

APPLICATION DE L'ACP À UNE BASE DE DONNÉES DE LIQUIDES IONIQUES

MOHAMED CHERIF Yamina¹; BELARBI El Habib²; DEBDAB Mansour³;
ARBAOUI Khaldia⁴, AOUINE Rachida⁵

Laboratoire Synthèse et catalyse , Univ Tiaret BP78 RP 14000 Tiaret Algeria,

¹ nouraliman.m.h@gmail.com

² belarbi.hb@gmail.com

³ m.debdab@hotmail.fr

⁴ khalida.arbaoui@yahoo.fr

⁵ lokatchikifafouch@gmal.com

RESUME

L'objectif de cette étude est de mettre en évidence le facteur le plus influant sur la conductivité électrique d'un liquide ionique en appliquant l'analyse en composantes principales (ACP) . Pour atteindre cet objectif nous avons utilisé le logiciel XLSTAT que nous avons appliqué à une base de données de 500 liquides ioniques différents. Pour chaque liquide ionique, une dizaine de propriétés physico-chimiques sont répertoriées. Il ressort de cette étude que les liquides ioniques présentant une forte conductivité électrique sont caractérisés par une faible température de transition vitreuse, une forte température de cristallisation et un faible indice de réfraction. La conductivité d'un liquide ionique augmente avec la taille de l'anion et diminue avec l'augmentation de la taille du cation.

Mots clés : *liquides ioniques ; Data mining ; ACP, base des données ; XLSTAT.*

NOMENCLATURE

Symbols :

Td température de décomposition, K

Tc température de cristallisation, K

Tg température de transition vitreuse, K

Tf température de fusion, °C

F: largeur de la fenetre électrochimique, V

Cathod : limite du potentiel cathodique, V

Anod: limite du potentiel anodique, V

M: masse molaire, g

n indice de réfraction

Lettres grecques:

ρ masse volumique, g cm⁻³

σ conductivité électrique, S·m⁻¹

η viscosité, cp

1. INTRODUCTION

Les liquides ioniques qui sont des sels fondus à température ambiante et stables à l'air, présentent un intérêt technologique dans divers domaines^[1]. L'existence d'énormes possibilités de synthétiser de nouveaux liquides ioniques rend toute tentative expérimentale pour explorer ces nouvelles structures une tâche très fastidieuse et budgétivore, voire impossible. Le recours aux techniques d'exploration des données (data mining) s'avère indispensable plus que jamais. Dans ce contexte, plusieurs techniques ont été proposées et dont l'objectif est de trouver des règles, des tendances inconnues ou un nouveau savoir à partir d'un savoir déjà acquis^[2]. 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 d'observer certaines associations telles que l'influence de la température de transition vitreuse et la température de cristallisation, la fenêtre électrochimique et l'indice de réfraction sur la conductivité électrique des liquides ioniques.

2. METHODOLOGIE

2.1 Construction de la base de données :

Une base de données est un ensemble structuré et organisé permettant le stockage de grandes quantités d'informations afin d'en faciliter l'exploitation^[2,3].

Dans le cadre de cette étude, nous avons construit une base de données des propriétés physico-chimiques de 500 liquides ioniques, qui ont été collectées à partir de plusieurs sources documentaires.

La base des données n'étant pas complète et présente des données manquantes. Pour combler ces lacunes, l'une des méthodes suivantes est adoptée. remplacement par la moyenne.; remplacement aléatoire, utilisation de la régression. Dans notre cas c'est cette dernière méthode qui a été utilisée et qui consiste à estimer la valeur possible en utilisant les données des enregistrements complets et toutes les informations disponibles. C'est la méthode la plus précise; et la plus complexe à implanter.

2.2 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. C'est une méthode descriptive qui permet d'effectuer une représentation simplifiée d'une série de variables intercorrélées. Cette technique utilise comme principe une transformation des variables initiales en de nouvelles variables non corrélées. Ces nouvelles variables sont appelées les Composantes Principales. Chaque composante principale est une combinaison linéaire des variables initiales.

L'analyse en composantes principales est utilisée pour réduire la dimension (le nombre de variables) d'un problème. Cette diminution du nombre de variables doit s'effectuer en perdant un minimum d'informations. Le but de l'Analyse en Composantes Principales (ACP) est de condenser les données originelles en de nouveaux groupements, appelées nouvelles composantes. L'ACP permet d'explorer les liaisons entre variables et les ressemblances entre individus.

Elle permet aussi:

- De visualiser et analyser rapidement les corrélations entre les N variables,
- De 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,
- De 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. Pour appliquer cette méthode à notre base de données, nous avons utilisé le logiciel XLSTAT^[4,5].

3. RESULTATS

L'analyse en composante principale permet 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. L'analyse en composantes principales en utilisant le logiciel XLSTAT nous a permis de définir les axes principaux. Pour un plan factoriel donné, on regarde le pourcentage de l'information du nuage initial retranscrite sur le plan factoriel. Ainsi, un axe expliquant moins de 10% sera rarement intéressant ^[7], les axes sont rangés dans l'ordre décroissant, de telle manière que le premier plan factoriel constitué par les deux premiers axes factoriels soit toujours celui qui est le plus riche en renseignements sur les propriétés du nuage étudié. Dans notre exemple, les axes F1 (PC1) et F2 (PC2) sont les plus riches en information figure 02.

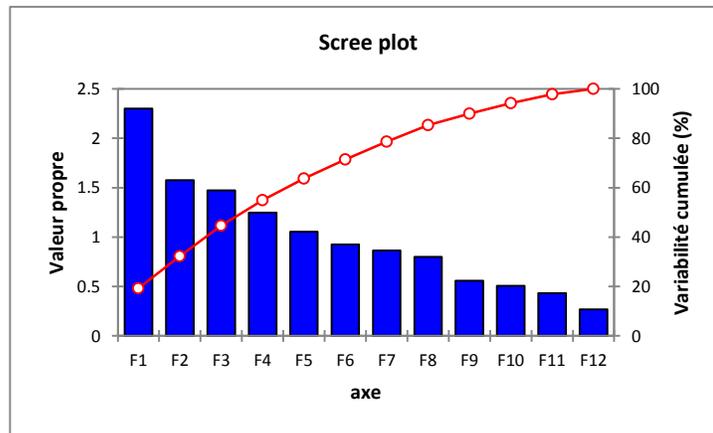


Figure 02 : Histogramme des valeurs propres. (Pourcentages de variance observée, pour les 12 axes principaux).

Le premier résultat intéressant à identifier suite à une analyse en composantes principales est la matrice des corrélations qui permet d'analyser les relations bilatérales existant entre les différentes variables retenues. Cette matrice est représentée sur le tableau 01.

Variables	Tc	Tf	Tg	Tm	σ	η	ρ	M	n	f	cathod	Anod
Tc	1	0,281	-0,003	0,142	0,102	0,068	0,286	0,043	0,005	-0,014	-0,007	-0,076
Tf	0,281	1	0,038	0,150	-0,157	0,046	0,066	-0,034	0,017	-0,038	0,028	-0,010
Tg	-0,003	0,038	1	0,115	-0,036	0,162	-0,250	0,330	0,007	0,036	0,005	-0,004
Tm	0,142	0,150	0,115	1	0,023	0,134	0,069	-0,033	0,046	-0,086	0,042	-0,080

σ	0,102	-0,157	-0,036	0,023	1	0,002	0,046	-0,113	-0,103	0,065	-0,066	-0,047
η	0,068	0,046	0,162	0,134	0,002	1	0,025	-0,076	-0,022	-0,011	0,009	-0,193
ρ	0,286	0,066	-0,250	0,069	0,046	0,025	1	-0,033	0,052	-0,011	-0,031	0,035
M	0,043	-0,034	0,330	-0,033	-0,113	-0,076	-0,033	1	0,018	0,042	-0,019	0,030
n	0,005	0,017	0,007	0,046	-0,103	-0,022	0,052	0,018	1	-0,614	0,562	0,014
f	-0,014	-0,038	0,036	-0,086	0,065	-0,011	-0,011	0,042	-0,614	1	-0,703	0,196
cathod	-0,007	0,028	0,005	0,042	-0,066	0,009	-0,031	-0,019	0,562	-0,703	1	-0,052
anod	-0,076	-0,010	-0,004	-0,080	-0,047	-0,193	0,035	0,030	0,014	0,196	-0,052	1

Tableau 01 : Matrice de corrélation (Pearson)

La diagonalisation de la matrice de corrélation donne les valeurs propres présentées dans le tableau 02. Chaque facteur est caractérisé par une valeur propre, une part de la variance totale (% de variabilité), et une part de la variance cumulée (% cumulée).

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Valeur propre	2,300	1,576	1,473	1,248	1,053	0,926	0,864	0,801	0,557	0,506	0,432	0,266
Variabilité (%)	19,165	13,131	12,276	10,398	8,771	7,713	7,199	6,675	4,640	4,216	3,598	2,218
% cumulé	19,165	32,296	44,572	54,970	63,741	71,454	78,652	85,327	89,967	94,183	97,782	100,000

Tableau 02 : les valeurs propre des axes principaux.

Ce tableau signifie que le premier axe (F1) permet d'expliquer (19,17%) de la variance totale du nuage de points, que le second axe (F2) permet d'expliquer (13,13%) de la variance totale. En projetant donc chaque individu sur un plan (F1, F2) on conserve 19,17+13,13 soit 32,30% de la variance totale.

Remarquons ici que la variance expliquée par l'axe F3 est faible (12,28 %) et F4 (10,40%), et remarquons aussi que l'essentiel de l'information est fourni par les 2 premiers axes, puisque la variance expliquée cumulée par est de 32,30%. Si les variables sont très corrélées entre elles, les premières valeurs propres sont importantes alors que si le lien linéaire entre les variables d'origines est faible, les pourcentages de variance expliquée par les axes seront faibles eux aussi.

La figure 02 nous permet d'évaluer les valeurs propres de manière schématique, La première composante principale (F1) correspond à la plus grande valeur propre, ainsi la deuxième composante (F2) correspond à la deuxième plus grande valeur propre. Généralement, l'étude se restreint à ces deux premiers axes, du fait qu'ils contiennent la majorité de la variance, le pourcentage de variance absorbée par les axes principaux est indiqué sur la figure (02).

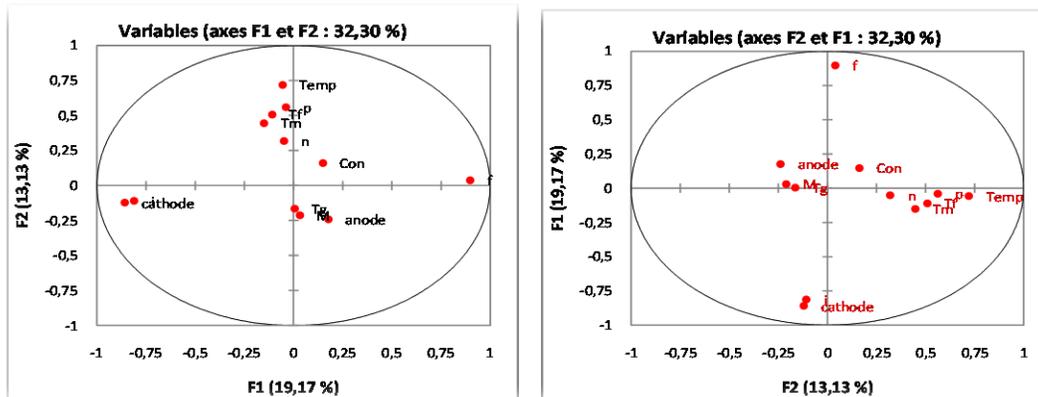


Figure 03-a : représentation des variables pour les liquide ioniques correspondant aux observations sur le plan factoriel (F1,F2).

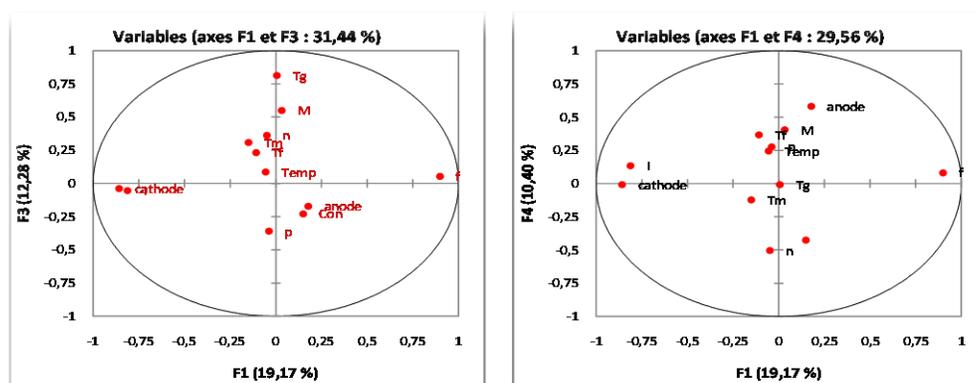


Figure 03-b : représentation des variables pour les liquide ioniques correspondant aux observations sur les plans factoriels (F1,F3) et (F1,F4).

A partir du tableau (02) et du graphique (02) , on remarque que ces valeurs sont décroissantes de F1 (2,3) à F9(0,3) et qu’à partir de F6 elles sont inférieures à 1.

L’analyse des résultats de l’ACP pour les liquides ioniques, est donnée par les graphes des observations "score plot" et des variables "loading plot". Le résultat des "Score plot" de cette analyse est montré sur les figures 03.

Sur la figure 03-a On peut constater que l’axe F1 est fortement corrélé positivement à la fenêtre électrochimique, aussi il est fortement corrélé négativement à limite cathodique et l’indice de réfraction. L’axe F2 est fortement corrélé négativement à la température de cristallisation et à la température de fusion et à la densité de liquides ioniques.

Pour confirmer le fait qu’une variable est fortement liée à une composante principale, il suffit de consulter le tableau 03 suivant des cosinus carrés.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Tc	0,003	0,518	0,008	0,060	0,077	0,011	0,043	0,038	0,239	0,002	0,000	0,000
Tf	0,012	0,258	0,054	0,135	0,201	0,007	0,136	0,046	0,148	0,002	0,001	0,000
Tg	0,000	0,027	0,662	0,000	0,038	0,024	0,000	0,031	0,000	0,210	0,007	0,000
Tm	0,023	0,199	0,096	0,015	0,002	0,312	0,019	0,323	0,002	0,009	0,000	0,000
σ	0,022	0,026	0,052	0,180	0,418	0,109	0,067	0,049	0,069	0,006	0,002	0,000
n	0,002	0,102	0,132	0,253	0,028	0,017	0,268	0,144	0,001	0,051	0,001	0,000
ρ	0,002	0,315	0,129	0,077	0,069	0,049	0,198	0,014	0,065	0,071	0,012	0,000
M	0,001	0,045	0,304	0,164	0,206	0,096	0,001	0,034	0,028	0,120	0,001	0,000
I	0,660	0,012	0,003	0,018	0,009	0,003	0,014	0,005	0,000	0,007	0,257	0,012
F	0,807	0,002	0,003	0,007	0,002	0,001	0,005	0,001	0,000	0,000	0,013	0,160
Cathod	0,737	0,015	0,001	0,000	0,002	0,003	0,001	0,012	0,000	0,009	0,135	0,085
Anod	0,031	0,058	0,029	0,339	0,001	0,293	0,112	0,104	0,004	0,017	0,004	0,009

Tableau 03 : Cosinus carrés des variables (propriétés physico-chimique)

Une variable est plus proche de l’axe si le cosinus carrés est élevé. Dans notre cas l’axe F1 est lié à n, f ,cathode les cosinus carrés respectifs n(0,66) ,f(0,81) ,cathod(0,74) , l’axe F2 est lié à Tc, Tf, p avec des cosinus carrés respectifs Tc (0,52), Tf (0,26) ,ρ(0,32), l’axe F3 est lié à Tg, M avec des cosinus carrés respectifs Tg(0,66), M(0,30) et l’axe F4 est lié à la limite anodique avec cosinus carrés anod(0,34) .

On peut remarquer que la conductivité d'une part, la masse molaire des liquides ioniques, l'indice de réfraction, la température de fusion et la température de transition vitreuse d'autre part sont inversement corrélés, ce qui nous permet de dire que les liquide ionique les plus visqueux ou de grande masse molaire ou d'indice de réfraction très important ou de température de transition vitreuse élevé et point de congélation élevé sont de faible conductivité électrique.

La température de cristallisation et la conductivité se trouvent dans le même coté cela signifie qu'ils sont fortement corrélés entre eux. Les liquide ionique de grande température de cristallisation seront les meilleurs conducteurs.

A première vue, et d'après l'analyse des résultats obtenus, il s'avère que l'ensemble des liquides ioniques se séparent nettement en quatre groupes (clusters). Le cluster 1 regroupe les liquides ioniques qui ont une forte conductivité électrique, la valeur de limite anodique élevée, de masse molaire faible et ont de température de transition vitreuse très faible. Le cluster 2 correspond à liquides ioniques plus denses et visqueux, de température de fusion très importante et de faible conductivité. Pour les liquides ioniques qui sont regroupé par le cluster 3, ils sont caractérisés par une température de transition vitreuse élevée, une densité élevée et une faible conductivité électrique. Le cluster 4 regroupe les liquides ioniques qui ont une valeur de limite cathodique très importante et indice de réfraction élevé.

CONCLUSION

D'après cette étude on peut conclure que la conductivité des liquides ioniques est liée à la variation de la taille des anions et les cations. Lorsque la taille de l'anion augmente la conductivité augmente par contre elle diminue par l'augmentation de la taille de cation.

On constate aussi que les liquides ioniques qui ont une forte conductivité parmi les 500 liquides ioniques étudiés sont caractérisés par une faible température de fusion, ils sont moins visqueux et possèdent une faible température de transition vitreuse.

REFERENCES

- [1] R. D. Rogers et K. R. Seddon, « Ionic liquids—solvents of the future? », *Science*, vol. 302, n° 5646, p. 792–793, 2003.
- [2] F. Gorunescu, *Data Mining: Concepts, models and techniques*, vol. 12. Springer Science & Business Media, 2011.
- [3] M. Ikhlef, « Conception et mise en place d'une base de données pour le service du récolement des dépôts au Musée du Louvre », 2011.
- [4] R. Féraud, M. Boullé, F. Clérot, et F. Fessant, « Vers l'exploitation de grandes masses de données. », in *EGC*, p. 241–252, 2008.
- [5] G. Grégoire, F.-X. Jollois, J.-F. Petiot, A. Qannari, S. Sabourin, P. Swertwaegher, J.-C. Turlot, V. Vandewalle, et S. Viguier-Pla, « Les logiciels et l'enseignement de la statistique dans les départements“ Statistique et Informatique Décisionnelle”(STID) des IUT », *Statistique et Enseignement*, vol. 2, n° 2, p. 5–24, 2012.