République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



UNIVERSITE YAHIA FARES DE MEDEA
Faculté de la Technologie
Département de Génie des Procédés et Environnement

MEMOIRE DE MASTER

***En vue de l'obtention du Diplôme de Master en Génie des* Procédés**

Option : Génie Chimique

Thème

Modélisation du coefficient d'Activité à dilution infinie des liquides ioniques dans différents solutés (eau, n-alcanes, cycloalcanes) en utilisant les machines à vecteur de support

Présenté par :
Khedidja Chouih

Dirigé par :
Dr. Chérif SI MOUSSA

Promotion 20017-2018

Remerciements

Remerciements à Allah le tout puissant pour la volonté, la santé et la patience qui m'procuré pour réaliser ce modeste travail

*Je tiens à remercier mon promoteur, **Dr. Chérif SI-MOUSSA** Pour son encadrement, sa disponibilité, et son suivi durant toute la période du projet.*

Je tiens à remercier les membres de Jury le Pr. Salah HANINI et le Dr. Maamar LAIDI pour avoir accepter de juger ce travail.

Je tiens aussi à témoigner ma reconnaissance à toute personne ayant aidée, de prés ou de loin, à l'aboutissement de ce travail.

En fin j'adresse les remerciements très sincères pour tous les enseignants qui ont contribué a notre formation et aux étudiants et étudiantes de ma promotion.

Merci à tous et à toutes

DEDICACE

Je dédie ce modeste travail :

Ma chère mère, pour ses sacrifices depuis qu'elle

M'a mis au monde. Et à mon père.

Ames chers frère Mohamed, Ibrahim, Abdelkader

Et me sœur Yamina

A tout la famille CHOUIH

A toutes les amies

Résumé

ملخص

نمذجة معامل النشاط التخفيف اللانهائي للسوائل الأيونية في المواد المذابة المختلفة (الماء ، الألكانات ، السيكلان ألكان) باستخدام آلات ناقلات الدعم:تمثل السوائل الأيونية جيلًا جديدًا من المذيبات الأقل سمية ، والأقل قابلية للاشتعال والأقل تلويثًا ، وتعتبر بدائل للمذيبات العضوية المتطايرة. يعتبر معامل النشاط التخفيف اللانهائي للمذيبات في السوائل الأيونية خاصية مهمة جدًا لتوازن الطور. تم تطوير قاعدة بيانات تتكون من 4882 نقطة قياس تجريبية لـمعامل النشاط التخفيف اللانهائي لِلماء و 5 ألكانات حلقية و 11 ألكانات عادية في 80 من السوائل الأيونية عند درجات حرارة مختلفة. تم حساب مواصفات ل 29 كتيون و 14 أنيون استنادًا إلى ملامح σ لنموذج كيمياء الكم COSMO-RS. تم استخدام معاملات أبراهام للعلاقة الخطية لطاقة التذويب (LRSE) لوصف المحاليل السبعة عشر. تم استخدام أسلوب Least Square Least Support (LSSVM) للتحديد الأمثل لأربع نماذج العلاقة لبيانات معامل النشاط التخفيف اللانهائي. تم اختيار درجة الحرارة، والوصفات 8 على أساس ملامح σ للكتيون وللأنيون ومعاملات أبراهام الخمسة (LRSE) كمتغيرات المدخلات. تم اختيار اللوغاريتم الطبيعي لِمعامل النشاط التخفيف اللانهائي كمتغير مخرجات. إن نماذج LSSVM المحسنة في هذا العمل لديها القدرة على إيجاد علاقة الارتباط بطريقة مقبولة باستخدام لمعامل النشاط التخفيف اللانهائي للمواد المذابة (ماء ، سيكلو ألكان وألكانات) في السوائل الأيونية مع خطأ نسبي متوسط بين 10.30٪ و 20.50٪ ومعامل ارتباط بين 0.9919 و 0.9643 لـ جميع بيانات التحقق من الصحة. هذه النتائج مشجعة للغاية لمشاريع مستقبلية لنمذجة الخصائص الفيزيائية والكيميائية الأخرى للسوائل الأيونية.

الكلمات الدالة: السوائل الأيونية، معامل نشاط التخفيف اللانهائي، ملامح σ، (LSSVM)، العلاقة الخطية لطاقة التذويب.

Résumé

Modélisation du coefficient d'Activité à dilution infinie des liquides ioniques dans différents solutés (eau, n-alcanes, cycloalcanes) en utilisant les machines à vecteur de support : Les liquides ioniques représentent une nouvelle génération de solvants moins toxiques, moins inflammables et moins polluants et sont considérés comme des alternatives aux solvants organiques volatils. Le coefficient d'activité de dilution infinie (CADI) des solutés dans les liquides ioniques est une propriété très importante pour les équilibres de phase. Une base de données de 4882 points de mesure expérimentale de CADI de l'eau, 5 cycloalcanes et 11 n-alcanes dans 80 liquides ioniques à différentes températures a été élaborée. Les descripteurs pour 29 cations et 14 anions ont été calculés sur la base des profils σ du modèle de chimie quantique COSMO-RS. Les paramètres d'Abraham de la relation linéaire de l'énergie de solvatation (RLES) ont été utilisés pour décrire les 17 solutés. La méthode LSSVM (Least Square Support Vector Machines) a été utilisée pour optimiser 4 modèles de corrélation du CADI de la base de données. La température normalisée, les 8 descripteurs basés sur les profils σ des cations et des anions et les cinq paramètres d'Abraham (RLES) sont choisis comme variables d'entrée. Le logarithme naturel du CADI est choisi comme variable de sortie. Les modèles LSSVM optimisés dans ce travail permettent de corréler de façon acceptable le CADI des solutés (eau, cycloalcanes et n-alcanes) dans les liquides ioniques avec une erreur relative moyenne comprise entre 10,30% et 20,50% et un coefficient de corrélation compris entre 0,9919 et 0,9643 pour toutes les données de validation. Ces résultats sont très encourageants pour les futurs projets de modélisation d'autres propriétés physiques et chimiques des liquides ioniques.

Mots clés: Liquides Ioniques, Coefficient d'Activité à Dilution Infinie, profils σ, LSSVM (Least Square Vector Support Machines), Relation Linéaire de l'Energie de Solvatation.

Abstract

Modeling of infinite dilution activity coefficient of ionic liquids in different solutes (water, n-alkanes, cycloalkanes) using support vector machines:Ionic liquids represent a new generation of less toxic solvents, less flammable and less polluting and are considered as alternatives to volatile organic solvents. The infinite dilution activity coefficient (IDAC) of solutes in ionic liquids is a very important property for phase equilibrium. A database of 4882 experimental measurement points of IDAC of water, 5 cycloalkanes and

11 n-alkanes in 80 ionic liquids at different temperatures was developed. The descriptors for 29 cations and 14 anions were calculated based on the σ profiles of the COSMO-RS quantum chemistry model. Abraham's parameters of the linear relationship of solvation energy (LRSE) were used to describe the 17 solutes. The Least Square Support Vector Machines (LSSVM) method was used to optimize 4 IDAC data correlation. Normalized temperature, 8 descriptors based on cation and anion σ profiles and Abraham's five parameters (LRSE) are chosen as input variables. The natural logarithm of the IDAC is chosen as output variable. The LSSVM models optimized in this work have the ability to correlate in an acceptable way the IDAC of solutes (water, cycloalkanes and normal alkanes) in ionic liquids with a mean relative error between 10.30% and 20.50% and a correlation coefficient between 0.9919 and 0.9643 for all validation data. These results are very encouraging for future modeling projects of other physical and chemical properties of ionic liquids.

Key words: Ionic Liquids, Infinite Dilution Activity Coefficient, σ profiles, Least Square Vector Support Machines (LSSVM), Linear Relationship of Solvation Energy.

Table des matières

Table des matières I
Abréviations V
Nomenclature et symbole VI
Liste des figures. VII
Liste des tableaux X

Introduction générale……..……1

Chapitre I : Généralités sur les liquides ioniques (LIs)

I.1. Introduction……..4
I.2. Définition…….5
I.3. Historique…….6
I.4. Structure des liquides ioniques……6
I.5. Synthèses……..……8
I.5.1 Réaction de quaternisation…….……...……...8
I.5.2 Réaction d'échange de l'anion…….……10
I.5.3 Purification et pureté des LI…….…11
I.5.4 Nomenclature et acronymes des LI……...12
I.6. Propriétés uniques des Liquides Ioniques……..13
I.7.Propriétés physico-chimiques des liquides ioniques……...13
I.7.1. Stabilité thermique et point de fusion……...14
I.7.2. La densité et la viscosité……..…14
I.7.3 Conductivité……..……15
I.7.4. Solubilité……...……...…15

I.7.5.Stabilité chimique……..……15
I.7.6.Toxicité et biodégradabilité……..…….15
I.7.7. Volatilité……..…….……16
I.8. Propriétés spécifiques pour le traitement par extraction……16
I.9. Applications……...16
I.9.1.Electrochimie……...17
I.9.2. Chimie Analytique……..18
I.9.3. Synthèse et catalyse……19
I.9.4. Biotechnologie……19
I.10.Utilisation des liquides ioniques dans des procédés de séparation……...19
I.10.1.Distillation extractive…….…….20
I.10.2.Extraction liquide-liquide…….……..20
I.10.2.1. Extraction des alcools pour la synthèse des biocarburant…….20
I.10.2.2.Extraction de composés aromatiques des vapeur issues de procédés de raffinage.21
I.10.2.3 Désulfuration des carburants…….……..21
I.11. Les Avantages des liquides ioniques……...22
I.11.Conclusion……23

Chapitre II : Coefficients d'activité à la dilution infinie(IDAC)

II.1.Introduction……...24
II.2.Coefficients d'activité à la dilution infinie…….24
II.2.1.Définition……..……..…….24
II.2.2. Importance et utilisation des données de coefficient d'activité de dilution infinie….……25
II.2.2.1. Synthèse, conception et optimisation des processus de séparation……..…..26
II.2.2.2. Calcul des constantes de la loi de Henry……..…..27
II.2.2.3. Prédire l'existence d'un azéotrope……..….28
II.2.2.4. Développement de méthodes prédictives pour G^E 28
II.2.2.5. Caractérisation du comportement pour les mélanges liquides……...28
II.2.3. Dépendance de la température du coefficient d'activité……29
II.2.4. Modèles de coefficients d'activité prédictive…….29
II.2.4.1. Méthodes de contribution de groupe. (GCMs)…….30
II.2.4.1.1.Le modèle ASOG (solution analytique du groupe)…….…...30
II.2.4.1.2.Le modèle UNIFAC et ses modifications……..…....31

II.2.4.2. Méthodes basées sur COSMO……..…….32

II.2.5. Techniques expérimentales pour les mesures d'IDAC……..34

II.2.5.1. Chromatographie gaz-liquide (GLC)……...…….34

II.2.5.2. La technique d'extraction de gaz inerte. (IGS)……...……..35

II.2.6. Les descripteurs moléculaires basés sur profils σ des…….…..…….36

II.3.Conclusion…….38

Chapitre III : SupportVector Machines (SVM)

III.1.Introduction……..39

III .2.Pourquoi les machines à vecteur de support……...40

III.3.Apprentissage statistique et SVM……41

III.3.1 Objectif de l'apprentissage statistique……...41

III.3.2 Théorie de Vapnik-Chervonenkis…….42

III.3.3 Marge et dimension de VC…….……...44

III .4.SVM principe de fonctionnement général……...46

III .4.1. Notions de base: Hyperplan, marge et support vecteur……...……...46

III .4.2. Pourquoi maximiser la marge ?.....48

III .4.3. Linéarité et non-linéarité……..…..49

III .4.4 Cas non linéaire……...…50

III .4.5 .Illustration de transformation de cas non linéaire : le cas XOR…….....50

III.5.Least Squares SVM (LSSVM)…….52

III.5.1 .Description de la méthode et résolution……...……52

III.5.1.1. Recherche de la solution par utilisation du théorème de représentation…….53

III.5.1.2. Recherche de la solution par exploitation des conditions de points selle……...55

III.5.1.3 .Recherche de la solution par résolution des équations de Karush, Kuhn et Tucker.....57

III.6. Les Machines à Vecteurs Supports (SVM) pour la régression……...59

III.6.1 .Formulation du problème dans le cas linéaire.…..……...…….…..…..59

III.6.2 .Extension au cas non linéaire.…..…….…..…..…..62

III.6.3 .Choix des hyperparamètres.…..……...…….…..…..…..63

III.6.4 .Choix de la fonction de coût.…..…..…..…….…..…......64

III.6.4.1 Fonctions de coût et modèles de bruit associés…….……64

III.6.4.2 Problème d'optimisation pour une fonction de coût générale…….……..66
III .7.Les domaines d'applications……...67
III .8.Conclusion……...67

Chapitre IV : Résultats et discussion

IV.1. Introduction…….……69
IV.2 .Elaboration de la base de données……..69
IV.2.1. Base de données du CADI…….…..70
IV.2.2 .Base de données des descripteurs moléculaires des cations et des anions…….…..73
IV.2.3 .Bases de données des descripteurs moléculaires des solutés…….…..79

IV.3. Modélisation par LSSVM……...81

IV.4. Choix des variables d'entrées…….83
IV.5. Normalisation des données…….84
IV.6. Division des données en ensembles apprentissage/test……...85
IV.7.Ajustement de la constante de régularisation γ et du paramètre kernel σ^2……85
IV.8. Résultats de corrélation des modèles optimisés……..86
IV.8.1. Résultats du modèle optimisé pour la base B1……...….86
IV.8.2. Résultats du modèle optimisé pour la base B2……87
IV.8.3. Résultats du modèle optimisé pour la base B3……...….88
IV.8.4. Résultats du modèle optimisé pour la base B4……....89
IV.8.5. Evaluation des erreurs des modèles optimisés……..…..90
IV. 9. Conclusion……...…..91
Conclusion générale..……92
Référence……...……..……94
ANNEXE…….....

Liste des Abréviations

Abréviation	signification
CADI	Coefficient d'activité à dilution infinie
CGL	**C**hromatographie **g**az-**l**iquide
COSMO-RS	COnductor comme Screening MOdel pour de vrais solvants Solvents
IGS	La technique d'extraction de gaz inerte
LIs	**L**iquides **I**oniques
LSSVM	moindres carrés machines de vecteur de support
MSE	erreur quadratique moyenne (Mean Square Error)
RL	**R**égression **l**inéaire
SVM	Soutenir les machines vectorielles (machine à vecteurs de support)

Nomenclature et symboles

Symbole	Désignation	Unité
A	Surface de pic soluté détectée par chromatographie en phase gazeuse	mV.min
B --	Deuxième coefficient viriel	$cm^3 mol^{-1}$
a	L'activité chimique	Pas d'unité
α	Pente	min^{-1}
$f\hat{}$	Fugacité pour les espèces en solution	*Pa*
f_i	Fugacité du corps pur	*Pa*
f_{i^*}	Fugacité du corps réel	*Pa*
G	L'énergie de Gibbs	Pas d'unité
H	La constante de la loi de Henry	Pas d'unité
H	Enthalpie	J
I	Correction de Poynting	Pas d'unité
K	Coefficient de partitionnement	Pas d'unité
k	Capacité	Pas d'unité
k i	Facteurs de correction	Pas d'unité
k_L	Coefficient de transfert de masse dans le liquide	$mol.s^{-1}.cm^2$
P	Pression	kPa
P^o or P*	pression de vapeur	kPa
P	Pression partielle	kPa
R	Constante de gaz	$J.mol^{-1} K^{-1}$
T	Température absolue	°K
T_b	Température du point d'ébullition	°K
x	fraction dans la phase liquide	Mole
y	fraction en phase vapeur	Mole
G^e	l'énergie libre de Gibbs d'excès.	Pas d'unité
α	Facteur de Séparation	Pas d'unité
k^{∞}_j	numériquement la quantité maximale d'espèce j	Pas d'unité
β^{∞}_{ij}	sélectivité limite des solvants s pour le système constitué des composants i et j	Pas d'unité
γ^{∞}_{ij}	coefficients d'activité limite de i et j respectivement dans le solvant	Pas d'unité

γ	Coefficient d'activité	Pas d'unité
σ	Densité surfacique de charge	e / nm^2

Exposants

Symbole	Désignation
sat	À la saturation
E	Propriétés excédentaires
exp	Expérimentalement déterminé
Cal	Calculer
V	Phase de vapeur
L	Phase liquide
G	Phase gazeuse

Indices

Symbole	Désignation
i	Propriétés liées au composant pur i
ij	Propriétés d'interaction
C	Propriétés critiques
O	Valeur initiale
F	Valeur finale
∞	Dilution infinie

Liste des figures

Figure	Titre de la figure	Page
Partie théorique		
Chapitre I		
Figure I.1	Classification des solvants selon le type de liaisons chimiques mis en jeu	**4**
Figure I.2	Photographie du chlorure de sodium (à gauche) et le liquide ionique [hmim][PF_6] (à droite) à température ambiante	**5**
Figure I.3	Structure chimique des principaux cations des LI rencontrés dans la littérature	**6**
Figure I.4	Différentes classes de liquides ioniques	**8**
Figure I.5	Exemples de réaction de quaternisation de l'imidazole (A et B) et du DBU (C et D).	**9**
Figure I.6	Synthèse du chlorure de I-butyl-3-méthylimidazolium par quaternisation.	**10**
Figure I.7	les anions les plus utilisés	**12**
Figure I.8	Nombre de publications par année portant sur les liquides ioniques	**13**
Figure I.9	Comparaison de l'état de l'eau et des liquides ioniques.	**14**
Figure I.9	Les domaines d'applications des liquides ioniques	**17**
Chapitre II		
Figure II .1	σ^- profiles de l'eau (H2O), Hydronium ion ($H3O^+$) et hydroxyde (OH^-)dans les modèles de type cosmo [15].	**33**
Figure II .2	σ-profile des cations:1-(2-hydroxyéthyl)-3-méthylimidazolium ([HOC2C1im]), 1-butyl-3-méthylimidazolium ([C 4 C 1 Im]) et des anions: méthanesulphonate ([MeSO3]) et acétate ([Ac]). [GONFA] [16].	**36**
Chapitre III		
Figure III.1	Principe des techniques SVM	**40**
Figure III.2	Exemple du problème de sur apprentissage.	**42**

Figure III.3	Illustration de l'inégalité (4.3)	**44**
Figure III.4	Classifieur linéaire et marge (4.3)	**45**
Figure III.5	Exemple d'un hyperplan séparateur	**46**
Figure III.6	Exemple de vecteurs de support	**47**
Figure III.7	Exemple de marge maximal (hyperplan valide)	**48**
Figure III.8	**a)** Hyperplan avec faible marge, **b)** Meilleur hyperplan séparateur	**48**
Figure III.9	Exemple de classification d'un nouvel élément	**49**
Figure III.10	**a)** Cas linéairement séparable, **b)** Cas non linéairement séparable	**49**
Figure III.11	Exemple de changement de l'espace de données	**50**
Figure III.12	Illustration de cas non linéairement séparable (le cas XOR)	**51**
Figure III.13	Illustration de passage d'un espace 2D à un espace 3D	**51**
Chapitre IV		
Figure IV.1	Profil σ de la molécule d'eau.	**74**
Figure IV.2	Profil σ des cations 1- (2-hydroxyéthyl) -3-méthylimidazolium ([IM-2OH,1]) et 1-butyl-3-méthyl imidazolium ([IM-4,1])	**75**
Figure IV.3	Profil σ des anions méthanesulfonate ([SO_3-1]) et l'acétate ([Ac]).	**76**
Figure IV.4	Ecran de numérisation du graphe de profil σ de l'anion dimethyl phosphate [PO2-O1,O1] en utilisant GetData Graph Digitizer 2.26.	**77**
Figure. IV.5	Diagrammes du nuage de points du logarithme népérien du CADI de l'eau dans les Lis (Base B1), (a) test, (b) apprentissage, (c) globale	87
Figure. IV.6	Diagrammes du nuage de points du logarithme népérien du CADI des cycloalcanes dans les Lis (Base B2), (a) test, (b) apprentissage, (c) globale	88
Figure. IV.7	Diagrammes du nuage de points du logarithme népérien du CADI des alcanes dans les Lis (Base B3), (a) test, (b) apprentissage, (c) globale	89
Figure. IV.8	Diagrammes du nuage de points du logarithme népérien du CADI de l'eau, des cycloalcanes et des alcanes dans les Lis (Base B4), (a) test, (b) apprentissage, (c) globale	90

Liste des tableaux

Tableau	Titre du tableau	Page
Partie théorique		
Chapitre II		
Tableau II.1	Avantages et inconvénients de la méthode de chromatographie liquide en phase gazeuse	**35**
Tableau II.2	Avantages et inconvénients de la méthode d'extraction au gaz inerte	**35**
Tableau II.3	Les profiles σ des cations: [HOC2C1im] et [C 4 C 1 im] et des anions :[MeSO3] et [Ac]	**38**
	Chapitre III	
Tableau III.1	les fonctions de coût usuelles ainsi que les distributions associées du bruit.	**65**
Chapitre IV		
Tableau IV.1	La liste des Lis sélectionnées, le nombre de point expérimentaux par liquide ionique et la référence bibliographique.	**70**
Tableau IV.2	Liste des cations	**72**
Tableau IV.3	Liste des anions	**73**
Tableau IV.4	Valeurs Sσ de deux cations et de deux anions typiques des Lis.	**76**
Tableau IV.5	Descripteurs moléculaires ($S_{\sigma i}$) des cations des Lis étudiés	79
Tableau IV.6	Descripteurs moléculaires ($S_{\sigma i}$) des anions des Lis étudiés	80
Tableau IV.7	Les valeurs numériques des descripteurs de solutés	80
Tableau IV.8	Taille des ensembles d'apprentissage et de test pour les 4 bases de données	85
Tableau IV.9	Les valeurs minimales et maximales des variables d'entrée et de sortie	85

Tableau IV.10	Résultats de l'ajustement optimal de la constante de régularisation γ et du paramètre kernel σ^2	85
Tableau IV.11	Erreur Relative Absolue Moyenne (ERAM) et coefficient de corrélation pour les modèles LSSVM optimisés	90

Introduction Générale

Les solvants moléculaires organiques sont utilisés dans différents procédés chimiques industriels avec des résultats techniques satisfaisants. Cependant, la plupart de ces produits chimiques représentent un risque pour la santé humaine et sont peu écologiques en raison de leur volatilité. Une bonne illustration du caractère destructeur des produits chimiques modernes est donnée, par exemple, des catastrophes du nuage toxique sur Seveso en Italie, 1976; des 3500 vies réclamées quand en 1984, l'usine de pesticide basée de Bhopal a libéré des émissions toxiques en Inde; et de l'explosion dévastatrice de l'usine chimique AZF dans la ville française de Toulouse en 2001.

De tels accidents contribuent à la sensibilisation accrue aux menaces environnementales et sanitaires résultant du grand nombre de processus industriels. Le développement de solvants moins polluants et de procédés sûrs est aujourd'hui une priorité pour l'agenda de la chimie verte. Il est urgent de repenser la conception des procédés chimiques dangereux afin que les jeunes générations héritent d'un environnement sain et sûr.

Ces dernières années, les liquides ioniques ont reçu le plus d'attention en tant que solutions de rechange aux solvants organiques volatils en chimie réactive, ainsi qu'aux procédés de séparation. Ceci est traduit par la tendance de croissance exponentielle du nombre de demandes de brevets et d'articles de journaux sur les liquides ioniques durant les deux dernières décennies (Fig. 1).

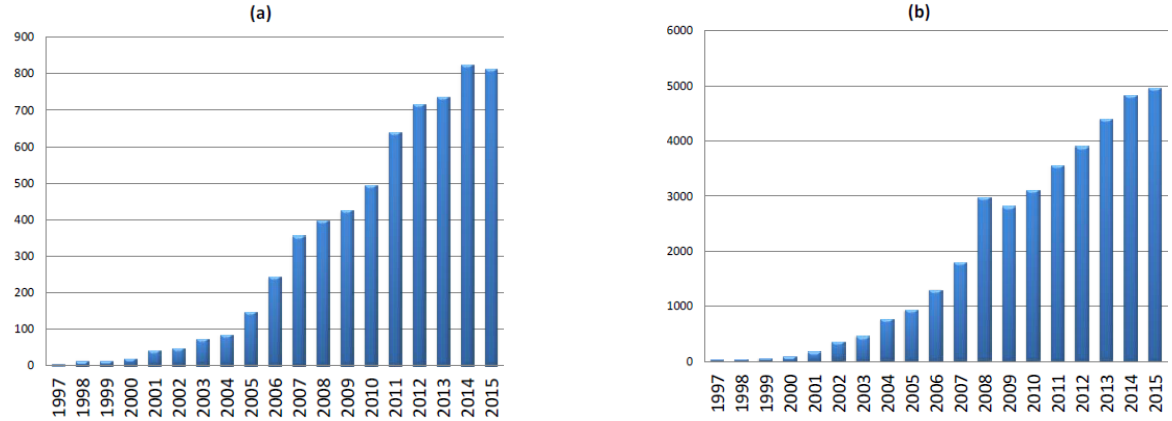


Fig. 1. Evolution du nombre total de demandes de brevets (a) et du nombre total des articles de journaux relatifs aux liquides ioniques chaque année (b).[https://doi.org/10.1016/j.seppur.2017.11.023]

Parmi ces publications un nombre très large rapporte les mesures systématiques des coefficients d'activité de dilution infinie (CADI) pour une large gamme de solutés organiques dans divers liquides ioniques afin d'évaluer leur efficacité en tant que solvant d'extraction. Cependant, environ 10^{18} liquides ioniques peuvent être obtenus par les différentes combinaisons cation-anion; ainsi, les

données expérimentales disponibles sur le CADI qui peuvent être utilisées pour guider les applications industrielles sont loin d'être adéquates. La détermination expérimentale CADI de tous les solutés dans les Lis envisageables prend du temps, coûteuse, laborieuse et impossible. En outre, les exigences sévères des conditions expérimentales ne peuvent être négligées. En conséquence, une compréhension préliminaire de la relation entre le CADI des solutés dans les liquides ioniques et la structure des IL et des solutés à une certaine température est nécessaire.

L'utilisation de méthodes alternatives et complémentaires à l'expérimentation, parmi lesquelles les relations quantitatives structure propriété/activité (QSPR/QSAR) sont devenues d'un grand intérêt et sont même recommandées dans les nouvelles règlementations afin d'obtenir les données nécessaires à l'enregistrement des substances. L'élaboration des modèles mathématiques QSPR/QSAR reliant les propriétés physicochimiques et les activités biologiques à la structure moléculaire permet, d'une part, d'expliquer l'origine de ces activités/propriétés et, d'autre part, de les prédire pour des molécules dont les données expérimentales ne sont pas disponibles.

L'utilisation du modèle COSMO-RS est particulièrement utile pour établir des modèles QSPR pour la modélisation des équilibres liquide-vapeur, liquide-liquide, mais aussi pour la prédiction des pressions de vapeur des composés purs, enthalpies, entropies de vaporisation, les propriétés physiques, chimiques et toxicologiques des Lis.

La modélisation des propriétés des liquides ioniques repose principalement sur l'utilisation des équations d'état, mais en raison des limites inhérentes de ces équations d'état, d'autres méthodes de modélisation peuvent être envisagées. Parmi ces alternatives, on cite les corrélations empiriques qui sont simples est très souvent efficaces et les méthodes de l'intelligence computationnelle telles les réseaux de neurones artificiels, les machines à vecteur de support, la logique floue qui sont par contre complexes, mais très robustes dans la description des variations non linéaires entre les variables, et présentent aujourd'hui un moyen de modélisation des données très utiles dans plusieurs disciplines.

Une des propriétés très importantes des liquides ioniques, qui a une importance particulière dans le domaine de séparation, est le coefficient d'activité à dilution infinie des différents solutés dans les liquides ioniques. Ce paramètre est très utile dans la caractérisation du comportement des mélanges liquides, l'estimation des solubilités mutuelles, la sélection des solvants pour les processus de distillation extractive, et le calcul des constantes de Henry et le coefficient de partage.

Le présent travail, considéré comme une suite d'un travail antérieur, à pour objectif la modélisation du coefficient d'activité a dilution infinie de l'eau et de quelques solutés organiques dans différents liquide ioniques. Cette contribution est matérialisée par l'élargissement de la base de

données expérimentale, le calcul et l'utilisation des descripteurs issus des profils σ du modèle COSMO-RS, l'utilisation des paramètres d'Abraham de la relation linéaire de l'énergie de solvatation et par l'utilisation de la technique de Least Square Support Vector Machines (LSSVM) comme méthode de modélisation.

Ce mémoire est divisé en 4 chapitres :

- Un premier chapitre qui donne un aperçu général sur les liquides ioniques.
- Un deuxième chapitre nous avons tenté de présenter de manière simple et complète le concept de coefficients d'activité à dilution infinie (CADI).
- Le troisième chapitre introduit les machines à vecteurs de support (Support Vector Machines, SVM).
- Le dernier chapitre présentant les résultats de la modélisation et interprétation de ces résultats.
- Enfin, ce mémoire s'achève sur une conclusion générale et des perspectives.

Chapitre I : Généralités sur les liquides ioniques

I.1.Introduction

Les solvants peuvent être classés en fonction du type de liaisons chimiques qu'ils mettent en jeu (Figure I.1):

- les liquides moléculaires dans lesquels interviennent seulement des liaisons covalentes.
- les liquides ioniques où prédominent les interactions ioniques.
- les liquides atomiques tels que le mercure ou les alliages d'alcalins liquides, siège de liaisons métalliques. (amalgame Na-K liquide à température ambiante)**[1]**

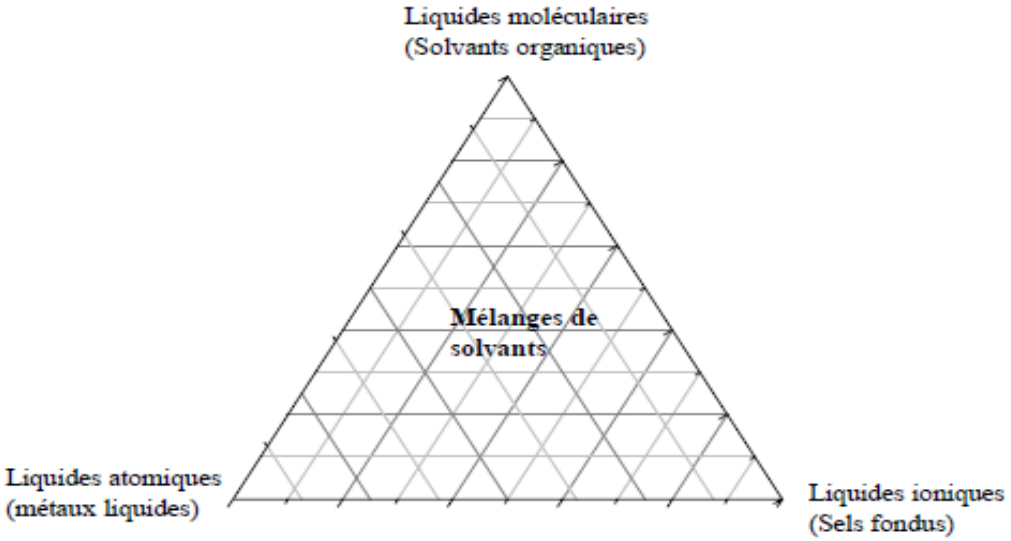


Figure I.1 : Classification des solvants selon le type de liaisons chimiques mis en jeu **[1]**

Les liquides ioniques (LI) sont capables de dissoudre un grand nombre de composés organiques ou inorganiques mais présentent des effets de solvants particuliers par rapport aux solvants organiques classiques et sont par conséquent considérés comme des solvantsnéotériques. Les propriétés physico-chimiques des LI telles que leur grande stabilité thermique, leur bonne conductivité électrique, leur viscosité raisonnable, leur large domaine liquide, leur faible tension de vapeur, leur haute conductivité thermique, en font un milieu réactionnel très intéressant pour la chimie verte permettant notamment de travailler à haute température avec de bonne dispersion de la chaleur.

De plus, les propriétés physiques et chimiques des LI sont modifiables par variation descombinaisons anion/cation et les LI sont souvent considérés comme des solvants modifiables. Cela a permis de concevoir des systèmes adaptés à des applications variées (batteries thermiques, électrochimie, synthèse organique, inorganique et enzymatique, catalyse...) mais également d'ajuster les propriétés physico-chimiques des LI en fonction des besoins.

Après avoir présenté les LI, leurs caractéristiques générales et leurs propriétés den solvants, ce chapitre bibliographique traitera de la problématique de l'utilisation des LI en tantque solvants de réaction catalytique.

La stratégie et la méthodologie des travaux entrepris pour étudier ces systèmes serontégalement exposées. **[1]**

I.2. Définition

Les liquides ioniques sont des composés totalement ioniques qui possèdent un point de fusion inférieur à 100°C. Dans le cas idéal, les liquides ioniques possèdent une tension de vapeur très faible voire non mesurable, cette nature non volatile offre un avantage certain pour la séparation des produits par distillation et évite l'exposition aux vapeurs non contrôlées. Ils sont une stabilité thermique élevée et sont relativement peu coûteux et faciles à synthétiser. Ils offrent une forte solvatation car c'est un milieu non coordinat dans lequel un certain nombre de solutés organiques et inorganiques peuvent être dissous. L'ensemble de ces propriétés en font des milieux de choix pour le développement de la chimie verte [3].

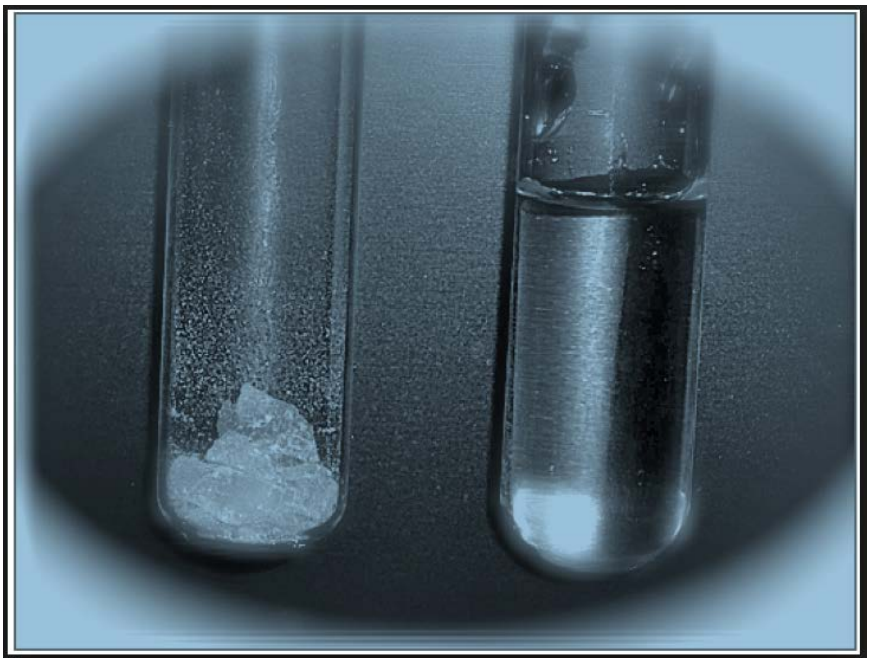


Figure I.2 : Photographie du chlorure de sodium (à gauche) et le liquide ionique [hmim][PF_6] (à droite) à température ambiante,[12]

I.3. Historique

En 1914, le premier liquide ionique à température ambiante synthétisé par Walden est lenitrate d'éthylammonium [EtNH3] + [NO3]- (Walden, 1914). Les premières utilisations des liquides ioniques étaient principalement dans le domaine de l'électrochimie dans les années1970. En 1992, une nouvelle génération de LIs plus stable à l'air et à l'eau est apparue. Depuis 1992, les LIs ont connu une évolution exponentielle en termesde publications. Une large gamme de LIs a été synthétisée incorporant différents anions telsque l'hexafluorophosphate, l'éthanoate, le trifluoroéthanoate, les alkylsulfate,l'hydrogensulfate, le nitrate, le dicyanamide, le trifluorométhanesulfonate ou encore lebis (trifluoromethylsulfonyl) amide. Ses dernières années, un autre groupede LI a été développé ; les LIs fonctionnalisés également appelés liquides ioniques à tâchesspécifiques (TaskSpecificIonicLiquids, TSILs) comportant un groupement fonctionnelattaché par liaison covalente au cation ou l'anion. [6]

I.4. Structure des liquides ioniques

Les liquides ioniques sont des sels organiques liquides se différenciant de l'ensemble des sels fondus par une température de fusion inférieure à 100°C (arbitrairement fixée en référence à la température d'ébullition de l'eau) mais un grand nombre d'entre eux sont liquides à température ambiante, appelée RTIL's («room températureionicliquids»). Ces solvants sont formés par l'association de cations organiques et d'anions;avec des variations presque infinies de structures aussi bien au niveau des anions que des cations. Parmi les cations les plus étudiés, on peut citer les ammoniums quaternaires et des composés aromatiques polycycliques tels que les noyaux imidazolium et pyrrolidinium, tandis que les cations alkylpyridinium, alkylphosphonium ou alkylsulfonium sont moins fréquemment rencontrés (Figure I.2).[3]

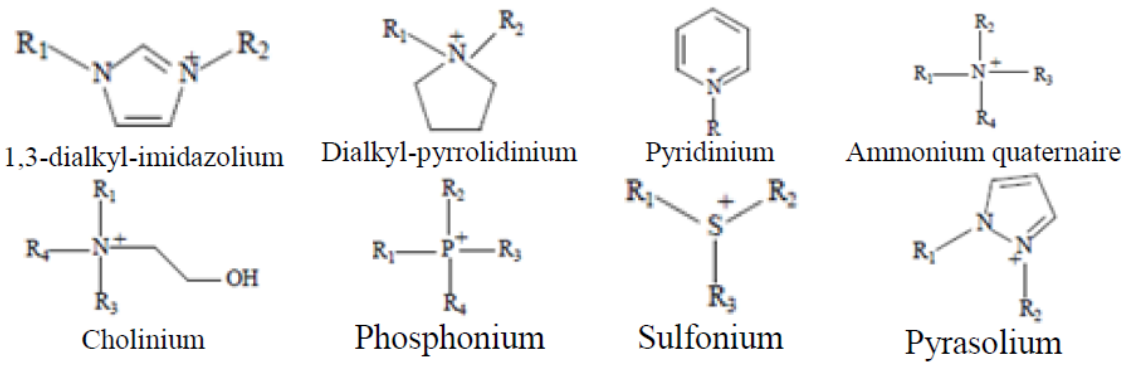


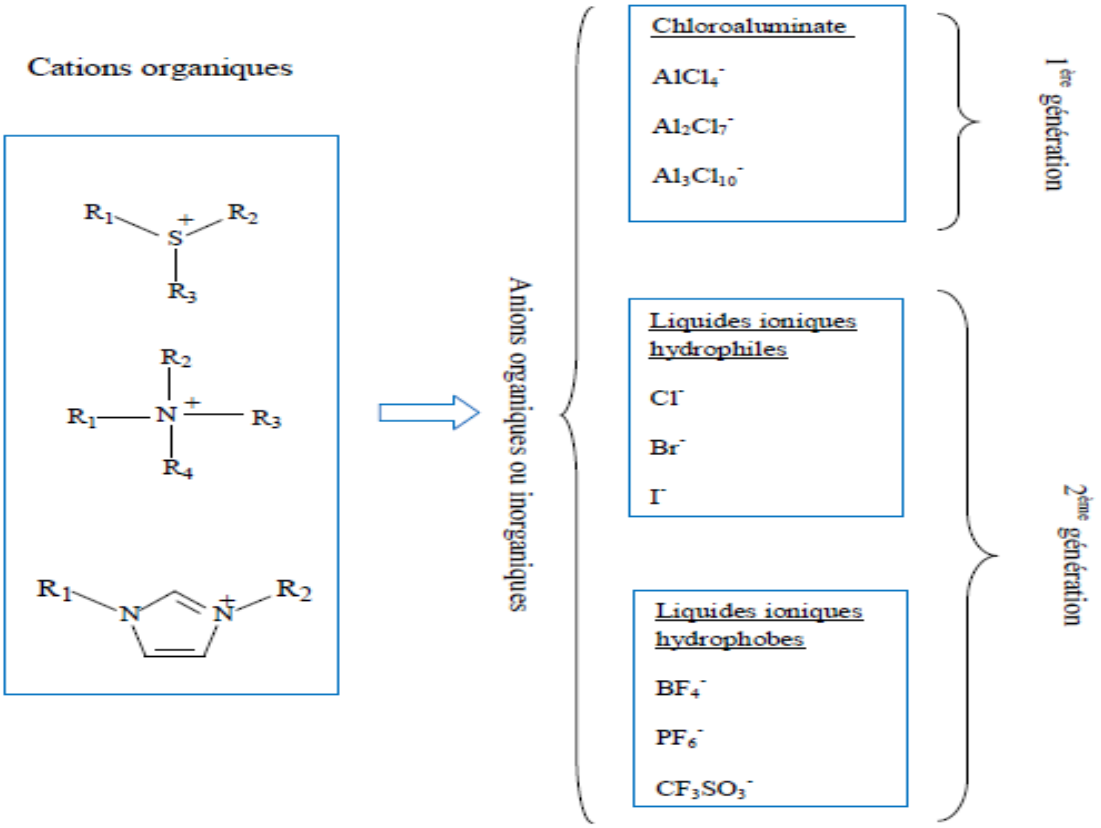
Figure I.3: Structure chimique des principaux cations des LI rencontrés dans la littérature. [3]

- Les contre ions les plus couramment utilisés sont soient des anions inorganiques tels que BF_4^-, PF_6^-, Cl^-, NO_3^-, $AlCl_4^-$, SbF_6^-et autres soient des anions organiques ($CF3SO3^-$, $(CF3SO2)_2N^-$, $CF3CO2^-$et $CH3CO2^-$et autres). Les anions tétrafluoroborate ($BF4^-$) ethexafluorophosphate ($PF6^-$) sont très utilisés en chimie organique ou organométallique car ils confèrent une solubilité recherchée aux espèces ioniques qui sont à la base detrès nombreux sels liquides. Les LI sont regroupés en trois familles: les LI de première génération, de deuxième génération et les LI à tâche spécifique ou de troisièmegénération (Figure I.3).

- Dans la littérature, les liquides ioniques de première génération sont caractérisés par la nature de leurs anions qui sont des chloroaluminates. En général, les liquides ioniques à anion-halogénure sont désignés comme étant ceux de première génération.

- Quant aux liquides ioniques de deuxième génération appelés également liquides ioniques stables à l'air et dans l'eau, ils possèdent des anions qui sont inorganiques ou organiques.

- L'appartenance à la famille des liquides ioniques de troisième génération correspond à une utilisation pour une tâche spécifique plus qu'à la nature des ions qui constituent ces liquides ioniques. En effet, les liquides ioniques à tâche spécifique sont caractérisés par le fait que le cation et/ou l'anion comportent un groupe fonctionnel.[3]



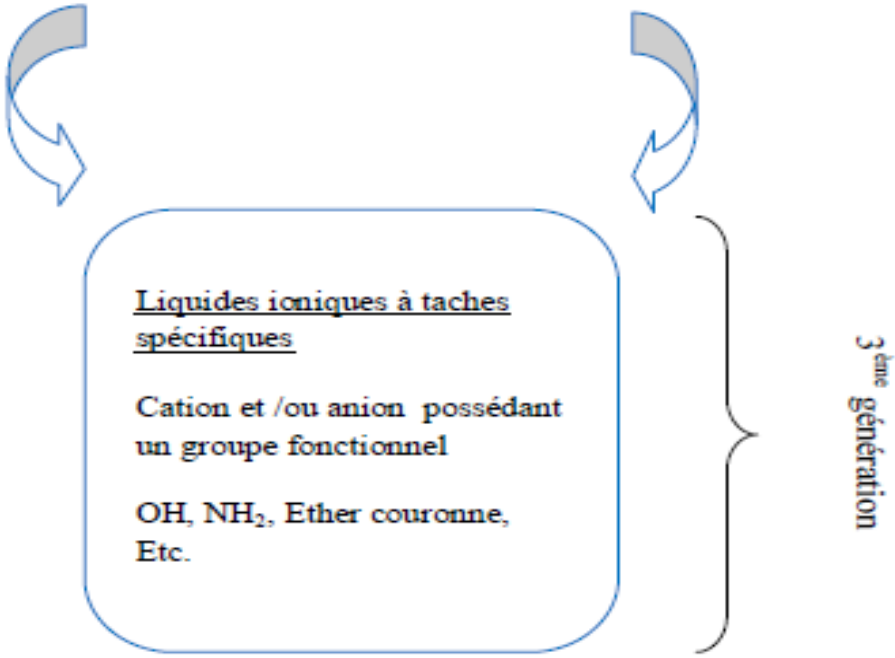


Figure I.4: Différentes classes de liquides ioniques [3].

I.5. Synthèses

Comme de nombreuses associations de cations et d'anions sont possibles, nous avons restreint cette partie bibliographique à la préparation de LI dérivés d'imidazoliums.

La synthèse de ces LI s'effectue en deux étapes :

- Laquaternisation du noyau imidazole
- L'échange d'anion

I.5.1 Réaction de quaternisation

Le principe consiste à quatemiser l'amine d'unalkylimidazole ou l'azote de l'amidine(DBU) en présence d'un halogénoalcane choisi en fonction de la structure ou des propriétésparticulières recherchées pour le liquide ionique final. En effet, il estpossible de choisir le cation parmi des composés aliphatiques ou aromatiques, de jouersur la longueur de la chaîne alkyle attachée (saturée ou insaturée) et de choisir l'anion(simple ou complexe). La réaction de quatemisation des imidazoles ou du DBU peut êtreeffectuée soit par réaction acide-base ou par réaction de nuc1éophile (Figure I.4).

Exemple de synthèse du chlorure de I-butyl-3-méthylimidazolium: On fait réagir1 équivalent de méthylimidazole avec 1,1 équivalent de chloroalcane sous atmosphèred'azote à 70°C pendant 72 heures (la fin de la réaction est donnée par la disparitioncomplète de la phase supérieure (Figure I.5). [10]

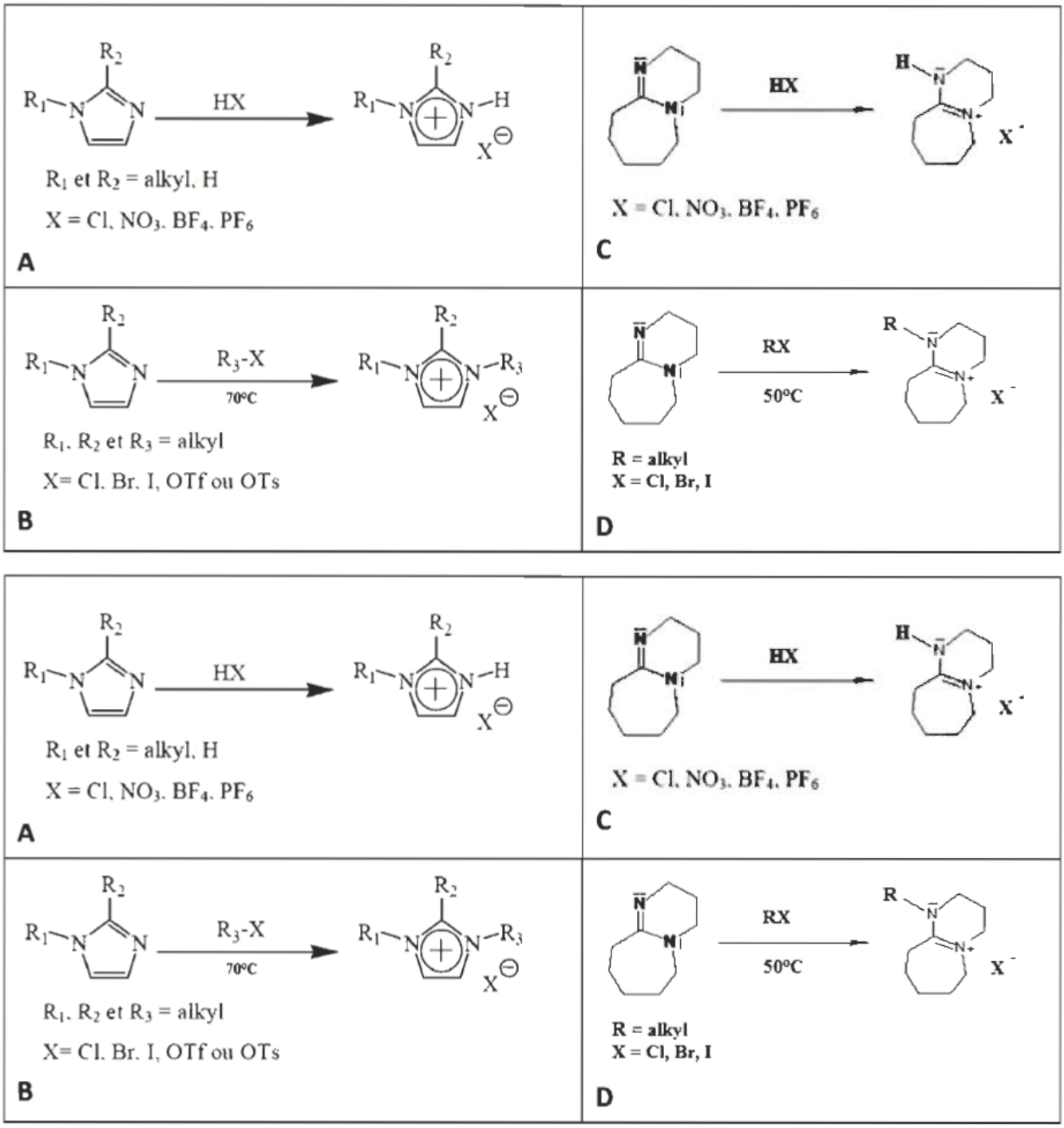


Figure I.5 : Exemples de réaction de quaternisation de l'imidazole (A et B)et du DBU (C et D). [10]



Figure I.6 : Synthèse du chlorure de I-butyl-3-méthylimidazolium par quaternisation. [10]

Après 72 heures, on obtient une solution impure, visqueuse et jaunâtre de [C4Ctimt [Cr] qui est mise au réfrigérateur pendant une nuit. La solution cristallise sous fonne decristaux blancs qui sont ensuite lavés 3 fois pour enlever les impuretés à l'acétated' éthyle. L'acétate d'éthyle est ensuite enlevé en chauffant la solution sous vide à 70°Cpendant environ 4 heures. Le produit obtenu est très hygroscopique. Ce type de réactionest facilitée dans le sens classique des substitutions nucléophiles:

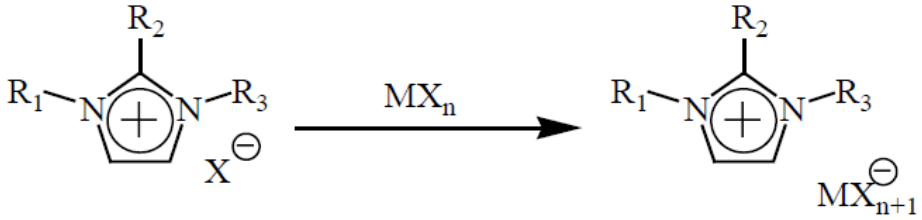
Cl >Br>I. fi estpossible de changer le contre ion du liquide ionique presque autant de fois qu'on ledésire.

Cette modification d'anion intervient en général dans la dernière étape et peutêtre facilement adaptée. Il faut noter cependant que la séparation du nouveau liquideavec le sel issu du changement d'anion est souvent difficile à réaliser. Cela est dû à la solubilité (partielle) de ces sels dans le nouveau liquide ionique [10]

I.5.2 Réaction d’échange de l’anion

La réaction d’échange de l’anion peut se diviser en deux catégories : traitement direct du sel d’imidazolium par un acide de Lewis ou réaction d’échange par métathèse d’anions.

Le traitement d’un halogénure d’imidazolium avec un acide de Lewis MXnconduit à la formation d’un contre-ion métallique. [1]



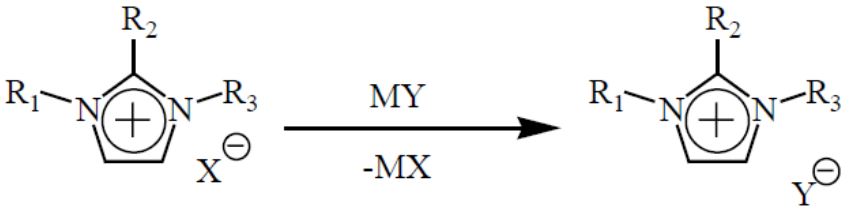
R_1, R_2 et R_3 = alkyl

X= Cl, Br, I

M = Al, Cu, Sn, Fe, Zn

Cette réaction est relativement exothermique et doit être réalisée en conditions anhydres.

-Il est possible de réaliser l'échange de l'anion des sels d'imidazoliums avec un autre sels inorganiques :



R_1, R_2 et R_3 = alkyl

X= Cl, Br, I

MY=$LiNTf_2$, NaOTf, $NaPF_6$, $NaBF_4$

Cette réaction conduit aux LI avec de hauts rendements et une très bonne pureté. L'inconvénient de cette technique est lié à l'échange incomplet des halogénures qui peut conduire à la contamination du LI. Par conséquent, un grand soin doit être apporté lors de la phase de lavage du LI. [1]

Les LI obtenus par ces voies de synthèse sont généralement des liquides incolores bien que les sels d'imidazoliums à base de PF6 ou BF4 puissent présenter une légère coloration jaune. Il est indispensable de caractériser la pureté de ces composés et il est parfois nécessaire de les purifier avant usage.[1]

I.5.3 Purification et pureté des LI

La pureté des LI est un paramètre important pour la plupart des applications car la présence d'impuretés peut gravement affecter les propriétés physico-chimiques des LIetla réactivité des systèmes catalytiques. [1]

Comme les LI sont très peu volatils, la purification par distillation n'est pas envisageable excepté dans des conditions sévères. La purification par chromatographie est également délicate dans la mesure où les LI ont tendance à s'absorber sur les phasesstationnaires. Par conséquent, un grand soin doit être apporté lors de la synthèse des LI afin de limiter tant que possible les impuretés présentes dans le produit final notamment les traces d'eau, d'halogénures et d'imidazole de départ.[1]

Il est possible de mesurer la quantité d'eau présente dans les LI par dosage Karl- Fischer, par spectroscopie infra-rouge,par spectroscopie UV ou par spectrométrie de masse. La détection de traces d'halogénures est généralement réalisée par un test au nitrate d'argent mais ces impuretés peuvent être quantifiées par analyse élémentaire, par chromatographie ionique, ou par électrophorèse capillaire. La présence d'imidazole de départ peut être détectée par un test colorimétrique au chlorure de cuivre(II). [1]

I.5.4 Nomenclature et acronymes des LI

La nomenclature des liquides ioniques varie énormément en fonction des communautés auxquels les articles s'adressent. Par exemple pour 1-éthyl-3- méthylimidazoliumethylsulfate on trouve l'abréviation : [EMIM] [EtSO4], d'autres utilisent le nombre de carbones portés par les différentes chaines alkyles suivant ou précédant l'abréviation du cœur du cation : [C2Mim] [EtSO4].donc il est malheureusement fréquent d'observer plusieurs noms pour le même composé. [2]

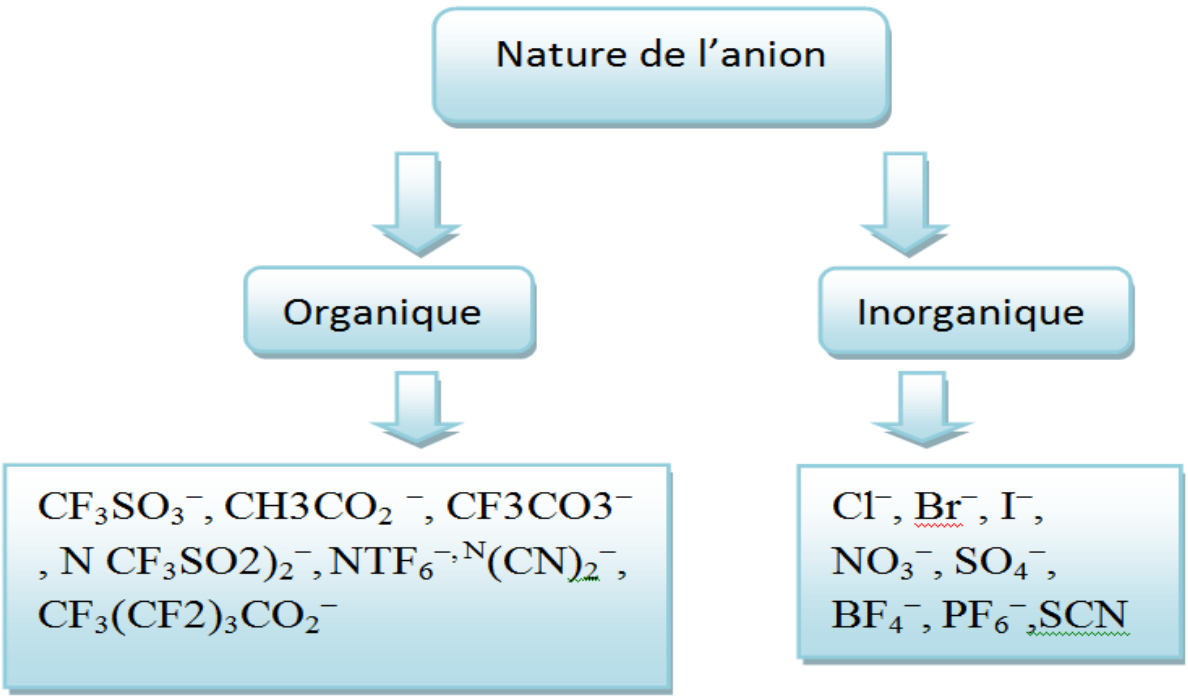


Figure I.7 : les anions les plus utilisés. [2]

I.6.Propriétés uniques des Liquides Ioniques

Des propriétés bien sélectives permettent de distinguer les LI des sels fondus ordinaires. Parmi les propriétés les plus importantes on peut citer les suivantes : Les LI sont caractérises par des températures de fusion plutôt basses. [4]

L'intérêt croissant pour les LIs est démontré par l'explosion bibliographique qui a eu lieu dans ce domaine depuis 15 ans .Ainsi le nombre de publication portant sur les LIs est passé de 50 par an en 1995 pour atteindre 450 par an en 2002, et finalement plus de 2400 en 2009 (figure I.7). Cette croissance remarquable est forcément due à la découverte par les industriels de l'importance des LIs dans beaucoup d'application en raison de leur propriété très remarquable [3].

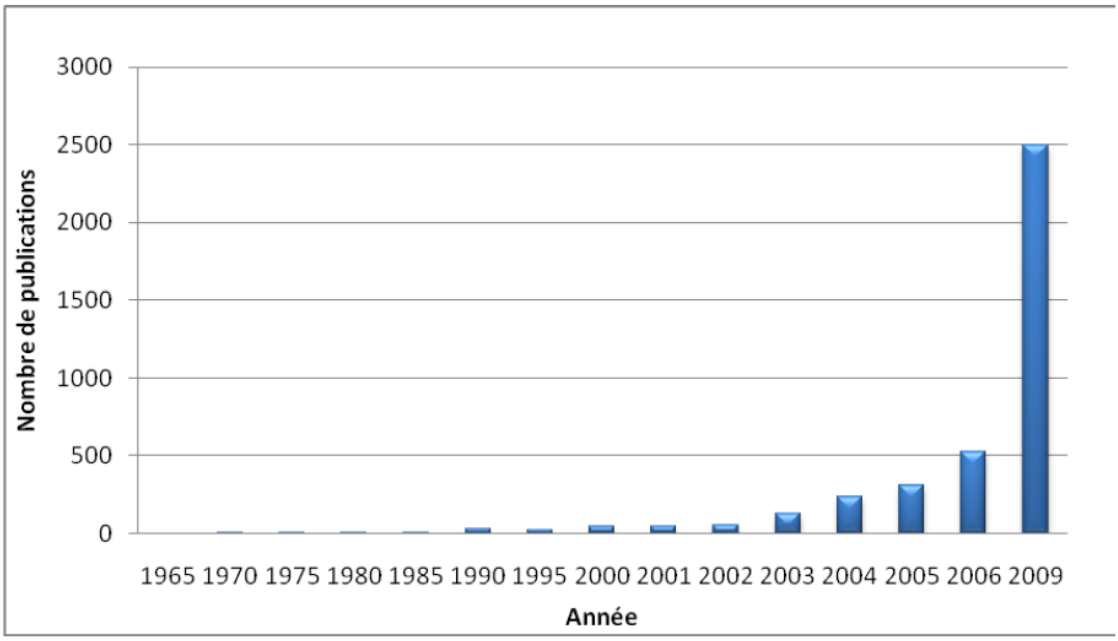


Figure I.8 : Nombre de publications par année portant sur les liquides ioniques [3].

I.7.Propriétés physico-chimiques des liquides ioniques

Les liquides ioniques ont une bonne résistance thermique et leurs tensions de vapeursont très basses et leurs domaines liquides à pression atmosphérique sont plus importantsque ceux de l'eau et des solvants organiques classiques (Figure I.8).

Ils sont ininflanunables. Ils peuvent être utilisés plusieurs fois. On peut aussi ajusterleurs propriétés physico-chimiques en changeant la nature de l'anion ou du cation. Lespremiers travaux sur les propriétés des liquides ioniques purs se sont intéressés sur lacorrélation entre la structure du liquide ionique et les propriétés physico-chimiques. Les impuretés peuvent changer considérablement les propriétés des

liquidesioniques. En effet, la présence d'impureté telle que l'eau peut changer la densité, laviscosité et le point de fusion.[10]

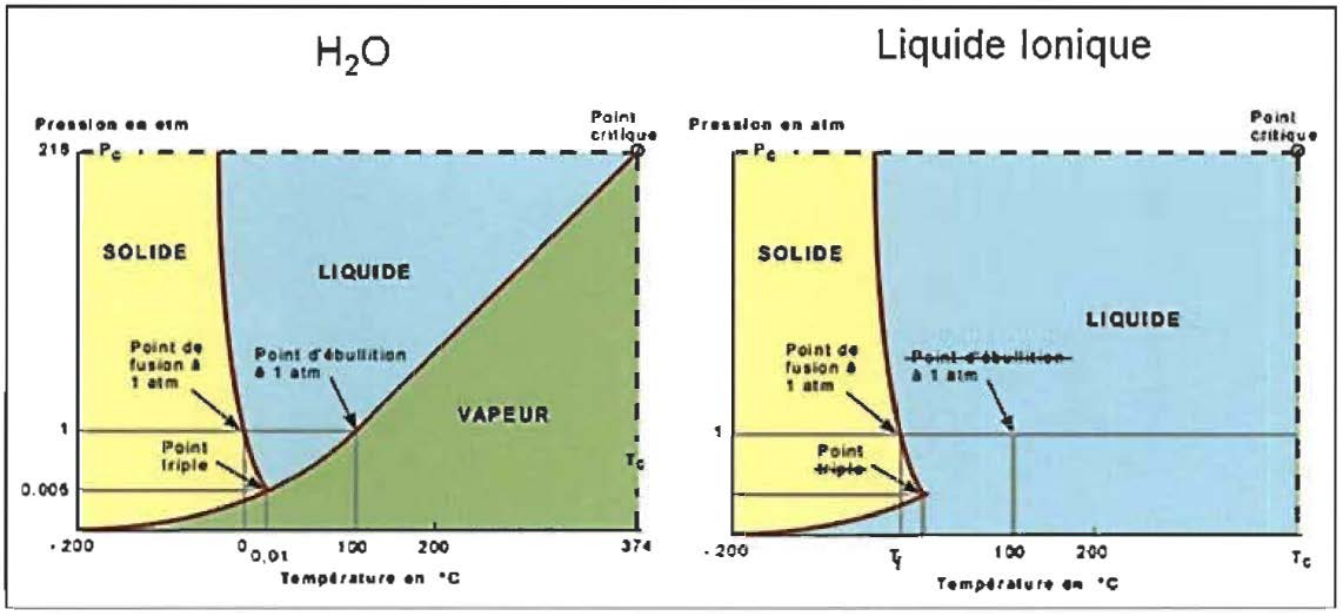


Figure I.9 : Comparaison de l'état de l'eau et des liquides ioniques.[10]

I.7.1. Stabilité thermique et point de fusion

Le point de fusion est un paramètre essentiel pour un liquide ionique, il a été remarqué dans de nombreuses études que sa valeur dépendrait à la fois du cation et de l'anion .Les LIs, caractérisés par des températures de fusion relativement basses et des températures de décomposition élevées (350 – 400°C), possèdent un très large domaine de température dans lequel ils sont à l'état liquide (de l'ordre de 200-300°C) et un domaine de stabilité thermique très grand [3].

I.7.2. La densité et la viscosité

Les LIs sont en général denses et visqueux. Ils ont une viscosité qui peut atteindre dix fois celle des solvants organiques ordinaires. La viscosité des LIs peut également changer avec la composition moléculaire de l'anion, ainsi elle augmente quasi-linéairement avec la langueur de la chaine alkyle du cation.

La densité des LIs est aussi fonction de la nature de l'anion. Elle est aussi une fonction décroissante de la langueur de la chaine alkyle portée par le cation, comme par exemple pour l'imidazole, La densité des LIs est aussi une fonction décroissante de la température [3].

I.7.3 Conductivité

Les liquides ioniques peuvent jouer un rôle de substituant dans les réactions électrochimiques. En effet, ils présentent une large gamme de conductivité de 0.1 à 20 mS.cm-1. Ce sont les liquides ioniques avec un cation imidazolium qui présentent les conductivités les plus importantes. Plusieurs paramètres peuvent affecter cette conductivité comme la taille de l'ion, la densité ou la viscosité [5].

I.7.4. Solubilité

Les LIs possèdent un grand pouvoir solvatant, qui leur permet de solubiliser une large gamme de composés organiques, inorganiques et également organométalliques, et sont liquides dans une large gamme de température. Ces deux dernières propriétés permettent de les envisager facilement comme solvants de réaction.

La plupart des liquides ioniques sont entièrement ou partiellement miscibles avec les solvants organiques polaires (méthanol, acétonitrile, tétrahydrofurane, dichlorométhane, acétone, etc.).
Des systèmes biphasés sont généralement formés avec les solvants organiques de la basse polarité (hexane, toluène, éthers alkyliques, …). [3]

I.7.5.Stabilité chimique

Les liquides ioniques de deuxième génération aussi appelés liquides ioniques « stables à l'air et à l'eau », sont beaucoup plus faciles à utiliser et à stocker. Néanmoins, certains de ces liquides ioniques de deuxième génération, notamment ceux associés à des anions de typesPF_6^-, et BF_4^-, s'hydrolysent partiellement pour former de l'acide fluorhydrique (HF), composé très corrosif et toxique.[3].

I.7.6.Toxicité et biodégradabilité

Les liquides ioniques sont souvent présentés comme des solvants de la chimie verte. Jusqu'à aujourd'hui la toxicité des LIs sont mal connue; ainsi certaines études ont été entreprises pour évaluer quelques propriétés toxicologiques. Des études systématiques, s'intéressant principalement aux LIs à base d'imidazolium, ou ammonium quaternaire ont été entreprises assez récemment.[3]démontrent que plus la chaîne alkyle latérale du cation est longue, plus la toxicité du LI est grande. L'introduction d'une chaine polaire réduit leur toxicité et augmente leur biodégradabilité. En revanche, aucune tendance n'a pu être dégagée quant à l'influence de la nature du cation:imidazolium, pyridinium, ou ammonium quaternaire, même si les cations non-aromatiques paraissent moins toxiques que les cations aromatiques [3].

I.7.7. Volatilité

Les liquides ioniques sont des solvants dont la pression de vapeur saturante est négligeable à température ambiante. Cela permet de les recycler facilement en distillant les autres produits directement à partir du mélange. Cependant cette pression de vapeur saturante est négligeable, et non nulle. Ainsi, dans certains cas précis, il est possible de séparer des liquides ioniques par distillation fractionnée à très faible pression et très haute température. Ces cas sont extrêmement rares et la majorité des liquides ioniques se dégradent au chauffage avant de pouvoir être distillé [3].

I.8. Propriétés spécifiques pour le traitement par extraction

Les solvants organiques moléculaires sont les solvants les plus couramment utilisés dans les procédés de traitement par extraction liquide-liquide de métaux contenus dans une phase aqueuse. Généralement dans un tel système, un extractant est rajouté au solvant pour l'extraction du cation métallique dans la phase organique. Ces systèmes d'extraction utilisent de grands volumes de solvants organiques, qui sont, contrairement aux liquides ioniques, volatils. Des travaux ont montré que l'efficacité d'extraction et la sélectivité de certains systèmes d'extraction utilisant les liquides ioniques étaient parfois supérieures aux systèmes utilisant des solvants organiques. Les liquides ioniques, contrairement aux solvants organiques classiques, ont en effet de très bonnes capacités à solvates les espèces ioniques. [3]

I.9. Les Applications

Les LIs sont utilisés dans de nombreux domaines (Figure I.9) comme l'électrochimie, la synthèse organique, l'extraction, la biocatalyse, la spectroscopie. Jusqu'à présent, leurs utilisations à l'échelle industrielle restent limitées en raison du manque de données expérimentales et du prix élevé de ces solvants [6]

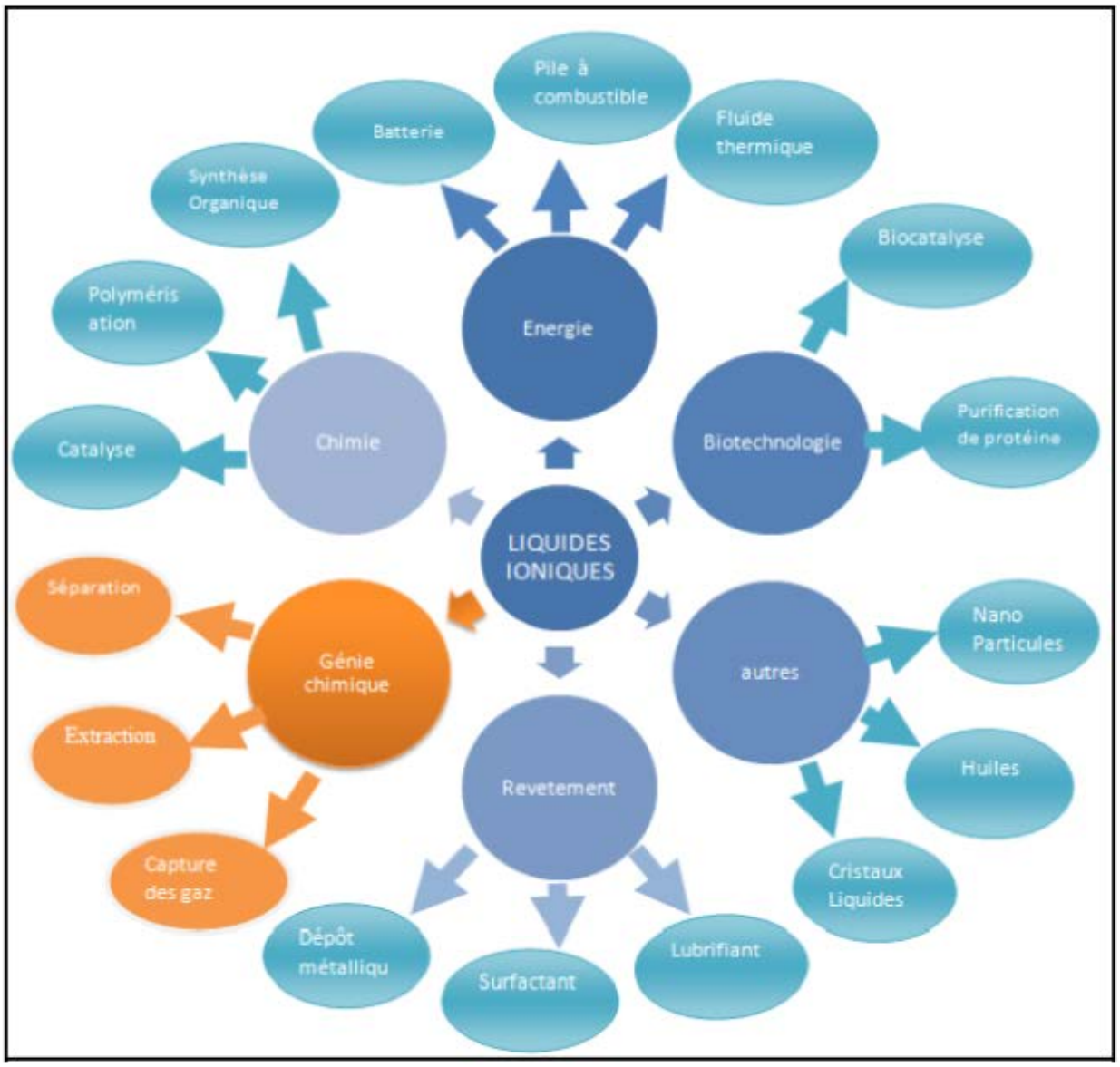


Figure I.9 : Les domaines d'applications des liquides ioniques. [6]

I.9.1. Electrochimie

La fenêtre électrochimique d'un solvant est un paramètre important en électrochimie. Les Lis ont un intérêt comme électrolyte dans les processus et dispositifs électrochimiques à cause de leurs larges fenêtres électrochimiques. Plusieurs Lis ont de larges fenêtres électrochimiques dans la gamme 5-6 V voire plus. Parmi les applications importantes des LIs dans le domaine électrochimique ; on trouve :

- Les batteries Li-ion: les LIs sont utilisées comme électrolytes dans les batteries Li-ion à cause de leur nature non inflammable pour remplacer les électrolytes organiques très volatiles et inflammables.

- Piles à combustible : Certains LIs présentent une conductivité supérieure à 10^{-2}S/cm àtempérature ambiante, valeur proche du niveau requis pour les applications de piles àcombustible. La plupart des investigations sont axés sur le développement de LIs à based'acide de Bronsted, synthétisés par un simple mélange d'amines ternaires et d'acides résultant en LIs avec proton libre. Ces LIs nommés, LIsprotiques ont été étudiés largement pour le transfert de protonsdans les électrolytes des piles à combustible.[6]

I.9.2.Chimie Analytique

De nombreuses applications ont été proposées dans le domaine de la chimie analytique. A ce jour, les liquides ioniques sont principalement utilisés en chromatographie en phase gazeuse et en électrophorèse capillaire (EC). Il reste encore de nombreuses possibilités à explorer dansle domaine de séparations analytiques.[6]

* Chromatographie en phase gazeuse (CPG):

Les LIs sont très étudiés en tant que phase stationnaire par chromatographie inverse en phasegazeuse. Cette technique intéressante est très utile pour étudier les propriétés physicochimiquesdes analytes et les interactions entre solutés et phase stationnaire.ont été les premiers à étudier les comportements desLIs ([BMIM]+[PF6]- et [BMIM]+[Cl]-) avec différents composés organiques. Les résultats ontmontré que chaque LI agissait comme une phase stationnaire faiblement polaire avec lescomposés non polaires et comme une phase polaire avec les composés polaires. Les phases stationnaires de LI permettent d'obtenir une très bonne sélectivité et des séparations plusrapides.

Plusieurs études ont été consacrées à l'utilisation de la CPG pour la détermination ducoefficient d'activité à dilution infinie dans les LIs. Ce paramètrepermet d'évaluer les interactions moléculaires entre le solvant et le soluté.[6]

* Electrophorèse capillaire :

Depuis les années 2000, une vingtaine de publications est sortie dans la littérature sur des applications en électrophorèse capillaire mettant en jeu des LIs. Ces applications qui ont été réalisées en phase aqueuse et non-aqueuse ont permis d'observer de nouveaux typesd'interaction. En effet, les liquides ioniques ont la faculté de solubiliser des composésorganiques et inorganiques et de générer des interactions particulières en solution, ilsprésentent donc un intérêt certain dans la recherche de nouveaux constituants d'électrolytepour l'électrophorèse capillaire. De nombreux travaux ont été effectués d'une part, pourmettre en place une méthode de caractérisation des liquides ioniques, et d'autre part, pourévaluer l'intérêt de ces composés comme nouveaux constituants d'électrolytes. [6]

I.9.3. Synthèse et catalyse

Il existe dans la littérature de nombreuses applications des LIs dans les domaines de synthèseorganique et de la catalyse. Cessolvants peuvent avoir plusieurs fonctions dans la catalyse, comme solvant, solvant et cosolvant, solvant et support, solvant et ligands. Ainsi, il est important de mieux comprendre les propriétés physico-chimiques qui peuvent affecter les réactions catalytiques et de choisir les meilleursLIs pour une réaction donnée.

Les LIs peuvent être utilisés comme milieux réactionnels pour les réactions organocatalysées. Ce potentiel a été démontré dans les réactions Diels-Alder où les Lis peuventformer une liaison hydrogène importante avec les réactifs et peuvent par la suite orienter la sélectivité de la réaction. Un exemple typique qui montre l'influence du LI sur le résultat de la réaction chimique est celui donnée par l'étude de la réaction entre le toluène etl'acide nitrique. En fonction du LI utilisé, les résultats observés montrent, la nitration du toluène dans [BMIM] + [OTf]-, l'oxydation du toluène en l'acide benzoïque dans[BMIM] + [OMs]-, ou halogénation dans [BMIM] +[Cl]-.

D'autre part, les LIs ont reçu une attention particulière comme solvants alternatifs etstabilisants pour la synthèse de nanomatériaux en raison de leur synthèse facile, leur stabilitéet leur faible pression de vapeur. La synthèsede nanoparticule (NP) dans le LI a été décrite en utilisant différentes techniques. Les pluslargement décrites sont la réduction chimique de sels métalliques et de complexesorganométalliques.[6]

I.9.4. Biotechnologie

En raison de leurs propriétés favorables, les LIs ont émergé avec succès le domaine debiotechnologie, comme milieu solvant pour les réactions enzymatiques et les applications à base de protéines. Ils stabilisent les protéines sur une large gammede température. Néanmoins, la stabilité thermique des protéines dépend du choix du LI. Eneffet, Candida antarctica Lipase B (CaLB) s'agrège dans [EMIM]+[N(CN)2]- et perd son activitéfonctionnelle . Par contre, on observe une bonne stabilité pourmonel line dans [BMPY]+[Tf2N]- et pour les enzymes alphachymotrypsinedans [EMIM]+[Tf2N]- et [BTMA]+[Tf2N]- ont également été rapportés .[6]

I.10.Utilisation des liquides ioniques dans les procédésde séparation

Les procédés de séparation sont au cœur de l'industrie chimique dans des domaines variés tels que la pétrochimie, l'hydrométallurgie, les industries pharmaceutique, agroalimentaireou encore nucléaire.

Grâce à leurs propriétés, les liquides ioniques sont envisagéspour deux techniques de séparation : la distillation extractive et l'extraction liquide [7]

I.10.1 Distillation extractive

La distillation extractive est un procédé de distillation qui permet de séparer, danscertains cas, les constituants d'un mélange ayant des points d'ébullition voisins ou formantun azéotrope. Un tiers corps, appelé entraîneur, est ajouté au mélange pour modifierla volatilité relative d'un des constituants. Cet entraîneur doit être sélectif vis-à-vis dusystème à séparer, peu volatil et il ne doit pas former d'azéotrope avec les constituants dumélange à traiter. Lorsqu'on ne connaît pas la totalité du diagramme de phases liquide vapeurdu mélange avec l'entraîneur, la sélectivité de ce dernier est calculée à partir descoefficients d'activité à dilution infinie des constituants du mélange.

De nombreuses études sur la détermination des coefficients d'activité à dilution infinie decomposés organiques dans les liquides ioniques ont montré que ces solvants présentent dessélectivités élevées pour de nombreux problèmes de séparation. L'utilisation deliquides ioniques pour la distillation extractive a déjà été étudiée dans la littérature. En effet, les liquides ioniques présentent de nombreux avantages pour la distillationextractive : ils ne sont pas volatils, ils sont capables de dissoudre une large gamme decomposés organiques, ils sont peu visqueux aux températures d'opération et ils sont peu corrosifs par rapport aux autres solvants. L'utilisation des liquides ioniques dans ce type deprocédés doit donc être envisagée, cependant l'étude de leurs propriétés physico-chimiqueset thermodynamiques reste nécessaire pour le dimensionnement du procédé. [7]

I.10.2 Extraction liquide-liquide

Les liquides ioniques sont également considérés pour l'extraction liquide-liquide de solutésd'intérêt dans diverses applications. Celles-ci se divisent essentiellement en troiscatégories [7]:

•Extraction des alcools pour la synthèse des biocarburants

Les biocarburants sont envisagés comme une alternative aux carburants d'origine fossile. Les alcools : méthanol, éthanol, 1-propanol et 1-butanol peuvent être produitsà partir de la biomasse par fermentation et sont donc considérés comme debons candidats pour ce type de carburant. Lors de sa production, l'alcool synthétisése trouve en mélange avec de l'eau et d'autres composés organiques.

Desprocédés de séparation sont donc nécessaires afin d'extraire l'alcool. La distillationest la technique généralement utilisée, mais elle est très coûteuse énergétiquementet recquiert une installation lourde

pour arriver à la pureté désirée. L'extractionliquide-liquide à l'aide de liquides ioniques est donc proposée comme procédéalternatif. Pour cela, des liquides ioniques hydrophobes constitués généralementde l'anion [Tf_2N] ou [PF_6] sont utilisés. L'étude du système {1-butanol +eau + [HMIM][eFAP]} montre que ce liquide ionique présente une très bonnesélectivité (de l'ordre de 300) ainsi qu'un important coefficient de partage pour le1-butanol. On peut donc envisager ce solvant pour ce type d'applications.[7]

•Extraction de composés aromatiques des vapeurs issues de procédés deraffinage

La séparation des composés aromatiques (benzène, toluène, éthylbenzène ou xylène) des hydrocarbures aliphatiques (butane au décane) est difficile car ces composés ont des températures d'ébullitions voisines et certains mélanges forment des azéotropes. L'un des procédés utilisés pour la séparation de ces mélanges est l'extraction liquide -liquideavec des solvants organiques polaires comme le sulfolane. Cependant, la purification et la régénération du sulfolane nécessitent une importante quantitéd'énergie. L'utilisation des liquides ioniques comme solvant de remplacement a doncété largement étudiée dans la littérature. La solubilité des n-alcanes est très faible dans les liquides ioniques, ce qui implique une sélectivité élevée.Ont démontré que le liquide ionique [EMIM][Tf2N] est plus performant que le sulfolanepour la récupération des composés aromatiques .[7]

•Désulfuration des carburants

La combustion des carburants est la source principale d'émissions de dioxyde desoufre, responsable des pluies acides. Une directive de l'Union Européenne a récemment fixé la teneur en soufre des carburants pétroliers : celle-ci ne doit pas excéder10 ppm. De telles spécifications imposent de réaliser une désulfuration de plus enplus poussée des produits pétroliers par des procédés performants et économiques. L'extraction des composés organiques soufrés (thiophène, dibenzothiophène) à l'aidede liquides ioniques a été proposée pour la première fois .etde nombreuses études ont suivi. Dans ces travaux, l'influence de la nature duliquide ionique sur l'efficacité de l'extraction a été examinée. D'après les résultatsobtenus, le choix du cation semble être plus important que le choix de l'anion pouroptimiser la sélectivité. A titre d'exemple, les liquides ioniques constitués du cationalkylméthylpyridinium présentent des taux d'extraction pour les composés soufréssupérieurs à 70 %. Ces cations couplés aux anions thiocyanate ou éthanoate offrent les plus grandes performances d'extraction.

Dans ces travaux, l'étude expérimentale des équilibres liquide-liquide de systèmes ternairesest réalisée afin d'évaluer thermodynamiquement la capacité du liquide ionique àextraire le soluté désiré. La modélisation de ces données est indispensable pour le dimensionnementet l'optimisation des procédés. La plupart des données expérimentales sontcorrélées à l'aide de modèles thermodynamiques

classiques de coefficients d'activités : NRTL, UNIQUAC ou électrolyte-NRTL. Ont démontréque ces modèles peuvent être utilisés pour prédire des équilibres liquide-liquide de systèmes ternaires, néanmoins ils doivent être améliorés. Des approches purement prédictivesont également été envisagées. Par exemple, le modèle COSMO-RS, qui consiste à caractériserle comportement d'un mélange à partir d'une description quantique des moléculesqui le composent, a déjà été utilisé pour prédire les équilibres liquide-liquide de systèmes ternaires impliquant les liquides ioniques mais les résultats ne sont pas satisfaisants. [7]

I.11. Les Avantages des liquides ioniques

Les liquides ioniques présentent :

-Un point de fusion variable en fonction des ions choisis **;**

-Une viscosité modulable (similaire à de l'huile d'olive) ;

-Une grande stabilité thermique (stables à haute température jusqu'à 200 °C ou 400 °C) ;

-Une grande stabilité chimique ;

-Une faible pression de vapeur saturante ;

-Une large fenêtre électrochimique ;

-Une polarité modulable (comprise entre l'acétonitrile et l'eau) ;

-Ininflammables (moins de contraintes pour la manipulation et le stockage) ;

-Hydrophiles et hydrophobes ;

-Bons conducteurs ;

-Mécaniquement stables vis-à-vis de l'eau et de l'oxygène ;

-Grand pouvoir de solvatation ;

-Meilleur rendement suivant les réactions ;

-Meilleure sélectivité en diminuant les produits secondaires suivant les réactions. Une solubilitéadaptable [11]

I.12.Conclusion

Dans ce chapitre, nous avons tenté de présenter de manière simple et complète leconcept les LIs sont des liquides formés généralement par un cation organique volumineux etun anion organique ou inorganique et ayant une température de fusion inférieure à 100°C malgré tout les Lis présentent des avantages:

-Un point de fusion variable en fonction des ions choisis.

-Une viscosité modulable (similaire à l'huile d'olive).

-Une grande stabilité thermique.

-Une grande stabilité chimique.

-Une faible pression de vapeur saturante.

-Un liquide ionique qui ne se dégrade pas sous les conditions de travail.

-Une solubilité adaptable.

Ces Propriétés ont éveille un grand intérêt du à leurs applications Potentielles dans desnombreux domaines industriels. Parmi ces applications, les procédés des séparations sont au cœur de l'industrie chimique tel que la distillation extractive et l'extraction liquide-liquide. [2]

Chapitre II : Coefficients d'activité à la dilution infinie(IDAC)

II.1.Introduction :

Le coefficient d'activité de dilution infinie est également appelé coefficient d'activité limitant. Ce nombre est une constante pour différents types de solutions et est utilisé pour calculer quand une solution atteindra une dilution infinie. Les coefficients de dilution infinie ont été calculés pour différents types de solutions et sont disponibles pour les scientifiques. La recherche publiée a décrit une méthode pour déterminer ces valeurs. Les chercheurs ont utilisé la chromatographie liquide gazeuse, une méthode qui utilise du papier ou des gaz pour absorber et séparer les gaz, les liquides et d'autres substances. De cette manière, les coefficients de dilution infinie pour un certain nombre de gaz aromatiques ont été déterminés. [13]

II.2. Coefficients d'activité de dilution infinie

II.2.1. Définition

Le coefficient d'activité γ_i de l'espèce i dans une solution est défini par le rapport suivant:

$$\gamma_i = \frac{\hat{f}_i}{x_i f_i} \qquad (2\text{-}1)$$

Où, $\hat{f}$ est la fugacité du composant i en solution, de sa fugacité pure et de sa fraction molaire. Le coefficient d'activité est également défini comme la constante de proportionnalité entre l'activité d'une substance et sa fraction molaire dans une solution donnée.

$$a_i = \gamma_i x_i \qquad (2\text{-}2)$$

Dans l'étude des équilibres de phase et donc la conception et l'optimisation de la séparation processus, le concept de coefficient d'activité est utilisé plutôt que d'activité [14].

II.2.2. Importance et utilisation des données de coefficient d'activité de dilution infinie

Le coefficient d'activité de dilution infinie γ_i^{∞}, aussi appelé coefficient d'activité limitant est la valeur limite du coefficient d'activité d'un soluté lorsque sa concentration tend vers zéro. a défini la région de dilution infinie comme une où chaque molécule du soluté est entouré de molécules du solvant seulement. Seules les interactions soluté-solvant ont lieu. Ce peut être supposé pour les fractions molaires comprises entre 10^{-7} et 10^{-4}, selon les composants sous enquête. La région de dilution infinie des solutions a attiré beaucoup d'intérêt pour le couple des raisons qui en soulignent la grande importance :

a) Le comportement du système dans les régions très diluées contribue à l'obtention de produits de pureté;

b) L'étape la plus difficile et la plus coûteuse d'un processus de séparation est la suppression du dernier des traces d'impureté;

c) Le plus grand écart par rapport à l'idéalité se produit dans les régions très diluées;

d) Les préoccupations environnementales sont basées sur les régions très diluées dans la phase gazeuse qui est l'air.

Discute des applications théoriques et pratiques de l'activité de dilution infinie valeurs de coefficient (IDAC) dans l'ingénierie chimique et environnementale. Les plus importants qui sont liés à l'ingénierie chimique sont énumérés ci-dessous [14]:

a) Synthèse, conception et optimisation des processus de séparation;

b) Calcul des constantes de la loi de Henry;

c) Prédire l'existence d'un azéotrope;

d) Développement de méthodes prédictives pour G^E

e) Caractérisation du comportement pour les mélanges liquides.

II.2.2.1. Synthèse, conception et optimisation des processus de séparation

La séparation est une étape critique dans la plupart des processus chimiques en raison de son impact sensible sur le coût global, ainsi que la qualité du produit. L'une des applications les plus importantes de Les IDAC sont trouvés dans les processus de séparation. Les valeurs IDAC sont utilisées dans la sélection préliminaire de solvant en vue de la distillation extractive, ainsi que l'extraction liquide-liquide.

Le meilleur agent de séparation, appelé entraineur, est celui qui permet la meilleure sélectivité et les valeurs de capacité. En d'autres termes, lorsque les solvants sont étudiés pour trouver le plus approprié agent de séparation pour un mélange contenant deux composants i et j, leur IDAC respectif les valeurs dans chaque solvant sont déterminées. L'équation (2-3) est ensuite utilisée pour calculer la sélectivité à dilution infinie β^{∞}_{ij} de chaque solvant par rapport au système i / j.

$$\beta^{\infty}_{ij} = \frac{\gamma^{\infty}_{is}}{\gamma^{\infty}_{js}} \qquad (2\text{-}3)$$

Est γ^{∞}_{is} et γ^{∞}_{js} représentent les coefficients d'activité limite de i et j respectivement dans le solvant, s. β^{∞}_{ij} est la sélectivité limite des solvants s pour le système constitué des composants i et j. Ainsi, les valeurs de sélectivité à dilution infinie sont importantes pour une sélection préliminaire d'un entraineur. Une deuxième quantité à regarder est la capacité du solvant à dilution infinie, k^{∞}_{j} qui est numériquement la quantité maximale d'espèce j qui peut être dissoute dans le solvant.

$$k_j^{\infty} = \frac{1}{\gamma_j^{\infty}} \qquad (2\text{-}4)$$

De plus, la capacité révèle la quantité de solvant utilisée pendant le processus de séparation. Le plus la capacité de solvant est grande, plus la quantité de solvant utilisée est faible. En principe, le solvant à L'utilisation comme agent de séparation est supposée avoir la plus grande valeur de sélectivité et de capacité. Mais dans la pratique, un compromis doit être trouvé, puisque ces deux quantités sont en variation inverse l'un par rapport à l'autre. Un critère qui intègre à la fois la sélectivité et la capacité est appelé indice de performance, PI donné par le produit de ces propriétés. Ainsi, à dilution infinie,

$$PI^{\infty} = \beta_{ij}^{\infty} . k_j^{\infty} \qquad (2\text{-}5)$$

La détermination de γ^{∞} permet d'évaluer l'utilisation potentielle de solvants. Par conséquent, il sert de critère de présélection. Autres critères conduisant au choix définitif d'un solvant comprennent: récupérabilité, coefficient de distribution, solubilité du solvant, densité, tension interfaciale, viscosité, coût, corrosivité, ininflammabilité, faible toxicité, etc .[14]

II.2.2.2. Calcul des constantes de la loi de Henry

Il y a une relation entre la constante de la loi de Henry d'un soluté dans un solvant et son coefficient d'activité limitant:

$$\gamma_i^{\infty} = \frac{H_i}{P_i^*} \qquad (2\text{-}6)$$

H_i représente la constante de la loi de Henry pour soluté i dans le solvant et P_i^* est la pression de vapeur de soluté à la température pour laquelle l'IDAC est déterminé. L'équation (2-6) montre que Les constantes de loi de Henry peuvent être trouvées à travers les valeurs de l'IDAC. Dans les cas où la température du système est au-dessus de la température critique du soluté, la constante de la loi de Henry, plutôt que la L'IDAC est utile. Calculs de conception pour les applications de séparation de gaz couramment impliquer l'utilisation des constantes de la loi de Henry [14].

II.2.2.3. Prédire l'existence d'un azéotrope

Les conditions d'apparition d'un azéotrope dans les systèmes binaires peuvent être établies en utilisant données sur le coefficient d'activité de dilution. Pour un système montrant déviation positive de la loi de Raoult, un azéotrope avec un minimum de température (pressure-maximum) a lieu quand

$$\ln \gamma_j^{\infty} > \frac{\ln P_i^{sat}}{\ln P_j^{sat}} > \ln \gamma_i^{\infty} \qquad (2\text{-}7)$$

Pour un système présentant une déviation négative de la loi de Raoult, un azéotrope avec un maximum de température (pression-minimum) a lieu lorsque:

$$\ln \gamma_j^{\infty} < \frac{\ln P_i^{sat}}{\ln P_j^{sat}} < \ln \gamma_i^{\infty} \qquad (2\text{-}8)$$

A partir des équations 2-7 et 2-8, il apparaît évidemment que la connaissance des valeurs IDAC peut aider à prédire les conditions menant à un mélange azéotropique. [14]

II.2.2.4. Développement de méthodes prédictives pour G^E

Les valeurs du coefficient d'activité de dilution infinie sont importantes pour le développement de nouvelles modèles thermodynamiques, l'ajustement des paramètres, ainsi que l'expansion de applicabilité aux modèles de contribution de groupe existants .A l'aide d'une corrélation appropriée, on peut utiliser un coefficient d'activité de dilution infinie valeurs à venir avec des données VLE dans la gamme de concentration intermédiaire. De plus, il est recommandé d'adapter simultanément les paramètres d'interaction de groupe à tous les VLE, LLE et données expérimentales sur le coefficient d'activité de dilution infinie [14].

II.2.2.5. Caractérisation du comportement pour les mélanges liquides

La valeur du coefficient d'activité de dilution infinie est une indication de la nature des interactions entre le soluté et les molécules de solvant. Une valeur IDAC élevée est synonyme d'interactions faible soluté-solvant, c'est-à-dire faible solubilité. L'IDAC La valeur est

également une mesure limitante de la non-réalité d'une solution. Pour un système idéal, $\gamma^{\infty}_i=1$ considérant qu'un écart par rapport à l'idéalité correspond aux valeurs IDAC plus petit que l'unité [14].

II.2.3. Dépendance de la température du coefficient d'activité

L'équation de Gibbs-Helmholtz (2-9) montre l'effet du changement de température sur l'activité coefficient.

$$\left[\frac{\partial\left(G^E/RT\right)}{\partial T}\right]_{P,x}=-\frac{H^E}{RT^2} \qquad (2\text{-}9)$$

Cela conduit aux relations significatives suivantes:

$$\left(\frac{\partial \ln \gamma_i}{\partial T}\right)_{P,x}=-\frac{H_i^E}{RT^2} \qquad (2\text{-}10)$$

$$\left(\frac{\partial \ln \gamma_i^{\infty}}{\partial(1/T)}\right)_{P,x}=\frac{H_i^{E,\infty}}{R} \qquad (2\text{-}11)$$

En utilisant les données γ^{∞} expérimentales, l'enthalpie partielle excédentaire à dilution infinie pour un soluté donné i, $H_i^{E,\infty}$ peut être directement obtenu à partir de la pente d'une droite issue de l'équation (2-11), où R est la constante de gaz. Il ne faut qu'une approximation expérimentale γ_i^{∞} qui est obtenue le régression linéaire,

$$\ln \gamma_i^{\infty}=A+\frac{B}{T} \qquad (2\text{-}12)$$

II.2.4. Modèles de coefficients d'activité prédictive

Les modèles thermodynamiques prédictifs sont très importants dans les processus de séparation elles permettent:

a) Une sélection rapide de solvants à coûts réduits ;

b) accéder à des données difficiles ou impossibles à obtenir expérimentalement;

c) Le développement de simulateurs de processus.

Il est au-delà de la portée de cette étude d'examiner tous les modèles thermodynamiques utilisés pour prédire coefficients d'activité de dilution infinie. Seuls les modèles les plus prometteurs utilisés pour les systèmes impliquant des liquides ioniques sont traités dans cette section. Les modèles les plus réussis utilisés par divers chercheurs pour prédire les coefficients d'activité de divers solutés dans les liquides ioniques peuvent être classés en deux catégories [14]:

a) les méthodes de contribution de groupe;

b) Méthodes basées sur COSMO.

II.2.4.1 Méthodes de contribution de groupe. (GCMs)

Ces méthodes sont basées sur le concept de contribution de groupe qui suppose que l'interaction l'énergie d'un système est la somme des énergies d'interaction des groupes fonctionnels, à condition correctement défini. Les molécules sont donc divisées en plus petites structures appelées groupes auxquels des paramètres sont affectés. Ces modèles de coefficients d'activité sont principalement basés sur la relation entre ce dernier et l'énergie libre excédentaire [14].

II.2.4.1.1.Le modèle ASOG (solution analytique du groupe)

Le modèle ASOG (solution analytique du groupe) a été développé qui a exploité des travaux antérieurs. Il a été redéfini qui ont modifié plusieurs paramètres pour tenir compte de l'effet des changements de température. Dans cette méthode, l'expression du coefficient d'activité est composée de deux parties: le terme de contribution de taille et celui d'interaction de groupe. ont utilisé le modèle ASOG pour corréler les données expérimentales du coefficient d'activité de dilution infinie du n-octane, du n-nonane et du n-décane dans le tétrafluoroborate de 4-méthyl-n-butylpyridinium liquide ionique dans la plage de température de 297 à 344 K. les résultats étaient en bon accord avec les mesures expérimentales. Douze paramètres d'interaction liés à trois groupes structuraux, CH2, pyridinium et BF4 ont été déterminés [14].

II.2.4.1.2 Le modèle UNIFAC et ses modifications

Après sa publication en (1975), le UNIFAC (Uniquac Quasi-Chemical Functional Activity Coefficient) a subi une pléthore de révisions et d'extensions qui sont décrites dans les revues. L'UNIFAC original, comme l'ont indiqué présentait les avantages suivants:

a) disponibilité par le biais de simulateurs de procédés commerciaux;

b) Fiabilité dans la prédiction des résultats VLE;

c) Applicabilité sur une large gamme de mélanges.

Les mêmes auteurs soulignent ses faiblesses comme suit:

a) Description incorrecte de la dépendance en température du coefficient d'activité et d'autres propriétés apparentées;

b) Résultats insatisfaisants lors de la prédiction des propriétés dans la région diluée, ainsi que pour les systèmes asymétriques.

Pour tenter de remédier à ces limitations, des versions plus avancées du modèle UNIFAC ont été développées, les plus importantes étant UNIFAC modifié (Dortmund) diffère du modèle UNIFAC original en ce qu'il utilise une partie combinatoire différente et prend en compte de manière fiable la dépendance en température des paramètres d'interaction de groupe. Ses développeurs ont également introduit de nouvelles quantités de van der Waals. Aucun modèle thermodynamique ne peut revendiquer autant de popularité que l'UNIFAC (Do). Ceci est attribué à la quantité de travaux de recherche et de publications qui s'y rapportent au cours des vingt dernières années. Récemment, Le groupe du professeur Gmehling a commencé à évaluer la capacité prédictive d'UNIFAC (Do) pour les systèmes impliquant des liquides ioniques. La matrice des paramètres de ce modèle a été étendue, permettant ainsi prédiction des coefficients d'activité de dilution infinie de:

a) n-alcanes et alk-1-ènes dans [BMPyr] [BTI], [HMPyr] [BTI] et [OMPyr] [BTI];

b) Méthylcyclohexane et toluène dans [HMIM] [BTI];

c) les n-alcanes, le benzène et le toluène dans [MMIM] [BTI] [BMIM] [BTI] [HMIM] [BTI], ainsi que le benzène et le toluène dans [BMIM] [BTI];

d) les n-alcanes, les alk-1-ènes, les cycloalcanes, les cycloalcènes, les aromatiques, les alcools, les cétones, les esters, les éthers et l'eau dans [HMIM] [BTI], [OMIM] [BTI], [BMPyr] [BTI];

e) Le mélange binaire de [BMIM] [BTI] et [EMIM] [BTI] .Il est clair que la matrice des paramètres reste très limitée quand on regarde le nombre d'ions impliqué dans la structure des liquides ioniques. La principale faiblesse des MCG est leur incapacité à traiter avec les isomères, les effets de proximité et les systèmes impliquant des groupes pour lesquels les données expérimentales ont jamais été compilé [14].

II.2.4.2. Méthodes basées sur COSMO

Le modèle de dépistage de type conducteur (COSMO) est une méthode prédictive pour la thermophysique propriétés utilisant une approche chimique quantique. Ses deux variantes les plus utilisées pour l'infini Les coefficients d'activité de dilution dans les liquides ioniques sont COSMO-RS (le criblage de type conducteur). Modèle pour les solvants réels), développé par Klamt (1995) et COSMO-SAC (The conducator-like modèle de criblage, coefficients d'activité de solvatation). Ces deux modèles impliquer la thermodynamique statistique, ainsi que le calcul chimique quantique. Ils nécessitent σ -profils de molécules de la même manière que les méthodes de contribution de groupe compter sur les paramètres d'interaction de groupe. Un σ -profil est défini par Klamt et Eckert (2000) comme –a fonction de distribution de la quantité de surface ayant une densité de charge de blindage entre σ et σ+ dσ". Les modèles COSMO sont paramétrés pour 9 éléments jusqu'à présent: hydrogène, carbone, l'azote, l'oxygène, le fluor, le chlore, le brome, l'iode et le soufre. Ils comptent sur un très petit nombre de paramètres qui ne sont pas spécifiques aux molécules types ou groupes fonctionnels. Cela représente un grand avantage par rapport aux GCM qui impliquer un grand nombre de groupes fonctionnels. Le temps de calcul long et les coûts de calcul élevés sont les principaux inconvénients de ces modèles COSMO. Dans un document discutant des comparaisons entre COSMO-RS et GCMs, en tenant compte d'un autre article à laquelle Klamt a lui-même contribué (Putnam et al., 2004) révèlent ce qui suit:

a) COSMO-RS obtient de mauvais résultats dans les limites d'applicabilité des GCM, plus en particulier, UNIFAC (Do);

b) COSMO-RS est recommandé en cas de paramètres GCM inconnus, isomères et effets de proximité.

Ce document a suscité une forte réaction du développeur de COSMO-RS qui trouvé ses conclusions biaisées afin de favoriser clairement l'UNIFAC modifié (Do). De point de vue neutre, il faut plus de temps pour confirmer si COSMO-RS est inférieure supérieur aux GCM. Plus important encore, des extensions sont en cours pour améliorer les deux modèles. Coefficients d'activité de dilution infinie d'un large spectre de solutés certains liquides ioniques ont été prédits en utilisant COSMO-RS. Dans une étude sur l'élimination des thiophène du gazole, COSMO-RS prédictions de l'infini coefficient d'activité de dilution des alcanes, alk-1-ènes et cycloalcanes dans [HMIM] [PF6], [HMIM] [BF4], [MOIM] [Cl] et [EMIM] [Tf_2N] sont signalés

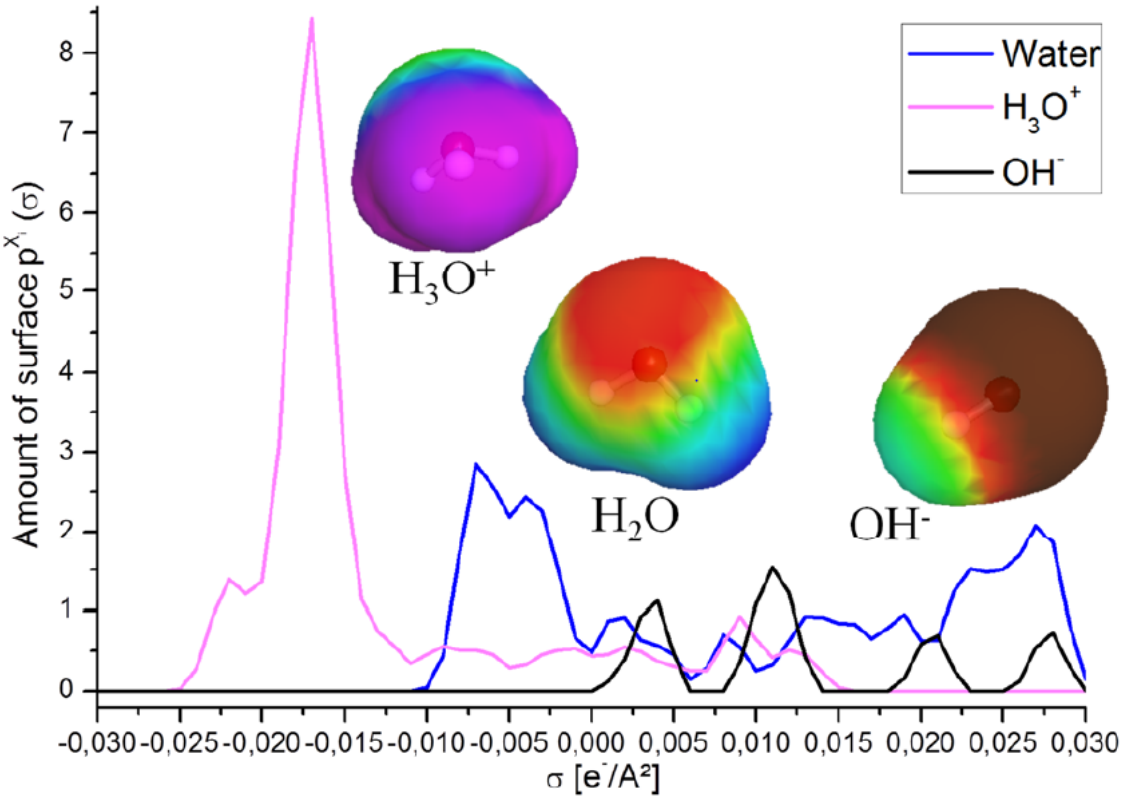


Figure II.1 : σ^{-} profiles de l'eau (H2O), Hydronium ion ($H3O^{+}$) et hydroxyde (OH^{-})dans les modèles de type cosmo [15].

II.2.5. Techniques expérimentales pour les mesures d'IDAC

Différentes méthodes expérimentales existent pour déterminer les coefficients d'activité de dilution infinie [14]:

a) Chromatographie gaz-liquide ;

b) Méthode d'extraction au gaz inerte ;

c) Méthode d'analyse de l'espace de tête ;

d) Chromatographie indirecte d'espace de tête.

e) méthode du point de rosée ;

f) Méthode des cellules statiques différentielles ;

g) Méthode d'ébulliométrie différentielle

h) Méthode de distillation de Rayleigh.

II.2.5.1. Chromatographie gaz-liquide (GLC)

Dans cette technique, le solvant, utilisé comme phase stationnaire, est chargé sur la colonne. Le soluté, qui est la phase mobile, est injecté dans un gaz porteur à travers la colonne. Le soluté élué est détecté et son temps de rétention, avec d'autres paramètres expérimentaux, est utilisé dans une équation pour calculer l'IDAC. La chromatographie gaz-liquide peut être utilisée pour les IDAC de solutés volatils solvants non volatils ou à faible volatilité. Il ne s'applique pas aux systèmes impliquant un mélange de les solvants. Le tableau II.1 résume les avantages et les inconvénients de Gas-Liquide Chromatographie [14].

Tableau II.1: Avantages et inconvénients de la méthode de chromatographie liquide en phase gazeuse. [14]

Avantages	**inconvénients**
La pureté de l'échantillon n'est pas nécessaire	La méthode est adaptée à la faible volatilité ou solvants non volatils seulement
La méthode est rapide autant de solutés peut être injecté immédiatement	L'IDAC du solvant dans le soluté ne peut être déterminé
Les systèmes réactifs peuvent être étudiés	

II.2.5.2. La technique d'extraction de gaz inerte. (IGS)

Dans cette méthode, un soluté hautement dilué est extrait d'une solution par un flux constant de gaz inerte dans des conditions isothermes. La vapeur quittant la cellule est périodiquement analysée par gaz Chromatographie. On observe que la zone de pic représentant la concentration de soluté dans le la phase vapeur diminue exponentiellement avec le temps. Dans cette condition particulière de l'exponentielle dilution, le coefficient d'activité de dilution infinie du soluté dans le solvant est lié à taux décroissant de sa surface de pic correspondante avec le temps. La méthode IGS est utilisée pour mesurer γ^{∞}de solutés volatils dans des solvants non volatils ou volatils. Les systèmes avec des mélanges de solvants peuvent être étudié. Le tableau II.2 présente les avantages et les inconvénients de cette technique [14].

Tableau II.2: Avantages et inconvénients de la méthode d'extraction au gaz inerte. [14]

Avantages	**inconvénients**
L'étalonnage du CPG n'est pas requis puisque seuls les rapports de surface de pic sont impliqués dans les calculs	Les solutés doivent être purifiés.
Un mélange de solvants peut être étudié.	Les systèmes avec des solvants à haute volatilité peuvent être mesurés

Les solutés à faible volatilité sont difficiles à mesurer.	Un bon contact gaz-liquide est essentiel pour obtenir des résultats fiables.
De nombreux solutés peuvent être étudiés la fois	

II.2.6. Les descripteurs moléculaires basés sur profils σ des Lis

La première étape du développement actuel du modèle QSPR consiste à construire des descripteurs moléculaires appropriés à partir des profils σ des cations et des anions des Lis étudiés. Généralement, le profil σ des cations se situe dans la région non polaire (-1.0 <σ <1.0) e / nm2), et la région donneur de liaison hydrogène (σ <-1.0 e / nm 2). Cependant, pour les anions, les pics de σ sont habituellement dans la région d'accepteur de liaison hydrogène (σ <1,0 e / nm2) et la région non polaire (-1,0 <σ <1,0) e / nm2). La position, la largeur et la hauteur des pics varient en fonction de la nature des cations et des anions, de la longueur de chaîne alkyle et de la présence de groupes fonctionnels dans la chaîne alkyle [16].

Les σ-profile de descripteurs (S σ) des deux cations et des deux anions comme exemples typiques utilisés dans cette étude avec les divisions de zone σ-profils sont présentés à la Figure II.2.

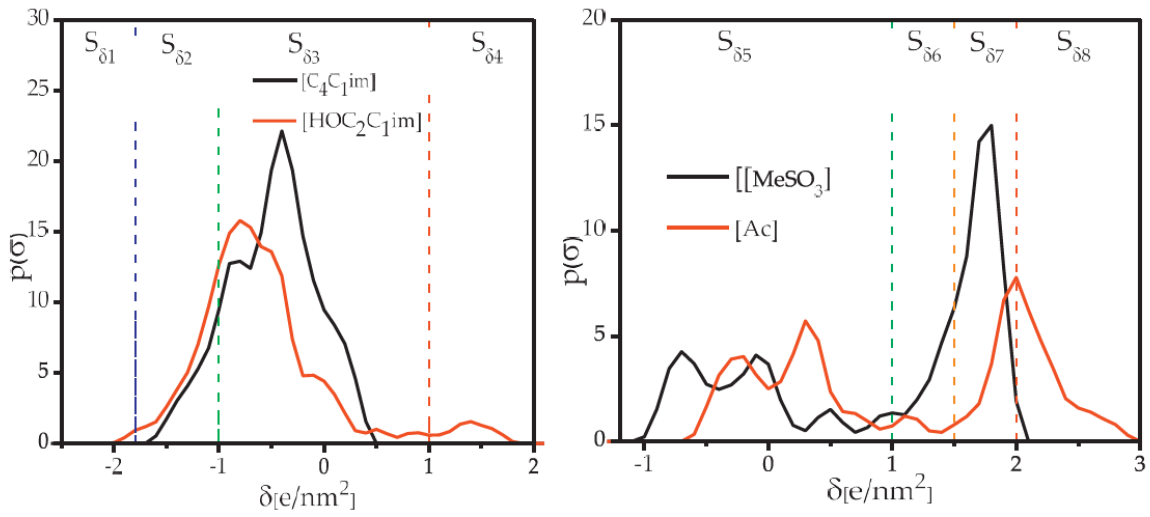


Figure II .2 : σ-profile des cations:1-(2-hydroxyéthyl)-3-méthylimidazolium ([HOC2C1im]), 1-butyl-3-méthylimidazolium ([C 4 C 1 Im]) et des anions : méthane

sulphonate ([MeSO3]) et acétate ([Ac]). [GONFA] [16].

La figure II.2 (a) montre le profil σ du cation 1-butyl-3-méthylimidazolium ([C 4 C 1 Im]) s'étend de -1,7 à 0,5 e / nm2 avec un pic maximal d'environ -0,5 e / nm2. Cela montre que [C 4 C 1 Im] a à la fois le caractère de donneur de liaison hydro-génique en raison de l'hydrogène acide attaché au cation d'imidazolium et du caractère non polaire en raison des Chaînes alkyle . Cependant, dans le cas du 1- (2-hydroxyéthyl) -3-méthylimidazolium ([HOC 2 C 1 Im]), le profil σ est plus étendu de (-2,1 à 1,9) e / nm2 et le pic maximal est également déplacé vers la région polaire (-0,9 e / nm2).

L'introduction de la fonction hydroxyle augmente la capacité du donneur de liaison hydrogène du cation 1- (2-hydroxyéthyl) -3-méthylimidazolium. Par conséquent, pour prendre en considération les effets des groupes fonctionnels, nous avons divisé les profils σ des cations comme le montre la Figure II.2 (a) et le tableau II.3 Ainsi, S σ1 représente un fort caractère de donneur de liaison hydrogène des cations, en particulier pour les cations fonctionnalisés; S σ2 représente le comportement faible des donneurs de liaison à l'hydrogène; S σ3 montre le caractère non polar des cations. S σ4 représente le comportement d'accepteur de liaison hydrogène des cations; Principalement pour les cations contenant un groupe fonctionnel avec un caractère accepteur de liaison hydrogène tel qu'un groupe fonctionnel hydroxyle. Les valeurs S σ des cations 1- (2-hydroxyéthyl) -3-méthylimidazolium et 1-butyl-3- méthylimidazolium sont fournies dans le tableau 1. Les valeurs de S σ indiquent que le cation de 1- (2-hydroxyéthyl) -3-méthylimidazolium a à la fois un donneur de liaison à hydrogène fort et un caractère accepteur alors que le 1-butyl-3-méthylimidazolium ne présente pas ce comportement.

Le profil σ des anions méthanesulfonate et acétate s'étendent dans la région accepteur de liaison hydrogène (σ> 1,0 e / nm2) et la région non polaire (-1,0 <σ <1,0) e / nm2 (figure II.2 (b)). Les pics dans la région non polaire sont dus à la chaîne méthylalkyle (CH3) des anions méthanesulfonate et acétate. En fonction de la capacité d'accepteur de liaison d'hydrogène de l'anion, le profil σ peut se déplacer et changer

À droite (régions acceptrices de liaison hydrogène forte). Par exemple, le profil σ de l'acétate s'est étendu à une région acceptrice de liaison hydrogène forte par rapport à l'anion méthanesulfonate (figure II.2 (b)) en raison de sa basicité plus forte par rapport à l'anion précédent. Ainsi, les valeurs numériques du descripteur moléculaire S σ8 sont plus élevées pour l'acétate que pour le méthanesulfonate. Le descripteur moléculaire S σ8 montre un comportement d'accepteur de liaison hydrogène fort des anions tandis que S σ5 et S σ6

indique le caractère non polaire des anions. S σ7 montrent le comportement faible accepteur de liaison d'hydrogène des anions. Les valeurs S σ du méthanesulfonate et de l'acétate sont données dans le tableau II .3. [GONFA

Tableau II .3 : Les profiles σ des cations: [HOC2C1im] et [C 4 C 1 Im] et des anions :[MeSO3] et [Ac].[16]

Cations				
	S_1 (e/nm^2) ($-2.5 \leq S_{\sigma 1} \leq -1.8$)	S_2(e/nm^2) ($-1.8 < S_{\sigma 1} \leq -1.0$)	S_3(e/nm^2) ($-1.0 < S_{\sigma 1} \leq 1.0$)	S_4(e/nm^2) ($1.0 < S_{\sigma 4} \leq 2.0$)
[HOC_2C_1im] [C_4C_1im]	0.0961 0	3.77365 2.62295	12.4437 17.60025	0.78105 0
Anions				
	S_5 (e/nm^2) ($-1.1 \leq S_{\sigma 5} \leq 1.0$)	S_6 (e/nm^2) ($1.0 < S_{\sigma 6} \leq 1.5$)	S_7 (e/nm^2) ($1.5 < S_{\sigma 7} \leq 2.0$)	S_8 (e/nm^2) ($2.0 < S_{\sigma 7} \leq 3.0$)
[MeSO 3] [Ac]	4.17565 4.3219	1.4724 0.39395	5.03015 1.7661	0.0952 2.55105

II.3. Conclusion

Le coefficient d'activité a dilution infinie présente un importance particulière pour [17] :

- caractériser le comportement des mélanges liquides
- ajustement des paramètres du modèle gE (par exemple Margules, van Laar, Wilson, NRTL, UNIQUAC)
- prédire l'existence d'un azéotrope
- l'estimation des solubilités mutuelles
- chromatographes analytiques
- criblage des solvants pour l'extraction et les procédés de distillation extractive
- le calcul des facteurs de séparation limites nécessaires pour la conception fiable des processus de distillation
- le calcul des constantes de Henry et des coefficients de partition
- le développement de méthodes prédictives

Chapitre III : Support Vector Machines (SVM)

III.1 .Introduction

L'origine des machines à vecteurs de support (**SVM**) remonte à 1975 lorsque Vapnik et Chervonenkis proposèrent le principe du risque structurel et la dimension **VC** pour caractériser la capacité d'une machine d'apprentissage. A cette époque, Ce principe n'a pas trouvé place et il n'existait pas encore un modèle de classification solidement appréhendé pour être utilisable. Il a fallu attendre jusqu'à l'an 1982 pour que Vapnik propose un premier classificateur basé sur la minimisation du risque structurel baptisé **SVM**. [19]

Les Support Vector Machines (SVM) sont des nouvelles techniques discriminantes dans la théorie de l'apprentissage statistique. Elles permettent d'aborder plusieurs problèmes divers et variés comme la régression, la classification, la fusion etc. [19]

Les SVM fournissent une approche très intéressante de l'approximation statistique. Souvent, le nombre des exemples pour l'apprentissage est insuffisant pour que les estimateurs fournissent un modèle avec une bonne précision. D'un autre côté, l'acquisition d'un grand nombre d'exemples s'avère être souvent très coû- teuse et peut même mener à des problèmes de sur-apprentissage dans le cas ou la capacité du modèle est très complexe. Pour ces deux raisons, il faut arriver à un compromis entre la taille des échantillons et la précision recherchée. Dans ces cas spécifiques comme la reconnaissance de formes, il serait intéressant de trouver une mesure de la fiabilité de l'apprentissage, et d'avoir une mesure du taux d'erreur qui sera commis durant la phase de test. Ces nouvelles techniques unifient deux théories : Minimisation du risque empirique et Capacité d'apprentissage d'une famille de fonctions. C'est la Minimisation du Risque Structurel. [19]

Le principe des SVM consiste à projeter les données de l'espace d'entrée (appartenant à deux classes différentes) non-linéairement séparables dans un espace de plus grande dimension appelé espace de caractéristiques de façon à ce que les données deviennent linéairement séparables. Dans cet espace, on construit un hyperplan optimal séparant les classes tel que [19] :

_ Les vecteurs appartenant aux différentes classes se trouvent de différents

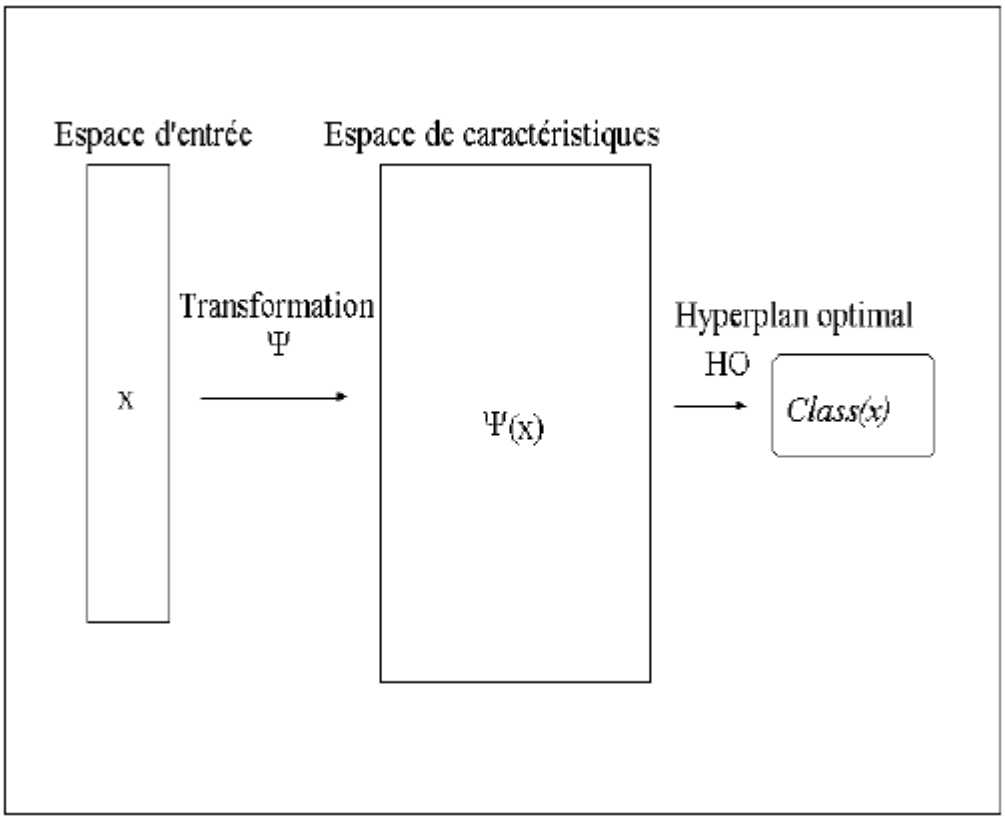


Figure III.1: Principe des techniques SVM [19]

Côtés de l'hyperplan.

_ La plus petite distance entre les vecteurs et l'hyperplan (la marge) soit maximale.

La Figure III.1 représente le principe de la technique SVM.

III.2. Pourquoi les Machines à Vecteurs de Support (SVM) ?

L'algorithme des machines à vecteurs de support a été développé dans les années 90 par le russe Vladimir Vapnik. Initialement, les SVM ont été développé comme un algorithme de classification binaire supervisée. Il s'avère particulièrement efficace de par le fait qu'il peut traiter des problèmes mettant en jeu de grands nombres de descripteurs, qu'il assure une solution unique (pas de problèmes de minimum local comme pour les réseaux de neurones) et il a fourni de bons résultats sur des problèmes réels [20].

III.3. Apprentissage statistique et SVM

La notion d'apprentissage étant importante, nous allons commencer par effectuer un rappel. L'apprentissage par induction permet d'arriver à des conclusions par l'examen d'exemples particuliers. Il se divise en apprentissage supervisé et non supervisé. Le cas qui concerne les SVM est l'apprentissage supervisé. Les exemples particuliers sont représentés par un ensemble de couples d'entrée/sortie. Le but est d'apprendre une fonction qui correspond aux exemples vus et qui prédit les sorties pour les entrées qui n'ont pas encore été vues. Les entrées peuvent être des descriptions d'objets et les sorties la classe des objets donnés en entrée [20].

III.3.1 Objectif de l'apprentissage statistique

Effectuer une classification consiste à déterminer une règle de décision capable, à partir d'observations externes, d'assigner un objet à une classe parmi plusieurs. Le cas le plus simple consiste à discriminer deux classes. D'une manière plus formelle, la classification bi-classe revient à estimer une fonction f : x → {+1,-1} à partir d'un ensemble d'apprentissage constitué de couples (xi, yi), qu'on suppose i.i.d. suivant une distribution de probabilité P(x, y) inconnue, tels que :

(xi, yi) ϵ X × Y où i =1,…, Nx et Y= {+1,-1}, de sorte à ce que f classifie correctement des exemples inconnus (x_t, y_t). Par exemple, on peut assigner x_t à la classe (+1) si f (x_t) ≥0, et à la classe (-1) sinon. Les exemples inconnus sont supposés suivre la même distribution de probabilité P(x, y) que ceux de l'ensemble d'apprentissage. La meilleure fonction f est celle obtenue en minimisant le risque :

$$R[f] = \int L\,[f(x), y]\, d\, P(x, y). \quad (4.1)$$

Où L désigne une fonction de coût, comme par exemple :

$$L\,[f(x),y] = (f(x)-y)^2$$

Malheureusement, le risque (4.1) ne peut être directement minimisé dans la mesure où la distribution de probabilité sous-jacente P(x, y) est inconnue. Aussi, on va chercher une fonction de décision proche de celle optimale à partir de dont on dispose, c'est-à dire l'ensemble d'apprentissage et la classe de fonctions F est à laquelle la solution f appartient. Pour ce faire, on approxime le minimum du risque théorique par le minimum du risque empirique qui s'écrit :

$$R_{emp}[f]=\frac{1}{Nx}\sum_{i=1}^{Nx} L\,[f(x_i),y_i]\ . \qquad (4.2)$$

Il est possible de donner des conditions au classifieur pour qu'asymptotiquement (si Nx→ ∞), le risque empirique (4.2) converge vers le risque (4.1). Cependant, si on dispose de peu d'exemples pour faire l'apprentissage (i.e Nx petit), on s'expose au risque de sur-apprentissage (Figure 4.1). Pour éviter le sur-apprentissage, on peut restreindre la complexité de la classe F à laquelle appartient *f*. Intuitivement, une fonction de décision simple (la classe la plus simple se constituant des fonctions linéaires) capable de discriminer correctement les données est préférable à une fonction complexe. Pour cela, on introduit un terme de régularisation pour limiter la complexité des fonctions de F.

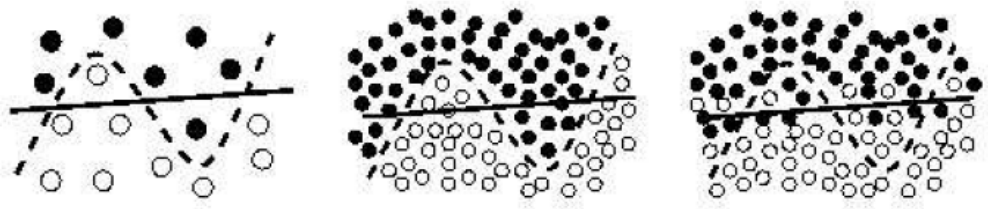


Figure III.2 : Exemple du problème de sur apprentissage. [20]

Etant donné un petit ensemble d'apprentissage (schéma de gauche), deux frontières de discrimination (représentées par les lignes continue et discontinue) sont possibles. La ligne discontinue est plus complexe mais minimise davantage le risque empirique. Seul un ensemble d'exemples plus grand permet de déterminer la meilleure des deux frontières de décision. S'il s'agit de la ligne discontinue, alors la ligne continue n'est pas suffisamment discriminante (schéma du milieu) ; s'il s'agit de la ligne continue, alors la ligne discontinue ne convient pas et caractérise un sur apprentissage (schéma de droite) [20].

III.3.2 Théorie de Vapnik-Chervonenkis

Une manière de contrôler la complexité d'une classe de fonctions est donnée par la théorie de Vapnik-Chervonenkis (VC) et le principe de minimisation du risque structurel. Ici, le concept de complexité de la fonction de décision f s'exprime par la dimension de VC (notée h) de la classe de fonctions F à laquelle appartient f.

Grossièrement, la dimension de VC mesure combien d'échantillons de l'ensemble d'apprentissage peuvent être séparés par toutes les classifications possibles issues des fonctions de la classe [20].

Considérons une famille imbriquée de classes de fonctions:

$F_1 \subset F_2 \subset \ldots \subset F_k$;

Avec une dimension de VC non-décroissante, et $f_1 \ldots f_k$ les fonctions minimisant le risque empirique dans chacune de ces classes.

La minimisation du risque structurel consiste à choisir la classe Fi (et la fonction *f*i) de sorte à ce qu'une borne supérieure de l'erreur de généralisation puisse être minimisée (grâce, par exemple, au théorème suivant) [20].

Théorème 1 : Soient h la dimension de VC de la classe de fonctions F, Remp[f] le risque empirique défini par (4.2) avec la fonction perte 0/1 (i.e. L [f (xi), yi] =H (y f(x)) Où H désigne la fonction de Heaviside). Pour tout $\delta>0$ et $f \in F$, l'inégalité bornant le risque

$$R[f] = R_{emp}[f] + \sqrt{\frac{h(\ln\frac{2Nx}{h}+1) - \ln(\frac{\delta}{4})}{Nx}} \qquad (4,3)$$

est vraie avec une probabilité de moins $(1-\delta)$ pour $Nx > h$.

Cette borne n'est qu'un exemple et des formulations du même type ont été démontrées pour d'autres fonctions perte et d'autre mesures de complexité. Le but recherché ici est de minimiser l'erreur de généralisation R[f] en obtenant un faible risque empirique R_{emp} [f] tout en gardant la plus petite classe de fonctions possible.

L'inégalité (4.3) fait apparaître deux cas extrêmes:

- une très petite classe de fonctions (par exemple F1) fait décroître rapidement le terme de complexité (celui en racine carrée), mais le risque empirique demeure grand,
- une très grande classe de fonctions (par exemple Fk) implique un risque empirique petit, mais le terme de complexité explose.

La meilleure classe de fonctions est généralement intermédiaire entre la plus petite et la plus grande, puisque nous cherchons d'avoir une fonction qui explique au mieux les données tout en préservant un faible risque empirique (Figure 4.2) [20].

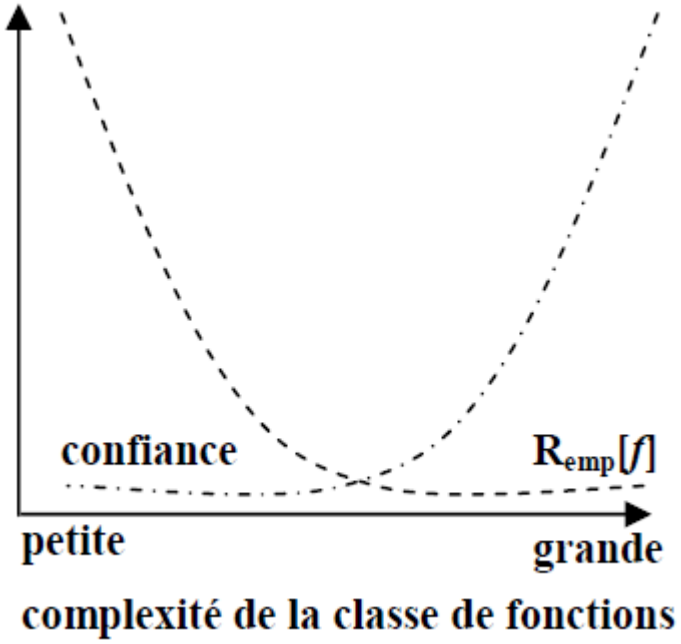


Figure III.3: Illustration de l'inégalité (4.3) [20].

La courbe croissante, appelée confiance, correspond à la borne supérieure du terme de complexité. Les comportements du terme de complexité et de l'erreur empirique sont clairement opposés. On recherche donc le meilleur compromis entre complexité et erreur empirique [20].

III.3.3 Marge et dimension de VC

Supposons pour l'instant que les échantillons de l'ensemble d'apprentissage sont séparables par un hyperplan (Figure 4.3), i.e on choisit des fonctions de décision de la forme :

$$f(x)= (w, x) +b \qquad (4.4)$$

La marge est la distance minimale entre les échantillons de l'ensemble d'apprentissage et la frontière de décision.

Il a été montré que pour la classe des hyperplans, la dimension de VC peut être bornée en fonction de la marge. La marge peut à son tour être mesurée grâce au vecteur poids

w: puisque nous supposons que les échantillons sont séparables, on peut redéfinir *w* et *b* de sorte à ce que les échantillons *x* les plus proches de l'hyperplan satisfassent

$$| (w, x)+b |=1$$

Considérons maintenant deux échantillons $x1$ et $x2$ de classes différentes telles qu'on ait. $|(w, x_1) + b| = +1$ et $|(w, x_2) + b| = -1$ La marge γ correspond alors à la distance entre $x1$ et $x2$ mesurée perpendiculairement à l'hyperplan :

$$\gamma = (w / \|w\|, x1 - x2) = 2/\|w\|$$

Les résultats liant la dimension de VC de la classe des hyperplans de séparation à la marge et à la longueur du vecteur poids w sont respectivement donnés par les inégalités suivantes :

Où R est le rayon de la plus petite boule englobant les données. Ainsi, en bornant la marge de la classe de fonction, on peut contrôler sa dimension de VC [20].

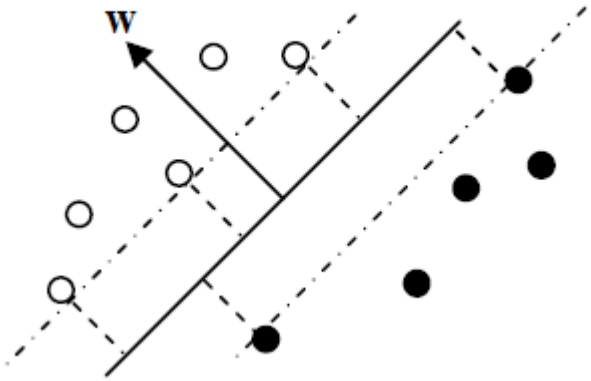


Figure III.4 : Classifieur linéaire et marge (4.3) [20].

Un classifieur linéaire est défini par un vecteur normal à l'hyperplan w et un biais b:la frontière de décision est $\{|x| (w, x)+b=0\}$ (ligne continue).Chacun des deux sous-espaces séparés par l'hyperplan correspond à une classe, i.e. $f(x)=\text{signe}((w, x)+b)$. La marge du classifieur linéaire est la distance minimale entre les échantillons de l'ensemble d'apprentissage et la frontière de décision. Sur le schéma, il s'agit de la distance entre la ligne continue et les lignes discontinues [20].

III.4. SVM principe de fonctionnement général

III.4.1 Notions de base: Hyperplan, marge et support vecteur

Pour deux classes d'exemples donnés, le but de SVM est de trouver un classificateur qui va séparer les données et maximiser la distance entre ces deux classes. Avec SVM, ce classificateur est un classificateur linéaire appelé hyperplan.

Dans le schéma qui suit, on détermine un hyperplan qui sépare les deux ensembles de points [20].

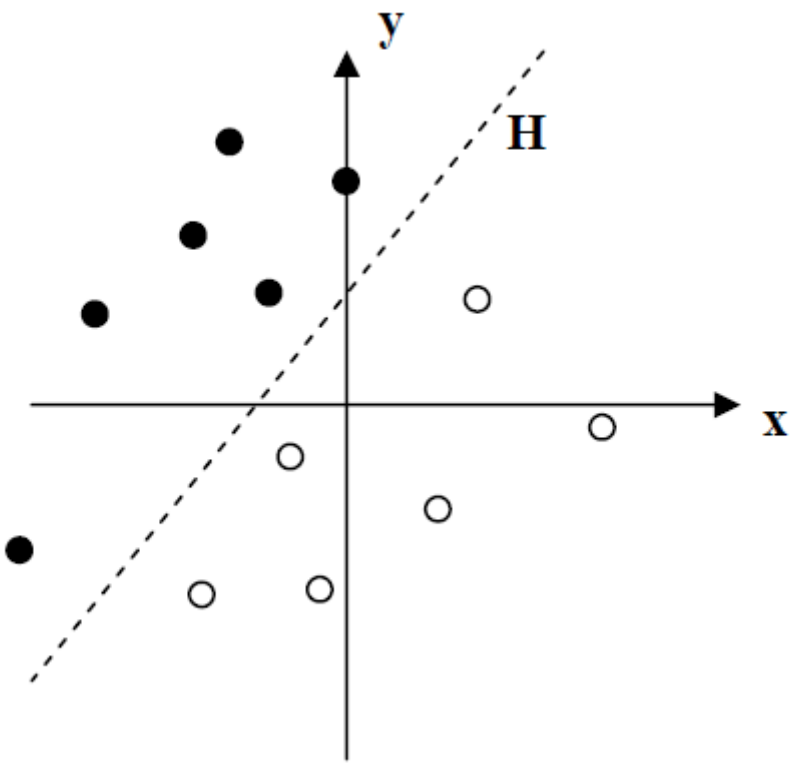


Figure III.5 : Exemple d'un hyperplan séparateur .[20]

Les points les plus proches, qui seuls sont utilisés pour la détermination de l'hyperplan, sont appelés vecteurs de support.

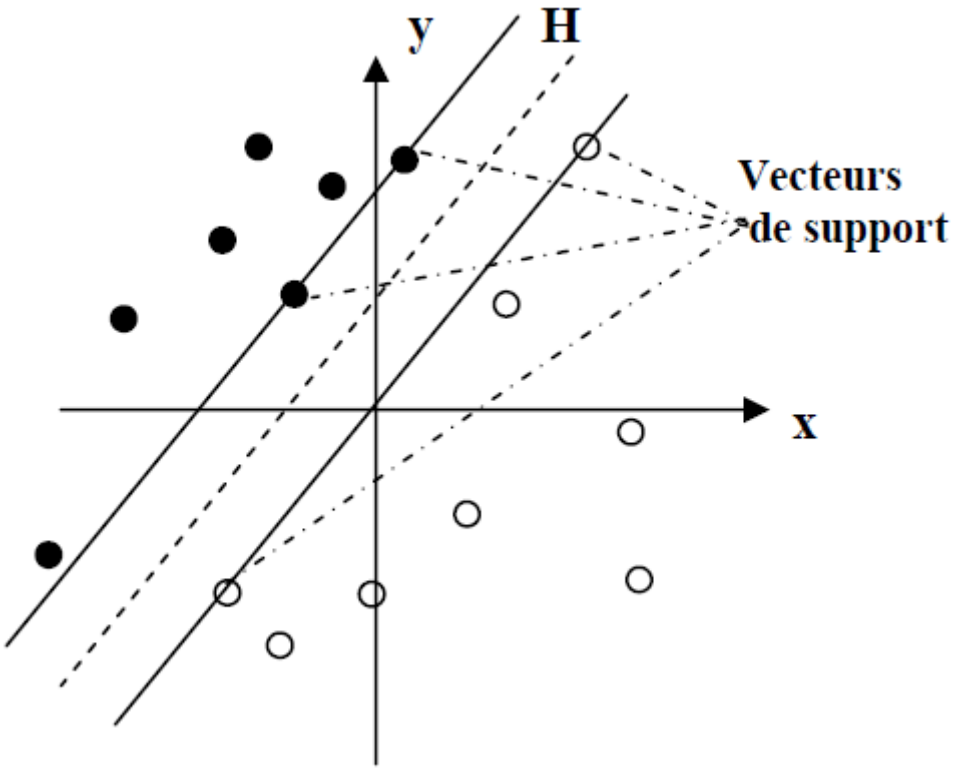


Figure III.6: Exemple de vecteurs de support [20].

Il est évident qu'il existe une multitude d'hyperplan valide mais la propriété remarquable des SVM est que cet hyperplan doit être optimal. Nous allons donc en plus chercher parmi les hyperplans valides, celui qui passe « au milieu » des points des deux classes d'exemples. Intuitivement, cela revient à chercher l'hyperplan le « plus sûr » [19].

En effet, supposons qu'un exemple n'ait pas été décrit parfaitement, une petite variation ne modifiera pas sa classification si sa distance à l'hyperplan est grande. Formellement, cela revient à chercher un hyperplan dont la distance minimale aux exemples d'apprentissage est maximale [19]

On appelle cette distance « marge » entre l'hyperplan et les exemples. L'hyperplan séparateur optimal est celui qui maximise la marge. Comme on cherche à maximiser cette marge, on parlera de séparateurs à vaste marge [20].

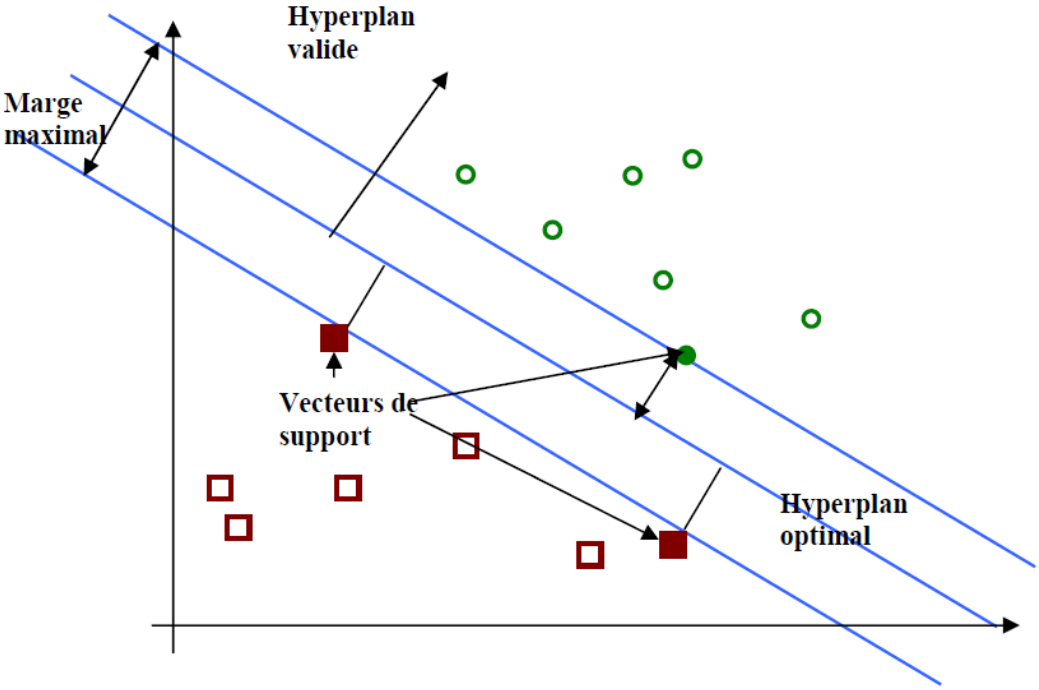


Figure III.7 : Exemple de marge maximal (hyperplan valide)[20].

III.4.2 Pourquoi maximiser la marge ?

Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsque l'on classe un nouvel exemple. De plus, si l'on trouve le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples. Dans le schéma qui suit, la partie droite nous montre qu'avec un hyperplan optimal, un nouvel exemple reste bien classé alors qu'il tombe dans la marge. On constate sur la partie gauche qu'avec une plus petite marge, l'exemple se voit mal classé [20].

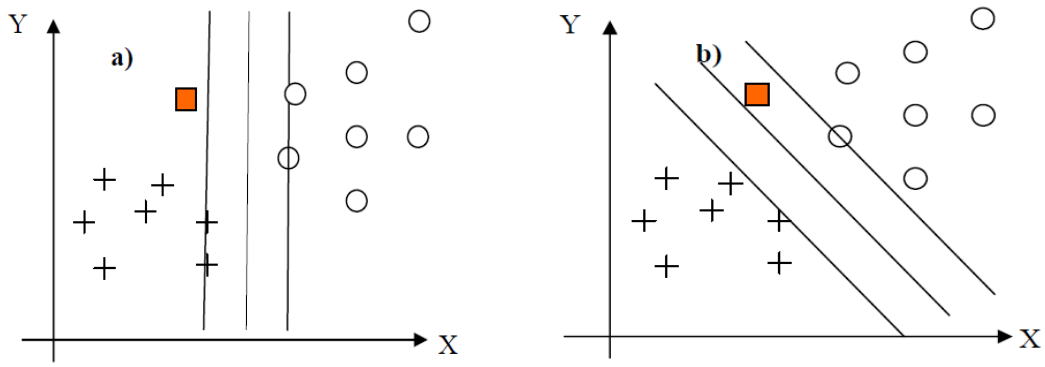


Figure III.8 :a) Hyperplan avec faible marge, **b)** Meilleur hyperplan séparateur [20].

En général, la classification d'un nouvel exemple inconnu est donnée par sa position par rapport à l'hyperplan optimal. Dans le schéma suivant, le nouvel élément sera classé dans la catégorie des « + ».

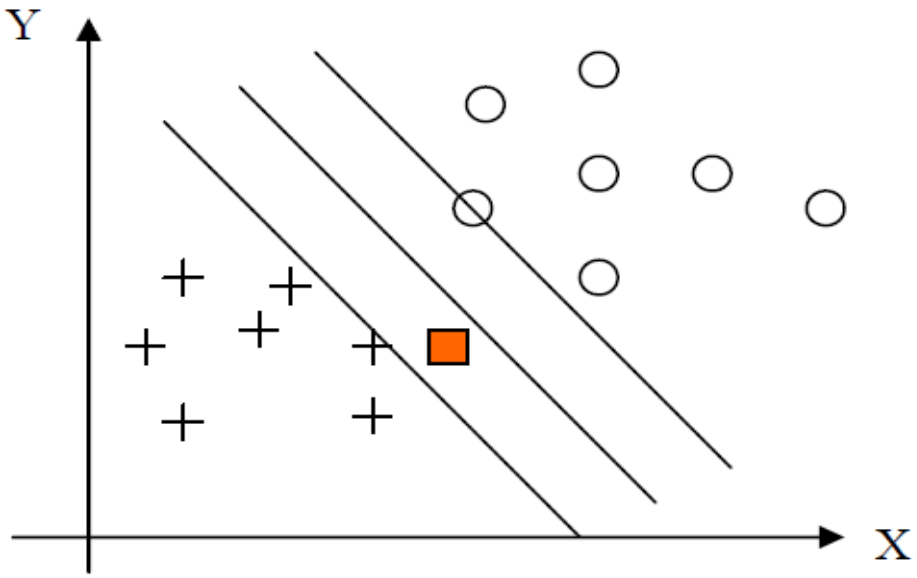


Figure III.9 : Exemple de classification d'un nouvel élément. [20].

III.4.3 Linéarité et non-linéarité

Parmi les modèles des SVM, on constate les cas linéairement séparable et les cas non linéairement séparable.Les premiers sont les plus simples de SVM car ils permettent de trouver facilement le classificateur linéaire. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données, le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables [19].

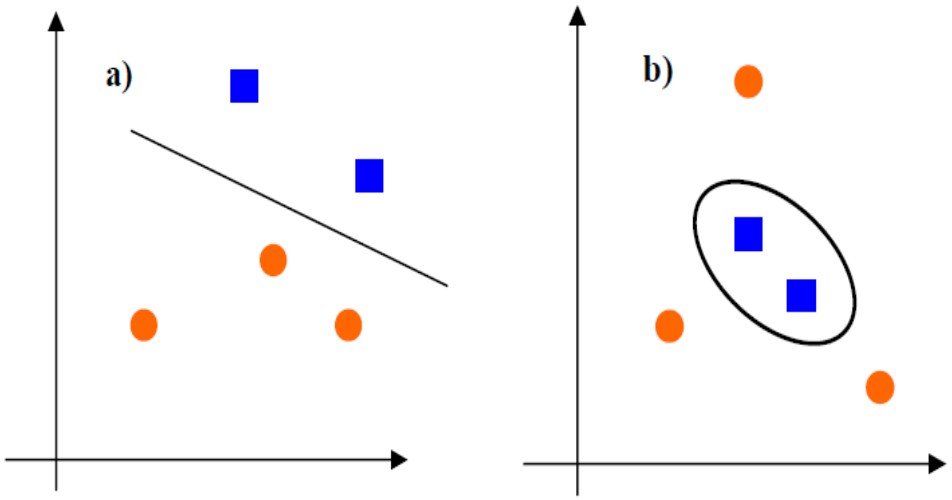


Figure III.10 : a) Cas linéairement séparable, **b)** Cas non linéairement séparable [20].

III.4.4 Cas non linéaire

Pour surmonter les inconvénients des cas non linéairement séparable, l'idée des SVM est de changer l'espace des données. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. On va donc avoir un changement de dimension. Cette nouvelle dimension est appelé « espace de redescription ».En effet, intuitivement, plus la dimension de l'espace de redescription est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée. Ceci est illustré par le schéma suivant [20]:

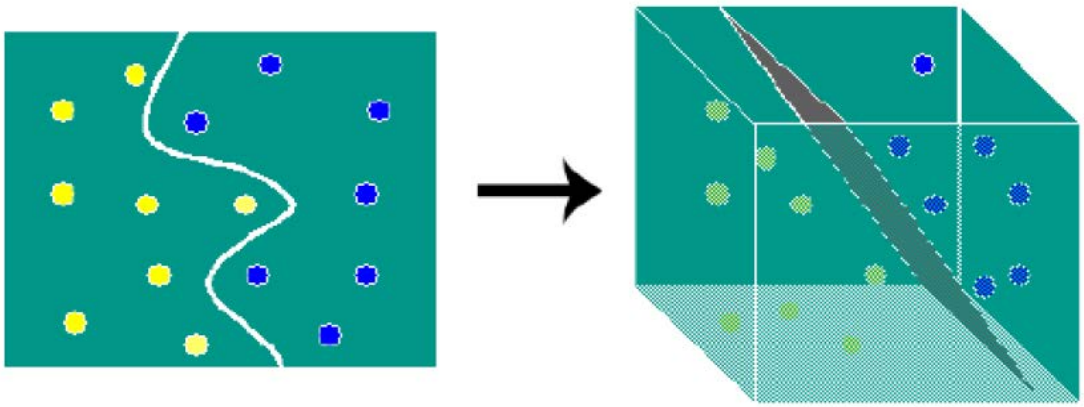


Figure III.11 : Exemple de changement de l'espace de données. [20].

On a donc une transformation d'un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un espace de redescription de plus grande dimension. Cette transformation non linéaire est réalisée via une fonction noyau [20].

En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et laplacien [21].

III.4.5 Illustration de transformation de cas non linéaire : le cas XOR

Le cas de XOR n'est pas linéairement séparable, si on place les points dans un plan à deux dimensions, on obtient la figure suivante :

Coordonnées des points : (0,0) ; (0,1) ; (1,0) ; (1,1)

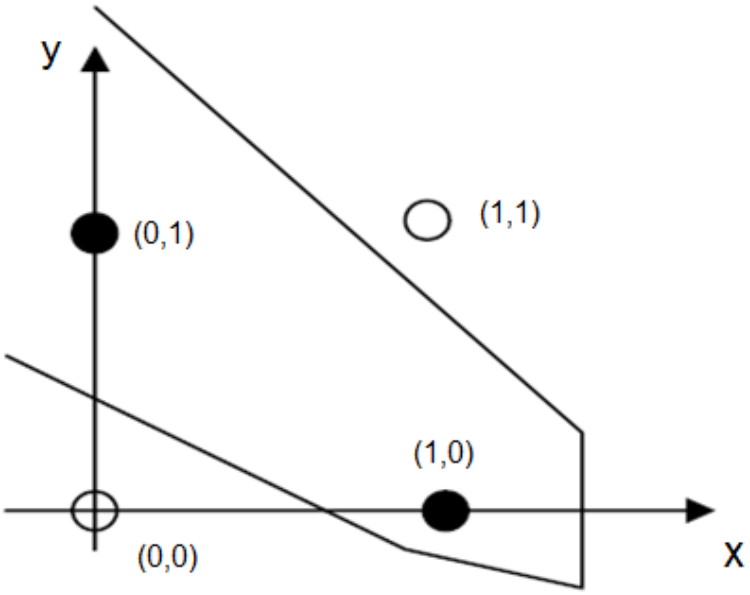


Figure III.12 : Illustration de cas non linéairement séparable (le cas XOR) [20].

Si on prend une fonction polynomiale (x, y) → (x, y, z) qui fait passer d'un espace d dimension 2 à un espace de dimension 3, on obtient un problème en trois dimensions linéairement séparable :

(0,0) → (0, 0,0) (1,0) → (1, 0,0)

(0,1) → (0, 1,0) (1,1) → (1, 1,1)

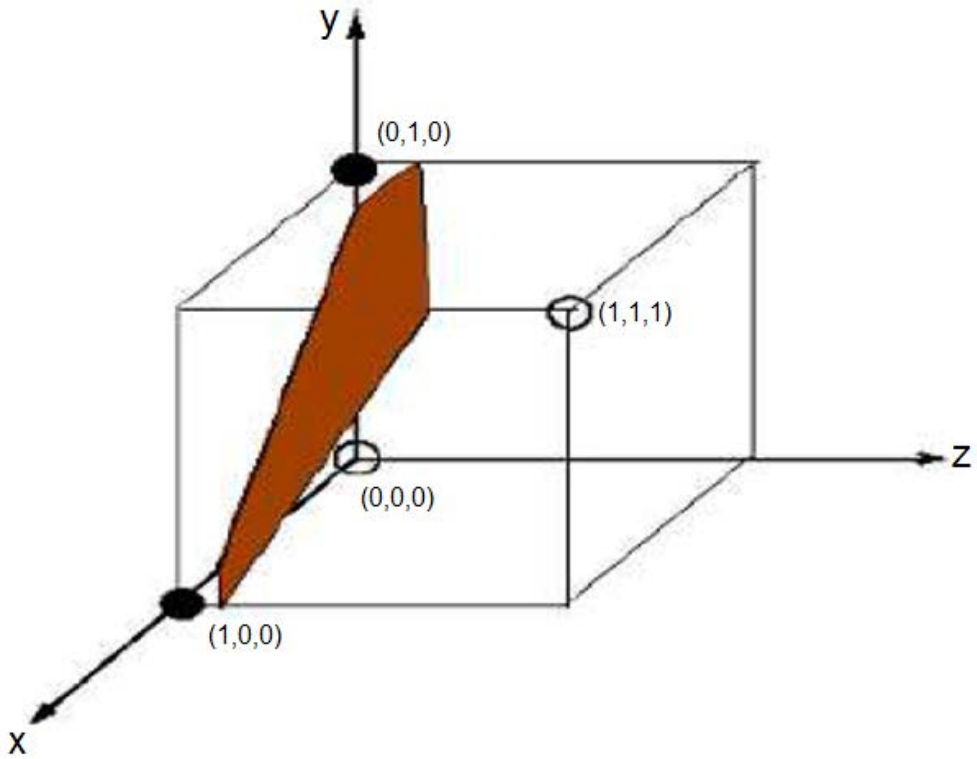


Figure III.13 : Illustration de passage d'un espace 2D à un espace 3D [20].

III.5. Least Squares SVM (LSSVM)

III.5.1 Description de la méthode et résolution

La régression de crête (« ridge regression »), méthode à laquelle appartiennent les LSSVM, est une technique de régression qui constitue une forme régularisée des moindres carrés ; il est possible de l'inscrire dans le cadre du principe de minimisation du risque structurel. [21]

On dispose d'un ensemble de données $\{(x_i, y_i) \in x * y\}_{i-1}^{N}$, où $X \subset R^d$

Et $Y \subset R$.

L'objectif est de trouver une fonction *f* dans un espace de Hilbert *H*, par minimisation du risque structurel (SRM), établissant une relation entre les variables **x** et la grandeur à modéliser *y*, à partir de l'ensemble de mesures *D*.

Plaçons-nous tout d'abord dans le cas linéaire : on cherche une fonction solution qui s'écrive sous la forme $f(\mathbf{x}) = \mathbf{w} * \mathbf{x} + b$, où **w** et *b* sont les paramètres du modèle, et où est le produit scalaire entre **w** et **x** dans l'espace *X*. En se référant à la relation (1), la fonctionnelle que l'on cherche à optimiser s'écrit [21]:

$$F[f] = \gamma \sum_{i=1}^{N} \left(y_i - f(\mathbf{x}_i)\right)^2 + \frac{1}{2}\|f\|_H^2 \qquad (3)$$

Où $\|\cdot\|_H \equiv \|\cdot\|$ *H* est une norme sur l'espace *H*.

Cette fonctionnelle (3) se traduit en le problème d'optimisation suivant :

$$\begin{aligned} &\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 + \gamma \sum_{i=1}^{N} \xi_i^2 \\ &\text{sous } \ \xi_i = y_i - f(\mathbf{x}_i) = y_i - \mathbf{w}\cdot\mathbf{x}_i - b, \ \forall i \in [\![1, N]\!] \end{aligned} \qquad (4)$$

Où le premier terme de la fonction objectif représente la norme de la fonction *f* dans l'espace *H* (terme de régularisation), et où le second terme représente la fonction de coût des moindres carrés au travers des variables d'erreurs *i*

ξ : on retrouve donc l'expression classique du problème d'optimisation des moindres carrés régularisés [21].

Nous allons montrer que ce problème d'optimisation (4) peut être résolu de différentes manières, comme décrit au paragraphe 2.4. [21]

III.5.1.1 Recherche de la solution par utilisation du théorème de représentation

En omettant le terme de biais *b* (nous reviendrons ultérieurement sur ce point), le problème d'optimisation (4) peut s'écrire :

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 + \gamma \sum_{i=1}^{N} (y_i - \mathbf{w}\cdot\mathbf{x}_i)^2 , \qquad (5)$$

Soit, de manière équivalente et sous forme matricielle :

$$\min_{\mathbf{w}} J(\mathbf{w}) = \frac{1}{4\gamma}\|\mathbf{w}\|^2 + \frac{1}{2}\|\mathbf{Y} - \mathbf{X}\mathbf{w}\|^2 \qquad (6)$$

$$\text{où } \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix}, \ \mathbf{Y} = [y_1 \dots y_N]^T .$$

Le problème (6) vérifie bien les hypothèses du théorème de représentation. On peut donc écrire que la solution *f* admet une représentation du type [21] :

$$f(\mathbf{x}) = \sum_{i=1}^{N} c_i \mathbf{x}_i \cdot \mathbf{x}$$

Ceci impose au vecteur **w** de s'écrire sous la forme :

$$\mathbf{w} = \sum_{i=1}^{N} c_i \mathbf{x}_i \equiv \mathbf{X}^T\mathbf{c}$$

$\mathbf{c} = [c_{1.....}c_N]^T$ La fonction objectif du problème d'optimisation (6) devient :

$$J(\mathbf{c}) = \frac{1}{2}\left\|\mathbf{Y} - \mathbf{X}\mathbf{X}^T\mathbf{c}\right\|^2 + \frac{1}{4\gamma}\left\|\mathbf{X}^T\mathbf{c}\right\|^2$$

En imposant la stationnarité de *J* vis-à-vis de **c**, on obtient :

$$\frac{\partial J}{\partial \mathbf{c}} = \mathbf{0} \Leftrightarrow \mathbf{X}\mathbf{X}^T\left(\frac{1}{2\gamma}\mathbf{I} + \mathbf{X}\mathbf{X}^T\right)\mathbf{c} = \mathbf{X}\mathbf{X}^T\mathbf{Y}$$

Où **I** est la matrice identité (*N*, *N*). Comme la matrice définie positive $\mathbf{X}\mathbf{X}^T$ est inversible, le minimum de *J*, qui correspond à la solution du problème d'optimisation (6), vérifie la relation suivante :

$$\left(\frac{1}{2\gamma}\mathbf{I}+\mathbf{X}\mathbf{X}^T\right)\mathbf{c}=\mathbf{Y} \qquad (7)$$

La recherche de la solution se ramène donc au calcul de l'inverse (ou de la pseudo inverse) d'une matrice qui ne fait intervenir que les produits scalaires entre éléments de l'espace de régression. [21]

Cette caractéristique de la solution va nous permettre de faire usage de l' « astuce des noyaux» pour étendre la méthode de régression au domaine non linéaire [17]. En réalisant une transformation d'espace préliminaire, définie par une fonction
$\varphi : X \rightarrow F$, où *F* est un espace image muni du produit scalair $\varphi(x)$, $\varphi(y)=K(x,y)$, le problème d'apprentissage devient [21] :

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2+\gamma\sum_{i=1}^{N}\left(y_i-\mathbf{w}\cdot\phi(\mathbf{x}_i)\right)^2 \qquad (8)$$

Où **w** appartient désormais à l'espace image *F*, de dimension potentiellement infinie.

Le résultat obtenu dans le cas linéaire nous permet d'affirmer que la solution du problème (8) est définie par la relation suivante [21] :

$$\left(\frac{1}{2\gamma}\mathbf{I}+\mathbf{K}\right)\mathbf{c}=\mathbf{Y} \qquad (9)$$

Ou $K=[k_{ij}=K(x_i,y_i)=(\varphi(x_i),(y_i))]^N_{ij=1}$ joue le rôle de la matrice $\mathbf{X}\mathbf{X}^T$ du cas linéaire. La fonction *f* solution s'écrit alors de la façon suivante :

$$f(\mathbf{x})=\sum_{i=1}^{N}c_i\langle\phi(\mathbf{x}_i),\phi(\mathbf{x})\rangle=\sum_{i=1}^{N}c_iK(\mathbf{x}_i,\mathbf{x}) \qquad (10)$$

En définitive, l'utilisation du théorème de représentation nous a permis, dans le cas linéaire, d'exprimer la solution du problème d'apprentissage comme une somme pondérée de produits scalaires sur les exemples de la base d'apprentissage. Ce résultat permet d'étendre la méthode de régression au cas non linéaire par une simple transformation d'espace, qui n'intervient qu'implicitement au travers de la définition d'une fonction noyau *K* (« astuce des noyaux »). [21]

Concernant le terme de biais *b*, que nous avons omis lors de la recherche de la solution, notons que celui-ci n'a que peu d'influence dans le cas non linéaire, puisqu'il revient à ajouter une dimension supplémentaire à un espace de dimension potentiellement infinie. [20]

III.5.1.2 Recherche de la solution par exploitation des conditions de points selle

Revenons à l'expression du problème initial (4), en présence du terme de biais, et après une transformation d'espace φ : $X \rightarrow F$ (nous nous plaçons donc directement dans le cas non linéaire et noterons $(\varphi(\mathbf{x}), \varphi(\mathbf{x})) \equiv (\varphi(x), \varphi(x))_F$ le produit scalaire dans F) [21] :

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 + \gamma \sum_{i=1}^{N} \xi_i^2$$

$$\text{sous } \xi_i = y_i - f(\mathbf{x}_i) = y_i - \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - b, \forall i \in [\![1, N]\!]$$

Nous allons à présent montrer qu'en exploitant les caractéristiques de ce problème d'optimisation, il est également possible d'aboutir à l'expression de la solution. [21]

Le Lagrangien du problème d'optimisation ci-dessus s'écrit :

$$L(\mathbf{w}, b, \boldsymbol{\xi}; \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 + \gamma \sum_{i=1}^{N} \xi_i^2 - \sum_{i=1}^{N} \left[\alpha_i \left(\xi_i - y_i + \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b \right) \right] \quad (11)$$

Où les coefficients $\{\alpha_i\}^N_{i-1}$, appelés multiplicateurs de Lagrange, sont des réels quelconques du fait des contraintes d'égalité du problème d'optimisation.

La convexité de la fonction objectif et la linéarité des contraintes d'égalité permettent d'affirmer l'existence d'un point selle en l'unique solution du problème [18]. Il est donc possible de rechercher le minimum vis-à-vis des variables primales du Lagrangien (11), avant de le rechercher vis-à-vis des variables duales (multiplicateurs de Lagrange). La stationnarité du Lagrangien vis-à-vis des variables primales s'écrit [21] :

$$\begin{cases} \dfrac{\partial L}{\partial \mathbf{w}} = 0 \Leftrightarrow \mathbf{w} - \sum_{i=1}^{N} \alpha_i \phi(\mathbf{x}_i) = 0 \Leftrightarrow \mathbf{w} = \sum_{i=1}^{N} \alpha_i \phi(\mathbf{x}_i) \\ \dfrac{\partial L}{\partial b} = 0 \Leftrightarrow \sum_{i=1}^{N} \alpha_i = 0 \\ \dfrac{\partial L}{\partial \xi_i} = 0 \Leftrightarrow 2\gamma\xi_i - \alpha_i = 0 \Leftrightarrow \xi_i = \dfrac{\alpha_i}{2\gamma}, \forall i \in [\![1, N]\!] \end{cases}$$

En substituant ces trois relations dans l'expression du Lagrangien primal on obtient le problème d'optimisation dual suivant :

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & W(\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle - \frac{1}{4\gamma}\sum_{i=1}^{N} \alpha_i^2 + \sum_{i=1}^{N} y_i \alpha_i \\ sous \quad & \sum_{i=1}^{N} \alpha_i = 0 \end{aligned} \tag{12}$$

Sous forme matricielle, celui-ci s'écrit :

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & W(\boldsymbol{\alpha}) = -\frac{1}{2}\boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} - \frac{1}{4\gamma}\boldsymbol{\alpha}^T \boldsymbol{\alpha} + \boldsymbol{\alpha}^T \mathbf{y} \\ sous \quad & \mathbf{1}^T \boldsymbol{\alpha} = 0 \end{aligned} \tag{13}$$

où $\mathbf{K} = \left[k_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \right]_{i,j=1}^{N}$, $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$,

$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T$, $\mathbf{1} = [1, \ldots, 1]^T$.

Une fois déterminés les coefficients *αi*, le terme de biais *b* peut être calculé sur chaque exemple (ou en moyennant sur l'ensemble des *N* exemples) par :

$$b = y_i - \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - \xi_i = y_i - \sum_{j=1}^{N} \alpha_j \langle \phi(\mathbf{x}_j), \phi(\mathbf{x}_i) \rangle - \frac{\alpha_i}{2\gamma}, \forall i \in [\![1, N]\!]$$

Le fait d'adopter un formalisme Lagrangien et d'exploiter les caractéristiques du problème d'optimisation permet donc d'aboutir à l'expression d'un nouveau problème d'optimisation dont la résolution est plus aisée (les contraintes étant plus simples à traiter).

Remarque : en l'absence erme de biais *b*, la contrainte

$$\sum_{i=1}^{N} \alpha_i = 0$$

Disparaît, et le problème d'optimisation est encore simplifié. La solution du problème d'optimisation s'obtient alors simplement en imposant la stationnarité au Lagrangien dual $W(\boldsymbol{\alpha})$:

$$\frac{\partial W}{\partial \boldsymbol{\alpha}} = 0 \Leftrightarrow \left(\mathbf{K} + \frac{1}{2\gamma} \mathbf{I} \right) \boldsymbol{\alpha} = \mathbf{y}$$

Où **I** est la matrice identité (N, N). On retrouve alors la relation (9), obtenue par utilisation du théorème de représentation en l'absence de biais, et l'on est ramené au calcul de la pseudo-inverse d'une matrice (N, N). [21]

III.5.1.3 Recherche de la solution par résolution des équations de Karush, Kuhn et Tucker

Une troisième approche est également envisageable : elle s'appuie à nouveau sur les caractéristiques du problème d'optimisation, mais les exploite à travers des conditions particulières, que nous allons décrire.

Revenons au problème initial :

$$\min_{\mathbf{w},b} \frac{1}{2} \|\mathbf{w}\|^2 + \gamma \sum_{i=1}^{N} \xi_i^2$$

$$\text{sous } \xi_i = y_i - f(\mathbf{x}_i) = y_i - \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - b, \ \forall i \in [\![1, N]\!]$$

Outre leur caractère différentiable, la convexité de la fonction objectif et des fonctions exprimant les contraintes d'égalité, nous permet d'affirmer que les conditions de Karush, Kuhn et Tucker (KKT) sont nécessaires et suffisantes [20].

En d'autres termes, $(\mathbf{w}, b)^T$ est minimum global du problème d'optimisation ci-dessus si et seulement si il existe des réels $\boldsymbol{\alpha}$ et b tels que [21] :

$$\begin{cases} \dfrac{\partial L}{\partial \mathbf{w}} = 0 \Leftrightarrow \mathbf{w} - \sum_{i=1}^{N} \alpha_i \phi(\mathbf{x}_i) = 0 \Leftrightarrow \mathbf{w} = \sum_{i=1}^{N} \alpha_i \phi(\mathbf{x}_i) \\ \dfrac{\partial L}{\partial b} = 0 \Leftrightarrow \sum_{i=1}^{N} \alpha_i = 0 \\ \dfrac{\partial L}{\partial \xi_i} = 0 \Leftrightarrow 2\gamma\xi_i - \alpha_i = 0 \Leftrightarrow \xi_i = \dfrac{\alpha_i}{2\gamma}, \forall i \in [\![1, N]\!] \\ \xi_i = y_i - \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - b, \forall i \in [\![1, N]\!] \end{cases}$$

En éliminant **w** et ξ dans les *N* dernières de ces équations, on se ramène ainsi à la résolution du système de *N*+1 équations linéaires suivant :

$$\begin{bmatrix} \mathbf{K} + \dfrac{1}{2\gamma}\mathbf{I} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix}$$

Où **I** est la matrice identité (*N*, *N*).

La fonction *f*, solution du problème initial, s'écrira finalement :

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b$$

Les LSSVM aboutissent donc à un développement de la solution sur l'ensemble des exemples d'apprentissage : du fait du choix de la fonction de coût, qui s'est traduit par des contraintes d'égalité dans la formulation du problème d'optimisation, tous les multiplicateurs de Lagrange sont non nuls. Dans ce cas, on dit que tous les exemples de l'ensemble d'apprentissage sont des vecteurs supports. Cette caractéristique peut devenir problématique, du point de vue du temps de calcul ou de la capacité mémoire, pour des bases de données de taille importante. Une variante, exposée ci-dessous, permet de ne prendre en considération, dans le développement de **w**, qu'un nombre limité d'exemples. [21]

III.6. Les Machines à Vecteurs Supports (SVM) pour la régression

III.6.1 Formulation du problème dans le cas linéaire

Dans le cas des LSSVM, nous allons à présent décrire les machines à vecteurs supports sous leur forme conventionnelle.

On dispose d'un ensemble de données

$$D = \left\{ \left(\mathbf{x}_i, y_i \right) \in X \times Y \right\}_{i=1}^{N}, \quad \text{où} \quad X \subset \mathbb{R}^d \text{ et } Y \subset \mathbb{R}.$$

L'objectif est de trouver une fonction *f* dans un espace de Hilbert *H*, par minimisation du risque structurel (SRM), établissant une relation entre les variables **x** et la grandeur à modéliser *y*, à partir de l'ensemble de mesures *D*. La fonctionnelle que l'on cherche à optimiser s'écrit :

$$F[f] = C\sum_{i=1}^{N} \left| y_i - f(\mathbf{x}_i) \right|_{\varepsilon} + \frac{1}{2}\|f\|_H^2 \qquad (17)$$

Où $\|\cdot\|_H$ *H* est une norme sur *H*, et où $\varepsilon\,|\cdot|_{\varepsilon}$ représente la fonction de coût ε-insensible suivante :

$$V\left(y, f(\mathbf{x})\right) = \left| y - f(\mathbf{x}) \right|_{\varepsilon} = \begin{cases} 0 & \text{si } |\mathbf{x}| < \varepsilon \\ |\mathbf{x}| - \varepsilon & \text{sinon} \end{cases}$$

Cette fonction de coût ne pénalise que les déviations supérieures à ε entre les sorties mesurées et estimées.

Nous allons, dans un premier temps, réaliser un modèle linéaire, en cherchant la fonction de régression *f* sous la forme suivante :

$f(\mathbf{x}) = \mathbf{w}.\mathbf{x} + \boldsymbol{b}$, où $\mathbf{w}.\mathbf{x}$ est un produit scalaire dans l'espace *X*.

En introduisant des variables d'erreurs ξ_i et ξ_i^* (pour les erreurs respectivement positives et négatives), on parvient à intégrer la fonction de coût ε-insensible au sein du problème d'optimisation suivant [21]:

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right)$$
$$\text{sous} \begin{cases} y_i - \mathbf{w}\cdot\mathbf{x}_i - b \le \varepsilon + \xi_i \\ \mathbf{w}\cdot\mathbf{x}_i + b - y_i \le \varepsilon + \xi_i^*, \forall i \in [\![1, N]\!] \\ \xi_i, \xi_i^* \ge 0 \end{cases} \quad (18)$$

Le premier terme de la fonction objectif limite la capacité de la fonction de régression (terme de régularisation), tandis que le second pénalise les erreurs supérieures à ε, qu'elles soient par valeur positive ou négative : ce compromis biais/variance est contrôlé par la constante C.

Le lagrangien associé à la fonction objectif sous contrainte ci-dessus est :

$$L\left(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\xi}^*; \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\eta}, \boldsymbol{\eta}^*\right) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) - \sum_{i=1}^{N}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right)$$
$$-\sum_{i=1}^{N}\alpha_i\left(\varepsilon + \xi_i - y_i + \mathbf{w}\cdot\mathbf{x}_i + b\right) - \sum_{i=1}^{N}\alpha_i^*\left(\varepsilon + \xi_i^* + y_i - \mathbf{w}\cdot\mathbf{x}_i - b\right) \quad (19)$$

où η_i ,η_i^* ,α_i ,α_i^* sont les multiplicateurs de Lagrange positifs (variables duales) associés aux contraintes inégalités du problème initial (18).

On peut montrer que, pour le problème d'optimisation ci-dessus (fonction de coût convexe, contraintes linéaires), les conditions complémentaires de KKT sont nécessaires et suffisantes.

De plus, l'existence d'un point selle impose que le minimum par rapport aux variables primales ($\mathbf{w}$, b, $\boldsymbol{\xi}$, $\boldsymbol{\xi}^*$) coïncide avec le maximum par rapport aux variables duales ($\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^*$, $\boldsymbol{\eta}$, $\boldsymbol{\eta}^*$). En imposant la stationnarité du lagrangien vis-à-vis des variables primales, on obtient :

$$\begin{cases} \dfrac{\partial L}{\partial b} = 0 \Leftrightarrow \displaystyle\sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right) = 0 \\ \dfrac{\partial L}{\partial \mathbf{w}} = 0 \Leftrightarrow \mathbf{w} = \displaystyle\sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right)\mathbf{x}_i \\ \dfrac{\partial L}{\partial \xi_i^{(*)}} = 0 \Leftrightarrow C = \alpha_i^{(*)} + \eta_i^{(*)}, \ \forall i \in [\![1, N]\!] \end{cases}$$

Où $\boldsymbol{\alpha}^{(*)}$ représente $\boldsymbol{\alpha}$ ou $\boldsymbol{\alpha}^*$.

En substituant les relations ci-dessus dans l'expression du Lagrangien (19), on aboutit au problème d'optimisation suivant, qui n'est fonction que des variables duales :

$$\max_{\alpha,\alpha^*} -\frac{1}{2}\sum_{i,j=1}^{N}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\mathbf{x}_i\cdot\mathbf{x}_j-\varepsilon\sum_{i=1}^{N}\left(\alpha_i+\alpha_i^*\right)+\sum_{i=1}^{N}y_i\left(\alpha_i-\alpha_i^*\right)$$

$$sous\begin{cases}\sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)=0\\ \alpha_i,\alpha_i^*\in[0,C]\end{cases} \qquad (20)$$

On est ainsi ramené à la résolution d'un problème d'optimisation quadratique, dont les contraintes d'égalité sont plus aisées à manipuler que celles du problème primal.

Les conditions complémentaires de KKT s'écrivent :

$$\begin{cases}\alpha_i\left(\varepsilon+\xi_i-y_i+\mathbf{w}\cdot\mathbf{x}_i+b\right)=0\\ \alpha_i^*\left(\varepsilon+\xi_i^*+y_i-\mathbf{w}\cdot\mathbf{x}_i-b\right)=0\\ \left(C-\alpha_i\right)\xi_i=0\\ \left(C-\alpha_i^*\right)\xi_i^*=0\end{cases},\forall i\in[\![1,N]\!]$$

Seuls les coefficients α_i relatifs aux exemples pour lesquels les contraintes sont saturées peuvent être non nuls : on appelle ces points les vecteurs supports. Les conditions de KKT permettent également de déterminer le biais *b*.

Après détermination des coefficients α_i, α_i par résolution du problème d'optimisation (20), la fonction *f* recherchée s'écrit :

$$f(\mathbf{x})=\sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)\mathbf{x}_i\cdot\mathbf{x}+b \qquad (21)$$

Cette fonction s'écrit donc comme une somme, pondérée par les coefficients, de produits scalaires sur l'ensemble des exemples de la base d'apprentissage. Cette caractéristique, qui est conforme au résultat énoncé dans la relation (2), va à présent être exploitée pour étendre ces résultats au cas non linéaire (« astuce des noyaux »).

Remarque : la différence fondamentale entre les SVM conventionnelles et les LSSVM introduites au paragraphe précédent réside dans la fonction de coût utilisée : plutôt

que d'employer la fonction de coût quadratique, on se sert ici d'une fonction de coût ε-insensible. [21]

III.6.2 Extension au cas non linéaire

Comme nous l'avons indiqué dans le paragraphe 2, il est possible de traiter le problème non linéaire de manière essentiellement identique à ce qui vient d'être fait dans le cas linéaire, en cherchant la solution dans un RKHS muni d'un noyau convenablement choisi. Ceci revient à effectuer un « prétraitement » des variables, qui consiste en une transformation non linéaire de l'espace des variables *X* vers un espace image *F*, grâce à une fonction $\varphi : X \rightarrow F$ associée au noyau *K*. Les résultats obtenus au paragraphe précédent dans l'espace initial *X* sont alors transposables à l'espace image *F*, le noyau choisi définissant la « distance » entre les objets dans l'espace des caractéristiques. [21]

Le problème d'optimisation (20) devient donc :

$$\max_{\alpha,\alpha^*} -\frac{1}{2}\sum_{i,j=1}^{N}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\left\langle\phi\left(\mathbf{x}_i\right),\phi\left(\mathbf{x}_j\right)\right\rangle-\varepsilon\sum_{i=1}^{N}\left(\alpha_i+\alpha_i^*\right)+\sum_{i=1}^{N}y_i\left(\alpha_i-\alpha_i^*\right)$$

$$sous\begin{cases}\sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)=0\\ \alpha_i,\alpha_i^*\in\left[0,C\right]\end{cases} \qquad (22)$$

Notons que la fonction de régression, qui par construction sera non linéaire vis-à-vis des entrées, reste toutefois linéaire en ses paramètres. La propriété de convexité du problème d'optimisation (22) garantit donc l'unicité de la solution. [21]

D'autre part, la transformation φ, qui n'intervient dans le problème (22) qu'à travers des produits scalaires dans l'espace image *F*, peut être définie de manière implicite au moyen d'une fonction noyau SV-admissible *K*, telle que :

$(\varphi(x_i), \varphi(x_j)) = (K(\boldsymbol{x_i}, \boldsymbol{x_j}))$.Comme indiqué dans le paragraphe 2, il n'est donc pas nécessaire de connaître explicitement la transformation φ : celle-ci est entièrement définie par le choix du noyau *K* qui lui est associé. [21]

La solution du problème d'optimisation (22) s'écrira finalement :

$$f\left(\mathbf{x}\right)=\left\langle\mathbf{w},\phi\left(\mathbf{x}\right)\right\rangle+b=\sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)K\left(\mathbf{x}_i,\mathbf{x}\right)+b \qquad (23)$$

III.6.3 Choix des hyperparamètres

La résolution des problèmes d'optimisation quadratique (20) ou (22) s'effectue à l'aide d'algorithmes classiques1 (on pourra se référer à [23] pour un inventaire des principaux algorithmes utilisés). Elle nécessite la définition d'un ensemble d'hyperparamètres, regroupés dans le vecteur $\boldsymbol{\theta}_{hyper}$, dont les composantes sont :

• ε : la valeur d'insensibilité

• C : le compromis entre régularisation et pénalisation des erreurs

• les différents paramètres relatifs au noyau utilisé (écart type des gaussiennes dans le cas de noyaux gaussiens)

Afin de choisir au mieux ces hyperparamètres, des méthodes classiques peuvent être mises en œuvre (notamment la validation croisée). Il est également possible, dans ce cas particulier des SVM, d'établir des bornes à certains indices de performance (erreur de validation, erreur de leave-one-out, …), et d'optimiser ceux-ci par une descente de gradient vis-à-vis des hyper paramètres [21].

La solution obtenue par un algorithme de SVM, c'est-à-dire l'ensemble des coefficients $\{\alpha, \alpha^*\}_{i=1}^{N}$, dépend évidemment des hyperparamètres utilisés. Toutefois, le résultat suivant simplifie grandement le problème [20].

Théorème : soit $\mathbf{v}_\theta$ un vecteur de dimension (n x 1) et $\mathbf{P}_\theta$ une matrice (n x n), dépendant continûment du paramètre θ. Considérons la fonction :

$$L(\theta) = \max_{\mathbf{x} \in F} \left(\mathbf{x}^T \mathbf{v}_\theta - \frac{1}{2} \mathbf{x}^T \mathbf{P}_\theta \mathbf{x} \right)$$

$$où\ F = \left\{ \mathbf{x} : \mathbf{b}^T \mathbf{x} = c, \mathbf{x} \geq 0 \right\}$$

Soit $\mathbf{x}_o$ la valeur en laquelle le maximum est atteint. Si ce maximum est unique, alors on a :

$$\frac{\partial L}{\partial \theta} = \mathbf{x}_o^T \frac{\partial \mathbf{v}_\theta}{\partial \theta} - \frac{1}{2} \mathbf{x}_o^T \frac{\partial \mathbf{P}_\theta}{\partial \theta} \mathbf{x}_o$$

En d'autres termes, on peut différentier L par rapport à θ comme si $\mathbf{x}_o$ ne dépendait pas de θ.

Ce résultat reste valable si certaines contraintes dans la définition de *F* sont supprimées. [21]

Il est donc possible de rechercher, en premier lieu, la solution SVM avec un jeu particulier d'hyper paramètres **h**yper, puis de chercher, par descente de gradient, le minimum d'un indice de performance convenable par rapport à **h**yper, par différentiation de celui-ci comme si les coefficients $\alpha_i^{(*)}$ ne dépendaient pas de **h**yper. [21]

III.6.4.Choix de la fonction de coût

III.6.4.1 Fonctions de coût et modèles de bruit associés

Nous avons souligné, au paragraphe 4.1, que la différence fondamentale entre les SVM et les LSSVM réside dans le choix de la fonction de coût. Indépendamment du fait que ce choix aboutit à des problèmes d'optimisation (et donc à des solutions) différents, nous allons montrer que, sous certaines conditions, le choix de la fonction de coût revient à faire une hypothèse sur la nature du bruit qui affecte les données.

Supposons que les données $D = \{(x_1, y_1) \in X*Y\}_{i=1}^{N}$ dont nous disposons sont engendrées par une fonction *f* purement déterministe, à laquelle s'ajoute un bruit additif, représenté par une variable aléatoire ξ suivant une distribution de probabilité *p* [21] *:*

$$Y_i = f(x_i) + \xi_i$$

Nous supposons de plus la fonction *f* connue : l'objectif est alors de déterminer une fonction de coût optimale, au sens où sa minimisation correspond au maximum de vraisemblance de l'estimation $D_f = \{(x_1, f(x_1)),\ldots, (x_N, f(x_N))\}$ connaissant *D*.

Sous les hypothèses de bruit additif et de réalisations indépendantes et identiquement Distribuées, la vraisemblance de l'estimation *Df* connaissant *D* est définie par :

$$P(D_f \mid D) = \prod_{i=1}^{N} P(f(\mathbf{x}_i) \mid (\mathbf{x}_i, y_i)) = \prod_{i=1}^{N} P(f(\mathbf{x}_i) \mid y_i)$$

L'estimation *Df* étant réalisée par la fonction génératrice des données *f*, la relation ci-dessus devient :

$$P\left(D_f \mid D\right)=\prod_{i=1}^{N} p\left(y_i - f\left(\mathbf{x}_i\right)\right)$$

où *p* est la distribution de probabilité associée au bruit qui entache les données *D*.

L'estimateur du maximum de vraisemblance est celui pour lequel $P(D_f/D)$ est maximum, soit, de manière équivalente, celui pour lequel

$-\log(D_f/D) = -\sum_{i=1}^{n} log(p\ (yi-f\ (xi)))$ est minimum.

$$V\ (y,\ f(x))= -\log(P(y-f(x)))$$

La fonction de coût optimale, au sens du maximum de vraisemblance, est alors définie comme :

$$V\ (y,\ f\ (\mathbf{x})) = -\log\ (p\ (y - f\ (\mathbf{x})))$$

L'utilisation de cette fonction de coût dans un problème d'apprentissage assure donc, en principe, la possibilité de retrouver la fonction génératrice des données *f*, à partir de l'ensemble de données *D*.

Le tableau III.1 : indique les fonctions de coût usuelles ainsi que les distributions associées du bruit. [21]

	Fonction de coût	**Distribution du bruit associée**
ε-insensible	$V(\xi)=\lvert\xi\rvert_\varepsilon$	$p(\xi)=\frac{1}{2(1+\varepsilon)}\exp\left(-\lvert\xi\rvert_\varepsilon\right)$
Laplacienne	$V(\xi)=\lvert\xi\rvert$	$p(\xi)=\frac{1}{2}\exp\left(-\lvert\xi\rvert\right)$
Gaussienne	$V(\xi)=\frac{1}{2}\xi^2$	$p(\xi)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\xi^2}{2}\right)$
Fonction de coût robuste de Huber	$V(\xi)=\begin{cases}\frac{1}{2\sigma}\xi^2 & \text{si } \lvert\xi\rvert\le\sigma\\ \lvert\xi\rvert-\frac{\sigma}{2} & \text{sinon}\end{cases}$	$p(\xi)\propto\begin{cases}\exp\left(-\frac{\xi^2}{2\sigma}\right) & \text{si } \lvert\xi\rvert\le\sigma\\ \exp\left(\frac{\sigma}{2}-\lvert\xi\rvert\right) & \text{sinon}\end{cases}$
Polynomiale	$V(\xi)=\frac{1}{p}\lvert\xi\rvert^p$	$p(\xi)=\frac{p}{2\Gamma(1/p)}\exp\left(-\lvert\xi\rvert^p\right)$
Polynomiale par morceaux	$V(\xi)=\begin{cases}\frac{1}{p\sigma^{p-1}}\xi^p & \text{si } \lvert\xi\rvert\le\sigma\\ \lvert\xi\rvert-\sigma\frac{p-1}{p} & \text{sinon}\end{cases}$	$p(\xi)\propto\begin{cases}\exp\left(-\frac{\xi^p}{p\sigma^{p-1}}\right) & \text{si } \lvert\xi\rvert\le\sigma\\ \exp\left(\sigma\frac{p-1}{p}-\lvert\xi\rvert\right) & \text{sinon}\end{cases}$

Il est à noter que le choix de la fonction ε-insensible de Vapnik, fréquemment utilisée dans la régression par SVM, est équivalent à une hypothèse de bruit additif et gaussien, dont la variance et la moyenne des gaussiennes sont des variables aléatoires. [21]

III.6.4.2 Problème d'optimisation pour une fonction de coût générale

Supposons que la fonction de coût V choisie vérifie les propriétés suivantes :

• V est symétrique

• la dérivée première de V présente au plus deux discontinuités à $\pm\varepsilon$

• V est nulle dans l'intervalle $[-\varepsilon, +\varepsilon]$

Nous pouvons alors écrire : $V(y, f(x))=V'(|y - f(x)|_{\varepsilon})$

En raisonnant de la même manière que précédemment, on aboutit au problème d'optimisation suivant (l'extension au cas non linéaire se faisant par introduction d'une fonction noyau en lieu et place des produits scalaires) [20]:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\left(\widetilde{V}(\xi_i)+\widetilde{V}(\xi_i^*)\right)$$

$$\text{sous}\begin{cases} y_i - \mathbf{w}\cdot\mathbf{x}_i - b \le \varepsilon + \xi_i \\ \mathbf{w}\cdot\mathbf{x}_i + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases}, \forall i \in [\![1, N]\!]$$

Ce problème d'optimisation peut se réécrire :

$$\max_{\boldsymbol{\alpha},\boldsymbol{\alpha}^*} \frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\mathbf{x}_i\cdot\mathbf{x}_j + \sum_{i=1}^{N}\left[y_i(\alpha_i-\alpha_i^*)-\varepsilon(\alpha_i+\alpha_i^*)\right] + C\sum_{i=1}^{N}\left[T(\xi_i)+T(\xi_i^*)\right]$$

$$sous\begin{cases} \sum_{i=1}^{N}(\alpha_i-\alpha_i^*)=0 \\ \boldsymbol{\alpha} \le C\dfrac{\partial\widetilde{V}}{\partial\boldsymbol{\xi}} \\ \boldsymbol{\xi} = \inf\left\{\xi \,\middle|\, C\dfrac{\partial\widetilde{V}}{\partial\boldsymbol{\xi}} \ge \boldsymbol{\alpha}\right\} \\ \boldsymbol{\alpha}, \boldsymbol{\xi} \ge 0 \end{cases} \qquad (24)$$

où $\mathbf{w} = \sum_{i=1}^{N}(\alpha_i-\alpha_i^*)\mathbf{x}_i, \quad T(\xi) = \widetilde{V}(\xi) - \xi\dfrac{\partial\widetilde{V}}{\partial\xi}(\xi)$

Remarque : la condition de nullité dans l'intervalle [-ε, + ε] peut être contournée par ajout de variables « élastiques» supplémentaires (« slack variables »). Il est également possible de définir une fonction de coût et une valeur de ε différentes pour chaque exemple. Plus de deux discontinuités sont possibles, au prix de multiplicateurs de Lagrange supplémentaires. [21]

III.7. Les domaines d'applications

SVM est une méthode de classification qui montre de bonnes performances dans la résolution de problèmes variés. Cette méthode a montré son efficacité dans de nombreux domaines d'applications tels que le traitement d'image, la catégorisation de textes ou le diagnostiques médicales et ce même sur des ensembles de données de très grandes dimensions.

La réalisation d'un programme d'apprentissage par SVM se ramène à résoudre un problème d'optimisation impliquant un système de résolution dans un espace de dimension conséquente. L'utilisation de ces programmes revient surtout à sélectionner une bonne famille de fonctions noyau et à régler les paramètres de ces fonctions. Ces choix sont le plus souvent faits par une technique de validation croisée, dans laquelle on estime la performance du système en la mesurant sur des exemples n'ayant pas été utilisés en cours d'apprentissage.

L'idée est de chercher les paramètres permettant d'obtenir la performance maximale. Si la mise en œuvre d'un algorithme de SVM est en général peu coûteuse en temps, il faut cependant compter que la recherche des meilleurs paramètres peut requérir des phases de test assez longues [21].

III.8.Conclusion

Dans ce chapitre, nous avons tenté de présenter de manière simple et résumé le concept de système d'apprentissage introduit par Vladimir Vapnik, les Machines à Vecteurs de Support. Nous avons donné une vision générale et une vision purement mathématiques des SVM. [20]

Cette méthode de classification est basée sur la recherche d'un hyperplan qui permet de séparer au mieux des ensembles de données. Nous avons exposé les cas linéairement séparable et les cas non linéairement séparables qui nécessitent l'utilisation de fonction noyau (kernel) pour changer d'espace. Cette méthode est

applicable pour des tâches de classification à deux classes, mais il existe des extensions pour la classification multi classe. [20]

Nous nous sommes ensuite intéressés aux différents domaines d'application. Il existe des extensions que nous n'avons pas présentées, parmi lesquelles l'utilisation des SVM pour des tâches de régression, c'est-à-dire de prédiction d'une variable continue en fonction d'autres variables, comme c'est le cas par exemple dans la prédiction de consommation électrique en fonction de la période de l'année, de la température, etc. [20]

Le champ d'application des SVM est donc large et représente une méthode de classification intéressante.

Chapitre IV : Résultats et Discussions

IV.1.Introduction

La littérature sur les liquides ioniques propose aujourd'hui une base de données de mesures expérimentales des différentes propriétés physico-chimiques et des équilibres de phases. L'une de ces propriétés physiques ayant une relation avec les équilibres de phase est le coefficient d'activité à dilution infinie, (CADI) (Infinite Diultion Activity Coefficient, IDAC) de l'eau et de plusieurs solutés organiques dans différents liquides ioniques. Parmi les méthodes de modélisation qui ont montré une capacité de corrélation des données dans plusieurs disciplines d'ingénierie est la méthode de régression hybride moindres carrés machines de vecteur de support (Least Square Support Vector Machine, LSSVM). Dans ce chapitre nous allons rapporter l'application de la technique LSSVM pour la régression du CADI de l'eau et quelques alcanes et cyclo-alcanes dans plusieurs familles de liquides ioniques.La méthodologie utilisée consiste des étapes suivantes :

- la collecte des données expérimentales du CADI et des descripteurs moléculaire des Lis et des solutés sélectionnés,
- choix des variables d'entrées,
- normalisation des données
- optimisation des paramètres du modèle LSSVM
- évaluation de la capacitéde corrélation du modèle optimisé.

Ce travail concrétise les recommandations d'un projet antérieur par l'élargissement de la base de données en termes de liquide ioniques sélectionnés, le nombre de solutés et la méthode de modélisation.

IV.2. Elaboration de la base de données

La base de données élaborée regroupe les données expérimentales du CADI de l'eau et de quelques solutés organiques dans différents Lis à différentes températures, les données des profiles sigma des cations et des anions constituants les Lis sélectionnés et les descripteurs moléculaires d'Abraham exprimant la relation linéaire de l'énergie de solvatation des solutés considérées.

IV.2.1. Base de données du CADI

Les données de mesures expérimentales du CADI d'un soluté donné dans un liquide ionique donné sont rapportées, généralement, dans la littérature en fonction de la température. Une très grande partie de ces données sont dues aux travaux d'U. Domańska et ses collaborateurs de la faculté de chimie de l'université de Varsovie en Pologne durant la période 2005-2016. Le tableau IV.1 montre la liste des 80 Lis sélectionnées, le nombre de point expérimentaux par liquide ionique et la référence bibliographique.

Tableau IV.1 Laliste des Lis sélectionnées, le nombre de point expérimentaux par liquide ionique et la référence bibliographique.

N°	Li	Nombre de points	Réf
1	[IM-1,1] [NTF2]	24	30
2	[IM-1,1] [PO2-O1,O1]	59	31-32
3	[IM-10,1] [BF4]	49	33
4	[IM-10,1] [TCB]	69	34
5	[IM-10,10,1] [NTF2]	27	35
6	[IM-1O6,1] [NTF2]	87	36
7	[IM-1O6,1O6] [NTF2]	88	36
8	[IM-2,1] [AC]	20	37
9	[IM-2,1] [BF4]	65	26-29
10	[IM-2,1] [CCN3]	72	38
11	[IM-2,1] [DCA]	39	39
12	[IM-2,1] [FAP]	107	41-40
13	[IM-2,1] [NTF2]	62	30-43-9-42
14	[IM-2,1] [OTF]	24	44
15	[IM-2,1] [PO2-O1,O1]	48	32
16	[IM-2,1] [SCN]	66	45
17	[IM-2,1] [SO3-1]	110	47-35-46
18	[IM-2,1] [SO3-PH1]	5	48
19	[IM-2,1] [TCB]	139	50-49-51
20	[IM-2,1] [TFA]	73	52
21	[IM-2OH,1] [BF4]	35	28
22	[IM-2OH,1][DCA]	50	53
23	[IM-2OH,1][FAP]	92	54-55
24	[IM-2OH,1][NTF2]	52	57
25	[IM-4,1] [AC]	15	58-59-60
26	[IM-4,1] [BR]	111	59-60
27	[IM-4,1] [CCN3]	54	61
28	[IM-4,1] [CL]	41	62-59-60
29	[IM-4,1] [DCA]	72	63
30	[IM-4,1] [NTF2]	121	30-64-66-65-58
31	[IM-4,1] [OTF]	166	67-69-59-70
32	[IM-4,1] [PO2-O1,O1]	40	62-60

33	[IM-4,1] [SCN]	69	70-59
34	[IM-4,1] [SO3-1]	45	62-58-60-59
35	[IM-4,1] [SO3-PH1]	41	71-59
36	[IM-4,1] [TFA]	5	59-60
37	[IM-4,1] [BF4]	111	86-85-25-26-66
38	[IM-6,1] [BF4]	40	25-26-27
39	[IM-6,1] [CL]	20	72
40	[IM-6,1] [FAP]	33	35
41	[IM-6,1] [NTF2]	70	74-73-79
42	[IM-6,1] [OTF]	80	76-79-69
43	[IM-6,1] [SCN]	69	78
44	[IM-6,1] [TCB]	72	76
45	[IM-6,1,1] [NTF2]	40	79
46	[MO-2O1,1] [FAP]	72	80
47	[MO-2O1,1] [NTF2]	72	80
48	[MO-4,1] [CCN3]	72	82
49	[N-2OH,1,1,1] [NTF2]	66	83
50	[P-4I,4I,4I,1] [SO3-PH1]	46	111
51	[PIP-3,1] [NTF2]	60	87
52	[PIP-4,1] [NTF2]	72	88
53	[PIP-4,1] [SCN]	49	89
54	[PIP-5,1] [NTF2]	66	90
55	[PIP-6,1] [NTF2]	72	90
56	[PY-2] [NTF2]	42	94
57	[PY-2] [TFA]	27	31
58	[PY-3OH] [FAP]	64	92
59	[PY-3OH] [NTF2]	81	93
60	[PY-4] [NTF2]	42	94
61	[PY-4,1[3]] [OTF]	64	95
62	[PY-4,1[3]] [SO3-PH1]	32	96
63	[PY-4,1[4]] [BF4]	42	23-24
64	[PY-4,1[4]] [CCN3]	54	61
65	[PY-4,1[4]] [DCA]	74	97
66	[PY-4,1[4]] [NTF2]	88	98
67	[PY-4,1[4]] [SCN]	68	99
68	[PY-6,1[3]] [SO3-PH1]	48	100
69	[PYR-2O1,1] [FAP]	92	132-55
70	[PYR-2O1,1] [NTF2]	72	101
71	[PYR-4,1] [CCN3]	72	102
72	[PYR-4,1] [DCA]	20	103
73	[PYR-4,1] [FAP]	66	104
74	[PYR-4,1] [NTF2]	57	74-105
75	[PYR-4,1] [OTF]	89	106-79
76	[PYR-4,1] [SCN]	61	99
77	[PYR-4,1] [TCB]	72	107-108
78	[QUINI-6] [SCN]	75	109
79	[QUINUC-8] [NTF2]	60	110
80	[S-2,2,2] [NTF2]	87	98

	Total	**4882**	

La liste des cations et des anions, avec abréviation et nomenclature, qui entre dans la constitution des 80 Lis sélectionnés sont cités dans les tableaux IV.2 et IV.3 respectivement.

Tableau IV.2 Liste des cations

N^0	**Nom de cations**	**Abréviation**	**Famille**
1	1,3-dimethyl-1H-imidazol-3-ium	[IM-1,1]	Imidazolium
2	1-decyl-3-methyl-1H-imidazol-3-ium	[IM-10,1]	
3	1,3-didecyl-2-methylimidazolium	[IM-10, 10,1]	
4	3-[(hexyloxy)methyl]-1-methyl-1H-imidazol-3-ium	[IM-1O6,1]	
5	1,3-bis[(hexyloxy)methyl]-1H-imidazol-3-ium	[IM-1O6,1O6]	
6	1-ethyl-3-methyl-1H-imidazol-3-ium	[IM-2,1]	
7	1-(2-hydroxyethyl)-3-methyl-1H-imidazol-3-ium	[IM-2OH,1]	
8	1-butyl-3-methyl-1H-imidazol-3-ium	[IM-4,1]	
9	3-hexyl-1-methyl-1H-imidazol-3-ium	[IM-6,1]	
10	3-hexyl-1,2-dimethyl-1H-imidazol-3-ium	[IM-6,1,1]	
11	4-(2-methoxyethyl)-4-methylmorpholin-4-ium	[MO-2O1,1]	Morpholinium
12	4-butyl-4-methylmorpholin-4-ium	[MO-4,1]	
13	(2-hydroxyethyl)trimethylazanium	[N-2OH,1,1,1]	Ammonium
14	methyltris(2-methylpropyl)phosphanium	[P-4I,4I,4I,1]	Phosphonium
15	1-methyl-1-propylpiperidin-1-ium	[PIP-3,1]	Piperidinium
16	1-butyl-1-methylpiperidin-1-ium	[PIP-4,1]	
17	1-methyl-1-pentylpiperidin-1-ium	[PIP-5,1]	
18	1-hexyl-1-methylpiperidin-1-ium	[PIP-6,1]	
19	1-ethylpyridin-1-ium	[PY-2]	Pyridinium
20	1-(3-hydroxypropyl)pyridin-1-ium	[PY-3OH]	
21	1-butylpyridin-1-ium	[PY-4]	
22	1-butyl-3-methylpyridin-1-ium	[PY-4,1[3]]	
23	1-butyl-4-methylpyridin-1-ium	[PY-4,1[4]]	
24	1-hexyl-3-methylpyridin-1-ium	[PY-6,1[3]]	
25	1-(2-methoxyethyl)-1-methylpyrrolidin-1-ium	[PYR-2O1,1]	Pyrrolidinium
26	1-butyl-1-methylpyrrolidin-1-ium	[PYR-4,1]	
27	2-hexylisoquinolin-2-ium	[QUINI-6]	Quinolinium
28	2-octylisoquinolin-2-ium	[QUINUC-8]	Bicyclique
29	triethylsulfanium	[S-2,2,2]	Sulfonium

Les 4882 points expérimentaux concernent le CADI de l'eau, de 10 alcanes (n-hexane, n-heptane, n-octane, n-nonane, n-décane, n-pentane, n-undecane, n-tridécane, n-tetradécane, n-dodécane) et de 6 cycloalcanes (cyclohexane, méthylcyclohexane, cyclopentane, cycloheptane, cyclooctane, méthylcyclopentane) dans les Lis sélectionnés.

Tableau IV.3 Liste des anions

N^0	**Nom de l'anions**	**Abréviation**
1	bis(pentafluroethylsulfonyl)azanide	[NTF2]
2	dimethyl phosphate	[PO2-O1,O1]
3	tetrafluoroborate	[BF4]
4	tetracyanoboranuide	[TCB]
5	Acetate	[Ac]
6	tricyanomethide	[CCN3]
7	dicyanoazanide	[DCA]
8	trifluorotris(pentafluoroethyl) phosphanuide	[FAP]
9	trifluoromethanesulfonate	[OTF]
10	thiocyanate	[SCN]
11	methanesulfonate	[SO3-1]
12	4-methylbenzene-1-sulfonate	[SO3-PH1]
13	trifluoroacetate	[TFA]
14	Chloride	[CL]

IV.2.2. Base de données des descripteurs moléculaires des cations et des anions

Suite aux résultats satisfaisant obtenu dans un travail antérieur [18], nous avons adoptés les mêmes descripteurs moléculaires c.-à-d. les profiles σ issue de la chimie quantiques. Ses descripteurs dérivent de la théorie COSMO-RS qui est considérée comme modèle de solvatation continue qui combine la théorie chimique quantique, les modèles de continuum diélectrique, les interactions de surface et la thermodynamique statistique. Il calcule les interactions moléculaires à partir des densités de charge de blindage (polarisation) sur les segments de surface moléculaires. Les densités de charge de polarisation sont calculées par des méthodes de chimie quantiques. Les calculs de chimie quantique fournissent un segment de surface discret autour d'une seule molécule intégrée dans un conducteur virtuel. La surface de chaque segment est caractérisée par sa surface et la densité de charge de blindage du segment, ce qui tient compte de l'interaction électrostatique de la molécule de soluté par son entourage (qui dans un conducteur virtuel est un criblage parfait) et de la rétro-polarisation de la molécule de soluté. La densité de charge de criblage est habituellement donnée dans le diagramme de probabilité de la distribution de la densité de charge statistique à la surface d'une molécule, et également connue sous le nom de profil σ. Le profil σ d'une molécule montre la probabilité de la quantité relative de surface avec la polarité σ sur la surface de la molécule, et il est considéré comme les propriétés caractéristiques de la molécule (Figure. IV.1).

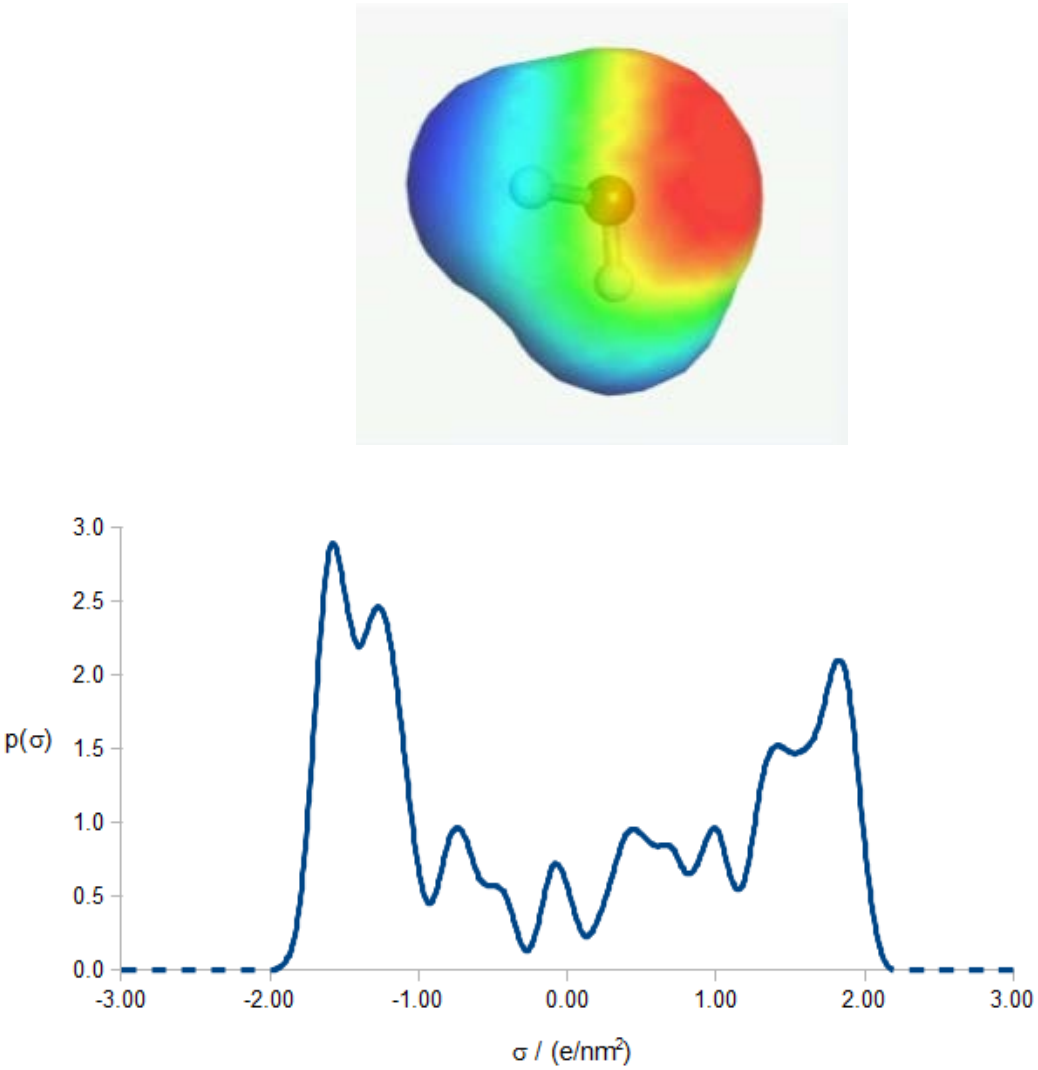


Figure IV.1: Profil σ de la molécule d'eau.

En d'autres termes, le profil σ d'une molécule comprend les informations chimiques principales nécessaires pour prédire les interactions électrostatiques, de liaison hydrogène et de dispersion possibles de la molécule dans un fluide. Les profils σ se sont révélés être un descripteur moléculaire efficace pour établir des modèles de QSPR capables de prédire les propriétés physiques, chimiques et toxicologiques des Lis.

La première étape du développement actuel du modèle, est la construction de descripteurs moléculaires appropriés à partir des profils σ des cations et des anions des Lis étudiés. Généralement, le profil σ des cations se situe dans la région non polaire ($-1,0 < \sigma < 1,0$) e/nm^2 et dans la région donneur de la liaison hydrogène ($\sigma < -1,0$ e/nm^2). Cependant, pour les anions, les pics du profil σ sont habituellement dans la région accepteur de la liaison hydrogène ($\sigma < 1,0$ e/nm^2) et dans la région non polaire ($-1,0 < \sigma < 1,0$) e/nm^2. La position, la largeur et la hauteur des pics varient en fonction de la nature des cations et des anions, de la longueur de la chaîne

alkyle et de la présence du groupe fonctionnel dans la chaîne alkyle. Le profil σ et les descripteurs (S_σ) des cations et des anions typiques utilisés dans cette étude ainsi que les divisions des aires des profils σ sont illustrés dans les Fig.IV.2 et IV.3 respectivement.

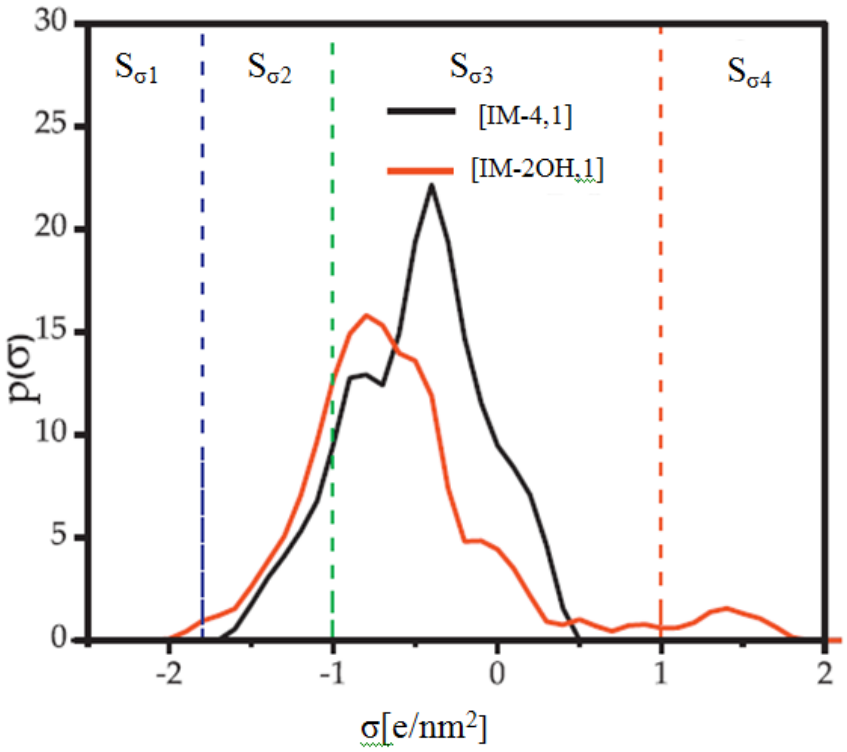


Figure. IV.2. Profil σ des cations 1- (2-hydroxyéthyl) -3-méthylimidazolium ([IM-2OH,1]) et 1-butyl-3-méthyl imidazolium ([IM-4,1])

Comme le montre la Fig.IV.2, le profil σ du cation 1-butyl-3-méthylimidazolium ([IM-4,1]) s'étend de -1,7 à 0,5 e/nm^2 avec un pic maximal à environ -0,5e/nm^2. Ceci montre que ([IM-4,1]) a le double caractère donneur de la liaison hydrogène en raison de l'hydrogène acide attaché au cation imidazolium et du caractère non polaire dû aux chaînes alkyles.Cependant, dans le cas du cation 1-(2-hydroxyéthyl) -3-méthylimidazolium ([IM-2OH,1]), le profil σ est plus étendu de (-2,1 à 1,9) e/nm^2 et le pic maximum est également décalé à la région polaire (-0,9 e/nm^2). L'introduction de la fonction hydroxyle augmente la capacité donneuse de liaison hydrogène du cation 1- (2-hydroxyéthyl) -3-méthylimidazolium. Par conséquent, pour prendre en compte les effets des groupes fonctionnels, nous avons divisé les profils σ des cations comme montré sur la Fig. IV.2

Ainsi, $S_{\sigma1}$ représente le caractère donneur de liaisons hydrogène fort des cations, spécifiquement pour les cations fonctionnalisés; $S_{\sigma2}$ représente le comportement donneur de liaison hydrogène faible; $S_{\sigma3}$ montre le caractère non-polaire des cations. $S_{\sigma4}$ représente le comportement d'accepteur de liaison hydrogène des cations; principalement pour des cations

contenant un groupe fonctionnel avec un caractère accepteur de liaison hydrogène tel qu'un groupe fonctionnel hydroxyle.Les valeurs $S_{\sigma i}$ des cations 1- (2-hydroxyéthyl) -3-méthylimidazolium et 1-butyl-3-méthylimidazolium sont regroupées dans le tableauIV.4. Les valeurs $S_{\sigma i}$ indiquent que le cation 1- (2-hydroxyéthyl) -3-méthylimidazolium posséde à la fois un fort caractère donneur et accepteur de liaison hydrogène tandis que le 1-butyl-3-méthylimidazolium ne présente pas ce comportement.

Tableau IV.4 :Valeurs S_σ de deux cations et de deux anions typiques des Lis.

	Cations			
	$S_{\sigma 1}(-2.5 \le \sigma_1 \le -1.8)$	$S_{\sigma 2}(-1.8 < \sigma_2 \le -1.0)$	$S_{\sigma 3}(-1.0 < \sigma_3 \le 1.0)$	$S_{\sigma 4}(1.0 < \sigma_4 \le 2.0)$
[IM-2OH,1]	0.0961	3.77365	12.4437	0.78105
[IM-4,1]	0	2.62295	17.60025	0
	Anions			
	$S_{\sigma 5}(-1.1 \le \sigma_5 \le 1.0)$	$S_{\sigma 6}(1.0 < \sigma_6 \le 1.5)$	$S_{\sigma 7}(1.5 < \sigma_7 \le 2.0)$	$S_{\sigma 8}(2.0 < \sigma_8 \le 3.0)$
[SO_3-1]	4.17565	1.4724	5.03015	0.0952
[Ac]	4.3219	0.39395	1.7661	2.55105

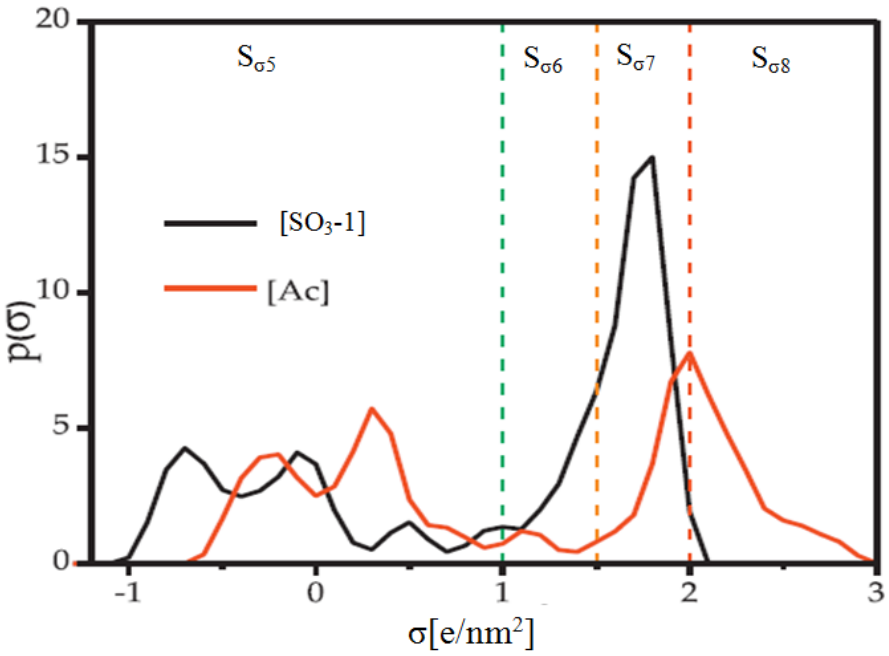


Figure. IV.3 Profil σ des anions méthanesulfonate ([SO_3-1]) et l'acétate ([Ac]).

Le profil σ des anions méthanesulfonate [SO_3-1] et acétate [Ac] s'étend dans la région accepteur de la liaison hydrogène (σ> 1,0 e/nm^2) et dans la région non polaire (-1,0 <σ <1,0) e/nm^2 (Fig. IV.3). Les pics dans la région non polaire sont dus à la chaîne méthylalkyle (CH3)

des anions méthanesulfonate et acétate. En fonction de la capacité de l'accepteur de liaison d'hydrogène de l'anion, le profil σ peut se déplacer vers la droite (régions d'accepteur de liaisons hydrogène fortes). Par exemple, le profil σ de l'acétate s'étend à la région de l'accepteur de liaison hydrogène fort par rapport à l'anion méthanesulfonate (Fig. IV.3) en raison de sa basicité plus forte par rapport à l'anion précédent. Ainsi, les valeurs numériques du descripteur moléculaire $S_{\sigma 8}$ sont plus élevées pour l'acétate que pour le méthanesulfonate. Le descripteur moléculaire $S_{\sigma 8}$ montre un comportement d'accepteur de liaison hydrogène fort des anions tandis que $S_{\sigma 5}$ indique le caractère non polaire des anions. $S_{\sigma 6}$ et $S_{\sigma 7}$ montrent le comportement non-polaire des anions. Les valeurs S_{σ} du méthane sulfonate et de l'acétate sont données dans le tableauIV.4.

Dans ce travail nous avons élaboré un programme MATLAB pour calculer les aires $S_{\sigma i}$/i=1..8 des cations et des anions considérés après numérisation des graphes de profiles σ obtenus à partir de la référence [18]. La figure IV.4 montre l'étape de numérisation du profil σ de l'anion dimethyl phosphate [PO2-O1,O1].

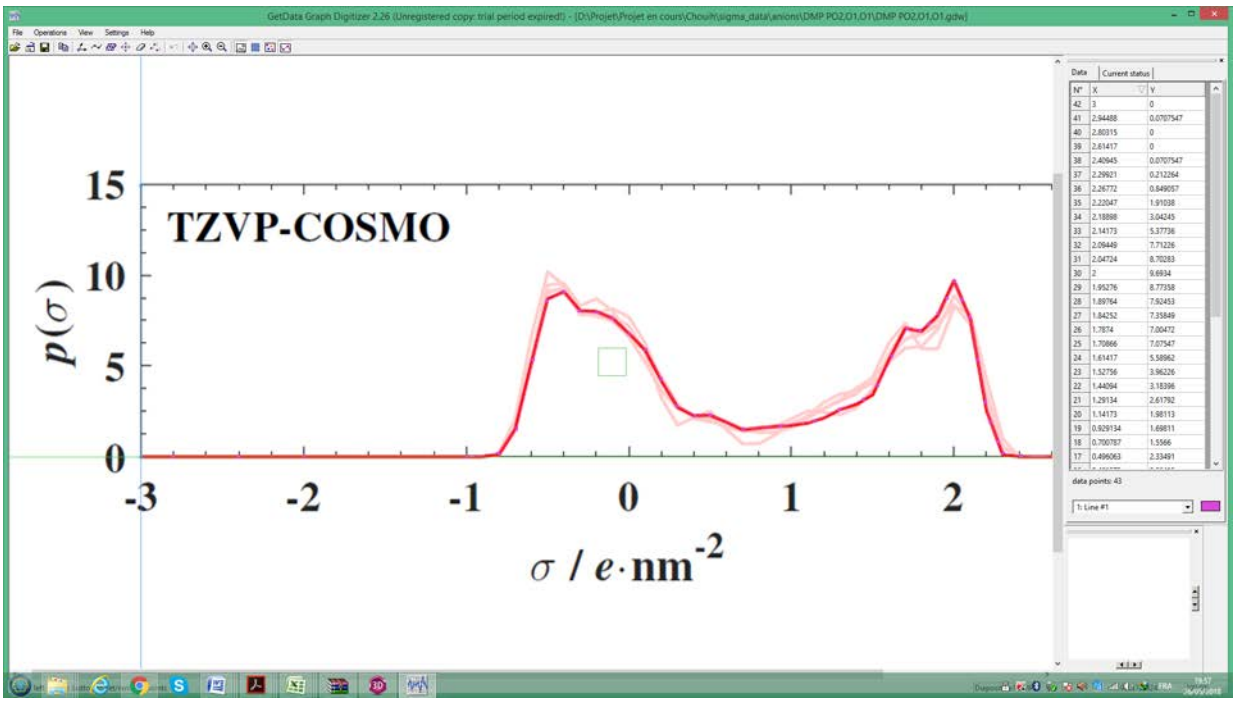


Figure.IV.4 : Ecran de numérisation du graphe de profil σ de l'anion dimethyl phosphate [PO2-O1,O1] en utilisant GetData Graph Digitizer 2.26.

Les données de numérisation sont utilisées ensuite pour une interpolation de spline cubique. Les résultats d'interpolation sont ensuite utilisés pour obtenir les $S_{\sigma i}$ par intégration numérique. Les codes MATLAB des fonctions get_sig_prof_cation etget_sig_prof_anion pour le calcul des $S_{\sigma i}$ des cations et des anions respectivement sont les suivantes :

```
function sigm0a=get_sig_prof_cation(f)
% f est le nom du fichier excel qui contient les données de numérisation
data=xlsread(f);%lecture des données de numérisation du graphe de profil σ du cation
x=data(:,1);
y=data(:,2);
pp=csapi(x,y);% interpolation spline cubique
sigma=zeros(1,4);
k=1;
sigma(k) = integral(@(x) ppval(pp,x),-2.5,-1.8); % calcul de S_σ1
sigma(k+1) = integral(@(x) ppval(pp,x),-1.8,-1); % calcul de S_σ2
sigma(k+2)  = integral(@(x) ppval(pp,x),-1,1); % calcul de S_σ3
sigma(k+3)  = integral(@(x) ppval(pp,x),1,2);% calcul de S_σ4
function sigma=get_sig_prof_anion(f)
% f est le nom du fichier excel qui contient les données de numérisation
data=xlsread(f);
x=data(:,1);
y=data(:,2);
pp=csapi(x,y); % interpolation spline cubique
sigma=zeros(1,4);
k=5;
sigma(k) = integral(@(x) ppval(pp,x),-1.1,1); % calcul de S_σ5
sigma(k+1) = integral(@(x) ppval(pp,x),1,1.5); % calcul de S_σ6
sigma(k+2)  = integral(@(x) ppval(pp,x),1.5,2); % calcul de S_σ7
sigma(k+3)  = integral(@(x) ppval(pp,x),2,3); % calcul de S_σ8
```

Les descripteurs fondés sur le profil σ ($S_{\sigma i}$) de tous les cations et les anions étudiés sont regroupés dans les tableaux IV.5 et IV.6 respectivement. Les descripteurs basés sur le profil σ (S_{σ}) ont été sélectionnés avec soin pour inclure les principales informations chimiques nécessaires pour prédire les interactions électrostatiques, de liaison hydrogène et de dispersion possibles des différents solutés dans les IL.

Tableau IV.5: Descripteurs moléculaires ($S_{\sigma i}$) des cations des Lis étudiés

Cations	$S_{\sigma 1}$	$S_{\sigma 2}$	$S_{\sigma 3}$	$S_{\sigma 4}$
[IM-1,1]	0	2,85905	11,50845	0
[IM-10,1]	0	2,59895	29,55795	0
[IM-10, 10,1]	0	1,25830	48,70780	0
[IM-1O6,1]	0	3,46090	23,24395	0,50665
[IM-1O6,1O6]	0,0004	3,81320	34,78790	1,02110
[IM-2,1]	0	2,67900	13,52630	0
[IM-2OH,1]	0,0961	3,77365	12,44370	0,78105
[IM-4,1]	0	2,62295	17,60025	0
[IM-6,1]	0	2,62745	21,56155	0
[IM-6,1,1]	0	1,89540	23,74280	0
[MO-2O1,1]	0,1012	4,0105	12,3263	1,33100
[MO-4,1]	0	2,88055	17,22055	0
[N-2OH,1,1,1]	0,11035	3,2872	11,11155	0,63090
[P-4I,4I,4I,1]	0	1,23355	27,26105	0
[PIP-3,1]	0	2,71690	16,7056	0
[PIP-4,1]	0	1,26445	20,29545	0
[PIP-5,1]	0	1,28835	22,08255	0
[PIP-6,1]	0	1,3368	23,9002	0
[PY-2]	0	3,18625	12,43015	0
[PY-3OH]	0,06575	3,87935	13,42925	0,92145
[PY-4]	0	2,9986	16,5229	0
[PY-4,1[3]]	0	2,37260	19,1217	0
[PY-4,1[4]]	0	1,99475	19,60785	0
[PY-6,1[3]]	0	2,35380	23,1401	0
[PYR-2O1,1]	0	1,52370	17,45210	0,33860
[PYR-4,1]	0	0,98555	20,82475	0
[QUINI-6]	0	1,59565	26,62895	0
[QUINUC-8]	0	2,25065	29,52255	0
[S-2,2,2]	0	1,89325	15,81125	0

IV.2.3. Bases de données des descripteurs moléculaires des solutés

Pour les solutés nous avons adopté une autre famille de modèles QSPR pouvant être utilisés pour décrire le CADI (γ^{∞}) qui sontbasés sur la relation linéaire de l'énergie de solvatation impliquant les descripteurs moléculaires d'Abraham. Généralement, il y a cinq descripteurs notés E, S, A, B et L implémentés dans ces corrélations, alors que chaque descripteur est lié à un type spécifique de propriétés moléculaires du soluté, à savoir, la dipolarité (E) (composé ayant moment dipolaire non nul), la polarisabilité (S) (la polarisabilité électrique est la tendance relative d'une distribution de charge, comme le nuage d'électrons d'un atome ou d'une molécule, et par conséquent de tout corps matériel, à faire déplacer ses charges par un champ électrique externe), l'acidité (A), la basicité (B) et la lipophilicité (L) (se réfère à la capacité d'un composé chimique à se dissoudre dans les graisses, les

huiles, les lipides et les solvants non polaires tels que l'hexane ou le toluène).Des études montrent que ces descripteurs peuvent être utilisés pour corréler les données de γ^{∞}avec précision en utilisant une régression linéaire multiplesimple:

$$RTln\gamma^{\infty} = c + eE + sS + aA + bB + lL \qquad (4.1)$$

Tableau IV.6 : Descripteurs moléculaires ($S_{\sigma i}$) des anions des Lis étudiés

Anions	$\mathbf{S_{\sigma 5}}$	$\mathbf{S_{\sigma 6}}$	$\mathbf{S_{\sigma 7}}$	$\mathbf{S_{\sigma 8}}$
[NTF2]	15,07195	5,2518	0,03085	0,00000
[PO2-O1,O1]	7,97735	1,2011	3,3661	1,50805
[BF4]	1,8414	7,4511	0,1052	0,00000
[TCB]	11,83765	4,04255	0,1932	0,00000
[AC]	4,3219	0,39395	1,7661	2,55105
[CCN3]	7,67725	4,5446	0,91695	0,00000
[DCA]	4,64265	2,98635	2,55155	0,00335
[FAP]	25,3723	0,513	0	0
[OTF]	6,9381	4,5559	1,2859	0
[SCN]	0,00000	2,9603	4,0302	1,6415
[SO3-1]	4,17565	1,4724	5,03015	0,0952
[SO3-PH1]	12,43575	1,7128	4,5087	0,01055
[TFA]	6,6641	0,9255	2,95225	0,38565
[CL]	0	0	4,6633	0,6177

Le tableauIV.7 regroupe les valeurs des descripteurs des solutés (paramètres de solvatation d'Abraham)

TableauIV.7: Les valeurs numériques des descripteurs de solutés [18].

Soluté		**E**	**S**	**A**	**B**	**L**
Alcanes	n-hexane	0	0	0	0	2,668
	n-heptane	0	0	0	0	3,130
	n-octane	0	0	0	0	3,667
	n-nonane	0	0	0	0	4,182
	n-décane	0	0	0	0	4,686
	n-pentane	0	0	0	0	2,162
	n-undécane	0	0	0	0	5,191
	n-tridécane	0	0	0	0	6,200
	n-tetradécane	0	0	0	0	6,700
	n-dodécane	0	0	0	0	5,696
Cycloalcanes	cyclohexane	0,305	0,100	0	0	2,964
	méthylcyclohexane	0,224	0,100	0	0	3,323
	cyclopentane	0,263	0,100	0	0	2,477
	cycloheptane	0,263	0,100	0	0	2,477
	cyclooctane	0,413	0,100	0	0	4,329
	méthylcyclopentane	0,225	0,100	0	0	2,816
Eau		0	0,450	0,820	0,350	0,260

IV.3. Modélisation par LSSVM

L'idée de machine à vecteurs de support (SVM) découle du problème de la classification des données, à savoir l'attribution d'une étiquette numérique (généralement d'un ensemble fini, par exemple ± 1) à une donnée d'entrée basée sur une fonction de décision. La forme finale du classificateur est obtenue au moyen d'un apprentissage supervisé, c'est-à-dire sur la base de quelques exemples représentant un comportement valable du problème considéré. Les SVM ont également été utilisés comme outils efficaces pour l'estimation de la fonction non linéaire. En ce qui concerne ce domaine particulier des applications, il a été bien établi dans la communauté d'apprentissage par machine que les SVM présentent plusieurs avantages par rapport aux autres algorithmes d'apprentissage automatique. Par exemple, les SVM ne nécessitent pas une étape similaire à une sélection de la topologie comme pour les réseaux de neurones artificiels (par exemple le nombre de couches ou de neurones cachés), ne présentent pas de risque élevé de minima local ou de sur-apprentissage,et ne sont pas sensibles aux estimations initiales.

En 1999, Suykens et Vandewalle ont proposé une version améliorée du SVM appelé (LSSVM). En particulier, LSSVM reformule faiblement le problème d'optimisation original de SVM en termes de méthode des moindres carrés et remplace les contraintes d'inégalité par des égalités respectives.Il s'est avéré que ces changements simplifient considérablement les calculs et réduisent le coût de calcul global. En effet, le modèle LSSVM final est donné comme une solution d'équations linéaires au lieu d'un problème de programmation quadratique requis dans les SVM conventionnels.

L'utilisation d'un algorithme d'apprentissage par machine, tel que LSSVM, pour le calcul du CADI (γ^{∞}) se présente sous la forme d'une fonction$f : \mathbb{R}^n \rightarrow \mathbb{R}$ représentant un ensemble de données :

$$D = \{(\mathbf{x}_i, y_i) : \mathbf{x}_i \in \mathbb{R}^n; y_i \in \mathbb{R}; i = 1, \ldots, N\} \qquad (4.2)$$

(où$\mathbb{R}^n$ désigne un ensemble de vecteurs colonnes de n-éléments de nombres réels) avec les relations de type :

$$y_i = f(\mathbf{x}_i; \mathbf{p}) + \varepsilon_i = \hat{y}_i + \varepsilon_i \qquad (i = 1, \ldots, N) \qquad (4.3)$$

Où $\hat{y} = f(\mathbf{x}; \boldsymbol{p})$désigne la valeur calculée en utilisant le modèle,εreprésente le terme de perturbation (résiduel), alors que **p** est le vecteur des paramètres (coefficients) du modèle *f*. En

général, le meilleur modèle (c-a-d la valeur optimale de **p**) est choisie pour que la somme des carrés des résidus (ou une fonction objective similaire) soit minimale.

LSSVM considère le problème d'approximation de l'ensemble de données D défini dans l'Eq. (4.2) avec la fonction suivante :

$$y_i = w^T \varphi(\mathrm{x}_i) + b + \varepsilon_i (i = 1, \dots N) \qquad (4.4)$$

où w $\in \mathbb{R}^h$ est le vecteur de poids de l'espace de haute dimension, b $\in \mathbb{R}$ est le terme de biaiset $\varphi : \mathbb{R}^h \to \mathbb{R}^n$. Le problème d'optimisation pour les valeurs finales de **w** et **b** est donné comme suit:

$$\min_{\mathbf{w},b,\varepsilon_1,\dots,\varepsilon_N} J(\mathbf{w}, b, \varepsilon_1, \dots, \varepsilon_N) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \frac{1}{2}\gamma \sum_{i=1}^{N} \varepsilon_i^2 \qquad (4.5)$$

Où $\gamma > 0$ est appelée la constante de régularisation, soumise à l'Eq. (4.4).

Le problème d'optimisation avec contraintes donné dans l'Eq. (4.5) est réécrit sous la forme de Lagrangien sans contraintes avec des multiplicateurs $\alpha_i (i = 1 \dots N)$.La différenciation du lagrangien par rapport à toutes les variables suivies de réarrangements algébriques simples conduit à une relation compacte entre le vecteur des poids et les multiplicateurs de Lagrange, à savoir :

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i \varphi(\mathbf{x}_i) \qquad (4.6)$$

D'autres transformations permettent de représenter la condition optimale comme les systèmes suivants d'équations linéaires :

$$\left[\begin{array}{c|c} 0 & \mathbf{1}_N^{\mathrm{T}} \\ \hline \mathbf{1}_N & \boldsymbol{\Omega} + \gamma^{-1}\mathbf{I}_N \end{array}\right] \left[\begin{array}{c} b \\ \hline \boldsymbol{\alpha} \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline \mathbf{y} \end{array}\right] \qquad (4.7)$$

Dans l'Eq. (4.7), $\mathbf{1}_N = [1, \dots, 1]^{\mathrm{T}}$, $\mathbf{y} = [y_1, \dots, y_N]^{\mathrm{T}}$, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^{\mathrm{T}}$ et $\boldsymbol{\Omega} \in \mathbb{R}_N^N$

Ω est une matrice symétrique avec : $\Omega_{ij} = \varphi(\mathbf{x}_j)^{\mathrm{T}} \varphi(\mathbf{x}_i)$

En remplaçant l'équation (4.6) dans l'équation (4.4) (en gardant cela à l'esprit ($\varepsilon_i = y_i - \hat{y}_i$) et en introduisant $K(\mathbf{x}, \mathbf{x}') \equiv \varphi(\mathbf{x})^T \varphi(\mathbf{x}')$ on aboutit à l'équation finale du modèle basé sur LSSVM, à savoir :

$$\hat{y}_i = \sum_{j=1}^{N} \alpha_i K(\mathbf{x}_i, \mathbf{x}_j) + b \qquad (i = 1, \ldots, N) \tag{4.8}$$

L'astuce du noyau (Kernel) permet d'éviter d'utiliser les produits scalaires de la fonction φet de les remplacer par une fonction noyau particulière. Plusieurs types de fonctions du noyau sont couramment utilisés dans la modélisation LSSVM. Dans ce travail, le noyau de la fonction de base radiale (RBF) est adopté :

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

Le paramètre σ^2 est considéré comme un degré de liberté supplémentaire, optimisé par un algorithme d'optimisation externe lors des calculs.

Les outils MATLAB pour les calculs LSSVM inclus dans la boîte à outils LS-SVMlab ont été utilisés dans ce travail. LS-SVMlab (version 1.8) [LSSVMlab. Publié en ligne. Accédé le 15 janvier 2018, URL: http: // www.esat.kuleuven.be/sista/lssvmlab/.]

Tout d'abord, un algorithme approprié implémenté dans LS-SVMlab détecte et redimensionne les variables continues, catégorielles et binaires.Ensuite, les paramètres de LSSVM (constante de régularisationγ et le paramètre kernelσ^2) sont déterminés par minimisation d'une mesure de performance sélectionnée; cette étape est appelée tuning. Les paramètres par défaut de LS-SVMlab, c'est-à-dire MSE (erreur quadratique moyenne) :

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{4.10}$$

Ont été adoptés pour ajuster le LSSVM. Les points de départ appropriés pour la constante de régularisationγ et le paramètre kernelσ^2 sont déterminés par une technique d'optimisation globale de pointe. Dans le cadre del'ajustement (tuning) LS-SVMlab, la qualité de l'ajustement est également vérifiée par validation croisée 5 fois.Après avoir ajusté la constante de régularisationγ et le paramètre kernelσ^2,α et b (le modèle final) sont calculés en résolvant des systèmes linéaires dans l'Eq (4.7). Cette étape est appelée apprentissage LSSVM.

IV.4. Choix des variables d'entrées

Dans ce travail nous avons opté pour la modélisation du CADI par la méthode LSSVM, telle que décrite plus haut, pour 4 ensembles de données : nous avons considéré la base B1 relative à l'eau, B2 celle des cycloalcanes, B3 celle des alcanes et B4 celles qui contient tous les solutés.

- Pour la base de données B1 : Le vecteur des variables d'entrée est constitué de la température et les profils σ des cations et des anions. La variable de sortie est le logarithme normal du CADI

$$\mathbf{x_i} = [1/\tilde{T}_i\,, S_{\sigma 1i}, S_{\sigma 2i}, S_{\sigma 3i}, S_{\sigma 4i}, S_{\sigma 5i}, S_{\sigma 6i}, S_{\sigma 7i}, S_{\sigma 8i}]^T \in \mathbb{R}^9$$

$$\mathbf{y_i} = \ln(\gamma_i^{\infty}) \in \mathbb{R}$$

Avec $\tilde{T}_i = 10^{-3} T_i$ T= la température en Kelvin et *i=1,2,..N.*

- Pour les bases B2, B3 et B4 : Le vecteur des variables d'entrée est constitué de la température et les profils σ des cations et des anions et les cinq paramètres d'Abraham des solutés.

$$\mathbf{x_i} = [1/\tilde{T}_i\,, S_{\sigma 1i}, S_{\sigma 2i}, S_{\sigma 3i}, S_{\sigma 4i}, S_{\sigma 5i}, S_{\sigma 6i}, S_{\sigma 7i}, S_{\sigma 8i}, E_i, S_i, A_i, B_i, L_i]^T \in \mathbb{R}^{14}$$

$$\mathbf{y_i} = \ln(\gamma_i^{\infty}) \in \mathbb{R}$$

Avec $\tilde{T}_i = 10^{-3} T_i$ T= la température en Kelvin et *i=1,2,..N.*

IV.5. Normalisation des données

La normalisation est une étape importante dans le processus de l'élaboration desdonnées. Elle devient dans la plus part des cas, une exigence pour les matrices des inputs,du fait qu'on utilise souvent des fonctions de transfert bornées de typesigmoïdes dans les modélisations statiques.
La normalisation utilisée a pour but d'avoir les données normalisées comprises entre[-1 ,1].Dans ce travail, on a utilisé la base de données qui est représenté en dix colonnes, correspondant respectivement aux descripteurs moléculaire, les températures et lescoefficients d'activités. Les valeurs numériques ont été normalisées afin d'améliorerl'optimisation. La normalisation se fait par la fonction ***mapminmax*** de MATLAB.Chaque paramètre des entrées et de sortie ($x_{réel}$) est normalisé (x_{norm}) par l'expressionsuivante:

$$x_{norm} = \frac{2(x_{réel} - x_{min})}{(x_{max} - x_{min})} - 1 \qquad (4.11)$$

Par cette équation les valeurs réelles de x comprises ($x_{min} \leq x_{réel} \leq x_{max}$) deviennent des valeurs normalisées comprises ($-1 \leq x_{norm} \leq +1$).

IV.6. Division des données en ensembles apprentissage/test

Pour les quatre bases de données B1à B4, les données ont été classées dans un fichier Excel par liquide ionique et par ordre croissant de la température. Ensuite les lignes impaires ont été choisies comme un ensemble d'apprentissage à partir desquels les paramètres α et $b$du modèle LSSVM sont déterminés et les lignes paires comme un ensemble de test pour valider le modèle optimisé.Les résultats de cette division son résumés dans le tableau IV.8 :

Tableau IV.8 Taille des ensembles d'apprentissage et de test pour les 4 bases de données

	B1	**B2**	**B3**	**B4**
Taille de l'ensemble d'apprentissage	167	909	1365	2441
Taille de l'ensemble de test	166	908	1365	2441
Total App.+test	333	1817	2732	4882

Le tableau IV.9 montre les valeurs minimales et maximales des variables d'entrée et de sortie pour les quatre bases considérées.

Tableau IV.9 Les valeurs minimales et maximales des variables d'entrée et de sortie

	Base B1 (Eau)														
	T (K)	$S_{\sigma 1}$	$S_{\sigma 2}$	$S_{\sigma 3}$	$S_{\sigma 4}$	$S_{\sigma 5}$	$S_{\sigma 6}$	$S_{\sigma 7}$	$S_{\sigma 8}$	E	S	A	B	L	$ln\gamma^{\infty}$
Min	293,15	0	0,98555	11,11155	0	0	0	0	0	-	-	-	-	-	-4.3428
Max	393	0,11035	4,01050	34,78790	1,331	25,3723	5,2518	5,03015	2,55105	-	-	-	-	-	2.2364
	Base B2 (Cycloalcanes)														
Min	293,15	0	0,9856	11,1116	0	0	0	0	0	0,224	0,10	0	0	2,4770	0.2311
Max	413,00	0,11035	4,0105	48,7078	1,331	25,3723	7,4511	5,0302	2,55105	0,413	0,10	0	0	4,3290	6.8394
	Base B3 (n-alcanes)														
Min	293,15	0	0,98555	11,11155	0	0	0	0	0	0	0	0	0	2.62	0.5653
Max	413,00	0,11035	4,01050	48,70780	1,331	25,3723	7,45110	5,03015	2,55105	0	0	0	0	6.20	8.3481
	Base B4 (Eau+cycloalcanes+n-alcanes)														
Min	293,15	0	0,98555	11,11155	0	0	0	0	0	0	0	0	0	0,26	-4.3428
Max	413.00	0,11035	4,0105	48,7078	1,331	25,3723	7,4511	5,03015	2,55105	0,413	0,45	0,82	0,35	6,7	8.3481

IV.7.Ajustement de la constante de régularisation γ et du paramètre kernel σ^2

Les résultats de l'ajustement optimal de la constante de régularisationγ et le paramètre kernelσ^2sont résumés sur le tableau IV.10

Tableau IV.10 Résultats de l'ajustement optimal de la constante de régularisation γ et du paramètre kernel σ^2

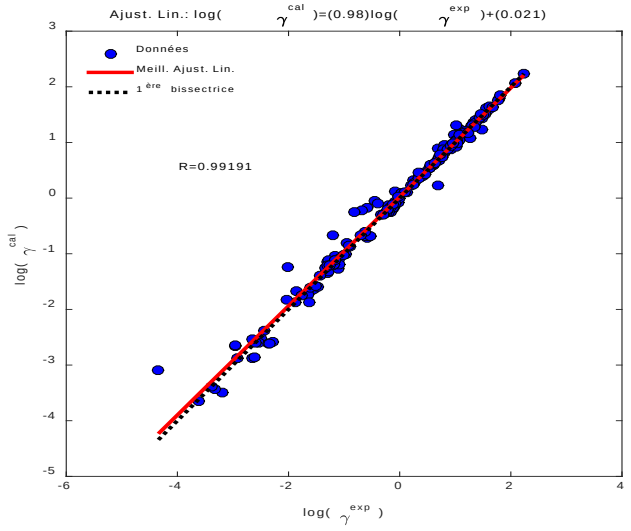
	B1	**B2**	**B3**	**B4**
Constante de régularisation γ	$2.92539.10^3$	$7.8944.10^4$	$5.8987.10^3$	$0.9607.10^3$
Paramètre kernel σ^2	12.5715	51.0036	64.5308	3.2663

IV.8. Résultats de corrélation des modèles optimisés

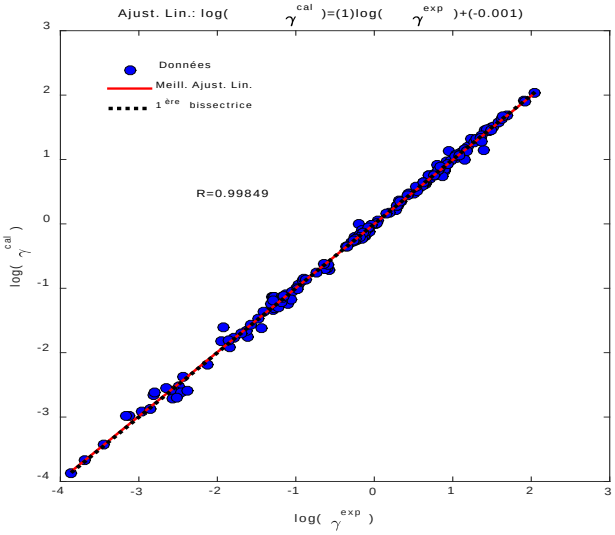
Les résultats de la modélisation LSSVM du CADI pour les 4 bases de données choisies sont présentés sous forme de digrammes de dispersion ou ne nuage de points qui montrent la qualité de concordance entre les valeurs calculées par le modèle et les valeurs expérimentales du CADI. Cette qualité est visualisée à travers l'alignement des points sur la première bissectrice pour l'ensemble des données d'apprentissage et de test, le coefficient de corrélation R et la droite du meilleur ajustement linaire des points. La meilleure concordance correspond à la droite du meilleur ajustement linéaire de pente égale ou proche à l'unité et de l'ordonnée à l'origine égale ou proche de zéro et un coefficient de corrélation proche de un.

IV.8.1. Résultats du modèle optimisé pour la base B1

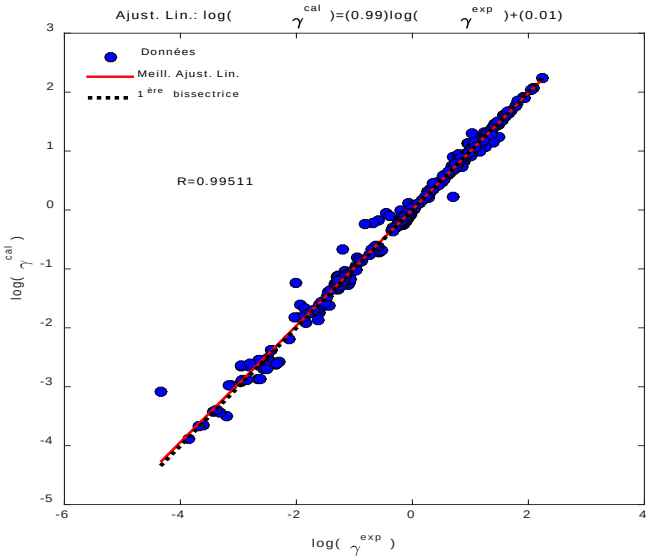
La capacité du modèle LSSVM pour la corrélation du CADI de l'eau dans les Lis de la base B1 est illustrée dans la Fig. IV.5. Il faut noter que bien que la majorité des points son alignés sur la première bissectrice pour l'ensemble de test (Fig. IV.5(c)), il y a cependant, un point qui se trouve en dehors de l'intervalle d'apprentissage pour lequel la valeur du CADI prédite par le modèle est quelque peu supérieure à la valeur expérimentale.



(a)



(b)

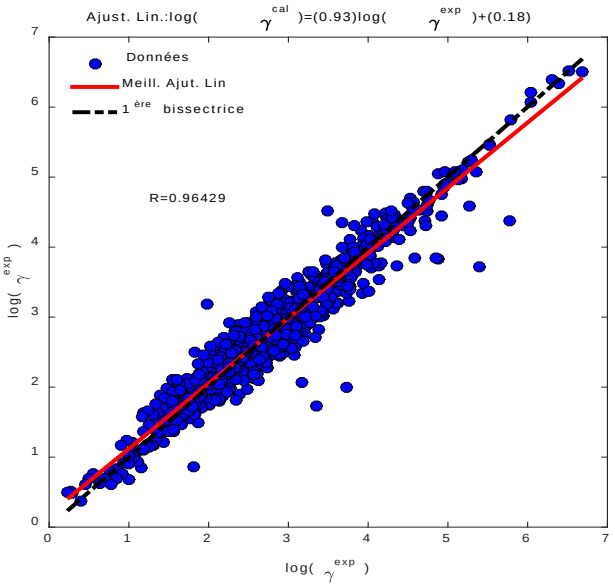


(c)

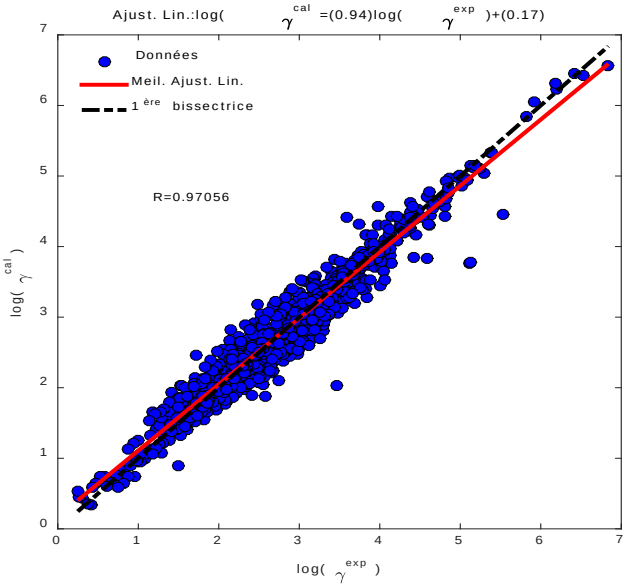
Figure. IV.5 Diagrammes du nuage de points du logarithme népérien du CADI de l'eau dans les Lis (Base B1), (a) test, (b) apprentissage, (c) globale

IV.8.2. Résultats du modèle optimisé pour la base B2

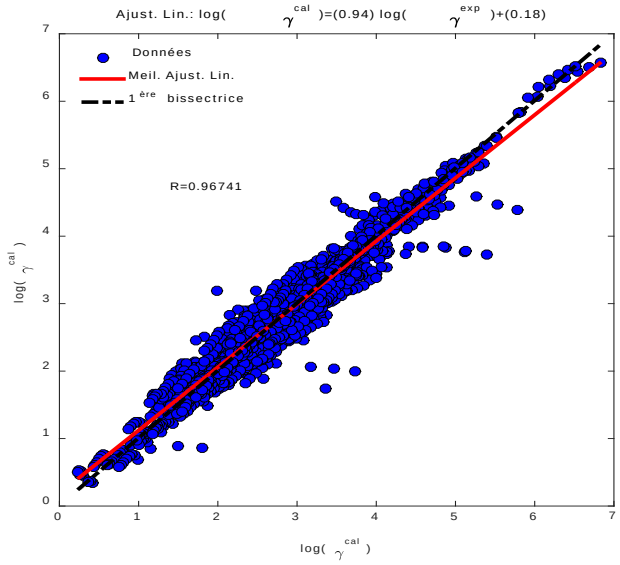
Pour la base de données des cycloalcanes la performance du modèle optimisé LSSVM est montrée sur la Fig. IV.6. Il est clair que la qualité de corrélation est légèrement inférieure à celle de la base B1. Ceci se traduit par une légère déviation de la droite du meilleur ajustement linéaire des points par rapport à la première bissectrice et coefficient de corrélation légèrement inférieur. Ceci peut être expliqué par le fait d'existence de données expérimentales différentes, publiées généralement par des équipes de recherche différentes, pour les mêmes conditions de température et pour le même système solvant-soluté. Cela dit, la qualité de corrélation peut être classée dans le domaine d'acceptabilité.



(a)



(b)

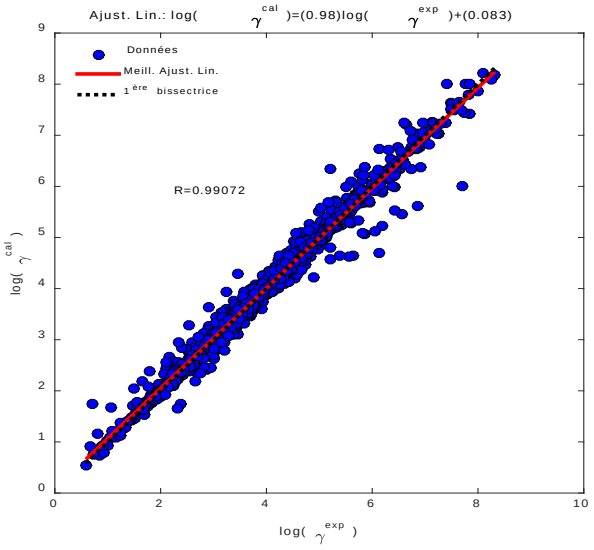


(c)

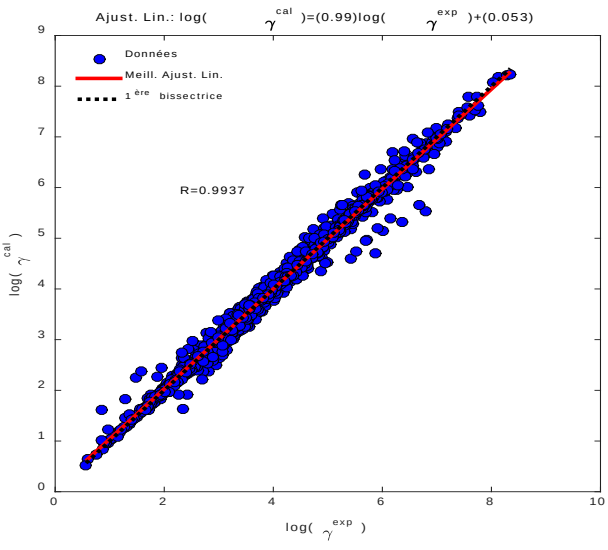
Figure. IV.6 Diagrammes du nuage de points du logarithme népérien du CADI des cycloalcanes dans les Lis (Base B2), (a) test, (b) apprentissage, (c) globale

IV.8.3. Résultats du modèle optimisé pour la base B3

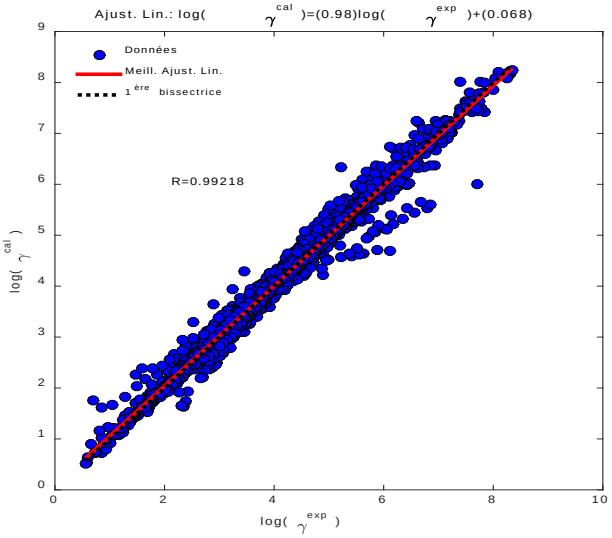
La Fig. IV.7 montre la qualité de corrélation du modèle optimisé LSSVM pour la base de données du CADI des alcanes dans les LIs. La performance est légèrement supérieure à celle du modèle optimisé à partir de la base de données des cycloalcanes malgré que la base nettement plus large. Ceci est visualisé par le nuage de point plus compact sur la première bissectrice. Cependant il faut remarquer l'existence de quelques points un peu plus dispersés par rapport à la première bissectrice, et qui peut être expliqué comme pur les cycloalcanes, le fait d'existence de données expérimentales différentes pour les mêmes conditions de température et pour le même système solvant-soluté.



(a)



(b)



(c)

Figure. IV.7 Diagrammes du nuage de points du logarithme népérien du CADI des alcanes dans les Lis (Base B3), (a) test, (b) apprentissage, (c) globale

IV.8.4. Résultats du modèle optimisé pour la base B4

Cette partie de l'étude à été réalisée pour évaluer l'effet de regrouper plusieurs familles de solutés en un seul modèle sur la qualité de corrélation. Comme le montre la Fig. IV.8 le modèle LSSVM optimisé à partir la base B4 présente une capacité de corrélation des données du CADI des 17 solutés considérés dans les 80 Lis très comparable aux modèles des bases B2 et B3 malgré la taille de la base. Il faut toutefois noteraussi l'existence du point aberrantconstaté sur la F.i.v. Même remarque peut être faite pour les quelques points un peu plus dispersés par rapport à la première bissectrice et qu'on a attribué a l'existence de données expérimentales différentes pour les mêmes conditions de température et pour le même système solvant-soluté. Les résultats du modèle optimisé pour à partir de la base B4 montrent que le fait de regrouper des données du CADI de plusieurs famille de solutés dans différents Lis ne diminue nullement la qualité de corrélation du modèle optimisé.

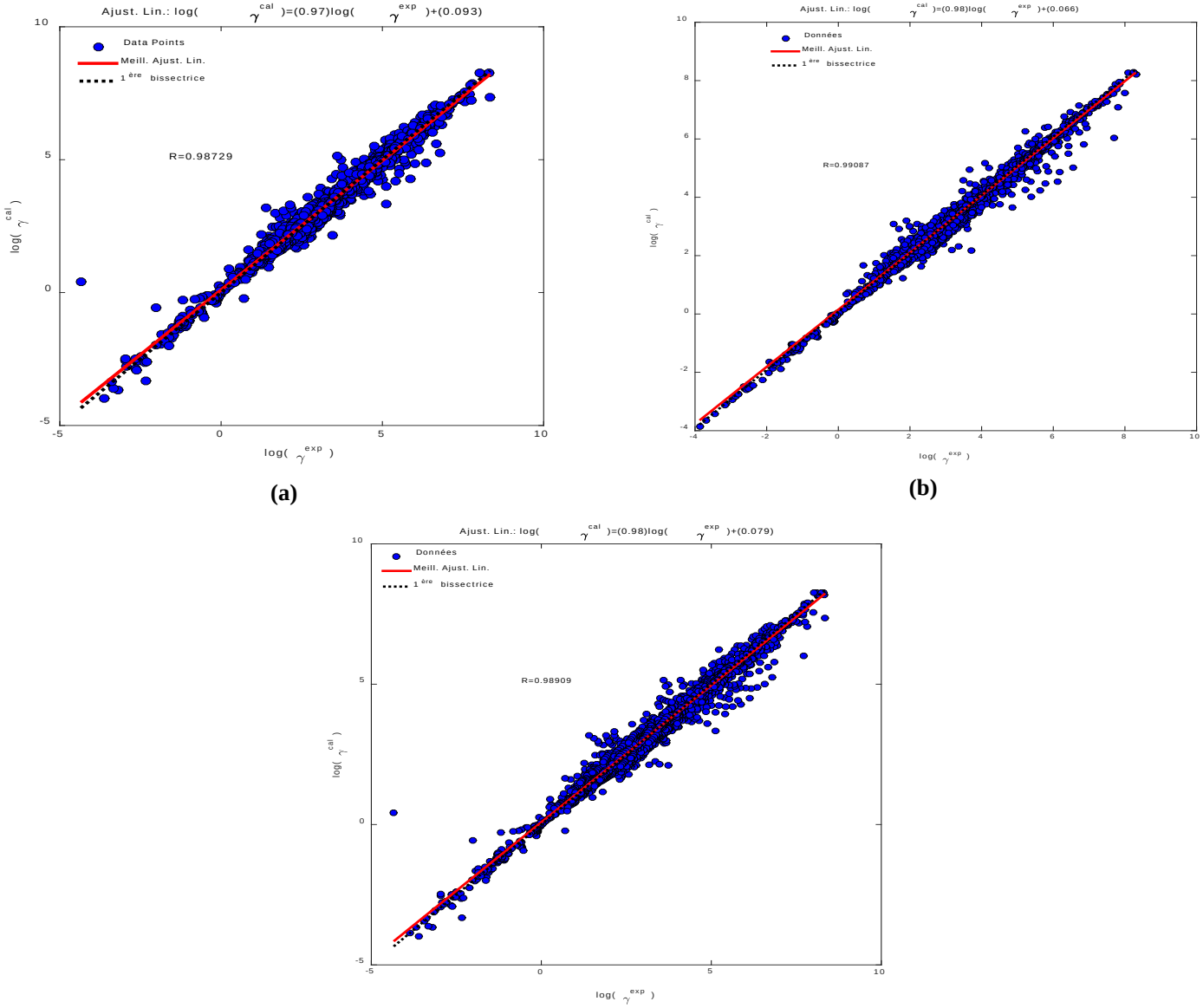


Figure. IV.8 Diagrammes du nuage de points du logarithme népérien du CADI de l'eau, des cycloalcanes et des alcanes dans les Lis (Base B4), (a) test, (b) apprentissage, (c) globale

IV.8.5. Evaluation des erreurs des modèles optimisés

L'évaluation de la performance des modèles optimisés a été réalisée en termes d'erreur relative absolue moyenne (ERAM) définie par l'équation:

$$EARM(\%) = \frac{100}{N}\sum_{i=1}^{N}\frac{\left|\gamma_i^{\infty cal} - \gamma_i^{\infty exp}\right|}{\gamma_i^{\infty exp}} \qquad (4.12)$$

Le tableau IV.11 résume le calcul de l'ERAM pour les modèles LSSVM optimisés.

Tableau IV.11: Erreur Relative Absolue Moyenne (ERAM) et coefficient de corrélation pour les modèles LSSVM optimisés

Bases	B1(333)		B2(1817)		B3(2732)		B4(4882)	
	ERAM%	R	ERAM%	R	ERAM%	R	ERAM%	R

Apprentissage	**4.76**	0.9985	**16.84**	0.9706	8.85	0.9937	13.57	0.9909
Test	**10.30**	0.9919	17.61	0.9643	11.26	0.9922	**20.50**	0.9873
Global	**7.52**	0.9952	**17.23**	0.9674	10.06	0.9907	17.03	0.9891

Le calcul d'erreur montre que la différence entre les ERAM des ensembles d'apprentissage et de test est plus grande pour les bases B1 et B4. Ceci est expliqué par l'existence du point aberrant dans les deux bases. Pour la base B2 cette différence est très faible alors que l'ERAM est la plus élevée.

IV. 9. Conclusion

La modélisation LSSVM du CADI de 17 solutés dans 80 LIs,résultant de la combinaison de 29 cations appartenant à 10 familles cations avec 14 anions, a donné des résultats satisfaisant. L'utilisation des descripteurs basés sur les profils σ du modèle COSMO-RS pour décrire les anions et les cations et les paramètres de la relation linéaire de l'énergie de solvatation d'Abrahampour décrire les solutés a permit la corrélation du CADI, des solutés considérés dans LIs de la base de données élaborée dans la cadre de ce travail, avec une ERAM comprise entre 7.52% et 17.23% selon la base de donnée considérée. L'existence de points aberrants, dues probablement aux erreurs d'impression ou de mesure, ainsi que l'existence de données expérimentales différentes, publiées généralement par des équipes de recherche différentes, pour les mêmes conditions de température et pour le même système solvant-soluté ont fait en sorte que l'ERAM soit élevée pour les quatre modèles optimisés. Cela dit les résultats obtenus peuvent êtreconsidérés comme acceptable comparativement aux résultats des études similaires bien que les bases de données ne soit pas les mêmes.

Conclusion générale

Dans ce travail, nous avons donné un aperçu général sur les liquides ioniques qui présentent une nouvelle génération de solvants moins toxique, moins inflammable et moins polluant. Les liquides ioniques, sels fondus qui satisfont ces critères, sont envisagés comme alternatives des solvants organiques volatiles.

Nous avons commencé notre travail par l'élaboration d'une base de données de mesures expérimentales du coefficient d'activité à dilution infinie de l'eau, de 5 cycloalcanes et de 11 n-alcanes dans 80 liquides ioniques constitués par la combinaison de 29 cations, appartenant à 10 familles de cations, avec 14 anions totalisant ainsi 4882 points expérimentaux. Nous avons ensuite calculé les descripteurs des 29 cations et 14 anions en se basant sur les profils σ du modèle de chimie quantique COSMO-RS. Pour compléter la base de données, nous avons ensuite choisi pour décrire les solutés, les paramètres d'Abraham de la relation linéaire de l'énergie de solvatation.

Afin d'évaluer l'effet de la base de données en termes de taille et la nature du soluté nous avons considéré 4 bases de données. La première concerne l'eau constituée de 333 points expérimentaux, la deuxième les cycloalcanes de 1817 points, la troisième les n-alcanes de 2732 points et la quatrième tous les solutés groupés constituée de 4882 points.

Nous avons ensuite optimisé pour chacune des bases de données un modèle LSSVM ayant comme variables d'entrée la température normalisée, les 8 descripteurs basés sur les profils σ des cations et des anions (4 pour le cation et 4 pour l'anion constituant le liquide ionique) pour la 1ère base de données (09 variables au total) et les cinq paramètres d'Abraham de la relation linéaire de l'énergie de solvatation (13 variables au total) pour les 3 autres bases de données. La variable de sortie des 4 modèles optimisés est le logarithme népérien du coefficient d'activité à dilution infinie du soluté dans le liquide ionique. Le processus d'optimisation est constitué de trois étapes ; la première concerne l'ajustement de deux paramètres du modèle LSSVM, constante de régularisation et le paramètre de la fonction radiale de base du Kernel, la deuxième étape concerne l'apprentissage en utilisant la moitié de la base de données afin de déterminer les paramètres α et b du modèle, la troisième étape concerne la validation du modèle en utilisant la deuxième moitié de la base de données.

Les résultats obtenus montrent, globalement, la capacité des 4 modèles optimisés à corréler d'une manière acceptable le coefficient d'activité à dilution infinie de solutés (eau, cycloalcanes et n-

alcanes) dans les liquides ioniques avec une erreur relative absolue moyenne comprise entre 10.30% et 20.50% et un coefficient de corrélation compris entre 0.9919 et 0.9643 pour l'ensemble des données de validation selon la base de donnée considérée. Ces résultats s'avèrent très satisfaisants du point de vue de génie des procédés et sont très encourageants pour des futurs projets de modélisation des propriétés chimiques, physiques, ou thermodynamiques ou d'activité des liquides ioniques à partir des descripteurs moléculaires basés sur les profils σ en utilisant les machines à vecteurs de support.

L'adaptation de procédés verts respectant l'environnent par l'utilisation des liquides ioniques nous impose la nécessité de consacrer beaucoup de recherches vers ce domaine qui est très prometteur, mais nécessite aussi lacollection le plus grand nombre d'information possibles sur les liquides ioniques afin de pouvoir les adhérer sur le terrain.

Comme perspectives nous suggérons les recommandations suivantes comme de nouvelles pistes d'investigation futures:

☐ Elargir la base de données de coefficient d'activité à dilution infinie pour d'autres solutés dans d'autres liquides ioniques.

☐ Elargir la base de données des descripteurs moléculaires basés sur les profils σ pour d'autre cations et anions.

☐ Elaborer d'autres bases de données des propriétés physiques des liquides ioniques telles que la viscosité et la tension superficielle dans le but d'établir des modèles de prédiction et/ou corrélation de ces deux propriétés vue leurs importances dans les processus de transfert de masse associés aux systèmes réactionnel et aux systèmes de séparation par distillation et par extraction.

Références

[1] **Catherine Santini,** Yves Chauvin, Les Liquides Ioniques, leur Utilisation et leur rolecomme solvants de reaction Catalytique, Thèse de doctorat, 2007, L'université Claude Bernard - Lyon 1

[2] **Melle Bouhassoun Wassila,** Synthese d'un liquide ionique sulfonate & application dans l'extraction de la (iii), Memoire de master, 2016, Universite Aboubekr Belkaid- Tlemcen, Algerie

[3] **Toubakh Meriem,** Etude de l'utilisation des liquides ioniques comme eco-solvants pour la récupération des composés organiques a l 'échelle industrielle, Mémoire de master, 2016, Universite Kasdi Merbah Ouargla, Algerie

[4] **Bachir Aoun,** Liquides ioniques: structure et dynamique, Thèse de doctorat, 2010, Université D'orléans, France

[6] **Amel Ayad,** Contribution à l'étude de la réduction des émissions de cov par les liquides ioniques, Thèse de doctorat, 2017, Universite Abou-Bekr Belkaid – Tlemcen, Algerie

[5] **Amine Bendaoud,** Fluide supercritique et liquide ionique comme plastifiants de polymères biosourcés: application à l'amidon et à l'acétate de cellulose, Thèse de doctorat, 2014, L'universite Jean Monnet

[7] **Anne-Laure Revelli,** Etude thermodynamique des liquides ioniques : applications à la protection de l'environnement, Thèse de doctorat, 2010,De Nancy université institut national polytechnique de lorraine

[8] **Grover., B. Singh., M. Bakshi., S. Singh.,** Quantitative structure- Property relationship in pharmaceutical research Part2. Pharm. (2000)

[9] **Klamt, A. and Eckert, F.**: COSMO-RS: un nouveau procédé efficient pour la prédiction a priori de données thermophysiques de liquides. fluid phase equilibria 172: 43-72 (2000)

[10] **Amadou Diop,** Extraction, Depol ymerisation et valorisation de la lignine kraft de la liqueur noire, Thèse de doctorat, 2014, L'université du Québec À Trois-Rivières,Canada

[11] **Amich, G.Benrouissi,A,** Modélisation neuronale des coefficients d'activité à la dilution infinie de l'eau dans les liquides ioniques en utilisant des profils sigma comme descripteurs moléculaires, Memoire de master, 2016, universite yahia fares de medea, Algerie

[13] **Emmanuel Billy,** Application des liquides ioniques à la valorisation des métaux précieux par une voie de chimie verte, Thèse de doctorat, 2012, L'université De Grenoble

[14] **Kaniki Armel Tumba,** Infinite dilution activity coefficient measurements of organic solutes in fluorinated ionic liquids by gas-liquid chromatography and the inert gas stripping method, Mémoire de Master , *2010,* University Of Kwazulu-Natal

[15] **Gagandeep Singh Saggu,** Postdoctoral Research Scientist at National Institutes of Health, 2017

[16] **Girma Gonfa a,∗, Mohamad Azmi Bustam a, Azmi M. Shariffa, Nawshad Muhammad b,Sami Ullah ,** Quantitative structure–activity relationships (QSARs) for estimation of activity coefficient at. Infinite dilution of water in ionic liquids for natural gas dehydration, journal of the Taiwan institute of chemical engineer 66 (2019) 222-229

[17] **J.Gmehling,** 2012,Chemie-Webmaster Stand: 06.04.2018

[18**]Kamil Paduszynski,** J. Chem. Inf. Model., 2016, 56 (8), pp 1420–1437 **DOI:** 10.1021/acs.jcim.6b00166.

[19] **Djedi Sara,** Etude comparative de PCA et KPCA associées au SVM en biométrie, Memoire de master, 2012, Université Mohamed Khider Biskra

[20] **Jamal Kharroubi,** Etude de techniques de classement "Machines à vecteurs supports" pour la vérification automatique du locuteur, Thèse de doctorat, 2002, l'Ecole nationale supérieure des télécommunications

[21] **M. Zaiz Faouzi,** Les Supports vecteurs machines (svm) pour la reconnaissance des caractères manuscrits arabes, Mémoire de master, 2010, Université Mohamed Khider – BISKRA , Algerie

[22] **Marc Lucea,** Modélisation dynamique par réseaux de neurones et machines à vecteurs supports: contribution à la maîtrise des émissions polluantes de véhicules automobiles**,** Thèse de doctorat**,** 2006, L'universite Paris 6

[23] **Heintz, Kulikov & Verevkin,** Propriétés thermodynamiques des mélanges contenant des liquides ioniques. 1. Coefficients d'activité à la dilution infinie d'alcanes, d'alcènes et d'alkylbenzènes dans le tétrafluoroborate de 4-méthyl-n-butylpyridinium en utilisant une chromatographie gaz-liquide, 2001, J. Chem. Eng. Data, 46,1526 1529

[24] **Inoue, Iwai, Yasutake, Honda & Arai,** Mesure des coefficients d'activité de dilution infinie des n-alcanes dans le tétrafluoroborate de 4-méthyl-n-butylpyridinium par chromatographie gaz-liquide,2007, Fluid Phase Equilib,251,17 23

[25] **Bahlmann, Nebig & Gmehling,** 2009, Fluid Phase Equilib,282,113 116

[26] **Foco, Bottini, Quezada, {de la Fuente} & Peters,** Coefficients d'activité à dilution infinie dans les liquides ioniques de tétrafluoroborate de 1-alkyl-3-méthylimidazolium, 2006, J. Chem. Eng. Data 51,1088 1091

[27] **Letcher, Soko, Reddy & Deenadayalu,** Détermination des coefficients d'activité à la dilution infinie de solutés dans le tétrafluoroborate de 1-hexyl-3-méthylimidazolium liquide ionique par chromatographie gaz-liquide aux températures 298,15 K et 323,15 K, 2003, J. Chem. Eng. Data 48,1587 1590

[28] **Zhang, Wang & Li,** Coefficients d'activité à dilution infinie d'alcanes, d'alcènes et de benzènes alkylés dans le tétrafluoroborate de 1- (2-hydroxyéthyl) -3-méthylimidazolium par chromatographie gaz-liquide, 2009, J. Chem. Eng. Data, 54,2887 2890

[29] **Ge, Wang, Wu & Zhou,** 2008, J. Chem. Eng. Data, 53,846 849

[30] **Krummen, Wasserscheid & Gmehling,** Mesure des coefficients d'activité à dilution infinie dans les liquides ioniques à l'aide de la technique de dilution, 2002, J. Chem. Eng. Data, 47, 1411 1417

[31] **Kato & Gmehling,** Coefficients d'activité à dilution infinie de différents solutés dans l'ionique liquides [MMIM] + [CH3SO4] -, [MMIM] + [CH3OC2H4SO4] -, [MMIM] + [(CH3) 2P04] -, [C5H5NC2H5] + [(CF3SO2) 2N] - et [C5H5NH] + [C2H5OC2H4OSO3] -,2004, Fluid Phase Equilib, 226,37 44

[32] **Ge, Lu, Liu, Li, Chen & Xiong,** 2015, J. Chem. Thermodyn, 91,279 285

[33] **Li, Wang, Li & Tian,** 2011, Ind. Eng. Chem. Res, 5,10755 10764

[34] **Domańska & Marciniak,** Propriétés physicochimiques et coefficients d'activité à dilution infinie pour les solutés organiques et l'eau dans le tétracyanoborate de 1-décyl-3-méthylimidazolium liquide ionique,2010, J. Phys. Chem. B, 114,16542 16547

[35] **Moise, Mutelet, Jaubert, Grubbs, Acree & Baker,** Coefficients d'activité à la dilution infinie de composés organiques dans quatre nouveaux liquides ioniques à base d'imidazolium,2011, J. Chem. Eng. Data, 56,3106 3114

[36] **Domańska & Marciniak,** Coefficients d'activité à des mesures de dilution infinies pour les solutés organiques et l'eau dans les liquides ioniques 1-hexyloxyméthyl-3-méthyl-imidazolium et 1,3-dihexyloxyméthyl-imidazolium bis (trifluorométhylsulfonyl) -imide - L'influence des cations,2009, Fluid Phase Equilib, 286,154 161

[37] **Rosenboom, Afzal & Prausnitz,** 2012, J. Chem. Thermodyn, 47,320 327

[38] **Karpińska, Wlazło & Domańska,** 2017, J. Mol. Liq, 225,382 390

[39] **Mutelet, Revelli, Jaubert, Sprunger, Acree & Baker,** 2010, J. Chem. Eng. Data, 55 ,234 242

[40] **Wlazło, Marciniak & Letcher,** 2014, J. Solution Chem, 44,413 430

[41] **Yan, Yang, Liu, Liu, Tan & Welz-Biermann,** 2010, Fluid Phase Equilib, 298,287 292

[42] **Deenadayalu, Letcher & Reddy,** 2005, J. Chem. Eng. Data, 50,105 108

[43] **Heintz, Kulikov & Verevkin,** Propriétés thermodynamiques des mélanges contenant des liquides ioniques. 2. Coefficients d'activité à la dilution infinie d'hydrocarbures et de solutés polaires dans le bis (trifluorométhyl-sulfonyl) amidure de 1-méthyl-3-éthyl-imidazolium et dans le bis (trifluorométhyl-sulfonyl) amidure de 1,2-diméthyl-3-éthyl-imidazolium. Chromatographie gaz-liquide, 2002, J. Chem. Eng. Data, 47,894 899

[44] **Olivier, Letcher, Naidoo & Ramjugernath,** 2010, J. Chem. Thermodyn, 42,78 83

[45] **Domańska & Marciniak,** Mesures des coefficients d'activité à dilution infinie des aromatiques et les hydrocarbures aliphatiques, les alcools et l'eau dans la nouvelle ionique liquide [EMIM] [SCN] utilisant GLC,2008, J. Chem. Thermodyn, 40,860 866

[46] **Domańska & Królikowski,** Mesures de coefficients d'activité à dilution infinie pour les solutés organiques et l'eau dans le tétracyanoborate de 1-hexyl-3-méthylimidazolium liquide ionique,2012, J. Chem. Thermodyn, 47,389 396

[47] **Blahut, Sobota, Dohnal & Vrbka,** 2010, Fluid Phase Equilib, 299,198 206

[48] **Mutelet & Jaubert,** Mesures précises des propriétés thermodynamiques des solutés dans les liquides ioniques par chromatographie en phase gazeuse inverse, 2006, J. Chromatography A, 1102, 256 267

[49] **Domańska, Królikowska, Acree & Baker,** Thermodynamique et coefficients d'activité à des mesures de dilution infinies pour les solutés organiques et l'eau dans le liquide ionique tétracyanoborate de 1-butyl-1-méthylpyrrolidinium,2011, J. Chem. Thermodyn, 43,1810 1817

[50] **Yan, Yang, Liu, Wang, Tan & Welz-Biermann,** 2010, J. Chem. Thermodyn, 42,1415 1422

[51] **Blahut, Dohnal & Vrbka,** 2012, J. Chem. Thermodyn, 47,100 108

[52] **Domańska & Marciniak,** Coefficients d'activité à des mesures de dilution infinie pour les solutés organiques et l'eau dans le trifluoroacétate de 1-éthyl-3-méthylimidazolium liquide ionique,2007, J. Phys. Chem. B, 111,11984 11988

[53] **Paduszyński, Królikowska & Królikowski,** 2017, J. Mol. Liq, in preparation

[54] **Marciniak & Wlazło,** Coefficients d'activité à dilution infinie et propriétés physicochimiques pour les solutés organiques et l'eau dans le liquide ionique 1- (2-hydroxyéthyl) - 3-méthylimidazolium trifluorotris (perfluoroéthyl) phosphate,2013, J. Chem. Thermodyn, 64,114 119

[55] **Singh, Bahadur, Naidoo, Redhi & Ramjugernath,** 2016, J. Mol. Liq, 220,33 40

[56] **Orfao, Dohnal & Blahut,** Coefficients d'activité de dilution infinie de composés organiques volatils dans deux liquides ioniques composés de l'anion tris (pentafluoroéthyl) trifluorophosphate ([FAP]) et d'un cation fonctionnalisé, 2013, J. Chem. Thermodyn, 65,53 64

[57] **Revelli, Mutelet, Jaubert, Garcia-Martinez, Sprunger, Acree & Baker,** 2010, J. Chem. Eng. Data, 55,2434 2443

[58] **Stark, Ondruschka, Zaitsau & Verevkin,** Produits chimiques de plate-forme dérivés de la biomasse: Études thermodynamiques sur l'extraction du 5-hydroxyméthylfurfural des liquides ioniques, 2012, J. Chem. Eng. Data, 57,2985 2991

[59] **Passos, Khan, Mutelet, Oliveira, Carvalho, Santos, Held, Sadowski, Freire & Coutinho,** Équilibre vapeur-liquide de l'eau + liquides ioniques à base d'alkylimidazolium: mesures et modélisation de la théorie des fluides associée à la chaîne perturbation Modélisation de la théorie des fluides 2014, Ind. Eng. Chem. Res, 53,3737 3748

[60] **Khan, Kurnia, Mutelet, Pinho & Coutinho,** Sonder les interactions entre les liquides ioniques et l'eau: approche chimique expérimentale et quantique,2014, J. Phys. Chem. B, 118,1848 1860

[61] **Lukoshko, Mutelet & Domańska,** Etude expérimentale et théorique de l'interaction entre les composés organiques et les liquides ioniques à base de tricyanométhanides,2015, J. Chem. Thermodyn, 85,49 56

[62] **Martins, Coutinho, Pinho & Domańska,** 2015, J. Chem. Thermodyn, 91,194 203

[63] **Domańska, Wlazło & Karpińska**, Coefficients d'activité à dilution infinie des solvants organiques et de l'eau dans le 1-butyl-3-méthylimidazolium dicyanamide. Une revue de la littérature sur la séparation hexane / hex-1-ène,2016, Fluid Phase Equilib, 417,50 61

[64] **Heintz, Casas, Nesterov, Emel'yanenko & Verevkin,** Propriétés thermodynamiques des mélanges contenant des liquides ioniques. 5. Coefficients d'activité à la dilution infinie d'hydrocarbures, d'alcools, d'esters et d'aldéhydes dans le bis (trifluorométhyl-sulfonyl) imide de 1-méthyl-3-butyl-imidazolium par chromatographie gaz-liquide,2005, J. Chem. Eng. Data

, 50,1510 1514

[65] **Singh, Bahadur, Naidoo, Redhi & Ramjugernath,** Application de 1-butyl-3-méthylimidazolium bis (trifluorométhylsulfonyl) imide liquide ionique pour les différents types de problèmes de séparation: Coefficients d'activité à des mesures de dilution infinies en utilisant la technique de chromatographie gaz-liquide,2016, J. Mol. Liq, 220,33 40

[66] **Zhang, Zhang, Qiao & Deng,** Solubilités des solutés gazeux et liquides et leur thermodynamique de solubilisation dans les nouveaux liquides ioniques à température ambiante à dilution infinie par chromatographie en phase gazeuse,2007, J. Chem. Eng. Data, 52,2277 2283

[67] **Domańska & Marciniak,** Coefficients d'activité à des mesures de dilution infinie pour les solutés organiques et l'eau dans le trifluorométhanesulfonate de 1-butyl-3-méthylimidazolium liquide ionique,2008, J. Phys. Chem. B, 112,11100 11105

[68] **Ge, Wang, Li & Wu,** 2007, J. Chem. Eng. Data, 52,2257 2260

[69] **Nebig & Gmehling,** Mesures de différentes propriétés thermodynamiques de systèmes contenant des liquides ioniques et prédiction de ces propriétés à l'aide de l'UNIFAC modifié,2010, Fluid Phase Equilib, 294,206 212

[70] **Domańska & Laskowska,** Coefficients d'activité à des mesures de dilution infinies pour les solutés organiques et l'eau dans le liquide ionique 1-butyl-1-méthylpyrrolidinium trifluorométhanesulfonate en utilisant GLC,2009, Fluid Phase Equilib, 278,97 102

[71] **Domańska & Królikowski,** Mesures par chromatographie gaz-liquide des coefficients d'activité à une dilution infinie de divers solutés organiques et de l'eau dans liquide ionique de tosylate de tri-iso-butylméthylphosphonium,2010, J. Chem. Thermodyn, 42,707 711

[72] **Li, Wang, Li, Deng, Zhang & Tian,** Détermination des paramètres thermodynamiques du chlorure de 1-hexyl-3-méthylimidazolium liquide ionique par chromatographie en phase gazeuse inverse,2014, J. Mol. Liq, 200 ,139 144

[73] **Heintz, Verevkin & Ondo,** 2006, J. Chem. Eng. Data, 51,648 655

[74] **Kato & Gmehling,** 2005, J. Chem. Thermodyn, 37,603 619

[75] **Liebert, Nebig & Gmehling,** Équilibre de phase expérimental et prédit et propriétés excédentaires pour les systèmes contenant des liquides ioniques,2008, Fluid Phase Equilib, 268,14 20

[76] **Yang, Wu, Ge, Wang & Li,** Coefficients d'activité à dilution infinie d'alcanes, d'alcènes et de benzènes alkyles dans le trifluorométhanesulfonate de 1-hexyl-3-méthylimidazolium par chromatographie gaz-liquide,2008, J. Chem. Eng. Data, 53,1220 1222

[77] **Domańska, Marciniak, Królikowska & Arasimowicz,** Mesures de coefficients d'activité à dilution infinie dans des mélanges de solvants avec des liquides ioniques à base de thiocyanate en utilisant la technique GLC,2010, J. Phys. Chem. B, 114,8460 8466

[78] **Domańska, Lukoshko & Wlazło,** 2012, J. Chem. Thermodyn, 54,20 27

[79] **Ge, Zhang, Li, Li, Zhang & Mu,** Thermodynamique et coefficients d'activité à dilution infinie pour les solutés organiques dans le liquide ionique 1-hexyl-2,3-diméthylimidazolium le bis (trifluorométhylsulfonyl) imide,2016, J. Chem. Thermodyn, 102,303 309

[80] **Wlazło & Marciniak,** Coefficients d'activité à dilution infinie et propriétés physicochimiques des solutés organiques et de l'eau dans le liquide ionique 4- (2-méthoxyéthyl) -4-méthylmorpholinium trifluorotris (perfluoroéthyl) phosphate,2012, J. Chem. Thermodyn, 54,366 372

[81] **Marciniak & Wlazło**, Coefficients d'activité à dilution infinie et propriétés physicochimiques pour les solutés organiques et l'eau dans le liquide ionique 4- (2-méthoxyéthyl) - 4-méthylmorpholinium bis (trifluorométhylsulfonyl) -amide,2012, J. Chem. Thermodyn, 47,382 388

[82] **Domańska & Lukoshko,** Thermodynamique et coefficients d'activité à dilution infinie pour les solutés organiques, l'eau et les diols dans le liquide ionique choline bis (trifluorométhylsulfonyl) imide,2014, J. Chem. Thermodyn, 77,63 70

[83] **Domańska, Papis & Szydłowski,** 2014, J. Chem. Thermodyn, 68,53 59

[84] **Domańska & Paduszyński,** Détermination des coefficients d'activité à dilution infinie de 35 solutés dans le liquide ionique, tosylate de 1-butyl-3-méthylimidazolium,
En utilisant la chromatographie gaz-liquide,2010, J. Chem. Eng. Data, 55,4817 4822

[85] **Zhou & Wang,** 2006, J. Chem. Eng. Data ,51,1698 1701

[86] **Revelli, Mutelet, Turmine, Solimando & Jaubert,**2008, J. Chem. Eng. Data,54,90 101

[87] **Domańska & Paduszyński,** Coefficients d'activité à des mesures de dilution infinie pour les solutés organiques et l'eau dans le thiocyanate de 1-hexyl-3-méthylimidazolium liquide ionique,2010, J. Chem. Eng. Data, 55,2532 2536

[88] **Paduszyński & Domańska,** Coefficients d'activité limitants et coefficients de partage gaz-liquide de différents solutés dans des liquides ioniques de pipéridinium: mesures et
Calculs LSER,2011, J. Phys. Chem. B, 115,8207 8215

[89] **Domańska & Królikowska,** Mesures de coefficients d'activité à dilution infinie de composés organiques et d'eau dans un liquide ionique à base d'isoquinolinium [C8iQuin] [NTf2] en utilisant GLC, 2011, J. Chem. Thermodyn, 43,499 504

[90] **Paduszyński & Domańska,** Etude expérimentale et théorique sur l'activité de dilution infinie coefficients de différents solutés dans les liquides ioniques pipéridinium,2013, J. Chem. Thermodyn, 60,169 178

[91] **Yan, Liu, Yang, Liu, Tan & Welz-Biermann,** 2010, J. Chem. Thermodyn, 42,817 822

[92] **Marciniak & Wlazło,** Coefficients d'activité à des mesures de dilution infinie pour les solutés organiques et l'eau dans le liquide ionique 1- (3-Hydroxypropyl) pyridinium Trifluorotris (perfluoroéthyl) phosphate,2010, J. Phys. Chem. B, 114,6990 6994

[93] **Marciniak,** Coefficients d'activité à dilution infinie et propriétés physicochimiques
solutés organiques et eau dans le liquide ionique 1- (3-hydroxypropyl) pyridinium
bis (trifluorométhylsulfonyl) -amide,2011, J. Chem. Thermodyn., 43,1446 1452

[94]**Yan, Liu, Yang, Liu, Tan & Welz-Biermann,** 2010, J. Chem. Thermodyn, 42,817 822

[95] **Marciniak & Wlazło,** Coefficients d'activité à des mesures de dilution infinie pour les solutés organiques et l'eau dans le trifluorométhanesulfonate de 1-butyl-3-méthyl-pyridinium liquide ionique,2010, J. Chem. Eng. Data, 55,3208 3211

[96] **Letcher, Ramjugernath, Królikowski, Laskowska, Naidoo & Domańska,** Coefficients d'activité à des mesures de dilution infinie pour des solutés organiques dans le liquide ionique tosylate de N-butyl-4-méthylpyridinium en utilisant GLC à T = (328.15, 333.15, 338.15 et 343.15) K,2009, Fluid Phase Equilib, 276,31 36

[97] **Królikowski & Królikowska,** L'étude des coefficients d'activité à dilution infinie pour les solutés organiques et l'eau dans le 1-butyl-4-méthylpyridinium dicyanamide, [B4MPy] [DCA] en utilisant GLC,2014, J. Chem. Thermodyn, 68,138 144

[98] **Domańska & Marciniak,** Coefficients d'activité à des mesures de dilution infinies pour les solutés organiques et l'eau dans le liquide ionique 4-méthyl-N-butyl-pyridinium
bis (trifluorométhylsulfonyl) -imide, 2009, J. Chem. Thermodyn, 41, 1350 1355

[99] **Domańska & Królikowska**, Mesures de coefficients d'activité à dilution infinie de solutés organiques et d'eau dans le bis {(trifluorométhyl) sulfonyl} imide de 1-propyl-1-méthylpipéridinium liquide ionique utilisant g.l.c,2010, J. Chem. Thermodyn, 42,1361 1366

[100] **Domańska & Królikowski,** Thermodynamique et coefficients d'activité à la dilution infinie Mesures pour les solutés organiques et l'eau dans le tosylate de N-hexyl-3-méthylpyridinium liquide ionique,2011, J. Phys. Chem. B, 155,7397 7404

[101] **Marciniak & Wlazło,** Coefficients d'activité à dilution infinie et propriétés physicochimiques des solutés organiques et de l'eau dans le liquide ionique 1- (2-méthoxyéthyl) -1-méthylpyrrolidinium bis (trifluorométhylsulfonyl) -amide,2012, J. Chem. Thermodyn, 54,90 96

[102] **Domańska & Lukoshko,** Mesures des coefficients d'activité à dilution infinie pour l'organique solutés et eau dans le liquide ionique 1-butyl-1-méthylpyrrolidinium tricyanométhanide, 2013, J. Chem. Thermodyn, 66,144 150

[103] **Blahut & Dohnal,** Interactions des Composés Organiques Volatils avec le Liquide Ionique 1-Butyl-1-methylpyrrolidinium Dicyanamide,2011, J. Chem. Eng. Data, 56,4909 4918

[104] **Domańska, Lukoshko & Królikowski,** Mesures des coefficients d'activité à dilution infinie pour les solutés organiques et l'eau dans le liquide ionique 1-butyl-1-méthylpyrrolidinium
le tris (pentafluoroéthyl) trifluorophosphate ([BMPYR] [FAP]),2012, Chem. Eng. J, 183,261 270

[105] **Mutelet, Hassan, Stephens, Acree & Baker,** Coefficients d'activité à dilution infinie pour les solutés organiques dissous dans trois liquides ioniques de bis (trifluorométhylsulfonyl) imide de 1-

alkyl-1-méthylpyrrolidinium portant des chaînes latérales linéaires courtes de trois à cinq carbones,2013, J. Chem. Eng. Data, 58,2210 2218

[106] **Domańska, Redhi & Marciniak,** Mesures de coefficients d'activité à dilution infinie d'hydrocarbures aliphatiques et aromatiques, d'alcools, de thiophène, de tétrahydrofurane, de MTBE et d'eau dans un liquide ionique [BMIM] [SCN] en utilisant GLC,2009, J. Chem. Thermodyn, 41,645 650

[107] **Domańska, Królikowski & Acree,** Mesures des coefficients d'activité à la dilution infinie pour les solutés organiques et l'eau dans le thiocyanate de 1-butyl-1-méthylpipéridinium liquide ionique,2011, J. Chem. Eng. Data, 56,124 129

[108] **Blahut & Dohnal,** 2013, J. Chem. Thermodyn, 57,344 354

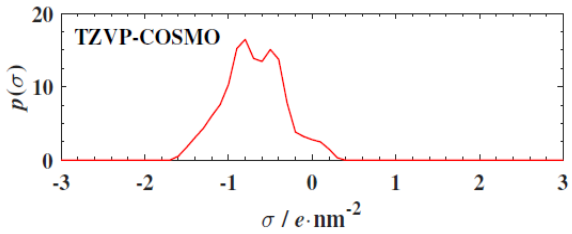
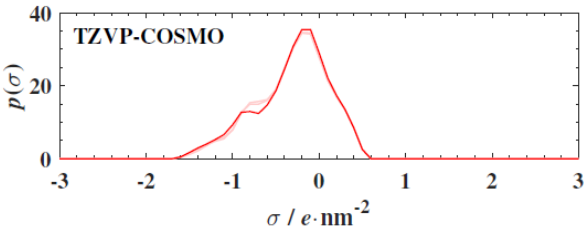
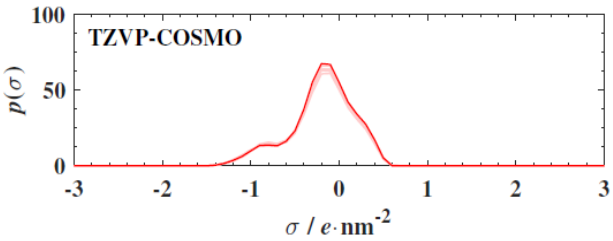
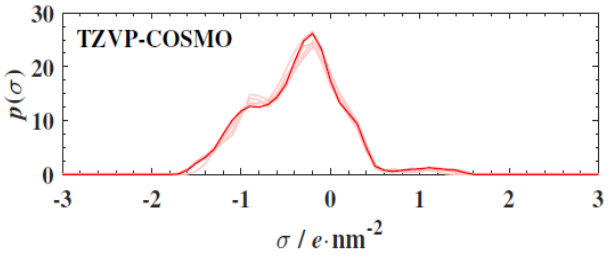
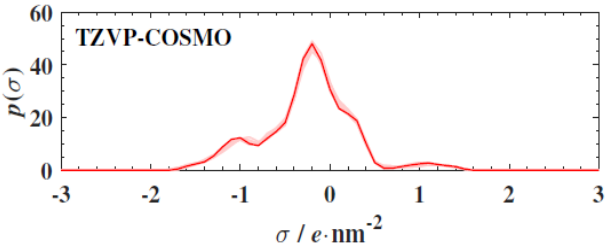
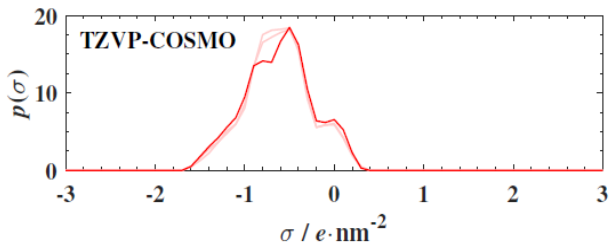
[109] **Królikowska, Karpińska & Królikowski,** La mesure des coefficients d'activité a été réalisée pour la détermination de solutés organiques et d'eau dans le thiocyanate de N-hexylisoquinolinium, [HiQuin] [SCN] en utilisant GLC,2013, J. Chem. Thermodyn, 62,1 7

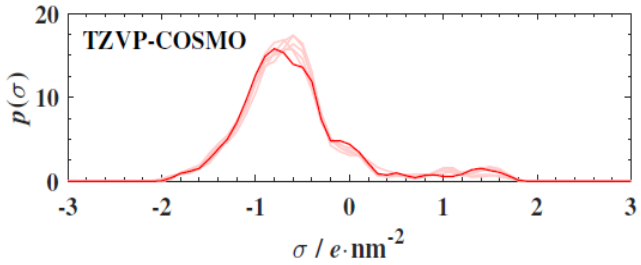
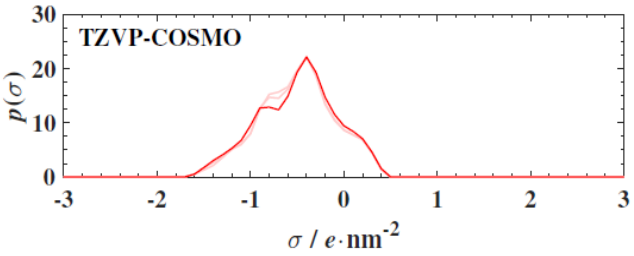
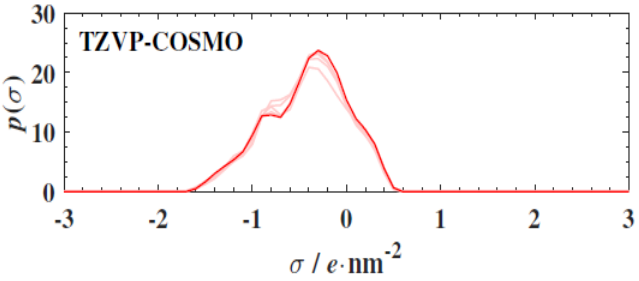
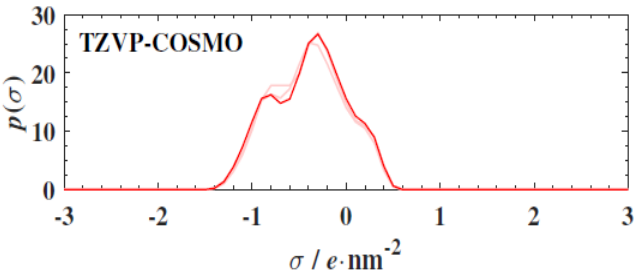
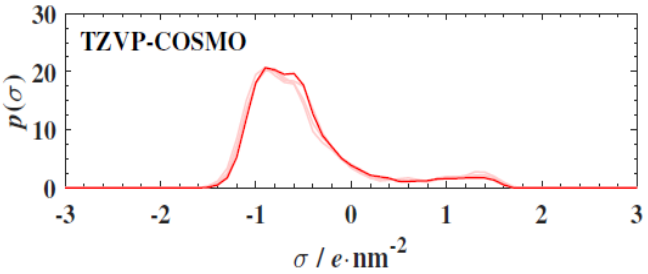
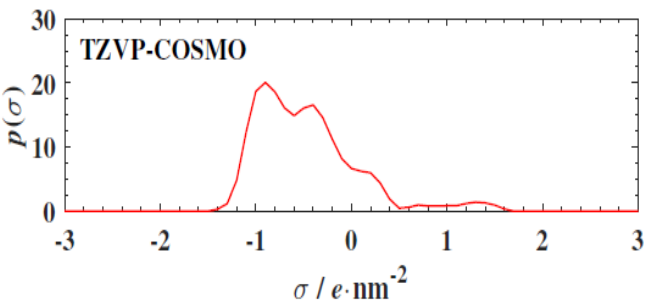
[110] **Ayad, Mutelet, Negadi, Acree, Jiang, Lu, Wagle & Baker,** 2016, J. Mol. Liq, 215,176 184

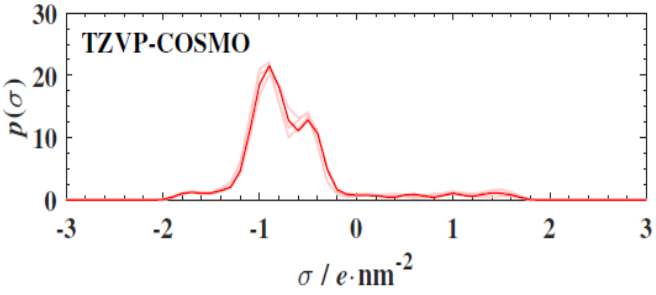
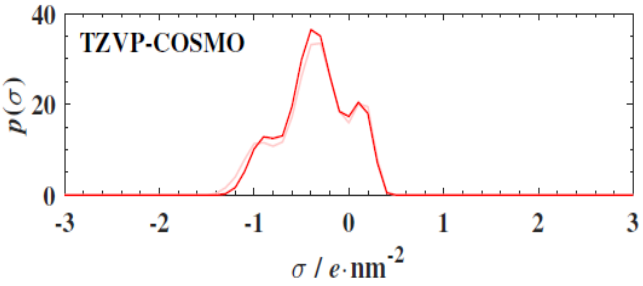
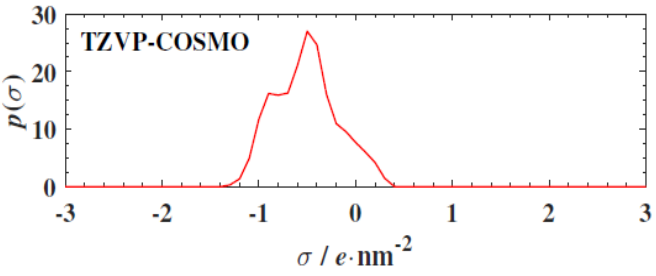
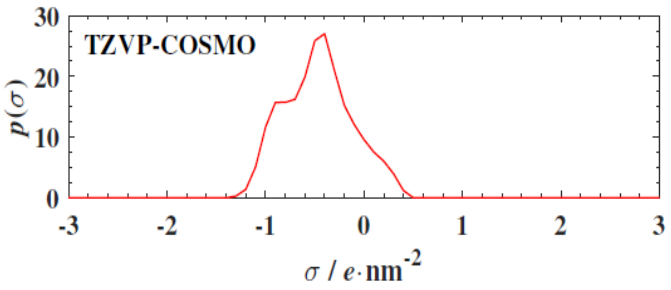
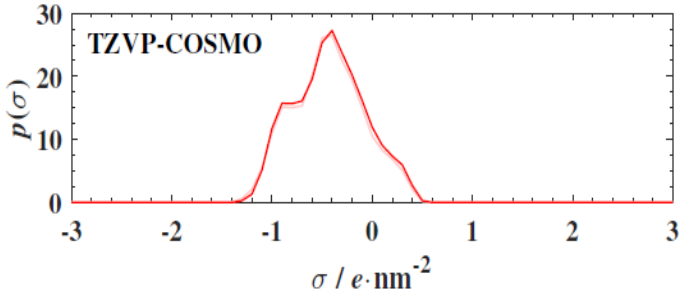
[111] **Domańska & Paduszyński,** 2010, J. Chem. Eng. Data, 55,4817 4822

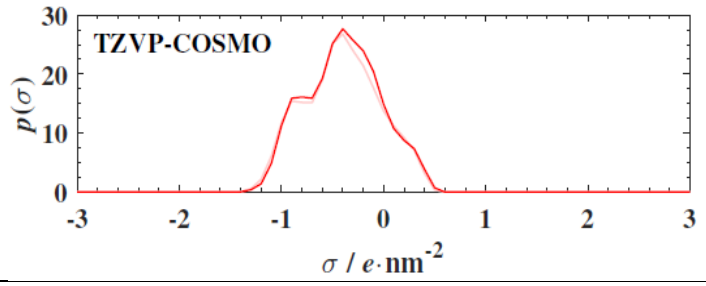
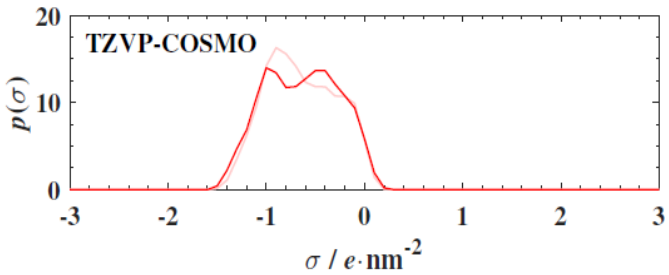
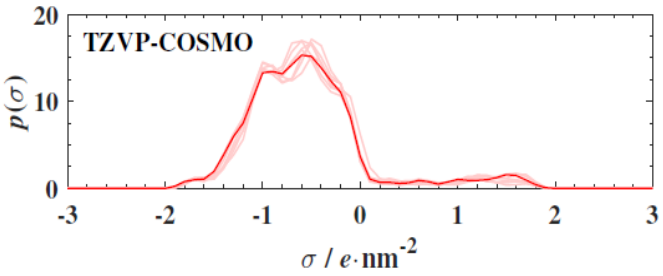
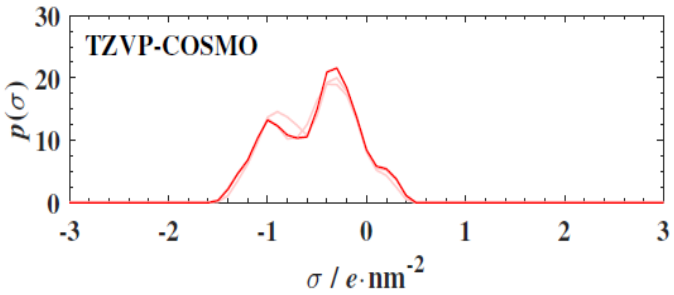
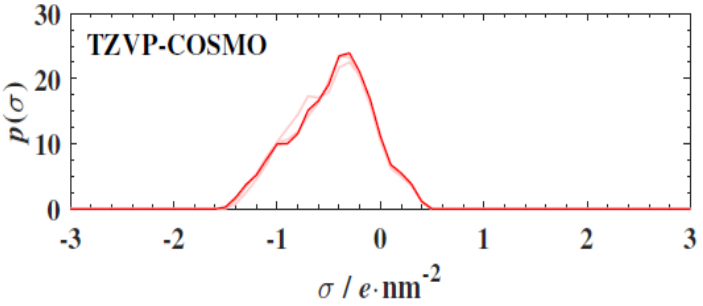
ANNEXE

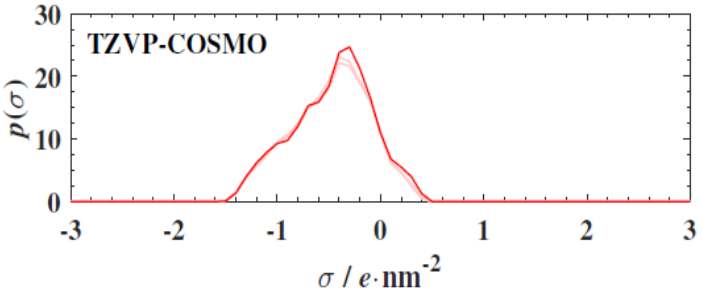
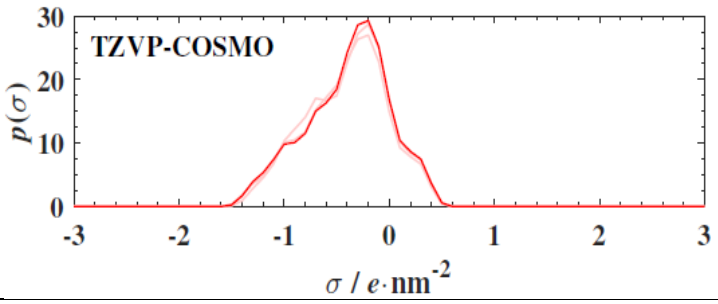
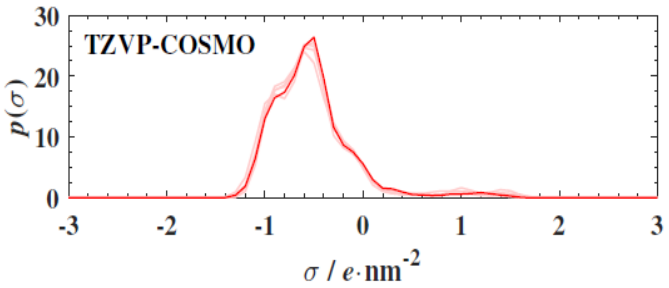
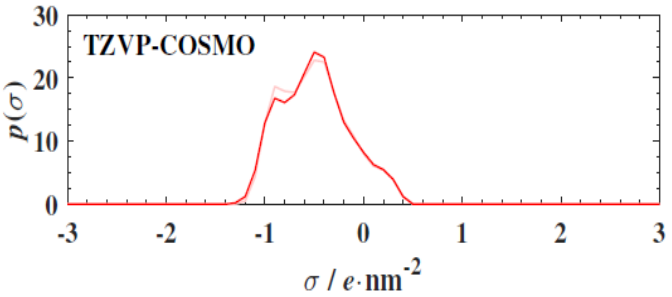
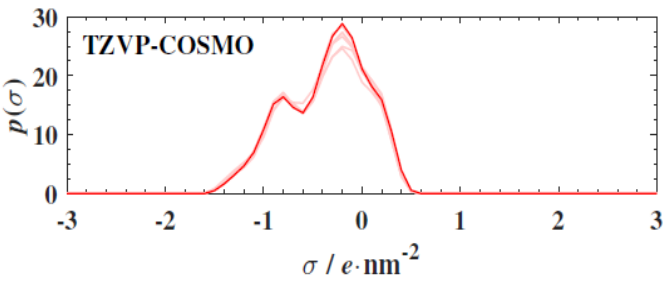
Tableau A.1 :Profils σ des cations

N^0	Nom de cations	Profil σ
1	1,3-dimethyl-1H-imidazol-3-ium	
2	1-decyl-3-methyl-1H-imidazol-3-ium	
3	1,3-didecyl-2-methylimidazolium	
4	3-[(hexyloxy)methyl]-1-methyl-1H-imidazol-3-ium	
5	1,3-bis[(hexyloxy)methyl]-1H-imidazol-3-ium	
6	1-ethyl-3-methyl-1H-imidazol-3-ium	

7	1-(2-hydroxyethyl)-3-methyl-1H-imidazol-3-ium	
8	1-butyl-3-methyl-1H-imidazol-3-ium	
9	3-hexyl-1-methyl-1H-imidazol-3-ium	
10	3-hexyl-1,2-dimethyl-1H-imidazol-3-ium	
11	4-(2-methoxyethyl)-4-methylmorpholin-4-ium	
12	4-butyl-4-methylmorpholin-4-ium	

13	(2-hydroxyethyl)trimethylazanium	
14	methyltris(2-methylpropyl)phosphanium	
15	1-methyl-1-propylpiperidin-1-ium	
16	1-butyl-1-methylpiperidin-1-ium	
17	1-methyl-1-pentylpiperidin-1-ium	

18	1-hexyl-1-methylpiperidin-1-ium	
19	1-ethylpyridin-1-ium	
20	1-(3-hydroxypropyl)pyridin-1-ium	
21	1-butylpyridin-1-ium	
22	1-butyl-3-methylpyridin-1-ium	

23	1-butyl-4-methylpyridin-1-ium	
24	1-hexyl-3-methylpyridin-1-ium	
25	1-(2-methoxyethyl)-1-methylpyrrolidin-1-ium	
26	1-butyl-1-methylpyrrolidin-1-ium	
27	2-hexylisoquinolin-2-ium	

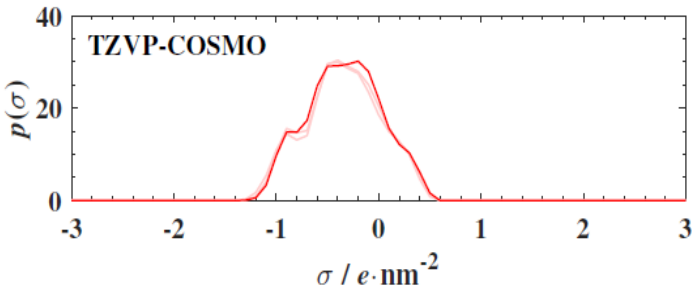
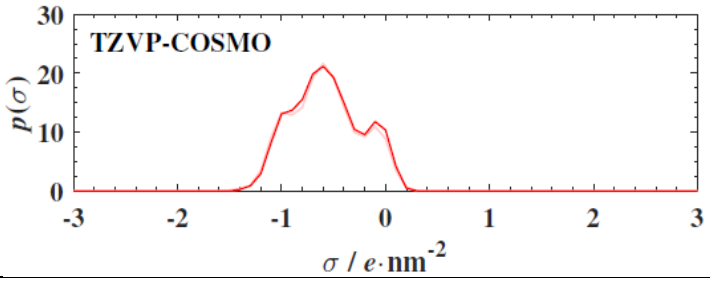
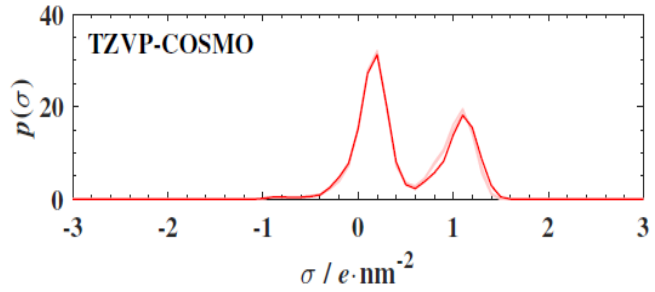
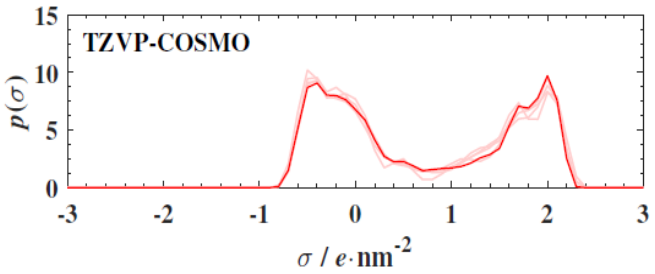
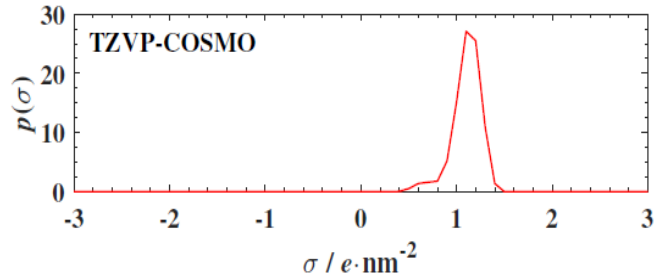
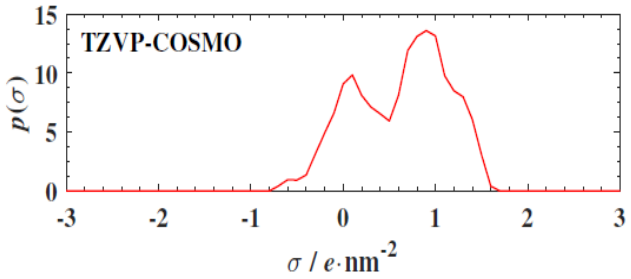
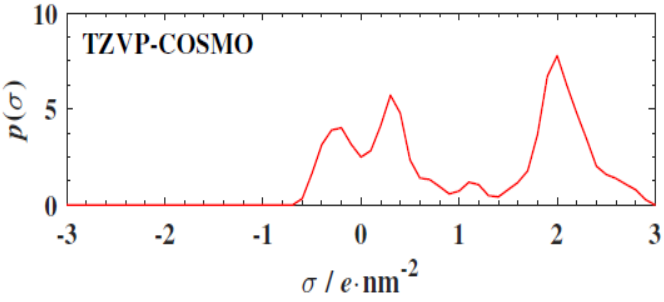
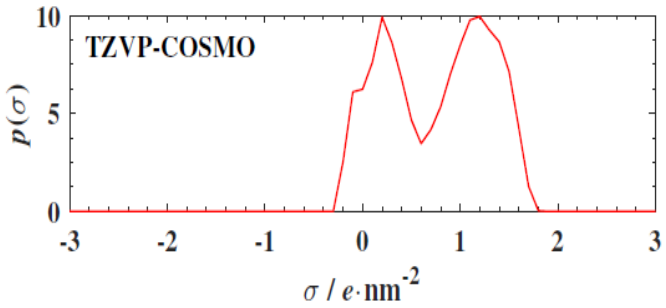
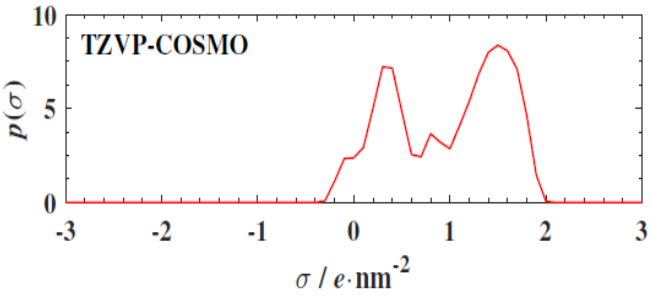
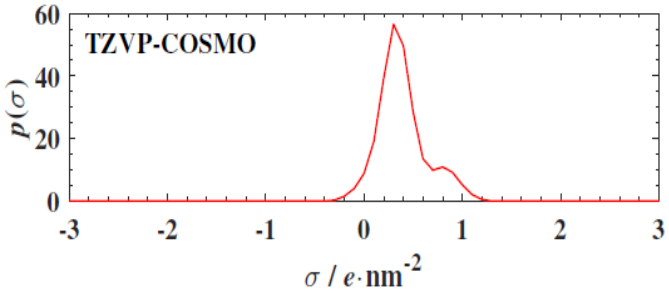
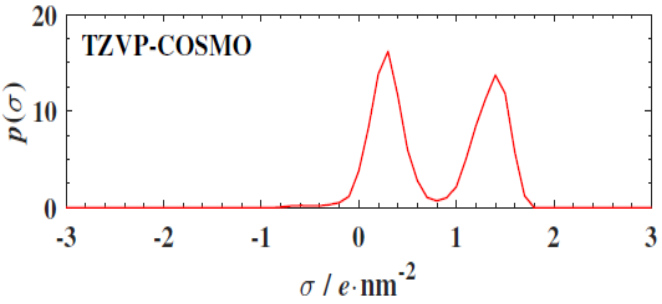
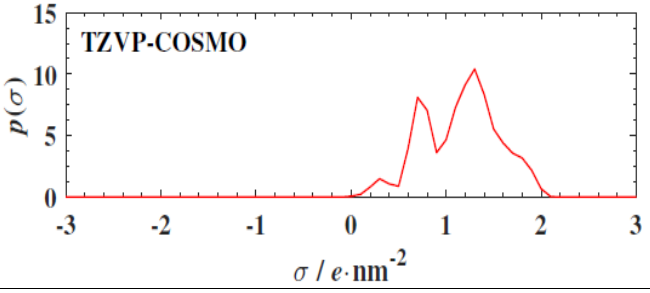
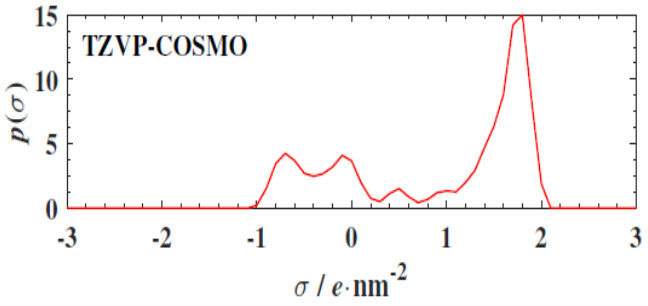
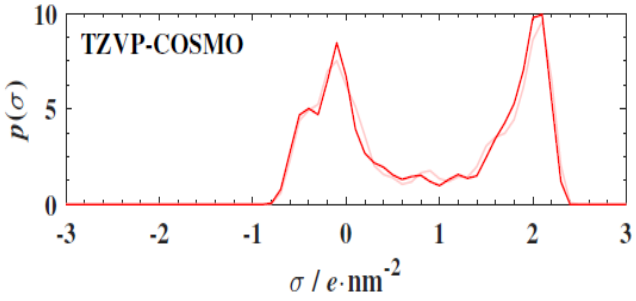
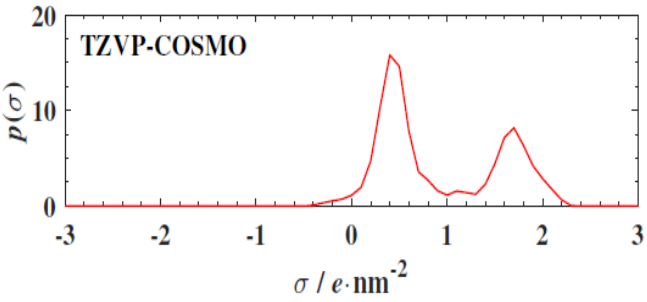
28	2-octylisoquinolin-2-ium	
29	triethylsulfanium	

Tableau A.2 : Profils σ des anions

N^0	Nom de l'anions	Profils σ
1	bis(pentafluroethylsulfonyl)azanide	
2	dimethyl phosphate	
3	tetrafluoroborate	
4	tetracyanoboranuide	
5	Acetate	

6	tricyanomethide	
7	dicyanoazanide	
8	trifluorotris(pentafluoroethyl) phosphanuide	
9	trifluoromethanesulfonate	
10	thiocyanate	

11	methanesulfonate	
12	4-methylbenzene-1-sulfonate	
13	trifluoroacetate	
14	chloride	