

Université du Québec  
Institut National de la Recherche Scientifique (INRS)  
Centre Énergie Matériaux Télécommunications

## **ANALYSE DES SPECTRES LIBS PAR CHIMIOMÉTRIE**

Par

Jordan Fernandes

Maitrise présentée pour l'obtention du grade de  
Maître ès Sciences (M.Sc.)  
en sciences de l'énergie et des matériaux

### **Jury d'évaluation**

Président du jury et  
examineur interne

Prof. Jean-Pierre Matte  
INRS-EMT

Co-directeur de recherche

Prof. Bruno Bousquet  
CELIA

Directeur de recherche

Prof. François Vidal  
INRS-EMT

## Résumé

Dans ce travail, plusieurs méthodes ont été utilisées pour déterminer la concentration en or d'échantillons rocheux au moyen de la technique LIBS. Cette technique est basée sur l'interprétation du rayonnement émis par le plasma produit par laser à la surface du matériau à analyser. Pour déterminer la concentration de certains éléments chimiques comme l'or, l'approche la plus couramment employée en LIBS consiste à utiliser une courbe de calibration reliant l'intensité d'une raie spectrale de l'élément d'intérêt à sa concentration. L'inconvénient de cette approche est que la courbe de calibration doit être établie à l'aide d'un matériau dont le spectre est proche de celui du matériau à analyser. Comme le minerai d'or peut contenir de nombreuses phases de compositions chimiques différentes, plusieurs courbes de calibration seraient en principe nécessaires. En contraste avec cette approche dite univariée, les approches multivariées supervisées permettent en principe de contourner le problème des multiples courbes de calibration. Elles sont basées sur l'exploitation, par extrapolation ou généralisation, d'un apprentissage qui met en correspondance les spectres d'échantillons de référence avec leur concentration d'or.

Dans ce travail, nous constatons d'abord les limitations des méthodes univariées pour prédire la concentration d'or de différents matériaux de référence homogènes, composés de poudres compactées, dont la concentration d'or est connue. La présence de fer dans les matériaux (un élément de concentration variable dans le minerai d'or) crée une grande disparité dans les spectres normalisés qui soulève des difficultés pour calibrer les mesures de concentration. Les méthodes multivariées supervisées PLS (Partial Least Square) et ANN (Artificial Neural Network) ont ensuite été explorées pour déterminer la concentration d'or dans une série de roches taillées en parallélépipèdes de quelques cm de côtés, dont la concentration volumique en or a été déterminée par des méthodes classiques de laboratoire.

Dans un premier temps, les méthodes PLS et ANN ont été appliquées aux spectres obtenus en moyennant des milliers de spectres prélevés à autant de points à la surface des roches. La comparaison entre la concentration volumique du laboratoire et la concentration surfacique obtenue par LIBS ne s'est pas avérée satisfaisante en raison d'écarts aberrants dans plusieurs cas.

Ensuite, une approche hybride mettant en œuvre à la fois une approche multivariée et des courbes de calibration a été explorée. Cette fois, les méthodes multivariées étaient non-supervisées et ont été utilisées pour classer le spectre issu de chaque tir laser sur la roche selon sa ressemblance avec le spectre de l'un des 5 matériaux de référence dont nous disposons. Les méthodes multivariées employées sont ACP (Analyse en Composantes Principales) et ANN. (Une méthode basée sur la corrélation entre les spectres a aussi été utilisée pour produire des cartes permettant de visualiser le degré de ressemblance entre les spectres locaux et les spectres de référence.) Ensuite, la courbe de

calibration correspondant à ce matériau de référence a été utilisée pour déterminer la concentration d'or associée à au spectre prélevé sur la roche. La moyenne des concentrations ainsi déterminées pour chaque tir laser (des milliers) à la surface de la roche a ensuite été comparée avec la mesure du laboratoire. Ici encore, la comparaison a été plus ou moins satisfaisante, ce qui laisse la porte grande ouverte pour des développements additionnels dans le but d'aboutir à des résultats pratiques fiables.

## Abstract

In this work, several methods were used to determine the gold concentration of rock samples using the LIBS technique. This technique is based on the interpretation of the radiation emitted by the plasma produced by laser on the surface of the material to be analyzed. To determine the concentration of certain chemical elements such as gold, the most common approach used in LIBS consists in using a calibration curve linking the intensity of a spectral line of the element of interest to its concentration. The disadvantage of this approach is that the calibration curve must be established using a material whose spectrum is close to that of the material to be analyzed. Since gold ore can contain many phases of different chemical compositions, several calibration curves would in principle be necessary. In contrast to this so-called univariate approach, supervised multivariate approaches allow in principle to bypass the problem of multiple calibration curves. They are based on the exploitation, by extrapolation or generalization, of a learning process that maps the spectra of reference samples to their gold concentration.

In this work, we first note the limitations of univariate methods to predict the gold concentration of different homogeneous reference materials, composed of compacted powders, whose gold concentration is known. The presence of iron in the materials (an element of variable concentration in gold ore) creates a large disparity in the normalized spectra that raises difficulties in calibrating the concentration measurements. Supervised multivariate PLS (Partial Least Square) and ANN (Artificial Neural Network) methods were then explored to determine the concentration of gold in a series of rocks cut into parallelepipeds of a few cm sides, whose volume concentration of gold was determined by conventional laboratory methods.

In a first step, PLS and ANN methods were applied to the spectra obtained by averaging thousands of spectra taken from as many points on the surface of the rocks. The comparison between the laboratory volume concentration and the LIBS surface concentration was not satisfactory due to outliers in several cases.

Next, a hybrid approach implementing both a multivariate approach and calibration curves was explored. This time, the multivariate methods were unsupervised and were used to classify the spectrum from each laser shot on the rock according to its similarity to the spectrum of one of the 5 reference materials available to us. The multivariate methods used are PCA (Principal Component Analysis) and ANN. (A method based on the correlation between the spectra was also used to produce maps allowing to visualize the degree of similarity between the local spectra and the reference spectra). Then, the calibration curve corresponding to this reference material was used to determine the gold concentration associated with the spectrum collected from the rock. The average of the concentrations thus determined for each laser shot (thousands) at the rock surface was then compared

with the laboratory measurement. Again, the comparison was more or less satisfactory, which leaves the door wide open for further development in order to achieve reliable practical results.

## Table des matières

Résumé.....	2
Abstract .....	4
Table des figures .....	7
Table des tableaux.....	9
Introduction .....	10
Chapitre 1 : LIBS, contexte et étude univariée.....	12
1.1 Principe de fonctionnement du LIBS .....	12
1.2 Contexte et mesures .....	13
1.3 Étude univariée .....	18
1.3.1 Courbes d'étalonnage des échantillons de référence .....	20
1.3.2 Test des courbes de calibration.....	26
Chapitre 2 : Outils de chimiométrie .....	28
2.1 Analyse en Composantes Principales (ACP) .....	28
2.2 Régression aux moindres carrés partiels (PLS).....	29
2.3 Réseaux de neurones artificiels (ANN).....	31
2.3.1 Transmission du signal .....	33
2.3.2 Retro-propagation de l'erreur .....	34
Chapitre 3 : Étude multivariée.....	36
3.1 Analyse en composantes principales des échantillons .....	37
3.2 Analyse quantitative des échantillons par PLS.....	40
3.3 Analyse quantitative par ANN .....	44
Chapitre 4 : Méthodes de quantification hybrides.....	48
4.1 Cartes de corrélation des spectres d'évaluation avec les échantillons de calibration .....	48
4.2 Sélection des courbes de calibration adaptées aux échantillons par ACP :.....	51
4.3 Sélection des courbes de calibration adaptées aux échantillons par ANN :.....	52
Conclusion.....	55
Références .....	56

## Table des figures

Figure 1. 1 Schéma général de fonctionnement d'une expérience LIBS [12].....	12
Figure 1. 2 Évolution temporelle du spectre LIBS dans de l'eau contenant des traces de Ca, Na, Li et K. Le temps d'intégration est de 1 ms. [2].....	13
Figure 1. 3 Dispositif expérimental LIBS utilisé dans le cadre de la quantification de la concentration d'or dans les roches minières. [3] .....	15
Figure 1. 4 Comparaison du spectre simulé par NIST data base LIBS et du spectre expérimental WMS-1 à 500 ppm (en bleu) (a) avec zoom et (b) sans zoom. Les pics de Fe II sont en violet, les pics de Fe I sont en jaune, les pics de Mn II sont en rouge et les pics d'Au I sont en vert.....	17
Figure 1. 5 Variation du rapport Fe II 271.44 nm sur Fe I 271.904 nm en fonction de la température.....	18
Figure 1. 6 (a) Spectres moyens normalisés par rapport à l'aire spectrale de tous les échantillons de référence à une concentration de 200 ppm dans la gamme spectrale 265 – 271 nm. Les flèches bleues représentent les raies de Fer II, les flèches oranges les raies d'Al I, les flèches vertes les raies de Fer I et la flèche violette la raie d'Au I à 267.59 nm (b) Zoom sur la raie d'or à 267.59 nm et la raie de Fe à 267.9 nm.....	19
Figure 1. 7 Comparaison des courbes de calibration de tous les lots de référence. Le background se situe à 267,37 nm. ....	21
Figure 1. 8 (a) Courbe de calibration des échantillons WMS-1 normalisée par rapport à l'aire de l'ensemble du domaine spectral. (b) Évaluation du modèle de prédiction par rapport aux concentrations de référence.....	21
Figure 1. 9 (a) Courbe de calibration des échantillons QC normalisés par rapport à l'aire de l'ensemble du domaine spectral. (b) Évaluation du modèle de prédiction par rapport aux concentrations de référence.....	23
Figure 1. 10 (a) Courbe de calibration des échantillons WMS-1 et PTC-1b normalisée par rapport à l'aire de l'ensemble du domaine spectral. (b) Évaluation du modèle de prédiction par rapport aux concentrations de référence.....	24
Figure 1. 11 (a) Courbe de calibration des échantillons JG-3, LDI-1 et QC normalisée par rapport à l'aire de l'ensemble du domaine spectral. (b) Évaluation du modèle de prédiction par rapport aux concentrations de référence.....	25
Figure 1. 12 (a) Prédiction des concentrations des échantillons QC à l'aide du modèle de calibration JG-3. (b) Prédiction des concentrations des échantillons WMS-1 à l'aide du modèle de calibration JG-3.....	26
Figure 2. 1 ACP appliquée sur des données LIBS provenant de [3].....	29
Figure 2. 2 Réseau de neurones à 3 couches : la couche d'entrée, la couche intermédiaire et la couche de sortie. Les poids $W_{ij}$ relient les neurones $i$ de la couche d'entrée ( $i = 1, \dots, I$ ) aux neurones $j$ de la couche intermédiaire ( $j = 1, \dots, J$ ). Les poids $W_{jk}$ relient les neurones $j$ de la couche intermédiaire aux neurones $k$ la couche de sortie ( $k = 1, \dots, K$ ). [1].....	32
Figure 2. 3 Schéma de fonctionnement du réseau de neurones. L'erreur est propagée dans le sens inverse de calcul du réseau. ....	33
Figure 2. 4 Évolution du RMSE par rapport au nombre d'itérations pour le lot de validation en rouge et le lot d'apprentissage en bleu. Le minimum de la courbe rouge correspond au critère d'arrêt de l'apprentissage. [8] .....	35
Figure 3. 1 Comparaison entre le spectre normalisé par rapport à l'aire spectrale de la face 3 de l'échantillon Lar-37 (a) non corrigé en longueurs d'onde et (b) corrigé en longueurs d'onde par une transformation linéaire, et ceux des échantillons de calibration WMS-1 et JG-3 à 50 ppm. ....	37
Figure 3. 2 Scores des échantillons de calibration suivant les deux premières composantes principales. Chaque point correspond à un échantillon de son lot de calibration.....	39

<b>Figure 3. 3 (a) Loadings de la première composante principale suivant la longueur d'onde. (b) Loadings de la seconde composante principale suivant la longueur d'onde.....</b>	<b>39</b>
<b>Figure 3. 4 Scores des échantillons de calibration et des échantillons d'évaluation suivant les deux premières composantes principales calculées avec les échantillons de calibration. Chaque point d'un lot de calibration représente un échantillon de référence. Chaque point d'un échantillon d'évaluation représente le spectre moyen de chacune des 6 faces qui le composent. ....</b>	<b>40</b>
<b>Figure 3. 5 Résultats RMSECV sur 12 composantes principales par PLS. ....</b>	<b>41</b>
<b>Figure 3. 6 (a) Concentrations d'or prédites par le modèle PLS à 6 composantes principales en utilisant tous les lots de calibration pauvres en fer (JG-3, LDI-1 et QC). (b) Poids calculés par PLS sur les lots de calibration pauvres en fer avec 6 composantes principales.....</b>	<b>41</b>
<b>Figure 3. 7 Résultats de RMSECV sur 7 composantes principales par PLS. ....</b>	<b>43</b>
<b>Figure 3. 8 (a) Concentrations d'or prédites avec le modèle PLS à 4 composantes principales en utilisant tous les lots de calibration. (b) Poids calculés par PLS sur les lots de calibration avec 4 composantes principales. ....</b>	<b>43</b>
<b>Figure 3. 9 Résultat de RMSECV et RMSEC selon le nombre d'itérations par validation croisée. ....</b>	<b>46</b>
<b>Figure 4. 1 Cartes de corrélation entre la face 1 de Lar-37 et (a) WMS-1, (b) JG-3, (c) JG-3 et WMS-1. Les nombres sur les axes dénotent les coordonnées des tirs laser dans la matrice rectangulaire.....</b>	<b>49</b>
<b>Figure 4. 2 Carte de corrélation entre la face 1 de Gol-70 et (a) WMS-1, (b) JG-3, (c) JG-3 et WMS-1.....</b>	<b>50</b>
<b>Figure 4. 3 (a) Carte de corrélation entre la face 2 de West-30 et (a) WMS-1, (b) JG-3, (c) JG-3 et WMS-1, (d) PTC-1b, (e) LDI-1, (f) QC. ....</b>	<b>50</b>
<b>Figure 4. 4 Histogramme du choix de la courbe de calibration la plus adaptée par ANN. Les chiffres 1, 2, 3, 4 et 5 correspondent respectivement à WMS-1, JG-3, LDI-1, PTC-1b et QC. ....</b>	<b>54</b>

## Table des tableaux

Tableau 1. 1 Échantillons de calibration utilisées ainsi que leurs compositions élémentaires. ....	16
Tableau 1. 2 Valeurs de températures et de densités utilisées pour la simulation du spectre NIST. .....	17
Tableau 1. 3 Comparaison des valeurs de concentration de référence aux valeurs prédites par le modèle pour WMS-1. ....	22
Tableau 1. 4 Comparaison des valeurs $R^2$ , de RMSEC et de RMSECV des courbes de calibration pour les différents échantillons de référence.....	22
Tableau 1. 5 Comparaison des valeurs de concentration de référence aux valeurs prédites par le modèle pour QC.....	23
Tableau 1. 6 Comparaison des valeurs de concentration de référence avec les valeurs prédites par le modèle JG-3, LDI-1 et QC.....	26
Tableau 3. 1 Échantillons utilisés pour évaluer les résultats obtenus par la méthode LIBS par comparaison avec les mesures de laboratoire. ....	36
Tableau 3. 2 Comparaison entre les concentrations mesurées par les laboratoires et les concentrations obtenues par la méthode PLS pour les échantillons test en utilisant pour l'apprentissage les échantillons de calibration JG-3, LDI-1 et QC.....	42
Tableau 3. 3 Comparaison entre les concentrations mesurées par le laboratoire et les concentrations obtenues par la méthode PLS pour les échantillons d'évaluation en utilisant pour l'apprentissage tous les échantillons de calibration.....	44
Tableau 3. 4 Comparaison entre les concentrations mesurées par le laboratoire et les concentrations obtenues par la méthode ANN pour les échantillons d'évaluation en utilisant pour l'apprentissage tous les échantillons de calibration.....	46
Tableau 4. 1 Dimensions du scan laser et du nombre de tirs faits sur la surface des échantillons Casper-17, Gol-70 et West-30.....	48
Tableau 4. 2 Résultat de la prédiction des échantillons de test par la méthode hybride en calculant la moyenne de la concentration obtenue pour chaque tir sur toutes les face des échantillons de test. .....	52
Tableau 4. 3 Nombre de tirs projetés sur la surface des échantillons test.....	53
Tableau 4. 4 Résultat de la quantification par la sélection des courbes de calibration univariées adaptées aux échantillons d'évaluation par ANN.....	54

## Introduction

L'industrie minière joue un rôle important dans la société. Elle réalise un chiffre d'affaires mondial supérieur à 600 milliard d'euros ce qui constitue une activité fructueuse et indispensable. Elle est soumise à une demande croissante de la part des industriels, des sociétés liées à la production d'énergie ou encore des sociétés de constructions de bâtiments.

Pour répondre à cette demande, l'industrie minière n'a pas d'autres choix que de prendre plus de risques, d'essayer de produire plus. Néanmoins cette augmentation a un prix. En effet, le fait d'augmenter la production pousse les industries minières à consommer de plus en plus d'énergie et indirectement à produire plus de déchets. Cet impact écologique est inévitable et ne fera que croître au cours du temps.

Une des solutions à ces problèmes consisterait à déterminer avec plus de précision les sites de forage à haute teneur en minerai. Dans l'idéal il faudrait que cette méthode de sélection soit rapide, précise et peu coûteuse. Il existe à ce jour plusieurs méthodes d'analyse élémentaire pour parvenir à cette sélection des meilleurs sites, comme la fluorescence à rayon X (XRF), la spectroscopie d'émission atomique d'un plasma créé par couplage inductif (ICP-AES), la spectroscopie de masse d'un plasma créé par couplage inductif (ICP-MS), ou encore la spectroscopie d'émission atomique d'un plasma créé par laser (LIBS).

Dans cette étude, nous avons choisi la LIBS qui est aujourd'hui exploitée dans un très grand nombre d'applications et dont la pertinence pour les analyses géologiques a d'abord été démontrée par l'instrument ChemCam installé à bord du rover Curiosity et depuis février 2021 par l'instrument SuperCam installé à bord du rover Perseverance, tous deux mis en œuvre par la NASA sur Mars. Dans le contexte qui nous concerne, la LIBS semble pertinente pour prendre une première décision quant à la rentabilité espérée d'un site minier, même si elle ne permet d'analyser qu'en surface les échantillons de roche pour en estimer la concentration de l'élément d'intérêt dans le minerai. L'échantillonnage des points d'analyse LIBS sera donc inévitablement une étape clé pour que quelques analyses fournissent une information à grande échelle. Nous ne discuterons pas ce point-là et nous centrerons la discussion sur les analyses elles-mêmes.

Pour déterminer la concentration de l'élément chimique d'intérêt, l'analyste LIBS a recours à la construction d'une courbe d'étalonnage au moyen de matériaux de référence de même nature que ceux qui seront ensuite analysés et couvrant une gamme de concentrations suffisante pour décrire le contexte d'étude. Cette courbe, qui correspond au signal LIBS (intensité d'une raie atomique) en fonction de la concentration, ne prend pas en compte les effets de matrices, c'est-à-dire la diversité des compositions chimiques des échantillons naturels comme les roches. Des techniques multivariées de chimiométrie permettent en principe de prendre en compte ces effets.

La chimiométrie est décrite comme l'application de méthodes mathématiques et statistiques dans le but d'extraire le maximum d'informations de mesures issues d'analyses chimiques. Ces méthodes peuvent être supervisées comme la régression aux moindres carrés partiels (PLS) ou non supervisées comme l'analyse en composantes principales (PCA). De plus, elles peuvent être non linéaires comme les réseaux de neurones artificiels (ANN).

Les travaux de ce mémoire ont été effectués à l'Institut National de la Recherche Scientifique (INRS) de Varennes, Québec, au Conseil National de Recherche Canada (CNRC) de Boucherville, Québec, et au Centre Lasers Intenses et Applications (CELIA) à l'université de Bordeaux, France. Ils ont consisté à traiter des spectres LIBS enregistrés sur des échantillons de roches provenant de plusieurs mines d'or au Québec dans le but de quantifier le plus précisément possible la concentration d'or présente dans les roches minières.

Dans le chapitre 1 nous définirons le principe de fonctionnement de la LIBS, le contexte d'étude, l'expérience ainsi que l'étude univariée des échantillons. Puis dans le chapitre 2 nous allons expliquer de manière théorique les méthodes de chimiométrie utilisées, soit PCA, PLS et ANN. Ensuite dans le chapitre 3 nous appliquerons ces méthodes multivariées sur les échantillons de roches minières. Enfin, dans le chapitre 4 nous utiliserons des méthodes hybrides pour analyser la concentration en or de ces échantillons.

# Chapitre 1 : LIBS, contexte et étude univariée

## 1.1 Principe de fonctionnement du LIBS

La LIBS consiste à créer un plasma à la suite d'une ablation laser puis à collecter et à analyser spectralement la lumière émise par celui-ci. Le laser délivre des impulsions de quelques millijoules qui permettent d'atteindre une température de plasma de l'ordre de 10 000 K. Les ions et les atomes présents dans le plasma dans différents états excités émettent des rayonnements brefs caractéristiques des transitions radiatives détectées à l'aide d'un spectromètre optique (Figure 1. 1).

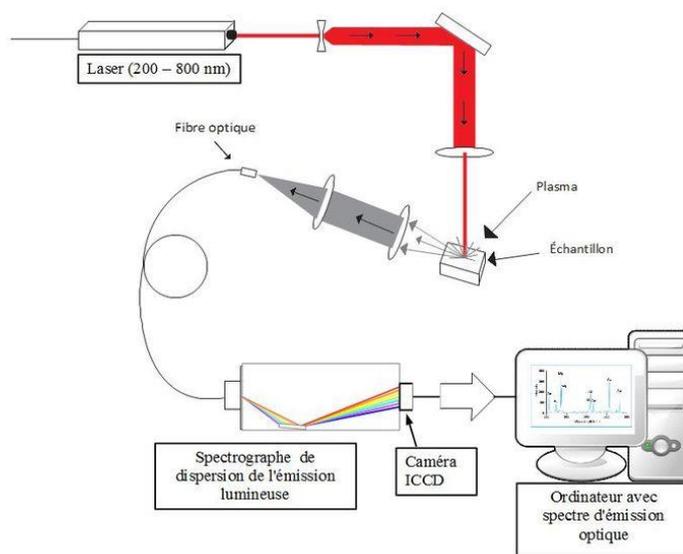


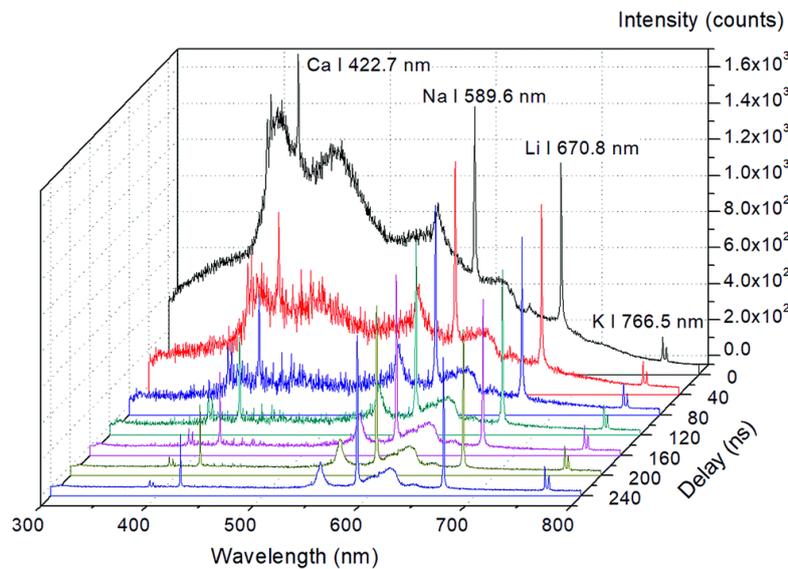
Figure 1. 1 Schéma général de fonctionnement d'une expérience LIBS [12]

Selon le délai d'acquisition du signal du spectromètre il y a deux types de rayonnement observés :

- Le rayonnement continu, qui est très intense suite à la formation du plasma. Il naît du rayonnement Bremsstrahlung (rayonnement de freinage des électrons) qui se produit lorsqu'un électron libre perd une partie de son énergie cinétique sous forme d'un photon en passant au voisinage d'un ion sans pour autant se recombiner. Le rayonnement continu provient aussi en partie de la recombinaison entre les électrons libres et les ions.
- Le rayonnement discret émerge lorsqu'on décale l'acquisition d'une centaine de nanosecondes. Il résulte de la désexcitation radiative des ions et atomes présents dans le plasma.

Pour éviter que le rayonnement continu masque le rayonnement discret il faut retarder le début de l'acquisition des spectres LIBS. Néanmoins il ne faut pas non plus trop retarder l'acquisition des mesures

car le plasma a une durée de vie allant de 100 ns à 50  $\mu$ s. Il faut donc trouver un compromis pour maximiser le signal des raies qui est à la base de la LIBS. Sur la figure 1.2, provenant de la référence [2], on observe la prédominance du rayonnement continu dans les premiers instants du plasma. Au cours du temps, ce rayonnement devient de moins en moins grand et le rayonnement discret est de plus en plus discernable avec pour meilleur délai d'acquisition environ 200 ns.



**Figure 1. 2** Évolution temporelle du spectre LIBS dans de l'eau contenant des traces de Ca, Na, Li et K. Le temps d'intégration est de 1 ms. [2]

Comme mentionné plus haut, une mesure quantitative de l'élément d'intérêt, ici l'or, nécessite l'établissement d'un modèle d'étalonnage au moyen de matériaux dont la concentration en or est connue et spécialement préparés à cet effet. Chaque tir laser sur la surface d'un échantillon quelconque permet ainsi de déterminer la concentration en or dans la région délimitée en première approximation par le point focal du laser. Dans le cas d'un échantillon où l'or est distribué d'une manière hétérogène, comme dans les roches minières, il est évidemment nécessaire d'effectuer de nombreux tirs laser sur sa surface pour déterminer sa teneur moyenne en or. Cette question est abordée dans [3] et [11] mais sort du cadre de ce travail.

## 1.2 Contexte et mesures

Dans le but de quantifier la concentration d'or dans les roches minières nous avons utilisé le dispositif expérimental présenté sur la figure 1.3. Il est composé d'un laser Nd :YAG (Continuum Surelite I) capable de délivrer des impulsions de 600 mJ à un taux de répétition de 10 Hz. La largeur à mi-hauteur temporelle de l'impulsion laser est de 8 ns. Le faisceau laser a un profil Gaussien de divergence 0,8 mrad et de diamètre 600  $\mu$ m.

Deux spectromètres ont été utilisés :

- Un spectromètre Czerny-Turner (modèle McPherson 207) de distance focale 0,67 m et d'ouverture numérique F/5.8. Il est équipé d'un réseau de diffraction de 2400 traits par millimètres blazé à 300 nm. Il permet de couvrir une gamme de longueurs d'onde allant de 180 à 650 nm. Sa dispersion linéaire est de 0.53 nm par millimètre à 267 nm. (La raie d'intérêt pour quantifier l'or se situe à 267.59 nm)
- Un spectromètre Czerny Turner (modèle Avantes) couvrant une gamme de longueur d'onde allant de 200 à 1100 nm. Il est équipé d'un réseau contenant 600 traits par millimètre blazé à 250 nm. Il dispose d'une moins bonne résolution que le modèle McPherson. Sa dispersion linéaire vaut 1.2 nm à 267 nm.

Une caméra ICCD (intensified charged-coupled device) Istar DH720-25H-05 de technologie Andor est couplé au spectromètre McPherson. Cette caméra possède sur sa surface photosensible 1024 x 256 pixels de dimension 26 x 26  $\mu\text{m}^2$ . Le spectromètre Avantes est couplé à une caméra linéaire CCD (Chip Sony ILX554) qui contient 2048 pixels de dimensions 14 x 56  $\mu\text{m}^2$ .

Le faisceau laser est focalisé sur la surface de l'échantillon par une lentille convergente de distance focale 20 cm. L'échantillon était monté sur une plateforme motorisée suivant 3 axes (Newport UTM 100 mm) et contrôlée par une télécommande programmable (Newport modèle ESP 300). L'émission du plasma est réfléchiée par un miroir dichroïque puis divisée en deux parties par un cristal diviseur de telle sorte que 8% de l'émission du plasma arrive au spectromètre Avantes et la partie restante au spectromètre McPherson par le biais d'un paquet de 25 fibres optiques de 100  $\mu\text{m}$  de diamètre.

Dans cette étude, l'énergie du laser était de 60 mJ et le diamètre du faisceau focalisé sur la cible était 600  $\mu\text{m}$ . La raie Au I 267.59 nm a été utilisée pour bâtir la courbe d'étalonnage permettant de quantifier la concentration d'or. Cette raie d'or a pour intérêt d'être la plus intense dans la gamme spectrale des spectromètres utilisés et de peu interférer avec les raies des autres éléments chimiques présents dans nos échantillons miniers, comme il est spécifié dans la référence [3]. Un bon rapport signal/bruit de cette raie d'or a été obtenu avec un temps d'intégration de 10  $\mu\text{s}$  et un délai avant la prise des mesures de 2  $\mu\text{s}$ .

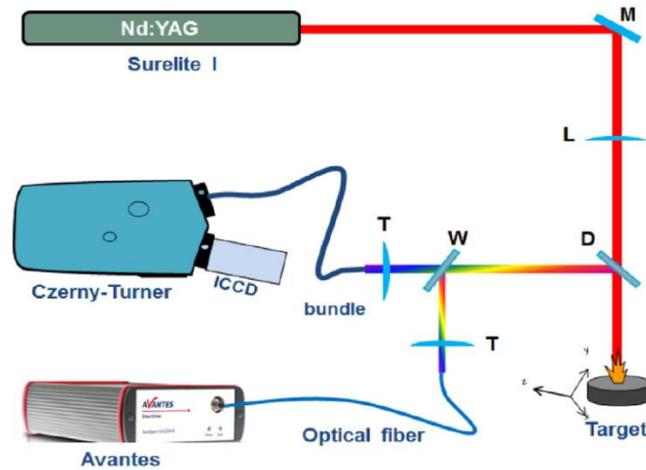


Figure 1. 3 Dispositif expérimental LIBS utilisé dans le cadre de la quantification de la concentration d'or dans les roches minières. [3]

Dans le but d'établir le modèle d'étalonnage, les mesures ont d'abord été prises sur des échantillons de calibration. Ces échantillons étaient sous forme de poudres compressées contenant des concentrations d'or quasi-uniformes. La taille des grains de la poudre est de l'ordre de  $1\ \mu\text{m}$ , soit beaucoup plus petite que le diamètre du point focal du laser ( $600\ \mu\text{m}$ ). La mesure de la concentration d'or dans ces échantillons a été faite par activation neutronique. Cinq types d'échantillons ont été utilisés, qui diffèrent entre eux par leur teneur en fer, silice et soufre :

- Les échantillons Quartz-Chlorite (QC) ont été dopés en or avec des concentrations allant de 0.08 à 296 ppm. Comme la répartition de l'or dans les roches apparaît sous forme de grains, d'un tir laser à un autre on peut obtenir des différences significatives en matière de signal LIBS. Si l'impulsion laser tombait sur une région où la concentration d'or est particulièrement élevée, la concentration maximale de prédiction de QC (296 ppm) pourrait ne pas être suffisante. C'est pour cela que nous avons créé dans nos laboratoires les échantillons de calibration WMS-1, JG-3, LDI-1 et PTC-1b, avec des concentrations d'or plus élevées, qui ne pouvaient se trouver facilement dans le commerce. De plus, pour enrichir notre étude, il était nécessaire d'utiliser des échantillons de compositions chimiques différentes, en particulier, en ce qui concerne la teneur en fer.
- Les échantillons WMS-1 et JG-3 ont été dopés en or au moyen de solutions d'or composées d'eau et de 4% de HCl. Les solutions d'or ont été tout d'abord mélangées uniformément avec les échantillons qui étaient sous forme de poudre. Le résultat de ce processus a ensuite été séché à  $60\ ^\circ\text{C}$  pendant 24 heures. Pour améliorer la cohésion, il y a eu un ajout de 4% de polyéthylène puis tout cela a été mélangé dans un malaxeur pendant 30 minutes. Enfin, les poudres des échantillons ont été compressées par une presse automatique à une pression de 0.71 GPa. Pour plus de détails, les étapes de fabrication des échantillons sont décrites dans la référence [4]. A

l'issu de ce processus, les échantillons WMS-1 avaient des concentrations d'or allant de 0.28 à 891 ppm (déterminées par les mesures d'un laboratoire certifié).

- Les échantillons JG-3 avaient une concentration en or de 0 à 932 ppm.
- Les échantillons PTC-1b et LDI-1 ont été dopée de la même manière que WMS-1 et JG-3 mais avec des solutions d'or ne contenant pas d'acide et sans polyéthylène pour minimiser l'altération et la contamination des échantillons. Les échantillons PTC-1b avaient une concentration d'or prévue de 25 à 500 ppm.
- Les échantillons LDI-1 avaient une concentration d'or prévue de 0 à 500 ppm.

Les concentrations en Si, S, Fe et Au des matériaux de calibration sont résumées dans le tableau 1.

QC		
[Si] 16.9%	[S] 0.84%	[Fe] 4.98%
[Au] 0.08, 5.2, 11.8, 31, 58, 119, 148, 198, 247, 296 ppm		
WMS-1		
[Si] 2.15%	[S] 33%	[Fe] 50,36%
[Au] 0.28, 23, 40, 92, 175, 421, 891 ppm		
JG-3		
[Si] 31.44%	[S] 0.006%	[Fe] 2.58%
[Au] 0, 24, 47, 85, 187, 458, 932 ppm		
PTC-1b		
[Si] 2.468%	[S] 29.95%	[Fe] 36.78%
[Au] 25, 50, 100, 200, 500 ppm		
LDI-1		
[Si] 22.79%	[S] 0.12%	[Fe] 5.38%
[Au] 0, 25, 50, 100, 200, 500 ppm		

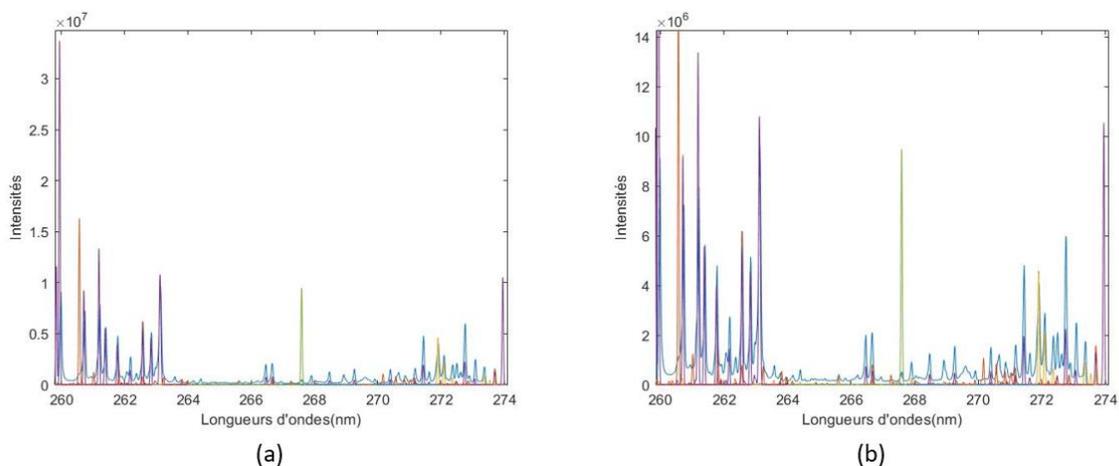
**Tableau 1. 1 Échantillons de calibration utilisées ainsi que leurs compositions élémentaires.**

N'ayant pas participé aux expériences, nous avons essayé de comprendre le spectre prélevé au cours de celles-ci. Dans la référence [11] la température et la densité électronique de WMS-1 avaient pour valeurs respectives 0.5407 eV et  $1.57 \times 10^{16} \text{ cm}^{-3}$ . Nous avons simulé le spectre expérimental avec NIST database LIBS [16]. Nous savions que pour un temps d'intégration de 10  $\mu\text{s}$ , le plasma LIBS ne peut pas être caractérisé par une valeur unique de température et de densité électronique. C'est pourquoi nous avons choisi de prendre 12 valeurs de températures et de densités électroniques représentant la variation de ces paramètres au cours du temps lors de la relaxation du plasma. Le choix de ces valeurs a été fait de façon à ce que la température et la densité électronique diminuent de manière linéaire. Cette approximation permet d'avoir un modèle simpliste qui borne grossièrement les paramètres plasmas. Ces valeurs sont décrites dans le tableau 1. 2.

Température (eV)	Densité électronique x 10 <sup>16</sup> (cm <sup>-3</sup> )
0.5286	2.9286
0.5715	3.5715
0.6143	4.2143
0.6572	4.8572
0.7	5.5000
0.7429	6.1429
0.7857	6.7857
0.8286	7.4286
0.8714	8.0714
0.9143	8.7143
0.9571	9.3571
1	10

**Tableau 1. 2 Valeurs de températures et de densités utilisées pour la simulation du spectre NIST.**

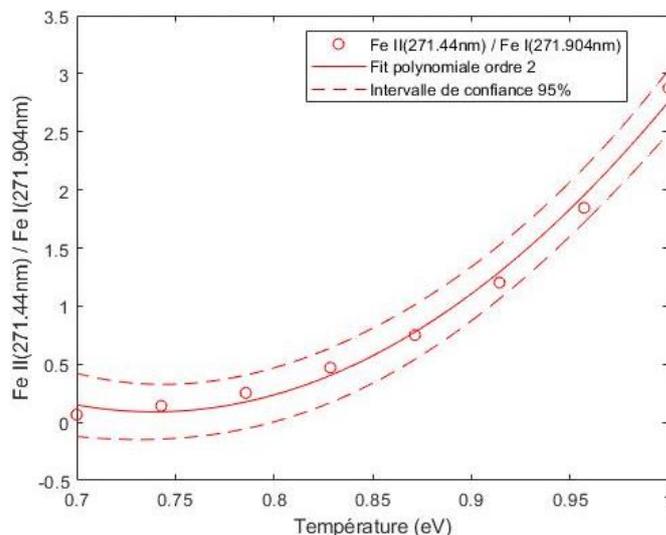
Pour simuler un spectre dans NIST-LIBS [16] nous avons besoin de connaître sa température, sa densité électronique mais aussi sa composition chimique. Nous connaissons la composition chimique de WMS-1, qui était majoritairement composé de fer et de soufre. Les valeurs de températures et de densités utilisées sont présentées dans le tableau 1. 2. Pour obtenir le spectre final de la figure 1. 4 nous avons sommé sans attribuer de poids relatifs tous les spectres simulés de valeur différentes de températures et de densités électroniques. Au total 12 spectres ont été sommés. Sur ce spectre nous retrouvons majoritairement les mêmes pics que ceux qui sont présents sur le spectre expérimental, ce qui nous assure que les éléments implémentés dans la simulation et leurs concentrations sont réalistes.



**Figure 1. 4 Comparaison du spectre simulé par NIST data base LIBS et du spectre expérimental WMS-1 à 500 ppm (en bleu) (a) avec zoom et (b) sans zoom. Les pics de Fe II sont en violet, les pics de Fe I sont en jaune, les pics de Mn II sont en rouge et les pics d’Au I sont en vert.**

La température du plasma pour WMS-1 a été trouvée par la méthode du diagramme de Boltzmann dans la référence [11]. Nous voulions comparer cette valeur au résultat obtenu par notre simulation. Pour cela nous avons calculé le rapport des raies Fe II à 271.44 nm sur Fe I à 271.904 nm du spectre expérimental qui vaut 1.08. Ce rapport a aussi été calculé pour les 12 spectres obtenus par simulation. Ces valeurs ont

été reportées sur la figure 1. 5. On notera qu'en dessous de 0.7 eV la valeur du rapport, restant la même, nous l'avons retirée de la figure 1. 5.



**Figure 1. 5** Variation du rapport Fe II 271.44 nm sur Fe I 271.904 nm en fonction de la température.

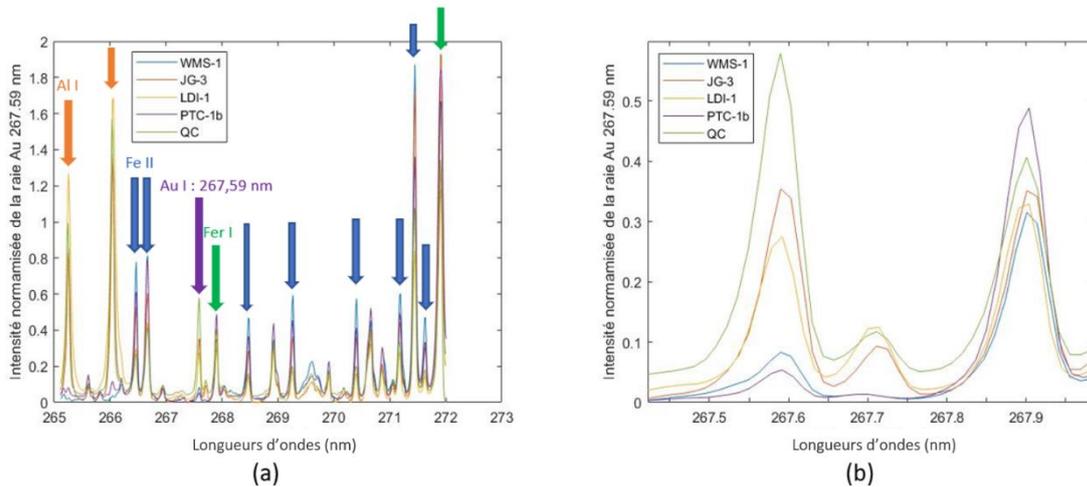
Une fois le rapport 1.08 reporté sur la figure 1.5, nous obtenons une température égale à 0.89 eV. Cette température est bien plus grande que celle trouvée dans la référence [11]. Cette différence se retrouve aussi dans la référence [13]. Il est mentionné que cet écart est dû aux espèces excitées qui ne respectent pas l'équilibre thermodynamique local en raison du caractère non-maxwellien de la distribution en énergie des électrons..

### 1.3 Étude univariée

La construction de la courbe d'étalonnage permet de prédire ensuite la concentration en analyte d'intérêt de n'importe quel échantillon inconnu. Dans le cas présent elle consiste à tracer l'intensité nette (intensité maximale moins le fond continu) de la raie d'or Au I à 267.59 nm des échantillons de référence en fonction de leur teneur en or.

La normalisation des données est une étape indispensable pour établir la courbe d'étalonnage. Elle permet de compenser les effets de la composition chimique de la matrice étudiée sur l'intensité de la raie d'or ainsi que des paramètres plasmas qui peuvent changer légèrement d'une mesure à une autre (fluence laser, rugosité de la surface, présence d'humidité, ...). Il y a plusieurs manières de normaliser les données provenant de mesures LIBS. Dans cette étude nous avons choisi de normaliser par rapport à l'aire totale du spectre. La figure 1.6 (a) montre les spectres normalisés des échantillons sur la gamme des longueurs d'onde sélectionnée près de la raie d'or à 267.59 nm. Sur la figure 1.6 (b) on voit que les spectres de WMS-1 et PTC-1b ont des valeurs en intensité pour la raie d'or moins grandes que celles de JG-3, LDI-1 et QC. Cela s'explique par la différence de la concentration en fer. En effet le fer est un

élément très émissif dans la gamme spectrale observée qui comporte de nombreuses raies. Ainsi l'aire spectrale de WMS-1 et PTC-1b, des matériaux riches en fer, étant plus grande, après normalisation l'intensité de la raie d'or est plus petite. Avant normalisation, chaque spectre a été moyenné sur 500 spectres obtenus par autant de tirs laser, ce qui permet de réduire le bruit par un facteur approximativement égal à la racine carrée de 500.



**Figure 1. 6 (a) Spectres moyens normalisés par rapport à l'aire spectrale de tous les échantillons de référence à une concentration de 200 ppm dans la gamme spectrale 265 – 271 nm. Les flèches bleues représentent les raies de Fer II, les flèches oranges les raies d'Al I, les flèches vertes les raies de Fer I et la flèche violette la raie d'Au I à 267.59 nm (b) Zoom sur la raie d'or à 267.59 nm et la raie de Fe à 267.9 nm.**

Suite à ces étapes de prétraitement, les courbes d'étalonnage de tous les échantillons de référence ont été tracées. L'étude univariée s'arrête à 500 ppm car la majorité de nos échantillons de référence sont bornés par cette valeur. Dans l'objectif d'évaluer les modèles d'étalonnage, nous avons choisi comme figure de mérite le coefficient de détermination  $R^2$ , l'erreur quadratique moyenne sur les échantillons de calibration RMSEC (root mean square error calibration) et de validation RMSECV (root mean square error cross-validation) obtenues par validation croisée « leave one out ». Cette méthode consiste à retirer un échantillon du lot de calibration pour le transférer dans le lot de validation. La courbe d'étalonnage est tracée avec le nouveau lot de calibration puis testée avec l'échantillon de validation choisi. Cela est répété autant de fois qu'il y a d'échantillons de validation. Une fois tous les échantillons de validation testés on calcule RMSECV. Ces quantités proviennent de la référence [14] et s'expriment respectivement

$$R^2 = \frac{\sum_c y_c - \bar{y}}{\sum_i y_i - \bar{y}}$$

où les  $y_c$  représentent les intensités obtenues par la régression de la concentration  $x_i$ , les  $y_i$  les intensités correspondantes de la raie d'or et  $\bar{y}$  leur valeur moyenne;

$$RMSEC = \sqrt{\frac{\sum_{p=1}^{N_c} (c_p - \hat{c}_p)^2}{N_c}}$$

$$RMSECV = \sqrt{\frac{\sum_{v=1}^{N_v} (c_v - \hat{c}_v)^2}{N_p}}$$

Dans ces expressions,  $c_p$  et  $c_v$  représentent la concentration de référence,  $\hat{c}_p$  et  $\hat{c}_v$  la concentration prédite par la fonction de fit choisie,  $N_c$  le nombre d'échantillons de référence et  $N_p$  le nombre d'échantillons de validation.

Si la valeur de RMSEC est proche de la valeur RMSECV, alors on peut estimer que le modèle est bon. Bien entendu, si ces deux valeurs sont grandes, le modèle reste mauvais.

Le signal LIBS de la raie d'intérêt  $\hat{y}_0$  de concentration  $x_0$  est compris dans un intervalle de confiance à 95% borné par  $\widehat{y_{0,min}}$  et  $\widehat{y_{0,max}}$ .

$$\widehat{y_{0,min}} = \text{fit choisi} - t \cdot s_{\hat{y}_0}$$

$$\widehat{y_{0,max}} = \text{fit choisi} + t \cdot s_{\hat{y}_0}$$

où  $t$  est le coefficient de Student et  $s_{\hat{y}_0}$  est l'écart type de  $\hat{y}_0$  défini par :

$$s_{\hat{y}_0} = s_e \left[ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]^{\frac{1}{2}}$$

Ici,  $s_e$  est l'écart type des résidus avec pour formule :

$$s_e = \left[ \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 3} \right]^{\frac{1}{2}}$$

où  $\hat{y}_i$  est la concentration prédite par le modèle.

### 1.3.1 Courbes d'étalonnage des échantillons de référence

Les courbes d'étalonnage des échantillons du lot de calibration sont représentées sur la figure 1. 7. On observe sur celles-ci que tous les échantillons, excepté QC, ont une tendance quadratique. Cela est dû à un phénomène physique qui arrive souvent en LIBS : l'auto-absorption des raies. Ce phénomène survient lorsque les raies émises par les atomes d'or dans le plasma sont réabsorbées par les atomes d'or dans l'état fondamental en périphérie du plasma où la température est plus basse. Il se produit lorsque la concentration d'or est telle que la section efficace de réabsorption devient significative. On constate que la raie d'or à 267.59 nm est auto-absorbée pour les plus hautes concentrations considérées. On distingue deux groupes de courbes sur cette figure : le groupe de courbe JG-3, QC et LDI-1 qui sont des

échantillons pauvres en fer et le groupe de courbe WMS-1 et PTC-1b qui sont des échantillons plutôt riches en fer. C'est la normalisation par rapport à l'aire spectrale qui crée cette démarcation.

Dans ce qui suit nous évaluons les modèles univariés, c'est-à-dire dépendant seulement de l'intensité de la raie d'or normalisée, pour chaque lot d'échantillons de référence.

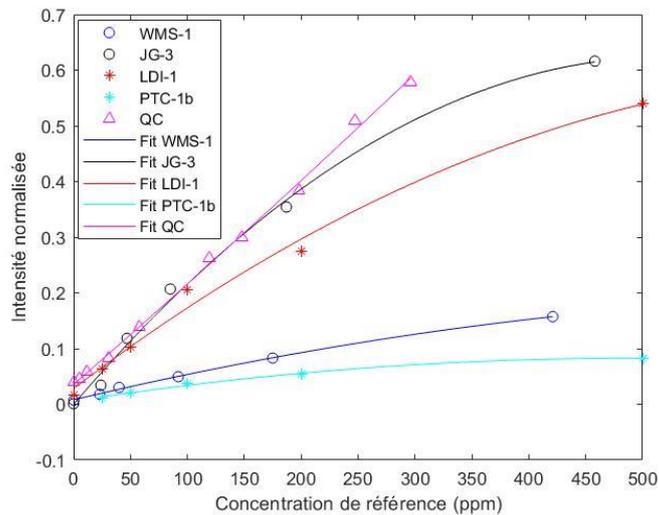


Figure 1. 7 Comparaison des courbes de calibration de tous les lots de référence. Le background se situe à 267,37 nm.

### 1.3.1.1 Échantillon de calibration WMS-1

La courbe de calibration de WMS-1 est tracée sur la figure 1. 8. On voit sur celle-ci que le meilleur modèle est quadratique avec  $R^2 = 0.9991$  et  $RMSEC = 3$  ppm.

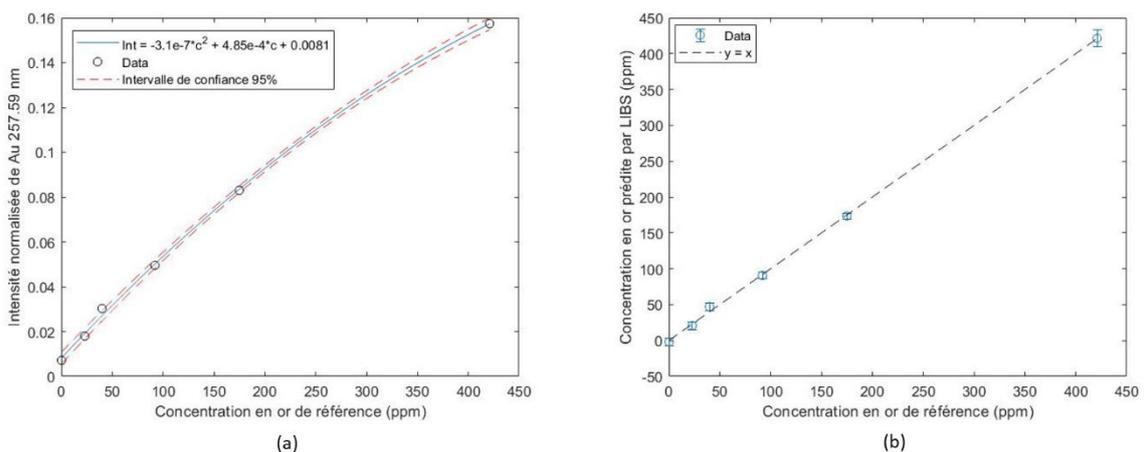


Figure 1. 8 (a) Courbe de calibration des échantillons WMS-1 normalisée par rapport à l'aire de l'ensemble du domaine spectral. (b) Évaluation du modèle de prédiction par rapport aux concentrations de référence.

D'après la figure 1. 8 et le tableau 1. 3, on constate que les valeurs prédites par ce modèle sont bonnes. Cela se remarque à l'intervalle de confiance qui est proche de la courbe d'étalonnage. Les incertitudes ont aussi été calculées à l'aide des intervalles de confiance. Suite à l'utilisation de la méthode de validation croisée nous trouvons que  $RMSECV = 23$  ppm. Cette valeur est assez grande, néanmoins il faut réaliser que nous ne disposons pas d'un assez grand nombre d'échantillons dans le lot de calibration WMS-1 ce qui affecte  $RMSECV$  car lorsqu'on retire un point du lot de calibration pour le mettre dans le lot de validation, la courbe de calibration s'en trouve considérablement modifiée. Le  $RMSECV$  prend ici une valeur qui est discutable en raison du petit nombre de données, et qui sera également discutable pour JG-3, LDI-1 et PTC-1b.

<b>WMS-1</b>	
Concentration de référence (ppm)	Concentration prédite par LIBS (ppm)
0.28	$-2 \pm 5$
23	$21 \pm 5$
40	$47 \pm 5$
92	$91 \pm 5$
175	$174 \pm 4$
421	$421 \pm 11$

**Tableau 1. 3 Comparaison des valeurs de concentration de référence aux valeurs prédites par le modèle pour WMS-1.**

Les résultats obtenus pour WMS-1 sont similaires aux résultats obtenus pour les échantillons JG-3, PTC-1b et LDI-1. Ils sont résumés dans le tableau 1. 4.

$R^2$	RMSEC (ppm)	RMSECV (ppm)
<b>WMS-1</b>		
0.9991	3	23
<b>JG-3</b>		
0.9956	7	29
<b>PTC-1b</b>		
0.9926	10	40
<b>LDI-1</b>		
0.9909	13	35

**Tableau 1. 4 Comparaison des valeurs  $R^2$ , de RMSEC et de RMSECV des courbes de calibration pour les différents échantillons de référence.**

La courbe de calibration des échantillons QC est différente des autres échantillons, et est décrite ci-dessous.

### 1.3.1.2 Échantillon de référence QC

Sur la figure 1. 9, on a l'impression que l'intensité de la raie d'or augmente linéairement avec la concentration d'or. L'effet d'auto-absorption ne s'est donc quasiment pas manifestée sur ce lot de calibration pour des raisons qui nous échappent. On a tout de même procédé à un fit polynomial donnant un coefficient de détermination meilleur qu'un fit linéaire de valeur 0.9977. Le coefficient de détermination ayant une valeur très proche de 1, l'erreur quadratique moyenne sur les échantillons de calibration est égale à 5 ppm. Lorsque qu'on procède à une validation par la méthode de la validation croisée, on obtient une erreur quadratique moyenne sur le lot de validation de 7 ppm. Cette valeur est proche de l'erreur quadratique moyenne du lot de calibration car il y a davantage de points dans le lot de calibration pour tracer la courbe d'étalonnage. Ainsi, la perte d'un échantillon de calibration dans son lot n'a pas un impact significatif sur la courbe d'étalonnage et la prédiction de l'échantillon de validation est relativement juste.

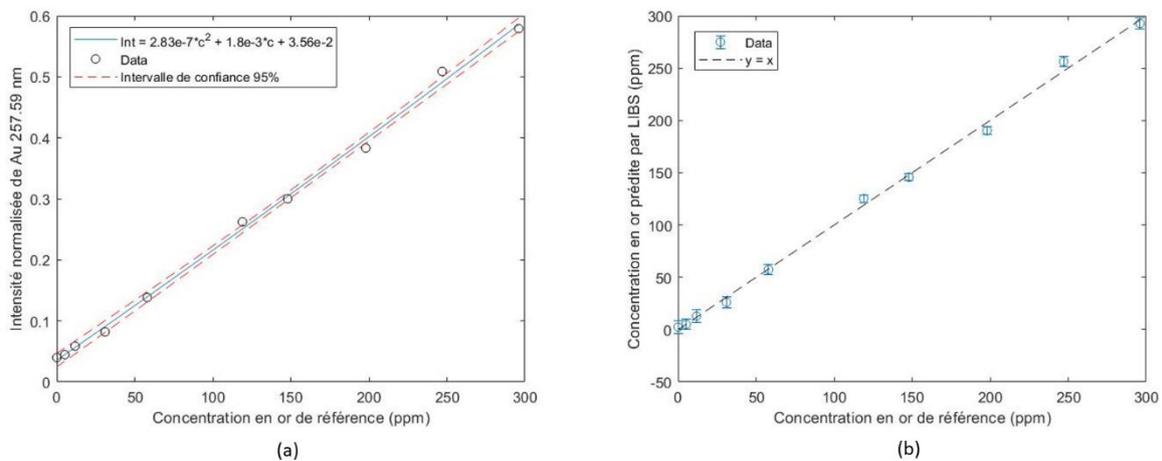


Figure 1. 9 (a) Courbe de calibration des échantillons QC normalisés par rapport à l'aire de l'ensemble du domaine spectral. (b) Évaluation du modèle de prédiction par rapport aux concentrations de référence.

QC	
Concentration de référence (ppm)	Concentration prédite par LIBS (ppm)
0.08	$2 \pm 4$
5.2	$5 \pm 5$
11.8	$13 \pm 7$
31	$26 \pm 31$
58	$57 \pm 4$
119	$125 \pm 10$
148	$146 \pm 1$
198	$190 \pm 4$
247	$256 \pm 14$
296	$293 \pm 2$

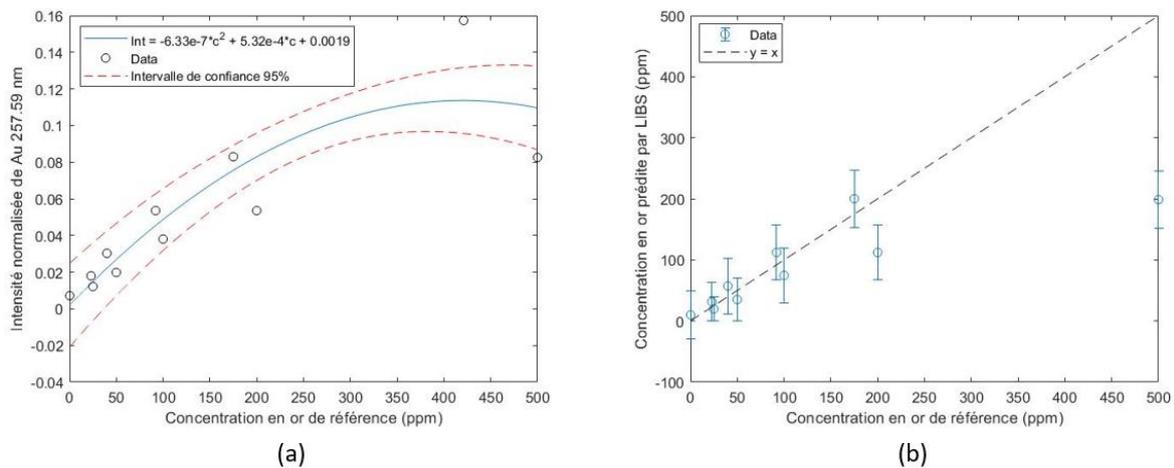
Tableau 1. 5 Comparaison des valeurs de concentration de référence aux valeurs prédites par le modèle pour QC.

Dans l'objectif de tracer les courbes de calibration avec plus de points nous avons décidé de regrouper les échantillons de référence qui ont une composition chimique similaire. Deux courbes de calibration seront tracées :

- La courbe de calibration créée avec WMS-1 et PTC-1b qui sont des échantillons riches en fer.
- La courbe de calibration JG-3, LDI-1 et QC qui sont des échantillons pauvres en fer.

### 1.3.1.3 Courbe de calibration avec WMS-1 et PTC-1b

La figure 1. 10 (a) montre l'allure de la courbe de calibration tracée avec les points appartenant aux échantillons WMS-1 et PTC-1b. On remarque que le phénomène d'auto-absorption est toujours présent et que le tracé de la courbe est mauvais, le coefficient de détermination vaut 0.7991 et RMSECV vaut 44 ppm. Si on regarde la figure 1. 10 (b) on voit que les points sont peu alignés avec la courbe  $y = x$  et les barres d'erreur sont grandes. Cela confirme que les échantillons de calibration sont mal prédits par ce modèle.

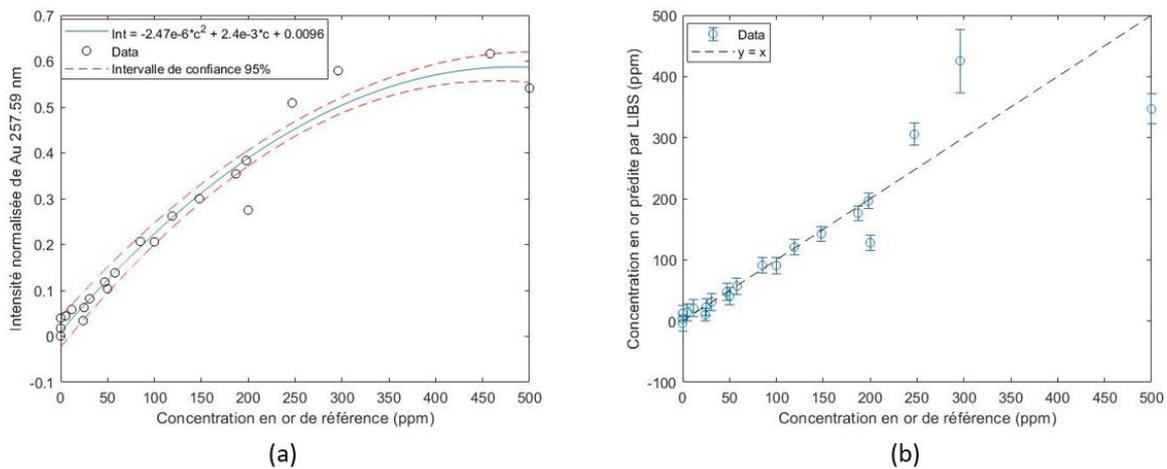


**Figure 1. 10 (a) Courbe de calibration des échantillons WMS-1 et PTC-1b normalisée par rapport à l'aire de l'ensemble du domaine spectral. (b) Évaluation du modèle de prédiction par rapport aux concentrations de référence.**

La réunion des échantillons PTC-1b et WMS-1, qui sont des échantillons riches en fer, ne donne pas de bons résultats. Pourtant étant de même nature, les effets de matrices ne devraient pas être aussi importants. Il n'est donc pas si aisé de réunir deux échantillons de calibration pour former une nouvelle courbe de calibration.

### 1.3.1.4 Courbe de calibration avec JG-3, LDI-1 et QC

Le rassemblement des échantillons de référence JG-3, LDI-1 et QC permet d'avoir une courbe de calibration avec 22 points (échantillons). On remarque sur la figure 1. 11 (a) le phénomène d'auto-absorption. Le tracé de la courbe est, cette fois-ci, bon puisque le coefficient détermination vaut 0.9634. L'intervalle de confiance à 95% est proche de la courbe ce qui manifeste une bonne précision. Les points de la figure 1. 11 (b) suivent bien la droite  $y = x$  et les barres d'erreurs sont généralement petites. On en conclut que les échantillons de calibration sont bien prédits par ce modèle quantitatif.



**Figure 1. 11 (a) Courbe de calibration des échantillons JG-3, LDI-1 et QC normalisée par rapport à l'aire de l'ensemble du domaine spectral. (b) Évaluation du modèle de prédiction par rapport aux concentrations de référence.**

On remarque sur le tableau 1. 6 que les valeurs sont en général bien prédites. La valeur des incertitudes est grande par rapport aux échantillons qui ont une faible concentration de référence mais est petite pour les échantillons qui ont une grande concentration de référence.

<b>JG-3</b>	
Concentration de référence (ppm)	Concentration prédite par LIBS (ppm)
0	$-4 \pm 13$
24	$23 \pm 14$
47	$41 \pm 14$
85	$91 \pm 13$
187	$128 \pm 12$
458	$432 \pm 15$
<b>LDI-1</b>	
0	$3 \pm 3$
25	$31 \pm 14$
50	$57 \pm 14$
100	$120 \pm 13$
200	$305 \pm 18$
500	$347 \pm 25$
<b>QC</b>	
0.08	$13 \pm 13$
0.8	$15 \pm 14$
5.2	$21 \pm 14$

11.8	$10 \pm 10$
31	$48 \pm 14$
58	$91 \pm 13$
119	$143 \pm 12$
148	$196 \pm 12$
198	$177 \pm 12$
247	$426 \pm 52$

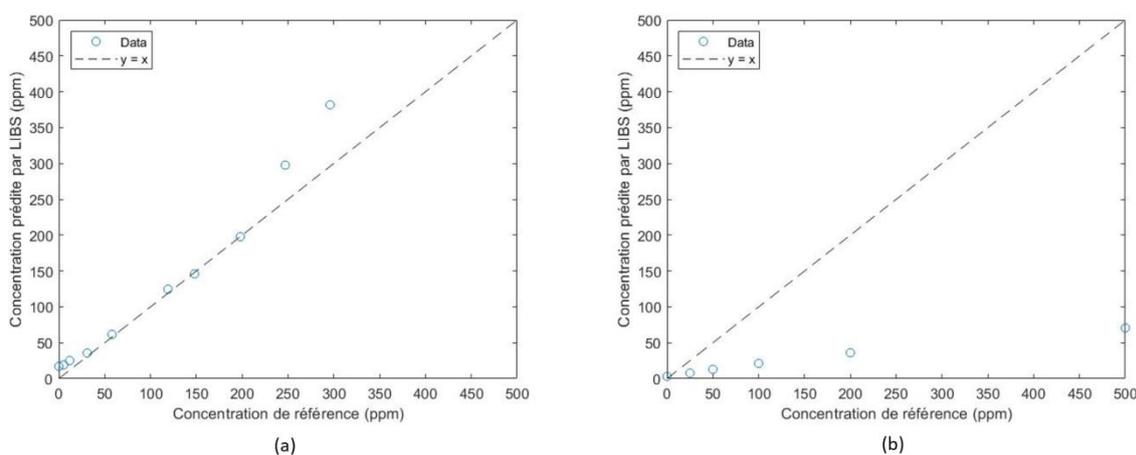
**Tableau 1. 6** Comparaison des valeurs de concentration de référence avec les valeurs prédites par le modèle JG-3, LDI-1 et QC.

La valeur RMSECV après validation croisée est égale à 32 ppm. Cette valeur est moins grande que pour le modèle créé avec WMS-1 et PTC-1b. La courbe de calibration faite avec les échantillons pauvres en fer est meilleure que la courbe de calibration des échantillons riches en fer.

### 1.3.2 Test des courbes de calibration

Nous avons montré précédemment que la méthode de validation croisée ne permettait pas d'évaluer la qualité de notre modèle d'étalonnage de façon très robuste en raison du petit nombre de données disponibles.

Dans le but d'augmenter le nombre de données à traiter de manière statistique, nous avons choisi de prédire, par le biais d'une courbe d'étalonnage construite à partir d'un lot de calibration présélectionné, la concentration en or d'échantillons issus d'un autre lot et de concentrations connues. Nous avons ainsi utilisé la courbe d'étalonnage construite à partir de JG-3 pour prédire les concentrations en or des échantillons QC et WMS-1.



**Figure 1. 12** (a) Prédiction des concentrations des échantillons QC à l'aide du modèle de calibration JG-3. (b) Prédiction des concentrations des échantillons WMS-1 à l'aide du modèle de calibration JG-3.

On observe sur la figure 1.12 (a) que les échantillons QC sont bien prédits jusqu'à 198 ppm par le modèle JG-3 tandis que les concentrations de ceux qui ont des concentrations plus élevées sont surestimées. C'est la tendance linéaire de la courbe de calibration de QC qui en est responsable. Les deux points correspondant aux concentrations d'or les plus élevées s'éloignent donc de la courbe de calibration de JG-3 qui est polynomiale, ce qui baisse la qualité de la prédiction.

Pour les échantillons WMS-1, les valeurs de concentration prédites avec la courbe d'étalonnage de JG-3 sont très largement sous-estimées. Cela s'explique par leur différence de composition chimique. En effet, JG-3 est un échantillon pauvre en fer, de même que QC, ainsi la prédiction de QC par rapport à JG-3 est acceptable. Au contraire, WMS-1 étant un échantillon riche en fer, il est très mal prédit par la courbe d'étalonnage de JG-3 à cause de la normalisation à l'aire totale du spectre qui est largement influencée par les nombreuses raies atomiques relatives au fer. Ce résultat confirme qu'un modèle d'étalonnage univarié ne convient que si les échantillons sont de matrices semblables.

Il est donc nécessaire d'établir des critères permettant de sélectionner le modèle qui convient le mieux à l'analyse quantitative d'un échantillon quelconque. Ces critères seront abordés et traités dans le chapitre 4. Dans l'idéal nous voudrions disposer d'une méthode de quantification qui permet de prédire la concentration en or d'un échantillon sans se préoccuper explicitement de sa composition chimique. Des outils chimiométriques permettent de prendre en compte la composition chimique dans l'élaboration d'une courbe de calibration unique. On les qualifie de méthodes multivariées car elles prennent en compte dans leur fonctionnement plusieurs raies spectrales et donc implicitement la composition chimique. Les outils chimiométriques que nous avons exploités dans le cadre de notre étude sont présentés dans le chapitre 2 et appliqués au traitement des spectres LIBS des échantillons de roches contenant de l'or dans le chapitre 3.

## Chapitre 2 : Outils de chimiométrie

### 2.1 Analyse en Composantes Principales (ACP)

L'analyse en composantes principales est une méthode d'analyse qualitative non supervisée (sans phase d'apprentissage). Elle permet de décrire une matrice de données  $X$  (intensités spectrales mesurées) de dimension  $(i, j)$  où  $i$  représente le nombre de mesures (nombre de spectres) et  $j$  le nombre de variables (nombre de longueurs d'ondes par spectre) dans un espace de dimension réduite  $n < j$  avec peu de perte d'information. Dans cet espace à  $n$  dimensions, la matrice  $X$  apparaît comme un ensemble de  $i$  points où chaque point représente un spectre et où la distance entre deux points constitue une mesure du degré de ressemblance des deux spectres. On peut alors évaluer le degré de similitude entre des spectres ainsi que de la composition chimique des échantillons dont ils sont issus et distinguer les spectres qui se démarquent grandement des autres, correspondant à des points aberrants dans l'espace en  $n$  dimensions.

Cette technique se base sur le calcul de la matrice de covariance. Si on prend une matrice de données  $X$  contenant  $i$  spectres et  $j$  longueurs d'onde (de dimension  $(i, j)$ ) dont on a soustrait à chaque colonne la valeur moyenne de la colonne, la matrice de covariance est :

$$\text{Cov}(X) = {}^tX \cdot X$$

Elle a pour propriété d'être une matrice carrée symétrique de dimension  $(j, j)$  avec pour valeurs sur la diagonale la variance de chaque variable, c'est-à-dire des intensités pour chaque longueur d'onde.

Le signe des valeurs hors de la diagonale nous renseigne sur les corrélations entre les variables (i.e. entre les intensités à deux longueurs d'ondes différentes). Une variance positive montre qu'il y a une corrélation positive entre les variables (les deux variables croissent ou décroissent en même temps). Au contraire, une corrélation négative montre que les variables sont anti-corrélées (quand une variable augmente, l'autre diminue).

Les  $j$  vecteurs propres de la matrice de covariance sont les composantes principales aussi appelées loadings. Par définition, les composantes principales sont orthogonales deux à deux et normalisées à 1. On peut démontrer qu'elles correspondent aux directions pour lesquelles la variance globale est maximale. Leurs valeurs propres sont associées à la quantité d'informations qu'elles retiennent sur la variance totale. Par exemple sur la figure 2. 1, la première composante représente 83.13% de la variance totale alors que la deuxième composante principale représente 9.07% de la variance totale. On peut ainsi visualiser sur ce plan 92.2% de la variance totale. Ces composantes principales peuvent souvent être reliées à la composition chimique des échantillons dont les spectres sont issus. Ici la première composante principale différencie les échantillons rocheux selon leur teneur en fer.

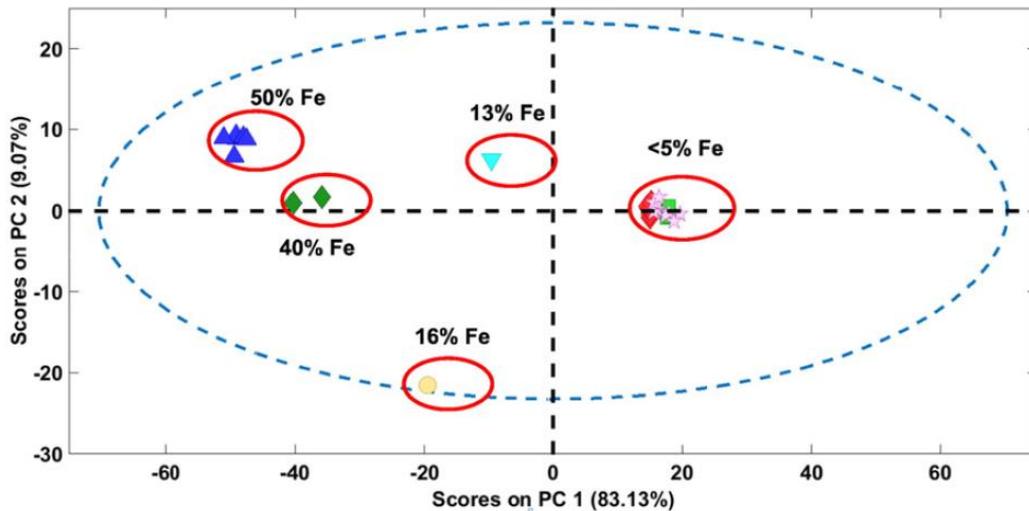


Figure 2. 1 ACP appliquée sur des données LIBS provenant de [3].

Les coordonnées des spectres dans la nouvelle base sont appelées scores. Cette décomposition se décrit comme ci :

$$X = T \cdot P + E$$

où T représente la matrice des scores ( $i, j$ ), P la matrice des composantes principales ( $j, j$ ) et E les résidus de la décomposition.

Cependant la difficulté de cette technique peut venir de la diagonalisation de matrices de grandes dimensions car cela demande des ressources informatiques conséquentes. Pour y remédier on utilise un algorithme itératif nommé NIPALS (Non Linear Partial Least Square). Il peut être arrêté dès que le nombre de composantes principales souhaitées pour bâtir un modèle multivarié est atteint. Il est décrit dans la référence [10].

L'ACP, bien que pratique pour obtenir des informations sur le lot d'échantillons dont nous disposons ne permet pas de quantifier leur concentration en or. La méthode de la régression aux moindres carrés partiels (PLS) le permet et sa description figure ci-dessous.

## 2.2 Régression aux moindres carrés partiels (PLS)

La régression aux moindres carrés partiels, appelée PLS pour « Partial Least Square », est une méthode supervisée quantitative. On dit que cette méthode est supervisée car elle repose sur une phase d'étalonnage. Il faut là encore calculer la matrice de covariance, ce qui la rend similaire à l'ACP. Elle donne en sortie la concentration de l'analyte souhaitée. La principale différence entre PLS et ACP est la phase d'apprentissage de PLS. En effet, la méthode PLS a pour objectif de trouver les scores qui ont la

plus grande covariance entre les spectres et la concentration de l'élément de référence. Cela permet de minimiser l'impact des raies qui varient indépendamment de la concentration.

La première étape de PLS consiste à décomposer la matrice des spectres et le vecteur des concentrations de référence comme ci :

$$X = T \cdot P + E \quad (2.0)$$

$$C = T \cdot p + e \quad (2.1)$$

où  $X(i, j)$  est la matrice des spectres,  $T(j, A)$  la matrice des scores avec  $A$  le nombre de colonnes appelées composantes ou facteurs,  $P(i, A)$  la matrice des vecteurs propres de la nouvelle base dans laquelle sont projetés les spectres,  $E$  et  $e$  les résidus de la décomposition,  $C(j, 1)$  le vecteur des concentrations et  $p(A, 1)$  le vecteur issu de la régression de  $C$  sur  $T$ . Le nombre de composantes  $A$  est toujours plus petit que le nombre de longueurs d'ondes  $j$  car, comme pour l'ACP, cette méthode permet la compression des données tout en minimisant la perte d'informations. Les formules (2.0) et (2.1) indiquent que la matrice des scores  $T(j, A)$  est la même pour les spectres et les concentrations. Il y a donc une corrélation de la variance associée aux spectres et aux concentrations permettant ainsi de prédire la concentration d'un spectre inconnu.

Pour appliquer PLS il faut centrer sur la moyenne la matrice des données  $X$  ainsi que le vecteur des concentrations de référence  $C$ . L'algorithme est itératif et se répète autant de fois qu'il y a de composantes choisies  $k = 1, \dots, K$ .

L'algorithme commence par le calcul de la quantité  $a$  qui rend le vecteur des loadings  $w_k$  égal à 1 :

$$a = \frac{1}{\sqrt{(C_{k-1}^T \cdot X_{k-1} \cdot X_{k-1}^T \cdot C_{k-1})}}$$

$$w_k = a \cdot X_{k-1}^T \cdot x_{k-1}$$

Cela s'enchaîne par le calcul des scores  $z_k$  et des loadings  $v_k$  de la matrice des spectres ainsi que le calcul des loadings de la concentration  $q_k$  :

$$z_k = X_{k-1} \cdot w_k$$

$$v_k = X_{k-1}^T \cdot \frac{z_k}{z_k^T \cdot z_k}$$

$$q_k = C_{k-1} \cdot \frac{z_k}{z_k^T \cdot z_k}$$

Puis on supprime l'effet de cette composante des données :

$$X_k = X_{k-1} - z_k \cdot v_k^T$$

$$C_k = C_{k-1} - z_k \cdot q_k$$

Le vecteur de régression  $b$  et le coefficient  $b_0$  sont utilisés pour calculer la concentration  $c$  des échantillons sont calculés comme ci :

$$b = W \cdot (Z^T \cdot W)^{-1} \cdot Q$$

$$b_0 = \bar{C} - x \cdot b$$

où  $x$  est le vecteur de la moyenne des colonnes composant  $X$  non centrée sur la moyenne.

Finalement la concentration d'un échantillon quelconque ayant un vecteur d'intensités  $x_t(1,j)$  est donnée par :

$$c = b_0 + b^T \cdot x_t$$

Les loadings permettent l'évaluation des poids de chaque longueur d'onde sur la prédiction des concentrations. Si pour une certaine longueur d'onde, la composante de  $b$  est grande alors elle comptera beaucoup dans la prédiction et si la composante de  $b$  a une valeur proche de 0 alors son impact sur la prédiction sera négligeable.

Le calcul des scores est aussi expliqué dans les références [5, 6, 7]. Le nombre de composantes principales est choisi par la méthode de validation croisée interne ou externe, ce qui sera expliquée dans la section 2.3.

Le point faible de cette méthode c'est qu'elle n'offre pas toujours de bons résultats, notamment lorsque le problème étudié est fortement non linéaire. Les réseaux de neurones artificiels étant un algorithme non linéaire permettent de résoudre ce genre de problèmes.

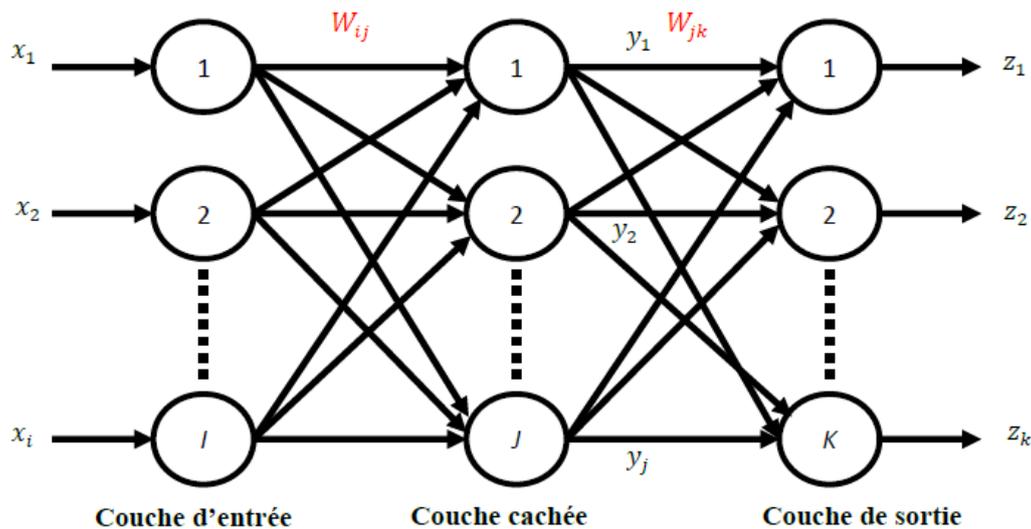
### 2.3 Réseaux de neurones artificiels (ANN)

Les réseaux de neurones artificiels sont une méthode inspirée du fonctionnement du cerveau humain (mémoire, vitesse d'apprentissage, ...) permettant de résoudre des problèmes complexes. Son acronyme ANN vient de « artificial neural network ».

Elle dispose de nombreux avantages qui la rendent extrêmement robuste. Premièrement, sa capacité à fournir selon le contexte une réponse linéaire (comme PCA et PLS) ou non linéaire est très intéressante dans la mesure où généralement les données évoluent de façon non linéaire, ce qui peut rendre les

modèles linéaires inefficaces. Deuxièmement, contrairement à PLS par exemple, il n'y a pas besoin de connaître la relation mathématique qui relie les données d'entrée aux données de sortie pour construire un modèle d'étalonnage. Troisièmement, le réseau de neurones s'adapte à un jeu de données différent de celui qui a servi pour l'apprentissage. Il peut donc fournir une réponse assez précise à un problème éloigné de ce qu'il a appris auparavant. Dans notre cas, si un échantillon test ne ressemble pas aux échantillons de calibration, l'ANN est en mesure de prédire sa concentration sur la base de ce qu'il a déjà vu. Enfin, on peut interpréter un résultat en sortie de l'ANN avec un certain taux de confiance.

Un réseau de neurones est composé de neurones interconnectés disposés en plusieurs couches (figure 2.2). Ces derniers sont reliés par des poids qui quantifient l'intensité des liens entre le neurone d'une couche et ceux de la couche suivante. Chaque neurone, sauf ceux de la couche d'entrée, possède aussi un seuil d'activation, ou biais, qui lui est propre. Il existe plusieurs types d'architectures pour les réseaux de neurones. Nous allons ici étudier les réseaux de neurones à 3 couches : une couche d'entrée, une couche intermédiaire, aussi appelée couche cachée, et enfin une couche de sortie. La couche d'entrée reçoit les données expérimentales. En général, sur cette couche il y a autant de neurones que de variables (par exemple, les longueurs d'ondes d'un spectre). La couche intermédiaire permet de connecter la couche d'entrée à la couche de sortie. La couche de sortie fournit la réponse du réseau de neurones. Elle contient un neurone si on souhaite analyser un seul élément chimique et plusieurs si on souhaite en étudier plusieurs.



**Figure 2. 2 Réseau de neurones à 3 couches : la couche d'entrée, la couche intermédiaire et la couche de sortie. Les poids  $W_{ij}$  relient les neurones  $i$  de la couche d'entrée ( $i = 1, \dots, I$ ) aux neurones  $j$  de la couche intermédiaire ( $j = 1, \dots, J$ ). Les poids  $W_{jk}$  relient les neurones  $j$  de la couche intermédiaire aux neurones  $k$  la couche de sortie ( $k = 1, \dots, K$ ). [1]**

L'algorithme de calcul de l'ANN est itératif. Les poids et les seuils d'activation sont initialisés aléatoirement de telle sorte que les valeurs de sortie lors de la première itération sont généralement très éloignées des valeurs de référence. L'erreur entre les valeurs de sortie du réseau de neurones et les valeurs de référence est ensuite propagée en sens inverse pour corriger les poids et les seuils d'activation afin de diminuer l'erreur (figure 2.3). Une fois les poids et seuils corrigés partiellement pour l'échantillon 1, on introduit les données de l'échantillon 2 et on recommence. La première itération passe donc en revue tous les échantillons du lot de calibration. Ce processus est répété plusieurs fois jusqu'à ce que l'erreur en sortie atteigne une valeur satisfaisante.

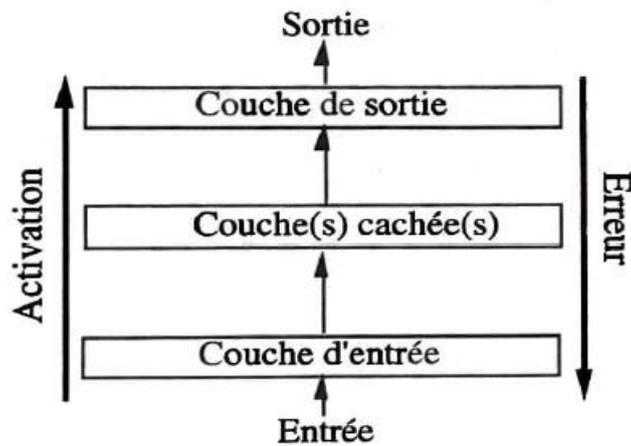


Figure 2. 3 Schéma de fonctionnement du réseau de neurones. L'erreur est propagée dans le sens inverse de calcul du réseau.

### 2.3.1 Transmission du signal

On dénote  $i$  les neurones de la couche d'entrée,  $j$  les neurones de la couche intermédiaire et  $k$  les neurones de la couche de sortie. Chaque valeur entrante  $x_i$  se voit appliquer un poids  $W_{ij}$  qui la relie à un neurone  $j$  de la couche intermédiaire. En entrée, les valeurs des neurones  $j$  de la couche intermédiaire sont données par la somme pondérée :

$$a_j = \sum_i W_{ij} \cdot x_i - \theta_j$$

où  $\theta_j$  est le seuil d'activation du neurone  $j$ . On note que le seuil d'activation peut être traité comme un poids  $W_{0j} = -\theta_j$  avec l'entrée  $x_0 = 1$ .

En sortie, les neurones  $j$  de la couche intermédiaire prennent les valeurs  $f(a_j)$  où  $f$  est la fonction d'activation qui prend des valeurs comprises entre 0 et 1. Il existe plusieurs fonctions d'activation possibles mais la plus utilisée est la fonction sigmoïde

$$f(x) = \frac{1}{1 + e^{-x}}$$

La fonction  $\frac{\tanh(x)+1}{2}$  est aussi fréquemment utilisée.

En entrée des neurones de sortie on a :

$$a_k = \sum_j W_{jk} \cdot a_j - \theta_k$$

Enfin, en sortie, les neurones de sortie prennent les valeurs  $f(a_k)$ .

### 2.3.2 Retro-propagation de l'erreur

Nous avons choisi la méthode dite online pour rétro-propager l'erreur. Sa simplicité de mise en œuvre ainsi que ses corrections fines ont motivé notre choix. En effet, cette méthode consiste à effectuer une rétropropagation de l'erreur pour chaque spectre présenté en entrée du réseau. Comme l'erreur est propagée de la couche de sortie à la couche d'entrée, on commence par corriger les poids de la couche de sortie.

Les nouveaux poids de la couche de sortie (incluant les seuils) sont calculés ainsi :

$$W_{jk}(n+1) = W_{jk}(n) + \eta \cdot \delta_k \cdot f(a_j) + \alpha \cdot [W_{jk}(n) - W_{jk}(n-1)]$$

où  $n$  dénote le numéro de l'itération. Le terme de correction  $\delta_k$  est calculé par

$$\delta_k = (t_k - f(a_k)) \cdot f'(a_k)$$

où  $t_k$  est la valeur de référence du neurone  $k$ . Une forme utile de la dérivée de la fonction sigmoïde est  $f'(x) = f(x) \cdot (1 - f(x))$ . Pour  $n = 0$ , on utilise  $w_{jk}(1) = W_{jk}(0) + \eta \cdot \delta_k \cdot a_j$ .

Les coefficients  $\eta$  et  $\alpha$  sont respectivement la vitesse d'apprentissage et l'effet de mémoire. Ces coefficients sont tous deux des constantes positives comprises entre 0.01 et 0.8 [8,9]. Le choix de leur valeur est important car cela affecte la convergence de l'algorithme. Par exemple, une vitesse d'apprentissage trop grande peut rendre l'algorithme instable, ce qui entraînera l'absence de convergence des poids. L'effet de mémoire permet de stabiliser le réseau de neurones. Il faut donc déterminer empiriquement les valeurs qui permettront d'obtenir les meilleures performances comme discuté plus bas.

La correction des poids de la couche intermédiaire suit des étapes similaires. Les nouveaux poids pour les neurones de la couche cachée sont obtenus par l'expression

$$W_{ij}(n + 1) = W_{ij}(n) + \eta \cdot \delta_j \cdot x_i + \alpha \cdot [W_{ij}(n) - W_{ij}(n - 1)]$$

Le terme de correction  $\delta_j$  de chaque neurone  $j$  de la couche cachée s'exprime

$$\delta_j = f'(a_j) \cdot \sum_k \delta_k \cdot W_{jk}(n) \text{ Pour}$$

Considérant que l'algorithme des réseaux de neurones est un algorithme itératif, la question qui se pose est : à quel moment doit-on arrêter l'apprentissage ? Le critère d'arrêt de l'apprentissage est basé sur la méthode de validation croisée externe. Pour ce faire, il faut séparer nos échantillons en deux lots, le lot d'apprentissage et le lot de validation. Le lot d'apprentissage permet de créer le modèle et le lot de validation, comme son nom l'indique, de valider le modèle. Les erreurs moyennes des lots d'apprentissage et de validation sont calculées à chaque itération avec les poids courants du réseau. En général, l'erreur moyenne du lot de d'apprentissage diminue continuellement avec les itérations alors que l'erreur moyenne du lot de validation passe par un minimum. À ce point, on bascule dans un régime de sur-apprentissage et il faut alors cesser d'itérer (voir figure 2.4). En cherchant à minimiser les RMSE des lots d'apprentissage et de validation, cette procédure permet d'optimiser les paramètres du réseau comme la vitesse d'apprentissage, l'effet mémoire, le nombre de neurones sur la couche cachée et même le nombre optimum d'itérations.

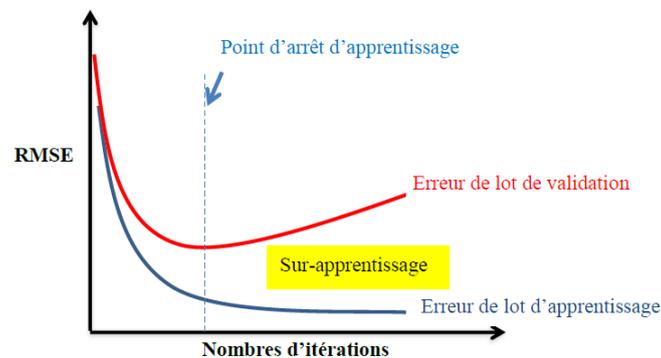


Figure 2. 4 Évolution du RMSE par rapport au nombre d'itérations pour le lot de validation en rouge et le lot d'apprentissage en bleu. Le minimum de la courbe rouge correspond au critère d'arrêt de l'apprentissage. [8]

### Chapitre 3 : Étude multivariée

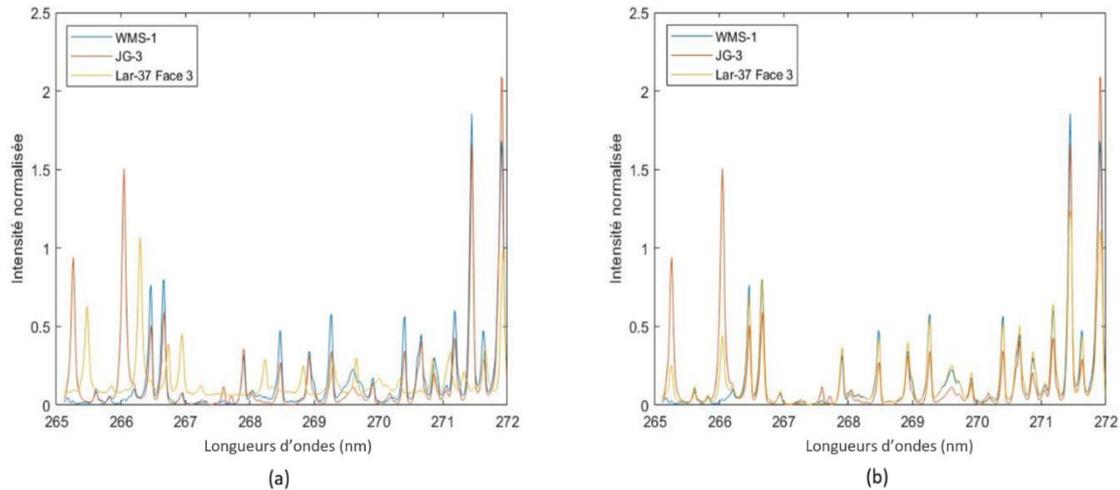
Dans l'objectif d'évaluer différents modèles de calibration, nous disposons d'échantillons prélevés dans différentes mines d'or du Québec sous forme de carottes de forage qui font environ 1 mètre de long sur quelques centimètres de diamètre. Une moitié de chaque carotte, coupée dans le sens de la longueur, a été analysée par le laboratoire du site minier. La gravimétrie et l'émission atomique sont les deux techniques conventionnelles qui ont été employées par les laboratoires. Un parallélépipède dont les côtés font quelques centimètres a ensuite été taillé dans l'autre portion de chaque carotte de forage. Ces échantillons ont d'abord été analysés par LIBS puis par le laboratoire certifié ALS, par les méthodes des gravimétrie et d'absorption atomique. À cause de la distribution hétérogène de l'or dans le minerai, les concentrations en or obtenues par les deux laboratoires et par LIBS sont généralement différentes. Cette hétérogénéité se manifeste d'une part par le fait que les laboratoires analysent un volume de la roche après pulvérisation et homogénéisation alors que la méthode LIBS analyse la surface de la roche, et d'autre part par le fait que les laboratoires n'analysent pas les mêmes échantillons. Le tableau 3.1 recense les échantillons qui ont été utilisés pour l'évaluation des modèles. Pour déduire la concentration en or d'un échantillon d'évaluation, les mesures LIBS des six faces composant l'échantillon ont été faites.

Mine de provenance	Échantillons d'évaluation	Laboratoire ALS (ppm)	Laboratoire du site minier (ppm)
Casa Berardi	Casber-17	40.5	5.28
Goldex	Gol-69	0.05	1.21
Goldex	Gol-70	0.78	2.35
LaRonde	Lar-37	1.7	0.44
LaRonde	Lar-45	0.22	2.11
Malartic	Mal-43-1	0.32	0.86
Malartic	Mal-43-2	0.37	0.86
Westwood	West-16	99.5	4.5
Westwood	West-19	1.08	2
Westwood	West-30	2.56	1.25

**Tableau 3.1 Échantillons utilisés pour évaluer les résultats obtenus par la méthode LIBS par comparaison avec les mesures de laboratoire.**

Les échantillons fournis ont l'avantage d'avoir été doublement analysés. Les disparités entre les résultats du laboratoire ALS et le laboratoire du site minier sont importantes pour certains échantillons comme West-16 où l'écart est de 95 ppm. Les concentrations faibles constituent une difficulté pour évaluer les modèles chimiométriques qui ont été élaborés avec des échantillons de plus grandes concentrations. En effet la prédiction de ces échantillons en utilisant leurs spectres moyens sera compliquée à cause de l'intensité de la raie d'or qui sera difficilement distinguable du bruit de fond. Ainsi, les résultats des prédictions seront à la hauteur de ce que seront capable de faire ces méthodes. Cependant, si on prédit la concentration de la raie d'or tir par tir, il y a la possibilité qu'à certains endroits la concentration en

or soit grande, ce qui rend les courbes de calibration plus adaptées pour la prédiction de ces tirs. Les erreurs moyennes quadratiques de prédiction RMSEP seront calculées par rapport aux concentrations fournies par le laboratoire ALS puisqu'elles concernent les mêmes échantillons que ceux mesurés par LIBS.



**Figure 3. 1** Comparaison entre le spectre normalisé par rapport à l'aire spectrale de la face 3 de l'échantillon Lar-37 (a) non corrigé en longueurs d'onde et (b) corrigé en longueurs d'onde par une transformation linéaire, et ceux des échantillons de calibration WMS-1 et JG-3 à 50 ppm.

En comparant les spectres des échantillons d'évaluation avec ceux des échantillons de calibration JG-3 et WMS-1 nous avons observé un décalage en longueur d'onde des raies par rapport à ces derniers. Celui-ci est probablement dû à une erreur ou à une négligence lors la prise des mesures. Nous l'avons corrigé numériquement en appliquant une transformation linéaire sur les échantillons test. Nous avons utilisé les pics d'intensités 266.6690 nm et 271.4470 nm des échantillons de référence pour calculer la distance entre ces mêmes pics des échantillons d'évaluation qui étaient décalés.

Les spectres utilisés par la suite seront les spectres moyennés sur le nombre de spectres prélevés sur chaque échantillon (un spectre par tir laser). Ils seront normalisés sur l'ensemble du domaine spectral et seront centrés sur leur moyenne.

Afin d'établir un modèle quantitatif multivarié, il est indispensable de bien connaître ses échantillons. Pour cela nous avons utilisé l'analyse en composantes principales (ACP).

### 3.1 Analyse en composantes principales des échantillons

Comme mentionné précédemment, l'étude des spectres par ACP permet de révéler facilement les groupes d'échantillons similaires du point de vue de leur composition chimique. Cela offre aussi l'opportunité d'identifier les raies qui varient le plus d'un spectre à l'autre et qui donnent ainsi une

information utile sur nos échantillons. Les échantillons de calibration sont composés de 509 variables dans la gamme 265.11 - 271.99 nm. Nous disposons de six échantillons pour WMS-1, JG-3 et LDI-1, cinq pour PTC-1b et dix pour QC.

L'ACP est effectuée sur la matrice  $X_0$  définie comme

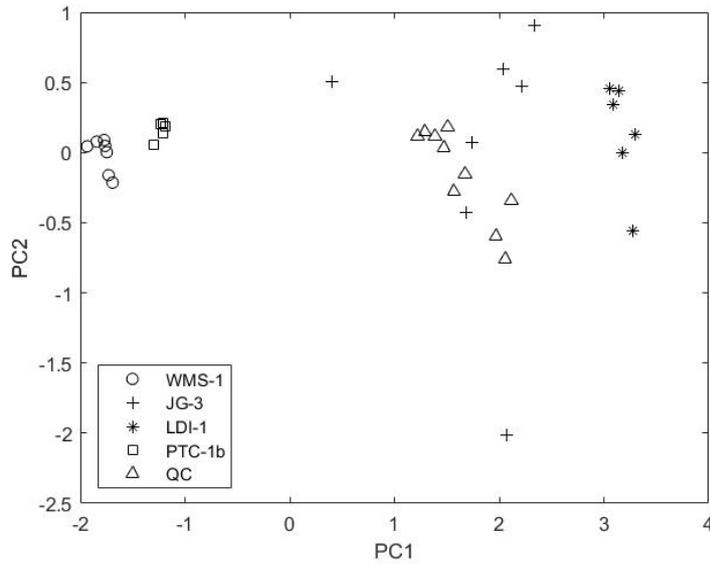
$$X_0 = X - \bar{X} \quad (3.1)$$

où  $X$  est la matrice de base (les lignes représentent les spectres et les colonnes représentent les longueurs d'onde) et  $\bar{X}$  est la matrice dont les colonnes sont formées de la valeur moyenne de la colonne correspondante de la matrice  $X$ .

La figure 3. 2 représente les scores des échantillons de calibration selon la première et la deuxième composante principale. La première composante principale recense 86.82% de la variance totale et la deuxième composante principale 11.02%. Ainsi ce plan contient 97.84% de la variance totale. On observe sur cette figure deux groupes d'échantillons :

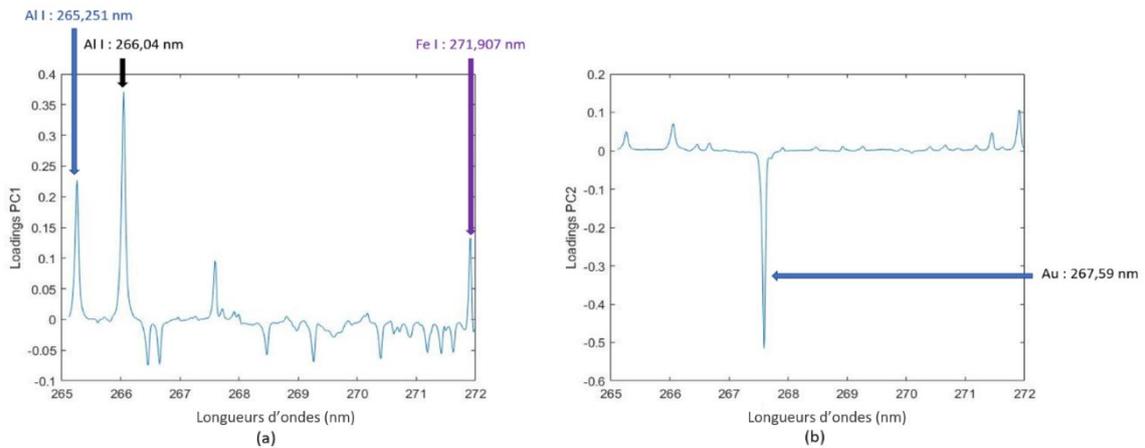
- Le groupe composé des échantillons WMS-1 et PTC-1b.
- Le groupe composé des échantillons LDI-1, JG-3 et QC.

Le premier groupe représente les échantillons qui sont riches en fer et le deuxième groupe les échantillons qui sont riches en silice, pauvres en fer. À première vue, la première composante principale répartit les échantillons de calibration suivant leur teneur en fer. Cependant, la figure 3. 3 (a) montre que les trois valeurs des vecteurs propres (loadings) de la première composante principale les plus hautes sont une raie de fer et deux raies d'aluminium. Par conséquent, la première composante principale ne répartit pas seulement les échantillons en fonction de leur contenu en fer, mais aussi en fonction de leur contenu en aluminium car l'intensité des raies d'aluminium est grande pour les échantillons riches en silice alors qu'elle est faible pour les échantillons riches en fer.



**Figure 3. 2** Scores des échantillons de calibration suivant les deux premières composantes principales. Chaque point correspond à un échantillon de son lot de calibration.

Nous avons aussi calculé les vecteurs propres de la deuxième composante principale. D'après la figure 3. 3 (b) la deuxième composante principale est majoritairement affectée par la raie d'or situé à 267.59 nm.



**Figure 3. 3** (a) Loadings de la première composante principale suivant la longueur d'onde. (b) Loadings de la seconde composante principale suivant la longueur d'onde.

Pour étudier les échantillons d'évaluation par ACP, nous avons projeté leurs spectres moyens sur les composantes principales calculées avec les échantillons de calibration. Les scores des échantillons d'évaluation sur la nième composante principale s'expriment :

$$S_i^{eval,n} = \sum_{j=1}^P (X_{ij}^{eval} - \overline{X_{ij}^{cal}}) \cdot v_j^{cal,n} \quad (3.2)$$

Ici,  $j = 1, \dots, P$  où  $P$  est le nombre de longueurs d'onde,  $i = 1, \dots, N'$  où  $N'$  est le nombre d'échantillons d'évaluation, les  $X_{ij}^{test}$  sont les éléments de la matrice  $N' \times P$  des échantillons d'évaluation, les  $\overline{X_{ij}^{cal}}$  sont

les éléments d'une matrice dont les colonnes sont la moyenne sur la colonne  $j$  de la matrice des échantillons de calibration (similaire à  $\bar{X}$  dans (3.1)) et les  $v_j^{cal,n}$  sont les composantes du nième vecteur propre de la matrice de covariance  $X_0 \cdot X_0$  des échantillons de calibration.

Les scores des échantillons d'évaluation sont représentés sur la figure 3.4. On observe que les échantillons d'évaluations sont encadrés par les échantillons de référence. Les échantillons Casber-17 West-16 et Lar-37 ont une teneur en fer comparable à WMS-1 et PTC-1b. Pour le reste, ils sont plutôt pauvres en fer.

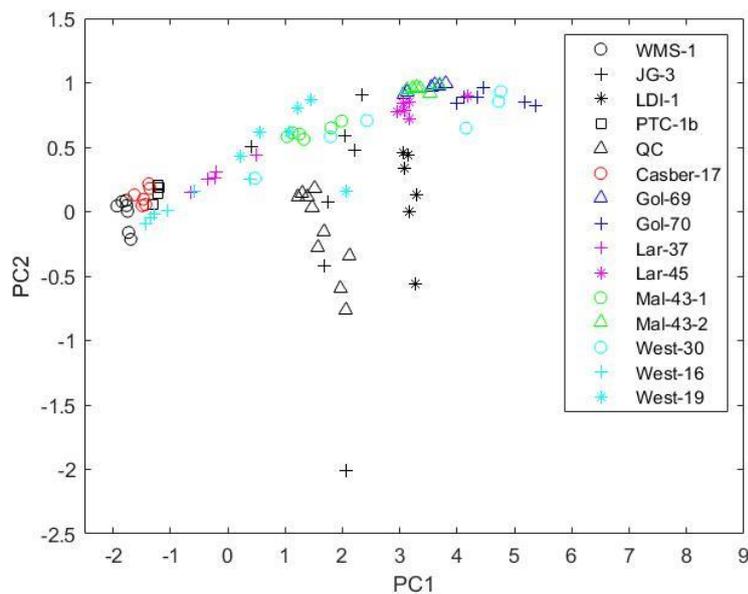


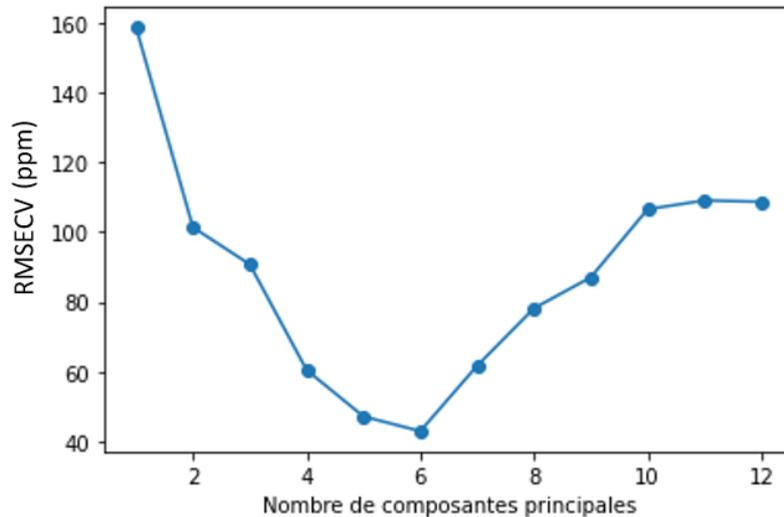
Figure 3. 4 Scores des échantillons de calibration et des échantillons d'évaluation suivant les deux premières composantes principales calculées avec les échantillons de calibration. Chaque point d'un lot de calibration représente un échantillon de référence. Chaque point d'un échantillon d'évaluation représente le spectre moyen de chacune des 6 faces qui le composent.

La connaissance de nos échantillons nous permet d'avancer sur l'analyse quantitative des données. Nous pouvons commencer à quantifier la concentration en or dans nos échantillons de test par PLS.

### 3.2 Analyse quantitative des échantillons par PLS

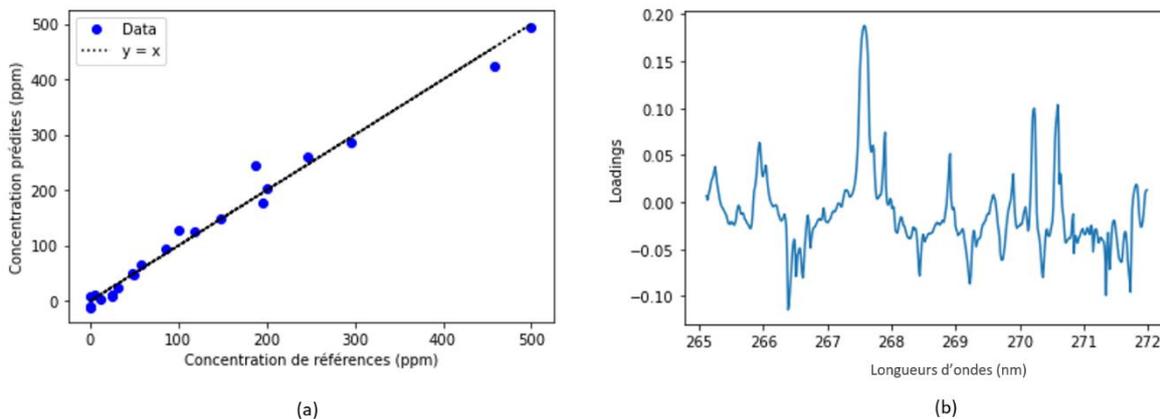
Précédemment nous avons vu que les échantillons étaient classifiés selon leur teneur en fer et en aluminium. La majorité des échantillons d'évaluation ressemble aux échantillons de calibration qui sont pauvres en fer c'est-à-dire JG-3, LDI-1 et QC. Nous avons décidé dans un premier temps de prédire les valeurs des échantillons d'évaluation par PLS en utilisant JG-3, LDI-1 et QC pour la phase d'apprentissage. La méthode de validation croisée « leave one out » a été utilisée pour évaluer la qualité du modèle quantitatif ainsi que le nombre de composantes principales qui étaient requis. Nous avons ici une matrice de données composé de 22 lignes (22 spectres – voir tableau 1, en excluant JG-3 à 932

ppm) et de 509 colonnes (509 longueurs d'ondes). La méthode de validation croisée nous atteste que le minimum de RMSECV est atteint pour 6 composantes principales avec une valeur de 43 ppm (figure 3.5).



**Figure 3. 5 Résultats RMSECV sur 12 composantes principales par PLS.**

Les concentrations prédites sont étalées suivant la courbe  $y = x$  de la figure 3.6 (a). Cela nous atteste que les prédictions sont précises. C'est donc un bon modèle de calibration qui n'est pas en état de sur-apprentissage car nous avons choisi le bon nombre de composantes principales par validation croisée. Les loadings de la figure 3.6 (b) ont un maximum à 267.59 nm qui est notre raie d'or, ce qui prouve que l'algorithme a bien reconnu la variation d'intensité en rapport à la variation de la concentration en or des échantillons. Cependant, ce pic est accompagné de plusieurs autres pics de grande amplitude.



**Figure 3. 6 (a) Concentrations d'or prédites par le modèle PLS à 6 composantes principales en utilisant tous les lots de calibration pauvres en fer (JG-3, LDI-1 et QC). (b) Poids calculés par PLS sur les lots de calibration pauvres en fer avec 6 composantes principales.**

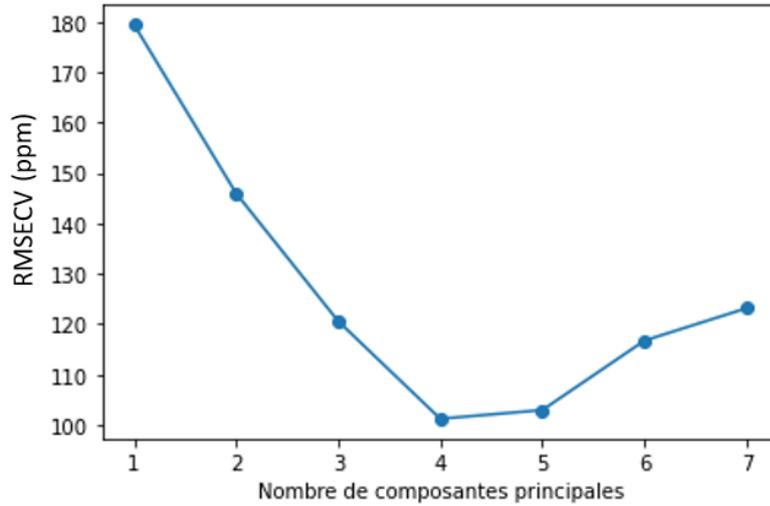
Les résultats des concentrations prédites par PLS pour des échantillons d'évaluation en utilisant 6 composantes principales sont disposés dans le tableau 3. 2. La valeur de concentration dans la dernière colonne est obtenue en calculant la concentration en or de toutes les faces composant un échantillon d'évaluation puis en calculant la moyenne pondérée en fonction du nombre de tirs laser sur les 6 faces. Les valeurs prédites par PLS sont toutes très surestimées par rapport à celles fournies par les analyses de laboratoire. L'erreur quadratique moyenne sur la prédiction des échantillons test par PLS vaut 109 ppm.

Échantillons d'évaluation	Laboratoire ALS (ppm)	Laboratoire du site minier (ppm)	Prédiction LIBS par PLS (ppm)
Casber-17	40.5	5.28	123
Gol-69	0.05	1.21	137
Gol-70	0.78	2.35	123
Lar-37	1.7	0.44	54
Lar-45	0.22	2.11	154
Mal-43-1	0.32	0.86	127
Mal-43-2	0.37	0.86	123
West-16	99.5	4.5	122
West-19	1.08	2	137
West-30	2.56	1.25	52

**Tableau 3. 2 Comparaison entre les concentrations mesurées par les laboratoires et les concentrations obtenues par la méthode PLS pour les échantillons test en utilisant pour l'apprentissage les échantillons de calibration JG-3, LDI-1 et QC.**

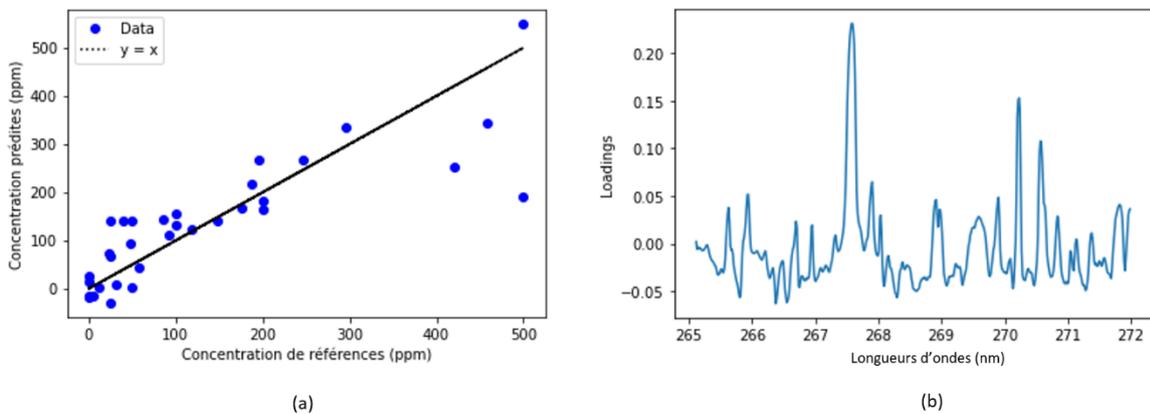
Les échantillons d'évaluation ont de faibles concentrations en or si bien qu'il est difficile pour l'algorithme de quantifier sa valeur. Ce manque de précision est potentiellement dû à l'ajout de polyéthylène dans la phase de production des données de calibration qui entraîne une variation anormale de l'intensité de certaines raies. Nous savons depuis le chapitre 2 que PLS calcule les scores de façon à maximiser la covariance entre les spectres et la concentration de l'élément de référence. Il est possible que la variation importante des raies selon la concentration ait été prise en compte dans la prédiction, ce qui affecte les résultats obtenus. Néanmoins, ne pouvant pas refaire les échantillons de calibration nous avons décidé de faire avec.

Dans l'espoir d'améliorer les résultats nous avons utilisé tous les échantillons de calibration pour l'apprentissage de PLS permettant ainsi de calculer de nouveaux loadings. Ainsi, nous avons maintenant une matrice de données comprenant les échantillons de calibration de dimension (33, 509) soit 33 spectres (voir Tableau 1, en excluant JG-3 à 932 ppm et WMS-1 à 891 ppm) de 509 longueurs d'ondes. Pour connaître le nombre idéal de composantes principales, nous avons utilisé la même méthode de validation croisée qu'auparavant. Le résultat est qu'il nous faut 4 composantes principales avec un RMSECV minimal égal à 101 ppm (figure 3.7).



**Figure 3. 7 Résultats de RMSECV sur 7 composantes principales par PLS.**

Les points de concentration sont moins alignés sur la courbe linéaire  $y = x$  qu'avant (figure 3.6 (a)). Il est clair que ce modèle de calibration est moins bien. Le plus grand loading observé sur la figure 3.8 (b) est toujours le loading de la raie d'or, ce qui nous annonce que l'algorithme a reconnu la variation d'intensité de la raie d'or en relation avec sa variation de concentration. Néanmoins nous avons toujours ce problème de la grande amplitude de plusieurs autres pics.



**Figure 3. 8 (a) Concentrations d'or prédites avec le modèle PLS à 4 composantes principales en utilisant tous les lots de calibration. (b) Poids calculés par PLS sur les lots de calibration avec 4 composantes principales.**

Les résultats de la quantification des échantillons test par PLS avec 4 composantes principales sont consignés dans le tableau 3. 3. Celui-ci nous montre aussi que les résultats prédits ont des valeurs bien plus grandes que les valeurs mesurées en laboratoire. Cependant, les concentrations prédites pour les échantillons Lar-37 et West-30 ont été largement améliorées. L'ajout d'échantillons riches en fer dans l'apprentissage de PLS améliore un peu les résultats pour Lar-37 qui est un échantillon de test riche en fer d'après la figure 3.4. Pour l'échantillon West-30, le point le plus proche de WMS-1 et PTC-1b sur

la figure 3.4 est la face qui contient le plus grand nombre de tirs (3200 tirs) et donc participe le plus au calcul de la moyenne pondérée de toutes ses faces. Les autres prédictions sont souvent moins bonnes qu'avant, comme pour l'échantillon Casber-17 dont la concentration prédite augmente de 44 ppm. Ce changement est surprenant dans la mesure où l'échantillon Casber-17 est l'échantillon le plus riche en fer du lot de test d'après l'ACP. L'ajout de WMS-1 et PTC-1b aurait dû améliorer aussi sa valeur. Le RMSEP est égal ici à 117 ppm. Cette valeur est plus grande qu'auparavant et on en déduit que faire l'apprentissage avec le lot de calibration entier pour PLS n'améliore pas toutes les prédictions.

Échantillons d'évaluation	Laboratoire ALS (ppm)	Laboratoire du site minier (ppm)	Prédiction LIBS par PLS (ppm)
Casber-17	40.5	5.28	167
Gol-69	0.05	1.21	141
Gol-70	0.78	2.35	136
Lar-37	1.7	0.44	33
Lar-45	0.22	2.11	178
Mal-43-1	0.32	0.86	130
Mal-43-2	0.37	0.86	134
West-16	99.5	4.5	127
West-19	1.08	2	124
West-30	2.56	1.25	18

**Tableau 3. 3 Comparaison entre les concentrations mesurées par le laboratoire et les concentrations obtenues par la méthode PLS pour les échantillons d'évaluation en utilisant pour l'apprentissage tous les échantillons de calibration.**

On conclut que malgré tous nos efforts, la méthode de quantification PLS a ses limites. Il est difficile de dire qu'elle n'est pas adaptée à la quantification de la raie d'or de nos échantillons de validation. Les échantillons ont une concentration trop faible, difficile à prédire. L'ajout de polyéthylène dans la création des échantillons de calibration contribue peut-être à dégrader les prédictions. Il reste néanmoins clair que l'utilisation de tous les échantillons de calibration avec des matrices différentes n'améliore pas les résultats. À cela s'ajoute le caractère non linéaire de l'évolution de l'intensité de la raie d'or vu dans le chapitre 1. La méthode PLS étant un algorithme linéaire n'est en principe pas capable de décrire ce caractère non linéaire. Pour vérifier cette hypothèse nous avons quantifié la raie d'or avec les réseaux artificiels de neurones qui est un algorithme non linéaire.

### **3.3 Analyse quantitative par ANN**

Les réseaux de neurones artificiels sont sensibles au prétraitement des données. Prétraiter les données d'apprentissage permet d'augmenter la précision du modèle et les chances que l'algorithme converge. Nous avons choisi comme prétraitement des données la normalisation min-max souvent utilisée pour les modèles prédictifs :

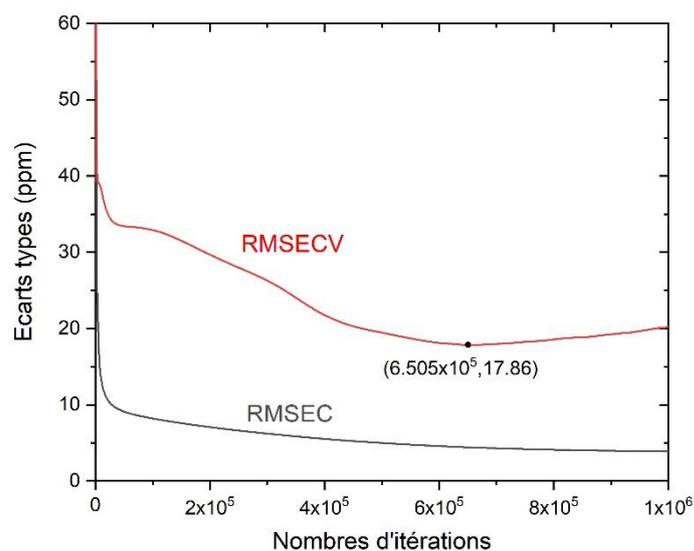
$$X_{\text{normalisé}} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

Les données normalisées  $X_{\text{normalisé}}$  sont comprises entre 0 et 1. Dans le cadre d'une étude par ANN, avoir des données normalisées de cette manière permet de rendre la rétropropagation de l'erreur plus stable et plus rapide que sans normalisation. Cependant, comme cette normalisation dépend des valeurs maximales et minimales des données, il faut faire attention à ce qu'il n'y ait pas de valeurs aberrantes, autrement cela produira du biais.

Pour procéder à une analyse quantitative des échantillons il faut choisir l'architecture du réseau de neurones que nous voulons employer pour prédire les valeurs. Nous avons utilisé un réseau de neurones à trois couches comme décrit auparavant. Cette architecture est généralement considérée comme adéquate dans la grande majorité des cas simples, comme ceux considérés ici [1]. La couche d'entrée compte autant de neurones que de variables soit 509 et la couche de sortie, un neurone. Le nombre de neurones de la couche cachée a été fixé à 3. Les poids, à la première itération se sont vu attribuer une valeur aléatoire comprise entre -0.5 et 0.5. La vitesse d'apprentissage est égale à 0.1 et le paramètre de mémoire 0.6.

Nous avons utilisé dans l'apprentissage tous les échantillons de calibration. Plus précisément, 5 échantillons de chaque lot de calibration ont été choisis pour l'apprentissage. Les échantillons restants ont servi à valider le modèle. Le nombre d'itérations optimal a été déterminé par la méthode de validation croisée.

Après l'apprentissage du réseau de neurones artificiels fait avec les échantillons de calibration, le meilleur nombre d'itérations est 650 500 (figure 3.9). L'erreur moyenne quadratique calculée avec ces paramètres est de 18 ppm.



**Figure 3. 9** Résultat de RMSECV et RMSEC selon le nombre d'itérations par validation croisée.

Les concentrations prédites avec l'ANN sont consignées dans le tableau 3.4. On remarque que les résultats fournis par ANN sont les plus proches des valeurs mesurées chez ALS de toutes celles obtenues jusqu'à présent. Les concentrations prédites ont tendance à être plus petites que les valeurs de laboratoire. La valeur finale de l'erreur quadratique moyenne de prédiction est égal à 18 ppm.

Échantillons d'évaluation	Laboratoire ALS (ppm)	Laboratoire du site minier (ppm)	Prédiction LIBS par ANN (ppm)
Casber-17	40.5	5.28	1.9
Gol-69	0.05	1.21	0.65
Gol-70	0.78	2.35	0.10
Lar-37	1.7	0.44	0.52
Lar-45	0.22	2.11	0.07
Mal-43-1	0.32	0.86	23.05
Mal-43-2	0.37	0.86	0.23
West-16	99.5	4.5	0.01
West-19	1.08	2	29.61
West-30	2.56	1.25	0.16

**Tableau 3. 4** Comparaison entre les concentrations mesurées par le laboratoire et les concentrations obtenues par la méthode ANN pour les échantillons d'évaluation en utilisant pour l'apprentissage tous les échantillons de calibration.

L'erreur quadratique des prédictions pour ANN est bien plus faible que par PLS (17.86 ppm au lieu de 109 ppm). Les prédictions pour Mal 43-1 et de West-19 sont au-dessus des valeurs de laboratoires. Cela constitue un problème majeur pour l'évaluation de la concentration en or sur site dans la mesure où le dispositif expérimental doit permettre d'évaluer si le forage à un endroit en vaut la peine.

On sait que les prédictions multivariées sont affectées par la gamme de longueurs d'ondes choisies. Ici, la majorité des pics qui composent cette gamme sont des raies de fer II. Les seules autres raies d'intérêt sont la raie d'or ainsi que les deux raies d'aluminium. Cette pauvreté d'information chimique sur notre fenêtre d'observation limite la précision des résultats prédit par l'ANN et des autres méthodes multivariées.

Dans le chapitre 1, nous avons vu que sans avoir une méthode de sélection, la prédiction par des courbes de calibration univariées était impossible. L'idée du chapitre 4 est d'introduire des méthodes « hybrides » qui permettent d'associer la bonne courbe de calibration univariée à un échantillon test pour prédire sa concentration.

## Chapitre 4 : Méthodes de quantification hybrides

Les méthodes hybrides sont des méthodes qui utilisent à la fois des méthodes multivariées et univariées. Nous avons vu précédemment que les prédictions faites avec les courbes univariées ne sont efficaces que si nous arrivons à associer la bonne courbe de calibration à l'échantillon test. L'idée de ce chapitre est d'utiliser les méthodes multivariées mentionnées précédemment pour associer la bonne courbe de calibration à un échantillon test. Nous avons testé cette sélection avec l'ACP et l'ANN.

### 4.1 Cartes de corrélation des spectres d'évaluation avec les échantillons de calibration

Nous avons toujours étudié les spectres moyens des échantillons test dans le but de réduire le bruit. Cependant nous savons très bien que la surface analysée n'est pas homogène. Pour être sûr que les échantillons de calibration sont adaptés à la prédiction en or des échantillons test nous avons décidé de créer des cartes de corrélation avec un échantillon de calibration pour chaque face d'un échantillon d'évaluation. La définition usuelle de la matrice de corrélation utilisée est la suivante :

$$corr(i, j) = \frac{cov(i, j)}{\sqrt{cov(i, i) \times cov(j, j)}}$$

Ici, cov est la matrice des covariances avec  $i, j = 1, \dots, N+1$  où  $N$  est le nombre de tirs effectués sur la face de l'échantillon d'évaluation. La matrice de covariance est donc de dimension  $(N+1, N+1)$ . Cette matrice compte  $N+1$  lignes et colonnes car aux  $N$  tirs nous ajoutons le spectre de l'échantillon de calibration avec lequel cet échantillon d'évaluation est comparé. La matrice  $cov(i, j)$  est centrée sur la moyenne. Cette matrice est une matrice carrée, symétrique avec des valeurs comprises entre -1 et 1. Les valeurs sur la diagonale valent 1. La position du tir laser étant connue, nous avons associé chaque coefficient de corrélation à sa position pour obtenir les images des figures 4. 1, 4. 2 et 4. 3. Les dimensions du scan laser ainsi que le nombre de tirs sont présentés dans le tableau 4. 1.

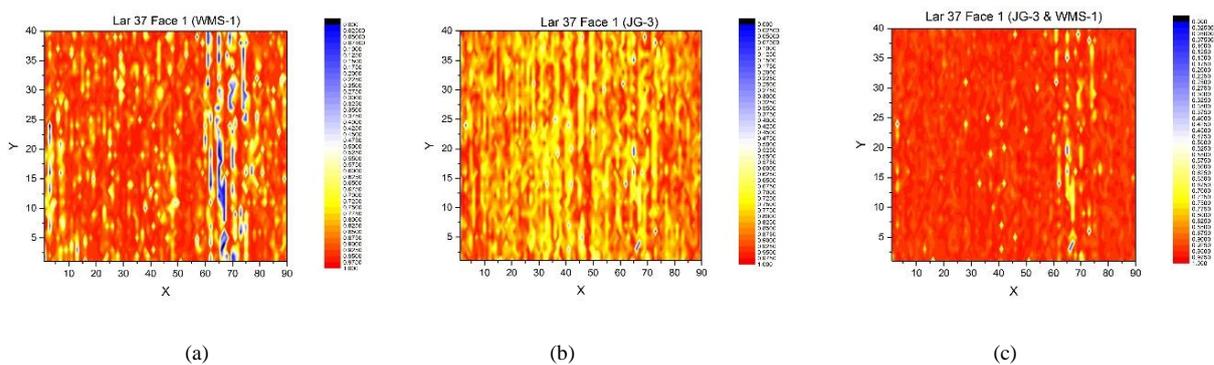
Echantillons d'évaluation	Dimension du scan laser (mm)	Nombre de tirs sur la face étudiée
Lar-37 face 1	90 x 40	3600
Gol-70 face 1	70 x 18	2100
West-30 face 2	70 x 45	3150

**Tableau 4. 1 Dimensions du scan laser et du nombre de tirs faits sur la surface des échantillons Casper-17, Gol-70 et West-30.**

Sur les cartes de corrélations présentées dans les figure 4.1, 4.2 et 4.3, les couleurs vont du bleu au rouge, la couleur bleue foncé correspondant à une corrélation proche de 0 et la couleur rouge vif à une corrélation maximum de 1. Le diamètre du point focal du laser est de 600  $\mu\text{m}$  et les tirs sont espacés de

800  $\mu\text{m}$ . La figure 4.1 montre que Lar-37 Face1 peut majoritairement être prédit par la courbe de calibration de WMS-1. Cela se remarque à la couleur rouge qui signifie que la corrélation est bonne. Néanmoins nous remarquons pour X entre les coordonnées 60 et 70 sur la figure 4.1 (a) des lisières bleues. Cette couleur bleue nous indique que WMS-1 n'est pas adapté pour prédire la concentration d'or à cet endroit. Sur la figure 4.1 (b) ces lisières bleues sont beaucoup moins prononcées. Cela signifie que dans cette zone, les spectres issus des tirs laser ressemblent plus à un spectre pauvre en fer et que l'utilisation de JG-3 est plus adaptée.

La figure 4.1 (c) montre le maximum des corrélations des figures 4.1 (a) et (b). On note que la figure est devenue généralement rouge, indiquant que sur cet échantillon, les courbes de calibration de WMS-1 et de JG-3 sont complémentaires et nécessaires pour prédire précisément la concentration en or de la face 1 de Lar-37. Il reste néanmoins de petites zones bleues qui indiquent que ni WMS-1 ni JG-3 ne ressemblent aux spectres prélevés à ces endroits.



**Figure 4. 1** Cartes de corrélation entre la face 1 de Lar-37 et (a) WMS-1, (b) JG-3, (c) JG-3 et WMS-1. Les nombres sur les axes dénotent les coordonnées des tirs laser dans la matrice rectangulaire.

La figure 4.2 montre les cartes de corrélation de l'échantillon Gol-70. La surface de la première face de cet échantillon d'après les cartes est majoritairement pauvre en fer. L'échantillon de calibration JG-3 permet de prédire un grand nombre de tirs contrairement à WMS-1. L'association de ces deux spectres de calibration ne permet pas de prédire toutes les zones observées puisque des zones bleutées subsistent sur la carte de corrélation conjointe de la figure 4.2 (c).

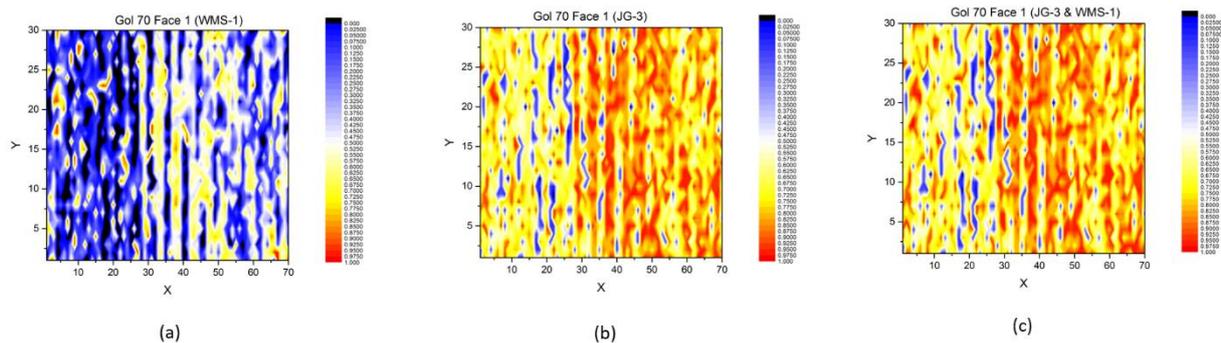


Figure 4. 2 Carte de corrélation entre la face 1 de Gol-70 et (a) WMS-1, (b) JG-3, (c) JG-3 et WMS-1.

L'échantillon test West-30 dispose de la matrice la plus complexe du lot d'échantillons de test. Nous voyons sur les cartes de corrélation produites avec WMS-1 et PTC-1b des figures 4.3 (a) et (d) une couleur noire que nous n'avons encore jamais vue. Cette couleur indique qu'il n'y a aucune corrélation entre le spectre test et le spectre de calibration. Ces zones noires deviennent bleues pour JG-3 et QC (figure 4.3(b), (f)) ce qui révèle une faible corrélation avec les échantillons pauvres en fer. Le meilleur résultat obtenu est avec l'échantillon de calibration LDI-1 où ces zones noires se transforment en bleu pâle (32% de de corrélation). Cela nous montre que pour cet échantillon, aucun de nos échantillons de calibration n'est adapté pour prédire dans sa globalité la concentration en or de celui-ci.

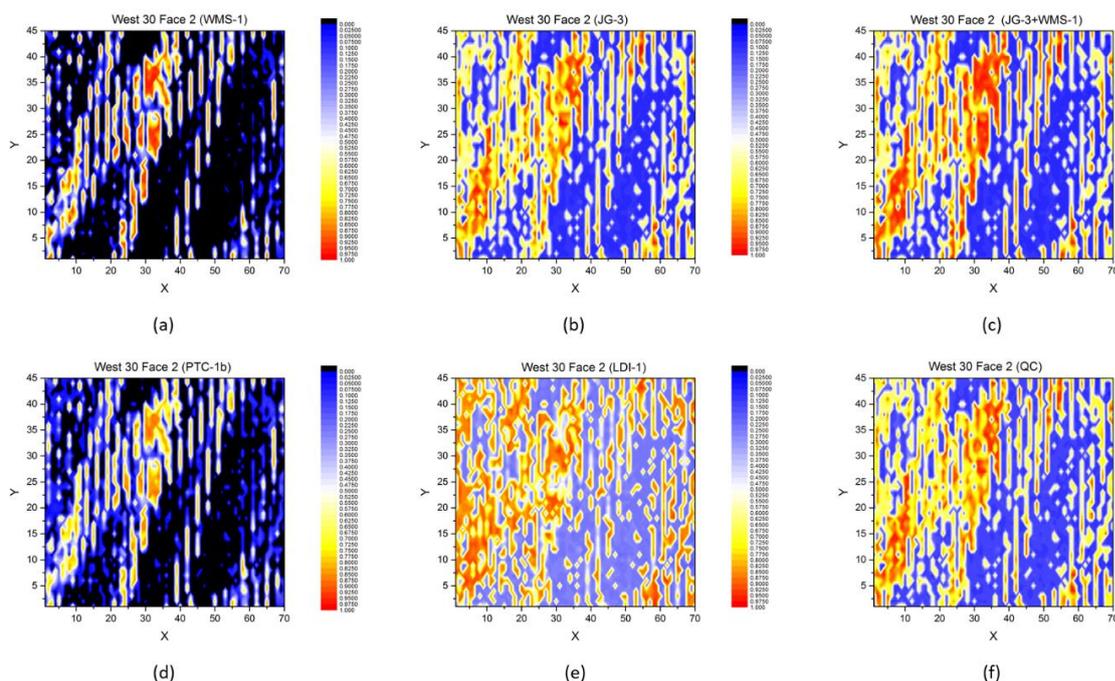


Figure 4. 3 (a) Carte de corrélation entre la face 2 de West-30 et (a) WMS-1, (b) JG-3, (c) JG-3 et WMS-1, (d) PTC-1b, (e) LDI-1, (f) QC.

Le calcul de la corrélation entre les échantillons de calibration et les échantillons de test permet d'associer la courbe de calibration qui lui est adaptée et de calculer la concentration en or. Le calcul de la concentration en or est possible mais n'a pas été réalisé dans ce mémoire en utilisant cette méthode

des corrélations des spectres. Néanmoins l'élaboration de ces cartes a permis de se rendre compte de l'hétérogénéité des roches étudiées.

Dans les prochaines sections, il s'agira d'utiliser les approches ACP et ANN afin de prédire la concentration en or de chaque tir en lui associant la bonne courbe de calibration, ce qui permettra de calculer la concentration en or moyenne de la face étudiée. Enfin pour obtenir la concentration moyenne de l'échantillon nous calculerons la moyenne pondérée des concentrations obtenues pour chaque face.

## 4.2 Sélection des courbes de calibration adaptées aux échantillons par ACP :

Comme nous avons vu dans le chapitre 3, l'analyse en composante principale ACP permet de regrouper les échantillons qui sont similaires. Cette propriété est très intéressante pour mieux connaître les échantillons analysés. L'idée de cette section est d'utiliser cette particularité pour associer à l'échantillon d'évaluation la bonne courbe de calibration. En effet si on calcule la distance entre le score de l'échantillon d'évaluation et le score d'un échantillon de calibration alors nous pouvons déterminer s'il est spectralement proche ou non de celui-ci. Cela peut être fait pour tous les échantillons de calibration et tous les échantillons d'évaluation de manière automatique en programmant un algorithme. La distance utilisée est la distance euclidienne. Entre un point  $A(x, y)$  et  $B(x', y')$ , elle est définie comme :

$$d_{AB,euclidienne} = \sqrt{(x - x')^2 + (y - y')^2}$$

Les données de calibration et des échantillons de test sont centrées sur la moyenne et normalisées par rapport à l'ensemble du domaine spectral.

L'algorithme procède en plusieurs étapes :

- Etape 1 : Calcul des scores des échantillons de calibration par ACP puis projection du spectre (tirs) de l'échantillon de d'évaluation pour obtenir son score dans la base des échantillons de calibration.
- Etape 2 : Calcul de la distance euclidienne entre le score du tir laser sur l'échantillon d'évaluation et tous les échantillons de calibration.
- Etape 3 : Utiliser la courbe de calibration de l'échantillon dont le score est le plus proche du score du tir laser en question. Prédire sa concentration en or et stocker sa valeur.
- Etape 4 : Répéter les étapes 1 à 3 jusqu'à que les concentrations en or de tous les tirs de la face aient été prédites.
- Etape 5 : Faire la moyenne des concentrations prédites des tirs de la face.

Cet algorithme est répété pour les six faces d'un échantillon de test et le calcul de la moyenne pondérée de ces six faces permet d'obtenir la valeur finale de la concentration du parallélépipède. Chaque tir de l'échantillon test est donc associé à l'échantillon de calibration qui lui ressemble le plus et prédit par celui-ci. Cela permet de prendre en compte l'hétérogénéité des surfaces analysées par LIBS. Les résultats obtenus figurent dans le tableau 4.2.

Échantillons d'évaluation	Laboratoire ALS (ppm)	Laboratoire du site minier (ppm)	Prédiction LIBS par méthode hybride (ppm)
Casber-17	40.5	5.28	26
Gol-69	0.05	1.21	3
Gol-70	0.78	2.35	10
Lar-37	1.7	0.44	14
Lar-45	0.22	2.11	2
Mal-43-1	0.32	0.86	12
Mal-43-2	0.37	0.86	4
West-16	99.5	4.5	35
West-19	1.08	2	81
West-30	2.56	1.25	15

**Tableau 4. 2 Résultat de la prédiction des échantillons de test par la méthode hybride en calculant la moyenne de la concentration obtenue pour chaque tir sur toutes les face des échantillons de test.**

Les résultats obtenus par cette méthode sont comparables à ceux de la méthode ANN puisque l'erreur RMSEP = 33 ppm. Néanmoins on trouve que la concentration de tous les échantillons d'évaluation excepté West-16 est plus grande que les valeurs de laboratoires ce qui constitue un problème pour une future quantification sur site de forage.

### **4.3 Sélection des courbes de calibration adaptées aux échantillons par ANN :**

Nous avons vu dans le chapitre 2 que les réseaux de neurones artificiels sont des algorithmes non linéaires. Ils permettent de classifier de façon plus fiable les données à caractère non linéaire que les algorithmes linéaires. Les nombres de tirs pour chaque face de chaque échantillon d'évaluation sont rangés dans le tableau 4.3. Nous observons qu'un grand nombre de tirs ont été faits sur la surface des échantillons d'évaluation. Casber-17 faces 5 et 6 ainsi que West-19 faces 5 et 6 ont un nombre de tirs assez faible par rapport aux autres faces qui les constituent. Ces derniers compteront moins dans la prédiction de la concentration en or en raison de la moyenne pondérée.

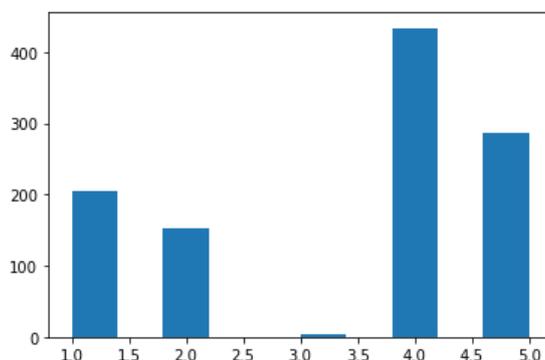
Nombre de tirs par faces pour les échantillons test					
Casber-17					
Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
1080	1080	600	600	160	160
Gol-69					
Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
1200	1200	1200	1200	900	920
Gol-70					
Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
2100	2450	1750	1750	875	796
Lar-37					
Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
3600	3600	900	720	320	400
Lar-45					
Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
2340	2160	2340	2160	576	576
Mal-43-1					
Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
1250	1250	1650	1650	750	750
Mal-43-2					
Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
1250	1250	1650	1254	750	750
West-16					
Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
2000	2000	2000	2000	1600	1600
West-19					
Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
1500	1200	680	680	300	200
West-30					
Face 1	Face 2	Face 3	Face 4	Face 5	Face 6
3200	3200	800	960	480	420

**Tableau 4. 3 Nombre de tirs projetés sur la surface des échantillons test.**

La classification du réseau de neurones a été faite par la fonction `MLPClassifier()` de la librairie `scikitlearn` [15].

Les paramètres optimaux du réseau de neurones sont trouvés par la fonction `GridSearchCV()` qui permet de trouver la famille de paramètres qui classifie avec le plus de précision. Dans cette fonction nous pouvons spécifier la méthode de sélection de la meilleure famille de paramètres. Nous avons choisi la méthode de validation croisée « leave one out ». Ce réseau de neurones sera constitué de 5 neurones de sorties (un pour chaque type d'échantillon de calibration) et 509 neurones sur la couche d'entrée. Après application de la fonction `GridSearchCV()` à nos échantillons d'apprentissage, nous trouvons comme meilleurs paramètres 8 neurones sur la couche cachée, un taux d'apprentissage de 0.001, l'effet de mémoire de 0.0001, la fonction d'activation ReLu (Rectified Linear Unit) et 231 itérations. Les résultats de la classification pour la face 1 de Casber-17 sont présentés dans l'histogramme de la figure 4. 4. Sur celle-ci nous voyons que la majorité des tirs laser ont été classés comme ressemblant à PTC-1b et QC.

Ces résultats sont similaires aux résultats de la classification ACP où nous avons vu que Casber-17 face 1 est riche en fer.



**Figure 4. 4** Histogramme du choix de la courbe de calibration la plus adaptée par ANN. Les chiffres 1, 2, 3, 4 et 5 correspondent respectivement à WMS-1, JG-3, LDI-1, PTC-1b et QC.

Les résultats de la prédiction des échantillons test par la classification ANN sont présentés dans le tableau 4.4. L'erreur moyenne quadratique RMSEP vaut 31 ppm. Les résultats sont meilleurs après classification par ANN que par ACP puisque le RMSEP après ANN est plus petit que pour ACP de 2 ppm. Il subsiste toujours des écarts importants entre les résultats des laboratoires et les prédictions. Seules les concentrations prédites de Casber-17 et West-16 sont moins grandes que les valeurs de référence. Les autres échantillons ont une concentration prédite plus grande que les concentrations analysées en laboratoire. La classification par ANN permet donc de réduire le RMSEP mais nous ne pouvons pas pour une analyse sur site se satisfaire de cette méthode, il faut l'améliorer.

Échantillons d'évaluation	Laboratoire ALS (ppm)	Laboratoire du site minier (ppm)	Prédiction LIBS par méthode hybride (ppm)
Casber-17	40.5	5.28	15
Gol-69	0.05	1.21	7
Gol-70	0.78	2.35	20
Lar-37	1.7	0.44	9
Lar-45	0.22	2.11	3
Mal-43-1	0.32	0.86	20
Mal-43-2	0.37	0.86	12
West-16	99.5	4.5	25
West-19	1.08	2	46
West-30	2.56	1.25	14

**Tableau 4. 4** Résultat de la quantification par la sélection des courbes de calibration univariées adaptées aux échantillons d'évaluation par ANN.

## Conclusion

Nous avons exploré différentes méthodes chimiométriques pour déterminer la concentration d'or dans les roches minières. Pour cela, nous avons utilisé des méthodes univariées, multivariées et hybrides. Les méthodes univariées ne permettent pas à elles seules de quantifier la concentration en or. Il faut y ajouter une stratégie de sélection pour y parvenir. Les méthodes multivariées ne fournissent pas des résultats satisfaisants à cause de la gamme de longueur d'onde étudiée qui ne dispose pas de raies riches en informations sur la composition chimique. À ce problème s'ajoute le caractère non linéaire des données qui n'améliore pas les prédictions pour la méthode linéaire PLS. Les réseaux de neurones artificiels ont fourni des prédictions acceptables mais pas pour tous les échantillons d'évaluation. Cette méthode reste prometteuse au même titre que les méthodes hybrides qui consistent à utiliser des méthodes multivariées pour sélectionner la courbe de calibration univariée adaptée à l'échantillon d'évaluation analysé. Elles ont permis de réduire le RMSEP de 109 ppm à 31 ppm pour la classification par ANN et de 109 à 18 ppm pour la prédiction ANN. Les méthodes de prédictions non linéaires fournissent les meilleurs résultats et sont donc les plus adaptées pour prédire la concentration en or de nos échantillons. L'élaboration des cartes de corrélations nous a permis de comprendre que pour quantifier précisément la concentration en or d'un échantillon rocheux, il faut des échantillons de calibration supplémentaires avec une composition chimique différente. Étudier les échantillons rocheux dans une autre gamme de longueurs d'ondes permettra d'augmenter le nombre de raies avec des informations géologiques pertinentes. L'extension de la gamme de longueurs d'ondes ainsi que l'élaboration de nouveaux échantillons de calibration permettront d'améliorer les résultats obtenus.

## Références

- [1] Josette El Haddad. Chimométrie appliquée à la spectroscopie de plasma induit par laser (LIBS) et à la spectroscopie terahertz. Other. Université Sciences et Technologies – Bordeaux I, 2013. French. <NNT: 2013BOR15206>. <tel-00959288>
- [2] Ye tian, Lintao Wang, Boyang Xue, Qian Chen and Ying li, Laser focusing geometry effects on laser-induced plasma and laser-induced breakdown spectroscopy in bulk water, *Journal of Analytical Spectrometry*, 118-126, 2018.
- [3] Kheireddine Rifai, Marcel Laflamme, Marc Constantin, François Vidal, Mohamad Sabsabi, Alain Blouin, Paul Bouchard, Konstantinos Fytas, Maryline Castello, Blandine Nguengang Kamwa, Analysis of gold in rock samples using laser-induced breakdown spectroscopy: Matrix and heterogeneity effects, *Spectrochimica Acta Part B*, 33-41, 2017
- [4] M. Castello. Développement de l'analyse de l'or des minerais par spectroscopie d'émission de plasma induit par laser (LIBS). M.Sc. Thesis, Université Laval, Québec, Canada (2018), 142 p.
- [5] David M. Haaland and Edward V. Thomas. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.*, 60 :1193–1202, 1988.
- [6] Harald Martens and Tormod Næs. *Multivariate calibration*. John Wiley & Sons, Chichester, New-York, Brisbane, Toronto, Singapore, 1989. 422 p.
- [7] Svante Wold, Michael Sjöström, and Lennart Eriksson. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.*, 58 :109–130, 2001.
- [8] S.O. Haykin, *Neural Networks and Learning Machines: International Version, Third Edition ed.*, Upper Saddle River [etc.] Pearson Education, 2009.
- [9] T.-C. Chang, R.-J. Chao, Application of back-propagation networks in debris flow prediction, *Engineering Geology*, 85 :270-280, 2006.
- [10] Einax, J.W. M.J. Adams: Chemometrics in Analytical Spectroscopy, 2nd edition. *Anal Bioanal Chem* **382**, 861–862 (2005). <https://doi.org/10.1007/s00216-005-3181-9>
- [11] Ismail Elhamdaoui , Kheireddine Rifai , Javed Iqbal , Nessrine Mohamed, Samira Selmani, Jordan Fernandes, Paul Bouchard, Marc Constantin, Marcel Laflamme, Mohamad Sabsabi, François Vidal, Measuring the concentration of gold in ore samples by laser-induced breakdown spectroscopy (LIBS) and comparison with the gravimetry/ atomic absorption techniques, *Spectrochimica Acta Part B* 183, 106256, 2021.
- [12] Jean-Baptiste Sirven, Bruno Bousquet, Lionel Canioni, Laurent Sarger, La Spectroscopie sur plasma induit par laser : une technique en devenir pour l'analyse sur site des sols pollués, *Spectra Analyse*, n°246 pages 29 – 32, (2005)
- [13] Olivier Barthélemy, Joëlle Margot, Stéphane Laville, Francois Vidal, Mohamed Chaker, Boris Le Drogoff, Tudor W. Johnston, and Mohamad Sabsabi, Investigation of the State of Local Thermodynamic Equilibrium of a Laser-Produced Aluminum Plasma, *Applied Spectroscopy*, Volume 59, number 4, 2005
- [14] Mermet, Jean-Michel *Spectrochimica Acta Part B* **65** : 509-523, 2010.
- [15] [Scikit-learn: Machine Learning in Python](#), Pedregosa *et al.*, *JMLR* 12 :2825-2830, 2011.
- [16] NIST database, <https://dx.doi.org/10.18434/T4W30F>, 2021