.

Wang J., Song Z., Chen L., Xu T., Deng L., Qi Z. Prediction of CO₂ solubility in deep eutectic solvents using random forest model based on COSMO-RS-derived descriptors. *Green Chemical Engineering* 2 (2021) 431–440.

Wang X., Chi Y., Mu T. A review on the transport properties of ionic liquids. *Journal of Molecular Liquids* 193 (2014) 262-266.

Wang X., Lu X., Zhou Q., Zhao Y., Li X., Zhang S. Database and new models based on a group contribution method to predict the refractive index of ionic liquids. *Physical Chemistry Chemical Physics* 19 (2017) 19967-19974.

Weiss V.C. Liquid–vapor equilibrium and critical parameters of the ionic liquid 1-butyl-3-methylimidazolium hexafluorophosphate from molecular dynamics simulations. *Journal of Molecular Liquids* 209 (2015) 745-752.

Welton T. Room-Temperature Ionic Liquids. *Solvents for Synthesis and Catalysis*. *Chemical Reviews* 99 (1999) 2071-2084.

Wu K.-J., Zhao C.-X., He C.-H. A simple corresponding-states group-contribution method for estimating surface tension of ionic liquids. *Fluid Phase Equilibria* 328 (2012) 42-48.

Wu K.-J., Zhao C.-X., He C.-H. Development of a group contribution method for determination of thermal conductivity of ionic liquids. *Fluid Phase Equilibria* 339 (2013) 10-14.

Wu K.-J., Chen Q.-L., He C.-H. Speed of sound of ionic liquids: Database, estimation, and its application for thermal conductivity prediction. *American Institute of Chemical Engineers Journal* 60 (2014) 1120-1131.

Wu K.-J., Luo H., Yang L. Structure-based model for prediction of electrical conductivity of pure ionic liquids. *American Institute of Chemical Engineers Journal* 62 (2016) 3751-3762.

Xia L., Pan Y., Zhao T., Sun X., Tao S., Chen Y., Xiang S. Estimating heat capacities of liquid organic compounds based on elements and chemical bonds contribution. *Chinese Journal of Chemical Engineering* 57 (2023) 30–38.

Xiong J., Zhu W., Li H., Xu Y., Jiang W., Xun S., Liu H., Zhao Z. Immobilized fenton-like ionic liquid: Catalytic performance for oxidative desulfurization. *American Institute of Chemical Engineers Journal* 59 (2013) 4696-4704.

Xu P., Du P.-X., Zong M.-H., Li N., Lou W.-Y. Combination of deep eutectic solvent and ionic liquid to improve biocatalytic reduction of 2-octanone with *Acetobacter pasteurianus* GIM1.158 cell. *Scientific Reports* 6 (2016) 26158.

Yadav J.S., Reddy B.V.S., Basak A.K., Narsaiah A.V. [Bmim]PF₆ and BF₄ ionic liquids as novel and recyclable reaction media for aromatic amination. *Tetrahedron Letters* 44 (2003) 2217-2220.

Yalamanchi K.K., van Oudenhoven V.C.O., Tutino F., Monge-Palacios M., Alshehri A., Gao X., Sarathy S.M. Machine Learning To Predict Standard Enthalpy of Formation of Hydrocarbons. *The Journal of Physical Chemistry A* 123 (2019) 8305–8313.

Ye C., Liu W., Chen Y., Yu L. Room-temperature ionic liquids: a novel versatile lubricant. *Chemical Communications* (2001) 2244-2245.

Ye C., Shreeve J.M. Rapid and Accurate Estimation of Densities of Room-Temperature Ionic Liquids and Salts. *Journal of Physical Chemistry A* 111 (2007) 1456-1461.

Yusuf F., Olayiwola T., Afagwu C. Application of Artificial Intelligence-based predictive methods in Ionic liquid studies: A review. *Fluid Phase Equilibria* 531 (2021) 112898.

Zaid H.F.M., Kait C.F., Mutalib M.I.A. Extractive deep desulfurization of diesel using choline chloride-glycerol eutectic-based ionic liquid as a green solvent. *Fuel* 192 (2017) 10-17.

Zábranský M., Růžička Jr. V. Estimation of the heat capacities of organic liquids as a function of temperature using group additivity: an amendment. *Journal of Physical and Chemical Reference Data* 33 (2004) 1071–1081.

Zendehboudi S., Rezaei N., Lohi A. Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. *Applied Energy* 228 (2018) 2539–2566.

Zhang F., O'Donnell L.J. Chapter 7 – Support vector regression. Machine Learning. Methods and Applications to Brain Disorders (2020) 123–140.

Zhang G., Hu M.Y., Patuwo B.E., Indro D.C. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. European Journal of Operational Research 116 (1999) 16-32.

Zhang Z., Li H., Chang H., Pan Z., Luo X. Machine learning predictive framework for CO₂ thermodynamic properties in solution. Journal of CO₂ Utilization 26 (2018) 152–159.

Zhou Y., Lin Z., Wu K., Xu G., He C. A Group Contribution Method for the Correlation of Static Dielectric Constant of Ionic Liquids. Chinese Journal of Chemical Engineering 22 (2014) 79-88.

Zhu C., Liu X., He M., Kontogeorgis G.M., Liang X. Heat capacities of fluids: the performance of various equations of state. Journal of Chemical and Engineering Data 65 (2020) 5654–5676.

Appendix A

Figure A.1 and Table A.1 exemplify the applications of ILs in the development of novel processes and products, thus contributing to the industrial sustainability and circular economy.

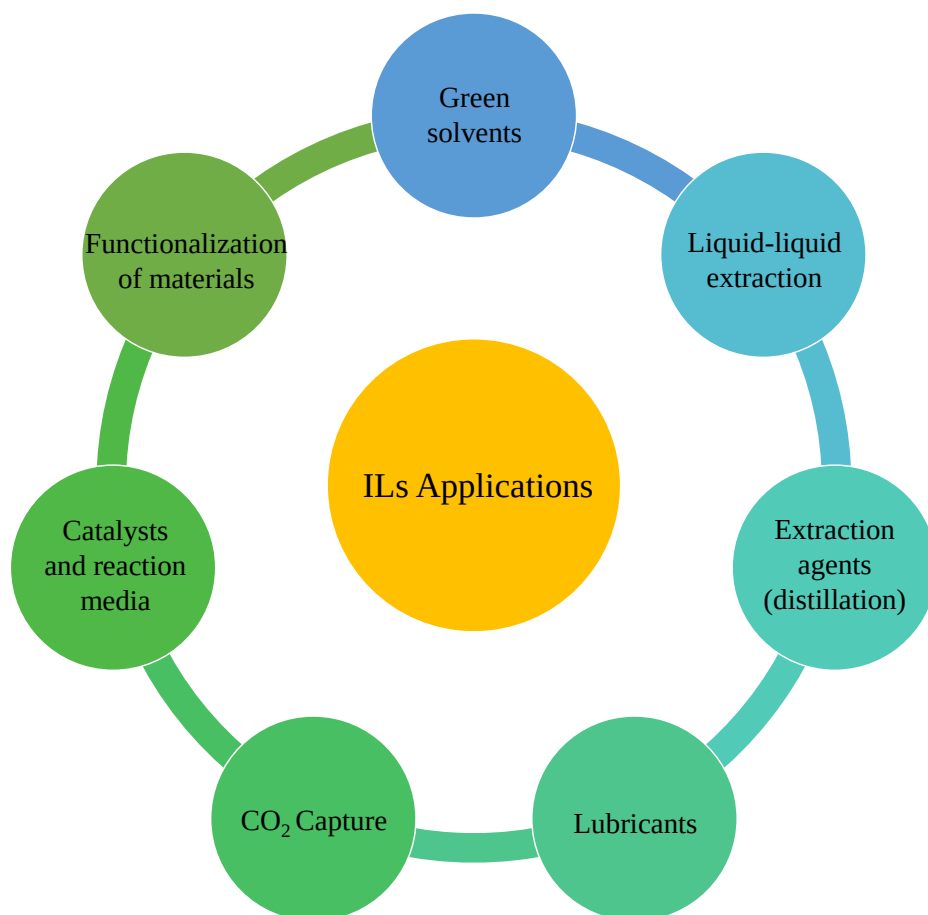


Figure A.1 Applications of ILs in Process Systems Engineering.

Table A.1 Different applications of ILs for the development of processes and products.

Application	Remarks	Reference
Solvents		
Reactive extraction of propionic acid.	This study indicates that the ILs 1-hexyl-3-methylimidazolium hexafluorophosphate and 1-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide can be successfully used as green extraction solvents for this process.	Ayan et al. (2020)
Process to transform α -pinene into α -terpineol.	IL allows to avoid the use of a new catalyst for each reaction. The benefits of using IL include an easy product recovery and it is not necessary to neutralize the product.	Aguilar-Elguezabal et al. (2020)
Extraction agents in distillation processes		
Production of anhydrous ethanol using an IL as solvent in an extractive distillation process.	The modeling and simulation of this process showed that a high purity and recovery of ethanol can be obtained with an easy of IL recovery.	Figuerola et al. (2012)
Extractive distillation of the azeotropic mixture of diisopropyl ether and isopropanol.	The application of IL improved the relative volatility of the compounds. Higher purities of the separation system were obtained in comparison to those obtained with a conventional solvent based on ethylene glycol.	Qi et al. (2019)
Liquid-liquid extraction processes		
Separation of hydrocarbons from their azeotropic mixtures containing ethanol via the application of an IL as extraction solvent.	This process was effective for the purification of hydrocarbons. In comparison to azeotropic or extractive distillation, this process is more environmentally friendly.	Sosa et al. (2020)
Directional Solvent Extraction with an IL for the seawater desalination	The application of IL allowed to obtain a higher water yield than that generated with the current process based on decanoic acid. It was also observed a reduction of the energy and exergy consumptions.	Guo et al. (2021)

Table A.1 (continuation).

<i>Application</i>	<i>Remarks</i>	<i>Reference</i>
Liquid-liquid extraction processes		
Extraction of viral particles	4 ILs showed better extraction behavior compared to currently available protocols for this extraction system. The application of IL is a time-saving and easy method compared to current protocols. Depending on the IL, it is potentially more economical than commonly used protocols.	Sommer et al. (2021)
Catalysis and reaction engineering		
Esterification of several carboxylate sodium salts with different alkyl halides.	ILs were used as effective reusable reaction media. The procedure was environmentally friendly compared to existing methods based on toxic solvents.	Brinchi et al. (2003)
Alkylation of isobutane with 2-butene catalyzed by a composite IL.	The composite IL showed an effective catalytic performance in a short contact time.	Liu et al. (2014)
CO₂ capture		
CO ₂ capture with hybrid imidazolium-based ILs and amine absorbents	The amine improved the CO ₂ absorption performance of ILs in terms of viscosity, CO ₂ loading and regeneration efficiency.	Xiao et al. (2019)
Development and evaluation of novel phase-change solvents based on ILs and other solvents for the CO ₂ capture.	The phase-change system showed better absorption performance than the aqueous solution of amines. It was identified an improvement of the regeneration characteristics.	Zhou et al. (2020)

Table A.1 (continuation).

<i>Application</i>	<i>Remarks</i>	<i>Reference</i>
Functionalization of materials		
Functionalization of chitosan using different ILs for the adsorption of hexavalent chromium.	This functionalized adsorbent showed better adsorption properties than raw chitosan due to the increment of electrostatic interactions between amino groups and hexavalent chromium. The adsorbent showed good reusability properties.	Elidório et al. (2020)
Synthesis and functionalization of multi-walled carbon nanotubes using ILs.	The functionalized carbon nanotubes were thermally stable with a higher surface area where they can be applied in the adsorption of metal ions.	Matandabuzo and Ajibade (2018)

Appendix B

Figure B.1 provides an overview of thermodynamic models for the estimation of ILs physicochemical properties.

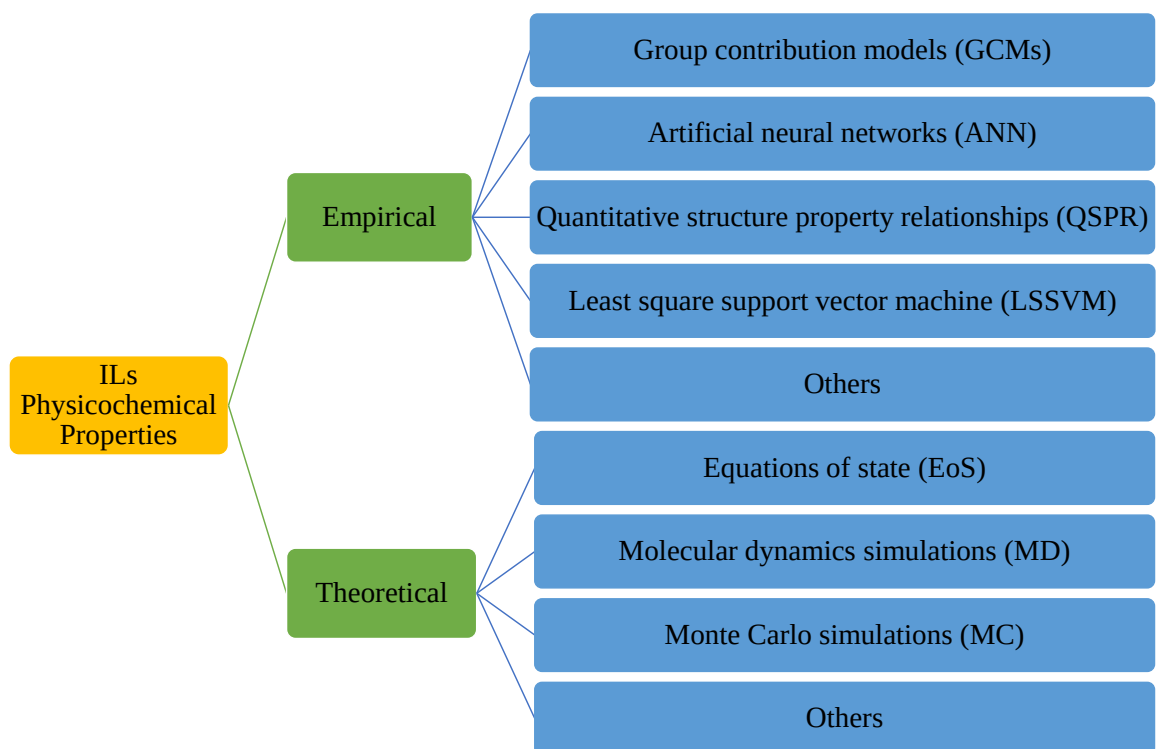


Figure B.1 Thermodynamic models for the estimation of ILs physicochemical properties.

Appendix C

Figure C.1 illustrates the concept of GCMs for the prediction of a given thermodynamic property using an IL molecule as example.

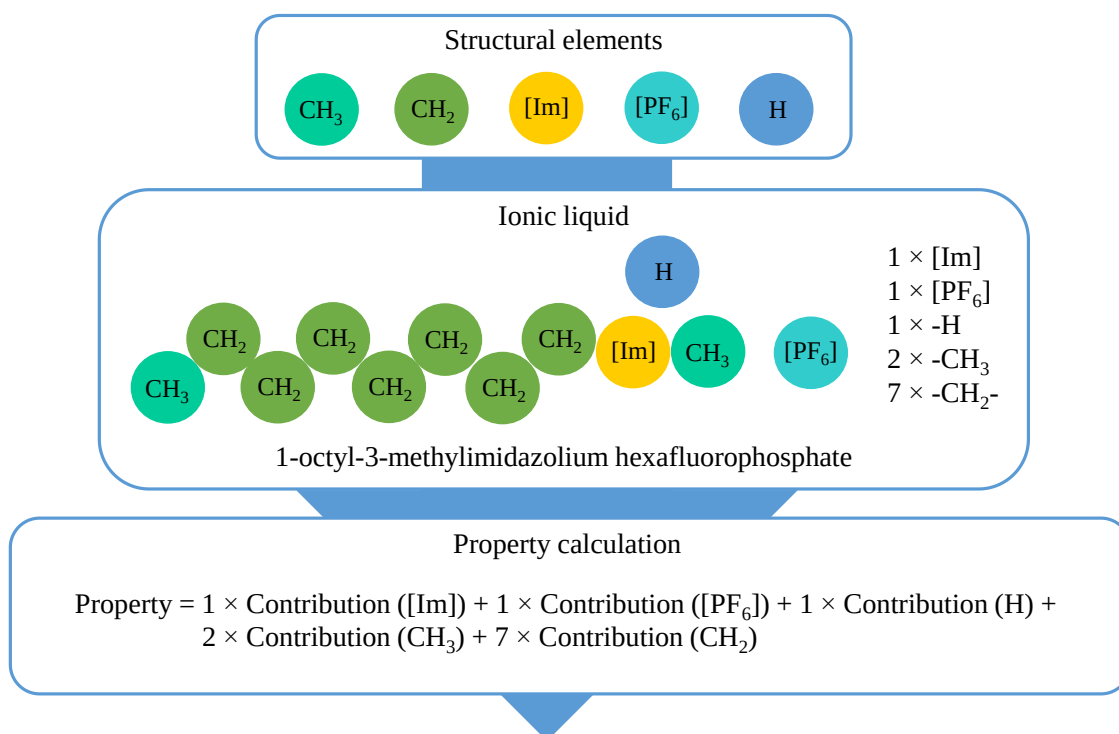


Figure C.1 Conceptualization of the calculation of a thermodynamic property using the concept of group contribution model.

Appendix D

Figure D.1 exemplifies the structural elements considered in the Meccano and Lego approaches.

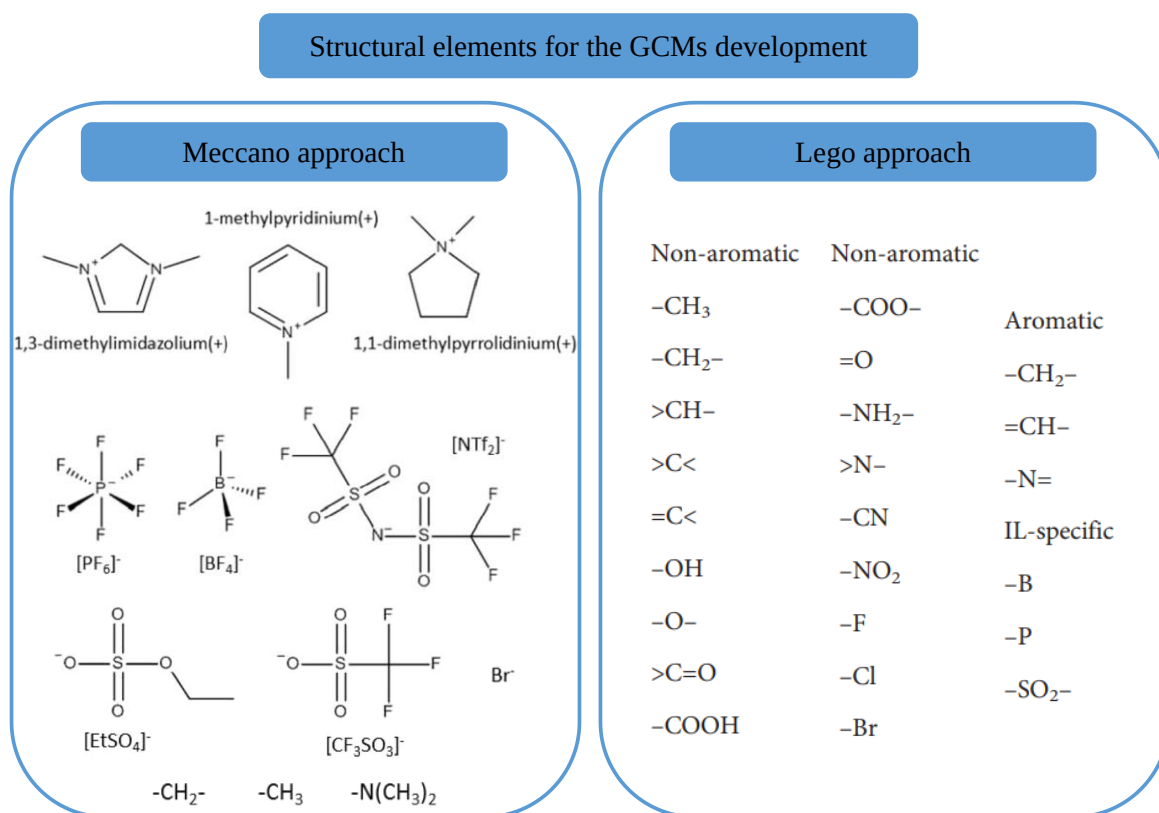


Figure D.1 Structural elements used in the Meccano and Lego approaches of GCMs.

Appendix E

The folder (Heat_capacity_calculation) contains a Matlab® file (*Cp_liquid.m*) for the estimation of liquid C_p of pure compounds using the best GCMs–ANNs model reported in the Chapter 3 of this thesis (i.e., the M_2 GCM-ANNs model with 1 hidden layer and 7 neurons). It is necessary to download the files (network.mat) and (inputs.xlsx) in the same folder of the file (*Cp_liquid.m*) to run the model written in Matlab® language.



Heat_capacity_calculation

The user should introduce the inputs (i.e., molecular weight, reduced temperature and frequency of each functional group that compose the molecule) in the Excel file (inputs.xlsx). Note that if any of the required values are not entered, an error will occur. The values of frequency of each functional group that compose the molecule are set by default as zero. The user should introduce the correct values in each case. The user must save the changes in the Excel file and then run the model written in Matlab® language. The result will be displayed in the Command Window of Matlab®.

Example. Calculation of liquid C_p of dicyclohexyl sulfide

Figure E.1 shows the molecular structure of dicyclohexyl sulfide and the functional groups that compose this molecule.

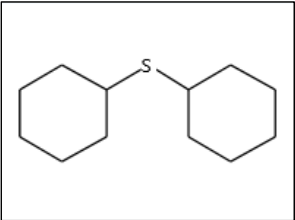
	Functional group	Frequency
	-S- (outside rings)	1
	-CH ₂ - (inside rings)	10
	>CH- (inside rings)	2

Figure E.1 Molecular structure of dicyclohexyl sulfide.

The input information requested by the Excel file is given below:

Molecular weight of the compound (g/mol): 198.37
Reduced temperature of the compound: 0.3691

Frequency of each functional group that compose the molecule:

-CH₃ (outside rings): 0

-CH₂- (outside rings): 0

>CH- (outside rings): 0

>NH (outside rings): 0

-NH₂ (outside rings): 0

-NO₂: 0

>C< (outside rings): 0

-F: 0

-Cl: 0

=CH₂ (outside rings): 0

=CH- (outside rings): 0

=C< (outside rings): 0

=C= (outside rings): 0

Triple bondC- (outside rings): 0

-OH (alcohol): 0

-O- (outside rings): 0

>C=O (outside rings): 0

-CHO (aldehyde): 0

-COOH (acid): 0

-COO- (ester): 0

=O: 0

>N- (outside rings): 0

-N= (outside rings): 0

-CN: 0

-Br: 0

-I: 0

Triple bondCH (outside rings): 0

-S- (outside rings): 1

-SH: 0

>Si< (outside rings): 0

=Pthree bonds: 0

-NO3: 0

>S=: 0

-SO2- (outside rings): 0

>Al-: 0

-CH2- (inside rings): 10

>CH- (inside rings): 2

=CH- (inside rings): 0

>C< (inside rings): 0

=C< (inside rings): 0

-O- (inside rings): 0

-OH (phenol): 0

>C=O (inside rings): 0

>NH (inside rings): 0

>N- (inside rings): 0

-N= (inside rings): 0

-S- (inside rings): 0

>Si< (inside rings): 0

-SO2- (inside rings): 0

The next result is obtained for this example:

The value of the heat capacity is: 3.109604e+05 J/kmol·K

The Excel file (Database_compounds.xlsx) contains all the compounds used in this study and the type and frequency of the functional groups that compose the molecules of tested compounds.



Database_compounds.xlsx

The Matlab[®] file (Weights_and_biases.mat) contains the values of model parameters (i.e., weights and biases) for the M₂ GCM-ANNs model to allow its standalone use.



Weights_and_biases.mat