

Université Mustapha Stambouli
Mascara



جامعة مصطفى سطمولي
معسكر

Faculté des sciences exactes

Département de chimie

Laboratoire de Chimie Organique, Macromoléculaire et des Matériaux

THESE de DOCTORAT

Spécialité : matériaux innovants et chimie verte

Intitulée

Développement de modèles prédictives pour la détermination des propriétés de transport des liquides ioniques

Présenté par : ZEBIDA Mohammed Amine

Le 19/11/2025

Devant le jury :

Président	REGUIG Bendoukha Abdelkarim	Pr	Université M. S. de Mascara
Examinateur	BOUDJELAL Mokhtar	Pr	Université M. S. de Mascara
Examinateur	GUELLA Soufiane	Pr	Université Djillali Liabes de Sidi Bel Abbes
Directeur de thèse	BENKOUIDER Ali Mustapha	Pr	Université M. S. de Mascara
Co-Directeur de thèse	YAHIAOUI Ahmed	Pr	Université M. S. de Mascara
Invité	ARGOUB kadda	MCA	Université M. S. de Mascara

Année universitaire : 2024-2025

REMERCIEMENTS

Ce travail de thèse a été réalisé au sein du laboratoire de Chimie organique macromoléculaire et des matériaux de la faculté des sciences exactes de l'université Mustapha Stambouli de mascara sous la direction du Professeur **BENKOUIDER Ali Mustapha**

L'aboutissement de cette thèse de doctorat marque l'achèvement d'un long parcours, à la fois scientifique et humain. Il serait impossible de clore ce travail sans exprimer ma profonde reconnaissance envers celles et ceux qui m'ont soutenu tout au long de ces années.

Je tiens tout d'abord à exprimer ma profonde reconnaissance et gratitude au professeur **BENKOUIDER Ali Mustapha** Professeur à l'université Mustapha Stambouli de mascara pour m'avoir accepté dans son équipe et d'avoir dirigé cette thèse. Qu'il me soit permis de remercier pour son encadrement scientifique, ses conseils précieux. J'ai très apprécié la confiance qu'il m'a accordée durant mon travail.

Je suis profondément reconnaissant à Monsieur **YAHIAOUI Ahmed** Professeur à l'université Mustapha Stambouli de mascara pour son accompagnement précieux.

je tiens a remercié très chaleureusement le professeur **HACHEMAOUI Aicha** directrice de laboratoire de Chimie organique macromoléculaire et des matériaux de la faculté des sciences exactes de l'université Mustapha Stambouli de mascara pour l'honneur de m'avoir accueillie.

J'adresse mes respectueuses grâcietés à Monsieur **REGUIG Bendoukha Abdelkarim** Professeur à l'Université de MASCRA pour m'avoir fait l'honneur de présider le jury de soutenance.

Mes sincères remerciements vont également à monsieur **Boudjelal Mokhtar** Professeur à l'université Mustapha Stambouli de Mascara, et Monsieur **Guella Soufiane** Professeur à l'Université de Sidi Bel Abbès pour avoir accepté d'examiner ce travail. Je tiens à remercier chaleureusement notre invité d'honneur Monsieur **Argoub Kadda** Docteur à l'université Mustapha Stambouli de Mascara pour son participation à ce jury et pour l'intérêt qu'il a porté à mes travaux.

Je voudrais aussi exprimer toute ma reconnaissance à **ma famille** qui tout au long de ces années d'études n'ont cessé de m'encourager

Pour n'oublier personne : Merci à tous ceux qui ont partagé avec moi un moment de travail ou de détente ces dernières années.

A

Mes très chers parents,

Ma petite famille

Mon frère, mes sœurs et mes amis. . .

الملخص:

يستكشف هذا العمل البحثي في إطار أطروحة الدكتوراه التنبؤ بخصائص النقل للسوائل الأيونية، وهي مركبات تتألف من أيونات ذات خصائص فريدة مثل الاستقرار الحراري وعدم النطاير والمرنة التركيبية العالية. تجعل هذه الخصائص السوائل الأيونية ذات قيمة كبيرة في تطبيقات مثل الكيمياء الكهربائية والتغذيف وإدارة الحرارة، غير أن التفاعلات المعقدة بينها تُعد التنبؤ بخصائص مثل التوصيلية الكهربائية والتوصيلية الحرارية واللزوجة. وإزاء محدودية الطرق التجريبية المكلفة والبطيئة، يعتمد هذا البحث على الذكاء الاصطناعي وتعلم الآلة لتطوير نماذج تنبؤية دقيقة من خلال استغلال قواعد بيانات غنية وخوارزميات متقدمة.

تعتمد المنهجية على إنشاء قواعد بيانات منقاة بعناية، تجمع آلاف نقاط القياس لعدة عائلات من السوائل الأيونية. لقد خضعت كل مجموعة من البيانات لمراجعة دقيقة تأخذ في الحسبان حالات عدم اليقين والظروف التجريبية (مثل درجة الحرارة والضغط)، وذلك لضمان موثوقية النماذج. تُوصف البنية الجزيئية باستعمال منهجيات مثل مساهمات المجموعات (GC)، التي تُجزئ الجزيئات إلى وحدات كيميائية، والعلاقات الكمية بنية-خاصية (QSPR)، التي تستخدم المتغيرات الوصفية لقياس الخصائص الجزيئية. تُعدّي هذه البيانات خوارزميات تعلم الآلة، بما في ذلك Deep Neural Networks وSupport Vector Machine (SVM) وRandom Forest وXGBoost (DNN). تقوم هذه الأدوات بنمذجة الروابط بين البنية الكيميائية وخصائص النقل، مقِمةً في الوقت نفسه الدقة وقابلية التفسير، وغالباً ما تتفوق على الأساليب التقليدية.

بالنسبة للتوصيلية الكهربائية، تُظهر النتائج أن نماذج DNN و SVM تحقق دقةً قريبةً من حدود عدم اليقين التجريبية، مع أخطاء نسبية متوسطة غالباً أقل من 3% ومعامل R^2 يقارب 0.99. أما بالنسبة للتوصيل الحراري، فتبرز منهجيات QSPR XGBoost (DNN) المقربون بـ QSPR خطأً متوسطاً في حدود 2.7%. وقد تم التحقق من صحة هذه النماذج بدقة باستخدام تقنيات التحقق المقاطع والتحقق الخارجي.

تقترح هذه الأطروحة أدوات تنبؤية قوية لتصميم السوائل الأيونية بشكل عقلاني، مما يقلل الاعتماد على القياسات التجريبية. وتشمل الأفاق المستقبلية توسيع قواعد البيانات لتشمل العوائل قليلة التمثيل (مثل الأنيونات النادرة والكاتيونات الجديدة)، والتوسيع إلى خصائص ناشئة أخرى، وتطوير نماذج هجينية تجمع بين الذكاء الاصطناعي والمحاكاة الذرية لزيادة دقة وشمولية التنبؤات.

الكلمات المفتاحية:

السوائل الأيونية، التوصيل الكهربائي، التوصيل الحراري، اللزوجة، مساهمات المجموعات، QSPR، الشبكات العصبية العميقية، آلات الدعم الناقل، XGBoost، الغابات العشوائية.

Résumé

Ce travail de thèse explore la prédiction des propriétés de transport des liquides ioniques, des composés constitués d'ions aux propriétés uniques comme la stabilité thermique, la non-volatilité et une grande modularité. Ces caractéristiques rendent les LI précieux pour des applications en électrochimie, catalyse ou gestion thermique, mais leurs interactions complexes compliquent la prédiction de propriétés telles que la conductivité électrique, la conductivité thermique et la viscosité. Face aux limites des méthodes expérimentales, coûteuses et lentes, cette recherche s'appuie sur l'intelligence artificielle (IA) et l'apprentissage automatique pour développer des modèles prédictifs précis, en exploitant des bases de données riches et des algorithmes avancés.

La méthodologie repose sur la constitution de bases de données soigneusement sélectionnées, regroupant des milliers de points de mesure pour diverses familles de LI. Chaque jeu de données a fait l'objet d'une révision rigoureuse prenant en compte les incertitudes et les conditions expérimentales (température, pression) afin d'assurer la fiabilité des modèles. Les structures moléculaires sont décrites via des approches comme les contributions de groupes (GC), qui fragmentent les molécules en unités chimiques, et les relations quantitatives structure-propriété (QSPR), qui utilisent des descripteurs pour quantifier les caractéristiques moléculaires. Ces données alimentent des algorithmes d'apprentissage automatique, notamment XGBoost, Random Forest, Support Vector Machine (SVM) et Deep Neural Networks (DNN). Ces outils modélisent les liens entre structures chimiques et propriétés de transport, offrant à la fois précision et interprétabilité, et surpassant souvent les approches traditionnelles.

Pour la conductivité électrique, les résultats démontrent que les modèles DNN et SVM atteignent des précisions proches des incertitudes expérimentales, avec des erreurs relatives moyennes souvent inférieures à 3 % et des coefficients R^2 avoisinant 0,99. Pour la conductivité thermique, les approches QSPR-XGBoost et QSPR-SVM se distinguent ($ERAM \approx 0,52\%$, $R^2 \approx 0,9981$), tandis que pour la viscosité, le DNN couplé aux QSPR obtient une erreur moyenne de l'ordre de 2,7 %. Les modèles ont été rigoureusement validés par des techniques de validation croisée et externe.

Cette thèse propose des outils prédictifs robustes pour la conception rationnelle des liquides ioniques, réduisant la dépendance aux mesures expérimentales. Les perspectives incluent l'élargissement des bases de données aux familles peu représentées (anions rares,

nouveaux cations), l'extension à d'autres propriétés émergentes et le développement de modèles hybrides combinant l'intelligence artificielle et simulations atomistiques pour accroître encore la précision et la portée des prévisions.

Mots clés :

liquides ioniques , conductivité électrique, conductivité thermique , viscosité , contributions de groupes, QSPR ,Réseaux de neurones profonds , machines à vecteurs de support , XGBoost, forêts aléatoires.

Abstract

This thesis explores the prediction of transport properties of ionic liquids (ILs) – compounds composed of ions that exhibit unique characteristics such as thermal stability, non-volatility, and high tunability. These characteristics make ILs valuable for applications in electrochemistry, catalysis, and thermal management, but the complex interactions between ions complicate the prediction of properties such as electrical conductivity, thermal conductivity, and viscosity. In light of the limitations of experimental methods, which are costly and time-consuming, this research relies on artificial intelligence (AI) and machine learning to develop accurate predictive models by leveraging rich datasets and advanced algorithms.

The methodology is based on carefully curated databases containing thousands of measurement points for various IL families. Each dataset underwent rigorous review to account for experimental uncertainties and conditions (temperature, pressure) in order to ensure the reliability of the models. Molecular structures are described using approaches such as group contributions (GC) and quantitative structure–property relationships (QSPR). GC methods fragment molecules into chemical units, while QSPR uses molecular descriptors to quantify structural characteristics. These data feed into machine learning algorithms, including XGBoost, Random Forest, Support Vector Machines (SVM), and Deep Neural Networks (DNN). These tools model the relationships between chemical structures and transport properties, offering both accuracy and interpretability, and often outperforming traditional approaches.

For electrical conductivity, the results demonstrate that DNN and SVM models achieve accuracies close to experimental uncertainties, with mean relative errors often below 3% and R^2 values around 0.99. For thermal conductivity, the QSPR XGBoost and QSPR

SVM approaches stand out (mean relative error $\approx 0.52\%$, $R^2 \approx 0.9981$), while for viscosity, the DNN coupled with QSPR achieves a mean error on the order of 2.7%. The models were rigorously validated using cross-validation and external validation techniques.

This thesis provides robust predictive tools for the rational design of ionic liquids, reducing reliance on experimental measurements. Future work includes expanding the databases to underrepresented families (rare anions, novel cations), extending to other emerging properties, and developing hybrid models that combine artificial intelligence with atomistic simulations to further improve the accuracy and scope of predictions.

Keywords:

ionic liquids, electrical conductivity, thermal conductivity, viscosity, group contributions, QSPR, deep neural networks, support vector machines, XGBoost, random forests.

SOMMAIRE

Table des matières

Remerciements	i
Sommaire	viii
Liste des figures	xii
Liste des Tableaux	xv
Nomenclature	xvi
Introduction générale	1
Chapitre I: Synthèse bibliographique et concept théorique	4
I.1 Introduction	4
I.1 Définition des liquides ioniques et défis de conception	5
I.2 Propriétés de transport	8
I.2.1 Définition de la conductivité électrique	9
I.2.2 Définition de la conductivité thermique	11
I.2.3 Définition de la viscosité	11
I.3 Mesure des propriétés de transport	13
I.3.1 Mesure de la Conductivité Électrique	13
I.3.2 mesure de la conductivité thermique	23
I.3.3 Mesure de la viscosité	29
I.4 Revue des méthodes de prédiction des propriétés de transport	37
I.4.1 Conductivité Électrique	37
I.1.1.A Méthodes de contributions de	41
I.4.2 Conductivité Thermique	43
I.4.3 Viscosité	48
I.5 Conclusion	62
I.6 Bibliographie	63
Chapitre II: Méthodologie et méthodes de modélisation	76

II.1	INTRODUCTION	76
II.2	IMPORTANCE DES BASES DE DONNÉES	77
II.2.1	Révision des données	78
II.3	MÉTHODES DE CONTRIBUTION DE GROUPE	80
II.3.1	Principe	80
II.3.2	Hiérarchie des groupes	81
II.3.3	Groupes fonctionnels	87
II.4	Relations Quantitatives Structure-Propriété.....	89
II.4.1	Principe	89
II.4.2	Descripteurs moléculaires	89
II.4.3	Empreintes digitales dans la modélisation prédictive	93
II.4.4	Réduction de la dimensionnalité des descripteurs.....	95
II.5	Modèles et régression.....	98
II.5.1	Modèles conventionnels.....	99
II.5.2	Algorithmes d'apprentissage automatique.....	99
II.5.3	Réglage des hyper-paramètres	110
II.6	Validation du modèle	111
II.6.1	Validation croisée et validation externe	113
II.7	Conclusion	115
II.8	Bibliographie.....	116
Chapitre III: prédition de la conductivité électrique des liquides ioniques		119
III.1	Introduction	119
III.2	base de données expérimentales.....	120
III.2.1	Collecte de données:	120
III.2.2	Révision des données	120
III.2.3	Analyse de données.....	126
III.3	Méthodologie	129
III.4	Résultats et discussion.....	131
III.4.1	Méthodes de contributions de groupes.....	131
III.4.2	Méthodes QSPR	140
III.4.3	Comparaison des modèles proposés par rapport aux ceux existants	150
III.5	Conclusion	153
III.6	Référence	154
Chapitre IV: prédition de la conductivité thermique des liquides ioniques.....		155
IV.1	Introduction	155
IV.2	Collecte et traitement des données.....	156
IV.2.1	Collecte de données:	156

IV.2.2	Révision des données	157
IV.2.3	Analyse de données.....	162
IV.3	Stratégie méthodologique de modélisation	165
IV.4	Résultats et discussion.....	166
IV.4.1	Méthode de contributions de groupes	167
IV.4.2	Méthodes QSPR.....	174
IV.4.3	Comparaison des modèles proposés par rapport aux ceux existants	183
IV.5	Conclusion	186
IV.6	Référence	187
Chapitre V: Prédiction de la viscosité des liquides ioniques		189
V.1	Introduction.....	189
V.2	base de données expérimentales.....	189
V.3	Méthodologique pour la modélisation.....	194
V.4	Résultats et discussion.....	196
V.4.1	Méthodes de Contributions de Groupes	196
V.4.2	Méthodes QSPR.....	202
V.4.3	Comparaison des modèles proposés par rapport aux ceux existants	208
V.5	Conclusion	210
V.6	Référence	212
Conclusion générale		214
Annexe 1 : Extraire de la base de données de la conductivité électrique.....		216

LISTE DES FIGURES

Figure 1:quelque propriété des liquides ioniques.....	6
Figure 2: diverses applications des IL.....	6
Figure 3:Nombre de publications par année portant sur les liquides ioniques (source sciencedirect)	7
Figure 4: Gradient de vitesse pour un fluide non newtonien tel que discuté par Masse	12
Figure 5 : Migration différente des ions lorsqu'un courant continu ou alternatif est appliqué à la cellule de mesure.....	15
Figure6: Schéma de principe de la méthode des deux pointes.....	17
Figure 7 : Schéma de principe de la méthode des quatre pointes.....	17
Figure 8 : Schéma de principe de la méthode de Van der Pauw.....	18
Figure 9 : Schéma d'un capteur toroïdal pour méthode inductive.	20
Figure 10 : Configuration capacitive avec électrodes externes.....	21
Figure 11 : Dispositif micro-ondes avec guide d'ondes et échantillon.	22
Figure 12 : Le principe de la méthode du fil chaud transitoire	25
Figure 13: Schéma de principe de méthode de la Source Plane Transitoire	25
Figure 14 : Le principe de la Plaque Chaude Gardée.....	27
Figure15 : Le principe de la Cylindres Concentriques.....	28
Figure 16: Quelques types de capillaires en verre : 1 : Capillaire d'Ostwald, 2 : Capillaire d'Ubbelohde, 3 : Capillaire de Cannon-Fenske, 4 : Capillaire de Houillon	30
Figure 17: Schéma illustrant les forces agissant sur une bille en chute ; F_g force de gravité, F_b force de flottabilité ascendante et F_d force de friction visqueuse (interne).....	31
Figure 18: Schéma d'un viscosimètre à corde vibrante	32
Figure 19: Schéma d'un viscosimètre à cristal vibrant et plaque vibrante	33
Figure 20: Types de viscosimètre rotatif ; (a) Viscosimètre à cylindre coaxial, (b) Viscosimètre à plan conique, (c) Viscosimètre plan	35
Figure 21:Algorithme de prétraitement des données	79
Figure 22: Exemple de découpe du 2-méthylnonane selon un schéma de découpe d'ordre zéro	82
Figure 23: Exemple de découpe de 2-Nonanon selon un schéma de découpe du premier ordre	82
Figure 24: Exemple de découpe d'Oxymethylene-Ether selon un schéma de découpe du second ordre	83
Figure 25: Découpe du composé L'acide 4-vinylbenzoïque par le Marrero et Gani.....	84
Figure 26: 1,9-nonadiol.....	86
Figure 27: N-Phényl-1,4-benzènediamine	86
Figure 28: Exemple de nouveau schéma de découpage utilisés pour représenter le 3-ethyl-1-methyl-1H-imidazol-3-ium L-tyrosinate.....	88
Figure 29: Représentation des méthodes QSPR.....	89
Figure 30: génération des modèles en utilisant des signatures moléculaires de hauteurs 0 à 3 pour chaque atome de la molécule	95
Figure 31: Schéma de fonctionnement Random Forest (RF).....	101
Figure 32: Schéma de fonctionnement de XGBoost [25].	104
Figure 33: Architecture réseau d'une régression SVM.....	106
Figure 34: représentation illustrative des réseaux de neurones.....	107

Figure 35: représentation d'un réseau de neurones avec trois couches cachées et un neurone de sortie.....	109
Figure 36: Représentation schématique de la validation croisée Leave-One-Out	114
Figure 37: Organigramme de la procédure de formation (validation croisée) et de test (validation externe).....	115
Figure 38: Familles chimiques cationiques et anioniques étudiées. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.....	124
Figure 39: Nombre d'LI selon différentes combinaisons cation-anion avant traitement des données. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.....	125
Figure 40: Nombre d'IL selon différentes combinaisons cation-anion après traitement des données. Les abréviations des familles chimiques sont détaillées dans le Tableau 5	125
Figure 41: Influence de la température sur la conductivité électrique des liquides ioniques....	127
Figure 42: Influence de la longueur de chaîne du cation sur la conductivité électrique des liquides ioniques (298k,101 kpa)	128
Figure 43: Influence de la pression sur la conductivité électrique des liquides ioniques (298k)	129
Figure 44: Comparaison des modèles de contribution de groupes : (a) Nuages de points des valeurs prédictes et expérimentales dans l'ensemble d'apprentissage et l'ensemble de test. (b) Distribution gaussienne des erreurs de prédiction dans l'ensemble d'apprentissage et l'ensemble de test. L'unité de conductivité électrique est (S/m).....	134
Figure 45: Distribution des écarts relatifs (ER) en fonction des familles cationiques et anioniques. La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieur et supérieur de la boîte représentent respectivement le premier et le troisième quartile. (a) Résultats du modèle GC-DNN. (b) Résultats du modèle GC-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.....	136
Figure 46: Distributions des écarts relatifs ER(%) en fonction des différentes familles de cations et d'anions. (a) modèle DNN, (b) modèle SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.....	138
Figure 47: Distributions des points de données en fonction des erreurs relatif moyen ER (%). (a) modèle DNN, (b) modèle SVM.	140
Figure 48: Comparaison des modèles QSPR sélectionnés pour prédire σ : (a) Nuages de points des valeurs prédictes et expérimentales dans l'ensemble d'apprentissage et l'ensemble de test. (b) Distribution gaussienne des erreurs de prédiction dans l'ensemble d'apprentissage et l'ensemble de test. L'unité de conductivité électrique est (S/m).....	144
Figure 49: Distribution des écarts relatifs (ER) en fonction des familles cationiques et anioniques. La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieur et supérieur de la boîte représentent respectivement le premier et le troisième quartile. (a) Résultats du modèle QSPR-DNN. (b) Résultats du modèle QSPR-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.....	146
Figure 50: Distributions des écarts relatifs ER(%) en fonction des différentes familles de cations et d'anions. (a) modèle QSPR-DNN, (b) modèle QSPR-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.....	148
Figure 51: Distributions des points de données en fonction des erreurs relatif moyen ER (%). (a) modèle DNN, (b) modèle SVM.	149
Figure 52 : Familles chimiques cationiques et anioniques considérées dans cette étude. Les abréviations des familles chimiques sont détaillées dans le Tableau 10.....	160
Figure 53 : Nombre de point de données expérimentales de la conductivité thermique selon différentes combinaisons cation-anion. Les abréviations des familles chimiques sont détaillées dans Tableau 10.	161

Figure 54 : Nombre de sous ensembles de valeurs expérimentales de la conductivité thermique selon différentes combinaisons cation-anion. Les abréviations des familles chimiques sont détaillées dans Tableau 10.	161
Figure 55 : Variation de la conductivité thermique (λ) des liquides ioniques en fonction de la température (en K) pour des ILs sélectionnés.	163
Figure 56 : Influence de la pression sur la conductivité thermique des liquides ioniques	164
Figure 57: Comparaison des modèles de contribution de groupes : (a) Nuages de points des valeurs prédictes et expérimentales dans l'ensemble d'apprentissage et l'ensemble de test. (b) Distribution gaussienne des erreurs de prédiction dans l'ensemble d'apprentissage et l'ensemble de test. L'unité de conductivité thermique est ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$).	168
Figure 58: Distribution des écarts relatifs (ER) en fonction des familles cationiques et anioniques. La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieur et supérieur de la boîte représentent respectivement le premier et le troisième quartile. (a) Résultats du modèle GC- XGBoost. (b) Résultats du modèle GC-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 10.....	170
Figure 59: Distributions des écarts relatifs ER(%) en fonction des différentes familles de cations et d'anions. (a) modèle GC- XGBoost, (b) modèle GC-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 10.....	172
Figure 60: Distributions des points de données en fonction des erreurs relatif moyen ER (%). (a) modèle GC-XGBoost, (b) modèle GC-SVM.	174
Figure 61: Comparaison des modèles QSPR sélectionnés pour prédire λ : (a) Nuages de points des valeurs prédictes et expérimentales dans l'ensemble d'apprentissage et l'ensemble de test. (b) Distribution gaussienne des erreurs de prédiction dans l'ensemble d'apprentissage et l'ensemble de test. L'unité de conductivité thermique est ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$).	177
Figure 62: Distribution des écarts relatifs (ER) en fonction des familles cationiques et anioniques. La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieur et supérieur de la boîte représentent respectivement le premier et le troisième quartile. (a) Résultats du modèle QSPR- XGBoost. (b) Résultats du modèle QSPR-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 10.....	179
Figure 63: Distributions des écarts relatifs ER(%) en fonction des différentes familles de cations et d'anions. (a) modèle GC- XGBoost, (b) modèle GC-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 10.....	181
Figure 64: Distributions des points de données en fonction des erreurs relatif moyen ER (%). (a) modèle QSPR-XGBoost, (b) modèle QSPR-SVM.	182
Figure 65. (a) Nombre de liquides ioniques à base d'imidazolium selon les différentes familles d'anions. (b) Nombre de points de données expérimentales selon les différentes familles d'anions. Les abréviations des familles chimiques sont détaillées dans le Tableau 15.	191
Figure 66 : Structures chimiques des familles anioniques et imidazolium. Les abréviations des familles chimiques sont détaillées dans le Tableau 15.....	192
Figure 67. Benchmark des modèles GC: (a) Corrélation prédit/expérimental (train/test), (b) Distribution normale des résidus (train/test)	198
Figure 68. : Distribution des erreurs relatifs (ER%) par familles anioniques pour les 3 modèles GC les plus performants (représentation en boîtes à moustaches). La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieurs et supérieurs de la boîte représentent respectivement le premier et le troisième quartile.	200
Figure 69. : Distribution de l'erreur relative moyenne (ER, %) des modèles de GC : (a) DNN, (b) SVM, (c) XGBoost.....	201

Figure 70. : Benchmark des modèles QSPR: (a) Corrélation prédit/expérimental (train/test), (b) Distribution normale des résidus (train/test)	204
Figure 71. : Distribution des erreurs relatifs par familles anioniques pour les 3 modèles QSPR les plus performants (représentation en boîtes à moustaches). La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieur et supérieur de la boîte représentent respectivement le premier et le troisième quartile.....	206
Figure 72. : Distribution de l'erreur relative moyenne (ER %) des modèles de QSPR : (a) DNN, (b) SVM, (c) XGBoost.....	208

LISTE DES TABLEAUX

Tableau 1:Comparaison entre les Méthodes Expérimentales et l'Intelligence Artificielle pour la recherche de nouveaux liquides ioniques.....	8
Tableau 2 :Découpage de la molécule 1,9-nonadiol en groupes	86
Tableau 3 :Découpage de la molécule N-Phényl-1,4-benzènediamine en groupes	87
Tableau 4 :Espace de recherche des hyper-paramètres et détermination de leur sélection optimale à l'aide de l'algorithme d'optimisation bayésien	111
Tableau 5: Résumé de la base de données sur la conductivité électrique par familles d'ions...	123
Tableau 6: Espace de recherche des hyperparamètres et détermination de leur sélection optimale à l'aide d'un algorithme d'optimisation bayésienne.	131
Tableau 7: Statistiques de performance des GC-modèles sur les jeux d'apprentissage et de test, exprimées en moyennes issues d'une validation croisée à 5-plis, répétée 25 fois selon des partitions aléatoires.	132
Tableau 8: Statistiques de performance des QSPR-modèles sur les jeux d'apprentissage et de test, exprimées en moyennes issues d'une validation croisée à 5-plis, répétée 25 fois selon des partitions aléatoires.	141
Tableau 9: Comparaison des performances des modèles proposés par rapport aux modèles existants.....	152
Tableau 10 : Résumé de la base de données sur la conductivité thermique par familles d'ions	159
Tableau 11: Espace de recherche des hyperparamètres et détermination de leur sélection optimale à l'aide d'un algorithme d'optimisation bayésienne.	166
Tableau 12: Statistiques de performance des GC-modèles sur les jeux d'apprentissage et de test, exprimées en moyennes issues d'une validation croisée à 5-plis, répétée 25 fois selon des partitions aléatoires.	167
Tableau 13: Statistiques de performance des QSPR-modèles sur les jeux d'apprentissage et de test, exprimées en moyennes issues d'une validation croisée à 5-plis, répétée 25 fois selon des partitions aléatoires.	175
Tableau 14: Comparaison des performances des modèles proposés par rapport aux modèles existants.....	185
Tableau 15: Résumé de la base de données de viscosité par familles d'ions	193
Tableau 16: Performances des modèles GC : moyennes issues d'une validation croisée 5 folds × 25 répétitions (ensembles d'entraînement et de test).....	196
Tableau 17: Performances des modèles QSPR : moyennes issues d'une validation croisée 5 folds × 25 répétitions (ensembles d'entraînement et de test).....	202
Tableau 18: Comparaison des performances des modèles proposés par rapport aux modèles existants.....	210

NOMENCLATURE

IL : ionique liquide

Ω : Ohm

V : volt

A : ampère

G : conductance

S : siemens

κ : conductivité

η : coefficient de viscosité

μ : viscosité dynamique

ν : viscosité cinématique

ρ : densité

THW : fil Chaud Transitoire

TLS : source de Ligne Transitoire

TPS : source de plan transitoire

HMF : détecteur de flux de chaleur

GC : contribution de groupe

VFT : L'équation de Vogel-Fulcher-Tammann

M_w : masse molaire

λ : conductivité thermique

GA : algorithme génétique

cp : capacité calorifique

EC : états correspondants
MLR : algorithme de régression linéaire multiple
SVM : algorithme de machine à vecteurs de support
FC-CS : contributions de fragments-états correspondants
EAM : erreur absolue moyenne
ERAM : erreur relative moyenne absolue
EQM : erreur quadratique moyenne
IQR : écart interquartile
ER : écart relatif
 R^2 : coefficient de détermination
FT : théorie du Frottement
FVT : théorie du Volume Libre
EFT : théorie du Fluide Étendu
SAFT : fluides associatifs statistiques
 k : constante de Boltzmann
SMILE : Système simplifié d'entrée de ligne pour la représentation moléculaire
QSPR : relation quantitative structure-propriété
QSAR : relation quantitative structure-activité
FFANN : réseau neuronal artificiel à trois couches
SVM : machine à vecteurs de support
LSSVM : machine à vecteurs de support des moindres carrés
RF : random Forest
XGBoost : Extreme Gradient Boosting
RBF : fonctions de base radiale
MLP : Multi-Layer Perceptron
DNN : réseau neuronal profond
ECFP : empreintes de connectivité étendue
PCA : l'analyse en composantes principales
PLS : moindres carrés partiels
ICA : L'analyse en composantes indépendantes

RFE : élimination récursive de caractéristiques

LASSO : Least Absolute Shrinkage and Selection Operator

INTRODUCTION GENERALE

INTRODUCTION GENERALE

Les liquides ioniques (IL) constituent une classe de composés exceptionnels, entièrement formés d'ions et caractérisés par des points de fusion inférieurs à 100°C. Découverts en 1914 avec le nitrate d'éthylammonium, ils ont récemment suscité un vif intérêt grâce à leurs propriétés physico-chimiques uniques : stabilité thermique et chimique, non-volatilité, large fenêtre électrochimique et modularité par le choix des cations et anions. Ces attributs en font des candidats idéaux pour de nombreuses applications industrielles et scientifiques, telles que l'électrochimie (batteries, supercondensateurs), la catalyse, les procédés de séparation, ou encore les fluides fonctionnels dans les systèmes énergétiques.

Cependant, exploiter pleinement le potentiel des IL nécessite une compréhension approfondie et une prédition précise de leurs propriétés de transport, notamment la conductivité électrique, la conductivité thermique et la viscosité. Ces propriétés, qui régissent les performances des IL dans des contextes pratiques, comme le transfert de charge, la gestion thermique ou la rhéologie, sont difficiles à prévoir en raison de la complexité des interactions ioniques, de la diversité des structures moléculaires et de leur sensibilité aux conditions thermodynamiques (température, pression). Les approches expérimentales traditionnelles, bien qu'essentielles, sont souvent limitées par leur coût, leur durée et la difficulté à explorer l'immense espace chimique des IL.

Dans ce contexte, les méthodes prédictives basées sur l'intelligence artificielle (IA), l'apprentissage automatique et la modélisation computationnelle émergent comme des outils puissants pour accélérer la découverte et l'optimisation des IL. En combinant des bases de données expérimentales étendues avec des descripteurs moléculaires pertinents

y compris des structures moléculaires fonctionnelles, ces approches permettent de modéliser les relations structure-propriété complexes et de fournir des prédictions précises et généralisables.

Cette thèse s'attache à répondre à ces défis en développant des modèles prédictifs robustes et fiables pour les propriétés de transport des IL. Ses objectifs principaux sont les suivants :

1. **Constitution de bases de données exhaustives** : Collecter et réviser des données expérimentales, couvrant plusieurs milliers de points de mesure pour différentes familles de LI, sur la conductivité électrique, la conductivité thermique et la viscosité, en tenant compte des incertitudes de mesure et des conditions expérimentales.
2. **Développement de modèles prédictifs avancés** : Implémentation et optimisation de méthodologies de modélisation, incluant les approches Relations Quantitatives Structure-Propriété (QSPR), Contributions de Groupes (CG) et une palette d'algorithmes d'IA : Extreme Gradient Boosting¹ (XGBoost), Random Forest² (RF), Support Vector Machine³ (SVM), Deep Neural Network⁴ (DNN).
3. **Validation et optimisation** : Évaluation comparative des performances des modèles selon la nature chimique des cations et anions, les conditions thermodynamiques (température, pression) et la diversité structurale, en analysant les sources d'incertitude et les limites de chaque approche.
4. **Recommandations d'utilisation et perspectives** : Proposition de recommandations pour l'application industrielle des modèles développés et perspectives d'extension aux propriétés émergentes et à d'autres types de liquides ioniques.

La thèse s'organise en cinq chapitres principaux :

- **Chapitre I : Synthèse bibliographique et concept théorique** : Une exploration des concepts théoriques des propriétés de transport, des méthodes expérimentales et des approches prédictives existantes, posant les bases du travail.

¹ Renforcement extrême du gradient

² Forêts aléatoires

³ Machines à vecteurs de support

⁴ Réseaux de neurones profonds

- **Chapitre II : Méthodologie et méthodes de modélisation :** Une présentation des outils de modélisation (bases de données, GC, QSPR, apprentissage automatique) et des techniques de validation employées.
- **Chapitre III : Prédiction de la conductivité électrique des liquides ioniques :** Développement de modèles basés sur 6014 points de données pour 373 IL, avec des performances remarquables via DNN et SVM.
- **Chapitre IV : Prédiction de la conductivité thermique des liquides ioniques :** Modélisation à partir de 1025 points pour 79 IL, avec des résultats exceptionnels grâce à l'approche XGBoost et SVM-QSPR.
- **Chapitre V : Prédiction de la viscosité des liquides ioniques :** Analyse de 7015 points pour 620 IL à base d'imidazolium, où le DNN excelle.

Les contributions majeures de ce travail résident dans le développement de modèles prédictifs atteignant des précisions proches des incertitudes expérimentales, surpassant les méthodes existantes grâce à des bases de données plus riches et à des approches d'apprentissage automatique avancées. Ces avancées offrent des outils précieux pour la conception rationnelle des nouveaux IL et ouvrent la voie à des recherches futures sur d'autres propriétés et applications

CHAPITRE I : SYNTHESE BIBLIOGRAPHIQUE ET CONCEPT THEORIQUE

CHAPITRE I: SYNTHESE BIBLIOGRAPHIQUE ET CONCEPT THEORIQUE

I.1 INTRODUCTION

Dans ce chapitre inaugural, notre étude s'articule autour d'une exploration structurée en trois axes complémentaires.

Premièrement, Nous traitons du potentiel des liquides ioniques (IL) dans les applications industrielles et de la nécessité d'utiliser la prédition et l'intelligence artificielle (IA) pour accélérer la découverte d'IL.

Dans un second temps, nous exposons les concepts théoriques fondamentaux régissant les propriétés de transport – conductivité électrique, conductivité thermique et viscosité – en clarifiant leurs définitions, leurs mécanismes physico-chimiques sous-jacents et les lois qui les gouvernent.

Dans un troisième temps, nous proposons une analyse critique et systématique des méthodes expérimentales employées pour la mesure de ces propriétés, mettant en lumière leurs avantages, leurs contraintes techniques et leurs domaines d'application privilégiés.

Complémentant ces aspects expérimentaux, notre travail se consacre à une analyse approfondie des approches prédictives dédiées aux propriétés de transport. À travers un examen détaillé de la littérature scientifique, nous cartographions l'éventail des modèles théoriques et des outils numériques disponibles, en discutant leur pertinence, leurs hypothèses fondatrices et leurs limites intrinsèques.

En intégrant ces trois dimensions – théorique, expérimentale et prédictive – ce chapitre vise à offrir au lecteur un cadre uniifié pour appréhender les défis et les avancées récentes dans l'étude des propriétés de transport, tout en posant les bases méthodologiques essentielles à la suite de cette thèse.

I.1 DEFINITION DES LIQUIDES IONIQUES ET DEFIS DE CONCEPTION

Définis comme sels organiques liquides à température ambiante, les liquides ioniques (LI) offrent des propriétés chimiques modulables (Figure 1). Cette adaptabilité permet leur emploi industriel dans plusieurs domaines comme solvants, électrolytes ou systèmes de séparation gazeuse, en particulier pour la capture du CO₂ [1] (Figure 2).

En principe, le réglage des propriétés IL pour une application donnée est possible par modification ou échange des cations et des anions. Le nombre d'IL possibles est estimé à plus de 10¹⁸, cependant, le nombre d'IL qui pourraient être construits en tenant compte uniquement des ions publiés à ce jour est d'environ 150 000, alors qu'environ 1 500 seulement sont disponibles dans le commerce [2]. Le nombre de publications portant sur ce sujet a augmenté de façon quasi exponentielle (voir Figure 3), ce qui témoigne d'un intérêt scientifique croissant.

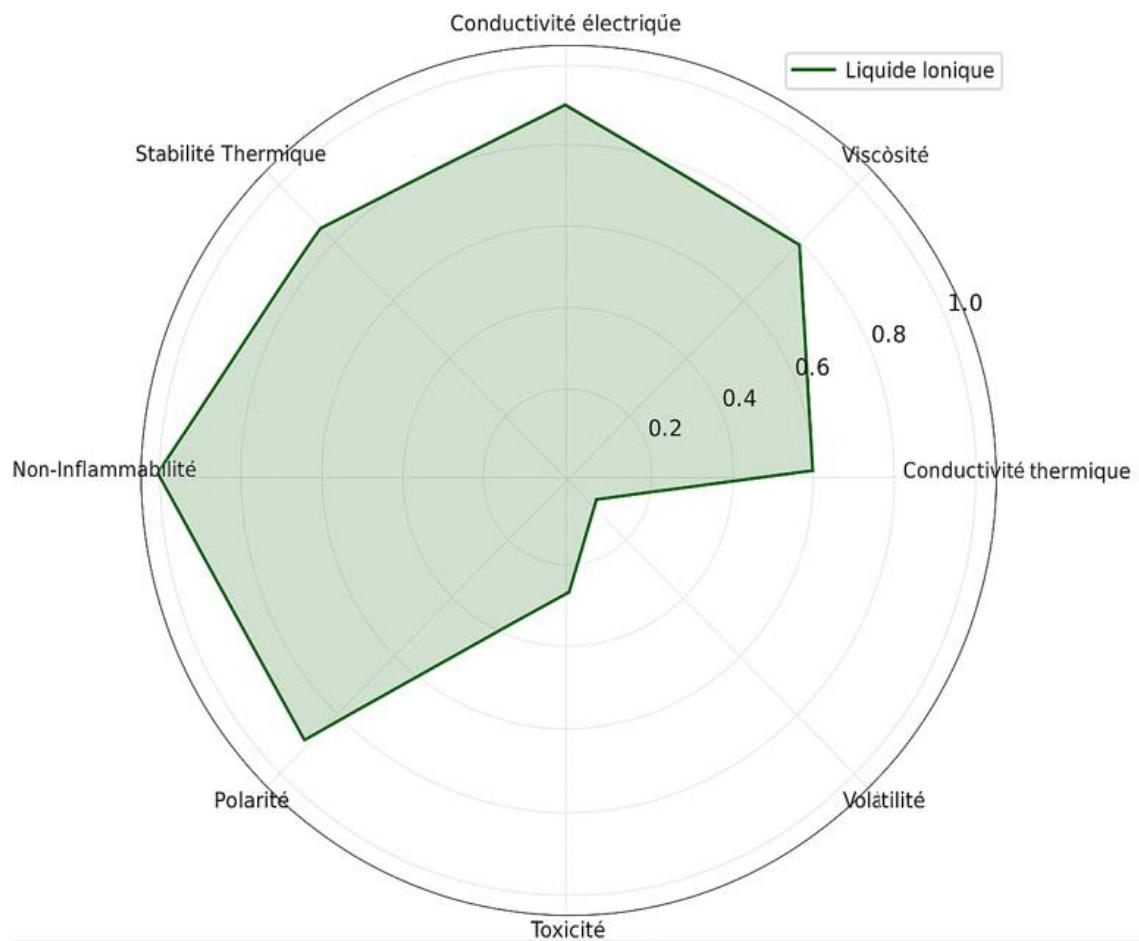


Figure 1: quelques propriétés des liquides ioniques

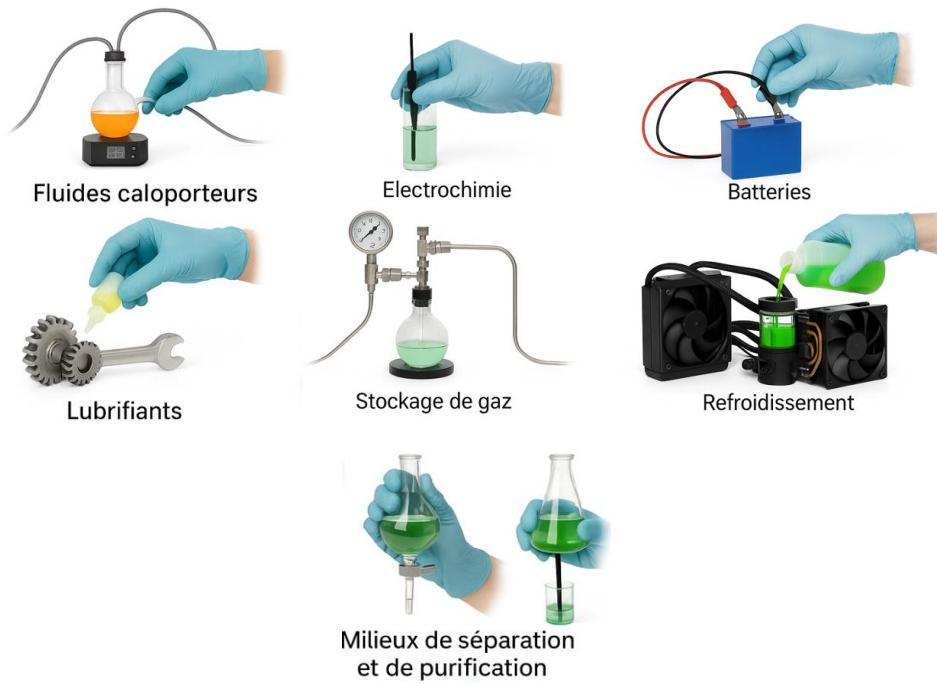


Figure 2: diverses applications des IL.

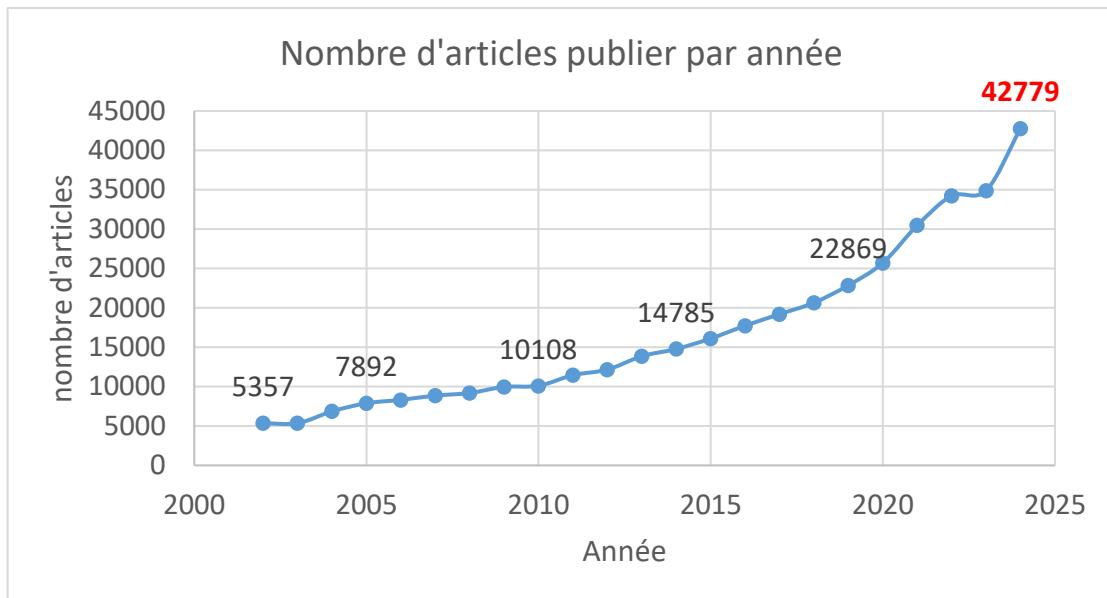


Figure 3: Nombre de publications par année portant sur les liquides ioniques (source sciencedirect)

Cependant, l'identification de l'IL optimal avec un ensemble spécifique de propriétés pour une application donnée est un défi majeur en raison des coûts élevés de synthèse et de caractérisation des matériaux pour étudier de nombreuses structures candidates au cours du processus de conception. En fait, les approches traditionnelles de développement de matériaux reposant sur des recherches par essais et erreurs et/ou systématiques pour identifier les conceptions optimales peuvent prendre des décennies pour introduire de nouveaux matériaux dans l'industrie. Il faut passer à un nouveau paradigme de la science, un paradigme basé sur la prédiction et l'intelligence artificielle (AI), afin d'accélérer ce processus et d'optimiser le développement des matériaux. Les techniques d'IA, notamment les réseaux de neurones et les modèles de machine learning, peuvent être utilisées pour prédire les propriétés des liquides ioniques en fonction de leur structure chimique. Cela permettrait d'identifier rapidement quels ions et combinaisons d'ions sont les plus susceptibles de posséder des propriétés spécifiques telles que la conductivité thermique, conductivité électrique, ou la viscosité, sans avoir à synthétiser et tester chaque composé individuellement.

Tableau 1: Comparaison entre les Méthodes Expérimentales et l'Intelligence Artificielle pour la recherche de nouveaux liquides ioniques

Critères	Méthodes expérimentales	Prédiction et Intelligence Artificielle (IA)
Temps de développement	Long (synthèse + caractérisation + tests)	Court (calculs et simulations rapides)
Coût	Élevé (matériaux, équipements, main-d'œuvre)	Réduit (besoin d'infrastructure informatique)
Exploration du domaine	Limitée à un nombre restreint d'échantillons	Large, peut tester virtuellement des milliers de combinaisons
Découverte de nouvelles structures	Par essais et erreurs, empirique	Guidée par optimisation et apprentissage automatique
Manipulation de substances	Risque réel lors des tests	Aucun risque direct
Adaptabilité aux conditions extrêmes	Parfois difficile à reproduire en laboratoire	Peut modéliser des conditions variées sans contraintes physiques
Impact environnemental	Consommation de ressources et déchets	Plus écologique,

I.2 PROPRIÉTÉS DE TRANSPORT

Les propriétés de transport jouent un rôle fondamental dans le comportement et les performances des liquides ioniques, tant dans les applications industrielles que dans la recherche en laboratoire. Leur maîtrise est cruciale pour garantir l'efficacité des procédés de transformation et pour développer de nouveaux matériaux aux caractéristiques sur-mesure. Dans un contexte scientifique, comprendre et contrôler ces paramètres permet d'innover en matière de conception de matériaux, en optimisant notamment les échanges d'énergie et de matière.

Parmi ces propriétés, la viscosité, la conductivité thermique et la conductivité électrique se distinguent particulièrement, aux côtés d'autres paramètres tels que la diffusivité et la

tension superficielle. Dans le cadre de cette thèse, nous concentrons notre étude sur ces trois propriétés essentielles, car elles représentent des leviers clés tant pour l'optimisation des performances des liquides ioniques que pour l'amélioration des technologies développées au laboratoire.

I.2.1 Définition de la conductivité électrique

La conductivité électrique correspond à la capacité d'un matériau à transporter un courant électrique. Pour que l'électricité se déplace, la présence de particules chargées est indispensable. On distingue généralement deux grands types de conducteurs selon la nature des particules responsables du transport de la charge.

Le premier groupe regroupe les conducteurs dans lesquels la charge est assurée par des électrons. Dans ces matériaux, notamment les métaux et le graphite, les atomes forment un réseau où les électrons de valence sont délocalisés, constituant ainsi un « nuage d'électrons » qui facilite la conduction du courant.

Le second groupe concerne les conducteurs ioniques, dans lesquels le courant est transporté non pas par des électrons, mais par des ions chargés. Dans ces systèmes, appelés électrolytes, le transfert de charge est étroitement lié au déplacement de la matière. L'ionisation dans ces matériaux s'opère généralement par dissolution dans un solvant polaire, tel que l'eau, ou par fusion, ce qui libère des ions mobiles capables de conduire l'électricité. Les liquides ioniques, objet de cette étude, constituent une classe particulière d'électrolytes où la conductivité est gouvernée par la dynamique des ions en milieu purement ionique, sans solvant moléculaire.

Selon la loi d'Ohm (1), la tension (U) établie à travers une solution est proportionnelle au courant circulant (I) :

$$U = R \times I \quad (1)$$

R = résistance (ohm, Ω) U = tension (volt, V), I = courant (ampère, A)

la résistance (R) est une constante de proportionnalité, qui peut être calculée d'après le courant circulant mesuré si une tension connue est appliquée :

$$R = \frac{U}{I} \quad (2)$$

La conductance (G) est définie comme l'inverse de la résistance :

$$G = \frac{I}{R} \quad (3)$$

G =conductance (Siemens, S)

Pour mesurer la résistance ou la conductance d'un échantillon, une cellule dite de mesure est nécessaire. La cellule de mesure est constituée d'au moins deux pôles de charge opposée. Le terme « électrode » s'emploie comme synonyme de pôle et les pôles forment ensemble une cellule de mesure. La cellule et le corps isolant qui maintient l'ensemble s'appellent le capteur. La mesure dépend de la géométrie de la cellule, qui est décrite par la constante de cellule (K). C'est le rapport de la distance (l) et de la surface (A) des pôles

$$K = \frac{l}{A} \quad (4)$$

K = constante de cellule (cm-1) l = distance entre les électrodes (cm) A = surface effective de contact des électrodes avec l'électrolyte (cm²) La conductance peut être transformée en conductivité normalisée (σ), qui est indépendante de la configuration de la cellule de mesure. Cela s'obtient en multipliant la conductance par la constante de cellule :

$$\sigma = G \times \frac{l}{A} = G \times K \quad (5)$$

Cette propriété (σ), exprimée généralement en S·m⁻¹ (Siemens par mètre), joue un rôle déterminant dans la conception et l'optimisation des applications électrochimiques impliquant ces milieux.

I.2.2 Définition de la conductivité thermique

La conductivité thermique (notée k , λ ou κ) quantifie la capacité d'un matériau à transférer l'énergie thermique par conduction. Exprimée en $\text{W}/(\text{m}\cdot\text{K})$, elle est définie comme la quantité de chaleur traversant une surface donnée par unité de temps, pour une différence de température donnée sur une épaisseur unitaire. Ce paramètre reflète l'efficacité du transfert d'énergie à l'échelle microscopique via les interactions entre molécules ou atomes.

La conductivité thermique se produit par agitation moléculaire et contact et ne résulte pas du mouvement en masse du solide lui-même. La chaleur se déplace le long d'un gradient de température, d'une zone de haute température et d'énergie moléculaire élevée vers une zone de température plus basse et d'énergie moléculaire inférieure. Le transfert de chaleur persiste jusqu'à ce que les deux objets atteignent la même température, connue sous le nom d'équilibre thermique. La vitesse à laquelle la chaleur se transfère dépend de l'ampleur du gradient de température et des caractéristiques thermiques spécifiques du matériau.

La conductivité thermique est quantifiée à l'aide du système international d'unités (unité SI) de $\text{W}/(\text{m}\cdot\text{K})$ (Watts par mètre par degré Kelvin) et est l'inverse de la résistivité thermique, qui mesure la capacité d'un objet à résister au transfert de chaleur. Il peut être calculé à l'aide de l'équation suivante :

$$K = \frac{Q \times L}{A(T_2 - T_1)} \quad (6)$$

Où, Q c'est le flux de chaleur (W), L c'est longueur ou épaisseur du matériau (m) , A c'est la surface du matériau (m^2), $T_2 - T_1$ c'est le gradient de température (K).

I.2.3 Définition de la viscosité

La viscosité, grandeur physique centrale dans l'étude des fluides, quantifie la résistance d'un matériau à l'écoulement ou à la déformation sous l'effet d'une contrainte tangentielle. Elle traduit les frottements internes résultant des interactions entre les constituants du fluide lors de leur mouvement relatif. Formellement, pour un écoulement laminaire, elle est définie par la loi de Newton :

$$\tau = -\eta \frac{\partial u}{\partial y} \quad (7)$$

Dans ce contexte, la constante η représente le coefficient de viscosité, également appelé viscosité dynamique. La notation avec le sous-script " y " dans le terme de contrainte de cisaillement indique la zone spécifique sur laquelle la contrainte de cisaillement est exercée. Cette zone est orientée perpendiculairement à l'axe des y et est située à une distance " y " de l'origine. Pendant ce temps, le deuxième sous-script " x " signifie la direction le long de laquelle la contrainte de cisaillement est appliquée. Le signe négatif dans l'équation signifie que la contrainte de cisaillement est exercée depuis une région de vitesse plus élevée vers une région de vitesse plus faible, reflétant le concept fondamental de la contrainte de cisaillement agissant pour égaliser les gradients de vitesse dans un milieu fluide. Dans ce contexte de modélisation, de nombreux fluides se conforment au principe de Newton et sont classés comme des fluides newtoniens. En revanche, les fluides non newtoniens présentent une relation plus complexe entre la contrainte de cisaillement et le gradient de vitesse, s'écartant de la linéarité simple (voir Figure 4).

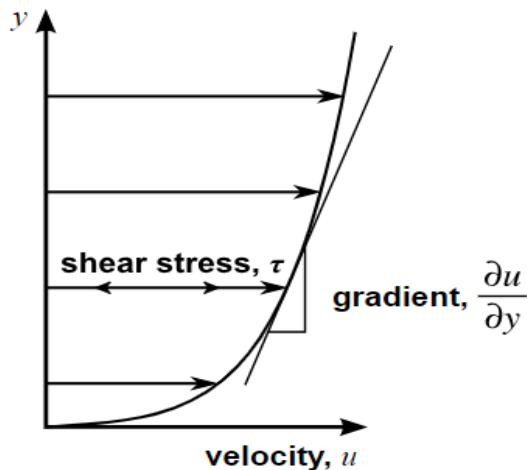


Figure 4: Gradient de vitesse pour un fluide non newtonien tel que discuté par Masse

Ainsi, la viscosité, par définition, émerge comme le rapport de la contrainte de cisaillement par unité de surface au gradient de vitesse. L'unité SI de viscosité dynamique est le pascal-seconde ($Pa \cdot s$). Ses autres unités sont le newton-seconde par mètre carré ($N.s.m^{-2}$) ou le poise (P). Il existe deux types de viscosité couramment utilisés : la viscosité dynamique (μ), également connue sous le nom de viscosité absolue, et la viscosité cinématique (ν). La viscosité cinématique, quant à elle, relie la viscosité dynamique à la densité du fluide. Elle reflète la résistance du fluide à l'écoulement. La

viscosité cinématique est obtenue en divisant la viscosité dynamique par la densité du fluide (ρ) et est exprimée en unités de mètres carrés par seconde (m^2/s). Tout au long de ce manuscrit, lorsqu'il est fait référence à "la viscosité" sans autre clarification, cela implique la viscosité dynamique.

I.3 MESURE DES PROPRIÉTÉS DE TRANSPORT

Dans le domaine de la mesure des propriétés de transport des liquides ioniques, une gamme diversifiée d'appareil de mesure existe, chacun utilisant des technologies distinctes pour déterminer des propriétés de transport des liquides ioniques [1,2]. Ces technologies présentent des degrés de précision et de sensibilité variables, ce qui pose des défis pour le développement de modèles prédictifs robustes. Une compréhension approfondie de ces aspects est primordiale pour le développement de modèles prédictive efficace pour les propriétés de transport des liquides ioniques.

Nous présentons ici un aperçu des principaux techniques des appareils de mesure et des incertitudes qui leur sont associés. Il est impératif de noter que bien que cette étude s'efforce de fournir une analyse approfondie, elle ne prétend pas être exhaustive dans sa portée.

I.3.1 Mesure de la Conductivité Électrique

La conductivité électrique constitue l'une des propriétés de transport essentielles des liquides ioniques, traduisant leur aptitude à conduire un courant. Toutefois, sa mesure se confronte à des défis liés aux spécificités physico-chimiques de ces liquides ioniques, telles que la viscosité [3], la structuration locale [4] et les interactions intermoléculaires [5][6]. Ces caractéristiques influent sur la répartition du champ électrique au sein de la cellule de mesure, générant des incertitudes quant aux paramètres géométriques (distance entre électrodes, surface effective) et aux phénomènes d'interface.

Pour réduire ces sources d'erreur et améliorer la reproductibilité, plusieurs stratégies expérimentales complémentaires ont été développées [7–9]. Ces stratégies, chacune présentant des avantages et des limites en fonction des spécificités des liquides ioniques. Elles peuvent être regroupées et hiérarchisées selon divers critères, tels que l'échelle de temps, la nature du signal appliqué ou encore la résolution spatiale. Parmi ces méthodes, on distingue:

I.3.1.1 Méthodes basées sur le type de signal électrique

1. Courant continu (CC) Méthodes

Les méthodes à courant continu (CC) constituent une approche classique et directe pour caractériser la conductivité des liquides ioniques. Elles reposent sur l'application d'un signal d'excitation constant (soit en tension, soit en courant) afin d'obtenir une réponse stationnaire du système. Cette simplicité méthodologique permet d'exploiter la loi d'Ohm pour déterminer la conductivité globale du milieu [9,10].

Dans le cadre des mesures CC, deux grandes approches se distinguent :

❖ Mesures en régime stationnaire :

L'échantillon est soumis à une tension ou un courant constant jusqu'à l'atteinte d'un état d'équilibre. Le courant résultant ou la chute de tension mesurée permet ensuite de déduire la conductivité. Cette méthode, bien que simple, est particulièrement sensible aux phénomènes de polarisation des électrodes et aux réactions électrochimiques indésirables, qui peuvent biaiser l'interprétation des résultats [9,11].

❖ Mesures en régime dynamique :

Des méthodes telles que la chronoamperométrie ou la chronopotentiométrie consistent à étudier la réponse du système suite à un pas brusque de potentiel ou de courant [12]. Ces approches fournissent des informations sur la cinétique de la migration ionique et sur les processus de relaxation, permettant ainsi d'approfondir la compréhension des mécanismes de transfert de charge aux interfaces électrode/échantillon [11].

2. Méthodes à courant alternatif (CA)

Les méthodes à courant alternatif (CA) forment une famille de techniques électrochimiques indispensables pour caractériser les liquides ioniques, dont la conductivité dépend de mécanismes complexes tels que la mobilité ionique, les interactions solvant-soluté et les phénomènes de relaxation diélectrique. Contrairement au courant continu (CC), qui génère des polarisations indésirables aux électrodes [8,9], les méthodes CA exploitent des signaux oscillants pour accéder à des informations dynamiques sur les phénomènes de transport ionique et les processus interfacés (Figure 5).

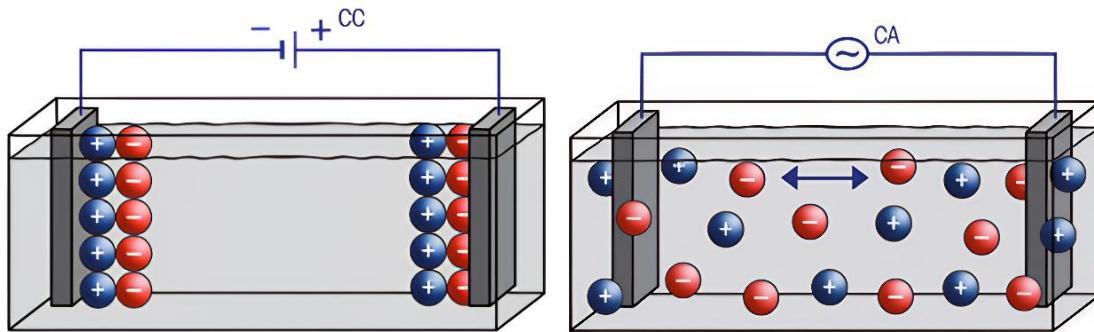


Figure 5 : Migration différente des ions lorsqu'un courant continu ou alternatif est appliqué à la cellule de mesure

Parmi ces approches, on distingue :

1. **La conductométrie en CA** : Mesure la conductance globale du milieu à une fréquence fixe (souvent 1 kHz), utilisée pour des analyses rapides mais moins discriminantes que l'EIS [13].
2. **La spectroscopie diélectrique** : Étudie la permittivité complexe, $\epsilon^*(\omega)$, sur une large gamme de fréquences (10^{-2} Hz – 10^9 Hz), révélant les processus de polarisation dipolaire et les relaxations structurales des LIIs [14]. Cette méthode permet une étude approfondie de la conductivité et des mécanismes ioniques.
3. **La voltampérométrie en CA** : Superpose un signal sinusoïdal à un balayage de potentiel en CC, permettant d'étudier les cinétiques de réaction aux électrodes, notamment dans les milieux visqueux comme les LIIs [9].
4. **La méthode d'impédance électrochimique (EIS)** : Approche phare, combinant une résolution fréquentielle étendue (10 mHz – 100 kHz) et une modélisation par circuits équivalents pour dissocier les contributions résistives (R), capacitives (C) et diffusives [15,16].

Si la conductométrie et la voltampérométrie en CA offrent des données simplifiées pour des applications industrielles (ex. : contrôle qualité), l'EIS se démarque par sa polyvalence. En couplant une excitation sinusoïdale ($V(t) = V_0 \sin(\omega t)$) à une analyse de la réponse en phase et amplitude, elle cartographie l'impédance complexe $Z(\omega)$, un véritable « ADN fréquentiel » du matériau. $Z(\omega)$ s'exprime par [17] :

$$Z(\omega) = \frac{I(\omega)}{V(\omega)} = ZRe + jZIm \quad (8)$$

où $\omega = 2\pi f$ est la pulsation, et $j2 = -1$. Cette représentation fréquentielle (généralement 10 mHz – 100 kHz) permet de dissocier les processus électrochimiques selon leurs constantes de temps caractéristiques [17].

L'EIS présente plusieurs avantages pour la mesure de la conductivité électrique. Elle permet une analyse détaillée des propriétés électrochimiques en fonction de la fréquence, ce qui est essentiel pour comprendre les mécanismes de transport ionique dans les liquides ioniques [9,15,16,18]. L'EIS est non invasive, garantissant que les échantillons restent intacts pendant l'analyse, et offre une grande précision dans la décomposition de l'impédance en résistance et réactance, permettant ainsi d'identifier les processus de transfert de charge et de diffusion[15]. Cependant, l'EIS a également quelques limites. Elle peut être sensible aux conditions expérimentales telles que la température et l'humidité [18]. De plus, la résolution des données peut être affectée par les effets de bruit de mesure, ce qui nécessite une calibration précise des appareils. Malgré ces limites, l'EIS reste une méthode puissante pour optimiser la compréhension des systèmes électrochimiques complexes.

I.3.1.2 Méthodes basées sur la configuration des électrodes (contact et sans contact)

1. Méthode de deux points (électrodes)

Dans la méthode des deux points, deux électrodes en contact direct avec l'échantillon assurent simultanément l'injection du courant et la mesure de la chute de potentiel [9]. Le courant I est appliqué via une électrode et la tension V mesurée sur l'ensemble du montage, permettant de calculer une résistance apparente $R = V/I$. Pour obtenir la conductivité électrique σ , on applique la formule $\sigma = (1/R) \times K$, où K représente le facteur géométrique, intégrant la distance entre électrodes et la surface de contact.

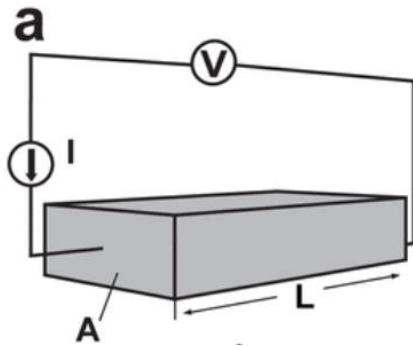


Figure 6: Schéma de principe de la méthode des deux pointes

Cette approche doit toutefois intégrer la contribution des résistances parasites (contacts, fils), souvent non négligeables. Pour en déduire la conductivité électrique, il est essentiel de corriger ces influences, particulièrement dans le cas des liquides ioniques, où la stabilisation des interfaces et le contrôle de la dispersion du liquide demeurent cruciaux [19]. Cette méthode, plus simple en configuration, nécessite ainsi des ajustements spécifiques pour garantir la précision des mesures.

2. Méthode des Quatre Pointes (Sonde de Van der Pauw)

La méthode des Quatre Pointes repose sur l'utilisation de quatre électrodes placées placés en colinéaires en contact avec l'échantillon [20]. Dans une configuration classique, deux électrodes servent à injecter un courant et les deux autres électrodes mesurent la chute de potentiel (Figure 7).

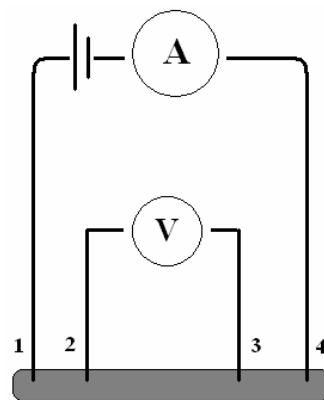


Figure 7 : Schéma de principe de la méthode des quatre pointes

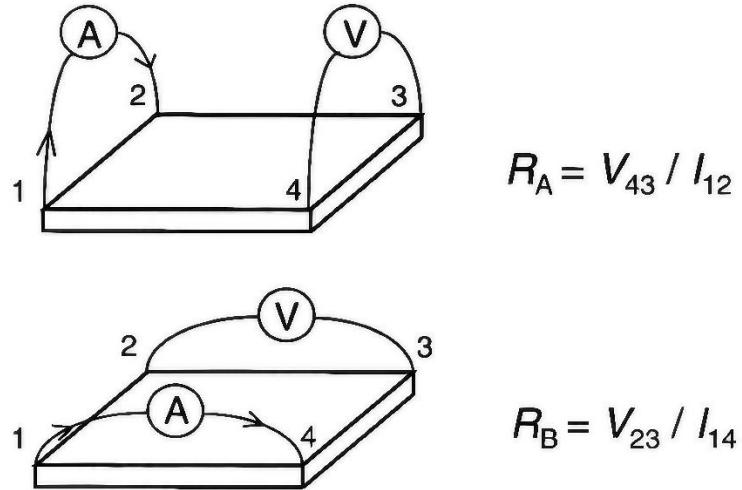


Figure 8 : Schéma de principe de la méthode de Van der Pauw

L’application de la méthode Van der Pauw [20] peut compenser les insuffisances des méthodes de mesure à deux et quatre points [33,82]. La méthode Van der Pauw diffère de la sonde à quatre points car les sondes ne sont pas colinéaires (Figure 8 : Schéma de principe de la méthode de Van der Pauw); les quatre sondes sont connectées aux coins de l’échantillon. Le courant est mesuré par un ampèremètre d’un côté, et la tension générée est mesurée du côté opposé par un voltmètre. Dans la première configuration, un courant I_{12} est appliqué entre les contacts 1 et 2, tandis que la tension V_{43} est mesurée entre les contacts 4 et 3. Cette configuration permet d’obtenir la résistance R_A définie par la relation $R_A = V_{43}/I_{12}$. De manière analogue, dans la seconde configuration, un courant I_{14} est injecté entre les contacts 1 et 4, et la tension V_{23} est mesurée entre les contacts 2 et 3. La résistance correspondante est alors définie par $R_B = V_{23}/I_{14}$.

La résistivité ρ de l’échantillon est calculée à partir de ces deux résistances, en appliquant la formule de Van der Pauw, qui s’exprime comme suit :

$$\exp\left(\frac{-\pi R_A}{\rho}\right) + \exp\left(\frac{-\pi R_B}{\rho}\right) = 1 \quad (9)$$

Cette équation permet d’évaluer la résistivité (et, par conséquent, la conductivité électrique) d’un échantillon à partir de mesures effectuées dans une géométrie symétrique ou arbitraire.

Cette méthode présente l'avantage de réduire l'influence des contacts ainsi que des imperfections géométriques, ce qui la rend particulièrement efficace dans des systèmes caractérisés par des interfaces complexes. En minimisant ces sources d'erreur, elle permet d'obtenir des mesures plus fiables et d'améliorer la précision des analyses [21]. Cependant, son application aux liquides ioniques nécessite une adaptation spécifique. En effet, il est indispensable de stabiliser les interfaces et de maîtriser la dispersion du liquide dans la cellule de mesure. Ces ajustements sont essentiels, car ils permettent de limiter les incertitudes supplémentaires qui pourraient compromettre la qualité des données obtenues.

3. Méthodes sans contact

Les méthodes sans contact offrent une alternative aux techniques traditionnelles en évitant tout contact direct entre les électrodes et le liquide ionique. Ces approches sont particulièrement avantageuses pour les échantillons sensibles à la contamination, corrosifs, ou nécessitant une mesure dans des conditions de pureté élevée. Elles éliminent également les problèmes liés à la polarisation des électrodes et à la dégradation des interfaces, fréquents dans les méthodes avec contact. Trois principales méthodes sont couramment utilisées : les méthodes inductives, capacitives et micro-ondes.

❖ Méthode inductive (capteur toroïdal)

La méthode inductive repose sur l'utilisation d'un capteur toroïdal composé de deux bobines magnétiquement couplées (Figure 9 : Schéma d'un capteur toroïdal pour méthode inductive.). Un courant alternatif est appliqué à la bobine primaire, générant un champ magnétique qui induit un courant circulaire (courants de Foucault) dans le liquide ionique. Ce courant induit dépend de la conductivité électrique σ du liquide. La bobine secondaire mesure ensuite le champ magnétique résultant, permettant de déduire σ à partir de la relation :

$$\sigma = k \cdot \frac{1}{L} \cdot \frac{I_{induit}}{f \cdot \mu_0 N^2 A} \quad (10)$$

où k est une constante sans dimension dépendant de la géométrie du système, L est la distance entre les bobines, f la fréquence du signal, μ_0 la perméabilité magnétique du vide, N le nombre de spires, et A la section transversale du tore [22].

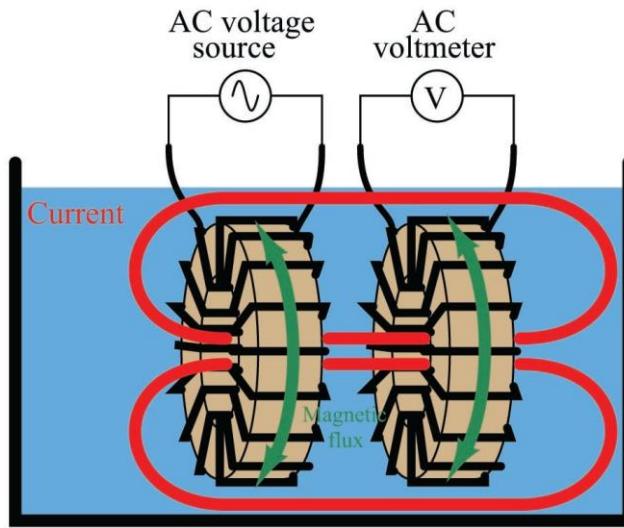


Figure 9 : Schéma d'un capteur toroïdal pour méthode inductive.

Cette méthode est idéale pour les liquides ioniques corrosifs ou visqueux, car le capteur peut être isolé chimiquement (ex. : revêtement en PTFE⁵). Cependant, une calibration précise est requise pour tenir compte de la géométrie du capteur et des variations de température, critiques pour les liquides ioniques dont la conductivité est fortement thermo-dépendante [23].

❖ Méthode capacitive

Dans cette approche, deux électrodes sont placées de part et d'autre d'un récipient diélectrique contenant le liquide ionique (Figure 9 : Schéma d'un capteur toroïdal pour méthode inductive.). Un signal alternatif est appliqué aux électrodes, créant un champ électrique qui traverse le liquide. La conductivité influence la composante resistive de l'impédance mesurée Z , liée à la constante diélectrique complexe $\epsilon^* = \epsilon' - j \frac{\sigma}{\omega \epsilon_0}$. La conductivité est extraite en analysant la réponse en fréquence, via la relation :

$$\sigma = \omega \epsilon_0 \epsilon'' \quad (11)$$

où ϵ'' représente la partie imaginaire de la permittivité, ω la pulsation, et ϵ_0 la permittivité du vide [14].

⁵ Polytétrafluoroéthylène

Les avantages incluent la compatibilité avec des échantillons encapsulés et l'absence d'usure des électrodes. Toutefois, la précision dépend fortement des propriétés diélectriques du récipient et de la stabilité thermique. Pour les liquides ioniques à forte conductivité, les effets de double couche électrique, bien que réduits, peuvent nécessiter des corrections à basse fréquence [24].

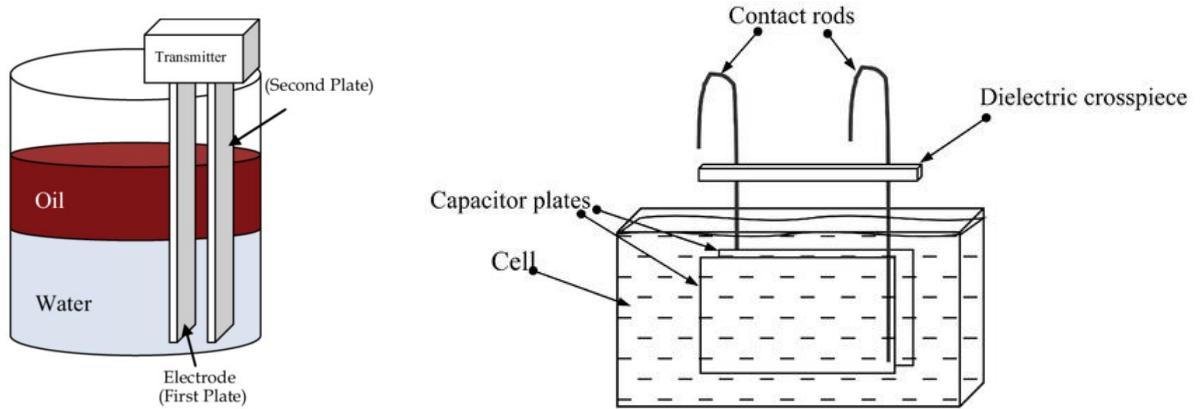


Figure 10 : Configuration capacitive avec électrodes externes.

❖ Méthode micro-ondes

Cette technique analyse l'atténuation et le déphasage d'un signal micro-ondes traversant le liquide (Figure 11). La conductivité électrique affecte directement l'absorption du rayonnement, décrite par l'équation :

$$\sigma = \omega \epsilon_0 \epsilon'' \tan\delta \quad (12)$$

où $\tan\delta$ est la tangente de perte, liée à l'énergie dissipée [25]. Un réseau de guides d'ondes ou un détecteur analyseur permet de mesurer ces paramètres, offrant une résolution élevée même pour des échantillons opaques.

Bien que coûteuse en instrumentation, cette méthode est non invasive et adaptée aux liquides ioniques dans des environnements contrôlés (ex. : réacteurs sous pression). Les défis incluent la calibration des effets de dispersion ionique et la compensation des interférences électromagnétiques [26].

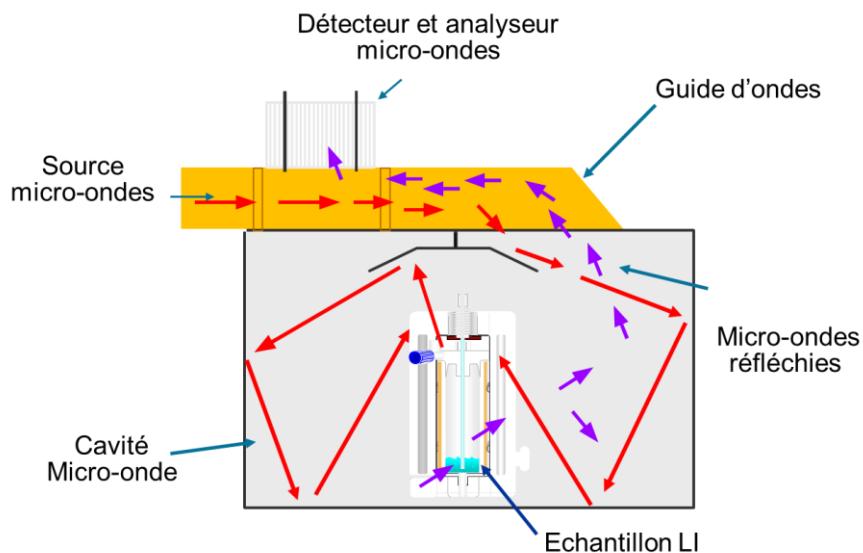


Figure 11 : Dispositif micro-ondes avec guide d'ondes et échantillon.

I.3.1.3 Facteurs critiques influençant la mesure de σ

La mesure précise de la conductivité électrique dans les LI est influencée par une multitude de facteurs expérimentaux et instrumentaux, dont la négligence peut introduire des erreurs significatives. Ces facteurs critiques, largement documentés dans la littérature, englobent des défis techniques, des biais méthodologiques et des limitations intrinsèques aux instruments.

1. Erreurs de calibration

La calibration constitue l'étape fondamentale pour garantir la fiabilité des conductimètres. Des normes de calibration inexactes ou dégradées (p. ex., solutions de KCl vieillissantes) induisent systématiquement des biais dans les mesures [27]. Une étude de McFarlane et al. [1] a montré que l'utilisation de solutions contaminées ou non certifiées pour étalonner les cellules de conductivité entraîne des erreurs relatives pouvant dépasser 5 % dans les LI hygroscopiques⁶. De plus, une procédure de calibration inappropriée modifie la constante de cellule, comme l'a souligné Horváth [28] dans son analyse des systèmes électrolytiques complexes.

⁶ La capacité d'un liquide ionique à absorber l'humidité présente dans l'air ambiant.

2. Dégradation et contamination des électrodes

Les électrodes sont vulnérables à l'encrassement, à la passivation ou à la corrosion, particulièrement dans les LI contenant des anions réactifs tels que $[\text{NTf}_2]^-$ ou $[\text{PF}_6]^-$ [1]. Une contamination par des résidus organiques ou des oxydes métalliques réduit la surface effective des électrodes, faussant la conductance mesurée [12]. Par ailleurs, un stockage inapproprié (p. ex., exposition à l'air libre) des électrodes en platine peut entraîner une oxydation superficielle, augmentant la résistance de contact [22].

3. Variabilité des conditions d'échantillon

La conductivité des LI est fortement dépendante de la température, avec des coefficients typiques de 2–3 % par $^{\circ}\text{C}$ [29]. Une compensation thermique inadéquate, notamment dans les instruments bas de gamme, génère des erreurs systématiques, comme observé par [30] pour les liquides ioniques à base d'imidazolium. De plus, l'inhomogénéité des échantillons, causée par une agitation insuffisante ou une contamination par l'humidité, modifie localement la concentration ionique, biaisant les mesures [31].

4. Limitations instrumentales et interférences

Les conductimètres basés sur des méthodes AC/DC low-cost présentent souvent une sensibilité réduite aux hautes conductivités ($> 20 \text{ mS/cm}$) en raison de phénomènes de polarisation non linéaire [32]. Les interférences électromagnétiques, courantes dans les environnements de laboratoire non blindés, perturbent les signaux de faible amplitude, comme l'ont rapporté. Enfin, une alimentation électrique instable affecte la précision des appareils portables, en particulier pour les mesures à haute résolution [32].

5. Biais opérationnels et analytiques

Des protocoles de mesure non standardisés, tels qu'une immersion variable des électrodes ou un temps de stabilisation écourté, introduisent une variabilité inter-utilisateur. Par exemple, Bonhôte et al. [33] ont démontré qu'une profondeur d'immersion inconstante dans les LI visqueux entraîne des écarts de $\pm 8\%$ sur la conductivité σ .

I.3.2 mesure de la conductivité thermique

La conductivité thermique (λ) des liquides ioniques joue un rôle central dans leur performance pour des applications énergétiques telles que les batteries, les supercondensateurs ou les systèmes de refroidissement. Cette propriété, qui reflète la

capacité du matériau à transférer l'énergie thermique sous un gradient de température, est influencée par des mécanismes complexes incluant la conduction ionique, les interactions moléculaires et les effets de convection ou de rayonnement [34]. Bien que de nombreuses méthodes a été développées pour mesurer λ , allant des approches classiques (fil chaud transitoire, cylindres concentriques ou plaques parallèles) aux techniques optiques avancées (diffusion de Rayleigh forcée, flash laser, spectroscopie par corrélation de photons, radiométrie photothermique, thermoréflectance transitoire, détection photothermique, interférométrie à cisaillement du front d'onde, flux de chaleur, calorimétrie CA, échange de chaleur photoacoustique et par rayonnement) [34,35], cette section se concentre sur celles adaptées aux spécificités des LI, en soulignant leurs avantages, limites et défis associés.

I.3.2.1 Méthodes Transitoires

1. Méthode de Fil Chaud Transitoire

Parmi les méthodes transitoires, la technique du fil chaud transitoire (Transient Hot-Wire, THW) se distingue comme référence pour sa précision ($\leq 1\%$) et son statut de méthode primaire⁷. Son principe repose sur l'application d'un courant électrique à un fil métallique mince immergé dans le liquide, permettant de mesurer l'évolution temporelle de la température (Figure 12). Cette méthode, validée par l'IUPAC pour des liquides de référence comme l'eau et le toluène, a été adaptée aux LI conducteurs grâce à l'isolation électrique du fil par des revêtements en oxyde métallique (ex. Ta_2O_5) ou des techniques de polarisation DC pour supprimer les courants parasites [34,36]. Cependant, sa mise en œuvre exige des corrections rigoureuses pour les effets de convection, de rayonnement à haute température [37].

⁷ Reconnues par des organismes comme le CIPM, elles atteignent la plus haute exactitude (incertitude $\leq 1\%$) et servent à établir des références métrologiques

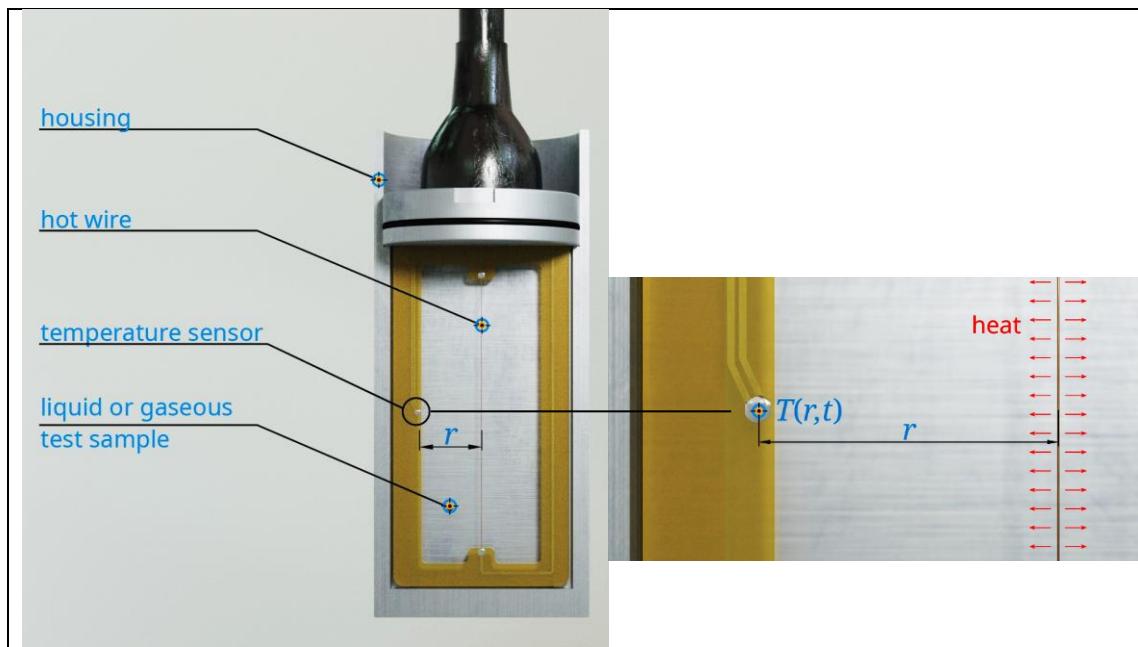


Figure 12 : Le principe de la méthode du fil chaud transitoire

2. Méthode de la Source Plane Transitoire

La méthode de la source plane transitoire (Transient Plane Source, TPS), ou Hot-Disk, repose sur l'utilisation d'une sonde circulaire qui joue simultanément le rôle d'élément chauffant et de capteur de température. Cette sonde est placée en contact direct avec l'échantillon à mesurer. Lors de l'expérience, un courant électrique pulsé est appliqué à la sonde, générant une dissipation de chaleur rapide qui se propage dans l'échantillon. La réponse thermique est ensuite enregistrée en temps réel, et la manière dont la température de la sonde augmente est directement liée aux propriétés thermiques du matériau environnant.

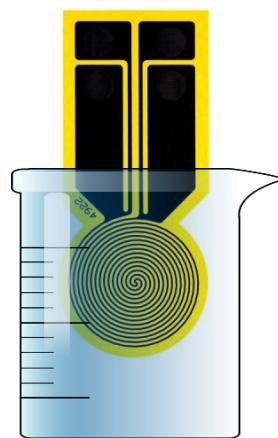


Figure 13: Schéma de principe de méthode de la Source Plane Transitoire

Cette méthode offre une alternative polyvalente pour mesurer simultanément λ et la diffusivité thermique. Bien que rapide et applicable aux solides et fluides, son incertitude plus élevée (3–5%) et sa sensibilité à l'hétérogénéité des nanofluides en restreignent la fiabilité pour les LI complexes [34,35].

I.3.2.2 Méthodes Stationnaires

Les techniques en régime permanent comprennent essentiellement deux types de géométrie : les plaques parallèles et les cylindres concentriques.

1. Plaques Parallèles

Cette méthode repose sur l'établissement d'un flux thermique unidirectionnel au sein de l'échantillon. Deux plaques métalliques parallèles encadrent le LI : une plaque chaude est chauffée par un élément chauffant interne, qui assure un gradient de température souhaité ; l'énergie est transférée à travers l'échantillon vers la plaque froide. Ce dispositif permet de calculer la conductivité thermique λ par l'équation suivante :

$$\lambda = \frac{\dot{q} \cdot d}{A \cdot \Delta T} \quad (13)$$

où \dot{q} représente le flux thermique par unité de longueur, d distance entre les 2 plaques, A la surface des plaques, et ΔT la différence de température entre les deux surfaces.

Cette méthode offre une précision de (± 2 –3 %) pour les échantillons stables et homogènes, grâce au contrôle rigoureux du flux thermique et à la suppression des pertes parasites [1,38,39]. Cependant, la présence de convection naturelle dans les LIIs peu visqueux perturbe le flux unidirectionnel, induisant des erreurs systématiques.

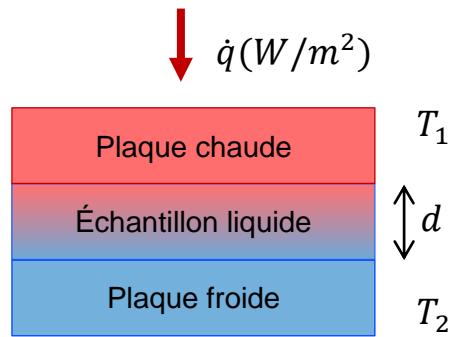


Figure 14 : Le principe de la Plaque Chaude Gardée

2. Cylindres Concentriques

Cette méthode utilise deux cylindres métalliques coaxiaux, séparés par un espace annulaire rempli du LI. Le cylindre intérieur chauffé, génère un flux thermique radial traversant l'échantillon vers le cylindre extérieur maintenu à une température contrôlée [40]. La conductivité thermique λ est déterminée via l'équation suivante :

$$\lambda = \frac{\dot{q} \cdot \ln(r_2/r_1)}{2\pi l \Delta T} \quad (14)$$

où \dot{q} représente le flux thermique par unité de longueur, r_1 et r_2 les rayons des cylindres intérieur et extérieur, l la longueur des cylindres, et ΔT la différence de température entre les deux surfaces.

Cette méthode offre une précision de ($\pm 2-3\%$) lorsque la géométrie est contrôlée par des mesures capacitatives, garantissant un écart constant entre les cylindres [34,35]. Elle a été validée pour les sels fondus (ex. NaNO_3 , KNO_3) et les LI hautement conducteurs, où elle offre une alternative fiable aux méthodes transitoires dans des conditions extrêmes [34,35].

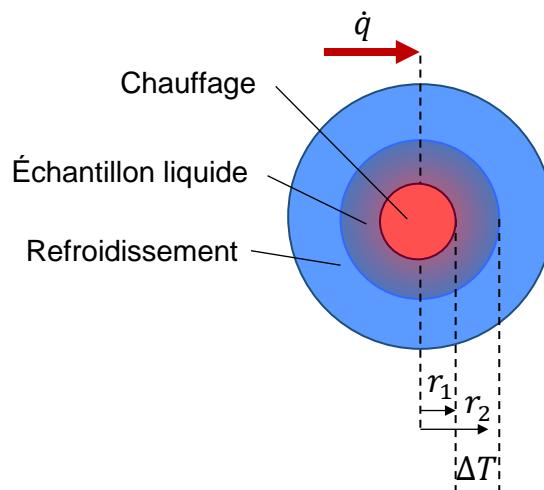


Figure 15 : Le principe de la Cylindres Concentriques

I.3.2.3 Autres Méthodes

Parmi les méthodes complémentaires, la **spectroscopie de diffusion Rayleigh forcée (FRS)** utilise un réseau laser pour mesurer la diffusivité thermique via la décroissance d'un motif de température, mais nécessite des colorants absorbants, limitant son usage [41,42]. La **spectroscopie de corrélation de photons (PCS)**, adaptée aux fluides transparents ($<200\text{ }^{\circ}\text{C}$), analyse les fluctuations de densité avec une précision de $\pm 2\text{--}3\text{ }%$, mais exclut les milieux opaques [43,44]. La méthode Laser Flash, basée sur une impulsion laser chauffant une face de l'échantillon, mesure la diffusivité thermique via la réponse thermique de la face opposée. Bien que rapide et adaptée aux hautes températures, elle est limitée par les effets de surface (ex. ondes de Marangoni dans les liquides) et nécessite des données précises de densité et capacité calorifique [35]. La **méthode 3ω** , basée sur un courant alternatif, évite la polarisation électrochimique dans les LI conducteurs, mais son incertitude ($\pm 5\text{ }%$) la réserve aux études préliminaires [35,45]. Ces approches, bien que spécialisées, répondent à des cas spécifiques (LI photosensibles, milieux confinés) où les méthodes conventionnelles échouent [46].

I.3.2.4 Facteurs critiques influençant la mesure de λ

Les LIIs présentent des défis particuliers lors de la mesure de leur conductivité thermique en raison de plusieurs facteurs :

La conductivité électrique élevée qui perturbe les mesures par court-circuit dans les méthodes utilisant des fils nus. Ce problème est contourné par l'isolation des sondes ou l'application d'un potentiel polarisant pour neutraliser les doubles couches électriques [34].

Les interactions ioniques induisent également un effet Soret en présence de gradients thermiques, modifiant localement la composition du liquide et faussant les résultats sans corrections appropriées [34].

La stabilité thermique limitée des LI (décomposition au-delà de 150 °C) et leur sensibilité aux impuretés (ex. traces d'eau) imposent un contrôle strict des conditions expérimentales et des matériaux de capteurs résistants à la corrosion, comme l'alumine (Al_2O_3) ou les revêtements céramiques [35].

I.3.3 Mesure de la viscosité

I.3.3.1 Viscomètres capillaires

Parmi les viscosimètres les plus répandus, on trouve les viscosimètres capillaires qui utilisent des tubes capillaires verticaux immergés dans un bain thermostaté pour évaluer la viscosité. Ces viscosimètres mesurent l'écoulement d'un fluide à travers un capillaire, nécessitant un fluide de référence calibré, généralement de l'eau, pour les déterminations de viscosité [47]. Bien que ces viscosimètres soient économiques, les tubes capillaires peuvent s'avérer difficiles à nettoyer, et la précision de mesure est compromise par l'humidité présente dans la colonne d'air du viscosimètre. Les incertitudes associées à ce type de viscosimètre proviennent principalement des fluctuations de température, des erreurs de lecture et de la taille du capillaire [47,48].

Traditionnellement utilisés pour les mesures de liquides, ces viscosimètres sont également adaptables pour la détermination de la viscosité des gaz. Il existe plusieurs types de viscosimètres capillaires, chacun conçu pour des applications spécifiques et offrant différents niveaux de précision et de sensibilité (Figure 16)[49]. Certains des types les plus importants comprennent le viscosimètre d'Ostwald, qui utilise un tube en U pour mesurer l'écoulement du fluide entre deux points ; le viscosimètre de Cannon-Fenske, qui est utilisé pour les liquides transparents et mesure l'écoulement à travers un tube capillaire ; le viscosimètre d'Ubbelohde, qui mesure l'écoulement à travers un tube capillaire à l'aide

d'une ampoule au fond ; et le viscosimètre de Houillon, utilisé pour mesurer la viscosité des liquides opaques. D'autres types comprennent les viscosimètres à niveau suspendu et les microviscosimètres, chacun ayant ses propres avantages et limitations en fonction de l'application et de la précision requise. Le choix du viscosimètre dépend des exigences spécifiques de l'application, telles que le type de fluide mesuré, la plage de viscosité et le niveau de précision souhaité.

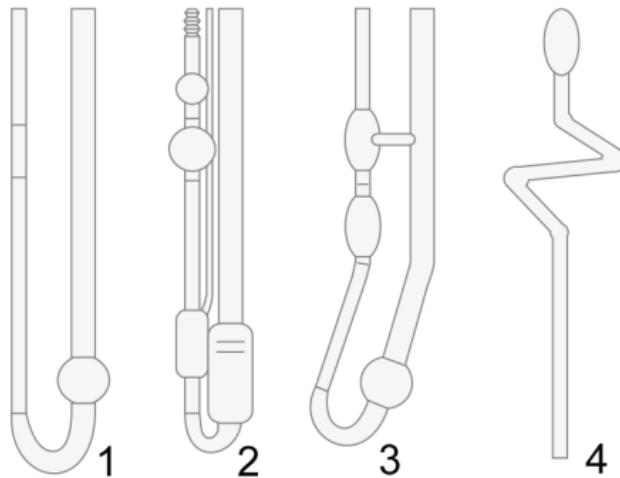


Figure 16: Quelques types de capillaires en verre : 1 : Capillaire d'Ostwald, 2 : Capillaire d'Ubbelohde, 3 : Capillaire de Cannon-Fenske, 4 : Capillaire de Houillon

I.3.3.2 viscomètres à billes ou à sphères

Le viscosimètre à bille ou à sphère en chute libre est un instrument largement utilisé pour mesurer la viscosité. Cette méthode consiste à chronométrier la chute libre d'un objet sphérique ou cylindrique à travers un échantillon de fluide sous l'influence de la gravité ou d'une force externe, telle qu'un piston. La vitesse de l'objet est directement liée à la viscosité du fluide [50,51]. Ces viscosimètres peuvent manipuler une large gamme de fluides, y compris les fluides non newtoniens, et peuvent être adaptés pour mesurer la viscosité de certains types de gaz [51]. Un défi majeur dans l'utilisation des viscosimètres à bille en chute libre est la mesure précise de la vitesse de l'objet. La plus grande source d'incertitude de mesure provient des erreurs de mesure de vitesse, qui peuvent être influencées par des facteurs tels que la viscosité du fluide, la température et la forme et la taille de l'objet en chute [51,52]. Des études ont montré que l'erreur dans la mesure de la viscosité avec ces instruments peut fluctuer de 4 % à des densités de fluide élevées à 6 % à des densités plus faibles [51,52]. L'entretien des viscosimètres à bille en chute libre peut

également être intensif, nécessitant un étalonnage régulier et une manipulation soigneuse. Pour les fluides viscoélastiques, le rapport de contrainte élastique doit être pris en compte, ce qui ajoute de la complexité au processus de mesure. Un étalonnage et un entretien minutieux sont essentiels pour garantir des mesures de viscosité précises et fiables.

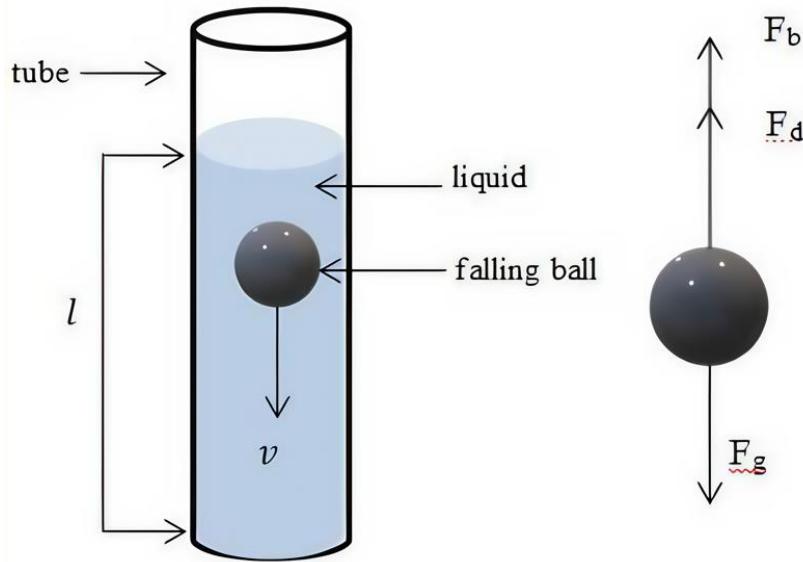


Figure 17: Schéma illustrant les forces agissant sur une bille en chute ; F_g force de gravité, F_b force de flottabilité ascendante et F_d force de friction visqueuse (interne).

I.3.3.3 viscomètres vibratoires ou oscillants

Les viscosimètres vibratoires ou oscillants sont une classe de viscosimètres qui fonctionnent sur le principe du mouvement oscillatoire. Ces appareils intègrent un élément vibratoire, tel qu'un fil, une lame ou un cristal, qui est immergé dans le fluide en cours d'examen. Le principe de leur fonctionnement repose sur l'analyse du comportement de ces éléments vibrants dans le fluide [53,54].

Dans le cas d'un viscosimètre à fil vibratoire, le concept opérationnel est basé sur l'étude de la réponse en fréquence aux oscillations transversales forcées d'un fil métallique. Ce fil, semblable à un diapason, est soumis à un champ magnétique (B) et immergé dans le milieu à étudier (voir Figure 18). Un courant électrique alternatif, étroitement aligné avec la fréquence de résonance mécanique du fil, induit un mouvement dans le fil. Ce mouvement génère une force électromotrice (force de Lorentz) aux bornes du fil, qui correspond à l'impulsion détectée. En balayant la fréquence électrique, on observe la

résonance. La fréquence de résonance, qui dépend de la densité du fluide, est mesurée expérimentalement, permettant ainsi la détermination de la viscosité [54].

Les viscosimètres à fil vibratoire ont été utilisés pour mesurer la viscosité d'une large gamme de fluides, sur une large gamme de températures et de pressions. Ces viscosimètres ont démontré leur capacité à mesurer des viscosités allant de 0,009 à 200 mPa·s dans les phases liquide et gazeuse [52,54,55]. Fonctionnant sur une plage de températures de 1,1 K [56] à 455 K [57] et à des pressions allant jusqu'à 1 GPa [58], les viscosimètres à fil vibratoire se sont révélés polyvalents dans leur application sur une large gamme de conditions expérimentales.

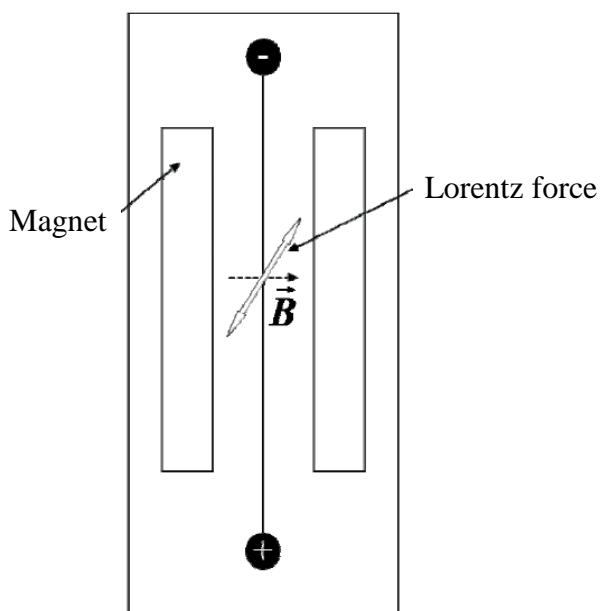


Figure 18: Schéma d'un viscosimètre à corde vibrante

Dans le cas d'un viscosimètre à cristal vibrant et d'un viscosimètre à plaque vibrante, ces deux types de viscosimètres à corps vibrant ont un principe de fonctionnement similaire [55]. Ce qui distingue cette technologie du viscosimètre à fil vibrant est sa remarquable polyvalence, lui permettant de fonctionner efficacement sur une vaste plage de viscosité allant de 0,01 à 105 mPa·s. Au cœur de son fonctionnement, le principe opérationnel repose sur le mouvement microscopique d'un cristal ou d'une tige métallique. Cela garantit que l'équilibre thermodynamique du fluide reste inchangé. Le corps vibrant comprend un composé piézoélectrique, généralement de forme cylindrique, traversé axialement par un courant électrique. Des électrodes intégrées dans le corps appliquent

une tension sinusoïdale, induisant des vibrations de torsion. Plongé dans le milieu étudié, le corps vibrant met en mouvement une onde de déformation, rapidement atténuée en raison de son déplacement perpendiculaire à la surface du cristal [51]. Les forces visqueuses exercées à cette interface modifient la fréquence de résonance du corps vibrant. La dérivation de la valeur de viscosité repose sur le même principe que celui du viscosimètre à fil vibrant.

Malgré l'efficacité de ces instruments, les viscosimètres vibratoires présentent généralement des limites [53,55]. Une limitation significative est le manque de champ de cisaillement défini dans ces instruments. Cela signifie que les forces de cisaillement agissant sur le fluide ne sont pas bien définies, les rendant inadaptés à la mesure de la viscosité des fluides dont les comportements d'écoulement sont inconnus. De plus, la précision de ces viscosimètres peut être affectée par des facteurs tels que les variations de température et la présence de particules dans le fluide.

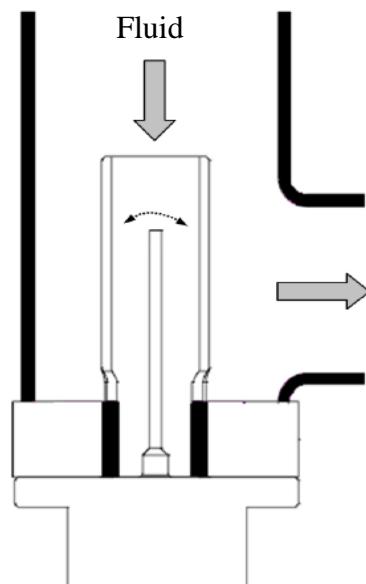


Figure 19: Schéma d'un viscosimètre à cristal vibrant et plaque vibrante

I.3.3.4 Viscomètres rotationnels

Les viscosimètres rotationnels figurent parmi les types de viscosimètres les plus répandus. Leur popularité découle de leurs structures de conception relativement simples et de leur principe de fonctionnement, qui consiste à mesurer la rotation d'un objet solide (tel qu'une bille, une pièce métallique moulée ou un axe cylindrique) lorsqu'il est immergé dans un

fluide visqueux sous l'application d'une force ou d'un couple connu nécessaire pour le faire tourner dans le fluide [59,60].

Il existe trois principales catégories de viscosimètres rotationnels : les viscosimètres à cylindre coaxial, communément appelés viscosimètres de Couette ; les viscosimètres à cône-plaque ; et les viscosimètres à plaque-plaque, également appelés viscosimètres à disque rotatif. Ces viscosimètres varient dans leurs éléments mobiles, qui peuvent être des cylindres, des cônes ou des disques (voir Figure 20). La mesure de la viscosité est intimement liée au couple appliqué à l'élément (G), à la vitesse angulaire de l'élément mobile (ω), au rayon extérieur de l'élément mobile (R_m), et au rayon intérieur de l'élément stationnaire (R_s).

Les viscosimètres rotationnels conviennent bien pour caractériser les milieux très visqueux, tels que les polymères fondus. Ils permettent également l'étude du comportement rhéologique de ces milieux grâce au contrôle du taux de cisaillement. Cependant, cette technique est moins adaptée pour les milieux à faible viscosité car le couple nécessaire à la rotation de l'élément mobile peut être insuffisant pour une mesure précise de la viscosité [59].

Ces instruments sont largement utilisés en raison de leurs structures de conception simples et de leur polyvalence. Cependant, leurs géométries complexes peuvent entraver la mesure efficace du comportement rhéologique de certains fluides, ce qui a un impact sur la précision des mesures de contrainte de cisaillement et peut entraîner une imprécision de mesure d'au moins 10 % [61]. De plus, les viscosimètres rotationnels nécessitent une immersion complète de leur élément rotatif dans le liquide d'essai, ce qui nécessite une quantité importante d'échantillon de test. Cela peut contribuer à un glissement efficace de particules plus importantes sur les parois de l'instrument, ce qui peut potentiellement affecter la précision des mesures de viscosité.

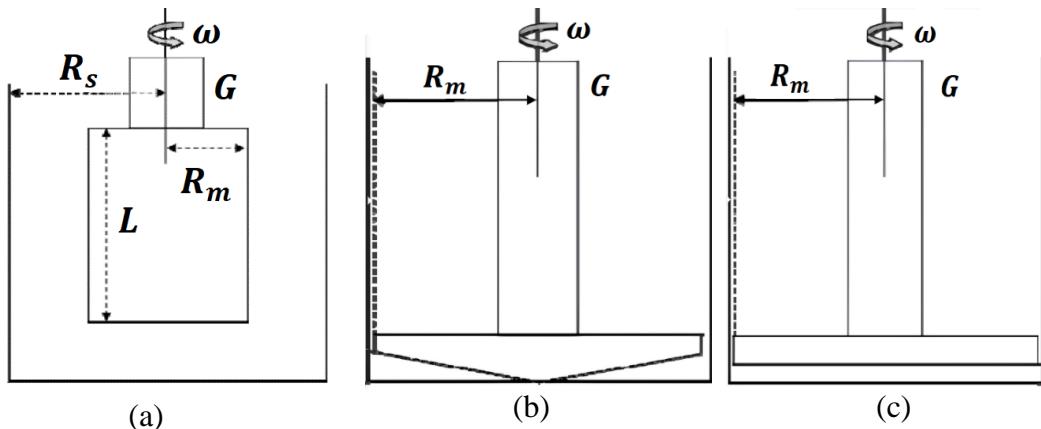


Figure 20: Types de viscosimètre rotatif ; (a) Viscosimètre à cylindre coaxial, (b) Viscosimètre à plan conique, (c) Viscosimètre plan

I.3.3.5 Viscomètres non conventionnels

Les viscosimètres non conventionnels regroupent toute une gamme de technologies innovantes qui s'écartent des méthodes de mesure traditionnelles. Ces viscosimètres offrent souvent des avantages uniques, tels que la capacité de mesurer la viscosité in situ, en temps réel, ou dans des fluides non newtoniens. Un exemple est la viscosimètre acoustique, qui mesure la viscosité en analysant la vitesse des ondes sonores se propageant à travers un fluide [62,63]. Cette technique est particulièrement utile pour les fluides opaques ou non homogènes et peut fournir des mesures dans des environnements difficiles [64]. Une autre approche non conventionnelle est la microrhéologie, qui mesure la viscosité à l'échelle microscopique en utilisant des techniques telles que le suivi de particules ou les pincettes optiques. Les microrhéomètres offrent une grande sensibilité et peuvent caractériser des fluides complexes, y compris ceux présentant des comportements rhéologiques non linéaires [65,66].

Les viscosimètres optiques représentent une autre catégorie de viscosimètre non conventionnelle, utilisant des méthodes basées sur la lumière telles que l'interférométrie ou la fluorescence pour mesurer la viscosité. Ces viscosimètres sont souvent non invasifs et peuvent fournir des informations sur la dynamique des fluides au niveau moléculaire [67]. Les viscosimètres basés sur les MEMS (systèmes micro-électro-mécaniques), exploitant les technologies micro fluidiques, offrent des avantages en termes de miniaturisation et d'intégration, permettant des mesures précises de la viscosité dans de petits volumes [68,69]. Les viscosimètres électromagnétiques, qui utilisent des champs électromagnétiques pour mesurer la viscosité, offrent une autre approche non

conventionnelle adaptée à une large gamme de fluides et de conditions [70,71]. Ces viscosimètres non conventionnels élargissent collectivement les possibilités de mesure de la viscosité, offrant des solutions innovantes pour diverses applications.

I.3.3.6 Facteurs critiques influençant la mesure de la viscosité

La mesure précise de la viscosité des liquides ioniques (LIs) est sensible à plusieurs facteurs expérimentaux et intrinsèques, qui doivent être rigoureusement contrôlés pour garantir la reproductibilité et la fiabilité des résultats. Ces facteurs incluent la température, l'hygroscopicité, le comportement rhéologique, la pureté des échantillons et les conditions de cisaillement.

1. Température

La viscosité des LIs suit une dépendance non linéaire à la température, modélisée par les équations d'Arrhenius ou de Vogel-Fulcher-Tamman. Une variation de ± 1 °C peut induire des écarts de 5–20 %, notamment pour les LIs à haute viscosité ($\eta > 100$ mPa·s). Un contrôle thermique précis (cellules Peltier, bains thermostatés) est indispensable [72].

2. Hygroscopicité et absorption d'humidité

De nombreux LIs, notamment ceux à base de cations fluorés (e.g., $[\text{BF}_4]^-$, $[\text{PF}_6]^-$), sont hygroscopiques. L'absorption d'eau modifie significativement leur viscosité en altérant les interactions ioniques et la structure du réseau [73]. Par exemple, une teneur en eau de 1 % en masse peut réduire la viscosité de $[\text{C}_4\text{mim}][\text{NTf}_2]$ de 30 % à 25 °C [74].

3. Rhéologie et cisaillement

Certains LIs présentent un comportement non newtonien (pseudoplastique), exigeant des rhéomètres à contraintes contrôlées pour des mesures fiables [75]. La viscosité doit être évaluée sur une gamme de taux de cisaillement ($0,1$ – 1000 s $^{-1}$), avec des géométries cône-plan pour éviter les artefacts [76].

4. Pureté des échantillons

Les impuretés (solvants résiduels, ions halogures, ou produits de dégradation) influencent la viscosité en perturbant les interactions cation-anion. Par exemple, la présence de 0,5 % de chlorure dans $[\text{C}_4\text{mim}][\text{Cl}]$ augmente sa viscosité de 15 % [31]. Une caractérisation préalable par RMN, ICP-OES ou chromatographie est essentielle [77].

I.4 REVUE DES METHODES DE PREDICTION DES PROPRIETES DE TRANSPORT

I.4.1 Conductivité Électrique

Dans cette section, nous présentons une revue des méthodes de prédition de la conductivité électrique des ILs, un domaine en pleine expansion compte tenu de l'intérêt croissant pour cette propriété. Les approches proposées dans la littérature se divisent principalement en trois catégories, chacune reposant sur des types d'entrées et des approches conceptuelles différentes : les méthodes théoriques, les approches fondées sur les relations quantitatives structure-propriété (QSPR) et les techniques basées sur les contributions de groupes.

I.4.1.1 Méthodes Théoriques

Les méthodes théoriques s'appuient sur des principes fondamentaux tels que la chimie quantique et les simulations de dynamique moléculaire pour prédire la conductivité électrique. Ces approches permettent de mieux comprendre les mécanismes qui régissent la conductivité dans les IL. Dans cette catégorie plus travaux ont été publiés, par exemple, une étude récente de dynamique moléculaire a révélé que, bien que la majorité des ions dans les ILs soient non conducteurs, environ 15 à 25 % existent en tant qu'ions libres capables de porter une charge - et cette fraction d'ions libres augmente avec la température. En outre, la conductivité du courant direct mesurée expérimentalement correspond bien aux prévisions théoriques basées sur l'équation de Nernst-Einstein, qui tient compte à la fois de la concentration d'ions libres et de leurs coefficients de diffusion [78] :

$$\sigma = \frac{F^2}{RT} \sum_i z_i^2 c_i D_i \quad (15)$$

Où F est la constante de Faraday, R est la constante des gaz, T est la température, z_i est le nombre de charge de l'ion i , c_i est sa concentration, et D_i est son coefficient de diffusion.

Eiden et al. [79] ont proposé une approche théorique pour prédire la conductivité électrique des ILs basée sur l'énergie de solvatation de Gibbs, calculée à l'aide de la

théorie de la fonctionnelle de la densité (DFT), le rayon moléculaire calculé à partir du volume moléculaire et le nombre de symétrie.

$$\ln\left(\frac{\sigma}{\sigma_0}\right) = k + l \ln(r_m^*) + m \ln(\sigma) + n \frac{\Delta G_{\text{solv},\infty}}{G_0} \quad (16)$$

Dans cette équation, r_m^* représente le rayon moléculaire calculé à partir du volume, σ est le nombre de symétrie moyen (défini par $\sigma = \sqrt{\sigma_+ \sigma_-}$, $\Delta G_{\text{solv},\infty}$ est l'énergie de solvatation de Gibbs calculée à une constante diélectrique infinie, et σ_0 ainsi que G_0 sont des constantes de normalisation. Les coefficients k , l , m et n sont des constantes déterminées par régression sur la base des données expérimentales.

L'étude a également exploré la dépendance en température des ILs en utilisant la relation Stokes-Einstein/Nernst-Einstein ou la règle de Walden. La méthode a été développée en utilisant 565 points de données pour 69 IL et a montré un meilleur R^2 de 0,91.

Wilenska et al. [80] ont présenté une méthode pour estimer la conductivité électrique des ILs basée sur leurs volumes moléculaires.

$$\sigma_{\text{calc}} = 3736.00109 \times \exp(-17.41385 V_m^{\text{cal}}) \quad (17)$$

Les auteurs ont employé la DFT pour calculer les volumes moléculaires et les ont corrélés avec des données expérimentales de conductivité à l'aide de l'équation de Vogel-Fulcher-Tammann. Cette méthode a été développée en utilisant seulement 37 points de données pour 19 IL et a abouti à une erreur relative absolue moyenne (ERAM) de 19 %. En général, le principal inconvénient de ces modèles basés sur la DFT est qu'ils requièrent une quantité significative de puissance de calcul et de ressources, ce qui peut limiter leur application pratique pour des systèmes de plus grande taille.

I.4.1.2 Relations Quantitatives Structure-Propriété (QSPR)

Le principe des méthodes de QSPR est d'établir une relation mathématique entre les propriétés moléculaires, englobant à la fois les aspects électroniques et géométriques (connues sous le nom de descripteurs), et la propriété recherchée.

Plusieurs modèles QSPR ont été proposés pour prédire la conductivité électrique des liquides ioniques en utilisant diverses représentations moléculaires et approches de régression [81–83], y compris récemment l'apprentissage automatique [84–88]. Tout récemment, en 2022, Abdullah et al. [84] ont proposé une approche QSPR combinant deux types différents de réseaux neuronaux pour prédire la conductivité électrique des LI. Dans ce travail, un réseau neuronal de type Graph Neural Network (GNN) a d'abord été utilisé pour générer des caractéristiques structurelles des anions et des cations. Ensuite, un perceptron multicouche (MLP) a été employé en séquence, en utilisant la température et 20 descripteurs moléculaires extraits du programme JRgui comme entrées. Les descripteurs moléculaires et structurels ont été obtenus à partir d'une base de données contenant 2684 points de données expérimentales couvrant 158 cations et 80 anions. Malheureusement, les auteurs n'ont pas fourni suffisamment d'informations sur les caractéristiques structurelles utilisées et, pire encore, ils n'ont pas présenté de données complémentaires concernant leur base de données. La même année, Baskin et al. [85] ont mené une vaste étude de benchmarking portant sur les méthodes d'apprentissage automatique basées sur QSPR et différentes représentations moléculaires pour prédire six propriétés physiques des LI, dont la conductivité électrique. Cette étude comprenait à la fois des méthodes d'apprentissage automatique traditionnelles et des architectures de réseaux neuronaux profonds, en s'appuyant sur cinq types de représentations moléculaires. Les résultats ont montré que la combinaison d'un réseau neuronal de type Transformer-convolutionnel et de SMILES comme représentation moléculaire surpassait les autres associations pour prédire la conductivité électrique. Cependant, cette meilleure combinaison présentait des performances relativement modestes, avec une EAM de 0,299 lg(S/m) et une valeur de R² de 0,594. Une limitation critique de cette étude réside dans l'utilisation de l'ensemble complet de la base de données ILThermo [89], qui comprend 7598 LI, sans aucun prétraitement pour éliminer doublons et valeurs aberrantes, ce qui peut réduire l'efficacité du modèle proposé. Par ailleurs, Dhakal et Shah [87] ont établi un autre modèle QSPR pour prédire la conductivité électrique des LI, basé uniquement sur un ensemble de 2869 points de données pour 397 LI couvrant 214 cations et 68 anions, enregistrées à la pression atmosphérique et sur une plage de température allant de 238 K à 472 K. Les auteurs ont utilisé 99 descripteurs générés avec le package RDKit et ont comparé trois modèles d'apprentissage automatique, à savoir la régression linéaire multiple (MLR), la forêt aléatoire (RF) et l'extrême gradient boosting (XGBoost). Le modèle XGBoost s'est révélé être le plus performant, bien qu'une dégradation de sa

capacité prédictive ait été observée pour les familles de cations représentées par une très petite part de l'ensemble de données. Le modèle XGBoost a ainsi obtenu une valeur de R^2 de 0,80 ainsi qu'une EAM de 0,14 lg(S/m) pour l'ensemble des données. De leur côté, Karakasidis et al. [90] présentent une double approche pour prédire la conductivité électrique des liquides ioniques (LI) en s'appuyant à la fois sur des techniques numériques d'apprentissage automatique et sur la régression symbolique. Les auteurs ont élaboré un modèle complet basé sur des données expérimentales, comprenant 2274 points extraits de la base de données NIST IL-Thermo. Ils se sont concentrés sur trois paramètres d'entrée clés, la température, la pression et la masse moléculaire, pour modéliser la conductivité électrique des LI. Dans l'approche numérique, plusieurs algorithmes d'apprentissage supervisé ont été implémentés, notamment la régression linéaire multiple (MLR), les k-plus proches voisins (KNN), les arbres de décision (DT), les forêts aléatoires (RF), le régressseur par gradient boosting (GBR) et le perceptron multicouche (MLP). Les résultats ont montré que le modèle MLR donnait un ajustement relativement médiocre en raison de la relation non linéaire entre la température et la conductivité, comme le suggèrent les modèles empiriques. En revanche, les méthodes basées sur les arbres (DT, RF, et en particulier GBR) ont obtenu d'excellentes performances, le modèle GBR atteignant une valeur de R^2 proche de 0,99, ce qui indique que les prédictions suivaient étroitement les données expérimentales. En complément des méthodes numériques, l'étude a utilisé la régression symbolique pour dériver des expressions analytiques décrivant le comportement de la conductivité. Cette approche analytique a généré plusieurs équations candidates, chacune présentant un degré de complexité variable, bien que ces modèles de régression symbolique n'aient pas atteint le même niveau de précision que les meilleurs modèles numériques.

Nous avons délibérément limité nos citations à ces références clés, afin de souligner les apports les plus significatifs et d'éviter toute surcharge d'informations. En effet, la littérature sur les méthodes QSPR appliquées aux liquides ioniques est vaste et en constante expansion. En privilégiant les travaux les plus représentatifs et méthodologiquement solides, nous visons à offrir au lecteur une synthèse ciblée et instructive.

I.1.1.A Méthodes de contributions de groupes

Les méthodes GC reposent sur le fractionnement du cation et de l'anion en fragments ou groupes moléculaires prédéfinis, chacun contribuant à la valeur globale de la propriété [13,14]. Cette approche présente l'avantage d'être simple, rapide et facile à mettre en œuvre, sans nécessiter de puissance de calcul importante ni de logiciel spécialisé [91–95]. Depuis les travaux pionniers de Matsuda et al. [96], Gardas et Coutinho [97,98] [22,23], Gharagheizi et al. [99] et d'autres, de nombreux efforts ont été consacrés au développement de modèles basés sur la contribution de groupe pour la conductivité électrique des LI [100–102]. Récemment, Chen et al. [101] ont proposé un modèle simple de contribution de groupe pour prédire la conductivité électrique des LI : un polynôme du second ordre dont les trois paramètres, spécifiques à chaque liquide ionique, sont estimés par GC. Ces paramètres de groupe (cation, anion et substituant) ont été calibrés sur 1 578 points de données expérimentales couvrant 77 LI. Un algorithme déterministe de régression linéaire multivariée a été appliqué, avec 1 121 points issus de 57 LI pour l'apprentissage et 457 points issus de 20 LI pour la validation, aboutissant à des ERAM de 3,36 % et 6,83 % en log-unités pour l'apprentissage et le test, respectivement. Cependant, les auteurs ont utilisé une base de données limitée par rapport à celle publiée en 2020, ce qui peut restreindre la portée et l'efficacité du modèle proposé.

L'an passé, Nakhaei-Kohani et al. [103] ont présenté deux approches, l'une basée sur la contribution de groupe et l'autre empirique, pour prédire la conductivité électrique des LI en fonction de la température, en intégrant quatre techniques de régression avancées relevant de l'apprentissage automatique : k-nearest neighbours (KNN), extreme gradient boosting (XGBoost), adaptive boosting-support vector regression (Ada-SVR) et adaptive boosting-decision tree (Ada-DT). Dans l'approche empirique, le poids moléculaire, la pression critique et la température critique ont été utilisés comme variables d'entrée. Les résultats montrent que, parmi les modèles basés sur la contribution de groupe, le modèle KNN est le plus précis pour la prédiction de la conductivité électrique, avec des valeurs de REQM et de R^2 respectivement de 0,1976 et 0,9935. Malheureusement, les auteurs ont utilisé une base de données restreinte de seulement 2 625 points couvrant 225 LI, sans fournir de données complémentaires.

Zhao et al. [100] ont développé le modèle UNIFAC-CONDUCT, une approche par contributions de groupe inspirée de leur précédent modèle UNIFAC-VISCO, pour prédire

la conductivité électrique des liquides ioniques purs. En s'appuyant sur un jeu de données de 784 mesures expérimentales de conductivité pour 38 LI différents, couvrant une large plage de températures (248–468 K) et de conductivités (0,002–14,54 S·m⁻¹), les auteurs ont établi trois jeux de paramètres distincts via l'optimisation de Marquardt du modèle suivant :

$$\ln \sigma = \sum_{i=1}^C x_i \ln \left(\sigma_i \cdot \frac{V_i}{V_m} \right) + \frac{g_C^E}{RT} - \frac{g_r^E}{RT} \quad (18)$$

$$\sigma_i = A_i \exp \left[\frac{B_i}{T - T_{0,i}} \right] \quad (19)$$

où, l'indice i représente le composant dans la LI pure (c-à-d le cation et l'anion) ; C est le nombre de composants dans la LI et est égal à 2 ; x_i est la fraction molaire de LI du cation/anion ($x_1 = x_2 = 0,5$) ; σ_i est la conductivité de l'ion et est exprimée à l'aide de l'équation VFT (l'équation ci-dessous) ; V_i est le volume molaire effectif de l'ion ; V_m est le volume molaire de l'IL pur ; g_C^E et g_r^E sont le terme combinatoire et le terme résiduel respectivement ; A_i , B_i et $T_{0,i}$ sont des paramètres ajustables déterminés pour chaque ion.

Leur méthode a affiché un écart moyen de seulement 2,3 %. De plus, l'étude a mis en évidence une excellente corrélation linéaire entre la conductivité calculée et l'inverse de la viscosité prédictive par le modèle UNIFAC-VISCO, confirmant ainsi la cohérence physique de l'approche via les équations de Nernst–Einstein et de Stokes–Einstein. Enfin, le modèle a été étendu avec succès à la prédiction de la conductivité d'un mélange binaire de LI avec une précision d'environ 4,1 %, soulignant son potentiel pour la conception pratique d'électrolytes.

Wu et al. [102] ont mis au point un modèle structure-basé de contribution de second ordre pour prédire la conductivité électrique des liquides ioniques purs sur de larges plages de température (238–484 K) et de conductivité (1,5×10⁻⁴–19,3 S·m⁻¹). S'appuyant sur une méta-analyse de 2 195 points de données issus de la littérature pour 188 LI, ils utilisent la forme Vogel–Fulcher–Tammann (VFT) :

$$\ln \sigma_{i,T} = A_i + \frac{B_i}{T - T_{0,i}} \quad (20)$$

$$A_i = A_0 + \sum_j n_{ij} W_j^A, B_i = B_0 + \sum_j n_{ij} W_j^B, T_{0,i} = T_0 + \sum_j n_{ij} W_j^{T_0} \quad (21)$$

où chaque paramètre dépendant du LI i est exprimé par contribution de groupe ; A_i , B_i et T_i , avec n_{ij} le nombre d'occurrences du groupe j dans le LI i et W les contributions de groupe . Grâce à une optimisation globale et à la validation sur un jeu de test indépendant (217 points, 11 LI), le modèle atteint des erreurs moyennes absolues en pourcentage de 6,12 % pour l'apprentissage et 5,10 % pour le test (6,02 % au total). Les auteurs montrent également que leur modèle reproduit fidèlement les grandes tendances : augmentation exponentielle de la conductivité avec la température, diminution de la conductivité pour des chaînes alkyles plus longues ou un volume molaire plus élevé, et effets de position des substituants dans les LI pyridiniques.

I.4.2 Conductivité Thermique

Les modèles de prédiction de la conductivité thermique des liquides ioniques (ILs) peuvent être classés en deux grandes catégories : les modèles semi-empiriques et les modèles empiriques.

I.4.2.1 Modèles Semi-Empiriques

Ces modèles combinent des bases théoriques avec des ajustements empiriques pour améliorer la précision des prédictions. Ils utilisent souvent des équations basées sur des relations physiques et chimiques bien établies, mais nécessitent des ajustements basés sur des données expérimentales.

Tomida et al. [105] ont présenté une corrélation dans la plage de température de 294 à 335 K et des pressions allant jusqu'à 20 MPa, en supposant que la masse molaire des ILs est deux fois leur valeur réelle. Par conséquent, cette corrélation proposée est seulement valide pour un nombre limité d'anions :

$$\log\left(\frac{M_w \lambda}{\mu}\right) = -0,004499 M_w + 1,9596 \quad (22)$$

où M_w , λ et μ désignent respectivement la masse molaire, la conductivité thermique et la viscosité des ILs. Pour prédire la conductivité thermique de 21 ILs, Shojaee et al. [109] ont obtenu une corrélation comme suit :

$$\lambda = \frac{x + y\left(\frac{M_w}{T}\right)^m + d\left(\frac{T_m}{T}\right)^n}{T} \quad (23)$$

où M_w , T , T_m et P sont respectivement la masse molaire, la température, le point de fusion et la pression. Les paramètres x , y , m , d et n du modèle ont été ajustés par un algorithme génétique (GA). Ils ont montré que l'erreur relative moyenne absolue (ERAM) pour les données d'entraînement est de 5,22% et pour les données de test est de 10,76%. Bien que ce modèle ait un MARE élevé, sa simplicité intrinsèque reste un avantage. La limitation la plus importante de ce modèle est l'inaccessibilité du point de fusion pour les ILs. Récemment, pour prédire la conductivité thermique de quatre ILs, une corrélation a été développée par Frez et al. [110] comme suit :

$$\lambda = x \cdot cp \cdot \rho \quad (24)$$

où x , cp et ρ désignent respectivement la diffusivité thermique, la capacité calorifique et la densité. Les liquides ioniques utilisés dans ce modèle comprennent : [EMIM][NTf2], [BMIM][BF4], [BMIM][NTf2] et [BMIM][PF6]. Ce modèle présente la limitation de nécessiter des paramètres comme la diffusivité thermique, la capacité calorifique et la densité, qui sont difficiles à obtenir [110].

I.4.2.1.A Relations quantitatives structure-propriété (QSPR)

Les modèles QSPR (Quantitative Structure-Property Relationship) sont des méthodes puissantes qui permettent de prédire les propriétés physiques et chimiques d'un composé, comme la conductivité thermique, en fonction de sa structure chimique. Ces modèles sont

particulièrement utiles dans le cas des liquides ioniques, car ils permettent de prédire les propriétés de substances dont l'expérimentation peut être coûteuse et complexe.

Lazzús[172] propose un modèle QSPR (relation quantitative structure-propriété) impliquant 11 descripteurs moléculaires pour estimer la conductivité thermique à différentes températures des liquides ioniques. Ce modèle a été construit à partir de 195 points de données expérimentales, couvrant une large plage de températures allant de 273 à 390 K. Les performances du modèle sont jugées excellentes, avec un écart absolu relatif moyen (ERAM) de 1,97 % et un coefficient de corrélation R^2 de 0,9919.

I.4.2.1.B Modèles de contribution de groupe

Les modèles de contribution de groupe pour la prédiction de la conductivité thermique des liquides ioniques font partie d'une approche empirique qui utilise les caractéristiques moléculaires des composants pour estimer leurs propriétés thermiques. Ces modèles sont basés sur l'idée que la conductivité thermique des liquides ioniques peut être prédictive en combinant les contributions des différents groupes fonctionnels présents dans la molécule.

Une corrélation a été proposée par Gardas et al. [104] pour prédire la conductivité thermique de 16 ILs avec 107 points de données dans la plage de température de 293 à 390 K, en dessous ou proche du point d'ébullition normal. Les résultats révèlent que la conductivité thermique diminue linéairement avec l'augmentation de la température selon:

$$\lambda(T) = X' - Y \cdot T \quad (25)$$

où X' et Y sont des constantes et T représente la température absolue. Le modèle linéaire a également montré une déviation relative absolue de moins de 1,06%.

Une relation a été suggérée par Tomida et al. [105] pour estimer les conductivités thermiques de trois ILs composés de [hmim][PF₆], [omim][PF₆] et [bmim][PF₆] à des pressions allant jusqu'à 20 MPa et des températures comprises entre 294 et 335 K. Il a été montré que les changements de la conductivité thermique par rapport à la pression peuvent être décrits par la relation suivante :

$$\lambda(T, P) = Z \cdot P + X' \quad (26)$$

où P représente la pression et X' , Z sont des constantes. Par conséquent, la combinaison des équations (25) (26) donne :

$$\lambda(T, P) = X - Y \cdot T + Z \cdot P \quad (27)$$

De plus, en considérant les cations et les anions dans les ILs, (27) peut être réécrite comme suit :

$$\lambda(T, P) = (x^+ - y^+ \cdot T + z^+ \cdot P) + (x^- - y^- \cdot T + z^- \cdot P) \quad (28)$$

où les composants ajustables des cations et des anions sont (x^+ , y^+ et z^+) et (x^- , y^- et z^-), respectivement. Cette méthode utilise la technique GCM pour prédire les paramètres des anions et des cations [106]. La technique GCM qui repose sur la régression multilinéaire [107] ajuste les données d'une propriété d'intérêt (y) en utilisant une combinaison linéaire d'un ou plusieurs descripteurs (x_j) (qui peuvent être obtenus par analyse de régression) :

$$y = \sum_{i=1, j=0} n_i x_j \quad (29)$$

où x_j est la contribution de chaque groupe et n_i est le nombre d'occurrences pour chaque groupe moléculaire.

Lazzús [106] a proposé une corrélation, Eq. (30), basée sur GCM et les équations mentionnées ci-dessus (Eq. (25) à (28)) pour estimer la conductivité thermique de 400 points de données expérimentales (41 ILs) dans les plages de température et de pression de 273–390 K et 100–20 000 kPa, respectivement. De plus, 300 points de données ont été sélectionnés comme ensembles de corrélation et 100 points de données comme ensembles de validation. En conséquence, la conductivité thermique pourrait être estimée par GCM pour plusieurs ILs avec une précision appropriée. Par exemple, la déviation relative

moyenne absolue (ERAM) pour les 300 points de données utilisés dans l'ensemble de corrélation est inférieure à 1,90%. Aussi, dans l'ensemble de validation, l'ERAM est d'environ 2,33% pour 100 données. Le GCM proposé par Lazzús est le suivant :

$$\begin{aligned} \lambda(T, P) = & (\sum_i m_i x_i^+ + \sum_j m_j x_j^-) - T \left| \sum_i m_i y_i^+ + \sum_i m_j y_j^- \right| \\ & + P \left| \sum_i m_i z_i^+ + \sum_j m_j z_j^- \right| \end{aligned} \quad (30)$$

où m_i et m_j désignent les groupes répétés i et j dans les IL, x^- , y^- , et z^- sont les contributions des groupes anions, et x^+ , y^+ , et z^+ désignent les contributions des groupes cations. Dans l'équation (30), l'unité de pression et de température est kPa et K, respectivement, où x^+ et x^- , y^+ et y^- , z^+ et z^- , et λ ont des unités de W/m·K, W/m²·K, W/kPa·m·K, et W/m·K, respectivement [106]. La conductivité thermique a été prédite par Tomida et al. [35] pour le tétrafluoroborate de 1-butyl-3-méthylimidazolium ([C4mim][BF₄]) dans la plage de pression entre 0,1 et 20 MPa et la plage de température de 294 à 334 K, avec une déviation maximale de 0,6% comme suit [18] :

$$\lambda_0 = 177 \times 10^{-3} \left(\frac{W}{m \cdot K} \right) - 2,5 \times 10^{-5} \left(\frac{W}{m \cdot K} \right) \cdot T \quad (31)$$

$$\frac{\lambda - \lambda_0}{\lambda} = 2,30 \times 10^{-2} \ln \left(\frac{11,4 \text{ MPa} + P}{11,4 \text{ MPa} + 0,01} \right) \quad (32)$$

Pour démêler les dépendances de la conductivité thermique et de la viscosité des ILs,

De plus, Fröba et al. [108] ont proposé une formule empirique générale pour estimer la conductivité thermique basée sur ρ et Mw qui désignent respectivement la densité et la masse molaire pour neuf ILs à base d'imidazolium et un IL à base de tétraalkylammonium, dans la plage de 273,15 à 353,15 K à pression ambiante :

$$\lambda \cdot MW \cdot \rho = 0,1130 \left(\frac{g}{cm^3 \cdot Wm^{-1} \cdot k^{-1}} \right) \cdot M_w + 22,65 \left(\frac{g^2}{cm^3 \cdot W \cdot m^{-1} \cdot k^{-1} \cdot mol^{-1}} \right) \quad (33)$$

Les résultats montrent que la déviation moyenne absolue en pourcentage dans ce travail était d'environ 7%. Fröba et al. ont développé une GCM pour dix ILs à 1 atm et 293,15 K comme suit :

$$\lambda \cdot \rho = D + \frac{V}{M_w} \quad (34)$$

où $D=0,1130 \left(\frac{gW}{cm^3 \cdot Wm^{-1} \cdot k^{-1}} \right)$ $V=22,65 \left(\frac{g^2}{cm^3 \cdot W \cdot m^{-1} \cdot k^{-1} \cdot mol^{-1}} \right)$ et ρ et Mw désignent respectivement la densité et la masse molaire des ILs. La corrélation développée par Fröba et al. a également été appliquée à d'autres ILs à différentes températures ; cependant, les résultats n'étaient pas satisfaisants, montrant une déviation moyenne absolue de 8,15%.

I.4.3 Viscosité

La nature complexe de la viscosité des liquides continue de défier les chercheurs, malgré de nombreuses études. La viscosité des liquides a fait l'objet de nombreuses recherches, mais à ce jour, aucune théorie adéquate n'est disponible pour reproduire de manière satisfaisante les données expérimentales et être universellement valide pour tous les liquides. Toutes les théories utilisées pour les gaz denses et les liquides sont empiriquement modifiées, nécessitant souvent des ajustements pour s'adapter aux données expérimentales, ce qui les rend essentiellement semi-théoriques dans le meilleur des cas. Par conséquent, les modèles utilisés pour la prédition et la corrélation des viscosités dans les fluides peuvent être largement classés en deux groupes principaux : les modèles semi-empiriques et empiriques.

I.4.3.1 Modèles semi-empiriques

Les modèles semi-théoriques, qui combinent des principes théoriques avec les tendances des données expérimentales, sont généralement considérés comme produisant des

résultats satisfaisants, surtout en comparaison avec les modèles purement théoriques qui impliquent souvent des erreurs significatives dans l'estimation des viscosités des liquides. Ces modèles sont particulièrement répandus dans l'étude des gaz denses et des liquides. Ils impliquent généralement l'association d'un modèle de viscosité avec une méthode de prédiction de la densité ou du volume. Ces approches reposent couramment sur des principes d'états correspondants et des modèles de mécanique statistique, incluant l'approche résiduelle, la théorie des sphères dures et du puits carré ou des modifications de ces théories. Ces modèles dépendent intrinsèquement de la température et de la densité (ou du volume). Cependant, étant donné que la viscosité est souvent requise à des conditions de pression et de température spécifiques, l'association d'une méthode de prédiction de la densité avec le modèle de viscosité devient nécessaire. La précision de ces modèles dans la prédiction de la viscosité des liquides dépend fortement de la précision de la valeur de densité fournie, ce qui souligne l'importance de valeurs de densité précises dans ces calculs.

I.1.1.A.1 Approches des états correspondants

Le théorème des états correspondants (EC) dans la modélisation de la viscosité est basée sur l'idée que la propriété réduite d'un fluide est équivalente à celle d'un fluide de référence aux mêmes conditions réduites. Initialement observé par van der Waals, ce principe implique que deux fluides présentent la même viscosité réduite à une température et une pression réduites identiques. Un exemple de modèle d'états correspondants est le modèle Ely et Hanley, également connu sous le nom de modèle de prédiction des propriétés de transport (TRAPP), qui utilise le méthane comme fluide de référence. Le modèle relie la viscosité d'un fluide d'intérêt à celle du méthane en utilisant les propriétés critiques, la masse moléculaire et la densité. Il est représenté comme suit :

$$\eta_i(\rho, T) = \eta_0(\rho_0, T_0)(M_i/M_0)^{1/2} h_{i,0}^{-2/3} f_{i,0}^{1/2} \quad (35)$$

où η_i est la viscosité du fluide d'intérêt, ρ est sa densité, et T est la température. M_i et M_0 sont les masses moléculaires du fluide d'intérêt et du méthane, respectivement, tandis que $h_{i,0}^{-2/3}$ et $f_{i,0}^{1/2}$ sont des facteurs de forme déterminés empiriquement. Alors que le modèle TRAPP montre de bonnes performances pour les hydrocarbures non polaires droits, il éprouve des difficultés avec les hydrocarbures ramifiés, les hydrocarbures cycliques et

les fluides polaires, ce qui a conduit à des modifications par les chercheurs ultérieurs. Un autre modèle d'états correspondants est le modèle de Pedersen et Fredenslund, qui lie la viscosité du fluide d'intérêt à celle du méthane en fonction de la pression et de la température[119].

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:

$$\eta_x(P, T) = \left(\frac{T_{c,x}}{T_{c,0}}\right)^{-1/6} \left(\frac{P_{c,x}}{P_{c,0}}\right)^{2/3} \left(\frac{M_x}{M_0}\right)^{1/2} \left(\frac{\alpha_x}{\alpha_0}\right) \eta_0(P^*, T^*) \quad (36)$$

où $T_{c,x}$ et $T_{c,0}$ sont les températures critiques du fluide d'intérêt et du fluide de référence, $P_{c,x}$ et $P_{c,0}$ sont les pressions critiques, M_x et M_0 sont les masses moléculaires, et α_x et α_0 sont les coefficients de couplage rotationnel de Tham-Gubbins pour le fluide d'intérêt et le fluide de référence, respectivement. La viscosité du fluide de référence est désignée par $\eta_0(P^*, T^*)$ où P^* et T^* sont la pression et la température réduites, respectivement. Ce modèle montre une bonne corrélation pour la plupart des viscosités de composants purs et de mélanges, mais présente des écarts plus importants pour les alcanes lourds et les mélanges binaires avec des tailles de molécules significativement différentes. Des extensions pour améliorer les prédictions à des températures réduites inférieures à 0,47 ont été proposées par Pedersen et Fredenslund[120] .

De plus, le modèle de Teja et Rice [121] étend le principe des états correspondants généralisés aux viscosités de mélanges liquides, en utilisant deux fluides de référence non sphériques et un paramètre de réduction. L'équation du modèle est donnée par:

$$\ln(\eta\xi) = \ln(\eta\xi)^{r_1} + \frac{(\omega - \omega^{r_1})}{(\omega^{r_2} - \omega^{r_1})} [(\ln \eta\xi)^{r_2} - (\ln \eta\xi)^{r_1}] \quad (37)$$

Ici, les exposants r_1 et r_2 font référence à deux fluides de référence non sphériques similaires aux composants purs ou aux composants clés d'intérêt dans les mélanges. ω représente le facteur acentrique de Pitzer, et ξ est le paramètre de réduction donné par $\xi = \frac{V_c^{2/3}}{(MT_c)^{1/2}}$. Teja et al. ont reformulé la méthode Teja-Rice en termes de pression critique

(P_c) plutôt que de volume critique (V_c) , tenant compte de l'incertitude souvent associée à la mesure des volumes critiques. Le paramètre de réduction révisé est exprimé comme $\xi = T_c^{1/6} / (M^{1/2} P_c^{2/3})$. Cependant, une limitation notable de ce modèle est l'exigence de fluides de référence qui ressemblent étroitement au fluide d'intérêt, ce qui rend difficile son obtention.

Les données de viscosité pour ces fluides de référence sont facilement disponibles. Par conséquent, une interpolation ou une extrapolation peut être nécessaire pour obtenir ces données. Aasberg-Petersen et al.[122] ont ensuite affiné l'approche Teja-Rice en utilisant deux fluides de référence (méthane et décane) pour prédire les viscosités des hydrocarbures, démontrant des prédictions raisonnables pour les fractions de pétrole léger mais une précision moindre pour les fractions lourdes. Contrairement au modèle Teja-Rice, qui utilise le facteur acentrique comme paramètre d'interpolation, ce modèle utilise la masse moléculaire, car le facteur acentrique diminue avec l'augmentation du poids moléculaire pour les fractions d'huile lourde. Des extensions et des modifications de ces modèles continuent d'être développées pour améliorer leur applicabilité et leur précision dans un plus large éventail de conditions et de types de fluides.

Plusieurs approches basées sur la théorie des champs (FT) ont été utilisées pour prédire la viscosité des liquides ioniques. Par exemple, Huang et al.[123] ont développé une méthode basée sur la méthode des contributions de groupe et le principe des états correspondants pour prédire les propriétés critiques des liquides ioniques (LI). Ils ont classé 46 fragments spécifiquement pour les LI, en tenant compte de leurs caractéristiques ioniques, et ont déterminé les incrémentations de fragments correspondantes en utilisant des données de densité expérimentales. Les résultats indiquent que cette méthode est fiable, avec un écart relatif absolu moyen de moins de 4%.

Zhao et al.[124] ont proposé deux modèles pour prédire la viscosité des liquides ioniques à base d'imidazolium (LI). Le premier modèle intègre la méthode des contributions de fragments-états correspondants (FC-CS) avec un algorithme de régression linéaire multiple (MLR), tandis que le deuxième modèle utilise un algorithme de machine à vecteurs de support (SVM). Dans la méthode FC-CS, le volume critique pseudo et le facteur de compressibilité (V_c et Z_c), ainsi que la température de point d'ébullition (T_b), sont calculés et utilisés pour prédire la viscosité avec les algorithmes MLR et SVM. Ils ont validé les modèles en utilisant un grand ensemble de données de 1079 points de

données expérimentaux pour 45 LI à base d'imidazolium, couvrant une large gamme de pressions et de températures. L'écart relatif absolu moyen (ERAM) pour l'ensemble des données était de 24,2% pour MLR et de 3,95% pour SVM. Le modèle non linéaire développé par l'algorithme SVM a surpassé le modèle linéaire créé par MLR, ce qui indique que l'algorithme SVM est plus fiable pour prédire la viscosité des LI à base d'imidazolium.

I.1.1.A.2 Approche résiduelle

L'approche résiduelle dans les calculs de viscosité exprime la viscosité d'un fluide comme une propriété résiduelle, $\Delta\eta$, par rapport à la limite de densité nulle ou à la viscosité des gaz dilués η_0 , où $\eta - \eta_0 = \Delta\eta$. Cette approche est particulièrement utile pour les liquides et les fluides supercritiques, où le terme des gaz dilués est négligeable. Le terme résiduel $\Delta\eta$ domine pour les viscosités des liquides dans ces cas. Diverses méthodes ont été proposées dans la littérature pour calculer ce terme, avec trois modèles principaux étant la Théorie du Frottement (FT), la Théorie du Volume Libre (FVT), et la Théorie du Fluide Étendu (EFT).

❖ Théorie du frottement (FT)

Quiñones-Cisneros et al.[125] ont développé le modèle FT, reliant la viscosité résiduelle à une prédition de l'équation d'état (EoS) des contributions de pression répulsive et attractive. Ce modèle, basé sur des concepts de mécanique classique, a été testé avec succès sur des n-alcanes et leurs mélanges sur une large gamme de conditions.

Plusieurs approches basées sur la FT ont été utilisées pour prédire la viscosité des liquides ioniques. Shen et al.[126] ont été parmi les premiers à appliquer cette théorie du frottement aux liquides ioniques, en particulier [Cnmim][BF4], [Cnmim][PF6], et [Cnmim][NTf2], en utilisant ePC-SAFT pour calculer les différentes contributions de pression. Abolala et al.[127] ont également appliqué une approche similaire, en substituant la théorie du frottement par la théorie du volume libre, et ont estimé les pressions répulsives et attractives à partir de l'équation d'état SAFT. Ils ont constaté que les forces répulsives à courte portée l'emportent sur les forces attractives à longue portée à haute pression, conduisant à un modèle simplifié avec moins de paramètres ajustables.

Haghbakhsh et al.[128] ont obtenu une meilleure corrélation avec l'équation PC-SAFT par rapport à Abolala et al., avec un modèle qui incluait le terme d'attraction du deuxième

ordre et une dépendance de température plus complexe pour les coefficients de frottement. Macias-Salinas[129] a développé un modèle de viscosité basé sur la théorie du frottement en parallèle avec la théorie d'Eyring, négligeant le terme d'attraction du deuxième ordre et simplifiant les expressions pour les coefficients de frottement pour réduire le nombre de paramètres ajustables à six. Ces modèles ont été testés sur divers liquides ioniques et comparés à d'autres modèles dans la littérature, montrant des performances globalement comparables, avec une certaine sous-estimation de la viscosité à haute pression et/ou basse température. La difficulté dans l'optimisation des paramètres est notée, attribuée à la corrélation entre les paramètres et leur dégénérescence.

❖ Théorie des volumes libres (FVT)

Dérivée de l'idée selon laquelle la viscosité dépend de l'espace libre dans un fluide, la FVT a été utilisée par Batschinski[130], Cohen et Turnbull[131], et Doolittle[132]. Malgré le besoin d'ajustement des données expérimentales, elle a montré une bonne précision prédictive pour divers fluides, notamment lorsqu'elle est utilisée avec des écarts absolus moyens inférieurs à environ 10 % sur de vastes plages de température et de pression.

Plusieurs approches basées sur la FVT ont été utilisées pour prédire la viscosité des liquides ioniques. Mac-Dowell et al. ont été parmi les premiers à utiliser cette approche pour les liquides ioniques, en particulier [Cnmim]Cl et [Cnmim][CH₃SO₄], en utilisant l'équation d'état soft-SAFT pour calculer la densité. Ils ont obtenu des résultats prometteurs avec un Écart Relatif Absolu Moyen (ERAM) de 8,48 % pour [Cnmim]Cl et 2,97 % pour [Cnmim][CH₃SO₄]. Polishuk et Yitzhak ont appliqué une méthode similaire en utilisant l'équation d'état SAFT+cubic pour les composés organiques, y compris [Cnmim][BF₄] ou [Cnmim][PF₆], mais ont rencontré de grandes déviations, ce qui les a amenés à explorer la Corrélation Modifiée Yarranton-Satyro, qui a donné des résultats nettement meilleurs.

Llovell et Vega[133] ont adapté un modèle précédemment développé pour les n-alcanes aux liquides ioniques [Cnmim][BF₄], mais ont rencontré des défis dans l'optimisation des paramètres lors de l'utilisation des données de viscosité à pression atmosphérique sur une large plage de température. Shen et al.[126] ont affiné la stratégie d'optimisation des paramètres du modèle FVT pour les liquides ioniques [Cnmim][BF₄], [Cnmim][PF₆], et

[Cnmim][NTf2], en linéarisant les paramètres par rapport au poids moléculaire du cation pour réduire le nombre total de paramètres.

Dans l'ensemble, ces méthodes démontrent l'application de la FVT et de divers modèles d'équation d'état pour prédire la viscosité des liquides ioniques, mettant en évidence l'importance de l'optimisation des paramètres et de la sélection du modèle pour des prédictions précises

❖ Théorie des fluides élargie (EFT)

Yarranton et Satyro [134] ont proposé un modèle EFT corrélant la viscosité résiduelle à la densité basé sur le principe de fluidité de Hildebrand. Ce modèle considère l'augmentation de la viscosité avec l'expansion du fluide due à l'augmentation de la distance entre les molécules de fluide. Mis en œuvre avec succès dans des simulateurs de processus commerciaux, il fournit un outil pratique pour prédire la viscosité des hydrocarbures et de leurs mélanges[135][136].

❖ Méthodes de mise à l'échelle

La mise à l'échelle de l'entropie excédentaire, telle que proposée par Rosenfeld[137], établit une relation entre la viscosité et l'entropie résiduelle pour les fluides atomiques simples, dérivée de simulations de dynamique moléculaire. D'autre part, la mise à l'échelle thermodynamique, introduite par Ashurst et Hoover[138], relie la viscosité à une seule variable généralisée comprenant la densité, la température et l'expansion du volume. Bien que ces méthodes aient montré leur succès dans la prédiction de la viscosité pour une gamme de fluides, elles peuvent ne pas être adaptées aux liquides à liaison hydrogène.

I.1.1.A.3 Théorie de la sphère dure (théorie Enskog modifiée)

Enskog a été le premier à développer une théorie cinétique pour prédire l'effet de la pression sur la viscosité des gaz, basée sur l'équation de Boltzmann et l'hypothèse selon laquelle les gaz denses sont constitués de molécules sphériques dures. Cette théorie[139], communément appelée théorie de Chapman-Enskog, suppose que les collisions moléculaires sont impulsionales. La viscosité dans ce modèle est donnée par une expression approximative impliquant la viscosité du gaz dilué η_0 , radiale.

la fonction de distribution au contact (Ψ), la densité molaire (ρ), et le volume occupé par les sphères dures (b).

$$\frac{\eta}{\eta_0} = \left[\frac{1}{\Psi} + 0.8b\rho + 0.761\Psi(b\rho)^2 \right] \quad (38)$$

Alder et al.[140] ont ensuite montré que des écarts significatifs peuvent survenir entre la théorie d'Enskog et les calculs de dynamique moléculaire pour un fluide de sphères dures, surtout dans la région de haute densité. Des modifications du modèle de sphères dures ont été proposées pour tenir compte de la haute densité, comme la théorie des sphères dures rugueuses de Chandler[141], qui prend en compte le couplage entre les mouvements de translation et de rotation des molécules.

Dymond [142] a proposé des corrections à la théorie d'Enskog pour inclure les forces attractives des molécules réelles, tandis que Chung et al. [143] ont étendu la technique de Chung-Lee-Starling pour inclure les fluides denses[144], incorporant une fonction empiriquement corrélée de densité et de température. La méthode proposée par Chung-Ajlan-Lee-Starling (CALS)[144] a été testée sur des fluides purs, montrant une précision prometteuse. D'autres approches basées sur la théorie des sphères dures incluent celles d'Assael et al[145], Bleazard et Teja[146], Ciotta et al[147], et Riesco et Vesovic[148]. Des extensions de ces modèles aux mélanges ont également été présentées, la méthode Vesovic-Wakeham (VW) [149] étant une manière fiable de prédire la viscosité des mélanges liquides. Récemment, de Wijn et al[150]. ont étendu le modèle VW aux mélanges liquides constitués de longues molécules en chaîne en utilisant l'approche de la théorie des fluides associatifs statistiques (SAFT), qui n'a aucun paramètre ajustable dans ses règles de mélange thermodynamiquement cohérentes. Cette approche a été appliquée et ajustée pour prédire la viscosité des liquides ioniques. Initialement développée par Assael et al[57]. au début des années 1990 pour divers composés organiques, la procédure de corrélation a ensuite été testée par Gaciño et al[151]. en 2014 pour 19 liquides ioniques purs de types imidazolium, pyrrolidinium et phosphonium. La relation entre le volume molaire caractéristique V_0 et la température est décrite par un modèle à 4 paramètres.

Hossain et Teja [152] ont suivi une approche similaire pour 48 liquides ioniques sur de larges plages de température et de pression, en optimisant deux ensembles de paramètres basés sur des données de viscosité à pression atmosphérique et à haute pression. Les deux

ensembles de paramètres ont montré une bonne extrapolation à haute pression. Hosseini et al.[153] ont utilisé une méthode différente, calculant la viscosité d'un fluide de sphères dures lisses. Ils ont obtenu une corrélation linéaire entre les valeurs des deux paramètres V_0 et R_η et la masse molaire pour une famille de liquides ioniques avec un anion commun. Hosseini et al.[153] ont étendu leur méthode à un mélange binaire de sphères dures en utilisant des lois de mélange. Cette approche n'a été testée que pour le mélange binaire de liquides ioniques [C4mim][PF6]⁺ [C4mim][BF4] et a montré de bonnes capacités prédictives. Cependant, il y avait une erreur dans le manuscrit de Hosseini et al.[154] et un erratum n'a été publié que un an après la publication de l'article de Akbari et al[155].

I.1.1.A.4 Modèles à puits carrés

Davis et al.[156] ont dérivé une expression pour la viscosité des fluides denses de molécules interagissant avec un potentiel de puits carré, qui tient compte à la fois de la répulsion et de l'attraction. L'équation dérivée est complexe et comprend des paramètres tels que la viscosité dynamique (η_{SW}), la viscosité du gaz dilué (η_0), la densité du fluide (ρ), le diamètre de répulsion (σ_1), la profondeur du puits énergétique (ε), la fonction de distribution radiale (g), le paramètre de largeur de puits carré (R), le covolume (b), la constante de Boltzmann (k), et la température (T). Ils ont également introduit ψ_1 et ψ_2 en tant que fonctions liées à la profondeur du puits énergétique et à la température.

McLaughlin et Davis [157] ont étendu ce modèle aux mélanges binaires, tandis que Du et Guo [158] ont proposé une modification semi-théorique du modèle de puits carré et ont dérivé une expression empirique pour la fonction de distribution radiale. Monnery et al[159] ont développé un modèle semi-théorique pour estimer les viscosités en phase liquide et gazeuse basées sur la théorie du puits carré. Galliero et al[160]. ont utilisé des simulations de dynamique moléculaire pour prédire la viscosité des fluides Lennard-Jones, révélant que le modèle Lennard-Jones de base tend à sous-estimer la viscosité, surtout à des températures plus basses.

Récemment, Zerón et al[161] ont adapté ce modèle en utilisant une combinaison de simulation de dynamique moléculaire et de la version continue du potentiel intermoléculaire de puits carré (CSW). Cette étude a validé l'utilisation du potentiel intermoléculaire CSW dans les simulations de dynamique moléculaire pour émuler avec précision les propriétés de transport SW.

I.1.1.A.5 Relations quantitatives structure-propriété (QSPR)

La Relation Quantitative Structure-Property (QSPR) est une méthode qui prédit les propriétés des molécules en fonction de leur structure. Cette méthode trouve son origine dans la Relation Quantitative Structure-Activity (QSAR), qui était initialement utilisée pour prédire l'activité biologique des molécules pharmaceutiques. Les modèles QSPR varient en complexité en fonction du type de descripteurs utilisés, qui peuvent être constitutionnels, topologiques, géométriques ou basés sur des calculs de mécanique quantique. Les descripteurs constitutionnels sont les plus simples, dérivés de la formule empirique d'une substance, tels que le nombre de donneurs de liaison hydrogène. Les descripteurs topologiques décrivent la forme et le degré de ramifications d'une molécule en utilisant sa structure 2D. Les descripteurs géométriques, calculés à partir de la structure 3D, comprennent des propriétés telles que le volume de van der Waals et la surface. Les descripteurs quantiques, qui nécessitent des calculs de structure électronique complexes, sont utilisés pour prédire diverses propriétés chimiques ou biologiques. Ces descripteurs, tels que l'énergie de l'Orbital Moléculaire Occupé le Plus Élevé ou le moment dipolaire, sont dérivés de la structure électronique d'une molécule. Le calcul de ces descripteurs moléculaires, généralement effectué avec des logiciels commerciaux tels que CODESSA et DRAGON, et la sélection des plus significatifs d'entre eux sont donc l'un des principaux inconvénients de l'approche[162]. La viscosité, une propriété de transport, est particulièrement difficile à prédire à l'aide de QSPR[163]. Divers modèles QSPR utilisant des descripteurs moléculaires ont été développés pour prédire les viscosités des liquides. Initialement, les prédictions étaient basées sur les viscosités à une température fixe ou sur d'autres propriétés telles que la pression de vapeur. Ivanciu et al. ont créé un modèle de Régression Linéaire Multiple à cinq paramètres à partir de 337 composés organiques, avec une erreur standard logarithmique de 0,371. Ce modèle prédit la viscosité en utilisant l'équation suivante :

$$\begin{aligned} \ln \eta = & -2.814 + 3.387C + 8.858 \times 10^{-3}MW \\ & + 3.919 \times 10^{-1}O - 8.486 T + 6.684 \times 10^{-1}Y \end{aligned} \quad (39)$$

Ici, C représente la surface chargée de donneurs de liaison hydrogène (ordre 2), MW est le poids moléculaire, O est l'indice de Randic (ordre 3), T est l'indice de réactivité électrophile maximal pour un atome de C , et Y est la population électronique orbitale

atomique maximale. En utilisant 361 composés organiques, un autre modèle de Régression Linéaire Multiple a été proposé par Katritzky et al[164]. avec une erreur standard de 0,22 (unités logarithmiques). Ce modèle prédit la viscosité ($\ln\eta$) en utilisant l'équation :

$$\begin{aligned}\ln\eta = & -10.3 + 1.77\text{HDCA}(2) + 0.0000557G_I \\ & + 2.78N_{\text{rings}} + 20.2\text{FPSA}(3) + 0.0897E_{\min}(\text{C})\end{aligned}\quad (40)$$

Ici, $\text{HDCA}(2)$ est la surface chargée de donneurs de liaison hydrogène (ordre 2), G_I est l'indice gravitationnel, N_{rings} est le nombre relatif d'anneaux dans la molécule, $\text{FPSA}(3)$ est la surface chargée positive fractionnaire (ordre 3), et $E_{\min}(\text{C})$ est l'énergie d'état atomique minimale pour un atome de C .

Différents modèles ont été développés pour prédire la viscosité des liquides ioniques en fonction de la température[165]. Généralement, les modèles utilisent la température comme descripteur, mais cela ne fournit pas toujours une compréhension claire de son impact sur la viscosité. Lors du classement des descripteurs en fonction de leur influence, la température émerge souvent comme l'un des facteurs les plus significatifs. Par exemple, dans leur modèle, Tochigi et Yamamoto [113] ont utilisé une température de référence (la température à laquelle la viscosité expérimentale est mesurée) comme descripteur pour prédire la viscosité des liquides ioniques. Ils ont ensuite démontré comment la viscosité prédictive varie en fonction de cette température de référence. Barycki et son équipe [166] ont formulé une approche pour prédire la viscosité des liquides ioniques à une température spécifique en se basant sur la viscosité à 25°C, déterminée à l'aide d'un modèle QSPR qu'ils ont développé. De plus, ils ont pris en compte la densité calculée par des simulations de dynamique moléculaire. Cette approche repose sur la compréhension selon laquelle la viscosité et la densité sont interconnectées par le concept de volume libre. Les chercheurs Bini et ses collègues[167] ont développé deux corrélations distinctes en utilisant différents descripteurs pour les températures élevées (353 K) et basses (293 K). L'erreur plus importante à basse température suggère que le modèle pourrait ne pas capturer pleinement le mécanisme de transport à ces températures. Dans une étude similaire, Yu et al[168]. ont examiné l'importance des différentes interactions intermoléculaires à huit températures différentes. Ils ont constaté que les interactions électrostatiques affectent la viscosité sur l'ensemble de la plage de température étudiée (de 283 à 343 K). À des

températures plus basses, les interactions de Van der Waals et les liaisons hydrogène sont dominantes, tandis que les effets stériques deviennent plus significatifs à des températures plus élevées. Cela indique que les effets enthalpiques ont un plus grand impact sur la viscosité à basse température, et que les effets entropiques augmentent avec la température, ce qui est conforme aux principes thermodynamiques. Malgré leurs avantages, les modèles QSPR obtenus peuvent ne pas être universellement applicables à la même famille de liquides ioniques, surtout dans les gammes de viscosité élevées et basses simultanément[163].

Au cours des dernières décennies, de nombreux chercheurs ont utilisé des modèles non linéaires avancés pour décrire la viscosité en utilisant différents descripteurs, allant de modèles simples à complexes. Ces modèles utilisent des techniques d'apprentissage automatique, notamment des réseaux neuronaux et des machines à vecteurs de support (SVM). Ces modèles de pointe ont démontré un potentiel pour dépasser certaines des contraintes des modèles QSPR conventionnels, en fournissant des prédictions plus précises dans un plus large éventail de conditions et de familles de composés. Ce type de technologie permet de repousser les limites, facilitant la description de la relation non linéaire entre la température, les distributeurs et la viscosité.

Gharagheizi et al[169]. ont présenté un modèle QSPR pour prédire la viscosité liquide de composés organiques purs à 25 °C. Ce modèle, un réseau neuronal artificiel à trois couches (FFANN), a été développé en utilisant huit paramètres sélectionnés par la méthode de régression linéaire multivariée basée sur l'algorithme génétique (GA-MLR) parmi plus de 3000 descripteurs potentiels. La capacité prédictive du modèle a été évaluée à l'aide de plusieurs méthodes statistiques, avec une racine carrée de l'erreur moyenne et une erreur moyenne absolue de pourcentage du modèle trouvées respectivement à 0,34 et 7 %. Très récemment, en 2021, Zhang et al. ont proposé un modèle QSPR pour la prédiction de la viscosité des liquides ioniques en fonction de la température et de la pression, basé sur des descripteurs de norme dérivés de la matrice de distribution atomique des graphes moléculaires et des géométries moléculaires. Les auteurs ont présenté le même modèle non linéaire que celui publié dans leur travail précédent, où ils ont supposé une dépendance non linéaire de la viscosité en fonction de la température et linéaire en fonction de la pression. La base de données utilisée pour déterminer les descripteurs moléculaires contenait 7342 points de données expérimentales pour 351 liquides ioniques couvrant la plage de température de 253 à 438 K et la plage de pression

de 0,06 à 300 MPa. Leur modèle a été évalué en utilisant une validation interne (validation croisée leave-one-out- LOOCV) et une procédure externe dans laquelle l'ensemble de données a été divisé de manière aléatoire en ensemble d'entraînement (80 %) et ensemble de test (20 %). Les valeurs ERAM étaient respectivement de 4,76, 4,84 et 4,74 % pour LOOCV, l'entraînement et les tests, pour les viscosités rapportées en unités logarithmiques. Lorsque l'ERAM global est recalculé, en utilisant les données supplémentaires fournies par les auteurs, une valeur d'environ 7,5 % est obtenue. Il convient de noter que l'application de ce modèle est relativement difficile, car elle nécessite une optimisation géométrique des ions et un logiciel avec des compétences particulièrement élevées pour évaluer les descripteurs. La même année, Ding et al. ont proposé quatre modèles QSPR pour prédire la viscosité des liquides ioniques en utilisant quatre différentes façons de représenter les liquides ioniques, à savoir les empreintes moléculaires, les descripteurs moléculaires, l'ajout des empreintes moléculaires du cation et de l'anion et la combinaison des empreintes moléculaires et des descripteurs moléculaires. Les corrélations ont été établies en utilisant un algorithme d'apprentissage automatique basé sur les arbres avec un renforcement par gradient. Le modèle QSPR développé en utilisant la combinaison d'empreinte moléculaire et de descripteur moléculaire a montré de meilleurs résultats avec un coefficient de corrélation (R^2) de 0,908. Les auteurs ont appliqué avec succès la méthode SHAP (explication additive shapely) pour interpréter les contributions de chaque empreinte digitale ou descripteur dans leur meilleur modèle, mais malheureusement ils n'ont pas rapporté de données supplémentaires sur leur base de données ou les résultats de prédiction. Malgré ses avantages, les modèles QSPR, le succès d'un modèle QSPR dépend fortement de la sélection de descripteurs moléculaires appropriés tels que conformationnels, électroniques, mécaniques quantiques, spatiaux, topologiques, thermodynamiques, et bien d'autres[170]. La sélection de descripteurs non pertinents ou redondants peut avoir un impact négatif sur les performances du modèle. De plus, un défi significatif se présente lorsque certains descripteurs chimiques ne sont pas directement accessibles pour les composés étudiés ou nécessitent l'utilisation de logiciels complexes. Cela peut constituer un obstacle substantiel à la mise en œuvre réussie d'un modèle QSPR.

I.1.1.A.6 Modèles de contribution de groupe

Les modèles de contribution de groupe sont des outils précieux pour prédire la viscosité de mélanges chimiques complexes. Ces modèles exploitent le principe selon lequel la

viscosité d'un composé peut être approximée en additionnant les contributions de ses groupes fonctionnels individuels ou de ses fragments structuraux. En attribuant des valeurs spécifiques de viscosité à chaque groupe ou fragment, ces modèles peuvent estimer la viscosité d'un mélange en fonction de sa composition. Les modèles de contribution de groupe ont été largement utilisés dans diverses industries, notamment le génie chimique, la pharmacie et les sciences alimentaires, en raison de leur capacité à fournir des prédictions de viscosité rapides et fiables pour une large gamme de composés. Cette approche présente l'avantage d'être simple, très rapide et facile à appliquer sans nécessiter une grande puissance de calcul et ne nécessite pas de logiciel spécialisé pour sa mise en œuvre. Divers modèles de contribution de groupe ont été développés pour prédire les viscosités des liquides. Les modèles les plus importants de cette catégorie sont ceux présentés par Paduszyński[171], Chen[171] , Lazzús[172], Baghban[165] ,Paduszyński [171]et Gharagheizi [118]. Cinq ans après la publication de leur remarquable modèle de GC [171] pour la prédiction de la viscosité des liquides ioniques, Paduszyński et ses collaborateurs ont proposé un nouveau GC très étendu pour le calcul de la viscosité des liquides ioniques en fonction de la température sur la base de la plus grande base de données comprenant les données pour 1978 liquides ioniques distincts, soit 15 372 points de données. Le modèle a été établi en tentant de combiner deux des trois algorithmes d'apprentissage automatique courants, à savoir la régression linéaire multiple pas à pas, le réseau neuronal artificiel à propagation avant (FFANN) et la machine à vecteurs de support des moindres carrés (LSSVM). Les techniques de validation interne et externe ont été utilisées pour sélectionner le meilleur modèle et vérifier la stabilité du modèle. Le schéma de fragmentation UNIFAC est utilisé pour fragmenter les cations et les anions en leurs groupes. Le modèle final implique un terme de référence basé sur FFANN pour prédire la valeur de la viscosité à la température de référence $T_0=298$ K, et un terme de correction basé sur LSSVM pour prédire l'effet de T autre que T_0 . En utilisant toutes les données dans la phase d'entraînement, ce modèle montre une erreur relative moyenne absolue (ERAM) globale entre expérimentale et calculée de 32,3 %. Bien que les résultats obtenus par Paduszyński semblent être la limite qui peut actuellement être atteinte, il convient de noter que leur modèle a été dérivé sur la base de bases de données étendues. La même année, Chen et al. ont présenté une série de modèles de GC simples estimant diverses propriétés des liquides ioniques, y compris la viscosité. La formulation mathématique du modèle de viscosité est la suivante :

$$\ln \frac{\eta}{R_{0\eta}} = A_\eta + B_\eta \frac{100}{T} + C_\eta \left(\frac{100}{T}\right)^2 \quad (41)$$

$R_{0\eta}$ est un paramètre ajustable et A_η , B_η et C_η sont des paramètres spécifiques à chaque liquide ionique qui sont estimés à partir de GC. Cette équation a été développée en utilisant uniquement 1090 points de données pour 76 liquides ioniques couvrant 36 cations et 16 anions, dont 778 points de données (56 liquides ioniques) ont été utilisés dans la régression et les données restantes ont été utilisées pour les tests. Les valeurs de ERAM entre la viscosité calculée et expérimentale, rapportées en unités ln, étaient de 3,36 % et 5,63 % pour l'entraînement et les tests, respectivement. Il convient de mentionner que les auteurs ont utilisé un ensemble limité de groupes fonctionnels complexes qui ne permettent pas de décrire une grande variété de liquides ioniques possibles. En effet, chaque structure anionique des 16 anions est proposée comme groupe fonctionnel, et pour décrire les 36 cations, 8 groupes fonctionnels complexes sont proposés ainsi que trois groupes substitués pour décrire l'effet de la longueur de la chaîne latérale.

I.5 CONCLUSION

En conclusion, ce chapitre a fourni un aperçu complet des concepts théoriques, des techniques expérimentales et des méthodes prédictives pertinentes pour l'étude des propriétés de transport (conductivité électrique, conductivité thermique, viscosité) dans les composés organiques purs, couvrant à la fois les composés non ioniques et ioniques. Nous avons exploré les principes fondamentaux sous-jacents à ces propriétés, élucidé les différentes méthodes expérimentales utilisées pour leur mesure et passé en revue les approches de prédiction. Par ailleurs, **l'impact des incertitudes de mesure sur la modélisation prédictive des propriétés de transport** a été mis en évidence, soulignant l'importance de leur prise en compte pour garantir la fiabilité des modèles. Dans les chapitres suivants, nous approfondirons les modèles prédictifs spécifiques et leurs applications, en nous appuyant sur les connaissances fondamentales établies ici.

I.6 BIBLIOGRAPHIE

- [1] D.R. MacFarlane, M. Kar, J.M. Pringle, *Fundamentals of Ionic Liquids*, Wiley, 2017. <https://doi.org/10.1002/9783527340033>.
- [2] F. Philippi, D. Rauber, J. Zapp, R. Hempelmann, Transport properties and ionicity of phosphonium ionic liquids, *Physical Chemistry Chemical Physics* 19 (2017) 23015–23023. <https://doi.org/10.1039/C7CP04552B>.
- [3] F. Philippi, D. Rauber, K.L. Eliasen, N. Bouscharain, K. Niss, C.W.M. Kay, T. Welton, Pressing matter: why are ionic liquids so viscous, *Chem Sci* 13 (2022) 2735. <https://doi.org/10.1039/D1SC06857A>.
- [4] J.G. McDaniel, A. Yethiraj, Understanding the Properties of Ionic Liquids: Electrostatics, Structure Factors, and Their Sum Rules, *Journal of Physical Chemistry B* 123 (2019) 3499–3512. https://doi.org/10.1021/ACS.JPCB.9B00963/SUPPL_FILE/JP9B00963_SI_002.PDF.
- [5] Physico-chemical properties and environmental impact of ionic liquids - TEL - Thèses en ligne, (n.d.). <https://theses.hal.science/tel-00669538/> (accessed March 23, 2025).
- [6] Q. Zhou, X. Lu, S. Zhang, L. Guo, *Physicochemical Properties of Ionic Liquids*, Wiley Blackwell 6 9781118438633 (2014) 275–307. <https://doi.org/10.1002/9781118839706.CH11>.
- [7] Y. SINGH, ELECTRICAL RESISTIVITY MEASUREMENTS: A REVIEW, [Https://Doi.Org/10.1142/S2010194513010970](https://doi.org/10.1142/S2010194513010970) 22 (2013) 745–756. <https://doi.org/10.1142/S2010194513010970>.
- [8] Z. Xie, H. Ventura, M. Ardanuy, A Review of the Electrical Conductivity Test Methods for Conductive Fabrics, *Textiles* 2024, Vol. 4, Pages 284-308 4 (2024) 284–308. <https://doi.org/10.3390/TEXTILES4030017>.
- [9] B. Allen J., F. Larry R., *Electrochemical Methods: Fundamentals and Applications*, 2nd Edition, Wiley, 2000.
- [10] A. Boruń, Conductance and ionic association of selected imidazolium ionic liquids in various solvents: A review, *J Mol Liq* 276 (2019) 214–224. <https://doi.org/10.1016/J.MOLLIQ.2018.11.140>.
- [11] P.T. Kissinger, W.R. Heinema, *Laboratory Techniques in Electroanalytical Chemistry*, Revised and Expanded, CRC Press, 2018. <https://doi.org/10.1201/9781315274263>.
- [12] Allen J. Bard and Larry R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, New York: Wiley, 2001, 2nd ed., *Russian Journal of Electrochemistry* 2002 38:12 38 (2002) 1364–1365. <https://doi.org/10.1023/A:1021637209564>.
- [13] C.M.A. Brett, Electrochemical Impedance Spectroscopy in the Characterisation and Application of Modified Electrodes for Electrochemical Sensors and Biosensors, *Molecules* 2022, Vol. 27, Page 1497 27 (2022) 1497. <https://doi.org/10.3390/MOLECULES27051497>.

- [14] K. Friedrich, S. Andreas, *Broadband Dielectric Spectroscopy*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. <https://doi.org/10.1007/978-3-642-56120-7>.
- [15] K. Ariyoshi, Z. Siroma, A. Mineshige, M. Takeno, T. Fukutsuka, T. Abe, S. Uchida, *Electrochemical Impedance Spectroscopy Part 1: Fundamentals*, *Electrochemistry* 90 (2022) 102007–102007. <https://doi.org/10.5796/ELECTROCHEMISTRY.22-66071>.
- [16] N.O. Laschuk, E.B. Easton, O. V. Zenkina, Reducing the resistance for the use of electrochemical impedance spectroscopy analysis in materials chemistry, *RSC Adv* 11 (2021) 27925. <https://doi.org/10.1039/D1RA03785D>.
- [17] A.C. Lazanas, M.I. Prodromidis, *Electrochemical Impedance Spectroscopy-A Tutorial*, *ACS Measurement Science Au* 3 (2023) 162–193. <https://doi.org/10.1021/ACSMEASURESCIAU.2C00070>.
- [18] C. Gabrielli, M. Keddam, H. Takenouti, Spectroscopie d'impédance électrochimique appliquée à l'étude de la corrosion, *Matériaux & Techniques* 95 (2007) 385–395. <https://doi.org/10.1051/MATTECH:2008035>.
- [19] L.J. van der PAUW, A METHOD OF MEASURING SPECIFIC RESISTIVITY AND HALL EFFECT OF DISCS OF ARBITRARY SHAPE, in: *Semiconductor Devices: Pioneering Papers*, WORLD SCIENTIFIC, 1991: pp. 174–182. https://doi.org/10.1142/9789814503464_0017.
- [20] Y. Lu, L.M. Santino, S. Acharya, H. Anandarajah, J.M. D'Arcy, Studying Electrical Conductivity Using a 3D Printed Four-Point Probe Station, *J Chem Educ* 94 (2017) 950–955. https://doi.org/10.1021/ACS.JCHEMED.7B00119/SUPPL_FILE/ED7B00119_SI_004.ZIP.
- [21] I. Kazani, G. de Mey, C. Hertleer, J. Banaszczuk, A. Schwarz, G. Guxho, L. Van Langenhove, About the collinear four-point probe technique's inability to measure the resistivity of anisotropic electroconductive fabrics, *Textile Research Journal* 83 (2013) 1587–1593. <https://doi.org/10.1177/0040517512452951>.
- [22] P. Ripka, A. Tipek, *Modern Sensors Handbook*, in: *Modern Sensors Handbook*, Wiley-ISTE, 2010. <https://doi.org/10.1002/9780470612231>.
- [23] D.R. MacFarlane, M. Forsyth, E.I. Izgorodina, A.P. Abbott, G. Annat, K. Fraser, On the concept of ionicity in ionic liquids, *Physical Chemistry Chemical Physics* 11 (2009) 4962. <https://doi.org/10.1039/b900201d>.
- [24] V. Lockett, R. Sedev, S. Harmer, J. Ralston, M. Horne, T. Rodopoulos, Orientation and mutual location of ions at the surface of ionic liquids, *Physical Chemistry Chemical Physics* 12 (2010) 13816. <https://doi.org/10.1039/c0cp00683a>.
- [25] D.M. Pozar, *Microwave Engineering*, 4th Edition, 2012. <https://www.wiley.com/en-us/Microwave+Engineering%2C+4th+Edition-p-9780470631553> (accessed March 29, 2025).
- [26] Y. Guo, Z. Zhang, C. Xue, J. Zheng, Z. Hai, S. Zhang, J. Liu, W. Zhang, J. Jing, B. Yao, Y. Xu, W. Ouyang, T. Song, B. Shi, Advanced Approaches to Immersion Nondestructive Measurements of Ionic Conductivity Gradients in Multi-phase Aqueous Environments – A Review, *IEEE Sens J* (2025) 1–1. <https://doi.org/10.1109/JSEN.2025.3549520>.

- [27] Daniel C. Harris, Charles A. Lucy, Quantitative Chemical Analysis, 10 th, Macmillan Learning, 2020.
- [28] Ari L. Horvath, Handbook of Aqueous Electrolyte Solutions: Physical Properties, Estimation and Correlation Methods, illustrée, 1985.
- [29] D.R. MacFarlane, N. Tachikawa, M. Forsyth, J.M. Pringle, P.C. Howlett, G.D. Elliott, J.H. Davis, M. Watanabe, P. Simon, C.A. Angell, Energy applications of ionic liquids, *Energy Environ. Sci.* 7 (2014) 232–250. <https://doi.org/10.1039/C3EE42099J>.
- [30] J. Leys, M. Wübbenhorst, C. Preethy Menon, R. Rajesh, J. Thoen, C. Glorieux, P. Nockemann, B. Thijs, K. Binnemans, S. Longuemart, Temperature dependence of the electrical conductivity of imidazolium ionic liquids, *Journal of Chemical Physics* 128 (2008). <https://doi.org/10.1063/1.2827462/978734>.
- [31] R.D. Rogers, K.R. Seddon, Ionic Liquids - Solvents of the Future?, *Science* (1979) 302 (2003) 792–793. <https://doi.org/10.1126/SCIENCE.1090313/ASSET/113B68CF-EEA0-43BB-B405-30789656C992/ASSETS/SCIENCE.1090313.FP.PNG>.
- [32] Douglas A. Skoog, Donald M. West, F. James Holler, Stanley R. Crouch, Douglas A. Skoog, Donald M. West, F. James Holler, and Stanley R. Crouch: Fundamentals of analytical chemistry, 9th ed., international ed., 9th ed., Springer, 2013. <https://doi.org/10.1007/S00216-013-7242-1>.
- [33] P. Bonhôte, A.P. Dias, N. Papageorgiou, K. Kalyanasundaram, M. Grätzel, Hydrophobic, Highly Conductive Ambient-Temperature Molten Salts, *Inorg Chem* 35 (1996) 1168–1178. <https://doi.org/10.1021/IC951325X/ASSET/IMAGES/MEDIUM/IC951325XN00001.GIF>.
- [34] C.A. Nieto de Castro, M.J. V. Lourenço, Towards the Correct Measurement of Thermal Conductivity of Ionic Melts and Nanofluids, *Energies* 2020, Vol. 13, Page 99 13 (2019) 99. <https://doi.org/10.3390/EN13010099>.
- [35] V.M.B. Nunes, M.J. V. Lourenço, F.J. V. Santos, M.L.S.M. Lopes, C.A. Nieto de Castro, Accurate Measurement of Physicochemical Properties on Ionic Liquids and Molten Salts, in: Molten Salts and Ionic Liquids, Wiley, 2010: pp. 229–263. <https://doi.org/10.1002/9780470947777.ch17>.
- [36] Y. Nagasaka, A. Nagashima, Absolute measurement of the thermal conductivity of electrically conducting liquids by the transient hot-wire method, *J Phys E* 14 (1981) 1435. <https://doi.org/10.1088/0022-3735/14/12/020>.
- [37] G.J. Tertsinidou, C.M. Tsolakidou, M. Pantzali, M.J. Assael, L. Colla, L. Fedele, S. Bobbo, W.A. Wakeham, New measurements of the apparent thermal conductivity of nanofluids and investigation of their heat transfer capabilities, *J Chem Eng Data* 62 (2017) 491–507. https://doi.org/10.1021/ACS.JCED.6B00767/ASSET/IMAGES/JE-2016-00767Q_M004.GIF.
- [38] ASTM C177-19, Standard Test Method for Steady-State Heat Flux Measurements and Thermal Transmission Properties by Means of the Guarded-Hot-Plate Apparatus, (2019). <https://doi.org/10.1520/C0177-19>.

- [39] D. Rooney, J. Jacquemin, R. Gardas, *Thermophysical Properties of Ionic Liquids*, Springer, Berlin, Heidelberg, 2009. https://doi.org/10.1007/128_2008_32.
- [40] H.C.B. Costa, D.O. Silva, L.G.M. Vieira, A numerical and experimental analysis of the methodology of thermal conductivity measurements in fluids by concentric cylinders, *Heat and Mass Transfer/Waerme- Und Stoffuebertragung* 55 (2019) 669–683. <https://doi.org/10.1007/S00231-018-2448-6/METRICS>.
- [41] Y. Nagasaka, T. Hatakeyama, M. Okuda, A. Nagashima, Measurement of the thermal diffusivity of liquids by the forced Rayleigh scattering method: Theory and experiment, *Review of Scientific Instruments* 59 (1988) 1156–1168. <https://doi.org/10.1063/1.1139743>.
- [42] Y. Nagasaka, N. Nakazawa, A. Nagashima, Experimental determination of the thermal diffusivity of molten alkali halides by the forced Rayleigh scattering method. I. Molten LiCl, NaCl, KCl, RbCl, and CsCl, *Int J Thermophys* 13 (1992) 555–574. <https://doi.org/10.1007/BF00501941/METRICS>.
- [43] Vincenzo. Bianco, Oronzio. Manca, Sergio. Nardini, K.. Vafai, *Heat transfer enhancement with nanofluids*, CRC Press, 2015. https://books.google.com/books/about/Heat_Transfer_Enhancement_with_Nanofluid.html?hl=fr&id=v5S9BwAAQBAJ (accessed April 1, 2025).
- [44] F.E.B. Bioucas, M.H. Rausch, T.M. Koller, A.P. Fröba, Guarded Parallel-Plate Instrument for the Determination of the Thermal Conductivity of Gases, Liquids, Solids, and Heterogeneous Systems, *Int J Heat Mass Transf* 212 (2023) 124283. <https://doi.org/10.1016/I.IJHEATMASSTRANSFER.2023.124283>.
- [45] H.-F. Lee, *A MODIFIED 3w METHOD FOR THERMAL CONDUCTIVITY MEASUREMENT OF ONE-DIMENSIONAL NANOSTRUCTURES*, The Pennsylvania State University, 2009.
- [46] C.S.G.P. Queirós, M.J.V. Lourenço, S.I. Vieira, J.M. Serra, C.A. Nieto De Castro, New portable instrument for the measurement of thermal conductivity in gas process conditions, *Review of Scientific Instruments* 87 (2016). <https://doi.org/10.1063/1.4952745/906219>.
- [47] J. Kestin, M. Sokolov, W. Wakeham, Theory of capillary viscometers, *Applied Scientific Research* 27 (1973) 241–264. <https://doi.org/10.1007/BF00382489/METRICS>.
- [48] R.F. Berg, E.F. May, M.R. Moldover, Viscosity ratio measurements with capillary viscometers, *J Chem Eng Data* 59 (2014) 116–124. https://doi.org/10.1021/JE400880N/ASSET/IMAGES/MEDIUM/JE-2013-00880N_0005.GIF.
- [49] C. Bowlt, A simple capillary viscometer, *Phys Educ* 10 (1975) 102. <https://doi.org/10.1088/0031-9120/10/2/011>.
- [50] M. Brizard, M. Megharfi, E. Mahé, C. Verdier, Design of a high precision falling-ball viscometer, *Review of Scientific Instruments* 76 (2005). <https://doi.org/10.1063/1.1851471/926196>.

- [51] A.A.H. Pádua, D. Tomida, C. Yokoyama, E.H. Abramson, R.F. Berg, E.F. May, M.R. Moldover, A. Laesecke, Viscometers, Experimental Thermodynamics Volume IX (2014) 96–131. <https://doi.org/10.1039/9781782625254-00096>.
- [52] Viscosity of Liquids, Springer Netherlands, Dordrecht, 2007. <https://doi.org/10.1007/978-1-4020-5482-2>.
- [53] Dabir S. Viswanath, Tushar K. Ghosh, Dasika H. L. Prasad, Nidamarty V.K. Dutt, Kalipatnapu Y. Rani, Viscosity of Liquids, Springer Netherlands, Dordrecht, 2007. <https://doi.org/10.1007/978-1-4020-5482-2>.
- [54] S. V. Gupta, Oscillating Viscometers, in: 2014: pp. 107–136. https://doi.org/10.1007/978-3-319-04858-1_5.
- [55] S. V. Gupta, Viscometry for Liquids, Springer International Publishing, Cham, 2014. <https://doi.org/10.1007/978-3-319-04858-1>.
- [56] J.T. Tough, W.D. McCormick, J.G. Dash, Viscosity of liquid He II, Physical Review 132 (1963) 2373–2378. <https://doi.org/10.1103/PhysRev.132.2373>.
- [57] M.J. Assael, N.K. Dalaouti, V. Vesovic, Viscosity of Natural-Gas Mixtures: Measurements and Prediction, Int J Thermophys 22 (2001) 61–71. <https://doi.org/10.1023/A:1006784814390>.
- [58] F.P. Equilibria, 301 Elsevier Science Publishers, 79 (1992) 301–311.
- [59] R.L. Powell, Rotational Viscometry, Rheological Measurement (1993) 247–296. https://doi.org/10.1007/978-94-017-2898-0_9.
- [60] Y. Kraftmakher, Rotational viscometers—a subject for student projects, Phys Educ 45 (2010) 622. <https://doi.org/10.1088/0031-9120/45/6/007>.
- [61] A.S. Morris, R. Langari, Summary of Other Measurements, 2016. <https://doi.org/10.1016/b978-0-12-800884-3.00021-6>.
- [62] M. Dzida, E. Zorębski, M. Zorębski, M. Żarska, M. Geppert-Rybczyńska, M. Chorążewski, J. Jacquemin, I. Cibulka, Speed of Sound and Ultrasound Absorption in Ionic Liquids, Chem Rev 117 (2017) 3883–3929. https://doi.org/10.1021/ACS.CHEMREV.5B00733/SUPPL_FILE/CR5B00733_SI_001.PDF.
- [63] M. HOUMMADY, D. HAUDEN, F. BASTIEN, CAPTEURS À ONDES ACOUSTIQUES TRANSVERSES HORIZONTALES. POSSIBILITÉS ET LIMITES POUR LA MESURE DE LA VISCOSITÉ, Le Journal de Physique IV 02 (1992) C1-657-C1-660. <https://doi.org/10.1051/jp4:19921142>.
- [64] Assia OUDINA, Etude théorique et expérimentale de la propagation des ondes ultrasonores dans les fluides visqueux, UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE De HOUARI BOUMEDIENNE , 2005.
- [65] F. Del Giudice, A Review of Microfluidic Devices for Rheological Characterisation, Micromachines 2022, Vol. 13, Page 167 13 (2022) 167. <https://doi.org/10.3390/MI13020167>.
- [66] A. Matsumoto, F. Del Giudice, R. Rotrattanadumrong, A.Q. Shen, Rheological Scaling of Ionic-Liquid-Based Polyelectrolytes in Ionic Liquid Solutions, Macromolecules 52 (2019)

- 2759–2771.
https://doi.org/10.1021/ACS.MACROMOL.8B02544/ASSET/IMAGES/MEDIUM/MA-2018-025445_M017.GIF.
- [67] R. Kuriyama, T. Nakagawa, K. Tatsumi, K. Nakabe, Two-dimensional fluid viscosity measurement in microchannel flow using fluorescence polarization imaging, *Meas Sci Technol* 32 (2021) 095402. <https://doi.org/10.1088/1361-6501/ABECCB>.
- [68] T.M. Squires, S.R. Quake, Microfluidics: Fluid physics at the nanoliter scale, *Rev Mod Phys* 77 (2005) 977–1026. <https://doi.org/10.1103/RevModPhys.77.977>.
- [69] P. Singh, K. Sharma, I. Puchades, P.B. Agarwal, A comprehensive review on MEMS-based viscometers, *Sens Actuators A Phys* 338 (2022) 113456. <https://doi.org/10.1016/J.SNA.2022.113456>.
- [70] K. Sherniyaz, K. Kaisarali, B. Akhat, S. Zdenka, S. Ibragim, Viscometer with Remote Electromagnetic Control, 2023 3rd International Conference on Robotics, Automation and Artificial Intelligence, RAAI 2023 (2023) 118–123. <https://doi.org/10.1109/RAAI59955.2023.10601296>.
- [71] A. Hans, Electromagnetic in-line measurement of viscosity, *Conference Record - IEEE Instrumentation and Measurement Technology Conference 2* (2002) 1023–1028. <https://doi.org/10.1109/IMTC.2002.1007095>.
- [72] Peter Wasserscheid, Thomas Welton, Ionic Liquids in Synthesis, Wiley, 2007. <https://doi.org/10.1002/9783527621194>.
- [73] R.D. Rogers, K.R. Seddon, Ionic Liquids - Solvents of the Future?, *Science* (1979) 302 (2003) 792–793. <https://doi.org/10.1126/SCIENCE.1090313/ASSET/113B68CF-EEA0-43BB-B405-30789656C992/ASSETS/SCIENCE.1090313.FP.PNG>.
- [74] M. Freemantle, T. Welton, R.D. Rogers, An Introduction to Ionic Liquids, The Royal Society of Chemistry, 2009. <https://doi.org/10.1039/9781839168604>.
- [75] V.I. Pârvulescu, C. Hardacre, Catalysis in ionic liquids, *Chem Rev* 107 (2007) 2615–2665. https://doi.org/10.1021/CR050948H/ASSET/CR050948H.FP.PNG_V03.
- [76] De Howard A. Barnes, John Fletcher Hutton, K. Walters, An Introduction to Rheology, Volume 3, Elsevier, 1989.
- [77] J.P. Hallett, T. Welton, Room-temperature ionic liquids: Solvents for synthesis and catalysis. 2, *Chem Rev* 111 (2011) 3508–3576. https://doi.org/10.1021/CR1003248/SUPPL_FILE/CR1003248_SI_001.PDF.
- [78] G. Feng, M. Chen, S. Bi, Z.A.H. Goodwin, E.B. Postnikov, N. Brilliantov, M. Urbakh, A.A. Kornyshev, Free and Bound States of Ions in Ionic Liquids, Conductivity, and Underscreening Paradox, *Phys Rev X* 9 (2019) 021024. <https://doi.org/10.1103/PHYSREVX.9.021024/FIGURES/10/MEDIUM>.
- [79] P. Eiden, S. Bulut, T. Köchner, C. Friedrich, T. Schubert, I. Krossing, In Silico Predictions of the Temperature-Dependent Viscosities and Electrical Conductivities of Functionalized and Nonfunctionalized Ionic Liquids, *J Phys Chem B* 115 (2011) 300–309. <https://doi.org/10.1021/jp108059x>.

- [80] D. Wileńska, I. Anusiewicz, S. Freza, M. Bobrowski, E. Laux, S. Uhl, H. Keppner, P. Skurski, Predicting the viscosity and electrical conductivity of ionic liquids on the basis of theoretically calculated ionic volumes, *Mol Phys* 113 (2015) 630–639. <https://doi.org/10.1080/00268976.2014.964344>.
- [81] Z.K. Koi, W.Z.N. Yahya, K.A. Kurnia, Prediction of ionic conductivity of imidazolium-based ionic liquids at different temperatures using multiple linear regression and support vector machine algorithms, *New Journal of Chemistry* 45 (2021) 18584–18597. <https://doi.org/10.1039/D1NJ01831K>.
- [82] K. Tochigi, H. Yamamoto, Estimation of Ionic Conductivity and Viscosity of Ionic Liquids Using a QSPR Model, *The Journal of Physical Chemistry C* 111 (2007) 15989–15994. <https://doi.org/10.1021/jp073839a>.
- [83] F. Gharagheizi, M. Sattari, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, D. Richon, A “non-linear” quantitative structure–property relationship for the prediction of electrical conductivity of ionic liquids, *Chem Eng Sci* 101 (2013) 478–485. <https://doi.org/10.1016/J.CES.2013.07.007>.
- [84] M. Abdullah, K. Chellappan Lethesh, A.A.B. Baloch, M.O. Bamgbopa, Comparison of molecular and structural features towards prediction of ionic liquid ionic conductivity for electrochemical applications, *J Mol Liq* 368 (2022) 120620. <https://doi.org/10.1016/J.MOLLIQ.2022.120620>.
- [85] I. Baskin, A. Epshtain, Y. Ein-Eli, Benchmarking machine learning methods for modeling physical properties of ionic liquids, *J Mol Liq* 351 (2022) 118616. <https://doi.org/10.1016/J.MOLLIQ.2022.118616>.
- [86] P. Dhakal, J.K. Shah, Developing machine learning models for ionic conductivity of imidazolium-based ionic liquids, *Fluid Phase Equilib* 549 (2021) 113208. <https://doi.org/10.1016/J.FLUID.2021.113208>.
- [87] P. Dhakal, J.K. Shah, A generalized machine learning model for predicting ionic conductivity of ionic liquids, *Mol Syst Des Eng* 7 (2022) 1344–1353. <https://doi.org/10.1039/D2ME00046F>.
- [88] P. Nancarrow, A. Al-Othman, D.K. Mital, S. Döpking, Comprehensive analysis and correlation of ionic liquid conductivity data for energy applications, *Energy* 220 (2021) 119761. <https://doi.org/10.1016/J.ENERGY.2021.119761>.
- [89] Q. Dong, C.D. Muzny, A. Kazakov, V. Diky, J.W. Magee, J.A. Widgren, R.D. Chirico, K.N. Marsh, M. Frenkel, ILThermo: A free-access web database for thermodynamic properties of ionic liquids, *J Chem Eng Data* 52 (2007) 1151–1159. <https://doi.org/10.1021/JE700171F>.
- [90] T.E. Karakasidis, F. Sofos, C. Tsonos, The Electrical Conductivity of Ionic Liquids: Numerical and Analytical Machine Learning Approaches, *Fluids* 7 (2022). <https://doi.org/10.3390/fluids7100321>.
- [91] K. Argoub, A.M. Benkouider, A. Yahiaoui, F. Bagui, Estimation and uncertainty analysis of standard enthalpy of formation in the liquid state by third-order-group-contribution method, *Fluid Phase Equilib* 520 (2020) 112644. <https://doi.org/10.1016/j.fluid.2020.112644>.

- [92] F.Z. Serat, A.M. Benkouider, A. Yahiaoui, F. Bagui, Nonlinear group contribution model for the prediction of flash points using normal boiling points, *Fluid Phase Equilib* 449 (2017) 52–59. <https://doi.org/10.1016/j.fluid.2017.06.008>.
- [93] A.M. Benkouider, R. Kessas, S. Guella, A. Yahiaoui, F. Bagui, Estimation of the enthalpy of vaporization of organic components as a function of temperature using a new group contribution method, *J Mol Liq* 194 (2014) 48–56. <https://doi.org/10.1016/j.molliq.2014.01.006>.
- [94] K. Argoub, A.M. Benkouider, A. Yahiaoui, R. Kessas, S. Guella, F. Bagui, Prediction of standard enthalpy of formation in the solid state by a third-order group contribution method, *Fluid Phase Equilib* 380 (2014) 121–127. <https://doi.org/10.1016/j.fluid.2014.08.003>.
- [95] A.D. Boualem, K. Argoub, A.M. Benkouider, A. Yahiaoui, K. Toubal, Viscosity prediction of ionic liquids using NLR and SVM approaches, *J Mol Liq* 368 (2022) 120610. <https://doi.org/https://doi.org/10.1016/j.molliq.2022.120610>.
- [96] H. Matsuda, H. Yamamoto, K. Kurihara, K. Tochigi, Computer-aided reverse design for ionic liquids by QSPR using descriptors of group contribution type for ionic conductivities and viscosities, *Fluid Phase Equilib* 261 (2007) 434–443. <https://doi.org/https://doi.org/10.1016/j.fluid.2007.07.018>.
- [97] R.L. Gardas, J.A.P. Coutinho, Group contribution methods for the prediction of thermophysical and transport properties of ionic liquids, *AIChE Journal* 55 (2009) 1274–1290. <https://doi.org/https://doi.org/10.1002/aic.11737>.
- [98] J.A.P. Coutinho, R.L. Gardas, Predictive Group Contribution Models for the Thermophysical Properties of Ionic Liquids, in: *Ionic Liquids: From Knowledge to Application*, American Chemical Society, 2009: pp. 385–401. <https://doi.org/doi:10.1021/bk-2009-1030.ch025>.
- [99] F. Gharagheizi, P. Ilani-Kashkouli, M. Sattari, A.H. Mohammadi, D. Ramjugernath, D. Richon, Development of a LSSVM-GC model for estimating the electrical conductivity of ionic liquids, *Chemical Engineering Research and Design* 92 (2014) 66–79. <https://doi.org/10.1016/J.CHERD.2013.06.015>.
- [100] N. Zhao, J. Jacquemin, The development of the UNIFAC-CONDUCT model as a novel approach for the estimation of the conductivity of pure ionic liquids, *Fluid Phase Equilib* 449 (2017) 60–67. <https://doi.org/https://doi.org/10.1016/j.fluid.2017.06.010>.
- [101] Y. Chen, Y. Cai, K. Thomsen, G.M. Kontogeorgis, J.M. Woodley, A group contribution-based prediction method for the electrical conductivity of ionic liquids, *Fluid Phase Equilib* 509 (2020) 112462. <https://doi.org/https://doi.org/10.1016/j.fluid.2020.112462>.
- [102] K.-J. Wu, H. Luo, L. Yang, Structure-based model for prediction of electrical conductivity of pure ionic liquids, *AIChE Journal* 62 (2016) 3751–3762. <https://doi.org/https://doi.org/10.1002/aic.15429>.
- [103] R. Nakhaei-Kohani, S. Ali Madani, S.-P. Mousavi, S. Atashrouz, A. Abedi, A. Hemmati-Sarapardeh, A. Mohaddespour, Machine learning assisted Structure-based models for predicting electrical conductivity of ionic liquids, *J Mol Liq* 362 (2022) 119509. <https://doi.org/https://doi.org/10.1016/j.molliq.2022.119509>.

- [104] R.L.G. and J.A.P. Coutinho, PVAm–PIP/PS composite membrane with high performance for CO₂/N₂ separation, *AIChE Journal* 55 (2009) 1274–1290. <https://doi.org/10.1002/aic>.
- [105] D. Tomida, S. Kenmochi, T. Tsukada, K. Qiao, C. Yokoyama, Thermal conductivities of [bmim][PF₆], [hmim][PF₆], and [omim][PF₆] from 294 to 335 K at pressures up to 20 MPa, *Int J Thermophys* 28 (2007) 1147–1160. <https://doi.org/10.1007/s10765-007-0241-8>.
- [106] J.A. Lazzús, A group contribution method to predict the thermal decomposition temperature of ionic liquids, *J Mol Liq* 168 (2012) 87–93. <https://doi.org/10.1016/j.molliq.2012.01.011>.
- [107] J.A. Lazzús, A group contribution method to predict the melting point of ionic liquids, *Fluid Phase Equilib* 313 (2012) 1–6. <https://doi.org/10.1016/j.fluid.2011.09.018>.
- [108] A.P. Fröba, M.H. Rausch, K. Krzeminski, D. Assenbaum, P. Wasserscheid, A. Leipertz, Thermal conductivity of ionic liquids: Measurement and prediction, *Int J Thermophys* 31 (2010) 2059–2077. <https://doi.org/10.1007/s10765-010-0889-3>.
- [109] S.A. Shojaee, S. Farzam, A.Z. Hezave, M. Lashkarbolooki, S. Ayatollahi, A new correlation for estimating thermal conductivity of pure ionic liquids, *Fluid Phase Equilib* 354 (2013) 199–206. <https://doi.org/10.1016/j.fluid.2013.06.004>.
- [110] C. Frez, G.J. Diebold, C.D. Tran, S. Yu, Determination of thermal diffusivities, thermal conductivities, and sound speeds of room-temperature ionic liquids by the transient grating technique, *J Chem Eng Data* 51 (2006) 1250–1255. <https://doi.org/10.1021/je0600092>.
- [111] A.P. Abbott, Application of hole theory to the viscosity of ionic and molecular liquids, *ChemPhysChem* 5 (2004) 1242–1246. <https://doi.org/10.1002/cphc.200400190>.
- [112] H. Zhao, Z.C. Liang, F. Li, An improved model for the conductivity of room-temperature ionic liquids based on hole theory, *J Mol Liq* 149 (2009) 55–59. <https://doi.org/10.1016/j.molliq.2009.08.002>.
- [113] K. Tochigi, H. Yamamoto, Estimation of ionic conductivity and viscosity of ionic liquids using a QSPR model, *Journal of Physical Chemistry C* 111 (2007) 15989–15994. <https://doi.org/10.1021/jp073839a>.
- [114] N. Zhao, J. Jacquemin, The development of the UNIFAC-CONDUCT model as a novel approach for the estimation of the conductivity of pure ionic liquids, *Fluid Phase Equilib* 449 (2017) 60–67. <https://doi.org/10.1016/j.fluid.2017.06.010>.
- [115] Y. Chen, Y. Cai, K. Thomsen, G.M. Kontogeorgis, J.M. Woodley, A group contribution-based prediction method for the electrical conductivity of ionic liquids, *Fluid Phase Equilib* 509 (2020) 112462. <https://doi.org/10.1016/j.fluid.2020.112462>.
- [116] P. Nancarrow, A. Al-Othman, D.K. Mital, S. Döpking, Comprehensive analysis and correlation of ionic liquid conductivity data for energy applications, *Energy* 220 (2021). <https://doi.org/10.1016/j.energy.2021.119761>.
- [117] K.J. Wu, C.X. Zhao, C.H. He, Development of a group contribution method for determination of thermal conductivity of ionic liquids, *Fluid Phase Equilib* 339 (2013) 10–14. <https://doi.org/10.1016/j.fluid.2012.11.024>.

- [118] F. Gharagheizi, P. Ilani-Kashkouli, M. Sattari, A.H. Mohammadi, D. Ramjugernath, D. Richon, Development of a LSSVM-GC model for estimating the electrical conductivity of ionic liquids, *Chemical Engineering Research and Design* 92 (2014) 66–79. <https://doi.org/10.1016/j.cherd.2013.06.015>.
- [119] K.S. Pedersen, A. Fredenslund, P.L. Christensen, P. Thomassen, Viscosity of crude oils, *Chem Eng Sci* 39 (1984) 1011–1016. [https://doi.org/10.1016/0009-2509\(84\)87009-8](https://doi.org/10.1016/0009-2509(84)87009-8).
- [120] A. Fredenslund, D.- Lyngby, K.S. Pedersen, D.- Lyngby, A Thermodynamic Model for Predicting Wax Formation in Crude Oils, 34 (1988) 1937–1942.
- [121] A.S. Teja, P. Rice, Generalized Corresponding States Method for the Viscosities of Liquid Mixtures, *Industrial and Engineering Chemistry Fundamentals* 20 (1981) 77–81. <https://doi.org/10.1021/i100001a015>.
- [122] K. Aasberg-Petersen, K. Knudsen, A. Fredenslund, Prediction of viscosities of hydrocarbon mixtures, *Fluid Phase Equilib* 70 (1991) 293–308. [https://doi.org/10.1016/0378-3812\(91\)85041-R](https://doi.org/10.1016/0378-3812(91)85041-R).
- [123] Y. Huang, H. Dong, X. Zhang, C. Li, S. Zhang, A new fragment contribution-corresponding states method for physicochemical properties prediction of ionic liquids, *AIChE Journal* 59 (2013) 1348–1359.
- [124] Y. Zhao, X. Zhang, L. Deng, S. Zhang, Prediction of viscosity of imidazolium-based ionic liquids using MLR and SVM algorithms, *Comput Chem Eng* 92 (2016) 37–42. <https://doi.org/10.1016/j.compchemeng.2016.04.035>.
- [125] S.E. Quiñones-Cisneros, C.K. Zéberg-Mikkelsen, E.H. Stenby, The friction theory (f-theory) for viscosity modeling, *Fluid Phase Equilib* 169 (2000) 249–276. [https://doi.org/10.1016/S0378-3812\(00\)00310-1](https://doi.org/10.1016/S0378-3812(00)00310-1).
- [126] G. Shen, C. Held, J.P. Mikkola, X. Lu, X. Ji, Modeling the viscosity of ionic liquids with the electrolyte perturbed-chain statistical association fluid theory, *Ind Eng Chem Res* 53 (2014) 20258–20268. <https://doi.org/10.1021/ie503485h>.
- [127] M. Abolala, K. Peyvandi, F. Varaminian, Modeling the viscosity of pure imidazolium-based ionic liquids using SAFT-VR-Mie EoS, *Fluid Phase Equilib* 394 (2015) 61–70. <https://doi.org/10.1016/j.fluid.2015.02.034>.
- [128] R. Haghbakhsh, K. Parvaneh, A. Shariati, 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, *Ind Eng Chem Res* 56 (2017) 2247–2258. <https://doi.org/10.1021/acs.iecr.6b04193>.
- [129] R. Macías-Salinas, Viscosity Modeling of Ionic Liquids Using the Friction Theory and a Simple Cubic Equation of State, *Ind Eng Chem Res* 57 (2018) 1109–1120. <https://doi.org/10.1021/acs.iecr.7b04252>.
- [130] A.J. Batschinski, Untersuchungen Aber die innere Reibnng der Flüssigkeiten. I, *Zeitschrift Für Physikalische Chemie* 84U (1913) 643–706. <https://doi.org/10.1515/zpch-1913-8442>.
- [131] M.H. Cohen, D. Turnbull, Molecular transport in liquids and glasses, *J Chem Phys* 31 (1959) 1164–1169. <https://doi.org/10.1063/1.1730566>.

- [132] A.K. Doolittle, Studies in newtonian flow. II. the dependence of the viscosity of liquids on free-space, *J Appl Phys* 22 (1951) 1471–1475. <https://doi.org/10.1063/1.1699894>.
- [133] F. Llorell, E. Valente, O. Vilaseca, L.F. Vega, Modeling complex associating mixtures with [C_n-mim][Tf₂N] ionic liquids: Predictions from the soft-SAFT equation, *Journal of Physical Chemistry B* 115 (2011) 4387–4398. <https://doi.org/10.1021/jp112315b>.
- [134] H.W. Yarranton, M.A. Satyro, Expanded fluid-based viscosity correlation for hydrocarbons, *Ind Eng Chem Res* 48 (2009) 3640–3648. <https://doi.org/10.1021/ie801698h>.
- [135] M.A. Satyro, H.W. Yarranton, Expanded fluid-based viscosity correlation for hydrocarbons using an equation of state, *Fluid Phase Equilib* 298 (2010) 1–11. <https://doi.org/10.1016/j.fluid.2010.06.023>.
- [136] H. Loria, H. Motahhari, M.A. Satyro, H.W. Yarranton, Process simulation using the expanded fluid model for viscosity calculations, *Chemical Engineering Research and Design* 92 (2014) 3083–3095. <https://doi.org/10.1016/j.cherd.2014.06.019>.
- [137] Y. Rosenfeld, Relation between the transport coefficients and the internal entropy of simple systems, *Phys Rev A (Coll Park)* 15 (1977) 2545–2549. <https://doi.org/10.1103/PhysRevA.15.2545>.
- [138] W.T. Ashurst, W.G. Hoover, Dense-fluid shear viscosity via nonequilibrium molecular dynamics, *Phys Rev A (Coll Park)* 11 (1975) 658–678. <https://doi.org/10.1103/PhysRevA.11.658>.
- [139] S.G. (1990) Chang, R. P., & Rhee, Scholar (28), (1990).
- [140] A. Baskaran, J.W. Dufty, J.J. Brey, Transport coefficients for the hard-sphere granular fluid, *Phys Rev E Stat Nonlin Soft Matter Phys* 77 (2008) 1–28. <https://doi.org/10.1103/PhysRevE.77.031311>.
- [141] D. Chandler, Rough hard sphere theory of the self-diffusion constant for molecular liquids, *J Chem Phys* 62 (1975) 1358–1363. <https://doi.org/10.1063/1.430647>.
- [142] J.H. Dymond, Corrections to the Enskog theory for viscosity and thermal conductivity, *Physica B+C* 144 (1987) 267–276. [https://doi.org/10.1016/0378-4363\(87\)90009-X](https://doi.org/10.1016/0378-4363(87)90009-X).
- [143] T.C. Horng, M. Ajlan, L.L. Lee, K.E. Starling, M. Ajlan, Generalized Multiparameter Correlation for Nonpolar and Polar Fluid Transport Properties, *Ind Eng Chem Res* 27 (1988) 671–679. <https://doi.org/10.1021/ie00076a024>.
- [144] T.H. Chung, L.L. Lee, K.E. Starting, Applications of Kinetic Gas Theories and Multiparameter Correlation for Prediction of Dilute Gas Viscosity and Thermal Conductivity, *Industrial and Engineering Chemistry Fundamentals* 23 (1984) 8–13. <https://doi.org/10.1021/i100013a002>.
- [145] M.J. Assael, J.H. Dymond, M. Papadaki, P.M. Patterson, Correlation and prediction of dense fluid transport coefficients. I. n-alkanes, *Int J Thermophys* 13 (1992) 269–281. <https://doi.org/10.1007/BF00504436>.

- [146] J.G. Bleazard, A.S. Teja, Extension of the rough hard-sphere theory for transport properties to polar liquids, *Ind Eng Chem Res* 35 (1996) 2453–2459. <https://doi.org/10.1021/ie9507585>.
- [147] F. Ciotta, J.P.M. Trusler, V. Vesovic, Extended hard-sphere model for the viscosity of dense fluids, *Fluid Phase Equilib* 363 (2014) 239–247. <https://doi.org/10.1016/j.fluid.2013.11.032>.
- [148] N. Riesco, V. Vesovic, Extended hard-sphere model for predicting the viscosity of long-chain n-alkanes, *Fluid Phase Equilib* 425 (2016) 385–392. <https://doi.org/10.1016/j.fluid.2016.06.034>.
- [149] V. Vesovic, W.A. Wakeham, Prediction of the viscosity of fluid mixtures over wide ranges of temperature and pressure, *Chem Eng Sci* 44 (1989) 2181–2189. [https://doi.org/10.1016/0009-2509\(89\)85152-8](https://doi.org/10.1016/0009-2509(89)85152-8).
- [150] A.S. De Wijn, N. Riesco, G. Jackson, J.P. Martin Trusler, V. Vesovic, Viscosity of liquid mixtures: The Vesovic-Wakeham method for chain molecules, *Journal of Chemical Physics* 136 (2012). <https://doi.org/10.1063/1.3685605>.
- [151] F.M. Gaciño, M.J.P. Comuñas, J. Fernández, S.K. Mylona, M.J. Assael, Correlation and prediction of dense fluid transport coefficients. IX. ionic liquids, *Int J Thermophys* 35 (2014) 812–829. <https://doi.org/10.1007/s10765-014-1626-0>.
- [152] M.Z. Hossain, A.S. Teja, Correlation and Prediction of the Transport Properties of Ionic Liquids, *Int J Thermophys* 37 (2016) 1–17. <https://doi.org/10.1007/s10765-015-2021-1>.
- [153] S.M. Hosseini, M.M. Alavianmehr, J. Moghadasi, Transport properties of pure and mixture of ionic liquids from new rough hard-sphere-based model, *Fluid Phase Equil* 429 (2016) 266–274. <https://doi.org/10.1016/j.fluid.2016.09.004>.
- [154] S.M. Hosseini, Erratum to “Transport properties of pure and mixture of ionic liquids from new rough hard-sphere-based model” (*Fluid Phase Equil.* (2016) 429 (266–274) (S0378381216304381)(10.1016/j.fluid.2016.09.004)), *Fluid Phase Equil* 458 (2018) 300. <https://doi.org/10.1016/j.fluid.2017.12.013>.
- [155] F. Akbari, M.M. Alavianmehr, R. Behjatmanesh Ardakani, D. Mohammad-Aghaie, Thermophysical properties of ionic liquids and their mixtures from a new equation of state, *Ionicics (Kiel)* 24 (2018) 1357–1369. <https://doi.org/10.1007/s11581-017-2310-8>.
- [156] H.T. Davis, S.A. Rice, J. V. Sengers, On the kinetic theory of dense fluids. IX. The fluid of rigid spheres with a square-well attraction, *J Chem Phys* 35 (1961) 2210–2233. <https://doi.org/10.1063/1.1732234>.
- [157] I.L. McLaughlin, H. Ted Davis, Kinetic theory of dense fluid mixtures. I. Square-well model, *J Chem Phys* 45 (1966) 2020–2031. <https://doi.org/10.1063/1.1727886>.
- [158] L.G. Du, T.M. Guo, A semi-theoretical viscosity model for non-polar liquids, *The Chemical Engineering Journal* 47 (1991) 163–167. [https://doi.org/10.1016/0300-9467\(91\)85022-N](https://doi.org/10.1016/0300-9467(91)85022-N).
- [159] W.D. Monnery, W.Y. Svrcek, A.K. Mehrotra, Viscosity: A Critical Review, *Canadian Journal of Chemical Engineering* 73 (1995) 3–40.

- [160] G. Galliero, C. Nieto-Draghi, C. Boned, J.B. Avalos, A.D. Mackie, A. Baylaucq, F. Montel, Molecular dynamics simulation of acid gas mixtures: A comparison between several approximations, *Ind Eng Chem Res* 46 (2007) 5238–5244. <https://doi.org/10.1021/ie061616l>.
- [161] I.M. Zerón, M. Cueto-Mora, F.J. Blas, Transport properties of the square-well fluid from molecular dynamics simulation, *Mol Phys* (n.d.) e2302385. <https://doi.org/10.1080/00268976.2024.2302385>.
- [162] H. Guerrand, M. Puchault, M. Vaultier, Ionic liquids, *Green Proc. Eng* 267 (2015).
- [163] C. Nieto-Draghi, G. Fayet, B. Creton, X. Rozanska, P. Rotureau, J.C. De Hemptinne, P. Ungerer, B. Rousseau, C. Adamo, A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes, *Chem Rev* 115 (2015) 13093–13164. <https://doi.org/10.1021/acs.chemrev.5b00215>.
- [164] A.R. Katritzky, K. Chen, Y. Wang, M. Karelson, B. Lucic, N. Trinajstic, T. Suzuki, G. Schüürmann, Prediction of liquid viscosity for organic compounds by a quantitative structure-property relationship, *J Phys Org Chem* 13 (2000) 80–86. [https://doi.org/10.1002/\(SICI\)1099-1395](https://doi.org/10.1002/(SICI)1099-1395).
- [165] A.F. Bouarab, J.P. Harvey, C. Robelin, Viscosity models for ionic liquids and their mixtures, *Physical Chemistry Chemical Physics* 23 (2021) 733–752. <https://doi.org/10.1039/d0cp05787h>.
- [166] M. Barycki, A. Sosnowska, A. Gajewicz, M. Bobrowski, D. Wileńska, P. Skurski, A. Giełdoń, C. Czaplewski, S. Uhl, E. Laux, T. Journot, L. Jeandupeux, H. Keppner, T. Puzyn, Temperature-dependent structure-property modeling of viscosity for ionic liquids, *Fluid Phase Equilib* 427 (2016) 9–17. <https://doi.org/10.1016/j.fluid.2016.06.043>.
- [167] R. Bini, M. Malvaldi, W.R. Pitner, C. Chiappe, QSPR correlation for conductivities and viscosities of low-temperature melting ionic liquids, *J Phys Org Chem* 21 (2008) 622–629. <https://doi.org/10.1002/poc.1337>.
- [168] G. Yu, L. Wen, D. Zhao, C. Asumana, X. Chen, QSPR study on the viscosity of bis(trifluoromethylsulfonyl)imide-based ionic liquids, *J Mol Liq* 184 (2013) 51–59. <https://doi.org/10.1016/j.molliq.2013.04.021>.
- [169] F. Gharagheizi, S.A. Mirkhani, M.H. Keshavarz, N. Farahani, K. Tumba, A molecular-based model for prediction of liquid viscosity of pure organic compounds: A quantitative structure property relationship (QSPR) approach, *J Taiwan Inst Chem Eng* 44 (2013) 359–364. <https://doi.org/10.1016/j.jtice.2012.12.015>.
- [170] scholar (29), (n.d.).
- [171] K. Paduszyński, Extensive databases and group contribution QSPRs of ionic liquids properties. 1. Density, *Ind Eng Chem Res* 58 (2019) 5322–5338.
- [172] J.A. Lazzús, G. Pulgar-Villarroel, Estimation of thermal conductivity of ionic liquids using quantitative structure-property relationship calculations, *J Mol Liq* 211 (2015) 981–985. <https://doi.org/10.1016/j.molliq.2015.08.037>.

CHAPITRE II : METHODOLOGIE ET METHODES DE MODELISATION

CHAPITRE II: METHODOLOGIE ET METHODES DE MODELISATION

II.1 INTRODUCTION

Ce chapitre sert de guide fondamental, explorant diverses méthodologies et méthodes de modélisation essentielles pour comprendre le processus complexe de développement de modèles. En commençant par une introduction, le chapitre aborde l'importance des bases de données dans la modélisation, en soulignant l'importance de la révision des données pour garantir précision et fiabilité. Il explore ensuite les méthodes de contribution de groupe, en expliquant les principes de ces techniques et la structure hiérarchique des groupes utilisés dans les schémas de fragmentation.

Le chapitre examine plus en détail les Relations Quantitatives Structure-Propriété (QSPR), en éclairant les principes fondamentaux, les descripteurs moléculaires impliqués et les empreintes digitales dans la modélisation prédictive. Les techniques de réduction de dimensionnalité, les méthodes de transformation, la sélection de caractéristiques et les techniques de régression parcimonieuse sont également examinées en détail.

Ensuite, le chapitre élucide les modèles conventionnels et les algorithmes d'apprentissage automatique utilisés dans la prédiction des propriétés de transport, y compris la Forêt Aléatoire, le Gradient Boosting Extrême, la Machine à Vecteurs de Support Non-Linéaire et le Réseau de Neurones. Enfin, le chapitre se termine par une discussion sur la validation des modèles, couvrant les paramètres de validation statistique, la validation croisée et la validation externe.

II.2 IMPORTANCE DES BASES DE DONNÉES

Les bases de données utilisées pour la construction d'un modèle empirique ou semi-théorique doivent être composées de données expérimentales fiables et précises. Les erreurs associées à ces données se propagent dans le modèle final, puisque les paramètres du modèle sont ajustés en fonction de ces données. Ainsi, des données de haute qualité avec des incertitudes minimales sont nécessaires pour limiter les erreurs expérimentales et assurer la robustesse statistique du modèle.

De plus, les données doivent être obtenues selon un protocole expérimental standardisé pour garantir leur validité et leur comparabilité. Les conditions expérimentales, telles que la température et la pression, peuvent avoir une influence significative sur les valeurs de propriété mesurées. Par conséquent, il est crucial de définir précisément les conditions expérimentales pour chaque point de données inclus dans la base de données.

Une base de données de qualité dans les modèles empiriques ou semi-théoriques doit également être représentative de la diversité des valeurs que la propriété étudiée peut prendre. Cela permet de déterminer le domaine d'applicabilité du modèle final, c'est-à-dire la gamme de valeurs pour lesquelles le modèle est fiable et précis. Un domaine d'applicabilité plus large permet au modèle de prédire avec précision une plus grande variété de valeurs.

En outre, la distribution des données dans la base de données doit être aussi homogène et normale que possible. Les méthodes statistiques utilisées pour développer des modèles empiriques ou semi-théoriques sont généralement basées sur des distributions normales, et les valeurs aberrantes peuvent fausser les résultats du modèle.

Le choix des molécules incluses dans la base de données est également crucial pour la performance du modèle empirique ou semi-théorique. Il est recommandé d'inclure des molécules avec des mécanismes uniques et des structures moléculaires diverses. Cela permet d'obtenir un modèle plus précis et généralisable. Cependant, inclure des molécules avec des structures similaires peut également améliorer la performance du modèle, même si cela peut réduire son domaine d'applicabilité.

II.2.1 Révision des données

Le prétraitement des données joue un rôle crucial dans le développement de modèles prédictifs. Cette étape vise à améliorer la qualité des données en éliminant les erreurs, en réduisant le bruit et en mettant en évidence les caractéristiques importantes des données. Un aspect important du prétraitement des données est la gestion des valeurs aberrantes, qui peuvent déformer les résultats du modèle. Ces valeurs aberrantes étaient probablement dues à des impuretés dans les échantillons, aux procédures de purification ou à la diversité des techniques expérimentales utilisées pour mesurer les propriétés de transport des composés organiques.

Dans le cadre de cette thèse, nous avons développé une procédure innovante pour le prétraitement des données et la détection des valeurs aberrantes. Elle se compose principalement de sept étapes consécutives.

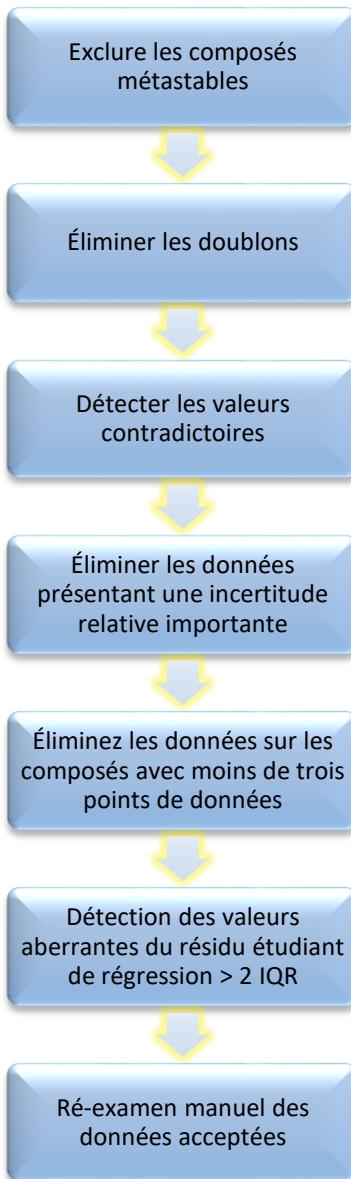


Figure 21:Algorithme de prétraitement des données

La première étape a consisté à exclure les composés métastables. La deuxième étape visait à supprimer les doublons dans chaque ensemble de données. La troisième étape impliquait l'identification des valeurs présentant un comportement contradictoire, c'est-à-dire des valeurs multiples pour le même liquide à des températures et des pressions identiques. Dans de tels cas, une seule valeur représentant la moyenne des valeurs a été retenue. La quatrième étape visait à éliminer les données avec une incertitude relative dépassant 100%, tandis que la cinquième étape visait à éliminer les données pour les composés ayant moins de trois points de données. Lors de la sixième étape, une régression a été effectuée sur les données restantes pour chaque composé à pression atmosphérique (100-101,325 kPa) en utilisant l'équation de Vogel-Fulcher-Tammann. Par la suite, une valeur

aberrante était considérée comme tout point de données ayant un résidu studentisé dépassant deux fois l'intervalle interquartile (IQR), et cette étape a été itérée jusqu'à ce qu'aucune valeur aberrante ne soit détectée. Dans la dernière étape de ce processus, les données acceptées ont été examinées manuellement pour éliminer les données potentiellement inappropriées.

II.3 MÉTHODES DE CONTRIBUTION DE GROUPE

II.3.1 Principe

Les méthodes de contribution de groupe représentent une approche puissante pour prédire les propriétés de transport des composés organiques. Ces méthodes fonctionnent sur le principe que les propriétés de transport d'un composé peuvent être approximées en additionnant les contributions de ses groupes fonctionnels ou fragments moléculaires constitutifs. En attribuant des contributions spécifiques à des groupes individuels en fonction de leurs caractéristiques structurelles et chimiques, les méthodes de contribution de groupe offrent une approche systématique pour prédire les propriétés de transport sur une large gamme de molécules organiques[2][3].

L'essence des méthodes de contribution de groupe réside dans l'expression des propriétés de transport d'un composé comme une combinaison linéaire des contributions de divers groupes fonctionnels ou fragments moléculaires présents dans la molécule. Chaque groupe fonctionnel se voit attribuer une valeur de contribution caractéristique, généralement dérivée de données expérimentales ou de calculs théoriques. La propriété de transport globale du composé est ensuite calculée en additionnant les contributions de tous les groupes constitutifs[4].

La détermination des valeurs de contribution pour chaque groupe fonctionnel implique des corrélations empiriques avec les données de propriété de transport expérimentales pour un ensemble diversifié de composés [5]. Des techniques statistiques telles que l'analyse de régression sont couramment utilisées pour établir des relations entre les contributions des groupes individuels et la propriété de transport globale des composés. De plus, des calculs de chimie quantique et des simulations moléculaires peuvent être utilisés pour affiner les valeurs de contribution et améliorer la précision prédictive[6].

Les méthodes de contribution de groupe trouvent une application large dans la modélisation prédictive des propriétés de transport en raison de leur simplicité et de leur polyvalence [7]. Ces méthodes permettent l'estimation des propriétés de transport pour des composés où les données expérimentales sont limitées ou indisponibles, ce qui les rend des outils inestimables pour le génie chimique, la conception de procédés et le développement de formulations. De plus, les méthodes de contribution de groupe peuvent être intégrées dans des modèles computationnels plus larges pour prédire les propriétés de transport des mélanges complexes ou de systèmes polymères[8].

Malgré leurs avantages, il est important de reconnaître les limites des méthodes de contribution de groupe. La précision des prédictions dépend fortement de la qualité et de l'exhaustivité des valeurs de contribution attribuées à chaque groupe[9]. De plus, l'applicabilité des méthodes de contribution de groupe peut être limitée à des classes spécifiques de composés ou à certaines conditions de fonctionnement. Par conséquent, une attention particulière doit être accordée à la sélection des valeurs de contribution appropriées et à la validation des modèles prédictifs par rapport aux données expérimentales[10].

II.3.2 Hiérarchie des groupes

Les méthodes de contribution de groupe sont basées sur le principe que la propriété globale d'un composé peut être estimée en additionnant les contributions des groupes fonctionnels individuels présents dans la molécule. La précision de ces méthodes dépend en grande partie du type de schéma de fragmentation utilisé pour identifier les groupes fonctionnels.

II.3.2.A Schéma de fragmentation de l'ordre zéro

Le schéma de fragmentation de l'ordre zéro représente la forme la plus simple, où la molécule est fragmentée en atomes individuels, et la contribution de chaque atome à la propriété globale est considérée. Cette approche fournit une estimation basique des propriétés de transport en se contentant de citer le nombre d'atomes dans le composé. Bien que simple, le schéma de fragmentation de l'ordre zéro manque de la capacité à capturer les interactions structurelles et chimiques entre les atomes au sein de la molécule[11].

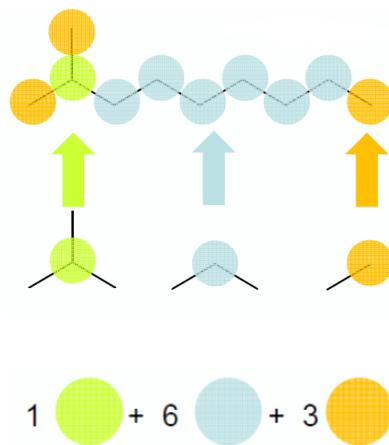


Figure 22: Exemple de découpe du 2-méthylnonane selon un schéma de découpe d'ordre zéro

II.3.2.B Schéma de fragmentation de premier ordre

Le schéma de découpage de premier ordre, illustré par la méthode de Joback [12], va un peu plus loin en considérant les groupes fonctionnels au lieu des atomes individuels. Un groupe fonctionnel est un groupe spécifique d'atomes au sein d'une molécule qui est responsable d'un certain motif de réactivité. En considérant les groupes fonctionnels, le schéma de découpage de premier ordre peut fournir des prédictions plus précises que le schéma de l'ordre zéro. Cependant, il ne prend toujours pas en compte les interactions entre différents groupes fonctionnels.

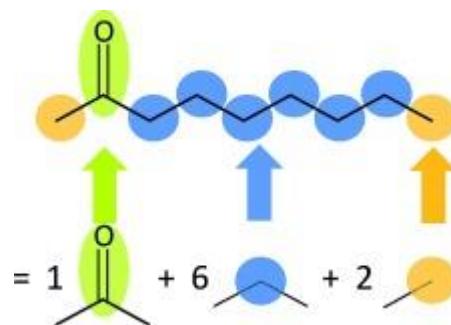


Figure 23: Exemple de découpe de 2-Nonanon selon un schéma de découpe du premier ordre

II.3.2.C Méthode de fragmentation de second ordre

Le schéma de découpage de second ordre, tel que la méthode de Benson, considère non seulement les groupes fonctionnels mais aussi leurs interactions adjacentes. Ce schéma

divise la molécule en sous-structures qui incluent deux groupes fonctionnels ou plus connectés. En prenant en compte les interactions entre les groupes fonctionnels adjacents, le schéma de découpage de second ordre peut fournir des prédictions encore plus précises[13].

Dans la fragmentation de second ordre, un groupe est constitué d'un atome central avec une valence strictement supérieure à 1, et d'un ensemble de liaisons qu'il forme avec ses voisins. Un groupe est noté comme suit :

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

où X représente l'atome central auquel sont liés i atome(s) A , j atome(s) B , k atome(s) C et 1 atome(s) D .

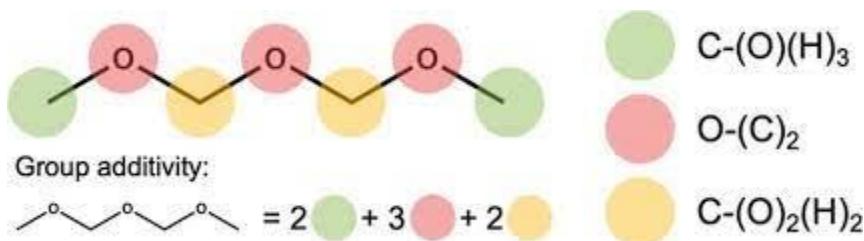


Figure 24: Exemple de découpe d'Oxymethylene-Ether selon un schéma de découpe du second ordre

II.3.2.D Méthode de fragmentation de troisième ordre

La méthode la plus largement utilisée dans le domaine des méthodes de contribution de groupe et ayant l'un des meilleurs schémas de partitionnement de structure moléculaire est la méthode de troisième ordre développée par Marrero et Gani (Marrero, 2001)[12] pour estimer les propriétés physico-chimiques et thermodynamiques avec un large éventail d'applications, traitant à la fois des molécules simples et des molécules multifonctionnelles ou même complexes.

La technique de fragmentation de troisième ordre d'une structure moléculaire est considérée comme un ensemble de trois types de groupes : groupes de premier ordre, groupes de second ordre et groupes de troisième ordre.

Le principe de cette méthode est présenté dans la représentation GC du composé L'acide 4-vinylbenzoïque (Figure 25).

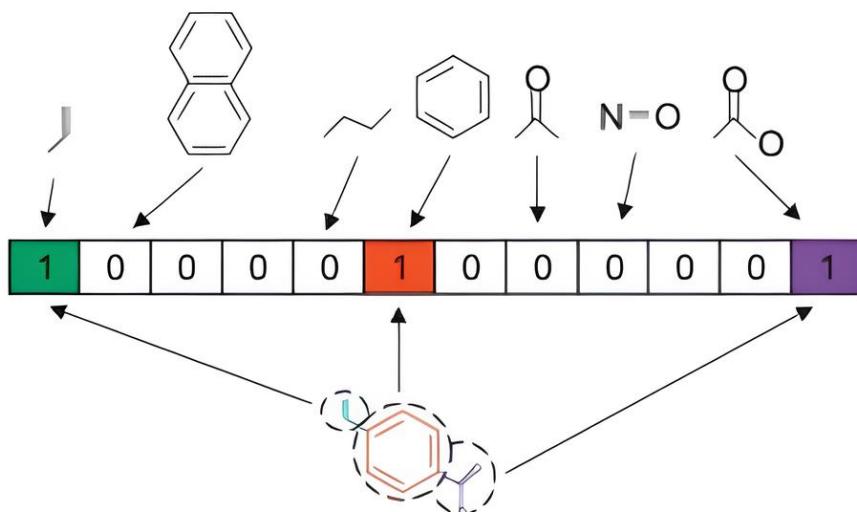


Figure 25: Découpe du composé L'acide 4-vinylbenzoïque par le Marrero et Gani

II.3.2.D.1 Groupes de premier ordre

Les groupes de premier ordre sont de nombreux groupes simples qui permettent de décrire une grande variété de composés organiques, y compris les hydrocarbures aromatiques et aliphatiques, les alcools, les phénols, les cétones, les aldéhydes, les acides, les esters, les éthers, les amines, les anilines, et de nombreuses autres classes de composés organiques. Cet ensemble de groupes doit également décrire les composés polyfonctionnels, grands, complexes et hétérocycliques. L'ensemble des groupes de premier ordre doit permettre de distinguer les groupes apparaissant dans des structures cycliques, aromatiques et acycliques. Par exemple, les groupes de premier ordre distinguent les atomes de carbone partagés par différents noyaux aromatiques dans le même système fusionné (par exemple, le naphtalène), les atomes de carbone partagés par des noyaux aromatiques et non aromatiques dans le même système fusionné (par exemple, l'indane), et tout autre carbone aromatique substitué qui n'appartient pas aux catégories précédentes (par exemple, le biphenyle). Les groupes de premier ordre doivent décrire la molécule entière. En d'autres termes, il ne doit pas y avoir de fragment d'une molécule donnée qui ne puisse être représenté par des groupes de premier ordre. Cependant, les groupes de premier ordre proposés décrivent partiellement les effets de proximité et la différence entre les isomères. Pour cette raison, l'ensemble des groupes de premier ordre est destiné à traiter des composés simples et monofonctionnels, et d'autres niveaux de groupes sont nécessaires pour fournir des estimations plus précises pour des complexes plus complexes.

II.3.2.D.2 Groupes de second ordre

Les groupes de second ordre permettent une meilleure description des composés polyfonctionnels et une différenciation entre les isomères. Ces groupes visent à décrire les fragments moléculaires qui ne peuvent pas être adéquatement représentés par les groupes de premier ordre, ce qui conduit à de mauvaises estimations au premier niveau. Dans ce cas, le second niveau utilise des groupes qui fournissent une meilleure représentation des composés polyfonctionnels. Par exemple, des groupes tels que $OH - CH_n - COO$, $CHm(NH_n) - COOH$ ou $NH_2 - CH_m - CHn - COOH$ ont été définis pour offrir une meilleure description des lactates et des acides aminés, qui aboutissent généralement à de mauvaises estimations en utilisant seulement les groupes de premier ordre. D'autres fragments moléculaires qui ne sont pas entièrement décrits par les groupes de premier ordre incluent les substituants alicycliques, pour lesquels des groupes tels que $CH_{cyc} - OH$, $CH_{cyc} - NH_2$, $CH_{cyc} - COO$, $CH_{cyc} - COOH$ ou $CH_{cyc} - SH$ ont été inclus dans l'ensemble de second ordre.

Comme suggéré ci-dessus, l'ensemble des groupes de second ordre devrait faciliter la différenciation entre les isomères. En conséquence, des groupes spécifiques sont fournis dans ce but. Ces groupes permettent de distinguer les isomères non seulement dans les alcènes, les alkynes et autres structures à chaîne ouverte, mais aussi dans les composés aromatiques, pour lesquels des groupes spéciaux tels que AROMARINGs1s2, AROMARINGs1s3 ont été ajoutés.

Cependant, les groupes de second ordre proposés ne peuvent pas fournir une bonne représentation des composés contenant plus d'un cycle et, dans certains cas, des composés polyfonctionnels à chaîne ouverte avec plus de quatre atomes de carbone dans la chaîne principale. Pour cette raison, un niveau supplémentaire est nécessaire pour permettre une meilleure description de ces types de composés.

II.3.2.D.3 Groupes de troisième ordre

Ces groupes permettent une représentation assez détaillée des systèmes d'anneaux aromatiques fusionnés, des systèmes d'anneaux aromatiques et non aromatiques fusionnés, et des systèmes d'anneaux non fusionnés connectés par des chaînes où divers groupes fonctionnels peuvent apparaître. Les critères utilisés pour identifier les groupes de troisième ordre sont analogues à ceux utilisés pour les groupes de second ordre.

Pour illustrer les différents schémas de découpage mentionnés ci-dessus, deux exemples illustratifs sont fournis.

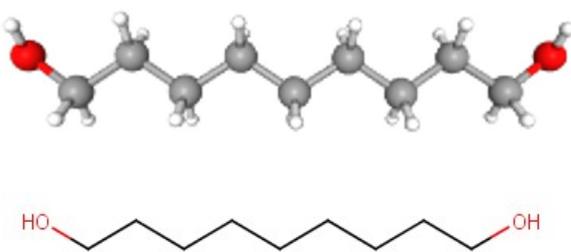


Figure 26: 1,9-nonadiol

Le découpage de cette molécule donne :

Tableau 2 : Découpage de la molécule 1,9-nonadiol en groupes

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

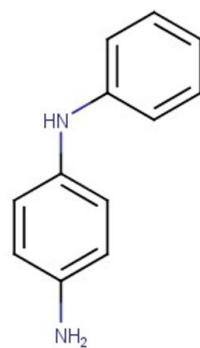
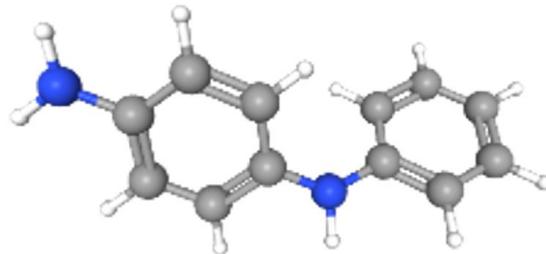


Figure 27: N-Phényl-1,4-benzènediamine

Le découpage de cette molécule donne :

Tableau 3 :Découpage de la molécule N-Phényl-1,4-benzènediamine en groupes

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 (differet rings)	1

II.3.3 Groupes fonctionnels

Pour l'estimation des propriétés des IL, plusieurs méthodes de contribution de groupe ont été proposées dans la littérature[14]. Ces méthodes ont l'avantage d'être rapides sans nécessiter de ressources informatiques importantes. Cependant, la précision de bon nombre de ces méthodes est encore discutable et leur applicabilité est limitée en raison de la simplification excessive des schémas de fragmentation résultant de la simple division de la structure moléculaire en cations et anions au lieu de groupes fonctionnels plus petits et clairement identifiés. Dans le cas des LI, qui sont des fluides très complexes, l'identification des groupes fonctionnels ne doit pas se faire avec une simple approximation du premier ordre car les effets de structure et de proximité seront négligés.

Dans ce travail, un nouveau schéma de fragmentation a été proposé, motivé par des travaux sur la prédiction des propriétés des liquides ioniques [15]et liquides non ioniques[16]. Ce nouveau schéma permet de décrire une grande variété d'IL et de distinguer les isomères. Chaque structure moléculaire d'une IL est considérée comme une collection de trois ensembles de groupes : les groupes cationiques, les groupes anioniques et les groupes substituants. Les groupes cationiques ou anioniques sont les groupes représentant des fragments chargés des ions, et les groupes substituants sont ceux représentant des fragments neutres de différents types de chaînes latérales. Ce concept a été mis en œuvre dans d'autres travaux [15].Dans ce travail, chaque ensemble de groupes est composé de deux types de groupes, à savoir les groupes de premier ordre et les groupes de second ordre. Les groupes de premier ordre sont proposés pour décrire une grande variété d'IL et capter les effets de proximité, tandis que le rôle des groupes de second ordre est de combler le manque d'informations structurelles qui n'est pas suffisamment

décrit par les groupes de premier ordre. Ces groupes de second ordre sont utilisés comme correction empirique des effets de proximité et de structure.

Ce nouveau schéma de fragmentation, simple à appliquer, évite les situations de chevauchement de groupes de premier ordre (chaque atome doit appartenir à un et un seul groupe de premier ordre). Figure 28 montre des exemples de groupes de premier et de second ordre utilisés pour représenter le 1-butyl-3-méthyl-1H-imidazol-3-ium 2-amino-3-hydroxypropanoate, [im-4,1] [ser]. Par exemple, pour représenter l'influence de la longueur de la chaîne dans ce cation, le groupe de second ordre $CH_3[CH_2]_n[aN]^+$ (n en 4, ..., 8) est considéré. Il convient de noter que tous les anions plus petits et certains des anions plus grands (complexes) ont été pris comme groupe de premier ordre sans les fragmenter. Dans cette étude, un total de 127 groupes ont été identifiés pour décrire toutes les structures moléculaires des IL dans la base de données. La liste de groupes comprend 52 groupes cationiques, 44 anioniques et 31 substituants, constitués de 79 groupes de premier ordre et 48 groupes de second ordre.

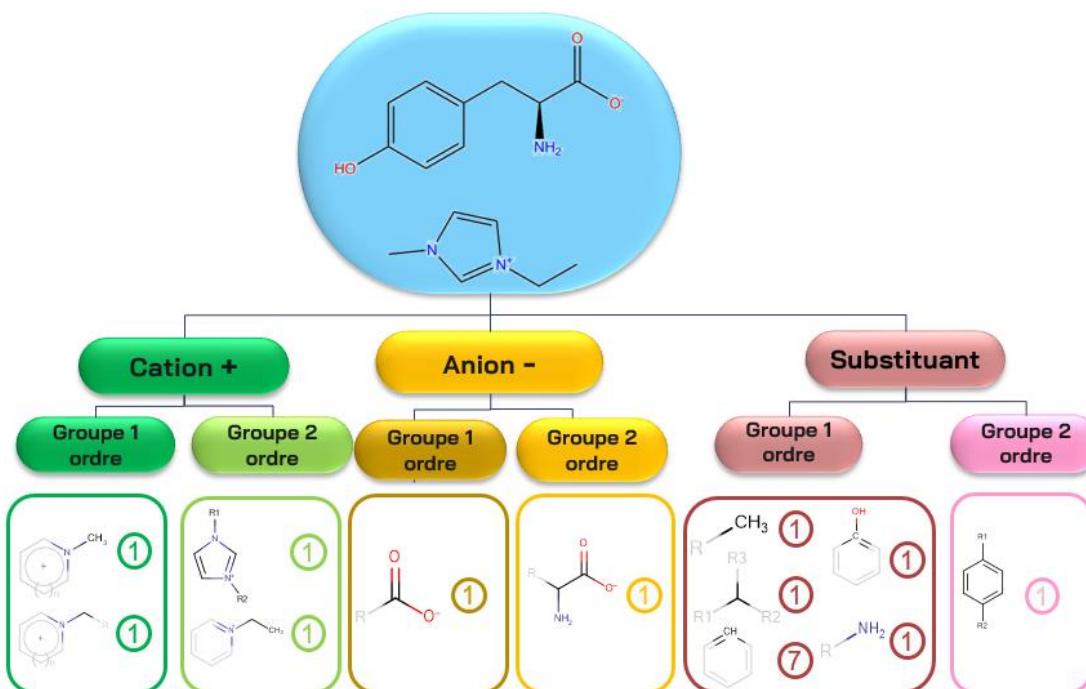


Figure 28: Exemple de nouveau schéma de découpage utilisés pour représenter le 3-éthyl-1-méthyl-1H-imidazol-3-ium L-tyrosinate

II.4 RELATIONS QUANTITATIVES STRUCTURE-PROPRIÉTÉ

II.4.1 Principe

Dans le domaine de la chimie computationnelle et de la modélisation prédictive, le principe des méthodes de Relation Quantitative Structure-Propriété (QSPR) est d'établir une relation mathématique entre les propriétés moléculaires, englobant à la fois les aspects électroniques et géométriques, connus sous le nom de descripteurs, et un observable macroscopique tel qu'une propriété physicochimique, une activité biologique ou une toxicité. Cette relation se formule comme suit :

$$\text{Propriété} = f(\text{Descripteurs})$$

Après l'établissement et la validation réussis de cette relation sur un ensemble de données de validation, il devient possible de l'utiliser pour prédire les propriétés de nouvelles molécules, même celles ne disposant pas de données expérimentales, ou pour des molécules qui n'ont pas encore été synthétisées. De plus, dans certains scénarios, de tels modèles peuvent faciliter une compréhension plus approfondie des phénomènes moléculaires impliqués dans la propriété d'intérêt.

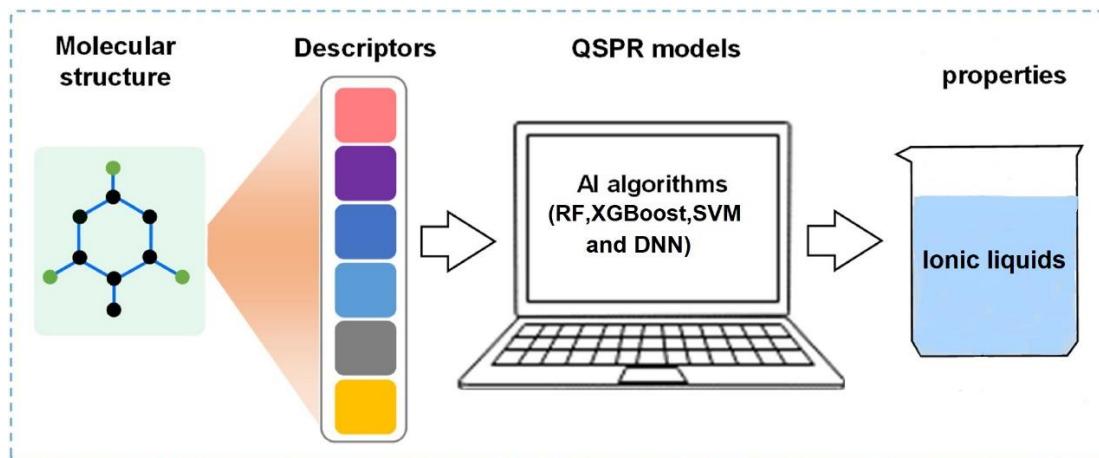


Figure 29: Représentation des méthodes QSPR

II.4.2 Descripteurs moléculaires

Les descripteurs moléculaires jouent un rôle crucial dans la modélisation quantitative des relations entre la structure chimique des composés et leurs propriétés physicochimiques. Ces descripteurs, également appelés variables descriptives, représentent diverses caractéristiques et attributs des molécules, tels que leur taille, leur forme, leur polarité, leur charge électrique, et bien d'autres. L'utilisation judicieuse de ces descripteurs permet de capturer les informations pertinentes nécessaires pour la prédiction précise des propriétés des composés chimiques.

Depuis des décennies, la recherche s'est concentrée sur le développement de descripteurs moléculaires complets visant à décrire en profondeur les structures moléculaires. Actuellement, la littérature répertorie des milliers de ces descripteurs [17]. Pour des informations détaillées sur ces descripteurs, il est possible de se référer aux travaux de Karelson (2012) [17]. Les descripteurs peuvent être classés comme suit : constitutionnels, topologiques, géométriques et quantiques[18].

II.4.2.A Descripteurs constitutionnels

Les descripteurs constitutionnels, largement utilisés en raison de leur simplicité conceptuelle et computationnelle, se concentrent sur la composition élémentaire et la connectivité atomique des molécules. Ils ne tiennent pas compte des aspects géométriques ou électroniques. Ces descripteurs incluent divers paramètres tels que la masse moléculaire, le nombre d'atomes (C, H, O, N, etc.), le nombre de groupes fonctionnels (CO, COOH, OH, NO₂, etc.) et le nombre de cycles (qu'ils soient polyatomiques ou non, aromatiques ou non).

Ces descripteurs forment la base des approches de contribution de groupe, où l'effet des substituants est introduit comme une valeur tabulée. Bien qu'ils soient utilisés pour obtenir les modèles QSAR/QSPR les plus simples, ils peuvent poser des problèmes pour une interprétation correcte des mécanismes d'interaction en jeu pour la propriété étudiée. Par exemple, ils ne permettent pas de distinguer les isomères.

II.4.2.B Descripteurs topologiques

Les descripteurs topologiques quantifient la structure moléculaire en examinant les connexions spatiales entre les atomes et les liaisons au sein d'une molécule. Ces

descripteurs capturent des informations importantes sur la connectivité atomique et la topologie de la molécule, ce qui est crucial pour modéliser les propriétés chimiques et physiques des composés organiques.

Quelques exemples de descripteurs topologiques couramment utilisés incluent :

Nombre de liaisons simples, doubles et triples : Ce descripteur compte le nombre de liaisons simples, doubles et triples présentes dans la molécule, fournissant une indication de la complexité structurelle de la molécule.

Distances entre les atomes : Les distances entre les atomes dans une molécule fournissent des informations sur la géométrie spatiale de la molécule et la proximité des différents groupes fonctionnels.

Indices topologiques : Les indices topologiques, tels que l'indice de Wiener et l'indice de Randic, sont des mesures mathématiques de la complexité topologique d'une molécule. Ils capturent des caractéristiques telles que la symétrie, la compacité et la connectivité de la molécule.

Nombre de cycles : Ce descripteur compte le nombre de cycles présents dans la molécule, fournissant des informations sur sa rigidité et sa stabilité conformationnelle.

En utilisant ces descripteurs topologiques, il est possible de caractériser quantitativement la structure moléculaire et de les utiliser pour prédire diverses propriétés des composés organiques, y compris les propriétés de transport. Ces descripteurs sont largement utilisés dans les domaines de la chimie computationnelle, de la modélisation moléculaire et de la conception de médicaments pour comprendre les relations structure-activité des composés chimiques.

II.4.2.C Descripteurs géométriques

Les descripteurs géométriques sont des paramètres utilisés pour caractériser la forme, la taille et l'arrangement spatial des molécules. Ces descripteurs fournissent des informations précieuses sur la géométrie tridimensionnelle des structures moléculaires, essentielles pour comprendre leurs propriétés chimiques et physiques. Des exemples de descripteurs géométriques fréquemment utilisés en chimie computationnelle et en chimiométrie incluent :

Surface totale : Ce descripteur quantifie la surface totale de la molécule, fournissant des informations sur son accessibilité pour les interactions avec d'autres molécules.

Volume : Les descripteurs de volume mesurent le volume total occupé par la molécule dans l'espace tridimensionnel, offrant des informations sur sa taille et sa masse.

Rayon de giration : Le rayon de giration est une mesure de la distribution de la masse autour du centre de masse de la molécule. Il fournit des informations sur la compacité et la flexibilité de la molécule.

Moments d'inertie : Les moments d'inertie caractérisent la distribution de la masse autour des axes principaux de rotation de la molécule. Ils offrent des informations sur la dynamique rotationnelle et la symétrie de la molécule.

Centre de masse : Ce descripteur indique le point dans l'espace où la masse de la molécule est concentrée, fournissant des informations sur sa position spatiale globale.

Les descripteurs géométriques sont des outils précieux pour étudier la conformation moléculaire, l'amarrage moléculaire et les simulations de dynamique moléculaire. En quantifiant les propriétés géométriques des molécules, ces descripteurs contribuent au développement de modèles prédictifs pour divers processus chimiques et biologiques.

II.4.2.D Descripteurs quantiques

Les descripteurs quantiques sont des paramètres dérivés de calculs de mécanique quantique qui fournissent des informations sur les propriétés moléculaires, telles que la structure électronique, les niveaux d'énergie et les orbitales moléculaires. Ces descripteurs sont précieux pour comprendre le comportement quantique des molécules et prédire leurs propriétés chimiques et physiques[5].

Quelques exemples de descripteurs quantiques couramment utilisés incluent :

Énergies HOMO et LUMO : Les énergies de la plus haute orbitale moléculaire occupée (HOMO) et de la plus basse orbitale moléculaire inoccupée (LUMO) fournissent des informations sur la structure électronique et la stabilité des molécules.

Potentiel d'ionisation (PI) : L'énergie requise pour retirer un électron d'une molécule, représentant sa tendance à subir une ionisation.

Affinité électronique (AE) : Le changement d'énergie lorsqu'un électron est ajouté à une molécule neutre, indiquant sa capacité à accepter des électrons.

Coefficients des orbitales moléculaires : Ces coefficients quantifient la contribution de chaque orbitale atomique à l'orbitale moléculaire globale, fournissant des informations sur la délocalisation électronique et les motifs de liaison.

Électronégativité : Dérivée des énergies des orbitales moléculaires ou des propriétés atomiques, l'électronégativité décrit la capacité d'un atome à attirer des électrons dans une molécule.

Moment dipolaire : Les calculs quantiques peuvent déterminer l'ampleur et la direction du moment dipolaire, qui caractérise la polarité d'une molécule.

Distribution de charge : Les descripteurs quantiques peuvent révéler la distribution de la densité électronique et des charges partielles au sein d'une molécule, influençant sa réactivité et ses interactions intermoléculaires.

Ces descripteurs quantiques sont cruciaux pour comprendre le comportement moléculaire, la réactivité et les interactions dans divers processus chimiques et physiques. Une compilation exhaustive de descripteurs quantiques tridimensionnels peut être trouvée dans une multitude de références bibliographiques[19][17]. Il convient de noter que le domaine de la mécanique quantique, en particulier en ce qui concerne les descripteurs quantiques tridimensionnels, nécessite l'utilisation de programmes de calcul fiables et sophistiqués.

II.4.3 Empreintes digitales dans la modélisation prédictive

En plus des descripteurs moléculaires, les empreintes digitales servent d'outils indispensables en chimiométrie et dans les études de relation quantitative structure-activité (QSAR), contribuant de manière significative aux tâches de modélisation prédictive. Les empreintes digitales représentent les structures moléculaires sous un format binaire ou catégorique, encodant des motifs structurels, des sous-structures et des propriétés moléculaires. Elles jouent un rôle crucial dans les recherches de similarité, les tâches de regroupement et de classification, complétant les informations fournies par les descripteurs moléculaires[18].

Dans cette thèse, l'utilisation de différents types d'empreintes digitales en complément des descripteurs moléculaires sera explorée pour développer des modèles prédictifs robustes des propriétés de transport.

En incorporant les deux types de représentations, les modèles visent à capturer une gamme plus large d'informations structurelles et basées sur les propriétés.

Dans la littérature scientifique, plusieurs types d'empreintes digitales sont couramment utilisés en chimiométrie et dans les études QSAR/QSPR, notamment les empreintes digitales des paires d'atomes, les empreintes digitales de Morgan, les clés MACCS et les empreintes pharmacophores. Dans cette thèse, nous nous concentrerons sur les empreintes digitales les plus fréquemment utilisées dans les tâches de modélisation prédictive, à savoir les empreintes digitales des paires d'atomes et les empreintes digitales de Morgan. Pour des détails complets sur ces empreintes digitales, nous recommandons de se référer aux publications suivantes[20][21].

II.4.3.A Empreintes digitales de Morgan

Les empreintes digitales de Morgan, également connues sous le nom d'empreintes de connectivité étendue (ECFP), sont des empreintes circulaires qui encodent des informations structurelles basées sur les environnements atomiques dans un rayon spécifié [20]. Les empreintes ECFP capturent des caractéristiques structurelles locales et globales, ce qui les rend polyvalentes pour diverses applications en chimiométrie.

Le principe de l'algorithme d'empreinte digitale de Morgan repose sur le concept de "rayon". L'algorithme commence par attribuer un identifiant initial à chaque atome de la molécule, généralement basé sur le type de l'atome. Ensuite, à chaque itération ou "rayon", l'algorithme considère le voisinage circulaire autour de chaque atome. Ce voisinage inclut l'atome lui-même et ses voisins immédiats à la première itération, les voisins de ces voisins à la deuxième itération, et ainsi de suite. La taille du voisinage s'élargit à chaque itération, d'où le terme "rayon".

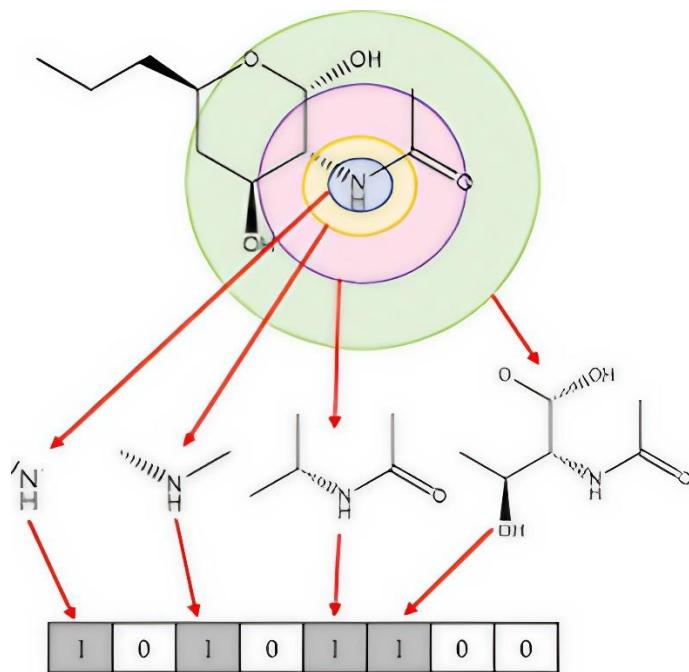


Figure 30: génération des modèles en utilisant des signatures moléculaires de hauteurs 0 à 3 pour chaque atome de la molécule

II.4.3.B Empreintes digitales des paires d'atomes

Les empreintes digitales des paires d'atomes encodent la présence ou l'absence de paires d'atomes à une certaine distance l'un de l'autre dans une molécule. Elles sont efficaces pour capturer les caractéristiques structurelles locales et sont largement utilisées dans la recherche de similarité et la modélisation QSAR/QSPR. Les empreintes digitales peuvent être représentées sous forme de vecteurs binaires ou de vecteurs de comptage. Dans un vecteur binaire, chaque position de bit correspond à la présence (1) ou à l'absence (0) d'une paire d'atomes particulière à une certaine distance. Dans un vecteur de comptage, chaque position stocke le nombre de paires d'atomes particulières à une certaine distance[21].

Cette méthode fournit une représentation détaillée de la structure de la molécule, capturant les positions relatives des atomes au sein de la molécule. Cependant, il est important de noter que les empreintes digitales des paires d'atomes ne considèrent pas les chemins entre les paires d'atomes, uniquement les distances.

II.4.4 Réduction de la dimensionnalité des descripteurs

Lorsque l'on traite une multitude de descripteurs, le potentiel de redondance augmente, entraînant des problèmes de colinéarité. Il est crucial que les paramètres sélectionnés soient non seulement significatifs mais aussi facilement interprétables d'un point de vue chimique. Les descripteurs sont considérés comme plus pertinents lorsqu'ils offrent des aperçus des mécanismes sous-jacents. L'objectif est de construire des modèles QSPR qui soient à la fois simples et transparents, mais aussi compréhensibles d'un point de vue phénoménologique[22].

Au cœur de cette démarche se trouve le principe de parcimonie, où la simplicité prévaut parmi les explications également plausibles d'un phénomène donné. Cependant, cette quête de simplicité ne doit pas impliquer une compromission informationnelle. En modélisation QSPR, elle plaide plutôt pour des modèles avec un minimum de paramètres tout en maximisant l'information encapsulée dans la propriété étudiée. De plus, une variable ne doit être retenue que si son exclusion diminue significativement les performances du modèle. Il faut toutefois veiller à ne pas perdre d'informations cruciales lors du processus de simplification du modèle[17].

Diverses techniques existent pour réduire le nombre de descripteurs inclus dans les modèles. Ces techniques se répartissent en trois grandes catégories : les méthodes de transformation, les méthodes de sélection de caractéristiques et les méthodes de régression parcimonieuses. Chacune de ces approches joue un rôle crucial dans la simplification de l'ensemble des descripteurs et l'amélioration de l'interprétabilité et de l'efficacité des modèles QSPR.

II.4.4.A Méthodes de transformation

La première catégorie de techniques de réduction de la dimensionnalité inclut les méthodes de transformation telles que l'analyse en composantes principales (PCA) et les moindres carrés partiels (PLS)[17]. La PCA transforme l'ensemble original de descripteurs en un nouvel ensemble de variables orthogonales appelées composantes principales, tandis que la PLS trouve un ensemble de variables latentes optimisées pour maximiser la covariance entre les descripteurs et la propriété d'intérêt. La PCA et la PLS réduisent la dimensionnalité de l'espace des descripteurs tout en préservant autant d'informations que possible, ce qui les rend efficaces pour la modélisation QSPR.

En plus de la PCA et de la PLS, plusieurs techniques peuvent être considérées. L'analyse en composantes indépendantes (ICA) se distingue par sa capacité à démêler des descripteurs multivariés en composantes indépendantes, réduisant ainsi l'espace des descripteurs en identifiant les composantes les plus indépendantes expliquant la variation des données. La factorisation en matrices non négatives (NMF) est une autre option viable, car elle factorise des données de haute dimension en deux matrices, l'une contenant des éléments non négatifs, facilitant l'extraction des caractéristiques tout en préservant les contraintes de non-négativité. Le PCA à noyau offre une extension non linéaire du PCA en projetant les données dans un espace de dimension supérieure à l'aide de méthodes à noyau avant d'appliquer la PCA. Chacune de ces techniques présente des avantages et des limitations uniques, ce qui rend le choix dépendant des caractéristiques des données et des objectifs spécifiques de l'étude.

II.4.4.B Sélection de caractéristiques

En plus des techniques de réduction de la dimensionnalité comme la PCA et la PLS, les méthodes de sélection de caractéristiques peuvent également être utilisées pour identifier les descripteurs les plus pertinents pour la modélisation QSPR. Ces méthodes évaluent l'importance de chaque descripteur individuellement et sélectionnent un sous-ensemble de descripteurs en fonction de leur pertinence pour la propriété prédite. Les approches courantes de sélection de caractéristiques incluent l'élimination récursive de caractéristiques (RFE), qui élimine de manière itérative les descripteurs les moins importants, la sélection de caractéristiques univariées, qui évalue chaque caractéristique individuellement à l'aide de tests statistiques univariés (par exemple, t-test, ANOVA, score de Fisher) et sélectionne les caractéristiques les plus fortement liées à la variable cible, et la régression linéaire multiple par étapes, qui combine la sélection progressive et l'élimination régressive, ajoutant ou supprimant des caractéristiques à chaque étape en fonction d'un critère choisi jusqu'à ce qu'aucune amélioration supplémentaire ne soit observée[23].

II.4.4.C Techniques de régression parcimonieuse

La troisième catégorie de techniques de réduction de la dimensionnalité inclut les techniques de régression parcimonieuse, qui visent à sélectionner un sous-ensemble des descripteurs les plus pertinents tout en encourageant la parcimonie dans les coefficients

de régression. Ces techniques, telles que LASSO (Least Absolute Shrinkage and Selection Operator), Elastic Net et la régularisation L1, pénalisent la taille absolue des coefficients de régression, sélectionnant ainsi efficacement un sous-ensemble de descripteurs avec des coefficients non nuls [24][25]. Les techniques de régression parcimonieuse peuvent aider à réduire le surajustement et à améliorer l'interprétabilité des modèles QSPR en identifiant les descripteurs les plus importants pour prédire la propriété d'intérêt[26].

II.5 MODÈLES ET RÉGRESSION

Le développement de modèles CG ou QSPR présente un défi significatif en raison des différences d'échelle inhérentes entre les structures moléculaires et les propriétés macroscopiques. Les structures moléculaires sont détaillées de manière complexe, existant à l'échelle moléculaire, tandis que les propriétés à prédire résident à un niveau macroscopique. Cette disparité d'échelle complique le processus de corrélation entre les deux, nécessitant une considération et une adaptation minutieuses des techniques de modélisation[1].

De plus, les incertitudes jouent un rôle crucial dans la modélisation CG ou QSPR, découlant à la fois des structures moléculaires, influencées par le niveau de précision des calculs, et des données expérimentales, qui peuvent varier en raison des protocoles de mesure. Ces incertitudes ajoutent une couche de complexité au processus de modélisation, nécessitant des méthodes robustes pour les gérer et les prendre en compte de manière efficace.

Un autre aspect crucial de la modélisation CG ou QSPR est la gestion de grands ensembles de données. Avec une multitude de descripteurs et de molécules à analyser, la sélection des paramètres structurels les plus pertinents à partir de l'ensemble complet disponible devient une tâche difficile. Bien que de nombreux outils et méthodologies existent à cette fin, le principal défi réside dans l'identification de l'approche la plus appropriée pour dériver un modèle fiable à partir de l'ensemble de données disponible.

La sélection d'une approche de modélisation appropriée dépend des caractéristiques spécifiques du système étudié. Étant donné la variabilité des structures moléculaires et des propriétés, il est essentiel de choisir une approche qui caractérise au mieux le

comportement du système. Ce processus de sélection implique l'évaluation de différentes stratégies de modélisation pour déterminer la méthode la plus efficace et la plus fiable pour tirer des informations significatives des données disponibles.

Les paragraphes suivants décrivent les principales méthodes de régression utilisées dans la modélisation CG et QSPR, qui se répartissent en deux catégories distinctes : les méthodes conventionnelles et les algorithmes d'apprentissage automatique.

II.5.1 Modèles conventionnels

Dans le domaine de la prédiction des propriétés de transport des composés organiques, les modèles traditionnels ou conventionnels ont été largement utilisés. Ceux-ci incluent les techniques de régression linéaire et non linéaire, qui établissent des relations entre les propriétés de transport et les variables telles que la température et la structure chimique (descripteurs ou contributions de groupe) [27][28]. L'équation d'Arrhenius, l'équation de Vogel-Fulcher-Tamman et d'autres modèles empiriques et semi-empiriques sont souvent utilisés pour décrire la dépendance des propriétés de transport à la température[29] . Ces modèles sont souvent basés sur des relations mathématiques non linéaires, telles que les fonctions puissance et exponentielle, pour caractériser le comportement des propriétés de transport des composés organiques. Des techniques telles que l'algorithme de moindres carrés non linéaires de Levenberg-Marquardt et l'algorithme des moindres carrés réitérés sont fréquemment appliqués pour optimiser les hyper-paramètres du modèle et améliorer la précision prédictive[24][30].

Il est important de noter que, bien que ces modèles traditionnels aient été efficaces dans de nombreux cas, ils peuvent avoir du mal à capturer le comportement non linéaire ou non idéal de certains composés organiques, en particulier les composés complexes ou nouveaux. Cela a conduit à l'exploration d'autres approches, telles que les algorithmes d'apprentissage automatique, pour améliorer la précision de la prédiction des propriétés de transport.

II.5.2 Algorithmes d'apprentissage automatique

Dans le domaine de la modélisation prédictive des propriétés physiques et chimiques des composés organiques, divers algorithmes d'apprentissage automatique ont été employés

pour développer des modèles de régression précis et robustes. Cette section discute de quatre méthodes d'apprentissage automatique parmi les plus utilisées dans ce contexte : Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine non linéaire (SVM) et réseau neuronal (NN).

II.5.2.A Random Forest

Random Forest (RF) est une méthode d'apprentissage par ensemble puissante largement utilisée pour les tâches de régression. Elle appartient à la classe des méthodes basées sur les arbres de décision et offre plusieurs avantages par rapport aux techniques de régression traditionnelles[22]. RF fonctionne en construisant une multitude d'arbres de décision pendant la phase d'entraînement[31] (Figure 31). Chaque arbre est construit en utilisant un sous-ensemble aléatoire des données d'entraînement et un sous-ensemble aléatoire des caractéristiques en utilisant la procédure de bootstrap, également connue sous le nom de bagging (bootstrap aggregation) [7] .Cette randomisation aide à réduire le surajustement et améliore la généralisation du modèle. Mathématiquement, RF peut être exprimée comme suit :

$$y(x) = \left(\frac{1}{N}\right) \sum_{i=1}^N T_{i(x)} \quad (42)$$

où $y(x)$ est la sortie prédite pour l'entrée x , $T_{i(x)}$ est la sortie du i -th arbre de décision dans la forêt, et N est le nombre total d'arbres dans la forêt.

Les RF pour la régression ont été utilisées avec succès pour construire des modèles GC non linéaires et des modèles QSPR en chimiométrie. Les RF disposent également de plusieurs hyper-paramètres ajustables qui peuvent être définis pour optimiser leurs performances. Le choix optimal de quelques hyper-paramètres que nous avons soigneusement pris en compte sont : *NumLearningCycles* and *MinLeafSize*.

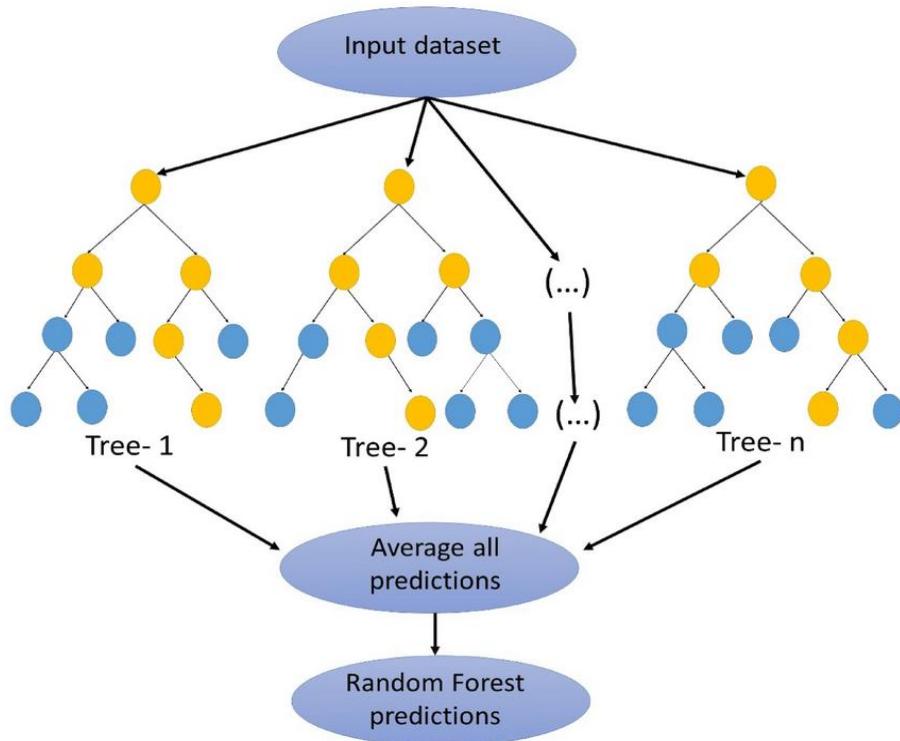


Figure 31: Schéma de fonctionnement Random Forest (RF).

II.5.2.B Boosting de Gradient Extrême

Le Boosting de Gradient Extrême (XGBoost) est une méthode d'ensemble basée sur les arbres de décision qui utilise le boosting, une technique permettant de combiner des apprenants faibles pour former un apprenant fort. XGBoost a gagné en popularité en raison de ses excellentes performances et de sa scalabilité [32][25].

La Figure 32 illustre la structure fondamentale de XGBoost, où X représente soit la matrice descripteur dans le cadre de l'approche QSPR, soit la matrice d'apparition de groupes dans le cadre de l'approche par contribution de groupes pour ce modèle, et y désigne la propriété de transport d'un composé. Dans le contexte du boosting, XGBoost utilise une stratégie d'entraînement additive. Chaque arbre est construit en apprenant à partir du résidu δ de l'arbre précédent. La prédition de la k -th itération, notée $\hat{y}_i^{(k)}$, est calculée comme la somme de la prédition de l'itération précédente, $\hat{y}_i^{(k-1)}$, et de la sortie de l'arbre actuel, $f_k(x_i)$.

XGBoost vise à optimiser le modèle à chaque itération pour réduire l'erreur de prédiction. La prédiction finale, \hat{y}_i , est produite en combinant les sorties de tous les arbres en utilisant une sommation pondérée :

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (43)$$

Ici, F est l'espace des fonctions qui inclut tous les arbres de régression, et K est le nombre total d'arbres dans le modèle. Pour entraîner chaque arbre et apprendre la fonction f_k , XGBoost définit une fonction objectif avec régularisation :

$$\mathcal{L}(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^K \Omega(f_k) \quad (44)$$

où θ est le paramètre du modèle, n est le nombre d'échantillons, $l(y_i, \hat{y}_i^{(t)})$ est la fonction de perte qui mesure la différence entre la valeur prédite $\hat{y}_i^{(t)}$ et la valeur réelle y_i , K est le nombre d'arbres, f_k est le k -th arbre. Le terme de régularisation $\Omega(f_k)$ pénalise les modèles complexes en ajoutant un terme de pénalité à la fonction objectif qui augmente avec la complexité du modèle. Ce terme prend la forme de :

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^K \omega_j^2 \quad (45)$$

où T est le nombre de nœuds terminaux dans l'arbre f_k , ω_j est le poids du j -th nœud terminal, et γ et λ sont des hyper-paramètres qui contrôlent la force de la régularisation. Le premier terme pénalise le nombre de nœuds terminaux dans l'arbre, tandis que le second terme pénalise les poids des nœuds terminaux.

L'algorithme fonctionne en ajoutant itérativement des arbres de décision au modèle, chaque arbre apprenant à partir des erreurs des arbres précédents. À chaque itération,

l'algorithme calcule le gradient et le Hessien de la fonction de perte par rapport à la valeur prédictée, et construit un nouvel arbre pour minimiser la fonction objectif. Le nouvel arbre est ensuite ajouté au modèle, et les valeurs prédictées sont mises à jour. Ce processus continue jusqu'à ce qu'un critère d'arrêt soit atteint.

Bien que XGBoost dispose de nombreux hyper-paramètres ajustables, il peut être démontré qu'une partie significative de ces hyper-paramètres n'est pas trop sensible à notre problème de modélisation. Cependant, le choix optimal de quelques hyper-paramètres standard peut significativement améliorer la précision de prédiction des modèles GC basés sur XGBoost. Voici les hyper-paramètres que nous avons soigneusement considérés :

MinLeafSize: Nombre minimum d'observations par noeud feuille, également connu sous le nom de `min_child_weight`. Chaque feuille a au moins *MinLeafSize* observations par feuille d'arbre.

NumLearningCycles: Le nombre de cycles d'apprentissage pour le boosting, également connu sous le nom de `num_round`.

LearnRate: Taux d'apprentissage pour le shrinkage, également connu sous le nom de `learning_rate`, il contrôle la taille des pas pendant les itérations de boosting.

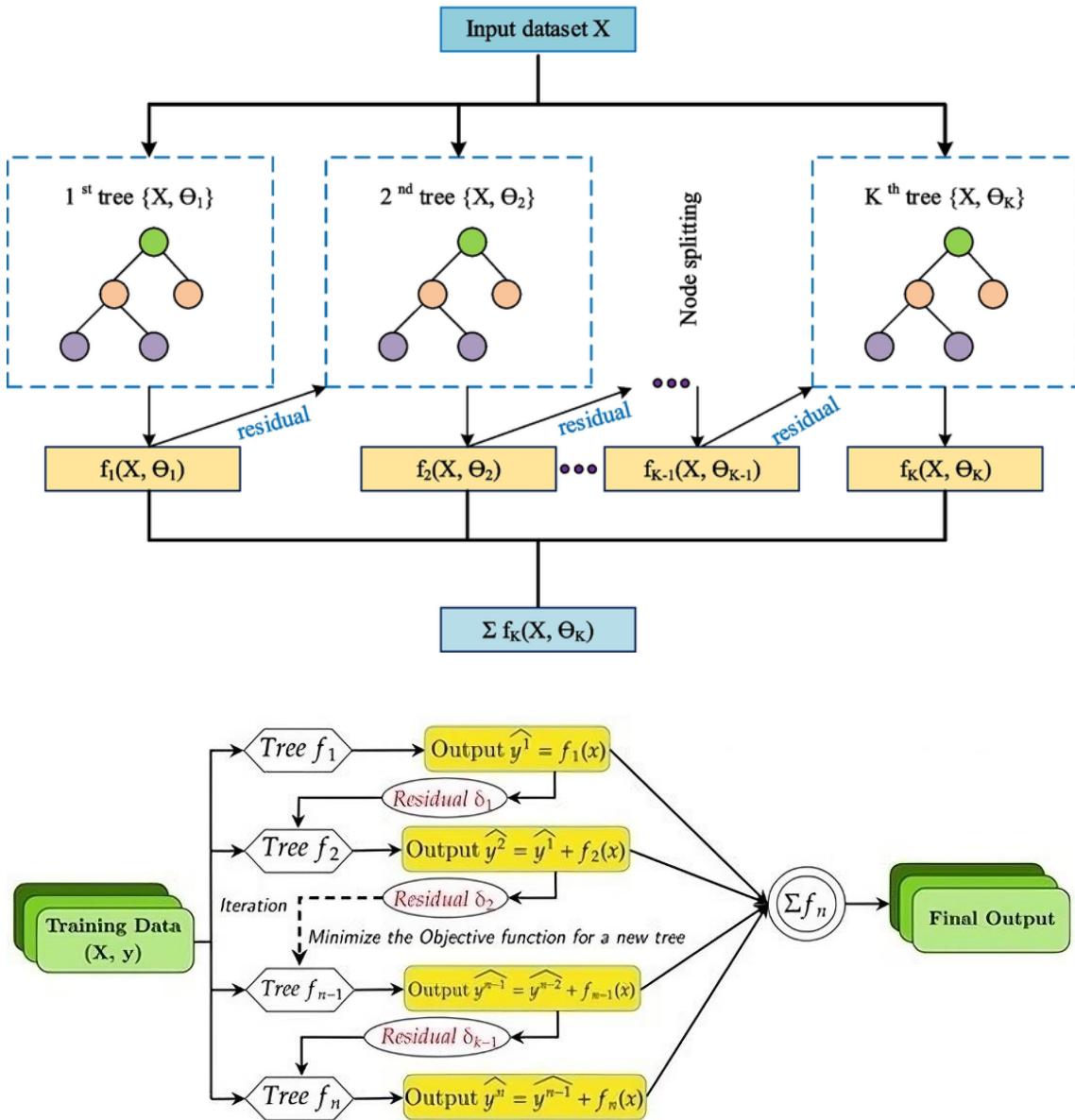


Figure 32: Schéma de fonctionnement de XGBoost [25].

II.5.2.C Non-linear Support Vector Machine

Les modèles de régression par machine à vecteurs de support (SVM) non linéaires sont un type d'algorithme d'apprentissage automatique utilisé pour prédire des valeurs de sortie continues en fonction d'un ensemble de variables d'entrée. [25] Ces modèles de régression peuvent capturer des relations complexes et non linéaires en utilisant des fonctions noyaux pour mapper les variables d'entrée dans un espace de caractéristiques de dimension supérieure. L'idée de base de la régression SVM est de trouver un hyperplan dans l'espace des caractéristiques qui sépare au mieux les données en deux régions, une

pour les valeurs de sortie positives et une pour les valeurs de sortie négatives. La distance entre l'hyperplan et les points de données les plus proches dans chaque région, appelés vecteurs de support, est maximisée. L'hyperplan peut ensuite être utilisé pour prédire la valeur de sortie pour de nouveaux points de données d'entrée [33][34].

Étant donné un ensemble d'entraînement $\{(x_1, y_1), \dots, (x_n, y_n)\}$, où $x_i \in R^d$ et $y_i \in R$, nous voulons apprendre une fonction $f: R^d \rightarrow R$ qui peut prédire la sortie y pour une nouvelle entrée x [35].

La fonction de décision du modèle de régression SVM peut être écrite comme (Figure 33)

$$f(x) = b + \sum_{i=1}^n \alpha_i K(x, x_i) \quad (46)$$

où b est le terme de biais, $K(x, x_i)$ est la fonction noyau qui mappe les données d'entrée dans un espace de caractéristiques de dimension supérieure, et α_i sont les coefficients des vecteurs de support. La fonction noyau $K(x, x_i)$ est définie comme :

$$K(x, x_i) = \varphi(x) \cdot \varphi(x_i) \quad (47)$$

où $\varphi(x)$ est une fonction de mapping de caractéristiques qui mappe les données d'entrée x dans un espace de caractéristiques de dimension supérieure.

Les coefficients α_i sont obtenus en résolvant le problème d'optimisation suivant :

$$\begin{aligned} & \min \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i y_i \\ & \text{subject to: } \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \in [0, C] \\ & \quad |y_i - f(x_i)| \leq \varepsilon \end{aligned} \quad (48)$$

où C est un hyper-paramètre de régularisation, également connu sous le nom de BoxConstraint, qui contrôle le compromis entre la maximisation de la marge et la minimisation de l'erreur de prédiction, et y_i est la valeur de sortie pour le point de données d'entrée x_i . Le paramètre ε est couramment connu sous le nom de hyper-paramètre d'insensibilité dans la régression SVM, et sa sélection nécessite une considération attentive pour équilibrer le compromis entre la complexité du modèle et la précision. En plus de C et ε , qui servent de hyper-paramètres pour un modèle SVM, la fonction noyau et l'échelle du noyau sont également utilisés comme hyper-paramètres pour affiner le SVM. Les fonctions noyaux les plus couramment utilisées dans les problèmes de GC et QSPR sont les fonctions de base radiale (RBF). Dans ce travail, le noyau RBF a été retenu après avoir testé plusieurs fonctions, donné par :

$$K(x_n, x) = \exp(-\gamma \cdot \|x_n - x\|^2) \quad (49)$$

où γ est le paramètre d'échelle du noyau (hyper-paramètre) du RBF.

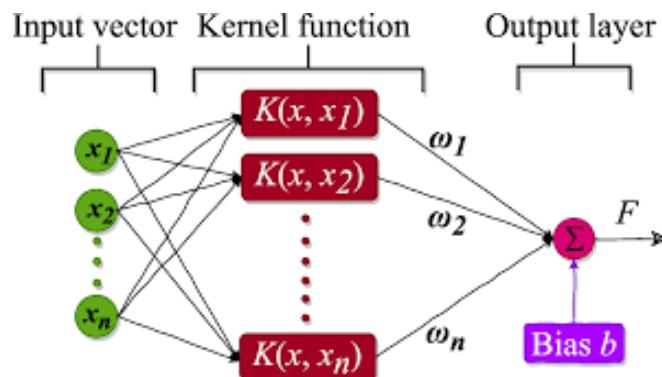


Figure 33: Architecture réseau d'une régression SVM

II.5.2.D Réseau de neurones

Les réseaux de neurones, également connus sous le nom d'algorithmes de perceptron multicouche, sont un outil puissant pour développer des modèles prédictifs, y compris ceux pour prédire des propriétés de transport des composés organiques. Les MLP font partie d'une classe plus large de réseaux de neurones artificiels conçus pour imiter le traitement de l'information des cerveaux animaux (Figure 34), en particulier en termes d'apprentissage à partir des données.

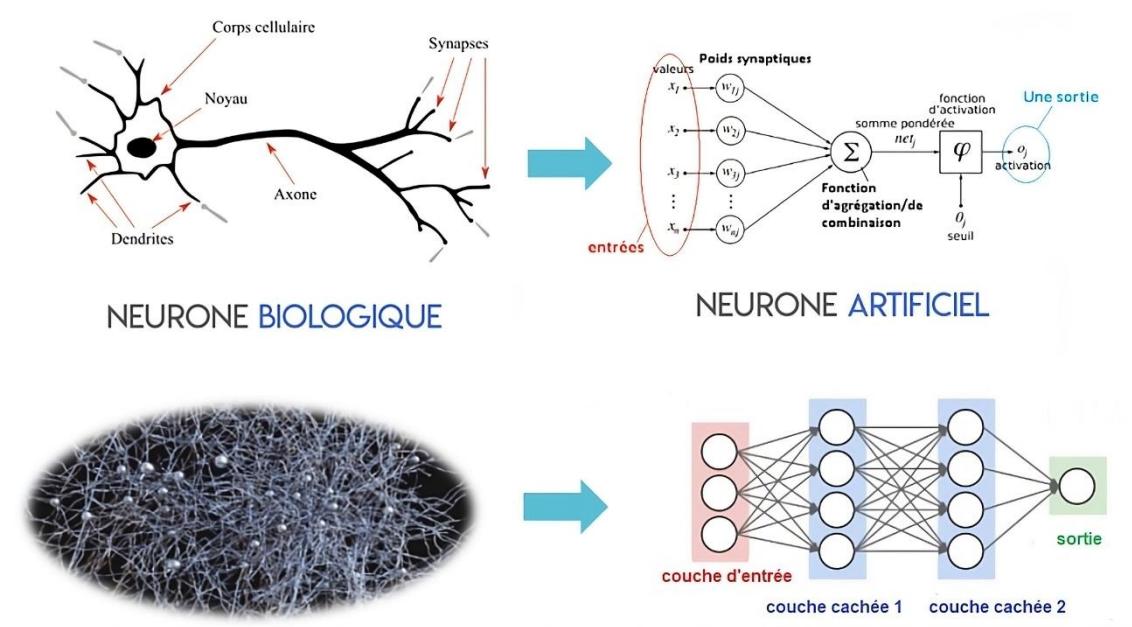


Figure 34: représentation illustrative des réseaux de neurones

Un MLP se compose d'une couche d'entrée, d'une ou plusieurs couches cachées et d'une couche de sortie. Chaque couche contient plusieurs nœuds (neurones), et chaque nœud est connecté aux nœuds des couches adjacentes. Les connexions ont des poids et des biais associés, qui sont ajustés pendant l'entraînement pour optimiser les performances du modèle [32].

Considérons un problème de régression où nous visons à prédire les propriétés de transport des composés organiques. Pour un MLP à une seule couche, la sortie du réseau peut être représentée comme :

$$y(\mathbf{x}) = \sum_{i=1}^D w_i x_i + b \quad (50)$$

où \mathbf{x} est le vecteur d'entrée représentant les caractéristiques du composé, w_i sont les poids associés à chaque caractéristique, b est le terme de biais, et D est le nombre de caractéristiques.

Dans un perceptron multicouche, la sortie de chaque nœud dans une couche cachée est calculée comme :

$$a_j^{(1)} = \sum_{i=1}^D w_{ji}^{(1)} x_i + b_j^{(1)} \quad (51)$$

où $a_j^{(1)}$ est l'activation du nœud j dans la première couche cachée, $w_{ji}^{(1)}$ sont les poids reliant la couche d'entrée à la première couche cachée, $b_j^{(1)}$ est le terme de biais pour le nœud j , et x_i sont les entrées provenant de la couche d'entrée.

L'activation $a_j^{(1)}$ est ensuite transformée en utilisant une fonction d'activation non linéaire $f(\cdot)$ pour produire la sortie de la couche cachée :

$$z_j^{(1)} = f(a_j^{(1)}) \quad (52)$$

Pour la couche de sortie, le processus est similaire, avec la sortie de chaque nœud dans la couche de sortie calculée comme :

$$a_k^{(2)} = \sum_{j=1}^m w_{kj}^{(2)} z_j^{(1)} + b_k^{(2)} \quad (53)$$

où $a_k^{(2)}$ est l'activation du nœud k dans la couche de sortie, $w_{kj}^{(2)}$ sont les poids reliant la couche cachée à la couche de sortie, $b_k^{(2)}$ est le terme de biais pour le nœud k , et $z_j^{(1)}$ sont les sorties de la couche cachée.

La sortie finale du réseau est ensuite calculée comme :

$$y(x) = f_{out}(a_{out}) \quad (54)$$

où a_{out} est l'activation du nœud de sortie et $f(\cdot)$ est une fonction d'activation appropriée, telle que la fonction identité pour les tâches de régression.

Les poids et les biais dans un MLP sont généralement appris à partir des données en utilisant un algorithme d'optimisation tel que la descente de gradient stochastique ou l'algorithme quasi-Newton limité-mémoire de Broyden-Fletcher-Goldfarb-Shanno [27]. Dans ce travail, ce dernier algorithme est considéré.

Il existe plusieurs hyper-paramètres qui peuvent être ajustés pour obtenir le meilleur modèle. Les plus importants, qui donnent un meilleur niveau de prédiction pour la plupart des modèles GC basés sur les réseaux de neurones profonds, sont expliqués ici :

N_Hidden (Nombre de couches cachées) : C'est le nombre de couches entre les couches d'entrée et de sortie.

Layer_N_Size (Nombre de neurones par couche) : C'est le nombre de neurones dans la N th couche cachée.

Activations (Fonctions d'activation) : Chaque neurone applique généralement une fonction d'activation à son entrée. Les choix courants incluent ReLU, sigmoïde et tanh. Le choix de la fonction d'activation peut affecter les performances du réseau.

Lambda (Force du terme de régularisation) : hyper-paramètre de régularisation qui contrôle la complexité du modèle en ajoutant une pénalité pour les poids des hyper-paramètres.

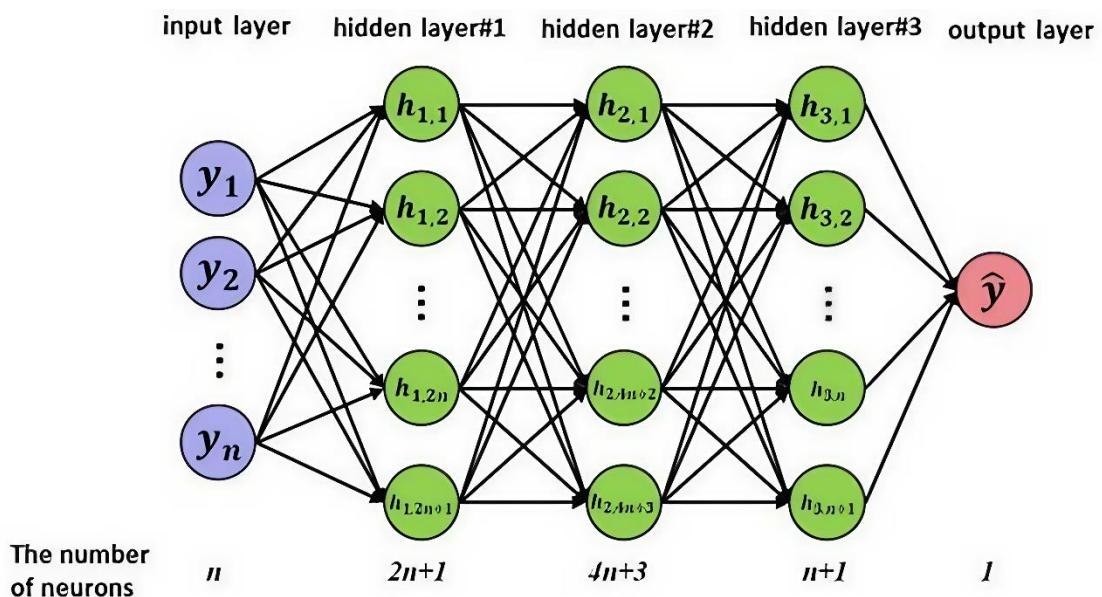


Figure 35: représentation d'un réseau de neurones avec trois couches cachées et un neurone de sortie.

II.5.2.D.1 Réseau de neurones profond

Dans ce travail, un réseau neuronal profond (DNN) entièrement connecté basé sur l'architecture de perceptron multicouche (MLP) est utilisé. Le réseau comprend plusieurs couches entièrement connectées, la première couche étant directement connectée à l'entrée du réseau, c'est-à-dire aux données du prédicteur. Chaque couche suivante reçoit une entrée de la couche précédente et applique une multiplication matricielle de poids suivie d'une addition de vecteur de biais. Une fonction d'activation est appliquée à la sortie de chaque couche entièrement connectée à l'exception de la dernière.[1].

Mathématiquement, un MLP à n couches peut être représenté comme:

$$y = f_n(W_n f_{\{n-1\}}(W_{\{n-1\}} \dots f_2(W_2 f_1(W_1 x + b_1) + b_2) \dots) + b_n) \quad (55)$$

où x est l'entrée du réseau, W_i et b_i sont les poids et biais pour la couche i , et f_i est la fonction d'activation utilisée dans la couche i . La sortie y est la prédiction finale du réseau.

Les poids et les biais dans un MLP sont généralement appris à partir de données à l'aide d'un algorithme d'optimisation tel que la descente de gradient stochastique ou le quasi-Newton de Broyden-Fletcher-Goldfarb-Shanno à mémoire limitée [23]. Dans ce travail, le dernier algorithme est considéré.

II.5.3 Réglage des hyper-paramètres

Tous les modèles de régression ont été implémentés à l'aide de la Statistics & Machine Learning Toolbox de MATLAB (MathWorks, Inc., version 2021b, licence académique) et tous les modèles utilisent des hyper-paramètres par défaut définis dans les fonctions de la bibliothèque, sauf indication contraire explicite. Chaque type de modèle d'apprentissage automatique utilisé dans cette étude comporte plusieurs dizaines d'hyper-paramètres réglables. Cependant, nous nous sommes concentrés spécifiquement sur les hyper-paramètres qui peuvent présenter des écarts par rapport aux valeurs par défaut dans les problèmes GC et QSPR, et nous les avons examinés en détail.

Afin d'affiner les hyper-paramètres pour une performance optimale, nous avons utilisé l'algorithme d'optimisation bayésienne très efficace [36]. Cet algorithme, connu pour ses capacités remarquables, utilise les résultats passés pour sélectionner intelligemment le prochain candidat hyper-paramètre, assurant un processus de recherche plus efficace. La méthode du CV à 5 fois et l'erreur quadratique moyenne (REQM) ont été appliquées comme mesure de la qualité de la performance du modèle pendant la phase de formation. Le Tableau 4 fournit les hyper-paramètres optimisés et leur espace de recherche pour chaque modèle.

Tableau 4 :Espace de recherche des hyper-paramètres et détermination de leur sélection optimale à l'aide de l'algorithme d'optimisation bayésien

Models	Search space	Selected parameters
XGBoost	$MinLeafSize = [1 \ 100]$, $NumLearningCycles = [10 \ 1000]$, $LearnRate = [1e-3, 1]$.	$MinLeafSize = 1$, $NumLearningCycles = 476$, $LearnRate = 0.38$.
RF	$MinLeafSize = [1 \ 100]$, $NumLearningCycles = [10 \ 1000]$.	$MinLeafSize = 1$, $NumLearningCycles = 800$
SVM	$C = [1 \ 500]$, $\varepsilon = [0.0001 \ 1]$, $\gamma = [0.1 \ 100]$.	$C = 294$, $\varepsilon = 0.023$, $\gamma = 3.072$
DNN	$N_Hidden = [1, 5]$, $Layer_N_Size = [2, 50]$, $Activations = [\text{ReLU, sigmoid, and tanh}]$, $Lambda = [0, 100]$.	$N_Hidden = 2$, $Layer_N_Size = [28 \ 10]$, $Activations = \text{tanh}$, $Lambda = 2.29e-07$.

II.6 VALIDATION DU MODÈLE

Une fois le modèle construit, il est important de le valider pour garantir sa robustesse et sa capacité à prédire avec précision les propriétés de transport des composés organiques. Cela peut se faire en divisant le jeu de données en un ensemble d'entraînement et un ensemble de test, et en évaluant les performances du modèle sur l'ensemble de test à l'aide de plusieurs paramètres de validation statistique.

Paramètres de validation statistique

Les performances des modèles proposés sont analysées, évaluées et discutées en termes de quatre mesures statistiques, à savoir, l'erreur absolu moyen (EAM), l'erreur absolu relatif moyen (ERAM), racine d'erreur quadratique moyenne (REQM) et le coefficient de détermination (R^2). Ces indicateurs de performance statistique sont exprimés comme suit:

$$EAM = \frac{1}{N} \sum_i |(y_i^{exp} - y_i^{pred})| \quad (56)$$

$$REQM = \sqrt{\frac{1}{N} \sum_i (y_i^{exp} - y_i^{pred})^2} \quad (57)$$

$$ERAM = \frac{1}{N} \sum_i \left| \frac{y_i^{exp} - y_i^{pred}}{y_i^{exp}} \right| \times 100\% \quad (58)$$

$$R^2 = 1 - \left(\frac{\sum_i (y_i^{pred} - y_i^{exp})^2}{\sum_i (y_i^{exp} - \mu^{exp})^2} \right) \quad (59)$$

où i est le i ème élément de données dans le jeu de données, N le nombre total de jeux de données, y_i^{exp} les valeurs expérimentales, y_i^{pred} les valeurs prédictives, μ^{exp} , la valeur moyenne calculée sur toutes les valeurs expérimentales.

L'erreur absolu moyen (EAM), tel que défini par sa formule (voir équation (56)), représente la moyenne des différences absolues entre les valeurs prédictives et expérimentales. Lors de la comparaison de plusieurs valeurs d'EAM, le modèle associé à l'EAM le plus bas est considéré comme le meilleur, car il indique une correspondance plus étroite entre les valeurs prédictives et observées. L'un des principaux avantages de l'utilisation de l'EAM est sa robustesse contre les valeurs aberrantes ou les erreurs très importantes, car il prend en compte les valeurs absolues des écarts plutôt que leurs valeurs carrées.

L'erreur quadratique moyenne (REQM), telle que définie par sa formule (voir équation (57)), représente la racine carrée de la moyenne des différences au carré entre les valeurs prédictives et expérimentales. Dans le domaine de la régression simple ou multiple, de

nombreux algorithmes sont conçus pour minimiser la REQM afin d'améliorer la précision du modèle. Lors de la comparaison des valeurs de REQM, une REQM plus faible indique un modèle plus précis. Cependant, il est crucial de reconnaître que la REQM est plus sensible aux valeurs aberrantes par rapport à racine d'erreur absolu moyen (REAM), car elle met les écarts au carré avant de prendre la moyenne. Cette sensibilité souligne l'importance de comprendre la distribution des données et l'impact des valeurs aberrantes sur les performances du modèle.

L'erreur absolu relatif moyen (ERAM), tel que défini par sa formule (voir équation (58)), est une métrique utilisée pour évaluer la précision d'un modèle prédictif en mesurant la moyenne des différences absolues entre les valeurs prédites et réelles, relatives aux valeurs réelles. Il est défini comme la moyenne des différences absolues entre les valeurs prédites et réelles divisées par les valeurs réelles, exprimée en pourcentage. Un ERAM plus faible indique un modèle plus précis. Contrairement à la REQM, qui met les erreurs au carré, l'ERAM fournit une mesure plus directe de la magnitude des erreurs par rapport aux valeurs réelles, ce qui le rend utile pour comprendre les performances du modèle en termes de précision relative. Cependant, comme la REQM, l'ERAM est sensible aux valeurs aberrantes et doit être interprété avec prudence, surtout dans les cas où les valeurs aberrantes peuvent avoir un impact significatif sur le calcul global de l'erreur.

Le coefficient de détermination, communément noté R^2 (voir équation (59)), sert de métrique cruciale pour évaluer l'efficacité d'un modèle prédictif. Il encapsule la proportion de la variance totale exhibée par la propriété sous considération qui est élucidée par le modèle lui-même. Essentiellement, R^2 est utilisé comme référence pour évaluer la capacité du modèle à décrire la distribution des points de données dans l'espace des paramètres. Les valeurs de R^2 se situent généralement entre 0 et 1, offrant une échelle complète pour évaluer les performances du modèle. Une valeur de 0 implique que le modèle n'explique aucune des variances dans la propriété, le rendant essentiellement insignifiant. À l'inverse, une valeur de R^2 de 1 indique un ajustement parfait, suggérant que le modèle capture précisément toute la variabilité présente dans le jeu de données.

II.6.1 Validation croisée et validation externe

Dans le cadre du développement de modèles prédictifs GC/QSPR pour la propriété des composés organiques, la validation croisée est une technique essentielle utilisée pour

évaluer les performances et la fiabilité de ces modèles. La validation croisée consiste à diviser le jeu de données en plusieurs sous-ensembles, ou plis, avec un pli réservé aux tests tandis que le modèle est entraîné sur les plis restants. Ce processus est répété plusieurs fois, chaque fois en utilisant un pli différent comme ensemble de test. Les résultats sont ensuite moyennés sur tous les plis pour obtenir une estimation plus robuste des performances du modèle. Les méthodes de validation croisée courantes incluent la validation croisée k-plis, où le jeu de données est divisé en k sous-ensembles de taille approximativement égale, et la validation croisée "leave-one-out" (LOOCV), où chaque point de données est utilisé comme échantillon de test une fois, le modèle étant entraîné sur tous les autres échantillons (Figure 36). Dans l'ensemble, la validation croisée aide à évaluer la capacité de généralisation du modèle, en s'assurant qu'il fonctionne bien sur des données non vues. Elle aide également à détecter le surajustement, où le modèle fonctionne bien sur les données d'entraînement mais mal sur de nouvelles données.

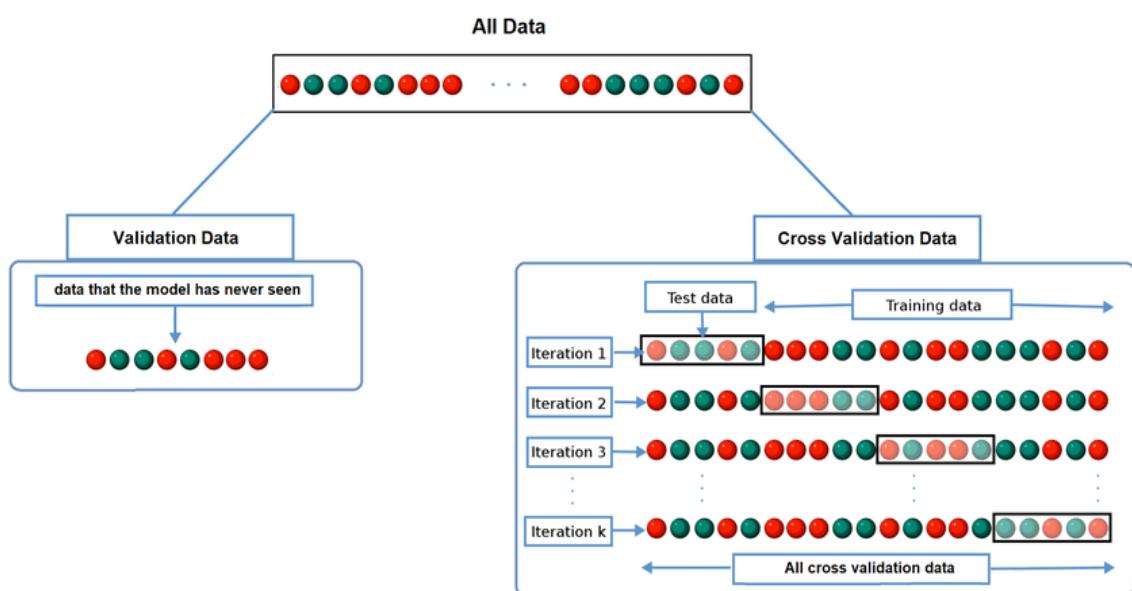


Figure 36: Représentation schématique de la validation croisée Leave-One-Out

De même, la validation peut être catégorisée en validation interne et externe, chacune ayant un objectif distinct dans l'évaluation des performances du modèle. La validation interne est utilisée pour l'optimisation du modèle, en s'assurant que le modèle est ajusté pour atteindre les meilleures performances possibles [37]. En revanche, la validation externe consiste à isoler les données de validation (également connues sous le nom de jeu de données de test) du processus de construction et d'optimisation du modèle (voir Figure

37). Cette approche vise à reproduire les conditions réelles dans lesquelles le modèle sera déployé, fournissant une évaluation plus précise de ses performances dans des applications pratiques. La validation externe est cruciale pour confirmer que le modèle peut bien se généraliser à des données non vues et fonctionner efficacement dans des environnements de production [38].

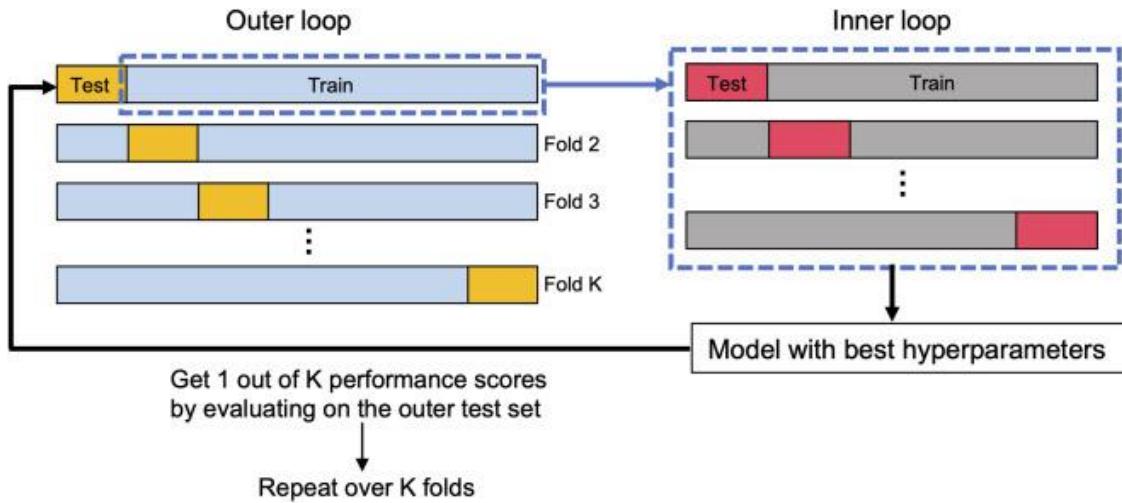


Figure 37: Organigramme de la procédure de formation (validation croisée) et de test (validation externe)

II.7 CONCLUSION

En conclusion, le chapitre 2 a fourni un aperçu complet de la méthodologie et des méthodes de modélisation essentielles pour le développement de modèles prédictifs des propriétés de transport des composés organiques. Le chapitre a souligné l'importance des bases de données et de la révision des données pour garantir l'exactitude et la fiabilité des modèles. Il a également exploré les méthodes de contribution de groupe, mettant en lumière la structure hiérarchique des groupes et leur rôle dans les schémas de fragmentation. Le chapitre a ensuite discuté des relations quantitatives structure-propriété (QSPR), en insistant sur l'importance des descripteurs moléculaires et des empreintes digitales dans la modélisation prédictive. De plus, les modèles conventionnels et les algorithmes d'apprentissage automatique tels que la forêt aléatoire, le boosting par gradient extrême, la machine à vecteurs de support non linéaire et le réseau de neurones ont été examinés en détail. Enfin, le chapitre a abordé la validation des modèles, y compris les paramètres de validation statistique et les techniques de validation croisée.

II.8 BIBLIOGRAPHIE

- [1] M.A. Zebida, K. Argoub, A.M. Benkouider, A. Yahiaoui, K. Toubal, A. Hachemaoui, Machine learning coupled with group contribution for predicting the electrical conductivity of ionic liquids with experimental accuracy, *Fluid Phase Equilib.* 579 (2024) 114024. <https://doi.org/10.1016/j.fluid.2024.114024>.
- [2] K. Argoub, A.M. Benkouider, A. Yahiaoui, F. Bagui, Estimation and uncertainty analysis of standard enthalpy of formation in the liquid state by third-order-group-contribution method, *Fluid Phase Equilib.* 520 (2020) 112644. <https://doi.org/10.1016/j.fluid.2020.112644>.
- [3] A.M. Benkouider, R. Kessas, S. Guella, A. Yahiaoui, F. Bagui, Estimation of the enthalpy of vaporization of organic components as a function of temperature using a new group contribution method, *J. Mol. Liq.* 194 (2014) 48–56. <https://doi.org/10.1016/j.molliq.2014.01.006>.
- [4] G. Shen, C. Held, J.P. Mikkola, X. Lu, X. Ji, Modeling the viscosity of ionic liquids with the electrolyte perturbed-chain statistical association fluid theory, *Ind. Eng. Chem. Res.* 53 (2014) 20258–20268. <https://doi.org/10.1021/ie503485h>.
- [5] A.F. Bouarab, J.P. Harvey, C. Robelin, Viscosity models for ionic liquids and their mixtures, *Phys. Chem. Chem. Phys.* 23 (2021) 733–752. <https://doi.org/10.1039/d0cp05787h>.
- [6] F. Akbari, M.M. Alavianmehr, R. Behjatmanesh Ardakani, D. Mohammad-Aghaie, Thermophysical properties of ionic liquids and their mixtures from a new equation of state, *Ionics (Kiel)*. 24 (2018) 1357–1369. <https://doi.org/10.1007/s11581-017-2310-8>.
- [7] M. Barycki, A. Sosnowska, A. Gajewicz, M. Bobrowski, D. Wileńska, P. Skurski, A. Giełdoń, C. Czaplewski, S. Uhl, E. Laux, T. Journot, L. Jeandupeux, H. Keppner, T. Puzyń, Temperature-dependent structure-property modeling of viscosity for ionic liquids, *Fluid Phase Equilib.* 427 (2016) 9–17. <https://doi.org/10.1016/j.fluid.2016.06.043>.
- [8] R. Haghbakhsh, K. Parvaneh, A. Shariati, 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, *Ind. Eng. Chem. Res.* 56 (2017) 2247–2258. <https://doi.org/10.1021/acs.iecr.6b04193>.
- [9] F. Gharagheizi, S.A. Mirkhani, M.H. Keshavarz, N. Farahani, K. Tumba, A molecular-based model for prediction of liquid viscosity of pure organic compounds: A quantitative structure property relationship (QSPR) approach, *J. Taiwan Inst. Chem. Eng.* 44 (2013) 359–364. <https://doi.org/10.1016/j.jtice.2012.12.015>.
- [10] F.Z. Serat, A.M. Benkouider, A. Yahiaoui, F. Bagui, Nonlinear group contribution model for the prediction of flash points using normal boiling points, *Fluid Phase Equilib.* 449 (2017) 52–59. <https://doi.org/10.1016/j.fluid.2017.06.008>.
- [11] K. Argoub, A.M. Benkouider, A. Yahiaoui, R. Kessas, S. Guella, F. Bagui, Prediction of standard enthalpy of formation in the solid state by a third-order group contribution method, *Fluid Phase Equilib.* 380 (2014) 121–127. <https://doi.org/10.1016/j.fluid.2014.08.003>.
- [12] J. Marrero, R. Gani, Group-contribution based estimation of pure component properties, *Fluid Phase Equilib.* 183 (2001) 183–208.

- [13] K.J. Wu, C.X. Zhao, C.H. He, Development of a group contribution method for determination of thermal conductivity of ionic liquids, *Fluid Phase Equilib.* 339 (2013) 10–14. <https://doi.org/10.1016/j.fluid.2012.11.024>.
- [14] K. Paduszyński, Extensive Databases and Group Contribution QSPRs of Ionic Liquid Properties. 3: Surface Tension, *Ind. Eng. Chem. Res.* 60 (2021) 5705–5720. <https://doi.org/10.1021/acs.iecr.1c00783>.
- [15] K. Paduszyński, Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity, *Ind. Eng. Chem. Res.* 58 (2019) 17049–17066. <https://doi.org/10.1021/acs.iecr.9b03150>.
- [16] A.M. Benkouider, R. Kessas, S. Guella, A. Yahiaoui, F. Bagui, Estimation of the enthalpy of vaporization of organic components as a function of temperature using a new group contribution method, *J. Mol. Liq.* 194 (2014) 48–56.
- [17] M. Dehmer, K. Varmuza, D. Bonchev, Edited by Titles of the series “ Quantitative and Network Biology ” Applied Statistics for Network Biology Drug Design of Zinc-Enzyme Inhibitors Molecular Descriptors for Chemoinformatics Molecular Modelling for Beginners, 2012.
- [18] Danishuddin, A.U. Khan, Descriptors and their selection methods in QSAR analysis: paradigm for drug design, *Drug Discov. Today.* 21 (2016) 1291–1302. <https://doi.org/10.1016/j.drudis.2016.06.013>.
- [19] L. Wang, J. Ding, L. Pan, D. Cao, H. Jiang, X. Ding, Quantum chemical descriptors in quantitative structure–activity relationship models and their applications, *Chemom. Intell. Lab. Syst.* 217 (2021) 104384. <https://doi.org/10.1016/j.chemolab.2021.104384>.
- [20] D. Rogers, M. Hahn, Extended-Connectivity Fingerprints, (2010) 742–754.
- [21] D.H. Smith, R.E. Carhart, R. Venkataraghavan, Atom Pairs as Molecular Features in Structure-Activity Studies: Definition and Applications, *J. Chem. Inf. Comput. Sci.* 25 (1985) 64–73. <https://doi.org/10.1021/ci00046a002>.
- [22] V. Svetnik, A. Liaw, C. Tong, J. Christopher Culberson, R.P. Sheridan, B.P. Feuston, Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling, *J. Chem. Inf. Comput. Sci.* 43 (2003) 1947–1958. <https://doi.org/10.1021/ci034160g>.
- [23] M. Li, J. Chen, Y. Huang, M. Li, X. Lin, T. Qiu, Reusable and efficient heterogeneous catalysts for biodiesel production from free fatty acids and oils: Self-solidifying hybrid ionic liquids, *Energy.* 211 (2020) 118631. <https://doi.org/10.1016/j.energy.2020.118631>.
- [24] I. Daubechies, R. Devore, M. Fornasier, C.S. Güntürk, Iteratively reweighted least squares minimization for sparse recovery, *Commun. Pure Appl. Math.* 63 (2010) 1–38. <https://doi.org/10.1002/cpa.20303>.
- [25] F. Cheng, C. Yang, C. Zhou, L. Lan, H. Zhu, Y. Li, Simultaneous determination of metal ions in zinc sulfate solution using UV–VIS spectrometry and SPSE-xgboost method, *Sensors (Switzerland)*. 20 (2020) 1–14. <https://doi.org/10.3390/s20174936>.
- [26] M. Panik, Nonlinear Regression, 2020. <https://doi.org/10.1201/9781420091984-21>.
- [27] M.P.W. and T.W. Maase, Ionic Liquids in Synthesis Edited, 2007. <https://doi.org/10.1002/9783527621194.ch8>.
- [28] A.T.K. and A. Mehrkesh, Computer-Aided Design of Tailor-Made Ionic Liquids, *AIChE J.* 59

- (2013) 4627–4640. <https://doi.org/10.1002/aic>.
- [29] P. Nancarrow, A. Al-Othman, D.K. Mital, S. Döpking, Comprehensive analysis and correlation of ionic liquid conductivity data for energy applications, *Energy*. 220 (2021). <https://doi.org/10.1016/j.energy.2021.119761>.
- [30] P. Nancarrow, A. Al-Othman, D.K. Mital, S. Döpking, Comprehensive analysis and correlation of ionic liquid conductivity data for energy applications, *Energy*. 220 (2021). <https://doi.org/10.1016/j.energy.2021.119761>.
- [31] X. Zhang, H. Shen, T. Huang, Y. Wu, B. Guo, Z. Liu, H. Luo, J. Tang, H. Zhou, L. Wang, W. Xu, G. Ou, Improved random forest algorithms for increasing the accuracy of forest aboveground biomass estimation using Sentinel-2 imagery, *Ecol. Indic.* 159 (2024) 111752. <https://doi.org/10.1016/j.ecolind.2024.111752>.
- [32] R. Nakhaei-Kohani, S. Ali Madani, S.P. Mousavi, S. Atashrouz, A. Abedi, A. Hemmati-Sarapardeh, A. Mohaddespour, Machine learning assisted Structure-based models for predicting electrical conductivity of ionic liquids, *J. Mol. Liq.* 362 (2022). <https://doi.org/10.1016/j.molliq.2022.119509>.
- [33] C. Nieto-Draghi, G. Fayet, B. Creton, X. Rozanska, P. Rotureau, J.C. De Hemptinne, P. Ungerer, B. Rousseau, C. Adamo, A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes, *Chem. Rev.* 115 (2015) 13093–13164. <https://doi.org/10.1021/acs.chemrev.5b00215>.
- [34] F. Gharagheizi, P. Ilani-Kashkouli, M. Sattari, A.H. Mohammadi, D. Ramjugernath, D. Richon, Development of a LSSVM-GC model for estimating the electrical conductivity of ionic liquids, *Chem. Eng. Res. Des.* 92 (2014) 66–79. <https://doi.org/10.1016/j.cherd.2013.06.015>.
- [35] L. Song, H. Zhang, J. Chen, Z. Li, M. Guan, H. Qiu, Imidazolium ionic liquids-derived carbon dots-modified silica stationary phase for hydrophilic interaction chromatography, *Talanta*. 209 (2020) 120518. <https://doi.org/10.1016/j.talanta.2019.120518>.
- [36] J. Fan, J. Zeng, A Levenberg-Marquardt algorithm with correction for singular system of nonlinear equations, *Appl. Math. Comput.* 219 (2013) 9438–9446. <https://doi.org/10.1016/j.amc.2013.03.026>.
- [37] J.A. Lazzús, Estimation of the thermal conductivity $\lambda(T, P)$ of ionic liquids using a neural network optimized with genetic algorithms, *Comptes Rendus Chim.* 19 (2016) 333–341. <https://doi.org/10.1016/j.crci.2015.09.010>.
- [38] J. Albert, K. Müller, Thermal conductivity of Ionic Liquids: An estimation approach, *Chem. Eng. Sci.* 119 (2014) 109–113. <https://doi.org/10.1016/j.ces.2014.08.023>.

CHAPITRE III : PREDICTION DE LA CONDUCTIVITE ELECTRIQUE DES LIQUIDES IONIQUES

CHAPITRE III: PREDICTION DE LA CONDUCTIVITE ELECTRIQUE DES LIQUIDES IONIQUES

III.1 INTRODUCTION

Les liquides ioniques (LI) constituent une classe de composés aux propriétés physico-chimiques uniques, notamment leur conductivité électrique, qui les rend particulièrement attractifs pour des applications en électrochimie, catalyse et procédés de séparation [1][2] [3]. Cependant, la prédiction précise de cette propriété reste un défi majeur en raison de la complexité des interactions ioniques et de la diversité des structures moléculaires impliquées[4][5]. Ce chapitre vise à développer et évaluer des modèles prédictifs robustes pour la conductivité électrique des LI, en exploitant des approches computationnelles avancées et une base de données expérimentales extensive.

Ce travail s'inscrit dans la continuité des recherches existantes tout en apportant des innovations méthodologiques significatives. Il poursuit trois objectifs principaux : (i) constituer une base de données complète et fiable, (ii) développer et comparer des modèles prédictifs basés sur l'apprentissage automatique et des méthodes conventionnelles, et (iii) analyser les performances de ces modèles en fonction des familles chimiques et des conditions expérimentales. La structure du chapitre reflète ces ambitions : la section sur la collecte et le traitement des données détaille la constitution d'une base de 6014 points pour 373 LI, suivie d'une analyse des influences de la température, de la structure ionique et de la pression. La méthodologie expose les cinq approches de modélisation déployées ; Extreme Gradient Boosting, Random Forest, Support Vector Machine (SVM), réseaux de neurones profonds (DNN) et modèle conventionnel, tandis que les résultats et discussions évaluent leurs performances et

contributions par rapport à la littérature. Une conclusion est avancée à la fin de ce chapitre.

III.2 BASE DE DONNÉES EXPÉRIMENTALES

III.2.1 Collecte de données:

Les données utilisées dans cette étude ont été collectées à partir de la base de données en accès libre ILThermo (version 2.0), qui contient les propriétés thermodynamiques des liquides ioniques (LI). Nous avons développé un script Python maison basé sur la bibliothèque Python pyILT2 pour extraire automatiquement les données de la base de données ILThermo (consultée le 2 octobre 2022). Ainsi, une vaste base de données sur la conductivité électrique a été obtenue à partir de la base de données en ligne ILThermo, contenant **7 598 points de données expérimentales** correspondant à **520 LI différents** appartenant à **12 et 14 familles chimiques différentes de cations et d'anions**, respectivement. Environ 95,5 % des points de données ont été enregistrés sans répétition.

La base de données a été organisée selon la structure de la base ILThermo. Pour chaque liquide ionique, les données ont été regroupées en jeux de données numérotés selon leur source bibliographique d'origine. Chaque jeu inclut des informations sur l'état physique du LI, la préparation et la pureté des échantillons, ainsi que la technique expérimentale utilisée pour mesurer la conductivité électrique. Cette organisation est essentielle pour une revue, interprétation et évaluation précise des données.

Enfin, la base de données a été indexée par nom de liquide et par jeu de données, incluant des informations clés telles que l'abréviation, la formule chimique, l'incertitude, la source bibliographique, la méthode expérimentale, ainsi que les conditions de température et de pression. Pour une liste détaillée de cette base et des informations supplémentaires, le lecteur est invité à consulter l'annexe 1.

III.2.2 Révision des données

Avant d'élaborer les nouveaux modèles, les données expérimentales brutes téléchargées ont été analysées et examinées en profondeur en raison de certaines divergences dans les données déclarées. Ces écarts étaient probablement dus aux impuretés dans les échantillons, aux procédés de purification ou à la diversité des techniques expérimentales

utilisées pour mesurer la conductivité électrique des IL. Pour y remédier, une procédure de révision des données comprenant sept étapes consécutives a été adoptée.

La première étape consistait à exclure les cristaux et les liquides métastables. La deuxième étape visait à éliminer les doublons dans chaque ensemble de données. La troisième étape consistait à trouver des valeurs avec un comportement contradictoire, c'est-à-dire des valeurs multiples pour le même liquide à des températures et pressions identiques. Dans ce cas, une seule valeur représentant la moyenne des valeurs a été retenue. La quatrième étape visait à éliminer les données avec une incertitude relative de plus de 100 %, tandis que la cinquième étape visait à éliminer les données liquides avec moins de trois points de données. Dans la sixième étape, une régression a été effectuée sur les données restantes pour chaque IL à la pression atmosphérique (100-101,325 kPa) en utilisant l'équation de Vogel-Fulcher-Tammann. Ensuite, une valeur aberrante a été considérée comme n'importe quel point de données avec un résidu studentisé en dehors de 2 fois la plage interquartile (IQR), et cette étape a été répétée jusqu'à ce qu'aucune valeur aberrante n'ait été détectée. Dans la dernière étape de ce flux de travail, les données à des pressions autres que atmosphériques ont été vérifiées manuellement et les données acceptées à l'aide de ce flux de travail ont également été réexaminées manuellement pour éliminer les données potentiellement inappropriées [6].

Après la procédure de révision, une base de données de conductivité électrique relativement étendue et diversifiée a été générée, composée de **6014 points** de données expérimentales pour **373 LI de 183 cations et 80 d'anions**. Au cours des étapes précédentes, les nombres de points suivants ont été éliminés : 430, 152, 180, 23, 149 et 650, respectivement. Des renseignements supplémentaires sur la régression de la sixième étape ont été fournis en détail dans l'annexe 2.

Afin de permettre aux lecteurs de suivre l'avancement de la révision des données pour chaque point de données, une valeur comprise entre 0 et 7 est attribuée à chaque point de données de la base de données en fonction de l'étape de révision pertinente. Une valeur de 0 indique que le point de données n'est pas rejeté, tandis qu'une valeur de n signifie que le point de données est rejeté à l'étape *n*. Ces valeurs ont été compilées dans un vecteur appelé « **Révision** », qui a été intégré à la base de données. Une nouvelle colonne nommée « **Elect_After_Modif** » a également été introduite pour stocker la moyenne des valeurs en conflit détectées à la troisième étape.

A l’issu de l’étape de révision, la base de données de conductivité électrique est présentée en fonction des familles d’ions, voir Tableau 5. Elle comprend des informations telles que le nombre de points de données, les ensembles de données, le nombre des LI, la plage de température, la plage de pression, la plage de conductivité électrique et la plage d’incertitude de conductivité électrique. Comme on peut le voir sur le Tableau 5, la base de données comprend une vaste gamme de valeurs de conductivité électrique, allant de 1,18 E-05 à 144,6 S/m, des températures de 219,15 à 495 K et des pressions de 89,3 à 250900 kPa. La Figure 38 montre les structures chimiques des familles de cations et d’anions. Pour une compréhension plus complète, les Figure 39 et Figure 40 présentent des distributions détaillées des LI entre diverses combinaisons de familles de cations et d’anions chimiques, avant et après le traitement des données. Il convient toutefois de noter que cette répartition n’est pas uniforme, certaines combinaisons cation-anion étant plus répandues que d’autres. Les combinaisons cation-anion les plus courantes sont celles contenant de l’imidazolium, du pyridinium, du pyrrolidinium, de l’ammonium, des bistriflamides (dérivés NTf₂) et des dérivés BF₄.

Tableau 5: Résumé de la base de données sur la conductivité électrique par familles d'ions

IL family	Data points	Data sets	Molecules	ILs	T min (K)	T max (K)	P min (kPa)	P max (kPa)	σ min (S/m)	σ max (S/m)	$\Delta\sigma$ min (S/m)	$\Delta\sigma$ max (S/m)
Cations												
ammonium (n)	687	81	38	63	234.15	443.1	100	1.0E+02	1.3E-04	8.8E+00	0.0E+00	2.0E-01
guanidinium (guan)	28	4	4	4	283.15	313.15	100	101.3	3.4E-02	1.6E-01	4.0E-04	2.0E-03
imidazolium (im)	3162	314	47	150	238.05	495.35	89.3	2.5E+05	4.2E-04	1.4E+02	0.0E+00	2.5E+00
morpholinium (mo)	108	14	10	13	293.18	358	100	101.3	3.2E-04	7.1E-01	6.0E-06	4.0E-02
phosphonium (p)	331	28	9	24	219.15	373	100	2.5E+05	1.2E-05	2.2E+00	1.2E-06	2.7E-01
piperidinium (pip)	94	10	7	9	293.18	361.15	100	1.0E+02	8.8E-04	1.2E+00	6.0E-06	1.4E-01
pyridinium (py)	700	67	32	48	263	438.15	100	1.0E+02	2.0E-03	5.2E+00	8.0E-05	2.4E-01
pyrrolidinium (pyr)	657	65	20	38	238.15	448.15	100	1.0E+02	2.1E-04	1.2E+01	3.0E-06	8.4E-01
sulfonium (s)	84	10	6	10	258.15	468.15	100	101.3	6.7E-02	9.2E+00	1.5E-03	1.4E-01
triazolium (trz)	139	12	7	11	298	353.2	100	101.3	1.4E-02	1.6E+00	6.0E-04	1.6E-01
Anions												
BF ₄ derivatives (BF ₄)	1079	110	25	50	238.05	468.15	100	2.0E+05	2.1E-04	1.9E+01	0.0E+00	2.5E+00
NTf ₂ derivatives (NTf ₂)	2385	230	126	133	234.15	495.35	100	2.5E+05	8.3E-04	1.0E+01	0.0E+00	4.0E-01
PF ₆ derivatives (PF ₆)	352	26	8	10	263	468.15	100	2.0E+05	3.3E-03	8.1E+00	1.0E-04	2.0E-01
aminoacids (aa)	176	22	7	18	278	378	100	101	1.3E-04	2.6E+00	1.0E-05	7.2E-02
azides (N ₃)	13	1	1	1	303.15	368.15	100	101	4.9E-02	1.5E+00	1.9E-03	6.1E-02
carboxylates (RCO ₂)	475	53	18	37	248.15	418.15	89.3	1.0E+02	2.9E-04	6.9E+00	0.0E+00	3.0E-01
dicyanamides (dca)	404	49	35	35	238.15	468.15	100	1.0E+02	4.0E-03	1.5E+01	5.0E-04	4.4E-01
inorganics (x)	301	25	16	23	219.15	484.1	100	1.0E+02	1.2E-05	1.4E+02	0.0E+00	8.4E-01
metal complexes (ML _n)	67	8	8	8	283.15	343.15	100	1.0E+02	3.4E-02	1.1E+00	4.0E-04	6.5E-02
methanides (CR ₃)	16	2	2	2	293.25	379.25	100	101.3	7.6E-01	4.6E+00	2.2E-02	1.0E-01
phosphates (RPO ₄)	191	16	16	16	288.15	393.2	100	101.3	1.2E-03	1.5E+00	1.0E-04	1.5E-01
sulfates (RSO ₄)	199	24	10	14	263.1	433.1	100	1.0E+02	5.0E-03	6.9E+00	7.5E-04	1.4E+00
sulfonates (RSO ₃)	356	42	20	26	258.1	468.15	100	1.0E+02	4.2E-04	8.8E+00	0.0E+00	2.2E-01

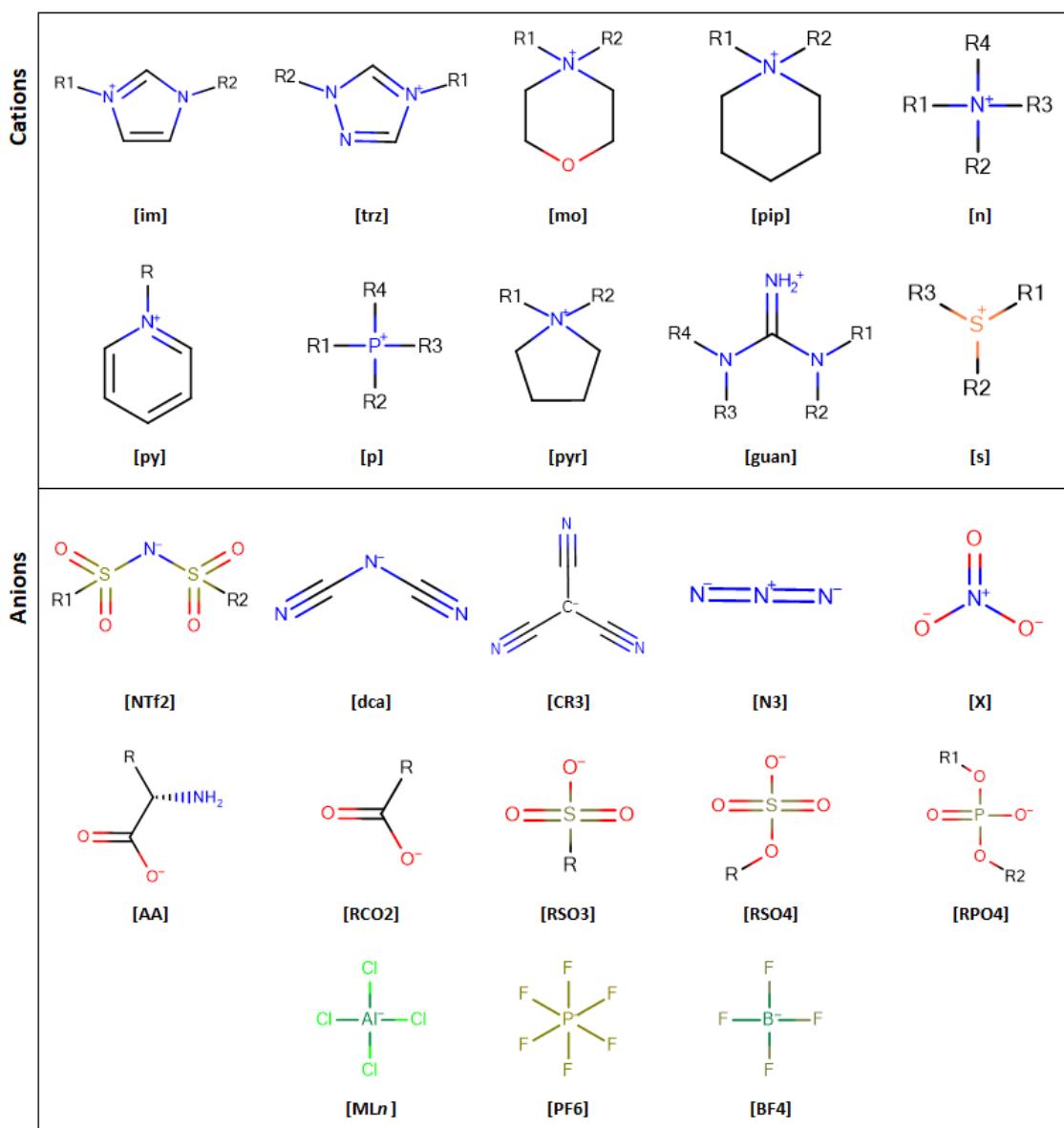


Figure 38: Familles chimiques cationiques et anioniques étudiées. Les abréviations des familles chimiques sont détaillées dans le Tableau 5

[im]	46	41	27	25	20	14	12	10	7	5	4	3	1	
[n]	42	21	8	4	26	12	3	1					1	
[py]	28	7				3	3	1	7					
[pyr]	22	3	1	1	1	1			6	1	4			
[p]	17	2	4	3	7	2	1			3				
[trz]	7	1		1			2							
[s]	6							4						
[pip]	3	2								4				
[cs]	3							2						
[mo]	1	1						7		4				
[guan]			1		1					4			1	
	[NTf ₂]	[BF ₄]	[aa]	[x]	[RCO ₂]	[RSO ₃]	[RSO ₄]	[dca]	[PF ₆]	[RPO ₄]	[ML _n]	[CR ₃]	[N3]	[hca]

Figure 39: Nombre d'LI selon différentes combinaisons cation-anion avant traitement des données. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.

[im]	40	30	16	12	11	10	8	7	6	4	3	2	1	
[py]	28	7		3	1	7	2							
[n]	20	6	3	14	1	1	11	6				1		
[pyr]	20	3	1	1			6	1	1	1		4		
[trz]	7	1	1					2						
[s]	6					4								
[p]	5		2	7	1			2	4	3				
[pip]	3	2								4				
[mo]	1	1				7				4				
[guan]									4					
	[NTf ₂]	[BF ₄]	[x]	[RCO ₂]	[RSO ₄]	[dca]	[RSO ₃]	[aa]	[PF ₆]	[ML _n]	[RPO ₄]	[CR ₃]	[N3]	

Figure 40: Nombre d'IL selon différentes combinaisons cation-anion après traitement des données. Les abréviations des familles chimiques sont détaillées dans le Tableau 5

III.2.3 Analyse de données

Dans cette section, nous analysons l'influence de la température, de la structure ionique et de la pression sur la conductivité électrique des liquides ioniques. Ces analyses sont essentielles afin de dégager les tendances générales, pour comprendre les mécanismes sous-jacents qui régissent la conductivité électrique de ces liquides et pour guider le développement de modèles prédictifs précis.

III.2.3.A Effet de la température sur la conductivité électrique

En général, la conductivité électrique des liquides ioniques augmente de manière marquée avec la température, en raison de l'accroissement de l'agitation thermique et de la diminution concomitante de la viscosité, qui facilite la mobilité des ions. Pour illustrer cette tendance, la Figure 41 compare la variation de σ en fonction de T pour la série de cations imidazolium [im-n,1] associée à l'anion NTf_2 ($n = 1, 2, 4, 6$ et 8) sur une plage s'étendant de 263 K à 373 K .

Sur l'ensemble de cette plage, on observe un accroissement quasi-exponentiel de la conductivité, typique d'une dépendance de type Vogel–Fulcher–Tammann (VFT). Par ailleurs, pour une température donnée (p. ex. 298 K à 101 kPa), la conductivité suit strictement la séquence : [im-1,1][ntf2] > [im-2,1][ntf2] > [im-4,1][ntf2] > [im-6,1][ntf2] > [im-8,1][ntf2], c'est-à-dire que l'allongement progressif de la chaîne alkyle se traduit par une baisse systématique de σ .

Cette diminution s'explique d'une part par l'effet stérique accru, qui augmente la viscosité et donc la résistance au flux ionique, et d'autre part par l'effet hydrophobe, par lequel les longues chaînes alkyles favorisent l'auto-assemblage micellaire et réduisent le nombre d'espèces réellement mobiles dans le liquide ionique.

Ainsi, pour la série [im-n,1][NTf_2], σ décroît typiquement d'environ $0,9\text{ S/m}$ pour $n = 1$ à l'ordre de $0,1311\text{ S/m}$ pour $n = 8$ à 298 K . Cette double influence de la température et de la structure ionique constitue un signal robuste et quantifiable, qui sert de socle pour le développement de méthodes prédictives.

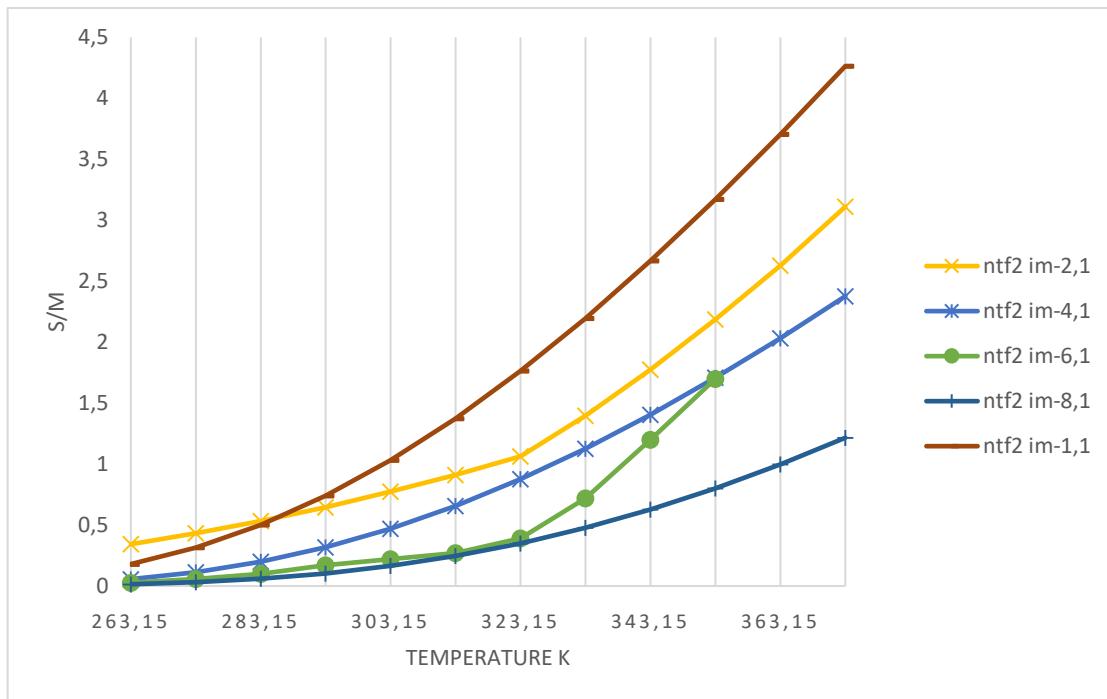


Figure 41: Influence de la température sur la conductivité électrique des liquides ioniques

III.2.3.B Effet de la structure ionique sur les conductivités électriques

La conductivité électrique des liquides ioniques dépend fortement de la structure ionique, notamment de la taille et de la nature des cations qui les composent. Cet effet est ici analysé en examinant l'influence de la longueur de la chaîne alkyle n sur différentes familles de cations – imidazolium [im- $n,1$], pyrrolidinium [pyr- $n,1$], pyridinium [py- n], ammonium [$n-n,2,2,2$] et triazolium [trz124- $n,1$] – tout en conservant l'anion NTf₂ fixe. La Figure 42 présente la variation de la conductivité électrique σ en fonction de n (nombre de carbones de 1 à 8), mesurée à 298 K et 101 kPa, offrant une vue comparative des tendances au sein de ces séries cationiques.

On observe, pour chaque type de cation, une diminution quasi-exponentielle de σ . Les cations imidazolium restent globalement les plus conducteurs, suivis de près par les pyrrolidiniums, tandis que les triazoliums les moins mobiles.

Ces données mettent en évidence la sensibilité de σ aux structures moléculaires des LI et de la longueur de la chaîne alkyle des cations, des paramètres clés dans la modélisation des propriétés des LI.

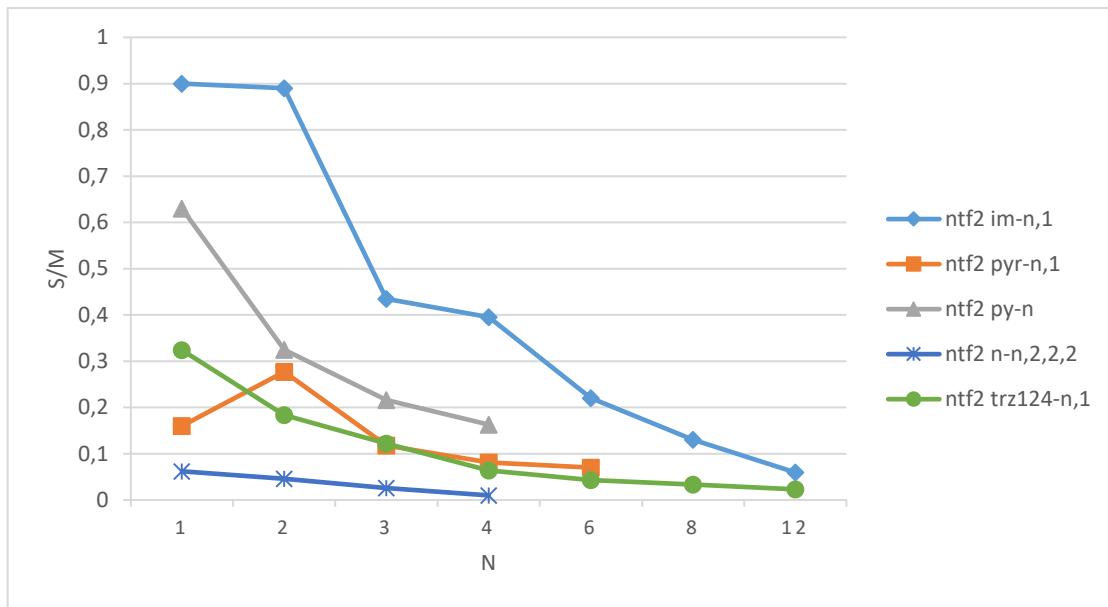


Figure 42: Influence de la longueur de chaîne du cation sur la conductivité électrique des liquides ioniques (298K, 101 kPa)

III.2.3.C Effet de la pression sur la conductivité électrique d'Ils

La conductivité électrique des liquides ioniques décroît généralement lorsque la pression augmente, en raison de la réduction du volume libre et de l'accroissement des interactions ion-ion qui entravent la mobilité. Pour illustrer cette tendance, la Figure 43 présente l'évolution de σ en fonction de la pression pour plusieurs LI représentatifs : [im-4,1][NTf₂], [im-4,1][BF₄], [im-6,1][PF₆], [im-4,1][PF₆], [im-8,1][BF₄], [im-8,1][PF₆] et [P-5,2,2,2][NTf₂], sur une plage allant de 0,1 MPa à 250 MPa.

Sur l'ensemble de cette plage, on observe une diminution progressive de la conductivité, souvent caractérisée par une dépendance de type logarithmique ou exponentiel inverse. Cette baisse s'explique principalement par l'augmentation de la viscosité sous pression, qui freine le déplacement des ions, et par une réduction de l'espace libre disponible, limitant leur diffusion. Ainsi, pour [im-4,1][ntf2], σ passe typiquement de 0,403 S/m à 0,1 MPa à environ 0,15 S/m à 98,1 MPa. L'influence combinée de la pression et de la composition ionique offre un cadre solide pour prédire le comportement des liquides ioniques sous contraintes mécaniques.

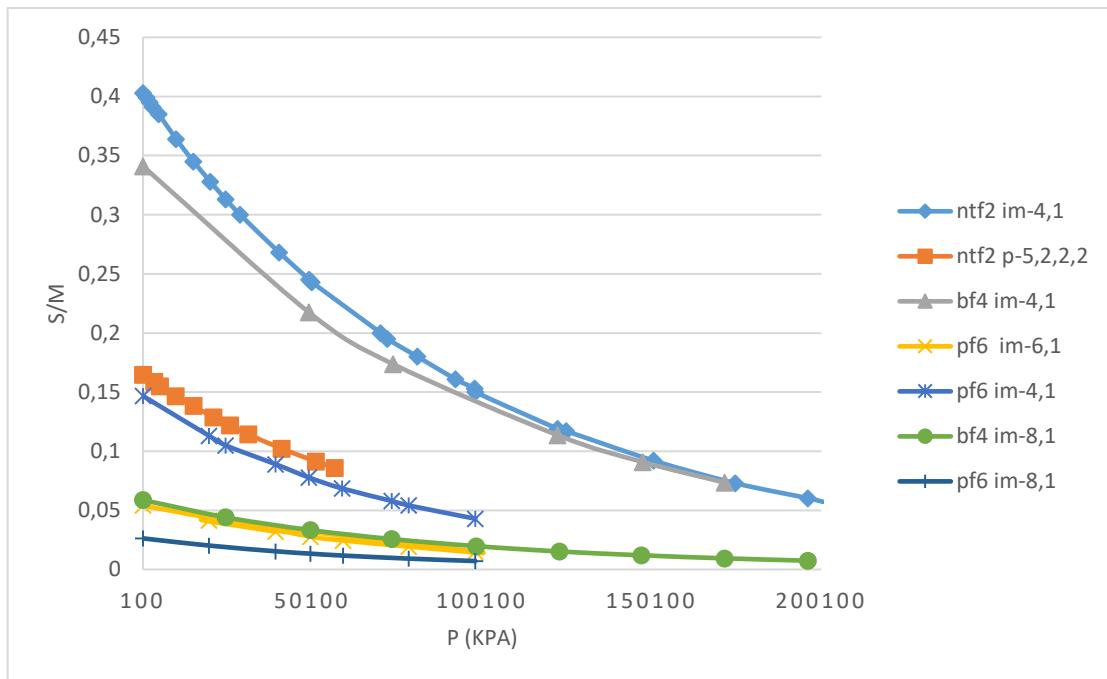


Figure 43: Influence de la pression sur la conductivité électrique des liquides ioniques (298k)

III.2.3.D Conclusion :

Les analyses menées confirment que la conductivité des ILs est fortement régie par la température (effet activateur), la structure ionique (effet stérique et hydrophobe) et la pression (effet inhibiteur). Ces observations guideront la construction de modèles prédictifs intégrant simultanément ces trois variables pour optimiser leurs performances prédictives en conditions réelles.

III.3 MÉTHODOLOGIE

Dans ce travail, cinq approches de modélisation ont été déployées pour prédire la conductivité électrique des liquides ioniques : Extreme Gradient Boosting, Random Forest, Support Vector Machine, réseau de neurones profond et un modèle conventionnel (voir cf. II.3 Modèles et régression).

Dans le cas de modèle conventionnel, plusieurs formes non-linéaires de modèles conventionnels ont été développés et évalués[7], et le meilleur modèle a été sélectionné pour déterminer l'expression optimale représentant la dépendance température-pression de la conductivité électrique via des contributions de groupes structuraux ou des

descripteurs. Après une évaluation minutieuse, l'équation suivante a finalement été sélectionnée:

$$\ln\left(\frac{\varepsilon}{R_0}\right) = A + B \frac{100}{T} + C \left(\frac{100}{T}\right)^2 + D \frac{100}{P}$$

où R_0 est le paramètre ajustable et A, B, C et D sont la contribution de tous les groupes, tant de premier ordre que de second ordre, présents dans les cations, anions et substituants. T est la température en kelvins (K) et P la pression en kilopascals (kPa).

Ces cinq approches de modèles, exploitent une double approche computationnelle pour générer leurs entrées structurales à savoir : Contributions de groupes et les descripteurs QSPR. À ces informations s'ajoutent les variables thermodynamiques ; la température T (en K) et la pression P (en kPa). Ces types informations jouent un rôle crucial en transformant les caractéristiques structurales et physico-chimiques des LI en données numériques exploitables.

Tous les modèles, QSPR ou contributions de groupes, ont été développés avec la Statistics & Machine Learning Toolbox de MATLAB (v. 2021b, licence académique), en conservant par défaut la majorité des hyper-paramètres (cf. section II.5.3) fournis par la bibliothèque. Seuls les paramètres ayant un impact notable dans les contextes de QSPR et CG ont été passés en revue. Pour optimiser ces réglages, nous avons recouru à l'optimisation bayésienne [36], qui propose de manière itérative les configurations les plus prometteuses. La qualité des modèles a été évaluée par validation croisée à 5 plis répétée 25 fois, en minimisant l'erreur quadratique moyenne.

L'optimisation et le réglage des hyper-paramètres pour les quatre modèles étudiés (Extreme Gradient Boosting, Forêt Aléatoire, Machine à Vecteurs de Support, Réseau de Neurones Profond) a montré une convergence remarquable des performances entre les approches QSPR et contribution de groupes. Pour faciliter une comparaison méthodologique rigoureuse, les résultats ont été standardisés. Le Tableau 6 synthétise les hyperparamètres optimisés retenus et leurs domaines de recherche.

Tableau 6: Espace de recherche des hyperparamètres et détermination de leur sélection optimale à l'aide d'un algorithme d'optimisation bayésienne.

Modèles	Espace de recherche	Paramètres sélectionnés
XGBoost	$MinLeafSize = [1 \text{ } 100]$, $NumLearningCycles = [10 \text{ } 1000]$, $LearnRate = [1e-3, 1]$.	$MinLeafSize = 1$, $NumLearningCycles = 476$ $LearnRate = 0.38$.
RF	$MinLeafSize = [1 \text{ } 100]$, $NumLearningCycles = [10 \text{ } 1000]$.	$MinLeafSize = 1$, $NumLearningCycles = 800$
SVM	$C = [1 \text{ } 500]$, $\varepsilon = [0.0001 \text{ } 1]$, $\gamma = [0.1 \text{ } 100]$.	$C = 294$, $\varepsilon = 0.023$, $\gamma = 3.072$
DNN	$N_Hidden = [1, 5]$, $Layer_N_Size = [2, 50]$, $Activations = [\text{ReLU, sigmoid, tanh}]$, $Lambda = [0, 100]$.	$N_Hidden = 2$, $Layer_N_Size = [28 \text{ } 10]$, $Activations = \text{tanh}$, $Lambda = 2.29e-07$.

III.4 RÉSULTATS ET DISCUSSION

Cette section présente et discute les résultats des modèles proposés. Les modèles proposés sont testés et évalués à l'aide des indicateurs statistiques cités dans la section II.6. De plus, les résultats des prédictions sont classés par familles d'IL et par combinaisons de cations-anions, ce qui permet non seulement de mieux comprendre la précision des modèles, mais aussi d'aider les utilisateurs à appliquer les modèles proposés de façon appropriée.

Les résultats de prédiction sont organisés en deux catégories :

- Méthodes basées sur les contributions de groupes, qui exploitent directement ces contributions comme variables d'entrée,
- Approches QSPR, qui s'appuient sur des descripteurs moléculaires pour alimenter les modèles.

III.4.1 Méthodes de contributions de groupes

Dans cette étude, **127 groupes** ont été identifiés pour décrire toutes les structures moléculaires des LI de la base de données de la conductivité électrique. Cette liste

comprend : **52** groupes **cationiques**, **44** groupes **anioniques**, **31** groupes **substituants**, répartis en 79 groupes de premier ordre et 48 groupes de second ordre. Pour obtenir des informations détaillées sur le schéma de fragmentation et les groupes considérés dans cette étude, voir (section II.3.3) et le travail [1].

Le Tableau 7 montre une comparaison des performances prédictives pour les cinq modèles proposés basées sur l'approche de CG. Les mesures statistiques ont été moyennées sur une validation croisée 5 plis répétée aléatoirement 25 fois. Les résultats montrent que le modèle DNN est le modèle le plus précis pour cette tâche. Le modèle SVM est une seconde proche, tandis que le modèle XGBoost et le modèle RF sont moins précis. Le modèle conventionnel est le modèle le moins précis. Ces résultats soulignent la supériorité des modèles d'apprentissage automatique par rapport aux approches conventionnelles, soulignant leur capacité à fournir une précision accrue.

Tableau 7: Statistiques de performance des GC-modèles sur les jeux d'apprentissage et de test, exprimées en moyennes issues d'une validation croisée à 5-plis, répétée 25 fois selon des partitions aléatoires.

Models	EAM_{train} (S/m)	EAM_{test} (S/m)	$ERAM_{train}$ (%)	$ERAM_{test}$ (%)	$REQM_{train}$	$REQM_{test}$	R^2_{train}	R^2_{test}
XGBoost	0.062	0.117	5.48	11.27	0.477	1.122	0.9809	0.9194
RF	0.196	0.246	18.51	33.26	1.955	2.092	0.7906	0.7034
SVM	0.034	0.046	3.45	5.47	0.107	0.156	0.9991	0.9974
DNN	0.032	0.042	3.12	4.60	0.098	0.140	0.9992	0.9977
Conventional	0.308	0.329	38.89	123.66	3.037	2.829	0.226	0.313

Il convient de noter que les résultats d'erreur obtenus par les modèles SVM et DNN se situent dans la plage des **incertitudes expérimentales**. Plus précisément, l'incertitude moyenne des données expérimentales après traitement (révision des données) est de 0,037 S/m, tandis que les modèles SVM et DNN montrent des EAM de 0,046 et 0,042 S/m, respectivement. Cet accord entre les résultats du modèle et les incertitudes expérimentales renforce la fiabilité de ces modèles dans la prédiction de la conductivité électrique.

Les modèles les plus performants de chaque catgorie, sélectionnés à partir des 25 itérations d'une validation croisée en 5 plis, sont conservés pour une analyse plus approfondie des résultats. Des informations détaillées sur ces modèles retenus, incluant tous les paramètres nécessaires à la prédiction de la conductivité électrique, sont disponibles dans les Annexe X.

La Figure 44a fournit une représentation visuelle des valeurs au carré R (R^2) obtenues par chaque modèle sélectionné dans l'ensemble d'entraînement et l'ensemble de test. On peut observer que les points obtenus à partir des modèles DNN et SVM présentent un degré d'accord plus élevé avec les valeurs cibles de conductivité électrique, dépassant les performances des autres modèles. Un examen plus approfondi des résultats présentés à la Figure 44a confirme les résultats présentés au Tableau 7. De plus, la Figure 44b présente une comparaison des distributions gaussiennes pour les cinq modèles, en mettant l'accent à la fois sur l'ensemble de formation et l'ensemble de tests. Cette analyse révèle des variations significatives en termes de dispersion ainsi que les valeurs de probabilité maximales à travers les modèles. Toutes les distributions présentent des formes symétriques, indiquant une distribution équilibrée des erreurs autour de la moyenne.

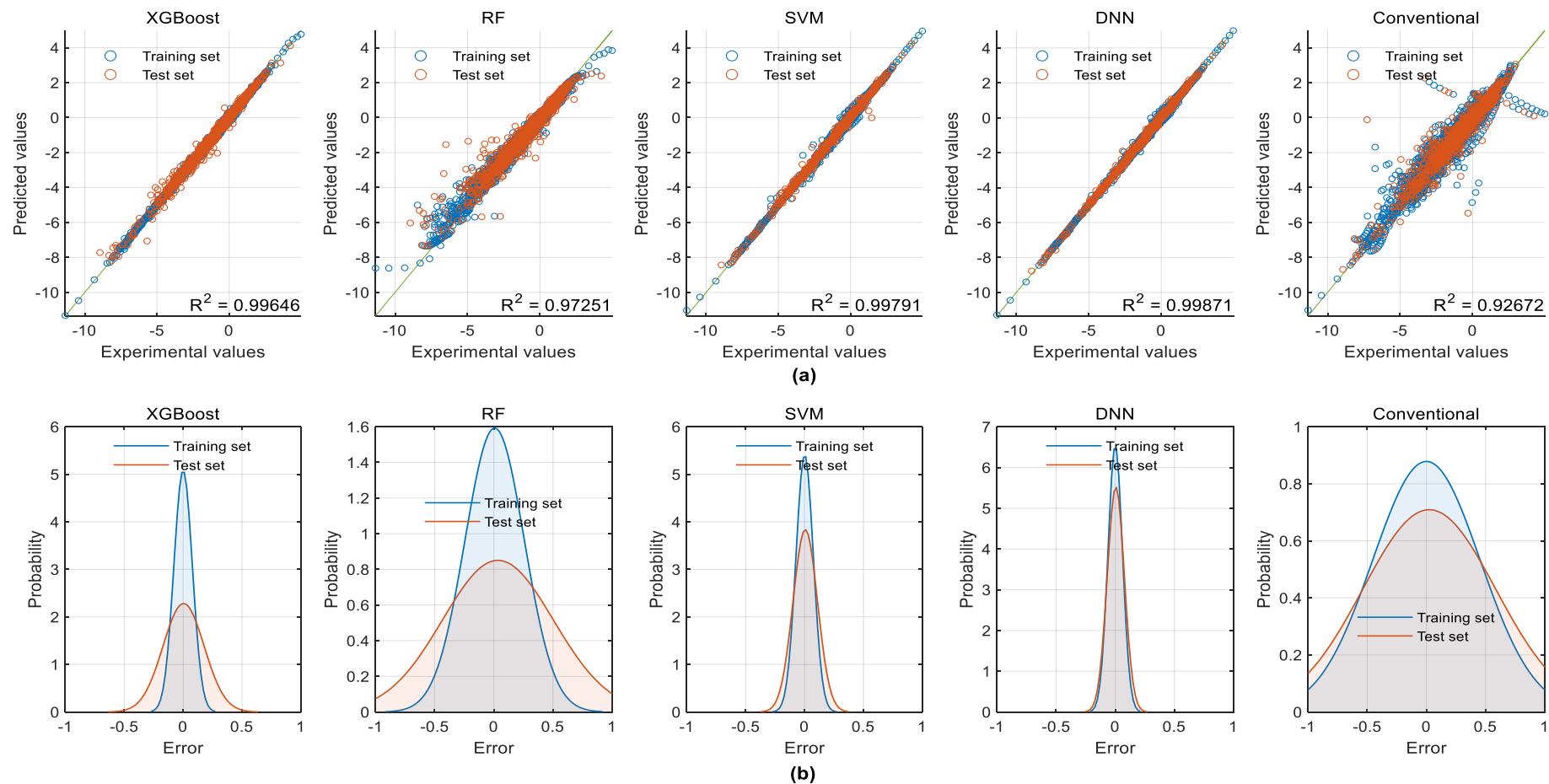


Figure 44: Comparaison des modèles de contribution de groupes : (a) Nuages de points des valeurs prédictes et expérimentales dans l'ensemble d'apprentissage et l'ensemble de test. (b) Distribution gaussienne des erreurs de prédiction dans l'ensemble d'apprentissage et l'ensemble de test. L'unité de conductivité électrique est (S/m).

Les modèles DNN et SVM présentent des distributions plus étroites et plus concentrées, accompagnées de probabilités maximales plus élevées, indiquant un niveau plus élevé de précision dans leurs prédictions. En revanche, les modèles RF et Conventionnel montrent des distributions plus larges avec des probabilités maximales plus petites, suggérant une plus grande variabilité dans leurs profils d'erreur. Il convient de noter que les distributions dans les ensembles d'entraînement semblent être plus étroites, plus concentrées et accompagnées de probabilités maximales plus élevées par rapport aux ensembles d'essais. Cependant, cette tendance est moins prononcée pour les modèles DNN et SVM, indiquant leur puissance prédictive plus élevée. Toutes ces constatations concordent avec les classements de rendement présentés dans le Tableau 7 et soulignent la supériorité des modèles DNN et SVM en termes de précision et de répartition des erreurs.

Pour améliorer la lisibilité et l'exhaustivité du document, nous allons maintenant procéder dans une analyse détaillée axée uniquement sur les deux modèles les plus performants : DNN et SVM. Ces modèles ont été minutieusement évalués à travers diverses combinaisons de familles cationiques et anioniques, permettant une évaluation complète de leurs capacités prédictives.

III.4.1.A Ecarts relatifs selon les familles cationiques et anioniques

La Figure 45 illustre la distribution des écarts relatifs (ER) entre différentes familles cationiques et anioniques pour les modèles GC-DNN (Figure 45a) et GC-SVM (Figure 45b). La marque centrale de chaque boîte correspond à la médiane des (ER), tandis que les bords supérieurs et inférieurs de la boîte représentent les 75e et 25e percentiles. Pour renforcer la clarté des boîtes à moustaches, seules les valeurs d' (ER) comprises dans un intervalle de $\pm 5\%$ sont incluses. Dans cette figure (Figure 45), les familles sont énumérées par ordre décroissant selon le nombre de LI, allant de l'imidazolium [im] au guanidinium [guan] pour les cations, et des dérivés de bistriflamide [NTf₂] aux azides [N₃] pour les anions. À partir de cette figure, il devient évident que la majorité des médianes des différentes familles chimiques sont centrées, ce qui indique une distribution équilibrée des valeurs ER pour les deux modèles, en particulier pour le modèle DNN. Cependant, il y a quelques familles situées au sommet des deux figures qui dévient du centre. Cet écart peut être attribué à la présence d'un petit nombre de points de données décrivant ces familles spécifiques. En outre, nous observons que les bords supérieurs et inférieurs des boîtes sont confinés dans une plage ne dépassant pas 2,5%, ce qui indique

une distribution assez compacte des valeurs ER. Cela met en évidence la cohérence et le regroupement étroit de la majorité des points de données au sein de chaque famille chimique.

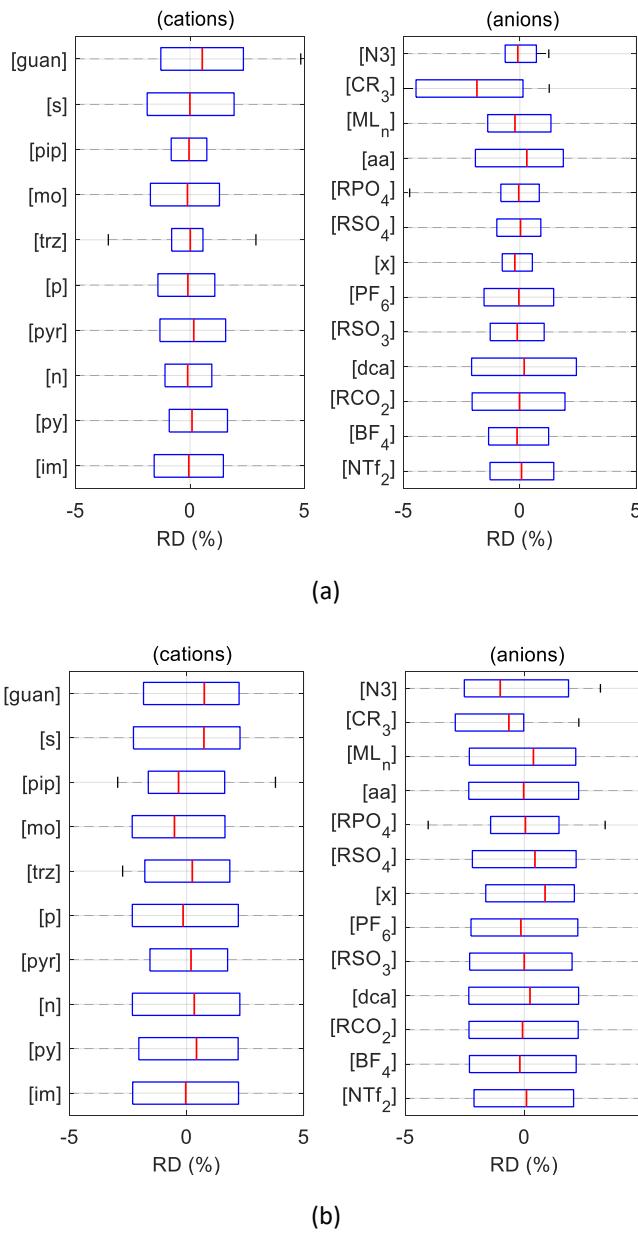


Figure 45: Distribution des écarts relatifs (ER) en fonction des familles cationiques et anioniques. La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieur et supérieur de la boîte représentent respectivement le premier et le troisième quartile. (a) Résultats du modèle GC-DNN. (b) Résultats du modèle GC-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.

III.4.1.B Écarts relatifs selon les combinaisons cation-anion

La Figure 46 présente la distribution des écarts relatifs ER pour différentes combinaisons de familles chimiques de cations et d'anions, obtenue avec les modèles DNN et SVM. L'analyse des résultats permet de formuler plusieurs observations. Premièrement, les deux modèles présentent des variations d'ER selon les combinaisons cation-anion. Les ER varient de 0,3 à 11,1 pour le DNN et de 1,2 à 11,7 pour le SVM. Pour le DNN, plus de 75 % des ER sont inférieurs ou égaux à l'ER moyen des combinaisons cation-anion. De même, plus de 72,5 % des ER du SVM sont inférieurs ou égaux à sa moyenne. Deuxièmement, certaines combinaisons cation-anion affichent systématiquement des ER plus élevés dans les deux modèles. Par exemple : Les cations phosphonium associés aux acides aminés montrent des ER notables de 10,8 (DNN) et 11,7 (SVM). Les cations morpholinium combinés aux anions dicyanamide présentent des ER de 11,1 (DNN) et 9,7 (SVM).

Il est important de mentionner que ces valeurs élevées proviennent exclusivement d'ensembles de données limités comprenant un maximum de 7 liquides ioniques distincts. Néanmoins, les deux modèles fournissent des valeurs moyennes raisonnables de ER pour chaque famille cationique ou anionique, allant de 1,2 à 5,5 pour le modèle DNN et de 1,4 à 6,3 pour le modèle SVM. En fin de compte, ces résultats soulignent l'importance de considérer le contexte chimique spécifique lors de l'utilisation de modèles DNN et SVM pour prédire la conductivité électrique.

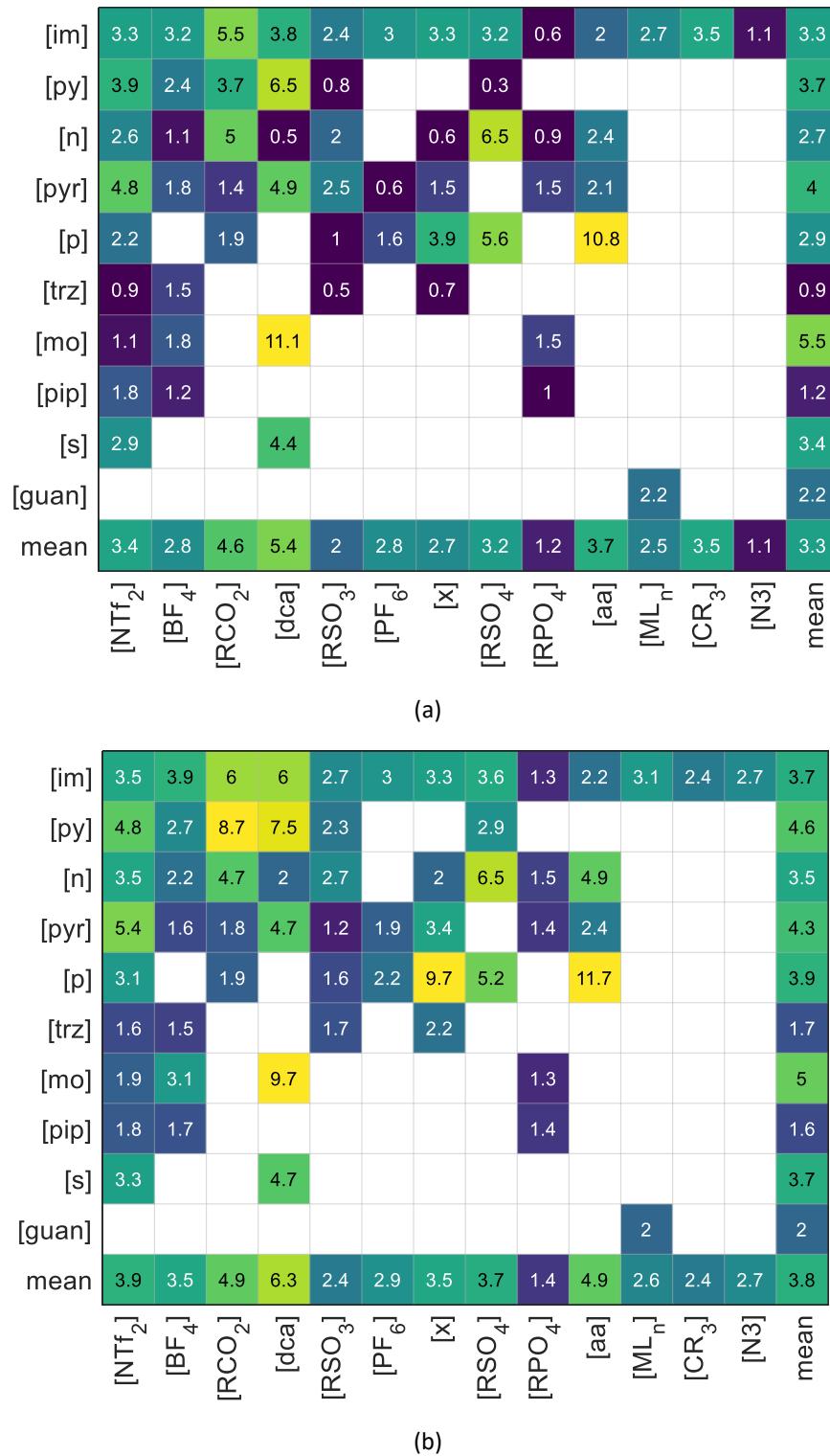


Figure 46: Distributions des écarts relatifs ER(%) en fonction des différentes familles de cations et d'anions. (a) modèle DNN, (b) modèle SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.

III.4.1.C Distributions d'écart relatifs moyens de prédition

Pour obtenir un aperçu plus complet de la distribution des erreurs de prédition découlant des meilleurs modèles DNN et SVM, nous avons affiché la distribution des points de données par rapport à la plage des écarts relatifs moyens (ERM). Comme le montre la Figure 47a pour le modèle DNN, le ERM de 2 481 points de données (soit 41,3 % sur le nombre total de 6 014 points de données) est inférieur à 1 %. De plus, il y a 73,2 % (voir les trois premiers intervalles) et 92,4 % (voir les sept premiers intervalles) de points de données pour lesquels le ERM est inférieur à 3 % et 10 %, respectivement. Il n'y a que dix points de données pour lesquels le modèle montre un ERM entre 50 et 86 % et seulement deux points de données pour lesquels le modèle montre un ERM entre 86 et 117 %. Des tendances similaires sont évidentes dans le modèle SVM représenté à la Figure 47b, dans lequel seuls 25 points de données présentent un ERM allant de 50 à 86 %, et seulement 8 points de données se situent dans la plage ERM de 86 à 213 %. Le point commun entre ces grands écarts est qu'ils appartiennent à des combinaisons cationiques et anioniques qui ont un nombre de points de données relativement faible.

En conclusion, sur la base des informations fournies par les Figure 47a et Figure 47b, les deux modèles, DNN et SVM, présentent des prouesses prédictives significatives. Remarquablement, **81,3 %** et **83,9 %** des erreurs relatives de prédictions se situent dans la plage d'incertitude expérimentale (qui est de 4,38 %) pour le DNN et le SVM, respectivement.

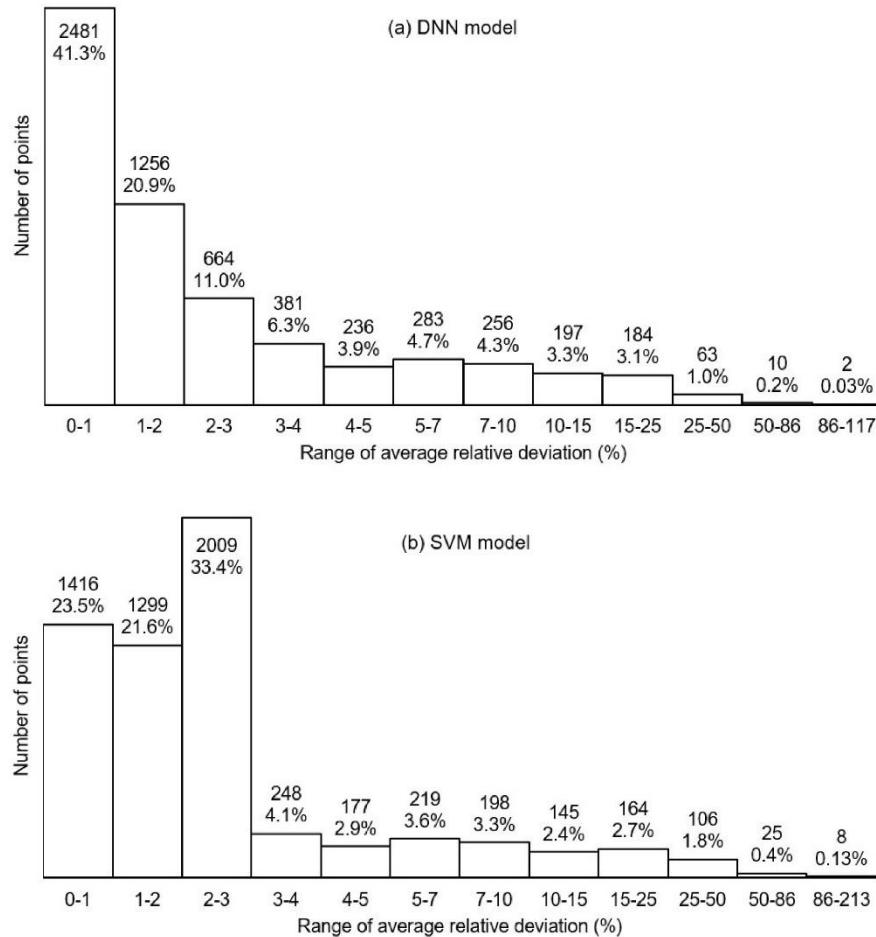


Figure 47: Distributions des points de données en fonction des erreurs relatifs moyens ER (%). (a) modèle DNN, (b) modèle SVM.

III.4.2 Méthodes QSPR

Dans l'approche QSPR, chaque liquide ionique est représenté par un vecteur de 163 descripteurs moléculaires calculés à partir de RDKit, incluant des indices topologiques, géométriques et électroniques. Ces données quantitatives capturent la taille, la polarité, la flexibilité et la distribution de charges de la molécule, fournissant un point d'entrée riche et multidimensionnel aux algorithmes.

Nous avons délibérément choisi de ne pas mettre en œuvre une procédure d'optimisation exhaustive des descripteurs clés (telle qu'une sélection wrapper, embedded ou génétique) afin de garantir une comparaison équitable entre les approches QSPR et contributions de groupes. En effet, limiter a priori le vecteur d'entrée risquerait d'éliminer des variables faiblement corrélées mais porteuses d'interactions non linéaires essentielles, notamment pour les SVM et DNN dont la performance dépend de la richesse des représentations.

L’analyse des statistiques de performance (Tableau 8) révèle que les modèles SVM et DNN conservent leur supériorité, avec des EAM_{test} respectifs de 0,044 S/m et 0,049 S/m et des ERAM_{test} de 4,51 % et 4,73 %. Le SVM atteint même un coefficient de détermination en test R²_{test} de 0,9992, soulignant son exceptionnel pouvoir de généralisation. Ces résultats contrastent avec ceux de la méthode GC, où le DNN dominait, suggérant que les descripteurs QSPR, riches en informations électroniques, sont optimisés plus efficacement par les algorithmes SVM. DNN demeurent un cran en dessous (EAM_{test} ≈ 0,044 S/m et 0,049 S/m).

Tableau 8: Statistiques de performance des QSPR-modèles sur les jeux d’apprentissage et de test, exprimées en moyennes issues d’une validation croisée à 5-plis, répétée 25 fois selon des partitions aléatoires.

Models	EAM _{train} (S/m)	EAM _{test} (S/m)	ERAM _{train} (%)	ERAM _{test} (%)	REQM _{train}	REQM _{test}	R ² _{train}	R ² _{test}
XGBoost	0.038	0.127	3.98	10.29	0.128	1.683	0.9981	0.9371
RF	0.142	0.306	13.72	21.24	1.503	4.206	0.7830	0.4255
SVM	0.033	0.044	3.29	4.51	0.105	0.208	0.9988	0.9992
DNN	0.040	0.049	3.80	4.73	0.123	0.321	0.9984	0.9987

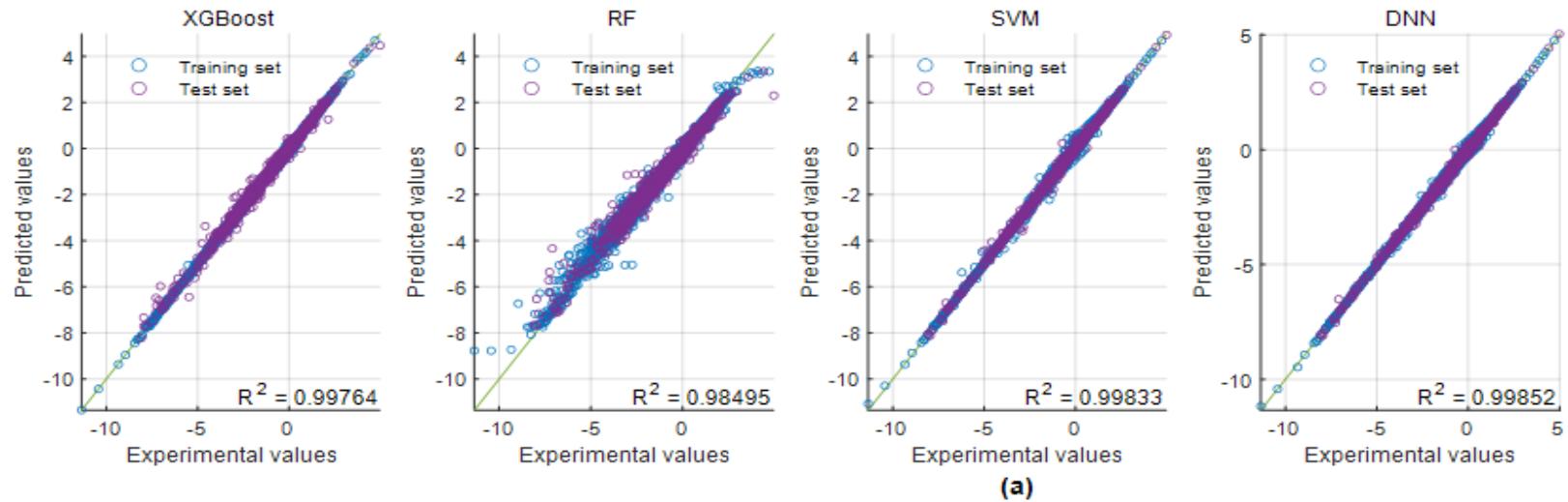
La confrontation des deux stratégies met en évidence un léger avantage de l’approche QSPR sur la méthode par contributions de groupes pour la majorité des modèles, sans toutefois bouleverser l’ordre de performance. Pour XGBoost et Random Forest, la QSPR fait passer l’ERAM_{test} de 11,27 % à 10,29 % et de 33,26 % à 21,24 %, respectivement. Ce gain modeste reflète la capture plus fine des effets de la topologie moléculaire grâce aux descripteurs continus. En revanche, SVM et DNN affichent des écarts mineurs entre les deux approches (EAM_{test} variant de 0,046 S/m à 0,044 S/m pour SVM, et de 0,042 S/m à 0,049 S/m pour DNN), tous deux toujours dans la marge d’incertitude expérimentale, ce qui souligne la robustesse de ces modèles face à la nature des entrées.

Cette analyse comparative montre que, bien que la QSPR procure un léger gain de performance aux modèles arborescents (XGBoost, RF) et SVM, c’est toutefois le DNN basé sur les contributions de groupes qui délivre la meilleure précision globale ; par ailleurs, cette dernière approche conserve un avantage notable en termes de simplicité computationnelle et d’interprétabilité.

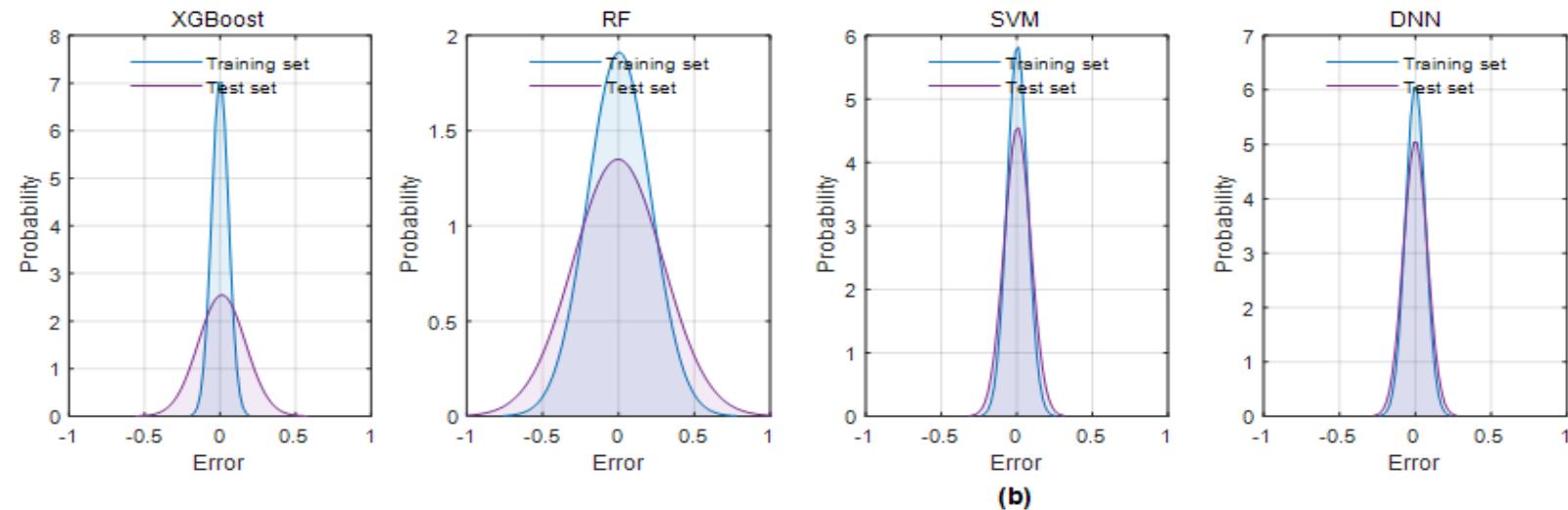
La Figure 48.a illustre la corrélation entre valeurs prédictes et valeurs expérimentales pour chacun des quatre modèles étudiés. Les graphes relatifs au SVM et au DNN montrent un

alignement quasi parfait des points sur la droite d'identité ($y = x$), avec des coefficients de détermination R^2 supérieurs à 0,998 tant en phase d'entraînement qu'en phase de test. XGBoost affiche lui aussi une très forte concordance ($R^2 \approx 0,9976$), bien qu'un léger écart apparaisse aux extrémités de la plage de conductivité. En revanche, le modèle RF présente une dispersion plus marquée, notamment sur l'ensemble de test ($R^2 \approx 0,985$), traduisant un léger surapprentissage et une sensibilité moindre aux valeurs extrêmes. La seconde partie de la Figure 48b compare les densités de probabilité des erreurs de prédiction pour les jeux d'entraînement (bleu) et de test (orange). Les distributions associées au SVM et au DNN sont étroitement centrées sur zéro et présentent une faible variance, signe d'une excellente robustesse et d'une stabilité entre les deux phases. XGBoost conserve une erreur moyenne de même ordre tout en montrant une variance légèrement supérieure. Le RF se caractérise par une distribution plus large et des queues épaisses, reflétant une hétérogénéité accrue des écarts.

Pour garantir une lecture fluide et une couverture approfondie, nous concentrerons dorénavant notre analyse QSPR sur les deux modèles les plus performants, à savoir le SVM et le DNN.



(a)



(b)

Figure 48: Comparaison des modèles QSPR sélectionnés pour prédire σ : (a) Nuages de points des valeurs prédictives et expérimentales dans l'ensemble d'apprentissage et l'ensemble de test. (b) Distribution gaussienne des erreurs de prédiction dans l'ensemble d'apprentissage et l'ensemble de test. L'unité de conductivité électrique est (S/m)

III.4.2.A Ecarts relatifs selon les familles cationiques et anioniques

La Figure 49 met en lumière la variabilité des écarts relatifs (ER) obtenus par les modèles QSPR-DNN (Figure 49a) et QSPR-SVM (Figure 49b) au sein de chaque famille chimique. Les boîtes à moustaches restreignent ici les ER à $\pm 5\%$ pour améliorer la lisibilité, la ligne médiane indique la tendance centrale et les quartiles délimitent l'étendue interquartile (IQR) pour chaque catégorie de cation ou d'anion (voir Tableau 5 pour la légende).

Globalement, les médianes des ER sont très proches de zéro, ce qui confirme l'absence de biais systématique selon les types de cation et d'anion. On observe néanmoins que le modèle QSPR-SVM produit des boîtes plus étroites (IQR typiquement inférieur à 2 %), témoignant d'une meilleure homogénéité des écarts au sein des familles les plus représentées (p.ex. Imidazolium [im], [NTf₂]). En revanche, le QSPR-DNN présente légèrement des IQR plus larges pour certaines familles rares (par exemple Guanidinium [guan], les complexes métalliques [MLn]), signe d'une sensibilité accrue aux jeux de données peu denses.

Les moustaches des deux modèles ne dépassent quasiment jamais $\pm 3\%$, soulignant la robustesse de l'approche QSPR pour l'ensemble des familles chimiques. Quelques valeurs extrêmes apparaissent néanmoins pour les dérivés peu fréquents (notamment les azides [N₃] et les complexes métalliques [MLn]), ces outliers résultant vraisemblablement du faible effectif. Par comparaison avec les modèles CG, la méthode QSPR conserve une cohérence globale similaire, tout en offrant un léger raffinement des distributions d'erreur pour les modèles SVM grâce à la richesse des 163 descripteurs moléculaires.

La dissymétrie modérée des boîtes et l'étroitesse des moustaches confirment que l'information topologique et électronique capturée par les descripteurs QSPR permet aux SVM et DNN d'atteindre une performance prédictive uniforme sur l'ensemble des familles ioniques. Les rares écarts majeurs soulignent toutefois l'intérêt d'un échantillonnage plus dense pour les catégories sous-représentées.

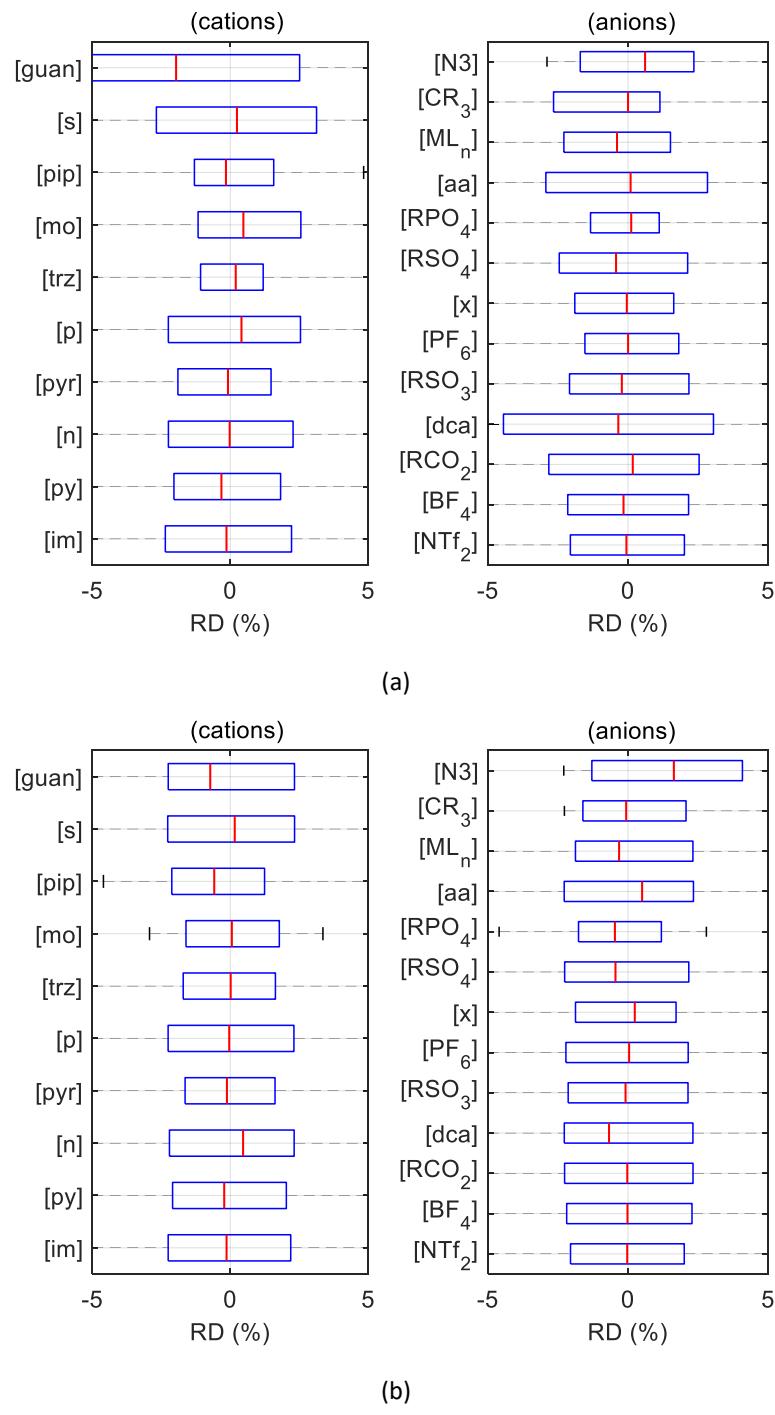


Figure 49: Distribution des écarts relatifs (ER) en fonction des familles cationiques et anioniques. La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieur et supérieur de la boîte représentent respectivement le premier et le troisième quartile. (a) Résultats du modèle QSPR-DNN. (b) Résultats du modèle QSPR-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.

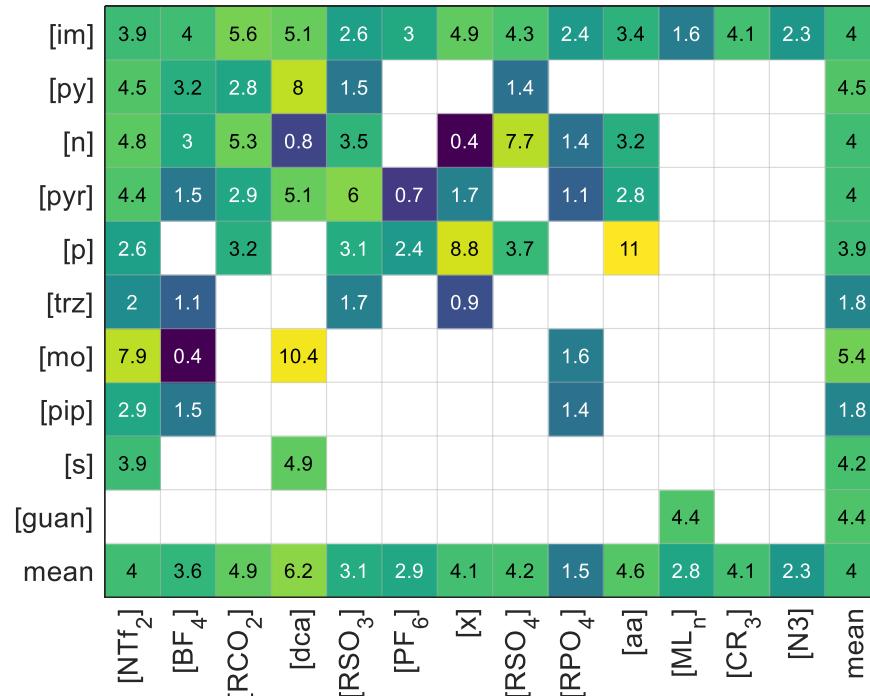
III.4.2.B Écarts relatifs selon les combinaisons cation-anion

La Figure 50 présente deux matrices de chaleur illustrant la distribution des écarts relatifs (ER %) pour chacune des combinaisons de familles cationiques et anioniques, respectivement pour le modèle QSPR-DNN (a) et le modèle QSPR-SVM (b). Chaque cellule renseigne l'écart moyen en pourcentage observé sur l'ensemble de test pour une paire spécifique ; les moyennes par famille (lignes et colonnes « mean ») offrent une vue d'ensemble de la performance globale.

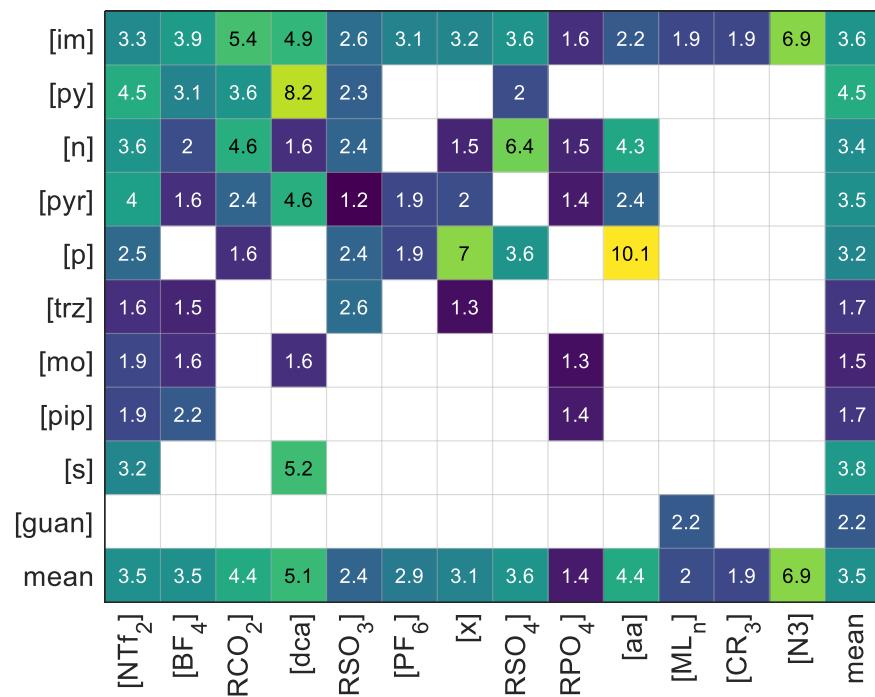
Pour le QSPR-DNN, les maxima d'erreur apparaissent notamment pour les paires Morpholinium [mo]–Carboxylates [dca] (10,4 %) et Phosphonium [p]– sulfonates [RSO₃] (3,1 %), tandis que plusieurs combinaisons courantes (Imidazolium [im], Pyridinium [py], Ammonium [n]) restent sous la barre des 5 %. En comparaison, le QSPR-SVM limite les écarts extrêmes : le pic ER de 10,1 % se rencontre pour Phosphonium [p]– phosphates [aa] et seuls quelques dérivés peu représentés dépassent 6 %. Globalement, les moyennes par famille sont plus homogènes sous SVM (mean cations \simeq 3,5 %; mean anions \simeq 3,5 %) que sous DNN (mean cations \simeq 4,0 %; mean anions \simeq 4,0 %), témoignant d'une meilleure uniformité de la QSPR-SVM.

Il est à noter que l'abondance des jeux de données influe clairement sur la fiabilité des estimations : les familles surreprésentées ; imidazolium (3162 points), NTf₂ (2385 points) ou pyridinium (700 points), affichent systématiquement des ER < 5 %, attestant de la robustesse des prédictions. À l'inverse, les complexes métalliques ([MLn], 67 points), génèrent quelques valeurs extrêmes 6,9 %.

En fin de compte, ces cartes thermiques mettent en évidence que, pour prédire la conductivité électrique, il est essentiel de tenir compte du contexte chimique spécifique des cations et anions tout en assurant un échantillonnage suffisamment dense pour renforcer la fiabilité des combinaisons marginales ; elles illustrent par ailleurs la stabilité légèrement supérieure du modèle QSPR-SVM face à la diversité des familles ioniques, guidant ainsi les priorisations expérimentales futures.



(a)



(b)

Figure 50: Distributions des écarts relatifs ER(%) en fonction des différentes familles de cations et d'anions. (a) modèle QSPR-DNN, (b) modèle QSPR-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 5.

III.4.2.C Distributions d'écart relatifs moyens de prédition

L'analyse conjointe des distributions des erreurs relatives moyennes (ER %) pour le DNN et le SVM révèle deux profils complémentaires : on observe que le DNN maximise la proportion de prédictions très précises (27,0 % des points à $\text{ER} \leq 1\%$ et 62,4 % à $\text{ER} \leq 3\%$) mais présente une longue queue d'erreurs plus importantes (6,6 % au-delà de 10 % et 1,3 % d'outliers $> 25\%$), tandis que le SVM concentre 80,2 % des observations dans l'intervalle 0–3 % (pic à 2–3 % avec 33,8 %) et limite drastiquement les erreurs extrêmes (moins de 4,6 % au-delà de 10 % et 2,1 % au-delà de 25 %), traduisant une régularisation qui privilégie la robustesse aux dépens d'une fraction d'erreurs ultra-fines ($< 1\%$). Cette dualité met en lumière un compromis fondamental : le DNN excelle en précision locale grâce à sa richesse non linéaire, mais souffre d'une variabilité ponctuelle, alors que le SVM, par son noyau fixe et sa forte pénalisation, délivre une distribution d'erreurs plus resserrée et maîtrisée

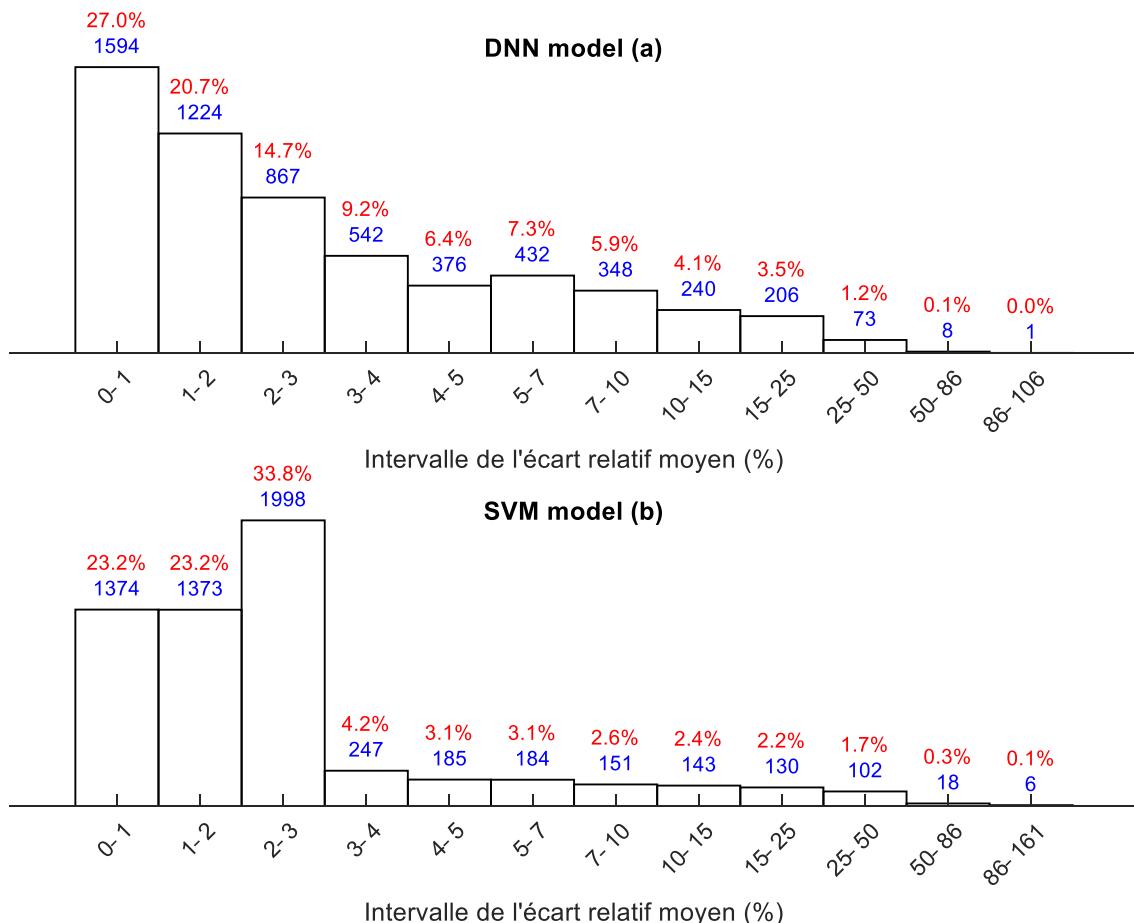


Figure 51: Distributions des points de données en fonction des erreurs relatifs moyens ER (%). (a) modèle DNN, (b) modèle SVM.

En conclusion, sur la base des informations fournies par la Figure 51, les deux modèles, DNN et SVM, présentent des profils d'erreurs complémentaires : le DNN maximise la précision fine en concentrant plus d'un quart de ses prédictions sous 1 % d'écart au prix d'une queue d'outliers, tandis que le SVM, par sa régularisation, offre une robustesse exceptionnelle en limitant drastiquement les erreurs extrêmes, mais lisse davantage ses prédictions avec un pic d'écart autour de 2–3 %.

III.4.3 Comparaison des modèles proposés par rapport aux ceux existants

Le Tableau 9 présente une comparaison quantitative entre les modèles développés dans le cadre de ce travail et ceux déjà publiés dans la littérature. Les méthodes sont classées selon le nombre décroissant de points de données afin de mettre en perspective à la fois la performance prédictive et la robustesse statistique de chaque approche.

Les quatre modèles proposés dans cette étude ; GC/DNN, GC/SVM, QSPR/DNN et QSPR/SVM, ont été entraînés et testés sur un ensemble de données exceptionnel comprenant 7598 points et 520 LI, surpassant en échelle la plupart des travaux précédents. Cette analyse repose sur quatre métriques principales : l'erreur absolue relative moyenne (ERAM), l'erreur absolue moyenne (EAM), racine de l'erreur quadratique moyenne (REQM) et le coefficient de détermination (R^2), permettant une évaluation complète de la précision et de la robustesse des modèles.

Les modèles proposés affichent des performances remarquables. Le modèle GC/DNN atteint un ERAM de 4,60%, un EAM de 0,042 S/m, un REQM de 0,140 et un R^2 de 0,9977, démontrant une précision exceptionnelle et une excellente capacité à capturer les relations complexes des données. Le modèle GC/SVM, avec un ERAM de 5,47 %, un EAM de 0,046 S/m, un REQM de 0,156 et un R^2 de 0,9974, suit de près, tandis que les modèles QSPR/DNN (ERAM: 4,73 %, EAM : 0,049 S/m, REQM : 0,321, R^2 : 0,9987) et QSPR/SVM (ERAM: 4,51 %, EAM : 0,044 S/m, REQM : 0,208 , R^2 : 0,9992) confirment l'efficacité des approches combinant GC ou QSPR avec des techniques d'apprentissage automatique avancées.

En comparaison, les modèles existants présentent des performances variables, souvent limitées par des ensembles de données moins vastes ou des méthodes moins sophistiquées. Par exemple, Baskin et al. (2022)[8], avec le modèle QSPR/TransCNF

appliqué à 7598 points, obtiennent un EAM de 0,299, un REQM de 0,461 et un R² de 0,594, des valeurs nettement inférieures à celles des modèles proposés. De même, Abdullah et al. (2022), avec leur approche QSPR/GNN+MLP sur 2684 points, rapportent un EAM de 0,470 S/m, un REQM de 0,677 et un R² de 0,937, loin derrière la précision des modèles DNN et SVM de cette thèse. Le modèle QSPR/GBR de Karakasidis et al. (2022) montre des lacunes encore plus marquées, avec un ERAM de 271,6 % pour 2274 points, illustrant les limites du gradient boosting dans ce contexte.

Certains travaux antérieurs atteignent des performances compétitives, mais sur des ensembles de données plus restreints. Chen et al. (2020)[9], par exemple, rapportent un ERAM de 3,62 %, un EAM de 0,03 S/m, un REQM de 0,074 et un R² de 0,9973 pour 1578 points (57 ILs), tandis que Gharagheizi et al. (2014)[10], avec un modèle GC/LSSVM, obtiennent un ERAM de 3,3 % et un R² de 0,997 pour 1077 points (54 ILs). Ces résultats, bien que solides, manquent de généralisabilité en raison de la taille et de la diversité limitées de leurs données par rapport aux 520 ILs étudiés ici. Les approches traditionnelles, comme le modèle UNIFAC GC de Zhao et al. (2017)[11] (ERAM : 4,1 % pour 784 points) ou celui de Gardas et al. (2009)[12] (ERAM : 4,57 % pour 307 points), restent également en deçà des performances des modèles proposés.

Enfin, des études comme celle de Zi Kang Koi et al. (2021), avec un ERAM de 7,15 % et un R² de 0,9906 pour 239 points, ou celle de Nakhaei-Kohani et al. (2022)[8], avec un REQM de 0,1976 et un R² de 0,9935 pour 2625 points, soulignent les défis de la prédiction précise sur des bases de données réduites. En revanche, les modèles de cette thèse, bénéficiant d'un ensemble de données vaste et diversifié, maintiennent une précision élevée et une robustesse inégalée, surpassant la majorité des approches existantes.

Tableau 9: Comparaison des performances des modèles proposés par rapport aux modèles existants.

Authors	Year	Number of data points	Number of ILs	Method/technique	ERAM (%)	EAM S/m	REQM	R ²
This work	2025	7598	520	GC/DNN	4,60	0.042	0.140	0.9977
This work	2025	7598	520	GC/SVM	5,47	0.046	0.156	0.9974
This work	2025	7598	520	QSPR/DNN	4,73	0,049	0,321	0,9987
This work	2025	7598	520	QSPR/SVM	4,51	0.044	0.208	0.9992
Igor baskin et al. [8]	2022	7598	474	QSPR/ TransCNF	_ ^a	0.299	0.461	0.594
Dhakal and Shah [13]	2022	2869	397	QSPR/ XGBoost	_ ^a	0.018 ^b	0,032 ^b	0.9972
Abdullah et al.[14]	2022	2684	_ ^a	QSPR/GNN+ MLP	_ ^a	0.470	0.677	0.937
Nakhaei-Kohani et al. [15]	2022	2625	225	GC/KNN	_ ^a	_ ^a	0.1976	0.9935
Karakasidis et al. [16]	2022	2274	_ ^a	QSPR/GBR	271.6	0.137	_ ^a	0.9888
Ke-Jun Wu [17]	2016	2195	188	GC	6.02	0.093	0.271	0.9895
Chen et al. [9]	2020	1578	57	GC	3.62	0.03	0.074	0.9973
Dhakal and Shah [18]	2021	1323	89	QSPR/SVM	4.13 ^b	0.017 ^b	0.057 ^b	0.993
Gharagheizi et al. [10]	2014	1077	54	GC/ LSSVM	3.3	_ ^a	0.16	0.997
Gharagheizi et al. [19]	2013	977	54	QSPR/LSSVM	1.9	_ ^a	0.06	0.997
Nan Zhao et al. [11]	2017	784	38	UNIFAC GC	4.1	_ ^a	_ ^a	_ ^a
Gardas et al. [12]	2009	307	15	GC	4.57	_ ^a	_ ^a	_ ^a
Zi Kang Koi et al. [20]	2021	239	25	COSMO-RS/SVM	7.15	_ ^a	_ ^a	0.9906

^a n'est pas mentionné dans le document original et ne peut pas être calculé. ^b σ donné en unités logarithmiques. GBR : gradient boosting regressor.

III.5 CONCLUSION

Ce chapitre a proposé une étude approfondie de la prédiction de la conductivité électrique des liquides ioniques, reposant sur une base de données diversifiée de 6014 points pour 373 LI et des modèles de pointe. La collecte et la révision rigoureuse des données expérimentales ont permis d'établir un ensemble fiable, tandis que l'analyse des effets de la température, de la structure ionique et de la pression a fourni des insights clés pour orienter la modélisation. Les approches basées sur les contributions de groupes (CG) et les descripteurs QSPR, combinées à des techniques comme les DNN et SVM, ont démontré une précision remarquable, avec des erreurs souvent dans la plage des incertitudes expérimentales.

Les résultats soulignent la supériorité des modèles DNN et SVM, tant en termes de précision que de généralisabilité, surpassant les approches existantes grâce à un corpus de données plus vaste et à des méthodes avancées. Ces avancées offrent des outils prédictifs prometteurs pour des applications industrielles et scientifiques. Toutefois, des limites subsistent, notamment pour les familles chimiques sous-représentées, suggérant la nécessité d'un échantillonnage plus dense et d'explorations de modèles hybrides. En somme, ce chapitre pose des bases solides pour des recherches futures visant à optimiser la prédiction des propriétés des LI dans des contextes variés.

III.6 RÉFÉRENCE

- [1] M. Armand, F. Endres, D.R. MacFarlane, H. Ohno, B. Scrosati, Ionic-liquid materials for the electrochemical challenges of the future, *Nat. Mater.* 8 (2009) 621–629. <https://doi.org/10.1038/nmat2448>.
- [2] Z. Chen, J. Chen, Y. Qiu, J. Cheng, L. Chen, Z. Qi, Z. Song, Prediction of Electrical Conductivity of Ionic Liquids: From COSMO-RS Derived QSPR Evaluation to Boosting Machine Learning, *ACS Sustain. Chem. Eng.* 12 (2024) 6648–6658. <https://doi.org/10.1021/acssuschemeng.4c00307>.
- [3] M. Hosseinzadeh, A. Hemmati-Sarapardeh, F. Ameli, F. Naderi, M. Dastgahi, A computational intelligence scheme for estimating electrical conductivity of ternary mixtures containing ionic liquids, *J. Mol. Liq.* 221 (2016) 624–632.
- [4] P. Nancarrow, A. Al-Othman, D.K. Mital, S. Döpking, Comprehensive analysis and correlation of ionic liquid conductivity data for energy applications, *Energy* 220 (2021). <https://doi.org/10.1016/j.energy.2021.119761>.
- [5] Z. Shi, F. Song, C. Ji, Y. Xiao, C. Peng, H. Liu, QSPR Models Based on Multiple Linear Regression and Back-Propagation Artificial Neural Network for Predicting Electrical Conductivity of Ionic Liquids, *Ind. Eng. Chem. Res.* 63 (2024) 4571–4584.
- [6] M.A. Zebida, K. Argoub, A.M. Benkouider, A. Yahiaoui, K. Toubal, A. Hachemaoui, Machine learning coupled with group contribution for predicting the electrical conductivity of ionic liquids with experimental accuracy, *Fluid Phase Equilib.* 579 (2024) 114024. <https://doi.org/10.1016/j.fluid.2024.114024>.
- [7] R. Datta, R. Ramprasad, S. Venkatram, Conductivity prediction model for ionic liquids using machine learning, *J. Chem. Phys.* 156 (2022) 214505.
- [8] I. Baskin, A. Epshtain, Y. Ein-Eli, Benchmarking machine learning methods for modeling physical properties of ionic liquids, *J. Mol. Liq.* 351 (2022) 118616. <https://doi.org/10.1016/j.molliq.2022.118616>.
- [9] Y. Chen, Y. Cai, K. Thomsen, G.M. Kontogeorgis, J.M. Woodley, A group contribution-based prediction method for the electrical conductivity of ionic liquids, *Fluid Phase Equilib.* 509 (2020) 112462. <https://doi.org/10.1016/j.fluid.2020.112462>.
- [10] F. Gharagheizi, P. Ilani-Kashkouli, M. Sattari, A.H. Mohammadi, D. Ramjugernath, D. Richon, Development of a LSSVM-GC model for estimating the electrical conductivity of ionic liquids, *Chem. Eng. Res. Des.* 92 (2014) 66–79. <https://doi.org/10.1016/j.cherd.2013.06.015>.
- [11] N. Zhao, J. Jacquemin, The development of the UNIFAC-CONDUCT model as a novel approach for the estimation of the conductivity of pure ionic liquids, *Fluid Phase Equilib.* 449 (2017) 60–67. <https://doi.org/10.1016/j.fluid.2017.06.010>.
- [12] R.L.G. and J.A.P. Coutinho, PVAm–PIP/PS composite membrane with high performance for CO₂/N₂ separation, *AIChE J.* 55 (2009) 1274–1290. <https://doi.org/10.1002/aic>.
- [13] P. Dhakal, J.K. Shah, A Generalized Machine Learning Model for Predicting Ionic Conductivity for Ionic Liquids Design, System, and Application Statement A Generalized Machine Learning Model for Predicting Ionic Conductivity for Ionic Liquids' †, (2022).
- [14] M. Abdullah, K.C. Lethesh, A.A.B. Baloch, M.O. Bamgbopa, Comparison of molecular and structural features towards prediction of ionic liquid ionic conductivity for electrochemical applications, *J. Mol. Liq.* 368 (2022) 120620.

- [15] R. Nakhaei-Kohani, S. Ali Madani, S.P. Mousavi, S. Atashrouz, A. Abedi, A. Hemmati-Sarapardeh, A. Mohaddespour, Machine learning assisted Structure-based models for predicting electrical conductivity of ionic liquids, *J. Mol. Liq.* 362 (2022). <https://doi.org/10.1016/j.molliq.2022.119509>.
- [16] T.E. Karakasidis, F. Sofos, C. Tsonos, The Electrical Conductivity of Ionic Liquids: Numerical and Analytical Machine Learning Approaches, *Fluids.* 7 (2022). <https://doi.org/10.3390/fluids7100321>.
- [17] K.J. Wu, H. Luo, L. Yang, Structure-based model for prediction of electrical conductivity of pure ionic liquids, *AIChE J.* 62 (2016) 3751–3762. <https://doi.org/10.1002/aic.15429>.
- [18] P. Dhakal, J.K. Shah, Developing machine learning models for ionic conductivity of imidazolium-based ionic liquids, *Fluid Phase Equilib.* 549 (2021). <https://doi.org/10.1016/j.fluid.2021.113208>.
- [19] F. Gharagheizi, S.A. Mirkhani, M.H. Keshavarz, N. Farahani, K. Tumba, A molecular-based model for prediction of liquid viscosity of pure organic compounds: A quantitative structure property relationship (QSPR) approach, *J. Taiwan Inst. Chem. Eng.* 44 (2013) 359–364. <https://doi.org/10.1016/j.jtice.2012.12.015>.
- [20] Z.K. Koi, W.Z.N. Yahya, K.A. Kurnia, Prediction of ionic conductivity of imidazolium-based ionic liquids at different temperatures using multiple linear regression and support vector machine algorithms, *New J. Chem.* 45 (2021) 18584–18597.

CHAPITRE IV : PREDICTION DE LA CONDUCTIVITE THERMIQUE DES LIQUIDES IONIQUES

CHAPITRE IV: PREDICTION DE LA CONDUCTIVITE THERMIQUE DES LIQUIDES IONIQUES

IV.1 INTRODUCTION

Les liquides ioniques (LI) constituent une classe de fluides à base d'ions fondus affichant des propriétés physico-chimiques remarquables : stabilité thermique et chimique, non-volatilité, large fenêtre électrochimique, et modifiabilité structurale.[1][2] Outre leur conductivité électrique, ces liquides présentent une conductivité thermique (λ) souvent faible mais néanmoins cruciale pour des applications variées ; caloporeurs pour le transfert de chaleur, fluides de refroidissement dans les systèmes énergétiques, supports thermiques en capture d'énergie, ou encore milieux réactionnels où l'évacuation de la chaleur joue un rôle clé [3][4][5]. Cependant, la prédiction fiable de la conductivité thermique des LI reste un défi, en raison des interactions ioniques complexes, de la diversité des architectures moléculaires et de la dépendance marquée aux conditions thermodynamiques (température, pression) [6][7].

Ce chapitre s'inscrit dans la continuité méthodologique du Chapitre III, qui portait sur la prédiction de la conductivité électrique, en adoptant la même structure : collecte et traitement des données, analyse exploratoire des données, méthodologie de modélisation, résultats et discussion et finalement conclusion.

L'objectif premier de ce chapitre est donc d'élaborer des modèles robustes et généralisables pour la conductivité thermique des LI, en s'appuyant sur une base de données étendue et sur des méthodes d'apprentissage automatique de pointe, afin de fournir aux expérimentateurs et aux ingénieurs des outils prédictifs fiables pour la conception et l'optimisation de fluides ioniques à usage thermique.

IV.2 COLLECTE ET TRAITEMENT DES DONNEES

IV.2.1 Collecte de données:

La base de données expérimentales destinée à la modélisation de la conductivité thermique des liquides ioniques a été constituée à partir de la base de données spécialisée ILThermo (version 2.0), largement utilisée pour les propriétés thermodynamiques des LI, à l'aide d'un script Python. Les 1 092 points de mesure retenus couvrent 88 liquides ioniques différents, issus de sept grandes familles de cations, voir Figure 66, (notamment imidazoliums, pyridiniums, ammoniums quaternaires, phosphoniums, sulfoniums, pyrrolidinium , et piperidinium) et de treize familles d'anions (borates, hexafluorophosphates, bis(trifluorométhylsulfonyl)imides, sulfonates, sulfates, dicyanamide ,aminoacids ,carboxylates ,Methanides ,inorganics, metal complexes ,phosphates ,phosphinate).

Afin de garantir la qualité et la comparabilité des données, seuls les liquides ioniques de pureté $\geq 99\%$ ont été conservés, limitant ainsi les biais potentiels liés aux impuretés. Les mesures de conductivité thermique, converties systématiquement en $\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$, couvrent une plage de températures de 273 K à 438 K, majoritairement réalisées à pression atmosphérique (0,1 MPa), avec quelques points supplémentaires jusqu'à 65 MPa pour explorer l'influence de la pression. Les différentes techniques expérimentales (hot-wire transient, méthode 3ω , calorimétrie à balayage, ...) ont été soigneusement annotées afin d'identifier d'éventuels écarts méthodologiques.

Chaque point de mesure est accompagné d'une fiche complète mentionnant le nom et la formule chimique du liquide ionique, ses familles cationique et anionique, la température et la pression de mesure, la valeur expérimentale de la conductivité thermique, la valeur d'incertitude de mesure, la méthode expérimentale utilisée et la référence bibliographique.

Grâce à cette collecte rigoureuse, la base de données de 1 092 points pour 88 liquides ioniques offre une couverture chimique et thermodynamique étendue, constituant un socle solide pour l'analyse exploratoire et le développement des modèles prédictifs décrits dans les sections suivantes.

IV.2.2 Révision des données

La phase de révision visait à garantir la fiabilité et la cohérence des données expérimentales de conductivité thermique. Pour ce faire, nous avons appliqué la même méthodologie de révision en sept étapes successives, telle qu'exposée au Chapitre II (cf. II.2.1) et mise en œuvre dans le Chapitre III.

La première étape a consisté à supprimer 13 points de données présentant des données métastables c'est-à-dire les cristaux et les liquides métastables. La seconde a ciblé les mesures dupliquées ou redondantes issues de sources multiples, entraînant l'élimination de 20 points supplémentaires pour ne conserver qu'une seule valeur représentative par condition expérimentale. Enfin, dans la troisième étape, 34 points supplémentaires ont été supprimés après une vérification croisée des valeurs contradictoires (valeurs multiples pour le même liquide à des températures et des pressions identiques). Aucun point de donnée n'a été supprimé au cours des étapes 4 à 7.

Au terme de ces trois étapes de révision, la base de données compte désormais **1 025 points** de données expérimentales, couvrant 79 liquides ioniques. Ces liquides sont répartis sur 46 cations et 32 anions distincts, assurant une diversité chimique suffisante pour l'apprentissage et la robustesse des modèles développés dans les sections suivantes. Le Tableau 10 présente en détail la composition finale de la base de données après révision, en indiquant pour chaque famille cationique et anionique le nombre de liquides ioniques correspondants ainsi que le nombre de points de mesure associés. On y trouve également la répartition des 1 025 mesures finales selon les plages de température (273–438 K) et les conditions de pression (0,1–65 MPa). Ce tableau fournit ainsi un aperçu clair de la diversité chimique et thermodynamique couverte par l'étude. La Figure 53 présente, sous la forme d'une matrice, le nombre de points de mesure de conductivité thermique disponibles pour chaque combinaison cation–anion après le traitement des données. De même, la Figure 54 illustre, également en matrice, le nombre de sous-ensembles⁸ de valeurs expérimentales étudiés pour chaque couple cation–anion suite à ce même traitement. Cette visualisation met en évidence les déséquilibres

⁸ Un ensemble de données issus de la même source bibliographie et appartenant à une famille chimique donnée.

d'échantillonnage initiaux entre familles chimiques et sert de point de départ pour la phase de révision et d'homogénéisation des données.

Tableau 10 : Résumé de la base de données sur la conductivité thermique par familles d'ions

IL family	Data points	Data sets	ILs	T min (K)	T max (K)	P min (kPa)	P max (kPa)	λ min (W·m ⁻¹ ·K ⁻¹)	λ max (W·m ⁻¹ ·K ⁻¹)	$\Delta\lambda$ min (W·m ⁻¹ ·K ⁻¹)	$\Delta\lambda$ max (W·m ⁻¹ ·K ⁻¹)
Cations											
imidazolium (im)	597	70	43	293	357.76	100	6.5E+04	9.4E-02	21E-02	0.0E+00	2.0E-02
phosphonium (p)	153	21	18	288	358.21	101.325	101.325	1.21E-02	1.63E-02	2E-03	2.0E-02
pyrrolidinium (pyr)	78	04	04	293	438.07	100	6.5E+04	1.05E-01	1.55 E-01	3.6E-03	2.0E-02
pyridinium (py)	44	08	08	300	343.15	100	101.325	1.176E-01	2.05E-01	02E-03	1.0E-02
ammonium (n)	38	06	04	273.15	353.15	101.325	101.325	1.242E-01	2.55E-01	1.0E-03	1.0E-02
piperidinium (pip)	36	01	01	283.15	313.15	100	6.5E+04	1.032E-01	1.23E-01	8.3E-03	9.9E-03
sulfonium (s)	36	01	01	283.15	313.15	100	6.5E+04	1.136E-01	1.31 E-01	9.1E-03	1.1E-02
Anions											
NTf ₂ derivatives (NTf ₂)	290	30	19	293	438.07	100	6.5E+04	1.032E-01	1.44 E-01	0.0E+00	2.0E-02
BF ₄ derivatives	117	14	11	293	353.15	100	2.0E+04	1.44 E-01	1.99 E-01	2.0E-03	1.0E-02
sulfates (RSO ₄)	103	10	05	293	353.51	100	6.5E+04	1.53E-01	2.05E-01	2.0E-03	2.0E-02
sulfonates (RSO ₃)	76	08	03	293	355.51	100	6.5E+04	1.37E-01	2.121E-01	4.2E-03	2.0E-02
dicyanamide	63	05	03	323	353.15	100	6.5E+04	1.42E-01	2.053E-01	6.4 E-03	2.0E-02
aminoacids (aa)	62	10	08	298.	353.15	101.325	101.325	1.46E-01	2.55E-01	1.0E-03	1.0E-02
carboxylates (RCO ₂)	61	08	07	288	358.21	100	101.325	1.48E-01	2.137E-01	6.4 E-03	1.5 E-02
PF ₆ derivatives (PF ₆)	55	07	06	293	355.07	100	2.0E+04	0.94 E-01	1.48E-01	06E-03	2.0E-02
methanides	49	06	05	273.15	353.15	101.325	101.325	1.593E-01	1.966E-01	5.4 E-03	1.0E-02
inorganics (x)	36	05	05	293	334.5	100	2.0E+04	1.47E-01	2.1E-01	5.0 E-03	2.0E-02
metal complexes	36	01	01	283.15	313.15	100	6.5E+04	1.31E-01	1.51E-01	1.1E-02	1.2E-02
phosphates (RPO ₄)	19	06	03	300	353.15	100	101.325	1.718E-01	2.38E-01	6.4 E-03	2.0E-02
phosphinate	09	01	01	282.47	353.62	101.325	101.325	1.31E-01	1.37E-01	2.0E-02	2.0E-02

Cation					
	[Im]	[py]	[pip]	[pyr]	
	[n]	[p]		[s]	
Anion					
	[RSO3]	[RCO2]	[X]	[BF4]	[PF6]
	[RPO3]	[RPO4]	[RSO4]	[AA]	[SbF6]
	[dca]	[NTF2]	[CR3]		

Figure 52 : Familles chimiques cationiques et anioniques considérées dans cette étude.

Les abréviations des familles chimiques sont détaillées dans le Tableau 10.

[im]	126	87	81	71	62	44	14	39		49	33	18	
[p]	20		15	6		6	44	9	34				9
[pyr]	62				6			7					
[n]	21			6		13			19				
[py]	21	27	7										
[s]	36												
[pip]	33												
	[NTf ₂]	[BF ₄]	[RSO ₄]	[RSO ₃]	[dca]	[x]	[RCO ₂]	[PF ₆]	[aa]	[CR ₃]	[ML _n]	[RPO ₄]	[PO ₂]

Figure 53 : Nombre de point de données expérimentales de la conductivité thermique selon différentes combinaisons cation-anion. Les abréviations des familles chimiques sont détaillées dans Tableau 10.

[im]	21	33	3	2	2	20	1	10		26	1	2	
[p]	2		1	6		1	5	1	30				9
[pyr]	2				1			1					
[n]	21			6		13			19				
[py]	18	19	1										
[s]	1												
[pip]	1												
	[NTf ₂]	[BF ₄]	[RSO ₄]	[RSO ₃]	[dca]	[x]	[RCO ₂]	[PF ₆]	[aa]	[CR ₃]	[ML _n]	[RPO ₄]	[PO ₂]

Figure 54 : Nombre de sous ensembles⁹ de valeurs expérimentales de la conductivité thermique selon différentes combinaisons cation-anion. Les abréviations des familles chimiques sont détaillées dans Tableau 10.

⁹ Un ensemble de données issus de la même source bibliographie et appartenant à une famille chimique donnée.

IV.2.3 Analyse de données

IV.2.3.A Effet de la température sur la conductivité thermique d'ILs

Les données expérimentales montrent que, pour la majorité des ILs, la conductivité thermique diminue légèrement avec l'augmentation de la température. Cette tendance, bien que modérée, est cohérente avec le comportement des liquides où l'agitation thermique perturbe les mécanismes de transfert de chaleur. Par exemple, sur une plage de température de 273 K à 373 K, les ILs à base d'imidazolium, comme [im-4,1,1][NTf₂], présentent une décroissance quasi-linéaire de λ , tandis que les ILs à base de pyrrolidinium, comme [py-4,1,4][NTf₂], montrent une diminution moins prononcée. Ces différences reflètent l'influence de la structure du cation sur la réponse thermique.

Pour illustrer ces observations, la Figure 55 présente la conductivité thermique en fonction de la température pour plusieurs LI représentatifs. Les courbes mettent en évidence une décroissance de λ (en W·m⁻¹·K⁻¹) avec la température, avec des variations subtiles selon la famille chimique. Cette diminution est toutefois plus ou moins accentuée selon les familles chimiques : les imidazoliums à chaînes alkyles courtes perdent environ 0,07 W·m⁻¹·K⁻¹ sur la plage étudiée, tandis que les ammoniums quaternaires, dont les réseaux ioniques sont plus compacts, ne voient λ chuter que de 0,015 W·m⁻¹·K⁻¹. Ces fluctuations soulignent l'importance de la densité locale et de l'organisation moléculaire dans le transfert thermique.

Ces observations soulignent l'importance de prendre en compte la sensibilité thermique propre à chaque combinaison cation–anion pour modéliser avec précision la conductivité thermique.

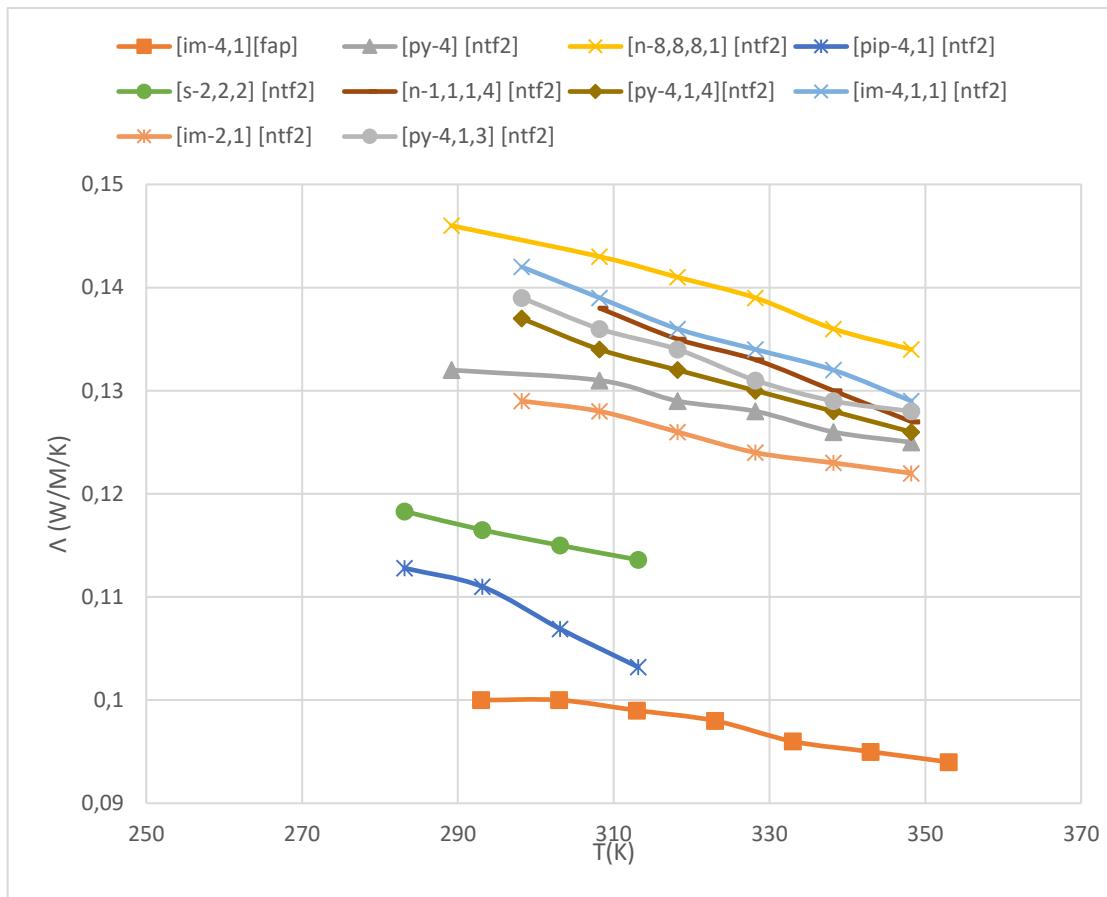


Figure 55 : Variation de la conductivité thermique (λ) des liquides ioniques en fonction de la température (en K) pour des ILs sélectionnés.

IV.2.3.B Effet de la pression sur la conductivité thermique d'Ils

L'influence de la pression sur la conductivité thermique (λ) des liquides ioniques a été évaluée à partir d'un sous-ensemble de mesures réalisées jusqu'à 65 MPa (Figure 56), à une température constante (par exemple, 298 K). À pression atmosphérique (0,1 MPa), les variations de λ sont essentiellement dictées par la température, comme vu précédemment. En revanche, sous haute pression, on observe systématiquement une augmentation de λ : typiquement, une élévation de la pression de 0,1 MPa à 65 MPa conduit à un accroissement de λ de 5 à 15 % selon les liquides ioniques considérés. Cette tendance s'explique par la réduction du volume libre et la densification du fluide, favorisant un meilleur couplage des modes vibratoires entre ions et, par conséquent, un transport thermique plus efficace.

Les effets de pression diffèrent toutefois selon les architectures ioniques. Les familles dont les cations présentent des chaînes alkyles longues (ammoniums quaternaires)

montrent une sensibilité modérée, avec une augmentation de λ comprise entre 0,010 et 0,015 $\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$ pour $\Delta P \approx 65 \text{ MPa}$. En revanche, les imidazoliums et pyrrolidiniums à chaînes courtes bénéficient davantage de la compaction, avec un gain de λ pouvant atteindre 0,020 $\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$ sur la même plage de pression. Cette hétérogénéité reflète le rôle clé du volume libre initial et de la structuration micro-étriquée du fluide : plus le volume libre est élevé à basse pression, plus la densification induite par la pression a un impact fort sur la facilitation du transport d'énergie.

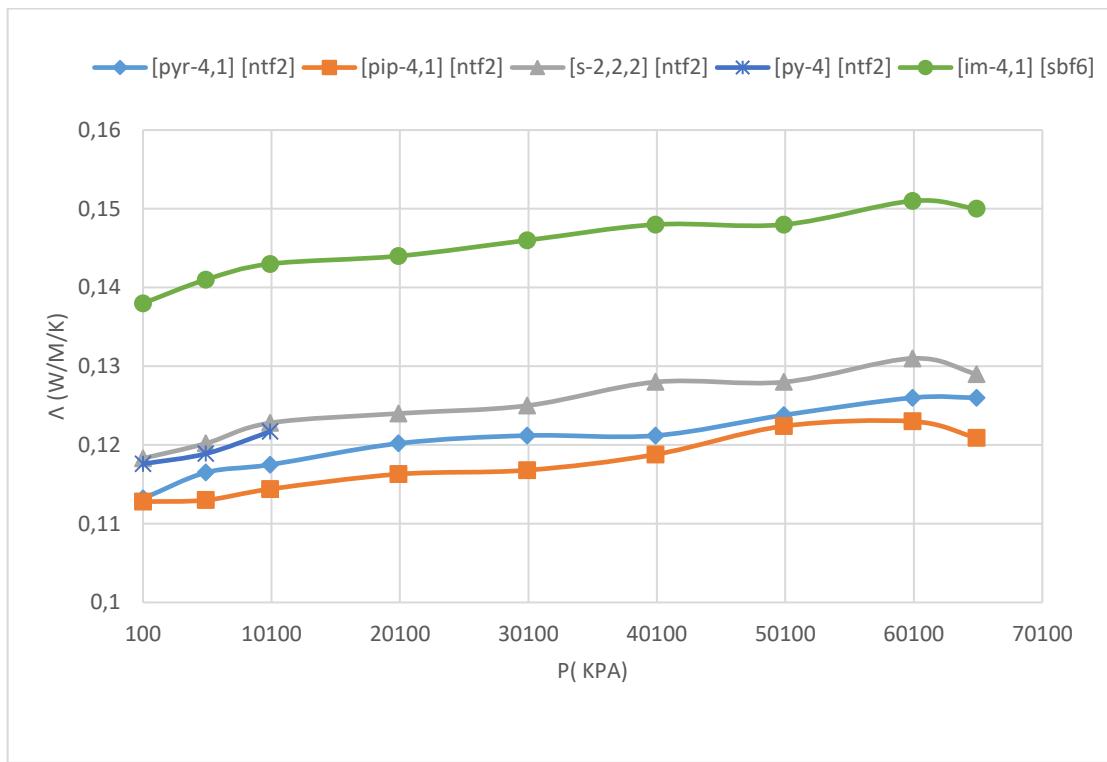


Figure 56 : Influence de la pression sur la conductivité thermique des liquides ioniques

IV.2.3.C Conclusion :

L'analyse conjointe des effets de la température et de la pression sur la conductivité thermique des liquides ioniques met en évidence deux comportements complémentaires. D'une part, la plupart des LI étudiés présentent une légère diminution de λ avec l'élévation de la température, due à l'accroissement du volume libre et à la désorganisation des réseaux ioniques, qui limitent le couplage vibratoire. D'autre part, l'application d'une pression élevée favorise la densification du fluide, renforçant le transport thermique par réduction du volume libre et amélioration du couplage des modes vibratoires, avec des gains de conductivité thermique allant jusqu'à 15 %.

Ces variations sont fortement modulées par la nature des cations et des anions : les liquides à chaînes alkyles courtes (imidazoliums, pyrrolidiniums) sont particulièrement sensibles à la fois à la température et à la pression, tandis que les systèmes à chaînes longues (ammoniums quaternaires) présentent des réponses plus atténuées.

IV.3 STRATÉGIE MÉTHODOLOGIQUE DE MODÉLISATION

Dans ce travail, nous avons déployé cinq approches de modélisation pour prédire la conductivité thermique λ des liquides ioniques, à savoir Extreme Gradient Boosting (XGBoost), Forêt aléatoire (RF), Machine à vecteurs de support (SVM), réseau de neurones profond (DNN) et un modèle conventionnel semi-empirique. Chacune de ces méthodes s'appuie sur deux types d'entrées structurales ; contributions de groupes (CG) contenant des cations, anions et substituants, et descripteurs QSPR calculés via RDKit, enrichies par les variables thermodynamiques de température T et de pression P . Pour les CG, on a fragmenté chaque molécule en 63 groupes (26 cationiques, 25 anioniques, 12 substituants), répartis en contributions de premier et de second ordre. Pour les QSPR, un vecteur de 163 descripteurs a été calculé pour chaque LI, encapsulant : taille, polarité, distribution de charge et groupes fonctionnels moléculaire.

Pour le modèle conventionnel, après avoir testé plusieurs lois (Arrhenius, Vogel–Fulcher–Tammann, puissances en P), la relation finalement adoptée tient compte à la fois de l'activation thermique et de la compressibilité :

$$\ln(\lambda) = A + \frac{B}{T - C} + D \cdot \log(P)$$

où T est la température en kelvins (K) et P la pression en kilopascals (kPa), et A, B, C et D sont la contribution de tous les groupes.

Il est important de souligner que le modèle conventionnel est retenu uniquement dans le cadre de l'approche CG, car la décomposition en contributions de groupes permet d'exprimer directement les coefficients de la loi semi-empirique à partir des fragments moléculaires, ce qui n'est pas praticable avec l'approche QSPR fondée sur un vecteur dense de descripteurs non localisés.

Le jeu de données final, constitué de 1 025 mesures pour 79 liquides ioniques, a été divisé en utilisant 5 plis de manière stratifiée, garantissant la diversité des familles

cationiques/anioniques et des conditions thermodynamiques. Les modèles XGBoost, RF, SVM et DNN ont tous été développés sous MATLAB (Statistics & Machine Learning Toolbox v. 2021b), tandis que le modèle conventionnel a été ajusté par régression non linéaire.

L'optimisation des hyper-paramètres pour les quatre modèles d'apprentissage automatique a été conduite au moyen d'une optimisation bayésienne, avec un espace de recherche défini empiriquement pour chaque algorithme. Les domaines de recherche et les valeurs retenues sont synthétisés dans le Tableau 11. Ces réglages ont permis d'atteindre une convergence rapide et une qualité prédictive homogène entre CG et QSPR, tout en évitant le surapprentissage grâce à la validation croisée à 5 plis répétée.

Tableau 11: Espace de recherche des hyperparamètres et détermination de leur sélection optimale à l'aide d'un algorithme d'optimisation bayésienne.

Modèles	Espace de recherche	Paramètres sélectionnés
XGBoost	$MinLeafSize = [0.1 \text{ } 100]$, $NumLearningCycles = [5 \text{ } 1000]$, $LearnRate = [1e-4, 1]$.	$MinLeafSize = 1$, $NumLearningCycles = 133$ $LearnRate = 0.327$.
RF	$MinLeafSize = [1 \text{ } 100]$, $NumLearningCycles = [10 \text{ } 1000]$.	$MinLeafSize = 1$, $NumLearningCycles = 499$
SVM	$C = [1 \text{ } 400]$, $\varepsilon = [0.0001 \text{ } 1]$, $\gamma = [0.1 \text{ } 100]$.	$C = 46$, $\varepsilon = 1.3056 \text{ e-4}$, $\gamma = 6.55$
DNN	$N_Hidden = [1, 5]$, $Layer_N_Size = [1, 60]$, $Activations = [\text{ReLU, sigmoid, tanh}]$, $Lambda = [0, 10]$.	$N_Hidden = 2$, $Layer_N_Size = [20 \text{ } 5]$, $Activations = \text{tanh}$, $Lambda = 0.517e-10$.

IV.4 RÉSULTATS ET DISCUSSION

Les performances des cinq stratégies modélisation (XGBoost, RF, SVM, DNN, modèle conventionnel) ont été évaluées sur les deux jeux de données (calibration et validation externe) au moyen des indicateurs MEA, EQM, ERAM et R^2 . Les résultats se répartissent selon les approches CG et QSPR, mettant en lumière les compromis entre complexité des descripteurs et précision prédictive.

IV.4.1 Méthode de contributions de groupes

Dans le cadre des modèles CG, le Tableau 12 synthétise les performances moyennées d'une validation croisée 5-plis répétée 25 fois. Les algorithmes XGBoost et SVM se distinguent par des EAM en test de seulement 0,003 ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$), avec des ERAM = 1,62 % et des $R^2 = 0.9788$. Le modèle DNN suit de près ($EAM_{test} = 0,003$, $ERAM_{test} = 1,64 \%$, $R^2_{test} = 0,98$), tandis que la Forêt aléatoire affiche une précision légèrement réduite ($EAM_{test} = 0,004$, $ERAM_{test} = 2,24\%$, $R^2_{test} = 0,972$). Le modèle conventionnel, bien qu'offrant encore un ajustement correct ($EAM_{test} = 0,004$, $ERAM_{test} = 2,49 \%$, $R^2_{test} = 0,9583$), reste le moins précis du lot, illustrant la valeur ajoutée des méthodes d'apprentissage automatique.

Tableau 12: Statistiques de performance des GC-modèles sur les jeux d'apprentissage et de test, exprimées en moyennes issues d'une validation croisée à 5-plis, répétée 25 fois selon des partitions aléatoires.

Models	EAM_{train} ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	EAM_{test} ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	ERAM _{train} (%)	ERAM _{test} (%)	REQM _{train}	REQM _{test}	R^2_{train}	R^2_{test}
XGBoost	0.001	0.003	0.74	1.62	0.002	0.004	0.9966	0.9788
RF	0.003	0.004	1.67	2.24	0.004	0.005	0.9818	0.9728
SVM	0.002	0.003	1.07	1.60	0.003	0.004	0.9831	0.9784
DNN	0.002	0.003	1.16	1.64	0.003	0.004	0.9861	0.9801
Conventional	0.003	0.004	2.17	2.49	0.005	0.006	0.9673	0.9583

Visuellement, les nuages de points de la Figure 57a mettent en évidence l'excellent accord entre valeurs prédites et expérimentales pour les modèles XGBoost et SVM, tant en apprentissage qu'en test, tandis que la distribution gaussienne des erreurs (Figure 57b) confirme la concentration des écarts autour de zéro et l'absence de biais systématique. La forme quasi-gaussienne des histogrammes et l'asymétrie minimale indiquent également une bonne hétérogénéité des erreurs.

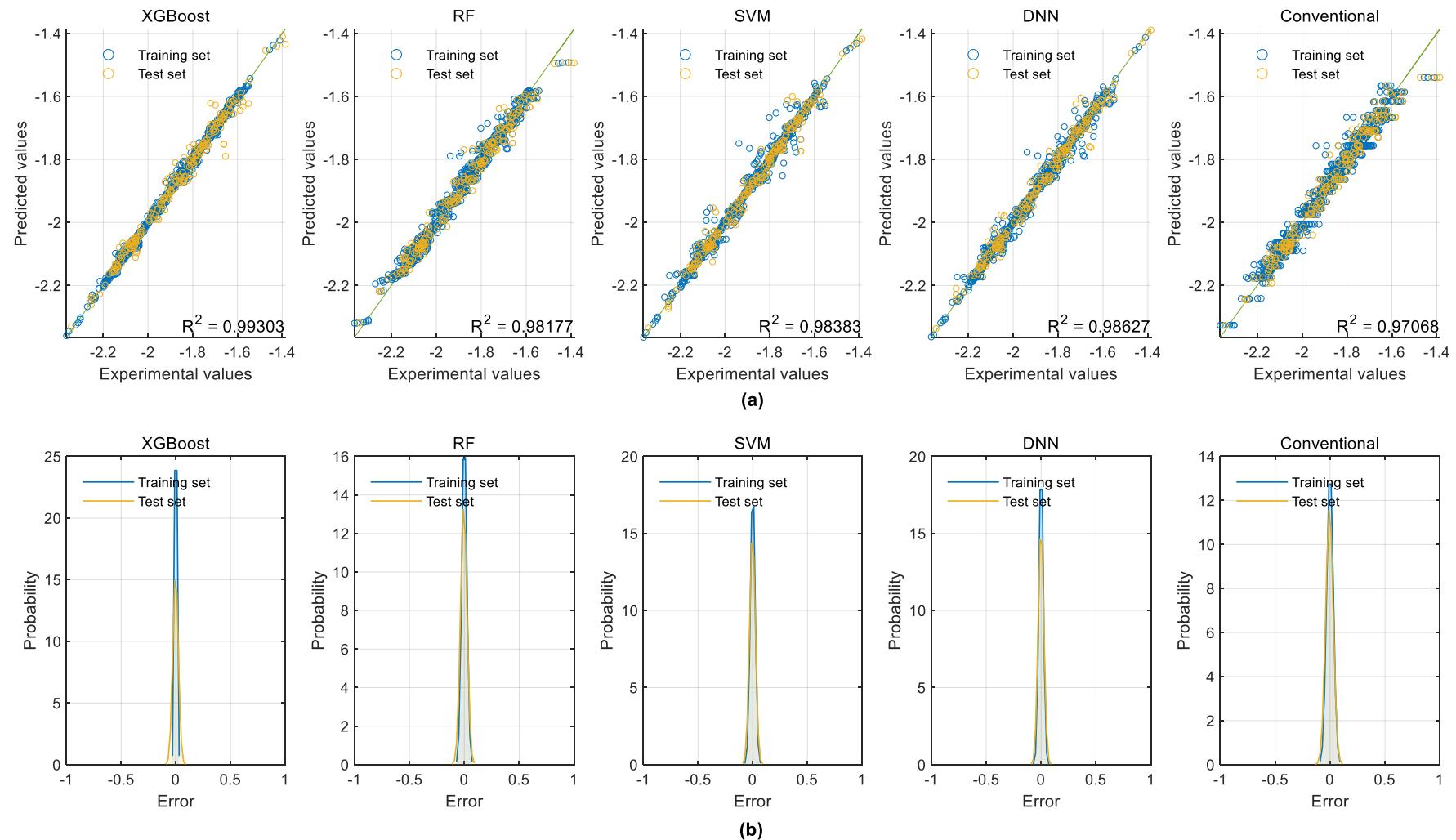


Figure 57: Comparaison des modèles de contribution de groupes : (a) Nuages de points des valeurs prédictives et expérimentales dans l'ensemble d'apprentissage et l'ensemble de test. (b) Distribution gaussienne des erreurs de prédiction dans l'ensemble d'apprentissage et l'ensemble de test. L'unité de conductivité thermique est ($W \cdot m^{-1} \cdot K^{-1}$).

Pour améliorer la clarté et l'exhaustivité de ce chapitre, nous allons à présent concentrer notre analyse sur les deux modèles les plus performants : GC-XGBoost et GC-SVM. Ces derniers ont été évalués de manière approfondie à travers différentes combinaisons de familles cationiques et anioniques, offrant ainsi une vision complète de leurs capacités prédictives.

IV.4.1.A Ecarts relatifs selon les familles cationiques et anioniques

La Figure 58 présente, pour les deux modèles GC-XGBoost et GC-SVM, la distribution des écarts relatifs (ER) de la prédiction de la conductivité thermique λ en fonction des différentes familles cationiques et anioniques (les familles sont détaillées dans le Tableau 10). Pour chaque boîte à moustaches, la ligne médiane indique la médiane des ER, et les extrémités de la boîte correspondent respectivement aux premier et troisième quartiles. Les moustaches sont limitées à $\pm 5\%$ pour améliorer la lisibilité.

Les médianes des ER sont très proches de zéro pour la majorité des familles, ce qui témoigne de l'absence de biais systématique selon la nature ionique, tant pour XGBoost que pour SVM. Les boîtes (IQR) restent confinées dans $\pm 2\%$ pour la plupart des familles cationiques majeures (Imidazolium, Pyridinium, Pyrrolidinium) et anioniques (NTf₂, BF₄, PF₆), indiquant une bonne cohérence des prédictions.

Le modèle GC-SVM présente globalement des boîtes légèrement plus étroites que GC-XGBoost, reflétant une homogénéité accrue des erreurs au sein des familles. Pour les familles sous-représentées (p. ex. ammonium, aminoacides ([aa])), on observe des IQR plus larges et quelques valeurs extrêmes localisées en bord de boîte, liées au faible nombre de points de données disponibles.

Les familles très abondantes (e.g. Imidazolium : 597 points, NTf₂ : 290 points) affichent systématiquement des ER < 2 %, assurant la robustesse des prédictions sur ces catégories. Par ailleurs, la famille des aminoacides ([aa]) se distingue par une boîte à moustaches légèrement plus large reflétant une plus grande dispersion des écarts relatifs. Ce comportement suggère que la modélisation de la conductivité thermique pour les aminoacides mérite une attention particulière.

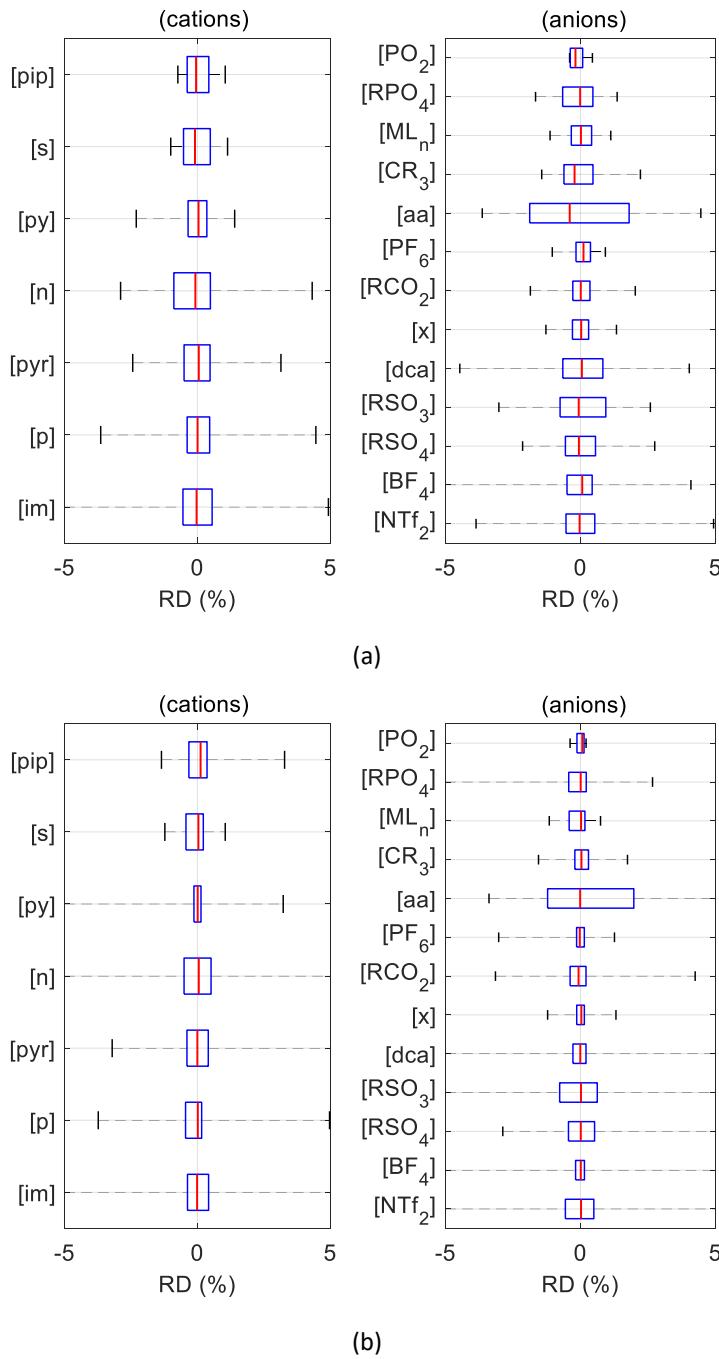


Figure 58: Distribution des écarts relatifs (ER) en fonction des familles cationiques et anioniques. La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieur et supérieur de la boîte représentent respectivement le premier et le troisième quartile. (a) Résultats du modèle GC-XGBoost. (b) Résultats du modèle GC-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 10.

IV.4.1.B Écarts relatifs selon les combinaisons cation-anion

La Figure 59 présente deux matrices de chaleur illustrant la distribution des écarts relatifs moyens (ER %) pour chacune des combinaisons de familles cationiques et anioniques,

respectivement pour le modèle GC-XGBoost (a) et le modèle GC-SVM (b). Chaque cellule de la matrice indique l'écart relatif moyen pour une paire cation-anion spécifique. Les valeurs moyennes par famille de cations et d'anions sont également fournies dans les lignes et colonnes intitulées "mean", offrant une vue d'ensemble de la performance globale des modèles.

Pour le modèle GC-XGBoost, les ER s'étendent approximativement de 0,2 % à 2.6 % avec plus de 70% des combinaisons présentent un $ER \leq 0.8\%$, tandis que pour GC-SVM la plage va de 0,1 % à 3 % avec plus de 70% des combinaisons présentent un $ER \leq 1.1\%$, témoignant d'une large majorité de prédictions très précises pour les deux modèles.

Pour les deux modèles, certaines paires ioniques montrent systématiquement des déviations plus élevées à savoir Phosphonium [p] – aminoacides [aa] atteint un ER moyen proche de 2.1 %, et Ammonium [n] – aminoacides [aa] présente également des valeurs extrêmes, jusqu'à 2.6 %. Ce comportement s'explique probablement par deux facteurs imbriqués : d'une part, il souligne que la prédiction de la conductivité thermique des aminoacides mérite une attention particulière, et d'autre part, il reflète les limites de la qualité et de la quantité des données expérimentales disponibles pour ces couples.

Les combinaisons les mieux pourvues en données (p.ex. imidazolium [im]-bis(trifluorométhylsulfonyl)imide NTf₂, plus de 126 points) montrent systématiquement des $ER = 1.1\%$ pour GC-XGBoost et 1.5% pour GC-SVM, confirmant la robustesse des prédictions.

L'analyse des combinaisons cation-anion met en évidence la nécessité d'un échantillonnage plus dense pour les paires Phosphonium [p] – aminoacides [aa], Ammonium [n] – aminoacides [aa].

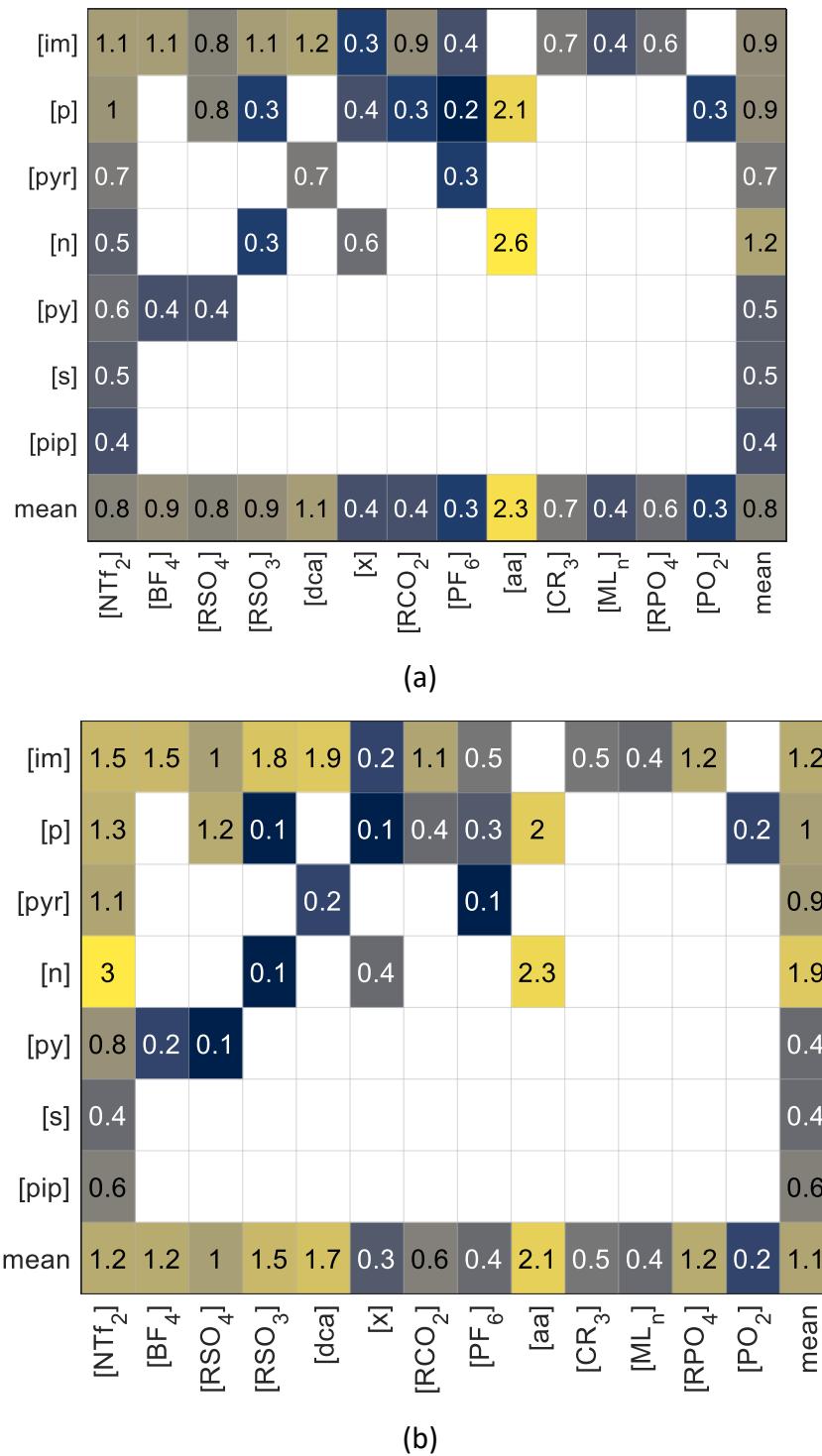


Figure 59: Distributions des écarts relatifs ER(%) en fonction des différentes familles de cations et d'anions. (a) modèle GC-XGBoost, (b) modèle GC-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 10.

IV.4.1.C Distributions d'écart relatifs moyens de prédiction

La Figure 60 illustre la distribution des points de données en fonction des plages d'écart relatifs moyens (ERM) pour les modèles GC-XGBoost (a) et GC-SVM (b). Chaque histogramme représente le nombre de points de données dont les ERM se situent dans des intervalles spécifiques, allant de 0 % à plus de 10 %, avec des paliers de 1 % pour les erreurs faibles et des regroupements pour les erreurs plus élevées. Cette représentation vise à évaluer la précision globale des modèles et à quantifier la proportion de points de données pour lesquels les prédictions sont particulièrement fiables.

Pour le modèle GC-XGBoost, les résultats montrent que 70,1 % des points de données (soit 719 points sur 1 025) présentent un ERM inférieur à 1 %, ce qui témoigne d'une précision exceptionnelle pour une part significative de la base de données. 89,2 % des points ont un ERM inférieur à 2 %, et 98,6 % un ERM inférieur à 5 %, démontrant une forte capacité à produire des prédictions précises dans la majorité des cas. Seuls 1,4 % des points affichent un ERM supérieur à 5 %, avec des erreurs maximales atteignant 13 % pour quelques cas isolés, souvent liés à des combinaisons cation-anion peu représentées dans les données.

Pour le modèle GC-SVM, la distribution des erreurs est encore plus concentrée ; 70,2 % des points de données ont un ERM inférieur à 1 %, dépassant légèrement les performances de XGBoost dans cette plage. 88,2 % des points présentent un ERM inférieur à 3 %, et 94,6 % un ERM inférieur à 5 %, indiquant une excellente maîtrise des prédictions sur une vaste majorité des données. Le pourcentage de points avec un ERM supérieur à 10 % est limité à 1,3 %, avec des erreurs maximales ne dépassant pas 21 %, ce qui reflète une robustesse accrue face aux cas complexes.

Ces distributions des écarts relatifs moyens pour GC-XGBoost et GC-SVM attestent d'une précision remarquable dans la prédiction de la conductivité thermique des liquides ioniques.

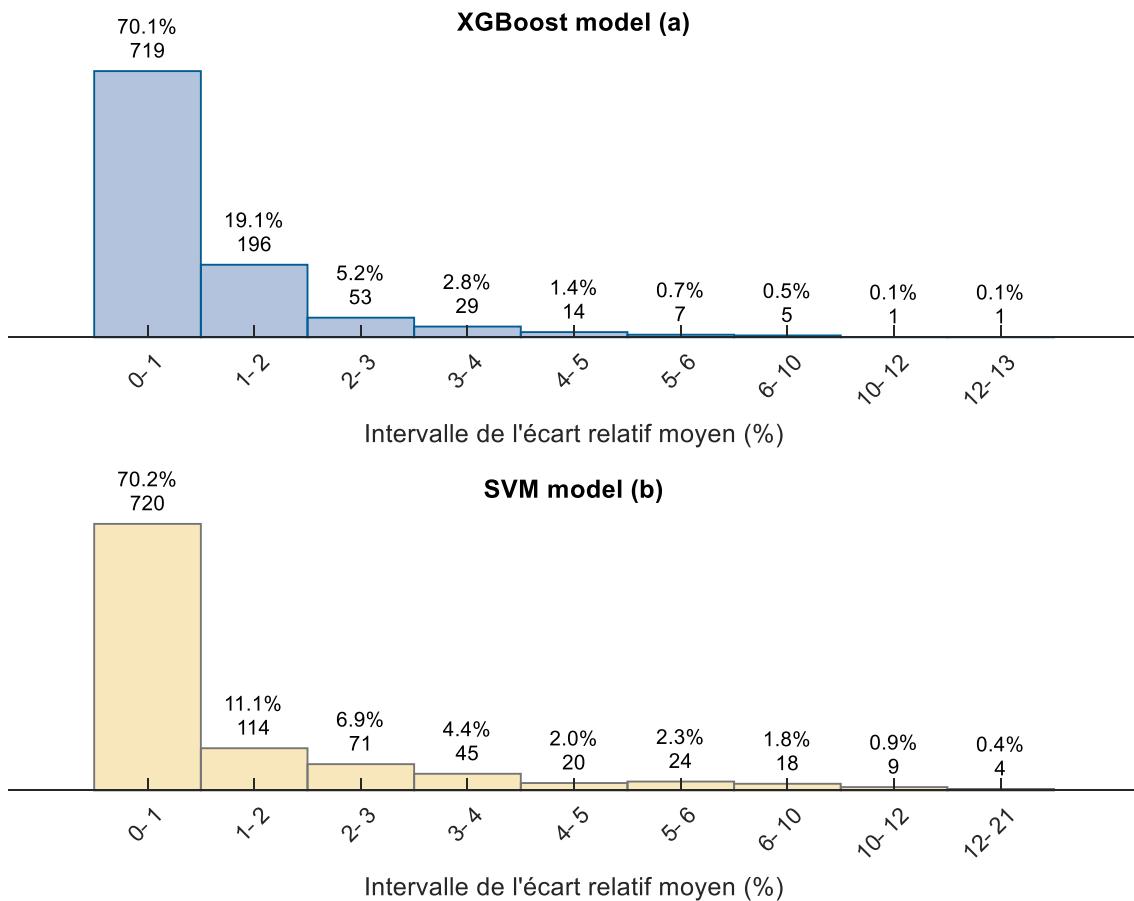


Figure 60: Distributions des points de données en fonction des erreurs relatifs moyens ER (%). (a) modèle GC-XGBoost, (b) modèle GC-SVM.

IV.4.2 Méthodes QSPR

Dans cette section, nous présentons les résultats obtenus avec les modèles basés sur l'approche QSPR. Chaque liquide ionique est décrit par un vecteur de 163 descripteurs moléculaires générés via la bibliothèque RDKit. Ces descripteurs incluent, entre autres, des indices de connectivité, des descripteurs de taille et de forme, des indicateurs de polarité, ainsi que des propriétés quantiques simplifiées, et fournissent une représentation multidimensionnelle riche pour les algorithmes d'apprentissage automatique.

Afin de garantir une comparaison équitable avec les méthodes par contributions de groupes, aucune procédure de sélection exhaustive (feature selection) n'a été appliquée a priori aux descripteurs QSPR. En effet, réduire le nombre de descripteurs pourrait éliminer des variables faiblement corrélées mais cruciales pour capturer les interactions non linéaires, particulièrement dans les modèles comme les XGBoost et les SVM, dont la performance repose sur la richesse des représentations.

Il est à noté que, lors du calcul des descripteurs QSPR avec RDKit, 7 liquides ioniques (88 points de données) n'ont pas pu être traités (structures non reconnues ou stéréochimie complexe) et ont dû être retirés. Le jeu de données est passé de 1 025 à 937 points de mesure, garantissant des descripteurs complets et fiables pour l'entraînement des modèles.

Les mêmes algorithmes que dans la section précédente ont été déployés : XGBoost, RF, SVM et DNN. Chaque modèle a été entraîné et évalué sur le jeu de données, en utilisant une validation croisée à 5 plis répétée 25 fois pour garantir la robustesse des résultats.

Le Tableau 13 synthétise les performances moyennes obtenues sur les jeux d'apprentissage et de test exprimées en ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$) pour la conductivité thermique :

Tableau 13: Statistiques de performance des QSPR-modèles sur les jeux d'apprentissage et de test, exprimées en moyennes issues d'une validation croisée à 5-plis, répétée 25 fois selon des partitions aléatoires.

Models	EAM_{train} ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	EAM_{test} ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	$ERAM_{train}$ (%)	$ERAM_{test}$ (%)	$REQM_{train}$	$REQM_{test}$	R^2_{train}	R^2_{test}
XGBoost	0.001	0.002	0.45	1.5	0.001	0.004	0.9986	0.9787
RF	0.002	0.003	1.41	2.01	0.003	0.005	0.9867	0.9713
SVM	0.001	0.002	0.92	1.52	0.003	0.005	0.9866	0.9728
DNN	0.002	0.002	1.05	1.64	0.003	0.005	0.9888	0.9633

On constate que les modèles XGBoost et SVM sont les plus performants en généralisation, avec des EAM_{test} de 0,002 ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$) et des $ERAM_{test}$ proches de 1,5 % chacune. Le DNN suit de près, tandis que la Forêt aléatoire reste légèrement derrière, avec une $ERAM_{test}$ de 2,01 %. Ce léger avantage de XGBoost et SVM illustre la capacité des algorithmes arborescents et à noyau à exploiter efficacement la richesse des descripteurs QSPR pour la prédiction de la conductivité thermique.

La Figure 61 met en évidence, pour les deux modèles QSPR retenus (XGBoost et SVM), à la fois la qualité d'ajustement aux données expérimentales et la distribution des erreurs de prédiction, tant sur l'ensemble d'apprentissage que sur l'ensemble de test. Les points de l'ensemble d'apprentissage s'alignent presque parfaitement sur la diagonale $y = x$, traduisant un ajustement quasi sans biais. Pour l'ensemble de test, on observe un léger élargissement du nuage. Cette dispersion légèrement plus importante en test se traduit

numériquement par un coefficient de détermination R_{test}^2 de 0,9787 pour XGBoost et 0,9728 pour SVM, contre respectivement 0,9986 et 0,9866 en apprentissage.

Les histogrammes d'erreur centrés sur zéro confirment l'absence de biais (moyenne $\simeq 0$). L'écart-type des erreurs en apprentissage est plus faible que celui en test, ce qui reflète une légère perte de précision hors de l'échantillon d'entraînement. Les deux distributions restent étroites et symétriques, indiquant que plus de 95 % des prédictions s'écartent de moins de $\pm 0,002$ ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$) des valeurs expérimentales. La forte linéarité et la faible dispersion observées signifient que les descripteurs QSPR et l'algorithme XGBoost capturent de manière particulièrement efficace les relations structure–propriété pour la conductivité thermique, avec un $ERAM_{test}$ de 1,5% contre 1,52% pour le SVM.

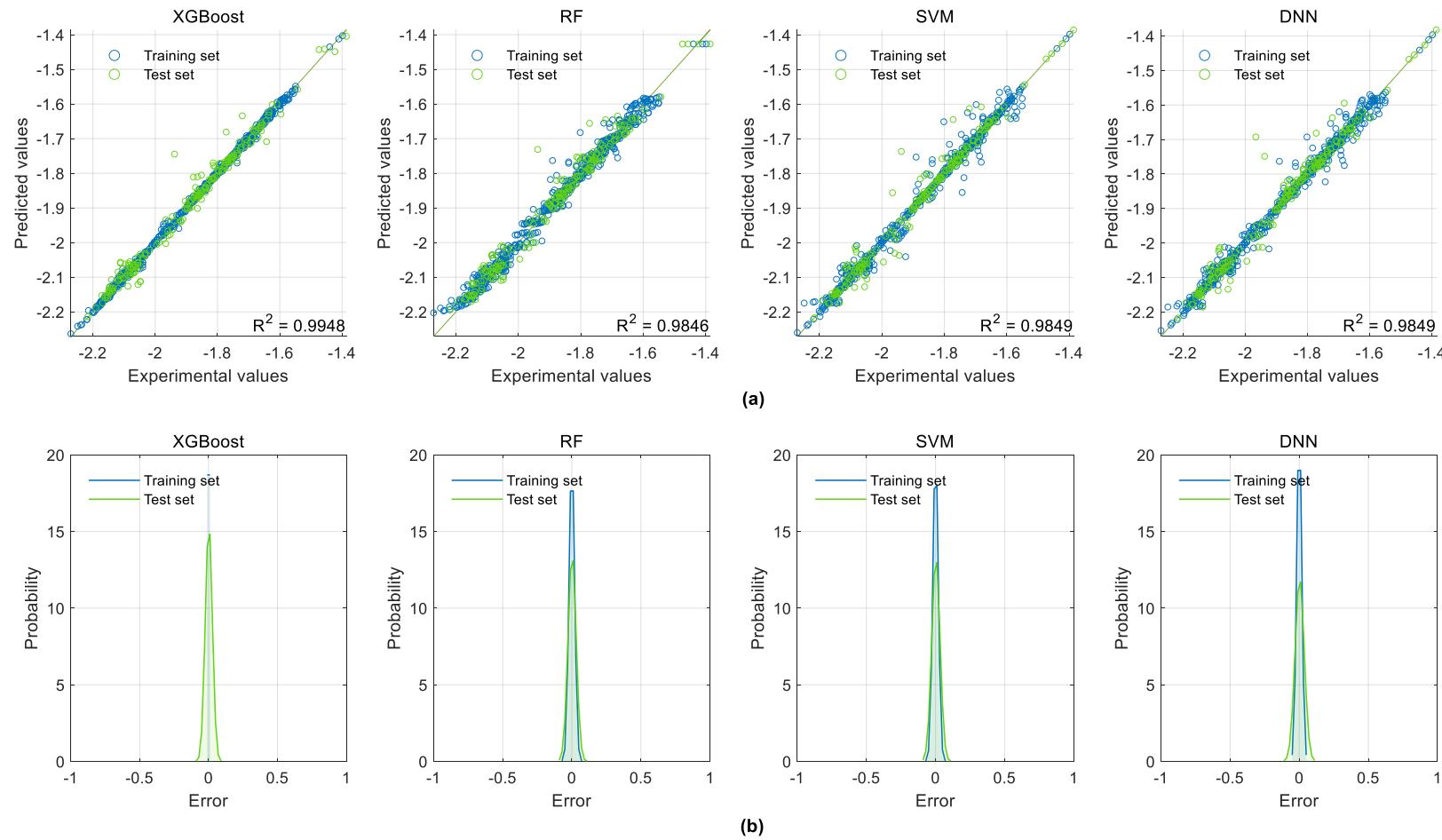


Figure 61: Comparaison des modèles QSPR sélectionnés pour prédire λ : (a) Nuages de points des valeurs prédictes et expérimentales dans l'ensemble d'apprentissage et l'ensemble de test. (b) Distribution gaussienne des erreurs de prédiction dans l'ensemble d'apprentissage et l'ensemble de test. L'unité de conductivité thermique est ($W \cdot m^{-1} \cdot K^{-1}$).

IV.4.2.A Ecarts relatifs selon les familles cationiques et anioniques

Dans cette sous-section, nous analysons la distribution des écarts relatifs (ER) des prédictions de conductivité thermique pour les modèles QSPR- XGBoost et QSPR-SVM, en fonction des familles chimiques des cations et des anions. Cette analyse vise à évaluer la performance des modèles au sein de chaque famille et à identifier les éventuelles disparités ou biais liés à la nature chimique des liquides ioniques.

La Figure 62 compare les écarts relatifs (ER%) de prédition de la conductivité thermique pour chaque famille de cations et d'anions, en distinguant les résultats obtenus par les modèles XGBoost et SVM.

Globalement, les médianes des ER pour les deux modèles sont très proches de zéro pour la plupart des familles, ce qui indique l'absence de biais systématique dans les prédictions, quel que soit le type de cation ou d'anion. Cela confirme que les modèles QSPR-XGBoost et QSPR- SVM parviennent à ajuster correctement la conductivité thermique sur l'ensemble des familles chimiques étudiées. Pour les boîtes sont plus étroites (IQR typiquement inférieur à 2 %) pour la plupart des familles.

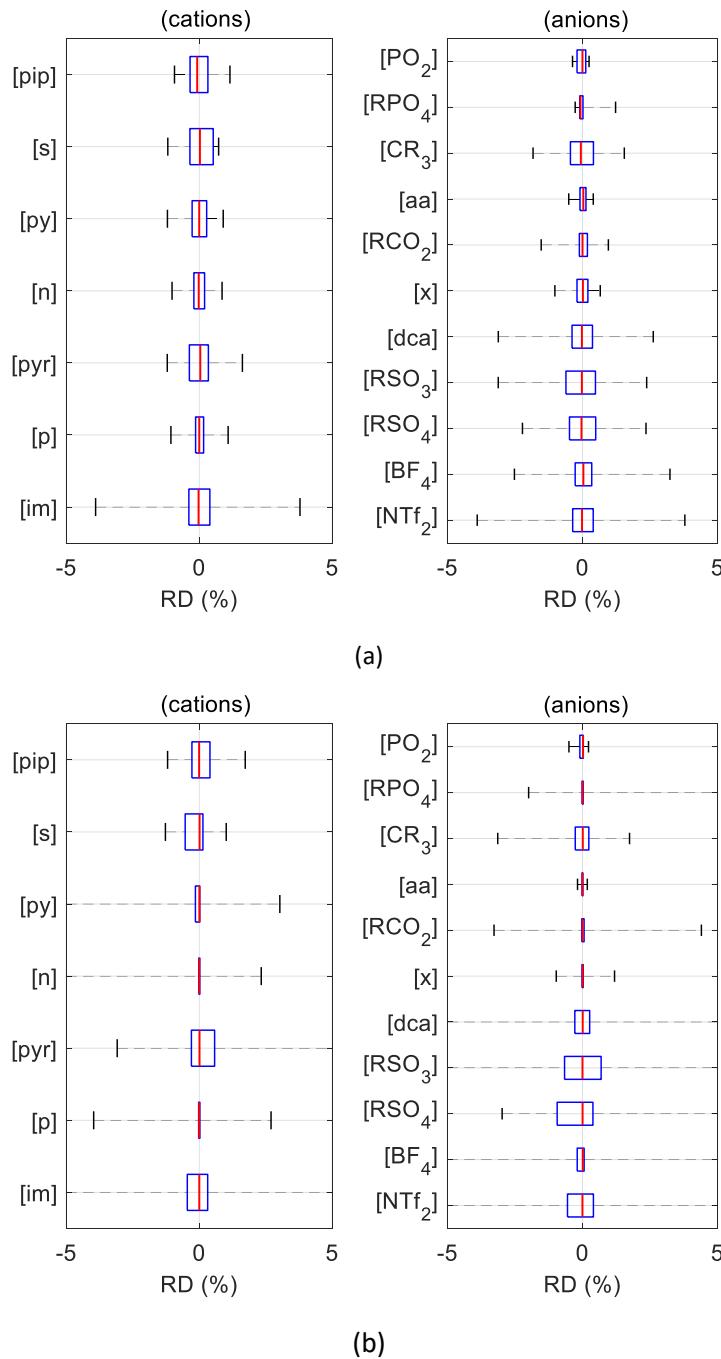


Figure 62: Distribution des écarts relatifs (ER) en fonction des familles cationiques et anioniques. La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieur et supérieur de la boîte représentent respectivement le premier et le troisième quartile. (a) Résultats du modèle QSPR-XGBoost. (b) Résultats du modèle QSPR-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 10.

IV.4.2.B Écarts relatifs selon les combinaisons cation-anion

Dans cette sous-section, nous examinons les écarts relatifs (ER) des prédictions de conductivité thermique obtenues à partir des modèles QSPR-XGBoost et QSPR-SVM, en

fonction des combinaisons spécifiques de cations et d'anions. Cette analyse permet d'identifier les forces et les limites des modèles en lien avec les propriétés chimiques des liquides ioniques étudiés.

La Figure 63a illustre ces écarts relatifs sous la forme d'une carte de chaleur (heatmap) pour le modèle QSPR-XGBoost. Les écarts relatifs varient de 0.1% à 0.9%, avec une majorité de valeurs concentrées entre 0.1% et 0.4%, ce qui reflète une performance globalement satisfaisante des modèles. Cependant, certaines combinaisons se distinguent par des écarts plus élevés mais qui reste très correcte. Par exemple, la paire [im]-[RSO₃] affiche un écart de 0.9%, le plus élevé observé, tandis que [im]-[NTf₂] atteint 0.8%, indiquant des prédictions qui reste correcte pour ces combinaisons. À l'opposé, les paires comportant des anions inorganiques [x], des carboxylates [RCO₂], des acides aminés [aa], des phosphates [RPO₄] et des phosphinates [PO₂] présentent des écarts de l'ordre de 0,2 à 0,3 %, témoignant d'une précision remarquable.

Pour le modèle QSPR-SVM, la Figure 63b montre des écarts relatifs variant de 0 % à 3 %, avec la plupart des valeurs comprises entre 0,2 % et 0,5 %. Cela témoigne d'une performance globalement satisfaisante du modèle. Toutefois, la combinaison des dérivés d'ammonium [n]-bistriflamide [NTf₂] se distingue avec une déviation ER de 3 %, soulignant une difficulté pour le modèle QSPR-SVM à prédire avec précision cette propriété spécifique.

Les cartes de chaleur de la Figure 63 démontrent que les modèles QSPR offrent des prédictions fiables pour la plupart des combinaisons cation-anion, avec des écarts relatifs généralement faibles.

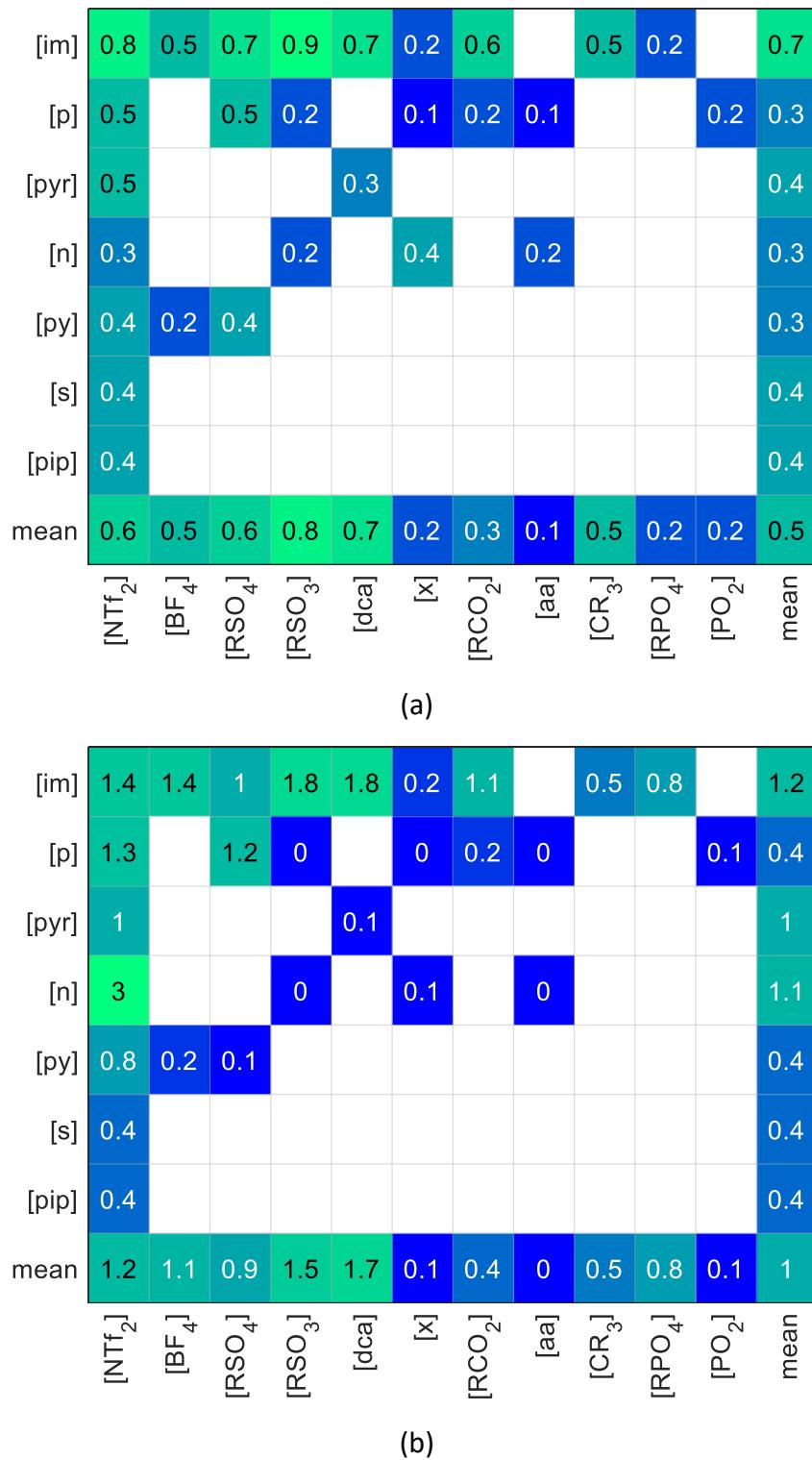


Figure 63: Distributions des écarts relatifs ER(%) en fonction des différentes familles de cations et d'anions. (a) modèle GC-XGBoost, (b) modèle GC-SVM. Les abréviations des familles chimiques sont détaillées dans le Tableau 10.

IV.4.2.C Distributions d'écart relatifs moyens de prédition

La Figure 64 présente la répartition des points de données selon leurs plages d'écart relatifs pour les modèles QSPR-XGBoost (a) et QSPR-SVM (b). Chaque histogramme illustre le nombre de points dont les écarts relatifs sont compris dans des intervalles définis.

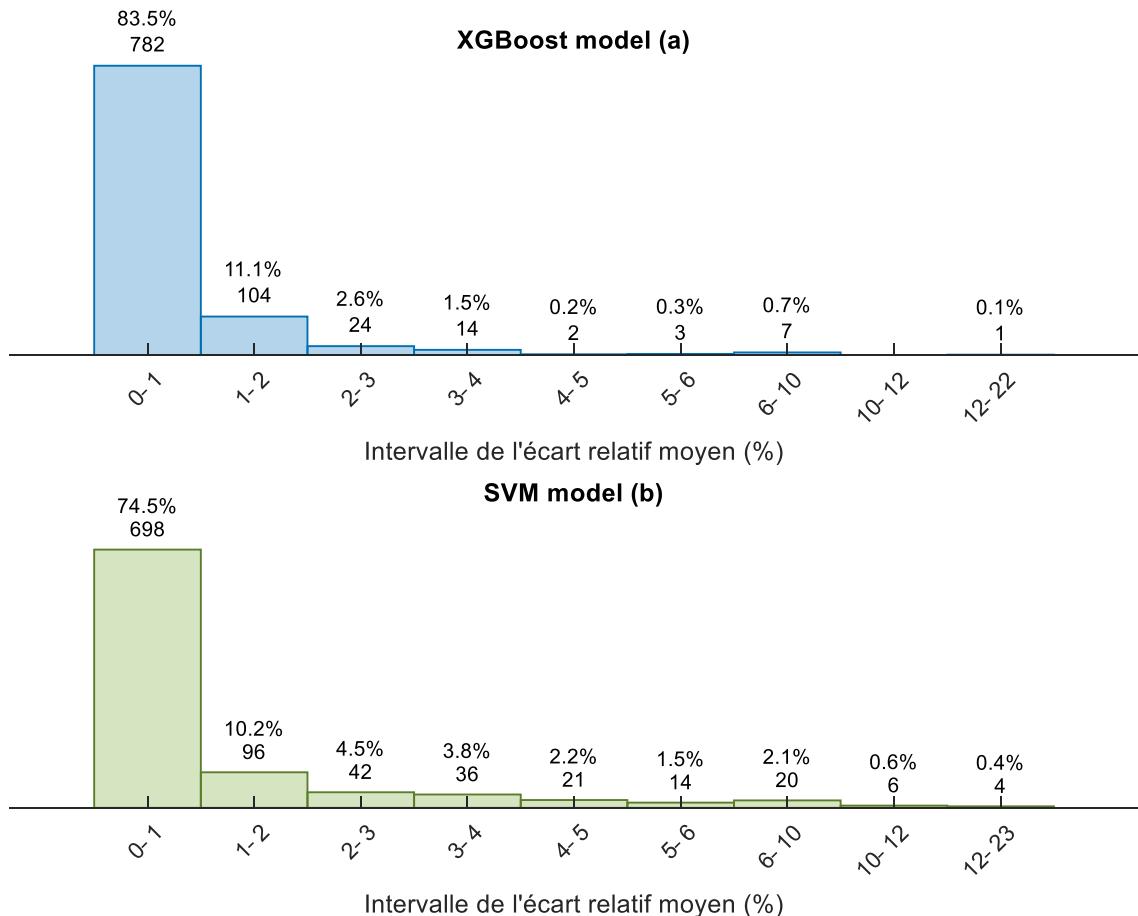


Figure 64: Distributions des points de données en fonction des erreurs relatifs moyens ER (%). (a) modèle QSPR-XGBoost, (b) modèle QSPR-SVM.

Le modèle XGBoost affiche une plus grande proportion de prédictions dans l'intervalle 0-1% (83,5%) par rapport au modèle SVM (74,5%). Cela indique que XGBoost est globalement plus précis, ses prédictions étant plus souvent très proches des valeurs réelles. Pour XGBoost, les fréquences diminuent rapidement après l'intervalle 1-2%, avec seulement 0,1% dans chaque intervalle au-delà de 7%. Cela montre que les erreurs importantes sont rares. Pour SVM, la distribution présente une queue plus longue, avec des pourcentages non négligeables dans les intervalles supérieurs (par exemple, 2,2%

dans 4-5%, 1,5% dans 5-6%, et 0,4% dans 12-23%). Cela suggère que SVM produit plus fréquemment des erreurs.

Cette figure met en évidence la supériorité du modèle XGBoost sur le modèle SVM en termes de précision et de consistance. Avec 83,5 % de ses prédictions dans l'intervalle 0-1 % contre 74,5 % pour SVM, et une queue de distribution beaucoup plus courte, XGBoost démontre une performance plus fiable et une meilleure maîtrise des erreurs. Bien que les deux modèles soient performants, XGBoost se distingue comme le choix optimal pour cette tâche spécifique.

IV.4.3 Comparaison des modèles proposés par rapport aux ceux existants

Dans cette section, nous comparons les performances de nos modèles développés (Ce travail) avec celles d'approches publiées dans la littérature, voir Tableau 14, en nous appuyant sur les indicateurs classiques que sont le nombre de points de données, le nombre d'ions liquides (ILs) couverts, la technique employée, Erreur absolue relative moyenne (ERAM), l'erreur moyenne absolue (EAM), l'erreur quadratique moyenne (EQM) et le coefficient de détermination (R^2). Nos quatre modèles s'appuient sur deux jeux de données de taille inédite ; 1 025 points pour les modèles GC-XGBoost et GC-SVM, et 937 points pour les modèles QSPR-XGBoost et QSPR-SVM, et englobent respectivement 79 et 73 ILs. À titre de comparaison, les études antérieures rapportent des jeux de données oscillant entre 107 et 606 points de mesure et ne couvrent que 11 à 50 ILs. Cette augmentation substantielle de la quantité et de la diversité des données constitue un atout majeur pour la robustesse et la généralisabilité de nos prédictions.

En termes de précision, les modèles basés sur l'algorithme XGBoost se distinguent nettement. Le modèle QSPR-XGBoost atteint une erreur ERAM de seulement 1,52 %, une EAM de 0,002 et une REQM de 0,004, avec un coefficient de R^2 de 0,9787. Ces performances surpassent largement le meilleur résultat antérieur, obtenu par Ren Wan et al. (2022) avec 2,18 % d'ERAM et $R^2 = 0,9713$, ou encore le QSPR linéaire de Lazzús (2015) affichant 1,97 % d'ERAM pour $R^2 = 0,9919$. Notre second meilleur modèle, GC-XGBoost (ERAM = 0,82 %, $R^2 = 0,9788$), demeure plus performant que la plupart des approches antérieures. Même le modèle de Juan A. Lazzús (2016), avec un ERAM de

0,84 % et un R² de 0,9963, bien qu'efficace, repose sur un ensemble plus limité (400 points, 41 ILs) et n'égale pas la précision des modèles actuels.

Les modèles GC-SVM et QSPR-SVM présentent quant à eux des résultats légèrement moins précis que leurs homologues XGBoost, avec respectivement 1,60 % et 1,52 d'ERAM, tout en conservant d'excellents scores de R² (0,9784 et 0,9728). Même le modèle de Juan A. Lazzús (2016), avec un ERAM de 0,84 % et un R² de 0,9963, bien qu'efficace, repose sur un ensemble plus limité (400 points, 41 ILs) et n'égale pas la précision des modèles actuels. Cela confirme que, bien que les SVM soient compétitives, les techniques de XGBoost s'avèrent plus efficaces pour capturer les relations non linéaires complexes inhérentes aux descripteurs moléculaires de type graphe ou QSPR.

Ces améliorations découlent de l'utilisation de techniques avancées d'apprentissage automatique, notamment XGBoost, SVM et DNN, associées à des entrées de modèle de type GC et QSPR. Ces approches permettent de modéliser des relations complexes dans les données, surpassant les méthodes concourantes. Ces résultats mettent en lumière le potentiel des techniques modernes d'apprentissage automatique pour résoudre des défis en chimie et en science des matériaux. Par ailleurs, l'emploi d'un ensemble de données plus large et diversifié améliore la généralisation des modèles, les rendant particulièrement adaptés à des applications pratiques dans l'étude des liquides ioniques.

Tableau 14: Comparaison des performances des modèles proposés par rapport aux modèles existants.

Authors	Year	number of data points	number of ILs	method/technique	ERAM (%)	EAM ($\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)	REQM	R^2
Ce travail	2025	1025	79	GC-XGBoost	1,62	0.003	0.004	0.9788
Ce travail	2025	1025	79	GC-SVM	1.60	0.003	0.004	0.9784
Ce travail	2025	937	73	QSPR-XGBoost	1.50	0.002	0.004	0.9787
Ce travail	2025	937	73	QSPR-SVM	1,52	0.002	0.005	0.9728
Ren Wan et al[8]	2022	606	44	(COSMO-SAC)-MLR	2.18		0.0043	0.9713
S.P.Mousavi et al [9]	2020	504	50	RBF-GSA		1.051		
Wu et al. [10]	2012	286	36	GC	1.66	2.322		
Gardas-Cotinho [11]	2009	107	16		13.4	4.021		
Lazzús [12]	2015	195	11	QSPR	1.97	1.480		0.9919
Albert[13]	2014	375	39	GC	7.96	1.613		0.81
Albert[14]	2014	372	39		2.5	3.281		
Shojaee[15]	2013	209	21		7.3	6.911		
Reza Haghbakhsh,[4]	2017	378	44		1 .0			
Saeid Atashrouz[16]	2016	216	22	CSA-LSSVM			0.0020	0.994
Juan A. Lazzús[17]	2016	400	41	ANN-GA	0.84			0.9963

IV.5 CONCLUSION

Ce chapitre a exploré la prédiction de la conductivité thermique des liquides ioniques en s'appuyant sur une approche méthodique combinant collecte rigoureuse de données, analyse exploratoire et modélisation avancée. À partir d'une base de données initiale de 1 092 points de mesure couvrant 88 liquides ioniques, réduite à 1 025 points pour 79 LI après révision, l'étude a permis de mettre en lumière les effets des conditions thermodynamiques sur cette propriété. Les analyses ont révélé que la conductivité thermique diminue légèrement avec l'augmentation de la température, en raison de l'agitation thermique, tandis qu'elle augmente sous l'effet de la pression, grâce à la densification du fluide, avec des variations modulées par la structure chimique des cations et anions.

Pour prédire cette propriété, cinq stratégies de modélisation ont été déployées : les contributions de groupes (CG) et les descripteurs QSPR, associés à des algorithmes d'apprentissage automatique tels que XGBoost, SVM, RF et DNN, ainsi qu'un modèle semi-empirique conventionnel. Les résultats ont démontré la supériorité des modèles XGBoost et SVM, notamment avec l'approche QSPR, qui ont atteint des performances exceptionnelles (ERAM de 1,5 % pour QSPR-XGBoost et R^2 de 0,9787). Ces modèles surpassent les approches existantes dans la littérature, grâce à leur capacité à capturer les relations complexes entre la structure moléculaire et la conductivité thermique, et à l'utilisation d'un ensemble de données plus large et diversifié (937 à 1 025 points contre 107 à 606 dans les études précédentes).

En comparaison avec les travaux antérieurs, cette étude se distingue par une précision accrue et une meilleure généralisabilité, ouvrant ainsi de nouvelles perspectives pour la conception et l'optimisation des liquides ioniques dans des applications thermiques, telles que les caloporteurs ou les fluides de refroidissement. À l'avenir, élargir les données expérimentales concernant les familles d'anions peu couvertes (dicyanamides, complexes métalliques, sulfates et phosphates) et intégrer des variables structurales émergentes (quantum-descripteurs plus fins, nouveaux groupes fonctionnels) permettra d'améliorer encore la précision et la portée de ces modèles.

IV.6 RÉFÉRENCE

- [1] C. Wang, X. Wei, X. Jin, J. Li, M. He, Developing a two-grade model for the thermal conductivity of ionic liquids and their mixtures, *Chem. Eng. Sci.* 290 (2024) 119881. <https://doi.org/10.1016/j.ces.2024.119881>.
- [2] A.P. Fröba, M.H. Rausch, K. Krzeminski, D. Assenbaum, P. Wasserscheid, A. Leipertz, Thermal conductivity of ionic liquids: Measurement and prediction, *Int. J. Thermophys.* 31 (2010) 2059–2077. <https://doi.org/10.1007/s10765-010-0889-3>.
- [3] A.Z. Zhao, J.E. Garay, High temperature liquid thermal conductivity: A review of measurement techniques, theoretical understanding, and energy applications, *Prog. Mater. Sci.* 139 (2023) 101180. <https://doi.org/10.1016/j.pmatsci.2023.101180>.
- [4] R. Haghbakhsh, S. Raeissi, A novel correlative approach for ionic liquid thermal conductivities, *J. Mol. Liq.* 236 (2017) 214–219. <https://doi.org/10.1016/j.molliq.2017.03.117>.
- [5] L.H. Soares, R. Guirardello, M.P. Rolemburg, A simple group contribution model to predict thermal conductivity of pure ionic liquids, *Chem. Eng. Trans.* 74 (2019) 1195–1200. <https://doi.org/10.3303/CET1974200>.
- [6] R.M. Almeida, M.J. V. Lourenço, C. Nieto de Castro, The thermal conductivity of ionic liquids. Experiment and molecular interpretation, *J. Mol. Liq.* 397 (2024). <https://doi.org/10.1016/j.molliq.2024.124123>.
- [7] H. Yang, R.C. Gallagher, A.T. Phan, P. Chartrand, A.E. Gheribi, A predictive approach for the compositional and temperature representation of thermal conductivity in multicomponent molten salt systems for advanced energy applications, *Mater. Today Energy*. 38 (2023) 101441. <https://doi.org/10.1016/j.mtener.2023.101441>.
- [8] R. Wan, M. Li, F. Song, Y. Xiao, F. Zeng, C. Peng, H. Liu, Predicting the Thermal Conductivity of Ionic Liquids Using a Quantitative Structure–Property Relationship, *Ind. Eng. Chem. Res.* 61 (2022) 12032–12039. <https://doi.org/10.1021/acs.iecr.2c01964>.
- [9] S.P. Mousavi, S. Atashrouz, F. Rezaei, M.E. Peyvastegan, A. Hemmati-Sarapardeh, A. Mohaddespour, Modeling thermal conductivity of ionic liquids: A comparison between chemical structure and thermodynamic properties-based models, *J. Mol. Liq.* 322 (2021) 114911. <https://doi.org/10.1016/j.molliq.2020.114911>.
- [10] K.J. Wu, C.X. Zhao, C.H. He, Development of a group contribution method for determination of thermal conductivity of ionic liquids, *Fluid Phase Equilib.* 339 (2013) 10–14. <https://doi.org/10.1016/j.fluid.2012.11.024>.
- [11] R.L.G. and J.A.P. Coutinho, PVAm–PIP/PS composite membrane with high performance for CO₂/N₂ separation, *AIChE J.* 55 (2009) 1274–1290. <https://doi.org/10.1002/aic>.
- [12] J.A. Lazzús, G. Pulgar-Villarroel, Estimation of thermal conductivity of ionic liquids using quantitative structure-property relationship calculations, *J. Mol. Liq.* 211 (2015) 981–985. <https://doi.org/10.1016/j.molliq.2015.08.037>.
- [13] J. Albert, K. Müller, Thermal conductivity of Ionic Liquids: An estimation approach, *Chem. Eng. Sci.* 119 (2014) 109–113. <https://doi.org/10.1016/j.ces.2014.08.023>.
- [14] J. Albert, K. Müller, A group contribution method for the thermal properties of ionic liquids, *Ind. Eng. Chem. Res.* 53 (2014) 17522–17526. <https://doi.org/10.1021/ie503366p>.

- [15] S.A. Shojaee, S. Farzam, A.Z. Hezave, M. Lashkarbolooki, S. Ayatollahi, A new correlation for estimating thermal conductivity of pure ionic liquids, *Fluid Phase Equilib.* 354 (2013) 199–206. <https://doi.org/10.1016/j.fluid.2013.06.004>.
- [16] S. Atashrouz, A. Hemmati-Sarapardeh, H. Mirshekar, B. Nasernejad, M. Keshavarz Moraveji, On the evaluation of thermal conductivity of ionic liquids: Modeling and data assessment, *J. Mol. Liq.* 224 (2016) 648–656. <https://doi.org/10.1016/j.molliq.2016.09.106>.
- [17] J.A. Lazzús, Estimation of the thermal conductivity $\lambda(T, P)$ of ionic liquids using a neural network optimized with genetic algorithms, *Comptes Rendus Chim.* 19 (2016) 333–341. <https://doi.org/10.1016/j.crci.2015.09.010>.

CHAPITRE V : PREDICTION DE LA VISCOSITE DES LIQUIDES IONIQUES

CHAPITRE V: PREDICTION DE LA VISCOSITE DES LIQUIDES IONIQUES

V.1 INTRODUCTION

Le Chapitre 5 se consacre à la prédiction de la viscosité des liquides ioniques à base d'imidazolium, une propriété physico-chimique clé qui joue un rôle déterminant dans de nombreuses applications industrielles et scientifiques. Inscrit dans une démarche globale visant à élaborer des modèles prédictifs précis et fiables pour ces composés, ce chapitre explore des approches novatrices basées sur l'apprentissage automatique et l'analyse de la structure moléculaire. Plus précisément, il examine deux méthodes complémentaires : les Contributions de Groupes (GC) et les relations quantitatives structure-propriété (QSPR), appliquées à une vaste base de données expérimentales comprenant 7015 points pour 620 liquides ioniques à base d'imidazolium, reflétant une diversité de structures anioniques.

Ce chapitre propose une analyse approfondie de cinq approches de modélisation avancées : Deep Neural Network (DNN), la Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost) et Random Forest (RF) et un modèle semi-empirique conventionnel, en évaluant leurs performances respectives.

V.2 BASE DE DONNÉES EXPÉRIMENTALES

L'essentiel des données viscosimétriques provient des références validées par Paduszyński [1], couvrant les publications de 1984 à 2018. Cette sélection, exclusivement centrée sur les sels d'imidazolium combinés à différentes familles anioniques, rassemble 8038 mesures expérimentales pour 676 liquides ioniques (LI) uniques à pression

atmosphérique, intégralement compilées dans un fichier Excel. L'extension de cette base, toujours dédiée aux systèmes imidazolium/anions variés, a été réalisée via ILThermo (NIST-ILs) [2] avec des données récentes (2018 à ce jour), soit 905 points supplémentaires pour 81 LI, accompagnés de leurs incertitudes.

Cette base de données actualisée a été analysée et révisée selon la procédure suivante : La première étape a consisté à exclure les phases cristallines et les liquides métastables. La deuxième étape visait à supprimer les entrées en double au sein de chaque ensemble de données. La troisième étape a impliqué d'identifier les mesures incohérentes, c'est-à-dire plusieurs valeurs pour un même liquide à des températures identiques, et de ne conserver que la moyenne de ces valeurs. La quatrième étape a éliminé les points de données présentant une incertitude relative supérieure à 100 %, tandis que la cinquième étape a écarté les liquides pour lesquels il restait moins de trois points de données. Lors de la sixième étape, les données restantes pour chaque LI à pression atmosphérique ont été ajustées à l'aide de l'équation de Vogel–Fulcher–Tammann. Tout point de données présentant un résidu studentisé dépassant deux fois l'écart interquartile (IQR) alors été signalé comme valeur aberrante, et ce cycle d'ajustement-élimination des valeurs aberrantes a été répété jusqu'à ce qu'aucune autre valeur aberrante ne soit détectée. Dans l'étape finale de la procédure, l'ensemble de données a été examiné manuellement afin de supprimer toute entrée potentiellement inappropriée.

Ce processus a généré une base exhaustive de 7015 points expérimentaux pour 620 LI (Figure 65), catégorisés en 15 familles anioniques (Figure 66). La Figure 65(a) quantifie le nombre des liquides, tandis que la Figure 65(b) en détaille les occurrences des points de mesure. Les abréviations des familles chimiques sont détaillées dans le Tableau 15. Ce tableau résume, pour les différentes familles anioniques des liquides ioniques à base d'imidazolium, le nombre de points de données et de jeux de données collectés, ainsi que la diversité moléculaire et les plages de température et de viscosité couvertes. Comme on peut le constater, l'ensemble de données inclut une grande variété d'anions, les plus nombreux étant NTf_2 , PF_6^- , BF_4^- , les sulfonates, les sulfates et les anions inorganiques. Parmi les anions, les bistriflamides (NTf_2) et l'hexafluorophosphate (PF_6^-) sont les plus représentés dans la base de données, couvrant respectivement une fraction de 35 % et 16 % de l'ensemble des liquides ioniques. Les températures mesurées s'étendent généralement d'environ 253 K à 573 K, et les viscosités couvrent plusieurs ordres de

grandeur (de moins de $0.841 \text{ mPa}\cdot\text{s}$ à plus de $2.3\times 10^5 \text{ mPa}\cdot\text{s}$), ce qui met en lumière la variabilité rhéologique extrême des liquides ioniques en fonction de leur structure ionique. Cette abondance de données justifie le recours à des méthodes statistiques et de modélisation avancées pour en extraire des tendances fiables.

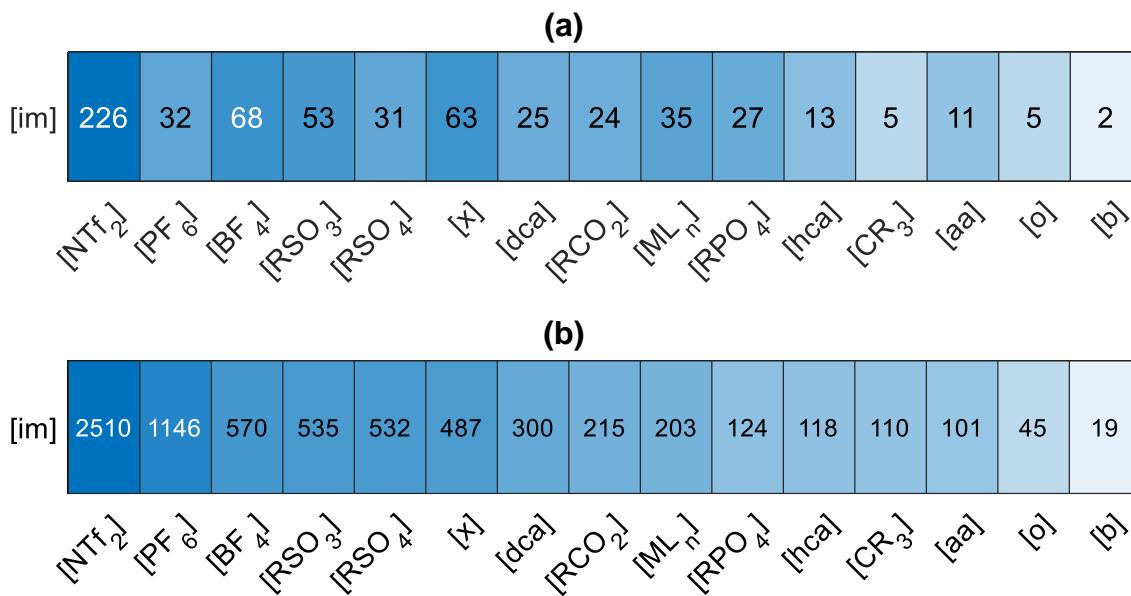


Figure 65. (a) Nombre de liquides ioniques à base d'imidazolium selon les différentes familles d'anions. (b) Nombre de points de données expérimentales selon les différentes familles d'anions. Les abréviations des familles chimiques sont détaillées dans le Tableau 15.

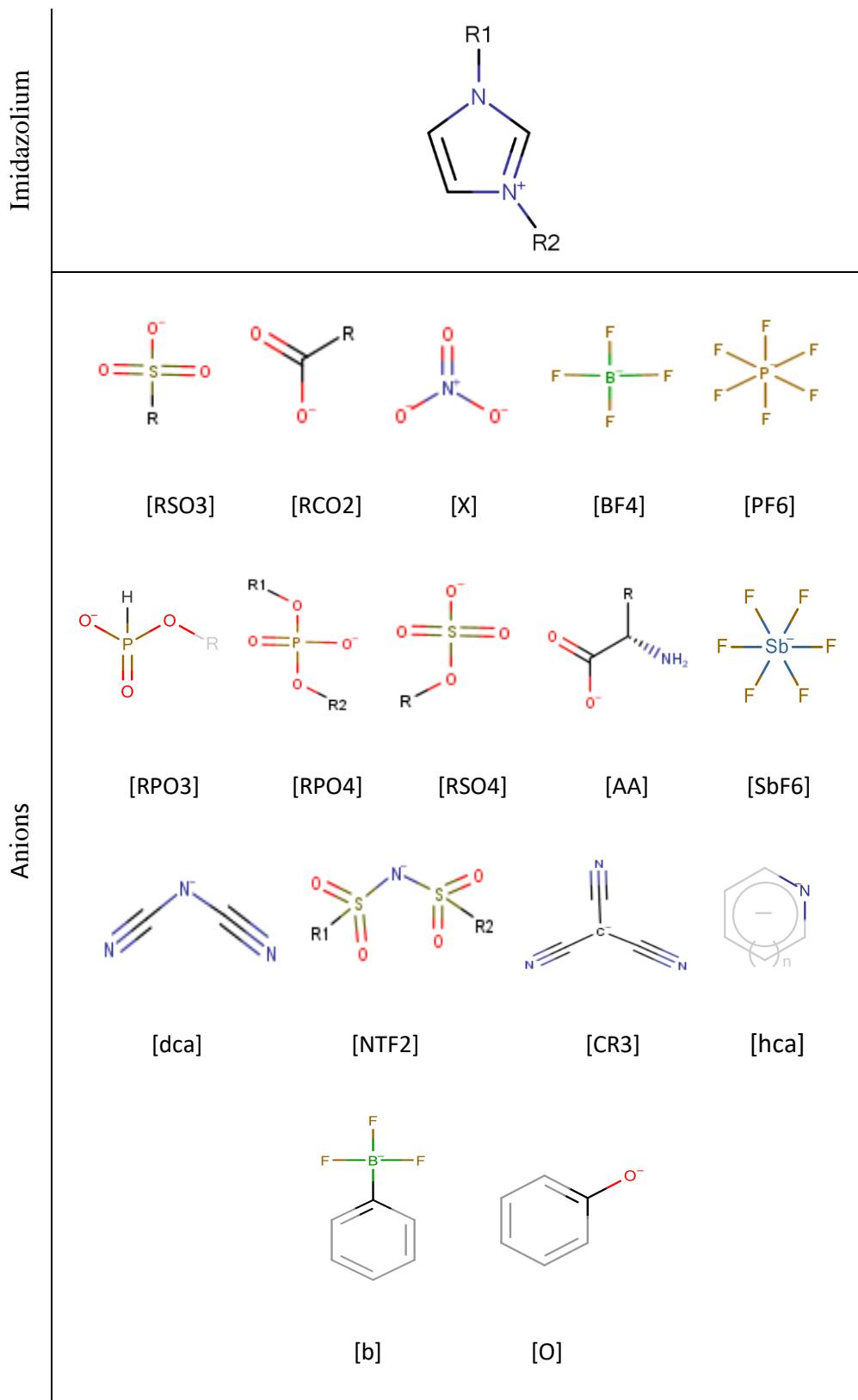


Figure 66 : Structures chimiques des familles anioniques et imidazolium. Les abréviations des familles chimiques sont détaillées dans le Tableau 15.

Tableau 15: Résumé de la base de données de viscosité par familles d'ions

IL family	Data points	Data sets	Molecules	ILs	T min (K)	T max (K)	min (mPa.s)	max (mPa.s)
NTf ₂ derivatives [NTf ₂]	2510	283	203	226	258.15	573	0.841	9053.1
PF ₆ derivatives [PF ₆]	1146	50	23	32	262.57	413.85	4.603	10508
BF ₄ derivatives [BF ₄]	570	91	26	68	273.15	388.04	3.289	1036.98
sulfonates [RSO ₃]	535	62	29	53	263.15	373.15	8.24703	19212
sulfates [RSO ₄]	532	56	16	31	253.15	395.32	7.2	16145
inorganics [x]	487	71	43	63	278.15	393.15	5.4846	19734.1
dicyanamides [dca]	300	36	25	25	258.15	372	3.64	370
carboxylates [RCO ₂]	215	34	14	24	263.15	408.15	4.589	12300
metal complexes [ML _n]	203	35	14	35	263.1	368.1	4.18455	690.88
phosphates [RPO ₄]	124	30	20	27	273.15	373.15	9.289	23508
heterocyclic amines [hca]	118	13	7	13	278.15	343.15	8.6	14600
methanides [CR ₃]	110	12	2	5	278.15	363.15	4.6415	480
aminoacids [aa]	101	11	3	11	293	373.15	13.3	2853
alcoholates [o]	45	5	5	5	293.15	373.15	6.012	350.9
organic borates [b]	19	2	2	2	263.15	353.15	10.6525	2412.3
Totalité de la base de données	7015	791	258	620	253.15	573	0.841	23508

V.3 MÉTHODOLOGIQUE POUR LA MODÉLISATION

Ce travail a employé cinq approches de modélisation distinctes pour prédire la viscosité des liquides ioniques : *Extreme Gradient Boosting* (XGBoost), *Random Forest* (RF), *Support Vector Machine* (SVM), *Deep Neural Network* (DNN) et un modèle semi-empirique conventionnel. Pour plus de détaillé voir la section II.3, *Modèles et Régression*).

Les quatre approches d'apprentissage automatique (XGBoost, RF, SVM, DNN) ont combiné deux types d'entrées complémentaires :

1. Contributions de Groupes (GC) : Les molécules ont été fragmentées en **223** groupes prédefinis (**77 cationiques, 83 anioniques, 63 substituants**), catégorisés en contributions du premier ordre et du second ordre.
2. Descripteurs QSPR : Un vecteur de 163 descripteurs moléculaires calculés à l'aide de RDKit. Ces entrées structurales ont été complétées par la variable thermodynamique température (T).

La détermination de la forme fonctionnelle optimale du modèle semi-empirique conventionnel a fait l'objet d'une démarche systématique. En prenant pour point de départ l'équation VTF (Vogel-Tamman-Fulcher), plusieurs formulations non linéaires décrivant la dépendance $\eta(T)$ ont été évaluées sur la totalité des données. Le choix final s'est porté sur une équation alliant robustesse corrélative et pouvoir d'extrapolation, dont l'architecture emprunte aux fonctions sigmoïdes.

$$\ln(\eta) = \sum_{i=1}^3 b_i \frac{\sum_{j=1}^K N_j a_{i,j} + \alpha_i T}{1 + |\sum_{j=1}^K N_j a_{i,j} + \alpha_i T|} + c \quad (60)$$

Dans l'équation (X), b_i et c sont des paramètres ajustables, A_i représente la contribution de tous les groupes dans le i -ième terme de cette équation, incluant à la fois les groupes du premier ordre et du second ordre présents dans les noyaux cationiques, anioniques et substituants. N_j désigne le nombre d'occurrences du $j^{ième}$ groupe individuel, $a_{i,j}$ est la contribution du $j^{ième}$ groupe dans le $i^{ième}$ terme de cette équation. K est le nombre total de groupes fonctionnels et T est la température en kelvins (K). α_i sont des paramètres ajustables dans le i -ième terme de cette équation.

À noter que le modèle conventionnel est exclusivement appliqué à l'approche par contributions de groupes car la fragmentation moléculaire permet d'assigner directement les coefficients de la loi semi-empirique à des groupes chimiques spécifiques. En revanche, l'approche QSPR (s'appuyant sur un vecteur dense de descripteurs moléculaires globaux) ne se prête pas à une interprétation physique directe.

Les modèles QSPR et GC ont été construits sous MATLAB (v. 2023b) via sa Statistics and Machine Learning Toolbox (licence académique). Les paramètres par défaut de la bibliothèque (cf. section II.5.3) ont été conservés pour la plupart des hyperparamètres. Seuls ceux ayant un impact significatif dans les contextes QSPR et GC ont été optimisés. Pour ce réglage, une optimisation bayésienne [36] a été utilisée pour proposer itérativement les configurations les plus prometteuses. La qualité des modèles a été évaluée via 25 répétitions de validation croisée 5-folds, en minimisant l'erreur quadratique moyenne.

L'exploration des espaces de recherche et la sélection finale des hyperparamètres, déterminées par optimisation bayésienne, sont synthétisées comme suit (voir section II.5.3). Pour XGBoost, l'espace de recherche couvrait MinLeafSize¹⁰ (0,1–100), NumLearningCycles¹¹ (5–1000) et LearnRate¹² (1×10^{-4} –1), conduisant à une configuration optimale de MinLeafSize=1, NumLearningCycles=133 et LearnRate=0,327. L'optimisation de Random Forest (RF) a exploré MinLeafSize (1–100) et NumLearningCycles (10–1000), retenant finalement MinLeafSize=1 avec NumLearningCycles=499. Le réglage des paramètres du SVM s'est concentré sur C (1–400), ϵ (0,0001–1) et γ (0,1–100), aboutissant à C=46, $\epsilon=1,3056\times10^{-4}$ et $\gamma=6,55$. Enfin, pour le réseau de neurones profond (DNN), la recherche incluait les couches cachées (1–5), neurones par couche (1–60), fonctions d'activation (ReLU/sigmoïde/tangente hyperbolique) et le terme de régularisation Lambda (0–10), convergeant vers une architecture à 2 couches (20 et 5 neurones), une activation tanh et Lambda= $0,517\times10^{-10}$.

¹⁰ MinLeafSize : Taille minimale des feuilles terminales.

¹¹ NumLearningCycles : Nombre d'arbres (estimateurs) dans le modèle.

¹² LearnRate : Taux d'apprentissage.

V.4 RÉSULTATS ET DISCUSSION

Cette section présente et analyse les performances prédictives des modèles développés, évaluées à l'aide des métriques statistiques décrites dans (chapitre II.6). Les résultats sont organisés selon les deux approches de modélisation :

1. Les Contributions de Groupes, où les fragments moléculaires servent de variables d'entrée directes ;
2. L'approche QSPR basée sur des descripteurs moléculaires calculés.

Une classification complémentaire par familles de liquides ioniques et par anion a été mise en œuvre. Cette structuration des résultats guide les utilisateurs vers l'application optimale des modèles pour des systèmes ioniques spécifiques.

V.4.1 Méthodes de Contributions de Groupes

Dans cette section, nous analysons les performances des modèles de Contributions de Groupes (GC) pour la prédiction de la viscosité des liquides ioniques à base d'imidazolium. Cinq modèles ont été évalués : XGBoost, RF, SVM, DNN et un modèle conventionnel. Les résultats, issus d'une validation croisée 5-plis répétée 25 fois, sont présentés dans le Tableau 16 et illustrés par la Figure 67.

Tableau 16: Performances des modèles GC : moyennes issues d'une validation croisée 5 folds × 25 répétitions (ensembles d'entraînement et de test)

Models	EAM_{train}	EAM_{test}	$ERAM_{train}$ (%)	$ERAM_{test}$ (%)	$REQM_{train}$	$REQM_{test}$	R^2_{train}	R^2_{test}
XGBoost	14.08	39.97	4.32	8.12	83.85	273.0	0.9947	0.9511
RF	88.32	100.2	11.9	18.7	632.9	628.9	0.8534	0.8514
SVM	23.41	31.05	3.78	7.96	217.5	220.9	0.9712	0.9660
DNN	4.95	14.31	2.24	5.76	26.32	110.3	0.995	0.9900
Conventionnel	129.3	126.3	26.7	27.4	861.0	789.2	0.6588	0.7214

Le Tableau 16 rapporte les métriques de performance pour chaque modèle GC, incluant l'erreur absolue moyenne (EAM), l'erreur absolue relative moyenne (ERAM), racine d'erreur quadratique moyenne (REQM) et le coefficient de corrélation (R^2), calculés pour les ensembles d'entraînement et de test. Le modèle DNN se distingue par sa performance optimale, affichant la plus faible erreur $EAM_{test} = 14.31$ (en test) et la plus basse erreur $ERAM_{test} = 5.76\%$, signes d'une haute précision prédictive. Sa robustesse est confirmée

par l'écart minimal entre les scores R^2 d'entraînement (0.995) et de test (0.990), indiquant un surapprentissage maîtrisé. Le SVM démontre également une compétitivité notable avec des métriques équilibrées ($EAM_{test} = 31.05$, $ERAM_{test} = 7.96\%$, $R^2_{test} = 0.966$) et un écart train-test modéré, bien que légèrement inférieur au DNN.

En revanche, XGBoost révèle des faiblesses en généralisation malgré un excellent R^2_{train} (0.9947), subissant une chute marquée en test ($R^2_{test} = 0.9511$) et une augmentation significative de l'EAM (de 14.08 à 39.97), trahissant un surapprentissage prononcé. Le modèle Random Forest (RF) et le modèle conventionnel présentent des limitations critiques : le RF montre une EAM_{test} élevée (100.2) et un R^2_{test} faible (0.8514), tandis que le modèle conventionnel enregistre les performances les plus médiocres ($EAM_{test} = 126.3$, $ERAM_{test} = 27.4\%$, $R^2_{test} = 0.7214$), confirmant la supériorité systématique des approches par apprentissage automatique.

La hiérarchie des modèles (DNN > SVM > XGBoost > RF > Conventionnel) souligne l'avantage des architectures profondes et des modèles à noyau pour cette tâche. Le surapprentissage observé chez XGBoost plaide pour un renforcement des mécanismes de régularisation dans les méthodes arborescentes. Enfin, le DNN émergeant comme la méthodologie optimale pour une modélisation non-linéaire exigeantes, bien que le SVM constitue une alternative viable pour des implémentations moins complexes.

La Figure 67 illustre deux aspects clés : (a) la corrélation entre les valeurs prédictives, par les modèles de GC, et expérimentales, et (b) la distribution des résidus, pour les ensembles d'entraînement et de test. Concernant la corrélation (Figure 67a), les graphiques montrent une forte corrélation pour le DNN et le SVM, avec des points alignés près de la diagonale, indiquant des prédictions précises. XGBoost présente une corrélation correcte, mais avec plus de dispersion sur l'ensemble de test. Le RF et le modèle conventionnel montrent une dispersion significative, reflétant leur moindre précision. Pour la distribution des résidus (Figure 3b), les résidus du DNN et du SVM sont centrés autour de zéro avec une faible variance, ce qui est idéal. XGBoost affiche une distribution légèrement plus large, tandis que le RF et le modèle conventionnel ont des distributions étalées, confirmant leurs erreurs plus élevées. Ces observations visuelles corroborent les résultats du Tableau 2, soulignant la supériorité du GC-DNN et du GC-SVM.

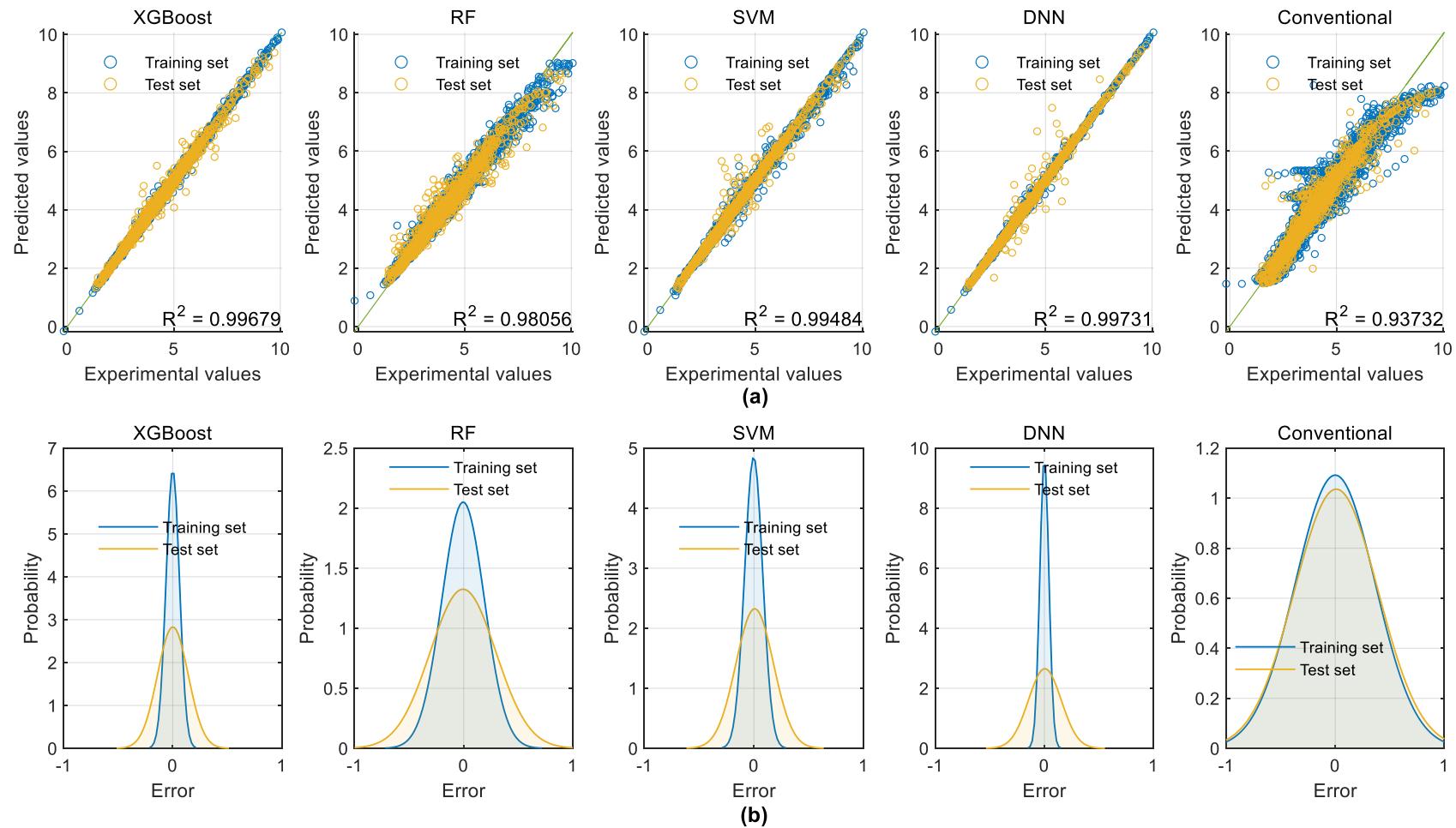


Figure 67. Benchmark des modèles GC: (a) Corrélation prédit/expérimental (train/test), (b) Distribution normale des résidus (train/test)

V.4.1.A Ecarts relatifs selon les familles anioniques

La Figure 68 présente la distribution des erreurs relatives (ER%) pour les trois modèles de Contributions de Groupes (GC) les **plus performants**, appliqués à différentes familles anioniques. Les boîtes à moustaches illustrent la médiane, les premier et troisième quartiles, ainsi que les valeurs aberrantes pour chaque famille, offrant une vue d'ensemble de la précision prédictive des modèles en fonction des structures anioniques des LI.

Les modèles, notamment le DNN, le SVM et XGBoost, montrent des performances variables selon les familles anioniques. Le DNN surpasse généralement les autres modèles, affichant des erreurs médianes plus faibles et une meilleure consistance across les familles anioniques. Le SVM reste compétitif pour les familles bien représentées comme [NTf2], et rencontre des limitations spécifiques face aux amines hétérocycliques [hca] et aux borates organiques [b], où sa capacité prédictive chute significativement, soulignant la sensibilité des modèles aux particularités structurales de ces groupes minoritaires.

Pour des familles comme les [NTf2] (bistriflamides) et les [PF6] (hexafluorophosphates), les erreurs médianes sont faibles (souvent inférieures à 5 %), avec des distributions resserrées, témoignant d'une excellente capacité prédictive. Ces familles, bien représentées dans la base de données, bénéficient probablement d'un apprentissage robuste grâce à la disponibilité de données et à la régularité de leurs structures. À l'inverse, certaines familles, comme les sulfates [RSO₄] et les inorganiques [x], présentent des écarts interquartiles larges pour les trois meilleurs modèles. Ces points dispersés pourraient refléter des liquides ioniques aux caractéristiques inhabituelles (par exemple, conditions expérimentales atypiques).

Ces résultats de la Figure 4 révèlent que la précision des modèles de GC dépend fortement de la famille anionique, avec une efficacité marquée pour les anions courants ([NTf2], [PF6]). Le DNN se positionne comme le modèle le plus performant, soulignant l'importance d'une approche basée sur des réseaux profonds pour capturer les relations structure-propriété dans les liquides ioniques.

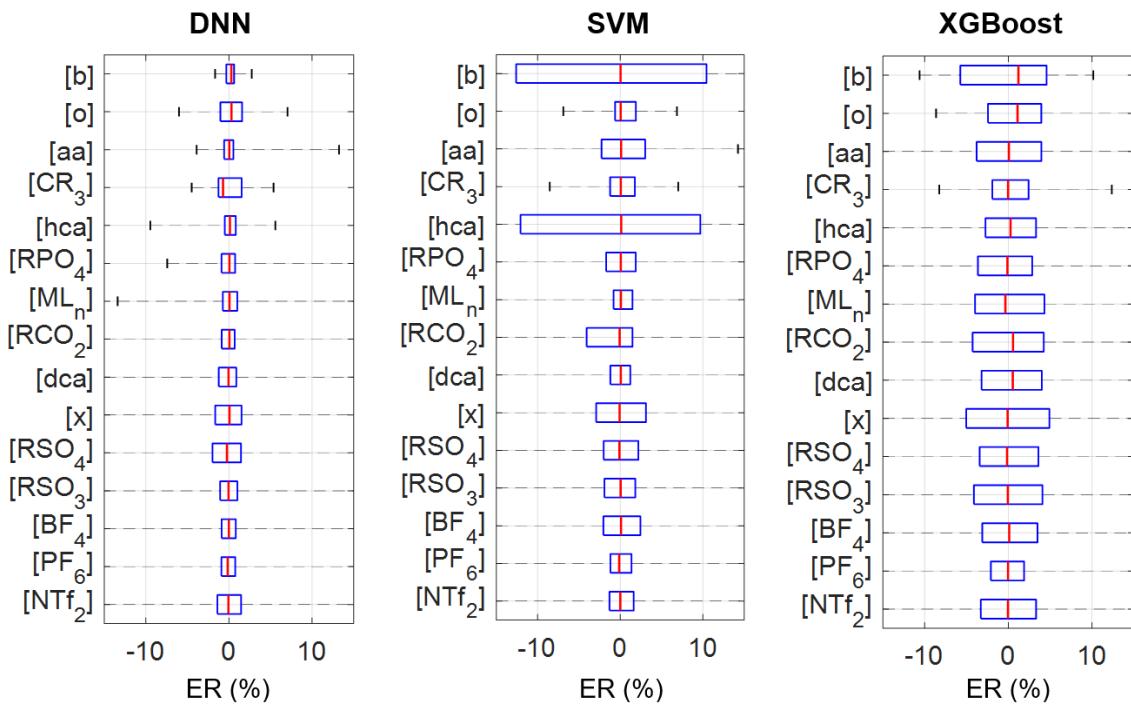


Figure 68. : Distribution des erreurs relatifs (ER%) par familles anioniques pour les 3 modèles GC les plus performants (représentation en boîtes à moustaches). La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieurs et supérieurs de la boîte représentent respectivement le premier et le troisième quartile.

V.4.1.B Distributions d'écart relatifs moyens de prédiction

Les modèles DNN, SVM et XGBoost ont été évalués en termes de distribution de l'erreur relative moyenne (ER) de leurs prédictions, comme illustré dans la Figure 69. L'analyse de ces distributions révèle des performances variables entre les modèles, en particulier concernant leur capacité à prédire avec une faible erreur.

Le modèle DNN (Figure 69a) présente la meilleure performance globale en termes de précision, avec 66,7% des prédictions ayant une ER comprise entre 0 et 2%. Cela représente 4680 cas, ce qui est significativement plus élevé que les autres modèles dans cette même tranche d'erreur. De plus, 20,6% des prédictions (1442 instances) se situent dans la fourchette d'erreur de 2-5%. Cela indique que le modèle DNN est capable de fournir des prédictions très précises pour la grande majorité des cas. À mesure que l'intervalle d'erreur augmente, le pourcentage de prédictions diminue considérablement, avec seulement 8,6% pour 5-10%, et des pourcentages marginaux pour les erreurs supérieures à 10%.

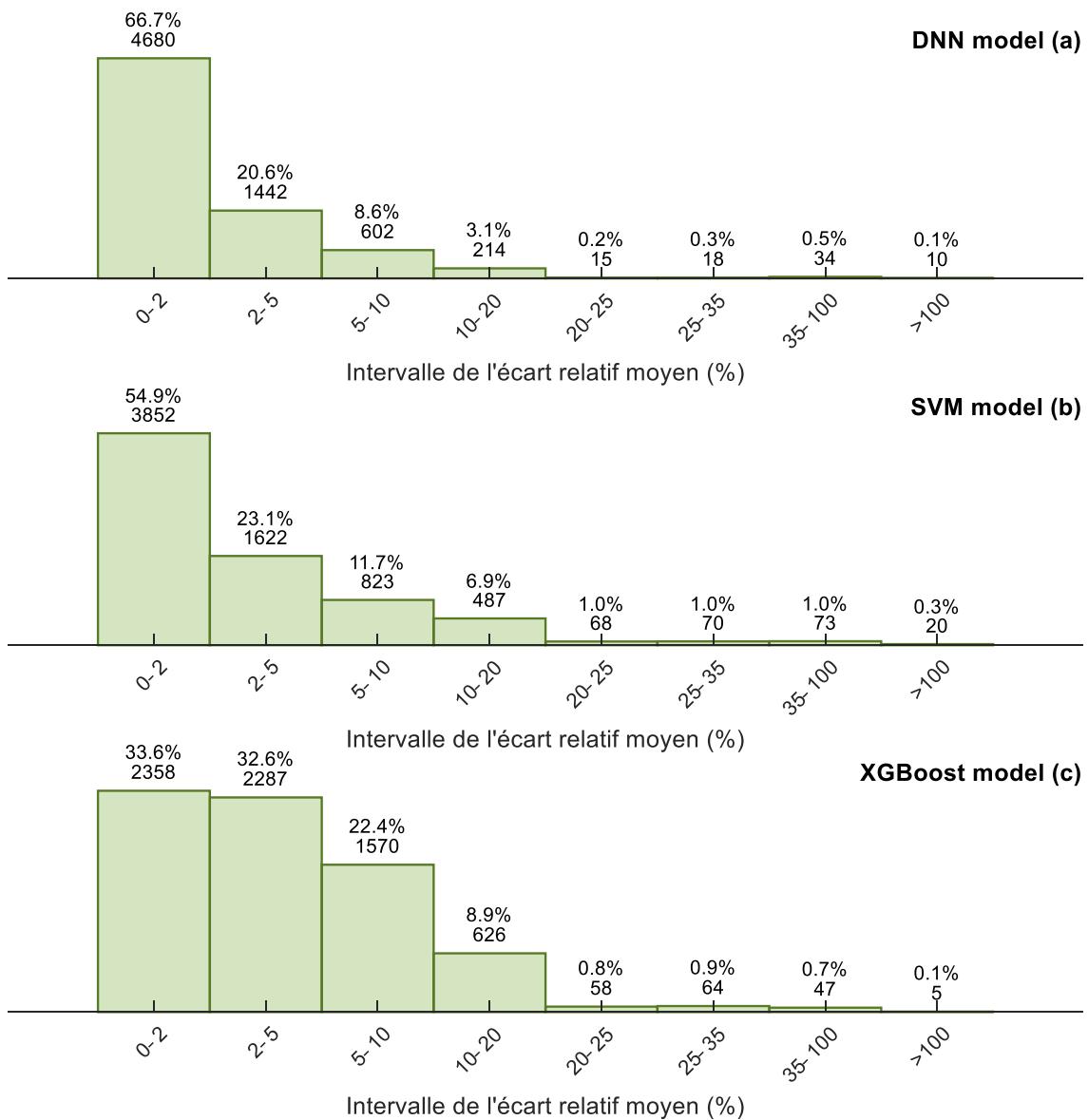


Figure 69. : Distribution de l'erreur relative moyenne (ER, %) des modèles de GC : (a) DNN, (b) SVM, (c) XGBoost.

En comparaison, le modèle SVM (Figure 69b) montre une performance légèrement inférieure à celle du DNN pour les erreurs très faibles. Il obtient 54,9% de prédictions avec une ER entre 0 et 2% (3852 cas), ce qui est respectable mais reste moins élevé que le DNN. Cependant, il présente une proportion légèrement plus élevée de prédictions dans l'intervalle d'erreur de 2-5% (23,1%, soit 1622 cas) par rapport au DNN. Pour les intervalles d'erreur plus grands (5-10% et 10-20%), le SVM maintient des pourcentages de prédictions plus élevés que le DNN, suggérant une distribution de l'erreur un peu plus étalée.

Le modèle XGBoost (Figure 69c) se distingue par une distribution d'erreur plus large par rapport aux deux autres modèles. Il n'obtient que 33,6% des prédictions avec une ER entre 0 et 2% (2358 cas), ce qui est nettement inférieur aux modèles DNN et SVM. Cependant, il compense en partie avec une proportion importante de prédictions dans l'intervalle 2-5% (32,6%, soit 2287 cas) et 5-10% (22,4%, soit 1570 cas). Cela indique que le modèle XGBoost, bien que moins performant pour les erreurs très faibles, conserve une performance décente pour des erreurs moyennes. Ces résultats soulignent l'importance de choisir le modèle approprié en fonction de la tolérance à l'erreur et des exigences spécifiques de l'application.

V.4.2 Méthodes QSPR

Les résultats des méthodes QSPR, présentés dans le Tableau 3 et la Figure 6, fournissent une évaluation complète des performances des modèles XGBoost, RF, SVM et DNN pour la prédiction de la viscosité. L'analyse porte sur les corrélations entre valeurs prédictes et expérimentales, ainsi que sur la distribution des erreurs résiduelles.

Tableau 17: Performances des modèles QSPR : moyennes issues d'une validation croisée 5 folds × 25 répétitions (ensembles d'entraînement et de test)

Modèle QSPR	EAM_{train}	EAM_{test}	$ERAM_{train}$ (%)	$ERAM_{test}$ (%)	$REQM_{train}$	$REQM_{test}$	R^2_{train}	R^2_{test}
XGBoost	10.94	39.31	2.79	9.28	139.68	244.19	0.9873	0.9485
RF	73.65	79.43	9.68	15.88	562.92	584.98	0.8948	0.8097
SVM	18.97	25.29	2.64	7.48	164.81	199.71	0.9858	0.9650
DNN	4.85	16.02	1.99	5.38	30.12	142.2	0.9994	0.9828

Le Tableau 17 met en évidence la supériorité du modèle DNN. Avec un R^2_{test} de 0.9828 et un R^2_{train} de 0.9994, le DNN démontre une excellente capacité à capturer la variance des données, tant sur l'ensemble d'entraînement que de test. Ses erreurs moyennes absolues (EAM_{test} de 16.02) et relatives ($ERAM_{test}$ de 5.38%) sont les plus faibles parmi tous les modèles, indiquant une grande précision des prédictions. Le $REQM_{test}$ de 142.2 confirme également cette performance. La Figure 70a corrobore ces observations, montrant une corrélation quasi parfaite entre les valeurs prédictes et expérimentales pour le DNN, avec des points de données très proches de la ligne de régression $y = x$. De même, la Figure 70b illustre une distribution des résidus pour le DNN très centrée autour de zéro, avec une faible dispersion, ce qui est caractéristique d'un modèle bien ajusté.

Le modèle SVM arrive en deuxième position en termes de performance. Il affiche un R^2_{test} de 0.9650 et un R^2_{train} de 0.9858, indiquant une bonne capacité prédictive. Ses valeurs d'erreur (EAM_{test} de 25.29, $ERAM_{test}$ de 7.48%, et $REQM_{test}$ de 199.71) sont supérieures à celles du DNN, mais demeurent compétitives. Les graphiques de la Figure 70a pour le SVM montrent également une forte corrélation, bien que les points puissent être légèrement plus dispersés que pour le DNN. La distribution des résidus (Figure 70b) est également bien centrée, mais avec une dispersion légèrement plus large que celle du DNN, ce qui correspond à des erreurs légèrement plus élevées.

Le modèle XGBoost présente des performances respectables, avec un R^2_{test} de 0.9485 et un R^2_{train} de 0.9873. Ses erreurs (EAM_{test} de 39.31, $ERAM_{test}$ de 9.28%, et $REQM_{test}$ de 244.19) sont plus élevées que celles du DNN et du SVM, mais demeurent acceptables pour de nombreuses applications. La Figure 70a pour XGBoost révèle une bonne corrélation entre les valeurs prédites et expérimentales, bien qu'avec une dispersion plus prononcée des points par rapport aux deux meilleurs modèles. La distribution des résidus (Figure 70b) montre une dispersion plus large, suggérant des erreurs plus fréquentes et potentiellement plus importantes.

En revanche, le modèle RF est le moins performant parmi les quatre modèles évalués. Son R^2_{test} de 0.8097 et son R^2_{train} de 0.8948 sont significativement inférieurs aux autres modèles. Les valeurs d'erreur sont également les plus élevées (EAM_{test} de 79.43, $ERAM_{test}$ de 15.88%, et $REQM_{test}$ de 584.98). La Figure 70a pour RF illustre une dispersion notable des points autour de la ligne de régression, ce qui est cohérent avec un R^2 plus faible. La distribution des résidus (Figure 70b) est la plus large, indiquant une plus grande variabilité et magnitude des erreurs de prédiction.

En comparant les modèles QSPR proposés, le modèle DNN se révèle être le choix optimal pour les prédictions QSPR dans cette étude, offrant la plus haute précision et la plus faible erreur. Le SVM est une alternative solide, tandis que le XGBoost offre des performances acceptables. Le Random Forest, bien qu'utile dans d'autres contextes, semble moins adapté pour cette tâche QSPR spécifique, du moins avec les configurations et données utilisées.

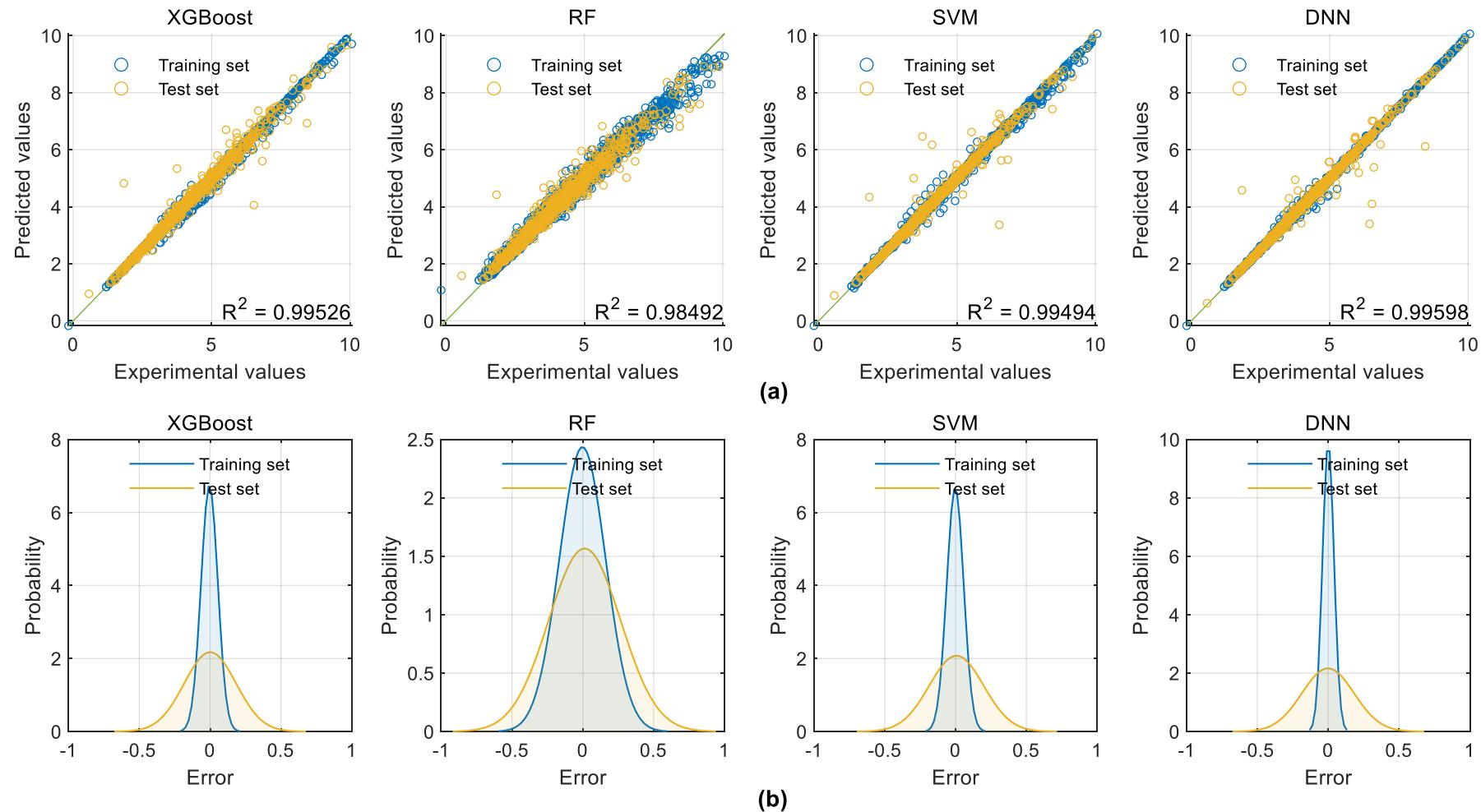


Figure 70. : Benchmark des modèles QSPR: (a) Corrélation prédit/experimental (train/test), (b) Distribution normale des résidus (train/test)

V.4.2.A Ecarts relatifs selon les familles anioniques

L'analyse de la Figure 71, qui présente la distribution des erreurs relatives (ER) par familles anioniques pour les trois modèles QSPR les plus performants (DNN, SVM et XGBoost), fournit des informations cruciales sur la robustesse et la spécificité de chaque modèle face à différentes chimies anioniques. La représentation en boîtes à moustaches permet de visualiser la médiane, l'étendue interquartile et la dispersion globale des erreurs.

Le modèle DNN (colonne de gauche) démontre une performance remarquablement stable et précise sur l'ensemble des familles anioniques. Pour la plupart des anions, les boîtes à moustaches sont très courtes et centrées autour de 0% d'ER. Cela indique une faible dispersion des erreurs et une médiane proche de zéro, signifiant que le DNN prédit avec une grande exactitude pour ces familles. Les exceptions sont minimes, avec des boîtes très légèrement plus larges pour la famille hexafluorophosphates [PF₆], mais l'ER reste généralement dans une plage étroite, typiquement entre -3% et 3%. Cela confirme la robustesse et la généralisabilité du modèle DNN à travers diverses structures anioniques.

Le modèle SVM (colonne centrale) présente une performance globalement bonne, mais avec une variabilité légèrement plus prononcée que le DNN pour certaines familles anioniques. Si de nombreuses familles affichent des boîtes à moustaches étroites et centrées autour de 0%, on observe des dispersions plus importantes pour des anions comme amines hétérocycliques [hca] ou hexafluorophosphates [PF₆]. Pour ces familles, la plage d'ER est plus étendue, indiquant une précision légèrement moindre ou une plus grande variabilité des prédictions. Cependant, la médiane reste généralement proche de zéro pour la majorité des familles, ce qui suggère que les erreurs ne sont pas systématiquement biaisées.

Le modèle XGBoost (colonne de droite) montre la plus grande variabilité dans la distribution des erreurs relatives selon les familles anioniques. Bien que certaines familles soient prédites avec une bonne précision (boîtes étroites), d'autres, comme phosphates [RPO₄] ou hexafluorophosphates [PF₆], présentent des boîtes à moustaches considérablement plus larges et parfois décalées par rapport à zéro. Par exemple, pour [RPO₄], la boîte est très étendue et sa médiane semble très négative, suggérant une sous-prédition pour cette famille. Cela indique que le XGBoost, bien que performant

globalement, peut-être plus sensible à la spécificité de certaines structures anioniques, entraînant des erreurs plus importantes ou plus dispersées pour ces groupes.

Le modèle DNN se révèle être le plus cohérent et le plus précis sur l'ensemble des familles anioniques, avec des erreurs relatives minimales et une faible dispersion. Le SVM maintient un bon niveau de performance, mais avec quelques variations de précision entre les familles. Le XGBoost, bien que puissant, montre des fluctuations plus marquées dans la distribution de ses erreurs relatives en fonction des familles anioniques, soulignant l'importance d'une évaluation spécifique par groupe chimique pour ce type de modèle.

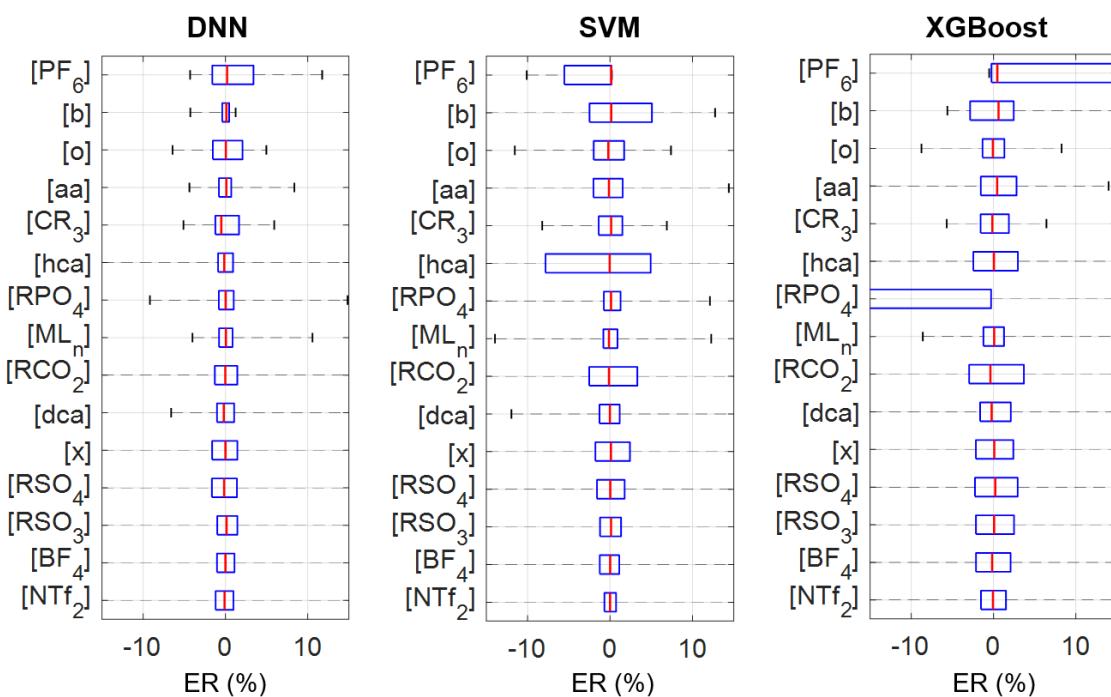


Figure 71. : Distribution des erreurs relatifs par familles anioniques pour les 3 modèles QSPR les plus performants (représentation en boîtes à moustaches). La marque centrale, correspondant à la médiane, donne une indication de la tendance centrale des données. Les bords inférieur et supérieur de la boîte représentent respectivement le premier et le troisième quartile.

V.4.2.B Distributions d'écart relatif moyen de prédition

Les performances des modèles QSPR basés sur DNN, SVM et XGBoost ont été comparées au travers de la distribution de l'erreur relative moyenne (ER) de leurs prédictions (Figure 72). L'analyse met en lumière des profils distincts selon les algorithmes et leur aptitude à produire des prédictions à faible erreur.

Le modèle DNN (Figure 72a) se distingue par la proportion la plus élevée de prédictions avec une très faible erreur. En effet, 69,5% de ses prédictions (4030 cas) se situent dans l'intervalle d'ER de 0 à 2%. Cela démontre une excellente capacité du DNN à prédire avec une grande précision. De plus, 20,6% des prédictions (1194 cas) ont une ER comprise entre 2 et 5%. À mesure que l'erreur augmente, les pourcentages diminuent drastiquement, indiquant que la majorité des prédictions du DNN sont très proches des valeurs réelles. Les erreurs supérieures à 10% sont marginales, soulignant la robustesse de ce modèle pour les applications QSPR.

Le SVM (Figure 72b) affiche des résultats solides, bien que légèrement en retrait par rapport au DNN sur les erreurs les plus basses : 64,8 % de ses prédictions (3 760 cas) présentent une ER comprise entre 0 et 2 %. Pour la plage d'erreur 2–5 %, il atteint 19,0 % (1 100 cas). Si ces performances restent remarquables, elles traduisent toutefois une dispersion un peu plus prononcée que celle du DNN. Au-delà, la proportion de prédictions décroît progressivement pour les intervalles supérieurs, mais de façon plus modérée que pour le DNN, ce qui témoigne d'une répartition d'erreurs légèrement plus étendue.

Quant au modèle XGBoost (Figure 72c), il affiche une distribution d'erreur plus large par rapport aux modèles DNN et SVM. Seulement 52,4% de ses prédictions (3039 instances) se situent dans la plage d'ER de 0 à 2%. Cependant, il compense en partie avec une proportion significative de prédictions dans l'intervalle d'erreur de 2 à 5% (30,0%, soit 1739 instances) et de 5 à 10% (11,4%, soit 660 instances). Cela indique que le modèle XGBoost, bien qu'étant moins performant pour des erreurs très faibles, offre une prédition raisonnablement précise pour un plus grand nombre de cas avec des erreurs moyennes. Les erreurs plus importantes, bien que présentes, restent minoritaires.

En comparant la Figure 69 (modèles de GC) et la Figure 72 (modèles de QSPR), on observe des tendances similaires en termes de performance relative des modèles (DNN étant le plus précis, suivi par SVM, puis XGBoost), mais avec des niveaux de précision légèrement meilleurs pour les modèles de QSPR. Le DNN, en particulier, montre une capacité supérieure à confiner les erreurs dans l'intervalle 0-2% pour les modèles de QSPR (69,5%) par rapport aux modèles de GC (66,7%), suggérant que les descripteurs utilisés dans les QSPR permettent une prédition marginalement plus fine. Cependant, dans les deux cas, le DNN démontre une nette supériorité pour minimiser l'erreur relative moyenne, tandis que le XGBoost, bien que moins performant sur les très faibles erreurs,

maintient une distribution plus large de ses prédictions, avec une proportion plus élevée d'erreurs moyennes.

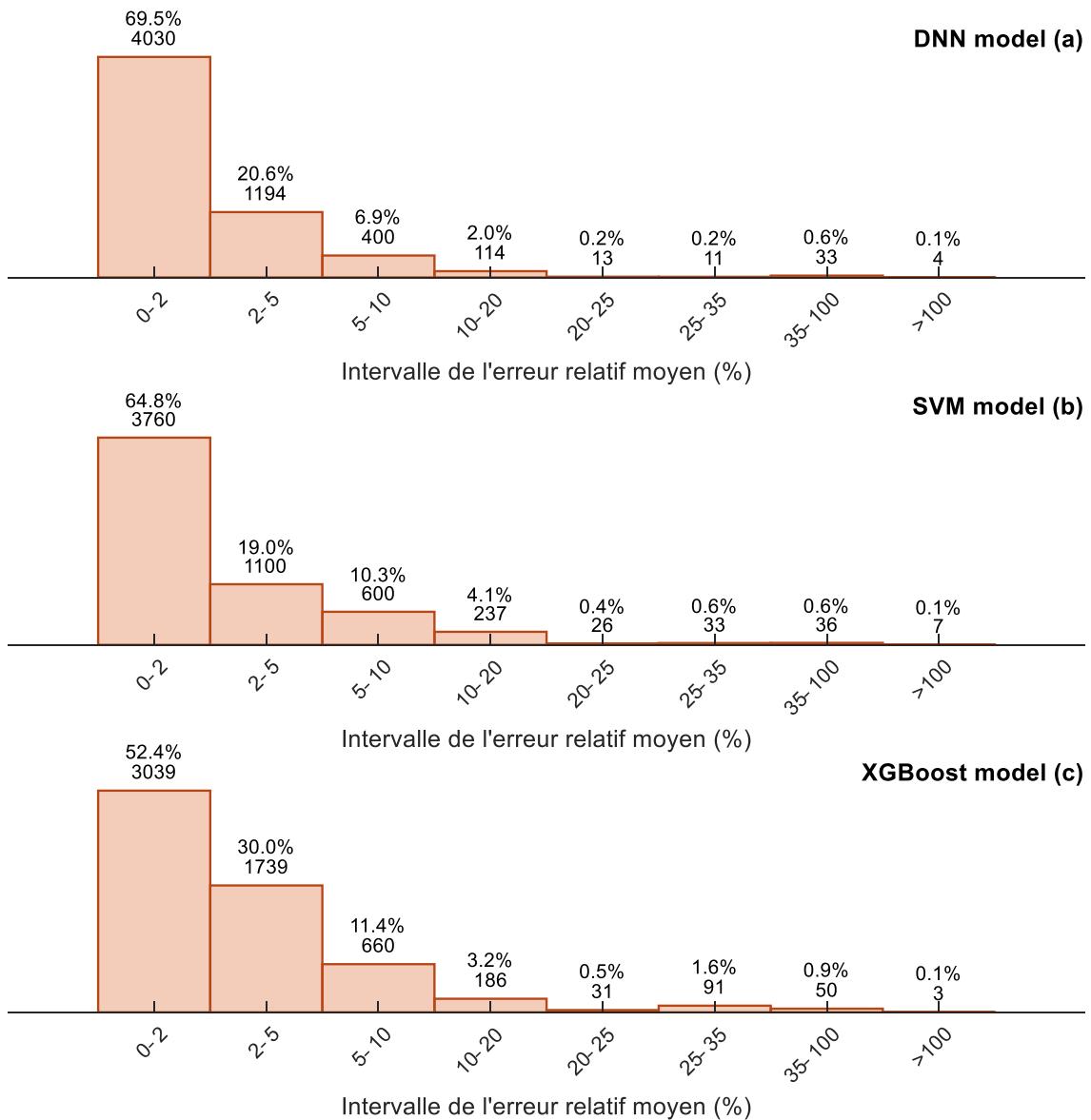


Figure 72. : Distribution de l'erreur relative moyenne (ER %) des modèles de QSPR :
(a) DNN, (b) SVM, (c) XGBoost.

V.4.3 Comparaison des modèles proposés par rapport aux ceux existants

Le Tableau 18 présente l'erreur relative absolue moyenne (ERAM) obtenue par les modèles développés dans ce travail, comparée à un ensemble représentatif de méthodes publiées dans la littérature. Nos modèles QSPR/DNN et GC/DNN, respectivement entraînés sur 5 800 et 7 015 points de données, atteignent des ERAM remarquablement

basses de 2,67 % et 2,94 %. Ces valeurs témoignent non seulement de l'efficacité des réseaux de neurones profonds (DNN) couplés à descripteurs globaux (QSPR) ou fragmentaires (GC), mais aussi de la richesse et de la couverture étendue de notre base de données (5 800–7 015 cas), qui dépasse en volume la plupart des études antérieures.

À titre de comparaison, les approches classiques de corrélation ou de régression linéaire montrent des performances globalement moins satisfaisantes. Par exemple, Zhang et al. [19] rapportent une ERAM de 7,5 % pour un modèle QSPR/LR entraîné sur 7 342 points, tandis que Valderrama et al. [31] atteignent 5,4 % avec une Équation d'État cubique couplée à VPT sur 3 857 cas. Les méthodes GC/FFANN-LSSVM de Paduszyński [3] enregistrent une erreur moyenne élevée de 32,28 % malgré un nombre de 13 539 données, ce qui souligne la difficulté de généralisation des réseaux peu profonds face à une grande diversité structurale.

Nos modèles à base de SVM et XGBoost montrent quant à eux des ERAM intermédiaires. La version QSPR/SVM (5 800 points) et QSPR/XGBoost (5 800 points) obtiennent toutes deux 6,24 %, tandis que GC/SVM (7 015 points) atteint 4,61 %. Ces résultats se situent dans la fourchette des performances rapportées par Haghbakhsh & Raeissi [56] (6,9 %) et Baghban et al. [15] (7,9 % en GC/LSSVM), mais restent significativement supérieurs aux modèles plus anciens de Gharagheizi et al. [8] (31,0 % en GC/MLR) ou de Yan et al. [20] (19,1 % en QSPR/MLR).

Deux études méritent une mention particulière : Darabi & Zare [61], qui obtiennent une ERAM de 2,69 % sur 1 461 points via une corrélation linéaire avancée, et He et al. [23], qui atteignent 1,92 % avec une combinaison Équation d'État cubique + Eyring appliquée aux données GC (1 070 cas). Ces méthodes rivalisent avec nos meilleurs résultats, mais leur domaine d'application et la nature des liquides étudiés sont plus restreints, ce qui limite leur généralisabilité. En revanche, nos DNN bénéficient d'une portée plus large, couvrant un spectre multiple de familles ioniques et de conditions expérimentales.

Ainsi, les modèles DNN proposés surpassent la majorité des approches existantes en termes de précision globale, tout en s'appuyant sur une base de données plus conséquente et diversifiée. Les algorithmes SVM et XGBoost offrent quant à eux un compromis intéressant entre robustesse et complexité, bien qu'ils affichent une dispersion d'erreur légèrement plus marquée. Ces résultats soulignent l'intérêt d'intégrer des architectures

profondes et des descripteurs complets pour atteindre une performance prédictive optimale dans l'étude des liquides ioniques.

Tableau 18: Comparaison des performances des modèles proposés par rapport aux modèles existants.

Auteurs	N° de points de données	Méthode/technique	Années	ERAM (%)
Ce travail	5800	QSPR/DNN	2025	2.67
Ce travail	7015	GC/DNN	2025	2.94
Ce travail	5800	QSPR/SVM	2025	6.24
Ce travail	5800	QSPR/ XGBoost	2025	6.24
Ce travail	7015	GC/SVM	2025	4.61
Paduszyński [3]	13539	GC/FFANN-LSSVM	2019	32.28
Zhang et al. [19]	7342	QSPR/LR	2021	7.5
Valderrama et al. [31]	3857	cubic EoS + VPT	2019	5.4
Yan et al. [20]	3228	QSPR/MLR	2018	19.1
Haghbakhsh & Raeissi [56]	2073	correlation/ANN	2019	6.9
Baghban et al. [15]	1672	GC/LSSVM	2017	7.9
Gharagheizi et al. [8]	1672	GC/MLR	2012	31.0
Kang et al. [59]	1502	QSPR/ELM	2017	20
Zhao et al. [60]	1502	QSPR/MLR	2015	59.4
Darabi and Zare [61]	1461	correlation/LR	2020	2.69
Yongsheng et al. [45]	1079	correlation/SVM	2016	7.62
He et al. [23]	1070	cubic EoS + Eyring/GC	2018	1.92

V.5 CONCLUSION

Ce chapitre démontre la supériorité des approches d'apprentissage automatique, en particulier le Deep Neural Network (DNN), pour la prédiction haute fidélité de la viscosité des liquides ioniques à base d'Imidazolium. Le modèle DNN, couplé aux descripteurs QSPR ou GC, atteint des erreurs relatives moyennes remarquables (2,67 % et 2,94 % respectivement), surpassant significativement les modèles conventionnels et les méthodes existantes de la littérature. Le SVM émerge comme une alternative compétitive, tandis que le XGBoost et le RF présentent des limitations en généralisation. Les analyses par familles anioniques confirment la robustesse du DNN face à la diversité structurale, bien que des défis persistent pour les anions rares (e.g., borates organiques). Comparativement, l'approche QSPR offre une précision marginalement supérieure, mais la méthode GC fournit une interprétation chimique plus directe via les contributions de

groupes. Ces modèles, validés sur une base de données étendue et diversifiée, constituent des outils prédictifs puissants pour la conception rationnelle de LI. Des perspectives futures incluent l'extension à d'autres cations, l'intégration des effets de pression et l'optimisation d'architectures neuronales pour capturer davantage de complexité structure-propriété.

V.6 RÉFÉRENCE

- [1] C. Wang, X. Wei, X. Jin, J. Li, M. He, Developing a two-grade model for the thermal conductivity of ionic liquids and their mixtures, *Chem. Eng. Sci.* 290 (2024) 119881. <https://doi.org/10.1016/j.ces.2024.119881>.
- [2] A.P. Fröba, M.H. Rausch, K. Krzeminski, D. Assenbaum, P. Wasserscheid, A. Leipertz, Thermal conductivity of ionic liquids: Measurement and prediction, *Int. J. Thermophys.* 31 (2010) 2059–2077. <https://doi.org/10.1007/s10765-010-0889-3>.
- [3] A.Z. Zhao, J.E. Garay, High temperature liquid thermal conductivity: A review of measurement techniques, theoretical understanding, and energy applications, *Prog. Mater. Sci.* 139 (2023) 101180. <https://doi.org/10.1016/j.pmatsci.2023.101180>.
- [4] R. Haghbakhsh, S. Raeissi, A novel correlative approach for ionic liquid thermal conductivities, *J. Mol. Liq.* 236 (2017) 214–219. <https://doi.org/10.1016/j.molliq.2017.03.117>.
- [5] L.H. Soares, R. Guirardello, M.P. Rolemburg, A simple group contribution model to predict thermal conductivity of pure ionic liquids, *Chem. Eng. Trans.* 74 (2019) 1195–1200. <https://doi.org/10.3303/CET1974200>.
- [6] R.M. Almeida, M.J. V. Lourenço, C. Nieto de Castro, The thermal conductivity of ionic liquids. Experiment and molecular interpretation, *J. Mol. Liq.* 397 (2024). <https://doi.org/10.1016/j.molliq.2024.124123>.
- [7] H. Yang, R.C. Gallagher, A.T. Phan, P. Chartrand, A.E. Gheribi, A predictive approach for the compositional and temperature representation of thermal conductivity in multicomponent molten salt systems for advanced energy applications, *Mater. Today Energy.* 38 (2023) 101441. <https://doi.org/10.1016/j.mtener.2023.101441>.
- [8] R. Wan, M. Li, F. Song, Y. Xiao, F. Zeng, C. Peng, H. Liu, Predicting the Thermal Conductivity of Ionic Liquids Using a Quantitative Structure–Property Relationship, *Ind. Eng. Chem. Res.* 61 (2022) 12032–12039. <https://doi.org/10.1021/acs.iecr.2c01964>.
- [9] S.P. Mousavi, S. Atashrouz, F. Rezaei, M.E. Peyvastegan, A. Hemmati-Sarapardeh, A. Mohaddespour, Modeling thermal conductivity of ionic liquids: A comparison between chemical structure and thermodynamic properties-based models, *J. Mol. Liq.* 322 (2021) 114911. <https://doi.org/10.1016/j.molliq.2020.114911>.
- [10] K.J. Wu, C.X. Zhao, C.H. He, Development of a group contribution method for determination of thermal conductivity of ionic liquids, *Fluid Phase Equilib.* 339 (2013) 10–14. <https://doi.org/10.1016/j.fluid.2012.11.024>.
- [11] R.L.G. and J.A.P. Coutinho, PVAm–PIP/PS composite membrane with high performance for CO₂/N₂ separation, *AIChE J.* 55 (2009) 1274–1290. <https://doi.org/10.1002/aic>.
- [12] J.A. Lazzús, G. Pulgar-Villarroel, Estimation of thermal conductivity of ionic liquids using quantitative structure-property relationship calculations, *J. Mol. Liq.* 211 (2015) 981–985. <https://doi.org/10.1016/j.molliq.2015.08.037>.
- [13] J. Albert, K. Müller, Thermal conductivity of Ionic Liquids: An estimation approach, *Chem. Eng. Sci.* 119 (2014) 109–113. <https://doi.org/10.1016/j.ces.2014.08.023>.
- [14] J. Albert, K. Müller, A group contribution method for the thermal properties of ionic liquids, *Ind. Eng. Chem. Res.* 53 (2014) 17522–17526. <https://doi.org/10.1021/ie503366p>.

- [15] S.A. Shojaee, S. Farzam, A.Z. Hezave, M. Lashkarbolooki, S. Ayatollahi, A new correlation for estimating thermal conductivity of pure ionic liquids, *Fluid Phase Equilib.* 354 (2013) 199–206. <https://doi.org/10.1016/j.fluid.2013.06.004>.
- [16] S. Atashrouz, A. Hemmati-Sarapardeh, H. Mirshekar, B. Nasernejad, M. Keshavarz Moraveji, On the evaluation of thermal conductivity of ionic liquids: Modeling and data assessment, *J. Mol. Liq.* 224 (2016) 648–656. <https://doi.org/10.1016/j.molliq.2016.09.106>.
- [17] J.A. Lazzús, Estimation of the thermal conductivity $\lambda(T, P)$ of ionic liquids using a neural network optimized with genetic algorithms, *Comptes Rendus Chim.* 19 (2016) 333–341. <https://doi.org/10.1016/j.crci.2015.09.010>.

CONCLUSION GENERALE

CONCLUSION GENERALE

Cette thèse a permis de franchir une étape significative dans la modélisation des propriétés de transport des liquides ioniques à savoir ; la conductivité électrique, la conductivité thermique et la viscosité. Ces propriétés sont essentielles pour optimiser l'utilisation des IL dans des applications industrielles et scientifiques variées. En combinant des bases de données exhaustives avec des approches de modélisation avancées, notamment les contributions de groupes, les QSPR et des algorithmes d'apprentissage automatique (Réseaux de neurones profonds, machines à vecteurs de support, XGBoost, forêts aléatoires), ce travail a abouti à des modèles prédictifs robustes et précis.

Les principaux apports de cette recherche peuvent être résumés en trois points clés :

1. **Construction de bases de données riches et diversifiées** : Des ensembles de données complets ont été constitués, comprenant 6014 points pour la conductivité électrique (373 IL), 1025 points pour la conductivité thermique (79 IL) et 7015 points pour la viscosité (620 IL à base d'imidazolium). Ces bases, les plus vastes et les plus à jour à ce jour comparativement aux études antérieures, ont servi de fondation solide pour les modèles développés
2. **Performances prédictives exceptionnelles** : Les modèles, en particulier ceux basés sur les réseaux de neurones profonds (DNN) et les machines à vecteurs de support (SVM), ont atteint des précisions remarquables, avec des erreurs relatives moyennes souvent inférieures à 3 % et des coefficients de détermination (R^2) proches de 0,99. Ces résultats surpassent les méthodes existantes, grâce à l'utilisation de données étendues et d'algorithmes sophistiqués.

3. Validation rigoureuse : Les modèles ont été soumis à des tests de validation croisée et externe, confirmant leur capacité à généraliser sur des IL variés, bien que des limites subsistent pour certaines familles chimiques sous-représentées (anions rares, nouveaux cations). Les résultats ont été présentés en fonction des familles de liquides ioniques étudiées et des combinaisons cationiques et anioniques.

L'importance de ces résultats réside dans leur contribution à la conception rationnelle des liquides ioniques. En offrant des outils prédictifs fiables, cette thèse réduit la dépendance aux mesures expérimentales coûteuses, accélérant ainsi l'innovation dans des domaines stratégiques. Les modèles développés permettent de prédire avec précision les propriétés de transport, facilitant la sélection ou la synthèse de nouveaux IL adaptés à des besoins spécifiques.

Pour l'avenir, plusieurs perspectives se dessinent. L'élargissement des bases expérimentales, l'extension aux autres classes de LI et aux propriétés émergentes (tension interfaciale, conductivité ionique sélective) ainsi que l'essor de modèles hybrides (couplant apprentissage profond et simulations atomistiques) constituent des axes prometteurs pour renforcer encore la précision et la portée de ces méthodes prédictives.

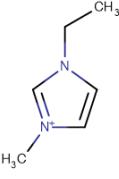
Cette thèse a établi une avancée majeure dans la prédiction des propriétés des liquides ioniques, en combinant des approches novatrices et des données exhaustives pour produire des outils à la fois précis et pratiques. Ces contributions ouvrent la voie à de nouvelles recherches et applications, renforçant le rôle des IL dans les technologies durables et innovantes.

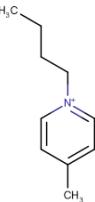
ANNEXE

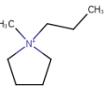
ANNEXE 1 : EXTRAIRE DE LA BASE DE DONNEES DE LA CONDUCTIVITE ELECTRIQUE

Structure	Abbreviation	Formula	Name
	[pyr-4,1] [fsa]	C ₉ H ₂₀ F ₂ N ₂ O ₄ S ₂	1-butyl-1-methylpyrrolidinium bis(fluorosulfonyl)amide
	Method		Reference
	Alternating current cell with electrodes		Makino, T.; Kanakubo, M.; Umecky, T.; Suzuki, A.; Nishida, T.; T
Temperature, (K)	Pressure, (kPa)	Electrical conductivity, (S/m)	Uncertainty, (S/m)
273.15	100	0.236	0.009
278.15	100	0.295	0.012
283.15	100	0.363	0.015
288.15	100	0.439	0.018
293.15	100	0.525	0.021
298.15	100	0.619	0.025
303.15	100	0.722	0.029
313.15	100	0.959	0.038
323.15	100	1.231	0.049
333.15	100	1.538	0.062
343.15	100	1.865	0.075
353.15	100	2.23	0.089

Structure	Abbreviation	Formula	Name
	[pyr-1o1,1] [fsa]	C ₇ H ₁₆ F ₂ N ₂ O ₅ S ₂	1-(methoxymethyl)-1-methylpyrrolidinium bis(fluorosulfonyl)amide
	Method		Reference
	Alternating current cell with electrodes		Makino, T.; Kanakubo, M.; Umecky, T.; Suzuki, A.; Nishida, T.; T
Temperature, (K)	Pressure, (kPa)	Electrical conductivity, (S/m)	Uncertainty, (S/m)
273.15	100	0.46	0.018
278.15	100	0.554	0.022
283.15	100	0.659	0.026
288.15	100	0.772	0.031
293.15	100	0.898	0.036
298.15	100	1.033	0.041
303.15	100	1.178	0.047
313.15	100	1.492	0.06
323.15	100	1.841	0.074
333.15	100	2.222	0.089
343.15	100	2.636	0.1
353.15	100	3.068	0.12

Structure	Abbreviation	Formula	Name
	[im-2,1] [cl]	C ₆ H ₁₁ CIN ₂	1-ethyl-3-methylimidazolium chloride
Method			
Alternating current cell with electrodes		Reference	
Temperature, (K)	Pressure, (kPa)	Electrical conductivity, (S/m)	Uncertainty, (S/m)
289	101.325	0.0025	0.0007
295.8	101.325	0.0249	0.0019
302.9	101.325	0.129	0.004
308.7	101.325	0.194	0.005
312.8	101.325	0.256	0.006
317.7	101.325	0.344	0.008
323.7	101.325	0.487	0.011
328.4	101.325	0.628	0.014
333.5	101.325	0.802	0.018
337.8	101.325	0.975	0.021
341.2	101.325	1.12	0.02
345.6	101.325	1.33	0.03
350.1	101.325	1.57	0.03
355.1	101.325	1.87	0.04
360.4	101.325	2.15	0.05
366.7	101.325	2.67	0.06

Structure	Abbreviation	Formula	Name
	[py-4,1[4]] [bf4]	C ₁₀ H ₁₆ BF ₄ N	1-butyl-4-methylpyridinium tetrafluoroborate
Method			
Conductimeter from CRISON, model GLP31		Reference	
Temperature, (K)	Pressure, (kPa)	Electrical conductivity, (S/m)	Uncertainty, (S/m)
278.15	101.325	0.0381	0.0012
280.65	101.325	0.0485	0.0015
283.15	101.325	0.0601	0.0019
285.65	101.325	0.075	0.0024
288.15	101.325	0.0911	0.0029
290.65	101.325	0.1127	0.0036
293.15	101.325	0.1318	0.0042
295.65	101.325	0.1574	0.005
298.15	101.325	0.1854	0.0059
300.65	101.325	0.218	0.007
303.15	101.325	0.254	0.008
305.65	101.325	0.297	0.009
308.15	101.325	0.339	0.011
310.65	101.325	0.39	0.012
313.15	101.325	0.439	0.014
315.65	101.325	0.497	0.016

Structure	Abbreviation	Formula	Name
	[pyr-3,1] [la]	C ₁₁ H ₂₃ NO ₃	N-methyl-N-propylpyrrolidinium lactate
Method			
Alternating current cell with electrodes		Reference	
Temperature, (K)	Pressure, (kPa)	Electrical conductivity, (S/m)	Uncertainty, (S/m)
278	101.325	0.04	0.01
282	101.325	0.06	0.01
288	101.325	0.09	0.01
294	101.325	0.14	0.01
298	101.325	0.2	0.01
302	101.325	0.23	0.01
308	101.325	0.32	0.01
313	101.325	0.42	0.01
317	101.325	0.5	0.01
323	101.325	0.62	0.01
328	101.325	0.72	0.02
333	101.325	0.87	0.02
338	101.325	1	0.02
342	101.325	1.18	0.02
349	101.325	1.36	0.03