

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE



UNIVERSITE ABOU BEKR BELKAÏD DE TLEMCEM  
FACULTE DES SCIENCES  
DEPARTEMENT DE PHYSIQUE



DIVISION ETUDE ET PREDICTION DES MATERIAUX  
UNITE DE RECHERCHE MATERIAUX ET ENERGIES  
RENOUVELABLES (URMER)

# THÈSE

Pour obtenir le grade de

**Docteur en physique**

Spécialité : Physique et chimie de la matière condensée

**Présentée par :**

**M<sup>elle</sup>. Nadéra Settouti**

Sur le Thème

**Le réductionnisme en chimie quantique et les grandes contradictions du tableau périodique avec la mécanique quantique et le rôle du data mining**

Soutenue le : / / 2015 devant le jury composé de:

|                    |                                |                 |
|--------------------|--------------------------------|-----------------|
| Président          | <b>Pr. Ghouti MERAD</b>        | (Univ. Tlemcen) |
| Directeur de Thèse | <b>Pr. Hafid AOURAG</b>        | (Univ. Tlemcen) |
| Examinateur        | <b>Pr. Houda Imane FARAOUN</b> | (Univ. Tlemcen) |
| Examinateur        | <b>Pr. Tahar SAHRAOUI</b>      | (Univ. Blida)   |
| Examinateur        | <b>Pr. Mohammed HADJI</b>      | (Univ. Blida)   |
| Examinateur        | <b>Pr. Lahcene Larabi</b>      | (Univ. Tlemcen) |

---

## Remerciements

---

*En tout premier lieu, je remercie ALLAH le Tout-puissant de m'avoir donné la santé et tout dont je necessitais pour l'accomplissement de cette thèse, et de m'avoir aidé à dépasser toutes les difficultés.*

*Je tiens à remercier mon directeur de thèse, **Monsieur le Professeur Hafid AOURAG** Professeur à l'université de Tlemcen et directeur général de la recherche scientifique et du développement technologique pour m'avoir donné l'opportunité de travailler dans son équipe en me proposant le sujet, ainsi pour ses conseils qui m'ont été d'une grande utilité et clairvoyance dans la rédaction de cette thèse. En plus ce travail m'a permis de prendre contact avec le monde de la recherche, qui a été pour moi extrêmement enrichissant sur bien des plans.*

*Je remercie aussi **Mr le Professeur Ghouti MERAD** professeur à l'université de Tlemcen et Directeur du Laboratoire d'Etude et Prédiction des matériaux, de m'avoir accepté au sein de son laboratoire, qui nous a initiés à la recherche, on a bénéficié de son savoir et de son savoir-faire au cours de nombreuses années, et qui m'a fait l'honneur de présider ce jury.*

*Mes remerciements les plus sincères vont à l'ensemble des membres du jury ; tout d'abord à **Melle Houda Imane FARAOUN** Professeur à l'Université de Tlemcen, **Mr Tahar SAHRAOUI** Professeur à l'Université de Blida, et à **Mr Mohammed HADJI** Professeur à l'Université de Blida, et à **Mr Lahcene Larabi** Professeur à l'Université de Tlemcen d'avoir accepté d'examiner ce travail et de faire partie de ce jury.*

*J'adresse aussi mes remerciements à tous mes collègues de l'URMER.*

*Mes plus grand remerciements vont naturellement aussi à mes très chers parents, mes chers frères et sœurs, qui m'ont toujours encouragé et soutenu en toutes circonstances.*

*A mon père et ma mère  
la lumière de ma vie  
qui m'ont appris dès mon très jeune âge  
la foi et le respect de la science  
A mes sœurs et mes frères qui m'ont  
toujours encouragé et soutenu  
A mes très chers neveux et nièces  
Je dédie ce modeste travail*

*« La connaissance pratique, c'est d'acquérir un contenu de pensée relative à l'agir. » (Avicenne, Ibn Sina)*

*« Une théorie qui ne serait pas réfutable n'appartient pas au domaine scientifique » Karl Popper*

*« Ce qu'il faut, c'est voir. Et voir c'est chercher à voir, derrière toutes les images, toutes les représentations, derrière toutes les visions, ce qui ne se laisse pas voir, mais qui se laisse chercher. » (Philippe Quéau)*

*« Aucune loi de la nature n'a d'importance scientifique, à moins qu'elle ne soulève des questions auxquelles l'expérience puisse répondre ».*

*Dimitri I. Mendeleiev*

*« Personne ne croit une nouvelle théorie sauf le théoricien qui l'a développée. Tout le monde croit des résultats expérimentaux sauf l'expérimentateur qui les a obtenus. »*

*J.C.H Spence (1992)*

# Table des matières

|   |    |
|---|----|
| <b>Introduction générale</b>  | 01 |
| <b>Chapitre I : Le réductionnisme en chimie quantique</b>                   |    |
| I.1. Introduction .....   | 05 |
| I.2. Définition du réductionnisme.....                                      | 05 |
| I.3. Les différentes formes de réductionnisme.....                          | 05 |
| I.4. Réductionnisme et complexité .....                                     | 06 |
| I.5. Aspects du réductionnisme .....  | 07 |
| I.6. Le réductionnisme, une démarche incontournable.....                    | 07 |
| I.7. L'idée de réductionnisme.....  | 08 |
| I.8. Les types de réduction.....  | 09 |
| I.8.1. Réduction verticale.....   | 09 |
| I.8.2. Réduction horizontale.....   | 10 |
| I. 9. Comparaison entre la physique et la chimie.....                       | 10 |
| I. 10. Le monde quantique.....  | 11 |
| I.10.1. La réduction de la chimie à la mécanique quantique.....             | 12 |
| I.10.2. Le tableau périodique et le réductionnisme en chimie quantique..... | 14 |
| I. 11. Conclusion.....  | 15 |
| I. 12. Références.....  | 18 |
| <b>Chapitre II: Le Tableau périodique</b>                                   |    |
| II. 1. Introduction.....  | 20 |
| II. 2. Historique .....   | 20 |
| II.2.1. Lothar Meyer ou Dimitri Mendeleïev.....                             | 20 |
| II.2.2. Le tableau périodique des éléments.....                             | 21 |
| II.2.3. Quantification de l'énergie.....                                    | 23 |
| II.2.4. Niveaux énergétiques des orbitales atomiques.....                   | 24 |
| II.2.5. La configuration électronique d'un atome.....                       | 24 |
| II.2.5.1. Principe d'exclusion de Pauli.....                                | 25 |
| II.2.5.2. La règle de Klechkowski.....                                      | 25 |
| II.2.5.3. Règle de Hund.....  | 26 |
| II. 3. Principe de la classification moderne.....                           | 26 |
| II.3.1. La loi de la périodicité des propriétés.....                        | 27 |

|  |    |
|--|----|
| II.3.2. La loi périodique.....   | 28 |
| II.3.3. La compréhension du système périodique.....                          | 28 |
| II.4. Analyse du tableau périodique.....                                     | 29 |
| II.4.1. Les anomalies du tableau périodique.....                             | 29 |
| II. 4.2. Propriétés physiques et chimiques des éléments.....                 | 30 |
| II.4.2.1. L'électronégativité.....   | 30 |
| II.4.2.2. Le rayon atomique .....  | 30 |
| II.4.2.3. L'énergie (ou potentiel) d'ionisation .....                        | 31 |
| II.4.2.4. L'affinité électronique.....                                       | 31 |
| II.5. Conclusion.....  | 32 |
| II.6. Références .....   | 33 |
| <b>Chapitre III: Le data mining et ses techniques d'analyse de données</b>   |    |
| III.1. Introduction.....   | 34 |
| III.2. Différentes définitions du Data Mining.....                           | 34 |
| III.3. Historique.....   | 34 |
| III.4. La démarche du Data Mining.....                                       | 35 |
| III.4.1. Valorisation des données.....                                       | 36 |
| III.4.2. Définition d'un modèle .....  | 36 |
| III.4.3. Les modèles prédictifs .....  | 37 |
| III.4.4. La segmentation.....  | 37 |
| III.4.5. La description (ou identification de modèles fréquents).....        | 38 |
| III.4.6. Les deux types de techniques du Data Mining.....                    | 38 |
| III.4.6.1. Les techniques descriptives.....                                  | 38 |
| III.4.6.2. Les techniques prédictives.....                                   | 38 |
| III.4.6.3. Comparaison entre le Data Mining et la statistique.....           | 38 |
| III.4.6.4 Que peut-on trouver et prouver réellement avec le Data Mining ?... | 39 |
| III.4.6.5 Conclusions .....  | 39 |
| III.5. L'Analyse en Composantes Principales.....                             | 40 |
| III.5.1. Introduction.....   | 40 |
| III.5.2. Historique.....   | 40 |
| III.5.3. L'analyse de données.....   | 40 |
| III.5.4. La nature des données.....  | 41 |
| III.5.4.1. Définition des variables.....                                     | 41 |

|   |    |
|---|----|
| III.5.4.2. Définition des individus.....  | 41 |
| III.5.4.3. Le centrage et la réduction des données brutes.....  | 41 |
| III. 5.4.3.1. Les paramètres de dispersion.....   | 42 |
| III. 5.4.3.2. Caractéristique de la tendance centrale.....  | 42 |
| III. 5.4.3.3. Caractéristique de la dispersion.....   | 42 |
| III. 5.4.3.4. La variance.....  | 42 |
| III. 5.4.3.5. L'écart type .....  | 43 |
| III. 5.5. Qu'est-ce qu'une base de données ?.....   | 43 |
| III. 5.6. Principe de l'Analyse en Composantes Principales (ACP).....   | 43 |
| III. 5.7. Le processus de l'Analyse en Composantes Principales.....   | 44 |
| III. 5.7.1. Construction des axes principaux, démarche géométrique....  | 45 |
| III. 5.7.2. Calcul matriciel.....   | 46 |
| III. 5.8. Conclusion.....   | 47 |
| III. 6. La méthode des moindres carrés partiels (PLS).....  | 48 |
| III.6.1. Introduction-Problématique .....   | 48 |
| III.6.2. Historique .....   | 48 |
| III.6.3. Définition de la régression PLS ?.....   | 49 |
| III.6.4. Les objectifs de la régression PLS.....  | 49 |
| III.6.5. Les versions de la régression PLS.....   | 49 |
| III.7. Le principe de la régression PLS.....  | 50 |
| III.7.1. Données et notations.....  | 50 |
| III.7.2. La méthode PLS .....   | 50 |
| III.7.3.Indépendance des composantes dans la régression PLS.....  | 51 |
| III.8. Les avantages de la régression PLS.....  | 52 |
| III.9. Conclusion .....   | 53 |
| III.10. Références .....  | 54 |
| <b>Chapitre IV: Résultats et discussions</b>  |    |
| IV.1. Une étude comparative des propriétés physiques et mécaniques du lutétium avec celles des métaux de transition par l'approche du datamining..... | 57 |
| IV.1.1. Introduction.....   | 57 |
| IV.1.2. La base de données.....   | 59 |
| IV.1.3. Techniques de calculs.....  | 60 |
| IV.1.4. L'analyse en composantes principales (ACP) .....  | 60 |

|   |            |
|---|------------|
| IV.1.5. La régression par les Moindres Carrés Partiels (PLS).....   | 66         |
| IV.1.6. Conclusion.....   | 68         |
| IV.1.7. Références.....   | 69         |
| IV.2. Une étude comparative des propriétés physiques et mécaniques de l'hydrogène en appliquant les techniques de recherche du data mining..... | 71         |
| IV.2.1. Introduction.....   | 71         |
| IV.2.2. Les bases de données .....  | 72         |
| IV.2.2.1 La bases de données des métaux alcalins y compris l'hydrogène  | 72         |
| IV.2.2.2. L'analyse en composantes principales (ACP).....   | 73         |
| IV.2.2.3 La bases de données du groupe de carbone y compris l'hydrogène.....  | 75         |
| IV.2.2.4. L'analyse en composantes principales (ACP) .....  | 75         |
| IV.2.2.5. La bases de données des halogènes y compris l'hydrogène.....  | 78         |
| IV.2.2.6. L'analyse en composantes principales (ACP) .....  | 78         |
| IV.2.3. La régression par les Moindres Carrés Partiels (PLS).....   | 81         |
| IV.2.4. Conclusion .....  | 85         |
| IV.2.5. Références .....  | 86         |
| IV.3. Investigation des propriétés structurales et mécaniques des hydrures alcalins par les calculs ab initio et l'approche du datamining.....  | 104        |
| IV.3.1. Introduction.....   | 104        |
| IV.3.2. Base de données .....   | 105        |
| IV.3.3. L'analyse en composantes principales (ACP).....   | 107        |
| IV.3.3.1. Les propriétés structurales .....   | 107        |
| IV.3.3.2. Les propriétés mécaniques.....  | 107        |
| IV.3.4. Analyse par l'approche data mining.....   | 108        |
| IV.3.5. Conclusion.....   | 112        |
| IV.3.6. Références.....   | 113        |
| <b>Conclusion générale.....</b>   | <b>112</b> |

## Table des figures

| <b>Figures</b>       | <b>Titres</b>  | <b>Pages</b> |
|----------------------|--|--------------|
| <b>Figure II.1</b>   | La première forme du système périodique de Mendeleev datant du 17 février 1869, intitulée « une tentative au système d'éléments, se basant sur leur masse ». | <b>22</b>    |
| <b>Figure II.2</b>   | Dimitri Ivanovitch Mendeleev   | <b>22</b>    |
| <b>Figure II.3</b>   | Hierarchie des niveaux énergétiques des orbitales atomiques (visualisée par la règle de Klechkowski).  | <b>26</b>    |
| <b>Figure III.1</b>  | Technologies et modèle général du data mining.   | <b>37</b>    |
| <b>Figure III.2</b>  | Illustration en 2D de la réduction de la dimension. La première composante contient le maximum d'information (maximum de variance).                          | <b>44</b>    |
| <b>Figure III.3</b>  | Représentation en 2D des 2 vecteurs propres et des 2 valeurs propres associés aux deux nouvelles composantes.  | <b>45</b>    |
| <b>Figure III.4</b>  | Projection des individus dans le plan principal.   | <b>46</b>    |
| <b>Figure IV.1.1</b> | Le graphe des observations des métaux de transition de la troisième période y compris le lutétium.   | <b>61</b>    |
| <b>Figure IV.1.2</b> | Le graphe des variables des métaux de transition de la troisième période y compris le lutétium.  | <b>64</b>    |
| <b>Figure IV.1.3</b> | Le graphe des observations (t1, t2) des métaux de transition de la troisième période y compris le lutétium.  | <b>65</b>    |
| <b>Figure IV.2.1</b> | Le graphe des observations des métaux alcalins y compris l'hydrogène.  | <b>71</b>    |
| <b>Figure IV.2.2</b> | Le graphe des variables des métaux alcalins y compris l'hydrogène.   | <b>72</b>    |
| <b>Figure IV.2.3</b> | Le graphe des observations du groupe du carbone y compris l'hydrogène.   | <b>75</b>    |
| <b>Figure IV.2.4</b> | Le graphe des variables du groupe du carbone y compris l'hydrogène.  | <b>75</b>    |

|                      |   |            |
|----------------------|---|------------|
| <b>Figure IV.2.5</b> | Le graphe des observations du groupe des halogènes y compris l'hydrogène.   | <b>76</b>  |
| <b>Figure IV.2.6</b> | Le graphe des variables du groupe des halogènes y compris l'hydrogène.  | <b>77</b>  |
| <b>Figure IV.2.7</b> | Le cercle de corrélation entre les propriétés mécaniques, et les propriétés physiques dans les deux composantes PLS ( $t_1$ , $t_2$ ) (la base de données des métaux alcalins).   | <b>81</b>  |
| <b>Figure IV.2.8</b> | Le cercle de corrélation entre les propriétés mécaniques, et les propriétés physiques dans les deux composantes PLS ( $t_1$ , $t_2$ ) (la base de données du groupe de carbone).  | <b>81</b>  |
| <b>Figure IV.2.9</b> | Le cercle de corrélation entre les propriétés mécaniques, et les propriétés physiques dans les deux composantes PLS ( $t_1$ , $t_2$ ) (la base de données des halogènes).   | <b>82</b>  |
| <b>Figure IV.3.1</b> | Structure cristalline de la structure NaCl (B1) des hydrures alcalins MH dans les conditions ambiantes. Les cercles bleus représentent $M = \text{Li, Na, K, Rb, Cs}$ , et les cercles violets représentent les atomes d'hydrogène. | <b>102</b> |
| <b>Figure IV.3.2</b> | Le graphe des observations des hydrures alcalins.   | <b>108</b> |
| <b>Figure IV.3.3</b> | Le graphe des variables des hydrures alcalins.  | <b>108</b> |
| <b>Figure IV.3.4</b> | Le graphe de «Biplot » représentant les hydrures alcalins et les propriétés.  | <b>109</b> |

## Liste des tableaux

| Tableaux              | Titres   | Pages      |
|-----------------------|--|------------|
| <b>Tableau II.1</b>   | Désignation des valeurs du nombre quantique azimutal, $l$ .  | <b>24</b>  |
| <b>Tableau II.2</b>   | Niveaux énergétiques des orbitales atomiques.  | <b>24</b>  |
| <b>Tableau IV.1.1</b> | La base de données incluant les propriétés physiques et mécaniques des métaux de transition de la troisième période y compris le lutétium.   | <b>58</b>  |
| <b>Tableau IV.1.2</b> | La matrice de corrélation des métaux de transition de la troisième période y compris le lutétium.  | <b>60</b>  |
| <b>Tableau IV.1.3</b> | Les propriétés physiques du groupe 3 (IIIB), y compris le lutécium. Les quantités indiquées dans le tableau sont les suivantes: l'électronégativité ( $\chi$ ), le rayon atomique ( $r$ ), 1 <sup>er</sup> potentiel d'ionisation (eV), 2 <sup>ème</sup> potentiel d'ionisation (eV), le point de fusion ( $^{\circ}$ C).  | <b>63</b>  |
| <b>Tableau IV.2.1</b> | La base de données incluant les propriétés physiques et mécaniques des métaux alcalins y compris l'hydrogène.  | <b>70</b>  |
| <b>Tableau IV.2.2</b> | La base de données incluant les propriétés physiques et mécaniques du groupe de carbone y compris l'hydrogène.   | <b>73</b>  |
| <b>Tableau IV.2.3</b> | La base de données incluant les propriétés physiques et mécaniques des halogènes y compris l'hydrogène.  | <b>76</b>  |
| <b>Tableau IV.3.1</b> | Les valeurs calculées de paramètre de maille $a_0$ (Å), le module de compressibilité ( $B_0$ ) (GPa) et la dérivé première de la pression ( $B'_0$ ) des hydrures alcalins dans la structure NaCl (B1), avec quelques autres résultats.  | <b>103</b> |
| <b>Tableau IV.3.2</b> | Les Constantes mécaniques et élastiques des hydrures alcalins dans la structure NaCl (B1), avec quelques autres résultats, le module de compressibilité ( $B_0$ ), module de cisaillement ( $G$ ), constantes élastiques ( $C_{ij}$ ), le module d'Young ( $E$ ) sont en (GPa). Le rapport $B/G$ , le facteur d'anisotropie ( $A$ ) et le coefficient de Poisson ( $\nu$ ) sont des paramètres sans dimension. | <b>104</b> |

## INTRODUCTION GENERALE

## Introduction générale

La découverte de la structure de l'atome au début du vingtième siècle a marqué une grande étape dans l'avancée des connaissances. De nombreux phénomènes physiques et chimiques demeuraient sans interprétation tels que l'effet Compton, l'effet Zeeman, ...*etc.* Ce qui a conduit les physiciens et chimistes à abandonner le modèle classique de la mécanique newtonienne et à introduire un nouveau modèle qui est la mécanique quantique, ce qui a permis une interprétation nouvelle de la classification périodique [1].

La mécanique quantique a été la théorie la plus spectaculaire réussie dans l'histoire de la science. L'impact de la théorie a été ressenti dans des domaines tels que la physique du solide, et la science des matériaux [2].

L'évolution de l'histoire de la chimie au XIXe siècle nous montre à quel point une classification devenait nécessaire. La classification telle que Mendeleiev l'a publiée en 1869 puis en 1871 a stimulé la recherche en chimie et de nombreuses questions fondamentales se sont appuyées sur cette formidable base théorique que renferme ce tableau. C'est actuellement un point de repère universel pour la communauté des chimistes [3].

Le tableau périodique reste, après plus d'un siècle, le plus important moyen de comparaison entre les éléments chimiques. Il nous permet d'étudier rationnellement la grande variété des substances que nous trouvons dans la nature [4].

La matière condensée est formée par l'agrégation d'atomes, et les propriétés de l'agrégat dépendent de la structure électronique des atomes qui le constituent. Elle est donc d'une grande importance, ce qui est essentiel pour la classification des propriétés physiques des éléments solides comme il l'est pour ses atomes constitutifs [5].

La classification périodique des éléments exprime une profonde vérité physique et permet d'unifier un grand nombre de connaissances portant sur la matière. Ce classement facilite la compréhension des groupes d'éléments en montrant la relation entre leurs propriétés chimiques et leur structure atomique. De plus, la connaissance du tableau périodique permet de prévoir les formules des composés et les types de liaisons unissant les composants d'une molécule [4]. Les éléments dans le tableau périodique sont rangés par ordre croissant du

numéro atomique  $Z$ , selon la configuration électronique des atomes [6], en respectant la règle de Klechkowski dans le remplissage des orbitales par énergie croissante.

Le tableau périodique est pour nous, de toute évidence, le guide le plus sûr dans l'étude de la chimie. Mais il comporte certaines anomalies ou contradictions, qui sont les éléments qui ne respectent pas la règle de Klechkowski, et qui présentent des configurations électroniques faisant exception à la norme, tels que :

- L'hydrogène qui ne s'ajuste harmonieusement à aucun groupe, c'est un non-métal est il est placé généralement dans la même colonne que le lithium qui est un métal très réactif. L'hydrogène est parfois placé en tête du groupe 7, avec les halogènes. Dans certains tableaux périodiques, il est même placé dans les deux groupes à la fois [7].
- L'Hélium se trouve dans le bloc p alors qu'il est de type s.
- Le lutécium et le lawrencium sont placés dans le bloc f, alors que leur couche f est déjà saturée, ils contiennent 1 électron 5d ou 6d. Ils constituent bien le début des éléments de transition.

Les tableaux périodiques habituels de Mendeleïev des éléments chimiques sont déjà à 97% en accord avec la mécanique quantique. Trois éléments seulement qui ne le sont pas, et sont en désaccord avec le principe d'exclusion de Pauli. Afin d'assurer la cohérence, il est proposé de placer l'hélium à côté de l'hydrogène dans le bloc s. Le lutécium et le lawrencium dans le bloc  $d$  des métaux de transition, pour que le tableau périodique soit conforme à 100% avec la mécanique quantique [8].

Dans cette thèse on a étudié les contradictions existantes dans le tableau périodique, et on les a confronté moyennant l'approche data mining et ses techniques, telles que l'analyse en composantes principales (ACP) et la régression (PLS), qui sont employées pour la première fois pour l'étude de ce problème, et ce qui fait l'originalité de ce travail.

Cette thèse est organisée comme suit:

- Le premier chapitre est consacré à la présentation générale du réductionnisme en chimie quantique, ses différentes formes, ses aspects, les types de réduction, la comparaison entre la

physique et la chimie, la réduction de la chimie à la mécanique quantique, et le lien existant entre le tableau périodique et le réductionnisme en chimie quantique.

- Le deuxième chapitre est consacré à l'étude du tableau périodique, son historique, la loi de la périodicité des propriétés, son évolution, et ses anomalies.

- Le troisième chapitre est consacré à la définition du datamining et ses différentes techniques, l'analyse en composantes principales (ACP) et la régression PLS (la régression par les moindres carrés partiels), le principe de ces méthodes, et leurs objectifs.

- Le quatrième chapitre présente les résultats liés à l'application de l'approche datamining, et ses techniques telles que : l'analyse en composantes principales (ACP) et la régression (PLS), qui ont été appliquées sur les différentes bases de données contenant les différentes familles du tableau périodique qui présentent des contradictions avec la mécanique quantique.

À la fin de cette thèse, une conclusion générale fait un récapitulatif de l'essentiel des principaux résultats obtenus tout en dégageant des perspectives pour la poursuite de cette investigation.

**Références**

- [1] André Durupthy, Jacques Estienne, Magali Giacino, Alain Jaubert, Claude Mesnil, « Physique, Chimie et sciences de l'ingénieur » (PCSI), I.S.B.N. 978-2-0118-1752-5. [www.hachette-education.com](http://www.hachette-education.com).
- [2] Eric R. Scerri, « Collected papers on philosophy of chemistry », Imperial College Press, (2008). ISBN-13 978- 1-84816- 137-5.
- [3] Jean-François Le Maréchal, « La classification périodique des éléments chimiques », Université Lyon 2 - UMR GRIC équipe COAST (1999).
- [4] Le repaire des sciences, <http://www.lerepairedessciences.fr>
- [5] H.P. Myers, « Introductory Solid State Physics », 2<sup>nd</sup> ed, Taylor & Francis e-Library, (2009), ISBN 0-203-21255-X.
- [6] <http://www.web-sciences.com>.
- [7] Carole McQuarrie, Donald A. McQuarrie, Peter A. Rock, « Chimie générale », Editions De Boeck Université 1<sup>st</sup> ed., (1992).
- [8] Bernard SCHAEFFER, Periodic table quantum mechanics consistent, <https://www.scribd.com/doc/5994795/Periodic-table-quantum-mechanics-consistent>.

## CHAPITRE I

### Le réductionnisme en chimie quantique

## **I.1. Introduction**

Bien qu'il soit évident que les tentatives pour expliquer la chimie en termes de la théorie quantique ne peuvent apparaître suite au développement de la mécanique quantique au XXe siècle, ce type de réductionnisme ne date pas d'aujourd'hui.

En général, le réductionnisme est l'idée qu'une science, telle que la chimie, peut être réduite plus à une science "fondamentale", dans ce cas, à la physique.

Comme nous le verrons dans ce qui suit, la conception même du réductionnisme supposant une hiérarchie dans les sciences entre les sciences plus fondamentales, généralement la physique et les mathématiques, et ceux qui sont considérées comme des dérivés ou des épiphénomènes, à commencer par la chimie, et passant par la biologie aux sciences humaines comme la sociologie ou la psychologie [1].

## **I. 2. Définition du réductionnisme**

Le réductionnisme consiste à rendre compte de la grande variété des phénomènes naturels par un petit nombre d'entités « élémentaires » et de lois « fondamentales ». Son but, selon le physicien français Jean Perrin, est d'« *expliquer le visible compliqué par de l'invisible simple* », faisant de la simplicité une marque de vérité [2].

Réduire, dans un sens général, c'est rapetisser ; pour les sciences, on comprend le plus souvent que réduire, c'est rendre compte d'une réalité dans les termes des propriétés des éléments qui la constituent [3].

## **I. 3. Les différentes formes de réductionnisme**

Le réductionnisme est un des thèmes les plus importants de l'épistémologie. Il existe plusieurs formes de réductionnisme :

- La première, dite **ontologique**, consiste simplement à admettre que les lois d'une discipline s'appliquent à une autre. (Dans le cadre qui nous intéresse ici, cela veut dire admettre que toutes les lois de la physique, et elles seules, sont valables dans les systèmes chimiques et dans l'ensemble de la chimie.)
- Le deuxième un niveau de réduction **méthodologique**, qui consiste à admettre que l'on peut tout déduire dans une discipline à partir d'une autre. (Ici cela voudrait dire que l'on peut

analyser et comprendre le comportement d'un système chimique à partir des seules lois de la physique.)

- Il y a enfin un réductionnisme **épistémologique** (qui consisterait à admettre que la chimie n'est qu'une branche de la physique). Il n'est jamais simple de mettre tout le monde d'accord sur les réductions existantes entre deux champs disciplinaires. On peut cependant reconnaître que la chimie est réductible à la physique du point de vue ontologique et méthodologique, mais pas du point de vue épistémologique. La chimie n'est certainement pas une branche de la physique, mais la mécanique quantique, et elle seule, permet de calculer avec une grande précision les liaisons chimiques [4].

#### **I. 4. Réductionnisme et complexité**

L'essence des débats sur le réductionnisme est la question suivante: la connaissance d'un objet complexe se réduit-elle à la seule connaissance de ses composants? Cette conception de la réduction du complexe au simple a déjà été réfutée par Aristote (dans la métaphysique), qui remarquait que la syllabe BA n'était pas identique à B et A, mais avait un contenu informatif qui allait au-delà de celui des deux lettres qui la composent. Elle est autre chose. Il y a **émergence** d'une propriété nouvelle. La définition de ce que sont des objets ou des systèmes complexes demande à être précisée. Pour Gallacher et Apenzeller, dans leur introduction au dossier déjà mentionné de Science (1999): «Un système complexe est celui dont les propriétés ne sont pas complètement expliquées par la connaissance et la compréhension de ses composants.» Cette définition identifie les objets complexes et les objets pour lesquels il est nécessaire d'aller au-delà d'une analyse purement réductionniste pour les comprendre. Selon cette définition, une syllabe ou un mot sont déjà des objets complexes.

Une autre définition empirique d'un système complexe est celle, d'un système dont les composants sont nombreux et en interaction. D'autres définitions sont données dans un dossier («Complex systems», Nature **410**, 2001, p. 241-284.) consacré aux systèmes complexes, principalement physiques, paru dans Nature, comme par exemple celle d'un système opérant à la limite du chaos. On trouve de nombreux exemples de complexité dans un livre récent, totalement consacré à ce thème (Réda Benkirane, 2002). En fait, il semble que l'on parle de choses assez différentes quand un physicien, un informaticien ou un biologiste utilisent le terme de complexité. Ce qui est commun est sans doute ce que Chaitin, dans cet ouvrage, appelle complexité algorithmique, qui est essentiellement la quantité d'informations

(les composants mais aussi toutes les interactions pouvant exister entre eux) nécessaires pour décrire le système.

L'un des objectifs de la science est d'essayer de surmonter les problèmes posés par cette complexité. En physique, dans le cas du problème à N corps, la théorie du chaos permet de surmonter la complexité.

La complexité est, comme dit plus haut, une notion très ancienne; mais elle a connu récemment un développement important dont témoigne la publication, en décembre 2003, dans la revue *Pour la science*, d'un numéro spécial qui lui est totalement consacré sous le titre «La complexité: la science du XXI<sup>e</sup> siècle». Ce concept a des applications dans tous les domaines, y compris les sciences humaines [4].

### **I. 5. Aspects du réductionnisme**

La question du réductionnisme s'est posée à tous les niveaux de la classification des sciences: peut-on réduire les mathématiques à la logique, la chimie à la physique, les sciences de la vie à la physico-chimie, la psychologie et la sociologie à la biologie? Historiquement, elle a acquis son vocabulaire propre («réduire», «ramener», tant en langue française qu'allemande) au milieu du XIX<sup>e</sup> siècle.

Quelles sont les raisons de la persistance, aujourd'hui, de la querelle du réductionnisme, entamée il y a au moins un siècle et demi? Cette querelle assurément n'est pas vide de tout contenu, car elle possède des motivations clairement identifiées. Il est pourtant difficile d'en déceler le fondement réel, et plus difficile encore d'en démontrer l'utilité [4].

### **I.6. Le réductionnisme, une démarche incontournable**

La physique corpusculaire, et la mécanique quantique ont conduit à diversifier les concepts et les critères mêmes de la réalité, eu égard notamment à la « non-séparabilité » absolue entre l'espace et le temps, entre la matière et l'énergie...

Il n'empêche, toute description de la réalité semble rester fidèle à un principe hiérarchique visant à définir des niveaux supérieurs à partir de niveaux plus fondamentaux. Ainsi, et pour rester dans le domaine de la physique, « toute la réalité matérielle est censée s'expliquer au niveau fondamental par des quarks, des leptons (électrons, muons et tauons) et quatre interactions fondamentales qui rendent compte des forces exercées entre les particules de la

matière ». Il en va de même pour la chimie, pour la biologie, qui peuvent être donc « réduites » selon cette même logique à la physique. Le réductionnisme gouverne aujourd'hui toute démarche scientifique ; ce qui permet en outre d'effectuer les ponts nécessaires entre les diverses disciplines que croise par exemple la modélisation moléculaire : physique, chimie, biologie, médecine... ce qui lui vaut d'être un mode de réduction emblématique.

La conception réductionniste vise donc toujours plus l'élémentaire, le niveau fondamental, afin d'assurer toujours davantage une « maîtrise du réel », et par suite, une efficacité de l'intervention humaine sur ce même réel [5].

### **I.7. L'idée de réductionnisme**

Les réductionnistes, ont la vertu d'essayer d'être aussi clairs et explicites que possibles. La première précision qu'ils apportent est qu'il ne faut pas parler de concepts ou de propriétés (comme par exemple la température) sans les placer à l'intérieur d'une théorie. Ce que l'on met en rapport, ce sont toujours des théories. La réduction d'une théorie T à une théorie S établit une telle relation entre deux théories.

De façon informelle, on peut dire qu'une théorie T est réductible à une théorie S si toute observation que l'on peut expliquer par T peut être expliquée aussi par S. (C'est une relation asymétrique ; toute observation explicable par S n'est pas nécessairement explicable par T.) Il faut ajouter, pour qu'il y ait un sens à parler de réduction, que l'on suppose que le vocabulaire de T (la théorie que l'on réduit) contient des termes que S ne contient pas, c'est-à-dire que T parle d'objets dont S ne parle pas et/ou à l'aide de concepts que S ne connaît pas. C'est, de nouveau, une de ces trivialités dont les conséquences ne sont pas triviales. Evidemment, le vocabulaire de S peut contenir des termes que celui de T ne contient pas, de sorte que l'on peut dire plus simplement que ce qu'on demande est que le vocabulaire de T ne soit pas inclus dans celui de S.

Les réductionnistes admettent l'existence de niveaux et, associée à chaque niveau, une liste de propriétés caractéristiques de ce niveau. Nous pouvons alors, en suivant Broad<sup>1</sup> et en utilisant une terminologie modifiée, distinguer trois sortes de lois. (i) Les lois 'intra-niveaux' qui concernent les propriétés caractéristiques d'un niveau. Elles sont valables à ce niveau et à des niveaux supérieurs, mais pas à des niveaux inférieurs (par exemple une loi qui relie la température et l'entropie d'un système). (ii) Les lois 'inter-niveaux' qui relient des propriétés

---

<sup>1</sup> BROAD, C.D. *The mind and its place in nature*. London : Routledge & Kegan Paul, (1925) pp. 77-78.

définies à un niveau  $i$  donné à celles définies au niveau  $i-1$  (par exemple une loi qui relie l'énergie cinétique à la température). (iii) Des lois 'trans-niveaux' qui seraient valables non pas pour un niveau spécifique, ni pour une paire spécifique de niveaux adjacents, mais pour tout niveau et toute paire de niveaux adjacents.

Les premières ('intra') n'offrent à mon avis pas de problèmes de principe, mais bien les dernières ('trans'). En voici un exemple. Oppenheim et Putnam proposent l'échelle suivante comme « système de niveaux réductifs » : (1) particules élémentaires, (2) atomes, (3) molécules, (4) cellules, (5) êtres vivants multicellulaires, (6) groupes sociaux. Laissant de côté tout ce que les niveaux supérieurs peuvent avoir de problématique, constatons qu'ils obéissent manifestement aux deux lois trans-niveaux : (i) toute entité d'un niveau quelconque, excepté le niveau le plus bas, peut être décomposée en des entités du niveau sous-jacent, (ii) aucune entité d'un niveau quelconque ne possède des parties qui appartiennent à un niveau supérieur. Ce sont sans doute des lois 'trans' pour système de niveaux, qui d'ailleurs été construit conformément à ces lois. Ce ci montre simplement que la notion de lois trans-niveaux n'est pas absurde même si l'on n'en trouve pas d'intéressantes.

Les lois 'inter-niveaux' de quelle nature sont-elles, quels sont leurs ingrédients? Contentons nous ici de prendre comme exemple une loi qui relie la mécanique à la thermodynamique, et intéressons nous plus particulièrement et de façon tout à fait sommaire à cette loi des gaz parfaits, selon laquelle le produit de la pression,  $P$ , et du volume,  $V$ , d'une quantité donnée de gaz est proportionnel à sa température,  $T$  :  $PV \propto T$ . Comment peut-on déduire cette loi de la mécanique à laquelle obéissent les molécules qui composent le gaz ? Quels sont les ingrédients de cette déduction ? [6].

## **I.8. Les types de réduction**

Il semble raisonnable de distinguer, la réduction "horizontale" d'une théorie à une autre plus englobante, dont la première devient un cas particulier, et la réduction "verticale" d'une théorie à une autre portant sur des objets considérés comme plus fondamentaux parce que plus élémentaires.

### **I.8.1. Réduction verticale**

La réduction de la chimie à la mécanique quantique est un des exemples le plus souvent cités. « Les lois physiques fondamentales nécessaires pour la théorie mathématique de la plus grande partie de la physique et de la totalité de la chimie sont ainsi complètement

connues, et la difficulté est seulement que l'application de ces lois conduit à des équations beaucoup trop compliquées pour qu'on puisse les résoudre » (Dirac, 1929).

### I.8.2. Réduction horizontale

Pour prendre un exemple connu, la mécanique quantique a-t-elle englobé la mécanique classique en sorte de la rendre superflue, ou en tout cas de la réduire au statut d'un cas particulier, d'un cas limite? On sait que Bohr<sup>2</sup> n'était pas de cet avis. Il attachait une importance considérable au passage à la limite où la constante de Planck tend vers zéro, qu'il avait théorisé sous le nom de principe de correspondance. Mais ce n'était pour lui que l'une des faces de la relation entre la mécanique quantique et la physique classique : l'autre face étant que les concepts classiques restaient, et devaient rester indispensables à la description des expériences. « En fait, nous devons comprendre que l'interprétation non ambiguë de n'importe quelle mesure doit être articulée essentiellement en termes des théories physiques classiques, et nous pouvons dire en ce sens que le langage de Newton et Maxwell restera toujours le langage des physiciens » [7].

## I. 9. Comparaison entre la physique et la chimie

En 1669, Bernard Le Bovier De Fontenelle, le premier et le plus célèbre secrétaire perpétuel de l'académie royale des sciences à Paris, avait proposé la comparaison suivante: Par ses opérations évidentes, la chimie résout les corps en un certain nombre de principes réels et bruts; les sels, les soufres, etc. tandis que par ses spéculations délicates, la physique agit conformément aux principes comme agit la chimie sur les corps, les résolvant en d'autres principes encore plus simples, petits corps façonnés et déplacés dans un nombre infini de manières: c'est la différence principale entre la physique et la chimie, l'esprit de la chimie est plus confus, plus dense; où les principes sont mélangés l'un avec l'autre, tandis que l'esprit de la physique est plus clair, plus simple, moins encombré, et, enfin, va droit aux origines des choses, tandis que l'esprit de la chimie ne va pas à la fin.

Ainsi, alors que la chimie est souvent perçue comme visant à fournir les principes généraux régissant les transformations matérielles, c'est la physique qui est engagée dans la recherche des constituants finaux de la nature et des lois fondamentales régissant leur comportement.

---

<sup>2</sup> BOHR N, 1931, Maxwell and modern theoretical physics, in Nature (Suppl.) 128, 691-692. Et Niels Bohr, Collected Works, Amsterdam, North-Holland, vol. 6, 1985, pp. 357-360.

En effet, comme une série d'études récentes sur la chimie du XVIIIe siècle a montré, ces positions ont été considérées comme étant entièrement compatibles. Ainsi, Daniel Sennert, par exemple, explique un certain nombre d'opérations chimiques en termes d'ajout ou de retrait d'atomes qui possèdent leurs propres propriétés chimiques caractéristiques. Nous devons donc conclure, que ni la discontinuité ni l'homogénéité des matières sous-entendues par atomisme ne fournit les raisons suffisantes pour distinguer une vue philosophique caractéristique du physicien de celle du chimiste.

C'est trop superficiel d'essayer de distinguer la chimie de la physique suivant la ligne des éléments contre les atomes, entre une philosophie de matière caractérisée par des principes inhérents ou une vision basée sur la matière homogène mais discontinue. D'abord, le désaccord entre ces deux traditions a précédé l'apparition des disciplines modernes de la physique et de la chimie. En second lieu, et de manière plus significative, les chimistes ont toujours été confrontés à une multiplicité d'interprétations des phénomènes chimiques, et un grand nombre d'entre eux n'ont pas hésité à allier la doctrine corpusculaire avec celle des éléments ou des principes.

Au milieu du dix-neuvième siècle, la définition d'un composé par la nature et la proportion de ses constituants étaient défiées par un paradigme structural qui a tenu compte de l'arrangement des atomes en molécules. Sur ce modèle, c'est l'arrangement physique des éléments constitutifs qui explique les propriétés du composé [1].

## **I. 10. Le monde quantique**

La révolution quantique, conduite vers le milieu des années vingt par de jeunes physiciens est peut-être, parmi les nombreuses révolutions dont est jalonnée l'histoire des sciences, celle qui a le plus radicalement et profondément bouleversé l'image que les physiciens se faisaient du monde. La difficulté qu'ils eurent à l'accepter fut parfois si grande que certains, et non des moindres puisque Einstein fut l'un d'entre eux, ne parvinrent pas à la surmonter. « Celui qui n'est pas choqué par la théorie quantique », disait Bohr, « ne l'a pas comprise ». Elle remettait profondément en question les concepts auxquels les avait habitués la mécanique rationnelle et que la relativité n'avait pas vraiment bousculés : un monde fait de points matériels doués à tout moment d'une masse, d'une position et d'une vitesse bien définies. Elle introduisait à sa place la notion d'état d'un système et en faisait l'élément de base de son processus d'abstraction et de mathématisation.

Aujourd'hui, si elle est toujours aussi choquante et indigeste – et par conséquent difficile à enseigner aux jeunes étudiants – elle est unanimement acceptée par les hommes de science. Les anciens concepts de la mécanique rationnelle sont compris comme n'étant que des approximations, en quelque sorte des illusions de la mécanique quantique. Il a fallu, pour parvenir à ce résultat, des années d'observations et d'expériences qui ont donné de la théorie tant d'éclatantes confirmations que le doute n'est plus permis [8].

### **I. 10. 1. La réduction de la chimie à la mécanique quantique**

Les concepts et théories que les chimistes ont développés, sans contradiction avec la mécanique quantique mais en renonçant aux solutions exactes et en usant d'approximations justifiées par les données expérimentales, sont-ils légitimes ?

Même avec les techniques actuelles, on ne peut obtenir des solutions exactes de l'équation de Schrödinger que dans des cas extrêmement particuliers : atomes à un électron, molécules diatomiques homonucléaires à un électron. De plus, la fonction d'onde, définie dans un espace de configuration abstrait, n'a pas de signification physique simple : elle oblige le chimiste à renoncer à toute tentative de voir ce que les atomes et les molécules sont en train de « faire » – et qui serait possible seulement dans l'espace à trois dimensions. Il en résulte une curieuse dichotomie en chimie quantique : les approches ad hoc, simples et qualitatives, très utilisées par les chimistes, coexistent avec des approches rigoureuses ou ab initio, dépourvues de concepts et d'images simples. Il n'y a guère d'espoir, concluaient ces auteurs, que l'abîme séparant ces deux types d'approches puisse être jamais franchi (« There appears hardly any hope of ever bridging the chasm between the two »). Autrement dit, la réduction de la chimie à la mécanique quantique est impossible, jusqu'à nouvel ordre. Certes, la mécanique quantique joue un rôle fondamental en chimie, mais le concept de réduction est incapable de décrire ce rôle. Non seulement les solutions exactes des équations quantiques sont presque toujours inaccessibles, mais encore si nous les possédions nous ne saurions pas quoi en faire. Les solutions approchées, élaborées en tenant compte des données empiriques, nous aident bien mieux à comprendre. Pourquoi tenter de calculer exactement telle grandeur chimique, quand on peut la mesurer ?

Un autre physicien, Philip W. Anderson, a publié une intéressante critique du réductionnisme sous le titre *More is different* (Anderson. 1972). Il est vrai qu'il se dit « réductionniste » : mais cela ne signifie pour lui rien de plus que la compatibilité au sens de Leggett : les lois de la physique sont universelles, en ce sens qu'aucun phénomène connu ne

leur échappe. Mais, ajoute Anderson, on ne peut en aucune manière passer du réductionnisme au “constructionnisme” : « La capacité de tout réduire à des lois fondamentales simples n’implique pas la capacité de partir de ces lois pour reconstruire l’univers. En fait, plus la physique des particules élémentaires nous dit de choses sur la nature des lois fondamentales, moins celles-ci semblent pertinentes pour les problèmes bien réels du reste de la science, et moins encore pour ceux de la société ». Après avoir analysé divers exemples (molécules, solides, noyaux), Anderson énonce qu’en général la relation entre un système et ses parties à sens unique : la synthèse est presque impossible, tandis que l’analyse peut être non seulement possible, mais féconde à tous égards. Et il conclut : « L’arrogance du physicien des particules est peut-être derrière nous (“ le reste est de la chimie”, a dit le découvreur du positron) [7].

En 1929, Paul Dirac, l’un des créateurs de la mécanique quantique, avait déclaré que les lois physiques fondamentales nécessaires pour la théorie mathématique d’une grande partie de physique et de la totalité de la chimie sont maintenant complètement connues. La seule difficulté, selon Dirac, c’était que l’application exacte de ces lois a mené aux équations beaucoup trop compliquées pour être soluble. Ceci et d’autres déclarations similaires ont clairement démontré l’ambition des physiciens pour réduire tous les phénomènes chimiques aux mécanismes de la mécanique quantique des atomes. L’argument lui-même est tout à fait franc. En principe, la chimie est entièrement circonscrite par la description de la distribution d’électron exprimée par l’équation de Schrödinger. Avec le principe établi, le reste de la chimie est une question de détails. Quoique des solutions exactes pour l’équation de Schrödinger ne sont connues que pour les atomes simples comme l’hydrogène et l’hélium avec seulement un ou deux électrons, il est raisonnable de supposer qu’avec l’amélioration des outils mathématiques, et en particulier, l’énorme multiplication de la puissance de calcul apportée par l’ère numérique, approximations ou solutions de mieux en mieux pour tous les autres éléments et composés suivra. Ainsi, la chimie pourrait, en principe au moins, être déduite de la mécanique quantique. Les concepts d’électrons, des protons, des orbitales, et les niveaux d’énergie qui sont déployés pour expliquer les phénomènes chimiques réussissent à le faire seulement grâce à l’utilisation judicieuse des hypothèses complémentaires. Selon la théorie quantique, le carbone devrait être bivalent, mais ceci n’arrête pas la majeure partie de ses combinaisons étant tétravalent. L’hypothèse de l’hybridation des orbitales explique cette tétravalence, mais, tout en étant compatible avec la théorie quantique, elle demeure ad hoc. En outre, alors que la systématisation des théories en physique pourrait bien avoir réussi à intégrer certains éléments de chimie, il n’a pas mis un terme au développement des théories chimiques indépendantes. C’est pourquoi Monique Lévy discute à la fin d’une " réduction par

synthèse " qui est une forme de réduction qui réserve un rôle fondamental pour la discipline "réduite " la relation entre la microstructure et les propriétés macroscopiques qui forme l'axe central de la physique de l'état solide, et qui a profondément modifié l'approche de tous les chimistes travaillant sur des matériaux solides [1].

Ces dernières années, la réduction de la chimie a été traitée de différentes façons. De nombreuses études se sont concentrées sur la réduction inter-théorique entre les théories de la chimie et les théories de la physique (Bunge, 1982; Primas, 1983). D'autres ont discuté de la réduction de la chimie d'une manière naturaliste, en examinant la question comment certaines propriétés généralement moléculaires tels que les angles de liaison peuvent être déduites de la mécanique quantique dans un modèle *ab initio* ou si le système périodique peut être déduit de la même façon de la mécanique quantique [9].

### **I. 10. 2. Le tableau périodique et le réductionnisme en chimie quantique**

Le tableau périodique des éléments a eu une influence profonde sur le développement de la chimie et de la physique moderne. Dans la chimie son influence est bien connue et indéniable. Les fonctions du système périodique comme un principe unificateur continue à guider la recherche des chimistes de jour en jour dans de nombreux domaines spécialisés. L'influence de la classification périodique sur le développement de la physique et en particulier la mécanique quantique n'est pas si bien connue mais également indéniable. La mécanique quantique fait partie de la tradition réductionniste dans la science moderne [9].

En 1939, Georges Champetier a suggéré que le tableau périodique avait efficacement été réinventé en raison de la théorie atomique. Selon ses propres paroles, "la théorie atomique est maintenant devenue le fondement le plus solide de la chimie théorique et expérimentale. La classification périodique a pris une importance beaucoup plus grande à la suite de la clarification de nos idées concernant la structure atomique". Champetier a suggéré que la réforme pourrait être poussée plus loin en redéfinissant l'élément en termes de nombre combiné de protons et de neutrons dans le noyau, donnant à chaque " isotope " sa propre place dans un nouveau tableau. En dépit de cette réorientation proposée du tableau périodique autour des découvertes en physique atomique, Champetier n'a jamais suggéré qu'il pourrait être possible de déduire la chimie de la physique, voyant le futur rapport entre les deux disciplines en termes de collaboration plutôt que de domination. Ainsi, la chimie des solides continue à devenir de plus en plus orientée vers la science des matériaux, qui étudie - ou, de plus en plus, conçoit - des matériaux en termes de leurs propriétés, développant les matériaux

qui sont particulièrement bien adaptés à des applications données. De ce point de vue, le chimiste peut voir le tableau périodique comme une boîte à outils bien organisée, où chaque colonne est comme un tiroir, avec chaque élément fournissant un certain réseau de propriétés. Tandis qu'un groupe fournit une gamme des métaux ioniques qui sont de bons conducteurs, ils peuvent être employés en même temps que les oxydes d'un autre groupe pour faire une pile. D'autres groupes fournissent des semi-conducteurs, et les halogènes, par exemple, assurent une gamme d'additifs pour faire les lampes fluorescentes de diverses couleurs. Le tableau périodique peut être vu comme un entrepôt accessiblement organisé, ou un catalogue à partir duquel on peut commander les éléments de base pour construire de nouveaux matériaux qui remplissent les fonctions désirées. C'est certainement l'impression donnée par le chimiste Michel Pouchard: Le chimiste est principalement l'architecte de la matière aussi bien que son maçon ; sa balance est celle du nanomètre, ses briques environ les cent éléments dans le système périodique de Mendeleïev, et son ciment est leurs électrons de valence [1].

## **I. 11 Conclusion**

On objectera aisément à ce qui précède que même si la réduction n'existe pas pour le moment elle peut jouer le rôle d'un idéal, d'un but à atteindre. N'a-t-elle pas stimulé le développement de la chimie et de la physique au 19<sup>e</sup> siècle ? N'inspire-t-elle pas aujourd'hui la biologie? Examinons cela. L'idée d'une réduction verticale qui, de l'antiquité aux lumières, a été une thèse philosophique (l'atomisme) s'incarne au 19<sup>e</sup> siècle dans l'hypothèse atomique des chimistes, puis dans la théorie cinétique des gaz, dans la mécanique statistique, et finalement dans la "réalité moléculaire" (selon l'expression de Jean Perrin). La théorie des quanta puis la mécanique quantique donnent à l'atomisme des physiciens et des chimistes une justification plus profonde ; elles expliquent que, comme le disait Bohr, il existe dans la nature une tendance à produire des formes déterminées et à les faire réapparaître, encore et toujours, même lorsqu'elles ont été perturbées ou détruites. Mais, en triomphant, l'atomisme a perdu sa simplicité originelle. Les atomes ne sont plus insécables et inaltérables. L'existence des molécules n'est plus un postulat de base de la chimie, mais un résultat de la théorie quantique, valable sous certaines réserves. La postérité de l'atomisme s'est scindée en deux branches bien distinctes : d'un côté la théorie savante, et les techniques récentes de manipulation d'atomes ou de molécules presque isolées ; de l'autre, l'idée de réduction verticale qui n'est plus une thèse philosophique ou une hypothèse scientifique, mais est devenue une croyance.

Imaginons que la chimie, la physique du solide, la thermodynamique, la mécanique classique enfin, soient toutes remplacées par la mécanique quantique, devenue entièrement calculable après l'avènement d'ordinateurs d'un type nouveau. L'industrie informatique réaliserait des chiffres d'affaires fabuleux. Mais l'idéal d'intelligibilité qui a présidé à la naissance de la science serait mort et enterré, au moins dans les sciences physiques ; car la mécanique quantique, qui calcule par ses propres moyens et fournit des résultats d'une exactitude numérique étonnante, ne peut guère se passer des béquilles de la physique classique quand on lui demande de nous faire comprendre ce qui se passe. Par suite, les sciences physiques, débarrassées désormais de toute incertitude et de toute nouveauté véritable, sombreraient dans un océan d'ennui. En somme, même si la réduction devenait possible dans les sciences physiques, elle ne les aiderait pas à remplir ce qui doit rester la tâche fondamentale de toute science : nous aider à comprendre les phénomènes [7]. La mécanique quantique a montré à quel point la séparation entre la physique et la chimie était arbitraire [10].

Friedrich Paneth a clairement défini le statut épistémologique du concept chimique de l'élément dans une conférence portant ce titre qu'il a donné en 1931. Ici, il a soulevé deux questions. La première rappelle le problème d'Aristote avec « les mixtes véritables », en quel sens peut-on supposer que les éléments persistent dans les composés? " Sa réponse - basée sur une étude historique des notions d'éléments et des atomes - est claire: puisque la chimie est concernée par les qualités secondaires des substances, les chimistes devraient assumer la persistance des qualités dans les composés. Il a proposé d'appeler ce qui persiste " substance de base " alors qu'il réservait le terme " substance simple " pour la matière montrant des manifestations phénoménologiques associées à cette substance de base abstraite, renouvelant de ce fait la distinction conceptuelle de Mendeleev. La deuxième question de Paneth était: « s'il est vrai ou non que la chimie doit et va se dissoudre dans la physique? " Ici, il a traité la possibilité de réduire la chimie à la physique, un développement qui avait été déjà prédit par certains physiciens.

La théorie quantique offre clairement des avantages considérables pour l'enseignement de la chimie et, en particulier, pour la présentation du tableau périodique, avec son explication du comment et du pourquoi les propriétés chimiques des divers éléments obéissent à un modèle global. Néanmoins, ceci ne signifie pas que les chimistes peuvent s'en sortir seuls avec la mécanique quantique. Le chimiste ne peut pas s'isoler de ces développements et de se contenter d'exécuter et de décrire les réactions. Il ne pourra pas tirer parti de ses succès précédents à moins qu'il assimile les disciplines du physicien ou accepte de collaborer avec

lui. Cette évolution de la chimie vers la physico-chimie est est le trait le plus caractéristique de ces vingt dernières années [1].

**I. 12. Références**

- [1] Bernadette Bensaude-Vincent, Jonathan Simon, « CHEMISTRY THE IMPURE SCIENCE », Imperial College Press, (2008), ISBN-13 978-1-84816-225-9.
- [2] J. Perrin, « *Les Atomes* », Flammarion, (1991).
- [3] Charles Widmer, « Droits de l'homme et sciences de l'homme pour une Ethique anthropologique », Librairie Droz S. A., Genève - Paris (2007).
- [4] Bernard JACROT, Eva PEBAY-PEYROULA, Régis MACHE, Claude DEBRU, « PHYSIQUE ET BIOLOGIE, UNE INTERDISCIPLINARITÉ COMPLEXE », EDP Sciences, (2006), ISBN 2-86883-892-8.
- [5] Lambert D., 1999. « Sciences et théologie Les figures d'un dialogue », Ed. Lessius, Presses Universitaires de Namur, 220p.
- [6] Edgar Ascher, « L'idée d'émergence », Archives de psychologie, **52**, (1984), 175-193.
- [7] À René Thom, « Passion des formes : dynamique qualitative, sémiophysique et intelligibilité », Vol 1, ENS Editions, (1994) 828 p.
- [8] Pierre Darriulat, « Réflexions sur la science contemporaine », EDP Sciences (2007), ISBN 978-2-86883-964-0.
- [9] Eric R. Scerri, « Collected papers on philosophy of chemistry », Imperial College Press, (2008). ISBN-13 978- 1-84816- 137-5.
- [10] Champ Vallon, « La totalité », vol 5, (2002), Ed. Champ Vallon, Presses Universitaires de France, ISBN 2-87673-355-2.

## CHAPITRE II

Le tableau périodique



## II.1. Introduction

Le tableau périodique se classe comme l'une des idées les plus fécondes et unificatrices dans toute la science moderne [1]. L'élément chimique est proche d'une abstraction mathématique. Ainsi, il devient un instrument qui nous permet de construire une série qui est supposée se cacher derrière. Cette abstraction permet au chimiste de décrire l'ordre qui existe derrière la multiplicité apparente chaotique du monde phénoménal. La plus grande valeur du concept de Mendeleev d'un élément est sa puissance de construire la série apparentée. Les éléments chimiques sont efficacement "désincarnés" les entités abstraites qui sont attachées au vrai monde par le réseau des rapports décrits par le système périodique. Ce rapport abstrait-concret est représenté par la position de l'élément dans le tableau périodique, qui forme alternativement la base pour une multiplicité d'interprétation [2].

## II.2. Historique

Tout a commencé par la notion d'élément, puis l'analyse de la réactivité des éléments chimiques a amené bon nombre de scientifiques à proposer des systèmes de classement basés sur l'analogie des réactivités. Le tout premier système de classement, publié en 1829, celui de **Johann Döbereiner** de l'Université de Iéna et Sa "**Loi des triades**". **La première classification périodique** qu'il nomma "**la vis tellurique**", fut celle de **Beguyer de Chancourtois** en 1862. Dès 1857, **William Odling** s'intéressa au classement des éléments connus en développant la **règle des triades**. Il publia en 1864 une classification assez proche de celle que Dmitri Mendeleïev publia cinq ans plus tard.

En 1863, **John Newlands** publia dans "Chemical News" une classification de 56 éléments en 11 groupes basés sur la similitude de leurs propriétés chimiques, il observa aussi que dans un même groupe deux éléments consécutifs avait une masse atomique différente d'un multiple de huit, Newlands introduisit une **loi des octaves** laissant entendre que tous les huit éléments on retrouve des propriétés chimiques similaires.

### II.2.1. Lothar Meyer ou Dimitri Mendeleïev ?

L'allemand Meyer et le russe Mendeleïev ont établi, chacun de leur côté, un classement des éléments chimiques. Meyer proposa une classification périodique dès 1864, mais celle-ci ne fut publiée qu'en 1870, un an après celle de Mendeleïev.

Le classement se fit sur la base de la masse atomique croissante (les numéros atomiques n'étaient pas encore inventés) en conjonction avec la similitude des propriétés chimiques. L'originalité des tableaux de Mendeleïev et de Meyer fut de laisser des emplacements vides, prévoyant ainsi la prochaine découverte de nouveaux éléments [3].

C'est au chimiste russe, **Dimitri Ivanovitch Mendeleïev** (1834-1907), que revient le mérite d'avoir structuré une classification cohérente de l'ensemble des éléments. En rédigeant un volume de chimie, il se rend compte qu'en plaçant les éléments en ordre croissant de leur masse atomique, il apparaît une tendance en ce qui concerne les propriétés, tendance qui se répète à plusieurs reprises. En 1869, il présenta une première version de son tableau périodique. Celui-ci contient alors les 63 éléments connus à cette époque. Mendeleïev disposa son tableau de manière que la périodicité apparût nettement. En classant les éléments verticalement (ils sont disposés horizontalement dans la classification actuelle). Les rangées horizontales se succèdent régulièrement au fur et à mesure que certaines propriétés chimiques et physiques se répètent. Dans les rangées verticales, on retrouve des éléments possédant à peu près les mêmes propriétés chimiques et des ressemblances dans leurs propriétés physiques. Pour respecter la loi périodique à laquelle Mendeleïev croyait fermement, il dut parfois modifier l'ordre déterminé par la progression des masses atomiques et laisser certaines cases vides. Jusque là, très peu de scientifiques acceptaient les idées de Mendeleïev. Mais lorsque ces éléments prédits par ce dernier furent découverts, présentant de surcroît des propriétés très proches de celles qu'il avait prévues, les scientifiques reconnurent l'utilité de son tableau périodique. Bien que la classification de Mendeleïev marquât un net progrès sur tous les autres essais de classification d'alors, il reste qu'elle contenait certaines anomalies dues à des masses atomiques encore mal déterminées à l'époque [4].

### **II.2.2. Le tableau périodique des éléments**

Le tableau périodique des éléments répertorie les éléments chimiques selon l'ordre croissant de leur nombre atomique. Le tableau périodique est constitué de rangées horizontales, nommées périodes et de rangées verticales, nommées colonnes (groupes d'éléments).

Il est important de noter que:

- les périodes sont en nombre de 7;
- les groupes d'éléments sont en nombre de 18; dont les propriétés présentent certaines tendances [5].

|        |          |             |           |            |            |
|--------|----------|-------------|-----------|------------|------------|
|        |          |             | Ti = 50   | Zr = 90    | ? = 180.   |
|        |          |             | V = 51    | Nb = 94    | Ta = 182.  |
|        |          |             | Cr = 52   | Mo = 96    | W = 186.   |
|        |          |             | Mn = 55   | Rh = 104,4 | Pt = 197,4 |
|        |          |             | Fe = 56   | Ru = 104,4 | Ir = 198.  |
|        |          |             | Ni = 59   | Pd = 106,6 | Os = 199.  |
|        |          |             | Cu = 63,4 | Ag = 108   | Hg = 200.  |
| H = 1  | Be = 9,4 | Mg = 24     | Zn = 65,2 | Cd = 112   |            |
|        | B = 11   | Al = 27,4   | ? = 68    | Ur = 116   | Au = 197?  |
|        | C = 12   | Si = 28     | ? = 70    | Sn = 118   |            |
|        | N = 14   | P = 31      | As = 75   | Sb = 122   | Bi = 210?  |
|        | O = 16   | S = 32      | Se = 79,4 | Te = 128?  |            |
|        | F = 19   | Cl = 35,5   | Br = 80   | J = 127    |            |
| Li = 7 | Na = 23  | K = 39      | Rb = 85,4 | Cs = 133   | Tl = 204.  |
|        |          | Ca = 40     | Sr = 87,6 | Ba = 137   | Pb = 207.  |
|        |          | ? = 45      | Ce = 92   |            |            |
|        |          | ? Er = 56   | La = 94   |            |            |
|        |          | ? Yt = 60   | Di = 95   |            |            |
|        |          | ? In = 75,6 | Th = 118? |            |            |

Figure. II.1 : La première forme du système périodique de Mendeleev datant du 17 février 1869, intitulée « une tentative au système d'éléments, se basant sur leur masse.



Figure. II.2 : Dimitri Ivanovitch Mendeleev [1].

Il en résulte que nous pouvons acquérir une connaissance raisonnable d'un élément si nous savons à quel groupe il appartient et quels sont ses voisins [6].

- les groupes d'éléments contiennent des éléments chimiques qui présentent des propriétés chimiques similaires;
- le 1<sup>er</sup> groupe d'éléments regroupe les **métaux alcalins** ;
- le 2<sup>e</sup> groupe d'éléments regroupe les **métaux alcalino-terreux** ;
- le 17<sup>e</sup> groupe d'éléments regroupe les **halogènes** ;
- le 18<sup>e</sup> groupe d'éléments regroupe les **gaz rares** ;
- les groupes d'éléments 3 à 12 compris regroupent les **métaux de transition** [5].
- Le long bloc situé au-dessous est constitué par les métaux de **transition interne**. La rangée supérieure de ce bloc la 6<sup>e</sup> période, est constituée par les **lanthanides** et la rangée inférieure la 7<sup>e</sup> période, par les actinides [6].

Les colonnes sont occupées de telle manière que chacune d'entre elles renferme tous les éléments dont les atomes ont la même configuration électronique de valence. Comme ce sont les électrons de valence qui sont responsables des propriétés chimiques des éléments, une colonne regroupe tous les membres d'une même famille chimique.

On distingue également des blocs, chaque bloc correspond au remplissage d'un type de sous-couche :

- Le bloc *s* correspond au remplissage de sous-couches *s* (colonnes 1 et 2).
- Le bloc *p* à celui des sous-couches *p* (colonnes 13 à 18)
- Le bloc *d* à celui des sous-couches *d* (colonnes 3 à 12)
- Le bloc *f* à celui des sous-couches *f* (les deux lignes sous le tableau) [7].

### II.2.3. Quantification de l'énergie

#### Les nombres quantiques

En mécanique quantique, l'état d'un électron d'un atome peut être décrit à l'aide de quatre nombres dits *quantique* et notés:  $n$ ,  $l$ ,  $m_l$  et  $m_s$ .

- $n$  est appelé nombre quantique principal. C'est un nombre entier positif.
- $l$  est appelé nombre quantique secondaire ou azimutal. C'est un nombre entier positif ou nul inférieur ou égal à  $n-1$ :  $0 \leq l \leq n-1$  [7].

Les valeurs du nombre quantique azimutal,  $l$ , sont désignées respectivement par des lettres empruntées à la spectroscopie [6].

|                  |          |          |          |          |          |
|------------------|----------|----------|----------|----------|----------|
| $l$              | 0        | 1        | 2        | 3        | 4        |
| niveau d'énergie | <i>s</i> | <i>p</i> | <i>d</i> | <i>f</i> | <i>g</i> |

**Tableau II.1 :** Désignation des valeurs du nombre quantique azimutal,  $l$  [7].

#### **II.2.4. Niveaux énergétiques des orbitales atomiques**

Pour les atomes polyélectroniques, les niveaux énergétiques des orbitales atomiques dépendent à la fois des valeurs de  $n$  et de  $l$ . Autrement dit, les orbitales caractérisées par la même valeur de  $n$  mais par des valeurs différentes de  $l$ , possèdent des énergies différentes.

- les niveaux énergétiques des orbitales atomiques, précisés par une des valeurs de  $n$ , sont appelés couches électroniques. Selon la valeur de  $n$ , les couches électroniques sont désignées par une des lettres majuscules suivantes : K, L, M, N, O, P, Q.

- les niveaux énergétiques des orbitales atomiques, précisés par les diverses valeurs de  $l$ , sont appelés sous-couches électroniques.

|                     |   |
|---------------------|---|
| Couche électronique | sous-couches électroniques  |
| K ( $n = 1$ )       | $l$ sous-couche $s$ ( $l = 0$ )   |
| L ( $n = 2$ )       | $l$ sous-couche $s$ ( $l = 0$ ) et $l$ sous-couche $p$ ( $l = 1$ )  |
| M ( $n = 3$ )       | $l$ sous-couche $s$ ( $l = 0$ ), $l$ sous-couche $p$ ( $l = 1$ ), $l$ sous-couche $d$ ( $l = 2$ )                                     |
| N ( $n = 4$ )       | $l$ sous-couche $s$ ( $l = 0$ ), $l$ sous-couche $p$ ( $l = 1$ ), $l$ sous-couche $d$ ( $l = 2$ ), et $l$ sous-couche $f$ ( $l = 3$ ) |

**Tableau II.2 :** Niveaux énergétiques des orbitales atomiques [5].

#### **II.2.5. La configuration électronique d'un atome**

Pour établir la configuration électronique d'un atome polyélectronique dans son état fondamental, trois règles doivent être appliquées: le principe de Pauli, la règle de Klechkowski et la règle de Hund.

### II.2.5.1. Principe d'exclusion de Pauli

Dans un édifice monoatomique, deux électrons ne peuvent pas avoir leurs quatre nombres quantiques ( $n, l, m_l, m_s$ ) identiques.

Ainsi, deux électrons qui occupent une sous-couche 1s sont décrits par le quadruplet (1, 0, 0, + 1/2) pour l'un et (1, 0, 0, - 1/2) pour l'autre.

Ils ont la même énergie, mais leurs nombres magnétiques sont de spin opposés; on dit que leurs spins sont *anti-parallèles* ou *appariés*.

#### **Une orbitale atomique ne peut donc pas décrire plus de deux électrons.**

Ainsi, une sous-couche *s* peut contenir au plus deux électrons, une sous-couche *p* au plus six électrons, une sous-couche *d* au plus dix électrons, une sous-couche *f* au plus quatorze électrons.

- Une sous-couche qui contient le maximum d'électrons qui lui est permis de contenir est dite saturée [7]. On peut résumer le principe d'exclusion de Pauli comme ceci: Dans chaque couche, l'on peut mettre au maximum  $2n^2$  électrons. (où  $n$  est le numéro de la couche) [8].

### II.2.5.2. La règle de Klechkowski

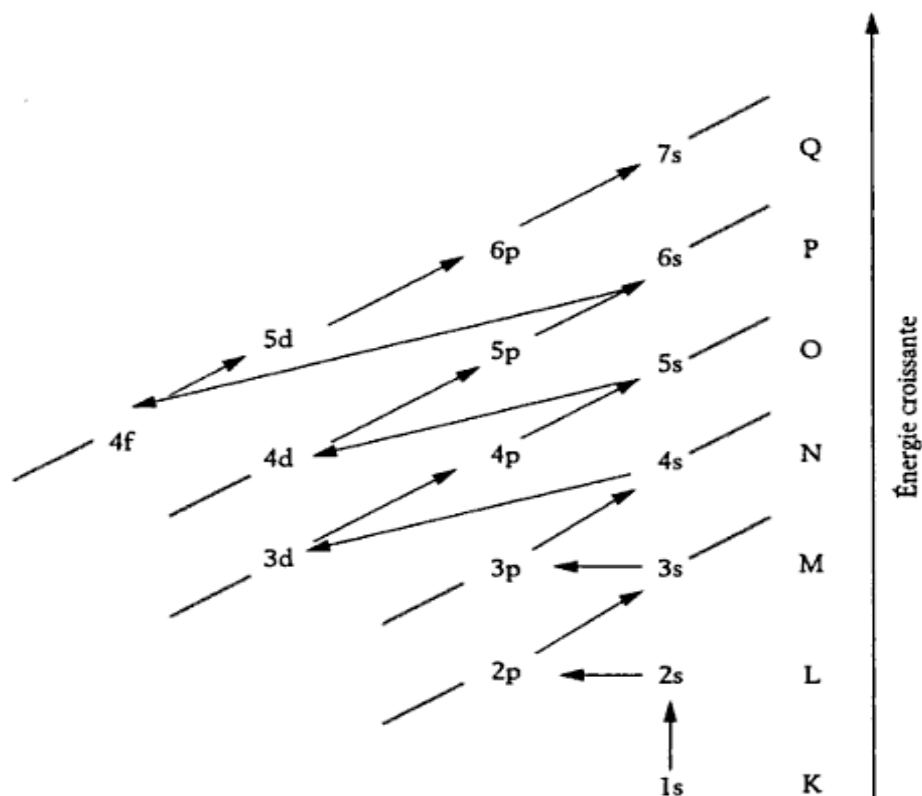
À l'état fondamental, l'énergie de l'atome est minimale : cela correspond à une occupation des niveaux d'énergie électronique les plus bas. La règle de Klechkowski (Appelé principe de stabilité) est une règle empirique qui permet de retrouver l'ordre de remplissage des sous-couches afin d'obtenir la configuration électronique d'un atome dans son état fondamental. Dans l'ordre :

$1s < 2s < 2p < 3s < 3p < 4s < 3d < 4p < 5s < 4d < 5p < 6s < 4f < 5d < 6p < 7s < 5f < 6d < 7p < 8s < \dots$  [9].

Dans un atome polyélectronique, l'ordre de remplissage des sous-couches (caractérisées par les nombres quantiques  $n$  et  $l$ ) est celui pour lequel la somme ( $n + l$ ) croît. Quand deux sous-couches ont la même valeur pour la somme ( $n + l$ ), la sous-couche qui est occupée la première est celle dont le nombre quantique principal  $n$  est le plus petit [7].

#### Remarque

La règle de Klechkowski est parfois appelée principe de construction (ou principe de l'Aufbau) [7].



**Figure. II.3. :** Hiérarchie des niveaux énergétiques des orbitales atomiques (visualisée par la règle de Klechkowski) [5].

### II.2.5.3. Règle de Hund

Dans des orbitales de même énergie (dans une même couche), les électrons se placent de préférence avec leurs spins parallèles (même chiffre) dans des orbitales différentes. Ceci s'expliquant par un gain d'énergie [8].

## II.3. Principe de la classification moderne

Cette classification comporte 7 lignes (ou périodes) et 18 colonnes. Les éléments sont rangés dans chaque ligne par ordre croissant du numéro atomique  $Z$ , selon la configuration électronique des atomes [10], suivant la règle de Klechkowski où le remplissage des orbitales se fait par valeurs croissantes des niveaux énergétiques  $n + l$ .

### II.3.1. La loi de la périodicité des propriétés

Bien avant la découverte de la nature atomique de la matière, les chimistes avaient remarqué que, lorsque les éléments sont classés par ordre croissant des « masses équivalentes » (masses atomiques), une périodicité de leurs propriétés apparaissait. C'est Mendeleïev qui proposa la classification qui devait se trouver expliquée par la découverte de la structure atomique de la matière.

La comparaison des couches de valence (C.V.) des différents éléments conduit très vite à constater que leur structure électronique se répète pour des groupes d'éléments (familles), la seule différence étant la valeur du nombre quantique principal « n » impliqué dans la couche de valence. Ainsi, à une colonne donnée, correspondent des éléments ayant une couche de valence de même type, c'est-à-dire possédant des propriétés comparables. La structure du tableau périodique des éléments traduit ces propriétés remarquables. Dans ce tableau, les éléments situés dans une même colonne constituent une famille d'éléments : ceux-ci ont la même structure de la couche de valence. Ceux qui sont situés sur une même ligne constituent une période : leur couche de valence correspond à une même valeur du nombre quantique « n » [11].

Il existe une corrélation entre la position d'un élément chimique dans le tableau périodique et sa configuration des électrons de valence (électrons de la couche la plus externe) [5].

Par ailleurs, quatre remarques permettent de comprendre l'évolution des propriétés des éléments en fonction de leur position dans le tableau :

1 - Les éléments possédant une couche de valence 0 sont particulièrement stables (gaz rares ou gaz nobles). Une conséquence est que les éléments auront tendance à réaliser cette configuration de couche de valence soit en captant, soit en perdant des électrons ; ils forment alors des ions.

2 - Moins des éléments possèdent d'électrons sur leur couche de valence, plus ils ont tendance à perdre des électrons pour réaliser la configuration de la couche de valence du gaz rare les précédant dans la classification : ils ont donc tendance à donner des ions positifs (cations). Ces éléments sont appelés « **métaux** ».

3 - Plus des éléments possèdent d'électrons sur leur couche de valence, plus ils ont tendance à gagner des électrons pour réaliser la configuration de la couche de valence du gaz rare les suivant dans la classification : ils ont tendance à donner des ions négatifs (anions). Ces éléments sont les « **non-métaux** ».

4 - Plus un élément est « gros », c'est-à-dire plus il y a d'électrons dans son nuage électronique entre un électron de valence et le noyau, moins l'interaction entre noyau et électrons de valence sera forte. Autrement dit, quand, dans une même famille, on se déplace de l'élément le plus léger vers les éléments plus lourds, le caractère métallique devient plus prononcé et celui de non-métal l'est de moins en moins [11].

### **II.3.2. La loi périodique**

La loi périodique indique qu'après certains intervalles réguliers, les éléments chimiques montrent une répétition approximative dans leurs propriétés. Par exemple, le fluor, le chlore, et le brome, qui appartiennent au groupe VII (des halogènes), partagent la propriété de former des sels cristallins blancs de formule générale  $\text{NaX}$  avec le sodium. Cette répétition périodique de propriétés est le fait essentiel qui sous-tend tous les aspects du système périodique. Cette présentation de la loi périodique soulève des questions philosophiques intéressantes. Tout d'abord, la périodicité entre les éléments n'est ni constante ni exacte. La première ligne comporte deux éléments, la deuxième et la troisième chacune contient huit, la quatrième et la cinquième contient 18, et ainsi de suite. Parmi les éléments, cependant, non seulement la longueur de période varie, mais également la périodicité n'est pas exacte. Les éléments dans chaque colonne de la classification périodique ne sont pas en exacte récurrence les uns des autres. Le système périodique est la notion la plus abstraite soutenant qu'il existe une relation fondamentale entre les éléments [1].

### **II.3.3. La compréhension du système périodique**

Les deux théories importantes de la physique moderne sont la théorie de la relativité d'Einstein et la mécanique quantique. La théorie quantique est en fait née dans l'année 1900, Environ 14 années avant la découverte du numéro atomique. Elle a été appliquée à des atomes par Niels Bohr, qui a poursuivi la notion que les similitudes entre les éléments dans n'importe quel groupe du tableau périodique pourrait s'expliquer qu'ils ont un nombre égal d'électrons périphériques. L'explication la plus récente du système périodique est donnée en termes du nombre d'orbitales peuplées par les électrons, cette explication dépend de la configuration électronique de l'atome [1].

## II.4. Analyse du tableau périodique:

### II.4.1. Les anomalies du tableau périodique

L'Analyse par période du tableau périodique révèle quelques anomalies et contradictions. Les anomalies du tableau périodique sont les éléments qui ne respectent pas la règle de Klechkowski, dans le remplissage des orbitales par énergie croissante. Beaucoup d'anomalies dans les groupes peuvent être expliquées par la taille, ou au privilège des orbitales (d ou f) à moitié ou totalement remplies. Parmi ces éléments qui présentent des configurations électroniques faisant exception à la norme:

#### La première période : H, He

Le nombre quantique principal  $n = 1$  n'autorise que la sous-couche  $1s$ , de nombre quantique secondaire  $l=0$ . L'hydrogène  ${}^1\text{H}$ , de configuration électronique  $1s^1$ , et l'hélium  ${}^2\text{He}$ , de configuration électronique  $1s^2$ .

**L'hydrogène** se place dans la case de gauche, au-dessus de la famille des alcalins, bien que ses propriétés diffèrent largement de celles des alcalins. L'hydrogène est particulier il n'est conforme à aucun groupe, dans certains tableaux il est placé au dessus du groupe 1, dans d'autres dans le groupe 17, et d'autres encore le placent dans ces deux groupes. Dans certains tableaux il n'est placé dans aucun groupe, on le place tout seul au sommet du tableau périodique [6].

L'Hélium se trouve dans le bloc p avec les gazs « rares » (colonne 18), alors qu'il est de type s, et qui devait être dans la seconde colonne comme l'inciterait un classement continu par valeurs croissantes de Z.

#### La quatrième période :

Dans cette période deux éléments font exception à la règle de Klechkowski lors de l'occupation du sous-niveau  $3d$  : le chrome et le cuivre qui présentent respectivement les configurations de valence  $3d^5 4s^1$  et  $3d^{10} 4s^1$  au lieu des configurations  $3d^4 4s^2$  et  $3d^9 4s^2$  attendues. Elles s'expliquent par le fait que les sous-couches saturées ou à demi remplies procurent une stabilisation particulière aux configurations correspondantes.

#### La cinquième période :

Les exceptions à la règle de Klechkowski sont ici plus fréquentes qu'au cours de la Quatrième période, puisqu'au nombre de six, car les énergies des orbitales  $5s$  et  $4d$  sont très proches.

### La sixième période :

Le lutécium et le lawrencium sont placés dans le bloc f, alors que leur couche f est déjà saturée, ils contiennent 1 électron 5d ou 6d. Ils constituent bien le début des éléments de transition.

## **II.4.2. Propriétés physiques et chimiques des éléments**

### **II.4.2.1. L'électronégativité**

Il a été possible de chiffrer le comportement de l'élément, par rapport à l'électron grâce au concept d'électronégativité qui mesure la tendance d'un atome, à l'état fondamental, à retenir ses électrons de valence et à attirer des électrons supplémentaires [11].

L'électronégativité est une grandeur relative qui traduit l'aptitude d'un atome  $B$  à attirer vers lui le doublet électronique qui l'associe à un autre atome  $A$  [7].

L. Pauling a introduit le concept d'électronégativité, qui est un concept fondamental pour la chimie, dès 1932, il l'a défini comme le pouvoir d'un atome, de capter un électron lorsqu'il est impliqué dans une liaison chimique, au sein d'une molécule. Néanmoins, son expression en fonction de grandeurs mesurables n'est pas évidente et un certain nombre de relations ont été proposées [12]. L'électronégativité croît lorsqu'on se déplace de la gauche vers la droite et du bas vers le haut de la classification périodique.

Le fluor est l'élément le plus électronégatif et le césium le moins électronégatif. Bien qu'elle ne soit déterminée qu'empiriquement, l'électronégativité constitue une notion fondamentale en chimie, en particulier pour l'étude de la réactivité des composés [7]. L'électronégativité gouverne une grande partie de la chimie des éléments [1].

L'effet de la diagonale peut être expliqué comme le résultat de plusieurs tendances opposées. Quand on descend chaque groupe, l'électronégativité diminue. Mais, quand on se déplace à travers le tableau, la même propriété augmente. Si l'on se déplace en diagonale les deux tendances s'annulent mutuellement, et il y a peu de changement en électronégativité. De même, les tendances de l'énergie d'ionisation et du rayon atomique changent légèrement selon la diagonale [1].

### **II.4.2.2. Le rayon atomique**

On peut tout d'abord considérer un atome libre et définir un rayon atomique comme étant la distance au centre du noyau du maximum de la densité radiale de l'orbitale occupée la

plus externe. Ces maxima ont été calculés par différents auteurs, selon la méthode de Hartree-Fock-Slater. On trouve ainsi 110 pm pour le potassium dont l'orbitale la plus externe est l'orbitale 4s (rayon orbitalaire). Mais, les orbitales monoatomiques des atomes libres n'ont aucune raison de se retrouver dans les descriptions de la molécule, et encore moins sous forme liée dans un solide ; l'atome peut perdre ou gagner partiellement des électrons et se déformer sous l'action d'un champ électrique, en particulier celui des entités chargées qui l'entourent [11].

#### II.4.2.3. L'énergie (ou potentiel) d'ionisation

Les énergies d'ionisation des éléments chimiques dépendent de leur configuration électronique à l'état fondamental. Autrement dit, les énergies de liaison des éléments ont un caractère périodique.

Par définition, l'énergie d'ionisation,  $I$ , d'un atome neutre ou d'un ion est l'énergie minimale qu'il faut fournir à l'atome ou à l'ion, se trouvant à l'état gazeux dans son état fondamental, pour leur arracher un électron, et former un ion positif [5].

C'est l'énergie correspondant à la réaction :



Ce qu'on appelle l'énergie ou le potentiel de première ionisation ( $I_1$ ) De même,  $I_2$ , potentiel de deuxième ionisation, est l'énergie nécessaire pour expulser un deuxième électron, et ainsi de suite [12].

- L'énergie de première ionisation est toujours positive, elle est d'autant plus grande que l'électron arraché est plus fortement lié au noyau [7].

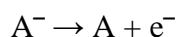
- Les énergies d'ionisation d'un atome gazeux augmentent dans l'ordre :  $I_1 < I_2 < I_3 < \dots < I_n$  [5].

Remarque : Un atome gazeux est libre de toute interaction.

#### II.4.2.4. L'affinité électronique

L'affinité électronique  $A$  est égale à l'énergie nécessaire pour arracher un électron à l'anion gazeux.

C'est l'énergie correspondant à la réaction :



Cette réaction peut être exothermique ou endothermique. Par convention,  $A$  est positive si la réaction est exothermique.

## II.5. Conclusion

À mesure qu'augmentait le nombre d'éléments connus, les chimistes du XIX<sup>e</sup> siècle éprouvaient le besoin de les ordonner. On connaissait les masses atomiques et on avait déjà l'habitude de classer les éléments d'après l'augmentation de cette caractéristique. La simple progression des masses atomiques ne pouvait toutefois expliquer logiquement les différents comportements des éléments [4].

Il est en effet une sorte de miracle que la mécanique quantique explique le tableau périodique. Mais nous ne devrions pas se laisser séduire en croyant que c'est une explication déductive. Une chose qui est claire: est que la tentative d'expliquer les détails de la classification périodique continue à défier l'ingéniosité des physiciens quantiques et chimistes quantiques et que le tableau périodique continuera de présenter un test concernant la pertinence des nouvelles méthodes développées en chimie quantique. La réduction de la chimie à la mécanique quantique n'a ni complètement échoué, comme certains philosophes de la science ont revendiqué, et elle n'a pas été une réussite complète, comme certains historiens contemporains ont revendiqué [1].

**II.6. Références**

- [1] Eric R. Scerri, « The periodic table : its story and its significance », Published by Oxford University Press, Inc, (2007).
- [2] Bernadette Bensaude-Vincent, Jonathan Simon, « Chemistry the impure science », Imperial College Press, (2008), ISBN-13 978-1-84816-225-9.
- [3] Union des professeurs de physique et de chimie, <http://www.udppc.asso.fr/>.
- [4] Le repaire des sciences, <http://www.lerepairedessciences.fr>
- [5] Melania Kiel, Chimie générale, ed. ESTEM, (2003). ISBN : 2 84371 219 X.
- [6] Peter Atkins, Loretta Jones, « Chemistry. Molecules, Matter and Change », 3rd. ed., W. H. Freeman and Company, New York, New York and Basingstoke (1997). ISBN 2-7445-0028-3.
- [7] André Durupthy, Jacques Estienne, Magali Giacino, Alain Jaubert, Claude Mesnil, « Physique, Chimie et sciences de l'ingénieur » (PCSI), I.S.B.N. 978-2-0118-1752-5. [www.hachette-education.com](http://www.hachette-education.com)
- [8] Miseur Ludovic, « La Chimie.net », <<http://www.lachimie.net>>, (Novembre 2014).
- [9] Andrew Ede, « The chemical element : a historical perspective », Greenwood Publishing Group, Inc (2006). <http://www.sciences-en-ligne.com>.
- [10] <http://www.web-sciences.com>.
- [11] Jacques Le Coarer, « Chimie le minimum à savoir », EDP Sciences, (2003).
- [12] J. -F. Marucco, « Chimie des solides », EDP Sciences, (2004), ISBN: 2-86883-673-9.

## CHAPITRE III

Le data mining et ses techniques d'analyse de données

### **III. 1. Introduction**

Le data mining, dans sa forme et compréhension actuelle, à la fois comme champ scientifique et industriel, est apparu au début des années 90. On peut voir le data mining comme une nécessité imposée par le besoin des entreprises de valoriser les données qu'elles accumulent dans leurs bases. En effet, le développement des capacités de stockage et les vitesses de transmission des réseaux ont conduit les utilisateurs à accumuler de plus en plus de données. Certains experts estiment que le volume des données double tous les ans. Que doit-on faire avec des données coûteuses à collecter et à conserver? [1].

L'exploration de données est le processus de découvrir les modèles pertinents, intéressants et novatrices, ainsi que des modèles descriptifs, compréhensibles, et prédictifs de données à grande échelle [2].

### **III.2. Différentes définitions du Data Mining**

#### Première définition

Le Data Mining est un processus non-trivial d'identification de structures inconnues, valides et potentiellement exploitables dans les bases de données (Fayyad, 1996) [3].

#### Deuxième définition

Le Data Mining est une partie (ou une étape) du processus d'extraction de connaissances à partir de données (ECD) en Français, ou (KDD) (Knowledge discovery in Databases) en anglais.

#### Troisième définition

Le data mining est un processus qui fait intervenir des méthodes et des outils issus de différents domaines de l'informatique, de la statistique ou de l'intelligence artificielle en vue de découvrir des connaissances utiles [2].

### **III.3. Historique**

L'expression « *data mining* » est apparue vers le début des années 1960 et avait, à cette époque, un sens péjoratif. En effet, les ordinateurs étaient de plus en plus utilisés pour toutes sortes de calculs qu'il n'était pas envisageable d'effectuer manuellement jusque là.

Certains chercheurs ont commencé à traiter sans *a priori* statistiquement les tableaux de données relatifs à des enquêtes ou des expériences dont ils disposaient. Comme ils constataient que les résultats obtenus, loin d'être aberrants, étaient encourageants, ils furent incités à systématiser cette approche opportuniste.

Cette attitude opportuniste face aux données coïncida en France avec la diffusion dans le grand public de l'analyse de données dont les promoteurs, comme Jean-Paul Benzecri, ont également dû subir dans les premiers temps les critiques venant des membres de la communauté des statisticiens.

Le succès de cette démarche empirique ne s'est malgré tout pas démenti. L'analyse des données s'est développée et son intérêt grandissait en même temps que la taille des bases de données. Vers la fin des années 1980, des chercheurs en base de données, tel que Rakesh Agrawal, ont commencé à travailler sur l'exploitation du contenu des bases de données volumineuses, convaincus de pouvoir valoriser ces masses de données dormantes. Ils utilisèrent l'expression « *database mining* » mais, celle-ci étant déjà déposée par une entreprise (*Database mining workstation*), ce fut « *data mining* » qui s'imposa. En mars 1989, Shapiro Piatetski proposa le terme « *knowledge discovery* » à l'occasion d'un atelier sur la découverte des connaissances dans les bases de données.

La communauté de « *data mining* » a initié sa première conférence en 1995 à la suite de nombreux *workshops* sur le KDD entre 1989 et 1994. En 1998, s'est créé, sous les auspices de l'ACM, un chapitre spécial baptisé ACM-SIGKDD, qui réunit la communauté internationale du KDD. La première revue du domaine « *Data mining and knowledge discovery journal* » publiée par « Kluwers » a été lancée en 1997 [2].

#### **III.4. La démarche du Data Mining**

Le Data Mining est une démarche plus qu'une théorie, et qui comprend les étapes suivantes :

- 1- Formalisation des objectifs
- 2- Acquisition des données
- 3- Préparation des données
- 4- Apprentissage – application des méthodes
- 5- Interprétation – explication
- 6- Evaluation et validation
- 7- Déploiement [4].

### III.4.1. Valorisation des données

Au-delà de l'origine de la paternité de l'expression « *data mining* », nous allons maintenant nous intéresser à l'émergence de ce champ à la fois technologique et scientifique. L'exploitation des données pour en extraire des connaissances est une préoccupation constante de l'être humain car elle est une condition essentielle de son évolution. L'homme a toujours mémorisé sur des supports différents des informations qui lui ont permis d'inférer des lois. La biologie, la physique, la chimie ou la sociologie, pour ne citer que ces disciplines, font largement usage de l'approche empirique pour découvrir des lois et ou faire ressortir des éléments structurants dans des populations. La statistique est devenue une science dont l'objet est de donner un cadre rigoureux à la démarche empirique. C'est au sein de la statistique et du domaine des bases de données que le *data mining* a puisé ses outils. Les historiens des sciences, s'ils ne l'ont pas déjà fait, vont certainement nous proposer de nombreuses théories pour mieux situer le *data mining* dans le domaine des sciences [2].

### III.4.2. Définition d'un modèle

Construire des modèles a toujours été une activité des statisticiens. Un modèle est un résumé global des relations entre les variables, permettant de comprendre des phénomènes, et d'émettre des prévisions.

Le Data Mining ne traite pas d'estimation et de tests les modèles préspecifiés, mais elle découvre des modèles à l'aide d'un processus de recherche algorithmique d'exploration de modèles. En Data Mining les modèles ne sont pas issus d'une théorie mais de l'exploration des données. [5]

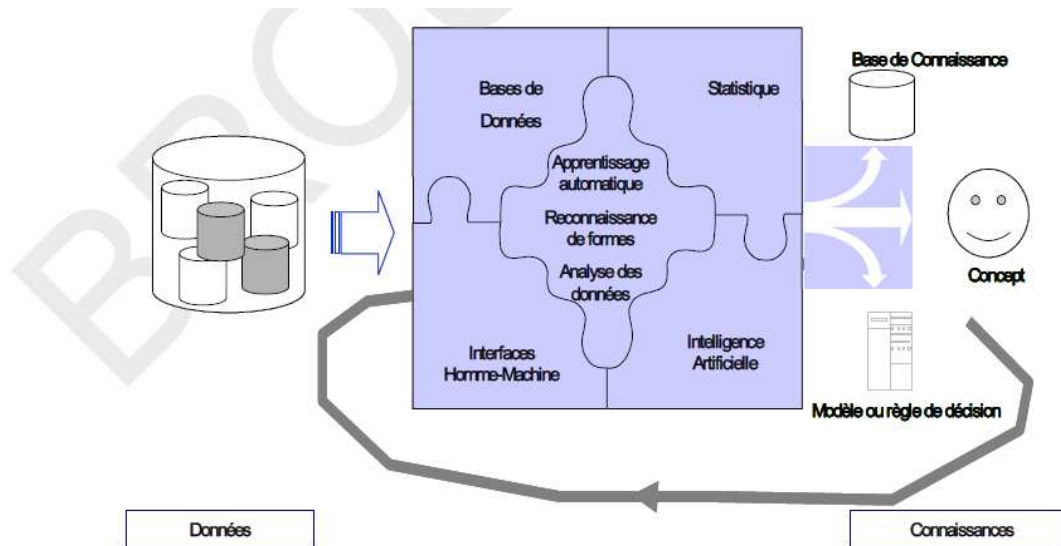


Figure III.1: Technologies et modèle général du data mining [2].

### III.4.3. Les modèles prédictifs

Ils ont pour but de déterminer une fonction (ou un modèle) qui associe des entrées et des sorties. Les régressions, les arbres de décision, sont les méthodes les plus utilisées. Ces méthodes sont des méthodes supervisées en ce sens qu'il faut préciser quelles sont les entrées et quelle est la sortie à prédire.

En fonction de la nature de la variable de sortie, deux sous catégories d'outils coexistent :

- **Si la variable de sortie est de type discret**, la **classification** aura pour rôle de construire le modèle qui permettra de classer correctement les enregistrements, c'est-à-dire d'assigner des catégories prédéfinies aux données.
- **Si la variable de sortie est de type continu**, l'**estimation** consistera à compléter une valeur manquante dans un champ particulier en fonction des autres champs de l'enregistrement. Les outils statistiques usuels de régression sont les plus employés.

### III.4.4. La segmentation

Est un apprentissage non supervisé (on ne définit pas ce qui est "entrées" et ce qui est "sorties") qui vise à identifier des ensembles d'éléments qui partagent certaines similarités. Les algorithmes de segmentation maximisent l'homogénéité à l'intérieur de chaque ensemble et maximisent l'hétérogénéité entre les ensembles. Différentes méthodes sont utilisées pour définir ces groupes.

### **III.4.5. La description (ou identification de modèles fréquents)**

Elle consiste à expliquer les relations existantes dans les données.

- Les méthodes cherchent à identifier les associations entre des variables.
- L'analyse des liens et les techniques de visualisations sont couramment utilisées dans ce but.
- Les techniques de visualisation sont utilisées pour simplifier la compréhension des données à l'aide de représentations graphiques adaptées.
- Les règles d'associations représentent des combinaisons de variables avec des niveaux prédéfinis de régularité [6].

### **III.4.6. Les deux types de techniques du Data Mining**

#### **III.4.6.1. Les techniques descriptives**

- Visent à mettre en évidence des informations présentes mais cachées par le volume des données.
- Réduisent, résument, synthétisent les données.
- Il n'y a pas de variable « cible » à prédire.

#### **III.4.6.2. Les techniques prédictives**

- Visent à **extrapoler de nouvelles informations** à partir des informations présentes (c'est le cas du scoring).
- Expliquent les données.
- Il y a une variable « cible » à prédire [7].

#### **III.4.6.3 Comparaison entre le data Mining et la statistique**

Cependant on s'aperçoit bien vite que les principales applications du Data Mining concernent plus la recherche de modèles que de structures comme les définit D.J.Hand « Un modèle est une représentation simplifiée des relations existantes entre les variables dans un but de synthèse et de prévision » [8 ,9]

Modéliser a toujours été une activité essentielle des statisticiens, et on peut se demander en effet ce qu'il y'a de nouveau avec le Data Mining.

- La réponse est qu'en Data Mining le modèle provient des données et n'est pas choisi a priori.
- Dans la pratique statistique habituelle, le modèle découle d'une théorie (économique, physique, biologique....) et le but est d'estimer et de tester les paramètres du modèle.
- En Data Mining le modèle final vient après une exploration combinatoire d'un grand nombre de modèles [10].

#### **III.4.6.4 Que peut-on trouver et prouver réellement avec le Data Mining ?**

Il faut donc se garder de certaines illusions : découvrir des structures « inattendues » est également une idée trompeuse.

On a d'autant plus de chances de trouver quelque chose d'intéressant que l'on connaît mieux ses données. [10]

#### **III.4.6.5 Conclusions**

- Le Data Mining est une discipline née en dehors de la statistique, dans le but de valoriser les bases de données.
- Le Data Mining offre des perspectives nouvelles pour la statistique et répond au défi du traitement des gigabases de données [5].

### **III.5. L'Analyse en Composantes Principales**

#### **III.5.1. Introduction**

L'Analyse en Composantes Principales (Principal Component Analysis) est une méthode statistique de projection de données multidimensionnelles, utilisée pour la réduction de la dimension de données, elle fait partie du groupe des méthodes descriptives appelées méthodes factorielles. La méthode est basée sur l'hypothèse qu'il existe de fortes corrélations entre les données étudiées. Le but de l'ACP est de condenser les données originelles en de nouveaux groupements, appelées nouvelles composantes, de façon à ce qu'elles ne présentent plus de corrélation entre elles et soient ordonnées en terme de pourcentage de variance apportée par chaque composante [11, 12, 13].

#### **III.5.2. Historique**

C'est H.Hotelling (1930) qui avait posé les fondements de l'analyse en composantes principales [11], en développant les travaux de C.Spearman [14], et de K.Pearson [12] qui dataient du début du siècle. Jusqu'aux années 60, en particulier par Jean-Paul Benzécri qui a beaucoup exploité les aspects géométriques et les représentations graphiques. [15] ces méthodes étaient perfectionnées et s'enrichissaient de variantes mais toutes restaient inabordables pour les praticiens car elles nécessitaient une masse considérable de calculs. Aujourd'hui, grâce à l'outil informatique, ces méthodes sont de plus en plus utilisées.

#### **III.5.3. L'analyse de données**

Aujourd'hui, l'analyse de données est couramment utilisée dans un bon nombre d'activités et de disciplines de base comme : la physique, la chimie,...etc. [11,14], et qui utilisent des « faits numériques » qu'on appelle **données**.

Avant toute analyse statistique, il est indispensable de consacrer du temps à examiner les données et d'en faire un tableau facilement exploitable.

La statistique classique s'est axée sur l'étude d'un nombre restreint de paramètres mesurés sur un petit ensemble d'individus .Elle a développé les notions d'estimations et de tests fondés sur des hypothèses probabilistes très restrictives. Les méthodes d'analyse de données permettent une étude globale des individus et des variables en utilisant

généralement des représentations graphiques suggestives. Les données peuvent être analysées selon plusieurs points de vue. La recherche des ressemblances ou des différences entre individus peut être un objet de l'analyse. En utilisant des méthodes factorielles comme l'Analyse en Composantes Principales, il est possible de représenter les proximités entre individus sur un même graphique [16].

#### **III.5.4. La nature des données** [17]

##### **III.5.4.1. Définition des variables**

Les variables correspondent aux intitulés des colonnes du tableau de données, elles sont également appelées paramètres ou caractères ; elles forment un espace dans lequel chaque individu est représenté par un point. Les coordonnées des points sont égales aux valeurs respectives des variables.

##### **III.5.4.2. Définition des individus**

L'individu est l'entité de base sur laquelle l'observateur réalise un certain nombre de mesures. Il est parfois appelé critère. Chaque individu est défini à partir de plusieurs **variables** quantitatives et/ou qualitatives.

- Une variable est quantitative lorsqu'elle prend ses valeurs sur une échelle numérique, plus précisément lorsque l'ensemble des valeurs qu'elle prend sur les individus est inclus dans l'ensemble des nombres réels et que l'on peut effectuer sur cette variable des opérations algébriques habituelles : additions, multiplications ...etc.
- Une variable est qualitative lorsqu'elle prend des modalités non numériques : appréciations, couleurs, descriptions, ... etc.

##### **III.5.4.3. Le centrage et la réduction des données brutes**

Dans ce paragraphe, nous présentons un exemple de codage très fréquemment utilisé dans le cadre de l'analyse de données : le codage des données brutes en données centrées et centrées réduites.

- Pour obtenir des données centrées, il faut remplacer  $x_{ij}$  par  $(x_{ij} - \bar{x}_j)$  où  $\bar{x}_j$  représente la moyenne arithmétique des valeurs prises par la variable  $x_j$ .

- Pour obtenir des données centrées réduites notées  $X_{ij}$ , on divise par l'écart type des valeurs prises par la variable  $x_j$ .

Ce qui donne :  $X_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma(x_j)}$

### III. 5.4.3.1. Les paramètres de dispersion

Les paramètres de dispersion sont des nombres qui mesurent la dispersion (ou la répartition) des individus autour d'un paramètre de position (par exemple la moyenne arithmétique). Ils s'expriment dans la même unité que les données et permettent de comparer des séries de même nature [18,19].

### III. 5.4.3.2. Caractéristique de la tendance centrale

Pour chacune des valeurs  $X_{ij}$ , on définit la **moyenne arithmétique** des valeurs observées ou mesurées à partir de la formule :

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

### III. 5.4.3.3. Caractéristique de la dispersion

La notion de dispersion complète les informations sur la tendance centrale. Il arrive que l'on traduise la dispersion des données à partir de l'**étendue** :

$$\text{Étendue} = X_{ij \max} - X_{ij \min}$$

### III. 5.4.3.4. La variance

La variance est un indicateur de dispersion reflétant l'importance des fluctuations des valeurs prise par une variable  $x_j$  de la série autour de leur moyenne arithmétique. Elle s'écrit :

$$\text{Var}(X_j) = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

- Pour les données centrées réduites, la **variance de chaque variable est égale à 1**.

### III. 5.4.3.5. L'écart type

Le calcul de l'étendue ne prend en compte que deux valeurs observées ; de plus il s'agit des valeurs extrêmes. On préfère caractériser la dispersion à partir de **l'écart type** expérimental

$\sigma(x_j)$  :

$$\sigma(x_j) = [\text{Var}(x_j)]^{1/2}$$

### III. 5.5. Qu'est-ce qu'une base de données ?

Une base de données (son abréviation est BD, en anglais DB, *database*) est une entité dans laquelle il est possible de stocker des données de façon structurée et avec le moins de redondance possible. Ces données doivent pouvoir être utilisées par des programmes, et par des utilisateurs différents. [20]

Ces bases de données sont généralement construites dans le but soit d'archiver les informations relatives à un domaine particulier, soit d'apporter une réponse à un problème précis, à l'image des réponses issues de sondages.

### III. 5.6. Principe de l'Analyse en Composantes Principales (ACP)

Les données sont rangées dans un tableau comportant souvent un grand nombre d'individus **n**, et un grand nombre de variables **p** ; la simple lecture de ce tableau ne permet pas de saisir l'essentiel des informations qu'il contient. L'analyse en composantes principales est une méthode de projection qui résume et décrit ce tableau de données. Si un tableau de données comporte **p** variables quantitatives, on dira que les individus sont représentés dans un « **espace à p dimensions** ».

Le but de l'A.C.P est de trouver des « espaces de dimensions plus petites », dans lesquels il est possible d'observer « au mieux » les individus [21], et de mieux connaître les données sur lesquelles on travaille, et à détecter éventuellement des valeurs suspectes, et aidera à formuler des hypothèses qu'il faudra étudier à l'aide de modèles et d'études statistiques inférentielles.

### III. 5.7. Le processus de l'Analyse en Composantes Principales [11, 12, 13]

Le processus d'Analyse en Composantes Principales consiste dans un premier temps à calculer la matrice de covariance  $\mathbf{R}$  entre les  $p$  vecteurs initiaux. Le terme général de la matrice de covariance  $\mathbf{R}$  s'écrit sous la forme :

$$R_{ij} = \sum_K (r_{ik} - \bar{r}_k) (r_{jk} - \bar{r}_k) \quad (1)$$

$r_{ik}$  : Le coefficient de corrélation entre le vecteur  $i$  et le vecteur  $k$

$\bar{r}_k$  : La moyenne des coefficients  $r_{ik}$ ;  $\forall i$ .

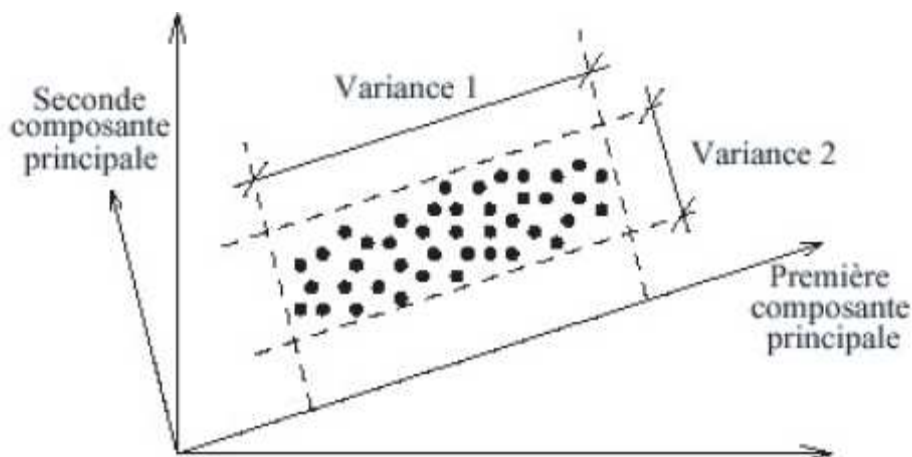
La matrice de covariance  $\mathbf{R}$  est alors diagonalisée :

$$\mathbf{R} = \mathbf{V} \mathbf{U} \mathbf{V}^t \quad (2)$$

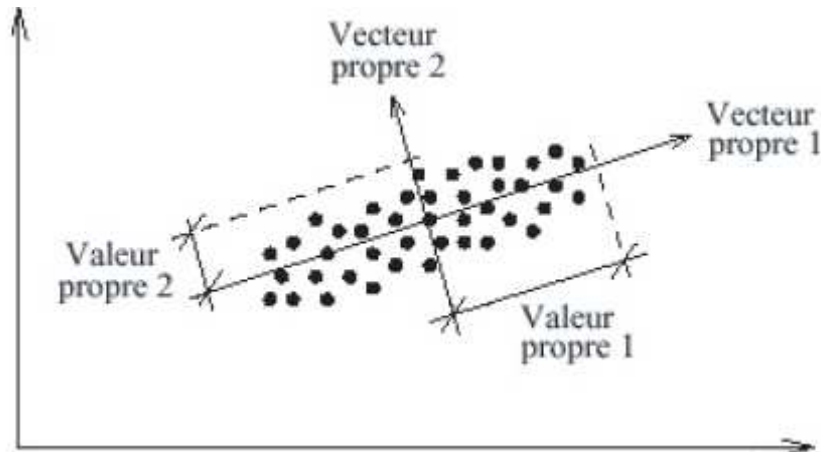
$\mathbf{V}$  : Une matrice contenant les vecteurs propres  $\vec{v}_i$  de  $\mathbf{R}$ , c'est-à-dire les coordonnées des  $p$  nouveaux axes.

Ces vecteurs propres représentent les nouvelles composantes (**Figure III.3.**).

$\mathbf{U}$  : Une matrice diagonale contenant les valeurs propres  $\lambda_i$  de  $\mathbf{R}$ , c'est-à-dire la variance des  $p$  nouvelles composantes (**Figure III.2. et III.3.**). Les valeurs propres sont rangées par ordre décroissant.



**Figure III.2 :** Illustration en 2D de la réduction de la dimension. La première composante contient le maximum d'information (maximum de variance). [11, 12, 13]



**Figure III.3 :** Représentation en 2D des 2 vecteurs propres et des 2 valeurs propres associées aux deux nouvelles composantes. [11, 12, 13]

- Enfin, les valeurs propres  $\lambda_i$  peuvent être normalisées pour obtenir les pourcentages de concentration  $\delta_i$  de l'information dans chaque nouvelle composante :

$$\delta_i = \frac{\lambda_i}{\sum_k \lambda_k} \times 100 \quad (3)$$

- La nature de l'information contenue dans chaque composante principale peut être connue en calculant la matrice de corrélation variables-composantes  $G$ .

Les éléments  $G_{ij}$  de la matrice  $G$  représentent les coefficients de corrélation entre les composantes initiales et les nouvelles composantes et se calculent à partir de :

$$G_{ij} = (\lambda_i)^{1/2} \vec{v}_j \quad (4)$$

Le processus de réduction de la dimension consiste à conserver les  $k$  ( $k < p$ ) premières nouvelles composantes, qui seront alors appelées composantes principales. Aucune règle générale n'existe quant au choix du nombre  $k$  de composantes principales. La valeur  $k$  peut être déterminée par l'étude des valeurs propres ou par l'étude visuelle des images :

- Kaiser (1958) suggère de conserver les  $k$  premières composantes ayant une valeur propre supérieure à 1.

### III. 5.7.1. Construction des axes principaux, démarche géométrique

Les droites et les plans issus d'une ACP ne sont pas réalisés avec les variables initiales mais avec des « indices synthétiques » obtenus à partir de « combinaisons linéaires » des variables initiales. Parmi tous les indices possibles, l'ACP recherche d'abord celui qui

permet de voir au mieux les individus c'est-à-dire celui pour lequel la variance des individus est maximale : cet indice est appelé **1<sup>ère</sup> composante principale** portée par le 1<sup>er</sup> axe principal appelé axe 1 ou **PC1**. Une certaine proportion de la variation totale des individus est expliquée par cette composante principale.

Ensuite une **deuxième composante principale** est recherchée, la variance (**PC2**) doit être, à son tour, la plus grande possible. Ainsi, cette deuxième composante principale fournit la plus grande information possible complémentaire à la première.

Le processus se déroule jusqu'à l'obtention de la  $p^{\text{ième}}$  et dernière composante principale, les parts d'informations expliquées par chacune d'elles devenant de plus en plus faibles (avec  $p$  variables, on obtient  $p$  composantes principales).

- L'intersection de tous les axes principaux se situe au centre de gravité du nuage de points [22].

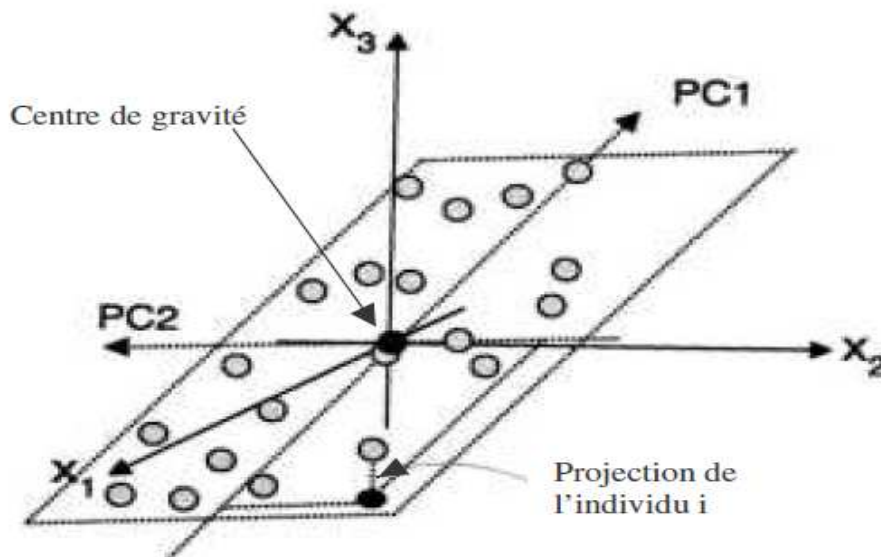


Figure III.4 : Projection des individus dans le plan principal. [22]

### III. 5.7.2. Calcul matriciel

Pour obtenir les composantes principales de façon analytique, on procède en deux étapes :

- 1- Elaboration de la **matrice des corrélations**.
- 2- Faire une **diagonalisation**, qui permet d'obtenir une **matrice de vecteurs propres**, qui contient les coefficients des combinaisons linéaires des  $p$  variables actives et la **matrice « diagonale » de valeurs propres** [23].

- **Les valeurs propres** qui représentent les variances des individus sur les axes principaux correspondants.
- **Les vecteurs propres** permettent le calcul des composantes principales et donc le calcul des coordonnées des variables et des individus sur les nouveaux axes principaux [24].

### **III. 5.8. Conclusion**

L'ACP est une technique mathématique qui réduit la dimensionnalité d'un système complexe de corrélations à un nombre réduit de dimensions, néanmoins représentatives de la majorité de la variance totale des données. On passe d'un certain nombre de variables potentiellement corrélées à un plus petit nombre de variables décorrélées, les "Composantes Principales".

Ensuite, un changement de base est effectué à la nouvelle base des "Composantes Principales" afin de représenter les données dans le nouvel espace. L'utilisateur choisit le nombre d'axes utilisés pour la représentation, en se basant par exemple sur le pourcentage de variance représenté.

### **III.6. La méthode des moindres carrés partiels (PLS)**

#### **III.6.1. Introduction-Problématique**

De nombreux problèmes peuvent être décrits sous la forme d'un modèle de régression, où l'on possède des variables  $X$  sur lesquelles on peut plus ou moins agir et des variables  $Y$  que l'on ne peut qu'observer.

L'objectif est alors de décrire les relations entre  $Y$  et  $X$ , en l'absence de modèle théorique. Le problème est que le nombre de variables  $X$  est souvent très important par rapport au nombre d'observations. La régression PLS est une méthode d'analyse des données spécifiquement construite pour l'étude de ce type de problème. [25]

#### **III.6.2. Historique**

La régression PLS (Partial Least Squares regression) en anglais, ou (La régression des moindres carrés partiels) en Français, est une technique d'analyse et de prédiction relativement récente. Elle a été conçue pour faire face aux problèmes résultants de l'insuffisance de l'utilisation de la régression linéaire classique, qui trouve ses limites dès lors que l'on cherche à modéliser des relations entre des variables pour lesquelles il y a peu d'individus, ou beaucoup de variables explicatives en comparaison au nombre d'individus, ou encore lorsque les variables explicatives sont fortement corrélées entre elles.

La régression PLS tire son origine des sciences sociales (plus précisément des sciences économiques, Herman Wold 1966 [27], mais devient très populaire en chimie grâce au fils d'Herman, Svante.

**1966** : Herman Wold publie un ouvrage introduisant le PLS, alors appelé (Non Linear Iterative Partial Least Squares) NIPALS.

**1983** : Svante Wold (le fils d'Herman) et Harold Mertens [27] adaptent NIPALS au problème de régression avec trop de prédicteurs et appellent PLS cette adaptation de l'algorithme.

**1990** : Stone et Brooks [28] introduisent le PLS dans le contexte du "Continuum Regression"

En France, Michel Tenenhaus s'est beaucoup intéressé à cette méthode et a réalisé de nombreux travaux à ce sujet. Par ailleurs, Tenenhaus M. pense que la régression PLS pourrait connaître les mêmes succès qu'en chimie si elle était utilisée dans d'autres domaines. [25]

### **III.6.3. Définition de la régression PLS ?**

La régression PLS est une méthode statistique permettant d'identifier des relations entre plusieurs variables. Il y a toujours, d'une part, les variables explicatives (notées généralement  $x_1, \dots, x_p$ ), et les variables expliquées (notées généralement  $y_1, \dots, y_q$ ). Ces variables sont, dans une régression PLS, toutes étudiées sur les mêmes « individus ».

On distingue la régression PLS univariée, ou « régression PLS1 », de la régression PLS multivariée, appelée également « régression PLS2 ». Dans le premier cas, la régression ne porte que sur une seule variable expliquée. Dans le second, il peut y avoir plusieurs variables expliquées. [25]

### **III.6.4. Les objectifs de la régression PLS**

La régression PLS répond à deux objectifs :

- La prédiction : Relier un bloc de variables à expliquer Y à un bloc de variables explicatives X.
- La description : Décrire simultanément les relations entre les variables composant le bloc des X, et les variables composant le bloc des Y, et les liaisons entre X et Y. [25, 29, 30].

### **III.6.5. Les versions de la régression PLS**

Il y'a différentes version de la régression PLS qui sont proposées en fonction de l'objectif poursuivi:

- PLS1** : Une variable cible Y quantitative est à expliquer, modéliser, prévoir par p variables explicatives quantitatives  $X^j$ .
- PLS2** : Version canonique. Mettre en relation un ensemble de q variables quantitatives  $Y^k$  et un ensemble de p variables quantitatives  $X^j$ .
- PLS2** : Version régression. Chercher à expliquer, modéliser un ensemble de q variables  $Y^k$  par un ensemble de p variables explicatives quantitatives  $X^j$ .
- PLS-DA** : Version discriminante. Cas particulier du cas précédent. La variable Y qualitative à q classes est remplacée par q variables indicatrices (dummy variables) de ces classes [31].

### III.7. Le principe de la régression PLS [32]

#### III.7.1. Données et notations

Les notations de la régression PLS sont relativement bien standardisées dans la littérature, et sont :

$N$  = Nombre d'individus

$M$  = Nombre de variables indépendantes  $X_j$

$P$  = Nombre de variables dépendantes  $Y_k$

$A$  = Nombre de composantes retenues

$X$  = Matrice des données ( $N \times M$ ) pour les variables indépendantes

$Y$  = Matrice des données ( $N \times P$ ) pour les variables dépendantes

$E_0$  = Matrice des variables  $X_j$  centrées-réduites

$F_0$  = Matrice des variables  $Y_k$  centrées-réduites

$E_h$  = Matrice des résidus de la décomposition de  $E_0$  en utilisant  $h$  composantes

$E_{hj}$  =  $j$ -ième colonne de  $E_h$

$F_h$  = Matrice des résidus de la décomposition de  $F_0$  en utilisant  $h$  composantes

$F_{hk}$  =  $k$ -ième colonne de  $F_h$

#### III.7.2. La méthode PLS

Il s'agit d'étudier les liaisons entre les variables dépendantes  $Y_k$  et les variables indépendantes  $X_j$  en prenant en compte les relations internes entre les variables de chaque groupe. L'algorithme de la régression PLS est itératif, il est décrit dans ce qui suit :

Etape 0 : On part des tableaux  $E_0$  et  $F_0$ .

Etape 1 : On construit une combinaison linéaire  $u_1$  des colonnes  $F_0$  et une combinaison linéaire  $t_1$  des colonnes  $E_0$ , maximisant  $\text{cov}(u_1, t_1) = \text{cor}(u_1, t_1) \cdot [\text{var}(u_1) \cdot \text{var}(u_2)]^{1/2}$

- On obtient donc deux variables  $u_1$  et  $t_1$  aussi corrélées que possible et résumant au mieux les deux tableaux  $E_0$  et  $F_0$ .

- On construit ensuite les régressions :

$$E_0 = t_1 p'_1 + E_1$$

$$F_0 = t_1 r'_1 + F_1$$

Etape 2 : On reprend l'étape 1 en remplaçant les tableaux  $E_0$  et  $F_0$  par  $E_1$  et  $F_1$ . On obtient donc deux nouvelles composantes :  $u_2$ , combinaison linéaire des colonnes de  $F_1$ ,  $t_2$ , combinaison linéaire des colonnes de  $E_1$ .

- D'où les décompositions obtenues par régression :

$$E_0 = t_1 p'_1 + t_2 p'_2 + E_2$$

$$F_0 = t_1 r'_1 + t_2 r'_2 + F_2$$

On itère la procédure jusqu'à ce que les composantes  $t_1, \dots, t_A$  expliquent suffisamment  $F_0$ . Les composantes  $t_h$  sont des combinaisons linéaires des colonnes d' $E_0$ , et non corrélées entre elles.

De la décomposition  $F_0 = t_1 r'_1 + \dots + t_h r'_h + F_h$ , on peut donc déduire les équations de régression PLS :

$$Y_k = \beta_{k0} + \beta_{k1} X_1 + \dots + \beta_{kM} X_M + F_{hk}$$

### **III.7.3. Indépendance des composantes dans la régression PLS**

L'une des propriétés primordiales d'une régression PLS est l'indépendance des composantes  $t_1, t_2, \dots, t_h$  formées à partir des variables explicatives. En effet, la première composante  $t_1$  est formée à partir des variables explicatives, en leur donnant certains coefficients sur la base de leur covariance avec la variable expliquée «  $y$  » (ou de leur coefficient de corrélation avec la variable  $y$  si les variables sont centrées réduites). Pour se faire, la variable  $t_1$  sera représentative d'une partie de la variance des variables explicatives.

Bien entendu, si «  $y$  » n'est pas une combinaison linéaire des variables explicatives, et qu'il y a plus d'une variable explicative dans l'analyse (et qu'aucune de ces variables n'est combinaison linéaire des autres), la variable  $t_1$  sera insuffisante pour expliquer toute la variance de «  $y$  », de même qu'elle sera insuffisante pour expliquer toute la variance des variables explicatives, et toute la covariance des variables explicatives avec «  $y$  ».

Il en demeurera un résidu. La variance de «  $y$  » ne sera pas totalement expliquée par la variance de  $t_1$ . Il y a un moyen d'améliorer le pouvoir explicatif du modèle. Pour cela, on s'intéresse aux résidus, qui ont été « oubliés » par la première composante. Cette première composante est indépendante des résidus. Or, on se sert de ces résidus pour construire la seconde composante  $t_2$ , qui sera par la même occasion indépendante de  $t_1$ . La composante  $t_2$  s'intéressera donc à la variance de «  $y$  » qui n'est pas expliquée par  $t_1$ . Les résidus qui en résulteront, qui sont donc indépendants de  $t_2$ , et indépendants de  $t_1$  (ils sont le résultat d'une régression sur des résidus qui sont déjà indépendants de  $t_1$ ), serviront à la création de  $t_3$ , ( $t_3$

sera donc indépendante de  $t_1$  et  $t_2$ ). Il en ira de même pour toutes les composantes, qui seront toutes indépendantes entre elles. [25]

### **III.8. Les avantages de la régression PLS**

La régression PLS peut se caractériser comme étant une méthode robuste, fiable, s'appliquant dans de nombreux cas, et étant plus générale que ne l'est la régression linéaire. [33], c'est une technique très puissante pour la modélisation des relations structurelles sur variables latentes (concepts, facteurs). [25, 29, 30].

Ses principaux avantages, peuvent se résumer ainsi :

- La régression PLS fonctionne bien sur un échantillon de taille faible, pouvant même être inférieur au nombre de variables explicatives.
- La régression PLS permet de compenser, partiellement, une baisse de qualité de l'échantillon.
- La régression PLS permet d'éviter certains problèmes engendrés par la multicolinéarité des variables.
- La régression PLS, dans son approche multivariée, permet d'expliquer plusieurs variables endogènes.
- La régression PLS, dans un algorithme plus général, permet de créer un modèle en tenant compte des individus présentant certaines données manquantes, sans avoir recours à des méthodes d'estimation des données manquantes. [33]
- Avec la régression PLS on a la possibilité de travailler sur des données individuelles et de faire de la prévision. [25, 29, 30].
- La simplicité de son algorithme : ni inversion, ni diagonalisation de matrices, mais seulement une succession de régressions simples, autrement dit des calculs de produits scalaires. On peut donc traiter de très grands ensembles de données.
- La régression PLS donne en pratique d'excellentes prévisions, même dans le cas d'un petit nombre d'observations et d'un grand nombre de variables. [34,35]

**III.9. Conclusion**

La régression PLS permet de rapprocher le statisticien du chercheur dans les problèmes de modélisation. En général ce dernier souhaite en effet conserver dans son modèle toutes les variables importantes tout en obtenant des équations de régression cohérentes. [32].

Bien que cette méthode ait connu la plupart de ses succès dans le domaine de la chimie, on peut penser qu'elle pourrait facilement être transposée à d'autres domaines, particulièrement dans ceux où le nombre d'individus actifs est faible (en comparaison au nombre de variables explicatives) et où les variables explicatives sont significativement corrélées entre-elles. [33]

**III.10. Références**

- [1] Djamel Abdelkader ZIGHED, Ricco RAKOTOMALALA, « Extraction de connaissances à partir de données (ECD) - Objet du data mining », Techniques de l'Ingénieur, (2002).
- [2] Mohammed J. Zaki and Wagner Meira Jr., « Data mining and analysis, Fundamental Concepts and Algorithms », Cambridge University Press, 1st ed., (2014).
- [3] Fayyad U.M., Piatetsky-Shapiro G., Smyth P., and Uthurusamy R. (eds.) *Advances in Knowledge Discovery and Data Mining*. Menlo Park, California: AAAI Press, 1996.
- [4] Gilbert Saporta, Chaire de Statistique Appliquée, <http://cedric.cnam.fr/~saporta/DM>.
- [5] Equipe de recherche en Ingénierie des Connaissances, Laboratoire ERIC, <http://chirouble.univ-lyon2.fr/~ricco/cours/>.
- [6] Bruno Agard, Andrew Kusiak, “ Exploration des bases de données industrielles à l'aide du data mining – Perspectives ”, 9ème Colloque National Aip Primeca, La Plagne - 5-8 avril 2005.
- [7] Stéphane Tufféry - Data Mining – (2006) - <http://data.mining.free.fr>.
- [8] Hand.D.J, “ Why Data Mining is more than statistics write large ”, Bulletin of the International Statistical Institute, 52nd Session .Helsinki, Vol. 1, (1999) 433-436. <http://www.stat.fi/isi99/index.html>
- [9] Hand.D.J, “ Methodological issues in Data Mining ”, Compstat 2000 : Proceedings in Computational Statistics .ed. J.G. Bethlehem and P.G.M.vau der Heijden, Physica-Verlag, (2000) 77-85.
- [10] Saporta G, “ Data Mining and Official Statistics ” .Quinta Conferenza Naionale di Statistica, ISTAT, Roma (2000).
- [11] H.Hotelling, “ Analysis of Complex Statistical Variables Into Principal Components ”, Journal of Educational Psychology, 1933, vol. 24,417-441 ,408-520.
- [12] K.Pearson, “ On Lines and Planes of Closest Fit to System of Points in Space ”, Philos.Mag., (1901), n° 11,559-572.
- [13] Kaiser H. “ The Varimax Criterion for Analytic Rotation in Factor Analysis”, Psychometrika 23, (1958). 187-200.
- [14] C.Spearman, “ General intelligence objectively determined and measured ”, American Journal of Psychology, 1904, vol.28, 129-149.
- [15] C. Duby, S. Robin, "Analyse en Composantes Principales", Département O.M.I.P (2006).

- [16] M.Volle, “ Analyse des données ”, Collection « économie et statistiques avancées », Ed, Economica .1993.
- [17] J-M Bouroche, G.Saporta, “ L’analyse des données ”, Ed .Que sait- je ? , 2002, 3-4.
- [18] P.Casin, “ Analyse des données et des panels de données ”, Ed. DeBoeck, Université perspectives marketing, 1999.
- [19] G.Chauvat, J-Preau, Statistiques descriptives, Ed Armand colin Editeur, 2002.
- [20] Site internet : <http://www.commentcamarche.net/> .
- [21] P.Fenelon., “ Qu’est ce que l’analyse des données ? ”, Ed.Lefonen, Paris, 1981.
- [22] G.Phillippeau, “ Comment interpréter les résultats d’une analyse en composantes principales ? ”, Collection STAT-ITCF-Ed.1992.
- [23] R.Rodriguez Herrera, D.Salles.Le Gac, “ Initiation à l’analyse factorielle des données ”, Ellipse, 2002.
- [24] Site internet : <http://www.lemoal.org/spss/> .
- [25] Tenenhaus M., « *La Régression PLS: Théorie et Pratique* », Editions Technip, (2002).
- [26] Wold, H. (1966), « *Estimation of principal component and related models by iterative least squares, Multivariate Analysis* », ed. P.R. Krishnaiah, Newyork : Academic Press, 391-420.
- [27] Stone, M. & Brooks, R. (1990), « *Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares, and principal components regression* », Journal of the Royal Statistical Society, Series B, 52(2), 237-269.
- [28] Wold, S., Martens, H. & Wold, H. (1983), « *The multivariate calibration problem in chemistry solved by the PLS method* », in Proc. Conf. Matrix Pencils, Ruhe A..(Eds), March 1982, Lecture Notes in Mathematics, Springer Verlag, Heidelberg, PP. 286-293.
- [29] ECSI ‘2000: « Use of PLS Path Modelling to estimate the European Consumer Satisfaction Index (ECSI) model », *MP Bayol, C Tellier, MTenenhaus* ([www.cisia.com/cisia/download/PLS\\_CISIA.doc](http://www.cisia.com/cisia/download/PLS_CISIA.doc)).
- [30] PLS' 2001 : « 2ème Symposium International sur les Méthodes PLS », Capri Palace, Île de Capri (Naples, Italie), (2001), CISIA·CERESTA, HEC et l'Université de Naples "Federico II".
- [31] M. Tenenhaus, La régression PLS : théorie et applications, Technip, 1998.
- [32] Tenenhaus, J.-P.Gauchi, C.Ménardo, « Régression PLS et applications », Revue de statistique appliquée, tome 43, n°1,(1995) , P.7-63.
- [33] Site internet : <http://www.renaudloup.be/refinternet.htm> .

[34] Tenenhaus M., « *La Régression PLS: Théorie et Pratique* », Editions Technip, (1998).

[35] Esposito Vinzi V., Lauro C., « PLS Regression and Classification, PLS and Related Methods, Proceedings of the PLS'03 », International Symposium, 45-56, Decisia , Paris, (2003).

## CHAPITRE IV

Résultats et discussions

Dans ce chapitre nous présentons, les résultats obtenus par l'application des techniques de data mining telles que l'Analyse en composantes principales (ACP) et la régression (PLS), qui sont des techniques d'analyse de données multivariées permettant d'étudier les corrélations existantes entre les différentes variables des bases de données, Ce chapitre se divise en quatre parties :

Dans la première partie nous avons réalisé une étude comparative des propriétés physiques et mécaniques du Lutétium avec ceux des métaux de transition par l'approche de datamining.

Tandis que dans la deuxième partie, on a étudié les propriétés physiques et mécaniques de l'hydrogène, et on les a comparée avec ceux du groupe des métaux alcalins, avec les éléments du groupe de carbone, et avec ceux du groupe des halogènes, en utilisant l'approche data mining.

Enfin dans la troisième partie on a étudié les propriétés structurales et mécaniques des hydrures alcalins en appliquant les calculs de premier principe et qu'on a analysé par l'approche de datamining.

## **Première partie**

### **IV.1. Une étude comparative des propriétés physiques et mécaniques du lutétium avec celles des métaux de transition par l'approche du datamining**

#### **IV.1.1. Introduction**

La position du lutétium dans le tableau périodique a suscité beaucoup de critiques de la part des chercheurs et a été considérée une de ses anomalies. Le but principal de ce travail est de démontrer que le lutécium Lu ( $Z = 71$ ) a des propriétés communes avec la troisième rangée du bloc  $d$  des métaux de transition. Il y'a beaucoup de chercheurs qui ont examiné la position de lutétium dans le tableau périodique, par différents moyens. En 1967, H. Merz et K. Ulmer avaient effectué une étude spectroscopique de rayonnement X montrant distinctement, que Lu, doit se placer à gauche de Hf (en dessous de l'yttrium), Les propriétés de Lu sont plus similaires à ceux de Hf que La, le Lutétium (Lu) peut donc être considéré comme un métal de transition et pas La [13]. Jensen (1982) a attribué Lu au groupe III, en

dessous de Y [16]. Scerri (2009) a estimé que le lutétium, doit être placé dans la même colonne du tableau périodique que le scandium, l'yttrium et le lawrencium [17].

Dans cette étude on a essayé d'aborder ce problème, en investiguant les propriétés physiques et mécaniques du lutétium, et les comparer avec ceux de la troisième période des métaux de transition. Pour réaliser ce travail on a constitué une base de données de 14 éléments y compris le lutétium (Cs, Ba, Hf, Ta, W, Re, Os, Ir, Pt, Au, Tl, Pb, Bi, **Lu**) et incluant les variables (propriétés) suivantes : la polarizabilité ( $\alpha$ ), l'électronégativité ( $\chi$ ), la masse atomique ( $m$ ), le numéro atomique ( $Z$ ), le rayon atomique ( $r$ ), le volume atomique ( $V$ ), l'affinité électronique ( $E_A$ ), l'énergie d'ionization ( $I$ ), l'énergie totale ( $E_{\text{tot}}$ ), l'énergie de cohésion ( $E_{\text{coh}}$ ), le module de compressibilité ( $B$ ), le module de cisaillement ( $G$ ), le module de Young ( $E$ ), le coefficient de Poisson ( $\mu$ ), le rapport ( $B/G$ ), les constantes élastiques ( $C_{11}$ ,  $C_{12}$ ,  $C_{44}$ ), la pression de Cauchy ( $C_{12}$ - $C_{44}$ ) (Tableau IV.1.)

## IV.1.2. Base de données

| Propriété                  | Cs                  | Ba                  | Lu                  | Hf                  | Ta                  | W                   | Re                  | Os                  | Ir                  | Pt                  | Au                  | Tl                  | Pb                 | Bi                  |
|----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|--------------------|---------------------|
| $\alpha$                   | 59,42 <sup>i</sup>  | 39,7 <sup>i</sup>   | 21,9 <sup>i</sup>   | 16,2 <sup>i</sup>   | 13,1 <sup>i</sup>   | 11,1 <sup>i</sup>   | 9,7 <sup>i</sup>    | 8,5 <sup>i</sup>    | 7,6 <sup>i</sup>    | 6,5 <sup>i</sup>    | 5,8 <sup>i</sup>    | 7,6 <sup>i</sup>    | 6,8 <sup>i</sup>   | 7,4 <sup>i</sup>    |
| $\chi$                     | 0,86 <sup>h</sup>   | 0,97 <sup>h</sup>   | 1,14 <sup>h</sup>   | 1,23 <sup>h</sup>   | 1,33 <sup>h</sup>   | 1,40 <sup>h</sup>   | 1,46 <sup>h</sup>   | 1,52 <sup>h</sup>   | 1,44 <sup>h</sup>   | 1,44 <sup>h</sup>   | 1,42 <sup>h</sup>   | 1,44 <sup>h</sup>   | 1,55 <sup>h</sup>  | 1,67 <sup>h</sup>   |
| <b>m</b>                   | 132,91 <sup>a</sup> | 137,33 <sup>a</sup> | 174,97 <sup>a</sup> | 178,49 <sup>a</sup> | 180,95 <sup>a</sup> | 183,84 <sup>a</sup> | 186,21 <sup>a</sup> | 190,23 <sup>a</sup> | 192,22 <sup>a</sup> | 195,08 <sup>a</sup> | 196,97 <sup>a</sup> | 204,38 <sup>a</sup> | 207,2 <sup>a</sup> | 208,98 <sup>a</sup> |
| <b>Z</b>                   | 55                  | 56                  | 71                  | 72                  | 73                  | 74                  | 75                  | 76                  | 77                  | 78                  | 79                  | 81                  | 82                 | 83                  |
| <b>r</b>                   | 2,60 <sup>c</sup>   | 2,15 <sup>c</sup>   | 1,734 <sup>j</sup>  | 1,55 <sup>c</sup>   | 1,45 <sup>c</sup>   | 1,35 <sup>c</sup>   | 1,35 <sup>c</sup>   | 1,30 <sup>c</sup>   | 1,35 <sup>c</sup>   | 1,35 <sup>c</sup>   | 1,35 <sup>c</sup>   | 1,90 <sup>c</sup>   | 1,80 <sup>c</sup>  | 1,60 <sup>c</sup>   |
| <b>V</b>                   | 115,89 <sup>f</sup> | 62,47 <sup>f</sup>  | 29,53 <sup>g</sup>  | 22,41 <sup>f</sup>  | 18,31 <sup>f</sup>  | 16,23 <sup>f</sup>  | 14,99 <sup>f</sup>  | 14,36 <sup>f</sup>  | 14,62 <sup>f</sup>  | 15,83 <sup>f</sup>  | 18,19 <sup>f</sup>  | 31,25 <sup>f</sup>  | 31,93 <sup>f</sup> | 35,38 <sup>g</sup>  |
| <b>EA</b>                  | 0,47 <sup>i</sup>   | 0,15 <sup>h</sup>   | 0,34 <sup>i</sup>   | 0,1 <sup>b</sup>    | 0,32 <sup>h</sup>   | 0,82 <sup>h</sup>   | 0,15 <sup>h</sup>   | 1,1 <sup>h</sup>    | 1,57 <sup>h</sup>   | 2,13 <sup>h</sup>   | 2,31 <sup>i</sup>   | 0,2 <sup>h</sup>    | 0,36 <sup>h</sup>  | 0,95 <sup>h</sup>   |
| <b>I</b>                   | 3,89 <sup>a</sup>   | 5,21 <sup>a</sup>   | 5,43 <sup>a</sup>   | 6,83 <sup>a</sup>   | 7,55 <sup>a</sup>   | 7,86 <sup>a</sup>   | 7,83 <sup>a</sup>   | 8,44 <sup>a</sup>   | 8,97 <sup>a</sup>   | 8,96 <sup>a</sup>   | 9,23 <sup>a</sup>   | 6,11 <sup>a</sup>   | 7,42 <sup>a</sup>  | 7,29 <sup>a</sup>   |
| <b>E<sub>tot</sub> [1]</b> | -7550,56            | -7880,11            | -13848,23           | -14317,49           | -14795,88           | -15283,45           | -15780,24           | -16286,30           | -16801,68           | -17326,58           | -17860,79           | -18956,96           | -19518,99          | -20090,41           |
| <b>E<sub>coh</sub></b>     | 0,80 <sup>d</sup>   | 1,90 <sup>d</sup>   | 4,43 <sup>d</sup>   | 6,44 <sup>d</sup>   | 8,10 <sup>d</sup>   | 8,90 <sup>d</sup>   | 8,03 <sup>d</sup>   | 8,17 <sup>d</sup>   | 6,94 <sup>d</sup>   | 5,84 <sup>d</sup>   | 3,81 <sup>d</sup>   | 1,88 <sup>d</sup>   | 2,03 <sup>d</sup>  | 2,18 <sup>d</sup>   |
| <b>B</b>                   | 2,03 <sup>e</sup>   | 10,3 <sup>e</sup>   | 47,6 <sup>i</sup>   | 108,9 <sup>e</sup>  | 200,1 <sup>e</sup>  | 323,2 <sup>e</sup>  | 371,7 <sup>e</sup>  | 417,8 <sup>e</sup>  | 355 <sup>e</sup>    | 278,3 <sup>e</sup>  | 173,2 <sup>e</sup>  | 35,9 <sup>e</sup>   | 42,9 <sup>e</sup>  | 31,5 <sup>e</sup>   |
| <b>G</b>                   | 0,65 <sup>h</sup>   | 4,84 <sup>h</sup>   | 27,1 <sup>h</sup>   | 53,0 <sup>h</sup>   | 64,7 <sup>h</sup>   | 152 <sup>h</sup>    | 180 <sup>h</sup>    | 223 <sup>h</sup>    | 209 <sup>h</sup>    | 60,9 <sup>h</sup>   | 26 <sup>h</sup>     | 2,67 <sup>h</sup>   | 5,54 <sup>h</sup>  | 12,8 <sup>h</sup>   |
| <b>E</b>                   | 1,69 <sup>h</sup>   | 12,8 <sup>h</sup>   | 68,4 <sup>h</sup>   | 138 <sup>h</sup>    | 185 <sup>h</sup>    | 407 <sup>h</sup>    | 520 <sup>h</sup>    | 559 <sup>h</sup>    | 528 <sup>h</sup>    | 170 <sup>h</sup>    | 78 <sup>h</sup>     | 7,89 <sup>h</sup>   | 15,8 <sup>h</sup>  | 34 <sup>h</sup>     |
| $\mu$                      | 0,30 <sup>h</sup>   | 0,28 <sup>h</sup>   | 0,27 <sup>h</sup>   | 0,29 <sup>h</sup>   | 0,35 <sup>h</sup>   | 0,28 <sup>h</sup>   | 0,26 <sup>h</sup>   | 0,25 <sup>h</sup>   | 0,26 <sup>h</sup>   | 0,39 <sup>h</sup>   | 0,42 <sup>h</sup>   | 0,45 <sup>h</sup>   | 0,44 <sup>h</sup>  | 0,33 <sup>h</sup>   |
| <b>B/G</b>                 | 3,12                | 2,13                | 1,76                | 2,05                | 3,09                | 2,13                | 2,07                | 1,87                | 1,70                | 4,57                | 6,66                | 13,45               | 7,74               | 2,46                |
| <b>C11</b>                 | 1,60 <sup>h</sup>   | 12,6 <sup>h</sup>   | 86,2 <sup>h</sup>   | 181 <sup>h</sup>    | 264 <sup>h</sup>    | 523 <sup>h</sup>    | 616 <sup>h</sup>    | 729,6 <sup>f</sup>  | 580 <sup>h</sup>    | 347 <sup>h</sup>    | 191 <sup>h</sup>    | 41,9 <sup>h</sup>   | 48,8 <sup>h</sup>  | 63,4 <sup>h</sup>   |
| <b>C12</b>                 | 0,99 <sup>h</sup>   | 8,0 <sup>h</sup>    | 32,0 <sup>h</sup>   | 77 <sup>h</sup>     | 158 <sup>h</sup>    | 203 <sup>h</sup>    | 273 <sup>h</sup>    | 230,0 <sup>f</sup>  | 242 <sup>h</sup>    | 251 <sup>h</sup>    | 162 <sup>h</sup>    | 36,6 <sup>h</sup>   | 41,4 <sup>h</sup>  | 24,5 <sup>h</sup>   |
| <b>C44</b>                 | 1,44 <sup>h</sup>   | 9,5 <sup>h</sup>    | 26,8 <sup>h</sup>   | 55,7 <sup>h</sup>   | 82,6 <sup>h</sup>   | 160 <sup>h</sup>    | 161 <sup>h</sup>    | 251,5 <sup>f</sup>  | 256 <sup>h</sup>    | 76,5 <sup>h</sup>   | 42,2 <sup>h</sup>   | 7,20 <sup>h</sup>   | 14,8 <sup>h</sup>  | 11,5 <sup>h</sup>   |
| <b>C12-C44</b>             | -0,45               | -1,5                | 5,2                 | 21,3                | 75,4                | 43                  | 112                 | -21,5               | -14                 | 174,5               | 119,8               | 29,4                | 26,6               | 13                  |

<sup>[1]</sup> Ref [1], <sup>a</sup> Ref [2], <sup>b</sup> Ref [3], <sup>c</sup> Ref [4], <sup>d</sup> Ref [5], <sup>e</sup> Ref [6], <sup>f</sup> Ref [7], <sup>g</sup> Ref [8], <sup>h</sup> Ref [9], <sup>i</sup> Ref [10], <sup>j</sup> Ref [11].

**Tableau IV.1.1.** La base de données incluant les propriétés physiques et mécaniques des métaux de transition de la troisième période y compris le lutétium.

### **IV.1.3. Techniques de calculs**

Dans ce travail, nous utilisons l'approche data mining afin d'explorer les corrélations existantes entre les données. Ce qui exige de puissants outils d'analyse statistique tels que l'ACP et la PLS. L'ACP est probablement la technique statistique multivariée la plus populaire utilisée dans de nombreux domaines scientifiques, comme elle peut être utilisée pour la classification des données, l'étude de similitudes, etc. [18-19]. L'ACP est une approche statistique puissante pour l'analyse des propriétés des matériaux; elle a été largement utilisée pour étudier de nombreux problèmes en physique et en science des matériaux [20]. Il ya une relation étroite entre la PLS et l'ACP, car elles ont des informations intéressantes en commun [21].

### **IV.1.4. L'analyse en composantes principales (ACP)**

Parmi les résultats obtenus par l'analyse en composantes principales (ACP): La matrice de corrélation, le graphe des observations, et le graphe des variables. Ce travail a été réalisé en exécutant l'ACP en utilisant les données numériques structurées en 14 observations par rapport à 19 variables ( $\alpha$ ,  $\chi$ , m, Z, r, V,  $E_A$ , I,  $E_{tot}$ ,  $E_{Coh}$ , B, G, E,  $\mu$ , B/G, C11, C12, C44, C12-C44). Pour une visualisation de la discrimination de données, l'ACP représente les variables (dix-neuf propriétés) et les observations ( $n = 14$ ) à l'aide du graphe des observations et du graphe des variables dans les espaces de dimensions déterminées par des PCs (composantes principales) ayant des valeurs propres  $> 1,0$ , selon la règle de Kaiser. Le graphe des variables représente les variables les plus importantes, et le graphe des observations montre les similitudes et les différences entre les échantillons ( les éléments dans notre cas) [22]. L'ACP génère un ensemble de nouveaux axes orthogonaux ou variables, appelé les composantes principales (PCs), à partir des variables d'origine. Les axes générés sont orthogonaux et non corrélés entre eux. La première composante (PC1) explique l'écart maximal (variance) dans l'ensemble de données ainsi que sa direction, et ainsi de suite. Les PCs sont définis par des vecteurs qui sont les vecteurs propres de la matrice de variance-covariance. L'écart (la variance) le long du vecteur est défini comme étant la valeur propre. Les observations (ou les variables) correspondantes aux PCs sont obtenues à partir de la matrice de corrélation. Chaque variable révèle la contribution de la variable d'origine par rapport aux composantes principales. La réduction de la variance est atteinte lorsqu'on néglige les directions selon lesquelles la variance des échantillons est insignifiante [22].

L'idée principale de la technique ACP est de réduire la dimension d'un ensemble de données qui contient un grand nombre de variables interdépendantes, en conservant autant que possible la variation trouvée dans l'ensemble de données. Cela rend la visualisation, ainsi que la prévision et la classification plus facile [23-24]. L'ACP résume les informations fournies par les différentes variables dans un simple graphique avec une perte minimale d'information pour évaluer les relations entre les variables et les échantillons (les éléments, dans notre cas) [25], elle identifie les motifs dans un ensemble de données et comprime les données en réduisant le nombre de dimensions en fonction de leurs similarités et de leur différences [21]. Quand une signification physique des composantes principales est trouvée, la technique ACP peut découvrir la structure cachée des données et révéler les similitudes entre les variables et les éléments [19].

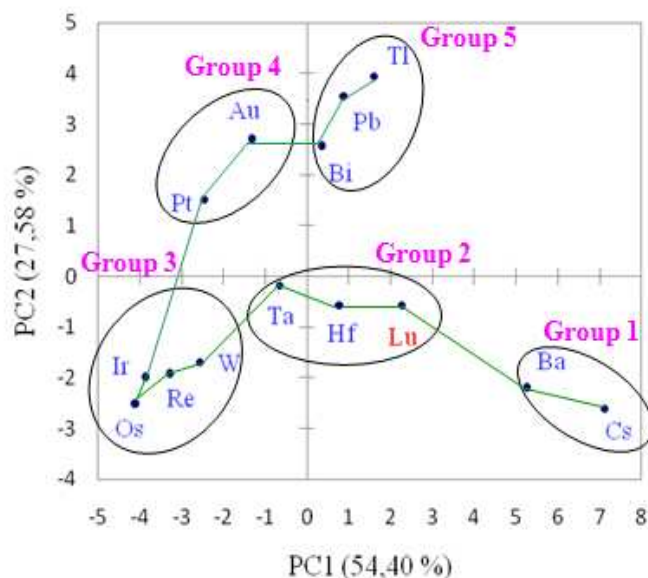
La matrice des corrélations nous aident à déterminer les variables qui sont corrélées entre elles, et le degré de leur corrélation.

En analysant la matrice des corrélations ci-dessus, on déduit l'existence de forte corrélation entre les propriétés physiques ( $\alpha$ ,  $\chi$ ,  $m$ ,  $Z$ , et  $r$ ) ( $> 0,86$ ) et entre les propriétés mécaniques ( $B$ ,  $G$ , et  $E$ ) ( $> 0,94$ ).

| Variables | $\alpha$     | $\chi$   | m        | Z        | r        | V        | EA       | I        | Etot     | Ecoh     | B           | G           | E        | $\mu$    | B/G         | C11      | C12      | C44      | C12-C44  |
|-----------|--------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-------------|-------------|----------|----------|-------------|----------|----------|----------|----------|
| $\alpha$  | <b>1</b>     | -0,92    | -0,94    | -0,94    | 0,86     | 0,93     | -0,38    | -0,84    | 0,92     | -0,46    | -0,49       | -0,37       | -0,37    | -0,32    | -0,25       | -0,44    | -0,54    | -0,38    | -0,37    |
| $\chi$    | <b>-0,92</b> | <b>1</b> | 0,95     | 0,94     | -0,73    | -0,74    | 0,36     | 0,77     | -0,95    | 0,31     | 0,43        | 0,35        | 0,35     | 0,31     | 0,24        | 0,41     | 0,45     | 0,36     | 0,25     |
| m         | <b>-0,94</b> | 0,95     | <b>1</b> | 1,00     | -0,68    | -0,75    | 0,35     | 0,71     | -1,00    | 0,21     | 0,29        | 0,18        | 0,18     | 0,48     | 0,41        | 0,25     | 0,34     | 0,20     | 0,29     |
| Z         | <b>-0,94</b> | 0,94     | 1,00     | <b>1</b> | -0,69    | -0,76    | 0,36     | 0,71     | -1,00    | 0,23     | 0,30        | 0,19        | 0,20     | 0,46     | 0,40        | 0,26     | 0,35     | 0,21     | 0,30     |
| r         | <b>0,86</b>  | -0,73    | -0,68    | -0,69    | <b>1</b> | 0,94     | -0,47    | -0,91    | 0,64     | -0,79    | -0,77       | -0,64       | -0,66    | 0,08     | 0,18        | -0,73    | -0,80    | -0,65    | -0,42    |
| V         | 0,93         | -0,74    | -0,75    | -0,76    | 0,94     | <b>1</b> | -0,31    | -0,82    | 0,71     | -0,69    | -0,63       | -0,52       | -0,53    | -0,06    | -0,01       | -0,59    | -0,66    | -0,52    | -0,36    |
| EA        | -0,38        | 0,36     | 0,35     | 0,36     | -0,47    | -0,31    | <b>1</b> | 0,70     | -0,37    | 0,14     | 0,41        | 0,21        | 0,19     | 0,22     | 0,01        | 0,31     | 0,51     | 0,29     | 0,45     |
| I         | -0,84        | 0,77     | 0,71     | 0,71     | -0,91    | -0,82    | 0,70     | <b>1</b> | -0,69    | 0,60     | 0,74        | 0,57        | 0,57     | 0,14     | -0,02       | 0,66     | 0,81     | 0,61     | 0,49     |
| Etot      | 0,92         | -0,95    | -1,00    | -1,00    | 0,64     | 0,71     | -0,37    | -0,69    | <b>1</b> | -0,16    | -0,25       | -0,15       | -0,15    | -0,50    | -0,43       | -0,21    | -0,31    | -0,16    | -0,29    |
| Ecoh      | -0,46        | 0,31     | 0,21     | 0,23     | -0,79    | -0,69    | 0,14     | 0,60     | -0,16    | <b>1</b> | 0,87        | 0,82        | 0,84     | -0,47    | -0,47       | 0,86     | 0,83     | 0,80     | 0,23     |
| B         | -0,49        | 0,43     | 0,29     | 0,30     | -0,77    | -0,63    | 0,41     | 0,74     | -0,25    | 0,87     | <b>1</b>    | 0,94        | 0,95     | -0,39    | -0,35       | 0,99     | 0,97     | 0,93     | 0,27     |
| G         | -0,37        | 0,35     | 0,18     | 0,19     | -0,64    | -0,52    | 0,21     | 0,57     | -0,15    | 0,82     | <b>0,94</b> | <b>1</b>    | 1,00     | -0,59    | -0,45       | 0,98     | 0,83     | 0,99     | -0,05    |
| E         | -0,37        | 0,35     | 0,18     | 0,20     | -0,66    | -0,53    | 0,19     | 0,57     | -0,15    | 0,84     | <b>0,95</b> | <b>1,00</b> | <b>1</b> | -0,59    | -0,45       | 0,98     | 0,85     | 0,98     | 0,00     |
| $\mu$     | -0,32        | 0,31     | 0,48     | 0,46     | 0,08     | -0,06    | 0,22     | 0,14     | -0,50    | -0,47    | -0,39       | -0,59       | -0,59    | <b>1</b> | <b>0,87</b> | -0,49    | -0,23    | -0,55    | 0,43     |
| B/G       | -0,25        | 0,24     | 0,41     | 0,40     | 0,18     | -0,01    | 0,01     | -0,02    | -0,43    | -0,47    | -0,35       | -0,45       | -0,45    | 0,87     | <b>1</b>    | -0,40    | -0,26    | -0,42    | 0,20     |
| C11       | -0,44        | 0,41     | 0,25     | 0,26     | -0,73    | -0,59    | 0,31     | 0,66     | -0,21    | 0,86     | 0,99        | 0,98        | 0,98     | -0,49    | -0,40       | <b>1</b> | 0,91     | 0,96     | 0,13     |
| C12       | -0,54        | 0,45     | 0,34     | 0,35     | -0,80    | -0,66    | 0,51     | 0,81     | -0,31    | 0,83     | 0,97        | 0,83        | 0,85     | -0,23    | -0,26       | 0,91     | <b>1</b> | 0,83     | 0,49     |
| C44       | -0,38        | 0,36     | 0,20     | 0,21     | -0,65    | -0,52    | 0,29     | 0,61     | -0,16    | 0,80     | 0,93        | 0,99        | 0,98     | -0,55    | -0,42       | 0,96     | 0,83     | <b>1</b> | -0,08    |
| C12-C44   | -0,37        | 0,25     | 0,29     | 0,30     | -0,42    | -0,36    | 0,45     | 0,49     | -0,29    | 0,23     | 0,27        | -0,05       | 0,00     | 0,43     | 0,20        | 0,13     | 0,49     | -0,08    | <b>1</b> |

Tableau IV.1.2 : La matrice de corrélation des métaux de transition de la troisième période y compris le lutétium.

Dans ce travail, l'analyse en composantes principales définit deux composantes principales qui représentent toutes les deux 81.98 % de la variance totale dans la base de données. PC1 capture 54.40% de la variance dans l'ensemble de données, et PC2 capture 27.58 % de la variance comme indiqué dans (Figure IV.1.1).



**Figure IV.1.1.** : Le graphe des observations des métaux de transition de la troisième période y compris le lutétium.

Le graphe des observations pour PC1 vs PC2, (**Figure IV.1.1**), révèle cinq groupes distincts d'éléments. Le groupe 1 (Cs, Ba) est lié aux éléments du bloc s, le groupe 2 (Ta, Hf, Lu) et le groupe 3 (Ir, Os, Re, W) contiennent des éléments appartenant au bloc d, le groupe 4 (Pt, Au)

contient des métaux précieux, et le groupe 5 (Tl, Pb, Bi) comprend des métaux post-transition, dont leur propriétés sont déterminées par leurs sous-couches p partiellement remplies. Nous notons que l'axe PC1 sépare les groupes 1, 2, 3 des groupes 4, 5. Il ya une grande similitude entre les éléments de chaque groupe. En regardant le groupe 2 (**Figure IV.1.1**) et (**Tableau IV.1.1**), nous notons que les distances séparant les points représentant les éléments (Ta, Hf, Lu) sont très petites, ce qui indique qu'ils ont des propriétés similaires à celles de l'hafnium (et à celles des métaux de transition), ils partagent plusieurs caractéristiques entre eux, parce qu'ils ont la même configuration de la couche de valence, qui est égal à 2.

Empiriquement, il a été démontré que l'électron externe du lutétium se situe dans la première sous-couche 5d; il doit marquer le début de la troisième série d'éléments de transition (c'est à

dire la sixième période du tableau périodique), et en dessous de l'yttrium (Y) dans le groupe 3 (IIIB), car il est un métal de type d<sup>1</sup>. L'emplacement d'un élément dans le tableau périodique dépend de sa configuration électronique, de sorte que les éléments d'une même rangée ont la même configuration de la couche de valence. Les électrons de valence déterminent la structure et les propriétés chimiques, mais la configuration électronique détermine l'organisation du tableau périodique des éléments ainsi que leur propriétés. Ce qui explique les tendances périodiques dans le comportement des éléments chimiques. Les propriétés des éléments sont déterminés par la taille (n), la forme (l) des orbitales, et le numéro atomique (charge nucléaire). Il y'a plutôt un grand nombre de preuves chimiques et physiques indiquant que le lutétium, doit être placé dans le même groupe que Sc et Y [12]. Cette question a été débattue par plusieurs chercheurs ([13], [14] et [27] à [31]) et d'une manière plus convaincante par Jensen en 1982 [14]. Une autre justification pour placer le lutétium (Lu) en dessous de l'yttrium (Y) dans le groupe 3 (IIIB), est que cet arrangement donne des tendances périodiques similaires dans plusieurs propriétés physiques (point de fusion, rayons atomiques, somme des deux premiers potentiels d'ionisation, et l'électronégativité), comme on le voit en (Tableau IV.1.3.). Sc, Y, et Lu ont tous une structure hexagonale compacte (hcp), à température ambiante [12]. Placer les éléments dans le tableau périodique est actuellement acceptée selon les observations empiriques suivantes: numéro atomique, les propriétés, les tendances périodiques, et la configuration électronique de l'état fondamental des atomes [32].

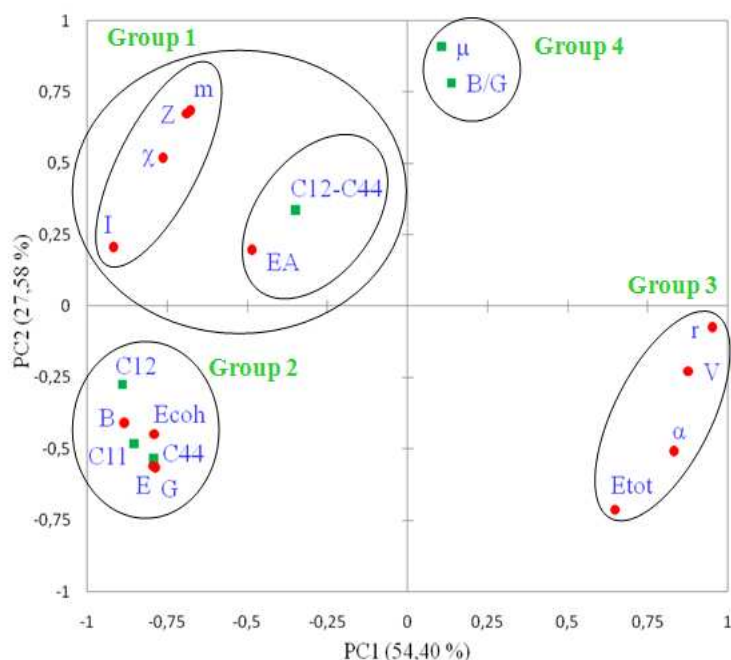
Dans le graphe des variables (Figure IV.1.2), 4 principaux groupes peuvent être observés: le groupe 1 contient principalement les propriétés physiques qui peuvent être décomposées en deux sous-groupes, le premier sous-groupe contient les propriétés atomiques (m, Z, I,  $\chi$ ) et le second contient la pression Cauchy C12-C44, et E<sub>A</sub>. Le deuxième groupe comprend les propriétés mécaniques B, E, G et C11, C12, C44, y compris l'énergie de cohésion E<sub>coh</sub>. Le troisième comporte des propriétés physiques telles que le rayon atomique r, le volume atomique V, la polarisabilité  $\alpha$ , et l'énergie totale E<sub>tot</sub>, et le quatrième groupe inclut le rapport B/G, ainsi que le coefficient de Poisson  $\mu$ . Les propriétés qui ont les valeurs des composantes principales (PC) identiques sont significativement corrélées et sont similaires les unes aux autres; Mais ceux qui ont des valeurs différentes de PC sont inversement corrélées et sont donc différentes. Dans le groupe 1, une similitude significative est observée entre m et Z, aussi dans le groupe 2 entre E, G et C44, les points indiquant les éléments sont proches les uns des autres (Figure IV.1.2), ce qui révèle l'importance de ces propriétés. La PCA sépare automatiquement les données. En effet, nous voyons que PC2 sépare clairement les propriétés physiques des propriétés atomiques, ce qui signifie une forte différence entre ces deux types

de propriétés. Les variables au sein du même groupe sont corrélées. Les groupes 1 et 3 sont inversement corrélés, et ainsi pour les groupes 2 et 4. Une augmentation de la masse atomique  $m$  et le numéro atomique  $Z$  se traduira par une augmentation de  $I$ , et  $\chi$  et une diminution de  $r$ ,  $V$ ,  $E_{\text{tot}}$ , et  $\alpha$ . Cependant, le module d'Young  $E$ , le module de cisaillement  $G$ , et le module de compressibilité  $B$  sont significativement corrélés aux constantes élastiques  $C_{11}$ ,  $C_{12}$ ,  $C_{44}$  et à l'énergie de cohésion  $E_{\text{coh}}$ .

| Propriété                               | Sc                 | Y                  | Lu                 |
|---|--------------------|--------------------|--------------------|
| Structure                               | hcp                | hcp                | hcp                |
| Point de fusion ( $T_{\text{melt}}$ )   | 1,541 <sup>k</sup> | 1,522 <sup>k</sup> | 1,663 <sup>k</sup> |
| Electronégativité ( $\chi$ )            | 1,20 <sup>h</sup>  | 1,11 <sup>h</sup>  | 1,14 <sup>h</sup>  |
| Rayons atomique ( $r$ )                 | 1,60 <sup>c</sup>  | 1,80 <sup>c</sup>  | 1,734 <sup>j</sup> |
| 1 <sup>er</sup> potentiel d'ionisation  | 6,56 <sup>h</sup>  | 6,22 <sup>h</sup>  | 5,43 <sup>a</sup>  |
| 2 <sup>ème</sup> potentiel d'ionisation | 12,80 <sup>h</sup> | 12,24 <sup>h</sup> | 13,9 <sup>h</sup>  |

<sup>a</sup> Ref [2], <sup>c</sup>Ref [4], <sup>h</sup> Ref [9], <sup>j</sup> Ref [11], <sup>k</sup> Ref [12].

**Tableau IV.1.3.** Les propriétés physiques du groupe 3 (IIIB), y compris le lutécium. Les quantités indiquées dans le tableau sont les suivantes: l'électronégativité ( $\chi$ ), le rayon atomique ( $r$ ), 1<sup>er</sup> potentiel d'ionisation, 2<sup>ème</sup> potentiel d'ionisation (eV), le point de fusion ( $^{\circ}$  C).



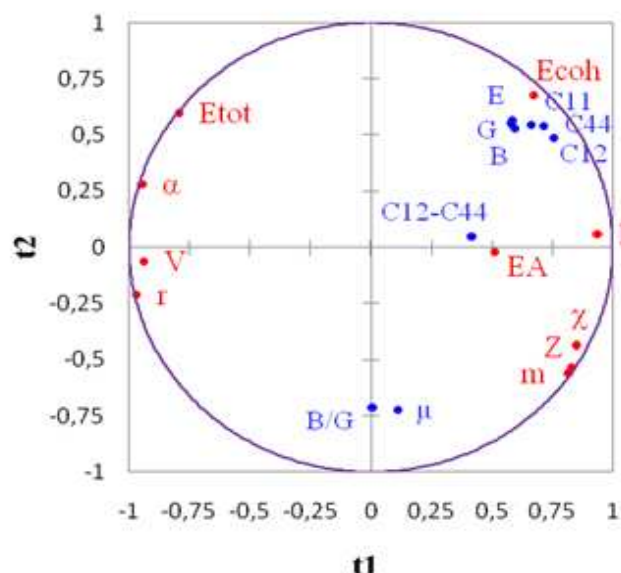
**Figure IV.1.2 :** Le graphe des variables des métaux de transition de la troisième période y compris le lutétium.

#### IV.1.5. La régression par les Moindres Carrés Partiels (PLS)

Alors que l'APC est utile pour évaluer l'impact relatif de plusieurs paramètres sur les propriétés, la régression PLS peut être utilisée comme un outil d'analyse exploratoire pour sélectionner les variables prédictives appropriés et d'identifier les valeurs aberrantes avant la régression linéaire classique. Une combinaison des techniques de l'APC et de la PLS permet d'explorer un large intervalle de nouvelles tendances et de corrélations, qui n'ont pas été établies avant [23]. La PLS comprend une large catégorie de méthodes de modélisation des relations entre les ensembles de variables observées en utilisant les variables latentes. La régression, inclut les tâches de classification, les techniques de réduction des dimensions et les outils de modélisation. L'idée principale de la méthode PLS est que les données observées sont générées par un système qui est entraîné par un petit nombre de variables latentes (*i.e.* pas directement observé ou mesuré). L'algorithme PLS devient un outil réussi standard pour traiter de larges problèmes de données chimiques. Généralement la PLS crée des vecteurs d'observations orthogonaux (appelés vecteurs latents ou composantes), et qui contiennent la covariance maximale entre les différents groupes de variables. Il existe différentes techniques pour obtenir les vecteurs latents PLS, ce qui donne une variante de PLS [26].

Nous étudions les relations entre les deux blocs de variables; entre les propriétés mécaniques ( $E$ ,  $B$ ,  $G$ ,  $\mu$ ,  $B/G$ ,  $C_{11}$ ,  $C_{12}$ ,  $C_{44}$ ,  $C_{12}-C_{44}$ ) et les propriétés structurales et atomiques ( $\alpha$ ,  $\chi$ ,  $m$ ,  $Z$ ,  $r$ ,  $V$ ,  $E_A$ ,  $I$ ,  $E_{\text{coh}}$ ,  $E_{\text{tot}}$ ) en utilisant l'approche PLS. Les résultats PLS, présentés dans (**Figure IV.1.3**), indiquent le cercle de corrélation qui explore la relation entre les propriétés mécaniques, atomiques et structurales, ainsi qu'entre les deux composantes PLS ( $t_1$ ,  $t_2$ ) qui capturent le maximum de la variance dans la base de données. Les résultats peuvent être expliqués de la même manière, comme dans l'analyse en composantes principales (ACP).

Pour les propriétés, tels que  $C_{12}-C_{44}$  et l'affinité d'électrons  $E_A$ , situées au centre de la carte (le cercle de corrélation), où les corrélations sont faibles. Cela signifie que leurs propriétés ne sont pas bien représentés sur les deux premières composantes. Les propriétés  $C_{12}-C_{44}$  et  $E_A$  expliquent à peine les tendances des éléments; ils n'ont pas un effet important qui pourrait facilement influencer les tendances des éléments; ils ne sont pas importantes. On constate de fortes corrélations entre  $m$  et  $Z$  d'une part, ainsi qu'entre les constantes élastiques  $C_{11}$ ,  $C_{12}$ ,  $C_{44}$ , et  $E$ ,  $G$ ,  $B$ , d'autre part. En outre, une corrélation négative existe entre  $E_{\text{tot}}$  et  $m$ ,  $Z$ , ainsi qu'entre l'énergie d'ionisation  $I$  et le volume atomique  $V$ . Les variables  $B/G$  et  $\mu$  sont séparées des autres variables parce qu'elles sont moins corrélées à ces variables d'un côté, et aux composantes PLS ( $t_1$ ,  $t_2$ ) d'un autre côté. Les propriétés situées sur le périphérie du cercle de corrélation sont significatives, c'est à dire, ils contribuent fortement à la discrimination.



**Figure IV.1.3 :** Le graphe des observations ( $t_1$ ,  $t_2$ ) des métaux de transition de la troisième période y compris le lutétium.

**IV.1.6. Conclusion**

Dans cette partie, nous avons introduit une nouvelle approche pour analyser et comparer les propriétés physiques et mécaniques de lutétium avec des éléments de transition de la troisième rangée du tableau périodique, en utilisant des méthodes d'analyse multivariée tels que l'ACP et la régression PLS. L'ACP nous a permis de classer les propriétés et les éléments, qui ont contribué à visualiser la position des éléments l'un par rapport à l'autre, ainsi que les relations existantes entre leurs propriétés. Les résultats obtenus par la PLS est en bonne cohérence avec les résultats obtenus par l'ACP. Un intérêt majeur dans cette méthodologie est donnée par le cadre unitaire, fondée sur les principes de PLS, ce qui permet de découvrir les relations structurelles existantes entre les données et les propriétés. Nos résultats montrent que les propriétés de lutétium (Lu) sont semblables à ceux de l'hafnium (Hf). Nous avons constaté que nos résultats sont généralement en bon accord avec de nombreuses études déjà publiées sur ce sujet expérimentaux et théoriques. Il est difficile pour les scientifiques de comprendre les interactions entre un grand nombre de variables, donc nous avons utilisé l'ACP et la PLS comme des outils pour identifier les structures existantes dans un ensemble de données, ainsi que pour le dépistage, et d'extraire des informations utiles à partir de données.

**IV.1.7. Références**

- [1] NIST (National Institute of Standards and Technology), (2013). [www.nist.gov](http://www.nist.gov).
- [2] R.A. Dragoset, A. Musgrove, C.W. Clark, and W.C. Martin, *Periodic Table: Atomic Properties of the Elements* (Version 4), NIST SP 966 (2003).  
<<http://physics.nist.gov/PT>>. NIST (National Institute of Standards and Technology), Gaithersburg, MD.
- [3] E.C.M. Chen, E.S. Chen, Molecular electron affinities and the calculation of the temperature dependence of the electron-capture detector response. *Journal of Chromatography A*, 2004, vol. 1037, pp. 83–106.
- [4] J.C. Slater, *J. Chem. Phys.*, 1964, vol. 39, pp. 3199, <http://www.webelements.com/>.
- [5] Charles Kittel, *Introduction to Solid State Physics*, 8th ed., Hoboken, NJ: John Wiley & Sons, Inc, 2005.
- [6] Y. Makino, S. Miyake, Estimation of bulk moduli of compounds by empirical relations between bulk modulus and interatomic distance. *Journal of Alloys and Compounds*, 2000, vol.313, pp. 235–241.
- [7] S.L. Shang, A. Saengdeejing, Z.G. Mei, D.E. Kim, H. Zhang, S. Ganeshan, Y. Wang, Z.K. Liu, First-principles calculations of pure elements: Equations of state and elastic stiffness constants. *Computational Materials Science*, 2010, vol. 48, pp. 813–826.
- [8] The Massachusetts Institute of Technology, 1996. <<http://www.chemcool.com>> (2013).
- [9] Martienssen Warlimont., *Springer handbook of condensed matter and materials Data*, Volume 1-Hans Warlimont, Springer Berlin Heidelberg, 2005.
- [10] David R. Lide, ed., *CRC Handbook of Chemistry and Physics*, 90th Ed, CRC Press/Taylor and Francis, Boca Raton, FL, 2010.
- [11] S. Cotton, *Lanthanide and Actinide Chemistry*, John Wiley & Sons Ltd, England, 2006.
- [12] K. A. Gschneidner, Jr., J.-C. G. Bünzli and V. K. Pecharsky, *Handbook on the physics and chemistry of rare earths*, vol 41, North-Holland, Elsevier, 2011.
- [13] H. Merz, K. Ulmer, Position of lanthanum and lutetium in the periodic table. *Physics Letters*, 1967, vol. 26, pp. 1-2.
- [14] Jensen, W. B. *J. Chem. Educ.*, 1982, vol. 59, pp. 634–636.
- [15] Glossaire de la modélisation des données, site: <http://www.aiaccess.net/>
- [16] Michael Laing, A revised periodic table: with the lanthanides repositioned. *Foundations of Chemistry*, 2005, vol. 7, pp. 203–233.

- [17] Eric Scerri, Explaining the periodic table, and the role of chemical triads. *Found Chem.* 2010, vol. 12, pp. 69–83.
- [18] Hervé Abdi, Lynne J. Williams, Principal component analysis. *WIREs Computational Statistics.* 2010, vol. 2, pp. 433–459.
- [19] A. Lengyel, K.Héberger, L. Paksy, O.Bánhidi, R. Rajkó, Prediction of ozone concentration in ambient air using multivariate methods. *Chemosphere.* 2004, vol. 57, pp. 889– 896.
- [20] Scott R. Broderick, Hafid Aourag, Krishna Rajan, Data mining of Ti–Al semi-empirical parameters for developing reduced order models. *Physica B,* 2011, vol. 406, pp. 2055–2060.
- [21] Hai-Ni Qu, Guo-Zheng Li, Wei-Sheng Xu, An asymmetric classifier based on partial least squares. *Pattern Recognition,* 2010, vol.43, pp. 3448–3457.
- [22] Eui-Cheol Shin, Brian D. Craft, Ronald B. Pegg, R. Dixon Phillips, Ronald R. Eitenmiller, Chemometric approach to fatty acid profiles in Runner-type peanut cultivars by principal component analysis (PCA). *Food Chemistry,* 2010, vol. 119, pp. 1262– 1270.
- [23] Krishna Rajan, Materials informatics. *Materials Today.* 2005, vol.8, pp. 38–45.
- [24] I.T. Jolliffe, Principal Component Analysis, Springer series in statistics, 2<sup>nd</sup> ed, Springer-Verlag New York, Inc, 2002.
- [25] Ivano Marchi, Serge Rudaz, Maurice Selman, Jean-Luc Veuthey, Evaluation of the influence of protein precipitation prior to on-line SPE–LC–API/MS procedures using multivariate data analysis. *Journal of Chromatography B.* 2007, vol. 845, pp. 244–252.
- [26] Roman Rosipal, Nicole Krämer, Overview and recent advances in partial least squares, C. Saunders et al. (Eds.), 2006, pp. 34–51.
- [27] Scerri, E. R. *J. Chem. Educ,* 1991, vol. 68, 122–126.
- [28] Jensen, W. B. *J. Chem. Educ,* 2003 vol. 80, pp. 952–961.
- [29] Luder, W. F., *The Electron-Repulsion Theory of the Chemical Bond;* Reinhold:New York, 1967.
- [30] Hamilton, D. C. *Amer. J. Phys,*1965, vol. 33, pp. 637–640.
- [31] Matthias, B. T.; Zacharisen, W. H.; Webb, G. W.; Engelhardt. J. J. *Phys. Rev. Letters,* 1967, vol. 18, pp. 781–784.
- [32] Laurence Lavelle, Lanthanum (La) and Actinium (Ac) Should Remain in the d- Block. *Journal of Chemical Education* N<sup>o</sup> 11, 2008, vol. 85, pp. 1482–1483.

## Deuxième partie

### **IV.2. Une étude comparative des propriétés physiques et mécaniques de l'hydrogène en appliquant les techniques de recherche du data mining**

#### **IV.2.1. Introduction**

La position de l'hydrogène dans le tableau périodique a été le sujet de beaucoup de recherches et de débats, c'est une de ses grandes contradictions. Parmi tous les éléments dans la nature, l'hydrogène est le plus répandu sur terre c'est le dixième élément plus lourd en classant les éléments par masse. L'hydrogène existe dans toutes les formes de vie [18]. Placer l'hydrogène dans un tableau a été problématique depuis le premier tableau de Dmitri Mendeleev en 1869, où l'hydrogène a été considéré comme un élément spécial. En 1871, Mendeleev avait placé l'atome d'hydrogène au sommet de la première colonne du tableau périodique. Bayley (1882) et Thomsen (1895) ont donné une forme pyramidale au tableau, où ils ont placé l'hydrogène en haut au centre et ils l'ont isolé contre les éléments de la deuxième période du tableau périodique. En 1922, Bohr avait ajouté les gaz nobles, en mettant l'hélium à côté de l'hydrogène. En 1967, Sanderson a utilisé l'électronégativité comme un critère; il a placé l'hydrogène dans la partie supérieure du tableau périodique chevauchant le bore et le carbone parce qu'il pensait que cette position était la plus appropriée. Greenwood et Earnshaw (1984) ont placé H et He ensemble, en haut et à part du reste des éléments et il a considéré l'hélium comme un gaz rare. En 2003, Kaesz et Atkins ont placé séparément l'atome d'hydrogène, au milieu des groupes 7 et 8 des métaux de transition. En 2003, Cronyn a placé l'hydrogène au-dessus du carbone [21]. Ainsi, après plus de 130 ans, la position de l'hydrogène dans le tableau périodique est encore un sujet de confusion [22].

Dans cette partie on va essayer de confronter ce problème, en comparant les propriétés physiques et mécaniques de l'hydrogène avec ceux du groupe des métaux alcalins, ceux du groupe du carbone, et avec ceux du groupe des halogènes, en utilisant les techniques du data mining : l'Analyse en composantes principales (ACP) et la régression (PLS). Pour cela nous avons construit trois bases de données (**Tableau IV.2.1**, **Tableau IV.2.2**, et le **Tableau IV.2.3**). La première base de données est constituée de six (6) observations (éléments), représentant les métaux alcalins, y compris l'hydrogène, et incluant 16 variables

(propriétés) (les propriétés physiques et mécaniques des éléments de la base de données). Les propriétés et leurs abréviations sont: la polarisabilité ( $\alpha$ ), l'électronégativité ( $\chi$ ), la masse atomique ( $m$ ), le numéro atomique ( $Z$ ), le rayon atomique ( $r$ ), le volume atomique ( $V$ ), l'affinité électronique ( $E_A$ ), l'énergie d'ionisation ( $I$ ), l'énergie totale ( $E_{tot}$ ), la conductivité électrique ( $\sigma$ ), le module de compressibilité ( $B$ ), le module de Young ( $E$ ), les constantes élastiques ( $C11$ ,  $C12$ ,  $C44$ ), la pression de Cauchy ( $C12$ - $C44$ ). La deuxième base de données est composée de six (6) observations. Elle contient les éléments du groupe de carbone, y compris l'hydrogène, et seize 16 variables identiques à celles de la première base de données. La troisième base de données contient six observations qui sont les éléments du groupe VII (les halogènes), plus l'hydrogène, et dix (10) variables, représentant les propriétés physiques et la conductivité électrique. Dans cette troisième base de données nous avons inclus seulement ces propriétés à cause de la non-disponibilité des données pour les autres propriétés.

L'objectif principal de ce travail est d'étudier les tendances dans les propriétés des matériaux pour les trois bases de données en appliquant l'ACP qui va révéler la composition dissimulée des données et montrera les similitudes existantes entre les variables et les éléments, afin d'identifier le groupe qui partage le plus de propriétés avec l'hydrogène.

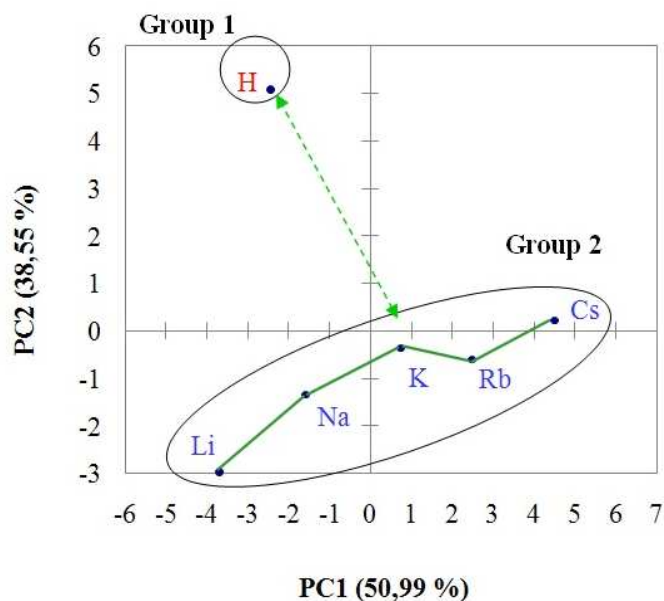
## IV.2.2. Bases de données

### IV.2.2.1 La base de données des métaux alcalins y compris l'hydrogène

| Propriété     | H                     | Li                      | Na                      | K                       | Rb                      | Cs                     |
|---------------|-----------------------|-------------------------|-------------------------|-------------------------|-------------------------|------------------------|
| $\alpha$      | 0.667 <sup>h</sup>    | 24.33 <sup>h</sup>      | 24.11 <sup>h</sup>      | 43.4 <sup>h</sup>       | 47.3 <sup>h</sup>       | 59.42 <sup>h</sup>     |
| $\chi$        | 2.20 <sup>g</sup>     | 0.97 <sup>g</sup>       | 1.01 <sup>g</sup>       | 0.91 <sup>g</sup>       | 0.89 <sup>g</sup>       | 0.86 <sup>g</sup>      |
| $m$           | 1.00794 <sup>a</sup>  | 6.941 <sup>a</sup>      | 22.989770 <sup>a</sup>  | 39.0983 <sup>a</sup>    | 85.4678 <sup>a</sup>    | 132.90545 <sup>a</sup> |
| $Z$           | 1                     | 3                       | 11                      | 19                      | 37                      | 55                     |
| $r$           | 0.25 <sup>b</sup>     | 1.45 <sup>b</sup>       | 1.80 <sup>b</sup>       | 2.20 <sup>b</sup>       | 2.35 <sup>b</sup>       | 2.60 <sup>b</sup>      |
| $V$           | 92.23 <sup>f</sup>    | 20.299 <sup>e</sup>     | 37.144 <sup>e</sup>     | 73.736 <sup>e</sup>     | 90.810 <sup>e</sup>     | 115.892 <sup>e</sup>   |
| $E_A$         | 0.754 <sup>h</sup>    | 0.618 <sup>h</sup>      | 0.5479 <sup>h</sup>     | 0.5015 <sup>h</sup>     | 0.4859 <sup>h</sup>     | 0.4716 <sup>h</sup>    |
| $I$           | 13.5984 <sup>a</sup>  | 5.3917 <sup>a</sup>     | 5.1391 <sup>a</sup>     | 4.3407 <sup>a</sup>     | 4.1771 <sup>a</sup>     | 3.8939 <sup>a</sup>    |
| $E_{tot}$ [1] | -12.137               | -199.600                | -4393.105               | -16278.245              | -79903.602              | -205465.784            |
| $\sigma$      | 16 x 10 <sup>3i</sup> | 11.7 x 10 <sup>6f</sup> | 20.1 x 10 <sup>6f</sup> | 16.4 x 10 <sup>6f</sup> | 47.8 x 10 <sup>6f</sup> | 5.3 x 10 <sup>6f</sup> |
| $B$           | 0.2 <sup>c</sup>      | 11.6 <sup>d</sup>       | 6.81 <sup>d</sup>       | 3.18 <sup>d</sup>       | 3.14 <sup>d</sup>       | 2.03 <sup>d</sup>      |
| $E$           | 0.2 <sup>g</sup>      | 11.5 <sup>g</sup>       | 6.80 <sup>g</sup>       | 3.52 <sup>g</sup>       | 2.35 <sup>g</sup>       | 1.69 <sup>g</sup>      |
| $C11$         | 0.042 <sup>g</sup>    | 13.4 <sup>g</sup>       | 7.59 <sup>g</sup>       | 3.69 <sup>g</sup>       | 2.96 <sup>g</sup>       | 1.60 <sup>g</sup>      |
| $C12$         | 0.18 <sup>g</sup>     | 11.3 <sup>g</sup>       | 6.33 <sup>g</sup>       | 3.18 <sup>g</sup>       | 2.44 <sup>g</sup>       | 0.99 <sup>g</sup>      |
| $C44$         | 0.11 <sup>g</sup>     | 9.6 <sup>g</sup>        | 4.30 <sup>g</sup>       | 1.90 <sup>g</sup>       | 1.60 <sup>g</sup>       | 1.44 <sup>g</sup>      |
| $C12$ - $C44$ | 0.07                  | 1.70                    | 2.03                    | 1.28                    | 0.84                    | -0.45                  |

<sup>[1]</sup> Ref [5], <sup>a</sup> Ref [6], <sup>b</sup> Ref [7], <sup>c</sup> Ref [8], <sup>d</sup> Ref [9], <sup>e</sup> Ref [4], <sup>f</sup> Ref [10], <sup>g</sup> Ref [11], <sup>h</sup> Ref [12], <sup>i</sup> Ref [13] at 162 GPa and 6500 K.

**Tableau IV.2.1.** La base de données incluant les propriétés physiques et mécaniques des métaux alcalins y compris l'hydrogène.

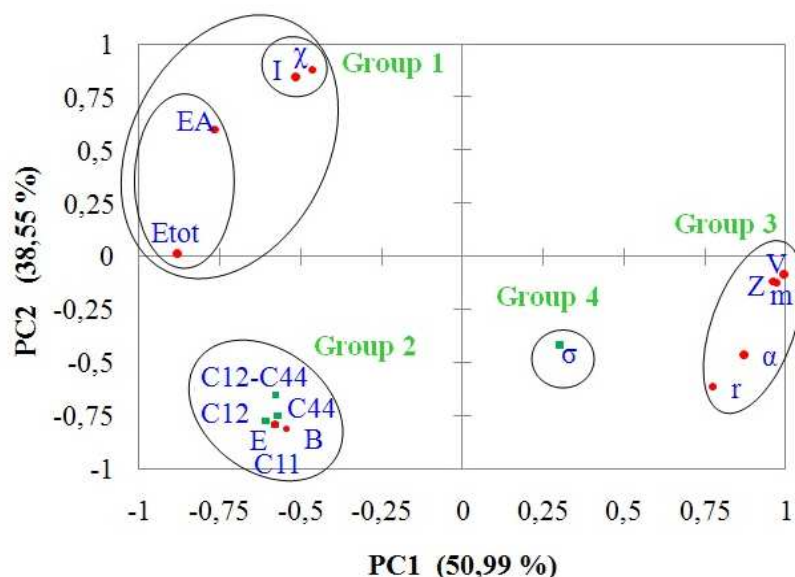
IV.2.2.2. L'analyse en composantes principales (ACP)

**Figure IV.2.1.** : Le graphe des observations des métaux alcalins y compris l'hydrogène.

Dans ce travail, l'analyse en composantes principales de cette base de données définit deux composantes principales qui représentent toutes les deux 89.54% de la variance totale dans la base de données, PC1 capture 50.99% de la variance dans l'ensemble de données, et PC2 capture 38.55% de la variance comme indiqué dans (**Figure IV.2.1.**).

Les autres composantes principales ne sont pas prises en considération parce qu'elles ne fournissent pas d'informations significatives. L'objectif du graphe des observations est de visualiser les positions du comportement des variables dans un espace bidimensionnel, ainsi que leurs relations équivalentes. Les variables à proximité l'une de l'autre et éloignées de l'origine du graphe sont positivement corrélées; tandis que celles qui s'opposent les unes aux autres sont négativement corrélées (ou inversement proportionnelles). Dans le graphe des variables, les corrélations entre deux variables sont expliquées par l'intermédiaire de leur angle par rapport au centre. Le coefficient de corrélation entre deux variables est déterminé au moyen du cosinus de l'angle formée par leurs vecteurs correspondants sur le graphe. Étant donné que le cosinus d'un angle plat est égale à -1, ils ont une corrélation divergente. Les variables qui ne sont pas corrélées forment entre elles un angle droit dont le cosinus est nul. De même, le cosinus d'un angle de degré  $0^\circ$  est 1, ce qui signifie que la corrélation est totale entre les variables [16].

Le graphe des observations pour PC1 vs PC2, (**Figure IV.2.1.**), révèle deux groupes distincts d'éléments. Groupe 1 constitué des métaux alcalins (Li, Na, K, Rb, Cs), et le groupe 2 est constitué de l'élément de l'hydrogène (H). Nous notons que l'axe PC1 sépare les groupes 1, 2. Il ya une ressemblance significative entre les éléments dans le groupe 2. Une distance importante est constatée entre le point représentant H et les points représentant les métaux alcalins, ce qui signifie que l'hydrogène est bien différent des métaux alcalins, et qu'il n'est pas conforme au premier groupe. Une nette différence entre l'hydrogène et les métaux alcalins apparaît dans la profondeur de leurs puits d'énergie, avec leurs potentiels d'ionisation égaux à 13,6 eV et 5,3 eV pour l'hydrogène et le lithium, respectivement [1]. D'après les données fournies dans le **Tableau IV.2.1**, toutes les propriétés de l'hydrogène (H) sont très différentes de celles des métaux alcalins à l'exception de l' $E_A$ . L'emplacement de l'hydrogène dans le tableau périodique avec les métaux alcalins est due à sa configuration électronique (monovalent) et non à ses caractéristiques chimiques. Les électrons de valence déterminant la chimie, et la configuration électronique reflète l'organisation du tableau périodique. Les métaux alcalins sont parmi les meilleurs donneurs d'électrons; ils ont à peu près la même électronégativité  $\chi$ , et qui est proche de 1 (**Tableau IV.2.1**). Le caractère métallique d'un élément est principalement due à sa structure et ses caractéristiques physiques [2].



**Figure IV.2.2 :** Le graphe des variables des métaux alcalins y compris l'hydrogène.

### IV.2.2.3 La base de données du groupe de carbone y compris l'hydrogène

| Propriété            | H                     | C                       | Si                   | Ge                  | Sn                     | Pb                     |
|----------------------|-----------------------|-------------------------|----------------------|---------------------|------------------------|------------------------|
| $\alpha$             | 0.667 <sup>h</sup>    | 1.76 <sup>h</sup>       | 5.38 <sup>h</sup>    | 6.07 <sup>h</sup>   | 7.7 <sup>h</sup>       | 6.8 <sup>h</sup>       |
| X                    | 2.20 <sup>g</sup>     | 2.50 <sup>g</sup>       | 1.74 <sup>g</sup>    | 2.02 <sup>g</sup>   | 1.72 <sup>g</sup>      | 1.55 <sup>g</sup>      |
| m                    | 1.00794 <sup>a</sup>  | 12.0107 <sup>a</sup>    | 28.0855 <sup>a</sup> | 72.64 <sup>a</sup>  | 118.710 <sup>a</sup>   | 207.2 <sup>a</sup>     |
| Z                    | 1                     | 6                       | 14                   | 32                  | 50                     | 82                     |
| R                    | 0.25 <sup>b</sup>     | 0.70 <sup>b</sup>       | 1.10 <sup>b</sup>    | 1.25 <sup>b</sup>   | 1.45 <sup>b</sup>      | 1.80 <sup>b</sup>      |
| V                    | 92.23 <sup>f</sup>    | 5.68 <sup>f</sup>       | 20.09 <sup>f</sup>   | 22.58 <sup>f</sup>  | 27.07 <sup>f</sup>     | 31.931 <sup>e</sup>    |
| E <sub>A</sub>       | 0.754 <sup>h</sup>    | 1.26 <sup>g</sup>       | 1.39 <sup>g</sup>    | 1.23 <sup>g</sup>   | 1.11 <sup>g</sup>      | 0.364 <sup>g</sup>     |
| I                    | 13.5984 <sup>a</sup>  | 11.2603 <sup>a</sup>    | 8.1517 <sup>a</sup>  | 7.8994 <sup>a</sup> | 7.3439 <sup>a</sup>    | 7.4167 <sup>a</sup>    |
| E <sub>tot</sub> [1] | -12.137               | -1018.436               | -7842.444            | -56432.436          | -163814.961            | -531150.838            |
| $\sigma$             | 16 x 10 <sup>3i</sup> | 0.07 x 10 <sup>6f</sup> | 4 x 10 <sup>2f</sup> | 3 <sup>f</sup>      | 8.7 x 10 <sup>6f</sup> | 4.8 x 10 <sup>6f</sup> |
| B                    | 0.2 <sup>c</sup>      | 444 <sup>g</sup>        | 98.0 <sup>g</sup>    | 74.9 <sup>g</sup>   | 56,6 <sup>g</sup>      | 42.9 <sup>d</sup>      |
| E                    | 0.2 <sup>g</sup>      | 545 <sup>g</sup>        | 112 <sup>g</sup>     | 79.9 <sup>g</sup>   | 52.9 <sup>g</sup>      | 15.8 <sup>g</sup>      |
| C11                  | 0.042 <sup>g</sup>    | 1076.4 <sup>g</sup>     | 165.77 <sup>g</sup>  | 124.0 <sup>g</sup>  | 72.30 <sup>g</sup>     | 48.8 <sup>g</sup>      |
| C12                  | 0.18 <sup>g</sup>     | 124.5 <sup>g</sup>      | 63.9 <sup>g</sup>    | 48.3 <sup>g</sup>   | 59.40 <sup>g</sup>     | 41.4 <sup>g</sup>      |
| C44                  | 0.11 <sup>g</sup>     | 578 <sup>g</sup>        | 79.6 <sup>g</sup>    | 67.1 <sup>g</sup>   | 22.03 <sup>g</sup>     | 14.8 <sup>g</sup>      |
| C12-C44              | 0.07                  | -453.5                  | -15.7                | -18.8               | 37.37                  | 26.6                   |

<sup>[1]</sup> Ref [5], <sup>a</sup> Ref [6], <sup>b</sup> Ref [7], <sup>c</sup> Ref [8], <sup>d</sup> Ref [9], <sup>e</sup> Ref [4], <sup>f</sup> Ref [10], <sup>g</sup> Ref [11], <sup>h</sup> Ref [12], <sup>i</sup> Ref [13] at 162 GPa and 6500 K.

**Tableau IV.2.2 :** La base de données incluant les propriétés physiques et mécaniques du groupe de carbone y compris l'hydrogène.

### IV.2.2.4. L'analyse en composantes principales (ACP)

D'après le graphe ACP des observations, représentant l'ensemble des données du groupe de carbone (**Figure IV.2.3**), nous notons que les deux PC les plus importants représentent 86,10% de l'ensemble de la variabilité dans l'ensemble de données (PC1 capte 62,00%, et PC2 24,10% de la variance totale. On remarque deux groupes distincts d'éléments; le groupe 1 contenant l'hydrogène H, et le groupe 2 est constitué des éléments du groupe IV (C, Si, Ge, Sn, Pb), celui-ci peut être divisée en trois sous-groupes. Le premier contient le carbone C qui est un non-métal et un mauvais conducteur d'électricité, comme l'hydrogène. La conductivité électrique est une propriété importante dans l'étude des matériaux elle nous informe pas uniquement sur la conduction électrique d'un métal, mais aussi sur sa structure et ses propriétés mécaniques. (Mais en dépit de l'absence d'électrons libres, l'hydrogène peut être

métallisé sous des pressions extrêmement élevées, et des températures basses); dans cet état, il est considéré comme métallique. Les propriétés atomiques déterminent si un élément chimique dans le tableau périodique est un métal ou un non métal. L'hydrogène liquide est considéré comme un métal, tandis que l'hydrogène solide n'atteint pas l'état métallique, même à haute pression [3].

Les allotropes de certains éléments montrent plus visiblement le comportement métallique ou non métallique que d'autres. Le diamant est une forme allotropique du carbone, il n'est pas métallique; cependant le carbone graphite est le seul élément non métallique et qui est conducteur de l'électricité.

Le deuxième sous-groupe contient le silicium et le germanium **Ge**, **Si** (**Figure IV.2.3**). Les propriétés physiques du germanium sont très similaires à ceux du silicium. Dans leur état solide élémentaire, ils sont tous les deux métalloïdes, parce qu'ils se comportent comme des semi-conducteurs, mais le silicium est principalement non-métallique. Leurs conductivités électriques peuvent être affectées à des degrés divers par le dopage. Le troisième sous-groupe contient l'étain (**Sn**) et le plomb **Pb**, qui sont certainement métalliques. Malgré leur adhésion aux tendances périodiques, les propriétés du groupe de carbone varient considérablement. Le carbone est un non-métal. L'étain et le plomb se comportent tout à fait comme des métaux. En ce qui concerne la conductivité électrique, les éléments peuvent être divisés en trois classes: 'les métaux', 'semi-métaux' et 'non-métaux' [2]. Les métaux ont une grande conductivité; elle peut être supérieure à  $10^5 \Omega^{-1} \cdot m^{-1}$ . Les semi-métaux ont des conductivités modérées à faibles; ils sont donc des conducteurs électriques intermédiaires [17]. Le semi-métal le plus courant est le silicium **Si**. La propriété la plus utile des semi-métaux est leur capacité variable à conduire le courant électrique. Les propriétés physiques des non-métaux sont opposées à celles des métaux. En regardant (**Figure IV.2.3**), nous notons que la distance entre le point représentant l'hydrogène **H** et les points représentant le groupe de carbone est plus courte que celle qui se trouve dans (**Figure IV.2.1**) avec les métaux alcalins. Cela signifie que l'hydrogène **H** partage plusieurs propriétés avec les éléments du groupe de carbone et a de nombreuses similitudes avec eux. L'hydrogène partage plusieurs caractéristiques chimiques avec le carbone et la famille du silicium. D'après le **Tableau IV.2.2**, l'hydrogène et les éléments du groupe IV ont des électronégativités similaires.

En thermodynamique, il est bien connu que l'électronégativité affecte les propriétés chimiques des éléments. On remarque que les propriétés d'hydrogène tel que l'énergie d'ionisation et l'affinité électronique sont similaires à ceux des éléments du groupe de

carbone. L'objectif de la classification périodique est de refléter les propriétés des éléments en fonction de leur nombre atomique et de leur périodicité [1].

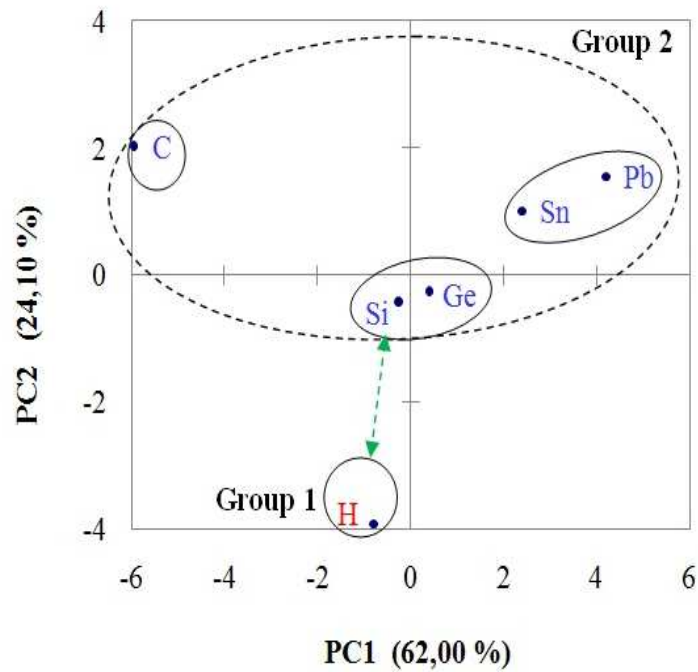


Figure IV.2.3. : Le graphe des observations du groupe du carbone y compris l'hydrogène.

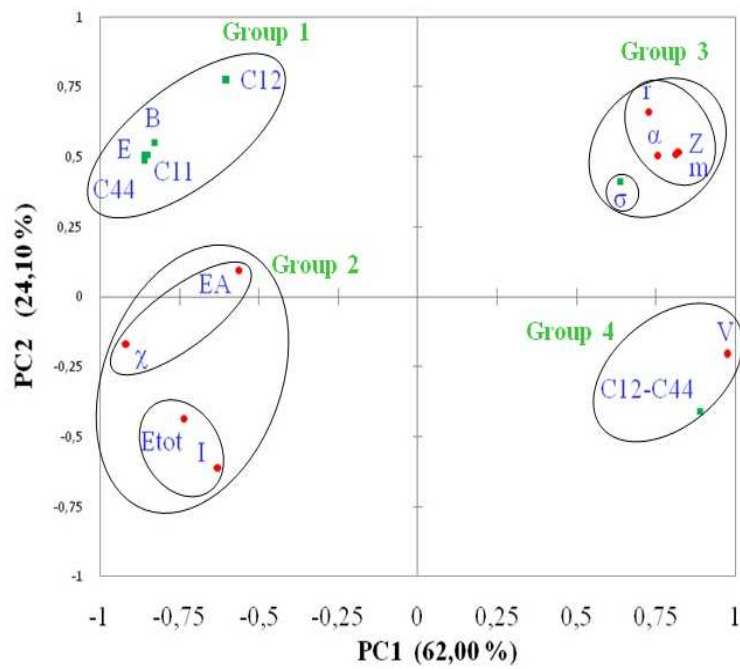


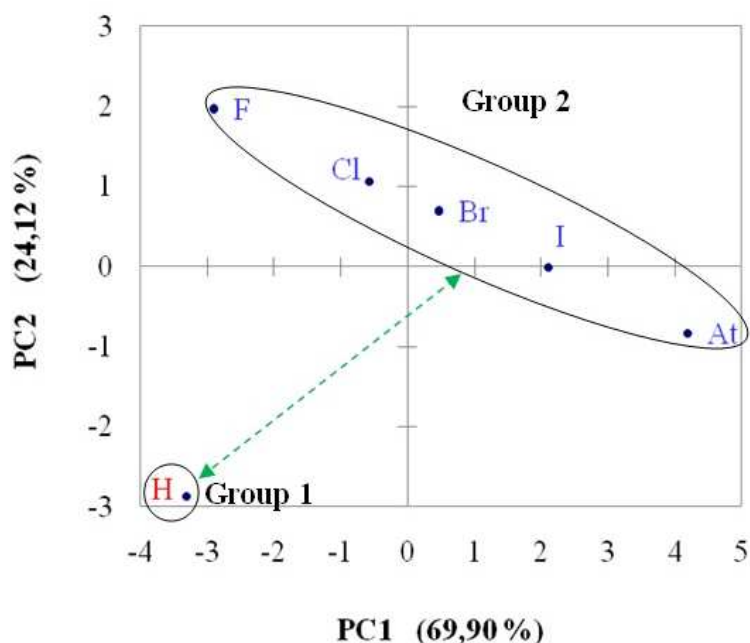
Figure IV.2.4. : Le graphe des variables du groupe du carbone y compris l'hydrogène.

IV.2.2.5. La bases de données des halogènes y compris l'hydrogène

| Propriété     | H                     | F                       | Cl                       | Br                      | I                       | At                  |
|---------------|-----------------------|-------------------------|--------------------------|-------------------------|-------------------------|---------------------|
| $\alpha$      | 0.667 <sup>h</sup>    | 0.557 <sup>h</sup>      | 2.18 <sup>h</sup>        | 3.05 <sup>h</sup>       | 4.7 <sup>h</sup>        | 6.0 <sup>h</sup>    |
| $\chi$        | 2.20 <sup>g</sup>     | 4.10 <sup>g</sup>       | 2.83 <sup>g</sup>        | 2.74 <sup>g</sup>       | 2.21 <sup>g</sup>       | 1.90 <sup>g</sup>   |
| $m$           | 1.00794 <sup>a</sup>  | 18.9984 <sup>a</sup>    | 35.453 <sup>a</sup>      | 79.904 <sup>a</sup>     | 126.9045 <sup>a</sup>   | 210 <sup>a</sup>    |
| $Z$           | 1                     | 9                       | 17                       | 35                      | 53                      | 85                  |
| $r$           | 0.25 <sup>b</sup>     | 0.50 <sup>b</sup>       | 1.00 <sup>b</sup>        | 1.15 <sup>b</sup>       | 1.40 <sup>b</sup>       | 1.27 <sup>b</sup>   |
| $V$           | <sup>f</sup> 92.23    | 28.39 <sup>f</sup>      | 37.69 <sup>f</sup>       | 39.03 <sup>f</sup>      | 42.74 <sup>f</sup>      | 49.82 <sup>f</sup>  |
| $E_A$         | 0.754 <sup>h</sup>    | 3.40 <sup>g</sup>       | 3.61 <sup>g</sup>        | 3.36 <sup>g</sup>       | 3.06 <sup>g</sup>       | 2.8 <sup>g</sup>    |
| $I$           | 13.5984 <sup>a</sup>  | 17.4228 <sup>a</sup>    | 12.9676 <sup>a</sup>     | 11.8138 <sup>a</sup>    | 10.4513 <sup>a</sup>    | 9.5352 <sup>b</sup> |
| $E_{tot} [1]$ | -12.137               | -2696.709               | -12481.164               | -69951.739              | -188164.803             | -578569.435         |
| $\sigma$      | 16 x 10 <sup>3i</sup> | 1.0 x 10 <sup>-4j</sup> | 1.0 x 10 <sup>-14k</sup> | 1.0 x 10 <sup>-4j</sup> | 1.0 x 10 <sup>-4j</sup> | 0 <sup>j</sup>      |

<sup>[1]</sup> Ref [5], <sup>a</sup> Ref [6], <sup>b</sup> Ref [7], <sup>f</sup> Ref [10], <sup>g</sup> Ref [11], <sup>h</sup> Ref [12], <sup>i</sup> Ref [13] at 162 GPa and 6500 K, <sup>j</sup> Ref [14], <sup>k</sup> Ref [15] at 203,15 K.

**Tableau IV.2.3.** La base de données incluant les propriétés physiques et mécaniques des halogènes y compris l'hydrogène.

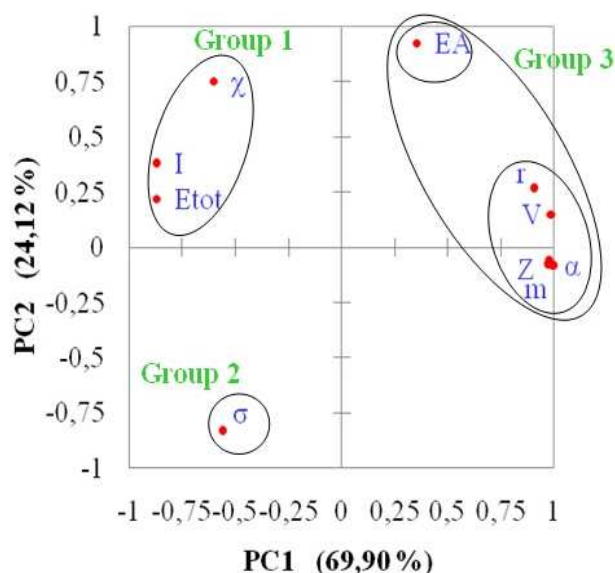
IV.2.2.6. L'analyse en composantes principales (ACP)

**Figure IV.2.5:** Le graphe des observations du groupe des halogènes y compris l'hydrogène.

Le graphe des observations du groupe des halogènes y compris l'hydrogène (**Figure IV.2.5**) génère deux pertinentes composantes principales qui représentent 94,02% de l'ensemble de la

variance dans la base de données; la première composante principale (PC1) capture 69,90% et la seconde (PC2) caractérise 24,12% de la variance totale.

D'après la (Figure IV.2.5) l'analyse ACP révèle deux groupes distincts d'éléments, l'hydrogène H que nous appellerons le groupe 1, et les halogènes (F, Cl, Br, I, A) le groupe 2. Il apparaît que l'hydrogène est très éloigné des halogènes. Cela signifie que l'hydrogène et les halogènes sont différents, et qu'ils ne partagent pas les mêmes propriétés (Tableau IV.2.3.), à l'exception de l'électronégativité  $\chi$  et de l'énergie d'ionisation I. En plus, l'hydrogène a une conductivité électrique  $\sigma$  supérieure à celle des halogènes, qui sont de mauvais conducteurs de l'électricité.



**Figure IV.2.6.** Le graphe des variables du groupe des halogènes y compris l'hydrogène.

En regardant le graphe des variables ACP des métaux alcalins y compris l'hydrogène dans (Figure IV.2.2.) quatre groupes distincts peuvent être observés: le Groupe 1, qui contient en particulier les propriétés physiques, et qui peut être décomposé en deux sous-groupes; le premier est constitué des propriétés atomiques tels que l'énergie d'ionisation I et l'électronégativité  $\chi$ , et le second de l'affinité d'électrons  $E_A$  et de l'énergie totale  $E_{tot}$ ; le deuxième groupe (Groupe 2) inclut les propriétés mécaniques B, E et C11, C12, C44 et la pression de Cauchy C12-C44, le troisième (groupe 3) se compose des propriétés physiques comme le rayon atomique r, le volume atomique V, la polarisabilité  $\alpha$  ainsi que la masse atomique m, et du numéro atomique Z. le quatrième groupe (groupe 4) contient seulement la

conductivité électrique  $\sigma$ . PC2 sépare clairement les propriétés physiques des propriétés atomiques (**Figure IV.2.2**). Les propriétés qui ont les mêmes valeurs de PC, tel que déterminé par les différents groupes, sont étroitement liées, tandis que celles avec des valeurs opposées de PC ont une corrélation divergente. Dans le groupe des métaux alcalins, les propriétés varient comme suit: une ressemblance importante peut être observée entre  $m$ ,  $Z$ ,  $V$  dans le Groupe 3, et entre toutes les propriétés du groupe 2; les points présentant les éléments sont très proches les uns des autres (**Figure IV.2.2**). Le point de  $m$  est le plus proche de celui de  $Z$ , indiquant son importance; La même remarque peut être faite pour le point de  $C_{11}$  et celui de  $E$ . Les groupes 1 et 3 sont inversement corrélés, donc pour les groupes 1 et 4, plusieurs corrélations peuvent être identifiées. Une augmentation de  $\chi$  et  $I$  se traduira par une augmentation d' $E_A$  et  $E_{tot}$  et une diminution de  $r$ ,  $V$ ,  $\alpha$ ,  $m$ ,  $Z$  et  $\sigma$ . En même temps,  $E$  et  $B$  sont corrélées et étroitement liées aux constantes élastiques  $C_{11}$ ,  $C_{12}$ ,  $C_{44}$  et à la pression de Cauchy  $C_{12}$ - $C_{44}$ .

Dans le graphe des variables du groupe de carbone y compris l'hydrogène la (**Figure IV.2.4**), quatre groupes distincts peuvent être observés: le premier (groupe 1) contient  $C_{11}$ ,  $C_{12}$ , et  $C_{44}$ , en plus de  $B$  et  $E$ . Le deuxième groupe (Groupe 2) peut être décomposé en deux sous-groupes; un contient  $E_A$  et  $\chi$ , et l'autre  $I$  et  $E_{tot}$ . Dans le troisième groupe (groupe 3) il y'a deux sous-groupes; l'un comprend  $r$ ,  $\alpha$ ,  $m$  et  $Z$  et l'autre  $\sigma$ ; Le quatrième groupe (groupe 4) contient  $V$  et  $C_{12}$ - $C_{44}$ . Dans le groupe de carbone, les propriétés varient comme suit: dans le groupe 1, une forte corrélation est observée entre  $B$ ,  $E$  et  $C_{11}$ ,  $C_{12}$ ,  $C_{44}$ ; tandis que  $E$  et  $B$  sont corrélées et étroitement liées aux constantes élastiques  $C_{11}$ ,  $C_{12}$ ,  $C_{44}$ . De même,  $m$  et  $Z$  dans le groupe 3 sont également fortement corrélées. Le point de  $m$  est le plus proche de celui de  $Z$ ; La même remarque peut être faite pour le point de  $E$  et ceux de  $C_{11}$ ,  $C_{44}$ . Les groupes 1 et 4 sont inversement corrélés. De même, les groupes 2 et 3 présentent une corrélation inverse, et plusieurs corrélations peuvent être mentionnées. Une augmentation de  $\chi$  et  $E_A$  se traduira par une hausse des  $I$  et  $E_{tot}$  et une baisse de  $r$ ,  $\alpha$ ,  $m$ ,  $Z$  et  $\sigma$ . Par conséquent, une augmentation de  $V$  conduira à une diminution de  $B$ , et une augmentation de  $E$  provoquera une augmentation de  $B$ . En outre, une augmentation de  $B$  donne une diminution de  $C_{12}$ - $C_{44}$ .

En regardant le graphe ACP des variables du groupe des halogènes y compris l'hydrogène (**Figure IV.2.6**), trois groupes distincts peuvent être spécifiés: le premier groupe (Groupe 1) contient les propriétés  $I$ ,  $\chi$  et  $E_{tot}$ , le deuxième groupe (Groupe 2) contient  $\sigma$ ; tandis que le (groupe 3) peut être décomposé en deux sous-groupes, l'un contient les propriétés  $r$ ,  $V$ ,  $\alpha$ ,  $m$ , et  $Z$ , et l'autre  $E_A$ .

Après examination de la **figure IV.2.6**, nous remarquons que les propriétés  $I$ ,  $\chi$  et  $E_{\text{tot}}$  sont fortement corrélées à  $\sigma$ ; tandis que les groupes 2 et 3 ont des valeurs inverses de PC, ce qui signifie que les propriétés contenues dans le groupe 3 sont inversement corrélées à  $\sigma$ . Il est facile de noter que ces propriétés sont liées à la conductivité électrique, qui est une caractéristique importante dans le tableau périodique des éléments.

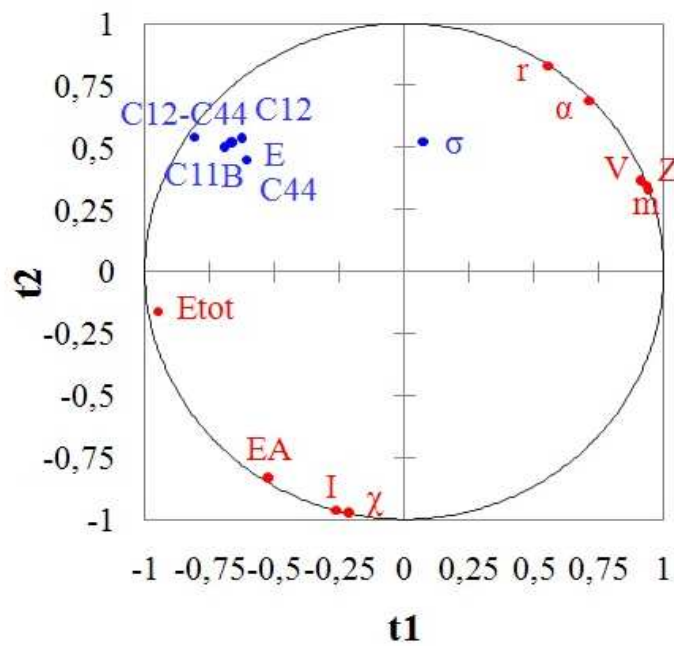
### **IV.2.3. La régression par les Moindres Carrés Partiels (PLS)**

Cette régression est basée sur la relation entre les propriétés mécaniques ( $E$ ,  $B$ ,  $\sigma$ ,  $C_{11}$ ,  $C_{12}$ ,  $C_{44}$ ,  $C_{12}-C_{44}$ ) et les propriétés physiques ( $\alpha$ ,  $\chi$ ,  $m$ ,  $Z$ ,  $r$ ,  $V$ ,  $E_A$ ,  $I$ ,  $E_{\text{tot}}$ ). L'approche PLS est appliquée pour élucider les relations entre ces propriétés. Elle crée un ensemble de composantes, en commençant par une base de données composées de diverses observations caractérisées par différentes variables. C'est une technique de réduction des données, pour réduire les variables  $X$  à une série de composantes non apparentées qui illustrent la variation dans les données. La régression PLS est d'un intérêt particulier car elle peut analyser les données corrélées et les variables redondantes; elle peut aussi modéliser les différentes caractéristiques en même temps. Les résultats de la PLS pour les trois bases de données sont présentés dans les figures : (**Figure IV.2.7**, **Figure IV.2.8** et **Figure IV.2.9**). Ils sont obtenus en utilisant le deux premières composantes ( $t_1, t_2$ ). Le nombre de facteurs est automatiquement sélectionné; la première composante PLS est plus fiable que les composantes suivantes. Finalement, le choix des deux composantes PLS permet de faire de pratiques visualisations. Les variables  $X$  et  $Y$  sont visualisées sur un graphe, nommé " le cercle de corrélation ", en fonction de leurs relations avec les facteurs PLS,  $t_1$  et  $t_2$ , qui pourrait être interprété de la même manière que pour le graphe des variables dans l'ACP; les cosinus des angles formés par les vecteurs-variables dans le plan indiquent les corrélations entre les variables d'origine [19]. Les variables les plus importantes se trouvent à la périphérie du cercle de corrélation, alors que les variables, qui n'apportent que de petites ou aucune contribution à la discrimination, sont à proximité du centre du cercle [20].

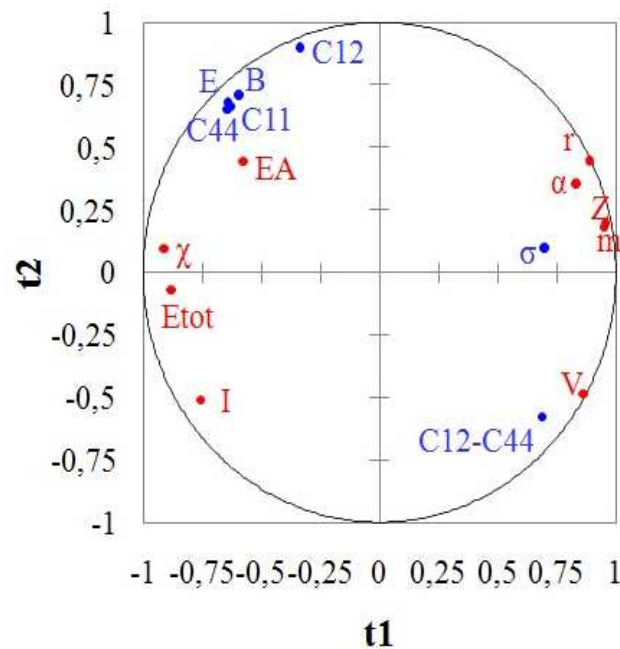
Pour la base de données des métaux alcalins, une régression PLS des 7 variables  $Y$  en fonction des 9 variables  $X$ , en utilisant les deux premiers facteurs  $t_1$  et  $t_2$ , est réalisée comme le montre la **Figure IV.2.7**. Le graphe obtenu est le cercle de corrélation entre les variables des propriétés mécaniques, et celles des propriétés physiques dans les deux composantes PLS ( $t_1, t_2$ ). Ceci permet la visualisation des similitudes entre les propriétés physiques des variables ( $x_i$ ) et les deux facteurs PLS  $t_1$  et  $t_2$  d'une part, et entre les propriétés physiques des variables

( $Y_i$ ) et  $t_1$ ,  $t_2$  d'autre part. Nous trouvons de fortes corrélations entre  $m$ ,  $Z$ ,  $V$ , d'une part, et  $\chi$ ,  $I$  de l'autre part; ainsi entre les constantes d'élasticité  $C_{11}$ ,  $C_{12}$ , et  $C_{44}$  d'un côté et  $E$ ,  $B$  et la pression de Cauchy  $C_{12}$ - $C_{44}$  d'un autre côté. Une corrélation négative est également notée entre  $E_{\text{tot}}$  et  $m$ ,  $Z$ , et la même tendance est observée entre  $r$  et  $E_A$ . En ce qui concerne les variables explicatives ( $x_i$ ), la conductivité  $\sigma$  électrique n'est pas bien représentée dans les deux premières composantes; elle est à proximité du centre du cercle de corrélation. Ceci peut s'expliquer par le fait que cette variable définit à peine les tendances des propriétés des éléments. Elle est moins corrélée aux variables  $X$ , et donc semble être moins corrélée aux facteurs PLS  $t_1$  et  $t_2$ . En regardant le graphe PLS représentant la base de données du groupe de carbone (**Figure IV.2.8**), il y'a de fortes corrélations observées entre  $m$ ,  $Z$  d'un côté et entre les constantes élastiques  $C_{11}$ ,  $C_{12}$ ,  $C_{44}$ , et  $E$ ,  $B$  d'un autre côté. Il semble également en (**Figure IV.2.8**) qu' $E_{\text{tot}}$  est corrélée négativement à  $m$  et  $Z$ . Une tendance similaire est observée entre  $V$  et  $E_A$ . La conductivité électrique  $\sigma$  est à proximité du centre du cercle de corrélation, donc elle apporte une contribution réduite à la discrimination.

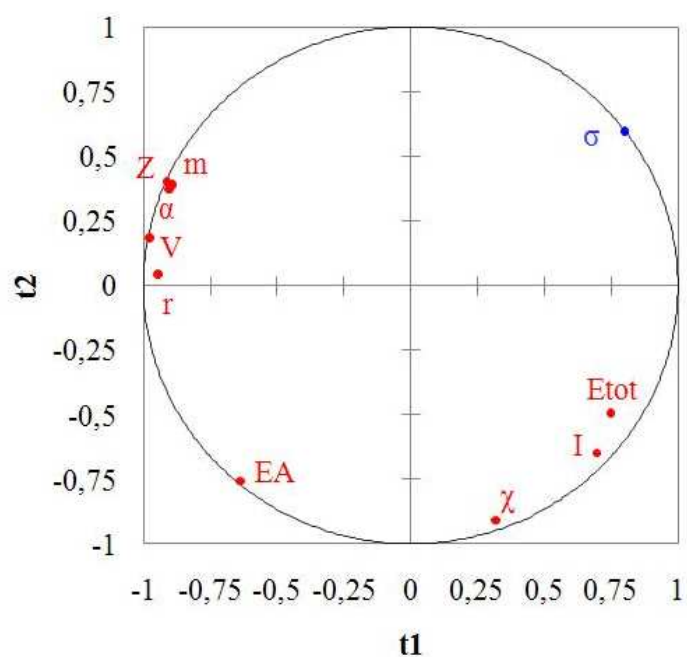
Pour la base de données des halogènes, la régression PLS dont la variable  $Y$  est la conductivité électrique  $\sigma$  est en fonction des variables  $X$ , qui sont les 9 propriétés physiques ( $\alpha$ ,  $\chi$ ,  $m$ ,  $Z$ ,  $r$ ,  $V$ ,  $E_A$ ,  $I$ ,  $E_{\text{tot}}$ ), avec  $t_1$  et  $t_2$ , les résultats sont visualisés dans le cercle de corrélation dans la **Figure IV.2.9**. Un premier aperçu indique que toutes les propriétés se situent à la périphérie du cercle, ce qui signifie qu'elles sont toutes des variables discriminantes et participantes.



**Figure IV.2.7.** Le cercle de corrélation entre les propriétés mécaniques, et les propriétés physiques dans les deux composantes PLS ( $t_1$ ,  $t_2$ ) (la base de données des métaux alcalins).



**Figure IV.2.8.** Le cercle de corrélation entre les propriétés mécaniques, et les propriétés physiques dans les deux composantes PLS ( $t_1$ ,  $t_2$ ) (la base de données du groupe de carbone).



**Figure IV.2.9.** Le cercle de corrélation entre les propriétés mécaniques, et les propriétés physiques dans les deux composantes PLS ( $t_1$ ,  $t_2$ ) (la base de données des halogènes).

#### IV.2.4. Conclusion

L'atome d'hydrogène a un seul électron, donc on pourrait penser que ses propriétés chimiques sont limitées, mais il a diverses propriétés chimiques et peut être combiné avec presque tous les éléments. L'hydrogène est différent des autres éléments de la classification périodique, car il possède des propriétés exceptionnelles, il a été placé à la tête du groupe des alcalins, car il a un seul électron de valence. Cependant, cette position ne révèle pas ses propriétés chimiques et physiques [18].

Par conséquent, d'après les résultats représentés par les graphes des observations dans les figures IV.2.1, IV.2.3., et IV.2.5, on peut en déduire que l'hydrogène partage certaines propriétés avec certains éléments et groupes du tableau périodique, tels que le groupe du carbone, mais pas totalement; à vrai dire l'hydrogène est un élément à part, et ne s'ajuste harmonieusement à aucun groupe.

Dans cette étude, la méthode d'extraction des données a été utilisée comme une analyse exploratoire des données. C'est une approche qui devient de plus en plus importante dans la recherche, en utilisant les différentes méthodes statistiques multivariées telles que l'ACP et la régression PLS pour visualiser les positions des éléments l'un par rapport à l'autre dans chaque base de données, et pour étudier les relations entre leurs propriétés. Ce travail montre que les outils statistiques tels que l'ACP combinée avec la PLS peuvent être très utiles pour identifier les corrélations entre les groupes de variables dans un ensemble de données afin de déterminer les variables (propriétés) qui devraient être utilisées pour classer les éléments. Cette étude présente une nouvelle méthode pour étudier le comportement d'un élément dans le tableau périodique, et les tendances de leurs propriétés à l'aide de l'exploration de données et de ses techniques appropriées. Nous avons trouvé que l'hydrogène partage certaines similitudes avec certains éléments et groupes dans le tableau périodique tels que le groupe de carbone. En fait, l'hydrogène est toujours considéré comme un élément à part, distinct et spécial.

**IV.2.5. Références**

- [1] Marshall W. Cronyn, The Proper Place for Hydrogen in the Periodic Table, *Journal of Chemical Education*. N° 8. **80**, 947-951 (2003).
- [2] Albert Hérold, An arrangement of the chemical elements in several classes inside the periodic table according to their common properties, *C. R. Chimie*. **9** (2006) 148–153.
- [3] P. P. Edwards, M. T. J. Lodge, F. Hensel and R. Redmer, a metal conducts and a non-metal doesn't...', 941-965368 *Phil. Trans. R. Soc. A*. **368** (2010) 941-965.
- [4] S.L. Shang, A. Saengdeejing, Z.G. Mei, D.E. Kim, H. Zhang, S. Ganeshan, Y. Wang, Z.K. Liu, First-principles calculations of pure elements: Equations of state and elastic stiffness constants, *Computational Materials Science*. **48** (2010) 813–826.
- [5] NIST (National Institute of Standards and Technology), [www.nist.gov](http://www.nist.gov) (Accessed June 10, 2012).
- [6] R.A. Dragoset, A. Musgrove, C.W. Clark, and W.C. Martin, (2003), *Periodic Table: Atomic Properties of the Elements (Version 4)*, NIST SP 966. <<http://physics.nist.gov/PT>>. NIST (National Institute of Standards and Technology), Gaithersburg, MD, (2012).
- [7] J.C. Slater, *J. Chem. Phys.* **39** (1964) 3199, <http://www.webelements.com/>
- [8] Charles Kittel, *Introduction to Solid State Physics*, 8th edition. Hoboken, (NJ: John Wiley & Sons), Inc, 2005.
- [9] Y. Makino, S. Miyake, Estimation of bulk moduli of compounds by empirical relations between bulk modulus and interatomic distance, *Journal of Alloys and Compounds*. **313** (2000) 235–241.
- [10] The Massachusetts Institute of Technology. <<http://www.chemicool.com>>. (Accessed 20 July 2012).
- [11] Martienssen Warlimont, *Springer handbook of condensed matter and materials data* Volume 1- Hans Warlimont, Springer Berlin Heidelberg (2005).
- [12] David R. Lide, ed., *CRC Handbook of Chemistry and Physics*, 90th Edition, (CRC Press/Taylor and Francis, Boca Raton, FL), 2010.
- [13] V. Ya. Ternovoi, A. A. Pyalling, D. N. Nikolaev, and S. V. Kvitov, Conductivity of multiple shock compressed hydrogen along 135 and 180 GPa isobars, *CP1195*, shock compression of condensed matter. (2009) 915-918.
- [14] <http://easycalculation.com/chemistry/periodic-table.php>. (Accessed 29 June 2013).

- [15] Peter Schmittinger, Chlorine, Principles and industrial practice, 1st Edition (2000), (Wiley-VCH Verlag GmbH, D-69469 Weinheim) Germany.
- [16] Shin, E. C., Craft, B. D., Pegg, R. B., R. Phillips, D., and Eitenmiller, R. R. (2010), “Chemometric approach to fatty acid profiles in Runner-type peanut cultivars by principal component analysis (PCA)”, *Food Chemistry*, 119, 1262–1270.
- [17] Wiberg, E., Wiberg, N., and Holleman, A. F. (2001), « *Inorganic chemistry* », Academic Press.
- [18] Atkins, P.W., Overton, T.L., Rourke, J.P., Weller, M.T., and Armstrong, F.A. (2010). *Inorganic Chemistry*, 5th ed, (Oxford University Press).
- [19] Tenenhaus, M., Pagès, J., Ambroisine, L., and Guinot, C. (2005), « PLS methodology to study relationships between hedonic judgements and product characteristics », *Food Quality and Preference*, 16, 315–325.
- [20] Sjöström, M., and Wold, S. (1986), “Partial Least Square Discriminant Plots”, *Pattern Recognition In Practice II*. pp. 461-470.
- [21] Laing, M. (2007), « Where to put hydrogen in a periodic table?, *Foundations of Chemistry* », 9, 127–137.
- [22] Cronyn, M. W. (2003), « The Proper Place for Hydrogen in the Periodic Table », *Journal of Chemical Education*, N° 8. 80, 947-951.

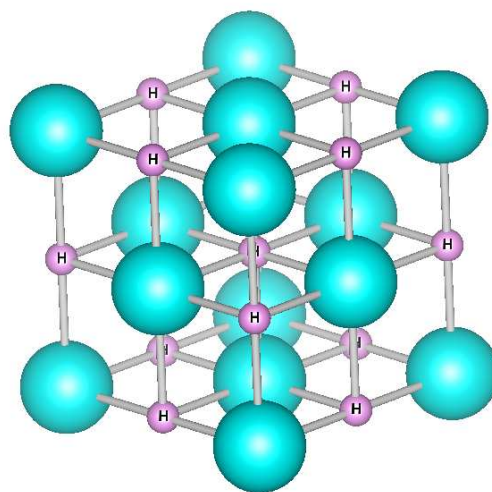
## Troisième partie

### IV.3. Investigation des propriétés structurales et mécaniques des hydrures alcalins par les calculs ab initio et l'approche du datamining

#### IV.3.1. Introduction

Les hydrures sont des matériaux prometteurs; qui peuvent être utilisés dans les batteries rechargeables, les cellules, les réfrigérateurs, les capteurs, les commutateurs optiques, ainsi que dans le stockage de chaleur, en industrie nucléaire et dans la purification de l'hydrogène. Par conséquent, il est nécessaire d'étudier leurs propriétés, pour pouvoir concevoir des hydrures appropriés pour les applications spéciales [1].

L'étude des hydrures métalliques est importante car elle peut aider à comprendre les propriétés de l'hydrogène qui améliore les propriétés de stockage [2]. Les hydrures que l'hydrogène forme par réaction avec les métaux alcalins (Li, Na, K, Rb, et Cs) sont stoechiométriquement stable. Les hydrures alcalins MH (M = Li, Na, K, Rb, et Cs) représentent un domaine intéressant de la recherche, parce qu'ils sont utiles dans la fusion ce qui permet d'acquérir des carburants de haute énergie [3]. Depuis la première étude expérimentale par rayonnement X, en 1931, il a été démontré [4] que les hydrures alcalins cristallisent dans la structure NaCl (B1) à température ambiante et pression ambiante, comme la structure de Li, Na, K, Rb, et Cs [5; 6].



**Figure IV.3.1.** Structure cristalline de la structure NaCl (B1) des hydrures alcalins MH dans les conditions ambiantes. Les cercles bleus représentent M = Li, Na, K, Rb, Cs, et les cercles violets représentent les atomes d'hydrogène.

L'objectif de ce travail est d'examiner les propriétés structurales et mécaniques des hydrures alcalins obtenues par les calculs ab initio en utilisant l'approche datamining, et la technique d'Analyse en Composantes Principales (ACP) pour explorer de nouvelles relations afin de révéler les tendances dans la base de données et montrer les similitudes et les corrélations existantes entre les variables (les propriétés) et les hydrures alcalins.

Pour réaliser cette étude, nous avons construit une base de données (**Tableau IV.3.2**) à partir des propriétés mécaniques calculées, par les calculs ab initio implémentés dans le code VASP [7; 8] dans le cadre de la Théorie de la Fonctionnelle de la Densité DFT [9; 10], les énergies d'échange et de corrélation des structures ont été traitées par l'approximation du gradient généralisé (GGA) [11] et le procédé de projection des ondes augmentées (PAW) [12; 13; 14], ça confirme la précision des méthodes tout électrons, avec l'efficacité de la méthode des pseudo-potentiels [15]. La fonctionnelle d'échange et de corrélation est adoptée pour décrire l'interaction d'échange-corrélation. Le test de convergence a donné une énergie de coupure de 312,50 eV lorsque la base d'onde plane est utilisée. Suivant le schéma Monkhorst-Pack [16], pour l'intégration nous avons utilisé une maille de (13×13×13) points-k dans la première zone de Brillouin, et qui a été utilisé pour tous les calculs. Dans tous les cas, les optimisations de géométrie ont été effectuées et les convergences ont été assumées lorsque les forces sur les atomes étaient inférieures à 1 MeV / Å.

### IV.3.2. Base de données

|                   | LiH                    | NaH                     | KH                      | RbH                     | CsH                      |
|-------------------|------------------------|-------------------------|-------------------------|-------------------------|--------------------------|
| $a_0(\text{Å})$   | 4.02 <sup>a</sup>      | 4.833 <sup>a</sup>      | 5.699 <sup>a</sup>      | 6.068 <sup>a</sup>      | 6.437 <sup>a</sup>       |
|                   | 4.057[53] <sup>b</sup> | 4.870[55] <sup>b</sup>  | 5.70[57] <sup>b</sup>   | 6.04[59] <sup>b</sup>   | 6.388 [61] <sup>b</sup>  |
| $B_0(\text{GPa})$ | 4.02 [54] <sup>c</sup> | 4.834[56] <sup>c</sup>  | 5.699 [58] <sup>c</sup> | 6.075 [60] <sup>c</sup> | 6.429 [62] <sup>c</sup>  |
|                   | 34.614 <sup>a</sup>    | 22.507 <sup>a</sup>     | 13.017 <sup>a</sup>     | 10.944 <sup>a</sup>     | 8.686 <sup>a</sup>       |
|                   | 34.7[55] <sup>b</sup>  | 19.40 [55] <sup>b</sup> | 15.6 [66] <sup>b</sup>  | 10.00[66] <sup>b</sup>  | 8.00 [61] <sup>b</sup>   |
| $B'_0$            | 34.30[58] <sup>c</sup> | 22.60[62] <sup>c</sup>  | 13.30 [58] <sup>c</sup> | 11.1 [62] <sup>c</sup>  | 8.80 [58] <sup>c</sup>   |
|                   | 3.141 <sup>a</sup>     | 3.522 <sup>a</sup>      | 3.539 <sup>a</sup>      | 3.649 <sup>a</sup>      | 3.508 <sup>a</sup>       |
|                   | 3.40 [67] <sup>b</sup> | 4.40 [55] <sup>b</sup>  | 4.00 [66] <sup>b</sup>  | 3.9 [66] <sup>b</sup>   | 4.00 [70] <sup>b</sup>   |
|                   | 3.3[68] <sup>c</sup>   | 3.62 [69] <sup>c</sup>  | 3.70 [58] <sup>c</sup>  | 3.47 [58] <sup>c</sup>  | 3.0365 [71] <sup>c</sup> |

<sup>a</sup> Le présent travail

<sup>b</sup> Valeurs expérimentales

<sup>c</sup> Valeurs théoriques

**Tableau IV.3.1:** Les valeurs calculées du paramètre de maille  $a_0$  (Å), le module de compressibilité ( $B_0$ ) (GPa) et la dérivé première de la pression ( $B'_0$ ) des hydrures alcalins dans la structure NaCl (B1), avec quelques autres résultats.

|                | LiH   | NaH   | KH  | RbH   | CsH   |
|----------------|---|---|---|---|---|
| C11(GPa)       | 56.97 <sup>a</sup><br>65.31 [72] <sup>b</sup>                           | 49.90 <sup>a</sup><br>47.3 [73] <sup>b</sup><br>48.3 [74] <sup>c</sup>  | 23.49 <sup>a</sup><br>23.0 [60] <sup>c</sup>  | 21.914 <sup>a</sup><br>24.6 [58] <sup>c</sup> | 15.817 <sup>a</sup><br>16.4 [60] <sup>c</sup> |
| C12(GPa)       | 23.43 <sup>a</sup><br>17.5 [75] <sup>b</sup><br>22.3 [76] <sup>c</sup>  | 8.81 <sup>a</sup><br>8.0[60] <sup>c</sup>                               | 7.78 <sup>a</sup><br>8.35 [71] <sup>c</sup>   | 5.458 <sup>a</sup><br>4.6 [58] <sup>c</sup>   | 5.12 <sup>a</sup><br>6.761 [71] <sup>c</sup>  |
| C44(GPa)       | 46.156 <sup>a</sup><br>46.2 [73] <sup>b</sup><br>46.0 [54] <sup>c</sup> | 22.969 <sup>a</sup><br>22.5 [73] <sup>b</sup><br>22.7 [58] <sup>c</sup> | 11.804 <sup>a</sup><br>12.9 [58] <sup>c</sup> | 9.299 <sup>a</sup><br>8.2 [58] <sup>c</sup>   | 7.397 <sup>a</sup><br>6.7 [60] <sup>c</sup>   |
| <i>G</i> (GPa) | 30.768 <sup>a</sup>   | 21.966 <sup>a</sup><br>23 [71] <sup>c</sup>                             | 10.025 <sup>a</sup><br>13 [71] <sup>c</sup>   | 8.85 <sup>a</sup><br>10.2 [71] <sup>c</sup>   | 6.49 <sup>a</sup><br>9.15 [71] <sup>c</sup>   |
| <i>E</i> (GPa) | 71.20 <sup>a</sup>  | 49.72 <sup>a</sup><br>53 [71] <sup>c</sup>                              | 23.931 <sup>a</sup>                           | 20.91 <sup>a</sup>                            | 15.587 <sup>a</sup>                           |
| $\nu$          | 0.157 <sup>a</sup><br>0.11 [71] <sup>c</sup>                            | 0.13 <sup>a</sup><br>0.13 [71] <sup>c</sup>                             | 0.19 <sup>a</sup><br>0.21 [71] <sup>c</sup>   | 0.18 <sup>a</sup><br>0.23 [71] <sup>c</sup>   | 0.20 <sup>a</sup><br>0.22 [71] <sup>c</sup>   |
| <i>A</i>       | 2.75 <sup>a</sup>   | 1.11 <sup>a</sup><br>0.88 [71] <sup>c</sup>                             | 1.50 <sup>a</sup><br>1.27 [71] <sup>c</sup>   | 1.13 <sup>a</sup><br>1.18 [71] <sup>c</sup>   | 1.38 <sup>a</sup><br>1.18 [71] <sup>c</sup>   |
| <i>B/G</i>     | 1.125 <sup>a</sup>  | 1.024 <sup>a</sup><br>1.13 [71] <sup>c</sup>                            | 1.29 <sup>a</sup><br>1.2 [71] <sup>c</sup>    | 1.236 <sup>a</sup><br>1.37 [71] <sup>c</sup>  | 1.33 <sup>a</sup><br>1.31 [71] <sup>c</sup>   |

<sup>a</sup> Le présent travail

<sup>b</sup> Valeurs expérimentales

<sup>c</sup> Valeurs théoriques

**Tableau IV.3.2:** Les Constantes mécaniques et élastiques des hydrures alcalins dans la structure NaCl (B1), avec quelques autres résultants, le module de compressibilité ( $B_0$ ), module de cisaillement ( $G$ ), constantes élastiques ( $C_{ij}$ ), le module d'Young ( $E$ ) sont en (GPa). Le rapport  $B/G$ , le facteur d'anisotropie ( $A$ ) et le coefficient de Poisson ( $\nu$ ) sont des paramètres sans dimension.

### IV.3.3. L'analyse en composantes principales (ACP)

#### IV.3.3.1. Les propriétés structurales

Les paramètres structuraux des hydrures alcalins (LiH, NaH, KH, RbH et CsH) ont été optimisés, et l'énergie totale a été étudiée en fonction du volume. Les paramètres de la structure, module de compressibilité ( $B_0$ ), la dérivé première de la pression ( $B'_0$ ) ont été déterminés en ajustant les énergies totales à l'équation d'état de Murnaghan [17]; les valeurs sont données dans le (Tableau IV.3.1). A partir de ces valeurs, on peut noter que les valeurs calculées sont en bon accord avec ceux obtenues dans les travaux expérimentaux et théoriques précédents; cela prouve la validité de notre modèle théorique. L'approximation du gradient généralisé (GGA) estime la fiabilité de l'énergie totale à la fois sur la densité d'électrons et de son gradient. La GGA améliore généralement les paramètres structuraux du matériau [18].

#### IV.3.3.2. Les propriétés mécaniques

Le cristal cubique possède trois constantes élastiques indépendantes,  $C_{11}$ ,  $C_{12}$  et  $C_{44}$  [19] qui ont été déterminées par ajustement des énergies totales d'un cristal contraint à une équation de contraintes polynomiale du quatrième ordre, selon la loi de Hooke [20]. Trois types de contraintes, à savoir la variation de volume, les contraintes de cisaillement tétragonal et rhomboédrique et le volume conservé, ont été appliquées à la structure optimisée. Les constantes élastiques sont calculées et rapportées dans le (Tableau IV.3.2), avec quelques autres données disponibles, qui indiquent que les structures cubiques pour chaque hydrure alcalin sont mécaniquement stables parce qu'elles satisfont les critères de stabilité du cristal cubique [21].

$$C_{44} > 0 \quad (\text{IV-1})$$

$$C_{11} > |C_{12}| \quad (\text{IV-2})$$

$$C_{11} + 2C_{12} > 0 \quad (\text{IV-3})$$

Le module de cisaillement (Tableau IV.3.2) joue un rôle dominant dans la prédiction de la dureté du matériau. Il a été trouvé que LiH a le module de cisaillement le plus élevé parmi les hydrures alcalins, ce qui signifie que LiH a une grande dureté. Par conséquent, LiH est plus résistant que les autres hydrures alcalins. Le rapport  $B/G$  est généralement utilisé pour prédire le comportement ductile et fragile des matériaux. Le matériau est ductile lorsque

$B/G > 1,75$  [22], et il a un caractère fragile quand  $B/G \leq 1,75$  [23]. Un tel comportement est indirectement lié au facteur d'anisotropie. D'après les valeurs de  $B/G$  calculées, présentées dans le (**Tableau IV.3.2**), on constate qu'elles sont toutes inférieures à 1,75, ce qui indique que les hydrures alcalins se comportent d'une manière fragile dans les conditions ambiantes, et peuvent être donc considérés comme des matériaux fragiles. Le module d'Young  $E$  est connue comme une quantité de mesure de la rigidité d'un matériau élastique [24]. En analysant les valeurs de  $E$  présentées dans le (**Tableau IV.3.2**), on peut noter que le **LiH** est le plus rigide parmi les hydrures alcalins. La rigidité des hydrures alcalins diminue de **LiH** à **CsH**. Comme le montrent les tableaux **IV.3.1** et **IV.3.2**, les valeurs calculées du module de compressibilité, du module de cisaillement, et du module d'Young diminuent selon l'ordre **LiH** < **NaH** < **KH** < **RbH** < **CsH**. Le rapport de Poisson  $\nu$ , qui varie généralement de -1 à 0,5, est utilisé pour estimer la stabilité du cristal contre le cisaillement [38],  $\nu$  mesure la résistance du matériau au cisaillement [36], il fournit plus d'informations sur les caractéristiques des forces de liaison que tout autre constante élastique [48]; il est généralement égale à 1/3 pour les matériaux ductiles et moins de 1/3 pour les matériaux fragiles [49]. Les valeurs du rapport de Poisson  $\nu$ , figurant dans le **tableau IV.3.2**, sont inférieures à 1/3, ce qui indique que les hydrures alcalins sont des matériaux fragiles.

L'anisotropie  $A$  (également appelée le rapport d'anisotropie de Zener) est une mesure du degré d'anisotropie dans les solides [25]. Elle est égale à 1 pour un matériau isotrope. Lorsque  $A$  est inférieur ou supérieur à 1,  $A$  est utilisée pour mesurer le degré d'anisotropie élastique [19]. Selon les valeurs d'anisotropie calculées et indiqués dans le **tableau IV.3.2**, on peut noter que les hydrures alcalins sont caractérisés par une anisotropie élastique.

#### **IV.3.4. Analyse par l'approche data mining**

Les valeurs du module de compressibilité ( $B$ ), le module de cisaillement ( $G$ ), le module de Young ( $E$ ), le coefficient de Poisson ( $\nu$ ), le rapport d'anisotropie ( $A$ ), le rapport  $B/G$ , et les constantes élastiques ( $C_{11}$ ,  $C_{12}$  et  $C_{44}$ ) calculées par les calculs ab initio et affichées dans le **tableau IV.3.2**, les hydrures alcalins (**LiH**, **NaH**, **KH**, **RbH** et **CsH**) constituent la base de données analysées par l'approche data mining et la technique d'ACP. L'analyse en composantes principales (ACP) effectuée sur la base de données, a donné deux importantes composantes principales PC, où PC1 capture 83,07% de la variance totale dans la base de données, alors que PC2 capture 16,84%, comme le montre (**Figure IV.3.2**). Le graphe des observations des deux premières composantes principales pour les hydrures alcalins (**Figure**

IV.3.2), révèle deux groupes distincts d'hydrures. Groupe 1 comprend LiH et NaH ayant une valeur positive de PC1, tandis que le groupe 2 contient KH, RbH et CsH qui ont une valeur négative de PC1. Il est à noter que l'axe de PC2 sépare le groupe 1 du groupe 2. Le groupe 1 contient les hydrures alcalins avec des valeurs élevées du module de B, du module de cisaillement, du module d'Young, et des constantes élastiques (C11, C12 et C44) et de faibles valeurs du coefficient de Poisson et du ratio B/G; contrairement au groupe 2 qui contient les hydrures ayant de faibles valeurs du module de compressibilité, du module de cisaillement, du module d'Young, des constantes élastiques (C11, C12 et C44) et des valeurs élevées du coefficient de Poisson et du rapport B/G. En outre, les hydrures alcalins du groupe 2 ont des valeurs très proches de B, G, E, C11, C12, C44,  $\nu$ , et B/G. En regardant le graphe d'ACP des variables (**Figure IV.3.3**), on peut noter que les propriétés mécaniques, à savoir B, G, E, C11, C12, C44, et A, ont les mêmes valeurs de la composante principale (PCs), ce qui signifie qu'elles sont fortement corrélées parce que leurs points sont proches les uns des autres; cela révèle l'importance de ces propriétés. Celles-ci sont inversement corrélées à  $\nu$  et B/G, car elles ont des valeurs de PC opposées. L'axe PC2 sépare clairement B, G, E, C11, C12, C44, A de  $\nu$  et B/G; cela exprime la forte différence entre elles. G représente la dureté contrairement à B/G et  $\nu$  qui expriment la ductilité; donc une augmentation de la dureté, conduira à une diminution de la ductilité (soit une augmentation de la fragilité), ce qui prouve que les hydrures alcalins sont des matériaux fragiles. Aussi, si B/G diminue et le module d'Young augmente, le matériau se transformerait en moins ductile. Cela nécessite d'augmenter la rigidité et à diminuer la fragilité. Le biplot montré en (**Figure IV.3. 4**) révèle que le premier groupe des hydrures (LiH, NaH) est lié à B, G et E, indiquant leur grande dureté et rigidité, tandis que ceux du groupes 2 (KH, RbH, CsH) sont corrélés à B/G et  $\nu$ ; ce qui indique la faible ductilité et la haute fragilité de ces hydrures. Les résultats de l'analyse de données ACP sont en bon accord avec les résultats obtenus par les calculs ab initio et avec ceux déterminés théoriquement et expérimentalement.

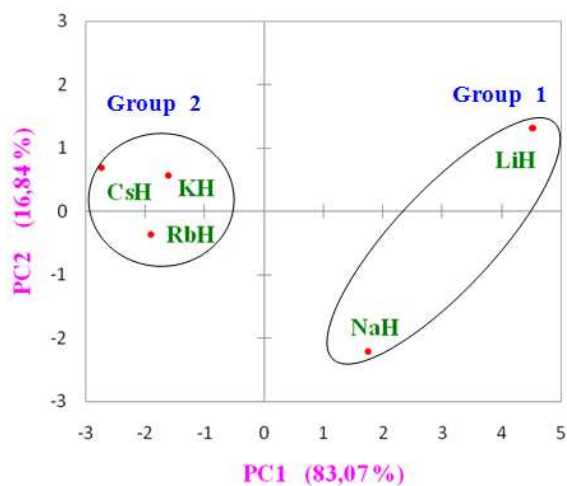


Figure IV.3.2. Le graphe des observations des hydrures alcalins

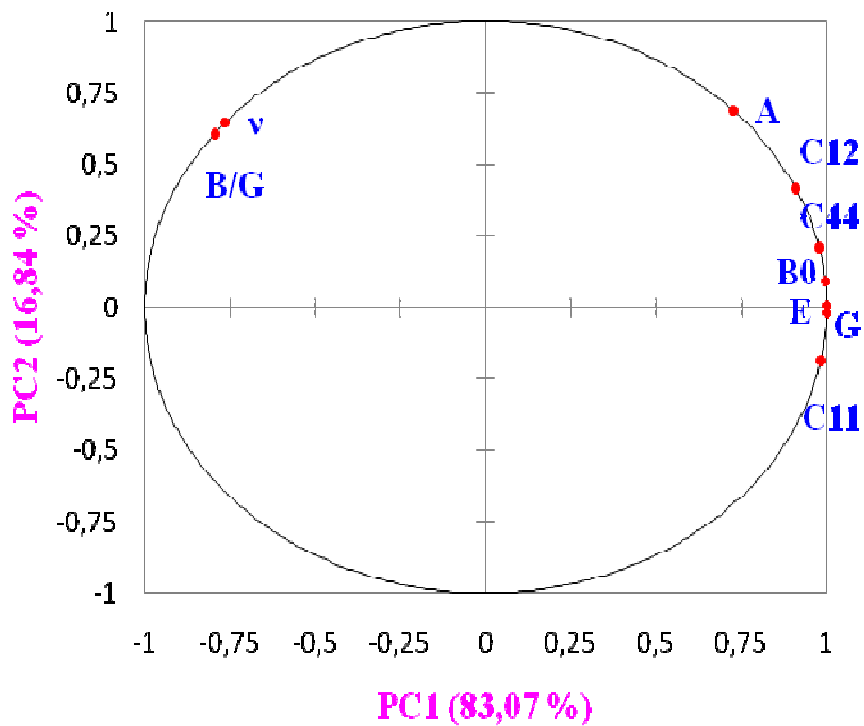


Figure IV.3.3. Le graphe des variables des hydrures alcalins.

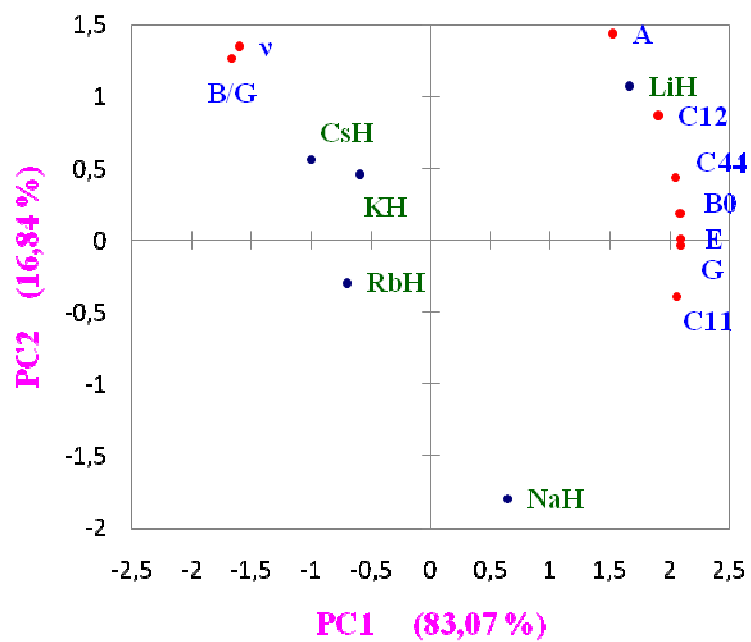


Figure IV.3.4. Le graphe de « Biplot » représentant les hydrures alcalins et les propriétés.

### IV.3.5. Conclusion

En résumé, les propriétés structurales et mécaniques des hydrures alcalins (LiH, NaH, KH, RbH et CsH) ont été étudiées par les calculs ab initio. Les calculs des énergies totales, des géométries optimisées et des constantes élastiques ont été effectués en utilisant le logiciel VASP, ainsi que la méthode PAW GGA. Le data mining est une nouvelle approche qui a été employée pour analyser les résultats obtenus, il est considéré comme l'une des approches fiables pour les problèmes d'optimisation, l'ACP est une des techniques de data mining largement utilisée pour l'analyse des données. Les propriétés calculées sont en bon accord avec les données expérimentales et théoriques disponibles. Les constantes élastiques calculées indiquent que les hydrures alcalins sont mécaniquement stables. L'analyse de G/B révèle que les hydrures alcalins sont des matériaux fragiles dans les conditions ambiantes. Compte tenu des valeurs du module de cisaillement (G), du module de compressibilité (B), et du module d'Young (E), il a été constaté que LiH est le plus rigide et le plus fragile parmi les hydrures étudiés. Les résultats obtenus montrent que les hydrures possèdent des propriétés intéressantes qui peuvent être très utiles dans certaines applications industrielles, en particulier LiH qui est un matériau prometteur.

Ce document présente l'aptitude des techniques d'exploration de données, telles que l'ACP, à analyser et à obtenir des informations précieuses à partir d'un ensemble de données. En outre, il peut être conclu que les résultats de l'analyse des données, sont en bon accord avec les résultats obtenus par les calculs ab initio et ceux rapportés dans la littérature.

**IV.3.6. Références**

- [1] L. George, R. Hrubciak, K. Rajan, S. K. Saxena, Principal component analysis on properties of binary and ternary hydrides and a comparison of metal versus metal hydride properties, *Journal of Alloys and Compounds* 478 (2009) 731–735.
- [2] L. George, S. K. Saxena, Structural stability of metal hydrides, alanates and borohydrides of alkali and alkali- earth elements: A review, *International Journal of Hydrogen Energy* 35 (2010) 5454-5470.
- [3] X.W. Sun, Q.F. Chen, X.R. Chen, L.C. Cai, F.Q. Jing, Ab initio study of phase transition and bulk modulus of NaH, *Journal of Solid State Chemistry* 184 (2011) 427–431.
- [4] E. Zintl, A. Harder, *Z. Phys. Chem. Abt. B* 14 (1931) 265.
- [5] X.W. Sun, L.C. Cai, Q.F. Chen, X.R. Chen, F.Q. Jing, Structural, thermodynamic, electronic, and optical properties of NaH from first-principles calculations, *Materials Chemistry and Physics* 133 (2012) 346– 355.
- [6] D. Colognesi, M. Zoppi, Hydrogen vibrational dynamics in ionic metal hydrides revealed through inelastic neutron scattering, *Notiziari Neutroni E Luce Disincrotrone*, 11 (2006) 4-13.
- [7] G. Kresse, J. Furthmuller, *Comput. Mater. Sci.* 6 (1996) 15.
- [8] <<http://cms.mpi.univie.ac.at/vasp>>.
- [9] P. Hohenberg, W. Kohn, *Physical Review* 136 (1964) B864–B871.
- [10] W. Kohn, L.J. Sham, *Physical Review* 140 (1965) A1133–A1138.
- [11] J. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, C. Fiolhais, *Phys. Rev. B* 46, 6671 (1992).
- [12] G. Kresse, D. Joubert, *Phys. Rev. B* 59 (1999) 1758.
- [13] S. Shang, A.J. Bottger, M.P. Steenvoorden, M.W.J. Craje, *Acta Mater.* 54 (2006) 2407–2417.
- [14] P.E. Blöchl, O. Jepsen, O.K. Andersen, *Physical Review B* 49 (1994) 16223–16233.
- [15] S.L. Shang, A. Saengdeejing, Z.G. Mei, D.E. Kim, H. Zhang, S. Ganeshan, Y. Wang, Z.K. Liu, First-principles calculations of pure elements: Equations of state and elastic stiffness constants, *Computational Materials Science* 48 (2010) 813–826.
- [16] H.J. Monkhorst, J.D. Pack, *Phys. Rev. B* 13 (1976) 5188–5192.
- [17] F.D. Murnaghan, *Proc. Natl. Acad. Sci. USA* Vol. 30, (1944) 244.

- [18] R. Miloua, Z. Kebbab, N. Benramdane, M. Khadraoui, F. Chiker, Ab initio prediction of elastic and thermal properties of cubic TiO<sub>2</sub>, *Computational Materials Science* 50 (2011) 2142–2147.
- [19] S. Boucetta, F. Zegrar, Density functional study of elastic, mechanical and thermodynamic properties of MgCu with a CsCl-type structure, *Journal of Magnesium and Alloys* 1 (2013) 128-133.
- [20] J.F. Nye, *Propriétés Physiques des Matériaux*, Dunod, Paris, 1961.
- [21] D.C. Wallace, *Thermodynamics of Crystals*, Wiley, New York, (1972).
- [22] S.F. Pugh, *Philos. Mag.* 45 (1954) 823.
- [23] P. Ravindran, L. Fast, P.A. Korzhavyi, B. Johansson, J. Wills, O. Eriksson, *J. Appl. Phys.* 84 (1998) 4891.
- [24] M.F. Yan, H.T. Chen, Structural, elastic and electronic properties of Cr<sub>2</sub>N: A first-principles study, *Computational Materials Science* 88 (2014) 81–85.
- [25] C. Zener, *Elasticity and Anelasticity of Metals*, University of Chicago Press, Chicago, 1948.

## CONCLUSION GÉNÉRALE

Dans ce travail on déduit comment la combinaison de deux domaines importants : la mécanique quantique et le data mining et ses méthodes (ACP et PLS), nous a permis d'étudier un phénomène qui reste insoluble jusque-là pour les physiciens et les chimistes, qui est les contradictions du tableau périodique avec la mécanique quantique. Le placement de certains éléments dans le tableau périodique a été problématique depuis le premier tableau de Dimitri Mendeleev en 1869.

Le tableau périodique des éléments, est la généralisation la plus importante en chimie, il est souvent considéré comme un cas représentatif particulier dans l'étude de la relation entre la chimie et la physique, et comme l'une des idées les plus unificatrices dans toute la science moderne.

La tentative d'expliquer les détails du tableau périodique continue à défier l'ingéniosité des physiciens quantiques et chimistes quantiques [2]. La mécanique quantique a la capacité d'expliquer et reproduire la chimie, ce qui a permis aux physiciens d'aller au-delà du tableau périodique des atomes neutres à l'état fondamental, et ce qui pourrait être interprété comme une déclaration obtenue d'une réduction de la chimie à la physique [1]. La classification périodique n'a pas encore sa forme définitive, et n'est pas entièrement cohérente avec la mécanique quantique, à cause des anomalies qu'elle contient.

L'originalité de ce travail persiste dans l'usage de l'approche data mining et de ses techniques, l'Analyse en composantes principales (ACP) et la régression (PLS), qui sont appliquées pour la première fois dans l'étude de ce problème, et qui nous ont permis de visualiser les corrélations existantes dans chaque base de données, et de voir de près le problème.

Par conséquent, les résultats obtenus montrent que le lutétium partage beaucoup de similitudes avec les métaux de transition de la troisième période du tableau périodique, par contre l'hydrogène partage certaines propriétés avec certains groupes du tableau périodique, tel que le groupe du carbone, mais pas tout à fait; à vrai dire l'hydrogène est un élément à part, et n'est conforme à aucun groupe.

Enfin on a calculé les propriétés structurales et mécaniques des hydrures alcalins par les calculs ab initio implémentées dans le code VASP, dans le cadre de la Théorie de la Fonctionnelle de la Densité DFT, les résultats obtenus ont été analysés en utilisant l'approche

datamining, et la technique d'Analyse en Composantes Principales (ACP) pour explorer les similitudes et les corrélations existantes entre ce type d'hydrures. On a trouvé que les hydrures alcalins sont des matériaux fragiles dans les conditions ambiantes, et que LiH est le plus rigide et le plus fragile parmi ces hydrures, LiH est un matériau prometteur en industrie des batteries rechargeables, et dans le stockage d'énergie.

Les résultats obtenus sont très intéressants, du point de vue scientifique, ils ont été approuvés par de grands chercheurs travaillant sur ce problème. Cela prouve l'importance de l'approche data mining et de ses techniques, l'Analyse en composantes principales (ACP) et la régression (PLS) dans l'analyse et traitement de données.

Ce travail ouvre des horizons à l'usage de l'approche data mining et les techniques d'analyse multivariée pour explorer et étudier des problèmes qui semblent sans issue.

**Références**

- [1] V. N. Ostrovsky, « Physical Explanation of the Periodic Table », Ann. N. Y. Acad. Sci. **988** (2003) 182-192.
- [2] Eric R. Scerri, « Collected papers on philosophy of chemistry », Imperial College Press, (2008). ISBN-13 978- 1-84816- 137-5.

La découverte de la structure de l'atome au début du vingtième siècle a marqué une grande étape dans l'avancée des connaissances. De nombreux phénomènes physiques et chimiques demeuraient sans interprétation tels que l'effet Compton, l'effet Zeeman,...*etc.* Ce qui a conduit les physiciens et chimistes à abandonner le modèle classique de la mécanique newtonienne et à introduire un nouveau modèle qui est la mécanique quantique, ce qui a permis une interprétation nouvelle de la classification périodique [1].

La mécanique quantique a été la théorie la plus spectaculaire réussie dans l'histoire de la science. L'impact de la théorie a été ressenti dans des domaines tels que la physique du solide, et la science des matériaux [2].

Bien qu'il soit évident que les tentatives pour expliquer la chimie en termes de la théorie quantique ne peuvent apparaître suite au développement de la mécanique quantique

au XXe siècle, ce type de réductionnisme ne date pas d'aujourd'hui. En général, le réductionnisme est l'idée qu'une science, telle que la chimie, peut être réduite plus à une science "fondamentale", dans ce cas, à la physique. la conception même du réductionnisme supposant une hiérarchie dans les sciences entre les sciences plus fondamentales, généralement la physique et les mathématiques, et ceux qui sont considérées comme des dérivés ou des épiphénomènes, à commencer par la chimie, et passant par la biologie aux sciences humaines comme la sociologie ou la psychologie [9].

On objectera aisément à ce qui précède que même si la réduction n'existe pas pour le moment elle peut jouer le rôle d'un idéal, d'un but à atteindre. N'a-t-elle pas stimulé le développement de la chimie et de la physique au 19<sup>e</sup> siècle ? N'inspire-t-elle pas aujourd'hui la biologie? Examinons cela. L'idée d'une réduction verticale qui, de l'antiquité aux lumières, a été une thèse philosophique (l'atomisme) s'incarne au 19<sup>e</sup> siècle dans l'hypothèse atomique des chimistes, puis dans la théorie cinétique des gaz, dans la mécanique statistique, et finalement dans la "réalité moléculaire" (selon l'expression de Jean Perrin). La théorie des quanta puis la mécanique quantique donnent à l'atomisme des physiciens et des chimistes une justification plus profonde ; elles expliquent que, comme le disait Bohr, il existe dans la nature une tendance à produire des formes déterminées et à les faire réapparaître, encore et toujours, même lorsqu'elles ont été perturbées ou détruites. Mais, en triomphant, l'atomisme a perdu

sa simplicité originelle. Les atomes ne sont plus insécables et inaltérables. L'existence des molécules n'est plus un postulat de base de la chimie, mais un résultat de la théorie quantique, valable sous certaines réserves. La postérité de l'atomisme s'est scindée en deux branches bien distinctes : d'un côté la théorie savante, et les techniques récentes de manipulation d'atomes ou de molécules presque isolées ; de l'autre, l'idée de réduction verticale qui n'est plus une thèse philosophique ou une hypothèse scientifique, mais est devenue une croyance.

Imaginons que la chimie, la physique du solide, la thermodynamique, la mécanique classique enfin, soient toutes remplacées par la mécanique quantique, devenue entièrement calculable après l'avènement d'ordinateurs d'un type nouveau. L'industrie informatique réaliserait des chiffres d'affaires fabuleux. Mais l'idéal d'intelligibilité qui a présidé à la naissance de la science serait mort et enterré, au moins dans les sciences physiques ; car la mécanique quantique, qui calcule par ses propres moyens et fournit des résultats d'une exactitude numérique étonnante, ne peut guère se passer des béquilles de la physique classique quand on lui demande de nous faire comprendre ce qui se passe. Par suite, les sciences physiques, débarrassées désormais de toute incertitude et de toute nouveauté véritable, sombreraient dans un océan d'ennui. En somme, même si la réduction devenait possible dans les sciences physiques, elle ne les aiderait pas à remplir ce qui doit rester la tâche fondamentale de toute science : nous aider à comprendre les phénomènes [10]. La mécanique quantique a montré à quel point la séparation entre la physique et la chimie était arbitraire [11].

Friedrich Paneth a clairement défini le statut épistémologique du concept chimique de l'élément dans une conférence portant ce titre qu'il a donné en 1931. Ici, il a soulevé deux questions. La première rappelle le problème d'Aristote avec « les mixtes véritables », en quel sens peut-on supposer que les éléments persistent dans les composés? " Sa réponse - basée sur une étude historique des notions d'éléments et des atomes - est claire: puisque la chimie est concernée par les qualités secondaires des substances, les chimistes devraient assumer la persistance des qualités dans les composés. Il a proposé d'appeler ce qui persiste " substance de base " alors qu'il réservait le terme " substance simple " pour la matière montrant des manifestations phénoménologiques associées à cette substance de base abstraite, renouvelant de ce fait la distinction conceptuelle de Mendeleev. La deuxième question de Paneth était: « s'il est vrai ou non que la chimie doit et va se dissoudre dans la physique? " Ici, il a

traité la possibilité de réduire la chimie à la physique, un développement qui avait été déjà prédit par certains physiciens.

La théorie quantique offre clairement des avantages considérables pour l'enseignement de la chimie et, en particulier, pour la présentation du tableau périodique, avec son explication du comment et du pourquoi les propriétés chimiques des divers éléments obéissent à un modèle global. Néanmoins, ceci ne signifie pas que les chimistes peuvent s'en sortir seuls avec la mécanique quantique. Le chimiste ne peut pas s'isoler de ces développements et de se contenter d'exécuter et de décrire les réactions. Il ne pourra pas tirer parti de ses succès précédents à moins qu'il assimile les disciplines du physicien ou accepte de collaborer avec lui. Cette évolution de la chimie vers la physico-chimie est le trait le plus caractéristique de ces vingt dernières années [9].

Le tableau périodique se classe comme l'une des idées les plus fécondes et unificatrices dans toute la science moderne [12]. L'élément chimique est proche d'une abstraction mathématique. Ainsi, il devient un instrument qui nous permet de construire une série qui est supposée se cacher derrière. Cette abstraction permet au chimiste de décrire l'ordre qui existe derrière la multiplicité apparente chaotique du monde phénoménal. La plus grande valeur du concept de Mendeleev d'un élément est sa puissance de construire la série apparentée. Les éléments chimiques sont efficacement "désincarnés" les entités abstraites qui sont attachées au vrai monde par le réseau des rapports décrits par le système périodique. Ce rapport abstrait-concret est représenté par la position de l'élément dans le tableau périodique, qui forme alternativement la base pour une multiplicité d'interprétation [13].

À mesure qu'augmentait le nombre d'éléments connus, les chimistes du XIX<sup>e</sup> siècle éprouvaient le besoin de les ordonner. On connaissait les masses atomiques et on avait déjà l'habitude de classer les éléments d'après l'augmentation de cette caractéristique. La simple progression des masses atomiques ne pouvait toutefois expliquer logiquement les différents comportements des éléments [14].

Il est en effet une sorte de miracle que la mécanique quantique explique le tableau périodique. Mais nous ne devrions pas se laisser séduire en croyant que c'est une explication déductive. Une chose qui est claire: est que la tentative d'expliquer les détails de la classification périodique continue à défier l'ingéniosité des physiciens quantiques et chimistes quantiques et que le tableau périodique continuera de

présenter un test concernant la pertinence des nouvelles méthodes développées en chimie quantique. La réduction de la chimie à la mécanique quantique n'a ni complètement échoué, comme certains philosophes de la science ont revendiqué, et elle n'a pas été une réussite complète, comme certains historiens contemporains ont revendiquée [12].

Le tableau périodique se classe comme l'une des idées les plus fécondes et unificatrices dans toute la science moderne [9]. L'élément chimique est proche d'une abstraction mathématique. Ainsi, il devient un instrument qui nous permet de construire une série qui est supposée se cacher derrière. Cette abstraction permet au chimiste de decrier l'ordre qui existe derrière la multiplicité apparente chaotique du monde phénoménal. La plus grande valeur du concept de Mendeleev d'un élément est sa puissance de construire la série apparentée. Les éléments chimiques sont efficacement "désincarnés" les entités abstraites qui sont attachées au vrai monde par le réseau des rapports décrits par le système périodique. Ce rapport abstrait-concret est représenté par la position de l'élément dans le tableau périodique, qui forme alternativement la base pour une multiplicité d'interprétation [10].

À mesure qu'augmentait le nombre d'éléments connus, les chimistes du XIX<sup>e</sup> siècle éprouvaient le besoin de les ordonner. On connaissait les masses atomiques et on avait déjà l'habitude de classer les éléments d'après l'augmentation de cette caractéristique. La simple progression des masses atomiques ne pouvait toutefois expliquer logiquement les différents comportements des éléments [12].

Il est en effet une sorte de miracle que la mécanique quantique explique le tableau périodique. Mais nous ne devrions pas se laisser séduire en croyant que c'est une explication déductive. Une chose qui est claire: est que la tentative d'expliquer les détails de la classification périodique continue à défier l'ingéniosité des physiciens quantiques et chimistes quantiques et que le tableau périodique continuera de présenter un test concernant la pertinence des nouvelles méthodes développées en chimie quantique. La réduction de la chimie à la mécanique quantique n'a ni complètement échoué, comme certains philosophes de la science ont revendiqué, et elle n'a pas été une réussite complète, comme certains historiens contemporains ont revendiquée [9].

L'évolution de l'histoire de la chimie au XIXe siècle nous montre à quel point une classification devenait nécessaire. La classification telle que Mendeleiev l'a publiée en 1869 puis en 1871 a stimulé la recherche en chimie et de nombreuses questions fondamentales se sont appuyées sur cette formidable base théorique que renferme ce tableau. C'est actuellement un point de repère universel pour la communauté des chimistes [3].

Le tableau périodique reste, après plus d'un siècle, le plus important moyen de comparaison entre les éléments chimiques. Il nous permet d'étudier rationnellement la grande variété des substances que nous trouvons dans la nature [4].

La matière condensée est formée par l'agrégation d'atomes, et les propriétés de l'agrégat dépendent de la structure électronique des atomes qui le constituent. Elle est donc d'une grande importance, ce qui est essentiel pour la classification des propriétés physiques des éléments solides comme il l'est pour ses atomes constitutifs [5].

La classification périodique des éléments exprime une profonde vérité physique et permet d'unifier un grand nombre de connaissances portant sur la matière. Ce classement facilite la compréhension des groupes d'éléments en montrant la relation entre leurs propriétés chimiques et leur structure atomique. De plus, la connaissance du tableau périodique permet de prévoir les formules des composés et les types de liaisons unissant les composants d'une molécule [4]. Les éléments dans le tableau périodique sont rangés par ordre croissant du numéro atomique  $Z$ , selon la configuration électronique des atomes [6], en respectant la règle de Klechkowski dans le remplissage des orbitales par énergie croissante.

Le tableau périodique est pour nous, de toute évidence, le guide le plus sûr dans l'étude de la chimie. Mais il comporte certaines anomalies ou contradictions, qui sont les éléments qui ne respectent pas la règle de Klechkowski, et qui présentent des configurations électroniques faisant exception à la norme, tels que :

- L'hydrogène qui ne s'ajuste harmonieusement à aucun groupe, c'est un non-métal est il est placé généralement dans la même colonne que le lithium qui est un métal très réactif. L'hydrogène est parfois placé en tête du groupe 7, avec les halogènes. Dans certains tableaux périodiques, il est même placé dans les deux groupes à la fois [7].

- L'Hélium se trouve dans le bloc p alors qu'il est de type s.
- Le lutécium et le lawrencium sont placés dans le bloc f, alors que leur couche f est déjà saturée, ils contiennent 1 électron 5d ou 6d. Ils constituent bien le début des éléments de transition.

Les tableaux périodiques habituels de Mendeleïev des éléments chimiques sont déjà à 97% en accord avec la mécanique quantique. Trois éléments seulement qui ne le sont pas, et sont en désaccord avec le principe d'exclusion de Pauli. Afin d'assurer la cohérence, il est proposé de placer l'hélium à côté de l'hydrogène dans le bloc s. Le lutécium et le lawrencium dans le bloc *d* des métaux de transition, pour que le tableau périodique soit conforme à 100% avec la mécanique quantique [8].

Dans cette thèse on a étudié les contradictions existantes dans le tableau périodique, et on les a confronté moyennant l'approche du data mining et ses techniques, telles que l'analyse en composantes principales (ACP) et la régression (PLS), qui sont employées pour la première fois pour l'étude de ce problème, et ce qui fait l'originalité de ce travail.

Le data mining, dans sa forme et compréhension actuelle, à la fois comme champ scientifique et industriel, est apparu au début des années 90. On peut voir le data mining comme une nécessité imposée par le besoin des entreprises de valoriser les données qu'elles accumulent dans leurs bases. En effet, le développement des capacités de stockage et les vitesses de transmission des réseaux ont conduit les utilisateurs à accumuler de plus en plus de données. Certains experts estiment que le volume des données double tous les ans. Que doit-on faire avec des données coûteuses à collecter et à conserver? [15].

L'ACP est une technique mathématique qui réduit la dimensionnalité d'un système complexe de corrélations à un nombre réduit de dimensions, néanmoins représentatives de la majorité de la variance totale des données. On passe d'un certain nombre de variables potentiellement corrélées à un plus petit nombre de variables décorréées, les "Composantes Principales".

Ensuite, un changement de base est effectué à la nouvelle base des "Composantes Principales" afin de représenter les données dans le nouvel espace. L'utilisateur

choisit le nombre d'axes utilisés pour la représentation, en se basant par exemple sur le pourcentage de variance représenté.

L'exploration de données est le processus de découvrir les modèles pertinents, intéressants et novatrices, ainsi que des modèles descriptifs, compréhensibles, et prédictifs de données à grande échelle [16].

La régression PLS permet de rapprocher le statisticien du chercheur dans les problèmes de modélisation. En général ce dernier souhaite en effet conserver dans son modèle toutes les variables importantes tout en obtenant des équations de régression cohérentes. [17].

Bien que cette méthode ait connu la plupart de ses succès dans le domaine de la chimie, on peut penser qu'elle pourrait facilement être transposée à d'autres domaines, particulièrement dans ceux où le nombre d'individus actifs est faible (en comparaison au nombre de variables explicatives) et où les variables explicatives sont significativement corrélées entre-elles. [18]

De nombreux problèmes peuvent être décrits sous la forme d'un modèle de régression, où l'on possède des variables  $X$  sur lesquelles on peut plus ou moins agir et des variables  $Y$  que l'on ne peut qu'observer.

L'objectif est alors de d'écrire les relations entre  $Y$  et  $X$ , en l'absence de modèle théorique. Le problème est que le nombre de variables  $X$  est souvent très important par rapport au nombre d'observations. La régression PLS est une méthode d'analyse des données spécifiquement construite pour l'étude de ce type de problème. [19]

Cette thèse est organisée comme suit:

- Le premier chapitre est consacré à la présentation générale du réductionnisme en chimie quantique, ses différentes formes, ses aspects, les types de réduction, la comparaison entre la physique et la chimie, la réduction de la chimie à la mécanique quantique, et le lien existant entre le tableau périodique et le réductionnisme en chimie quantique.

- Le deuxième chapitre est consacré à l'étude du tableau périodique, son historique, la loi de la périodicité des propriétés, son évolution, et ses anomalies.

- Le troisième chapitre est consacré à la définition du datamining et ses différentes techniques, l'analyse en composantes principales (ACP) et la régression PLS (la régression par les moindres carrés partiels), le principe de ces méthodes, et leurs objectifs.

- Le quatrième chapitre présente les résultats liés à l'application de l'approche du datamining, et ses techniques telles que : l'analyse en composantes principales (ACP) et la régression (PLS), qui ont été appliqués sur les différentes bases de données contenant les différentes familles du tableau périodique qui présentent des contradictions avec la mécanique quantique. Ce chapitre se divise en trois parties:

Dans la première partie nous avons réalisé une étude comparative des propriétés physiques et mécaniques du lutétium avec ceux des métaux de transition par l'approche de datamining.

Tandis que dans la deuxième partie, on a étudié les propriétés physiques et mécaniques de l'hydrogène, et on les a comparées avec ceux du groupe des métaux alcalins, avec les éléments du groupe de carbone, et avec ceux du groupe des halogènes, en utilisant l'approche de data mining.

Enfin dans la troisième partie on a étudié les propriétés structurales et mécaniques des hydrures alcalins en appliquant les calculs de premier principe et qu'on a analysé par l'approche de datamining.

Dans ce travail on déduit comment la combinaison de deux domaines importants : la mécanique quantique et le data mining et ses méthodes (ACP et PLS), nous a permis d'étudier un phénomène qui reste insoluble jusque-là pour les physiciens et les chimistes, qui est les contradictions du tableau périodique avec la mécanique quantique. Le placement de certains éléments dans le tableau périodique a été problématique depuis le premier tableau de Dimitri Mendeleev en 1869.

Le tableau périodique des éléments, est la généralisation la plus importante en chimie, il est souvent considéré comme un cas représentatif particulier dans l'étude de

la relation entre la chimie et la physique, et comme l'une des idées les plus unificatrices dans toute la science moderne.

La tentative d'expliquer les détails du tableau périodique continue à défier l'ingéniosité des physiciens quantiques et chimistes quantiques [21]. La mécanique quantique a la capacité d'expliquer et reproduire la chimie, ce qui a permis aux physiciens d'aller au-delà du tableau périodique des atomes neutres à l'état fondamental, et ce qui pourrait être interprétée comme une déclaration obtenue d'une réduction de la chimie à la physique [20]. La classification périodique n'a pas encore sa forme définitive, et n'est pas entièrement cohérente avec la mécanique quantique, à cause des anomalies qu'elle contient.

L'originalité de ce travail persiste dans l'usage de l'approche du data mining et de ses techniques, l'Analyse en composantes principales (ACP) et la régression (PLS), qui sont appliquées pour la première fois dans l'étude de ce problème, et qui nous ont permis de visualiser les corrélations existantes dans chaque base de données, et de voir de près le problème.

Par conséquent, les résultats obtenus montrent que le lutétium partage beaucoup de similitudes avec les métaux de transition de la troisième période du tableau périodique, par contre l'hydrogène partage certaines propriétés avec certains groupes du tableau périodique, tels que le groupe du carbone, mais pas tout à fait; à vrai dire l'hydrogène est un élément à part, et n'est conforme à aucun groupe.

Enfin on a calculé les propriétés structurales et mécaniques des hydrures alcalins par les calculs ab initio implémentées dans le code VASP, dans le cadre de la Théorie de la Fonctionnelle de la Densité DFT, les résultats obtenus ont été analysés en utilisant l'approche de datamining, et la technique d'Analyse en Composantes Principales (ACP) pour explorer les similitudes et les corrélations existantes entre ce type d'hydrures. On a trouvé que les hydrures alcalins sont des matériaux fragiles dans les conditions ambiantes, et que LiH est le plus rigide et le plus fragile parmi ces hydrures, LiH est un matériau prometteur en industrie des batteries rechargeables, et dans le stockage d'énergie.

Les résultats obtenus sont très intéressants, du point de vue scientifique, ils ont été approuvés par de grands chercheurs travaillant sur ce problème. Cela prouve l'importance de l'approche du data mining et de ses techniques, l'Analyse en composantes principales (ACP) et la régression (PLS) dans l'analyse et traitement de données.

Ce travail ouvre des horizons à l'usage de l'approche de data mining et les techniques d'analyse multivariée pour explorer et étudier des problèmes qui semblent sans issue.

## **Références**

- [1] André Durupthy, Jacques Estienne, Magali Giacino, Alain Jaubert, Claude Mesnil, « Physique, Chimie et sciences de l'ingénieur » (PCSI), I.S.B.N. 978-2-0118-1752-5. [www.hachette-education.com](http://www.hachette-education.com).
- [2] Eric R. Scerri, « Collected papers on philosophy of chemistry », Imperial College Press, (2008). ISBN-13 978-1-84816-137-5.
- [3] Jean-François Le Maréchal, « La classification périodique des éléments chimiques », Université Lyon 2 - UMR GRIC équipe COAST (1999).
- [4] Le repaire des sciences, <http://www.lerepairedessciences.fr>
- [5] H.P. Myers, « Introductory Solid State Physics », 2<sup>nd</sup> ed, Taylor & Francis e-Library, (2009), ISBN 0-203-21255-X.
- [6] <http://www.web-sciences.com>.
- [7] Carole McQuarrie, Donald A. McQuarrie, Peter A. Rock, « Chimie générale », Editions De Boeck Université 1<sup>st</sup> ed., (1992).
- [8] Bernard SCHAEFFER, Periodic table quantum mechanics consistent, <https://www.scribd.com/doc/5994795/Periodic-table-quantum-mechanics-consistent>.
- [9] Bernadette Bensaude-Vincent, Jonathan Simon, « CHEMISTRY THE IMPURE SCIENCE », Imperial College Press, (2008), ISBN-13 978-1-84816-225-9.
- [10] À René Thom, « Passion des formes : dynamique qualitative, sémiophysique et intelligibilité », Vol 1, ENS Editions, (1994) 828 p.
- [11] Champ Vallon, « La totalité », vol 5, (2002), Ed. Champ Vallon, Presses Universitaires de France, ISBN 2-87673-355-2.
- [12] Eric R. Scerri, « The periodic table : its story and its significance », Published by Oxford University Press, Inc, (2007).
- [13] Bernadette Bensaude-Vincent, Jonathan Simon, « Chemistry the impure science », Imperial College Press, (2008), ISBN-13 978-1-84816-225-9.
- [14] Le repaire des sciences, <http://www.lerepairedessciences.fr>
- [15] Djamel Abdelkader ZIGHED, Ricco RAKOTOMALALA, « Extraction de connaissances à partir de données (ECD) - Objet du data mining », Techniques de l'Ingénieur, (2002).

- [16] Mohammed J. Zaki and Wagner Meira Jr., « Data mining and analysis, Fundamental Concepts and Algorithms », Cambridge University Press, 1st ed., (2014).
- [17] Tenenhaus, J.-P. Gauchi, C. Ménardo, « Régression PLS et applications », Revue de statistique appliquée, tome 43, n°1, (1995), P.7-63.
- [18] Site internet : <http://www.renaudloup.be/refinternet.htm> .
- [19] Tenenhaus M., « *La Régression PLS: Théorie et Pratique* », Editions Technip, (2002).
- [20] V. N. Ostrovsky, « Physical Explanation of the Periodic Table », Ann. N. Y. Acad. Sci. **988** (2003) 182-192.
- [21] Eric R. Scerri, « Collected papers on philosophy of chemistry », Imperial College Press, (2008). ISBN-13 978- 1-84816- 137-5.

# A Study of the Physical and Mechanical Properties of Lutetium Compared with Those of Transition Metals: A Data Mining Approach

NADERA SETTOUTI<sup>1,2</sup> and HAFID AOURAG<sup>1</sup>

1.—LEPM, URMER, Department of Physics, University Abou Bakr Belkaid, 13000 Tlemcen, Algeria. 2.—e-mail: nadera\_settouti@yahoo.fr

In this article, we study the physical and mechanical properties of lutetium, which will be compared with the elements of the third-row transition metals (Cs, Ba, Hf, Ta, W, Re, Os, Ir, Pt, Au, Tl, Pb, and Bi). Data mining is an ideal approach for analyzing the information and exploring the hidden knowledge among the data. The purpose of the data mining scheme is to identify and classify the effects of the relationships existing between properties. The results of the investigation are presented by means of multivariate modeling methods, such as the principal component analysis and the partial least squares regression to discover the implicit, yet meaningful, relationship between the elements of the data set, and to locate correlations between the properties of the materials. In this study, we present a data mining approach to discover such unusual correlations between properties of the elements. When comparing the properties of the transition metals with those of lutetium, our results show that lutetium shares many properties and similarities with the transition metals of the sixth row in the periodic table and can be well described as a transition metal.

## INTRODUCTION

The classification of the chemical elements is a precise and clear result of the periodic table, despite the few anomalies that still persist. The periodic table is a physical representation of two laws: the periodic law and the periodic system, which are more fundamental than the periodic table.<sup>1</sup>

The periodic table of elements reflects a distinguished example of classification; each column contains a series of elements that have the same number of valence electrons. Therefore, the study of the arrangement of the elements in different groups, according to their principal common properties, is scientifically and practically interesting and should be carried out.<sup>2</sup>

The transition elements, also called transition metals, are not very different from each other. They have similar properties and behave similarly.<sup>3</sup> Their *d* subshell is partially filled of electrons, and this represents the difference between a transition metal and the other elements in the periodic table. The transition metals are arranged in four rows

that correspond to the principal energy levels of *d* subshells that are progressively filled with electrons.<sup>4</sup> The main purpose of this work is to demonstrate that lutetium Lu ( $Z = 71$ ) has common properties with the third row of the *d*-block transition metals.

Many researchers have considered the position of lutetium in the periodic table before, by different ways. In 1967, H. Merz and K. Ulmer carried out an x-ray spectroscopic investigation showing, distinctly, that Lu, to the left of Hf (below Yttrium), fits much better with its properties than La does, and so it may be considered as a transition metal but not La.<sup>5</sup> Jensen (1982) assigned Lu to Group III, below Y.<sup>6</sup> Scerri (2009) believed that lutetium should be placed in the same column of the periodic table as scandium, yttrium, and lawrencium.<sup>7</sup>

In this work, we use a data mining approach to expose the correlations between the data. This approach needs powerful statistical analysis tools such as principal component analysis (PCA) and partial least squares regression (PLS) to study the correlations between the properties and the

relationships between elements. PCA is probably the most popular multivariate statistical technique used in many scientific fields, as it can be used for data classification, investigation of similarities, etc.<sup>8,9</sup> PCA is a powerful statistical approach for the analysis of material properties; it has been widely used to study many problems in physics and materials science.<sup>10</sup> There is a close connection between PLS and PCA, as they have some interesting information in common.<sup>11</sup>

## DATA AND COMPUTATIONAL DETAILS

### Data Description

To identify the trends and clusters in materials properties data, we build a database for 14 elements of the third-row transition elements (Cs, Ba, Hf, Ta, W, Re, Os, Ir, Pt, Au, Tl, Pb, and Bi), including lutetium, with their physical and mechanical properties, as seen in (Table I).

The properties and their abbreviations are as follows: polarizability ( $\alpha$ ), electronegativity ( $\chi$ ), atomic weight ( $m$ ), atomic number ( $Z$ ), atomic radius ( $r$ ), atomic volume ( $V$ ), electron affinity ( $E_A$ ), and ionization energy ( $I$ ), as well as total energy ( $E_{tot}$ ), cohesive energy ( $E_{coh}$ ), bulk modulus ( $B$ ), shear modulus ( $G$ ), Young's modulus ( $E$ ), Poisson number ( $\mu$ ), ratio ( $B/G$ ), elastic stiffness constants (C11, C12, and C44), and Cauchy pressure (C12–C44). We have included only these properties in the current study due to the unavailability of enough data on other properties.

### Data Mining Methods

The data mining tasks can be classified into two groups: descriptive and predictive. This can be accomplished using different techniques based on data mining functionalities such as clustering, classification, and so on, to determine the types of patterns to be mined. In this work, we use the descriptive data mining method, which entails exploration of patterns and relationships existing among data. The fundamental descriptive functions are summarization, clustering, association, and discovery.<sup>12</sup> PCA and PLS are appropriate techniques to carry out this data mining task.

#### *Principal Component Analysis Technique*

This was achieved by running PCA using the numerical data structured in 14 observations versus 19 variables ( $\alpha$ ,  $\chi$ ,  $m$ ,  $Z$ ,  $r$ ,  $V$ ,  $E_A$ ,  $I$ ,  $E_{tot}$ ,  $E_{coh}$ ,  $B$ ,  $G$ ,  $E$ ,  $\mu$ ,  $B/G$ , C11, C12, C44, and C12–C44). For a visualization of the data discrimination, PCA plots mapped variables (19 properties) and observations ( $n = 14$ ) using loadings and scores in dimensional spaces determined by principal components (PCs) with eigenvalues  $> 1.0$ , based on Kaiser's rule. The loading plot depicts the most important variables, and the score plot shows similarities and differences between samples (i.e., the elements in our data set).<sup>13</sup> PCA generates a set of new orthogonal axes

or variables, called the PCs, from the original variables. The generated axes are orthogonal and uncorrelated with one another. The first PC (PC1) explains the maximal variance in the data set as well as its direction, and so on. The PCs are defined by vectors that are the eigenvectors of the variance–covariance matrix. The variance along the vector is defined as the eigenvalue. The loadings and the scores corresponding to the PCs are obtained from the correlation matrix. Each loading of variables reveals the contribution of the original variable to the PC. A variance reduction is reached when neglecting the directions along which the variance of samples is insignificant.<sup>13</sup>

The key idea of PCA technique is to reduce the dimensionality of a data set that contains a large number of interrelated variables by retaining as much as possible of the variation encountered in the data set. This reduction is achieved by transforming the original variables into a smaller new set of variables called the principal components, which are uncorrelated and ordered in such a way that the first principal component captures most of the variation present in all the original variables. This makes visualization, prediction, and classification easier.<sup>14,15</sup> PCA summarizes the information provided by various variables in a simple graphical display with minimal loss of information to assess the relationships between variables and samples.<sup>16</sup> PCA identifies patterns in a data set and compresses the data by reducing the number of dimensions based on their similarities and differences.<sup>11</sup> When a physical significance to principal components is found, the PCA technique can discover the hidden data structure and reveal similarities among variables and elements.<sup>9</sup>

#### *Partial Least Squares Regression*

Although PCA is useful in estimating the relative impact of multiple parameters on properties, PLS regression can be used as an exploratory analysis tool to select suitable predictor variables and identify outliers before the classic linear regression. Using a combination of PCA and PLS techniques helps explore a broader range of new trends and correlations, which have not been established before.<sup>14</sup> PLS comprises a wide category of methods for modeling relations between sets of observed variables by using the latent variables. It includes regression, classification tasks, dimension reduction techniques, and modeling tools. The main idea of PLS methods is that the observed data are generated by a system that is driven by a small number of latent (not directly observed or measured) variables. The PLS algorithm becomes a successful standard tool for processing wide chemical data problems; it facilitates many applications in other scientific fields. Generally, PLS creates orthogonal score vectors (called latent vectors or components) by finding the latent vectors that contain the maximal covariance between different

Table I. Physical and mechanical properties of the third-row transition elements including lutetium

| Property              | Cs               | Ba               | Lu               | Hf               | Ta               | W                | Re               |
|-----------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| $\alpha$              | 59.42 (Ref. 27)  | 39.7 (Ref. 27)   | 21.9 (Ref. 27)   | 16.2 (Ref. 27)   | 13.1 (Ref. 27)   | 11.1 (Ref. 27)   | 9.7 (Ref. 27)    |
| $\chi$                | 0.86 (Ref. 28)   | 0.97 (Ref. 28)   | 1.14 (Ref. 28)   | 1.23 (Ref. 28)   | 1.33 (Ref. 28)   | 1.40 (Ref. 28)   | 1.46 (Ref. 28)   |
| $m$                   | 132.91 (Ref. 29) | 137.33 (Ref. 29) | 174.97 (Ref. 29) | 178.49 (Ref. 29) | 180.95 (Ref. 29) | 183.84 (Ref. 29) | 186.21 (Ref. 29) |
| $Z$                   | 55               | 56               | 71               | 72               | 73               | 74               | 75               |
| $r$                   | 2.60 (Ref. 30)   | 2.15 (Ref. 30)   | 1.734 (Ref. 31)  | 1.55 (Ref. 30)   | 1.45 (Ref. 30)   | 1.35 (Ref. 30)   | 1.35 (Ref. 30)   |
| $V$                   | 115.89 (Ref. 32) | 62.47 (Ref. 32)  | 29.53 (Ref. 33)  | 22.41 (Ref. 32)  | 18.31 (Ref. 32)  | 16.23 (Ref. 32)  | 14.99 (Ref. 32)  |
| $E_A$                 | 0.47 (Ref. 27)   | 0.15 (Ref. 28)   | 0.34 (Ref. 27)   | 0.1 (Ref. 34)    | 0.32 (Ref. 28)   | 0.82 (Ref. 28)   | 0.15 (Ref. 28)   |
| $I$                   | 3.89 (Ref. 29)   | 5.21 (Ref. 29)   | 5.43 (Ref. 29)   | 6.83 (Ref. 29)   | 7.55 (Ref. 29)   | 7.86 (Ref. 29)   | 7.83 (Ref. 29)   |
| $E_{\text{tot}}^{35}$ | -7550.56         | -7880.11         | -13848.23        | -14317.49        | -14795.88        | -15283.45        | -15780.24        |
| $E_{\text{coh}}$      | 0.80 (Ref. 36)   | 1.90 (Ref. 36)   | 4.43 (Ref. 36)   | 6.44 (Ref. 36)   | 8.10 (Ref. 36)   | 8.90 (Ref. 36)   | 8.03 (Ref. 36)   |
| $B$                   | 2.03 (Ref. 37)   | 10.3 (Ref. 37)   | 47.6 (Ref. 27)   | 108.9 (Ref. 37)  | 200.1 (Ref. 37)  | 323.2 (Ref. 37)  | 371.7 (Ref. 37)  |
| $G$                   | 0.65 (Ref. 28)   | 4.84 (Ref. 28)   | 27.1 (Ref. 28)   | 53.0 (Ref. 28)   | 64.7 (Ref. 28)   | 152 (Ref. 28)    | 180 (Ref. 28)    |
| $E$                   | 1.69 (Ref. 28)   | 12.8 (Ref. 28)   | 68.4 (Ref. 28)   | 138 (Ref. 28)    | 185 (Ref. 28)    | 407 (Ref. 28)    | 520 (Ref. 28)    |
| $\mu$                 | 0.30 (Ref. 28)   | 0.28 (Ref. 28)   | 0.27 (Ref. 28)   | 0.29 (Ref. 28)   | 0.35 (Ref. 28)   | 0.28 (Ref. 28)   | 0.26 (Ref. 28)   |
| B/G                   | 3.12             | 2.13             | 1.76             | 2.05             | 3.09             | 2.13             | 2.07             |
| C11                   | 1.60 (Ref. 28)   | 12.6 (Ref. 28)   | 86.2 (Ref. 28)   | 181 (Ref. 28)    | 264 (Ref. 28)    | 523 (Ref. 28)    | 616 (Ref. 28)    |
| C12                   | 0.99 (Ref. 28)   | 8.0 (Ref. 28)    | 32.0 (Ref. 28)   | 77 (Ref. 28)     | 158 (Ref. 28)    | 203 (Ref. 28)    | 273 (Ref. 28)    |
| C44                   | 1.44 (Ref. 28)   | 9.5 (Ref. 28)    | 26.8 (Ref. 28)   | 55.7 (Ref. 28)   | 82.6 (Ref. 28)   | 160 (Ref. 28)    | 161 (Ref. 28)    |
| C12-C44               | -0.45            | -1.5             | 5.2              | 21.3             | 75.4             | 43               | 112              |

| Property              | Os               | Ir               | Pt               | Au               | Tl               | Pb              | Bi               |
|-----------------------|------------------|------------------|------------------|------------------|------------------|-----------------|------------------|
| $\alpha$              | 8.5 (Ref. 27)    | 7.6 (Ref. 27)    | 6.5 (Ref. 27)    | 5.8 (Ref. 27)    | 7.6 (Ref. 27)    | 6.8 (Ref. 27)   | 7.4 (Ref. 27)    |
| $\chi$                | 1.52 (Ref. 28)   | 1.44 (Ref. 28)   | 1.44 (Ref. 28)   | 1.42 (Ref. 28)   | 1.44 (Ref. 28)   | 1.55 (Ref. 28)  | 1.67 (Ref. 28)   |
| $m$                   | 190.23 (Ref. 29) | 192.22 (Ref. 29) | 195.08 (Ref. 29) | 196.97 (Ref. 29) | 204.38 (Ref. 29) | 207.2 (Ref. 29) | 208.98 (Ref. 29) |
| $Z$                   | 76               | 77               | 78               | 79               | 81               | 82              | 83               |
| $r$                   | 1.30 (Ref. 30)   | 1.35 (Ref. 30)   | 1.35 (Ref. 30)   | 1.35 (Ref. 30)   | 1.90 (Ref. 30)   | 1.80 (Ref. 30)  | 1.60 (Ref. 30)   |
| $V$                   | 14.36 (Ref. 32)  | 14.62 (Ref. 32)  | 15.83 (Ref. 32)  | 18.19 (Ref. 32)  | 31.25 (Ref. 32)  | 31.93 (Ref. 32) | 35.38 (Ref. 33)  |
| $E_A$                 | 1.1 (Ref. 28)    | 1.57 (Ref. 28)   | 2.13 (Ref. 28)   | 2.31 (Ref. 27)   | 0.2 (Ref. 28)    | 0.36 (Ref. 28)  | 0.95 (Ref. 28)   |
| $I$                   | 8.44 (Ref. 29)   | 8.97 (Ref. 29)   | 8.96 (Ref. 29)   | 9.23 (Ref. 29)   | 6.11 (Ref. 29)   | 7.42 (Ref. 29)  | 7.29 (Ref. 29)   |
| $E_{\text{tot}}^{35}$ | -16286.30        | -16801.68        | -17326.58        | -17860.79        | -18956.96        | -19518.99       | -20090.41        |
| $E_{\text{coh}}$      | 8.17 (Ref. 36)   | 6.94 (Ref. 36)   | 5.84 (Ref. 36)   | 3.81 (Ref. 36)   | 1.88 (Ref. 36)   | 2.03 (Ref. 36)  | 2.18 (Ref. 36)   |
| $B$                   | 417.8 (Ref. 37)  | 355 (Ref. 37)    | 278.3 (Ref. 37)  | 173.2 (Ref. 37)  | 35.9 (Ref. 37)   | 42.9 (Ref. 37)  | 31.5 (Ref. 37)   |
| $G$                   | 223 (Ref. 28)    | 209 (Ref. 28)    | 60.9 (Ref. 28)   | 26 (Ref. 28)     | 2.67 (Ref. 28)   | 5.54 (Ref. 28)  | 12.8 (Ref. 28)   |
| $E$                   | 559 (Ref. 28)    | 528 (Ref. 28)    | 170 (Ref. 28)    | 78 (Ref. 28)     | 7.89 (Ref. 28)   | 15.8 (Ref. 28)  | 34 (Ref. 28)     |
| $\mu$                 | 0.25 (Ref. 28)   | 0.26 (Ref. 28)   | 0.39 (Ref. 28)   | 0.42 (Ref. 28)   | 0.45 (Ref. 28)   | 0.44 (Ref. 28)  | 0.33 (Ref. 28)   |
| B/G                   | 1.87             | 1.70             | 4.57             | 6.66             | 13.45            | 7.74            | 2.46             |
| C11                   | 729.6 (Ref. 32)  | 580 (Ref. 28)    | 347 (Ref. 28)    | 191 (Ref. 28)    | 41.9 (Ref. 28)   | 48.8 (Ref. 28)  | 63.4 (Ref. 28)   |
| C12                   | 230.0 (Ref. 32)  | 242 (Ref. 28)    | 251 (Ref. 28)    | 162 (Ref. 28)    | 36.6 (Ref. 28)   | 41.4 (Ref. 28)  | 24.5 (Ref. 28)   |
| C44                   | 251.5 (Ref. 32)  | 256 (Ref. 28)    | 76.5 (Ref. 28)   | 42.2 (Ref. 28)   | 7.20 (Ref. 28)   | 14.8 (Ref. 28)  | 11.5 (Ref. 28)   |
| C12-C44               | -21.5            | -14              | 174.5            | 119.8            | 29.4             | 26.6            | 13               |

The quantities given in the table are as follows:  $\alpha$  is the polarizability ( $\text{\AA}^3$ ),  $\chi$  is the electronegativity (eV),  $m$  is the atomic weight (amu),  $U$ ,  $Z$  is the atomic number,  $r$  is the atomic radius ( $\text{\AA}$ ),  $V$  is the atomic volume ( $\text{\AA}^3/\text{atom}$ ),  $E_A$  is the electron affinity (eV),  $I$  is the ionization energy (eV),  $E_{\text{tot}}$  is the total energy (eV),  $E_{\text{coh}}$  is the cohesive energy (eV),  $B$  is the bulk modulus (GPa),  $G$  is the shear modulus (GPa),  $E$  is the Young's modulus (GPa),  $\mu$  is the Poisson number, the ratio  $B/G$ , elastic stiffness (C11, C12, and C44) (GPa), and the Cauchy pressure (C12-C44) (GPa).

sets of variables. Various PLS techniques are used to obtain latent vectors, which gives a variant of PLS.<sup>17</sup>

## RESULTS AND DISCUSSION

### Principal Component Analysis (PCA)

In this work, PCA of the database yielded two significant PCs that together account for 81.98% of the total variance in the data set, where PC1 contributes

to 54.40% of the total, whereas PC2 contributes 27.58%, as shown in (Fig. 1). The objective of the loading plot is to visualize the position of the variables (i.e., the properties in our data set), with respect to one another in a two-dimensional space, and their corresponding correlations. By comparing the distances between the points representing the properties, we can deduce that the variables, represented by points that are close to each other and far from the origin, are positively correlated, whereas those opposite one another on the plot are negatively correlated (or inversely proportional). For instance, variables like  $E_{tot}$ ,  $m$ , and  $Z$  can be seen in (Fig. 2). The loading plot explains the relationships relating two variables by their angles at center on the plot. The correlation coefficient between two variables represents the cosine of the angle between their corresponding vectors on the plot. In Fig. 2, the cosine of  $180^\circ$  (i.e., the angle between  $I$  and  $V$  on the PC1–PC2 plot) is  $-1$ , which means that they are negatively correlated. Based on this mathematical rule, the uncorrelated variables are at right angles to each other (cosine  $90^\circ = 0$ , therefore not correlated). Similarly, the cosine of  $0^\circ$  is 1, which indicates a positive correlation between the variables.<sup>13</sup> The plot of the scores for PC1 versus PC2, shown in Fig. 1 reveals five distinct groups of elements. Group 1 (Cs and Ba) is related to elements of block s, groups 2 (Ta, Hf, and Lu) and 3 (Ir, Os, Re, and W) contain elements that belong to block d, group 4 (Pt, Au) contains elements that are precious metals, and group 5 (Tl, Pb, and Bi)

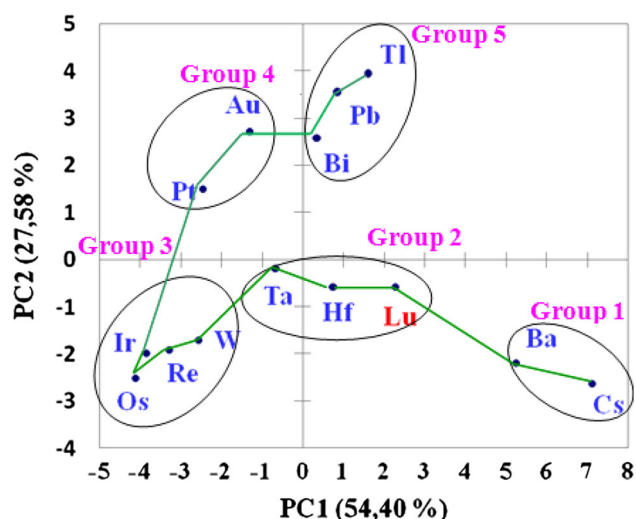


Fig. 1. PCA scores plot from the database of Table 1 containing 14 elements of the third-row transition elements, including lutetium.

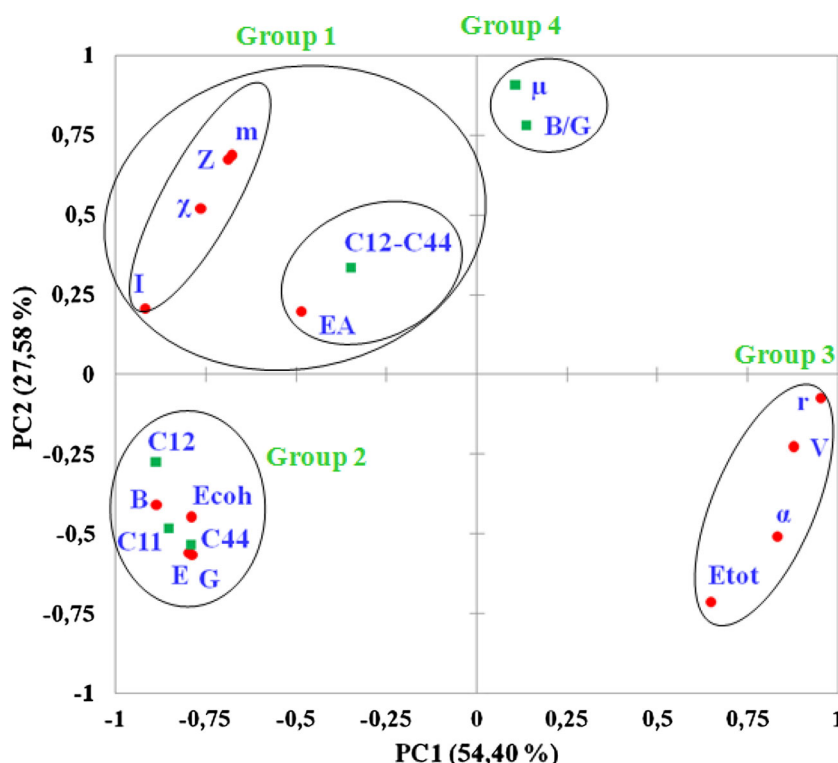


Fig. 2. PCA loadings plot from the database of Table 1, with 19 properties.

includes elements of post-transition metals, whose properties are determined by their partially filled  $p$ -subshells. We note that the axis PC1 separates clusters 1, 2, and 3 from 4 and 5. There is a close similarity between elements in each group. Looking at group 2 in Fig. 1 and Table I, we note that the distances separating the points representing the elements Ta, Hf, and Lu are small, which indicates that they have properties similar to those of hafnium (i.e., transition metals), and they share several attributes with them because they have the same valence shell configuration, which is equal to 2. Empirically, it was shown that lutetium's outermost electron is the first electron in the  $5d$  subshell; it must mark the beginning of the third series of transition elements (i.e., the sixth row of the periodic table), below yttrium (Y) in group 3 (IIIB) because it is a  $d^1$  metal. The location of an element on the periodic table is based on its electron configuration, so the elements in the same row (family) have the same valence shell configuration. The valence electrons determine the chemistry, but the electron configurations determine the organization of the periodic table as well as the properties of the elements. These are the reasons for the periodic trends in the behavior of chemical elements. The properties of the elements are determined by the size ( $n$ ), shape ( $l$ ) of orbitals, and atomic number (nuclear charge). A great deal of chemical and physical evidence shows that lutetium should be placed in the same group as Sc and Y.<sup>18</sup> This issue was debated by several researchers (Refs.5 and 19–24) and discussed in a more persuasive way by Jensen in 1982.<sup>19</sup> Another justification for placing lutetium (Lu) below yttrium (Y) in group 3 (IIIB) is that this arrangement gives similar periodic trends in several physical properties (melting point, atomic radii, sum of the first two ionization potentials, and electronegativity), as seen in (Table II). Sc, Y, and Lu all have an hexagonal closed-packed (hcp) structure at an ambient temperature.<sup>18</sup>

Placing the elements in the periodic table is currently accepted according to the following empirical observations: atomic number, properties, periodic trends, and ground-state electron configuration of atoms.<sup>25</sup>

In the PCA loadings plot (Fig. 2), four main clusters can be observed: Cluster 1 contains mainly the

physical properties that can be decomposed into two subgroups. The first subgroup contains the atomic properties ( $m$ ,  $Z$ ,  $I$ , and  $\chi$ ) and the second one contains the Cauchy pressure C12–C44, and  $E_A$ . The second cluster group involves the mechanical properties  $B$ ,  $E$ ,  $G$ , C11, C12, and C44 including the structural property  $E_{\text{coh}}$ . The third one comprises physical properties such as the atomic radius  $r$ , atomic volume  $V$ , polarizability  $\alpha$ , and total energy  $E_{\text{tot}}$ , and the fourth cluster group consists of the ratios  $B/G$  as well as the Poisson's ratio  $\mu$ . The properties that have identical PC values are significantly correlated and are similar to each other; however those which have different PC values are inversely correlated and are therefore different. Within group 1, the closest resemblance can be observed between  $m$  and  $Z$  on the one hand, and among  $E$ ,  $G$  and C44 on the other hand. In group 2, their points are close to each other (Fig. 2), which reveals the importance of these properties. The PCA automatically separates data. Indeed, we see that PC2 clearly separates the physical from the atomic properties, which means a strong difference between these two properties. Variables within the same cluster group are correlated. Groups 1 and 3 are inversely correlated, as are groups 2 and 4. It is possible to identify several relationships. An increase in the atomic weight  $m$  and the atomic number  $Z$  will result in an increase in  $I$  and  $\chi$ , as well as a decrease in  $r$ ,  $V$ ,  $E_{\text{tot}}$ , and  $\alpha$ . However, Young's modulus  $E$ , the shear modulus  $G$ , and the bulk modulus  $B$  are correlated and tightly bound to the elastic stiffness C11, C12, C44, and the cohesive energy  $E_{\text{coh}}$ . By means of PCA, we create a map that indicates how the parameters and properties will qualitatively change when changing the chemical elements.

### Partial Least Squares

We study the relationships between two blocks of variables, that is, between the mechanical properties ( $E$ ,  $B$ ,  $G$ ,  $\mu$ ,  $B/G$ , C11, C12, C44, and C12–C44) and the atomic and structural properties ( $\alpha$ ,  $\chi$ ,  $m$ ,  $Z$ ,  $r$ ,  $V$ ,  $E_A$ ,  $I$ ,  $E_{\text{coh}}$ , and  $E_{\text{tot}}$ ), using the PLS approach. The PLS results, which were presented in (Fig. 3), show the correlation circle that explores the relationship among the mechanical, atomic, and structural properties, as well as between the two PLS

**Table II. Physical properties of group 3 (IIIB), including lutetium**

| Property<br>Structure    | Sc<br>hcp       | Y<br>hcp        | Lu<br>hcp       |
|--------------------------|-----------------|-----------------|-----------------|
| $T_{\text{melt}}$        | 1.541 (Ref. 18) | 1.522 (Ref. 18) | 1.663 (Ref. 18) |
| $\chi$                   | 1.20 (Ref. 28)  | 1.11 (Ref. 28)  | 1.14 (Ref. 28)  |
| $r$                      | 1.60 (Ref. 30)  | 1.80 (Ref. 30)  | 1.734 (Ref. 31) |
| First ionization energy  | 6.56 (Ref. 28)  | 6.22 (Ref. 28)  | 5.43 (Ref. 29)  |
| Second ionization energy | 12.80 (Ref. 28) | 12.24 (Ref. 28) | 13.9 (Ref. 28)  |

The quantities given in the table are as follows: the melting point ( $^{\circ}\text{C}$ ),  $\chi$  is the electronegativity (eV),  $r$  is the atomic radius ( $\text{\AA}$ ), and the ionization energy (eV).

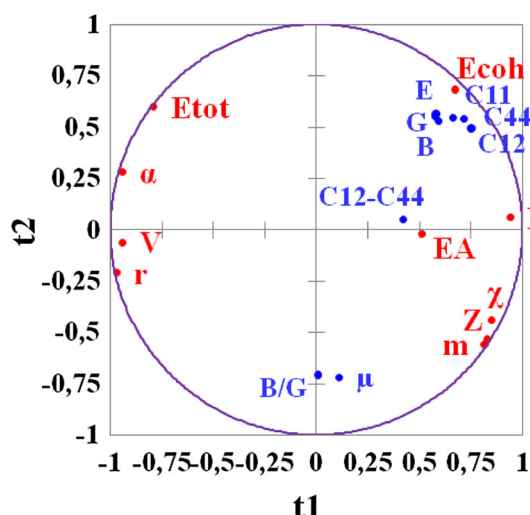


Fig. 3. Variable map in PLS regression of mechanical properties according to atomic and structural properties.

components ( $t_1$ ,  $t_2$ ), which captured the maximum of the variance within the data set. The results can be explained in the same way, as in the PCA. The cosines of the angles of the vectors representing the variables (properties) in the plane of the graph in Fig. 3 designate the correlations that exist between the properties.<sup>26</sup> We can see that for properties such as C12–C44 and the electron affinity  $E_A$ , displayed at the center of the map (the correlation circle), the correlations are low. This means that their properties are not well represented on the first two dimensions. The properties C12–C44 and  $E_A$  barely explain the trends of the elements; they do not have a strong effect that could easily influence the elements' trend; they are not important. We notice strong correlations between  $m$  and  $Z$  on the one hand, as well as between the elastic constants C11, C12, and C44, and  $E$ ,  $G$ , and  $B$ , on the other hand. Furthermore, a negative correlation is found between  $E_{tot}$  and  $m$ ,  $Z$ , as well as between the ionization energy  $I$  and the atomic volume  $V$ . The variables  $B/G$  and  $\mu$  are separated from the other variables because they are less correlated to these variables and to PLS components ( $t_1$ ,  $t_2$ ). The properties located on the periphery of the correlation circle are significant; i.e., they strongly contribute to the discrimination.

## CONCLUSION

In this work, we introduced a new approach to analyze and compare the physical and mechanical properties of lutetium with elements belonging to the third-row transition elements, using multivariate analysis methods such as PCA and PLS regression. PCA allowed classifying both the properties and the elements, which helped us to visualize the position of the elements relative to each other, as well as the relationships existing between their properties. The use of PLS regression fits well

with PCA results. A major interest in this methodology is given by the unitary framework, based on PLS principles, which allows to uncover the structural relationships existing among data and properties. Our results show that the properties of lutetium (Lu) are similar to those of hafnium (Hf). We found that our results are generally in good agreement with many research studies previously published on these subjects. It is difficult for scientists to understand the interactions between large numbers of variables, so we used PCA and PLS as tools for identifying the structures existing in a data set, as well as for screening and extracting useful information from data.

## REFERENCES

1. E.R. Scerri, *J. Sci. Educ.* 12, 4 (2011).
2. A. Hérold, *C. R. Chim.* 9, 148 (2006).
3. M-L. Kamberg, *The Transition Elements: The 37 Transition Metals*, 1st ed. (New York: The Rosen Publishing Group Inc., 2010).
4. M. Halka and B. Nordstrom, *Periodic Table of the Elements—Transition Metals* (New York: Facts On File Inc, 2011).
5. H. Merz and K. Ulmer, *Phys. Lett.* 26, 1 (1967).
6. M. Laing, *Found. Chem.* 7, 203 (2005).
7. E. Scerri, *Found. Chem.* 12, 69 (2010).
8. H. Abdi and L.J. Williams, *WIREs Comput. Stat.* 2, 433 (2010).
9. A. Lengyel, K. Héberger, L. Paksy, O. Bánhidi, and R. Rajkó, *Chemosphere* 57, 889 (2004).
10. S.R. Broderick, H. Aourag, and K. Rajan, *Phys. B* 406, 2055 (2011).
11. H.-N. Qu, G.-Z. Li, and W.-S. Xu, *Pattern Recognit.* 43, 3448 (2010).
12. G. Köksal, I. Batmaz, and M.C. Testik, *Exp. Syst. Appl.* 38, 13448 (2011).
13. E.-C. Shin, B.D. Craft, R.B. Pegg, R.D. Phillips, and R.R. Eitenmiller, *Food Chem.* 119, 1262 (2010).
14. K. Rajan, *Mater. Today* 8, 38 (2005).
15. I.T. Jolliffe, *Principal Component Analysis, Springer series in statistics*, 2nd ed. (New York: Springer, 2002).
16. I. Marchi, S. Rudaz, M. Selman, and J.-L. Veuthey, *J. Chromatogr. B* 845, 244 (2007).
17. R. Rosipal and N. Krämer, *Subspace, Latent Structure and Feature Selection Lecture Notes in Computer Science*, Vol. 3940, eds. C. Saunders, M. Grobelnik, S. Gunn, and J.S. Taylor (New York: Springer, 2006), pp. 34–51.
18. K.A. Gschneidner Jr, J.-C.G. Bünzli, and V.K. Pecharsky, *Handbook on the Physics and Chemistry of Rare Earths*, Vol. 41 (Dordrecht, The Netherlands: North-Holland, Elsevier, 2011).
19. W.B.J. Jensen, *Chem. Educ.* 59, 634 (1982).
20. E.R.J. Scerri, *Chem. Educ.* 68, 122 (1991).
21. W.B.J. Jensen, *Chem. Educ.* 80, 952 (2003).
22. W.F. Luder, *The Electron-Repulsion Theory of the Chemical Bond* (New York: Reinhold, 1967).
23. D.C. Hamilton, *Am. J. Phys.* 33, 637 (1965).
24. B.T. Matthias, W.H. Zacharisen, G.W. Webb, and J.J. Engelhardt, *Phys. Rev. Lett.* 18, 781–784 (1967).
25. L. Lavelle, *J. Chem. Educ.* 85, 1482 (2008).
26. M. Tenenhaus, J. Pagès, L. Ambroisine, and C. Guinot, *Food Qual. Prefer.* 16, 315 (2005).
27. D.R. Lide, eds., *CRC Handbook of Chemistry and Physics*, 90th ed. (Boca Raton, FL: CRC Press/Taylor and Francis, 2010).
28. W. Martienssen and H. Warlimont, eds., *Springer Handbook of Condensed Matter and Materials Data*, Vol. 1 (Berlin/Heidelberg, Germany: Springer, 2005).

A Study of the Physical and Mechanical Properties of Lutetium Compared with Those of Transition Metals: A Data Mining Approach

29. R.A. Dragoset, A. Musgrove, C.W. Clark, and W.C. Martin, *Periodic Table: Atomic Properties of the Elements* (Version 4), NIST SP 966 (Gaithersburg, MD: National Institute of Standards and Technology, 2003), <http://physics.nist.gov/PT>.
30. J.C. Slater, *J. Chem. Phys.* 39, 3199 (1964), <http://www.webelements.com/>.
31. S. Cotton, *Lanthanide and Actinide Chemistry* (London, U.K.: Wiley, 2006).
32. S.L. Shang, A. Saengdeejing, Z.G. Mei, D.E. Kim, H. Zhang, S. Ganeshan, and Y. Wang, *Comput. Mater. Sci.* 48, 813 (2010).
33. D. Stewart, *Chemicool Periodic Table* (Cambridge, MA: The Massachusetts Institute of Technology, 2013), <http://www.chemicool.com>.
34. E.C.M. Chen and E.S. Chen, *J. Chromatogr. A* 1037, 83 (2004).
35. NIST (National Institute of Standards and Technology), (2013), [www.nist.gov](http://www.nist.gov).
36. C. Kittel, *Introduction to Solid State Physics*, 8th ed. (Hoboken, NJ: Wiley, 2005).
37. Y. Makino and S. Miyake, *J. Alloy. Compd.* 313, 235 (2000).

## Abstract

The periodic table of elements, is the most important generalization in chemistry, it is often considered as a special representative case in the study of the relationship between chemistry and physics. The study of the anomalies in the periodic table has been the subject of many researches, such as the study of the positions of lutetium and hydrogen, but no one could clear up the problem. In this work we tried to confront this problem, by applying the data mining approach and its techniques of multivariate analysis such as Principal Component Analysis (PCA) and regression (PLS). The obtained results indicate the significative similarity shared between lutetium and transition metals of the third row of the periodic table, which shows that the position of lutetium must be below yttrium and next hafnium, while for hydrogen it was found that it did not conform to any group of the periodic table, it is a special element.

Key words: Periodic table- Hydrogen- Lutetium- data mining approach- PCA- PLS

## Résumé

Le tableau périodique des éléments, est la généralisation la plus importante en chimie, il est souvent considéré comme un cas représentatif particulier dans l'étude de la relation entre la chimie et la physique. L'étude des contradictions dans le tableau périodique a été le sujet de beaucoup de recherches, tels que l'étude de la position du lutétium et la position de l'hydrogène, mais sans pouvoir éclaircir le problème. Dans ce travail on a essayé de confronter ce problème, moyennant l'approche data mining, et en appliquant les techniques de l'analyse multivariée telles que l'Analyse en composantes principales (ACP) et la régression (PLS). Les résultats obtenus indiquent la grande similitude entre le lutétium et les métaux de transition de la troisième période du tableau périodique, ce qui montre que la position du lutétium doit être en dessous de l'yttrium à côté du hafnium, tandis que pour l'hydrogène on a trouvé qu'il n'est conforme à aucun groupe du tableau périodique, c'est un élément spécial.

Mots clés : Le tableau périodique- Le lutétium- l'hydrogène- Le data mining- L'ACP- La PLS

## ملخص

الجدول الدوري للعناصر (periodic table)، هو التعميم الأكثر أهمية في الكيمياء، ويعتبر في غالب الأحيان على أنه حالة خاصة ممثلة للعلاقة بين الكيمياء والفيزياء. وكانت دراسة التناقضات الموجودة في الجدول الدوري الموضوع الذي خص بالكثير من البحوث، مثل دراسة مكان اللوتيسيوم (Lu) والهيدروجين (H)، ولكن دون جدوى. في هذا العمل حاولنا مواجهة هذه المشكلة، باستعمال مقارنة استخراج البيانات (data mining) وتطبيق تقنيات التحليل متعدد المتغيرات مثل التحليل لى مكونات أساسية (PCA) والتراجع (PLS) وتشير النتائج إلى تشابه وثيق بين اللوتيسيوم و المعادن الانتقالية المنتمية الى السطر الثالث من الجدول الدوري، مما يدل على أن مكان اللوتيسيوم يجب أن يكون أسفل من الإيتريوم (Y) و بجانب الهافنيوم (Hf)، في حين قد وجد أن الهيدروجين لا يتوافق مع أي مجموعة من الجدول الدوري، بل هو عنصر خاص.

**كلمات مفتاحية:** الجدول الدوري للعناصر، اللوتيسيوم، الهيدروجين، مقارنة استخراج البيانات