

الجمهورية الجزائرية الديمقراطية الشعبية

People's Democratic Republic of Algeria

و البحث العلمي وزارة التعليم العالي

Ministry of Higher Education and Scientific Research

MUSTAPHA Stambouli University of

Mascara



جامعة مصطفى اسطembولي

معسكر

Faculty of Exact Sciences

Departement of chemistry

Laboratory of Organic Chemistry, Macromolecular and Materials (LCOMM)

DOCTORATE THESIS

Speciality :Innovative Materials and Green Chemistry

Title

Development of Predictive Models for the viscosity of
organic compounds vapours and liquids

Presented by : Boualem Affaf Djihed

19/11/2025

In front of the jury:

Chairwoman	HACHEMAOUI Aicha	Professor	University of Mascara-Mustapha Stambouli
Reviewer	GHERRAS Hammou	MCA	University of Mascara-Mustapha Stambouli
Reviewer	GUELLA Soufiane	Professor	University of Djillali liabes- Sidi BelAbbes
Guest	ARGOUB Kadda	MCA	University of Mascara-Mustapha Stambouli
Supervisor	Ali Mustapha Benkouider	Professor	University of Mascara

Academic year : 2025-2026

ACKNOWLEDGMENTS

Completing this PhD has been one of the most challenging and rewarding experiences of my life, and it would not have been possible without the support, guidance, and encouragement of many individuals.

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Ali Mustapha Benkouider, for their invaluable mentorship, insightful feedback, and unwavering support throughout the course of my research. Your guidance helped shape this thesis and sharpen my thinking, and I am truly thankful for the opportunity to learn under your supervision. Prof. Ali Mustapha Benkouider has also shown me the value of good leadership and management, in addition to academic considerations.

I would also like to thank my thesis committee and the faculty members in the Faculty of Exact Sciences at University Mustapha Stambouli of Mascara for their constructive feedback and for providing a stimulating academic environment.

To the jury I would also like to thank the professor Hachemaoui Aicha at the university of Mustapha Stambouli of Mascara for the honour of chairing the defence jury.

My sincere thanks also go to Mr Gherras Hamou Lecturer A at the university of Mustapha Stambouli of Mascara and Mr .Guella Sofiane professor at the university of Djillali Liabes of Sidi Bel Abbes, and Mr Argoub Kada Lecturer at the university of Mustapha Stambouli of Mascara for agreeing to examine this work.

To my colleagues and friends in the lab and department—thank you for the discussions, collaborations, and shared moments of frustration and triumph. Your camaraderie made this journey far more enjoyable.

To my family, your constant love and belief in me have been my foundation. Thank you for your patience, sacrifices, and encouragement during the difficult times. A special thanks to my parents and siblings for being my anchor and for understanding the countless hours dedicated to this work.

This thesis is not only the result of individual effort but also a collective journey, and I am profoundly grateful to everyone who was part of it.

TABLE OF CONTENTS

Acknowledgments.....	3
list of figures:	8
List of tables.....	10
General introduction.....	11
Chapter I: Theoretical concept and Bibliographical summary	14
I.1 Introduction	14
I.2 Viscosity definition	15
I.3 Viscosity measurement.....	17
I.3.1 Capillary viscometers	17
I.3.2 Ball or phere viscometers.....	19
I.3.3 Vibrational or oscillational viscometers.....	20
I.3.4 Rotational viscometers	23
I.3.5 Non conventional viscometers	24
I.3.6 Conclusion	25
I.4 Causes of errors in viscometers.....	25
I.5 Temperature and pressure dependence	26
I.6 Understanding the Differences in Gas and Liquid Behavior	27
I.7 A REVIEW OF PRACTICAL PREDICTIVE AND CORRELATIVE METHODS FOR ESTIMATING VISCOSITY	28
I.7.1 Semi- empirical models.....	28
I.7.1.A Corresponding states approaches.....	29

I.7.1.B	Residual Approach	32
I.7.1.B.1	Friction Theory (FT):	32
I.7.1.B.2	Free-Volume Theory (FVT):.....	33
I.7.1.B.3	Expanded Fluid Theory (EFT):	33
I.7.1.B.4	Scaling Methods:.....	34
I.7.1.C	Hard-sphere theory (Modified Enskog Theory)	34
I.7.1.D	Square-well models.....	35
I.7.2	Empirical models.....	36
I.7.2.A	Quantitative Structure-Property Relationships (QSPR).....	36
I.7.2.B	Group Contribution models.....	39
I.8	Conclusion	41
I.9	Bibliography	42
Chapitre II:	Methodology and Modelling methods.....	53
II.1	Introduction	53
II.2	Database importance.....	53
II.2.1	Data revision.....	54
II.3	Group contribution methods	56
II.3.1	Principle.....	56
II.3.2	Group's hierarchy.....	57
II.3.2.A	Zero-order fragmentation scheme.....	57
II.3.2.B	First-order fragmentation scheme	58
II.3.2.C	Second-order fragmentation method	58
II.3.2.D	Third-order fragmentation method	59
II.3.2.D.1	First-order groups	60
I.1.1.A.1	Second-order groups.....	61
II.3.2.D.2	Third-order groups.....	61
II.4	Quantitative Structures-Property Relationships	63
II.4.1	Principle.....	63
II.4.2	Molecular descriptors	63
II.4.2.A	Constitutional descriptors	64
II.4.2.B	Topological descriptors	64
II.4.2.C	Geometric descriptors.....	65
II.4.2.D	Quantum descriptors	66
II.4.3	Fingerprints in Predictive Modeling.....	67
II.4.3.A	Morgan fingerprints	67
II.4.3.B	Atom pairs fingerprints:	68

II.4.4	Dimensionality Reduction of Descriptors [84].....	69
II.4.4.A	Transformation Methods	69
II.4.4.B	Feature Selection	70
II.4.4.C	Sparse Regression Techniques.....	70
II.5	Models and regression.....	71
II.5.1	Conventional models	71
II.5.2	Machine learning algorithms.....	72
II.5.2.A	Random Forest.....	72
II.5.2.B	Extreme Gradient Boosting	73
II.5.2.C	Non-linear Support Vector Machine	76
II.5.2.D	Neural Network.....	78
II.6	Model validation.....	81
II.6.1	Statistical validation parameters	81
II.6.2	Cross-validation and external validation.....	83
II.7	Conclusion	85
II.8	Bibliography	86
Chapitre III:	prediction of the viscosity of ionic liquids	97
III.1	Introduction	97
III.2	Experimental database	98
III.3	methodological strategy for modelling.....	104
III.4	Results and discussion	106
III.4.1	Relative errors by cationic and anionic families	112
III.4.2	Relative errors according to cation-anion combinations	115
III.4.2.A	Performance by Cationic Families.....	116
III.4.2.B	Performance by Anionic Families.....	118
III.4.2.C	Synthesis of Findings	118
III.4.3	Distributions of relative prediction errors.....	119
III.4.4	Comparison of the proposed models with existing ones	122
III.5	Conclusion	125
III.6	Bibliography	126
Chapitre IV:	Viscosity prediction of non-ionic organic compounds	136
IV.1	Introduction	136
IV.2	Database: Data Collection and Revision	137
IV.3	methodological strategy for modelling.....	140
IV.4	Results and discussion	142
IV.4.1	Relative errors by chemical families	148

IV.4.2	Distributions of relative prediction errors.....	150
IV.5	Conclusion	154
	General conclusion	156

LIST OF FIGURES:

Figure 1 :	Gradient of velocity for a Newtonian fluid as discussed by Massey	16
Figure 2 :	Gradient of velocity for a non-Newtonian fluid as discussed by Massey	16
Figure 3 :	Example of a capillary viscometer with temperature control.....	18
Figure 4 :	Some types of glass capillaries: 1: Ostwald capillary, 2: Ubbelohde capillary, 3: Cannon-Fenske capillary, 4: Houillon capillary	19
Figure 5 :	Example of a falling ball viscometer with thermal regulation (phywe.fr).....	20
Figure 6 :	Diagram illustrating affecting forces on a falling ball; F_g gravity force, F_b upward buoyant and F_d force of viscous (internal) friction.....	20
Figure 7 :	Diagram of a vibrating wire viscometer.....	21
Figure 8 :	Diagram of a vibrating crystal and vibrating plate viscometer	22
Figure 9 :	Types of rotational viscometer; (a) Coaxial cylinder viscometer, (b) Cone-plane viscometer, (c) Planar viscometer	24
Figure 10 :	Effect of pressure on viscosity of selected organic liquids at room temperature [6].	27
Figure 11 :	Data pre-processing algorithm	55

Figure 12 : Example of the cutting of 2-methylnonane using a first-order cutting scheme	57
Figure 13 : Example of the cutting of 2-Methylbenzoic acid using a first-order cutting scheme	58
Figure 14 : Example of the cutting of 2-methylpropylbenzene using a second-order cutting scheme	59
Figure 15 : Cutting of the methoxylor compound by the Marrero and Gan [77].	60
Figure 16 : 1,9-nonadiol	62
Figure 17 : N-Phenyl-1,4-benzenediamine	62
Figure 18 : Representation of QSPR methods	63
Figure 19 : Random Forest (RF) operating diagram.	73
Figure 20 : XGBoost operating diagram.	76
Figure 21 : Network architecture of a SVM regression.....	78
Figure 22 : illustrative representation of neural networks	79
Figure 23 : representation of a neural network with two hidden layers and one output neuron.	81
Figure 24 : Schematic representation of leave-one-out cross-validation	84
Figure 25 : Flowchart of the training (cross-validation) and testing (external validation) procedure	85
Figure 27. Number of data points according to different cation-anion combinations. Empty field means that the experimental data have been not available yet. The abbreviation of chemical families are detailed in Table 1.	100
Figure 28 : Cationic and anionic chemical families considered in this study. The abbreviations of the chemical families are detailed in Table 1.....	72
Figure 29: Scatter plots of predicted and experimental values in the training and test sets. (a) Contribution groups. (b) QSPR.....	79
Figure 30: Gaussian distribution of prediction errors in the training set and the test set (a) Contribution groups. (b) QSPR.....	81
Figure 31: Distribution of relative errors (RE) according to cationic and anionic families. The central mark, corresponding to the median, gives an indication of the central tendency of the data. The lower and upper edges of the box represent the first and third quartiles respectively. (a) QSPR- SVM model, (b) QSPR-DNN model, (c) GC- SVM model. Empty field means that the experimental data have been not available yet. The abbreviations of the chemical families displayed on the axes are detailed in Table 1.....	83
Figure 32: Relative error ER(%) distributions as a function of the combination of cation and anion families. (a) QSPR- SVM model, (b) QSPR-DNN model, (c) GC- SVM model. Empty field means that the experimental data have been not available yet. The abbreviations of the chemical families displayed on the axes are detailed in Table	87
Figure 33: Distribution of data points as a function of mean relative error ER (%).....	94
Figure 34: Scatter plots of predicted and experimental values in the training and test sets. (a) Contribution groups. (b) QSPR.....	126
Figure 35: Gaussian distribution of prediction errors in the training set and the test set (a) Contribution groups. (b) QSPR.....	127
Figure 16: Distribution of data points as a function of mean relative error ER - GC models (%).....	13
Figure 37: Distribution of data points as a function of mean relative error ER - QSPR models (%).....	133

LIST OF TABLES

Table 1: Cutting the 1,9-nonadiol molecule into groups	62
Table 2: Cutting the N-Phenyl-1,4-benzenediamine molecule into groups	62
Table 3: Summary of the viscosity database by ion families.....	73
Table 4: Model performance statistics on the training and test sets, expressed as averages from a 5-fold cross-validation, repeated 25 times according to random partition	78
Table 5: Comparison of performance of proposed models against existing models.....	96
Table 6: Number of viscosity data by sub-family of organic compounds.....	119
Table 7: Overview of experimental viscosity data for non-ionic compounds classified by chemical family.....	120
Table 8: Model performance statistics on the training and test sets, expressed as averages from a 5-fold cross-validation, repeated 25 times according to random partitions.....	123
Table 9: Mean Relative Error of GC Models (XGBoost, RF, SVM, DNN) by Major Chemical Family.....	128
Table10: Mean Relative Error of QSPR Models (XGBoost, RF, SVM, DNN) by Major Chemical Family.....	129

GENERAL INTRODUCTION

Viscosity, a fundamental fluid property, plays an essential role in numerous scientific and industrial domains, influencing phenomena ranging from molecular transport and reaction kinetics to process design and optimization. In chemistry, it governs liquid flow behavior, affecting chemical synthesis efficiency as well as the performance of lubricants and coatings. In physics, it lies at the heart of fluid dynamics, while in engineering, it is crucial for designing systems involving heat transfer, mass transfer, or fluid flows. The ability to accurately predict viscosity is therefore of paramount importance. However, this task remains complex due to the subtle interplay between molecular interactions, structural diversity of compounds, and their sensitivity to thermodynamic conditions such as temperature and pressure. Traditional experimental methods, though reliable, are often costly, time-consuming, and poorly suited for high-throughput screening of new compounds or large-scale applications – justifying the development of innovative predictive models.

This doctoral thesis addresses this critical need by leveraging advanced computational techniques, particularly machine learning algorithms and Quantitative Structure-Property

Relationships (QSPR), to predict the viscosity of organic compounds. These approaches capture complex relationships between molecular structure and macroscopic properties, thereby offering an efficient alternative to experimental measurements. The work focuses on two distinct classes of organic compounds: ionic liquids (ILs) and non-ionic organic compounds. Ionic liquids, characterized by their weakly coordinated ions, exhibit exceptional properties such as thermal stability, non-volatility, and structural tunability, making them attractive for applications in green chemistry, electrochemistry, and materials science. However, predicting their viscosity remains challenging due to complex ionic interactions and the diversity of possible ion pairs. Non-ionic organic compounds, meanwhile, encompass a vast range of molecules used in pharmaceutical, petrochemical, and materials industries, where viscosity prediction is equally crucial yet complicated by molecular diversity and nonlinear temperature dependencies.

In this context, this thesis aims to bridge these gaps by developing robust and generalizable predictive models based on integrating machine learning (ML) methods with advanced molecular representations. Our approach rests on three pillars:

1. Utilization of novel databases (15,251 experimental data points for ILs; 9,715 points for non-ionics) covering unprecedented thermodynamic ranges (74–1,120 K) and exceptional structural diversity.
2. Synergy between molecular descriptors: Group Contribution (GC) methods for chemical interpretability and Quantitative Structure-Property Relationship (QSPR) descriptors derived from deep learning to capture global structural effects.
3. Rigorous evaluation of ML algorithms (XGBoost, Random Forest, nonlinear SVM, deep neural networks) compared to conventional semi-empirical models.

Structured into four chapters, this thesis begins with a theoretical and bibliographic framework (Chapter I), followed by a methodological analysis of modeling tools (Chapter II). Chapters III and IV detail viscosity prediction for ionic liquids and non-ionic compounds respectively, demonstrating the superiority of QSPR models coupled with nonlinear algorithms (XGBoost, DNN), achieving up to 60% reduction in mean relative errors for complex structures.

This research provides scientists and engineers with high-fidelity predictive tools capable of replacing costly experimental measurements and accelerating the rational design of

task-specific fluids. Emerging prospects – GC/QSPR hybridization, extension to multicomponent systems – underscore the transformative potential of machine learning for predicting thermophysical properties.

CHAPTER I: THEORETICAL CONCEPT AND BIBLIOGRAPHICAL SUMMARY

I.1 INTRODUCTION

In this inaugural chapter, our presentation embarks on a twofold exploration. First, we look at the fundamental theoretical concepts governing the viscosity of pure organic compounds, clarifying the basic concepts and principles. Secondly, we undertake an exhaustive review of the literature concerning the various experimental techniques used to measure this essential property.

Beyond the realm of experimentation, our study becomes an in-depth investigation of the different approaches and methods used in viscosity prediction. Through a meticulous review of the existing literature, we delineate the landscape of predictive techniques, describing their scope, applicability and inherent limitations. This in-depth review covers both non-ionic and ionic compounds.

By engaging in this comprehensive review, our objective is to furnish readers with a comprehensive understanding of the theoretical underpinnings, experimental methodologies, and predictive strategies central to the study of viscosity in pure organic compounds.

I.2 VISCOSITY DEFINITION

When you apply a shearing stress (a force that causes layers to slide upon each other) to a fluid, it will start to move. The speed of the fluid's movement will change across its depth, being fastest where the stress is applied.

The viscosity of the fluid is a measure of how much it resists this movement. It's calculated by dividing the shear stress by the change in velocity across the fluid's depth.

You can think of viscosity as the “thickness” of a fluid, or how much it resists flow. It's also a measure of the friction within the fluid. For example, water, which flows easily, has a low viscosity and is considered “thin”. On the other hand, vegetable oil, which flows less easily, has a high viscosity and is considered “thick”.

In any flowing fluid, different layers move at different speeds, creating resistance between them that opposes any force applied (see Figure 2). Isaac Newton suggested that for fluid flowing in straight, parallel layers, the shear stress, denoted as τ , between layers is directly proportional to the velocity gradient in the direction perpendicular to the layers (as shown in Equation 1).

$$\tau = -\eta \frac{\partial u}{\partial y} \quad (1)$$

In this context, the constant η represents the coefficient of viscosity, also referred to as dynamic viscosity. The notation with the subscript "y" in the shear stress term indicates the specific area over which the shear stress is exerted. This area is oriented perpendicular to the y-axis and is located at a distance "y" from the origin. Meanwhile, the second subscript "x" signifies the direction along which the shear stress is applied. The negative sign in the equation signifies that the shear stress is exerted from a region of higher velocity to one of lower velocity, reflecting the fundamental concept of shear stress acting to equalize velocity gradients within a fluid medium. In this modelling context, many fluids conform to Newton's principle and are classified as Newtonian fluids.

Conversely, non-Newtonian fluids demonstrate a more intricate relationship between shear stress and velocity gradient, deviating from simple linearity (refer to Figure 2).

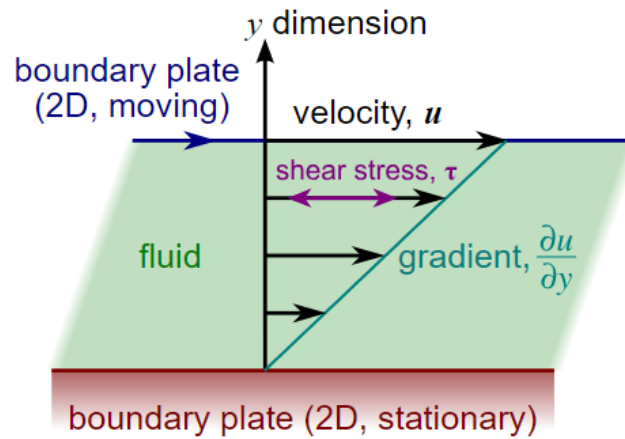


Figure 2 : Gradient of velocity for a Newtonian fluid as discussed by Massey

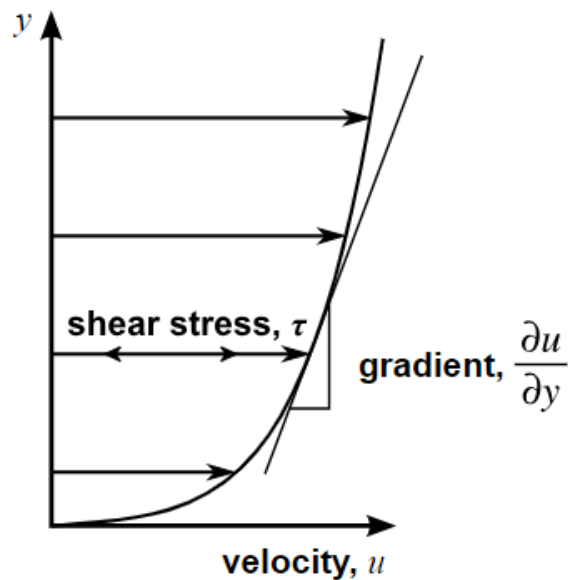


Figure 3 : Gradient of velocity for a non-Newtonian fluid as discussed by Massey

So, viscosity, by definition, emerges as the ratio of shearing stress per unit area to velocity gradient. The SI unit of dynamic viscosity is the Pascal-second ($Pa \cdot s$). Its other units are newton-second per square meter ($N \cdot s \cdot m^{-2}$) or poiseuille (PI).

There are two commonly used types of viscosity: dynamic viscosity (μ), also known as absolute viscosity and kinematic viscosity (ν). Kinematic viscosity, on the other hand, relates dynamic viscosity to the fluid's density. It reflects the fluid's resistance to flow. Kinematic viscosity is obtained by dividing the dynamic viscosity by the fluid's density

(ρ) and is expressed in units of square meters per second (m^2/s). Throughout this manuscript, when referring to "viscosity" without further clarification, it shall imply dynamic viscosity.

I.3 VISCOSITY MEASUREMENT

In the realm of viscosity measurement, a diverse array of viscometers exists, each employing distinct technologies to ascertain viscosity. These technologies exhibit varying degrees of precision and sensitivity, thereby presenting challenges for the development of robust predictive models. A comprehensive understanding of these aspects is paramount for the development of effective predictive models for viscosity. Here we present an overview of the main viscometry technologies and the uncertainties associated with them. It is imperative to note that while this study endeavors to provide a thorough analysis, it does not profess to be exhaustive in its scope.

I.3.1 Capillary viscometers

Among the most prevalent viscometers are capillary viscometers which utilize vertical capillary tubes immersed in a temperature-controlled bath for viscosity assessment. These viscometers gauge the flow of a fluid through a capillary, necessitating a calibrated reference fluid, typically water, for viscosity determinations. However, no single tube accommodates all viscosity ranges (Figure 5).

While these viscometers are cost-effective, the capillary tubes can prove challenging to clean, and measurement precision is compromised by humidity within the viscometer's air column. The uncertainties associated with this viscometer type stem primarily from temperature fluctuations, reading errors, and capillary size.



Figure 4 : Example of a capillary viscometer with temperature control.

Traditionally utilized for liquid measurements, these viscometers are also adaptable for gas viscosity determination. There are several types of capillary viscometers, each designed for specific applications and offering different levels of precision and sensitivity (Figure 5). Some of the most important types include the Ostwald viscometer, which uses a U-shaped tube to measure the flow of fluid between two points; the Cannon-Fenske viscometer, which is used for transparent liquids and measures flow through a capillary tube; the Ubbelohde viscometer, which measures flow through a capillary tube using a bulb at the bottom; and The Houillon viscometer, used for measuring the viscosity of opaque liquids. Other types include suspended level viscometers and microviscometers, each with its own advantages and limitations depending on the application and required accuracy. The choice of viscometer depends on the specific requirements of the application, such as the type of fluid being measured, the viscosity range, and the desired level of accuracy.

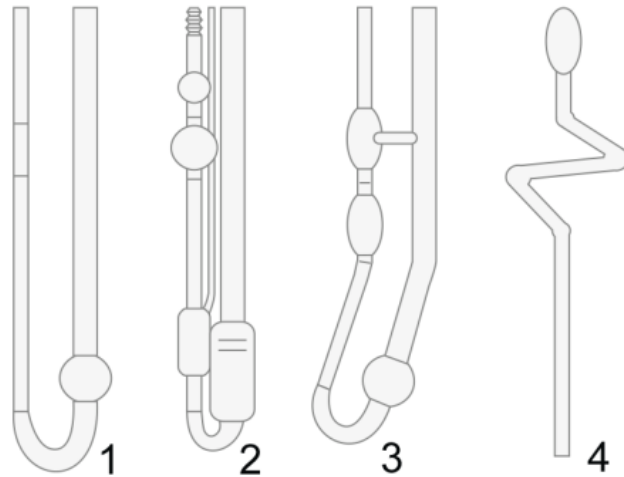


Figure 5 : Some types of glass capillaries: 1: Ostwald capillary, 2: Ubbelohde capillary, 3: Cannon-Fenske capillary, 4: Houillon capillary

I.3.2 Ball or phere viscometers

The falling ball or phere viscometer is a widely used instrument for measuring viscosity. This method involves timing the free-fall of a spherical or cylindrical object through a sample fluid under the influence of gravity or an external force, such as a piston. The object's velocity is directly related to the viscosity of the fluid. These viscometers can handle a wide range of fluids, including non-Newtonian fluids, and can be adapted to measure the viscosity of some types of gas.

One key challenge in using falling ball viscometers is the accurate measurement of the object's velocity. The largest source of measurement uncertainty arises from velocity measurement errors, which can be influenced by factors such as the fluid's viscosity, temperature, and the shape and size of the falling object. Studies have shown that the error in measuring viscosity with these instruments can fluctuate from 4% at high fluid densities to 6% at lower densities.

Maintenance of falling ball viscometers can also be intensive, requiring regular calibration and careful handling. For viscoelastic fluids, the elastic stress ratio must be considered, adding complexity to the measurement process. Careful calibration and maintenance are essential to ensure accurate and reliable viscosity measurements.



Figure 6 : Example of a falling ball viscometer with thermal regulation (phywe.fr).

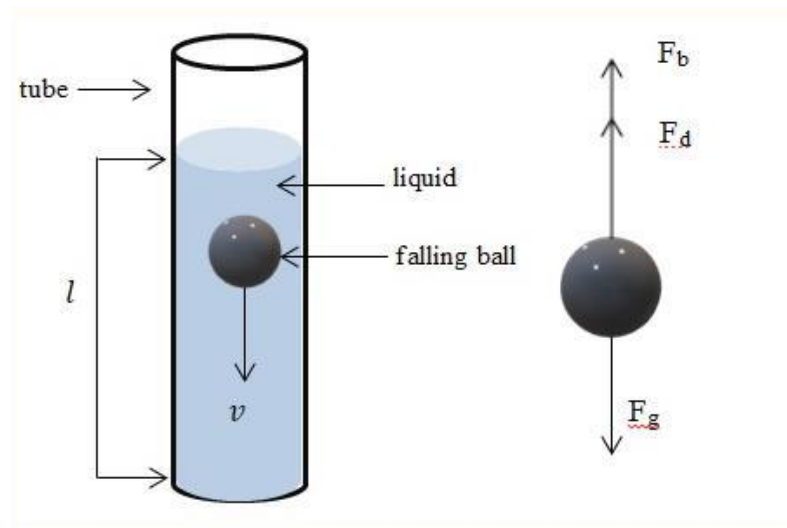


Figure 7 : Diagram illustrating affecting forces on a falling ball; F_g gravity force, F_b upward buoyant and F_d force of viscous (internal) friction.

I.3.3 Vibrational or oscillational viscometers

Vibrational or oscillational viscometers are a class of viscometers that operate on the principle of oscillatory motion. These devices incorporate a vibratory element, such as a wire, blade, or crystal, that is immersed in the fluid under examination. The principle behind their operation lies in analyzing the behavior of these vibrating elements within the fluid.

In the case of a vibrational wire viscometer, the operational concept is based on studying the frequency response to forced transverse oscillations of a metallic wire. This wire, akin to a tuning fork, is subjected to a magnetic field (B) and immersed in the medium under investigation (see Figure 8). An alternating electric current, closely aligned with the wire's mechanical resonance frequency, induces a movement in the wire. This movement generates an electromotive force (Lorentz force) at the wire's terminals, which corresponds to the detected momentum. By sweeping the electric frequency, resonance is observed. The resonance frequency, which is contingent on the fluid's density, is experimentally measured, thereby enabling the determination of viscosity.

Vibrating wire viscometers have been used to measure the viscosity of a wide range of fluids, over a wide range of temperatures and pressures. These viscometers have demonstrated their ability to measure viscosities ranging from 0.009 to 200 mPa·s in both liquid and gas phases. Operating over a temperature range from 1.1 K [1] to 455 K [2], and at pressures up to 1 GPa [3], the vibrating wire viscometers have proved versatile in their application across a wide range of experimental conditions.

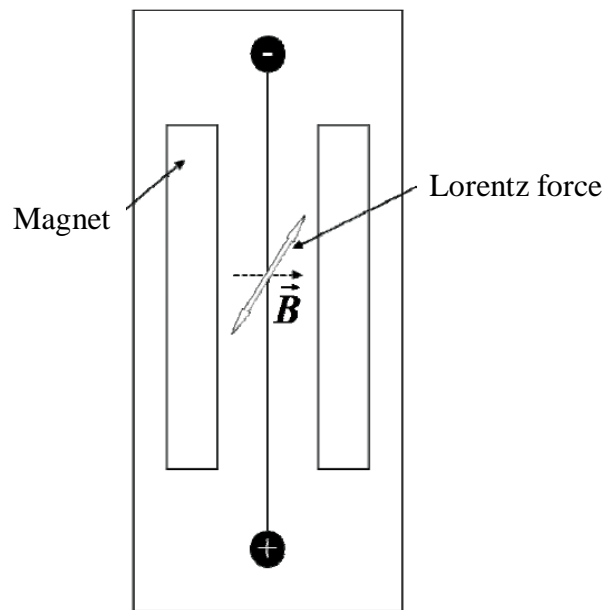


Figure 8 : Diagram of a vibrating wire viscometer

In the case of a vibrating crystal and vibrating plate viscometer, these two types of vibrating body viscometer have a similar operating principle. What sets this technology

apart from the vibrating wire viscometer is its remarkable versatility, allowing it to operate effectively across an extensive viscosity range spanning from 0.01 to 105 mPa·s.

At its core, the operational principle relies on the microscopic movement of a crystal or metallic rod. This ensures that the thermodynamic equilibrium of the fluid remains unaffected. The vibrating body comprises a piezoelectric compound, typically cylindrical in shape, traversed axially by an electrical current. Electrodes embedded within the body apply a sinusoidal voltage, inducing torsional vibrations. Submerged within the studied medium, the vibrating body sets in motion a distortion wave, swiftly attenuated due to its perpendicular displacement to the crystal's surface. Viscous forces exerted at this interface alter the resonant frequency of the vibrating body. Deriving the viscosity value relies on the same principle as that of the vibrating wire viscometer.

Despite the effectiveness of these instruments, vibrational viscometers generally have limitations. One significant limitation is the lack of a defined shear field in these instruments. This means that the shear forces acting on the fluid are not well-defined, making them unsuitable for measuring the viscosity of fluids with unknown flow behaviors. Additionally, the accuracy of these viscometers can be affected by factors such as temperature variations and the presence of particulate matter in the fluid.

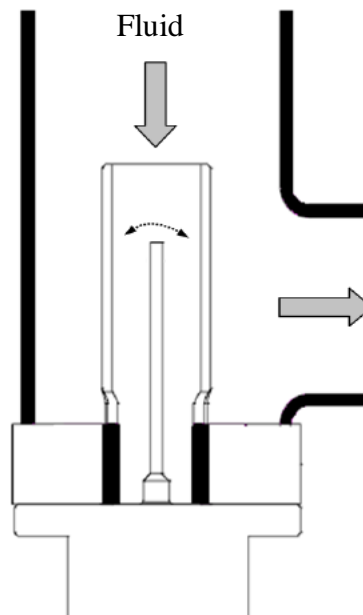


Figure 9 : Diagram of a vibrating crystal and vibrating plate viscometer

I.3.4 Rotational viscometers

Rotational viscometers are among the most prevalent types viscometers. Their popularity stems from their relatively simple design structures and their operational principle, which involves measuring the rotation of a solid object (such as a bob, metallic moulding, or cylindrical axle) when immersed in a viscous fluid under the application of a known force or torque required to rotate it within the fluid.

There are three main categories of rotational viscometers: coaxial cylinder viscometers, commonly known as Couette viscometers; cone-plate viscometers; and plate-plate viscometers, referred to as rotating disk viscometers. These viscometers vary in their mobile elements, which can be cylinders, cones, or disks (see Figure 10). The viscosity measurement is intricately linked to the torque applied to the element (G), the angular velocity of the mobile element (ω), the outer radius of the mobile element (R_m), and the inner radius of the stationary element (R_s).

Rotational viscometers are well-suited for characterizing highly viscous media, such as molten polymers. They also enable the study of the rheological behavior of these media through the control of shear rate. However, this technique is less suitable for low-viscosity media because the torque required for the rotation of the mobile element may be insufficient for precise viscosity measurement.

These instruments are widely used due to their simple design structures and versatility. However, their complex geometries can hinder the effective measurement of the rheological behavior of certain fluids, which has an impact on the accuracy of shear stress measurements and can result in a measurement inaccuracy of at least 10% [4]. Additionally, rotational viscometers require full immersion of their rotating element in the test liquid, leading to a significant amount of testing sample being needed. This can contribute to an effective slip of larger particles to the walls of the instrument, potentially affecting the accuracy of viscosity measurements.

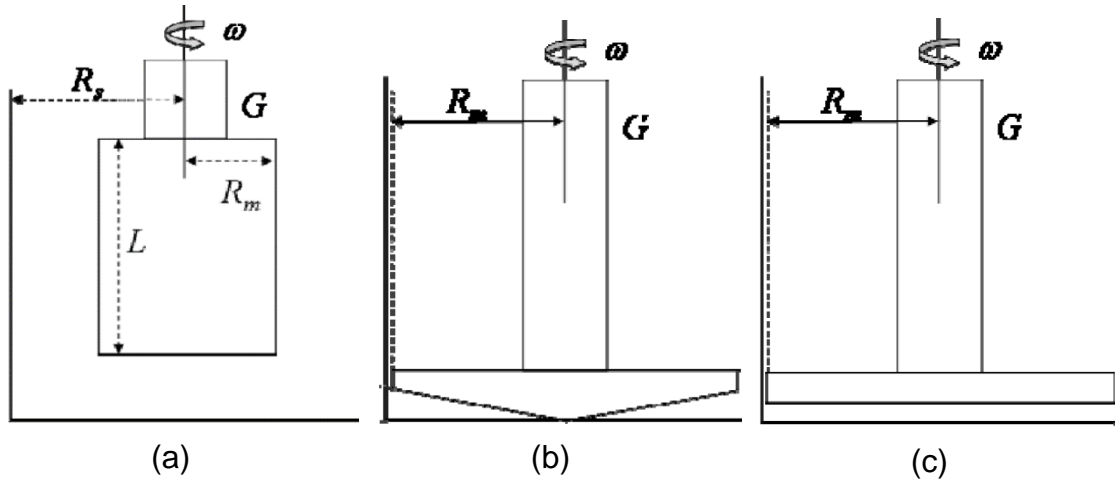


Figure 10 : Types of rotational viscometer; (a) Coaxial cylinder viscometer, (b) Cone-plane viscometer, (c) Planar viscometer

I.3.5 Non conventional viscometers

Unconventional viscometers encompass a range of innovative technologies that depart from traditional measurement methods. These viscometers often offer unique advantages, such as the ability to measure viscosity in situ, in real-time, or in non-Newtonian fluids. One example is acoustic viscometry, which measures viscosity by analyzing the speed of sound waves propagating through a fluid. This technique is particularly useful for opaque or non-homogeneous fluids and can provide measurements in challenging environments. Another unconventional approach is microrheology, which measures viscosity at the microscale using techniques such as particle tracking or optical tweezers. Microrheometers offer high sensitivity and can characterize complex fluids, including those with nonlinear rheological behaviors.

Optical viscometers represent another category of unconventional viscometry, utilizing light-based methods such as interferometry or fluorescence to measure viscosity. These viscometers are often non-invasive and can provide insights into fluid dynamics at the molecular level. MEMS (Micro-Electro-Mechanical Systems)-based viscometers, leveraging microfluidic technologies, offer miniaturization and integration benefits, enabling precise viscosity measurements in small volumes. Electromagnetic viscometers, which use electromagnetic fields to measure viscosity, provide another unconventional approach suitable for a wide range of fluids and conditions. These unconventional viscometers collectively expand the possibilities for viscosity measurement, offering innovative solutions for diverse applications.

I.3.6 Conclusion

The existence of multiple measurement technologies for viscosity introduces varying degrees of uncertainty associated with each method. This diversity in uncertainty intervals can complicate predictive modeling efforts, as different viscometers may yield different results for the same fluid. Factors such as measurement precision, sensitivity to environmental conditions, and calibration requirements contribute to the overall uncertainty in viscosity measurements. For instance, capillary viscometers may have uncertainties related to temperature fluctuations and capillary size, while falling ball viscometers may face challenges with velocity measurement errors and maintenance requirements. Therefore, when developing predictive models for viscosity, it is essential to account for the uncertainties inherent in each measurement technology to ensure the accuracy and reliability of the models.

I.4 CAUSES OF ERRORS IN VISCOMETERS

Having acknowledged the existence of multiple viscosity measurement technologies and the varying degrees of uncertainty they introduce, it's important to note that there are several potential sources of error in viscometers. These errors can be broadly classified into the following categories [5]:

1. **End/edge effect:** This occurs due to energy losses at the fluid entrance and exit of the main test geometry.
2. **Kinetic energy losses:** Loss of pressure to kinetic energy can lead to errors in measurement.
3. **Secondary flow:** Unwanted secondary flow, vortices, etc., can cause energy loss and increase with the Reynolds number.
4. **Nonideal geometry:** Deviations from ideal shape, alignment, and finish can introduce errors.
5. **Shear rate non-uniformity:** This is particularly important for non-Newtonian fluids.
6. **Temperature variation and viscous heating:** Variation in temperature, both in time and space, can influence the measured viscosity.
7. **Turbulence:** Partial and/or local turbulence can develop even at low Reynolds numbers.
8. **Surface tension:** Differences in interfacial tensions can affect measurements.

9. **Elastic effects:** Structural and fluid elastic effects can introduce errors.
10. **Miscellaneous effects:** These depend on the test specimen and can include factors like melt fracture, thixotropy, and rheopexy.

I.5 TEMPERATURE AND PRESSURE DEPENDENCE

As previously discussed, the viscosity behaviors of liquids and gases are governed by distinct mechanisms, leading to observable differences in their responses to temperature variations.

In gases, the tangential force between adjacent layers arises from the transfer of momentum caused by the movement of individual molecules across layers. This movement results in collisions between molecules, with each molecule traversing freely over a certain distance before colliding with another. Consequently, the collective effect of these molecular interactions contributes to the overall viscosity of the gas. Thus, as temperature rises, the viscosity of gases shows an opposite tendency to that of liquids, which typically increases with temperature and often increasing further under higher pressure conditions.

Conversely, the same conceptual framework does not apply seamlessly to liquids. In liquid environments, molecules tend to occupy longer-lived structures and engage in more sustained interactions, punctuated by intermittent shifts to adjacent positions, often referred to as "jumps." These movements occur within a complex network of intermolecular forces, which significantly influence the overall behavior and viscosity of the liquid. Thus, the dynamics of liquid viscosity are characterized by a unique interplay of molecular arrangements and interactions. The viscosity of liquids generally decreases with increasing temperature. This is because the kinetic energy of the molecules in the liquid increases, leading to greater molecular movement and weaker intermolecular forces. It's important to note that the temperature dependence of viscosity varies among different liquids and can be influenced by factors such as molecular structure, polarity, and intermolecular interactions. On the other hand, the pressure dependence of liquid viscosity is less straightforward compared to its temperature dependence. In general, an increase in pressure tends to increase the viscosity of liquids, although the magnitude of this effect can vary (refer to Figure 11)

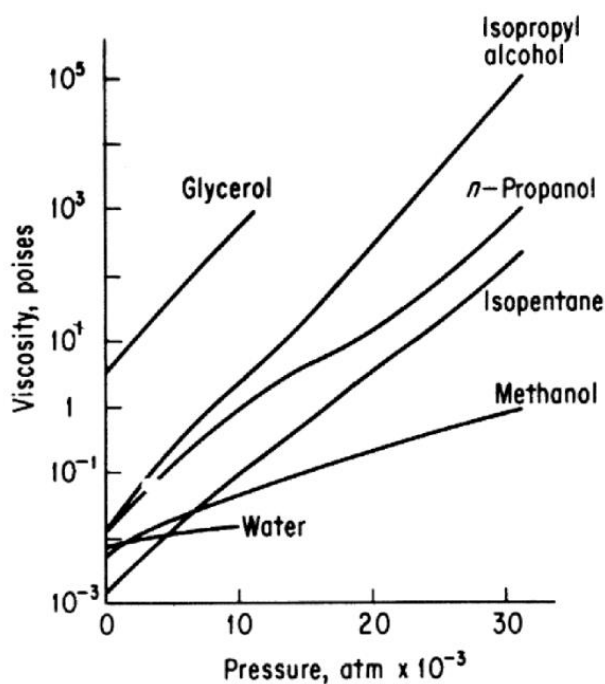


Figure 11 : Effect of pressure on viscosity of selected organic liquids at room temperature [6].

I.6 UNDERSTANDING THE DIFFERENCES IN GAS AND LIQUID BEHAVIOR

Gases and liquids exhibit fundamentally different behaviors in this aspect, requiring distinct approaches for their characterization. Gases, especially at low densities and high temperatures (as observed in ideal gases), behave in a relatively simple manner. Their viscosity can be effectively described by an equation that considers the average distance molecules travel between collisions (mean free path) and the transfer of momentum during these interactions. This approach aligns well with the Boltzmann equation, which statistically describes the behavior of dilute gases, particularly monatomic gases (those containing single atoms).

In contrast, liquids and dense gases involve a more complex interplay of molecular forces, making the Boltzmann equation inadequate for capturing their viscosity. These forces, acting as the invisible glue that binds liquids together, encompass a diverse spectrum. Short-range interactions, such as repulsive forces and hydrogen bonding, operate at close quarters, while long-range electrostatic forces, arising from multipole moments, exert their influence over broader distances. Furthermore, unlike gases where viscosity diminishes with rising temperatures, liquids exhibit the opposite trend. Additionally,

liquid viscosity is intricately linked to their density, a stark contrast to the behavior observed in gases. All these fundamental differences highlight the inapplicability of gas models to describe liquid behavior and emphasize the need for more sophisticated approaches to accurately predict the viscosity of liquids.

In the scope of our doctoral thesis, while the prediction of gas viscosity holds significance, our focus has been exclusively on developing models for predicting liquid viscosity, thereby setting aside gas-specific models. This decision is influenced by several pertinent factors. Firstly, there is a notable need to address gaps in predicting liquid viscosity, particularly for ionic liquids, where current models may exhibit deficiencies due to the absence of a universally accepted theory for liquid viscosity. Moreover, viscosity theories applied to gases have reached a certain level of maturity, with well-established and widely accepted models.

I.7 A REVIEW OF PRACTICAL PREDICTIVE AND CORRELATIVE METHODS FOR ESTIMATING VISCOSITY

The complex nature of liquid viscosity continues to challenge researchers, despite extensive studies. The viscosity of liquids has been the subject of numerous investigations, yet to date, no adequate theory is available to satisfactorily reproduce experimental data and be universally valid for all liquids. All theories used for dense gases and liquids are empirically modified, often requiring adjustments to fit experimental data, rendering them essentially semi-theoretical at best. Consequently, models used for the prediction and correlation of viscosities in fluids can be broadly categorized into two main groups: semi-empirical and empirical models.

I.7.1 Semi-empirical models

Semi-theoretical models, which blend theoretical principles with experimental data trends, are generally considered to yield satisfactory results, especially when compared to purely theoretical models that often involve significant errors in estimating liquid viscosities. These models are particularly prevalent in the study of dense gases and liquids. They typically involve coupling a viscosity model with a density or volume prediction method. Such approaches commonly rely on corresponding states principles and statistical mechanics models, including residual approach, hard sphere, and square

well theory or modifications of these theories. These models are inherently dependent on temperature and density (or volume). However, since viscosity is often required at specific pressure and temperature conditions, coupling a density prediction method with the viscosity model becomes necessary. The accuracy of these models in predicting liquid viscosity is highly dependent on the accuracy of the density input, emphasizing the importance of precise density values in these calculations.

I.7.1.A Corresponding states approaches

The corresponding states (CS) method in viscosity modeling is based on the idea that the reduced property of a fluid is equivalent to that of a reference fluid at the same reduced conditions. Originally observed by van der Waals, this principle implies that two fluids exhibit the same reduced viscosity at identical reduced temperature and pressure. One example of a corresponding states model is the Ely and Hanley model [7], also known as the transport Properties Prediction (TRAPP) model, which utilizes methane as the reference fluid. The model relates the viscosity of a fluid of interest to that of methane using critical properties, molecular mass, and density. It is represented as:

$$\eta_i(\rho, T) = \eta_0(\rho_0, T_0)(M_i/M_0)^{1/2} h_{i,0}^{-2/3} f_{i,0}^{1/2} \quad (2)$$

where η_i is the viscosity of the fluid of interest, ρ is its density, and T is the temperature. M_i and M_0 are the molecular masses of the fluid of interest and methane, respectively, while $h_{i,0}^{-2/3}$ and $f_{i,0}^{1/2}$ are shape factors determined empirically. While the TRAPP model shows good performance for straight nonpolar hydrocarbons, it struggles with branched hydrocarbons, cyclic hydrocarbons, and polar fluids, leading to modifications by subsequent researchers.

Another corresponding states model is the Pedersen and Fredenslund model [8], which links the viscosity of the fluid of interest to that of methane based on pressure and temperature. The model relates the viscosity of the fluid of interest (subscript x) to that of the reference fluid (subscript 0) in terms of pressure and temperature. It is represented by the equation:

$$\eta_x(P, T) = \left(\frac{T_{c,x}}{T_{c,0}} \right)^{-1/6} \left(\frac{P_{c,x}}{P_{c,0}} \right)^{2/3} \left(\frac{M_x}{M_0} \right)^{1/2} \left(\frac{\alpha_x}{\alpha_0} \right) \eta_0(P^*, T^*) \quad (3)$$

where $T_{c,x}$ and $T_{c,0}$ are the critical temperatures of the fluid of interest and reference fluid, $P_{c,x}$ and $P_{c,0}$ are the critical pressures, M_x and M_0 are the molecular masses, and α_x and α_0 are the Tham-Gubbins rotational coupling coefficients for the fluid of interest and reference fluid, respectively. The viscosity of the reference fluid is denoted as $\eta_0(P^*, T^*)$ where P^* and T^* are reduced pressure and temperature, respectively. This model shows good correlation for most pure component and mixture viscosities but exhibits larger deviations for heavy alkanes and binary mixtures with significantly different molecule sizes. Extensions to improve predictions at reduced temperatures below 0.47 have been proposed by Pedersen and Fredenslund [9].

Additionally, the Teja and Rice model [10] extends the generalized corresponding states principle to liquid mixture viscosities, using two nonspherical reference fluids and a reducing parameter. The model equation is given by:

$$\ln(\eta\xi) = \ln(\eta\xi)^{r_1} + \frac{(\omega - \omega^{r_1})}{(\omega^{r_2} - \omega^{r_1})} [(\ln \eta\xi)^{r_2} - (\ln \eta\xi)^{r_1}] \quad (4)$$

Here, superscripts r_1 and r_2 refer to two nonspherical reference fluids similar to the pure component or key components of interest in mixtures. ω represents the Pitzer acentric factor, and ξ is the reducing parameter given by Teja et al. reformulated the Teja-Rice method in terms of critical pressure (P_c) instead of critical volume (V_c), considering the uncertainty often associated with the measurement of critical volumes. The revised reducing parameter is expressed as

$$\xi = T_c^{1/6} / (M^{1/2} P_c^{2/3}) \quad (5)$$

However, a notable limitation of this model is the requirement for reference fluids that closely resemble the fluid of interest, making it challenging to obtain viscosity data for these reference fluids readily. Consequently, interpolation or extrapolation may be necessary for obtaining this data.

Aasberg-Petersen et al. [11] further refined the Teja-Rice approach by using two reference fluids (methane and decane) for predicting hydrocarbon viscosities, demonstrating reasonable predictions for light petroleum fractions but less accuracy for heavy fractions. Unlike the Teja-Rice model, which uses the acentric factor as the interpolation parameter, this model uses the molecular mass, as the acentric factor decreases with increasing molecular weight for heavy oil fractions. Extensions and modifications to these models continue to be developed to improve their applicability and accuracy across a wider range of conditions and fluid types.

Several approaches based on FT have been used to predict the viscosity of ionic liquids. For instance, Huang et al [12] developed a method based on the group contribution method and the corresponding states principle to predict the critical properties of ionic liquids (ILs). They classified 46 fragments specifically for ILs, taking into account their ionic features, and determined the corresponding fragment increments using experimental density data. The results indicate that this method is reliable, with an average absolute relative deviation of less than 4%.

Zhao et al. [13] proposed two models for predicting the viscosity of imidazolium-based ionic liquids (ILs). The first model integrates the fragment contribution-corresponding states (FC-CS) method with a multiple linear regression (MLR) algorithm, while the second model utilizes a support vector machine (SVM) algorithm. In the FC-CS method, pseudo-critical volume and compressibility factor (V_c and Z_c), as well as the boiling point temperature (T_b), are calculated and used to predict viscosity with MLR and SVM algorithms. They validated the models using a large dataset of 1079 experimental data points for 45 imidazolium-based ILs, covering a wide range of pressure and temperature. The average absolute relative deviation (AARD) for the entire dataset was found to be 24.2% for MLR and 3.95% for SVM. The nonlinear model developed by the SVM algorithm outperformed the linear model created by MLR, indicating that the SVM algorithm is more reliable for predicting the viscosity of imidazolium-based ILs.

I.7.1.B Residual Approach

The residual approach in viscosity calculations expresses the viscosity of a fluid as a residual property, $\Delta\eta$, from the zero-density limit or dilute gas viscosity, η_0 , where $\eta - \eta_0 = \Delta\eta$. This approach is particularly useful for liquids and supercritical fluids, where the dilute gas term is negligible. The residual term $\Delta\eta$ dominates for liquid viscosities in these cases. Various methods have been proposed in the literature for calculating this term, with three prominent models being Friction Theory (FT), Free-Volume Theory (FVT), and Expanded Fluid Theory (EFT).

I.7.1.B.1 Friction Theory (FT):

Quiñones-Cisneros et al. [14] developed the FT model, linking residual viscosity to an equation of state (EoS) prediction of repulsive and attractive pressure contributions. This model, based on classical mechanics concepts, has been successfully tested with n-alkanes and their mixtures over a wide range of conditions.

Several approaches based on FT have been used to predict the viscosity of ionic liquids. Shen et al. [15] were among the first to apply this friction theory to ionic liquids, specifically [Cnmim][BF₄], [Cnmim][PF₆], and [Cnmim][NTf₂], using ePC-SAFT to calculate the different pressure contributions. Abolala et al. [16] also applied a similar approach, substituting the friction theory for the free volume theory, and estimated the repulsive and attractive pressures from the SAFT equation-of-state. They found that short-range repulsive forces prevail over long-range attractive forces at high pressure, leading to a simplified model with fewer adjustable parameters.

Haghighbakhsh et al. [17] achieved a better correlation with the PC-SAFT equation compared to Abolala et al., with a model that included the second-order attraction term and a more complex temperature dependence for the friction coefficients. Macias-Salinas [18] developed a viscosity model based on friction theory in parallel with Eyring's theory, neglecting the second-order attraction term and simplifying the expressions for the friction coefficients to reduce the number of adjustable parameters to six. These models were tested on various ionic liquids and compared with other models in the literature, showing comparable performance overall, with some underestimation of viscosity at high pressure and/or low temperature. The difficulty in parameter optimization is noted, attributed to the correlation between parameters and their degeneracy.

I.7.1.B.2 Free-Volume Theory (FVT):

Originating from the idea that viscosity depends on free space in a fluid, FVT has been utilized by Batschinski [19], Cohen and Turnbull [20], and Doolittle [21]. Despite the need for experimental data fitting, it has shown good predictive accuracy for various fluids, especially when used with average absolute deviations of less than $\sim 10\%$ over wide temperature and pressure ranges.

Several approaches based on FVT have been used to predict the viscosity of ionic liquids. Mac-Dowell et al. were among the first to use this approach for ionic liquids, specifically [Cnmim]Cl and [Cnmim][CH₃SO₄], using the soft-SAFT equation-of-state to calculate density. They achieved promising results with an Average Absolute Relative Deviation (AARD) of 8.48% for [Cnmim]Cl and 2.97% for [Cnmim][CH₃SO₄]. Polishuk and Yitzhak applied a similar method using the SAFT+cubic equation of state for organic compounds, including [Cnmim][BF₄] or [Cnmim][PF₆], but encountered large deviations, leading them to explore the Modified Yarranton-Satyro Correlation, which yielded significantly better results.

Llovel and Vega [22] adapted a model previously developed for n-alkanes to [Cnmim][BF₄] ionic liquids, but faced challenges in parameter optimization when using viscosity data at atmospheric pressure over a wide temperature range. Shen et al [23] refined the parameter optimization strategy of the FVT model for [Cnmim][BF₄], [Cnmim][PF₆], and [Cnmim][NTf₂] ionic liquids, linearizing parameters with respect to cation molecular weight to reduce the overall number of parameters.

Overall, these methods demonstrate the application of FVT and various equation-of-state models to predict the viscosity of ionic liquids, highlighting the importance of parameter optimization and model selection for accurate predictions.

I.7.1.B.3 Expanded Fluid Theory (EFT):

Yarranton and Satyro [24] proposed an EFT model correlating residual viscosity with density based on the Hildebrand fluidity principle. This model considers the increase in viscosity with the expansion of fluid due to the increasing distance between fluid molecules. Successfully implemented in commercial process simulators, it provides a practical tool for predicting viscosity of hydrocarbons and their mixtures [25] [26].

I.7.1.B.4 Scaling Methods:

Excess Entropy Scaling, as proposed by Rosenfeld [27], establishes a relationship between viscosity and residual entropy for simple atomic fluids, derived from molecular dynamics simulations. On the other hand, Thermodynamic Scaling, introduced by Ashurst and Hoover [28], relates viscosity to a single generalized variable including density, temperature, and volume expansion. While these methods have shown success in predicting viscosity for a range of fluids, they may not be suitable for hydrogen-bonded liquids.

I.7.1.C Hard-sphere theory (Modified Enskog Theory)

Enskog was the first to develop a kinetic theory for predicting the effect of pressure on the viscosity of gases, based on the Boltzmann equation and the assumption that dense gases consist of hard spherical molecules [29]. This theory, commonly known as Chapman-Enskog theory [30], assumes that molecular collisions are impulsive. The viscosity in this model is given by an approximate expression involving the dilute gas viscosity η_0 , radial distribution function at contact (Ψ), the molar density (ρ), and the volume occupied by the hard spheres (b).

$$\frac{\eta}{\eta_0} = \left[\frac{1}{\Psi} + 0.8b\rho + 0.761\Psi(b\rho)^2 \right] \quad (6)$$

Alder et al. [31] later showed that significant deviations can occur between Enskog's theory and molecular dynamics calculations for a hard-sphere fluid, especially in the high-density region. Modifications to the hard-sphere model were proposed to account for high density, such as Chandler's rough-hard-sphere theory [32], which considers the coupling between translational and rotational motions of molecules.

Dymond [33] proposed corrections to Enskog's theory to include the attractive forces of real molecules, while Chung et al [34]. extended the Chung-Lee-Starling technique [35] to include dense fluids, incorporating an empirically correlated function of density and temperature. The proposed Chung-Ajlan-Lee-Starling (CALS) method [35] was tested on pure fluids, showing promising accuracy.

Other approaches based on hard-sphere theory include those by Assael et al [36], Bleazard and Teja [37], Ciotta et al. [38], and Riesco and Vesovic [39]. Extensions of these models to mixtures have also been presented, with the Vesovic-Wakeham (VW) method [39] being a reliable way to predict the viscosity of liquid mixtures. Recently, de Wijn et al. [40] extended the VW model to liquid mixtures consisting of long, chain-like molecules using the statistical associating fluid theory (SAFT) approach, which has no adjustable parameters in its thermodynamically consistent mixing rules.

This approach has been applied and adjusted for predicting the viscosity of ionic liquids. Initially developed by Assael et al. [41] in the early 1990s for various organic compounds, the correlation procedure was later tested by Gaciño et al. [42] in 2014 for 19 pure ionic liquids of imidazolium, pyrrolidinium, and phosphonium types. The relationship between the characteristic molar volume V_0 and temperature is described by a 4-parameter model.

Hossain and Teja [43] followed a similar approach for 48 ionic liquids over wide ranges of temperature and pressure, optimizing two sets of parameters based on viscosity data at atmospheric pressure and high pressure. Both sets of parameters showed good extrapolation at high pressure. Hosseini et al. [44] used a different method, calculating the viscosity of a smooth hard-spheres fluid. They obtained a linear correlation between the values of both parameters V_0 and R_η and the molar mass for a family of ionic liquids with a common anion.

Hosseini et al. [44] extended their method to a binary mixture of hard spheres using mixing laws. This approach was tested only for the binary mixture of ionic liquids [C4mim][PF6]+ [C4mim][BF4] and showed good predictive capability. However, there was an error in the manuscript of Hosseini et al., and an erratum [45] was not published until one year after the publication of the article of Akbari et al. [46]

I.7.1.D Square-well models

Davis et al. [47] derived an expression for the viscosity of dense fluids of molecules interacting with a square-well potential, which accounts for both repulsion and attraction. The derived equation is complex and includes parameters such as dynamic viscosity (η_{SW}), dilute gas viscosity (η_0), fluid density (ρ), repulsion diameter (σ_1), energy well depth (ϵ), radial distribution function (g), square-well width parameter (R), covolume

(b), Boltzmann constant (k), and temperature (T). They also introduced ψ_1 and ψ_2 as functions related to the energy well depth and temperature.

McLaughlin and Davis [48] extended this model to binary mixtures, while Du and Guo [49] proposed a semi-theoretical modification to the square-well model and derived an empirical expression for the radial distribution function. Monnery et al. [50] developed a semi-theoretical model to estimate liquid and gas phase viscosities based on the square-well theory. Galliero et al. [51] used molecular dynamic simulations to predict the viscosity of Lennard-Jones fluids, revealing that the basic Lennard-Jones model tends to underestimate viscosity, especially at lower temperatures.

Recetely, Zerón et al. [52] adapted this model Using a combination of molecular dynamics simulation and the continuous version of the square-well (CSW) intermolecular potential. This study validated the use of the CSW intermolecular potential in molecular dynamics simulations for accurately emulating SW transport properties.

I.7.2 Empirical models

I.7.2.A Quantitative Structure-Property Relationships (QSPR)

Quantitative Structure-Property Relationship (QSPR) is a method that predicts the properties of molecules based on their structure. This method originated from Quantitative Structure-Activity Relationship (QSAR), which was initially used to predict the biological activity of pharmaceutical molecules. QSPR models vary in complexity, depending on the type of descriptors used, which can be constitutional, topological, geometrical, or based on quantum mechanics calculations.

The calculation of these molecular descriptors, generally calculated with commercial software such as CODESSA and DRAGON, and the selection of the most significant of them is therefore one of the main drawbacks of approach [53].

Viscosity, a transport property, is particularly challenging to predict using QSPR [54]. Various QSPR models using molecular descriptors have been developed to predict liquid viscosities. Initially, predictions were based on viscosities at a fixed temperature or other properties like vapor pressure. Ivanciuc et al. created a five-parameter Multiple Linear

Regression (MLR) model from on 337 organic compounds, with a log standard error of 0.371. This model predicts viscosity using the following equation:

$$\ln \eta = -2.814 + 3.387C + 8.858 \times 10^{-3}MW + 3.919 \times 10^{-1}O - 8.486 T + 6.684 \times 10^{-1}Y \quad (7)$$

Here, C represents the hydrogen-bonding donor charged surface area (order 2), MW is the molecular weight, O is the Randic index (order 3), T is the maximum electrophilic reactivity index for a C atom, and Y is the maximum atomic orbital electronic population.

Using 361 organic compounds, another MLR model was proposed by Katritzky et al. [55] with a standard error of 0.22 (log units). This model predicts viscosity ($\ln \eta$) using equation :

$$\ln \eta = -10.3 + 1.77HDCA(2) + 0.0000557G_I + 2.78N_{rings} + 20.2FPSA(3) + 0.0897E_{min}(C) \quad (8)$$

Here, $HDCA(2)$ is the hydrogen-bonding donor charged surface area (order 2), G_I is the gravitational index, N_{rings} is the relative number of rings in the molecule, $FPSA(3)$ is the fractional positive partial charged surface area (order 3), and $E_{min}(C)$ is the minimum atomic state energy for a C atom.

Different models have been developed to predict ionic liquid viscosity as function of temperatures [56]. Typically, models use temperature as a descriptor, but this doesn't always provide a clear understanding of its impact on viscosity. When ranking descriptors based on their influence, temperature often emerges as one of the most significant factors. For instance, in their model, Tochigi and Yamamoto [57] used a reference temperature (the temperature at which the experimental viscosity is measured) as a descriptor to predict for ionic liquid viscosity. They then demonstrated how the predicted viscosity varies as a function of this reference temperature.

Barycki and his team [58] formulated an approach to predict viscosity for ionic liquids at a specific temperature based on the viscosity at 25°C, which is determined using a QSPR model they developed. Additionally, they considered the density calculated through molecular dynamics simulations. This approach relies on the understanding that viscosity and density are interconnected through the concept of free volume.

Researchers Bini and colleagues [59] developed two distinct correlations using different descriptors for high (353 K) and low (293 K) temperatures. The larger error at lower temperatures suggests that the model may not fully capture the transport mechanism at these temperatures. In a similar study, Yu et al. [60] examined the significance of different intermolecular interactions at eight different temperatures. They found that electrostatic interactions affect viscosity across the entire temperature range studied (from 283 to 343 K). At lower temperatures, Van der Waals interactions and hydrogen bonds are dominant, while steric effects become more significant at higher temperatures. This indicates that enthalpic effects have a greater impact on viscosity at lower temperatures, and entropic effects increase with temperature, which aligns with thermodynamic principles. Despite their advantages, the obtained QSPR models may not be universally applicable to the same family of ionic liquids, especially in the high and low viscosity ranges simultaneously [61].

Over the last few decades, many researchers have used advanced non-linear models to describe viscosity using different descriptors, ranging from simple to complex models. These models leverage machine learning techniques, including neural networks and Support Vector Machines (SVMs). These cutting-edge models have demonstrated potential in surpassing some of the constraints of conventional QSPR models, providing more precise predictions across a wider array of conditions and compound families. This kind of technology enables the boundaries to be pushed further, facilitating the description of the non-linear relationship between temperature, descriptors, and viscosity.

Gharagheizi et al. [62] presented a QSPR model for predicting the liquid viscosity of pure organic compounds at 25 °C. This model, a three-layer Feed Forward Artificial Neural Networks (FFANN), was developed using eight parameters selected by the genetic algorithm-based multivariate linear regression (GA-MLR) method from over 3000 potential descriptors. The model's predictive capability was evaluated using several statistical methods, with the root mean square of error and the average absolute percent error of the model found to be 0.34 and 7%, respectively.

Very recently, in 2021, Zhang et al. [63] proposed a QSPR model, for the prediction of ILs viscosity as a function of temperature and pressure, based on norm descriptors derived from atomic distribution matrix of molecular graphs and molecular geometries. The authors presented the same nonlinear model as published in their previous work [20],

where they assumed a nonlinear dependence of viscosity on temperature and linear on pressure. The database used to determine the molecular descriptors contained 7342 experimental data points for 351 ILs covering the temperature range 253–438 K and pressure range 0.06–300 MPa. Their model were evaluated using internal validation (leave-one-out cross validation- LOOCV) and external procedure in which the data set was randomly divided into training set (80%) and testing set (20%). The AARE values were 4.76, 4.84 and 4.74% for LOOCV, training and testing, respectively, for viscosities reported in log-units. When the overall AARE is recalculated, using the supplementary data provided by the authors, a value of approximately 7.5% is obtained. It should be noted that the application of this model is relatively difficult, as it requires geometric optimization of the ions and software with particularly high skills to evaluate the descriptors. In same year, Ding et al. [64] proposed four QSPR models to predict the viscosity of ILs using four different ways to represent the ILs, viz. molecular fingerprints, molecular descriptors, addition of cation and anion molecular fingerprints and combination of molecular fingerprints and molecular descriptors. The correlations were established using tree-based machine learning algorithm with a gradient boosting. The QSPR model developed using combination of molecular fingerprint and molecular descriptor showed better results with a correlation coefficient (R^2) of 0.908. The authors successfully applied the SHAP (shapely additive explanation) method to interpret the contributions of each fingerprint or descriptor in their best model, but unfortunately they did not report any supplementary data about their database or prediction results.

Despite its advantages, QSPR models, the success of a QSPR model heavily depends on the selection of appropriate molecular descriptors such as conformational, electronic, quantum mechanical, spatial, topological, thermodynamic, and many more [65]. The selection of irrelevant or redundant descriptors can negatively impact the performance of the model. Additionally, a significant challenge presents itself when certain chemical descriptors are either not directly accessible for the compounds being studied or necessitate the use of complex software. This can pose a substantial obstacle in the successful implementation of a QSPR model.

I.7.2.B Group Contribution models

Group contribution models are valuable tools in predicting the viscosity of complex chemical mixtures. These models leverage the principle that the viscosity of a compound

can be approximated by summing the contributions of its individual functional groups or structural fragments. By assigning specific viscosity values to each group or fragment, these models can estimate the viscosity of a mixture based on its composition. Group contribution models have been widely used in various industries, including chemical engineering, pharmaceuticals, and food science, due to their ability to provide quick and reliable viscosity predictions for a wide range of compounds. This approach has the advantage of being simple, very fast and easy to apply without requiring a great deal of computing power and does not require specialized software to implement [66] [67] [68] [69] [70] .

Various group contribution models have been developed to predict liquid viscosities (see Table x). The most important models in this category are those presented by Paduszyński [71], Chen [72], Baghban [73], Lazzús [74], Paduszyński [75] and Gharagheizi [62]. Five years after publishing their remarkable GC model [75] for the prediction of ILs viscosity, the co-authored Paduszyński proposed a new very extensive GC for calculating viscosity of ILs as a function of temperature based on the largest data including the data for 1978 distinct ILs, i.e. 15 372 data points [76]. The model was established by trying to combine two of the three common machine-learning algorithms namely Stepwise Multiple Linear Regression, Feed-Forward Artificial Neural Network (FFANN) and Least-Squares Support Vector Machine (LSSVM). Both internal and external validation techniques have been used to select the best model and check the model's stability. The UNIFAC fragmentation scheme is used to fragment cations and anions into its groups. The final model involves FFANN-based reference term for predicting the value of viscosity at reference temperature $T_0=298$ K, and LSSVM-based correction term for predicting the effect of T other than T_0 . Using all the data in the training phase this model shows an overall average absolute relative error (AARE) between experimental and calculated of 32.3%. Although the results obtained by Paduszyński seem to be the limit that can currently be achieved, one should note that their model was derived on the basis on extensive databases. In same year, Chen et al. [72] presented a series of simple GC models estimating various properties of ILs including viscosity. The mathematical formulation of the viscosity model is as follows:

$$\ln \frac{\eta}{R_{0\eta}} = A_{\eta} + B_{\eta} \frac{100}{T} + C_{\eta} \left(\frac{100}{T} \right)^2 \quad (9)$$

$R_{0\eta}$ is an adjustable parameter and A_η, B_η and C_η are parameters specific to each ionic liquid that are estimated from GC. This equation was developed using only 1090 data points for 76 ILs covering 36 cations and 16 anions, of which 778 data points (56 ILs) were used in the regression and the remaining data were used for testing. The values of AARE between calculated and experimental viscosity, reported in ln-units, were 3.36% and 5.63% for training and testing, respectively. It is worth mentioning that authors have used a limited set of complex functional groups that do not allow describing a wide variety of possible ILs. Indeed, each anionic structure of the 16 anions is proposed as a functional group, and to describe the 36 cations 8 complex functional groups are proposed along three substituted groups to describe the effect of side chain length.

I.8 CONCLUSION

In conclusion, this chapter has provided a comprehensive overview of the theoretical concepts, experimental techniques, and predictive methods relevant to the study of viscosity in pure organic compounds, covering both non-ionic and ionic compounds. We have explored the fundamental principles underlying viscosity, elucidated the various experimental methods used for viscosity measurement, and reviewed the different approaches for viscosity prediction. In the following chapters, we will delve deeper into specific predictive models and their applications, building upon the foundational knowledge established here.

I.9 BIBLIOGRAPHY

- [1] Tough, J. T. McCormick, W. D. Dash, J. G. « Viscosity of liquid He II,» *Physical Review*, vol. 132, n° 16, p. 2373, 1963.
- [2] Assael, M. J. Dalaouti, N. K. Vesovic, V. « Viscosity of natural-gas mixtures: measurements and prediction,» *International journal of thermophysics*, vol. 22, pp. 61-71, 2001.
- [3] P.S. van der Gulik, R. Mostert and H.R. van den Berg « The viscosity of methane at 273 K up to 1 GPa,» *Fluid phase equilibria*, vol. 79, pp. 301-311, 1992.
- [4] R. L. Alan S. Morris, «Chapter 21 - Summary of Other Measurements,» chez *Measurement and Instrumentation (Second Edition)*, Academic Press, 2016, pp. 633-672.
- [5] M. NIZAM, «MEASUREMENT OF VISCOSITY OF DIFFERENT TYPES OF FLUID,» Fluid mechanics (code: CLB 11003) , University of Kuala Lumpur, 2024.
- [6] Poling, B. E., Prausnitz, J. M., John Paul, O. C., & Reid, R. C, The properties of gases and liquids, McGraw-Hill, New York,: The McGraw-Hill Companies, Inc, 1977.
- [7] Ely, J. F., & Hanley, H. J. M, «Prediction of transport properties.1 Viscosity of fluids and mixtures,» *Industrial & Engineering Chemistry Fundamentals*, vol. 20, n° 14, pp. 323-332, 1981.
- [8] F. A. Pedersen K.S., « Viscosity of Crude Oil,» *Chem. Eng. Sci*, pp. 39, 6, 1011-1016, 1984.
- [9] Pedersen, K. S., & Fredenslund, A. A. G. E, «An improved corresponding states model for the prediction of oil and gas viscosities and thermal conductivities,» *Chemical Engineering Science*, vol. 42, n° 11, pp. 182-186, 1987.
- [10] Teja, A. S., & Rice, P, «Generalized corresponding states method for the viscosities of liquid mixtures,» *Industrial & Engineering Chemistry Fundamentals*, vol. 20, n° 11, pp. 77-81, 1981.
- [11] Aasberg-Petersen, K., Knudsen, K., & Fredenslund, A., «Prediction of viscosities of hydrocarbon mixtures,» *Fluid Phase Equilibria*, vol. 70, n° 12-3, pp. 293-308, 1991.
- [12] Huang, Y., Dong, H., Zhang, X., Li, C., & Zhang, S, «A new fragment contribution-corresponding states method for physicochemical properties prediction of ionic liquids,» *Thermodynamics and Molecular-Scale Phenomena*, vol. 59, n° 14, pp. 1348-1359, 2013.
- [13] Zhao, Y., Zhang, X., Deng, L., & Zhang, S., «Prediction of viscosity of imidazolium-based ionic liquids using MLR and SVM algorithms,» *Computers & Chemical Engineering*, vol. 92, pp. 37-42, 2016.

- [14] Quiñones-Cisneros, S. E., Zéberg-Mikkelsen, C. K., & Stenby, E. H., « The friction theory (f-theory) for viscosity modeling,» *Fluid Phase Equilibria*, vol. 169, n° %12, pp. 249-276, 2000.
- [15] Shen, G., Held, C., Mikkola, J. P., Lu, X., & Ji, X. , «Modeling the Viscosity of Ionic Liquids with the Electrolyte Perturbed-Chain Statistical Association Fluid Theory,» *Industrial & Engineering Chemistry Research*, vol. 53.52, p. 20258–20268, 2014.
- [16] Abolala, M., Peyvandi, K., & Varaminian, F «Modeling the viscosity of pure imidazolium-based ionic liquids using SAFT-VR-Mie EoS,» *Fluid Phase Equilibria*, vol. 394, pp. 61-70, 2015.
- [17] Haghbakhsh, R., Parvaneh, K., & Shariati, A, «Viscosities of Pure Ionic Liquids Using Combinations of Free Volume Theory or Friction Theory with the Cubic, the Cubic Plus Association, and the Perturbed-Chain Statistical Associating Fluid Theory Equations of State at High Pressures,» *Industrial & Engineering Chemistry Research*, vol. 56, n° %18, p. 2247–2258, 2017.
- [18] R. Macías-Salinas, «Viscosity Modeling of Ionic Liquids Using the Friction Theory and a Simple Cubic Equation of State,» *Industrial & Engineering Chemistry Research*, vol. 57, n° %13, p. 1109–1120, 2018.
- [19] A. J. Batschinski, «Untersuchungen Aber die innere Reibnng der Flüssigkeiten. I,» *Zeitschrift für Physikalische Chemie*, vol. 84U, n° %11, pp. 643-706, 1913.
- [20] M. H. Cohen et D. Turnbull, «Molecular Transport in Liquids and Glasses,» *The journal of Chemical Physics*, vol. 31, n° %15, p. 1164–1169, 1959.
- [21] A. K. Doolittle, «Studies in Newtonian Flow. II. The Dependence of the Viscosity of Liquids on Free-Space,» *Journal of Applied physics*, vol. 22, n° %112, p. 1471–1475, 1951.
- [22] Llorell, F., Valente, E., Vilaseca, O., & Vega, L. F, «Modeling Complex Associating Mixtures with [Cn-mim][Tf2N] Ionic Liquids: Predictions from the Soft-SAFT Equation,» *The Journal of Physical Chemistry B*, vol. 115, n° %115, p. 4387–4398, 2011.
- [23] Shen, G., Held, C., Mikkola, J. P., Lu, X., & Ji, X, «Modeling the Viscosity of Ionic Liquids with the Electrolyte Perturbed-Chain Statistical Association Fluid Theory,» *Industrial & Engineering Chemistry Research*, vol. 53, n° %152, p. 20258–20268, 2014.
- [24] Yarranton, H. W., & Satyro, M. A, «Expanded Fluid-Based Viscosity Correlation for Hydrocarbons,» *Industrial & Engineering Chemistry Research*, vol. 48, n° %17, p. 3640–3648, 2009.
- [25] Satyro, M. A., & Yarranton, H. W, «Expanded fluid-based viscosity correlation for hydrocarbons using an equation of state,» *Fluid Phase Equilibria*, vol. 298, n° %11, pp. 1-11, 2010.

- [26] Loria, H., Motahhari, H., Satyro, M. A., & Yarranton, H. W, «Process simulation using the expanded fluid model for viscosity calculations,» *Chemical Engineering Research and Design*, vol. 92, n° 112, pp. 3083-3095, 2014.
- [27] Y. Rosenfeld, «Relation between the transport coefficients and the internal entropy of simple systems,» *Physical Review A*, vol. 15, n° 16, pp. 2545-2549, 1977.
- [28] Ashurst, W. T., & Hoover, W. G, « Dense-fluid shear viscosity via nonequilibrium molecular dynamics,» *Physical Review A*, vol. 11, n° 12, p. 658, 1975.
- [29] Enskog, D, «Kinetic theory of heat conductivity, viscosity and diffusion in certain condensed gases and liquids,» *Kgl Svenska Vetenskaps Akad Handl*, vol. 63, pp. 1-44, 1922.
- [30] Chapman, S., & Cowling, T. G, *The mathematical theory of non-uniform gases: an account of the kinetic theory of viscosity, thermal conduction and diffusion in gases*, Cambridge university press, 1990.
- [31] Baskaran, A., Dufty, J. W., & Brey, J. J, «Transport coefficients for the hard-sphere granular fluid,» *Physical Review E*, vol. 77, n° 13, p. 031311, 2008.
- [32] D. Chandler, «Rough hard sphere theory of the self-diffusion constant for molecular liquids,» *J. Chem. Phys*, vol. 62, p. 1358–1363, 1975.
- [33] J. H. DYMOND, « Corrections to the Enskog theory for viscosity and thermal conductivity,» *Physica B+ C*, vol. 144, n° 13, pp. 267-276., 1987.
- [34] Chung, T. H., Ajlan, M., Lee, L. L., & Starling, K. E, «Generalized multiparameter correlation for nonpolar and polar fluid transport properties,» *Industrial & Engineering Chemistry Research*, vol. 27, n° 14, p. 671–679, 1988.
- [35] Chung, T. H., Lee, L. L., & Starling, K. E, «Applications of kinetic gas theories and multiparameter correlation for prediction of dilute gas viscosity and thermal conductivity,» *Industrial & Engineering Chemistry Fundamentals*, vol. 23, n° 11, pp. 8-13, 1984.
- [36] Assael, M. J., Dymond, J. H., Papadaki, M., & Patterson, P. M., «Correlation and prediction of dense fluid transport coefficients. I. n-alkanes,» *Correlation and prediction of dense fluid transport coefficients. I. n-alkanes*, vol. 13, p. 269–281, 1992.
- [37] Bleazard, J. G., & Teja, A. S, «Extension of the Rough Hard-Sphere Theory for Transport Properties to Polar Liquids,» *Industrial & Engineering Chemistry Research*, vol. 35, n° 17, p. 2453–2459, 1996.
- [38] Ciotta, F., Trusler, J. M., & Vesovic, V, «Extended hard-sphere model for the viscosity of dense fluids,» *Fluid Phase Equilibria*, vol. 363, pp. 239-247, 2014.
- [39] V. V. Nicolas Riesco, «Extended hard-sphere model for predicting the viscosity of long-chain n-alkanes,» *Fluid Phase Equilibria*, vol. 425, pp. 385-392, 2016.

- [40] de Wijn, A. S., Riesco, N., Jackson, G., Martin Trusler, J. P., & Vesovic, V., «Viscosity of liquid mixtures: The Vesovic-Wakeham method for chain molecules,» *The Journal of Chemical Physics*, vol. 136, n° 17, 2012.
- [41] Assael, M. J., Dymond, J. H., Papadaki, M., & Patterson, P. M., «Correlation and prediction of dense fluid transport coefficients. I. n-alkanes,» *International Journal of Thermophysics*, vol. 13, p. 269–281, 1992.
- [42] Gaciño, F. M., Comuñas, M. J., Fernández, J., Mylona, S. K., & Assael, M. J., «Correlation and Prediction of Dense Fluid Transport Coefficients. IX. Ionic Liquids,» *International Journal of Thermophysics*, vol. 35, p. 812–829, 2014.
- [43] Hossain, M. Z., & Teja, A. S., «Correlation and Prediction of the Transport Properties of Ionic Liquids,» *International Journal of Thermophysics*, vol. 37, 2016.
- [44] Hosseini, S. M., Alavianmehr, M. M., & Moghadasi, J., «Transport properties of pure and mixture of ionic liquids from new rough hard-sphere-based model,» *Fluid Phase Equilibra*, vol. 429, pp. 266–274, 2016.
- [45] S. M. Hosseini, «Erratum to “Transport properties of pure and mixture of ionic liquids from new rough hard-sphere-based model,» *Fluid Phase Equilibra*, vol. 429, p. 266–274, 2016.
- [46] Akbari, F., Alavianmehr, M. M., Behjatmanesh Ardakani, R., & Mohammad-Aghaie, D., «Thermophysical properties of ionic liquids and their mixtures from a new equation of state,» *Ionics*, vol. 24, p. 1357–1369, 2018.
- [47] H. T. Davis, S. A. Rice et J. V. Sengers, «On the Kinetic Theory of Dense Fluids. IX. The Fluid of Rigid Spheres with a Square-Well Attraction,» *The Journal of Chemical Physics*, vol. 35, n° 16, p. 2210–2233, 1961.
- [48] I. L. McLaughlin et H. T. Davis, «Kinetic Theory of Dense Fluid Mixtures. I. Square-Well Model,» *The Journal of Chemical Physics*, vol. 45, n° 16, p. 2020–2031, 1966.
- [49] Du, L. G., & Guo, T. M., «A semi-theoretical viscosity model for non-polar liquids,» *The Chemical Engineering Journal*, vol. 47, n° 13, pp. 163–167, 1991.
- [50] Monnery, W. D., Svrcek, W. Y., & Mehrotra, A. K., «Viscosity: A critical review of practical predictive and correlative methods,» *The Canadian Journal of Chemical Engineering*, vol. 73, n° 11, pp. 3–40, 1995.
- [51] Galliero, G., Nieto-Draghi, C., Boned, C., Avalos, J. B., Mackie, A. D., Baylaucq, A., & Montel, F., «Molecular Dynamics Simulation of Acid Gas Mixtures: A Comparison between Several Approximations,» *Industrial & Engineering Chemistry Research*, vol. 46, n° 115, p. 5238–5244, 2007.
- [52] Zerón, I. M., Cueto-Mora, M., & Blas, F. J., «Transport properties of the square-well fluid from molecular dynamics simulation,» *Molecular Physics*, vol. 122, p. 19–20, 2024.
- [53] S. Handy, *Ionic Liquids - Current State of the Art*, Germany: first ed., IntechOpen, 2015.

- [54] Nieto-Draghi, C., Fayet, G., Creton, B., Rozanska, X., Rotureau, P., de Hemptinne, J. C., ... & Adamo, C; «A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes,» *Chemical Reviews*, vol. 115, n° %124, p. 13093–13164, 2015.
- [55] Katritzky, A. R., Chen, K., Wang, Y., Karelson, M., Lucic, B., Trinajstić, N., ... & Schüürmann, G, «Prediction of liquid viscosity for organic compounds by a quantitative structure–property relationship,» *Journal of Physical Organic Chemistry*, vol. 13, n° %11, pp. 80-86, 2000.
- [56] Bouarab, A. F., Harvey, J. P., & Robelin, C, «Viscosity models for ionic liquids and their mixtures,» *Physical Chemistry Chemical Physics*, n° %12, 2021.
- [57] Tochigi, K., & Yamamoto, H, «Estimation of Ionic Conductivity and Viscosity of Ionic Liquids Using a QSPR Model,» *The Journal of Physical Chemistry C*, vol. 111, n° %143, 2007.
- [58] Barycki, M., Sosnowska, A., Gajewicz, A., Bobrowski, M., Wileńska, D., Skurski, P., ... & Puzyn, T., «Temperature-dependent structure-property modeling of viscosity for ionic liquids,» *Fluid Phase Equilibria*, vol. 427, pp. 9-17, 2016.
- [59] Bini, R., Malvaldi, M., Pitner, W. R., & Chiappe, C, «QSPR correlation for conductivities and viscosities of low-temperature melting ionic liquids,» *Journal of Physical Organic Chemistry*, vol. 21, pp. 7-8, 2008.
- [60] Yu, G., Wen, L., Zhao, D., Asumana, C., & Chen, X, «QSPR study on the viscosity of bis(trifluoromethylsulfonyl)imide-based ionic liquids,» *Journal of Molecular Liquids*, vol. 184, pp. 51-59, 2013.
- [61] Nieto-Draghi, C., Fayet, G., Creton, B., Rozanska, X., Rotureau, P., de Hemptinne, J. C., ... & Adamo, C, «A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes,» *Chemical Reviews*, vol. 115, n° %124, p. 13093–13164, 2015.
- [62] 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*, vol. 80, pp. 326-333, 2012.
- [63] Zhang, S., Jia, Q., Yan, F., Xia, S., & Wang, Q, «Evaluating the properties of ionic liquid at variable temperatures and pressures by quantitative structure–property relationship (QSPR),» *Chemical Engineering Science*, vol. 231, p. 116326, 2021.
- [64] Ding, Y., Chen, M., Guo, C., Zhang, P., & Wang, J, «Molecular fingerprint-based machine learning assisted QSAR model development for prediction of ionic liquid properties,» *Journal of Molecular Liquids*, vol. 326, p. 115212, 2021.
- [65] Wadhwa, P., & Mittal, A, *Quantitative Structure-Property Relationship (QSPR) Modeling Applications in Formulation Development*, Singapore: Springer, Singapore, 2022.

- [66] Argoub, K., Benkouider, A. M., Yahiaoui, A., & Bagui, F, «Estimation and uncertainty analysis of standard enthalpy of formation in the liquid state by third-order-group-contribution method,» *Fluid Phase Equilibria*, vol. 520, p. 112644–112644, 2020.
- [67] Serat, F. Z., Benkouider, A. M., Yahiaoui, A., & Bagui, F, «Nonlinear group contribution model for the prediction of flash points using normal boiling points,» *Fluid Phase Equilibria*, vol. 449, p. 52–59, 2017.
- [68] Guella, S., Argoub, K., Benkouider, A. M., Yahiaoui, A., Kessas, R., & Bagui, F, «Artificial Neural Network-Group Contribution Method for Predicting Standard Enthalpy of Formation in the Solid State: C–H, C–H–O, C–H–N, and C–H–N–O Compounds,» *International journal of thermophysics*, vol. 36, n° %110-11, p. 2820–2832, 2015.
- [69] Benkouider, A. M., Kessas, R., Guella, S., Yahiaoui, A., & Bagui, F, «Estimation of the enthalpy of vaporization of organic components as a function of temperature using a new group contribution method,» *Journal of molecular liquids*, vol. 194, p. 48–56, 2014.
- [70] Argoub, K., Benkouider, A. M., Yahiaoui, A., Kessas, R., Guella, S., & Bagui, F, «Prediction of standard enthalpy of formation in the solid state by a third-order group contribution method,» *Fluid phase equilibria*, vol. 380, p. 121–127, 2014.
- [71] K. Paduszyński, «Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity,» *Industrial & engineering chemistry research*, vol. 58, n° %136, p. 17049–17066, 2019.
- [72] Chen, Y., Kontogeorgis, G. M., & Woodley, J. M, «Group Contribution Based Estimation Method for Properties of Ionic Liquids,» *Industrial & Engineering Chemistry Research*, vol. 58, n° %110, p. 4277–4292, 2019.
- [73] Baghban, A., Kardani, M. N., & Habibzadeh, S, «Prediction viscosity of ionic liquids using a hybrid LSSVM and group contribution method,» *Journal of Molecular Liquids*, vol. 236, p. 452–464, 2017.
- [74] Lazzús, J. A., & Pulgar-Villarroel, G, «A group contribution method to estimate the viscosity of ionic liquids at different temperatures,» *Journal of molecular liquids*, vol. 209, pp. 161–168,, 2015.
- [75] , «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*, vol. 54, n° %15, p. 1311–1324.
- [76] K. Paduszyński, «Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity,» *Industrial & engineering chemistry research*, vol. 58, n° %136, p. 17049–17066, 2019.
- [77] Cao, X., Gong, M., Tula, A., Chen, X., Gani, R., & Venkatasubramanian, V, «An Improved Machine Learning Model for Pure Component Property Estimation,,» *Engineering*, vol. 39, pp. 61-73, 2024.

- [78] Varmuza, K., Dehmer, M., & Bonchev, D, Statistical Modelling of Molecular Descriptors in QSAR/QSPR, Wiley-VCH Verlag GmbH & Co. KGaA, 2012.
- [79] Khan, A. U, «Descriptors and their selection methods in QSAR analysis: paradigm for drug design,,» *Drug Discovery Today,,* vol. 21, n° 18, pp. 1291-1302, 2016.
- [80] Wang, L., Ding, J., Pan, L., Cao, D., Jiang, H., & Ding, X, «Quantum chemical descriptors in quantitative structure–activity relationship models and their applications,,» *Chemometrics and Intelligent Laboratory Systems,,* vol. 217, p. 104384, 2021.
- [81] Rogers, D., & Hahn, M., «Extended-Connectivity Fingerprints,,» *Journal of Chemical Information and Modeling*, vol. 50, n° 15, p. 742–754, 2010.
- [82] Carhart, Raymond E., SMITH, Dennis H., et VENKATARAGHAVAN, RENGACHARI, «Atom pairs as molecular features in structure-activity studies: definition and applications,,» *Journal of Chemical Information and Computer Sciences*, vol. 25, n° 12, p. 64–73, 1985.
- [83] Pérez-Nueno, V. I., Rabal, O., Borrell, J. I., & Teixidó, J, «APIF: A New Interaction Fingerprint Based on Atom Pairs and Its Application to Virtual Screening,,» *Journal of Chemical Information and Modeling*, vol. 49, n° 15, p. 1245–1260, 2009.
- [84] R. Davronov et S. Kushmuratov, «Comparative analysis of QSAR feature selection methods,,» *AIP Conference Proceedings*, vol. 3004, n° 11, 2024.
- [85] Zebida, M. A., Argoub, K., Benkouider, A. M., Yahiaoui, A., Toubal, K., & Hachemaoui, A, «Machine learning coupled with group contribution for predicting the electrical conductivity of ionic liquids with experimental accuracy,,» *Fluid Phase Equilibria,,* vol. 579, pp. 0378-3812, 2024.
- [86] Seber, G. A, Nonlinear Regression, 1989 John Wiley & Sons, Inc, 1989.
- [87] Daubechies, I., DeVore, R., Fornasier, M., & Güntürk, C. S., «Iteratively reweighted least squares minimization for sparse recovery,,» *Pure and Applied Mathematics*, vol. 63, n° 11, pp. 1-38, 2010.
- [88] Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P, «Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling,,» *Journal of Chemical Information and Computer Sciences*, vol. 43, n° 16, 2003.
- [89] Zhang, X., Shen, H., Huang, T., Wu, Y., Guo, B., Liu, Z., ... & Ou, G, «Improved random forest algorithms for increasing the accuracy of forest aboveground biomass estimation using Sentinel-2 imagery,,» *Ecological Indicators*, vol. 159, 2024.

- [90] Wang, C., Wei, X., Jin, X., Li, J., & He, M., «Developing a two-grade model for the thermal conductivity of ionic liquids and their mixtures,» *Chemical Engineering Science*, vol. 290, p. 119881, 2024.
- [91] Fröba, A. P., Rausch, M. H., Krzeminski, K., Assenbaum, D., Wasserscheid, P., & Leipertz, A., « Thermal conductivity of ionic liquids: Measurement and prediction,» *Int. J. Thermophys*, vol. 31, p. 2059–2077, 2010.
- [92] Zhao, A. Z., & Garay, J. E., «High temperature liquid thermal conductivity: A review of measurement techniques, theoretical understanding, and energy applications,» *Prog. Mater. Sci*, vol. 139, p. 101180, 2023.
- [93] Haghbakhsh, R., & Raeissi, S., «A novel correlative approach for ionic liquid thermal conductivities,» *J. Mol. Liq*, vol. 236, p. 214–219, 2017.
- [94] Soares, L. H., Guirardello, R., & Rolemberg, M. P., « A simple group contribution model to predict thermal conductivity of pure ionic liquids,» *Chem. Eng. Trans*, vol. 74, p. 1195–1200, 2019.
- [95] Almeida, R. M., Lourenço, M. J. V., & de Castro, C. N., «The thermal conductivity of ionic liquids. Experiment and molecular interpretation,» *J. Mol. Liq*, vol. 397, 2024.
- [96] Yang, H., Gallagher, R. C., Phan, A. T., Chartrand, P., & Gheribi, A. E., « A predictive approach for the compositional and temperature representation of thermal conductivity in multicomponent molten salt systems for advanced energy applications,» *Mater. Today Energy*, vol. 38, p. 101441, 2023.
- [97] Wan, R., Li, M., Song, F., Xiao, Y., Zeng, F., Peng, C., & Liu, H., « Predicting the Thermal Conductivity of Ionic Liquids Using a Quantitative Structure–Property Relationship,» *Ind. Eng. Chem. Res*, vol. 61, p. 12032–12039, 2022.
- [98] Kazakov, A. M. J. W., Magee, J. W., Chirico, R. D., Paulechka, E., Diky, V., Muzny, C. D., ... & Frenkel, M., «NIST Standard Reference Database 147: NIST Ionic Liquids Database-(ILThermo),» National Institute of Standards and Technology. Gaithersburg MD, v2.0, 2021. [En ligne]. Available: <http://ilthermo.boulder.nist.gov>.
- [99] Snoek, J., Larochelle, H., & Adams, R. P., *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, Nevada: Curran Associates Inc, 2012.
- [100] Gelbart, M. A., Snoek, J., & Adams, R. P., *Proceeding of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, Quebec City, Quebec, Canada: AUAI Press, 2014.
- [101] Han, C., Yu, G., Wen, L., Zhao, D., Asumana, C., & Chen, X
, «Data and QSPR study for viscosity of imidazolium -based ionic liquids,» *Fluid Phase Equilibra*, vol. 300, p. 95, 2011.
- [102] Valderrama, J. O., Cardona, L. F., & Rojas, R. E., « Correlation and prediction of ionic liquids viscosity using Valderrama-Patel-Teja cubic equation of state and the geometric

- similitude concept, Part I: Pure ionic liquids,» *Fluid Phase Equilibra*, vol. 497, p. 164, 2019.
- [103] He, M., Zhu, C., & Liu, X., « Estimating the viscosity of ionic liquid at high pressure using Eyring's absolute rate theory,» *Fluid Phase Equilib*, vol. 458, p. 170, 2018.
- [104] Darabi, L., & Zare, M., «High correlate simple equations for temperature and pressure dependence of the viscosity of ionic liquids,» *Chem. Phys*, vol. 539, 2020.
- [105] Yan, F., He, W., Jia, Q., Wang, Q., Xia, S., & Ma, P., « Prediction of ionic liquids viscosity at variable temperatures and pressures,» *Chem. Eng. Sci.*, vol. 184, p. 134, 2018.
- [106] Kang, X., Zhao, Z., Qian, J., & Muhammad Afzal, R., «Predicting the Viscosity of Ionic Liquids by the ELM Intelligence Algorithm,» *Ind. Eng. Chem. Res*, vol. 56, p. 11344, 2017.
- [107] Sun, Y., Chen, M., Zhao, Y., Zhu, Z., Xing, H., Zhang, P., ... & Ding, Y., «Machine learning assisted QSPR model for prediction of ionic liquid's refractive index and viscosity: The effect of representations of ionic liquid and ensemble model developement,» *J. Mol. Liq*, vol. 333, 2021.
- [108] T. Mezger, *The Rheology Handbook*, Vincentz Network: 3rd revised ed. Hanove, 2011.
- [109] M. Rao, « Rheology of liquid foods,» *Journal Tenure studies*, p. pp. 135 – 168., 1997.
- [110] T. C. Scott et R. Wham, «Surface area generation and droplet size control in solvent extraction systems utilizing high intensity electric fields,» *U.S. Patent 4767515.12*, 1988.
- [111] Ptasinski, K. J., & Kerkhof, P. J. A. M., «Electric field driven separations: Phenomena and applications.,» *Sep. Sci. Technol.*, 1992.
- [112] H. Xie, L. Li, Y. Sun, Y. Wang, S. Gao, Y. Tian, X. Ma, C. Guo, F. Bo et L. Zhang, « An Available Strategy for Nasal Brain Transport of Nanocomposite Based on PAMAM Dendrimers via In Situ Gel,» *Nanomaterials*, pp. 9, 147, 2019.
- [113] B. Vigani, S. Rossi, G. Milanesi, M. Bonferoni, G. Sandri, G. Bruni et F. Ferrari, « Electrospun Alginate Fibers: Mixing of Two Different Poly (ethylene oxide) Grades to Improve Fiber Functional Properties.,» *Nanomaterials*, pp. 8, 971, 2018.
- [114] P. Mishra, S. Mukherjee, S. Nayak et A. Panda, «A brief review on viscosity of nanofluids,» *Int. Nano Lett*, pp. 4, 109–120., 2014.
- [115] A. Einstein, «Eineneuebestimmung der molekuldimensionen.,» *Ann. Phys*, pp. 324, 289–306., 1906.
- [116] T. Phuoc et M. Massoudi, « Experimental observations of the effects of shear rates and particle concentration on the viscosity of Fe₂O₃–deionized water nanofluids.,» *Int. J. Therm. Sci*, pp. 48, 1294–1301, 2009.
- [117] Beggs, H. D., & Robinson, J. R., « Estimating the Viscosity of Crude Oil Systems,» *J. Petrol. Technol*, pp. 9, 1140-1141, 1995.

- [118] F. F. Petrosky G.E., «Viscosity Correlations for Gulf of Mexico Crudes Oils,,» 1995.
- [119] Labedi, R, « Improved Correlations for Predicting the Viscosity of Light Crudes,» *J. Petrol. Sci. Eng.*, pp. 8, 221-234, 1992.
- [120] Khan, S. A., Al-Marhoun, M. A., Duffuaa, S. O., & Abu-Khamsin, S. A, «Viscosity Correlations for Saudi Arabian Crude Oils,» p. 15 720., 1987.
- [121] V. d. W. J.D, Leyden, 1873.
- [122] Lohrenz, J., Bray, B. G., & Clark, C. R, « Calculating Viscosities of Reservoir Fluids from their Composition,» *J. Petrol. Technol*, pp. 1171-1176, 1964.
- [123] K. H. Little J.E., «A Correlation of the Viscosity of Hydrocarbon Systems with Pressure, Temperature and Composition,,» *Soc. Petrol. Eng. J. AIME,,* pp. 243, 157-162, 1968.
- [124] Guo, X. Q., Wang, L. S., Rong, S. X., & Guo, T. M, « Viscosity Model Based on Equations of State for Hydrocarbon Liquids and Gases,» *Fluid Phase Equilib*, pp. 139, 405-421, 1997.
- [125] T. A. Patel N.C., « A New Cubic Equation of State for Fluids and Fluid Mixtures,» *Chem. Eng. Sci*, pp. 37, 463-473., 1982.
- [126] Ungerer, P., *Thermodynamics: Applications in Chemical Engineering and the Petroleum Industry,,* paris: Ed. Technip, Paris, 2003.
- [127] 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*, vol. 80, p. 326–333, 2015.
- [128] Zhang, S., Jia, Q., Yan, F., Xia, S., & Wang, Q., «Evaluating the properties of ionic liquid at variable temperatures and pressures by quantitative structure–property relationship (QSPR),» *Chem. Eng. Sci*, vol. 231, 2021.

CHAPITRE II: METHODOLOGY AND MODELLING METHODS

II.1 INTRODUCTION

This chapter serves as a foundational guide, delving into various methodologies and modeling methods essential for understanding the intricate process of model development. Beginning with an introduction, the chapter navigates through the significance of databases in modeling, emphasizing the importance of data revision to ensure accuracy and reliability. It then delves into group contribution methods, elucidating the principles behind these techniques and the hierarchical structure of groups utilized in fragmentation schemes.

The chapter further explores Quantitative Structure-Property Relationships (QSPR), shedding light on the fundamental principles, molecular descriptors involved and fingerprints in predictive modeling. Dimensionality reduction techniques, transformation methods, feature selection, and sparse regression techniques are also examined in detail.

Next, the chapter elucidates conventional models and machine learning algorithms utilized in viscosity prediction, including Random Forest, Extreme Gradient Boosting, Non-linear Support Vector Machine, and Neural Network. Finally, the chapter concludes with a discussion on model validation, covering statistical validation parameters, cross-validation, and external validation.

II.2 DATABASE IMPORTANCE

The databases used for the construction of an empirical or semi-theoretical model must be composed of reliable and precise experimental data. The errors associated with these data propagate into the final model, since the model parameters are adjusted based on these data. Thus, high-quality data with minimal uncertainties are necessary to limit experimental errors and ensure the statistical robustness of the model.

Additionally, the data must be obtained following a standardized experimental protocol to ensure their validity and comparability. Experimental conditions, such as temperature and pressure, can have a significant influence on the measured property values. Therefore, it is crucial to precisely define the experimental conditions for each data point included in the database.

A quality database in empirical or semi-theoretical models must also be representative of the diversity of values that the studied property can take. This allows determining the applicability domain of the final model, i.e., the range of values for which the model is reliable and accurate. A broader applicability domain allows the model to accurately predict a greater variety of values.

Furthermore, the distribution of data in the database should be as homogeneous and normal as possible. The statistical methods used to develop empirical or semi-theoretical models are generally based on normal distributions, and outliers can skew the model results.

The choice of molecules included in the database is also crucial for the performance of the empirical or semi-theoretical model. It is recommended to include molecules with unique mechanisms and diverse molecular structures. This allows obtaining a more precise and generalizable model. However, including molecules with similar structures can also improve the model's performance, even if it may reduce its applicability domain.

II.2.1 Data revision

Data preprocessing plays a crucial role in the development of predictive models. This step aims to enhance data quality by eliminating errors, reducing noise, and highlighting important data features. An important aspect of data preprocessing is managing outliers, which can distort model results. These outliers were likely due to impurities in the samples, purification procedures, or the diversity of experimental techniques used to measure the viscosity of organic compounds.

In the context of this thesis, we have developed an innovative procedure for data preprocessing and outlier detection. It mainly consists of seven consecutive steps.

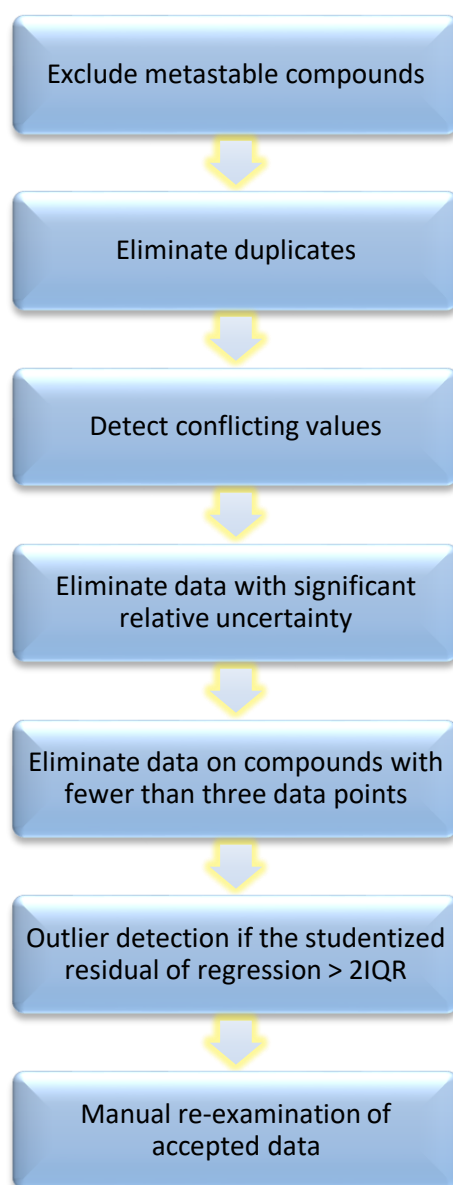


Figure 12 : Data pre-processing algorithm

The first step involved excluding metastable compounds. The second step aimed to remove duplicates in each dataset. The third step involved identifying values exhibiting contradictory behavior, i.e., multiple values for the same liquid at identical temperatures and pressures. In such cases, a single value representing the average of the values was retained. The fourth step aimed to eliminate data with a relative uncertainty exceeding 100%, while the fifth step aimed to eliminate data for compounds with fewer than three data points. In the sixth step, a regression was performed on the remaining data for each compound at atmospheric pressure (100-101,325 kPa) using the Vogel-Fulcher-Tammann equation. Subsequently, an outlier was considered as any data point with a studentized residual exceeding two times the interquartile range (IQR), and this step was

iterated until no outliers were detected. In the final step of this process, the data accepted through this process were manually reviewed to eliminate potentially inappropriate data.

II.3 GROUP CONTRIBUTION METHODS

II.3.1 Principle

Group contribution methods represent a powerful approach for predicting the viscosity of organic compounds. These methods operate on the principle that the viscosity of a compound can be approximated by summing contributions from its constituent functional groups or molecular fragments. By assigning specific contributions to individual groups based on their structural and chemical characteristics, group contribution methods offer a systematic approach to predict viscosity across a wide range of organic molecules.

The essence of group contribution methods lies in expressing the viscosity of a compound as a linear combination of contributions from various functional groups or molecular fragments present in the molecule. Each functional group is attributed a characteristic contribution value, typically derived from experimental data or theoretical calculations. The overall viscosity of the compound is then calculated by summing the contributions from all constituent groups.

The determination of contribution values for each functional group entails empirical correlations with experimental viscosity data for a diverse set of compounds. Statistical techniques such as regression analysis are commonly employed to establish relationships between the contributions of individual groups and the overall viscosity of the compounds. Furthermore, quantum chemical calculations and molecular simulations may be utilized to refine contribution values and enhance predictive accuracy.

Group contribution methods find widespread application in predictive modeling of viscosity due to their simplicity and versatility. These methods enable the estimation of viscosity for compounds where experimental data is limited or unavailable, making them invaluable tools for chemical engineering, process design, and formulation development. Moreover, group contribution methods can be integrated into larger computational models to predict the viscosity of complex mixtures or polymer systems.

Despite their advantages, it is important to recognize the limitations of group contribution methods. The accuracy of predictions heavily relies on the quality and completeness of the contribution values assigned to individual groups. Additionally, the applicability of group contribution methods may be constrained to specific classes of compounds or under certain operating conditions. Hence, careful consideration should be given to the selection of appropriate contribution values and validation of predictive models against experimental data.

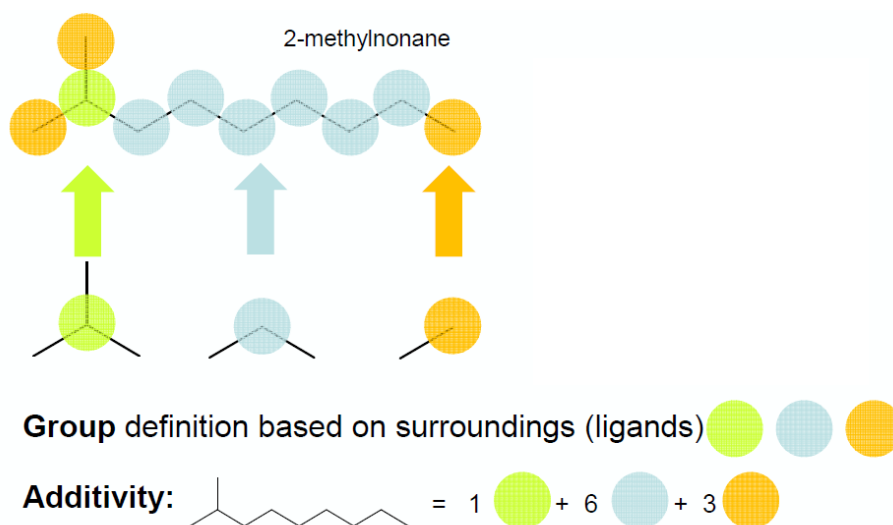


Figure 13 : Example of the cutting of 2-methylnonane using a first-order cutting scheme

II.3.2 Group's hierarchy

Group contribution methods are based on the principle that the overall property of a compound can be estimated by summing up the contributions of individual functional groups present in the molecule. The accuracy of these methods largely depends on the type of cutting scheme used to identify the functional groups.

II.3.2.A Zero-order fragmentation scheme

The zero-order cutting scheme represents the simplest form, where the molecule is fragmented into individual atoms, and the contribution of each atom to the overall property is considered. This approach provides a basic estimation of viscosity by merely quoting the number of atoms in the compound. While straightforward, the zero-order cutting scheme lacks the ability to capture the structural and chemical interactions between atoms within the molecule.

II.3.2.B First-order fragmentation scheme

The first order cutting scheme, exemplified by the Joback method [], takes a step further by considering functional groups instead of individual atoms. A functional group is a specific group of atoms within a molecule that is responsible for a certain reactivity pattern. By considering functional groups, the first order cutting scheme can provide more accurate predictions than the zeroth order scheme. However, it still does not consider the interactions between different functional groups.

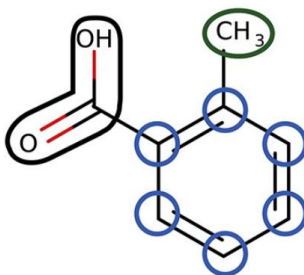


Figure 14 : Example of the cutting of 2-Methylbenzoic acid using a first-order cutting scheme

II.3.2.C Second-order fragmentation method

The second order cutting scheme, such as the Benson method, considers not only the functional groups but also their adjacent interactions. This scheme divides the molecule into substructures that include two or more connected functional groups. By considering the interactions between adjacent functional groups, the second order cutting scheme can provide even more accurate predictions.

In second-order fragmentation, a group is made up of a central atom with a valence strictly greater than 1, and a set of bonds that it forms with its neighbors. A group is denoted as follows:

$$X - (A)_i(B)_j(C)_k(D)_l$$

where X represents the central atom to which are bonded i atom(s) A , j atom(s) B , k atom(s) C and one atom(s) D .

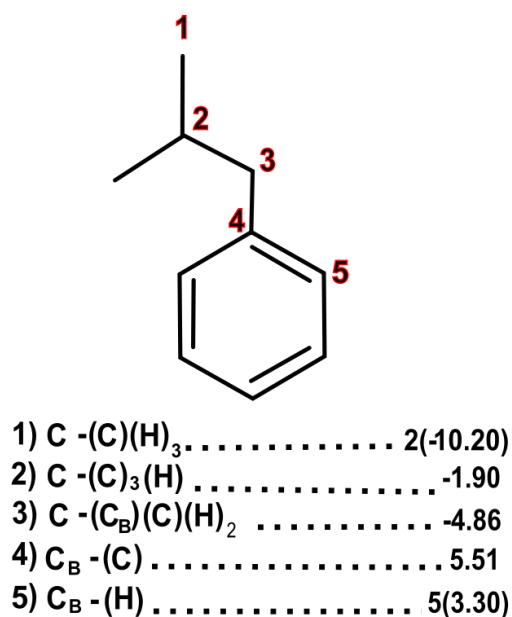


Figure 15 : Example of the cutting of 2-methylpropylbenzene using a second-order cutting scheme

II.3.2.D Third-order fragmentation method

The most widely used method in the field of group contribution methods and having one of the best molecular structure partitioning schemes is the third-order method developed by Marrero and Gani (Marrero, 2001) for estimating physicochemical and thermodynamic properties with a wide range of applications and dealing with both simple molecules and large multifunctional or even complex molecules.

The third-order fragmentation technique of a molecular structure is considered as a set of three types of groups: first-order groups, second-order groups and third-order groups.

The principle of this method is presented in the GC representation of the compound methoxylor (Figure 16). The molecular structure of methoxylor is depicted using a 424-dimensional vector, which is segmented into three group orders [77].

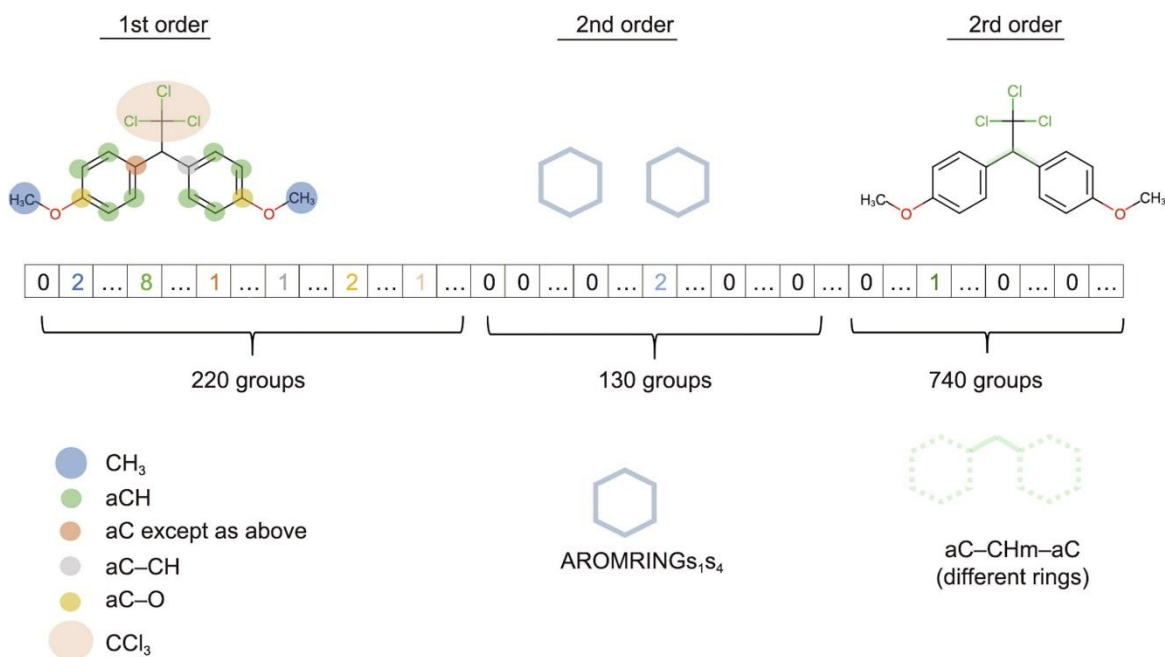


Figure 16 : Cutting of the methoxylor compound by the Marrero and Gan [77].

II.3.2.D.1 First-order groups

First order groups are numerous simple groups that allow the description of a wide variety of organic compounds, including aromatic and aliphatic hydrocarbons, alcohols, phenols, ketones, aldehydes, acids, esters, ethers, amines, anilines, and many other classes of organic compounds. This set of groups should also describe polyfunctional, large, complex, and heterocyclic compounds. The set of first order groups should allow the distinction between groups that appear in cyclic, aromatic, and acyclic structures. For example, first order groups distinguish between carbon atoms shared by different aromatic nuclei in the same fused system (e.g., naphthalene), carbon atoms shared by aromatic and non-aromatic nuclei in the same fused system (e.g., indane), and any other substituted aromatic carbon that does not belong in the previous categories, (e.g., biphenyl). First order groups must describe the entire molecule. In other words, there should not be a fragment of a given molecule that cannot be represented by first order groups. However, the proposed first order groups partially describe proximity effects and the difference between isomers. For this reason, the set of first order groups is intended to deal with simple and monofunctional compounds and other levels of groups are necessary to provide more accurate estimates for more complex complexes.

I.1.1.A.1 Second-order groups

Second-order groups allow for a better description of polyfunctional compounds and differentiation between isomers. These groups aim to describe molecular fragments that cannot be adequately represented by first-order groups, which leads to poor estimation at the first level. In this case, the second level utilizes groups that provide a better representation of polyfunctional compounds. For example, groups such as $OH - CH_n - COO$, $CH_m(NH_n) - COOH$ ou $NH_2 - CH_m - CH_n - COOH$ have been defined to offer a better description of lactates and amino acids, which typically result in poor estimations using only first-order groups. Other molecular fragments that are not fully described by first-order groups include alicyclic substituents, for which groups such as $CH_{cyc} - OH$, $CH_{cyc} - NH_2$, $CH_{cyc} - COO$, $CH_{cyc} - COOH$ ou $CH_{cyc} - SH$ have been included in the second-order set.

As suggested above, the set of second-order groups should facilitate differentiation between isomers. Consequently, specific groups are provided with this goal in mind. These groups allow for the distinction of isomers not only in alkenes, alkynes, and other open-chain structures, but also in aromatic compounds, for which special groups such as AROMARINGs1s2, AROMARINGs1s3 have been added.

However, the proposed second-order groups are unable to provide a good representation of compounds containing more than one cycle and, in some cases, polyfunctional open-chain compounds with more than four carbon atoms in the main chain. For this reason, an additional level is required to enable a better description of these types of compounds.

II.3.2.D.2 Third-order groups

These groups allow for a fairly detailed representation of fused aromatic ring systems, fused aromatic and non-aromatic ring systems, and unfused ring systems connected by chains where various functional groups may appear. The criteria used for identifying third-order groups are analogous to those used for second-order groups.

To illustrate the various cutting schemes mentioned above, two illustrative examples are provided.



Figure 17 : 1,9-nonadiol

The cutting of this molecule gives:

Table 1: Cutting the 1,9-nonadiol molecule into groups

Method	Groups	Number of contributions
Ordre 1	OH	2
	CH ₂	9
Ordre 2	Aucun group de second ordre	0
Ordre 3	HO-(CH _n) _m -OH (m > 2)	1

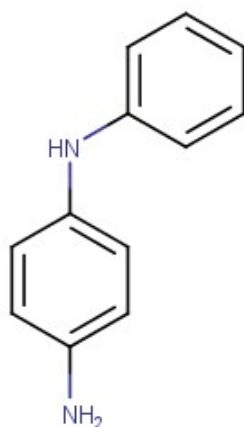


Figure 18 : N-Phenyl-1,4-benzenediamine

The cutting of this molecule gives:

Table 2: Cutting the N-Phenyl-1,4-benzenediamine molecule into groups

Method	Groups	Number of contributions
Ordre 1	aC	8
	aC	1
	aC-NH ₂	1
	aC-NH	1
Ordre 2	AROMARINGs1s4	1
Ordre 3	aC-NH-aC (different rings)	1

II.4 QUANTITATIVE STRUCTURES-PROPERTY RELATIONSHIPS

II.4.1 Principle

In the realm of computational chemistry and predictive modeling, the principle of Quantitative Structure-Property Relationship (QSPR) methods is to establish a mathematical relationship between molecular properties, encompassing both electronic and geometric aspects, known as descriptors, and a macroscopic observable such as physicochemical property, biological activity or toxicity. This relationship is formulated as follows:

$$\text{Property} = f(\text{Descriptors})$$

After the successful establishment and validation of this relationship on a validation dataset, it becomes feasible to employ it for predicting the properties of new molecules, even those lacking experimental data, or for molecules that have not yet been synthesized. Furthermore, in certain scenarios, such models can facilitate a deeper comprehension of the molecular phenomena involved in the property of interest.

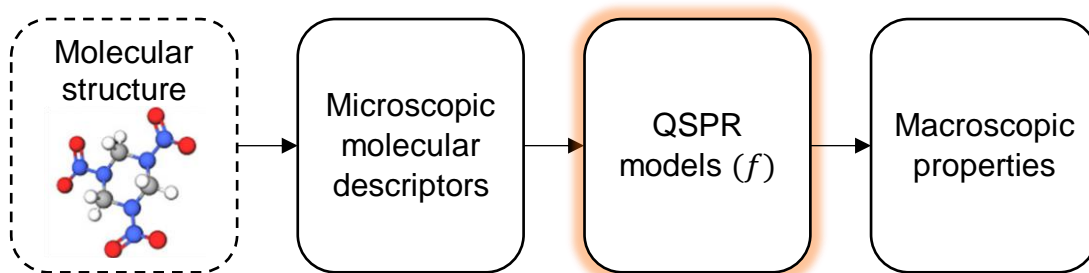


Figure 19 : Representation of QSPR methods

II.4.2 Molecular descriptors

Molecular descriptors play a crucial role in the quantitative modeling of relationships between the chemical structure of compounds and their physicochemical properties. These descriptors, also referred to as descriptive variables, represent various features and attributes of molecules, such as their size, shape, polarity, electric charge, and many others. The judicious use of these descriptors allows for capturing the relevant information necessary for the precise prediction of chemical compound properties.

For decades, research has focused on developing comprehensive molecular descriptors aimed at thoroughly describing molecular structures. Currently, the literature lists thousands of these descriptors (Karelson, 2012) [78]. For detailed information on these descriptors, reference can be made to Karelson's work (2012). The descriptors may be categorized as follows: constitutional, topological, constitutional and quantum [79].

II.4.2.A Constitutional descriptors

Constitutional descriptors, which are widely used because of their conceptual and computational simplicity, focus on the elemental composition and atomic connectivity of molecules. They take no account of geometric or electronic aspects. These descriptors include various parameters such as molecular mass, the number of atoms (C, H, O, N, etc.), the number of functional groups (CO, COOH, OH, NO₂, etc.) and the number of rings (whether polyatomic or not, aromatic or not)

These descriptors form the basis of contribution group approaches, where the effect of substituents is introduced as a tabulated value. Although they are used to obtain the simplest QSAR/QSPR models, they can pose problems for a proper interpretation of the interaction mechanisms at play for the property under study. For example, they do not allow isomers to be distinguished.

II.4.2.B Topological descriptors

Topological descriptors quantify the molecular structure by examining the spatial connections between atoms and bonds within a molecule. These descriptors capture important information about atomic connectivity and the topology of the molecule, which is crucial for modeling the chemical and physical properties of organic compounds.

Some examples of commonly used topological descriptors include:

1. Number of single, double, and triple bonds: This descriptor counts the number of single, double, and triple bonds present in the molecule, providing an indication of the structural complexity of the molecule.
2. Distances between atoms: Distances between atoms in a molecule provide information about the spatial geometry of the molecule and the proximity of different functional groups.

3. Topological indices: Topological indices, such as the Wiener index and the Randic index, are mathematical measures of the topological complexity of a molecule. They capture features such as symmetry, compactness, and connectivity of the molecule.
4. Number of cycles: This descriptor counts the number of cycles present in the molecule, providing information about its rigidity and conformational stability.

By using these topological descriptors, it is possible to quantitatively characterize the molecular structure and use them to predict various properties of organic compounds, including viscosity. These descriptors are widely used in the fields of computational chemistry, molecular modeling, and drug design to understand the structure-activity relationships of chemical compounds.

II.4.2.C Geometric descriptors

Geometric descriptors are parameters used to characterize the shape, size, and spatial arrangement of molecules. These descriptors provide valuable information about the three-dimensional geometry of molecular structures, which is essential for understanding their chemical and physical properties. Examples of geometric descriptors frequently employed in computational chemistry and cheminformatics encompass:

1. Surface area: This descriptor quantifies the total surface area of the molecule, providing insights into its accessibility for interactions with other molecules.
2. Volume: Volume descriptors measure the total volume occupied by the molecule in three-dimensional space, offering information about its overall size and bulkiness.
3. Radius of gyration: The radius of gyration is a measure of the distribution of mass around the center of mass of the molecule. It provides information about the molecule's compactness and flexibility.
4. Moments of inertia: Moments of inertia characterize the distribution of mass around the principal axes of rotation of the molecule. They offer insights into the molecule's rotational dynamics and symmetry.

5. Center of mass: This descriptor indicates the point in space where the mass of the molecule is concentrated, providing information about its overall spatial position.

Geometric descriptors are valuable tools for studying molecular conformation, molecular docking, and molecular dynamics simulations. By quantifying the geometric properties of molecules, these descriptors contribute to the development of predictive models for various chemical and biological processes.

II.4.2.D Quantum descriptors

Quantum descriptors refer to parameters derived from quantum mechanical calculations that provide insights into molecular properties, such as electronic structure, energy levels, and molecular orbitals. These descriptors are valuable for understanding the quantum behavior of molecules and predicting their chemical and physical properties.

Some examples of commonly used quantum descriptors include:

1. HOMO and LUMO energies: The Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) energies provide information about the electronic structure and stability of molecules.
2. Ionization potential (IP): The energy required to remove an electron from a molecule, representing its tendency to undergo ionization.
3. Electron affinity (EA): The energy change when an electron is added to a neutral molecule, indicating its ability to accept electrons.
4. Molecular orbital coefficients: These coefficients quantify the contribution of each atomic orbital to the overall molecular orbital, providing insights into electronic delocalization and bonding patterns.
5. Electronegativity: Derived from molecular orbital energies or atomic properties, electronegativity describes the ability of an atom to attract electrons in a molecule.
6. Dipole moment: Quantum calculations can determine the magnitude and direction of the dipole moment, which characterizes the polarity of a molecule.

7. Charge distribution: Quantum descriptors can reveal the distribution of electron density and partial charges within a molecule, influencing its reactivity and intermolecular interactions.

These quantum descriptors are crucial for understanding molecular behavior, reactivity, and interactions in various chemical and physical processes. An extensive compilation of three-dimensional quantum descriptors can be discovered within a multitude of bibliographic references [80], [78]. It should be noted that the field of quantum mechanics, particularly when it comes to three-dimensional quantum descriptors, requires the use of reliable and sophisticated calculation programmers.

II.4.3 Fingerprints in Predictive Modeling

In addition to molecular descriptors, fingerprints serve as indispensable tools in cheminformatics and quantitative structure-activity relationship (QSAR) studies, contributing significantly to predictive modeling tasks. Fingerprints represent molecular structures in a binary or categorical format, encoding structural patterns, substructures, and molecular properties. They play a crucial role in similarity searching, clustering, and classification tasks, complementing the information provided by molecular descriptors.

In this thesis, the utilization of different types of fingerprints alongside molecular descriptors will be explored to develop robust predictive models for viscosity prediction. By incorporating both types of representations, the models aim to capture a broader range of structural and property-based information.

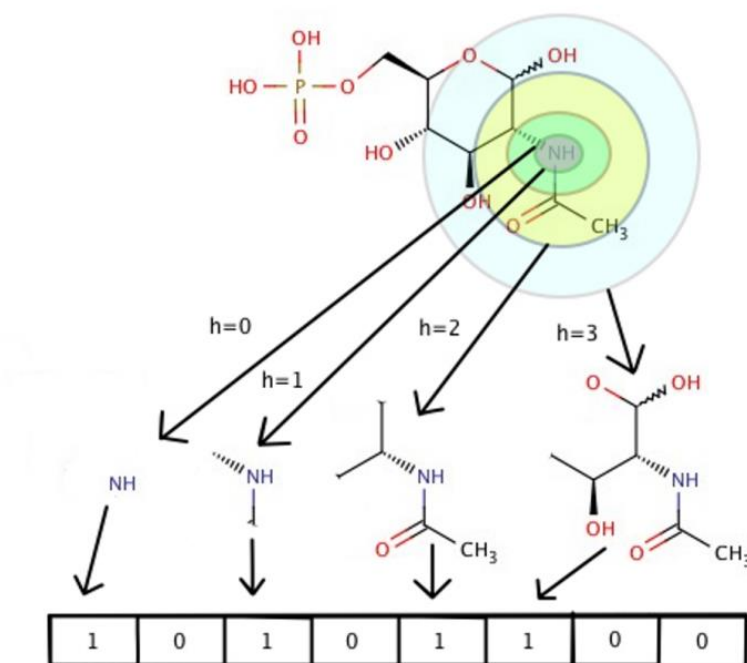
In scientific literature, multiple types of fingerprints are commonly employed in cheminformatics and QSAR/QSPR studies, including atom-pair fingerprints, Morgan fingerprints, MACCS Keys, and pharmacophore fingerprints. In this thesis, we focus on the most frequently utilized fingerprints in predictive modeling tasks, namely atom-pair fingerprints and Morgan fingerprints. For comprehensive details on these fingerprints, we recommend referring to the following publications [81] [82].

II.4.3.A Morgan fingerprints

Morgan fingerprints, also known as Extended Connectivity Fingerprints (ECFP), are circular fingerprints that encode structural information based on atom environments

within a specified radius [81]. ECFP fingerprints capture local and global structural features, making them versatile for various cheminformatics applications.

The principle of the Morgan Fingerprint algorithm involves the concept of a ‘radius’. The algorithm starts by assigning an initial identifier to each atom in the molecule, typically based on the atom’s type. Then, in each iteration or ‘radius’, the algorithm considers the circular neighborhood around each atom. This neighborhood includes the atom itself and its immediate neighbors in the first iteration, the neighbors of those neighbors in the second iteration, and so on. The size of the neighborhood expands with each iteration, hence the term ‘radius’.



II.4.3.B Atom pairs fingerprints:

Atom pairs fingerprints encode the presence or absence of pairs of atoms at a certain distance from each other in a molecule. They are efficient for capturing local structural features and are widely used in similarity searching and QSAR/QSPR modeling. The fingerprints can be represented as either bit vectors or count vectors. In a bit vector, each bit position corresponds to the presence (1) or absence (0) of a particular atom pair at a certain distance. In a count vector, each position stores the count of a particular atom pair at a certain distance [83].

This method provides a detailed representation of the molecule's structure, capturing the relative positions of atoms within the molecule. However, it's important to note that Atom Pairs Fingerprints do not consider the paths between atom pairs, only the distances.

II.4.4 Dimensionality Reduction of Descriptors [84]

When dealing with a multitude of descriptors, the potential for redundancy arises, leading to collinearity issues. It is crucial that the selected parameters are not only meaningful but also easily interpretable from a chemical standpoint. Descriptors are deemed more pertinent when they offer insights into underlying mechanisms. The aim is to construct QSPR models that are not only simple and transparent but also comprehensible from a phenomenological perspective.

At its core, the endeavor embodies the principle of parsimony, wherein simplicity reigns supreme among equally plausible explanations for a given phenomenon. However, this pursuit of simplicity need not entail an informational compromise. Rather, in QSPR modeling, it advocates for models to harbor minimal parameters while maximizing information encapsulated within the property under scrutiny.

Furthermore, a variable should only be retained if its exclusion significantly diminishes the model's performance. Care must be taken, however, to avoid losing crucial information during the model simplification process.

Various techniques exist to reduce the number of descriptors included in the models. These techniques fall into three primary categories: transformation methods, feature selection methods, and sparse regression methods. Each of these approaches plays a crucial role in streamlining the descriptor set and enhancing the interpretability and efficiency of QSPR models.

II.4.4.A Transformation Methods

The first class of dimensionality reduction techniques includes transformation methods such as Principal Component Analysis (PCA) and Partial Least Squares (PLS). PCA transforms the original set of descriptors into a new set of orthogonal variables called principal components, while PLS finds a set of latent variables that are optimized to maximize the covariance between the descriptors and the property of interest. Both PCA

and PLS reduce the dimensionality of the descriptor space while preserving as much information as possible, making them effective for QSPR modeling. In addition to PCA and PLS, several techniques can be considered. Independent Component Analysis (ICA) stands out for its ability to untangle a multivariate descriptors into independent components, thus reducing the descriptor space by pinpointing the most independent components that explain data variation. Non-negative Matrix Factorization (NMF) is another viable option, as it factors high-dimensional data into two matrices, with one matrix holding non-negative elements, facilitating feature extraction while preserving non-negativity constraints. Kernel PCA offers a non-linear extension to PCA by projecting data into a higher-dimensional space using kernel methods before applying PCA. Each of these techniques presents unique advantages and limitations, making the choice contingent upon the data's characteristics and the specific objectives of the study.

II.4.4.B Feature Selection

In addition to dimensionality reduction techniques like PCA and PLS, feature selection methods can also be used to identify the most relevant descriptors for QSPR modeling. These methods evaluate the importance of each descriptor individually and select a subset of descriptors based on their relevance to the property being predicted. Common feature selection approaches include Recursive Feature Elimination (RFE), which iteratively removes the least important descriptors, Univariate Feature Selection, which evaluates each feature individually using univariate statistical tests (e.g., t-test, ANOVA, Fisher Score) and selects features that are most strongly related to the target variable, and Stepwise Multiple Linear Regression, which combines forward selection and backward elimination, adding or removing features at each step based on a chosen criterion until no further improvement is observed.

II.4.4.C Sparse Regression Techniques

The third class of dimensionality reduction techniques includes Sparse Regression Techniques, which aim to select a subset of the most relevant descriptors while encouraging sparsity in the regression coefficients. These techniques, such as LASSO (Least Absolute Shrinkage and Selection Operator), Elastic Net and L1-Regularization, penalize the absolute size of the regression coefficients, effectively selecting a subset of descriptors with non-zero coefficients. Sparse regression techniques can help reduce

overfitting and improve the interpretability of QSPR models by identifying the most important descriptors for predicting the property of interest.

II.5 MODELS AND REGRESSION

Developing CG or QSPR models presents a significant challenge due to the inherent scale differences between molecular structures and macroscopic properties. Molecular structures are intricately detailed, existing at the molecular scale, while properties to be predicted reside at a macroscopic level. This disparity in scales complicates the process of correlating the two, requiring careful consideration and adaptation of modeling techniques [85].

Additionally, uncertainties play a crucial role in GC or QSPR modeling, arising from both molecular structures, influenced by the level of calculation precision, and experimental data, which can vary due to measurement protocols. These uncertainties add a layer of complexity to the modeling process, requiring robust methods to handle and account for them effectively.

Another critical aspect of GC or QSPR modeling is the management of large datasets. With a multitude of descriptors and molecules to analyze, selecting the most relevant structural parameters from the complete set available becomes a challenging task. While numerous tools and methodologies exist for this purpose, the key challenge lies in identifying the most suitable approach to derive a reliable model from the dataset at hand.

The selection of an appropriate modeling approach depends on the specific characteristics of the system under study. Given the variability in molecular structures and properties, it is essential to choose an approach that best characterizes the system's behavior. This selection process involves evaluating different modeling strategies to determine the most effective and reliable method for deriving meaningful insights from the available data.

The following paragraphs describe the main regression methods used in CG and QSPR modelling, which fall into two distinct categories: conventional methods and machine learning algorithms.

II.5.1 Conventional models

In the realm of predicting the viscosity of organic compounds, traditional or conventional models have been widely used. These include linear and non-linear regression techniques, which establish relationships between viscosity and variables such as temperature and chemical structure (descriptors or group contributions). The Arrhenius equation, Vogel-Fulcher-Tamman equation, and other empirical and semi-empirical models are often used to describe the temperature dependence of viscosity. These models are often based on mathematical non-linear relationships, such as power-law and exponential functions, to characterize the viscosity behavior of organic compounds. Techniques such as the Levenberg-Marquardt nonlinear least squares algorithm [86] and the Iteratively Reweighted Least Squares algorithm [87] are frequently applied to optimize model parameters and enhance predictive accuracy.

It is important to note that while these traditional models have been effective in many cases, they can struggle to capture the non-linear or non-ideal behaviour of some organic compounds, particularly complex or novel compounds. This has led to the exploration of other approaches, such as machine learning algorithms, to improve the accuracy of viscosity prediction.

II.5.2 Machine learning algorithms

In the field of predictive modeling for the physical and chemical properties of organic compounds, various machine learning algorithms have been employed to develop accurate and robust regression models. This section discusses four prominent machine learning methods used in this context: Random Forest (RF), Extreme Gradient Boosting (XGBoost), Nonlinear Support Vector Machine (SVM), and Neural Network (NN).

II.5.2.A Random Forest

Random Forest (RF) is a powerful ensemble learning method widely used for regression tasks. It belongs to the class of decision tree-based methods and offers several advantages over traditional regression techniques [88]. RF operates by constructing a multitude of decision trees during the training phase [89] (Figure 20). Each tree is built using a random subset of the training data and a random subset of the features using the bootstrap procedure, also known as bagging (bootstrap aggregation). This randomness helps to

reduce overfitting and improves the generalization of the model. Mathematically, RF can be expressed as:

$$y(x) = \left(\frac{1}{N}\right) \sum_{i=1}^N T_{i(x)} \quad (10)$$

where $y(x)$ is the predicted output for input x , $T_{i(x)}$ is the output of the i -th decision tree in the forest, and N is the total number of trees in the forest.

RF for regression have been successfully used for building non-linear GC models and QSPR in chemoinformatics. RF also has several adjustable parameters that can be set to optimise its performance. The optimal choice of a few hyper parameters that we have carefully considered are: *NumLearningCycles* and *MinLeafSize*.

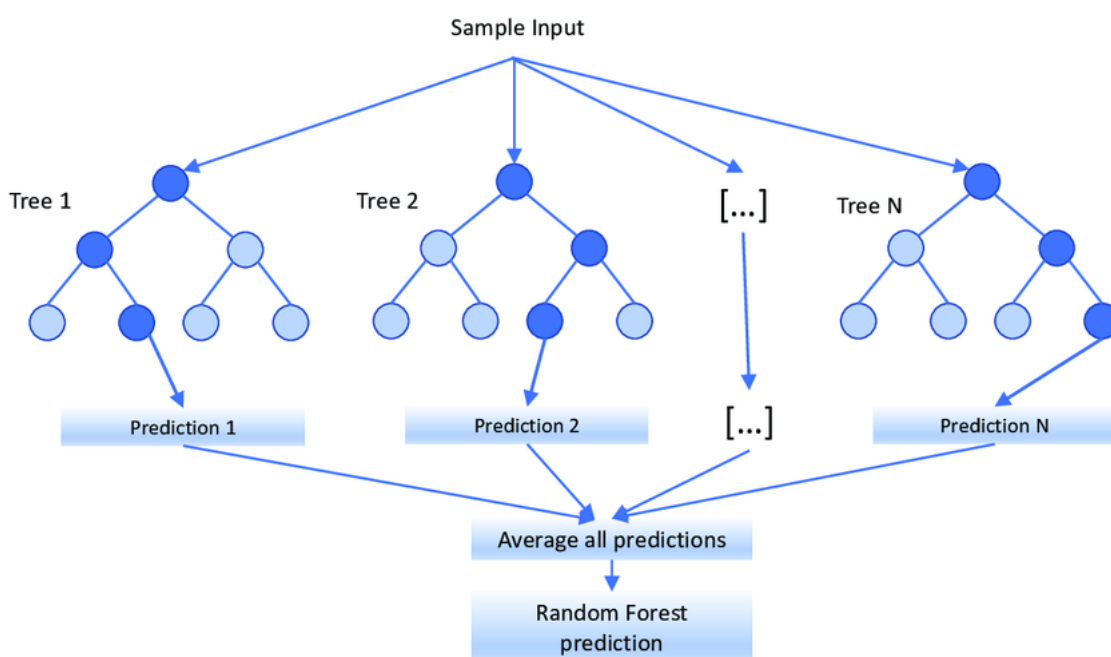


Figure 20 : Random Forest (RF) operating diagram.

II.5.2.B Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a decision-tree-based ensemble method that utilizes boosting, a technique for combining weak learners to form a strong learner. XGBoost has gained popularity due to its excellent performance and scalability.

Figure 21 depicts the fundamental structure of XGBoost, where X represents either the descriptor matrix in the context of the QSPR approach or the group appearance matrix in the context of the group contribution approach to this model, and y denotes the viscosity of a compound. In the context of boosting, Extreme Gradient Boosting (XGBoost) employs an additive training strategy. Here, each tree is built by learning from the residual δ of the preceding tree. The prediction of the k -th iteration, denoted as $\hat{y}_i^{(k)}$, is calculated as the sum of the prediction from the previous iteration, $\hat{y}_i^{(k-1)}$, and the output of the current tree, $f_k(x_i)$.

XGBoost aims to optimize the model at each iteration to reduce the prediction error. The final prediction, \hat{y}_i , is produced by combining the outputs of all trees using a weighted summation:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (11)$$

Here, F is the space of functions that includes all regression trees, and K is the total number of trees in the model. To train each tree and learn the function f_k , XGBoost defines an objective function with regularization:

$$\mathcal{L}(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^K \Omega(f_k) \quad (12)$$

where θ is the model parameter, n is the number of samples, $l(y_i, \hat{y}_i^{(t)})$ is the loss function that measures the difference between the predicted value $\hat{y}_i^{(t)}$ and the true value y_i , t is the number of trees, f_k is the k -th tree. The regularization term $\Omega(f_k)$ penalizes complex models by adding a penalty term to the objective function that increases as the complexity of the model increases. This term takes the form of:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^K \omega_j^2 \quad (13)$$

where T is the number of terminal nodes in the tree f_k , ω_j is the weight of the j -th terminal node, and γ and λ are hyperparameters that control the strength of the regularization. The first term penalizes the number of terminal nodes in the tree, while the second term penalizes the weights of the terminal nodes.

The algorithm works by iteratively adding decision trees to the model, with each tree learning from the errors of the previous trees. In each iteration, the algorithm calculates the gradient and Hessian of the loss function with respect to the predicted value, and constructs a new tree to minimize the objective function. The new tree is then added to the model, and the predicted values are updated. This process continues until a stopping criterion is met.

Despite XGBoost having numerous adjustable parameters, it can be demonstrated that a significant portion of these parameters are not overly sensitive to our modelling problem. However, the optimal choice of a few standard parameters can significantly improve the prediction accuracy of XGBoost-based GC models. Below are the hyper parameters that we have carefully considered:

- *MinLeafSize*: Minimum number of leaf node observations, also known as *min_child_weight*. Each leaf has at least *MinLeafSize* observations per tree leaf.
- *NumLearningCycles*: The number of rounds for boosting, also known as *num_round*.
- *LearnRate*: Learning rate for shrinkage also known as *learning_rate*, it controls the step size during boosting iterations.

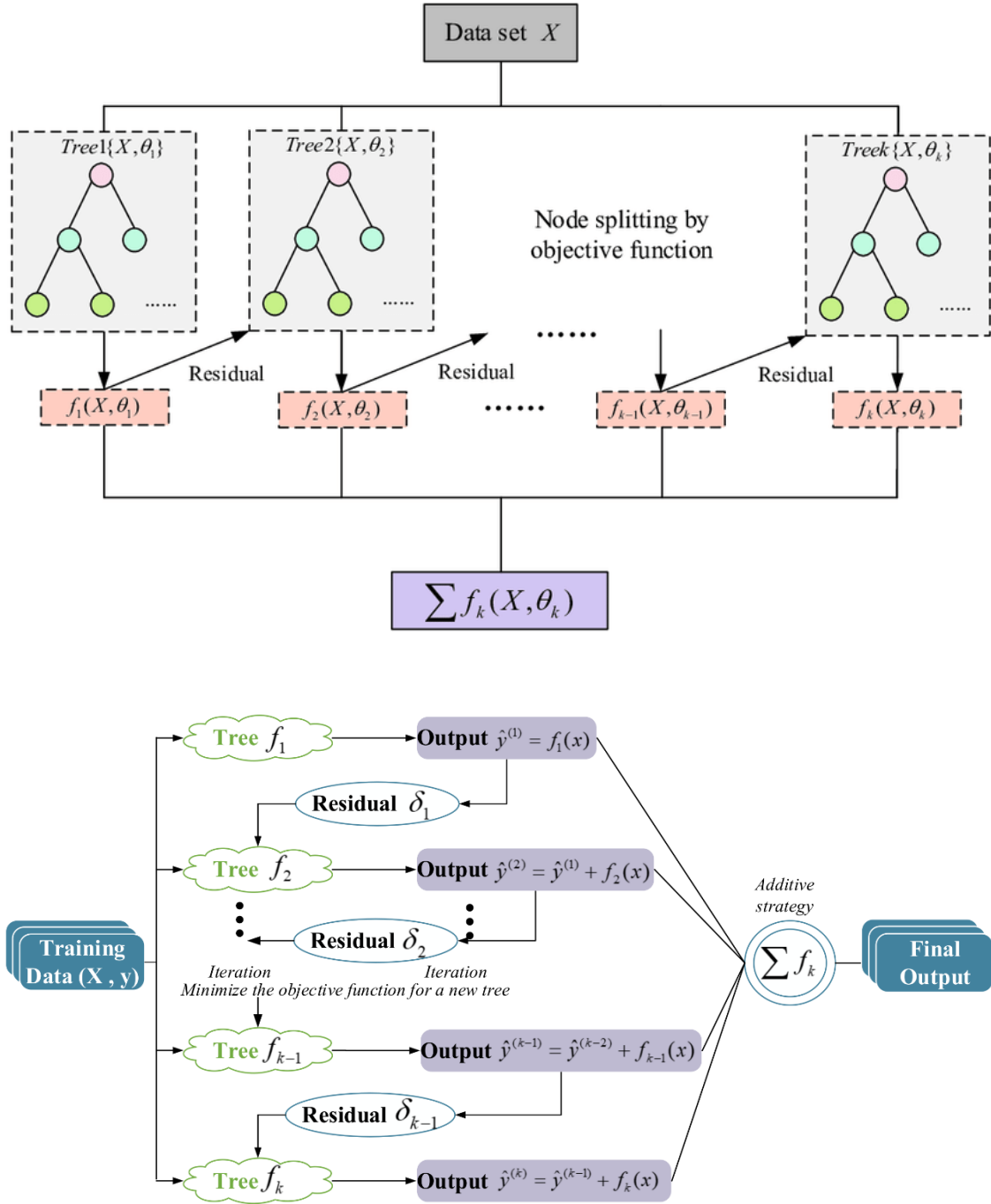


Figure 21 : XGBoost operating diagram.

II.5.2.C Non-linear Support Vector Machine

Nonlinear Support Vector Machine (SVM) regression models are a type of machine learning algorithm used for predicting continuous output values based on a set of input variables. These SVM regression models can capture complex, nonlinear relationships by using kernel functions to map the input variables to a higher-dimensional feature space.

The basic idea behind SVM regression is to find a hyperplane in the feature space that best separates the data into two regions, one for positive output values and one for negative output values. The distance between the hyperplane and the closest data points in each region, called support vectors, is maximized. The hyperplane can then be used to predict the output value for new input data points.

Given a training set $\{(x_1, y_1), \dots, (x_n, y_n)\}$, where $x_i \in R^d$ and $y_i \in R$, we want to learn a function $f: R^d \rightarrow R$ that can predict the output y for a new input x .

The decision function of the SVM regression model can be written as (Figure 22):

$$f(x) = b + \sum_{i=1}^n \alpha_i K(x, x_i) \quad (14)$$

where b is the bias term, $K(x, x_i)$ is the kernel function that maps the input data into a high-dimensional feature space, and α_i are the coefficients of the support vectors.

The kernel function $K(x, x_i)$ is defined as:

$$K(x, x_i) = \varphi(x) \cdot \varphi(x_i) \quad (15)$$

where $\varphi(x)$ is a feature mapping function that maps the input data x into a higher-dimensional feature space.

The coefficients α_i are obtained by solving the following optimization problem:

$$\begin{aligned} \min & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i y_i \\ \text{subject to: } & \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \in [0, C] \\ & |y_i - f(x_i)| \leq \varepsilon \end{aligned} \quad (16)$$

where C is a regularization parameter, it is also known as the BoxConstraint which controls the trade-off between maximizing the margin and minimizing the prediction error, and y_i is the output value for the input data point x_i . The ε parameter is commonly known as the insensitivity parameter in SVM regression, and its selection requires careful

consideration to balance the trade-off between model complexity and accuracy. In addition to C and ε , which serve as hyper-parameters for an SVM model, the kernel function and kernel scale are also used as hyper-parameters to fine-tune the SVM. The most commonly used kernel functions in GC and QSPR problems are the radial basis function (RBF). In this work, the RBF kernel was retained after testing several functions, given by:

$$K(x_n, x) = \exp(-\gamma \cdot \|x_n - x\|^2) \quad (17)$$

where γ is kernel scale parameter (hyper-parameter) of the RBF.

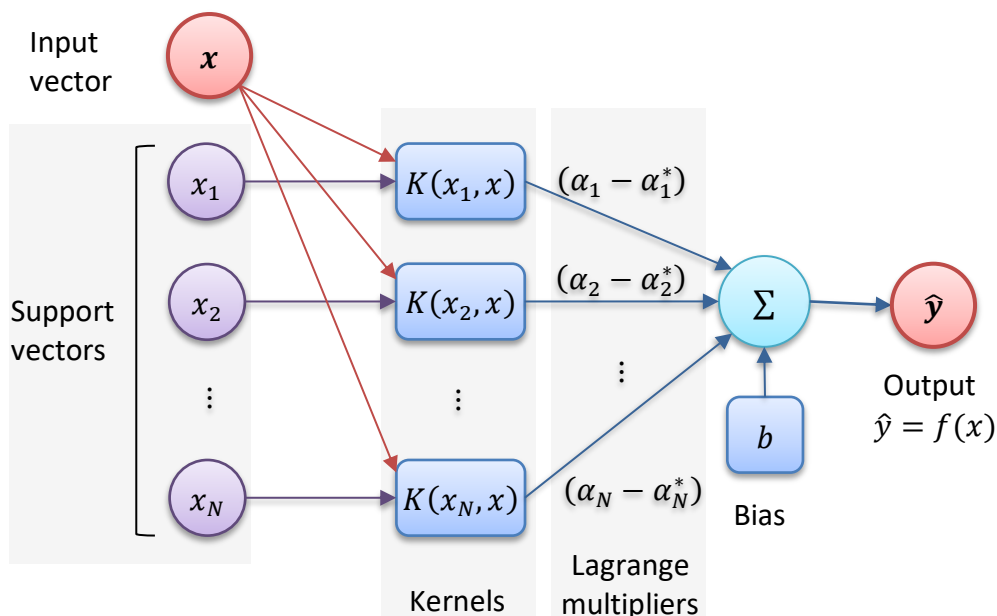


Figure 22 : Network architecture of a SVM regression

II.5.2.D Neural Network

Neural networks, also known as multi-layer perceptron algorithms, are a powerful tool for developing predictive models, including those for predicting the viscosity of organic compounds. MLPs are part of a broader class of artificial neural networks designed to mimic the information processing of animal brains (Figure 23), particularly in terms of learning from data.

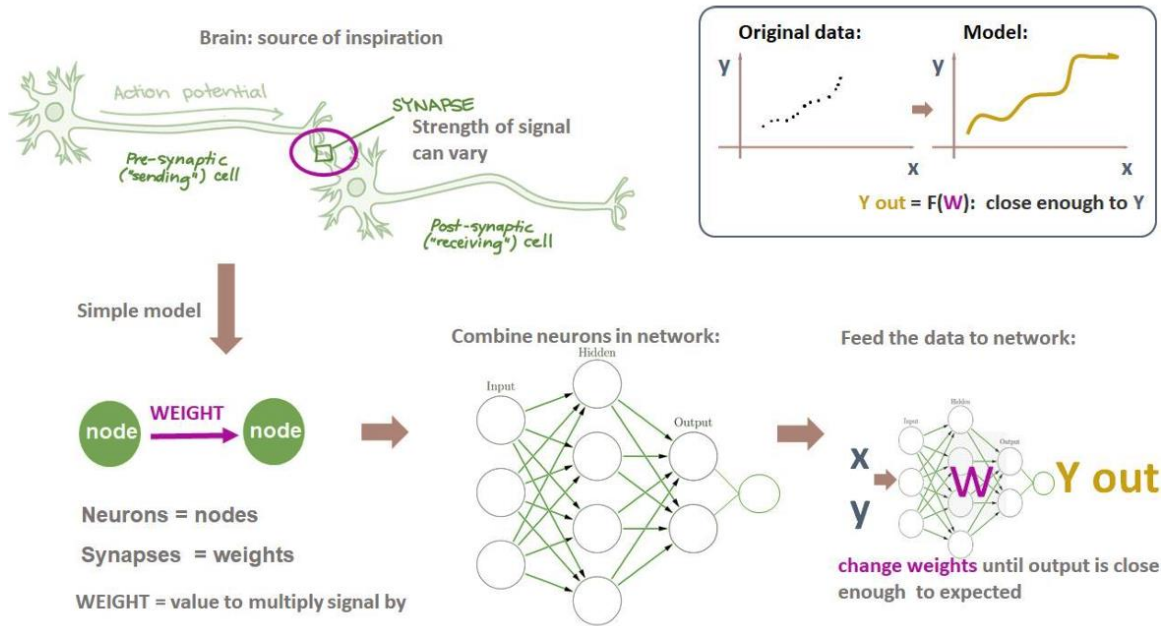


Figure 23 : illustrative representation of neural networks

An MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer contains multiple nodes (neurons), and each node is connected to nodes in the adjacent layers. The connections have associated weights and biases, which are adjusted during training to optimize the model's performance.

Let's consider a regression problem where we aim to predict the viscosity of organic compounds. For a single-layer MLP, the output of the network can be represented as:

$$y(x) = \sum_{i=1}^D w_i x_i + b \quad (18)$$

where x is the input vector representing the features of the compound, w_i are the weights associated with each feature, b is the bias term, and D is the number of features.

In a multilayer perceptron, the output of each node in a hidden layer is calculated as:

$$a_j^{(1)} = \sum_{i=1}^D w_{ji}^{(1)} x_i + b_j^{(1)} \quad (19)$$

where $a_j^{(1)}$ is the activation of node j in the first hidden layer, $w_{ji}^{(1)}$ are the weights connecting the input layer to the first hidden layer, $b_j^{(1)}$ is the bias term for node j , and x_i are the inputs from the input layer.

The activation $a_j^{(1)}$ is then transformed using a nonlinear activation function $f(\cdot)$ to produce the output of the hidden layer:

$$z_j^{(1)} = f(a_j^{(1)}) \quad (20)$$

For the output layer, the process is similar, with the output of each node in the output layer calculated as:

$$a_k^{(2)} = \sum_{j=1}^m w_{kj}^{(2)} z_j^{(1)} + b_k^{(2)} \quad (21)$$

where $a_k^{(2)}$ is the activation of node k in the output layer, $w_{kj}^{(2)}$ are the weights connecting the hidden layer to the output layer, $b_k^{(2)}$ is the bias term for node k , and $z_j^{(1)}$ are the outputs from the hidden layer.

The final output of the network is then calculated as:

$$y(x) = f_{out}(a_{out}) + b \quad (22)$$

where a_{out} is the activation of the output node and $f(\cdot)$ is an appropriate activation function, such as the identity function for regression tasks.

The weights and biases in an MLP are typically learned from data using an optimization algorithm such as stochastic gradient descent or limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newton [27]. In this work, the last algorithm is considered.

There are several parameters that can be adjusted to obtain the best model. The most important of them, which give a better level of prediction for most DNN-based GC models, are explained here:

- *N_Hidden* (Number of hidden layers): This is the number of layers between the input and output layers.
- *Layer_N_Size* (Number of neurons per layer): This is the number of neurons in the *N*th hidden layer.
- *Activations* (Activation functions): Each neuron typically applies an activation function to its input. Common choices include ReLU, sigmoid, and tanh. The choice of activation function can affect the network's performance.
- *Lambda* (Regularization term strength): Regularization parameter that controls the complexity of the model by adding a penalty for parameter weights.

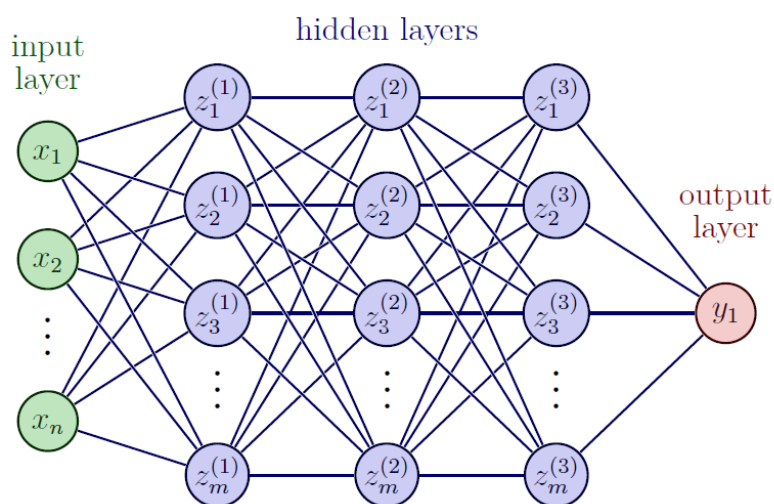


Figure 24 : representation of a neural network with two hidden layers and one output neuron.

II.6 MODEL VALIDATION

Once the model has been built, it is important to validate it to ensure its robustness and ability to accurately predict the viscosity of organic compounds. This can be done by dividing the dataset into a training set and a test set, and evaluating the performance of the model on the test set using several statistical validation parameters.

II.6.1 Statistical validation parameters

The performance of proposed models is analysed, evaluated and discussed in terms of four statistical measures, namely, Average absolute deviation (AAE), Average absolute relative deviation (AARD), Mean squared error (MSE) and coefficient of determination (R^2). These statistical performance indicators are expressed as follows.

$$AAE = \frac{1}{N} \sum_i |(y_i^{exp} - y_i^{pred})| \quad (23)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_i (y_i^{exp} - y_i^{pred})^2} \quad (1)$$

$$AARE = \frac{1}{N} \sum_i \left| \frac{y_i^{exp} - y_i^{pred}}{y_i^{exp}} \right| \times 100\% \quad (25)$$

$$R^2 = 1 - \left(\frac{\sum_i (y_i^{pred} - \mu^{exp})^2}{\sum_i (y_i^{exp} - \mu^{exp})^2} \right) \quad (26)$$

where, i be the i th data item in the dataset, N the total number of data sets, y_i^{exp} , the experimental values, y_i^{pred} , the predicted values, μ^{exp} , the mean value calculated over all the experimental values.

The average absolute error (AAE), as defined by its formula (eq23.), represents the average of the absolute differences between predicted and experimental values. When comparing multiple AAE values, the model associated with the lowest AAE is considered the best, as it indicates a closer alignment between predicted and observed values. One of the key advantages of using AAE is its robustness against outliers or very large errors, as it considers the absolute values of deviations rather than their squared values.

Root Mean Squared Error (RMSE), as defined by its formula (eq24.), represents the square root of the average of the squared differences between predicted and experimental values. In the realm of simple or multiple regression, many algorithms are designed to minimize RMSE to enhance model accuracy. When comparing RMSE values, a lower RMSE is indicative of a more accurate model. However, it's crucial to acknowledge that RMSE is more sensitive to outliers compared to average absolute deviation (AAD), as it squares the differences before taking the mean. This sensitivity underscores the importance of understanding the data distribution and the impact of outliers on model performance.

The Average Absolute Relative Error (AARE), as defined by its formula (eq25.), is a metric used to evaluate the accuracy of a predictive model by measuring the average of the absolute differences between predicted and actual values, relative to the actual values. It is defined as the average of the absolute differences between the predicted and actual values divided by the actual values, expressed as a percentage. A lower AARE indicates a more accurate model. Unlike RMSE, which squares the errors, AARE provides a more direct measure of the magnitude of errors relative to the actual values, making it useful for understanding the model's performance in terms of relative accuracy. However, like RMSE, AARE is sensitive to outliers and should be interpreted with caution, especially in cases where outliers may significantly impact the overall error calculation.

The coefficient of determination, commonly denoted as R^2 (see equation 26), serves as a crucial metric in assessing the efficacy of a predictive model. It encapsulates the proportion of the total variance exhibited by the property under consideration that is elucidated by the model itself. Essentially, R^2 is used as a benchmark to assess the model's ability to delineate the distribution of data points in parameter space. R^2 values typically range between 0 and 1, offering a comprehensive scale for evaluating model performance. A value of 0 implies that the model fails to explain any of the variance in the property, rendering it essentially meaningless. Conversely, a R^2 value of 1 indicates a perfect fit, suggesting that the model precisely captures all the variability present in the dataset.

II.6.2 Cross-validation and external validation

In the context of developing GC/QSPR predictive models for organic compound propriety, cross-validation is a vital technique used to assess the performance and reliability of these models. Cross-validation involves dividing the dataset into multiple subsets, or folds, with one fold reserved for testing while the model is trained on the remaining folds. This process is repeated multiple times, each time using a different fold as the test set. The results are then averaged over all folds to obtain a more robust estimate of model performance. Common cross-validation methods include k-fold cross-validation, where the dataset is divided into k subsets of approximately equal size, and leave-one-out cross-validation (LOOCV), where each data point is used as a test sample once, with the model trained on all other samples (Figure 25). Overall, cross-validation helps to assess the generalization ability of the model, ensuring that it performs well on

unseen data. It also helps to detect overfitting, where the model performs well on the training data but poorly on new data.

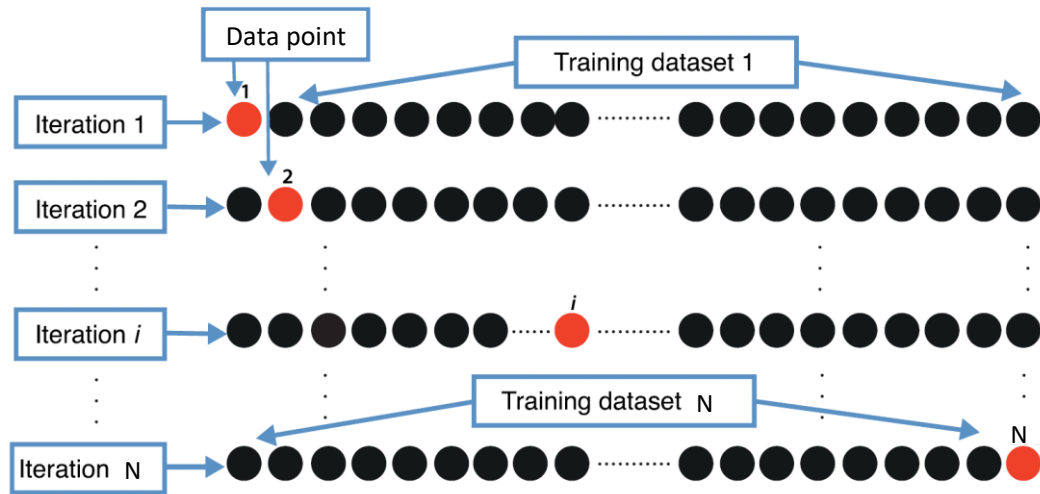


Figure 25 : Schematic representation of leave-one-out cross-validation

Similarly, validation can be categorized into internal and external validation, each serving a distinct purpose in assessing model performance. Internal validation is utilized for model optimization, ensuring that the model is fine-tuned to achieve the best possible performance. On the other hand, external validation involves isolating validation data (also known as the test dataset) from the model construction and optimization process (see Figure 26). This approach aims to replicate the real-world conditions under which the model will be deployed, providing a more accurate assessment of its performance in practical applications. External validation is crucial for confirming that the model can generalize well to unseen data and operate effectively in production environments.

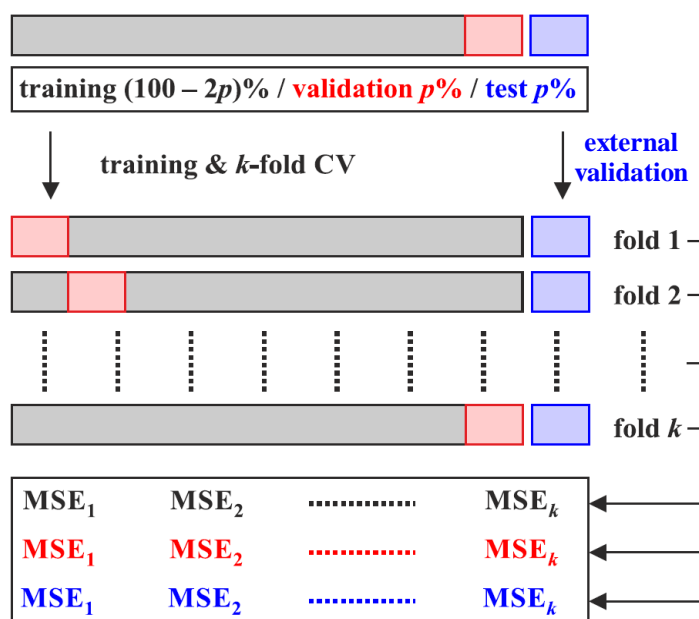


Figure 26 : Flowchart of the training (cross-validation) and testing (external validation) procedure

II.7 CONCLUSION

In conclusion, Chapter 2 has provided a comprehensive overview of the methodology and modeling methods essential for the development of predictive models for the viscosity of organic compounds. The chapter highlighted the importance of databases and data revision in ensuring the accuracy and reliability of models. It also explored group contribution methods, shedding light on the hierarchical structure of groups and their role in fragmentation schemes. The chapter further discussed Quantitative Structure-Property Relationships (QSPR), emphasizing the significance of molecular descriptors and fingerprints in predictive modeling. Additionally, conventional models and machine learning algorithms such as Random Forest, Extreme Gradient Boosting, Non-linear Support Vector Machine, and Neural Network were examined in detail. Finally, the chapter discussed model validation, including statistical validation parameters and cross-validation techniques.

II.8 BIBLIOGRAPHY

- [1] Tough, J. T. McCormick, W. D. Dash, J. G. « Viscosity of liquid He II,» *Physical Review*, vol. 132, n° 16, p. 2373, 1963.
- [2] Assael, M. J. Dalaouti, N. K. Vesovic, V. « Viscosity of natural-gas mixtures: measurements and prediction,» *International journal of thermophysics*, vol. 22, pp. 61-71, 2001.
- [3] P.S. van der Gulik, R. Mostert and H.R. van den Berg « The viscosity of methane at 273 K up to 1 GPa,» *Fluid phase equilibria*, vol. 79, pp. 301-311, 1992.
- [4] R. L. Alan S. Morris, «Chapter 21 - Summary of Other Measurements,» chez *Measurement and Instrumentation (Second Edition)*, Academic Press, 2016, pp. 633-672.
- [5] M. NIZAM, «MEASUREMENT OF VISCOSITY OF DIFFERENT TYPES OF FLUID,» Fluid mechanics (code: CLB 11003) , University of Kuala Lumpur, 2024.
- [6] Poling, B. E., Prausnitz, J. M., John Paul, O. C., & Reid, R. C, The properties of gases and liquids, McGraw-Hill, New York, : The McGraw-Hill Companies, Inc, 1977.
- [7] Ely, J. F., & Hanley, H. J. M, «Prediction of transport properties.1 Viscosity of fluids and mixtures,» *Industrial & Engineering Chemistry Fundamentals*, vol. 20, n° 14, pp. 323-332, 1981.
- [8] F. A. Pedersen K.S., « Viscosity of Crude Oil,» *Chem. Eng. Sci*, pp. 39, 6, 1011-1016, 1984.
- [9] Pedersen, K. S., & Fredenslund, A. A. G. E, «An improved corresponding states model for the prediction of oil and gas viscosities and thermal conductivities,» *Chemical Engineering Science*, vol. 42, n° 11, pp. 182-186, 1987.
- [10] Teja, A. S., & Rice, P, «Generalized corresponding states method for the viscosities of liquid mixtures,» *Industrial & Engineering Chemistry Fundamentals*, vol. 20, n° 11, pp. 77-81, 1981.
- [11] Aasberg-Petersen, K., Knudsen, K., & Fredenslund, A., «Prediction of viscosities of hydrocarbon mixtures,» *Fluid Phase Equilibria*, vol. 70, n° 12-3, pp. 293-308, 1991.
- [12] Huang, Y., Dong, H., Zhang, X., Li, C., & Zhang, S, «A new fragment contribution-corresponding states method for physicochemical properties prediction of ionic liquids,» *Thermodynamics and Molecular-Scale Phenomena*, vol. 59, n° 14, pp. 1348-1359, 2013.
- [13] Zhao, Y., Zhang, X., Deng, L., & Zhang, S., «Prediction of viscosity of imidazolium-based ionic liquids using MLR and SVM algorithms,» *Computers & Chemical Engineering*, vol. 92, pp. 37-42, 2016.

- [14] Quiñones-Cisneros, S. E., Zéberg-Mikkelsen, C. K., & Stenby, E. H, « The friction theory (f-theory) for viscosity modeling,» *Fluid Phase Equilibria*, vol. 169, n° %12, pp. 249-276, 2000.
- [15] Shen, G., Held, C., Mikkola, J. P., Lu, X., & Ji, X. , «Modeling the Viscosity of Ionic Liquids with the Electrolyte Perturbed-Chain Statistical Association Fluid Theory,» *Industrial & Engineering Chemistry Research*, vol. 53.52, p. 20258–20268, 2014.
- [16] Abolala, M., Peyvandi, K., & Varaminian, F «Modeling the viscosity of pure imidazolium-based ionic liquids using SAFT-VR-Mie EoS,» *Fluid Phase Equilibria*, vol. 394, pp. 61-70, 2015.
- [17] Haghighbakhsh, R., Parvaneh, K., & Shariati, A, «Viscosities of Pure Ionic Liquids Using Combinations of Free Volume Theory or Friction Theory with the Cubic, the Cubic Plus Association, and the Perturbed-Chain Statistical Associating Fluid Theory Equations of State at High Pressures,» *Industrial & Engineering Chemistry Research*, vol. 56, n° %18, p. 2247–2258, 2017.
- [18] R. Macías-Salinas, «Viscosity Modeling of Ionic Liquids Using the Friction Theory and a Simple Cubic Equation of State,» *Industrial & Engineering Chemistry Research*, vol. 57, n° %13, p. 1109–1120, 2018.
- [19] A. J. Batschinski, «Untersuchungen Aber die innere Reibnng der Flüssigkeiten. I,» *Zeitschrift für Physikalische Chemie*, vol. 84U, n° %11, pp. 643-706, 1913.
- [20] M. H. Cohen et D. Turnbull, «Molecular Transport in Liquids and Glasses,» *The journal of Chemical Physics*, vol. 31, n° %15, p. 1164–1169, 1959.
- [21] A. K. Doolittle, «Studies in Newtonian Flow. II. The Dependence of the Viscosity of Liquids on Free-Space,» *Journal of Applied physics*, vol. 22, n° %112, p. 1471–1475, 1951.
- [22] Llorell, F., Valente, E., Vilaseca, O., & Vega, L. F, «Modeling Complex Associating Mixtures with [Cn-mim][Tf2N] Ionic Liquids: Predictions from the Soft-SAFT Equation,» *The Journal of Physical Chemistry B*, vol. 115, n° %115, p. 4387–4398, 2011.
- [23] Shen, G., Held, C., Mikkola, J. P., Lu, X., & Ji, X, «Modeling the Viscosity of Ionic Liquids with the Electrolyte Perturbed-Chain Statistical Association Fluid Theory,» *Industrial & Engineering Chemistry Research*, vol. 53, n° %152, p. 20258–20268, 2014.
- [24] Yarranton, H. W., & Satyro, M. A, «Expanded Fluid-Based Viscosity Correlation for Hydrocarbons,» *Industrial & Engineering Chemistry Research*, vol. 48, n° %17, p. 3640–3648, 2009.
- [25] Satyro, M. A., & Yarranton, H. W, «Expanded fluid-based viscosity correlation for hydrocarbons using an equation of state,» *Fluid Phase Equilibria*, vol. 298, n° %11, pp. 1-11, 2010.

- [26] Loria, H., Motahhari, H., Satyro, M. A., & Yarranton, H. W, «Process simulation using the expanded fluid model for viscosity calculations,» *Chemical Engineering Research and Design*, vol. 92, n° 112, pp. 3083-3095, 2014.
- [27] Y. Rosenfeld, «Relation between the transport coefficients and the internal entropy of simple systems,» *Physical Review A*, vol. 15, n° 16, pp. 2545-2549, 1977.
- [28] Ashurst, W. T., & Hoover, W. G, « Dense-fluid shear viscosity via nonequilibrium molecular dynamics,» *Physical Review A*, vol. 11, n° 12, p. 658, 1975.
- [29] Enskog, D, «Kinetic theory of heat conductivity, viscosity and diffusion in certain condensed gases and liquids,» *Kgl Svenska Vetenskaps Akad Handl*, vol. 63, pp. 1-44, 1922.
- [30] Chapman, S., & Cowling, T. G, *The mathematical theory of non-uniform gases: an account of the kinetic theory of viscosity, thermal conduction and diffusion in gases*, Cambridge university press, 1990.
- [31] Baskaran, A., Dufty, J. W., & Brey, J. J, «Transport coefficients for the hard-sphere granular fluid,» *Physical Review E*, vol. 77, n° 13, p. 031311, 2008.
- [32] D. Chandler, «Rough hard sphere theory of the self-diffusion constant for molecular liquids,» *J. Chem. Phys*, vol. 62, p. 1358–1363, 1975.
- [33] J. H. DYMOND, « Corrections to the Enskog theory for viscosity and thermal conductivity,» *Physica B+ C*, vol. 144, n° 13, pp. 267-276., 1987.
- [34] Chung, T. H., Ajlan, M., Lee, L. L., & Starling, K. E, «Generalized multiparameter correlation for nonpolar and polar fluid transport properties,» *Industrial & Engineering Chemistry Research*, vol. 27, n° 14, p. 671–679, 1988.
- [35] Chung, T. H., Lee, L. L., & Starling, K. E, «Applications of kinetic gas theories and multiparameter correlation for prediction of dilute gas viscosity and thermal conductivity,» *Industrial & Engineering Chemistry Fundamentals*, vol. 23, n° 11, pp. 8-13, 1984.
- [36] Assael, M. J., Dymond, J. H., Papadaki, M., & Patterson, P. M., «Correlation and prediction of dense fluid transport coefficients. I. n-alkanes,» *Correlation and prediction of dense fluid transport coefficients. I. n-alkanes*, vol. 13, p. 269–281, 1992.
- [37] Bleazard, J. G., & Teja, A. S, «Extension of the Rough Hard-Sphere Theory for Transport Properties to Polar Liquids,» *Industrial & Engineering Chemistry Research*, vol. 35, n° 17, p. 2453–2459, 1996.
- [38] Ciotta, F., Trusler, J. M., & Vesovic, V, «Extended hard-sphere model for the viscosity of dense fluids,» *Fluid Phase Equilibria*, vol. 363, pp. 239-247, 2014.
- [39] V. V. Nicolas Riesco, «Extended hard-sphere model for predicting the viscosity of long-chain n-alkanes,» *Fluid Phase Equilibria*, vol. 425, pp. 385-392, 2016.

- [40] de Wijn, A. S., Riesco, N., Jackson, G., Martin Trusler, J. P., & Vesovic, V., «Viscosity of liquid mixtures: The Vesovic-Wakeham method for chain molecules,» *The Journal of Chemical Physics*, vol. 136, n° 17, 2012.
- [41] Assael, M. J., Dymond, J. H., Papadaki, M., & Patterson, P. M., «Correlation and prediction of dense fluid transport coefficients. I. n-alkanes,» *International Journal of Thermophysics*, vol. 13, p. 269–281, 1992.
- [42] Gaciño, F. M., Comuñas, M. J., Fernández, J., Mylona, S. K., & Assael, M. J., «Correlation and Prediction of Dense Fluid Transport Coefficients. IX. Ionic Liquids,» *International Journal of Thermophysics*, vol. 35, p. 812–829, 2014.
- [43] Hossain, M. Z., & Teja, A. S., «Correlation and Prediction of the Transport Properties of Ionic Liquids,» *International Journal of Thermophysics*, vol. 37, 2016.
- [44] Hosseini, S. M., Alavianmehr, M. M., & Moghadasi, J., «Transport properties of pure and mixture of ionic liquids from new rough hard-sphere-based model,» *Fluid Phase Equilibra*, vol. 429, pp. 266-274, 2016.
- [45] S. M. Hosseini, «Erratum to “Transport properties of pure and mixture of ionic liquids from new rough hard-sphere-based model,» *Fluid Phase Equilibra*, vol. 429, p. 266–274, 2016.
- [46] Akbari, F., Alavianmehr, M. M., Behjatmanesh Ardakani, R., & Mohammad-Aghaie, D., «Thermophysical properties of ionic liquids and their mixtures from a new equation of state,» *Ionics*, vol. 24, p. 1357–1369, 2018.
- [47] H. T. Davis, S. A. Rice et J. V. Sengers, «On the Kinetic Theory of Dense Fluids. IX. The Fluid of Rigid Spheres with a Square-Well Attraction,» *The Journal of Chemical Physics*, vol. 35, n° 16, p. 2210–2233, 1961.
- [48] I. L. McLaughlin et H. T. Davis, «Kinetic Theory of Dense Fluid Mixtures. I. Square-Well Model,» *The Journal of Chemical Physics*, vol. 45, n° 16, p. 2020–2031, 1966.
- [49] Du, L. G., & Guo, T. M., «A semi-theoretical viscosity model for non-polar liquids,» *The Chemical Engineering Journal*, vol. 47, n° 13, pp. 163-167, 1991.
- [50] Monnery, W. D., Svrcek, W. Y., & Mehrotra, A. K., «Viscosity: A critical review of practical predictive and correlative methods,» *The Canadian Journal of Chemical Engineering*, vol. 73, n° 11, pp. 3-40, 1995.
- [51] Galliero, G., Nieto-Draghi, C., Boned, C., Avalos, J. B., Mackie, A. D., Baylaucq, A., & Montel, F., «Molecular Dynamics Simulation of Acid Gas Mixtures: A Comparison between Several Approximations,» *Industrial & Engineering Chemistry Research*, vol. 46, n° 115, p. 5238–5244, 2007.
- [52] Zerón, I. M., Cueto-Mora, M., & Blas, F. J., «Transport properties of the square-well fluid from molecular dynamics simulation,» *Molecular Physics*, vol. 122, p. 19–20, 2024.
- [53] S. Handy, *Ionic Liquids - Current State of the Art*, Germany: first ed., IntechOpen, 2015.

- [54] Nieto-Draghi, C., Fayet, G., Creton, B., Rozanska, X., Rotureau, P., de Hemptinne, J. C., ... & Adamo, C; «A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes,» *Chemical Reviews*, vol. 115, n° %124, p. 13093–13164, 2015.
- [55] Katritzky, A. R., Chen, K., Wang, Y., Karelson, M., Lucic, B., Trinajstić, N., ... & Schüürmann, G, «Prediction of liquid viscosity for organic compounds by a quantitative structure–property relationship,» *Journal of Physical Organic Chemistry*, vol. 13, n° %11, pp. 80-86, 2000.
- [56] Bouarab, A. F., Harvey, J. P., & Robelin, C, «Viscosity models for ionic liquids and their mixtures,» *Physical Chemistry Chemical Physics*, n° %12, 2021.
- [57] Tochigi, K., & Yamamoto, H, «Estimation of Ionic Conductivity and Viscosity of Ionic Liquids Using a QSPR Model,» *The Journal of Physical Chemistry C*, vol. 111, n° %143, 2007.
- [58] Barycki, M., Sosnowska, A., Gajewicz, A., Bobrowski, M., Wileńska, D., Skurski, P., ... & Puzyn, T., «Temperature-dependent structure-property modeling of viscosity for ionic liquids,» *Fluid Phase Equilibria*, vol. 427, pp. 9-17, 2016.
- [59] Bini, R., Malvaldi, M., Pitner, W. R., & Chiappe, C, «QSPR correlation for conductivities and viscosities of low-temperature melting ionic liquids,» *Journal of Physical Organic Chemistry*, vol. 21, pp. 7-8, 2008.
- [60] Yu, G., Wen, L., Zhao, D., Asumana, C., & Chen, X, «QSPR study on the viscosity of bis(trifluoromethylsulfonyl)imide-based ionic liquids,» *Journal of Molecular Liquids*, vol. 184, pp. 51-59, 2013.
- [61] Nieto-Draghi, C., Fayet, G., Creton, B., Rozanska, X., Rotureau, P., de Hemptinne, J. C., ... & Adamo, C, «A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes,» *Chemical Reviews*, vol. 115, n° %124, p. 13093–13164, 2015.
- [62] 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*, vol. 80, pp. 326-333, 2012.
- [63] Zhang, S., Jia, Q., Yan, F., Xia, S., & Wang, Q, «Evaluating the properties of ionic liquid at variable temperatures and pressures by quantitative structure–property relationship (QSPR),» *Chemical Engineering Science*, vol. 231, p. 116326, 2021.
- [64] Ding, Y., Chen, M., Guo, C., Zhang, P., & Wang, J, «Molecular fingerprint-based machine learning assisted QSAR model development for prediction of ionic liquid properties,» *Journal of Molecular Liquids*, vol. 326, p. 115212, 2021.
- [65] Wadhwa, P., & Mittal, A, Quantitative Structure-Property Relationship (QSPR) Modeling Applications in Formulation Development, Singapore: Springer, Singapore, 2022.

- [66] Argoub, K., Benkouider, A. M., Yahiaoui, A., & Bagui, F, «Estimation and uncertainty analysis of standard enthalpy of formation in the liquid state by third-order-group-contribution method,» *Fluid Phase Equilibria*, vol. 520, p. 112644–112644, 2020.
- [67] Serat, F. Z., Benkouider, A. M., Yahiaoui, A., & Bagui, F, «Nonlinear group contribution model for the prediction of flash points using normal boiling points,» *Fluid Phase Equilibria*, vol. 449, p. 52–59, 2017.
- [68] Guella, S., Argoub, K., Benkouider, A. M., Yahiaoui, A., Kessas, R., & Bagui, F, «Artificial Neural Network-Group Contribution Method for Predicting Standard Enthalpy of Formation in the Solid State: C–H, C–H–O, C–H–N, and C–H–N–O Compounds,» *International journal of thermophysics*, vol. 36, n° %110-11, p. 2820–2832, 2015.
- [69] Benkouider, A. M., Kessas, R., Guella, S., Yahiaoui, A., & Bagui, F, «Estimation of the enthalpy of vaporization of organic components as a function of temperature using a new group contribution method,» *Journal of molecular liquids*, vol. 194, p. 48–56, 2014.
- [70] Argoub, K., Benkouider, A. M., Yahiaoui, A., Kessas, R., Guella, S., & Bagui, F, «Prediction of standard enthalpy of formation in the solid state by a third-order group contribution method,» *Fluid phase equilibria*, vol. 380, p. 121–127, 2014.
- [71] K. Paduszyński, «Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity,» *Industrial & engineering chemistry research*, vol. 58, n° %136, p. 17049–17066, 2019.
- [72] Chen, Y., Kontogeorgis, G. M., & Woodley, J. M, «Group Contribution Based Estimation Method for Properties of Ionic Liquids,» *Industrial & Engineering Chemistry Research*, vol. 58, n° %110, p. 4277–4292, 2019.
- [73] Baghban, A., Kardani, M. N., & Habibzadeh, S, «Prediction viscosity of ionic liquids using a hybrid LSSVM and group contribution method,» *Journal of Molecular Liquids*, vol. 236, p. 452–464, 2017.
- [74] Lazzús, J. A., & Pulgar-Villarroel, G, «A group contribution method to estimate the viscosity of ionic liquids at different temperatures,» *Journal of molecular liquids*, vol. 209, pp. 161–168,, 2015.
- [75] , «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*, vol. 54, n° %15, p. 1311–1324.
- [76] K. Paduszyński, «Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity,» *Industrial & engineering chemistry research*, vol. 58, n° %136, p. 17049–17066, 2019.
- [77] Cao, X., Gong, M., Tula, A., Chen, X., Gani, R., & Venkatasubramanian, V, «An Improved Machine Learning Model for Pure Component Property Estimation,,» *Engineering*, vol. 39, pp. 61-73, 2024.

- [78] Varmuza, K., Dehmer, M., & Bonchev, D, Statistical Modelling of Molecular Descriptors in QSAR/QSPR, Wiley-VCH Verlag GmbH & Co. KGaA, 2012.
- [79] Khan, A. U, «Descriptors and their selection methods in QSAR analysis: paradigm for drug design,,» *Drug Discovery Today,,* vol. 21, n° 18, pp. 1291-1302, 2016.
- [80] Wang, L., Ding, J., Pan, L., Cao, D., Jiang, H., & Ding, X, «Quantum chemical descriptors in quantitative structure–activity relationship models and their applications,,» *Chemometrics and Intelligent Laboratory Systems,,* vol. 217, p. 104384, 2021.
- [81] Rogers, D., & Hahn, M., «Extended-Connectivity Fingerprints,,» *Journal of Chemical Information and Modeling*, vol. 50, n° 15, p. 742–754, 2010.
- [82] Carhart, Raymond E., SMITH, Dennis H., et VENKATARAGHAVAN, RENGACHARI, «Atom pairs as molecular features in structure-activity studies: definition and applications,,» *Journal of Chemical Information and Computer Sciences*, vol. 25, n° 12, p. 64–73, 1985.
- [83] Pérez-Nueno, V. I., Rabal, O., Borrell, J. I., & Teixidó, J, «APIF: A New Interaction Fingerprint Based on Atom Pairs and Its Application to Virtual Screening,,» *Journal of Chemical Information and Modeling*, vol. 49, n° 15, p. 1245–1260, 2009.
- [84] R. Davronov et S. Kushmuratov, «Comparative analysis of QSAR feature selection methods,,» *AIP Conference Proceedings*, vol. 3004, n° 11, 2024.
- [85] Zebida, M. A., Argoub, K., Benkouider, A. M., Yahiaoui, A., Toubal, K., & Hachemaoui, A, «Machine learning coupled with group contribution for predicting the electrical conductivity of ionic liquids with experimental accuracy,,» *Fluid Phase Equilibria,,* vol. 579, pp. 0378-3812, 2024.
- [86] Seber, G. A, Nonlinear Regression, 1989 John Wiley & Sons, Inc, 1989.
- [87] Daubechies, I., DeVore, R., Fornasier, M., & Güntürk, C. S., «Iteratively reweighted least squares minimization for sparse recovery,,» *Pure and Applied Mathematics*, vol. 63, n° 11, pp. 1-38, 2010.
- [88] Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P, «Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling,,» *Journal of Chemical Information and Computer Sciences*, vol. 43, n° 16, 2003.
- [89] Zhang, X., Shen, H., Huang, T., Wu, Y., Guo, B., Liu, Z., ... & Ou, G, «Improved random forest algorithms for increasing the accuracy of forest aboveground biomass estimation using Sentinel-2 imagery,,» *Ecological Indicators*, vol. 159, 2024.

- [90] Wang, C., Wei, X., Jin, X., Li, J., & He, M., «Developing a two-grade model for the thermal conductivity of ionic liquids and their mixtures,» *Chemical Engineering Science*, vol. 290, p. 119881, 2024.
- [91] Fröba, A. P., Rausch, M. H., Krzeminski, K., Assenbaum, D., Wasserscheid, P., & Leipertz, A., « Thermal conductivity of ionic liquids: Measurement and prediction,» *Int. J. Thermophys*, vol. 31, p. 2059–2077, 2010.
- [92] Zhao, A. Z., & Garay, J. E., «High temperature liquid thermal conductivity: A review of measurement techniques, theoretical understanding, and energy applications,» *Prog. Mater. Sci*, vol. 139, p. 101180, 2023.
- [93] Haghbakhsh, R., & Raeissi, S., «A novel correlative approach for ionic liquid thermal conductivities,» *J. Mol. Liq*, vol. 236, p. 214–219, 2017.
- [94] Soares, L. H., Guirardello, R., & Rolemberg, M. P., « A simple group contribution model to predict thermal conductivity of pure ionic liquids,» *Chem. Eng. Trans*, vol. 74, p. 1195–1200, 2019.
- [95] Almeida, R. M., Lourenço, M. J. V., & de Castro, C. N., «The thermal conductivity of ionic liquids. Experiment and molecular interpretation,» *J. Mol. Liq*, vol. 397, 2024.
- [96] Yang, H., Gallagher, R. C., Phan, A. T., Chartrand, P., & Gheribi, A. E., « A predictive approach for the compositional and temperature representation of thermal conductivity in multicomponent molten salt systems for advanced energy applications,» *Mater. Today Energy*, vol. 38, p. 101441, 2023.
- [97] Wan, R., Li, M., Song, F., Xiao, Y., Zeng, F., Peng, C., & Liu, H., « Predicting the Thermal Conductivity of Ionic Liquids Using a Quantitative Structure–Property Relationship,» *Ind. Eng. Chem. Res*, vol. 61, p. 12032–12039, 2022.
- [98] Kazakov, A. M. J. W., Magee, J. W., Chirico, R. D., Paulechka, E., Diky, V., Muzny, C. D., ... & Frenkel, M., «NIST Standard Reference Database 147: NIST Ionic Liquids Database-(ILThermo),» National Institute of Standards and Technology. Gaithersburg MD, v2.0, 2021. [En ligne]. Available: <http://ilthermo.boulder.nist.gov>.
- [99] Snoek, J., Larochelle, H., & Adams, R. P., *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, Nevada: Curran Associates Inc, 2012.
- [100] Gelbart, M. A., Snoek, J., & Adams, R. P., *Proceeding of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, Quebec City, Quebec, Canada: AUAI Press, 2014.
- [101] Han, C., Yu, G., Wen, L., Zhao, D., Asumana, C., & Chen, X
, «Data and QSPR study for viscosity of imidazolium -based ionic liquids,» *Fluid Phase Equilibra*, vol. 300, p. 95, 2011.
- [102] Valderrama, J. O., Cardona, L. F., & Rojas, R. E., « Correlation and prediction of ionic liquids viscosity using Valderrama-Patel-Teja cubic equation of state and the geometric

- similitude concept, Part I: Pure ionic liquids,» *Fluid Phase Equilibra*, vol. 497, p. 164, 2019.
- [103] He, M., Zhu, C., & Liu, X, « Estimating the viscosity of ionic liquid at high pressure using Eyring's absolute rate theory,» *Fluid Phase Equilib*, vol. 458, p. 170, 2018.
- [104] Darabi, L., & Zare, M., «High correlate simple equations for temperature and pressure dependence of the viscosity of ionic liquids,» *Chem. Phys*, vol. 539, 2020.
- [105] Yan, F., He, W., Jia, Q., Wang, Q., Xia, S., & Ma, P., « Prediction of ionic liquids viscosity at variable temperatures and pressures,» *Chem. Eng. Sci.*, vol. 184, p. 134, 2018.
- [106] Kang, X., Zhao, Z., Qian, J., & Muhammad Afzal, R, «Predicting the Viscosity of Ionic Liquids by the ELM Intelligence Algorithm,» *Ind. Eng. Chem. Res*, vol. 56, p. 11344, 2017.
- [107] Sun, Y., Chen, M., Zhao, Y., Zhu, Z., Xing, H., Zhang, P., ... & Ding, Y, «Machine learning assisted QSPR model for prediction of ionic liquid's refractive index and viscosity: The effect of representations of ionic liquid and ensemble model developement,» *J. Mol. Liq*, vol. 333, 2021.
- [108] T. Mezger, *The Rheology Handbook*, Vincentz Network: 3rd revised ed. Hanove, 2011.
- [109] M. Rao, « Rheology of liquid foods,» *Journal Tenure studies*, p. pp. 135 – 168., 1997.
- [110] T. C. Scott et R. Wham, «Surface area generation and droplet size control in solvent extraction systems utilizing high intensity electric fields,» *U.S. Patent 4767515.12*, 1988.
- [111] Ptasinski, K. J., & Kerkhof, P. J. A. M, «Electric field driven separations: Phenomena and applications.,» *Sep. Sci. Technol.*, 1992.
- [112] H. Xie, L. Li, Y. Sun, Y. Wang, S. Gao, Y. Tian, X. Ma, C. Guo, F. Bo et L. Zhang, « An Available Strategy for Nasal Brain Transport of Nanocomposite Based on PAMAM Dendrimers via In Situ Gel,» *Nanomaterials* , pp. 9, 147, 2019.
- [113] B. Vigani, S. Rossi, G. Milanesi, M. Bonferoni, G. Sandri, G. Bruni et F. Ferrari, « Electrospun Alginate Fibers: Mixing of Two Different Poly (ethylene oxide) Grades to Improve Fiber Functional Properties.,» *Nanomaterials* , pp. 8, 971, 2018.
- [114] P. Mishra, S. Mukherjee, S. Nayak et A. Panda, «A brief review on viscosity of nanofluids,» *Int. Nano Lett*, pp. 4, 109–120., 2014.
- [115] A. Einstein, «Eineneuebestimmung der molekuldimensionen.,» *Ann. Phys*, pp. 324, 289–306., 1906.
- [116] T. Phuoc et M. Massoudi, « Experimental observations of the effects of shear rates and particle concentration on the viscosity of Fe₂O₃–deionized water nanofluids.,» *Int. J. Therm. Sci*, pp. 48, 1294–1301, 2009.
- [117] Beggs, H. D., & Robinson, J. R., « Estimating the Viscosity of Crude Oil Systems,» *J. Petrol. Technol*, pp. 9, 1140-1141, 1995.

- [118] F. F. Petrosky G.E., «Viscosity Correlations for Gulf of Mexico Crudes Oils,,» 1995.
- [119] Labedi, R, « Improved Correlations for Predicting the Viscosity of Light Crudes,» *J. Petrol. Sci. Eng.*, pp. 8, 221-234, 1992.
- [120] Khan, S. A., Al-Marhoun, M. A., Duffuaa, S. O., & Abu-Khamsin, S. A, «Viscosity Correlations for Saudi Arabian Crude Oils,» p. 15 720., 1987.
- [121] V. d. W. J.D, Leyden, 1873.
- [122] Lohrenz, J., Bray, B. G., & Clark, C. R, « Calculating Viscosities of Reservoir Fluids from their Composition,» *J. Petrol. Technol*, pp. 1171-1176, 1964.
- [123] K. H. Little J.E., «A Correlation of the Viscosity of Hydrocarbon Systems with Pressure, Temperature and Composition,,» *Soc. Petrol. Eng. J. AIME,,* pp. 243, 157-162, 1968.
- [124] Guo, X. Q., Wang, L. S., Rong, S. X., & Guo, T. M, « Viscosity Model Based on Equations of State for Hydrocarbon Liquids and Gases,» *Fluid Phase Equilib*, pp. 139, 405-421, 1997.
- [125] T. A. Patel N.C., « A New Cubic Equation of State for Fluids and Fluid Mixtures,» *Chem. Eng. Sci*, pp. 37, 463-473., 1982.
- [126] Ungerer, P., Thermodynamics: Applications in Chemical Engineering and the Petroleum Industry,, paris: Ed. Technip, Paris, 2003.
- [127] 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*, vol. 80, p. 326–333, 2015.
- [128] Zhang, S., Jia, Q., Yan, F., Xia, S., & Wang, Q., «Evaluating the properties of ionic liquid at variable temperatures and pressures by quantitative structure–property relationship (QSPR),» *Chem. Eng. Sci*, vol. 231, 2021.

CHAPITRE III: PREDICTION OF THE VISCOSITY OF IONIC LIQUIDS

III.1 INTRODUCTION

Ionic liquids (ILs) are distinguished by their arrangement of weakly coordinated ions, endowing this class of compounds with exceptional physicochemical properties: thermal and chemical stability, non-volatility, a wide electrochemical window, and structural tunability [90] [91]. Among their key characteristics, viscosity (η) plays a central role in determining flow behavior, affecting ionic diffusion, electrical and thermal conductivity, as well as the efficiency of mixing and mass-transfer processes [92] [93]. For these reasons, precise control and prediction of IL viscosity are indispensable in numerous applications: heat-transfer fluids in exchangers, “green” solvents for organic syntheses, high-performance electrolytes for batteries and supercapacitors, and reaction media with high ionic load [94] [95] .

Yet reliable prediction of IL viscosity remains a major challenge. Complex ionic interactions, the vast diversity of cationic and anionic structures, and the sensitivity of η to thermodynamic conditions (temperature, pressure) make it difficult to establish simple, universal correlations [96] [97]. Classical approaches based on empirical or semi-empirical models quickly reach their limits when confronted with novel ion pairs or extended temperature ranges.

In this chapter, we have structured our approach in five stages: data collection and preprocessing, exploratory analysis, model development, presentation of results and discussion, and conclusion.

The primary objective of this chapter is therefore to develop robust, generalizable models for IL viscosity, relying on an extensive database and state-of-the-art machine-learning

methods, in order to provide experimentalists and engineers with reliable predictive tools for the design and optimization of ionic fluids.

III.2 EXPERIMENTAL DATABASE

The main objective of this chapter was to extend the range of applicability of the proposed model as well as to increase the accuracy of predictions over wider temperature ranges compared to other concurrent models described in the chapter II. To achieve this goal, an extensive and diverse experimental database was developed.

A significant part of the viscosity database is retrieved from the data accepted and not excluded in Paduszyński's paper [71], i.e. data published between 1984 and early 2018. Thus, a total of 13539 experimental data points for 1596 distinct ILs are considered at atmospheric pressure. This entire database is presented in the Microsoft Excel file. This extensive database was updated from the ILThermo (NIST-ILs) database [98] with viscosity data published between 2018 and today. The total updated data were 1712 data points for 119 distinct ILs recorded with uncertainties. This updated database was analysed and revised according to the following workflow: The first step was to exclude crystalline phases and metastable liquids. The second step aimed to remove duplicate entries within each dataset. The third step involved identifying inconsistent measurements—that is, multiple values for the same liquid at identical temperatures—and retaining only the average of those values. The fourth step eliminated data points with a relative uncertainty greater than 100 %, while the fifth step discarded liquids for which fewer than three data points remained. In the sixth step, the remaining data for each IL at atmospheric pressure were fitted using the Vogel–Fulcher–Tammann equation. Any data point with a studentized residual outside twice the interquartile range (IQR) was then flagged as an outlier, and this fitting–outlier removal cycle was repeated until no further outliers were detected. In the final step of the workflow, the dataset was manually reviewed to remove any potentially inappropriate entries.

[im]	226	32	24	68	31	53	63	25	13	35	27	11	5	5	2
[n]	117	1	97	26	23	15	9	15	1	6	5	17			1
[p]	33	4	12	5	2	17	18	8	33	2	8	25		19	3
[py]	68	4	1	10	12	6	5	5		9			4		
[pyr]	43	4	17	12	5	1	13	6		1	5		1		
[pip]	33	1	8	7	1	4	3	1			3				
[mo]	9	1	1	5	2			8		1	3		4		
[cprop]	20							13							
[s]	16							8							
[guan]	17		1		2	1		1					1		
[azp]	6		2			2									
[trz]	12			1		2	2								
[pz]	5			2		5		7							
[bic]	1						3	6							
[thur]	9				2										
[cs]	6						1	3							
[thz]	5														
[ox]	4			8											
[camd]	4					2									
[amd]	3														
[pipz]			1		1										
	[NTf ₂]	[PF ₆]	[RCO ₂]	[BF ₄]	[RSO ₄]	[RSO ₃]	[x]	[dca]	[hca]	[ML _n]	[RPO ₄]	[aa]	[CR ₃]	[o]	[b]

Figure 26. Number of ILs according to different cation-anion combinations. Empty field means that the experimental data have been not available yet. The abbreviation of chemical families are detailed in Table .

A new extensive viscosity database was obtained, containing 15251 experimental data points corresponding to 1654 various ILs (see Figure 26) belonging to 21 and 15 different chemical families of cations and anions, respectively (see Figure). Figure 26 displays the number of ionic liquids (ILs) based on combinations of cation and anion chemical families, while Figure 27 displays the corresponding number of data points.

The abbreviation of chemical families are detailed in Table . This table summarizes, for the various cationic and anionic families of ionic liquids, the number of data points and datasets collected, as well as the molecular diversity and the ranges of temperature and viscosity covered. As can be seen, the data set includes a wide variety of cations, the most numerous being imidazolium, ammonium, phosphonium pyridinium and pyrrolidinium. Among the anions, bistriflamides (NTf₂) and carboxylates are the most important in the

database as they cover a fraction of 39% and 10% of all ILs, respectively. Overall, the imidazolium/NTf₂ classes dominate the database both in volume and in chemical diversity, whereas some of the rarer families (oxazolidinium, amidium, piperazinium, organic borates) remain sparsely represented. Measured temperatures typically span from approximately 253 K to 423 K, and viscosities cover several orders of magnitude (from below 1 mPa·s to over 10⁵ mPa·s), highlighting the extreme rheological variability of ionic liquids as a function of ionic structure. This wealth of data justifies the use of advanced statistical and modeling methods to extract reliable trends.

[im]	2510	1146	215	570	532	535	487	300	118	203	124	101	110	45	19
[n]	788	6	536	33	172	125	96	17	8	40	7	95			6
[p]	287	209	158	49	13	70	184	30	321	24	75	127		67	15
[py]	636	25	13	281	140	67	22	74		90			46		
[pyr]	597	54	194	24	63	19	61	29		3	53		6		
[pip]	326	6	75	19	10	52	23	15			39				
[mo]	72	6	26	5	21			72		7	39		36		
[cprop]	147							100							
[s]	170							47							
[guan]	197		1		14	1		1					1		
[azp]	78		30			30									
[trz]	89			8		16	16								
[pz]	19			6		11		72							
[bic]	10						27	54							
[thur]	52				12										
[cs]	23						6	31							
[thz]	30														
[ox]	4			8											
[camd]	6					4									
[amd]	6														
[pipz]			4		1										
	[NTf ₂]	[PF ₆]	[RCO ₂]	[BF ₄]	[RSO ₄]	[RSO ₃]	[X]	[dca]	[hca]	[ML _n]	[RPO ₄]	[aa]	[CR ₃]	[o]	[b]

Figure 277. Number of data points according to different cation-anion combinations. Empty field means that the experimental data have been not available yet. The abbreviation of chemical families are detailed in Table .

Table 3: Summary of the viscosity database by ion families

IL family	Data points	Data sets	Molecules	ILs	T min (K)	T max (K)	η min (mPa.s)	η max (mPa.s)
Cations								
Imidazolium (im)	7015	791	258	620	253.15	573	0.84	23508
Ammonium (n)	1929	351	164	333	273.15	573	1.2	14433
Phosphonium (p)	1629	196	64	189	253.15	374.15	3.56	128800
Pyridinium (py)	1394	145	76	124	273.15	373.15	5.49	9946
Pyrrolidinium (pyr)	1103	129	40	108	258.15	393	1.15	12220
Piperidinium (pip)	565	65	33	61	283	413.2	5.76	47700
Morpholinium (mo)	284	35	16	34	283.2	368.2	9.1	3278.8
Cyclopropanium (cprop)	247	33	20	33	293	363.15	5.57	332
Sulfonium (s)	217	25	19	24	253.2	398.15	3.76	509
Guanidinium (guan)	215	24	18	23	293	353.15	8.76	673.4
Azepanium (azp)	138	10	6	10	289.55	368.1	10.64	797.5
Triazolium (trz)	129	17	12	17	298.1	353.2	9.83	1451.7
Pyrazolium (pz)	108	19	15	19	278.15	353.2	6.1	2131.24
Bicyclic (bic)	91	10	7	10	283	373.15	7.72	6473.9
Thiouronium (thur)	64	11	10	11	278.15	353.2	5.2	356
Cyclic sulfonium (cs)	60	10	8	10	253.2	353.1	7.82	1289.2
Thiazolium (thz)	30	5	5	5	293.15	333.15	26.1	948
Oxazolidinium (ox)	12	12	4	12	298.2	298.2	90	731
Cyclic amidium (camd)	10	6	4	6	298.2	323.2	14.6	85.1
Amidium (amd)	6	3	3	3	298.2	323.2	16.7	66.9
Piperazinium (pipz)	5	2	2	2	298.15	313.15	80.93	680
Anions								

NTf ₂ derivatives (NTf ₂)	6047	736	576	637	253.2	573	0.84	12220
PF ₆ derivatives (PF ₆)	1452	68	36	47	258.15	413.85	4.6	41714
Carboxylates (RCO ₂)	1252	178	76	164	263.15	423.15	1.16	47700
BF ₄ derivatives (BF ₄)	1003	181	65	144	273.15	388.04	3.29	6120
Sulfates (RSO ₄)	978	108	58	81	253.15	423.15	7.2	16145
Sulfonates (RSO ₃)	930	117	76	108	263.15	423.15	3.1	21829
Inorganics (x)	922	127	90	117	253.15	393.15	1.18	128800
Dicyanamides (dca)	842	122	106	106	253.2	373.15	3.64	6473.9
Heterocyclic amines (hca)	447	47	22	47	278.15	363.15	8.35	14600
Metal complexes (ML _n)	367	54	31	54	263.1	373.15	4.18	28230
Phosphates (RPO ₄)	337	54	42	51	273.15	373.2	9.28	23508
Aminoacids (AA)	323	55	16	53	283.15	373.15	3.5	10643.8
Methanides (CR ₃)	199	22	12	15	278.15	368.15	4.64	1124
Alcoholates (O)	112	24	7	24	293.15	373.15	6.01	984.3
Organic borates (B)	40	6	5	6	263.15	363.2	10.65	2412.3

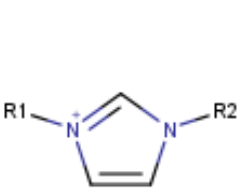
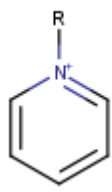
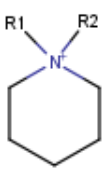
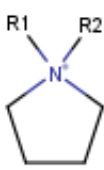
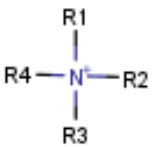
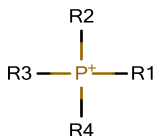
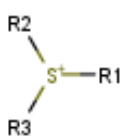
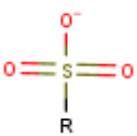
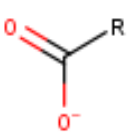
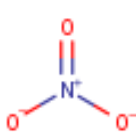


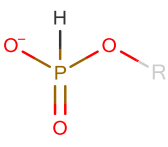
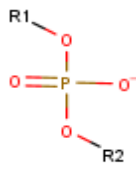
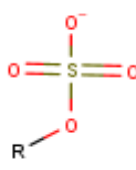
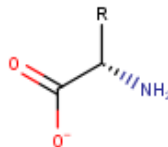
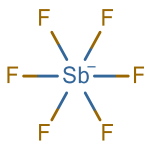
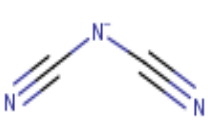
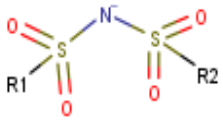
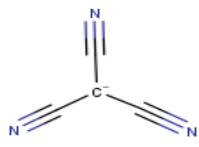
Cation	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>[Im]</p> </div> <div style="text-align: center;">  <p>[py]</p> </div> <div style="text-align: center;">  <p>[pip]</p> </div> <div style="text-align: center;">  <p>[pyr]</p> </div> </div> <div style="display: flex; justify-content: space-around; align-items: center; margin-top: 20px;"> <div style="text-align: center;">  <p>[n]</p> </div> <div style="text-align: center;">  <p>[p]</p> </div> <div style="text-align: center;">  <p>[s]</p> </div> </div>
Anion	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>[RSO3]</p> </div> <div style="text-align: center;">  <p>[RCO2]</p> </div> <div style="text-align: center;">  <p>[X]</p> </div> <div style="text-align: center;">  <p>[BF4]</p> </div> <div style="text-align: center;">  <p>[PF6]</p> </div> </div> <div style="display: flex; justify-content: space-around; align-items: center; margin-top: 20px;"> <div style="text-align: center;">  <p>[RPO3]</p> </div> <div style="text-align: center;">  <p>[RPO4]</p> </div> <div style="text-align: center;">  <p>[RSO4]</p> </div> <div style="text-align: center;">  <p>[AA]</p> </div> <div style="text-align: center;">  <p>[SbF6]</p> </div> </div> <div style="display: flex; justify-content: space-around; align-items: center; margin-top: 20px;"> <div style="text-align: center;">  <p>[dca]</p> </div> <div style="text-align: center;">  <p>[NTF2]</p> </div> <div style="text-align: center;">  <p>[CR3]</p> </div> </div>

Figure 28 : Cationic and anionic chemical families considered in this study. The abbreviations of the chemical families are detailed in Table .

III.3 METHODOLOGICAL STRATEGY FOR MODELLING

In this work, five modeling approaches were deployed to predict the viscosity of ionic liquids: Extreme Gradient Boosting (XGBoost), Random Forest (RF), Support Vector Machine (SVM), Deep Neural Network (DNN), and a conventional semi-empirical model (see Section II.5, Models and Regression).

Each approach relies on two types of structural inputs: Group Contributions (GC) encompassing cations, anions, and substituents, and QSPR (Quantitative Structure-Property Relationship) descriptors calculated using RDKit. These inputs were augmented with the thermodynamic variable temperature (T). For the GC approach, each molecule was fragmented into 127 groups (52 cationic, 44 anionic, 31 substituents), categorized into first-order and second-order contributions. For QSPR, a vector of 163 descriptors was computed for each ionic liquid, encapsulating molecular size, polarity, charge distribution, and functional groups.

For the conventional semi-empirical model the selection of the best shape of the model has been performed. Starting from the Vogel–Tammann–Fulcher (VTF) equation, several non-linear functional forms, representing the temperature dependence of the viscosity, η , have been tested and examined to find the best final expression. In this step, all data points in the database are used. Finally, the following equation was chosen due to its correlative power and extrapolation ability. It is mainly inspired by sigmoid transfer function used in artificial neural networks.

$$\ln(\eta) = \sum_{i=1}^3 b_i \frac{A_i}{1 + |A_i|} + c \quad (27)$$

$$A_i(T) = \sum_{j=1}^K N_j a_{i,j} + \alpha_i T \quad (28)$$

In equation (27), b_i and c are adjustable parameters, A_i is the contribution of all groups in i^{th} term of equation (28), including both first- and second-order groups that occurred in cationic, anionic and substituent cores. In equation (28), N_j denote the number of occurrences of j^{th} individual group, $a_{i,j}$ is the contribution of j^{th} group in i^{th} term of equation (28). K is the total number of functional groups and T is temperature in units of kelvin (K). α_i are adjustable parameters in i^{th} term of equation (28).

Note that the conventional model is exclusively applied to the GC method: molecular fragmentation enables direct assignment of the semi-empirical law's coefficients to specific chemical groups. In contrast, the QSPR approach (relying on a dense vector of global molecular descriptors) does not allow for such direct physical interpretation.

The development of all models, whether QSPR or group contributions, was carried out using MATLAB's Statistics & Machine Learning Toolbox (v. 2023b, academic licence). Default hyperparameter settings from the library were retained for most parameters. Only hyperparameters demonstrating significant impact in QSPR and GC contexts were optimized. For this tuning, Bayesian optimization [99] [100] was employed to iteratively propose the most promising configurations. Model quality was evaluated through 25 repeats of 5-fold cross-validation, minimizing mean squared error.

Hyperparameter optimization across all four models (XGBoost, RF, SVM, DNN) revealed remarkable performance convergence between QSPR and GC methodologies. Results were standardized to enable rigorous cross-approach comparison. The exploration of search spaces and final hyperparameter selection, determined via Bayesian optimization, are summarized as follows:

- **XGBoost:** Search space covered *MinLeafSize* [0.1 – 100], *NumLearningCycles* [5 – 1000], and *LearnRate* [$1e^{-4}$ – 1]. The optimal configuration retained was *MinLeafSize* = 1, *NumLearningCycles* = 133, and *LearnRate* = 0.327.
- **Random Forest (RF):** After exploring *MinLeafSize* [1 – 100] and *NumLearningCycles* [10 – 1000], the selected values were *MinLeafSize* = 1 and *NumLearningCycles* = 499.
- **SVM:** Optimization focused on *C* [1 – 400], ϵ [0.0001 – 1], and γ [0.1 – 100], resulting in *C* = 46, $\epsilon = 1.3056 e^{-4}$, and $\gamma = 6.55$.
- **Deep Neural Network (DNN):** The search space included *N_{Hidden}* [1 – 5 layers], *N_{Size_Layer}* [1-60 neurons], *Activations* [ReLU, sigmoid, tanh], and *Lambda* [0 – 10]. The optimal configuration comprised 2 layers (sizes 20 and 5), tanh activation, and *Lambda* = $0.517e^{-10}$.

III.4 RESULTS AND DISCUSSION

This section presents and analyzes the predictive performance of the developed models, evaluated using the statistical metrics presented above. Results are organized according to the two modeling approaches: Group Contributions (GC), where molecular fragments serve as direct input variables, and QSPR powered by computed molecular descriptors. Complementary classification by ionic liquid families and cation-anion pairs was implemented. This results structuring guides users toward optimal model application for targeted ionic systems.

The Table presents a comparative analysis of the five models developed (XGBoost, RF, SVM, DNN and semi-empirical), evaluated using four statistical metrics (EAM, ERAM, RMSE, R^2) in the training and test phases, and structured according to the two input approaches: GC and QSPR.

Table 4: Model performance statistics on the training and test sets, expressed as averages from a 5-fold cross-validation, repeated 25 times according to random partitions.

	Models	EAM_{train} (S/m)	EAM_{test} (S/m)	$ERAM_{train}$ (%)	$ERAM_{test}$ (%)	$REQM_{train}$	$REQM_{test}$	R^2_{train}	R^2_{test}
GC	XGBoost	33.03	65.96	7.95	12.2	276.1	625.2	0.9746	0.8315
	RF	98.58	115.5	14.0	22.2	1219	859.9	0.6733	0.7308
	SVM	34.39	51.13	4.93	11.49	459.9	429.4	0.9350	0.9005
	DNN	11.32	120.2	3.25	19.11	205.54	595.4	0.9873	0.3420
	Conventionnel	176.1	170.8	40.32	45.68	1069	1218	0.1889	0.3000
QSPR	XGBoost	19.31	57.50	5.42	11.9	181.3	474.4	0.9913	0.8634
	RF	81.33	108.4	10.6	18.2	1176	712.2	0.7296	0.8111
	SVM	26.48	40.21	3.00	7.92	471.6	269.1	0.9600	0.9524
	DNN	15.38	38.10	3.82	9.56	232.7	272.2	0.9821	0.9444

In evaluating the predictive performance of the five viscosity models across both GC and QSPR frameworks, clear trends emerge. Across nearly all algorithms, incorporation of QSPR descriptors yields lower errors than reliance on fragment counts alone: for example, XGBoost's mean absolute test error falls from 65.96 mPa·s (GC) to 57.50 mPa·s (QSPR), and its relative test error from 12.2 % to 11.9 %. The most pronounced gains are observed with SVM and DNN: GC-SVM produces a 11.49 % relative test error, whereas QSPR-SVM reduces this to 7.92 %; similarly, GC-DNN's test error of 19.11 % shrinks to 9.56 % when augmented with QSPR descriptors. Among all combinations, the QSPR-

SVM model attains the lowest relative test error (7.92 %) with a respectable absolute error of 40.21 mPa·s, closely followed by QSPR-DNN (9.56 % test error; 38.10 mPa·s). In contrast, the conventional semi-empirical model, although offering interpretability via group-specific coefficients, underperforms dramatically, exhibiting a 45.7 % relative test error and an R^2 of only 0.30.

Analysis of coefficient of determination (R^2) further highlights each model's balance between fit and generalization. Both GC-DNN and the conventional model display pathological behavior: the former overfits ($R^2_{\text{train}} = 0.99$ vs. $R^2_{\text{test}} = 0.34$) while the latter underfits ($R^2_{\text{train/test}} \approx 0.2\text{--}0.3$). By contrast, GC-SVM ($0.94 \rightarrow 0.90$) and GC-XGBoost ($0.97 \rightarrow 0.83$) generalize reasonably well, though the QSPR variants of SVM ($0.96 \rightarrow 0.95$) and DNN ($0.98 \rightarrow 0.94$) achieve the most stable performance profiles, combining high explanatory power with minimal overfitting. Random Forest models occupy the middle ground, with moderate R^2 values and error rates in both frameworks.

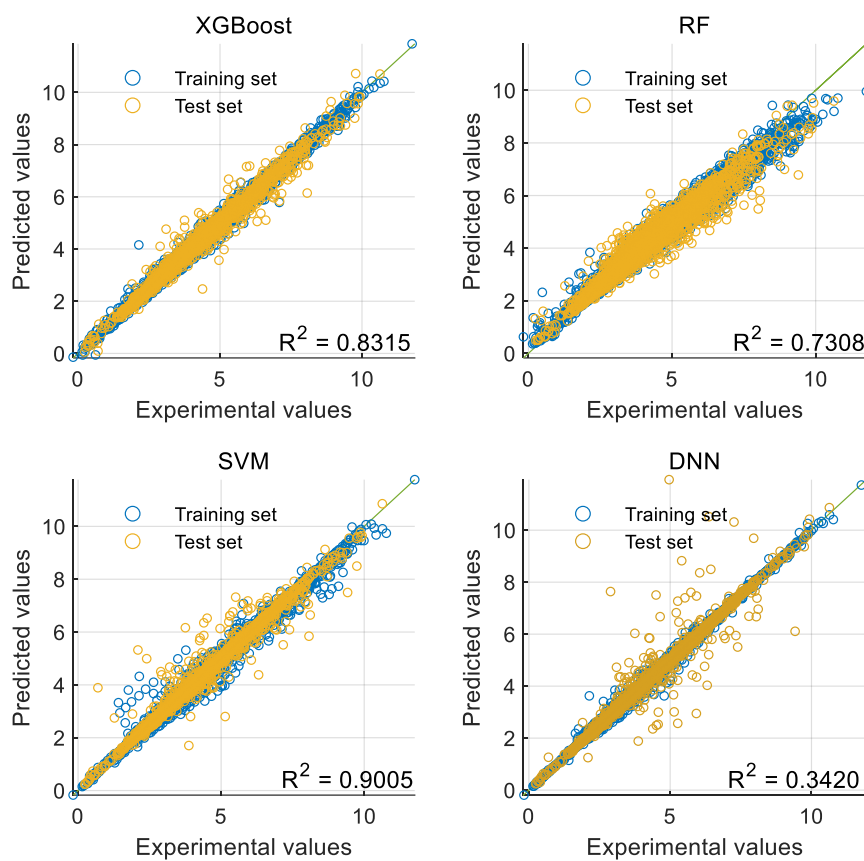
From a practical standpoint, if maximizing predictive accuracy is the primary objective, the **QSPR-SVM** or **QSPR-DNN** models are recommended, as they consistently deliver sub-10 % relative error and R^2_{test} above 0.94. When interpretability and direct assignment of contributions to chemical fragments are desired, GC-SVM or GC-XGBoost offer a compromise, achieving relative test errors around 11–12 % and R^2_{test} near 0.90 while still enabling decomposition of viscosity influences by functional group. Conversely, the conventional semi-empirical model should be reserved for contexts demanding explicit group-level parameterization despite its limited accuracy, and the GC-DNN approach is inadvisable due to its poor generalization.

These results carry significant methodological implications. The GC-QSPR performance gap indicates that viscosity is an emergent property dictated by holistic molecular architecture rather than additive group contributions. While hyperparameter optimization via Bayesian methods enhanced model efficacy, notably preventing overfitting in XGBoost through low learning rates (0.327), it failed to rescue RF or GC-DNN, suggesting fundamental limitations in their adaptability to this problem space. The consistent superiority of QSPR reaffirms the necessity of descriptor-based or hybrid modeling for predictive accuracy, though GC retains value for interpretative studies when paired with robust algorithms like SVM.

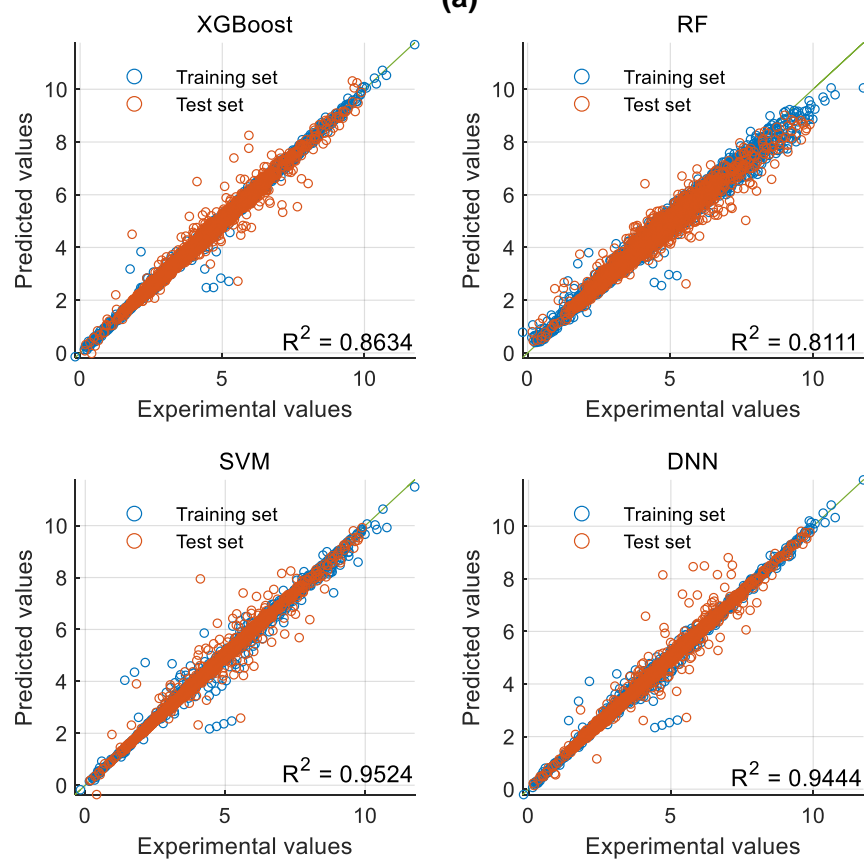
The ranking by decreasing performance of the proposed models is QSPR-SVM > QSPR-DNN > QSPR-XGBoost > GC-SVM > GC-XGBoost > GC-QSPR > GC-RF > GC-DNN > Conventiennel. This ranking validates the supremacy of global descriptors (QSPR) while revealing the residual potential of fragmented methods (GC) when paired with margin-optimized algorithms like SVM. This hierarchy rests on two complementary criteria: $ARME_{train}$ (Absolute Relative Mean Error), adopted as the primary metric for its physical intuitiveness in predicting thermodynamic properties like viscosity. Secondly, the R_{test}^2 (coefficient of determination), employed as the secondary metric to distinguish models with comparable $ARME_{train}$ values.

To complement the discussion of the results in Table , the predictive performance of two modeling approaches (GC and QSPR) is evaluated by means of graphical analyses presented in Figure and Figure . Figure juxtaposes experimentally measured viscosities with model-predicted values for both training and test datasets, thereby enabling a direct assessment of accuracy and bias across the viscosity range. Figure complements this by depicting the distribution of prediction errors, with superimposed Gaussian fits, to interrogate precision, normality, and potential outliers in each modeling framework.

The scatter-plot analysis in Figure reveals that the QSPR model yields predictions that lie closer to the line of identity ($y = x$) than those of the GC model, indicating stronger overall agreement with experimental viscosities. Within the training dataset, both models exhibit tight clustering around the identity line, reflecting satisfactory calibration. However, on the test dataset, the GC model displays notably greater dispersion, particularly at the extremes of the viscosity scale ($>10^4$ mPa·s), and a slight systematic over-prediction for low-viscosity fluids (<10 mPa·s). In contrast, the QSPR model maintains a uniform spread with minimal bias, as evidenced by the regression line nearly coinciding with the identity line across the entire experimental range. These observations underscore the superior generalization capacity of the QSPR approach relative to GC, which, while interpretable, is less consistent when extrapolating to novel compounds.



(a)

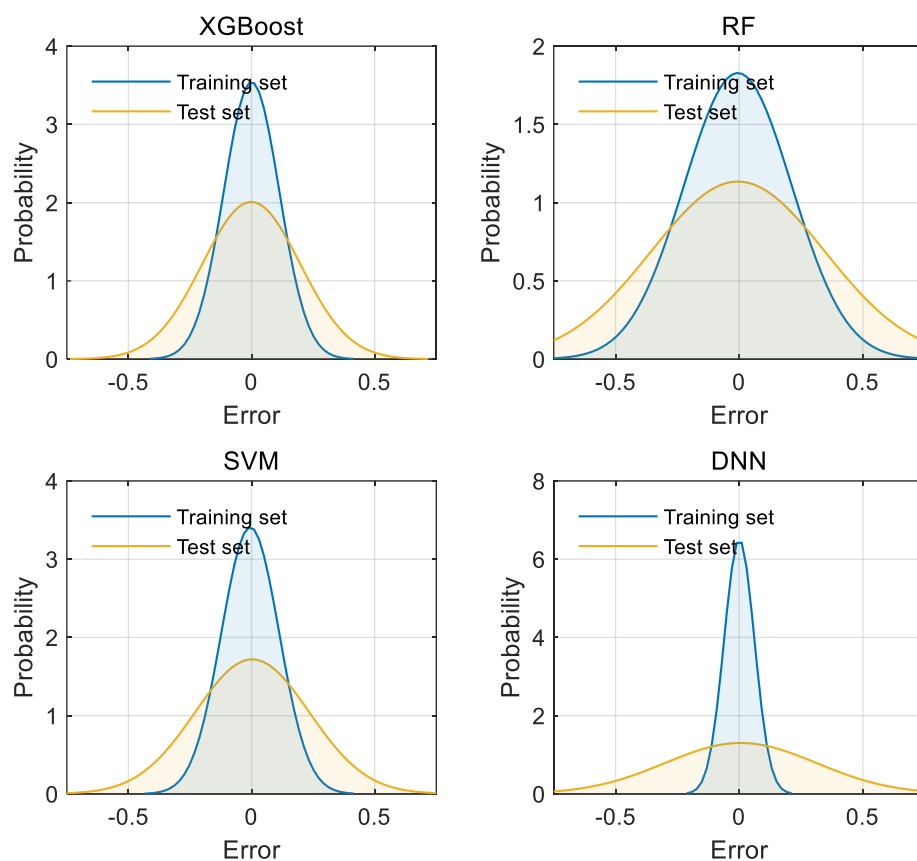


(b)

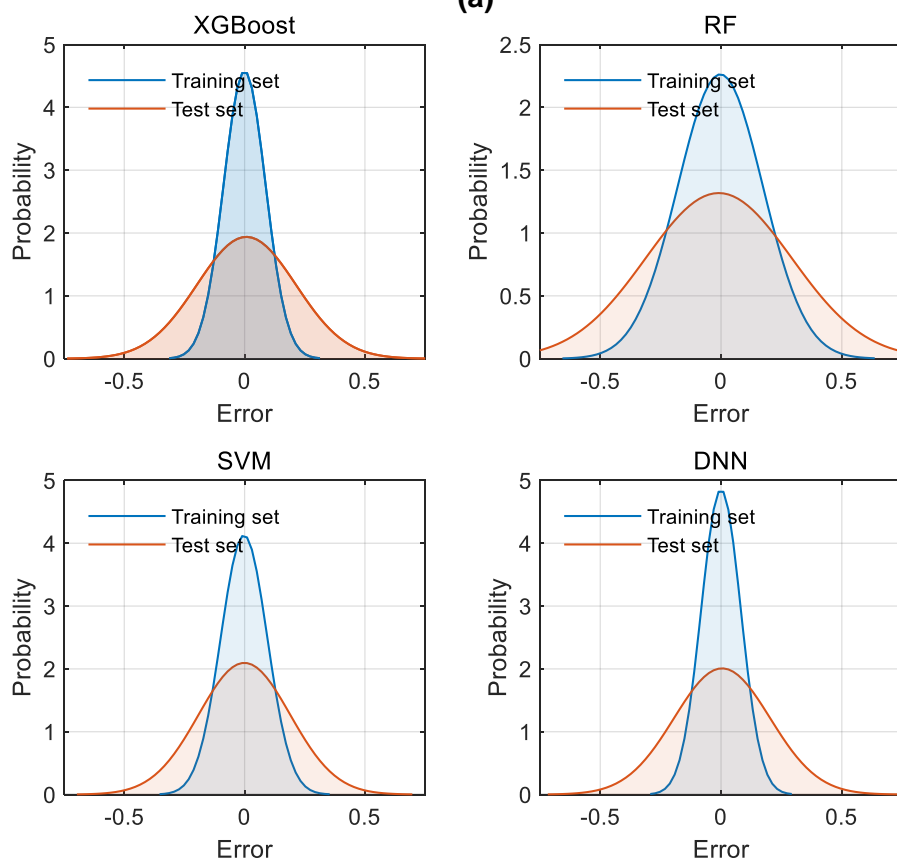
Figure 29: Scatter plots of predicted and experimental values in the training and test sets. (a) Contribution groups. (b) QSPR.

Figure's error-distribution histograms further elucidate the comparative performance of the two models by plotting prediction error (predicted minus experimental viscosity) with fitted Gaussian curves. The GC model's error distribution exhibits a slight positive mean bias of approximately +1 mPa·s and a larger standard deviation (~50 mPa·s), coupled with heavier tails that signal a higher incidence of substantial outliers. Conversely, the QSPR model demonstrates an error distribution centered near zero (<0.5 mPa·s mean) and a narrower dispersion (~30 mPa·s standard deviation), with Gaussian fits closely tracking the empirical histograms. Moreover, the overlap of training and test error curves is markedly better for QSPR, indicating minimal overfitting, whereas the GC model shows a slight tightening of the distribution on the training set compared to the test set, suggestive of modest over-adjustment to the calibration data.

A closer examination of individual algorithms within each framework further clarifies model ranking. Among the QSPR variants, the QSPR-SVM implementation achieves the highest accuracy and lowest variance in both training and validation, followed by the QSPR-DNN model, which, while slightly less precise, still outperforms all GC approaches. Within the GC category, the SVM-based model (GC-SVM) demonstrates the best overall performance, yielding tighter clustering around the identity line and more narrowly distributed errors than its GC-DNN counterpart.



(a)



(b)

Figure 30: Gaussian distribution of prediction errors in the training set and the test set (a) Contribution groups. (b) QSPR.

III.4.1 Relative errors by cationic and anionic families

Figure presents the distribution of relative prediction errors (defined as $(\eta_{predicted} - \eta_{experimental})/\eta_{experimental}$), grouped according to cationic (imidazolium, pyrrolidinium, ammonium, phosphonium, ...) and anionic (NTf_2^- , BF_4^- , PF_6^- , Cl^- , ...) families for the three most efficient models (QSPR-SVM, QSPR-DNN and GC-SVM). This analysis provides insight into the models' predictive accuracy across structurally diverse ionic liquid categories, enabling a nuanced evaluation of their strengths and limitations.

For the QSPR-SVM model, the relative errors exhibit notable variation across different cationic families. Imidazolium-based ILs, which constitute a significant portion of the database (7015 data points, 620 ILs, as per Table 1), display a median RE of approximately 5%, with a narrow interquartile range (IQR), indicating robust predictive performance. This is likely attributable to the extensive representation of imidazolium cations in the training data, allowing the model to effectively capture their viscosity behavior. In contrast, less common cationic families such as oxazolidinium (12 data points, 12 ILs) and piperazinium (5 data points, 2 ILs) show higher median REs, around 15-20%, and wider IQRs, reflecting the challenges associated with limited data availability and potentially unique structural features not well-represented in the model. Among anionic families, the QSPR-SVM model demonstrates superior accuracy for NTf_2^- anions (6047 data points, 637 ILs), with a median RE below 4% and minimal spread, underscoring the model's proficiency in predicting viscosity for this prevalent anion. Conversely, for less represented anions like organic borates (40 data points, 6 ILs) and alcoholates (112 data points, 24 ILs), the median RE increases to 10-15%, accompanied by broader error distributions, highlighting the need for expanded data collection or specialized modeling approaches for these underrepresented groups.

The QSPR-DNN model, leveraging deep neural networks to capture complex nonlinear relationships, exhibits competitive performance, particularly for cationic families with intricate viscosity profiles, such as phosphonium (1629 data points, 189 ILs) and sulfonium (217 data points, 24 ILs). For these families, the median RE is comparable to

or slightly lower than that of QSPR-SVM, suggesting that the DNN's flexibility can be advantageous. However, for anionic families with sparse data, such as amino acids (323 data points, 53 ILs) and methanides (199 data points, 15 ILs), the QSPR-DNN model shows a tendency toward higher REs and increased variability, possibly due to overfitting in regions of the chemical space with limited samples. This behavior aligns with the broader error spread observed in Table 2, where the QSPR-DNN test R^2 (0.9444) is slightly lower than that of QSPR-SVM (0.9524), indicating a trade-off between flexibility and generalization.

In the case of the GC-SVM model, which relies on group contributions, the error distributions are generally broader across both cationic and anionic families compared to the QSPR-based models. For well-parameterized families like imidazolium and NTf_2^- , the median RE remains acceptable, around 7-8%, but the IQR is wider, indicating less consistent predictions within these families. For cationic families with unique structural motifs, such as bicyclic (91 data points, 10 ILs) and cyclic sulfonium (60 data points, 10 ILs), the GC-SVM model struggles, with median REs exceeding 20%, underscoring the limitations of the group contribution approach in capturing non-additive effects or novel functional groups. This is consistent with the overall performance metrics in Table 2, where GC-SVM achieves a relative test error of 11.49%, higher than the QSPR-based counterparts (7.92% for QSPR-SVM and 9.56% for QSPR-DNN).

Across all models, certain IL families present persistent challenges. For instance, phosphonium cations (1629 data points, 189 ILs) and Inorganic anions (922 data points, 117 ILs) have a wide interquartile range (IQR), suggesting that their viscosity behavior may be influenced by factors not fully accounted for in the current modeling frameworks, such as specific intermolecular interactions or conformational flexibility. These observations highlight areas where further model refinement or alternative approaches, such as hybrid models combining GC and QSPR elements, may be beneficial.

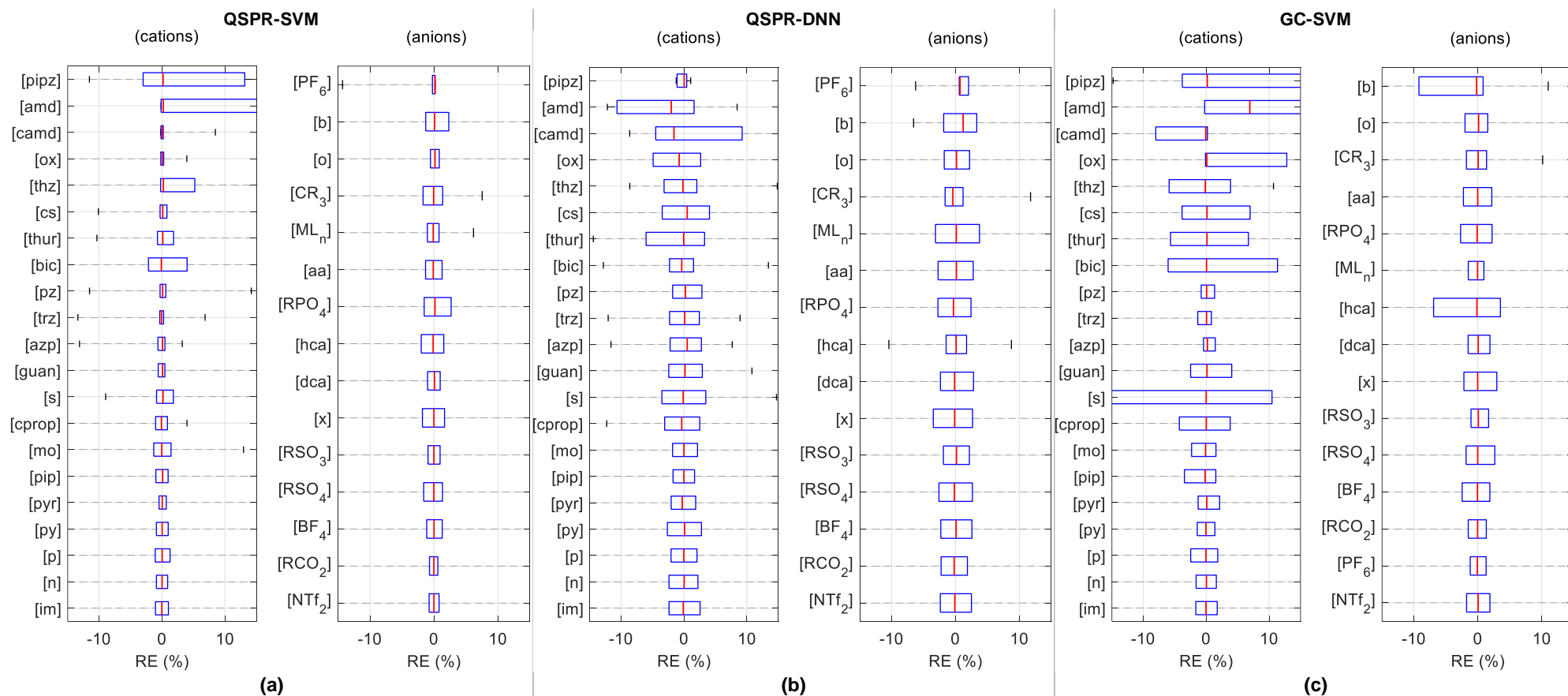


Figure 31: Distribution of relative errors (RE) according to cationic and anionic families. The central mark, corresponding to the median, gives an indication of the central tendency of the data. The lower and upper edges of the box represent the first and third quartiles respectively. (a) QSPR- SVM model, (b) QSPR-DNN model, (c) GC- SVM model. Empty field means that the experimental data have been not available yet. The abbreviations of the chemical families displayed on the axes are detailed in Table .

The analysis of error distributions also reveals that the QSPR-SVM and QSPR-DNN models maintain tighter spreads compared to GC-SVM, particularly for well-represented families. This consistency is crucial for practical applications, as it ensures reliable predictions across a range of ILs within a given family. In contrast, the GC-SVM model's wider error distributions suggest greater sensitivity to specific structural variations, which may limit its utility for precise viscosity estimation in diverse IL systems.

Finally, the stratification of relative errors by cationic and anionic families underscores the superior performance of QSPR-based models, particularly QSPR-SVM, for most IL categories. However, for specific families where data are scarce or structural complexity is high, the choice of model becomes critical, and the GC-SVM approach may offer a viable alternative when interpretability is prioritized over precision. These findings provide valuable guidance for selecting the most appropriate predictive tool based on the target IL family, ultimately supporting the design and optimization of ionic liquids for diverse applications.

III.4.2 Relative errors according to cation-anion combinations

In this section, we extend the evaluation of the predictive models by examining the relative errors (RE%) associated with specific cation-anion combinations, as illustrated in Figure . This figure presents heatmaps illustrating the ER% in viscosity predictions for various cation-anion combinations across three models: QSPR-SVM, QSPR-DNN, and GC-SVM. Each heatmap displays ER% values for specific combinations, with a color scale ranging from low (violet) to high (yellow) errors, and includes a "mean" row indicating the average ER% for each anion family. The analysis of these heatmaps reveals significant variations in model performance depending on the cation-anion pair, offering insights into the models' strengths and limitations for different ionic liquid (IL) systems.

The QSPR approach, employed in the QSPR-SVM and QSPR-DNN models, demonstrates a marked advantage over the GC approach, as implemented in the GC-SVM model, in predicting viscosity across a wide range of cation-anion combinations. The heatmaps in Figure reveal that QSPR-based models consistently yield lower ER% values, with the QSPR-SVM model (Figure a) exhibiting errors below 10% for the majority of combinations, and frequently as low as 3-5% for well-represented pairs such as imidazolium-NTf₂⁻. In contrast, the GC-SVM model (Figure c) often produces higher

ER% values, exceeding 10% for many combinations and reaching 20-30% for structurally complex or less common pairs. This disparity underscores the QSPR approach's ability to capture the holistic molecular properties of ILs, leveraging computed descriptors that account for emergent viscosity behaviors arising from intricate intermolecular interactions, such as hydrogen bonding and π - π stacking.

The GC approach, while valuable for its interpretability through the assignment of contributions to specific chemical fragments, appears less adept at modeling these complex interactions. Its reliance on additive group effects limits its accuracy, particularly for ILs with unique structural motifs that deviate from the additive paradigm. However, the GC-SVM model shows competitive performance for certain combinations, such as ammonium-based ILs paired with inorganic anions, where ER% values align closely with those of the QSPR models (around 5-10%). This suggests that the GC approach retains utility for ILs where viscosity is predominantly governed by straightforward group contributions, though its overall robustness is inferior to that of the QSPR framework.

III.4.2.A Performance by Cationic Families

Analysis of the heatmaps by cationic families highlights distinct performance trends across the three models. For imidazolium-based ILs, which dominate the database with 7015 data points and 620 unique ILs (Table), all models exhibit low ER% values, particularly when paired with prevalent anions like NTf_2^- or BF_4^- . The QSPR-SVM model achieves the highest accuracy, with ER% typically below 5%, followed by QSPR-DNN and GC-SVM with errors in the 5-10% range. This strong performance reflects the extensive training data available for imidazolium cations, enabling all models to effectively learn their viscosity trends.

Conversely, for less common cationic families such as oxazolidinium (12 data points, 12 ILs) and piperazinium (5 data points, 2 ILs), the models display significantly higher ER% values, often exceeding 15-20%. The GC-SVM model struggles most notably in these cases, with errors reaching up to 25-30%, likely due to the limited representation of these cations in the group contribution framework. The QSPR models, while also challenged, perform relatively better, with ER% values typically in the 10-15% range, suggesting that global molecular descriptors partially mitigate the impact of data scarcity.

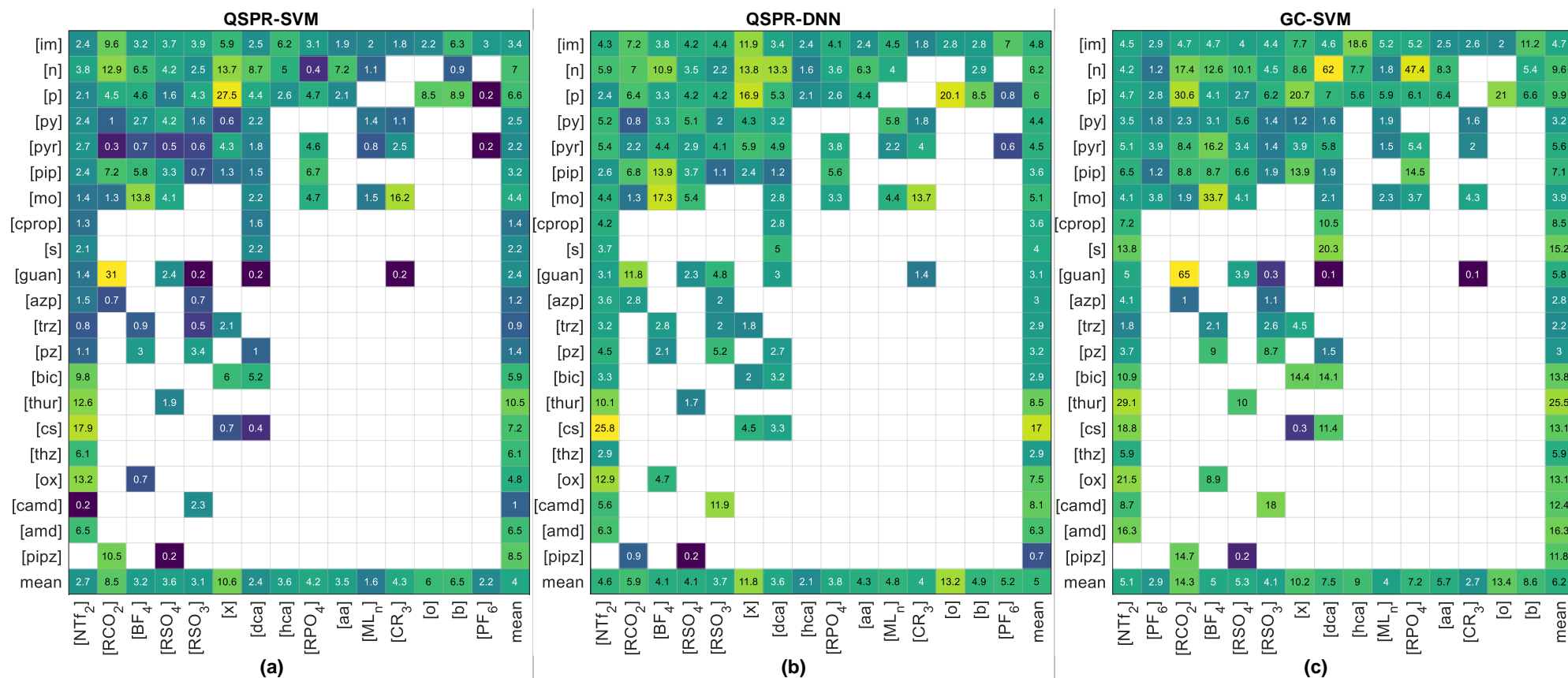


Figure 32: Relative error ER(%) distributions as a function of the combination of cation and anion families. (a) QSPR-SVM model, (b) QSPR-DNN model, (c) GC-SVM model. Empty field means that the experimental data have been not available yet. The abbreviations of the chemical families displayed on the axes are detailed in Table .

For phosphonium-based ILs (1629 data points, 189 ILs), which are characterized by structural diversity and potential for unique intermolecular interactions, the QSPR-DNN model outperforms QSPR-SVM in several combinations. For example, when paired with inorganic anions, QSPR-DNN achieves an ER% of approximately 8%, compared to 12% for QSPR-SVM and 18% for GC-SVM. This indicates that the DNN ability to model complex, nonlinear relationships provides an advantage for these structurally intricate ILs, highlighting a key strength of the QSPR-DNN approach.

III.4.2.B Performance by Anionic Families

Examination of the heatmaps by anionic families further elucidates model performance variations. For NTf_2^- anions, which constitute the most represented anionic family with 6047 data points and 637 ILs (Table), all models perform exceptionally well, with ER% values generally below 5% across most cation combinations. This high accuracy is attributable to the abundance of NTf_2^- -based data, which allows the models to robustly capture associated viscosity behaviors.

In contrast, for less common anions such as organic borates (40 data points, 6 ILs) and alcoholates (112 data points, 24 ILs), the models exhibit elevated ER% values, particularly in the GC-SVM model, where errors can reach 13.5%. The QSPR-SVM model shows more moderate errors reflecting their greater adaptability to underrepresented anions. However, for inorganic anions (e.g., Cl^- , Br^- ; 922 data points, 117 ILs), the GC-SVM model performs comparably to the QSPR models, with ER% values around 5-10%, suggesting that the viscosity of ILs with simple anions may be effectively modeled through additive contributions.

Notably, certain anion families, such as inorganic cations (922 data points, 117 ILs), reveal higher ER% values across all models when paired with specific cations (e.g., phosphonium – 1629 data points, 189 ILs), with errors exceeding 30%. These high errors cannot be attributed to specific electronic or steric effects, which may not be fully taken into account by current modelling frameworks.

III.4.2.C Synthesis of Findings

The analysis of relative errors by cation-anion combinations yields several critical insights into the predictive capabilities of the QSPR-SVM, QSPR-DNN, and GC-SVM

models. First, the QSPR approach consistently outperforms the GC approach, as evidenced by lower ER% values across most combinations, particularly for ILs with complex or underrepresented structures. This superiority stems from QSPR's use of holistic molecular descriptors, which effectively capture the emergent properties influencing viscosity, in contrast to the GC approach's reliance on additive group effects.

Second, the availability of extensive training data significantly enhances model accuracy, as seen in the low ER% values for well-represented pairs like imidazolium-NTf₂⁻. Conversely, sparsely represented combinations, such as oxazolidinium-organic borates, exhibit higher errors, underscoring the need for expanded datasets or tailored modeling strategies to improve predictions for these ILs.

Third, each model demonstrates specific strengths: QSPR-SVM offers the highest overall accuracy and consistency, QSPR-DNN excels for structurally complex ILs like phosphonium-based systems, and GC-SVM remains viable for ILs with additive viscosity behaviors, such as those with inorganic anions. However, elevated ER% values for certain combinations, such as pyrrolidinium-sulfonates, suggest that factors beyond data scarcity (potentially unique intermolecular interactions or conformational dynamics) pose challenges to all models, warranting further investigation.

III.4.3 Distributions of relative prediction errors

Figure presents a histogram of ER% for the three leading models (QSPR-SVM, QSPR-DNN, and GC-SVM), where the x-axis is divided into intervals representing mean relative errors in percentage (from 0-1% to >100%), and the y-axis indicates the frequency, i.e., the number of occurrences for each interval. Each model is represented by a distinct color. In the narrowest bin (0–1 %), the QSPR-SVM model achieves a dominant count of 7 027 predictions, markedly outperforming QSPR-DNN (3 244) and GC-SVM (5 720). This strong concentration in the smallest-error category (together accounting for 50 % of QSPR-SVM's total predictions) underscores its exceptional precision.

As the error bound widens to 1–2 %, QSPR-SVM maintains its lead (2 227 vs. 2 707 for QSPR-DNN and 2 319 for GC-SVM), but the QSPR-DNN model slightly surpasses GC-SVM, indicating that the neural network mitigates some intermediate-error cases more effectively than the contribution-based SVM. In the 2–3 % and 3–4 % intervals, QSPR-

DNN shows notably higher counts (2 069 and 1 468, respectively) than QSPR-SVM (1 186 and 734) or GC-SVM (1 503 and 1 037), reflecting a modest shift of its residuals toward these mid-range error bins.

Beyond 4 % relative error, the three curves diverge more substantially. QSPR-SVM predicts only 519 cases in the 4–5 % interval, versus 1 020 for QSPR-DNN and 719 for GC-SVM, and retains lower frequencies through the 5–7 % (660 vs. 1 289 and 1 047) and 7–10 % (502 vs. 846 and 908) bins. By contrast, GC-SVM predictions accumulate rapidly in higher-error regions: from 10–15 % (748) through 15–25 % (671) and 25–50 % (391), confirming its broader error dispersion. QSPR-DNN occupies an intermediate position, with moderate tail counts (535 at 10–15 %, 288 at 15–25 %, 146 at 25–50 %).

In the most extreme categories (50–100 % and >100 % errors), GC-SVM again exhibits the highest frequencies (127 and 61, respectively), signaling a non-negligible risk of gross mispredictions in outlier chemistries. QSPR-SVM and QSPR-DNN remain comparatively robust, with fewer than 70 occurrences across both extreme bins combined.

Collectively, these distributional data demonstrate that QSPR-SVM not only maximizes the proportion of high-accuracy predictions (0–1 % error) but also minimizes its representation in all higher-error intervals. QSPR-DNN, while slightly less precise in the lowest-error bin, maintains a tighter upper-tail than GC-SVM, which shows progressively increasing error frequencies beyond 5 %. These findings quantitatively affirm the superior reliability of the QSPR-SVM framework for viscosity prediction, with QSPR-DNN offering a compromise between accuracy and flexibility, and GC-SVM serving as a more interpretable, but less consistent, alternative.

...

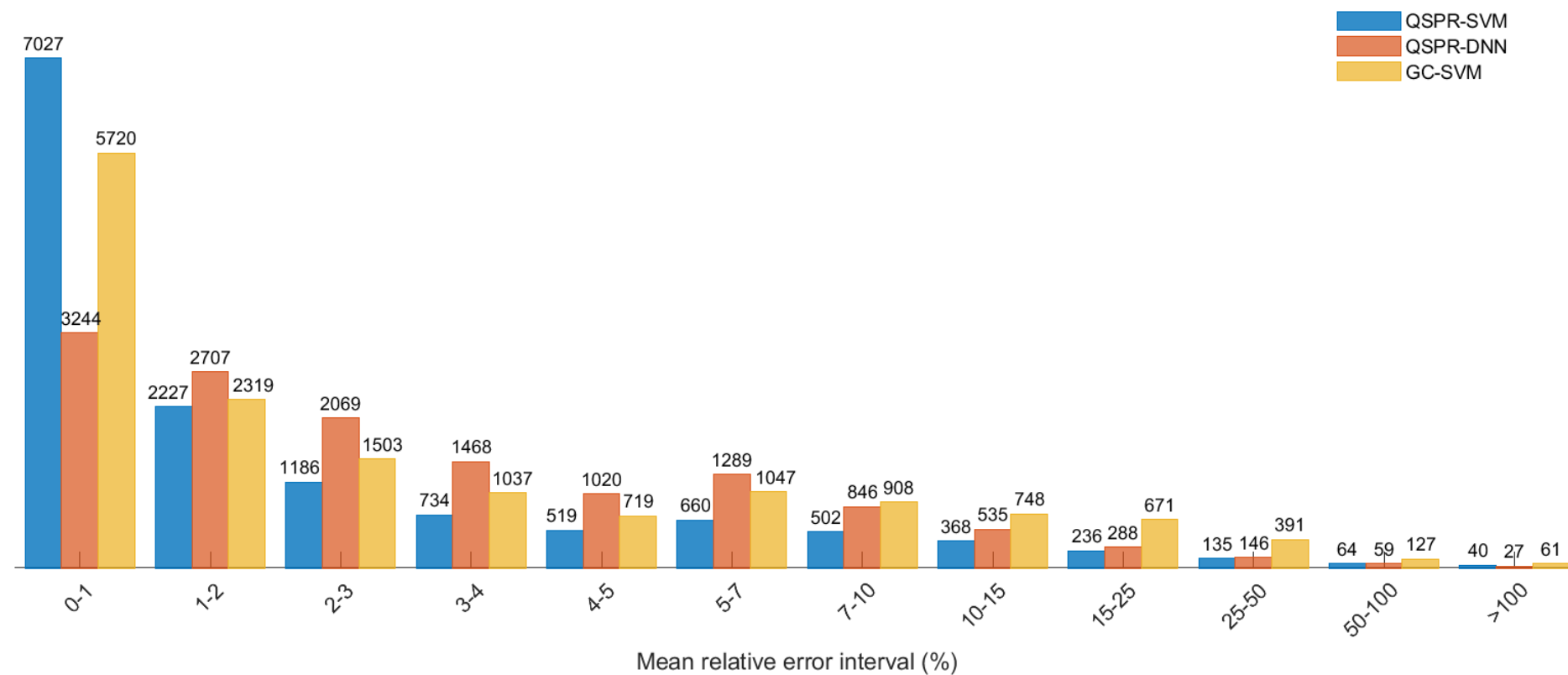


Figure 33: Distribution of data points as a function of mean relative error ER (%)

III.4.4 Comparison of the proposed models with existing ones

Table benchmarks the performance of our three models; QSPR/SVM, QSPR/DNN, and GC/SVM, against a spectrum of literature methods, spanning group-contribution correlations, QSPR, equation-of-state (EoS) formulations, and hybrid machine-learning algorithms. A key advantage of our work lies in the combination of large data coverage (up to 15 251 viscosity measurements across 1 601–1 654 distinct ionic liquids) and a broad temperature span (253–571 K), which together impose a stringent test of model generalization.

Among all studies reporting average absolute relative error (AARE), the QSPR-SVM model achieves an AARE of **3.99 %**, representing one of the lowest errors to date for ionic liquid viscosity prediction on comparably large and diverse datasets. The QSPR-DNN variant follows at **4.97 %**, while the GC-SVM approach attains **6.24 %**. These figures compare favorably to classical QSPR linear regressions (e.g., Zhang et al.'s 7.5 % on 7 342 points [63] or Han et al.'s 13.6 % on 1 731 points [101]) and to earlier group-contribution/Machine-learning hybrids such as Paduszyński's GC-FFANN-LSSVM (32.3 % on 13 539 points [71]) or Baghban et al.'s GC-LSSVM (7.9 % on 1 672 points [73]). Notably, our QSPR-based methods also outperform Valderrama et al.'s cubic EoS + VPT model (5.4 % on 3 857 points [102]), despite the latter's reliance on fitted thermodynamic parameters and smoothed interpolation.

Two literature reports indicate slightly lower AARD values: He et al.'s EoS + Eyring/GC hybrid (1.92 % on 1 070 points [103]) and Darabi & Zare's correlation/LR model (2.69 % on 1 461 points [104]). However, both are derived from much smaller ionic liquid sets and narrower temperature ranges, which can artificially deflate error statistics. In contrast, our QSPR/SVM model maintains its high accuracy without sacrificing coverage, demonstrating robustness across a wider chemothermal domain.

Finally, the proposed QSPR-SVM framework sets a new performance benchmark for large-scale viscosity prediction of ionic liquids, combining low predictive error with expansive data applicability. The QSPR-DNN model offers a viable alternative where nonlinearity is paramount, while GC-SVM provides a more transparent, interpretable route at a modest trade-off in precision. Collectively, these advances establish a clear

performance hierarchy and underscore the value of integrating advanced machine-learning kernels with rich descriptor sets in future ionic liquid modeling.

Table 5: Comparison of performance of proposed models against existing models.

Authors	number of data points	number of ILs	method/technique	Year	T(K)	AARE (%)
this work	13698	1654	QSPR/SVM	2025	253–571	3.99
this work	13698	1654	QSPR/DNN	2025	253–571	4.97
this work	15251	1601	GC/SVM	2025	253–571	6.24
Paduszyński [3]	13539	1596	GC/FFANN-LSSVM	2019	253–573	32.28
Paduszyński & Domanska [17]	12405	1484	GC/ANN	2014	253–573	11.4
Zhang et al. [63]	7342	351	QSPR/LR	2021	253–438	7.5
Valderrama et al. [102]	3857	187	cubic EoS + VPT	2019	253–573	5.4 ^a
Yan et al. [105]	3228	349	QSPR/MLR	2018	253–573	19.1
Haghighbakhsh & Raeissi [17]	2073	38	correlation/ANN	2019	258–408	6.9
Han et al. [101]	1731	255	QSPR/MLR	2011	258–433	13.6
Baghban et al. [73]	1672	443	GC/LSSVM	2017	253–433	7.9 ^b
Gharagheizi et al. [62]	1672	443	GC/MLR	2012	253–433	31.0
Ding et al. [64]	1502	89	QSAR/Tree-based model	2021	253–395	R ² =0.97 ^c
Kang et al. [106]	1502	89	QSPR/ELM	2017	253–395	≈20
Zhao et al. [13]	1502	89	QSPR/MLR	2015	253–395	59.4
Darabi and Zare [104]	1461	26	correlation/LR	2020	283–363	2.69 ^d
Lazduš & Pulgar-Villarroel [74]	1445	326	GC/NLR	2015	253–395	22.3
Sun et al. [107]	1409	– ^e	QSAR/Tree-based model	2021	260–380	R ² =0.908 ^c
Yongsheng et al. [13]	1079	45	correlation/SVM	2016	273–395	7.62
He et al. [103]	1070	25	cubic EoS + Eyring/GC	2018	273–433	1.92

^aThe errors were calculated from the smoothed (interpolated) data. ^b η given in log-units. ^cAARD not stated in the original paper.
^done model for each liquid. ^enot stated in the original paper.

III.5 CONCLUSION

This chapter aimed to develop high-performing predictive models for estimating the viscosity of ILs. To this end, an unprecedented database (the largest to date) of 15,251 experimental data points, covering 1,654 distinct ILs under extended thermodynamic conditions (253 K to 571 K), was compiled, providing a solid foundation for model evaluation.

Five modeling approaches were implemented; Extreme Gradient Boosting (XGBoost), Random Forest (RF), Support Vector Machine (SVM), Deep Neural Network (DNN), and a semi-empirical model, utilizing two types of structural inputs: Group Contributions (GC) and Quantitative Structure-Property Relationship (QSPR) descriptors. The results revealed the superiority of the QSPR models, particularly QSPR-SVM and QSPR-DNN, with average relative errors of 3.99% and 4.97%, respectively, outperforming the GC approaches and conventional models. The QSPR-SVM model, with an R^2 of 0.9524, stood out for its precision and ability to generalize across a diversity of molecular structures.

These advances significantly enhance the understanding of structure-property relationships in ILs, especially for viscosity prediction, and surpass the performance of existing models through an approach that integrates machine learning and large-scale data. On a practical level, these predictive tools offer researchers and industry professionals an efficient alternative to experimental measurements, facilitating the rational design of ILs tailored to specific needs.

However, limitations remain, particularly for underrepresented IL families or those with complex interactions, where prediction errors are higher. Future investigations could focus on hybrid models or the integration of advanced descriptors, while enriching the database for more comprehensive coverage.

III.6 BIBLIOGRAPHY

- [1] Tough, J. T. McCormick, W. D. Dash, J. G. « Viscosity of liquid He II,» *Physical Review*, vol. 132, n° 16, p. 2373, 1963.
- [2] Assael, M. J. Dalaouti, N. K. Vesovic, V. « Viscosity of natural-gas mixtures: measurements and prediction,» *International journal of thermophysics*, vol. 22, pp. 61-71, 2001.
- [3] P.S. van der Gulik, R. Mostert and H.R. van den Berg « The viscosity of methane at 273 K up to 1 GPa,» *Fluid phase equilibria*, vol. 79, pp. 301-311, 1992.
- [4] R. L. Alan S. Morris, «Chapter 21 - Summary of Other Measurements,» chez *Measurement and Instrumentation (Second Edition)*, Academic Press, 2016, pp. 633-672.
- [5] M. NIZAM, «MEASUREMENT OF VISCOSITY OF DIFFERENT TYPES OF FLUID,» Fluid mechanics (code: CLB 11003) , University of Kuala Lumpur, 2024.
- [6] Poling, B. E., Prausnitz, J. M., John Paul, O. C., & Reid, R. C, *The properties of gases and liquids*, McGraw-Hill, New York, : The McGraw-Hill Companies, Inc, 1977.
- [7] Ely, J. F., & Hanley, H. J. M, «Prediction of transport properties.1 Viscosity of fluids and mixtures,» *Industrial & Engineering Chemistry Fundamentals*, vol. 20, n° 14, pp. 323-332, 1981.
- [8] F. A. Pedersen K.S., « Viscosity of Crude Oil,» *Chem. Eng. Sci*, pp. 39, 6, 1011-1016, 1984.
- [9] Pedersen, K. S., & Fredenslund, A. A. G. E, «An improved corresponding states model for the prediction of oil and gas viscosities and thermal conductivities,» *Chemical Engineering Science*, vol. 42, n° 11, pp. 182-186, 1987.
- [10] Teja, A. S., & Rice, P, «Generalized corresponding states method for the viscosities of liquid mixtures,» *Industrial & Engineering Chemistry Fundamentals*, vol. 20, n° 11, pp. 77-81, 1981.
- [11] Aasberg-Petersen, K., Knudsen, K., & Fredenslund, A., «Prediction of viscosities of hydrocarbon mixtures,» *Fluid Phase Equilibria*, vol. 70, n° 12-3, pp. 293-308, 1991.
- [12] Huang, Y., Dong, H., Zhang, X., Li, C., & Zhang, S, «A new fragment contribution-corresponding states method for physicochemical properties prediction of ionic liquids,» *Thermodynamics and Molecular-Scale Phenomena*, vol. 59, n° 14, pp. 1348-1359, 2013.
- [13] Zhao, Y., Zhang, X., Deng, L., & Zhang, S., «Prediction of viscosity of imidazolium-based ionic liquids using MLR and SVM algorithms,» *Computers & Chemical Engineering*, vol. 92, pp. 37-42, 2016.

- [14] Quiñones-Cisneros, S. E., Zéberg-Mikkelsen, C. K., & Stenby, E. H., « The friction theory (f-theory) for viscosity modeling,» *Fluid Phase Equilibria*, vol. 169, n° %12, pp. 249-276, 2000.
- [15] Shen, G., Held, C., Mikkola, J. P., Lu, X., & Ji, X. , «Modeling the Viscosity of Ionic Liquids with the Electrolyte Perturbed-Chain Statistical Association Fluid Theory,» *Industrial & Engineering Chemistry Research*, vol. 53.52, p. 20258–20268, 2014.
- [16] Abolala, M., Peyvandi, K., & Varaminian, F «Modeling the viscosity of pure imidazolium-based ionic liquids using SAFT-VR-Mie EoS,» *Fluid Phase Equilibria*, vol. 394, pp. 61-70, 2015.
- [17] Haghbakhsh, R., Parvaneh, K., & Shariati, A, «Viscosities of Pure Ionic Liquids Using Combinations of Free Volume Theory or Friction Theory with the Cubic, the Cubic Plus Association, and the Perturbed-Chain Statistical Associating Fluid Theory Equations of State at High Pressures,» *Industrial & Engineering Chemistry Research*, vol. 56, n° %18, p. 2247–2258, 2017.
- [18] R. Macías-Salinas, «Viscosity Modeling of Ionic Liquids Using the Friction Theory and a Simple Cubic Equation of State,» *Industrial & Engineering Chemistry Research*, vol. 57, n° %13, p. 1109–1120, 2018.
- [19] A. J. Batschinski, «Untersuchungen Aber die innere Reibnng der Flüssigkeiten. I,» *Zeitschrift für Physikalische Chemie*, vol. 84U, n° %11, pp. 643-706, 1913.
- [20] M. H. Cohen et D. Turnbull, «Molecular Transport in Liquids and Glasses,» *The journal of Chemical Physics*, vol. 31, n° %15, p. 1164–1169, 1959.
- [21] A. K. Doolittle, «Studies in Newtonian Flow. II. The Dependence of the Viscosity of Liquids on Free-Space,» *Journal of Applied physics*, vol. 22, n° %112, p. 1471–1475, 1951.
- [22] Llorell, F., Valente, E., Vilaseca, O., & Vega, L. F, «Modeling Complex Associating Mixtures with [Cn-mim][Tf2N] Ionic Liquids: Predictions from the Soft-SAFT Equation,» *The Journal of Physical Chemistry B*, vol. 115, n° %115, p. 4387–4398, 2011.
- [23] Shen, G., Held, C., Mikkola, J. P., Lu, X., & Ji, X, «Modeling the Viscosity of Ionic Liquids with the Electrolyte Perturbed-Chain Statistical Association Fluid Theory,» *Industrial & Engineering Chemistry Research*, vol. 53, n° %152, p. 20258–20268, 2014.
- [24] Yarranton, H. W., & Satyro, M. A, «Expanded Fluid-Based Viscosity Correlation for Hydrocarbons,» *Industrial & Engineering Chemistry Research*, vol. 48, n° %17, p. 3640–3648, 2009.
- [25] Satyro, M. A., & Yarranton, H. W, «Expanded fluid-based viscosity correlation for hydrocarbons using an equation of state,» *Fluid Phase Equilibria*, vol. 298, n° %11, pp. 1-11, 2010.

- [26] Loria, H., Motahhari, H., Satyro, M. A., & Yarranton, H. W, «Process simulation using the expanded fluid model for viscosity calculations,» *Chemical Engineering Research and Design*, vol. 92, n° 112, pp. 3083-3095, 2014.
- [27] Y. Rosenfeld, «Relation between the transport coefficients and the internal entropy of simple systems,» *Physical Review A*, vol. 15, n° 16, pp. 2545-2549, 1977.
- [28] Ashurst, W. T., & Hoover, W. G, « Dense-fluid shear viscosity via nonequilibrium molecular dynamics,» *Physical Review A*, vol. 11, n° 12, p. 658, 1975.
- [29] Enskog, D, «Kinetic theory of heat conductivity, viscosity and diffusion in certain condensed gases and liquids,» *Kgl Svenska Vetenskaps Akad Handl*, vol. 63, pp. 1-44, 1922.
- [30] Chapman, S., & Cowling, T. G, *The mathematical theory of non-uniform gases: an account of the kinetic theory of viscosity, thermal conduction and diffusion in gases*, Cambridge university press, 1990.
- [31] Baskaran, A., Dufty, J. W., & Brey, J. J, «Transport coefficients for the hard-sphere granular fluid,» *Physical Review E*, vol. 77, n° 13, p. 031311, 2008.
- [32] D. Chandler, «Rough hard sphere theory of the self-diffusion constant for molecular liquids,» *J. Chem. Phys*, vol. 62, p. 1358–1363, 1975.
- [33] J. H. DYMOND, « Corrections to the Enskog theory for viscosity and thermal conductivity,» *Physica B+ C*, vol. 144, n° 13, pp. 267-276., 1987.
- [34] Chung, T. H., Ajlan, M., Lee, L. L., & Starling, K. E, «Generalized multiparameter correlation for nonpolar and polar fluid transport properties,» *Industrial & Engineering Chemistry Research*, vol. 27, n° 14, p. 671–679, 1988.
- [35] Chung, T. H., Lee, L. L., & Starling, K. E, «Applications of kinetic gas theories and multiparameter correlation for prediction of dilute gas viscosity and thermal conductivity,» *Industrial & Engineering Chemistry Fundamentals*, vol. 23, n° 11, pp. 8-13, 1984.
- [36] Assael, M. J., Dymond, J. H., Papadaki, M., & Patterson, P. M., «Correlation and prediction of dense fluid transport coefficients. I. n-alkanes,» *Correlation and prediction of dense fluid transport coefficients. I. n-alkanes*, vol. 13, p. 269–281, 1992.
- [37] Bleazard, J. G., & Teja, A. S, «Extension of the Rough Hard-Sphere Theory for Transport Properties to Polar Liquids,» *Industrial & Engineering Chemistry Research*, vol. 35, n° 17, p. 2453–2459, 1996.
- [38] Ciotta, F., Trusler, J. M., & Vesovic, V, «Extended hard-sphere model for the viscosity of dense fluids,» *Fluid Phase Equilibria*, vol. 363, pp. 239-247, 2014.
- [39] V. V. Nicolas Riesco, «Extended hard-sphere model for predicting the viscosity of long-chain n-alkanes,» *Fluid Phase Equilibria*, vol. 425, pp. 385-392, 2016.

- [40] de Wijn, A. S., Riesco, N., Jackson, G., Martin Trusler, J. P., & Vesovic, V., «Viscosity of liquid mixtures: The Vesovic-Wakeham method for chain molecules,» *The Journal of Chemical Physics*, vol. 136, n° 17, 2012.
- [41] Assael, M. J., Dymond, J. H., Papadaki, M., & Patterson, P. M., «Correlation and prediction of dense fluid transport coefficients. I. n-alkanes,» *International Journal of Thermophysics*, vol. 13, p. 269–281, 1992.
- [42] Gaciño, F. M., Comuñas, M. J., Fernández, J., Mylona, S. K., & Assael, M. J., «Correlation and Prediction of Dense Fluid Transport Coefficients. IX. Ionic Liquids,» *International Journal of Thermophysics*, vol. 35, p. 812–829, 2014.
- [43] Hossain, M. Z., & Teja, A. S., «Correlation and Prediction of the Transport Properties of Ionic Liquids,» *International Journal of Thermophysics*, vol. 37, 2016.
- [44] Hosseini, S. M., Alavianmehr, M. M., & Moghadasi, J., «Transport properties of pure and mixture of ionic liquids from new rough hard-sphere-based model,» *Fluid Phase Equilibra*, vol. 429, pp. 266–274, 2016.
- [45] S. M. Hosseini, «Erratum to “Transport properties of pure and mixture of ionic liquids from new rough hard-sphere-based model,» *Fluid Phase Equilibra*, vol. 429, p. 266–274, 2016.
- [46] Akbari, F., Alavianmehr, M. M., Behjatmanesh Ardakani, R., & Mohammad-Aghaie, D., «Thermophysical properties of ionic liquids and their mixtures from a new equation of state,» *Ionics*, vol. 24, p. 1357–1369, 2018.
- [47] H. T. Davis, S. A. Rice et J. V. Sengers, «On the Kinetic Theory of Dense Fluids. IX. The Fluid of Rigid Spheres with a Square-Well Attraction,» *The Journal of Chemical Physics*, vol. 35, n° 16, p. 2210–2233, 1961.
- [48] I. L. McLaughlin et H. T. Davis, «Kinetic Theory of Dense Fluid Mixtures. I. Square-Well Model,» *The Journal of Chemical Physics*, vol. 45, n° 16, p. 2020–2031, 1966.
- [49] Du, L. G., & Guo, T. M., «A semi-theoretical viscosity model for non-polar liquids,» *The Chemical Engineering Journal*, vol. 47, n° 13, pp. 163–167, 1991.
- [50] Monnery, W. D., Svrcek, W. Y., & Mehrotra, A. K., «Viscosity: A critical review of practical predictive and correlative methods,» *The Canadian Journal of Chemical Engineering*, vol. 73, n° 11, pp. 3–40, 1995.
- [51] Galliero, G., Nieto-Draghi, C., Boned, C., Avalos, J. B., Mackie, A. D., Baylaucq, A., & Montel, F., «Molecular Dynamics Simulation of Acid Gas Mixtures: A Comparison between Several Approximations,» *Industrial & Engineering Chemistry Research*, vol. 46, n° 115, p. 5238–5244, 2007.
- [52] Zerón, I. M., Cueto-Mora, M., & Blas, F. J., «Transport properties of the square-well fluid from molecular dynamics simulation,» *Molecular Physics*, vol. 122, p. 19–20, 2024.
- [53] S. Handy, *Ionic Liquids - Current State of the Art*, Germany: first ed., IntechOpen, 2015.

- [54] Nieto-Draghi, C., Fayet, G., Creton, B., Rozanska, X., Rotureau, P., de Hemptinne, J. C., ... & Adamo, C; «A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes,» *Chemical Reviews*, vol. 115, n° %124, p. 13093–13164, 2015.
- [55] Katritzky, A. R., Chen, K., Wang, Y., Karelson, M., Lucic, B., Trinajstić, N., ... & Schüürmann, G, «Prediction of liquid viscosity for organic compounds by a quantitative structure–property relationship,» *Journal of Physical Organic Chemistry*, vol. 13, n° %11, pp. 80-86, 2000.
- [56] Bouarab, A. F., Harvey, J. P., & Robelin, C, «Viscosity models for ionic liquids and their mixtures,» *Physical Chemistry Chemical Physics*, n° %12, 2021.
- [57] Tochigi, K., & Yamamoto, H, «Estimation of Ionic Conductivity and Viscosity of Ionic Liquids Using a QSPR Model,» *The Journal of Physical Chemistry C*, vol. 111, n° %143, 2007.
- [58] Barycki, M., Sosnowska, A., Gajewicz, A., Bobrowski, M., Wileńska, D., Skurski, P., ... & Puzyn, T., «Temperature-dependent structure-property modeling of viscosity for ionic liquids,» *Fluid Phase Equilibria*, vol. 427, pp. 9-17, 2016.
- [59] Bini, R., Malvaldi, M., Pitner, W. R., & Chiappe, C, «QSPR correlation for conductivities and viscosities of low-temperature melting ionic liquids,» *Journal of Physical Organic Chemistry*, vol. 21, pp. 7-8, 2008.
- [60] Yu, G., Wen, L., Zhao, D., Asumana, C., & Chen, X, «QSPR study on the viscosity of bis(trifluoromethylsulfonyl)imide-based ionic liquids,» *Journal of Molecular Liquids*, vol. 184, pp. 51-59, 2013.
- [61] Nieto-Draghi, C., Fayet, G., Creton, B., Rozanska, X., Rotureau, P., de Hemptinne, J. C., ... & Adamo, C, «A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes,» *Chemical Reviews*, vol. 115, n° %124, p. 13093–13164, 2015.
- [62] 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*, vol. 80, pp. 326-333, 2012.
- [63] Zhang, S., Jia, Q., Yan, F., Xia, S., & Wang, Q, «Evaluating the properties of ionic liquid at variable temperatures and pressures by quantitative structure–property relationship (QSPR),» *Chemical Engineering Science*, vol. 231, p. 116326, 2021.
- [64] Ding, Y., Chen, M., Guo, C., Zhang, P., & Wang, J, «Molecular fingerprint-based machine learning assisted QSAR model development for prediction of ionic liquid properties,» *Journal of Molecular Liquids*, vol. 326, p. 115212, 2021.
- [65] Wadhwa, P., & Mittal, A, *Quantitative Structure-Property Relationship (QSPR) Modeling Applications in Formulation Development*, Singapore: Springer, Singapore, 2022.

- [66] Argoub, K., Benkouider, A. M., Yahiaoui, A., & Bagui, F, «Estimation and uncertainty analysis of standard enthalpy of formation in the liquid state by third-order-group-contribution method,» *Fluid Phase Equilibria*, vol. 520, p. 112644–112644, 2020.
- [67] Serat, F. Z., Benkouider, A. M., Yahiaoui, A., & Bagui, F, «Nonlinear group contribution model for the prediction of flash points using normal boiling points,» *Fluid Phase Equilibria*, vol. 449, p. 52–59, 2017.
- [68] Guella, S., Argoub, K., Benkouider, A. M., Yahiaoui, A., Kessas, R., & Bagui, F, «Artificial Neural Network-Group Contribution Method for Predicting Standard Enthalpy of Formation in the Solid State: C–H, C–H–O, C–H–N, and C–H–N–O Compounds,» *International journal of thermophysics*, vol. 36, n° %110-11, p. 2820–2832, 2015.
- [69] Benkouider, A. M., Kessas, R., Guella, S., Yahiaoui, A., & Bagui, F, «Estimation of the enthalpy of vaporization of organic components as a function of temperature using a new group contribution method,» *Journal of molecular liquids*, vol. 194, p. 48–56, 2014.
- [70] Argoub, K., Benkouider, A. M., Yahiaoui, A., Kessas, R., Guella, S., & Bagui, F, «Prediction of standard enthalpy of formation in the solid state by a third-order group contribution method,» *Fluid phase equilibria*, vol. 380, p. 121–127, 2014.
- [71] K. Paduszyński, «Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity,» *Industrial & engineering chemistry research*, vol. 58, n° %136, p. 17049–17066, 2019.
- [72] Chen, Y., Kontogeorgis, G. M., & Woodley, J. M, «Group Contribution Based Estimation Method for Properties of Ionic Liquids,» *Industrial & Engineering Chemistry Research*, vol. 58, n° %110, p. 4277–4292, 2019.
- [73] Baghban, A., Kardani, M. N., & Habibzadeh, S, «Prediction viscosity of ionic liquids using a hybrid LSSVM and group contribution method,» *Journal of Molecular Liquids*, vol. 236, p. 452–464, 2017.
- [74] Lazzús, J. A., & Pulgar-Villarroel, G, «A group contribution method to estimate the viscosity of ionic liquids at different temperatures,» *Journal of molecular liquids*, vol. 209, pp. 161–168,, 2015.
- [75] , «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*, vol. 54, n° %15, p. 1311–1324.
- [76] K. Paduszyński, «Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity,» *Industrial & engineering chemistry research*, vol. 58, n° %136, p. 17049–17066, 2019.
- [77] Cao, X., Gong, M., Tula, A., Chen, X., Gani, R., & Venkatasubramanian, V, «An Improved Machine Learning Model for Pure Component Property Estimation,,» *Engineering*, vol. 39, pp. 61-73, 2024.

- [78] Varmuza, K., Dehmer, M., & Bonchev, D, Statistical Modelling of Molecular Descriptors in QSAR/QSPR, Wiley-VCH Verlag GmbH & Co. KGaA, 2012.
- [79] Khan, A. U, «Descriptors and their selection methods in QSAR analysis: paradigm for drug design,,» *Drug Discovery Today,,* vol. 21, n° 18, pp. 1291-1302, 2016.
- [80] Wang, L., Ding, J., Pan, L., Cao, D., Jiang, H., & Ding, X, «Quantum chemical descriptors in quantitative structure–activity relationship models and their applications,,» *Chemometrics and Intelligent Laboratory Systems,,* vol. 217, p. 104384, 2021.
- [81] Rogers, D., & Hahn, M., «Extended-Connectivity Fingerprints,,» *Journal of Chemical Information and Modeling*, vol. 50, n° 15, p. 742–754, 2010.
- [82] Carhart, Raymond E., SMITH, Dennis H., et VENKATARAGHAVAN, RENGACHARI, «Atom pairs as molecular features in structure-activity studies: definition and applications,,» *Journal of Chemical Information and Computer Sciences*, vol. 25, n° 12, p. 64–73, 1985.
- [83] Pérez-Nueno, V. I., Rabal, O., Borrell, J. I., & Teixidó, J, «APIF: A New Interaction Fingerprint Based on Atom Pairs and Its Application to Virtual Screening,,» *Journal of Chemical Information and Modeling*, vol. 49, n° 15, p. 1245–1260, 2009.
- [84] R. Davronov et S. Kushmuratov, «Comparative analysis of QSAR feature selection methods,,» *AIP Conference Proceedings*, vol. 3004, n° 11, 2024.
- [85] Zebida, M. A., Argoub, K., Benkouider, A. M., Yahiaoui, A., Toubal, K., & Hachemaoui, A, «Machine learning coupled with group contribution for predicting the electrical conductivity of ionic liquids with experimental accuracy,,» *Fluid Phase Equilibria,,* vol. 579, pp. 0378-3812, 2024.
- [86] Seber, G. A, Nonlinear Regression, 1989 John Wiley & Sons, Inc, 1989.
- [87] Daubechies, I., DeVore, R., Fornasier, M., & Güntürk, C. S., «Iteratively reweighted least squares minimization for sparse recovery,,» *Pure and Applied Mathematics*, vol. 63, n° 11, pp. 1-38, 2010.
- [88] Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P, «Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling,,» *Journal of Chemical Information and Computer Sciences*, vol. 43, n° 16, 2003.
- [89] Zhang, X., Shen, H., Huang, T., Wu, Y., Guo, B., Liu, Z., ... & Ou, G, «Improved random forest algorithms for increasing the accuracy of forest aboveground biomass estimation using Sentinel-2 imagery,,» *Ecological Indicators*, vol. 159, 2024.

- [90] Wang, C., Wei, X., Jin, X., Li, J., & He, M., «Developing a two-grade model for the thermal conductivity of ionic liquids and their mixtures,» *Chemical Engineering Science*, vol. 290, p. 119881, 2024.
- [91] Fröba, A. P., Rausch, M. H., Krzeminski, K., Assenbaum, D., Wasserscheid, P., & Leipertz, A., « Thermal conductivity of ionic liquids: Measurement and prediction,» *Int. J. Thermophys*, vol. 31, p. 2059–2077, 2010.
- [92] Zhao, A. Z., & Garay, J. E., «High temperature liquid thermal conductivity: A review of measurement techniques, theoretical understanding, and energy applications,» *Prog. Mater. Sci*, vol. 139, p. 101180, 2023.
- [93] Haghbakhsh, R., & Raeissi, S., «A novel correlative approach for ionic liquid thermal conductivities,» *J. Mol. Liq*, vol. 236, p. 214–219, 2017.
- [94] Soares, L. H., Guirardello, R., & Rolemberg, M. P., « A simple group contribution model to predict thermal conductivity of pure ionic liquids,» *Chem. Eng. Trans*, vol. 74, p. 1195–1200, 2019.
- [95] Almeida, R. M., Lourenço, M. J. V., & de Castro, C. N., «The thermal conductivity of ionic liquids. Experiment and molecular interpretation,» *J. Mol. Liq*, vol. 397, 2024.
- [96] Yang, H., Gallagher, R. C., Phan, A. T., Chartrand, P., & Gheribi, A. E., « A predictive approach for the compositional and temperature representation of thermal conductivity in multicomponent molten salt systems for advanced energy applications,» *Mater. Today Energy*, vol. 38, p. 101441, 2023.
- [97] Wan, R., Li, M., Song, F., Xiao, Y., Zeng, F., Peng, C., & Liu, H., « Predicting the Thermal Conductivity of Ionic Liquids Using a Quantitative Structure–Property Relationship,» *Ind. Eng. Chem. Res*, vol. 61, p. 12032–12039, 2022.
- [98] Kazakov, A. M. J. W., Magee, J. W., Chirico, R. D., Paulechka, E., Diky, V., Muzny, C. D., ... & Frenkel, M., «NIST Standard Reference Database 147: NIST Ionic Liquids Database-(ILThermo),» National Institute of Standards and Technology. Gaithersburg MD, v2.0, 2021. [En ligne]. Available: <http://ilthermo.boulder.nist.gov>.
- [99] Snoek, J., Larochelle, H., & Adams, R. P., *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, Nevada: Curran Associates Inc, 2012.
- [100] Gelbart, M. A., Snoek, J., & Adams, R. P., *Proceeding of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, Quebec City, Quebec, Canada: AUAI Press, 2014.
- [101] Han, C., Yu, G., Wen, L., Zhao, D., Asumana, C., & Chen, X
, «Data and QSPR study for viscosity of imidazolium -based ionic liquids,» *Fluid Phase Equilibra*, vol. 300, p. 95, 2011.
- [102] Valderrama, J. O., Cardona, L. F., & Rojas, R. E., « Correlation and prediction of ionic liquids viscosity using Valderrama-Patel-Teja cubic equation of state and the geometric

- similitude concept, Part I: Pure ionic liquids,» *Fluid Phase Equilibra*, vol. 497, p. 164, 2019.
- [103] He, M., Zhu, C., & Liu, X, « Estimating the viscosity of ionic liquid at high pressure using Eyring's absolute rate theory,» *Fluid Phase Equilib*, vol. 458, p. 170, 2018.
- [104] Darabi, L., & Zare, M., «High correlate simple equations for temperature and pressure dependence of the viscosity of ionic liquids,» *Chem. Phys*, vol. 539, 2020.
- [105] Yan, F., He, W., Jia, Q., Wang, Q., Xia, S., & Ma, P., « Prediction of ionic liquids viscosity at variable temperatures and pressures,» *Chem. Eng. Sci.*, vol. 184, p. 134, 2018.
- [106] Kang, X., Zhao, Z., Qian, J., & Muhammad Afzal, R, «Predicting the Viscosity of Ionic Liquids by the ELM Intelligence Algorithm,» *Ind. Eng. Chem. Res*, vol. 56, p. 11344, 2017.
- [107] Sun, Y., Chen, M., Zhao, Y., Zhu, Z., Xing, H., Zhang, P., ... & Ding, Y, «Machine learning assisted QSPR model for prediction of ionic liquid's refractive index and viscosity: The effect of representations of ionic liquid and ensemble model developement,» *J. Mol. Liq*, vol. 333, 2021.
- [108] T. Mezger, *The Rheology Handbook*, Vincentz Network: 3rd revised ed. Hanove, 2011.
- [109] M. Rao, « Rheology of liquid foods,» *Journal Tenure studies*, p. pp. 135 – 168., 1997.
- [110] T. C. Scott et R. Wham, «Surface area generation and droplet size control in solvent extraction systems utilizing high intensity electric fields,» *U.S. Patent 4767515.12*, 1988.
- [111] Ptasinski, K. J., & Kerkhof, P. J. A. M, «Electric field driven separations: Phenomena and applications.,» *Sep. Sci. Technol.*, 1992.
- [112] H. Xie, L. Li, Y. Sun, Y. Wang, S. Gao, Y. Tian, X. Ma, C. Guo, F. Bo et L. Zhang, « An Available Strategy for Nasal Brain Transport of Nanocomposite Based on PAMAM Dendrimers via In Situ Gel,» *Nanomaterials* , pp. 9, 147, 2019.
- [113] B. Vigani, S. Rossi, G. Milanesi, M. Bonferoni, G. Sandri, G. Bruni et F. Ferrari, « Electrospun Alginate Fibers: Mixing of Two Different Poly (ethylene oxide) Grades to Improve Fiber Functional Properties.,» *Nanomaterials* , pp. 8, 971, 2018.
- [114] P. Mishra, S. Mukherjee, S. Nayak et A. Panda, «A brief review on viscosity of nanofluids,» *Int. Nano Lett*, pp. 4, 109–120., 2014.
- [115] A. Einstein, «Eineneuebestimmung der molekuldimensionen.,» *Ann. Phys*, pp. 324, 289–306., 1906.
- [116] T. Phuoc et M. Massoudi, « Experimental observations of the effects of shear rates and particle concentration on the viscosity of Fe₂O₃–deionized water nanofluids.,» *Int. J. Therm. Sci*, pp. 48, 1294–1301, 2009.
- [117] Beggs, H. D., & Robinson, J. R., « Estimating the Viscosity of Crude Oil Systems,» *J. Petrol. Technol*, pp. 9, 1140-1141, 1995.

- [118] F. F. Petrosky G.E., «Viscosity Correlations for Gulf of Mexico Crudes Oils,,» 1995.
- [119] Labedi, R, « Improved Correlations for Predicting the Viscosity of Light Crudes,» *J. Petrol. Sci. Eng.*, pp. 8, 221-234, 1992.
- [120] Khan, S. A., Al-Marhoun, M. A., Duffuaa, S. O., & Abu-Khamsin, S. A, «Viscosity Correlations for Saudi Arabian Crude Oils,» p. 15 720., 1987.
- [121] V. d. W. J.D, Leyden, 1873.
- [122] Lohrenz, J., Bray, B. G., & Clark, C. R, « Calculating Viscosities of Reservoir Fluids from their Composition,» *J. Petrol. Technol*, pp. 1171-1176, 1964.
- [123] K. H. Little J.E., «A Correlation of the Viscosity of Hydrocarbon Systems with Pressure, Temperature and Composition,,» *Soc. Petrol. Eng. J. AIME,,* pp. 243, 157-162, 1968.
- [124] Guo, X. Q., Wang, L. S., Rong, S. X., & Guo, T. M, « Viscosity Model Based on Equations of State for Hydrocarbon Liquids and Gases,» *Fluid Phase Equilib*, pp. 139, 405-421, 1997.
- [125] T. A. Patel N.C., « A New Cubic Equation of State for Fluids and Fluid Mixtures,» *Chem. Eng. Sci*, pp. 37, 463-473., 1982.
- [126] Ungerer, P., Thermodynamics: Applications in Chemical Engineering and the Petroleum Industry,, paris: Ed. Technip, Paris, 2003.
- [127] 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*, vol. 80, p. 326–333, 2015.
- [128] Zhang, S., Jia, Q., Yan, F., Xia, S., & Wang, Q., «Evaluating the properties of ionic liquid at variable temperatures and pressures by quantitative structure–property relationship (QSPR),» *Chem. Eng. Sci*, vol. 231, 2021.

CHAPITRE IV: VISCOSITY PREDICTION OF NON-IONIC ORGANIC COMPOUNDS

IV.1 INTRODUCTION

Viscosity is a critical thermophysical property in the design and optimization of chemical processes, influencing mass/heat transfer, fluid flow, and equipment selection. For non-ionic organic compounds (widely used in pharmaceutical, petrochemical, and materials industries), accurate viscosity prediction remains challenging due to molecular diversity and nonlinear temperature dependence. While experimental methods are reliable, they are costly and impractical for rapid screening of novel compounds.

This chapter presents a multi-algorithm modeling framework to predict liquid and gas-phase viscosity of non-ionic organics across broad thermodynamic ranges (74–1120 K). By combining complementary molecular representations, group contributions (GC) and RDKit-calculated QSPR descriptors, with five state-of-the-art methods (XGBoost, Random Forest, SVM, Deep Neural Network, and a semi-empirical model), we establish a robust predictive architecture. Our rigorously curated database, derived from DIPPR 801, encompasses 9,715 experimental points for 668 molecules spanning 20 chemical families (alcohols, polyols, halocarbons, etc.), offering unprecedented structural diversity.

Systematic performance analysis, via statistical metrics (MAE, AARE, RMSE, R^2), chemical family benchmarking, and error distribution profiling, reveals the synergistic potential of global descriptors (QSPR) and nonlinear algorithms (XGBoost, DNN) to

capture molecular complexity. These results enable high-fidelity predictive tools for process engineering.

IV.2 DATABASE: DATA COLLECTION AND REVISION

When modeling the viscosity of non-ionic organic compounds, collecting and revising data are key steps to ensure the models are strong and trustworthy. This section explains the main features of the database used for this purpose, focusing on the quality and importance of the data. The viscosity values in this study come only from experiments and are taken from the DIPPR 801 database (Design Institute for Physical Properties), which is widely respected around the world for its thorough and accurate physical-chemical data.

At first, the DIPPR 801 database had 27,915 measurement points, including 2,729 in the gas phase and 25,186 in the liquid phase. It covered a wide range of pure organic chemical families. These starting data included different sub-families like alkanes, ketones, aliphatic alcohols, halogenated compounds, and complex polyfunctional structures, as shown in Table . However, to keep the data consistent and meet the specific needs of the modeling, a strict preprocessing method was applied. This method, described in Chapter II, Section II.2.1, had seven steps: removing duplicates, checking how temperature and viscosity relate, and filtering out odd values using statistical and thermodynamic rules.

After this revision, and to keep improving, the simulated data originally in the DIPPR 801 database was removed, leaving only the experimental measurements. The final database now has 9,715 valid data points from 668 different molecules, as summarized in

Table 7. This carefully prepared database stands out for its wide variety of structures, with temperatures ranging from 74 K to 1120 K and viscosities from 4.28×10^{-7} mPa·s to 6710 mPa·s. The extreme values come mainly from polyols (highest viscosity) and polyfunctional compounds (lowest viscosity), showing the richness and complexity of the chemical profiles included.

To make tracking and deep analysis easier, the database was carefully organized. Each molecule's entry groups data sets by their original bibliographic source and always includes this information: physical state, chemical family and sub-family, systematic name, chemical formula, CAS number, SMILES notation, temperature, viscosity value,

and reference. This setup allows for detailed analyses that combine experimental factors and molecular descriptions, creating a solid base for future modeling work.

Table 6: Number of viscosity data by sub-family of organic compounds

Chemical Subfamilies	N°	Chemical Subfamilies	N°	Chemical Subfamilies	N°
1-Alkenes	550	Isocyanates/Diisocyanates	134	Other Ethers/Diethers	52
Aldehydes	890	Ketones	941	Other Hydrocarbon Rings	26
Aliphatic Ethers	773	N-Alcohols	578	Other Polyfunctional C, H, O	368
Anhydrides	180	N-Aliphatic Acids	167	Peroxides	168
Aromatic Alcohols	315	N-Aliphatic Primary Amines	256	Polyfunctional Acids	122
Aromatic Amines	547	N-Alkanes	798	Polyfunctional Amides/Amines	407
Aromatic Chlorides	367	N-Alkylbenzenes	940	Polyfunctional C, H, N, Halide, (O)	97
Aromatic Esters	614	Naphthalenes	328	Polyfunctional C, H, O, Halide	829
Aldehydes	56	Nitriles	524	Polyfunctional C, H, O, N	533
Aliphatic Ethers	68	Naphthalenes	21	Polyfunctional C, H, O, S	242
Aromatic Alcohols	10	Nitriles	67	Polyfunctional Esters	429
Aromatic Esters	87	Other Aliphatic Acids	387	Polyfunctional Nitriles	102
C, H, Br Compounds	513	Other Aliphatic Alcohols	1167	Polyols	440
C, H, F Compounds	517	Other Alkanes	571	Peroxides	41
C, H, F Compounds	962	Other Alkylbenzenes	987	Polyfunctional Amides/Amines	20
C, H, Multihalogen Compounds	580	Other Amines, Imines	582	Polyfunctional C, H, O, Halide	42
C, H, Multihalogen Compounds	591	Other Ethers/Diethers	575	Polyfunctional C, H, O, N	36
C, H, No2 Compounds	245	Other Hydrocarbon Rings	113	Polyfunctional Esters	168
C1/C2 Aliphatic Chlorides	753	Other Monoaromatics	470	Polyols	107
Cycloaliphatic Alcohols	170	Other Polyfunctional C, H, O	870	Silanes/Siloxanes	894
Dialkenes	577	Other Polyfunctional Organics	10	Sulfides/Thiophenes	1431
Diphenyl/Polyaromatics	82	Other Saturated Aliphatic Esters	547	Silanes/Siloxanes	295
Epoxides	307	Other Aliphatic Acids	21	N-Aliphatic Acids	42
Epoxides	97	Other Aliphatic Alcohols	53		
Glycerides	28	Other Amines, Imines	38		

Table 7: Overview of experimental viscosity data for non-ionic compounds classified by chemical family

Chemical families	Data points	Molecules	T min (K)	T max (K)	η min (mPa.s)	η max (mPa.s)
Acides & Anhydrides	414	24	199.85	1000	3.53E-06	0.0602
Alcohols	1163	51	155.95	1000	2.91E-06	44.2
Aldehydes	81	16	149.78	1000	3.25E-06	0.0447
Alkanes	818	25	85.47	1000	2.56E-06	1.94
Alkenes	350	21	88	1000	2.87E-06	0.0153
Autres	1292	100	116	1000	3.73E-06	3.96
Autres Hydrocarbures	20	2	193	1000	3.95E-06	0.00294
Benzenes & Aryles	1002	32	159.7	1000	3.33E-06	95.1
Epoxides	160	10	130	1000	2.95E-06	0.0129
Esters	352	36	161.5	1000	2.42E-06	64.5
Ethers	317	37	127.93	1000	3.34E-06	21.3
Halogénés	1904	69	74	1000	3.01E-06	0.143
Ketones	260	22	167.15	1000	1.66E-06	0.035
Naphthalenes & Polyaryles	138	11	196.97	1000	3.05E-06	0.0443
Nitriles	164	15	161.3	1000	2.60E-06	2.31
Polyfonctionnels	555	65	173	1000	4.28E-07	100
Polyols & Glycerides	250	16	182.15	1120	4.13E-06	6710
Silanes/Siloxanes	323	44	116.34	1000	2.79E-06	0.0325
Sulfides/Thiophenes	152	20	156.12	1000	3.66E-06	0.078
Total	9 715	668	74	1120	4.28E-07	6710

IV.3 METHODOLOGICAL STRATEGY FOR MODELLING

This study uses five different modeling approaches to predict the viscosity of non-ionic organic compounds (liquids and gases): Extreme Gradient Boosting (XGBoost), Random Forest (RF), Support Vector Machine (SVM), Deep Neural Network (DNN), and a conventional semi-empirical model (see Section II.3, “Models and Regression,” for details). Each method systematically employs two complementary structural representations: group contributions (GC) and QSPR descriptors calculated with RDKit, augmented by the temperature (T).

For the GC approach, each molecule was broken down into 191 groups, organized as first-, second- and third-order contributions. For the QSPR approach, a vector of 178

descriptors was computed per molecule, capturing size, polarity, charge distribution and molecular functional groups.

In the conventional model, after testing several laws (Arrhenius, Vogel–Fulcher–Tammann), we settled on an equation that balances correlation strength with extrapolation ability, using a sigmoid-style form:

$$\ln(\eta) = \sum_{i=1}^3 b_i \frac{\sum_{j=1}^K N_j a_{i,j} + \alpha_i T}{1 + |\sum_{j=1}^K N_j a_{i,j} + \alpha_i T|} + c \quad (29)$$

This equation, b_i and c are adjustable parameters, A_i is the contribution of all groups in i^{th} term of this equation, including both first- and second-order groups. In this equation, N_j denote the number of occurrences of j^{th} individual group, $a_{i,j}$ is the contribution of j^{th} group in i^{th} term of this equation. K is the total number of functional groups and T is temperature in units of kelvin (K). α_i are adjustable parameters in i^{th} term of this equation.

Note that this conventional model applies only to the group-contribution approach, because the molecular fragmentation lets us assign semi-empirical coefficients directly to discrete structural units (functional groups and their interactions). In contrast, the QSPR approach relies on a dense global descriptor vector and cannot be interpreted in a localized physico-chemical way.

The QSPR and GC models were developed in MATLAB 2023b using the Statistics and Machine Learning Toolbox under an academic license. To ensure a standardized workflow, most toolbox hyperparameters remained at their default values, and only those demonstrating a significant impact on model performance were tuned. A Bayesian optimization strategy was implemented to iteratively identify the best configurations. Model robustness and accuracy were assessed through 25 repeats of 5-fold cross-validation, with mean squared error as the primary minimization criterion.

The search for search spaces and the final selection of hyperparameters, determined by Bayesian optimisation, are summarised as follows (see Section x). For XGBoost, the search covered minimum MinLeafSize¹ from 1 to 100, number of NumLearningCycles²

¹ MinLeafSize : Taille minimale des feuilles terminales.

² NumLearningCycles : Nombre d'arbres (estimateurs) dans le modèle.

(trees) from 5 to 1000, and LearnRate³ between 1×10^{-4} and 1. Bayesian optimization found the optimal settings to be a MinLeafSize of 1, NumLearningCycles = 400, and a LearnRate of 0.15. In the Random Forest models, the search space included MinLeafSize from 1 to 100 and 10 to 1000 for NumLearningCycles; the best configuration used a MinLeafSize of 1 and NumLearningCycles = 581. Tuning of the SVM focused on the regularization parameter C (1–400), the epsilon-tube width ϵ ($0,0001$ –1) and the kernel scale γ (0,1–100). The optimization converged to $C=75$, $\epsilon=1,8 \times 10^{-2}$ et $\gamma=9,98$. Finally, for the deep neural network, the search ranged over 1 to 5 hidden layers, 1 to 60 neurons per layer, activation functions (ReLU, sigmoid, tanh), and regularization strength Lambda (0–10). The best architecture consisted of two hidden layers with 40 and 10 neurons respectively, a tanh activation, and $\text{Lambda} = 10^{-10}$.

IV.4 RESULTS AND DISCUSSION

In this section, we present and analyze the predictive performance of the models developed to estimate the viscosity of non-ionic organic compounds in the gas phase and liquid phase, based on the statistical metrics described in Section X. The results are organized according to two distinct modeling approaches: the Group Contribution (GC) method, where molecular fragments serve as direct input variables, and the QSPR approach, which utilizes calculated molecular descriptors. Five modeling techniques were implemented for this purpose: XGBoost, Random Forest (RF), SVM, DNN Network, and a conventional semi-empirical model. The performance of these models was rigorously evaluated using a 5-fold cross-validation procedure repeated 25 times, ensuring the stability and reliability of the results. The main statistical metrics; Mean Absolute Error (MAE), Absolute Average Relative Error (AARE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2), are reported in Table for the training and test sets. Furthermore, the performance on the test set is visually summarized in Figure and Figure via scatter plots comparing experimental and predicted values and Gaussian distribution plots of prediction errors for both modeling strategies.

Table 8 provides a detailed comparison of the model performance statistics for the GC and QSPR approaches. For the GC method, the XGBoost model demonstrated superior predictive accuracy, with a MAE of 0.290 on the training set and 0.426 on the test set,

³ LearnRate : Taux d'apprentissage.

coupled with AARE values of 4.30% and 7.43%, respectively. The RMSE values were 0.547 (training) and 0.802 (test), while the R^2 values reached 0.9385 (training) and 0.8723 (test), indicating excellent agreement between the predicted and experimental values. The SVM model also exhibited good performance under the GC approach, with a MAE of 0.242 (training) and 0.304 (test), AARE values of 3.25% and 5.14%, and R^2 values of 0.8836 (training) and 0.8649 (test). In contrast, the conventional semi-empirical model showed the weakest performance, with a MAE of 0.628 (training) and 0.651 (test), AARE values of 13.49% and 19.48%, and R^2 values of 0.7888 (training) and 0.7841 (test), highlighting its limitations compared to the machine learning models. For the QSPR approach, the XGBoost model again stood out, with a MAE of 0.283 (training) and 0.389 (test), AARE values of 3.85% and 5.50%, and R^2 values of 0.9378 (training) and 0.8723 (test). The DNN model also showed robust performance under the QSPR framework, with a MAE of 0.273 (training) and 0.358 (test), AARE values of 3.25% and 4.51%, and R^2 values of 0.9180 (training) and 0.8638 (test). Overall, the QSPR approach generally offered slightly higher predictive accuracy than the GC method, particularly for the XGBoost and DNN models, as evidenced by the lower AARE values and higher R^2 values on the test sets.

Table 8: Model performance statistics on the training and test sets, expressed as averages from a 5-fold cross-validation, repeated 25 times according to random partitions.

	Models	EAM_{train}	EAM_{test}	$ERAM_{train}$ (%)	$ERAM_{test}$ (%)	$REQM_{train}$	$REQM_{test}$	R^2_{train}	R^2_{test}
GC	XGBoost	0.290	0.426	4.30	7.43	0.547	0.802	0.9385	0.8723
	RF	0.373	0.498	7.81	14.19	0.674	0.878	0.9103	0.8487
	SVM	0.242	0.304	3.25	5.14	0.761	0.830	0.8836	0.8649
	DNN	0.260	0.356	3.16	4.75	0.627	0.781	0.9187	0.8793
	Conventionnel	0.628	0.651	13.49	19.48	1.011	1.042	0.7888	0.7841
QSPR	XGBoost	0.283	0.389	3.85	5.50	0.549	0.769	0.9378	0.8723
	RF	0.298	0.391	4.78	6.01	0.615	0.799	0.9233	0.8626
	SVM	0.300	0.347	5.20	5.53	0.805	0.870	0.8712	0.8414
	DNN	0.273	0.358	3.25	4.51	0.629	0.796	0.9180	0.8638

Figure provides a visual representation of the predictive performance through scatter plots of experimental values versus predicted values for the training and test sets, categorized according to the GC (Figure a) and QSPR (Figure b) approaches. In Figure a (GC method), the scatter plots for each model—XGBoost, RF, SVM, and DNN—display the training set data as blue circles and the test set data as yellow circles, plotted against

a diagonal line representing perfect prediction. The XGBoost plot shows a tight clustering of points along the diagonal, with an R^2 of 0.92466, indicating excellent predictive accuracy and minimal deviation between experimental and predicted values. The RF model, with an R^2 of 0.89749, exhibits slightly greater dispersion, particularly for the test set, reflecting its higher AARE of 14.19% as reported in Table. The SVM ($R^2 = 0.87971$) and DNN ($R^2 = 0.91169$) models also show good agreement between experimental and predicted values, although the SVM plot reveals a slightly wider dispersion of test set points, consistent with its RMSE of 0.830 on this set. In Figure b (QSPR method), the scatter plots for the same models are presented, with the training set as green circles and the test set as pink circles. The XGBoost model again demonstrates the best performance, with an R^2 of 0.92502 and a highly linear distribution of points along the diagonal, corroborating its low test set AARE of 5.50%. The RF ($R^2 = 0.91153$) and DNN ($R^2 = 0.90736$) models also display strong predictive capabilities, although the RF plot indicates a slightly greater dispersion of test set points, consistent with its AARE of 6.01%. The SVM model, with an R^2 of 0.86843, shows the most notable dispersion among the QSPR models, particularly for the test set, which is consistent with its relatively higher RMSE of 0.870. Across both figures, the scatter plots confirm the superior performance of the XGBoost and DNN models under the GC and QSPR approaches, with the QSPR method generally offering a slight advantage in terms of predictive accuracy, as evidenced by the tighter clustering of points and higher R^2 values.

Figure presents the Gaussian distributions of prediction errors for the training and test sets, evaluating four machine learning models (XGBoost, Random Forest, SVM, and DNN) under two modelling approaches: Group Contributions (GC) and QSPR. In both approaches, the error distributions for the training set consistently exhibit higher peaks and reduced widths compared to those of the test set, reflecting the expected fitting to the training data. Under the GC approach, XGBoost demonstrates the most concentrated distributions with density peaks of 0.7 (training) and 0.6 (test) and an exceptional R^2 of 0.92466, while RF ($R^2=0.89749$), SVM ($R^2=0.87971$), and DNN ($R^2=0.91169$) show wider distributions (peaks $\approx 0.45/0.35$) suggesting increased predictive variance, particularly pronounced for RF. Under the QSPR approach, XGBoost maintains optimal performance ($R^2=0.92502$), RF significantly improves its R^2 (0.91513 vs. 0.89749 under GC) despite persistent variance, whereas SVM records the largest performance drop ($R^2=0.86543$), likely due to sensitivity to complex molecular descriptors. DNN retains

notable robustness ($R^2=0.90736$), albeit lower than its GC score. Comparatively, XGBoost and DNN emerge as the most performant and stable models, with errors strongly centered around zero under both approaches. The differential impact of QSPR descriptors, beneficial for RF but penalizing for SVM, highlights the importance of compatibility between model architecture and data nature. These results quantitatively corroborate the metrics reported in Table and confirm XGBoost's superior predictive capability for estimating the viscosity of ionic liquids.

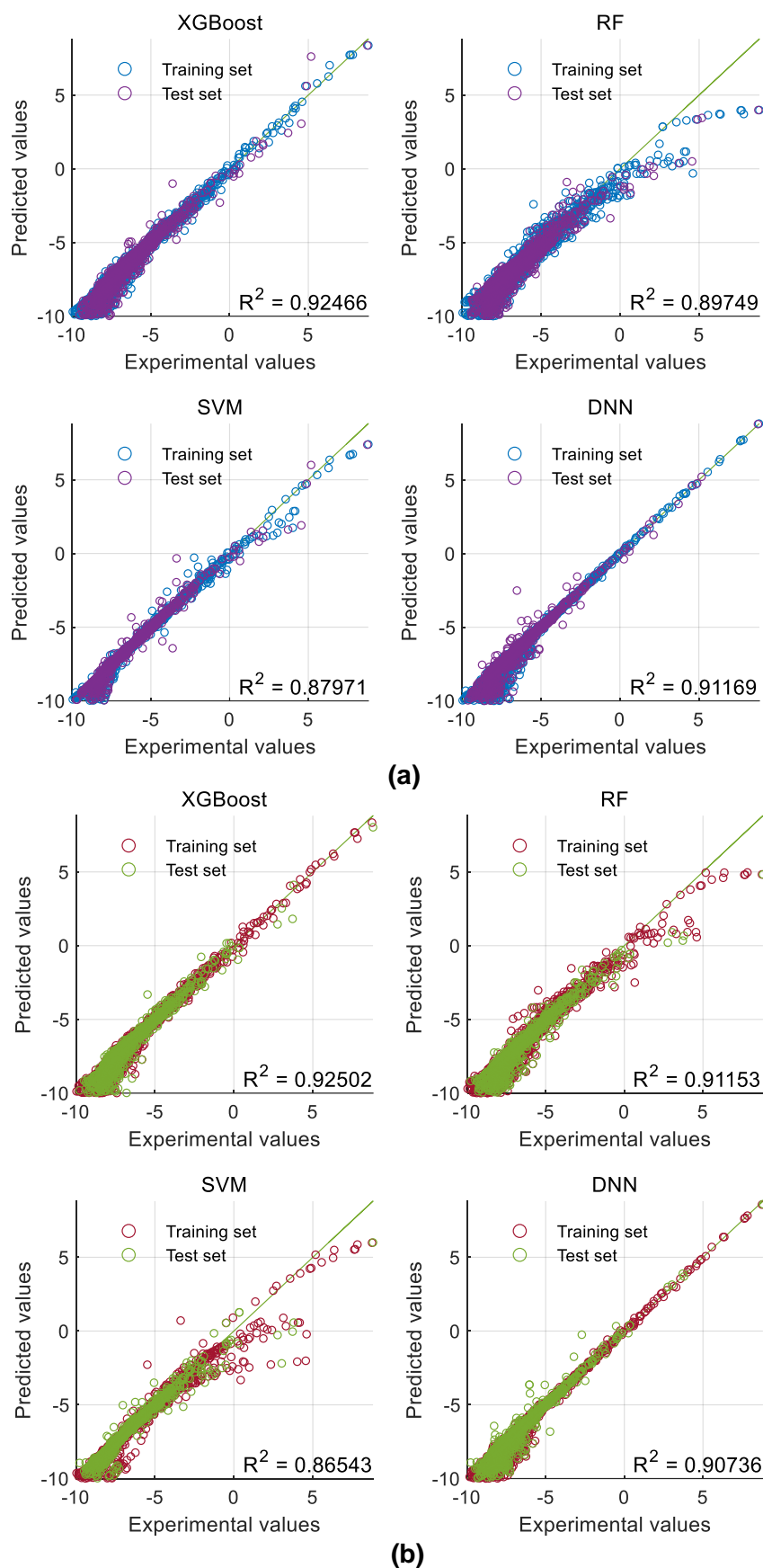


Figure 34: Scatter plots of predicted and experimental values in the training and test sets. (a) Contribution groups. (b) QSPR..

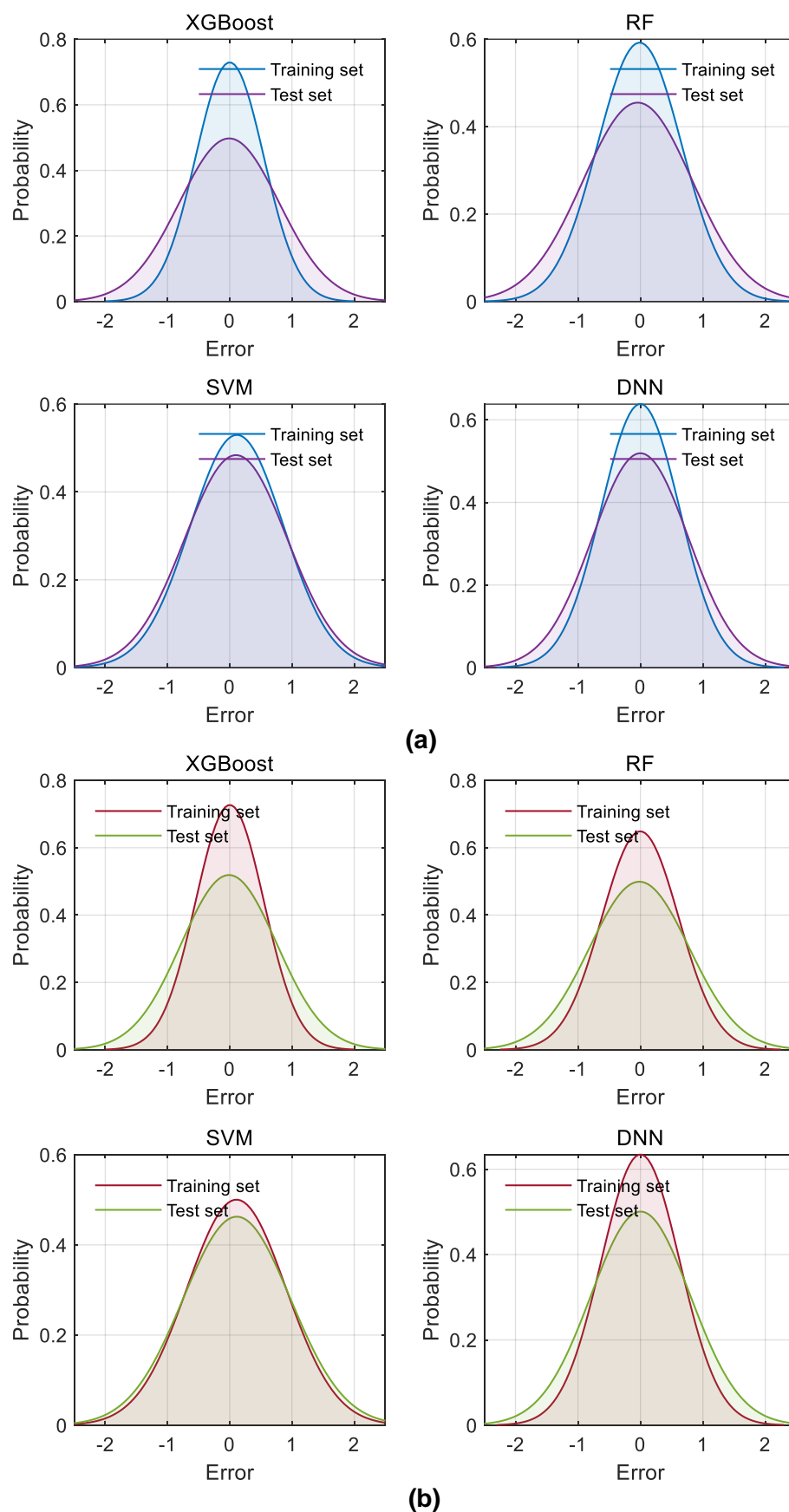


Figure 35: Gaussian distribution of prediction errors in the training set and the test set (a) Contribution groups. (b) QSPR.

IV.4.1 Relative errors by chemical families

For the GC models (Table), mean relative errors vary significantly across chemical families. The families with the highest errors are Polyols & glycerides (11.5% for XGBoost, peaking at 23.3% for RF), Polyfunctionals (8.8% for XGBoost, up to 32.2% for RF), and Alkanes (7.6% for XGBoost). Conversely, the most accurate families are Epoxides (1.8% for XGBoost), Naphthalenes & polyaryls (1.7%), Other hydrocarbons (2.1%), and Silanes/siloxanes (2.0%). The RF model consistently shows the highest errors, particularly for Polyfunctionals (32.2%) and Esters (18.3%). The DNN excels for several complex families like Polyfunctionals (1.8%), Esters (1.6%), and Polyols & glycerides (2.1%), often outperforming XGBoost and SVM.

Table 9: Mean Relative Error of GC Models (XGBoost, RF, SVM, DNN) by Major Chemical Family

GrandFamily	DataPoints	XGBoost (%)	RF(%)	SVM (%)	DNN (%)
Acides & anhydrides	414	3.9	4.7	2.6	4.3
Alcohols	1163	6.7	16.1	5.1	4.1
Aldehydes	81	3.1	3.1	2.3	0.7
Alkanes	818	7.6	9.2	6.5	7.9
Alkenes	350	3.1	4.0	1.9	1.8
Autres	1292	3.4	4.2	2.3	3.0
Autres hydrocarbures	20	2.1	1.6	0.3	0.5
Benzenes & aryles	1002	3.4	4.2	2.3	2.6
Epoxides	160	1.8	2.9	0.6	0.7
Esters	352	7.0	18.3	6.9	1.6
Ethers	317	2.8	4.0	1.8	1.8
Halogénés	1904	5.0	5.9	3.9	4.9
Ketones	260	2.1	2.3	1.2	1.4
Naphthalenes & polyaryles	138	1.7	2.0	0.5	0.8
Nitriles	164	2.8	3.1	1.2	1.7
Polyfunctionnels	555	8.8	32.2	7.3	1.8
Polyols & glycerides	250	11.5	23.3	4.7	2.1
Silanes/siloxanes	323	2.0	3.2	0.5	1.6
Sulfides/thiophenes	152	1.9	2.8	1.1	2.2

For the QSPR models (Table), overall improvement is observed for several families problematic under GC. Errors for Alcohols decrease significantly (XGBoost drops from 6.7% to 5.0%, DNN from 4.1% to 2.8%). Polyfunctionals see drastic error reduction with XGBoost (from 8.8% to 3.3%) and DNN (from 1.8% to 1.6%). Esters also benefit from

QSPR (XGBoost decreases from 7.0% to 4.1%, DNN from 1.6% to 2.6%). However, Polyols & glycerides remain challenging to model (XGBoost at 10.5%, RF at 17.0%), though DNN maintains high accuracy (1.8%). Families already accurate under GC (Epoxides, Naphthalenes & polyaryls, Other hydrocarbons, Silanes/siloxanes) retain very low errors (<2.5%). The DNN model confirms its robustness, frequently displaying the lowest errors (notably for Alcohols, Polyfunctionals, Polyols & glycerides, and Esters), while SVM struggles with certain families like Esters (13.2%) and Polyols & glycerides (22.2%).

Table10: Mean Relative Error of QSPR Models (XGBoost, RF, SVM, DNN) by Major Chemical Family

GrandFamily	Data Points	XGBoost (%)	RF (%)	SVM (%)	DNN (%)
Acides & anhydrides	414	3.7	3.8	2.9	4.2
Alcohols	1161	5.0	6.7	8.3	2.8
Aldehydes	81	2.3	2.3	2.7	1.6
Alkanes	817	7.3	8.9	7.9	8.0
Alkenes	350	2.5	2.2	2.5	2.0
Autres	1302	3.4	3.5	3.1	3.3
Autres hydrocarbures	20	1.6	1.0	0.4	0.4
Benzenes & aryles	1002	2.9	3.4	4.9	2.7
Epoxides	160	2.0	1.4	0.8	0.7
Esters	329	4.1	7.2	13.2	2.6
Ethers	317	2.5	2.2	1.9	2.1
Halogénés	1904	5.1	5.2	4.4	5.2
Ketones	260	2.6	1.6	1.4	1.7
Naphthalenes & polyaryles	138	1.7	1.7	1.3	0.8
Nitriles	172	2.9	2.4	1.4	2.1
Polyfunctionnels	553	3.3	7.3	6.9	1.6
Polyols & glycerides	250	10.5	17.0	22.2	1.8
Silanes/siloxanes	388	2.0	1.6	0.9	1.5
Sulfides/thiophenes	188	2.9	2.8	2.2	3.3

The analysis of relative errors by chemical family reveals significant performance variations across models and approaches. Overall, the QSPR approach demonstrates clear superiority over the GC method for complex chemical families (polyols, polyfunctionals, esters), sometimes halving errors (e.g., polyfunctionals' error dropping from 8.8% to 3.3% with XGBoost). The DNN emerges as the most robust model in both approaches, consistently delivering the lowest errors for critical families (polyols, esters, aldehydes), especially under QSPR where it achieves notable accuracy ($\leq 2.8\%$). Conversely, RF

shows vulnerability to complex structures (error >30% for polyfunctionals under GC), while SVM exhibits inconsistencies (error exceeding 22% for polyols under QSPR). The best-predicted families remain epoxides, naphthalenes, and silanes (errors $\leq 2.5\%$), while polyols and alkanes persist as the most challenging to model. These results underscore the crucial importance of combining the QSPR-based approach with the DNN algorithm to optimize accuracy based on chemical nature, paving the way for targeted modeling by functional families.

IV.4.2 Distributions of relative prediction errors

The analysis of the distributions of relative prediction errors is a critical step in assessing the accuracy and reliability of predictive models. By examining how errors are distributed across various ranges, we gain insights into the strengths and limitations of each model and its ability to generalize to unseen data. In this section, we focus on the performance of the three best-performing models; XGBoost, SVM, and DNN, applied to two distinct modeling approaches: the GC method and the QSPR approach. These results are illustrated in two histograms (referred to as Figure 284 and Figure), which depict the frequency of mean relative prediction errors (in percentage) for each model across predefined error intervals. The Random Forest (RF) model, identified as the least precise among the four implemented techniques, is excluded from this discussion.

Figure 286 presents the distribution of relative prediction errors for the GC approach. This histogram reveals distinct performance patterns among the three models. In the lowest error range (0-1%), GC-SVM stands out with an impressive 6968 predictions, far surpassing GC-DNN (2720) and GC-XGBoost (1588). This suggests that GC-SVM excels at delivering highly accurate predictions for a significant portion of the dataset. However, as the error range increases, GC-XGBoost demonstrates a more consistent performance. For instance, in the 1-2% interval, GC-XGBoost achieves 2070 predictions compared to 812 for both GC-SVM and GC-DNN, and in the 2-3% interval, it records 1439 predictions against 344 for GC-SVM and 574 for GC-DNN. This trend continues across moderate error ranges, with GC-XGBoost maintaining higher frequencies up to the 15-25% interval (444 predictions) compared to GC-SVM (109) and GC-DNN (324). Conversely, GC-DNN shows a notable presence in higher error ranges, such as 10-15% (409 predictions) and 25-50% (268 predictions), indicating a tendency toward larger errors in some cases. All models exhibit minimal predictions in extreme error ranges

(>50%), though GC-DNN and GC-XGBoost have slightly more instances in the >100% range (17 and 16, respectively) compared to GC-SVM (3). These observations highlight GC-SVM's strength in minimizing low errors and GC-XGBoost's balanced performance across a broader spectrum.

Figure illustrates the distribution of relative prediction errors for the QSPR approach. Here, QSPR-SVM leads in the 0-1% error interval with 6017 predictions, followed by QSPR-DNN (4615) and QSPR-XGBoost (3092), reinforcing SVM's capability for high-precision predictions. However, in the subsequent error ranges, QSPR-XGBoost and QSPR-DNN exhibit stronger performances. For example, in the 1-2% interval, QSPR-XGBoost records 2263 predictions compared to 1111 for QSPR-SVM and 1872 for QSPR-DNN, and in the 2-3% interval, it has 1201 predictions against 788 for QSPR-SVM and 500 for QSPR-DNN. QSPR-XGBoost continues to perform well in moderate ranges, such as 7-10% (509 predictions) compared to QSPR-SVM (263) and QSPR-DNN (361). Notably, QSPR-XGBoost has no predictions in the >100% range, contrasting with QSPR-DNN (51) and QSPR-SVM (13), indicating its superior ability to avoid extreme errors. QSPR-DNN, while competitive in the 0-1% range, shows greater variability, with a significant number of predictions in higher error intervals, such as >100%. These results suggest that QSPR-SVM is highly effective for low-error predictions, while QSPR-XGBoost offers robustness across a wider range of errors.

The comparison of the two figures reveals key differences in model performance between the GC and QSPR approaches. In the GC approach, GC-SVM's dominance in the 0-1% range likely stems from its ability to effectively capture the linear or near-linear relationships inherent in group contribution methods. However, its performance wanes in higher error ranges, suggesting limited flexibility for more complex patterns. GC-XGBoost, with its capacity to model non-linear interactions, provides a more balanced error distribution, making it a versatile choice for the GC approach. In the QSPR approach, QSPR-SVM retains its strength in the lowest error range, but QSPR-XGBoost emerges as the most robust model overall, avoiding extreme errors and maintaining solid performance across moderate ranges. QSPR-DNN, despite its precision in the 0-1% range, exhibits greater variability, including a higher incidence of large errors, which may reflect overfitting or sensitivity to the complexity of QSPR descriptors. These findings underscore the trade-offs between precision at low error levels and consistency across broader error ranges.

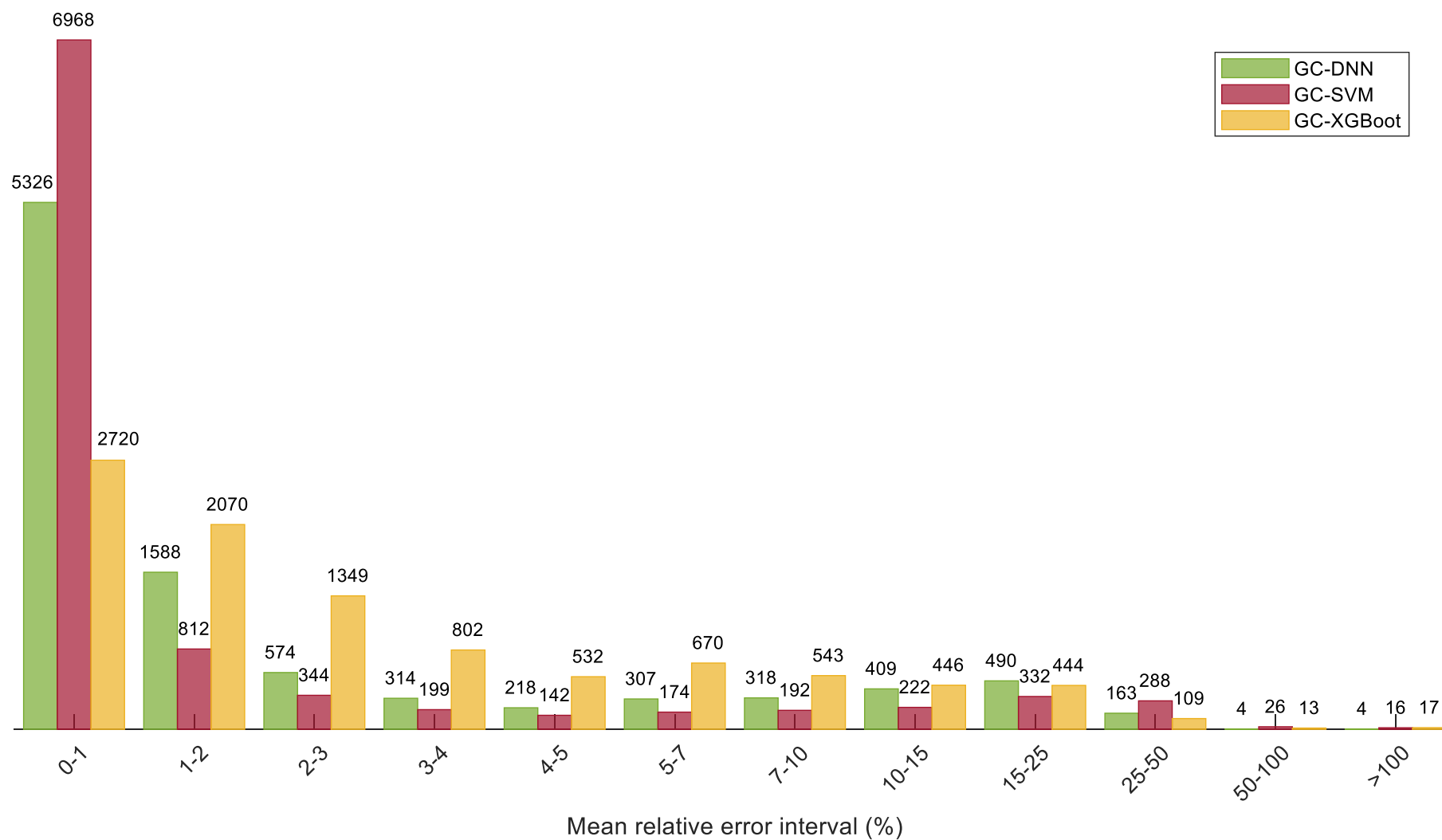


Figure 286: Distribution of data points as a function of mean relative error ER - GC models (%)

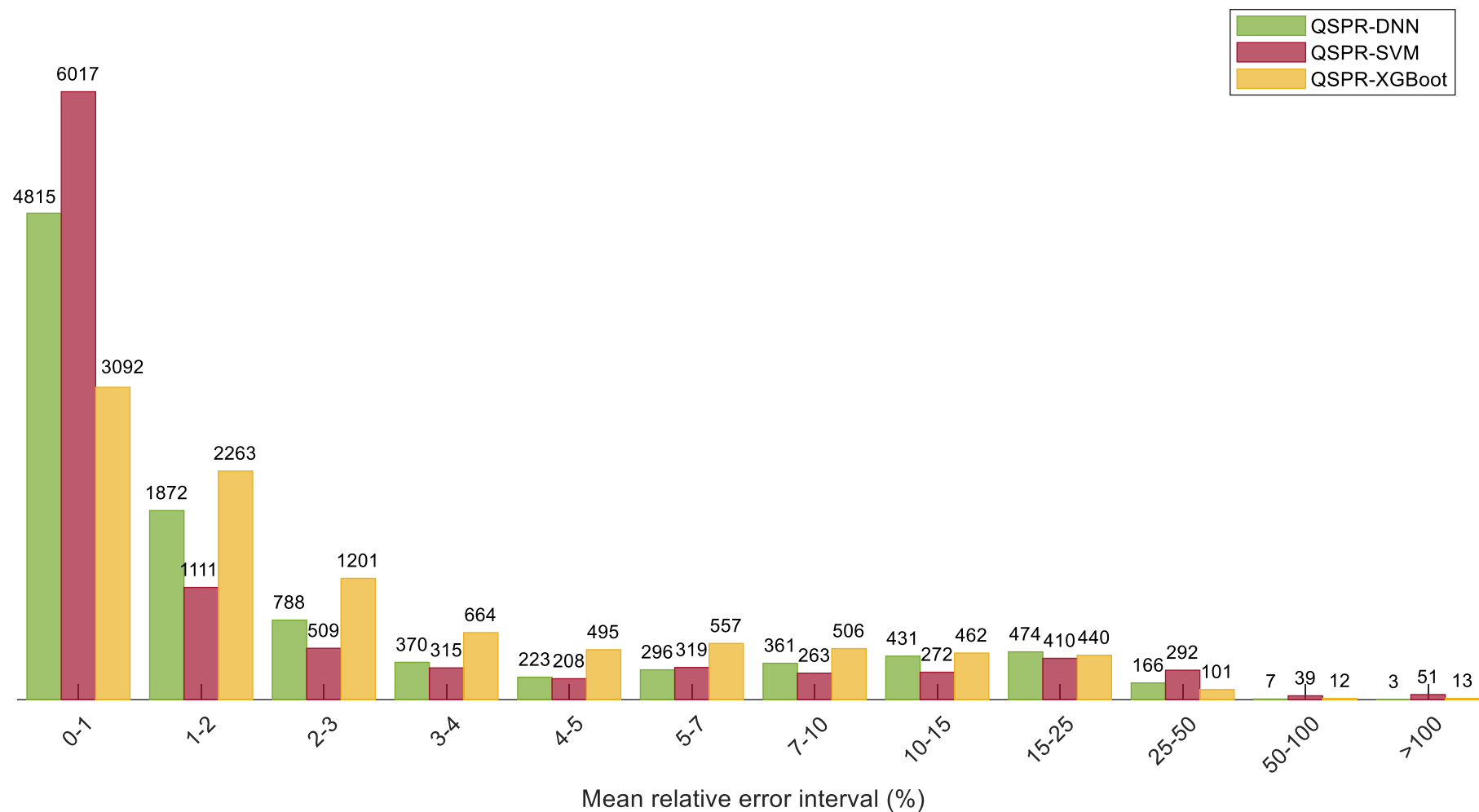


Figure 37: Distribution of data points as a function of mean relative error ER - QSPR models (%)

IV.5 CONCLUSION

This chapter has established a robust framework for predicting the viscosity of non-ionic organic compounds across an extensive thermodynamic range (74–1120 K). By systematically evaluating five modeling techniques (XGBoost, RF, SVM, DNN, and semi-empirical) with two complementary molecular representations (group contribution and QSPR descriptors) on 9,715 experimental data points (DIPPR 801) spanning 668 compounds and 20 chemical families, key insights emerge. The QSPR approach consistently outperformed group contribution methods, particularly for complex structures like polyols and polyfunctional compounds, reducing mean relative errors by up to 60% in critical families. Among algorithms, XGBoost and DNN demonstrated superior accuracy and robustness, achieving test-set AARE $\leq 5.50\%$ and $R^2 \geq 0.87$ with QSPR inputs, while the conventional model showed significant limitations (AARE $> 19\%$).

Notably, chemical family analysis revealed pronounced performance variations: Epoxides, naphthalenes, and silanes were predicted with exceptional accuracy (AARE $< 2.5\%$), whereas polyols and alkanes remained challenging despite DNN-QSPR's significant error reduction (e.g., polyols: 1.8% AARE). Error distribution studies further confirmed the stability of QSPR-based models, with XGBoost exhibiting zero extreme errors ($> 100\%$). These findings underscore that combining global molecular descriptors (QSPR) with nonlinear algorithms (XGBoost/DNN) effectively captures intricate structure-property relationships overlooked by localized fragmentation methods.

For future work, hybrid GC/QSPR frameworks and extension to multicomponent systems represent promising paths to enhance predictive scope. The models developed here provide reliable tools for process optimization, demonstrating that machine learning-driven molecular modeling can overcome traditional limitations in thermophysical property prediction.

GENERAL CONCLUSION

Viscosity, as a fundamental physical property of fluids, plays a decisive role in a multitude of scientific and industrial domains. In chemistry, it influences reaction kinetics, separation processes, and product formulation, while in physics, it is central to the analysis of fluid dynamics. In the field of engineering, viscosity is an essential parameter for designing systems involving heat or mass transfer, as well as for optimizing industrial processes. However, despite its crucial importance, the accurate prediction of the viscosity of organic compounds remains a major challenge. Complex molecular interactions and the structural diversity of the studied compounds complicate this task, while traditional experimental methods, although reliable, are often costly, time-consuming, and unsuitable for high-throughput screening or large-scale applications. This observation highlights the necessity of developing innovative predictive approaches capable of overcoming these limitations and providing robust, generalizable viscosity estimates.

In this context, this thesis lies at the intersection of computational chemistry and machine learning, with the ambition of predicting the viscosity of pure organic compounds, whether ionic or non-ionic. Ionic liquids (ILs), owing to their unique properties such as thermal stability, non-volatility, and tunability, are promising candidates for applications in green chemistry, electrochemistry, and materials science. Yet, their viscosity remains difficult to anticipate due to the complexity of ionic interactions and the diversity of possible ion pairs. Non-ionic organic compounds, widely used in the pharmaceutical, petrochemical, and materials industries, also present predictive challenges related to their molecular heterogeneity and the non-linear variations of their properties with temperature. Thus, this research focuses on a central problem: designing accurate and versatile predictive models for the viscosity of these two classes of compounds, surpassing conventional approaches through innovative methods.

To address this problem, this study adopts a multi-algorithmic strategy, integrating Group Contribution (GC) techniques, Quantitative Structure-Property Relationships (QSPR),

and advanced machine learning algorithms such as XGBoost, Random Forest, Support Vector Machine (SVM), and Deep Neural Networks (DNN). This approach relies on an extensive database, comprising 15,251 experimental data points for 1,654 distinct ionic liquids (from 253 K to 571 K) and 9,715 points for 668 non-ionic compounds spanning 20 chemical families (from 74 K to 1,120 K). The objectives of this thesis are threefold: (1) to develop and compare predictive models for the viscosity of ILs and non-ionic compounds, (2) to evaluate the effectiveness of the different molecular representations and algorithms employed, and (3) to propose reliable tools for the rational design of organic compounds. The methodology relies on rigorous validation, including statistical metrics, cross-validation, and chemical family-based analysis, to ensure the robustness of the results.

The contributions of this work are both theoretical and practical. Firstly, this thesis offers predictive models of remarkable accuracy, notably achieving a mean relative error of 3.99% and an R^2 of 0.9524 for ILs using the QSPR-SVM model, outperforming existing methods. For non-ionic compounds, approaches based on QSPR and non-linear algorithms significantly reduce errors, by up to 60% in some complex cases. Furthermore, this research highlights the superiority of global descriptors (QSPR) over fragmentation methods (GC) for capturing molecular interactions. Finally, the developed tools constitute an efficient alternative to experimental measurements, facilitating the design of compounds for targeted applications. Beyond these advances, this work opens perspectives for hybrid models and their extension to multicomponent systems, thereby strengthening the role of machine learning in the study of thermophysical properties.