

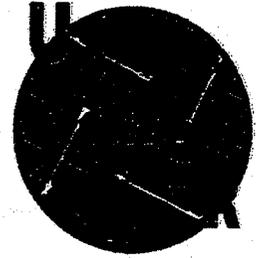
MS/ 530 - 06/01

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE.
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE.
UNIVERSITE ABOU-BAKR BELKAID DE TLEMCCEN.



FACULTE DES SCIENCES

LEPM Département de physique



Laboratoire d'Etude et Prédiction des Matériaux
Unité de recherche Matériaux et Energies Renouvelables

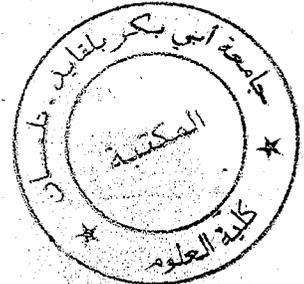
MEMOIRE DE MASTER
EN

Inscrit Sous le n°
Date le 09 JUL 2012
Code: 7599

PHYSIQUE de la MATIERE CONDENSEE et des SEMI-
CONDUCTEURS

Par :

Mme Mammour Fatiha



THEME

*La théorie de la topologie de la densité de
charge dans le cadre de la DFT*

Soutenu le : 30 / 06/ 2012 devant le jury :

Dr. H. I. FARAOUN
Pr. H. AOURAG
Pr. G. MERAD

Président.
Encadreur.
Examineur.

Université de Tlemcen.
Université de Tlemcen.
Université de Tlemcen.

Année universitaire 2011/2012



Remerciements

Nous remercions "ALLAH " le tout puissant, de nous avoir donné la volonté et le courage de terminer mon travail, et de toujours nous guider vers le bon et le droit chemin.

Nous tenons à remercier tous ceux et celles qui nous ont aidés à réaliser ce travail, et leurs présenter toute notre gratitude.

Je tiens à exprimer ma reconnaissance à **Mr. H. AOURAG**, Professeur à l'université de Tlemcen, pour avoir dirigé ce mémoire, pour son suivi permanent, ses lectures attentives, ses conseils judicieux et le soutien constant qu'il m'a prodigué au cours de l'élaboration de ce travail.

J'exprime mes sincères remerciements à **Melle. H. I. FARAOUN**, Maître de conférences à l'université de Tlemcen d'avoir accepté de présider le Jury de cette thèse.

J'exprime mes sincères remerciements à monsieur **G. MERAD**, Professeur à l'Université de Tlemcen, d'avoir accepté de faire partie du Jury.

Je tiens vivement à exprimer mes remerciements à **Mr ZENASNI Hocine** pour l'aide précieuse qu'elle n'a cessé de m'apporter tout le long de ce travail à travers ses conseils et l'attention qu'elle a apporté à mon travail, j'y suis reconnaissante.

J'adresse aussi mes remerciements les plus chaleureux à mon cher mari par leurs conseils, leurs prières, et leurs encouragements.

Je tiens également à exprimer mes remerciements à tous les étudiants de Master en physique de la matière condensée et de semi conducteurs (2011) sans exception.

Enfin, je tiens à remercier toutes les personnes qui de loin ou de près, m'ont aidé par leurs conseils, leurs critiques ou leur encouragements.

Dédicaces

Avec grande joie et cœur ouvert je dédie ce modeste travail .

Surtout à mes très chers parents

A mon mari et ses parents

A ma sœur et son mari

A mes frères

A tout mes amies sans exception

Fatiha

Table des matières

Introduction générale.....1

Chapitre 1 Synthèse Bibliographique

I- Introduction2

II- Intermétalliques composés.....2

 II.1. Application des intermétalliques3

 II.2. Applications pratiques d'intermétalliques :.....3

 II.3. Exemples d'alliages intermétalliques.....3

III- Succès et limites de la DFT4

 III.1. Densité de charge électronique4

 III.2. Liaisons chimiques.....6

 III.2.1. Liaison covalente.....6

 III.2.2. Liaison ionique :.....6

 III.2.3. La liaison métallique7

 III.2.4. Liaison de Van der Waals :.....7

Références.....8

Chapitre II Datamining

I- Introduction9

II- Datamining9

 II.1. Définition.....9

 II.2. Description d'un projet de Datamining9

 ✓ Obtention de données significatives9

 ✓ Analyse des données9

 ✓ Exploitation des données10

 II.3. Principe de datamining.....10

 II.4. Méthodes statistiques.....11

 II.4.1. L'analyse en composantes principales11

 II.4.1.1. Introduction.....11

 ✓ Qu'est-ce que l'ACP ?.....11

 II.4.1.2. Tableau de données.....12

 II.4.1.3. Choix de l'origine13

 II.4.1.4. Formulation mathématique de l'ACP14

 II.4.1.5. Principe de l'analyse en composante principale.....14

 II.4.2. La régression (PLS)15

 II.4.2.1. Introduction.....15

 ✓ Qu'est-ce que la régression PLS ?15

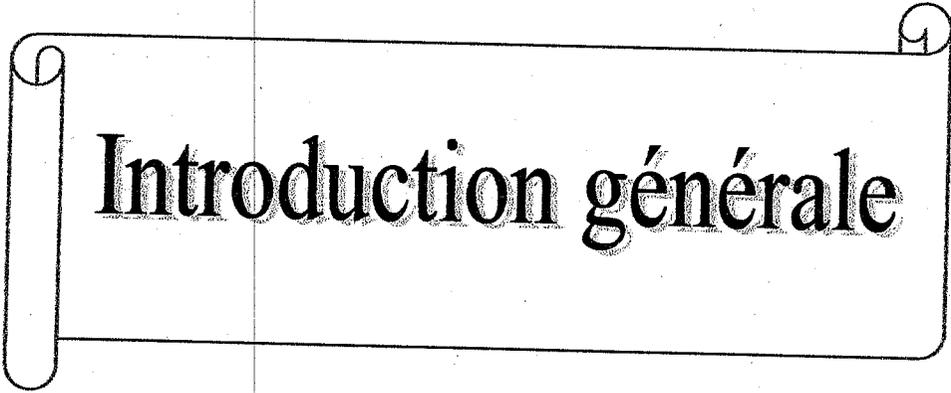
✓ Etablissement du modèle PLS1 simple.....	16
✓ Etablissement du modèle de La régression PLS2	17
II.4.2.2. Interprétation du modèle PLS	17
II.4.2.3. Intérêt de la Régression PLS	18
II.4.2.4. Conclusion.....	18
Référence	19

Chapitre III
Résultats et interprétation

I- Introduction :	20
II-Résultats	20
II.1. Définition de la base de données.....	20
II.2. Base de données (1).....	20
II.2.1. L'analyse en composantes principales (ACP).....	21
1. Choix des axes.....	22
2. Matrice de corrélation.....	22
3. Interprétation des données.....	23
II.2.2- Résultats graphiques de la régression PLS.....	26
1. Paramètres et équation du modèle.....	26
2. prédictions à la variable «B ₁ »... ..	26
II.3. Base de données (2).....	30
II.3.1. L'analyse en composantes principales (ACP)....	31
1. Matrice de corrélation.....	31
2. Interprétation des données.....	32
II.3.2. Résultats graphiques de la régression PLS.....	34
1. Paramètres et équation du modèle.....	34
2. Prédications à la variable «B ₂ ».....	34
III-Conclusion.....	37
Références.....	38
Conclusion générale.....	39

Annexe

I-Introduction.....	40
II- Faire une Analyse en Composantes Principales (ACP) avec XLSTAT.....	40
III- Faire une régression PLS avec XLSTAT.....	42



Introduction générale

Introduction générale

Les composés intermétalliques sont d'intérêt pour les scientifiques et les ingénieurs en raison de leurs propriétés supérieures, électriques, chimiques, et la stabilité mécanique à haute température.

La densité de charge électronique constitue un moyen topologique pour identifier la nature de la liaison.

Dans ce travail nous allons essayer de tester la corrélation entre les propriétés structurales (module de compressibilité) et la densité moyenne des électrons, pour des composés binaires intermétalliques.

Cette thèse s'articule autour de trois chapitres :

Le premier chapitre est consacré à la présentation bibliographique des composés intermétalliques, et leurs propriétés ainsi que les rappels sur le principe de la théorie de la fonctionnelle de la densité (DFT). Ensuite, nous aborderons le concept de la densité électronique, et des notions de bases sur les liaisons.

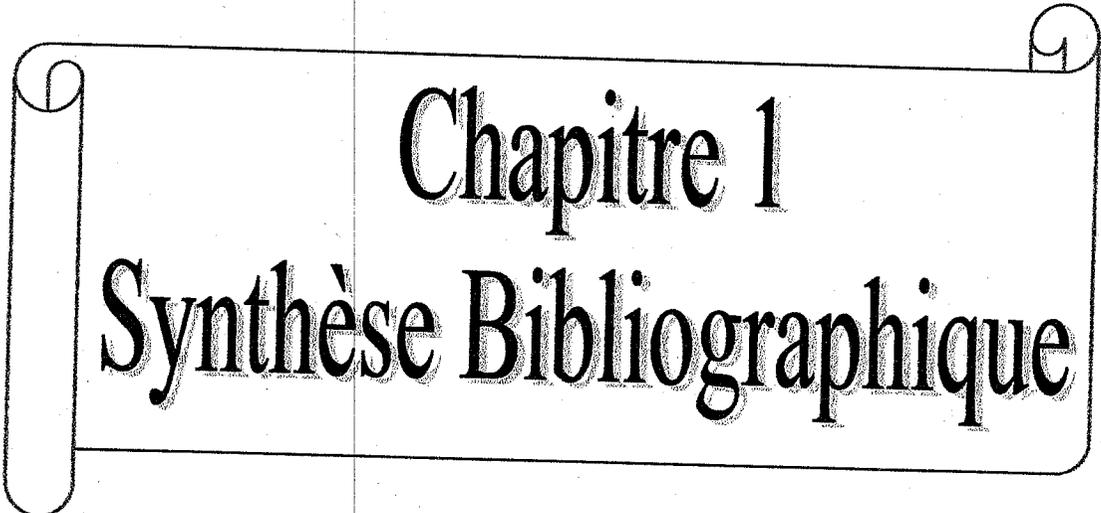
Le second chapitre développe les techniques de datamining nous insistons:

L'ACP (analyse en composante principale) et la PLS (Partial Least Squares régression).

Enfin, dans le troisième chapitre, nous présenterons les résultats obtenus à partir de notre base de données par les méthodes ACP et PLS. Nous discuterons ces résultats pour tirer quelques conclusions, et nous ferons des comparaisons avec les connaissances théoriques établies.

Ce mémoire se termine par une conclusion générale récapitulative sur les principaux résultats obtenus.

En annexe, sera présenter le logiciel utilisé dans nos calculs, qui est un logiciel statistique d'analyse des données pour Microsoft Excel, intitulé XLSTAT.



Chapitre 1
Synthèse Bibliographique

I. Introduction :

Un alliage forme un intermétallique à certaines compositions et conditions de pression. Il s'agit d'une alternance d'atomes qui forment une structure périodique. Les intermétalliques se forment lorsque les atomes de l'alliage ont une forte différence d'électronégativité ; la densité électronique devient alors hétérogène, la liaison n'est plus strictement métallique mais devient partiellement covalente ou ionique. [1, 2]

Initialement, les alliages intermétalliques se formaient des précipités dans un alliage plus complexe (par exemple dans les super-alliages à base nickel comme les Inconels). Ces précipités jouent un rôle important dans le renforcement mécanique des alliages (durcissement structural). Par la suite, ces phases ont été isolées, caractérisées, et on a commencé dans les années 1960 à les utiliser comme constituant massif de pièces et non plus comme phase mineure.

II. Intermétalliques composés :

Lorsque deux éléments A et B sont miscibles en toutes proportions et forment une solution solide continue pour certains domaines de concentration, il peut se former des sur structurés. Les composés A_3B , AB , AB_3 sont des exemples de ces nouvelles phases. Le nom de phases intermédiaires est plutôt réservé à celles qui possèdent une structure différente de celle des métaux de base ou des solutions solides terminales. On parle de composé intermétallique lorsque la phase intermédiaire n'existe qu'à l'état ordonné. Ces composés ont un caractère métallique plus ou moins marqué. Leur stabilité dépend de différents facteurs : facteur de valence (ou de concentration électronique), facteur de taille, facteur électrochimique (différence d'électronégativité).

Ces composés présentent des propriétés intéressantes pour de nombreuses applications. Les composés intermétalliques sous forme de particules dispersées dans une matrice métallique (phénomène de précipitation) modifient de façon notable les propriétés de cette matrice. De nombreux alliages sont durcis par précipitation : précipitation des carbures (aciers spéciaux), des composés intermétalliques (alliages légers, aciers à hautes performances). Les eutectiques contrôlés possèdent de remarquables propriétés mécaniques : ce sont des composés intermétalliques sous forme de longues baguettes ou de grandes lamelles, régulièrement disposées parallèlement à une même direction, dans une matrice plus ductile (principe du renforcement par fibres). Les composés à grande maille ont souvent une très grande dureté,

accompagnée d'une fragilité très marquée ; ils sont souvent utilisés sous forme de pièces frittées (carbures pour outil de coupe). Des composés covalents comme InSb et GaAs sont utilisés comme semi-conducteurs, d'autres (de type A_3B) ont des températures critiques de supraconduction élevées : 18,05 K pour Nb_3Sn et 17,9 K pour V_3Si . [3,4]

II.1. Application des intermétalliques : [5]

Intermétalliques ont donné lieu à divers développements de nouveaux matériaux. En voici quelques exemples des matériaux de stockage d'hydrogène dans les piles au nickel-métal hydrure. Ni_3Al , qui est dans les familiers à base de nickel des superalliages, et des aluminures de titane différents ont également suscité l'intérêt pour les applications de pales de turbine, tandis que le second est également utilisé en très petites quantités pour affinement du grain des alliages de titane. Siliciures, intermétalliques impliquant silicium, impliquant de nombreux éléments ont été étudiés pour, et certains utilisés comme, la barrière et les couches de contact en microélectronique.

Une grande partie du travail a été fait dans les 10 à 15 dernières années pour développer et caractériser intermétalliques et de développer des technologies de traitement. En réponse à la nécessité d'une faible densité, les alliages à hautes performances pour une utilisation dans les composants de cellules et les moteurs à turbine, par exemple des efforts concertés ont été réalisés ces dernières années pour améliorer les propriétés des alliages intermétalliques, en particulier alliages à base d'aluminures , par exemple, $TiAl$, Ti_3Al , $NiAl$, Ni_3Al , $FeAl$.

II.2. Applications pratiques d'intermétalliques :

A la fois terrestres et aérospatiales; poly phasiques alliages intermétalliques à usage dentaire; applications de matériaux magnétiques durs basés sur intermétalliques, en particulier ceux comportant des métaux des terres rares.

II.3. Exemples d'alliages intermétalliques :

- Aluminures de fer : $FeAl$ B_2 , Fe_3Al .
- Aluminures de nickel : $NiAl$, Ni_3Al .
- Aluminures de titane : $TiAl$
- Alliages de terres rares pour applications magnétiques :

- samarium-fer $\text{Sm}_2\text{Fe}_{17}$
- samarium-cobalt
- samarium-fer-cobalt
- gadolinium-fer $\text{Gd}_2\text{Fe}_{17}$
- samarium-fer-gadolinium SmGdFe_{17}

III. Succès et limites de la DFT :

La DFT permet souvent d'obtenir, à plus faible coût, des résultats avec une bonne précision. De plus, elle peut être utilisée pour étudier des systèmes relativement gros, contenant jusqu'à plusieurs centaines d'électrons, ce que les calculs post-Hartree-Fock ne peuvent pas traiter. Ceci explique que la DFT soit aujourd'hui très utilisée pour étudier les propriétés de systèmes moléculaires complexes ou même biologiques.

Les nombreux travaux effectués ces dernières années montrent que les calculs DFT donnent de bons résultats pour les états fondamentaux de systèmes très divers (métalliques, ioniques, organo-métalliques, métaux de transition...) pour de nombreuses propriétés (structures moléculaires, fréquences de vibration, potentiels d'ionisation...).

Toutefois, ces méthodes souffrent encore de plusieurs défauts. Ainsi, il semblerait, jusqu'à preuve du contraire, que les forces de dispersion ou Van Der Waals ne sont pas encore traitées correctement en DFT sauf dans les fonctionnelles développées spécialement. [6,7]

III.1. Densité de charge électronique :

La densité de charge joue un rôle primordial, car elle fournit une bonne description des différentes propriétés chimiques et physiques des solides tels que les liaisons qui s'établissent entre les constituants d'un matériau et aussi l'origine des propriétés optiques, électroniques et structurales. Elle est traitée dans l'espace directe et donne ainsi une idée réelle des distributions de charge électroniques. Il est donc évident de dire que la densité de charge électronique représente un pont entre le concept des fonctions d'états dans l'espace de Hilbert et le modèle physique de la matière dans l'espace direct.

Il est aisé de constater que les théoriciens ont réalisé de grandes avancées dans les travaux basés sur l'étude des propriétés électroniques, optiques ainsi que sur les propriétés de liaison.

La plus part des méthodes de calculs implique le calcul des fonctions d'ondes dans l'espace réciproque (espace des K) ; en particulier, la réponse spectrale des fonctions nécessite souvent une détermination des courbes de dispersions comme $E_n(k)$. D'un autre côté, les chimistes se sont concentrés sur les propriétés de liaison en utilisant l'espace réel où on décrit les liaisons et les orbitales ainsi que les longueurs de liaisons empiriques.

La probabilité de la fonction des électrons se trouvant dans une région d'espace de volume $d\Omega$ est donnée par : $\int |\Psi_{n,k}(r)|^2 d\Omega$ pour des électrons dans la bande de valence n, et dans l'état k. Lorsque plusieurs états k sont considérés, la distribution de charge appelée aussi densité, peut être décrite, en particulier la densité de charge de la bande de valence est donnée par :

$$\rho_n(\mathbf{r}) = \sum_k e |\Psi_{n,k}(\mathbf{r})|^2 \quad (I-1)$$

La sommation porte sur tous les états de la zone de Brillouin pour une bande donnée n. Pour obtenir une convergence pour $|\Psi_{n,k}(r)|^2$, il est nécessaire d'étendre les ondes planes et d'évaluer les fonctions dans une partie de la zone de Brillouin contenant plus de points.

La densité de charge totale est obtenue en sommant les densités de charge de toutes les bandes de valence tel que :

$$\rho(\mathbf{r}) = \sum_n \rho_n(\mathbf{r}) \quad (I-2)$$

La somme porte sur toutes les bandes occupées.

La fonction $\rho_n(\mathbf{r})$ donne des informations sur les propriétés de chaque bande individuellement, elle peut servir alors à connaître le rôle de chaque bande dans la nature de la liaison au sein du matériau. Cependant, la densité totale $\rho(\mathbf{r})$ représente d'une façon directe la distribution électronique.

En général, le rapport entre les liaisons et le tracé de $\rho(\mathbf{r})$ est presque directement évident. La fonction de la densité de charge étant tridimensionnelle, pour une meilleure visibilité, il convient de tracer $\rho(\mathbf{r})$ dans un plan et suivant une direction donnée. Le plan choisi doit contenir les atomes et la liaison.

III.2. Liaisons chimiques : [8]

Il y a différentes liaisons chimiques en fonction de la forme du nuage électronique et de leurs niveaux d'énergie. Dans les solides, on distingue quatre types de liaisons : covalente, ionique, métallique, et la liaison de Van Der Waals.

III.2.1. Liaison covalente :

C'est une liaison chimique dans laquelle chacun des atomes liés met en commun un ou plusieurs électrons d'une de ses couches externes afin de former un doublet d'électrons liant les deux atomes (Fig I.1). C'est une des forces qui produit l'attraction mutuelle entre atomes.

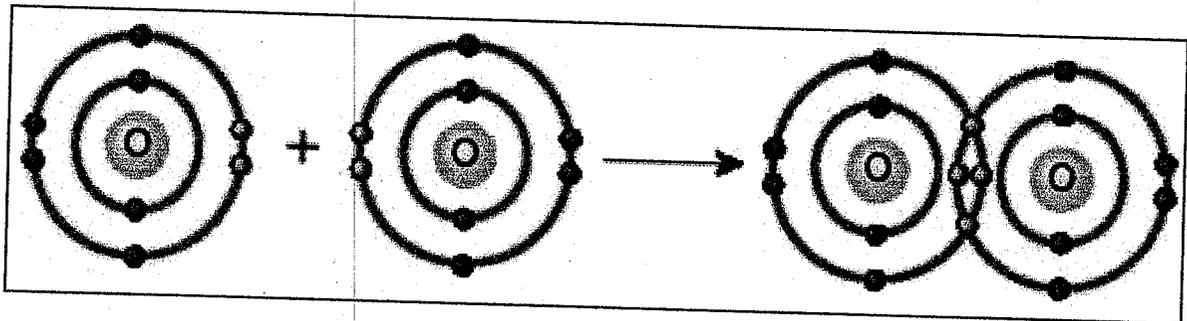


Fig I.1 : Schéma illustrant le partage des électrons.

La liaison covalente implique généralement, le partage équitable d'une seule paire d'électrons, appelé doublet liant; chaque atome fournissant un électron ; la paire d'électrons étant délocalisée dans l'espace se situant entre les deux noyaux atomiques.

III.2.2. Liaison ionique :

On appelle liaison ionique le lien causé par le transfert d'un ou plusieurs électrons de valence (de la couche électronique externe) entre un atome (donneur) et un autre atome (receveur). La différence d'électronégativité des atomes présents dans cette liaison est supérieure à 1,6 (cette limite est conventionnelle).

La liaison ionique est plus fréquente lorsqu'un ou chacun des deux atomes sont des métaux. En effet, un ou plusieurs électrons sont simplement transférés d'un atome à l'autre faisant qu'un des atomes devienne un ion positif et l'autre un ion négatif (Fig I.2).

Les liaisons ioniques se caractérisent par le fait que deux atomes ne partagent pas mais échangent des électrons. Ces échanges conduisent à la perte d'un ou de plusieurs électrons pour l'un des deux atomes et pour le gain proportionnel d'électrons pour l'autre atome.

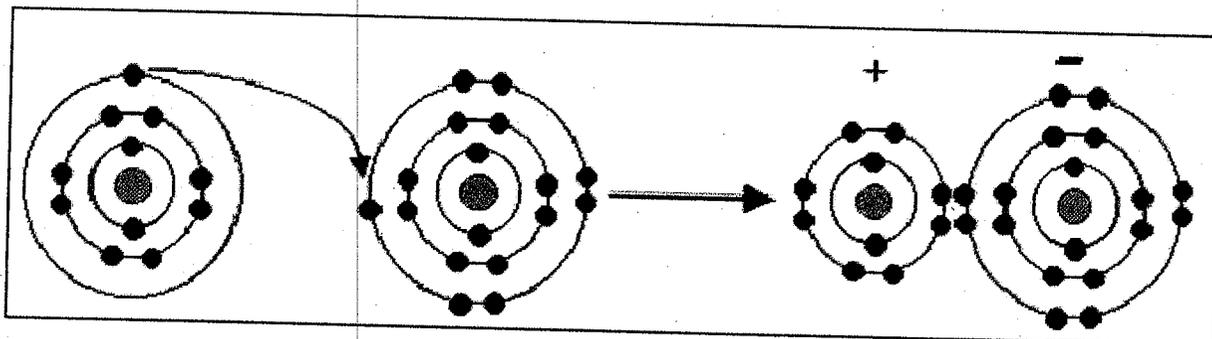


Fig I.2: le transfert d'électron d'un atome à l'autre.

III.2.3. La liaison métallique :

Ce type de liaison concerne un très grand nombre d'atomes (typiquement plusieurs millions voire plus). Ces atomes mettent en commun un ou plusieurs électrons, appelés « électrons libres », ce sont ces électrons qui permettent la conduction électrique.

Une liaison métallique n'existe que dans un solide (ou à l'état liquide) tout comme la liaison ionique. Dans la liaison métallique, les électrons sont mobiles dans tout un réseau alors que dans une liaison ionique, les charges sont presque statiques.

III.2.4. Liaison de Van der Waals :

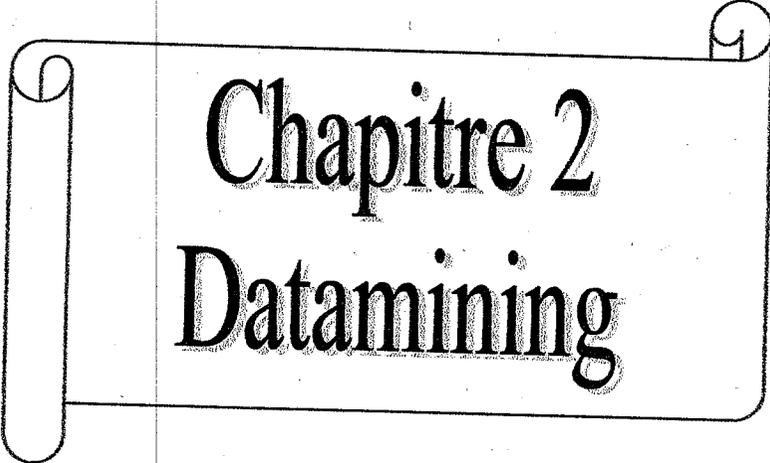
C'est une interaction de faible intensité entre atomes, molécules, ou une molécule et un cristal. Elle est due aux interactions entre les moments dipolaires électriques des deux atomes mis en jeu. Aucun électron n'est mis en commun entre les deux atomes.

Dans cette liaison, la paire d'électrons se déplace. Les deux atomes vont donc porter en alternance et de façon transitoire, une charge positive et une charge négative.

En fin la densité de charge électronique constitue un moyen topologique pour identifier la nature de la liaison. Jusqu'à ce jour, toute l'information que nous pouvons déduire à partir de la connaissance de la densité de charge est qualitative.

Références :

- [1] J. Philibert, A. Vignes, Y. Bréchet, P. Combrade, *Métallurgie — Du minerai au matériau*, Dunod, 2002, p. 309
- [2] *Intermetallic Compounds, Vol. 3: Progress*, edited by J. H. Westbrook and R.L.Fleischer, John Wiley & Sons, Chichester 2002, 1086 pages.
- [3] *Intermetallic Compounds — Principles and Practice*, edited by J. H. Westbrook and R. L. Fleischer, John Wiley & Sons, Chichester 1995, 2 volumes.
- [4] G. Sauthoff, *Intermetallics*, Wiley-VCH, Weinheim 1995, 165 pages.
- [5] F. Moret, R. Baccino, P. Martel et L. Guetaz
Propriétés et applications des alliages intermétalliques B2-FeAl
CEA/CEREM, DEM/Section de Génie des Matériaux CENG, 38054 Grenoble cedex 9,
France
- [6] P. Hohenberg and W. Khon, *Phys. Rev.* B136, 864 (1964).
- [7] W. Khon and L.J. Sham, *Phys. Rev.* A140, 1133 (1965).
- [8] Linus Pauling, "The nature of the chemical bond. Application of results obtained from the quantum mechanics and from a theory of paramagnetic susceptibility to the structure of molecules." February 17, 1931. *J. Am. Chem. Soc.* 53 (April 1931): 1367-1400.



Chapitre 2
Datamining

I. Introduction :

L'exploitation de données (Datamining), connue aussi sous le nom de fouille de données : est un nouvel axe de recherche qui a pour objet l'extraction du savoir à partir de grandes quantités de données.

Cette dernière est un processus d'analyse dont l'approche est différente de celle utilisée en statistique, Il est nécessaire que l'on établisse une hypothèse de départ qu'il s'agira de vérifier.

Le Datamining se propose de découvrir automatiquement des modèles intéressants à partir d'un ensemble de données de grande taille. Ceci est possible en transformant en informations et en connaissances les données stockées au préalable dans des bases de données structurées.

II. Datamining :

II.1. Définition :

C'est l'ensemble des techniques et de méthodes du domaine des statistiques, des mathématiques et de l'informatique permettant l'extraction, à partir d'un important volume de données brutes, des connaissances originales auparavant inconnues. Il s'agit de « *fouilles* » visant à découvrir « *de l'information cachée* » que des données renferment et que l'on découvre à la recherche d'association, de corrélation, de relations ou régularités.

II.2. Description d'un projet de Datamining :

Un projet d'analyse de données est représenté par ces trois étapes :

✓ **Obtention de données significatives** : La réalisation d'un système de collecte ou de fabrication de données, mais lorsque celles-ci n'existent pas, plus souvent, il s'agit simplement d'élaborer quelque requêtes dans des bases de données et de mettre en forme ces données.

✓ **Analyse des données** : elle met en jeu le cœur du datamining, elle permet de déterminer des relations, de réaliser des prédictions, de faire émerger une ou plusieurs classifications, et cela grâce aux multiples techniques du datamining telles :

- Etude des covariances.
- Régression multi variable.
- Clustering (maximum de vraisemblance, analyse discriminante, hiérarchique).

✓ *Exploitation des données* : c'est l'exploitation des résultats.

II.3. Principe de datamining:

Le Datamining est un processus cyclique (pas linéaire) d'extraction de connaissances, elle doit permettre de capitaliser les connaissances acquises sous forme de connaissances explicites en structurant suivant les étapes principales suivantes :

- formaliser le problème.
- Accéder aux données appropriées.
- préparer et construire la base de données.
- Modéliser: utiliser des programmes adéquats.
- évaluation: extraction du savoir.
- Déployer: exploiter et échanger la connaissance découverte avec d'autres applications.

La figure (II.1) montre le principe de déroulement du Datamining.

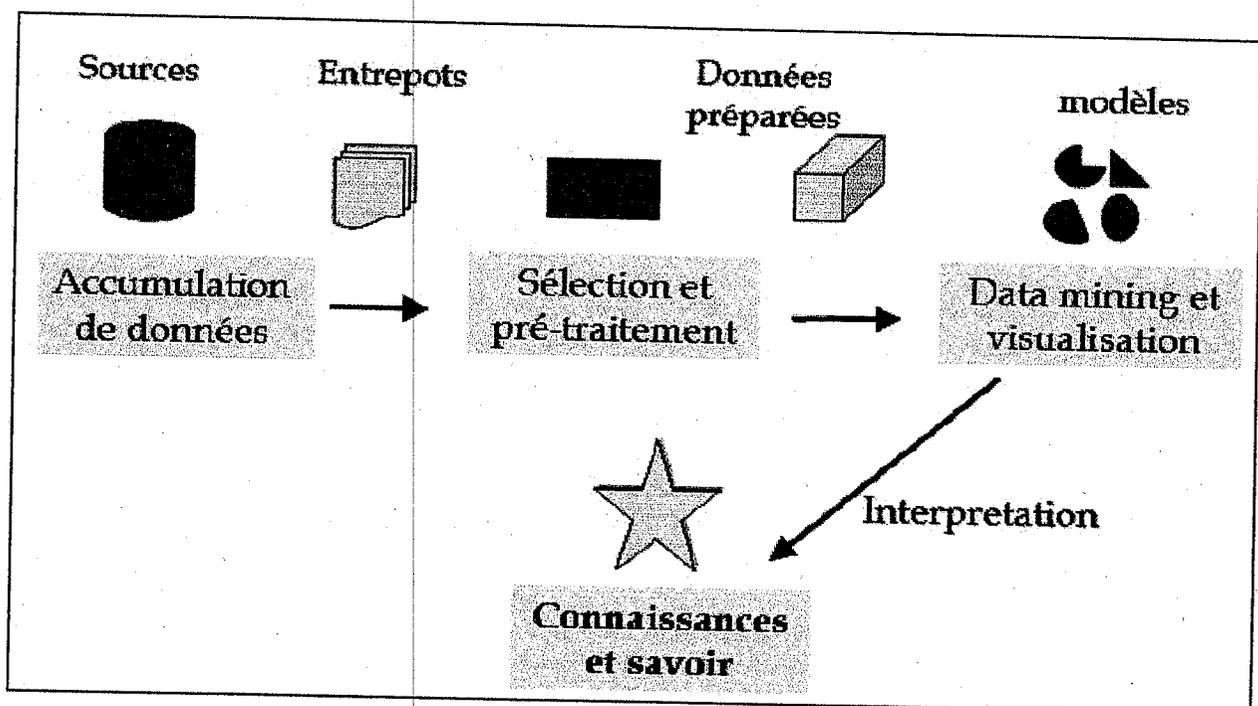


Figure II.1 : Les différentes étapes de déroulement du Datamining.

II.4. Méthodes statistiques :

Le datamining se propose d'utiliser un ensemble de techniques issues de discipline scientifiques diverses (statistiques, intelligence artificielle, base de données) pour construire des modèles à partir de données, et extraire de ces données un maximum de connaissances utiles. Dans notre travail on a utilisé deux de ces méthodes, l'ACP (analyse en composante principale) et la PLS (Partial Least Squares régression).

II.4.1. L'analyse en composantes principales :

II.4.1.1. Introduction :

L'Analyse en Composantes principales (ACP) fait partie du groupe des méthodes descriptives multidimensionnelles appelées méthodes factorielles. Ces méthodes qui sont apparues au début des années 30 ont été surtout développées en France dans les années 60, en particulier par Jean-Paul Benzécri qui a beaucoup exploité les aspects géométriques et les représentations graphiques. Dans la mesure où ce sont des méthodes descriptives, elles ne s'appuient pas sur un modèle probabiliste, mais elles dépendent d'un modèle géométrique. L'ACP propose, à partir d'un tableau rectangulaire de données comportant les valeurs de " p " variables quantitatives pour " n " unités (appelées aussi individus), des représentations géométriques de ces unités et de ces variables. Ces données peuvent être issues d'une procédure d'échantillonnage ou bien de l'observation d'une population toute entière. [1,2]

✓ *Qu'est-ce que l'ACP ?*

L'analyse en composantes principales ("*Principal Component Analysis*"), en abrégé ACP, est une technique descriptive permettant d'étudier les relations qui existent entre les variables, sans tenir compte, a priori, d'une quelconque structure [3-4], son utilisation pour l'exploitation des données remonte au début du siècle dernier. Elle est principalement issue des travaux de psychologues américains [Pearson 1901], [Spearman 1904] [5-6].

Le but de l'ACP est d'identifier la structure de dépendance entre des observations multi variables, afin d'obtenir une description ou une représentation compacte de ces dernières.

II.4.1.2. Tableau de données : [4]

Les données sont les mesures effectuées sur " n " unités $\{u_1, u_2, \dots, u_i, \dots, u_n\}$. Les " p " variables quantitatives qui représentent ces mesures sont $\{v_1, v_2, \dots, v_j, \dots, v_p\}$.

Le tableau des données brutes à partir du quel on va faire l'analyse est noté X et à la forme suivante :

$$X = \begin{array}{c} u_1 \\ u_2 \\ \cdot \\ u_i \\ \cdot \\ u_n \end{array} \begin{array}{cccccc} v_1 & v_2 & \dots & v_j & \dots & v_p \\ \left[\begin{array}{cccccc} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{array} \right. \end{array}$$

On peut représenter chaque unité par le vecteur de ses mesures sur les " p " variables :

$$U_i = [x_{i1} \quad x_{i2} \quad \dots \quad x_{ij} \quad \dots \quad x_{ip}] \quad \text{ce qui donne} \quad U_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \cdot \\ x_{ij} \\ \cdot \\ x_{ip} \end{bmatrix}$$

Alors U_i est un vecteur de \mathbb{R}^p .

De façon analogue, on peut représenter chaque variable par un vecteur de \mathbb{R}^n dont les composantes sont les valeurs de la variable pour les " n " unités :

$$V_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \cdot \\ x_{ij} \\ \cdot \\ x_{nj} \end{bmatrix}$$

Pour avoir une image de l'ensemble des unités, on se place dans un espace affine en choisissant comme origine un vecteur particulier de \mathbb{R}^p , par exemple le vecteur

dont toutes les coordonnées sont nulles. Alors, chaque unité sera représentée par un point dans cet espace. L'ensemble des points qui représentent les unités est appelé traditionnellement "nuage des individus".

En faisant de même dans \mathbb{R}^n , chaque variable pourra être représentée par un point de l'espace affine correspondant. L'ensemble des points qui représentent les variables est appelé "nuage des variables".

On constate, que ces espaces étant de dimension supérieure en général à 2 et même 3, on ne peut visualiser ces représentations. L'idée générale des méthodes factorielles est de trouver un système d'axes et de plans tels que les projections de ces nuages de points sur ces axes et ces plans permettent de reconstituer les positions des points les uns par rapport aux autres, c'est-à-dire avoir des images les moins déformées possible.

II.4.1.3. Choix de l'origine :

Le point o correspondant au vecteur de coordonnées toutes nulles, n'est pas forcément une origine satisfaisante, car si les coordonnées des points du nuage des individus sont grandes, le nuage est éloigné de cette origine. Il apparaît plus judicieux de choisir une origine liée au nuage lui-même : le centre de gravité du nuage G (Fig. II.2). [4,7]

Pour l'ACP on choisit de donner le même poids à tous les individus. Le centre de gravité G du nuage des individus est alors le point dont les coordonnées sont les valeurs moyennes des variables. [4]

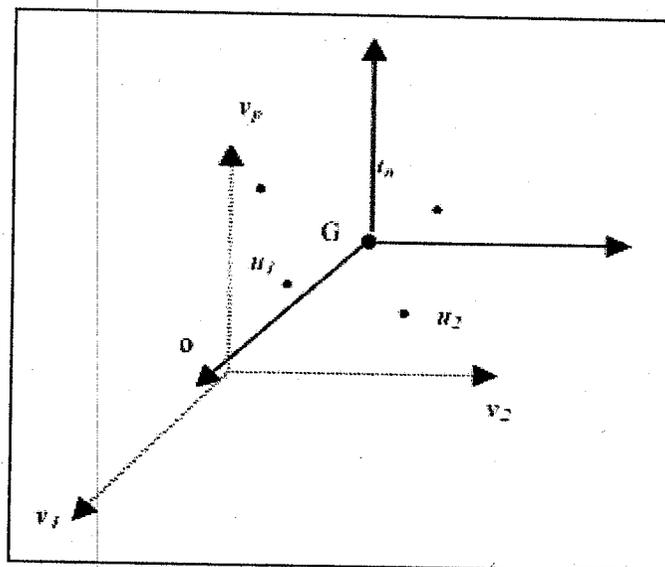


Figure II.2: choix d'origine

II.4.1.4. Formulation mathématique de l'ACP : [5, 6, 8, 9]

L'ACP est un simple changement de repère : sa grande force est de pouvoir ainsi recourir à l'algèbre linéaire comme outil mathématique principal, et de recevoir une interprétation géométrique simple.

Etant donné un ensemble d'observations décrites par des variables exclusivement numériques $\{x_1, x_2, \dots, x_p\}$, en plaçant en ligne les individus et en colonnes les variables. Soit X ce tableau de données.

$$X = \text{individus} \dots \begin{matrix} & \begin{matrix} 1 & \dots & p \end{matrix} \\ \begin{matrix} 1 \\ \vdots \\ n \end{matrix} & \begin{bmatrix} x_1^1 & \dots & x_1^p \\ \vdots & & \vdots \\ x_n^1 & \dots & x_n^p \end{bmatrix} \end{matrix}$$

Par convention, on placera en exposant ce qui se rapporte aux individus : le premier individu est donc x^1 , et en indice ce qui se rapporte aux variables : x_1 désigne la première variable. x_1^1 désigne la valeur numérique prise par la première variable, pour le premier individu. Et notre tableau de données regroupe toutes les valeurs prises par tous les individus (de 1 à n) par p variables, l'ACP a pour objectif de décrire ce même ensemble de données par de nouvelles variables en nombre réduit. Ces nouvelles variables seront des combinaisons linéaires des variables originales, et porteront le nom de Composantes Principales (CP).

En général, la réduction du nombre de variables utilisées pour décrire un ensemble de données provoque une perte d'information. L'ACP procède de façon à ce que cette perte d'information soit la plus faible possible. L'analyse en composantes principales peut donc être vue comme une technique de réduction de la dimensionnalité.

II.4.1.5. Principe de l'analyse en composante principale : [9]

L'idée de base de l'ACP est de réduire la dimension de la matrice des données, en retenant le plus possible les variations présentes dans le jeu des données de départ.

Cette réduction ne sera possible que si les variables initiales ne sont pas indépendantes et ont des coefficients de corrélation entre elles non nuls. Ces variables initiales sont transformées en de nouvelles variables, appelés composantes principales. Elles sont obtenues par combinaisons linéaires des précédentes et sont ordonnées et non corrélées entre elles.

Du point de vue géométrique l'ACP consiste à effectuer une certaine rotation du repère des variables autour de leurs valeurs moyennes. Cette rotation transforme les n variables corrélées en " L " variables non corrélées. Notons que ce sont justement ces variables transformées que l'on a nommé les composantes principales.

II.4.2. La régression (PLS) :

II.4.2.1. Introduction :

Cette méthode fut proposée par Wold S., Albano C., Dunn III W.J., Esbensen K., Hellberg S., Johansson E. et Sjöström M. en 1983 [10]. Elle fut principalement développée par Svante Wold, qui a dirigé de nombreuses thèses portant sur le sujet. De plus, ce dernier, associé à Nouna Kettaneh-Wold et à plusieurs collaborateurs, ont développé le logiciel d'analyse des données SIMCA-P, logiciel centré sur la régression PLS.

La régression PLS (*Partial Least Squares regression*) 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 (le nombre de variables explicatives pouvant excéder très largement le nombre d'individus), ou encore lorsque les variables explicatives sont fortement corrélées entre elles.

✓ *Qu'est-ce que la régression PLS ?* [11]

La régression PLS est une technique récente qui généralise et combine les caractéristiques de l'analyse sur composantes principales et de la régression multiple. Elle est particulièrement utile quand on a besoin de prédire un ensemble de variables dépendantes à partir d'un ensemble très grand de variables explicatives (prédicteurs) qui peuvent être très fortement corrélées entre elles.

Quand les prédicteurs sont peu nombreux, non significativement colinéaires et ont une relation connue avec les réponses, alors la régression linéaire multiple est la meilleure méthode pour utiliser les données. Cependant si l'une de ces trois conditions n'est pas vérifiée, la régression linéaire multiple peut être inefficace et inappropriée.

La méthode PLS est donc une méthode pour construire des modèles de prédiction quand les facteurs sont nombreux et très colinéaires. Notons que cette méthode met l'accent sur la

prédiction de la réponse et pas nécessairement sur la mise en évidence d'une relation entre les variables.

Ceci signifie que la PLS n'est pas appropriée pour désigner les variables ayant un effet négligeable sur la réponse, mais quand le but est la prédiction et qu'il n'y a pas besoin de limiter le nombre de variables mesurées, la PLS est un outil très utile.

On distingue deux types de PLS :

- La régression PLS uni variée, ou « régression PLS1 » : la régression PLS1 ne porte que sur une seule variable expliquée.
- La régression PLS multi variée, appelée également « régression PLS2 » : la régression PLS il peut y avoir plusieurs variables expliquées.

✓ Établissement du modèle PLS1 simple :

Elle comprend le calcul de la matrice T , dont les éléments sont les scores et les colonnes les composantes, obtenue par l'application d'une relation de la forme :

$$T = X W \quad (II.1)$$

Avec : W : matrice des poids.

X : matrice des données prédictives centrées, comprenant n lignes et m colonnes.

Dans la régression PLS, le calcul des composantes T se fait en tenant compte de la variable à prédire y . Plus précisément, on cherche à effectuer une double modélisation correspondant aux deux relations :

$$X = TP + R \quad (II-2)$$

$$y = Tq + f \quad (II-3)$$

Avec :

R : matrice des résidus, associée à la prédiction de X

f : vecteur des résidus associé à la prédiction de y .

✓ Etablissement du modèle de La régression PLS2 :

La régression PLS2 s'applique dans le cas où plusieurs variables sont à prédire à partir de la même matrice de données prédictives X . Les données à prédire forment maintenant la matrice Y , comprenant n lignes et r colonnes.

Considérons les données initiales centrées X_0 et Y_0 . On cherche à calculer une première composante t_1 , combinaison linéaire des colonnes de X_0 qui soit à la fois représentative de X_0 et de Y_0 . Pour cela, on s'intéresse tout d'abord à deux combinaisons linéaires associées respectivement à X_0 et à Y_0 soit :

$$t_1 = X_0 w_1 \quad \text{et} \quad u_1 = Y_0 c_1 \quad (\text{II-4})$$

Avec w_1 et c_1 vecteurs de norme unitaire

Le critère à optimiser est une extension, dans le cas où Y est multivariée, de celui adopté pour la régression PLS1. On détermine w_1 et c_1 de manière à maximiser la covariance entre t_1 et u_1 .

On effectue à nouveau une double modélisation :

$$X_0 = t_1 p_1 + X_1 \quad \text{et} \quad Y_0 = t_1 q_1 + Y_1 \quad (\text{II-5})$$

Les éléments de q_1 sont les coefficients des régressions linéaires simples effectuées entre t_1 et chaque colonne de Y_0 .

II.4.2.2. Interprétation du modèle PLS :

Le modèle PLS, nous permet de former de nouvelles variables x , (t_k), qui sont des combinaisons linéaires des x et qui sont utilisées comme prédicateurs de Y . Ainsi PLS est basé sur un modèle linéaire.

Tous les paramètres, t , u , w , et q sont déterminés par un algorithme PLS. Les scores t et u contiennent des informations sur les objets et les corrélations/ non corrélations entre ces différents objets [12].

Les poids w et q donnent des informations sur la manière dont la quelle les différentes variables se combinent pour former des relations quantitatives entre X et Y , et par conséquent fournir une interprétation des scores t et u . Ces poids w expriment à la fois les corrélations positives entre X et Y et les corrélations nécessaires pour prédire Y à partir de X .

La régression PLS permet ainsi de conserver toutes les variables explicatives tout en obtenant une équation de régression cohérente.

II.4.2.3. Intérêt de la Régression PLS :

La régression PLS pour le modèle linéaire s'applique où l'on possède peu d'observations sur des variables très corrélées et en très grand nombre, c'est le cas pathologique de la régression linéaire, là où les méthodes habituelles ne fonctionnent pas (n étant très petit on ne peut pas tendre vers l'infini).

II.4.2.4. Conclusion :

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

Référence :

- [1] I.T. Jolliffe "Principal component analysis", Springer-Verlag, New York, 1986.
- [2] K.I. Diamantaras, S.Y. Kung, "Principal component neural networks. Theory and applications", John Wiley and Sons , 1996.
- [3] H. Hotelling, Journal of Educational Psychology, 24 (1933) 417- 441
"Analysis of a complex of statistical variables into principal components",
- [4] C. Duby, S. Robin, "Analyse en Composantes Principales", Département O.M.I.P.(2006).
- [5] Escoffier B., Pagès J. (1990) –«Analyses factorielles simples et multiples ». Dunod, Paris.
- [6] Escoffier B., Pagès J. (1997) – « Initiation aux traitements statistiques, méthodes », méthodologies. Dida Stat, PUR, Rennes.
- [7] Lindsey Smith, " A tutorial on principal component Analysis ", (2002).
- [8] Lebart L., Morineau A., Piron M. (1995) – Dunod, Paris
« Statistique exploratoire multidimensionnelle ».
- [9] Qin S. J. and Dunia R., "Determining the number of principal components for best reconstruction", Proc. of the 5th IFAC Symposium on Dynamics and Control of Process Systems, PP. 359-364, Corfu, Greece, 1988. [9] Saporta G. (1990) – « Probabilités, analyse des données et statistique ». Dunod, Paris
- [10] S. Wold, M. Sjostrom, L. Eriksson, "PLS-regression : a basic tool of chemometrics ", Chemometrics and Intelligent Laboratory Systems, 58 (2001) 109-130.
- [11] Juan Rosas Cédric Lacour, « Etude, classification par spectroscopie proche infra rouge (NIR) des produits cosmétiques », thèse de l'université de Brest soutenue 2007
- [12] A. Hoskuldsson, "Variable and subset selection in PLS regression", Chemometrics and Intelligent Laboratory Systems 55 (2001) 23–38.

Chapitre 3

Résultats et interprétations



I. Introduction :

Dans ce chapitre les calculs ont été faits avec la technique ACP (L'analyse en composantes principales) et la technique des moindres carrés partiels PLS (*Partial Least Squares*), en utilisant le logiciel XLSTAT décrit en annexe.

La ACP est un outil puissant pour l'analyse des matériaux et a été employé pour résoudre des problèmes en science des matériaux.

La ACP est efficace pour la découverte des informations de base qui peuvent être perdues quand on a une grande quantité de données à exploiter.

La régression PLS est une technique d'analyse et de prédiction relativement récente.

II. Résultats :**II.1. Définition de la base de données :**

Une base de données, usuellement abrégée en BD ou BDD, est un ensemble structuré et organisé permettant le stockage de grandes quantités d'informations afin d'en faciliter l'exploitation. [1]

II.2. Base de données (1):

Cette base contient les composés binaires et leurs propriétés de la densité électronique n_{ws} , le volume molaire V_M et le module de compressibilité B_1 .

Nombre	Alliages	n_{ws}	V_M	B_1	N
1	Al ₅₂ Ru ₄₈	4,36	8,71	165,8	37,99
2	Al ₅₀ Ru ₅₀	4,43	8,66	170,6	38,45
3	Al ₄₀ Ru ₆₀	4,79	8,49	195,4	40,74
4	Be ₁₇ Nb ₂	4,61	5,5	117,1	25,38
5	Be ₁₂ Nb	4,62	5,33	114,1	24,67
6	Be ₁₂ Ti	4,51	5,28	107,6	23,84
7	Be13Zr	4,39	5,44	105,1	23,92
8	CaAl ₂	1,95	12,54	47,7	24,47
9	Co ₂ Zr	4,25	8,64	156,3	36,74
10	Cr ₂ Nb	4,86	8,4	199,1	40,9
11	Mo _{48,7} Nb	4,98	10,04	249	50,02
12	MoSi ₂	3,98	11,18	177,5	44,55
13	AlAs	2,52	10,98	70	27,72
14	GaAs	2,33	11,79	64,2	27,51
15	GaSb	1,93	15,09	56,6	29,21
16	CdS	1,76	14,57	45,2	25,66
17	CdSe	1,62	15,28	40,2	24,78
18	CdTe	1,4	17,58	34,7	24,69
19	HgSe	1,59	15,79	42,2	25,18
20	HgTe	1,42	18,05	36,6	25,72

Tableau III.1 : Les valeurs de n_{ws} , B_1 et V_M pour les composés binaires [2]

❖ **Remarque :** La densité électronique calculée à partir de la formule suivante :

$$n_{ws} = N/V_M \quad \text{Tel que : } N : \text{ le nombre des électrons de valences [3]}$$

II.2.1.L'analyse en composantes principales (ACP):

L'analyse en composantes principales (ACP) nous permet de tracer les trois graphes : observations (scores plot) et variables (loadings plot) et Biplot. Le graphe des observations donne des informations sur les échantillons et traite la façon dont ils se comportent, le graphe des variables nous informe sur les relations qui existent entre les différentes propriétés, et enfin le graphe des biplot traite les relations existant entre les individus qui ont utilisés et leurs propriétés.

1. Choix des axes :

Pour un plan factoriel donné, on regarde le pourcentage de l'information du nuage initial retranscrite par le plan factoriel. Ainsi, un axe expliquant moins de 10% sera rarement intéressant, les axes sont rangés dans l'ordre décroissant, de telle manière que le premier plan factoriel - constitué par les deux premiers axes principaux soit toujours celui qui est le plus riche en renseignements sur les propriétés du nuage étudié. Dans notre exemple, les axes (PC1) et (PC2) sont les plus riches en information (fig. III.1).

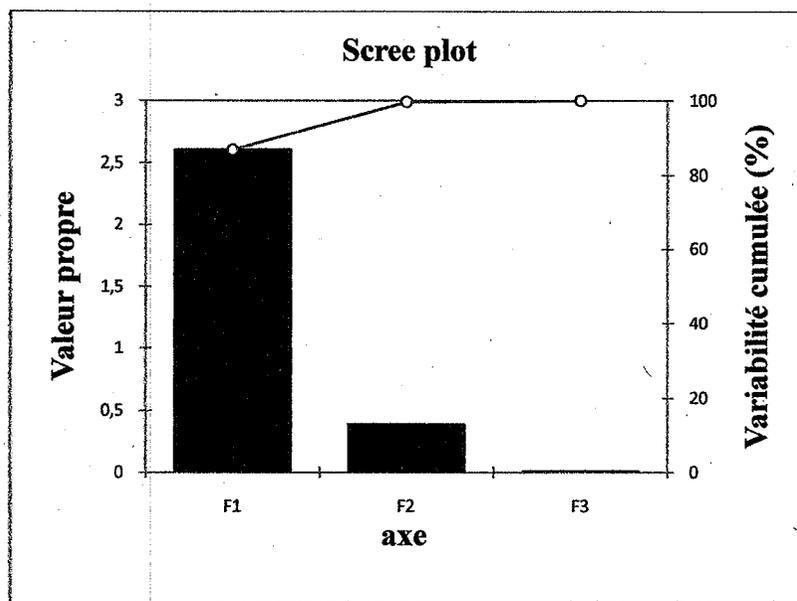


Figure III.1 : Représentation graphique des axes

2. Matrice de corrélation :

Le tableau III.2 représente la matrice de corrélation des composantes binaires utilisé dans la base de données :

Variables	n_{ws}	V_M	B_1
n_{ws}	1	-0.898	0,885
V_M	-0.898	1	-0.614
B_1	0,885	-0.614	1

Tableau III.2 : Matrice de corrélation

D'après ce tableau, on remarque bien que la densité électronique n_{ws} est corrélée à B_1 avec de valeur est égale à 0.885, et le volume molaire V_M est inversement corrélée à (B_1, n_{ws}) avec des valeurs est (-0.614, -0.898).

3. Interprétation des données :

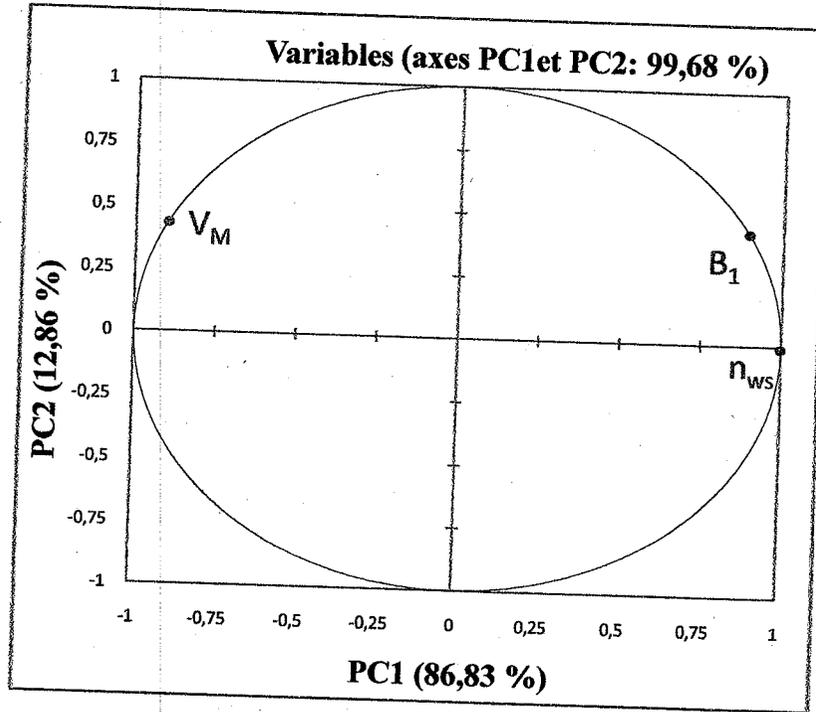


Figure III.2 : Graphe des variables (Loadings plot)

A partir de ce graphe nous pouvons constater :

✓ La disposition des différentes propriétés introduites dans la base de données dans le plan formé par les deux composantes principales PC1 et PC2. On remarque que PC1 capture 12.86 % de la variabilité de la base de données, et PC2 capture 86.83 %.

✓ La densité électronique n_{ws} et le module de compressibilité B_1 sont fortement corrélés entre eux, et les deux propriétés sont inversement corrélées avec le volume molaire V_M si l'on considère leur disposition par rapport à l'axe PC1.

L'origine de ce comportement est expliquée par le fait que B_1 représente la rigidité et la compressibilité de la structure. Autant sa valeur est élevée autant il est difficile de comprimer le matériau.

Les trois propriétés ont une valeur PC2 positive ce qui est indiquent des corrélations entre ces derniers.

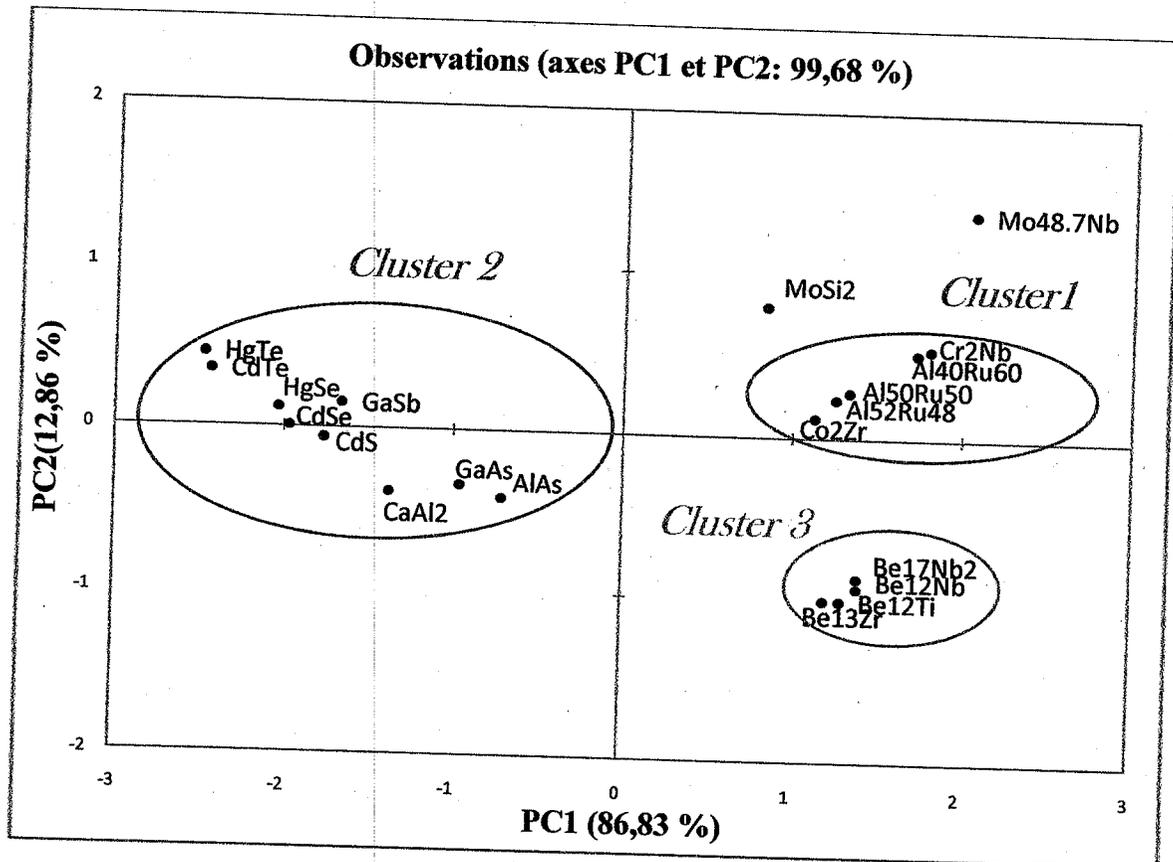


Figure III.3 : Graphe des observations (Score plot)

La figure III.3 présente le graphe des observations (Score plot).

Elle montre la répartition dans le plan formé par PC1 et PC2 des variables ou des différents matériaux.

La première constatation que nous pouvons faire est que les carbures se séparent en trois parties (clusters) bien distinctes, la 1^{er} et la 3^{ème} avec des valeurs de PC1 positives et l'autre avec des valeurs de PC1 négatives.

En analysant l'ensemble de ces trois clusters qui est représenté sur la figure III.3 comme la façon suivant :

Cluster 1 : On remarque bien que les binaires qui se trouvent dans ce cluster sont des intermétallique sont très corrélées, avec des valeurs PC1 positif.

D'après la BD nous pouvons voir que ces éléments (Cr₂Nb, Al₄₀Ru₆₀, Al₅₂Ru₄₈, Co₂Zr, Al₅₀Ru₅₀) à un module de compressibilité très grand avec des valeurs (199.1, 195.4, 165.8, 156.3, 170.6) respectivement.

Cluster 2 : Dans le 2^{ème} cluster, on ne trouve que les semi-conducteurs (HgTe, CdTe, HgSe, GaSb, CdSe, CdS, GaAs, AlAs, CaAl₂), ces composés sont corrélés, avec des valeurs PC1 négatif.

Cluster 3 : Le troisième cluster contient tous les composés à base de Beryllium (Be₁₇Nb₂, Be₁₂Nb, Be₁₂Ti, Be₁₃Zr) nous pouvons remarquer que ces dernières sont fortement corrélées, et ont un volume molaire très petit d'après la BD.

Enfin nous remarquons bien un composé très distinct, le Mo_{48.7}Nb qui se trouve très loin de l'origine.

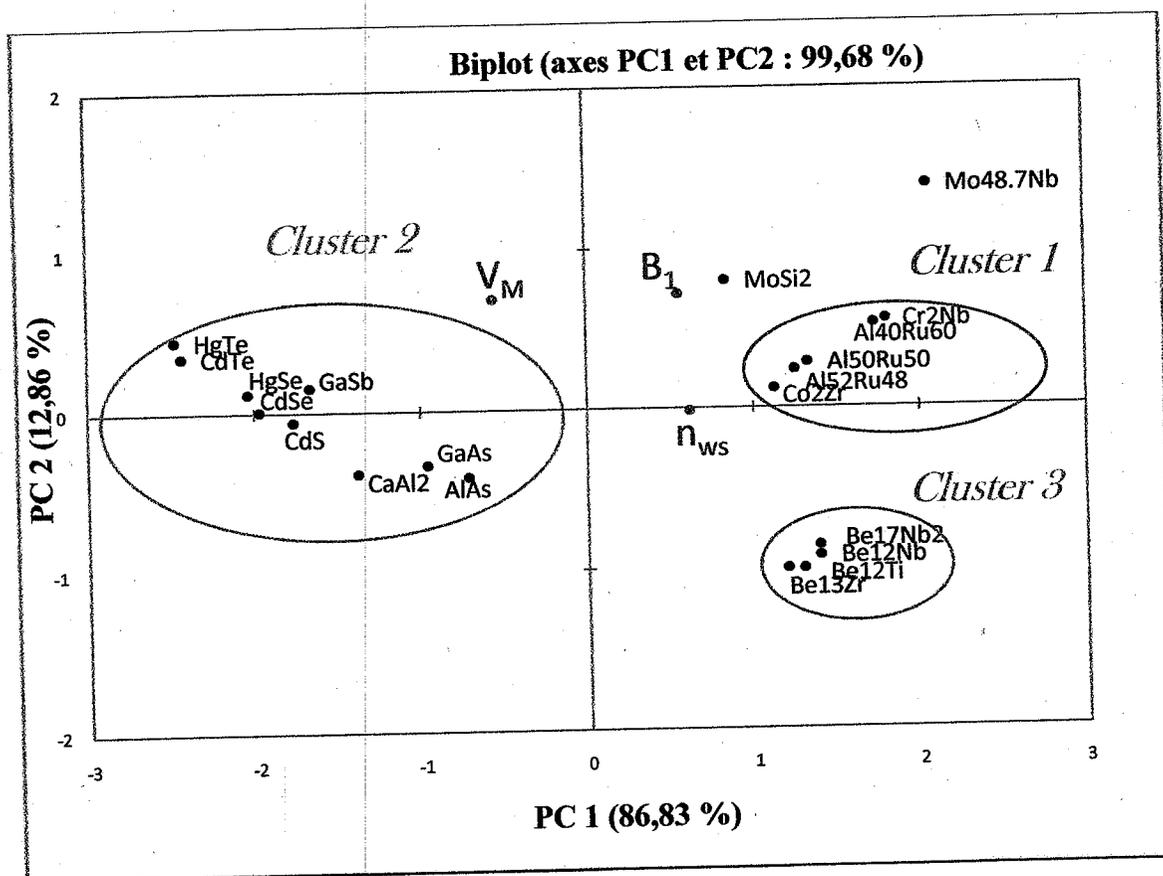


Figure III.4: Graphe superposant les échantillons et les variables (Biplot)

Dans ce graphe on y trouve la disposition des différents propriétés introduites dans la base données (1) et les éléments binaires que nous avons utiliser.

Comme le graphe précédent les individus sont séparé en trois cluster : Ceux avec un PC1 positif, que nous notons cluster 1, et cluster 3, ceux avec un PC1 négatif que nous notons cluster 2.

- La première contient les éléments ont un grand B_1 , et une grande densité électronique n_{ws} par contre un volume molaire V_M petit, on peut déduire que ces composantes sont des rigidités importantes.
- Nous pouvons aussi noter le comportement du $Mo_{48.7}Nb$ qui est pratiquement une valeur très grand de $B_1=249$, ceci signifie que leurs rigidités plus importantes.
- Les éléments qui se trouvent dans cluster 2 (HgTe, CdTe, HgSe, CdS, GaSb, CdSe, $CaAl_2$, GaAs, AlAs) ont un grand volume molaire (18.05,17.58, 15.28, 14.57, 15.09, 15.28, 12.54, 11.79, 10.98) respectivement, ça veut dire qu'il est facilement de comprimer ce matériaux.
- Cluster 3 contient des individus possèdent un PC2 négatif et un PC1 positif, ce qui est indique que ces élément contient un grand " n " et un petit (V_M, B_1), donc il est difficile de comprimer ce matériaux.

II.2.2-Résultats graphiques de la régression PLS :

1. Paramètres et équation du modèle :

Les paramètres du modèle correspondant à chacune des variables sont affichés sur le tableau suivant :

Variabes	B_1
Constante	-321,314
n_{ws}	80,846
V_M	15,005

Tableau III.3 : paramètres du modèle

✓ Equation du modèle :

$$B_1 = - 321,314437809139 + 80, 8461315585036 n_{ws} + 15, 0050162595656 V_M$$

2. prédictions à la variable « B_1 » :

Le modèle PLS permet de donner le graphe correspondant aux prédictions permettant d'identifier la qualité du modèle.

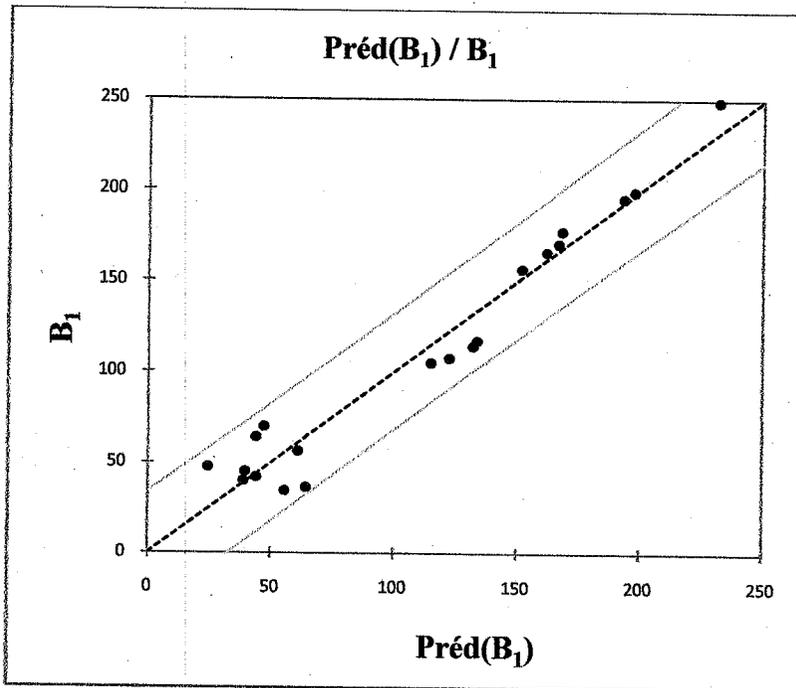
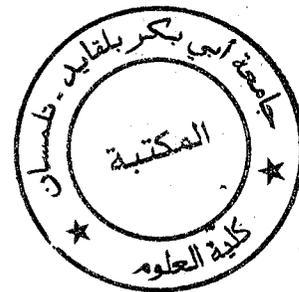


Figure III.5: Graphe de prédictions correspondant à la variable «B₁»

L'analyse du modèle correspondant à la première variable « B₁ » nous permet de conclure que l'ensemble des composantes binaires est significatif et influence le modèle.

Nous remarquons dans le graphe de prédictions que la variation est linéaire entre les valeurs de « B₁ » prédites en fonction de celle utilisée en entrée, qui est identique avec la relation du modèle précédent.



Observation	B_1	Préd(B_1)	Ecart-type sur la préd. (Moyenne)	Ecart-type sur la préd. (Observation)
Al ₅₂ Ru ₄₈	165,800	161,868	4,510	16,175
Al ₅₀ Ru ₅₀	170,600	166,777	4,688	16,225
Al ₄₀ Ru ₆₀	195,400	193,331	5,878	16,608
Be ₁₇ Nb ₂	117,100	133,914	6,148	16,705
Be ₁₂ Nb	114,100	132,171	6,343	16,778
Be ₁₂ Ti	107,600	122,528	6,669	16,904
Be ₁₃ Zr	105,100	115,227	6,766	16,943
CaAl ₂	47,700	24,498	6,167	16,713
Co ₂ Zr	156,300	151,925	4,223	16,097
Cr ₂ Nb	199,100	197,640	6,084	16,682
Mo _{48,7} Nb	249,000	231,950	8,848	17,877
MoSi ₂	177,500	168,209	5,593	16,509
AlAs	70,000	47,173	5,582	16,506
GaAs	64,200	43,966	5,379	16,438
GaSb	56,600	61,144	5,064	16,338
CdS	45,200	39,598	5,293	16,410
CdSe	40,200	38,933	5,512	16,482
CdTe	34,700	55,658	6,820	16,964
HgSe	42,200	44,160	5,625	16,520
HgTe	36,600	64,328	7,343	17,181

Tableau III.4 : Tableau des résultats prédictions

Cependant on nous referons au tableau des résultats prédit nous remarquons de forte variations entre la valeur réel et la valeur prédite ceci on forte l'idée que nous remarquons que l'ensemble des alliages ne se comportent pas de la même façon puisque l'ACP à démontré qu'il existe trois clusters comportement différents.

A cet effet il est impossible de trouver un model juste pour l'ensemble de ces alliages. Toutefois, il est peu être possible d'identifier un model pour chaque cluster.

A cet effet nous avons tracé différentes courbes de module de compressibilité en fonction de V_M et de n_{ws} .

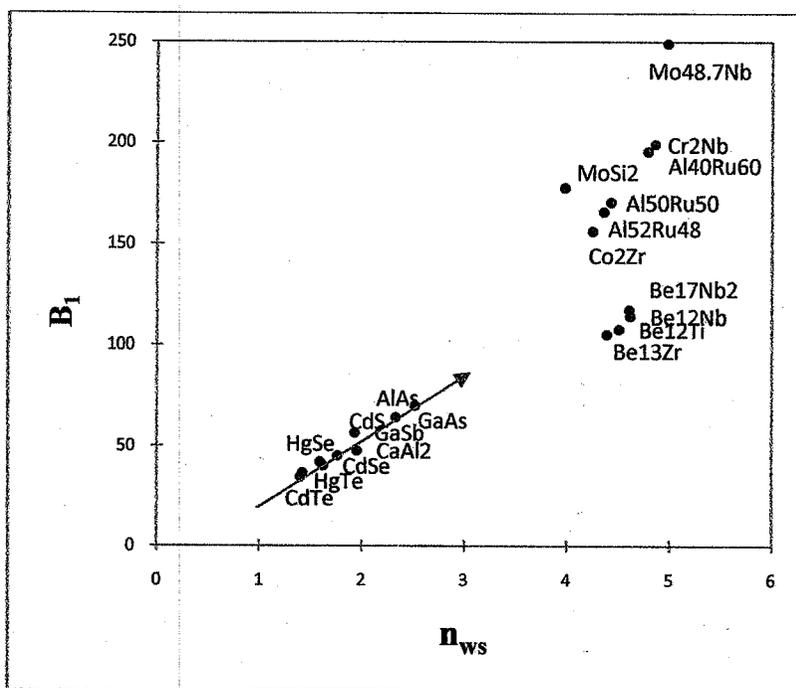


Figure III.6 : Courbe de B_1 en fonction de n_{ws}

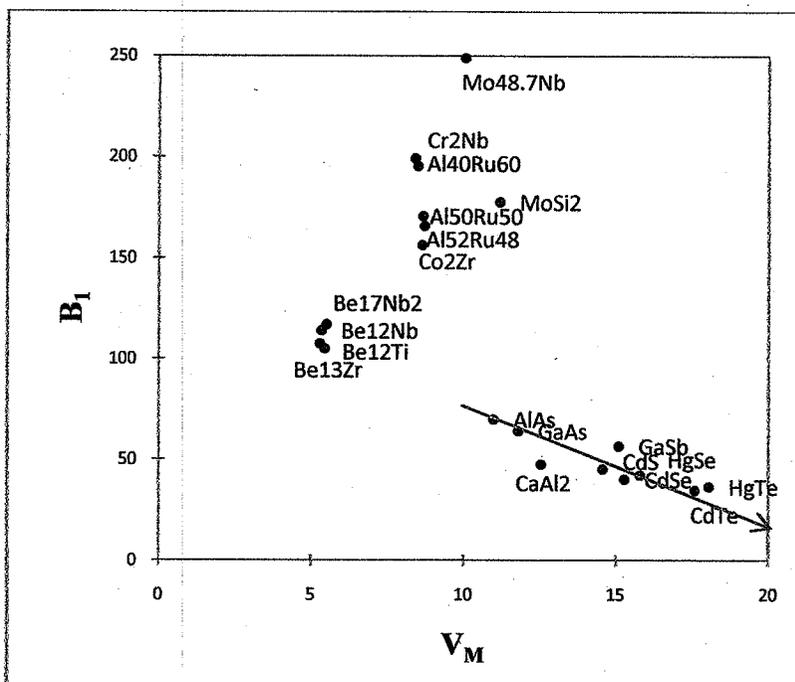


Figure III.7 : Courbe de B_1 en fonction de V_M

Les figures montre encore une fois l'existence de trois clusters et on remarque que c'est les mêmes que ceux me dit par l'ACP.

Ainsi il est impossible de tracer un model linéaire pour l'ensemble et que pour certains alliages tel que les semi conducteurs, la loi peu être linéaire.

II.3. Basse de données (2) :

Cette basse contient les même binaires de la première basse et leur propriétés suivantes :

n_{ws} : densité électronique.

V_M : volume molaire.

E_{coh} : l'énergie de cohésion.

ΔH : l'enthalpie de formation.

B_2 : module de compressibilité.

Nombre	Alliages	n_{ws}	V_M	ΔH	E_{coh}	B_2
1	Al ₅₂ Ru ₄₈	4,36	8,71	48	530	180,8
2	Al ₅₀ Ru ₅₀	4,43	8,66	48,4	536,9	186
3	Al ₄₀ Ru ₆₀	4,79	8,49	46,7	567,5	211,5
4	Be ₁₇ Nb ₂	4,61	5,5	14	377,2	121,4
5	Be ₁₂ Nb	4,62	5,33	10,2	361,8	17,3
6	Be ₁₂ Ti	4,51	5,28	12,1	343,5	111,4
7	Be ₁₃ Zr	4,39	5,44	17,4	357,6	110,2
8	CaAl ₂	1,95	12,54	13,7	291	50
9	Co ₂ Zr	4,25	8,64	56,9	540,6	172,8
10	Cr ₂ Nb	4,86	8,4	9,7	516,3	202,8
11	Mo _{48,7} Nb	4,98	10,04	8,4	701,4	252
12	MoSi ₂	3,98	11,18	39,9	556,5	190,2
13	AlAs	2,52	10,98	8,8	315	72
14	GaAs	2,33	11,79	13,2	291,3	67,1
15	GaSb	1,93	15,09	-0,4	267,6	56,5
16	CdS	1,76	14,57	169,1	362,6	66,3
17	CdSe	1,62	15,28	83,7	258,2	53,2
18	CdTe	1,4	17,58	36	197,5	41
19	HgSe	1,59	15,79	71,7	222,7	55,7
20	HgTe	1,42	18,05	26,9	164,9	42,6

Tableau III.5: Les valeurs de n_{ws} , B_2 , V_M , ΔH , E_{coh} pour les composé binaires [2]

II.3.1. L'analyse en composantes principales (ACP):**1. Matrice de corrélation :**

Le tableau III.6 représente la matrice de corrélation des composantes binaires utilisé dans la base de données :

Variabes	n_{ws}	V_M	B_2	ΔH	E_{coh}
n_{ws}	1	-0,898	0,757	-0,326	0,809
V_M	-0,898	1	-0,463	0,347	-0,557
B_2	0,757	-0,463	1	-0,083	0,934
ΔH	-0,326	0,347	-0,083	1	-0,026
E_{coh}	0,809	-0,557	0,934	-0,026	1

Tableau III.6 : Matrice de corrélation

D'après ce tableau, on remarque bien que le module de compressibilité B_2 et l'énergie de cohésion E_{coh} sont fortement corrélés avec de valeur 0.934, et les deux propriétés sont corrélés avec la densité électronique n_{ws} avec les valeurs (0.757,0.809) respectivement, et la corrélation inverse entre (B_2, n_{ws}, E_{coh}) et V_M avec des valeurs respectives (-0.463,-0.898, - 0.557).

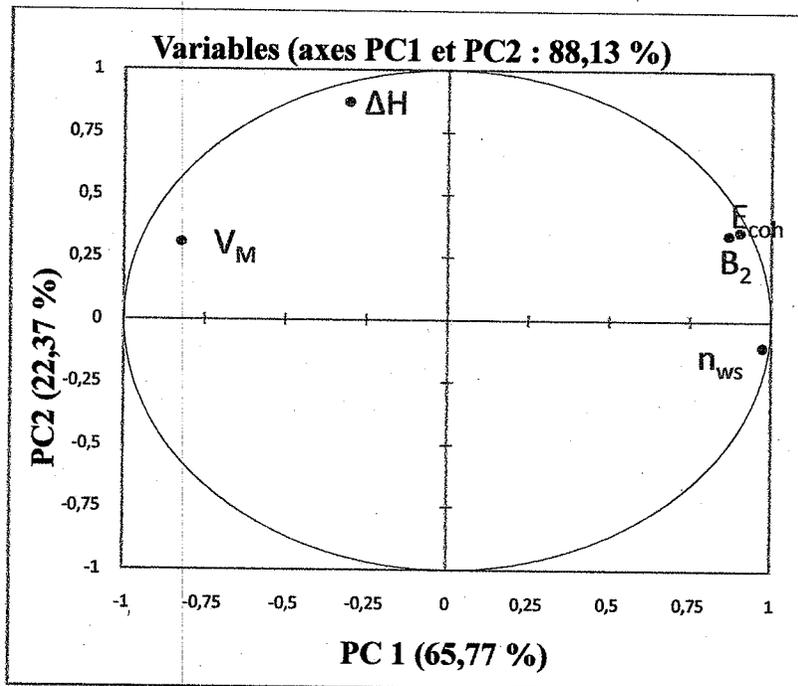
2. Interprétation des données :

Figure III.8 : Graphe des variables

Les axes principaux sont les axes PC1 capture 65.77 % et PC2 capture 22.37 %.

La figure III.7 représenté un utile pour interpréter la signification des axes.

La première remarque que nous avons remarquée une corrélation intéressante qui est la corrélation entre l'énergie de cohésion E_{coh} et le module de compressibilité B_2 , sont corrélées en même temps avec la densité électronique n_{ws} par rapport à l'axe PC1.

Dans notre cas l'axe PC1 est clairement lié aux V_M ce dernier est inversement corrélées avec les propriétés B_2 , n_{ws} et E_{coh} ce résultat est prévue parce que lorsque le module de compressibilité B_2 augmente le volume molaire diminue.

Nous savons que l'énergie de cohésion est fortement dépendante de la densité électronique, plus la densité est forte plus l'énergie est important.

On peut voir aussi dans le même graphe que l'axe PC2 représente l'augmentation de l'enthalpie de formation ΔH dans son côté positif, de façon contraire ΔH diminue dans le côté négatif de PC2.

Nous remarquons aussi que l'enthalpie de formation corrélée à la densité, la formation de ces alliages dépend finalement de ceux densité.

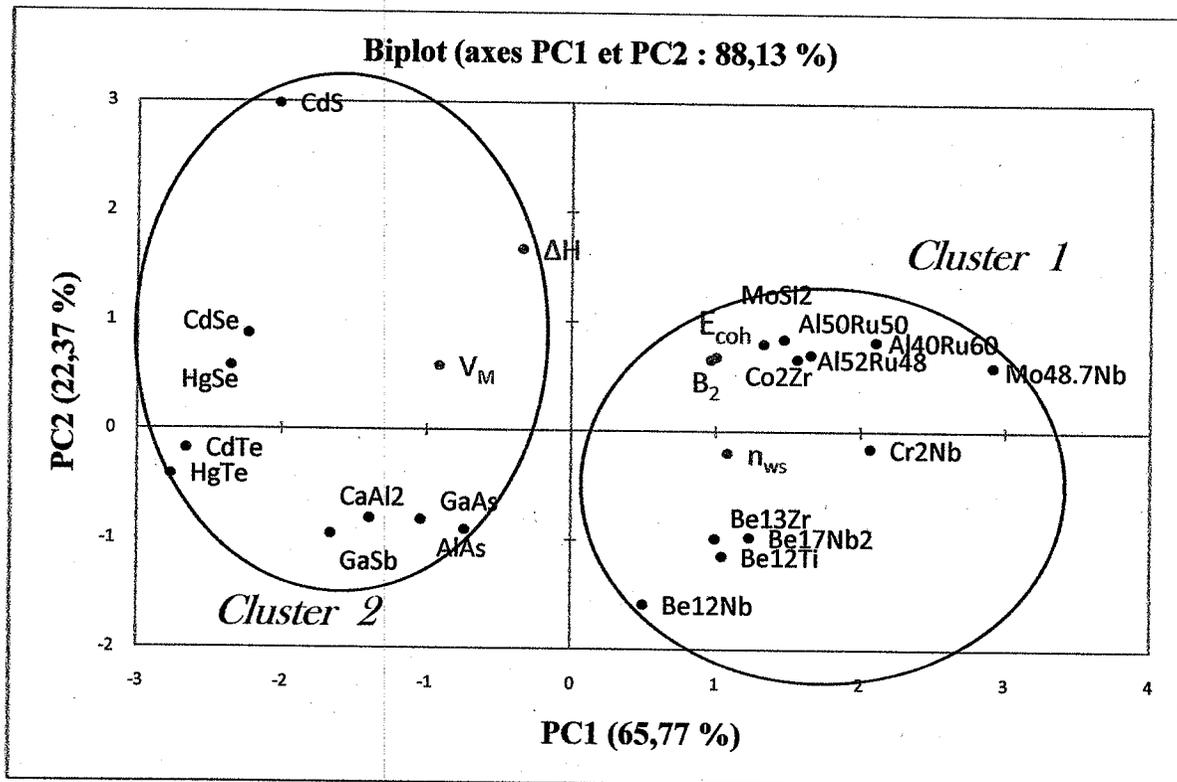


Figure III.9 : Graphe superposant les échantillons et les variables

En analysant la figure III.9, nous constatons deux groupes. L'un avec un PC1 positif, que nous notons cluster 1, et l'autre avec un PC1 négatif que nous notons cluster 2.

Dans un premier aperçu, il s'avère que dans le premier cluster, on ne trouve des composantes binaires fortement corrélées caractérisé par des trois propriétés très grande (E_{coh} , B_2 et n_{ws}), le deuxième c'est l'ensemble des binaires corrélées caractériser par V_M très grand.

Nous avons voir aussi que l'enthalpie de formation ΔH soit liée avec l'axe PC2, donc il y'a aucun sens.

Ce pendant il est impossible de trouver un modèle regroupant l'ensemble des alliages du faible que certains sont des semi-conducteurs (HgTe, CdTe, HgSe, CdS, GaSb, CdSe, CaAl₂, GaAs, AlAs), et d'autre des intermétalliques (Cr_2Nb , Al₄₀Ru₆₀, Al₅₂Ru₄₈, Co₂Zr, Al₅₀Ru₅₀, Be₁₇Nb₂, Be₁₂Nb, Be₁₂Ti, Be₁₃Zr) donc la nature de leurs liaisons ne sont pas les mêmes, et nous remarquons cela dans le graphe de l'ACP.

II.3.2. Résultats graphiques de la régression PLS :

1. Paramètres et équation du modèle :

Les paramètres du modèle correspondant à chacune des variables sont affichés sur le tableau suivant :

Variable	B ₂
Constante	-152,092
n _{ws}	15,200
V _M	5,331
ΔH	-0,129
E _{coh}	0,416

Tableau III.7 : paramètres du modèle

✓ Equation du modèle :

$$B_2 = -152,091525107306 + 15,2002648444432 n_{ws} + 5,33131941938124 V_M - 0,128837564803109 \Delta H + 0,416329981307281 E_{coh}$$

2. Prédications à la variable «B₂»:

Le modèle PLS permet de donner le graphe correspondant aux prédictions permettant d'identifier la qualité du modèle.

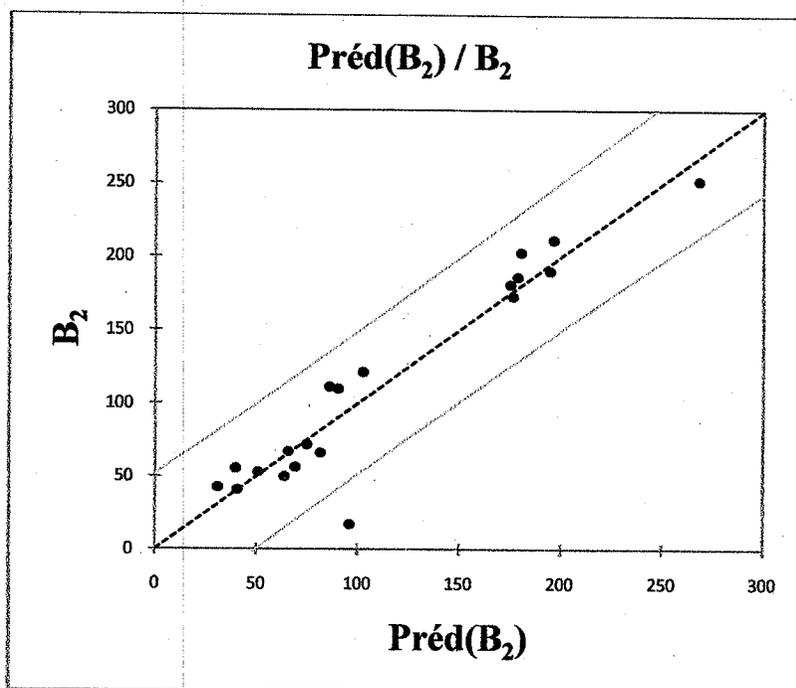


Figure III.10 : Graphe de prédictions correspondant à la variable «B₂»

L'analyse du modèle correspondant à la deuxième variable « B_2 » nous permet de conclure que les binaires est significatif et influencé le modèle .Nous remarquons dans le graphe des prédictions que la variation est linéaire entre les valeurs de « B_2 » prédites en fonction de celle utilisée en entrée.

Observation	B_2	Préd(B_2)	Ecart-type sur la préd. (Moyenne)	Ecart-type sur la préd. (Observation)
Al ₅₂ Ru ₄₈	180,800	175,088	7,950	25,840
Al ₅₀ Ru ₅₀	186,000	178,707	8,167	25,908
Al ₄₀ Ru ₆₀	211,500	196,231	9,131	26,228
Be ₁₇ Nb ₂	121,400	102,540	10,752	26,835
Be ₁₂ Nb	17,300	95,864	11,375	27,090
Be ₁₂ Ti	111,400	86,061	11,923	27,325
Be ₁₃ Zr	110,200	90,278	11,298	27,058
CaAl ₂	50,000	63,791	8,132	25,897
Co ₂ Zr	172,800	176,309	8,535	26,026
Cr ₂ Nb	202,800	180,266	8,643	26,061
Mo _{48,7} Nb	252,000	268,064	16,326	29,514
MoSi ₂	190,200	194,557	9,836	26,481
AlAs	72,000	74,761	7,603	25,736
GaAs	67,100	65,758	7,757	25,781
GaSb	56,500	69,156	11,283	27,052
CdS	66,300	81,513	20,165	31,798
CdSe	53,200	50,708	10,076	26,571
CdTe	41,000	40,500	11,021	26,944
HgSe	55,700	39,737	9,713	26,436
HgTe	42,600	30,910	11,989	27,354

Tableau III.8 : Valeurs prédites des binaires correspondent

D'après de ce tableau des résultats prédit nous remarquons de forte variations entre la valeur réel et la valeur prédite ceci signifier que l'ensemble des alliages ne se comportent pas de la même façon.

Puisque l'ACP sépare les semi-conducteurs des intermétalliques par rapport à l'axe PC1, de cette façon nous ne pouvons pas trouver un model unique.

Et cela nous pouvons le confirmer à tracer les graphes de B_2 en fonction de E et de ΔH .

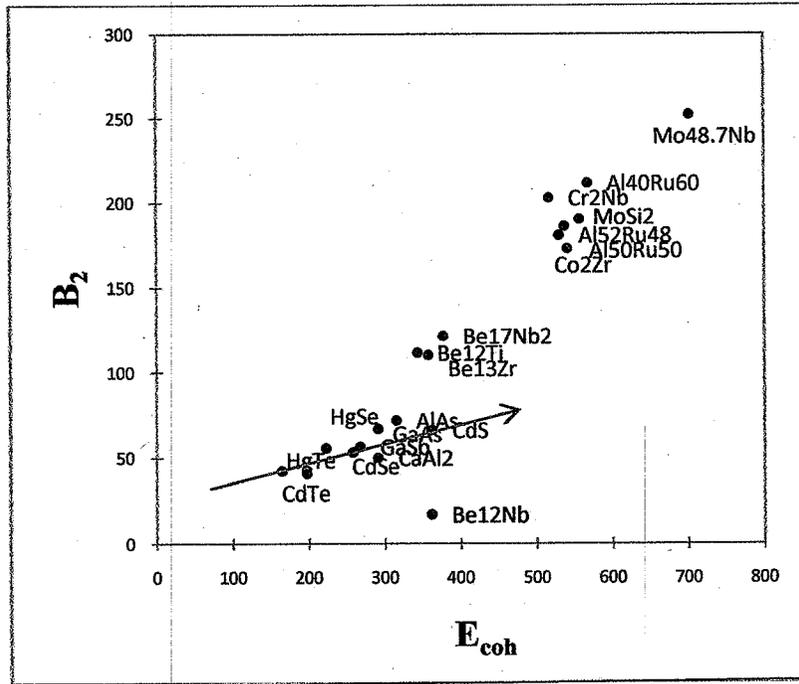


Figure III.11 : Courbe de B_2 en fonction de E_{coh}

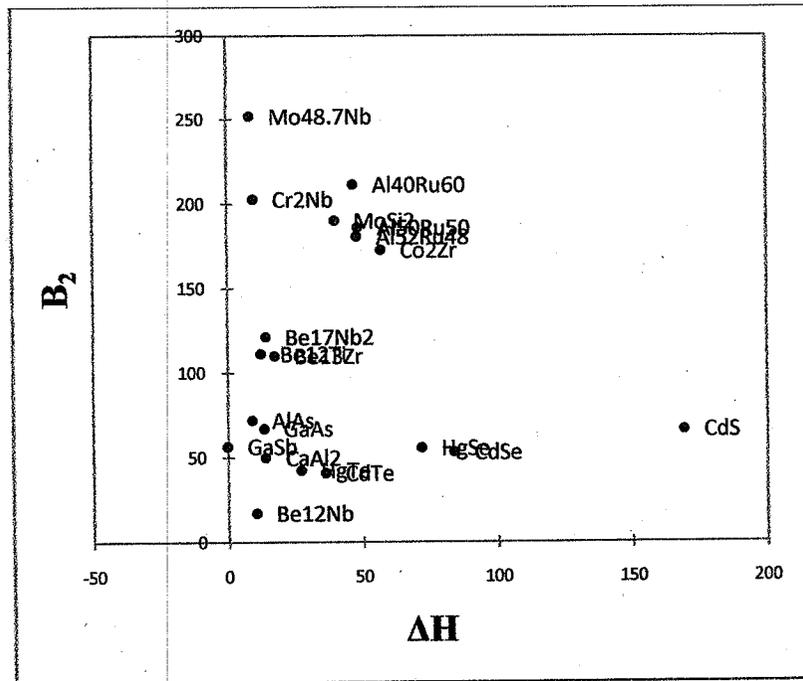


Figure III.12 : Courbe de B_2 en fonction de ΔH .



De la même façon précédente il est impossible de tracer un model linéaire pour l'ensemble des intermétalliques binaires et que pour certains alliages tel que les semi conducteurs, la loi peu être linéaire.

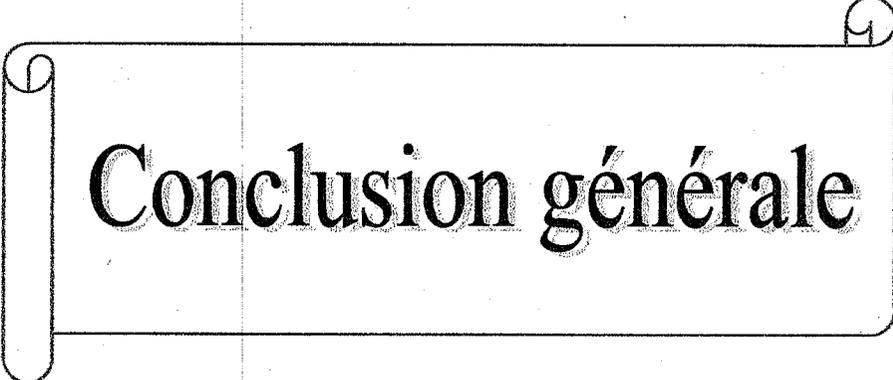
III-Conclusion :

L'utilisation de la ACP et de la PLS dans cette étude démontre comment l'informatique peut être employée pour utiliser l'information pour déterminer ce qui est nécessaire et utile, et employer alors cette connaissance pour la conception des matériaux.

Et on peut aussi découvrir de nouveaux schémas, et de nouvelles corrélations entre les différentes propriétés des matériaux.

Références:

- [1] C. Duby, S. Robin, "Analyse en Composantes Principales", Département O.M.I.P(2006).
- [2] Institute of High Performance Computing of Singapore, 89C Science Park Drive,
#02-11/12 The Rutherford Singapore Science Park 1, Singapore 118261, Singapore
Received April 24, 2001
- [3] LiCH, Wup. Mater 2001; 13:4642 Miedema, A.R.Physica B 1992, 182.1.

A decorative scroll graphic with a black outline and a white fill. The scroll is oriented horizontally and has a small circular flourish at each end, suggesting it is a rolled-up document. The text is centered within the scroll.

Conclusion générale

Conclusion générale

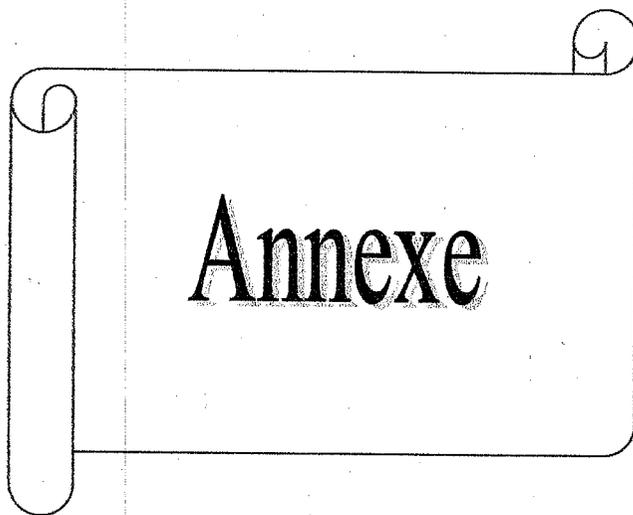
Le travail présenté dans ce mémoire consiste à collecter des données sur les propriétés structurales en utilisant une des approches du « DATAMINING ».

Pour ce résultat on applique l'analyse en composantes principales (ACP) et des moindres carrés partiels (PLS) sur ces données, pour ressortir à la fin avec des conclusions et des observations qui reflètent les résultats théoriques.

L'objectif principal de cette étude est de démontrer que pour un simple calcul ACP il est possible de différencier entre deux types de matériaux différents. Dans notre cas des semi conducteurs et des intermétalliques.

Effectivement notre étude a démontré que ces matériaux se comportent différemment et sont divisés en deux parties catégories cluster 1 et cluster 2.

Les techniques ACP et PLS nous ont permis de visualiser notre base de données sur des graphiques et d'étudier les relations qui existent entre les variables qui forment cette base de données, elle permet aussi d'identifier la structure de dépendance entre les observations, afin d'obtenir une description ou une représentation, sous une projection des données sur différents axes, appelés axes principaux (PC1, et PC2).



Annexe

I- Introduction:

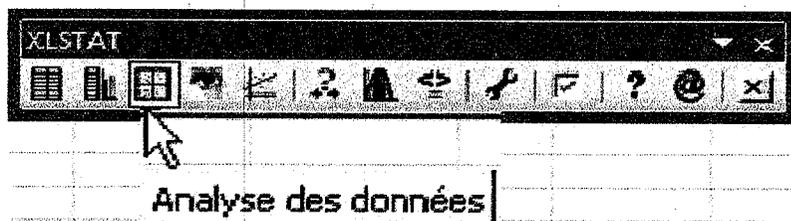
Le logiciel XLSTAT-Pro est un outil statistique d'analyse de données pour Microsoft Excel le plus complet et le plus utilisé, et il est destiné aux tests multidimensionnels sur les moyennes et les variances. XLSTAT-Pro offre de très nombreuses fonctionnalités qui font d'Excel un outil performant et facile d'accès pour répondre à la majorité de nos besoins en analyse de données. XLSTAT fonctionne avec toutes les versions d'Excel.

Ce logiciel peut remplir plusieurs fonctions y compris l'analyse en composantes principales (ACP) et ceci en faisant rentrer les variables explicatives (les X du modèle) qui sont indiquées par les valeurs de densités de charge dans notre cas, et les variables dépendantes (les Y du modèle) indiquées par les propriétés (B, V_M , n_{ws} , ΔH , E.....ect).

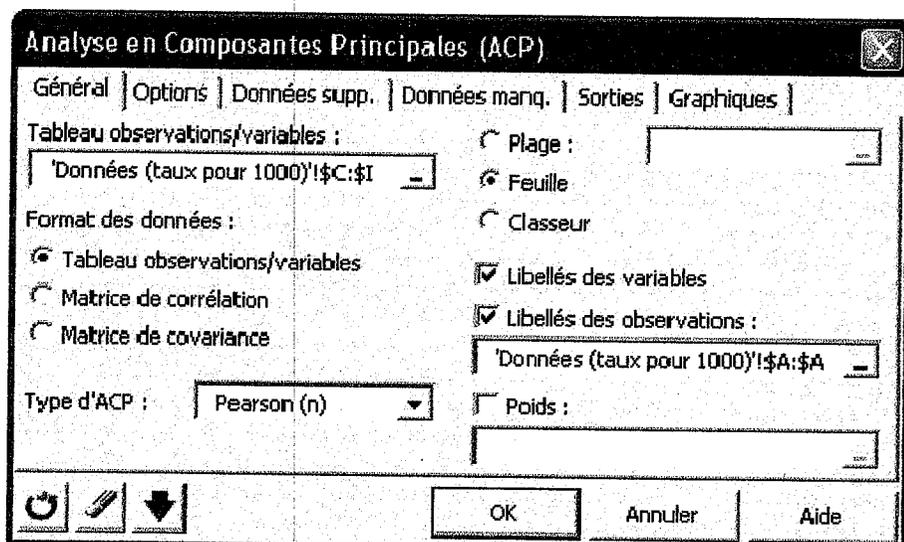
Dans cette annexe nous détaillons le mode l'utilisation de ce logiciel.

II- Faire une Analyse en Composantes Principales (ACP) avec XLSTAT :

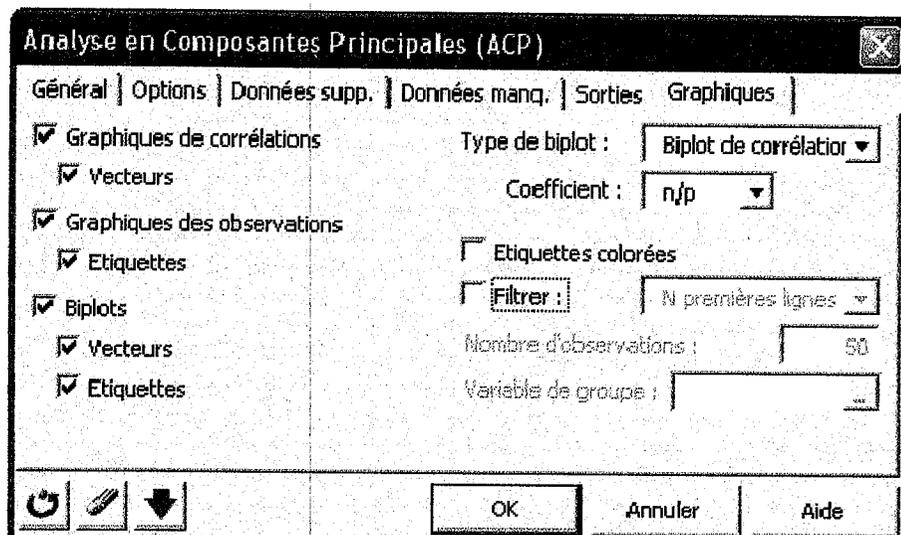
Nous cliquons sur le menu XLSTAT /Analyse de données/Analyse en Composantes Principales, ou cliquons sur le bouton correspondant de la barre "Analyse de données" (voir ci-dessous).



Une fois le bouton cliqué, la boîte de dialogue correspondant à l'Analyse en composantes principales apparaît. Nous pouvons alors sélectionner les données sur la feuille Excel.



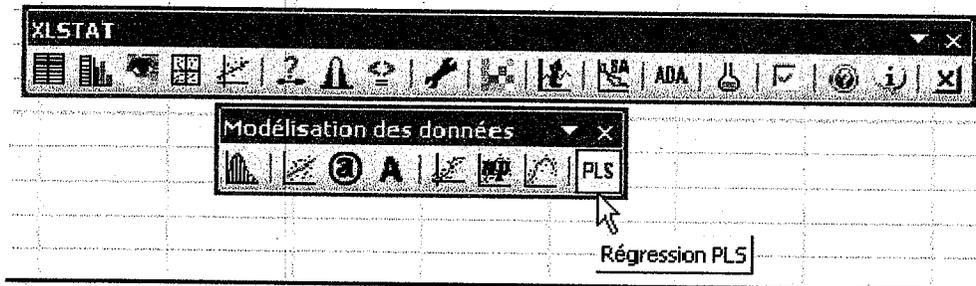
Dans l'onglet "Graphiques", les options d' "Etiquettes" sont toutes activées afin que les libellés des variables et des observations soient bien affichés. L'option de filtrage des observations à afficher est aussi désactivée afin d'afficher toutes les observations. Lorsqu'il y a beaucoup d'observations, il est conseillé de ne pas afficher les étiquettes pour accélérer la génération du graphique, et de ne pas afficher toutes les observations afin de rendre le graphique plus lisible.



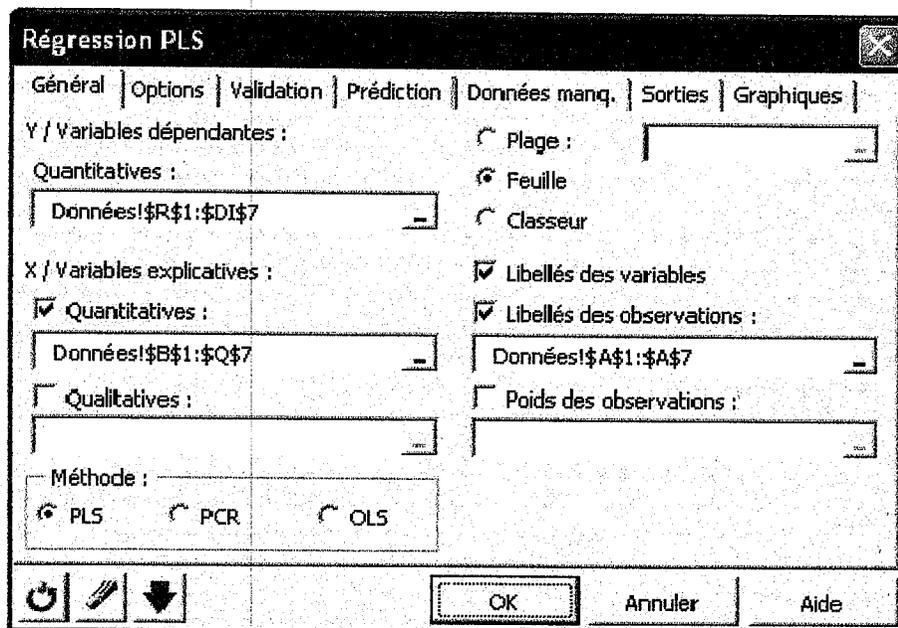
- Le premier résultat intéressant est la matrice des corrélations. Les tableaux des valeurs et vecteurs propres sont ensuite affichés. Tous les graphiques et tableaux sont mentionnés dans la partie résultat.

III- Faire une régression PLS avec XLSTAT :

Nous lançons XLSTAT en cliquant sur le bouton de la barre d'outils Excel, puis nous sélectionnons la commande Régression PLS de la barre d'outils "Modélisation des données".



Une fois que nous avons cliqué sur le bouton, la boîte de dialogue apparaît. Sélectionnons au niveau des "Variables dépendantes" (les "Y" du modèle), les propriétés des matériaux. Ce sont en effet les données que l'on veut expliquer à travers des variables explicatives quantitatives (les "X" du modèle) que sont les composantes binaires.



- Après quelques statistiques de base sur les différentes variables sélectionnées et la matrice de corrélations correspondante, les résultats propres à la régression PLS sont affichés. Tous les graphiques de corrélations et tableaux correspondant sont mentionnés dans la partie résultat.