

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Ibn Khaldoun – Tiaret



Faculté des Sciences de la Matière
Département de Physique

Mémoire

Présenté par :

M^{elle} : ARBAOUI KHALDIA

M^{elle} : AOUINE RACHIDA

Pour obtenir le diplôme de

Master

Filière : Physique
Spécialité: Génie des matériaux

Sujet :

Application d'analyse en composantes principales (ACP) à une base de données de liquides ioniques

Soutenu le : 09 juin 2015

Devant le jury:

M.HAOUZI Ahmed	Prof.	U IK Tiaret	Président
M. DEBDAB Mansour	MCA	UIK Tiaret	Examineur
M. BENAICHATA	MAA	UIK Tiaret	Examineur
Lazreg			
BELARBI El Habib	Prof.	UIK Tiaret	Encadreur

REMERCIEMENTS

*Ce travail a été réalisé au laboratoire synthèse et catalyse de l'université **IBN KHALDOUN** de **TIARET**.*

*Nous tenons à remercier vivement **M. BELARBI H.***

*Professeur à l'université **IBN KHALDOUN** **TIARET** pour avoir*

dirigé ce travail de master II et de nous avoir fait partager

son savoir faire et sa rigueur scientifique, ainsi que ses multiples compétences.

Qu'il soit assuré de toute notre gratitude et de notre reconnaissance.

*Nous tenons à remercier vivement. Monsieur le Professeur **HAOUZI A.** pour*

Avoir bien voulu présider le jury de ce mémoire.

*Nous tenons à remercier chaleureusement **M. DEBDAB M.** et*

***M. BENAICHATA L.** enseignants chercheurs à l'Université Ibn Khaldoun Tiaret pour leurs conseils et orientations et pour avoir bien voulu évaluer ce travail.*

Il est également venu le temps de remercier tous les gens

qui nous ont conseillées, nous ont aidées et nous ont encouragées.

Enfin, que toutes les personnes qui nous ont aidées d'une manière ou d'une autre, trouvent ici l'expression de notre sincère reconnaissance.

Liste des figures

Liste des figures

Figure	Titre de figures	page
Figure I-1	Exemple de technique non-supervisé (classification)	05
Figure II-1	Les liquides ioniques sont un sous ensemble des sels fondus	18
Figure II-2	Exemples des cations et des anions constituant les liquides ioniques	19
Figure III-1	Figure qui représente le logiciel XLSTAT	43
Figure VI-1	Répartition de l'inertie entre les axes	52
Figure VI-2	Le cercle de corrélation (plan factorielle F1, F2)	53
Figure VI-3	représentation des observations dans le plan factoriel (F1, F2)	54
Figure VI-4	le cercle de corrélation (plan factoriel F1, F3)	55
Figure VI-5	représentation observations dans plan factoriel (F1, F3)	55
Figure VI-6	le cercle de corrélation (plan factoriel F2, F3)	56
Figure VI-7	Représentation des observations dans plan factorielle (F2, F3)	56
Figure VI-8	Carte factorielle des propriétés physique-chimique des liquides ioniques	58

Liste des tableaux

Tableaux	Titre des tableaux	page
Tableau III-1	Les propriétés physique-chimique des liquides ioniques	39-42
Tableau IV-1	Valeurs propres	51
Tableau IV-2	Classification des trois groupes principaux à base de données	59
Tableau IV-3	Représentation le 1 ^{er} groupe des liquides ioniques	59
Tableau IV-4	Représentation le 2 ^{ème} groupe des liquides ioniques	60
Tableau IV-5	Représentation le 3 ^{ème} groupe des liquides ioniques	60

Liste des abréviations

LI : Liquide ionique	
cations	anions
<p>IM : imidazolium MIM : methylimidazolium MMIM : 1,3-dimethylimidazolium EMIM : 1-ethyl-3-methylimidazolium EMMIM : 1-ethyl-2,3-dimethylimidazolium PMIM : 1-propyl-3-methylimidazolium PMMIM : 1-propyl-2,3-dimethylimidazolium BMIM : 1-butyl-3-methylimidazolium BMMIM : 1-butyl-2,3-dimethylimidazolium HMIM : 1-hexa-3-methylimidazolium OMIM ou C₈MIM : 1-octyl-3-methylimidazolium C₁₀MIM : 1-decyl-3-methylimidazolium C₁₂MIM : 1-dodecyl-3-methylimidazolium EtChol : ethylcholine , cPhChol : phenylcholine</p>	<p>NO₃⁻ : nitrate SO₄²⁻ : sulfate BF₄⁻ : tétrafluoroborate PF₆⁻ : hexafluorophosphate CH₃CO²⁻ : acétate CF₃CO²⁻ : trifluoroacétate CF₃SO³⁻ ou Tf : trifluorométhylsulfonate N(CF₃SO²⁻)₂ ou NTf₂ : bis(trifluorométhylsulfonyl)imide N(CN)₂ : dicyanamide. Cl⁻ : chlorure Br⁻ : bromure I⁻ : iodure</p>

Tables des matières

Remerciements

Liste des figures

Liste des tableaux

Liste des abréviations

Sommaire

Introduction générale

01

Chapitre I techniques du Data mining

I.1.Introduction	03
I.2.Définition	04
I.3.Principe	04
I.4 .Type de techniques de datamining	05
I. 4.1.Techniques descriptives	05
I.4.2.Techniques prédictive	06
I.5. Conclusion	07
I.2. Méthode d'analyse en composante principales (ACP)	07
I.2.1.Introduction	07
I.2.2. Principe en d'analyse en composantes principales	09
I.2.3. Identification du modèle ACP	11
I.2.4. Détermination de la structure du modèle ACP	13

I.2.4.1. Pourcentage cumulé de la variance totale (PCV)	14
I.2.4.2. Moyenne des valeurs propres	14
I.2.5. Conclusion	15
Références	16

Chapitre II généralités Les liquides ioniques

II.1. Introduction	18
II.2. Définition	18
II.3. Historique	19
II.4. Principales propriétés physico-chimiques	20
II.4.1. Transition vitreuse	20
II.4.2. Température de fusion	21
II.4.3. Stabilité thermique	22
II.4.4. Densité	22
II.4.5. Viscosité	23
II.4.6. Polarité	23
II.4.7. Propriétés électrochimique	24
II.4.7.1. La fenêtre électrochimique	25
II.4.7.2. La conductivité	25
II.5. Application des liquides ioniques	26
II.5.1. Application en électrochimie	26
II.5.2. Application en synthèse organique et en catalyse	26

II.5.3. Application dans le domaine des procédés de séparation et de l'analyse	27
II.6. Conclusion	28
Références	29

Chapitre III présentation du Logiciel XLSTAT

III.1 Introduction	39
III.2 Base de données des liquides ioniques	39-42
III.3 Logiciel XLSTAT	42
III.3.1 Application de l'analyse en composantes principales (ACP)	43
III.3.2 Paramétrer une analyse en composantes principales	44-48
III.4 Conclusion	48
Références bibliographiques	49

Chapitre IV Résultats et Discussion

IV-1-Introduction	51
IV-2- Résultats de l'ACP	51
IV-3-discussions	52-59
IV-4 Conclusion	61
Conclusion générale	63

ANNEXE

Introduction générale

Le domaine de recherche dans les procédés chimiques, biochimiques et environnementaux constitue l'un des domaines les plus vastes du développement scientifique et industriel.

Dans ce travail, nous proposons une classification des liquides ioniques qui est liée à leurs propriétés physico-chimiques à savoir la conductivité électrique et molaire, la densité, la viscosité et le point de fusion.

Une base de données des liquides ioniques est analysée par une technique d'exploration des données communément appelée « data mining ». Cette technique est constituée d'un ensemble de méthodes et d'algorithmes destinés à l'analyse des données dans un objectif de trouver des règles, des tendances inconnues ou en résumé, un nouveau savoir à partir d'un savoir déjà acquis.

Parmi les méthodes utilisées dans ce contexte, la méthode d'analyse en composantes principales (ACP) a été choisie pour être appliquée à une base de données de liquides ioniques qui diffèrent soit par la nature de l'anion ou celle du cation. L'application de l'ACP permettra de tirer certaines associations telles que l'influence de la viscosité ou la densité sur la conductivité électriques des liquides ioniques.

Le mémoire présenté est subdivisé en 4 chapitres répartis comme suit :

Nous présenterons tout d'abord dans le chapitre I, la technique de data mining.

Le second chapitre sera dédié aux rappels bibliographiques concernant les liquides ioniques et leurs propriétés physico-chimiques et application dans différents domaines.

Le chapitre III est consacré à la description du logiciel XLSTAT avec la description des outils utilisés.

Le dernier chapitre résume les principaux résultats obtenus par l'application de l'ACP ainsi que leur discussion.

Enfin, on termine avec une conclusion générale sur le travail réalisé.

Techniques du Data Mining

I. 1. Introduction

Le datamining est une technique d'exploration de données, aussi connue sous les noms fouille de données, ou Extraction de Connaissances (*ECD* en français, de l'anglais *KDD*), a pour objet l'extraction d'un *savoir* ou d'une connaissance à partir de grandes quantités de données, par des méthodes automatiques ou semi-automatiques [1].

Il a vocation à être utilisé dans un environnement professionnel, et se distingue de l'analyse de données par des techniques de la statistique par les points suivants :

- Contrairement aux méthodes statistiques, le Datamining ne nécessite jamais que l'on établisse une hypothèse de départ qu'il s'agira de vérifier. C'est à partir des données elles-mêmes que se dégageront les corrélations intéressantes, l'algorithme n'étant là que pour les découvrir et non pour les établir. Il se situe à la croisée des statistiques, de l'intelligence artificielle.

- Les connaissances extraites par le Datamining ont vocation à être intégrées dans un schéma organisationnel. Le Datamining impose donc d'être capable d'utiliser de manière opérationnelle les résultats des analyses effectuées, souvent dans des délais très courts. Le processus d'analyse doit permettre à l'organisation une réactivité (très) importante.

- Les données traitées sont issues des systèmes de stockage en place dans une base de données et sont ainsi hétérogènes, multiples, plus ou moins structurées. Pour résumer ce sont des données dont la raison d'être n'est *a priori* pas l'analyse. Cela impose de disposer de systèmes performants de préparation ou de manipulation de données.

Le Datamining se propose donc de transformer en information, de grands volumes de données qui peuvent être stockées dans des bases de données relationnelles, mais qui peuvent aussi être récupérées de sources riches plus ou moins structurées comme Internet.

I.2. Définition

Le Datamining est l'ensemble des méthodes, et algorithmes destinés à l'exploitation et l'analyse de données, en vue de détecter dans ces données des règles, des associations, des tendances inconnues des structures particulières restituant de façon concise l'essentiel de l'information utile pour aider une décision, en d'autre terme c'est l'acquisition d'un nouveau savoir à partir d'un savoir déjà acquis.

I. 3. Principe

Le Datamining est un processus d'analyse dont l'approche est différente de celle utilisée en *statistique*. Cette dernière présuppose en général que l'on se *fixe une hypothèse* que les données vont nous permettre ou non de confirmer. Au contraire, le *Data mining* adopte une démarche beaucoup plus empirique et essaye ainsi de faire *émerger*, à partir des données brutes, des hypothèses que l'expérimentateur peut ne pas soupçonner, et dont il aura à valider la pertinence.

Plus qu'une théorie normalisée, le Datamining est un processus d'extraction de connaissances en suivant les étapes principales suivantes :

- Formaliser un problème que l'on cherche à résoudre en terme de données
- Accéder aux données idoines quelles qu'elles soient
- Préparer les données en vue des traitements et utilisations futurs
- Modéliser les données en leur appliquant des algorithmes d'analyse
- Evaluer et valider les connaissances ainsi extraites des analyses

Bien sûr ce processus n'est pas linéaire, il se veut cyclique, le Datamining participant dès lors à une rationalisation avancée du stockage de l'information et des données. Autrement dit, le Datamining ne consiste pas en une succession d'études ad hoc mais doit bien permettre de capitaliser des connaissances acquises sous forme de connaissances explicites, donc de structurer les contenus nécessaires à l'ingénierie des connaissances.

Le Datamining se propose Alors d'utiliser un ensemble d'algorithmes issus de disciplines scientifiques diverses (statistiques, intelligence artificielle, base de données) pour construire des modèles à partir des données, c'est-à-dire trouver des schémas « intéressants » (des 'patterns' ou motifs de conception en français) selon des critères fixés au départ, et extraire de ces données un maximum de connaissances utiles.

I. 4. Type de techniques de datamining

Résoudre une problématique avec un processus de Datamining impose généralement l'utilisation d'un grand nombre de méthodes et algorithmes différents. On peut distinguer deux grandes familles [2]:

I. 4.1. Techniques descriptives

Elles permettent de travailler sur un ensemble de données dans lequel aucune des données n'a d'importance particulière par rapport aux autres, c'est-à-dire un ensemble de données dans lequel aucune variable n'est considérée comme la cible, elle est utilisée pour dégager d'un ensemble d'individus des groupes homogènes, afin de réaliser une compression d'informations (fig.I-1).

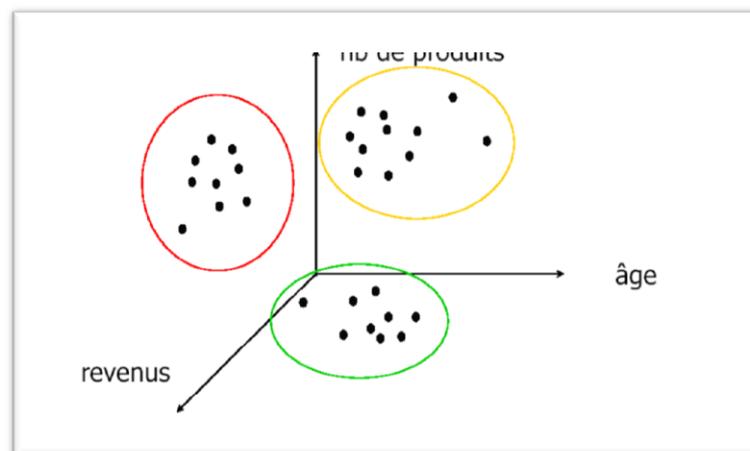


Fig. I-1- Exemple de technique non-supervisé (Classification)

On peut citer quelques techniques disponibles pour ces méthodes:

- L'Analyse en Composantes principales ;
- Techniques utilisées classiquement dans le monde des statistiques : classification ascendante hiérarchique, k-means et les nuées dynamiques (Recherche des plus proches voisins), les classifications mixtes (Birch...), les classifications relationnelles...

I. 4.2. Techniques prédictives

Ces méthodes servent à expliquer et/ou de prévoir un ou plusieurs phénomènes observables et effectivement mesurés. Concrètement, elles vont s'intéresser à une ou plusieurs variables de la base de données définies comme étant les cibles de l'analyse. Dont voici la principale technique existante :

- Arbres de décision (Arbre de décision).
- Réseau de neurones ;
- Régressions linéaires et non linéaires au sens large :(Régression logistique binaire ou multinomiale, Analyse discriminante linéaire ou quadratique, modèle linéaire généralisé, régression PLS, régressions non paramétriques, équations structurelles..).
- Les algorithmes génétiques.
- Inférence bayésienne (Réseau bayésien).

I.5. Conclusion

Le datamining est une discipline née en dehors de la statistique, dans la communauté des bases de données et de l'intelligence artificielle dans le but de valoriser les bases de données, il offre des perspectives nouvelles et répond au défi du traitement des giga bases de données, il cherche à découvrir des structures inconnus et utiles de modèle issus de l'exploitation des données et non d'une théorie, à l'aide de processus de recherche algorithmique, linéaire ou non, explicite ou implicite (réseaux de neurones, Algorithmes génétiques, arbres de décision, réseaux Bayesiens,...) [3].

I.2 .Méthode d'Analyse en Composantes Principales (ACP)

I.2.1. Introduction

Les statistiques peuvent être vues en fonction de l'objectif fixé ; classiquement les méthodes statistiques sont employées soit pour explorer les données (nommée statistique exploratoire) soit pour prédire un comportement. La statistique exploratoire s'appuie sur des techniques descriptives et graphiques. Elle est généralement décrite par la statistique descriptive qui regroupe des méthodes exploratoires simples, uni- ou bidimensionnelle (moyenne, moments, quantiles, variance, corrélation, ...) et la statistique exploratoire multidimensionnelle. L'analyse de données s'inscrit dans ce cadre de la statistique exploratoire multidimensionnelle.

L'Analyse en Composante Principale (ACP) en anglais (*Principal Component Analysis PCA*) est une **technique descriptive** son utilisation remonte au début du siècle dernier pour l'exploitation des données ; Permettant d'étudier les relations qui existent entre les variables, sans tenir compte, à priori, d'une quelconque structure [4].

Le but de l'ACP est d'identifier la structure de dépendance entre des observations multi variables afin d'obtenir une description ou une représentation compacte de ces derniers (variables). Son utilisation a été restreinte à une projection des données sur différents axes factoriels.

L'ACP est une technique de projection orthogonale linéaire qui projette les observations multidimensionnelles représentées dans un sous-espace de dimension m (m étant le nombre de variables observées) dans un sous-espace de dimension inférieur ($l < m$) en maximisant la variance des projections.

L'ACP peut être considéré comme une technique de minimisation de l'erreur quadratique d'estimation ou une technique de maximisation de la variance des projections (ces deux critères sont équivalents).

L'ACP est utilisé comme un outil de modélisation des relations entre les différentes grandeurs représentant le comportement d'un processus quelconque.

L'estimation des paramètres du modèle *ACP* est effectuée par calcul des **valeurs** et **vecteurs** propres de la matrice de corrélation des données, cependant, pour la détermination de la structure du modèle, il faut déterminer le nombre de composante à retenir dans ce modèle, pour cette raison plusieurs critères de sélection du nombre de composante, dans notre travail le critère que nous adoptons est basé sur le principe de la variance de reconstitution, pour la sélection du nombre de composante à retenir dans le modèle *ACP* car il exploite les redondances entre les variables. Ce nombre permet d'identifier le modèle *ACP*.

Historique

La représentation brute de l'information conduit à des vecteurs de caractéristiques de grandes dimensions, ce qui peut poser des problèmes de complexité de calcul et de capacité de stockage. Ainsi, il est parfois préférable de chercher à réduire la dimensionnalité d'un problème de classification pour améliorer ses performances. Le mathématicien Bellman, père de la programmation dynamique, a introduit l'expression «malédiction de la dimensionnalité» (curse of dimensionality), pour signifier que représenter les formes par des vecteurs de taille importante est source de problème ^[5].

Il est évident que l'on se trouve d'emblée confronté à un problème de grande dimension. Plusieurs solutions sont préconisées dans la littérature pour réduire la taille de sa dimension Minami 1999[6]. Dans ce travail, nous avons utilisé l'analyse en composantes principales. L'analyse en composantes principales (ACP) est une technique descriptive permettant d'étudier les relations qui existent entre les variables,

sans tenir compte, a priori, d'une quelconque structure [7],[8]. Son utilisation pour l'exploitation des données remonte au début du siècle dernier. Elle est principalement issue des travaux de psychologues américains [9], [10], [11]. Le but de l'ACP est d'identifier la structure de dépendance entre des observations multi variables, afin d'obtenir une description ou une représentation compacte de ces dernières. L'idée de base de l'ACP est de réduire la dimension de la matrice des données, en retenant le plus possible les variations présentes dans le jeu des données de départ. Cette réduction ne sera possible que si les variables initiales ne sont pas indépendantes et ont des coefficients de corrélation entre elles non nuls. Ces variables initiales sont transformées en de nouvelles variables, appelées composantes principales. Elles sont obtenues par combinaisons linéaires des précédentes et sont ordonnées et non corrélées entre elles. L'analyse en composantes principales cherche à identifier les vecteurs propres ortho normaux et leurs valeurs propres correspondantes de la matrice de dispersion des variables originelles.

Les vecteurs propres ortho normaux sont utilisés pour construire les composantes principales et les valeurs propres sont les variances des composantes principales correspondantes [7].

I.2.2. Principe de l'Analyse en Composantes Principales

Dans le cadre de mémoire, nous aborderons l'ACP comme une technique de réduction

T de description des échantillons. La théorie sous-jacente à l'analyse en composantes principales est vaste, nous ne passerons donc en revue que les points les plus importants. Du point de vue géométrique l'ACP consiste à effectuer une certaine rotation du repère des variables autour de leurs valeurs moyennes.

Cette rotation transforme les n variables corrélées en l variables non corrélées. Notons que ce sont justement ces variables transformées que l'on a nommées les composantes principales. Supposons qu'on a un vecteur de données aléatoires $X_K = [X_1, \dots, X_m]^t \in \mathbb{R}^m$ de moyenne nulle $\varepsilon\{x(k)\} = 0$ et de matrice de covariance ou d'auto-corrélation :

$$\Sigma = \varepsilon\{xx^t\} \in \mathbb{R}^{m \times m} \quad (\text{I.1})$$

‘t Avec $\varepsilon = \frac{1}{m-1}$

En analyse en composantes principales, un vecteur caractéristique $t \in \mathcal{R}^l$ est associé à chaque vecteur de données dont il optimise la représentation au sens de la minimisation de l'erreur d'estimation de \mathbf{x} ou la maximisation de la variance de \mathbf{t} . Les vecteurs \mathbf{t} et \mathbf{x} sont liés par une transformation linéaire $\mathbf{t} = \mathbf{P}'\mathbf{x}$,

Où la matrice de transformation $\mathbf{P} \in \mathcal{R}^{m \times l}$ vérifie la condition d'orthogonalité $\mathbf{P}'\mathbf{P} = \mathbf{I}_l$. Les colonnes de la matrice \mathbf{P} forment les vecteurs de base orthonormés d'un sous-espace \mathcal{R}^l de représentation réduite des données. La transformation linéaire s'apparente ainsi à une projection de l'espace des données de dimension m vers un sous-espace orthogonal de dimension l .

Les composantes t_j , avec $j = 1, \dots, l$, du vecteur caractéristique \mathbf{t} représentent les composantes projetées du vecteur de données \mathbf{x} dans ce sous-espace.

Au sens de l'ACP, la projection \mathbf{P} est optimale si l'erreur quadratique d'estimation des vecteurs de données \mathbf{x} est minimale. Ce problème d'optimisation s'exprime par :

$$P_{opt} = \underset{P}{\operatorname{argmin}} J_e(P) \quad (\text{I.2})$$

Où J_e représente le critère d'erreur d'estimation de l'ACP. Sous la contrainte d'orthogonalité de la matrice de projection $\mathbf{P}'\mathbf{P} = \mathbf{I}_l$, ce critère aura la forme :

$$\begin{aligned} J_e(\mathbf{P}) &= \mathcal{E}\{\|\mathbf{X} - \widehat{\mathbf{X}}\|^2\} \\ (\text{I.3}) \quad &= \mathcal{E}\{\|\mathbf{X} - \mathbf{P}\mathbf{P}'\mathbf{X}\|^2\} \\ &= \mathcal{E}\{(\mathbf{X} - \mathbf{P}\mathbf{t})'(\mathbf{X} - \mathbf{P}\mathbf{t})\} \\ &= \mathcal{E}\{\mathbf{X}'\mathbf{X} - 2\mathbf{t}'\mathbf{t} + \mathbf{t}'\mathbf{P}\mathbf{P}'\mathbf{t}\} \\ &= \mathcal{E}\{\operatorname{trace}(\mathbf{X}\mathbf{X}') - \mathbf{t}'\mathbf{t}\} \\ &= \operatorname{trace}(\boldsymbol{\Sigma}) - \mathcal{E}\{\mathbf{t}'\mathbf{t}\} \end{aligned}$$

Notons que la trace d'une matrice carrée est définie par la somme de ces éléments diagonaux. Du fait que la matrice de covariance Σ est indépendante de la matrice des paramètres \mathbf{P} , minimiser J_e revient à maximiser le second terme J_v de son expression :

$$J_v(\mathbf{P}) = \varepsilon\{\mathbf{t}^t \mathbf{t}\} = \sum_{j=1} \varepsilon\{t_j^2\} \quad (\text{I.4})$$

Ainsi, la minimisation de l'erreur quadratique d'estimation de \mathbf{x} est équivalente à la maximisation de la variance des projections \mathbf{t}_j des données. En conséquence :

$$\mathbf{P}_{opt} = \text{argmin}_{\mathbf{P}} J_e(\mathbf{P}) = \text{argmax}_{\mathbf{P}} J_v(\mathbf{P}) \quad (\text{I.5})$$

Le problème de l'ACP, considéré sous l'angle de la maximisation de la variance de projection des données, est celui de la détermination des vecteurs propres de la matrice de covariance Σ .

I.2.3. Identification du modèle ACP

Soit $\mathbf{x}_i = [\mathbf{x}_{i1}, \dots, \mathbf{x}_{in}]^t$ un vecteur d'observations de n variables aléatoires.

On notera $\bar{\mathbf{X}}$ le vecteur correspondant à la moyenne des observations de \mathbf{x} et Σ la matrice de covariance de ces m observations. L'estimation des paramètres du modèle ACP se résume en une estimation des valeurs et vecteurs propres de la matrice de covariance Σ . Puisque la matrice de covariance Σ est carrée de dimension $n \times n$ et symétrique, on peut, selon l'algèbre linéaire, vérifier une relation de type :

$$\mathbf{P}^t \Sigma \mathbf{P} = \mathbf{L} \quad (\text{I.6})$$

Où \mathbf{L} est une matrice diagonale et \mathbf{P} est une matrice ortho normale. Les éléments diagonaux de $\mathbf{L} = \{l_1, \dots, l_n\}$ sont les valeurs propres de la matrice de covariance Σ alors que les vecteurs colonnes de \mathbf{P} sont les vecteurs propres de Σ .

Les valeurs propres l_i peuvent être obtenues par la solution de l'équation caractéristique

$$\det[\Sigma - l_i \mathbf{I}] = 0 \quad (\text{I.7})$$

Pour les n valeurs propres. Connaissant les valeurs propres l_i , on peut déduire les vecteurs propres de la matrice de covariance par la solution de

$$(\Sigma - l_i I)A_i = 0 \quad (\text{I.8})$$

Et puis,

$$u_i = \frac{A_i}{\sqrt{A_i^t A_i}} \quad (\text{I.9})$$

Pour les n vecteurs propres. Notons que l'équation (I.9) est aussi appelée équation de normalisation. Il est intéressant de constater qu'ici les éléments des vecteurs propres u_i sont les cosinus directeurs des axes générés par une rotation dans l'espace des observations.

Nous pouvons donc calculer tout un ensemble de valeurs à partir de ces nouveaux axes. En d'autres termes, on peut transformer les variables $\mathbf{x} = [x_1, \dots, x_n]^t$ en un nouvel ensemble de variables $\mathbf{y} = [y_1, \dots, y_n]^t$ par l'application de

$$\mathbf{y}_i = \mathbf{u}_i^t (\mathbf{X} - \bar{\mathbf{X}}) \quad (\text{I.10})$$

Y_i est la $i^{\text{ème}}$ composante principale de l'ensemble des observations. Les nouvelles variables y_i ont évidemment une moyenne nulle et elles possèdent une variance donnée par l_i (valeur propre de Σ). Les vecteurs de colonnes \mathbf{u}_i de \mathbf{P} sont ortho normaux car ils sont orthogonaux et de modules unitaires. Donc, si l'on utilise ces vecteurs pour obtenir les composantes principales, tel que montré par l'équation (II.10), on obtiendra des composantes principales qui ne seront pas corrélées et de variances égales aux valeurs propres de la matrice de covariance Σ .

En appliquant une transformation sur les vecteurs colonnes \mathbf{u}_i , on peut obtenir de nouvelles propriétés intéressantes. L'une de ces transformations consiste à effectuer une mise à l'échelle de \mathbf{u}_i par les valeurs propres de la matrice de covariance Σ :

$$\mathbf{V}_i = \mathbf{u}_i \sqrt{l_i} \quad (\text{I.11})$$

Les composantes obtenues par l'utilisation de $\mathbf{V} = [v_1, \dots, v_n]^t$ seront de même unité que les variables originales. Nous retrouvons également une transformation dans

laquelle les vecteurs u_i sont divisés par les valeurs propres de la matrice de covariance S :

$$W_i = \frac{u_i}{\sqrt{l_i}} \quad (\text{I.12})$$

Dans ce cas, les composantes obtenues seront de variance unitaire. Enfin, nous soulignons que l'on peut avoir une mesure de variabilité expliquée par chacune des composantes principales en utilisant les vecteurs transformés de l'équation (III.11) selon

$$var = v_i v_i^t \text{ avec } i=1, \dots, n. \quad (\text{I.13})$$

Les valeurs obtenues par l'équation (I.13) peuvent être utilisées pour indiquer le pourcentage de la variabilité des caractéristiques représentées par chacune des composantes principales.

I.2.4. Détermination de la structure du modèle ACP

L'analyse en composantes principales recherche une approximation de la matrice initiale des données X par une matrice de rang inférieur issue d'une décomposition en valeurs singulières.

La question qui se pose alors, et qui a été largement débattue dans la littérature, concerne le choix du nombre de composantes principales qui doit être retenu. De nombreuses règles sont proposées pour déterminer le nombre de composantes à retenir [8, 9, 10,11]. La plupart de ces règles sont heuristiques et donnent un nombre de composants subjectif. Toutefois, dans le cadre de l'application de l'ACP à la réduction des données, le nombre de composantes a un impact significatif sur la phase finale de classification. Si peu de composantes sont utilisées, on risque de perdre des informations contenues dans les données de départ en projetant certains échantillons dans le sous-espace des résidus et donc avoir des erreurs de modélisation, ce qui provoque une mauvaise caractérisation.

Si par contre beaucoup de composantes sont utilisées, il y a risque d'avoir des composantes retenues (les composantes correspondantes aux valeurs propres les plus faibles parmi celles retenues dans le modèle) qui sont porteuses de redondance, ce qui est indésirable.

I.2.4.1. Pourcentage cumulé de la variance totale (PCV)

L'idée de base de cette approche est que chaque composante principale est représentative d'une portion de la variance des mesures de l'information étudiée. Les valeurs propres de la matrice de corrélation sont des mesures de cette variance et peuvent donc être utilisées dans la sélection du nombre de composantes principales. Pour faire ce choix, il faut estimer le pourcentage de la variance totale qu'on veut conserver. Le nombre de composantes est alors le plus petit nombre pris de telle sorte que ce pourcentage soit atteint ou dépassé; les composantes sont choisies successivement dans l'ordre des variances décroissantes. Le pourcentage de variance expliquée par les premières composantes est donné par :

$$PCV(l) = 100 \left(\frac{\sum_{j=1}^l \lambda_j}{\sum_{j=1}^m \lambda_j} \right) \% \quad (\text{I.14})$$

La variance du bruit étant inconnue a priori, la décision basée seulement sur le pourcentage de la variance expliquée est un peu arbitraire. Sa capacité à fournir le nombre correct de composantes principales dépendra fortement du rapport signal sur bruit.

I.2.4.2. Moyenne des valeurs propres

Cette règle consiste à ne prendre en considération que les composantes pour lesquelles la valeur propre est supérieure à la moyenne arithmétique de toutes les valeurs propres. En particulier, si on travaille sur les données centrées réduites, cela revient à négliger les composantes dont la variance est inférieure à l'unité

$$\left(\frac{1}{m} \text{trace}(\Sigma) > 1 \right) \quad (\text{I.15})$$

Dans le cas du modèle de calcul à partir de la matrice de covariance, la moyenne arithmétique des valeurs propres est donnée par $\frac{1}{m} \text{trace}(\Sigma)$

I.2.5. Conclusion

Ce que nous pouvons retenir de cette méthode est que l'ACP est une technique utile pour la compression et la classification des données. Le problème consiste à réduire la dimensionnalité d'un ensemble des données (échantillon) en trouvant un nouvel ensemble de variables plus petit que l'ensemble originel des variables, qui néanmoins contient la plupart de l'information sur l'échantillon. Par information nous voulons parler de la variation présente dans l'échantillon et donnée par les corrélations entre les variables originelles. Les nouvelles variables, appelées composantes principales (PC), sont non corrélées, et sont ordonnées par fraction de l'information totale que chacune contient [7,8].

En termes précis, l'usage principal de l'ACP est la réduction de la dimensionnalité tout en conservant le plus d'information possible.

1^{ière} PC: contient la plus grande quantité de variation.

2^{nde} PC: contient en deuxième lieu la plus grande quantité de variation.

Quasiment toutes les technique de datamining restent en suspens et ces méthodes ont encore beaucoup de progrès à faire.

Cependant, il est clair à la lumière des résultats incontestablement utiles de l'utilisation de ces techniques dans divers domaines de la recherche, qu'elles offrent un moyen puissant pour l'analyse dont on ne doit pas manquer l'utilisation en science des matériaux pour profiter de leurs prérogatives tout en demeurant prudents quant aux interprétations des résultats.

Références

- [1] D. T. Larose ET J. Wiley, *Data mining methods and models*. Wiley Online Library, 2006.
- [2] S. Tufféry, « DATA MINING-SCORING », *Bases de données et gestion de la relation client*, 2002.
- [3] G. Saporta, « Liaisons entre plusieurs ensembles de variables et codage de données qualitatives », 1975.
- [4] M. F. Haraket, *Détection et Localisation de Défauts par Analyse en Composant Principales*. Doctorat de l'Institut National Polytechnique de Lorraine, 2003.
- [5] R. E. Bellman, R. E. Bellman, R. E. Bellman, et R. E. Bellman, *Adaptive control processes: A guided tour*. Princeton University Press, 1966.
- [6] K. Minami, H. Nakajima, et T. Toyoshima, « Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network », *Biomedical Engineering, IEEE Transactions on*, vol. 46, n^o. 2, p. 179–185, 1999.
- [7] I. T. Jolliffe et MyiLibrary, *Principal component analysis*, vol. 2. Wiley Online Library, 2002.
- [8] K. I. Diamantaras et S. Y. Kung, *Principal component neural networks: theory and applications*. Wiley, 1996.
- [9] K. P. FRS, « LIII. On lines and planes of closest fit to systems of points in space », *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, n^o. 11, p. 559–572, 1901.
- [10] C. Spearman, « “ General Intelligence,” Objectively Determined and Measured », *The American Journal of Psychology*, vol. 15, n^o. 2, p. 201–292, 1904.
- [11] H. Hotelling, « Analysis of a complex of statistical variables into principal components. », *Journal of educational psychology*, vol. 24, n^o. 6, p. 417, 1933.

Généralités sur les liquides ioniques

II.1. Introduction

Les solvants peuvent être classés en fonction du type de liaisons chimiques qu'ils mettent en jeu

- ✓ les liquides moléculaires dans lesquels interviennent des liaisons covalents
- ✓ 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.

II.2. Définition

Les liquides ioniques sont des sels fondus se différenciant de l'ensemble des sels fondus par une température de fusion inférieure à 100°C. [Fig. II-1]

Les liquides ioniques sont constitués d'un cation le plus souvent organique, associé à un anion organique ou inorganique

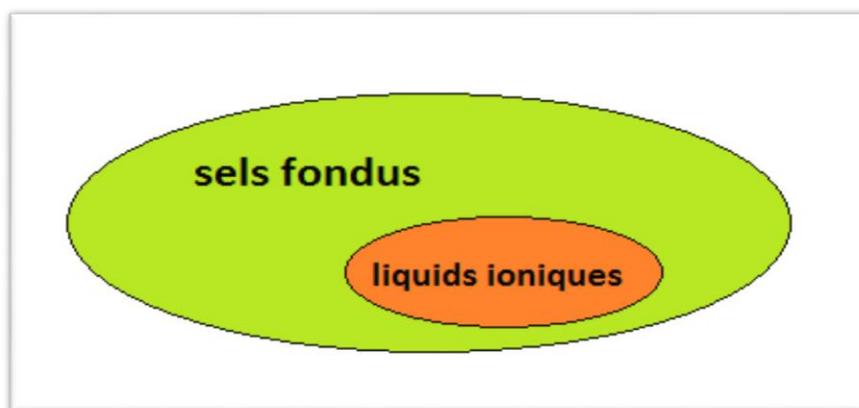


Fig. II-1- les liquides ioniques sont un sous ensemble des sels fondus

II.3. HISTORIQUE

La description du premier liquide ionique date du milieu du 19^{ème} siècle lors de la réaction de Friedel Crafts entre le benzène et le chlorométhane catalysée par un acide de Lewis, $AlCl_3$ [1], une seconde phase apparaît sous la forme d'une « huile rouge ». La structure de ce composé sera identifiée plus tard par le Pr. Jerry Atwood [2] à l'université du Missouri grâce à l'apparition de la RMN et correspond au complexe intermédiaire, jusqu'alors présumé, de la réaction de Friedel Crafts : le sel d'heptadichloroaluminate.

Des sels liquides de nitrate d'alkyl-ammoniums furent ensuite découverts au début du 20^{ème} siècle en particulier le nitrate d'éthylammonium (Walden, 1914)[3] dont le point de fusion est de $12^\circ C$. Et ce sel liquide n'avait pas attiré l'attention des chimistes. Ce n'est qu'après la deuxième guerre mondiale qu'une application de ces sels a été rapportée.

Harly [4] a aussi récrit l'utilisation de chloroaluminates de pyridinium comme électrolytes dans l'électrodéposition de l'aluminium. Ces travaux en ensuite été repris durant les années quatre-vingt par Hussey[5] et Osteryong[6].

Les liquides ioniques se sont progressivement immiscés dans la chimie organique (classique) notamment grâce aux sels de dialkylimidazolium associé à leur anion chloroaluminates qui sont utilisés comme catalyseurs dans les réactions de Friedel-Crafts.

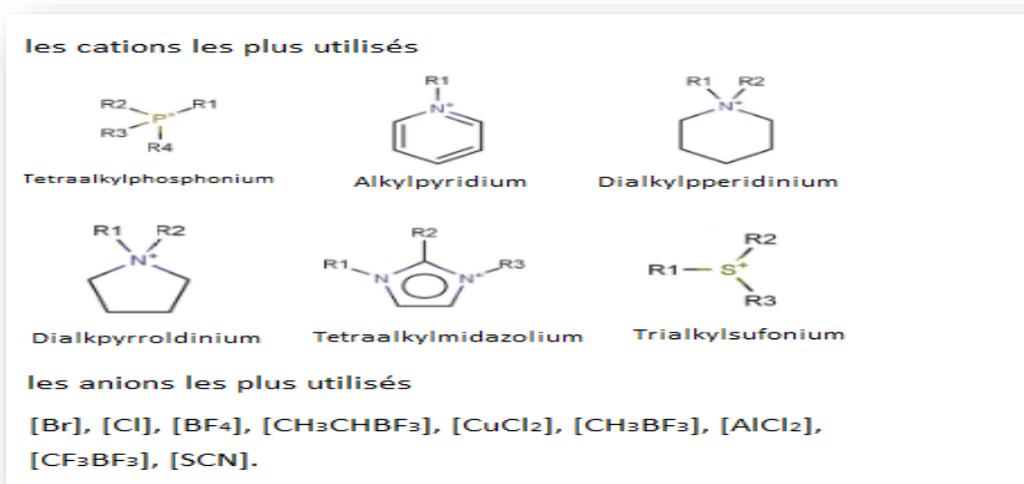


Fig.II-2 - Exemples des cations et des anions constituant les liquides ioniques

II.4. Principales propriétés physico-chimiques

La connaissance des propriétés physico-chimie des liquides ioniques est importante pour évaluer et sélectionner les liquides ioniques pour chaque application aussi bien que la conception de processus. Ils sont récemment devenus très populaire en tant que dissolvants potentiels pour des applications industrielles dans beaucoup différentes disciplines de la science et d'environnement.

La recherche, menée sur les propriétés des liquides ioniques purs, à porte dans un premier temps sur la compréhension de la relation entre la structure du cation et de l'anion, et leurs propriétés physico-chimiques. Très rapidement, il est apparu que la pureté des liquides ioniques était un paramètre influençant de façon très importante les mesures de leurs propriétés physico-chimiques. La présence de contaminants tels que l'eau ou/et les ions halogénure, a un effet significatif sur les propriétés physico-chimiques telles que la densité, la viscosité, le point de fusion et la conductivité [25,18].

Malheureusement, les données concernant la pureté des liquides ioniques ne sont pas systématiquement mentionnées dans la littérature car les méthodes de détermination des impuretés sont encore en cours de recherche. Ce manque de données peut, en grande partie expliquer la grande variabilité des valeurs des propriétés physico-chimiques des liquides ioniques purs issues de la littérature [3].

II.4.1. Transition vitreuse

Tous les liquides ioniques ne cristallisent pas et nombreux sont ceux qui ne présentent qu'une phase vitreuse, même en descendant à des températures très basses. Ces matériaux sont de plus très faciles à sur fondre, sur de grandes plages de températures pouvant aller jusqu'à 200 °C [56], même pour de grands volumes [55], [23], [56], [35]. À l'état solide, certains présentent plusieurs transitions de phase [49] dont quelquefois des états cristallins, voire plusieurs points de fusion [57].

Dans la série C_nC_1 -im ou pyridinium (contre-ions BF_4^- , Cl^- , Br^- , Tf_2N^- et $CF_3SO_3^-$ selon les études), les liquides présentent une phase cristal liquide au-dessus de $n = 12$ [17], [58],[44], [59]. La gamme de température sur laquelle un liquide ionique est également un cristal liquide augmente avec la longueur de la chaîne alkyle [16] et la stabilisation de cette phase se fait selon l'ordre [58] :

$\text{Cl}^- > \text{Br}^- > \text{BF}_4^- > \text{CF}_3\text{SO}_3^- > \text{Tf}_2\text{N}^-$

Une autre étude a mis en évidence l'ordre [59] :

$\text{Cl}^- > \text{Br}^- > \text{PF}_6^-$

II.4.2. Température de fusion

Le critère clé pour l'évaluation d'un liquide ionique est, par définition, son point de fusion. Un sel fondu est défini liquide ionique lorsque son point de fusion est inférieur à 100°C.

Le point de fusion est difficile à corréler avec la composition chimique. Les principaux facteurs qui influencent le point de fusion des liquides ioniques sont : la distribution de charge sur les ions, la possibilité de liaisons hydrogène, la symétrie des ions et les interactions de Van der Waals. Un succès modeste a été obtenu en utilisant des descripteurs moléculaires modélisés par ordinateur pour prévoir le point de fusion des bromures d'imidazolium [60].

La **température de fusion** est gouvernée par la structure et les interactions entre ions. Le volumineux cation organique des liquides ioniques engendre peu d'interactions, d'où un point de fusion bas. Si la longueur ou le volume de la chaîne alkyle greffée sur le squelette imidazolium augmente, ce phénomène est amplifié et la température de fusion diminue.

Cependant, ceci n'est pas une règle générale et, pour certains composés à base polycyclique, la température de fusion augmente avec la longueur des chaînes alkyle greffées [61] tandis que l'on note un minimum pour des tétraalkylammonium aliphatiques symétriques ou non, en fonction de la longueur de la chaîne [62], [63]. Étude a conclu que la raison principale des bas points de fusion des liquides ioniques était leur plus ou moins grande incapacité à s'ordonner en un réseau compact [27]. D'autres études ont mis en exergue l'effet de l'asymétrie du cation, celle-ci abaissant la température de fusion pour des cations imidazolium ou ammonium [56], [63] bien que ceci ne soit pas, là encore, une règle générale : certains liquides ioniques avec un cation symétrique à base imidazolium présentent des températures de fusion relativement basses, pouvant aller jusqu'à -27°C pour une chaîne en C₁₀ [64]. Enfin, l'alkylation de la position 2 du cycle imidazolium conduit à une augmentation de la température de fusion [56], [65]. Inversement, à cation constant, l'anion Tf₂N⁻ abaisse fortement la température de fusion. Ceci n'est pas toujours suffisant pour obtenir des liquides ioniques (selon notre terminologie) dans le cas de tétraalkylammoniums symétriques ou non [63][66], [67]

II.4.3. Stabilité thermique

Un grand nombre de ces sels, ayant un anion faiblement nucléophile, montre une stabilité thermique exceptionnelle permettant leur utilisation pour des applications à des

températures supérieures à 250 °C et dans certains cas supérieures à 400 °C dans l'air ou dans une atmosphère inerte [29].

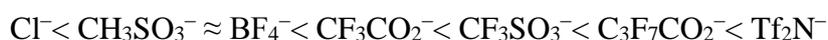
La température maximale d'emploies liquides ioniques est donnée par la température de décomposition et non d'ébullition comme c'est généralement le cas pour les autres liquides. À squelette organique constant, c'est essentiellement la nature du contre-anion qui détermine cette décomposition. Ainsi, les anions engendrant les plus faibles interactions intermoléculaires induisent les températures de décomposition les plus élevées, avec l'ordre [65] :



En revanche, des substituant alkyles ramifiés sur le cation abaissent la température de décomposition [65]. Celle-ci, qui a souvent lieu vers 350 à 400 °C, est due à l'élimination du substituant alkyle sur la position N du squelette imidazolium. Ces deux phénomènes (bas point de solidification et haut point de décomposition) expliquent la grande gamme de température accessible.

II.4.4. Densité

La densité de certains liquides ioniques peut être très proche de celle de l'eau ($d = 1,08$ pour $\text{C}_8\text{C}_1\text{-imBF}_4$ à 25 °C [70], voire 1,02 à 25 °C pour certains LI très spécifiques [87]). Cependant, la plupart des liquides ioniques ont une densité située dans la gamme 1,20 à 1,50 (à 25 °C) (voir tableau 3). Pour un anion fixé, l'augmentation de la chaîne alkyle du cation diminue la densité [27], [63], [72] et, pour un cation donné, une synthèse de différentes mesures permet de montrer que l'augmentation de la densité suit l'ordre [64], [65], [74],[67], [75], [76], [77], [78] :



La densité à température ambiante des liquides ioniques de la classe 1,3-dialkylimidazolium est en général plus grande que celle de l'eau (0,9 à 1,6 g.cm⁻³). Aussi bien pour les liquides ioniques hydrophiles qu'hydrophobes, elle diminue presque linéairement avec la longueur de la chaîne alkyle du cation [49]. De plus, il est à noter que l'augmentation de la teneur en eau provoque une diminution de la densité [40].

II.4.5. Viscosité

L'un des inconvénients majeurs des liquides ioniques est leur viscosité élevée, environ de deux ordres de grandeur supérieure à celle de l'eau (0,89 cP (centipoise) à $T = 25\text{ °C}$) ou des solvants organique classiques (acétonitrile : 0,34 cP à $T = 25\text{ °C}$). Parmi les plus faible viscosités recensées, on trouve $C_2C_1\text{-imTf}_2\text{N}$ ($\eta = 25\text{ cP}$ à $T = 25\text{ °C}$ [72]) et les viscosités les plus fortes rapportées à ce jour, toujours 25 °C , sont de l'ordre de 800 cP pour $C_9C_1\text{-imPF}_6$ [72] et 1 110 cP pour $C_4C_1\text{-imI}$ [77]. Comme tous les paramètres des liquides ioniques, la viscosité est très sensible à la quantité d'eau présente [77] et à la présence d'ions chlorures [77].

La viscosité des liquides ioniques est déterminée essentiellement par leur tendance à former des liaisons hydrogènes et par la force des interactions de Van der Waals [79,23]. La délocalisation de la charge sur l'anion semble favoriser une viscosité faible par l'affaiblissement de la liaison hydrogène avec le cation (exemple de l'anion NTf_2^-) [80,81]. La longueur de la chaîne alkyle portée par le cation est un paramètre influent sur la viscosité.

À anion constant, la viscosité augmente si la chaîne latérale sur le cation augmente et les liquides ioniques basés sur PF_6^- semblent toujours plus visqueux que ceux basés sur Tf_2N^- [59]. Dans le cas de BF_4^- les teneurs en eau dans les échantillons mesurés sont extrêmement variables mais le liquide ionique composé de ce contre-ion se situe vraisemblablement dans une gamme de viscosité légèrement inférieure à celle du liquide comportant PF_6^- (pour un même cation) [82], [23], [72], [35], [24], [71], [83].

À cation constant, la viscosité des dialkylimidazolium augmente dans le sens :
 $\text{Tf}_2\text{N}^- < \text{CF}_3\text{CO}_2^- \approx \text{CF}_3\text{SO}_3^- < \text{CH}_3\text{CO}_2^- \approx \text{CH}_3\text{SO}_3^- < \text{PF}_6^- < \text{Cl}^-$

II.4.6. Polarité

Les liquides ioniques sont des sels organiques à caractère très polaire. De nombreuses études sur la polarité des liquides ioniques ont été effectuées. A titre d'exemple, le $[\text{bmim}][\text{PF}_6]$ possède une polarité voisine du 1-butanol [84], et celle du $[\text{bmim}][\text{BF}_4]$ est proche du méthanol .

Leurs propriétés de solvations sont à relier à la propension qu'ont ces sels à agir en tant que donneur et accepteur de liaison hydrogène. En faisant varier la longueur des chaînes des groupements alkyles, on peut augmenter la lipophilie de ses liquides ioniques et ainsi moduler leurs propriétés de ses composés.

II.4.7. Propriétés électrochimiques

À l'origine, le développement des liquides ioniques basés sur des chloroaluminates est intimement lié à l'amélioration des propriétés électrochimiques des sels fondus, notamment dans le cadre d'applications pour les batteries [4], [75]. Les liquides ioniques possèdent en effet des propriétés qui les rendent très intéressants comme nouveaux électrolytes pour les procédés électrochimiques.

Ils possèdent un domaine de stabilité électrochimique élevé, compris suivant les liquides ioniques entre 3 V et 5,4 V et atteignant jusqu'à 6 V pour des tétraalkylammonium, qui sont de ce point de vue bien supérieurs aux imidazolium [63]. La conductivité d'un solvant est un critère fondamental pour les applications électrochimiques. Les liquides ioniques présentent une grande conductivité ionique, généralement de l'ordre de $10^{-1} \cdot S \cdot m^{-1}$ [77]. Bonhote et coll. rapportent la relation entre la conductivité et différentes propriétés [23] :

$$\sigma = \frac{yF^2d}{(6\pi N_A M_w \eta)[(\zeta_a r_c)^{-1} + (\zeta_c r_a)^{-1}]}$$

Notation : viscosité (η), masse molaire (M_w), nombre d'Avogadro (N_A), nombre de faraday (F), densité (d), degré de dissociation ($0 < y < 1$), rayon de l'anion et du cation (r_a , r_c respectivement), facteur de micro viscosité corrigée de l'anion et du cation (ζ_a , ζ_c respectivement).

Il apparait que la viscosité n'est pas le seul paramètre ayant une influence sur la valeur de la conductivité : il faut aussi prendre en considération la taille et la masse moléculaire des ions, qui ont également un effet important.

Ce phénomène a souvent été évoqué, mais n'a jamais été confirmé [80].

II.4.7.1. la fenêtre électrochimique

Un critère clé de sélection d'un solvant pour les études électrochimiques est la stabilité électrochimique du solvant [44,45].

Ceci est plus clairement manifesté par la gamme de tensions sur lequel le solvant est électrochimiquement inerte.

Ce potentiel électrochimique fenêtre utile dépend de la stabilité d'oxydation et de réduction du solvant

Dans le cas des liquides ioniques, la fenêtre de potentiel dépend principalement de la résistance du cation à la réduction et la résistance de l'anion à l'oxydation (une exception

Notable à cette règle est dans les liquides ioniques acides chloroaluminate ou la réduction des espèces heptachloroaluminate, $[Al_2Cl_7]^-$, est le processus limitant cathodique).

En outre, la présence des impuretés peut jouer un rôle important dans la limitation de la fenêtre de potentiel des liquides ioniques.

II.4.7.2. La Conductivité

La conductivité d'un électrolyte est une mesure du nombre de porteurs de charges (disponibles) et de leurs mobilités. Les liquides ioniques possèdent une forte concentration en ions ce qui suggère une haute conductivité. En réalité, ils sont significativement moins conducteurs que les électrolytes aqueux concentrés. La haute viscosité des liquides ioniques a un impact majeur sur la conductivité (inversement proportionnelle). Bien qu'il existe une corrélation forte, la viscosité seule ne peut pas expliquer l'ensemble des disparités des valeurs de conductivité.

II.5.Applications des liquides ioniques

Les liquides ioniques ont été développés il y a plus d'une vingtaine d'années dans le domaine de l'électrochimie pour la recherche de nouveaux systèmes d'énergie. Ces nouveaux milieux ont ensuite connu un grand intérêt dans les domaines de la synthèse organique et de la catalyse.

II.5.1.Applications en électrochimie

Les propriétés des liquides ioniques telles que leur très large domaine d'électroactivité, leur forte conductivité, et leur grande stabilité thermique ont fait de ces nouveaux milieux des candidats de choix dans la recherche de nouveaux systèmes d'énergie (cellule photovoltaïque, batterie...).[40,85]

II.5.2.Applications en synthèse organique et en catalyse

La synthèse organique et la catalyse sont certainement les deux domaines en expansion dans l'utilisation des liquides ioniques. Il existe de nombreuses applications des liquides ioniques dans ces domaines [40,86]. D'un point de vue chimique, le principal potentiel des liquides ioniques est d'augmenter le rendement et la cinétique de la réaction et d'améliorer la chimio- et les régions électroactivité par rapport aux solvants organiques. D'un point de vue pratique et économique, la grande variété de liquides ioniques permet d'améliorer les réactions selon les propriétés propres à chaque liquide. De plus, il est possible de séparer plus facilement le produit de la réaction et le catalyseur utilisé permettant ainsi un possible recyclage des liquides ioniques (aspect environnemental).

La réaction de Diels – Alder est l'une des réactions de condensation carbone – carbone les plus utilisées en chimie organique. La réaction réalisée dans les liquides ioniques est un peu plus rapide que dans l'eau et beaucoup plus rapide que dans les solvants organiques, mais l'ajout d'un acide de Lewis permet une amélioration de la rapidité et de la sélectivité de la réaction. Avec cette propriété, Lee et coll. ont mis en évidence l'intérêt des liquides ioniques chloroaluminate dans ce type de réactions [87]. La réaction catalytique d'hydrogénation d'une double liaison carbone – carbone par transition d'un complexe métallique, est l'une des réactions les plus étudiées en catalyse

homogène [46]. De nombreuses études ont montré que l'utilisation des liquides ioniques améliorerait les rendements et la sélectivité de ce type de réactions [88, 89].

II.5.3. Applications dans le domaine des procédés de séparation et de l'analyse

Les différentes propriétés originales des liquides ioniques, présentent un très grand intérêt dans le domaine des procédés de séparation et de l'analyse chimique. Leurs capacités à dissoudre des composés organiques apolaires aussi bien que des composés inorganiques ioniques en ont fait des milieux de choix pour les sciences séparatives [90, 91, 92].

Les propriétés physico-chimiques des liquides ioniques telles que la pression de vapeur négligeable, la non-miscibilité avec d'autres solvants, et la bonne solubilité des composés organiques et inorganiques, en font un milieu de choix pour les techniques d'extraction liquide-liquide, l'extraction de composés organiques et inorganiques ioniques et de micro-extraction en phase liquide (LPME) l'extraction de 45 polluants environnementaux (benzène, toluène, HAP, phénols...) [93].

II.6. CONCLUSION

Les liquides ioniques peuvent paraître déroutants, du fait du manque de accumulées mais ils ne sont en réalité que des solvants différents dont les propriétés peuvent être aussi bien (ou aussi mal) rationalisées que celles de solvants plus usuels. Leurs propriétés, font que ces milieux commencent à être reconnus comme des solvants à part entière et non plus seulement comme des curiosités de laboratoire. De fait, de très nombreux travaux sont actuellement en cours, ce qui est avéré par le nombre croissant d'articles les concernant paraissant dans les revues scientifiques.

Trois axes majeurs d'études sont abordés de front :

- La synthèse de nouveaux liquides ioniques fin de répondre à des besoins bien précis
- L'étude des propriétés physicochimiques des liquides ioniques ;
- L'emploi des liquides ioniques dans des procédés ou dans des domaines d'activités très divers, dans certains cas pour répondre à des problèmes environnementaux, en remplaçant les solvants organiques par des liquides ioniques, et dans un nombre croissant d'autres cas, pour tirer directement parti des propriétés des liquides ioniques, en vue d'applications ciblées.

Ainsi, le domaine d'étude et d'application des liquides ioniques est-il vaste et fécond. Si l'on considère la récence de leur apparition, il n'est pas douteux que de très nombreuses années d'études seront encore nécessaires avant de maîtriser ces nouveaux solvants. En effet, vu la très grande quantité de liquides ioniques potentiellement accessibles, l'obtention de données précises nécessitera du temps et, pour l'heure, seules des informations parcellaires sont disponibles.

Ces propriétés ont éveillé un grand intérêt dû à leurs applications potentielles dans des nombreux domaines industriels. Dans le chapitre suivant, nous allons proposer une nouvelle méthode pour identifier les propriétés physiques et chimiques des liquides ioniques, avec un outil de data mining qui est dans notre cas le logiciel XLSTAT.

References bibliographies

- [1] M. P. Bollin, "The Production of Fatty Acid Methyl Esters in Lewis Acidic Ionic Liquids", the University of Toledo, 2011.
- [2] MihkelKoel "Ionic Liquids in Chemical Analysis" Analytical Chemistry. by Taylor & Francis Group, LLC, 2009.
- [3]M. Antonietti, D. Kuang, B. Smarsly, and Y. Zhou , "Ionic liquids for the convenient synthesis of functional nanoparticles and other inorganic nanostructures ", *Angewandte Chemie International Edition* , vol.43, no.38, pp.4988-4992,2004.
- [4] J.H. Huang, H.L.Xie, Zheng, Y.L.Wu, and W.X.Zhu, "Synthesis and Liquefaction Properties of L-Aanine Hydrosulfate Ionic liquid" *Advanced Materials Research*, vol.476, pp. 1351-1354, 2012.
- [5] J. S. Wilkes et M. J. Zaworotko, « Air and water stable 1-ethyl-3-methylimidazolium based ionic liquids », DTIC Document, 1992.
- [6] J. S. Wilkes, J. A. Levisky, R. A. Wilson, et C. L. Hussey, « Dialkylimidazoliumchloroaluminate melts: a new class of room-temperature ionic liquids for electrochemistry, spectroscopy and synthesis », *Inorganic Chemistry*, vol. 21, n°. 3, p. 1263–1264, 1982.
- [7] F. Endres et S. Z. El Abedin, « Air and water stable ionic liquids in physical chemistry », *Phys. Chem. Chem. Phys.*, vol. 8, n°. 18, p. 2101–2116, 2006.
- [8] A. Berthod et S. Carda-Broch, « Use of the ionic liquid 1-butyl-3-methylimidazolium hexafluorophosphate in countercurrent chromatography », *Analytical and bioanalytical chemistry*, vol. 380, n°. 1, p. 168–177, 2004.
- [9] C. L. Hussey, « Room temperature haloaluminate ionic liquids. Novel solvents for transition metal solution chemistry », 1988.
- [10] T. Welton et others, « Room-temperature ionic liquids. Solvents for synthesis and catalysis », *Chemical Reviews*, vol. 99, p. 2071–2084, 1999.
- [11] M. Deetlefs et K. R. Seddon, « Improved preparations of ionic liquids using microwave irradiation », *Green Chemistry*, vol. 5, n°. 2, p. 181–186, 2003.
- [12] R. S. Varma, « Chem. Com-mun. 2001, 643-644; RS Varma, VV Namboodiri », *Tetrahedron Lett*, vol. 43, p. 5381–5383, 2002.
- [13] J. M. Lévêque, J. L. Luche, C. Pétrier, R. Roux, et W. Bonrath, « An improved preparation of ionic liquids by ultrasound », *Green Chem.*, vol. 4, n°. 4, p. 357–360, 2002.
- [14] R. S. Varma et V. V. Namboodiri, « An expeditious solvent-free route to ionic liquids

using microwaves », *Chemical Communications*, n° 7, p. 643–644, 2001.

[15] V. Lecocq, A. Graille, C. C. Santana, A. Bedouin, Y. Chauvin, J. M. Basset, L. Ariel, D. Boucher, et B. Feet, « Synthesis and characterization of ionic liquids based upon 1-butyl-2, 3-dimethylimidazolium chloride/ZnCl₂ », *New J. Chem.*, vol. 29, n° 5, p. 700–706, 2005.

[16] A. Abbott, « Electrochemistry in media of low dielectric constant », *Chem. Soc. Rev.*, vol. 22, n° 6, p. 435–440, 1993.

[17] C. M. Gordon, J. D. Hombre, A. R. Kennedy, et K. R. Seddon, « Ionic liquid crystals: hexafluorophosphate salts », *J. mater. Chem.*, vol. 8, n° 12, p. 2627–2636, 1998.

[18] J. G. Huddleston et R. D. Rogers, « Room temperature ionic liquids as novel media for ‘clean ‘liquid–liquid extraction », *Chem. Commun.*, n° 16, p. 1765–1766, 1998.

[19] L. Cammarata, S. G. Kazarian, P. A. Salter, et T. Welton, « Molecular states of water in room temperature ionic liquids », *Phys. Chem. Chem. Phys.*, vol. 3, n° 23, p. 5192–5200, 2001.

[20] J. S. Wilkes et M. J. Zaworotko, « Air and water stable 1-ethyl-3-methylimidazolium based ionic liquids », DTIC Document, 1992.

[21] J. D. Hombre et K. R. Seddon, « The phase behaviour of 1-alkyl-3-methylimidazolium tetrafluoroborates; ionic liquids and ionic liquid crystals », *J. Chem. Soc., Dalton Trans.*, n° 13, p. 2133–2140, 1999.

[22] N. L. Lancaster, T. Welton, et G. B. Young, « A study of halide nucleophilicity in ionic liquids », *J. Chem. Soc., Perkin Trans. 2*, n° 12, p. 2267–2270, 2001.

[23] P. Bonhôte, A. P. Dias, N. Papageorgiou, K. Kalyanasundaram, et M. Grätzel, « Hydrophobic, highly conductive ambient-temperature molten salts », *Inorganic Chemistry*, vol. 35, n° 5, p. 1168–1178, 1996.

[24] J. Fuller, R. T. Carlin, et R. A. Osteryoung, « The Room Temperature Ionic Liquid 1-Ethyl-3-methylimidazolium Tetrafluoroborate: Electrochemical Couples and Physical Properties », *Journal of the Electrochemical Society*, vol. 144, p. 3881, 1997.

[25] K. R. Seddon, A. Stark, M. J. Torres, et others, « Influence of chloride, water, and organic solvents on the physical properties of ionic liquids », *Pure Appl. Chem*, vol. 72, n° 12, p. 2275–2287, 2000.

[26] D. R. MacFarlane, J. Golding, S. Forsyth, M. Forsyth, et G. B. Deacon, « Low viscosity ionic liquids based on organic salts of the dicyanamide anion », *Chemical communications*, n° 16, p. 1430–1431, 2001.

[27] A. S. Larsen, J. D. Hombre, F. S. Tham, et C. A. Reed, « Designing ionic liquids: imidazolium melts with inert carborane anions », *Journal of the American Chemical Society*,

vol. 122, n°. 30, p. 7264–7272, 2000.

[28] M. Hasan, I. V. Kozhevnikov, M. R. . Siddiqui, A. Steiner, et N. Winterton, « Gold compounds as ionic liquids. Synthesis, structures, and thermal properties of N, N'-dialkylimidazoliumtetrachloroaurate salts », *Inorganic Chemistry*, vol. 38, n°. 25, p. 5637–5641, 1999.

[29] P. Wasserscheid, W. Keim, et others, « Ionic liquids-new“ solutions” for transition metal catalysis », *Angewandte Chemie*, vol. 39, n°. 21, p. 3772–3789, 2000.

[30] P. A. . Suarez, V. M. Selbach, J. E. .Dullius, S. Einloft, C. Piatnicki, D. S. Azambuja, R. F. de Souza, et J. Dupont, « Enlarged electrochemical window in dialkyl-imidazoliumcation based room-temperature air and water-stable molten salts », *Electrochimicaacta*, vol. 42, n°. 16, p. 2533–2535, 1997.

[31] C. Hardacre, J. D. Hombre, M. Nieuwenhuyzen, et T. G. .Youngs, « Structure and solvation in ionic liquids », *Accounts of chemical research*, vol. 40, n°. 11, p. 1146–1155, 2007.

[32] L. Crowhurst, P. R. Mawdsley, J. M. Perez-Arlandis, P. A. Salter, et T. Welton, « Solvent–solute interactions in ionic liquids », *Physical Chemistry Chemical Physics*, vol. 5, n°. 13, p. 2790–2794, 2003.

[33] P. Kölle et R. Dronskowski, « Hydrogen bonding in the crystal structures of the ionic liquid compounds butyldimethylimidazolium hydrogen sulfate, chloride, and chloroferrate (II, III) », *Inorganic chemistry*, vol. 43, n°. 9, p. 2803–2809, 2004.

[34] A. Arce, M. J. Earle, S. P. Katdare, H. Rodríguez, et K. R. Seddon, « Mutually immiscible ionic liquids », *Chemical Communications*, n°. 24, p. 2548–2550, 2006.

[35] A. Noda, K. Hayamizu, et M. Watanabe, « Pulsed-gradient spin-echo ¹H and ¹⁹F NMR ionic diffusion coefficient, viscosity, and ionic conductivity of non-chloroaluminate room-temperature ionic liquids », *The Journal of Physical Chemistry B*, vol. 105, n°. 20, p. 4603–4610, 2001.

[36] K. R. Seddon, A. Stark, M. J. Torres, et others, « Influence of chloride, water, and organic solvents on the physical properties of ionic liquids », *Pure Appl. Chem*, vol. 72, n°. 12, p. 2275–2287, 2000.

[37] C. Dagueuet et P. J. Dyson, « Inhibition of catalytic activity in ionic liquids: Implications for catalyst design and the effect of cosolvents », *Organometallics*, vol. 23, n°. 26, p. 6080–6083, 2004.

[38] L. Magna, Y. Chauvin, G. P. Niccolai, et J. M. Basset, « The importance of imidazolium substituents in the use of imidazolium-based room-temperature ionic liquids as solvents for

palladium-catalyzed telomerization of butadiene with methanol », *Organometallics*, vol. 22, n°. 22, p. 4418–4425, 2003.

[39] C. D. Tran, S. H. De Paoli Lacerda, et D. Oliveira, « Absorption of water by room-temperature ionic liquids: effect of anions on concentration and state of water », *Applied spectroscopy*, vol. 57, n°. 2, p. 152–157, 2003.

[40] H. Olivier-Bourbigou et L. Magna, « Ionic liquids: perspectives for organic and catalytic reactions », *Journal of Molecular Catalysis A: Chemical*, vol. 182, p. 419–437, 2002.

[41] V. Najdanovic-Visak, J. M. S. .Esperança, L. P. .Rebelo, M. N. da Ponte, H. J. .Guedes, K. R. Seddon, et J. Szydłowski, « Phase behaviour of room temperature ionic liquid solutions: an unusually large co-solvent effect in (water+ ethanol) », *Physical Chemistry Chemical Physics*, vol. 4, n°. 10, p. 1701–1703, 2002.

[42] R. P. Swatloski, J. D. Hombre, et R. D. Rogers, « Ionic liquids are not always green: hydrolysis of 1-butyl-3-methylimidazolium hexafluorophosphate », *Green Chem.*, vol. 5, n°. 4, p. 361–363, 2003.

[43] I. Billard, G. Moutiers, A. Labet, A. El Azzi, C. Gaillard, C. Mariet, et K. Lützenkirchen, « Stability of divalent europium in an ionic liquid: Spectroscopic investigations in 1-methyl-3-butylimidazolium hexafluorophosphate », *Inorganic chemistry*, vol. 42, n°. 5, p. 1726–1733, 2003.

[44] J. D. Hombre et K. R. Seddon, « The phase behaviour of 1-alkyl-3-methylimidazolium tetrafluoroborates; ionic liquids and ionic liquid crystals », *J. Chem. Soc., Dalton Trans.*, n°. 13, p. 2133–2140, 1999.

[45] S. G. Cull, J. D. Hombre, V. Vargas-Mora, K. R. Seddon, et G. J. Lye, « Room-temperature ionic liquids as replacements for organic solvents in multiphase bioprocess operations », *Biotechnology and bioengineering*, vol. 69, n°. 2, p. 227–233, 2000.

[46] M. J. Earle, J. M. S. . Esperança, M. A. Gilea, J. N. . Lopes, L. P. .Rebelo, J. W. Magee, K. R. Seddon, et J. A. Widegren, « The distillation and volatility of ionic liquids », *Nature*, vol. 439, n°. 7078, p. 831–834, 2006.

[47] P. Stepnowski, J. Nichthauser, W. Mroziak, et B. Buszewski, « Usefulness of π - π aromatic interactions in the selective separation and analysis of imidazolium and pyridinium ionic liquid cations », *Analytical and bioanalytical chemistry*, vol. 385, n°. 8, p. 1483–1491, 2006.

[48] L. Xiao et K. E. Johnson, « Electrochemistry of 1-butyl-3-methyl-1H-imidazolium tetrafluoroborate ionic liquid », *Journal of the Electrochemical Society*, vol. 150, p. E307, 2003.

- [49] J. F. Huang et I. W. Sun, « Electrochemical studies of tin in zinc chloride-1-ethyl-3-methylimidazolium chloride ionic liquids », *Journal of the Electrochemical Society*, vol. 150, p. E299, 2003.
- [50] J. D. Hombre, K. R. Seddon, et R. Wareing, « A simple colorimetric method for the quality control of 1-alkyl-3-methylimidazolium ionic liquid precursors », *Green Chemistry*, vol. 3, n° 1, p. 33–36, 2001.
- [51] J. Ranke, S. Stolte, R. Störmann, J. Arning, et B. Jastorff, « Design of Sustainable Chemical Products The Example of Ionic Liquids », *Chem. Rev.*, vol. 107, p. 2183, 2007.
- [52] B. Jastorff, R. Störmann, J. Ranke, K. Mölter, F. Stock, B. Oberheitmann, W. Hoffmann, J. Hoffmann, M. Nüchter, B. Ondruschka, et others, « How hazardous are ionic liquids? Structure– activity relationships and biological testing as important elements for sustainability evaluation », *Green Chem.*, vol. 5, n° 2, p. 136–142, 2003.
- [53] B. Jastorff, K. Mölter, P. Behrend, U. Bottin-Weber, J. Filser, A. Heimers, B. Ondruschka, J. Ranke, M. Schaefer, H. Schröder, et others, « Progress in evaluation of risk potential of ionic liquids—basis for an eco-design of sustainable products », *Green Chem.*, vol. 7, n° 5, p. 362–372, 2005.
- [54] M. Smiglak, W. M. Reichert, J. D. Hombre, J. S. Wilkes, L. Sun, J. S. Thrasher, K. Kirichenko, S. Singh, A. R. Katritzky, et R. D. Rogers, « Combustible ionic liquids by design: is laboratory safety another ionic liquid myth? », *Chem. Commun.*, n° 24, p. 2554–2556, 2006.
- [55] L. C. Branco, J. N. Rosa, J. J. Moura Ramos, et C. A. Afonso, « Preparation and characterization of new room temperature ionic liquids », *Chemistry-A European Journal*, vol. 8, n° 16, p. 3671–3677, 2002.
- [56] H. L. Ngo, K. LeCompte, L. Hargens, et A. B. McEwen, « Thermal properties of imidazolium ionic liquids », *Thermochimica Acta*, vol. 357, p. 97–102, 2000.
- [57] U. Domanska, E. Bogel-Lukasik, et R. Bogel-Lukasik, « Solubility of 1-dodecyl-3-methylimidazolium chloride in alcohols (C2-C12) », *The Journal of Physical Chemistry B*, vol. 107, n° 8, p. 1858–1863, 2003.
- [58] A. E. Bradley, C. Hardacre, J. D. Hombre, S. Johnston, S. E. J. McMath, et M. Nieuwenhuyzen, « Small-angle X-ray scattering studies of liquid crystalline 1-alkyl-3-methylimidazolium salts », *Chemistry of materials*, vol. 14, n° 2, p. 629–635, 2002.
- [59] J. C. ADRIAN, C. Hardacre, J. D. Hombre, M. Nieuwenhuyzen, et R. S. KENNETH, « Molecular layering and local order in thin films of 1-alkyl-3-methylimidazolium ionic liquids using X-ray reflectivity », *Molecular Physics*, vol. 99, n° 10, p. 795–800, 2001.

- [60] A. R. Katritzky, R. Jain, A. Lomaka, R. Petrukhin, M. Karelson, A. E. Visser, et R. D. Rogers, « Correlation of the melting points of potential ionic liquids (imidazolium bromides and benzimidazolium bromides) using the CODESSA program », *Journal of chemical information and computer sciences*, vol. 42, n° 2, p. 225–231, 2002.
- [61] A. E. Visser, J. D. Hombre, et R. D. Rogers, « Hydrophobic ionic liquids incorporating N-alkylisoquinoliniumcations and their utilization in liquid–liquid separations », *Chemical Communications*, n° 23, p. 2484–2485, 2001.
- [62] H. Matsumoto, H. Kageyama, et Y. Miyazaki, « Room temperature molten salts based on tetraalkylammoniumcations and bis (trifluoromethylsulfonyl) imide », *Chemistry Letters*, vol. 30, n° 2, p. 182–183, 2001.
- [63] J. Sun, M. Forsyth, et D. R. MacFarlane, « Room-temperature molten salts based on the quaternary ammonium ion », *The Journal of Physical Chemistry B*, vol. 102, n° 44, p. 8858–8864, 1998.
- [64] S. V. Dzyuba et R. A. Bartsch, « New room-temperature ionic liquids with C2-symmetrical imidazoliumcations Electronic supplementary information (ESI) available: 1H NMR and IR spectra and elemental analysis results consistent with the indicated structures for all of the [(Cn) 2-Im] Br and [(Cn) 2-Im] PF6 compounds. See <http://www.rsc.org/suppdata/cc/b1/b104512c> », *Chemical Communications*, n° 16, p. 1466–1467, 2001.
- [65] R. Hagiwara et Y. Ito, « Room temperature ionic liquids of alkylimidazoliumcations and fluoroanions », *Journal of Fluorine Chemistry*, vol. 105, n° 2, p. 221–227, 2000.
- [66] K. Xu, M. S. Ding, et T. R. Jow, « Quaternary onium salts as nonaqueous electrolytes for electrochemical capacitors », *Journal of the Electrochemical Society*, vol. 148, p. A267, 2001.
- [67] J. Sun, D. R. MacFarlane, et M. Forsyth, « Synthesis and properties of ambient temperature molten salts based on the quaternary ammonium ion », *Ionics*, vol. 3, n° 5, p. 356–362, 1997.
- [68] R. Sheldon, « Catalytic reactions in ionic liquids », *Chem. Commun.*, n° 23, p. 2399–2407, 2001.
- [69] D. S. . Wong, J. P. Chen, J. M. Chang, et C. H. Chou, « Phase equilibria of water and ionic liquids [emim][PF₆] and [bmim][PF₆] », *Fluid phase equilibria*, vol. 194, p. 1089–1095, 2002.
- [70] J. G. Huddleston, A. E. Visser, W. M. Reichert, H. D. Willauer, G. A. Broker, et R. D. Rogers, « Characterization and comparison of hydrophilic and hydrophobic room temperature ionic liquids incorporating the imidazoliumcation », *Green Chem.*, vol. 3, n° 4, p. 156–164,

2001.

[71] P. A. Z. Suarez, S. Einloft, J. E. L. Dullius, R. F. De Souza, et J. Dupont, « Synthesis and physical-chemical properties of ionic liquids based on 1-n-butyl-3-methylimidazolium cation », *Journal de Chimie Physique et de PhysicoChimieBiologique*, vol. 95, n°. 7, p. 1626–1639, 1998.

[72] J. Pernak, A. Czepukowicz, et R. Pozniak, « New ionic liquids and their antielectrostatic properties », *Industrial & engineering chemistry research*, vol. 40, n°. 11, p. 2379–2383, 2001.

[73] S. Chun, S. V. Dzyuba, et R. A. Bartsch, « Influence of structural variation in room-temperature ionic liquids on the selectivity and efficiency of competitive alkali metal salt extraction by a crown ether », *Analytical chemistry*, vol. 73, n°. 15, p. 3737–3741, 2001.

[74] J. Lu, C. L. Liotta, et C. A. Eckert, « Spectroscopically probing microscopic solvent properties of room-temperature ionic liquids with the addition of carbon dioxide », *The Journal of Physical Chemistry A*, vol. 107, n°. 19, p. 3995–4000, 2003.

[75] Y. S. Fung et R. Q. Zhou, « Room temperature molten salt as medium for lithium battery », *Journal of power sources*, vol. 81, p. 891–895, 1999.

[76] V. M. Hultgren, A. W. . Mariotti, M. Alan, et A. G. Wedd, « Reference potential calibration and voltammetry at macrodisk electrodes of metallocene derivatives in the ionic liquid [bmim][PF6] », *Analytical chemistry*, vol. 74, n°. 13, p. 3151–3156, 2002.

[77] B. M. Choudary, M. L. Kantam, A. Rahman, C. V. Reddy, et K. K. Rao, « The First Example of Activation of Molecular Oxygen by Nickel in Ni-Al Hydrotalcite: A Novel Protocol for the Selective Oxidation of Alcohols », *Angewandte Chemie*, vol. 113, n°. 4, p. 785–788, 2001.

[78] P. A. . Suarez, C. S. Consorti, R. F. Souza, J. Dupont, et R. S. Gonçalves, « Electrochemical behavior of vitreous glass carbon and platinum electrodes in the ionic liquid 1-n-butyl-3-methylimidazolium trifluoroacetate », *Journal of the Brazilian Chemical Society*, vol. 13, n°. 1, p. 106–109, 2002.

[79] W. Xu, L. M. Wang, R. A. Nieman, et C. A. Angell, « Ionic liquids of chelated orthoborates as model ionic glassformers », *The Journal of Physical Chemistry B*, vol. 107, n°. 42, p. 11749–11756, 2003.

[80] C. Chiappe et D. Pieraccini, « Ionic liquids: solvent properties and organic reactivity », *Journal of physical organic chemistry*, vol. 18, n°. 4, p. 275–297, 2005.

[81] Y. François, K. Zhang, A. Varenne, et P. Gareil, « New integrated measurement protocol using capillary electrophoresis instrumentation for the determination of viscosity, conductivity and absorbance of ionic liquid–molecular solvent mixtures », *Analyticachimicaacta*, vol. 562,

n° 2, p. 164–170, 2006.

[82] D. Swartling, L. Ray, S. Compton, et D. Ensor, « Preliminary investigation into modification of ionic liquids to improve extraction parameters », *SAAS Bull.: Biochem. Biotechnol.*, vol. 13, p. 1–6, 2000.

[83] A. B. McEwen, H. L. Ngo, K. LeCompte, et J. L. Goldman, « Electrochemical properties of imidazolium salt electrolytes for electrochemical capacitor applications », *Journal of the Electrochemical Society*, vol. 146, p. 1687, 1999.

[84] J. S. Wilkes, « Properties of ionic liquid solvents for catalysis », *Journal of Molecular Catalysis A: Chemical*, vol. 214, n° 1, p. 11–17, 2004.

[85] M. C. Buzzeo, R. G. Evans, et R. G. Compton, « Non-haloaluminate room-temperature ionic liquids in electrochemistry—A review », *ChemPhysChem*, vol. 5, n° 8, p. 1106–1120, 2004.

[86] C. Baudequin, J. Baudoux, J. Levillain, D. Cahard, A. C. Gaumont, et J. C. Plaquevent, « Ionic liquids and chirality: opportunities and challenges », *Tetrahedron: Asymmetry*, vol. 14, n° 20, p. 3081–3093, 2003.

[87] « Synthesis and Activity of a New Generation of Ruthenium-Based Olefin Metathesis Catalysts Coordinated with 1, 3-Dimesityl-4, 5-dihydroimidazol-2-ylidene Ligands \S ».

[88] Y. Chauvin, L. Müssmann, et H. Olivier, « Flüssige 1, 3-Dialkylimidazoliumsalze als Lösungsmittel für die Katalyse in Zweiphasensystemen: durch Rhodiumkomplexe katalysierte Hydrierung, Isomerisierung und Hydroformylierung von Alkenen », *Angewandte Chemie*, vol. 107, n° 23–24, p. 2941–2943, 1995.

[89] P. A. . Suarez, J. E. .Dullius, S. Einloft, R. F. De Souza, et J. Dupont, « The use of new ionic liquids in two-phase catalytic hydrogenation reaction by rhodium complexes », *Polyhedron*, vol. 15, n° 7, p. 1217–1219, 1996.

[90] A. L. Monteiro, F. K. Zinn, R. F. de Souza, et J. Dupont, « Asymmetric hydrogenation of 2-arylacrylic acids catalyzed by immobilized Ru-BINAP complex in 1-n-butyl-3-methylimidazolium tetrafluoroborate molten salt », *Tetrahedron: Asymmetry*, vol. 8, n° 2, p. 177–179, 1997.

[91] J. F. Liu, J. A. Jonsson, et G. B. Jiang, *Trac-Trends Anal. Chem.*, 2005.

[92] J. L. Anderson, D. W. Armstrong, et G. T. Wei, « Ionic liquids in analytical chemistry », *Analytical chemistry*, vol. 78, n° 9, p. 2892–2902, 2006.

[93] J. Liu, Y. Chi, et G. Jiang, « Screening the extractability of some typical environmental pollutants by ionic liquids in liquid-phase microextraction », *Journal of separation science*, vol. 28, n° 1, p. 87–91, 2005.

Présentation du logiciel XLSTAT

III.1. Introduction

Les mesures expérimentales des propriétés physico-chimiques des composés organiques et liquides ioniques présentent de grands obstacles pour les chercheurs lors de la conception et de la modélisation des procédés chimiques. Donc afin de limiter ce problème, plusieurs techniques ont été proposées, parmi lesquelles les techniques du data mining qui deviennent indispensables.

Dans ce chapitre nous allons présenter l'outil que nous avons utilisé pour arriver à cet objectif qui est une partie du logiciel XLSTAT que nous avons appliqué à une base de données de liquides ioniques constituée d'une soixantaine de composé différents.

III.2. Base de données des liquides ioniques

Une base de données des propriétés physico-chimiques des liquides ioniques, qui ont été collectées à partir de plusieurs sources documentaires fiables (SNDL : Système National de Documentation en Ligne) couvrant la période de 1982 à 2012. La base contient 1774 entrées, réparties sur 28 propriétés physico-chimiques. Presque tous les types de propriétés physico-chimiques importantes sont recueillis dans cette base de données. [1]

Cation	Anion	Température l'étude (k°)	Point de fusion	conductivité σ (S, m-1)	Viscosité (cp)	densité ρ (g, cm-1)	conductivité molaire (cm ² , S, mol)
[1-butyl-1méthylpyrolidinium]	CF ₃ SO ₂) 2N	298		0,22	85	1,41	
[1-méthyl-1propylepyrolidinium]	CF ₃ SO ₂) 2N	298		0,14	63	1,45	
B3	BF1	298		0,13	103	1,22	
B3	CF ₃ SO ₂) 2N	303		0,22	57	1,44	0,75

BEIM	CF3CO2	293		0,25	89	1,183	0,63
BEIM	CH3SO3	298		0,55		1,14	0,56
BEIM	CF3SO2	298	2	0,27		1,27	0,12
BEIM	C4F9SO3	293	21	0,053	323	1,427	0,64
BEIM	CF3SO2) 2N	293		0,41	48	1,404	0,17
BEIM	CF3SO2) 2N	293		0,28	83	1,428	1,27
BEIM	BF4	295		0,023			0,76
BEIM	PF5	295		0,077			
CF3CH2MIM	CF3SO2) 2N	293		0,098	248	1,656	
(CH1)3S	HBr2	298		3,4	20.5	1,74	0,25
(CH1)3S	H2Br3	298		5,6	3.3	1,79	4,62
(CH1)3S	Al2CL7	295		0,55			8,41
(CH1)3S	Al2Cl3Br	298		0,421	54.9	1,59	
(CH1)3S	Al2Br2	298		0,141	138	1,58	1,19
(CH1)3S	CF3SO2) 2N	318		0,71	30	1,64	1,85
(CH3)2(C2H5) (CH3O)	BF1C2H4) N	298		0,17			1,94
DCIM	CF3SO3	298				1,1	
EEM	CF3CO2	293		0,74	43	1,26	
EEM	CF3SO3	293	23	0,75	53	1,33	1,41
EEM	CF3SO2) 2N	293	14	0,85	35	1,452	1,55
EEMIM	CF3SO2)	293	-3	0,62	36	1,432	2,37

	2N						
EMIM	BF1	298		1,4	1,43	1,24	1,82
EMIM	BF4	298		1,1	32		2,24
EMIM	BF1	299		1,3	43		
EMIM	BF1	303		2			
EMIM	BF1	299		0,52			
EMIM	PF3	293		0,23	162		
EMIM	CH3CO2	293		0,93	35	1,385	1,67
EMIM	C3F2CO2	293		0,27	105	1,45	0,6
EMIM	CH3SO3	298		0,27	100	1,24	0,45
EMIM	CF3SO3	293	-9	0,85	45	1,39	1,61
EMIM	CF3SO2) 2N	295		0,92	43	1,38	1,73
EMIM	CF3SO2	293	-9	0,83	34	1,52	2,27
EMIM	C2F5SO2)2N	299		0,34	61		
EMIM	CF2SO2	293		0,32	88	1,485	0,87
EMIMIM	CF3SO2 N	293		0,64	61	1,334	1,32
EMIMIM	CF3SO2) 2N	293		0,63	37	1,47	1,82
HMIM	PF5	293				1,307	
MeoEtMIM	CF3SO2) 2N	293		0,35	74	1,354	0,77
MeoEtMIM	CF3SO2) 2N	293		0,42	54	1,496	1,18
(nC3H2)CH3N)	CF3SO2)	298		0,33	72	1,44	0,88

	2N						
OMIM	PF1	293				1,237	
PMIM	PF1	293				1,33	
PMIM	BF2	295		0,53		1,33	
PMIM	PF5	308		0,05			
PMIM	CF3SO2) 2N	299		0,3	60		
PMIM	BF2	293				1,21	
PMIM	PF3	293		0,14	312	1,36	
PMIM	PF5	295		0,106			
PMIM	CF3CO2	293		0,32	73	1,209	0,61
PMIM	C3F7CO2	293		0,32	73	1,209	0,61
PMIM	CF3SO3	293		0,1	182	1,33	0,26
PMIM	C4F3SO4	293	16	0,041	373	1,427	0,13
PMIM	CF3SO2) 2N	293	-4	0,39	52	1,429	1,14

Tableau III-1- Les propriétés physique chimique des liquides ioniques. [1]

III.3. Logiciel XLSTAT

XLSTAT est un outil d'analyse de données et de statistiques pour Microsoft Excel. Il offre de très nombreuses fonctionnalités qui font d'Excel un outil performant et facile d'accès pour répondre à la majorité des besoins en analyse de données et modélisation.

Toutes les fonctions XLSTAT sont accessibles à partir d'une icône qui est ajoutée à la barre de menus de l'Excel. L'utilisation d'Excel comme interface rend ce logiciel simple à utiliser et efficace. La qualité des calculs est quant à elle identique à celle des logiciels scientifiques les plus renommés, et XLSTAT couvre l'essentiel des besoins du statisticien qu'il soit expert ou débutant.

Des modules optionnels répondent à des besoins plus spécifiques (contrôle statistique des procédés, analyse des effets de doses en chimie et pharmacologie). Il fonctionne avec toutes les versions d'Excel, depuis la version 97 jusqu'à la version 2007 sous les environnements Windows et Mac.

XLSTAT est l'outil d'analyse des données et de statistiques pour Microsoft Excel le plus utilisé parce que c'est un logiciel sûr, rapide, facile à installer et à utiliser. [2]

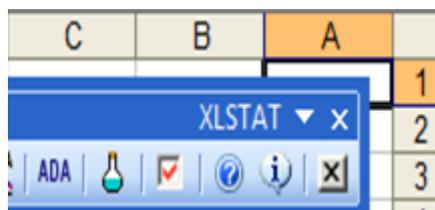


Figure.III-1 -Figure qui représente le logiciel XLSTAT.

III.3.1. Application de l'analyse en composantes principales (ACP)

L'Analyse en Composantes Principales (ACP) est une méthode très efficace pour l'analyse de données quantitatives (continues ou discrètes) se présentant sous la forme de tableaux à M observations / N variables. Elle permet de :

- visualiser et analyser rapidement les corrélations entre les N variables,
- visualiser et analyser les M observations initialement décrites par N variables sur un graphique à deux ou trois dimensions, construit de manière à ce que la dispersion entre les données soit aussi bien préservée que possible,
- construire un ensemble de P facteurs non corrélés

Les limites de l'Analyse en Composantes Principales viennent du fait que c'est une méthode de projection, et que la perte d'information induite par la projection peut entraîner des interprétations erronées. Des astuces permettent cependant d'éviter ces inconvénients.

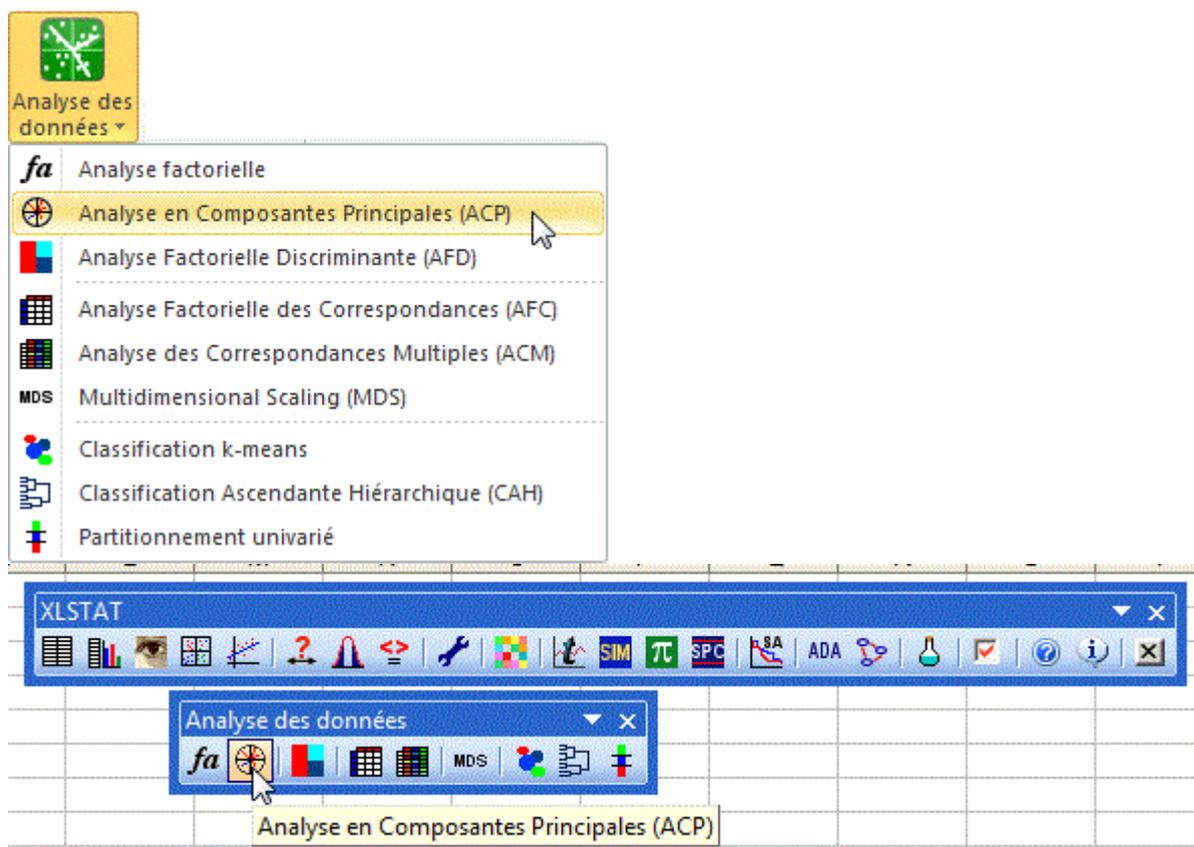
Il est également important de noter que l'ACP est un outil de **statistique exploratoire** et ne permet pas généralement de tester des hypothèses. Un avantage de cet

aspect exploratoire est que l'ACP peut être exécutée à différentes reprises, en éliminant / rajoutant des observations ou des variables, à condition que ces manipulations soient justifiées dans les interprétations.[2]

III.3.2. Paramétrer une analyse en composantes principales

Pour exécuter une analyse en composante principale avec le logiciel XLSTAT, il faut passer par les étapes suivantes :

XLSTAT / Analyse de données / Analyse en Composantes Principales, ou cliquez sur le bouton correspondant de la barre **Analyse de données** (voir ci-dessous). [4]



Une fois le bouton cliqué, la boîte de dialogue correspondant à l'**Analyse en composantes principales** apparaît.

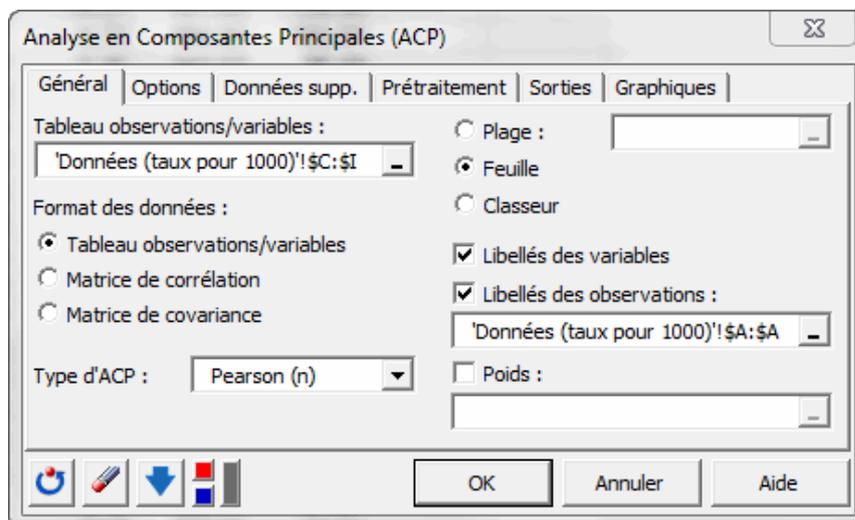
Vous pouvez alors sélectionner les données sur la feuille Excel.

Il y a plusieurs façons de sélectionner les données dans les boîtes de dialogue XLSTAT,

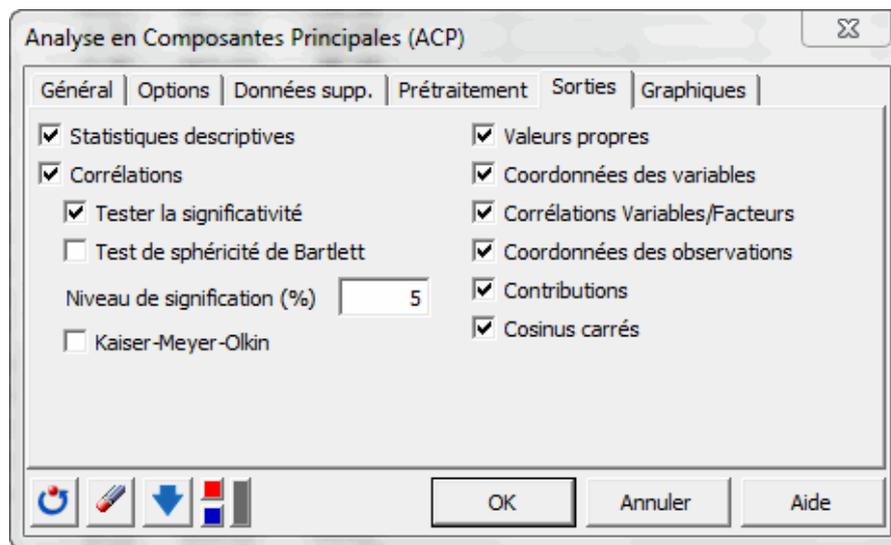
L'option Libellés des variables est activée, car la première ligne de données contient le nom des variables.

Le Format des données choisi ici est Observations/Variables car c'est bien le format des données de départ.

Le Type d'ACP choisi est Pearson (n), ce qui signifie que les calculs seront basés sur une matrice composée des coefficients de corrélation de Pearson, le coefficient de Pearson étant le coefficient de corrélation classiquement utilisé. Les matrices de covariance allouent plus de poids aux variables ayant des variances élevées. Les corrélations de Spearman peuvent être plus appropriées lorsque l'ACP est exécuté sur des variables aux distributions différentes. Les corrélations polychoriques sont adaptées aux variables ordinales.



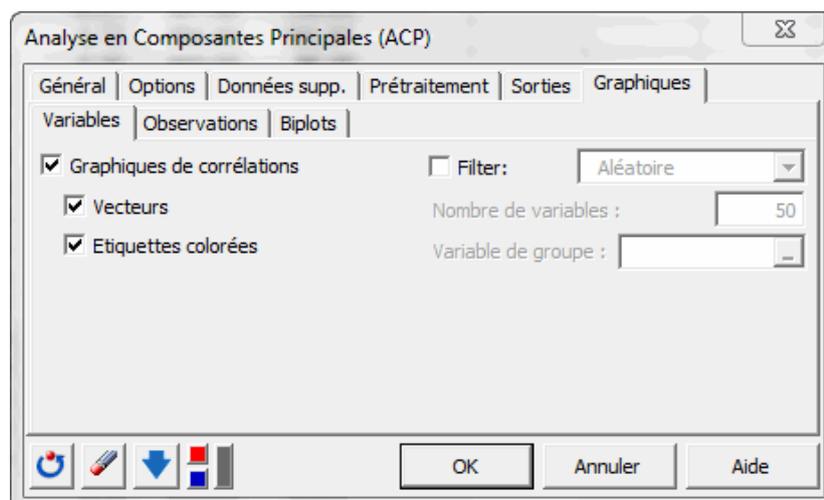
Dans l'onglet Sorties, on a choisi d'activer l'option Tester la significativité pour afficher en gras les corrélations significativement différentes de 0.

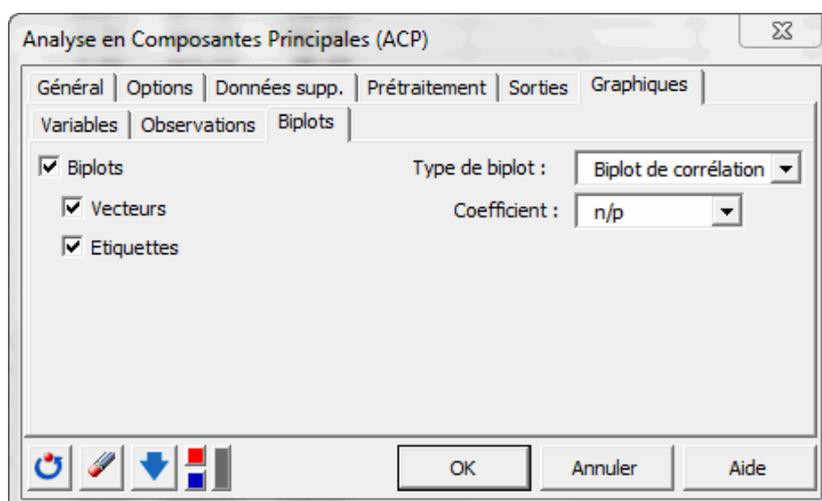
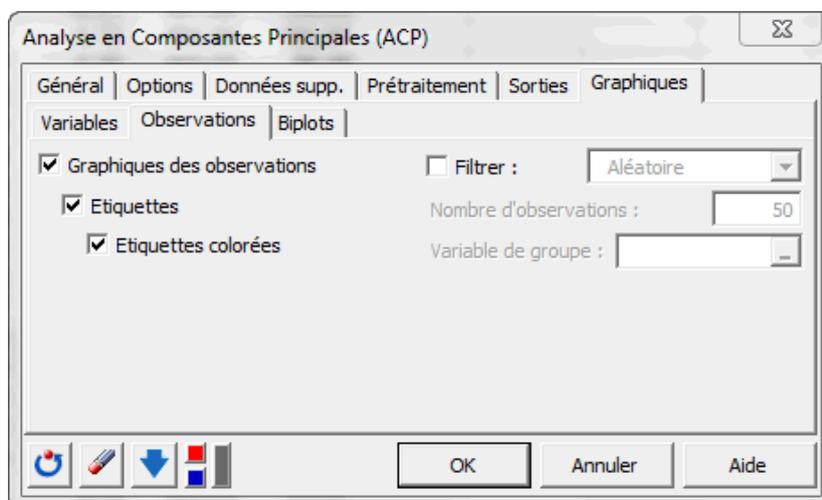


Dans l'onglet Graphiques, les options d'Étiquettes sont toutes activées afin que les libellés des variables et des observations soient bien affichés.

L'option de filtrage des observations à afficher est aussi désactivée afin d'afficher toutes les observations.

Lorsqu'il y a beaucoup d'observations, il est conseillé de ne pas afficher les étiquettes pour accélérer la génération du graphique, et de ne pas afficher toutes les observations afin de rendre le graphique plus lisible.

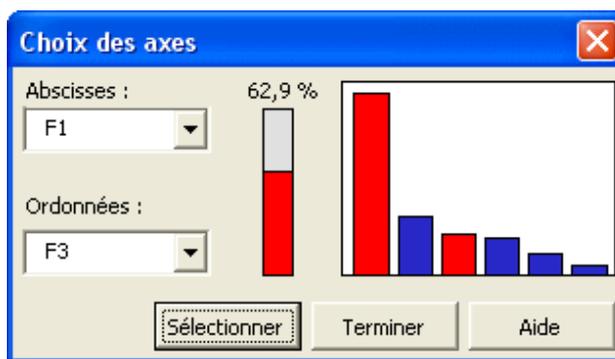
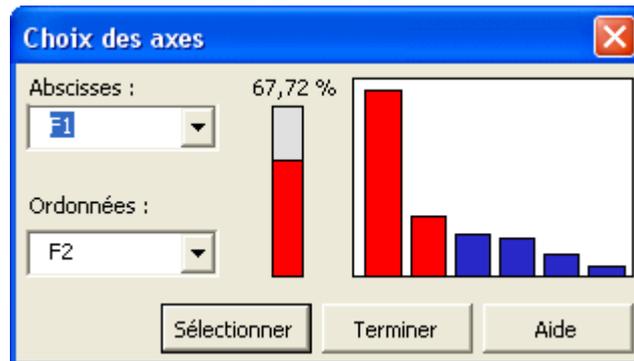




Les calculs commencent lorsque vous cliquez sur le bouton **OK**.

Si vous avez choisi dans le panneau des options XLSTAT l'option Demander la confirmation des sélections, XLSTAT vous demande de confirmer le nombre de lignes et de colonnes des

sélections. Puis une nouvelle boîte vous permet de choisir les axes pour lesquels les graphiques doivent être affichés.



III.4. Conclusion

Dans ce chapitre, nous avons présenté la base de données des liquides ioniques sur laquelle nous avons appliqué une analyse de type ACP à travers le logiciel XLSTAT. La partie du logiciel concernant cette analyse a été présentée. Les résultats obtenus et leurs interprétations seront exposés dans le chapitre suivant.

Références bibliographies

- [1] Wasserscheid, T. Welton, Ionic liquids in synthesis, Wiley-VCH, New-York, 2003
- [2] TALI Halima (2012/2013) thèse de master Développement d'une méthode prédictive de calcul des propriétés des liquides ioniques
- [3] Modèle mathématique <http://WWW.answers.com economic.order.quantity>.
- [4] http://eire.census.gov/popest/states_dataset.csv).

Résultats et discussions

IV.1.Introduction

Le but de ce travail, est le développement d'un modèle mathématique basé sur les données expérimentales disponibles en fonction de la structure moléculaire, permettant ainsi d'estimer, d'étudier et d'identifier les propriétés physico-chimiques des liquides ioniques.

Dans ce chapitre nous allons présenter quelques résultats obtenus par la méthode ACP en utilisant le logiciel XLSTAT.

IV.2.Résultats de l'ACP

L'application de l'ACP sur la base de données des liquides ioniques conçue et décrite dans le chapitre précédent a permis d'obtenir les résultats suivants :

	F1	F2	F3	F4	F5	F6
Valeur propre	2,5454	1,2651	0,9825	0,7430	0,3720	0,0919
Variabilité (%)	42,4238	21,0855	16,3749	12,3829	6,2007	1,5322
% cumulé	42,4238	63,5093	79,8842	92,2671	98,4678	100,0000

Tableau. IV-1-Valeurs propres.

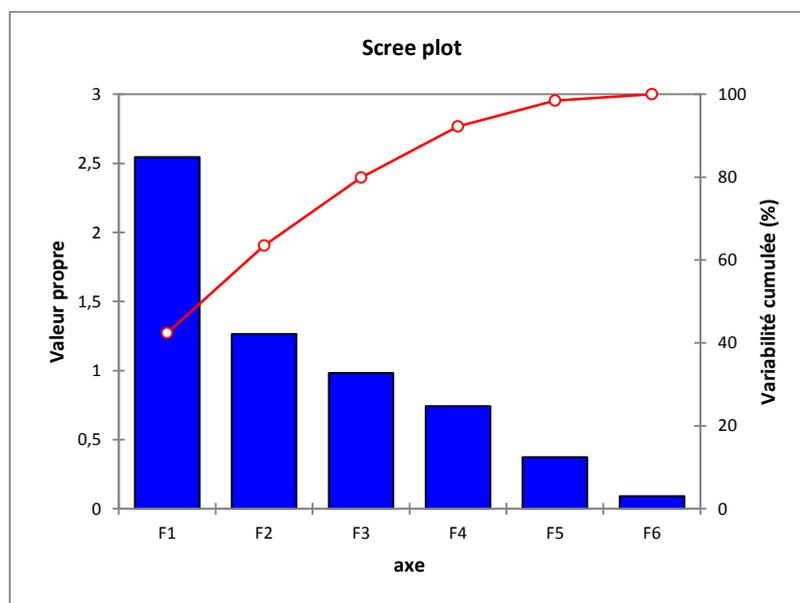


Figure .IV-1-Répartition de l’inertie entre les axes.

IV.3.Discussions

L'analyse du tableau des valeurs propres et la figure du scree plot permet de constater que la majeure partie de l'information concernant les propriétés des liquides ioniques est expliquée par les trois ou quatre premiers axes factoriels un pourcentage cumulé de 92.26% est obtenu. Pour simplifier l'étude, on limite le nombre de composantes à 3. Dans le plan factoriel F1xF2, F1xF3 et F2xF3 les valeurs propres des trois composantes F1, F2 et F3 et leur contribution à l'inertie totale sont représentées dans le Tableau IV-1. Dans ce cas, les trois axes pris en considération pour décrire les corrélations entre les variables liées aux liquides ioniques, détiennent à eux seuls 79.88 % de l'information totale avec respectivement 42.42 % pour l'axe 1, 20,08 % pour l'axe 2 et 16,37% pour l'axe 3 donc on confirme notre choix des trois composantes principales.

La figure IV-2 représente le cercle de corrélation entre les variables concernant les deux composantes F1 et F2 avec des participations respectives de 42,42% et 21,09%

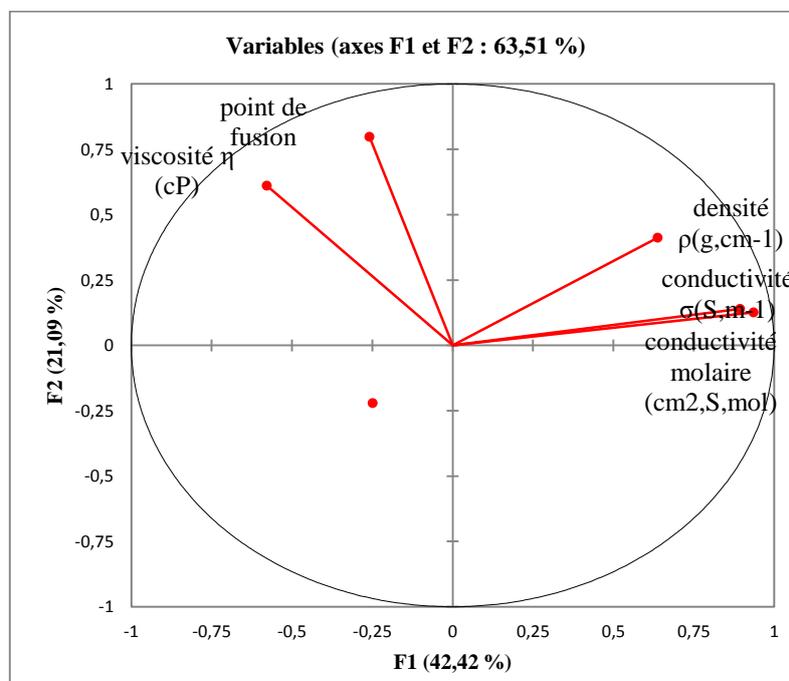


Figure - IV-2 - le cercle de corrélation (plan factoriel F1, F2).

La figure IV-2 est un graphique particulier à la méthode ACP et qui est le cercle des corrélations (cercle sur les axes F1 et F2). Il correspond à une projection des variables initiales sur un plan à deux dimensions constitué par les deux premiers facteurs. On remarque que sur l'axe F1 la conductivité et la conductivité molaire sont très bien représentées avec des corrélations respectives de 0.87 et 0.91. La densité et la viscosité sont moins bien représentées avec des corrélations de 0.58 et 0.64 ainsi que le point de fusion avec une corrélation très faible de 0.24.

Pour le deuxième axe, les corrélations les plus importantes concernant la viscosité et le point de fusion avec des valeurs respectives de 0,58 et 0,79.

La figure IV-3 présente toutes les observations dans le plan factoriel F1 x F2. En considérant la composante F1 et en tenant compte de la conductivité ou la conductivité molaire, on constate que

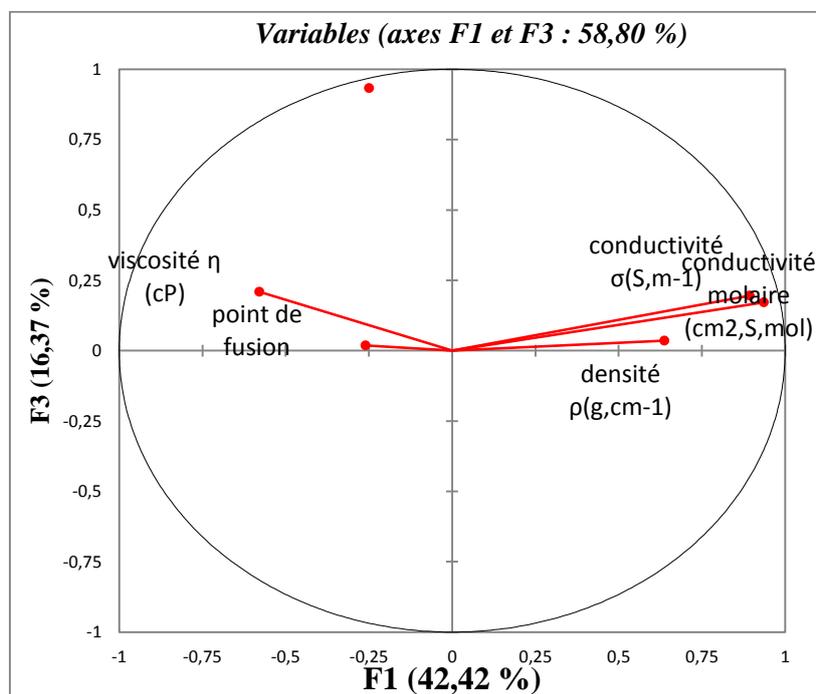


Figure- IV-4- le cercle de corrélation (plan factoriel F1, F3).

On remarque que la plupart des variables son faiblement représenté sur l'axe F3.

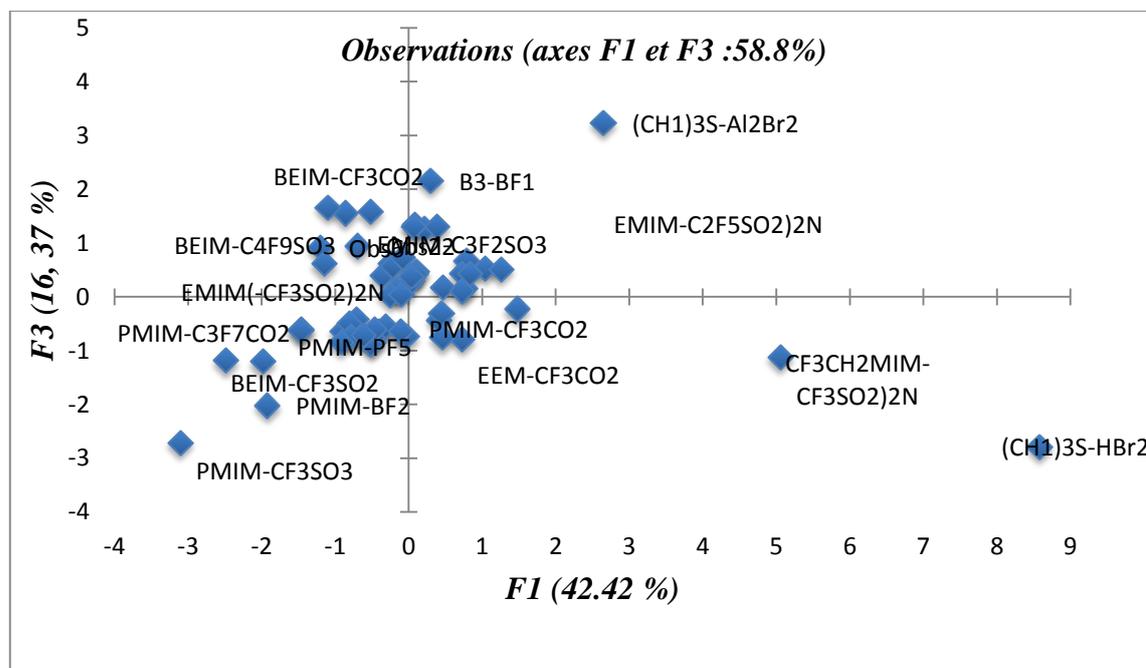


Figure-IV-5- représentation observations dans plan factoriel (F1, F3).

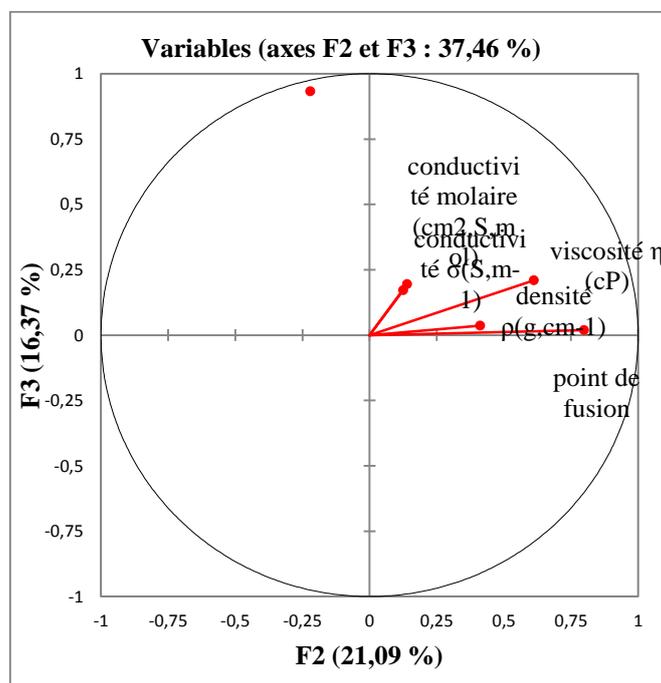


Figure- IV-6- le cercle de corrélation (plan factoriel F2, F3).

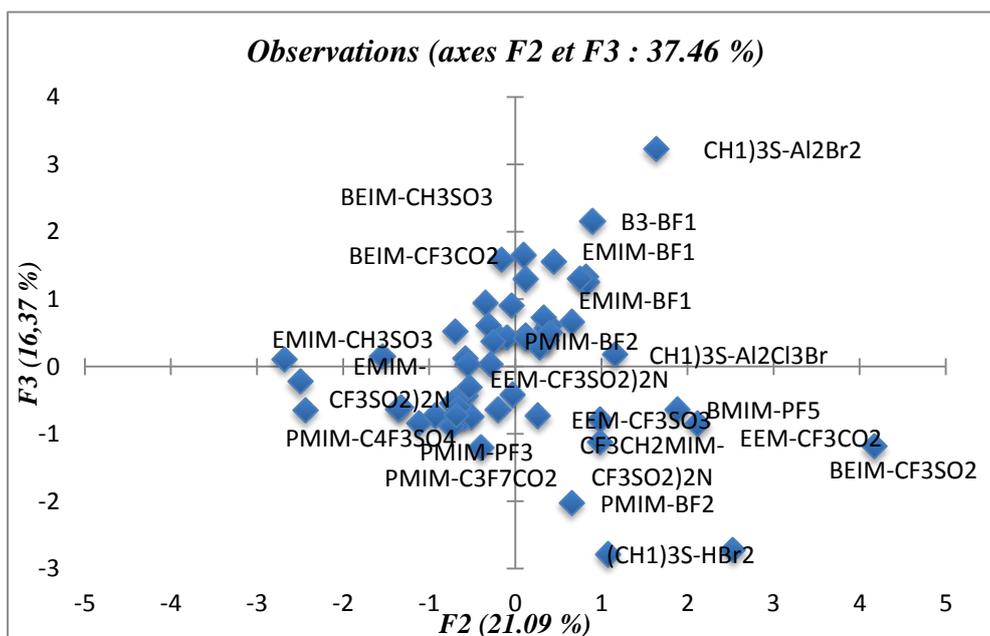


Figure-IV-7- représentation des observations dans plan factoriel (F2, F3).

On remarque dans ce cercle de corrélation le majeure partie des informations est expliquée par le premier axe factoriel entre les variables liées aux paramètres déterminé à eux seuls 37.46%. La projection de l'ensemble des paramètres sur le plan formé par les deux composants principaux (axe2, axe3) est illustrée par la figure IV-6. On remarque aussi sur ce cercle de corrélation que le point de fusion est bien représenté par rapport à l'axe F2. On constate que sur cet axe, le 1-butyl-2 ethylimidazolium trifluorométhyle sulfonyle (BEIM-CF₃SO₂), s'oppose au 1-ethyl-3-methylimidazolium sulfanate (EMIM-CH₃SO₃.)

On observe dans la figure IV-8 que la conductivité est très corrélée à 1ère composante, pour les liquides ioniques dont la position est proche de zéro la conductivité est faible et on remarque quand la taille de l'anion augmente la conductivité augmente par contre si la longueur de la chaîne alkylque du cation augmente la conductivité électrique diminue ce qui est accompagné d'une augmentation de la viscosité et du point de fusion.

On remarque que les liquides ioniques de viscosité et point de fusion élevés auront des conductivités très faibles sachant que la conductivité est liée à la mobilité des porteurs de charge.

La figure IV-8 (carte factorielle des propriétés physique-chimique des liquides ioniques) montre les trois groupes de liquides ioniques très importants par leur propriétés physiques et chimiques :

	Conductivité électrique-et molaire	Densité	Viscosité	Point de fusion
Groupe 1	+	+	-	-
Groupe 2	-	-	+	+
Groupe 3	-	-	-	-

Tableau IV-2 - classification des trois groupes principaux à base de données

Le 1^{er} groupe est le bon candidat pour une haute conductivité molaire et électrique et qui est caractérisé par une densité importante à la température d'étude. Les éléments appartenant à ce groupe sont, à titre d'exemple :

LI	conductivité σ (S, m ⁻¹)	conductivité molaire (cm ² , S, mol)	Densité ρ (g, cm ⁻¹)
(CH ₁) ₃ S-H ₂ Br ₃	5.6	8.41	1.79
(CH ₁) ₃ S-HBr ₂	3.4	4.62	1.74
(CH ₁) ₃ S-Al ₂ Cl ₃ Br	0.42	1.19	1.59
EMIM-CH ₃ CO ₂	0.93	1.67	1.38
EEMIM-CF ₃ SO ₂) ₂ N	0.62	1.82	1.43
EMIM-BF ₁	1.4	2.24	1.2
EMIM-BF ₄	1.1	2.24	1.24

Tableau IV-3- représentation le1^{er} groupe des liquides ioniques

Le 2^{ème} groupe est caractérisé par une forte viscosité et un point de fusion haut ce qui conduit à une faible conductivité et une faible densité comme c'est le cas de:

LI	viscosité η (cP)	point de fusion
PMIM-CF ₃ SO ₃	182	
CF ₃ CH ₂ MIM-(CF ₃ SO ₂) ₂ N	248	
BEIM-C ₄ F ₉ SO ₃	323	21
PMIM-PF ₃	312	
PMIM-C ₄ F ₃ SO ₄	373	16
MeoEtMIM-CF ₃ SO ₃	74	
B ₃ -(CF ₃ SO ₂) ₂ N	103	

Tableau IV-4-réprésentation le 2^{ème} groupe des liquides ioniques

Le 3^{ème} groupe représente des espèces à basse conductivité électrique et molaire et une faible viscosité. On cite dans cette catégorie:

LI	Conductivité σ (S,m-1)	viscosité η (cP)	Densité ρ (g,cm-1)
BEIM-BF ₄	0.023	-	-
HMIM-PF ₅	-	-	1.3
PMIM-PF ₅	0.05	-	-
PMIM-(CF ₃ SO ₂) ₂ N	0.39	52	1.42
DCIM-CF ₃ SO ₃	-	-	-

Tableau IV-5- représentation le 3^{ème} groupe des liquides ioniques

Conclusion

L'analyse des résultats graphiques montre que :

- L'application de l'ACP à une base de données de liquides ioniques en transformant les variables liées entre elles dites corrélées en statistique, en variables nommées composantes principales aux axes principaux, nous a permis de classer les liquides ioniques étudiés en trois groupes comme le montre le tableau IV-2
- Les conductivités électrique et molaire sont fortement liées à la taille de l'anion ; quand la taille de l'anion augmente la conductivité augmente. De même pour la densité des liquides ioniques.
- La viscosité augmente quand la chaîne alkyle du cation est importante. Ce-ci a un effet inverse sur la conductivité.

Conclusion générale

Conclusion générale

Dans ce travail, on propose une nouvelle technique d'estimation des propriétés physico-chimiques des liquides ioniques basée sur les techniques du data mining en particulier, la méthode d'analyse en composantes principale (ACP).

L'application de cette méthode permet de faire une classification de ces composés en mettant en évidence les paramètres importants telles que la conductivité électrique et molaire, viscosité et la densité et leur lien avec la structure chimique.

Pour avoir les résultats de cette estimation on a utilisé le logiciel XLSTAT.

Les conclusions qui peuvent être tirées de cette étude se résument à :

- L'application de l'ACP à une base de données de liquides ioniques nous a permis de classer les liquides ioniques étudiés en trois groupes différents, caractérisé chacun par des propriétés particulières (conductivité forte, viscosité faible) ; (conductivité faible, viscosité forte), (conductivité faible, densité faible)
- La conductivité électrique et molaire est fortement liée à la taille de l'anion ; quand la taille de l'anion augmente la conductivité augmente. De même pour la densité des liquides ioniques.
- La viscosité augmente quand la chaîne alkyle du cation est importante. Ceci a un effet inverse sur la conductivité.

Une analyse plus fine des résultats obtenus permettra certainement de trouver d'autres corrélations et prédire le comportement physico-chimique de liquides ioniques réels et hypothétiques qui pourront être ciblés dans le cadre de certaines applications technologiques.

I.1. Data mining

I.1.1. Introduction

Le datamining est une technique d'exploration de données, aussi connue sous les noms fouille de données, ou Extraction de Connaissances (*ECD* en français, de l'anglais *KDD*), a pour objet l'extraction d'un *savoir* ou d'une connaissance à partir de grandes quantités de données, par des méthodes automatiques ou semi-automatiques [1].

Il a vocation à être utilisé dans un environnement professionnel, et se distingue de l'analyse de données par des techniques de la statistique par les points suivants :

- Contrairement aux méthodes statistiques, le Datamining ne nécessite jamais que l'on établisse une hypothèse de départ qu'il s'agira de vérifier. C'est à partir des données elles-mêmes que se dégageront les corrélations intéressantes, l'algorithme n'étant là que pour les découvrir et non pour les établir. Il se situe à la croisée des statistiques, de l'intelligence artificielle.

- Les connaissances extraites par le Datamining ont vocation à être intégrées dans un schéma organisationnel. Le Datamining impose donc d'être capable d'utiliser de manière opérationnelle les résultats des analyses effectuées, souvent dans des délais très courts. Le processus d'analyse doit permettre à l'organisation une réactivité (très) importante.

- Les données traitées sont issues des systèmes de stockage en place dans une base de données et sont ainsi hétérogènes, multiples, plus ou moins structurées. Pour résumer ce sont des données dont la raison d'être n'est *a priori* pas l'analyse. Cela impose de disposer de systèmes performants de préparation ou de manipulation de données.

Le Datamining se propose donc de transformer en information, de grands volumes de données qui peuvent être stockées dans des bases de données relationnelles, mais qui peuvent aussi être récupérées de sources riches plus ou moins structurées comme Internet.

I.1.2. Définition

Le Datamining est l'ensemble des méthodes, et algorithmes destinés à l'exploitation et l'analyse de données, en vue de détecter dans ces données des règles, des associations, des tendances inconnues des structures particulières restituant de façon concise l'essentiel de l'information utile pour aider une décision, en d'autre terme c'est l'acquisition d'un nouveau savoir à partir d'un savoir déjà acquis.

I.1.3. Principe

Le Datamining est un processus d'analyse dont l'approche est différente de celle utilisée en *statistique*. Cette dernière présuppose en général que l'on se *fixe une hypothèse* que les données vont nous permettre ou non de confirmer. Au contraire, le *Datamining* adopte une démarche beaucoup plus empirique et essaye ainsi de faire *émerger*, à partir des données brutes, des hypothèses que l'expérimentateur peut ne pas soupçonner, et dont il aura à valider la pertinence.

Plus qu'une théorie normalisée, le Datamining est un processus d'extraction de connaissances en suivant les étapes principales suivantes :

- Formaliser un problème que l'on cherche à résoudre en terme de données
- Accéder aux données idoines quelles qu'elles soient
- Préparer les données en vue des traitements et utilisations futurs
- Modéliser les données en leur appliquant des algorithmes d'analyse
- Evaluer et valider les connaissances ainsi extraites des analyses

Bien sûr ce processus n'est pas linéaire, il se veut cyclique, le Datamining participant dès lors à une rationalisation avancée du stockage de l'information et des données. Autrement dit, le Datamining ne consiste pas en une succession d'études ad hoc mais doit bien permettre de capitaliser des connaissances acquises sous forme de connaissances explicites, donc de structurer les contenus nécessaires à l'ingénierie des connaissances.

Le Datamining se propose Alors d'utiliser un ensemble d'algorithmes issus de disciplines scientifiques diverses (statistiques, intelligence artificielle, base de données) pour construire des modèles à partir des données, c'est-à-dire trouver des schémas « intéressants » (des 'patterns' ou motifs de conception en français) selon des critères fixés au départ, et extraire de ces données un maximum de connaissances utiles.

I.1.4. Type de techniques de datamining

Résoudre une problématique avec un processus de Datamining impose généralement l'utilisation d'un grand nombre de méthodes et algorithmes différents. On peut distinguer deux grandes familles [2]:

I.1.4.1. Techniques descriptives

Elles permettent de travailler sur un ensemble de données dans lequel aucune des données n'a d'importance particulière par rapport aux autres, c'est-à-dire un ensemble de données dans lequel aucune variable n'est considérée comme la cible, elle est utilisée pour dégager d'un ensemble d'individus des groupes homogènes, afin de réaliser une compression d'informations (fig.I-1).

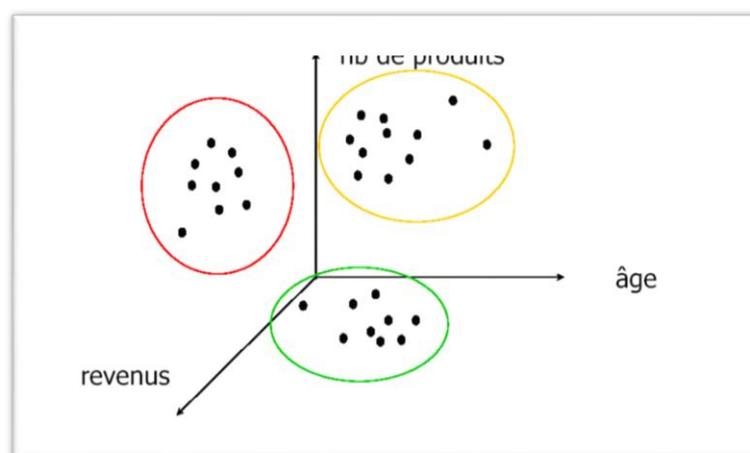


Fig. I-1- Exemple de technique non-supervisé (Classification)

On peut citer quelques techniques disponibles pour ces méthodes:

- L'Analyse en Composantes principales ;
- Techniques utilisées classiquement dans le monde des statistiques : classification ascendante hiérarchique, k-means et les nuées dynamiques (Recherche des plus proches voisins), les classifications mixtes (Birch...), les classifications relationnelles...

I.1.4.2. Techniques prédictives

Ces méthodes servent à expliquer et/ou de prévoir un ou plusieurs phénomènes observables et effectivement mesurés. Concrètement, elles vont s'intéresser à une ou plusieurs variables de la base de données définies comme étant les cibles de l'analyse. Dont voici la principale technique existante :

- Arbres de décision (Arbre de décision).
- Réseau de neurones ;
- Régressions linéaires et non linéaires au sens large :(Régression logistique binaire ou multinomiale, Analyse discriminante linéaire ou quadratique, modèle linéaire généralisé, régression PLS, régressions non paramétriques, équations structurelles..).
- Les algorithmes génétiques.
- Inférence bayésienne (Réseau bayésien).

I.1.5. Conclusion

Le datamining est une discipline née en dehors de la statistique, dans la communauté des bases de données et de l'intelligence artificielle dans le but de valoriser les bases de données, il offre des perspectives nouvelles et répond au défi du traitement des giga bases de données, il cherche à découvrir des structures inconnus et utiles de modèle issus de l'exploitation des données et non d'une théorie, à l'aide de processus de recherche algorithmique, linéaire ou non, explicite ou implicite (réseaux de neurones, Algorithmes génétiques, arbres de décision, réseaux Bayesiens,...) [3].

I.2 .Méthode d'Analyse en Composantes Principales (ACP)

I.2.1. Introduction

Les statistiques peuvent être vues en fonction de l'objectif fixé ; classiquement les méthodes statistiques sont employées soit pour explorer les données (nommée statistique exploratoire) soit pour prédire un comportement. La statistique exploratoire s'appuie sur des techniques descriptives et graphiques. Elle est généralement décrite par la statistique descriptive qui regroupe des méthodes exploratoires simples, uni- ou bidimensionnelle (moyenne, moments, quantiles, variance, corrélation, ...) et la statistique exploratoire multidimensionnelle. L'analyse de données s'inscrit dans ce cadre de la statistique exploratoire multidimensionnelle.

L'Analyse en Composante Principale (ACP) en anglais (*Principal Component Analysis PCA*) est une **technique descriptive** son utilisation remonte au début du siècle dernier pour l'exploitation des données ; Permettant d'étudier les relations qui existent entre les variables, sans tenir compte, à priori, d'une quelconque structure [4].

Le but de l'ACP est d'identifier la structure de dépendance entre des observations multi variables afin d'obtenir une description ou une représentation compacte de ces dernières (variables). Son utilisation a été restreinte à une projection des données sur différents axes factoriels.

L'ACP est une technique de projection orthogonale linéaire qui projette les observations multidimensionnelles représentées dans un sous-espace de dimension m (m étant le nombre de variables observées) dans un sous-espace de dimension inférieur ($l < m$) en maximisant la variance des projections.

L'ACP peut être considéré comme une technique de minimisation de l'erreur quadratique d'estimation ou une technique de maximisation de la variance des projections (ces deux critères sont équivalents).

L'ACP est utilisé comme un outil de modélisation des relations entre les différentes grandeurs représentant le comportement d'un processus quelconque.

L'estimation des paramètres du modèle *ACP* est effectuée par calcul des **valeurs** et **vecteurs** propres de la matrice de corrélation des données, cependant, pour la détermination de la structure du modèle, il faut déterminer le nombre de composante à retenir dans ce modèle, pour cette raison plusieurs critères de sélection du nombre de composante, dans notre travail le critère que nous adoptons est basé sur le principe de la variance de reconstitution, pour la sélection du nombre de composante à retenir dans le modèle *ACP* car il exploite les redondances entre les variables. Ce nombre permet d'identifier le modèle *ACP*.

Historique

La représentation brute de l'information conduit à des vecteurs de caractéristiques de grandes dimensions, ce qui peut poser des problèmes de complexité de calcul et de capacité de stockage. Ainsi, il est parfois préférable de chercher à réduire la dimensionnalité d'un problème de classification pour améliorer ses performances. Le mathématicien Bellman, père de la programmation dynamique, a introduit l'expression «malédiction de la dimensionnalité» (curse of dimensionality), pour signifier que représenter les formes par des vecteurs de taille importante est source de problème [5].

Il est évident que l'on se trouve d'emblée confronté à un problème de grande dimension. Plusieurs solutions sont préconisées dans la littérature pour réduire la taille de sa dimension Minami 1999[6]. Dans ce travail, nous avons utilisé l'analyse en

composantes principales. L'analyse en composantes principales (ACP) est une technique descriptive permettant d'étudier les relations qui existent entre les variables, sans tenir compte, a priori, d'une quelconque structure [7],[8]. Son utilisation pour l'exploitation des données remonte au début du siècle dernier. Elle est principalement issue des travaux de psychologues américains [9], [10], [11]. Le but de l'ACP est d'identifier la structure de dépendance entre des observations multivariées, afin d'obtenir une description ou une représentation compacte de ces dernières. L'idée de base de l'ACP est de réduire la dimension de la matrice des données, en retenant le plus possible les variations présentes dans le jeu des données de départ. Cette réduction ne sera possible que si les variables initiales ne sont pas indépendantes et ont des coefficients de corrélation entre elles non nuls. Ces variables initiales sont transformées en de nouvelles variables, appelées composantes principales. Elles sont obtenues par combinaisons linéaires des précédentes et sont ordonnées et non corrélées entre elles. L'analyse en composantes principales cherche à identifier les vecteurs propres orthonormaux et leurs valeurs propres correspondantes de la matrice de dispersion des variables originelles.

Les vecteurs propres orthonormaux sont utilisés pour construire les composantes principales et les valeurs propres sont les variances des composantes principales correspondantes [7].

I.2.2. Principe de l'Analyse en Composantes Principales

Dans le cadre de mémoire, nous aborderons l'ACP comme une technique de réduction

T de description des échantillons. La théorie sous-jacente à l'analyse en composantes principales est vaste, nous ne passerons donc en revue que les points les plus importants. Du point de vue géométrique l'ACP consiste à effectuer une certaine rotation du repère des variables autour de leurs valeurs moyennes.

Cette rotation transforme les n variables corrélées en l variables non corrélées. Notons que ce sont justement ces variables transformées que l'on a nommées les composantes principales. Supposons qu'on a un vecteur de données aléatoires $X_K = [X_1, \dots, X_m]^t \in \mathcal{R}^m$ de moyenne nulle $\varepsilon\{x(k)\} = 0$ et de matrice de covariance ou d'auto-corrélation :

$$\Sigma = \varepsilon\{xx^t\} \in \mathcal{R}^{m \times m} \quad (\text{I.1})$$

Avec $\varepsilon = \frac{1}{m-1}$

En analyse en composantes principales, un vecteur caractéristique $t \in \mathcal{R}^l$ est associé à chaque vecteur de données dont il optimise la représentation au sens de la minimisation de l'erreur d'estimation de \mathbf{x} ou la maximisation de la variance de \mathbf{t} . Les vecteurs \mathbf{t} et \mathbf{x} sont liés par une transformation linéaire $\mathbf{t} = \mathbf{P}^t \mathbf{x}$,

Où la matrice de transformation $\mathbf{P} \in \mathcal{R}^{m \times l}$ vérifie la condition d'orthogonalité $\mathbf{P}^t = \mathbf{I}_l$. Les colonnes de la matrice \mathbf{P} forment les vecteurs de base orthonormés d'un sous-espace \mathcal{R}^l de représentation réduite des données. La transformation linéaire s'apparente ainsi à une projection de l'espace des données de dimension \mathbf{m} vers un sous-espace orthogonal de dimension l .

Les composantes t_j , avec $j = 1, \dots, l$, du vecteur caractéristique \mathbf{t} représentent les composantes projetées du vecteur de données \mathbf{x} dans ce sous-espace.

Au sens de l'ACP, la projection \mathbf{P} est optimale si l'erreur quadratique d'estimation des vecteurs de données \mathbf{x} est minimale. Ce problème d'optimisation s'exprime par :

$$P_{opt} = \underset{P}{\operatorname{argmin}} J_e(P) \quad (\text{I.2})$$

Où J_e représente le critère d'erreur d'estimation de l'ACP. Sous la contrainte d'orthogonalité de la matrice de projection $\mathbf{P}^t \mathbf{P} = \mathbf{I}_l$, ce critère aura la forme :

$$\begin{aligned} J_e(\mathbf{P}) &= \mathcal{E} \left\{ \|\mathbf{X} - \widehat{\mathbf{X}}\|^2 \right\} \\ (\text{I.3}) \quad &= \mathcal{E} \{ \|\mathbf{X} - \mathbf{P}\mathbf{P}^t \mathbf{X}\|^2 \} \\ &= \mathcal{E} \{ (\mathbf{X} - \mathbf{P}\mathbf{t})^t (\mathbf{X} - \mathbf{P}\mathbf{t}) \} \\ &= \mathcal{E} \{ (\mathbf{X}^t \mathbf{X} - 2\mathbf{t}^t \mathbf{t} + \mathbf{t}^t \mathbf{P}\mathbf{P}^t \mathbf{t}) \} \\ &= \mathcal{E} \{ \operatorname{trace}(\mathbf{X}\mathbf{X}^t) - \mathbf{t}^t \mathbf{t} \} \end{aligned}$$

$$= \text{trace}(\Sigma) - \mathcal{E}\{t^t t\}$$

Notons que la trace d'une matrice carrée est définie par la somme de ces éléments diagonaux. Du fait que la matrice de covariance Σ est indépendante de la matrice des paramètres \mathbf{P} , minimiser J_e revient à maximiser le second terme J_v de son expression :

$$J_v(\mathbf{P}) = \mathcal{E}\{t^t t\} = \sum_{j=1} \mathcal{E}\{t_j^2\} \quad (\text{I.4})$$

Ainsi, la minimisation de l'erreur quadratique d'estimation de \mathbf{x} est équivalente à la maximisation de la variance des projections t_j des données. En conséquence :

$$\mathbf{P}_{opt} = \text{argmin}_e(\mathbf{P}) = \text{argmax}_v(\mathbf{P}) \quad (\text{I.5})$$

Le problème de l'ACP, considéré sous l'angle de la maximisation de la variance de projection des données, est celui de la détermination des vecteurs propres de la matrice de covariance Σ .

I.2.3. Identification du modèle ACP

Soit $\mathbf{x}_k = [\mathbf{x}_{k1}, \dots, \mathbf{x}_{kn}]^t$ un vecteur d'observations de n variables aléatoires.

On notera \bar{X} le vecteur correspondant à la moyenne des observations de \mathbf{x} et Σ la matrice de covariance de ces m observations. L'estimation des paramètres du modèle ACP se résume en une estimation des valeurs et vecteurs propres de la matrice de covariance Σ . Puisque la matrice de covariance Σ est carrée de dimension $n \times n$ et symétrique, on peut, selon l'algèbre linéaire, vérifier une relation de type :

$$P^t \Sigma P = L \quad (\text{I.6})$$

Où L est une matrice diagonale et P est une matrice orthonormale. Les éléments diagonaux de $L = \{l_1, \dots, l_n\}$ sont les valeurs propres de la matrice de covariance Σ alors que les vecteurs colonnes de P sont les vecteurs propres de Σ .

Les valeurs propres l_i peuvent être obtenues par la solution de l'équation caractéristique

$$\det[\Sigma - l_i I] = 0 \quad (\text{I.7})$$

Pour les n valeurs propres. Connaissant les valeurs propres l_i , on peut déduire les vecteurs propres de la matrice de covariance par la solution de

$$(\Sigma - l_i I)A_i = 0 \quad (\text{I.8})$$

Et puis,

$$u_i = \frac{A_i}{\sqrt{A_i^t A_i}} \quad (\text{I.9})$$

Pour les n vecteurs propres. Notons que l'équation (I.9) est aussi appelée équation de normalisation. Il est intéressant de constater qu'ici les éléments des vecteurs propres u_i sont les cosinus directeurs des axes générés par une rotation dans l'espace des observations.

Nous pouvons donc calculer tout un ensemble de valeurs à partir des ces nouveaux axes. En d'autres termes, on peut transformer les variables $\mathbf{x} = [x_1, \dots, x_n]$ en un nouvel ensemble de variables $\mathbf{y} = [y_1, \dots, y_n]$ par l'application de

$$y_i = \mathbf{u}_i^t (\mathbf{X} - \bar{\mathbf{X}}) \quad (\text{I.10})$$

Y_i est la $i^{\text{ème}}$ composante principale de l'ensemble des observations. Les nouvelles variables y_i ont évidemment une moyenne nulle et elles possèdent une variance donnée par l_i (valeur propre de Σ). Les vecteurs de colonnes \mathbf{u}_i de \mathbf{P} sont ortho normaux car ils sont orthogonaux et de modules unitaires. Donc, si l'on utilise ces vecteurs pour obtenir les composantes principales, tel que montré par l'équation (I.10), on obtiendra des composantes principales qui ne seront pas corrélées et de variances égales aux valeurs propres de la matrice de covariance Σ .

En appliquant une transformation sur les vecteurs colonnes \mathbf{u}_i , on peut obtenir de nouvelles propriétés intéressantes. L'une de ces transformations consiste à effectuer une mise à l'échelle de \mathbf{u}_i par les valeurs propres de la matrice de covariance Σ :

$$\mathbf{V}_i = \mathbf{u}_i \sqrt{l_i} \quad (\text{I.11})$$

Les composantes obtenues par l'utilisation de $\mathbf{V} = [v_1, \dots, v_n]^t$ seront de même unité que les variables originales. Nous retrouvons également une transformation dans laquelle les vecteurs u_i sont divisés par les valeurs propres de la matrice de covariance S :

$$W_i = \frac{u_i}{\sqrt{l_i}} \quad (\text{I.12})$$

Dans ce cas, les composantes obtenues seront de variance unitaire. Enfin, nous soulignons que l'on peut avoir une mesure de variabilité expliquée par chacune des composantes principales en utilisant les vecteurs transformés de l'équation (III.11) selon

$$var = v_i v_i^t \text{ avec } i=1, \dots, n. \quad (\text{I.13})$$

Les valeurs obtenues par l'équation (I.13) peuvent être utilisées pour indiquer le pourcentage de la variabilité des caractéristiques représentées par chacune des composantes principales.

I.2.4. Détermination de la structure du modèle ACP

L'analyse en composantes principales recherche une approximation de la matrice initiale des données X par une matrice de rang inférieur issue d'une décomposition en valeurs singulières.

La question qui se pose alors, et qui a été largement débattue dans la littérature, concerne le choix du nombre de composantes principales qui doit être retenu. De nombreuses règles sont proposées pour déterminer le nombre de composantes à retenir [8, 9, 10, 11]. La plupart de ces règles sont heuristiques et donnent un nombre de composants subjectif. Toutefois, dans le cadre de l'application de l'ACP à la réduction des données, le nombre de composantes a un impact significatif sur la phase finale de classification. Si peu de composantes sont utilisées, on risque de perdre des informations contenues dans les données de départ en projetant certains échantillons dans le sous-espace des résidus et donc avoir des erreurs de modélisation, ce qui provoque une mauvaise caractérisation.

Si par contre beaucoup de composantes sont utilisées, il y a risque d'avoir des composantes retenues (les composantes correspondantes aux valeurs propres les plus faibles parmi celles retenues dans le modèle) qui sont porteuses de redondance, ce qui est indésirable.

I.2.4.1. Pourcentage cumulé de la variance totale (PCV)

L'idée de base de cette approche est que chaque composante principale est représentative d'une portion de la variance des mesures de l'information étudiée. Les valeurs propres de la matrice de corrélation sont des mesures de cette variance et peuvent donc être utilisées dans la sélection du nombre de composantes principales. Pour faire ce choix, il faut estimer le pourcentage de la variance totale qu'on veut conserver. Le nombre de composantes est alors le plus petit nombre pris de telle sorte que ce pourcentage soit atteint ou dépassé; les composantes sont choisies successivement dans l'ordre des variances décroissantes. Le pourcentage de variance expliquée par les premières composantes est donné par :

$$PCV(l) = 100 \left(\frac{\sum_{j=1}^l \lambda_j}{\sum_{j=1}^m \lambda_j} \right) \% \quad (\text{I.14})$$

La variance du bruit étant inconnue a priori, la décision basée seulement sur le pourcentage de la variance expliquée est un peu arbitraire. Sa capacité à fournir le nombre correct de composantes principales dépendra fortement du rapport signal sur bruit.

I.2.4.2. Moyenne des valeurs propres

Cette règle consiste à ne prendre en considération que les composantes pour lesquelles la valeur propre est supérieure à la moyenne arithmétique de toutes les valeurs propres. En particulier, si on travaille sur les données centrées réduites, cela revient à négliger les composantes dont la variance est inférieure à l'unité

$$\left(\frac{1}{m} \text{trace}(\Sigma) = 1 \right) \quad (\text{I.15})$$

Dans le cas du modèle de calcul à partir de la matrice de covariance, la moyenne arithmétique des valeurs propres est donnée par $\frac{1}{m} \text{trace}(\Sigma)$

I.2.5. Conclusion

Ce que nous pouvons retenir de cette méthode est que l'ACP est une technique utile pour la compression et la classification des données. Le problème consiste à réduire la dimensionnalité d'un ensemble des données (échantillon) en trouvant un nouvel ensemble de variables plus petit que l'ensemble originel des variables, qui néanmoins contient la plupart de l'information sur l'échantillon. Par information nous voulons parler de la variation présente dans l'échantillon et donnée par les corrélations entre les variables originelles. Les nouvelles variables, appelées composantes principales (PC), sont non corrélées, et sont ordonnées par fraction de l'information totale que chacune contient [7,8].

En termes précis, l'usage principal de l'ACP est la réduction de la dimensionnalité tout en conservant le plus d'information possible.

1ière PC: contient la plus grande quantité de variation.

2nde PC: contient en deuxième lieu la plus grande quantité de variation.

Quasiment toutes les technique de datamining restent en suspend et ces méthodes ont encore beaucoup de progrès à faire.

Cependant, il est clair à la lumière des résultats incontestablement utiles de l'utilisation de ces techniques dans divers domaines de la recherche, qu'elles offrent un moyen puissant pour l'analyse dont on ne doit pas manquer l'utilisation en science des matériaux pour profiter de leurs prérogatives tout en demeurant prudents quant aux interprétations des résultats.

Références

- [1] D. T. Larose ET J. Wiley, *Data mining methods and models*. Wiley Online Library, 2006.
- [2] S. Tufféry, « DATA MINING-SCORING », *Bases de données et gestion de la relation client*, 2002.
- [3] G. Saporta, « Liaisons entre plusieurs ensembles de variables et codage de données qualitatives », 1975.
- [4] M. F. Haraket, *Détection et Localisation de Défauts par Analyse en Composant Principales*. Doctorat de l'Institut National Polytechnique de Lorraine, 2003.
- [5] R. E. Bellman, R. E. Bellman, R. E. Bellman, et R. E. Bellman, *Adaptive control processes: A guided tour*. Princeton University Press, 1966.
- [6] K. Minami, H. Nakajima, et T. Toyoshima, « Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network », *Biomedical Engineering, IEEE Transactions on*, vol. 46, n° 2, p. 179–185, 1999.
- [7] I. T. Jolliffe et MyiLibrary, *Principal component analysis*, vol. 2. Wiley Online Library, 2002.
- [8] K. I. Diamantaras et S. Y. Kung, *Principal component neural networks: theory and applications*. Wiley, 1996.
- [9] K. P. FRS, « LIII. On lines and planes of closest fit to systems of points in space », *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, n° 11, p. 559–572, 1901.
- [10] C. Spearman, « “ General Intelligence,” Objectively Determined and Measured », *The American Journal of Psychology*, vol. 15, n° 2, p. 201–292, 1904.
- [11] H. Hotelling, « Analysis of a complex of statistical variables into principal components. », *Journal of educational psychology*, vol. 24, n° 6, p. 417, 1933.

Résumé

Cette étude l'application de méthode d'analyse en composantes principale (ACP) sur une base de données de liquides ioniques afin de faire une classification de ces composés en mettant en évidence les paramètres importants telles que la conductivité électrique et molaire, viscosité et la densité et leur lien avec la structure chimique.

Les résultats obtenus permettent de retenir les points suivantes :

- L'application de l'ACP nous a permis de classer les liquides ioniques étudiés en trois groupes différents, caractérisé chacun par des propriétés particulières (conductivité forte, viscosité faible) ; (conductivité faible, viscosité forte), (conductivité faible, densité faible)
- Les conductivités électrique et molaire sont fortement liée à la taille de l'anion ; quand la taille de l'anion augmente la conductivité augmente. De même pour la densité des liquides ioniques.
- La viscosité augmente quand la chaîne alkyle du cation est importante. Ceci a un effet inverse sur la conductivité.

Mot de clé : base de données, liquide ionique, data mining, ACP, XLSTAT