



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

## **MEMOIRE DE MAGISTER**

**EN PHYSIQUE ET CHIMIE DE LA MATIERE CONDENSEE**

**Présentée par :**

**M<sup>elle</sup> : LAROUSSE SAMIA**

***Thème :***

**Construction D'une Carte Intermétalliques à Partir Du  
Caractère De La Liaison**

Soutenue en 2010 devant la commission d'examen :

<b>Pr. G.Merad</b>	<b>Président</b>	<b>Professeur</b>	Université de Tlemcen
	<b>Examineur</b>	<b>Professeur</b>	Université de Sidi-Bel-Abbès
<b>M<sup>elle</sup> H. FARAOUN</b>	<b>Examineur</b>	<b>Maître de Conférences</b>	Université de Tlemcen
<b>Pr. H.Aourag</b>	<b>Encadreur</b>	<b>Professeur</b>	Université de Tlemcen

# Remerciements

*Je remercie en premier lieu Dieu le tout puissant de nous avoir accordé la puissance et la et la volonté pour terminer ce travail.*

*Ce travail a été réalisé dans le laboratoire de recherche d'Etudes et Prédiction des Matériaux (LEPM) affilié à l'Unité de Recherche des Matériaux et des Energies Renouvelables (URMER) de la faculté des sciences, sous la direction de monsieur **H. AOURAG**, Professeur à l'Université de Tlemcen.*

*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 monsieur **G. MERAD**, Professeur à l'Université de Tlemcen, d'avoir accepté de présider le Jury de cette thèse.*

*Je remercie sincèrement **Mr. Benzair.Abdenour**, Professeur à l'université d Sidi Bel Abbes ,et **Melle. H. I. FARAOUN**, Maître de conférences à l'université de Tlemcen d'avoir accepter de faire partie du Jury.*

*Je tiens également à exprimer mes remerciements aux membres du laboratoire de recherche d'Etudes et Prédiction des Matériaux (LEPM)*

## *Dédicaces*

*Je dédie cette œuvre avec profond sentiment de reconnaissance à deux personnes les plus cher au monde Mes Parents qui sans leurs sacrifices et leurs compréhension, je ne serais jamais là ; merci chère mère, merci cher père.*

*Je dédie aussi à :*

*Mon frère Hocine et sa femme.*

*Mes sœurs et leurs maris ; Fatma ; Rachida ; Amaria.*

*A ma sœur Noria.*

*A mes nièces et nouveaux.*

*A ma grand-mère Haná « Fatima » ; à ma tante « Aicha ».*

*A ma chère amie Benhmed Samira.*

*A ceux qui n'ont pas hésité à mes tendres la main et ont fait preuve d'amitié avec remerciements : Fatiha, Fatima Zohra.*

*A tous les étudiants du Magister et doctorant de laboratoire «URMER ».*

# Sommaire

Introduction générale .....[01]

## **Chapitre I : revue bibliographique**

**I-1 Les intermétalliques .....[05]**

*I-1-1 Progrès des études sur les composés intermétalliques.....[05]*

*I-1-2 Généralités sur les intermétalliques.....[6 ,8]*

*I-1-2 Les intermétalliques les plus étudiés .....[9]*

**I-2 Liaisons Entre Atomes.....[10,12]**

1-2-1 La liaison métallique.

1-2-2 Liaison covalente.

1-2-3 Liaison ionique.

**I-3 Propriétés des matériaux et types de liaisons.....[12 ,13]**

**I-4 Les différentes phases des composés intermétalliques.....[13]**

*I-4-1 Composés électroniques (phases de Hume – Rothery)....[13]*

*I-4-2 Composés typologiquement compacts.....[14]*

**I-4-2-1 Les Phases de laves.....[14,15]**

*A - Structure type  $MgCu_2$  ou structure C15....[15]*

*B- Structure type  $MgZn_2$  ou structure C14.....[16]*

*C- Structure type  $MgNi_2$  ou structure C36.....[17]*

**I-4-2-2 Les composés A15.....[17]**

*I-4-3 Structure  $L1_0$  (AuCu).....[18]*

*I-4-4 Structure B1 (NaCl). .....[18]*

*I-4-5 Structure B2.....[19]*

**I-5 Quelques propriétés des intermétalliques .....[19]**

I-5-1 Propriétés mécaniques.....	[19]
I-5-2 Propriétés physiques.....	[20]
I-5-3 Propriétés magnétiques.....	[21]
I-5-4 Propriétés électriques .....	[21]

## ***Chapitre II : les techniques d'exploitation des données (Datamining)***

### ***II-1 Datamining.***

II-1-1 Historique du datamining.....	[25,26]
II-1-2 Définition du datamining .....	[26]
II-1-3 Techniques du datamining .....	[27]
➤ <i>Les techniques descriptives</i> .....	[27]
➤ <i>Les techniques prédictives</i> .....	[27]
II-1-4 Outils du datamining.....	[27,28]
II-1-5 Les application du datamining.....	[28]
II-1-6 Avantages et inconvénients du datamining.....	[29]
<b><i>II-2 Conclusion</i></b> .....	[29]

## ***Chapitre III : Analyse en composantes principales ACP et la Régression PLS***

<b><i>III-1 L'Analyse en composantes principales</i></b> .....	[32, 33]
III-1-1 Introduction.....	[32]
III-1-2 Définition.....	[33]
<b><i>III-2 Principe de l'ACP</i></b> .....	[33]

<b>III-2-1</b> formulation mathématique d'une ACP.....	[34]
<b>III-2-2</b> Le problème.....	[34]
<b>III-3</b> <i>La détermination des facteurs et des composantes principales</i> .....	[35]
<b>III-3-1</b> Dans l'espace des individus.....	[36]
<b>III-4</b> <i>Le nombre maximal d'étapes</i> .....	[37,38]
<b>III-5</b> <i>Les valeurs propres</i> .....	[38]
<b>III-6</b> <i>La part de variance expliquée par un axe</i> .....	[39]
<b>III-7</b> <i>Les graphiques et les aides à l'interprétation</i> .....	[40]
<b>III-7-1</b> Représentation des individus.....	[41]

## **Présentation de la régression PLS**

<b>III -9</b> <i>La régression PLS</i> .....	[44,45]
<b>III-9-1</b> Introduction.....	[45]
<b>III-9-2</b> Contexte historique.....	[45]
<b>III-10</b> <i>Régression PLS</i> .....	[46]
<b>III-10-1</b> La régression linéaire.....	[46,47]
<b>III-10-2</b> les avantages de la régression PLS.....	[47]
<b>III-9-3</b> La méthode NIPALS.....	[47,48]
<b>III-11</b> <i>Comparaison</i> .....	[48]
<b>III-11-1</b> La Régression Linéaire Multiple (MLR).....	[48]

<b>III-11-2</b> La régression sur composantes principales.....	[49]
<b>III-12</b> <i>Intérêt de la Régression PLS</i> .....	[49]
A- Établissement du modèle PLS simple .....	[49,50]
B- La méthode .....	[51]
C-Propriétés des composantes $t_1, \dots, t_A$ .....	[52]
<b>III-13</b> <i>Interprétation du modèle PLS</i> .....	[52,53]
<b>III-13-1</b> Choix du nombre de composantes PLS .....	[53]
A-Critère d'Ajustement.....	[53]
B-Critère de Validation Croisée.....	[54]
C-Critère3 : critère de validation externe et Prédiction...	[54]
<b>III-9</b> <i>Conclusion</i> .....	[55]

## ***Chapitre IV: Résultats et discussions***

<b>IV-1</b> <i>Introduction</i> .....	[59]
<b>IV-2</b> <i>La base de données</i> .....	[59]
<b>IV .3</b> <i>Technique de calculs :</i>	
<b>IV.3</b> <i>.L'application de L'analyse en composantes principales (ACP)</i> [60,67]	
<b>IV.4</b> <i>Le modèle PLS (Partial Least Squares)</i> .....	[68]
<b>IV.4.1</b> <i>Application de la régression PLS</i> .....	[68,69]
<b>IV.4.2</b> <i>Qualité du modèle</i> .....	[69 ,71]
<b>IV.4.3</b> <i>Les vecteurs du poids W et C</i> .....	[72,73]

<i>IV.4.4</i>	<i>Mesure de l'importance des variables</i> .....	[73,74]
<i>IV.4.5</i>	<i>Paramètre et équation du modèle</i> .....	[74,76]
<i>IV.4.6</i>	<i>L'ajustement du modèle</i> .....	[77,78]
<i>IV.5</i>	<i>Les Résidus et Prédictions</i> .....	[79,86]
<i>IV.5.1</i>	<i>La validation des équations pour la base des données</i> .....	[87]
<i>IV.6</i>	<i>Prédiction des propriétés des intermétalliques ternaires</i> .....	[87, 108]

# *Introduction générale*

Les intermétalliques sont des composés obtenus par combinaison de deux ou plusieurs éléments métalliques et qui présentent une structure cristalline différant de celle des éléments qui les composent. La nature non directionnelle des liaisons métalliques est perdue partiellement donnant lieu à des structures chimiquement ordonnées qui procurent à ces composés des propriétés particulières. Initialement, ces composés étaient employés comme phases durcissantes, puis ils ont été isolés afin de les développer comme principal composant dans différentes applications.

Dans cette étude nous nous sommes intéressés à la construction d'une carte des alliages intermétalliques à partir des résultats de l'ACP (Principal Component Analysis) est une méthode statistique de projection de données multidimensionnelles, utilisée pour la réduction de la dimensionnalité des données. Le but de l'Analyse en Composantes Principales (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. Ainsi, la première nouvelle composante contient les informations relatives au pourcentage maximal de variance, la deuxième contient les informations relatives au pourcentage de variance suivant. Le processus est répété jusqu'à l'obtention de la dernière nouvelle composante, et du PLS (Partial Least Squares), qui est une méthode prédictive de régression linéaire à plusieurs variables. Nous a permis d'obtenir une loi linéaire reliant des variables réponses aux variables explicatives. Sur la base d'une grande diversité de propriétés physiques. Ce travail a pour objectif l'étude de l'influence des concentrations sur les propriétés des différents alliages intermétalliques.

Dans un milieu isotrope il y a deux modules d'élasticité indépendants: le module de cisaillement (G) et le module de compressibilité (B). Ces quantités peuvent être reliées aux constantes élastiques du cristal simple en utilisant différentes techniques. Le module de cisaillement isotrope pour un grand nombre de matériaux et un indicateur de leur dureté mécanique H. Le module de compressibilité isotrope représente une mesure de la force des liaisons atomiques dans le cristal, et il est proportionnel à l'énergie cohésive, c'est le deuxième critère utilisé pour caractériser une dureté élevée.

Le rapport B/G est un paramètre de mesure de ductilité des solides. La ductilité est caractérisée par un rapport B/G élevé ( $>1,75$ ), alors qu'un petit rapport B/G ( $<1,75$ ) indique la fragilité des matériaux. Un autre critère physique, utilisé pour estimer la tendance de ductilité

de certains matériaux cubiques est nommé la pression de Cauchy  $C_{12}$ - $C_{44}$  il est basé sur le degré du caractère angulaire de la liaison chimique. En effet, les matériaux ductiles ont des valeurs positives de pression de Cauchy, qui correspondent à une liaison métallique plus isotrope. D'autre part, les matériaux fragiles montrent des valeurs négatives de pression de Cauchy, qui résultent du caractère plus angulaire de la liaison.

Pour la conception de nouveaux matériaux, non seulement la détermination des données par des calculs ou à partir des expériences est importante, mais aussi la manière d'analyser ces données d'une façon efficace et complète est également nécessaire, et de comprendre les corrélations entre les différentes propriétés, et l'utilisation de ces corrélations pour concevoir des matériaux avec des propriétés désirées.

Dans ce mémoire nous avons utilisé un axe de recherche très étendu : la datamining, qui est novateur en science des matériaux, qui utilise deux techniques : l'analyse en composante principale (ACP), et la régression PLS en un outil de prédiction de nouveaux matériaux avec des propriétés désirées, et ceci en exploitant les données déjà existantes tant sur le plan expérimental que théorique.

Le manuscrit présente dans ce premier chapitre une revue bibliographique qui précise le contexte de l'étude et les bases théoriques nécessaires pour mieux comprendre son déroulement. Dans le chapitre II nous donnons une idée générale sur l'ensemble des techniques de datamining, leur historique, Techniques du datamining (*descriptives*, *prédictives*), Outils du datamining, les applications du datamining, les avantages et les inconvénients du datamining. La troisième chapitre présentera le principe de l'analyse en composantes principales (ACP) et le principe de la régression PLS (Partial Least Squares). L'identification du modèle ACP repose, généralement, sur deux étapes, la première consiste à déterminer la structure du modèle alors que la seconde consiste en l'estimation des paramètres. La deuxième étape est très simple et se ramène à un calcul de valeurs et vecteurs propres, consiste à déterminer le nombre de composantes principales à retenir dans le modèle ACP, pour la régression PLS : Un petit aperçu sur l'historique et le développement de la méthode est d'abord présenté. Ensuite une interprétation sur la régression PLS.

Dans le chapitre quatre nous présentons, les résultats liés au développement de nouvelles approches du Datamining, l'Analyse en composantes principales (ACP) et la Régression PLS, des techniques informatiques qui ont été appliquées sur les intermétalliques pour prédire des nouveaux matériaux.

Enfin, une conclusion générale récapitule les principaux résultats obtenus et propose des perspectives de travaux futurs.

## **Chapitre I : Données Bibliographiques**

**C**e chapitre s'intéresse dans un premier lieu au progrès des études sur les composés intermétalliques. Après, une citation des intermétalliques les plus étudiés. Ensuite, une présentation générale des intermétalliques, les différentes phases des composés intermétalliques (Composés électroniques (phases de Hume – Rothery), Composés typologiquement compacts ....). Enfin on mentionne quelques propriétés des intermétalliques (physiques, mécaniques, électriques, magnétiques).

**I-1 Les intermétalliques ..... [05,12].**

**I-4 Les différentes phases des composés intermétalliques ..... [13,18].**

**I-5 Quelques propriétés des intermétalliques ..... [19,21].**

**I-1 Les intermétalliques :****I-1-1 Progrès des études sur les composés intermétalliques :**

Depuis les dernières décennies, les composés intermétalliques sont considérés comme une classe de matériaux fondamentaux à part entière. Bien que la première étude sur les intermétalliques ait été développée par Kurnakov et al, au cours de l'année 1916, c'est à partir des années 50 que les études ont réellement commencé. La grande fragilité des intermétalliques a contribué à freiner leur développement au cours des années 60 [1].

Avant les années 1970, très peu de progrès ont été faits dans les études et les développements sur les composés intermétalliques en raison de leur fragilité intrinsèque à la température ambiante, cela limitait leurs utilisations industrielles.

A la fin des années 1970, quelques progrès remarquables sur les études de certains composés intermétalliques ont déclenché à nouveau l'intérêt du monde entier pour les alliages intermétalliques [2]. Aoki et Izumi ont montré que quelques intermétalliques fragiles pouvaient être rendus plus ductiles par des additions de bore [3]. En 1976, des chercheurs américains ont trouvé que, par addition d'éléments et par la technique de métallurgie des poudres, la ductilité et la résistance d'alliages à base de TiAl et de Ti<sub>3</sub>Al pouvaient être améliorées [4]. Ensuite, ils ont trouvé qu'à la température ambiante, après avoir substitué partiellement au Co du Ni ou du Fe, le composé Co<sub>3</sub>Al pouvait être transformé d'une structure hexagonale à une structure L1<sub>2</sub>, ce qui présente une certaine ductilité. En 1979, les chercheurs américains et japonais ont indiqué presque en même temps que, par l'addition du bore, la ductilité du composé Ni<sub>3</sub>Al pouvait être significativement améliorée [5].

Depuis ces 20 dernières années, les composés intermétalliques sont largement étudiés, les plus grands progrès concernent les composés intermétalliques de types A<sub>3</sub>B et AB dans les trois systèmes tels que NiAl, TiAl, et FeAl. Les composés intermétalliques tels que NiAl, Ni<sub>3</sub>Al, TiAl et Ti<sub>3</sub>Al présentent une bonne tenue à haute température, ils pourraient donc être utilisés comme matériaux de structure mécanique. Actuellement, certains éléments de turboréacteurs fabriqués en alliages de Ni<sub>3</sub>Al, de TiAl et de Ti<sub>3</sub>Al sont entrain d'être testés en dimensions réelles. En particulier, les composants en alliages de Ti<sub>3</sub>Al sont déjà utilisés en pratique. Les composés de FeAl seront également prometteurs dans le domaine industriel

grâce à leurs bas coûts d'accès, à leurs ressources mondiales et à leurs hautes résistances à l'oxydation. Ils attirent donc beaucoup de chercheurs dans le monde entier [6].

Les intermétalliques ne forment pas un groupe homogène de matériaux, ils ne présentent pas des propriétés élémentaires simples. Cependant ils sont très souvent caractérisés par une grande résistance à l'oxydation et à la corrosion aux températures élevées, ainsi que par une grande résistance mécanique combinée à une faible densité. Ce sont ces propriétés qui vont attirer l'attention des chercheurs vers les aluminures de titane, de nickel, de fer et de niobium, les siliciures de molybdène, de nickel, de niobium, la phase de Laves  $\text{Cr}_2\text{Nb}$  et les intermétalliques du groupe du platine (Pt, Pd, Rh, Ru, Ir, et Re). Les siliciures de molybdène, principalement  $\text{MoSi}_2$  et  $\text{Mo}_5\text{Si}_3$ , sont parmi les intermétalliques les plus étudiés [7].

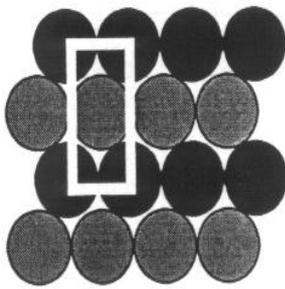
Bien qu'à hautes températures le composé intermétallique  $\text{MoSi}_2$  soit plus résistant à l'oxydation que le  $\text{Mo}_5\text{Si}_3$ , c'est ce dernier  $\text{Mo}_5\text{Si}_3$  qui est le composé le plus étudié du fait de sa bonne résistance au fluage, sa résistance à l'oxydation pouvant être améliorée par des additions de bore.

Ces dernières années, de nouveaux systèmes de composés intermétalliques ayant des structures beaucoup plus complexes et des points de fusion beaucoup plus élevés tels que le  $\text{NbAl}$ , le  $\text{BeNb}$  et le  $\text{MoSi}$  deviennent une nouvelle orientation d'étude et une nouvelle tendance de développement comme de nouveaux systèmes de matériaux de structure en température [8].

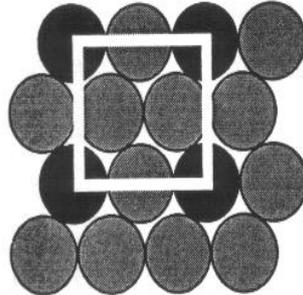
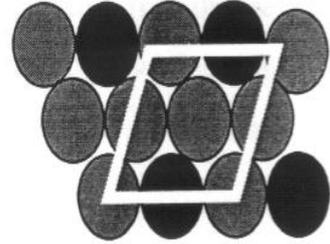
### **I-1-2 Généralités sur les intermétalliques :**

Lorsque deux éléments A et B sont miscibles en toutes proportions et forment une solution solide continue pour certain domaine de concentration, il peut se former des structures. Les composés  $\text{A}_3\text{B}$ ,  $\text{AB}$ ,  $\text{AB}_3$  sont des exemples de ces nouvelles phases. Le nom de phases intermédiaires est plutôt réservé à celles qui possèdent des structures différentes de celles 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 [9], d'où elle peut être définie comme une phase ordonnée d'un matériau formé entre deux ou plusieurs éléments métalliques et présente une structure cristallographique différente de celles des éléments qui la composent. Dans ces composés, les forces de liaison entre atomes de natures différentes

(A-B) sont plus fortes que celles entre atomes similaires (A-A ou B-B), leur température de fusion est élevée et leur structure cristallographique ordonnée (**Fig. I-1**) [10].



Type AB(AuCu)

Type AB<sub>3</sub>(TiAl<sub>3</sub>)Type AB<sub>3</sub>(AuCu<sub>3</sub>)

**Figure I-1** : Solution ordonné des matériaux intermétalliques.

Cependant, leur application industrielle est limitée par leur fragilité à température ambiante qui rend aussi plus délicats les procédés de fabrication [1 ,4].

Leur stabilité dépend de différent facteurs : facteurs de valence (ou de concentration électronique), facteurs de Taille, facteur électrochimique (différences d'électronégativités)[9].

Pour préciser les termes, nous parlerons de « composé intermétallique » quand la composition stœchiométrique est atteinte, et de « phase ou alliage intermétallique » s'il agit d'une solution solide sur la base du composé.

Les composés intermétalliques peuvent être soit stœchiométriques soit non stœchiométriques :

**\*\* Les composés stœchiométriques** sont des solides monophasés qui se forment par réaction entre deux éléments. Leurs compositions et leurs structures sont bien définies: ils apparaissent comme des lignes verticales dans le diagramme de phase, divisant le diagramme en sous domaines.

**\*\* Les composés intermétalliques non stœchiométriques** (ou phases intermédiaires) sont des phases uniques qui apparaissent dans les régions centrales d'un diagramme de phase, ils peuvent tolérer des variations de composition chimique. Ils apparaissent comme une région dans les diagrammes de phase [11]

Les alliages classiques sont formés d'une solution solide désordonnée d'un ou plusieurs éléments. Contrairement à la majorité des alliages métalliques où, après l'insertion des éléments d'addition, les liaisons entre les atomes de l'élément de base restent métalliques et relativement faibles, dans les composés intermétalliques l'insertion d'éléments d'addition change substantiellement la nature des liaisons entre les atomes. La nature non directionnelle de la liaison métallique est partiellement perdue et les liaisons peuvent prendre un caractère mixte, partiellement métallique, covalent ou ionique, ce qui confère au matériau des liaisons plus fortes. Alternativement, on peut aussi trouver des liaisons entièrement métalliques qui sont plus fortes entre des atomes de nature différente qu'entre les atomes de même nature. Ici, les atomes d'éléments prennent des positions préférentielles à l'intérieur de la maille cristalline - un atome est préférentiellement entouré par des atomes de types différents ; ce qui conduit à une structure ordonnée dont les modes de fracture et les propriétés mécaniques seront particulières [8].

Les différents types de liaison rencontrés confèrent aux intermétalliques des propriétés qui sont intermédiaires entre celles des métaux et celles des céramiques (Tableau 1).

**Tableau I-1 :** Propriétés comparées des métaux, composés intermétalliques et céramiques [12]

ALLIAGES METALLIQUES	COMPOSES INTERMETALLIQUES	CERAMIQUES
densités élevées Module d'élasticité intermédiaire ductile (à l'ambiante) résistance à la traction et à la compression modérément haute (à ambiante) résistance à chaud limitée par la température de fusion résistance à l'oxydation à chaud faible conductivité électrique élevée résistance à la rupture élevée (à ambiante)	densités intermédiaires module d'élasticité assez élevé peu ductile (à ambiante) résistance à la traction variable et résistance à la compression assez élevée (à ambiante) résistance à chaud élevée au voisinage du point de fusion résistance à l'oxydation à chaud assez élevée conductivité électrique moyenne résistance à la rupture faible (à ambiante)	densités faibles module d'élasticité élevé fragile (à ambiante) résistance à la traction variable et résistance à la compression élevée (à ambiante) résistance à chaud très élevée (pas vrai en fluage) résistance à l'oxydation à chaud élevée conductivité électrique très faible résistance à la rupture faible (à ambiante)

Ces matériaux sont intéressants à plus d'un titre. Outre le fait qu'ils possèdent souvent un point de fusion élevé (un grand nombre de composés intermétalliques ont des températures de fusion supérieures à 2000 °C). Stoloff et al. ont décrit les applications industrielles des composés intermétalliques (Tableau 2).

**Tableau I- 2 :** Applications des intermétalliques [13]

APPLICATIONS DES INTERMETALLIQUES	
Structural	automobile
	Aérospatial
Magnétique	
Stockage d'énergie	Batteries
	stockage d'hydrogène
	éléments de chauffe
Outils et matrices	
Matériel de four	Résistant à la corrosion
	tuyauterie pour des industries chimiques
	revêtement
	enduits
Dispositifs électroniques	

**I-1-2 Les intermétalliques les plus étudiés :**

De nombreuses études ont été consacrées aux aluminures de titane ( $Ti_3Al$ ,  $TiAl$ ) et de nickel ( $Ni_3Al$ ,  $NiAl$ ). **Le  $TiAl$  semble être un composé adéquat uniquement pour des applications dans les turbines à basse pression et les parties statiques du moteur, en raison de son faible point de fusion (1500 °C) et malgré sa faible ductilité et faible ténacité à la rupture à la température ambiante.** Le  $NiAl$  a aussi des propriétés intéressantes, notamment un point de fusion sensiblement plus élevé que celui du  $TiAl$  (~1650 °C), une bonne conductivité thermique, une faible densité et une bonne résistance intrinsèque à l'oxydation. Celle-ci est encore améliorée par des additions de tantale et de chrome. Ainsi une bonne résistance à des températures aussi élevées que 1000 °C a été obtenue notamment lors d'essais dans des turbines stationnaires. Sa ténacité à la température ambiante et sa résistance au fluage doivent encore être améliorées [4 ,12].

Ces aluminures sont de plus en plus étudiés comme **successeurs des superalliages** notamment en raison de leur **faible densité**. Mais il semble que pour des températures supérieures les microstructures à plus fort potentiel soient celles présentant une ou des phases intermétalliques en équilibre avec un métal ou alliage réfractaire [7]. **Le système Mo-Si a été également très étudié. La phase MoSi<sub>2</sub> dont le point de fusion est élevé (2030 °C) a une excellente résistance à l'oxydation à l'air même pour des températures de l'ordre de (1600 – 1700 °C). Son utilisation reste cependant limitée à cause de sa fragilité (ductilité faible) à basse température, de sa faible résistance mécanique au-dessus de 1000 °C et d'un taux de fluage élevé au-delà de 1200 °C. Cependant ses propriétés mécaniques peuvent être améliorées et son oxydation catastrophique évitée soit par des modifications structurales induites par un changement de méthode d'élaboration, soit par une pré-oxydation à haute température ou par l'insertion d'éléments d'addition.**

## **I-2 LIAISONS ENTRE ATOMES :**

Les liaisons entre atomes assurent la cohésion du matériau et en particulier sa déformabilité, sa conductivité (thermique, électrique...), ses propriétés magnétiques, sa fragilité, sa densité... La « taille » d'un atome ou d'un ion dépend du type de liaison concerné, à travers les positions relatives les plus probables de deux entités voisines (minimum d'énergie). Le rayon atomique augmente lorsque l'on descend le long d'une colonne de la classification périodique; les anions sont d'une manière générale beaucoup plus « encombrants » que les cations, ce qui explique l'arrangement des solides ioniques à l'échelle atomique. On distingue ci-dessous les liaisons selon la nature de l'interaction concernée [14].

### **1-2-1 La liaison métallique :**

Elle est assurée par la mise en commun des électrons de la couche périphérique de l'atome en un nuage d'électrons où ils sont libres et délocalisés. Elle est caractérisée par le fait que les électrons de liaison sont libres. On peut considérer le métal comme un empilement d'ions positifs entourés par un nuage d'électrons. Ces électrons libres expliquent, par exemple, les propriétés de conductivité thermique et électrique des métaux.

L'arrangement cristallin des matériaux métalliques résulte des forces d'attraction et de répulsion qui s'exercent entre des ions positifs et le nuage d'électrons délocalisés. Les structures cristallines qui en découlent sont en règle générale hautement symétriques, de compacité élevée et peuvent être considérées comme des empilements de sphères. C'est une liaison non directionnelle.

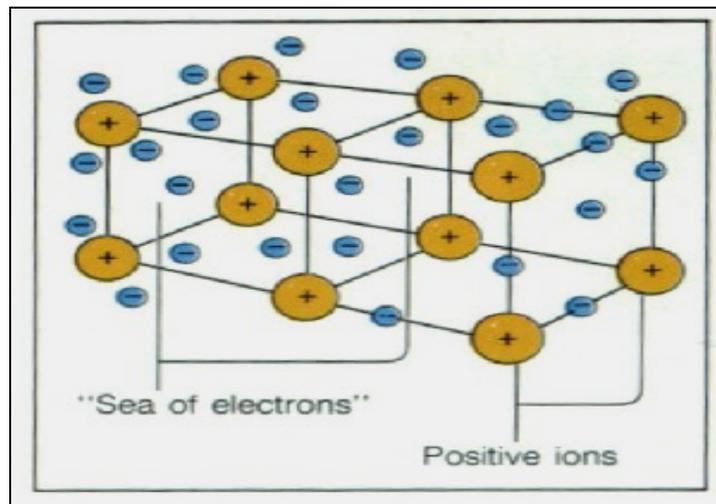


Figure 1.2.1 : Liaison métallique

### 1-2-2 Liaison covalente :

La liaison covalente intervient entre deux atomes qui se partagent un ou plusieurs électrons afin de compléter leur couche électronique extérieure à huit électrons.

C'est une liaison directionnelle, ce qui aura des conséquences importantes pour certaines caractéristiques des matériaux.

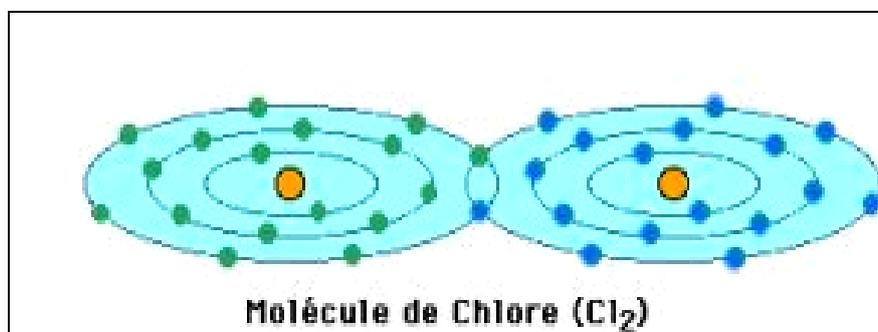


Figure 1.2.2 : Liaison covalente

**1-2-3 Liaison ionique :**

Comme la liaison covalente, la liaison ionique est une liaison de forte intensité, puisqu'une fois liés, les atomes ont une couche électronique externe complète à 8 électrons. En se liant entre eux par des liaisons ioniques, les atomes perdent ou gagnent un ou plusieurs électrons et deviennent ainsi des ions négatifs ou positifs (anions ou cations).

C'est une liaison non directionnelle, puisqu'en fait elle est le résultat d'une attraction électrostatique entre des charges électriques (anions et cations) de signes opposés.

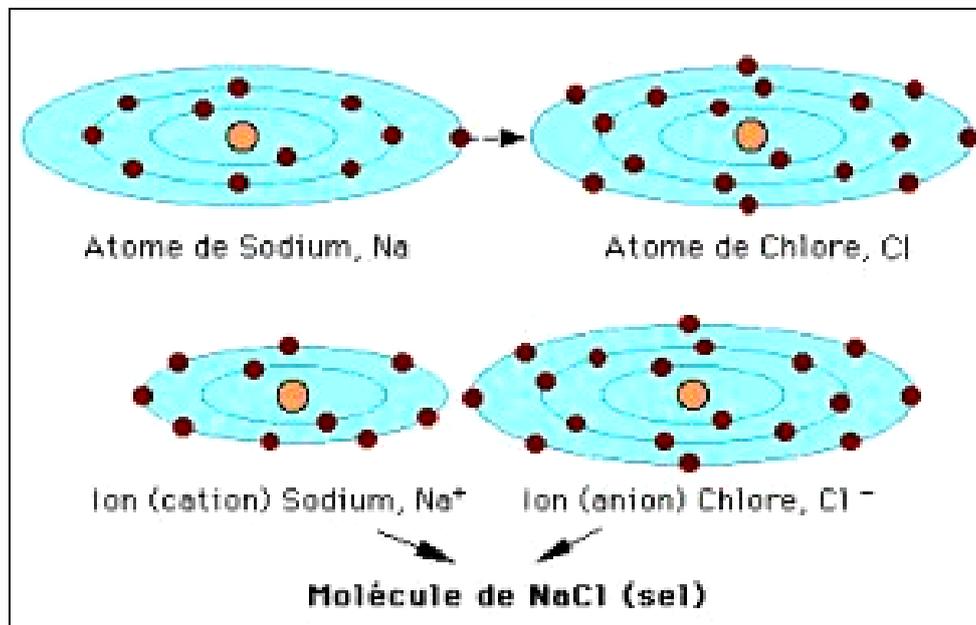


Figure 1.2.3 : Liaison ionique

**I-3 Propriétés des matériaux et types de liaison : [15]**

Des propriétés telles que la fragilité et la ductilité peuvent être facilement déduites de la nature des liaisons d'un matériau. Ainsi, puisque dans les matériaux à liaisons covalentes les valeurs des angles entre les atomes ne peuvent guère varier (liaisons directionnelles), la déformation plastique de ces matériaux est à peu près impossible : ils sont fragiles. Par conséquent, un tel matériau soumis à une contrainte se rompt généralement de façon fragile

par bris des liaisons, sans qu'il y ait possibilité d'un déplacement permanent des atomes les uns par rapport aux autres.

Par contre, dans les matériaux métalliques, où les ions positifs baignent dans un nuage d'électrons délocalisés, un déplacement relatif permanent des atomes les uns par rapport aux autres est possible et assure une certaine ductilité.

Les conductibilités thermique et électrique sont deux propriétés qui dépendent de la nature des liaisons. Elles sont dues principalement à la possibilité de déplacement des électrons sous l'effet d'un champ électrique ou d'un gradient de température. Plus ces électrons sont libres de se mouvoir, plus les conductibilités électrique et thermique sont élevées. C'est pourquoi les métaux dont les électrons de valence sont délocalisés sont bien meilleurs conducteurs d'électricité et de chaleur que les matériaux covalents ou ioniques, dans lesquels la localisation des électrons de valence est très marquée.

La très forte intensité des liaisons covalente et ionique fait en sorte que les céramiques ont des températures de fusion élevées et une grande inertie chimique, ce qui en fait des matériaux de choix pour des applications à haute température.

#### **I-4 Les différentes phases des composés intermétalliques :**

Les composés intermétalliques peuvent être classés en deux grandes catégories selon la stabilité du composés (la stabilité dépend de la structure électronique : la liaison chimique ...), sa composition chimique, et sa structure cristallographie.

##### **I-4-1 Composés électroniques (phases de Hume – Rothery) : [16]**

Ce sont des composés dont les formules et les structures sont déterminées par la concentration en atomes et en électrons de valence (d'où leur nom de composés électroniques) autour des nombres caractéristiques suivantes :

**La phase beta  $\beta$  :** cubique centré, caractérisée par un rapport 3/2 (exp : CuBe, CuZn).

**La phase gamma  $\gamma$  :** cubique complexe, caractérisée par un rapport 21/13 (exp : Cu<sub>9</sub>Al<sub>4</sub>, Cu<sub>5</sub>Zn<sub>8</sub>, Cu<sub>31</sub>Sn<sub>8</sub>).

La phase epsilon  $\epsilon$  : hexagonale, caractérisée par un rapport 7/4 (exp :  $\text{CuZn}_3$ ,  $\text{Cu}_5\text{Al}_3$ ,  $\text{Cu}_3\text{Sn}$ ).

#### I-4-2 Composés typologiquement compacts : [16]

Friauf a introduit en 1927 l'idée de considérer la structure de composés comme des polyèdres formés par des groupes de 16 atomes en un arrangement de compacité maximale de sphères dures de même diamètre dans le système cubique. Les polyèdres sont, en fait, des tétraèdres dont les quatre sommets ont été tronqués. J.Kasper et Frank considèrent que l'interaction entre les proches voisins peut être décrite par des configurations géométriques. L'espace géométrique est rempli avec des polyèdres convexes appelés les polyèdres de Kasper. Ces polyèdres s'interpénètrent de façon que chaque atome situé en un sommet d'un polyèdres se trouve au centre d'un autre polyèdre. La structure peut être décrite par un empilement de tétraèdres plus au moins distordus.

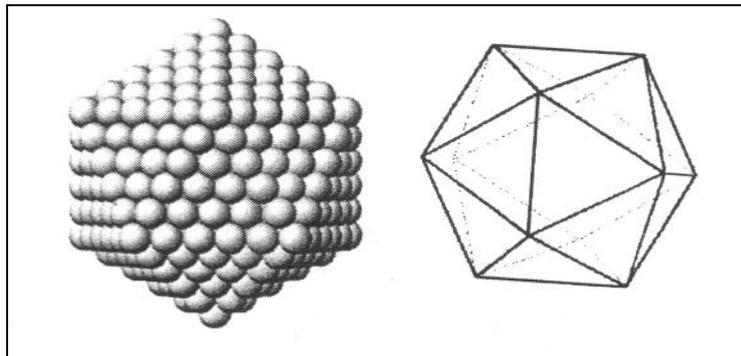


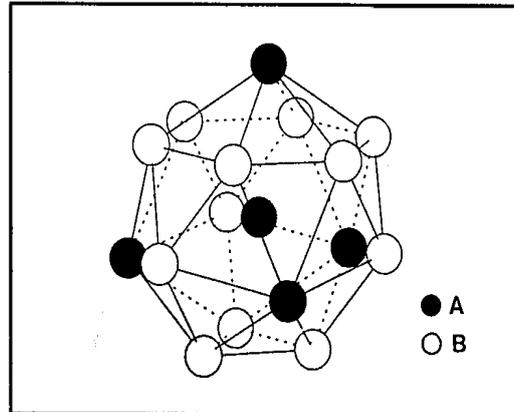
Figure I.4.2 : Icosaèdre régulière de Kasper

Relèvent de cette description de nombreux composés intermétalliques, qui appartiennent à différents groupes. Parmi eux citons les phases de Laves et les composés A15.

#### I-4-2-1 Les Phases de laves : [17]

Les phases de laves sont des composés intermétalliques de stœchiométrie  $\text{AB}_2$  cristallisant dans l'une des trois structures type :  $\text{MgCu}_2$  (cubique),  $\text{MgZn}_2$  ou  $\text{MgNi}_2$  (hexagonales). Elles correspondent en effet à des empilements compacts pour un rapport des rayons atomiques  $R_A/R_B = 1,225$ .

Toutes ces phases sont caractérisées par des polyèdres de coordinations identiques pour l'atome A, appelés polyèdres de Friauf. Cet environnement est représenté sur la figure 1.4.2.1



**Figure I.4.2.1 :** environnement de l'atome A dans les phases de laves

Les trois structures peuvent être décrites comme un empilement de ces polyèdres. C'est la symétrie des empilements qui génère les différences entre les trois structures.

Une seconde façon de représenter ces structures est de les décrire comme des empilements de couches identiques et de décalées selon abcabc.....pour la structure type "MgCu<sub>2</sub>", ababab....pour la structure type MgZn<sub>2</sub> abacabac..... pour la structure type MgNi<sub>2</sub>. Ces plans sont parallèles aux plans (111) de la structure cubique (MgCu<sub>2</sub>) et parallèles aux plans (001) des structures hexagonales (MgZn<sub>2</sub> et MgNi<sub>2</sub>).

### **A - Structure type MgCu<sub>2</sub> ou structure C15 :**

La figure .A. désigne la structure C15 ou les atomes B s'arrangent en tétraèdres, chaque tétraèdre étant joint par les sommets. Les atomes A occupent alors les espaces libres entre les tétraèdres formant entre eux un réseau de type diamant. Le réseau est cubique à faces centrées et il y a 8 motifs AB<sub>2</sub> par maille.

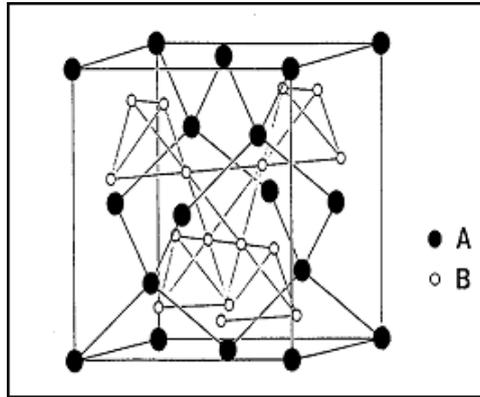


Figure .A. : structure type  $MgCu_2$  ou structure C15

**B- Structure type  $MgZn_2$  ou structure C14 :**

Dans cette structure le réseau est hexagonal. Les atomes B s'arrangent toujours en tétraèdres mais ces tétraèdres sont alternativement joints par les bases et par les sommets. Il ya cette fois 4 motifs par maille. La **figure** représente la structure C14.

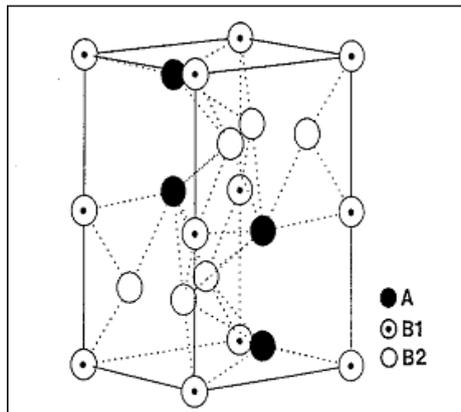
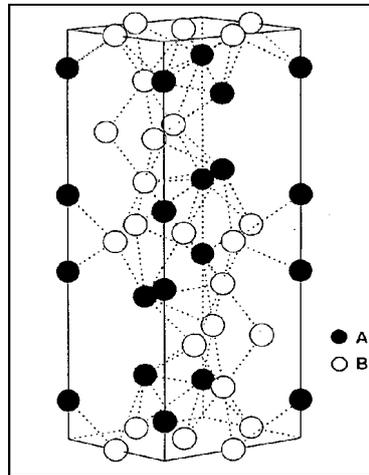


Figure .B. : structure type  $MgZn_2$  ou structure C14

**C- Structure type  $MgNi_2$  ou structure C36 :**

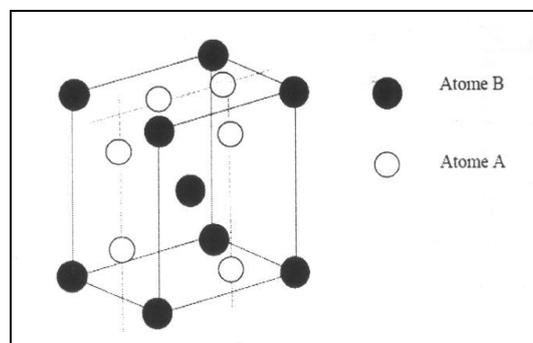
Cette structure est une combinaison des deux précédentes. Le réseau représenté sur la figure est hexagonal avec une maille doublée selon c par rapport à la structure C14.



**Figure.C. :** Structure type  $MgNi_2$  ou structure C36 :

**I-4-2-2 Les composés A15 : [18]**

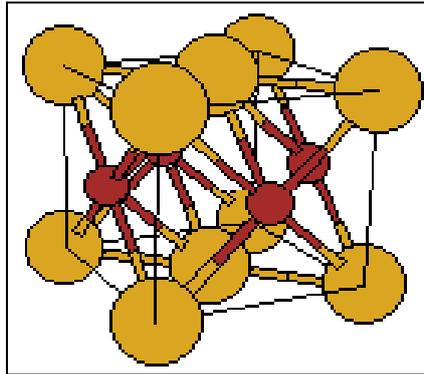
Les composés A15 ont une structure cristallographique d'un corps de composition  $A_3B$ . Ces composés sont dits géométriquement compacts et non topologiquement compacts parce qu'ils sont constitués par la superposition de plans dans lesquels la meilleure capacité est réalisée dans les deux dimensions du plan. La disposition des atomes dans ces plans présente une symétrie ternaire (plan T) ou binaire (plan R). Les atomes A forment des chaînes parallèles aux trois directions cristallographiques (100), (010) et (001).



**Figure.I.4.2.2 :** maille élémentaire d'un composé A15 (composition  $A_3B$ )

**I-4-3 Structure L1<sub>0</sub> (AuCu) : [19]**

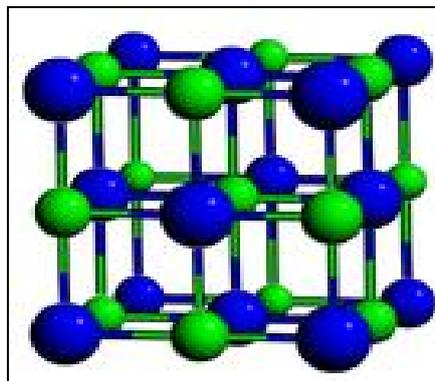
C'est une structure dérivée de la structure cubique centrée. Les atomes "Au" se trouvent aux sommets et au centre de deux faces opposées, les atomes "Cu" se trouvent au centre des autres faces, chaque atome "Au" a quatre atomes plus proches voisins de la même espèce et huit de l'autre espèce.



**Figure I.4.3 :** structure L1<sub>0</sub> de type AuCu.

**I-4-4 Structure B1 (NaCl) : [19]**

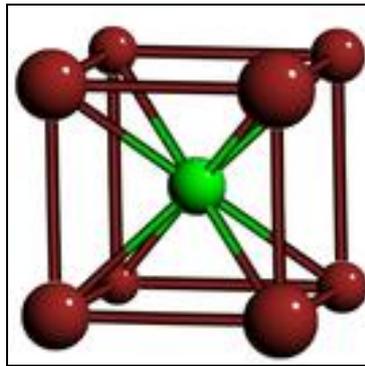
La structure B1 est caractérisée par une répartition octaédrique des atomes. Les atomes "Na" se placent en  $(0, 0, 0)$ ,  $(0, 1/2, 1/2)$ ,  $(1/2, 1/2, 0)$ ,  $(1/2, 0, 1/2)$  et les atomes Cl se placent en  $(1/2, 1/2, 1/2)$ ,  $(0, 1/2, 0)$ ,  $(0, 0, 1/2)$ ,  $(1/2, 0, 0)$ .



**Figure I.4.4 :** structure B1 de type NaCl.

**I-4-5 Structure B2 : [19]**

La structure B2 : structure dérivée de la structure cubique centrée, les atomes de Cs se placent au (0,0,0) et les atomes de Cl se placent en (1/2, 1/2, 1/2).



**Figure I.4.5 :** structure B2 de type CsCl.

**I-5 Quelques propriétés des intermétalliques :**

la variété de ces matériaux et le nombre élevé de leurs structures a pour conséquence une richesse et une diversité considérable de leurs propriétés :

**I-5-1 Propriétés mécaniques : [18]**

Parmi les propriétés attribuées aux intermétalliques en ingénierie sont leur résistance spécifique élevée (module d'élasticité/densité) et leurs grands modules d'élasticité à température élevée.

Le module d'élasticité pour plusieurs intermétalliques est compris entre 150 GPa à 400 GPa pour les matériaux qui ont de basses densités entre 2 à 5g/cm<sup>3</sup>.

Une grande rigidité combinée avec une basse densité est rare dans un métal normal. Cette combinaison des propriétés produit une rigidité spécifique montrée dans le tableau [I-5-1].

La grande rigidité minimise la flexibilité et la distorsion des pièces sous chargées, ce qui facilite pour l'ingénieur l'établissement des montages de pièces qui fonctionnent compatiblement.

L'habilité de maintenir un grand module d'élasticité à température élevée est particulièrement valable et ici encore les intermétalliques ont un avantage majeur par rapport aux matériaux ordinaires. Le module d'élasticité des matériaux ordinaires diminue plus rapidement contrairement aux intermétalliques autant la température s'élève.

Cela résulte de leurs liaisons partielles non métalliques et leurs températures de fusion élevées.

**Tableau [I-5-1] :** Comparaison entre le module d'élasticité, densité, résistance spécifique pour les métaux et les composés intermétalliques :

Métal ou composés	module d'élasticité E à 20°C(GPa)	Densité P (Kg /m <sup>3</sup> )	résistance spécifique E /p (Mg/kg)
W	411	19, 25	21, 4
Al	70	2710	25, 8
Fe	208	7874	26, 4
NiAl	184	5850	31, 5
TiSi <sub>2</sub>	272	4390	62, 0
MoSi <sub>2</sub>	440	6200	71, 0

**I-5-2 Propriétés physiques : [20]**

Lors de la formation des composés intermétalliques les principales propriétés physiques à savoir la dureté, la conductivité thermique et électrique, la susceptibilité magnétique..., seront modifiées de manière significatives de celles de leurs métaux constituants .prenant à titre 1-) d'exemple les propriétés physiques du Cu/Sn, Cu/Ni.(tableau-I-5-2).

Property	Cu <sub>6</sub> Sn <sub>5</sub>	Cu <sub>3</sub> Sn	Ni <sub>3</sub> Sn <sub>4</sub>
Vickers Hardness (kg/mm <sup>2</sup> )	378±55	343±47	365±7
Toughness MPa•m <sup>-1/2</sup>	1.4±0.3	1.7±0.3	1.2±0.1
Youngs Modulus GPa	85.56±1/65	108.3±4.4	133.3±5.6
Shear Modulus GPa	50.21	42.41	45.0
Thermal Expansion x10 <sup>-6</sup> /°C	16.3±.03	19.0±0.3	13.7±0.3
Thermal Diffusivity cm <sup>2</sup> /sec	0.145±0.015	0.24±0.024	0.083±0.008
Heat Capacity J/gm/deg	0.286±0.012	0.326±0.012	0.272±0.012
Resistivity μohm-cm	17.5±0.1	8.93±0.02	28.5±0.1
Density gm/cc	8.28±0.02	8.9±0.02	8.65±0.02
Thermal Conductivity watt/cm-deg	0.341±0.051	0.704±0.098	0.196±0.019

Pour une bonne appréciation des valeurs du tableau(I-5-1), les propriétés des métaux ont été présentés dans le tableau -2- pour le cuivre, le nickel et le tin.

La dureté des intermétalliques est plus grandes que celles des métaux indiquant leur grande fragilité. De même le module d'élasticité est plus grand pour les intermétalliques, provoquant une ductilité inférieure. Le coefficient de dilatation thermique est similaire à celui des métaux. D'autre part les conductivités thermiques et électriques sont généralement basses pour ces alliages.

### **I-5-3 Propriétés magnétiques : [18]**

Plusieurs composés sont ferromagnétiques, comme ceux à base de métaux ferromagnétiques .Ces composés présentent des propriétés intéressantes pour de nombreuses applications.les composés intermétalliques sous forme de particules dispersées dans une matrice métalliques (phénomène de précipitation) sont la source de durcissement structurale. L'orsque les composés intermétalliques sont disséminés dans une matrice métallique pure ou solution, ils frient le mouvement des dioslocations, leurs utilisation comme constituant principale d'alliages à haute résistance à la température élevée fait l'objet de nombreuses recherches, Certain composés intermétalliques sont ferromagnétiques comme  $\text{SmCo}_5$ ....

Le tableau I-5-3présente des différentes propriétés des composés intermétalliques .

### **Le tableau [I-5-3] : Différentes propriétés des composés intermétalliques. [20]**

Propriétés	composés
Propriétés mécaniques	$\text{Ti}_3\text{Al}$ , $\text{TiAl}$ , $\text{Ni}_3\text{Al}$ , $\text{NiAl}$ , $\text{Fe}_3\text{Al}$ , $\text{FeAl}$ , $\text{MoSi}_2$
Hydrogen storage	$\text{LaNi}_5$ , $\text{TiFe}$ , $\text{Mg}_2\text{Ni}$
Shape memory	$\text{TiNi}$
Thermoelectricity	$\text{FeSi}_2$ , $\text{Mg}_2\text{Si}$ , $\text{MnSi}_2$
Ferromagnétique	$\text{SmCo}_5$
Supraconducteur	$\text{Nb}_3\text{Al}$

### **I-5-4 propriétés électriques :**

Ils ont une liaison métalliques avec une densité électronique souvent anisotrope(caractère partiellement covalent de la liaison),Certains composés de structure A15 sont des supraconducteurs comme le matériau  $\text{Nb}_3\text{Al}$  au dessous de 20 k.

## Référence:

- [1] Intermetallic alloy development, Par National Research Council (U.S.). Panel on Intermetallic Alloy Development, National Research Council.
- [2] Intermetallic Alloy development, Par National Research Council (U.S.). Panel on Intermetallic Alloy Development, National Research Council.1997.
- [3] Structure-property relations in nonferrous metals, Par Alan M. Russell, Kok Loong Lee.
- [4] Modern physical metallurgy and materials engineering, Par R. E.Smallman, Ray J.Bishop.
- [5] Improvement of room temperature ductility of  $L1_2$  intermetallic compound  $Ni_3Al$  by boron addition .Nippon Kinzoku Gakkaichi, Aoki, K, and O Izumi.1979.
- [6] Metal and ceramic matrix composites, Par B.Cantor, Fionn Dunne, Ian Stone, Institute of physics (Great Britain).
- [7] Formation and stability of refractory metal diborides in an  $Fe_3Al$  Matrix, T. Doucakis, K.S. Kumar.
- [8] Memoire Magister, *DEVELOPPEMENT D'UN POTENTIEL INTERATOMIQUE POUR LES SYSTEMES Fe-Al-C et Fe-Al-B*, Par SI ABDELKADER Hayet.
- [9] [www.wikipedia.org](http://www.wikipedia.org).
- [10] Properties of ordered intermetallic alloys: first-principles and approximate methods, M. J. Mehl, D. J. Singh and D. A. Papaconstantopoulos.
- [11] AoKi K .and Izumi O .Improvement in Room Temperature Ductility of the  $L1_2$  Type Intermetallic Compound  $Ni_3Al$  by Boron addition .Nippon Kinzoku Gakkaishi.
- [12] Structure-property relations in nonferrous metals. Par Alan M. Russell, Kok Loong Lee.
- [13]Stoloff N.S.,Liu C.T.and Deevi S.C.Emerging applications of intermetallics.Intermetallics,2000.

- [14] Y. Ikematsu, T. Hanamura, H. Morikawa, M. Tanino and J. Takamura, “Oxidation Resistance Phosphorus-Doped Titanium Aluminides,” *Intermetallic Compounds - Structure and Mechanical Properties (JIMIS-6)*, ed: O. Izumi, JIM, Sendai, Japan, 1991, pp. 191– 195.
- [15] Y. W. Kim, “Intermetallic Alloys Based on Gamma Titanium Aluminide,” *JOM*, July 1989, pp. 24– 30.
- [16] M. Yamaguchi and H. Inui, “TiAl Compounds for Structural Applications,” *Structural Intermetallics*, eds: R. Darolia, J. J. Lewandowski, C. T. Liu, P. L. Martin, D. B. Miracle and M. V. Nathal, TMS, Warrendale PA, 1993, pp. 127– 142.
- [17]: Effect of electron concentration on the phase stability of NbCr<sub>2</sub>-based laves phase alloys ,J.H.Zhu,P.K.Liaw,C.T.Liu.
- [18] Sauthoff G .Multiphase intermetallic alloys for structural applications.*Intermetallics*,2000.
- [19] <http://cst-www.nrl.navy.mil/lattice>.
- [20]The mechanics of solder alloy interconnects Par D.R.Frear,Steve Burchett,Harold S .Morgan,John H .Lau

## ***Chapitre II : les techniques d'exploitation des données (Datamining)***

**C**e chapitre est consacré aux techniques d'exploitation des données (DATAMINING): Historique du datamining , Définition , Techniques du datamining (*descriptives , prédictives*), Outils du datamining, les applications du datamining ,les avantages et les inconvénients du datamining et enfin une petite conclusion.

***II-1 Datamining***..... [25 ,29].

***II-2 Conclusion***..... [29].

II-1 DATAMINING :

II-1-1 Historique du datamining :

L'expression (data mining) est apparue vers le début des années 1960. Le datamining est connu en France sous le terme d'ECD (Extraction de Connaissances à partir de Données). En effet, les ordinateurs devenaient 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 à priori statistique 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 tout au contraire prometteurs, ils furent incités à systématiser cette approche opportuniste. Les statisticiens officiels considéraient toutefois cette démarche comme peu scientifique et utilisèrent alors les termes (data mining) ou (data fishing) pour les critiquer.

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 [1].

Le succès de cette démarche empirique ne s'est pas démenti malgré tout. 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 comme par exemple celles des tickets de caisses de grandes surfaces, convaincus de pouvoir valoriser ces masses de données dormantes. Ils utilisèrent l'expression (data base mining) mais, celle-ci étant déjà déposée par une entreprise (Data base mining Workstation), ce fut (data mining) qui s'imposa [2].

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. Actuellement, les termes data mining et knowledge discovery in data bases (KDD, ou ECD en français) sont utilisés plus ou moins indifféremment. Nous emploierons par conséquent l'expression "data mining", celle-ci étant la plus fréquemment employée dans la littérature.

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 créée, 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 [3].

La naissance du data mining est essentiellement due à la conjonction des deux facteurs suivants :

- ✓ l'accroissement exponentiel dans les entreprises, de données liées à leur activité (données sur la clientèle, les stocks, la fabrication, la comptabilité ...) qu'il serait dommage de jeter car elles contiennent des informations-clé sur leur fonctionnement (...) stratégiques pour la prise de décision.
- ✓ les progrès très rapides des matériels et des logiciels.

L'objectif poursuivi par le data mining est donc celui de la valorisation des données [4].

### **II-1-2 Définition du datamining :**

On peut donner plusieurs définitions en ce qui concerne le datamining : Le Datamining ou fouille des données correspond à l'ensemble des techniques et des méthodes qui à partir de données permettent d'obtenir des connaissances exploitables. Son utilité est grande dès lors que l'entreprise possède un grand nombre d'informations stockées sous forme de bases de données [5].

D'après SAS-INSTITUTE qui est une société spécialisée dans le décisionnel et donc le datamining, il s'agit du processus de sélection, exploration, modification et modélisation de grandes bases de données afin de découvrir des relations entre les données jusqu'alors inconnues.

Selon Frawley et Piatetski-Shapiro on peut le définir comme l'extraction d'informations ou de connaissances originales, auparavant inconnues, potentiellement utiles à partir de gros volumes de données [6].

David Hand (1998) en donne la définition suivante: « Data Mining consists in the discovery of interesting, unexpected, or valuable structures in large data sets» [7].

**II-1-3 Techniques du datamining : [1]**

Les principales techniques du Datamining sont :

➤ **Les techniques descriptives :**

Les techniques descriptives (ou exploratoires) visent à mettre en évidence des informations présentes mais cachées par le volume des données. On peut citer quelques techniques disponibles pour ces techniques :

- ✓ L'analyse en composantes principales ACP.
- ✓ Analyse factorielle des correspondances AFC.
- ✓ Analyse des correspondances multiples ACM ; etc. ....

➤ **Les techniques prédictives : [2].**

Les techniques prédictives (ou explicatives) visent à extrapoler de nouvelles informations à partir des informations présentes, ces nouvelles informations pouvant être qualitatives (classement) ou quantitatives (prédiction). Parmi les principales techniques existantes :

- ✓ Régression linéaire simple PLS et multiple.
- ✓ Intelligence artificielle : Réseaux de neurones.
- ✓ Techniques heuristiques modernes : Les algorithmes génétiques.
- ✓ Méthodes Statistiques : Arbres de décision (Méthodes CHAID, AID,...),
- ✓ Inférence bayésienne (Réseau bayésien).

**II-1-4 Outils du datamining : [1,3]**

Les outils du Datamining doivent offrir des richesses analytiques d'un niveau équivalent aux outils statistiques traditionnels.

Ils sont destinés à des utilisateurs "métier" sans compétences statistiques ou informatiques particulières qui doivent répondre à des objectifs opérationnels précis (cibler les prospects, fidéliser les clients,...).

Ils doivent donc offrir une ergonomie, une convivialité, une facilité de mise en oeuvre et d'interprétation les rendant accessibles à des utilisateurs non statisticiens.

Cependant, ces outils ne doivent pas se résumer à de simples "presse-bouton", ils doivent permettre un "déverrouillage de l'automatisation" : laisser les possibilités pour l'utilisateur de saisir ses propres paramètres. Les résultats fournis par l'outil doivent être clairs et compréhensibles : ils ne doivent pas contenir trop de termes techniques statistiques par exemple.

Les outils de Datamining sont généralement utilisés par deux catégories de personnes, l'analyste (le statisticien, le consultant, l'analyste financier,...) et l'utilisateur final (en général ni statisticien, ni informaticien).

L'analyste sait comment interpréter les données, pratique occasionnellement la programmation mais n'est pas un programmeur. Il appréhende le Datamining de manière épisodique et spécifique : pour mener une étude marketing par exemple.

L'utilisateur final possède une connaissance empirique des données qu'il traite. Cette connaissance suffit souvent à diriger l'outil Datamining qu'il utilise. Ainsi, sans connaissance particulière en informatique ou en statistique, l'utilisateur final peut utiliser en continu une solution Datamining adaptée à ses besoins.

L'informaticien est à même de développer des outils de Datamining même s'il ne possède que quelques connaissances sommaires sur les statistiques. Par exemple, toutes les méthodes issues de la théorie de l'information ont un contenu statistique faible alors que l'algorithmique est souvent très important. L'informaticien pourra aussi utiliser un outil de Datamining de la même manière qu'un utilisateur final.

### **II-1-5 Les application du datamining : [6,7].**

Les applications du Datamining sont multiples, elles concernent: la grande distribution, la vente par correspondance, les opérateurs de télécommunications, les banques et assurances, etc. Le domaine majeur où le Datamining a prouvé son efficacité est la gestion de la relation client (CRM ou Customer Relationship Management). En effet, le Datamining permet par une meilleure connaissance de la clientèle d'accroître les ventes.

Le Datamining est donc devenu nécessaire car :

- ✓ Les volumes de données sont trop importants pour un traitement à l'aide de techniques d'analyses classiques
- ✓ L'utilisateur final n'est pas en général statisticien de métier.

**II-1-6 Avantages et inconvénients du datamining : [8].**

➤ **Avantage:**

- ✓ Obtention de connaissances exploitables.
- ✓ Possibilité d'exploitation pour un utilisateur final ayant peu de connaissances en informatiques.
- ✓ Le Datamining peut s'appliquer dans beaucoup de domaines.

➤ **Inconvénient :**

- ✓ Efficace à condition de posséder beaucoup d'informations stockées dans des bases de données.

**II-2 Conclusion:**

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. Elle présente l'avantage de trouver des corrélations informelles entre les données.

## Références :

- [1] The elements of statistical learning - Data Mining, Inference and Prediction », T. Hastie, R. Tibshirani, J. Friedman, Springer 2001.
- [2] Le Data mining », R. Lefebure et G. Venturi, ed. Eyrolles, 2001. Peu technique, point de vue général, très bon recul, complet.
- [3] Minami K., Nakajima H., Mark, R., and Toyoshima, T., "*Realtime discrimination of ventricular tachyarrhythmia with fourier-transform neural network*", IEEE Trans. Biomed Eng., vol. 46, no 2, Feb 1999.
- [4] Gilbert Saporta Data mining: une nouvelle façon de faire de la statistique?.
- [5] Besse, P., and Ferre, L., "*Sur l'usage de la validation croisée en analyse en composantes principales*", Revue de Statistique Appliquée, XLI (1), pp. 71-76, 1993.
- [6] Stéphane Tufféry - Data Mining - <http://data.mining.free.fr>.
- [7] "Data Mining: Concepts and Techniques" by Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers, August 2000. 550 pages. ISBN 1-55860-489-8
- [8] « Data Mining: Practical machine learning tools and techniques with Java implementations », I. Witten and E. Frank, Morgan Kaufman Pub., 2000.

## ***Chapitre III : Analyse en composantes principales ACP et la Régression PLS***

**C**e large chapitre est consacré aux principes de l'analyse en composante principale ACP. L'identification du modèle ACP repose, généralement, sur deux étapes, la première consiste à déterminer la structure du modèle alors que la seconde consiste en l'estimation des paramètres. La deuxième étape est très simple et se ramène à un calcul de valeurs et vecteurs propres, aussi le principe de la régression PLS (Partial Least Square). Un petit aperçu sur l'histoire et le développement de la méthode. Ensuite, une interprétation du modèle PLS est présentée. Quelques critères pour le choix du nombre de composantes sont ensuite présentés pour déterminer la bonne complexité du modèle.

**III-1 Analyse En Composante Principale(ACP) ..... [32,41].**

**III-8 La Régression De Moindre Carrée Partiel(PLS). . . . . [42,52].**

**III-9 Conclusion..... [53].**

**III-1 L'Analyse en composantes principales :****III-1-1 Introduction :**

L'Analyse en composantes principales ("*Principal Component Analysis*"), appelée aussi «analyse géométrique des données » ou « analyse des corrélations », en abrégé ACP est sans doute la méthode d'analyse des données la plus connue et la plus utilisée , qui fut introduite en 1901 par K. Pearson et développée par H. Hotelling en 1933. Nécessitant d'importants calculs numériques, l'ACP n'est devenue une technique opérationnelle qu'à partir des années 1960, avec le développement des moyens de calcul informatique [1].

L'Analyse en composantes principales permet d'effectuer une représentation simplifiée d'une série de variables intercorrélées. Cette technique utilise comme principe une transformation des variables initiales en de nouvelles variables non corrélées. Ces nouvelles variables sont appelées les Composantes Principales. Chaque Composante principale est une combinaison linéaire des variables initiales [2].

L'ACP a pour objet de résumer de grands ensembles de données quantitatives. Ces données sont rangées dans un tableau comportant un grand nombre d'individus et/ou un grand nombre de variables et la simple lecture de ce tableau ne permet pas de saisir l'essentiel des informations qu'il contient [3].

**III-1-2 Définition :**

L'analyse en composantes principales (ACP) est une méthode descriptive qui permet d'effectuer une représentation simplifiée d'une série de variables intercorrélées. Cette technique utilise comme principe une transformation des variables initiales en de nouvelles variables non corrélées. Ces nouvelles variables sont appelées les Composantes Principales. Chaque composante principale est une combinaison linéaire des variables initiales. La mesure de la quantité d'informations représente sa variance. Les variances associées à chaque composante principale sont classées par ordre décroissant. La composante principale la plus informative est donc la première, et la moins informative la dernière [4]. La méthode est basée sur l'hypothèse qu'il existe de fortes corrélations entre les données étudiées. On passe d'un certain nombre de variables potentiellement corrélées à un plus petit nombre de variables non corrélées, les "Composantes Principales". La 1ère "Composante Principale" absorbe le plus de variance possible, la 2ème "Composante Principale" absorbe le plus

de variance possible parmi la variance restante, etc.. L'ACP Permet d'analyser des données multivariées et de les visualiser sous forme de nuages de points dans des espaces géométriques [5].

L'analyse en composantes principales est utilisée pour réduire la dimension (le nombre de variables) d'un problème. Cette diminution du nombre de variables doit s'effectuer en perdant un minimum d'informations. Le but de l'Analyse en Composantes Principales (ACP) est de condenser les données originelles en de nouveaux groupements, appelées nouvelles composantes [2 ,4].

L'ACP se divise en étapes :

- ✓ Normalisation des données pour être indépendants des unités des P paramètres.
- ✓ Calcul d'une matrice de similarité C (bien souvent la corrélation).
- ✓ Recherche des éléments propres de C, qui donnent les axes principaux.
- ✓ Représentation des individus dans le nouvel espace (en ne considérant que les valeurs propres) [6].

L'A.C.P. permet d'explorer les liaisons entre variables et les ressemblances entre individus.

#### **Les applications de L'ACP sont très nombreuses ; ainsi :**

➤ En sciences expérimentales, on dispose souvent d'un ensemble de mesure sous forme d'une population donnée. L'ACP décrit de façon synthétique les individus et décrit de façon synthétique les relations entre les variables, etc.. [5].

#### **III-2 Principe de l'ACP :**

Quand les p variables décrivant les n individus d'une population sont toute numériques, chaque individu peut être représenté par un point dans un espace  $R^p$  à p dimensions. L'ensemble des individus est un « nuage de point ».

#### **➤ Objectif de l'ACP :**

L'ACP propose, à partir d'une matrice (n,p) comportant les valeurs correspondant à n individus sur p variables ,de déterminer quelles combinaisons de variables caractérisent au mieux les individus.

**III-2-1 formulation mathématique d'une ACP :[6,9]**

Soit  $X$  un tableau à  $n$  lignes et  $m$  colonnes. La ligne  $i$  d'écrit la valeur prise par l'individu  $i$  par  $m$  variables quantitatives. Les données sont centrées et réduites, c'est-à-dire que chaque variable à une moyenne nulle et une variance égale à 1.

$$X = \begin{bmatrix} X_{11} & \dots & X_{1j} & \dots & X_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ X_{i1} & \dots & X_{ij} & \dots & X_{im} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ X_{n1} & \dots & X_{nj} & \dots & X_{nm} \end{bmatrix}$$

On note  $X_j$  le vecteur – colonne constitué par les éléments de la colonne  $j$  de  $X$ .  $X_{ij}$  désigne l'élément de  $X$  situé à l'intersection de la ligne  $i$  et de la colonne  $j$ , c'est-à-dire la valeur de l'individu  $i$  pour la variable  $X_j$  [6].

$n$ :objets(ou individus),  $m$  : descripteurs(ou variables).

**III-2-2 Le problème :[10]**

Il s'agit de synthétiser les données contenues dans le tableau  $X$  ; pour cela, on construit un petit nombre de variables nouvelles ;  $C^1, C^2 \dots$  appelées **Composantes Principales** permettant de saisir l'essentiel du tableau  $X$ .

**La première étape :**

On détermine une variable synthétique  $C^1$ , la première composante principale, combinaison linéaire des variables  $X_j$ .

$$C^1 = a_1^1 X_1 + \dots \dots \dots a_j^1 X_j + \dots \dots \dots a_m^1 X_m.$$

Ce qui signifie que la valeur de  $C^1$  pour l'individu  $i$  est donnée par :

$$C_i^1 = a_1^1 X_{i1} + \dots \dots + a_j^1 X_{ij} + \dots \dots + a_m^1 X_{im}$$

La première composante principale ne suffit généralement pas à résumer de façon satisfaisante les données du tableau X.

**L'Etape K :**

On construit la composante d'ordre K,

$$C^k = a_1^k X_1 + \dots + a_j^k X_j + \dots + a_m^k X_m$$

Matriciellement,  $C^K = X a^K$  ou  $a^K$  est un vecteur-colonne à m éléments, l'élément d'ordre j étant égal à  $a_j^K$  ; ce vecteur  $a^K$  est appelé facteur d'ordre K(ou k ième facteur).

Les facteurs fournissent un système de « poids » pour les variables : certains « poids »  $a_j^K$  sont négatifs, d'autres positives ; en fait, ce qui importe n'est pas la valeur de chacun de ces poids, mais le rapport de ces « poids » les uns par rapport aux autres .Si on multiplie un facteur par une constante non nulle, ces rapports sont inchangés .Les facteurs sont donc définis à une constante multiplicative près ; aussi, on impose une contrainte de normalisation pour chacun des facteurs :

$$\sum_{j=1}^m (a_j^k)^2 = 1$$

Les composantes principales sont des variables de moyenne nulle, puisque les variables d'origine sont centrées ; la valeur pour l'individu i de la composante principale k est :

$$C^K = a_1^k X_{i1} + \dots + a_j^k X_{ij} + \dots + a_m^k X_{im}$$

**III-3 La détermination des facteurs et des composantes principales :[1,9]**

L'analyse en composantes principales (ACP) peut être présentée dans deux espaces : celui des individus et celui des variables

**III-3-1 Dans l'espace des individus :**

Dans cet espace, les n individus forment un nuage de points ; les variables étant centrées, l'origine O du repère est le centre du nuage.

L'objet de l'ACP est de décrire de façon synthétique la dispersion du nuage de points.

**A l'étape 1 :** l'ACP détermine l'axe D<sub>1</sub> passant par l'origine selon lequel la dispersion du nuage des points est maximale ; cet axe D<sub>1</sub> passe au (plus près) du nuage des points, c.-à-d. la moyenne des carrés des distances entre les n points et l'axe D<sub>1</sub> est minimale.

**Preuve :**

Les coordonnées de l'individu i sont données par le vecteur (X<sub>i1</sub>, ..., X<sub>ij</sub>, ..., X<sub>im</sub>) et par conséquent la coordonnée de la projection de l'individu i sur l'axe D<sub>1</sub> est :

$$C_i^1 = a_1^1 X_{i1} + \dots + a_j^1 X_{ij} + \dots + a_m^1 X_{im}$$

Qui est le produit scalaire de (X<sub>i1</sub>, ..., X<sub>ij</sub>, ..., X<sub>im</sub>)

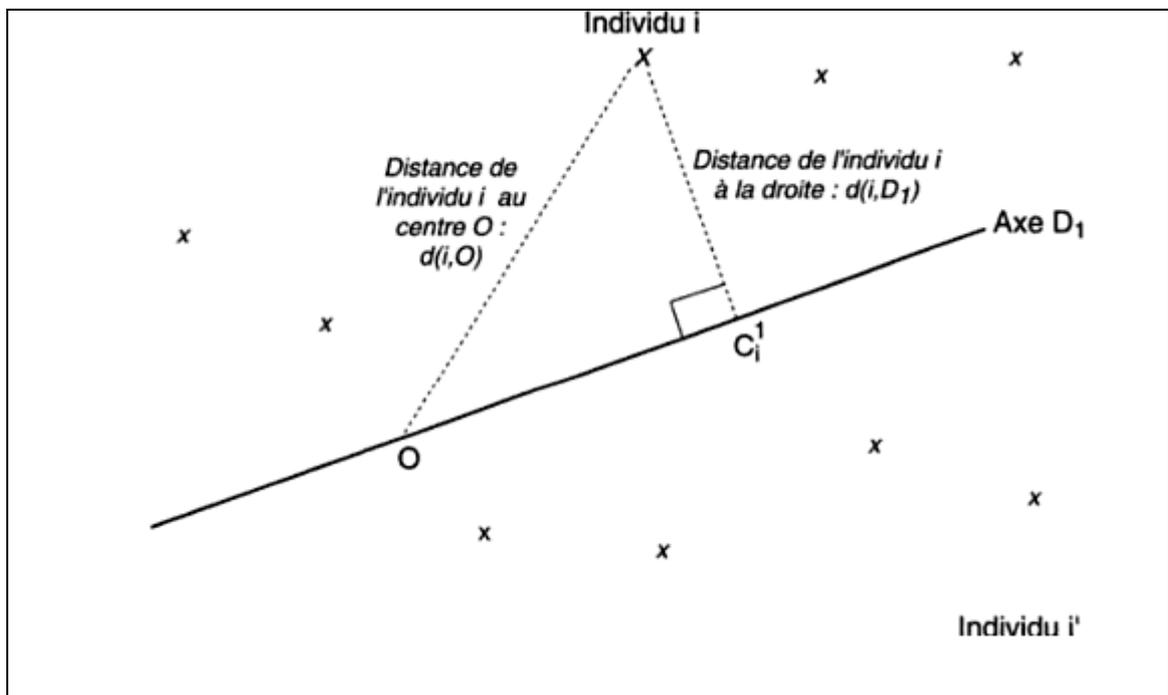


Figure [1-3-1] : Projection d'un individu sur D<sub>1</sub>

$d(i, O)$  désigne la distance entre le point  $i$  et l'origine

$d(i, D_1)$  désigne la distance entre le point  $i$  et l'axe  $D_1$

Cette distance est calculée perpendiculairement à l'axe  $D_1$  d'après le théorème de Pythagore :

$$d^2(i, O) = (C_i^1)^2 + d^2(i, D_1)$$

et par conséquent :

$$\frac{1}{n} \sum_{i=1}^n d^2(i, O) = \frac{1}{n} \sum_{i=1}^n (C_i^1)^2 + \frac{1}{n} \sum_{i=1}^n d^2(i, D_1)$$

Par conséquent,  $C^1$  est le premier vecteur propre de  $1/n XX'$ , donc  $C^1$  est le meilleur résumé possible de l'ensemble des variables de départ ; mais ne décrit pas totalement ces variables et il est nécessaire de calculer d'autres composantes principales.

**A l'étape 2 :** l'ACP détermine  $C^2$  qui doit être aussi le meilleur résumé possibles des variables  $X_j$ , mais en complétant les informations fournies par  $C^1$  : ceci signifie que  $C^2$  doit être non corrélé à  $C^1$ . le problème est donc à l'étape 2 de déterminer  $C^2$  tel que  $\sum_{j=1}^m R^2(C^2, X_j)$  ait une valeur maximale sous la contrainte :  $R(C^1, C^2)=0$  .On continue ainsi de suite et :

**A l'étape k :** l'ACP détermine une variable synthétique  $C^k$  ,résumant le mieux possible les variables de départ, et non corrélée aux(k-1) premières composantes principales, c'est-à-dire détermine  $C^k$  tel que  $\sum_{j=1}^m R^2(C^k, X_j)$  ait une valeur maximale sous les k-1 contraintes :

$$R(C^k, C^r)=0 \text{ pour } r < k$$

$C^k$  est alors le vecteurs propre de  $\frac{1}{n} XX'$  associé à ses k ième valeurs propre la plus élevée.

**Preuve :**

Il suffit de reprendre le même raisonnement à l'étape 1 pour l'étape k.

Notons  $B_k$  la valeur propre d'ordre  $k$  de  $\frac{1}{n}XX'$

Comme  $\frac{1}{n}XX' a^k$  et donc  $\frac{1}{n}XX' C^k = B_k C^k$  car  $C^k = X a^k$

par conséquent,  $B_k$  est aussi la  $k$  ième valeur propre de  $\frac{1}{n}XX'$ , et les deux approches conduisent bien aux mêmes résultats.

**III-4 Le nombre maximal d'étapes : [5]**

Dans l'espace des individus, le nombre maximum d'étapes est  $m$ , puisqu'après  $m$  étapes, on a obtenu une nouvelle base de l'espace des individus : il n'est pas possible de déterminer un vecteur orthogonal aux  $m$  vecteur  $u^1, \dots, u^m$ .

De façon symétrique, le nombre maximal d'étapes dans l'espace des variables est  $n$ .

Par conséquent, le nombre maximal d'étapes est le plus petit des nombres  $n$  et  $m$  soit  $\min(m, n)$ . Généralement  $n$  est supérieur à  $m$ . Généralement  $n$  est supérieur à  $m$  : il ya plus d'observations que de variables.

**III-5 Les valeurs propres : [2,4]**

Des résultats des paragraphes I-3 et I-4 ,on déduit :

$\frac{1}{n}X'X$  et  $\frac{1}{n}XX'$  ont les mêmes valeurs propres  $B_1, \dots, B_m$ ; comme  $\frac{1}{n}XX'$  est une matrice de dimension  $n \times n$  qui a le même rang que  $\frac{1}{n}X'X$ , ses  $n-m$  dernières valeurs propres sont nulles.

Deux autre résultats concernant les valeurs propres sont à noter :

Si  $s$  valeurs propres  $B_{m-s+1}, \dots, B_m$  sont nulles ,alors il existe  $s$  relations linéaires exactes entre les variables.

**Preuve :**

Si  $(\mathbf{a}^k)' X' X \mathbf{a}^k = 0$ , ceci signifie que la variance de la variable  $X \mathbf{a}^k$  est nulle, donc que cette variable est nulle, c'est-à-dire que  $\sum_{j=1}^m a_j^k X_j = 0$ .

D'autre part :  $\sum_{k=1}^m B_k = \text{tr} \left( \frac{1}{n} X' X \right) = m$

La somme des valeurs propres d'une matrice est égale à la trace de cette matrice ; la matrice  $\frac{1}{n} X' X$  est la matrice de corrélations entre  $m$  variables : sa diagonale comporte donc  $m$  éléments qui sont tous égaux à 1.

**III-6 La part de variance expliquée par un axe : [6]**

L'ACP comporte  $m$  étapes ; chaque étape fournit un « résumé » du tableau  $X$  moins intéressant que celui obtenu à l'étape précédente. La question qui se pose comment mesurer la « qualité » du résumé obtenu à l'étape  $k$  ?

*A l'étape  $K$* , le critère de l'ACP dans l'espace des individus est la maximisation sous contrainte de

$$\frac{1}{n} \sum_{i=1}^n (C_1^k)^2$$

A l'optimum :  $\frac{1}{n} \sum_{i=1}^n (C_1^k)^2 = \frac{1}{n} (\mathbf{a}^k)' X' X \mathbf{a}^k = (\mathbf{a}^k)' B_k \mathbf{a}^k = B_k$  Car  $\mathbf{a}^k$  est un vecteur normé.

$B_k$  est la variance du nuage expliquée par l'axe  $k$ , c'est-à-dire la variance de  $C^k$ .

Si on considère les  $m$  axes déterminés par l'ACP,  $\sum_{k=1}^m B_k = m$  et donc :

La part de variance expliquée par l'axe  $k$  est égale à  $B_k/m$ .

Souvent en analyse de données, le terme « inertie » est utilisé pour désigner la variance et on parlera donc de pourcentage d'inertie expliqué par un axe au lieu de part de variance expliquée par un axe.

De façon symétrique, dans l'espace des variables, on remarque que :

$$\sum_{j=1}^m R^2(\mathbf{C}^k, X_j) = \frac{(\mathbf{C}^k)' \mathbf{X} \mathbf{X}'}{n(\mathbf{C}^k)' \mathbf{C}^k} = \frac{(\mathbf{C}^k)' \mathbf{B}_K \mathbf{C}^k}{(\mathbf{C}^k)' \mathbf{C}^k}$$

Et  $\frac{\mathbf{B}_K}{m}$  est aussi la moyenne des carrés des corrélations entre les variables d'origines et l'axe k.

Si  $B_k$  a une valeur faible, cela signifie que la variance expliquée par l'axe k est faible ; par conséquent, l'information apportée par l'étape k est peu utile pour résumer les données de départ.

La variance expliquée étant décroissante pour des étapes successives, l'information apportée aux étapes d'ordre supérieur à k est, elle aussi, peu utile, par conséquent :

Pour résumer le tableau de départ, on n'utilisera donc que le résultat des premières étapes, qui correspondent à des variances expliquées importantes.

### **III-7 Les graphiques et les aides à l'interprétation : [8,3]**

L'information pertinente est donc celle donnée par les premières étapes ; il s'agit maintenant d'analyser simultanément les résultats de ces premières étapes, dans l'espace des individus et dans l'espace des variables.

Cette analyse se fait en établissant des cartes des proximités entre les individus et des cartes des corrélations entre les variables ; ces cartes sont obtenues à partir des résultats des premières étapes, en considérant les projections sur des plans définis par des axes obtenus à deux étapes distinctes.

#### **III-7-1 Représentation des individus : [9,4]**

Si on désire une représentation plane des individus, la meilleure sera celle réalisée grâce aux deux premières composantes principales et pour cela on analyse les résultats fournis par les 2 étapes r et s. La représentation des individus est obtenue en projetant chaque individu i sur le plan engendré par les facteurs  $a^r$  et  $a^s$  : on obtient ainsi une description du nuage de points en projection sur le plan défini par les vecteurs  $a^r$  et  $a^s$ .

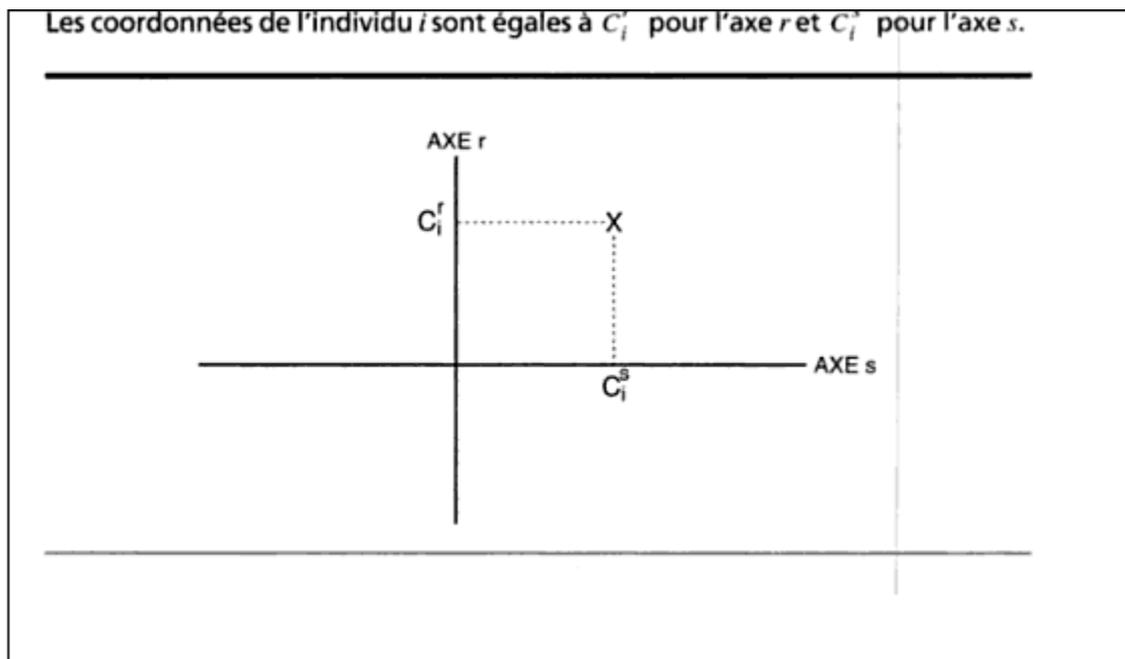
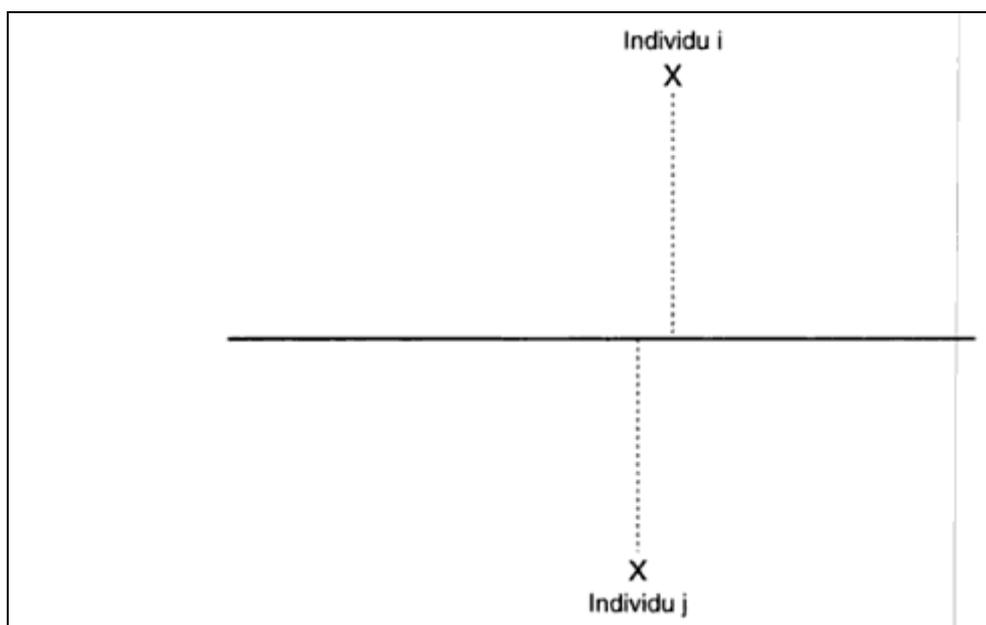


Figure [ ] : représentation des individus.

L'interprétation des résultats consiste à déterminer les caractéristiques du nuage de points décrites par les composantes principales, c'est –à-dire à repérer quelles particularités des individus sont mises en évidence à chaque étape.

Lors de l'interprétation, il faut garder constamment à l'esprit que l'on n'est pas en présence du nuage initial, mais d'une projection de ce nuage et que des erreurs de perspective peuvent être commises :



Si on ne considère que 2 composantes principales  $C^r$  et  $C^s$

$$R^2(X_j, C^r) + R^2(X_j, C^s) \leq 1$$

La somme des carrés des coordonnées de variables  $X_j$  est inférieure ou égale à 1, c'est-à-dire que le point représentatif de  $X_j$  est situé à l'intérieur d'un cercle de rayon 1 et de centre l'origine.

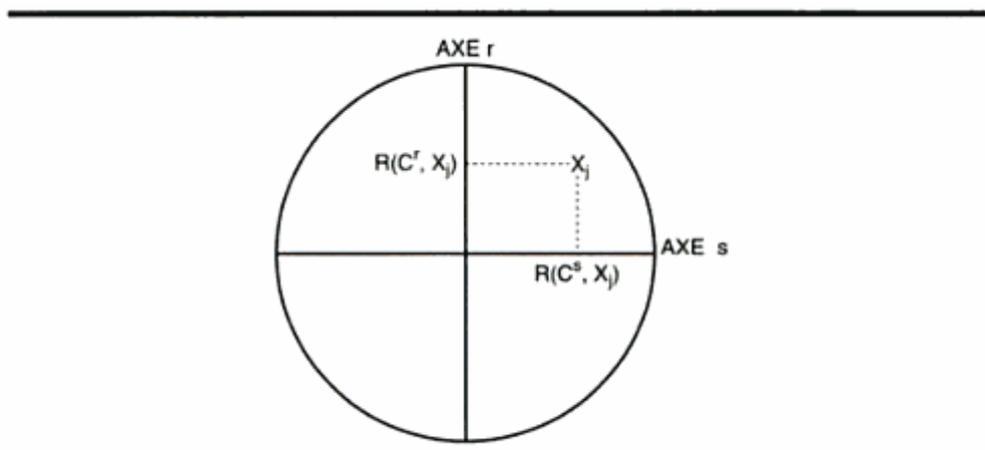


FIGURE 2.D. Cercle des corrélations

- ✓ Si le point  $X_j$  est proche du bord du cercle, ceci signifie que la variable  $X_j$  est très proche du plan défini par les composantes principales  $r$  et  $s$  puisque les coefficients de corrélations avec les autres composantes principales sont alors très faibles. Et donc :
- ✓ Si deux variables  $X_j$  et  $X_j'$  sont proches du bord du cercle, alors l'angle au centre  $G$  entre ces deux variables est proche de l'angle que font ces variables entre elles dans l'espace des variables, et le cosinus de cet angle  $G$  est approximativement égal à leur coefficient de corrélation.

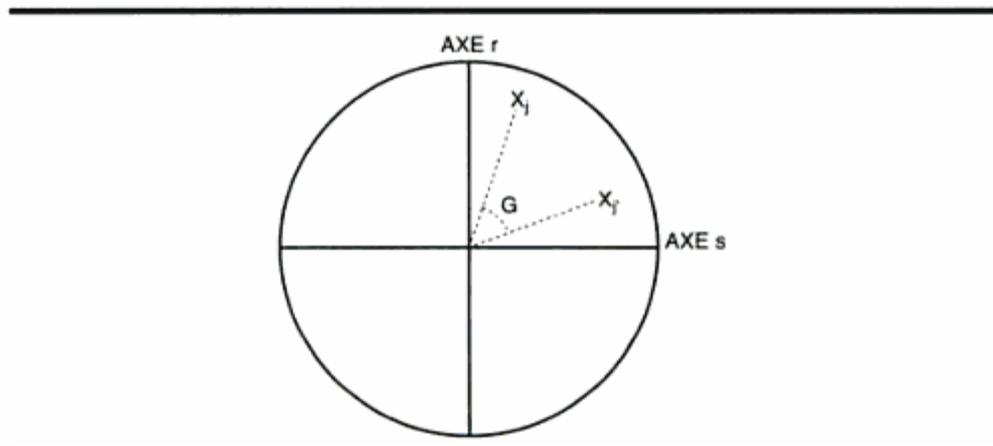


FIGURE 2.E. Angle entre deux variables dans le cercle des corrélations

Éléments en ligne droite

- ✓ Ainsi, soient deux variables proches du bord du cercle ;si elles sont proches l'une de l'autre, alors elles sont très fortement corrélées ;au contraire, si l'angle au centre est un angle droit, ces variables ont une corrélation nulle. Enfin, si ces deux variables sont opposés par rapport à l'origine, elles ont une corrélation proche de -1.
- ✓ La représentation des variables permet ainsi de saisir les relations linéaires existant entre ces variables. En d'autres termes, les cercles de corrélations des axes principaux décrivent l'essentiel de la matrice des corrélations entre les variables.

## Présentation de la régression PLS

### III -9 La régression PLS :

#### III-9-1 Introduction :

L'analyse statistique est un large domaine recouvrant des techniques d'analyse de plus en plus nombreuses. Ces nouvelles techniques se développent continuellement, pour faire face à différents problèmes. Les attentes envers ces analyses sont de plus en plus élevées, et on cherche à les rendre de plus en plus efficaces, et de plus en plus adaptées à des situations concrètes, parfois très spécifiques. Ainsi, lorsque l'on tente d'expliquer une variable par plusieurs autres variables (la première étant la variable expliquée, ou endogène, et les autres étant les variables explicatives, ou exogènes), on ne cherche pas seulement à obtenir un modèle minimisant les erreurs d'estimations des individus actifs (individus à partir desquels le modèle a été construit), on cherche également à obtenir un modèle qui soit facilement interprétable, et qui permette d'effectuer des prévisions sur des individus (ou des entrées) pour lesquels on ne connaît pas la valeur de la variable explicative. Il faut, bien évidemment, que ces prévisions soient les plus proches possibles de la réalité. Il faut également que les modèles soient stables, c'est-à-dire que les chances d'obtenir un modèle trop éloigné de la réalité soient minimales, car on ne peut pas toujours comparer les valeurs estimées aux valeurs réelles, dont on ne dispose pas (à priori), puisqu'on cherche à les estimer. Il faut parfois même que ce modèle remplisse ces conditions alors que l'on dispose de très peu d'individus actifs, alors même que le nombre de variables explicatives est très élevé, ce qui rend pourtant, d'un point de vue théorique, la construction d'un modèle, représentatif de la réalité, très délicate. C'est précisément ce à quoi tente de répondre la régression PLS.

Dans un cadre très général appelé Partial Least Squares (PLS), Herman et Svante Wold ont proposé des méthodes d'analyse des données permettant d'étudier J bloc de variables observées sur les mêmes individus.

La méthode NIPALS ( Nonlinear estimation by iterative Partial Least Squares ), proposée par Wold (1966) , permet d'étudier un seul bloc de variables (J=1). Elle conduit à l'analyse en composantes

principales lorsque les données sont complètes, mais fonctionne également lorsqu'il y a des données manquantes.

La régression PLS permet de relier un bloc de variables à expliquer à un bloc de variables explicatives ( $J=2$ ). Elle a été proposée par Wold, Martens (1983). On obtient les composantes PLS par applications successives de l'analyse factorielle inter-batteries de Tucker (1958). L'utilisation des principes de l'algorithme NIPALS permet le traitement des données manquantes. Il peut y avoir beaucoup plus de variables que d'observations. La régression PLS est sans doute actuellement la meilleure réponse au problème de la multicollinéarité en régression multiple.

Le cas de  $J$  blocs a été étudié dans le cadre de la modélisation de relations structurelles sur variables latentes (Path models with latent variables).

### **III-9-2 Contexte historique :**

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 résultant 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. La méthode PLS répond précisément à ses problèmes [2].

La régression PLS est née de l'association de l'algorithme NIPALS (Non linear Iterative Partial Least Squares) développée par H. Wold (1966) pour l'analyse en composantes principales et de l'approche PLS proposée par H. Wold (1975) pour l'estimation des modèles d'équations structurelles sur variables latentes. Cet algorithme est essentiel à la compréhension de la régression PLS car il permet de comprendre la logique de l'écriture de l'algorithme de régression PLS [3]. En 1966, Herman Wold publie un ouvrage introduisant le PLS, alors appelée Non Linear Iterative Partial Least Squares NIPALS. En 1983, Savante Wold (le fils d'Herman) et Harold Mertens adaptent NIPALS au problème de régression avec trop de prédicteurs et appellent PLS cette adaptation de l'algorithme ; en suite en 1990, Stone et Brooks introduisent le PLS dans le contexte du "Continuum Regression" (elle ajoute un paramètre continu, autorisant la méthode de modélisation à varier continuellement entre MLR "Multiple Linear Regression" ( $\lambda = 0$ ), PLS ( $\lambda = 0.5$ ) et PCR "Principal Component Regression" ( $\lambda = 1$ ). Ce qui signifie une première projection sérieuse

du PLS dans un contexte statistique [2]. On distingue habituellement en régression PLS le cas où il y a une seule variable Y de celui où il y en a plusieurs. La régression PLS1 correspond à la première situation et la régression PLS2 à la seconde. L'algorithme de régression PLS a été proposé initialement par Wold, Martens, Wold (1983) et Wold, Albano, Dunn III, Esbensen, Hellberg, Johansson, Sjöström (1983) [4]. Donc on pose la question : **Qu'est-ce que la régression PLS ?**

### **III-10 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 » [5]. On distingue la régression PLS uni variée, ou « régression PLS1 », de la régression PLS multi variée, 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 (et, même si l'algorithme de la régression PLS multi variée est présenté différemment de celui de la version simple, il constitue une généralisation de ce dernier dans la mesure où les résultats sont équivalents lorsque la régression PLS multivariée ne porte que sur une seule variable expliquée) [4,5]. La régression PLS s'inscrit dans la catégorie des régressions linéaires. Dans ce cas on peut poser deux questions : qu'est-ce que une régression linéaire ? et dans quel cas en utilise-t-on cette régression ?

#### **III-10-1 La régression linéaire :**

La régression linéaire est un outil (mathématique) statistique qui permet de définir une loi linéaire entre deux variables intervenant dans un même phénomène, elle est appliquée lorsque deux variables quantitatives continues semblent liées l'une à l'autre par une relation linéaire. Une relation linéaire entre  $x$  et  $y$  (les variables) est de la forme  $y = a.x + b$  où  $a$  et  $b$  sont des réels. Le but de la régression est donc d'expliquer les valeurs et les variations d'une ou plusieurs variables expliquées (les «  $y$  ») par les valeurs et les variations d'une ou plusieurs variables explicatives (les «  $x$  »). En effet, le but premier de la régression n'est pas de s'intéresser à un individu en particulier, mais à un individu « abstrait », pour lequel les relations entre les variables sont des relations valables « en moyenne », peu importe les valeurs prises par les variables explicatives. Cette régression linéaire a un objectif simple : trouver les coefficients, pour chaque variable explicative, qui minimisent les

écarts, pour la variable expliquée, entre les valeurs estimées par le modèle, et les valeurs observées dans la pratique, pour l'échantillon donné sur lequel est effectué la régression. Il s'agit de minimiser la somme des résidus (mis au carré, dans le simple but d'éviter la compensation systématique des erreurs positives et négatives), ou, dit autrement, de maximiser le coefficient de corrélation.

### **III-10-2 les avantages de la régression PLS :[10 ,12,14]**

Pourquoi on utilise la régression PLS ? Qu'est-ce qu'elle apporte de plus que la régression linéaire ?

Les avantages de la régression PLS sont nombreux :

- La régression PLS peut traiter des cas où les individus seraient moins nombreux que les variables explicatives. La régression linéaire ne peut le faire.
- Dans le cas de régression PLS multi variée (régression PLS2), il peut y avoir plusieurs variables expliquées.
- La régression PLS, étant basée sur l'algorithme NIPALS, permet de travailler sur des échantillons même si certaines données manquent pour certains individus pour certaines variables, et ce sans même avoir à estimer au préalable les données en question.
- Lorsque les variables explicatives sont fortement corrélées entre-elles, la régression linéaire devient très peu pertinente, au sens où les coefficients qui en ressortent deviennent très instables lorsque l'on « bruite » les données (on fait varier, de manière aléatoire et très légère, les données de l'échantillon). La régression PLS, basée sur des critères de covariance, est considérée comme étant plus robuste. Les coefficients demeurent stables et gardent une certaine significativité, même en présence de corrélations fortes entre les variables.

### **III-9-3 La méthode NIPALS[9 ,11]**

L'algorithme NIPALS (Non linear estimation by Itérative Partial Least Square) est une application de la formule de reconstruction des données en Analyse en Composantes Principales (ACP) adaptée au cas de données manquantes (Tenenhaus).

Étant donnée une matrice centrée de taille  $n \times p$  notée  $X=(X_{ij})$  et de rang  $a$  la formule de décomposition de l'analyse en composantes principales,  $X = \sum_{h=1}^a t_h p_h'$ , est caractérisée par les

vecteurs  $t=(t_1, \dots, t_1, \dots, t_1)'$  et  $p=(p_1, \dots, p_1, \dots, p_1)'$  appelés respectivement les composantes principales et les vecteurs directeurs des axes principaux. Les variables  $x_j$  s'expriment en fonction des composantes  $t_1, \dots, t_a$  :

$$X_j = \sum_{k=1}^a P_{kj} t_k \quad j=1, \dots, P$$

L'algorithme NIPALS donne une estimation des composantes principales et des axes principaux utilisant de manière itérative des régressions linéaires. La formule de reconstitution d'ordre  $h$  ( $h \leq a \leq p$ ) permet alors d'estimer les données manquantes par :

$$X_{ij} = \sum_{k=1}^h t_{ik} p_{kj}$$

La complexité de cet algorithme est d'ordre  $O(a \times n \times p \times C)$ , où  $C$  est le nombre maximal de boucles nécessaires pour obtenir la convergence dans le calcul des axes principaux.

**III-11 Comparaison[6]**

Comparé à d'autres méthodes de régression pour des données colinéaires, il a été établi que le plus grand avantage de PLSR est que "l'information dans la variable Y est utilisée", mais c'est une notion qu'il est difficile d'établir statistiquement. Un avantage plus évident est que la méthode rend possible la combinaison de la prédiction avec l'étude d'une structure jointe latente dans les variables X et Y. Ainsi, la méthode demande souvent moins de composantes que PCR pour donner une bonne prédiction. Avec la facilité d'implémentation, cela a rendu cette méthode très populaire en chimie, en particulier en spectroscopie, mais les applications dans d'autres domaines sont aussi très répandues.

**III-11-1 La Régression Linéaire Multiple (MLR) : [1,4]**

Si le nombre de facteurs A extrait est plus grand ou égal au rang de l'espace de l'échantillon X, alors PLS est équivalent MLR.

**III-11-2 La régression sur composantes principales : [1,4]**

PLS est une technique quantitative de décomposition spectrale étroitement liée à la régression sur composantes principale (PCR). Cependant, dans PLS, la décomposition est faite d'une manière légèrement différente.

Au lieu de décomposer d'abord la matrice spectrale  $X$  en un ensemble de vecteurs propres et de scores, et de les régresser contre les  $Y$  dans une étape séparée, PLS utilise l'information de  $Y$  en même temps que le processus de décomposition.

Ceci implique que les variables expliquant le mieux les  $Y$  seront plus fortement pondérées. De plus, les vecteurs propres et les scores calculés en utilisant PLS seront différents de ceux de PCR. L'idée principale de PLS est de donner le plus d'information possible sur  $Y$  dans les premiers vecteurs construits.

### **III-11-3 La régression Ridge :[2]**

La régression Ridge est une technique originaire des statistiques permettant de manipuler la colinéarité en régression. Elle est probablement la plus grande rivale de PLS en terme de flexibilité et de robustesse du modèle prédictif. Mais cette méthode n'a pas de procédure de réduction de dimension.

### **III-12 Intérêt de la Régression PLS :[6,8]**

La régression PLS pour le modèle linéaire s'applique au champ d'application 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).

#### **A- Établissement du modèle PLS simple**

La régression PLS présente de nombreux points communs avec la RCP. Elle comprend également une étape de calcul d'une matrice  $T$ , dont les éléments sont les *scores* et les colonnes les *composantes*, obtenue par l'application d'une relation de la forme :

$$T = XW \quad (1)$$

Avec

$W$  matrice des poids ou loadings

$X$  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 :

$$X = TP + R \tag{2}$$

$$Y = Tq + F \tag{3}$$

Avec

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

$F$ , vecteur des résidus associé à la prédiction de  $y$ .

La première étape consiste à calculer  $t_1$ , la première composante. On estime ensuite les paramètres des modèles de (2) et (3) à une seule composante, soit :

$$X = t_1 p_1 + R_1 \text{ et } y = t_1 q_1 + f_1 \tag{4}$$

Avec

$t_1$ , de dimension  $n \times 1$

$p_1$ , de dimension  $1 \times m$

$q_1$ , de dimension  $1 \times 1$  (nombre).

Le deuxième modèle associé à  $X$  ainsi obtenu indique qu'une ligne  $x_i$  d'indice  $i$  de  $X$  est égale à la somme de deux « signaux purs » représentés par les vecteurs  $p_1$  et  $p_2$  pondérés par les valeurs  $t_{1i}$  et  $t_{2i}$ , soit :

$$x = t_{1i} p_1 + t_{2i} p_2 \tag{5}$$

L'introduction de nouvelles composantes dans le modèle se fait selon la même procédure : ayant un modèle à  $k$  composantes (ou  $k$  dimensions), on crée un nouveau modèle à  $k + 1$  dimensions en calculant tout d'abord une nouvelle composante  $t_k$ , puis les paramètres des deux modèles couplés :

$$X = t_1 p_1 + t_2 p_2 + \dots + t_k p_k + t_{k+1} p_{k+1} + R_{k+1} \tag{6}$$

$$y = t_1 q_1 + t_2 q_2 + \dots + t_k q_k + t_{k+1} q_{k+1} + f_{k+1} \tag{7}$$

Ce qui s'écrit, de manière équivalente, sous les formes matricielles (6), et (7) .

À une étape k donnée, la composante tk est déterminée à partir des résidus R<sub>k-1</sub> et f<sub>k-1</sub> de l'étape précédente.

**B -La méthode :**

La méthode PLS est une méthode de régression linéaire de c variables réponses sur p variables explicatives toutes mesurées sur les mêmes n individus. Les tableaux des observations, notés respectivement Y et X, de dimensions n×c et n× p, sont supposés centrés et éventuellement réduits par rapport aux poids (p1, . . . , pn). On note D = diag(p1, . . . , pn) la matrice diagonale des poids [7].

L'intérêt de la méthode comparée à la régression sur composantes principales (RCP), réside dans le fait que les composantes PLS sur les X, notées t, sont calculées "dans le même temps" que des régressions partielles sont exécutées. Cette simultanéité leur confère un meilleur pouvoir prédictif que celles de la RCP. La question est donc d'examiner comment cette simultanéité est mise en œuvre.

Notons E<sub>0</sub> = X et F<sub>0</sub> = Y les tableaux centrés et réduits au sens de D qui en général est égal à  $\frac{1}{n} I_n$ .

La méthode procède par étapes successives permettant le calcul des composantes principales. On notera A le nombre total d'étapes, c'est à dire de composantes indicées par k=1, 2,.....A .

**Description de la kieme étape :**

Notons  $t = E_{k-1}w$  et  $u = F_{k-1}q$  les combinaisons linéaires colonnes des matrices centrées E<sub>k-1</sub> et F<sub>k-1</sub>, associées respectivement aux vecteurs des poids w et q. La covariance entre t et u s'écrit comme le D-produit scalaire

$$cov(t, u) = (t, u)_D = w^t E_{k-1}^T D F_{k-1} q$$

Le carré de la D-norme associée fournit la variance  $\|t\|^2 = var(t)$

L'étape k, se décompose en deux parties :

La première partie : fournit les composantes  $t_{k-1} = E_{k-1}w_k$  et  $u_k = F_{k-1}q_k$  par le calcul des poids optimaux  $w_k$  et  $q_k$ .

$$(w, q) = \arg \max cov(t, u) = w^t E_{k-1}^t D F_{k-1} q$$

Sous les contraintes  $\|w\|^2 = \|q\|^2 = 1$

la deuxième étape : actualise les matrices des prédicteurs et des réponses  $E_k$  et  $F_k$  comme résidus de la régression sur  $t_k$ .

$$E_k = E_{k-1} - P_{t_k} E_{k-1} \text{ et } F_k = F_{k-1} - P_{t_k} F_{k-1}$$

Où  $P_{t_k} = \frac{t_k t_k^t}{var(t_k)}$  est la matrice  $n \times n$  de projection D-orthogonale sur  $t_k$ .

**C-Propriétés des composantes  $t_1, \dots, t_A$**

Le modèle PLS linéaire trouve quelques nouvelles variables. Ces nouvelles variables sont appelées les X-scores notées par  $t_k (k = 1, 2, 3, \dots, A)$ . Ces composantes PLS  $t_1, t_2, \dots, t_A$  sont des combinaisons linéaires des colonnes de  $E_0$ , non corrélées entre elle, résumant au mieux  $E_0$  tout en expliquant autant que possible  $F_0$ . Ces composantes PLS sont donc analogues à des composantes principales des  $X_1, \dots, X_m$  expliquant au mieux les variables  $Y_1, \dots, Y_p$ .

La non corrélation ou D-orthogonalité, mutuelle entre les composantes  $t_1, \dots, t_A$  a de multiples conséquences. On peut montrer ainsi par récurrence que  $t_k$  appartient à  $ImX$  espace vectoriel engendré par les prédicteurs.

La non corrélation implique en outre que  $\sum_{k=1}^{k-1} P_{t_k} = P_{t_k} = P_{T_A}$ , ou  $P_{T_A}$  est le projecteur orthogonal sur la matrice  $T_A = [t_1, \dots, t_A]$ .

Enfin, dernière conséquence de la non corrélation, la décomposition des variances totales des réponses et des prédicteurs fournit deux critères pour le choix du nombre A de composantes ;

**III-13 Interprétation du modèle PLS**

Une façon à voir PLS, c'est qu'il forme des nouvelles variables  $x, (t_k)$ , qui sont des combinaisons linéaires des  $x$  et qui sont utilisées comme prédicteurs de  $Y$ . Ainsi PLS est basé sur un modèle linéaire.

Tout 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 []. 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 des corrélations positives entre  $X$  et  $Y$  et les corrélations nécessaires pour prédire  $Y$  à partir de  $X$ .

### **III-13-1 Choix du nombre de composantes PLS :**

En tout modèle empirique, il est essentiel de déterminer la bonne complexité du modèle. Avec un grand nombre de variables  $X$  corrélées, il y a un fort risque d'ajustement du modèle, c'est-à-dire obtenir un modèle bien ajusté avec peu ou pas de pouvoir prédictif. Par conséquent, un strict critère de l'importance prédictive de chaque composante PLS est nécessaire, et ça s'arrête quand les composants commencent à être non significatives. La borne supérieure de  $A$  sera évidemment le rang de la matrice de départ, c'est à dire

$A = \text{rang}(X)$ . On aura des critères permettant de déterminer  $A$  de façon raisonnable

#### **Critères du choix de $A$ : [11]**

- ❖ **Critères 1** : Ajustement ("FIT") sur les données.
- ❖ **Critères 2** : Prédiction interne ("Validation Croisée").
- ❖ **Critères 3** : **Prédiction** externe sur un jeu de données test ("Validation Externe").

On ne veut pas avoir un sur ajustement donc le FIT ne devra pas être trop grand ( $A$  trop grand) car cela dégradera la prédiction. Donc il faut une balance entre le critère "ajustement" et le critère "prédiction" (interne ou externe).

#### **A-Critère d'Ajustement [14]**

Le nombre total  $A$  d'axes ou de composantes (la dimension du modèle) est le super-paramètre de la méthode. Lors de l'exécution de l'algorithme, il est utile de voir évoluer, à chaque nouvel axe construit, la reconstruction de la variance des prédicteurs et des réponses. La stratégie consiste à s'arrêter lorsque le pourcentage de la variance de  $X$  est suffisamment grand, pour un gain faible dans le pourcentage de variance de  $Y$ . Ce critère est donc basé sur l'ajustement des données

**B-Critère de Validation Croisée[12]**

La validation croisée est un critère très populaire pour le choix du nombre de composantes dans un modèle PLS. La base de cette méthode est d'estimer les mesures d'un jeu de données de validation à partir d'un modèle qui a été calculé à partir d'un jeu de données d'identification et de comparer ces estimations avec les valeurs mesurées.

Cette procédure de validation croisée est basée sur la minimisation de la quantité *PRESS* (*Prediction Sum of Squares*). Cette quantité se présente comme la somme des carrés des erreurs entre les données observées et celles prédites ou estimées par le modèle obtenu à partir d'un jeu d'identification différent.

Après avoir enlevé un individu  $i$  (une ligne) aux matrices  $X$  et  $Y$  (ou 10% des individus si leur nombre est élevé), on calcule  $B_A^{(i)}$  qui représente la matrice des coefficients du modèle construit avec les individus restants. Puis, on calcule l'erreur de prédiction faite sur l'individu  $i$

$$E_A^{(i)} = Y_i - X_i B_A^{(i)}$$

L'erreur de prédiction est définie par le PRESS (*Prediction Sum of Squares*) :

$$PRESS(A) = \frac{1}{n} \sum_{i=1}^n \|E_A^{(i)}\|^2$$

On choisit donc  $A$  tel que  $PRESS(A)$  soit le plus petit possible .

**C-Critère 3 : critère de validation externe et prédiction : [12 ,13]**

Pour utiliser ce critère de choix de  $A$  , on doit posséder outre l'échantillon d'apprentissage (  $X$ ,  $Y$  ) sur les individus, un échantillon test (  $X_{test}$ ,  $Y_{test}$  ) sur  $N$  observations mesurées sur les mêmes variables. Différents modèles PLS bases sur (  $X$ ,  $Y$  ) sont calculés pour différentes dimensions. La valeur optimale de  $A$  correspond à la plus petite des erreurs de prédiction calculée sur (  $X_{test}$ ,  $Y_{test}$  ).

**III-9 Conclusion :**

L'Analyse en Composantes Principales (PCA) et des moindres carrés partiels(PLS) qui sont des techniques informatiques d'analyse des données.

L'Analyse en Composantes Principales (Principal Component Analysis) est une méthode statistique de projection de données multidimensionnelles, utilisée pour la réduction de la dimensionnalité des données. Le but de l'Analyse en Composantes Principales (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. Ainsi, la première nouvelle composante contient les informations relatives au pourcentage maximal de variance, la deuxième contient les informations relatives au pourcentage de variance suivant. Le processus est répété jusqu'à l'obtention de la dernière nouvelle composante.

La méthode des moindres carrés partiels PLS (Partial Least Squares), qui est une méthode prédictive de régression linéaire à plusieurs variables. Nous a permis d'obtenir une loi linéaire reliant des variables réponses aux variables explicatives .

## Référence :

### POUR L'ANALYSE EN COMPOSANTES PRICIPALES ACP

- [1] Analyse en Composantes Principales (PCA-principal component analysis) Nicolas Turenne INRA.
- [2] Pierre-Louis Gonzalez : Analyse en Composantes Principales.
- [3] Analyse en composantes principales de l'échelle cognitive de Bumby (1996) *Véronique Muschang*<sup>1</sup>, *Joanne L. Rouleau*<sup>2</sup>, *Katia Lavallée*<sup>3</sup>
- [4] Hotelling K (1933). Analysis of a Complex of Statistical Variables Into Principal Components, *Journal of Educational Psychology* pp. 417{441.
- [5] J. Pages, B. Escofier : Analyses factorielles simples et multiples : Objectifs, m'ethodes et interpr'etation. DUNOD, 1990.
- [6] Etude de données d'expression par Analyse en Composantes Principales Fonctionnelles BARRA VINCENT
- [7] Kaiser H (1958). The Varimax Criterion for Analytic Rotation in Factor Analysis., *Psychometrika* 23, 187.
- [8] P.L. Gonzalez, R.Cleroux et B.Rioux. « Selecting the best subset of variables in principal component analysis » *COMPSTAT* 1990 Physica-verlag Heidelberg for IASC. (International Association for statistical computing). 1990
- [9] Analyse en Composantes Principales (PCA-principal component analysis) Nicolas Turenne INRA.2006
- [10] Analyse en Composantes Principales C. Duby, S. Robin D'épartement O.M.I.P.10 Juiller 2006
- [11] Exploration Statistique .ALAIN BACCINI & PHILIPPE BESSE.Version Juin 2007.

**POUR LA REGRESSION PLS**

- [1] S. Wold, A. Ruhe, H. Wold, W.J. Dunn, "The collinearity problem in linear regression", Computational Statistics Data Analysis 5 (1984) 735–743.
- [2] A. Hoskuldsson, "Variable and subset selection in PLS regression", Chemometrics and Intelligent Laboratory Systems 55 (2001) 23–38.
- [3] A. J. Burnham, J. F. MacGregor, R. Viveris, " Latent variable regression tools", Chemometrics and Intelligent Laboratory Systems 48 (1999) 167–180.
- [4] H. Abdi, (in press 2003), "Partial Least Squares (PLS) Regression", In M. Lewis-Beck, A.T. Bryman, Futing (Eds) : Encyclopedia for Research methods for the Social Sciences. Thousand Oaks (CA) : Sage.
- [5] R. Manne, "Analysis of two partial least squares algorithms for multivariate calibration" Chemometrics and Intelligent Laboratory Systems 1 (1987) 187–197.
- [6] Jolliffe I.T., "Principal component analysis", Springer-Verlag, New York, 1986
- [7] Hotelling H., "Analysis of a complex of statistical variables into principal components", Journal of Educational Psychology, Vol. 24, pp. 417- 441 1933.
- [8] S. Wold, M. Sjostrom, L. Eriksson, "PLS-regression : a basic tool of chemometrics ", Chemometrics and Intelligent Laboratory Systems, 58 (2001) 109-130.
- [9] Tenenhaus M., La régression PLS Théorie et pratique. Editions Technip, 1998.
- [10] M. TENHAUS, J.-P. GAUCHI, C. MENARDO. Régression PLS et applications ,Revue de statistique appliqué ,tome 43, n°1(1995), p.7-63 .
- [11] Marx B.D. (1996) : Iteratively Reweighted Partial Least Squares Estimation for Generalized Linear Regression. Technometrics, vol. 38, n°4, pp. 374-381.
- [12] Bastien P., Tenenhaus M. (2001) : PLS generalized linear regression. Application to the analysis of life time data. In PLS and Related Methods, Proceedings of the PLS'01 International Symposium, Esposito Vinzi V., Lauro C., Morineau A. & Tenenhaus M. (Eds). CISIA-CERESTA Editeur, Paris, p. 131-140.
- [13] Höskuldsson A. (2001) : Variable and subset selection in PLS regression. Chemometrics & Intelligent Laboratory Systems, 55, 23-38.
- [14] Himes D.M, and al., "Determination of the number of principal component for disturbance detection and isolation", Proceedings of ACC, Baltimore, 1994.

## Chapitre IV: Résultats et discussions

**L**es chapitres précédents concernent une base de données sur les intermétalliques(chapitre I),les techniques techniques d'exploitation des données (DATAMINING) (chapitre II), la troisième chapitre consacré aux principe de l'analyse en composante principale ACP dans un première étapes ,dans une deuxième étape est très simple et se ramène à un calcul de valeurs et vecteurs propres, aussi le principe de la régression PLS (Partial Least Square).Dans cette partie nous présentons, les résultats liés à l'utilisation de l'analyse en composantes principales (ACP) et la régression (PLS),appliquées sur les intermétalliques ,afin de prédire de nouveaux matériaux.

<i>IV-1 Introduction</i> .....	[59]
<i>IV.3 Technique de calculs (Applications de L'ACP)</i> .....	[59,67]
<i>IV.4 Le modèle PLS (Partial Least Squares)</i> .....	[68,108]

**IV-1 Introduction :**

Afin de concevoir de nouveaux matériaux, l'identification des paramètres régissant le caractère de la liaison dans les intermétalliques sera d'une grande utilité académique. Dans cette partie nous présentons, les résultats liés à l'utilisation de l'Analyse en composantes principales (ACP) et la régression (PLS), appliquées sur un ensemble de composés intermétalliques. Les calculs ont été faits avec ces deux techniques en utilisant le logiciel XLSTAT.

**IV-2 La base de données :**

Dans le cadre de cette étude, nous avons construit une base de donnée comprenant : la distance interatomique ( $d$ ), les constantes élastiques ( $C_{ij} \dots$ ), le module de cisaillement ( $G$ ), le module de compression ( $B$ ), le module de Young ( $E$ ) et la température de fusion  $T_f$ . La base de données est représentée sur le tableau IV-2-a.

**Tableau [IV-2-a] :** Base de données utilisée dans ce travail.

	$T_f$ (°K)	$d$ (Å)	$E$ (GPa)	$B$ (GPa)	$G$ (GPa)	$B/G$ (GPa)	$C_{11}$ (GPa)	$C_{12}$ (GPa)	$C_{44}$ (GPa)	$C_{12} - C_{44}$ (GPa)
<b>FeAl</b>	1330	2,5	150	136	82	1,65	181	114	127	-13
<b>Fe<sub>3</sub>Al</b>	1540	2,51	140	158	19	8,31	184	145	160	-15
<b>NiAl</b>	1640	1,31	170	166	82,74	2,00	211	143	112	31
<b>Ni<sub>3</sub>Al</b>	1390	1,4	180	171	77,00	2,22	221	146	124	22
<b>TiAl</b>	1460	3,46	175	109,78	74,8	1,46	125	99	122	-23
<b>Ti<sub>3</sub>Al</b>	1600	1,66	146	106	57,00	1,86	176	87,8	62,4	25,4
<b>Al<sub>3</sub>Ti</b>	1350	1,92	156	110	69,00	1,594	177	77	85	-8
<b>MoSi<sub>2</sub></b>	2020	2,71	439,7	209,7	191,1	1,097	417	104,2	204,2	-100
<b>VSi<sub>2</sub></b>	1677	2,5	342,6	167,2	147,9	1,13	357,8	50,6	135,7	-85,1
<b>NbSi<sub>2</sub></b>	1923	2,58	362,8	151	153,2	0,985	384	75,9	146,8	-70,9
<b>TaSi<sub>2</sub></b>	1922	1,86	359	192,5	151	1,274	375,3	78,4	143,7	-65,3
<b>CrSi<sub>2</sub></b>	1490	2,47	354,6	172	153,3	1,12	372,2	45,3	149,1	-103,8
<b>WSi<sub>2</sub></b>	2150	1,5	467,9	222,4	203,6	1,092	442,8	121,7	211,6	-89,9
<b>TiSi<sub>2</sub></b>	1500	2,57	278,1	146,8	116,9	1,594	278,6	140	877	-737

Référence [2], Référence [4] .

Référence [7], Référence [8].

Le traitement de ce tableau peut être fait sous **XLSTAT** qui permet d'appliquer sur **EXCEL** les méthodes d'analyse des données.

**IV.3 Technique de calculs :****IV.3.1 application de L'analyse en composantes principales (ACP) :**

Le premier résultat intéressant à identifier suit à une analyse en composantes principales est la matrice des corrélations qui permet d'analyser les relations bilatérales existant entre les différentes variables retenues. Cette matrice est représentée sur le tableau IV-3-b.

**Tableau [IV-3-b] : Matrice de corrélation**

Variables	T <sub>f</sub>	D	E	B	G	B/G	C <sub>11</sub>	C <sub>12</sub>	C <sub>44</sub>	C <sub>12</sub> -C <sub>44</sub>
T <sub>f</sub>	1	-0,135	0,823	0,727	0,766	-0,235	0,806	-0,107	-0,021	0,002
D	-0,135	1	0,069	-0,261	0,031	0,039	-0,058	-0,274	0,202	-0,260
E	0,823	0,069	1	0,779	0,973	-0,457	0,974	-0,351	0,180	-0,251
B	0,727	-0,261	0,779	1	0,734	-0,087	0,814	0,112	0,074	-0,058
G	0,766	0,031	0,973	0,734	1	-0,628	0,940	-0,375	0,158	-0,232
B/G	-0,235	0,039	-0,457	-0,087	-0,628	1	-0,395	0,471	-0,048	0,134
C <sub>11</sub>	0,806	-0,058	0,974	0,814	0,940	-0,395	1	-0,350	0,143	-0,212
C <sub>12</sub>	-0,107	-0,274	-0,351	0,112	-0,375	0,471	-0,350	1	0,337	-0,174
C <sub>44</sub>	-0,021	0,202	0,180	0,074	0,158	-0,048	0,143	0,337	1	-0,986
C <sub>12</sub> -C <sub>44</sub>	0,002	-0,260	-0,251	-0,058	-0,232	0,134	-0,212	-0,174	-0,986	1

➤ La seconde étape consiste à diagonaliser la matrice de corrélations. Dans le cas présent, la matrice est de dimension 10 x 10, elle admet donc 9 valeurs propres positives ou nulles. Cependant, seules les premières valeurs propres ont un intérêt, les autres n'ayant qu'un faible pouvoir descriptif.

➤ La matrice de corrélation permet d'observer la relation positive forte entre la température de fusion et le module de compression B (0.727). Aussi une forte corrélation entre le module de Young et la température de fusion (0.823). On constate aussi la relation négative entre le module de cisaillement G et le rapport B/G (-0.628). forte corrélation entre G et le module de Young E (0,973). Toutes ces corrélations entre variables vont conditionner la composition des axes factoriels dont le sens et la signification s'interpréteront en fonction de leur corrélation avec chaque variable.

➤ La diagonalisation de la matrice de corrélation donne les valeurs propres présentées dans le tableau [IV-3-c]. Chaque facteur est caractérisé par une valeur propre, une part de la variance totale (% de variance), et une part de la variance cumulée (% cumulée).

- ✓ La deuxième ligne indique les valeurs propres de la matrice de corrélation.
- ✓ La troisième ligne nous renseigne sur le pourcentage expliqué par chaque valeur propre.

**Tableau [IV-3-c] :** Les valeurs propres de la matrice de corrélations

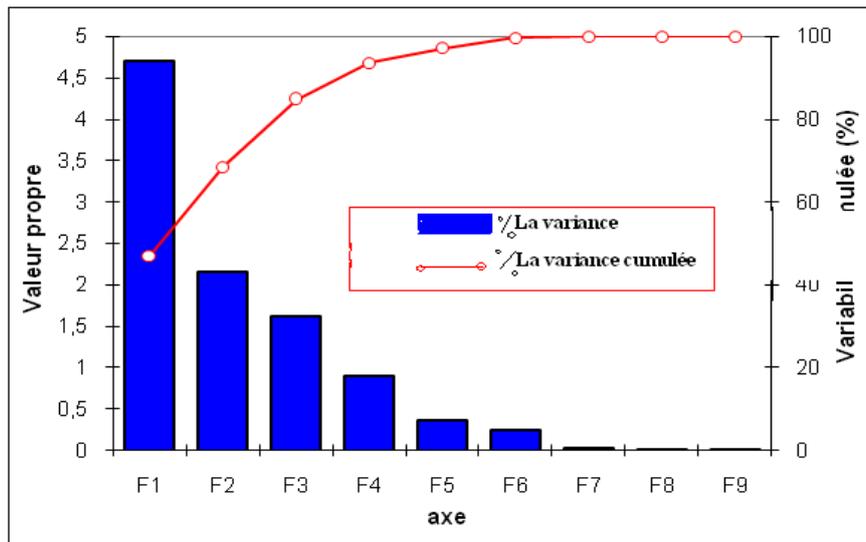
Qualité de l'ACP	F1	F2	F3	F4	F5	F6	F7	F8	F9
Valeur propre	4,696	2,156	1,619	0,889	0,353	0,250	0,024	0,011	0,003
% de Variabilité	46,963	21,559	16,190	8,891	3,528	2,500	0,235	0,105	0,028
% cumulé	46,963	68,522	84,712	93,604	97,131	99,631	99,867	99,972	100,0

➤ Ce tableau signifie que le premier axe (F1) permet d'expliquer (46,96%) de la variance totale du nuage de points, que le second axe (F2) permet d'expliquer (21,55) de la variance totale. En projetant donc chaque individu sur un plan (F1, F2) on conserve 46,96+21,55 soit 68,52% de la variance totale.

Remarquons ici que la variance expliquée par l'axe F3 est faible (16,19 %), et remarquons aussi que l'essentiel de l'information est fourni par les 2 premiers axes, puisque la variance expliquée cumulée par est de 68,52%. Les 2 axes sont donc retenus ici.

De façon plus générale, pour décider s'il ya lieu de retenir ou non un axe, il sera tenu compte à la fois du pourcentage de la variance expliquée par l'axe, et du pourcentage cumulé de la variance expliquée par cet axe et les précédents. Si les variables sont très corrélées entre elles, les premières valeurs propres sont importantes alors que si le lien linéaire entre les variables d'origines est faible, les pourcentages de variance expliquée par les axes seront faibles eux aussi.

➤ La figure IV-3-c nous permet d'évaluer les valeurs propres de manière schématique, La première composante principale (PC1) correspond à la plus grande valeur propre, ainsi la deuxième composante (PC2) correspond à la deuxième plus grande valeur propre. Généralement, l'étude se restreint à ces deux premiers axes, du fait qu'ils contiennent la majorité de la variance, le pourcentage de variance absorbée par les axes principaux est indiqué sur la figure (IV-3-c).



**Figure (IV-3-c) :** Histogramme des valeurs propres. (Pourcentages de variance absorbée, pour les 9 axes principaux).

Les axes sont indépendants les uns des autres et se croisent au centre d'un nuage de point correspondant aux individus statistiques (les composés intermétalliques) disséminés dans un espace dont le nombre de dimensions est égal au nombre de variables (dans ce cas 9 dimensions).

A partir du tableau et du graphique (IV-3-c), on remarque que ces valeurs sont décroissantes de F1(4,696) à F9(0,003) et qu'à partir F4 elles sont inférieures à 1. Remarquons aussi que les valeurs propres sont bien différenciées, ce qui traduit une hiérarchisation nette de l'information.

**(%)Variance** : correspond à la part de variance totale prise en compte par le facteur : à partir de l'ensemble F<sub>1</sub> : on remarque qu'il existe un rapport de 1 à 9 entre la valeur propre (4.696) et la part de variance (46.963). La part de variance totale prise en compte par un facteur est en effet définie comme le rapport entre la valeur propre qui caractérise le facteur et la somme des valeurs propres : pour F<sub>1</sub> cela donne (%)Variance = 4.696 / 10 = 46.96% ce qui signifie qu'à lui seul, le premier facteur résume près de 50% de l'information totale.

Plus généralement :

$$(\%)Variance(F_K) = R_K = U_K / I$$

**R<sub>K</sub>** est la part de variance totale prise en compte par le **K<sup>ième</sup>** facteur.

**U<sub>K</sub>** est la valeur propre associée au **K<sup>ième</sup>** facteur.

**I** est l'inertie totale du nuage de points

**(%)Cumulé :** correspond à la part de variance totale prise en compte jusqu'au facteur K. Pour le facteur ou axe factoriel  $F_1$ , la part de variance (4.696) et la part de variance cumulée sont identiques. En revanche dès  $F_2$ , elles sont différentes (%) Variance= 21.559 tandis que (%) cumulée=68.522 ce qui signifie que dès le premier facteur plus de la moitié de l'information est représentée. Si les 3 premiers facteurs sont retenus – autrement dit tous ceux dont la valeur propre est supérieure à 1, alors près de 85% de l'information totale sera représentée.

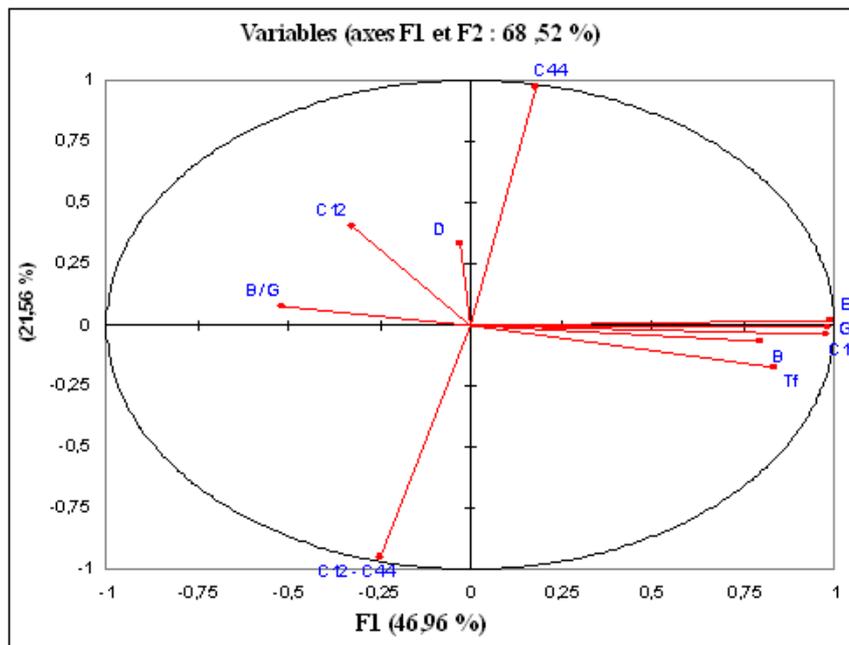
- Un autre résultat intéressant c'est le graphe des variables « loadings plot » sur la Figure [IV-3-d] (Ce graphique donne la représentation des variables dans le plan). Le graphe des variables, nous donnent les informations sur les relations qui existent entre les propriétés.
- Pour confirmer le fait qu'une variable est fortement liée à une composante principale, il suffit de consulter le tableau suivant de cosinus carrés :

	PC <sub>1</sub>	PC <sub>2</sub>	PC <sub>3</sub>	PC <sub>4</sub>	PC <sub>5</sub>	PC <sub>6</sub>	PC <sub>7</sub>	PC <sub>8</sub>	PC <sub>9</sub>
T <sub>f</sub>	0,698	0,030	0,071	0,027	0,022	0,150	0,002	0,000	0,000
D	0,001	0,108	0,396	0,427	0,063	0,006	0,000	0,000	0,000
E	0,979	0,000	0,002	0,013	0,000	0,000	0,001	0,004	0,001
B	0,632	0,004	0,259	0,017	0,000	0,078	0,008	0,000	0,000
G	0,967	0,000	0,013	0,003	0,007	0,006	0,001	0,001	0,002
B/G	0,269	0,006	0,266	0,372	0,084	0,002	0,000	0,001	0,000
C <sub>11</sub>	0,952	0,001	0,002	0,005	0,027	0,001	0,007	0,004	0,000
C <sub>12</sub>	0,104	0,161	0,599	0,001	0,130	0,002	0,003	0,000	0,000
C <sub>44</sub>	0,033	0,948	0,001	0,012	0,004	0,002	0,000	0,000	0,000
C <sub>12</sub> -C <sub>44</sub>	0,061	0,897	0,011	0,012	0,017	0,003	0,001	0,000	0,000

**Tableau [IV. 3.e] :** Cosinus carrés des variables

- ✓ Si le cosinus carrés est élevé, une variable est plus liée à l'axe. Dans notre cas l'axe PC<sub>1</sub> est lié à G, C<sub>11</sub>, E, T<sub>f</sub> et B avec les cosinus carrés G(0,96), C<sub>11</sub>(0,95), E(0,97), T<sub>f</sub>(0,69) et B(0,63), l'axe PC<sub>2</sub> est lié à C<sub>44</sub>, C<sub>12</sub>-C<sub>44</sub>, avec des cosinus carrés C<sub>44</sub>(0,94), C<sub>12</sub>-C<sub>44</sub> (0,89), et contrairement une variable est moins liée à l'axe si son cosinus carrés est proche de zéro. Dans ce cas la distance interatomique d, le rapport B/G et C<sub>12</sub> sont peu liés à l'axe PC<sub>1</sub>.
- ✓ Si le cosinus carrés est proche de "1" on a une très bonne représentation.
- ✓ Si le cosinus carrés est supérieur à "0,5" on a une représentation acceptable.
- ✓ Si le cosinus carrés est inférieur à "0,5" on a une mauvaise représentation.

- Sur la figure [VI.3.d] On peut constater que l'axe PC1 est fortement corrélé positivement à Tf ,B,G,E, aussi il est fortement corrélé négativement à B/G. tandis que l'axe PC2 est fortement corrélé positivement à d ,d'autre part il est fortement corrélé négativement à C11-C44.
- On remarque que l'axe PC1 concentre à lui seul presque de la moitié de l'ensemble de l'information (46,96%); et L'axe PC2 regroupe (21,56%) de la variance.



**Figure [IV.3.d]** : graphe des variables pour les intermétalliques correspondant aux Observations, chaque point correspond à une propriété.

On peut remarquer que, la distance interatomique  $d$  et  $B$  sont inversement corrélés, ce qui nous permet de dire que les matériaux avec une petite distance interatomique sont incompressibles.

On outre le module de cisaillement " $G$ " et le module de Young " $E$ " sont fortement liés ceci nous permet de dire que les matériaux durs sont rigides, aussi, il est claire que " $G$ " et " $B/G$ " sont inversement corrélés. Etant donné que le module du cisaillement ( $G$ ) est un critère de mesure de dureté, et le rapport ( $B/G$ ) est un critère de mesure de ductilité, donc ceci confirme que si la dureté augmente la ductilité diminue.



La liaison interatomique dans les composés intermétalliques affecte leurs modules d'élasticité, ce dernier augmente avec l'augmentation de la température de fusion. Le module d'élasticité des intermétalliques des disilicides est localisé dans la partie supérieure ce qui indique qu'ils ont une température de fusion plus élevée que celle des autres intermétalliques. Ce qui implique qu'ils sont plus rigides que les autres.

✓ L'ACP nous a permis aussi de faire, une autre représentation graphique en faisant une partition de l'ensemble des individus en fonction de leurs coordonnées sur les deux axes. La figure [IV.3.g] correspond au graphe des observations pour les 14 intermétalliques différents.

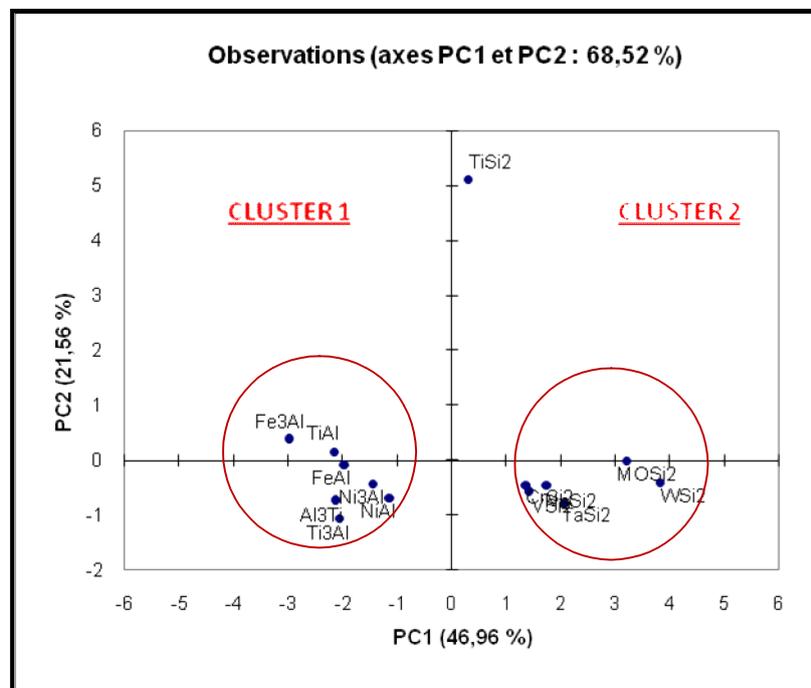


Figure [IV.3.f] : graphe des observations (scores plot).

En première vue, il s'avère que l'ensemble des intermétalliques se séparent en deux groupes (cluster). De plus on constate que le cluster 1 avec un PC1 négatif correspond à des intermétalliques à base d'aluminium [Fe<sub>3</sub>Al, TiAl, FeAl, Ni<sub>3</sub>Al, NiAl, Al<sub>3</sub>Ti, Ti<sub>3</sub>Al] qui ont un petit module de cisaillement G, un petit module de Young E et une petite température de fusion. Aussi les matériaux de ce cluster sont fortement liés au module de Cauchy qui est un

indicateur de la ductilité, en effet, les matériaux ductiles ont des valeurs positives de pression de Cauchy, qui correspondent à une liaison métallique plus isotrope.

En ce qui concerne le groupe 2, avec un PC1 positif regroupe des intermétalliques à base de silicium [ $\text{MoSi}_2$ ,  $\text{CrSi}_2$ ,  $\text{VSi}_2$ ,  $\text{NbSi}_2$ ,  $\text{TaSi}_2$ ,  $\text{WSi}_2$ ] avec un grand  $G$ , un grand  $E$ , une température de fusion élevée, et un module de Cauchy plus négatif donc ces matériaux sont fragiles. Cependant nous remarquons l'isolement du  $\text{TiSi}_2$  par rapport aux deux clusters. Ce qui pourrait aussi classer le  $\text{TiSi}_2$  comme un composé particulier. En effet la particularité de cet élément réside dans la faible valeur de la pression de Cauchy. Ce qui attribue une forte directionnalité de la liaison à la limite de la liaison covalente. Ceci peut être vérifié en fracture de la structure cristalline d'équilibre du  $\text{TiSi}_2$ .

✓ Ces clusters sont compatibles avec les différentes valeurs de  $B$  et de  $G$  comme il est montré dans la figure suivante qui représente les variations de  $B$  en fonction de  $G$  et montre les valeurs de dureté et de ductilité.

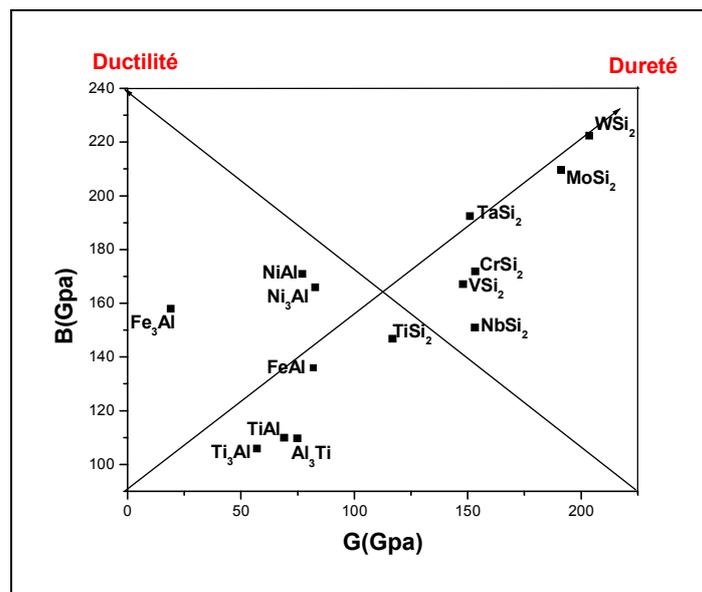


Figure IV-2-f : Valeurs des modules de cisaillement et de compressibilité

Ce que l'on peut retenir de cette analyse concernant les intermétalliques c'est que quelques propriétés dépendent généralement du type de liaison établie entre les atomes. Ainsi la fragilité sera associée aux matériaux covalents et ioniques alors que la ductilité sera une conséquence directe de la liaison métallique. Nous remarquons aussi que le  $\text{TiSi}_2$  présente un compromis entre dureté et ductilité.

## IV.4 Le modèle PLS (Partial Least Squares) :

### IV.4.1 Application de la régression PLS :

Régression PLS : méthode de régression linéaire permettant de relier un ensemble de variables à prédire à un ensemble de variables explicatives quand le nombre de variables à expliquer et ou explicatives est grand par rapport au nombre d'individus. Cette méthode est notamment adaptée à la construction de modèles lorsque les variables explicatives sont fortement corrélées entre elles. elle permet en effet d'obtenir une certaine cohérence entre les coefficients de corrélation et les coefficients de régression. dans notre cas la régression pls permet de trouver la relation entre les propriétés élastiques (le module de compressibilité B, le module de cisaillement G ,les constante élastiques ...),et les propriétés mécaniques (module de Yong ...)

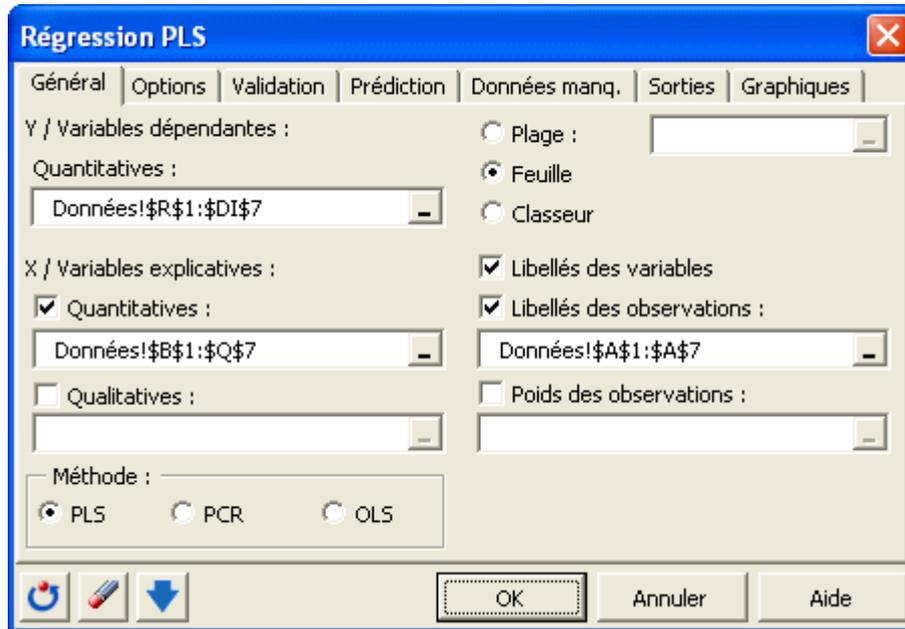
En applique la méthode PLS sur des ensembles de données pour chaque composés intermétalliques, nous avons relié la matrice X des variables prédictors (variables indépendantes : les composés intermétalliques) à la matrice Y (variables dépendantes : que sont les propriétés de ces composés).

Pour activer la boîte de dialogue de la régression PLS, lancez XLSTAT en cliquant sur le bouton de la barre d'outils Excel, puis sélectionnez la commande Régression PLS de la barre d'outils "modélisation des données ".

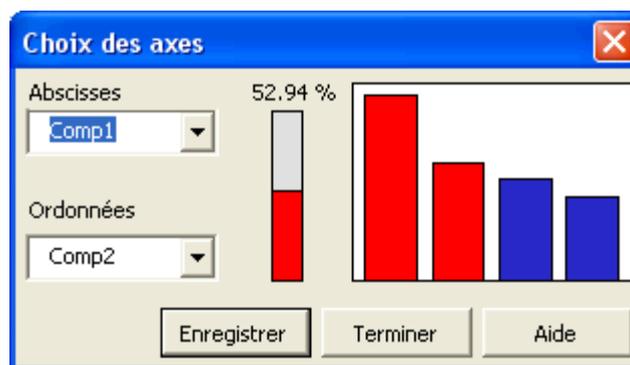


Une fois que vous avez cliqué sur le bouton, la boîte de dialogue apparaît. Sélectionnez au niveau des "variables dépendantes"(les "Y" du modèle), les propriétés des composés intermétalliques .Ce sont en effet les données que l'on veut expliquer au travers des "variables explicatives quantitatives "(les "X" du modèle)" que sont les composés intermétalliques .Puis

Cliquez sur l'onglet "Options" et vérifiez que "Automatique" est bien la condition d'arrêt sélectionnée. Enfin ,dans l'onglet "Graphiques"



Les calculs, extrêmement rapides, commencent lorsque l'on a appuyé sur "OK". L'affichage des résultats est interrompu pour vous permettre de choisir les axes de représentation des résultats.



Il suffit de cliquer sur "Terminer" pour que les graphiques concernent uniquement les deux premiers axes. L'affichage des résultats prend ensuite environ 2 minutes du fait de la présence de 14 variables dépendantes et des très nombreux tableaux et graphiques affichés.

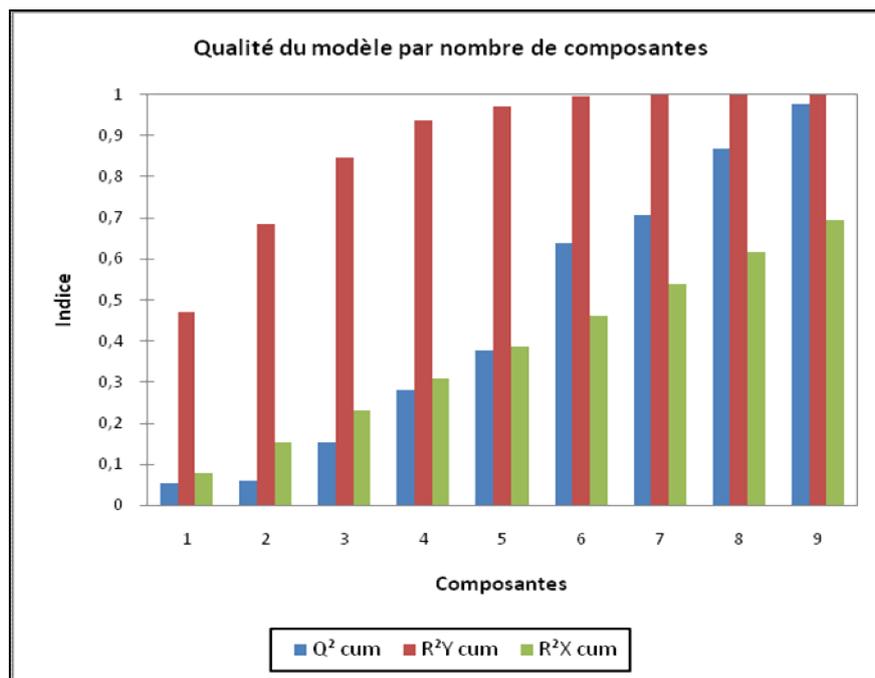
Après quelques statistiques de base sur les différentes variables sélectionnées (les variables explicatives sont en rouge, et les variables dépendantes en bleu) et la matrice de corrélations correspondante, les résultats propres à la régression PLS sont affichés.

#### IV.4.2 Qualité du modèle :

Le premier tableau et le graphique correspondant permettent de visualiser la qualité de la régression PLS en fonction du nombre de composantes retenues.

**Tableau [IV.4 .2 .a]:** les différents composants retenue par la régression PLS, et les valeurs des indices correspondants.

Indice	Compo1	Compo2	Compo 3	Compo4	Compo5	Compo6	Compo7	Compo8	Compo9
$Q^2 \text{ cum}$	0,052	0,061	0,154	0,281	0,376	0,638	0,706	0,869	0,978
$R^2 Y \text{ cum}$	0,470	0,685	0,847	0,936	0,971	0,996	0,999	1,000	1,000
$R^2 X \text{ cum}$	0,077	0,154	0,231	0,308	0,385	0,462	0,538	0,615	0,692

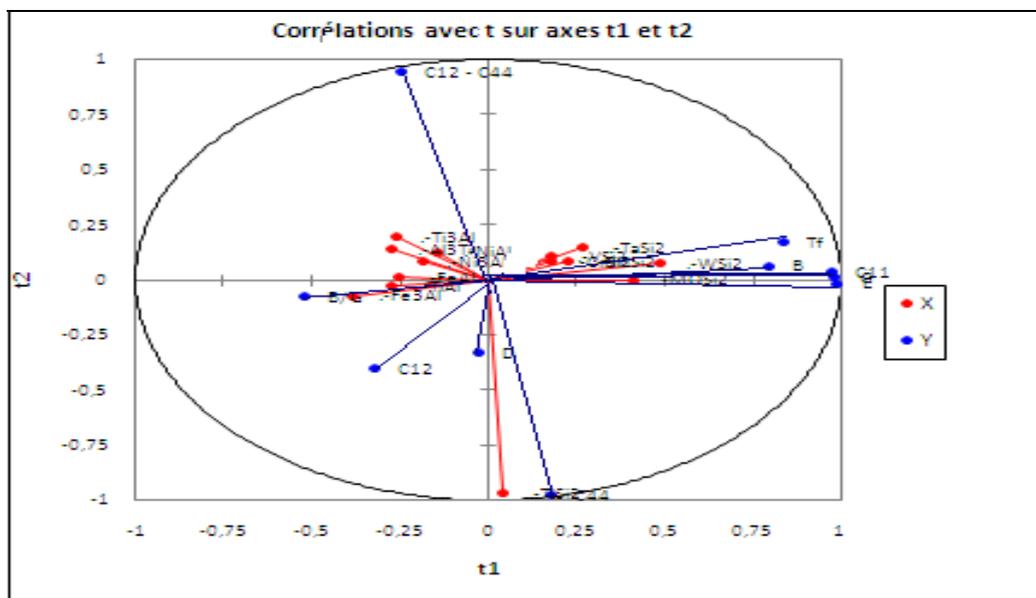


**Figure [IV.4 .2.a]:** qualité de la régression PLS en fonction du nombre de composante retenues .

➤ L'indice "Q<sup>2</sup>" cumulé est une mesure globale de la qualité de l'ajustement et de la qualité prédictive des 14 modèles. XLSTAT-PLS a automatiquement retenu 9 composantes. On voit que l'indice Q<sup>2</sup> varie considérablement. Cela suggère que la qualité de l'ajustement peut être très variable en fonction des propriétés du matériau.

➤ Les "R<sup>2</sup>Y" cum et "R<sup>2</sup>X" cum qui correspondent aux corrélations entre les composantes et les variables de départ sont proches de un dès la quatrième composante, ce qui indique que les composantes sont à la fois bien représentatives des "X" et des "Y".

✓ Le premier graphique des corrélations permet de visualiser sur les deux premières composantes générées par la régression PLS les corrélations entre "X" et les "Y" du modèle.



**Figure [IV.4 .2.b]:** Représentation des observations à partir des composantes t<sub>1</sub> et t<sub>2</sub> de la régression PLS

Le cercle des corrélations de la figure [IV.4.2.b] traduit bien les corrélations à chaque groupe de variable X et Y.

✓ Les variables dépendantes "(les "Y" du modèle), sont globalement bien corrélées avec les variables explicatives "(les "X" du modèle)"(les composés intermétalliques).

✓ On notera une forte corrélation positives entre le module de Yong E, le module de compressibilités B , la température de fusion T<sub>f</sub> ,et le module de cisaillement G ,et une corrélation bien étendue négative entre les variables [D , B /G,C<sub>12</sub>,C<sub>12</sub>-C<sub>44</sub>].Remarquant que les variables explicatives [TaSi<sub>2</sub>,WSi<sub>2</sub>,MoSi<sub>2</sub>,VSi<sub>2</sub>,NbSi<sub>2</sub>] sont bien liées aux variables [B

,E,T<sub>f</sub>,G,C<sub>11</sub>] ce qui indique que ces matériaux sont à la fois durs et rigides . [FeAl , Fe<sub>3</sub>Al, NiAl,Ni<sub>3</sub>Al,TiAl,Ti<sub>3</sub>Al, Al<sub>3</sub>Ti] sont liées aux variables [B /G,D,C<sub>12</sub>,C<sub>12</sub>-C<sub>44</sub>] sachant que B /G<1,75 et un pression de Cauchy C<sub>12</sub>-C<sub>44</sub><0 ce qui indique la fragilités de ces matériaux . En ce qui concerne les variables explicatives, on note que seule l'intermétallique TiSi<sub>2</sub> est mal représentée sur le graphique. Ce qui se comprend bien puisque elle n'a pas d'implication directe sur les critères influençant.

#### IV.4.3 Les vecteurs du poids W et C :

**Tableau [IV.4.3] :** Poids des variables explicatives dans la construction des composantes PLS.

variable	w <sub>1</sub> *	w <sub>2</sub> *	w <sub>3</sub> *	w <sub>4</sub> *	w <sub>5</sub> *	w <sub>6</sub> *	w <sub>7</sub> *	w <sub>8</sub> *	w <sub>9</sub> *
FeAl	-0,242	0,015	-0,112	-0,046	0,246	-0,332	-0,125	0,225	0,602
Fe <sub>3</sub> Al	-0,363	-0,068	0,435	0,703	-0,307	0,073	-0,035	-0,043	-0,045
NiAl	-0,140	0,123	0,357	-0,298	0,178	0,014	0,251	0,242	0,279
Ni <sub>3</sub> Al	-0,180	0,078	0,334	-0,287	0,065	-0,440	-0,242	0,105	-0,612
TiAl	-0,268	-0,029	-0,393	0,250	0,570	0,026	0,232	-0,308	-0,277
Ti <sub>3</sub> Al	-0,254	0,191	-0,072	-0,256	-0,125	0,548	0,035	-0,016	-0,106
Al <sub>3</sub> Ti	-0,265	0,135	-0,212	-0,247	-0,221	0,072	-0,133	-0,372	0,127
MoSi <sub>2</sub>	0,394	0,000	0,042	0,262	0,334	-0,095	0,065	0,085	-0,029
VSi <sub>2</sub>	0,173	0,105	-0,274	0,083	-0,263	-0,101	0,210	0,103	-0,088
NbSi <sub>2</sub>	0,215	0,083	0,194	0,127	0,079	0,379	-0,606	0,442	-0,086
TaSi <sub>2</sub>	0,256	0,142	0,059	-0,037	-0,167	0,113	0,578	0,238	-0,165
CrSi <sub>2</sub>	0,169	0,081	-0,313	0,042	-0,424	-0,443	-0,088	-0,131	-0,037
WSi <sub>2</sub>	0,470	0,074	0,372	-0,095	0,128	0,113	-0,174	-0,598	0,185
TiSi <sub>2</sub>	0,037	-0,932	-0,028	-0,200	-0,094	0,085	0,031	0,028	-0,014

Les poids  $W_{hi}^*$  traduisent l'importance de chaque variable  $X_i$  dans la construction des  $t_h$ , c'est-à-dire des composantes issues de la régression qui résument au mieux les  $X_i$  tout en présentant une bonne capacité à prédire les  $Y_i$ .

Nous remarquons ainsi que les variables explicatives  $X_2, X_8, X_{13}$  sont celles qui interviennent le plus dans la construction de la composante 1.

➤ La carte superposant les variables dépendantes sur les vecteurs "c" et les variables explicatives sur les vecteurs "w\*" permet de visualiser la relation globale entre les variables, sachant que les "w\*" sont représentatifs du poids des variables dans les modèles.

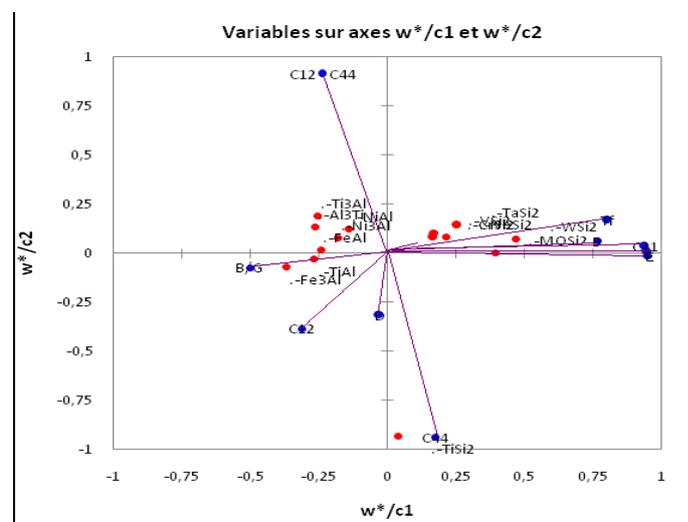


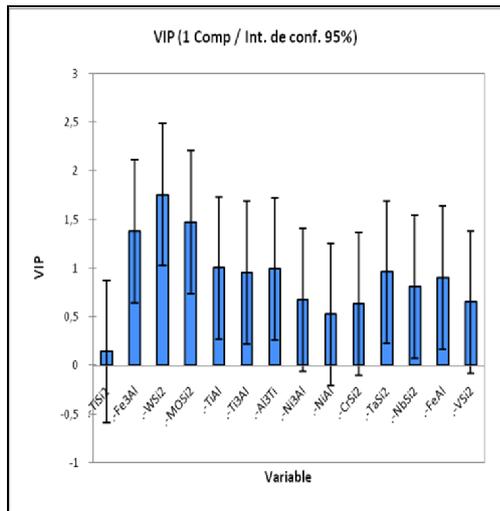
Figure [IV.4.3]: Graphe des variable pour les intermétalliques.

➤ Si l'on projète une variable explicative sur le vecteur d'une variable dépendante ; on a une idée de son poids dans le modèle concernant cette même variable dépendante.

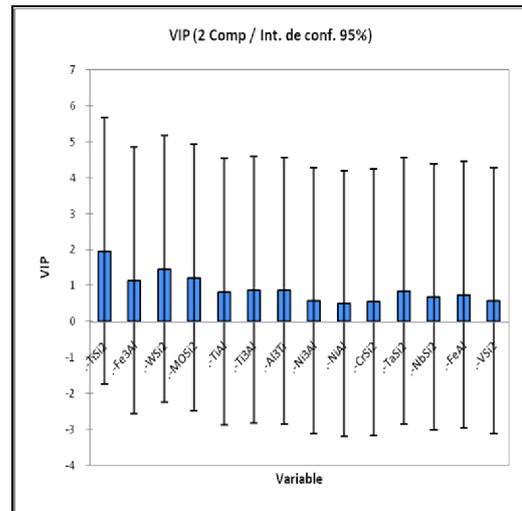
➤ La figure [IV.4.3] montre la première composante PLS1 dominée par les propriétés  $T_f, B, G, E, C_{11}$  dans le sens positifs et la deuxième composante PLS2 est fortement corrélée positivement à  $C_{12}-C_{44}$ . on remarque aussi que B est inversement corrélé avec B/G.

#### IV.4.4 Mesure de l'importance des variables :

Les figures [IV.4.4] (VIP pour les deux première composantes) qui présente les VIP (Variable Importance for the Projection) pour chacune des variables explicatives, sur chacun des modèles avec un nombre croissant de composantes. Cela permet d'identifier rapidement quelles sont les variables explicatives les plus importantes sur l'ensemble des modèles.



La figure [IV.4 .4.a]



La figure [IV.4 .4.b]

➤ La figure [IV.4 .4.a]: sur la première composante  $Fe_3Al$ ,  $WSi_2$ ,  $MoSi_2$ ,  $TiAl$ ,  $Ti_3Al$ ,  $TaSi_2$  apparaissent comme étant peu influentes sur le modèle.

➤ La figure [IV.4 .4.b]: les variables  $TiSi_2$ ,  $WSi_2$ ,  $MoSi_2$  sont significatives et les plus influencent sur le modèle .

#### IV.4.5 Paramètre et équation du modèle :

En utilisant la méthode des moindres carrés partiels (PLS), nous avons relié la matrice X des variables explicatives à la matrice Y des variables dépendantes (réponse) .Une fois que notre méthode est appliquée nous obtenant des équation qui décrit le modèle de prédiction.

Le tableau [IV.4.5] des paramètres des modèles correspondant à chacune des variables dépendantes. Les équations sont ensuite affichées afin de faciliter un éventuel utilisateur ultérieur.

Variable	T <sub>f</sub>	D	E	B	G	B/G	C <sub>11</sub>	C <sub>12</sub>	C <sub>44</sub>	C <sub>12</sub> -C <sub>44</sub>
constante	1642,2	2,21	265,83	158,45	112,75	1,95	278,76	101,993	190,036	-88,043
FeAl	-312,28	0,28	-115,8	-22,45	-30,75	-0,306	-97,76	12,007	-63,036	75,043
Fe <sub>3</sub> Al	-102,28	0,299	-125,8	-0,456	-93,75	6,354	-94,76	43,007	-30,036	73,043
NiAl	-2,286	-0,90	-95,83	7,544	-30,01	0,044	-67,76	41,007	-78,036	119,043
Ni <sub>3</sub> Al	-252,28	-0,81	-85,83	12,544	-48,45	0,264	-57,76	44,007	-66,036	110,043
TiAl	-182,28	1,249	-90,83	-48,67	-37,95	-0,496	-153,7	-2,993	-68,036	65,043
Ti <sub>3</sub> Al	-42,28	-0,55	-119,8	-52,45	-55,75	-0,096	-102,7	-14,193	-127,636	113,443
Al <sub>3</sub> Ti	-292,2	-0,29	-109,8	-48,45	78,347	-0,362	-101,7	-24,993	-105,036	80,043
MoSi <sub>2</sub>	377,714	0,499	173,864	51,244	78,347	-0,859	138,23	2,207	14,164	-11,957
VSi <sub>2</sub>	34,714	0,289	76,764	8,744	35,147	-0,826	79,036	-51,393	-54,336	2,943
NbSi <sub>2</sub>	280,714	0,369	96,964	-7,456	40,447	-0,971	105,236	-26,093	-43,236	17,143
TaSi <sub>2</sub>	279,714	-0,351	93,164	34,044	38,247	-0,682	96,536	-23,593	-46,336	22,743
CrSi <sub>2</sub>	-152,28	0,259	88,764	13,544	40,547	-0,836	93,436	-56,693	-40,936	-15,757
WSi <sub>2</sub>	507,714	-0,711	202,064	63,944	90,847	-0,864	164,036	19,707	21,564	-1,857
TiSi <sub>2</sub>	-142,28	0,359	12,264	-11,65	4,147	-0,362	-0,164	38,007	686,964	-648,957

Le tableau [IV.4.5] : Paramètres du modèle

### *Les équations du modèle pour les composés intermétalliques :*

$$\mathbf{Tf} = 1642,2857 - 312,2857 \mathbf{FeAl} - 102,2857 \mathbf{Fe}_3\mathbf{Al} - 2,2857 \mathbf{NiAl} - 252,2857 \mathbf{Ni}_3\mathbf{Al} - 182,2857 \mathbf{TiAl} - 42,2857 \mathbf{Ti}_3\mathbf{Al} - 292,2857 \mathbf{Al}_3\mathbf{Ti} + 377,7142 \mathbf{MoSi}_2 + 34,7142 \mathbf{VSi}_2 + 280,7142 \mathbf{NbSi}_2 + 279,7142 \mathbf{TaSi}_2 - 152,2857 \mathbf{CrSi}_2 + 507,714 \mathbf{WSi}_2 - 142,2857 \mathbf{TiSi}_2.$$

$$\mathbf{D} = 2,2107 + 0,2892 \mathbf{FeAl} + 0,2992 \mathbf{Fe}_3\mathbf{Al} - 0,9007 \mathbf{NiAl} - 0,8107 \mathbf{Ni}_3\mathbf{Al} + 1,2492 \mathbf{TiAl} - 0,5507 \mathbf{Ti}_3\mathbf{Al} - 0,2907 \mathbf{Al}_3\mathbf{Ti} + 0,4992 \mathbf{MoSi}_2 + 0,2892 \mathbf{VSi}_2 + 0,3692 \mathbf{NbSi}_2 - 0,3507 \mathbf{TaSi}_2 + 0,2592 \mathbf{CrSi}_2 - 0,7107 \mathbf{WSi}_2 + 0,3592 \mathbf{TiSi}_2.$$

$$\mathbf{E} = 265,8357 - 115,8357 \mathbf{FeAl} - 125,8357 \mathbf{Fe}_3\mathbf{Al} - 95,8357 \mathbf{NiAl} - 85,8357 \mathbf{Ni}_3\mathbf{Al} - 90,8357 \mathbf{TiAl} - 119,8357 \mathbf{Ti}_3\mathbf{Al} - 109,8357 \mathbf{Al}_3\mathbf{Ti} + 173,8642 \mathbf{MoSi}_2 + 76,7642 \mathbf{VSi}_2 + 96,9642 \mathbf{NbSi}_2 + 93,1642 \mathbf{TaSi}_2 + 88,7642 \mathbf{CrSi}_2 + 202,0642 \mathbf{WSi}_2 + 12,2642 \mathbf{TiSi}_2.$$

$$\mathbf{B} = 158,4557 - 22,4557 \mathbf{FeAl} - 0,4557 \mathbf{Fe}_3\mathbf{Al} + 7,5442 \mathbf{NiAl} + 12,5442 \mathbf{Ni}_3\mathbf{Al} - 48,6757 \mathbf{TiAl} - 52,4557 \mathbf{Ti}_3\mathbf{Al} - 48,4557 \mathbf{Al}_3\mathbf{Ti} + 51,2442 \mathbf{MoSi}_2 + 8,7442 \mathbf{VSi}_2 - 7,4557 \mathbf{NbSi}_2 + 34,0442 \mathbf{TaSi}_2 + 13,5442 \mathbf{CrSi}_2 + 63,9442 \mathbf{WSi}_2 - 11,6557 \mathbf{TiSi}_2.$$

$$\mathbf{G} = 112,7528 - 30,7528 \mathbf{FeAl} - 93,7528 \mathbf{Fe}_3\mathbf{Al} - 30,0171 \mathbf{NiAl} - 35,7528 \mathbf{Ni}_3\mathbf{Al} - 37,9528 \mathbf{TiAl} - 55,7528 \mathbf{Ti}_3\mathbf{Al} - 43,7528 \mathbf{Al}_3\mathbf{Ti} + 78,3471 \mathbf{MoSi}_2 + 35,1471 \mathbf{VSi}_2 + 40,4471 \mathbf{NbSi}_2 + 38,2471 \mathbf{TaSi}_2 + 40,5471 \mathbf{CrSi}_2 + 90,8471 \mathbf{WSi}_2 + 4,1471 \mathbf{TiSi}_2.$$

$$\mathbf{B/G} = 1,9561 - 0,3061 \mathbf{FeAl} + 6,3538 \mathbf{Fe}_3\mathbf{Al} + 4,3857 \mathbf{E-02NiAl} + 0,2638 \mathbf{Ni}_3\mathbf{Al} - 0,4961 \mathbf{TiAl} - 9,6142 \mathbf{E-02Ti}_3\mathbf{Al} - 0,3621 \mathbf{Al}_3\mathbf{Ti} - 0,8591 \mathbf{MoSi}_2 - 0,8261 \mathbf{VSi}_2 - 0,9711 \mathbf{NbSi}_2 - 0,6821 \mathbf{TaSi}_2 - 0,8361 \mathbf{CrSi}_2 - 0,8641 \mathbf{WSi}_2 - 0,3621 \mathbf{TiSi}_2.$$

$$\mathbf{C}_{12} = 101,9928 + 12,0071 \mathbf{FeAl} + 43,0071 \mathbf{Fe}_3\mathbf{Al} + 41,0071 \mathbf{NiAl} + 44,0071 \mathbf{Ni}_3\mathbf{Al} - 2,9928 \mathbf{TiAl} - 14,1928 \mathbf{Ti}_3\mathbf{Al} - 24,9928 \mathbf{Al}_3\mathbf{Ti} + 2,2071 \mathbf{MoSi}_2 - 51,3928 \mathbf{VSi}_2 - 26,0928 \mathbf{NbSi}_2 - 23,5928 \mathbf{TaSi}_2 - 56,6928 \mathbf{CrSi}_2 + 19,7071 \mathbf{WSi}_2 + 38,0071 \mathbf{TiSi}_2.$$

$$\mathbf{C}_{44} = 190,0357 - 63,0357 \mathbf{FeAl} - 30,0357 \mathbf{Fe}_3\mathbf{Al} - 78,0357 \mathbf{NiAl} - 66,0357 \mathbf{Ni}_3\mathbf{Al} - 68,0357 \mathbf{TiAl} - 127,6357 \mathbf{Ti}_3\mathbf{Al} - 105,0357 \mathbf{Al}_3\mathbf{Ti} + 14,1642 \mathbf{MoSi}_2 - 54,3357 \mathbf{VSi}_2 - 43,2357 \mathbf{NbSi}_2 - 46,3357 \mathbf{TaSi}_2 - 40,9357 \mathbf{CrSi}_2 + 21,5642 \mathbf{WSi}_2 + 686,9642 \mathbf{TiSi}_2.$$

$$\mathbf{C}_{12} - \mathbf{C}_{44} = -88,0428 + 75,0428 \mathbf{FeAl} + 73,0428 \mathbf{Fe}_3\mathbf{Al} + 119,0428 \mathbf{NiAl} + 110,0428 \mathbf{Ni}_3\mathbf{Al} + 65,0428 \mathbf{TiAl} + 113,4428 \mathbf{Ti}_3\mathbf{Al} + 80,0428 \mathbf{Al}_3\mathbf{Ti} - 11,9571 \mathbf{MoSi}_2 + 2,9428 \mathbf{VSi}_2 + 17,1428 \mathbf{NbSi}_2 + 22,7428 \mathbf{TaSi}_2 - 15,7571 \mathbf{CrSi}_2 - 1,8571 \mathbf{WSi}_2 - 648,9571 \mathbf{TiSi}_2.$$

Ces modèles est étudier les comportements des composés en fonction de la variation des concentrations des différents composés intermétalliques. Suivant ces modèles nous remarquons que la température de fusion croit avec l'augmentation de la teneur en composés

intermétalliques dicilicides et présente une faible dépendance sur les autres composés intermétalliques, d'un autre côté, le module de Young atteint sa valeur minimale pour des faibles concentrations on Ni<sub>3</sub>Al.

#### **IV.4.6 L'ajustement du modèle :**

Pour chaque modèle PLS nous affichons le tableau des coefficients d'ajustement et le tableau des coefficients normalisés. On obtient les mêmes coefficients d'ajustement du modèle pour tous les variables.

**Tableau IV.4.6-A : Coefficients d'ajustement (Variable Tf) :**

<b>Observation</b>	14,000
<b>Somme des poids</b>	14,000
<b>DDL</b>	4,000
<b>R<sup>2</sup></b>	1,000
<b>Ecart-type</b>	0,000
<b>MCE</b>	0,000
<b>RMCE</b>	0,000

Tableau IV.4.6-B : Coefficients normalisés (variables) .

Variable	Coefficient	Ecart-type	Borne inférieure (95%)	Borne supérieure (95%)
FeAl	-0,261	0,260	-0,770	0,249
Fe <sub>3</sub> Al	-0,283	0,281	-0,834	0,268
NiAl	-0,216	0,219	-0,645	0,213
Ni <sub>3</sub> Al	-0,193	0,199	-0,582	0,196
TiAl	-0,204	0,209	-0,613	0,205
Ti <sub>3</sub> Al	-0,270	0,268	-0,796	0,257
Al <sub>3</sub> Ti	-0,247	0,248	-0,733	0,239
MoSi <sub>2</sub>	0,391	0,391	-0,376	1,158
VSi <sub>2</sub>	0,173	0,190	-0,201	0,546
NbSi <sub>2</sub>	0,218	0,232	-0,236	0,672
TaSi <sub>2</sub>	0,210	0,224	-0,229	0,648
CrSi <sub>2</sub>	0,200	0,215	-0,221	0,621
WSi <sub>2</sub>	0,454	0,447	-0,422	1,331
TiSi <sub>2</sub>	0,028	0,083	-0,135	0,190

L'analyse du modèle correspondant a chaque variable ; nous permettent de montrer une bonne qualité du modèle.

**IV.5 Les Résidus et Prédications :**

La régression PLS, nous donne des tableaux de prédictions et résidus qui permettent de vérifier que les valeurs d'une propriétés sont très bien reproduites par ces variables.

**Coefficients normalisés (Variables D)**

Observation	Prédit D	Résidus
FeAl	2,500	0,000
Fe <sub>3</sub> Al	2,510	0,000
NiAl	1,310	0,000
Ni <sub>3</sub> Al	1,400	0,000
TiAl	3,460	0,000
Ti <sub>3</sub> Al	1,660	0,000
Al <sub>3</sub> Ti	1,920	0,000
MoSi <sub>2</sub>	2,710	0,000
VSi <sub>2</sub>	2,500	0,000
NbSi <sub>2</sub>	2,580	0,000
TaSi <sub>2</sub>	1,860	0,000
CrSi <sub>2</sub>	2,470	0,000
WSi <sub>2</sub>	1,500	0,000
TiSi <sub>2</sub>	2,570	0,000

**Coefficients normalisés (Variables T<sub>f</sub>)**

Observation	Prédit T <sub>f</sub>	Résidus
FeAl	1330,01	0,000
Fe <sub>3</sub> Al	1540,00	0,000
NiAl	1640,00	0,000
Ni <sub>3</sub> Al	1390,00	0,000
TiAl	1460,00	0,000
Ti <sub>3</sub> Al	1600,00	0,000
Al <sub>3</sub> Ti	1350,00	0,000
MoSi <sub>2</sub>	2020,00	0,000
VSi <sub>2</sub>	1677,00	0,000
NbSi <sub>2</sub>	1923,00	0,000
TaSi <sub>2</sub>	1922,00	0,000
CrSi <sub>2</sub>	1490,00	0,000
WSi <sub>2</sub>	2150,00	0,000
TiSi <sub>2</sub>	1500,00	0,000

**Coefficients normalisés (Variables E)**

Observation	Prédit E	Résidus
FeAl	150,00	0,000
Fe <sub>3</sub> Al	140,00	0,000
NiAl	170,00	0,000
Ni <sub>3</sub> Al	180,00	0,000
TiAl	175,00	0,000
Ti <sub>3</sub> Al	146,00	0,000
Al <sub>3</sub> Ti	156,00	0,000
MoSi <sub>2</sub>	439,70	0,000
VSi <sub>2</sub>	342,60	0,000
NbSi <sub>2</sub>	362,80	0,000
TaSi <sub>2</sub>	359,00	0,000
CrSi <sub>2</sub>	354,60	0,000
WSi <sub>2</sub>	476,90	0,000
TiSi <sub>2</sub>	278,10	0,000

**Coefficients normalisés (Variables B)**

Observation	Prédit B	Résidus
FeAl	136,00	0,000
Fe <sub>3</sub> Al	158,00	0,000
NiAl	166,00	0,000
Ni <sub>3</sub> Al	171,00	0,000
TiAl	109,78	0,000
Ti <sub>3</sub> Al	106,00	0,000
Al <sub>3</sub> Ti	110,00	0,000
MoSi <sub>2</sub>	209,70	0,000
VSi <sub>2</sub>	167,20	0,000
NbSi <sub>2</sub>	151,00	0,000
TaSi <sub>2</sub>	192,50	0,000
CrSi <sub>2</sub>	172,00	0,000
WSi <sub>2</sub>	222,40	0,000
TiSi <sub>2</sub>	146,80	0,000

**Coefficients normalisés (Variables G)**

Observation	Prédit G	Résidus
FeAl	82,000	0,000
Fe <sub>3</sub> Al	19,000	0,000
NiAl	82,740	0,000
Ni <sub>3</sub> Al	77,000	0,000
TiAl	74,800	0,000
Ti <sub>3</sub> Al	57,000	0,000
Al <sub>3</sub> Ti	69,000	0,000
MoSi <sub>2</sub>	191,10	0,000
VSi <sub>2</sub>	147,90	0,000
NbSi <sub>2</sub>	153,20	0,000
TaSi <sub>2</sub>	151,00	0,000
CrSi <sub>2</sub>	153,30	0,000
WSi <sub>2</sub>	203,60	0,000
TiSi <sub>2</sub>	116,90	0,000

**Coefficients normalisés (Variables B/G)**

Observation	Prédit B/ G	Résidus
FeAl	1,650	0,000
Fe <sub>3</sub> Al	8,310	0,000
NiAl	2,000	0,000
Ni <sub>3</sub> Al	2,220	0,000
TiAl	1,460	0,000
Ti <sub>3</sub> Al	1,860	0,000
Al <sub>3</sub> Ti	1,504	0,000
MoSi <sub>2</sub>	1,097	0,000
VSi <sub>2</sub>	1,130	0,000
NbSi <sub>2</sub>	0,985	0,000
TaSi <sub>2</sub>	1,274	0,000
CrSi <sub>2</sub>	1,120	0,000
WSi <sub>2</sub>	1,092	0,000
TiSi <sub>2</sub>	1,594	0,000

**Coefficients normalisés (Variables C11)**

Observation	Prédit C11	Résidus
FeAl	181,00	0,000
Fe <sub>3</sub> Al	184,00	0,000
NiAl	211,00	0,000
Ni <sub>3</sub> Al	221,00	0,000
TiAl	125,00	0,000
Ti <sub>3</sub> Al	176,00	0,000
Al <sub>3</sub> Ti	177,00	0,000
MoSi <sub>2</sub>	417,00	0,000
VSi <sub>2</sub>	357,00	0,000
NbSi <sub>2</sub>	384,00	0,000
TaSi <sub>2</sub>	375,30	0,000
CrSi <sub>2</sub>	372,20	0,000
WSi <sub>2</sub>	442,80	0,000
TiSi <sub>2</sub>	278,60	0,000

**Coefficients normalisés (Variables C12)**

Observation	Prédit C12	Résidus
FeAl	114,00	0,000
Fe <sub>3</sub> Al	145,00	0,000
NiAl	143,00	0,000
Ni <sub>3</sub> Al	146,00	0,000
TiAl	99,000	0,000
Ti <sub>3</sub> Al	87,800	0,000
Al <sub>3</sub> Ti	77,000	0,000
MoSi <sub>2</sub>	104,20	0,000
VSi <sub>2</sub>	50,600	0,000
NbSi <sub>2</sub>	75,900	0,000
TaSi <sub>2</sub>	78,400	0,000
CrSi <sub>2</sub>	45,300	0,000
WSi <sub>2</sub>	121,70	0,000
TiSi <sub>2</sub>	140,00	0,000

**Coefficients normalisés (Variables C44)    Coefficients normalisés (Variables C11-C44)**

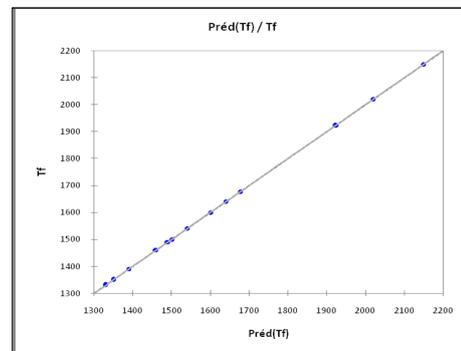
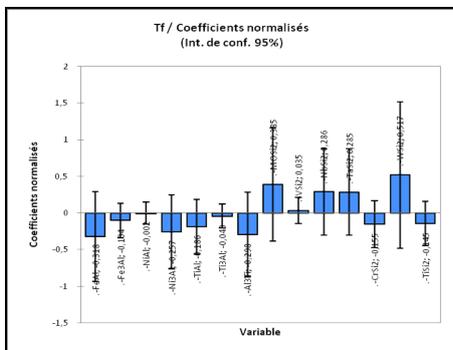
Observation	Prédit C44	Résidus
FeAl	127,00	0,000
Fe <sub>3</sub> Al	160,00	0,000
NiAl	112,00	0,000
Ni <sub>3</sub> Al	124,00	0,000
TiAl	122,00	0,000
Ti <sub>3</sub> Al	62,400	0,000
Al <sub>3</sub> Ti	85,000	0,000
MoSi <sub>2</sub>	204,20	0,000
VSi <sub>2</sub>	135,70	0,000
NbSi <sub>2</sub>	146,80	0,000
TaSi <sub>2</sub>	143,70	0,000
CrSi <sub>2</sub>	149,10	0,000
WSi <sub>2</sub>	211,60	0,000
TiSi <sub>2</sub>	877,00	0,000

Observation	Prédit C44	Résidus
FeAl	-13,000	0,000
Fe <sub>3</sub> Al	-15,000	0,000
NiAl	31,000	0,000
Ni <sub>3</sub> Al	22,000	0,000
TiAl	-23,000	0,000
Ti <sub>3</sub> Al	25,400	0,000
Al <sub>3</sub> Ti	-08,000	0,000
MoSi <sub>2</sub>	-100,00	0,000
VSi <sub>2</sub>	-85,100	0,000
NbSi <sub>2</sub>	-70,900	0,000
TaSi <sub>2</sub>	-65,300	0,000
CrSi <sub>2</sub>	-103,80	0,000
WSi <sub>2</sub>	-89,900	0,000
TiSi <sub>2</sub>	-737,00	0,000

✓ D’après les tableaux des Coefficients normalisés des différentes variables, nous remarquons que les résidus sont nuls pour tous les éléments ,ce qui nous fait dire que le modèle est bon , et ce qui permet de vérifier aussi que les valeurs de toutes les propriétés (D,E,B,T<sub>f</sub>,G,B /G,C11,C12,C44,C12-C44) sont très bien reproduites par ces variables.

✓ Le modèle PLS permet aussi de donner les graphes correspondants aux coefficients normalisés et les graphes des résidus et prédictions permettant d’identifier la qualité du modèle. Nous prenons les graphiques ci-dessous qui correspondent aux coefficients normalisés pour le modèle avec 2 composantes et ceci pour les différentes variables (T<sub>f</sub>,D,E,G,B,B/G, C<sub>11</sub>, C<sub>12</sub>, C<sub>44</sub>, C<sub>12</sub>- C<sub>44</sub>).

**Variable T<sub>f</sub>:**

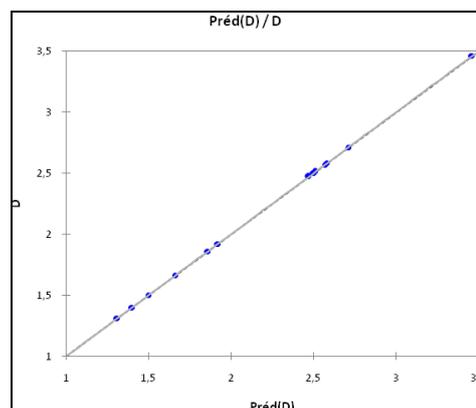
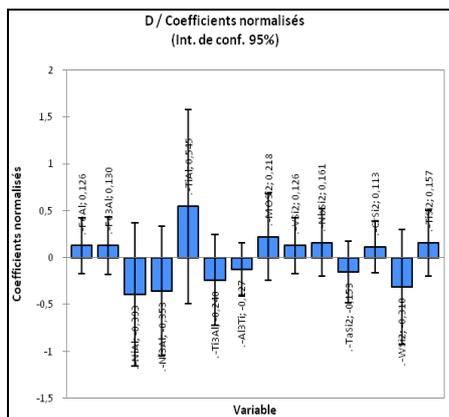


**La figure [IV.3 .7] :**Graphe de résidus et prédiction correspond aux variable T<sub>f</sub>

✓ Le graphique correspond aux coefficients normalisés pour  $T_f$ : On voit que l'ensemble des composés intermétalliques [WSi<sub>2</sub>, MoSi<sub>2</sub>, FeAl, Ni<sub>3</sub>Al] sont significatives et influencent le modèle.

✓ Dans le graphe de prédiction : la variation est linéaire entre la valeur de  $T_f$  prédite et celle utilisée en entrée, ce qui indique bien la bonne qualité du modèle.

**Variable D :**

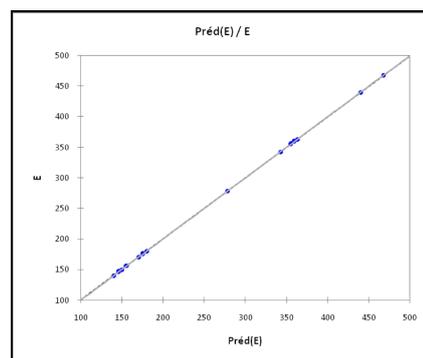
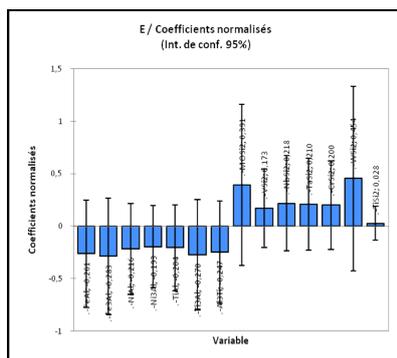


La figure [IV.3.8] : Graphe de résidus et prédiction correspond aux variable D.

✓ Le graphique correspond aux coefficients normalisés pour "D" : nous permet de conclure que l'ensemble des composés intermétalliques [NiAl, Ni<sub>3</sub>Al, TiAl, Ti<sub>3</sub>Al, WSi<sub>2</sub>] sont significatives et influencent le modèle.

✓ Dans le graphe de prédiction : la variation est linéaire entre la valeur de D prédite et celle utilisée en entrée, ce qui indique bien la bonne qualité du modèle.

**Variable E :**

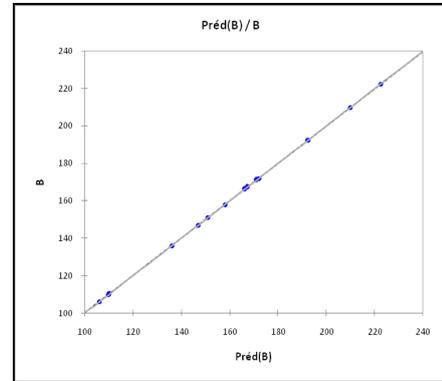
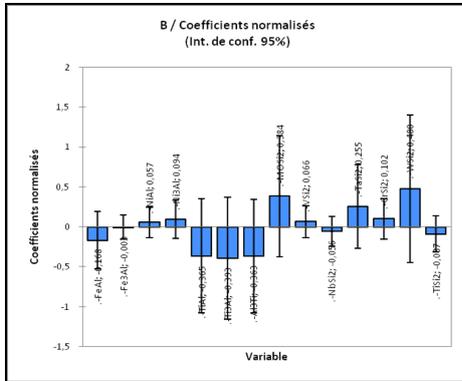


La figure [IV.3.9] : Graphe de résidus et prédiction correspond aux variable E.

✓ Le graphique correspond aux coefficients normalisés pour "E" : on voit que l'ensemble des composés intermétalliques [WSi<sub>2</sub>, MoSi<sub>2</sub>, Ti<sub>3</sub>Al, Ni<sub>3</sub>Al] sont significatives et influencent le modèle.

✓ Dans le graphe de prédiction : la variation est linéaire entre la valeur de E prédite et celle utilisée en entrée, ce qui indique bien la bonne qualité du modèle.

**Variable B :**

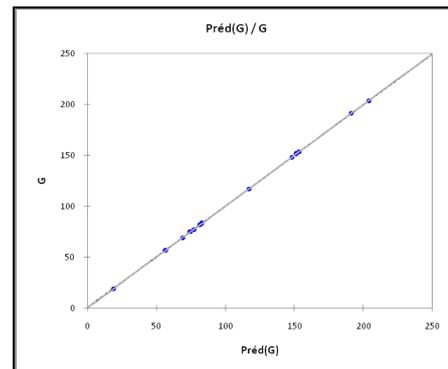
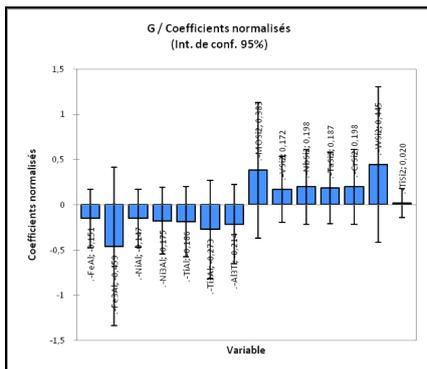


La figure [IV.3 .10] :Graphe de résidus et prédiction correspond aux variable B.

✓ L'analyse du modèle correspond "B" nous permet de conclure que l'ensemble des composés intermétalliques [WSi<sub>2</sub>, NiAl, Ni<sub>3</sub>Al , Al<sub>3</sub>Ti] sont significatives et influencent le modèle.

✓ Dans le graphe de prédiction : la variation est linéaire entre la valeur de "B" prédite et celle utilisée en entrée, ce qui indique bien la bonne qualité du modèle.

**Variable G :**

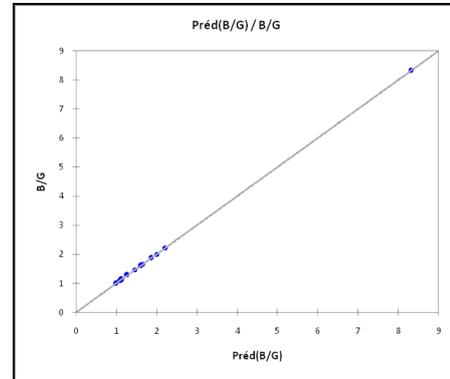
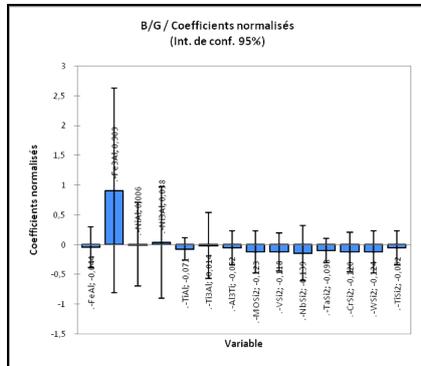


La figure [IV.3 .11] :Graphe de résidus et prédiction correspond aux variable G.

✓ L'analyse du modèle correspond "G" nous permet de conclure que l'ensemble des composés intermétalliques [Fe<sub>3</sub>Al,MoSi<sub>2</sub> WSi<sub>2</sub>,NiAl,Ni<sub>3</sub>Al ,Al<sub>3</sub>Ti] sont significatives et influencent le modèle.

✓ Dans le graphe de prédiction : la variation est linéaire entre la valeur de G prédites et celle utilisée en entrée, ce qui indique que notre modèle est bon .

**Variable B/G :**

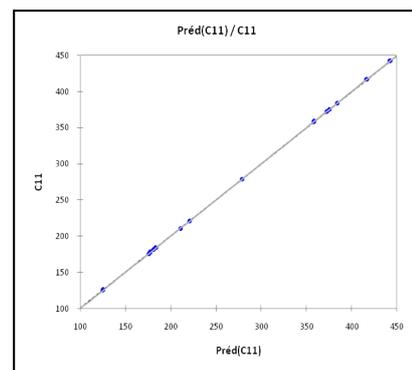
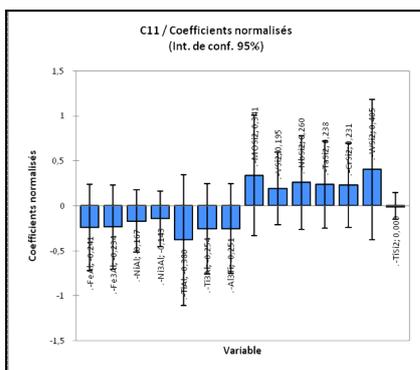


La figure [IV.3 .12] :Graphe de résidus et prédiction correspond aux variable B/G.

✓ Le graphique correspond aux coefficients normalisés pour le modèle à la sixième variable E : on voit que seul le composé intermétallique [Fe<sub>3</sub>Al] sont significatives et influence le modèle.

✓ Dans le graphe de prédiction : la variation est linéaire entre la valeur de B/G prédites et celle utilisée en entrée, ce qui indique que notre modèle est bon.

**Variable C<sub>11</sub> :**

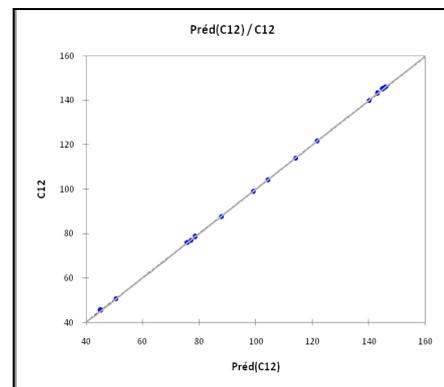
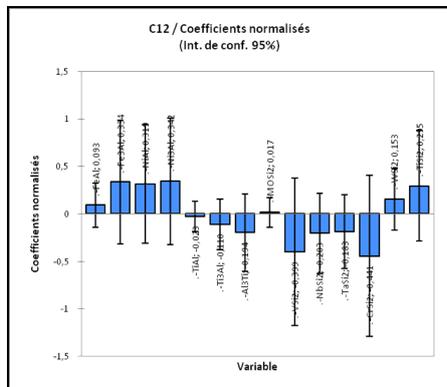


La figure [IV.3 .13] :Graphe de résidus et prédiction correspond aux variable C<sub>11</sub> .

✓ Le graphique correspond aux coefficients normalisés pour "C<sub>11</sub>" : on voit que les composés intermétalliques [NiAl,Ni<sub>3</sub>Al,Fe<sub>3</sub>Al,WSi<sub>2</sub>,MoSi<sub>2</sub>] sont significatives et influence le modèle.

✓ Remarquant aussi que la variation est linéaire entre les valeurs de C<sub>11</sub> prédite en fonction de celle utilisée en entrée, ce qui implique la bonne qualité du modèle.

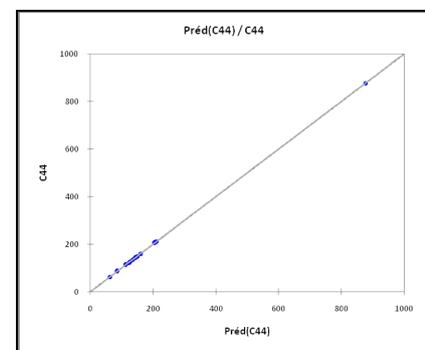
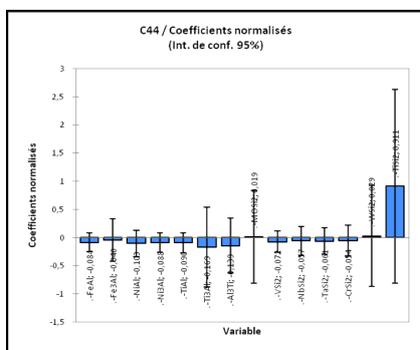
**Variable C<sub>12</sub> :**



La figure [IV.3 .14] :Graphe de résidus et prédiction correspond aux variable C<sub>12</sub> .

- ✓ Le graphique correspond aux coefficients normalisés pour "C<sub>12</sub>" : on voit que les composés intermétalliques [NiAl,Ni<sub>3</sub>Al,Fe<sub>3</sub>Al,VS<sub>2</sub>,CrSi<sub>2</sub>,MoSi<sub>2</sub>] sont significatives et influence le modèle.
- ✓ Remarquant aussi que la variation est linéaire entre les valeurs de C<sub>12</sub> prédite en fonction de celle utilisée en entrée, ce qui implique la bonne qualité du modèle.

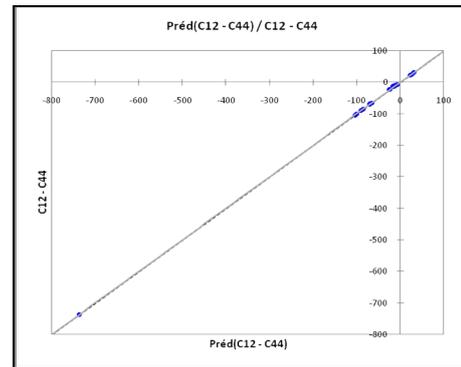
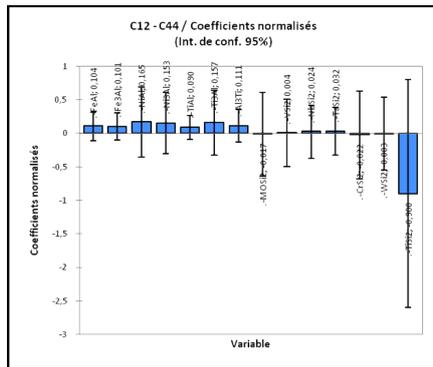
**Variable C<sub>44</sub> :**



La figure [IV.3 .14] :Graphe de résidus et prédiction correspond aux variable C<sub>44</sub>

- ✓ Le graphique correspond aux coefficients normalisés pour "C<sub>44</sub>" : on voit que seul le composé intermétallique [TiSi<sub>2</sub>] sont significatives et influence le modèle.
- ✓ Dans le graphe de prédiction : la variation est linéaire entre la valeur de C<sub>44</sub> prédites et celle utilisée en entrée, ce qui indique que notre modèle est bon.

**Variable C<sub>12</sub>-C<sub>44</sub> :**



**La figure [IV.3 .15] :**Graphe de résidus et prédiction correspond aux variable C<sub>12</sub>.C<sub>44</sub>

- ✓ L'analyse du modèle correspond C<sub>12</sub>-C<sub>44</sub> nous permet de conclure que l'ensemble des composés intermétalliques [TiAl,TiSi<sub>2</sub>] sont significatives et influencent le modèle.
- ✓ Dans le graphe de prédiction : la variation est linéaire entre la valeur de C<sub>12</sub>-C<sub>44</sub> prédite et celle utilisée en entrée, ce qui indique que notre modèle est bon.

#### **IV.6 Prédiction des propriétés des intermétalliques ternaires :**

L'utilisation de PLS ici 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 de nouveaux matériaux. Dans notre cas, la méthode PLS a été utilisée pour prédire les propriétés structurales des ternaires des intermétalliques. Alors notre approche essaye de prédire de nouveaux ternaires sans faire de calculs ni expérience et avec une vitesse importante.

La prédiction a été réalisée sur la base des matériaux binaires de la base de données du tableau. Ils est possible de prédire ces quantités et ceci sans avoir aucune information sur les propriétés structurales, mécanique (Module de Young) et thermique (Température de Fusion) des ternaires des intermétalliques.

En se basant sur ces résultats, nous avons commencé à prédire les propriétés pour les composés intermétalliques. Les résultats sont présentés dans les tableaux ( ) qui comportent les valeurs prédites, des nouveaux alliages intermétalliques ternaires, (Température de fusion, distance interatomique, le module de Young, le module de compressibilité B, le module du cisaillement et le rapport B/G), et ceci pour les différentes concentrations.

Nous avons prédit les propriétés des nouveaux alliages pour les différentes concentrations.

**Tableaux IV -6-A : Alliages type  $A_{1-x}B_xAl$   $0 \leq x \leq 1$** **Pour  $x=0,25$** 

	T <sub>f</sub>	D	E	B	G	B/G
FeTiAl	1409,50	3,22	168,7	116,35	80,21	1,45
TiFeAl	1427,51	3,21	150,01	116,33	102,10	1,13
FeNiAl	1407,50	2,51	155,01	143,50	67,17	2,13
NiFeAl	1562,5	1,62	165,00	158,49	82,55	1,9
TiNiAl	1505,00	2,92	173,75	123,83	76,78	1,61
NiTiAl	1595,00	1,85	171,25	151,94	80,75	1,88

**Pour  $x=0,75$** 

	T <sub>f</sub>	D	E	B	G	B/G
<b>FeTiAl</b>	1518,64	3,22	168,75	116,33	76,6	1,52
<b>TiFeAl</b>	1427,5	2,74	156,25	129,44	80,2	1,61
<b>FeNiAl</b>	1407,5	1,60	165,00	158,5	82,55	1,92
<b>NiFeAl</b>	1408,64	2,20	155,00	143,50	82,18	1,74
<b>TiNiAl</b>	1595,0	1,84	171,25	151,94	80,75	1,88
<b>NiTiAl</b>	1505,0		173,75	123,83	76,78	1,61

**Pour  $x=0,85$** 

	T <sub>f</sub>	D	E	B	G	B/G
<b>FeTiAl</b>	1440,50	3,32	171,25	113,71	75,88	1,49
<b>TiFeAl</b>	1349,50	2,64	234,83	132,06	80,92	1,63
<b>FeNiAl</b>	1593,5	1,50	167,00	161,50	133,65	1,20
<b>NiFeAl</b>	1377,18	2,32	153,00	140,49	82,11	1,71
<b>TiNiAl</b>	1613,00	1,63	170,75	157,56	81,54	1,93
<b>NiTiAl</b>	1487,68	3,13	174,25	118,21	75,99	1,55

Pour  $x=0,7$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>FeTiAl</b>	1421,00	3,17	167,50	117,64	76,96	1,52
<b>TiFeAl</b>	1369,00	2,49	157,50	128,13	79,84	1,60
<b>FeNiAl</b>	1547,00	1,66	164,00	156,99	82,51	1,90
<b>NiFeAl</b>	1423,00	2,14	156,00	144,99	82,22	1,76
<b>TiNiAl</b>	1586,00	1,66	171,50	149,13	80,35	1,85
<b>NiTiAl</b>	1514,00	2,81	173,50	126,64	77,18	1,64

Tableaux IV -6-B : Alliages type  $A_{3-x}B_xAl$   $0 \leq x \leq 3$ Pour  $x=0,25$  :

	<b>T<sub>f</sub></b>	<b>d</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>FeNiAl</b>	1527,50	2,41	143,44	159,07	24,80	6,41
<b>NiFeAl</b>	1402,70	1,49	176,76	170,96	72,22	2,36
<b>FeTiAl</b>	1545,08	2,44	140,62	153,68	22,25	6,90
<b>TiFeAl</b>	1592,22	1,73	145,61	110,37	53,92	2,04
<b>NiTiAl</b>	1407,68	1,42	177,08	177,08	75,87	2,83
<b>TiNiAl</b>	1582,61	1,64	111,45	111,45	58,72	1,89

Pour  $x=0,5$  :

	<b>T<sub>f</sub></b>	<b>d</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>FeNiAl</b>	1515,28	2,32	146,76	160,15	28,72	5,57
<b>NiFeAl</b>	1415,15	1,58	173,44	168,82	67,41	2,50
<b>FeTiAl</b>	1550,06	2,36	141,12	158,39	25,40	6,23
<b>TiFeAl</b>	1590,08	1,80	145,12	114,68	50,74	2,26
<b>NiTiAl</b>	1425,11	1,44	174,46	160,19	73,71	2,17
<b>TiNiAl</b>	1565,18	1,62	151,76	116,84	60,37	1,93

Pour  $x=0,75$ 

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>FeNiAl</b>	1502,5	2,23	150,0	159,99	33,5	4,77
<b>NiFeAl</b>	1427,5	1,68	170,0	167,74	62,5	2,68
<b>FeTiAl</b>	1555,0	2,30	141,5	145,0	28,50	5,08
<b>TiFeAl</b>	1585,0	1,87	144,5	118,78	47,51	2,50
<b>NiTiAl</b>	1442,5	1,46	171,5	154,75	72,00	2,14
<b>TiNiAl</b>	1547,5	1,60	154,5	122,25	62,00	1,97

Pour  $x=1,00$ :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>FeNiAl</b>	1467,09	2,41	143,44	159,07	24,80	6,41
<b>NiFeAl</b>	1402,70	1,49	176,76	170,96	72,22	2,36
<b>FeTiAl</b>	1545,08	2,44	140,62	153,68	22,25	6,90
<b>TiFeAl</b>	1592,22	1,73	145,61	110,37	53,92	2,04
<b>NiTiAl</b>	1407,68	1,42	177,08	177,08	75,87	2,83
<b>TiNiAl</b>	1582,61	1,64	111,45	111,45	58,72	1,89

Pour  $x=1,25$ :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>FeNiAl</b>	1450,08	2,05	156,76	163,40	43,13	3,78
<b>NiFeAl</b>	1452,65	1,40	163,44	165,57	52,79	3,13
<b>FeTiAl</b>	1565,06	1,63	126,88	136,36	32,39	4,20
<b>TiFeAl</b>	1575,08	2,01	143,62	127,68	41,24	3,09
<b>NiTiAl</b>	1624,69	1,50	165,94	143,94	68,71	2,09
<b>TiNiAl</b>	1513,51	1,23	160,26	133,09	65,37	2,03

Tableaux IV -6-C : Alliages de type  $Mo_{1-x}X_xSi_2$   $0 < x < 1$ (X=V,Nb,Ta, Cr,W,Ti)Pour  $x=0,25$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>MoVSi<sub>2</sub></b>	2217,53	2,65	415,42	198,99	180,29	1,10
<b>MoNbSi<sub>2</sub></b>	1995,74	2,67	420,47	195,02	181,62	1,07
<b>MoTaSi<sub>2</sub></b>	1995,50	2,58	419,52	205,39	181,07	1,13
<b>MoCrSi<sub>2</sub></b>	1887,49	2,64	418,42	200,27	181,64	1,10
<b>MoWSi<sub>2</sub></b>	2052,49	2,40	446,74	212,87	194,22	1,09
<b>MoTiSi<sub>2</sub></b>	1889,99	2,67	399,29	193,97	172,54	1,12

Pour  $x=0,75$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>MoVSi<sub>2</sub></b>	1762,74	2,55	357,87	177,82	158,69	1,12
<b>MoNbSi<sub>2</sub></b>	1947,24	2,61	382,02	169,40	162,67	1,04
<b>MoTaSi<sub>2</sub></b>	1946,49	2,30	379,17	171,26	161,02	1,06
<b>MoCrSi<sub>2</sub></b>	1622,49	2,53	375,87	181,42	162,74	1,11
<b>MoWSi<sub>2</sub></b>	1863,64	1,8	309,30	219,22	200,47	1,09
<b>MoTiSi<sub>2</sub></b>	1629,99	2,60	318,49	162,52	135,44	1,19

Pour  $x=0,1$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>MoVSi<sub>2</sub></b>	1985,69	2,68	429,98	205,53	186,77	1,10
<b>MoNbSi<sub>2</sub></b>	2010,29	2,69	432,00	203,82	187,30	1,08
<b>MoTaSi<sub>2</sub></b>	2010,19	2,62	431,62	207,97	187,08	1,11
<b>MoCrSi<sub>2</sub></b>	1966,99	2,67	431,18	205,92	187,31	1,09
<b>MoWSi<sub>2</sub></b>	2032,99	2,58	442,51	210,96	192,34	1,09
<b>MoTiSi<sub>2</sub></b>	1967,99	2,69	423,53	203,40	183,67	1,10

Pour  $x=0,15$

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>MoVSi<sub>2</sub></b>	1968,55	2,67	425,13	203,32	184,62	1,10
<b>MoNbSi<sub>2</sub></b>	2005,45	2,69	428,16	200,89	185,41	1,08
<b>MoTaSi<sub>2</sub></b>	2005,29	2,58	427,59	207,19	185,08	1,11
<b>MoCrSi<sub>2</sub></b>	1940,99	2,67	426,93	204,04	185,42	1,10
<b>MoWSi<sub>2</sub></b>	2039,49	2,53	531,67	211,60	192,97	1,09
<b>MoTiSi<sub>2</sub></b>	1941,99	2,68	415,46	203,76	179,96	1,13

Tableaux IV -6-D : Alliages de type  $Nb_{1-x}X_xSi_2$   $0 \leq x < 1$ (X=V, Mo, Ta, Cr, W, Ti)Pour  $x=0,25$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>NbVSi<sub>2</sub></b>	1861,49	2,56	357,74	155,04	151,87	1,02
<b>NbMoSi<sub>2</sub></b>	1947,24	2,61	382,02	165,67	162,67	1,02
<b>NbTaSi<sub>2</sub></b>	1922,74	2,40	361,85	161,37	152,65	1,05
<b>NbCrSi<sub>2</sub></b>	1814,74	2,27	360,74	156,25	143,08	1,09
<b>NbWSi<sub>2</sub></b>	1979,74	2,31	389,13	168,85	165,79	1,02
<b>NbTiSi<sub>2</sub></b>	1817,25	2,57	341,68	149,94	144,12	1,04

Pour  $x=0,5$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>NbVSi<sub>2</sub></b>	1799,99	2,54	352,69	159,09	150,55	1,06
<b>NbMoSi<sub>2</sub></b>	1971,49	2,64	401,24	180,35	172,15	1,04
<b>NbTaSi<sub>2</sub></b>	1922,49	2,22	360,89	171,75	152,09	1,13
<b>NbCrSi<sub>2</sub></b>	1706,49	2,52	358,69	161,49	153,25	1,05
<b>NbWSi<sub>2</sub></b>	2036,49	2,04	415,34	186,69	178,39	1,04
<b>NbTiSi<sub>2</sub></b>	1711,49	2,58	320,45	148,9	135,04	1,10

Pour  $x=0,75$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>NbVSi<sub>2</sub></b>	1738,49	2,52	347,65	163,12	149,25	1,09
<b>NbMoSi<sub>2</sub></b>	1995,75	2,68	420,47	195,00	181,62	1,07
<b>NbTaSi<sub>2</sub></b>	1922,25	2,04	359,94	182,10	151,55	1,20
<b>NbCrSi<sub>2</sub></b>	1598,25	2,50	356,64	166,72	153,27	1,08
<b>NbWSi<sub>2</sub></b>	2093,24	1,78	441,62	204,52	190,99	1,07
<b>NbTiSi<sub>2</sub></b>	1605,74	2,57	299,27	147,82	125,97	1,17

Pour  $x=0,03$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>NbVSi<sub>2</sub></b>	1925,90	2,58	356,10	152,76	154,33	0,98
<b>NbMoSi<sub>2</sub></b>	1915,61	2,57	362,19	151,48	153,04	0,97
<b>NbTaSi<sub>2</sub></b>	1922,29	2,55	362,68	152,24	153,13	1,00
<b>NbCrSi<sub>2</sub></b>	1910,01	2,57	362,55	151,62	153,20	0,98
<b>NbWSi<sub>2</sub></b>	1929,80	2,54	365,95	153,14	154,71	0,96
<b>NbTiSi<sub>2</sub></b>	1910,30	2,58	360,25	150,87	152,11	0,99

Tableaux IV -6-E : Alliages de type  $Ta_{1-x}X_xSi_2$   $0 \leq x \leq 1$ (X=V, Mo, Nb, Cr, W, Ti)Pour  $x=0,9$ :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>TaVSi<sub>2</sub></b>	1701,49	2,12	344,23	169,72	148,21	1,14
<b>TaMoSi<sub>2</sub></b>	2010,19	2,31	422,62	207,97	187,09	1,11
<b>TaNbSi<sub>2</sub></b>	1922,89	2,20	362,42	155,15	152,98	1,01
<b>TaCrSi<sub>2</sub></b>	1533,19	2,09	355,04	174,05	153,07	1,13
<b>TaWSi<sub>2</sub></b>	2127,20	1,22	457,00	219,40	198,34	1,10
<b>TaTiSi<sub>2</sub></b>	1542,19	2,18	286,19	151,36	120,31	1,25

Pour  $x=0,25$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>TaVSi<sub>2</sub></b>	1860,74	2,02	354,89	211,70	150,22	1,40
<b>TaMoSi<sub>2</sub></b>	1946,49	2,07	379,17	196,79	154,24	1,27
<b>TaNbSi<sub>2</sub></b>	1922,25	2,04	359,94	182,12	151,54	1,20
<b>TaCrSi<sub>2</sub></b>	1813,99	2,01	357,89	187,37	151,57	1,23
<b>TaWSi<sub>2</sub></b>	1978,99	1,77	386,22	199,97	164,15	1,21
<b>TaTiSi<sub>2</sub></b>	1816,49	2,03	338,77	181,07	142,47	1,27

Tableaux IV -6-F : Alliages de type  $V_{1-x}X_xSi_2$   $0 \leq x \leq 1$ (X=Nb, Mo, Ta, Cr, W, Ti)Pour  $x=0,15$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>VMoSi<sub>2</sub></b>	1728,45	2,53	357,16	191,96	154,37	1,24
<b>VNbSi<sub>2</sub></b>	1713,89	2,51	336,63	164,77	148,69	1,11
<b>VTaSi<sub>2</sub></b>	1713,75	2,40	336,06	170,99	148,36	1,15
<b>VCrSi<sub>2</sub></b>	1648,94	2,50	335,40	167,92	148,70	1,13
<b>VWSi<sub>2</sub></b>	1747,95	2,35	352,39	175,48	156,25	1,12
<b>VTiSi<sub>2</sub></b>	1650,45	2,51	323,92	164,14	143,25	1,14

Pour  $x=0,7$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>VMoSi<sub>2</sub></b>	1917,09	2,64	410,56	196,94	178,14	1,10
<b>VNbSi<sub>2</sub></b>	1849,20	2,55	356,74	155,85	151,60	1,02
<b>VTaSi<sub>2</sub></b>	1848,49	2,05	354,08	184,90	150,06	1,23
<b>VCrSi<sub>2</sub></b>	1546,09	2,47	350,99	170,55	151,67	1,12
<b>VWSi<sub>2</sub></b>	2008,09	1,80	430,30	205,83	186,88	1,10
<b>VTiSi<sub>2</sub></b>	1553,09	2,54	297,45	152,91	126,20	1,21

Tableaux IV -6-G : Alliages de type  $Cr_{1-x}X_xSi_2$   $0 \leq x \leq 1$ (X=Nb, Mo, Ta, V, W, Ti)Pour  $x=0,7$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>CrMoSi<sub>2</sub></b>	1860,99	2,63	344,62	198,38	179,76	1,10
<b>CrNbSi<sub>2</sub></b>	1620,89	2,50	346,19	168,63	149,52	1,13
<b>CrTaSi<sub>2</sub></b>	1793,09	2,55	360,34	157,29	153,23	1,02
<b>CrVSi<sub>2</sub></b>	1792,39	2,04	357,67	186,34	151,69	1,22
<b>CrWSi<sub>2</sub></b>	1951,99	1,80	433,90	207,27	188,51	1,09
<b>CrTiSi<sub>2</sub></b>	1497,00	2,54	301,05	154,35	127,82	1,20

Pour  $x=0,09$  :

	<b>T<sub>f</sub></b>	<b>D</b>	<b>E</b>	<b>B</b>	<b>G</b>	<b>B/G</b>
<b>CrMoSi<sub>2</sub></b>	1537,67	2,49	362,25	175,39	156,70	1,11
<b>CrVSi<sub>2</sub></b>	1506,82	2,47	353,51	171,56	152,81	1,12
<b>CrTaSi<sub>2</sub></b>	1528,87	2,41	354,99	173,84	153,09	1,13
<b>CrNbSi<sub>2</sub></b>	1529,96	2,47	355,33	171,45	153,29	1,11
<b>CrWSi<sub>2</sub></b>	1549,39	2,38	364,74	176,53	157,82	1,11
<b>CrTiSi<sub>2</sub></b>	1491,47	2,48	347,71	169,73	150,02	1,13

Tableaux IV -6-H : Alliages de type  $W_{1-x}X_xSi_2$   $0 \leq x \leq 1$ (X=V, Mo, Nb, Cr, Ti)Pour x=0,9:

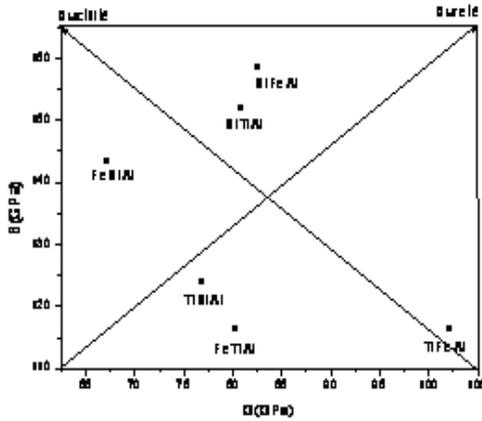
	T <sub>f</sub>	D	E	B	G	B/G
WMoSi <sub>2</sub>	2146,09	1,53	467,05	222,01	203,23	1,09
WVSi <sub>2</sub>	2135,80	1,54	464,14	220,74	201,93	1,09
WNbSi <sub>2</sub>	2143,18	1,53	464,74	220,25	202,08	1,08
WTaSi <sub>2</sub>	2143,15	1,51	464,63	221,50	202,02	1,08
WCrSi <sub>2</sub>	2130,19	1,52	464,50	220,88	202,09	1,06
WTiSi <sub>2</sub>	2130,49	1,53	462,20	220,13	200,99	1,09

➤ Ainsi pour l'ensemble de ces prédictions nous avons essayer de donner un ensemble de propriétés afin d'identifier les ternaires les plus intéressants. Ils est toute fois important de faire savoir que la majorité sont purement d'ordre hypothétique et que la technologie présentes aujourd'hui n'a pas pu les synthétiser.

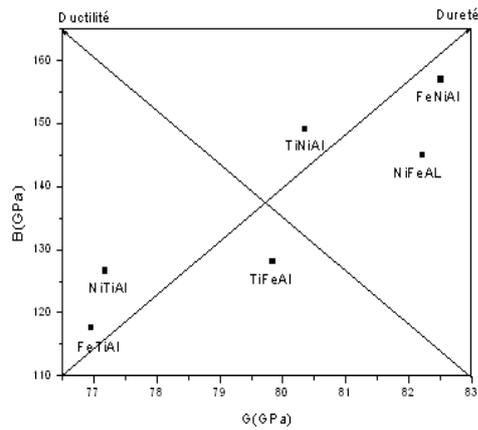
A cet effet il a été difficile de trouver des données sur l'ensemble de ces propriétés. Ainsi nous pouvons à l'heure actuelle valider notre modèle. Donc il est juste a titre indicatif .Toute fois afin d'analyser les propriétés de ces alliages nous avons tracé les graphes donnons B en fonction de G .Les résultats sont donnés sur les courbes (IV -6- (A,B,C,D,E,F,G)

**Graphes IV-6- A : Alliages type  $A_{1-x}B_xAl$   $0 \leq x \leq 1$**

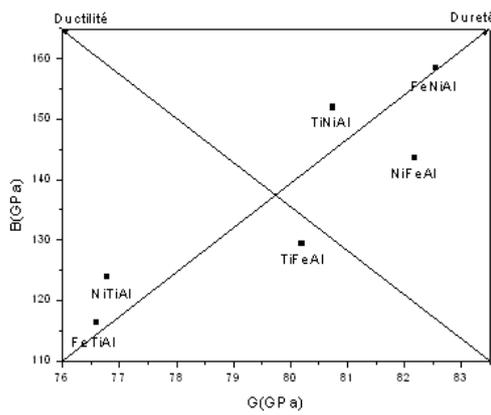
**Pour  $x=0,25$**



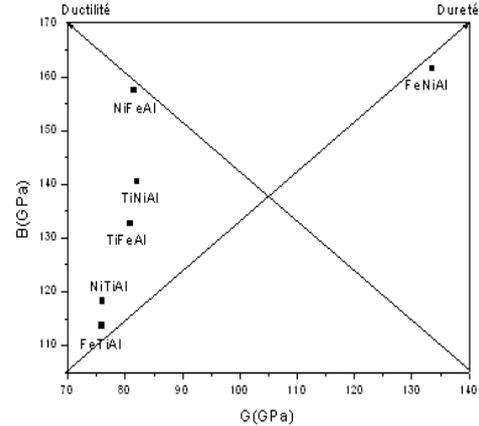
**Pour  $x=0,7$**



**Pour  $x=0,75$**



**Pour  $x=0,85$**



**Grappe IV-6- A1 : Grappe Pour toute les concentrations :**

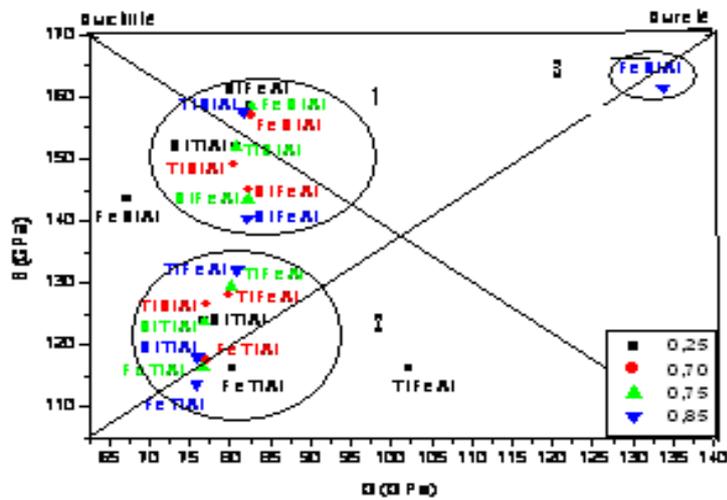


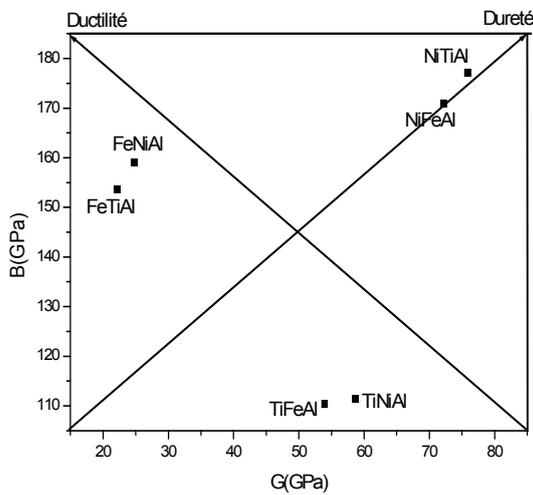
Figure : Grappe récapitulative des différentes concentrations.

✓ Nous remarquons que la concentration joue un rôle important quant aux propriétés de ces alliages.

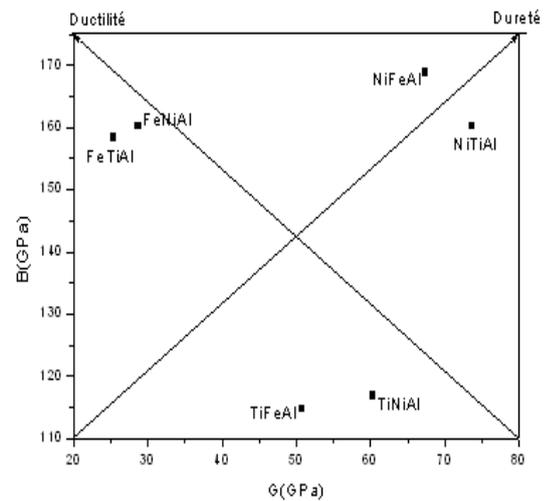
- Nous observons un cluster (1) montrant des alliages fortement ductiles.
- Un cluster (2) des alliages très peu durs et un peu ductile, et en cas exceptionnel(cluster 3) a large dureté le FeNiAl.

**Graphes IV -6- B : Alliages type  $A_{3-x}B_xAl$   $0 \leq x \leq 3$**

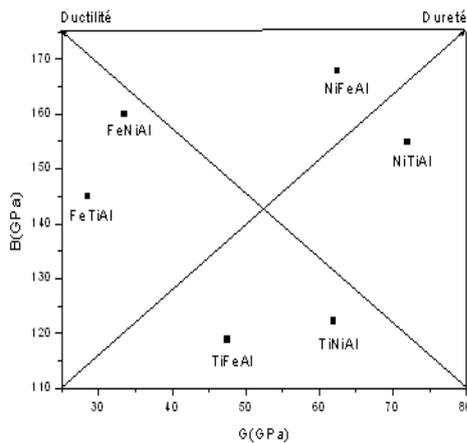
**Pour  $x = 0,25$  :**



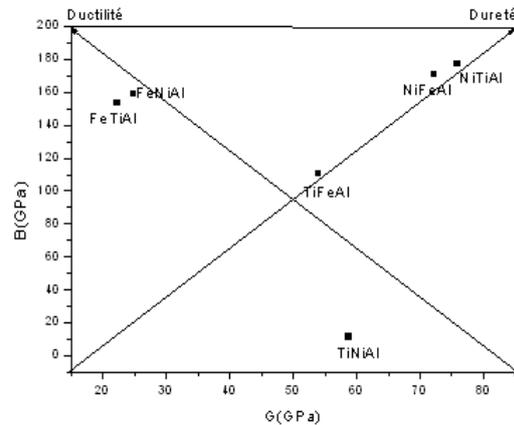
**Pour  $x=0,5$  :**



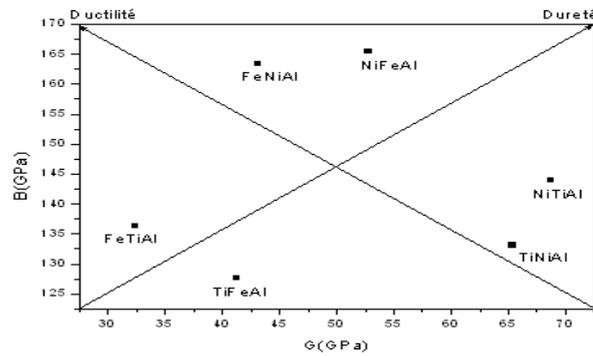
**Pour  $x=0,75$  :**



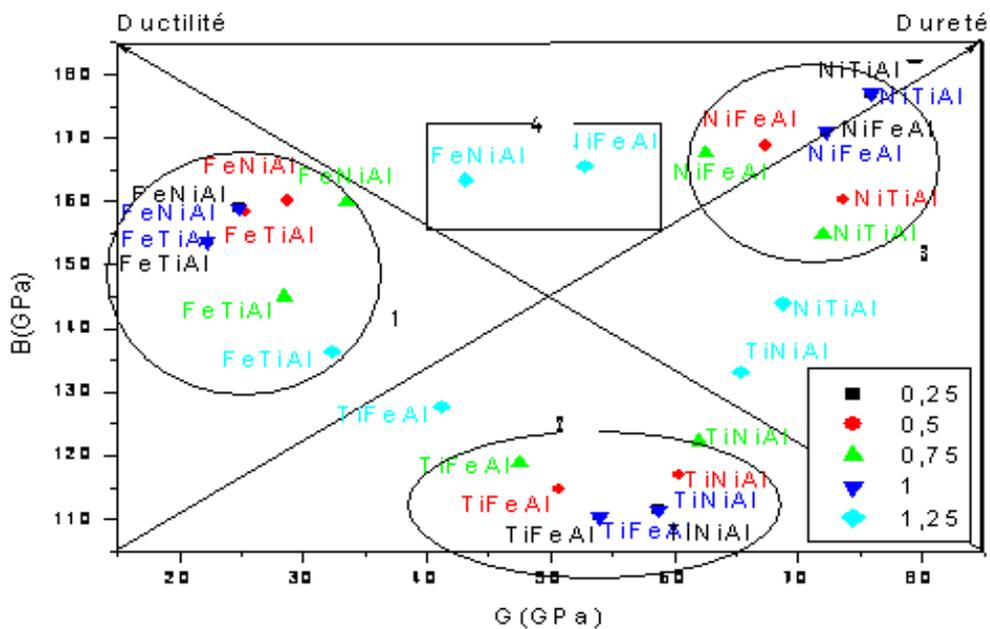
**Pour  $x=1$  :**



Pour  $x=1,25$  :



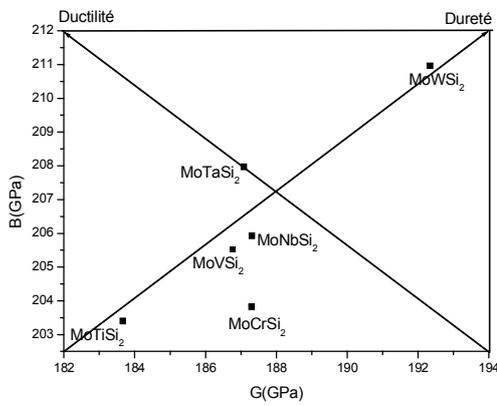
Graph : Pour toute les concentrations :



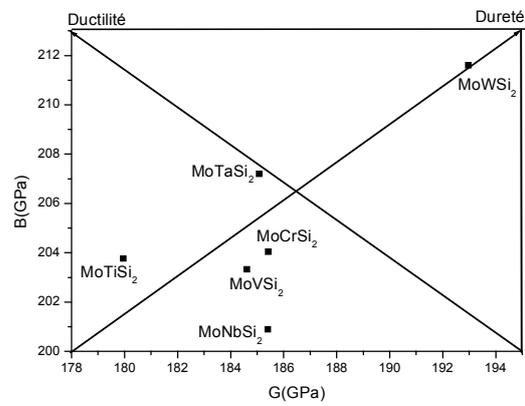
✓ Suivant ces cinq graphes nous pouvons remarqués que les matériaux à base de "Fe"(cluster 1) ( $Fe_{3-x}Ni_xAl$  ;  $Fe_{3-x}Ti_xAl$ ) sont ductiles et que les matériaux à base de Ni(cluster 2) ( $Ni_{3-x}Fe_xAl$  ;  $Ni_{3-x}Ti_xAl$ ) sont dures .Cluster 4 montre aussi les deux matériaux ( $Fe_{3-x}Ni_xAl, Ni_{3-x}Fe_xAl$ ) sont à la fois dure et ductile.

**Graphes IV-6- C : Alliages de type  $Mo_{1-x}X_xSi_2$   $0 \leq x < 1$**

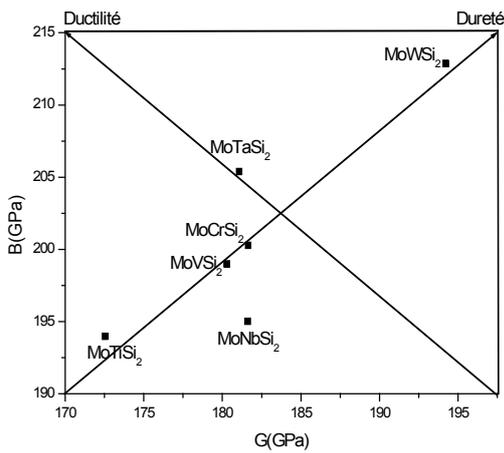
**Pour  $x=0,1$  :**



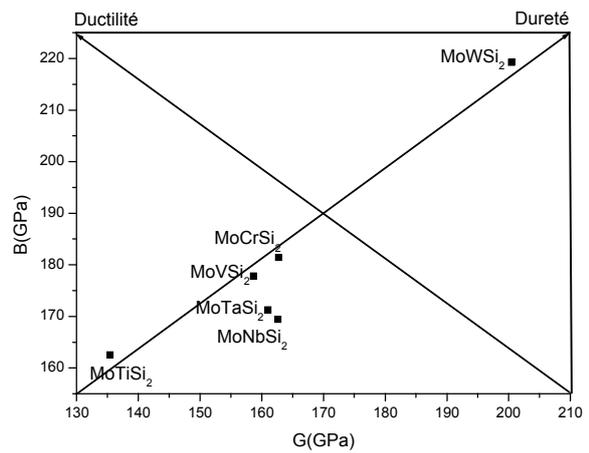
**Pour  $x=0,15$  :**



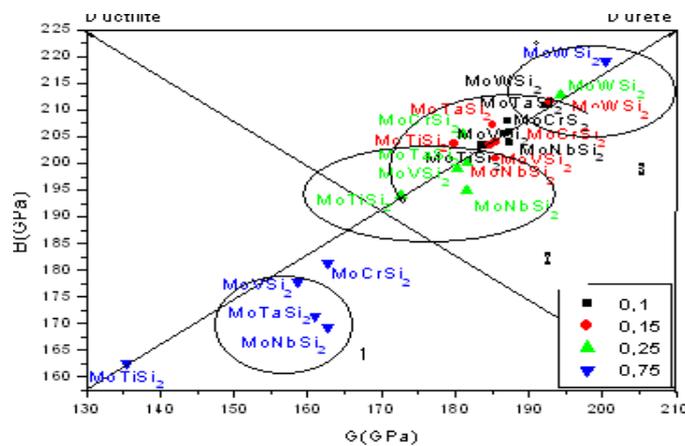
**Pour  $x=0,25$**



**Pour  $x=0,75$**



**Pour toute les concentrations :**



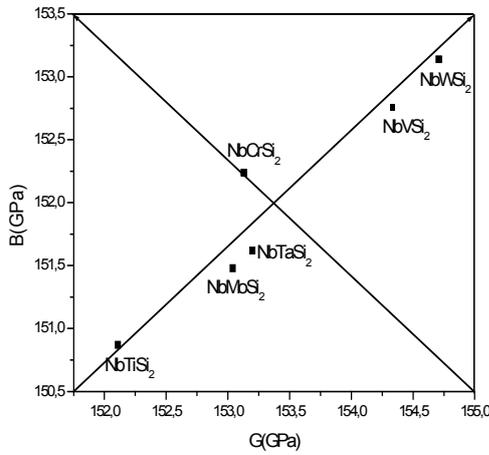
**Figure : Récapitulatif pour les différentes concentrations.**

✓ D'après le graphe ci-dessus, nous ont voit trois clusters :

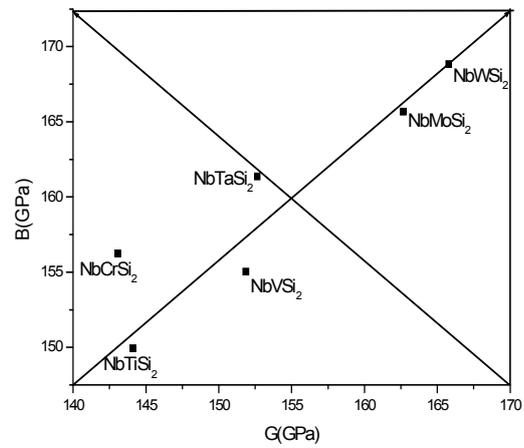
.Le premier cluster : présente des éléments que n'ont pas des propriétés mécaniques intéressants, d'autre le cluster "2" montre des alliages a forte dureté et ductilité en même temps. Le cluster "3" : ce sont des alliages a forte dureté ,nous remarquons qu'il n'existe pas des corrélations avec la concentration.

**Graphes IV -6- D : Alliages type  $Nb_{1-x}X_xAl$   $0 \leq x \leq 1$**

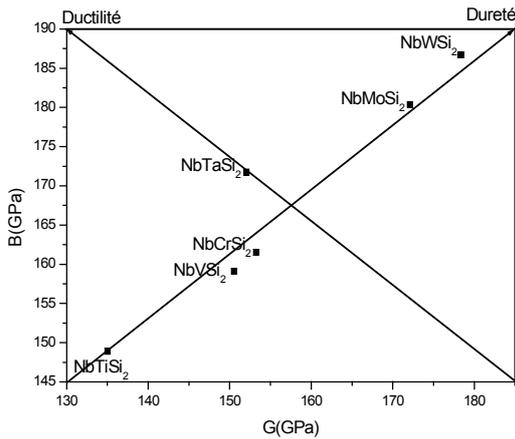
**Pour  $x=0,03$  :**



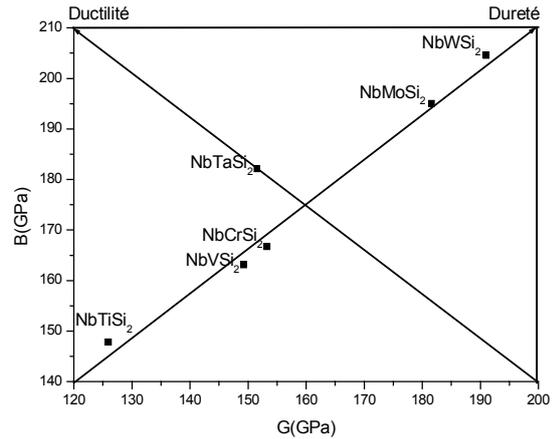
**Pour  $x=0,25$  :**



**Pour  $x=0,5$  :**



**Pour  $x=0,75$  :**



**Grappe : Pour toutes les concentrations :**

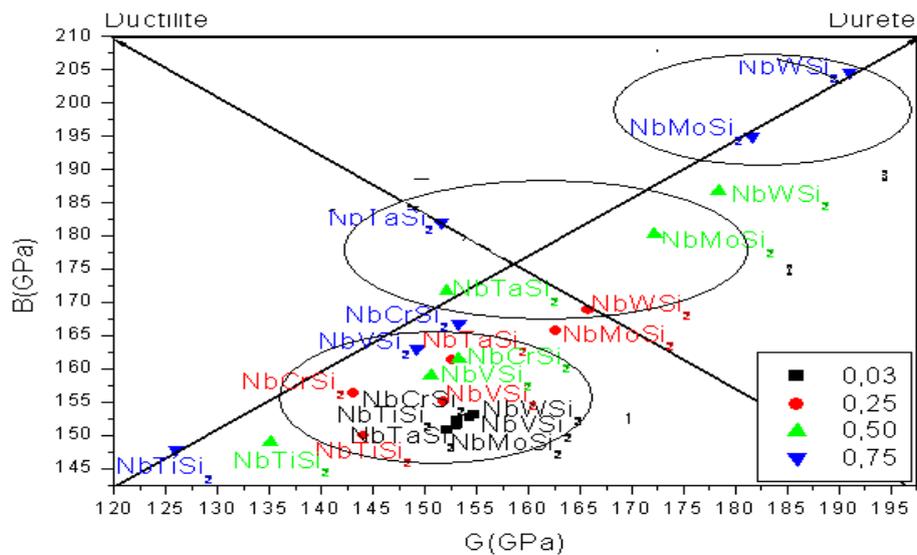


Figure : Grappe récapitulative des différentes concentrations.

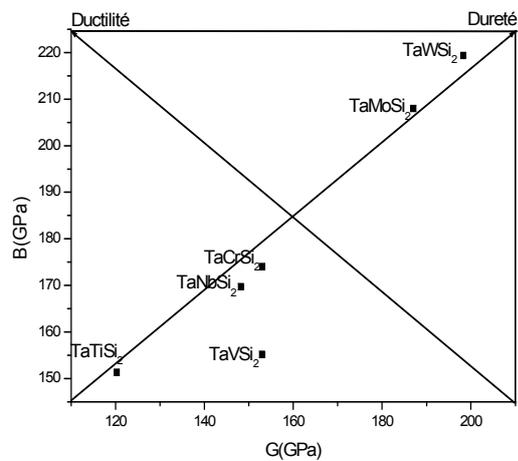
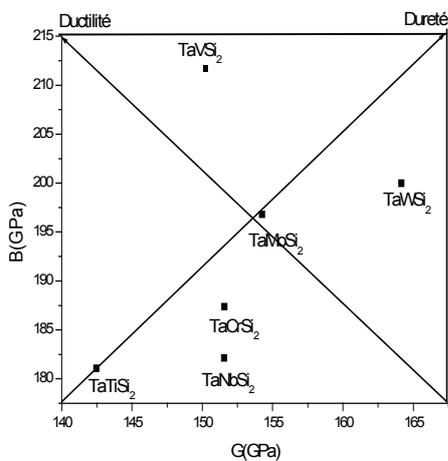
✓ D'après le graphe ci-dessus, nous ont voit trois clusters :

Le premier cluster présente des éléments que n'ont pas des propriétés mécaniques intéressants, d'autre le cluster "2" montre des alliages a forte dureté et ductilité en même temps. Le cluster "3" : ce sont des alliages a forte dureté .

**Graphes IV-6- E : Alliages de type  $Ta_{1-x}X_xSi_2$   $0 \leq x \leq 1$**

Pour  $x=0,25$ :

Pour  $x=0,9$ :



**Graphe : Pour toutes les concentrations :**

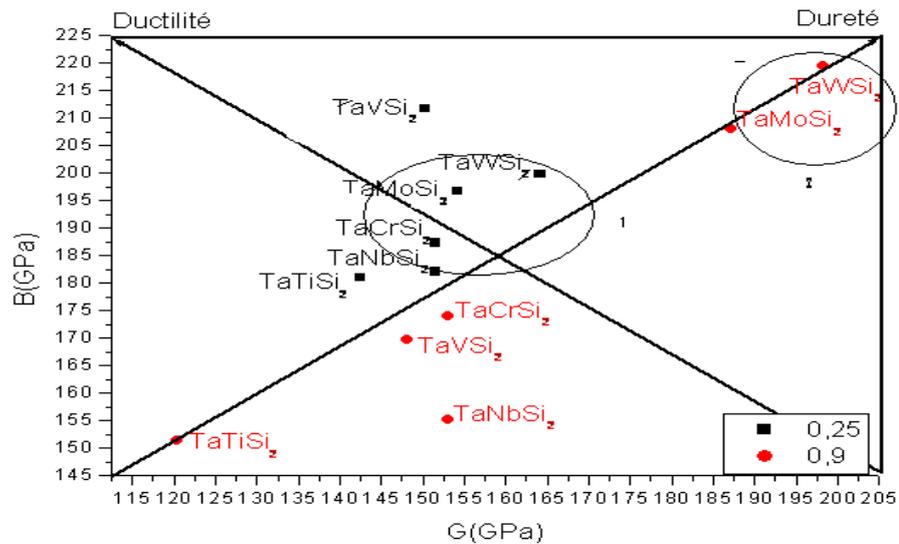


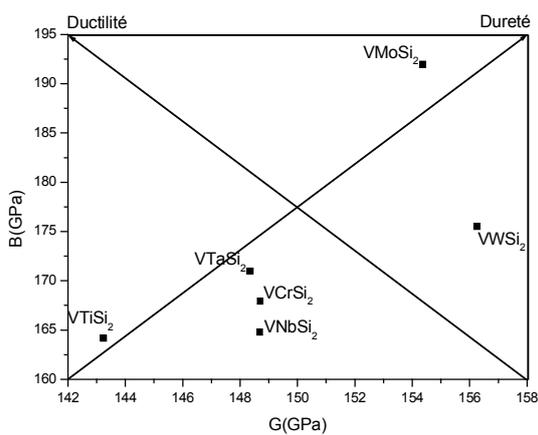
Figure : Graphe récapitulative des différentes concentrations

✓ D'après le graphe ci-dessus, nous ont voit deux clusters :

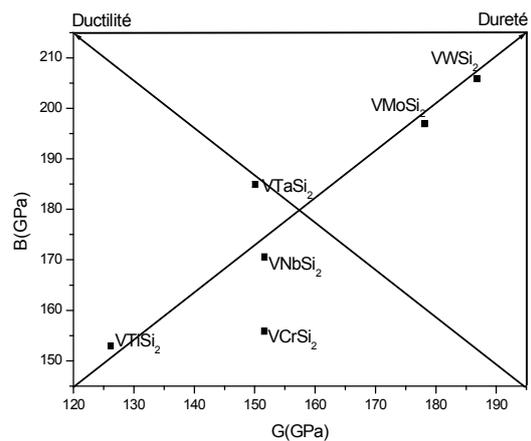
Le premier cluster montre des alliages a forte dureté et ductilité en même temps. Le cluster "2" : ce sont des alliages a forte dureté .

**Graphes IV-6-F : Alliages de type  $V_{1-x}X_xSi_2$   $0 \leq x \leq 1$**

Pour  $x=0,15$  :



Pour  $x=0,7$  :



**Graphe : Pour les deux concentrations :**

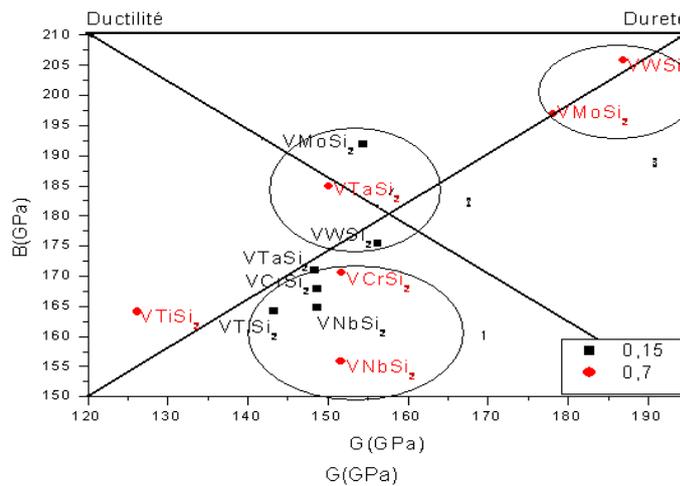


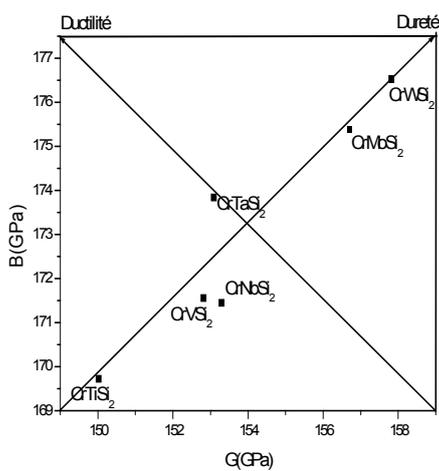
Figure : Graphe récapitulative des différentes concentrations

✓ D'après le graphe ci-dessus, nous ont voit trois clusters :

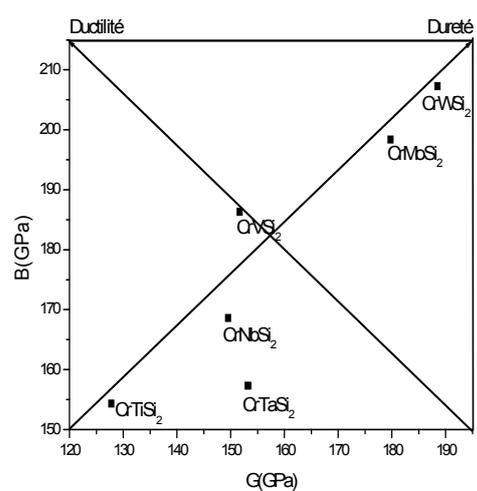
Le premier cluster présente des éléments que n'ont pas des propriétés mécaniques intéressants, d'autre le cluster "2" montre des alliages a forte dureté et ductilité en même temps. Le cluster "3" : ce sont des alliages a forte dureté.

**Graphes IV-6- G : Alliages de type  $Cr_{1-x}X_xSi_2$   $0 \leq x \leq 1$**

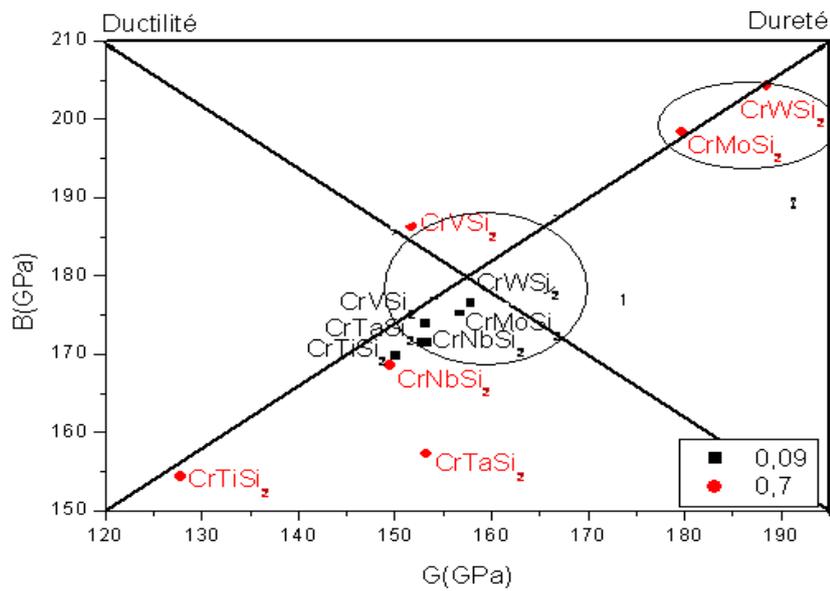
Pour  $x=0,09$  :



Pour  $x=0,7$  :



**Grappe : Pour les deux concentrations :**



**Figure :** Grappe récapitulative des différentes concentrations.

✓ D'après le graphe ci-dessus, nous ont voit deux clusters :

Le premier cluster montre des alliages à forte dureté et ductilité en même temps. Le cluster "2" : ce sont des alliages a forte dureté.

Si nous analysons avec beaucoup d'intérêt les différents graphes nous remarquent que les alliages concernant une forte concentration de W et Mo confèrent une dureté exceptionnelle aux composés.

## *Conclusion Générale*

La question principale de notre travail est: comment `a partir d'une série de propriétés (descripteurs) obtenues par un ensemble des composés intermétalliques peut-on avoir :

- ✓ des variables résumant le mieux l'information portée par ces propriétés ?
- ✓ un bilan de liaison entre les variables ?
- ✓ une représentation plane optimale des individus ?

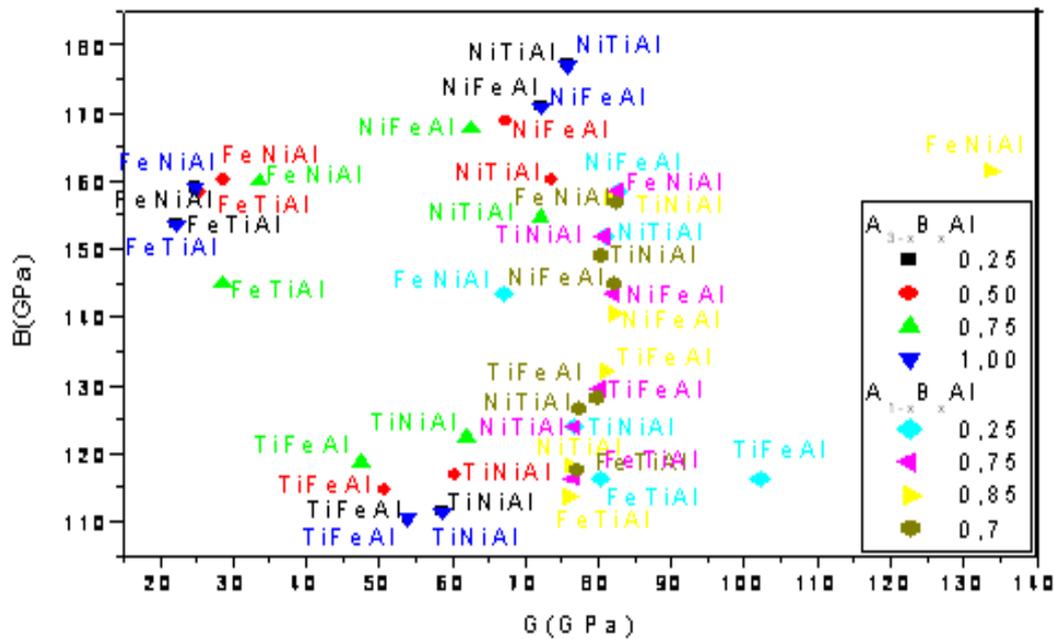
L'Analyse en Composantes Principales (Principal Component Analysis) est une méthode statistique de projection de données multidimensionnelles, utilisée pour la réduction de la dimensionnalité des données. Le but de l'Analyse en Composantes Principales (ACP) est de condenser les données originelles en 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. Ainsi, la première nouvelle composante contient les informations relatives au pourcentage maximal de variance, la deuxième contient les informations relatives au pourcentage de variance suivant. Le processus est répété jusqu'à l'obtention de la dernière nouvelle composante.

La méthode des moindres carrés partiels PLS (**P**artial **L**east **S**quares), qui est une méthode prédictive de régression linéaire à plusieurs variables. Nous a permis d'obtenir une loi linéaire reliant des variables réponses aux variables explicatives.

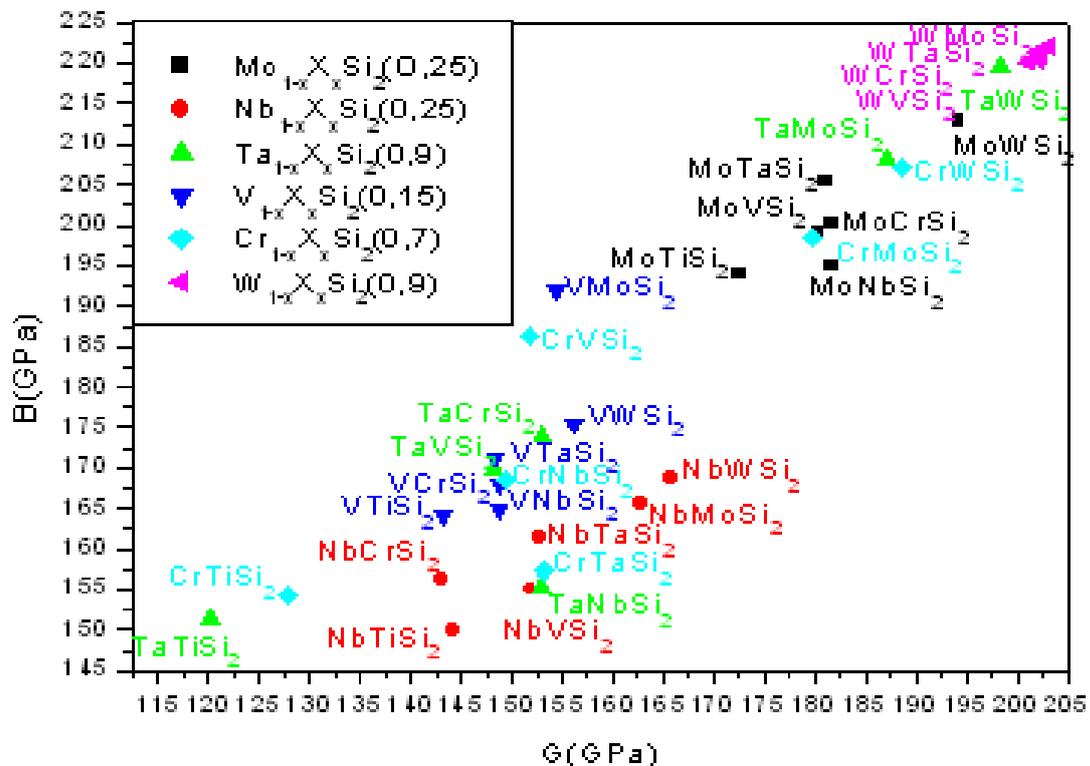
Cette étude nous a permis d'étudier l'ensemble des propriétés mécaniques d'un grand nombre de composés ternaire et de pouvoir les classer. Ainsi sur la base de l'étude par l'ACP nous avons pu ainsi mettre en valeur certains corrélations entre ces propriétés. Enfin nous avons étendu notre étude en utilisant la PLS.

Cette technique nous a permis de mettre en place un modèle pour chaque propriétés ce modèle a ensuite été utilisé pour prédire les propriétés pour un grand nombre d'alliage ternaire. Ainsi sur la base de ces résultats nous avons une cartographie de l'ensemble des propriétés des alliages ternaires a base de  $A_{1-x}B_xAl$ ,  $A_{3-x}B_xAl$ ,  $Mo_{1-x}X_xSi_2$ ,  $Nb_{1-x}X_xAl$ ,  $Ta_{1-x}X_xSi_2$ ,  $V_{1-x}X_xSi_2$ ,  $Cr_{1-x}X_xSi_2$  pour les différents concentrations (voir figure VI-7).

Carte a base de l'Aluminium (Al) :

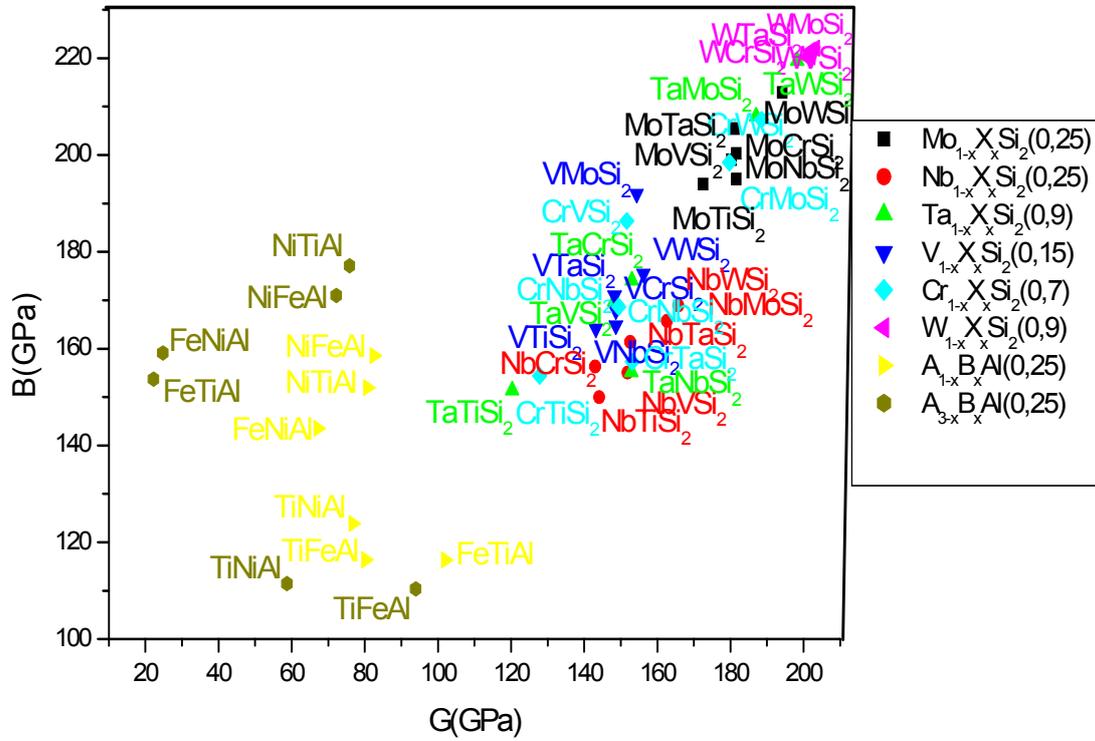


carte a base de silicium :



i

Carte pour toute les alliages intermétalliques prédites :



**Abstract:**

In this thesis, we present a novel approach based on data mining for the design of hard and ultra hard coatings using elastic properties, bulk and shear modulus of intermetallics compounds. Principal component analysis (PCA) and partial least squares (PLS), informatics techniques, have been applied to predict new multilayer ultra hard coatings materials.

**Résumé :**

Dans ce mémoire, nous présentons une nouvelle approche basée sur le datamining, pour la conception de nouveaux matériaux et ceci en utilisant les propriétés élastiques, le module de compressibilité et le module du cisaillement des intermétalliques. L'analyse en composantes principales (ACP) et la régression aux moindres carrés partiels (PLS), des techniques d'informatique, ont été appliquées pour prédire de nouveaux matériaux.

Cette étude nous a permis d'étudier l'ensemble des propriétés mécaniques d'un grand nombre de composés ternaire et de pouvoir les classer. Ainsi sur la base de l'étude par l'ACP nous avons pu ainsi mettre en valeur certaines corrélations entre ces propriétés. Enfin nous avons étendu notre étude en utilisant la PLS.

Cette technique nous a permis de mettre en place un modèle pour chaque propriétés ce modèle a ensuite été utilisé pour prédire les propriétés pour un grand nombre d'alliage ternaire. Ainsi sur la base de ces résultats nous avons une cartographie de l'ensemble des propriétés des alliages ternaires a base de  $A_{1-x}B_xAl$ ,  $A_{3-x}B_xAl$ ,  $Mo_{1-x}X_xSi_2$ ,  $Nb_{1-x}X_xAl$ ,  $Ta_{1-x}X_xSi_2$ ,  $V_{1-x}X_xSi_2$ ,  $Cr_{1-x}X_xSi_2$  pour les différents concentrations

## **Abstract:**

In this thesis, we present a novel approach based on data mining for the design of hard and ultra hard coatings using elastic properties, bulk and shear modulus of intermetallics compounds. Principal component analysis (PCA) and partial least squares (PLS), informatics techniques, have been applied to predict new multilayer ultra hard coatings materials.

The hardness and ductility trends in these materials are identified using.

This study UNO une Permis d'étudier l'ensemble des mécaniques d'Propriétés de l'ONU Grand Nombre de Composés ternaire et de Pouvoir les classer. Ainsi soit sur la base de l'ACP par l'étude pu UNO Avons AINSI law juror Certaine value corrélations Entre Propriétés de la SCÉ. Avons ENFIN UNO Etendu Notre study Exécuter la PLS la.

technique This UNO une Permis de mettre en lieu Modèle des Nations Unies verser Every Propriétés CE Modèle une salle de bains Été Utiliser Pour prédire les POUR UN Propriétés Nombre d'alliage ternaire grand. Ainsi soit sur la base de la SCÉ Résultats UNO Avons UNE cartographie de l'ensemble des Propriétés ternaires des Alliages d'une base de  $A_{1-x}B_xAl$ ,  $A_{3-x}B_xAl$ ,  $Mo_{1-x}X_xSi_2$ ,  $Nb_{1-x}X_xAl$ ,  $Ta_{1-x}X_xSi_2$ ,

$V_{1-x}X_xSi_2$ ,  $Cr_{1-x}X_xSi_2$  Verser les différents concentrations

## **Résumé :**

Dans ce mémoire, nous présentons une nouvelle approche basée sur le datamining, pour la conception de nouveaux matériaux et ceci en utilisant les propriétés élastiques, le module de compressibilité et le module du cisaillement des intermétalliques. L'analyse en composantes principales (ACP) et la régression aux moindres carrés partiels (PLS), des techniques d'informatique, ont été appliquées pour prédire de nouveaux matériaux.

Cette étude nous a permis d'étudier l'ensemble des propriétés mécaniques d'un grand nombre de composés ternaire et de pouvoir les classer. Ainsi sur la base de l'étude par l'ACP nous avons pu ainsi mettre en valeur certains corrélations entre ces propriétés. Enfin nous avons étendu notre étude en utilisant la PLS.

Cette technique nous a permis de mettre en place un modèle pour chaque propriétés ce modèle a ensuite été utilisé pour prédire les propriétés pour un grand nombre d'alliage ternaire. Ainsi sur la base de ces résultats nous avons une cartographie de l'ensemble des propriétés des alliages ternaires a base de  $A_{1-x}B_xAl$ ,  $A_{3-x}B_xAl$ ,  $Mo_{1-x}X_xSi_2$ ,  $Nb_{1-x}X_xAl$ ,  $Ta_{1-x}X_xSi_2$ ,

$V_{1-x}X_xSi_2$ ,  $Cr_{1-x}X_xSi_2$  pour les différents concentrations

## **الخلاصة:**

ف - ي ه - ذه الم - ذكرة، نع - رض صد - فحة جديد - ة نعتم - د عا - مي - DATAMINING ل - وازم جديد - دة و له - ذا نس - تعمل الخ - واصل

L'analyse en composantes principales (ACP) et la régression aux moindres carrés partiels (PLS) هي علاقات إعلامية تستعمل من أجل توقع معادن جديدة.

قمنا بهذا العمل من أجل دراسة الخواص الميكانيكية لموكب ثلاثي المواد و ترتيبها. و هذا في قاعدة دراسة ب:ACP، و جعلنا هذا في قيمة بعض الصلات بين الخواص.

، هذه التقنية تجعلنا نضع مكان لكل نموذج، هذا النموذج نستعمله من أجل توقع PLS و في النهاية نوسع هذه الدراسة باستعمال  $A_{1-x}B_xAl$  الخواص لمجموعة كبيرة لثلاثية المعادن و هذا في قاعدة النتائج نجد الخرائطية لمجموعة من ثلاثية المعادن المكونة ب:

$V_{1-x}X_xSi_2$ ,  $Cr_{1-x}X_xSi_2$ ,  $Nb_{1-x}X_xAl$ ,  $Ta_{1-x}X_xSi_2$ ,  $A_{3-x}B_xAl$ ,  $Mo_{1-x}X_xSi_2$ ,

من أجل مختلف التراكيز.