



AGROLIBS Spectroscopie de plasma induit par laser au service de la gestion durable des ressources agricoles

Julian Guezenoc

► To cite this version:

Julian Guezenoc. AGROLIBS Spectroscopie de plasma induit par laser au service de la gestion durable des ressources agricoles. Physique des plasmas [physics.plasm-ph]. Université de Bordeaux, 2020. Français. NNT : 2020BORD0207 . tel-03104031

HAL Id: tel-03104031

<https://theses.hal.science/tel-03104031>

Submitted on 8 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE PRÉSENTÉE A L'UNIVERSITE DE BORDEAUX

POUR OBTENIR LE GRADE DE

**DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE DES SCIENCES PHYSIQUES ET DE L'INGÉNIEUR

SPÉCIALITÉ : Doctorat en laser, matière et nanosciences

Par Julian GUÉZÉNOC

AGROLIBS

**Spectroscopie de plasma induit par laser au service de la gestion durable des
ressources agricoles**

Sous la direction de : Bruno BOUSQUET

Et de Anne GALLET-BUDYNEK

Soutenue le 27 Novembre 2020

Membres du jury :

Mme SERGENT, Michelle	Professeur, Université d'Aix-Marseille	Rapporteuse
M. MOTTO-ROS, Vincent	Maître de conférences, Université Claude Bernard Lyon 1	Rapporteur
Mme MARIN, Béatrice	Professeur des Universités, Université de Reims Champagne-Ardenne	Examinatrice
Mme EL HADDAD, Josette	Chargée de Recherche, CNRC	Examinatrice
M. RODRIGUEZ, Vincent	Professeur, Université de Bordeaux	Président du jury
M. BOULET, Jean-Claude	Ingénieur de Recherche, INRAE Montpellier	Examineur
M. BOUSQUET, Bruno	Professeur, Université de Bordeaux	Directeur de thèse
Mme GALLET-BUDYNEK, Anne	Ingénieur, INRAE Bordeaux-Aquitaine	Co-directrice de thèse
Mme DENAIX, Laurence	Directeur de recherche, INRAE Bordeaux-Aquitaine	Invitée

Remerciements

Cette thèse a été menée au Centre Lasers Intenses et Applications (CELIA) de Talence ainsi qu'à l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) de Villenave d'Ornon. Je remercie tout d'abord la Région Nouvelle-Aquitaine pour avoir financé mon travail de thèse. Je tiens à remercier également Eric MEVEL et Laurence DENAIX, directeur et directrice du CELIA et de l'INRAE respectivement, de m'avoir accueilli dans leur laboratoire.

Je suis honoré de recevoir l'expertise sur mon mémoire de thèse de Monsieur Vincent MOTTO-ROS et Madame Michelle SERGENT qui sont à la fois membres du jury et rapporteurs. Je souhaite remercier les membres du jury d'avoir accepté d'analyser le contenu de mon mémoire et d'assister à la soutenance, avec ou sans visio en ces temps compliqués, à savoir Mesdames Béatrice MARIN et Josette EL HADDAD et Messieurs Vincent RODRIGUEZ et Jean-Claude BOULET.

J'adresse un immense MERCI à Monsieur Bruno BOUSQUET et Madame Anne GALLET-BUDYNEK, directeur et directrice de mon travail de thèse. Voilà quatre ans que je travaille à vos côtés et je ne m'en lasse pas. A travers un soutien, une confiance et un désir de m'impliquer dans des expériences toutes enrichissantes (thèse en 180 secondes, AGU, les mesures en grotte, que de chemin parcouru !), j'arrive à voir à quel point je ne pouvais pas trouver meilleur tutorat et à quel point j'ai progressé grâce à vous. Encore Merci !

Je souhaite remercier également l'ensemble de l'équipe du CELIA (Sophie HEURTEBISE, Sonia SENUT, Céline RUTH) pour leur aide constante dans les démarches administratives et les équipes de recherche de l'INRAE (Laurence DENAIX, Cécile CORIOU, Alain MOLLIER) et de l'USRAVE (Pierre MASSON, Hugues ROUSSARIE), toujours disponibles lorsqu'il s'agit d'apporter de précieux conseils.

Je tiens à remercier aussi bien les permanents du SLAM (Lionel CANIONI, Inka MANEK-HONNINGER, Yannick PETIT) que les « moins permanents » (doctorants, techniciens, stagiaires) (Laura, Alain, Arthur, Elise) pour votre accueil, votre bienveillance et votre humour qui font de ce bâtiment vétuste qu'est le A4, un espace de travail particulièrement agréable ! J'ai une pensée toute particulière pour le trio magique que nous formions avec Wendé et Delphine aka la lazy team, l'année passée dans le même bureau n'est parsemée que de très bons souvenirs... Je remercie également Léna, dont j'admire et félicite le parcours et qui m'a toujours donné des conseils avisés, de l'époque où j'étais stagiaire jusqu'à la rédaction de mon mémoire de thèse.

Je tiens tout particulièrement à remercier Etonam et Thomas, qui ont aidé au bon développement de ce travail de thèse à travers leur sujet de stage. Je leur souhaite un très beau parcours. Je ne peux que remercier les doctorants de l'INRAE (David, Fred, Mohammed) pour leur bonne humeur communicative et les nombreuses séances de foot endiablées.

En tant que Breton ayant besoin de s'échapper de l'air Bordelais parfois irrespirable, je me dois de remercier l'ensemble de la J.United, collectif breton composé de gens formidables particulièrement portés par les joies simples de la vie. Le festival des anniversaires organisé par

vos soins en pleine contrée bretonne a été une véritable bouffée d'air frais durant la rédaction de ce mémoire.

Malgré une chaleur difficilement supportable de mai à octobre et des termes tels que « chocolatine » ou « poche », l'air Bordelais n'est pas si désagréable car j'y ai rencontré des personnes formidables que je considère aujourd'hui comme mes « z » amis (Mimic, Rudolphe, Lenny et Paul). J'ai aussi une certaine pensée pour un certain Damien, avec qui j'ai vécu 2 ans et avec qui nous avons tutoyé les tréfonds de la colocation.

J'ai évidemment une pensée pour toute ma famille, qui malgré ma faible présence ces derniers mois, a toujours compris et supporté mes choix. Je pense tout particulièrement à ma mère qui a toujours su trouver les mots justes pour me remotiver et à mon frère pour son soutien.

Je remercie finalement Vivy, qui a survécu au confinement et à la rédaction à mes côtés, et dont la présence au quotidien me fait le plus grand bien.

Table des matières

Table des matières	i
Table des figures	v
Table des figures - Annexe	xi
Table des tableaux	xiii
Table des tableaux - Annexe	xiv
Introduction	1
Chapitre I. Démarche analytique générale appliquée à des matrices agronomiques : sols et plantes	6
I.1) Corpus	7
I.1.i. Sélection des sols agricoles	7
I.1.ii. Sélection des échantillons de plantes	8
I.1.ii.a) Corpus de calibration	8
I.1.ii.b) Corpus de validation	9
I.1.iii. Echantillons pour l'étude de la faisabilité de mesures LIBS <i>in situ</i>	10
I.2) Mesures des concentrations de référence	11
I.2.i. Mesures par ICP-AES	11
I.2.i.a) Principe	11
I.2.i.b) Analyse des concentrations de référence de nos échantillons	12
I.2.ii. Mesures par XRF portable	13
I.3) Méthode LIBS	15
I.3.i. Principe	15
I.3.ii. Instrument LIBS portable	17
I.3.iii. Contrôle de l'instrument portable Z-300	20
I.3.iv. Analyse LIBS sur pastille : un protocole établi par plan d'expérience	22
I.3.v. Prétraitement des spectres LIBS	25
Chapitre II. LIBS portable et plantes : vers des modèles universels ?	28
II.1) L'observation des données	29
II.1.i. Identification spectrale	29

II.1.ii. Analyse en composantes principales, principe et calcul	32
II.1.ii.a) Principe général et application à la spectroscopie LIBS	32
II.1.ii.b) Le calcul mis en œuvre dans l'ACP	34
II.1.iii. ACP centrée sur les plantes	36
II.1.iii.a) ACP et information physique sur le plasma	38
II.1.iii.b) ACP et information chimique	40
II.1.iii.c) Détection de raies auto-absorbées par ACP	43
II.2) L'analyse quantitative des plantes	44
II.2.i. Normalisation	44
II.2.i.a) Etat de l'art de la normalisation en LIBS	45
II.2.i.b) Figures de mérite pour la sélection de la meilleure normalisation	46
II.2.i.c) Application aux échantillons de plantes	49
II.2.ii. Construction des modèles de régression	49
II.2.ii.a) Calcul par normalisation SNV	49
II.2.ii.b) Potassium sans normalisation	52
II.2.ii.c) Normalisation par étalon interne (Argon)	53
II.2.iii. Validation des modèles d'étalonnage	58
II.3) Conclusion sur l'analyse LIBS des échantillons de plante	60
II.3.i. Analyse qualitative LIBS	60
II.3.i.a) Détection	60
II.3.i.b) Analyse en composantes principales	61
II.3.ii. Analyse quantitative LIBS	61
II.3.ii.a) Calibration	61
II.3.ii.b) Validation	62
II.3.iii. L'apport de la XRF	64
II.3.iii.a) Détection des <i>outliers</i>	64
II.3.iii.b) Complémentarité des deux méthodes	65
Chapitre III. LIBS portable et sols agricoles : des modèles représentatifs des sols agricoles français ?	67
III.1) L'analyse qualitative des sols	69

III.1.i. Identification spectrale	69
III.1.ii. Analyse en composantes principales	70
III.1.ii.a) Décalage en longueur d'onde	70
III.1.ii.b) ACP et information chimique	72
III.1.ii.c) ACP et information physique	76
III.2) L'analyse quantitative des sols agricoles	78
III.2.i. Analyse univariée	79
III.2.i.a) Manque de corrélation entre le signal LIBS et la concentration	79
III.2.i.b) L'exception du Calcium	80
III.2.ii. Analyse multivariée	84
III.2.ii.a) Principe de la régression aux moindres carrés partiels	85
III.2.ii.b) Algorithme utilisé en PLS	86
III.2.ii.c) Construction du modèle d'étalonnage par PLS du magnésium	88
III.2.ii.d) Quantification des éléments d'intérêt	95
III.2.ii.e) Validation des modèles d'étalonnage	99
III.3) Conclusion de l'étude sur les sols agricoles	100
III.3.i. Analyse qualitative LIBS	100
III.3.i.a) Identification spectrale	100
III.3.i.b) Analyse en composantes principales	101
III.3.ii. Analyse quantitative LIBS	102
III.3.ii.a) Régression univariée	102
III.3.ii.b) Régression PLS	102
III.3.ii.c) Comparaison avec une étude similaire	104
III.3.ii.d) L'apport de la XRF	104
Chapitre IV. Du laboratoire au terrain ?	108
IV.1) Analyse <i>in situ</i>	108
IV.1.i. Protocole expérimental pour l'analyse des plantes	109
IV.1.ii. Protocole expérimental pour l'analyse des sols	109

IV.1.iii. Signaux obtenus sur les feuilles de blé intactes	109
IV.1.iv. Signaux obtenus sur les sols frais	112
IV.2) Influence de facteurs extérieurs sur le signal LIBS d'un échantillon de sol	113
IV.2.i. Principe du plan d'expériences	113
IV.2.ii. Réalisation d'un plan factoriel complet à 3 facteurs à 2 niveaux	113
IV.2.ii.a) Sélection de l'échantillon de sol	114
IV.2.ii.b) Sélection des modalités du plan d'expériences	114
IV.2.ii.c) Campagnes de mesures LIBS	117
IV.2.iii. Résultats	117
IV.2.iii.a) Graphe des effets	117
IV.2.iii.b) Graphe des interactions	121
IV.2.iii.c) Validation statistique : ANOVA	122
IV.2.iv. Evaluation et correction de l'effet de l'humidité sur le signal LIBS	125
IV.2.iv.a) Protocole expérimental	125
IV.2.iv.b) Evaluation du signal en fonction du taux d'humidité	126
IV.2.iv.c) Potentiel de la normalisation pour réduire les effets de l'humidité	127
IV.3) Conclusion sur l'effet des paramètres extérieurs sur le signal LIBS	128
III.3.i. Mesures <i>in situ</i>	128
III.3.ii. Effet des facteurs extérieurs sur le signal	129
III.3.iii. Focus sur l'effet de l'humidité et comment le compenser	130
III.3.iv. Recommandations	130
Conclusion	133
Annexe - Chapitre I	136
Annexe - Chapitre II	142
Annexe - Chapitre III	148
Annexe - Chapitre IV	154
Bibliographie	157

Table des figures

Figure 1-1 : Origine géographique des 62 échantillons de sols sélectionnés dans les analyses LIBS.	8
Figure 1-2 : Montage typique d'un dispositif ICP-AES [41].	12
Figure 1-3 : Montage typique d'un dispositif XRF portable [45].	14
Figure 1-4 : Photo de l'instrument de XRF portable, S1 800 Titan, développé par Bruker [46] et utilisé dans le cadre de cette étude.	14
Figure 1-5 : Schéma de principe d'une analyse LIBS (source Wikimedia Commons – Fanny Bachelier, 2016)	16
Figure 1-6 : Photo montrant l'instrument Z300 a) sur un châssis pour un usage en conditions de laboratoire, et b) en main pour un usage en dehors du laboratoire (notons que cette photo sert uniquement d'illustration et ne correspond pas à une mise en œuvre réaliste pour l'analyse in-situ d'un sol).	17
Figure 1-7 : Interface utilisateur sur PC de l'instrument Z-300.	18
Figure 1-8 : Comparaison du spectre LIBS du même échantillon de sol obtenu avec le pistolet Z-300 sous air (trait noir) et sous Argon (trait rouge), à pression ambiante.	19
Figure 1-9 : Carte de contrôle du pistolet LIBS Z-300, construite sur l'intensité nette de la raie du Fer à 302,08 nm. Le calcul des traits horizontaux et leur signification (LSC, LSS, Moy, LIS, LIC) est explicité dans le texte.	21
Figure 1-10 : Surfaces de réponse obtenues à l'aide du plan Box-Behnken pour l'intensité nette de la raie du silicium à 288 nm pour l'échantillon de sol 152 avec a) le nombre de zones en fonction du nombre de spots, b) le nombre de zones en fonction du nombre de tirs et c) le nombre de spots en fonction du nombre de tirs. Les cercles rouges représentent les conditions d'analyse conservées pour l'analyse des sols et des plantes par le Z-300.	24
Figure 1-11 : Problèmes constatés lors de l'extraction des spectres LIBS bruts du pistolet Z-300 avec a) un exemple de décalage en longueur d'onde pour le doublet du calcium à 315,92 et 317,96 nm et b) un exemple de recouvrement spectral.	26
Figure 1-12 : Spectre LIBS d'un sol zoomé sur la raie de Mg I à 285, 21 nm.	27
Figure 2-1 : Spectre LIBS typique moyen d'un échantillon de plante (pastille), ici une feuille de hêtre (IPE 157), avec l'attribution des éléments d'intérêt et les raies d'émission associées.	30

Figure 2-2 : Projection des données initiales (une bolée de cidre) sur le plan (1,2) et sur le plan (1,3) (illustration par Tony Richer).	32
Figure 2-3 : Schéma de de la réduction de dimension appliquée à des données LIBS.	34
Figure 2-4 : Ecriture matricielle de l'ACP.	34
Figure 2-5 : Schéma présentant l'intérêt de centrer les données pour une ACP.	37
Figure 2-6 : Pourcentage de variance expliquée par l'ACP sur les plantes en fonction du nombre de composantes ajouté. Les boîtes à moustache ont pu être calculées à l'aide du processus de cross-validation décrit dans le texte.	37
Figure 2-7 : Figures de mérite de l'ACP réalisée sur les échantillons de plantes avec a) le graphe des scores de la composante 1 en fonction de la composante 2 et b) le graphe des loadings de la composante 1.	38
Figure 2-8 : Graphe de l'aire totale sous le spectre en fonction de l'intensité nette de la raie de Ar à 763 nm pour les échantillons de plante analysés par LIBS.	40
Figure 2-9 : Graphe des loadings de l'ACP réalisée sur les échantillons de plantes pour la composante 2.	41
Figure 2-10 : Graphe des loadings zoomé sur les raies influentes pour la composante 4, 5 et 6 de l'ACP réalisée sur les plantes.	42
Figure 2-11 : extrait du graphe des loadings de la composante 3 zoomé sur le doublet du calcium à 393-396 nm.	43
Figure 2-12 : Présentation des quatre méthodes de normalisation les plus utilisées en LIBS ainsi que les précautions à prendre avant de les appliquer.	46
Figure 2-13 : Schéma récapitulant la stratégie de sélection de la meilleure normalisation.	48
Figure 2-14 : Courbe d'étalonnage pour le calcium dans les plantes a) non-normalisée et b) normalisée SNV pour la raie Ca II à 315,9 nm. Les cercles rouges indiquent les outliers observés sur l'étalonnage. L'ellipse verte indique la présence d'un gradient d'intensités pour une même concentration en calcium sur les données nettes.	50
Figure 2-15 : Extraits de spectres LIBS des pastilles de plante présentant un gradient d'intensités pour une même valeur de concentration en calcium a) signaux brut non-normalisés et b) signaux normalisés SNV.	51
Figure 2-16 : Droite de calibration finale pour le calcium dans les plantes. Le trait en pointillés représente la droite de régression linéaire entre le signal LIBS et la concentration mesurée.	51
Figure 2-17 : Courbe d'étalonnage pour l'intensité nette de la raie KI à 766,5 nm dans les échantillons de plantes sélectionnés. Les cercles rouges indiquent les outliers.	52

Figure 2-18 : Droite de calibration finale pour le potassium dans les plantes. Le trait en pointillés représente la droite de régression linéaire entre le signal LIBS (intensité nette de la raie de K I à 766,5 nm) et la concentration de référence. 53

Figure 2-19 : Courbe d'étalonnage pour le fer (aire nette de la raie Fe I à 302,1 nm) dans les plantes a) non-normalisée et b) normalisée par l'Argon (intensité nette de la raie Ar I à 763 nm) et pour l'aluminium (intensité nette de la raie Al I à 308,2 nm) c) non-normalisée et b) normalisée par l'Argon (intensité nette de la raie Ar I à 763 nm). 54

Figure 2-20 : Courbe d'étalonnage pour le magnésium dans les plantes à partir de la raie d'émission de Mg I à 285,2 nm a) aire nette sous le pic non-normalisée et b) aire nette sous le pic normalisée par l'Argon (intensité de la raie Ar I à 763 nm). Les ellipses rouges indiquent les outliers. 55

Figure 2-21 : Droite de calibration pour le magnésium (intensité nette de la raie de Mg I à 285,2 nm normalisée par l'intensité nette de la raie de Ar I à 763 nm) dans les plantes pour 13 échantillons. Le trait en pointillés représente la droite de régression linéaire entre le signal LIBS et la concentration mesurée. 56

Figure 2-22 : Courbe d'étalonnage pour le phosphore (intensité nette de la raie de P I à 213,6 nm) dans les plantes a) non-normalisée et b) normalisée par l'intensité nette de la raie de Ar I à 763 nm. Le trait en pointillés représente la droite de régression linéaire entre le signal LIBS et la concentration. 56

Figure 2-23 : Courbe d'étalonnage pour le sodium (intensité nette de la raie à 589,6 nm) dans les plantes a) non-normalisée et b) normalisée par l'intensité nette de la raie Ar I à 763 nm. Le trait en pointillés représente la droite de régression linéaire entre le signal LIBS et la concentration. 57

Figure 2-24 : Concentrations en carbone (%masse) mesurées pour les échantillons de plantes. 58

Figure 2-25 : Courbe d'étalonnage obtenue sur les mesures XRF en fonction des concentrations mesurées par ICP-AES dans les plantes pour a) le calcium et b) le potassium. Les cercles rouges représentent l'échantillon outlier 165. L'ellipse verte indique la présence d'un gradient en concentrations. 65

Figure 3-1 : Diagramme ternaire des textures pour les sols de cette étude issu du réseau QUASAPROVE. 68

Figure 3-2 : Spectre LIBS typique d'un sol (texture équilibrée) du réseau QUASAPROVE, avec l'attribution des éléments d'intérêt pour les raies associées. 69

Figure 3-3 : Graphe des loadings a) pour la totalité des variables et b) pour deux sous-ensembles de variables dans les gammes spectrales 380- 440 nm, et 750- 780 nm de la composante 3 révélant le décalage en longueur d'onde. Le rectangle en transparence orange indique la zone spectrale impactée par le décalage. 71

Figure 3-4 : Extraits de spectres LIBS de deux échantillons de sols illustrant la dérive en longueur d'onde du pistolet Z300. 72

Figure 3-5 : Graphe du pourcentage de variance expliquée en fonction du nombre de composantes pour le modèle ACP calculé sur la gamme 188-350 nm. 73

Figure 3-6 : Graphe des loadings de la 1^{ère}, 3^{ème} et 4^{ème} composante du modèle ACP calculé sur la gamme 188-350 nm. 74

Figure 3-7 : a) Graphe des scores des composantes 1-3 du modèle ACP calculé à partir des spectres LIBS des pastilles de sols et b) diagramme ternaire de la composition chimique des sols construit à partir des concentrations mesurées par ICP-AES. 75

Figure 3-8 : a) Graphe des loadings de la composante 2 et b) graphe des scores des composantes 1-2 du modèle ACP calculé sur les spectres (188-350 nm) de pastilles de sols. 77

Figure 3-9 : Modèle de régression univariée obtenu sur les sols agricoles en LIBS pour a) le magnésium et b) le fer. 79

Figure 3-10 : Modèle de régression univariée obtenu pour le magnésium dans les sols agricoles en LIBS après a) normalisation SNV et b) normalisation par l'aire totale. 80

Figure 3-11 : Modèle de régression univariée obtenu pour le calcium dans les sols agricoles en LIBS a) sans normalisation et b) après normalisation SNV. Les ellipses rouges indiquent les échantillons considérés comme outliers. 81

Figure 3-12 : Schéma permettant de mettre en évidence l'augmentation de la linéarité de la régression univariée à partir de la raie du calcium à 315,9 nm lorsque l'on utilise une longueur d'onde associée à l'aile de la raie (curseur vert) au lieu de celle attribuée à l'intensité maximale (curseurs rouge et orange). 83

Figure 3-13 : Modèle de régression univariée final obtenu pour le calcium dans les sols agricoles par LIBS. 83

Figure 3-14 : Schéma de principe de l'analyse univariée et multivariée. 85

Figure 3-15 : Tableau de bord des figures de mérite pour le modèle PLS du magnésium calculé sur tout le spectre avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) la régression entre la concentration mesurée par la méthode de référence (ICP-AES) et la concentration prédite par le modèle et c) le graphe du poids des variables associé à la première composante. Les ellipses rouges représentent les zones jugées « anormales » sur les figures de mérite et sont discutées dans le texte. 89

Figure 3-16 : Schéma de principe de la sélection de variables en analyse PLS. 90

Figure 3-17 : Tableau de bord des figures de mérite pour le modèle PLS final du magnésium avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) la régression entre la concentration mesurée par la méthode de référence (ICP-AES) et la concentration prédite par le modèle pour quatre composantes et c) le graphe des coefficients associé à la quatrième composante. 92

Figure 3-18 : Tableau de bord des figures de mérite pour le modèle PLS à 15 composantes avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) la

régression entre la concentration mesurée par la méthode de référence (ICP-AES) et la concentration prédite par le modèle à 15 composantes et c) le graphe des coefficients associé à la quatrième (noir) et quinzième composante (rouge). 94

Figure 3-19 : Modèle de régression obtenu pour le magnésium après analyse LIBS des échantillons de sols agricoles a) univarié et b) PLS à 4 composantes avec sélection de variables. 95

Figure 3-20 : Tableau de bord des figures de mérite pour le modèle PLS final du fer à 4 composantes avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) la régression entre la concentration mesurée par la méthode de référence (ICP-AES) et la concentration prédite par le modèle PLS à 4 composantes et c) le graphe des coefficients associé à la quatrième composante. 96

Figure 3-21 : Tableau de bord des figures de mérite pour le modèle PLS final du sodium à 4 composantes avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) la régression entre la concentration mesurée par la méthode de référence (ICP-AES) et la concentration prédite par le modèle PLS à 4 composantes et c) le graphe des coefficients associé à la quatrième composante. 97

Figure 3-22 : Figures de mérite pour le modèle PLS du potassium calculé en prenant en compte l'entièreté du spectre LIBS avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) le graphe des coefficients pour la composante 6. 98

Figure 3-23 : Graphe de la concentration en sodium en fonction de la concentration en potassium mesurées par ICP (g/kg). 98

Figure 3-24 : Modèle univarié du calcium pour les pastilles de sols agricoles en utilisant : a) le système XRF portable et b) le pistolet LIBS. 105

Figure 3-25 : Modèle d'étalonnage du magnésium pour les pastilles de sols agricoles en utilisant : a) le système XRF portable et b) le pistolet LIBS. 105

Figure 3-26 : Modèles d'étalonnage du calcium et du fer pour les pastilles de sols agricoles en utilisant : a) le système XRF portable et b) en utilisant la LIBS portable. 106

Figure 3-27 : Modèle d'étalonnage XRF de Mn et Zn pour les pastilles de sols agricoles en utilisant le système XRF portable. 106

Figure 4-1 : Exemples de spectres LIBS obtenus sur une feuille de blé fraîche prélevée sur une parcelle de Qualiagro. 110

Figure 4-2 : Photo d'une feuille de blé réalisée après analyse LIBS. Les rectangles rouges indiquent les 2 zones d'analyse LIBS effectuées. 111

Figure 4-3 : Photo de la fenêtre de tir lors de l'analyse LIBS des feuilles de blé avec le Z-300. 111

Figure 4-4 : photo d'un même sol directement prélevé du terrain dans deux zones différentes, avec deux taux d'humidité différents. 114

Figure 4-5 : vue de la surface de la pastille du sol 538 après une étape de tamisage à 2 mm) à gauche et finement broyée ($< 250 \mu\text{m}$) à droite. Encadré blanc : particule minérale de type feldspath non broyée. 115

Figure 4-6 : photo représentant le passage de la poudre de sol 538 à la pastille à l'aide de la presse manuelle. 116

Figure 4-7 : Graphe des effets pour le silicium (Si I à 288,2 nm) obtenu à la suite du plan d'expériences réalisé sur le sol 538. 118

Figure 4-8 : Graphe des effets pour le calcium obtenu à la suite du plan d'expériences réalisé sur le sol 538. 119

Figure 4-9 : Comparaison entre le spectre moyen d'une zone de la surface de la pastille avec agrégat (trait pointillé rouge) et le spectre moyen d'une zone en dehors d'un agrégat (trait noir). 120

Figure 4-10 : Graphique des interactions pour le silicium obtenu à la suite du plan d'expérience réalisé sur le sol 538. 122

Figure 4-11 : Graphiques utilisés pour évaluer la pertinence des résultats issus du plan d'expériences calculé sur l'intensité de la raie du silicium (288,2 nm) avec a) le graphe d'auto-corrélation, b) le Q-Q plot et c) les valeurs standardisées vs les valeurs prédites. 123

Figure 4-12 : Graphiques montrant l'évolution des intensités nettes de la raie de C à 193 nm et de l'aire nette totale sous le spectre en fonction du pourcentage d'humidité contenu dans la pastille de sol. Les barres d'erreur représentent l'écart-type de l'intensité calculé à partir des 5 zones d'analyse réalisées par taux d'humidité. 126

Figure 4-13 : photos de la fenêtre de tir lors de l'analyse LIBS d'un sol présentant un taux d'humidité de a) 15% et b) 30%. 127

Figure 4-14 : Intensités nettes de plusieurs raies élémentaires (Si I à 288 nm, Al I à 309 nm, C I à 193 nm et K à 766 nm) en fonction du taux d'humidité contenu dans la pastille de sol a) non normalisées et b) normalisées SNV. 127

Table des figures - Annexe

Figure 1-1 : Photo de la presse manuelle hydraulique utilisée pour le pastillage des échantillons de sol.	136
Figure 1-2 : Carte de contrôle de l'étendue pour l'instrument Z-300.	137
Figure 1-3 : Plan de Box-Behnken pour 3 facteurs.	137
Figure 2-1 : Graphe des scores sur la composante 4 et 6 de l'ACP réalisée sur les échantillons de plantes.	143
Figure 2-2 : Graphe des loadings zoomé de la composante 7 (trait noir) et 8 (trait rouge) de l'ACP réalisée sur les échantillons de plantes.	143
Figure 2-3 : Comparaison entre deux spectres zoomés de plante présentant soit des raies de Ca correctement définies (trait rouge) soit des raies auto-absorbées (trait noir). Les largeurs à mi-hauteur associées sont également indiquées.	144
Figure 2-4 : Spectres des échantillons de plantes zoomés sur une plage spectrale ne comprenant que des raies du fer.	144
Figure 2-5 : Spectres d'un échantillon de plante zoomé sur le doublet d'Al I interféré par le doublet de Ca II.	145
Figure 2-6 : Courbe d'étalonnage du fer obtenue sur les mesures XRF en fonction des concentrations mesurées par ICP-AES dans les plantes.	146
Figure 2-7 : Courbe d'étalonnage du phosphore obtenue sur les mesures XRF en fonction des concentrations mesurées par ICP-AES dans les plantes.	146
Figure 2-8 : Courbe d'étalonnage de l'aluminium obtenue sur les mesures XRF en fonction des concentrations mesurées par ICP-AES dans les plantes.	147
Figure 3-1 : Texture dominante de l'horizon supérieur des sols agricoles français ainsi que le diagramme ternaire associé [79].	148
Figure 3-2 : Graphe des scores de l'ACP calculée sur les spectres LIBS moyens des pastilles de sols pour a) la composante 1 et 4 et b) la composante 3 et 4.	149
Figure 3-3 : Diagramme ternaire des textures pour les sols de l'étude proposée par Gouegel et al en 2019 [82].	150
Figure 3-4 : Modèle de régression univariée obtenu sur les sols agricoles en LIBS pour a) l'aluminium et b) le potassium.	151
Figure 3-5 : Auto-absorption des raies du potassium pour des spectres LIBS de sols observée à travers a) le doublet de K I à 766 et 769 nm et b) le graphe des coefficients zoomé sur le même doublet du potassium.	151

Figure 3-6 : Figures de mérite de la PLS calculée pour le fer en prenant en compte l'entièreté du spectre LIBS comme variables d'entrée avec a) l'évolution de la RMSECV (g/kg) en fonction du nombre de composantes et b) la régression entre la concentration mesurée en Fe par ICP et la concentration prédite par le modèle PLS pour 4 composantes. Le cercle vert indique le nombre de composantes jugé optimal pour construire le modèle et les cercles rouges entourent les échantillons considérés comme outliers. 152

Figure 3-7 : Figures de mérite de la PLS calculée pour le sodium en prenant en compte l'entièreté du spectre LIBS comme variables d'entrée avec a) l'évolution de la RMSECV (g/kg) en fonction du nombre de composantes et b) la régression entre la concentration mesurée en Na par ICP et la concentration prédite par le modèle PLS pour 4 composantes. Le cercle vert indique le nombre de composantes jugé optimal pour construire le modèle et les cercles rouges entourent les échantillons considérés comme outliers. 152

Figure 3-8 : Diagramme ternaire des textures pour les sols de cette étude issu du réseau QUASAPROVE avec en noir les échantillons de calibration et en rouge les échantillons de validation. 153

Figure 4-1 : Exemple de feuille de blé sélectionnée sur une parcelle de blé de Feucherolles pour analyse LIBS in-situ. 154

Figure 4-2 : photo de l'instrument LIBS Z-300 monté sur un support lors de la campagne de mesures réalisée sur des parcelles de blé Qualiagro. 154

Figure 4-3 : Comparaison entre un spectre LIBS zoomé d'une plante sous la forme d'une pastille sèche (trait noir) et d'une plante fraîche sans préparation (trait rouge). 155

Figure 4-4 : Comparatif des spectres LIBS d'une pastille du même sol analysée sous 3 conditions différentes d'humidité et zoomés sur quatre raies élémentaires, à savoir a) le carbone, b) le silicium, c) l'aluminium et d) le potassium. 157

Liste des tableaux

Tableau 1-1 : Numéro et nature des échantillons utilisés pour construire les modèles d'étalonnage pour les plantes. 9

Tableau 1-2: Caractéristiques du système Z-300 de SciAps. 17

Tableau 2-1 : Récapitulatif des concentrations mesurées par ICP-AES ou ICP-MS pour les échantillons certifiés du WEPAL et des concentrations « théoriques » observées en milieu naturel. 29

Tableau 2-2 : Valeurs de concentration mesurées par ICP-AES et prédites par LIBS des éléments Ca, Fe, Mg, Al, K, Na et P pour les 8 échantillons de validation. Les incertitudes représentent l'écart-type des valeurs prédites à partir de 3 réplicats. 59

Tableau 2-3 : Tableau récapitulatif des performances analytiques obtenues pour les échantillons de plantes avec l'instrument LIBS portable Z-300. 63

Tableau 3-1 : Récapitulatif des concentrations mesurées par ICP-AES dans les échantillons de sols issus des réseaux internes de l'INRAE et des concentrations « théoriques » observées en milieu naturel. 69

Tableau 3-2: Récapitulatif des prédictions obtenues avec les modèles PLS de Fe, Mg et Na pour les 11 échantillons de validation. L'abréviation HG est utilisée pour les échantillons qui présentent une concentration hors gamme d'étalonnage. 99

Tableau 3-3: Récapitulatif des performances des modèles multivariés construits pour les sols. 103

Tableau 4-1 : Plan factoriel complet défini pour l'étude de l'influence de paramètres extérieurs sur le signal LIBS. 116

Liste des tableaux - Annexe

Tableau 1-1 : Tableau de correspondance entre la taille de l'échantillon et la valeur du facteur à utiliser pour calculer les limites de contrôle de la carte de contrôle. 136

Tableau 1-2 : Plan d'expérience de Box-Benkhen construit pour définir le nombre de mesures à effectuer sur la surface d'une pastille de sol. 138

Tableau 2-1 : Numéro et nature des échantillons utilisés pour construire les modèles d'étalonnage pour les plantes. 142

Tableau 2-2 : Récapitulatif des éléments détectés par le Z-300 dans les plantes et les longueurs d'onde associées (nm), attribuées à l'aide de la base de données NIST et la base données AtomTrace. 142

Tableau 2-3 : Numéro et nature des échantillons utilisés dans le jeu de validation utilisé pour valider les modèles d'étalonnage. 145

Tableau 3-1 : Récapitulatif des éléments détectés par le Z-300 dans les sols et les longueurs d'onde associées (nm), attribuées à l'aide de la base de données NIST. 148

Tableau 3-2 : Correspondance entre les noms donnés aux échantillons de sols agricoles utilisés pour la validation des modèles PLS et leur labellisation dans le réseau QUASAPROVE. 153

Tableau 4-1 : Résultats expérimentaux obtenus pour le plan d'expérience réalisé à partir de la raie du silicium à 288,2 nm du spectre LIBS du sol 538 (QUASAPROVE). 155

Tableau 4-2 : Résultats expérimentaux obtenus pour le plan d'expérience réalisé à partir de la raie du calcium à 315,9 nm du spectre LIBS du sol 538 (QUASAPROVE). 156

Tableau 4-3 : Correspondance entre le volume d'eau pure ajouté sur les pastilles de sol de cressonnière et le taux d'humidité (%). 156

Introduction

Les écosystèmes terrestres sont de plus en plus sollicités pour faire face à une demande croissante de biomasse alimentaire et non-alimentaire de la population mondiale en constante augmentation. Par ailleurs, les derniers rapports du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) [1], mettent en évidence les conséquences néfastes d'un réchauffement planétaire de 1,5°C d'ici 2050, qui entraînerait une chute considérable des rendements des cultures, mais aussi le rôle crucial des sols pour lutter contre le changement climatique, puisqu'ils représentent un puits de carbone permettant l'absorption de 30 % des émissions humaines de CO₂ [1]. Dans un contexte de changement global du climat, mais aussi de raréfaction de certaines ressources minérales telles que le phosphore [2], l'optimisation de la gestion des ressources agricoles, et en particulier des éléments minéraux des sols, est devenu un enjeu majeur [3]. En effet, les éléments minéraux sont à la fois des ressources indispensables pour la croissance des végétaux et le stockage du carbone, des polluants potentiels de certains compartiments de l'environnement, et pour certains éléments, des contaminants potentiellement toxiques susceptibles d'être transférés depuis les végétaux jusqu'à l'homme.

La croissance des plantes est conditionnée par l'absorption de différents éléments nutritifs que l'on peut classer en plusieurs catégories [4, 5] : 1) les éléments majeurs ou macroéléments présents en quantité importante et essentiels (indispensables pour la croissance et le développement de la plante, dont la fonction ne peut être remplacée par un autre élément, avec une concentration dans la matière sèche comprise entre 0.1% et 4%) dans les plantes, tels que N, P, K, Ca, Mg, S ; les éléments C, H et O sont également considérés comme des macroéléments mais sont puisés directement dans l'air par la plante ; chaque élément essentiel joue un rôle déterminant dans le bon développement de la plante, comme par exemple au niveau de la synthèse de la chlorophylle (N, P, K, Mg, S, Cl) mais également pour d'autres processus comme la respiration cellulaire (P), le maintien du port de la plante (K, Mg) et des parois cellulaires (K, Ca, Mg) ou encore pour la croissance des jeunes racines (Ca) [5] ; 2) Les micronutriments, ou éléments mineurs, aussi appelés oligoéléments, sont des éléments importants dans la nutrition globale de la plante, présents en quantité très faible (concentration dans la matière sèche inférieure à 0,01%) : B, Cl, Cu, Fe, Mn, Mo et Zn ; ils sont nécessaires, mais deviennent toxiques pour les plantes au-delà d'un certain seuil ; 3) des éléments utiles non nécessaires (Na, Si, Co, Al, Se, Ti), uniquement pour certains types de plantes ou toxiques à partir d'une certaine concentration ; 4) des éléments traces non nécessaires et toxiques (ex : Cd, Pb, Cr, As), contaminants des plantes ou de l'environnement. Tous ces éléments sont principalement absorbés dans le sol par les racines et ont pour origine les processus de

formation des sols, mais également l'activité humaine, notamment les intrants agricoles (fertilisants minéraux, produits organiques recyclés).

Ainsi il est essentiel de pouvoir connaître au mieux la composition des sols et des végétaux, afin de proposer des pratiques de fertilisations qui puissent i) prévenir les phénomènes de carence et les phénomènes de toxicité qui inhiberaient le bon développement des plantes et impacteraient les productions végétales, à la fois en quantité et en qualité, ii) favoriser le stockage de carbone dans les sols pour contribuer à la réduction des émissions en CO₂, et iii) préserver l'environnement et la santé.

La mesure des concentrations de ces éléments nutritifs est réalisée en laboratoire par des méthodes analytiques telles que la spectrométrie d'absorption atomique (AAS) et plus généralement par spectrométrie à plasma à couplage inductif (ICP) [6]. La méthode ICP est une méthode de spectroscopie d'émission atomique (ICP-AES) permettant de réaliser une analyse multi-élémentaire, à des limites de détection de l'ordre du mg/kg au µg/kg, avec peu de problèmes d'interférences spectrales, pour les échantillons de sols et de plantes. Cette technique performante, s'est imposée comme la méthode de référence pour l'analyse des matrices agronomiques. En complément, il est parfois recommandé d'utiliser un instrument de spectrométrie de masse (ICP-MS) afin de pouvoir atteindre des limites de détection plus basses pour les éléments plus faiblement concentrés dans les sols ou les plantes, comme le cadmium par exemple (concentrations jusqu'au µg/kg). Cependant, bien que ces techniques soient performantes, elles nécessitent préalablement d'importantes campagnes d'échantillonnage sur le terrain, puis une préparation en laboratoire avant analyse particulièrement fastidieuse comprenant plusieurs étapes : tamisage (pour les sols), séchage, broyage, minéralisation de l'échantillon à l'aide de solutions d'acides concentrés. En résumé, la lourdeur, le coût et les délais nécessaires est un facteur limitant l'acquisition intensive de données multi-élémentaires. D'un point de vue opérationnel, il est difficile dans ces conditions de réaliser un diagnostic rapide sur le terrain d'une situation de carence ou de contamination, ou encore de caractériser la variabilité spatio-temporelle d'une parcelle agricole.

L'émergence récente de capteurs et de systèmes portatifs a permis de développer de nouvelles méthodes pour mesurer par exemple les propriétés du sol et compléter ou remplacer les techniques conventionnelles de laboratoire utilisées pour différents types d'analyses [7]. Parmi ces méthodes, on peut citer par exemple la spectrométrie dans le visible ou le proche infrarouge [8], la XRF (fluorescence aux rayons X) [9, 10], ou encore la LIBS (Laser-Induced Breakdown Spectroscopy, ou spectroscopie de plasma induit par laser) [11, 12], qui semblent donc promettre une avancée pour l'analyse élémentaire *in situ* rapide des sols et des plantes dans le cadre d'études agronomiques et/ou environnementales.

La LIBS est une méthode de spectroscopie d'émission atomique à partir d'un plasma créé par laser. Le laser impulsionnel, focalisé sur la surface de l'échantillon, entraîne la formation d'un plasma au sein duquel les ions et les atomes constitutifs du matériau se retrouvent dans des états excités. En se désexcitant, ils émettent un rayonnement dont le spectre est caractérisé par les raies d'émission atomiques associées aux différents chemins de désexcitation. Ce rayonnement est alors analysé à l'aide d'un spectromètre optique [13, 14] dans une gamme spectrale allant de l'ultraviolet jusqu'au proche infrarouge. La LIBS étant une méthode d'analyse rapide, elle permet a priori d'enregistrer un très grand nombre de spectres par échantillon et ainsi de reconsidérer la stratégie d'échantillonnage des parcelles agricoles. De plus, c'est une méthode

analytique que l'on peut qualifier de « verte », car sa mise en œuvre ne s'appuie sur aucun réactif chimique dangereux et qu'elle ne laisse aucun déchet. La LIBS est donc tout à fait adaptée à l'analyse de matrices agronomiques, tout particulièrement depuis la commercialisation de dispositifs portables performants [15]. La LIBS a été exploitée dans un grand nombre de domaines applicatifs [16] comme la métallurgie dans le but de contrôler en ligne la fabrication d'alliages ou de trier des matériaux, dans l'analyse d'objets issus du patrimoine culturel (peintures, métaux, céramiques) pour assurer sa conservation, dans la détection de pollution par les métaux lourds dans des échantillons environnementaux parmi d'autres [13, 16, 17]. L'exemple le plus connu du grand public est certainement l'instrument ChemCam [18], monté sur le rover Curiosity, qui analyse les sols et roches de la planète Mars à plusieurs mètres de distance. Enfin, la maturité de la LIBS qui a connu une croissance extraordinaire ces dernières années, a aussi permis d'enclencher une révolution en imagerie que ce soit pour des applications biomédicales, géologiques ou encore industrielles [19].

La capacité de la LIBS à produire des résultats quantitatifs fiables a constitué un débat au sein de la communauté scientifique pendant plusieurs années. En effet, la capacité de la LIBS à établir des résultats quantitatifs a longtemps été considérée comme son talon d'Achille [20]. Cette soi-disant faiblesse de la LIBS peut aujourd'hui être interprétée de différentes façons. D'une part, de nombreux dispositifs LIBS de laboratoire ne possédaient pas les propriétés nécessaires pour atteindre des performances analytiques suffisantes. Ceci était lié notamment à l'absence de contrôle des paramètres expérimentaux qui pouvaient ainsi subir des dérives sans que celles-ci soient identifiées, ni prises en compte. Ces dernières années, des systèmes LIBS de laboratoire ont intégré des outils de contrôle de l'instrument et ont permis d'atteindre des performances quantitatives très satisfaisantes. Les dispositifs LIBS de laboratoire commercialisés s'appuient eux aussi désormais sur ces contrôles pour offrir des conditions d'analyse robustes. Une fois réglée la question du contrôle de l'instrumentation, les résultats d'analyses LIBS peuvent encore être dégradés par ce qu'on appelle les effets de matrice. En effet, la LIBS est le plus souvent mise en œuvre pour analyser la surface d'un échantillon sans préparation particulière de ce dernier. Or, un échantillon donné fournira des spectres LIBS différents selon qu'il est plus ou moins compacté (poudre vs. pastille) ou encore qu'il est plus ou moins gorgé d'eau, pour ne citer que ces exemples. Ceci provient du fait que la première étape de l'analyse est l'ablation laser et que ce processus est très sensible aux propriétés physiques du matériau analysé. Une ablation différente entraîne un plasma différent et par conséquent un spectre LIBS différent.

Un grand nombre d'articles a été publié ces vingt dernières années dans le but de compenser ces effets de matrice, notamment par la mesure de paramètres physiques, jugés représentatifs de la qualité du plasma [21], comme le signal acoustique créé par l'onde de choc du laser sur le matériau [22] ou encore la quantité de matière ablatée [23]. Plusieurs stratégies de normalisation des spectres LIBS ont aussi été proposées [24] pour tenter de limiter l'impact des effets de matrice sur les résultats d'analyses. Ce sujet reste ouvert encore aujourd'hui car il a été constaté qu'aucune méthode de normalisation ne permettait de corriger les spectres LIBS de manière universelle. Cela implique de traiter des données pour lesquelles le processus de normalisation fait encore débat. On s'attend donc à des modèles d'étalonnage moins performants lorsqu'on fait face à des effets de matrice. Ajoutons à cela le fait que les instruments LIBS portables comme celui que nous avons mis en œuvre dans le cadre de cette étude sont potentiellement exposés à des dérives comme par exemple une dérive de l'étalonnage en longueur d'onde au cours d'une campagne d'analyses et on comprend alors pourquoi l'analyse LIBS par

construction d'un modèle d'étalonnage reste encore aujourd'hui un travail de recherche pour prendre en compte les spécificités liées à l'instrument, aux conditions de mesures et à la nature des échantillons.

En plus de l'approche par étalonnage, il existe aussi une approche dite sans étalonnage connue sous le nom de CF-LIBS pour *calibration-free* LIBS [25]. La CF-LIBS se base sur le fait que l'émission du plasma est très sensible aux changements de matrice de l'échantillon et que ces changements peuvent être parfaitement estimés à l'aide de l'analyse du spectre LIBS lui-même. Cependant, le plasma reste un objet très complexe et l'algorithme derrière la CF-LIBS est basé sur des hypothèses simplifiées pour décrire son état, à savoir : i) l'ablation laser est stœchiométrique, c'est-à-dire que la composition du plasma est représentative de la composition de l'échantillon non perturbé ; ii) le plasma est en équilibre thermodynamique local (LTE), c'est à dire dans la fenêtre spatiale et temporelle observée ; iii) le plasma peut être modélisé comme une source spatialement homogène et iv) les raies spectrales impliquées dans le calcul doivent être bien définies, c'est-à-dire libres de toute interférence ou de phénomène d'auto-absorption. Malgré des résultats particulièrement encourageants sur certaines matrices, et notamment les alliages métalliques, l'utilisation de la CF-LIBS n'a pas encore été généralisée à tous types de matrices et fait encore aujourd'hui l'objet de travaux de recherche [26]. Nous n'aborderons pas dans ce mémoire l'approche CF-LIBS à laquelle nous avons préféré l'approche avec étalonnage.

La méthode LIBS est par ailleurs pénalisée par un manque de standardisation des traitements des données appliqués aux spectres. Ce problème a été parfaitement mis en lumière par le biais d'un essai inter laboratoire qui a fait l'objet d'un article publié en 2018 par Motto-Ros et al. [27]. L'étude montre qu'en exploitant les mêmes spectres LIBS, et plus précisément la même raie de l'aluminium, 8 participants annoncent des valeurs de concentrations prédites en aluminium très différentes pour deux échantillons de verre. La source de ces variations est attribuée à la manière dont sont traités les spectres LIBS, différente pour chaque participant et plus précisément en faisant référence à la méthode de soustraction de la ligne de base, d'extraction de l'intensité de la raie considérée et de la construction du modèle de régression. Ce constat n'est en réalité pas si surprenant lorsqu'on sait que les utilisateurs de la LIBS viennent d'horizons variés (géologues, physiciens, chimistes...) et n'ont pas tous les mêmes exigences ni la même méthodologie pour traiter et interpréter leurs spectres. Alors que cette analyse critique concernait le cas de l'analyse univariée dans le cadre d'une gamme étalon exempte d'effets de matrices, d'autres études ont très largement utilisé des outils de chimiométrie pour prendre en compte ces effets de matrice [28] mais les articles associés sont souvent insuffisamment décrits et laissent penser que certains utilisateurs ne maîtrisent probablement pas les compétences pour paramétrer les modèles statistiques et les outils numériques utilisés, ou encore qu'ils s'en servent dans certains cas comme de simples boîtes noires [29]. Ces travaux transmettent malheureusement l'usage de mauvaises pratiques au sein de la communauté LIBS qui se trouve dans l'incapacité de comparer différents modèles par manque de méthode et de figures de mérite adaptées.

La LIBS suscite maintenant un intérêt croissant pour les applications concernant l'agriculture, avec notamment des études concernant entre autres la détermination du carbone, du phosphore, de l'azote et des éléments traces dans les sols agricoles [11, 30], les végétaux [10, 12, 31] ou les fertilisants [11, 32]. Les performances analytiques et particulièrement quantitatives de la LIBS s'améliorent sur ces échantillons particuliers, qui peuvent être complexes à analyser du

fait de la grande variété de leurs matrices [33]. Cependant peu d'études couvrent à la fois une large gamme d'échantillons et un large spectre d'éléments minéraux représentatifs de conditions réelles dans le domaine de l'agriculture. En effet, une des difficultés consiste à la fois à analyser des éléments majeurs (de l'ordre du g/kg) et des oligo-éléments et éléments traces (de l'ordre du mg/kg), à la fois pour les sols [34] et les végétaux [35]. Si la portabilité de la LIBS fait de cette méthode d'analyse une méthode potentielle de choix pour des applications agronomiques, et notamment pour le développement de l'agriculture de précision, très peu d'instruments portables ont été utilisés dans ces conditions. L'utilisation d'un instrument LIBS directement sur le terrain, pour des diagnostics rapides de déficience ou de contamination serait donc une avancée méthodologique importante, en rupture avec l'approche traditionnelle consistant à prélever des échantillons sur le terrain puis à les analyser en laboratoire. Cela ouvrirait également des perspectives pour prendre en compte la variabilité spatio-temporelle et établir des cartographies de la composition multi-élémentaires de parcelles, permettant de définir ensuite un échantillonnage optimal (smart sampling) pour les analyses de laboratoires.

L'objectif de l'étude présentée dans ce mémoire est donc d'évaluer les capacités de la méthode LIBS, et plus précisément d'un instrument portable, à réaliser des analyses multi-élémentaires quantitatives d'échantillons présentant un intérêt pour la filière agronomique, à savoir les plantes et les sols agricoles. Le mémoire est organisé de la façon suivante :

- Le **Chapitre I** détaille le soin apporté à la sélection des meilleurs échantillons de sol et de plante, leur préparation, la stratégie d'étalonnage établie pour chaque matrice et les prétraitements des spectres appliqués. Toutes les étapes indispensables à la conduite de campagnes de mesures LIBS de qualité sont ainsi présentées.
- Le **Chapitre II** se focalise sur l'analyse des échantillons de plante en conditions contrôlées de laboratoire. Une grande partie du chapitre est consacrée à la normalisation des spectres LIBS et à l'évaluation des modèles univariés.
- Le **Chapitre III** présente non seulement la méthodologie complète suivie dans le cadre de l'analyse des sols mais aussi les résultats obtenus à partir d'une approche multivariée.
- Le **Chapitre IV** est destiné à évaluer les performances de l'instrument portable lorsque les conditions idéales de préparation des échantillons ne sont pas satisfaites. L'impact des facteurs extérieurs sur le signal LIBS évalué à l'aide d'un plan d'expériences sera présenté et les corrections possibles seront présentées.

Chapitre I

Démarche analytique générale appliquée à des matrices agronomiques : sols et plantes

Ce travail a pour objectif de tester les performances de la méthode LIBS pour l'analyse multi-élémentaire de matrices agronomiques, principalement sols et végétaux, sur un plan à la fois qualitatif et quantitatif. Cette partie en détaille le contexte analytique. La construction de modèles quantitatifs performants, quelle que soit la matrice étudiée, passe par une sélection pertinente des échantillons d'étalonnage. Les corpus de sols et de végétaux sélectionnés ainsi que leur processus de préparation avant analyse seront donc présentés de manière détaillée.

Afin de disposer de valeurs de référence pour les concentrations mesurées, les échantillons de sols et de végétaux sont d'abord analysés au laboratoire par ICP-AES, puis par XRF à partir d'un instrument portable afin de disposer d'un second jeu de valeurs de références, correspondant cette fois à des conditions d'analyses de surface, plus proches des conditions d'analyse LIBS.

Puis, après avoir rappelé les principes de la méthode LIBS, nous présenterons ici les spécificités de l'appareil utilisé dans ce travail : il s'agit d'un instrument portable, permettant d'acquérir des données non seulement en conditions de laboratoire mais aussi à l'extérieur du laboratoire.

Nous décrirons les performances de mesure de l'instrument, ainsi que les traitements numériques appliqués aux spectres, permettant de garantir la qualité des analyses LIBS et des résultats obtenus à la suite de l'exploitation du signal.

Enfin, la sélection du meilleur protocole d'échantillonnage et de mesure adapté à l'analyse LIBS des matrices de végétaux et de sols considérés dans notre étude sera, quant à elle, déduite d'un plan d'expérience qui sera également présenté dans ce chapitre. De plus, sachant que le traitement des données LIBS représente une étape cruciale qui divise la communauté des analystes LIBS [27], nous avons décidé de présenter les étapes de traitement des spectres que nous avons choisies d'adopter de manière très détaillée. Plus précisément, nous développerons les étapes de prétraitement permettant de disposer de spectres prêts à être exploités via des modèles quantitatifs.

1) Corpus

L'analyse chimique des sols est difficile, quelle que soit la méthode d'analyse utilisée : les méthodes de routine comme l'ICP-AES nécessitent des digestions acides chronophages, la LIBS peut, elle, être impactée négativement par les variations de la taille du grain d'un sol à l'autre, générant des fluctuations importantes d'intensité pour les spectres obtenus [36]. Ce phénomène fait partie de ce qu'on appelle les « effets de matrice ». L'analyse des plantes par LIBS semble plus simple et moins sujette à ce type de fluctuations. Les stratégies d'étalonnage sélectionnées pour ces deux matrices sont donc bien distinctes et la façon de choisir les échantillons associés également. Plusieurs corpus ont donc été sélectionnés afin de répondre à des problématiques différentes. En effet, dans un premier temps, des échantillons ont été sélectionnés spécialement afin de caractériser qualitativement et quantitativement la matrice des sols agricoles et des plantes en conditions d'analyse optimales, c'est-à-dire en laboratoire avec des échantillons préparés. Dans un second temps, un autre corpus sera exploité dans le but d'évaluer la capacité du pistolet Z-300 à analyser des échantillons dans des conditions dégradées par rapport aux conditions idéales.

i. Sélection des sols agricoles

Afin de construire en laboratoire des modèles quantitatifs de prédiction des teneurs totales en éléments dans les sols, un premier corpus d'échantillons a été utilisé. Les sols sélectionnés pour notre étude sont issus du réseau Quasaprove, réseau de suivi de parcelles agricoles sur toute la France exploitées pour l'agriculture conventionnelle et biologique, avec des productions végétales de grandes cultures en plein champ. Il s'agit d'un observatoire de la contamination diffuse des grandes cultures. Il est constitué de 97 parcelles de blé dur, blé tendre et tournesol qui ont été suivies sur 33 sites, de 2010 à 2015, répartis sur l'ensemble du territoire français. L'étude à l'échelle nationale permet de couvrir la diversité des systèmes agricoles et de mettre en évidence des situations à risque pour mieux appréhender la contamination des cultures. Ce réseau est piloté par l'Association de Coordination Technique Agricole (ACTA), association gérée par les agriculteurs, et par l'unité ISPA (INRAE Bordeaux). Plus d'informations sur le réseau Quasaprove sont données sur le site officiel [37].

Ces sols, essentiellement limoneux et argilo-limoneux, sont représentatifs des niveaux de fertilité et de contamination diffuse rencontrés en grande culture [5]. La couche de surface des sols (0-30 cm) a été prélevée chaque année et un sous-échantillon de chaque échantillon obtenu pour un site a été analysé pour déterminer les caractéristiques pédologiques et de nombreuses teneurs totales en éléments, dont les éléments d'intérêt pour ce projet : le carbone (C), les éléments majeurs (N, Ca, K, Mg, P) et les éléments traces nécessaires ou toxiques (Al, Cd, Cu, Fe, Na, Zn, Mn) pour la nutrition des plantes. Après prélèvement, séchage et tamisage à 2mm, puis analyse par les méthodes de références, ces sols sont conservés dans la terrothèque de l'unité ISPA. La collection d'échantillons de sols utilisée pour cette étude est un sous-ensemble constitué de 62 échantillons : 51 sélectionnés afin de construire les modèles d'étalonnage des différents éléments d'intérêt et 11 autres, dits de validation et non utilisés dans la construction des modèles d'étalonnage, pour valider les performances analytiques de ces derniers. Ces échantillons ont été choisis selon leur disponibilité, le nombre et la nature des éléments analysés de manière fiable et la gamme des concentrations balayée. Une carte présentant la localisation géographique des échantillons sélectionnés est présentée en Figure 1-1 : ces échantillons

proviennent de sites répartis dans presque toutes les régions de France métropolitaine. Lors de la séparation des 62 échantillons en un jeu de calibration (51) et un jeu de validation (11), il a été vérifié que chaque sous-ensemble couvre une zone géographique aussi large que possible. La composition chimique de ces sols sera discutée plus loin (cf chapitre III), lors de la construction des modèles d'étalonnage. Avec un tel corpus d'échantillons, on prétend pouvoir construire des modèles quantitatifs utilisables largement pour des parcelles de grandes cultures en France.

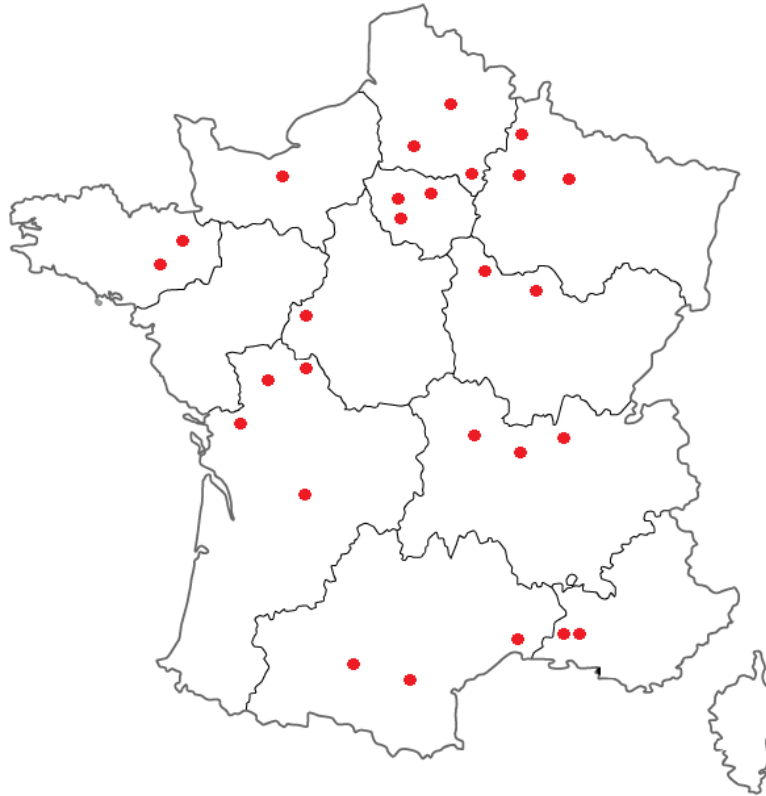


Figure 1-1 : Origine géographique des 62 échantillons de sols sélectionnés dans les analyses LIBS.

ii. Sélection des échantillons de plantes

a) Corpus de calibration

Nous avons sélectionné 19 échantillons de plante issus des programmes d'évaluation de Wageningen pour les laboratoires analytiques (WEPAL). Le WEPAL, organisation néerlandaise accréditée (ISO/IEC 17043), organise par le biais de la Wageningen Agricultural University et ce plusieurs fois par an, un programme d'évaluation pour les laboratoires d'analyse. Ce programme, nommé l'International Plant-Analytical Exchange (IPE), consiste en l'analyse de plusieurs échantillons de plantes par des laboratoires participants, selon leurs propres procédures, afin de comparer ensuite les valeurs de concentrations totales en éléments mesurés pour ces échantillons de plantes. Le but est d'assurer le contrôle qualité des procédures analytiques liées à ces types d'échantillons, c'est-à-dire d'assurer la production de données analytiques reproductibles et de confiance. Ainsi, 17 échantillons de plante de ce

programme, dont les concentrations élémentaires sont donc connues, ont été sélectionnés pour construire les modèles d'étalonnage dans le cadre de notre étude. Ces échantillons de plantes ont l'avantage de présenter des origines géographiques et des espèces très différentes, allant de la feuille de chêne à la pousse de maïs, en passant par la luzerne. Le tableau 1-1 décrit la diversité d'espèces de ces échantillons, indépendamment de leur origine géographique. Les valeurs de concentrations seront présentées juste avant la construction des modèles d'étalonnage (cf chapitre II). Pour de plus amples informations sur ces échantillons, leurs fiches descriptives (origine et concentrations mesurées par les laboratoires participants) peuvent être retrouvées sur le site du WEPAL [38].

Echantillon	Nature (espèce)
IPE 100	Herbe
IPE124 – 148 - 159	Luzerne
IPE 950	Melon
IPE163	Poivrier
IPE 175	Tubercule de tulipe
IPE 133	Pousse de maïs
IPE 156	Feuille de Chêne
IPE 157	Feuille de Hêtre
IPE 165	Feuille de Palmier
IPE 186	Paille de riz
IPE 130	Pomme de Terre
IPE 172	Laurier de cerise
IPE 207	Herbe à éléphant
IPE 198	Feuille de banane
IPE 199	Bois

Tableau 1-1 : Numéro et nature des échantillons utilisés pour construire les modèles d'étalonnage pour les plantes.

b) Corpus de validation

Une fois les modèles d'étalonnage établis à partir des 17 échantillons décrits dans le Tableau 1-1, ils sont validés à l'aide d'un jeu de 8 nouveaux échantillons de plantes d'espèces végétales variées, naturelles ou cultivées dont les concentrations des éléments d'intérêt ont été mesurées par ICP-AES (cf Tableau 2-3 en annexe du chapitre II). Ces valeurs ont été déterminées par ICP-AES dans le cadre d'une étude menée à l'Université Catholique de Louvain de 2015 à 2019 [39]. Ces échantillons destinés à la validation des modèles seront plus amplement décrits dans le chapitre II consacré à l'analyse de plantes par LIBS.

iii. Echantillons pour l'étude de la faisabilité de mesures LIBS *in situ*

Mesures sur le site expérimental de Qualiagro

Une série de tests préliminaires en conditions *in situ* a été réalisée sur le site expérimental de QualiAgro à Feucherolles (78), faisant partie du réseau SOERE PRO [40]. Il s'agit d'un réseau de sites expérimentaux piloté par l'INRAE en plein champ monitoré sur de longues durées et ayant pour objectif d'observer les effets de l'apport de Produits Résiduaux Organiques (PRO) sur les différents compartiments d'un système de grandes cultures. Ces sites, conçus à l'échelle de la parcelle, permettent de mesurer, depuis plus de 10 ans pour certains, l'évolution sur le long terme des différents compartiments (sol, plantes, eau, air, PRO) d'un agrosystème soumis à épandage. Le site de QualiAgro est divisé en 40 parcelles expérimentales de blé et de maïs, soumises à des apports répétés de compost de différentes natures et de fumier. L'intérêt de ces parcelles expérimentales est donc de pouvoir observer sur le long terme l'effet de ces différents apports répétés sur la croissance de la plante et sur l'évolution de différents indicateurs, dont les teneurs en éléments minéraux. Les parcelles sur lesquelles ont été prélevés les échantillons correspondent à différents types de traitements, chacun répétés sur 4 parcelles différentes : i) témoin (pas d'apports) ; ii) BIO (compost de biodéchets) ; iii) FUM (fumier) ; iv) DVB (compost de déchets verts et boues urbaines) et v) OMR (compost d'ordures ménagères résiduelles), sur une base d'un apport de 4t C/ ha.2ans, les quantités des autres éléments apportés étant dépendant de la composition chimique de chaque PRO.

Un échantillon de sol et de plante représentatif a été extrait pour 20 parcelles différentes. Plus précisément, pour chaque parcelle, 10 feuilles de blé et une carotte de sol de 8 cm de diamètre et de 5 cm de hauteur ont été prélevés, soit un total de 200 feuilles et 20 sols. Pour les plantes, deux zones d'analyse de 10 spots laser avec un tir par spot ont été effectuées directement sur chaque feuille de blé fraîche. Le spectre LIBS moyen de ces 2 zones * 10 spots * 10 feuilles, soit 200 spectres, a été considéré comme le spectre représentatif de l'ensemble de la parcelle. En ce qui concerne les sols, 2 zones d'analyse LIBS de 25 spots laser avec dix tirs par spot ont été effectuées directement sur la surface d'un sous-échantillon de sol. Le spectre LIBS moyen de ces 2 zones * 25 spots * 10 tirs, soit 500 spectres, a été considéré comme le spectre représentatif de l'ensemble de la parcelle. Le protocole ainsi que les conditions d'analyse pour ces échantillons seront détaillés dans le chapitre IV.

Effets des facteurs environnementaux sur les mesures

L'impact de facteurs environnementaux seuls ou combinés (humidité, compaction de la pastille, taille de grain) a été testé en conditions de laboratoire sur un sol du réseau Quasaprove : le sol 538, issu d'une parcelle de blé du Nord-Est de la France et qui présente une texture composée à 50% d'argile, 13% de sables et 37% de limons et une valeur de pH de 8,22 (cf l'étude du plan d'expériences du chapitre IV).

Enfin, l'impact du taux d'humidité dans les sols sur le signal LIBS a été étudié sur des sols de cresson cultivé en plein air et sous serre dans une cressonnière à La Capelle (Arveyres, 33) dans le cadre d'un projet interne à l'INRA. Dans les cressonnières, l'eau qui alimente le cresson circule le long de fossés avec une légère pente, pouvant induire une variation d'humidité des sols entre le haut et le bas de la pente. La culture sous serre induit également des variations

d'humidité. Un échantillon de ces sols de cresson, après extraction sur le terrain, a ensuite été conservé à 4°C au réfrigérateur, puis préparé et pastillé selon les modes opératoires décrits en 2) pour déterminer plus précisément la relation entre humidité du sol et signaux LIBS.

2) Mesures des concentrations de référence

- i. Mesures par ICP-AES
 - a) Principe

L'ICP-AES, à l'instar de la LIBS, est une méthode de spectroscopie d'émission atomique, permettant une analyse multi-élémentaire de la matrice étudiée. Elle s'est imposée comme la méthode de référence dans un bon nombre de domaines d'application [41] lorsqu'il s'agit de déterminer les teneurs totales en éléments, le domaine agronomique ne dérogeant pas à la règle. Cette omniprésence dans les laboratoires d'analyse s'explique par sa limite de détection faible, proche du $\mu\text{g/kg}$, sa robustesse, son caractère multi-élémentaire avec une grande diversité d'éléments détectables, et la grande linéarité de sa gamme de mesure. En effet, en terme de praticité, l'ICP-AES peut analyser à la fois des éléments majeurs et des éléments traces au cours d'une même analyse. L'intégralité des échantillons de l'étude présentée ici (sols et plantes) a donc été analysée, en premier lieu, par ICP-AES ou par ICP-MS mettant en jeu un spectromètre de masse lorsque la sensibilité de l'instrument était insuffisante. Les concentrations mesurées par cette méthode seront ensuite considérées comme les valeurs de référence qui permettront de construire les modèles d'étalonnage par régression à partir du signal LIBS.

L'échantillon, initialement en phase solide, puis placé en solution acide, est injecté dans le système sous la forme d'un aérosol à l'aide d'un nébuliseur, première étape qui est symbolisée par la partie « Introduction de l'échantillon » de la figure 1-2, présentant les composants d'un montage ICP-AES classique. Les fines gouttelettes sont ensuite transportées jusqu'à une torche à plasma à l'aide d'un flux gazeux d'argon, c'est la partie « Transfert d'énergie ». Dans la torche, le plasma est initié par une décharge électrique entre une cathode et une anode puis stabilisé par un champ magnétique. Exactement comme pour la LIBS, le plasma ICP permet d'exciter les atomes et les ions présents en solution. Ces atomes/ions, en retournant à leur état fondamental, émettent des photons caractéristiques des transitions entre les niveaux et ce rayonnement est ensuite analysé à l'aide d'un spectromètre optique (cf figure 1-2) et associé à un spectre de raies d'émissions atomiques dans le domaine UV-visible et dans une moindre mesure dans le proche infrarouge. La différence majeure entre ICP-AES et LIBS réside dans la nature de chaque plasma. Alors que le plasma ICP est stationnaire et parfaitement uniforme, le plasma LIBS est quant à lui hétérogène et transitoire avec une durée d'émission de quelques microsecondes. Il s'en suit une atomisation plus complète en ICP qu'en LIBS ainsi que l'absence d'auto-absorption [42]. La LIBS permet cependant de s'affranchir de la phase de préparation d'échantillon très fastidieuse de l'ICP qui consiste à minéraliser l'échantillon initialement en phase solide.

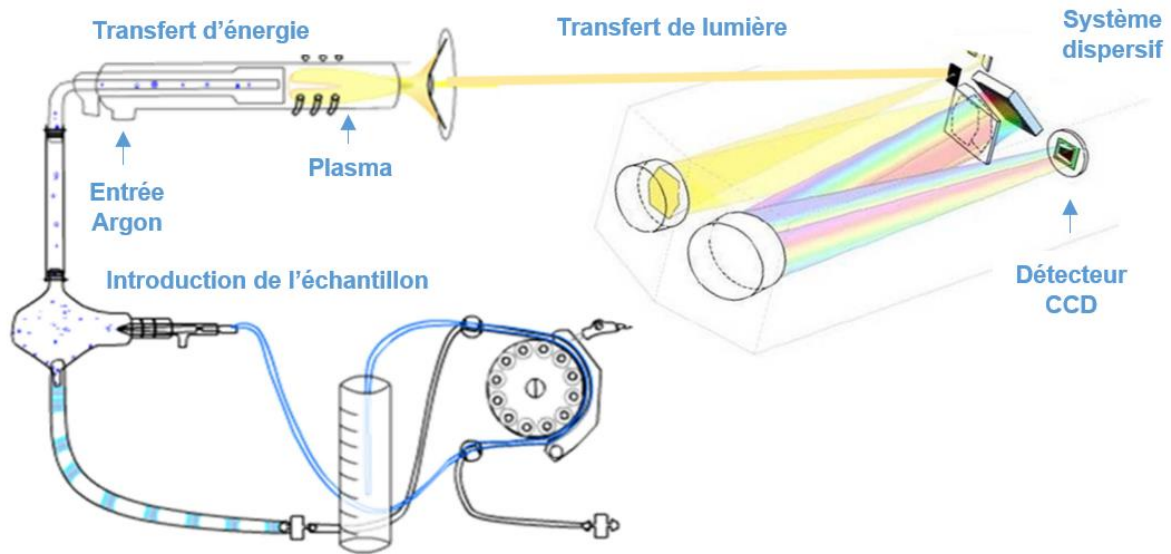


Figure 1-2 : Montage typique d'un dispositif ICP-AES [41].

b) Analyse des concentrations de référence de nos échantillons

Les dosages des éléments totaux des sols du réseau Quasaprove de cette étude ont donc été réalisés au LAS (Laboratoire d'Analyse de Sols d'Arras, INRAE, accrédité COFRAC), selon plusieurs étapes. Le dosage réalisé pour les végétaux a été réalisé par à l'Unité de Service et de Recherche de l'INRAE en Analyses Végétales et Environnementales (USRAVE). Le protocole est sensiblement le même pour l'analyse des végétaux que pour celle des sols si ce n'est que la quantité de matière mise en solution acide est plus importante (1g). L'ensemble des informations liées au dosage des plantes peut être retrouvé sur le site web de l'unité USRAVE (<https://www6.bordeaux-aquitaine.inrae.fr/usrave/Methodes-d-analyse/Vegetaux>).

- Concernant les sols, le protocole de préparation et d'analyse est le suivant :

Premièrement, les échantillons sont préparés selon la norme NF ISO 11464 qui consiste à :

- sécher les échantillons à l'air à une température inférieure à 40 °C
- les tamiser puis les broyer à 2 mm afin de récupérer une "terre fine".
- Un sous échantillon représentatif de cette terre fine est ensuite constitué et broyé à une granulométrie inférieure à 250 µm.

Deuxièmement, 250 mg de ce sous-échantillon sont mis en solution par les acides fluorhydrique (HF) et perchlorique (HClO₄) selon la norme NF X 31147, suite à une calcination à 450°C pendant 4 heures.

Finalement, les solutions acides sont analysées par ICP-AES sauf lorsque la sensibilité de l'instrument est insuffisante, comme dans le cas du Cd ou de Pb, où l'on a alors recours à l'ICP-MS.

Deux exceptions à ce qui vient d'être décrit sont à signaler : le Carbone et l'Azote. Dans les sols, ces éléments sont déterminés par combustion sèche à 1000°C en présence d'oxygène, en partant de 50 mg de sol broyé, puis par séparation par chromatographie en phase gazeuse des gaz formés (CO₂, N₂) et enfin par détection par un catharomètre, selon les normes NF ISO 10694 et NF ISO 13878.

- En ce qui concerne l'analyse des plantes certifiées du WEPAL, venant d'un programme d'évaluation impliquant plusieurs laboratoires, les mesures de concentrations proviennent donc de différentes procédures analytiques. Plus précisément, chaque laboratoire avait la possibilité de choisir la procédure de digestion et d'extraction qui semblait la plus adaptée selon l'élément observé mais également la méthode de détection. Le nombre de laboratoires et de protocoles expérimentaux employés dans cette étude étant très important, le descriptif de ces derniers ne sera pas développé dans cette partie. Il est cependant possible d'accéder à ces informations, via le rapport périodique des résultats d'analyse des IPE, disponible en libre accès sur le site du WEPAL [43].
- Finalement, les échantillons de validation pour les modèles d'étalonnage de plantes ont été analysés par ICP-AES à l'université catholique de Louvain (Belgique) après application du protocole suivant :
 - les échantillons de feuilles de plante ont été premièrement nettoyés à l'aide d'une solution d'éthanol à 70% afin d'éliminer toute trace de pollution.
 - ces feuilles ont ensuite été séchées à 70°C pendant 4 jours puis broyées.
 - une partie de ces feuilles broyées sont ensuite mis en solution par de l'acide nitrique à 15% (HNO₃) suite à une calcination à 450°C pendant 4 heures.

Toutes les informations concernant ces analyses ICP-AES peuvent être retrouvées dans le manuscrit de thèse de Charles Van Linden [39].

ii. Mesures par XRF portable

L'intérêt de réaliser une campagne de mesures XRF provient du fait que celle-ci s'est imposée au fil des ans comme étant une méthode très performante pour l'analyse des sols [44] et des plantes [10]. Même si elle n'est pas encore considérée aujourd'hui comme étant la méthode de référence des matrices agronomiques (la référence étant l'ICP-AES et MS selon les cas), la XRF permet d'analyser de façon directe des échantillons selon des étapes de préparation similaires à la LIBS et il est donc particulièrement intéressant de confronter les résultats obtenus par LIBS à ceux obtenus par XRF.

La XRF entre elle aussi dans la catégorie des spectroscopies d'émission atomique mais cette fois l'échantillon est bombardé par des photons qui se situent dans la gamme spectrale des rayons X comme le montre la figure 1-3. Les atomes qui absorbent l'énergie de ce rayonnement X éjectent des électrons de leur couche interne et se trouvent ainsi dans un état excité. Ils regagnent leur état fondamental grâce aux électrons des couches externes supérieures qui viennent combler le manque d'électrons au niveau des couches internes insaturées. Les électrons mis en jeu libèrent de l'énergie de manière radiative et on observe une émission de photons dans le domaine spectral des rayons X, qui est appelée fluorescence X. Chaque atome est caractérisé par une longueur d'onde d'émission caractéristique de ses niveaux d'énergie électroniques. La séparation des photons X est effectuée par diffraction sur un cristal et ces derniers sont détectés finalement par un compteur (SSD, Silicon Drift Detector ou diode Sipin), ce qui va permettre d'identifier l'atome en fonction de son énergie.

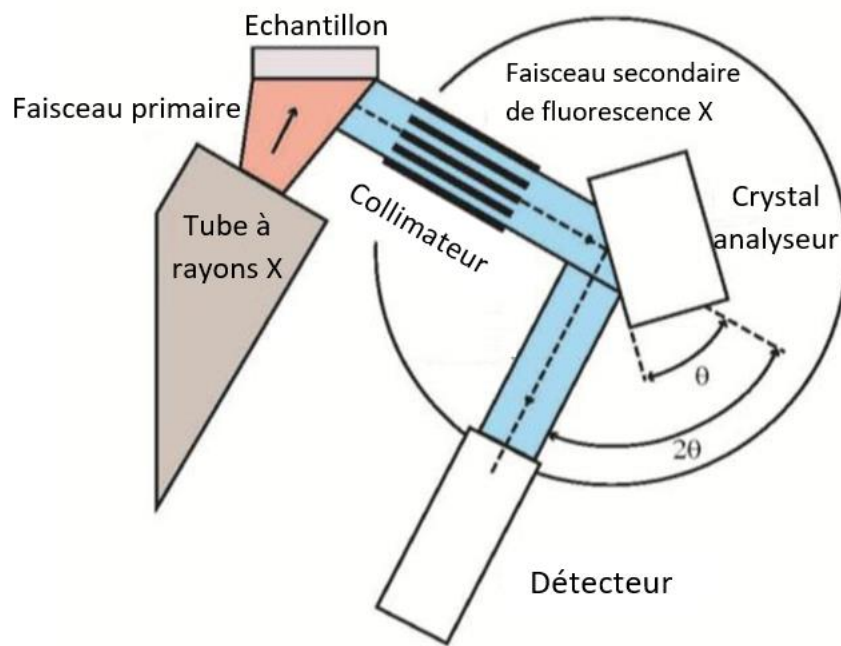


Figure 1-3 : Montage typique d'un dispositif XRF portable [45].



Figure 1-4 : Photo de l'instrument de XRF portable, SI 800 Titan, développé par Bruker [46] et utilisé dans le cadre de cette étude.

L'instrument XRF portable développé par Bruker et utilisé dans le cadre de cette étude présente les mêmes caractéristiques ergonomiques que l'appareil LIBS portable (cf partie 3ii de ce chapitre). En effet, le S1 800 Titan prend la forme d'un pistolet d'environ 1,5 kg et peut être fixé sur un châssis en conditions de laboratoire comme illustré par la figure 1-4. L'analyse XRF se fait également au contact de l'échantillon, tête vers le haut, à l'intérieur d'une enceinte fermée afin d'éviter au maximum la propagation des radiations. Les paramètres d'acquisition et l'analyse peuvent ensuite être commandés par un ordinateur portable à l'aide d'un logiciel fourni par le constructeur. Les principales caractéristiques de la XRF portable peuvent être consultées sur le site du constructeur (<https://www.bruker.com/products/x-ray-diffraction-and-elemental-analysis/handheld-xrf/s1-titan-series/technical-details.html>).

En ce qui concerne la préparation des échantillons de sols avant analyse XRF, elle est très simple. En effet, après séchage et tamisage (2 mm), les échantillons de sols sont broyés à une granulométrie inférieure à 250 µm. Ensuite, pour chaque échantillon, 3 g de cette fine poudre de sol sont utilisés pour fabriquer des pastilles de 32 mm de diamètre à l'aide d'un compresseur mécanique hydraulique (Specac, 25 tonnes). Une pression de 20 tonnes, durant 2 minutes, est appliquée sans liant afin d'obtenir des pastilles compactes et manipulables sans s'effriter. Un exemple de pastille est donné par la figure 1-1 en annexe. Une mesure XRF est finalement effectuée sur chaque face de la pastille, le diamètre de la surface analysée étant d'environ 13 mm. Il est cependant important de noter qu'en utilisant le pistolet XRF S1 800 Titan, l'analyste n'accède non pas à un spectre, mais directement aux valeurs de concentration des éléments totaux contenus dans l'échantillon. Ces valeurs de concentrations ont pu être prédites à l'aide d'une sélection d'échantillons adaptée et d'une calibration interne proposée par le constructeur. Une calibration interne était proposée pour les sols et une autre pour l'analyse des plantes. Aucun détail n'a été fourni par le constructeur sur la nature des échantillons et des algorithmes utilisés pour construire ces modèles internes de calibration.

En ce qui concerne la préparation et l'analyse des échantillons de plantes par XRF, elle a été réalisée par l'USRAVE et selon le protocole suivant :

- Les échantillons ont été séchés à 50°C.
- Ces échantillons séchés sont ensuite broyés soit au broyeur à couteaux (Retsch SM 2000) si ce sont des feuilles ou des tiges, soit au broyeur de type « moulin à café » (Ika M20) si ce sont des grains. La granulométrie finale obtenue est inférieure à 500 µm.
- La poudre obtenue est alors pastillée sous 22 tonnes à l'aide d'une pastilleuse semi-automatique Specac modèle Atlas T25.

3) Méthode LIBS

i. Principe

La LIBS est une technique de spectroscopie d'émission atomique dont la source d'excitation est un laser impulsif. Le laser est focalisé à la surface du matériau, où la matière est vaporisée si la densité surfacique d'énergie, ou fluence (exprimée en J/cm²), dépasse le seuil d'ablation du matériau. Le laser contribue ensuite à exciter et ioniser la vapeur qui se développe au-dessus de la surface de l'échantillon, initiant ainsi un micro-plasma au sein duquel toutes les espèces en présence sont excitées par collision avec les électrons libres du plasma ainsi que de manière radiative. Après l'impulsion laser, le plasma continue son expansion dans l'atmosphère

ambiante et se refroidit. Les atomes et ions présents en son sein se désexcitent alors en émettant un rayonnement caractéristique transporté par une fibre optique jusqu'à un spectromètre où il est décomposé en un spectre de raies dans la gamme spectrale allant de l'ultraviolet jusqu'au proche infrarouge. L'exploitation de ces raies, représentatives de la composition élémentaire du matériau analysé, permet de remonter à la concentration des éléments associés.

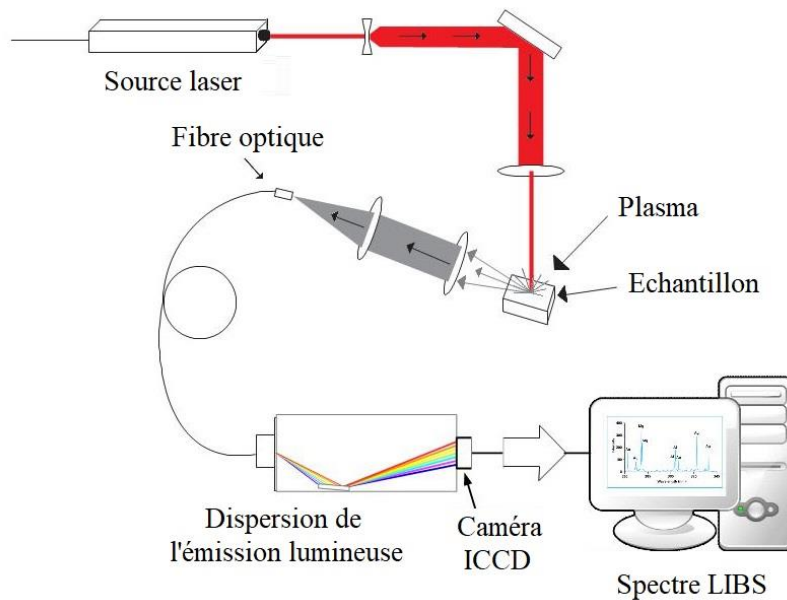


Figure 1-5 : Schéma de principe d'une analyse LIBS (source Wikimedia Commons – Fanny Bachelier, 2016)

La figure 1-5, schématisant le montage typique permettant de réaliser une analyse LIBS, permet de mettre en évidence les points-clés de la technique, à savoir le laser impulsionnel, à l'origine de l'ablation et de la création du plasma, et le spectromètre optique. Sur ce schéma, le faisceau laser est focalisé à la surface de l'échantillon à l'aide d'une lentille convergente. La lumière émise par le plasma est collectée à l'aide d'une lentille convergente puis transportée jusqu'à l'entrée du spectromètre par fibre optique. Le spectre LIBS contient essentiellement des raies atomiques correspondant aux atomes/ions présents dans l'échantillon. La LIBS est le plus souvent exploitée à partir d'impulsions laser de quelques nanosecondes et de quelques millijoules, et le spectre associé est généralement enregistré sur la gamme spectrale 200 et 900 nm. Cependant ces valeurs typiques varient sensiblement d'un instrument LIBS à un autre.

ii. Instrument LIBS portable



Figure 1-6 : Photo montrant l'instrument Z300 a) sur un châssis pour un usage en conditions de laboratoire, et b) en main pour un usage en dehors du laboratoire (notons que cette photo sert uniquement d'illustration et ne correspond pas à une mise en œuvre réaliste pour l'analyse in-situ d'un sol).

L'intégralité des données LIBS exploitées au cours de cette étude a été acquise à l'aide de l'instrument portable Z-300 conçu et fabriqué par SciAps (<https://www.sciaps.com/libshandheld-laser-analyzers/z-300/>) et commercialisé en France par Quantum RX dont les principales caractéristiques sont regroupées dans le tableau 1-2.

L'ensemble du dispositif LIBS, c'est-à-dire le laser, le système de focalisation et de collection ainsi que le spectromètre, est intégralement contenu dans le corps de l'appareil, présenté sous la forme d'un pistolet. L'ergonomie de ce dernier permet une utilisation contrôlée en conditions de laboratoire mais aussi une analyse directe sur le terrain comme présenté par la figure 1-6. Un support fixe en acier a été construit par le service mécanique du CELIA afin de pouvoir utiliser l'instrument sur une pailleasse, en position verticale, tête vers le bas, ce qui permet de limiter les projections de matière dans la tête de l'appareil et de maintenir le pistolet dans la même position et à la même hauteur à chaque analyse afin de travailler dans les conditions les plus répétables possibles. En laboratoire, l'instrument portable ne nécessite qu'un branchement sur prise électrique pour être fonctionnel et à un ordinateur pour définir les paramètres d'acquisition et contrôler la campagne de mesures. La capacité du Z-300 à fonctionner sous batterie lithium-ion, son faible poids (1.8 kg) et son ergonomie facilitent également son utilisation à la main directement sur le terrain.

Poids (kg)	Energie (mJ/impulsion)	Durée des impulsions (ns)	Diamètre spot (μm)	Fluence laser (J/cm^2)	Fréquence (Hz)
1.8	5-6	2	50	250	10

Longueur d'onde (nm)	Gamme spectrale (nm)	Résolution spectrale (nm)	Gaz d'utilisation
1064	190-950	0.1	Argon

Tableau 1-2: Caractéristiques du système Z-300 de SciAps.



Figure 1-7 : Interface utilisateur sur PC de l'instrument Z-300.

Le pistolet Z-300 fonctionne au contact de l'échantillon afin de garantir à l'utilisateur les conditions de sécurité laser, ce qui signifie que lorsque la fenêtre de tir est au contact de l'échantillon, le laser est considéré comme focalisé de manière optimale, en considérant évidemment que la surface de l'échantillon est plane. Une caméra interne au système permet de visualiser l'échantillon en temps réel (partie gauche de la figure 1-7), ainsi que les cratères laissés par les tirs laser, facilitant le positionnement de l'échantillon par rapport à la fenêtre de tir. Il est ensuite possible d'ajuster certains paramètres d'acquisition avant analyse et d'observer les spectres LIBS en direct, à l'aide d'un logiciel développé par le constructeur (Profile Builder). La plupart de ces paramètres comme le délai, c'est-à-dire le temps entre le déclenchement du tir laser et l'ouverture du détecteur, peuvent être ajustés mais ont cependant déjà été optimisés par le constructeur et les valeurs par défaut conviennent à notre domaine d'application. En ce qui concerne l'exploitation du signal LIBS, le logiciel Profile Builder permet uniquement d'observer les spectres bruts, de les comparer et de les exporter. L'intégralité des traitements de données a donc dû être réalisé à l'aide d'un outil de programmation supplémentaire et nous avons opté pour Rstudio, logiciel en libre accès qui a été largement utilisé au sein de la communauté de chimiométrie. Ce travail de traitement des données numériques sera présenté dans la partie *Prétraitement des spectres LIBS* de ce chapitre. L'une des spécificités les plus intéressantes du Z-300 est sa capacité à fonctionner avec un flux de gaz d'argon puisqu'il a été établi que l'expansion du plasma LIBS dans un gaz différent de l'air pouvait améliorer les performances analytiques de la technique [47].

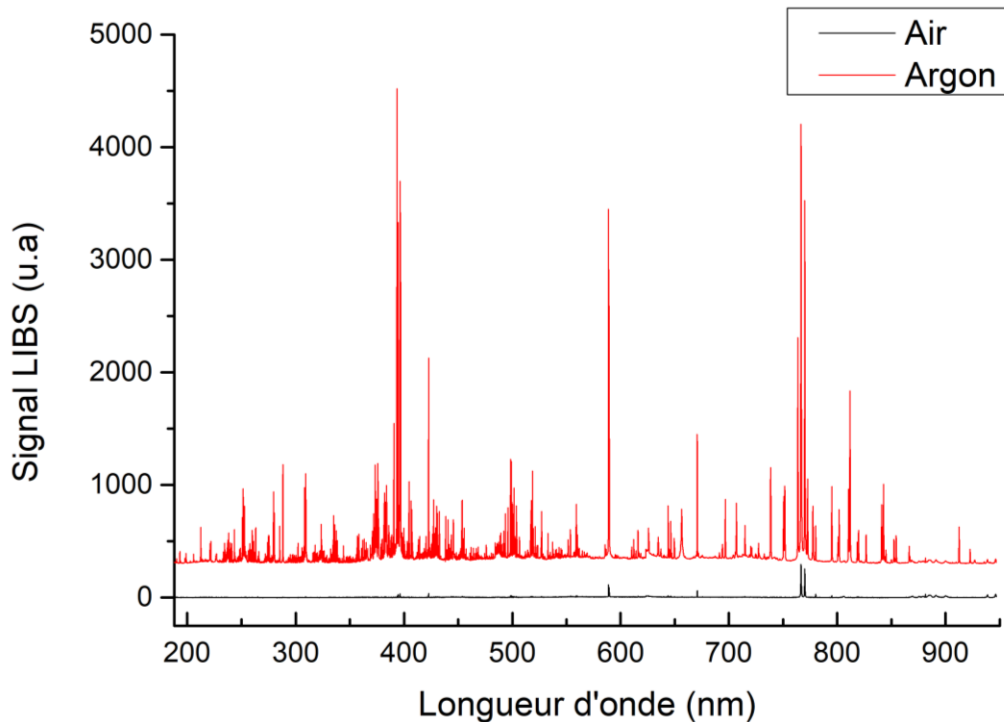


Figure 1-8 : Comparaison du spectre LIBS du même échantillon de sol obtenu avec le pistolet Z-300 sous air (trait noir) et sous Argon (trait rouge), à pression ambiante.

La figure 1-8 permet de comparer deux spectres LIBS d'un même échantillon de sol agricole représentatif de notre corpus, enregistrés par le même système Z-300 mais dans deux conditions d'analyse différentes : le premier (trait noir) est obtenu alors que le plasma se développe sous air à pression ambiante, et le deuxième (trait rouge) est obtenu alors qu'un flux d'argon d'environ 1 bar est appliqué sur la zone de l'échantillon analysée. On observe clairement sur la figure 1-8 que le spectre LIBS est beaucoup plus lumineux sous argon que sous air. Pour tirer le meilleur profit de ces conditions de mesure, alors que le pistolet Z300 était initialement conçu pour fonctionner exclusivement avec des petites cartouches d'argon non rechargeables, le fabricant SciAps, en réponse à notre demande, a ajouté à sa gamme de produits un adaptateur permettant de connecter le pistolet Z300 à une arrivée d'argon issue de bouteilles de plusieurs litres. Cette nouvelle option est très utile pour les analyses en laboratoire sachant que l'autonomie des cartouches d'argon non rechargeables est relativement faible, de l'ordre de 3 heures d'analyses.

Notons par ailleurs qu'un spectre LIBS acquis avec le Z-300 couvre la gamme spectrale allant de 180 à 950 nm. Une gamme aussi large est obtenue grâce à l'association de trois spectromètres compacts : le premier est dédié à la gamme spectrale de l'ultraviolet (de 180 à 350 nm), le deuxième au visible (de 350 à 600 nm) et le dernier au proche infrarouge (de 600 à 950 nm). Les organes de l'appareil (laser, dispositif optique et spectromètres) sont séparés de l'échantillon par une fenêtre en quartz qu'il convient de nettoyer régulièrement afin de rester dans des conditions de collection de signal optimales. En effet, la LIBS induit une ablation de la matière si bien que l'on constate régulièrement la présence d'un dépôt de particules sur la fenêtre de protection, ou encore la présence de buée ou de gouttelettes, rendant celle-ci plus ou moins opaque. L'instrument permet également de réaliser un étalonnage régulier et automatique de la longueur d'onde. Pour cela, le faisceau laser est envoyé non plus sur l'échantillon mais sur le métal de la tête de mesure qui encadre la zone d'analyse. Les raies d'émission de ce métal

étant connues, elles sont utilisées comme référence pour étalonner l'instrument si besoin. Cependant, l'interface utilisateur de l'instrument ne permet pas de vérifier que celui-ci fonctionne de manière robuste et c'est pourquoi nous avons choisi de développer notre propre carte de contrôle.

iii. Contrôle de l'instrument portable Z-300

La LIBS est de plus en plus exploitée dans le but de quantifier la composition de la matrice étudiée [17], si bien que le niveau d'exigence en ce qui concerne la stabilité de l'instrument utilisé et des conditions de mesure ne cesse d'augmenter. Les laboratoires d'analyses chimiques, en général, mettent en place une démarche qualité élevée, guidée par l'application de normes (généralement établies par l'organisation internationale de normalisation pour les laboratoires d'analyse - ISO) et validée par des organismes d'accréditation comme le Comité français d'accréditation (Cofrac). L'application de ces normes et la validation du respect de leur application permet d'assurer qu'un laboratoire est en mesure de proposer des résultats valides et fiables. Les résultats obtenus en LIBS sont issus de laboratoires de chimie pour certains mais aussi pour beaucoup d'entre eux de laboratoires de physique, de géologie, d'archéologie, de sciences criminelles et bien d'autres, qui, si on en croit les publications associées, n'ont pas tous le même niveau d'exigence en ce qui concerne les normes et la qualité des analyses.

Notons d'ailleurs qu'après plusieurs tentatives infructueuses de comparaisons interlaboratoires, Motto-Ros et al. ont proposé de séparer en deux blocs les sources de variabilité, celles liées à la mesure brute et celles liées au traitement des données [27]. Ils ont démontré que le traitement des données avait un impact très fort sur les résultats analytiques. Mais les mesures elles-mêmes peuvent présenter des fluctuations élevées ou même dans certains cas, un problème de reproductibilité [48].

La carte de contrôle, pratique usuelle en laboratoire et construite pour tout type d'instruments, de la balance de mesure aux appareils d'analyses chimiques (ICP, XRF, CPG...), permet de s'inscrire dans cette démarche qualité, qui fait encore défaut à la communauté LIBS de manière générale. La carte de contrôle est un outil graphique de suivi du bon fonctionnement d'un instrument ou d'un procédé, permettant de mettre en lumière une potentielle dérive instrumentale ou au contraire de valider la réalisation d'une nouvelle campagne de mesures. La construction de cet outil passe généralement par l'analyse régulière, et selon une certaine fréquence, d'un échantillon de référence, afin d'observer la variabilité des résultats obtenus c'est-à-dire leur moyenne, écart-type ou encore étendue. La carte finalement construite permet de distinguer la variabilité aléatoire jugée normale et liée à l'utilisation de l'instrument dans des conditions classiques de celle attribuable à une possible dérive de l'instrument. La carte de contrôle représente ainsi le carnet de santé de l'instrument. L'échantillon que nous avons sélectionné pour construire la carte de contrôle du Z-300 est un échantillon d'acier certifié de référence commercialisé par le BAM (Berlin). L'échantillon se présente sous la forme d'un cylindre de 50 mm de diamètre et 42 mm d'épaisseur. Le BAM est un organisme accrédité pour la production de matériaux de référence selon la norme ISO/IEC 17025 et les aciers qu'il produit sont réputés comme étant particulièrement homogènes, point indispensable lorsque l'on souhaite observer les variabilités instrumentales comme sur une carte de contrôle.

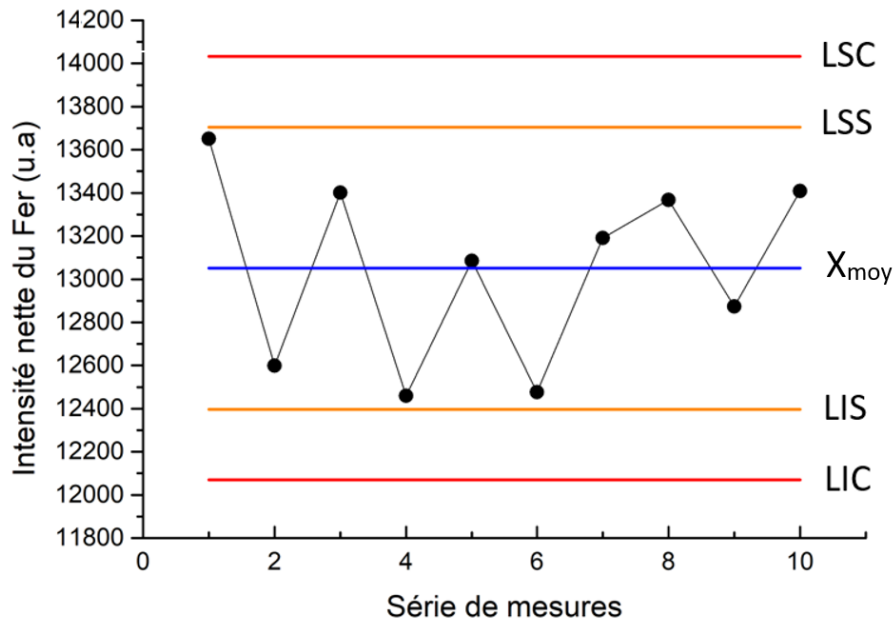


Figure 1-9 : Carte de contrôle du pistolet LIBS Z-300, construite sur l'intensité nette de la raie du Fer à 302,08 nm. Le calcul des traits horizontaux et leur signification (LSC, LSS, Moy, LIS, LIC) est explicité dans le texte.

Le but de la carte de contrôle pour cette étude est de vérifier le bon fonctionnement du système Z-300 à chaque nouvelle utilisation. Le spectre LIBS étant constitué de centaines de raies d'émission atomiques, la première étape consiste à sélectionner le signal qui sera utilisé pour la carte de contrôle. L'aire totale du spectre pourrait sembler être un bon choix puisqu'elle permet de prendre en compte les variations sur la totalité de la gamme spectrale. En effet, un plasma plus chaud sera associé à un spectre plus lumineux avec une augmentation du signal principalement dans l'UV. Cependant, afin d'être plus sensible aux variations relatives d'intensité entre les raies des atomes et des ions, il nous a semblé plus pertinent de prendre en compte l'intensité nette d'une raie d'émission du fer. Nous avons choisi la raie de Fe I à 302.08 nm car elle ne présente pas de saturation ni d'interférence avec une autre raie. L'intensité nette est obtenue après une simple soustraction de la ligne de base, comme décrit dans la partie *Prétraitement des spectres LIBS* de ce chapitre.

La carte est construite selon les étapes suivantes :

- A chaque journée de mesures, durant plusieurs mois, l'échantillon d'acier de référence a été analysé (2 fois le matin et 2 fois l'après-midi). Une moyenne a ensuite été calculée à partir de ces 4 points d'analyse sur l'intensité nette de la raie de Fe I à 302.08 nm.
- Il est ensuite possible, à l'aide de ces moyennes et plus précisément de la variabilité de ces moyennes, de calculer les limites de la carte de contrôle, limites qui vont permettre ensuite d'estimer si l'instrument est sous contrôle.
- Les fluctuations d'intensité peuvent ensuite être interprétées par rapport à ces limites.

Le calcul des limites de la carte de contrôle s'est basé sur les travaux de Shewhart et du Western Electric, *Statistical Quality Control Handbook* [49] :

— Représente la moyenne des intensités moyennes notée Moy

— Représente les limites inférieures (LIC) et supérieures (LSC) de contrôle calculées à l'aide de l'équation suivante :

Equation 1-1
$$LSC = X_{moy} + A_2 R_{moy} \quad \text{et} \quad LIC = X_{moy} - A_2 R_{moy}$$

avec A_2 un facteur déterminé selon la « taille de l'échantillon », c'est-à-dire le nombre de points analysés par observation. Ici, il est égal à 4 et la valeur du facteur A_2 associé est de 2,28, le tableau de correspondance des facteurs selon la taille de l'échantillon est donné par le tableau 1-1 en annexe. R_{moy} est la valeur moyenne des étendues calculées pour chaque observation, l'étendue étant la différence entre la plus grande et la plus petite valeur d'une observation.

— Représente les limites inférieures (LIS) et supérieures (LSS) de surveillance calculées par l'équation suivante :

Equation 1-2
$$LSS = X_{moy} + \frac{2}{3} (A_2 R_{moy}) \quad \text{et} \quad LIS = X_{moy} - \frac{2}{3} (A_2 R_{moy})$$

avec A_2 le facteur déterminé selon la taille de l'échantillon. Ici, il est égal à 4 et la valeur du facteur A_2 associé est de 2,28. R_{moy} est la valeur moyenne des étendues calculées pour chaque observation.

Il en résulte alors la carte de contrôle du pistolet portable Z-300 illustrée en figure 1-9. Plus précisément, cette figure représente une portion de la carte de contrôle correspondant à 10 jours de mesures. On remarque bien l'absence de point sortant des limites de contrôle et de surveillance, ainsi qu'une dispersion aléatoire des points autour de l'intensité nette moyenne de la raie du fer à 302,08 nm. Une tendance croissante ou décroissante des points aurait indiqué une dérive de l'instrument et une tendance systématiquement supérieure ou inférieure un problème de justesse instrumentale. L'instrument portable Z300 peut donc être considéré comme étant sous contrôle. Une carte de contrôle sur l'étendue a également été construite afin d'observer la variabilité intra-observation. Cette dernière présente également un profil classique et est présentée sur la figure 1-2 en annexe. Par conséquent, l'acier de référence est analysé avant chaque campagne de mesures proposée dans cette étude et la carte de contrôle complétée afin de garantir la validité des mesures.

iv. Analyse LIBS sur pastille : un protocole établi par plan d'expérience

Les deux instruments portables (LIBS et XRF) présentent un même mode de fonctionnement avec une analyse au contact de l'échantillon. De plus la XRF est une méthode non destructive. Enfin, il a déjà été démontré que l'ablation laser était plus efficace sur un échantillon compact que sur le même échantillon à l'état de poudre [50]. Les pastilles de sols et de plantes préparées pour la campagne de mesure XRF ont donc été ré-utilisées pour la campagne de mesure LIBS.

Le Z-300 permet de réaliser des micro-analyses avec un diamètre de spot laser de l'ordre de 50 microns (le diamètre du faisceau à rayons X du système portable XRF se rapprochant, lui, de 13 mm environ). Il est donc important, avec une pastille de sol de 32 mm de diamètre, de s'assurer que le spectre moyen soit bien représentatif de la composition du sol étudié. La

préparation de l'échantillon, via un broyage fin, permet déjà de s'affranchir de fortes fluctuations du signal LIBS liées à des problèmes d'hétérogénéité de l'échantillon mais il reste nécessaire d'évaluer la quantité de points d'analyse à effectuer sur la surface de la pastille pour être représentatif du sol. La manière d'établir le protocole expérimental en LIBS est rarement discutée en profondeur en littérature [51], se résumant généralement à décrire le nombre de spots laser réalisés à la surface de l'échantillon sans expliquer pourquoi ce nombre a été retenu. Dans le but de proposer une démarche objective afin de définir ce nombre, un plan d'expériences a été construit. Le principe général du plan d'expérience est particulièrement détaillé dans la partie *IV.2.i Principe du plan d'expérience*. Ici, ne sont présentés succinctement que les résultats expérimentaux qui ont permis d'évaluer quel nombre d'analyses LIBS est suffisant afin d'obtenir un spectre moyen représentatif d'une pastille de sol.

Le modèle de plan d'expérience sélectionné est celui de Box-Behnken pour 3 facteurs car ce dernier propose un faible nombre de mesures (12) par rapport à un plan factoriel complet par exemple qui, avec les mêmes conditions de facteur, en nécessiterait bien plus, à savoir 27. Les 3 facteurs sélectionnés sont les 3 paramètres modifiables lors d'une analyse avec le Z-300 : (1) le nombre de tirs réalisés par spot d'analyse ; (2) le nombre de spots réalisés par zone d'analyse et (3) le nombre de zones d'analyse symbolisés par le facteur 1, 2 et 3 respectivement. En effet, lorsque la fenêtre de tir (cf figure 1-3) du LIBS portable est au contact de l'échantillon, il est possible de réaliser soit un seul spot d'analyse, où peuvent être focalisés plusieurs tirs laser, soit de programmer une ligne ou un quadrillage de plusieurs spots séparés entre eux de quelques microns, permettant de couvrir une plus grande surface malgré le faible diamètre du spot. Lorsque l'on étudie l'influence d'un facteur, en général, on limite ses variations entre deux bornes appelées « niveaux ». Un niveau bas, codé -1 pour la borne inférieure et un niveau haut, codé +1 pour la borne supérieure. Au plan de Box-Behnken, un niveau intermédiaire, codé 0, a été ajouté à chaque facteur. L'ensemble de ces bornes définissent le domaine d'étude du plan d'expérience. En d'autres termes, c'est la zone de l'espace expérimental choisie par l'expérimentateur pour faire ses essais. Ces bornes codées sont donc à transformer en valeurs réelles par l'analyste selon ce qu'il souhaite observer et sa connaissance du procédé analytique. Un plan de Box-Behnken à 3 facteurs et 3 niveaux par facteur représente, en terme analytique, 12 mesures différentes auxquelles s'ajoute un nombre, fixé par l'analyste, de mesures au centre du domaine d'étude. Cette répétition au centre du domaine permet d'estimer une erreur expérimentale qui sera considérée similaire à tous les points de mesure [52]. La géométrie du plan Box-Behnken pour trois facteurs à trois niveaux est présentée en figure 1-3 en annexe. En ce qui concerne le choix du domaine d'étude, il a été décidé de se placer dans un contexte d'analyse rapide, propre à la LIBS portable, c'est-à-dire favorisant l'obtention d'un spectre moyen d'un sol en quelques secondes jusqu'à quelques minutes. Dans un souci de clarté, les bornes ont été fixées aux mêmes valeurs pour les 3 facteurs à savoir : 5 pour le niveau bas (-1), 10 pour le niveau intermédiaire (0) et 15 pour le niveau haut (+1). Un récapitulatif du plan final sous sa forme codée et réelle est proposé en annexe dans le tableau 1-2. Un spectre moyen est conservé par combinaison de facteurs, c'est-à-dire le spectre moyen calculé sur le nombre de zones/spots et tirs réalisés sur la surface de la pastille selon la combinaison de facteurs associée. Un sol agricole typique du jeu de données a été sélectionné pour réaliser les mesures du plan d'expériences. Il s'agit de l'échantillon n° 152 du réseau QUASAPROVE décrit précédemment et présentant une texture typique d'un sol agricole français, à savoir une composition majoritairement limono-argileuse. Plus précisément, plusieurs pastilles de ce sol ont été fabriquées selon la procédure détaillée précédemment afin de pouvoir réaliser toutes les zones d'analyse, la variabilité inter-pastille d'un même sol agricole broyé finement étant jugée négligeable. Les spectres moyens de plusieurs pastilles du même sol présentent en effet des intensités sensiblement similaires, avec des coefficients de variation inférieurs à 5%.

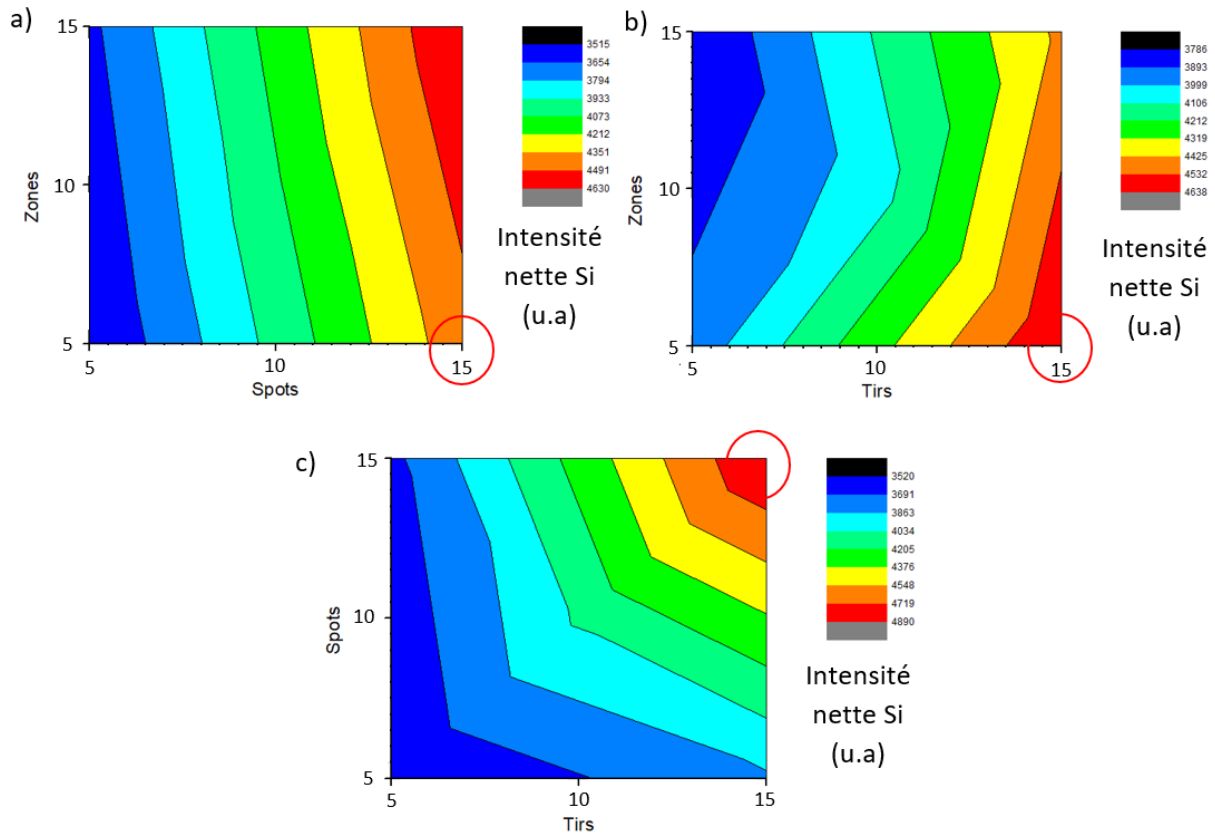


Figure 1-10 : Surfaces de réponse obtenues à l'aide du plan Box-Behnken pour l'intensité nette de la raie du silicium à 288 nm pour l'échantillon de sol 152 avec a) le nombre de zones en fonction du nombre de spots, b) le nombre de zones en fonction du nombre de tirs et c) le nombre de spots en fonction du nombre de tirs. Les cercles rouges représentent les conditions d'analyse conservées pour l'analyse des sols et des plantes par le Z-300.

Le résultat d'un plan d'expériences Box-Behnken peut se présenter sous la forme d'une surface de réponse en 2D comme montré sur la figure 1-10. On observe ici l'évolution de la réponse selon un espace à deux axes représentant deux facteurs, le troisième étant fixé sur sa valeur intermédiaire (ici 10). A l'ensemble de tous les points du domaine d'étude correspond un ensemble de réponses qui se localisent sur une surface appelée surface de réponse. La réponse sélectionnée pour notre étude est l'intensité nette de la raie du silicium à 288 nm qui est bien définie et ne présente pas de possible interférence avec un autre élément. Cependant, les mêmes profils de réponse ont été obtenus sur l'ensemble des raies du spectre LIBS. L'intensité de la raie de Si sélectionnée est donc considérée ici comme représentative du signal LIBS dans sa globalité. Les figures 1-10 a. et 1-10 b. montrent que l'intensité de la raie de Si, et donc de l'ensemble du spectre, par son profil en bandes verticales, est très similaire quand on réalise 5 ou 15 zones d'analyse. L'intensité se maintient entre 3500 et 4500 lorsque l'on réalise de 5 à 15 spots d'analyse par zone respectivement. Il ne semble donc pas utile, grâce au broyage fin permettant d'obtenir une pastille de sol homogène, de réaliser plus de 5 zones d'analyse afin d'obtenir un spectre moyen représentatif du sol étudié. En ce qui concerne le nombre de spots réalisés par zone d'analyse, il semblerait d'après les figures 1-10 a. et 1-10 c., que son augmentation permet d'obtenir des intensités plus fortes. Ce résultat ne paraît pas intuitif en LIBS, où pour un échantillon supposé homogène et plan, l'intensité du signal doit être identique sur toute la surface de l'échantillon. L'augmentation du signal ici a pourtant une explication qui est liée à l'instrument portable Z-300. En effet, le système de collection du signal du Z-300 permet de collecter plus de lumière lorsque le spot laser est proche du centre de la fenêtre de

tir. En pratique, les 5 spots représentent en fait une ligne de 5 spots laser située en bas de la fenêtre de tir. 10 spots représentent cette même ligne plus une nouvelle composée également de 5 spots située au-dessus de la première et plus proche du centre de la fenêtre de tir, idem pour 15 spots. Lorsque l'on réalise 15 spots de mesure, il y a donc plus de points d'analyse proches du centre de la fenêtre de tir qu'avec uniquement une ligne de 5 spots, permettant de collecter plus de lumière et donc d'obtenir un signal moyen plus intense.

En conclusion, les conditions optimales d'analyse d'un sol agricole avec le Z-300, c'est-à-dire permettant d'avoir un signal moyen intense et représentatif de l'échantillon étudié, correspondent à 5 zones d'analyse comprenant chacune 15 spots laser, de préférence centrés, et 15 tirs par spot.

Les campagnes de mesure pour l'étalonnage des sols agricoles et des plantes se sont donc appuyées sur les résultats de ce plan d'expériences. Une légère modification a cependant été effectuée concernant les sols par la réalisation de 5 zones d'analyse comprenant 12 spots laser et 15 tirs par spot. Le fait de ne réaliser que 12 spots, soit 3 lignes de 4 spots situées les unes au-dessus des autres, permet de plus espacer les spots entre eux et donc d'éviter un possible chevauchement entre cratères. En ce qui concerne l'analyse des pastilles de plantes, le nombre de zones analysées ne semblant pas être très influent sur l'obtention d'un spectre moyen représentatif et afin de réduire le temps d'analyse au maximum, le protocole LIBS est caractérisé par 3 zones d'analyse comprenant 12 spots laser et 15 tirs par spot par pastille de plante. En effet, les pastilles de plante ne présentant pas d'hétérogénéité à l'œil nu, à l'instar des sols, il n'a pas été jugé nécessaire de reproduire ce plan d'expériences pour les pastilles de plante. Les campagnes de mesure concernant les analyses en conditions *in-situ* seront détaillées et discutées dans le chapitre IV associé.

v. Prétraitement des spectres LIBS

Une fois le protocole expérimental déterminé et les mesures réalisées, il est nécessaire de s'assurer que les spectres sont prêts à être exploités. Les résultats obtenus en LIBS étant fortement dépendants non seulement des données brutes mais aussi des étapes de traitement [27], il est indispensable d'appliquer aux spectres LIBS certains pré-traitements adaptés avant de procéder à des opérations de traitement des données plus poussées.

A l'issue d'une mesure avec l'instrument portable Z300, il est possible d'extraire deux types de spectres LIBS, à savoir le spectre brut n'ayant subi aucun traitement ou un spectre corrigé par interpolation en interne par le logiciel d'acquisition des données fourni par le constructeur, SciAps. Les spectres corrigés présentant certaines anomalies, nous avons décidé de travailler sur les spectres bruts. Cependant, l'extraction des spectres bruts du Z-300 nécessite de tenir compte de deux problèmes. Le premier concerne un décalage en longueur d'onde (figure 1-11 a) et le deuxième un recouvrement spectral d'abord entre 300 et 400 nm et ensuite entre 550 et 650 nm (figure 1-11 b).

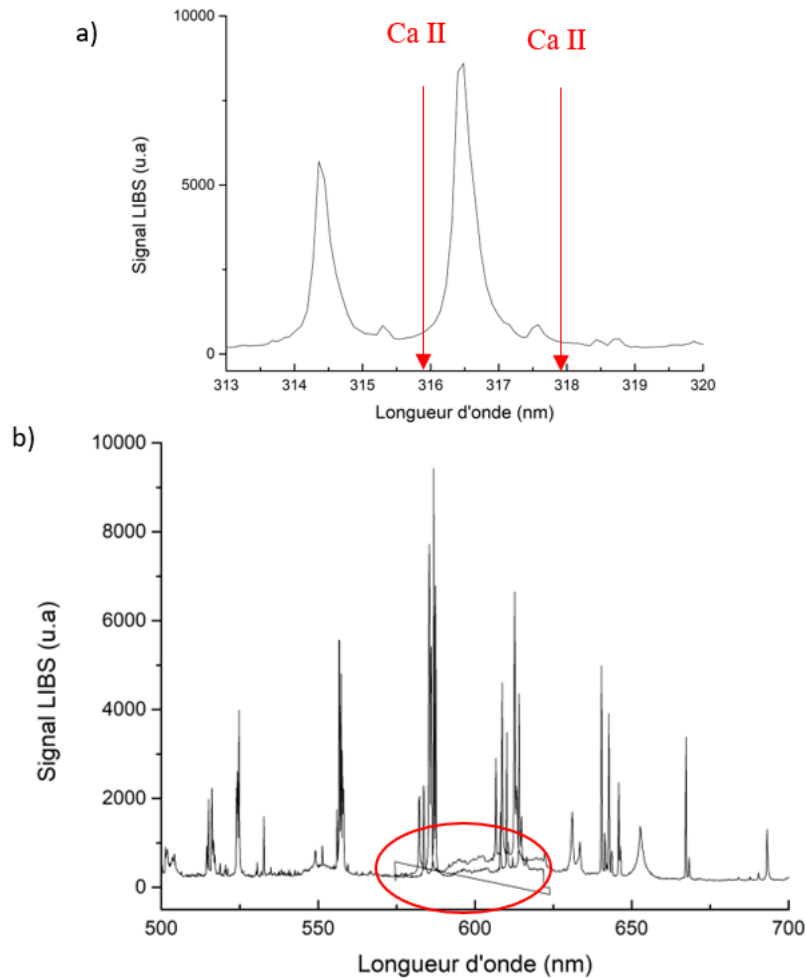


Figure 1-11 : Problèmes constatés lors de l'extraction des spectres LIBS bruts du pistolet Z-300 avec a) un exemple de décalage en longueur d'onde pour le doublet du calcium à 315,92 et 317,96 nm et b) un exemple de recouvrement spectral.

Le décalage en longueur d'onde correspond à un décalage constant sur l'ensemble de la plage spectrale de 1,5 nm par rapport aux valeurs théoriques de longueurs d'onde d'émission des raies spectrales. Un exemple est donné avec la figure 1-11 a. montrant le décalage du doublet de Ca II qui est attendu à 315,92 et 317,96 nm [53]. Ce décalage de 1,5 nm (soit de 18 points d'échantillonnage du signal) est constant et a été corrigé à l'aide d'un script R [54]. Le recouvrement spectral visible d'abord entre 300 et 400 nm et ensuite entre 550 et 650 nm correspond certainement au changement de spectromètre. Chacun de ces recouvrements reste inchangé sur tous les spectres. Notre script R comprend donc également des lignes de code permettant d'éviter ces recouvrements spectraux et d'obtenir au final un spectre « continu » de 188 à 950 nm. L'entièreté de ce script est disponible dans l'annexe I dédiée à ce chapitre. En résumé, après avoir rassemblé tous les spectres LIBS à corriger dans une même matrice, nous décalons nos valeurs d'intensité de 1,5 nm et nous supprimons les gammes de longueur d'onde affichant des valeurs d'intensité nulles à cause des superpositions spectrales. Le contenu du script est entièrement détaillé en annexe.

En ce qui concerne la soustraction de la ligne de base, de nombreuses stratégies ont déjà été appliquées [55]. Le délai de 250 ns fixé par le constructeur réduisant drastiquement la présence du fond (Bremmstrahlung) d'un spectre à l'autre sur le Z-300, il a été décidé de n'appliquer qu'une légère soustraction à chaque spectre à l'aide de la fonction « baseline.corr » du package

ptw. L'algorithme utilisé dans cette fonction calcule une ligne de base à l'aide d'une méthode asymétrique des moindres carrés et la soustrait ensuite aux données. Toutes les informations relatives à ce calcul sont fournies dans le descriptif du pack ptw [56].

Par ailleurs, les fluctuations expérimentales génèrent des fluctuations de signal qu'il est possible de réduire en partie en réalisant des moyennes statistiques sur un grand nombre de tirs laser. Au final, après correction des problèmes liés à l'instrumentation et soustraction de la ligne de base, les spectres sont moyennés de manière à réduire le bruit spectral et à obtenir un spectre moyen représentatif pour chaque échantillon. Les spectres individuels sont tout de même conservés pour calculer des coefficients de variation et des barres d'erreur pour les modèles d'étalonnage. Le fait de les conserver permet également d'observer les intensités zone par zone et de détecter la présence potentielle d'hétérogénéité en surface de la pastille analysée ou de fluctuations venant des conditions expérimentales. Un spectre LIBS, après avoir été prétraité par ces différentes étapes, est alors considéré comme prêt à être exploité dans un but qualitatif (identification spectrale, ACP) ou quantitatif (étalonnage univarié, multivarié).

En ce qui concerne l'extraction de l'intensité ou de l'aire nette des raies élémentaires des spectres LIBS pour la construction des modèles d'étalonnage univariés, elle a été réalisée de la manière la plus simple possible. En effet, l'intensité nette représente la différence entre l'intensité maximale du pic et la ligne de base estimée par l'analyste pour chaque raie élémentaire observée. Cette dernière étant soustraite efficacement pour la plupart des raies détectées, le calcul de l'intensité nette revient à extraire l'intensité maximale du pic. L'intensité nette de la raie de Mg I à 285,21 nm, à titre d'exemple, est représentée par la flèche verte sur la figure 1-12, la ligne de base étant représentée par une ligne bleue. L'aire nette est quant à elle obtenue en réalisant la somme des intensités nettes des longueurs d'onde acquises pour cette raie élémentaire. Cela représente la somme des flèches rouges et de la flèche verte sur la figure 1-12.

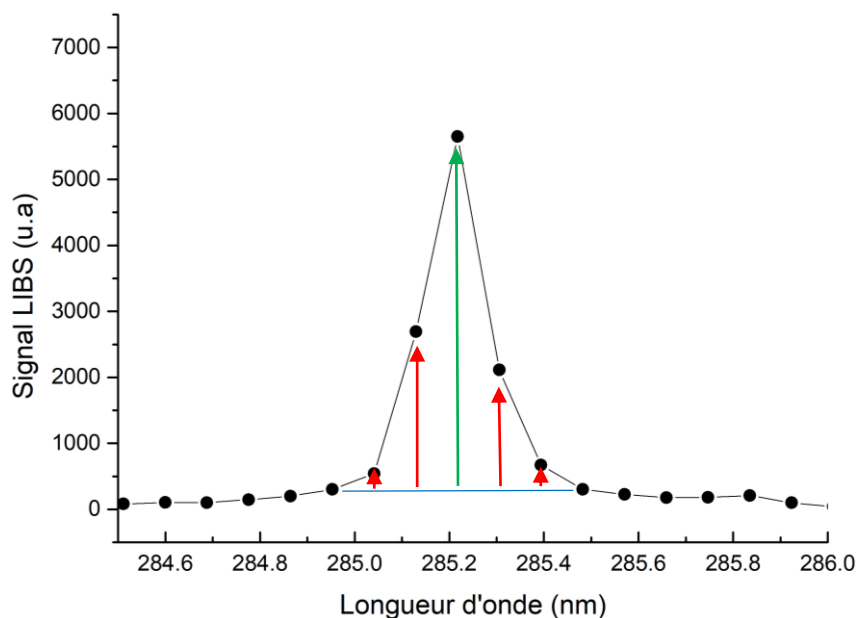


Figure 1-12 : Spectre LIBS d'un sol zoomé sur la raie de Mg I à 285, 21 nm.

Chapitre II

LIBS portable et plantes : vers des modèles universels ?

La stratégie d'étalonnage établie pour les plantes se base sur des échantillons d'espèces variées présentant des gammes de concentrations pertinentes. Il a donc été décidé d'analyser un jeu de données provenant du WEPAL, caractérisé par 17 échantillons certifiés de plantes (réf – IPE WEPAL) dont le numéro et la nature sont présentés dans le tableau 2-1 en annexe. Le jeu de données comprend aussi bien des feuilles ou du bois d'espèces ligneuses (chêne, hêtre) que d'espèces herbacées fourragères (luzerne) en passant par des espèces fruitières ou maraîchères (melon, cerise, banane). La sélection du jeu de données ne s'est pas faite uniquement à travers le critère de diversité des espèces présentes mais également par rapport aux gammes de concentrations observées. L'intérêt d'une telle démarche est de pouvoir quantifier la composition élémentaire des plantes pour des gammes de concentrations jugées réalistes, c'est-à-dire proches de celles observées en milieu naturel ou cultivé. Un récapitulatif des gammes de concentrations mesurées pour les échantillons de notre étude (concentration minimale, médiane et maximale du jeu de données) et « théoriques » est proposé par le tableau 2-1. La colonne « Gamme théorique » représente les gammes de concentrations moyennes jugées classiques en milieu naturel pour une plante présentant un bon statut nutritionnel. Les valeurs présentes dans cette colonne ont été déterminées selon plusieurs sources proposant le profil type du statut nutritionnel d'une plante [5, 57, 58]. Il est important de noter que le jeu d'échantillons sélectionné ici permet de bien encadrer les valeurs de concentrations observées dans la nature pour la grande majorité des éléments d'intérêt, à l'exception du phosphore pour lequel cette sélection d'échantillons ne permet pas d'atteindre les plus hautes valeurs de concentrations rencontrées en milieu naturel. Par ailleurs, la gamme de concentrations théoriques n'a également pas pu être déterminée pour l'aluminium et le cadmium. Ceci peut s'expliquer par le fait que ces deux éléments ne semblent pas jouer un rôle crucial dans le bon développement de la plante étant donné qu'ils sont considérés comme des éléments dits non essentiels. Cela signifie qu'ils ne participent à aucune réaction métabolique et leur absence ne conduit à aucune carence [58] mais que passée une certaine concentration ils peuvent devenir toxiques. C'est certainement pour cette raison qu'aucune donnée sur une gamme théorique jugée « normale » n'a pu être obtenue, la discussion concernant ces éléments étant principalement engagée lors de l'analyse d'échantillons de plante particulièrement pollués.

	Minimale	Médiane	Maximale	Gamme théorique
C (g/kg)	300	450	500	400 - 450
N (g/kg)	1,0	14	31	10 - 30
Al (g/kg)	0,02	0,2	2,0	
Ca (g/kg)	0,4	15	84	
Cd (g/kg)	0,01	0,1	1,0	
Cu (g/kg)	0,0009	0,005	0,01	
Fe (g/kg)	0,04	0,2	1,1	0,02 – 0,6
K (g/kg)	0,4	10	43	1,0 - 60
Mg (g/kg)	0,2	1,6	12	0,5 - 10
Mn (g/kg)	0,005	0,07	0,6	0,01 – 0,6
Na (g/kg)	0,05	0,1	3,3	0,05 – 0,8
P (g/kg)	0,05	2,0	6,0	0,5 – 10
Zn (g/kg)	0,009	0,03	0,1	0,01 – 0,25

Tableau 2-1 : Récapitulatif des concentrations mesurées par ICP-AES ou ICP-MS pour les échantillons certifiés du WEPAL et des concentrations « théoriques » observées en milieu naturel.

A la suite des mesures LIBS sur ces échantillons, une observation préalable des données est réalisée afin de bien identifier les zones spectrales présentant la plus grande variabilité entre les échantillons et de pouvoir ainsi mener une analyse quantitative de manière plus fine. Lors de la phase d'observation des données, il s'agit dans un premier temps d'identifier les éléments chimiques détectés par l'instrument LIBS puis d'évaluer les ressemblances et différences entre les spectres LIBS considérés à l'aide d'une Analyse en Composantes Principales (ACP). Les résultats d'ACP portant sur les spectres LIBS seront naturellement comparés à ceux relatifs aux données de référence obtenues par ICP-AES.

Une fois la phase d'observation des données achevée, on effectue une analyse quantitative basée sur une régression univariée entre le signal LIBS et les concentrations mesurées par la méthode de référence (ICP-AES) pour chaque élément d'intérêt détecté. L'importance de la normalisation dans la construction de ces modèles quantitatifs sera discutée. Les modèles d'étalonnage présentant les meilleures performances analytiques seront ensuite validés à l'aide d'un nouveau jeu de données, non utilisé pour leur construction et les performances analytiques des modèles conservés et leurs limitations seront discutées. La pertinence de confronter les données LIBS à des données XRF plutôt qu'à des données ICP-AES sera également abordée.

1) L'observation des données

i. Identification spectrale

Une fois l'étape de prétraitement (cf chapitre I) appliquée aux spectres LIBS bruts, il est possible d'attribuer à chaque raie spectrale un atome, qu'il soit neutre ou ionisé, en s'appuyant sur une base de données de raies d'émission atomiques, la plus connue et la plus fréquemment utilisée en LIBS étant celle de l'Institut National des Standards et de la Technologie (NIST), accessible gratuitement en ligne [53]. Il est possible d'y rechercher des éléments chimiques par

plage de longueurs d'onde et inversement, le tout en ayant des informations sur l'énergie d'ionisation de l'élément, l'intensité relative de la raie ou encore les niveaux de transition électronique. Récemment, la société AtomTrace a proposé un site ludique, présenté sous forme de tableau périodique, regroupant leur propre base de données ainsi que celle du NIST et de Kurucz afin d'attribuer rapidement une raie spectrale à un atome (<https://www.atomtrace.com/elements-database/>). L'attribution des raies de cette étude a donc été réalisée à l'aide de ces bases de données et facilitée par la connaissance du matériau analysé. En effet, plusieurs éléments peuvent émettre à des longueurs d'onde proches les uns des autres, ce qui induit un risque d'identification erronée. Ainsi, lorsque le contexte de l'analyse est connu et pris en compte, l'identification des raies d'émission s'en trouve grandement facilitée. D'autre part, il est rare qu'un élément ne présente qu'une seule raie d'émission détectable dans la gamme spectrale de l'instrument, et la détection simultanée de plusieurs raies d'un même élément permet d'établir des conclusions beaucoup plus robustes. Un spectre LIBS typique moyen d'un échantillon de plante du jeu de données (cf Tableau 2-1 en annexe) et plus précisément celui de la feuille de Hêtre (IPE 157), préparé sous forme de pastille sèche et compacte, est représenté sur la figure 2-1, ainsi que les éléments d'intérêt détectés et les raies d'émission mises en jeu pour l'identification.

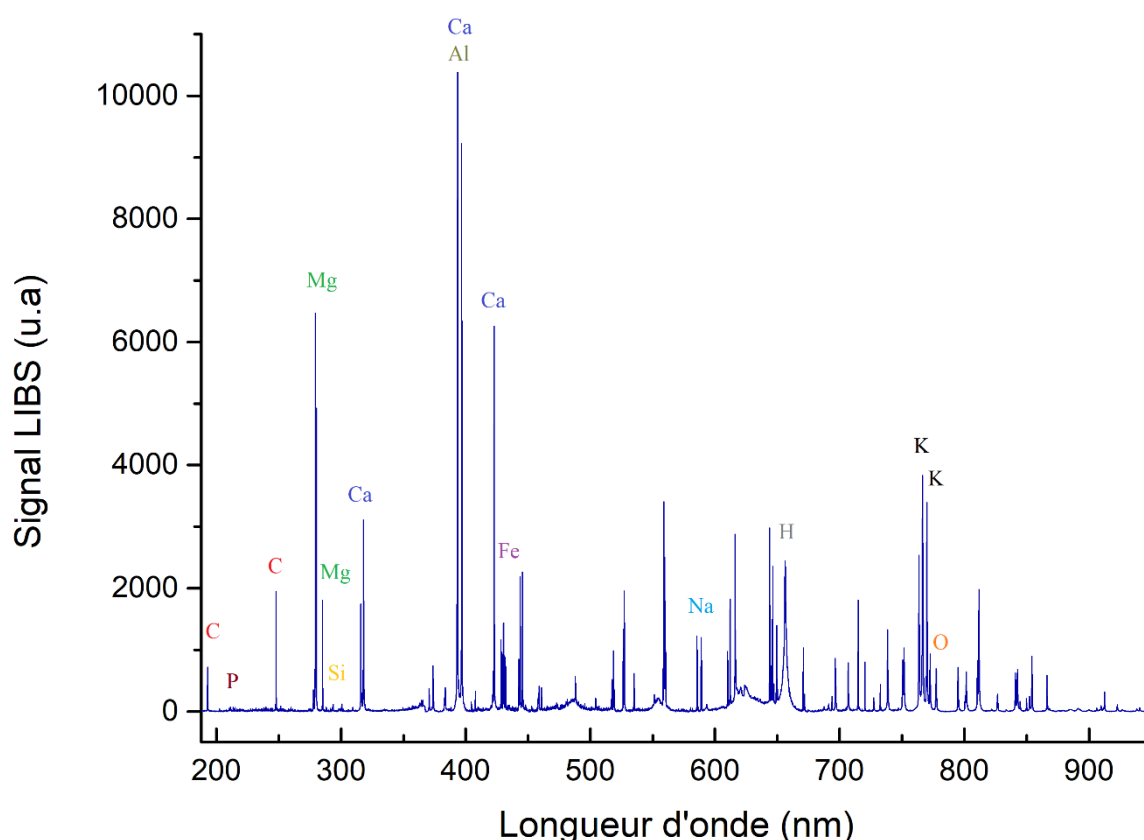


Figure 2-1 : Spectre LIBS typique moyen d'un échantillon de plante (pastille), ici une feuille de hêtre (IPE 157), avec l'attribution des éléments d'intérêt et les raies d'émission associées.

Les éléments d'intérêt détectés dans les échantillons de plantes de cette étude et reportés sur la figure 2-1 sont les suivants : C, Ca, Fe, K, Mg, Na, Si et P. Les raies d'émission associées sur la figure 2-1 sont les plus intenses pour chacun des éléments et celles restantes sont principalement associées au calcium et au magnésium ainsi qu'à l'argon qui est le gaz dans lequel le plasma se développe. Un inventaire complet des éléments observés, de leurs degrés d'ionisation et des longueurs d'onde associées est proposé en annexe par le tableau 2-2. Il est intéressant de noter que le pistolet Z-300 utilisé dans cette étude permet de détecter facilement les éléments décrits comme majeurs, c'est-à-dire les éléments concentrés de l'ordre du g/kg dans les plantes, à l'exception de l'azote.

En 2018, une étude menée par Jull et al. [59], et consacrée à la quantification des éléments nutritifs contenus dans des feuilles d'ivraie et de trèfle par LIBS, a permis d'expliquer pourquoi le cas de l'azote est très rarement discuté dans la cadre de l'analyse d'échantillons de plante par LIBS. La première raison est liée à la contribution de l'azote atmosphérique lorsque le système LIBS utilisé fonctionne sous atmosphère ambiante. Lorsque la LIBS est utilisée sous air, il n'est alors pas rare de voir du signal pour les raies élémentaires de l'azote mais ce signal est fortement représentatif de l'azote contenu dans l'air et non de celui contenu dans l'échantillon. La majorité des systèmes LIBS fonctionnant sous atmosphère ambiante, il est alors impossible d'exploiter les spectres dans le but d'observer le signal de l'azote. Cependant, il est possible, à l'aide de l'utilisation d'un autre gaz ambiant tel que l'argon, d'éliminer l'impact de l'azote atmosphérique et d'obtenir tout de même un signal, certes plus faible, mais associé cette fois-ci à l'azote contenu dans l'échantillon étudié. Malgré cet effort analytique, il reste cependant nécessaire d'utiliser un laser présentant une énergie assez élevée pour pouvoir observer un signal pour l'azote. Ici, les auteurs sont parvenus à obtenir un signal pour l'azote avec une énergie laser fixée à 100 mJ même si ce dernier reste faible d'après eux. C'est pour cela que, en ce qui nous concerne, malgré une expansion du plasma LIBS dans de l'argon, il n'est pas étonnant que l'on ne puisse pas détecter un signal lié à l'azote, le pistolet Z-300 disposant seulement d'une énergie laser de 6 mJ.

En ce qui concerne les signaux LIBS des éléments plus faiblement concentrés (mg/kg), dits « traces », comme Cd, Cu, Mn et Zn, ils ne sont pas détectés. Ces performances de détection ne concernent que les conditions de mesure associées au pistolet Z-300 et ne doivent en aucun cas être considérées comme celles de la technique LIBS en général. En effet, la LIBS a déjà pu montrer ses capacités à détecter ces éléments à des gammes de concentrations similaires à notre jeu de données dans les plantes mais à l'aide d'instruments de laboratoire opérant dans des conditions expérimentales très différentes (énergie laser, résolution spectrale, focalisation) [59, 60, 61].

L'identification des raies permet d'avoir une vision globale des éléments détectés mais pas de comprendre précisément ce qui différencie de manière spectrale une plante d'une autre. En effet, le nombre de longueurs d'onde (variables) est de 5843 pour le pistolet Z300 exploité pour cette étude, une observation individuelle de chacune des variables n'est pas réaliste. Il faut donc adopter une stratégie de compression des données et l'analyse en composantes principales (ACP) est l'outil que nous avons choisi d'adopter pour cela.

ii. Analyse en composantes principales, principe et calcul

a) Principe général et application à la spectroscopie LIBS

L'analyse en composantes principales (ACP) est certainement la méthode multivariée descriptive la plus répandue, permettant de représenter des observations issues d'un espace composé d'un grand nombre de dimensions dans un espace plus réduit. La représentation des données dans ce nouvel espace est un moyen de simplifier la réalité en décrivant un maximum de variabilité dans un espace réduit et donc en concentrant une information au départ diluée. La figure 2-2 schématise ce principe de réduction de dimensions où l'observation de départ est une photo d'une bolée de cidre breton. Cette dernière peut être décomposée et représentée selon trois axes qui permettent de comprendre rapidement les caractéristiques les plus notables de l'objet. Tout d'abord par une projection sur le plan (1,2), qui sont les deux axes expliquant un maximum la variabilité du jeu de données, ici on comprend directement avec cette projection que l'image représente un récipient avec une anse donc certainement une tasse ou un bol. La projection sur les axes suivants permet ensuite d'affiner la compréhension des données, ici d'attribuer une forme ronde à l'objet avec la présence d'une excroissance.

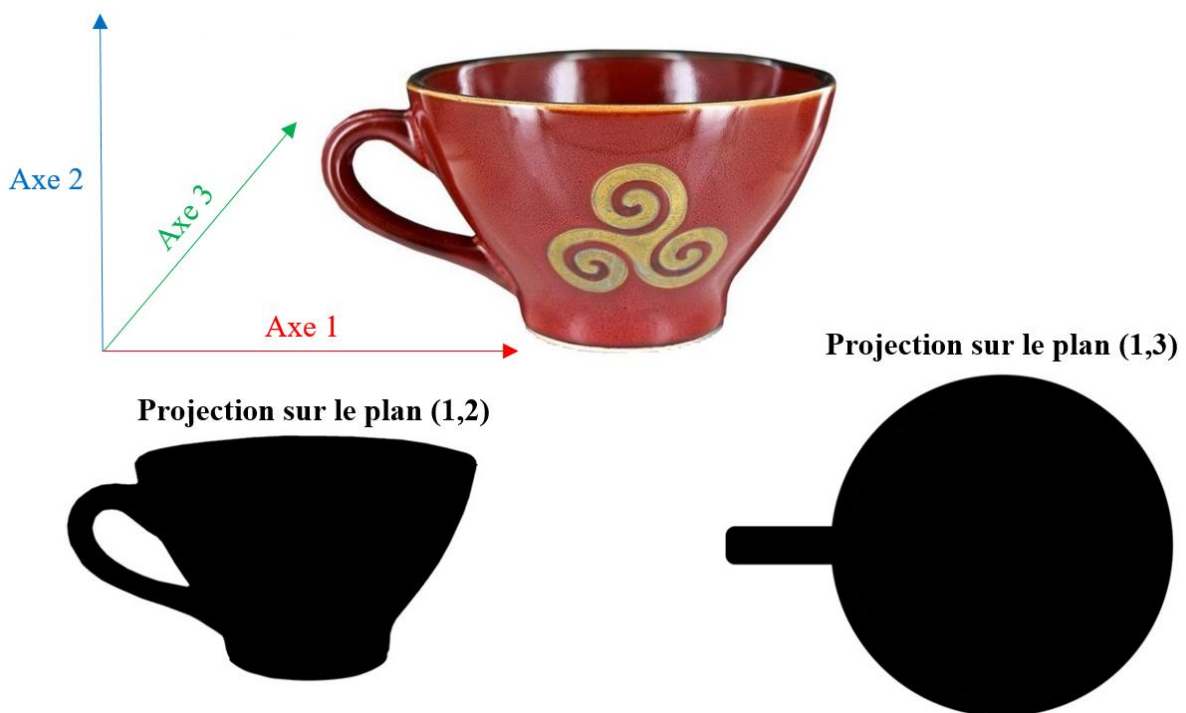


Figure 2-2 : Projection des données initiales (une bolée de cidre) sur le plan (1,2) et sur le plan (1,3) (illustration par Tony Richer).

L'ACP est très couramment appliquée à des données spectroscopiques [62, 63, 64] et donc aux spectres LIBS également [65]. Un jeu de données LIBS est présenté sous la forme d'une matrice contenant en ligne les observations et plus précisément dans notre cas, le spectre moyen de chaque échantillon. Les colonnes contiennent toutes les longueurs d'onde (nm) acquises par

l'instrument durant la mesure, dans notre cas de 188 à 950 nm. Un exemple est donné en figure 2-3 schématisant la réduction de dimensions par ACP appliquée à un jeu de données LIBS. On peut en effet remarquer la présence de plusieurs raies élémentaires associées au calcium entre 393 et 423 nm. L'intensité des raies détectées pour un même élément, ici le calcium, va donc varier de la même façon d'un échantillon à l'autre, on dit alors que ces longueurs d'onde, ou aussi appelées variables dans la matrice contenant les spectres moyens, sont corrélées entre elles. Ces variables corrélées vont ensuite être combinées afin de ne former qu'une seule variable latente, appelée composante principale. Il est cependant important de noter qu'une composante n'est pas forcément relative à des raies spectrales du même élément et peut permettre d'observer des corrélations ou anti-corrélations entre éléments, ou encore qu'une raie spectrale peut avoir du poids sur différentes composantes.

Pour résumer, les valeurs d'intensité pour chaque longueur d'onde représentent donc des variables quantitatives mesurées pour un certain nombre d'observations, défini par le nombre d'échantillons de plantes analysés, dans le cas où l'on ne considère qu'un spectre par échantillon. Le but de l'ACP sur les spectres LIBS va être d'identifier les combinaisons de variables/longueurs d'onde selon lesquelles les observations/spectres moyens se différencient le plus. Ces variables sont combinées selon leur covariance, définie par l'équation (2-1) suivante :

Equation 2-1
$$Cov(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

x et y représentant deux variables aléatoires et N le nombre d'échantillons. Plus précisément x_i représente la valeur d'intensité de la variable x pour l'échantillon i et $\bar{x}$ la moyenne des intensités de la variable x , même chose pour y_i et $\bar{y}$ mais pour la variable y . Lorsque que cette covariance est égale à 0, on considère alors que les variables ne sont pas corrélées. Au contraire, plus la covariance augmente et plus ces variables sont corrélées. Si elle est positive, cela signifie que lorsque la variable x augmente, la variable y augmente également. Si elle est négative, y diminue lorsque x augmente, les variables sont dites anti-corrélées. Les variables/longueurs d'onde corrélées entre elles qui permettent d'expliquer la variabilité du jeu de données sont alors combinées pour former les composantes principales (cf figure 2-3). La première composante symbolise donc un axe qui décrit la plus grande part de la variance du jeu de données, puis la deuxième composante permet d'expliquer la plus grande part de la variance résiduelle et ainsi de suite. La projection des observations, c'est-à-dire des spectres moyens, selon ces composantes permet ensuite d'observer si les échantillons forment des groupes distincts ou si l'un d'entre eux est anormalement éloigné des autres, aussi appelé *outlier*.

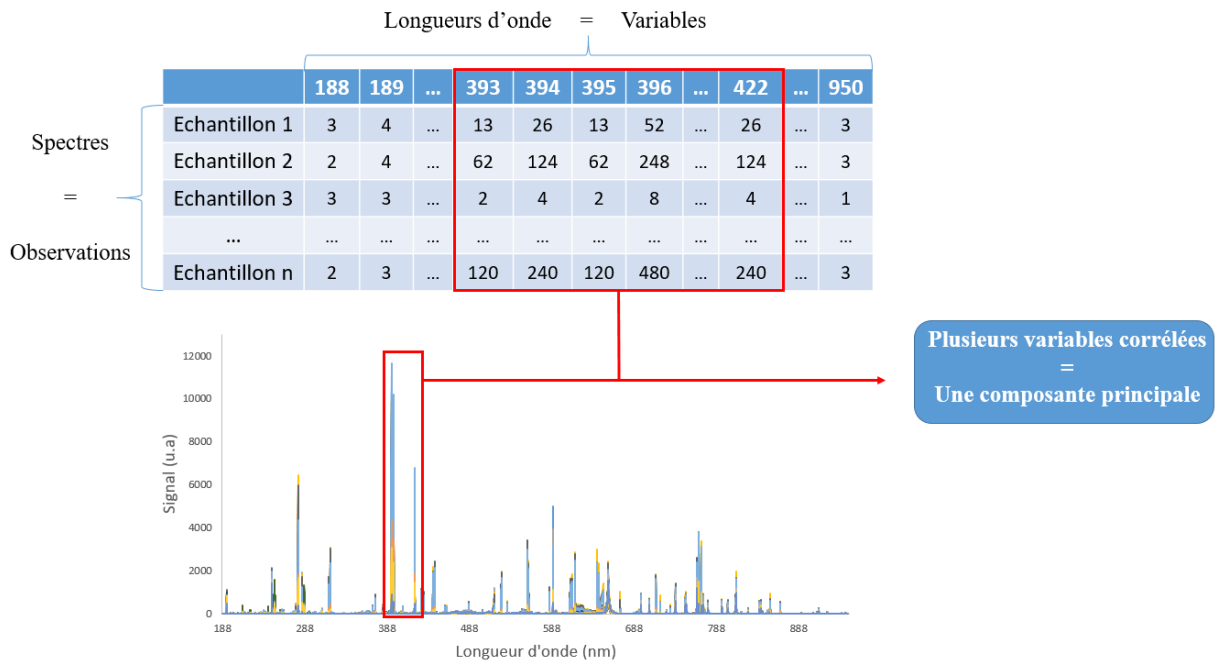


Figure 2-3 : Schéma de de la réduction de dimension appliquée à des données LIBS.

b) Le calcul mis en œuvre dans l'ACP

Le calcul d'ACP peut être décrit à l'aide de l'équation 2-2 et de la figure 2-4 :

Equation 2-2

$$X = TP^T + E$$

$$\begin{array}{c} X \\ \boxed{} \end{array} = \begin{array}{c} t_1 \\ \boxed{} \end{array} \begin{array}{c} p_1^T \\ \boxed{} \end{array} + \begin{array}{c} t_2 \\ \boxed{} \end{array} \begin{array}{c} p_2^T \\ \boxed{} \end{array} + \dots + \begin{array}{c} E \\ \boxed{} \end{array}$$

Figure 2-4 : Ecriture matricielle de l'ACP.

avec X la matrice des données, c'est-à-dire ici les spectres LIBS moyens en ligne, T la matrice des scores, P^T la matrice transposée des loadings et E la matrice des résidus. Les scores permettent de résumer comment sont reliés les observations, c'est-à-dire les spectres des échantillons, les uns aux autres. En effet, la colonne t_1 de la figure 2-4 représente les scores de la première composante, t_2 les scores de la deuxième jusqu'à la composante jugée optimale par l'analyste. Le fait de tracer les scores d'une composante par rapport à une autre, aussi appelé graphe des scores, permet ensuite de représenter la dispersion des observations/spectres des échantillons. Les loadings, représentés sur la figure 2-4 par p_1 pour la première composante, p_2 pour la deuxième et ainsi de suite, représentent le poids, ou autrement dit l'influence de chacune

des variables sur les scores. Le graphe des loadings permet de comprendre quelles variables/longueurs d'onde sont à l'origine de la dispersion des observations/échantillons sur le graphe des scores. Une fois ces variables déterminées, il est alors possible de comprendre quelles raies élémentaires permettent de distinguer les échantillons de plante entre eux. Finalement, la taille de la matrice E , qui représente la variance résiduelle qui n'est pas expliquée par l'ACP, dépend essentiellement du nombre de composantes sélectionnées pour expliquer le jeu de données.

L'ACP et ses composantes se déterminent à partir de la matrice des covariances qui est calculée de manière itérative pour les matrices de grandes dimensions et est calculée par le produit de la matrice X par sa matrice transposée X^T . La diagonale de cette nouvelle matrice représente les valeurs de variance de chaque variable. Il convient alors de calculer les valeurs et vecteurs propres de cette matrice de covariance, ce qui est possible, pour des matrices de grande dimension, à l'aide d'un calcul itératif. L'algorithme NIPALS, pour « nonlinear iterative partial least squares », est certainement l'algorithme le plus utilisé pour déterminer les composantes d'une ACP [42]. Ce calcul itératif a été appliqué à nos données à l'aide du pack « chemometrics » [66] sur Rstudio et plus précisément de la fonction `nipals`. Le modèle itératif NIPALS peut être interrompu à tout moment permettant à l'analyste de sélectionner le nombre de composantes A selon plusieurs critères fournis dans la partie *iii. ACP centrée sur les plantes* de ce chapitre. L'algorithme NIPALS fonctionne comme suit [42] :

L'équation 2-2 est réécrite comme ceci :

$$\text{Equation 2-3} \quad E_a = E_{a-1} - t_a p_a^T$$

Avec E_a les résidus de la nouvelle itération calculée à partir des résidus de l'ancienne E_{a-1} .

On peut aussi définir le vecteur propre p_a^T tel que :

$$\text{Equation 2-4} \quad E_{a-1} = t_a p_a^T$$

Qui peut aussi s'écrire :

$$\text{Equation 2-5} \quad t_a^T E_{a-1} = t_a^T t_a p_a^T$$

Et finalement :

$$\text{Equation 2-6} \quad p_a^T = \frac{t_a^T E_{a-1}}{t_a^T t_a}$$

On démarre alors avec l'itération 0 suivante :

$$\text{Equation 2-7} \quad E_0 = X$$

X étant la matrice des données d'entrée centrée, c'est-à-dire avec la valeur moyenne soustraite à toutes les données. On fixe alors la colonne de la matrice X qui présente la variance la plus élevée comme première colonne de la matrice t_1 . Le premier vecteur propre p_1 peut alors être calculé à l'aide de l'équation 2-6 :

$$\text{Equation 2-8} \quad p_1^T = \frac{t_1^T E_0}{t_1^T t_1}$$

p_1 est ensuite normée et le résidu à l'itération 1 peut ensuite être calculé à l'aide de l'équation 2-3 :

Equation 2-9
$$E_1 = E_0 - t_1 p_1^T$$

Le nouveau résidu E_1 permet alors de calculer la seconde colonne de la matrice des scores T soit t_2 :

Equation 2-10
$$t_2 = E_1 \frac{p_1}{p_1^T p_1}$$

La valeur de t_2 permet de calculer le vecteur p_2 et ainsi de suite.

Finalement on calcule les vecteurs propres à l'aide de cette relation :

Equation 2-11
$$\hat{t}_a = t_a^T t_a$$

Lorsque $(\hat{t}_a - \hat{t}_{a-1}) \leq 0,00001$ on estime alors que le calcul itératif a convergé.

iii. ACP centrée sur les plantes

L'ACP peut être calculée sur des données brutes, des données centrées ou encore des données centrées-réduites. Centrer les données signifie réaliser la soustraction de la valeur moyenne des intensités pour chaque variable. Cela revient à centrer sur 0 les points représentant les observations/échantillons dans les graphes des scores. Centrer-réduire les données signifie réaliser d'abord cette soustraction et de normaliser ensuite les données, de manière à ce que chaque colonne de la matrice des données, c'est-à-dire chaque variable, présente une moyenne égale à 0 et une variance égale à 1. L'intérêt de cette normalisation, en divisant chaque colonne du jeu de données par la valeur de l'écart-type de la variable, est de donner la même importance à chaque variable en dépit des valeurs initiales d'intensité. Cette normalisation a de l'intérêt lorsque le jeu de données contient des variables présentant des unités différentes comme par exemple un tableau récapitulant les performances de plusieurs athlètes selon plusieurs épreuves sportives. Le fait de comparer directement des temps au 100 mètres avec des distances de lancer de javelots n'aurait pas de sens et biaiserait les résultats de l'ACP. Cependant, dans le cadre d'une ACP réalisée sur des données spectrales, cette transformation des données pourrait avoir du sens après sélection des raies élémentaires les plus pertinentes mais pas dans le cadre d'une ACP réalisée sur le spectre entier. En effet cela signifierait donner autant d'importance à des zones spectrales ne contenant pas d'information chimique qu'à des plages spectrales contenant des raies élémentaires caractéristiques de la composition de l'échantillon [65]. L'ACP réalisée sur les données LIBS des échantillons de plantes a donc été calculée à partir des données centrées.

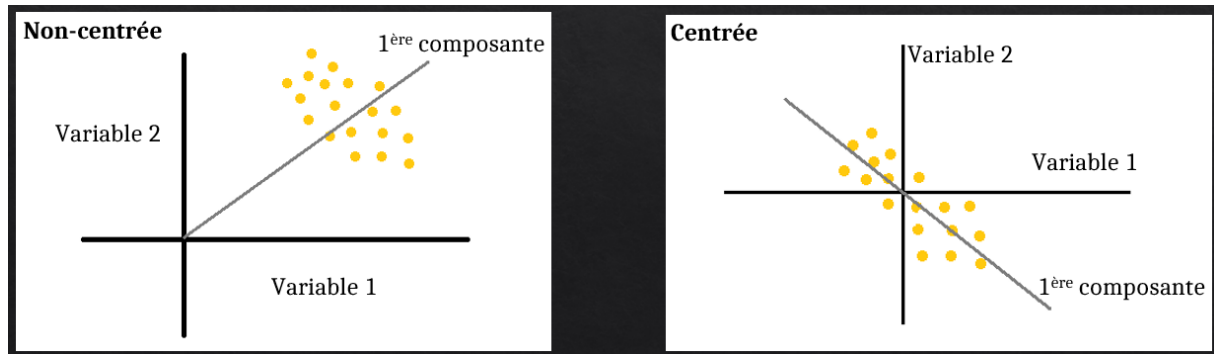


Figure 2-5 : Schéma présentant l'intérêt de centrer les données pour une ACP.

La première figure de mérite à observer à la suite d'un calcul ACP est le pourcentage de variance expliquée en fonction du nombre de composantes. Ce graphe obtenu pour les échantillons de plantes analysés par LIBS est donné en figure 2-6. Un tel graphe permet d'avoir une estimation rapide du nombre de composantes jugé suffisant à intégrer dans l'ACP afin d'expliquer à un niveau satisfaisant la globalité des variabilités spectrales du jeu de données. En règle générale, les projections sur les deux premières composantes sont observées car elles permettent d'expliquer la majorité de la variance. Cependant, même si les composantes suivantes n'expliquent individuellement que très peu de variance (entre 1 et 3%), elles peuvent contenir de l'information spectrale très importante et possiblement permettre de différencier les échantillons. Le pourcentage de variance expliquée a été calculé ici par cross-validation, c'est-à-dire que le jeu de données a été coupé en quatre segments : l'ACP a été réalisée sur trois d'entre eux et les scores du dernier segment ont été calculés par l'ACP réalisée sur les trois autres segments. Cette opération a été réalisée 50 fois avec une séparation des données différente à chaque fois. Ceci permet par la suite de calculer des barres d'erreur pour le pourcentage de variance expliquée, comme présenté par la figure 2-6.

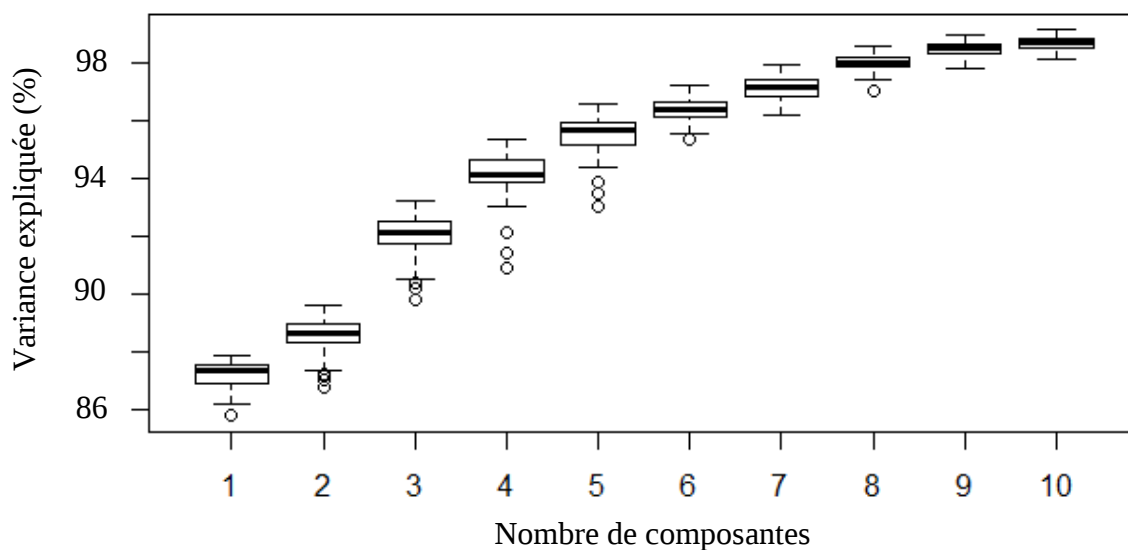


Figure 2-6 : Pourcentage de variance expliquée par l'ACP sur les plantes en fonction du nombre de composantes ajouté. Les boîtes à moustache ont pu être calculées à l'aide du processus de cross-validation décrit dans le texte.

Il est important de noter, à partir de la figure 2-6, que les deux premières composantes permettent déjà d'expliquer environ 89 % de la variance du jeu de données. Cette valeur semble ensuite atteindre un seuil autour de 97-98 % à partir de 5-6 composantes, signifiant donc que ce nombre de composantes semble suffisant pour comprendre la globalité des variabilités présentes dans les spectres LIBS entre les différents échantillons de plantes. Il convient donc de vérifier, à l'aide des graphes des scores et loadings, comment les échantillons de plantes se différencient en LIBS et à quelles variables/longueurs d'onde sont associées ces composantes. Il est dans un premier temps automatique de représenter la projection des points selon les deux premières composantes car elles expliquent la majorité de la variance (ici 89%).

a) ACP et information physique sur le plasma

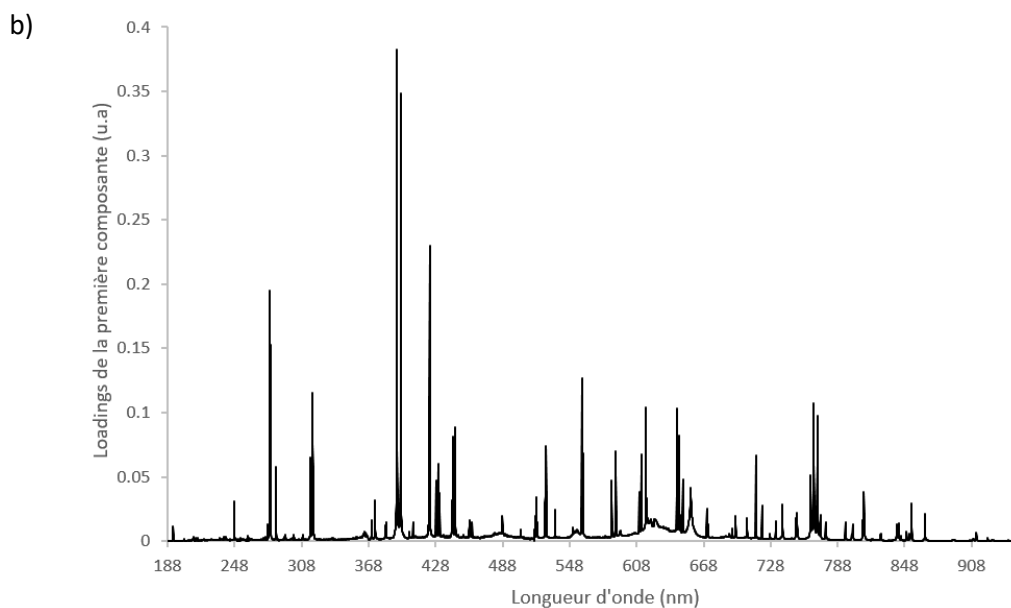
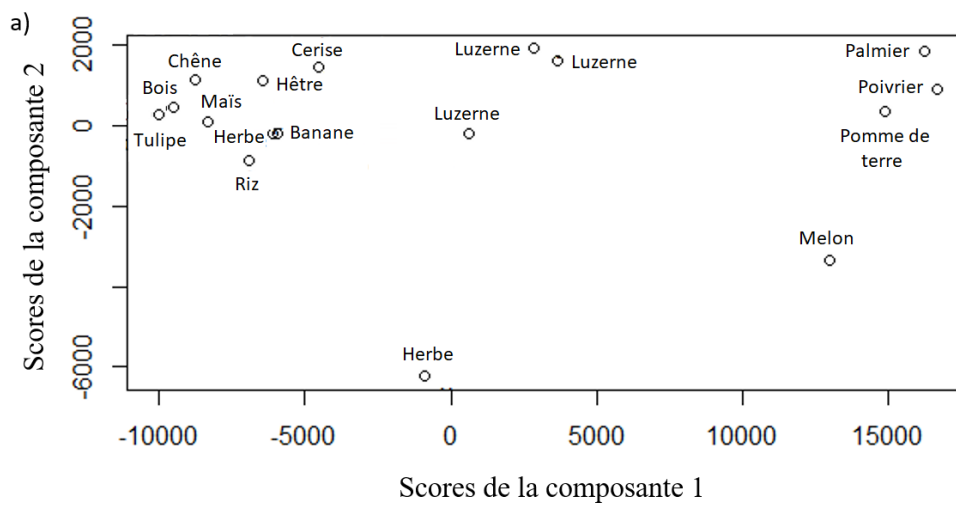


Figure 2-7 : Figures de mérite de l'ACP réalisée sur les échantillons de plantes avec a) le graphe des scores de la composante 1 en fonction de la composante 2 et b) le graphe des loadings de la composante 1.

Le graphe des scores sur les composantes 1 et 2, représenté sur la figure 2-7a., permet d'observer une bonne dispersion des échantillons le long des scores de la composante 1. En effet, il semble se dégager trois groupes selon cette première composante : le premier avec des scores négatifs, le deuxième avec une valeur de scores nulle ou positive et le dernier avec des scores positifs élevés. Cette dispersion des échantillons peut être spectralement comprise à l'aide tout d'abord du graphe des loadings de la première composante, donné en figure 2-7b. En effet, on peut remarquer que l'intégralité des raies élémentaires contenues dans le spectre LIBS d'une plante présente une valeur de loading positive, signifiant qu'en passant d'un échantillon à l'autre toutes les raies auraient tendance à augmenter de la même façon, peu importe la nature de l'élément. Ce phénomène n'a pas de sens chimique et est expliqué par un signal plus élevé probablement dû à une quantité de matière échantillonnée plus importante sur certains échantillons à cause d'une ablation laser plus efficace. La différence de quantité de matière ablatée d'un échantillon vient probablement des propriétés physiques de l'échantillon préparé sous forme de pastille, qui diffèrent d'un échantillon de plante à l'autre, ce qu'on appelle « l'effet de matrice ». Cette explication se confirme lorsque l'on observe la nature de chaque échantillon à l'aide de la figure 2-7a. qui met en lumière la formation de certains groupes. Les échantillons ligneux ou d'espèces ligneuses (chêne, bois, hêtre) ainsi que les échantillons de luzerne paraissent en effet se regrouper par famille selon les scores de la composante 1. Cependant le troisième groupe, présentant des valeurs positives de loadings pour la composante 1, est composé d'échantillons de plante d'espèces variées (Melon, Pomme de terre, Poivrier, Feuille de palmier). Lorsque l'on observe l'intensité d'une raie de l'argon (ici celle à 763 nm), il est pourtant clair que les conditions de plasma pour ces 4 échantillons semblent particulièrement différentes. On estime en effet que, l'argon étant le gaz dans lequel le plasma se forme pour nos conditions d'analyse, le niveau d'intensité LIBS des raies associées à cet élément nous permet alors d'avoir une indication sur l'émissivité du plasma. Le graphe de l'intensité de la raie de l'argon à 763 nm en fonction de l'aire totale sous le spectre, c'est-à-dire la somme de toutes les intensités contenues dans la gamme spectrale couverte par le pistolet Z-300, est représenté sur la figure 2-8. On peut y observer une distribution similaire à celle obtenue sur la composante 1 de l'ACP, ce qui n'est pas étonnant étant donné que la composante 1 représente la globalité des raies élémentaires détectées par le Z-300 ou autrement formulé l'aire totale sous le spectre. On remarque également une certaine corrélation entre l'aire totale sous le spectre et l'intensité nette de la raie de l'argon. A travers ce graphe, il peut donc être déduit que le plasma formé lors de l'analyse LIBS était de qualité différente (température, densité électronique) pour chaque famille d'échantillon et certainement plus intense lors de l'analyse de ces quatre échantillons (Melon, Pomme de terre, Poivrier, Feuille de palmier). Cette figure permet de mettre en évidence l'intérêt d'utiliser les raies élémentaires de l'argon comme outil afin d'évaluer les différences potentielles de qualité de plasma, et donc de quantité de lumière collectée entre échantillons. A la suite de ce constat, le signal LIBS de l'argon est exploité comme étalon interne dans le but de normaliser le signal LIBS dans la partie II.2.i *Normalisation*. Cependant, après observation des caractéristiques chimiques de ces échantillons et à cause d'un manque d'information sur le type de broyage effectué et sur la partie de la plante analysée, il n'a pas été possible de comprendre pourquoi ces quatre échantillons d'espèces variées se rassemblent en un même groupe.

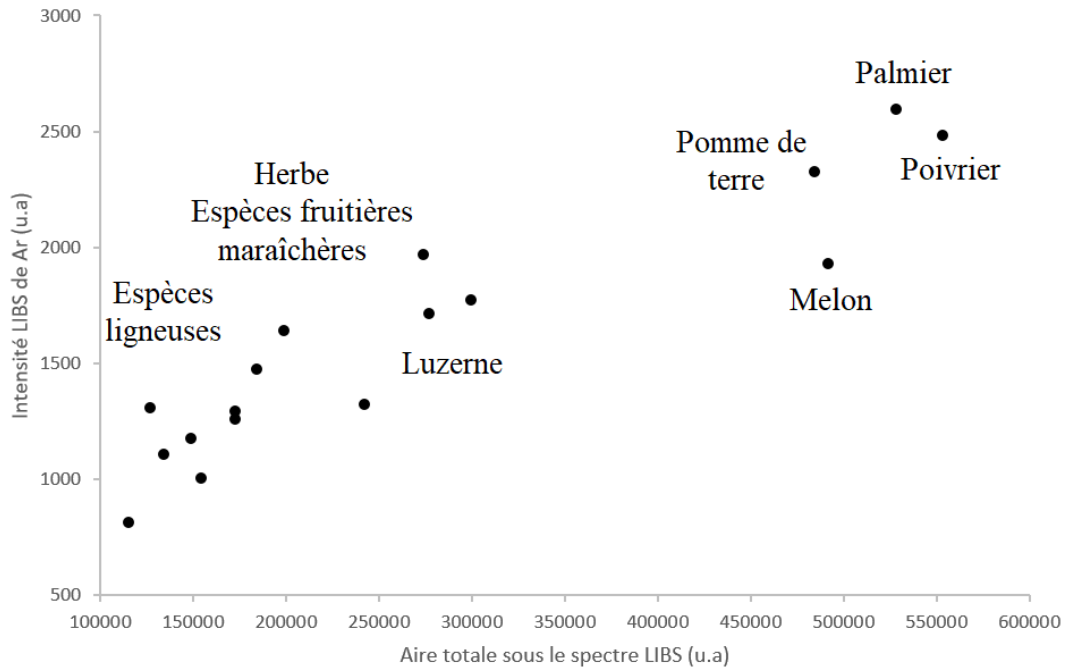


Figure 2-8 : Graphe de l'aire totale sous le spectre en fonction de l'intensité nette de la raie de Ar à 763 nm pour les échantillons de plante analysés par LIBS.

b) ACP et information chimique

La dispersion des points associée à la composante 2, d'après la figure 2-7a., est principalement décrite par un nuage de points autour de la même valeur et de deux points éloignés négativement de ce nuage, plus précisément l'échantillon de Melon (950) et celui de Graminée (100). Le graphe des loadings de la composante 2, en figure 2-9, est principalement marqué par une valeur de loading négative élevée pour le doublet du sodium à 589 nm, signifiant que plus l'échantillon présente une valeur de score négative sur la composante 2 et plus son signal LIBS relatif au sodium est intense. Les échantillons 950 et 100 présentent bien les valeurs d'intensité à 589 nm les plus élevées en bon accord avec les valeurs de concentration en sodium (3.0 et 3.3 g/kg respectivement), le reste des échantillons de cette étude étant concentré à des valeurs de l'ordre du mg/kg.

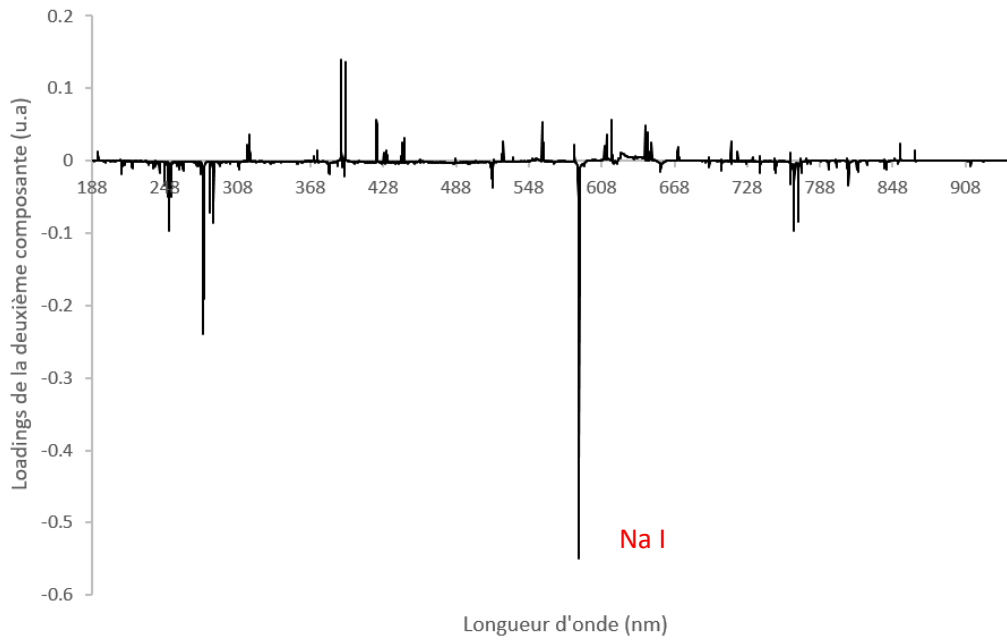


Figure 2-9 : Graphe des loadings de l'ACP réalisée sur les échantillons de plantes pour la composante 2.

Les composantes suivantes 4, 5 et 6 (expliquant à elles trois environ 4% de la variabilité des données) sont associées aux raies des éléments majeurs détectés restants tels que K, Mg ou encore Si d'après les graphes des loadings proposés en figure 2-10. Quoiqu'il en soit, la composante permettant de différencier au mieux les échantillons semble rester la première même si le graphe des scores de la composante 4 en fonction de la composante 6 permet d'observer une bonne dispersion des données. Le graphe est disponible en annexe sur la figure 2-1. Les composantes suivantes expliquent individuellement moins de 1 % de variance du jeu de données car elles sont attribuées à la variabilité aléatoire présente dans les zones bruitées sans signal du spectre. Ce phénomène, illustré sur la figure 2-2 en annexe se traduit par une allure bruitée des loadings et des formes sinusoïdales, c'est-à-dire des loadings présentant des valeurs successivement négatives, positives et nulles pour une même raie élémentaire.

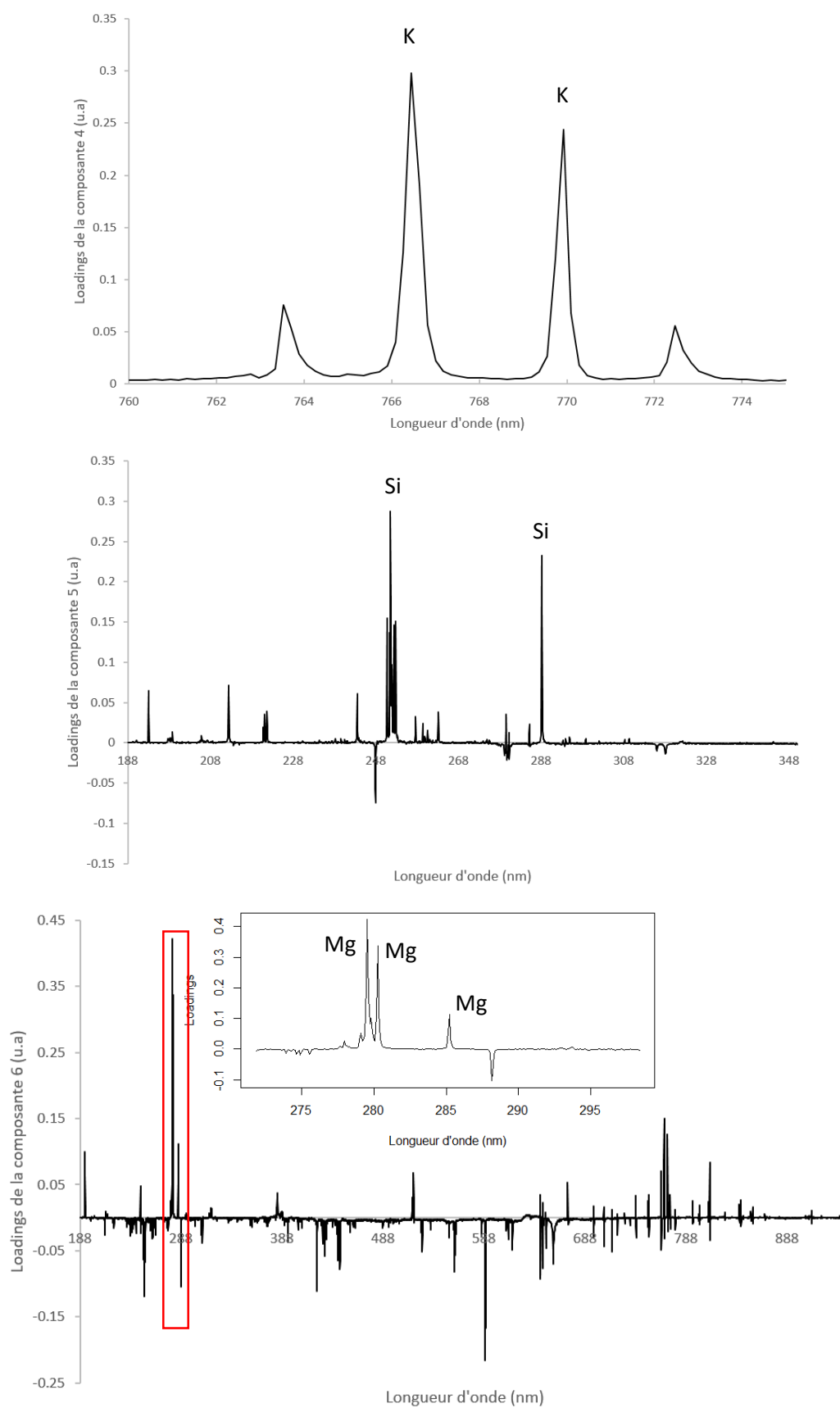


Figure 2-10 : Graphe des loadings zoomé sur les raies influentes pour la composante 4, 5 et 6 de l'ACP réalisée sur les plantes.

c) Détection de raies auto-absorbées par ACP

En plus de révéler des similitudes ou des différences entre les spectres liées aux propriétés physiques et/ou à la composition chimique des échantillons, l'ACP permet aussi de détecter la présence de données pouvant être considérées comme anormales dans les spectres. C'est tout l'intérêt de la composante 3 dans le modèle ACP présenté ici. Le graphe des loadings de la composante 3 montre clairement la contribution du doublet du calcium entre 393 et 397 nm (cf. figure 2-11). Cependant, on constate pour ces raies que certaines valeurs sont négatives alors que d'autres sont positives et que d'autre part, aucune autre raie d'émission du calcium n'est révélée par cette composante de l'ACP. On en déduit que les loadings présentés sur la figure 2-11 révèlent une anomalie pour ces raies du calcium avec notamment une anti-corrélation entre le sommet et les ailes de chaque pic.

Ce type de profil de loadings peut être attribuée au phénomène d'auto-absorption au sein du plasma, selon lequel certains photons émis sont réabsorbés plus loin dans le plasma. Sur le spectre LIBS, cela se traduit par un abaissement du maximum d'intensité de la raie d'émission et par son élargissement. La mesure des largeurs à mi-hauteur pour le doublet de Ca permet en effet d'observer une différence de valeur de 0,1 nm environ entre les raies d'émission classiques et celles auto-absorbées. A titre d'illustration, le signal de Ca pour deux échantillons présentant des largeurs à mi-hauteur différentes est donné en annexe sur la figure 2-3.

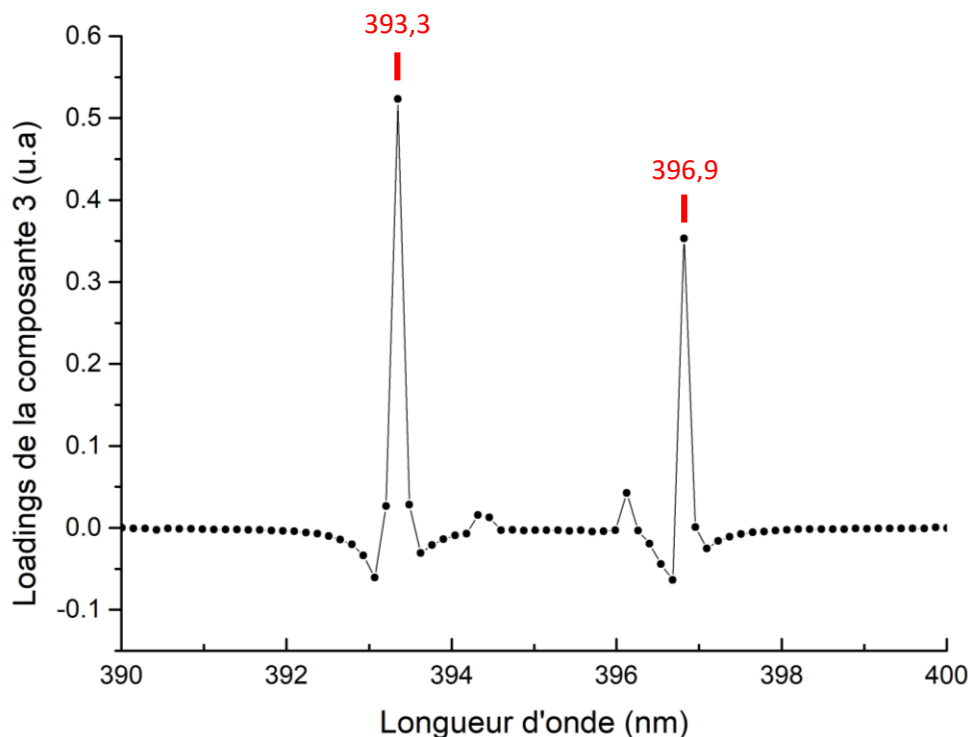


Figure 2-11 : extrait du graphe des loadings de la composante 3 zoomé sur le doublet du calcium à 393-396 nm.

L'analyse en composantes principales permet également de détecter si un échantillon peut être considéré comme « aberrant », on parle alors d'*outlier*. En ce qui concerne l'ACP réalisée sur les échantillons de plantes, aucun mauvais point de levier n'a été détecté et donc aucun point n'a été éliminé du jeu de données.

L'ACP a donc permis d'obtenir plusieurs types d'information sur les spectres LIBS des échantillons de plantes : i) une information relative à une possible différence de propriétés physiques entre échantillons ; ii) une information relative à la différence de composition chimique élémentaire des échantillons notamment pour les éléments K, Mg, Si, Na et Ca ; et iii) une information sur la présence de données « anormales » liées au phénomène d'auto-absorption et à éviter pour la suite de l'analyse. Une fois ces informations acquises, il est possible d'appliquer une stratégie de calibration pertinente pour l'analyse quantitative.

2) L'analyse quantitative des plantes

La stratégie d'étalonnage des échantillons de plante par LIBS est décrite dans le chapitre I mais peut être résumée à l'aide des étapes suivantes :

- Afin de bâtir les modèles d'étalonnage à partir des intensités LIBS, il est d'abord nécessaire de mesurer les concentrations en éléments totaux dans les échantillons de plante. Les concentrations de référence de ces échantillons ont été déterminées par ICP-AES (cf chapitre I). Une campagne de mesures a également été réalisée dans le but de pouvoir mesurer les concentrations par XRF à partir d'un instrument portable qui offre des conditions d'analyse proches de celles du pistolet Z-300. La discussion sur les résultats obtenus par XRF portable est proposée dans la partie *Conclusion sur l'analyse LIBS des échantillons de plante* de ce chapitre. Il est ensuite possible de créer ces modèles d'étalonnage par régression entre les intensités LIBS et les concentrations mesurées par les méthodes de référence.
- Une fois ces modèles d'étalonnage construits à partir du jeu de données initial (ici 17 échantillons de plante d'espèces variées), également appelé jeu de calibration, il est nécessaire de valider la fiabilité de ces modèles en prédisant la concentration d'un autre jeu de données, appelé jeu de validation, composé de 8 échantillons de plante n'ayant pas servis à construire les modèles de calibration.

i. Normalisation

L'analyse quantitative en spectroscopie repose sur la corrélation entre l'intensité du signal et la concentration dans l'échantillon. Cependant le signal exploité pour construire le modèle d'étalonnage est parfois sujet à des fluctuations qui détériorent les performances du modèle. Il est donc courant d'appliquer des étapes de prétraitement afin d'atténuer ces fluctuations, à travers des techniques de dé-bruitage ou de normalisation. En spectroscopie proche infrarouge par exemple, la granulométrie pour les solides ou la turbidité pour les liquides peuvent entraîner de telles fluctuations à travers un mécanisme de diffusion de la lumière. Dans ce cas, cela se

traduit par un effet multiplicatif sur l'intensité nette du signal mesuré et additif sur l'absorbance. Selon le cas, on calcule les fonctions dérivées à partir de l'algorithme Stavitsky Golay par exemple ou encore de simples opérations mathématiques comme celles conduisant à la normalisation SNV (Standard Normal Variate) [67]. En ICP-AES, dans un autre contexte, les fluctuations de signal liées aux fluctuations des conditions de plasma sont éliminées par le biais de l'utilisation d'un étalon interne [68, 69]. Cette stratégie est détaillée dans la partie *Etat de l'art de la normalisation en LIBS* de ce chapitre. La LIBS ne déroge pas à la règle du besoin de normalisation étant donné que le signal enregistré est lié aux conditions de création et d'expansion du plasma. Les fluctuations potentielles proviennent d'un changement des conditions de mesure (énergie laser) mais aussi des propriétés physiques de l'échantillon (rugosité de surface, absorption optique, diffusivité thermique...). La question de la normalisation en LIBS ne fait pas consensus mais elle est centrale dans le cadre de cette étude basée sur l'utilisation d'un instrument portable potentiellement exposé à plus de fluctuations qu'un instrument de laboratoire. C'est pour apporter des éléments de réponse à cette question de la normalisation que nous avons effectué un travail bibliographique qui a donné lieu à un article scientifique de revue. Les méthodes de normalisation qui y sont discutées concernent des stratégies de calcul des données spectrales et non pas l'utilisation de données supplémentaires telles que le signal acoustique, la température du plasma ou encore la quantité de matière ablatée. L'article complet, intitulé « Critical review and advices on spectral-based normalization methods for LIBS quantitative analysis » est disponible en annexe.

a) Etat de l'art de la normalisation en LIBS

Ce paragraphe est basé sur une compilation de références issue de l'article de revue de Guezenoc et al. (2019) [24], cité précédemment et disponible en annexe.

A l'instar des autres méthodes spectroscopiques, la question de la normalisation en LIBS est un sujet récurrent dans les publications scientifiques et reste encore aujourd'hui une question ouverte pour plusieurs raisons. Tout d'abord, dans les études ayant fait l'objet de publications, la normalisation a été appliquée dans des conditions de mesures très différentes (énergie laser, spectromètre, gaz ambiant, distance entre l'échantillon et le laser, nombre de tirs, nombre de zones analysées...), pour analyser des matériaux de natures diverses et avec des prétraitements de données significativement différents, rendant impossible toute comparaison. Ajouté à cela, on observe aussi parfois un manque de robustesse dû à de mauvaises pratiques, avec par exemple des modèles de calibration bâtis à partir de 3 points seulement, et trop peu de détails fournis sur le traitement des données. Il devient donc très compliqué de transposer une stratégie de normalisation d'un contexte analytique à un autre. Finalement, le modèle physique utilisé pour la méthode LIBS auto-calibrée (CF-LIBS) et décrivant l'émissivité du plasma pour chaque raie révèle que les variations d'intensité des raies en fonction de la température ne suivent pas un comportement linéaire [70]. La température du plasma ne varie pas non plus de façon linéaire avec l'énergie ou la fluence laser. Au final, les différentes raies atomiques ne varient pas de manière homothétique, ce qui laisse craindre que tout effort de normalisation soit vain et que, si une stratégie de normalisation est avantageuse pour une raie, il n'y a aucune raison qu'elle le soit pour les autres. Ainsi, notre étude conclut sur le fait que le choix de la normalisation la plus adaptée dépend de la raie d'émission atomique considérée.

Il ressort de cet article de revue 4 méthodes de normalisation qui sont les plus fréquemment appliquées en LIBS :

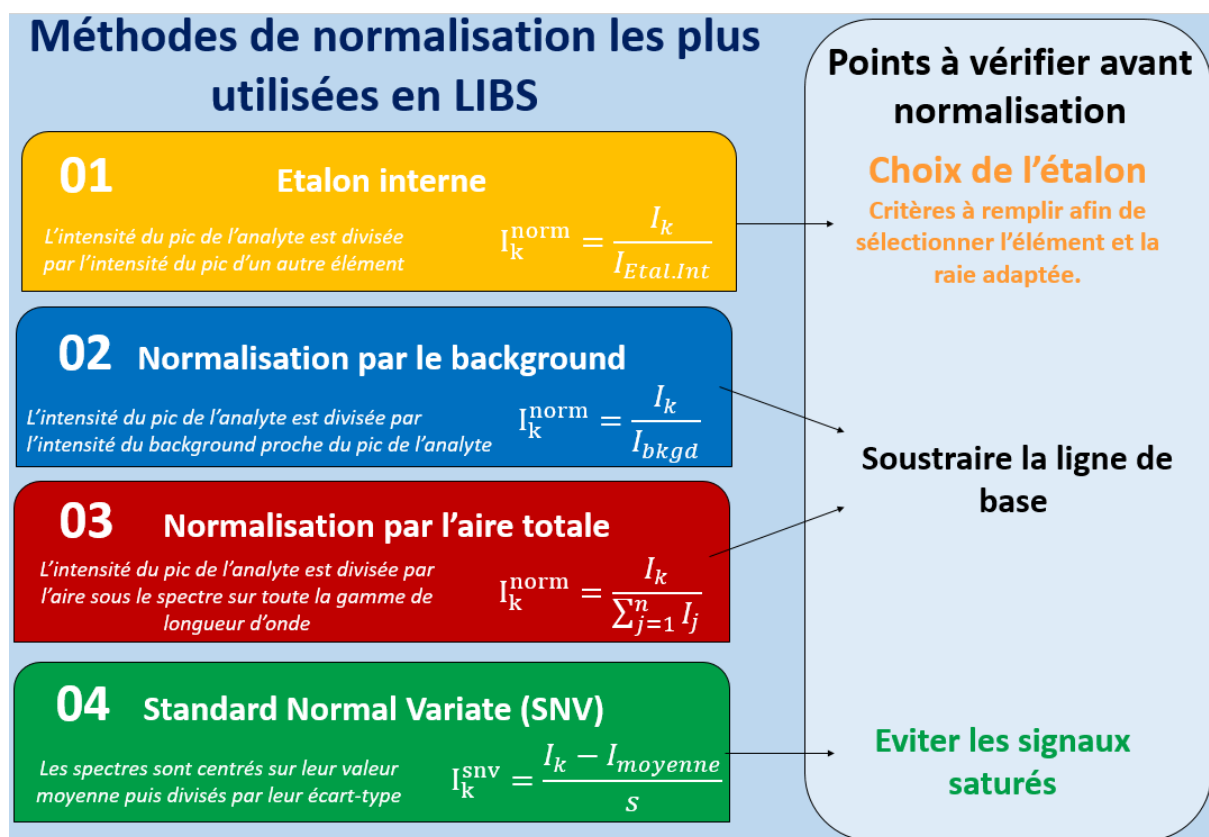


Figure 2-12 : Présentation des quatre méthodes de normalisation les plus utilisées en LIBS ainsi que les précautions à prendre avant de les appliquer.

Chacune de ces normalisations est appliquée à l'aide d'un calcul mathématique simple et en suivant quelques précautions. Le détail fin de ces précautions est fourni dans la version intégrale de l'article en annexe.

Notons enfin que certains modèles d'étalonnage ayant été construits à partir de données non normalisées ont parfois été décrits comme étant les plus performants. Cela pose évidemment la question du choix des figures de mérite permettant de comparer les modèles entre eux.

b) Figures de mérite pour la sélection de la meilleure normalisation

Nous avons proposé une stratégie analytique permettant de sélectionner la méthode de normalisation la plus adaptée. Il s'agit de comparer, pour chaque raie d'émission individuellement, les performances du modèle de calibration non-normalisé avec celles des modèles normalisés sur la base de plusieurs figures de mérite : le coefficient de détermination r^2 ; l'erreur du modèle à travers le calcul de l'erreur quadratique moyenne appliquée soit directement au jeu de calibration (RMSEC), soit selon une approche de validation croisée (RMSECV), soit enfin pour le jeu de validation n'ayant pas servi à la construction du modèle (RMSEP) ; et ses limites de détection (LOD) ou de quantification (LOQ). Nous avons confirmé, à travers un exemple pratique, que ces figures de mérite devaient toutes être calculées et que

l'observation d'une seule d'entre elles, comme le coefficient r^2 par exemple, ne suffisait pas à évaluer les performances d'un modèle voire même pouvait entraîner une mauvaise interprétation de celui-ci. Nous avons aussi mis l'accent sur le fait que la stratégie d'étalonnage devait suivre des bonnes pratiques afin d'atteindre des conclusions robustes. Il faut notamment un nombre suffisant d'échantillons de calibration avec des valeurs de concentration bien réparties sur toute la gamme visée. Ces recommandations peuvent sembler évidentes mais on constate qu'elles sont malheureusement très peu suivies dans les articles rapportant des résultats d'analyses LIBS. Voici les formules de calcul de ces figures de mérite telles que nous les avons définies ([24], article de revue sur la normalisation disponible en annexe) et appliquées pour l'analyse quantitative de cette étude :

- Le coefficient de corrélation r (le coefficient de détermination étant le carré du coefficient de corrélation r):

Equation 2-12

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$

avec x_i la concentration mesurée de l'échantillon i et $\bar{x}$ la concentration moyenne du jeu d'étalonnage, y_i l'intensité LIBS de l'échantillon i et $\bar{y}$ l'intensité LIBS moyenne du jeu d'étalonnage, N le nombre d'échantillons et s_x et s_y la valeur des écart-types sur les concentrations et l'intensité du signal respectivement. Lorsque r^2 vaut 1 la corrélation est maximale et lorsqu'il vaut 0 il y a absence de corrélation.

- L'erreur du modèle (RMSE) :

Equation 2-13

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}}$$

avec x_i la concentration prédite par le modèle pour l'échantillon i , $\hat{x}_i$ la concentration de référence de l'échantillon i et N le nombre d'échantillons. La RMSE permet d'évaluer la capacité de prédiction du modèle. Lorsque tous les échantillons de la calibration sont pris en compte, on la note RMSEC (racine-carré de l'erreur d'étalonnage). Cependant, la valeur de RMSEC est sensible aux points aberrants et peut ne pas être assez représentative de la véritable erreur du modèle. Il est alors possible de calculer la quantité RMSECV (racine-carré de l'erreur de validation croisée). La méthode de validation croisée qui est la plus utilisée lorsque le nombre d'échantillons est petit, ce qui est le cas pour notre étude, est la méthode Leave-One-Out (LOO). Cette méthode consiste à retirer un point du jeu de données puis à bâtir un modèle de calibration à partir des $N-1$ points restants et enfin à calculer l'erreur (RMSEC) pour ce modèle. Ensuite le point exclu du premier modèle est réintroduit dans le jeu de données et un autre point est retiré ; un nouveau modèle est alors construit et une nouvelle valeur de RMSEC est calculée sur ce nouveau jeu de données. Et ainsi de suite jusqu'à ce que tous les points aient été retirés. La valeur de RMSECV est finalement obtenue en calculant la moyenne de toutes les valeurs de RMSEC calculées au cours de ce processus. L'indicateur RMSECV, passant en revue l'impact de chaque point sur la valeur de coefficient de corrélation du modèle, permet de se rapprocher de la vraie valeur de l'erreur. Finalement, à l'aide du modèle d'étalonnage créé, il est possible de prédire les concentrations d'un nouveau jeu d'échantillons, non utilisé dans la construction du modèle. L'erreur calculée avec la même formule sur les valeurs de ce nouveau

jeu de données (jeu de validation) est appelée RMSEP (racine-carré de l'erreur de prédiction). Une valeur de RMSEP bien plus élevée que celle du RMSEC ou RMSECV signifie que les performances initiales du modèle ont été surestimées. La valeur de ces différents types d'erreur est censée être la plus basse possible.

- La limite de détection et de quantification :

Equation 2-14
$$LOD = \frac{3s}{a}$$

Equation 2-15
$$LOQ = \frac{10s}{a} = \frac{10}{3} LOD$$

avec s l'écart-type du bruit calculé à proximité de la raie d'émission considérée pour bâtir le modèle et a la pente de la droite de tendance du modèle d'étalonnage de l'analyte. Cette façon de calculer les limites appelée méthode de la pente a été discutée, ainsi que d'autres méthodes, en profondeur par J.-M. Mermet [71, 72] et a été sélectionnée dans le cadre de notre étude. La LOD et la LOQ permettent respectivement d'estimer à partir de quelle concentration l'instrument LIBS peut respectivement détecter et quantifier l'analyte. Notons cependant que dans le cadre de l'analyse par instrument portable d'échantillons naturels complexes, la limite de quantification n'est évidemment pas un indicateur prioritaire dans la mesure où l'on accède plutôt à une analyse semi-quantitative.

Finalement, à l'aide de l'observation de ces figures de mérite calculées par l'analyste, il est maintenant possible d'évaluer quel modèle d'étalonnage présente les meilleures performances et donc de sélectionner la meilleure méthode de normalisation. Cette méthode de sélection est résumée graphiquement sur la figure 2-13.

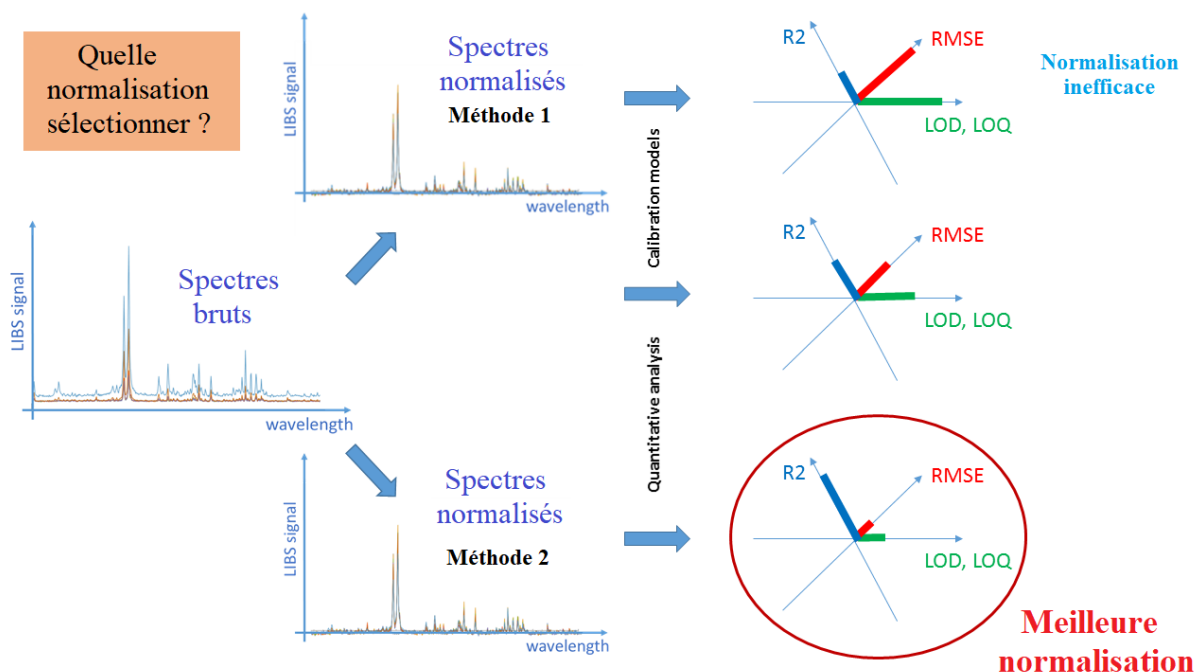


Figure 2-13 : Schéma récapitulant la stratégie de sélection de la meilleure normalisation.

Attention, selon le besoin de l'analyste, la décision de la meilleure stratégie de normalisation à adopter peut être différente. Par exemple, si celui-ci souhaite prédire de faibles concentrations plutôt que d'avoir un modèle robuste sur des concentrations plus élevées, il est plus intéressant alors d'atteindre la LOD la plus basse possible. Au contraire, s'il souhaite obtenir un modèle d'étalonnage particulièrement robuste sur l'ensemble de la gamme, il peut être plus intéressant pour lui de privilégier une valeur de RMSE faible. Dans le cadre de notre étude, impliquant l'utilisation d'un instrument portable et des échantillons complexes, nous avons décidé de privilégier un modèle présentant une valeur de RMSE la plus faible possible.

c) Application aux échantillons de plantes

La stratégie d'étalonnage et de sélection de la meilleure normalisation présentée ici a été appliquée à l'analyse LIBS quantitative des échantillons de plantes décrits plus haut. Plus précisément, 4 types de normalisation ont été retenues : la normalisation SNV, la normalisation par l'aire totale du spectre, la normalisation par étalon interne et la normalisation par le fond continu. Le choix de l'étalon interne s'est fait naturellement dans la mesure où le plasma LIBS se développe du gaz d'argon. On a ainsi choisi une raie d'émission de l'argon à 763 nm. En effet, on considère que l'émissivité des raies d'argon constitue un bon indicateur des propriétés du plasma. L'aire totale du spectre a été calculée en prenant la somme des intensités décrivant le spectre sur l'ensemble de la gamme de longueurs d'onde couverte par l'instrument (188 – 950 nm). En ce qui concerne le calcul de la normalisation SNV, il a été effectué après suppression des raies très émissives et auto-absorbées du calcium, plus précisément le doublet de Ca II à 393,3 et 396,9 nm et la raie de Ca I à 422,7 nm. Finalement, pour chaque élément étudié (dont toutes les raies d'émission ont été observées), ne seront présentées que la courbe d'étalonnage non-normalisée et celle normalisée présentant les figures de mérite les plus performantes, c'est-à-dire présentant le r^2 le plus élevé et l'erreur et la limite de détection les plus faibles. Les performances de calibration étant dépendantes du choix de la normalisation et de l'élément observé, comme le résumé de l'article de revue le souligne, les modèles d'étalonnages seront donc présentés par type de normalisation. Les barres d'erreur affichées sur les intensités LIBS correspondent à l'écart-type de l'intensité de la raie observée calculé sur les trois zones d'analyse effectuées par pastille. Celles affichées pour les valeurs de concentration mesurées par ICP-AES correspondent à l'écart-type calculé sur les valeurs de concentrations données par les différents laboratoires inclus dans le programme d'évaluation du WEPAL (cf chapitre I).

ii. Construction des modèles de régression

a) Calcul par normalisation SNV

Le choix de la raie sélectionnée pour réaliser la régression pour le calcium s'est basé sur le fait que beaucoup de signaux de Ca, comme révélé par l'ACP, présentent des profils auto-absorbés (le doublet à 393,3 – 396,9 nm et la raie à 422,7 nm en particulier). Le doublet de Ca II à 315,9 et 317,9 nm, a finalement été sélectionné car il correspond au signal le plus intense sans auto-absorption évidente. Les modèles d'étalonnage étant similaires pour les raies à 315,9 et 317,9 nm, seuls les résultats obtenus sur la raie à 315,9 nm seront présentés. Les courbes d'étalonnage

non-normalisée et normalisée SNV pour cette raie du calcium sont présentées sur la figure 2-14 a et b, respectivement.

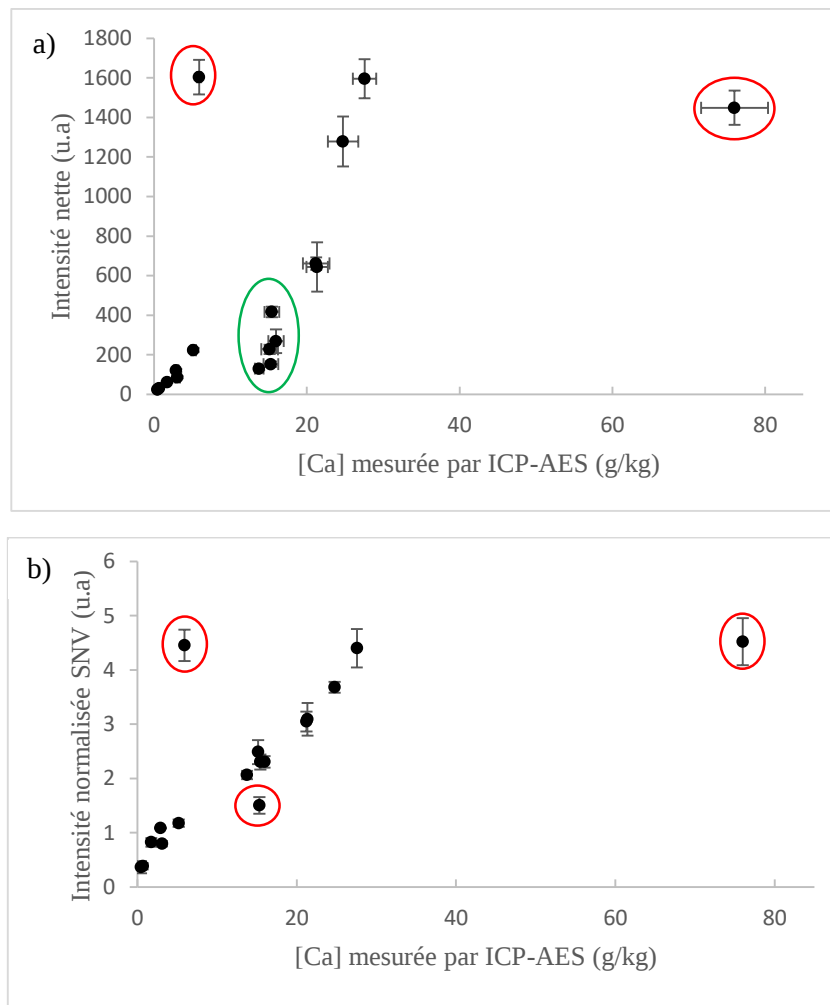


Figure 2-14 : Courbe d'étalonnage pour le calcium dans les plantes a) non-normalisée et b) normalisée SNV pour la raie Ca II à 315,9 nm. Les cercles rouges indiquent les outliers observés sur l'étalonnage. L'ellipse verte indique la présence d'un gradient d'intensités pour une même concentration en calcium sur les données nettes.

La courbe d'étalonnage obtenue sur les données nettes non normalisées met en lumière plusieurs anomalies. Premièrement, les intensités n'augmentent pas selon la même pente le long de la gamme d'étalonnage, laissant place à deux « régimes » de régression : le premier pour des valeurs de concentration entre 0 et 10 g/kg et le second pour des valeurs de concentration entre 10 et 30 g/kg, avec des pentes différentes. Deuxièmement, on observe la présence de 2 *outliers* indiqués par les ellipses rouges sur la figure 2-14a : la feuille de palmier à 5,9 g/kg et le melon à 76 g/kg. Enfin, un gradient d'intensités pour une même valeur de concentration (environ 16 g/kg) est signalé par l'ellipse verte sur la figure 2-14 a. La normalisation SNV, (cf. figure 2-14 b.), corrige en partie ces anomalies en livrant un seul régime de régression et en supprimant le gradient d'intensités qui était indiqué par l'ellipse verte sur la figure 2-14a à l'exception de l'échantillon 198 (banane), entouré en rouge sur la figure 2-14b. Notons que les deux *outliers* (palmier et melon) présents sur le premier modèle (figure 2-14a) restent des *outliers* dans le modèle normalisé SNV (figure 2-14b). Pour revenir au gradient d'intensité indiqué par l'ellipse verte (figure 2-14a), la figure 2-15 permet d'en faire une observation plus

approfondie. On constate (figure 2-15a) que les intensités brutes de la raie considérée sont significativement différentes mais que le fond varie aussi d'un spectre à l'autre. Cependant ces variations du pic et du fond ne permettent pas, de toute évidence, d'obtenir un signal net assez semblable pour ces quatre échantillons qui ont pourtant quasiment la même valeur de concentration en calcium. La normalisation SNV (figure 2-15b) supprime quant à elle les écarts entre trois des quatre spectres présentés, tandis que le spectre de l'échantillon de banane (198) reste à l'écart. Pour cet échantillon, une vérification de la valeur de concentration de référence serait à faire ainsi qu'une étude sur l'hétérogénéité et les propriétés de surface de la pastille analysée en LIBS. Dans l'immédiat, afin d'éviter d'impacter le modèle de régression final, cet échantillon de banane a été exclu du modèle d'étalonnage final pour le calcium, de même que les deux *outliers* (palmier, melon).

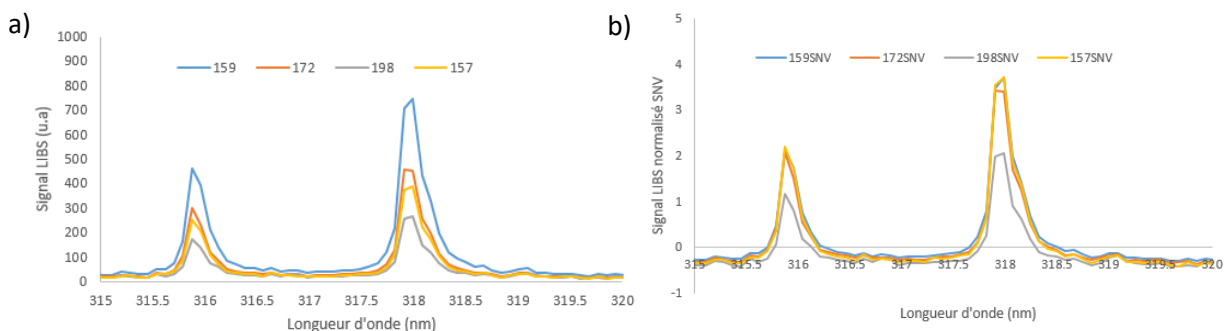


Figure 2-15 : Extraits de spectres LIBS des pastilles de plante présentant un gradient d'intensités pour une même valeur de concentration en calcium a) signaux brut non-normalisés et b) signaux normalisés SNV.

En ce qui concerne l'*outlier* correspondant à l'échantillon de melon (76 g/kg), il est plus de deux fois plus concentré que l'échantillon précédent dans la gamme d'étalonnage et peut présenter un effet de levier important sur la valeur du coefficient de corrélation [24].

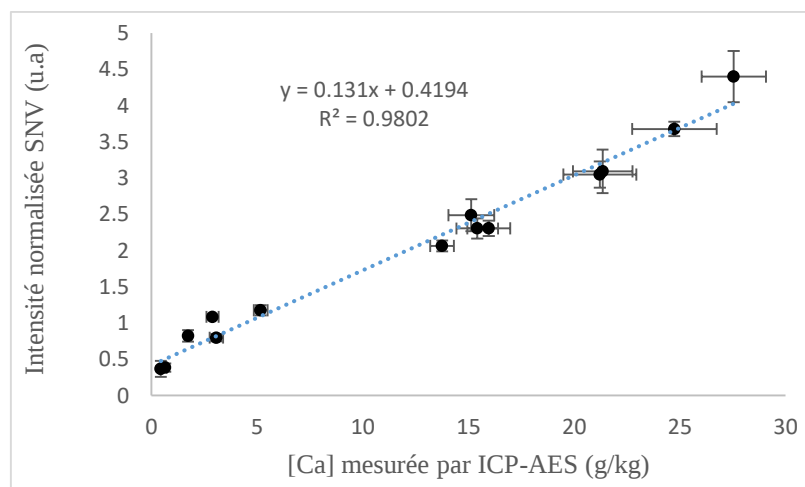


Figure 2-16 : Droite de calibration finale pour le calcium dans les plantes. Le trait en pointillés représente la droite de régression linéaire entre le signal LIBS et la concentration mesurée.

Le modèle SNV final pour le calcium (raie de Ca II à 315,9 nm), représenté sur la figure 2-16, est considéré comme le modèle présentant les meilleures performances avec : $r^2 = 0.98$, RMSE = RMSECV = 1.3 g/kg et LOD = 0.6 g/kg.

b) Potassium sans normalisation

Le seul signal du potassium détecté par le pistolet Z-300 pour les échantillons de plantes est le doublet de K I à 766,5 et 769,9 nm. Les performances analytiques obtenues sur les deux raies prises individuellement étant similaires, ne seront présentées que les résultats obtenus sur l'intensité nette de la raie à 766,5 nm. Le modèle de régression incluant les 17 échantillons de plantes initialement sélectionnés est représenté sur la figure 2-17.

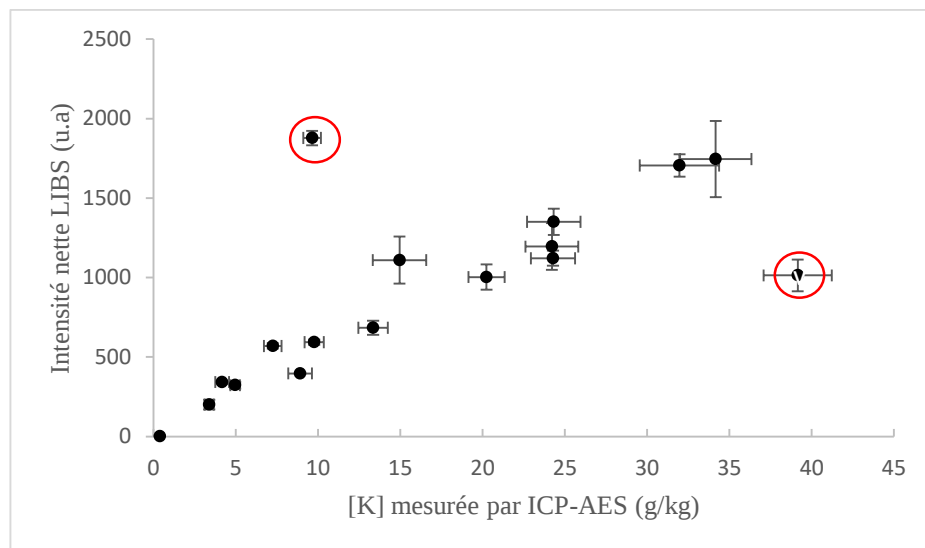


Figure 2-17 : Courbe d'étalonnage pour l'intensité nette de la raie KI à 766,5 nm dans les échantillons de plantes sélectionnés. Les cercles rouges indiquent les outliers.

Comme pour le calcium, deux *outliers* sont observés sur la courbe d'étalonnage du potassium alors que le reste des points montre une corrélation assez satisfaisante entre signal LIBS et concentration mesurée par ICP-AES. Il est intéressant de noter que les deux points aberrants détectés pour le potassium sont les mêmes que ceux observés pour l'étalonnage du calcium, confirmant leur statut d'*outlier*. Aucune normalisation n'a permis de corriger les écarts d'intensités de ces points ni d'améliorer la linéarité globale entre le signal et la concentration. Le modèle d'étalonnage final pour le potassium, présenté en figure 2-18, est donc celui basé sur les 15 spectres étalons non-normalisés, après exclusion des échantillons de palmier et de melon.

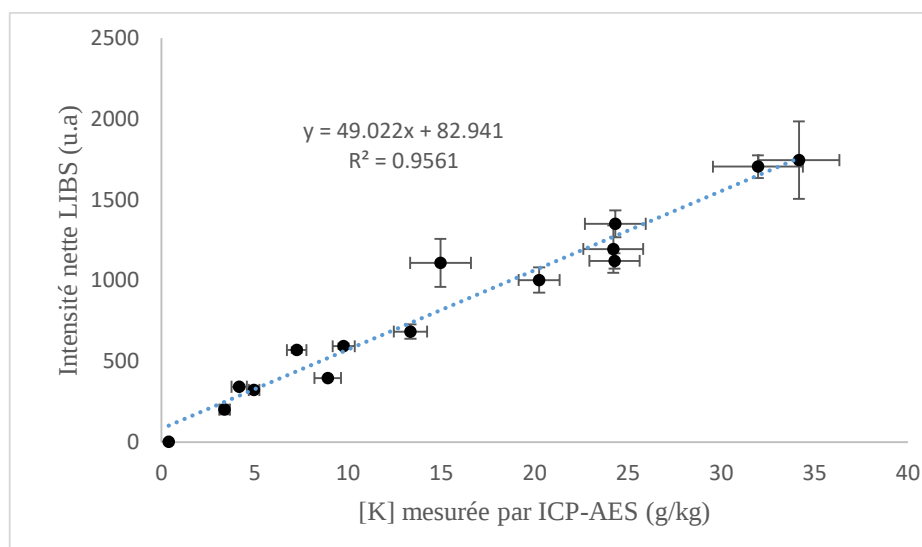


Figure 2-18 : Droite de calibration finale pour le potassium dans les plantes. Le trait en pointillés représente la droite de régression linéaire entre le signal LIBS (intensité nette de la raie de K I à 766,5 nm) et la concentration de référence.

La figure 2-18 permet donc d'observer une linéarité directe entre le signal LIBS et la concentration mesurée pour le potassium étant donné qu'aucune normalisation est appliquée. Le modèle final est caractérisé par : $r^2 = 0,96$, RMSE = 2,24 g/kg ; RMSECV = 2,22 g/kg et LOD = 0,43 g/kg.

c) Normalisation par étalon interne (Argon)

Les modèles d'étalonnage des éléments Fe, Al, Mg, P et Na ont présenté les meilleures performances à l'aide de la normalisation par étalon interne et plus précisément à l'aide de la raie de l'Argon I à 763,5 nm.

• Fer et aluminium

Les raies d'émission du fer sont très nombreuses dans un spectre LIBS mais dans notre cas, elles sont peu intenses. Un exemple de spectre LIBS dans une fenêtre spectrale comprenant uniquement des raies du fer est représenté sur la figure 2-4 en annexe. Il a donc été décidé d'étudier exclusivement la raie de Fe I à 302,1 nm qui est bien définie et libre de toute interférence. En ce qui concerne les raies de l'aluminium, le pistolet Z-300 permet de détecter les doublets de Al I à 394,4 et 396,2 nm et à 308,2 et 309,3 nm. Cependant, comme illustré par la figure 2-5 en annexe, le doublet à 394-396 nm est fortement interféré par le doublet très émissif de Ca II à 393,3 et 396,9 nm. Le modèle de régression a donc été construit à partir de l'autre doublet, à 308,2-309,3 nm. Enfin, pour le fer comme pour l'aluminium, le signal LIBS n'a pas pu être détecté pour 3 des 17 échantillons sélectionnés. Plus précisément cela concerne les trois échantillons ayant une concentration inférieure à 0,06 et 0,1 g/kg en fer et aluminium,

respectivement. Les modèles d'étalonnages sont donc construits sur les 14 échantillons offrant un signal utilisable.

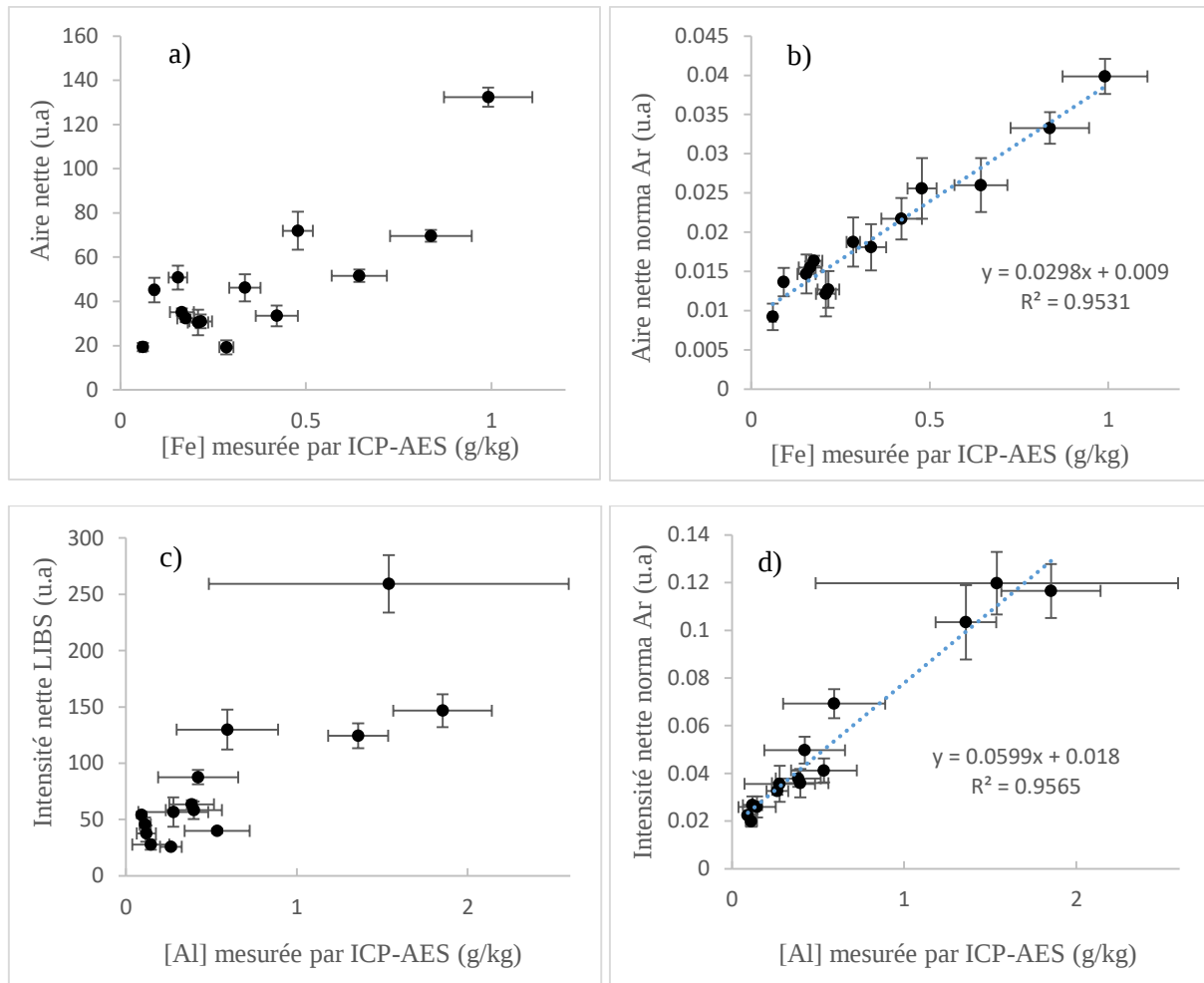


Figure 2-19 : Courbe d'étalonnage pour le fer (aire nette de la raie Fe I à 302,1 nm) dans les plantes a) non-normalisée et b) normalisée par l'Argon (intensité nette de la raie Ar I à 763 nm) et pour l'aluminium (intensité nette de la raie Al I à 308,2 nm) c) non-normalisée et b) normalisée par l'Argon (intensité nette de la raie Ar I à 763 nm).

Les figures 2-19 a. et 2-19 c. montrent que les modèles d'étalonnage pour le fer (aire nette de la raie de Fe I à 302,1 nm) et pour l'aluminium (intensité nette de la raie de Al I à 308,2 nm) ne sont pas exploitables. La normalisation à l'Argon (figures b et d) permet d'améliorer grandement ces deux modèles d'étalonnage. Notons par ailleurs la barre d'erreur très étendue pour la valeur déduite de l'analyse ICP-AES de l'échantillon le plus concentré en aluminium, susceptible d'introduire un biais sur le modèle d'étalonnage pour cet élément. On remarque enfin qu'en ce qui concerne l'aluminium, trois des échantillons sont beaucoup plus concentrés que ceux du reste de la gamme étalon et ont donc un effet de levier susceptible d'introduire un biais. Le modèle de l'aluminium est donc à considérer avec précautions et la valeur de $r^2 = 0,96$ est donc à nuancer. En effet, lorsqu'on ne prend pas en compte ces points aux plus fortes concentrations, on obtient un coefficient $r^2=0,80$. L'importance de points extrêmes sur la valeur

du r^2 est discutée en détails dans notre article consacré à la normalisation et disponible en annexe.

En ce qui concerne l'aluminium, le modèle final présente les valeurs de figures de mérite suivantes : un $r^2 = 0,96$, une $RMSE = RMSECV = 0,12$ g/kg et une $LOD = 0,12$ g/kg. Ces valeurs sont à nuancer étant donnée la répartition des points d'étalonnage.

Le modèle final pour le fer est quant à lui caractérisé par les figures de mérite suivantes : un $r^2 = 0,95$, une $RMSEC = 0,61$ g/kg, $RMSECV = 0,60$ g/kg et une $LOD = 0,11$ g/kg.

• Magnésium

Parmi toutes les raies d'émission détectées pour le magnésium, celle offrant la meilleure régression avec la concentration est la raie de Mg I à 285,2 nm. Cependant, comme l'illustre la figure 2-20 a., quatre *outliers* sont observés (ellipses rouges) et perturbent beaucoup l'allure générale de la courbe d'étalonnage bâtie à partir des 17 échantillons de plante sélectionnés. La normalisation à l'argon (raie Ar I à 763 nm) permet d'améliorer de manière significative le modèle d'étalonnage mais ne permet pas de totalement réintégrer les quatre points aberrants. Notons que parmi ces points, le plus concentré (10,6 g/kg) correspond au melon qui a déjà été considéré comme *outlier* pour les modèles du calcium et du potassium. Il a donc été logiquement décidé de l'éliminer du modèle d'étalonnage.

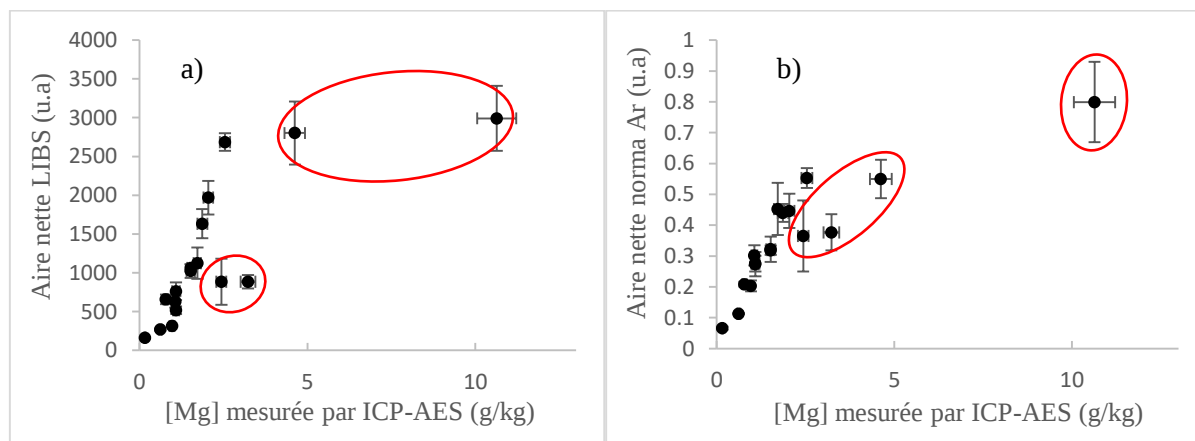


Figure 2-20 : Courbe d'étalonnage pour le magnésium dans les plantes à partir de la raie d'émission de Mg I à 285,2 nm a) aire nette sous le pic non-normalisée et b) aire nette sous le pic normalisée par l'Argon (intensité de la raie Ar I à 763 nm). Les ellipses rouges indiquent les outliers.

Finalement, après avoir supprimé les *outliers*, nous avons construit un modèle d'étalonnage sur la gamme de concentrations allant de 0,6 à 2,5 g/kg. Le modèle final est représenté sur la figure 2-21 et est caractérisé par un $r^2 = 0,94$, une $RMSEC = RMSECV = 0,14$ g/kg et une $LOD = 0,2$ g/kg.

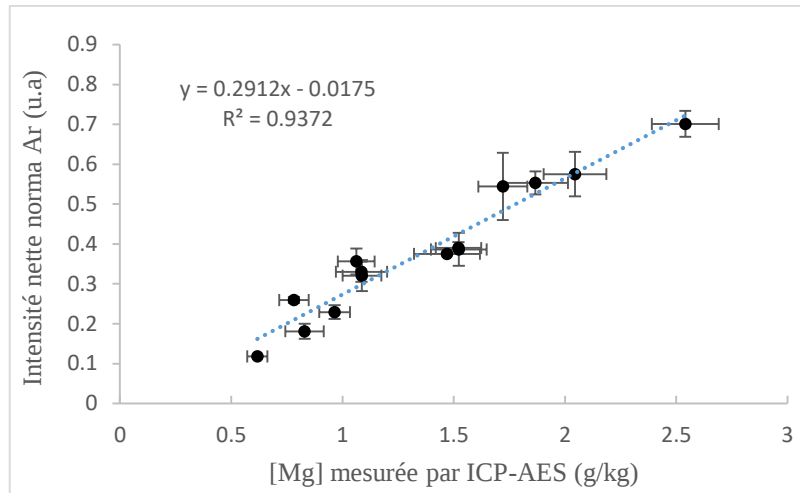


Figure 2-21 : Droite de calibration pour le magnésium (intensité nette de la raie de Mg I à 285,2 nm normalisée par l'intensité nette de la raie de Ar I à 763 nm) dans les plantes pour 13 échantillons. Le trait en pointillés représente la droite de régression linéaire entre le signal LIBS et la concentration mesurée.

- **Phosphore et Sodium**

Seul le doublet à 213,6 et 214,9 nm a pu être observé sur les spectres LIBS de plantes pour le phosphore. La raie à 213,6 nm étant sensiblement plus intense que l'autre, elle a donc été préférée pour construire le modèle d'étalonnage pour cet élément. Le signal de 3 échantillons, de concentration en phosphore inférieure à 0,7 g/kg, n'a pas été détecté.

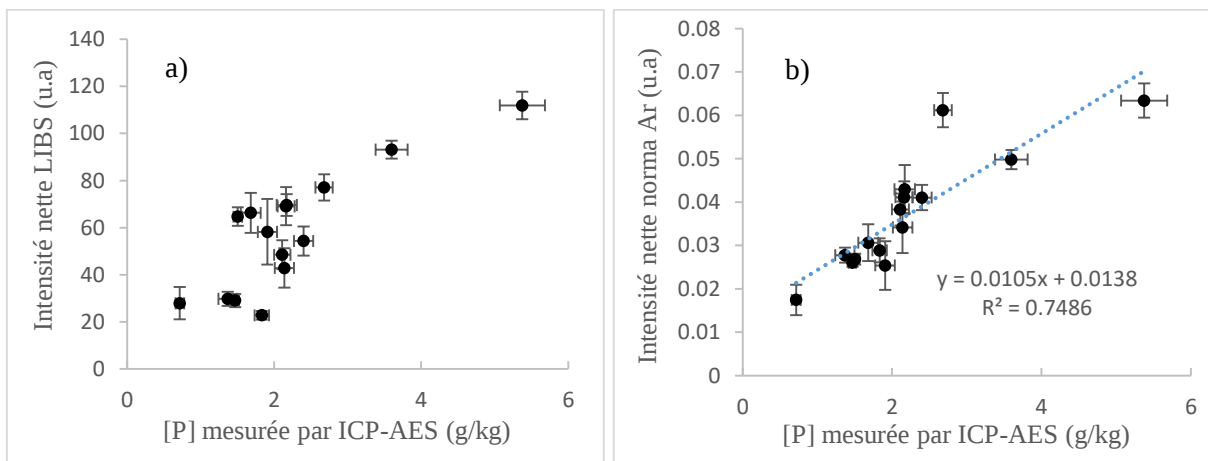


Figure 2-22 : Courbe d'étalonnage pour le phosphore (intensité nette de la raie de P I à 213,6 nm) dans les plantes a) non-normalisée et b) normalisée par l'intensité nette de la raie de Ar I à 763 nm. Le trait en pointillés représente la droite de régression linéaire entre le signal LIBS et la concentration.

La droite de régression finale, illustrée par la figure 2-22 b., obtenue pour le phosphore dans les plantes après normalisation à l'argon reste assez décevante. En effet, le coefficient de corrélation est inférieur à 0,8 mais surtout on observe un étalement vertical des points important, et contenant la majorité des points d'étalonnage, autour de la valeur de concentration de 2,1 g/kg malgré l'effort de normalisation. On peut conclure que, pour le phosphore, les analyses LIBS ne permettent pas de construire un modèle d'étalonnage fiable. Une alternative pourrait être de classer les échantillons analysés en deux classes, concentration inférieure ou supérieure à 2 g/kg. Le modèle final est représenté sur la figure 2-22b et est caractérisé par un $r^2 = 0.75$, une RMSEC = 0.61 g/kg, une RMSECV = 0.60 g/kg et une LOD = 0.7 g/kg.

Concernant le sodium, le seul signal détecté dans le spectre LIBS est le doublet de Na I à 589,0 et 589,6 nm. Les courbes d'étalonnage de la raie de Na I à 589,6 nm, d'abord non-normalisée puis normalisée par l'intensité de la raie de Ar I à 763 nm sont présentées sur la figure 2-23. Avant normalisation, on constate que le modèle est très influencé par les deux points associés à des valeurs de concentration supérieures à 3 g/kg, plus de trois fois plus élevées que les autres valeurs de concentration. Ainsi, nous présentons sur la figure 2-23b une courbe d'étalonnage après normalisation à l'argon sur une gamme de valeurs de concentration inférieures à 1 g/kg. Malgré cela, le modèle correspondant est caractérisé par un coefficient de corrélation $r^2 = 0,86$ très influencé par le point le plus concentré (autour de 1g/kg) tandis que la corrélation est très faible pour les plus basses valeurs de concentration et par une RMSEC = 0.11 g/kg, une RMSECV = 0.10 g/kg et une LOD = 0,3 g/kg.

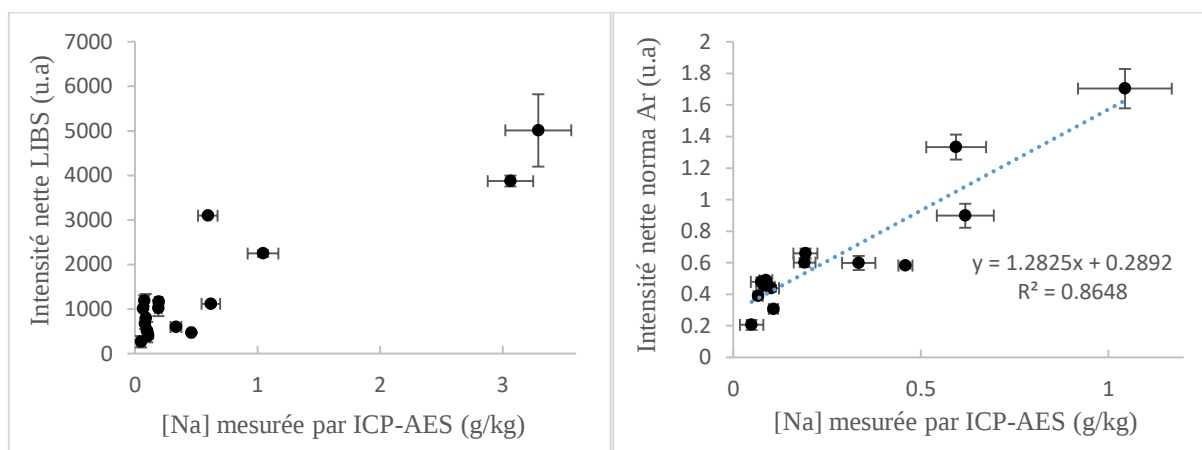


Figure 2-23 : Courbe d'étalonnage pour le sodium (intensité nette de la raie à 589,6 nm) dans les plantes a) non-normalisée et b) normalisée par l'intensité nette de la raie Ar I à 763 nm. Le trait en pointillés représente la droite de régression linéaire entre le signal LIBS et la concentration.

• Les éléments restants

L'élément carbone ne présente que très peu de variabilité dans le jeu de données sélectionné avec, comme le montre la figure 3-24 ci-dessous, des concentrations mesurées comprises entre environ 40 et 50%, ce qui n'est pas surprenant, sachant que le carbone, constituant majeur de la matière organique des végétaux et obtenu en quantité non limitante du CO₂ atmosphérique, est présent dans les plantes à une concentration moyenne peu variable d'environ 45 % [58]. Il n'y a donc pas d'intérêt de bâtir un modèle de régression à partir de ce jeu de données.

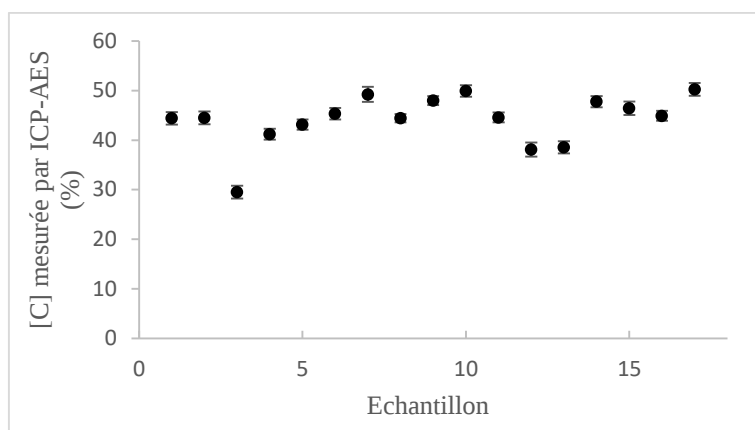


Figure 2-24 : Concentrations en carbone (%masse) mesurées pour les échantillons de plantes.

Enfin, les éléments tels que Mn, Zn, Cu et N, n'ont pas été détectés dans les spectres LIBS obtenus à l'aide du pistolet Z300 tandis que pour le silicium, détecté en LIBS, nous ne disposons pas des valeurs de référence. Il s'ensuit que, pour tous ces éléments, il n'est pas possible de construire un modèle d'étalonnage et donc de prédire des valeurs de concentration.

iii. Validation des modèles d'étalonnages

Il est de coutume, afin de valider les modèles d'étalonnages, de prédire la concentration d'échantillons non utilisés dans la construction des régressions. Le jeu de validation pour les plantes, comme décrit dans le chapitre I, est composé de 8 échantillons d'espèces différentes qui sont données dans le tableau 2-3 en annexe. Ce jeu de validation regroupe aussi bien des feuilles d'arbres issues de forêts diverses (myrcia, acacia, acajou) que des feuilles de canne à sucre ou de maïs. Ces échantillons ont été sélectionnés sur la base de leur disponibilité, de leurs concentrations en différents éléments contenues dans les gammes d'étalonnage construites à partir des 17 échantillons de plante décrits au chapitre I et de leur nature particulièrement variée, la plupart des échantillons n'étant pas présents dans le jeu de calibration. Le tableau 2-2 regroupe pour chaque élément d'intérêt, à savoir Ca, Fe, Mg, Al, K, Na et P, les valeurs de concentration mesurées par la méthode de référence (ICP-AES), celles prédites à partir des modèles d'étalonnage LIBS et l'incertitude sur ces valeurs prédites. Plus précisément, cette incertitude représente l'écart-type (en g/kg) des concentrations prédites obtenues sur trois réplicats réalisés par échantillon en LIBS. La notation HG signifie que la concentration de l'échantillon mesurée par ICP-AES est en dehors de la gamme d'étalonnage LIBS, si bien que la concentration ne peut donc pas être prédite par LIBS.

[Ca] g/kg	CFM1	CFV1	FAE2	FNF1	MAH3	MYR2	V463	IPE999
Mesurée par ICP-AES	2,9	2,3	8,7	6,4	11	4,8	2,6	19
Prédite par LIBS	3,1	2,4	9,7	7,7	9,8	5,1	3,8	18
Incetitude LIBS	1,1	0,5	0,8	1,6	1,3	0,6	0,4	1,8

[Fe] g/kg	CFM1	CFV1	FAE2	FNF1	MAH3	MYR2	V463	IPE999
Mesurée par ICP-AES	0,1	0,1	0,1	0,2	0,2	0,2	0,4	0,2
Prédite par LIBS	0,03	0,2	0,1	0,5	0,3	0,2	0,4	0,02
Incertitude LIBS	0,2	0,1	0,1	0,1	0,1	0,2	0,1	0,04
[Mg] g/kg	CFM1	CFV1	FAE2	FNF1	MAH3	MYR2	V463	IPE999
Mesurée par ICP-AES	1,8	1,7	1,5	2,1	3,0	2,1	1,5	2,2
Prédite par LIBS	1,7	1,6	1,6	2,4	HG	1,9	1,0	2,0
Incertitude LIBS	0,1	0,2	0,2	0,2		1,9	0,3	1,1
[Na] g/kg	CFM1	CFV1	FAE2	FNF1	MAH3	MYR2	V463	IPE999
Mesurée par ICP-AES	0,2	0,3	0,9	1,0	0,6	0,4	0,2	0,1
Prédite par LIBS	0,4	0,3	0,5	1,1	0,6	0,3	0,1	0,01
Incertitude LIBS	0,1	0,1	0,1	0,1	0,1	0,03	0,2	0,03
[P] g/kg	CFM1	CFV1	FAE2	FNF1	MAH3	MYR2	V463	IPE999
Mesurée par ICP-AES	0,5	1,7	0,3	0,6	0,2	0,2	2,4	2,8
Prédite par LIBS	HG	2,1	HG	HG	HG	HG	2,6	2,4
Incertitude LIBS		0,3					0,6	0,2
[Al] g/kg	CFM1	CFV1	FAE2	FNF1	MAH3	MYR2	V463	IPE999
Mesurée par ICP-AES	0,4	0,3	0,1	0,3	0,9	0,4	0,2	0,05
Prédite par LIBS	0,8	1,1	0,7	1,2	0,4	0,7	0,2	HG
Incertitude LIBS	0,1	0,2	0,3	0,4	0,1	0,1	0,04	
[K] g/kg	CFM1	CFV1	FAE2	FNF1	MAH3	MYR2	V463	IPE999
Mesurée par ICP-AES	5,4	15	2,5	3,7	1,8	2,5	19	16
Prédite par LIBS	8,5	22	0,6	13	3,2	1,4	21	11
Incertitude LIBS	1,7	0,4	0,2	1,3	0,3	1,2	0,8	0,7

Tableau 2-2 : Valeurs de concentration mesurées par ICP-AES et prédites par LIBS des éléments Ca, Fe, Mg, Al, K, Na et P pour les 8 échantillons de validation. Les incertitudes représentent l'écart-type des valeurs prédites à partir de 3 réplicats.

Pour les éléments reportés dans le tableau 2-2 sur fond vert, le modèle d'étalonnage LIBS est jugé satisfaisant ou validé, les valeurs prédites par LIBS sont proches de celles mesurées par ICP-AES et la valeur de RMSEP est acceptable, de l'ordre de celle de RMSEC ou de RMSECV.

A l'inverse, lorsque la valeur de RMSEP est très supérieure à celle de RMSEC ou RMSECV, le modèle n'est clairement pas validé et cela est indiqué par du rouge dans le tableau 2-2. Les valeurs de RMSEP sont toutes regroupées dans le tableau 2-3 présenté dans la partie *Conclusion sur l'analyse LIBS des échantillons de plante* de ce chapitre.

Enfin, certains modèles offrent des prédictions a priori satisfaisantes mais qui manquent cependant de robustesse. Ils sont indiqués en orange dans le tableau 2-2. C'est le cas des modèles associés aux éléments Na et P. Pour le phosphore, la majorité des échantillons de validation ont des valeurs de concentration en dehors de la gamme d'étalonnage. En ce qui concerne le sodium, si certaines valeurs prédites sont correctes, d'autres au contraire présentent un facteur 2, voire 10, par rapport aux valeurs issues des mesures ICP-AES (IPE999). Les discussions concernant ces résultats sont développées dans la section suivante.

3) Conclusion sur l'analyse LIBS des échantillons de plante

- i. Analyse qualitative LIBS
 - a) Détection

A travers les différents résultats obtenus dans ce chapitre, il est maintenant possible de dresser un bilan des performances analytiques obtenues à partir de l'instrument portable Z-300 de SciAps dans un contexte d'analyses contrôlées en laboratoire pour des échantillons de plantes préparés sous forme de pastilles. Tout d'abord, nous avons montré qu'il est relativement facile de détecter les raies d'émission atomique associées aux éléments pour lesquels les valeurs de concentration sont généralement de l'ordre du g/kg, à savoir Al, C, Ca, Na, Fe, K, Mg, Si et P. Seul l'élément majeur N, malgré sa concentration importante dans les végétaux, n'a pas été détecté dans les spectres, très certainement à cause de la contribution de l'azote atmosphérique et d'une énergie laser trop faible [59]. Les éléments pour lesquels les valeurs de concentration sont de l'ordre du mg/kg, voire moins parfois, à savoir Cd, Cu, Mn et Zn, n'ont pas pu être détectés dans nos conditions expérimentales pour les échantillons pastillés de plantes. Cependant, ces résultats ne doivent pas être considérés comme ceux de la méthode LIBS dans son ensemble mais uniquement comme ceux de l'instrument et des étapes de préparation mis en œuvres dans cette étude, puisque ces éléments sont couramment détectés par des instruments de laboratoire [59, 60, 61].

Une manière d'améliorer la quantité de signal collectée, en conservant le même système instrumental, serait d'optimiser le broyage des échantillons avant analyse. En effet, il a été démontré par une étude menée par Arantes de Carvalho et al. [73], qu'un broyage particulièrement fin d'un échantillon de feuille de boldo, avant pastillage, pouvait permettre d'obtenir des intensités LIBS significativement plus importantes. En effet, il est observé un signal LIBS de 1,6 à 1,9 fois plus important pour des raies élémentaires de Ca, K, P, Mg, B et Mn lorsque la taille de grain maximale passe de 150 à 50 μm . Ce phénomène est expliqué par une décomposition incomplète des particules larges présentes dans l'échantillon qui entraîne la présence d'un plus grand nombre d'atomes restant dans un état de non-émission dans le plasma, les intensités observées étant par conséquent moins élevées. Augmenter la finesse de broyage, qui dans notre étude est de 250 μm , permettrait potentiellement d'améliorer la sensibilité du

système à détecter des éléments traces ou d'augmenter la faible intensité de certaines raies élémentaires comme celles du phosphore.

Les autres façons d'optimiser la quantité de signal collectée nécessitent une modification des paramètres instrumentaux comme la fluence du laser, ce qui n'est pas possible avec le pistolet Z-300, ou une optimisation importante du système en appliquant un double-pulse laser [74] ou encore des nano-particules (NE-LIBS) à la surface de l'échantillon afin d'augmenter les intensités spectrales [75].

b) Analyse en composantes principales

Un modèle d'analyse en composantes principales appliquée au jeu de spectres de calibration a ensuite permis de retenir que : i) certains échantillons présentaient vraisemblablement des différences de qualité à cause d'un processus de broyage variable et/ou de leur nature intrinsèque, révélées par le biais de la composante 1 ; ii) les éléments présentant le plus de variabilité, à savoir Mg, Si, K, Na et Ca, permettent de différencier chimiquement les échantillons entre eux ; iii) certaines raies jugées anormales (le doublet de Ca à 393-396 nm et la raie de Ca I à 422,7 nm) à cause du phénomène d'auto-absorption sont à exclure en préalable à une approche quantitative. Le premier point reste cependant à nuancer et approfondir en incluant plus de diversité d'espèces et/ou plus de modalités de préparation d'échantillons de la même espèce dans le jeu de données par exemple. Ceci pourrait permettre de comprendre si les échantillons qui se regroupent selon la composante 1 se rassemblent par espèce à cause de leur nature (espèce ligneuse, fourragère, fruitière ou encore maraîchère) ou à cause d'un processus de broyage plus ou moins fin selon l'échantillon, ou les deux.

L'étude de Gomes et al. [76], consacrée à l'évaluation des méthodes de broyage pour l'analyse de pastilles de plantes diverses (feuilles de canne à sucre, d'oranger et de soja) par LIBS, montre que la méthode et les temps de broyage doivent être adaptés au type de plante afin d'obtenir le meilleur signal LIBS. Cette différence de traitement est rendue nécessaire par les propriétés physico-chimiques de la plante comme son contenu en fibres ou encore en cellulose, qui peut avoir un impact sur la porosité et la densité de la pastille. Cette étude rejoint les résultats obtenus par les travaux de Arantes de Carvalho [73] décrits précédemment, qui démontrent que les effets de matrice observés entre échantillons de plante sont minimisés, voire éliminés, pour des particules inférieures à 100 μm .

ii. Analyse quantitative LIBS a) Calibration

L'analyse quantitative de ces échantillons a été effectuée pour plusieurs éléments à l'aide d'une régression univariée à partir des concentrations mesurées par ICP-AES et après une étape indispensable de normalisation. Quatre méthodes de normalisation, à savoir la SNV, la normalisation par l'aire totale du spectre, la normalisation par étalon interne et la normalisation par le background ont été appliquées aux spectres LIBS des échantillons de plantes. Le choix de la meilleure normalisation, qui s'est révélé être dépendant de l'élément analysé, a été fait sur la base de plusieurs figures de mérite : le coefficient de détermination (r^2), les erreurs du modèle ainsi que la limite de détection (LOD). La construction d'un modèle d'étalonnage satisfaisant

($r^2 > 0,95$) a été possible pour Ca, Fe, K, Mg et Al bien que la gamme d'étalonnage pour Mg ait été réduite afin de conserver un modèle de régression exploitable. Au contraire, les modèles construits pour P et Na semblent moins convaincants. Le phosphore est caractérisé par une raie peu émissive et les échantillons analysés présentent des concentrations proches de la limite de détection. En ce qui concerne le sodium, son modèle d'étalonnage présente des valeurs extrêmes qui s'éloignent de la droite de régression. Le carbone n'a pas pu être quantifié à cause d'un manque de variabilité des concentrations mesurées. Il resterait cependant à comprendre la présence de certains *outliers* à travers potentiellement une limitation de l'efficacité de certaines normalisations ou des mesures de concentrations de référence à remettre en cause, même si ces dernières peuvent globalement être considérées comme fiables. L'analyse quantitative multivariée n'a également pas été pertinente, les éléments n'étant pas tous corrigés par la même méthode de normalisation et le nombre d'échantillons analysés étant relativement faible.

En règle générale, d'après un article de revue publié récemment par de Carvalho et al. [10], relatif à la construction de modèles d'étalonnage pour les plantes en LIBS, il est recommandé d'appliquer les deux stratégies univariée et multivariée, notamment lorsque l'on dispose de standards ou d'échantillons de référence certifiés avec des matrices similaires. En effet, il a été remarqué lors d'analyses LIBS d'échantillons de plante de même espèce, que les modèles d'étalonnage univariés et multivariés (PLS) présentaient des performances très similaires (LOD, erreur, r^2) [60, 77]. Au contraire, lorsque des espèces de plantes variées constituent le jeu de données, et présentent potentiellement des matrices différentes, l'approche multivariée (PLS) permet de compenser ces effets de matrice et d'obtenir de meilleures performances prédictives [73]. Dans le cadre de notre étude, les effets de matrice ont été corrigés à l'aide de la normalisation, procédé qui n'a été abordé qu'une fois et de manière succincte [77], à notre connaissance, pour l'analyse quantitative des échantillons de plante par LIBS.

b) Validation

Les modèles d'étalonnage construits ont finalement été validés à travers la prédiction en concentrations d'un nouveau jeu de données rassemblant des échantillons de plantes de différentes natures. Les concentrations en calcium, fer et magnésium ont été correctement prédites pour des gammes d'étalonnages jugées réalistes, c'est-à-dire représentatives de ce qu'on peut retrouver dans la nature. Le modèle du magnésium est cependant à nuancer car la gamme d'étalonnage de ce dernier a été réduite, ne couvrant pas toutes les concentrations observées dans la nature. Les modèles pour Na et P semblent également intéressants mais moins robustes, la majorité des échantillons de validation étant hors gamme pour le phosphore et la régression linéaire du sodium étant fragile ($r^2 < 0,9$). En revanche, les modèles d'étalonnage pour le potassium et l'aluminium ne sont pas considérés comme validés, les concentrations prédites étant significativement éloignées des concentrations mesurées. Cette différence est symbolisée par la valeur de la RMSEP, nettement supérieure aux valeurs de RMSEC et RMSECV des modèles associés, révélant des performances de quantification surestimées pour ces deux éléments. Cette non-validation peut être expliquée de deux façons différentes : i) pour l'aluminium, les valeurs d'incertitude observées sur les données ICP-AES sont très élevées, fragilisant la corrélation entre le signal LIBS et la concentration mesurée. De plus, la régression linéaire est fortement impactée par la présence des trois points les plus concentrés, faussant la linéarité entre les deux variables ; ii) le modèle du potassium n'est pas normalisé, indiquant

certainement l'aspect indispensable de la normalisation pour prédire la concentration d'espèces différentes. Les modèles normalisés pour K, dans notre contexte, n'étaient pas exploitables et il n'a pas été jugé pertinent de prédire les concentrations du jeu de validation.

Le récapitulatif des performances analytiques du Z-300 pour les échantillons de plantes est proposé par le tableau 2-3 suivant.

Elément	Ca	Fe	Mg	Na	P	K	Al	C	N, Cd, Cu, Mn, Zn
Déecté									
Quantifié									
Validé									
Min (g/kg)	0.4	0.1	0.6	0,05	0.7	0.4	0.1		
Max (g/kg)	28	1.0	2.5	1,0	5.4	34	1.9		
Méd (g/kg)	14	0.3	1.3	0,1	2.1	13	0.4		
Normalisation	SNV	E.I Ar	E.I Ar	E.I Ar	E.I Ar	Net	E.I Ar		
R ²	0.98	0.95	0.94	0.85	0.75	0.96	0.96		
RMSE (g/kg)	1.3	0.1	0.1	0.1	0.6	2.2	0.1		
RMSECV (g/kg)	1.3	0.1	0.1	0.1	0.6	2.2	0.1		
RMSEP (g/kg)	0.9	0.1	0.3	0.2		4.7	0.5		
LOD (g/kg)	0.6	0.1	0.1	0.3	0.7	0.4	0.1		

Tableau 2-3 : Tableau récapitulatif des performances analytiques obtenues pour les échantillons de plantes avec l'instrument LIBS portable Z-300.

L'analyse quantitative de pastilles de plantes avec un instrument LIBS portable, à notre connaissance, n'avait encore jamais réalisée auparavant [12], rendant difficile l'évaluation des performances analytiques obtenues avec ce type d'instrument. Cependant, les analyses d'échantillons de végétaux pastillés par des instruments LIBS de laboratoire sont réalisées depuis une dizaine d'années et sont donc bien documentées, ce qui nous permet d'avoir une base de comparaison. En effet, il a déjà été estimé que les méthodes LIBS et ICP-AES étaient comparables dans le but de diagnostiquer le statut des plantes, notamment à travers des limites de détection similaires entre les deux méthodes pour plusieurs éléments d'intérêt (P, K, Ca, Mg, Fe, Cu, Mn, Zn, Si et B) [31, 78]. Les différences entre les deux méthodes résident dans le fait que la LIBS permet de gagner en temps d'analyse et de préparation alors que l'ICP-AES permet de gagner en précision. Les valeurs de limite de détection obtenues dans le cadre de notre étude sont plus élevées que celles affichées par une méthode LIBS de laboratoire pour les éléments Ca, Fe et Mg, qui, pour ces éléments, affichent des valeurs de 0,01 ; 0,0004 et 0,02 g/kg respectivement [31]. Les instruments LIBS de laboratoire impliqués pour de tels résultats sont plus performants en termes d'énergie laser, de résolution spectrale et les raies élémentaires choisies sont parfois différentes de celles utilisées dans cette étude. Par exemple, la LOD théorique pour Ca a été obtenue *via* la raie de Ca I à 422 nm alors qu'elle est considérée comme auto-absorbée sur nos spectres LIBS. La LOD calculée sur nos données pour le potassium est quant à elle plus faible que celle observée sur certains instruments LIBS de laboratoire (2 g/kg [31], même si cette dernière a été mesurée sur une raie de K moins émissive (404 nm). Nous estimons que nos modèles sont relativement pertinents dans la prédiction de la concentration totale en éléments dans la plante pour les éléments d'intérêt quantifiés même s'ils n'atteignent pas les performances et la justesse d'un instrument LIBS de laboratoire ou ICP-AES.

iii. L'apport de la XRF

L'ensemble des modèles d'étalonnage construits à partir du signal LIBS pour cette étude ont pour référence les concentrations mesurées par ICP-AES car, comparé à des modèles construits en prenant les mesures XRF comme référence (non montrés ici), les modèles basés sur les données ICP étaient bien plus performants. Cependant, la comparaison des mesures LIBS et des mesures XRF a permis de montrer, selon les cas, la cohérence et/ou la complémentarité de ces deux méthodes de mesures rapides avec des instruments portables :

a) Détection des *outliers* :

En règle générale, la présence d'un *outlier* peut provenir d'une erreur pendant le processus de traitement des données avec une inversion de valeur ou encore un prétraitement non appliqué par exemple. Si aucune erreur de traitement ne semble être décelée, l'écartement de ce point par rapport aux autres peut avoir une nature analytique avec des conditions de mesure différentes lors de l'analyse de ces points. Si aucun paramètre instrumental ne peut être remis en cause, cela peut vouloir indiquer que l'échantillon jugé *outlier* est tout simplement différent des autres par sa composition physique ou chimique. C'est ce qui semble être le cas pour l'exemple de l'échantillon de feuille de palmier, détecté comme *outlier* pour le modèle d'étalonnage de Ca et de K notamment. En effet, aucune anomalie n'étant observée pendant l'analyse ou le traitement des données, il a été déduit que l'échantillon 165 présentait certainement une nature trop différente des autres échantillons. En effet, si on observe les modèles d'étalonnages des concentrations prédites par XRF en fonction des concentrations mesurées par ICP-AES, pour Ca et K, cet échantillon s'éloigne également très largement de la droite de régression. Ces deux modèles sont présentés par la figure 2-25.

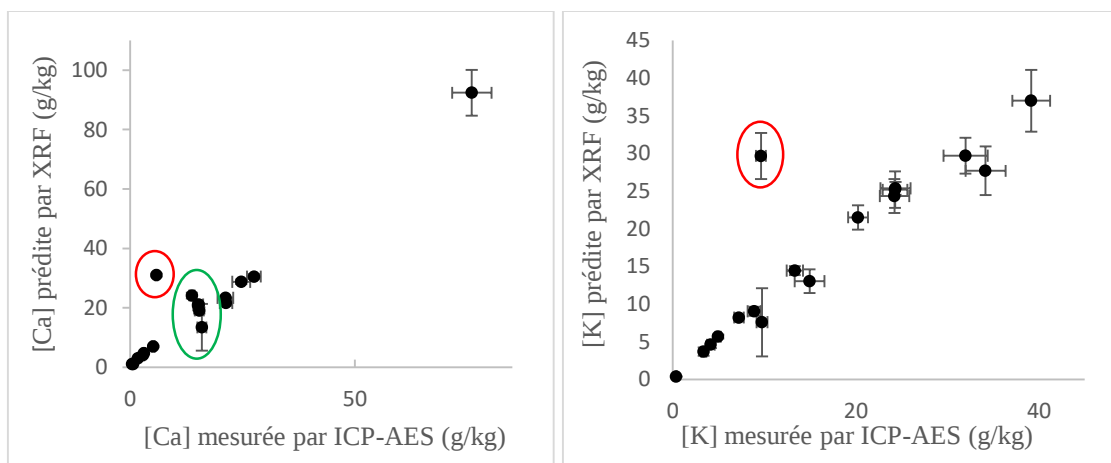


Figure 2-25 : Courbe d'étalonnage obtenue sur les mesures XRF en fonction des concentrations mesurées par ICP-AES dans les plantes pour a) le calcium et b) le potassium. Les cercles rouges représentent l'échantillon outlier 165. L'ellipse verte indique la présence d'un gradient en concentrations.

On peut observer, sur la figure 2-25, un éloignement de l'échantillon de feuille de palmier pour K et Ca aussi marqué que pour le modèle d'étalonnage obtenu en LIBS (cf figure 2-14 et 2-17). L'association des deux méthodes (LIBS et XRF) confirme bien la différence de cet échantillon par rapport aux autres. Egalement, on peut observer, via l'ellipse verte sur la courbe d'étalonnage du calcium, le même profil qu'en LIBS avec un gradient de prédictions pour une même concentration mesurée. Ceci souligne l'importance de pouvoir traiter directement les spectres afin d'améliorer l'étalonnage plutôt que de se fier à une calibration interne déjà établie. Ce dernier point ne remet pas en cause la méthode XRF mais plutôt la façon dont a été construit le modèle de calibration interne à l'instrument.

b) Complémentarité des deux méthodes :

Le fait d'analyser les échantillons de plantes par deux instruments portables très semblables mais impliquant des sources d'excitation distinctes a permis de mettre en lumière la complémentarité des deux méthodes. En effet, selon l'élément observé, trois cas de figure ont pu être observés :

- 1) Les deux méthodes permettent d'obtenir des modèles d'étalonnage similaires. C'est le cas du calcium, du potassium et du fer. La droite de calibration obtenue pour le fer peut être observée en annexe avec la figure 2-6. La comparaison des deux méthodes s'appuie essentiellement sur l'allure de la courbe d'étalonnage obtenue car les données acquises en XRF sont déjà des concentrations prédites d'une calibration interne pré-existante. L'élément N n'est pas détecté par les deux méthodes et la XRF ne permet pas d'obtenir un modèle exploitable pour Cu malgré sa détection.
- 2) La méthode XRF semble offrir des modèles d'étalonnages nettement plus performants qu'en LIBS. Cela concerne le phosphore et également le manganèse et le zinc qui n'ont pas été détectés en LIBS. Le modèle XRF du phosphore est présenté en figure 2-7 en annexe.
- 3) La méthode LIBS semble offrir des modèles d'étalonnages nettement plus performants qu'en XRF. Cela concerne l'aluminium et également le sodium qui n'est pas détecté à l'aide de la XRF portable. Le modèle XRF de l'aluminium est présenté en figure 2-8 en annexe.

La campagne de mesures XRF s'est donc révélée riche en informations, mettant en lumière une certaine complémentarité entre les deux méthodes portables, l'importance de pouvoir corriger les spectres et en confirmant l'existence de certains *outliers* dans le jeu de données. La comparaison des performances des deux instruments reste à nuancer car les résultats obtenus en XRF sont issus de prédictions de concentrations d'une calibration interne fournie par le constructeur alors que la démarche en LIBS est justement de construire cette calibration. Les résultats obtenus sur ces deux instruments ne sont donc pas totalement comparables.

Même si la XRF est implantée comme une méthode fiable pour l'analyse de pastilles de plante depuis une dizaine d'années, aussi bien en laboratoire que de manière portable, l'association de la XRF et de la LIBS, à travers ces résultats complémentaires et ceux disponibles dans des

travaux antérieurs [10], représente donc un fort potentiel analytique en ce qui concerne l'analyse des plantes en conditions d'analyse contrôlées et optimales.

La notion d'« universalité » des modèles d'étalonnage pouvant prédire la concentration en éléments dans un grand nombre d'espèces végétales reste encore à approfondir, le nombre d'échantillons analysés se révélant probablement assez faible pour répondre de manière satisfaisante à un tel objectif. Les modèles d'étalonnage doivent donc être complétés par d'autres espèces afin d'évaluer les limites de la normalisation et de leurs performances prédictives. Même si des modèles quantitatifs de prédiction n'ont pu être validés que pour les éléments Ca, Fe et Mg, ces résultats obtenus pour la première fois avec un appareil portable suggèrent des perspectives intéressantes pour l'utilisation de la normalisation en LIBS.

Nous allons voir dans le chapitre suivant, si des stratégies similaires peuvent s'appliquer à des matrices agricoles réputées plus complexes telles que les sols agricoles par exemple.

CHAPITRE III

LIBS portable et sols agricoles : des modèles représentatifs des sols agricoles français ?

La caractérisation par LIBS de sols représentatifs des sols de grandes cultures en France passe par une sélection pertinente d'échantillons. Le corpus d'échantillons (ici 62 sols) analysés doit être représentatif des sols de grandes cultures en France, aussi bien sur le plan de la texture, que sur celui des gammes de concentrations totales en élément. La texture, qui conditionne de nombreuses autres propriétés physiques et chimiques des sols, correspond à l'abondance relative dans la terre fine (constituants du sol < 2 mm) de trois grandes fractions classées selon leur granulométrie : i) les sables pour les particules ayant un diamètre compris entre 50 μm et 2 mm, ii) les limons pour les particules présentant un diamètre compris entre 50 et 2 μm et iii) les argiles pour les particules de taille < 2 μm . La texture dominante de l'horizon supérieur des sols agricoles français (figure 3-1 en annexe) est majoritairement de type équilibré, c'est-à-dire définie par une proportion similaire entre le pourcentage de sable, de limon et d'argile. La présence de sols de nature principalement argileuse (avec un pourcentage d'argile supérieur à 40%) ou limoneuse (avec un pourcentage de limon supérieur à 50%) sont observées largement dans le nord-est et nord de la France [79]. Il existe également dans une moindre mesure des sols sableux, composés à plus de 50% de sable, principalement dans les Landes. Les sols sélectionnés pour cette étude sont représentatifs des sols agricoles français en terme de texture (Figure 3-1) : ils sont principalement de textures limono-argileuse et argileuse, également de texture équilibrée, et très peu sont de texture sableuse.

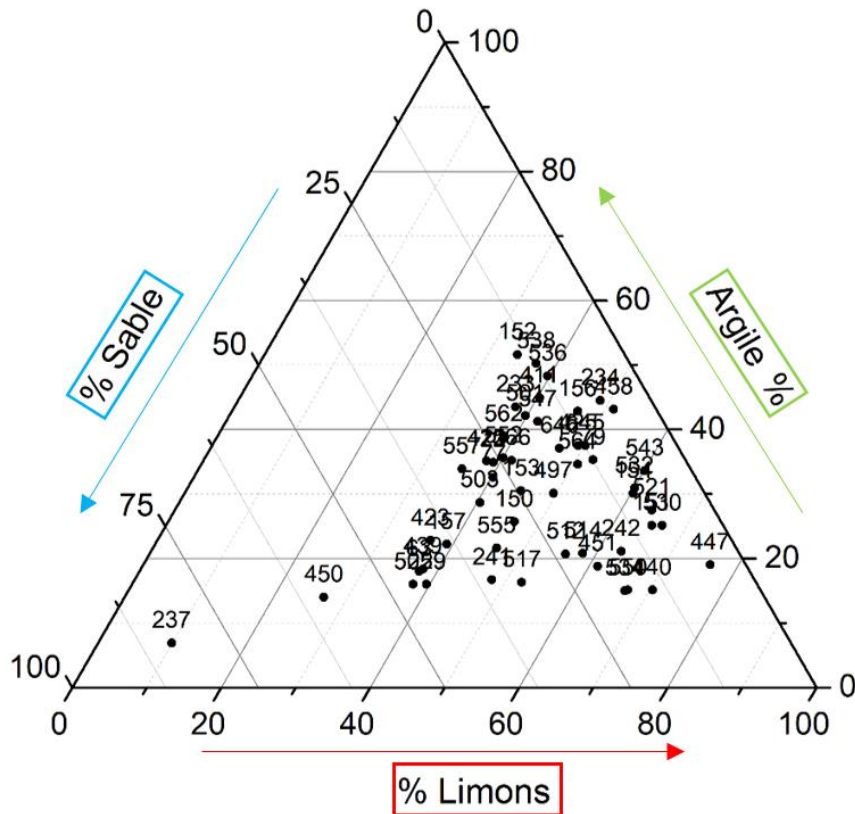


Figure 3-1 : Diagramme ternaire des textures pour les sols de cette étude issu du réseau QUASAPROVE.

En ce qui concerne les concentrations totales en éléments mesurées par ICP-AES dans les sols sélectionnés, on peut remarquer qu'elles sont dans le même ordre de grandeur que celles observées dans les sols français (tableau 3-1). Les gammes théoriques ont été déterminées à l'aide des cartes conçues par le Réseau de Mesures de la Qualité des Sols (RMQS) qui est un outil d'évaluation de la qualité des sols français sur le long terme. Ce réseau est composé de 2240 sites répartis uniformément sur le territoire français selon une maille carrée de 16 km de côté. Le sol au centre de chaque maille est prélevé, analysé et observé selon plusieurs protocoles [80] une fois tous les 15 ans. Le Groupement d'intérêt scientifique sol (Gis Sol) met ensuite à disposition des cartes représentant les valeurs mesurées de plusieurs paramètres physiques et chimiques selon la localisation du prélèvement. Les paramètres regroupent aussi bien des valeurs de pH, que de teneur en matière organique ou encore de concentration totale de plusieurs éléments chimiques. Les gammes théoriques indiquées dans le tableau suivant sont donc particulièrement représentatives du territoire français. On peut donc noter que les concentrations des sols sélectionnés couvrent globalement les gammes théoriques observées en France. Cependant, pour plusieurs éléments (N, Al, Ca, Fe, K et Mg), les sols sélectionnés ne permettent pas d'atteindre la concentration haute de la gamme théorique. Les concentrations hautes des gammes théoriques représentent en réalité les valeurs maximales observées pour chaque élément et non des valeurs moyennes. Ces dernières sont donc certainement plus faibles et donc proches des concentrations les plus élevées des sols sélectionnés.

Le jeu d'échantillons sélectionné ici représente donc bien la diversité des textures et des concentrations élémentaires que l'on peut observer au sein des sols français de grandes cultures agricoles.

	Minimale	Médiane	Maximale	Gamme théorique
C (g/kg)	7.7	16	41	12 - 45
N (g/kg)	0.6	1.4	4.1	1 - 16
Al (g/kg)	12	45	80	20 - 100
Ca (g/kg)	1.8	28	302	0.2 - 340
Cd (g/kg)	0.0001	0.0003	0.001	< 0.005
Cu (g/kg)	0.005	0.02	0.1	0.001 – 0.1
Fe (g/kg)	6.6	25	97	< 25 - 130
K (g/kg)	3.6	14	27	< 10 - 54
Mg (g/kg)	1.3	4.7	18	0.2 - 50
Mn (g/kg)	0.3	0.6	2.3	0.02 - 0.15
Na (g/kg)	0.6	3.5	16	0.2 - 30
P (g/kg)	0.9	2.2	8.0	0.2 – 0.9
Zn (g/kg)	0.03	0.07	0.2	0.03 – 0.26

Tableau 3-1 : Récapitulatif des concentrations mesurées par ICP-AES dans les échantillons de sols issus des réseaux internes de l'INRAE et des concentrations « théoriques » observées en milieu naturel [79].

1) L'analyse qualitative des sols

i. Identification spectrale

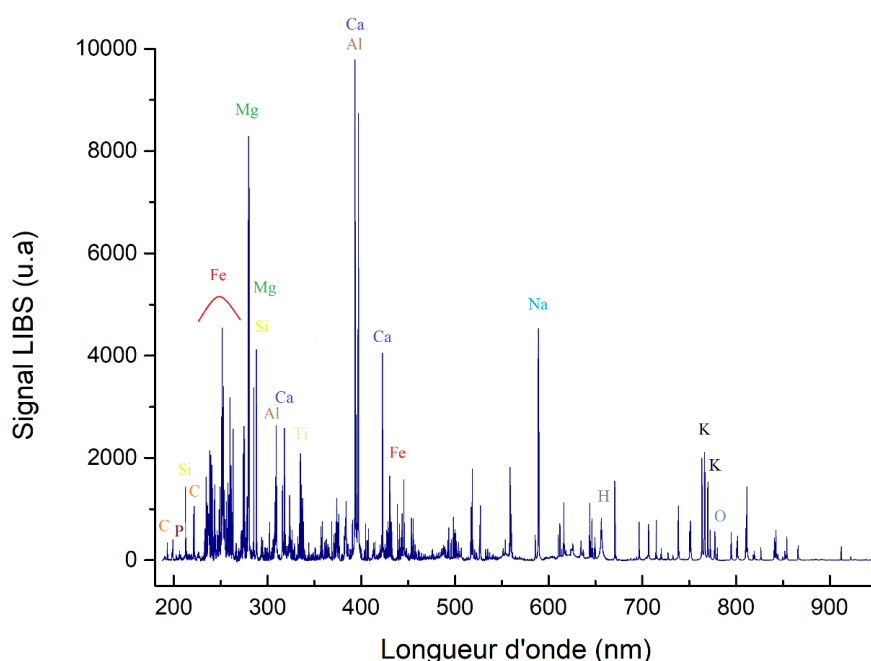


Figure 3-2 : Spectre LIBS typique d'un sol (texture équilibrée) du réseau QUASAPROVE, avec l'attribution des éléments d'intérêt pour les raies associées.

L'attribution des éléments pour chacune des raies élémentaires, à l'instar des spectres LIBS de plantes, a été réalisée à l'aide des différentes bases de données théoriques disponibles en ligne (NIST, AtomTrace, Kurucz) et de la connaissance du matériau analysé. La liste complète des éléments détectés et leurs degrés d'ionisation est fournie dans le tableau 3-1 en annexe. A partir

de la figure 3-2, on peut tirer sensiblement les mêmes enseignements que pour les échantillons de plante discutés dans le chapitre précédent, à savoir : i) la détection d'éléments d'intérêt concentrés de l'ordre du g/kg comme Al, C, Ca, Fe, K, Mg, Na, P et Si et ii) l'absence de signal pour les éléments dits « traces », concentrés de l'ordre du mg/kg tels que Cd, Cu, Mn et Zn. Il est intéressant également de noter la détection d'un triplet du titane autour de 334 nm et le fait que l'azote ne soit toujours pas détecté (l'élément étant moins concentré dans les sols que dans les plantes). Le spectre typique d'un sol présente cependant beaucoup plus de raies d'émission que celui d'un échantillon de plante, notamment entre 233 et 273 nm ainsi qu'autour de 430 nm. Ces nombreuses raies d'émission supplémentaires proviennent du fer dont la concentration moyenne est 100 fois supérieure dans les sols (30 g/kg) que dans les végétaux (0.3 g/kg).

ii. Analyse en composantes principales

Le principe de l'analyse en composantes principales (ACP) ainsi que l'algorithme NIPALS et les figures de mérite associées sont décrits en détail dans le chapitre II de ce mémoire. Il est simplement rappelé ici que l'ACP est d'abord calculée en prenant en compte les spectres LIBS moyens entiers de chaque échantillon, ici les pastilles de sols agricoles. Les spectres moyens sont centrés avant de lancer le calcul de l'ACP. La matrice des données utilisée pour le calcul ACP est ici composée de 5843 variables/longueurs d'onde et de 51 observations/spectres moyens.

a) Décalage en longueur d'onde

L'ACP calculée sur les spectres entiers a essentiellement permis de révéler la présence d'une anomalie sur la gamme visible et infrarouge du spectre. En revanche, la partie UV ne semble pas du tout impactée. Plus précisément, ce sont les loadings d'une des composantes qui permettent de mettre en évidence une certaine dérive spectrale (figure 3-3).

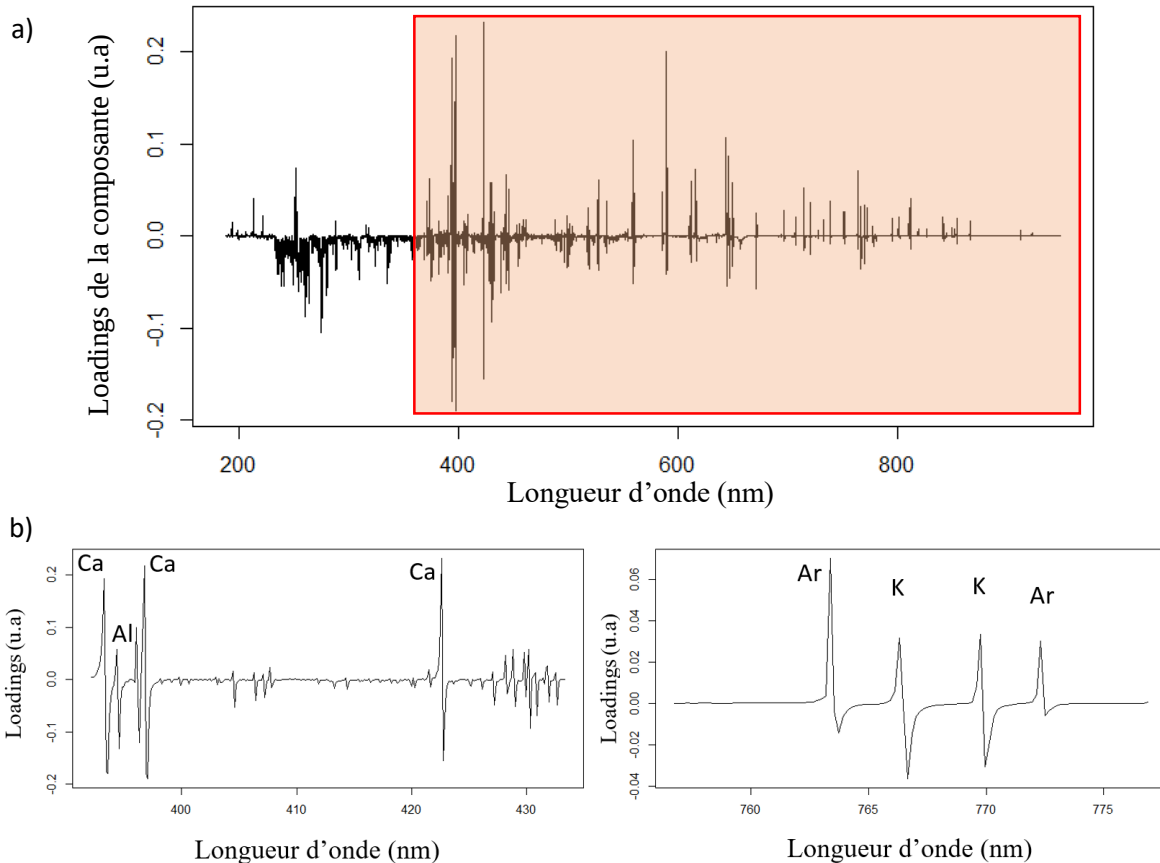


Figure 3-3 : Graphe des loadings a) pour la totalité des variables et b) pour deux sous-ensembles de variables dans les gammes spectrales 380- 440 nm et 750- 780 nm de la composante 3 révélant le décalage en longueur d'onde. Le rectangle en transparence orange indique la zone spectrale impactée par le décalage.

La figure 3-3 a. permet d'observer les valeurs des loadings de la composante 3 sur l'ensemble de la gamme de longueurs d'onde du pistolet Z-300 et la figure 3-3 b. offre un zoom de ces valeurs sur des plages spectrales du visible et de l'infrarouge présentant des allures de loadings jugées anormales. En effet, pour chacune des raies élémentaires du visible et de l'infrarouge, on observe une allure des loadings en forme de fonction dérivée d'un pic alors qu'en l'absence de dérive spectrale, l'allure de la courbe des loadings reproduit assez bien la forme des raies d'émission, en valeur positive ou négative, selon les cas. L'allure assimilée à une fonction dérivée se répète pour chaque raie et pour chaque élément et ne peut donc pas être associée à un élément en particulier. De plus, les profils en fonction dérivée ne sont observés que sur la gamme spectrale allant de 350 à 950 nm mais pas sur la gamme allant de 188 à 350 nm. Leur origine n'est donc pas de nature chimique mais le résultat de l'étalonnage interne de l'appareil. En effet, la campagne de mesures complète ayant été menée sur plusieurs jours afin d'analyser tous les sols, le pistolet Z-300 appliquait régulièrement une auto-calibration en longueur d'onde entraînant ainsi certains décalages en longueur d'onde. Le fait que le spectromètre inclus dans le système portable soit en réalité la juxtaposition de trois spectromètres distincts, couvrant les gammes UV pour le premier, visible pour le deuxième et IR pour le dernier, on comprend aisément que le décalage spectral constaté lors de la campagne de mesures n'affecte pas de la même façon ces trois spectromètres. Deux extraits de spectres LIBS moyens obtenus lors de deux jours différents de la campagne de mesure sont montrés sur la figure 3-4. Ils illustrent

clairement le décalage en longueur d'onde de l'instrument LIBS portable. Il est cependant tout à fait possible, bien que fastidieux, de recalcr en longueur d'onde chaque spectre LIBS [51] mais nous avons choisi ici une solution alternative.

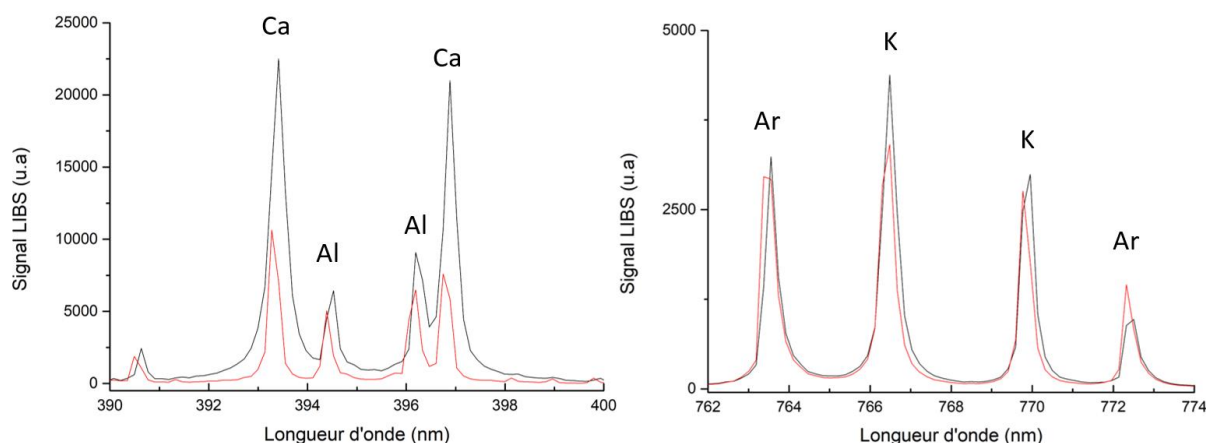


Figure 3-4 : Extraits de spectres LIBS de deux échantillons de sols illustrant la dérive en longueur d'onde du pistolet Z300.

En remarquant que la grande majorité des éléments (C, P, Mg, Ca, Si, Al, Fe) présentent des raies d'émission dans la partie UV (188-350 nm) non impactée par le décalage spectral, à l'exception de Na et K, nous avons décidé de calculer un nouveau modèle ACP en ne prenant en compte que les variables se trouvant dans cette gamme spectrale. Notons que ce modèle ACP est une version simplifiée de celui que l'on pourrait calculer à partir de la totalité des variables, une fois les spectres recalés en longueur d'onde.

b) ACP et information chimique

Le nouveau modèle ACP, calculé en ne prenant en compte que la partie UV (188-350 nm) des spectres LIBS des échantillons de sol, est caractérisé par une variance majoritairement expliquée (97 %) à l'aide de 4 composantes. De plus, comme le montre la figure 3-5, représentant l'évolution de la variance expliquée en fonction du nombre de composantes, la première composante explique déjà environ 62 % de la variance. Ensuite, les composantes 2 et 3 permettent individuellement d'expliquer 14 % de la variance restante pour au final atteindre un plateau à 97 % après l'ajout de la quatrième composante.

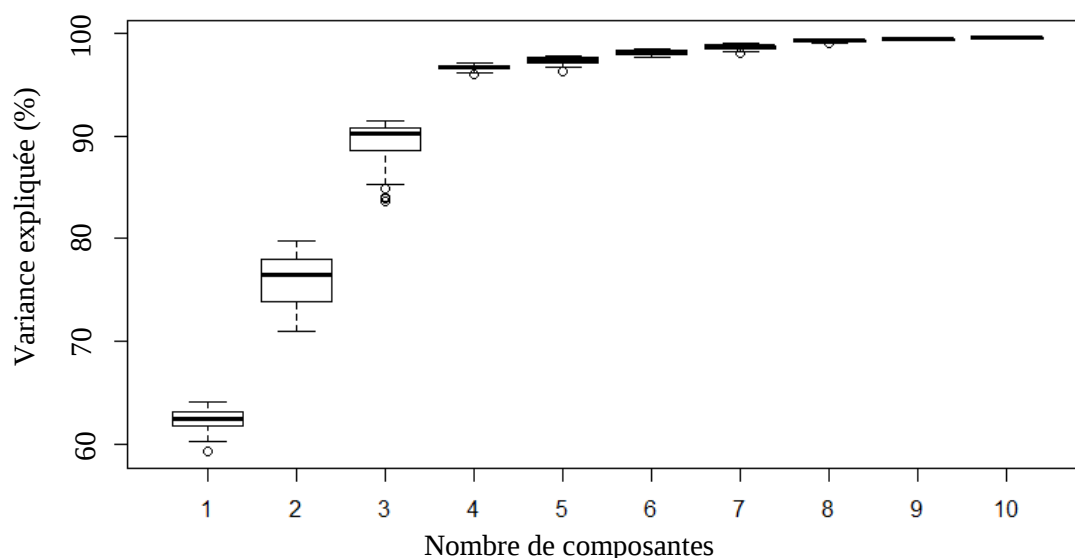


Figure 3-5 : Graphe du pourcentage de variance expliquée en fonction du nombre de composantes pour le modèle ACP calculé sur la gamme 188-350 nm.

Les 4 premières composantes permettent d'obtenir deux types d'information sur les données :
 i) une information chimique via les composantes 1, 3 et 4 et qui sera décrite dans ce paragraphe et ii) une information sur la nature physique des sols, via la deuxième composante qui sera décrite dans le paragraphe suivant. Lorsque l'on observe les graphes des loadings pour les composantes 1, 3 et 4 (figure 3-6), on remarque que la première composante est associée aux variations de signal du calcium, la troisième à celles du fer et la quatrième à celles du silicium et du magnésium. Plus précisément, la composante 4 présente une anti-corrélation entre les variations de signal du silicium et celles du magnésium.

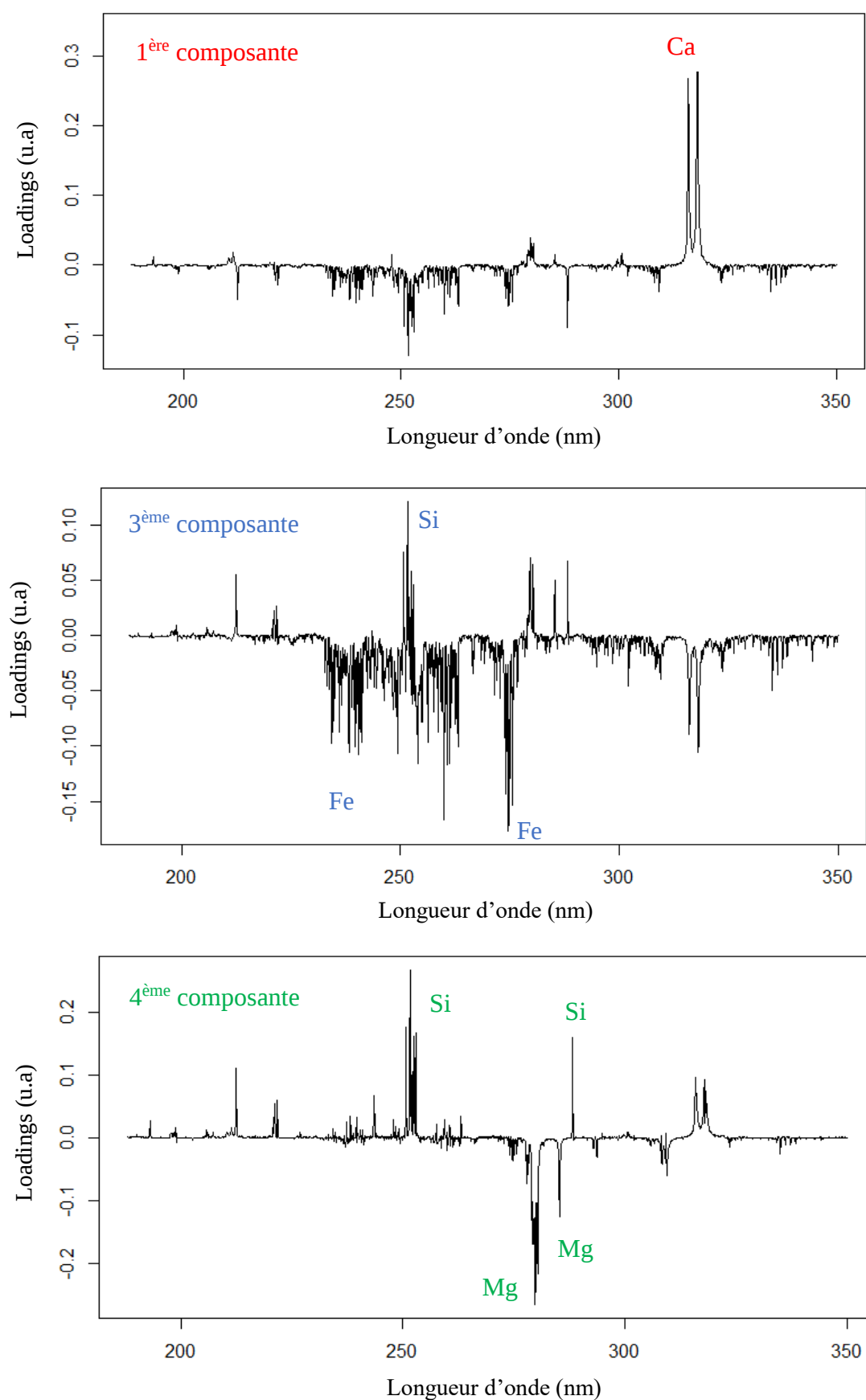


Figure 3-6 : Graphe des loadings de la 1^{ère}, 3^{ème} et 4^{ème} composante du modèle ACP calculé sur la gamme 188-350 nm.

Il est donc possible, à l'aide des composantes 1 et 3 et du graphe des scores, de proposer une représentation où les trois quarts (75 %) de la variance est expliquée. Le graphe des scores des composantes 1-3 (figure 3-7a) est comparé au diagramme ternaire construit à partir des concentrations mesurées par ICP-AES (Figure 3-7b). Le diagramme ternaire permet de différencier rapidement les sols entre eux selon trois pôles qui sont : le pôle calcaire (Ca+Mg), le pôle Aluminium (Al) et le pôle métaux (Fe). La concentration des autres éléments métalliques présentant des concentrations très faibles, voire négligeables, par rapport à la concentration en fer, ceux-ci n'ont pas été inclus dans le calcul du pôle métaux. Le silicium, à l'instar de l'aluminium, est un élément représentatif des sols argileux mais sa concentration dans les sols n'a pas été mesurée par ICP-AES. Le diagramme ternaire est un mode de représentation couramment utilisé pour décrire la répartition des propriétés granulométrique ou physico-chimique des sols [42].

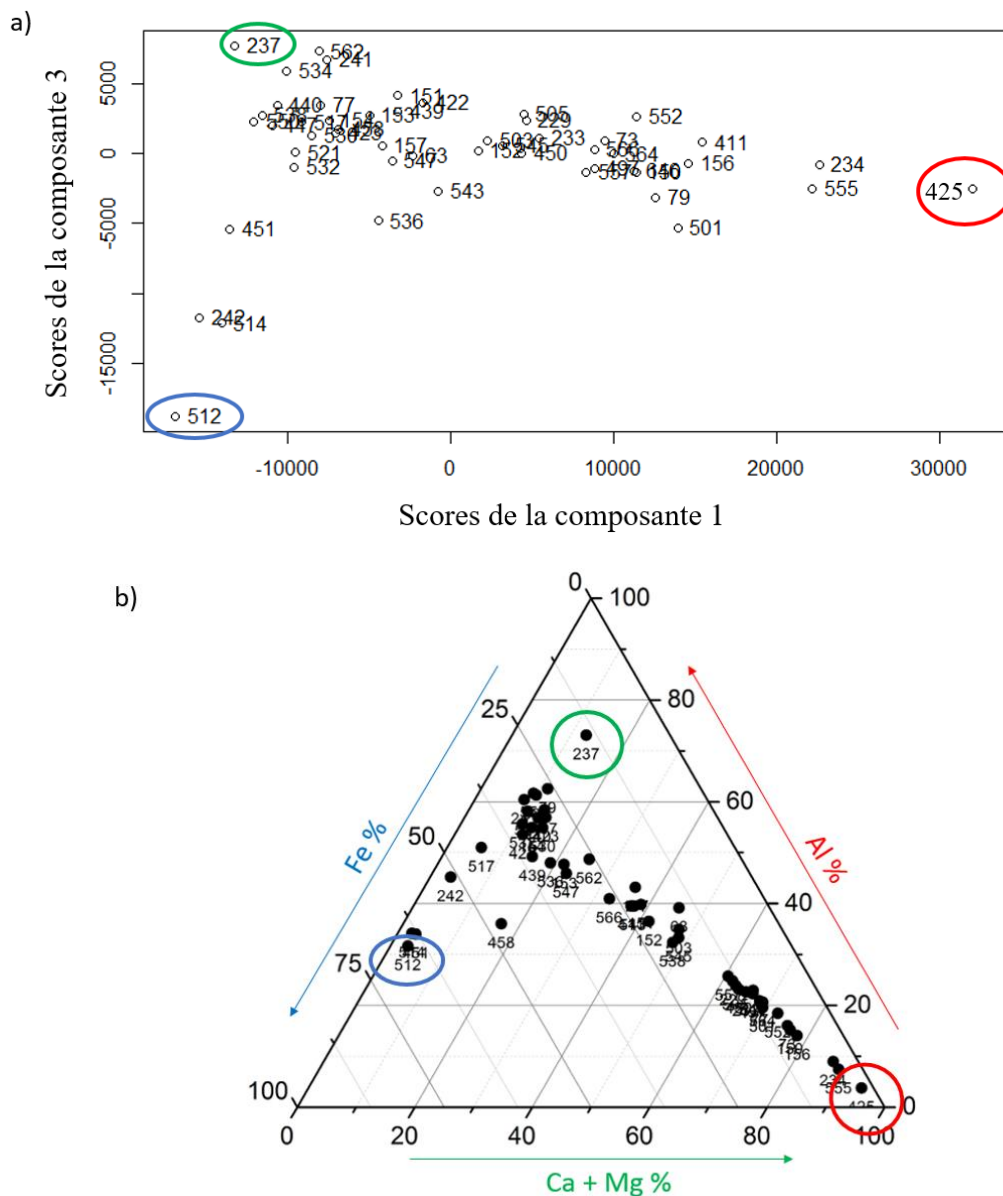


Figure 3-7 : a) Graphe des scores des composantes 1-3 du modèle ACP calculé à partir des spectres LIBS des pastilles de sols et b) diagramme ternaire de la composition chimique des sols construit à partir des concentrations mesurées par ICP-AES.

L'étalement des scores (spectres des échantillons) selon la première composante est principalement dû à la variabilité du doublet du calcium à 315,9 -317,9 nm. Il traduit le fait que le spectre moyen de l'échantillon 425, entouré en rouge sur la figure 3-7a, est caractérisé par une intensité maximale pour ce doublet, indiquant que cet échantillon est le plus concentré en calcium. Ce résultat est confirmé par le diagramme ternaire (figure 3-7b) sur lequel l'échantillon 425 constitue le pôle calcaire avec une concentration en calcium mesurée à 302 g/kg par ICP-AES.

Le même constat peut être dressé en observant simultanément la variabilité spectrale selon la composante 3 et la différence de concentrations pour le pôle métaux. Cette composante et ce pôle sont représentatifs de l'élément fer et on remarque bien, pour la LIBS et l'ICP-AES, que les échantillons aux extrêmes sont les mêmes. Plus précisément, on constate bien que l'échantillon 237 (entouré en vert sur les figures 3-7 a et b) est le moins intense en Fer en LIBS et également le moins concentré en Fer en ICP-AES. C'est le cas également pour l'échantillon 512 (entouré en bleu sur les figures 3-7 a et b) sauf qu'il est cette fois plus intense et plus concentré en fer. On peut donc conclure que le modèle ACP, calculé sur les spectres LIBS (188-350 nm) des pastilles de sols, permet de mettre en évidence une dispersion des données similaire à celle obtenue en prenant en compte les valeurs des concentrations mesurées par la méthode de référence, à savoir la méthode ICP-AES. Ce résultat démontre que, pour les sols agricoles, les deux méthodes de spectroscopie sont en excellent accord sur le plan de la description du jeu de données. Il est également possible de représenter les données selon les composantes 1 et 4 ou 3 et 4. Les graphes de scores associés sont présentés en annexe en figure 3-2.

c) ACP et information physique

La deuxième composante du modèle ACP calculé sur les spectres (188-350 nm) de sols permet d'expliquer 14 % de la variance des données. L'information associée à cette composante n'est pourtant pas de nature chimique. En effet, comme on peut l'observer sur la figure 3-8 a., présentant les valeurs de loadings de la composante 2, l'intégralité des raies élémentaires (pour C, Si, Ca, Mg, Fe, Al) du spectre forme une courbe entièrement négative. En d'autres termes, cela signifie qu'en passant d'un échantillon à l'autre toutes les raies augmentent ou diminuent de la même façon, peu importe la nature de l'élément, ce qui n'a aucun sens sur le plan de la variabilité de composition chimique entre les échantillons. La raison de l'obtention d'un tel graphe des loadings provient donc d'un autre phénomène, qui est ici de nature physique. En effet, comme précisé dans le paragraphe précédent, les sols agricoles de cette étude présentent des textures différentes. Le fonctionnement de la LIBS étant basé sur l'interaction du laser et de la matière analysée, l'ablation de cette dernière peut être plus ou moins efficace entraînant donc une quantité de lumière collectée variable selon la nature du matériau, c'est ce qu'on appelle plus communément les « effets de matrice » [36]. La lumière collectée étant différente d'un échantillon à l'autre, l'intensité de l'ensemble des raies détectées sera plus ou moins importante selon la nature du matériau. C'est le cas ici pour les sols, expliquant donc l'allure des loadings de la composante 2.

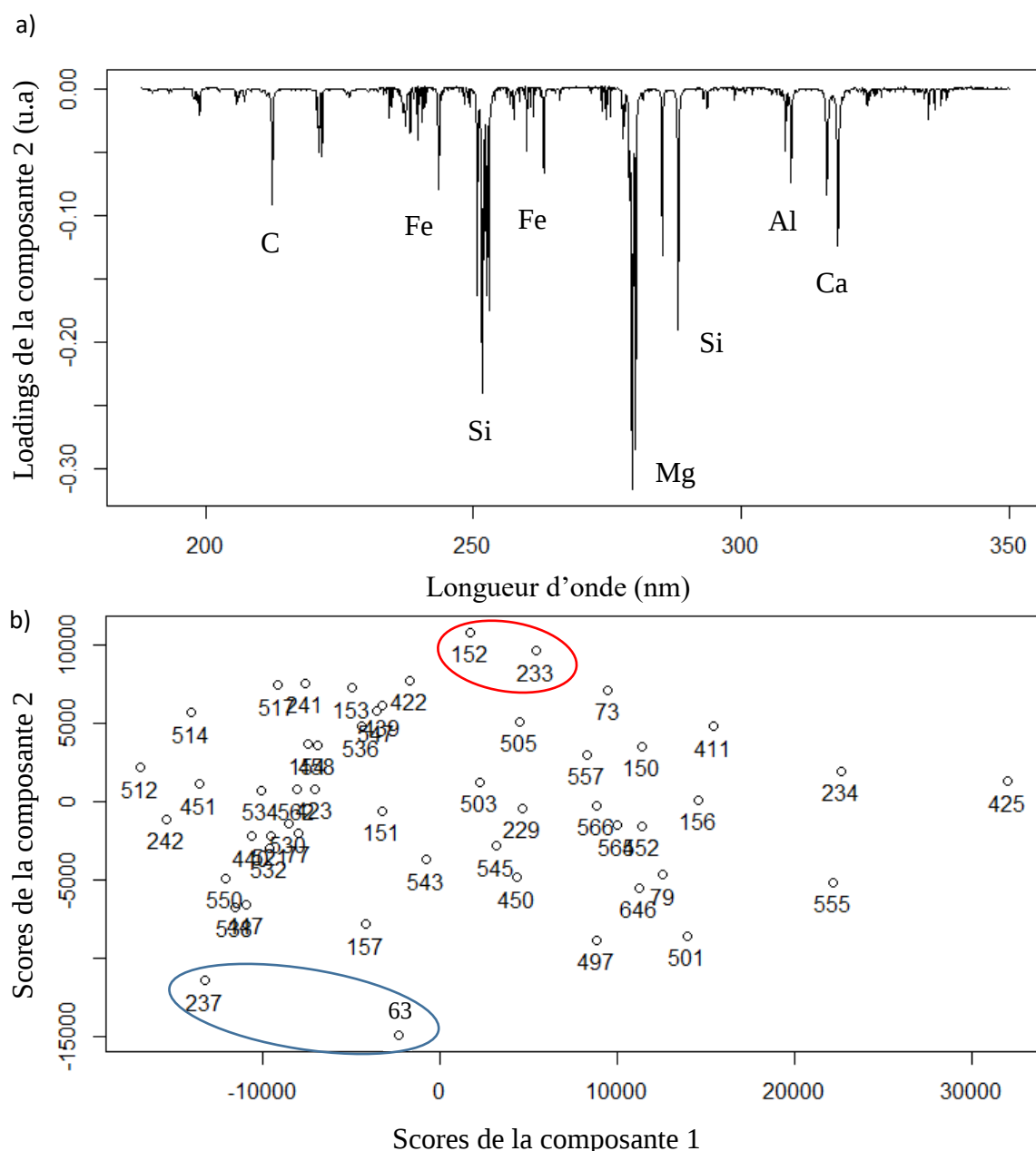


Figure 3-8 : a) Graphe des loadings de la composante 2 et b) graphe des scores des composantes 1-2 du modèle ACP calculé sur les spectres (188-350 nm) de pastilles de sols.

Cette association entre effets de matrice et résultats de l'ACP est renforcée par la figure 3-8 b.. En effet, d'après le graphe des scores en figure 3-8b., si on prend en compte les échantillons les plus éloignés selon la composante 2, on remarque qu'ils présentent des textures bien distinctes. Plus précisément, les échantillons 63 et 237 (entourés en bleu) sont essentiellement composés de sable (44 et 83 % respectivement) alors que les échantillons 152 et 233 (entourés en rouge) sont essentiellement argileux (52 et 44 % d'argile respectivement). Ces cas extrêmes illustrent le fait qu'il ne faut pas négliger l'influence de la texture dans les analyses LIBS. Cependant, la répartition des autres points ne traduit pas de corrélation évidente entre l'intensité du spectre LIBS et la texture du sol.

Le résultat n'est pas nouveau en LIBS, notamment avec l'existence de plusieurs articles associant l'intensité du signal collectée avec la texture du sol [36, 81, 82]. En effet en 2016, Villas-Boas et al. [81] ont proposé la construction de modèles multivariés, et plus précisément par régression aux moindres carrés partiels (PLRS), afin de pouvoir prédire le pourcentage de sable, de limons et d'argiles dans une soixantaine de sols agricoles brésiliens. Les modèles obtenus par PLS en prenant en compte l'ensemble des raies du spectre ou en sélectionnant certaines raies spécifiques permettent d'obtenir des coefficients de corrélation d'environ 0,89 et 0,90 respectivement entre les proportions prédites et mesurées. Les auteurs ont également montré une certaine corrélation (Na et Si) ou anti-corrélation (Fe) entre l'intensité de certains éléments et le pourcentage en sable contenu dans le sol. Gouegel et al., en 2019 [82] et dans une même mesure, a présenté une comparaison entre des modèles PLS et réalisés par un algorithme d'« elastic net » dans le but de prédire le pourcentage de sable, argile et limons de plusieurs sols du Québec. Il y a notamment été démontré que les intensités LIBS étaient globalement plus importantes pour des échantillons argileux et plus faibles pour les échantillons sableux, les sols limoneux présentant des intensités intermédiaires. Il est important de préciser, par leur diagramme ternaire proposé en annexe en figure 3-3, que leurs échantillons de sols présentent des textures bien plus diverses et extrêmes que les nôtres facilitant certainement les corrélations. Les sols de notre étude permettent tout de même d'observer une différence de signal selon leur texture mais pas d'aboutir à des résultats aussi clairs.

En conclusion, l'ACP sur les spectres LIBS s'est révélée être riche en informations : i) liées à des anomalies avec le décalage en longueur d'onde ; ii) chimiques avec une dispersion des échantillons en accord avec les mesures de référence ICP-AES ; iii) physiques, en révélant une ablation plus ou moins efficace liée à la texture du sol. Il reste à voir maintenant si ces particularités ont un impact sur la possibilité de construire des modèles quantitatifs.

2) L'analyse quantitative des sols agricoles

La construction et la validation des modèles d'étalonnage pour les sols est la même que pour les plantes, à savoir :

- Etablir la corrélation entre le signal LIBS de l'analyte et la mesure de sa concentration à l'aide la méthode de référence (ICP-AES) ;
- Evaluer quels sont les traitements de données permettant d'obtenir les modèles d'étalonnage les plus performants, c'est-à-dire affichant les valeurs optimales de figures de mérite (r^2 , RMSEC, RMSECV, RMSEP, LOD) ;
- Prédire la concentration élémentaire de nouvelles pastilles de sols agricoles n'ayant pas été utilisées dans la construction des modèles d'étalonnage afin de les valider.

i. Analyse univariée

a) Manque de corrélation entre le signal LIBS et la concentration

Nous avons vérifié que la régression univariée classique, c'est-à-dire la corrélation entre l'intensité ou l'aire nette de la raie d'émission de l'analyte et sa concentration mesurée par la méthode de référence (ICP-AES), ne fonctionnait pas du tout dans le cadre de notre étude sur les sols agricoles. On peut même parler d'un manque de corrélation. La figure 3-9 représente les modèles bâtis à partir des raies d'émission de deux éléments, le magnésium et le fer, choisis à titre d'exemples. On remarque, pour Mg premièrement, que les points ne s'alignent pas le long d'une droite de régression mais forment plutôt une sorte de nuage élargi de points. Ce nuage est caractérisé par des gradients d'intensité (étalement vertical) pour une même valeur de concentration et inversement, par des gradients de concentration (étalement horizontal) pour une valeur d'intensité donnée. En ce qui concerne Fe, bien que les trois points les plus concentrés semblent apporter de la variabilité logique, c'est-à-dire une augmentation de l'intensité proportionnelle à la concentration, le même constat que pour Mg peut être dressé pour le reste des points analytiques. Ce type de résultat a été obtenu pour l'ensemble des éléments d'intérêt dont certaines courbes d'étalonnage sont présentées en figure 3-4 en annexe.

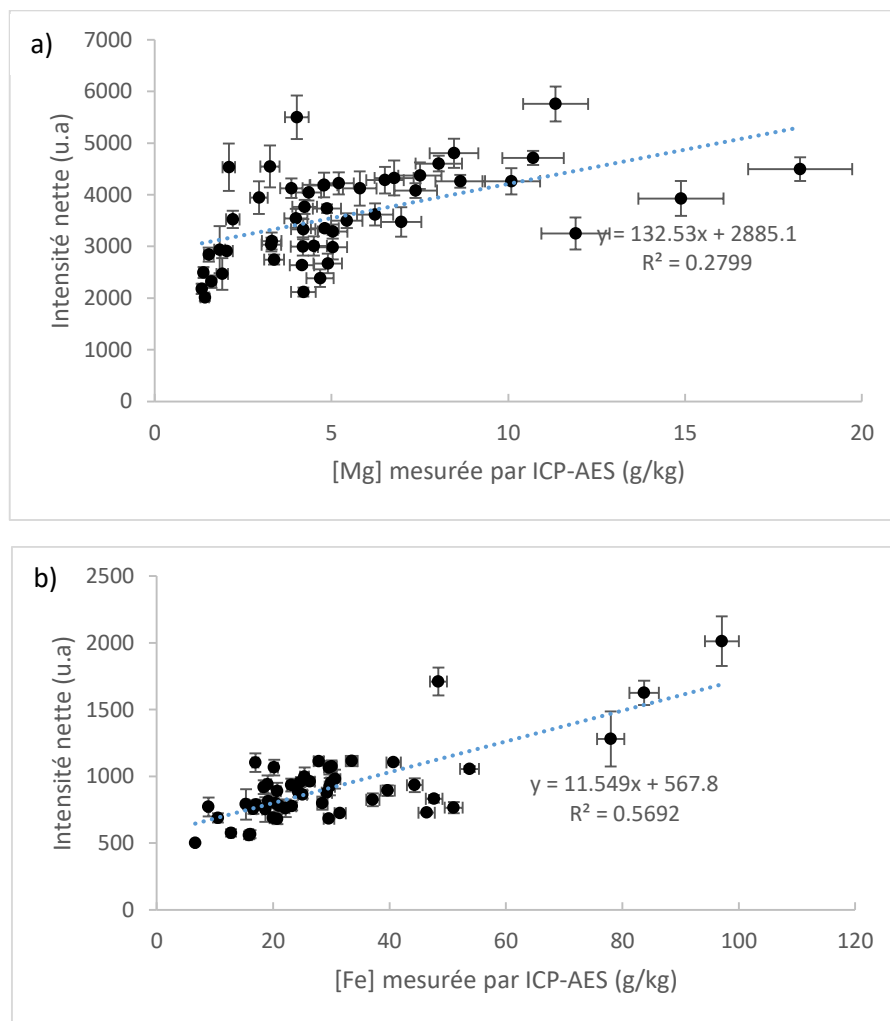


Figure 3-9 : Modèle de régression univariée obtenu sur les sols agricoles en LIBS pour a) le magnésium et b) le fer.

La normalisation spectrale se révélant efficace afin de corriger les fluctuations de plasma observées lors de la campagne de mesures des plantes, il était naturel d'évaluer si cette correction pouvait être efficace pour régler les problèmes d'effets de matrice rencontrés pour les sols. Les mêmes normalisations que pour les plantes ont donc été appliquées aux spectres de sols, à savoir : la normalisation SNV, la normalisation par étalon interne (en prenant la raie de l'argon à 763 nm comme étalon), la normalisation par l'aire totale et la normalisation par le fond continu. La correction des spectres par cette méthode s'est cependant révélée inefficace comme le montre la figure 3-10. Les courbes d'étalonnage après normalisation SNV et aire totale pour le magnésium y sont présentées, à titre d'exemple. Le même constat peut être dressé pour les modèles d'étalonnage univariés normalisés et non normalisés, c'est-à-dire l'observation d'un manque de corrélation entre le signal LIBS et la concentration mesurée traduit par une dispersion aléatoire des points de mesure. Les coefficients de corrélation, même s'ils ne sont pas ici très significatifs, sont également inférieurs à celui obtenu sur les données non-normalisées, confirmant que les tentatives de normalisation n'ont aucun effet dans ce contexte.

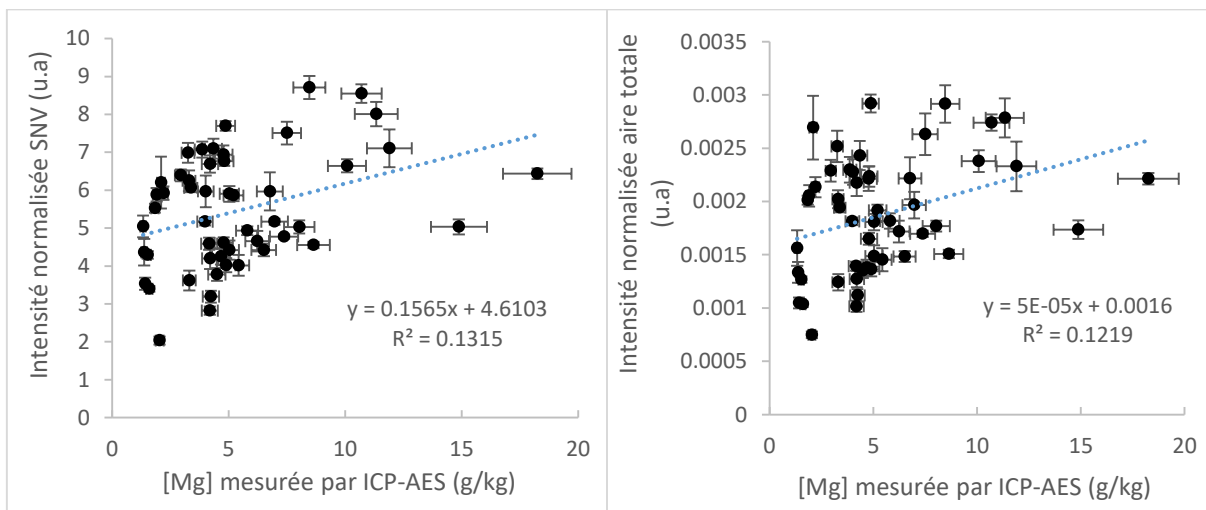


Figure 3-10 : Modèle de régression univariée obtenu pour le magnésium dans les sols agricoles en LIBS après a) normalisation SNV et b) normalisation par l'aire totale.

Ce type de résultat est observé quelle que soit la méthode de normalisation sélectionnée et l'élément quantifié, à l'exception du calcium.

b) L'exception du Calcium

Le calcium est un cas particulier car une certaine corrélation entre le signal LIBS et la concentration semble être déjà observée sur les données non normalisées malgré la présence de certains *outliers*. C'est ce que l'on observe sur la figure 3-11 a., montrant des points assez dispersés autour de la courbe de tendance linéaire de la régression entre l'intensité nette LIBS de la raie de Ca à 315,9 nm et la concentration mesurée. Malgré la dispersion de ces points dont

certaines pouvant être certainement considérés comme *outliers*, le coefficient de corrélation s'approche tout de même de 0,82. Les données normalisées, obtenues après application du calcul SNV (sans prendre en compte le doublet de Ca II à 393,3-396,6 nm et la raie de Ca I à 422,7 nm qui sont des raies très émissives), semblent également montrer une forte corrélation entre le signal et la concentration mais cette fois quadratique, comme représenté par la figure 3-11b. On remarque en effet qu'à l'exception de certains points aberrants entourés en rouge, l'intensité augmente avec la concentration jusqu'à atteindre un seuil autour de 8 en terme d'intensité normalisée. L'aspect quadratique d'une courbe de régression s'explique généralement en LIBS par une auto-absorption du signal lorsque ce dernier sature. C'est ici le cas du doublet de Ca II à 315,9-317,9 nm dans les spectres de plantes.

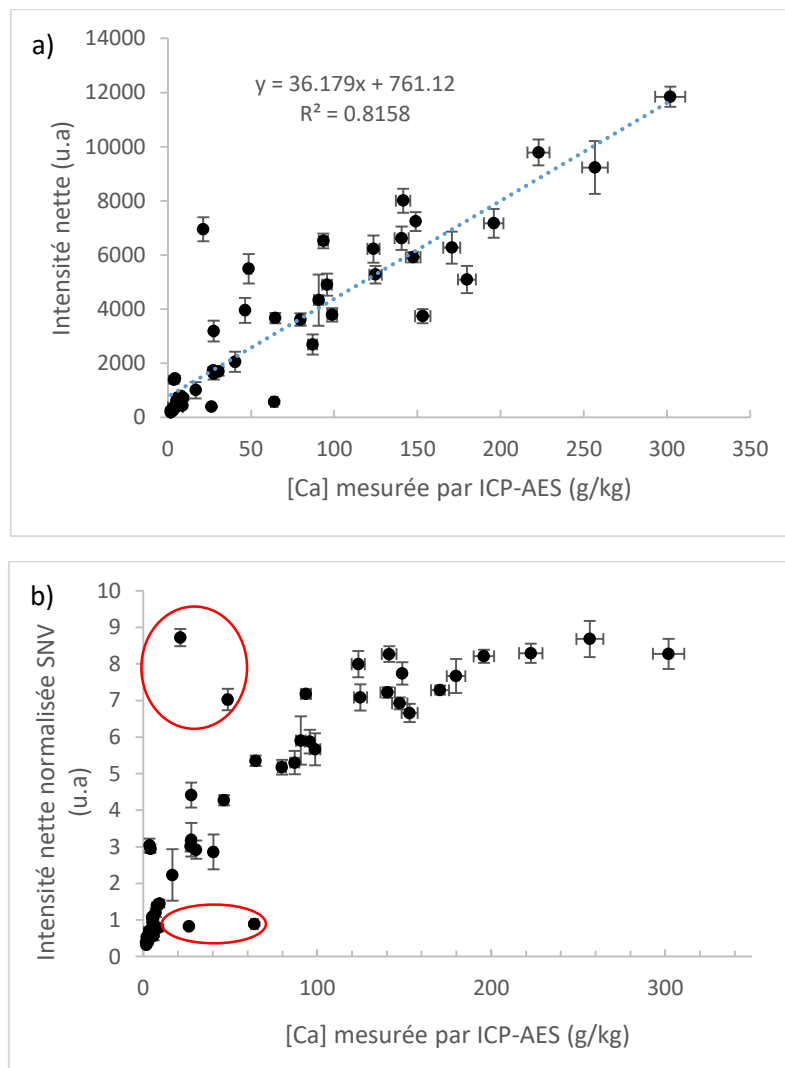


Figure 3-11 : Modèle de régression univariée obtenu pour le calcium dans les sols agricoles en LIBS
a) sans normalisation et b) après normalisation SNV. Les ellipses rouges indiquent les échantillons considérés comme outliers.

En règle générale, ces raies sont évitées pour exploiter des raies moins émissives mais mieux définies. Cependant, dans notre cas, les autres raies de Ca étant encore plus saturées ou décalées

en longueur d'onde, la raie à 315,9 nm est donc la plus facilement exploitable avec la raie à 317,9 nm présentant les mêmes propriétés.

Dans le cadre de notre étude, il n'a pas été décidé d'appliquer de prétraitement spectral tel qu'une méthode de fitting pour affiner l'allure de ce doublet de Ca dans le but de conserver les spectres dans leur forme la plus brute possible. Nous avons démontré [83] qu'il était possible d'atteindre les mêmes performances quantitatives issues de spectres LIBS traités par déconvolution qu'avec les mêmes spectres mais conservés dans leur forme la plus brute possible. Plus précisément, le but de cette étude était de construire un modèle quantitatif pour la raie élémentaire de Li I à 670,8 nm, fortement interférée par la raie de Ca I à 671,8 nm, dans des échantillons servant à calibrer l'instrument ChemCam, fixé sur le robot Curiosity dédié à l'analyse des sols sur Mars [84, 85]. Nous montrons dans ce cas qu'un modèle PLS appliqué aux données brutes est aussi performant qu'une extraction de signal utile par fit multi-pic.

Nous estimons, même si le cas de figure présent est différent et ne requiert pas l'utilisation d'une technique de fit aussi poussée mais uniquement un simple fit de Voigt par exemple, qu'il est tout à fait possible d'obtenir un modèle quantitatif satisfaisant en exploitant directement des raies élémentaires affichant un tel profil.

Nous avons décidé, au lieu de considérer l'intensité maximale ou l'aire nette de la raie de Ca comme signal à conserver pour bâtir le modèle de régression, de n'exploiter que l'aire nette de la raie de Ca sans prendre en compte les deux longueurs d'onde présentant les valeurs d'intensité les plus élevées ou de prendre l'intensité nette de la longueur d'onde associée à un bord de la raie. On remarque alors, à l'aide de la figure 3-12, que la courbure de la régression, qui est très marquée lorsqu'on prend comme signal l'intensité maximale (curseur rouge et curseur orange), diminue de plus en plus pour donner une régression beaucoup plus rectiligne lorsqu'on prend comme signal l'intensité sur une aile du pic (curseur vert).

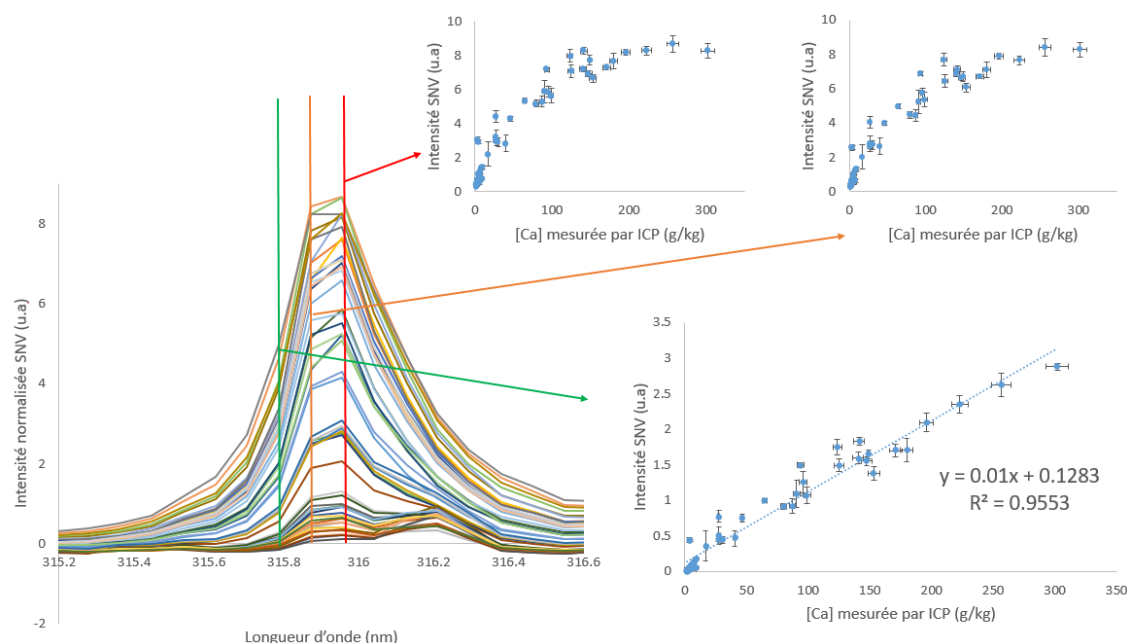


Figure 3-12 : Schéma permettant de mettre en évidence l'augmentation de la linéarité de la régression univariée à partir de la raie du calcium à 315,9 nm lorsque l'on utilise une longueur d'onde associée à l'aile de la raie (curseur vert) au lieu de celle attribuée à l'intensité maximale (curseurs rouge et orange).

Le modèle final pour le calcium, après avoir éliminé les *outliers* observés sur la figure 3-11b. et les échantillons sous la limite de détection, est caractérisé par : $r^2 = 0,96$, RMSEC = 16,8 g/kg et RMSECV = 16,7 g/kg , LOD = 14 g/kg. Le modèle de régression est illustré par la figure 3-13.

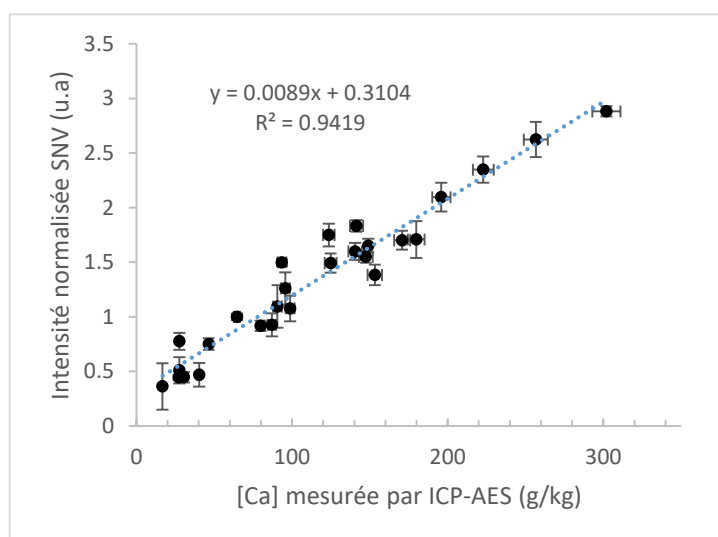


Figure 3-13 : Modèle de régression univariée final obtenu pour le calcium dans les sols agricoles par LIBS.

Le calcium reste cependant un cas particulier car il n'est pas possible, via un modèle univarié, de construire des modèles d'étalonnage pour l'ensemble des éléments détectés restants. Il est donc indispensable d'utiliser des méthodes de régression multivariées pour exploiter les spectres LIBS.

ii. Analyse multivariée

En ce qui concerne l'analyse univariée, dans le cadre de l'analyse des sols agricoles par LIBS, la proportionnalité entre le signal observé et la concentration de l'analyte n'est pas suffisamment élevée pour bâtir des modèles quantitatifs satisfaisants, et ce, même après la correction des spectres par normalisation. Ceci est dû à des effets de matrice importants entre les différents sols. Les variabilités de signal enregistrées sur tous les canaux des spectres LIBS permettent cependant de prendre en compte les multiples facettes du jeu de données [28] via une approche multivariée dont le principe général est schématisé sur la figure 3-14. L'analyse multivariée quantitative réalisée à partir d'un spectre LIBS consiste simplement en la prise en compte de plusieurs signaux spectraux : ceux liés à l'analyte dont on cherche à donner la concentration et d'autres, associés à des éléments différents de l'analyte. La méthode d'analyse multivariée la plus couramment utilisée dans un but quantitatif est la régression par moindres carrés partiels (PLSR) dont le principe est détaillé dans la partie suivante.

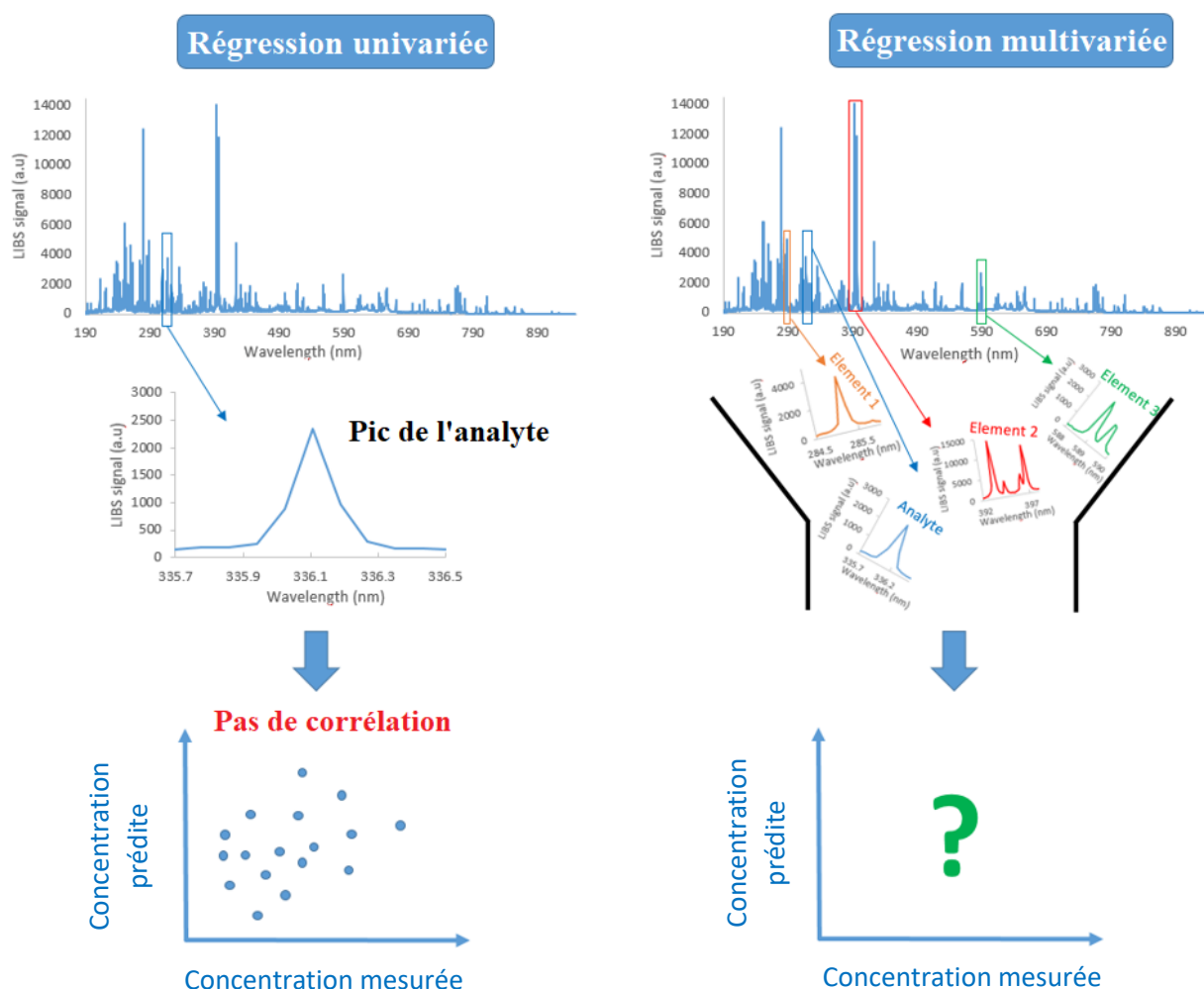


Figure 3-14 : Schéma de principe de l'analyse univariée et multivariée.

a) Principe de la régression aux moindres carrés partiels

La régression par moindres carrés partiels, appelée régression PLS pour Partial Least Squares, peut être exploitée à partir d'un jeu de données composé de variables explicatives X et d'une variable à expliquer Y . Dans un contexte spectroscopique, les variables explicatives sont les valeurs d'intensités acquises à chaque longueur d'onde du spectre et la variable à expliquer est la valeur de concentration de l'analyte sélectionné. La PLS fonctionne sur le même principe que l'ACP dans le sens où ce sont les covariances entre les variables qui sont évaluées et plus précisément la manière dont ces dernières permettent d'expliquer la variance de la concentration. En effet, là où en ACP on ne s'intéresse qu'à la matrice des covariances des données d'entrée X , on s'intéresse en PLS aux variables qui présentent la plus grande variance corrélée à la concentration de l'analyte. On considère alors que la PLS est une méthode d'analyse supervisée car son utilisation sous-entend une phase d'étalonnage avec des échantillons dont la concentration est connue, à la différence de l'ACP, qui est une méthode non-supervisée basée uniquement sur la variance des données d'entrée. La stratégie de calibration de la PLS pour les sols agricoles est d'ailleurs identique à celle menée pour les

échantillons de plantes en approche univariée, à savoir : la construction des modèles de régression à partir du jeu de données décrit jusqu'ici et la validation de ces modèles à partir d'échantillons qui n'ont pas été utilisés dans la construction des modèles de régression.

La régression PLS est particulièrement utile pour trouver rapidement quelles sont les variables explicatives corrélées à la concentration au sein d'un jeu de données présentant significativement plus de variables que d'observations. C'est tout particulièrement le cas des données spectroscopiques qui sont couramment traitées par PLS en spectroscopie Raman [86, 87], proche-infrarouge (NIR) [88, 89], en fluorescence par rayons X [90, 91] en LIBS [28] et bien d'autres. L'utilisation de la PLS induit cependant le fait que l'on construise des modèles linéaires. Pour des modèles non-linéaires, la régression PLS n'est alors plus adaptée et il est recommandé de s'orienter vers des algorithmes tels que les réseaux de neurones artificiels (ANN) ou encore les machines à vecteurs de support (SVM).

L'association de la PLS et de la LIBS a déjà été exploitée pour un nombre conséquent de matériaux différents allant du charbon, à la poudre de lait, en passant par l'encre jusqu'au sol Martien [28].

b) Algorithme utilisé en PLS

La PLS se base sur le même principe que l'ACP et l'algorithme NIPALS, aussi bien dans le fait d'étudier la covariance entre les variables mais également dans le fait de compresser les données dans un espace de faible dimensionnalité symbolisé par des composantes. Par conséquent, le principe de modèle de calcul itératif pour déterminer les scores et les loadings de chaque composante est le même sauf qu'ici on cherche à ce qu'ils permettent de décrire le maximum de covariance entre la matrice des données d'entrée X et la variable à expliquer Y . En d'autres termes, d'expliquer au maximum la covariance entre la matrice des spectres LIBS moyens X et la concentration de l'analyte à quantifier Y . On part donc des deux équations suivantes [42, 92]:

$$\text{Equation 3-1} \quad X = t_1 p_1^T + t_2 p_2^T + \dots + t_a p_a^T + E_a$$

$$\text{Equation 3-2} \quad y = t_1 q_1 + t_2 q_2 + \dots + t_a q_a + f_a$$

avec l'équation 3-1, la même écriture matricielle que pour l'ACP (cf chapitre II) et comprenant le produit entre les scores t et les loadings p et E la matrice des résidus. t_1 et p_1 représentent les scores et loadings de la première composante jusqu'à la composante « a » jugée finale. Dans l'équation 3-2, y représente les mêmes scores que X mais avec un vecteur q et des résidus f qui lui sont propres. A l'aide de cette double équation on peut maintenant maximiser la corrélation entre les spectres et les concentrations. Pour cela on fixe tout d'abord une condition initiale avec $E_0 = X$ et $f_0 = y$ afin de s'assurer que chaque composante explique un maximum de variance. Ceci donne, par réécriture des équations 3-1 et 3-2 et pour chaque composante :

$$\text{Equation 3-3} \quad E_a = E_{a-1} - t_a p_a^T$$

$$\text{Equation 3-4} \quad f_a = f_{a-1} - t_a q_a$$

Il est également possible de relier les deux résidus, pour chaque composante, à l'aide d'un nouveau terme appelé le poids W :

Equation 3-5
$$W_a = E_{a-1}^T f_{a-1}$$

A l'aide des conditions initiales déterminées précédemment et des équations établies, il est maintenant possible, en suivant l'ordre des équations 3-6 jusqu'à 3-11, de calculer les paramètres de la première composante :

Equation 3-6
$$W_1 = E_0^T f_0 = X^T y$$

Equation 3-7
$$t_1 = E_0 W_1 = X W_1$$

Equation 3-8
$$p_1 = \frac{E_0^T t_1}{t_1^T t_1}$$

Equation 3-9
$$q_1 = \frac{f_0^T t_1}{t_1^T t_1}$$

Equation 3-10
$$E_1 = E_0 - t_1 p_1^T$$

Equation 3-11
$$f_1 = f_0 - t_1 q_1$$

Il est ensuite possible de recommencer le processus à partir de l'équation 3-6 mais en déterminant les paramètres de la composante 2 à l'aide de ce qui a été obtenu pour la première composante et ainsi de suite jusqu'à la composante jugée finale.

La PLS calculant la covariance entre les spectres et la concentration, il convient de détailler les figures de mérite servant d'outils de décision pour déterminer le nombre de composantes optimales et caractériser les performances du modèle final obtenu :

- Le graphe de l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes principales incluses dans le modèle. L'erreur est ici calculée par cross-validation et plus précisément par la méthode du Leave-One-Out, décrite au chapitre II. La RMSECV est censée diminuer avec l'ajout de composantes au modèle PLS jusqu'à une valeur minimum à partir de laquelle elle se stabilise. Le nombre de composantes jugé optimal est le plus petit nombre qui permet d'atteindre cette valeur de RMSECV minimum.
- Le paramètre W , appelé poids des variables, peut être observé soit pour une composante donnée, on parle alors du graphe du poids des variables, soit pondéré par le nombre de composantes retenues, on parle alors de graphe des coefficients.
- Le graphe des concentrations prédites en fonction des concentrations connues. Ce dernier permet d'observer la régression entre les valeurs de concentrations connues (mesurées par ICP-AES) et les valeurs de concentrations prédites par le modèle PLS.

Les spectres de sols agricoles traités en PLS, pour les mêmes raisons que pour l'ACP, sont moyennés puis centrés. Les calculs ont été effectués à l'aide du logiciel R Studio [54] et plus précisément à l'aide du package « chemometrics » et de la fonction « pls ».

c) Construction du modèle d'étalonnage par PLS du magnésium

Lorsque l'analyse PLS est appliquée aux spectres LIBS, les publications associées fournissent rarement les détails suffisants. Pourquoi un tel nombre de composantes a été sélectionné ? Pourquoi ces variables ont été incluses dans le modèle et pas les autres ? La réponse à ces questions est pourtant fondamentale afin d'avoir une meilleure compréhension du jeu de données et de s'assurer de travailler avec des modèles d'étalonnage robustes. Ajouté à cela, l'utilisation de logiciels ou d'algorithmes dont on connaît peu le contenu [29] et il devient très vite compliqué de comprendre comment une stratégie d'analyse multivariée peut fonctionner.

En conséquence, nous avons décidé d'explicitier la construction d'un modèle PLS à partir de nos propres données, pas à pas, à travers l'exemple de l'élément magnésium dans les sols agricoles. Cependant, beaucoup d'informations sur la manière de construire un modèle méthodiquement et de le valider peuvent être trouvées dans la référence suivante [93]. Un premier modèle PLS a d'abord été calculé, après recalage en longueurs d'onde, en prenant en compte tous les échantillons de sols (51) et le spectre entier, c'est-à-dire de 188 à 950 nm, soit 5843 variables par spectre. Le tableau de bord, comprenant les figures de mérite décrites précédemment, est présenté en figure 3-15.

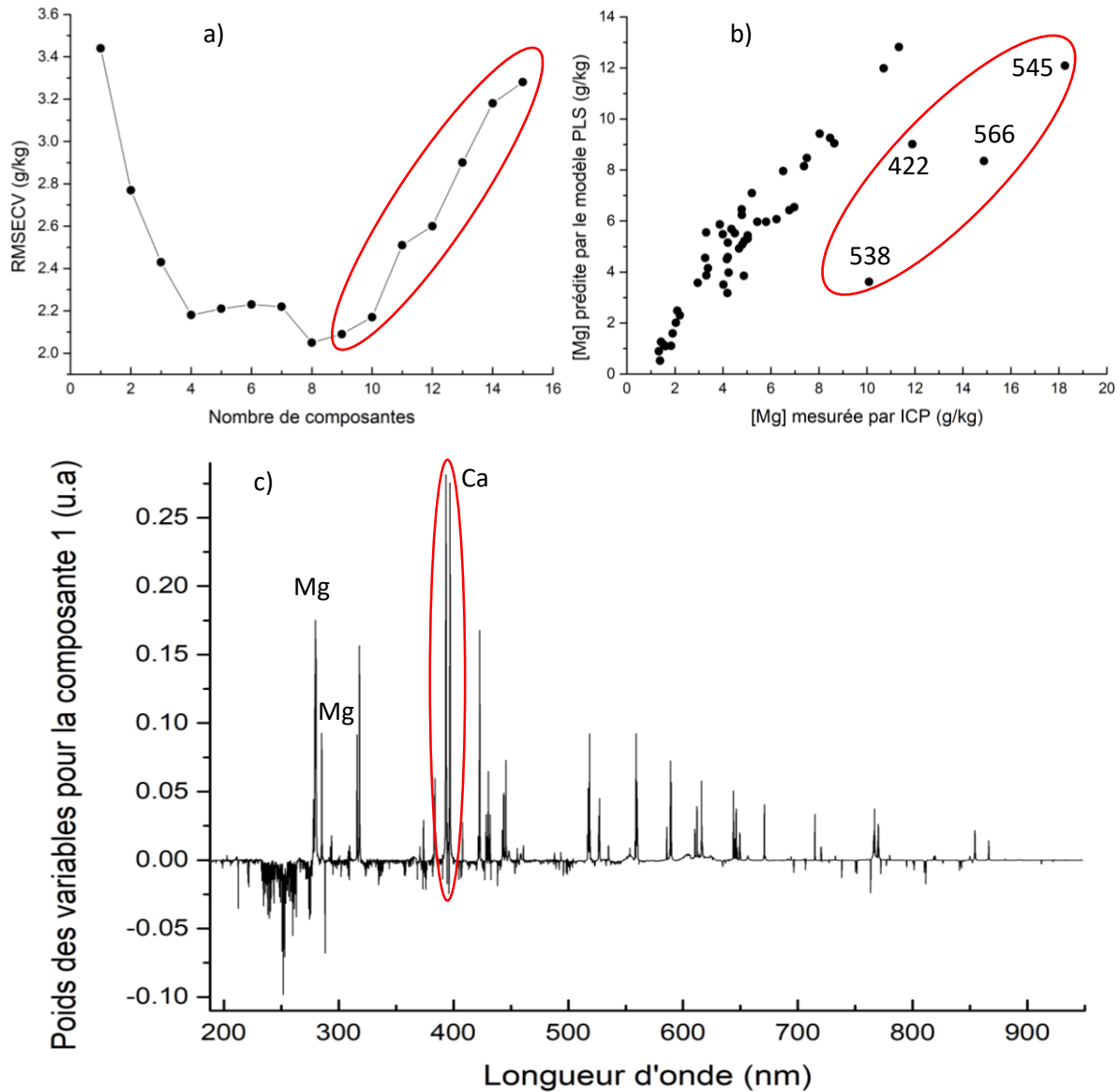


Figure 3-15 : Tableau de bord des figures de mérite pour le modèle PLS du magnésium calculé sur tout le spectre avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) la régression entre la concentration mesurée par la méthode de référence (ICP-AES) et la concentration prédite par le modèle et c) le graphe du poids des variables associé à la première composante. Les ellipses rouges représentent les zones jugées « anormales » sur les figures de mérite et sont discutées dans le texte.

La figure 3-15 montre 3 figures de mérite associées au modèle PLS dédié à la quantification du magnésium et construit en prenant en compte tout le spectre LIBS, soit 5843 variables. La figure 3-15a. représente l'évolution de l'erreur du modèle PLS (RMSECV) en fonction du nombre de composantes. On peut remarquer que cette erreur diminue logiquement avec l'ajout de composantes, car celui-ci permet d'expliquer de plus en plus de variance corrélée à la concentration en magnésium. L'erreur RMSECV atteint 2,2 g/kg, au bout de 4 composantes, valeur qui se stabilise sur les 3 composantes suivantes. Cependant, au contraire de l'évolution attendue marquée par une valeur d'erreur qui se stabilise avec l'ajout de composantes, celle-ci remonte ensuite jusqu'à 3,4 g/kg à la composante 15. Ce type de profil d'évolution d'erreur

traduit un manque de stabilité du modèle PLS qui trouve sa source dans la figure 3-15b. En effet, cette dernière illustre la régression entre la concentration en Mg mesurée par la méthode de référence (ici ICP-AES) et la concentration prédite par le modèle PLS. On y observe une certaine proportionnalité entre la concentration mesurée et celle prédite sur la majorité des échantillons à l'exception de 4 d'entre eux, indiqués par une ellipse rouge, qui se détachent très clairement de la tendance suivie par les autres points. Les intensités LIBS des raies du magnésium pour ces échantillons semblent en effet particulièrement basses par rapport à des échantillons présentant les mêmes concentrations. Il n'a cependant pas été possible, à travers l'observation des différents paramètres mesurés sur ces sols (pH, matière organique, texture...), des étapes de prétraitement ou encore du protocole analytique, de trouver une raison à ces écarts. Ces échantillons sont donc considérés comme *outliers* et rejetés pour la construction du modèle PLS pour le magnésium.

Finalement, il est aussi possible d'observer comment les variables (ici les intensités acquises à chaque longueur d'onde) impactent le modèle. On s'intéresse tout particulièrement ici, à travers la figure 3-15c., aux poids des variables sur la composante 1. En effet, par définition, la première composante est celle qui explique le plus de variance par rapport à la concentration de l'analyte. Il est logiquement attendu que les variables qui ont du poids sur la composante 1 soient celles associées aux raies de l'analyte. La figure 3-15c. montre en effet un certain poids positif des raies de Mg (à 280,3 nm et 285,2 nm) sur la composante 1 mais la valeur la plus importante de poids est attribuée au calcium et plus précisément au doublet de Ca II à 393-396nm (ellipse rouge). Ce doublet étant particulièrement sujet à des phénomènes d'auto-absorption, comme démontré lors du chapitre II dans le cas de l'analyse des plantes, ce résultat doit particulièrement alerter sur l'intérêt de procéder à une sélection des variables à inclure dans la construction du modèle PLS. Après vérification sur les spectres de sols, le doublet de Ca entouré en rouge semble en effet présenter des largeurs de raies à mi-hauteur variables d'un échantillon à l'autre. La composante 1, qui explique le plus de variance dans le modèle PLS est donc partiellement associée à une signature spectrale présentant des anomalies et pouvant biaiser la construction du modèle.

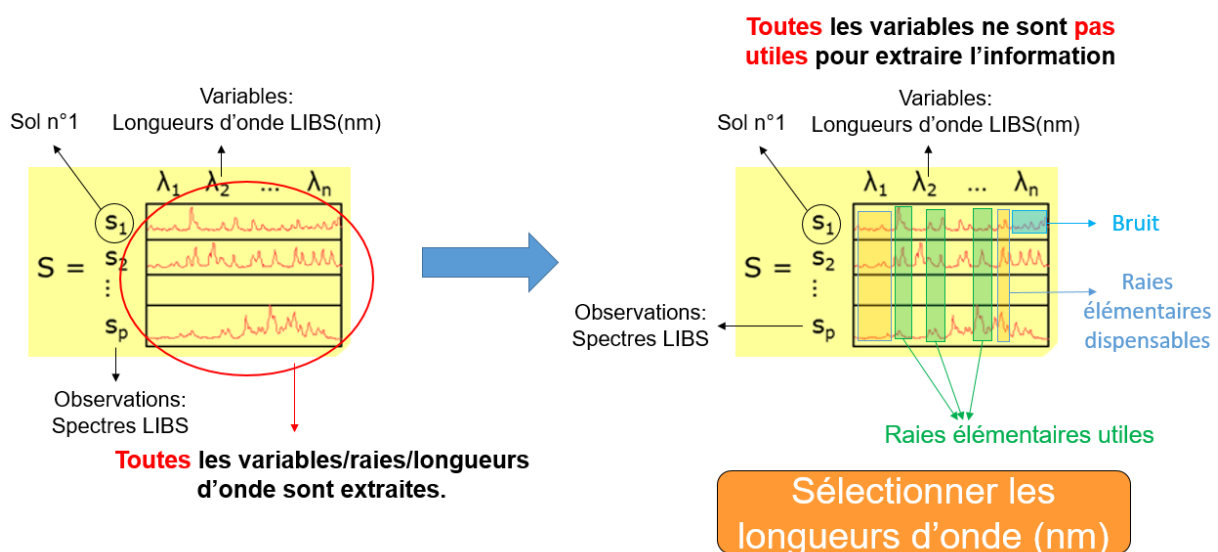


Figure 3-16 : Schéma de principe de la sélection de variables en analyse PLS.

La construction d'un modèle PLS consiste donc à sélectionner les variables d'entrée avec pertinence. Un schéma de principe de la sélection de variables est donné en figure 3-16. Au lieu de calculer un modèle PLS sur l'ensemble de la gamme spectrale à disposition, il est souhaitable de calculer un modèle PLS sur un nombre réduit de variables afin de simplifier le modèle, voire de l'optimiser [93]. Plusieurs méthodes sont employées pour déterminer quelles sont les variables pertinentes à conserver dans le modèle :

- Il est possible de sélectionner *a priori* les raies élémentaires à conserver sur la base de l'expertise de l'analyste et de sa connaissance des matériaux analysés [94, 95, 96]. Cela peut permettre de simplifier le modèle mais ne permet pas d'éviter l'utilisation de raies présentant une anomalie comme celles qui sont affectées par le phénomène d'auto-absorption ou encore celles qui sont affectées par un problème d'interférence spectrale.
- L'utilisation d'algorithmes comme *interval partial least squares* (iPLS) et ses dérivées (BiPLS, FiPLS, mIPW-PLS) qui consiste à faire une recherche exhaustive de la meilleure variable ou combinaison de variables en créant un modèle PLS pour chaque combinaison. Le modèle présentant l'erreur la plus faible est par la suite conservé et exploité [87].
- En exploitant les figures de mérite issues du calcul PLS comme le graphe des coefficients [97], le graphe du poids des variables ou encore celui des *Variable Importance in Projection* (VIP). Le graphe des VIP est un outil qui permet d'évaluer rapidement combien de variables et surtout lesquelles sont utiles afin d'expliquer les variabilités spectrales entre échantillons [98], et ce, à l'aide d'une somme pondérée des carrées du poids des variables issues du calcul PLS. En d'autres termes, les VIP permettent de représenter l'importance de chacune des variables d'entrée (ici les longueurs d'onde) pour caractériser les données spectrales et la concentration. En règle générale, une variable présentant une valeur de VIP inférieure à 1 symbolise une variable négligeable susceptible d'être retirée du jeu de données. En 2017 [99], nous avons montré l'importance de combiner le graphe des coefficients et celui des VIP dans le but de quantifier le potassium dans des échantillons de sols agricoles. Le système LIBS utilisé alors était différent de celui présenté dans cette étude. Le graphe des VIP nous avait permis de réaliser un premier balayage des raies inutiles et de ne conserver que les raies les plus influentes sur le modèle. Le graphe des coefficients nous avait ensuite permis de démontrer que les variables associées au centre des raies de K I à 766,5 et 769,9 nm n'étaient pas du tout corrélées à la concentration en potassium. Ceci s'expliquait par une saturation du signal probablement due au phénomène d'auto-absorption pour les fortes valeurs de concentrations. Le graphe des coefficients associé et la signature spectrale du doublet sont visibles en figure 3-5 en annexe. Au final, nous avons conclu que le modèle PLS ne prenant en compte que les ailes des raies du potassium présentait les meilleures capacités prédictives.

A la suite de ce qui a été observé en ACP et sur le graphe du poids des variables en figure 3-15c., nous avons décidé de calculer un nouveau modèle PLS à partir de spectres LIBS limités aux variables situées dans le domaine spectral allant de 188 à 350 nm. En effet, la majorité des raies élémentaires (C, P, Mg, Ca, Si, Al, Fe) sont présentes dans cette gamme spectrale et cela permet par ailleurs d'exclure certains signaux présentant des anomalies comme par exemple le doublet de Ca II à 393-396 nm. Enfin, après un recalage manuel en longueur d'onde de certains spectres, nous avons ajouté quelques variables à celles associées à l'intervalle 188-350 nm, ces variables permettant de prendre en compte le rôle du sodium (43 variables dans l'intervalle 588-

590 nm) et du potassium (66 variables dans l'intervalle 765-771 nm). Le tableau de bord de ce nouveau modèle, est présenté en figure 3-17.

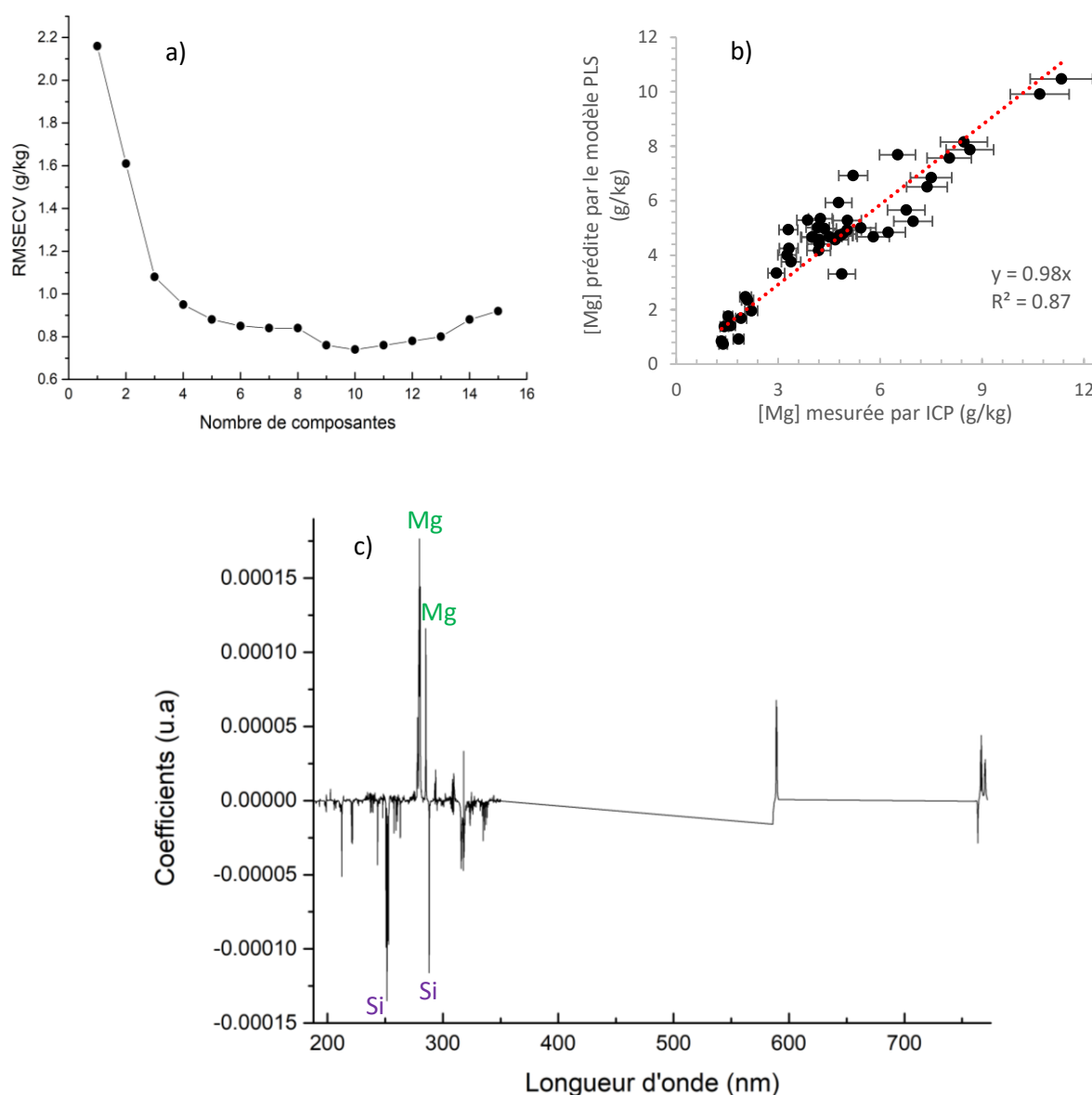


Figure 3-17 : Tableau de bord des figures de mérite pour le modèle PLS final du magnésium avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) la régression entre la concentration mesurée par la méthode de référence (ICP-AES) et la concentration prédite par le modèle pour quatre composantes et c) le graphe des coefficients associé à la quatrième composante.

En comparaison avec le modèle PLS calculé sur tout le spectre LIBS, on remarque que l'on atteint une erreur plus faible lorsque l'on observe l'évolution de la valeur de RMSECV en fonction du nombre de composantes sur le modèle tronqué à l'aide de la figure 3-17a. On obtient en effet une valeur d'erreur proche de 0,9-1,0 g/kg au bout de 4-5 composantes soit une valeur deux fois plus faible qu'en prenant en compte la totalité des variables. On remarque également que, cette fois-ci, la RMSECV se stabilise autour de 0,8-0,9 g/kg même après l'ajout de plus en

plus de composantes. Il a donc été décidé de fixer le nombre de composantes optimal à conserver à 4, car c'est à partir de cette composante que la valeur de RMSECV ne varie presque plus. Ensuite, pour 4 composantes, on observe bien, à l'aide de la figure 3-17b., que la régression linéaire (et passant par 0) entre la concentration mesurée par ICP-AES et celle prédite par le modèle PLS à partir des spectres LIBS est caractérisée par un coefficient de corrélation égal à 0,87. Finalement, le graphe des coefficients, présenté en figure 3-17c., qui est à ne pas confondre avec le graphe du poids des variables présenté en figure 3-15c., car le graphe des coefficients permet d'afficher les poids des variables pondéré du nombre de composantes (ici 4), révèle des valeurs de coefficient positives pour les raies de Mg et des valeurs négatives, aussi fortes, pour les raies de Si. Les doublets de Na et K montrent des valeurs positives mais nettement moins élevées. Cela signifie que pour 4 composantes, les variables qui pèsent le plus dans le modèle sont celles de Mg et Si, avec une anti-corrélation entre les deux éléments. Les concentrations de Si n'étant pas connues, il n'est cependant pas possible de vérifier si cette anti-corrélation est réelle. De plus, lorsque l'on calcule le graphe du poids des variables en fonction de la première composante de ce nouveau modèle, on peut maintenant y observer les valeurs de poids les plus élevées positivement pour les raies de Mg, alors que celles-ci étaient associées au doublet de Ca II à 393-396 nm pour le modèle prenant en compte la totalité du spectre comme démontré par la figure 3-15c. La première composante, soit celle qui permet d'expliquer le plus de variance par rapport à la concentration en magnésium, est donc maintenant logiquement associée aux raies élémentaires de l'analyte.

A ce stade de la construction du modèle, la question du nombre final de composantes à conserver est essentielle et ce point permet de discuter de ce qu'on appelle le sur-apprentissage ou overfitting en anglais [93]. En principe l'ajout de composantes permet en effet d'expliquer de plus en plus de variance liée à la concentration alors pourquoi ne pas en ajouter un maximum ? Tout simplement car on risque de se placer dans un cas de *sur-apprentissage*, c'est-à-dire que le modèle va faire en sorte de relier la variabilité aléatoire présente dans les spectres à ce qui reste à expliquer de la variance des données. Ceci va entraîner la création d'un modèle virtuellement parfait mais qui n'a en réalité aucune valeur prédictive. Un exemple est donné via le modèle du magnésium en figure 3-18. On y retrouve les figures de mérite habituelles si ce n'est que dans ce cas, le modèle construit comporte 15 composantes. On remarque alors, via la figure 3-18b. que la régression entre la concentration connue et prédite est quasiment parfaite avec un r^2 de 0,98. Ce résultat pourrait rassurer mais au contraire, lorsque l'on observe le graphe des coefficients à l'aide de la figure 3-18c., on remarque que l'allure du graphe des coefficients pour 15 composantes, entre 283 et 290 nm, révèle des anomalies pour la raie de Mg à 285,3 nm ainsi que pour la raie de Si à 288,2 nm. En effet, par comparaison avec le graphe des coefficients de la quatrième composante (en noir) proposée sur la même figure, là où celui-ci montre une allure bien définie de coefficients positifs pour Mg et négatifs pour Si, celui associé à la composante 15 (en rouge) présente des coefficients décrivant une courbe inattendue, notamment avec des changements de signe au sein d'une même raie spectrale. De plus, les zones spectrales ne présentant aucun signal sont associées à des valeurs de coefficient positives et négatives simultanément pour la composante 15 alors que pour la quatrième composante, ces zones sont logiquement représentées par une valeur de coefficient constamment nulle. Ce genre d'anomalie doit donc alerter l'analyste sur le fait de sélectionner moins de composantes pour bâtir son modèle PLS.

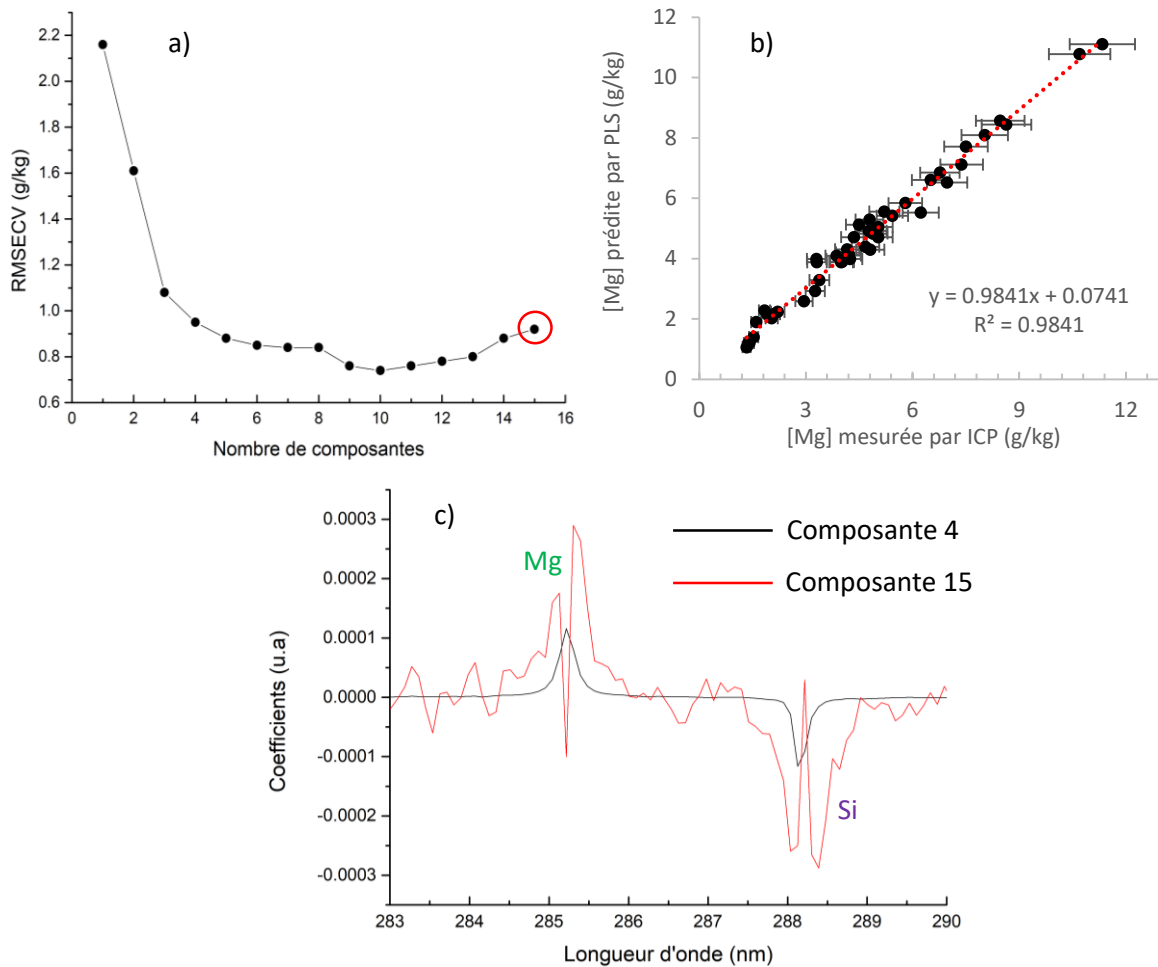


Figure 3-18 : Tableau de bord des figures de mérite pour le modèle PLS à 15 composantes avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) la régression entre la concentration mesurée par la méthode de référence (ICP-AES) et la concentration prédite par le modèle à 15 composantes et c) le graphe des coefficients associé à la quatrième (noir) et quinzième composante (rouge).

Le choix du nombre de composantes optimal à conserver se base alors sur le point d'inflexion de la courbe décrivant l'évolution de l'erreur RMSECV en fonction du nombre de composantes. Ici, ce point d'inflexion peut être discutable car selon l'analyste, il peut se trouver à 3, 4, 5 voire 6 composantes. C'est alors à l'analyste de trancher, en créant par exemple un modèle PLS pour chacune de ces 4 options et de sélectionner le modèle offrant les meilleures performances prédictives. Plus précisément, dans notre contexte d'étude, cette notion de meilleures performances prédictives est associée au modèle qui nous permettra d'obtenir les meilleures prédictions sur le jeu de validation, soit celui présentant l'erreur de prédiction (RMSEP) la plus faible. La réponse sera donnée dans la partie 2.ii.e. de ce chapitre.

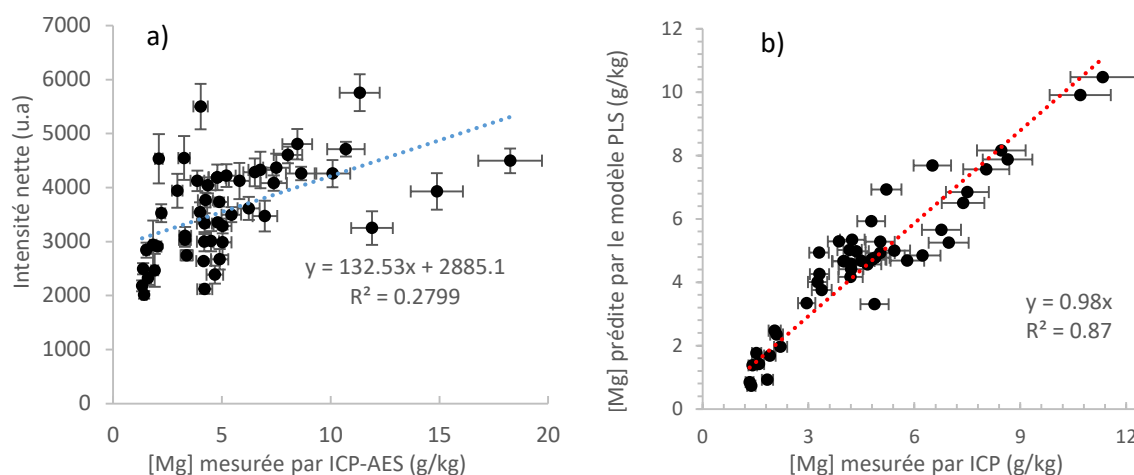


Figure 3-19 : Modèle de régression obtenu pour le magnésium après analyse LIBS des échantillons de sols agricoles a) univarié et b) PLS à 4 composantes avec sélection de variables.

Le passage de l'analyse univarié à la PLS, illustré par la figure 3-19, met clairement en lumière l'intérêt d'exploiter le caractère multivarié des spectres LIBS pour pouvoir quantifier les échantillons de sols agricoles. La même stratégie que celle décrite pour le magnésium a ensuite été utilisée pour l'analyse des autres éléments d'intérêt.

d) Quantification des éléments d'intérêt

La même stratégie d'étalonnage multivariée a été mise en œuvre pour le fer et le sodium et les figures de mérite des modèles PLS optimisés sont présentées dans les figures 3-20 et 3-21 respectivement. Ces modèles ont été obtenus après suppression des *outliers* (trois pour le modèle de Fe et six pour le modèle de Na) et après sélection de variables (les variables finales étant les mêmes que pour le modèle de Mg) en suivant le même protocole que celui décrit dans le point précédent. Les *outliers* peuvent être consultés en annexe à l'aide des figures 3-6 et 3-7. En ce qui concerne le fer, on obtient une RMSECV d'environ 6,4 g/kg pour un modèle à 4 composantes, jugé optimal. La régression linéaire entre les concentrations connues et prédites par PLS est caractérisée par un coefficient de corrélation égal à 0,88. Le graphe des coefficients en figure 3-20c. révèle des valeurs positives pour toutes les raies de Fe présentes dans l'U.V ainsi qu'une anti-corrélation avec les raies Si et dans une moindre mesure de Mg et Ca également.

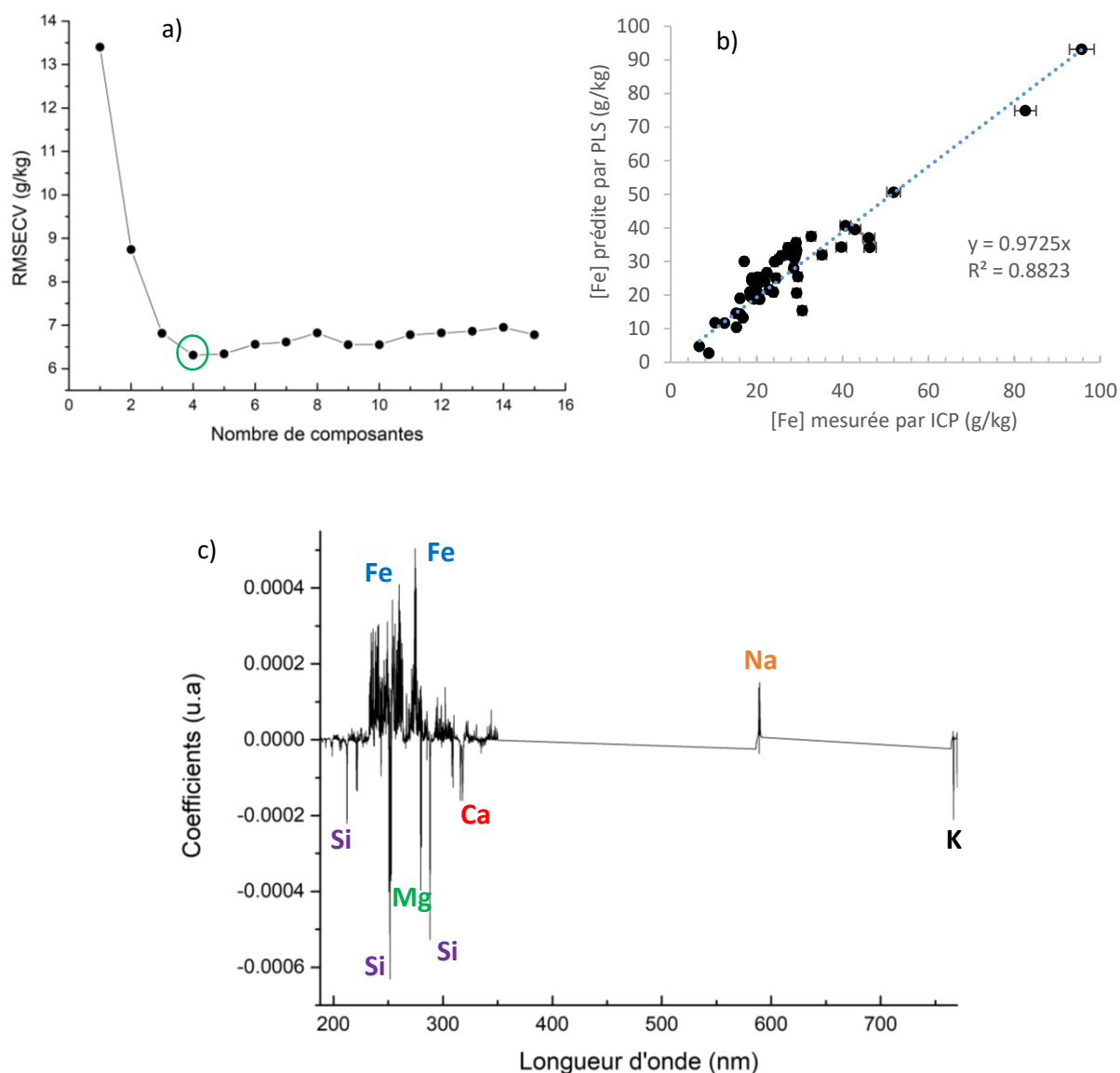


Figure 3-20 : Tableau de bord des figures de mérite pour le modèle PLS final du fer à 4 composantes avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) la régression entre la concentration mesurée par la méthode de référence (ICP-AES) et la concentration prédite par le modèle PLS à 4 composantes et c) le graphe des coefficients associé à la quatrième composante.

Le constat pour le sodium est assez similaire, en prenant les mêmes variables, avec un nombre de composantes optimal égal à 4, comme indiqué par un cercle vert sur la figure 3-21a pour une RMSECV cette fois de 0,95 g/kg. Le graphe des coefficients affiche quant à lui logiquement de fortes valeurs positives de coefficients pour les variables associées au doublet de Na (589,0-589,6 nm) ainsi que des valeurs moins élevées positives (Mg) et négatives (Si et Ca). La régression associée à ce modèle à 4 composantes permet d'obtenir une droite d'étalonnage pour le sodium avec un r^2 de 0,91.

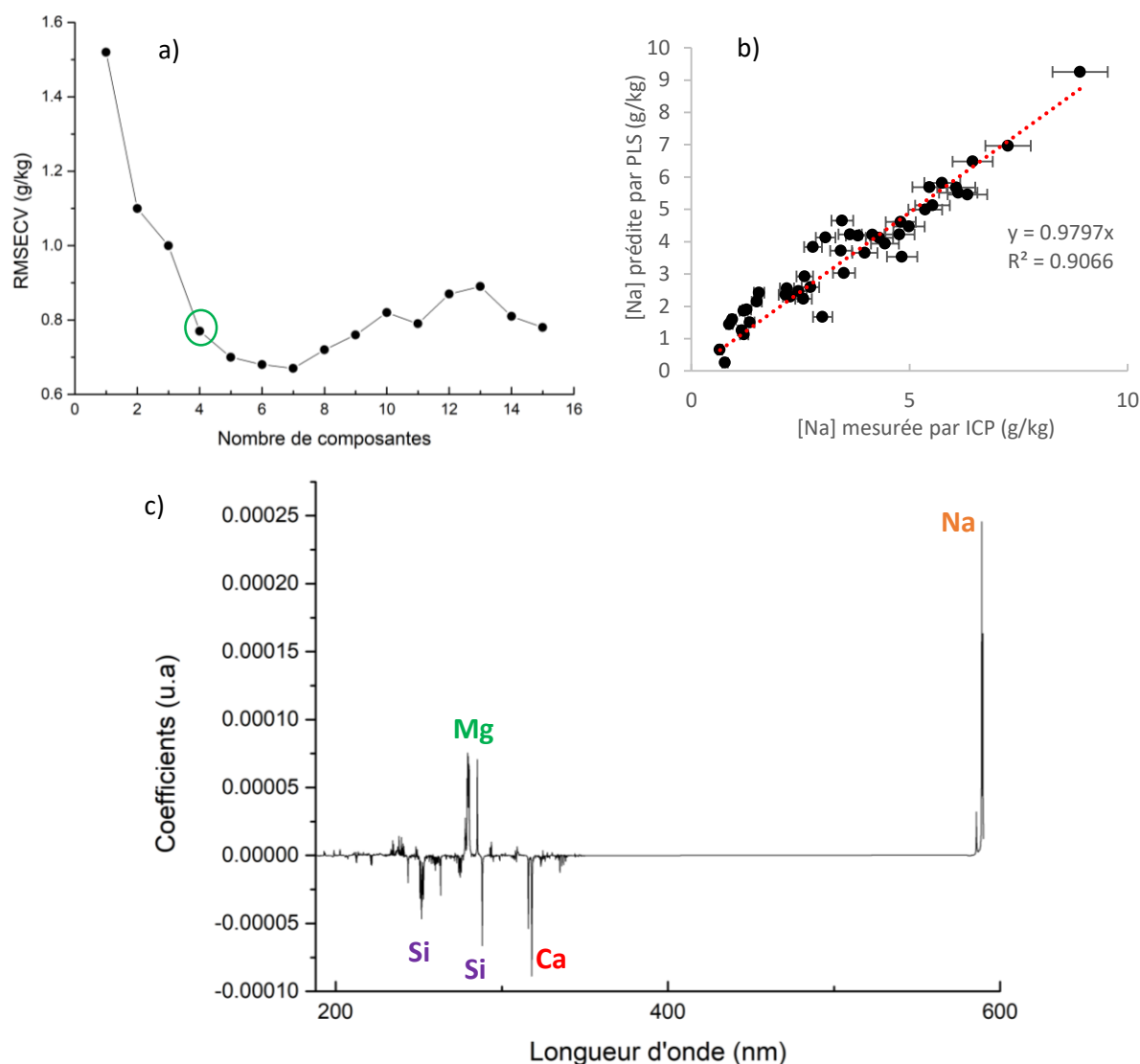


Figure 3-21 : Tableau de bord des figures de mérite pour le modèle PLS final du sodium à 4 composantes avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) la régression entre la concentration mesurée par la méthode de référence (ICP-AES) et la concentration prédite par le modèle PLS à 4 composantes et c) le graphe des coefficients associé à la quatrième composante.

Le fait de traiter les données LIBS de sols de manière multivariée, et plus précisément en utilisant la contribution d'autres raies d'émission associées à d'autres éléments comme Si, Mg, Na, Ca et Fe, permet de passer d'un étalonnage univarié inexploitable (cf la partie 2.i) à des modèles PLS satisfaisants pour Mg, Na et Fe. Cette méthode a cependant montré quelques limitations pour les éléments restants et notamment le potassium. On remarque en effet, à partir de la figure 3-22, que le modèle PLS relatif au potassium incluant toutes les variables du spectre LIBS présente une anomalie bien spécifique. Cette anomalie est particulièrement visible sur le graphe des coefficients, présenté par la figure 3-22b. et sélectionné pour un nombre de composantes jugé optimal, ici 6 (figure 3-22a.). Alors que l'on s'attend à un graphe des coefficients dominé par de fortes valeurs positives pour les variables associées à K, on remarque

que ce sont les variables liées au doublet de Na à 588 nm qui affichent, de loin, les valeurs les plus importantes. Et les valeurs de coefficient relatives au potassium ne sont finalement pas significativement plus élevés que d'autres éléments tels que Ca ou Mg. L'explication se trouve dans la composition des sols et dans la figure 3-23.

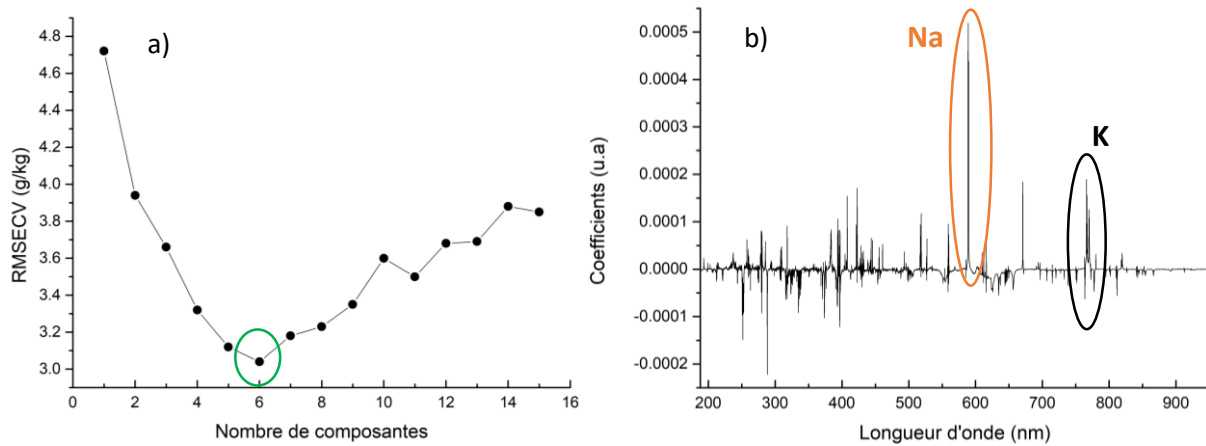


Figure 3-22 : Figures de mérite pour le modèle PLS du potassium calculé en prenant en compte l'entièreté du spectre LIBS avec a) l'évolution de l'erreur du modèle (RMSECV) en fonction du nombre de composantes, b) le graphe des coefficients pour la composante 6.

La forte contribution des variables associées au sodium sur l'explication des variabilités de la concentration en potassium provient tout simplement d'une corrélation entre la concentration en potassium et en sodium. La figure 3-23, illustrant la concentration en Na mesurée par ICP-AES en fonction de celle mesurée pour K, permet en effet de mettre en lumière une certaine corrélation entre les deux paramètres. On remarque bien que lorsque la concentration en K augmente, celle de Na aussi, ce qui est dû à la minéralogie des sols et possiblement de la présence en feldspaths, particulièrement riches en sodium et potassium. Le modèle PLS construit ne permet alors pas de quantifier réellement le potassium mais plutôt le sodium ce qui peut induire un biais dans les performances quantitatives du modèle.

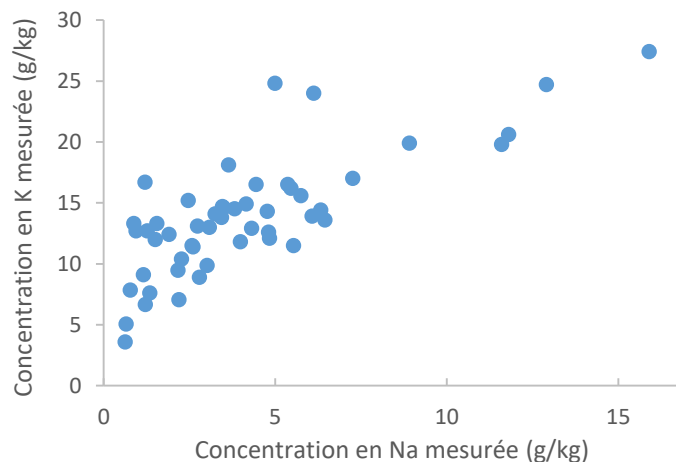


Figure 3-23 : Graphe de la concentration en sodium en fonction de la concentration en potassium mesurées par ICP (g/kg).

Pour preuve, un modèle PLS a été ensuite calculé à partir des mêmes variables auxquelles on a simplement retiré les variables associées au doublet de Na mais il n'a pas été possible d'établir dans ce cas la moindre corrélation entre le signal LIBS et la concentration en potassium.

En ce qui concerne les autres éléments, aucun modèle PLS exploitable n'a pu être calculé car soit aucun signal LIBS n'a été détecté (Cu, Zn, P, Mn), soit aucune corrélation n'a pu être établie entre le signal et la concentration (C, Al).

e) Validation des modèles d'étalonnage

Les échantillons utilisés pour valider les modèles de calibration sont issus de la même banque de données que les échantillons ayant servis pour construire les modèles PLS, à savoir des échantillons de sols agricoles issus du réseau QUASAPROVE. Plus précisément, 11 échantillons ont été sélectionnés sur la base de textures variées et de sorte à couvrir les gammes d'étalonnage des modèles construits. Le diagramme ternaire incluant les échantillons de validation est représenté sur la figure 3-8 en annexe. Les concentrations en Fe, Mg, Na des 11 échantillons ont donc été prédites par les modèles PLS associés et la concentration en Ca a été déterminée par le modèle univarié basé sur une normalisation SNV, présenté en 2.i.b. Les concentrations prédites pour chaque échantillon de validation (V1 à V11) sont reportées dans le tableau 3-2 ainsi que l'incertitude sur ces valeurs. La correspondance de ces noms d'échantillons avec les sols issus du réseau QUASAPROVE est fournie dans le tableau 3-2 en annexe. L'incertitude de chaque valeur prédite représente en réalité l'écart-type sur les concentrations prédites des répliquats (soit 5 par échantillon).

[Fe] g/kg	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
Connue	17.4	33.9	38.9	29.5	25.5	23.2	27.5	19.1	19.8	22.7	16
Prédite	15.7	32	32.3	31	23.5	27.9	26.7	18.9	27.3	24.7	9.5
Incertitude LIBS	1.19	2.46	0.75	1.79	2.78	1.77	3.06	2.09	1.81	1.07	2.02

[Mg] g/kg	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
Connue	2.63	5.00	14.8	1.61	3.48	3.23	3.99	2.99	2.87	3.62	3.61
Prédite	2.33	4.99	HG	0.05	4.26	4.81	4.93	3.76	3.89	5.24	3.90
Incertitude LIBS	0.62	0.32		0.05	0.30	0.28	0.15	0.5	0.23	0.20	0.61

[Na] g/kg	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
Connue	7.57	1.23	7.90	3.19	5.67	6.27	6.30	5.60	5.91	6.81	3.78
Prédite	7.09	1.85	9.23	3.31	5.94	7.07	5.45	6.29	6.87	6.93	5.04
Incertitude LIBS	0.81	0.37	0.69	0.38	0.45	0.40	0.52	0.61	0.54	0.87	0.32

Tableau 3-2: Récapitulatif des prédictions obtenues avec les modèles PLS de Fe, Mg et Na pour les 11 échantillons de validation. L'abréviation HG est utilisée pour les échantillons qui présentent une concentration hors gamme d'étalonnage.

Les modèles PLS de Fe, Mg et Na présentent des capacités prédictives satisfaisantes avec notamment l'obtention d'une RMSEP proche de 4,06, 1,04 et 0,79 g/kg respectivement. Ces valeurs sont en effet très proches de ce qui avait été obtenu en terme de RMSECV (6,3 g/kg pour Fe; 0,95 g/kg pour Mg et 0,76 g/kg pour Na) ce qui symbolise le fait que les modèles ne sont pas surévalués et sont particulièrement robustes. La valeur de RMSECV et RMSEP pour Fe est plus importante tout simplement parce que la gamme d'étalonnage du fer atteint des concentrations bien plus importantes que celles de Na et Mg.

En revanche, les prédictions concernant Ca à partir d'un modèle univarié (normalisation SNV) sont aberrantes, avec des concentrations prédites très éloignées des valeurs mesurées. La source de tels écarts n'a cependant pas encore pu être déterminée. Une limitation de la normalisation SNV est probable mais on notera en plus le fait qu'aucun modèle PLS n'ait pu fonctionner pour le calcium. Lors de notre tentative de modèle PLS pour le calcium, le modèle jugé optimal ne présentait qu'une seule composante, associée aux raies de Ca II à 316-318 nm. Dans de telles conditions, le caractère multivarié du modèle disparaît si bien que l'on retombe sur un modèle univarié.

3) Conclusion de l'étude sur les sols agricoles

- i. Analyse qualitative LIBS
 - a) Identification spectrale

L'utilisation du pistolet Z-300, à l'instar des échantillons de plante, nous a permis de détecter les éléments d'intérêt suivants dans les sols agricoles : Al, C, Ca, Fe, K, Mg, Na, P et Si, dont la concentration dans les sols est proche du g/kg. Le phosphore n'a pu être détecté que pour quelques échantillons seulement. L'azote (N), indispensable pour la nutrition des plantes, n'a pas été détecté du tout. Cet élément a rarement été étudié en LIBS dans les sols agricoles [11] à cause de la complexité de leur matrice et également de sa présence dans l'atmosphère. Un signal LIBS relatif à l'azote peut donc provenir autant de l'échantillon de sol que de l'air ambiant dans lequel se développe le plasma. Deux études uniquement se sont focalisées sur l'observation du signal de l'azote dans les sols agricoles par LIBS [100, 101] et ont utilisé l'argon comme gaz ambiant afin d'éliminer la contribution de l'azote atmosphérique. La première étude [100] a évalué qualitativement l'influence de l'énergie laser et du délai sur la quantité de signal LIBS observée pour deux raies d'émission de N (à 744,2 et 746,8 nm) avec l'aide d'un instrument LIBS de laboratoire pouvant atteindre une énergie laser 4 à 10 fois supérieure à celle du pistolet Z-300. La deuxième étude, menée en 2013 par Lu et al. [101], est consacrée à l'analyse quantitative de l'azote et du phosphore total dans 33 sols agricoles issus d'un même champ agricole chinois, 10 et 7 d'entre eux servant à la construction des modèles d'étalonnage univariés de N et P respectivement, les échantillons de sols restants constituant les jeux de validation. En ce qui concerne l'azote, la prédiction des concentrations des échantillons de validation s'est révélée particulièrement juste, avec un coefficient de corrélation de 0,98 entre les valeurs de concentration prédites par LIBS et les valeurs de référence. En revanche, le modèle pour P a été estimé moins performant et uniquement semi-quantitatif ($r^2 = 0,87$). L'étude laisse donc entrevoir un potentiel de la méthode dans le but de quantifier l'azote total dans les sols, il est cependant important de préciser que les sols étudiés sont issus du même

champ et possèdent donc des propriétés physiques similaires, au contraire des sols étudiés dans le cadre de notre étude.

A l'inverse, les éléments moins concentrés, dits éléments trace, ne sont pas détectés avec le Z-300. C'est le cas de Cd, Cu, Mn et Zn. Il est également très important de noter que ces résultats sont relatifs à l'instrument utilisé, ici un instrument portable, et ne reflètent pas les capacités de détection de la méthode LIBS en général. En effet, les éléments traces ont déjà été détectés à des concentrations similaires dans les sols par le passé par des systèmes LIBS de laboratoire permettant d'atteindre des énergies beaucoup plus élevées [74, 102, 103, 104, 105], et même à des concentrations encore plus faibles ($\mu\text{g/kg}$) notamment à l'aide de procédé d'analyse comme la NE-LIBS [106]. Plusieurs exemples d'instruments LIBS utilisés pour l'analyse d'échantillons de sols sont décrits dans l'article de Peng et al. (2016) [107].

b) Analyse en composantes principales

L'analyse en composantes principales a mis en lumière plusieurs points permettant de mieux comprendre la variabilité des spectres LIBS mais aussi les différences physiques et chimiques entre les pastilles de sols. En effet, à travers l'observation du graphe des loadings il a été possible de déceler la présence d'un décalage en longueur d'onde touchant les spectromètres couvrant les gammes du visible et du proche infrarouge. Il a également été possible de mettre en relation certaines variabilités des données LIBS avec les différences en composition chimique mesurées à l'aide de la méthode ICP-AES. Une distribution similaire des données a pu être obtenue en comparant un graphe des scores issu de l'ACP à partir des données LIBS et un diagramme ternaire chimique construit à partir des données ICP-AES, les deux mettant en jeu des variabilités dans les concentrations en calcium et en fer. Finalement, l'ACP a permis également de signaler des différences importantes sur les valeurs d'intensités touchant la globalité du spectre d'une pastille de sol à une autre. Ce phénomène peut être associé aux effets de matrice présents entre les différents sols, c'est-à-dire une quantité de signal récoltée par l'instrument variable selon la texture du sol analysé. Ce résultat fait directement écho à des travaux récents par Villas-Boas (2016) et al. [81] ou encore Gouegel et al. [82] qui ont, de manière sensiblement différente, établi une corrélation entre le signal LIBS et la texture du sol. Également, le même type de résultat a été observé en 2018 par Ruhlmann et al. [108], à travers une ACP réalisée sur 16 échantillons de sol de référence et 60 échantillons de sols agricoles allemands présentant des matrices différentes selon leurs teneurs en sable, limon et argile. L'ACP calculée sur les deux premières composantes a permis de mettre en lumière plusieurs familles d'échantillons, regroupés selon que leur texture soit à tendance majoritairement argileuse, sableuse ou limoneuse. A l'instar de notre étude, les auteurs précisent cependant qu'il n'est pas possible de réaliser une classification parfaite car le nombre d'échantillons par famille semble trop faible.

ii. Analyse quantitative LIBS
a) Régression univariée

En ce qui concerne l'analyse quantitative, au contraire des échantillons de plantes, la régression univariée s'est révélée insuffisante même avec l'aide d'une stratégie de normalisation. Une absence de corrélation entre le signal LIBS et la concentration mesurée par ICP-AES a en effet été constatée sur l'ensemble des éléments, à l'exception du calcium pour lequel un modèle de calibration univarié a pu être construit après normalisation SNV. Il a été décidé, sur la base de travaux précédents [83, 99], de sélectionner une variable sur une aile de la raie de Ca à 317 nm au lieu de choisir l'intensité nette maximale au centre de la raie, afin d'obtenir un modèle d'étalonnage linéaire exploitable. Cependant, il a été globalement observé que, pour l'analyse LIBS de sols présentant des matrices différentes, il était nécessaire d'opter pour une stratégie multivariée plutôt qu'univariée afin d'obtenir les modèles d'étalonnage les plus performants [108, 109, 110, 111]. L'étalonnage bâti pour le calcium montrant des tendances linéaires particulièrement encourageantes, il a été décidé d'appliquer une stratégie d'analyse multivariée linéaire et plus précisément la PLS, qui est certainement la méthode d'analyse multivariée quantitative la plus couramment utilisée en LIBS [28].

b) Régression PLS

L'analyse multivariée, à travers l'utilisation de la PLS, s'est révélée être indispensable afin de pouvoir construire des modèles d'étalonnage satisfaisants pour plusieurs éléments d'intérêt, en particulier Na, Mg et Fe. Les variables associées aux éléments Na, Mg, Fe, Si et Ca ressortent globalement comme les plus influentes sur les modèles PLS. Ces modèles ont été jugés satisfaisants sur la base de plusieurs critères : i) $0.85 < r^2 < 0.95$ avec des points analytiques bien répartis, ii) une erreur de modèle qui se stabilise après l'ajout d'un certain nombre de composantes et iii) de bonnes prédictions de concentrations pour un nouveau jeu de données de sols agricoles, plus précisément avec des modèles présentant des valeurs d'erreur de prédiction similaires pour les échantillons du jeu de validation (RMSEP) et celles obtenues par validation croisée (RMSECV). Un récapitulatif des performances de ces modèles est présenté dans le tableau 3-3. Les autres éléments n'ont soit pas été détectés (Cu, Zn, Mn) ou que pour de rares échantillons (P), soit aucune corrélation entre les données LIBS et les valeurs de concentration de référence obtenues par ICP-AES, n'a pu être établie (C, Al). Les modèles obtenus sur les sols agricoles avec le pistolet Z-300 dans des conditions optimales peuvent être considérés comme semi-quantitatifs. Ces derniers affichent en effet des valeurs de r^2 d'une valeur de 0,90 environ et ne sont validés que pour les 3 éléments Na, Fe et Mg.

Élément	Na	Fe	Mg	Ca	K	Al	C	N, Cd, Cu, Mn, Zn, P
Détecté								
Quantifié								
Validé								
Min (g/kg)	0.77	6.63	1.38	16.7				
Max (g/kg)	8.9	95.7	11.3	302				
Méd (g/kg)	3.4	24.1	4.35	120				
R ²	0.91	0.88	0.88	0.94				
RMSE (g/kg)	0.77	6.45	0.97	16.8				
RMSECV (g/kg)	0.76	6.30	0.95	16.7				
RMSEP (g/kg)	0.79	4.06	1.04					

Tableau 3-3: Récapitulatif des performances des modèles multivariés construits pour les sols.

Il est particulièrement compliqué d'évaluer les performances de tels modèles, l'analyse quantitative de sols agricoles par LIBS étant complexe et par conséquent peu discutée. Une grande partie des études quantitatives menées sur les sols agricoles par LIBS est focalisée sur la quantification du carbone sous ses différentes formes (totale, inorganique, organique), qui a fait l'objet d'un article de revue [30]. De plus, des articles se sont concentrés sur l'analyse quantitative de trois éléments différents dans les sols agricoles, à savoir Cu [112], Ca [108] et Mg [111]. Cependant, les performances quantitatives présentées dans ces articles restent particulièrement discutables. En effet, le premier propose un modèle quantitatif pour Cu bâti par ANN à l'aide de 59 sols agricoles brésiliens et présentant des textures différentes. Malgré des prédictions de concentrations plutôt correctes, le modèle d'étalonnage construit pour le cuivre est caractérisé par un nuage de points sur une grande partie de la gamme d'étalonnage et une valeur de $r^2 = 0,96$ fortement influencée par un point extrême. Dans une autre mesure, le modèle d'étalonnage construit par PLS pour le calcium afin de prédire la concentration de 60 sols agricoles allemands [108] n'est pas très convaincant, affichant des valeurs de prédictions négatives ou avec un facteur 2 par rapport aux mesures de référence. L'auteur conclut finalement sur l'aspect semi-quantitatif de son modèle d'étalonnage. Yongcheng et al., en 2017 [111], propose un modèle d'étalonnage multivarié non-linéaire pour quantifier le magnésium dans des sols issus de champs de riz alors que son modèle univarié construit avec l'intensité d'une raie de Mg montre clairement des tendances linéaires. Ces exemples servent ici à montrer la difficulté d'établir une stratégie d'analyse LIBS pour les sols, de construire les modèles d'étalonnage et de les valider. Finalement, plusieurs articles se sont également intéressés à l'analyse quantitative des contaminants (Cr, Cd, Cu, Pb) mais dans les sols en général [102, 108].

Les résultats obtenus dans notre étude et les études antérieures montrent que la LIBS, malgré des résultats parfois encourageants, ne semble pas être encore une méthode adaptée à l'analyse des sols, notamment à cause d'effets de matrice très importants et de la difficulté à pouvoir transposer un protocole analytique d'un cas particulier à un autre (instrument LIBS, traitement spectral, stratégie d'analyse). De plus, la méthode LIBS nous permet d'accéder à des corrélations avec la concentration totale en éléments dans les sols alors que la gestion de la fertilité des sols se base plutôt sur la connaissance d'une fraction biodisponible, c'est-à-dire une partie du total de l'élément présent dans un sol qui est susceptible d'être absorbé par les plantes.

c) Comparaison avec une étude similaire

Il existe néanmoins une étude antérieure qui est très similaire à la nôtre et qui a fait l'objet d'une publication très récemment. En effet, en 2020, Erler et al. [110] ont utilisé le même système LIBS portable Z-300 afin de détecter et déterminer la concentration en éléments nutritifs (Ca, K, Mg, Fe, N, P, Mn et Al) de 137 échantillons de sols issus de deux champs présentant des textures différentes. L'article compare aussi des modèles d'étalonnages bâtis pour ces éléments à l'aide de trois méthodes de régression : PLS, LASSO et GPR. Le modèle du calcium montre les meilleures performances avec un r^2 proche de 0,88 (de 0,85 avec la méthode LASSO à 0,90 avec la PLS). Les performances diminuent ensuite pour Mg et K avec des r^2 de 0,79 et 0,64 avec la méthode PLS et surtout la présence de points très dispersés dans le domaine de calibration. Les auteurs ont estimé également pouvoir détecter du signal, très faible, pour des raies élémentaires de N et P et pour des concentrations similaires aux nôtres. Cependant les corrélations affichées entre les valeurs prédites par le modèle PLS et les concentrations de référence ne sont pas convaincantes avec des coefficients de corrélation de 0,51 et 0,14 pour N et P, respectivement. Le modèle d'étalonnage du fer, à l'instar de Mg, semble quant à lui montrer un début de tendance linéaire avec un r^2 de 0,76 mais avec des points très dispersés.

En comparaison avec notre étude, cet article présente les mêmes conclusions, à savoir qu'il est possible de déterminer la concentration en éléments majeurs (Ca, Mg, Fe) dans les sols et que pour les échantillons moins concentrés ou pour les raies moins émissives (N et P), la tâche est bien plus compliquée. Malgré des conclusions semblables aux nôtres, l'étude conduite par Erler et al fournit trop peu d'informations sur les différentes étapes de traitement des données pour permettre une critique approfondie.

d) L'apport de la XRF

En ce qui concerne la campagne de mesures XRF des sols, à l'instar de l'analyse des plantes, elle s'est révélée particulièrement intéressante dans la détection d'*outliers* et complémentaire par rapport aux performances de la LIBS.

Premièrement, la campagne de mesures XRF permet d'observer les mêmes *outliers* qu'en LIBS notamment pour Ca comme le montre la figure 3-24. On peut en effet remarquer deux choses sur ces graphes : i. la présence d'un *outlier* commun aux deux méthodes, l'échantillon 411. Soit cet échantillon diffère des autres par ses propriétés physiques (pastillage, texture, broyage), soit la valeur de concentration du calcium par mesure ICP-AES est peut-être alors à remettre en question et ii. la présence d'un *outlier* en LIBS qui ne l'est pas en XRF, l'échantillon 538. La valeur de référence (ICP-AES) ne semble alors pas aberrante et le 538 pourrait être re-analysé par la LIBS portable.

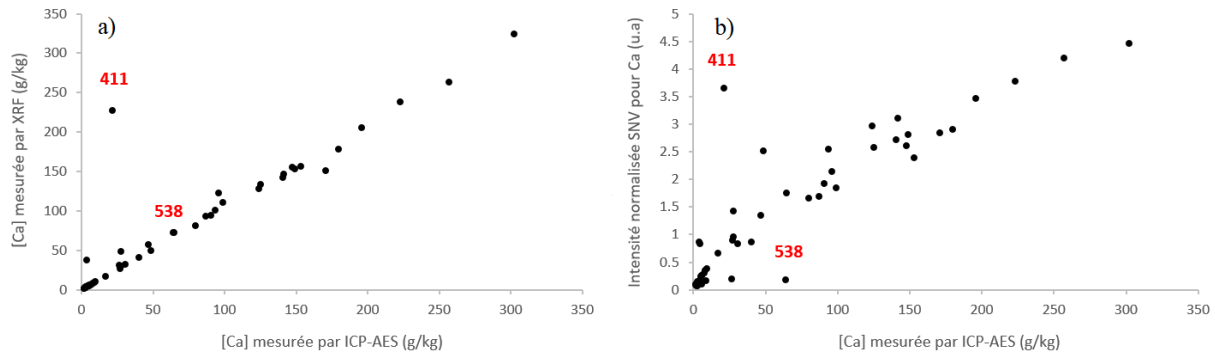


Figure 3-24 : Modèle univarié du calcium pour les pastilles de sols agricoles en utilisant : a) le système XRF portable et b) le pistolet LIBS.

Deuxièmement, les deux méthodes spectroscopiques LIBS et XRF paraissent complémentaires avec donc trois cas de figures :

- Les données LIBS traitées par PLS permettent d'atteindre dans certains cas des performances quantitatives que la XRF n'atteint pas. C'est le cas pour le magnésium pour lequel le comparatif entre XRF et LIBS est présenté sur la figure 3-25. La méthode XRF en général présente en effet des difficultés à mesurer les éléments légers, c'est-à-dire ceux présentant un numéro atomique faible (peut aller jusqu'à $Z=5$ ou $Z=11$ selon l'instrument). Le magnésium, d'un numéro atomique égal à 12, est considéré comme léger, il peut être détecté en XRF mais son signal reste faible. C'est pour cela que pour une concentration < 4 g/kg, quasiment aucun n'échantillon n'a pu être prédit par XRF.

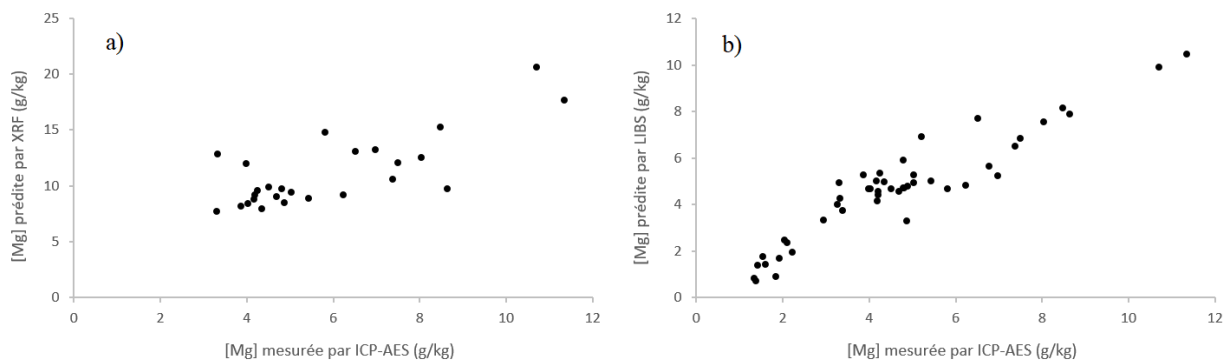


Figure 3-25 : Modèle d'étalonnage du magnésium pour les pastilles de sols agricoles en utilisant : a) le système XRF portable et b) le pistolet LIBS.

- Les deux méthodes LIBS et XRF présentent par ailleurs des modèles d'étalonnage linéaires exploitables et offrant des performances similaires (en sachant qu'on ne peut pas comparer sur les mêmes bases les deux modèles). C'est le cas pour le fer et le calcium dont les modèles de calibration sont présentés sur la figure 3-26.

LIBS portable et sols agricoles : des modèles représentatifs des sols agricoles français ?

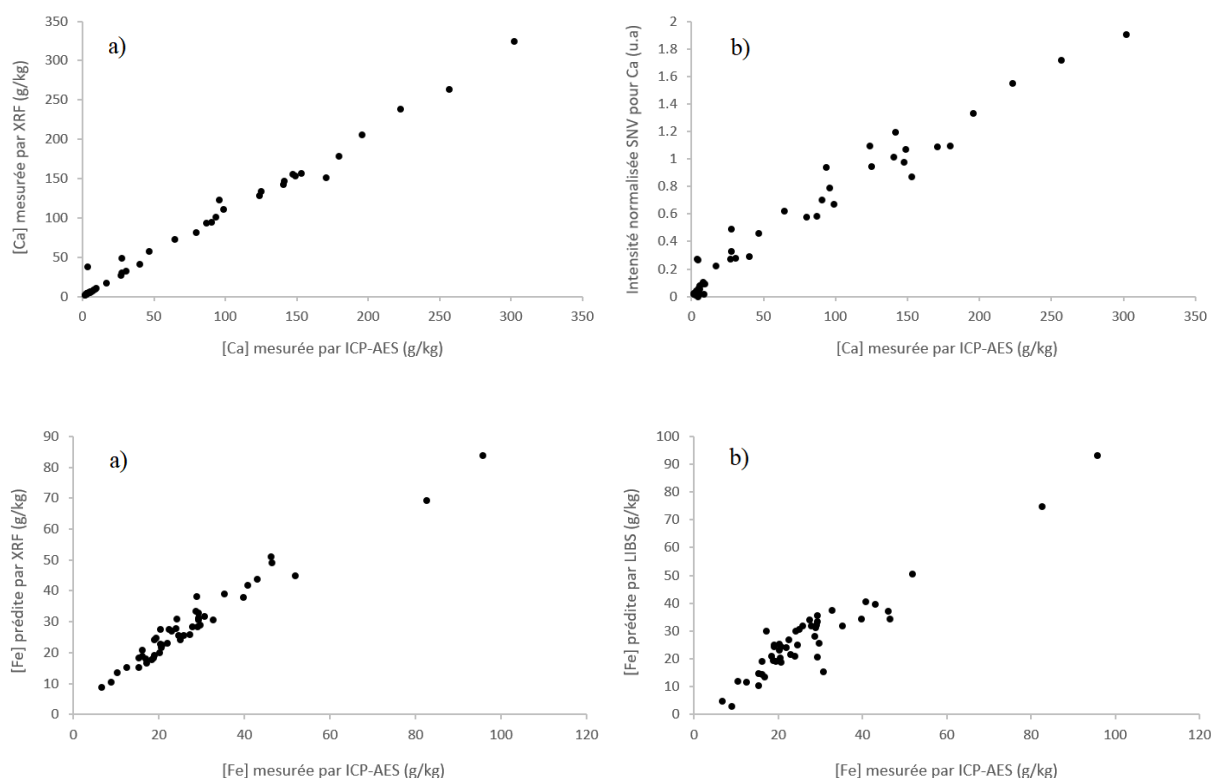


Figure 3-26 : Modèles d'étalonnage du calcium et du fer pour les pastilles de sols agricoles en utilisant : a) le système XRF portable et b) en utilisant la LIBS portable.

- La XRF portable permet de détecter et de quantifier certains éléments qui ne le sont pas en LIBS. C'est le cas de Zn, K (non montré ici) et Mn, comme présenté sur la figure 3-27.

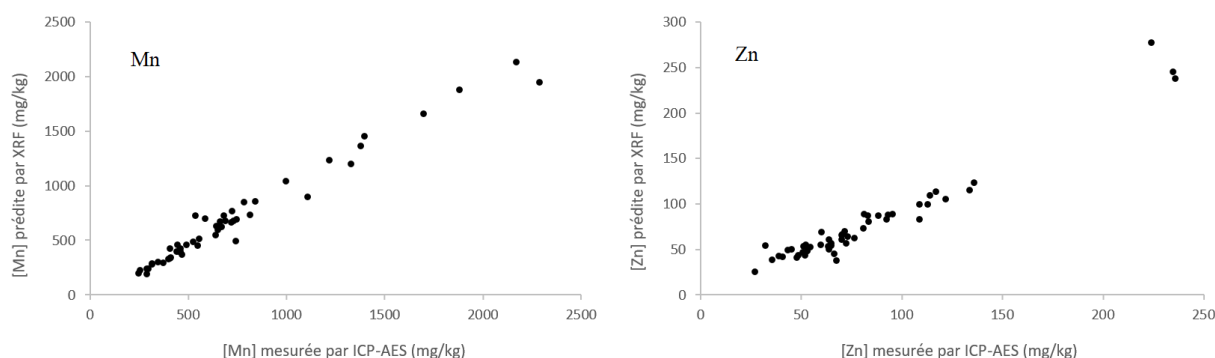


Figure 3-27 : Modèle d'étalonnage XRF de Mn et Zn pour les pastilles de sols agricoles en utilisant le système XRF portable.

Dans une démarche d'analyse multi-élémentaire, la combinaison des deux systèmes portables semble être une solution préférable à l'utilisation d'une seule méthode spectroscopique.

En résumé, les effets de matrice compliquent grandement l'analyse quantitative des sols agricoles et selon l'instrument, ne permettent d'obtenir que des capacités semi-quantitatives par

l'intermédiaire d'un traitement des spectres fin et multivarié. La méthode LIBS, via l'utilisation de systèmes de laboratoire plus performants et une préparation d'échantillon poussée, permet cependant d'atteindre des performances très encourageantes. Le chapitre suivant traite de l'analyse LIBS dans des conditions dégradées afin d'évaluer la possibilité de mettre en œuvre la technique sur le terrain.

CHAPITRE IV

Du laboratoire au terrain ?

La surveillance et l'optimisation des productions végétales est une question d'actualité cruciale [113] afin de pouvoir nourrir une population en constante croissance dans un contexte climatique de plus en plus difficile. Les techniques actuelles permettent une fertilisation uniforme d'un champ agricole alors que les sols présentent généralement des propriétés physiques, chimiques et biologiques hétérogènes [110]. Cette distribution hétérogène peut donc entraîner un sur ou sous-dosage de la fertilisation, néfaste aux productions végétales et à l'environnement. L'agriculture de précision, avec le développement constant de nouvelles technologies permettant l'accès à des informations précises à haut débit sur la variabilité des milieux, permet de moduler les pratiques agricoles selon la variabilité intra-parcellaire. Grâce au pistolet Z-300, il est possible d'analyser des matrices agricoles à la fois en conditions optimales de laboratoire (comme présenté dans les chapitres II et III), et dans des conditions au plus proche du terrain, permettant ainsi d'évaluer la pertinence de la méthode LIBS pour appréhender l'hétérogénéité des sols et des cultures dans le temps et l'espace, en matière de concentrations en éléments nutritifs par exemple. Dans ce chapitre, les capacités du système LIBS portable à réaliser des mesures directes sur le terrain seront évaluées via les résultats issus d'une campagne de mesures *in situ*. Une discussion plus approfondie sera ensuite présentée concernant les paramètres environnementaux les plus influents sur le signal LIBS lors de mesures *in situ*. L'influence de ces paramètres sera évaluée à l'aide d'un plan d'expériences. Finalement, des mesures correctives potentielles seront proposées dans le but de réduire les effets des paramètres les plus pénalisants pour les mesures.

1) Analyse *in situ*

Une campagne de mesures exploratoires LIBS a été réalisée sur des parcelles de blé du site expérimental de QualiAgro à Feucherolles (78) afin d'évaluer les performances analytiques de l'instrument portable Z-300 en conditions réelles. L'analyse sur site a donc consisté en la mise en place du matériel, la récolte des feuilles témoins et des échantillons de sols à analyser, le référencement de ces échantillons, les mesures LIBS et enfin le stockage des échantillons analysés.

i. Protocole expérimental pour l'analyse des plantes

Nous avons analysé 20 parcelles à partir de mesures aussi rapides que possible. Pour chaque parcelle, 10 feuilles de blé ont été prélevées soit un total de 200 feuilles. Les feuilles ont été sélectionnées proches du centre de la parcelle sur les plantes dont le blé commençait à être visible, une photo d'une feuille type est présentée en figure 4-1 en annexe. Deux zones d'analyse de 10 spots laser avec un tir par spot ont été effectuées directement sur chaque feuille de blé fraîche. Un spectre LIBS représentatif de l'ensemble de la parcelle a été calculé à partir de la moyenne de 200 spectres (2 zones x 10 spots x 10 feuilles). Au total, une journée d'analyse sur les feuilles de blé représente l'acquisition de 4000 spectres effectuée par un seul opérateur.

Il convient de préciser que les mesures LIBS ont été réalisées directement sur le site, mais en conditions fixes, c'est-à-dire avec le dispositif portable fixé sur son support posé sur une table directement à proximité des parcelles expérimentales. La photo de l'installation du système en conditions réelles est disponible en figure 4-2 en annexe.

De plus, sur chaque parcelle, la méthode ICP-AES nécessitant une plus grande quantité de matière pour réaliser une analyse, une trentaine de feuilles supplémentaires ont également été prélevées, séchées, broyées, puis analysées par ICP-AES selon les protocoles décrits dans le chapitre I.

ii. Protocole expérimental pour l'analyse des sols

En ce qui concerne les sols, les échantillons ont été prélevés à proximité du centre de chaque parcelle à l'aide d'un cylindre sur une profondeur de 30 cm. Les échantillons de sols ont ensuite été conservés en glacière puis au réfrigérateur à 4°C jusqu'à analyse, cette fois-ci en laboratoire.

En laboratoire, un sous-échantillon a été prélevé pour chaque parcelle et compacté à l'aide d'une spatule dans des boîtes de culture en plastique pour réaliser des analyses sur une surface plane et afin de mimer une analyse rapide dans des conditions de terrain. Deux autres sous-échantillons ont été prélevés puis séchés à 105°C pour déterminer leur taux d'humidité, et leur contenu en éléments totaux par ICP-AES (au laboratoire d'analyse INRAE du LAS), déterminé selon les méthodes décrites en chapitre I.

Deux zones d'analyse LIBS de 25 spots laser avec dix tirs par spot ont été effectuées directement sur chaque sous-échantillon de sol. Un spectre LIBS moyen représentatif de l'ensemble de la parcelle a été calculé à partir de l'obtention de 500 spectres (2 zones x 25 spots x 10 tirs).

iii. Signaux obtenus sur les feuilles de blé intactes

Le résultat le plus remarquable concernant l'analyse LIBS in situ des feuilles fraîches est le signal extrêmement faible et inexploitable pour plus des trois quarts des spectres enregistrés sur

chaque parcelle. Un exemple de signal exploitable (trait rouge) et non exploitable (trait noir) est présenté en figure 4-1. Le spectre exploitable permet d'observer les mêmes éléments significatifs que sur une pastille sèche de plante, à savoir : Al, C, Ca, K, Mg, Na, P et Si ainsi que Ar (associé aux conditions d'analyse), H et O. Certains spectres exploitables peuvent d'ailleurs présenter des intensités LIBS proches de celles obtenues sur des spectres LIBS de pastilles sèches de plantes, suggérant qu'il pourrait être possible de quantifier les éléments directement dans les feuilles fraîches, en utilisant les calibrations établies en laboratoire sur des pastilles sèches (cf chapitre II). Une comparaison spectrale entre une pastille sèche d'un échantillon de pousse de maïs, venant du jeu de données utilisé pour la validation des modèles d'étalonnage bâtis pour les plantes en chapitre II, et d'une feuille fraîche est présentée en figure 4-3 en annexe.

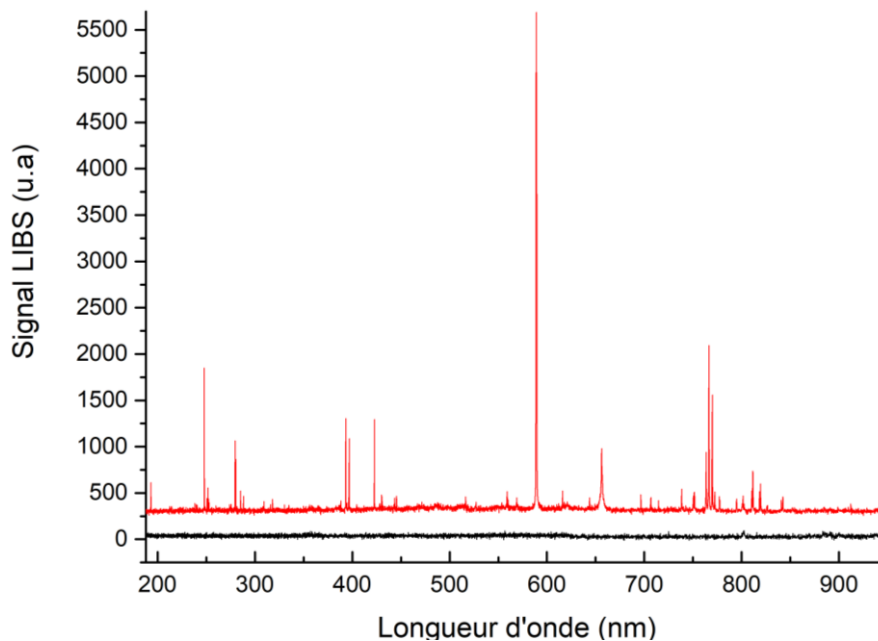


Figure 4-1 : Exemples de spectres LIBS obtenus sur une feuille de blé fraîche prélevée sur une parcelle de Qualiagro.

Après avoir éliminé les spectres pour lesquels il n'y a aucun signal, on constate par ailleurs qu'il y a une très grande variabilité entre les spectres d'un même échantillon, ce qui se traduit par un coefficient de variation (RSD) pouvant dépasser 100%. Il est donc impossible dans de telles conditions, de considérer que le spectre moyen est représentatif de l'échantillon considéré ou de la parcelle. Nous pensons qu'il existe deux causes à une telle variabilité :

- i. Les feuilles de blé sont très fines et sont transpercées au bout de deux tirs laser. Un seul tir laser par spot était donc réalisable afin de ne prendre en compte que le plasma formé sur la surface de la feuille. La figure 4-2 permet de voir les deux zones d'analyse encadrées en rouge. Or, en LIBS, il est courant de réaliser plusieurs tirs

laser sur un même point afin de réduire les fluctuations. Dans le cas présent, avec un seul tir laser par point d'analyse, la faible répétabilité des mesures n'est pas surprenante.



Figure 4-2 : Photo d'une feuille de blé réalisée après analyse LIBS. Les rectangles rouges indiquent les 2 zones d'analyse LIBS effectuées.

- ii. Les feuilles analysées présentaient un taux d'humidité très important (Figure 4-3). En effet, après mesures sur le terrain, puis stockage au réfrigérateur à 4°C (et donc probablement des pertes par évaporation), l'humidité relative mesurée indiquait encore des valeurs moyennes de 40% ; ces valeurs sont probablement déjà très inférieures à l'humidité réelle au moment de la réalisation des mesures LIBS, qui peut avoisiner pour le blé des valeurs plus proches de 60-70% [114]. On peut voir sur cette photo prise avec la caméra du Z300 au contact de la feuille analysée, que la zone d'analyse semble humide en certains points, ce qui est confirmé avec la présence de buée apparaissant après quelques tirs sur la fenêtre de quartz qui sépare l'échantillon du système optique. Ceci a impliqué un nettoyage régulier de cette fenêtre afin de ne pas impacter la bonne collection du signal, mais constitue un frein également à une collecte rapide et optimale des données spectrales.



Figure 4-3 : Photo de la fenêtre de tir lors de l'analyse LIBS des feuilles de blé avec le Z-300.

A partir des spectres LIBS enregistrés et présentant des fluctuations importantes, il n'est donc pas possible de calculer un spectre moyen qui soit représentatif de chaque parcelle. Et par conséquent il n'est pas possible d'exploiter les modèles de calibration préalablement bâtis à partir d'échantillons de plantes préparés sous forme de pastilles pour prédire la concentration en éléments totaux des feuilles fraîches.

iv. Signaux obtenus sur les sols frais

Les résultats obtenus sur les sols frais sont également difficilement exploitables et ce pour plusieurs raisons :

- Les sols frais présentaient un taux d'humidité moyen de 18%, générant des spectres d'intensités environ deux fois plus faibles que pour des pastilles sèches. Cependant, un signal LIBS a pu être obtenu pour chaque parcelle et plus précisément pour chaque zone d'analyse, permettant de pouvoir détecter les mêmes raies d'émission que pour une pastille de sol sèche. Il peut alors se poser la question, à l'instar des plantes, de pouvoir quantifier la composition en éléments totaux des sols frais à l'aide des modèles d'étalonnage construits à partir des pastilles sèches en chapitre III.
- Cependant, les analyses LIBS en laboratoire conduites sur des échantillons de sols prélevés *in situ* et compactés à l'aide d'une spatule, dans un contexte de mesures rapides, ont généré, comme cela a déjà été montré dans des travaux précédents [105], l'obtention d'intensités très fluctuantes d'un point à un autre, avec notamment des RSD atteignant 30 à 75%.
- Enfin, s'il était possible d'obtenir des valeurs quantitatives fiables à partir de spectres obtenus directement sur des sols frais, il aurait pu être envisageable d'utiliser ces résultats pour évaluer l'effet de différents apports selon les parcelles expérimentales. Or seuls les apports de C différaient significativement, selon les mesures réalisées par ICP. Dans le cadre de notre étude, aucun modèle quantitatif n'ayant pu être bâti en laboratoire pour le carbone sur les pastilles sèches de sols agricoles (chapitre III), une comparaison inter-parcelle, via la prédiction des concentrations en carbone des sols frais par des modèles bâtis en laboratoire, n'est alors pas possible. En ce qui concerne les éléments dont les modèles ont été bâtis avec succès en laboratoire, leurs concentrations ne présentaient que très peu de variabilité d'une parcelle à une autre ($2,5 \text{ g/kg} < [\text{Mg}] < 3,1 \text{ g/kg}$; $16,7 \text{ g/kg} < [\text{Fe}] < 18,9 \text{ g/kg}$; $6,8 \text{ g/kg} < [\text{Na}] < 7,2 \text{ g/kg}$).

En résumé, en dehors de l'observation de signal pour les feuilles de blé fraîches et les sols frais, les résultats issus de la campagne de mesures *in situ* se sont montrés difficilement exploitables. En effet, les mesures ne se sont pas montrées suffisamment répétables, aussi bien pour les feuilles de blé, pour lesquelles il n'était pas possible de faire plus d'un tir par point, que pour les sols, peu compacts du fait de la rapidité de leur préparation. Pour les deux types d'échantillons, la présence d'un taux d'humidité conséquent a considérablement réduit l'intensité globale des spectres LIBS.

Finalement, afin de bien comprendre les obstacles à l'analyse *in situ*, nous avons décidé d'étudier en laboratoire l'influence de la préparation des échantillons de sols sur le signal LIBS.

2) Influence de facteurs extérieurs sur le signal LIBS d'un échantillon de sol

Il est nécessaire de comprendre de manière rigoureuse quelle est l'influence de divers paramètres environnementaux pour pouvoir obtenir des résultats fiables lors d'analyses conduites sur le terrain. Pour cela, il a été choisi de conduire un plan d'expériences sur des pastilles de sols préparées en laboratoire, testant simultanément l'influence de plusieurs facteurs (humidité, taille des grains, compaction) susceptibles d'avoir un effet sur les signaux LIBS lors de mesures in situ. Ce type d'expérimentation, à notre connaissance, n'a encore jamais été conduit jusqu'à présent.

i. Principe du plan d'expérience

Le plan d'expérience est un outil de chimiométrie, couramment utilisé en chimie analytique [115] dans le but d'évaluer l'influence de plusieurs paramètres expérimentaux, appelés facteurs, sur le comportement d'une réponse, généralement un signal analytique, à l'aide d'un minimum de mesures. Les plans d'expériences peuvent être classés en deux catégories : i. les plans de criblage et ii. les méthodes de surfaces de réponses [116]. Les plans de criblage permettent de réaliser un nombre limité de mesures mais optimal, tout en maximisant l'information à extraire afin d'évaluer quels sont les facteurs ayant de l'influence sur le comportement d'une réponse. Une fois que les facteurs les plus influents sont déterminés et conservés, il est possible, à travers l'utilisation de surfaces de réponse, de déterminer les conditions expérimentales qui offrent les meilleures performances en terme de réponse analytique, selon les besoins de l'analyste. Cette méthode est particulièrement utilisée en chromatographie en phase liquide à haute performance (HPLC) qui est une technique de séparation dont l'efficacité dépend d'un nombre conséquent de facteurs (nature de la phase mobile, polarité, débit, pH, température de la colonne...) [117]. L'utilisation des plans d'expériences paraît alors essentielle. Les plans d'expériences ont également été utilisés en LIBS dans plusieurs études afin d'optimiser différents paramètres instrumentaux comme le délai, la taille du spot laser, l'énergie du laser ou encore la compaction dans le cadre d'analyses de pastilles [118, 119, 120, 121].

Ici, seule une analyse des facteurs les plus influents a été conduite, car il est déjà démontré en LIBS que la préparation de l'échantillon (c'est-à-dire le séchage, broyage et pastillage), permet d'obtenir un signal plus intense et répétable [11] pour l'analyse de sols.

ii. Réalisation d'un plan factoriel complet à 3 facteurs à 2 niveaux

Les sols en conditions réelles peuvent présenter des taux d'humidité variables ([50], figure 4-4), des hétérogénéités [122], différents types de textures [81, 82] et de niveaux de compaction [123]. Dans le but de tester l'influence que certains de ces paramètres peuvent avoir sur les mesures LIBS, nous avons choisi de conduire un plan d'expériences pour 3 facteurs sur des pastilles de sol en conditions de laboratoire. Trois facteurs ont été sélectionnés : i. le taux

d'humidité du sol, ii. la taille du grain (c'est-à-dire le diamètre maximal des particules de sol) et iii. le poids de pression appliqué pour former la pastille de sol.

Le choix du plan d'expérience le plus approprié se base en grande partie sur le nombre de facteurs à évaluer. Lorsque le nombre de facteurs est plus important, il est intéressant d'utiliser un plan factoriel fractionnaire [115, 124], qui permet de limiter le nombre total d'expériences conduites ou encore un plan de Box-Behnken comme proposé dans la partie I.3.iv. *Analyse LIBS sur pastille : un protocole établi par plan d'expérience*. Cependant, dans notre cas particulier, où nous souhaitons tester un petit nombre de facteurs (3) et où il est compliqué de proposer plus de deux modalités par facteur d'un point de vue pratique, il est assez facile de réaliser un plan d'expériences fiable et simple : le plan factoriel complet à deux modalités. Il convient également de préciser que l'utilisation de ce plan d'expérience, dans notre contexte d'étude, ne vise pas à déterminer finement comment un facteur peut faire varier la réponse expérimentale mais plutôt à comprendre rapidement si tel facteur est influent ou non sur le signal LIBS.

a) Sélection de l'échantillon de sol

Le plan d'expériences a été réalisé sur des pastilles de 350 mg du sol 538 du réseau QUASAPROVE. Ce sol est issu d'une parcelle de blé du Nord-Est de la France et présente une texture composée à 50% d'argile, 13% de sables et 37% de limons et une valeur de pH de 8,22 (il est également représenté sur le diagramme ternaire en Fig 3.1 du chapitre III). Le sol 538 est représentatif d'un sol français de grande culture et présente également des hétérogénéités importantes visibles à l'œil nu lorsqu'il est non broyé. La pastille est visible sur la présentation visuelle du facteur 2 en figure 4-5.

b) Sélection des modalités du plan d'expériences

Pour chaque facteur, deux modalités extrêmes (qui seront codées en -1 et +1 dans la représentation du plan d'expériences, voir tableau 4-1) ont été choisies.

Facteur 1 : Humidité



Figure 4-4 : photo d'un même sol directement prélevé du terrain dans deux zones différentes, avec deux taux d'humidité différents.

La figure 4-4 sert essentiellement à illustrer la différence visible à l'œil nu entre un même sol contenant deux taux d'humidité significativement différents. Elle ne représente pas les taux d'humidité appliqués au sol 538 pour le plan d'expériences. En effet, deux modalités extrêmes de 2% (-1) et 10% (+1) ont été choisies car certains tests préliminaires ont permis d'estimer que la valeur de 10% représentait le taux d'humidité maximal qui permettait d'analyser la pastille de sol au contact sans la détruire. Le niveau d'humidité le plus bas (2%) représente l'humidité résiduelle contenue dans la pastille de sol après la procédure de séchage (24 heures à une température de 40°C). Le taux d'humidité de 10% a été obtenu après ajout de 35 μL d'eau pure à la surface de la pastille via une micropipette jaugée.

Facteur 2 : taille du grain

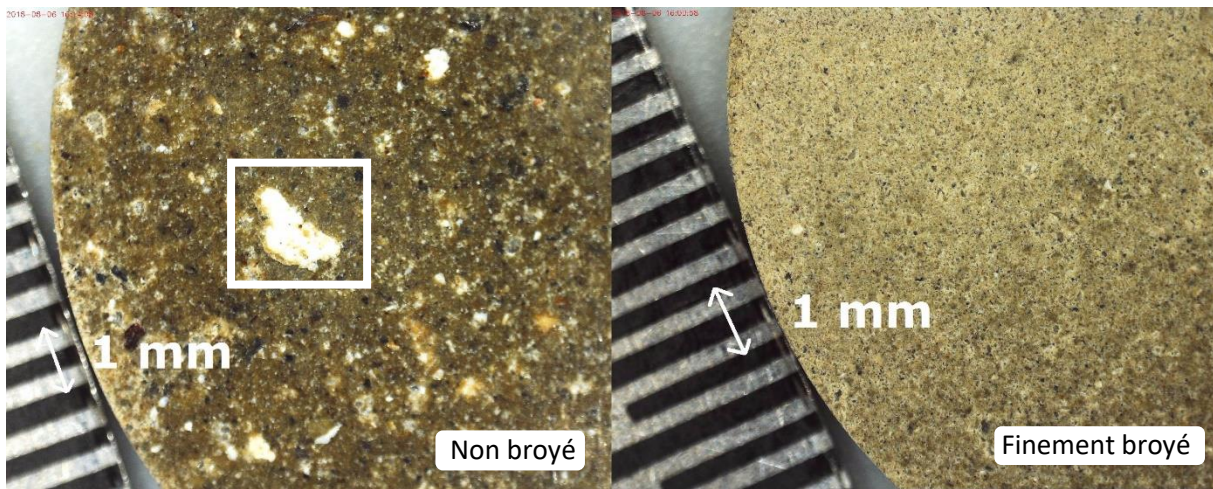


Figure 4-5 : vue de la surface de la pastille du sol 538 après une étape de tamisage à 2 mm) à gauche et finement broyée (< 250 μm) à droite. Encadré blanc : particule minérale de type feldspath non broyée.

En ce qui concerne la taille du grain, la première modalité choisie correspond à une taille maximale de 2 mm (modalité +1) : elle correspond à la maille du tamis conventionnellement utilisée dans les procédures de tamisage pour les sols agricoles. La deuxième modalité (-1) a été obtenue par broyage avec un broyeur planétaire à billes pour atteindre une taille maximale de grain de 250 microns. La pastille du sol 538 sans broyage et après une étape de broyage peut être observée en figure 4-5.

Facteur 3 : compaction

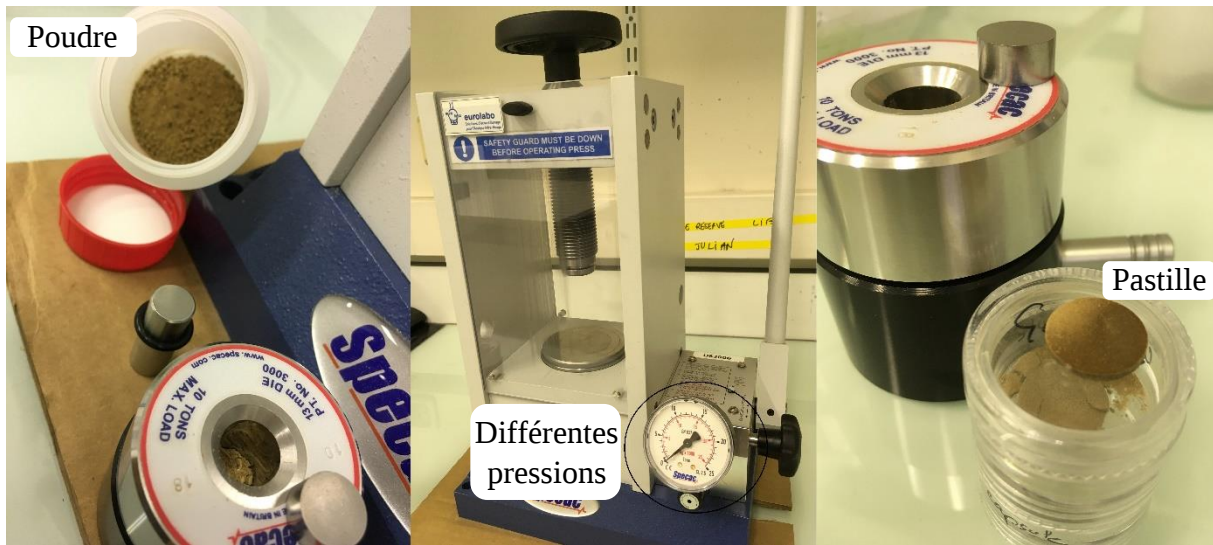


Figure 4-6 : photo représentant le passage de la poudre de sol 538 à la pastille à l'aide de la presse manuelle.

Les deux modalités choisies sont les suivantes : 2t de pression pour la modalité basse (la pastille de sol a été estimée trop fragile en dessous de cette valeur), et 8 t pour la modalité haute (valeur la plus proche du poids maximum que peut appliquer la presse manuelle). La figure 4-6 permet d'illustrer le passage de la poudre de sol à la pastille à l'aide de la presse manuelle.

Le plan d'expérience est alors représenté sous la forme d'un tableau contenant des variables codées (X1, X2, X3) pour chaque facteur sélectionné, avec ici deux niveaux (-1 et +1) par facteur (Tableau 4-1). Ces variables codées délimitent le domaine expérimental du plan d'expérience et sont ensuite transformées par l'analyste en valeurs réelles comme indiqué dans la moitié droite du tableau. Une ligne du tableau représente donc une combinaison de facteurs qui doit être réalisée et mesurée par LIBS. Par exemple, la première ligne du tableau est caractérisée par une pastille de sol construite avec un poids de pression de 2 tonnes, une taille de grain de 2 mm et un taux d'humidité de 10%. Cette pastille doit donc être construite selon ces modalités et analysée par LIBS.

facteurs	Matrice du plan			Plan d'expérience			
	X1	X2	X3	Humidité (%)	Taille du grain (mm)	Poids de pression (tonnes)	
x3	-1	-1	-1	2	0.25	2	35 µL d'eau ajouté à la surface
	+1	-1	-1	10	0.25	2	
	-1	+1	-1	2	2	2	
	+1	+1	-1	10	2	2	
	-1	-1	+1	2	0.25	8	Humidité résiduelle
	+1	-1	+1	10	0.25	8	
	-1	+1	+1	2	2	8	
	+1	+1	+1	10	2	8	

Tableau 4-1 : Plan factoriel complet défini pour l'étude de l'influence de paramètres extérieurs sur le signal LIBS.

c) Campagnes de mesures LIBS

Huit pastilles du même échantillon de sol combinant les différents facteurs et modalités ont été réalisées, correspondant aux huit lignes du tableau 4-1. Pour chaque cas, trois réplicats ont été réalisés afin d'observer la variance sur les résultats obtenus.

Sur chaque pastille, les analyses LIBS ont été réalisées sur 3 zones comprenant chacune 15 spots laser avec 5 tirs laser par spot. Le spectre LIBS moyen d'une pastille correspond à la moyenne de ces 225 spectres individuels.

iii. Résultats

Une fois les facteurs, le plan d'expérience et l'échantillon-test sélectionnés puis les mesures LIBS réalisées, il est essentiel de choisir la réponse analytique la plus adaptée. Le signal LIBS est un spectre de raies d'émission couvrant une très large gamme spectrale. La valeur de l'aire sous le spectre sur l'ensemble de la gamme spectrale a d'abord été testée comme réponse mais nous avons constaté qu'elle était très fortement influencée par les nombreuses raies du calcium contenues dans le spectre. La réponse analytique était donc représentative du calcium et non du signal global. Nous avons donc préféré comme réponse au plan d'expérience l'intensité nette de chaque élément détecté par le pistolet Z-300, à savoir la raie de Al I à 309,3 nm, de C I à 193,1 nm, de Ca II à 315,9 nm, de Fe I à 302,1 nm, de K I à 769,9 nm, de Mg I à 285,2 nm et de Si I à 288,2 nm.

a) Graphe des effets

Pour déterminer des profils de réponse pour chaque élément, nous avons observé la valeur moyenne de l'intensité du signal pour chaque modalité de facteur sur le graphe des effets pour chaque élément ou groupes d'éléments (figures 4-7 et 4-8). C'est-à-dire que pour chaque modalité -1 et +1 de chaque facteur a été calculée une moyenne des valeurs d'intensité nette de la raie sélectionnée pour les spectres enregistrés selon cette modalité. Les valeurs moyennes obtenues pour chaque modalité sont ensuite reliées par une droite : plus la pente de cette droite est grande, plus l'influence du facteur sur le signal est importante.

Deux profils de réponse ont été obtenus selon les éléments observés et sont présentés en figure 4-7 et 4-8.

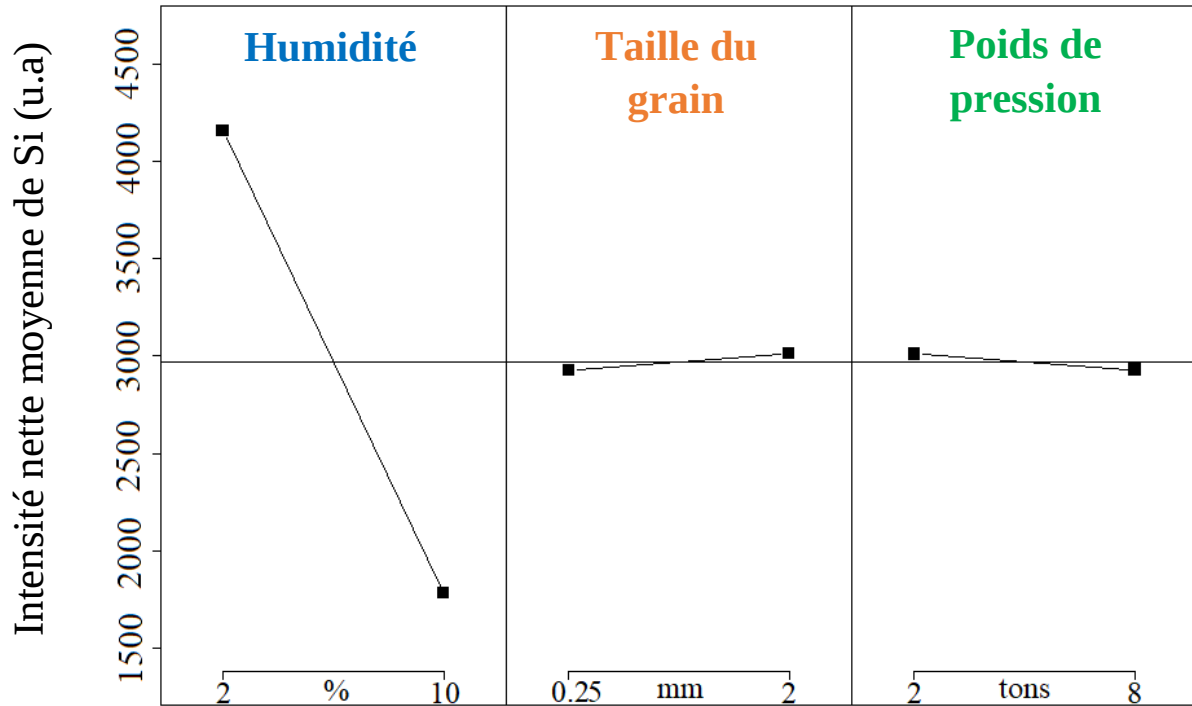


Figure 4-7 : Graphe des effets pour le silicium (Si I à 288,2 nm) obtenu à la suite du plan d'expériences réalisé sur le sol 538.

Le profil 1, illustré par la figure 4-7 a été obtenu pour les raies du silicium, de l'aluminium et du fer. La compaction et la taille de grain n'ont dans ce cas aucun effet sur le signal, au contraire de l'humidité, qui présente une très forte influence. On retrouve le fait que le signal LIBS diminue fortement lorsque le taux d'humidité dans le sol augmente. Ici, pour le silicium, lorsque la pastille est sèche (2%) la valeur de l'intensité nette de la raie à 288 nm est de 4200 alors que lorsqu'elle est humide (10%) elle est d'environ 1700. En revanche, lorsque l'on passe de la modalité basse à la modalité haute pour les deux autres facteurs, on observe une droite quasi horizontale signifiant que l'intensité nette moyenne est sensiblement la même d'une condition à l'autre. Cependant un deuxième type de profil, pour les éléments C, Ca, K et Mg, a pu être observé. Les résultats expérimentaux pour la raie du silicium à 288,2 nm et la raie du calcium à 315,9 nm, obtenus à partir de ce plan d'expérience, sont donnés par les tableaux 4-1 et 4-2 en annexe.

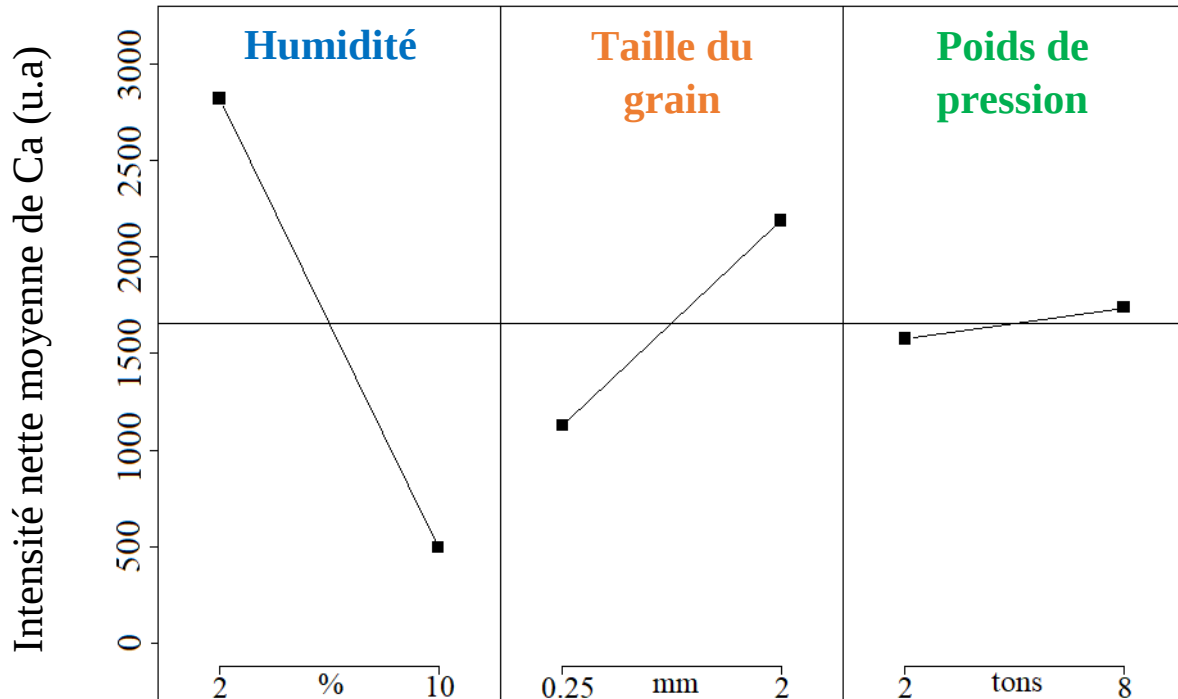


Figure 4-8 : Graphe des effets pour le calcium obtenu à la suite du plan d'expériences réalisé sur le sol 538.

Un deuxième type de profil (Figure 4-8), a été obtenu pour les éléments C, Ca, K et Mg. L'humidité présente toujours une très forte influence sur le signal moyen : en effet, l'intensité nette moyenne de la raie de Ca à 315 nm diminue de 2800 à 500 lorsque l'on humidifie la pastille. Le facteur lié à la compaction du sol n'a toujours aucun impact. En revanche, pour ce second type de profil, la taille de grain a une influence significative sur le signal. Le signal augmente d'une intensité moyenne d'environ 1000 à environ 2200 (soit un peu plus d'un facteur 2), lorsque l'on augmente la taille du grain de 0,25 à 2 mm. Cependant, si ce phénomène est observé uniquement avec ce résultat, il peut paraître contre-intuitif par rapport à de nombreuses études menées en LIBS démontrant l'obtention de signaux plus intenses lorsque le grain est plus fin [82], il convient alors de préciser ce que représente réellement le facteur taille de grain pour nos données. On peut expliquer un tel résultat par le mode de préparation de l'échantillon. En effet, pour la modalité -1 en taille de grain, le sol est broyé finement de sorte qu'aucun grain présentant un diamètre > 250 microns ne puisse être présent. En ce qui concerne la modalité +1, après un tamisage du sol à 2 mm, puisqu'aucune étape de broyage n'est mise en jeu, il est possible que la pastille contienne des grains présentant un diamètre allant jusqu'à 2 mm.

La figure 4-5 permet d'observer les différences entre la pastille d'un même échantillon de sol construite après une étape de broyage fin, dans la partie droite de la figure (modalité -1 de la taille de grain), ou directement après l'étape de tamisage à 2 mm et donc sans étape de broyage intermédiaire (partie gauche de la figure 4-5, modalité +1 de la taille de grain). La pastille non broyée présente un aspect beaucoup plus hétérogène avec la présence d'agrégats ou de particules minérales plus grossières (fig. 4-5, encadré blanc), qui sont difficiles à éviter lors de l'analyse LIBS.

La figure 4-9 permet de faire la comparaison entre un spectre moyen issu d'une zone d'analyse comprenant un agrégat riche en minéral (trait pointillé rouge), tel que celui encadré en figure 4-5, et un spectre moyen issu d'une zone d'analyse dite « neutre », c'est-à-dire ne présentant pas d'hétérogénéité visible à l'œil nu. Les spectres sont présentés uniquement dans la gamme entre 273-321 nm afin de pouvoir comparer facilement l'intensité des raies d'émission associées à Mg, Si, Al et Ca. On observe que toutes les raies d'émission ne sont pas impactées de la même façon : i. l'intensité des raies de Si et Al sont similaires que l'on analyse un agrégat ou une zone neutre, ii. l'intensité des raies de Mg et Ca sont bien plus élevées lorsque l'on analyse un agrégat. Ce phénomène ne peut pas être de nature instrumentale car il aurait affecté tout le spectre et non des raies élémentaires spécifiques. Cela signifie que les agrégats analysés sont riches en Ca et Mg. L'observation du reste du spectre (non présenté ici) montre que ces agrégats sont également riches en K et C.

Par conséquent, c'est la présence d'agrégats riches en Mg, Ca, C et K qui est à l'origine des deux profils de réponse différents observés dans le plan d'expériences. Ce type d'agrégat n'a été visible que pour des sols présentant une texture à tendance argileuse. Aucune hétérogénéité aussi importante n'a été observée, à l'œil nu, sur des sols de textures majoritairement limoneuses ou sableuses.

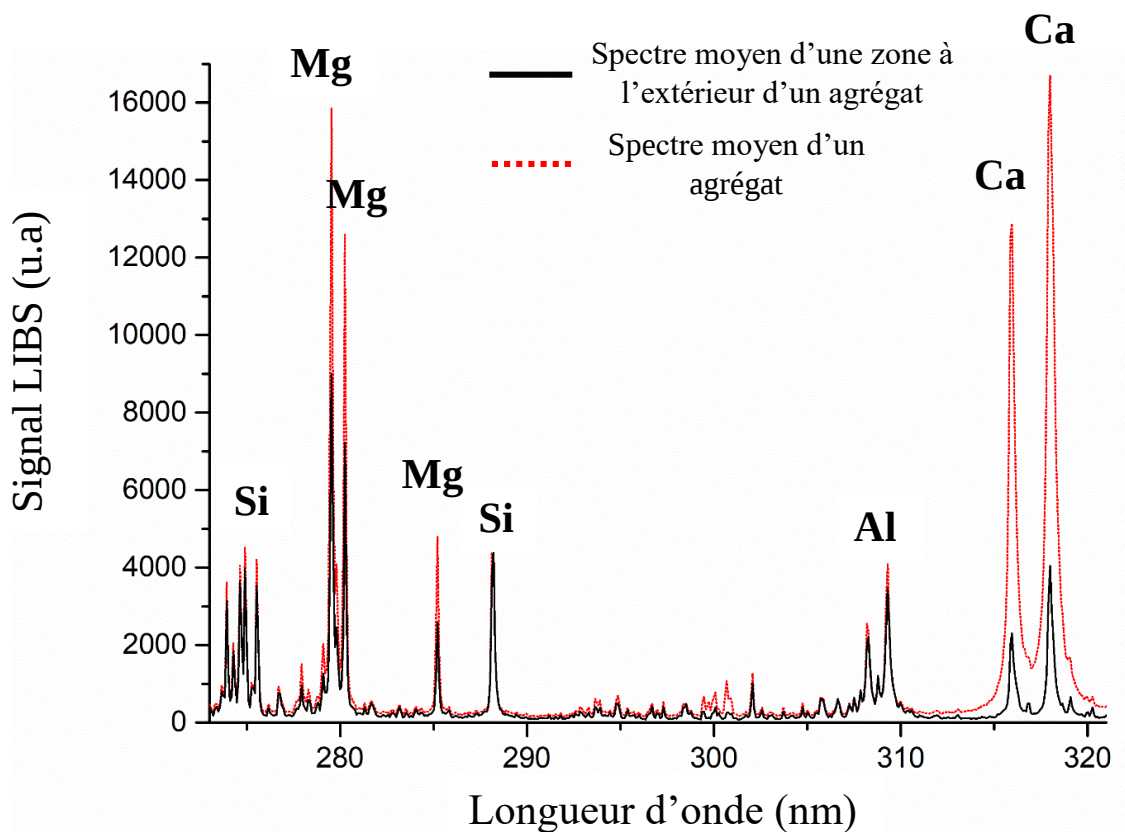


Figure 4-9 : Comparaison entre le spectre moyen d'une zone de la surface de la pastille avec agrégat (trait pointillé rouge) et le spectre moyen d'une zone en dehors d'un agrégat (trait noir).

La présence d'argiles et de matière organique peut en effet favoriser l'agrégation, avec des macro-agrégats ($> 250\mu\text{m}$) et des micro-agrégats ($< 250\mu\text{m}$), plus ou moins riches localement en éléments minéraux [125]. A l'échelle d'un spot LIBS (ici $50\ \mu\text{m}$), il peut donc être particulièrement difficile d'obtenir une mesure représentative du sol, d'autant plus s'il n'est pas broyé au préalable.

b) Graphe des interactions

Le plan factoriel complet permet également d'observer s'il existe des interactions entre les facteurs, c'est-à-dire si la baisse ou l'augmentation du signal est amplifiée lorsque deux facteurs sont combinés. Le graphique des interactions pour le silicium est donné en figure 4-10. et représente l'évolution de l'intensité nette moyenne de la raie observée en fonction des modalités -1 et +1 de deux facteurs. Par exemple la case numérotée 4 centrale de la première ligne permet d'observer l'interaction entre l'humidité et la taille du grain. La ligne pointillée rouge correspond aux intensités mesurées sur des pastilles à 2% d'humidité, et la ligne noire correspond aux intensités mesurées à 10% d'humidité ; ces deux lignes sont tracées à partir des deux modalités de taille du grain (0,25 et 2mm) en abscisse et sont parallèles. Cela signifie que la raie de Si évolue de la même façon lorsque l'on passe d'une pastille finement broyée à une pastille non broyée, que la pastille soit sèche ou humide. Une interaction forte entre facteurs aurait été symbolisée par deux droites qui se coupent. En ce qui concerne nos données, on remarque que l'ensemble des droites rouges et noires sont parallèles ou presque confondues et donc qu'il y a absence d'interaction entre les facteurs. Ce résultat illustré dans le cas du silicium en figure 4-10 a aussi été observé pour tous les autres éléments.

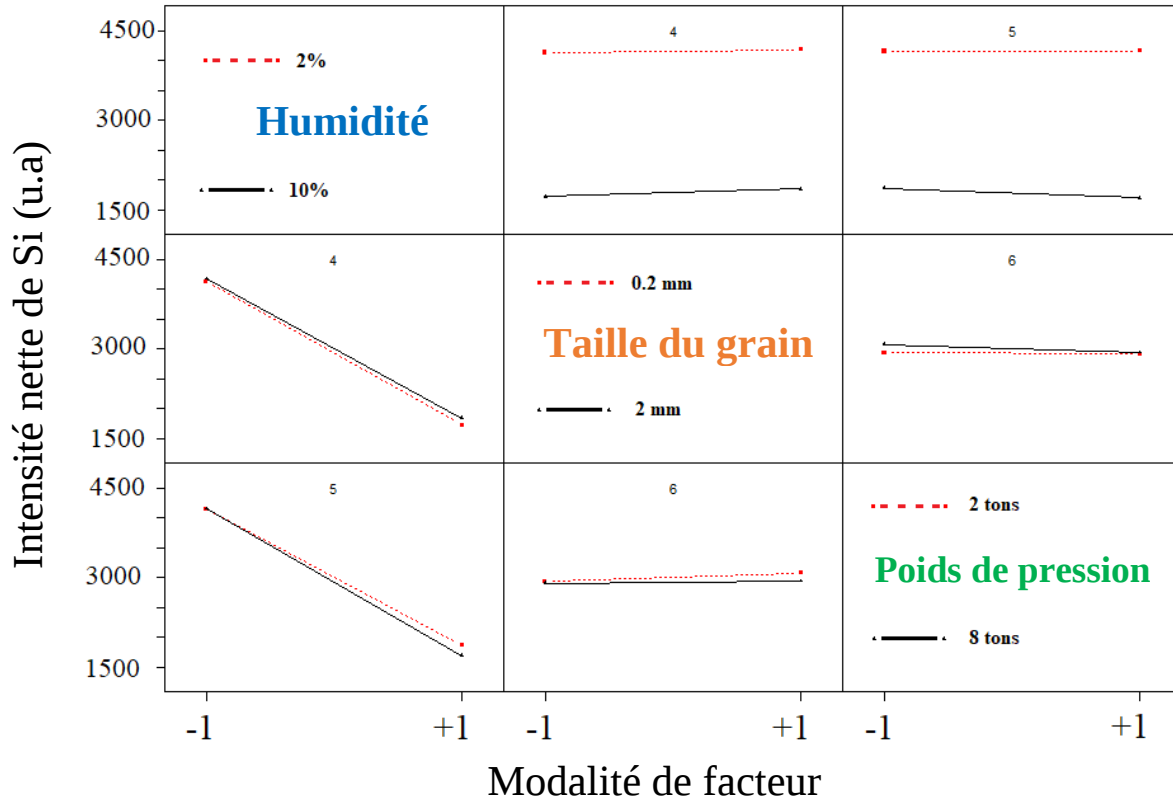


Figure 4-10 : Graphique des interactions pour le silicium obtenu à la suite du plan d'expérience réalisé sur le sol 538.

c) Validation statistique : ANOVA

Afin de valider ces résultats d'un point de vue statistique, nous avons réalisé une analyse de variance (ANOVA) [115]. L'intérêt d'une ANOVA est de s'assurer que l'influence d'un facteur sur la réponse est significative, en faisant la supposition que la variation de la réponse en fonction du changement du facteur est linéaire. Une régression multilinéaire pour 3 facteurs est alors réalisée à partir des données et de l'équation 4-1 :

$$\text{Equation 4-1} \quad y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \varepsilon$$

Avec y la réponse, ici l'intensité nette d'une raie élémentaire ; X_1 , X_2 , X_3 , les variables prédictives donc dans notre contexte les différents facteurs (humidité, taille du grain et poids de pression) qui prennent une valeur de modalité haute ou basse selon le point expérimental ; a_1 , a_2 , a_3 , les coefficients de régression et a_0 la valeur de y lorsque toutes les variables prédictives sont égales à 0. ε représente l'erreur aléatoire du modèle, appelée également résidu. Chaque point expérimental permet d'obtenir une valeur de la réponse, les coefficients sont les inconnues qu'il faut déterminer. L'interaction des facteurs n'a pas été prise en compte dans l'équation 4-1 car une ANOVA calculée en la prenant en compte permet d'obtenir les mêmes conclusions.

En résumé, pour que les résultats d'une telle régression puissent être interprétés, il convient de vérifier trois conditions par rapport aux résidus du modèle qui doivent : i. être indépendants, ii. suivre une distribution normale et iii. être homogènes. Ces conditions peuvent être vérifiées graphiquement et à l'aide de tests statistiques. Le résidu représente l'écart entre la valeur

observée via la mesure analytique (y_{mes}), ici la valeur d'intensité de la raie observée, et la valeur (y_{reg}) cette fois estimée par la régression 4-1.

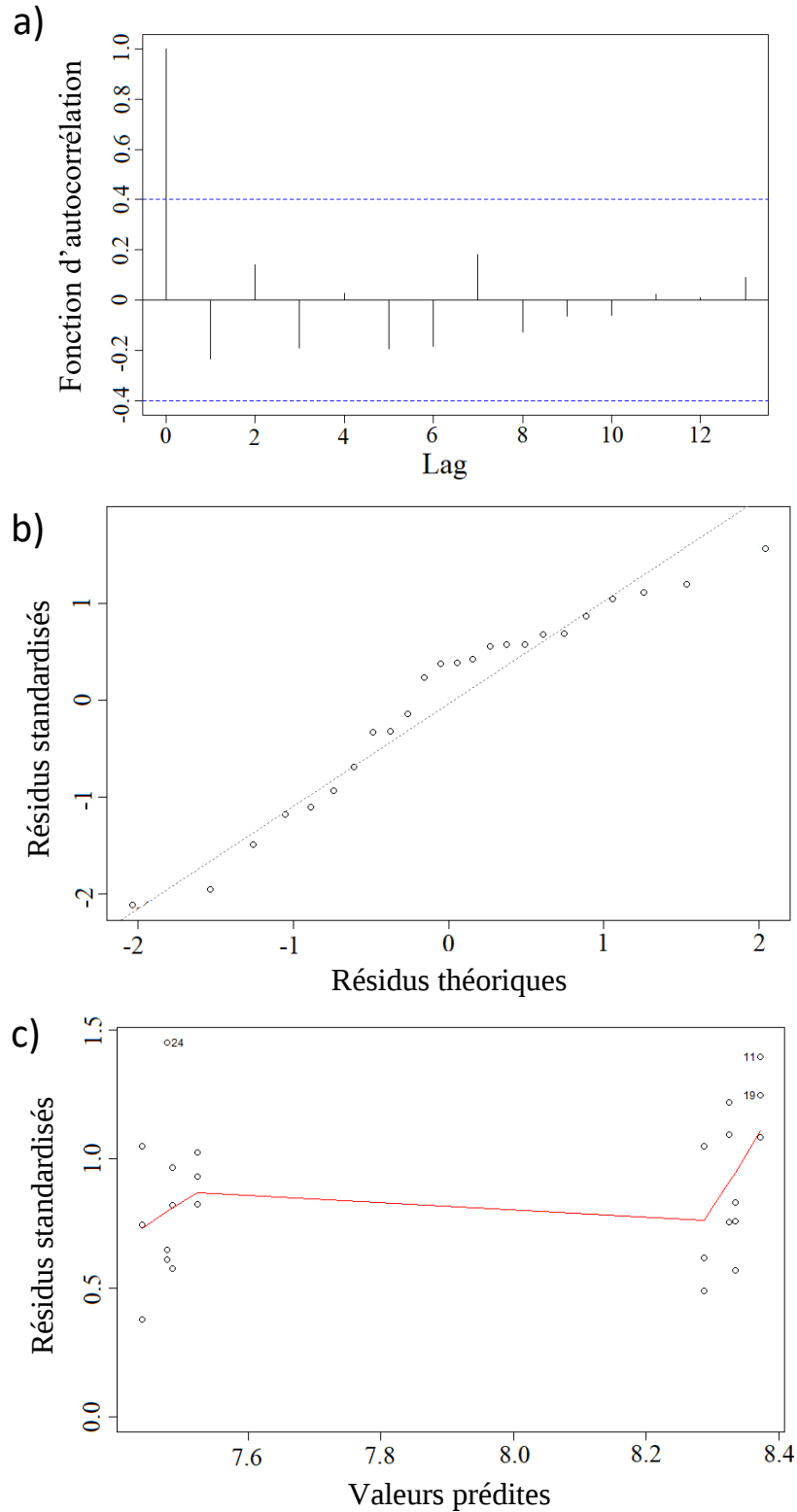


Figure 4-11 : Graphiques utilisés pour évaluer la pertinence des résultats issus du plan d'expériences calculé sur l'intensité de la raie du silicium (288,2 nm) avec a) le graphe d'auto-corrélation, b) le Q-Q plot et c) les valeurs standardisées vs les valeurs prédites.

Ainsi, trois graphiques présentés en figure 4-11 permettent respectivement d'évaluer l'indépendance, la normalité et l'homogénéité des résidus du modèle décrit par l'équation 4-1 et calculé pour l'intensité de la raie du silicium à 288 nm. Le graphe d'autocorrélation (Figure 4-11a.), permet d'évaluer si les résidus sont corrélés entre eux. En effet, la distribution des résidus étant supposée être aléatoire, une corrélation entre un résidu et un autre peut donc révéler une anomalie associée aux conditions expérimentales. Le graphe d'autocorrélation permet donc de mettre en lumière si un résidu est corrélé avec le suivant ($\text{lag}=1$) ou avec les prochains ($\text{lag}=2$ pour la mesure +2, $\text{lag}=3$ pour la mesure +3...). La ligne bleue en pointillés permet de représenter le seuil à partir duquel on considère qu'il y a présence d'une corrélation entre résidus, en dessous de ce seuil il y a absence de corrélation. La corrélation maximale de 1.0 pour un lag égal à 0 représente tout simplement la corrélation du résidu avec lui-même. Ici, dans le cas de la raie de Si, aucune corrélation entre résidus n'est révélée. Les résidus peuvent donc être considérés comme indépendants. Notons qu'un test de Durbin-Watson permet de confirmer cette conclusion.

La figure 4-11b. représente le Q-Q plot utilisé pour vérifier si les résidus suivent une distribution normale. Plus précisément, ce graphe permet de comparer les résidus issus du modèle de régression (points sur la figure) avec ceux attendus pour une distribution normale parfaite (ligne en pointillés). Les résidus se regroupent ici le long de la ligne de distribution normale parfaite. Les résidus suivent donc une distribution normale, ce qui est confirmé également par un test de Shapiro-Wilk.

Finalement, le dernier graphique 4-11c. compare la valeur des résidus standardisés, c'est-à-dire les résidus normalisés par l'écart-type de l'erreur, avec les valeurs prédites et permet d'évaluer l'homogénéité de la variance. Ce graphe peut être également complété avec un test de Bartlett qui a également été réalisé sur ces données. Les valeurs prédites sont obtenues à l'aide du modèle de prédiction décrit par l'équation 4-1. En résumé, on peut observer sur ce graphe la dispersion des résidus relative à chaque modalité de facteurs (8 colonnes composées de 3 points chacune), étant donné que seule l'humidité a de l'influence sur le signal du silicium, on n'observe finalement que 2 valeurs prédites (pour la modalité -1 et +1 de l'humidité). L'intérêt de ce graphe est d'évaluer si la dispersion des résidus est homogène entre les différentes modalités de facteur testées. En règle générale, plus la ligne rouge est horizontale et plus la distribution des résidus est homogène. Ici, avec seulement deux valeurs distinctes de valeurs ajustées, la ligne rouge ne peut être qu'horizontale et n'est donc pas très significative. A l'aide de ce graphe et du test de Bartlett, les résidus ont quand même été considérés comme homogènes.

L'ensemble de ces étapes statistiques est indispensable afin de s'assurer de la fiabilité du modèle et d'autoriser l'analyste à conclure sur l'influence des différents facteurs sur les mesures réalisées. Les résultats obtenus à partir de l'ANOVA (non montrés ici), en ce qui concerne Si, indiquent bien l'humidité comme seul facteur influent, avec une influence maximale. Ce résultat a été obtenu pour tous les éléments du profil 1 de réponse du plan d'expériences. Ceci signifie que pour les éléments du profil 1 (Al, Fe, Si), l'ANOVA permet de valider les résultats obtenus à l'issue du plan d'expériences, à savoir une forte influence de l'humidité sur le signal LIBS et une absence d'influence de la taille du grain et du poids de pression.

En revanche, en ce qui concerne les éléments du profil 2 (C, Ca, K, Mg), il n'a pas été possible de valider les résultats du plan d'expériences à l'aide de l'ANOVA. Plus précisément, il n'a pas été possible de prouver l'homogénéité de la variance. En effet, les résultats de l'ANOVA indiquent une forte influence de la taille du grain aussi importante que l'humidité sur le signal LIBS du calcium pour la raie Ca I à 315 nm. Cependant, d'un point de vue statistique, toutes les conditions concernant les résidus ne sont pas remplies. On remarque notamment que l'homogénéité des résidus n'est pas prouvée avec une ligne qui n'est pas horizontale sur le graphe des résidus standardisés vs les valeurs ajustées. Ce manque d'homogénéité a été confirmé par le test de Bartlett qui est négatif, même lorsque que l'on applique une transformation logarithmique aux données dans le but d'améliorer la linéarité du modèle.

Ce problème de variance des résidus pour les éléments du profil 2 provient du manque de répétabilité des analyses en présence d'agrégats. Ces agrégats sont de taille et de quantité variables d'une pastille à l'autre, et peuvent être inclus ou pas dans les zones d'analyse. Il est donc compliqué d'obtenir des analyses répétables et des résultats précis sur l'influence de ces agrégats, d'où cette non-validation des modèles, en tout cas avec le protocole choisi. Une augmentation du nombre de zones d'analyse pourrait améliorer la fiabilité de ces modèles mais la réalité du laboratoire, ici la fragilité des pastilles humides, ne le permettait pas. Cependant, l'effet de ces agrégats est observable et représente une importante source de fluctuations du signal pour certains éléments, mais ne peut pas être estimé de façon statistique.

Néanmoins, la présence d'humidité dans les sols a été jugée, d'après ces résultats, comme le paramètre extérieur impactant le plus le signal LIBS. Ces résultats ont été obtenus en conditions contrôlées de laboratoire, sur des pastilles de sol préparées, il est donc légitime d'envisager une influence encore plus grande de l'humidité en conditions de mesures directes sur le terrain. Nous nous sommes donc interrogés sur la possibilité de réduire et de compenser l'effet de l'humidité sur le signal.

iv. Evaluation et correction de l'effet de l'humidité sur le signal LIBS

Une expérience complémentaire a été conduite pour évaluer les effets de l'humidité sur le signal LIBS, mais cette fois-ci en testant des niveaux d'humidité plus proches des conditions réelles, car les sols peuvent connaître de nombreuses variations d'humidité en conditions naturelles. Le sol utilisé dans cette expérience provient d'une cressonnière, où l'eau circule le long de fossés avec une légère pente. Ils présentent donc une variation naturelle importante d'humidité entre le haut et le bas de la pente.

a) Protocole expérimental

Différentes quantités d'eau pure ont été ajoutées à des répliques du même sol après séchage et pastillage. Ainsi, six pastilles d'un même sol ont été préparées avec des niveaux croissants d'humidité (de 2% jusqu'à 25%), en ajoutant à la micropipette des volumes précis d'eau ultrapure sur des pastilles de sols de 350mg et 13mm de diamètre. La correspondance entre le

volume d'eau ajoutée et le pourcentage d'humidité est donnée dans le tableau 4-3 en annexe. Cinq zones d'analyse ont été réalisées par pastille comprenant chacune 25 spots laser avec 5 tirs par spot, à l'exception de la pastille contenant 25% d'humidité qui n'est caractérisée que par 2 zones d'analyse à cause de sa fragilité. Le spectre moyen de chaque pastille correspond donc à la moyenne de 625 spectres (et 250 pour la pastille à 25% d'humidité).

b) Evolution du signal en fonction du taux d'humidité

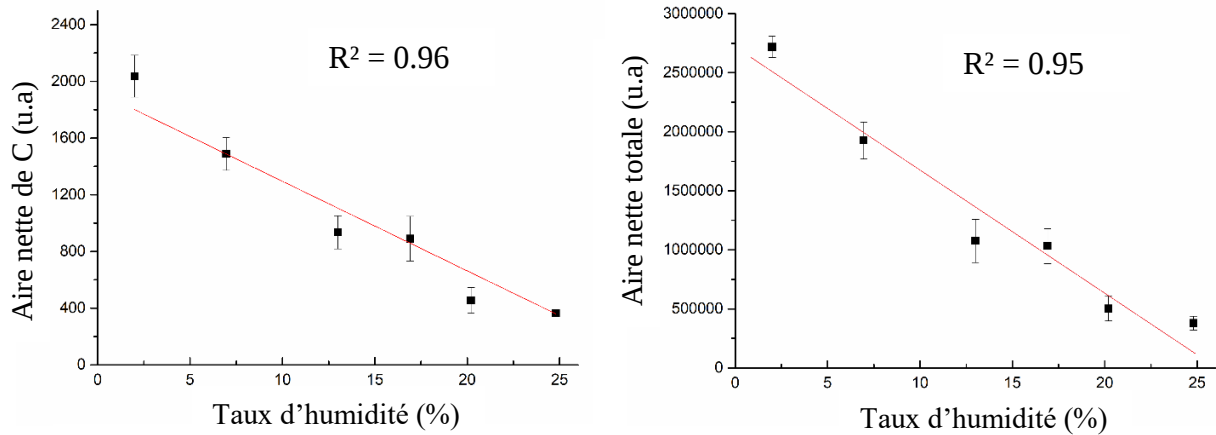


Figure 4-12 : Graphiques montrant l'évolution des intensités nettes de la raie de C à 193 nm et de l'aire nette totale sous le spectre en fonction du pourcentage d'humidité contenu dans la pastille de sol. Les barres d'erreur représentent l'écart-type de l'intensité calculé à partir des 5 zones d'analyse réalisées par taux d'humidité.

L'évolution du signal LIBS en fonction du taux d'humidité dans la pastille de sol est illustré par la figure 4-12 pour l'aire nette de la raie de C à 193.03 nm et l'aire nette totale sous le spectre. Toutes les raies élémentaires détectées présentent le même comportement que la raie du C (Figure 4-12a), et la diminution de l'aire totale sous le spectre (Figure 4-12b) montre que l'humidité impacte la globalité du spectre. A partir de la figure 4-12, on observe qu'une forte corrélation linéaire ($r^2 \geq 0.95$) existe entre le signal LIBS et le taux d'humidité, le signal diminuant linéairement avec l'augmentation du taux d'humidité. En effet, on peut noter une chute importante des intensités, de 2000 à 400 lorsque l'on atteint un taux d'humidité à 25% pour le signal de C et de 2750000 à 500000 pour l'aire totale, soit un facteur 5 environ.

Par ailleurs, au-delà de 25% d'humidité et dans nos conditions d'analyse, il a été observé que le plasma ne pouvait plus se former et donc que plus aucun signal ne pouvait être acquis. La figure 4-13 montre une photo de la surface de la pastille après l'impact du laser lors de l'analyse à 15% (figure 4-13a.) et 30% (figure 4-13b.) d'humidité. On remarque bien pour 15% d'humidité, 25 cratères formés individuellement par l'impact des tirs lasers alors que pour 30% d'humidité, on peut observer un impact très profond et unique avec une acquisition des données qui est stoppée à la moitié de la mesure. On peut donc conclure, avec le système portable Z-300, qu'il n'est plus possible de réaliser de mesure LIBS au-delà de 25% d'humidité contenue dans le sol.

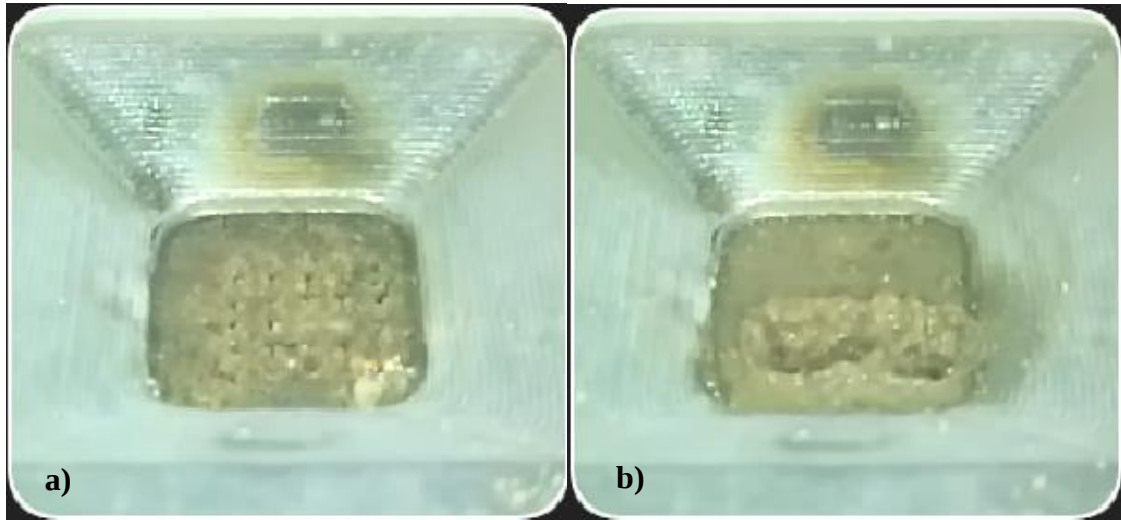


Figure 4-13 : photos de la fenêtre de tir lors de l'analyse LIBS d'un sol présentant un taux d'humidité de a) 15% et b) 30%.

L'effet de l'humidité sur le signal LIBS de sols a souvent été observé dans des études précédentes [102, 122, 123] limitant la prise de mesures dans des conditions d'analyse *in situ*. Généralement, il est alors conseillé d'envisager une préparation de l'échantillon avant analyse mais dans le contexte de notre étude (appareil portable, intérêt de la filière pour le développement de capteurs divers sur le terrain) nous avons essayé d'évaluer le potentiel de la normalisation dans le but de compenser l'effet de l'humidité.

c) Potentiel de la normalisation pour réduire les effets de l'humidité

Nous avons appliqué les mêmes méthodes de normalisation que celles exploitées dans le chapitre II pour traiter les spectres des pastilles humides : normalisations SNV, par l'aire totale sous le spectre et par étalon interne (Argon).

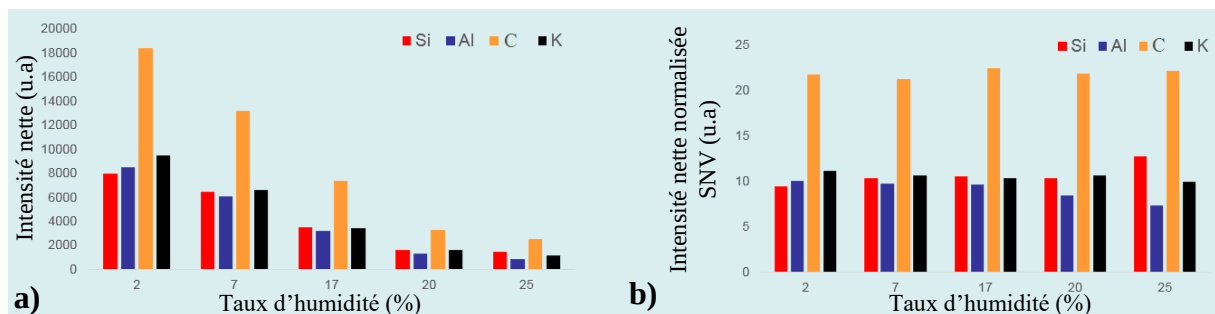


Figure 4-14 : Intensités nettes de plusieurs raies élémentaires (Si I à 288 nm, Al I à 309 nm, C I à 193 nm et K à 766 nm) en fonction du taux d'humidité contenu dans la pastille de sol a) non normalisées et b) normalisées SNV.

La figure 4-14 compare les intensités nettes de plusieurs raies élémentaires (Si I à 288 nm, Al I à 309 nm, C I à 193 nm et K à 766 nm) en fonction du taux d'humidité contenu dans le sol analysé i.) non normalisées (fig 4-14a) et ii) les intensités normalisées par SNV (fig 4-14b). Des performances similaires ont été obtenues avec la normalisation par l'aire totale sous le spectre. On remarque, comme démontré précédemment, que l'intensité nette de toutes les raies élémentaires diminue linéairement avec l'augmentation du taux d'humidité dans le sol. Cependant, lorsque les spectres sont normalisés, les intensités, pour un élément donné, semblent atteindre une valeur similaire quel que soit le niveau d'humidité, démontrant l'efficacité de la normalisation sur une large gamme de valeurs d'humidité. Un comparatif des spectres normalisés et non normalisés est donné sur la figure 4-4 en annexe. Ainsi, même si la correction n'est pas parfaite, on observe que la normalisation permet néanmoins de compenser l'effet de l'humidité en réduisant très nettement les différences d'intensités d'un taux d'humidité à l'autre. Il reste à vérifier si cette méthode de correction est généralisable pour tous types de sols et de contextes, et si on peut notamment l'utiliser pour des applications quantitatives. Les modèles d'étalonnage construits pour les sols dans cette étude, et présentés en chapitre III, n'étant pas normalisés et étant basés sur un calcul PLS, cela ne nous permet pas d'aller plus loin dans cette discussion.

3) Conclusion sur l'effet des paramètres extérieurs sur le signal LIBS

i. Mesures *in situ*

Les mesures LIBS sur site d'une culture de blé sur le site expérimental de QualiAgro se sont révélées être non concluantes, principalement à cause du taux d'humidité contenu dans les échantillons allant de 18% pour les sols à un minimum de 40% en moyenne pour les plantes. Avec un système portable au contact de l'échantillon, de la vapeur d'eau se concentre au niveau de la tête de l'appareil et empêche la bonne collection du signal. D'un point de vue instrumental, le nettoyage régulier de la fenêtre de tir pour pallier à ce problème est en contradiction avec le mode de mesure choisi, ayant pour objectif de caractériser rapidement la parcelle avec un instrument sur site. Cependant, à la fois pour les matrices « plantes » et « sols », du signal est détecté et permet d'observer les mêmes éléments qu'en laboratoire sur des pastilles préparées. La question de pouvoir quantifier les éléments totaux dans les plantes et les sols frais à l'aide des modèles quantitatifs bâtis sur les échantillons préparés en labo se pose alors. Cette discussion n'a pas été poussée plus loin, ces premières données obtenues via les mesures *in situ* étant n'étant pas réellement exploitables, du fait de la faiblesse du signal obtenu, et de l'importance des RSD mesurés.

La conclusion de cette étude en conditions réelles d'analyse n'est pas étonnante, la grande majorité des mesures LIBS étant habituellement réalisée en laboratoire, sur des échantillons préparés (séchage, broyage, pastillage). En ce qui concerne les sols, plusieurs travaux ont déjà été menés directement sur le terrain mais la conclusion aboutit toujours à un frein analytique venant de l'humidité et au besoin de préparer l'échantillon avant analyse afin d'obtenir des résultats fiables [50, 102, 126]. Le constat pour les plantes est sensiblement le même : pour traiter les spectres LIBS de plantes d'une manière quantitative, il est fortement conseillé de suivre les mêmes étapes de préparation des échantillons pour les sols (séchage, broyage,

pastillage) [10, 127], ainsi que nous l'avons montré au chapitre II. Plusieurs études ont cependant étudié la capacité de la méthode LIBS à analyser des feuilles fraîches ou *a minima* séchées (mais non broyées ni pastillées) mais elles confirment l'obtention de performances seulement qualitatives ou semi-quantitatives [59, 102, 127]. Une étude, menée par Guerra et al. en 2015 [127], estime que la LIBS présente un très gros potentiel pour l'analyse directe des plantes à travers un protocole d'analyse LIBS de plantes séchées. Cependant, même si les coefficients de corrélation entre la méthode LIBS et la méthode XRF, utilisée ici comme référence, pour la quantification en éléments nutritifs (P, K, Ca, Mg, Fe, Mn, Si) des plantes atteignent des valeurs $\geq 0,95$, ces dernières semblent fortement impactées par des points extrêmes.

En résumé, pour les deux matrices, le passage des conditions contrôlées en laboratoire à l'analyse de terrain est particulièrement compliqué avec une baisse de signal et de fortes fluctuations. Dans le but de comprendre plus en détails la source de ces fluctuations, une étude a été menée sur l'effet de plusieurs facteurs extérieurs (humidité, taille du grain et compacité) sur le signal LIBS d'un sol.

ii. Effet des facteurs extérieurs sur le signal

L'observation de l'effet de chaque facteur individuellement sur le signal étant une tâche fastidieuse, il a été décidé d'utiliser un plan d'expériences factoriel complet afin d'évaluer l'influence des paramètres extérieurs sur le signal LIBS. Plus précisément, les valeurs de taux d'humidité (de 2 à 10%), de taille du grain maximale (de 0.25 à 2 mm) et de poids de pression (de 2 tonnes à 8 tonnes) ont été testées pour la préparation d'une pastille de sol, typique d'un sol de grande culture française. Les résultats obtenus à la suite de ce plan d'expériences ont permis de mettre en lumière l'impact négatif de l'humidité sur l'intensité du signal mais également le danger d'analyser des échantillons présentant des hétérogénéités fortes au niveau de la taille du grain. En effet, il a été observé, pour toutes les raies élémentaires détectées dans le spectre LIBS du sol, que l'augmentation du taux d'humidité entraînait une diminution importante de l'intensité LIBS. De plus, les fluctuations observées pour l'intensité des raies de certains éléments (Ca, Mg, K, C), ont été attribuées à la présence d'agrégats lorsque l'échantillon de sol n'était pas broyé (taille du grain = 2 mm). Ce résultat n'est donc pas à confondre avec le constat, souvent démontré [36, 82], que le signal LIBS augmente lorsque la taille des grains diminue, ici on parle de taille de grain maximale et d'hétérogénéité de l'échantillon, à travers une accumulation de particules riches en Ca, Mg, K et C qui peuvent former des agrégats jusqu'à 2 mm de largeur.

En ce qui concerne la compacité du sol, elle n'a présenté aucune influence sur le signal acquis. Il a déjà été démontré que pour une pastille de sol, à partir d'un certain poids de pression (1 tonne), la température de plasma et la densité électronique restent identiques, autrement dit l'intensité reste la même [123]. Cependant, lorsque le sol est compacté manuellement, à l'aide d'une spatule par exemple, les fluctuations d'intensité d'une zone d'analyse à l'autre deviennent importantes, trop pour obtenir un spectre moyen représentatif du sol [105, 123].

L'humidité reste le facteur le plus impactant sur le signal LIBS même si la compacité du sol et la taille du grain sont à surveiller sous certaines conditions. Une étude, afin de comprendre

complètement l'influence du taux d'humidité sur le signal et comment le compenser, a finalement été menée.

iii. Focus sur l'effet de l'humidité et comment le compenser

Différents volumes d'eau pure (de 0 à 100 μL par pas de 20 μL) ont été ajoutés sur la surface d'une pastille de sol de cresson afin de mimer différents taux d'humidité (de 2 à 25%) dans le sol et d'observer le signal LIBS associé à chaque taux d'humidité. Il a été observé, pour l'ensemble des raies élémentaires détectées dans le spectre LIBS, une corrélation linéaire (avec une valeur de r^2 proche de 0,95) entre le taux d'humidité et l'intensité du signal LIBS. L'intensité du signal diminuant avec le taux d'humidité, avec environ un rapport de 5 entre la pastille présentant une humidité résiduelle de 2% et la pastille la plus humide (25% d'humidité). Malgré cette forte diminution, le signal LIBS reste détectable sur toute la gamme d'humidité, proche des taux d'humidité rencontrés dans la nature [128]. La LIBS semble donc suffisamment performante dans le but de détecter du signal dans les sols pour des taux d'humidité réalistes. Cependant, il a aussi été observé, à partir de 30% d'humidité et pour le pistolet Z-300, que le plasma ne pouvait plus se former et donc qu'aucun signal ne pouvait être acquis. Cette donnée peut donc être un frein analytique pour des mesures *in-situ* lors de précipitations importantes.

La méthode la plus recommandée pour réduire ces effets d'humidité consiste à optimiser la préparation de l'échantillon [50, 105, 126, 129] mais un autre moyen de correction du signal, encore jamais utilisé, a été testé à travers l'utilisation de la normalisation. Les spectres générés à la suite de l'étude sur l'humidité ont été normalisés selon les techniques de normalisation développées en chapitre III, à savoir la normalisation SNV, par l'aire totale et par étalon interne (et plus précisément une raie de l'argon). Il a été observé, après l'application de la normalisation SNV et celle par l'aire totale, que les intensités qui diminuaient avec le taux d'humidité présentaient maintenant la même valeur d'intensité quel que soit le taux d'humidité appliqué. Ce cas d'étude n'a été appliqué qu'à un seul type de sol et l'efficacité de la normalisation n'a pas encore été évaluée pour des variétés de sols/plantes différentes. Il semble que la normalisation puisse présenter un certain potentiel pour réduire les effets de l'humidité sur le signal mais cette approche reste exploratoire et à développer car peu exploitée [130].

iv. Recommandations

A travers cette étude, nous pouvons recommander 3 modalités d'utilisation de la LIBS afin d'atteindre des mesures quantitatives fiables sur les matrices des sols et des plantes. Deux de ces modalités (2 et 3) nécessitent de faire d'abord progresser les réflexions concernant la normalisation des spectres pour compenser l'effet de l'humidité et la préparation rapide des échantillons sur le terrain.

① L'échantillon de sol ou de plante est transféré en laboratoire pour être préparé puis analysé par LIBS. Cette modalité permet de réduire considérablement la préparation de l'échantillon et le temps d'analyse par rapport à une méthode classique telle que l'ICP-AES.

Cette approche reste basée sur une démarche d'analyse en laboratoire et ne peut répondre à des questions d'agriculture de précision et d'aide à l'échantillonnage qu'avec des modes opératoires de collecte et transfert des échantillons au laboratoires ultra-rapides, suivis d'analyses LIBS conduites par un instrument extrêmement performant en termes de cadence analytique.

② Les modèles d'étalonnages sont bâtis en conditions de laboratoire sur des échantillons préparés (séchage/broyage/tamissage) à l'aide d'un instrument portable et avec une stratégie de calibration en cohérence avec les sols/plantes à caractériser. Les mesures sur échantillons inconnus sont ensuite réalisées directement sur le terrain sans préparation. Ce mode d'utilisation nécessiterait d'avoir réussi à corriger l'effet de l'humidité par une technique de normalisation adaptée et induirait surement des performances uniquement semi-quantitatives. A posteriori cela permettrait de multiplier les mesures et d'utiliser l'instrument LIBS comme une aide à l'échantillonnage avant une analyse plus poussée.

③ Les modèles d'étalonnages sont bâtis en conditions de laboratoire sur des échantillons préparés (séchage/broyage/tamissage) à l'aide d'un instrument portable et avec une stratégie de calibration en cohérence avec les sols/plantes à caractériser comme pour la modalité 2. Les mesures sur échantillons inconnus sont ensuite réalisées directement sur le terrain mais cette fois en intégrant une préparation d'échantillon rapide. Cette dernière modalité d'utilisation nécessite de pouvoir déplacer sur site certains outils utiles afin d'améliorer les conditions d'analyses comme une pastilleuse portable. Cette dernière peut permettre potentiellement de réduire en grande partie l'humidité résiduelle du sol par percolation et donc d'obtenir un signal plus exploitable. Cependant, même dans ces conditions, la pastille fraîche ne serait certainement pas identique à celle préparée en laboratoire (notamment humidité, granulométrie, compaction différentes), impliquant toujours une méthode de normalisation (PLS, SNV, aire totale...).

Conclusion

Le travail de recherche présenté dans le cadre de ce mémoire de thèse s'inscrit dans la problématique de production raisonnée et durable de ressources agricoles dans un contexte de demande mondiale sans cesse croissante. Plus précisément, nous avons évalué les performances d'un instrument LIBS portable (modèle Z300, SciAps), sachant que la LIBS, spectroscopie d'émission atomique à partir d'un plasma induit par laser, qui offre des analyses multi-élémentaires rapides, pourrait permettre de reconsidérer les approches actuelles d'échantillonnage des parcelles agricoles, aussi bien spatialement que temporellement. Notre évaluation a été réalisée à partir d'un corpus de 62 sols et d'un corpus de 25 végétaux, non seulement dans des conditions idéales de laboratoire afin d'établir les performances limites, mais aussi dans conditions réelles afin de mettre en évidence l'influence des facteurs limitants.

A l'aide des analyses LIBS, les éléments les plus concentrés dans les sols et les plantes tels que C, P, K, Ca, Mg, Fe, Al et Na, ont été détectés à l'inverse de ceux moins concentrés tels que Mn, Cu, Zn, et Cd. Cependant, ce résultat ne concerne que le pistolet Z-300 utilisé dans le cadre de cette étude et non les performances atteintes par la méthode LIBS en général.

Les spectres LIBS ont d'abord été traités par analyse en composantes principales (ACP). Concernant les sols, le graphe des scores dans le plan PC1-PC2 a montré une distribution des échantillons semblable à celle observée dans un diagramme ternaire dérivé des données de concentration de référence. En ce qui concerne les plantes, l'ACP a permis de révéler des différences physiques entre échantillons, certainement dues au type de broyage ou à leur nature intrinsèque, ou les deux.

A propos de l'analyse quantitative par LIBS, nous avons bâti des modèles d'étalonnage à partir de régression entre les valeurs du signal LIBS et celles des concentrations de référence issues des mesures ICP-AES. Pour évaluer les performances des différents modèles, nous avons sélectionné trois indicateurs : le coefficient de détermination r^2 , la limite de détection LOD et l'erreur de prédiction calculée à partir de l'erreur quadratique moyenne RMSE portant sur les données du jeu de calibration uniquement (RMSEC), ou encore issue d'un processus de validation croisée (RMSECV). Enfin, l'erreur de prédiction a aussi été calculée à partir des échantillons d'un jeu de validation (RMSEP) n'ayant pas été pris en compte pour la construction du modèle.

Concernant les plantes, nous avons montré qu'une approche univariée était tout à fait satisfaisante, après une étape de normalisation des spectres LIBS. Nous avons, à cette occasion, montré que, parmi les quatre méthodes de normalisation couramment appliquées en LIBS, le meilleur choix dépendait de la raie d'émission considérée. Nous avons ainsi pu bâtir des modèles d'étalonnage pour les éléments P, K, Ca, Mg, Fe, Al et Na. Mais, alors que les modèles concernant les éléments Ca, Mg, Fe, ont donné des performances de validation acceptables ou encourageantes (Na et P), le modèle relatif à l'aluminium n'a pas pu être considéré exploitable. Nous pouvons finalement conclure que le corpus d'échantillons de plantes considéré dans le

Conclusion

cadre de cette étude est vraisemblablement trop restreint pour que les modèles puissent être généralisés.

Concernant les sols agricoles, nous avons montré que l'analyse univariée était totalement inadaptée, et ce, malgré les différentes stratégies de normalisation des spectres LIBS. La seule exception à cette règle a été observée pour le calcium. Nous avons finalement choisi de bâtir des modèles de régression par moindres carrés partiels (PLS) afin d'exploiter la richesse des spectres LIBS, et plus précisément l'influence de raies d'émission atomiques autres que celles associées à l'analyte. Les modèles PLS relatifs aux éléments Mg, Fe et Na ont montré des performances satisfaisantes lors de leur validation, tandis que celui relatif à Mg ne s'est pas montré suffisamment robuste lors de la validation. A l'inverse, les modèles relatifs aux éléments P et N n'ont pas pu être construits parce que le signal associé à l'élément considéré était très faible (P), voire totalement absent (N). Enfin, bien que le signal relatif aux éléments C et Al soit détecté, aucune corrélation entre signal et concentration n'a été observée, vraisemblablement à cause d'effets de matrice très pénalisants pour ces deux éléments. Ce bilan conduit à conclure que les données LIBS des sols agricoles traitées par PLS permettent d'atteindre des performances semi-quantitatives. Il est cependant important de préciser que ce type de résultat est attendu lorsque le corpus est constitué d'échantillons de différentes textures, et ce n'est que lorsque les échantillons de sols présentant des matrices similaires qu'on peut espérer atteindre des performances quantitatives.

Tandis que la technique ICP-AES concerne des analyses après minéralisation des échantillons solides, nous avons souhaité évaluer la technique XRF, permettant d'analyser les échantillons de sols et de plantes avec le même niveau de préparation que celui exigé par la LIBS. Il est intéressant de noter que les *outliers* révélés en LIBS sont également considérés *outliers* par XRF, soit à cause des propriétés physiques de l'échantillon ou d'une valeur de référence issue des mesures ICP-AES à remettre en cause. Par ailleurs, nous avons observé que pour certains éléments, le modèle LIBS était meilleur que le modèle XRF, tandis que pour d'autres, l'inverse se produit. La combinaison des deux méthodes – LIBS et XRF – pourrait donc représenter une bonne stratégie pour l'analyse multiélémentaire rapide des sols agricoles et des plantes.

Au-delà des expériences de laboratoire qui ont permis d'évaluer les performances limites de l'instrument LIBS utilisé dans des conditions optimales, nous avons aussi mené une campagne de mesures exploratoire directement sur le terrain, et plus précisément sur des parcelles expérimentales de blé du site Qualiagro. Les mesures consistaient en l'analyse LIBS de feuilles de blé fraîches et de sols frais sur ces mêmes parcelles. Nous avons remarqué une absence de signal sur la majorité des spectres LIBS des feuilles fraîches, expliquée par le taux d'humidité de l'ordre de 45% dans ce cas. Concernant les sols, ce taux n'étant que de l'ordre de 18%, il a été possible de détecter un signal LIBS et d'observer les mêmes éléments (C, P, K, Ca, Mg, Fe, Al et Na) que sur pastilles sèches malgré des intensités globalement moins élevées. Nous avons également observé, sur les spectres de sols et de plantes présentant du signal, de fortes fluctuations d'intensité (avec parfois une RSD > 100 %) d'une zone d'analyse à l'autre dues à l'absence de préparation des échantillons. Nous n'avons donc pas pu exploiter plus amplement les spectres LIBS, indiquant que l'analyse LIBS directe sur un échantillon agricole non préparé n'était pas adaptée.

Face aux difficultés liées à l'analyse LIBS directe, nous avons souhaité mieux comprendre comment le signal LIBS est impacté par des facteurs externes liés à la préparation de

Conclusion

l'échantillon et à son environnement. Pour cela, nous avons fait appel à un plan d'expériences factoriel complet sur le signal LIBS d'une pastille de sol, les trois facteurs étant l'humidité, la taille de grain maximale et la compacité de la pastille. Nous avons observé deux types de réponses en fonction de la raie d'émission sélectionnée : pour les raies d'émission associées à Si, Al, Fe, nous avons observé comme unique effet, une diminution significative de l'intensité avec l'augmentation du taux d'humidité, tandis que pour les raies relatives à C, Ca, K et Mg, la taille de grain a en plus été identifiée comme un paramètre influant. Cependant, ce résultat a été interprété comme une conséquence de la formation d'agrégats riches en ces éléments lorsque le broyage n'est pas suffisant, si bien que seul le taux d'humidité a été considéré comme facteur ayant un impact significatif. Nous avons observé que le signal LIBS baissait de façon linéaire lorsque le taux d'humidité augmentait et disparaissait totalement aux alentours de 30% d'humidité. Nous avons cependant montré la possibilité de compenser efficacement l'effet de l'humidité dans le sol par normalisation. Mais ce résultat reste néanmoins à consolider sur un ensemble de sols couvrant toute une gamme d'étalonnage afin de démontrer qu'un modèle quantitatif pourrait être bâti quel que soit le taux d'humidité des échantillons. Il est aussi à nuancer car nous avons ajouté des quantités d'eau choisies à des pastilles de sols secs, ce qui nous éloigne des conditions naturelles.

Les suites à cette étude peuvent être de différentes natures :

- Les modèles d'étalonnage peuvent être améliorés par la prise en compte de nouveaux échantillons dans le jeu de calibration ainsi que par la compréhension des échantillons outlier.
- Exploiter d'autres modèles de régression afin de pouvoir optimiser les modèles d'étalonnage ou de prendre en compte une potentielle relation non-linéaire entre le signal et la concentration de référence.
- Améliorer l'ergonomie de l'instrument et du logiciel en dialogue avec le constructeur.
- Les performances de ce LIBS portable restent à évaluer sur un autre type de matrice d'intérêt pour la filière agronomique : les matières fertilisantes, qui englobent aussi bien des fertilisants minéraux que des ressources organiques ou minérales de plus en plus variées issues du recyclage. La réalisation d'analyses chimiques multiélémentaires concomitantes du sol, du fertilisant et de la plante doit permettre de suivre et de comprendre les transferts couplés d'éléments dans les agroécosystèmes, depuis l'application du fertilisant jusqu'à la quantité de nutriments extraites par la plante dans le sol. Cependant, cette dernière discussion implique de connaître la concentration en nutriments biodisponibles dans le sol et non la concentration totale élémentaire. Or, par définition, la LIBS permet de mesurer la concentration élémentaire, il conviendrait donc de réaliser des études menant à établir la corrélation entre le signal LIBS et la concentration biodisponible pour répondre plus complètement aux attentes des acteurs de la filière agronomique. La corrélation entre le signal et la concentration biodisponible a déjà été l'objet de plusieurs travaux en LIBS mais la manière d'extraire l'information liée à cette concentration spécifique du sol reste encore une discussion ouverte chez les agronomes. Finalement, d'après la littérature, la recherche de présence de signatures moléculaires dans les spectres LIBS (CN, CaO) de matrices agricoles pourrait représenter le vrai intérêt de la méthode.

Annexe – Chapitre I

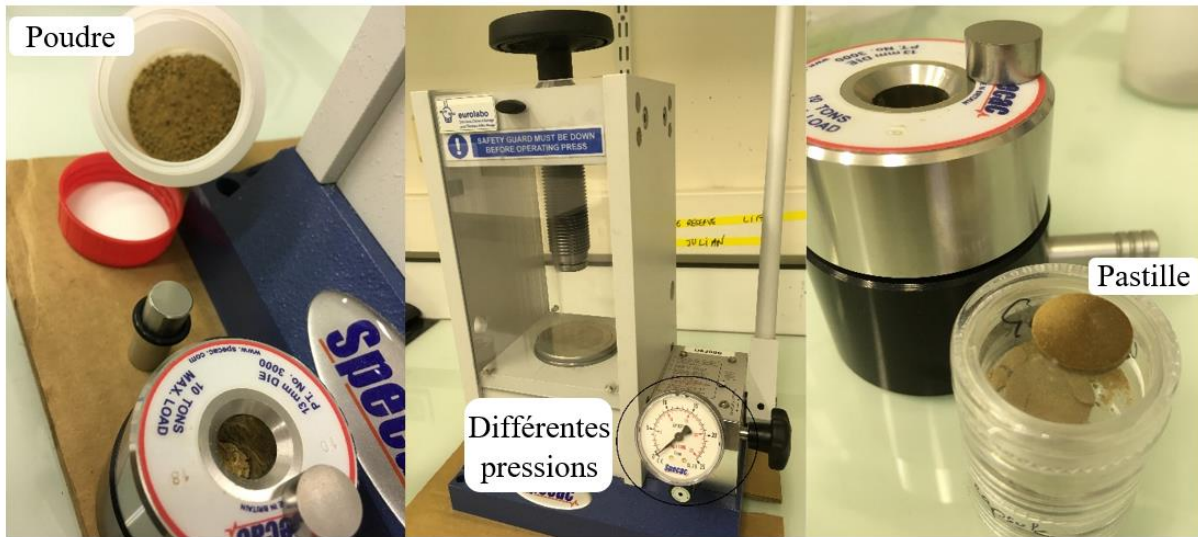


Figure 1-1 : Photo de la presse manuelle hydraulique utilisée pour le pastillage des échantillons de sol.

n	A_2
2	1,88
3	1,02
4	0,73
5	0,58
6	0,48
7	0,42
8	0,37
9	0,34

Tableau 1-1 : Tableau de correspondance entre la taille de l'échantillon et la valeur du facteur à utiliser pour calculer les limites de contrôle de la carte de contrôle.

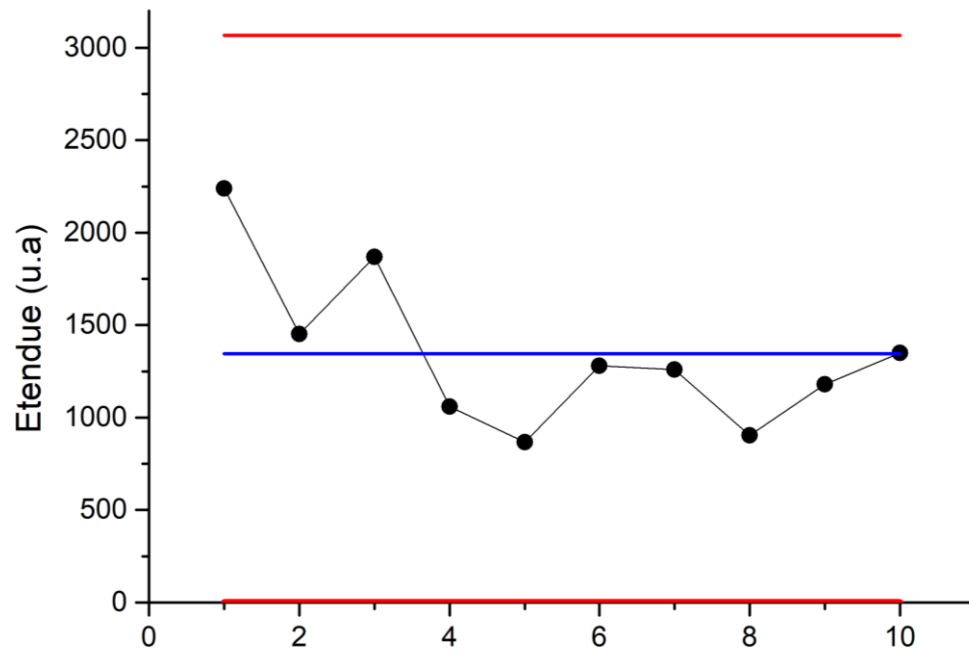


Figure 1-2 : Carte de contrôle de l'étendue pour l'instrument Z-300.

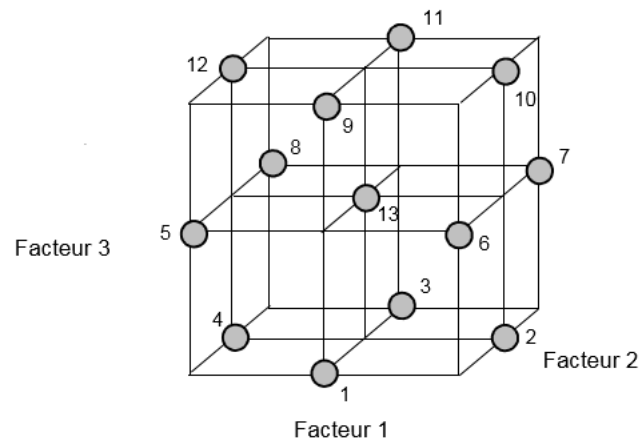


Figure 1-3 : Plan de Box-Behnken pour 3 facteurs.

Annexe

Numéro d'essai	Facteur 1	Facteur 2	Facteur 3	Tirs	Spots	Zones
<i>1</i>	<i>0</i>	<i>-1</i>	<i>-1</i>	<i>10</i>	<i>5</i>	<i>5</i>
<i>2</i>	<i>1</i>	<i>0</i>	<i>-1</i>	<i>15</i>	<i>10</i>	<i>5</i>
<i>3</i>	<i>0</i>	<i>1</i>	<i>-1</i>	<i>10</i>	<i>15</i>	<i>5</i>
<i>4</i>	<i>-1</i>	<i>0</i>	<i>-1</i>	<i>5</i>	<i>10</i>	<i>5</i>
<i>5</i>	<i>-1</i>	<i>-1</i>	<i>0</i>	<i>5</i>	<i>5</i>	<i>10</i>
<i>6</i>	<i>1</i>	<i>-1</i>	<i>0</i>	<i>15</i>	<i>5</i>	<i>10</i>
<i>7</i>	<i>1</i>	<i>1</i>	<i>0</i>	<i>15</i>	<i>15</i>	<i>10</i>
<i>8</i>	<i>-1</i>	<i>1</i>	<i>0</i>	<i>5</i>	<i>15</i>	<i>10</i>
<i>9</i>	<i>0</i>	<i>-1</i>	<i>1</i>	<i>10</i>	<i>5</i>	<i>15</i>
<i>10</i>	<i>1</i>	<i>0</i>	<i>1</i>	<i>15</i>	<i>10</i>	<i>15</i>
<i>11</i>	<i>0</i>	<i>1</i>	<i>1</i>	<i>10</i>	<i>15</i>	<i>15</i>
<i>12</i>	<i>-1</i>	<i>0</i>	<i>1</i>	<i>5</i>	<i>10</i>	<i>15</i>
<i>13</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>10</i>	<i>10</i>	<i>10</i>
<i>14</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>10</i>	<i>10</i>	<i>10</i>

Tableau 1-2 : Plan d'expérience de Box-Benkhen construit pour définir le nombre de mesures à effectuer sur la surface d'une pastille de sol.

Script Rstudio créé afin de corriger les spectres bruts issus du Z-300 après analyse

Ce script a été réalisé avec l'aide de Thomas Guzman dans le cadre de son stage de Master consacré au sujet suivant : « L'accumulation de silicium par les plantes en condition de stress hydrique : développement d'une technique d'analyse rapide de silicium par LIBS »

Le but de ce script Rstudio est de pouvoir corriger en quelques secondes les spectres LIBS issus du Z-300 après analyse. Celui-ci comprend plusieurs étapes distinctes :

1) Création d'une matrice composée de tous les spectres LIBS à corriger

Pour ce faire, il faut tout d'abord créer une fonction qui nous permet, à partir d'un dossier, d'importer les spectres dans une matrice vide. Cette fonction s'appelle *Assemblage_spectres*

Création de la fonction

```
Assemblage_spectres <- function(entree){
```

```
  setwd(entree)
```

entree représente le dossier dans lequel tous les spectres individuels sont réunis, dans notre cas, un spectre se présente sous la forme d'un fichier csv. setwd permet de définir dans quel dossier va être créée la matrice comprenant tous les spectres

```
  library(stringr)
```

le package stringr est utile pour préparer ou nettoyer des données

```
  files_names <- list.files(entree)
```

cette ligne permet de définir le nom des fichier

```
  nb_files <- length(files_names)
```

cette ligne permet de définir le nombre de fichiers/spectres à extraire, ici ce sera le nombre de fichiers contenus dans le dossier sélectionné

```
  Intensites <- matrix(NA, nrow = 8264, ncol = nb_files)
```

cette ligne sert à définir les dimension de la matrice des spectres à créer, c'est-à-dire son nombre de lignes et de colonnes. Un spectre issu du Z-300 est représenté par 8264 longueurs d'onde acquises soit autant de lignes pour la matrice à créer.

```
  colnames(Intensites) <- str_c(str_sub(gsub("_PixelData.csv", "", files_names), 1, 22),  
str_sub(gsub("_PixelData.csv", "", files_names), -10))
```

ici est défini le nom de chaque colonne de la matrice, à savoir le nom de chaque spectre. Ici sera sélectionné le nom du fichier csv associé sans son extension (_PixelData.csv)

On vient ici de créer la matrice vide qui va pouvoir accueillir tous les spectres que l'on souhaite. Il faut maintenant la remplir.

```
for (i in 1:nb_files) {  
  on crée à cet instant une boucle for, c'est un moyen d'appliquer une ligne de code à plusieurs entités individuellement  
  
  Intensites[,i] <- read.csv(files_names[i])[,3]  
  
  on fixe que la matrice des intensités va être composée de la troisième colonne de chaque fichier csv présent dans le dossier. Cette troisième colonne est bien celle des intensités acquises par le Z-300  
  
}  
  
Ondes <- read.csv(files_names[1])[,2]  
  
ici on fixe que la colonne des longueurs d'onde de la matrice finale représente la deuxième colonne du premier fichier csv du dossier soit le premier spectre. La deuxième colonne est en effet la colonne des longueurs d'onde acquises par le Z-300.  
  
Spectres_a_corriger <- cbind(Ondes, Intensites)  
  
on fusionne à l'aide de la fonction cbind la colonne des longueurs d'onde et celles des intensités afin d'obtenir la matrice des spectres à corriger.  
  
}
```

Application de la fonction *Assemblage_spectres*

```
Assemblage_spectres("D:/Dossier_LIBS")
```

2) Il est maintenant possible de corriger les spectres LIBS issus du Z-300

```
wv<-as.matrix(Spectres_a_corriger [,])  
  
On isole la colonne des longueurs d'onde wv  
  
X<-as.matrix(Spectres_a_corriger [,])  
  
On isole la matrice des intensités X  
  
wvcorr1<-as.matrix(wv[19:2024,])  
wvcorr2<-as.matrix(wv[2099:4097,])  
wvcorr3<-as.matrix(wv[4361:6198,])
```

On conserve les plages spectrales de chaque spectromètre qui ne sont pas touchées par le recouvrement (wvcorr1, 2 et 3)

```
wvcorr<-rbind(wvcorr1,wvcorr2,wvcorr3)
```

On accole les plages spectrales tronquées à l'aide de la fonction rbind

```
Xcorr1<-as.matrix(X[1:2024,])
```

```
Xcorr2<-as.matrix(X[2099:4097,])
```

```
Xcorr3<-as.matrix(X[4361:6180,])
```

```
Xcorr<-rbind(Xcorr1,Xcorr2,Xcorr3)
```

On procède de la même façon mais pour la matrice des intensités

```
Spectres_corriges <- cbind(wvcorr,Xcorr)
```

On rassemble les longueurs d'onde et les intensités afin d'obtenir la matrice finale corrigée

```
sortie <- write.csv(Spectres_corriges,"Spectres corriges")
```

Finalement, on l'exporte sous le format que l'on souhaite, ici le format csv

Annexe - Chapitre II

Echantillon	Nature (espèce)
IPE 100	Herbe
IPE 124 – 148 - 159	Luzerne
IPE 950	Melon (Cucumis melo)
IPE 163	Poivrier
IPE 175	Tubercule de tulipe
IPE 133	Pousse de maïs (maïs Zea)
IPE 156	Feuille de Chêne
IPE 157	Feuille de Hêtre
IPE 165	Feuille de Palmier
IPE 186	Paille de riz
IPE 130	Pomme de Terre
IPE 172	Laurier de cerise (Prunus laurocerasus)
IPE 207	Herbe à éléphant
IPE 198	Feuille de banane (Musa paradisiaca)
IPE 199	Bois

Tableau 2-1 : Numéro et nature des échantillons utilisés pour construire les modèles d'étalonnage pour les plantes.

Elément	C I	Si I	P I	Fe I	Fe II	Mg I	Mg II	Al I	Ca I	Ca II
Longueur d'onde (nm)	193,1	212,4				285,2		308,2		315,9
	247,9	243,5	213,6		234,4	383,7	279,6	309,3	422,7	317,9
	370,6	251,6	214,9	302,1	238,2	517,3	279,8	394,4	443,5	393,3
	373,7	288,2			239,6	518,4	280,3	396,2	526,5	396,9
Elément	Na I	K I	H I	Li I	O I	Ar				
Longueur d'onde (nm)	588	766,5				696,5				
	589,6	769,9	656,3	670,8	777	714,7				
						763,5				

Tableau 2-2 : Récapitulatif des éléments détectés par le Z-300 dans les plantes et les longueurs d'onde associées (nm), attribuées à l'aide de la base de données NIST et la base données AtomTrace.

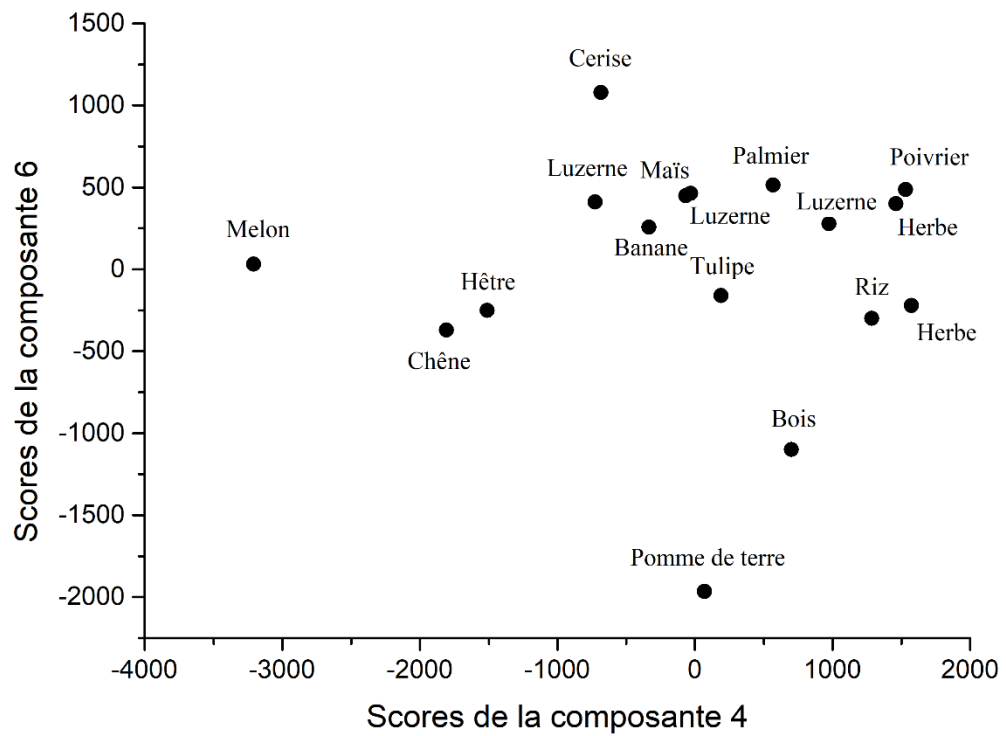


Figure 2-1 : Graphe des scores sur la composante 4 et 6 de l'ACP réalisée sur les échantillons de plantes.

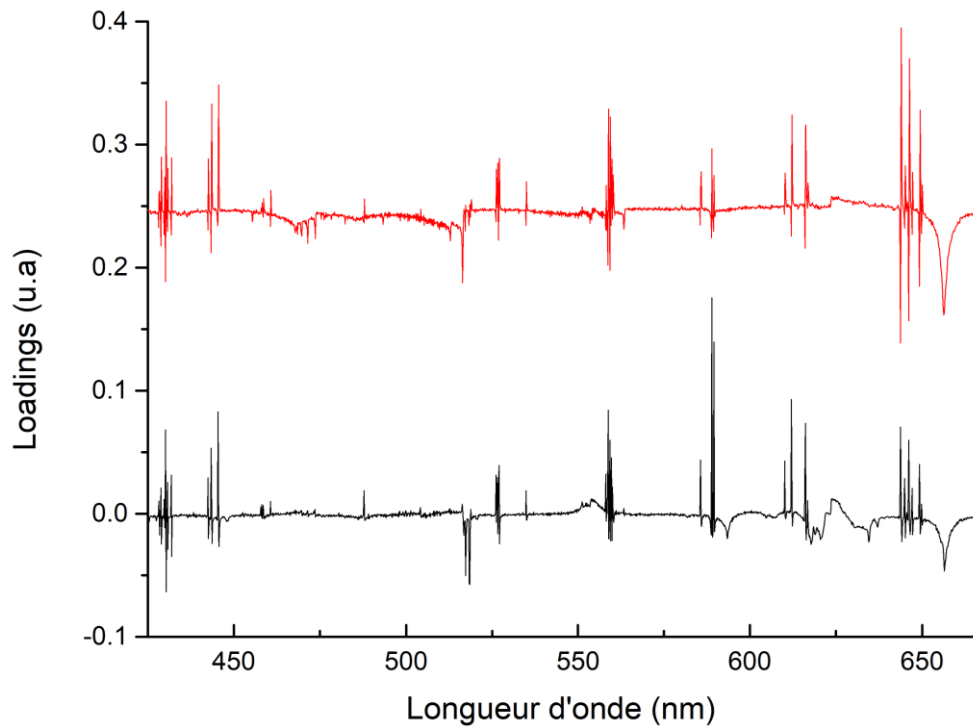


Figure 2-2 : Graphe des loadings zoomé de la composante 7 (trait noir) et 8 (trait rouge) de l'ACP réalisée sur les échantillons de plantes.

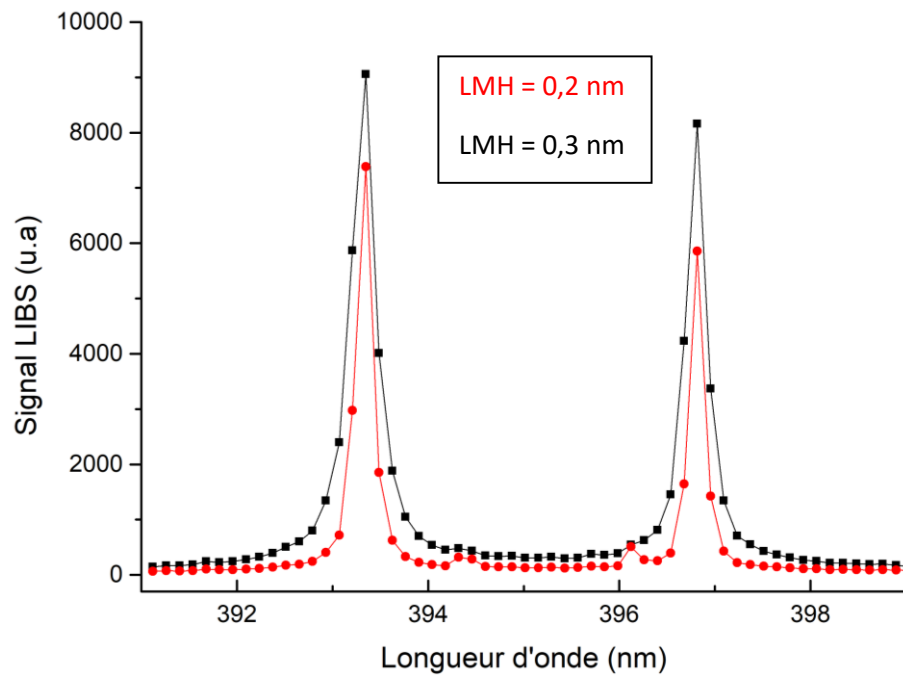


Figure 2-3 : Comparaison entre deux spectres zoomés de plante présentant soit des raies de Ca correctement définies (trait rouge) soit des raies auto-absorbées (trait noir). Les largeurs à mi-hauteur associées sont également indiquées.

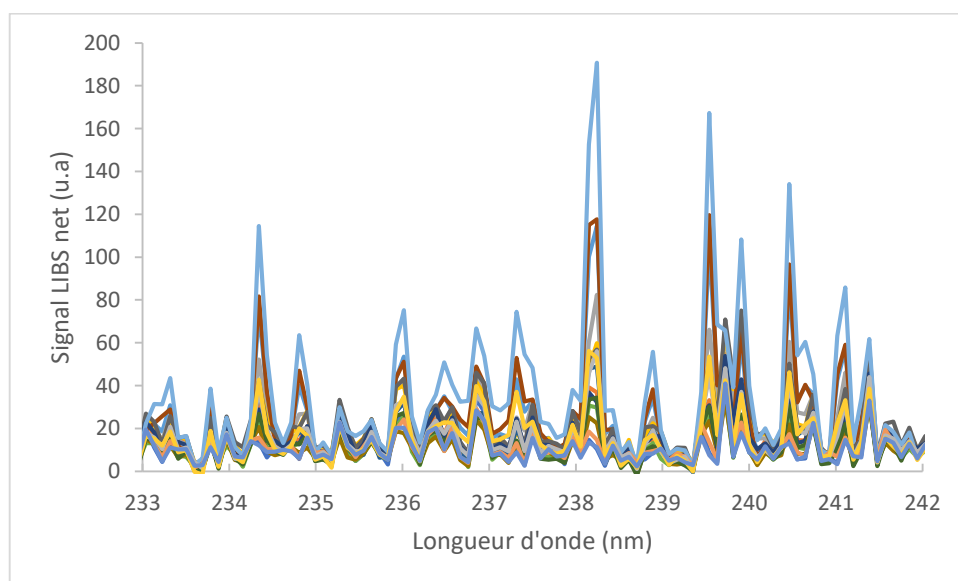


Figure 2-4 : Spectres des échantillons de plantes zoomés sur une plage spectrale ne comprenant que des raies du fer.

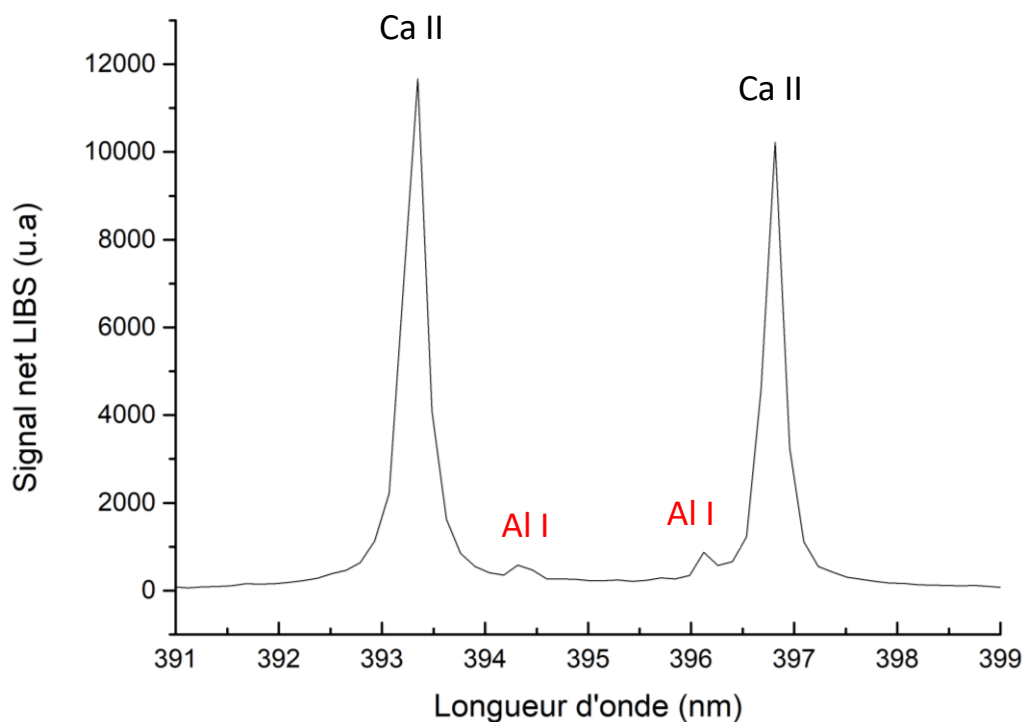


Figure 2-5 : Spectres d'un échantillon de plante zoomé sur le doublet d'Al I interféré par le doublet de Ca II.

Echantillon	Nature
CFM1	Feuille morte canne à sucre
CFV1	Feuille verte canne à sucre
FAE2	Forêt autre espèce
FNF1	Forêt non foliaire
MAH3	Feuille de Mahogany
MYR2	Feuille de Myrcia
V463	Maïs
IPE999	Feuille d'acacia

Tableau 2-3 : Numéro et nature des échantillons utilisés dans le jeu de validation utilisé pour valider les modèles d'étalonnage.

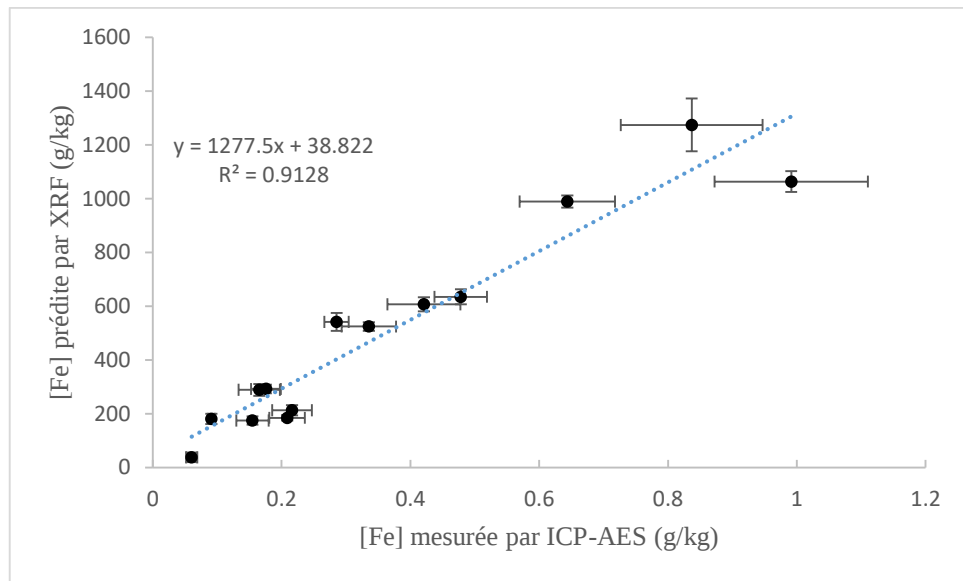


Figure 2-6 : Courbe d'étalonnage du fer obtenue sur les mesures XRF en fonction des concentrations mesurées par ICP-AES dans les plantes.

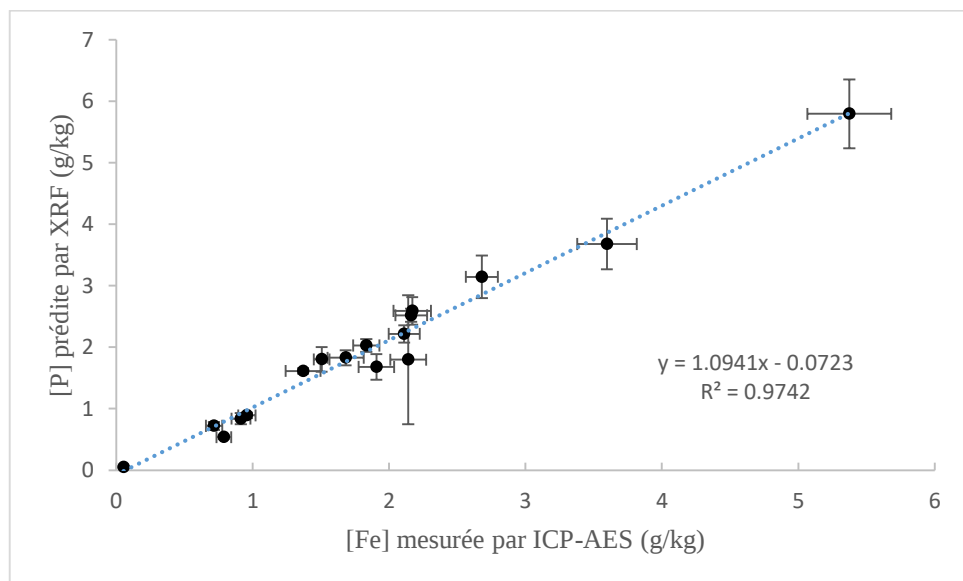


Figure 2-7 : Courbe d'étalonnage du phosphore obtenue sur les mesures XRF en fonction des concentrations mesurées par ICP-AES dans les plantes.

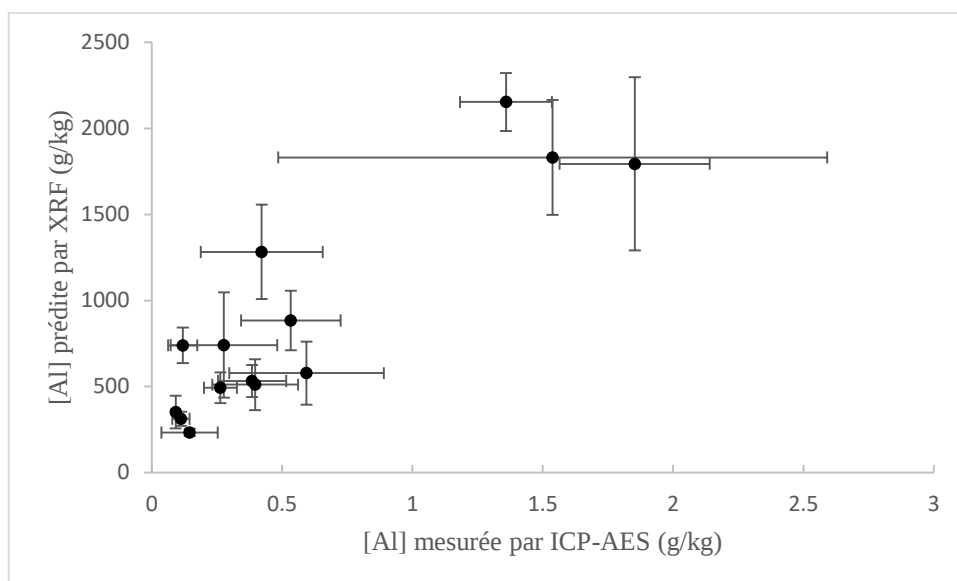


Figure 2-8 : Courbe d'étalonnage de l'aluminium obtenue sur les mesures XRF en fonction des concentrations mesurées par ICP-AES dans les plantes.



Invited Review

Critical review and advices on spectral-based normalization methods for LIBS quantitative analysis

Julian Guezenoc^{a,b,1}, Anne Gallet-Budynek^b, Bruno Bousquet^{a,*}^a CELIA, UMR CNRS 5107, Université de Bordeaux, 351 Cours de la Libération, F-33405 Talence Cedex, France^b ISPA, Bordeaux Sciences Agro, INRA, 33140 Villenave d'Ornon, France

ARTICLE INFO

Keywords:

LIBS
Normalization
Assessment
Good practices
Figures of merit

ABSTRACT

As it is the case for any spectroscopic technique, laser-induced breakdown spectroscopy (LIBS) is strongly influenced by the signal fluctuations, and the LIBS spectra need to be normalized to obtain enhanced analytical performance. Nowadays, normalization in LIBS remains an open question and, in the present review, the normalization methods commonly applied to LIBS are presented and discussed, in particular those based on background, total area, internal standard, and Standard Normal Variate. We emphasize that the figures of merit, namely the coefficient of determination, the root-mean square error of prediction and the limit of quantification used to assess the advantages of processing normalized instead of non-normalized LIBS spectra, in a context of quantification, must be calculated in a rigorous way to be able to draw conclusions. We thus propose advices and good practices to achieve a rigorous comparison between quantitative models involving various normalization approaches, the final choice of the best normalization being ultimately driven by the analytical context. In order to take the best advantage from normalization in LIBS and thus increase the analytical performance of this technique, we encourage the analyst to thoroughly compare different normalization methods.

1. Introduction

As it is the case for any spectroscopic technique, laser-induced breakdown spectroscopy (LIBS) is strongly influenced by signal fluctuations [1]. Indeed, the LIBS spectra are correlated to the plasma properties themselves correlated to: i) the measurement conditions such as laser energy, lens-to-sample distance, and spectrometer calibration, and ii) the physical properties of the sample such as roughness, hardness, and porosity.

Reducing the signal fluctuations allows increasing the quality in LIBS, which is of major importance for the development of this technique. The first strategy in order to reduce fluctuations is to build LIBS experiments with a high degree of control and stabilization. In this case, stabilizing the laser energy, the lens-to-sample distance as well as the spectrometer calibration allows for significantly reducing the signal fluctuations [2]. However, all the laboratory setups don't share the same level of control and thus don't allow for reaching the same reduction of signal fluctuations. In addition, in the case of on-site analyses, involving compact LIBS systems [3], for either stand-off or handheld operation, the degree of control and stabilization is expected to be much lower, and the related level of fluctuations much higher

compared to laboratory instruments. In this context, normalization approaches have been proposed in order to reduce the fluctuations observed in the raw LIBS data. A first type of normalization consists in monitoring other data in addition to the LIBS spectrum, either simultaneously, like the acoustic signal induced by the shock-wave [4–6], or after, like the ablated mass [7]. Then, the LIBS signal is corrected thanks to this additional information, namely divided by a factor related to this external parameter. Then, if the variation of the monitored parameter is correlated with those of the LIBS signal, this normalization process is expected to efficiently reduce the signal fluctuations. However, most of the LIBS instruments don't allow to monitor additional parameters during or after the measurement and the analyst has to treat the raw LIBS signals, without any extra information. In this last case, normalization simply consists in a mathematical transformation of the raw data.

Data normalization is expected to have a positive impact on both quantification and sorting models. Nevertheless, we decided to focus the present discussion only on quantification since this is, by principle, the top objective of any analytical method. In addition, quantification in LIBS can be addressed by building either a calibration model from a series of samples with known concentration values, or a physical model

* Corresponding author.

E-mail addresses: julian.guezenoc@u-bordeaux.fr (J. Guezenoc), bruno.bousquet@u-bordeaux.fr (B. Bousquet).¹ First author

from the plasma properties and spectroscopic data, without any calibration, and well-known as calibration-free (CF-LIBS) [8]. It is interesting to note that signal fluctuations due to a change in plasma properties are naturally taken into account by CF-LIBS. Thus, any change in electron density and temperature induces complex modifications of the calculated LIBS spectra, which cannot be described by a simple multiplicative coefficient. More precisely, the components of the LIBS spectra, namely the continuum as well as the ionic, neutral, and molecular emission lines are changing in non-proportional ways according to the changes in the plasma properties. The CF-LIBS approach thus includes a kind of self-normalization process that enables to correctly interpret each single LIBS spectrum individually. Consequently, in the present review, we will focus our discussion only on the case of calibration models. In this case, the transformations to be applied to the raw LIBS spectra for normalization purposes are quite basic, but to date there is no normalization approach considered as being more efficient than the others to treat the LIBS data. Even more, the advantage of normalization in LIBS is sometimes questionable, making this topic still an open question. In the critical review proposed here, a state-of-the-art about normalization in LIBS is presented in the first section, with a particular emphasis on the four methods of normalization the most often adopted in LIBS, specifically normalization by: i) background, ii) total area, iii) internal standard, and iv) standard normal variate (SNV). Other factors were also tested to normalize the LIBS spectra such as the Euclidian norm, the maximum and the minimum intensity values of each individual spectrum [9–11]. However, the related normalization approaches can be considered as quite similar to the four methods previously mentioned, and they are finally rarely used and not detailed in this review. Then, in the second section of this review, in order to properly compare the performance of the different normalization approaches between them, good practices and significance of selected figures of merit, namely the coefficient of determination (R^2); the root mean square error (RMSE) declined as the root mean square error of calibration (RMSEC) when strictly applied to the calibration set, the root mean square error of prediction (RMSEP) when strictly applied to the validation set not involved in the construction of the model, and the root mean square error of cross-validation (RMSECV) after a cross-validation process; the limit of detection (LOD), and the limit of quantification (LOQ) are presented as a key-step to make accurate decisions. Finally, we have decided to illustrate the present discussion through the quantitative analysis of a selected dataset after different normalization approaches, and we have accurately compared the related figures of merit to conclude if a given normalization method can be finally recommended in a general context.

2. State of the art

The need for normalization concerns not only LIBS but all spectroscopic techniques. And the most efficient strategies of spectra normalization in general are well documented in the scientific literature and routinely applied, especially for Near-Infrared Spectroscopy (NIR), and inductively-coupled plasma atomic emission spectroscopy (ICP-AES), selected here as two emblematic examples of the spectroscopic techniques, and detailed below.

In the case of NIR, the spectra can significantly be influenced by light scattering through a baseline change, known as the multiplicative effect, and non-linear variations of intensity. Therefore, the most used normalization techniques in NIR are divided into two groups [12]: i) the scatter-correction methods, including *Standard Normal Variate* (SNV) and *Multiplicative Scatter Correction*, and ii) the signal derivative based on algorithms such as *Savitzky-Golay* or *Norris-Williams*. The review made by Rinnan et al. [12] about the pre-processing methods in NIR, theoretically describes each of these normalization techniques as well as their advantages and limitations. Nowadays, all the NIR spectra are systematically normalized and the normalization process is well-established [13,14].

In the case of ICP-AES, the plasma fluctuations are eliminated by the use of the internal standard method [15]. It means that a chemical element is chosen for its compatibility with the matrix of interest, and by the presence of an atomic emission line exhibiting good signal-to-noise ratio and without any risk of spectral interference with emission lines related to other elements. The selected element is then used as a dopant and introduced under the same concentration into each solution prepared for quantification. In this context, the elements Yttrium (Y) and Scandium (Sc) are often selected as internal standards for these reasons and also because they are absent from almost all the *real-life* samples. Then, the emission line selected for the normalization process is exploited as follows: its peak intensity is measured for each sample, including the blank; and then, the ratio to the intensity of the blank is calculated for this emission line for each sample. Finally, the peak intensities related to all the other elements are multiplied by this ratio. This method is well-established and has been routinely applied to ICP-AES for many years [16].

Regarding LIBS, it is worth noticing that, despite many attempts of normalization during these last years, there is no clear preference for one type of normalization, the analysts adopting the method considered as the most efficient for their specific application. In 2010, Zorov et al. [5] described the strategies of normalization to be applied to LIBS and other spectroscopic techniques based on laser sampling, such as ICP-AES, atomic absorption spectroscopy and laser-enhanced ionization spectroscopy. They presented examples of LIBS data successfully normalized [17,18] as well as cases for which the normalization was not efficient [19,20], confirming that determining the right normalization method for LIBS is not an easy task. In 2012, Hahn and Omenetto [4] emphasized that it was almost impossible to apply the internal standard method to normalize the LIBS data since many important criteria couldn't be satisfied. Furthermore, Tognoni and Cristoforetti [6] proposed three types of normalization approaches, namely the internal standard method, the normalization by the background and finally a normalization based on the calculation of the plasma properties, similar to the CF-LIBS method. The authors finally recommended to select the most relevant normalization strategy among the three, based on each single experimental context.

To go further into details, an extended selection of articles dealing with normalization in LIBS is presented in Table 1. The aim of this list is to: i) show which normalization methods have been investigated and finally selected, ii) point out that some authors have demonstrated the advantage of treating normalized compared to uncorrected data, while some others haven't, and iii) present the figures of merit that have been used to determine the prediction ability of each model, and thus the advantage of applying normalization.

Seven normalization methods are reported in Table 1, namely normalization to the maximum, minimum, background, total area, Euclidian norm, internal standard and SNV. Details about the calculations related to these methods can be found in the corresponding references. Table 1 also reveals that the comparison between normalized and uncorrected data has not been systematically presented, some of the authors assuming that normalization was expected to provide a priori better analytical results than uncorrected data, but without verifying if this statement is always true. The figures of merit that have been used to assess the normalization methods are also given in this table. First, the reduction of the variations of the LIBS signals was presented by some authors [17,21–24]. The coefficient of determination R^2 has been very often used, either alone, or in complement to other figures of merit such as the root-mean square error (RMSE) declined as RMSEC (calibration set), RMSEP (validation set), and RMSECV (cross-validation), as well as the limits of detection (LOD) and quantification (LOQ). It is worth noticing that these figures of merit should be calculated with an extreme thoroughness since they allow not only for demonstrating if the normalization enables to reach better analytical performances than by uncorrected data, but also to determine if a normalization method provides a better correction than the others. In addition, it is worth

Table 1

Selection of articles dealing with normalization in LIBS. The investigated and selected normalization methods are presented as well as the figures of merit used to assess the prediction ability of each model. IS: internal standard; SNV: standard normal variate; Imax: maximum intensity; Imin: minimum intensity.

Investigated normalization method	Selected normalization method	Comparison normalized/uncorrected data	Figures of merit	References
IS, Background, Total area	IS	yes	RMSECV, LOD	2018, Thomas et al. [27]
IS, Total area, Imax, Norm	IS	yes	R ² , RMSEC	2019, Andrade et al. [28]
Background	Background	no	R ² , LOD	2011, Dell'Aglio et al. [30]
Background	Background	no	R ² , RMSEC	2009, Senesi et al. [31]
Total area	Total area	no	R ² , RSD	2014, Fabre et al. [22]
Total area	Total area	no	R ² , RMSE	2017, Payré et al. [58]
Total area	Total area	yes	R ² , RMSECV	2018, Takahashi et al. [26]
IS	IS	no	RSD	2008, Juvé et al. [21]
IS	IS	yes	R ²	2011, Gupta et al. [45]
IS applied to Fe, Mg, Al, Si	IS (Fe line)	yes	R ²	2009, Kwak et al. [35]
IS applied to Ba, Si, B	IS (Ba line)	no	R ² , RMSECV	2014, Sarkar et al. [36]
SNV	SNV	no	RSD	2011, Ismael et al. [23]
SNV, IS	SNV	yes	R ² , LOD	2015, Syvilay et al. [42]
Total area, IS, Imax, norm	Element dependent	yes	RMSEC, RMSECV, RMSEP, LOD	2016, Castro et al. [10]
Total area, Background	Total area	yes	RSD, RMSEP	2001, Body and Chadwick [17]
Total area, IS, Calibration-free	Element dependent	yes	R ²	2006, Sallé et al. [44]
Total area, IS, Imax, norm	Element dependent	yes	R ² , RMSEC, RMSEP	2017, Augusto et al. [11]
IS, Imax, norm	Element dependent	yes	RMSEC, RMSECV, RMSEP, LOD, LOQ	2016, Andrade et al. [43]
Total area, Background, IS	IS, total area	no	R ² , RMSEP, RSD	2016, Karki et al. [24]
Imin, Imax, IS, norm	Norm	no	REP, RMSEP	2015, Sarkar et al. [9]

pointing out that very recent publications [26–28] were partly dedicated to the selection of the best normalization approach, revealing it is still an open question. From Table 1, it is also interesting to notice that the best normalization method seems to depend on the element and the sample. However, in order to push further the discussion, the principles and advantages of the four most common normalization methods applied to LIBS are presented hereafter.

2.1. Normalization to the background

The background emission is theoretically considered to be a relevant indicator to monitor the plasma properties. In 2008, De Giacomo et al. [32] demonstrated the existence of a correlation between the intensity of the continuum radiation, i.e. the background intensity, and the plasma density, itself correlated to the amount of emitters in the plasma. In the case of normalization to the background, the peak intensity value of the emission line of interest is divided by the value of the background emission. This one is recorded at the vicinity of the emission line of interest, either from the intensity value at a single wavelength, or from the average value over a selected spectral window [6]. It is worth pointing out that the background emission must be carefully recorded. In practice, it means that the detector dark current must be separately recorded and then subtracted prior to such signal processing. This point has been deeply discussed in ref. [6, 63] and is crucial to take into account the actual background signal. Zorov et al. [5] attributed the first results about the use of normalization to the background to Xu et al. [29], but the corresponding work was quickly disapproved by other authors [19], which clearly revealed a lack of generalization ability. However, this type of normalization has been successfully applied to several cases [5,30,31]. In the frame of analysis of soils polluted with heavy metals, Dell'Aglio et al. [30] adopted the normalization to the background for emission lines related to several elements (Cr, Cu, Pb, V, Zn) in order to reduce matrix effects. They concluded that the correlation between the data derived from ICP-AES and the normalized LIBS signals was quite satisfactory. Senesi et al. [31] also adopted the normalization to the background to quantify the concentrations of heavy metals in soils, with a particular focus on chromium. However, despite significant reduction of the fluctuations and high correlation between the normalized LIBS and ICP-AES data, it is very important to notice that only 3 to 4 points were taken into

account to build the quantitative models in this study. The conclusions are thus likely to be non-robust.

2.2. Normalization to the total area

In the case of normalization to the total area, for each LIBS spectrum, the peak intensity value of the emission line related to the analyte is divided by the value of area of the spectrum over the whole spectral range. Here again, the detector dark current must be separately measured and subtracted prior to such signal processing, to deal with real LIBS spectra. The total area is calculated as the sum of all the intensity levels describing the LIBS spectrum, once the detector dark current has been removed, and is considered to be a good indicator of the plasma properties. Thus, Body and Chadwick [17] demonstrated the existence of a strong correlation between this area and the laser energy. Then, Fabre et al. [22] applied this normalization method to the analysis of Martian rocks and soils by ChemCam onboard the MSL Curiosity rover. In this case, three separate spectrometers were operated and the normalization to the total area was implemented by considering only the spectrometer for which the emission line of the analyte was detected. Therefore, all the results of quantification obtained from the ChemCam data, and based on Partial Least Squares regression (PLS) for the major elements and univariate analysis for the minor and traces, were entirely based on normalized data. In 2010, Zorov et al. [5] reported that J.A. Bolger [33] proposed the normalization method to the total area for the first time in LIBS, which allowed to improve the linearity of the calibration curves of Cu, Fe, Ni, Mn and Cr in mineral rocks. However, it is interesting to notice again that the calibration curves were built from a very small number of points, making the conclusion of this study potentially non-robust. Finally, Yu et al. [25], in the context of sorting certified soils by chemometric tools, examined the compensation of matrix effects and changes in experimental conditions, mainly temperature, through the normalization to the total area. They presented this normalization as a standard step of pre-processing but without comparing the corresponding analytical results to those obtained from uncorrected data. More generally, most of the articles about the normalization to the total area have concluded that this pre-processing step provides a benefit.

2.3. Normalization to an internal standard

The internal standard (IS) method is frequently adopted by the LIBS analysts. It consists in dividing the peak intensity (or area) of the emission line related to the analyte by the peak intensity (or area) of a selected emission line related to the internal standard. The concentration of the internal standard is supposed to remain constant for the whole set of samples, or at least to be known by the analyst. The underlying hypothesis is that the variations of the peak intensity related to the internal standard are only due to changes in the plasma properties. This is expected to be true when all the samples considered to build a calibration model are characterized by a single matrix in order to avoid any risk of bias linked to matrix effects. But it is interesting to notice also that this normalization method can be indifferently applied to any kind of matrix [21,36,45,60–62]. Moreover, it requires the use of another analytical technique to measure the concentration values of the internal standard inside each sample.

The list of criteria defined by Barnett et al. [34] to be verified in order to identify a reliable internal standard, have been applied to LIBS [4]. Three of these criteria are related to the choice of the element playing the role of internal standard, since the analyte and the internal standard are expected to share similar: i) volatilization rates; ii) ionization energies; and iii) atomic weights. Then, three other criteria are focused on the selection of the spectral line, which is expected to: i) have the same excitation energy than the one of the analyte; ii) be not affected by self-absorption; iii) have a similar intensity as the one of the analyte. In practice, this list of criteria seems to be almost impossible to fulfill, especially in the case of natural and complex matrices, driving to conclude that the internal standard method is generally not ideal to LIBS analysis. Despite this statement, several studies have concluded on the advantages of applying the internal standard method to LIBS, even if the choice of the internal standard according to the criteria previously mentioned was not always clearly presented. Thus, Juvé et al. [21], selected the spectral line of carbon at 247.86 nm considered as an internal standard, to process the space-resolved analysis of trace elements (Mg, Al, Ca, Ti, Mn, Fe) in fresh vegetables. The choice of carbon was based on the fact that it was found under constant concentration in the studied samples but no reference to the selection criteria neither for the elements nor for the spectral lines can be found in their article. Thus, the advantage of adopting internal standard normalization reported in this article might be due to chance. Similarly, Kwak et al. [35] did not consider the criteria related to the selection of internal standard. In the case of quantification of arsenic in soils, they normalized the intensity of the As I line at 228.812 nm to the one of the Fe I line at 248.8 nm and also to the Mg I line at 285.2 nm, then the Al I line at 309.3 nm, and also the Si I line at 288.2 nm. The only criterion to select these lines was the increase of the coefficient of determination. Indeed, starting from 0.22 without normalization, $R^2 = 0.96$ after a normalization to Fe, 0.91 with Mg, 0.75 with Al, and 0.58 with Si, demonstrating the advantage of normalizing the LIBS signal to the Fe I line at 248.8 nm to reach the best performance for the calibration model. Moreover, Sarkar et al. [36], in the context of quantification of uranium in a barium borosilicate glass matrix, selected barium as internal standard, after comparison with other elements, such as silicon and boron. Then, they selected the emission line of Ba I at 649.786 nm, which was satisfying to all the given criteria. Finally, by comparing the figures of merit (R^2 , RMSECV and the slope of the calibration curve), obtained after this normalization to the selected Ba I line, to the ones derived from other elements and spectral lines, they concluded that the best normalization according to theoretical criteria was also the best one from the experimental results.

To conclude, normalization to an internal standard has been demonstrated to reduce the LIBS signal fluctuations and thus increase the analytical performance of the related quantitative models. However, in addition to the cases exhibiting a successful effect of this type of normalization on the analytical performance, it is worth noticing that LIBS

is also often applied to analysis of samples for which it is impossible to find a proper internal standard, according to the list of criteria previously discussed. This normalization approach can't thus be considered as a general strategy for LIBS analysis, and should only be applied to specific cases.

2.4. Standard normal variate - SNV

Standard Normal Variate (SNV) is one of the most used normalization techniques in near-infrared spectroscopy (NIR) [37–39] and Raman spectroscopy [40,41]. SNV calculation is processed on each spectrum individually by centering the spectrum on the mean value and then dividing the corrected spectrum by the standard deviation of the original spectrum (Eq. (1)):

$$I_k^{\text{snv}} = \frac{I_k - I_{\text{mean}}}{s} \quad (1)$$

where I_k represents the intensity value at the wavelength k of the original spectrum; I_k^{snv} , the corresponding SNV normalized intensity value; I_{mean} , the mean of the intensity values of the original spectrum; and s , the standard deviation of the original spectrum.

In the context of in-situ quantitative LIBS analyses of polluted soils involving Pb, Cu and Fe, Ismaël et al. [23] obtained good correlations between the SNV-corrected LIBS data and the reference concentration values derived from ICP-AES, for Pb (405.78 nm), Cu (327.39 nm) and Fe (344.08 nm) and observed that SNV allowed to reduce the standard deviation after a series of repetitions. Then, Syvilay et al. [42] demonstrated in the context of quantitative analysis of Ag, Bi, Cu and Sn in lead samples, that the univariate models built from SNV-corrected data provided better figures of merit (R^2 and LOD) than those built from either raw data or data normalized to a selected emission line of lead, considered as internal standard. To conclude, the examples of LIBS studies reported in Table 1 indicate that SNV might be an efficient normalization technique for processing the LIBS data and we point out that the advantages of the SNV normalization is worth being accurately assessed in future works.

As a general conclusion, the four normalization methods presented in the present review have been successfully applied to LIBS data, meaning they have permitted to reduce the signal fluctuations and by consequence, improve the calibration performances.

3. Recommendations

To support the advantages of normalizing, we recommend to systematically compare the figures of merit related to the models built from either normalized or uncorrected data. Indeed, from Table 1, one can notice that this simple comparison was not made in all the cases. In addition, it is worth pointing out that, in some cases, normalization might negatively impact the analytical performance of the models [19,11,36,]. Based on the works reported in Table 1, it is not possible to conclude if one of the normalization methods might be advantageously adopted whatever the element and application. Thus, we recommend the LIBS analyst to test different normalization methods and then adopt the one offering the best analytical performance. This strategy of comparing different normalization methods has been adopted by some authors [9–11,24,43], who concluded that, in some cases, uncorrected data provide better analytical performances than normalized ones, and also that the normalization method should be selected on a case-to-case basis. Regarding the figures of merit calculated to assess the performance of each model, the coefficient of determination R^2 has been widely used. However, in some cases, R^2 was the only indicator considered for the assessment [35,44,45], which is clearly insufficient to establish robust conclusions, as it will be further discussed in this section.

3.1. Minimum number of points to build a reliable calibration model

In order to determine what is the best normalization approach for LIBS, we recommend to strictly apply the basic rules allowing to calculate correct values for the figures of merit dedicated to assess the analytical performances of the models. Our first advice concerns the minimum number of points (concentration values) to build a robust linear calibration model. This number is not defined in a unique way by the organizations in charge of standardization and traceability of chemical measurements. The ISO standard 11,095:1996 [46], namely the standard dedicated to linear calibration using reference materials published by the International Organization for Standardization (ISO), advises to include more than three points to build a reliable model. Moreover, through a guide named “The Fitness for Purpose of Analytical Methods” [47], the European network of organizations EURACHEM, promoting good quality practices about chemical measurements, recommends to include a minimum of 6 concentration levels plus blank. In the same way, a commission decision (2002/657/EC), edited by the commission of the European communities [48], about the performance of the analytical methods and the interpretation of results advises to include at least 5 points, i.e. concentration levels, including zero. Based on these references, the minimum number of points to build a reliable calibration model should be around 5 or 6. Furthermore, these points should be, ideally, equally distributed over the concentration range related to the calibration to obtain the most representative calibration model related to the analytical context. But, despite of its major importance for quality purpose, this last criterion is generally underestimated or simply not considered by the LIBS analysts. We have thus decided to illustrate how the use of an unappropriated dataset to build the calibration models might drive to non-robust conclusions.

3.2. Coefficient of determination, R^2

The coefficient of determination, R^2 , is the figure of merit the most commonly used (cf. Table 1) to assess the quality of a calibration model. However, this indicator is reliable only when the points are equally distributed over the concentration range and when there is no extreme point, very far from the other ones. At the opposite, if one of these conditions is not fulfilled, the decision based only on the parameter R^2 might be non-robust and this statement seems not always be taken into account in the LIBS literature [31,33], still nowadays [49].

Fig. 1 illustrates the case of univariate calibration models built from a set of 36 LIBS spectra. This dataset has been specially selected to illustrate the case of non-satisfying conditions to build a robust model. Thus, all the details about the samples, the analyte, and the experimental setup are voluntary not provided here, to highlight the general conclusions of this discussion and not consider it as a specific case-study. For each model displayed in Fig. 1, the baseline-subtracted signal (in arbitrary units, a.u.) was plotted on the vertical axis as a function of

the concentration of the analyte ($\mu\text{g/g}$) on the horizontal axis in three cases: a) without normalization; b) after normalization to the total area of the spectrum; c) after normalization by SNV. In agreement with the a priori expectation, the R^2 values are higher after normalization. However, it is worth noticing that, considering a range of calibration from 0 to 8000 $\mu\text{g/g}$, 89% of the points of the calibration set correspond to concentration values lower than 1500 $\mu\text{g/g}$. Moreover, for these points, whatever the normalization process, there is no correlation between the LIBS signal and the concentration values. In such conditions, the increase of the R^2 value from 0.52 (without normalization) to 0.84 (SNV) should be considered with caution, since the regression models over the full concentration range are, in this case, not robust. Indeed, none of the points from the calibration set is between 1500 and 2500 $\mu\text{g/g}$ and between 4500 and 7500 $\mu\text{g/g}$. Moreover, in such conditions, the coefficient of determination value is highly sensitive to extreme points, considered as outliers [50]. For instance, in this case, the last point of the calibration curve corresponds to a concentration value of 7500 $\mu\text{g/g}$, quite far from the previous one, which corresponds to 4500 $\mu\text{g/g}$. Indeed, after removing this last point from the calibration set, the new R^2 values become 0.38, 0.53 and 0.59, for the non-normalized data, data normalized to the total area, and data normalized by SNV, respectively. Finally, the example given here aims at warning the LIBS analyst about the danger of looking only at R^2 [51] to characterize the calibration performances of a model, whatever the method of normalization. It also means that outliers should induce additional investigation through either new measurements, or new steps in data processing.

3.3. Root Mean Squared Error of prediction (RMSE)

In addition to the R^2 parameter, the prediction ability has to be determined. It consists in calculating the common Root Mean Squared Error (RMSE) [52] resulting from the predicted concentration values from a series of samples. This indicator is expected to be as low as possible for a reliable model, and is mathematically written as in Eq. (2).

$$RMSE = \sqrt{\frac{\sum_{p=1}^N (c_p - \hat{c}_p)^2}{N}} \quad (2)$$

Where c_p is the predicted concentration value, $\hat{c}_p$ the reference concentration value of the sample p , and N the number of samples taken into account. It is worth noticing that, similarly to the case of R^2 , this figure of merit should be considered with extreme attention, as discussed below.

When all the samples belonging to the calibration (C) set are simultaneously taken into account, the resulting RMSE value is referred to as RMSEC. However, RMSEC might not be a reliable indicator of the calibration model in case of the presence of extreme points, namely points exhibiting very high concentration values compared to the other ones [52]. To overcome this drawback, the RMSE value should be

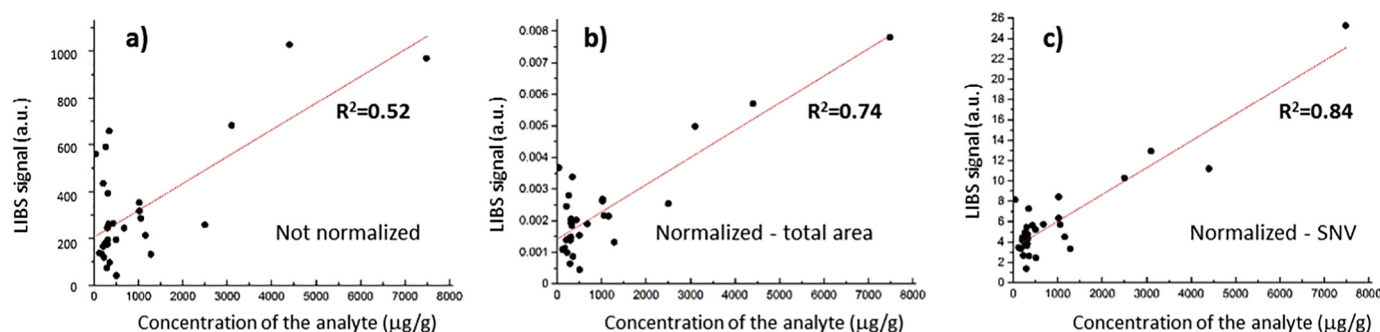


Fig. 1. Linear calibration models (red lines) obtained (after baseline-subtraction) for: a) not normalized data, b) data normalized to the total area, and c) data normalized by SNV, with the corresponding R^2 values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

calculated on the basis of a cross-validation (CV) process, and referred to as *RMSECV*. Among the different techniques of cross-validation, the Leave-One-Out (LOO) method is the most common. It consists in removing one point from the dataset and then calculate the *RMSEC* value from the regression built from the N-1 remaining points. Then the excluded point is reintroduced, another point is excluded, and a new *RMSEC* value is calculated, and so on, until all the points of the calibration set have been excluded. Finally, *RMSECV* is the average value of the N values of *RMSEC* calculated during the cross-validation process. *RMSECV* is expected to better reveal the limitations of the calibration model than *RMSEC*. When the calibration set is big, another method of cross-validation, known as *bootstrapping*, is generally preferred to the LOO, and consists in excluding not only one single point at a time but many. Finally, to fully assess a calibration model, it is necessary to predict the concentration values of a set of *known* samples different from the calibration set. Thus, it has been suggested [53] to split the original dataset into the calibration set (2/3) and the validation set (1/3), both of them equally spread over the concentration range. Then, the *RMSE* value is calculated over the validation set, and referred as *RMSEP*, with (P) related to the prediction of the concentration values of the samples belonging to the validation set. Comparing *RMSEC* and *RMSEP* allows to evaluate if there is a risk of overfitting (if $RMSEC \ll RMSEP$), before comparing different strategies of normalization. Note also that *RMSECV* and *RMSEP* are commonly used in the frame of multivariate regression, not only to evaluate the model performances but also to select the best number of components, which can easily become user dependent. Finally, the validation step in multivariate analysis is not the aim of this present review, but many information can be found in the following studies [52–54].

3.4. Limits of detection and quantification, LOD and LOQ

The limits of detection (LOD) and quantification (LOQ) have been discussed by Mermet in review papers [55,56] dedicated to atomic spectrometry. In LIBS, the LOD is the most frequently reported, and defined by Eq. (3) [56], for a regression model written as: $y = a + b \cdot x$.

$$LOD = \frac{3s}{b} \quad (3)$$

where s is the standard deviation of the background signal measured close to the peak of interest. Similarly, the limit of quantification is defined by Eq. (4) [55,56]:

$$LOQ = \frac{10s}{b} = \frac{10}{3} LOD \quad (4)$$

The limits of detection and quantification are calculated from the regression model. And, as already stated for the calculation of R^2 and *RMSE*, the calibration set needs to fulfill basic requirements to obtain reliable results for the values of these limits.

Finally, the figures of merit R^2 , *RMSE*, LOD (or LOQ) are perfect to

assess the analytical performance of the models as well as the advantage of normalization [57]. In the next section, we exploit these figures of merit to compare, for a dataset recorded in our laboratory with a handheld LIBS instrument, the analytical performance obtained by i) non-normalized spectra, ii) spectra normalized to the total area, and iii) spectra normalized by SNV.

4. Example of normalization

The general discussion about normalization in LIBS has been presented in the state-of-the-art of this review, and then important recommendations that deserve to be completed by a case-study. It aims at illustrating the question of normalization and allows us to propose a general guideline to assess and then report the analytical performance of the quantitative models based on different normalization methods. Let's mention that this guideline could efficiently be applied to any spectroscopic technique and normalization method but it appears to be of major importance for LIBS. We consider a first set of 14 samples of plants (calibration set), sampled from field crops in France, and stored in a sample bank managed by the French National Institute for Agricultural Research (INRA), in the frame of the French national observatory network Quasaprove [59]. More precisely, these samples were analyzed by a portable LIBS instrument and the analyses presented here were focused on the quantification of magnesium. Each sample was dried and then ground by using a planetary ball mill (PM 400, Retsch), allowing to obtain a grain size smaller than 250 μm . Then, each sample was prepared as a 12-mm diameter pressed pellet after 8 tons load during 3 min without using any additional binder. 10 locations were randomly chosen at the surface of the sample. And at each location, 3 laser shots were considered as pre-shots and then the LIBS spectra related to the following 7 shots were recorded. The 10 locations and the 7 spectra per location gives access to 70 LIBS spectra per sample. The average spectrum over the 70 spectra was considered for the analyses. At each wavelength, the RSD value from the 70 spectra has also been calculated and reveals that each sample can be considered as homogeneous, which confirms the observation one can make with the naked eyes.

Considering the quantification of magnesium as an example for the present guideline, the concentration values of the 14 samples composing the calibration set ranged from 0.7 to 8 $\mu\text{g/g}$, according to the reference ICP-AES analyses. In addition, 4 other samples of plant, from the same bank, and not used for the calibration, were selected to validate the calibration models, based on their availability and their concentration values measured by ICP-AES.

Fig. 2 displays calibration models built from; a) the baseline subtracted peak intensity value of Mg II at 280.27 nm, b) normalization by the total area of the spectrum, c) normalization by SNV. It should be emphasized that the present discussion is focused on the methodology one should apply to be able to objectively compare the different

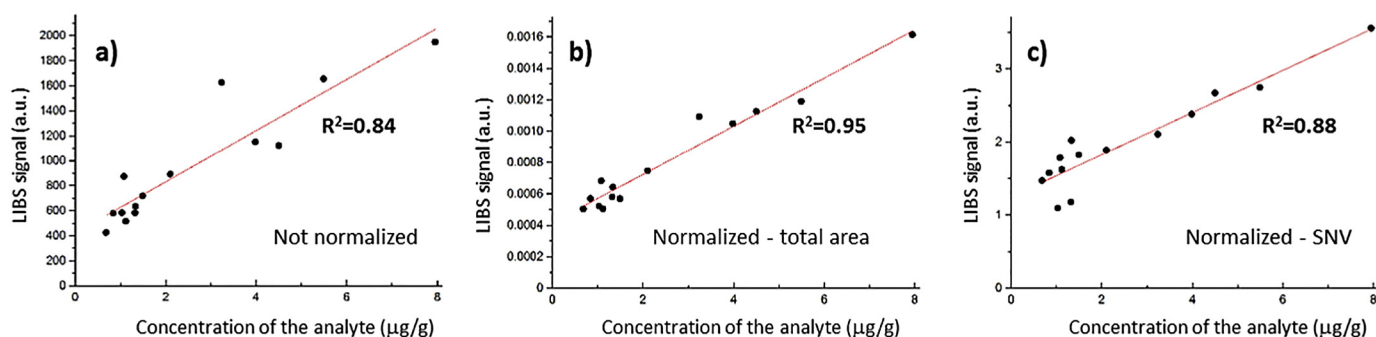


Fig. 2. Linear calibration models (red lines) obtained (after baseline-subtraction) for: a) not normalized data, b) data normalized to the total area, and c) data normalized by SNV, with the corresponding R^2 values. In this case-study, the analyte was magnesium and the samples were plants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

Figures of merit calculated to assess the performance of regression models involving different normalization strategies.

	R^2	RMSEC($\mu\text{g/g}$)	RMSECV($\mu\text{g/g}$)	RMSEP($\mu\text{g/g}$)	LOD($\mu\text{g/g}$)
Not-normalized	0.84	0.90	0.89	0.74	0.20
Total area	0.95	0.45	0.45	0.57	0.24
SNV	0.88	0.76	0.75	1.42	0.32

methods of normalization between them. Thus, even if the models based on normalization to the background and internal standard could have also been presented in Fig. 2, we have decided to select only two methods of normalization to illustrate our purpose. These two normalization methods were expected to reduce the fluctuations and thus to improve the analytical performance of the models, according to previous works [22–26,58], and this result is again verified in the case of plant samples displayed in Fig. 2. Indeed, the points related to the concentrations of magnesium larger than 3 $\mu\text{g/g}$ are much closer to the linear regression curve after normalization. At the opposite, two of the points related to concentration values close to 1 $\mu\text{g/g}$ and normalized by SNV are displayed far from the regression line and the other points.

Table 2 displays the values of the figures of merit dedicated to the assessment of the quantitative models, specifically R^2 , RMSEC, RMSECV, RMSEP, LOD (reminding that LOD and LOQ are closely linked, no need to report also the values of LOQ), calculated for the three regression models displayed in Fig. 2. As already discussed, even if the R^2 values are higher after normalization, it might be risky to draw a conclusion based on this single parameter. Indeed, one can observe in Table 2 that while R^2 is rising from 0.84 (without normalization) to 0.88 (SNV), RMSEP is unexpectedly increasing from 0.74 to 1.42, indicating that the regression model built after SNV normalization exhibits a lower predictive ability than the model built from non-normalized data. The SNV model might suffer from over-fitting since both the RMSEC and RMSECV values are half the RMSEP value. At the opposite, the values reported in Table 2 related to the normalization to the total area reveal for this model stronger correlation (higher R^2) and better prediction ability (lower errors of prediction). Finally, even if the values of R^2 , RMSEC, RMSECV, and RMSEP are better after normalization to the total area, the values of LOD are not significantly different from data without normalization and after normalization.

In the case of the models reported in Table 2, the normalization to the total area allows for building a calibration model with a better coefficient of determination and smaller prediction errors. This type of normalization might thus be adopted here even if it doesn't allow for lowering the limits of detection and quantification. At the opposite, the data normalized by SNV drive to worse prediction ability, namely a higher RMSEP value, than non-normalized data.

More generally, the normalization to the total area has been considered as valuable not only in the case reported in Table 2 but also in many previous studies. It is thus a priori expected that this type of normalization should improve the quantification ability of any LIBS analysis. However, we recommend to demonstrate its advantage a posteriori, from a comparison of the values taken by the figures of merit reported in Table 2. At the opposite, the normalization by SNV appears to be more questionable. Some previous studies have demonstrated its advantage but it was not the case in the frame of the case-study reported in Table 2.

We finally recommend, whatever the types of samples and elements to be quantified, to always compare the figures of merit related to the models calculated from not-normalized and normalized spectra, as displayed in Table 2. Let's point out that, in addition to the normalization to the total area and by SNV, the normalization to the background and to an internal standard should also be considered. Based on the methodology presented here, not only the best normalization

method can be objectively selected, but the calculated values of the figures of merit can also be directly exploited to report the analytical performance of the selected model.

5. Conclusion

LIBS spectra are often normalized in order to take into account signal fluctuations and thus obtain enhanced analytical performance. However, normalization in LIBS remains an open question and, in the present review, we have identified and discussed the normalization methods commonly applied to LIBS. We have more specifically detailed and discussed four of them, that we consider as the most relevant, namely i) background, ii) total area, iii) internal standard, and iv) SNV. We have also highlighted that the gain resulting from normalizing LIBS spectra was often not assessed and that the rare comparisons between different normalization methods were sometimes questionable. Indeed, this review has also revealed that basic requirements to calculate robust indicators were sometimes not fulfilled. A first case-study, based on the selection of LIBS spectra and concentration values unsuitable to build a robust calibration, has been discussed to emphasize the risk of taking a decision about the advantage of normalization from only the R^2 value. Finally, advices and good practices have been proposed in order to correctly calculate the relevant figures of merit dedicated to the assessment of the performance of the quantitative models. Obviously, an efficient normalization method is expected to provide higher R^2 values, and lower RMSE and LOD or LOQ values than those obtained without normalization. However, through a second case-study, related to the quantification of magnesium in plant samples, we have demonstrated that these three figures of merit were not always simultaneously improved. Then, considering that the analytical context naturally introduces some hierarchy within the figures of merit, in the context of quantification of major or abundant elements, reducing the RMSE value becomes the top priority and reducing the LOD or LOQ value can be of secondary importance. At the opposite, for the quantification of minor or trace elements, the normalization method should be able to reduce the LOD or LOQ value while keeping a good predictive ability, i.e. a small RMSE value. Based on this methodology, it becomes possible to objectively compare different normalization methods between each other and finally select the best one according to the analytical context. Future works may adopt this methodology in order to take the best advantage from normalization in LIBS, and thus increase the analytical performance of this technique.

Acknowledgements

The authors acknowledge the regional council of *Région Nouvelle Aquitaine* for the funding of J. Guezenc's doctoral position in the frame of the project AGROLIBS, and Dr. Laurence Denaix for having given access of samples from the Quasaprove network.

References

- [1] D.W. Hahn, N. Omenetto, Laser-induced breakdown spectroscopy (LIBS), part I: review of basic diagnostics and plasma-particle interactions: still-challenging issues within the analytical plasma community, *Appl. Spectrosc.* 64 (2010) 335A–366A.
- [2] V. Motto-Ros, E. Negre, F. Pelascini, G. Panczer, J. Yu, Precise alignment of the collection fiber assisted by real-time plasma imaging in laser-induced breakdown spectroscopy, *Spectrochim. Acta B* 92 (2014) 60–69.
- [3] G. Galbacs, A critical review of recent progress in analytical laser-induced breakdown spectroscopy, *Anal. Bioanal. Chem.* (2015), <https://doi.org/10.1007/s00216-015-8855-3>.
- [4] D.W. Hahn, N. Omenetto, Laser-induced breakdown spectroscopy (LIBS), part II: review of instrumental and methodological approaches to material analysis and applications to different fields, *Appl. Spectrosc.* 66 (2012) 347–419.
- [5] N.B. Zorov, A.A. Gorbatenko, T.A. Labutin, A.M. Popov, A review of normalization techniques in analytical atomic spectrometry with laser sampling: from single to multivariate correction, *Spectrochim. Acta B* 65 (2010) 642–657.
- [6] E. Tognoni, G. Cristoforetti, Signal and noise in laser induced breakdown spectroscopy: an introductory review, *Opt. Laser Technol.* 79 (2016) 164–172.
- [7] S.I. Gornushkin, I.B. Gornushkin, J.M. Anzano, B.W. Smith, J.D. Winefordner, Effective normalization technique for correction of matrix effects in laser-induced

- breakdown spectroscopy detection of magnesium in powdered samples, *Appl. Spectrosc.* 56 (4) (2002) 433–436.
- [8] E. Tognoni, G. Cristoforetti, S. Legnaioli, V. Palleschi, Calibration-free laser-induced breakdown spectroscopy: state of the art, *Spectrochim. Acta B* 65 (2010) 1–14.
 - [9] A. Sarkar, V. Karki, S.K. Aggarwal, G.S. Maurya, R. Kumar, A.K. Rai, X. Mao, R.E. Russo, Evaluation of the prediction precision capability of partial least squares regression approach for analysis of high alloy steel by laser induced breakdown spectroscopy, *Spectrochim. Acta B* 108 (2015) 8–14.
 - [10] J.P. Castro, E.R. Pereira-Filho, Twelve different types of data normalization for the proposition of classification, univariate and multivariate regression models for the direct analyses of alloys by laser-induced breakdown spectroscopy (LIBS), *J. Anal. At. Spectrom.* 31 (2016) 2005–2014.
 - [11] A. dos Santos Augusto, P.L. Barsanelli, F. Manhas Verbi Pereira, E.R. Pereira-Filho, Calibration strategies for the direct determination of Ca, K, and Mg in commercial samples of powdered milk and solid dietary supplements using laser-induced breakdown spectroscopy (LIBS), *Food Res. Int.* 94 (2017) 72–78.
 - [12] Å. Rinnan, F. Van den Berg, S. Balling Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, *Trends Anal. Chem.* 28 (10) (2009).
 - [13] W. Li, L. Xing, L. Fang, J. Wang, H. Qu, Application of near infrared spectroscopy for rapid analysis of intermediates of Tanreqing injection, *J. Pharm. Biomed. Anal.* 53 (2010) 350–358.
 - [14] D. Roman-Ospino Andrés, Singh Ravendra, Ierapetritou Maranthi, Ramachandran Rohit, Méndez Rafael, CarlosOrtega-Zuñiga, J. Muzzio Fernando, J. Románach Rodolfo, Near infrared spectroscopic calibration models for real time monitoring of powder density, *Int. J. Pharm.* (2016), <https://doi.org/10.1016/j.ijpharm.2016.08.029>.
 - [15] F. Pilon, A. Labet, K. Vielle, O. Vigneau, G. Granier, P. Biennu, J.M. Mermet, La validation de méthode en spectrométrie d'émission optique à source plasma (ICP-OES): de l'échantillon au résultat, *EDP Sciences*, 2017.
 - [16] M.F. Gazulla, M. Rodrigo, M. Orduna, M.J. Ventura, C. Andreu, High precision measurement of silicon in naphthas by ICP-OES using isooctane as diluent, *Talanta* 164 (2017) 563–569.
 - [17] D. Body, B.L. Chadwick, Optimization of the spectral data processing in a LIBS simultaneous elemental analysis system, *Spectrochim. Acta B* 56 (2001) 725–736.
 - [18] H. Kurniawan, M.M. Sulyanti, T.J. Lie, K. Kagawa, M.O. Tjia, Application of primary plasma standardization to Nd-YAG laser-induced shock wave plasma spectrometry for quantitative analysis of high concentration Au-Ag-Cu alloy, *Spectrochim. Acta B* 56 (2001) 1407–1417.
 - [19] I.B. Gornushkin, B.W. Smith, G.E. Potts, N. Omenetto, J.D. Winefordner, Some considerations on the correlation between signal and background in laser-induced breakdown spectroscopy using single-shot analysis, *Anal. Chem.* 71 (1999) 5447–5449.
 - [20] B.C. Castle, K. Talabardon, B.W. Smith, J.D. Winefordner, Variables influencing the precision of laser-induced breakdown spectroscopy measurements, *Appl. Spectrosc.* 52 (1998) 649–657.
 - [21] V. Juvé, R. Portelli, M. Boueri, M. Baudelet, J. Yu, Space-resolved analysis of trace elements in fresh vegetables using ultraviolet nanosecond laser-induced breakdown spectroscopy, *Spectrochim. Acta B* 63 (2008) 1047–1053.
 - [22] C. Fabre, A. Cousin, R.C. Wiens, A. Ollila, O. Gasnault, S. Maurice, V. Sautter, O. Forni, J. Lasue, R. Tokar, D. Vaniman, N. Melikechi, In situ calibration using univariate analyses based on the onboard ChemCam targets: first prediction of Martian rock and soil compositions, *Spectrochim. Acta B* (2014), <https://doi.org/10.1016/j.sab.2014.03.014>.
 - [23] A. Ismael, B. Bousquet, K. Michel-Le Pierrès, G. Travaillé, L. Canioni, S. Roy, In situ semi-quantitative analysis of polluted soils by laser-induced breakdown spectroscopy (LIBS), *Appl. Spectrosc.* 65 (2011) 467–473.
 - [24] V. Karki, A. Sarkar, M. Singh, G. Maurya, R. Kumar, A.K. Rai, S.K. Aggarwal, Comparison of spectrum normalization techniques for univariate of stainless steel by laser-induced breakdown spectroscopy, *Pramana* 86 (2016) 1313–1327.
 - [25] K.G. Yu, Y.R. Zhao, F. Liu, Y. He, Laser-induced breakdown spectroscopy coupled with multivariate chemometrics for variety discrimination of soil, *Sci. Rep.* 6 (2016) article number: 27574.
 - [26] T. Takahashi, B. Thornton, T. Sato, T. Ohki, K. Ohki, T. Sakka, Partial least squares regression calculation for quantitative analysis of metals submerged in water measured using laser-induced breakdown spectroscopy, *Appl. Opt.* 57 (20) (2018).
 - [27] N.H. Thomas, B.L. Ehlmann, D.E. Anderson, S.M. Clegg, O. Forni, S. Schröder, W. Rapin, P.-Y. Meslin, J. Lasue, D.M. Delapp, M.D. Dyar, O. Gasnault, R.C. Wiens, S. Maurice, Characterization of hydrogen in basaltic materials with laser-induced breakdown spectroscopy (LIBS) for application to MSL ChemCam data, *J. Geophys. Res.* 123 (2018), <https://doi.org/10.1029/2017JE005467>.
 - [28] D.F. Andrade, F.M. Fortunato, E.R. Pereira-Filho, Calibration strategies for determination of the In content in discarded liquid crystal displays (LCD) from mobile phones using laser-induced breakdown spectroscopy (LIBS), *Anal. Chim. Acta* (2019), <https://doi.org/10.1016/j.aca.2019.02.038>.
 - [29] L. Xu, V. Bulatov, V.V. Gridin, I. Schechter, Absolute analysis of particulate materials by laser-induced breakdown spectroscopy, *Anal. Chem.* 69 (1997) 2103–2108.
 - [30] M. Dell'Aglio, R. Gaudioso, G.S. Senesi, A. De Giacomo, C. Zacccone, T.M. Miano, O. De Pascale, Monitoring of Cr, Cu, Pb, V and Zn in polluted soils by laser induced breakdown spectroscopy (LIBS), *J. Environ. Monit.* 13 (2011) 1422.
 - [31] G.S. Senesi, M. Dell'Aglio, R. Gaudioso, A. De Giacomo, C. Zacccone, O. De Pascale, T.M. Miano, M. Capitelli, Heavy metal concentrations in soils as determined by laser-induced breakdown spectroscopy (LIBS), with special emphasis on chromium, *Environ. Res.* 109 (2009) 413–420.
 - [32] A. De Giacomo, M. Dell'Aglio, O. De Pascale, R. Gaudioso, A. Santagata, R. Teghil, Laser-induced breakdown spectroscopy methodology for the analysis of copper based alloys used in ancient artworks, *Spectrochim. Acta B* 63 (2008) 585–590.
 - [33] J.A. Bolger, Semi-quantitative laser-induced breakdown spectroscopy for analysis of mineral drill core, *Appl. Spectrosc.* 54 (2000) 181–189.
 - [34] W.B. Barnett, V.A. Fassel, R.N. Kniseley, Theoretical principles of internal standardization in analytical emission spectroscopy, *Spectrochim. Acta B* 23 (1968) 643.
 - [35] J.H. Kwak, C. Lenth, C. Salb, E.J. Ko, K.W. Kim, K. Park, Quantitative analysis of arsenic in mine tailing soils using double pulse-laser induced breakdown spectroscopy, *Spectrochim. Acta B* 64 (2009) 1105–1110.
 - [36] A. Sarkar, R.K. Mishra, C.P. Kaushik, P.K. Wattal, D. Alamelu, S.K. Aggarwal, Analysis of barium borosilicate glass matrix for uranium determination by using ns-IR-LIBS in air and Ar atmosphere, *Radiochim. Acta* 102 (9) (2014) 805–812.
 - [37] Y. Bi, K. Yuan, W. Xiao, J. Wu, C. Shi, J. Xia, G. Chu, G. Zhang, G. Zhou, A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation, *Anal. Chim. Acta* 909 (2016) 30–40, <https://doi.org/10.1016/j.aca.2016.01.010>.
 - [38] S. Buratti, N. Sinelli, E. Bertone, A. Venturello, E. Casiraghi, F. Geobaldo, Discrimination between washed Arabica, natural Arabica and Robusta coffees by using near infrared spectroscopy, electronic nose and electronic tongue analysis, *J. Sci. Food Agric.* (2014), <https://doi.org/10.1002/jsfa.6933>.
 - [39] Y. Guo, Y. Ni, S. Kokot, Evaluation of chemical components and properties of the jujube fruit using near infrared spectroscopy and chemometrics, *Spectrochim. Acta A* 153 (2016) 79–86.
 - [40] P. Heraud, B.R. Wood, J. Beardall, D. McNaughton, Effects of pre-processing of Raman spectra on in vivo classification of nutrient status of microalgal cells, *J. Chemom.* 20 (5) (2006) 193–197.
 - [41] S. Romero-Torres, J.D. Pérez-Ramos, K.R. Morris, E.R. Grant, Raman spectroscopy for tablet coating thickness quantification and coating characterization in the presence of strong fluorescent interference, *J. Pharmaceut. Biomed.* 41 (3) (2006) 811–819.
 - [42] D. Syvilay, N. Wilkie-Chancellor, B. Trichereau, A. Texier, L. Martinez, S. Serfaty, V. Detalle, Evaluation of the standard normal variate method for laser-induced breakdown spectroscopy data treatment applied to the discrimination of painting layers, *Spectrochim. Acta Part B* 114 (1 December 2015) 38–45.
 - [43] D.F. Andrade, E.R. Pereira-Filho, Direct determination of contaminants, major and minor nutrients in solid fertilizers using laser-induced breakdown spectroscopy (LIBS), *J. Agric. Food Chem.* 64 (41) (2016) 7890–7898.
 - [44] B. Sallé, J.L. Lacour, P. Mauchien, P. Fichet, S. Maurice, G. Manhès, Comparative study of different methodologies for quantitative rock analysis by laser-induced breakdown spectroscopy in a simulated Martian atmosphere, *Spectrochim. Acta B* 61 (2006) 301–313.
 - [45] G.P. Gupta, B.M. Suri, A. Verma, M. Sundaraman, V.K. Unnikrishnan, K. Alti, V.B. Kartha, C. Santhosh, Quantitative elemental analysis of nickel alloys using calibration-based laser-induced breakdown spectroscopy, *J. Alloys Compd.* 509 (2011) 3740–3745.
 - [46] AFNOR editions, Linear Calibration Using Reference Materials, (1996) (ISO 11095:1996, 29 pages).
 - [47] B. Magnusson, U. Örnemark, Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, 2nd ed, (2014) ISBN 978-91-87461-59-0. Available from www.eurachem.org.
 - [48] 2002/657/EC, Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (Text with EEA relevance) (notified under document number C), (2002), p. 3044.
 - [49] B.N. Bennett, M.Z. Martin, D.N. Leonard, E. Garlea, Calibration curves for commercial copper and aluminum alloys using handheld laser-induced breakdown spectroscopy, *Appl. Phys. B Lasers Opt.* 124 (2018) 42.
 - [50] J.K. Taylor, Quality Assurance of Chemical Measurements, Lewis Publishers, 1987, <https://doi.org/10.1201/9780203741610>.
 - [51] A. Golbraikh, A. Tropsha, Beware of q^2 , *J. Mol. Graph. Model.* 20 (2002) 269–276.
 - [52] J. El Haddad, L. Canioni, B. Bousquet, Good practices in LIBS analysis: review and advices, *Spectrochim. Acta B* 101 (2014) 171–182.
 - [53] F. Westad, F. Marini, Validation of chemometric models – a tutorial, *Anal. Chim. Acta* (2015), <https://doi.org/10.1016/j.aca.2015.06.056>.
 - [54] T. Zhang, H. Tang, H. Li, Chemometrics in laser-induced breakdown spectroscopy, *J. Chemom.* (2018), <https://doi.org/10.1002/cem.2983>.
 - [55] J.M. Mermet, Calibration in atomic spectrometry: a tutorial review dealing with quality criteria, weighting procedures and possible curvatures, *Spectrochim. Acta B* 65 (2010) 509–523.
 - [56] J.M. Mermet, Limit of quantification in atomic spectrometry: an unambiguous concept? *Spectrochim. Acta B* 63 (2008) 166–182.
 - [57] V. Motto-Ros, D. Syvilay, L. Bassel, E. Negre, F. Trichard, F. Pelascini, J. El Haddad, A. Harhira, S. Moncayo, J. Picard, D. Devismes, B. Bousquet, Critical aspects of data analysis for quantification in laser-induced breakdown spectroscopy, *Spectrochim. Acta B* 140 (2018) 54–64.
 - [58] V. Payré, C. Fabre, A. Cousin, V. Sautter, R.C. Wiens, O. Forni, O. Gasnault, N. Mangold, P.Y. Meslin, J. Lasue, A. Ollila, W. Rapin, S. Maurice, M. Nachon, L. le Deit, N. Lanza, S. Clegg, Alkali trace elements in Gale crater, Mars, with ChemCam: calibration update and geological implications, *J. Geophys. Res.* 122 (2017) 431–484.
 - [59] RMT Quasaprove, <http://www.quasaprove.org> (Last access date 05/12/2016).
 - [60] S. Moncayo, J.D. Rosales, R. Izquierdo-Hornillos, J. Anzano, J.O. Cáceres, Classification of red wine based on its protected designation of origin (PDO) using Laser-induced Breakdown Spectroscopy (LIBS), *Talanta* 158 (2016) 185–191.
 - [61] M.A. Ismail, H. Imam, A. Elhassan, W.T. Younis, M.A. Harith, LIBS limit of detection and plasma parameters of some elements in two different metallic matrices, *J. Anal. At. Spectrom.* 19 (2004) 489–494.
 - [62] J. Vrenegor, R. Noll, V. Sturm, Investigation of matrix effects in laser-induced breakdown spectroscopy plasmas of high-alloy steel for matrix and minor elements, *Spectrochim. Acta B* 60 (2005) 1083–1091.
 - [63] S. Schröder, P.-Y. Meslin, O. Gasnault, S. Maurice, A. Cousin, R.C. Wiens, et al., Hydrogen detection with ChemCam at Gale crater, *Icarus* 249 (2015) 43–61, <https://doi.org/10.1016/j.icarus.2014.08.029>.

Annexe - Chapitre III

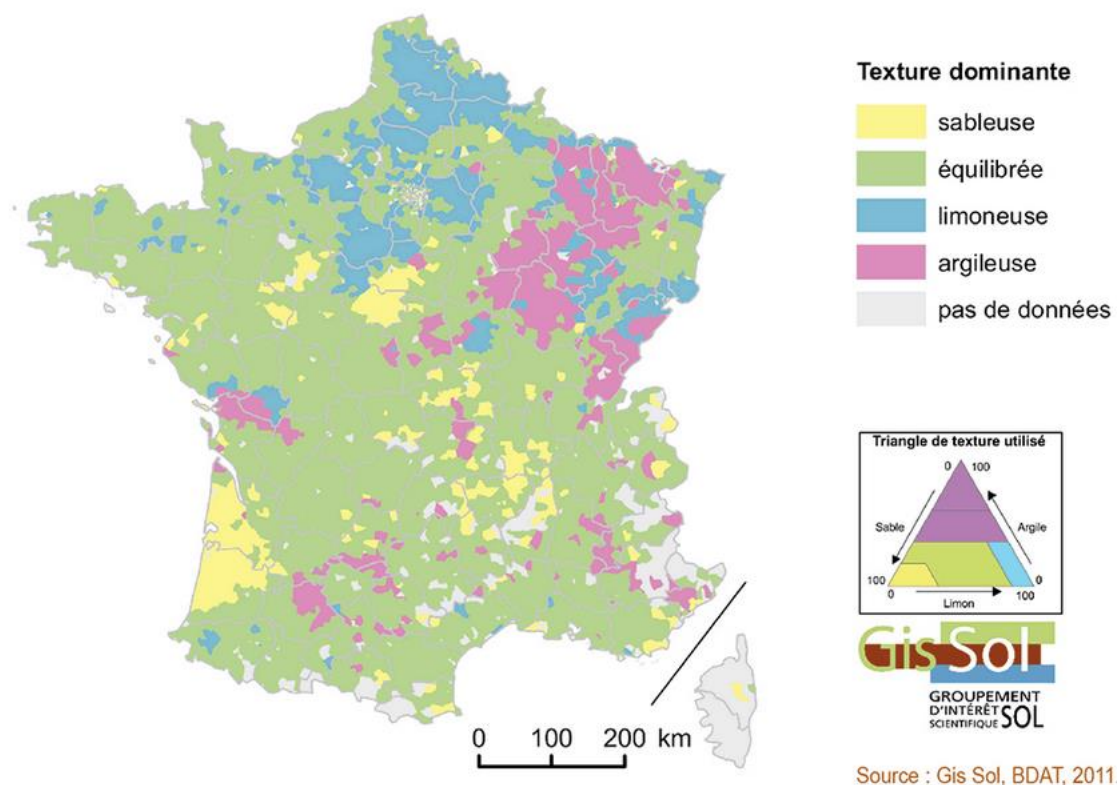


Figure 3-1 : Texture dominante de l'horizon supérieur des sols agricoles français ainsi que le diagramme ternaire associé [79].

Elément	C I	Si I	P I	Fe I	Fe II	Mg I	Mg II	Al I	Ca I	Ca II
Longueur d'onde (nm)	193,1	212,4 221,1 221,7 288,2	213,6 214,9	302,1 430	233- 276	285,2 517,3 518,4	279,6 279,8 280,3	308,2 309,3 394,4 396,2	422,7 443,5 526,5	315,9 317,9 393,3 396,9
Elément	Na I	K I	H I	Li I	O I	Ar	Ti II			
Longueur d'onde (nm)	588 589,6	766,5 769,9	656,3	670,8	777	696,5 714,7 763,5	334,9			

Tableau 3-1 : Récapitulatif des éléments détectés par le Z-300 dans les sols et les longueurs d'onde associées (nm), attribuées à l'aide de la base de données NIST.

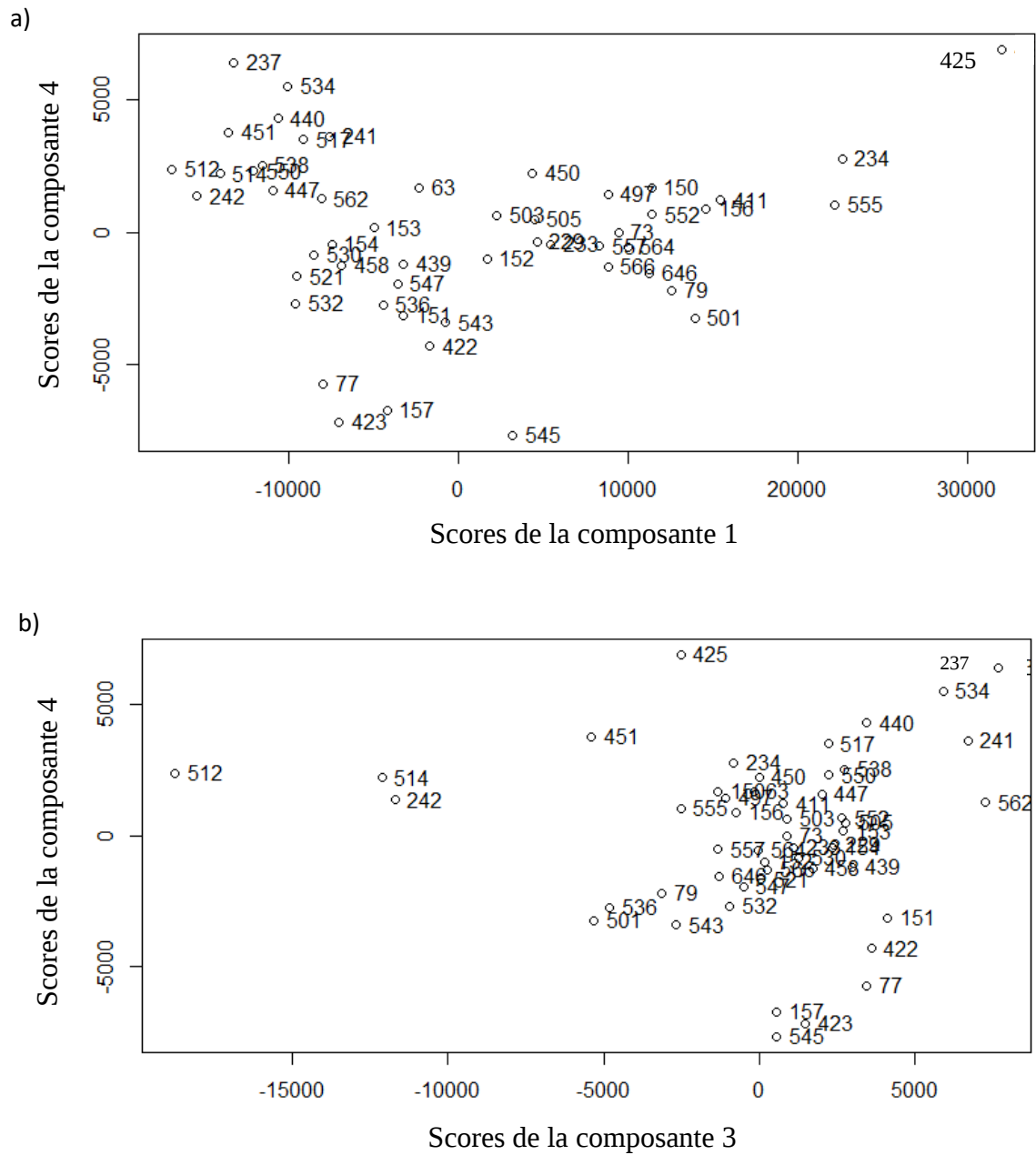


Figure 3-2 : Graphe des scores de l'ACP calculée sur les spectres LIBS moyens des pastilles de sols pour a) la composante 1 et 4 et b) la composante 3 et 4.

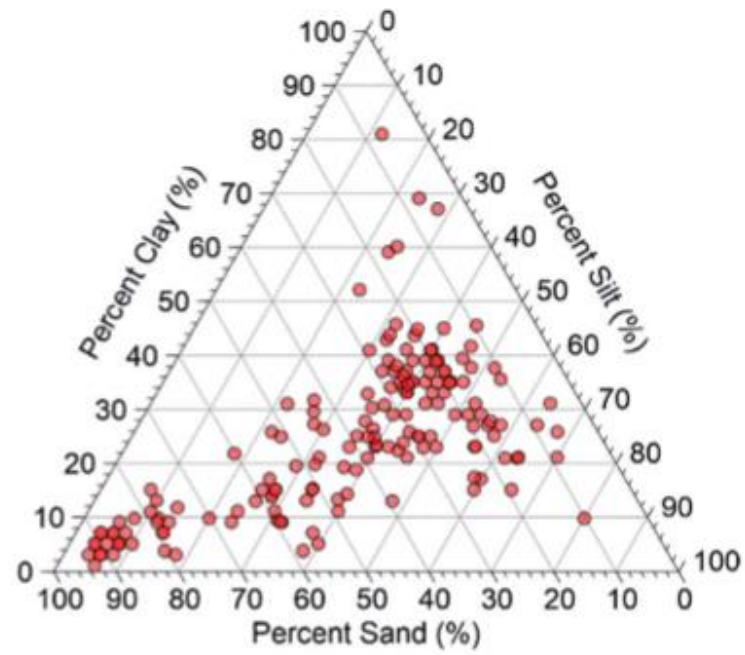


Figure 3-3 : Diagramme ternaire des textures pour les sols de l'étude proposée par Gouegel et al en 2019 [82].

Annexe

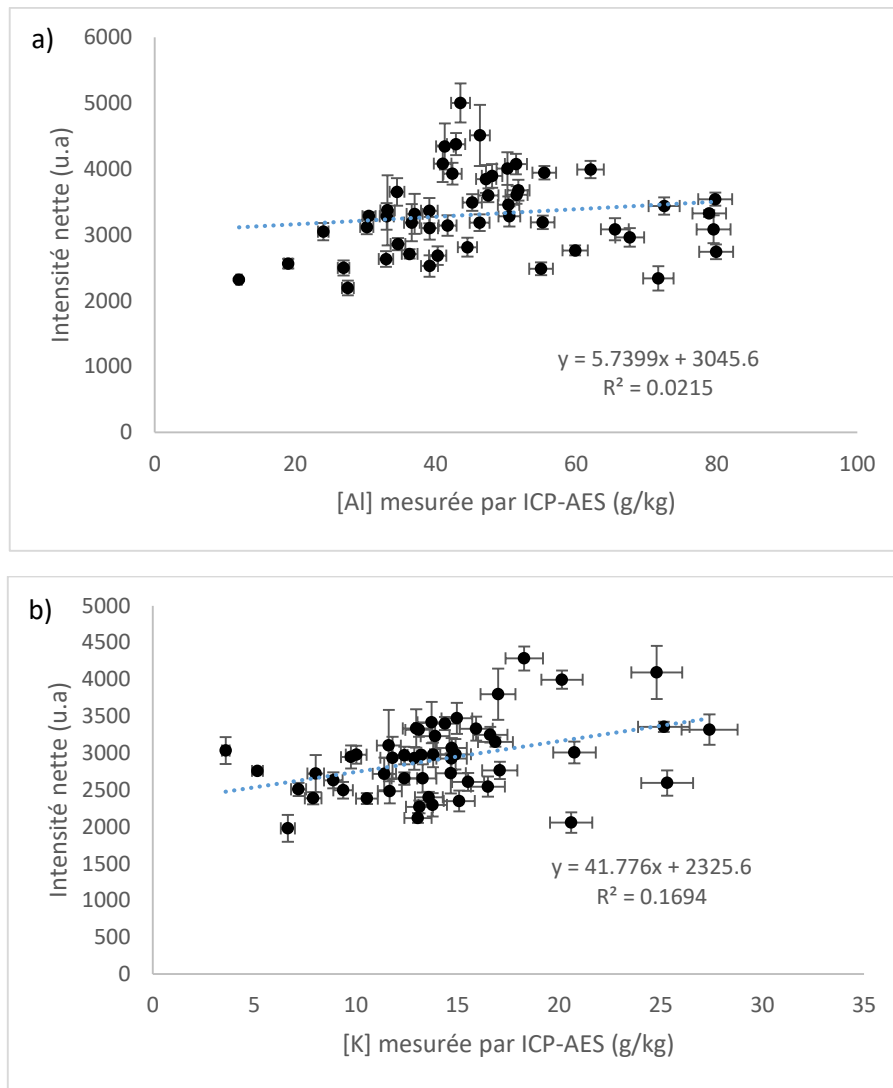


Figure 3-4 : Modèle de régression univariée obtenu sur les sols agricoles en LIBS pour a) l'aluminium et b) le potassium.

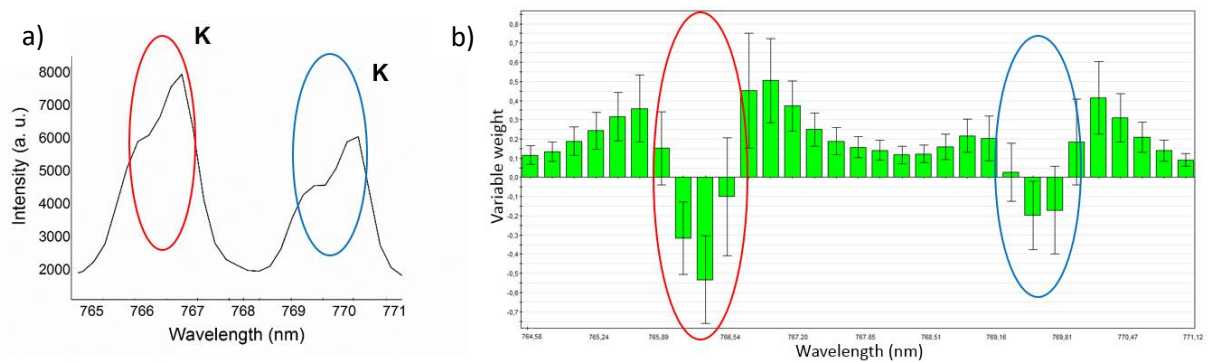


Figure 3-5 : Auto-absorption des raies du potassium pour des spectres LIBS de sols observée à travers a) le doublet de K I à 766 et 769 nm et b) le graphe des coefficients zoomé sur le même doublet du potassium.

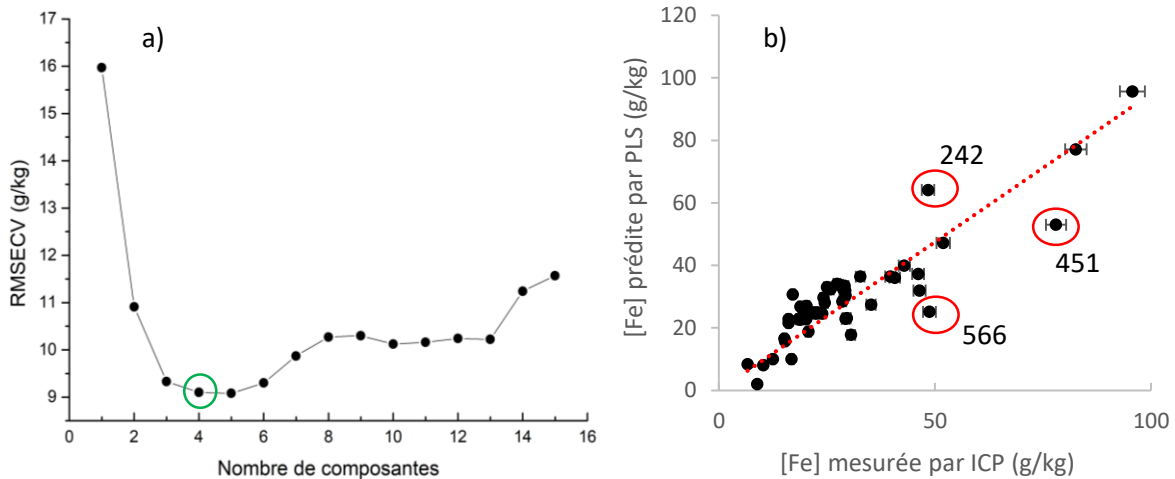


Figure 3-6 : Figures de mérite de la PLS calculée pour le fer en prenant en compte l'entièreté du spectre LIBS comme variables d'entrée avec a) l'évolution de la RMSECV (g/kg) en fonction du nombre de composantes et b) la régression entre la concentration mesurée en Fe par ICP et la concentration prédite par le modèle PLS pour 4 composantes. Le cercle vert indique le nombre de composantes jugé optimal pour construire le modèle et les cercles rouges entourent les échantillons considérés comme outliers.

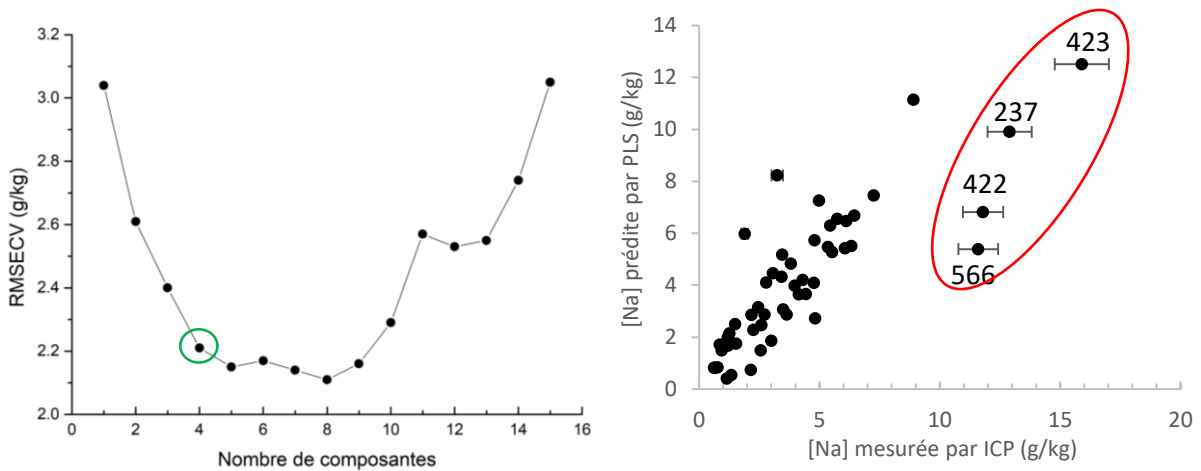


Figure 3-7 : Figures de mérite de la PLS calculée pour le sodium en prenant en compte l'entièreté du spectre LIBS comme variables d'entrée avec a) l'évolution de la RMSECV (g/kg) en fonction du nombre de composantes et b) la régression entre la concentration mesurée en Na par ICP et la concentration prédite par le modèle PLS pour 4 composantes. Le cercle vert indique le nombre de composantes jugé optimal pour construire le modèle et les cercles rouges entourent les échantillons considérés comme outliers.

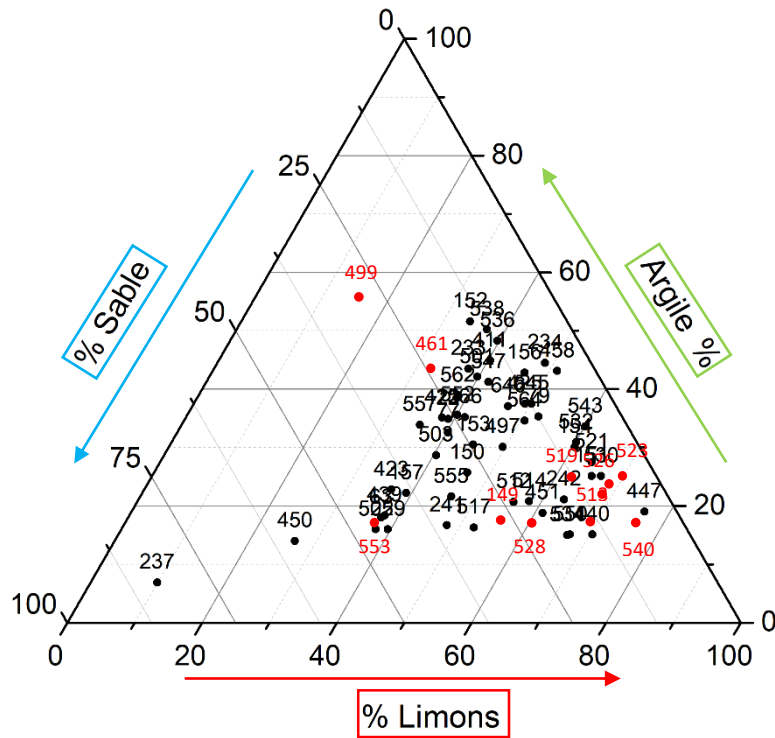


Figure 3-8 : Diagramme ternaire des textures pour les sols de cette étude issu du réseau QUASAPROVE avec en noir les échantillons de calibration et en rouge les échantillons de validation.

Echantillon de validation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
Identification réseau QUASAPROVE	149	461	499	513	519	523	526	528	540	549	553

Tableau 3-2 : Correspondance entre les noms donnés aux échantillons de sols agricoles utilisés pour la validation des modèles PLS et leur labellisation dans le réseau QUASAPROVE.

Annexe- Chapitre IV



Figure 4-1 : Exemple de feuille de blé sélectionnée sur une parcelle de blé de Feucherolles pour analyse LIBS in-situ.



Figure 4-2 : photo de l'instrument LIBS Z-300 monté sur un support lors de la campagne de mesures réalisée sur des parcelles de blé Qualiagro.

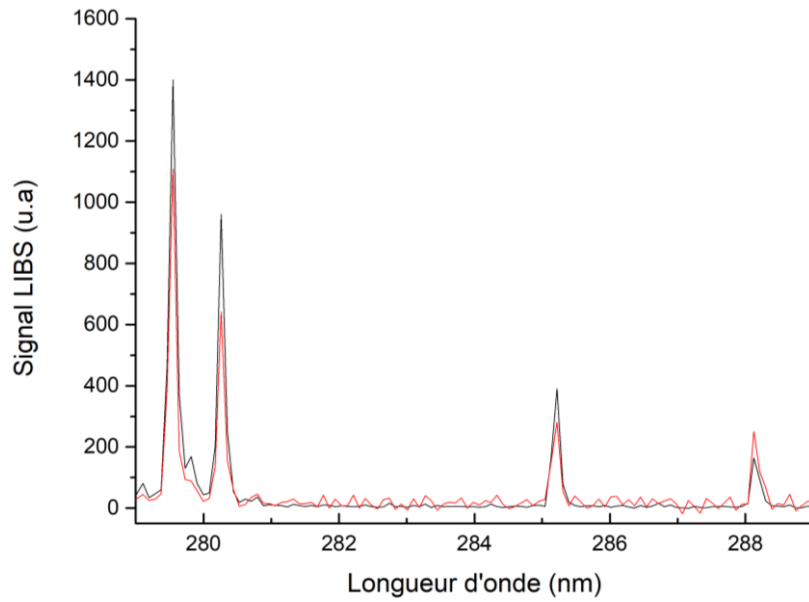


Figure 4-3 : Comparaison entre un spectre LIBS zoomé d'une plante sous la forme d'une pastille sèche (trait noir) et d'une plante fraîche sans préparation (trait rouge).

Facteurs			Intensité nette Silicium (288 nm)		
Humidité (%)	Taille de grain (mm)	Poids de pression (tonnes)	Pastille 1	Pastille 2	Pastille 3
2	0.25	2	3960	4339	4067
10	0.25	2	1872	1667	1741
2	2	2	3970	3755	4837
10	2	2	1997	1971	1945
2	0.25	8	4086	4044	4305
10	0.25	8	1772	1686	1573
2	2	8	4302	3708	4499
10	2	8	1817	1823	1519

Tableau 4-1 : Résultats expérimentaux obtenus pour le plan d'expérience réalisé à partir de la raie du silicium à 288,2 nm du spectre LIBS du sol 538 (QUASAPROVE).

Facteurs			Intensité nette Calcium (315 nm)		
Humidité (%)	Taille de grain (mm)	Poids de pression (tonnes)	Pastille 1	Pastille 2	Pastille 3
2	0.25	2	1888	1932	1777
10	0.25	2	375	333	335
2	2	2	2493	2765	4457
10	2	2	1020	789	739
2	0.25	8	1945	1850	1792
10	0.25	8	314	357	354
2	2	8	5505	3250	3923
10	2	8	245	552	518

Tableau 4-2 : Résultats expérimentaux obtenus pour le plan d'expérience réalisé à partir de la raie du calcium à 315,9 nm du spectre LIBS du sol 538 (QUASAPROVE).

Volume ajouté (µL)	0	20	40	60	80	100
Taux d'humidité (%)	2	7	13	17	20	25

Tableau 4-3 : Correspondance entre le volume d'eau pure ajouté sur les pastilles de sol de cressonnière et le taux d'humidité (%).

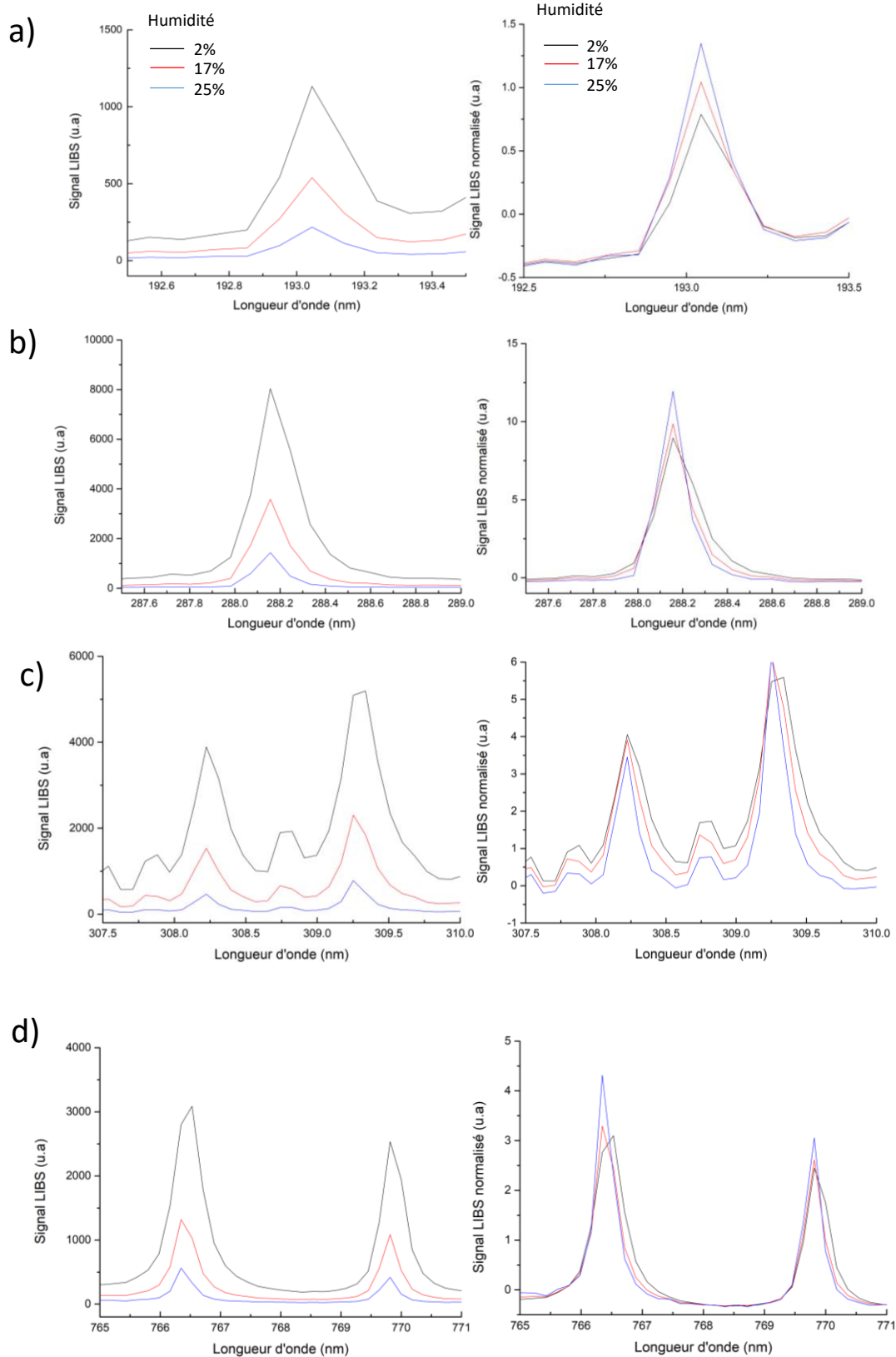


Figure 4-4 : Comparatif des spectres LIBS d'une pastille du même sol analysée sous 3 conditions différentes d'humidité et zoomés sur quatre raies élémentaires, à savoir a) le carbone, b) le silicium, c) l'aluminium et d) le potassium.

Bibliographie

- [1] Intergovernmental panel on climate change (IPPC), Global warming of 1,5°C, 2018.
- [2] D. Cordell, J.O. Drangert, S. White, The story of phosphorus : Global food security and food for thought, *Global Environmental Change*, 19 (2009), Issue 2, pp. 292-305.
- [3] P.J. Zarco-Tejada, L. Mateos, E. Fereres, F.J. Villalobos, New tools and methods in agronomy, *Principles of Agronomy for Sustainable Agriculture*, 2016, pp. 503-514.
- [4] C. Dordas, Role of nutrients in controlling plant diseases in sustainable agriculture. A review, *Agronomy for Sustainable Development*, 28 (2008), pp. 33-46.
- [5] Union des Industries de la Fertilisation (UNIFA) – Site Internet, La nutrition des plantes, Consulté sur <https://fertilisation-edu.fr/nutrition-des-plantes.html>.
- [6] M.R. Motsara, R.N. Roy, Guide to laboratory establishment for plant nutrient analysis, *Fertilizer and Plant Nutrition Bulletin (FAO)*, 19 (2008).
- [7] R.A.V. Rossell, V.I. Adamchuk, K.A. Sudduth, N.J. McKenzie, C. Lobsey, Chapter five – Proximal soil sensing : An effective approach for soil measurements in space and time, *Advances in Agronomy*, 113 (2011), pp. 243-291.
- [8] C. Nduwamungu, N. Ziadi, L-E. Parent, G.F. Tremblay, L. Thuriès, Opportunities for, and limitations of, near infrared reflectance spectroscopy applications in soil analysis : A review, *Canadian Journal of Soil Science*, 89 (2009), pp. 531-541.
- [9] F.J. Krug, M. Guerra, E. Almeida, G.G. Carvalho, P.F. Souza, L.C. Nunes, D.S. Júnior, Comparison of analytical performance of benchtop and handheld energy dispersive X-ray fluorescence systems for the direct analysis of plant materials, *Journal of Analytical Atomic Spectrometry*, 29 (2014), pp. 1667-1674.
- [10] G.G. Arantes de Carvalho, M.B. Bueno Guerra, A. Adame, C.S. Nomura, P.V. Oliveira, H.W. Pereira de Carvalho, D. Santos Jr, L.C. Nunes, F. José Krug, Recent advances in LIBS and XRF for the analysis of plants, *Journal of Analytical Atomic Spectrometry*, 33 (2018), Issue 6.
- [11] G. Nicolodelli, J. Cabral, C.R. Menegatti, B. Marangoni, G.S. Senesi, Recent advances and future trends in LIBS applications to agricultural materials and their food derivatives : An overview of developments in the last decade (2010-2019). Part I. Soils and fertilizers, *TrAC Trends in Analytical Chemistry*, 115 (2019), pp. 70-82.
- [12] G.S. Senesi, J. Cabral, C.R. Menegatti, B. Marangoni, G. Nicolodelli, Recent advances and future trends in LIBS applications to agricultural materials and their food derivatives : An overview of developments in the last decade (2010-2019). Part II. Crop plants and their food derivatives, *TrAC Trends in Analytical Chemistry*, 118 (2019), pp. 453-469.
- [13] D.W. Hahn, N. Omenetto, Laser-Induced Breakdown Spectroscopy (LIBS), Part II : Review of instrumental and methodological approaches to material analysis and applications to different fields, *Applied Spectroscopy*, 66 (2012), Issue 4, pp. 347-419.
- [14] F. Anabitarte, A. Cobo, J.M. Lopez-Higuera, Laser-Induced Breakdown Spectroscopy: Fundamentals, Applications, and Challenges, *ISRN Spectroscopy*, 2012.

- [15] J. Rakovsky, P. Cermak, O. Musset, P. Veis, A review of the development of portable laser induced breakdown spectroscopy and its applications, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 101 (2014), pp. 269-287.
- [16] N. Tilly, La LIBS : les applications d'un laser d'analyse, des systèmes nucléaires à l'exploration spatiale, *CEA – Service Information-Media*, 2014.
- [17] A.J.R. Bauer, S.G. Buckley, Novel applications of Laser-induced breakdown spectroscopy, *Applied Spectroscopy*, 71 (2017), Issue 4, pp. 553-566.
- [18] S. Maurice, et al., The ChemCam Instrument Suite on the Mars Science Laboratory (MSL) Rover : Science Objectives and Mast Unit Description, *Space Science Reviews*, 170 (2012), Issue 1-4, pp. 95-166.
- [19] L. Jolivet, M. Leprince, S. Moncayo, L. Sorbier, C.-P. Lienemann, V. Motto-Ros, Review of the recent advances and applications of LIBS-based imaging, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 151 (2019), pp. 41-53.
- [20] D.W. Hahn, N. Omenetto, Laser-Induced Breakdown Spectroscopy (LIBS), Part I : Review of basic diagnostics and plasma-particle interactions : still-challenging issues within the analytical plasma community, *Applied Spectroscopy*, 64 (2010), Issue 12, pp. 335A.
- [21] E. Tognoni, G. Cristoforetti, Signal and noise in laser induced breakdown spectroscopy: an introductory review, *Optics & Laser Technology*, 79 (2016), pp. 164-172.
- [22] N.B. Zorov, A.A. Gorbatenko, T.A. Labutin, A.M. Popov, A review of normalization techniques in analytical atomic spectrometry with laser sampling: From single to multivariate correction, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 65 (2010), pp. 642-657.
- [23] S.I. Gornushkin, I.B. Gornushkin, J.M. Anzano, B.W. Smith, and J.D. Winefordner, Effective normalization technique for correction of matrix effects in Laser-Induced breakdown spectroscopy detection of magnesium in powdered samples, *Applied Spectroscopy*, 56(4), 2002, pp. 433-436.
- [24] J. Guezenoc, A. Gallet-Budynek, B. Bousquet, Critical review and advices on spectral-based normalization methods for LIBS quantitative analysis, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 160 (2019).
- [25] E. Tognoni, G. Cristoforetti, S. Legnaioli, V. Palleschi, Calibration-free laser-induced breakdown spectroscopy : state of the art, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 65 (2010), Issue 1, pp. 1-14.
- [26] J. Hermann, E. Axente, F. Pelascini, V. Craciun, Analysis of multi-elemental thin films via calibration-free laser-induced breakdown spectroscopy, *Analytical Chemistry*, 91 (2019), Issue 3, pp. 2544-2550.
- [27] V. Motto-Ros, D. Syvilay, L. Bassel, E. Negre, F. Trichard, F. Pelascini, J. El Haddad, A. Harhira, S. Moncayo, J. Picard, D. Devismes, B. Bousquet, Critical aspects of data analysis for quantification in laser-induced breakdown spectroscopy, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 140 (2018), pp. 54-64.
- [28] T. Zhang, H. Tang, H. Li, Chemometrics in laser-induced breakdown spectroscopy, *Journal of Chemometrics*, 32 (2018), Issue 11.
- [29] A. Safi, B. Campanella, E. Grifoni, S. Legnaioli, G. Lorenzetti, S. Pagnotta, F. Poggialini, L. Ripoll-Seguer, M. Hidalgo, V. Palleschi, Multivariate calibration in laser-induced breakdown spectroscopy quantitative analysis : The dangers of a « black box » approach and how to avoid them, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 144 (2018), pp. 46-54.
- [30] G.S. Senesi, N. Senesi, Laser-induced breakdown spectroscopy (LIBS) to measure quantitatively soil carbon with emphasis on soil organic carbon. A review, *Analytica Chimica Acta*, 938 (2016), pp. 7-17.
- [31] D. Santos Jr, L.C. Nunes, G.G.A. de Carvalho, M. da Silva Gomes, P.F. de Souza, F. de Oliveira Leme, L.G.C. dos Santos, F. José Krug, Laser-induced breakdown spectroscopy for

- analysis of plant materials : A review, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 71-72 (2012), pp. 3-13.
- [32] Y. Groisman, M. Gaft, Online analysis of potassium fertilizers by laser-induced breakdown spectroscopy, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 65 (2010), pp. 744-749.
- [33] A. Segnini, A.A.P. Xavier, P.L. Otaviani-Junior, E.C. Ferreira, A.M. Watanabe, M.A. Sperança, G. Nicolodelli, P.R. Villas-Boas, P.P.A. Oliveira, D.M.B.P. Milori, Physical and Chemical Matrix Effects in Soil Carbon Quantification Using Laser-Induced Breakdown Spectroscopy, *American Journal of Analytical Chemistry*, 5 (2014), pp. 722-729.
- [34] M.E. Essington, G.V. Melnichenko, M.A. Stewart, R.A. Hull, Soil metals analysis using laser-induced breakdown spectroscopy (LIBS), *Soil Science Society of American Journal*, 73 (2009), Issue 5, pp. 1469-1478.
- [35] K. Devey, M. Mucalo, G. Rajendram, J. Lane, Pasture vegetation elemental analysis by laser-induced breakdown spectroscopy, *Communications in Soil Science and Plant Analysis*, 46 (2015), pp. 72-80.
- [36] S.A. Beal, A.M. Mossell, J.L. Clausen, Matrix and target particle-size effects on LIBS analysis of soils, *Engineer Research and Development Center*, 2020.
- [37] RMT Quasaprove, <http://www.quasaprove.org> (Last access date 05/12/2016).
- [38] IPE Reference Material WEPAL, <http://www.wepal.nl/website/products/RefMatIPE.htm> (Last access date 14/09/2020).
- [39] Charles Vander Linden, Impacts de cultures tropicales sur le cycle du silicium. Cas du bassin versant de la rivière des Pères et de la rivière Pérou, *Thèse de doctorat* (2019), Université catholique de Louvain.
- [40] Site de l'INRAE – Présentation du réseau SOERE-PRO, <https://www6.inrae.fr/qualiagro/Nos-partenaires/Le-reseau-SOERE-PRO> (Last access date 23/09/2015).
- [41] Agilent Technologies, ICP-OES Application eHandbook.
- [42] Josette El Haddad, Chimimétrie appliquée à la spectroscopie de plasma induit par laser (LIBS) et à la spectroscopie terahertz, *Thèse de doctorat* (2013), Université Sciences et Technologies – Bordeaux I.
- [43] IPE Periodic reports WEPAL, <http://www.wepal.nl/website/products/QuarterlyReport.htm> (Last access date 2016).
- [44] T.R. Tavares, J.P. Molin, L.C. Nunes, E.E.N. Alves, F.L. Melquiades, H.W.P. de Carvalho, A.M. Mouazen, Effect of X-Ray tube configuration on measurement of key soil fertility attributes with XRF, *Remote Sensing*, 12 (2020), Issue 6.
- [45] Valérie Thirion-Merle, Spectrométrie de fluorescence X. Circulation et provenance des matériaux dans les sociétés anciennes, *Editions des archives contemporaines* (2014), Collection Sciences Archéologiques.
- [46] Bruker, Reference Manual S1 TITAN, version 030.0058.04.0.
- [47] G. Galbacs, A critical review of recent progress in analytical laser-induced breakdown spectroscopy, *Analytical and Bioanalytical Chemistry*, 407 (2015), Issue 25, pp. 7537-7562.
- [48] Amina Ismaël, Une évaluation des performances analytiques de la spectroscopie sur plasma induit par laser (LIBS), *Thèse de doctorat* (2011), Université de Bordeaux I - Sciences Physiques et de l'ingénieur.
- [49] J.M. Gogue, Manuel du contrôle statistique de la qualité, *Western Electric*, 2012.
- [50] B. Bousquet, G. Travaillé, A. Ismaël, L. Canioni, K. Michel-Le Pierrès, E. Brasseur, S. Roy, I. Le Hecho, M. Larregieu, S. Tellier, M. Potin-Gautier, T. Boriachon, P. Wazen, A. Diard, S. Belbèze, Development of a mobile system based on laser-induced breakdown spectroscopy and dedicated to in situ analysis of polluted soils, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 63 (2008), Issue 10, pp. 1085-1090.

- [51] D. Syvilay, J. Guezenc, B. Bousquet, Guideline for increasing the analysis quality in laser-induced breakdown spectroscopy, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 161 (2019).
- [52] J. Goupy, Les plans d'expériences - Tutoriel, *Revue Modulad*, 34 (2006).
- [53] A. Kramida, Y. Ralchenko, J. Reader and NIST ASD Team, NIST Atomic Spectra Database, *National Institute of Standards and Technology*, Available online : <https://physics.nist.gov/asd>.
- [54] RStudio Team (2020), RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- [55] M.A. de Menezes Franco, D.M.B.P. Milori, P.R. Villas-Boas, Comparison of algorithms for baseline correction of LIBS spectra for quantifying total carbon in Brazilian soils, *arXiv eprints*, 2018.
- [56] T.G. Bloemberg, J. Gerretzen, H.J.P. Wouters, J. Gloerich, M. van Dael, H.J.C.T. Wessels, L.P. van den Heuvel, P.H.C. Eilers, L.M.C. Buydens, R. Wehrens, Improved Parametric Time Warping for Proteomics, *Chemometrics and Intelligent Laboratory Systems*, 104 (2010), Issue 1, pp. 65-74.
- [57] A.J. Miller, Plant Mineral Nutrition, *John Wiley & Sons* (2014).
- [58] E. Epstein, A.J. Bloom, Mineral nutrition of plants : principles and perspectives, *Sinauer Associates* (2005).
- [59] H. Jull, R. Künemeyer, P. Schaare, Nutrient quantification in fresh and dried mixtures of ryegrass and clover leaves using laser-induced breakdown spectroscopy, *Precision Agriculture*, 19 (2018), pp 823-839.
- [60] L.C. Nunes, J.W.B. Braga, L.C. Trevizan, P.F. de Souza, G.G.A. de Carvalho, D. Santos Jr, R.J. Poppi, F. José Krug, Optimization and validation of a LIBS method for the determination of macro and micronutrients in sugar cane leaves, *Journal of Analytical Atomic Spectrometry*, 25 (2010), Issue 9, pp 1453.
- [61] M. da Silva Gomes, G.G.A. de Carvalho, D. Santos Jr, F. José Krug, A novel strategy for preparing calibration standards of plant materials by laser-induced breakdown spectroscopy : A case study with pellets of sugar cane leaves, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 86 (2013), pp 137-141.
- [62] M. Fontalvo-Gomez, J.A. Colucci, N. Velez, R.J. Romanach, In-line near-infrared (NIR) and raman spectroscopy coupled with principal component analysis (PCA) for in situ evaluation of the transesterification reaction, *Applied Spectroscopy*, 67 (2013), Issue 10, pp 1142-1149.
- [63] X. Li, T. Yang, S. Li, D. Wang, Y. Song, S. Zhang, Raman spectroscopy combined with principal component analysis and k nearest neighbour analysis for non-invasive detection of colon cancer, *Laser Physics*, 26 (2016), Issue 3.
- [64] Y. Yucel, P. Sultanoglu, Characterization of Hatay honeys according to their multi-element analysis using ICP-OES combined with chemometrics, *Food Chemistry*, 140 (2013), Issue 1-2, pp. 231-237.
- [65] P. Porizka, J. Klus, E. Képes, D. Prochazka, D.W. Hahn, J. Kaiser, On the utilization of principal component analysis in laser-induced breakdown spectroscopy data analysis, a review, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 148 (2018), pp. 65-82.
- [66] P. Filzmoser, K. Varmuza, Multivariate Statistical Analysis in Chemometrics, 2017, version 1.4.2.
- [67] Å. Rinnan, F. Van den Berg, S. Balling Engelsen, Review of the most common preprocessing techniques for near-infrared spectra, *TrAC Trends in Analytical Chemistry*, 28 (2009), Issue 10, pp. 1201-1222.

- [68] F. Pilon, A. Labet, K. Vielle, O. Vigneau, G. Granier, P. Bienvenu, J.M. Mermet, La validation de méthode en spectrométrie d'émission optique à source plasma (ICPOES): de l'échantillon au résultat, *EDP Sciences*, 2017.
- [69] M.F. Gazulla, M. Rodrigo, M. Orduna, M.J. Ventura, C. Andreu, High precision measurement of silicon in naphthas by ICP-OES using isooctane as diluent, *Talanta*, 164 (2017) pp. 563–569.
- [70] S. Legnaioli, B. Campanella, S. Pagnotta, F. Poggialini, V. Palleschi, Chapter 24 – Self-calibrated methods for LIBS quantitative analysis, *Laser-induced Breakdown Spectroscopy*, 2020, pp. 561-580.
- [71] J.M. Mermet, Calibration in atomic spectrometry: a tutorial review dealing with quality criteria, weighting procedures and possible curvatures, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 65 (2010), pp. 509–523.
- [72] J.M. Mermet, Limit of quantification in atomic spectrometry: an unambiguous concept? *Spectrochimica Acta Part B : Atomic Spectroscopy*, 63 (2008), pp. 166–182.
- [73] G.G.A. de Carvalho, D. Santos Jr, M. da Silva Gomes, L.C. Nunes, M.B. Bueno Guerra, F. José Krug, Influence of particle size distribution on the analysis of pellets of plant materials by laser-induced breakdown spectroscopy, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 105 (2015), pp. 130–135.
- [74] G. Nicolodelli, G.S. Senesi, A.C. Ranulfi, B.S. Marangoni, A. Watanabe, V. de Melo Benites, P.P.A. de Oliveira, P. Villas-Boas, D.M.B.P. Milori, Double-pulse laser induced breakdown spectroscopy in orthogonal beam geometry to enhance line emission intensity from agricultural samples, *Microchemical Journal*, 133 (2017), pp. 272–278.
- [75] X. Zhao, S. Xu, T. Liu, P. Qiu, G. Qin, Moisture distribution in sloping black soil farmland during the freeze-thaw period in Northeastern China, *Water*, 11 (2019), Issue 3, pp. 536.
- [76] M. da Silva Gomes, D. Santos Jr, L.C. Nunes, G.G.A. de Carvalho, F. de Oliveira Leme, F. José Krug, Evaluation of grinding methods for pellets preparation aiming at the analysis of plant materials by laser-induced breakdown spectroscopy, *Talanta*, 85 (2011), Issue 4, pp. 1744-1750.
- [77] J.W.B. Braga, L.C. Trevizan, L.C. Nunes, I.A. Rufini, D. Santos Jr, F. José Krug, Comparison of univariate and multivariate calibration for the determination of micronutrients in pellets of plant materials by laser induced breakdown spectroscopy, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 65 (2010), Issue 1, pp. 66-74.
- [78] L.C. Trevizan, D. Santos Jr, R.E. Samad, N. Dias Vieira Jr, C.S. Nomura, L.C. Nunes, I.A. Rufini, F. José Krug, Evaluation of laser induced breakdown spectroscopy for the determination of macronutrients in plant materials, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 63 (2008), Issue 10, pp. 1151-1158.
- [79] Groupement d'Intérêt Scientifique Sol (Gis Sol), Base de données d'analyses des terres – BDAT, Consulté sur : <https://www.gissol.fr/le-gis/programmes/base-de-donnees-danalyses-des-terres-bdat-62>.
- [80] C. Jolivet, L. Boulonne, C. Ratié, Manuel du réseau de mesures de la qualité des sols, *Unité InfoSol*, édition 2006, INRA Orléans, France, 190p.
- [81] P.R. Villas-Boas, R.A. Romano, M.A. de Menezes Franco, E.C. Ferreira, E.D. Ferreira, S. Crestana, D.M.B.P. Milori, Laser-induced breakdown spectroscopy to determine soil texture : A fast analytical technique, *Geoderma*, 263 (2016), pp. 195-202.
- [82] C.L. Goueguel, A. Soumare, C. Nault, J. Nault, Direct determination of soil texture using laser-induced breakdown spectroscopy and multivariate linear regressions, *Journal of Analytical Atomic Spectrometry*, 34 (2019), Issue 8.
- [83] J. Guezenoc, V. Payré, C. Fabre, D. Syvilay, A. Cousin, A. Gallet-Budynek, B. Bousquet, Variable selection in laser-induced breakdown spectroscopy assisted by multivariate analysis :

- An alternative to multi-peak fitting, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 152 (2019), pp. 6-13.
- [84] A.M. Ollila, H.E. Newsom, B. Clark III, R.C. Wiens, A. Cousin, J.G. Blank, N. Mangold, V. Sautter, S. Maurice, S.M. Clegg, O. Gasnault, O. Forni, R. Tokar, E. Lewin, M.D. Dyar, J. Lasue, R. Anderson, S.M. McLennan, J. Bridges, D. Vaniman, N. Lanza, C. Fabre, N. Melikechi, G.M. Perrett, J.L. Campbell, P.L. King, B. Barraclough, D. Delapp, S. Johnstone, P.-Y. Meslin, A. Rosen-Gooding, J. Williams, The MSL science team, trace element geochemistry (Li, Ba, Sr, and Rb) using curiosity's ChemCam: early results for Gale crater from Bradbury landing site to Rocknest, *Journal of Geophysical Research Planets*, 119 (2014), pp. 255-285.
- [85] A. M. Ollila, J. Lasue, H.E. Newsom, R.A. Multari, R.C. Wiens, S.M. Clegg, Comparison of two partial least squares-discriminant analysis algorithms for identifying geological samples with the ChemCam laser-induced breakdown spectroscopy instrument, *Applied Optics*, 51 (2012), Issue 7, pp. B130-B142.
- [86] C.A.F. de Oliveira Penido, M.T. Tavares Pacheco, E.H. Novotny, I.K. Lednev, L. Silveira Jr, Quantification of cocaine in ternary mixtures using partial least squares regression applied to Raman and Fourier transform infrared spectroscopy, *Journal of Raman Spectroscopy*, 48 (2017), Issue 12, pp. 1732-1743.
- [87] A.M.T. Monfared, H. Anis, An improved partial least-squares regression method for Raman spectroscopy, *Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy*, 185 (2017), pp. 98-103.
- [88] D. Cozzolino, W.U. Cynkar, R.G. Damberg, M.D. Mercurio, P.A. Smith, Measurement of condensed tannins and dry matter in red grape homogenates using near infrared spectroscopy and partial least squares, *Journal of Agricultural and Food Chemistry*, 56 (2008), Issue 17, pp. 7631-7636.
- [89] C. Cui, T. Fearn, Comparison of partial least squares regression, least squares support vector machines and gaussian process regression for a near infrared calibration, *Journal of Near Infrared Spectroscopy*, 25 (2017), Issue 1, pp. 5-14.
- [90] A. Mir-Marqués, M. Martinez-Garcia, S. Garrigues, M. Luisa Cervera, M. de la Guardia, Green direct determination of mineral elements in artichokes by infrared spectroscopy and X-ray fluorescence, *Food Chemistry*, 196 (2016), pp. 1023-1030.
- [91] D. Xu, S. Chen, R.A. Viscarra-Rossel, A. Biswas, S. Li, Y. Zhou, Z. Shi, X-ray fluorescence and visible near infrared sensor fusion for predicting soil chromium content, *Geoderma*, 352 (2019), pp. 61-69.
- [92] I.S. Helland, On the structure of partial least squares regression, *Communications in Statistics - Simulation and Computation*, 17 (1988), pp. 581-607.
- [93] F. Westad, F. Marini, Validation of chemometric models – A tutorial, *Analytica Chimica Acta*, 893 (2015), pp. 14-24.
- [94] F.C. De Lucia Jr, J.L. Gottfried, C.A. Munson, A.W. Miziolek, Multivariate analysis of standoff laser-induced breakdown spectroscopy spectra for classification of explosive containing residues, *Applied optics*, 47 (2008), pp. G112-G121.
- [95] F.C. De Lucia, J.L. Gottfried, Influence of variable selection on partial least squares discriminant analysis models for explosive residue classification, *Spectrochimica Acta Part B: Atomic Spectroscopy*, 66 (2011), pp. 122-128.
- [96] B. Bousquet, J.-B. Sirven, L. Canioni, Towards quantitative laser-induced breakdown spectroscopy analysis of soil samples, *Spectrochimica Acta Part B: Atomic Spectroscopy*, 62 (2007), pp. 1582-1589.

- [97] F.B. Gonzaga, C. Pasquini, A Complementary Metal Oxide Semiconductor sensor array based detection system for Laser Induced Breakdown Spectroscopy: Evaluation of calibration strategies and application for manganese determination in steel, *Spectrochimica Acta Part B: Atomic Spectroscopy*, 63 (2008), pp. 56-63.
- [98] B.G. Oztoprak, J. Gonzalez, J. Yoo, T. Gulecen, N. Mutlu, R.E. Russo, O. Gundogdu, A. Demir, Analysis and classification of heterogeneous kidney stones using laser-induced breakdown spectroscopy (LIBS), *Applied spectroscopy*, 66 (2012), pp. 1353-1361.
- [99] J. Guezenoc, L. Bassel, A. Gallet-Budynek, B. Bousquet, Variable selection : A critical issue for quantitative laser-induced breakdown spectroscopy, *Spectrochimica Acta Part B: Atomic Spectroscopy*, 134 (2017), pp. 6-10.
- [100] D.M. Dong, C.J. Zhao, W.G. Zheng, X.D. Zhao, L.Z. Jiao, Spectral characterization of nitrogen in farmland soil by laser-induced breakdown spectroscopy, *Spectroscopy Letters*, 46 (2013), Issue 6, pp. 421-426.
- [101] C. Lu, L. Wang, H. Hu, Z. Zhuang, Y. Wang, R. Wang, L. Song, Analysis of total nitrogen and total phosphorus in soil using laser-induced breakdown spectroscopy, *Chinese Optics Letters*, 11 (2013), Issue 5.
- [102] G. Kim, J. Kwak, K.R. Kim, H. Lee, K.W. Kim, H. Yang, K. Park, Rapid detection of soils contaminated with heavy metals and oils by laser induced breakdown spectroscopy, *Journal of Hazardous Materials*, 263 (2013), Part II, pp. 754-760.
- [103] G.S. Senesi, M. Dell'Aglia, R. Gaudiuso, A. De Giacomo, C. Zaccone, O. De Pascale, T.M. Miano, M. Capitelli, Heavy metal concentrations in soils as determined by laser-induced breakdown spectroscopy (LIBS) with special emphasis on chromium, *Environmental Research*, 109 (2009), Issue 4, pp. 413-420.
- [104] A.M. Popov, T.A. Labutin, S.M. Zaytsev, I.V. Seliverstova, N.B. Zorov, I.A. Kal'ko, Y.N. Sidorina, I.A. Bugaev, Y.N. Nikolaev, Determination of Ag, Cu, Mo and Pb in soils and ores by laser-induced breakdown spectrometry, *Journal of Analytical Atomic Spectrometry*, 29 (2014), pp. 1925-1933.
- [105] D. Santos Jr, L.C. Nunes, L.C. Trevizan, Q. Godoi, F.O. Leme, J.W.B. Braga, F. José Krug, Evaluation of laser induced breakdown spectroscopy for cadmium determination in soils, *Spectrochimica Acta Part B: Atomic Spectroscopy*, 64 (2009), Issue 10, pp. 1073-1078.
- [106] M. Dell'Aglia, R. Alrifai, A. de Giacomo, Nanoparticle enhanced laser induced breakdown spectroscopy (NELBS), a first review, *Spectrochimica Acta Part B: Atomic Spectroscopy*, 148 (2018), pp. 105-112.
- [107] J. Peng, F. Liu, F. Zhou, K. Song, C. Zhang, L. Ye, Y. He, Challenging applications for multi-element analysis by laser-induced breakdown spectroscopy in agriculture : A review, *TrAC Trends in Analytical Chemistry*, 85 (2016), Part C, pp. 260-272.
- [108] M. Ruhlmann, D. Buchele, M. Ostermann, I. Bald, T. Schmid, Challenges in the quantification of nutrients in soils using laser-induced breakdown spectroscopy – A case study with calcium., *Spectrochimica Acta Part B: Atomic Spectroscopy*, 146 (2018), pp. 115-121.
- [109] G. Guo, G. Niu, Q. Shi, Q. Lin, D. Tian, Y. Duan, Multi-element quantitative analysis of soils by laser induced breakdown spectroscopy (LIBS) coupled with univariate and multivariate regression methods, *Analytical Methods*, 11 (2019), pp. 3006-3013.
- [110] A. Erler, D. Riebe, T. Beitz, H.G. Lohmannsroben, R. Gebbers, Soil nutrient detection for precision agriculture using handheld laser-induced breakdown spectroscopy (LIBS) and multivariate regression methods (PLSR, Lasso and GPR), *Sensors*, 20 (2020), Issue 2, pp. 418.
- [111] J. Yongcheng, S. Wen, Z. Baohua, L. Dong, Quantitative analysis of magnesium in soil by laser-induced breakdown spectroscopy coupled with nonlinear multivariate calibration, *Journal of Applied Spectroscopy*, 84 (2017), pp. 731-737.

- [112] E.C. Ferreira, D.M.B.P. Milori, E.J. Ferreira, R.M. da Silva, L. Martin-Neto, Artificial neural network for Cu quantitative determination in soil using a portable laser induced breakdown spectroscopy system, *Spectrochimica Acta Part B : Atomic Spectro*, 63 (2008), pp. 1216-1220.
- [113] R. Gebbers, V.I. Adamchuk, Precision agriculture and food security, *Science*, 327 (2010), Issue 5967, pp. 828-831.
- [114] C. Lugojan, S. Ciulca, Evaluation of relative water content in winter wheat, *Journal of Horticulture, Forestry and Biotechnology*, 15 (2011), Issue 2, pp. 173-177.
- [115] S.L.C. Ferreira, V.A. Lemos, V.S. de Carvalho, E.G.P. da Silva, A.F.S. Queiroz, C.S.A. Felix, D.L.F. da Silva, G.B. Dourado, R.V. Oliveira, Multivariate optimization techniques in analytical chemistry – an overview, *Microchemical Journal*, 140 (2018), pp. 176–182.
- [116] M. Mäkelä, Experimental design and response surface methodology in energy applications: a tutorial review, *Energy Conversion and Management*, 151 (2017), pp. 630-640.
- [117] P.K. Sahu, R.N. Rao, T. Cecchi, S. Swain, C.S. Patro, J. Panda, An overview of Experimental designs in HPLC method development and validation, *Journal of Pharmaceutical and Biomedical Analysis*, 147 (2018), pp. 590–611.
- [118] J.P. Castro, E.R. Pereira-Filho, Twelve different types of data normalization for the proposition of classification, univariate and multivariate regression models for the direct analyses of alloys by Laser-induced breakdown spectroscopy (LIBS), *Journal of Analytical Atomic Spectrometry*, 31 (2018), pp. 2005–2014.
- [119] D.F. Andrade, E.R. Pereira-Filho, P. Konieczynski, Comparison of ICP OES and LIBS analysis of medicinal herbs rich in flavonoids from eastern europe, *Journal of the Brazilian Chemical Society*, 28 (2017), pp. 838–847.
- [120] J.S. Cowpe, J.S. Astin, R.D. Pilkington, A.E. Hill, Application of response surface methodology to laser-induced breakdown spectroscopy : Influences of hardware configuration, *Spectrochimica Acta Part B: Atomic Spectroscopy*, 62 (2007), pp. 1335–1342.
- [121] R. Brennetot, J.L. Lacour, E. Vors, A. Rivoallan, D. Vailhen, S. Maurice, Mars analysis by laser-induced breakdown spectroscopy (MALIS): influence of mars atmosphere on plasma emission and study of factors influencing plasma emission with the use of Doehlert Designs, *Applied Spectroscopy*, 57 (2003), pp. 744–752.
- [122] L. Arroyo, T. Trejos, P.R. Gardinali, J.R. Almirall, Optimization and validation of a Laser Ablation Inductively Coupled Plasma Mass Spectrometry method for the routine analysis of soils and sediments, *Spectrochimica Acta Part B: Atomic Spectroscopy*, 64 (2009), pp. 16-25.
- [123] A.M. Popov, S.M. Zaytsev, I.V. Seliverstova, A.S. Zakuskin, T.A. Labutin, Matrix effects on laser-induced plasma parameters for soils and ores, *Spectrochimica Acta Part B : Atomic Spectroscopy*, 148 (2018), pp. 205-210.
- [124] P.W. Auraujo, R.G. Brereton, Experimental design I. Screening, *TrAC Trends in Analytical Chemistry*, 15 (1996), pp. 26-31.
- [125] C.J. Bronick, R. Lal, Soil structure and management: a review, *TrAC Trends in Analytical Chemistry*, 124 (2005), Issue 1-2, pp. 3-22.
- [126] R.S. Bricklemeyer, D.J. Brown, J.E. Barefield II, S.M. Clegg, Intact soil core total, inorganic, and organic carbon measurement using laser-induced breakdown spectroscopy, *Soil Science Society of America Journal*, 75 (2011), Issue 3, pp. 1006-1018.
- [127] M.B. Bueno Guerra, A. Adame, E. de Almeida, G.G. Arantes de Carvalho, M.A. Stolf Brasil, D. Santos Jr, F. José Krug, Direct analysis of plant leaves by EDXRF and LIBS : microsampling strategies and cross-validation, *Journal of Analytical Atomic Spectrometry*, 30 (2015), Issue 7.
- [128] Z. Zhang, Z. Zhang, P. Lu, G. Feng, W. Qi, Soil water-salt dynamics and maize growth as affected by cutting length of topsoil incorporation straw under brackish water irrigation, *Agronomy*, 10 (2020), Issue 2.

Bibliographie

- [129] X. Fu, F-J. Duan, T-T. Huang, L. Ma, J-J Jiang, Y-C Li, A fast variable selection method for quantitative analysis of soils using laser-induced breakdown spectroscopy, *Journal of Analytical Atomic Spectrometry*, 32 (2017), Issue 6, pp 1166-1176.
- [130] J. Peng, Y. He, L. Ye, T. Shen, F. Liu, W. Kong, X. Liu, Y. Zhao, Moisture influence reducing method for heavy metals detection in plant materials using laser-induced breakdown spectroscopy : a case study for chromium content detection in rice leaves, *Analytical Chemistry*, 89 (2017), pp 7593-7600.

AGROLIBS : Spectroscopie de plasma induit par laser au service de la gestion durable des ressources agricoles

RESUME Dans un contexte de changement climatique mondial, l'optimisation de la gestion des ressources agricoles, et en particulier des éléments minéraux des sols, est devenue un enjeu majeur. Les éléments minéraux sont à la fois des ressources indispensables pour la croissance des végétaux et le stockage du carbone, des polluants potentiels de certains compartiments de l'environnement, et des contaminants potentiellement toxiques susceptibles d'être transférés depuis les végétaux jusqu'à l'homme. Les méthodes actuelles de dosage de ces éléments sont très coûteuses en temps et en argent et d'un point de vue opérationnel, il est difficile dans ces conditions de réaliser un diagnostic fréquent sur le terrain d'une situation de carence ou de contamination, ou encore de caractériser la variabilité spatio-temporelle des caractéristiques d'une parcelle agricole.

L'émergence récente de capteurs et de systèmes portatifs a permis de développer de nouvelles méthodes pour mesurer les propriétés du sol et de la plante et compléter ou remplacer les techniques conventionnelles de laboratoire. C'est le cas de la LIBS, méthode de spectroscopie d'émission atomique à partir d'un plasma créé par laser, rapide et verte dans la mesure où elle ne s'appuie pas sur l'usage de réactifs chimiques dangereux. L'objectif de cette étude est d'évaluer les performances de la méthode LIBS, et plus précisément d'un instrument portable, appliquée à l'analyse multi-élémentaire quantitative d'échantillons de plantes et de sols agricoles. L'évaluation de cette méthode prend en compte la préparation des échantillons et les étapes de traitement des spectres LIBS. Elle a été mise en œuvre selon trois axes : i. La caractérisation des échantillons de plante d'espèces variées dans des conditions idéales, ii. La caractérisation des échantillons de sols représentatifs des grandes cultures françaises agricoles (blé, tournesol) toujours dans les conditions idéales et iii. L'étude de l'impact de facteurs dégradant les conditions de mesure afin d'évaluer la possibilité de mise en œuvre de la technique LIBS en dehors du laboratoire.

Notre étude, basée sur l'utilisation du pistolet LIBS Z300 de SciAps, a permis de détecter les éléments C, Ca, Fe, K, Mg, Na, Si et P, concentrés dans la gamme du g/kg dans les échantillons de sols et de plantes, aussi bien en conditions de laboratoire qu'en conditions volontairement dégradées. En revanche, ni l'azote, ni les éléments Cd, Cu, Mn et Zn n'ont pu être détectés dans ces échantillons. En ce qui concerne l'analyse des échantillons de plantes, les éléments Ca, Fe, Mg, Na, et P, ont été quantifiés à l'aide de modèles de régression univariés couplés à une stratégie de normalisation adaptée. L'analyse quantitative des sols a nécessité quant à elle, l'emploi de modèles PLS, afin de prendre en compte les effets de matrice, et d'obtenir des modèles d'étalonnage satisfaisants pour les éléments Mg, Na et Fe. En ce qui concerne les analyses en conditions dégradées, nous avons établi, à l'aide d'un plan d'expériences, que le taux d'humidité, supérieur à 40% dans les plantes et à 20% dans les sols était l'un des principaux freins aux analyses LIBS de terrain. Nous avons aussi montré que la baisse de signal LIBS observée lorsque le taux d'humidité augmente pouvait être corrigée en appliquant une étape de normalisation.

Le bilan de notre étude est partiellement satisfaisant sur le plan des performances quantitatives des modèles de régression, y compris dans les conditions idéales de laboratoire. La diversité des conditions physico-chimiques rencontrées sur le terrain fait que les analyses LIBS directes ne sont clairement pas réalistes. Le principe d'un laboratoire de campagne, avec une préparation simplifiée des échantillons, pourrait cependant être envisagé, afin de pouvoir placer la technique LIBS au cœur de nouvelles stratégies d'échantillonnage dans le contexte de l'agriculture de précision.

MOTS-CLES spectroscopie sur plasma induit par laser (LIBS) – agronomie – sols – plantes – analyse multi-élémentaire – chimiométrie

AGROLIBS : Laser-induced breakdown spectroscopy for the sustainable management of agricultural resources

ABSTRACT In a context of global climate change, optimizing the management of agricultural resources, and in particular the mineral elements in soils, has become a major challenge. Mineral elements are at the same time indispensable resources for plant growth and carbon storage, potential pollutants in certain environmental compartments, and potentially toxic contaminants that can be transferred from plants to humans. Current methods for the determination of these elements are very costly in terms of time and money, and from an operational point of view, it is difficult in these conditions to carry out a frequent diagnosis in the field of a deficiency or contamination situation, or to characterize the spatio-temporal variability of the characteristics of an agricultural land.

The recent emergence of portable sensors and systems has made it possible to develop new methods for measuring soil and plant properties and to complement or replace conventional laboratory techniques. This is the case of LIBS, a method of atomic emission spectroscopy from a laser-generated plasma, which is fast and green as it does not rely on the use of hazardous chemical reagents. The objective of this study is to evaluate the performance of the LIBS method, and more precisely of a portable instrument, applied to the quantitative multi-elemental analysis of plant and agricultural soil samples. The evaluation of this method takes into account the sample preparation and processing steps of LIBS spectra. It was implemented according to three axes: i. The characterization of plant samples of various species under ideal conditions, ii. The characterization of soil samples representative of the major French agricultural crops (wheat, corn, sunflower) always under ideal conditions and iii. The study of the impact of factors degrading the measurement conditions in order to evaluate the possibility of implementing the LIBS technique outside the laboratory.

Our study, based on the use of the SciAps Z300 LIBS, allowed the detection of the elements C, Ca, Fe, K, Mg, Na, Si and P, concentrated in the g/kg range in soil and plant samples, both in laboratory conditions and in deliberately degraded conditions. On the other hand, neither nitrogen nor the elements Cd, Cu, Mn and Zn could be detected in these samples. For the analysis of plant samples, the elements Ca, Fe, Mg, Na, and P were quantified using univariate regression models coupled with an adapted normalization strategy. Quantitative soil analysis required the use of PLS models to account for matrix effects. As regards the analyses in degraded conditions, we established, using an experimental design, that the moisture content, higher than 40% in plants and 20% in soils was one of the main obstacles to LIBS field analyses. We also showed that the drop in LIBS signal observed when the moisture content increases could be corrected by applying a normalization step.

The results of our study are partially satisfactory in terms of the quantitative performance of the regression models, even under ideal laboratory conditions. The diversity of physico-chemical conditions encountered in the field means that direct LIBS analyses are clearly not realistic. The principle of a field laboratory, with simplified sample preparation, could however be envisaged, in order to place the LIBS technique at the heart of new sampling strategies in the context of precision agriculture.

KEYWORDS laser-induced breakdown spectroscopy (LIBS) – agronomy – soils – plants – multi-elemental analysis – chemometrics