



الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE



UNIVERSITE AMAR TELIDJI – LAGHOUAT

DEPARTEMENT DES SCIENCES DE LA MATIERE
LABORATOIRE DE PHYSIQUE DES MATERIAUX

Mémoire de Master

Filière : Physique des Matériaux

Option : Physique des Matériaux

Présenté par :

Mlle. HAMDI Alhadja Mamma

THEME

**Etude par datamining de certaines propriétés physiques de
quelques borures, nitrures et carbures de fer**

Devant le jury :

HELIFA Bachir	M. C. A	Univ. de Laghouat	Président
GUEDDOUH Ahmed	M. C. A	Univ. de Laghouat	Examineur
LEFKAIER Iben Khaldoun	Professeur	Univ. de Laghouat	Encadreur
BENGHIA Ali	M. C. B	Univ. de Laghouat	Co-encadreur

juin 2020

Je dédie ce travail

A l'âme de mon cher grand père Bourzeg Aïssa,

A mes parents : HAMDİ Ahmed, Bourzeg Fatima Zohra

A mon frère et mes sœurs :

Abdelkader, Ayoub et sa petite famille,

Soumia et sa petite famille,

Ikram,

Salsabile Tasnime.

A ma Chère grande famille paternelle et maternelle,

A tous mes amis, particulièrement : G.Rym, S.Fadhila, K.Mahdjouba, B.Hafsa, C.Imane.

Remerciements

*Je remercie **ALLAH** le Tout-Puissant de m' avoir donné le courage, la volonté et la patience de terminer ce travail.*

*Ce travail a été réalisé dans le **Laboratoire de Physique des Matériaux** de l'université de Laghouat. Aussi, j'aimerai témoigner ma gratitude et mes remerciements sincères au directeur du laboratoire le professeur **LEFKAIER Iben Khaldoun**. Je tiens également à le remercier, ainsi que **M.BENGHIA Ali**, maître de conférences-B à l'université de Laghouat, pour avoir encadré tous les deux ce mémoire. Leurs patiences, leurs qualités humaines, leur disponibilité à tout temps et surtout leur aide précieuse tout le long de mon travail, m'ont été d'un grand soutien à réaliser ce travail. Je ne saurai les en remercier assez.*

*J'adresse mon respect et mes remerciements sincères à **M. HELIFA Bachir**, Maître de conférences-A à l'université de Laghouat, pour avoir accepté aimablement de présider le jury de délibération de ma soutenance.*

*J'exprime ma reconnaissance à **M. GADDOUH Ahmed**, Maître de Conférences-A à l'université de Laghouat, pour avoir accepté de juger ce travail.*

Ma gratitude sincère va à mes enseignants en licence et en master de physique du département des sciences de la matière pour les efforts consentis afin de nous assurer un enseignement de qualité.

*L'apport de **M.GUELLOUMA Younes**, Maître de conférences à l'université de Laghouat, dans la réalisation de ce travail m'a été très précieux et je le remercie infiniment.*

Que le responsable de la bibliothèque universitaire trouve ici toute ma gratitude pour m' avoir accompagné au cours des cinq années de ma formation.

Enfin, à tous ceux que je n'ai pas cités par oubli simplement, j'aimerai qu'ils acceptent mes sincères remerciements pour leur aide à la réalisation de ce travail de fin d'études.

Sommaire

Introduction générale.....	5
Chapitre I Notions de datamining	7
I.1 Notion du datamining.....	8
I.2 Analyse en composantes principales.....	10
I.2.1 Nuage des points.....	11
I.2.2 Cercle de corrélation.....	13
I.2.3 Biplot des variables et les individus	15
I.2.4 Clustering	16
II Chapitre II. Analyse ACP : Résultats et discussions	18
II.1 Propriétés étudiés familles de composés de fer Fe-X (X= B, C, N)	19
II.2 Outil d'analyse	22
II.3 Valeurs propres	23
II.4 Scree plot	24
II.5 Vecteurs propres	25
II.6 Coordonnées des variables	26
II.7 Cercle de corrélation.....	26
II.8 Contributions des variables	28
II.9 Cosinus carrés des variables.....	28
II.10 Coordonnées des observations	29
II.11 Contributions des observations	31
II.12 Cosinus carrés des observations.....	33
II.13 Nuages des points par R.....	35
II.14 Biplot des variables et des individus par R	37
II.15 Clustering (k-means) par R	40
II.16 Corrélations des variables avec le moment magnétique μ_B	44
Conclusion générale	48
Références bibliographiques	49

Liste des figures

Figure 1 Synoptique du Datamining.....	9
Figure 2 Tableau des données pour l'ACP.....	10
Figure 3 Projection d'un individu sur un axe.....	12
Figure 4 Représentation des coordonnées du nuage de points synthétisés par l' ACP.....	12
Figure 5 Projection du coefficient de corrélation sur le cercle de corrélation en 3D.....	13
Figure 6 La méthode d'interprétation des variables dans deux dimensions.....	14
Figure7 Méthode d'interprétation de des variables et des individus.....	16
Figure 8 l'interface de logiciel R.....	23
Figure 9 le script de la ligne de la commande de R.....	23
Figure 10 Le Scree plot de pourcentages des variabilités par R.....	25
Figure 11 Cercle de corrélation par R.....	27
Figure 12 Nuages de points des individus (Fe-X (X :B, C, N) Avec R.....	36
Figure 13 Biplot des variables et les individus.....	38
Figure 14 Biplot de classification (k-means) par R.....	41
Figure 15 Cercles de corrélations des variables. a) axe principaux F1 et F3 , b) axes principaux F2 et F3.....	46

Liste des tableaux

Tableau 1 Jeu des données des individus composants de (Fe-X(X: B, N,C) et leurs propriétés	20
Tableau 2 Valeurs propres obtenues par R.....	24
Tableau 3 Vecteurs propres des variables	25
Tableau 4 Coordonnées des variables	26
Tableau 5 Contributions des variables (%).	28
Tableau 6 Valeurs des cosinus carrés des variables	29
Tableau 7 Coordonnées des observations	30
Tableau 8 Valeurs des contributions des observations (%).	32
Tableau 9 Valeurs de cosinus carrés des observations.....	33
Tableau 10 Valeurs moyennes dans le Biplot	39
Tableau 11 Les classifications des groupes (K-means).....	43
Tableau 12 Coordonnées des centres de gravité des groupes	44

Liste de symboles

E	Module de Young
G	Module de cisaillement
K	Rapport G/B
B	Coefficient de compressibilité isostatique
V	Volume molaire
ν	Coefficient de Poisson
ρ	Densité
HV	Dureté de Vickers
μ_B	Moment magnétique

Introduction générale

De nos jours, le nombre d'expériences et de calcul par simulation des propriétés physiques et chimiques des corps solides sont énormes. Depuis des décennies que les chercheurs tentent par l'analyse de données d'en tirer profit, pour voir des comportements corrélés, des tendances par groupe de matériaux. De telles analyses peuvent conduire par projection à prédire d'autres matériaux nouveaux virtuels ayant des propriétés souhaités. Une démarche de la sorte est dite « datamining ». Elle permet donc de court-circuiter dans le temps la synthèse d'un matériau souhaité par méthode expérimentale couteuse en temps et en argent. L'application du datamining est faite depuis longtemps dans les domaines de la biologie, de l'économie, des phénomènes stochastiques etc.... Pour l'étude des matériaux solides, cette application a tardé de se faire. En 2005, pour la première fois le projet dit 'Génome des matériaux', qui n'est autre l'extraction de l'information utile à partir de l'application des méthodes d'analyse des données sur des méga-données de propriétés de matériaux solides, a été proposé en Algérie . Depuis lors, le projet Génome n'a pas eu une de suite en Algérie. En revanche, ce projet fut lancé aux USA depuis 2007. En juin 2011, à l'université de Carnegie-Mellon, le Président Obama fait un discours important où il a déclaré que *«le projet du génome des matériaux est le renouveau de la science pour les Etats-Unis et une révolution dans le domaine de la science»*. Les USA investissent dès lors 100 millions de dollars annuellement dans ce projet. Ceci dit, cet engagement démontre à quel point cette nouvelle vision de la science des matériaux qui est le datamining est importante pour le développement technologique.

Actuellement, l'une des méthodes de datamining la plus populaire est l'Analyse en Composantes Principales (ACP). Elle fait partie du groupe des méthodes descriptives multidimensionnelles, appelées méthodes factorielles. Ces méthodes, apparues au début des années 30, ont été surtout développées en France dans les années 60 [1], en particulier par Jean-Paul Benzecri qui a beaucoup exploité les aspects géométriques et les représentations graphiques. Cette méthode fut appliquée dans beaucoup de recherche de datamining des matériaux. Il est bien connu que l'ACP est une méthode mathématique intégrée dans plusieurs logiciels statistiques de programmation, telle que R et XLstat, afin d'en tirer des informations utiles à partir d'un grand jeu des données.

Notre projet de fin d'étude s'inscrit dans cette optique. Il vise à analyser par la méthode ACP un bon nombre de composés de fer « Fe-X (X=B, N, C) » et d'établir des corrélations entre certaines propriétés physiques, mécaniques et magnétiques et de classer en clusters les matériaux étudiés.

Il est important de noter que c'est la première fois que cette méthode est abordée dans un projet de fin d'étude en physique au département de SM.

Notre manuscrit est organisé en deux chapitres. Dans le premier, on y présente en bref le datamining, sa définition, ses principes et ses objectifs. Dans le second chapitre, on montre l'application de la méthode ACP sur la base de données riche que nous avons récolté de la littérature. Nous essayons ensuite d'en déduire de l'information utile qui fera l'objet ensuite d'une conclusion générale.

Chapitre I.

Notions de datamining

I.1 Notion du datamining

Le datamining, appelé Knowledge Discovery in Databases (KDD), est le domaine de la découverte d'informations nouvelles et potentiellement utiles à partir l'analyse de grandes quantités de données. L'exploration de données a été appliquée dans un grand nombre de domaines, notamment les ventes au détail, la bio-informatique et la lutte contre le terrorisme. Ces dernières années, il y a eu un intérêt croissant pour l'utilisation de l'exploration de données pour étudier les questions dans le cadre de la recherche scientifique [2].

Datamining est le processus d'exploration et d'analyse par des moyens automatiques ou semi-automatiques, de grandes quantités de données afin de découvrir des modèles et des règles significatifs. L'exploration de données consiste à résoudre des problèmes en analysant des données déjà présentes dans les bases de données, il est un sous-domaine interdisciplinaire de l'informatique qui implique un processus informatique de découverte de modèles de grands ensembles de données. L'objectif de ce processus d'analyse avancé est d'extraire des informations et les transformer vers une structure compréhensible pour une utilisation ultérieure [3].

L'exploration de données se compose de plusieurs éléments principaux:

- Extraire, charger, Stocker et gérer les données dans un système de base de données multidimensionnelle.
- Analyser les données par logiciel d'application.
- Présenter les données dans un format utile, tel qu'un graphique ou un tableau.
- Interpréter et conclure des résultats de travail.

Les fonctionnalités d'exploration de données sont utilisées pour spécifier le type de modèle à trouver dans les tâches d'exploration de données.

Les tâches d'exploration de données peuvent être classées en deux catégories principales descriptives et prédictives :

- **Descriptive** : ce terme est essentiellement utilisé pour produire la corrélation, la tabulation croisée, etc... . Il détermine ce qui s'est passé dans le passé. Il nécessite l'agrégation et l'exploration de données exact et bien précise.
- **Prédictive** : le but principal de cette exploitation minière est de dire quelque chose sur les résultats futurs et détermine ce qui peut arriver à l'avenir avec l'aide des analyses des données passées. L'approche prédictive utilise les fonctions de modélisation

prédictive, prévisions, simulation et alertes. Elle effectue l'induction sur les données actuelles et passées afin de pouvoir faire des prédictions.

Dans chaque catégorie il y a plusieurs méthodes. Aujourd'hui, on utilise le datamining descriptive qui contient deux méthodes : l'analyse de données et le clustering. La figure 1 présente la synoptique l'approche descriptive de l'ACP.

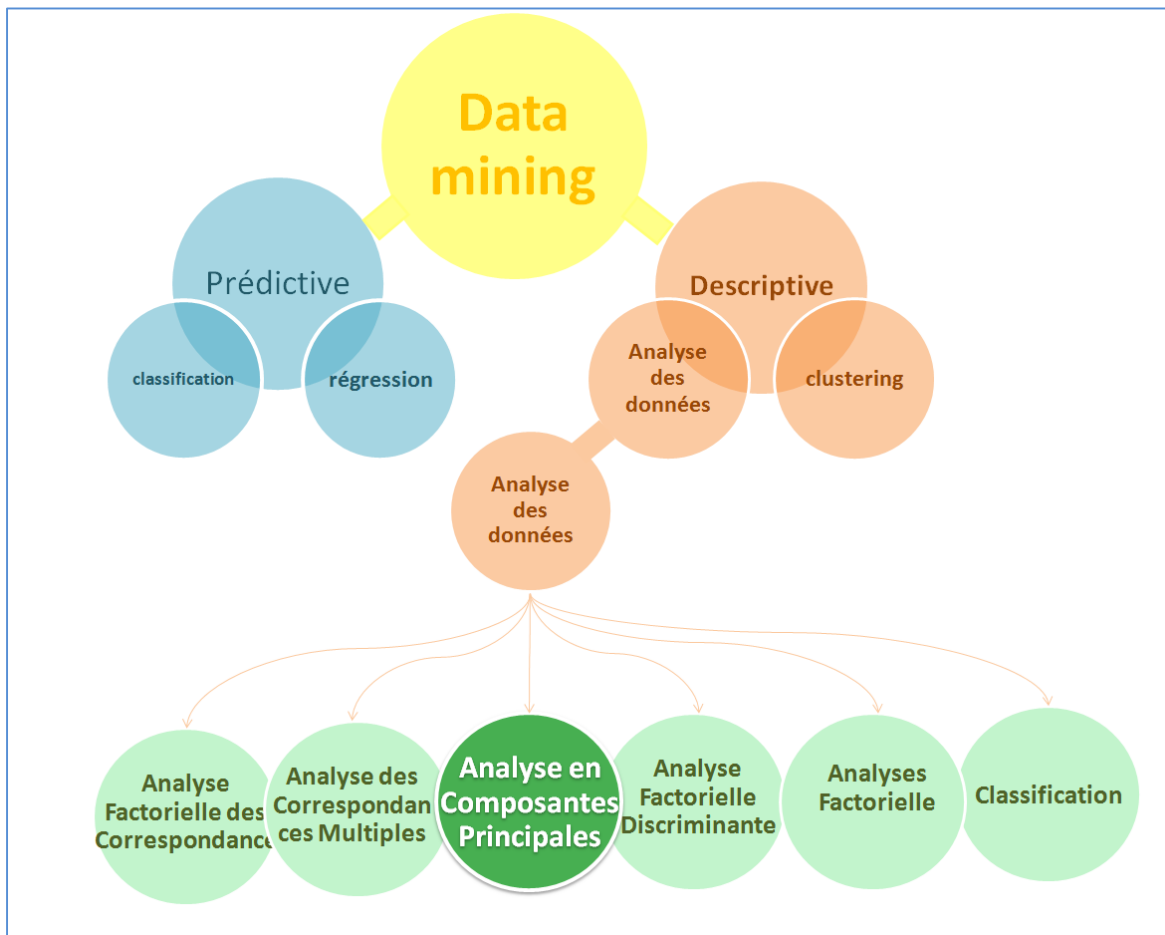


Figure 1. Synoptique du Datamining

I.2 Analyse en composantes principales

L'analyse en composantes principales (ACP) est conçue par Karl Pearson en 1901. L'ACP est une méthode de statistique exploratoire et descriptive multidimensionnelle. Cette méthode synthétise, résume, remarque les informations contenues dans un tableau de données quantitatives, représente et visualise par des graphique simples.

Les données pour l'ACP sont présentées sous forme de tableau des données (figure 2). Ces données sont constituées **d'individus** et de **variables** qui doivent être quantitatives, continues et corrélées entrent-elles.

x_{ik} sont des valeurs de la variable k pour l'individu i avec $k = 1, \dots, K$ et $i = 1, \dots, I$.

I représente à la fois le nombre d'individus et l'ensemble des individus $\{1, \dots, i, \dots, I\}$, de même

K représente le nombre de variables et l'ensemble des variables $\{1, \dots, k, \dots, K\}$ [4].

		variables		
		1	$\dots \dots \dots k$	$\dots \dots \dots K$
Individus	1			
	$\dots \dots \dots$		$\dots \dots \dots$	
	i	$\dots \dots \dots$	x_{ik}	$\dots \dots \dots$
$\dots \dots \dots$			$\dots \dots \dots$	
	I			

Figure02 .Tableau des données pour l'ACP.

Il y a des différentes façons d'étudier le tableau de la figure 2. On peut le considérer ce tableau comme un ensemble de lignes et chercher les différences et les ressemblances qu'il peut y avoir d'une ligne à l'autre ou bien on peut considérer le tableau comme un ensemble et colonne et chercher avoir les ressemblances entre colonne.

Pour l'étude les individus, on va chercher les ressemblances entre eux à partir le point de vu des variables et on construire des groupes des individus homogènes et identique.

On niveau des variables d'une façon symétrique, on va rechercher les ressemblances entre eux, (les variables constituer des informations à peu près identiques),

La ressemblance des variables dépend des liaisons entre les variables. La liaison la plus connue est la liaison linéaire. Elle est simple, très fréquente et résume de nombreuses corrélations.

La ressemblance entre des variables est mesurée par le coefficient de corrélation des variables. Si deux variables contiennent un coefficient de corrélation très élevé, cela veut dire que les deux variables apportent la même information. Aussi pour une bonne visualisation, on utilise un petit nombre d'indicateurs synthétiques pour priori (ex. la moyenne, l'écart-type) [5].

La moyenne :

$$\bar{x}_k = \frac{1}{I} \sum_{i=1}^I x_{ik} \quad (1)$$

L'écart-type :

$$s_k = \sqrt{\left[\frac{1}{I} \sum_{i=1}^I (x_{ik} - \bar{x}_k)^2 \right]} \quad (2)$$

Afin d'établir un lien entre les individus et des variables, nous construisons des groupes d'individus et nous allons chercher à caractériser des différentes classes d'individus par leurs variables.

Il y a deux façons de considérer le tableau de données. Soit nous le considérons comme un ensemble de lignes (étudier les individus), soit comme un ensemble de colonne (étudier les variables). L'objectif est d'avoir un nuage de points. Chaque point correspondant dans un espace de K dimensions [5] :

- K= 1 : représentation axiale ;
- K= 2 : nuage des points classique ;
- K= 3 : représentation en 3D (c'est difficile mais certains logiciels permettent de visualiser des dimensions en plan bien sûr) ;
- K= 4 : impossible de représenter, mais le concept mathématique est simple (grand K coordonnées sur grand K dimensions).

I.2.1 Nuage des points

Pour l'étude des individus, on utilise une méthode graphique de visualisation de nuage d'individus. A cette fin, nous définissons la notion de ressemblance. Quand deux individus se ressemblent, on prend des valeurs proches grâce à la définition de la distance entre deux individus (la distance carré de Pythagore).

La notion de ressemblance est donc définie par l'expression mathématique suivante:

$$d^2(i, i') = \sum_{k=1}^k (x_{ik} - x_{i'k})^2 \quad (3)$$

Pour trouver une bonne représentation, on va commencer par trouver le premier axe et aussi le premier facteur qui déforme le moins possible la projection du centre de gravité du nuage sur cet axe. La figure 3 illustre cette notion.

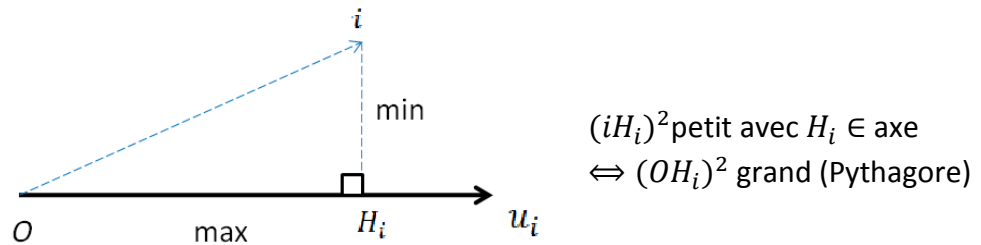


Figure 03. Projection d'un individu sur un axe

H_i est la projection des individus sur l'axe u_i , O le centre de gravité de nuage, iH_i^2 est l'écart entre l'individu i dans l'espace initial. L'axe u_i doit être contenu dans un plan qui soit le meilleur, celui qui assure le plus petit écart possible entre un individu et sa projection et la plus grande somme des distances OH_i^2 possible entre les individus.

Pour trouver le meilleur plan, on va chercher le premier meilleur axe, ensuite un axe orthogonal sur le premier et maximiser la somme $\sum_i (OH_i)^2$, ensuite le deuxième axe, puis un troisième et séquentiellement chercher les axes les uns après les autres et à chaque fois l'axe doit être orthogonal à tous les précédents. Ces axes sont appelés « axes principaux » et sont notés F_1, F_2, F_3, \dots .

On peut avoir une bonne connaissance des données à partir du jeu de données. On va vouloir utiliser les variables pour interpréter la variabilité des axes, On va considérer les coordonnées des individus sur les axes. Par exemple, si l'on considère deux axes orthogonaux F_1 et F_2 , la coordonnée de l'individu i sur le premier axe est noté F_{i1} et sur le deuxième axe elle est notée F_{i2} . Donc pour chaque individu nous collectons ses coordonnées sur l'axe horizontal et sur l'axe vertical. On peut ainsi constituer deux vecteurs F_1 et F_2 (figure 04) qui regroupent toutes les coordonnées de tous les individus sur le premier et un deuxième axe. Les F_{i1}, F_{i2}, \dots sont appelés les facteurs ou bien les dimensions de l'ACP (figure 04).

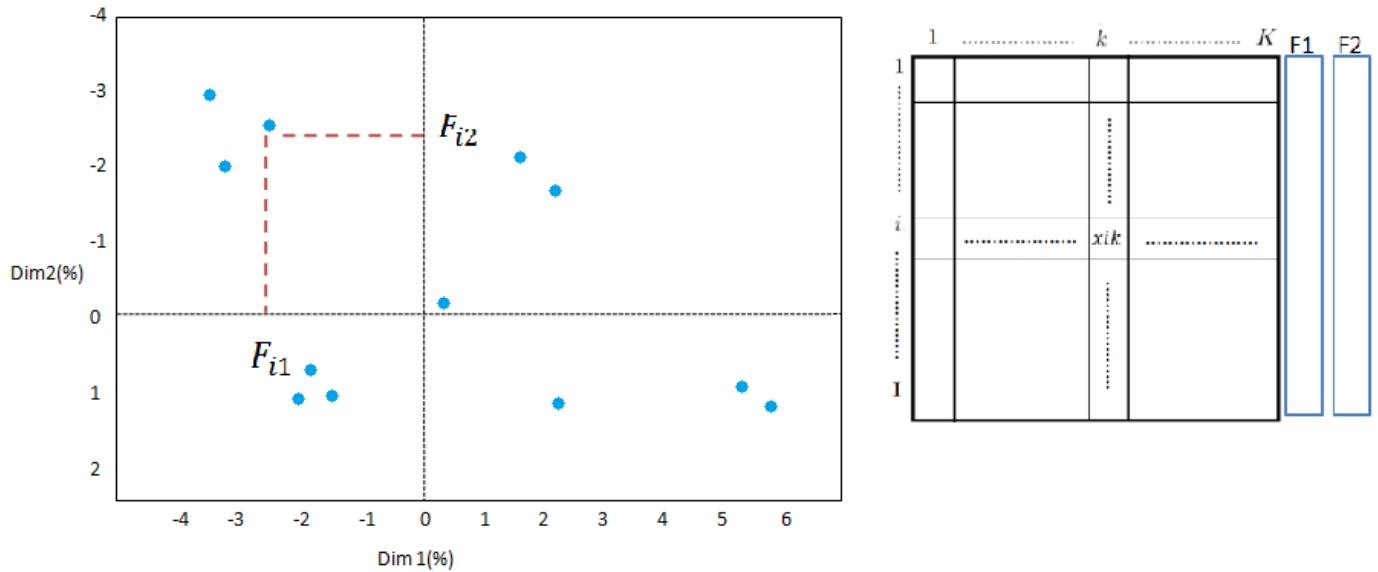
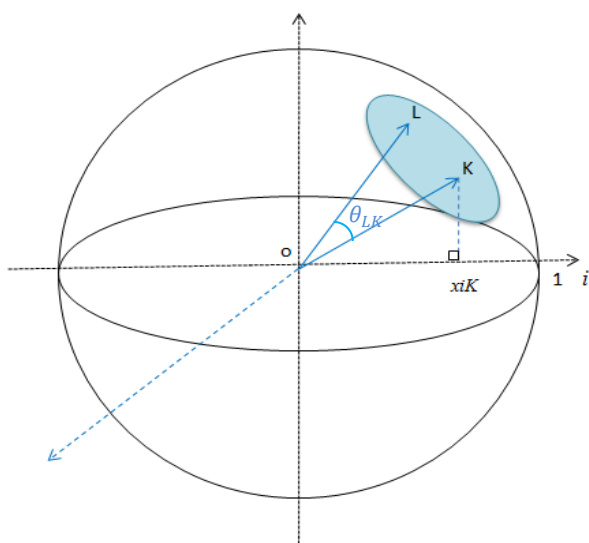


Figure 04. Représentation des coordonnées du nuage de points synthétisés par l'ACP

I.2.2 Cercle de corrélation

Pour l'étude des variables, on utilise le cercle de corrélation. Nous rappelons qu'une variable est un point dans un espace à grande dimension, par ce qu'il y a un grand nombre d'individus. C'est donc un vecteur de beaucoup de coordonnées de nuage et qui seront représentés dans un cercle de corrélation par des flèches en ACP.

Dans l'espace à 3 dimensions (choix de 3 axes F1, F2, F3), la corrélation est faite à partir des angles entre les flèches des variables (figure 5). On y définit ainsi le cosinus suivant :



$$\cos \theta_{LK} = \frac{\langle x_L, x_K \rangle}{\|x_L\| \|x_K\|} \quad (4)$$

$$\cos \theta_{LK} = \frac{\sum_{i=1}^I x_{iL} x_{iK}}{\sqrt{\sum_{i=1}^I x_{iL}^2} \sqrt{\sum_{i=1}^I x_{iK}^2}} \quad (5)$$

$$\cos \theta_{LK} = r(x_K, x_L) \quad (6)$$

Figure 5. Projection du coefficient de corrélation sur le cercle de corrélation en 3D

Deux variables K et L sont représenté par des flèches partant de l'origine des axes. Nous définissons le cosinus (θ_{LK}) (entre la flèche qui représente la variable L et flèche K) comme étant le produit scalaire entre la variable K est la variable L divisé par la norme de la variable K est la norme de la variable L (équation 4). Ce cosinus est noté $r(x_K, x_L)$ (équation 6) et qui est appelé « coefficient de corrélation » entre les vecteurs K et L.

Une projection sur un plan (choix d'axes F1 et F2 par exemple), nous obtenons un cercle des variables utile pour interpréter la signification des axes. Ce cercle aide à caractériser et à représenter les individus et aussi à donner une représentation optimale de nuage des variables est également une visualisation de l'ensemble des corrélations entre les variables grâce au cosinus des angles c'est aussi une visualisation de la matrice de corrélation.

Pour la projection, on a vu que dans l'espace le coefficient de corrélation entre les variables A et B par exemple est égal au cosinus de l'angle entre la variable A et B, soit $\cos \theta_{A,B} = r(A, B)$, mais attention cette égalité est vraie pour l'angle dans l'espace. Quand est-il dans lorsqu'on le visualise dans le plan de projection ?

La Figure 6 illustre un exemple de projection de variables dans l'espace est dans le plan de projection à deux dimensions (cercle de corrélation).

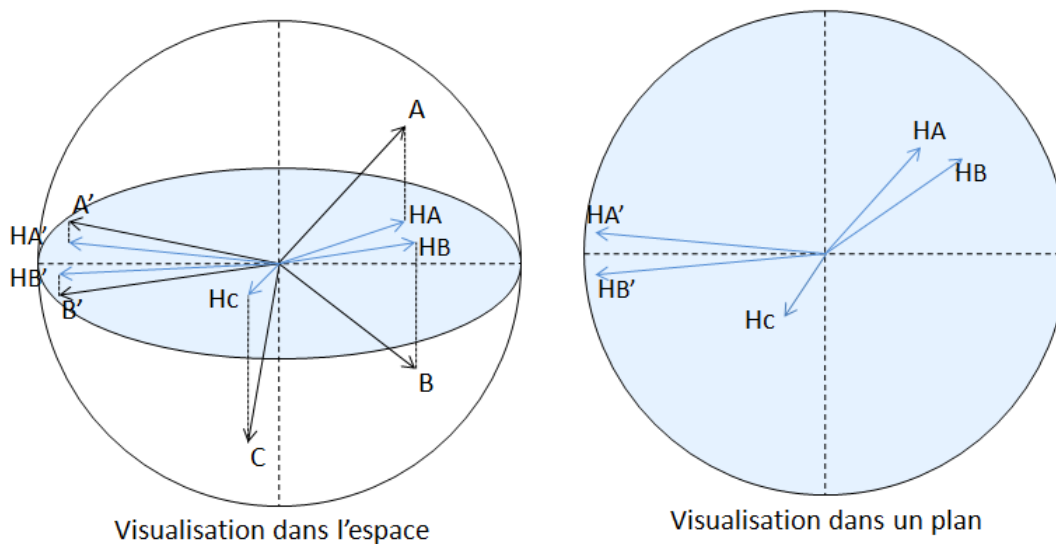


Figure 6. La méthode d'interprétation des variables dans deux dimensions

Dans cette Figure 6, le cosinus $\cos \theta_{A',B'}$ de l'angle entre vecteurs A' et B' est très proche du cosinus de l'angle entre la projection des vecteurs $\cos \theta_{HA',HB'}$, donc le $\cos \theta_{A',B'}$ est très proche du coefficient de corrélation. Par contre, quand les variables sont mal projetées, ce qui est le cas par exemple pour les variables A et B, l'angle ici dans le plan de projection est

petits alors que l'angle entre A et B dans l'espace est très grand. La même constatation est faite pour le vecteur C et sa projection.

Ainsi, pour une bonne interprétation il faut toujours que les variables soient bien projetées, *i.e.* $\cos \theta_{A,B} \approx \cos \theta_{HA,HB}$. Seules les variables bien projetées peuvent être interprétées.

L'interprétation du cercle de corrélation se fait comme suit :

- Si les vecteurs des variables sont courts (loin du bord du cercle), on ne peut pas l'interpréter parce qu'il y a une grande perte d'informations de ces variables.
- Si l'angle est presque droit entre deux vecteurs, il n'y a pas de corrélation entre les deux variables.
- Si les deux vecteurs sont dans la même direction, il y a une corrélation positive entre les deux variables. Si l'une des variables augmente l'autre croît.
- Si les deux vecteurs sont dans le sens opposé, il y a une corrélation négative. Si la première variable croît, la seconde décroît.

I.2.3 Biplot des variables et les individus

Ce type de graphe exploratoire, utilisé en statistique, est un graphique qui combine les individus et les variables sur un nuage des points selon des axes des variables à deux dimensions. Ainsi, il représente les nuages des individus et le cercle de corrélation sur un même graphe, d'où le nom de biplot.

Le graphique (figure 7) explique la méthode d'interprétation de biplot. C'est une méthode d'interprétation des individus et des variables en même temps. Dans cette figure on considère un exemple de 6 variables, présentés par des vecteurs V_1, V_2, V_3, V_4 et V_5 .

Dans la figure 7 nous remarquons qu'il y a une corrélation positive entre V_1 et V_2 , une autre entre V_3 et V_4 et une corrélation négative entre V_1 et V_5 et une autre entre V_2 et V_4 . Il n'y a pas de corrélation entre les vecteurs orthonormés V_1 et V_3 ou V_2 et V_3 .

Egalement dans la figure 7, les individus qui appartiennent à la région O et ses environs ont des valeurs moyennes en tous les variables. Les individus qui appartiennent à la région A et ses environs ont des fortes valeurs en V_3 . Les individus qui appartiennent à la région B et ses environs ont des fortes valeurs en V_1, V_2 et des faibles valeurs en V_4, V_5 parce que c'est une corrélation négative ainsi de suite pour les autres régions.

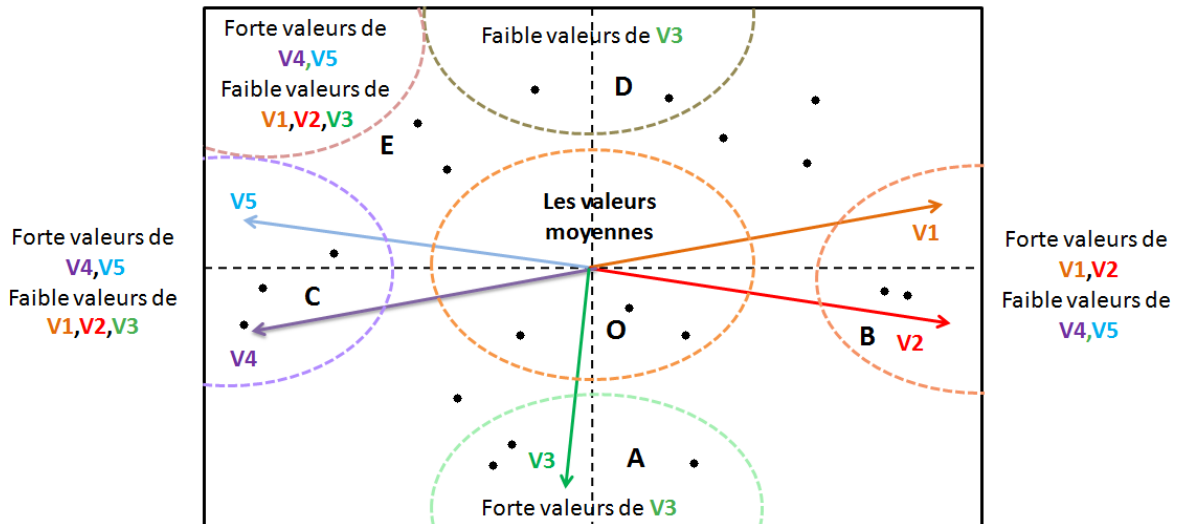


Figure 7. Méthode d'interprétation de des variables et des individus.

I.2.4 Clustering

Connu en langue française sous le nom de « classification hiérarchique », en anglais « *hierarchical clustering* », simplifié au mot le plus utilisé « clustering » est l'une des approches les plus importantes pour l'exploration des données multivariées. L'objectif est d'identifier des groupes (i.e., clusters) d'objets similaires dans un jeu de données. Le clustering retranscrit donc une démarche de classification automatique d'un ensemble de composantes décrites par leurs propriétés.

Brièvement, les deux stratégies de classification les plus courantes sont:

- **La classification ascendante hiérarchique (CAH)**, utilisée pour identifier des groupes d'observations similaires dans un jeu de données.
- **Le Partitionnement en K-moyennes (k-means en anglais)**, utilisé pour diviser un jeu de données en plusieurs groupes. C'est notre sujet d'intérêt.

La classification par la méthode des nuées dynamiques (*k-means*) a été développé par MacQueen en 1967 [6]. Cette méthode est une méthode itérative. Elle consiste à positionner au départ des barycentres au hasard dans les nuages des n points des individus de données (x_1, x_2, \dots, x_n) . Puis, par itération, on cherche à partitionner ces n points en k ensembles, tels que $S = \{S_1, S_2, \dots, S_k\}$ ($k \leq n$), tout en minimisant la distance entre les points à l'intérieur de

chaque partition par la formule suivante : $\arg \min_S \sum_{i=1}^k \sum_{x_j \in S_j} \|x_j - C_i\|$, où C_i est le barycentre des points dans l'ensemble des nuées S_i . Cela veut dire que se sont les points C_i qui vont se déplacer chaque fois dans chaque ensemble par chaque itération jusqu' à ce qu'il deviennent stables (convergence du calcul), assurant ainsi la minimisation des distances entre le centre de gravité et l'ensemble des points du nuage entourant dans la nuée considérée [7]. Ainsi, nous obtenons en fin de compte des regroupements en k clusters, dont le centre de gravité de chacun est C_i .

Chapitre II.

Analyse ACP : Résultats et discussions

I.3 Propriétés étudiées familles de composés de fer Fe-X (X= B, C, N)

Le choix de la famille des borures, des carbures et des nitrures de fer est fait en raison de leurs excellentes propriétés mécaniques. Ces composés de métal de transition (Fe) avec les éléments légers de 2ème ligne dans tableau périodique (B, C, N) possèdent des propriétés mécaniques et chimiques recherchés dans l'industrie. Les borures de fer offrent d'excellentes propriétés de surface en termes de dureté élevée et de stabilité mécanique à haute température et d'usure à la corrosion. De plus, le magnétisme du fer et de certains borures de fer jouent un rôle crucial pour le développement électrique et mécanique [6]. Le carbure de fer en l'occurrence la cémentite est un composé très dur mais très fragile et sa température de fusion est très élevée. Il est ferromagnétique à température ambiante [8]. Les nitrures de fer sont recommandés pour toutes pièces en acier, soumises à l'usure, au grippage, à la fatigue, telles que les cames, les vis d'extrusion, les outillages, les glissières, engrenages et toutes pièces en acier inoxydable.

Pour constituer une bonne base de données pour les trois familles Fe-X (X : B,C,N)), il est nécessaire de bien chercher dans la littérature le maximum de données physiques disponibles de manière à ce qu'elles soient de même nature pour ces trois familles. Le gap de données est exclu de la base de données convoitée.

Ainsi, nous avons sélectionné certaines propriétés mécaniques et structurales, en l'occurrence le système cristallin, le module de Young (E), le module de compressibilité isostatique (B) , le module de Coulomb (G), dit aussi module de cisaillement, la densité (ρ), la dureté de Vickers (HV), le volume molaire (V) , le rapport $K=G/B$, le coefficient de poisson (ν), et enfin le moment magnétique (μ_B).

Le tableau 1 présente les éléments sélectionnés et leurs propriétés. Il est important de noter que certaines valeurs du coefficient de poisson ν et la dureté HV ont été calculées à partir des relations suivantes [9]:

$$H_V = 2(K^2G)^{0.585} - 3 \quad (7)$$

$$\nu = (3B - 2G)/2(3B + G) \quad (8)$$

Le tableau 1 représente le jeu de données des éléments étudiés, soit 53 composés de Fe-X (X : B, C, N) [10].

Tableau 1. Jeu des données des individus composants de (Fe-X(X: B, N,C) et leurs propriétés [10].

N°	Individus	Groupe d'espace	E (GPa)	B (GPa)	G (GPa)	ρ	HV	V (Å ³)	K=G/B	ν	μ_B
1	Fe3B (tétraogonal)	$\bar{I}4$	343	255	135	7,58	13,686	156,33	0,528	0,275	5,868
2	FeB2 (hexagonal)	P6/mmm	323	260	125	5,9	11,288	21,82	0,480	0,293	0,783
3	Fe2B (tétraogonal)	P4/mmm	109	174	39	6,64	0,004	61,30	0,226	0,395	4,194
4	Fe3B (orthorhombique)	Pnma	342	254	134	7,54	13,685	157,14	0,529	0,275	6,069
5	FeB (orthorhombique)	Cmcm	375	239	151	6,89	19,024	32,11	0,632	0,239	1,253
6	FeB (orthorhombique)	Pmna	363	291	141	6,95	12,425	63,75	0,483	0,292	1,145
7	FeB4 (orthorhombique)	Pnnm	517	272	219	4,57	33,311	72,06	0,805	0,183	0,000
8	Fe2B7(orthorhombique)	Pbam	523	279	220	4,81	32,537	517,11	0,789	0,188	0,194
9	FeB(tétraogonal)	$I4_1/amd$	469	300	189	6,89	22,003	64,28	0,630	0,240	1,373
10	Fe 23 B 6(cubique)	$Fm\bar{3}m$	318	240	124	7,51	12,546	298,40	0,517	0,279	48,593
11	Fe 2 B (tétraogonal)	$I4/mcm$	358	227	145	7,51	18,702	54,18	0,638	0,237	3,700
12	FeN (tétraogonal)	$P4_2/mmc$	109	105	41	4,12	2,829	56,35	0,390	0,327	1,694
13	FeN (cubique)	$Pm\bar{3}m$	234	225	88	6,96	6,162	16,66	0,391	0,327	2,552
14	FeN(cubique)	$Fm\bar{3}m$	232	195	89	6,83	8,016	16,98	0,456	0,302	2,414
15	Fe4N(triclinique)	$R\bar{3}m$	264	243	100	8,11	7,490	48,58	0,412	0,319	0,422
16	FeN (hexagonal)	$P6_3/mmc$	418	276	167	7,14	19,219	32,50	0,605	0,248	1,738
17	Fe2N(triclinique)	$P\bar{3}m1$	253	230	96	7,19	7,407	29,03	0,418	0,317	3,977
18	FeN2 (orthorhombique)	Pnnm	282	230	109	6,03	10,031	92,37	0,475	0,295	0,008
19	Fe2N (triclinique)	$P\bar{3}1m$	238	244	89	7,41	5,460	84,50	0,364	0,338	3,001
20	Fe 4 N (cubique)	$Pm\bar{3}m$	156	193	57	7,95	2,107	49,58	0,295	0,366	0,027
21	FeN (cubique)	$F\bar{4}3m$	169	267	60	6,11	0,864	19,00	0,226	0,395	0,000
22	Fe 3 N(hexagonal)	$P6_322$	249	207	96	7,42	8,710	81,21	0,463	0,299	6,169
23	Fe 3 N (triclinique)	P312	242	200	93	7,45	8,639	323,62	0,467	0,298	5,961
24	Fe 8 N (tétraogonal)	$I4/mmm$	228	182	88	7,57	8,774	101,09	0,484	0,292	19,431
25	Fe 8 N 3 (monoclinique)	C2	247	213	94	7,44	8,031	327,27	0,443	0,307	14,878
26	Fe 12 N 5(triclinique)	$P\bar{3}1m$	268	238	102	7,44	8,094	330,45	0,428	0,313	20,951
27	Fe 24 N 11(monoclinique)	C2/c	270	241	103	7,44	8,103	667,48	0,427	0,313	38,126
28	Fe 2 N (orthorhombique)	$P2_12_12$	274	244	104	7,43	8,199	168,59	0,427	0,313	2,957
29	Fe 2 N (orthorhombique)	Pbcm	238	244	89	7,4	5,495	112,84	0,365	0,337	2,986
30	Fe 23 C 6(cubique)	$Fm\bar{3}m$	334	260	130	7,91	12,267	284,93	0,499	0,286	45,687
31	Fe 3 C (orthorhombique)	Pnma	216	223	81	7,83	4,950	152,24	0,362	0,339	5,597
32	Fe 3 C (hexagonal)	$P6_322$	235	198	90	7,33	8,169	81,31	0,457	0,302	6,001
33	Fe 5 C 2(monoclinique)	C2/2	240	198	90	7,79	8,120	129,33	0,456	0,302	8,320

34	Fe 5 C 2(monoclinique)	P2/c	369	278	144	7,76	14,025	259,57	0,519	0,279	8,475
35	Fe 2 C (orthorhombique)	Pnnm	263	221	101	7,21	8,912	56,98	0,457	0,302	3,245
36	Fe 7 C 3(orthorhombique)	Pnma	379	283	148	7,78	14,460	364,34	0,523	0,277	11,959
37	Fe 7 C 3 (hexagonal)	P6 ₃ mc	244	241	92	7,82	6,049	181,36	0,380	0,332	11,540
38	Fe 3 C (orthorhombique)	Pnma	185	180	69	7,21	4,825	165,51	0,385	0,329	5,674
39	Fe 4 C (monoclinique)	P2 ₁ /m	322	254	125	7,88	11,737	99,25	0,493	0,288	7,513
40	Fe 3 C (tétraogonal)	I4 ₂ m	352	274	137	7,99	12,857	149,27	0,501	0,285	5,598
41	Fe 3 C (monoclinique)	C2/m	343	266	134	7,83	12,666	152,38	0,502	0,285	4,779
42	Fe 4 C (tétraogonal)	I4/m	344	278	133	8,53	11,740	45,82	0,478	0,294	0,001
43	Fe 2 C(hexagonal)	P6/mmm	245	203	94	6,56	8,667	62,61	0,465	0,299	2,642
44	Fe 6 BC(orthorhombique)	Pmc2 ₁	347	262	136	7,7	13,374	154,44	0,518	0,279	11,616
45	Fe 46 (BC 3) 3(triclinique)	P1	326	252	127	7,74	12,267	581,21	0,504	0,284	94,626
46	Fe 23 (BC) 3(triclinique)	R3m	325	250	127	7,69	12,388	292,04	0,508	0,283	47,195
47	Fe 2 (CN 2) 3(triclinique)	R3c	230	153	92	3,46	12,634	222,14	0,604	0,249	2,000
48	FeCN 2 (hexagonal)	P6 ₃ /mmc	222	154	88	3,78	11,342	84,13	0,573	0,259	3,999
49	Fe(C 2 N3)3(orthorhom.)	Cmcm	99	59	40	1,45	8,097	581,95	0,681	0,222	5,000
50	Fe 4 CN 4(orthorhom.)	Pnma	290	234	112	6,32	10,374	306,46	0,479	0,293	5,113
51	Fe 2 (CN 2) 3 (triclinique)	R32	215	144	86	3,4	11,659	113,07	0,593	0,252	0,017
52	FeC 2 N 3 (monoclinique)	C2/c	204	133	82	3,07	11,958	263,80	0,616	0,244	4,000
53	Fe65(BN)6 (monoclin.)	Pm	244	197	94	6,95	9,078	903,40	0,478	0,294	150,88

I.4 Outil d'analyse

Le domaine de l'informatique il existe plusieurs logiciels pour utiliser et représenter les statistique de l'analyses de composantes principales, par exemple : S, SPSS, statistica, XLstat, R ,... etc.

Dans ce travail, nous avons opté pour le logiciel R pour analyser à travers l'ACP le jeu de données récolté.

R est un langage de programmation et un logiciel libre d'analyse statistique et graphique créé en 1993 par Ross Ihaka et Robert Gentleman à l'université d'Auckland (Nouvelle-Zélande). Développé par un groupe de développeurs depuis 1997, il voit sa première version officielle comme langage de programmation qu'en 2000. En pleine expansion, il est plus utilisé tant par le monde de l'entreprise, que par le monde de l'enseignement supérieur et de la recherche scientifique.

Le logiciel R est dédié à des méthodes d'analyses factorielles classiques (ACP, AFC ou ACM) et plus avancées (Analyse factorielle Multiple ou AFM Hiérarchique)... . Dans chaque méthode, il est possible d'ajouter des éléments supplémentaires: individus, variables quantitatives et/ou qualitatives. Les représentations graphiques sont au centre de chacune des analyses et de nombreuses options graphiques y sont disponibles [9, 10]. La spécificité de l'analyse des données est présente dans ce logiciel grâce à plusieurs bibliothèques de fonctions.

- **Exemple d'un script du programme R**

Comme exemple de l'interface de R, la figure 8 montre la façon de décrire les lignes des commandes d'insertion du fichier dans le programme et de lancer le calcul de clustering. Les résultats obtenus numériquement et graphiquement sont présentés dans la figure 9.

Tableau 2. Valeurs propres obtenues par R.

	Valeur propre	Variabilité (%)	Variab. cumulative
Dim1	4,4353	49,28	49,3
Dim2	2,7642	30,71	80,0
Dim3	1,1566	12,85	92,8
Dim4	0,443	4,92	97,8
Dim5	0,161	1,78	99,6
Dim6	0,034	0,04	99,9
Dim7	0,000	0,00	100,0
Dim8	0,000	0,00	100,0
Dim9	0,000	0,00	100,0

I.6 Scree plot

Le scree plot est un histogramme des valeurs propres ou des facteurs ou des composants principaux dans une analyse. Le graphique de la figure 10 représente la qualité de projection. Cette qualité de la projection est mesurée avec le pourcentage d'inertie dit aussi variabilité selon chaque axe ou par un plan. Le diagramme en barre des pourcentages d'inertie (variabilité) de chaque dimension de la figure 10 montre que le pourcentage de la première dimension, notée F1, est de l'ordre de 49.3 % et la 2ème dimension, notée F2, est de l'ordre de 30.7%, soit une variabilité cumulative de 80 % des informations qui sont contenu dans le jeu de données. Cela signifie que si l'on résume les 9 variables initiales de jeu données par les deux dimensions principales F1 et F2, nous récupérons 80% de l'information contenue dans tout le tableau 1. Autrement dit, nous avons un excellent résumé qui synthétise presque totalement les 9 variables. La perte d'information de 20% est répartie sur les 7 autres dimensions avec des valeurs propres relativement faibles (tableau 2), ce qui signifie qu'il y a très peu d'informations perdue et il n'est pas nécessaire d'aller visualiser ces dimensions.

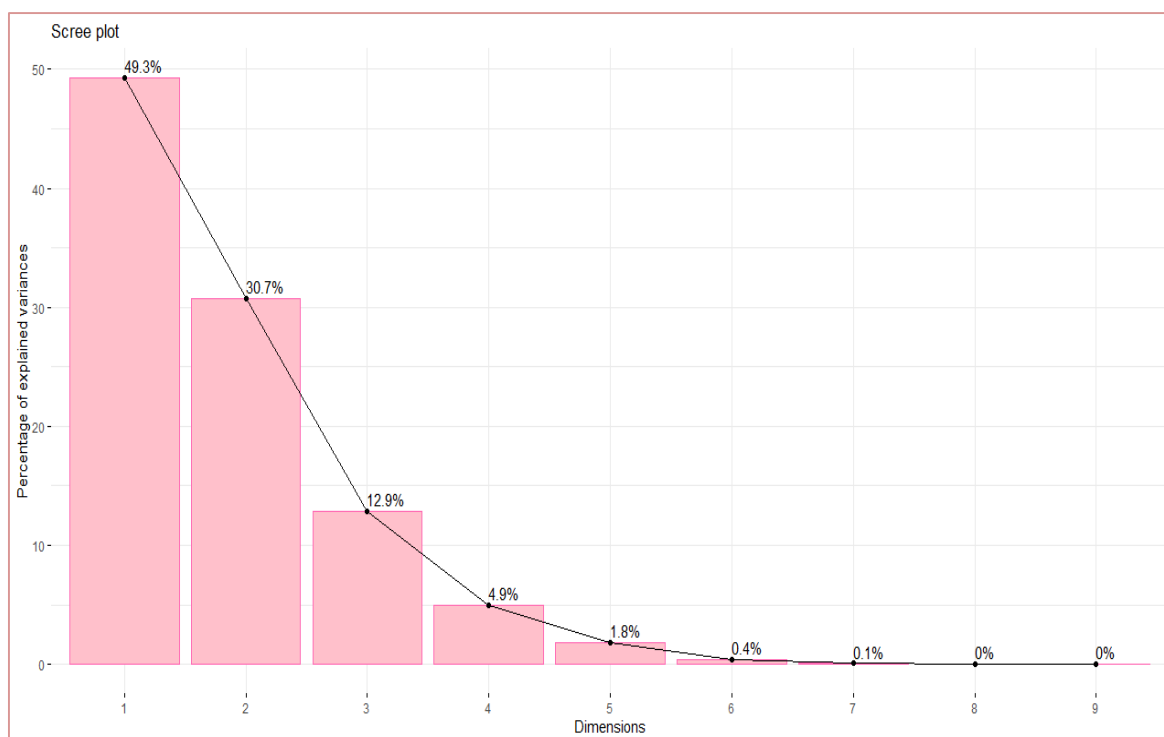


Figure 10. Le Scree plot de pourcentages des variabilités par R.

I.7 Vecteurs propres

Les valeurs propres sont les coordonnées artificielles sur les axes F1,F2,... des variables synthétisée par l'ACP. Le tableau 3 présente les vecteurs propres des 9 variables choisies.

Tableau 3. Vecteurs propres des variables.

Variables	F1	F2	F3	F4	F5	F6	F7	F8	F9
E	-0,4624	0,1149	0,0484	-0,1441	0,1054	-0,0071	-0,5402	0,1667	0,6474
B	-0,3200	0,4133	0,0669	-0,2793	0,3840	-0,5116	0,4782	0,0591	-0,0533
G	-0,4679	0,0755	0,0372	-0,1273	0,1264	0,1683	-0,4008	-0,3945	-0,6303
ρ	-0,1055	0,5376	0,1892	-0,1535	-0,7854	0,1144	0,1019	-0,0052	-0,0036
HV	-0,4587	-0,1236	-0,0333	0,1047	0,1064	0,6751	0,4743	0,2623	0,0385
V	0,1488	-0,3921	0,4094	-0,7981	-0,0090	0,1249	0,0525	0,0322	0,0059
K	-0,3332	-0,4209	-0,0113	0,0991	-0,2614	-0,2346	0,2720	-0,6470	0,2932
v	0,3320	0,4180	-0,0016	-0,0989	0,3333	0,4120	0,0809	-0,5697	0,3055
μ_B	0,0019	0,0164	0,8872	0,4403	0,1329	-0,0270	-0,0144	-0,0102	-0,0012

I.8 Coordonnées des variables

Ce sont les coordonnées dans le nouvel espace qui exprime les corrélations entre les variables d'origine selon les composantes dans le nouvel espace. Le tableau 4 représente les coordonnées des variables de notre étude.

Tableau 4. Coordonnées des variables.

	F1	F2	F3	F4	F5	F6	F7	F8	F9
E	-0,9737	0,1910	0,0520	-0,0960	0,0423	-0,0013	-0,0405	1,9430	3,1090
B	-0,6739	0,6872	0,0720	-0,1859	0,1540	-0,0943	0,0358	6,8936	-2,5605
G	-0,9854	0,1255	0,0400	-0,0848	0,0507	0,0310	-0,0300	-4,5985	-3,0268
ρ	-0,2222	0,8939	0,2035	-0,1022	-0,3150	0,0211	0,0076	-6,0233	-1,7340
HV	-0,9659	-0,2055	-0,0358	0,0697	0,0427	0,1244	0,0355	3,0579	1,8470
V	0,3134	-0,6520	0,4403	-0,5313	-0,0036	0,0230	0,0039	3,7506	2,8469
K	-0,7017	-0,6998	-0,0121	0,0660	-0,1048	-0,0432	0,0204	-7,5417	1,4078
v	0,6993	0,6950	-0,0017	-0,0659	0,1337	0,0759	0,0061	-6,6409	1,4671
μ_B	0,0041	0,0273	0,9542	0,2931	0,0533	-0,0050	-0,0011	-1,1860	-5,5373

I.9 Cercle de corrélation

Les coordonnées des variables peuvent être représentées dans un cercle de rayon 1 et dont les axes sont les deux axes principaux F1 et F2. Ce cercle est dit 'de cercle de corrélation' (voir §II.1.2). La figure 11 représente le cercle de corrélation de notre cas.

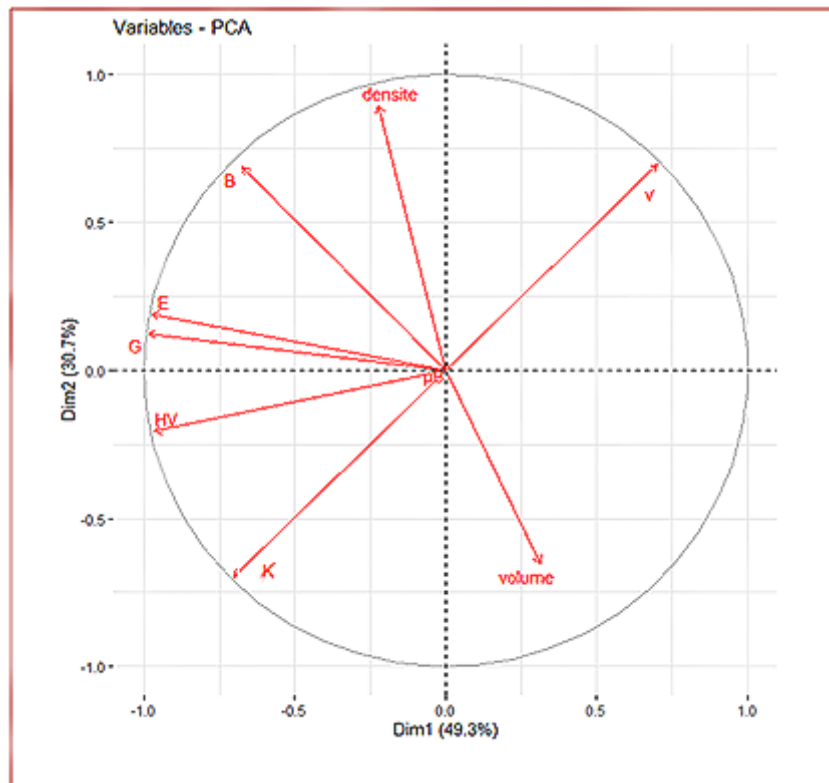


Figure 11. Cercle de corrélation par R

A partir du cercle de corrélation de la figure 11, nous observons :

- une forte corrélation positive entre le module de cisaillement G et le module de Young E.
- une corrélation positive entre le module de compressibilité isostatique B et les deux autres modules E et G.
- une forte corrélation positive entre le module de Young E, le module de cisaillement G et la dureté de Vickers HV,
- une corrélation positive entre la dureté de Vickers HV et le rapport K,
- une corrélation négative entre la densité ρ et le volume V.
- une corrélation négative entre le coefficient de Poisson ν et le rapport K.
- une corrélation négative entre volume V et les modules d'élasticité E et G.
- Aucune corrélation ou presque nulle entre :
 - le coefficient de Poisson ν et le module de Young E
 - le coefficient de Poisson ν et le module de compressibilité isostatique B
 - le coefficient de Poisson ν et le volume V.
 - le module de compressibilité isostatique B et le rapport K
 - la densité ρ et le rapport K

- la densité ρ et la dureté HV
- le volume V et la dureté HV.

Le vecteur de moment magnétique n'est pas bien représenté selon les axes principaux F1 et F2. Cela veut dire que les informations nécessaires et suffisantes de cette propriété ne sont pas portées sur ces deux axes et par suite nous ne pouvons faire aucune interprétation dans ces dimensions.

I.10 Contributions des variables

Les contributions (tableau 5) sont une aide à l'interprétation. Les variables ayant le plus influencé la construction des axes sont celles dont les contributions sont les plus élevées.

Tableau 5. Contributions des variables (%).

	F1	F2	F3	F4	F5	F6	F7	F8	F9
E	21,3777	1,3194	2,3422	2,0779	1,1105	0,0050	29,1848	2,7783	4,1912
B	10,2396	17,0820	4,4819	7,8010	14,7486	26,1743	22,8722	0,3497	2,8428
G	21,8923	0,5701	1,3803	1,6211	1,5974	2,8338	16,0613	15,5620	3,9724
ρ	1,1132	28,9066	3,5808	2,3577	61,6912	1,3081	1,0385	0,0027	1,3037
HV	21,0367	1,5281	1,1082	1,0959	1,1330	45,5723	22,4938	6,8815	1,4792
V	2,2141	15,3767	1,6762	63,6954	0,0081	1,5601	0,2761	0,1035	3,5144
K	11,1021	17,7172	1,2757	0,9828	6,8344	5,5015	7,3980	41,8570	8,5942
v	11,0240	17,4729	2,4242	0,9786	11,1107	16,9717	0,6545	32,4549	9,3325
μ_B	0,0004	0,0269	78,713	19,3896	1,7663	0,0730	0,0208	0,0104	1,3295

I.11 Cosinus carrés des variables

Un cosinus carré (noté \cos^2) (§ II.1.2, équation 4) représente la qualité de représentation des variables sur la carte de l'ACP. L'analyse des cosinus carrés des variables permet d'éviter des erreurs pendant les calculs des coefficients de corrélation et l'interprétation des variables dues à des effets de projection.

Un \cos^2 élevé indique une bonne représentation de la variable sur les axes principaux en considération. Dans ce cas, la variable est positionnée à proximité de la circonférence du cercle de corrélation.

Un faible \cos^2 indique que la variable n'est pas parfaitement représentée par les axes principaux. Dans ce cas, la variable est proche du centre du cercle (cas de μ_B).

Le tableau 6 présente les valeurs des cosinus carrés des variables de notre cas.

Tableau 6. Valeurs des cosinus carrés des variables.

Variables	F1	F2	F3	F4	F5	F6	F7	F8	F9
E	0,95	0,04	0,00	0,01	0,00	0,00	0,00	0,00	0,00
B	0,45	0,47	0,01	0,03	0,02	0,01	0,00	0,00	0,00
G	0,97	0,02	0,00	0,01	0,00	0,00	0,00	0,00	0,00
ρ	0,05	0,80	0,04	0,01	0,10	0,00	0,00	0,00	0,00
HV	0,93	0,04	0,00	0,00	0,00	0,02	0,00	0,00	0,00
V	0,10	0,43	0,19	0,28	0,00	0,00	0,00	0,00	0,00
K	0,49	0,49	0,00	0,00	0,01	0,00	0,00	0,00	0,00
v	0,49	0,48	0,00	0,00	0,02	0,01	0,00	0,00	0,00
μ_B	0,00	0,00	0,91	0,09	0,00	0,00	0,00	0,00	0,00

I.12 Coordonnées des observations

La représentation des variables diffère de celle des observations. Les observations sont représentées par leurs projections, alors que les variables sont représentées par leurs corrélations. Les coordonnées des observations sont les coordonnées artificielles des composés dans le nouvel espace. Ils sont représentés ensuite graphiquement par un nuage de points. Ils servent ensuite à faire du clustering. Le tableau 7 résume ces coordonnées pour les 54 composés de fer considérés.

Tableau 7. Coordonnées des observations.

N°	Individus	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	Fe3B (tet)	-1,3539	0,3597	-0,1180	-0,2686	-0,2368	-0,0637	-0,0109	-0,0003	0,0002
2	FeB2 (hex)	-0,6574	0,3461	-0,5898	-0,0758	0,7011	-0,2669	-0,1255	-0,0021	0,0010
3	Fe2B (tet)	4,1953	1,7104	-0,5632	0,2123	0,5275	0,5910	-0,0180	-0,0283	0,0118
4	Fe3B (ort)	-1,3369	0,3272	-0,1175	-0,2510	-0,2349	-0,0614	-0,0109	0,0000	0,0002
5	FeB(ort)	-2,5876	-0,6651	-0,4845	0,1617	-0,4134	0,0507	0,0431	-0,0051	-0,0009
6	FeB (orth)	-1,3562	0,8947	-0,3621	-0,4687	0,5288	-0,2995	-0,0928	-0,0091	0,0037
7	FeB4 (ort)	-6,0740	-2,3903	-0,6664	0,2184	0,5922	0,4641	0,0838	-0,0103	-0,0031
8	Fe2B7(ort)	-5,8832	-2,4832	-0,2391	-0,6208	0,6037	0,5400	0,0793	0,0161	0,0010
9	FeB(tet)	-4,0064	-0,1120	-0,3154	-0,4004	0,2756	-0,0118	-0,1251	-0,0228	-0,0002
10	Fe 23 B 6(cub)	-0,8352	0,1974	1,4582	0,3502	-0,1039	-0,0453	-0,0058	-0,0046	-0,0011
11	Fe 2 B (tét)	-2,4077	-0,6430	-0,3395	0,2212	-0,8365	0,1133	0,1194	-0,0043	-0,0022
12	FeN (tet)	3,4834	-0,9247	-1,0391	1,0878	0,3161	0,3281	-0,2796	-0,0069	-0,0051
13	FeN (cub)	1,2297	1,0326	-0,5012	0,0445	0,1804	-0,0464	-0,0245	-0,0020	-0,0020
14	FeN(cub)	0,8808	0,2423	-0,5707	0,3479	-0,2847	0,0312	-0,0385	0,0022	-0,0019
15	Fe4N(tri)	0,5652	1,3616	-0,3850	-0,2562	-0,2575	-0,0557	0,0420	-0,0028	-0,0029
16	FeN (hex)	-3,0621	-0,0499	-0,3590	-0,1905	-0,0394	-0,0344	-0,0618	-0,0153	-0,0002
17	Fe2N(tri)	0,7623	0,9321	-0,4029	0,0240	0,0305	-0,0660	-0,0203	-0,0018	-0,0023
18	FeN2 (orth)	0,0274	0,0896	-0,6072	0,0150	0,3411	-0,1291	-0,0856	0,0079	0,0008
19	Fe2N (tri)	1,3115	1,4812	-0,3488	-0,2587	0,2458	-0,0739	0,0416	0,0000	-0,0012
20	Fe4N (cub)	2,9813	1,7598	-0,5297	-0,0136	-0,2121	0,3565	0,0569	-0,0128	-0,0004
21	FeN(cub)	3,0994	2,3521	-0,6491	-0,3446	1,5250	-0,1181	0,2603	0,0000	-0,0047
22	Fe3N(hex)	0,5507	0,4228	-0,2967	0,1773	-0,4455	0,0371	0,0022	0,0023	-0,0047
23	Fe3N(tri)	0,7076	0,1432	-0,1098	-0,1698	-0,5500	0,1261	0,0352	0,0187	-0,0025
24	Fe8N(tet)	0,7376	0,0858	0,1531	0,5738	-0,7665	0,1104	0,0302	-0,0013	-0,0081
25	Fe8N3 (mono)	0,7859	0,4378	0,2286	-0,1477	-0,2835	0,0771	0,0188	0,0143	-0,0013
26	Fe12N5(tri)	0,5434	0,7708	0,4918	-0,2506	0,0443	-0,0297	0,0136	0,0102	0,0006
27	Fe24N11(mono)	0,6188	0,5462	1,3839	-0,5312	0,1618	0,0263	0,0408	0,0254	0,0042
28	Fe2N (ort)	0,4047	0,9493	-0,2647	-0,3414	0,0197	-0,0792	0,0056	0,0061	0,0000
29	Fe2N (orth)	1,3068	1,4507	-0,3260	-0,3053	0,2485	-0,0687	0,0434	0,0020	-0,0008
30	Fe23C6(cub)	-0,9807	0,6469	1,4278	0,1122	-0,0398	-0,0976	-0,0086	-0,0088	-0,0002
31	Fe3C(ort)	1,6684	1,3931	-0,1944	-0,2155	-0,1183	0,0826	0,0552	0,0019	-0,0020
32	Fe3C(hex)	0,8035	0,3560	-0,3312	0,2442	-0,4764	0,0671	-0,0015	0,0028	-0,0053
33	Fe5C2(mono)	0,7829	0,4760	-0,1567	0,1559	-0,6646	0,1077	0,0021	0,0119	0,0265

34	Fe5C2(mono)	-1,6638	0,6069	0,1275	-0,5992	-0,0429	-0,1079	-0,0253	-0,0003	0,0028
35	Fe2C (ort)	0,3685	0,5411	-0,4149	0,0726	-0,2060	-0,0308	-0,0186	0,0015	-0,0030
36	Fe7C3(ort)	-1,8135	0,5525	0,3531	-0,7543	0,0119	-0,0948	-0,0242	0,0034	0,0042
37	Fe7C3(hex)	1,1296	1,3981	0,0755	-0,2744	0,0252	-0,0290	0,0553	0,0022	-0,0013
38	Fe3C(ort)	2,0998	0,6409	-0,3264	0,1532	-0,3270	0,2321	-0,0221	0,0040	-0,0035
39	Fe4C(mono)	-0,8223	0,7503	-0,0835	-0,1992	-0,2695	-0,0902	-0,0013	-0,0042	-0,0012
40	Fe3C(tet)	-1,3383	0,8619	-0,0532	-0,4723	-0,1511	-0,1284	-0,0127	-0,0051	0,0009
41	Fe3C (mono)	-1,1948	0,7305	-0,1126	-0,4163	-0,1650	-0,1100	-0,0121	-0,0024	0,0007
42	Fe4C(tet)	-1,1270	1,3241	-0,2719	-0,5054	-0,3072	-0,1562	0,0149	-0,0112	-0,0008
43	Fe2C(hex)	0,6419	0,1231	-0,5372	0,2525	-0,1155	-0,0083	-0,0462	0,0052	-0,0033
44	Fe6BC(orth)	-1,3460	0,5427	0,1083	-0,2426	-0,1570	-0,0952	-0,0189	-0,0043	0,0003
45	Fe46(BC3)3(tricl)	-0,7971	0,2770	3,3579	0,5558	0,1925	-0,0540	-0,0052	-0,0068	0,0009
46	Fe23(BC)3(tri)	-0,8950	0,4112	1,4419	0,2376	-0,0678	-0,0695	-0,0090	-0,0064	-0,0006
47	Fe2(CN2)3(tri)	0,1594	-2,4699	-0,8657	0,8402	0,2548	-0,1019	0,0037	0,0213	0,0023
48	FeCN2 (hex)	0,4414	-2,0123	-0,8737	1,0159	0,2502	-0,1109	-0,0492	0,0123	-0,0017
49	Fe(C2N3)3(orth)	2,0292	-4,7831	-0,8826	1,3516	-0,2257	-0,4135	0,1865	-0,0372	0,0059
50	Fe4CN4(ort)	-0,0666	0,0161	-0,2064	-0,3029	0,2610	-0,0742	-0,0510	0,0186	0,0029
51	Fe2(CN2)3 (tri)	0,4583	-2,3924	-1,0513	1,0398	0,2218	-0,1259	-0,0187	0,0139	-0,0005
52	FeC2N3 (mono)	0,5235	-2,8923	-0,8459	1,0174	0,1750	-0,1243	0,0374	0,0180	0,0022
53	Fe65(BN)6 (mono)	0,8111	-0,4950	5,3979	1,4682	0,3635	0,0908	-0,0068	0,0009	-0,0013

I.13 Contributions des observations

Les contributions des observations diffèrent des contributions des variables (§IV.1.16). Elles concernent les contributions des individus observés dans le nuage des points. Elles constituent également une aide à l'interprétation. Les coordonnées des observations ayant le plus influencé à la construction des axes sont celles dont les contributions sont les plus élevées (tableau 8).

Tableau 8. Valeurs des contributions des observations (%).

N	Individus	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	Fe3B (tet)	0,7654	0,0867	0,0223	0,3015	0,6456	0,2216	0,0389	0,0015	0,0029
2	FeB2 (hex)	0,1805	0,0802	0,5569	0,0240	5,6600	3,8827	5,1993	0,0582	0,0850
3	Fe2B (tet)	7,3487	1,9599	0,5079	0,1884	3,2046	19,0449	0,1067	10,9421	11,1320
4	Fe3B (ort)	0,7462	0,0717	0,0221	0,2632	0,6351	0,2056	0,0394	0,0000	0,0021
5	FeB(ort)	2,7956	0,2964	0,3759	0,1093	1,9680	0,1402	0,6129	0,3570	0,0722
6	FeB (orth)	0,7679	0,5363	0,2099	0,9178	3,2204	4,8918	2,8455	1,1347	1,0762
7	FeB4 (ort)	15,4037	3,8277	0,7110	0,1994	4,0385	11,7449	2,3164	1,4509	0,7890
8	Fe2B7(ort)	14,4514	4,1309	0,0915	1,6103	4,1960	15,8947	2,0741	3,5512	0,0839
9	FeB(tet)	6,7019	0,0084	0,1593	0,6700	0,8748	0,0076	5,1638	7,0562	0,0037
10	Fe 23 B 6(cub)	0,2912	0,0261	3,4045	0,5124	0,1243	0,1118	0,0113	0,2835	0,0898
11	Fe 2 B (tét)	2,4205	0,2770	0,1845	0,2045	8,0568	0,6997	4,7053	0,2499	0,3923
12	FeN (tet)	5,0663	0,5729	1,7285	4,9439	1,1506	5,8672	25,8137	0,6534	2,1053
13	FeN (cub)	0,6313	0,7143	0,4022	0,0083	0,3747	0,1176	0,1985	0,0527	0,3223
14	FeN(cub)	0,3239	0,0393	0,5215	0,5057	0,9331	0,0531	0,4898	0,0689	0,2956
15	Fe4N(tri)	0,1334	1,2420	0,2373	0,2743	0,7637	0,1693	0,5828	0,1057	0,6722
16	FeN (hex)	3,9149	0,0017	0,2063	0,1516	0,0178	0,0646	1,2612	3,1863	0,0036
17	Fe2N(tri)	0,2426	0,5821	0,2599	0,0024	0,0107	0,2378	0,1363	0,0455	0,4403
18	FeN2 (orth)	0,0003	0,0054	0,5903	0,0009	1,3397	0,9081	2,4219	0,8448	0,0487
19	Fe2N (tri)	0,7182	1,4697	0,1948	0,2796	0,6955	0,2975	0,5715	0,0000	0,1178
20	Fe4N (cub)	3,7111	2,0746	0,4493	0,0008	0,5178	6,9281	1,0689	2,2369	0,0103
21	FeN(cub)	4,0108	3,7062	0,6745	0,4963	26,7809	0,7599	22,3741	0,0000	1,7466
22	Fe3N(hex)	0,1266	0,1198	0,1410	0,1314	2,2850	0,0752	0,0016	0,0698	1,7858
23	Fe3N(tri)	0,2091	0,0137	0,0193	0,1204	3,4838	0,8669	0,4085	4,7864	0,5149
24	Fe8N(tet)	0,2272	0,0049	0,0375	1,3756	6,7648	0,6646	0,3020	0,0225	5,3011
25	Fe8N3 (mono)	0,2579	0,1284	0,0837	0,0911	0,9256	0,3237	0,1172	2,8015	0,1378
26	Fe12N5(tri)	0,1233	0,3981	0,3872	0,2625	0,0226	0,0480	0,0614	1,4147	0,0329
27	Fe24N11(mono)	0,1599	0,1999	3,0663	1,1792	0,3013	0,0378	0,5490	8,8194	1,4283
28	Fe2N (ort)	0,0684	0,6037	0,1122	0,4870	0,0045	0,3417	0,0104	0,5148	0,0001
29	Fe2N (orth)	0,7130	1,4098	0,1701	0,3893	0,7109	0,2573	0,6220	0,0540	0,0480
30	Fe23C6(cub)	0,4016	0,2803	3,2638	0,0526	0,0183	0,5189	0,0246	1,0464	0,0027
31	Fe3C(ort)	1,1621	1,3002	0,0605	0,1940	0,1610	0,3722	1,0072	0,0487	0,3205
32	Fe3C(hex)	0,2695	0,0849	0,1756	0,2491	2,6138	0,2452	0,0007	0,1049	2,2959
33	Fe5C2(mono)	0,2559	0,1518	0,0393	0,1015	5,0868	0,6325	0,0015	1,9441	56,3561

34	Fe5C2(mono)	1,1558	0,2468	0,0260	1,5002	0,0212	0,6349	0,2120	0,0011	0,6236
35	Fe2C (ort)	0,0567	0,1962	0,2756	0,0220	0,4888	0,0519	0,1141	0,0299	0,7417
36	Fe7C3(ort)	1,3732	0,2045	0,1996	2,3771	0,0016	0,4904	0,1932	0,1584	1,4384
37	Fe7C3(hex)	0,5327	1,3096	0,0091	0,3147	0,0073	0,0459	1,0092	0,0646	0,1305
38	Fe3C(ort)	1,8409	0,2752	0,1705	0,0981	1,2309	2,9361	0,1610	0,2185	0,9939
39	Fe4C(mono)	0,2823	0,3771	0,0112	0,1658	0,8362	0,4440	0,0006	0,2434	0,1125
40	Fe3C(tet)	0,7478	0,4977	0,0045	0,9321	0,2628	0,8988	0,0531	0,3544	0,0661
41	Fe3C(mono)	0,5960	0,3575	0,0203	0,7242	0,3136	0,6592	0,0480	0,0756	0,0394
42	Fe4C(tet)	0,5303	1,1745	0,1183	1,0672	1,0864	1,3296	0,0732	1,7107	0,0513
43	Fe2C(hex)	0,1720	0,0101	0,4620	0,2664	0,1535	0,0038	0,7049	0,3698	0,9002
44	Fe6BC(orth)	0,7564	0,1973	0,0188	0,2459	0,2840	0,4946	0,1179	0,2547	0,0062
45	Fe46(BC3)3(tricl)	0,2653	0,0514	18,0530	1,2908	0,4267	0,1589	0,0089	0,6289	0,0643
46	Fe23(BC)3(tri)	0,3344	0,1133	3,3286	0,2359	0,0530	0,2630	0,0266	0,5632	0,0311
47	Fe2(CN2)3(tri)	0,0106	4,0868	1,2000	2,9494	0,7477	0,5660	0,0044	6,1750	0,4073
48	FeCN2 (hex)	0,0813	2,7129	1,2220	4,3124	0,7210	0,6711	0,7997	2,0459	0,2277
49	Fe(C2N3)3(orth)	1,7193	15,3269	1,2471	7,6326	0,5866	9,3216	11,4871	18,8851	2,8308
50	Fe4CN4(ort)	0,0019	0,0002	0,0682	0,3833	0,7847	0,2999	0,8575	4,7389	0,6963
51	Fe2(CN2)3 (tri)	0,0877	3,8343	1,7697	4,5173	0,5665	0,8645	0,1153	2,6344	0,0181
52	FeC2N3 (mono)	0,1144	5,6044	1,1457	4,3250	0,3528	0,8427	0,4625	4,4346	0,3941
53	Fe65(BN)6 (mono)	0,2747	0,1641	46,6494	9,0071	1,5213	0,4495	0,0152	0,0105	0,1364

I.14 Cosinus carrés des observations

La qualité de la représentation des points (individus) sur la carte ACP est désigné par le facteur cosinus carré. L'analyse des cosinus carrés dans le tableau 9 permet d'éviter des erreurs d'interprétation dues à des effets de projection (équation (6)). Dans le logiciel R la qualité graphique de la représentation des individus est faite par un *barcolor* des valeurs du cosinus carré (figure 12).

Tableau 9. Valeurs de cosinus carrés des observations.

Individus	F1	F2	F3	F4	F5	F6	F7	F8	F9
1 Fe3B (tet)	0,8693	0,0613	0,0066	0,0342	0,0266	0,0019	0,0001	0,0000	0,0000
2 FeB2 (hex)	0,2912	0,0807	0,2344	0,0039	0,3312	0,0480	0,0106	0,0000	0,0000
3 Fe2B (tet)	0,8180	0,1360	0,0147	0,0021	0,0129	0,0162	0,0000	0,0000	0,0000

4	Fe3B (ort)	0,8804	0,0527	0,0068	0,0310	0,0272	0,0019	0,0001	0,0000	0,0000
5	FeB(ort)	0,8840	0,0584	0,0310	0,0035	0,0226	0,0003	0,0002	0,0000	0,0000
6	FeB (orth)	0,5460	0,2376	0,0389	0,0652	0,0830	0,0266	0,0026	0,0000	0,0000
7	FeB4 (ort)	0,8448	0,1308	0,0102	0,0011	0,0080	0,0049	0,0002	0,0000	0,0000
8	Fe2B7(ort)	0,8264	0,1472	0,0014	0,0092	0,0087	0,0070	0,0001	0,0000	0,0000
9	FeB(tet)	0,9778	0,0008	0,0061	0,0098	0,0046	0,0000	0,0010	0,0000	0,0000
10	Fe 23 B 6(cub)	0,2326	0,0130	0,7092	0,0409	0,0036	0,0007	0,0000	0,0000	0,0000
11	Fe 2 B (tét)	0,8163	0,0582	0,0162	0,0069	0,0985	0,0018	0,0020	0,0000	0,0000
12	FeN (tet)	0,7809	0,0550	0,0695	0,0762	0,0064	0,0069	0,0050	0,0000	0,0000
13	FeN (cub)	0,5274	0,3719	0,0876	0,0007	0,0113	0,0008	0,0002	0,0000	0,0000
14	FeN(cub)	0,5685	0,0430	0,2387	0,0887	0,0594	0,0007	0,0011	0,0000	0,0000
15	Fe4N(tri)	0,1299	0,7541	0,0603	0,0267	0,0270	0,0013	0,0007	0,0000	0,0000
16	FeN (hex)	0,9817	0,0003	0,0135	0,0038	0,0002	0,0001	0,0004	0,0000	0,0000
17	Fe2N(tri)	0,3590	0,5368	0,1003	0,0004	0,0006	0,0027	0,0003	0,0000	0,0000
18	FeN2 (orth)	0,0015	0,0155	0,7116	0,0004	0,2246	0,0322	0,0142	0,0001	0,0000
19	Fe2N (tri)	0,4125	0,5261	0,0292	0,0160	0,0145	0,0013	0,0004	0,0000	0,0000
20	Fe4N (cub)	0,7144	0,2489	0,0226	0,0000	0,0036	0,0102	0,0003	0,0000	0,0000
21	FeN(cub)	0,5311	0,3059	0,0233	0,0066	0,1286	0,0008	0,0037	0,0000	0,0000
22	Fe3N(hex)	0,3784	0,2231	0,1099	0,0392	0,2476	0,0017	0,0000	0,0000	0,0000
23	Fe3N(tri)	0,5676	0,0232	0,0137	0,0327	0,3430	0,0180	0,0014	0,0004	0,0000
24	Fe8N(tet)	0,3616	0,0049	0,0156	0,2188	0,3904	0,0081	0,0006	0,0000	0,0000
25	Fe8N3 (mono)	0,6365	0,1976	0,0539	0,0225	0,0828	0,0061	0,0004	0,0002	0,0000
26	Fe12N5(tri)	0,2466	0,4963	0,2020	0,0525	0,0016	0,0007	0,0002	0,0001	0,0000
27	Fe24N11(mono)	0,1317	0,1026	0,6586	0,0971	0,0090	0,0002	0,0006	0,0002	0,0000
28	Fe2N (ort)	0,1302	0,7162	0,0557	0,0926	0,0003	0,0050	0,0000	0,0000	0,0000
29	Fe2N (orth)	0,4186	0,5158	0,0260	0,0228	0,0151	0,0012	0,0005	0,0000	0,0000
30	Fe23C6(cub)	0,2794	0,1215	0,5921	0,0037	0,0005	0,0028	0,0000	0,0000	0,0000
31	Fe3C(ort)	0,5760	0,4016	0,0078	0,0096	0,0029	0,0014	0,0006	0,0000	0,0000
32	Fe3C(hex)	0,5503	0,1080	0,0935	0,0508	0,1935	0,0038	0,0000	0,0000	0,0000
33	Fe5C2(mono)	0,4565	0,1687	0,0183	0,0181	0,3290	0,0086	0,0000	0,0001	0,0005
34	Fe5C2(mono)	0,7851	0,1045	0,0046	0,1018	0,0005	0,0033	0,0002	0,0000	0,0000
35	Fe2C (ort)	0,2089	0,4507	0,2649	0,0081	0,0653	0,0015	0,0005	0,0000	0,0000
36	Fe7C3(ort)	0,7653	0,0710	0,0290	0,1324	0,0000	0,0021	0,0001	0,0000	0,0000

37	Fe7C3(hex)	0,3847	0,5895	0,0017	0,0227	0,0002	0,0003	0,0009	0,0000	0,0000
38	Fe3C(ort)	0,8627	0,0804	0,0208	0,0046	0,0209	0,0105	0,0001	0,0000	0,0000
39	Fe4C(mono)	0,4948	0,4119	0,0051	0,0290	0,0531	0,0060	0,0000	0,0000	0,0000
40	Fe3C(tet)	0,6398	0,2654	0,0010	0,0797	0,0082	0,0059	0,0001	0,0000	0,0000
41	Fe3C(mono)	0,6529	0,2440	0,0058	0,0793	0,0125	0,0055	0,0001	0,0000	0,0000
42	Fe4C(tet)	0,3658	0,5050	0,0213	0,0736	0,0272	0,0070	0,0001	0,0000	0,0000
43	Fe2C(hex)	0,5182	0,0191	0,3629	0,0802	0,0168	0,0001	0,0027	0,0000	0,0000
44	Fe6BC(orth)	0,8194	0,1332	0,0053	0,0266	0,0112	0,0041	0,0002	0,0000	0,0000
45	Fe46(BC3)3(tricl)	0,0515	0,0062	0,9140	0,0250	0,0030	0,0002	0,0000	0,0000	0,0000
46	Fe23(BC)3(tri)	0,2571	0,0543	0,6674	0,0181	0,0015	0,0015	0,0000	0,0000	0,0000
47	Fe2(CN2)3(tri)	0,0033	0,7967	0,0979	0,0922	0,0085	0,0014	0,0000	0,0001	0,0000
48	FeCN2(hex)	0,0318	0,6620	0,1248	0,1687	0,0102	0,0020	0,0004	0,0000	0,0000
49	Fe(C2N3)3(orth)	0,1379	0,7662	0,0261	0,0612	0,0017	0,0057	0,0012	0,0000	0,0000
50	Fe4CN4(ort)	0,0206	0,0012	0,1976	0,4254	0,3160	0,0255	0,0120	0,0016	0,0000
51	Fe2(CN2)3(tri)	0,0257	0,6992	0,1350	0,1321	0,0060	0,0019	0,0000	0,0000	0,0000
52	FeC2N3(mono)	0,0263	0,8014	0,0686	0,0992	0,0029	0,0015	0,0001	0,0000	0,0000
53	Fe65(BN)6 (mono)	0,0203	0,0076	0,9011	0,0667	0,0041	0,0003	0,0000	0,0000	0,0000

I.15 Nuages des points par R

Etant donné que nous avons choisi deux axes principaux F1 et F2, les coordonnées des observations selon le tableau 7 de ces deux axes orthogonaux, peuvent être représentées sous forme graphique (figure 12). Nous obtenons ainsi un nuage de points qui représentent les individus (les composés de Fe). Ceci constitue l'un des objectifs principaux de l'ACP : le clustering (§ II.2). Rappelons que les axes F1 et F2 présentent des pourcentages d'inertie respectif 49.3% et 30.7% avec un cumul de 80% (tableau 2).

La qualité de représentations des individus (des composés considérés) est faite par des points de différentes couleurs qui correspondent au *barcolor* des cosinus carrés donnés par le logiciel R dans la carte ACP (figure 12). La représentation est meilleure, quand le \cos^2 est élevé (tableau 9 et figure 12).

I.16 Biplot des variables et des individus par R

Le biplot, comme son nom l'indique, présente sur un même graphique les nuages des individus et les flèches contenus dans le cercle de corrélation. Il s'agit donc d'une représentation graphique simultanée des individus (les composés de fer) et des variables (les propriétés physiques considérées) selon les axes principaux F1 et F2 (figure 13). Cela permet d'interpréter les groupes des individus avec leurs propriétés physiques respectifs. Le code de couleur dans le biplot sert à bien visualiser et à bien cerner et l'information pour une interprétation globale lisible.

D'après la figure 13, nous concluons ce qui suit :

- les données en code de couleur combiné au cercle de corrélation, nous permettent de dresser le tableau 10 dédiés aux composés dont les valeurs des propriétés physiques ont une représentativité de qualité moyenne dans l'espace d'axes principaux (représentativité ni trop forte ni trop faible) (voir aussi figure 7, chapitre I).

Tableau 10. Valeurs moyennées dans Biplot

Fe-N	Fe-C	Fe-B	Fe-B-N	Fe-B-C	Fe-C-N
Fe 24 N 11(mono)	Fe 23 C 6(cub)	Fe3B (orth)	Fe65(BN)6 (mono)	Fe 46 (BC 3) 3(tric)	Fe 4 CN 4(orth)
Fe 12 N 5(tric)	Fe 5 C 2(mono)	FeB (orth)		Fe 23 (BC) 3(tric)	
Fe 8 N 3 (mono)	Fe 7 C 3 (hex)	Fe 23 B 6(cub)		Fe 6 BC(orth)	
Fe 3 N (tric)	Fe 3 C (mono)	Fe3B (tét)			
Fe 2 N (orth)	Fe 3 C (tét)	FeB2 (hex)			
Fe 2 N (orth)	Fe 4 C (mono)				
Fe 8 N (tét)	Fe 3 C (hex)				
Fe2N (tric)	Fe 2 C(hex)				
Fe 3 N(hex)	Fe 2 C (orth)				
Fe2N(tric)	Fe 4 C (tét)				
FeN2 (orth)	Fe 7 C 3(orth)				
Fe4N(tric)	Fe 5 C 2(mono)				
FeN(cub)					
FeN (cub)					

Pour les autres composés ils se classent hors la zone des valeurs moyennes (figures 12 et 7).

En s'appuyant sur la figure 7, la figure 13 nous permet de faire les constatations suivantes :

- Les composés **Fe2B(tét)**, **FeB(ort)**, **Fe2B7(ort)** et **FeB4(ort)** sont des composés qui possèdent de fortes valeurs par rapport aux vecteurs de la dureté (HV) , le rapport (K), et de faibles valeurs par rapport au vecteur du coefficient de poisson (ν).
- Les composés **FeN (hex)** et **FeB(tet)** sont des composés qui ont de fortes valeurs par rapport aux vecteurs de module de Young (E), le module de cisaillement (G) et la dureté (HV).

- Les composés **FeCN₂(hex)**, **Fe₂(CN₂)₃ (tri)**, **Fe₂(CN₂)₃ (tri)**, **FeC₂N₃(mono)**, **Fe(C₂N₂)₃ (ort)** sont des composés qui ont de fortes valeurs par rapport au vecteur du volume, et des faibles valeurs par rapport aux vecteur du coefficient de Poisson(ν), de la dureté (HV) et du rapport (K).
- Les composés **FeN(tét)**, **Fe₃C (ort)**) sont des composés qui ont de faibles valeurs par rapport aux vecteurs du module de Young (E), du module de cisaillement (G), la dureté (HV) .
- Les composés **Fe₄N (cub)**, **FeN(cub)** et **Fe₂B(tet)** sont des composés qui ont de fortes valeurs par rapport au vecteur du coefficient de Poisson (ν), et des faibles valeurs par rapport la dureté (HV), le rapport (K).

Il faut rappeler ici que la notion de forte ou faible valeur concerne la représentativité de la variable (gradeur physique) par rapport aux axes principaux.

I.17 Clustering (k-means) par R

Le clustering fait par logiciel R par optimisation, nous donne quatre groupes classés par couleurs (figure 13).

Les composés ainsi définis dans ces groupes sont également classés pour plus de clarté dans le tableau 11. Le premier groupe contient 3 composés de Fe-B ; le 2ème groupe contient 5 composés Fe-CN et un Fe-N ; le 3ème groupe contient 16 composés de Fe-N et 7 composés de Fe-C et un de Fe-B, c'est le plus grand groupe ; le quatrième groupe contient 7 composés de Fe-B , 7 composés Fe-C, 3 composés de Fe-BC, un Fe-C, un Fe-N et un Fe-BN. Par ordre croissant en nombre de composés, nous pouvons faire le classement suivant :

groupe 1 < groupe 2 < groupe 4 < groupe 3

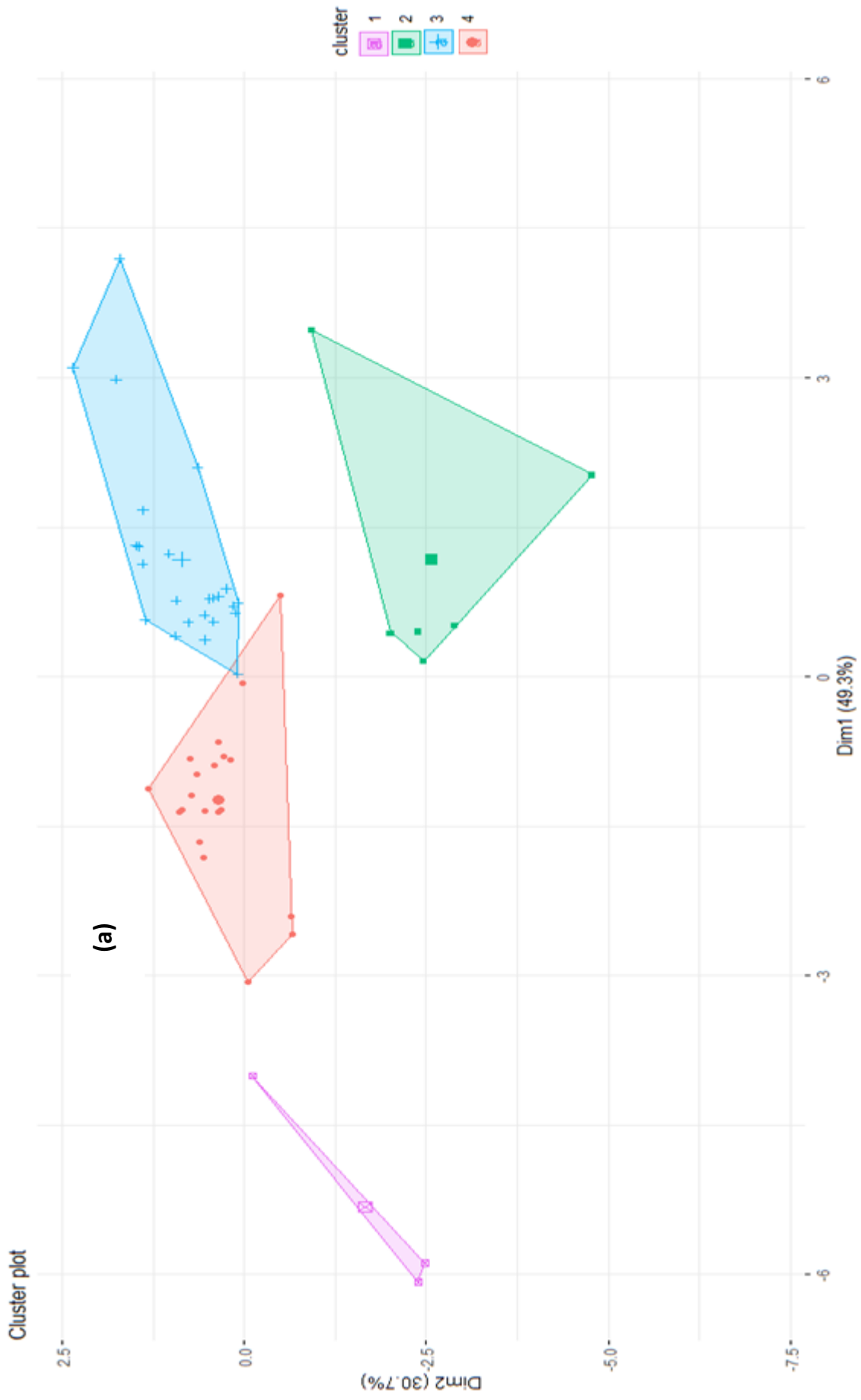


Figure 14 . Biplot de classification (*k-means*) par R.
a) clustering.

les groupes	1	2	3	4
le nombre	3	6	24	20
FeB4 (ort)	FeN (tet)	Fe2B (tet)	Fe3B (tet)	
Fe2B7(ort)	Fe 2 (CN 2) 3(tri)	FeN (cub)	FeB2 (hex)	
FeB(tet)	FeCN 2 (hex)	FeN(cub)	Fe3B (ort)	
	Fe(C 2 N 3) 3(orth)	Fe4N(tri)	FeB (ort)	
	Fe 2 (CN 2) 3 (tri)	Fe2N(tri)	FeB (orth)	
	FeC 2 N 3 (mono)	FeN2 (orth)	Fe 23 B 6(cub)	
		Fe2N (tri)	Fe 2 B (tét)	
		Fe 4 N (cub)	FeN (hex)	
		FeN (cub)	Fe 23 C 6(cub)	
		Fe 3 N(hex)	Fe 5 C 2(mono)	
		Fe 3 N (tri)	Fe 7 C 3(ort)	
		Fe 8 N (tet)	Fe 4 C (mono)	
		Fe 8 N 3 (mono)	Fe 3 C (tet)	
		Fe 12 N 5(tri)	Fe 3 C (mono)	
		Fe 24 N 11(mono)	Fe 4 C (tet)	
		Fe 2 N (ort)	Fe 6 BC(orth)	
		Fe 2 N (orth)	Fe 46 (BC 3) 3(tricl)	
		Fe 3 C (ort)	Fe 23 (BC) 3(tri)	
		Fe 3 C (hex)	Fe 4 CN 4(ort)	
		Fe 5 C 2(mono)	Fe 65 (BN) 6 (mono)	
		Fe 2 C (ort)		
		Fe 7 C 3 (hex)		
		Fe 3 C (ort)		
		Fe 2 C(hex)		

Rappelons que pour le logiciel R défini des groupes (dans notre cas) et leur affecte chacun initialement un centre de gravité par groupe et dont les coordonnées dans l'espace d'axes F1, F2, F3, sont aléatoires. Une optimisation des coordonnées de ce centre est faite par R jusqu'à la convergence. Après convergence des calculs itératifs d'optimisation, les coordonnées finales des centres de gravité sont données par le tableau 12. Dans notre cas, la représentativité des individus est suffisamment faite par les axes principaux F1 et F2 (un taux cumulé de 80%). Il faut noter que les calculs montrent que pour notre cas d'étude, le centre de gravité de chaque groupe n'est pas confondu avec un point du groupe, Par ailleurs, le clustering fait par R montre que les groupes ne sont pas spécifiques à un ou à un autre nature chimique de composés (tableau 11). Les borures de fer se classent dans les groupes 1 et 3 et plus dans le groupe 4. Les nitrures de fer se classent dans les groupes 2 et 4 et plus dans le groupe 3. Les carbures de fer se classent à égalité dans les groupes 3 et 4. Quant aux

composés ternaires du fer Fe-(C, B et N), ils sont classés presque exclusivement dans le groupe 2 et 4.

Tableau 12. Coordonnées des centres de gravité des groupes.

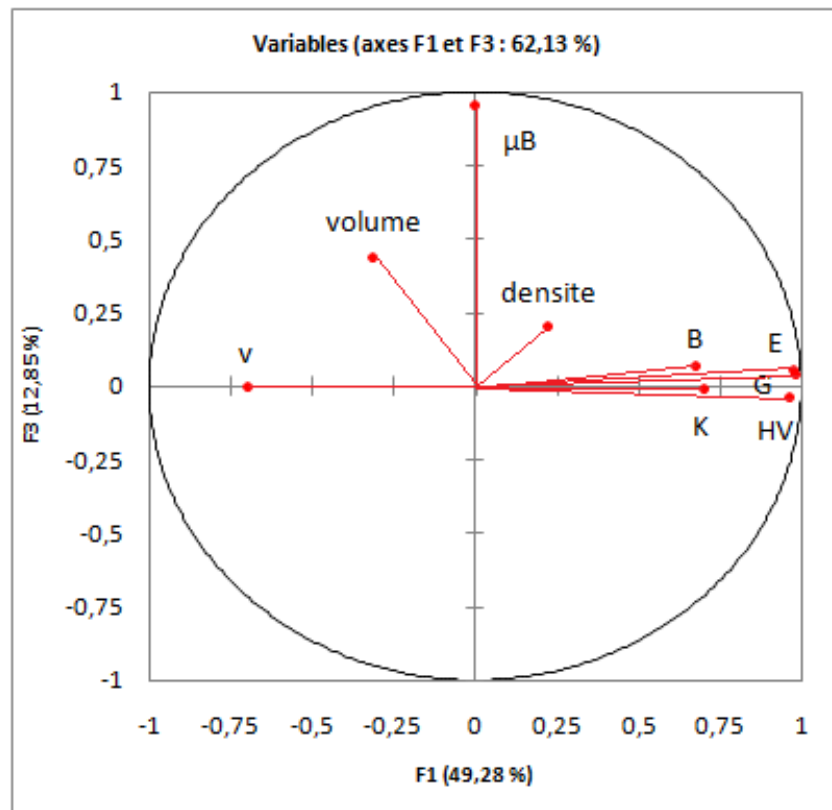
N° groupe	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	-1.0990	0.5528	0.3975	-0.2063	-0.0050	-0.1134	-0.0250	-0.0027	0.0009
2	-4.0035	-1.0572	-0.4007	-0.1017	0.0304	0.1870	0.0231	-0.0069	-0.0009
3	1.1751	0.86244	-0.2160	-0.0348	-0.0566	0.0463	0.0178	0.0029	-0.0003
4	3.1538	-3.8616	3.3926	-0.8029	0.1319	0.0656	-0.0237	-0.0010	-0.0019

I.18 Corrélations des variables avec le moment magnétique μ_B

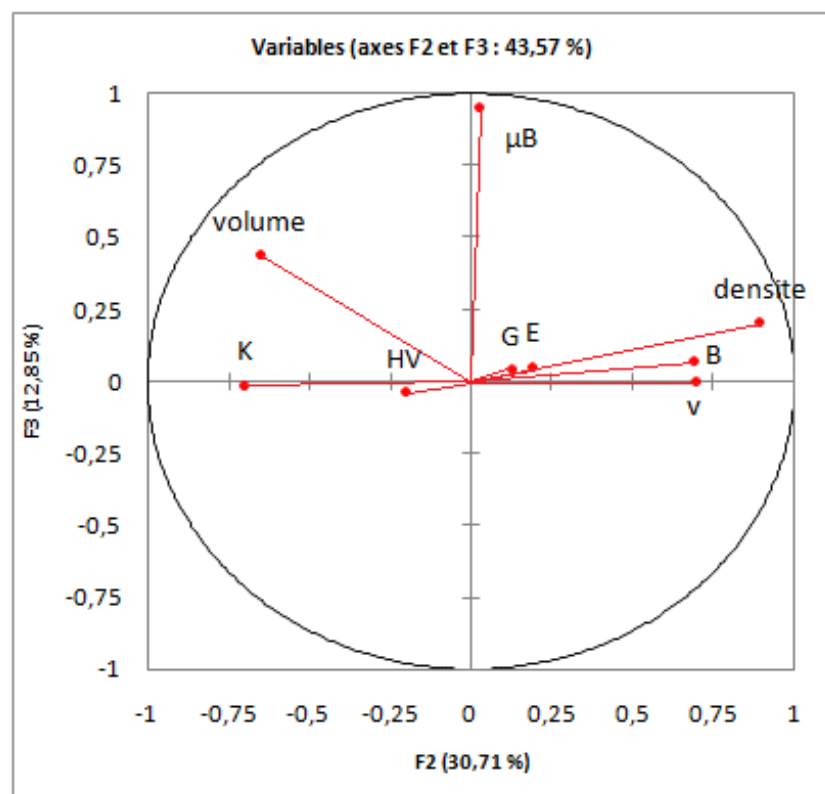
A travers l'analyse des données faite par l'ACP des 9 variables, en l'occurrence les constantes d'élasticité E, G, B, ν , les constantes de dureté K, HV, le volume molaire V, la densité ρ et le moment magnétique μ_B , nous avons interprétés les résultats obtenus selon les axes principaux F1 et F2 qui montrent une variabilité de 80% (tableau 2), témoignant ainsi de la bonne représentativité de l'information par ces deux axes. Toutefois, le moment magnétique n'a pas été pris en considération dans ces interprétations du fait de sa très faible contribution aux deux axes principaux considérés. En effet et d'après le tableau 5, la contribution cumulative de μ_B aux axes principaux F1 et F2 est presque nulle (0,003%). La plus grande valeur de la contribution du moment magnétique est relative à l'axe principal F3 et vaut 79%, suivi de l'axe F4 d'une valeur de 19%. Ainsi, l'axe F3 est suffisamment représentatif du moment magnétique. Toutefois, et d'après le tableau 2, l'axe F3 possède une variabilité de 13% ce qui témoigne qu'il n'est pas représentatif pour l'ensemble des variables. D'un autre côté, nous avons une situation contraire. Les axes F1 et F2 sont très représentatifs pour l'ensemble des variables avec une variabilité cumulative de 80% (tableau 2), alors qu'ils ne sont pas représentatif de μ_B . Dans cas, l'analyse ACP selon les axes F1 et F3 puis F2 et F3 constituent un bon compris entre perte et gain d'informations sur le globale des variables comparé à la variable μ_B .

Dans notre analyse ACP relative aux données du moment magnétique, nous nous sommes contentés de chercher seulement des corrélations entre μ_B et les huit autres variables. Pour cela et afin de cerner les pertes d'information des autres variables par rapport à l'axe

principal F3, nous considérons deux cercles de corrélation : le premier ayant comme axe principaux F1 et F3 et le second ayant pour axe principaux F2, F3 (figure 15).



(a)



(b)

Figure 15. Cercles de corrélations des variables.
a) axe principaux F1 et F3 b) axes principaux F2 et F3

La figure 15-a montre tout d'abord la perte d'information sur le volume V , la densité ρ , le module de compressibilité isostatique B et le rapport K . En revanche, le modules de Young E , le module de cisaillement G et la dureté Vickers HV possèdent une bonne représentativité et restent donc bien corrélés. Egalemeht, le coefficient de Poisson ν est assez bien représenté et corrélé négativement avec E , G et HV . Le moment magnétique n'est corrélé ni avec le module de Young, ni avec le module de cisaillement, ni avec le coefficient de Poisson et ni avec la dureté Vickers.

D'un autre coté, la figure 15-b montre une image inverse : perte d'information sur E , G , HV , et un gain de représentativité de V , K et ρ . La représentativité du coefficient de Poisson ν reste inchangée et est acceptable. De cette figure nous concluons que le moment magnétique n'est pas corrélés avec variables K , ρ , ν et très peu corrélé avec le volume molaire.

En conclusion, le moment magnétique μ_B n'est pas corrélé avec les 8 autres variables qui expriment la densité, l'élasticité et la dureté. Une telle constatation est prévisible du fait que le moment magnétique est de nature quantique (effets de spins) et c'est très peu probable qu'il y aura des corrélations entre μ_B et les autres propriétés physiques considérées qui, à l'échelle atomique, sont dictés par la nature des liaisons cristallines (interactions électrostatiques).

Conclusion générale

La conclusion finale est faite selon deux aspects : un aspect d'apport personnel et un autre scientifique relatif aux résultats des travaux de ce mémoire. Mais avant tout, il est important de signaler que la thématique du datamining est introduite pour la première fois dans un projet de fin d'études de master en physique au département des sciences de la matière.

Sur le plan apport personnel, ce projet de fin d'études m'a permis de m'initier sur le monde du datamining. J'ai ainsi pu me familiariser avec deux logiciels de traitement et analyse de données, à savoir le logiciel R et l'application XLSTAT. Cette dernière n'est pas abordée dans ce mémoire du fait que les résultats obtenus sont exactement les mêmes que ceux du logiciel R. De nouvelles connaissances scientifiques s'ajoutent donc à ma bibliothèque des connaissances acquises lors de ma formation en master.

Sur le plan scientifique, les résultats obtenus témoignent de l'efficacité de la méthode du datamining, relatif aux 53 matériaux choisis Fe-X (X: B, N, C), à décrire les corrélations existantes entre différentes propriétés physiques d'un part, et de classer les matériaux étudiés en des clusters de même comportement dans la carte de l'analyse des données d'autre part.

L'utilisation de la méthode de l'analyse en composantes principales (ACP) à travers le logiciel R nous a permis de décrypter dans un espace à 9 dimensions la base de données récoltée. L'ACP nous a permis également d'établir différentes corrélations linéaires entre les propriétés physiques considérées. Les résultats obtenus nous ont permis d'affirmer selon une analyse dans un espace plan F1 et F2, les corrélations suivantes :

- ✓ une corrélation linéaire positive entre les différentes propriétés élastiques E, G, K, et la dureté HV.
- ✓ une corrélation négative entre le coefficient de Poisson et les propriétés mécaniques (E, G, B, HV) et structurale (densité et volume).

En revanche, nous avons établi qu'aucune corrélation n'existe entre le moment magnétique et les autres propriétés physiques. Cette affirmation n'a pu être faite que lorsqu'on considère les trois premiers axes principaux F1, F2 et F3.

Par ailleurs, les résultats de Biplot nous ont permis de faire un clustering des composés du fer binaires et ternaires considérés selon quatre groupes. Chaque groupe n'est pas spécifique à une nature chimique unique des composés considérés.

Références bibliographiques

1. DUBY, C., & ROBIN, S. (2006). Analyse en composantes principales. Institut National Agronomique, Paris-Grignon, 80.
2. BAKER, R. S. J. D., et al. Data mining for education. International encyclopedia of education, 2010, vol. 7, no 3, p. 112-118.
3. Jain, N. & Srivastava, V. (2013). Data mining techniques: a survey paper. IJRET: International Journal of Research in Engineering and Technology, 2(11), 116-119.
4. MARTIN. A. L'analyse de données. Polycopié de cours, ENSIETA, septembre 2004.
5. Husson. F. Analyse en composantes principales (ACP). Laboratoire de mathématiques appliquées - Agrocampus Rennes.
6. Classification par la méthode des nuées dynamiques (K-MEANS) XLstat by addinsoft.
7. GUEDDOUH. A. Ab-initio calculation of the physical properties of some transition metal borides. 2017. Thèse de doctorat. University of Tlemcen.
8. The Magnetic Transition Temperature of Cementite. Proc. Phys. Soc. London. 2013.
9. Candan. A. Magnetic. Electronic. Mechanic. Anisotropic Elastic and Vibrational Properties of Antiferromagnetic Ru₂TGa (T = Cr. Mn. and Co) Heusler Alloys. Journal of Electronic Materials. 2019. 48(12): p. 7608-7622.
10. materials project . [https //materialsproject.org/](https://materialsproject.org/)
11. Analyse en Composantes Principales (ACP) avec Excel /XLstat support Center by Addinsoft.

Références des logiciel R et XLstat

12. François Husson. J.J.. Jérôme Pagès. Analyse de données avec R - Complémentarité des méthodes d'analyse factorielle et de classification. 2010.
13. R pour les débutants. Institut des Sciences de l'Evolution Université Montpellier II E-mail : paradis@isem.univ-montp2.fr
14. ACP - Analyse en Composantes Principales avec R: L'Essentiel. Statistical Tools For High-Throughput Data Analysis
15. The R Project for Statistical Computing.R-3.6.2.win
16. Principal ComponentAnalysis. Package 'factoextra' August 29, 2016
17. Principal ComponentAnalysis.Package 'FactoMineR'
18. Premiers pas en Machine Learning avec R. Volume 3 : k-means clustering
19. R Documentation and manuals
20. Cours de programmation sous R
21. Analyse en Composantes Principales (ACP) XLstat by addinsoft.
22. Classification par la méthode des nuées dynamiques (K-MEANS) XLstat by addinsoft.

Références des données des composés de Matériaux Project [10]

N°	Composé	URL du site	date de consultation
1	Fe3B (tet)	https://materialsproject.org/materials/mp-1181327/	Décembre2019
2	FeB2 (hex)	https://materialsproject.org/materials/mp-569376/	Décembre2019
3	Fe2B (tet)	https://materialsproject.org/materials/mp-568748/	Décembre2019
4	Fe3B (ort)	https://materialsproject.org/materials/mp-973682/	Décembre2019
5	FeB(ort)	https://materialsproject.org/materials/mp-1007881/	Décembre2019
6	FeB (orth)	https://materialsproject.org/materials/mp-20787/	Décembre2019
7	FeB4 (ort)	https://materialsproject.org/materials/mp-1079437/	Décembre2019
8	Fe2B7(ort)	https://materialsproject.org/materials/mp-1194531/	Décembre2019
9	FeB(tet)	https://materialsproject.org/materials/mp-1080525/	Décembre2019
10	Fe 23 B 6(cub)	https://materialsproject.org/materials/mp-542758/	Décembre2019
11	Fe 2 B (tét)	https://materialsproject.org/materials/mp-1915/	Décembre2019
12	FeN (tet)	https://materialsproject.org/materials/mvc-14553/	Décembre2019
13	FeN (cub)	https://materialsproject.org/materials/mp-1009019/	Décembre2019
14	FeN(cub)	https://materialsproject.org/materials/mp-1008929/	Décembre2019
15	Fe4N(tri)	https://materialsproject.org/materials/mp-1225203/	Décembre2019
16	FeN (hex)	https://materialsproject.org/materials/mp-12120/	Décembre2019
17	Fe2N(tri)	https://materialsproject.org/materials/mp-1224698/	Décembre2019
18	FeN2 (orth)	https://materialsproject.org/materials/mp-1080202/	Janvier 2020
19	Fe2N (tri)	https://materialsproject.org/materials/mp-248/	Janvier 2020
20	Fe 4 N (cub)	https://materialsproject.org/materials/mp-535/	Janvier 2020
21	FeN (cub)	https://materialsproject.org/materials/mp-6988/	Janvier 2020
22	Fe 3 N(hex)	https://materialsproject.org/materials/mp-1804/	Janvier 2020
23	Fe 3 N (tri)	https://materialsproject.org/materials/mp-673635/	Janvier 2020
24	Fe 8 N (tet)	https://materialsproject.org/materials/mp-555/	Janvier 2020
25	Fe 8 N 3 (mono)	https://materialsproject.org/materials/mp-673174/	Janvier 2020
26	Fe 12 N 5(tri)	https://materialsproject.org/materials/mp-27908/	Janvier 2020
27	Fe 24 N 11(mono)	https://materialsproject.org/materials/mp-684887/	Janvier 2020
28	Fe 2 N (ort)	https://materialsproject.org/materials/mp-684744/	Janvier 2020
29	Fe 2 N (orth)	https://materialsproject.org/materials/mp-21476/	Janvier 2020
30	Fe 23 C 6(cub)	https://materialsproject.org/materials/mp-1192884/	Janvier 2020
31	Fe 3 C (ort)	https://materialsproject.org/materials/mp-510623/	Janvier 2020
32	Fe 3 C (hex)	https://materialsproject.org/materials/mp-13154/	Janvier 2020
33	Fe 5 C 2(mono)	https://materialsproject.org/materials/mp-2794/	Janvier 2020
34	Fe 5 C 2(mono)	https://materialsproject.org/materials/mp-645339/	Janvier 2020
35	Fe 2 C (ort)	https://materialsproject.org/materials/mp-1871/	Janvier 2020
36	Fe 7 C 3(ort)	https://materialsproject.org/materials/mp-21717/	Janvier 2020
37	Fe 7 C 3 (hex)	https://materialsproject.org/materials/mp-18257/	Janvier 2020
38	Fe 3 C (ort)	https://materialsproject.org/materials/mp-613523/	Janvier 2020

39	Fe 4 C (mono)	https://materialsproject.org/materials/mp-1078579/	Janvier 2020
40	Fe 3 C (tet)	https://materialsproject.org/materials/mp-1205436/	Janvier 2020
41	Fe 3 C (mono)	https://materialsproject.org/materials/mp-1205444/	Janvier 2020
42	Fe 4 C (tet)	https://materialsproject.org/materials/mp-1070333/	Janvier 2020
43	Fe 2 C(hex)	https://materialsproject.org/materials/mp-568503/	Janvier 2020
44	Fe 6 BC(orth)	https://materialsproject.org/materials/mp-1225026/	Janvier 2020
45	Fe 46 (BC 3) 3(tricl)	https://materialsproject.org/materials/mp-1226402/	Janvier 2020
46	Fe 23 (BC) 3(tri)	https://materialsproject.org/materials/mp-1224758/	Janvier 2020
47	Fe 2 (CN 2) 3(tri)	https://materialsproject.org/materials/mp-1245918/	Janvier 2020
48	FeCN 2 (hex)	https://materialsproject.org/materials/mp-567933/	Janvier 2020
49	Fe(C 2 N 3) 3(orth)	https://materialsproject.org/materials/mp-1246055/	Février 2020
50	Fe 4 CN 4(ort)	https://materialsproject.org/materials/mp-1246069/	Février 2020
51	Fe 2 (CN 2) 3 (tri)	https://materialsproject.org/materials/mp-1245833/	Février 2020
52	FeC 2 N 3 (mono)	https://materialsproject.org/materials/mp-1245528/	Février 2020
53	Fe 65 (BN) 6 (mono)	https://materialsproject.org/materials/mp-1097708/	Février 2020

Résumé: Dans notre travail nous avons appliqué deux techniques du datamining, en l'occurrence l'Analyse en Composantes Principales (ACP) et le clustering (k-means), afin d'extraire, pour un ensemble de 53 composés à base de fer (borures, carbures, nitrides), des corrélations linéaires entre certaines propriétés mécaniques, tels que le module de Young (E), le module de d'élasticité isostatique (B), le module de cisaillement (G), la dureté de Vickers (HV), le coefficient de poisson (ν), le rapport $K=G/B$ et une des propriétés magnétiques : le moment magnétique (μ_B). L'analyse ACP est faite par le logiciel R. A partir des graphiques de l'ACP, un clustering et une classification sont faites. Nous avons établi des corrélations linéaires positives entre les propriétés élastique E, G, B, HV et des corrélations linéaires négatives entre le coefficient de Poisson de les propriétés (G, B, E, K, HV). En revanche, aucune corrélation linéaire n'existe entre le moment magnétique et les autres propriétés. Les 53 composés sont rassemblés en 4 groupes différents à partir de leurs propriétés physiques. Nous avons montré que chaque groupe n'est pas spécifique à une nature chimique donnée des composés considérés.

Mots clés : Datamining , Analyses en Composantes Principales, ACP, clustering, k-means, logiciel R ,borures de fer ,carbures de fer, nitrides de fer, modules d'élasticité, dureté, moment magnétique.

Abstract: In our work we applied two data mining techniques: Principal Component Analysis (PCA) and clustering (k-means), in order to extract for a set of 53 iron-based compounds (borides , carbides, nitrides) linear correlations between some mechanical properties, such as Young's modulus (E), bulk modulus (B), shear modulus (G), Vickers hardness (HV) , the Poisson's coefficient (ν), the ratio $K = G / B$ and the magnetic moment (μ_B). The PCA analysis is done by the R software. From PCA graphics, we built clustering and classification. We have established positive linear correlations between the elastic properties E, G, B, HV and negative linear correlations between the Poisson's ratio and the properties (G, B, E, K, HV). On the other hand, no linear correlation exists between the magnetic moment and the other properties. The 53 compounds are grouped into 4 different groups. We show that each group is not specific to a given chemical nature of the considered compounds.

Key words: Datamining, Principal Component Analysis, PCA, clustering, k-means, R software, iron borides, iron carbides, iron nitrides, elasticity modules, hardness, magnetic moment.

المخلص: تم في هذا العمل استخدام تقنيتين من تقنيات البحث في البيانات (تحليل المكونات الرئيسية (PCA) وخوارزمية التصنيف (clustering k-means))، من أجل دراسة 53 مركبًا (البوريدات الحديد، الكريبيدات الحديد، النتريدات الحديد)، وإيجاد الارتباطات الخطية بين بعض الخواص الميكانيكية، مثل معامل يونغ (E)، معامل الحجم المروني (B)، معامل القص (G)، صلابة فيكرز (HV)، معامل بواسون (ν)، النسبة ($K = G / B$) وأحد الخصائص المغناطيسية العزم المغناطيسي (μ_B). تم إجراء كلا التقنيتين بواسطة برنامج R وقُدِّمت النتائج على شكل رسومات بيانية.

تم اثبات وجود ارتباطات خطية إيجابية (متناسبة طردياً) بين الخصائص المرنة E، G، B، HV والارتباطات الخطية السلبية (متناسبة عكسياً) بين معامل بواسون (ν) والخصائص (G، B، E، K، HV). من ناحية أخرى، تين أنه لا يوجد أي ارتباط خطي بين العزم المغناطيسي وباقي الخصائص الأخرى المذكورة. تم تجميع المركبات في أربعة مجموعات مختلفة بناءً على خصائصها الفيزيائية وقد تبين أن تصنيف المجموعات ليس له أي علاقة بالطبيعة الكيميائية للمركب ما من بين المركبات المدروسة.

الكلمات المفتاحية: التنقيب في البيانات، تحليل المركبات الأساسية (ACP)، خوارزمية التصنيف (clustering k-means)، برنامج R، بوريدات الحديد، كريبيدات الحديد، نتريدات الحديد، الخصائص الميكانيكية، الصلابة، العزم المغناطيسي .