

Centre Eau Terre Environnement

**DÉVELOPPEMENT D'UNE APPROCHE ADAPTATIVE POUR
L'ESTIMATION DES COMPOSANTES DE LA TEXTURE DU SOL ET
SON CONTENU EN MATIÈRE ORGANIQUE PAR SPECTROSCOPIE
VISIBLE-PROCHE INFRAROUGE**

Par
Amal Hmaissia

Mémoire présenté pour l'obtention du grade de
Maître ès Sciences (M.Sc.)
en sciences de l'eau

Jury d'évaluation

Président du jury et
examineur interne

Saeid Homayouni
INRS-ETE, Université du Québec

Examineur externe

Viacheslav Adamchuk
Faculté des sciences agricoles et
environnementales, Université McGill

Directeur de recherche

Karem Chokmani
INRS-ETE, Université du Québec

Codirecteur de recherche

Athyna Cambouris
Agriculture et Agroalimentaire Canada

REMERCIEMENTS

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma gratitude.

Je tiens à exprimer ma reconnaissance à mon directeur de maîtrise, Karem Chokmani. Je le remercie de m'avoir encadré, orienté et aidé. Ses conseils précieux ont contribué à alimenter ma réflexion.

Je désire aussi remercier ma codirectrice de recherche Athyna Cambouris, qui m'a fourni les échantillons de sol nécessaires à la réussite de ce projet et qui a lu et corrigé mon mémoire de maîtrise.

Un grand merci à Anas, Hachem et Rachid qui par leurs paroles, leurs écrits, leurs conseils et leurs critiques ont guidé mes réflexions et ont accepté de me rencontrer et de répondre à mes questions durant mes recherches.

Je tiens à remercier aussi affectueusement mes chers parents Tayaa et Fatma, mon frère Sofien et ma petite Aya pour leur soutien moral inestimable qu'ils ont su m'apporter tout au long de ces deux années.

Un grand merci à mes amis qui ont été toujours là pour moi.

Finalement, j'aimerais exprimer ma gratitude à toutes les personnes, trop nombreux pour les citer, qui ont participé de loin ou de près dans l'avancement de mon analyse.

RÉSUMÉ

Le suivi continu de la qualité du sol représente un défi dans l'activité agricole. Pour subvenir à ce besoin, les scientifiques tentent à développer une méthode rapide et peu coûteuse de caractérisation des propriétés du sol. Ainsi, en se reposant sur la réponse spectrale du sol, la spectroscopie est devenue une méthode prometteuse pour quantifier ses paramètres. Cependant, la performance de cette méthode dépend de plusieurs facteurs comme l'humidité du sol. La présente étude vise à étudier quantitativement et qualitativement l'effet de la teneur en eau sur l'estimation des teneurs en argile, sable et matière organique par spectroscopie afin de développer une approche adaptative servant à leur estimation. Des mesures de réflectance au laboratoire sur 68 échantillons du sol ont été réalisées en variant la teneur en eau pour chaque échantillon. La classification des spectres selon le niveau de la teneur en eau (C1 : entre 0-10 %, C2 : entre 10-19 % et C3 : supérieure à 19 %) a permis une comparaison entre des modèles établis à partir de toute la base de données spectrales et ceux établis à partir des spectres de chaque classe ainsi qu'une deuxième comparaison inter classes des modèles d'estimation des propriétés étudiées. Après classification des spectres selon la teneur en eau des échantillons, la précision de l'estimation de la teneur en argile a été améliorée pour les classes C1 et C3 alors que pour la teneur en matière organique, l'amélioration de la précision des modèles a été observée chez les classes C1 et C2. Quant à la teneur en sable, seule la classe C3 présentait une meilleure performance après classification des spectres. Pour ce qui est comparaison inter classes, les échantillons les plus secs ont permis l'estimation la plus précise de la teneur en argile ($R^2 = 0.76$ et $RPD = 2.09$), contrairement à la teneur en sable où les échantillons les plus humides permettent la meilleure estimation de cette propriété ($R^2 = 0.81$ et $RPD = 1.86$). Pour la teneur en matière organique les échantillons appartenant à la classe C2 présentent le modèle le plus précis ($R^2 = 0.81$ et $RPD = 2.30$). L'étude de l'effet de la teneur en eau permettra de savoir si l'élimination de son influence sur les spectres améliorera ou non la précision d'estimation des variables texturale et organique du sol. Ainsi, les niveaux d'humidité du sol ayant un impact important sur la détermination des propriétés en question seront identifiés.

Mots-clés : spectroscopie, effet de la teneur en eau, approche adaptative, estimation des teneurs en argile, sable et matière organique

ABSTRACT

Continuous monitoring of soil quality is a challenge in agricultural activity. To achieve this need, scientists are trying to develop quick and inexpensive methods of characterizing soil properties. Thus, relying on the spectral response of the soil, spectroscopy has become a promising method for quantifying its parameters. However, this method remains sensitive to several factors such as soil moisture. The present study aims to study quantitatively and qualitatively the effect of water content on the estimation of clay, sand and organic matter content by spectroscopy in order to develop an adaptive approach for their estimation. Reflectance measurements in the laboratory on 68 soil samples were performed by varying the water content for each sample. The classification of the spectra according to the level of the water content (C1: between 0-10%, C2: between 10-19% and C3: greater than 19%) allowed a comparison between models established from the whole spectral database and those established from the spectra of each class as well as a second inter-class comparison of the models of estimation of the studied properties. After classification of the spectra according to the water content of the samples, the precision of the estimation of the clay content has been improved for classes C1 and C3 whereas for the organic matter content, the improvement of the precision of the models was observed in classes C1 and C2. As for the sand content, only class C3 presented a better performance after classification of the spectra. For inter-class comparison, the driest samples allowed the most precise estimation of the clay content ($R^2 = 0.76$ and $RPD = 2.09$), unlike the sand content where the wettest samples allow the best estimation of this property ($R^2 = 0.81$ and $RPD = 1.86$). For the organic matter content, samples belonging to class C2 present the most precise model ($R^2 = 0.81$ and $RPD = 2.30$). The study of the effect of the water content will make it possible to know whether the elimination of its influence on the spectra will improve or not the precision of estimation of the textural and organic variables of the soil. Thus, the soil moisture levels having a significant impact on the determination of the properties in question will be identified.

Keywords : spectroscopy, effect of water content, adaptive approach, estimation of clay, sand and organic matter content

TABLE DES MATIÈRES

REMERCIEMENTS	III
RÉSUMÉ	V
ABSTRACT	VII
TABLE DES MATIÈRES	IX
LISTE DES FIGURES.....	XII
LISTE DES TABLEAUX	XVII
LISTE DES ABRÉVIATIONS.....	XIX
1 INTRODUCTION.....	1
1.1 CONTEXTE	1
1.2 PROBLÉMATIQUE	1
1.3 OBJECTIFS	2
2 REVUE DE LITTÉRATURE	5
2.1 METHODES DE MESURE DES PROPRIETES DU SOL AU LABORATOIRE	5
2.1.1 <i>Échantillonnage du sol</i>	5
2.1.2 <i>Analyse des propriétés du sol</i>	6
2.2 MÉTHODES DE MESURE DES PROPRIÉTÉS DU SOL SUR TERRAIN.....	9
2.2.1 <i>Méthode électromagnétique</i>	9
2.2.2 <i>Méthode mécanique</i>	11
2.2.3 <i>Méthode électrochimique</i>	12
2.3 LA SPECTROSCOPIE DU SOL	14
2.3.1 <i>Définition de la réflexion de la lumière</i>	14
2.3.2 <i>Définition de la spectroscopie</i>	14
2.3.3 <i>La spectroscopie du sol</i>	15
2.3.4 <i>Particularités spectrales des propriétés de sol</i>	16
2.3.5 <i>Avantages et inconvénients de la spectroscopie du sol</i>	18
2.3.6 <i>Application de la spectroscopie</i>	19
3 MATÉRIELS ET MÉTHODES.....	27
3.1 PRESENTATION DE LA ZONE D'ETUDE	27

3.2	PROTOCOLE EXPERIMENTAL.....	29
3.2.1	<i>Préparation des échantillons de sol</i>	29
3.2.2	<i>Mesure de la réflectance</i>	31
3.2.3	<i>Mesure de la teneur en eau</i>	33
3.2.4	<i>Calcul des moyennes</i>	35
3.3	TRAITEMENT DES SPECTRES.....	36
3.3.1	<i>Nettoyage de la base des données spectrales</i>	38
3.3.2	<i>Exploration des données</i>	38
3.3.3	<i>Prétraitements des spectres</i>	41
3.3.4	<i>Estimation des propriétés de sol</i>	45
4	RÉSULTATS ET DISCUSSIONS	53
4.1	DESCRIPTION DE LA BASE DE DONNEES.....	53
4.1.1	<i>Propriétés du sol</i>	53
4.1.2	<i>Données spectrales</i>	53
4.2	RESULTATS DES TRAITEMENTS DES SPECTRES	54
4.2.1	<i>Exploration des spectres</i>	54
4.3	ESTIMATION DES PROPRIETES DU SOL.....	57
4.3.1	<i>Prétraitements appliqués</i>	57
4.3.2	<i>Choix des niveaux de la teneur en eau</i>	59
4.3.3	<i>Calibration des modèles de régression PLS</i>	61
5	CONCLUSION	89
6	BIBLIOGRAPHIE.....	93
7	ANNEXE	105

LISTE DES FIGURES

FIGURE 2.1 : DETERMINATION DE LA STRUCTURE DU SOL A PARTIR DE LA DISTRIBUTION DE LA TAILLE DES MOTTES POUR 3 COMPARTIMENTS DU NORD DE LA FRANCE (ADAPTEE DE ROGER-ESTRADE ET AL., 2004).	8
FIGURE 2.2 : SCHEMA D'UN CAPTEUR DE CONDUCTIVITE DE SOL (GEONICS EM38) EN MODE DIPOLE VERTICAL (TIREE DE SUDDUTH ET AL., 2001).	10
FIGURE 2.3 : SCHEMA D'UN PENETROMETRE (TIREE DE ANDERSON ET AL., 1980).	12
FIGURE 2.4 : SPECTRE ELECTROMAGNETIQUE AVEC LES DOMAINES DU VISIBLE ET PROCHE INFRAROUGES (ROSSEL <i>ET AL.</i> , 2006) ADAPTE DE MCBRATNEY ET AL. (2006).	15
FIGURE 2.5 : SIGNATURES SPECTRALES DES ECHANTILLONS DE SOL MONTRANT LES PICS D'ABSORPTION DE L'EAU A 1400 ET 1900 NM (NOCITA <i>ET AL.</i> , 2013).	16
FIGURE 3.1 : LOCALISATION DES CHAMPS EXPERIMENTAUX (A DROITE : CHAMPS 1 ET 2 SITUES A STE-CATHERINE-DE-LA-JACQUES-CARTIER, A GAUCHE CHAMPS 3 ET 4 SITUES A L'ÎLE D'ORLEANS).	27
FIGURE 3.2 : PRELEVEMENT D'ECHANTILLONS DE SOL PAR UNE TARIERE	28
FIGURE 3.3 : REPARTITION DE CHAQUE ECHANTILLON COMPOSITE EN 3 REPLICAS.	30
FIGURE 3.4 : SYSTEME D'OCEAN OPTICS DE MESURE DE LA REFLECTANCE DES ECHANTILLONS DE SOL COMPOSE D'UN SUPPORT DE LA SONDE DE LUMIERE (A), D'UNE LAMPE HALOGENE (B), DE DEUX SPECTROMETRES (C ET D), D'UNE SONDE DE REFLECTANCE (E) ET D'UNE REFERENCE DE CALIBRATION DES SPECTROMETRES (F).	32
FIGURE 3.5 : MESURE DU POIDS DES ECHANTILLONS DU SOL.	33
FIGURE 3.6 : ORGANIGRAMME DE DIFFERENTS TRAITEMENTS DES ECHANTILLONS DU SOL.	35
FIGURE 3.7 : ORGANIGRAMME DES DIFFERENTS TRAITEMENTS DES SPECTRES.	37
FIGURE 3.8 : PRINCIPE DE L'ACP MONTRANT LE CALCUL DES COMPOSANTES PRINCIPALES T1 ET T2 ET LA PROJECTION DES VARIABLES DANS L'ESPACE FORME PAR CES DEUX COMPOSANTES (ADAPTEE DE SEBASTIEN, 2016).	39
FIGURE 3.9 : PRINCIPE DE L'ALGORITHME K-MEANS MONTRANT LE PARTITIONNEMENT DE NUAGE DE POINTS EN CLASSES (ADAPTEE DE HTTPS://AWS.AMAZON.COM/FR/BLOGS/MACHINE-LEARNING/K-MEANS-CLUSTERING-WITH-AMAZON-SAGEMAKER/).	40

FIGURE 3.10 : PRINCIPES GEOMETRIQUES DE LA REGRESSION PLS (PARTIAL LEAST SQUARES OU PROJECTION TO LATENT STRUCTURES) MONTRANT LA MAXIMISATION DE LA CORRELATION ENTRE X ET Y (TIREE DE RUTLEDGE, NO DATE).....	48
FIGURE 4.1 : SPECTRES VISIBLES ET PROCHE INFRAROUGES AVANT (A) ET APRES (B) LA REDUCTION DES LONGUEURS D'ONDE PAR LE CALCUL DES MOYENNES DES REPLICATS.....	54
FIGURE 4.2 : SPECTRES MOYENS VISIBLES ET PROCHE INFRAROUGES DE 5 INTERVALLES DE NIVEAUX DE TENEUR EN EAU.	55
FIGURE 4.3 : PROJECTION EN 2-DIMENSIONS DES DEUX PREMIERES COMPOSANTES DE L'ACP DES SPECTRES DE CALIBRATION AVEC REPRESENTATION DES NIVEAUX D'HUMIDITE DE SOL EN COULEURS DIFFERENTES.	56
FIGURE 4.4 : SPECTRES VISIBLES (A) ET PROCHE INFRAROUGES (B) EN FONCTION DES MOYENNES DES REFLECTANCE POUR CHAQUE LONGUEUR D'ONDE.	57
FIGURE 4.5 : VARIATION DES VALEURS DE RMSECV AVEC L'ORDRE DE LA DERIVEE DE SAVITZKY-GOLAY ET DE LA TAILLE DE LA FENETRE.....	58
FIGURE 4.6 : SPECTRES VISIBLES (A) ET PROCHE INFRAROUGES (B) PRETRAITES.....	59
FIGURE 4.7 : PROJECTION EN 2-DIMENSIONS MONTRANT LES CLASSES SPECTRALES : C1, C2, C3 ET C4 ISSUES DE LA CLASSIFICATION NON SUPERVISEE DES SPECTRES (A) ET LES CLASSES D'HUMIDITE DU SOL : ECHANTILLONS DE TENEURS EN EAU ENTRE 0 ET 10 %, ENTRE 10 ET 19 % ET SUPERIEURES A 19 % ISSUES DE L'HOMOGENEISATION INTRA-CLASSES SPECTRALES DES TENEURS EN EAU (B).	60
FIGURE 4.8 : CORRELOGRAMME DE LA TENEUR EN EAU ET DES LONGUEURS D'ONDE.....	62
FIGURE 4.9 : CORRELOGRAMMES DE LA TENEUR EN ARGILE ET DES LONGUEURS D'ONDE POUR TOUS LE SOUS-ENSEMBLE DE CALIBRATION ET POUR LES TROIS CLASSES D'HUMIDITE DE SOL (0-10 %, 10-19 % ET >19 % DE LA TENEUR EN EAU DU SOL).	63
FIGURE 4.10 : CORRELOGRAMMES DE LA TENEUR EN SABLE ET DES LONGUEURS D'ONDE POUR TOUS LE SOUS-ENSEMBLE DE CALIBRATION ET POUR LES TROIS CLASSES D'HUMIDITE DE SOL (0-10 %, 10-19 % ET >19 % DE LA TENEUR EN EAU DU SOL).	63
FIGURE 4.11 : CORRELOGRAMMES DE LA TENEUR EN MATIERE ORGANIQUE ET DES LONGUEURS D'ONDE POUR TOUT LE SOUS-ENSEMBLE DE CALIBRATION ET POUR LES TROIS CLASSES D'HUMIDITE DE SOL (0-10 %, 10-19 % ET >19 % DE LA TENEUR EN EAU DU SOL).....	64

FIGURE 4.12 : VARIATION DES COEFFICIENTS DE DETERMINATION (R^2) ET RMSECV AVEC LES VARIABLES LATENTES LORS DE LA VALIDATION DES MODELES D'ESTIMATION DE LA TENEUR EN EAU, EN ARGILE, EN SABLE ET EN MATIERE ORGANIQUE. LES ELLIPSES DISCONTINUES REPRESENTENT LES COMPOSANTES PLS SELECTIONNEES POUR L'ESTIMATION DES PROPRIETES DE SOL ETUDIEES ET CORRESPONDANT A LA MEILLEURE COMBINAISON (R^2 LE PLUS ELEVE ET RMSECV LE PLUS FAIBLE).
65

FIGURE 4.13 : VARIATION DES COEFFICIENTS DE DETERMINATION (R^2) ET RMSECV AVEC LES VARIABLES LATENTES LORS DE LA VALIDATION DES MODELES D'ESTIMATION DE LA TENEUR EN ARGILE POUR LES TROIS CLASSES D'HUMIDITE DE SOL. LES ELLIPSES DISCONTINUES REPRESENTENT LES COMPOSANTES PLS SELECTIONNEES POUR L'ESTIMATION DE LA TEXTURE DU SOL ET CORRESPONDANT A LA MEILLEURE COMBINAISON (R^2 LE PLUS ELEVE ET RMSECV LE PLUS FAIBLE). C1, C2 ET C3 REPRESENTENT LES TROIS CLASSES D'HUMIDITE DU SOL (0-10 %, 10-19 % ET >19 % DE LA TENEUR EN EAU DU SOL).66

FIGURE 4.14 : VARIATION DES COEFFICIENTS DE DETERMINATION (R^2) ET RMSECV AVEC LES VARIABLES LATENTES LORS DE LA VALIDATION DES MODELES D'ESTIMATION DE LA TENEUR EN SABLE POUR LES TROIS CLASSES D'HUMIDITE DE SOL. LES ELLIPSES DISCONTINUES REPRESENTENT LES COMPOSANTES PLS SELECTIONNEES POUR L'ESTIMATION DE LA TEXTURE DU SOL ET CORRESPONDANT A LA MEILLEURE COMBINAISON (R^2 LE PLUS ELEVE ET RMSECV LE PLUS FAIBLE). C1, C2 ET C3 REPRESENTENT LES TROIS CLASSES D'HUMIDITE DU SOL (0-10 %, 10-19 % ET >19 % DE LA TENEUR EN EAU DU SOL).66

FIGURE 4.15 : VARIATION DES COEFFICIENTS DE DETERMINATION (R^2) ET RMSECV AVEC LES VARIABLES LATENTES LORS DE LA VALIDATION DES MODELES D'ESTIMATION DE LA TENEUR EN MATIERE ORGANIQUE POUR LES TROIS CLASSES D'HUMIDITE DE SOL. LES ELLIPSES DISCONTINUES REPRESENTENT LES COMPOSANTES PLS SELECTIONNEES POUR L'ESTIMATION DE LA TEXTURE DU SOL ET CORRESPONDANT A LA MEILLEURE COMBINAISON (R^2 LE PLUS ELEVE ET RMSECV LE PLUS FAIBLE). C1, C2 ET C3 REPRESENTENT LES TROIS CLASSES D'HUMIDITE DU SOL (0-10 %, 10-19 % ET >19 % DE LA TENEUR EN EAU DU SOL).67

FIGURE 4.16 : VALEURS DE TENEUR EN EAU OBSERVEES EN FONCTION DES COEFFICIENTS ISSUS DE L'APPLICATION DE LA PLS.69

FIGURE 4.17 : VALEURS DE TENEUR EN ARGILE OBSERVEES EN FONCTION DES COEFFICIENTS ISSUS DE L'APPLICATION DE LA PLS.....69

FIGURE 4.18 : VALEURS DE TENEUR EN SABLE OBSERVEES EN FONCTION DES COEFFICIENTS ISSUS DE L'APPLICATION DE LA PLS.....	70
FIGURE 4.19 : VALEURS DE TENEUR EN MATIERE ORGANIQUE OBSERVEES EN FONCTION DES COEFFICIENTS ISSUS DE L'APPLICATION DE LA PLS.	71
FIGURE 4.20 : VALEURS DE TENEUR EN ARGILE OBSERVEES EN FONCTION DES COEFFICIENTS ISSUS DE L'APPLICATION DE LA PLS POUR LES TROIS CLASSES D'HUMIDITE DU SOL C1 (0-10 %), C2 (10-19 %) ET C3 (>19 %).	73
FIGURE 4.21 : VALEURS DE TENEUR EN SABLE OBSERVEES EN FONCTION DES COEFFICIENTS ISSUS DE L'APPLICATION DE LA PLS POUR LES TROIS CLASSES D'HUMIDITE DE SOL C1 (0-10 %), C2 (10-19 %) ET C3 (>19 %).	74
FIGURE 4.22 : VALEURS DE TENEUR EN MATIERE ORGANIQUE OBSERVEES EN FONCTION DES COEFFICIENTS ISSUS DE L'APPLICATION DE LA PLS POUR LES TROIS CLASSES D'HUMIDITE DE SOL C1 (0-10 %), C2 (10-19 %) ET C3 (>19 %).	75
FIGURE 4.23 : TENEURS EN EAU ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE), (RMSE ET LE BIAIS SONT EN %).	77
FIGURE 4.24 : TENEURS EN ARGILE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE), (RMSE ET LE BIAIS SONT EN %).	78
FIGURE 4.25 : TENEURS EN SABLE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE), (RMSE ET LE BIAIS SONT EN %).	78
FIGURE 4.26 : TENEURS EN MATIERE ORGANIQUE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE), (RMSE ET LE BIAIS SONT EN %).	79
FIGURE 4.27 : TENEURS EN ARGILE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE) POUR LA CLASSE C1 (0-10 %), (RMSE ET LE BIAIS SONT EN %).	81
FIGURE 4.28 : TENEURS EN ARGILE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE) POUR LA CLASSE C2 (10-19 %), (RMSE ET LE BIAIS SONT EN %).	82
FIGURE 4.29 : TENEURS EN ARGILE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE) POUR LA CLASSE C3 (>19 %), (RMSE ET LE BIAIS SONT EN %).	82

FIGURE 4.30 : TENEURS EN SABLE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE) POUR LA CLASSE C1 (0-10 %), (RMSE ET LE BIAIS SONT EN %).	83
FIGURE 4.31 : TENEURS EN SABLE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE) POUR LA CLASSE C2 (10-19 %), (RMSE ET LE BIAIS SONT EN %).	83
FIGURE 4.32 : TENEURS EN SABLE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE) POUR LA CLASSE C3 (>19 %), (RMSE ET LE BIAIS SONT EN %).	84
FIGURE 4.33 : TENEURS EN MATIERE ORGANIQUE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE) POUR LA CLASSE C1 (0-10 %), (RMSE ET LE BIAIS SONT EN %).	85
FIGURE 4.34 : TENEURS EN MATIERE ORGANIQUE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE) POUR LA CLASSE C2 (10-19 %), (RMSE ET LE BIAIS SONT EN %).	85
FIGURE 4.35 : TENEURS EN MATIERE ORGANIQUE ESTIMEES EN FONCTION DE CELLES OBSERVEES AVANT (A GAUCHE) ET APRES SOUSTRACTION DU BIAIS (A DROITE) POUR LA CLASSE C3 (>19 %), (RMSE ET LE BIAIS SONT EN %).	86

LISTE DES TABLEAUX

TABLEAU 2.1 : SOURCES DES METHODES LES PLUS UTILISEES POUR LA PREDICTION DES PROPRIETES DE SOL.....	25
TABLEAU 3.1 : CLASSIFICATION ET TEXTURES DES ECHANTILLONS DE SOL.....	29
TABLEAU 3.2 : CARACTERISTIQUES DES ALGORITHMES DE PRETRAITEMENTS TESTES POUR LES SPECTRES VISIBLES ET PROCHE INFRAROUGE APPLIQUES POUR LE SOUS-ENSEMBLE DE CALIBRATION.	45
TABLEAU 4.1 : STATISTIQUE DESCRIPTIVE POUR LES DIFFERENTES PROPRIETES ETUDIEES.	53
TABLEAU 4.2 : VARIABLES LATENTES RETENUES POUR L'ESTIMATION DES PROPRIETES DE SOL ETUDIEES.	67

LISTE DES ABRÉVIATIONS

ACP : Analyse en Composantes Principales

CAC : Capacité d'Échange Cationique

CARS : Competitive Adaptive Reweighted Sampling

cm : centimètre

cov : covariance

GPS : Global Positioning System

MARS : Multivariate Adaptive Regression Splines

mm : millimètre, millimètre

nm : nanomètre

PLS : Partial Least Square

RMSECV : Root Mean Square Error of Cross Validation, Root Mean Square Error Cross Validation

RMSEP : Root Mean Squared Error of Prediction

RPD : Residual Prediction Deviation

SVN : Standard Normal Variate

1 INTRODUCTION

1.1 Contexte

Le sol est l'une des ressources indispensables dans le développement de l'agriculture et dans la préservation de l'environnement puisqu'il joue des rôles importants dans le stockage de l'eau, la régulation de la température de l'air et le stockage du carbone. En outre, le sol est le support qui fournit les éléments nutritifs aux plantes et abrite les microorganismes responsables de la décomposition de la matière organique (Brabant, 2008). Cependant, cette ressource est en déclin jour après jour laissant une part de 0.20 hectare disponible par habitant dans le monde en 2100 si on tient compte de la vitesse de croissance de la population et celle de la dégradation du sol (Brabant, 2008). Toutes ces constatations incitent à la préservation de cette ressource à travers le suivi continu de sa qualité. Mais les méthodes de détermination des propriétés du sol sont coûteuses, laborieuses et lentes. Pour cela, plusieurs méthodes reposant sur les réponses optiques du sol à des variations de ces propriétés physicochimiques ont été testées (Korsaeth, 2005; Crozier et al., 1999; Yegul et al., 2014). L'efficacité de la méthode spectroscopique pour l'atteinte de ces objectifs n'est plus à prouver (Li et al., 2015; Zornoza et al., 2008; Shi et al., 2015; Knox et al., 2015; Genot et al., 2011). Grâce à la simplicité de son application, son caractère non polluant (Rossel et al., 2006) et le coût faible associé à son utilisation, la spectroscopie peut devenir la méthode parfaite pour la quantification des propriétés du sol en continu.

1.2 Problématique

Malgré son potentiel à se substituer aux méthodes conventionnelles, la spectroscopie est sensible aux facteurs environnementaux. Ce qui explique en partie la différence de précision entre les estimations résultant de l'acquisition des spectres sur-le-champ et ceux au laboratoire. L'exemple de l'impact de la teneur en eau sur la signature spectrale a été largement étudié à cause de son effet significatif sur cette dernière qui peut être plus remarquable que celui des types du sol (Muller et Décamps, 2001). En effet, l'eau engendre une diminution générale de la réflectance ainsi que l'augmentation de l'absorption en profondeur et en largeur près de 1400 et 1900 nm (Ji et al., 2015). Pour des sols humides, la différence entre l'indice de réfraction du sol et celui de l'eau est inférieure à celle observée pour des sols secs entourés par l'air (Bach et Mauser, 1994). Cet abaissement dans l'indice de réfraction au niveau des points de contact sol-eau-air provoque une diminution de la diffusion de la lumière incidente par rapport aux sols

secs. En outre, la bande d'eau qui encercle les particules du sol produit une réflexion supplémentaire de l'énergie diffusée dans l'interface eau-air. Ce qui produit une énergie plus grande se propageant en profondeur dans le sol (Somers et al., 2010).

Dans la majorité des cas, les échantillons du sol subissent un séchage à l'air pour atteindre l'état sec avant l'acquisition de leurs spectres et leur analyse. Par la suite, les estimations des propriétés du sol à partir de leurs réponses spectrales déduites des échantillons secs ne peuvent pas être directement appliquées sur des échantillons collectés sur le terrain et ayant des niveaux d'humidité de sol variés. L'humidité du sol limite ainsi l'estimation de ses paramètres physicochimiques c'est pourquoi certains auteurs ont séparé les analyses selon des groupes de niveau de teneur en eau (Bach et Mauser, 1994), éliminé l'effet de l'eau sur l'estimation des paramètres du sol (Minasny et al., 2011) ou encore quantifié l'effet de l'eau au moment de la calibration des modèles d'estimation des propriétés de sol (Ji et al., 2015). Toutefois, la dernière étude s'est limitée à la quantification de la matière organique. En effet, à notre connaissance, aucune étude ne s'est intéressée à l'estimation conjointe de la texture et de la matière organique à partir des données spectrales tout en tenant compte du niveau de la teneur en eau de l'échantillon du sol analysé.

1.3 Objectifs

Dans cette étude, on s'intéresse donc au développement d'une approche d'estimation des paramètres de la texture du sol (argile et sable) et de son contenu en matière organique par spectroscopie visible et proche infrarouge. Cette approche, dite adaptative, prend en compte l'effet interférent de la teneur en eau du sol sur la réponse spectrale. Par conséquent, l'approche préconisée passe en premier lieu par la quantification de la teneur en eau du sol pour ensuite estimer la propriété de sol en question selon un modèle spécifique à ce niveau de teneur en eau.

Dans l'optique d'atteindre cet objectif, on a commencé par une revue de littérature qui rappelle les méthodes déjà utilisées pour la quantification des propriétés de sol et détaille l'allocation de la spectroscopie pour se substituer à ces outils. En deuxième lieu, on a expliqué la méthodologie adoptée pour la préparation des échantillons de sol, l'acquisition des spectres au laboratoire et le calcul de la teneur en eau ainsi que les étapes de traitements des spectres. Une grande importance a été accordée à l'amélioration de la qualité des spectres en sélectionnant la combinaison de prétraitements adéquate et en optimisant par la suite les paramètres de leurs algorithmes. L'analyse qualitative (étude de la corrélation entre chaque

propriété de sol et les longueurs d'onde, ACP et inspection visuelle des spectres) et quantitative (classification des spectres en groupes selon la teneur en eau des échantillons de sol, élaboration et amélioration des modèles d'estimation des propriétés de sol étudiées avant et après leur classification) des spectres a permis de souligner la nécessité d'élaboration de développement d'une approche adaptative pour la quantification des propriétés de sol par spectroscopie en vue de prendre en compte l'effet de l'eau sur celle-ci.

Puis, les résultats trouvés ont été présentés tout en expliquant leur fiabilité et limites à travers une analyse critique en présentant les sources d'erreurs liées à cette étude. Enfin, une conclusion rappelant l'objectif et les résultats les plus importants trouvés aussi bien que les limites et les perspectives de cette étude, a été proposée.

2 REVUE DE LITTÉRATURE

Dans cette partie, les méthodes de détermination de propriétés du sol sont réparties en deux catégories : les *méthodes conventionnelles* qui se basent sur des mesures dont la majorité s'effectue au laboratoire et des *méthodes optiques* qui sont appliquées directement sur terrain en se basant sur l'utilisation des capteurs proximaux.

2.1 Méthodes de mesure des propriétés du sol au laboratoire

Les méthodes conventionnelles de mesure des propriétés du sol peuvent être réparties en deux principales étapes : l'échantillonnage du sol et l'analyse des échantillons au laboratoire.

2.1.1 Échantillonnage du sol

Avant l'analyse du sol, l'étape primordiale est la sélection d'échantillons représentant la variabilité des propriétés du sol de la zone étudiée. Le diagnostic rapide de quelques paramètres du sol peut être réalisé sur le terrain avec des capteurs à main ou une tarière pour effectuer une caractérisation globale du sol. Cette caractérisation sert à calculer le nombre d'échantillons à prélever (Boone et al., 1999).

Pour diminuer le coût et le temps d'analyse du sol, il est privilégié de combiner des échantillons unitaires en un échantillon composite (Boone et al., 1999). Pour ce faire, il est nécessaire de respecter trois conditions :

- L'égalité des proportions des échantillons unitaires ;
- L'homogénéité de l'échantillon composite ;
- Une non-interaction significative entre les échantillons élémentaires.

Le temps et la fréquence d'échantillonnage varient selon les objectifs de l'étude. Par exemple, le pH et la teneur en nitrates présentent des variations inter- et intra-saisonnière également dépendantes des propriétés spécifiques du site. Donc, pour leur détermination, il faut effectuer des échantillonnages sous des conditions communes sur plusieurs années.

2.1.2 Analyse des propriétés du sol

Puisqu'il n'existe pas un protocole expérimental qui peut quantifier tous les paramètres du sol, cette partie sera subdivisée par propriétés. On verra d'abord la teneur en matière organique du sol, puis tour à tour, la texture, la structure et la teneur en eau du sol.

2.1.2.1 Matière organique

La détermination de la matière organique du sol se fait indirectement soit par destruction de la matière minérale ou par quantification des éléments chimiques présents constamment dans la matière organique comme le carbone ou l'azote (Schulte, 1984). Or, la quantité de l'azote change de façon importante en fonction de la source de la matière organique. Pour cela les méthodes souvent utilisées sont celles qui estiment le carbone total (Schulte, 1984). Selon Walkley et Armstrong (1993), les deux grands groupes de détermination de la teneur du sol en carbone totale sont :

- Les méthodes qui mesurent le dioxyde de carbone émis suite à l'oxydation de l'échantillon tel que la combustion sèche ;
- Les méthodes qui mesurent le pouvoir réducteur du sol par des agents oxydants comme la combustion humide avec l'aide chromique.

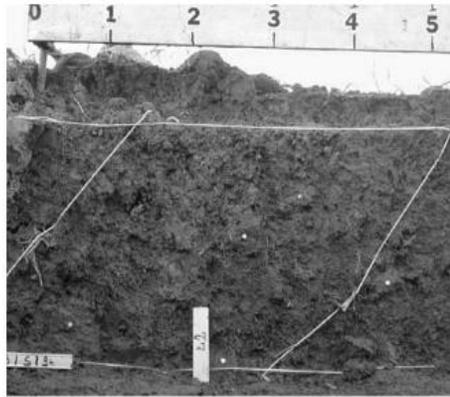
2.1.2.2 Texture du sol

Par définition, la texture du sol est « *les proportions relatives de limon, d'argile et de sable dans le sol* » (Taubner et al., 2009). Leur détermination repose le plus souvent sur des techniques de tamisage et de sédimentation ou d'analyse par diffraction laser (Taubner et al., 2009). Les techniques de sédimentation-tamisage doivent être lancées après séparation des particules grossières du sol avec utilisation de tamis pour retirer les particules dont le diamètre est supérieurs à 2 mm. Selon Kettler al. (2001), la deuxième étape consiste en la dispersion des particules en ajoutant un liquide, puis l'agitation de cette boue pendant 16 heures pour permettre d'isoler la fraction sablonneuse avec des tamis de maille supérieure à 0.053 mm. Le reste doit être remué, perturbé et laissé pendant 90 minutes au moins pour l'extraction de l'argile et 6 heures pour la fraction limoneuse. Après décantation, les masses d'argile et de limon doivent être séchées et pesées puis divisées par la masse de l'échantillon pour déterminer les teneurs en limon et en argile.

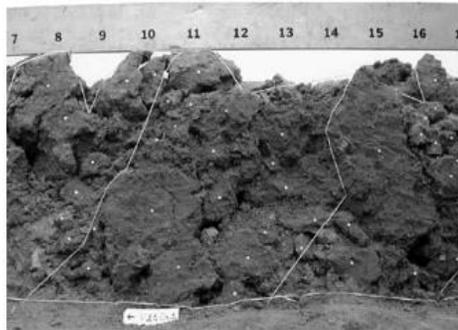
Il existe également une autre technique de mesure du diamètre des particules de sol basée sur la diffraction de la lumière. L'échantillon du sol à analyser doit d'abord être bien mélangé avec de l'eau pour mettre toutes les particules en suspension. Puis, des mesures de diffusion différentielle d'intensité de polarisation sont prises. Ces données optiques seront ensuite converties en tailles de particules à l'aide d'un modèle optique connu (Arriaga et al., 2006).

2.1.2.3 Structure du sol

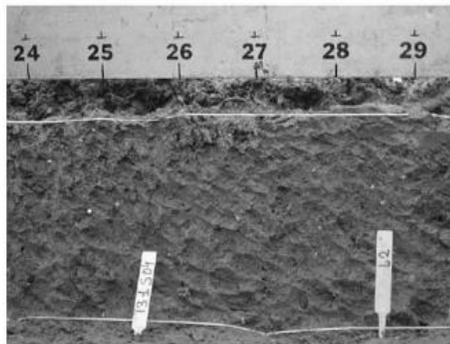
Par définition, la structure du sol est « *l'hétérogénéité spatiale des différentes composantes ou propriétés du sol* (Dexter, 1988) ». Ce paramètre peut être caractérisé de façon directe ou indirecte. Les *méthodes directes* sont basées sur l'étude morphologique des caractéristiques des unités structurales du sol. Ceci peut être fait en séparant des fragments stables du sol au laboratoire par tamisage et en étudiant leurs masses et tailles (Diaz-Zorita et al., 2002). Il est aussi possible d'extraire des mottes de sol et de les utiliser pour caractériser la structure du sol. Ceci se fait en se basant sur deux critères : la distribution des tailles des mottes et la classification des mottes selon leurs propriétés physiques internes (Figure 2.1) (Roger-estrade et al., 2004). Les *méthodes indirectes* sont plutôt basées sur l'estimation de la distribution des pores ou en étudiant la stabilité et la taille des particules. Puisque ces deux attributs peuvent avoir une corrélation avec la disposition des constituants du sol, ils sont souvent utilisés pour l'estimation de sa structure (Dexter, 1988).



Taille maximale des mottes $< 5\text{ cm}$
Structure fragmentaire



Taille moyenne des mottes $> 10\text{ cm}$
Structure fragmentaire grossière



Pas de distinction entre les mottes
Structure massive

Figure 2.1 : Détermination de la structure du sol à partir de la distribution de la taille des mottes pour 3 compartiments du nord de la France (adaptée de Roger-estrade et al., 2004).

2.1.2.4 Teneur en eau

La méthode standard de mesure de la teneur en eau s'appuie sur l'élimination de l'eau présente dans le sol et la quantification de cette masse. Ceci se fait soit par évaporation en utilisant la technique thermogravimétrique ou par des réactions chimiques en utilisant le carbure de calcium (Susha Lekshmi et al., 2014). Lors de la première technique, 100 g d'échantillon humide ou moins sont chauffés à l'étuve à 105 °C pendant 24 heures puis pesés à l'état sec, la différence entre le poids initial et le poids sec donne la quantité d'eau dans le sol (ASTMD, 2014). La deuxième technique consiste à mesurer la pression du gaz d'acétylène résultant de la réaction chimique entre le carbure de calcium et l'eau dans le sol (Blystone et al., 1962).

Quoique prouvées efficaces et précises, les méthodes usuelles de détermination des propriétés du sol sont lentes et coûteuses et font intervenir des éléments chimiques généralement polluants et ne permettent pas des mesures répétitives puisqu'elles changent certaines propriétés du sol. Pour cela, autres méthodes destinées à analyser des échantillons du sol sur le champ ont été développées.

2.2 Méthodes de mesure des propriétés du sol sur terrain

2.2.1 Méthode électromagnétique

Cette méthode repose sur la mesure de la capacité d'un sol donné à conduire le courant électrique et la variable physique utilisée pour quantifier cette capacité est soit la conductivité ou la résistivité électrique (Adamchuk et al., 2004). La détection de cette variable se fait en créant des charges électriques par une bobine placée à proximité du sol. Ces charges se transmettent dans le sol sous forme de boucles de courant dont la magnitude est directement proportionnelle à la conductivité électrique du sol. Chaque boucle génère un champ magnétique qui dépend du courant circulant dans la boucle et qui sera capté par une deuxième bobine de l'instrument, dite receveuse (Figure 2.2) (Yegul et al., 2014 ; Adamchuk et al., 2004).

Les valeurs de conductivité électrique sont influencées par les propriétés physicochimiques du sol ce qui permet la caractérisation de ses attributs à travers l'étude des relations empiriques entre eux et la conductivité électrique. Parmi les propriétés du sol caractérisées, on cite la capacité d'échange cationique, la teneur en argile, en limon et en eau (Yegul et al., 2014), en azote disponible (Korsaeth, 2005), le pH, en phosphore, en potassium et en matière organique (Tarr *et al.*, 2005). Des fortes relations ont été mises en évidence permettant ainsi de

caractériser la variabilité spatiale des propriétés de sol en dressant des cartes de conductivité donnant lieu à des zones de caractéristiques plus ou moins homogènes comme des cartes de texture de sol (Domsch et Giebel, 2004). Toutes ces études ont prouvé le potentiel de cette méthode à définir des différences et des similarités de propriétés de sol pour délimiter des zones homogènes de sol. Cependant, plusieurs limites liées à cette méthode ont été identifiées, notamment les effets de la température et de la teneur en eau du sol sur la variation de la conductivité électrique ainsi que la dérivée de l'instrument en fonction du temps et le décalage entre la position du récepteur GPS et celle dans laquelle se trouve le capteur électromagnétique (Sudduth et al., 2001). Domsch et Giebel (2004) ont révélé d'autres restrictions liées à l'applicabilité de cette méthode telle que l'uniformité de l'humidité de sol et la non-reproductibilité des relations entre la conductivité électrique et la texture.

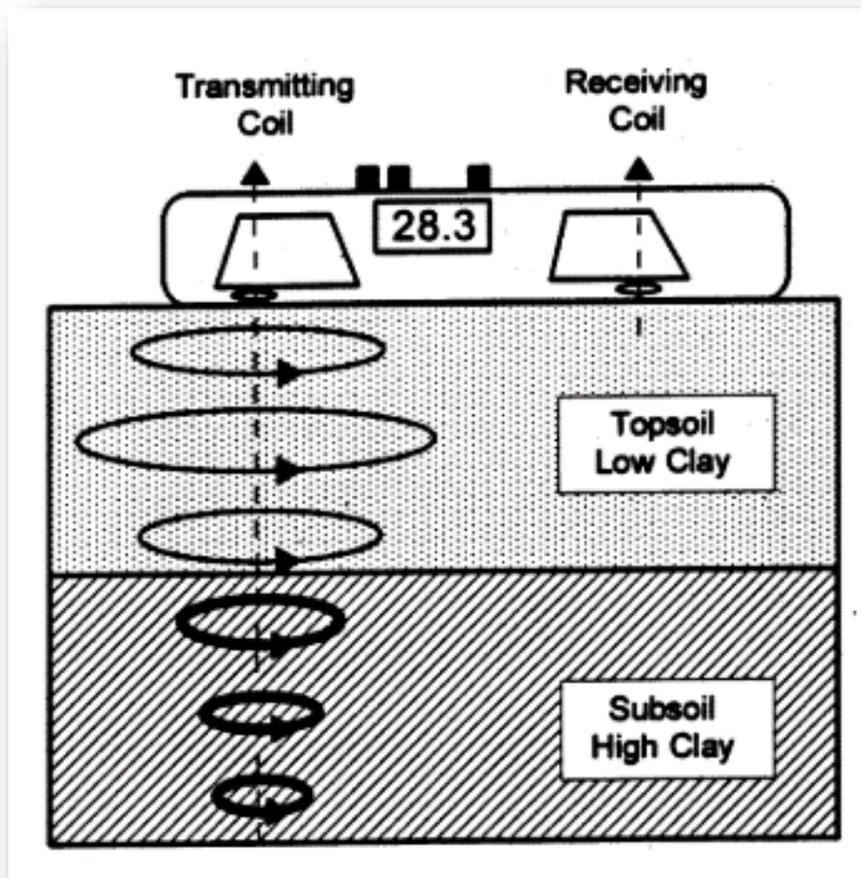


Figure 2.2 : Schéma d'un capteur de conductivité de sol (Geonics EM38) en mode dipôle vertical (tirée de Sudduth et al., 2001).

2.2.2 Méthode mécanique

Les méthodes mécaniques s'intéressent principalement à l'étude de l'intensité, de la position de la compaction et des propriétés intervenant dans le changement de la structure du sol. Ce paramètre physique a été proposé par Tollner et Verma (1984) et est défini comme étant la contrainte physique résultante de l'utilisation des équipements lourds de travail du sol ou de la compaction provoquée naturellement sur la croissance des racines des plantes. L'outil utilisé souvent pour évaluer le niveau de la compaction du sol est le pénétromètre (Figure 2.3) dont la caractéristique la plus importante est sa résistance exprimée en unité de masse par unité de surface. La résistance du pénétromètre est calculée grâce à une sonde conique capable d'estimer la distorsion d'une cellule de charge lorsqu'elle entre en contact avec le sol (Crozier, et al., 1999).

Bien que prouvé un outil efficace pour l'étude de la compaction du sol, le pénétromètre possède des limites liées à sa résistance qui ne peut pas être mesurée dans des parcelles drainées (Crozier et al., 1999). Selon Mulqueen et al. (1977), les valeurs des forces pouvant être estimées par le pénétromètre sont influencées par la teneur en eau et le développement de zones compactées autour du capteur, par conséquent, les valeurs données par le pénétromètre ne reflètent pas les vraies valeurs de force de pénétration.

Plusieurs tentatives ont été faites pour surmonter les limites de mesure de la compaction du sol telle que celle menée par Tollner et Verma (1984) qui ont développé un instrument qui estime les faibles niveaux de compaction du sol pour simuler le déplacement du sol par les racines. Loghavi et Khadem (2015) ont développé un prototype d'estimation de la résistance du sol composé de quatre pénétromètres horizontaux pour la caractérisation verticale de ce paramètre et la cartographie de la compaction du sol. Pour avoir une idée complète sur la compaction du sol et toutes les propriétés y attachées en temps réel, Lui et al. (1996) ont conçu un système composé d'un capteur diélectrique pour la mesure de la teneur en eau, un pistolet radar pour le calcul de la vitesse, un potentiomètre pour l'estimation de la profondeur et un récepteur GPS.

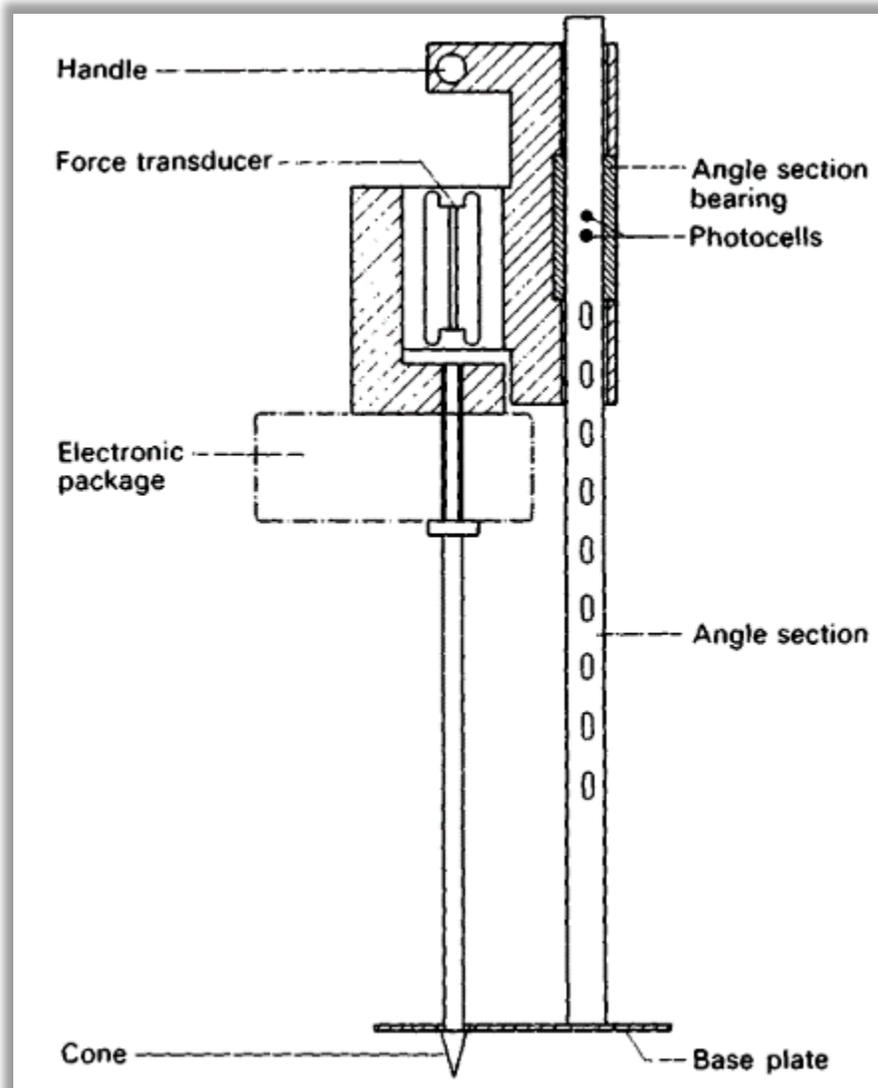


Figure 2.3 : Schéma d'un pénétromètre (tirée de Anderson et al., 1980).

2.2.3 Méthode électrochimique

La méthode électrochimique est souvent utilisée pour la caractérisation de la teneur du sol en nutriments. Elle a été développée grâce à la relation entre le potentiel électrique d'un conducteur du courant électrique et l'activité des ions présents dans une solution en contact avec le conducteur. Le plus souvent, les scientifiques utilisent des électrodes spécifiques à des ions bien déterminés pour mesurer leurs concentrations. Cette relation a été mise en évidence par Nernst et a été améliorée par Nikolskii pour inclure l'effet des ions présents dans la solution et auxquels l'électrode n'est pas spécifique (Sokalski et al., 2003; Birrell et Hummel, 2000).

Le paramètre utilisé pour la détermination de la teneur en nutriment est le potentiel électrique mesuré soit à l'aide des électrodes ou des transistors à effet de champ sélectifs aux ions. Dans le cas des électrodes, la somme des chutes de tension au niveau des jonctions solide-liquide, solide-solide et liquide-liquide est le potentiel électrique totale. Ce potentiel électrique est mesuré entre une électrode sélective aux ions et une électrode de référence dans une demi-cellule galvanique contenant une solution d'ions. Le potentiel mesuré le long de l'électrode sélective aux ions i par exemple renseigne sur son efficacité à évaluer l'activité de ces ions. Néanmoins, la présence d'autres ions dans la solution peut influencer les mesures de potentiel (Birrell et Hummel, 2000). Les électrodes sont le plus souvent intégrées dans des systèmes assurant la détermination en temps réel de la teneur du sol en nutriments comme les nitrates. Sibley (2008) a développé un système de cartographie des nitrates en temps réel. Grâce à la relation de proportionnalité entre le potentiel de l'électrode et le logarithme de la concentration de la solution de sol en ions nitrates, des cartes de teneurs en nitrate ont été construites. Adamchuk et al. (1999) ont utilisé une électrode de mesure de pH avec un outil d'échantillonnage de sol pour caractériser le pH de sol en temps réel.

Les transistors à effet de champ ont été développés en adhérant des membranes sélectives aux ions sur le matériau isolant des grilles du transistor. Les transistors opèrent dans des circuits pour la transformation rapide de l'impédance et le multiplexage de la sortie de chaque transistor (Tsukada et al., 1989). Comme les électrodes sélectives aux ions, la réponse des transistors à l'activité des ions dans la solution du sol est proportionnelle au logarithme de la concentration des ions étudiés (Birrell et Hummel, 2001). Ces outils sont fréquemment utilisés à cause de leur faible dimension et réponse rapide ainsi que pour leur fort rapport signal/bruit et leur capacité à déterminer la concentration en nitrates d'un faible volume d'échantillon (Birrell et Hummel, 2000). Ce système assure une mesure entre la solution contenant les ions étudiés et une solution de référence envoyée vers le capteur (Bergveld, 1991). Malgré l'efficacité des transistors à effet de champs sélectifs aux ions, leur dérivée après une longue durée de leur utilisation demeure une limite à considérer (Birrell et Hummel, 2001).

Les méthodes décrites précédemment permettent le partitionnement des parcelles en des zones de propriétés plus ou moins homogènes. Toutefois, elles sont incapables de remplacer les méthodes conventionnelles de quantification des propriétés de sol. Pour cette raison, l'apparition d'autres méthodes basées sur l'utilisation des capteurs comme la méthode optique s'avère une bonne alternative à celles décrites précédemment.

2.3 La spectroscopie du sol

2.3.1 Définition de la réflexion de la lumière

Pour bien comprendre le principe de la spectroscopie, la définition de la réflexion de la lumière s'avère une étape importante. Lorsqu'un groupe de photons se trouve en contact avec un milieu où l'indice de réfraction change, une partie des photons est réfléchi et une autre est réfractée dans le milieu. La réflexion (R) d'une lumière incidente à une surface plane est décrite par l'équation de Fresnel (Clark, 1993) :

$$R = \frac{(n-1)^2 + K^2}{(n+1)^2 + K^2} \quad (1)$$

Avec R : la réflexion de la lumière ;

n et K : sont respectivement la partie réelle et la partie imaginaire de l'indice de réfraction.

L'indice de réfraction est lié à la longueur d'onde de la lumière à travers l'indice d'absorption k qui est caractéristique du milieu traversé (Clark, 1993) :

$$k = \frac{4 \pi K}{\lambda} \quad (2)$$

2.3.2 Définition de la spectroscopie

« La signature spectrale des matières est définie par leur réflectance ou absorbance en fonction de la longueur d'onde dans le spectre électromagnétique » (Figure 2.4) et la spectroscopie est l'étude des signatures spectrales des objets (Stenberg et al., 2010). À une échelle moléculaire, l'énergie électromagnétique favorise la transition électronique de la molécule en provoquant la vibration des liaisons entre ses atomes. Ceci se manifeste par des figures caractéristiques qu'on appelle des figures fondamentales et autres secondaires dans la réponse de la matière à l'énergie reçue ou encore sa signature spectrale (Stenberg et al., 2010).

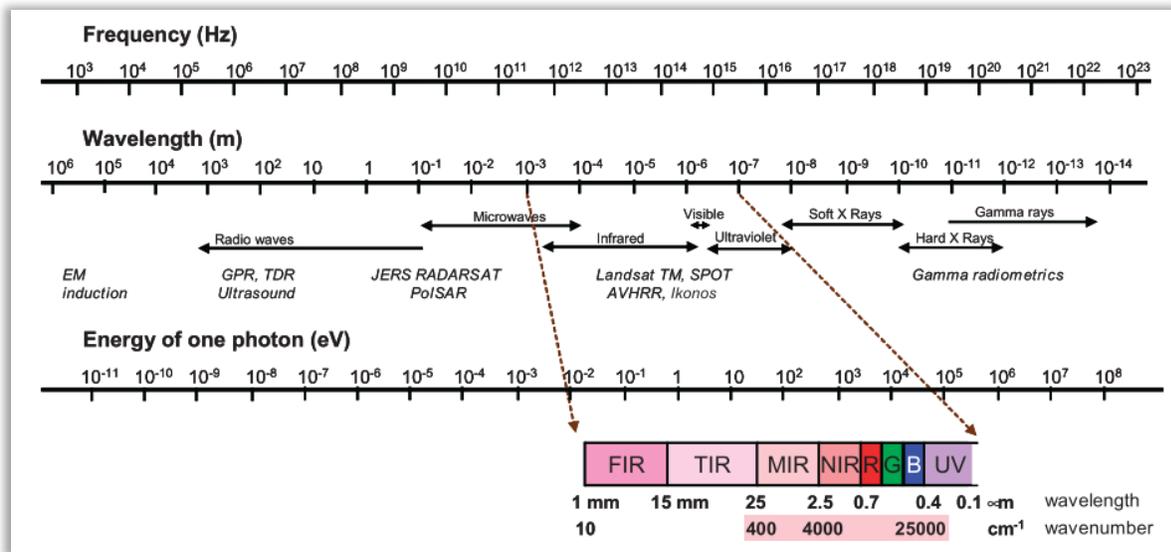


Figure 2.4 : Spectre électromagnétique avec les domaines du visible et proche infrarouges (Rossel et al., 2006) adapté de McBratney et al. (2006).

2.3.3 La spectroscopie du sol

Le début de l'utilisation de la spectroscopie proche infrarouge a débuté dans les années 1950 bien que la découverte du domaine de proche infrarouge dans la lumière par Herschel a été recensée dès 1800 (Blanco et Villarroya, 2002). Dans les années 1970, un groupe de chercheurs s'est servi de la spectroscopie infrarouge pour analyser des échantillons d'aliments. Dès lors, le progrès technologique des moyens électroniques ainsi que le développement des programmes efficaces capables de traiter l'information spectroscopique ont permis l'utilisation de cette méthode dans plusieurs domaines (Blanco et Villarroya, 2002).

Parmi ces domaines, la spectroscopie a été largement utilisée dans l'étude des propriétés de sol dans les dernières décennies (Genot et al., 2011; Knox et al., 2015; Shi et al., 2015; Brunet et al., 2007; Zornoza et al., 2008; Li et al., 2015; Pirie et al., 2006).

En effet, les liaisons des constituants chimiques du sol répondent à la lumière soit par flexion ou par étirement pour absorber des quantités spécifiques d'énergie et un spectre d'absorption ou de réflexion est déjà formé. La proportionnalité entre l'énergie et la longueur d'onde permet de distinguer des formes caractéristiques de la réponse des constituants du sol à la lumière dans des parties bien déterminées du spectre (Stenberg et al., 2010). L'absorption de la lumière se manifeste par des régions du spectre au niveau des longueurs d'onde bien précises avec un signal réduit de l'énergie réfléchi. Ces longueurs d'onde dépendent de la température, des

liaisons atomiques de l'élément chimique et aussi des groupes fonctionnels voisins. Ainsi, on peut identifier une panoplie de molécules dans le sol avec le même type de liaison (Stenberg et al., 2010).

2.3.4 Particularités spectrales des propriétés de sol

2.3.4.1 Teneur en eau

La liaison O-H de la molécule d'eau subit des vibrations lorsqu'exposée aux radiations de fréquences de la région de proche infrarouge. Ceci provoque une absorption de l'énergie dans cette partie du spectre (Hunt, 1977). L'eau présente dans le sol absorbe plus précisément aux alentours des bandes 1400 et 1900 nm (Figure 2.5), néanmoins, il existe d'autres bandes du domaine du visible-proche infrarouge où on trouve des faibles pics d'absorption de l'eau (Weidong et al., 2002).

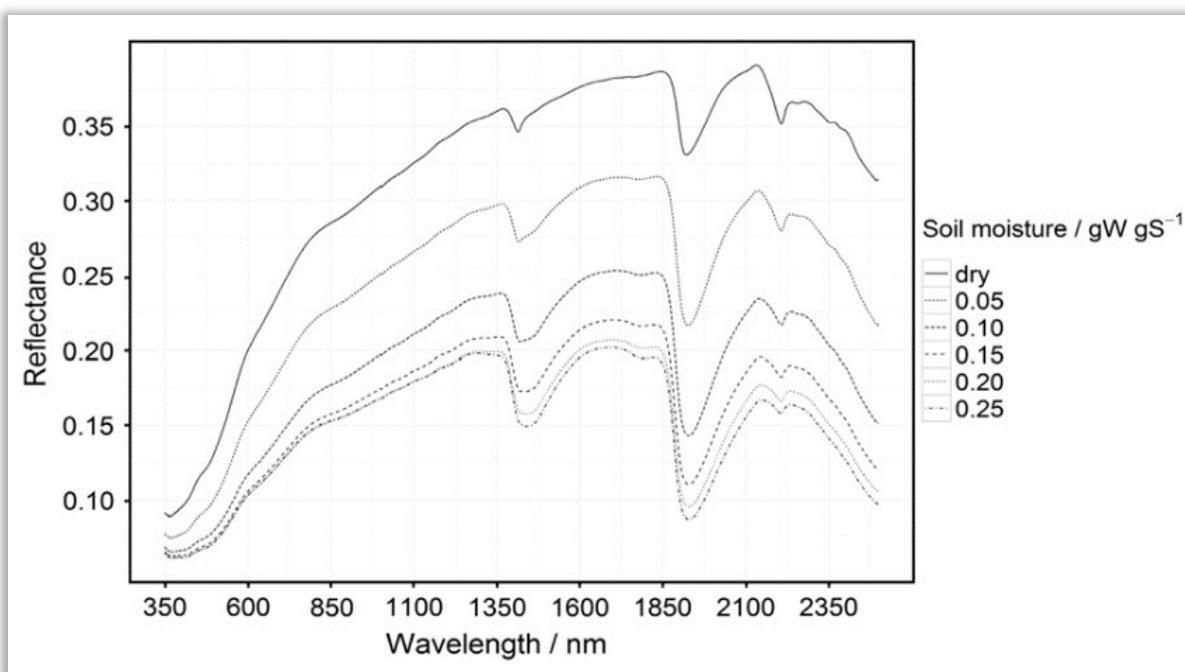


Figure 2.5 : Signatures spectrales des échantillons de sol montrant les pics d'absorption de l'eau à 1400 et 1900 nm (Nocita et al., 2013).

Bowers et Hanks (1965) affirment que les bandes proches de 1900 nm permettent une meilleure estimation de la teneur en eau du sol. Ceci résulte possiblement du fait que les bandes autour de 1400 nm renseignent sur l'absorption de l'ion hydroxyle qui peut être lié aux

minéraux d'argile et à l'eau incorporée dans la matrice d'argile (Clark, 1993 ; Clark et al., 1990). L'eau située dans les pores du sol agit en diminuant la valeur de l'albédo (Twomey et al., 1986) ce qui explique la diminution de l'allure générale du spectre autant que le sol est humide. Cette diminution est attribuée majoritairement au changement de l'indice de réfraction réel du milieu entourant les particules du sol de celui de l'air à celui de l'eau (Twomey et al., 1986). Malgré l'évidence de l'effet général de l'eau sur l'albédo, il n'existe aucune formule décrivant la corrélation entre l'humidité gravimétrique et volumétrique du sol et de l'albédo. Ceci parce que les valeurs de la porosité et de l'indice de réfraction des particules du sol sont sensibles aux types des sols. Par exemple, après séchage, l'augmentation de la réflectance chez des sols sablonneux est plus importante que celle observée chez l'argile de Kaolinite à cause de sa rétraction à l'état sec (Whalley et al., 1991)

2.3.4.2 Matière organique

Les groupements responsables de la détection de la matière organique sont la monohydrure d'azote (NH), le carbone-hydrogène (CH) et le monoxyde de carbone (CO). Les figures fondamentales d'absorption provoquées par la vibration de leurs liaisons ont lieu dans le moyen infrarouge alors que les vibrations secondaires et la combinaison entre elles ont lieu dans le domaine du visible et du proche infrarouge du spectre électromagnétique (Bokobza, 1998 ; Goddu et Delker, 1960). Le pic d'absorption du groupement CH a été observé au niveau de la longueur d'onde 3509 nm grâce à l'extraction et l'étude des acides humiques par Kang et Xing (2005). Ceci a confirmé les hypothèses de Clark (1993) qui a attribué le pic d'absorption à 3400 nm à l'étirement de la liaison C-H et ceux à 2300, 1700 et 1100 nm aux combinaisons des vibrations secondaires de ce groupement. Quoique les pics d'absorption de la matière organique dans les deux domaines visibles et proche infrarouges soient faibles, l'absorption globale de la matière organique est nette dans cette région (Baumgardner et al., 1985; Bowers et Hanks, 1965; Krishnan et al., 1979).

2.3.4.3 Eléments nutritifs

Il existe peu ou pas de figures d'absorption dans le domaine du visible-proche infrarouge qui permettent de caractériser la teneur du sol en nutriments comme l'azote (N), le phosphore (P), le potassium (K), le fer (Fe), le calcium (Ca), le sodium (Na) et le magnésium (Mg) (Maleki et al., 2006). Cependant, certains auteurs ont repéré de fortes corrélations entre les concentrations de ces éléments et des données spectrales. Parmi ces exceptions, on cite les

études effectuées par Ehsani et al. (1999) pour l'estimation de N (0.79-0.81) et par Ehsani et al. (1999) pour l'estimation de Fe (0.92), du cuivre Cu (0.87), du manganèse Mn (0.72), et du zinc Zn (0.72). Selon Stenberg et al., 2010, ces liens trouvés sont attribués à la combinaison des réponses d'autres éléments détectables spectralement et ceci varie d'une base de données à une autre. Maleki et al. (2006) proposent qu'il existe une corrélation entre le phosphore et le domaine du visible-proche infrarouge grâce à d'autres éléments présents dans le sol dont le phosphore est lié (Maleki et al., 2006).

2.3.4.4 Texture

Quant à la texture du sol, on s'intéressera principalement à la corrélation entre la teneur en argile et la réponse spectrale puisque cette fraction a été largement étudiée (Stenberg et al., 2010). Pour l'examiner, on se réfère à l'argile comme étant le minéral et non pas une particule du sol (diamètre inférieur à 2 mm), puisqu'il est possible de détecter la réponse des minéraux dans les spectres visibles et infrarouges (Stenberg et al., 2010). Lors d'une étude d'estimation de la teneur en argile, de la surface spécifique des paramètres reliés et de la capacité d'échange cationique (CAC) (Ben-dor et Banin, 1995), il a été trouvé que les bandes les plus importantes sont celles de l'absorption des groupements OH de l'eau, FeOH, Al et Mg constituant le réseau de cristaux de l'argile. Ceci a été confirmé après par Madejov et Komadel (2001). Pour cette raison, plusieurs chercheurs ont utilisé la capacité d'échange cationique et la surface spécifique des paramètres reliés pour l'estimation de la teneur en argile tels que Brown et al. (2006) et Chang et al. (2001). Cependant, ces observations ne peuvent pas être généralisées étant donné que Rossel et al. (2006) ont trouvé une corrélation de la CEC avec une réponse spectrale plus faible que celle de la teneur en argile dans le cas des sols australiens.

2.3.5 Avantages et inconvénients de la spectroscopie du sol

Depuis son apparition, la spectroscopie a été un outil mobilisé par un grand nombre de chercheurs pour l'étude des propriétés du sol. Selon Rossel et al. (2006), les raisons de l'utilisation intense de cette technique sont :

- La rapidité de l'obtention de résultat ;
- La possibilité de réutilisation des échantillons du sol déjà analysés puisqu'ils ne sont pas affectés par les analyses ;

- La préparation simple des échantillons (séchage à l'air, moulage et tamisage) ;
- La non-utilisation des substances dangereuses ;
- L'étude de plusieurs propriétés du sol en une seule acquisition ;
- Les analyses pouvant être effectuées sur-le-champ comme au laboratoire.

Néanmoins, l'utilisation de la spectroscopie est limitée par les petits changements du sol tels que la variation de sa structure, composition ou encore l'évolution chimique de ses constituants. Ceci a comme effet l'altération de la signature spectrale qui se manifeste par des décalages de la position des bandes d'absorption ou une modification de l'allure générale du spectre (Clark, 1993).

L'étude des propriétés du sol à travers la spécificité de la réponse spectrale des constituants du sol ne permet pas leur quantification. Ceci à cause du chevauchement des pics et des figures d'absorption (ou interférence de plusieurs constituants du sol). En outre, l'effet dispersif de quelques éléments présents dans le sol tel que le quartz rend l'interprétation des signatures spectrales complexes (Clark, 1993). Pour ces raisons, l'estimation des propriétés de sol est assurée en extrayant mathématiquement la partie du signal corrélée avec la propriété étudiée par des calibrations multivariées.

2.3.6 Application de la spectroscopie

2.3.6.1 Création et utilisation des données spectrales à une échelle globale

- **Construction des bibliothèques spectrales**

Plusieurs efforts ont été fournis pour collecter des spectres d'échantillons du sol de différentes propriétés et dans plusieurs régions du monde (Viscarra Rossel et al., 2016; Shepherd et Walsh, 2002; Ceddia, et al., 2017; Viscarra Rossel et al., 2008; Li et al., 2015). Il est nécessaire de respecter de nombreux aspects pour la construction d'une large base de données spectrales ou encore une bibliothèque spectrale. Un nombre d'échantillons de sol approprié pour bien représenter la variabilité dans les propriétés du sol de la région visée est un bon point de départ pour développer une bibliothèque spectrale. Il est également important de suivre des étapes bien précises pour l'échantillonnage, la préparation, le stockage et les protocoles d'analyse de sol. Dans une étude effectuée par Viscarra Rossel et al. (2008), la construction d'une bibliothèque

spectrale d'une région de l'Australie a été opérée en plusieurs étapes. Après l'échantillonnage du sol, les données spectrales ont été réduites en appliquant l'analyse en composante principale (ACP). Lors de la détection des valeurs atypiques, les spectres ont été acquis de nouveau pour la vérification de leur présence dans la librairie spectrale et ont été ensuite supprimés de la base de données. Il est important de souligner aussi que la construction d'une librairie spectrale se fait d'une façon continue. En effet, l'analyse de nouveaux échantillons n'appartenant à aucune population de spectres figurant dans la librairie spectrale, impose l'ajout de ces nouveaux échantillons à la librairie spectrale du pays ou de la région. Notamment, Brown (2007) a intégré de nouveaux échantillons locaux à la librairie spectrale de l'Uganda pour l'estimation de la teneur en argile et en carbone.

- **Estimation des propriétés du sol à partir des librairies spectrales**

La capacité des librairies spectrales à générer des modèles d'estimation a été évaluée dans plusieurs études. Terra et al. (2015) ont utilisé la régression multiple pour l'estimation complète des propriétés physiques (teneur en sable et argile), chimiques (concentration en Al^{3+} , en H^+ , la capacité en échange cationique, le pH, la concentration des oxydes des micronutriments et des macronutriments) et minéralogiques (activité des argiles, les teneurs en Fe_2O_3 et TiO_2) des échantillons du sol d'une librairie. Shepherd et Walsh (2002) se sont servis d'une librairie spectrale de l'est et de sud de l'Afrique pour l'estimation des propriétés des échantillons collectés et ont prouvé que la détection de valeurs atypiques améliore la précision des estimations d'une librairie spectrale.

Cependant, l'applicabilité des modèles établis à partir de ces larges bases de données sur d'autres échantillons n'est pas toujours garantie. Pour cela, d'autres chercheurs ont eu recours à la caractérisation d'échantillons de sols locaux en utilisant des librairies spectrales globales. Gogé et al. (2014) ont testé différentes stratégies pour la quantification des propriétés physicochimiques d'échantillons du sol à partir de la librairie spectrale de la France. Les modèles créés ont été déduits à partir de :

- La librairie spectrale seulement ;
- Des échantillons locaux seulement ;
- De la librairie spectrale combinée avec des sous-ensembles de la base de données locale.

Les résultats diffèrent d'une propriété à l'autre, s'il existe une corrélation entre une forte réponse spectrale et une propriété, des spectres d'échantillons locaux contribuent faiblement à la calibration du modèle de quantification des propriétés de ces échantillons. Pour cela, l'utilisation de moins d'échantillons locaux possible lors de l'estimation des propriétés sur une échelle locale à partir d'une base de données nationale est préférable. Aussi une grande taille ou un nombre d'échantillons important dans la librairie spectrale ne représente pas forcément un avantage pour la calibration d'un modèle à application locale. À travers l'utilisation de huit bibliothèques spectrales pour l'estimation de la teneur en carbone localement, Guerrero et al. (2016) ont proposé que l'adaptation des particularités spectrales des bibliothèques globales à celles des échantillons locaux abaisse le besoin à de larges bases de données. Contrairement, Brown et al. (2006) suggèrent que les larges bibliothèques spectrales (104 – 105 échantillons) sont exigées pour augmenter la reproductibilité des modèles. Gogé et al. (2012) ont fait une sélection raffinée pour identifier des échantillons dans la base de données globale similaire spectralement aux échantillons locaux. Pour chaque échantillon, un indice de similarité a permis d'isoler un groupe d'échantillons dans la librairie spectrale qui a servi à l'estimation des propriétés de cet échantillon. Dans le but d'appliquer un modèle construit à partir de la librairie spectrale de la Chine sur des échantillons du sol hautement humides, Ji et al. (2016), ont d'abord opté pour l'élimination de l'effet de la teneur en eau de ces échantillons.

Outre les caractéristiques de la librairie et sa correspondance avec des échantillons locaux, plusieurs chercheurs ont étudié l'effet de la méthode de calibration utilisée. Araújo et al. (2014) ont montré que l'utilisation d'une méthode d'apprentissage automatique ajoute une même précision à la calibration que celle ajoutée par la régression Partial Least Squares (PLS) associée à un partitionnement de la librairie spectrale. Araújo et al. (2014) ont prouvé que la combinaison de la régression PLS avec la localisation géographique des échantillons est plus performante que la régression PLS seule. De même Li et al. (2015) ont trouvé que la régression PLS non linéaire permet une meilleure quantification des concentrations en azote totale à partir d'une large librairie spectrale, comparée à une régression PLS linéaire.

2.3.6.2 Utilisation des données spectrales à une échelle locale

Comme pour les bibliothèques spectrales à nombre élevé d'échantillons, la spectroscopie a été largement étudiée à une échelle plus restreinte pour démontrer sa capacité à estimer les propriétés du sol. Que cela soit une propriété mesurable directement (par des protocoles définis au laboratoire) ou indirectement (par d'indices ou des fonctions de pédotransferts), le potentiel

de la spectroscopie à caractériser la qualité du sol a été testé par plusieurs chercheurs. Waruru et al. (2016) ont étudié la structure du sol en combinant la spectroscopie proche infrarouge et des propriétés liées à la formation d'agrégats du sol. Les résultats trouvés révèlent que la spectroscopie proche infrarouge permet l'estimation des fractions de macroagrégats et possède un potentiel moyen d'estimation de fraction d'argile colloïdale. Quant aux indices de fertilité du sol, Sadegh, et al. (2015) ont alloué la spectroscopie pour comparer la qualité de deux types de sols. De même pour les paramètres biologiques, Ludwig et al. (2017) ont évalué la possibilité d'estimer l'activité de neuf enzymes, le pH, la teneur en carbone organique, en azote total et la texture du sol. Bien qu'ils aient prouvé son potentiel à estimer la teneur en carbone organique et en azote total, la spectroscopie visible-proche infrarouge ne permet pas une estimation directe de l'activité enzymatique. Lors d'une étude effectuée par Comino et al. (2018), il a été démontré, qu'à partir de la spectroscopie, il est bien possible de caractériser qualitativement le statut biologique du sol en suivant le taux de recyclage des nutriments qui renseigne indirectement sur l'activité de quelques enzymes (la phosphatase d'acide, la phosphatase alcaline β -glucosidase, arylsulfatase et la déshydrogénase).

Malgré sa capacité à quantifier et qualifier les propriétés du sol, la précision des modèles développés à partir des données spectroscopiques de sol varie d'une base de données à une autre. Pour cette raison, l'influence de chaque composante de la spectroscopie du sol (le sol, le matériel d'acquisition et les variables d'acquisition et les traitements des spectres) a été mise en relief et questionnée par plusieurs auteurs.

- **Les propriétés et le traitement de sol**

En raison de sa forte absorption de l'énergie électromagnétique dans le moyen infrarouge, l'eau affecte en grande partie la calibration des autres propriétés du sol. Quoique les spectres proche infrarouges soient aussi influencés par ce composé, le domaine du moyen infrarouge est considéré non efficace dans le cas d'échantillons de sol très humides (Reeves, 2010). Nocita et al. (2012) se sont servis d'un indice normalisé d'humidité du sol pour isoler différentes classes spectrales et quantifier après l'effet de la teneur en eau sur la précision de l'estimation de la matière organique. Les résultats ont montré que la difficulté d'estimer la matière organique augmente avec l'augmentation de l'humidité du sol. Pour caractériser la réponse spectrale de la matière organique, Yeasmin et al. (2017) ont détruit la matière minérale du sol par l'acide hydrofluorique afin d'éliminer son impact sur le signal et quantifier précisément la matière organique dans le sol. La spectroscopie a aussi été utilisée par Rodionov et al. (2016) en vue de déterminer la limite de différenciation entre la biomasse végétale de la matière organique

dans le sol. Pour cela, la surestimation de la matière organique provoquée par la présence de biomasse végétale (paille et feuilles) a été quantifiée à travers des indices définis par d'autres chercheurs. Ceci sert à corriger la surestimation de la teneur en matière organique dans des conditions d'acquisition du sol sur le terrain. Les prétraitements du sol ont également un effet remarquable sur l'estimation de ses propriétés. Une comparaison faite par Gandariasbeitia et al. (2017) entre la calibration de modèles pour trois groupes d'échantillons : non prétraités, partiellement traités (frais et moulus) et prétraités (séchés à l'air, moulu et tamisé à 2 mm) a conduit à la conclusion que les prétraitements d'échantillons augmentent la précision des modèles d'estimation des attributs du sol.

- **Matériels et variables d'acquisition**

L'impact des spectromètres et des paramètres d'acquisition sur la qualité d'estimation a été l'objet de plusieurs études. Ceci a été mis en évidence en utilisant les mêmes spectromètres en suivant les mêmes protocoles expérimentaux ou en variant les configurations et les variables spectrales des spectromètres. Une comparaison faite par Pimstein (2011) entre les résultats d'acquisition de spectres par trois spectromètres identiques avec les mêmes manipulations des mêmes échantillons a bien montré l'influence des composantes électriques de chaque spectromètre sur les spectres. En effet, pour un même spectromètre, un même échantillon et entre une mesure et une autre, le signal subit des altérations qui se manifestent par un changement de la position et de la forme du pic d'absorption ou encore par l'intensité du signal. Pour pallier cette modification, Pimstein (2011) propose qu'elle soit quantifiée à l'aide d'un étalon interne pour être retranchée du signal après chaque mesure. En calibrant un modèle d'estimation des propriétés du sol, Ji et al. (2016) ont associé l'erreur de reproductibilité du modèle aux bruits optiques et électriques du spectromètre utilisé ainsi qu'à la microvariabilité du sol. La variation de la résolution spectrale et du ratio signal/bruit a été testée par Knadel et al. (2013) en utilisant trois spectromètres avec des configurations différentes. La qualité du signal a été faiblement affectée par cette variation surtout pour des figures d'absorption larges et prononcées et la résolution spectrale fine n'a pas amélioré considérablement la capacité des spectres à estimer la teneur en argile et en carbone organique. Pour cela, ils ont conclu qu'il est inutile d'utiliser des spectromètres avec des hautes résolutions spectrales tant que le protocole d'acquisition et de manipulation soit respecté. Cette constatation a été confirmée par Adeline et al. (2017) qui a évalué l'effet de la variation de la résolution spectrale sur la précision de prédiction de propriétés de sol et a trouvé que la précision du modèle calibré est stable pour une résolution spectrale entre 3 et 60 nm. Cependant, cette indépendance à la configuration

spectrale ne peut pas être généralisée et est liée à la réponse spectrale de la propriété du sol étudiée ainsi qu'à sa corrélation avec les autres propriétés. En ce qui concerne le domaine spectroscopique, une comparaison par Pirie et al. (2006) entre la capacité d'estimation de l'ultraviolet, du visible-proche infrarouge et du moyen infrarouge a montré que les propriétés du sol ont les réponses spectrales les plus fortes dans le moyen infrarouge. L'analyse qualitative des spectres proche infrarouge et moyen infrarouge par Rossel et al. (2006) fortifie cette constatation. Lors de la même étude, les modèles d'estimation de la texture du sol calibrés à partir des deux domaines proche infrarouge et moyen infrarouge ont améliorés la précision.

- **Traitements des spectres**

La précision de l'estimation et la généralisation d'un modèle de calibration dépendent des méthodes de traitement des spectres. Du partitionnement de la base de données en groupe d'échantillons (un pour la calibration et un autre pour la validation du modèle) jusqu'au choix de l'algorithme de modélisation des propriétés du sol, plusieurs options ont été évaluées pour la sélection de l'option la plus adéquate. La stratégie de sélection de données pour la calibration et la validation du modèle est presque la même (2/3 de la base de données pour la calibration et 1/3 pour la validation). Néanmoins, les échantillons de validation doivent avoir des caractéristiques différentes de celles de calibration pour augmenter l'applicabilité du modèle sur d'autres bases de données (Bellon-Maurel et McBratney, 2011). En plus, plusieurs algorithmes de modélisation des propriétés du sol ont été testés pour sélectionner celui ayant la meilleure performance. Bien que la PLS soit largement utilisée pour la modélisation des propriétés de sol (Nocita et al., 2013; Genot et al., 2011; Gras et al., 2014; Knox et al., 2015; Rodionov et al., 2016; Knadel et al., 2013), autres méthodes ont été prouvées avoir des résultats aussi précis comme les machines à vecteurs de support ou SVM, les modèles de régression cubiques (Morellos et al., 2016) et la régression PLS combinée avec la technique CARS (Vohland et al., 2014). Dans le Tableau 2.1, quelques études de la qualité du sol ayant utilisé des algorithmes de calibration connus ou en combinant entre eux sont citées.

Tableau 2.1 : Sources des méthodes les plus utilisées pour la prédiction des propriétés de sol

Algorithmes utilisés	Références
Réseaux de neurones artificiels	Daniel et al. (2003) ; Fidencio et al. (2002)
<i>Multivariate adaptive regression splines (MARS)</i>	Shepherd and Walsh (2002)
<i>Boosted regression trees (Brown)</i>	Brown et al. (2006)
PLSR-NN (PLSR combiné avec les réseaux de neurones artificiels)	Mouazen et al. (2010)

3 MATÉRIELS ET MÉTHODES

3.1 Présentation de la zone d'étude

Les échantillons du sol étudiés ont été prélevés par une équipe d'agriculture Canada de quatre champs : 2 situés à Sainte-Catherine-de-la-Jacques-Cartier et deux à L'Île d'Orléans (Figure 3.1). L'échantillonnage a été effectué avec une tarière (Figure 3.2) avec une grille d'échantillonnage triangulaire de 30m*30m.



Figure 3.1 : Localisation des champs expérimentaux (à droite : champs 1 et 2 situés à Ste-Catherine-de-la-Jacques-Cartier, à gauche champs 3 et 4 situés à l'Île d'Orléans).



Figure 3.2 : Prélèvement d'échantillons de sol par une tarière

Selon la classification canadienne des sols, ces échantillons ont été catégorisés en quatre séries du sol (Pont-Rouge, Morin, Orléans, St-Nicolas) appartenant à deux grandes classes (podzol humo-ferrique et brunisol dystrique). Des gammes importantes de textures et de types de sols des quatre champs (détaillées dans le Tableau 3.1) sont couvertes pour garantir une certaine variabilité de la base de données spectrales.

Tableau 3.1 : Classification et textures des échantillons de sol

Nom du site	Sainte-Catherine-de-la-Jacques-Cartier		Ile d'Orléans	
	1	2	3	4
Champ	1	2	3	4
Type du sol	Sol sableux	Sol sableux	Sol loameux	Sol graveleux
Texture du sol	Sable loameux grossier	Sable loameux grossier	Loam sableux grossier	Loam sableux grossier
Série de sol	Pont-Rouge	Morin	Orléans	St-Nicolas
Classification taxonomique	Podzol humo-ferrique orthique	Podzol humo-ferrique orthique	Brunisol dystrique éluvie	Podzol humo-ferrique orthique

3.2 Protocole expérimental

3.2.1 Préparation des échantillons de sol

Avant d'être analysés, les échantillons du sol ont été séchés à l'air et tamisés à deux millimètres pour leur homogénéisation. Puisque les volumes des échantillons disponibles ne suffisent pas pour remplir les boîtes de pétri, des échantillons composites ont été constitués par le mélange de deux échantillons unitaires. La sélection a été faite en se basant sur trois critères :

- La proximité géographique des échantillons ;
- La similarité des propriétés ;
- La représentativité de toutes les zones d'aménagement de chaque champ (ces zones ont été délimitées par l'équipe d'agriculture Canada en se basant sur les teneurs en sable des échantillons (Annexe 1)).

Au final 68 échantillons composites ont été constitués. Chaque deux échantillons ont été mélangés pour former l'échantillon de sol composite qui a été réparti en trois boîtes de pétri en plastique de dimension 60*60*15 mm et dont les poids ont été mesurés à l'avance (Figure 3.3).

La décomposition de chaque échantillon en 3 réplicas sert à tenir compte des erreurs causées par l'opérateur lors de la variation des teneurs en eau des échantillons. Les surfaces des échantillons ont été aplaties avec une spatule métallique tout en évitant de serrer les particules du sol. Cette étape diminue les effets multiplicatif et additif associés à la rugosité de la surface de contact entre échantillon et sonde de réflectance.



Figure 3.3 : Répartition de chaque échantillon composite en 3 réplicas.

Puis, les boîtes ont été couvertes par le géotextile (une membrane relativement mince permettant la rétention de toutes les particules de sol et ne laisser passer que l'eau). Les boîtes couvertes ont été ensuite placées à l'envers dans un bac rempli d'eau distillée et ont été maintenues dans cet état pendant une heure. Cette opération a permis la saturation du sol par capillarité sans toucher les surfaces des boîtes (là où les mesures de réflectance seront effectuées). Les échantillons ont été ensuite mis sur une grille métallique pendant 48 heures pour que l'eau se draine et que le sol atteigne la teneur en eau à la capacité au champ. La durée de drainage de l'eau gravitaire est comprise entre 24 et 48 heures (Susha Leskshmi et al., 2014) et le choix de 48 heures a été fait après des essais en suivant les étapes décrites ci-dessus, car les sols avec des niveaux d'humidité importants (saturés d'eau) n'étaient pas favorables à la mesure de la réflectance.

3.2.2 Mesure de la réflectance

La réflectance des échantillons du sol a été mesurée grâce à un système d'appareils disponibles commercialement auprès de la compagnie Ocean Optics (Ocean Optics, 2019). Ce système est composé de :

- Une lampe halogène HL-2000 (Figure 3.4.b) qui représente la source de lumière ;
- Une sonde de réflexion Premium (Figure 3.4.e) connectée d'un bout à 8 fibres optiques qui acheminent les photons de lumière de la lampe halogène vers elle, et de l'autre bout à une seule fibre optique dont le rôle est le transfert de la quantité d'énergie réfléchie par l'échantillon du sol (captée par la sonde) vers le spectromètre ;
- Un support sombre en contact avec la surface de l'échantillon du sol (Figure 3.4.a) et sur lequel la sonde a été fixée pour empêcher la lumière parasite de l'atteindre. Ce support sert aussi à garder une distance constante (1.7 cm) entre la sonde et la surface de sol pendant les acquisitions des spectres ;
- Deux spectromètres permettant de mesurer la réflectance sur deux plages spectrales différentes. Un spectromètre opère dans le domaine du visible et une partie du proche infrarouge (340 – 1038 nm) avec une résolution spectrale de 0.4 nm (Figure 3.4.c). L'autre spectromètre opère dans le proche infrarouge (900 - 1700 nm) avec une résolution spectrale de 3 nm (Figure 3.4.d) ;
- Un logiciel d'acquisition (OceanOptics, 2013), de lecture et de stockage de spectres qui permet d'ajuster la vitesse, le nombre d'acquisitions et d'autres paramètres permettant l'optimisation de la qualité du signal et la réduction du bruit ;
- Une référence (Figure 3.4.f) caractérisée par une réflectance totale (réflectance égale à 100%) pour la calibration des spectromètres. Pour ce faire, un premier spectre de cette référence est acquis dans le but de définir le seuil maximal de réflectance lorsque la source de lumière est allumée et un deuxième spectre est acquis lorsque cette dernière est éteinte pour définir le seuil minimal de réflectance (0%).

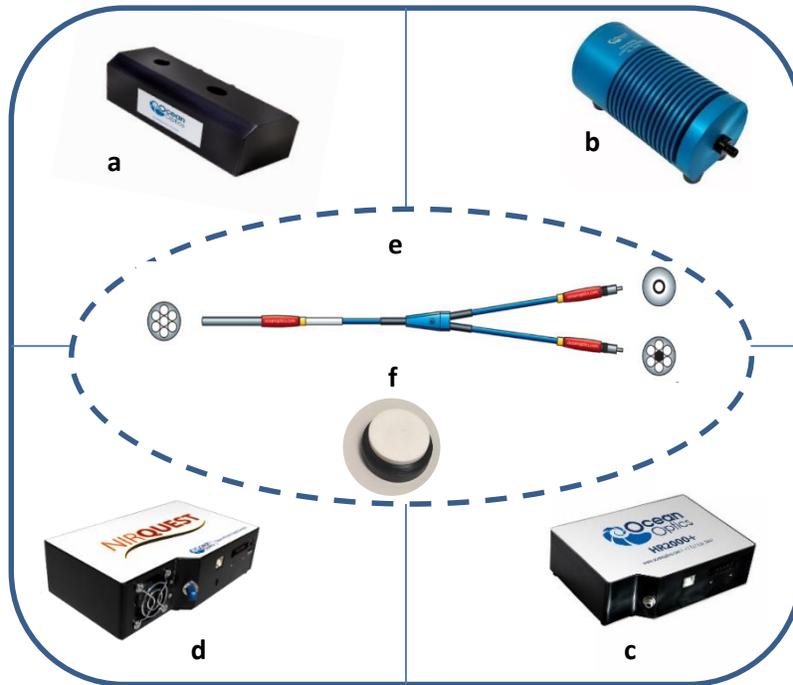


Figure 3.4 : Système d’Ocean Optics de mesure de la réflectance des échantillons de sol composé d’un support de la sonde de lumière (a), d’une lampe halogène (b), de deux spectromètres (c et d), d’une sonde de réflectance (e) et d’une référence de calibration des spectromètres (f).

Les paramètres d’acquisition ont été fixés de sorte que le logiciel calcule la moyenne de deux acquisitions internes afin de réduire le ratio signal/bruit. Pour chacun des trois réplicats, on a pris quatre mesures dans quatre points différents (à peu près un point chaque 45° du cercle de la surface de sol). Après 20 mesures de réflectance, le logiciel a été calibré du nouveau avec la référence pour enlever la dérivée des composantes électrique.

3.2.3 Mesure de la teneur en eau



Figure 3.5 : Mesure du poids des échantillons du sol.

Le poids de chaque réplicat a été mesuré par une balance électronique (Figure 3.5). Les échantillons ont été chauffés à l'étuve à 43 °C pour garder toujours une teneur en eau plus ou moins homogène de toutes les particules du sol et ont passé par les mêmes étapes de mesure de poids et de réflectance chaque trois heures jusqu'à ce que 24 heures passent après les premières mesures. Ceci a été fait pour amener les échantillons du sol à l'état sec, bien que selon la littérature (Susha Lekshmi et al., 2014), cet état ne soit atteint qu'après chauffage du sol à l'étuve à 105 °C pendant 24 heures et non pas à 43 °C. Pour cela, des tests de vérification ont été effectués en gardant le sol plus que 24 h à l'étuve à 43 °C. Dans ces conditions, aucune variation de poids de sol n'a été observée après 24h. Le calcul de la teneur en eau pour chaque mesure de poids et pour chaque réplicat, a été fait en utilisant l'équation suivante :

$$\theta = \frac{\text{poids Humide} - \text{poids sec}}{\text{poids sec}} * 100 \quad (3)$$

Avec :

θ : la teneur en eau volumétrique en pourcentage.

Poids humide : les poids des échantillons dans tous les niveaux d'humidité autre que l'état sec mesuré en gramme.

Poids sec : les poids des échantillons après 24 heures de chauffage dans l'étuve à 43 °C mesuré en gramme.

La Figure 3.6 résume les différentes manipulations des échantillons du sol et décrites dans les sections 3.2.1, 3.2.2, 3.2.3.

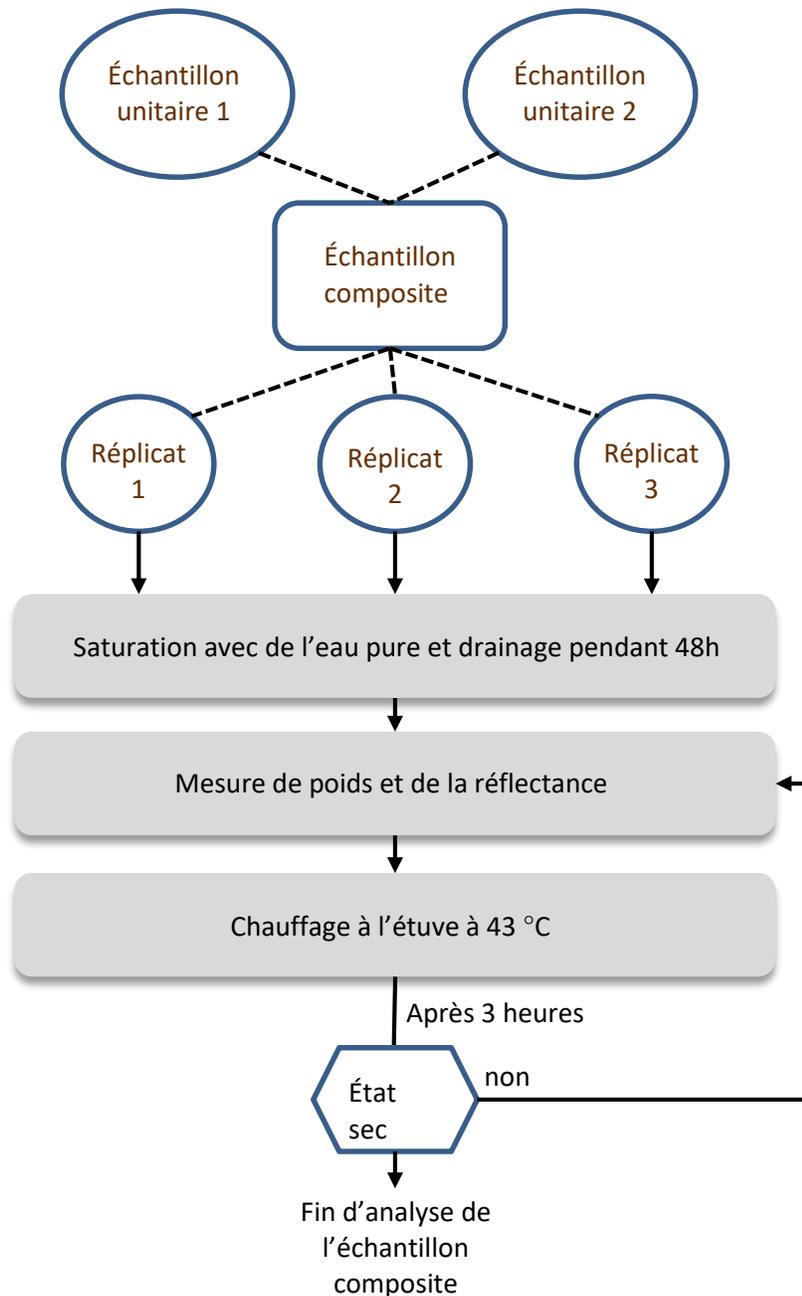


Figure 3.6 : Organigramme de différents traitements des échantillons du sol.

3.2.4 Calcul des moyennes

On a supposé que les échantillons unitaires étaient homogènes et possédaient des volumes égaux. Ces deux hypothèses ont été émises pour assigner des teneurs en argile, sable et matière organique moyennes à chaque échantillon composite à partir de celles des échantillons unitaires. Dans cette étude, on s'est limité à la quantification de deux paramètres de la texture

du sol seulement (argile et sable) puisque la détermination de pourcentage de limon peut être effectuée en soustrayant la somme des pourcentages d'argile et de sable de 100.

Quant aux mesures de réflectance, la moyenne des quatre spectres acquis lors de chaque mesure de teneur en eau pour chaque réplikat, a été calculée. De plus, les spectres et les valeurs d'humidité du sol de chaque trois réplikats de même échantillon composite ont été moyennés. Ceci a été effectué dans le but de diminuer la redondance de l'information tout en tenant compte des erreurs qui peuvent se produire entre une manipulation d'un réplikat et une autre.

3.3 Traitement des spectres

La Figure 3.7 résume les différents traitements effectués sur les spectres. La base de données a subi quatre types de traitements :

- La préparation effectuée sur des spectres bruts ;
- L'exploration effectuée sur des spectres moyens avec les longueurs d'onde réduites ;
- Le prétraitement effectué sur des spectres moyens avec les longueurs d'onde réduites ;
- L'estimation de l'humidité ainsi que les variables de la texture du sol et son contenu en matière organique effectuée sur des spectres prétraités. Il est à souligner que lors de l'étape de quantification des propriétés de sol à partir des spectres, la base de données a été partagée en deux : le 3/4 des échantillons pour la calibration des modèles d'estimation et 1/4 pour leur validation. Ceci a été fait en classant les spectres en ordre croissant selon les valeurs d'humidité du sol et en retenant pour chaque quatre échantillons le quatrième pour la validation.

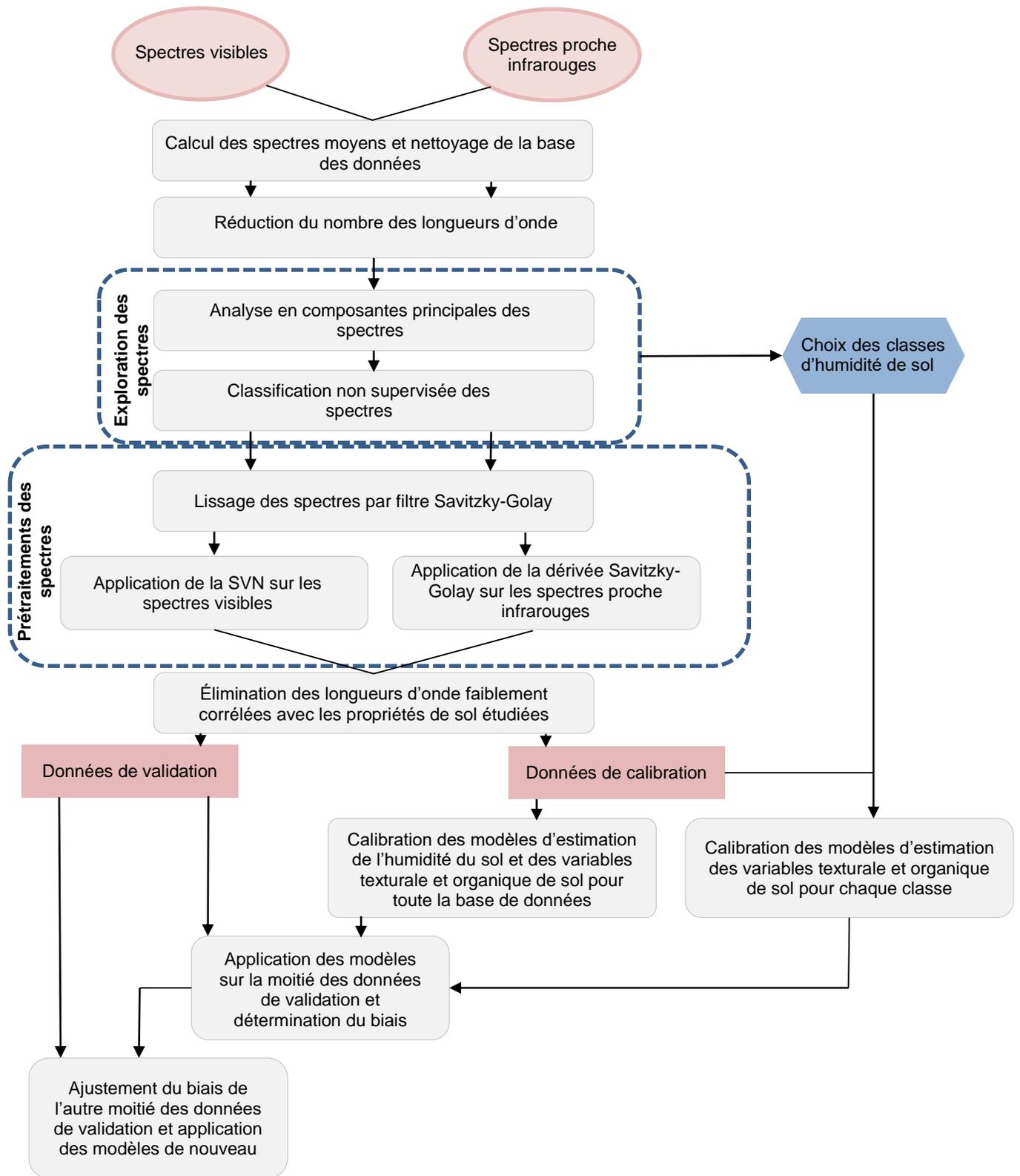


Figure 3.7 : Organigramme des différents traitements des spectres.

3.3.1 Nettoyage de la base des données spectrales

Puisque les spectres visibles et proche infrarouges ont été acquis séparément, les observations avec un spectre visible ou proche infrarouge manquant ont été supprimées de la base de données laissant un nombre total d'observations de 1025 avec un nombre total de longueurs d'onde de 1974. Ensuite, les échantillons dont le nombre de réplicats était inférieur à 3 ont été supprimés lors de l'étape de calcul des moyennes de leurs spectres. De plus, le nombre de longueurs d'onde a été réduit à un quart, à un sixième, à un huitième et à un dixième pour chaque domaine afin d'effectuer une amélioration préliminaire de la qualité du signal. Si on suppose que le facteur de réduction est n , cette opération consiste à assigner la moyenne des réflectances j jusqu'à $j+(n-1)$ à la longueur d'onde j . Le choix du facteur de réduction appliqué sur les longueurs d'onde visibles et celles proche infrarouges s'est appuyé sur l'allure des spectres (la disparition des rugosités) sous la condition que la résolution spectrale reste supérieure à 10 nm (pour rester toujours dans le domaine de l'hyperspectrale). Finalement, La base de données finale utilisée dans cette étude avait une taille de 258 longueurs d'onde et 288 observations.

3.3.2 Exploration des données

Cette étape a permis l'identification des structures menant à l'élaboration des hypothèses. Ces hypothèses aident par la suite à guider les étapes d'extraction de l'information des paramètres d'échantillons du sol à partir de leurs propriétés spectrales. L'application des méthodes d'exploration des données a été opérée sur le logiciel The Unscrambler X 10.5.1 sur les spectres ayant subi la réduction du nombre des longueurs d'onde ainsi que la réduction du bruit (la méthode sera abordée en détail dans la section de réduction du bruit) seulement afin d'investiguer la possibilité de mettre en évidence des structures spectrales pouvant renseigner sur les propriétés de sol, en particulier la teneur en eau.

3.3.2.1 Analyse en composante principale des spectres

Dans un premier lieu, une analyse en composante principale (ACP) a été appliquée sur les spectres. L'idée de base de l'ACP est la réduction de la dimensionnalité des données dans lesquelles il y a un nombre important de variables interreliées tout en gardant le maximum de variance présente dans les nouvelles variables (composantes principales). Cette réduction est faite en projetant les variables sur un nouvel espace de telle façon que ces dernières soient non

corrélées et dont les premiers retiennent la plus grande quantité d'information présente dans les variables originales (Jolliffe, 2002) (Figure 3.8).

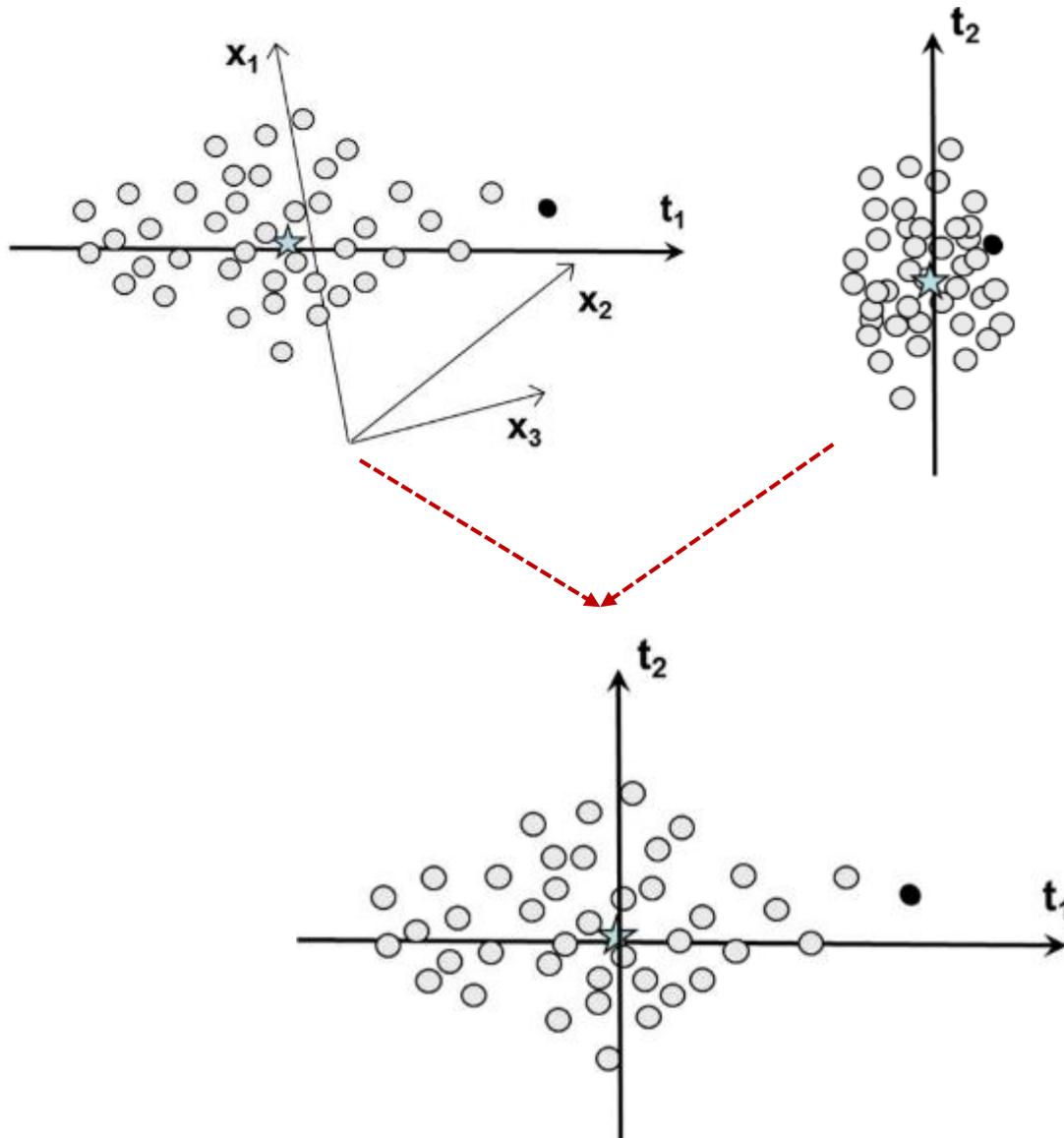


Figure 3.8 : Principe de l'ACP montrant le calcul des composantes principales t_1 et t_2 et la projection des variables dans l'espace formé par ces deux composantes (adaptée de Sébastien, 2016).

Cette méthode sert donc à concentrer l'information stockée dans les larges bases de données en la représentant dans un espace réduit. Ceci permet aussi la mise en relief des ressemblances des échantillons donc facilite l'identification de groupes d'individus ainsi que la détermination des liens entre les variables (Sébastien, 2016).

3.3.2.2 Classification non supervisée des spectres

Ce type de classification a été appliqué pour séparer les spectres en des groupes ayant des propriétés spectrales plus ou moins similaires. Pour ce faire, on a utilisé un algorithme de partitionnement de données non hiérarchique appelé l'algorithme des K-means ou des centres mobiles. Cette méthode part de la supposition qu'on peut agréger la base de données en un nombre K de classes. Les principes de calcul de cet algorithme peuvent être résumés en quatre étapes élémentaires (Macqueen, 1967) (Figure 3.9) :

- Choisir arbitrairement K individus qui seront les centres (ou les noyaux) des classes ;
- Assigner chaque individu de la base de données à une des K classes dont le noyau est le plus proche ;
- Recalculer les nouveaux centres de gravité de chaque classe ;
- Vérifier si les centres et les classes changent ou le nombre d'itérations maximale est atteint. Si aucune de ces deux conditions n'est vérifiées, le calcul se poursuit en recommençant par l'étape 2.

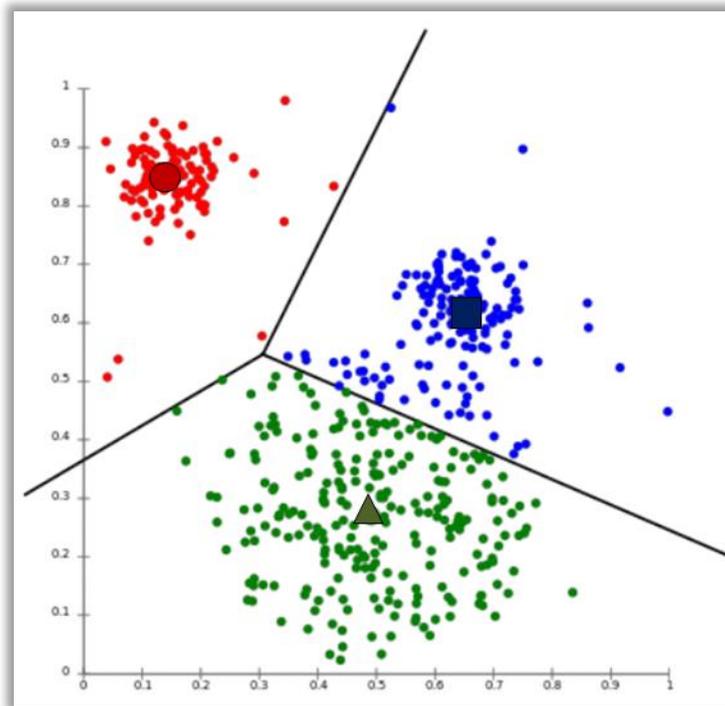


Figure 3.9 : Principe de l'algorithme K-means montrant le partitionnement de nuage de points en classes (adaptée de <https://aws.amazon.com/fr/blogs/machine-learning/k-means-clustering-with-amazon-sagemaker/>).

Le nombre de classes initiales a été fixé après l'inspection visuelle des spectres projetés sur l'espace des deux premières composantes principales. Le nombre maximal d'itérations a été arbitrairement fixé à 50. Après l'application d'une deuxième ACP sur les spectres classés par K-means, la visualisation du nuage de points en couleurs différentes selon le niveau d'humidité du sol a permis de décider le nombre de classes d'humidité du sol ainsi que d'identifier leurs bornes.

3.3.3 Prétraitements des spectres

Plusieurs facteurs affectent les formes et les magnitudes des spectres. Ces facteurs peuvent être induits soit par les instruments de mesure de réflectance, les particules de sol (formes et tailles) ou par l'opérateur. Si on suppose que Y_i^* est le spectre acquis d'un échantillon i , a_i est la valeur de l'effet additif, m_i est la valeur de l'effet multiplicatif et Y_i est la signature spectrale de l'échantillon i (Helland et al., 1995), alors :

$$Y_i^* = a_i + m_i * Y_i \quad (4)$$

À ceci s'ajoutent des erreurs attribuées le plus souvent aux effets des composantes du spectromètre qui se traduisent par des fluctuations du signal appelées bruits. L'objectif des prétraitements est d'éliminer ces termes pour extraire l'information spectrale qui reflète les propriétés de l'échantillon.

3.3.3.1 Réduction du bruit

Une première inspection visuelle des spectres a pu montrer les parties largement affectées par des erreurs de mesures. Pour cela, les spectres du visible ainsi que ceux du proche infrarouge ont subi la suppression des deux limites (les spectres gardés sont ceux qui sont supérieurs à 431.72 nm et inférieurs à 898.75 nm pour le domaine du visible et supérieurs à 900.92 nm et inférieurs à 1716,85 nm pour le domaine du proche infrarouge). En outre, les prétraitements ont été effectués sur chaque domaine séparément vu la différence entre eux en matière de résolution spectrale et même de qualité de signal (les spectres visibles sont plus bruités que ceux du proche infrarouge).

Pour ce qui suit, la réduction de la plus grande quantité possible du bruit (ou le lissage des spectres) a été effectuée en utilisant le filtre Savitzky-Golay. La définition de la fenêtre mobile est un bon point de départ pour comprendre cette méthode de lissage de données. Pour un graphe présentant des réflectances Y en axe d'ordonnées et des longueurs d'onde j dans l'axe des abscisses, la fenêtre de convolution (mobile) est un bloc de nombres de convolution (C) qui sont multipliés par les ordonnées de spectre originales et dont leur somme est divisée par la taille (N) de la fenêtre. Ce nombre, qui est le centre de la fenêtre, remplace la valeur de la réflectance du spectre original et l'opération est répétée jusqu'à la fin du spectre. L'équation mathématique décrivant ce processus est la suivante (Savitzky et Golay, 1964) :

$$Y_j^* = \frac{\sum_{i=-m}^{i=m} Y_{i+j} C_i}{N} \quad (5)$$

m est l'indice dans la fenêtre de convolution et le signe $*$ désigne la valeur de la réflectance lissée. Le filtre Savitzky-Golay ou des moindres carrées consiste à supposer que chaque partie du spectre est une fonction polynomiale (s'exprimant sous la forme de $y = a_0 + \sum_{i=1}^{i=n} a_i * x^i$, n est le degré de polynôme fixé par l'opérateur) et remplacer la valeur centrale dans la fenêtre mobile (la réflectance Y_j) par une valeur estimée à partir des points voisinant le point central. Cette valeur de Y est calculée après avoir déterminé les coefficients a_i de polynôme qui s'approche le plus possible de la partie du spectre à lisser en remplaçant les abscisses x_i des points de part et d'autre du centre de la fenêtre. Cette méthode a l'avantage de réduction du bruit en conservant le plus possible l'allure du spectre (Savitzky et Golay, 1964).

3.3.3.2 Réduction des effets additifs et multiplicatifs

L'opération de la réduction d'effets additifs et multiplicatifs des spectres est délicate, car une partie de l'information stockée dans le signal peut être perdue si on applique les mauvais algorithmes de prétraitements. Pour cela, l'identification du type de l'effet présent dans les spectres visibles et proche infrarouge a suivi la réduction du bruit. Une méthode simple pour ce faire est la représentation de tous les spectres en fonction du spectre moyen (remplacer les longueurs d'onde dans l'axe des abscisses par le spectre moyen) pour chaque domaine spectral étudié. La forme conique de ce graphe caractérise la présence d'un effet multiplicatif alors qu'une forme en millefeuille caractérise la présence d'un effet additif (Roger et Ecartot, 2016). Après avoir identifié les effets additifs et multiplicatifs sur les spectres, leur réduction a

été effectuée en utilisant deux algorithmes : la dérivée Savitzky-Golay et la SVN (Standard Normal Variate).

Les méthodes de dérivation des spectres sont connues par leur capacité à enlever les effets additifs. Si on admet que cet effet est de forme linéaire, alors la fonction de cette perturbation est un polynôme de degré 1 et par suite une dérivation du spectre réduit la ligne de base responsable des effets additifs. Cette méthode augmente en contrepartie le bruit, car la dérivation tend à amplifier les pics du signal. Ce qui explique l'utilisation de l'algorithme de la dérivée de Savitzky-Golay pour le lissage simultané de nos spectres (Eilers et Boelens, 2005).

Pour la réduction simultanée des effets additifs et multiplicatifs, on a appliqué la SVN qui consiste à centrer chaque spectre autour de sa moyenne (soustraire de chaque spectre la valeur moyenne des réflectances dans toutes les longueurs d'onde) et à le réduire en le divisant par son écart-type. La SVN est un algorithme qui effectue un centrage-réduction des spectres, mais selon les lignes de la base de données. La réflectance de chaque échantillon transformée par SVN est calculée selon la formule suivante (Barnes et al., 1989 ; Randolph, 2005) :

$$Y_i^* = \frac{Y_i - \bar{Y}}{\sqrt{\frac{\sum(Y_i - \bar{Y})^2}{N}}} \quad (6)$$

Avec :

Y_i^* : la réflectance de l'échantillon i ayant subi la transformation par la SVN

Y_i : la réflectance brute de l'échantillon i

\bar{Y} : La réflectance moyenne de chaque spectre

N : le nombre de longueurs d'onde

3.3.3.3 Optimisation des prétraitements de spectres

Les paramètres variés de l'algorithme Savitzky-Golay étaient la taille de la fenêtre et le degré de polynôme. En ce qui concerne l'élimination de l'effet additif, les paramètres à optimiser lors de l'application de la dérivée Savitzky-Golay, étaient la taille de la fenêtre, le degré de polynôme et l'ordre de la dérivée. Dans le but de choisir les paramètres optimaux des algorithmes sélectionnés, plusieurs essais ont été effectués en modifiant à chaque fois l'un des paramètres

rapportés dans ce paragraphe. La SVN a été appliquée sur les spectres visibles seulement vu la difficulté de l'identification du type d'effet perturbateur des spectres. Finalement, 10 et 16 prétraitements ont touché les spectres visibles et proche infrarouges respectivement en utilisant le logiciel The Unscrambler X 10.5.1 (Tableau 3.2). Les critères de choix des paramètres de prétraitements étaient : la qualité visuelle des spectres (la présence/absence et l'amplitude de l'effet du bruit, la présence/absence des pics d'absorption caractéristiques du sol et cités dans la littérature) et la capacité des spectres à estimer la teneur en eau en comparant les valeurs de RMSECV (Root Mean Squared Error of Cross Validation) lors de la validation croisée de modèle de calibration (à aborder plus en détails dans la section estimation des propriétés de sol).

$$RMSECV = \sqrt{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n}} \quad (7)$$

Avec : Y_i est la teneur en eau mesurée pour l'échantillon i

\hat{Y}_i est la valeur de teneur en eau de l'échantillon i estimée par le modèle

n est le nombre d'échantillons

Le RMSECV reflète la précision d'un modèle donné à estimer une variable expliquée (dans ce cas la teneur en eau) à partir des variables explicatives (les longueurs d'onde) lors de la validation de sa construction. Plus le RMSECV est faible, plus le modèle décrit bien les variations de la variable expliquée.

Tableau 3.2 : Caractéristiques des algorithmes de prétraitements testés pour les spectres visibles et proche infrarouge appliqués pour le sous-ensemble de calibration.

Algorithme	Variable	Visible	Proche infrarouge
Réduction	Facteur de réduction	4, 6, 8 et 10	4, 6, 8 et 10
Savitzky-Golay (lissage)	Taille de la fenêtre	5, 7, 9, 11 et 13	5, 7, 9 et 11
Savitzky-Golay (lissage)	Degré du polynôme	2	2
Dérivée Savitzky Golay	Taille de la fenêtre	-	5, 7, 9 et 11
Dérivée Savitzky Golay	Degré du polynôme	-	2
	Ordre de la dérivée	-	1 et 2
SVN	-	Appliquée	Non appliquée

3.3.4 Estimation des propriétés de sol

3.3.4.1 Sélection des variables indépendantes (longueurs d'onde)

La sélection des longueurs d'onde à l'entrée des modèles est une étape répondant au besoin de la diminution du nombre de variables indépendantes (les longueurs d'onde) ayant peu ou pas de liaison avec la variable dépendante (la propriété de sol étudiée). Pour ce faire, les coefficients de corrélation entre chaque propriété et toutes les longueurs d'onde ont été représentés en fonction de celles-ci :

$$\text{cor}(Y,X) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (8)$$

Avec : cov (X,Y) est la covariance entre la propriété étudiée et les longueurs d'onde

$$\text{cov}(X,Y) = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}) \quad (9)$$

N est le nombre d'observations ;

\bar{X} et \bar{Y} sont respectivement les moyennes de X et Y ;

σ_x et σ_y sont respectivement les écarts type de X et Y :

$$\sigma_x = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N} \quad (10)$$

Puisqu'on a trouvé des corrélations négatives, on a calculé les valeurs absolues des coefficients de corrélation et on a gardé les longueurs d'onde dont la valeur absolue de la corrélation avec la propriété étudiée est supérieure à 30 % (pour garder un bon nombre de longueurs d'onde).

3.3.4.2 Calibrations des modèles d'estimation des propriétés étudiées

La calibration des modèles à partir des données spectroscopiques représentait un défi puisque le nombre de variables indépendantes (les longueurs d'onde) est élevé et que celles qui sont successives sont souvent corrélées. La régression PLS (partial least squares ou projection to latent structures) est une méthode de calibration multivariée qui a été largement utilisée pour la modélisation des propriétés de sol en utilisant leurs spectres (Nocita et al., 2013; Genot et al., 2011 ; Gras et al., 2014; Knox et al., 2015; Rodionov et al., 2016; Knadel et al., 2013). Le concept de cette méthode est adapté à l'estimation d'une variable dépendante (dans ce cas les propriétés de sol) à partir d'un nombre important de variables indépendantes corrélées (les longueurs d'onde) (Wold et al., 2001). En effet, la procédure de modélisation de lien entre X et Y est contrôlée par un petit nombre de variables latentes. Ces variables sont estimées comme des moyennes pondérées (combinaisons linéaires) des variables observées apparentes. La PLS repose ainsi sur l'extraction de l'information des variables indépendantes et dépendantes simultanément en gardant le maximum de lien entre eux à travers leur projection sur un nouvel espace constitué par les variables latentes (Geladi et Kowalski, 1986) (Figure 3.10).

On estime que les longueurs d'onde (variables indépendantes) X et les propriétés de sol (variables dépendantes) Y peuvent être décortiquées comme suit :

$$X = T P^T + E \quad (11)$$

$$Y = U R^T + F \quad (12)$$

Avec P et R sont les vecteurs de contributions factorielles (loadings) de X et Y respectivement, T et U sont les matrices des coordonnées factorielles (scores) et E et F sont les matrices des résidus (Rutledge, no date).

La PLS est un processus itératif fondé sur deux hypothèses : les scores T de X permettent d'estimer avec une bonne précision Y ; il existe une relation linéaire entre T et U (Rutledge, no date). La construction du nouvel espace formé par les variables latentes est composée de 5 étapes (Rännar et al., 1994) :

- (1) Faire une première estimation de vecteur des scores u_1 de Y : $u_1 = y_i$
- (2) Déterminer le poids w qui relie les scores de Y à X ; $w_1 = u_1^T X / (u_1^T u_1)$
- (3) Calculer le score t_1 de X à partir de w_1 ; $t_1 = X w_1$
- (4) Déterminer le vecteur des loadings de Y r_1 à partir de t_1 (les scores de X) ; $r_1 = t_1^T Y / (t_1^T t_1)$
- (5) Utiliser les vecteurs des loadings r_1 pour calculer la variable latente $u_2 = Y r_1^T / \|r_1 r_1^T\|$

Lors de ces étapes, les vecteurs U sont utilisés pour calculer T qui sont eux aussi utilisés pour déterminer après les vecteurs U. Ces calculs sont répétitifs jusqu'à ce qu'on n'observe plus de changement significatif de u et t entre une itération et une autre. Après avoir retenu la variable latente 1, on calcule la contribution de chaque variable indépendante X à la construction de cette variable, les résidus de X et les résidus de Y qui seront soustraits de leurs matrices et on recommence le calcul de la variable latente suivante en ayant comme entrée les matrices E et F qui représentaient les matrices X et Y ayant les variances résiduelles non utilisées pour le calcul de la première variable latente (Lindgren et al., 1993). Le résultat final de la PLS est un ensemble de matrices servant au calcul des coefficients du modèle de prédiction B pour chaque variable latente (Geladi et Kowalski, 1986) :

$$B = W (P^T W)^{-1} R^T \quad (13)$$

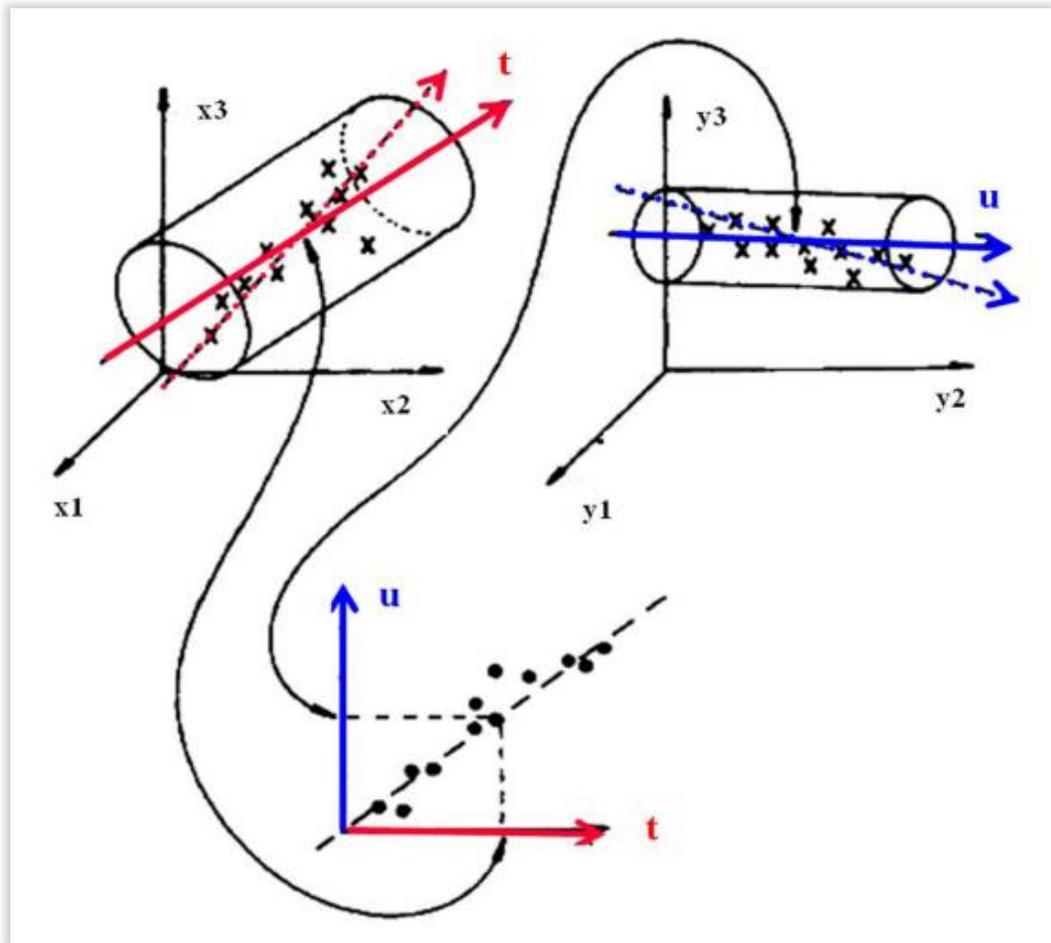


Figure 3.10 : Principes géométriques de la régression PLS (Partial Least Squares ou Projection to Latent Structures) montrant la maximisation de la corrélation entre X et Y (tirée de Rutledge, no date).

3.3.4.3 Validation des modèles

- **Validation interne des modèles**

La décomposition de la base de données spectrales en sous-ensemble de calibration et de validation du modèle est nécessaire pour évaluer sa qualité. Aussi, la validation interne (validation croisée) du modèle en utilisant le sous-ensemble de calibration assure une estimation optimale de la complexité du modèle ainsi qu'une estimation préliminaire des erreurs d'estimation des propriétés de nouveaux échantillons (Peter et al., 2009). La validation croisée est utilisée d'une façon importante pour la sélection du modèle de calibration multivarié le plus performant (Xu et al., 2004). Ce type de validation consiste à partager le sous-ensemble de

spectres de calibration en deux parties une pour la construction du modèle et une pour l'application de l'équation du modèle. Ensuite, le carré de la différence entre les valeurs réelles et les valeurs estimées est calculé lors de la validation interne ou le RMSECV et cette valeur est utilisée pour comparaison ultérieure. Ces étapes sont répétées avec des partitions différentes de la base de données de calibration. La condition d'arrêt est l'application de toutes les partitions possibles, comme cela, tous les échantillons sont utilisés pour la validation interne. À la fin de cette procédure, plusieurs modèles sont obtenus et celui ayant la RMSECV minimale est retenu (Peter et al., 2009 ; Xu et al., 2004).

- **Choix des variables latentes optimales**

La variance des variables indépendantes et dépendantes récupérée varie avec les variables latentes. Pour cette raison, il est nécessaire de sélectionner celle qui maximise à la fois la variance expliquée de X et Y et la corrélation entre eux. Cette étape permet de choisir les coefficients optimaux calculés lors de l'application de la PLS puisque ces coefficients changent avec la variable latente. Leur sélection a été effectuée en calibrant des modèles d'estimation des propriétés de sol étudiées (avec un nombre maximal de variables latentes de 7) et retenant celles qui correspondent au plus grand R^2 et le plus petit RMSECV lors de la validation croisée des modèles.

3.3.4.4 Estimation des propriétés de sol

Pour effectuer la validation externe des modèles, 25 % des échantillons ont été retirés de sorte qu'ils soient représentatifs de la population. Ce sous-ensemble a été ensuite décomposé en deux groupes en sélectionnant un échantillon et laissant le suivant pour la validation finale des modèles calibrés, toujours après avoir classé les échantillons par ordre croissant d'humidité du sol. Il suffit après d'appliquer les coefficients de régression sur les longueurs d'onde de ce sous-ensemble.

- **Test des modèles**

Les modèles construits ont été appliqués sur la moitié de sous-ensemble de calibration afin de calculer des paramètres servant à effectuer une première évaluation de la qualité des modèles et comparer entre leurs précisions. Les paramètres de performance des modèles utilisés dans cette étude sont le coefficient de détermination R^2 , RMSEP (Root Mean Squared Error of Prediction), le biais et le RPD (Residual Prediction Deviation).

R^2 reflète le pourcentage de la variance totale dans les valeurs de la propriété étudiée et expliquée par l'équation du modèle (Nagelkerke, 1991) :

$$R^2 = 1 - \frac{\sum_n (Y_{obs,i} - Y_{est,i})^2}{\sum_n (Y_{obs,i} - \bar{Y})^2} \quad (14)$$

Avec $Y_{obs,i}$ et $Y_{est,i}$ sont respectivement la valeur observée et la valeur estimée de la propriété de sol pour l'échantillon i . \bar{Y} est la moyenne des valeurs observées et n est le nombre d'observations. Dans certaines études, les conclusions élaborées à partir des modèles d'estimation des propriétés de sol se basent sur les valeurs de R^2 . Toutefois, ce paramètre n'est pas le plus adéquat pour ce faire puisqu'il est lié à la plage de données (Bellon-Maurel et McBratney, 2011).

RMSEP représente l'erreur absolue moyenne entre les valeurs estimées par le modèle et les valeurs réelles (Stone, 1974) :

$$RMSEP = \sqrt{\sum_{i=1}^n \frac{(Y_{obs,i} - Y_{est,i})^2}{n}} \quad (15)$$

Le biais est la différence entre la moyenne des valeurs estimées par les modèles et celle des valeurs observées (Bellon-Maurel et McBratney, 2011). Donc ce paramètre exprime l'erreur générale associée au modèle.

Le RPD a été utilisé dans plusieurs études pour l'évaluation de la performance des modèles calibrés puisque son calcul prend en considération la plage des valeurs observées :

$$RPD = \frac{\sigma_Y}{RMSEP} \quad (16)$$

Avec : σ_Y est l'écart type des valeurs observées

Selon Chang et Laird (2002), un modèle avec RPD inférieur à 1.4 est jugé avoir une faible performance, entre 1.4 et 2 est de moyenne performance et supérieur à 2 est un bon modèle.

- **Validation finale des modèles**

L'amélioration de la qualité des modèles a été effectuée en ajustant le biais (l'associer ou le retrancher) des valeurs estimées lors de la validation finale. Ensuite les quatre indicateurs de performance de modèles ont été recalculés pour quantifier l'effet de cette correction.

4 RÉSULTATS ET DISSCUSSIONS

4.1 Description de la base de données

4.1.1 Propriétés du sol

Le Tableau 4.1 présente une analyse descriptive des propriétés des échantillons étudiés après calcul des moyennes des réplicats. Une large plage de teneur en eau a été couverte lors de cette étude (écart-type égale à 8.42) afin de préciser l'effet de cette propriété sur la caractérisation des autres propriétés par la signature spectrale du sol. En ce qui concerne la texture du sol, on remarque la dominance de la fraction sablonneuse (moyenne égale à 74.21 %) dans les échantillons étudiés par rapport aux teneurs en argile (moyennes égale à 7.13 %) et en matière organique (moyenne égale à 4.34 %).

Tableau 4.1 : Statistique descriptive pour les différentes propriétés étudiées.

Propriété	Minimum	Maximum	Moyenne	Écart-type
Teneur en eau (%)	0.00	31.37	8.75	8.42
Teneur en argile (%)	2.72	13.23	7.13	2.36
Teneur en matière organique (%)	1.90	7.20	4.34	1.35
Teneur en sable (%)	54.25	91.10	74.21	9.07

4.1.2 Données spectrales

Après le calcul des spectres moyens, et la réduction des longueurs d'onde dans les deux domaines visible et proche infrarouge, une matrice composée de 258 longueurs d'onde (colonnes) et 288 observations (lignes) a été retenue. La comparaison entre les spectres de

tous les réplicats et ceux moyennés (Figure 4.1) montre la réduction de la dimensionnalité de la base des données sans perte d'informations.

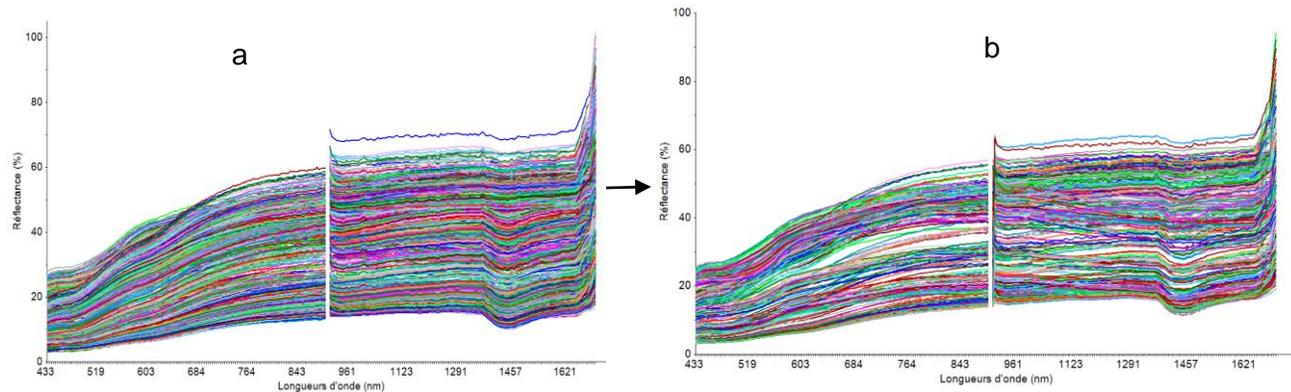


Figure 4.1 : Spectres visibles et proche infrarouges avant (a) et après (b) la réduction des longueurs d'onde par le calcul des moyennes des réplicats

4.2 Résultats des traitements des spectres

4.2.1 Exploration des spectres

4.2.1.1 Inspection visuelle des spectres

Des spectres moyens des niveaux de teneurs en eau avec un pas de 5% ont été calculés dans le but de mettre en évidence l'effet de l'humidité sur la réponse spectrale des échantillons de sol. Le résultat illustré à la Figure 4.2 permet de soulever deux points (en accord avec ce qui a été rapporté dans la littérature (Weidong et al., 2002)) :

- L'augmentation de l'humidité du sol a un effet abaissant de l'albédo dans le domaine du visible et proche infrarouge avec une diminution claire de la pente des spectres entre 400 et 700 nm ;
- Le pic d'absorption caractéristique des molécules d'eau vers 1400 nm subit une diminution en largeur et en profondeur autant que la teneur en eau augmente.

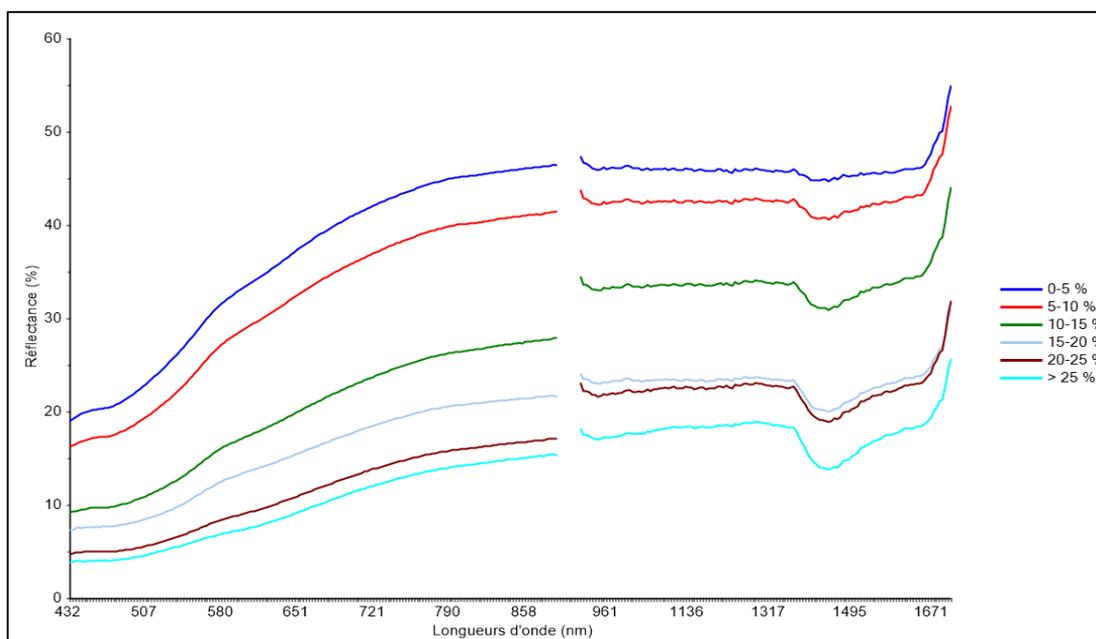


Figure 4.2 : Spectres moyens visibles et proche infrarouges de 5 intervalles de niveaux de teneur en eau.

4.2.1.2 ACP des spectres

À cette étape, l'ACP a été appliquée sur les spectres non prétraités du sous-ensemble de calibration (utilisé aussi pour la classification des spectres selon la teneur en eau en gardant le maximum d'homogénéité spectrale). Seule la réduction de longueurs d'onde a été appliquée pour réduire la dimensionnalité de la base de données. L'objectif de la visualisation du nuage de points des spectres dans l'espace formé par les deux premières composantes principales en différentes couleurs selon l'humidité du sol (Figure 4.3) est d'identifier le nombre de sous-ensembles de spectres.

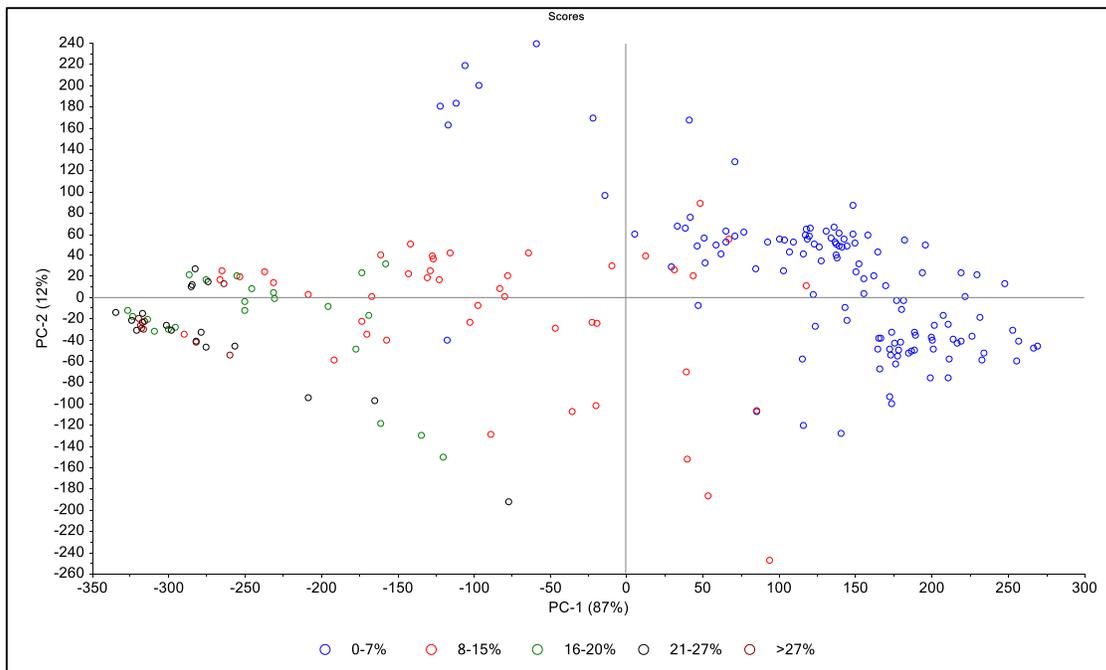


Figure 4.3 : Projection en 2-dimensions des deux premières composantes de l'ACP des spectres de calibration avec représentation des niveaux d'humidité de sol en couleurs différentes.

Les deux premières composantes récupèrent 99 % de la variance totale des spectres du sous-ensemble de calibration (Annexe 2). Les niveaux d'humidité du sol ont été arbitrairement groupés en cinq intervalles. Leurs bornes ont été choisies après des essais d'homogénéisation de chaque groupe de points. Pour chaque nuage de points élémentaire, on a essayé de garder le plus grand nombre de scores (points) dans un même intervalle d'humidité du sol. D'après la Figure 4.3, on a pu remarquer que plus les scores diminuent selon la première composante plus la teneur en eau diminue. En outre, deux observations importantes peuvent être reportées :

- Il est difficile d'identifier des seuils de séparabilité des spectres à partir des niveaux de teneur en eau des échantillons autant que l'humidité augmente (les spectres des échantillons les plus secs représentés en bleu dans la Figure 4.3, constitue un groupe homogène comparés à ceux correspondant aux échantillons humides représentés en violet, noir et vert). Ceci est principalement dû à l'effet microscopique de l'eau qui change la structure et la couleur du sol (Ji et al., 2015).
- Les scores présentent des dispersions plus ou moins importantes (certains points sont très distants les uns des autres) reflétant des écarts importants dans les réponses spectrales entre un échantillon et un autre.

De ce fait, il a été difficile de différencier avec précision des groupes de spectres.

4.3 Estimation des propriétés du sol

4.3.1 Prétraitements appliqués

Comme décrit dans la section 3.3.3, les spectres visibles et proche infrarouges du sous-ensemble de calibration ont subi la suppression du bruit et la réduction d'un ou des deux types d'effets additifs et multiplicatifs. Le choix du facteur de réduction ainsi que de l'ordre du polynôme et la taille de la fenêtre de l'algorithme Savitzky-Golay se sont reposés sur la diminution de l'effet du bruit dans les spectres sans affecter les pics d'absorption. Dans l'intention de distinguer entre l'effet additif et multiplicatif des signaux, des graphes de réflectance en fonction du spectre moyen ont été dressés et présentés dans la Figure 4.4 :

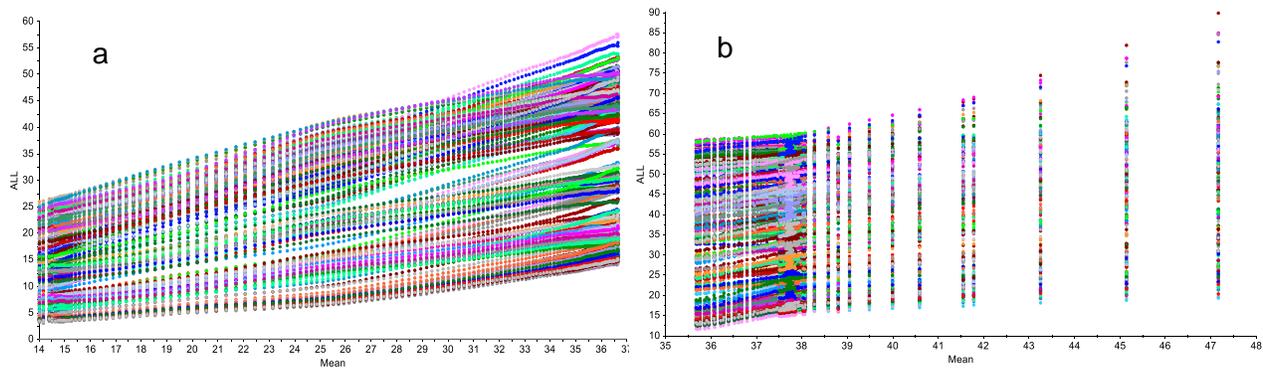


Figure 4.4 : Spectres visibles (a) et proche infrarouges (b) en fonction des moyennes des réflectances pour chaque longueur d'onde.

Selon la Figure 4.4.b, la forme observée dans le cas des spectres du proche infrarouge est en millefeuille avec accentuation de cet effet avec l'augmentation de la valeur de longueur d'onde indiquant la présence d'un effet additif croissant, ce qui justifie le choix de la dérivation des spectres proche infrarouges. Quant aux spectres visibles (Figure 4.4.a), la courbe des spectres en fonction du spectre moyen n'a montré aucune forme caractéristique pouvant renseigner sur le type d'effet sur les signaux. Pour ce fait, la SNV, qui élimine simultanément les effets additifs et multiplicatifs, a été appliquée.

Une fois les signatures spectrales lissées, la sélection des variables de la dérivée de Savitzky-Golay a été basée sur la comparaison des RMSECV de l'estimation de la teneur en eau à partir des spectres proche infrarouges. Le graphe illustré à la Figure 4.5 présente la variation des valeurs de RMSECV en fonction de l'ordre de la dérivée et de la taille de la fenêtre :

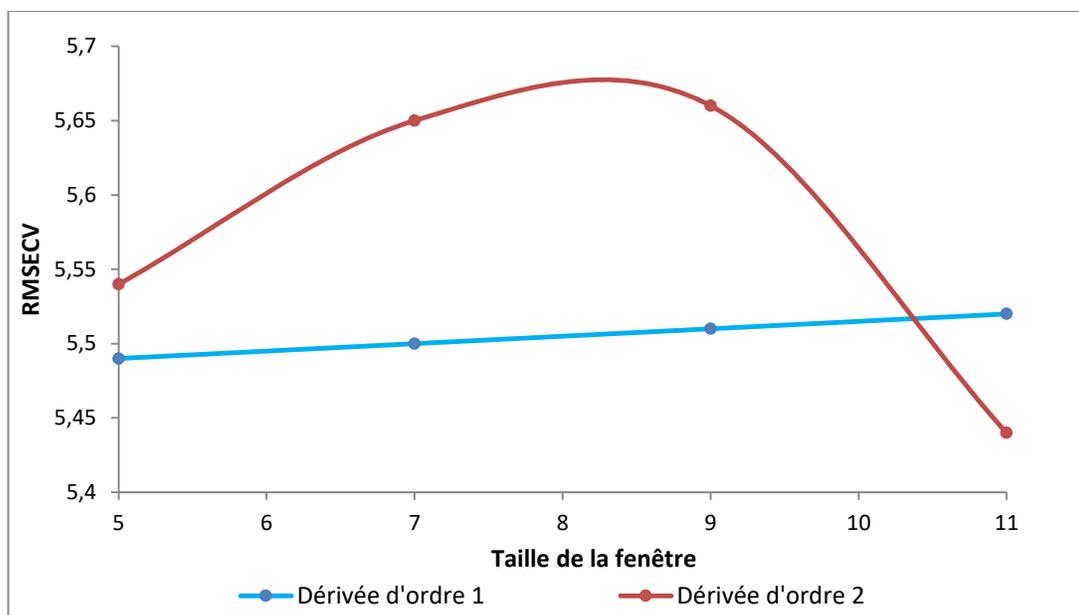


Figure 4.5 : Variation des valeurs de RMSECV avec l'ordre de la dérivée de Savitzky-Golay et de la taille de la fenêtre.

Selon la Figure 4.5, les valeurs de RMSECV associées à une dérivée de premier ordre sont presque constantes. Une dérivation du deuxième ordre des spectres diminue leur capacité à estimer l'humidité de sol au fur et à mesure que la taille de la fenêtre augmente jusqu'à atteindre une valeur de 9. Ceci peut être expliqué par la création de nouveaux pics de bruit. Une augmentation de la taille de la fenêtre pour le lissage des données (taille de fenêtre égale à 11) diminue ces erreurs. Ainsi, l'ordre de la dérivée de cette taille de fenêtre a été retenu. En résumé, les prétraitements appliqués sont :

- Pour le domaine du visible : Réduction du nombre de longueurs d'onde au dixième, filtre Savitzky-Golay avec un polynôme de deuxième degré et une fenêtre de taille égale à 11 et application de la SVN ;
- Pour le domaine du proche infrarouge : Réduction du nombre de longueurs d'onde au quart, filtre Savitzky-Golay avec un polynôme du deuxième degré et une fenêtre de taille égale à 11 et une dérivée Savitzky-Golay d'ordre deux, avec un polynôme du deuxième degré et une taille de fenêtre égale à 11.

Les spectres de sous-ensemble de calibration (visibles et proche infrarouges) ayant subis tous ces prétraitements sont présentés dans la Figure 4.6. Ces prétraitements retenus ont été ensuite appliqués sur le sous-ensemble de la base de validation.

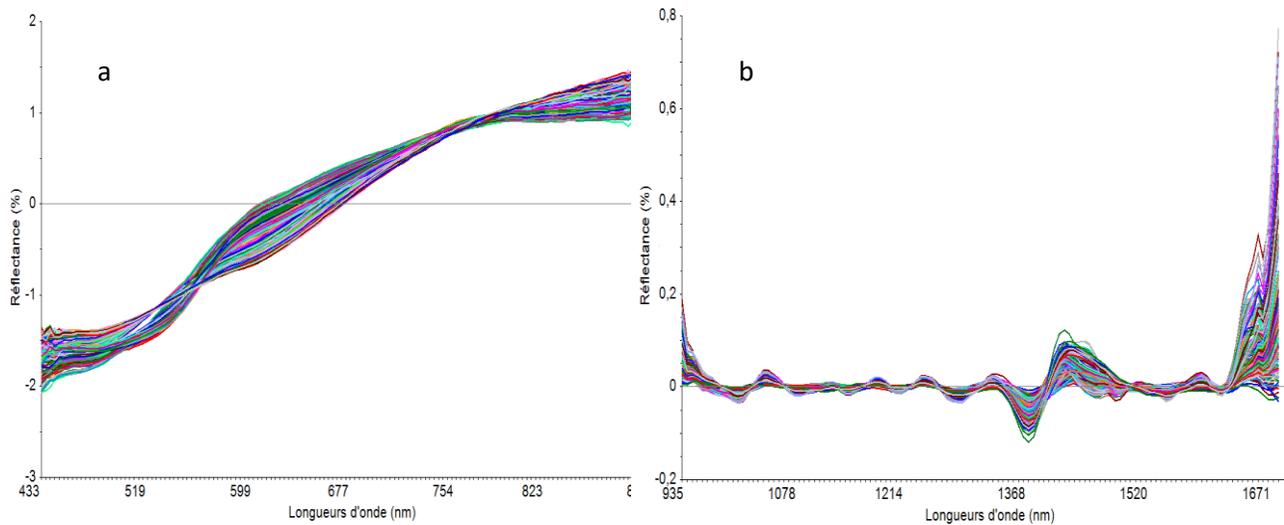


Figure 4.6 : Spectres visibles (a) et proche infrarouges (b) prétraités.

4.3.2 Choix des niveaux de la teneur en eau

À partir de l'ACP effectuée sur les spectres (Figure 4.3), quatre amas de points ont été différenciés c'est pourquoi on a supposé qu'il existe quatre classes spectrales. L'objectif de cette étape était de créer des classes de teneur en eau en optimisant l'homogénéisation spectrale intra-classe. Ainsi, une classification non supervisée des spectres a été effectuée. Dans l'intention d'étudier la relation entre le degré de distinction des spectres en classes et les plages de valeurs d'humidité de sol, les classes des spectres obtenues ainsi que les scores de teneurs en eau sont présentés en couleurs différentes après avoir effectué une ACP (Figure 4.7) :

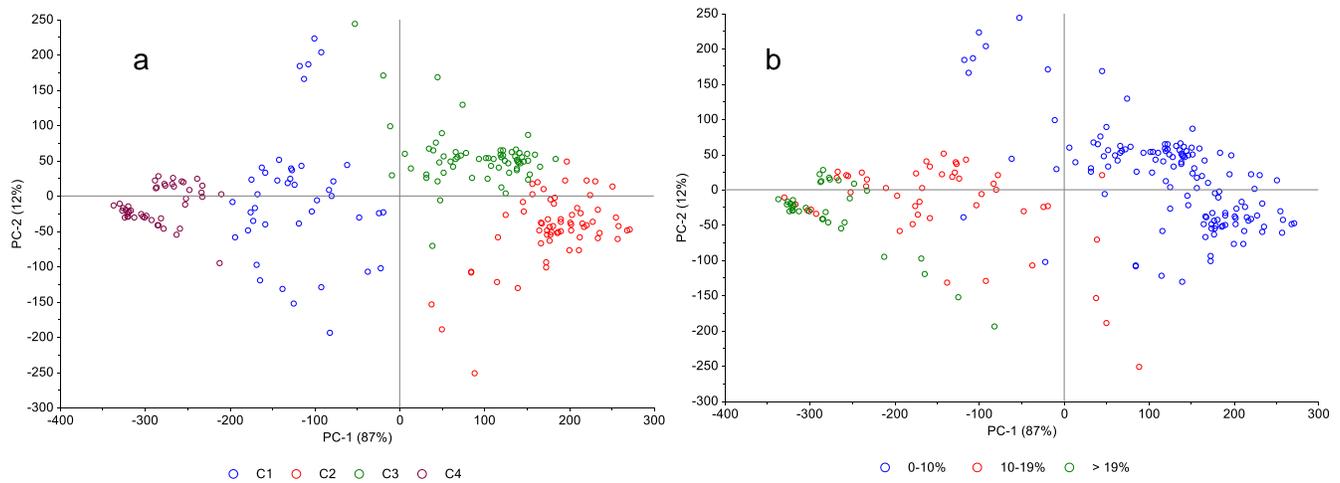


Figure 4.7 : Projection en 2-dimensions montrant les classes spectrales : C1, C2, C3 et C4 issues de la classification non supervisée des spectres (a) et les classes d'humidité du sol : échantillons de teneurs en eau entre 0 et 10 %, entre 10 et 19 % et supérieures à 19 % issues de l'homogénéisation intra-classes spectrales des teneurs en eau (b).

Le défi lié à cette étape résulte de l'importance des différences entre les spectres (nuage de points dispersé) ainsi que l'interférence d'autres propriétés. En effet, il existe des échantillons du sol dont l'écart entre leurs teneurs en eau sont non négligeable et ayant des réponses spectrales similaires (appartenant à la même classe spectrale) et *vice versa*. Ceci peut être expliqué par l'effet de la distance entre la sonde de réflectance et l'échantillon du sol qui change avec l'humidité du sol. Pour des sols humides, la surface des échantillons est relativement compacte, par conséquent, la distance entre la sonde et les particules du sol ne change pas entre une mesure et une autre. Par contre, la texture sablonneuse des échantillons du sol rend leur surface granulaire à l'état sec et par la suite, la sonde s'enfonce plus dans les particules du sol augmentant alors l'albédo général entre une mesure de réflectance et une autre. Par exemple, il a été difficile de définir deux groupes de teneur en eau appartenant aux classes spectrales C2 et C3 sur la Figure 4.7.a. Pour cela, ces deux classes ont été fusionnées en une seule. Le choix des bornes des niveaux d'humidité du sol a été fait après des essais d'homogénéisation de chaque classe spectrale, en gardant dans la même classe spectrale le plus grand nombre d'échantillons avec des valeurs de teneurs en eau appartenant à l'intervalle choisi. Cette opération a abouti à trois groupes de teneurs en eau : entre 0 et 10 %, entre 10 et 19 % et supérieur à 19 %.

4.3.3 Calibration des modèles de régression PLS

Avant de calibrer les modèles d'estimation de l'humidité, des composantes de la texture du sol et de la teneur en matière organique, les longueurs d'onde les moins corrélées avec chaque propriété ont été supprimées et seules les variables latentes optimales ont été retenues. Dans chacune de ces deux étapes, deux séquences de traitements ont été effectuées : une avec tous les spectres du sous-ensemble de la base de la calibration et une avec trois groupes de spectres pour chaque classe de teneur en eau. Ceci sert à l'étude de la capacité des spectres à estimer la teneur en eau et à l'étude de l'effet de la décomposition de la base de données sur la précision de l'estimation des variables texturales du sol.

4.3.3.1 Choix des variables indépendantes (longueurs d'onde)

Le corrélogramme à la Figure 4.8 permet de visualiser les corrélations entre l'humidité de sol et chaque longueur d'onde alors que les Figures 4.9, 4.10 et 4.11 présentent les corrélogrammes des teneurs en argile, sable et matière organique respectivement pour tout le sous-ensemble de calibration et pour chaque classe d'humidité du sol. Pour la teneur en eau, il existe peu de longueurs d'onde dont la corrélation avec l'humidité du sol est proche de 0. De plus, l'humidité du sol a l'intensité du lien le plus fort avec les données spectrales reflétant la forte réponse spectrale de l'eau. Pour les autres propriétés de sol étudiées, la teneur en sable présente les extrema de corrélation avec les longueurs d'onde les plus importants tandis que la teneur en matière organique possède les extrema les moins importants.

Globalement, les corrélations sont relativement meilleures pour les composantes de la texture du sol et la matière organique après classification des spectres en fonction de la teneur en eau (corrélogrammes en bleu, rouge et vert dans les Figures 4.9, 4.10 et 4.11) qu'avant la classification (corrélogrammes en pointillés dans les mêmes Figures). En outre, les longueurs d'onde où le coefficient de corrélation présente un pic, changent très peu avec la classe d'humidité du sol. Toutefois, la force de corrélation des variables de texture et de la matière organique et les données spectrales change. L'effet de l'eau sur la corrélation entre les propriétés du sol et les spectres n'est pas uniforme puisqu'il existe des longueurs d'onde où la valeur absolue du coefficient de corrélation augmente ou diminue avec la classe d'humidité de sol. Par exemple, la force de corrélation (en valeur absolue) entre la matière organique et les données spectrales augmente considérablement pour les échantillons de teneur en eau entre 10 et 19 % (Figure 4.11). Ces changements auront certainement une influence sur l'estimation

des composantes de la texture et de la matière organique à partir des données spectrales. Cette influence sera examinée plus en détail dans les parties qui suivent.

Aussi, dans le cas des quatre propriétés de sol, l'information est partagée sur plusieurs longueurs d'onde, en particulier dans le domaine du proche infrarouge (Figures 4.8, 4.9, 4.10 et 4.11). Pour cette raison, on a opté pour l'élimination des longueurs d'onde ayant de faibles corrélations avec la propriété étudiée au lieu de sélectionner les longueurs d'onde dont la relation avec les variables dépendantes est forte (les valeurs absolues des coefficients de corrélation sont maximales).

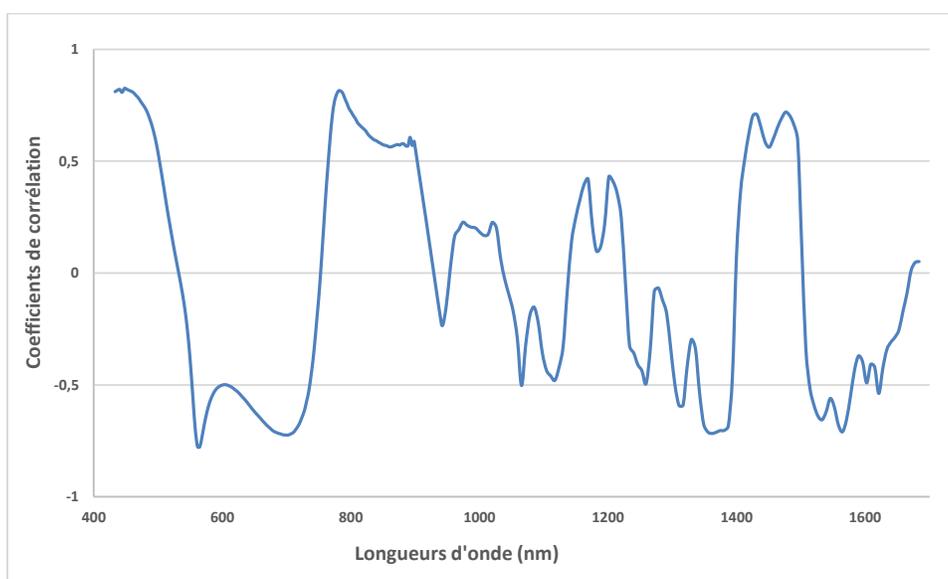


Figure 4.8 : Corrélogramme de la teneur en eau et des longueurs d'onde.

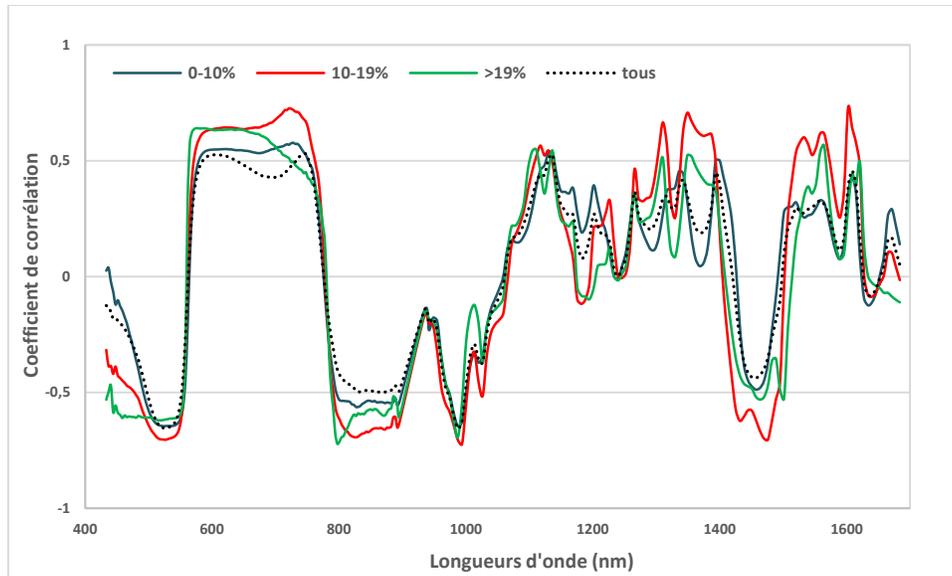


Figure 4.9 : Corrélogrammes de la teneur en argile et des longueurs d'onde pour tous le sous-ensemble de calibration et pour les trois classes d'humidité de sol (0-10 %, 10-19 % et >19 % de la teneur en eau du sol).

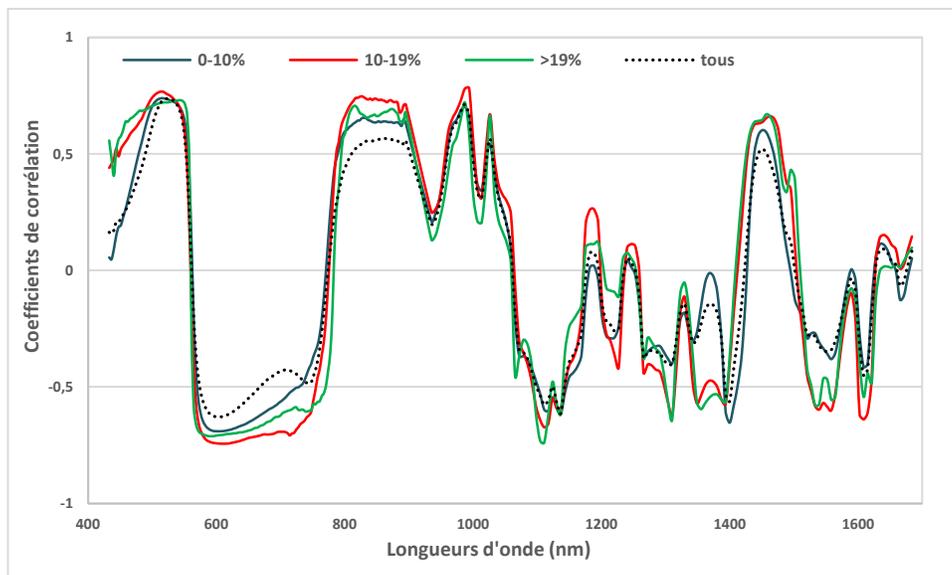


Figure 4.10 : Corrélogrammes de la teneur en sable et des longueurs d'onde pour tous le sous-ensemble de calibration et pour les trois classes d'humidité de sol (0-10 %, 10-19 % et >19 % de la teneur en eau du sol).

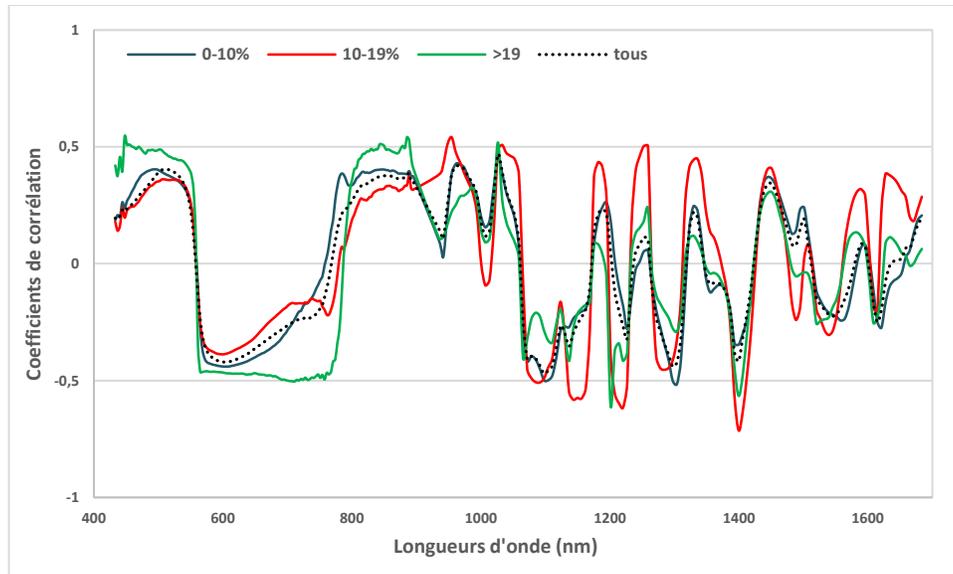


Figure 4.11 : Corrélogrammes de la teneur en matière organique et des longueurs d'onde pour tout le sous-ensemble de calibration et pour les trois classes d'humidité de sol (0-10 %, 10-19 % et >19 % de la teneur en eau du sol).

4.3.3.2 Choix des variables latentes optimales

Les composantes de la PLS affectent la précision des modèles générés. Pour étudier cet effet et choisir les variables latentes qui optimisent l'estimation des propriétés du sol, la variation de la RMSECV et du R^2 en fonction des variables latentes a été examinée lors de la validation des modèles d'estimation de la teneur en eau, en argile, en sable et en matière organique du sol. Avant classification des spectres en fonction de l'humidité du sol, les variables latentes 6 et 7 (ou composantes de PLS) présentent les R^2 les plus élevés et RMSECV les plus faibles dans la validation des modèles d'estimation de la teneur en eau ainsi que la teneur en sable, et la teneur en argile ainsi que la teneur en matière organique respectivement (Figure 4.12). Après classification des spectres en C1, C2 et C3 qui réfèrent aux échantillons d'humidité comprise entre 0 et 10 % ($0 \leq \theta < 10$), 10 et 19 % ($10 < \theta \leq 19$ %) et supérieure à 19 % ($\theta > 19$) respectivement, il a été possible d'étudier la composante de PLS permettant de récupérer le maximum de variance pour chaque classe de teneur en eau chez les composantes de la texture du sol et la quantité de matière organique. Quant à la teneur en argile, la première variable latente représente la variance la plus élevée des spectres de la classe C3 par rapport aux autres variables latentes. La troisième et la sixième composante permettent d'optimiser la précision des modèles d'estimation de la teneur en argile pour les classes C2 et C1 respectivement (Figure 4.13). Ceci peut être le résultat de l'effet de la teneur en eau sur les modèles d'estimation des trois propriétés étudiées. En effet, comme expliqué dans la section

3.3.4.2, la construction des modèles PLS aboutit à des coefficients dont le calcul est itératif. Après chaque étape, les résidus issus du calcul d'une composante PLS sont soustraits de la matrice utilisée pour le calcul de la composante PLS suivante. Ainsi, la variable latente 1 permet de récupérer la plus grande partie de variance des données dans la classe C3 (les échantillons les plus humides) ou encore il faut plus de calcul de variables latentes pour rendre les résidus dans la classe C1 (échantillons les plus secs) négligeables à l'opposé de la classe C3. La même observation peut être notée pour la teneur en sable (Figure 4.14) où le nombre d'itérations augmente avec la diminution de l'humidité du sol (la première composante pour la classe C3, la deuxième pour la classe C2 et la cinquième pour la classe C3). En ce qui concerne la matière organique (Figure 4.15), la variance des données de la classe C2 nécessite le plus grand nombre d'itérations (septième composante PLS) contrairement à la classe C1 (cinquième composante) et C3 (sixième composante). Toutes ces différences mettent en valeur l'effet de l'eau sur la variance spectrale qui explique les la texture du sol et son teneur en matière organique. Selon ces constatations, l'eau agit en réduisant la variance des spectres permettant de renseigner sur les propriétés de sol étudiées.

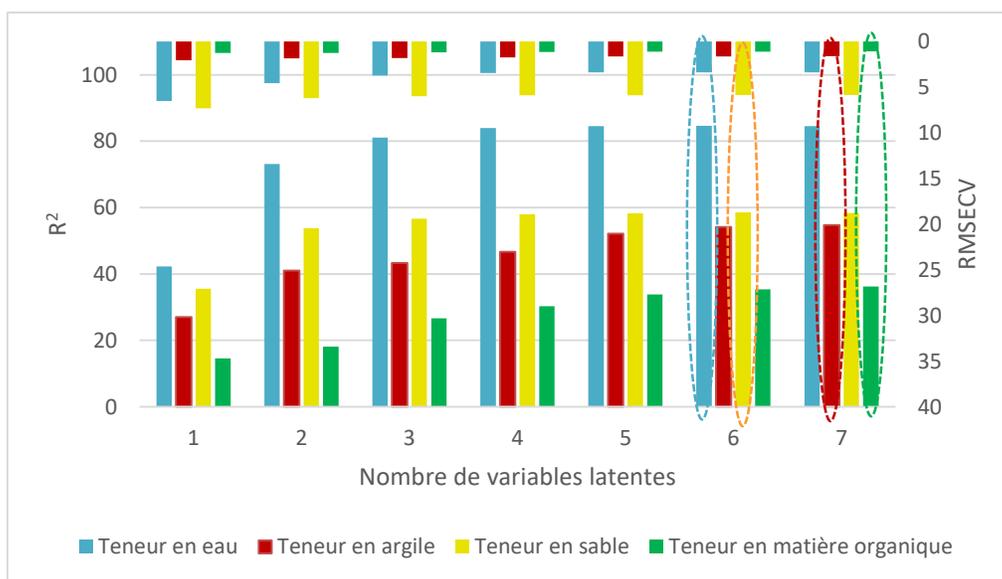


Figure 4.12 : Variation des coefficients de détermination (R^2) et RMSECV avec les variables latentes lors de la validation des modèles d'estimation de la teneur en eau, en argile, en sable et en matière organique. Les ellipses discontinues représentent les composantes PLS sélectionnées pour l'estimation des propriétés de sol étudiées et correspondant à la meilleure combinaison (R^2 le plus élevé et RMSECV le plus faible).

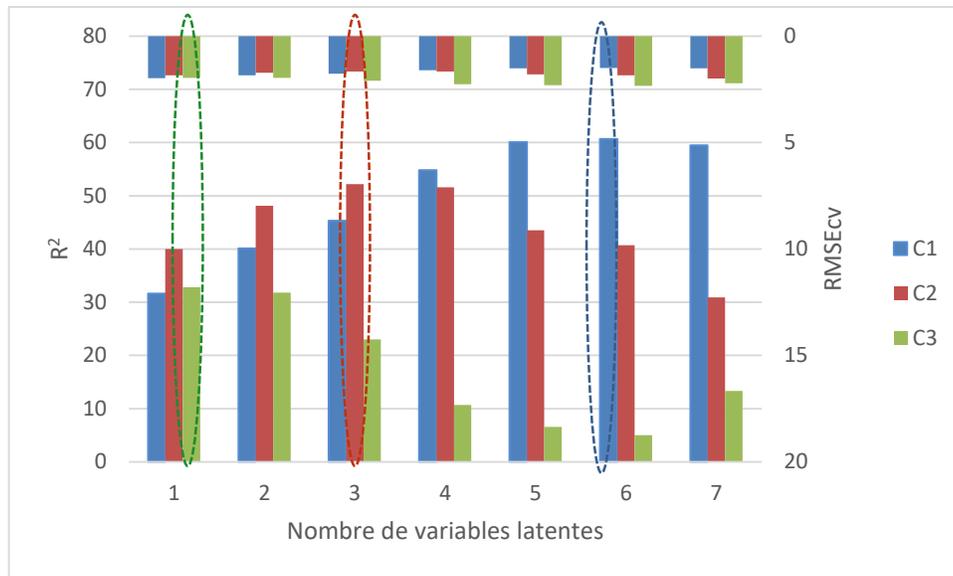


Figure 4.13 : Variation des coefficients de détermination (R^2) et RMSECV avec les variables latentes lors de la validation des modèles d'estimation de la teneur en argile pour les trois classes d'humidité de sol. Les ellipses discontinues représentent les composantes PLS sélectionnées pour l'estimation de la texture du sol et correspondant à la meilleure combinaison (R^2 le plus élevé et RMSECV le plus faible). C1, C2 et C3 représentent les trois classes d'humidité du sol (0-10 %, 10-19 % et >19 % de la teneur en eau du sol).

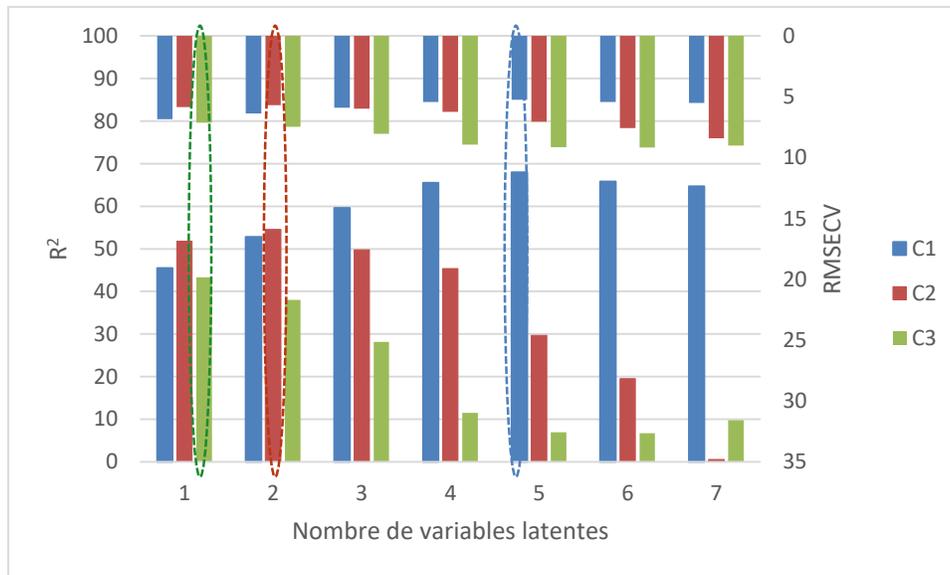


Figure 4.14 : Variation des coefficients de détermination (R^2) et RMSECV avec les variables latentes lors de la validation des modèles d'estimation de la teneur en sable pour les trois classes d'humidité de sol. Les ellipses discontinues représentent les composantes PLS sélectionnées pour l'estimation de la texture du sol et correspondant à la meilleure combinaison (R^2 le plus élevé et RMSECV le plus faible). C1, C2 et C3 représentent les trois classes d'humidité du sol (0-10 %, 10-19 % et >19 % de la teneur en eau du sol).

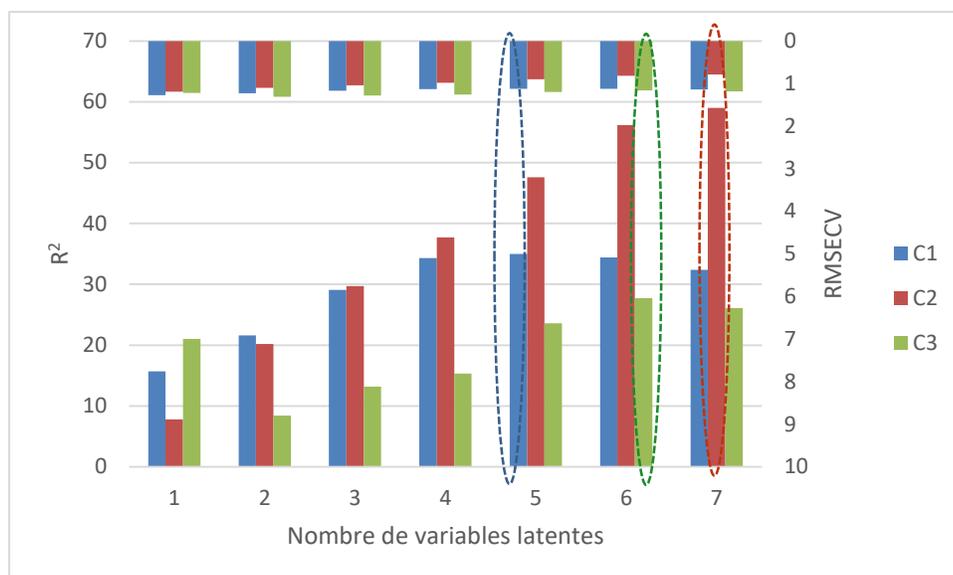


Figure 4.15 : Variation des coefficients de détermination (R^2) et RMSECV avec les variables latentes lors de la validation des modèles d'estimation de la teneur en matière organique pour les trois classes d'humidité de sol. Les ellipses discontinues représentent les composantes PLS sélectionnées pour l'estimation de la texture du sol et correspondant à la meilleure combinaison (R^2 le plus élevé et RMSECV le plus faible). C1, C2 et C3 représentent les trois classes d'humidité du sol (0-10 %, 10-19 % et >19 % de la teneur en eau du sol).

Finalement, les coefficients issus de l'application de la PLS sur les spectres pour l'estimation des propriétés de sol retenus sont ceux des variables latentes citées dans le Tableau 4.2 :

Tableau 4.2 : Variables latentes retenues pour l'estimation des propriétés de sol étudiées.

	Classe	Humidité de sol	Teneur en argile	Teneur en sable	Teneur en matière organique
Tout le sous-ensemble de la base de calibration	-	6	7	6	7
	C1	-	6	5	5
Spectres classés	C2	-	3	2	7
	C3	-	1	1	6

4.3.3.3 Calibration des modèles d'estimation de la teneur en eau et des teneurs du sol en argile, sable et matière organique

- **Modèles calibrés sans décomposition de la base de données de calibration**

Comme illustré dans la section 3.3, deux séquences d'estimation des paramètres de sol étudiés ont été effectuées : une avec tous les spectres de sous-ensemble de calibration et une après son partitionnement en trois groupes selon le niveau d'humidité du sol. Selon la Figure 4.16, le modèle d'estimation de la teneur en eau présente une faible performance dans la quantification des valeurs très faibles (voire nulles) de cette propriété. Cela peut être expliqué par une variance importante des propriétés spectrales des échantillons secs causée par les différences de leurs caractéristiques texturales et organiques.

Le modèle d'estimation de la teneur en eau est caractérisé par un coefficient de détermination R^2 de 0.88. L'objectif de cette étape est la vérification de l'intensité de l'impact de la teneur en eau sur la signature spectrale du sol après sélection des longueurs d'onde les plus riches en informations d'une part et pour vérifier si l'utilisation d'une partie du proche infrarouge seulement permet l'estimation de la teneur en eau d'autre part (puisque le pic d'absorption principal de l'eau est situé autour de 1900 nm (Bowers et Hanks, 1965)). Les points sont beaucoup plus dispersés autour de la droite dans les Figures 4.17, 4.18 et 4.19 (correspondant aux variables de la texture du sol et à la matière organique) que dans la Figure 4.16. Cela veut dire que pour la même valeur observée, différentes valeurs ont été estimées par les modèles dans le cas des teneurs en argile, sable et matière organique. Ceci est attendu puisque pour chaque échantillon ayant des caractéristiques texturales et organiques bien déterminés, des spectres ont été acquis dans différents niveaux d'humidité du sol.

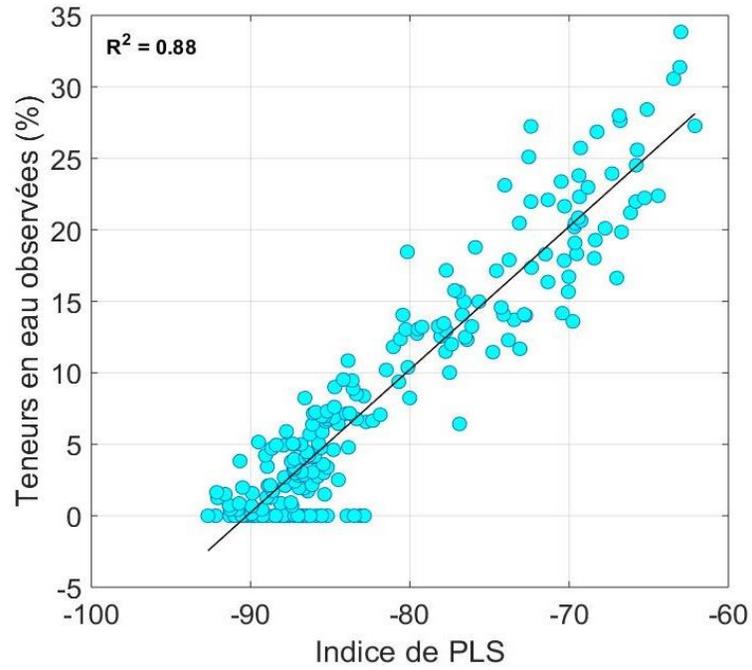


Figure 4.16 : Valeurs de teneur en eau observées en fonction des coefficients issus de l'application de la PLS.

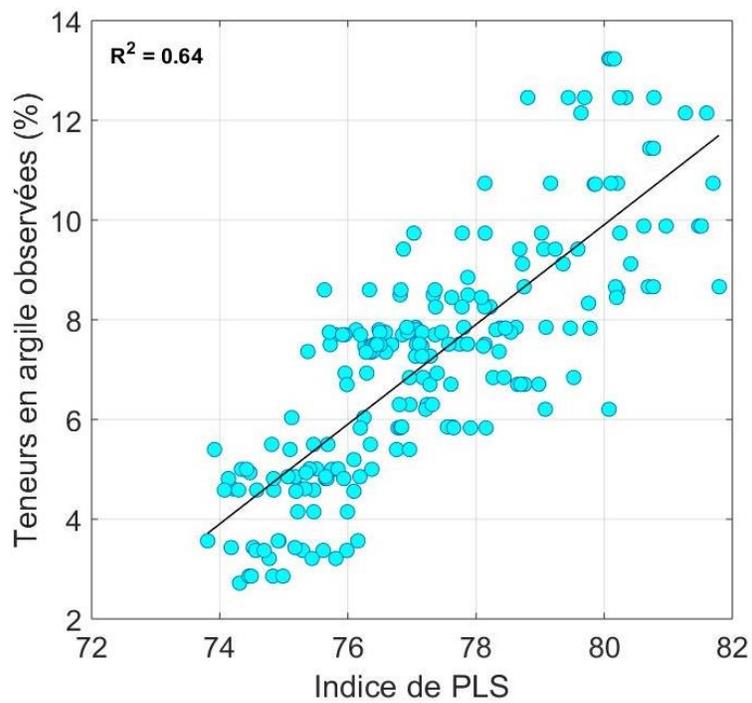


Figure 4.17 : Valeurs de teneur en argile observées en fonction des coefficients issus de l'application de la PLS.

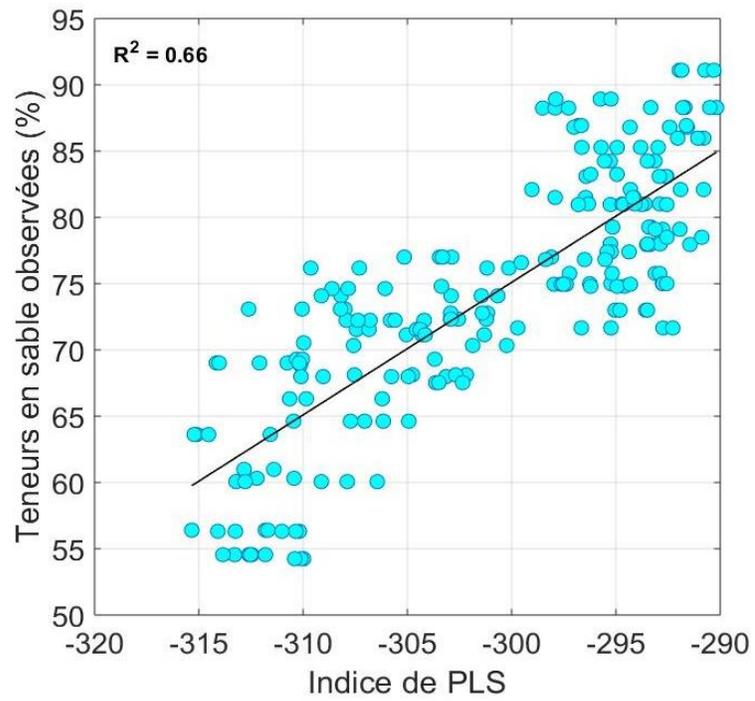


Figure 4.18 : Valeurs de teneur en sable observées en fonction des coefficients issus de l'application de la PLS.

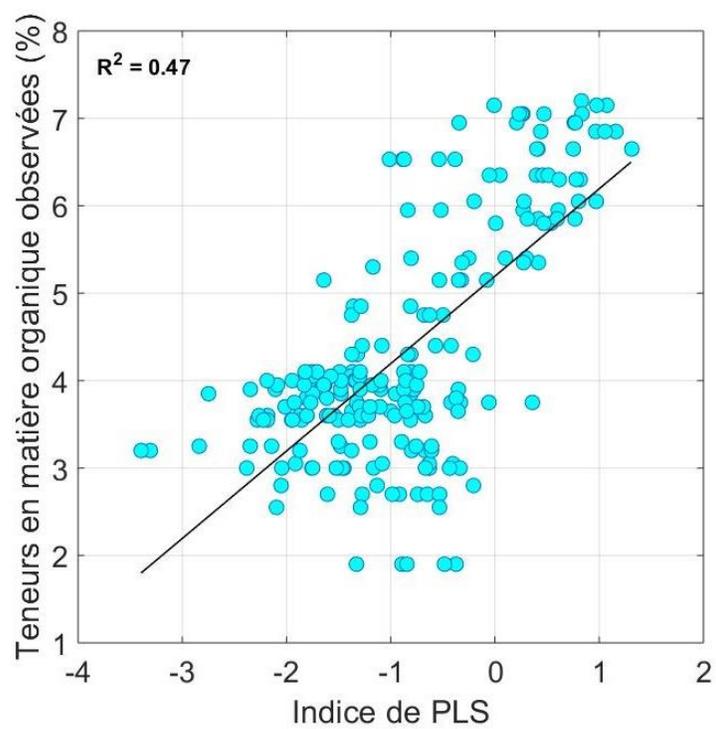


Figure 4.19 : Valeurs de teneur en matière organique observées en fonction des coefficients issus de l'application de la PLS.

- **Modèles calibrés après classification de la base de données de calibration**

Après décomposition des spectres de calibration en classes pour l'estimation de la texture du sol et de son contenu en matière organique, l'effet de la teneur en eau illustré dans les Figures 4.17, 4.18 et 4.19 a été atténué notamment pour la classe C1 (représentant les échantillons les plus secs) lors de la calibration des modèles d'estimation de la teneur en argile (Figure 4.20) (R^2 a passé de 0.64 avant classification à 0.70 pour la classe C1). La même constatation peut être repérée dans le cas de la teneur en sable (Figure 4.21) (R^2 a passé de 0.66 avant classification à 0.75 pour la classe C1). De plus, R^2 des modèles diminue lorsque que le niveau de teneur en eau augmente sauf pour la teneur en matière organique. Quant à cette propriété (Figure 4.22), l'atténuation de l'effet de l'eau est beaucoup plus remarquable pour les classes C2 et C3 (R^2 a passé de 0.47 avant classification à 0.79 et 0.84 respectivement pour les classes C2 et C3). Cela peut être expliqué soit par le fait que la corrélation entre les longueurs d'onde et la matière organique a augmenté pour les classes C2 et C3 (Figure 4.11) ou qu'elles sont contrôlées par deux patrons de modélisations différents c'est-à-dire que le fait de les estimer simultanément fait baisser le niveau de corrélation.

Selon l'analyse descriptive du sous-ensemble de calibration pour les trois classes (Annexe 3), les plages des valeurs observées des propriétés étudiées ne présentent pas des écarts importants entre une classe d'humidité du sol et une autre. Les différences entre les modèles déduits à partir de chaque classe sont donc majoritairement attribuées aux données spectrales.

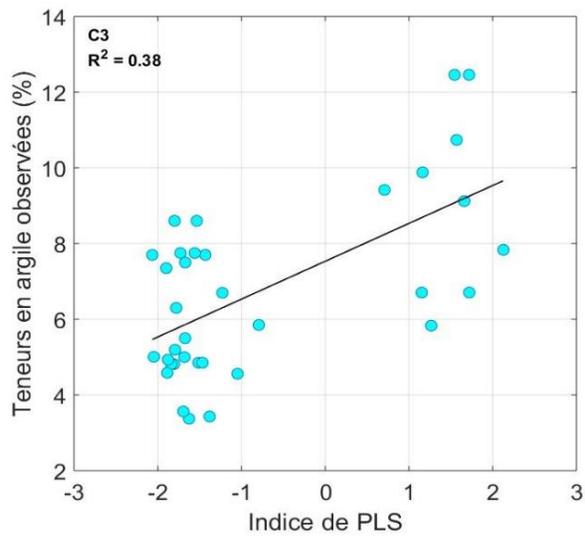
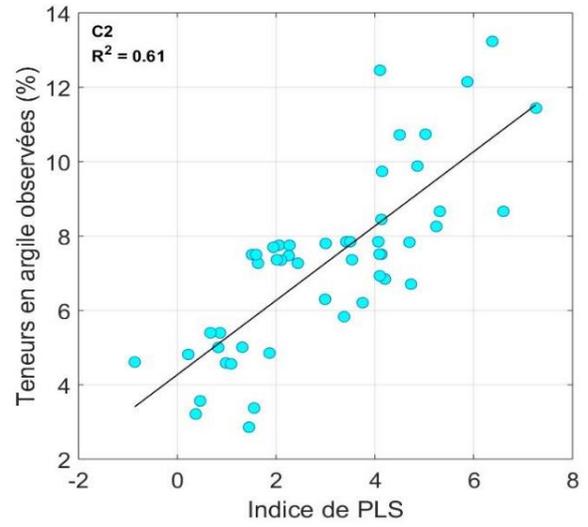
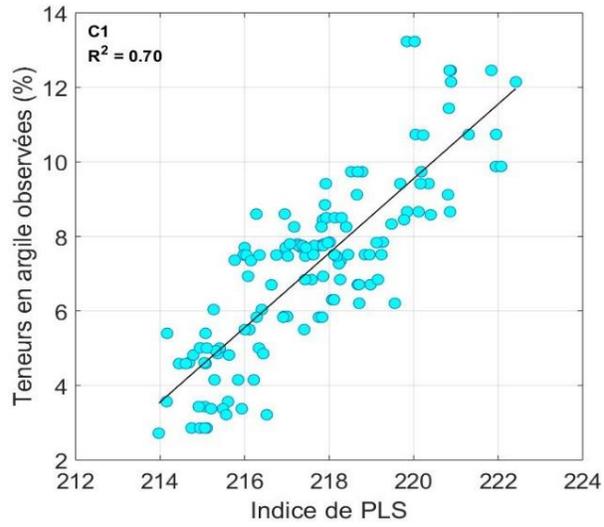


Figure 4.20 : Valeurs de teneur en argile observées en fonction des coefficients issus de l'application de la PLS pour les trois classes d'humidité du sol C1 (0-10 %), C2 (10-19 %) et C3 (>19 %).

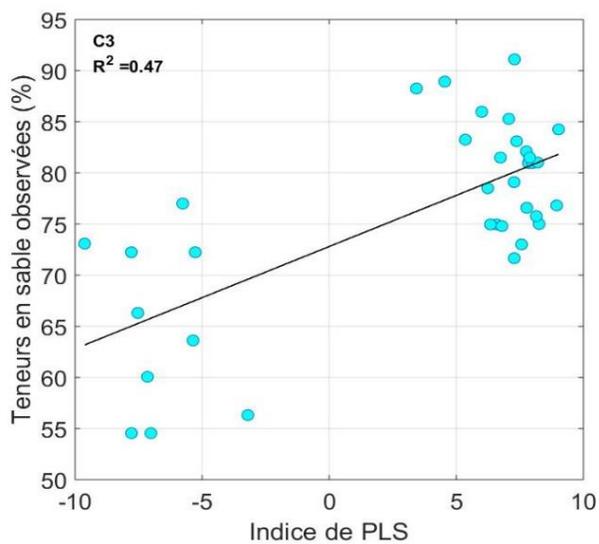
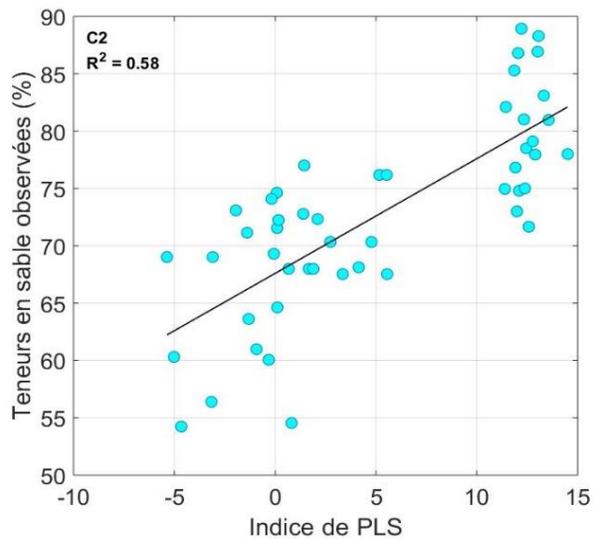
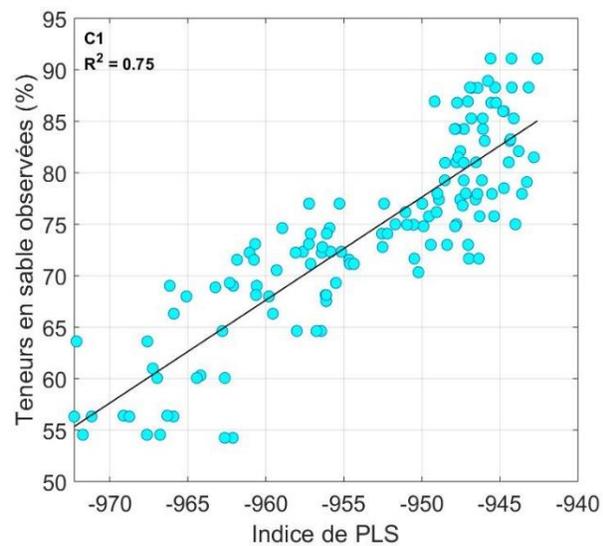


Figure 4.21 : Valeurs de teneur en sable observées en fonction des coefficients issus de l'application de la PLS pour les trois classes d'humidité de sol C1 (0-10 %), C2 (10-19 %) et C3 (>19 %).

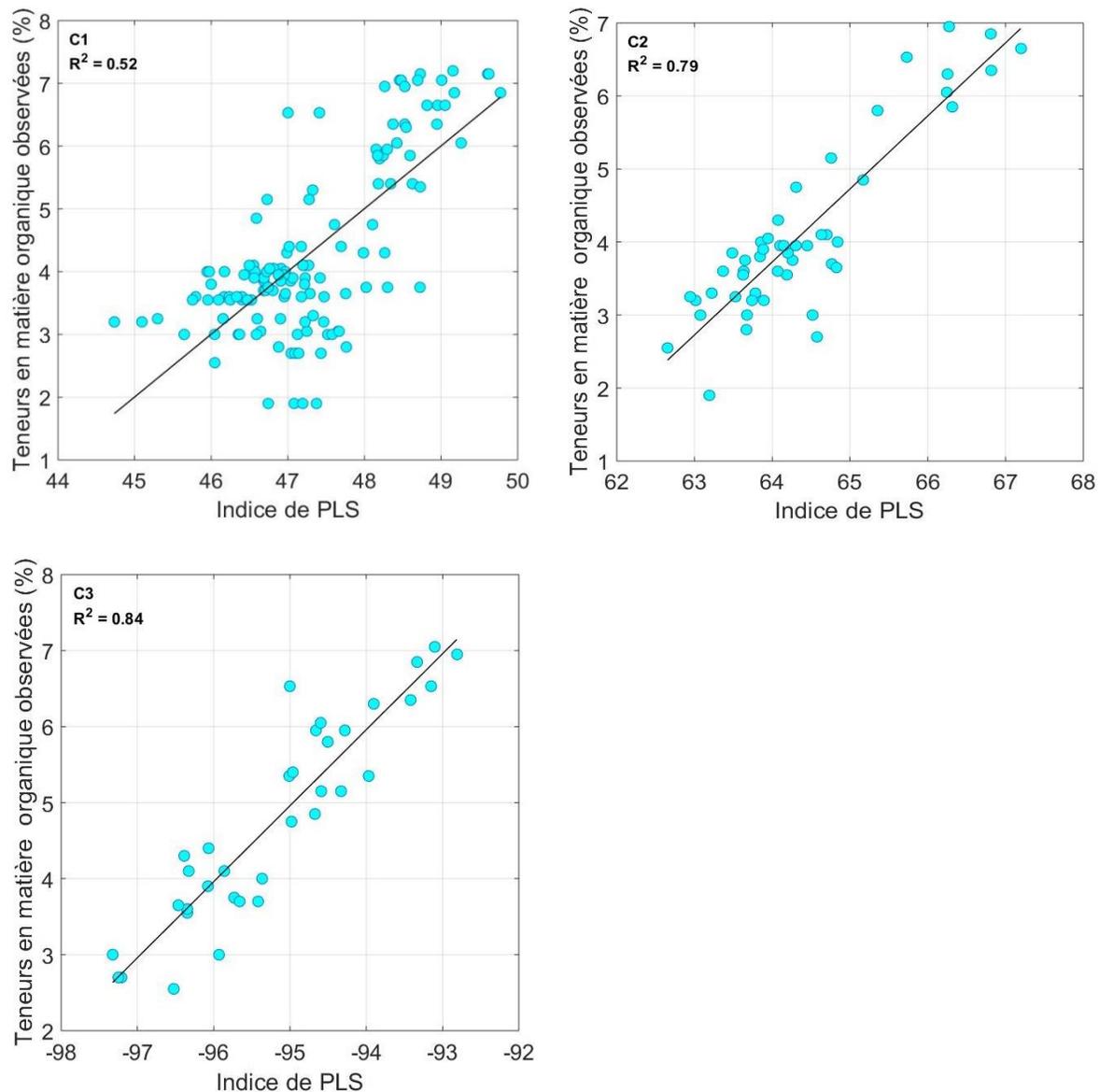


Figure 4.22 : Valeurs de teneur en matière organique observées en fonction des coefficients issus de l'application de la PLS pour les trois classes d'humidité de sol C1 (0-10 %), C2 (10-19 %) et C3 (>19 %).

Les parties suivantes de cette étude visaient l'application de ces modèles établis sur de nouveaux échantillons pour l'évaluation de la qualité des modèles calibrés et la comparaison entre eux en utilisant les coefficients de détermination R^2 , le RMSE de prédiction, le biais et le RPD. De plus, les longueurs d'onde ayant de faibles corrélations avec la propriété du sol étudiée supprimées des spectres ont été aussi supprimées de sous-ensemble de validation. Enfin, il est à souligner que ce sous-ensemble a été décomposé en deux groupes : un groupe

test pour la quantification du biais et un pour appliquer les modèles après l'ajustement de ce paramètre. Chaque groupe a été aussi subdivisé en trois selon les seuils d'humidité du sol fixés.

4.3.3.4 Validation des modèles d'estimation de la teneur en eau et des teneurs du sol en argile, sable et matière organique

- **Sans décomposition de la base des données de calibration**

Selon l'annexe 4, les plages des valeurs observées des propriétés étudiées pour le sous-ensemble de validation ne présentent pas également d'écarts importants entre une classe d'humidité du sol et une autre. Après subdivision de sous-ensemble de validation en deux, le nombre d'échantillons devient 22, 8 et 6 respectivement pour les classes C1, C2 et C3.

Lors de l'application des modèles d'estimation des propriétés de sol sur la moitié des données de validation, un biais systématique a été observé. Ceci peut être expliqué par l'effet de la distance entre la sonde et les particules du sol. En effet, la diminution de cette distance augmente l'albédo général des spectres sans affecter les pics d'absorption caractéristiques de chaque propriété de sol. Cet effet sera examiné plus en détail dans la partie suivante afin d'étudier la classe d'humidité du sol la plus impactée par ce changement ainsi que l'effet de l'ajustement du biais.

La comparaison entre les Figures 4.23, 4.24, 4.25 et 4.26 avant et après ajustement du biais, révèle l'amélioration de la performance des modèles d'estimation des propriétés de sol étudiées après diminution du biais. La surestimation de la teneur en eau a été corrigée en augmentant le biais qui a passé de -10.99 % pour les échantillons test à -1.11 % pour les échantillons de validation (Figure 4.23). Par conséquent, la précision du modèle de son estimation est devenue très bonne ($R^2 = 0.91$ et $RPD = 3.13$). L'intensité de la réponse spectrale de l'eau est très forte bien qu'on a utilisé une partie du proche infrarouge seulement (900.92-1716,85 nm) alors que le pic d'absorption principal de l'eau se situe autour de 1900 nm (Weidong et al., 2002). En comparant ces résultats avec ceux trouvés par Morellos et al. (2016) ($R^2 = 0.71$ et $RPD = 1.86$) qui ont acquis des spectres dans le domaine (305-2200 nm), on peut dire que le domaine utilisé dans cette étude permet l'estimation de la teneur en eau avec précision. Le modèle calibré dans cette étude pour quantifier l'humidité du sol présente par contre une limite d'estimation des faibles valeurs de cette propriété (voir des valeurs nulles).

Quant aux variables texturales et organique du sol, les modèles de quantification des teneurs en argile et sable présentent des bonnes précisions, après correction du biais, par rapport à

celui de la matière organique avec des R^2 égaux à 0.67 et 0.69 et RPD égaux à 1.76 et 1.66 respectivement pour les teneurs en argile et sable (Figures 4.24 et 4.25) comparés à R^2 et RPD égaux à 0.37 et 1.25 respectivement pour la teneur en matière organique (Figure 4.26). Ceci peut être causé par la différence de corrélation entre chacune des trois propriétés en question et les données spectrales (la force du lien entre la teneur en matière organique et les longueurs d'onde est la plus faible selon les Figures 4.9, 4.10 et 4.11. De plus, Les échantillons utilisés dans cette étude sont caractérisés par des teneurs en sable très importantes et pourcentages d'argile supérieurs aux pourcentages de matière organique selon les statistiques descriptifs présentés dans le tableau 4.1). Selon Stenberg et al. (2010), Les signaux de la matière organique dans les domaines du visible et proche infrarouge sont faibles en particulier dans des sols contenant de faibles quantités de matière organique et de fortes teneurs en matières minérales (argile et sable).

L'étude de l'effet de la classification des échantillons de sol en classes selon le niveau de teneur en eau sur la précision d'estimation des composantes de la texture, ainsi que la comparaison de performance inter classe sera examiné dans le paragraphe suivant.

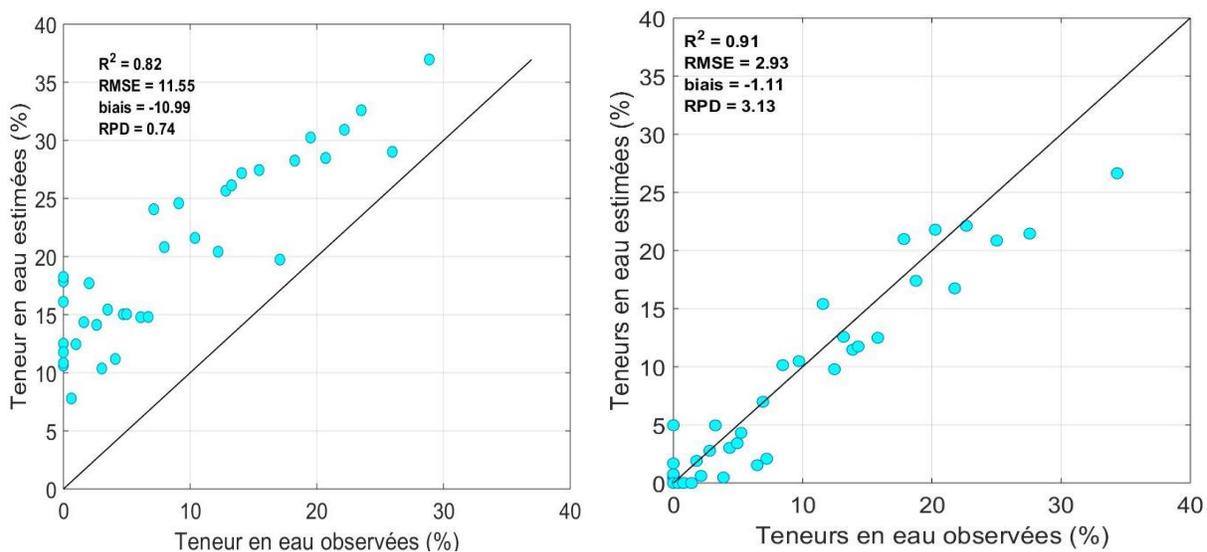


Figure 4.23 : Teneurs en eau estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite), (RMSE et le biais sont en %).

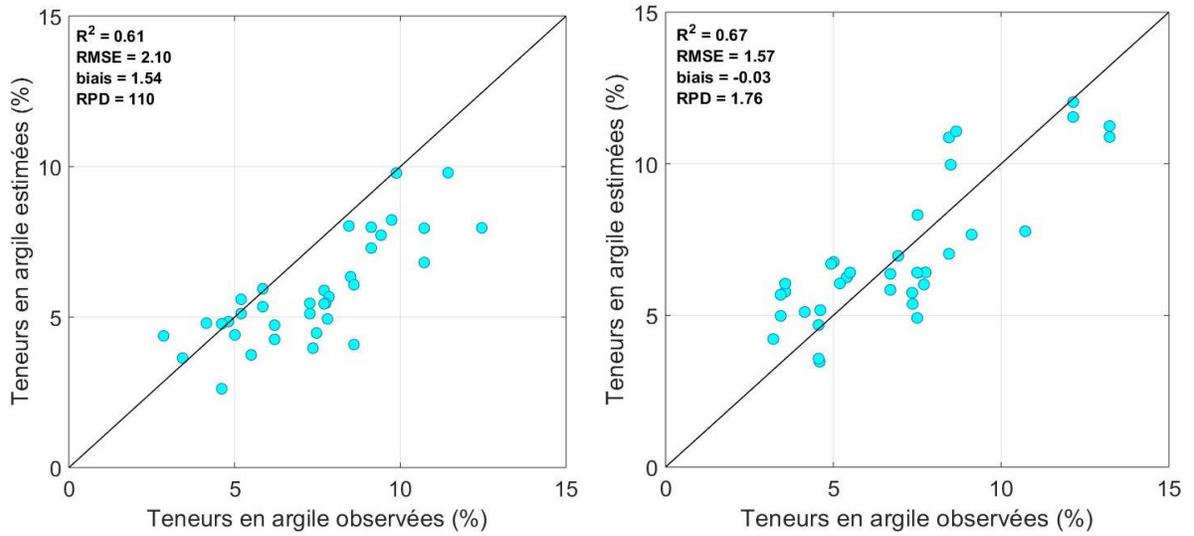


Figure 4.24 : Teneurs en argile estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite), (RMSE et le biais sont en %).

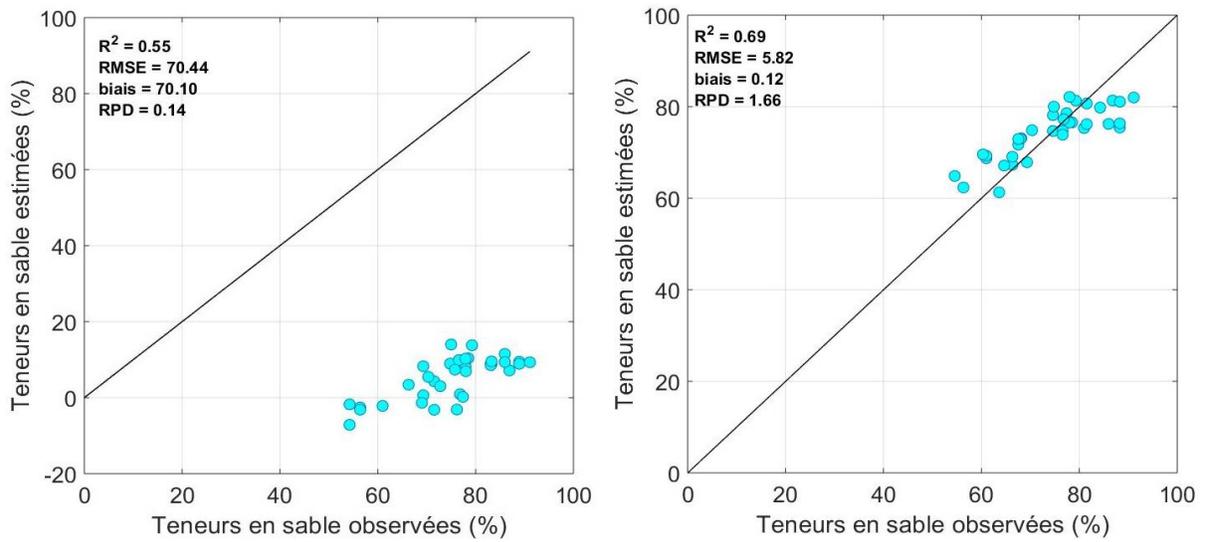


Figure 4.25 : Teneurs en sable estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite), (RMSE et le biais sont en %).

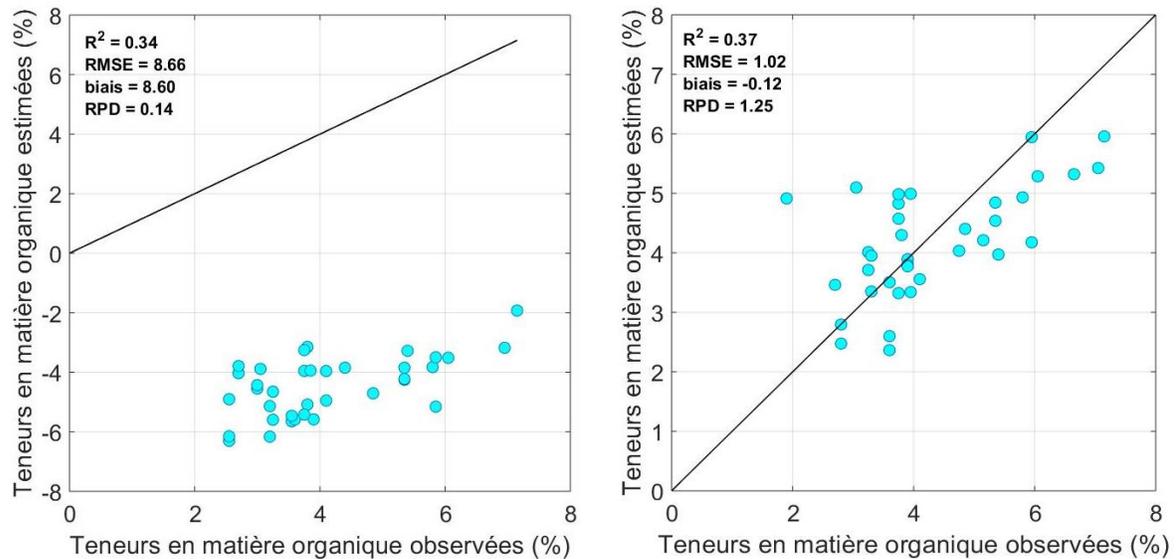


Figure 4.26 : Teneurs en matière organique estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite), (RMSE et le biais sont en %).

- **Après décomposition de la base des données de calibration**

L'importance de l'effet du biais dans la classe C1 (la plus sèche) pour les teneurs en sable et matière organique (4.33 et 4.30) et attendue puisque la diminution de la distance sonde-surface d'échantillon est plus poussée lorsque les échantillons sont secs. Cependant, la valeur du biais est importante dans la classe C2 pour la teneur en argile (Figure 4.28). Pour ce niveau d'humidité (10-19 %), la structure du sol change considérablement surtout pour les échantillons les plus argileux. En effet, au fur et à mesure que les échantillons seront chauffés, la structure des échantillons argileux devient compacte. Pour cette raison, on a essayé de dissocier les mottes formées, ce qui a créé des éléments structuraux compacts plus petits séparés par des espaces vides. Ce changement de structure et ces pores formés chez les échantillons de teneurs en argile relativement importantes ont changé les spectres et ont affecté ainsi la précision du modèle d'estimation de la teneur en argile pour la classe C2.

Selon la Figure 4.27, la classification des spectres selon le niveau de teneur en eau a amélioré la précision d'estimation de la teneur en argile pour la classe C1. Sans classification des échantillons, les valeurs calculées de R^2 et RPD sont respectivement 0.67 et 1.76 et après classification des spectres, R^2 obtenu est 0.76 et RPD est 2.09 pour la classe C1. En effet, les études menées par Ben-dor et Banin (1995) et Madejov et Komadel (2001) ont prouvées que les bandes correspondantes à l'absorption des groupements OH de l'eau ont des contributions

importantes des modèles d'estimation de la teneur en argile. De plus, les pics d'absorption de l'eau se trouvant autour de 1400 nm (figure 2.5) renseignent aussi sur l'ion hydroxyle qui peut être lié aux minéraux d'argile (Clark, 1993 ; Clark et al., 1990). Pour cela, le groupement des spectres en des classes selon la teneur en eau diminue son interférence et améliore ainsi l'estimation de la teneur en argile. La performance du modèle à estimer la quantité d'argile a baissé remarquablement pour la classe C2 en comparaison avec cette diminution pour la classe C3. Ceci rejoint les constatations de changement de structure qui augmente le biais.

En ce qui concerne la teneur en sable, la précision d'estimation de cette propriété diminuait autant que l'humidité augmente après correction du biais (Figures 4.30, 4.31 et 4.32). Seule la classe C3 (représentant des échantillons d'humidité supérieure à 19 %) présente une amélioration de qualité de modèle après décomposition des spectres (Figure 4.32) (R^2 et RPD ont passé de 0.69 et 1.66 respectivement avant décomposition des spectres à 0.81 et 1.86 respectivement).

Les classes dont la qualité de l'estimation de la teneur en matière organique a été améliorée après décomposition des spectres sont C1 et C2. R^2 et RPD ont passé de 0.37 et 1.25 en utilisant tous les échantillons pour l'estimation de la matière organique (Figure 4.26) à 0.55 et 1.4 (Figure 4.33) et 0.81 et 2.30 (Figure 4.34) respectivement pour les classes C1 et C2 après correction du biais. En ajoutant l'analyse des corrélogrammes de cette propriété avec les longueurs d'onde à cette observation, on peut dire que l'intensité du lien entre les données spectroscopiques et la matière organique pour la classe C2 présente de fortes valeurs par rapport à C1 et C3 surtout dans le domaine du proche infrarouge (Figure 4.11). Pour les échantillons du sol les plus humides (C3), le modèle d'estimation de la teneur en matière organique a subi une diminution de performance. En effet, l'eau agit en altérant le signal à cause de ses interactions avec d'autres composés de la matière organique (Reeves, 2010). La validation des modèles d'estimation du carbone organique construits à partir d'échantillons secs sur des échantillons d'humidité croissante dans une étude menée par Nocita et al. (2013), augmente l'erreur du modèle avec le niveau de la teneur en eau. Ceci plaide en faveur des résultats trouvés dans cette étude bien que dans notre cas, la classe C2 (représentant les échantillons moyennement humides) présente les meilleurs résultats par rapport à C1 et C3. Ceci est attendu puisque, comme pour le sable et l'argile, la différence importante entre les tailles des trois classes limite les résultats trouvés.

Globalement, la décomposition de la base des données selon la teneur en eau améliore la qualité d'estimation des variables texturales et organique au moins pour deux classes

d'humidité sauf pour le sable où seulement la classe C3 présente une augmentation de précision de l'estimation de cette propriété après ajustement du biais. Toutefois, il faut tenir compte des différences importantes des tailles des trois classes. Les classes C2 et C3 contiennent respectivement 8 et 6 échantillons comparées à la classe C1 contenant 22 échantillons. Par conséquent, les résultats observés dans le cas de la classe C1 reflète plus la performance réelle des modèles calibrés puisque la tendance sera visible et plus facilement interprétée pour un nombre important d'échantillons, alors qu'une taille faible des données de validation implique un alignement aléatoire fortement dépendant de la plage des valeurs de ces échantillons et non pas de la précision du modèle.

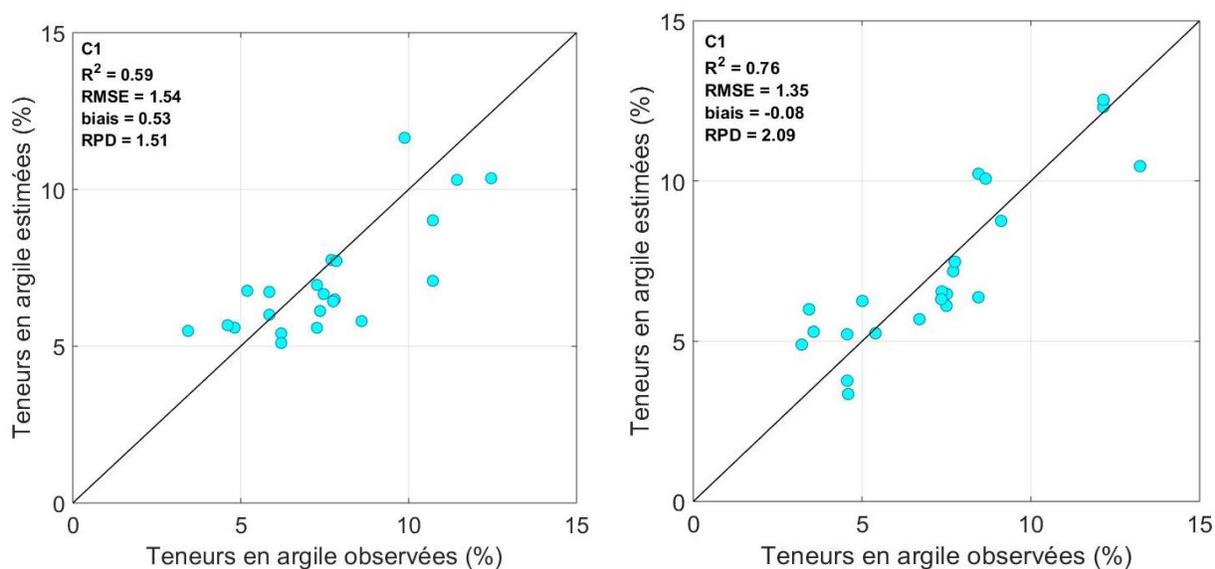


Figure 4.27 : Teneurs en argile estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite) pour la classe C1 (0-10 %), (RMSE et le biais sont en %).

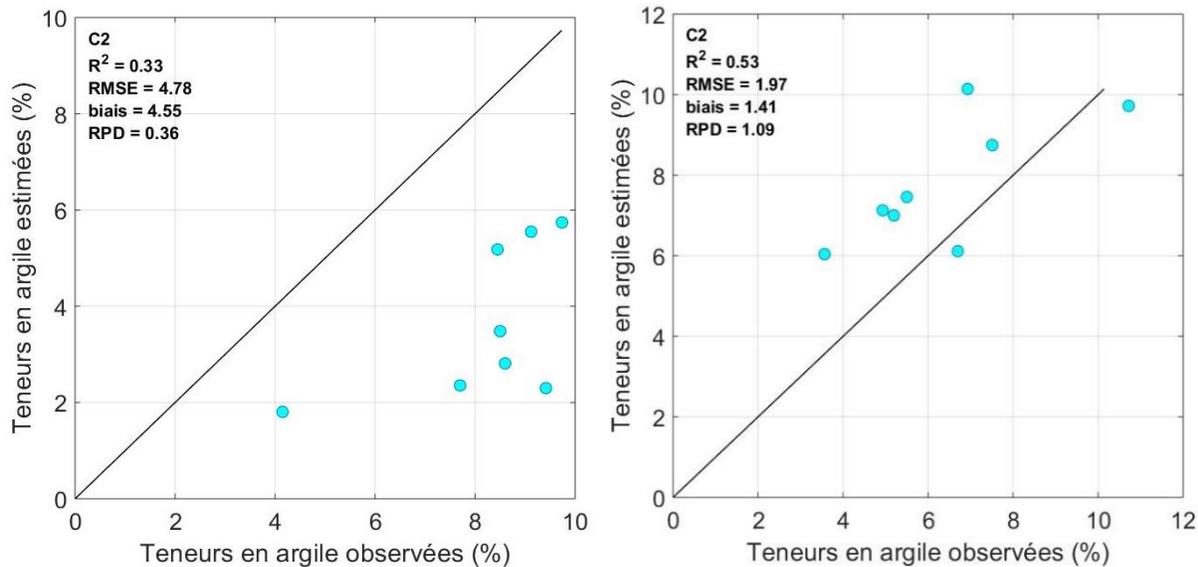


Figure 4.28 : Teneurs en argile estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite) pour la classe C2 (10-19 %), (RMSE et le biais sont en %).

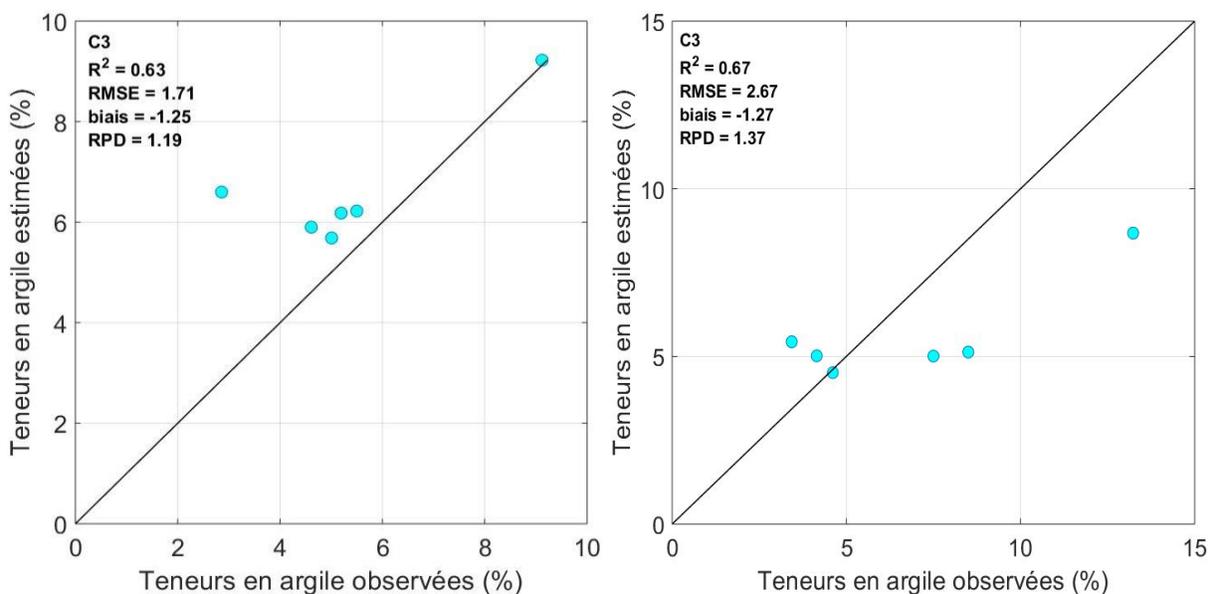


Figure 4.29 : Teneurs en argile estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite) pour la classe C3 (>19 %), (RMSE et le biais sont en %).

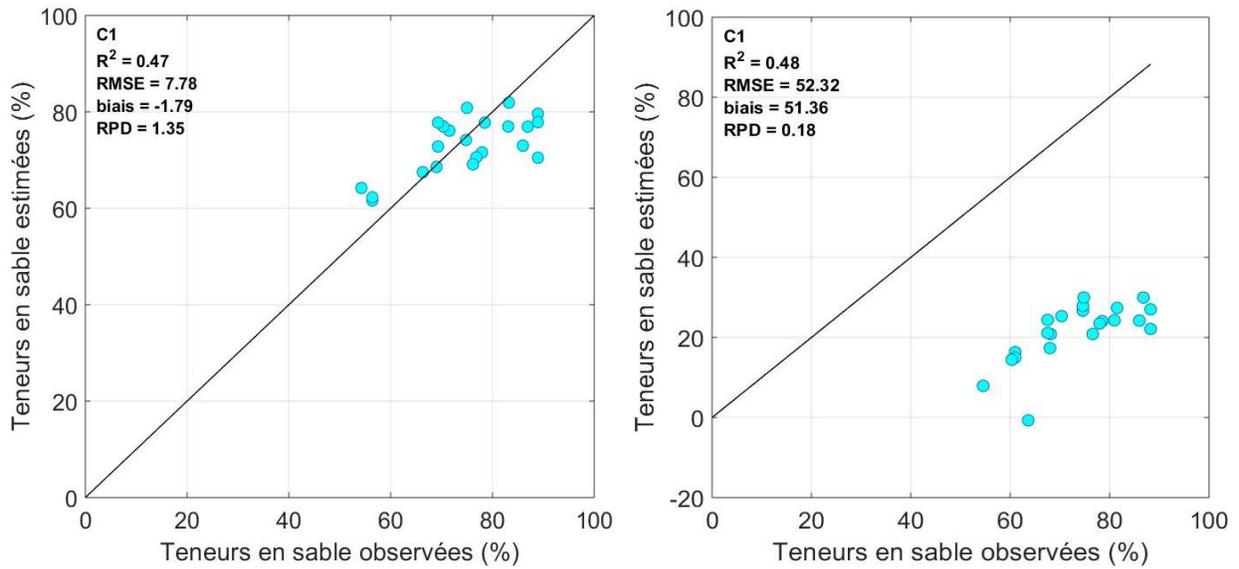


Figure 4.30 : Teneurs en sable estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite) pour la classe C1 (0-10 %), (RMSE et le biais sont en %).

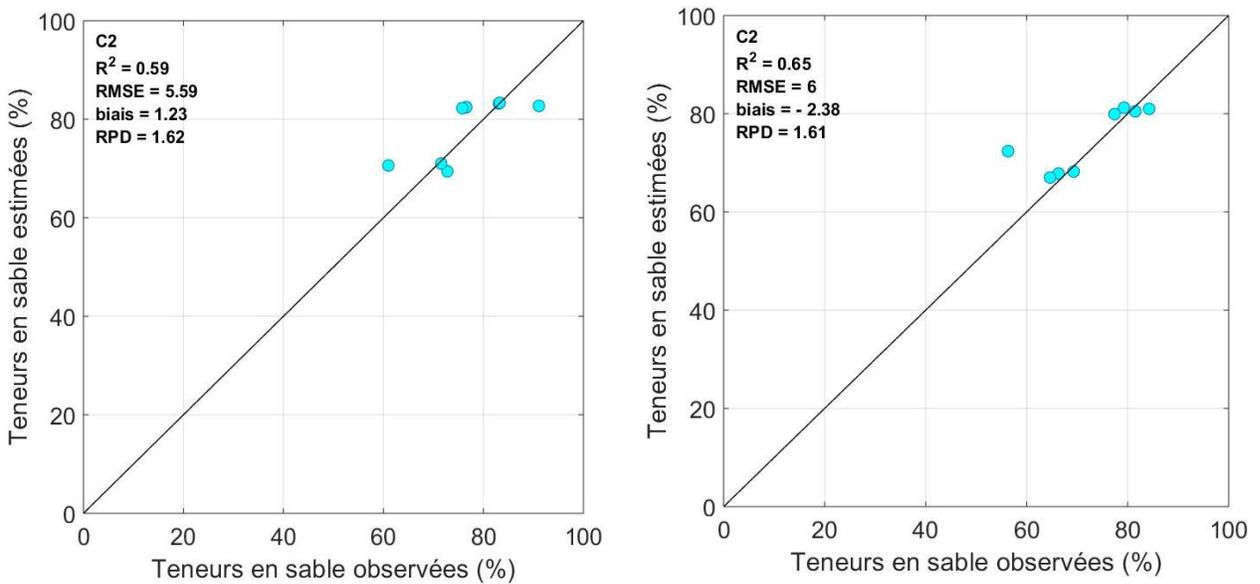


Figure 4.31 : Teneurs en sable estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite) pour la classe C2 (10-19 %), (RMSE et le biais sont en %).

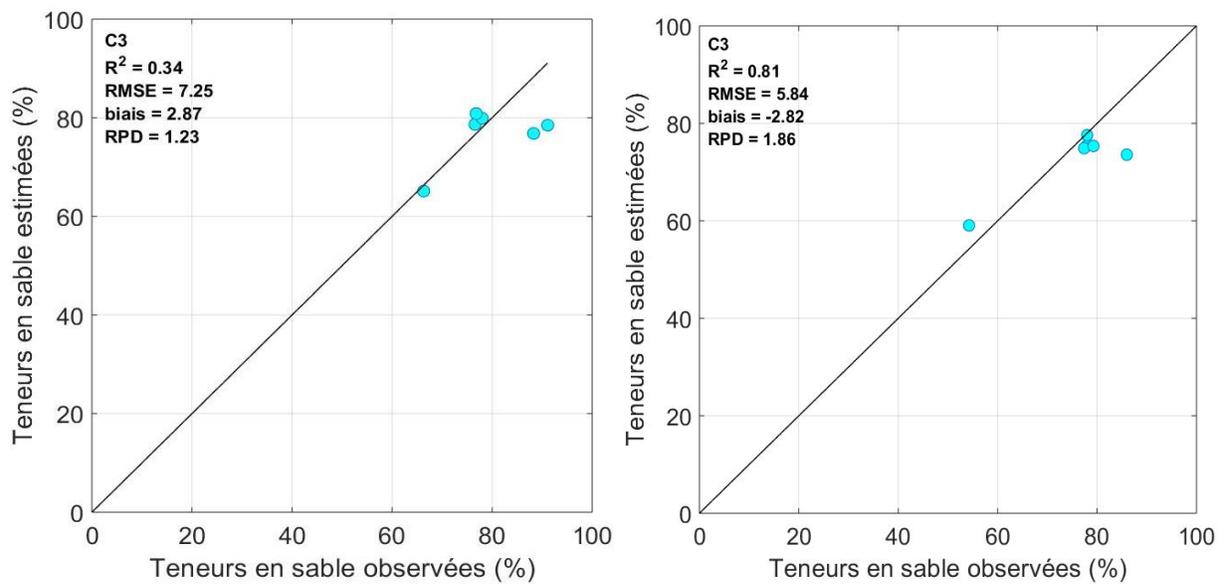


Figure 4.32 : Teneurs en sable estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite) pour la classe C3 (>19 %), (RMSE et le biais sont en %).

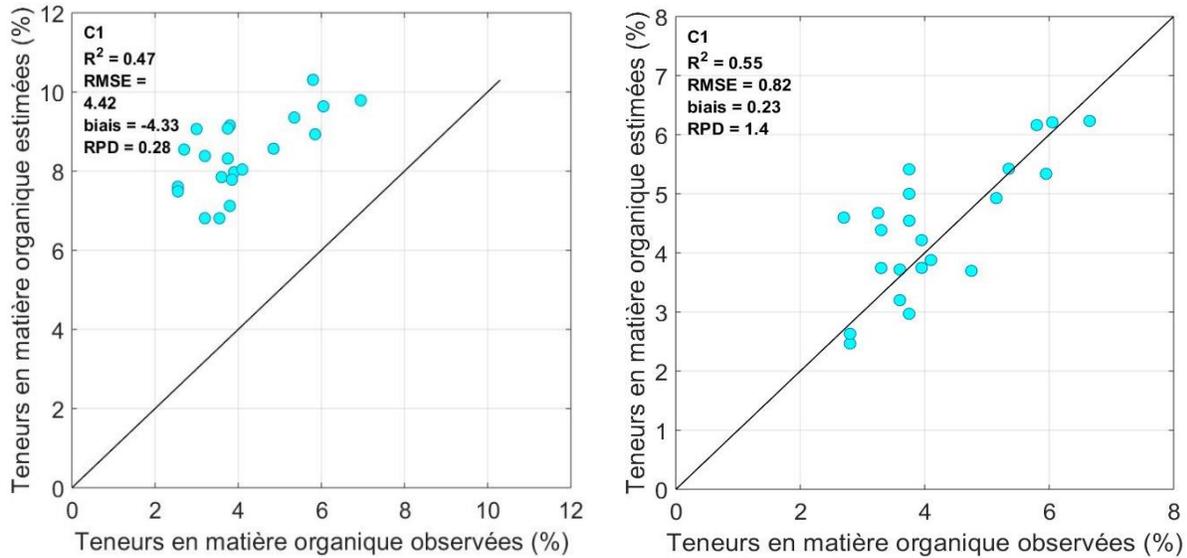


Figure 4.33 : Teneurs en matière organique estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite) pour la classe C1 (0-10 %), (RMSE et le biais sont en %).

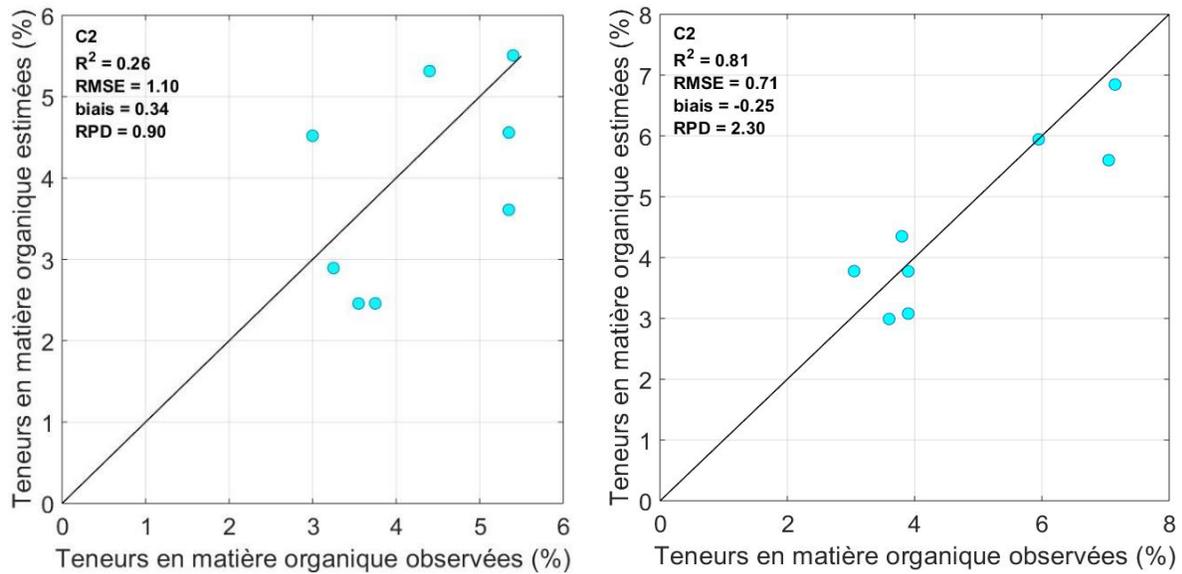


Figure 4.34 : Teneurs en matière organique estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite) pour la classe C2 (10-19 %), (RMSE et le biais sont en %).

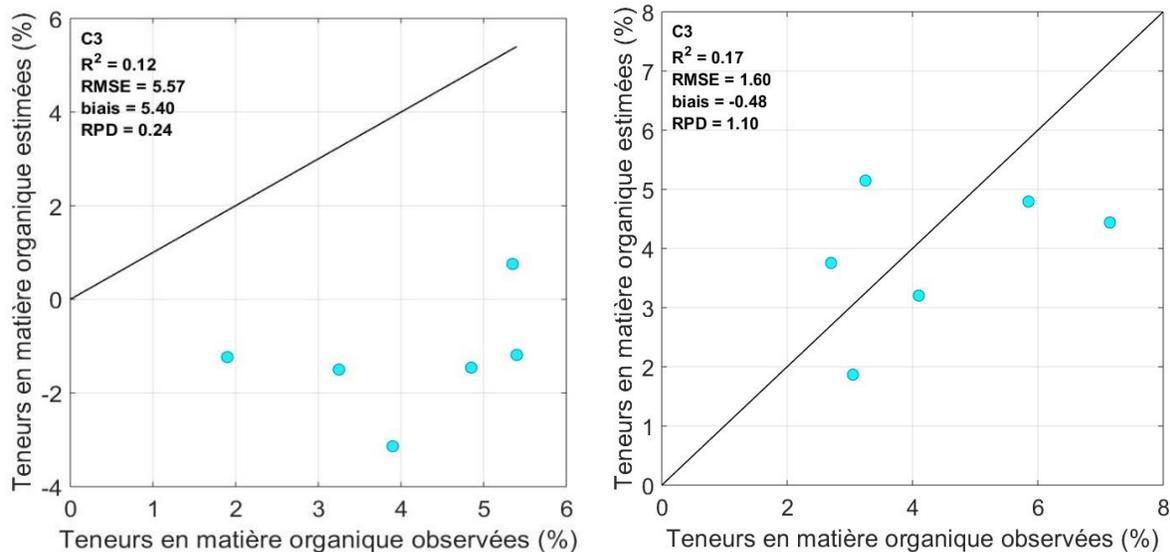


Figure 4.35 : Teneurs en matière organique estimées en fonction de celles observées avant (à gauche) et après soustraction du biais (à droite) pour la classe C3 (>19 %), (RMSE et le biais sont en %).

4.3.3.5 Sources d'erreurs

L'étude des sources d'erreurs ayant une influence potentielle sur les résultats trouvés peut un bon point de départ pour des travaux futurs. Les perturbations possibles ont été subdivisées en quatre : des erreurs liées au protocole expérimental, des erreurs liées aux propriétés initiales des échantillons de sol, des erreurs liées au matériel et des erreurs associées aux traitements des spectres.

- **Sources d'erreurs associées au protocole expérimental**

Les perturbations majeures ayant lieu lors du protocole expérimental ont touchées les propriétés de sol étudiées. La construction des échantillons composites a imposé l'utilisation d'une partie de chacun des échantillons unitaires. Par conséquent, ces parties peuvent ne pas représenter les mesures réelles des propriétés de chaque échantillon unitaire. De surcroît, les valeurs observées utilisées pour l'évaluation des modèles d'estimation des teneurs en argile, sable et matière organique du sol, sont les moyennes de chaque deux échantillons unitaires formant l'échantillon composite, utilisé pour l'acquisition des spectres et la calibration des modèles. Ceci pouvait avoir un effet sur les indicateurs de performances des modèles. D'autre part, le séchage des échantillons du sol à teneurs relativement importantes en argile a provoqué un compactage de la surface de sol. Lors de l'acquisition des spectres, les mottes

formées ont été décomposées ce qui a créé du vide pouvant avoir une influence sur l'intensité de l'énergie arrivée aux particules du sol et par la suite, sur les spectres acquis.

- **Sources d'erreurs liées aux propriétés initiales des échantillons de sol**

Il est à souligner que les échantillons contiennent des pourcentages importants de matière minérale (sable et argile) comparativement à la matière organique. Ceci abaisse possiblement la précision d'estimation de la matière organique. L'utilisation d'échantillons contenant de teneurs plus importantes en cette dernière peuvent valider ou non cette hypothèse. Afin de construire des modèles à forte reproductibilité, l'utilisation d'un sous-ensemble de validation provenant d'un autre champ est fortement recommandée.

- **Sources d'erreurs liées au matériel**

En ce qui concerne le matériel utilisé, on présume que l'acquisition des spectres dans une partie du proche infrarouge représentait une limite à l'estimation des propriétés étudiées puisqu'il existe des pics d'absorption intéressants dans le reste du domaine proche infrarouge (1700-2500 nm) pouvant augmenter la corrélation entre les longueurs d'onde et les caractéristiques de sol étudiées. En outre, l'utilisation d'un spectromètre avec une sonde de contact a provoqué un biais systématique chez les quatre propriétés en question. Il existe toutefois d'autres modèles de spectromètres faisant intervenir peu de contact avec les particules de sol.

- **Sources d'erreurs associées aux traitements des spectres**

Quant à la méthodologie adoptée pour le développement d'une approche adaptative d'estimation des variables texturale et organique du sol, la source d'incertitude principale est le nombre d'échantillons. Dans cette étude, la création de classes de teneurs en eau a été basée sur l'homogénéisation intra classe des spectres. Puisque pour chaque échantillon du sol, l'atteinte de l'état sec est certaine (humidité du sol égale à 0), la classe contenant d'échantillons à teneurs en eau faibles est de taille plus importante que les deux autres classes pour les sous-ensembles de calibration et de validation. Pour cela, l'ajout de teneurs en eau plus importantes est conseillé pour la construction de classes équilibrées en matière de nombre d'échantillons.

5 CONCLUSION

L'acquisition de spectres des échantillons du sol provenant de deux régions de Québec et ayant des caractéristiques texturale et organique différentes ainsi que différentes teneurs en eau a permis de développer une approche adaptative d'estimation des teneurs en argile, sable et matière organique. Ceci a été effectuée à travers l'étude qualitative et quantitative de l'effet de la teneur en eau sur l'estimation des teneurs en argile, sable et matière organique. Selon plusieurs travaux, l'humidité du sol est un facteur environnemental difficile à gérer à cause de son effet sur plusieurs propriétés de sol, tel que la teneur en carbone organique (Nocita *et al.*, 2013 ; Minasny *et al.*, 2011) et la teneur en azote total (Xiaofei *et al.*, 2015). L'analyse qualitative des spectres du sol groupés par niveaux de teneur en eau et des corrélations entre les trois paramètres du sol en question et des longueurs d'onde a abouti aux conclusions suivantes :

- L'albédo diminue non linéairement avec la teneur en eau. En supplément, l'augmentation de l'humidité du sol diminue en largeur et en profondeur le pic d'absorption caractéristique de l'eau se trouvant autour de 1400 nm ;
- La difficulté de distinction entre les spectres à partir des valeurs d'humidité du sol qui les correspondent augmente autant que les niveaux de cette propriété augmentent ;
- L'intensité du lien entre les teneurs du sol en argile, sable et matière organique et les données spectroscopiques change avec le niveau d'humidité du sol. Notamment, elle augmente pour des échantillons dont la teneur en eau est comprise entre 10 et 19 % dans le cas de la matière organique.

L'estimation des variables texturale et organique du sol en calibrant des modèles de régression PLS pour toute la base des données puis pour les trois classes de teneurs en eau choisies a conduit aux conclusions suivantes :

- Les données spectrales sont caractérisées par un biais systématique résultant du changement de l'albédo général des spectres engendré par la variation de la distance entre la sonde du spectromètre et la surface de contact avec le sol. Ceci peut être évité dans des prochaines études si on utilise un spectromètre qui n'implique pas un contact avec l'échantillon du sol ;
- La réponse spectrale de l'eau est très forte dans le domaine étudié (431.72 nm - 1716,849 nm), puisque le modèle construit pour son estimation a présenté une forte

performance ($R^2 = 0.91$ et $RPD = 3.13$). Toutefois, la teneur en eau des échantillons secs n'a pas été bien estimée par ce modèle ;

- La classification des spectres en groupes selon les valeurs de teneur en eau des échantillons du sol a ajouté plus de précision aux modèles pour les échantillons dont l'humidité du sol est entre 0-10 % et supérieure à 19 % lors de l'estimation de la teneur en argile et entre ceux ayant des teneurs en eau entre 0-10 % et 10-19 % lors de l'estimation de la matière organique. Alors que pour la teneur en sable, seul le modèle calibré sur des échantillons avec un niveau d'humidité supérieur à 19 % présente des résultats meilleurs que celui calibré sur tous les échantillons ;
- Le modèle d'estimation de la teneur en argile le plus performant a été observé chez les échantillons du sol les plus secs ($R^2 = 0.76$ et $RPD = 2.09$) alors que celui d'estimation de la teneur en sable a été observé pour des échantillons dont l'humidité de sol est supérieure à 19 % ($R^2 = 0.81$ et $RPD = 1.86$). Quant à la teneur en matière organique, les échantillons du sol dont l'humidité est entre 10-19 % présentent le modèle le plus précis ($R^2 = 0.81$ et $RPD = 2.30$).

Cette première expérience de laboratoire prouve que l'humidité du sol est un facteur limitant l'utilisation de spectroscopie pour la quantification des teneurs en argile et matière organique contrairement à la teneur en sable. Pour cela le développement d'une approche qui soit adaptative en prenant les effets interférents de l'eau est recommandé pour ajouter de la précision à l'estimation des paramètres étudiés. Les résultats d'estimation des quatre propriétés de sol étudiées sont satisfaisants avec une précision moins bonne pour la matière organique causée probablement par les propriétés initiales de sol (fortes teneurs en sable et argile par rapport à la matière organique) (Stenberg et al., 2010). Toutefois, comme toutes méthodes empiriques, les résultats dépendent en grande partie aux échantillons de sol utilisés dans cette étude. Par conséquent, les résultats ne peuvent pas être généralisés alors que la méthodologie, prouvée efficace, peut être appliquée sur d'autres échantillons, si on prend en considération les sources d'erreurs de cette étude. Par exemple, la possibilité de changement de la texture lors de l'acquisition des spectres (puisque une partie seulement des échantillons de sol a été utilisée pour l'analyse) et le changement de la structure du sol au fur et à mesure que sa teneur en eau diminue, ainsi que la construction d'échantillons composites pourraient avoir un effet sur les résultats trouvés. De plus, le nombre inégal des échantillons pour chaque classe d'humidité du sol limite la reproductibilité des modèles. Pour cela, on suggère un contrôle de ces effets perturbateurs en assurant une homogénéité plus importante des propriétés étudiées pour

chaque échantillon par l'utilisation d'échantillons unitaires dans des prochaines études (et non pas composites). En outre, la création de classes d'humidité de tailles presque égales ainsi que l'acquisition de spectres dans tout le domaine du visible-proche infrarouge (400-2500 nm) améliorera certainement la précision des modèles et leur applicabilité. L'extraction des informations pour l'estimation des propriétés de sol à partir des spectres est une méthode orientée données, il est alors préférable d'augmenter le nombre d'échantillons pour des études futures.

Pour augmenter l'efficacité de l'approche adaptative, on suggère un découpage de la base de données en utilisant plus de critères de sélection pour diminuer l'incertitude dans les zones de transition entre une classe et une autre et éviter l'effet escalier des seuils avec des limites non continues tel que proposé par El-Alem et al. (2019).

Ainsi, on peut programmer un spectromètre opérant sur le terrain capable de tenir compte de l'effet de l'humidité du sol sur la précision de quantification des propriétés de sol. De cette manière, les méthodes usuelles de détermination des propriétés de sol pourront être remplacées par la spectroscopie si on arrive à améliorer la performance des modèles calibrés sur des échantillons non prétraités et qui représentent bien la variabilité spatiale de la teneur en sable, limon et argile de la texture de même que le teneur en matière organique.

6 BIBLIOGRAPHIE

Adamchuk, V. I., Hummel, J. W., Morgan, M.T. et Updhayaya, S. K. (2004) 'On-The-Go Soil Sensors for Precision Agriculture', Elsevier, 44, pp. 71–91. doi: 10.1016/j.compag.2004.03.002.

Adamchuk, V. I., Morgan, M. T. et Ess, D. R. (1999) 'An Automated Sampling System for Measuring Soil pH', American Society of Agriculture Engineers, 42(765), pp. 885–891. doi: 10.13031/2013.13268.

Adeline, K. R., Gomez, C., Goretta, N. et Roger, J. M., . (2017) 'Predictive Ability of Soil Properties to Spectral Degradation from Laboratory Vis-NIR Spectroscopy Data', Geoderma. Elsevier B.V., 288, pp. 143–153. doi: 10.1016/j.geoderma.2016.11.010.

Araujo, S. R., Wetterlind, J., Dematte, A. M. et Stenberg, B. (2014) 'Improving the Prediction Performance of a Large Tropical Vis-NIR Spectroscopic Soil Library from Brazil By Clustering Into Smaller Subsets or Use of Data Mining', European Journal of Science, (September), pp. 718–729. doi: 10.1111/ejss.12165.

Arriaga, J. F. et Lowery, B. et Mays, M. (2006) 'A Fast Method for Determining Soil Particle Size Distribution Using a Laser Instrument', Soil Science, 171(9), pp. 663–674. doi: 10.1097/01.ss.0000228056.92839.88.

ASTMD (2016) 'Standard Test Methods for Laboratory Determination of Moisture (Moisture) Content of Soil', ASTM International, West Conshohocken, PA, 2008.

Bach, H. et Mauser, W. (1994) 'Modelling and Model Verification of The Spectral Reflectance of Soils under Varying Moisture Conditions', International Geoscience and Remote Sensing Symposium (IGARSS), 4(1 988), pp. 2354–2356. doi : 10.1109/IGARSS.1994.399735.

Barnes, J. R., Dhanoa, M. S. et Lister, S. J. (1989) 'Standard Normal Variate Transformation and De-trending of Near-infrared Diffuse Reflectance Spectra', Applied Spectroscopy, 43, pp. 772–777. doi : 10.1366/0003702894202201.

Baumgardner, M. F., Silva, L. R, Biehl, L. L. et Stoner, E. R. (1985) 'Reflectance Properties of Soils', 38, pp. 1–44. doi : 10.1016/S0065-2113(08)60672-0.

Bellon-maurel, V. et Mcbratney, A. (2011) 'Near-Infrared (NIR) and Mid-Infrared (MIR) Spectroscopic Techniques for Assessing The Amount of Carbon Stock in Soils-Critical Review and Research Perspectives', Soil Biology and Biochemistry. Elsevier Ltd, 43(7), pp. 1398–1410. doi: 10.1016/j.soilbio.2011.02.019.

- Ben-dor, E. et Banin, A. (1995) 'Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties', *Soil Science Society of America*, (59), pp. 364-372. doi : 10.2136/sssaj1995.03615995005900020014x.
- Bergveld, P. (1991) 'Future Applications of ISFETs*', *Sensors and Actuators*, (4), pp. 125–133. doi : 10.1016/0925-4005(91)80187-O.
- Birrell, S. J. et Hummel, J. W. (2000) 'Membrane Selection and ISFET Configuration Evaluation for Soil Nitrate Sensing Membrane Selection and ISFET Configuration Evaluation for Soil Nitrate', *Agriculture and Biosystems engineering*, (43), pp. 197–206. doi: 10.13031/2013.2694.Rights.
- Birrell, S. J. et Hummel, J. W. (2001) 'Real-time Multi ISFET / FIA Soil Analysis System with Automatic Sample Extraction', *Elsevier*, 32, pp. 45–67. doi : 10.1016/S0168-1699(01)00159-4.
- Blanco, M. et Villarroya, I. (2002) 'NIR Spectroscopy: A Rapid-Response Analytical Tool', *trends in analytical chemistry*, 21(4), pp. 240–250. doi : 10.1016/S0165-9936(02)00404-1.
- Blystone, J. R., Pelzner, A. et Steffens, G. P. (1962) 'Moisture Content Determination by The Calcium Carbide Gas Pressure Method', *Highway Research Board Bulletin*, (4), pp. 77–84.
- Bokobza, L. (1998) 'Near infrared spectroscopy', *Near Infrared Spectroscopy*, 6, pp. 3–17. doi : 10.1255/jnirs.116.
- Boone, R. D. et Grigal, D. F., Sollins, P., Ahrens, R. J., Armstrong, D. E. (1999) 'Soil Sampling, Preparation, Archiving and Quality Control', pp. 3–28. <https://andrewsforest.oregonstate.edu/sites/default/files/lter/pubs/pdf/pub2706.pdf>.
- Bowers, S. A., Hanks, R. J. (1965) 'Reflection of Radiant Energy from Soils', *Soil Science*, 100 (2), pp. 130–138. <https://www.ars.usda.gov/ARSUserFiles/30200525/870%20Reflection%20of%20radiant%20energy%20from%20soils.pdf>.
- Brabant, P. (2008) 'Activités humaines et dégradation des terres Indicateurs et indices'. https://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers15-05/010045188.pdf.
- Brown, D. J., Shephered, K. D., Walsh, M. G. et Mays, M. D. (2006) 'Global Soil Characterization with VNIR Diffuse Reflectance Spectroscopy', *Geoderma*, 132, pp. 273–290. doi: 10.1016/j.geoderma.2005.04.025.

- Brown, D. J. (2007) 'Using a Global VNIR Soil-Spectral Library for Local Soil Characterization and Landscape Modeling in a 2nd-order Uganda Watershed', *Geoderma*, 140, pp. 444–453. doi: 10.1016/j.geoderma.2007.04.021.
- Brunet, D., Barthès, B. G., Chotte, J. L. et Feller, C. (2007) 'Determination of Carbon and Nitrogen Contents in Alfisols, Oxisols and Ultisols from Africa and Brazil Using NIRS Analysis: Effects of Sample Grinding and Set Heterogeneity', *Geoderma*, 139(1–2), pp. 106–117. doi: 10.1016/j.geoderma.2007.01.007.
- Ceddia, M., Clingensmith, C. et Grunwald, S. (2017) 'Prediction of Soil Physical and Chemical Properties by Visible and Near-Infrared Diffuse Reflectance Spectroscopy in the Central Amazon', (March). doi: 10.3390/rs9040293.
- Chang, C., Laird, D. et Mausbach, M. J. (2001) 'Near-Infrared Reflectance Spectroscopy – Principal Components Regression Analyses of Soil Properties Near-Infrared Reflectance Spectroscopy – Principal Components', *Soil Science Society of America*, 65, pp. 480-490. doi: 10.2136/sssaj2001.652480x.Rights.
- Clark, R. N. King, T. V. et Klejwa, M. (1990) 'High Spectral Resolution Reflectance Spectroscopy of Minerals', *Journal of Geophysical research*, 95(89), pp. 12653-12680. doi : 10.1029/JB095iB08p12653.
- Clark, R. N. (1993) 'Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy', *Remote Sensing for the Earth Sciences*, pp. 3–52. <https://pdfs.semanticscholar.org/1e02/a41e9ff88cdf80bbe41bd81ecde71e779548.pdf>.
- Cockx, L., Van Meirvenne, M. et Hofman, G. (2005) 'Characterization of Nitrogen Dynamics in a Pasture Soil by Electromagnetic Induction', *Biology and Fertility of Soils*, 42(1), pp. 24–30. doi: 10.1007/s00374-005-0866-3.
- Comino, F., Aranda, V., Garcia-Ruiz, R., Ayora-Canada, M. J. et Dominguez-Vidal, A. (2018) 'Infrared Spectroscopy as a Tool for The Assessment of Soil Biological Quality in Agricultural Soils Under Contrasting Management Practices', *Ecological Indicators*. Elsevier, 87, pp. 117–126. doi: 10.1016/j.ecolind.2017.12.046.
- Crozier, C. R., Heiniger, R. W. et Bishop, M. (1999) 'Evaluating Soil Compaction with a Portable Electronic Cone Penetrometer', *Technology and Product Reports*, 2, pp. 440–443. doi : 10.21273/HORTTECH.9.3.440.

- Dexter, A. R. (1988) 'Advances in Characterization of Soil Structure', *Soil & Tillage Research*, 11, pp. 199–238. doi : 10.1016/0167-1987(88)90002-5.
- Diaz-Zorita, M., Perfect, E. et Grove, J. H. (2002) 'Disruptive Methods for Assessing Soil Structure', *Soil and Tillage Research*, 64, pp. 3–22. doi : 10.1016/S0167-1987(01)00254-9.
- Domsch, H. et Giebel, A. (2004) 'Estimation of Soil Textural Features from Soil Electrical Conductivity Recorded Using The EM38', *Precision Agriculture*, 5(4), pp. 389–409. doi: 10.1023/B:PRAG.0000040807.18932.80.
- Ehsani, M.R., Upadhyaya, S.K., Slaughter, D. et al. (1999) 'A NIR Technique for Rapid Determination of Soil Mineral Nitrogen', *Precision Agriculture* 1, pp. 219–236 doi:10.1023/A:1009916108990.
- Eilers, P. H. C. et Boelens, H. F. M. (2005) Baseline Correction with Asymmetric Least Squares Smoothing.
https://zanran_storage.s3.amazonaws.com/www.science.uva.nl/ContentPages/443199618.pdf.
- El-Alem, A., Chokmani, K., Laurion, I., El-Adlouni, S., Raymond, S. et Ratté-Fortin, C. (2019) 'Ensemble-Based Systems to Monitor Algal Bloom With Remote Sensing', *IEEE Transactions on Geoscience and Remote Sensing*. IEEE, 57(10), pp. 7955–7971. doi: 10.1109/TGRS.2019.2917636.
- Fidêncio, P. H., Poppi, R. J., De Ndrade, J. C et De Abreu, M. F. (2008) 'Use of Radial Basis Function Networks and Near-Infrared Spectroscopy for the Determination of Total Nitrogen Content in Soils from Sao Paulo State', *Analytical Sciences*, 24(July), pp. 945–948. doi : 10.2116/analsci.24.945.
- Gandariasbeitia, M., Besga, G., Albizu, I., Larregla, S. et Mendarte, S. (2017) 'Prediction of Chemical and Biological Variables of Soil in Grazing Areas with Visible- and Near-infrared Spectroscopy', *Geoderma*. Elsevier, 305(June), pp. 228–235. doi: 10.1016/j.geoderma.2017.05.045.
- Genot, V., Colinet, G., Bock, L., Vanvyve, D., Reusen, R. et Dardenne, P. (2011) 'Near Infrared Reflectance Spectroscopy for Estimating Soil Characteristics Valuable in The Diagnosis of Soil Fertility', *Journal of Near Infrared Spectroscopy*, 19(2), pp. 117–138. doi: 10.1255/jnirs.923.
- Goddu, R. F. et D. A. (1960) 'Spectra-Structure Correlations for the Near-Infrared Region', *Analytical Chemistry*, 32, pp. 140–141. doi : 10.1021/ac60157a048.

Gogé, F., Joffre, R., Jolivet, C., Ross, I. et Ranjard, L. (2012) 'Chemometrics and Intelligent Laboratory Systems Optimization Criteria in Sample Selection Step of Local Regression for Quantitative Analysis of Large Soil NIRS Database', *Chemometrics and Intelligent Laboratory Systems*. Elsevier B.V., 110(1), pp. 168–176. doi: 10.1016/j.chemolab.2011.11.003.

Gogé, F., Gomez, C., Jolivet, C. et Joffre, R. (2014) 'Which Strategy is Best to Predict Soil Properties of a Local Site from a National Vis – NIR Database?', *Geoderma*. Elsevier B.V., 213, pp. 1–9. doi: 10.1016/j.geoderma.2013.07.016.

Guerrero, C., Wetterlind, J., Stenberg, Bo., Mouazen, A. M, Gabarron-Galeote, M. A., Ruiz-Sinoga, J. D, Zornoza, R. et Rossel, V.A (2016) 'Do We Really Need Large Spectral Libraries for Local Scale SOC Assessment with NIR Spectroscopy?', *Soil & Tillage Research*. Elsevier B.V., 155, pp. 501–509. doi: 10.1016/j.still.2015.07.008.

Helland, I. S., Naes, T. et Isaksson, T. (1995) 'Related Versions of The Multiplicative Scatter Correction Method for Preprocessing Spectroscopic Data', *Chemometrics and intelligent laboratory systems*, 7439(95). doi : 10.1016/0169-7439(95)80098-T.

Hanks, R. J. et BOWERS, S. A. (1965) 'Reflection of Radiant Energy from Soils', *Soil Science*, 100 (2), pp. 130–138. <https://www.ars.usda.gov/ARSUserFiles/30200525/870%20Reflection%20of%20radiant%20energy%20from%20soils.pdf>.

Hunt, G. (1977) 'Spectral Signatures of Particulate Minerals in The Visible and Near Infrared', pp. 501–513. <https://courses.eas.ualberta.ca/eas451/hunt%201977.pdf>.

Ji, W., Adamchuk, V. I., Biswas, A., Dhawale, N. M., Sudarsan, B., Zhang, Y., Rossel, V. A. et Shi, Z. (2016) 'Assessment of Soil Properties in Situ Using a Prototype Portable MIR Spectrometer in Two Agricultural Fields', *Biosystems Engineering*. Elsevier Ltd, 152, pp. 14–27. doi:10.1016/j.biosystemseng.2016.06.005.

Ji, W., Li, S., Chen, S., Shi, Z., Rossel, V. A. et Mouazen, A. M. (2016) 'Prediction of Soil Attributes Using The Chinese Soil Spectral Library and Standardized Spectra Recorded at Field Conditions', *Soil & Tillage Research*. Elsevier B.V., 155, pp. 492–500. doi: 10.1016/j.still.2015.06.004.

Ji, W., Viscarra Rossel, R. A. et Shi, Z. (2015) 'Accounting for The Effects of Water and The Environment on Proximally Sensed Vis-NIR Soil Spectra and Their Calibrations', *European Journal of Soil Science*, 66(3), pp. 555–565. doi: 10.1111/ejss.12239.

Jolliffe, I. T. (2002) *Principal Component Analysis, Second Edition*.
[http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%20I.%20Principal%20Component%20Analysis%20\(2ed.,%20Springer,%202002\)\(518s\)_MVsa_.pdf](http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%20I.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%202002)(518s)_MVsa_.pdf).

Kang, S.H et Xing, B. (2005) 'Phenanthrene Sorption to Sequentially Extracted Soil Humic Acids and Humins', *American Chemical Society*, 39(1), pp. 134–140. doi : 10.1021/es0490828.

Kettler, T. A., Doran, J.W. et Gilbert, T. L. (2001) 'Simplified Method for Soil Particle-Size Determination to Accompany Soil-Quality Analyses', *Soil Science society of America*, 852 (13277), pp. 849–852. doi : 10.2136/sssaj2001.653849x.

Knadel, M., Stenberg, B., Deng, F., Thomson, A. et Greve, M. H. (2013) 'Comparing predictive abilities of three - visible-near infrared - spectrophotometers for soil organic carbon and clay - determination', *Journal of Near Infrared Spectroscopy*, 80(January), pp. 67–80. doi: 10.1255/jnirs.1035.

Knox, N. M., Grunwald, S., McDowell, M. L., Bruland, G. L., Myers, D. B. et Harris W. G. (2015) 'Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy', *Geoderma. Elsevier B.V.*, 239–240, pp. 229–239. doi: 10.1016/j.geoderma.2014.10.019.

Korsaeth, A. (2005) 'Soil apparent electrical conductivity (ECa) as a means of monitoring changes in soil inorganic N on heterogeneous morainic soils in SE Norway during two growing seasons', *Nutrient Cycling in Agroecosystems*, 72(3), pp. 213–227. doi: 10.1007/s10705-005-1668-6.

Krishnan, P., Alexander, D. J., Butler, B. J. et Hummel, J. W. (1979) 'Reflectance Technique for Predicting Soil Organic Matter'. *Soil Science Society of America*, 44 (6), pp. 1282-1285. doi:10.2136/sssaj1980.03615995004400060030x.

Kuang, B., Tekin, Y. et Mouazen, A. M. (2014) 'Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon , pH and clay content Soil & Tillage Research Comparison between artificial neural network and p', *Soil & Tillage Research. Elsevier B.V.*, 146(October 2018), pp. 243–252. doi: 10.1016/j.still.2014.11.002.

Li, S., Ji, W., Chen, S., Peng, J., Zhou, Y. et Shi, Z. (2015) 'Potential of VIS-NIR-SWIR Spectroscopy from the Chinese Soil Spectral Library for Assessment of Nitrogen Fertilization

Rates in the Paddy-Rice Region, China', *Remote Sensing*, pp. 7029–7043. doi: 10.3390/rs70607029.

Lindgren, F., Geladi, P. et Wold, S. (1993) 'The kernel algorithm for PLS', *Journal of Chemometrics*, 7(1), pp. 45–59. doi: 10.1002/cem.1180070104.

Loghavi, M. et Khadem, M. R. (2006) 'Development of a Soil Bin Compaction Profile Sensor', *Agric. Sci. Technol.*, 8, pp. 1–13. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1023.7883&rep=rep1&type=pdf>.

Ludwig, B., Vormstein, S., Neibuhr, J., Heinze, S., Marschner, B. et Vohland, M. (2017) 'Estimation accuracies of near infrared spectroscopy for general soil properties and enzyme activities for two forest sites along three transects', *Geoderma*. Elsevier B.V., 288, pp. 37–46. doi: 10.1016/j.geoderma.2016.10.022.

Lui, W., Upadhyaya, S. K. et Sbibusawa, S. (1996) 'Development of a Texture / Soil Compaction Sensor', *Precision Agriculture*, (1986), pp. 618–630. doi:10.2134/1996.precisionagproc3.c76.

Macqueen, J. (1967) 'Some Methods for Classification and Analysis of Multivariate Observations', 233(233), pp. 281–297. <https://projecteuclid.org/euclid.bsm/1200512992>.

Madejov, J. et Komadel, P. (2001) 'Baseline Studies of the Clay Minerals Society Source Clays: Infrared Methods', *Clays Clay Miner*, 49, pp. 410–432. doi: 10.1346/CCMN.2001.0490508.

Maleki, M. R., Holm, L. V, Ramon, H., Merckx, R., De Baerdmaeker, J. et Mouazen, A. M. (2006) 'Phosphorus Sensing for Fresh Soils using Visible and Near Infrared Spectroscopy', *ScienceDirect*, 95, pp. 425–436. doi: 10.1016/j.biosystemseng.2006.07.015.

McBratney, A. B., Mendonca Santos, M. L., Minansy, B. (2006) 'On digital soil mapping', *Science Direct*, 117, pp. 3-52. doi: 10.1016/S0016-7061(03)00223-4.

Minasny, B., B.McBratney, A., Bellon-Maurel, V., Roger, J. M., Gobrecht, A., Ferrand, L. et Joalland, S. (2011) 'Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon', *Geoderma*. Elsevier B.V., 167–168, pp. 118–124. doi: 10.1016/j.geoderma.2011.09.008.

Morellos, A., Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebenshon, J. Bill, R. et Mouazen, A. M. (2016) 'Machine learning based prediction of soil total nitrogen , organic carbon and moisture content by using VIS-NIR spectroscopy', *Biosystems Engineering*. Elsevier Ltd, 152, pp. 104–116. doi: 10.1016/j.biosystemseng.2016.04.018.

- Morellos, A., Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebenson, J., Bill, R. et Mouazen, A. M. (2016) 'Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy', *Biosystems Engineering*. Elsevier Ltd, 152, pp. 104–116. doi: 10.1016/j.biosystemseng.2016.04.018.
- Muller, E. et Décamps, H. (2001) 'Modeling soil moisture - Reflectance', *Remote Sensing of Environment*, 76(2), pp. 173–180. doi: 10.1016/S0034-4257(00)00198-X.
- Mulqueen, J., Strafford, J. V. et Tanner, D. W. (1977) 'Evaluation of Penetrometers for Measuring Soil Strength', *Journal of terramechanics*, 14(3), pp. 137–151. doi : 10.1016/0022-4898(77)90012-X.
- Nagelkerke, N. J. D. (1991) 'A note on a general definition of the coefficient of determination', *Biometrika*, 78(3), pp. 691–692. doi: 10.1093/biomet/78.3.691.
- Nocita, M., Stenvens, A., Noon, C. et Van Wesemael, B. (2012) 'Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy', *Geoderma*. Elsevier B.V. doi: 10.1016/j.geoderma.2012.07.020.
- OceanOptics (2013) 'OceanView Installation and Operation Manual'.
- Peter, F., Bettina, L. et Kurt, V. (2009) 'Repeated double cross validation for classification models', *Journal of Chemometrics*, 23(4), pp. 160–171. <http://file.statistik.tuwien.ac.at/filz/papers/JChem09.pdf>.
- Pimstein, A. (2011) 'Performance of Three Identical Spectrometers in Retrieving Soil Reflectance under Laboratory Conditions', 75(2). doi: 10.2136/sssaj2010.0174.
- Pirie, A., Balwant Singh, B. et Islam, K. (2006) 'Spectroscopic techniques to predict several soil properties', *Australian Journal of Soil Research*, 43, pp. 713–721.
- Randolph, T. W. (2005) 'Scale-based normalization of spectral data', *Cancer Biomarkers*,
- Rännar, S., Lindgren, F., Geladi, P. et Wold, S. (1994) 'A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm', *Journal of Chemometrics*, 8(2), pp. 111–125. doi : org/10.1002/cem.1180080204.
- Reeves, J. B. (2010) 'Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis : Where are we and what needs to be done?', *Geoderma*. Elsevier B.V., 158(1–2), pp. 3–14. doi: 10.1016/j.geoderma.2009.04.005.

- Robertson, G. P., Coleman, D. , Bledsoe, C. S. et Soliins, P. (1999) 'Soil Sampling, Preparation, Archiving and quality Control', in *Standard Soil Methods for Long-Term Ecological Research*, pp. 3–28. <https://andrewsforest.oregonstate.edu/sites/default/files/lter/pubs/pdf/pub2706.pdf>.
- Rodionov, A. Patzold, S., Welp, G., Pude, R. et Amelung, W. (2016) 'Proximal field Vis-NIR spectroscopy of soil organic carbon : A solution to clear obstacles related to vegetation and straw cover', *Soil & Tillage Research*. Elsevier B.V., 163, pp. 89–98. doi: 10.1016/j.still.2016.05.008.
- Roger-estrade, J. Richard, G., Caneill, J., Boizard, H., Coquet, Y., Defosse, P. et Manichon, H. (2004) 'Morphological characterisation of soil structure in tilled fields : from a diagnosis method to the modelling of structural changes over time', 79, pp. 33–49. doi: 10.1016/j.still.2004.03.009.
- Roger, J. M. et Ecartot, M. (2016) 'Grain 5 : Prétraitements 1', in *CheMoocs*, pp. 1–17. https://www.fun-mooc.fr/asset-v1:Agreenium+66002+session03+type@asset+block/Grain5_Pretraitements.pdf.
- Rutledge, D. (no date) 'Grain 8 : Régression PLS', in, pp. 2–11.
- Rossel, R. A. V., Walvoort, D. J. J., MacBratney, A. B., Janick, L. J. et Skjemstad, J. O. (2006) 'Visible , near infrared , mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties', 131, pp. 59–75. doi: 10.1016/j.geoderma.2005.03.007.
- Rossel, R. A. V., Mcclynn, R. N. et Mcbratney, A. B. (2006) 'Determining the composition of mineral-organic mixes using UV – vis – NIR diffuse reflectance spectroscopy', 137, pp. 70–82. doi: 10.1016/j.geoderma.2006.07.004.
- Rossel, R., Jeon, Y. S, Odeh, I., Mcbartney, A.. (2008) 'Using a legacy soil sample to develop a mid-IR spectral library', *Soil Research*, pp. 1–16.
- Sadegh, M., Rourke, S. M. O. et Holden, N. M. (2015) 'Geoderma Evaluation of soil quality for agricultural production using visible – near-infrared spectroscopy', *Geoderma*. Elsevier B.V., 243–244, pp. 80–91. doi: 10.1016/j.geoderma.2014.12.012.
- Savitzky, A. et Golay, M. J. E. (1964) 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures', 36(8), pp. 1627–1639.
- Schulte, E. E. (1984) 'Chapter 8 Recommended Soil Organic Matter Tests'.
- Sébastien, P. (2016) 'Grain 3 : ACP - 1', in *cheMoocs*, pp. 1–20.

- Shepherd, K. D. et Walsh, M. G. (2002) 'Development of Reflectance Spectral Libraries for Characterization of Soil Properties', Soil Science Society of America, (February 2015). doi: 10.2136/sssaj2002.0988.
- Shi, Z., Ji, W., Viscara Rossel, R. A., Chen, S. et Zhou, Y. (2015) 'Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis – NIR spectral library', European Journal of Soil Science, pp. 679–687. doi: 10.1111/ejss.12272.
- Sibley, K. J. (2008) 'Development and use of an automated on-the-go soil nitrate mapping system'.
- Sokalski, T., Lingenfelter, P. et Lewenstam, A. (2003) 'Numerical Solution of the Coupled Nernst - Planck and Poisson Equations for Liquid Junction and Ion Selective Membrane Potentials', American Chemical Society, pp. 2443–2452.
- Somers, B., Gysels V., Verstraeten, W. W., Delalieux, S. et Coppin, P. (2010) 'Modelling moisture-induced soil reflectance changes in cultivated sandy soils: A case study in citrus orchards', European Journal of Soil Science, 61(6), pp. 1091–1105. doi: 10.1111/j.1365-2389.2010.01305.x.
- Stenberg, B., Rossel, R. A. V. et Mouazen, A. M. (2010) 'Visible and near infrared spectroscopy in soil science', 107(10), pp. 1–44.
- Sudduth, K. a., Drummond, S. T. et Kitchen, N. R. (2001) 'Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture', Computers and Electronics in Agriculture, 31(3), pp. 239–264. doi: 10.1016/S0168-1699(00)00185-X.
- Susha Lekshmi, S. U., Singh, D. N. et Shojaei, M. (2014) 'A critical review of soil moisture measurement', Measurement. Elsevier Ltd, 54, pp. 92–105. doi: 10.1016/j.measurement.2014.04.007.
- Tarr, A. B. Moore, K. J., Bullock, D. G., Dixon, P. M. et Burras, C. L. (2005) 'Improving map accuracy of soil variables using soil electrical conductivity as a covariate', Precision Agriculture, 6(3), pp. 255–270. doi: 10.1007/s11119-005-1385-9.
- Taubner, H., Roth, B. et Tippkötter, R. (2009) 'Determination of soil texture : Comparison of the sedimentation method and the laser-diffraction analysis', InterScience, 172, pp. 161–171. doi: 10.1002/jpln.200800085.

- Terra, F. S., Demattê, J. A. M. et Viscarra, R. A. (2015) 'Geoderma Spectral libraries for quantitative analyses of tropical Brazilian soils : Comparing vis – NIR and mid-IR re fl ectance data', *Geoderma*. Elsevier B.V., 255–256, pp. 81–93. doi: 10.1016/j.geoderma.2015.04.017.
- Tollner, E. W. et Verma, B. P. (1984) 'Modified Cone Penetrometer for Measuring Soil Mechanical Impedance', pp. 331–336.
- Tsukada, K. Sebata, M., Miyahara, Y. et Miyagi, H. (1989) 'Long-life multiple-isfets with polymeric gates', *Sensors and Actuators*, 18, pp. 329–336.
- Twomey, S. A., Bohren, C. F. et Mergenthaler, J. L. (1986) 'Reflectance and albedo differences between wet and dry surfaces', *Optical Society of America*, 25(3), pp. 431–437.
- Viscarra Rossel, R. A., Jeon, Y. S., Odeh, O. A. et Mcbratney, A. B. (2008) 'Using a legacy soil sample to develop a mid-IR spectral library', *Australian Journal of Soil Research*, 46, pp. 1–16. doi: 10.1071/SR07099.
- Viscarra Rossel, R. A. et al. (2016) 'A global spectral library to characterize the world's soil', *Earth-Science Reviews*, 155(February), pp. 198–230. doi: 10.1016/j.earscirev.2016.01.012.
- Vohland, M., Ludwig, M., Thiele-Bruhn, S. et Ludwig, B. (2014) 'Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection', *Geoderma*. Elsevier B.V., 223–225, pp. 88–96. doi: 10.1016/j.geoderma.2014.01.013.
- Walkley, A., et Armstrong Black, I. (1993) 'An Examination of the DEGTJAREFF Method for Determining Soil Organic Matter, and a Proposed Modification of the Chromic Acid Titration Method', *Soil Science*, 37, pp. 29-38. doi : 10.1097/00010694-193401000-00003.
- Waruru, B. K., Shephered, K. D., Ndegwa, G. M. et Sila, A. M. (2016) 'Estimation of wet aggregation indices using soil properties and diffuse reflectance near infrared spectroscopy : An application of classification and regression tree analysis', *Biosystems Engineering*. Elsevier Ltd, 152, pp. 148–164. doi: 10.1016/j.biosystemseng.2016.08.003.
- Weidong, L., Baret, F., Xingfa, G, Qingxi, T., Lanfen, Z. et Bing, Z. (2002) 'Relating soil surface moisture to reflectance', *Remote sensing of environment*, 81, pp. 238–246.
- Wold, S., Sjöström, M. and Eriksson, L. (2001) 'PLS-regression: A basic tool of chemometrics', *Chemometrics and Intelligent Laboratory Systems*, 58(2), pp. 109–130. doi: 10.1016/S0169-7439(01)00155-1.
- Xiaofei, A. Minzan, L., Lihua, Z. et Hong, S. (2015) 'Eliminating the interference of soil moisture and particle size on predicting soil total nitrogen content using a NIRS-based portable detector',

Computers and Electronics in Agriculture. Elsevier B.V., 112, pp. 47–53. doi: 10.1016/j.compag.2014.11.003.

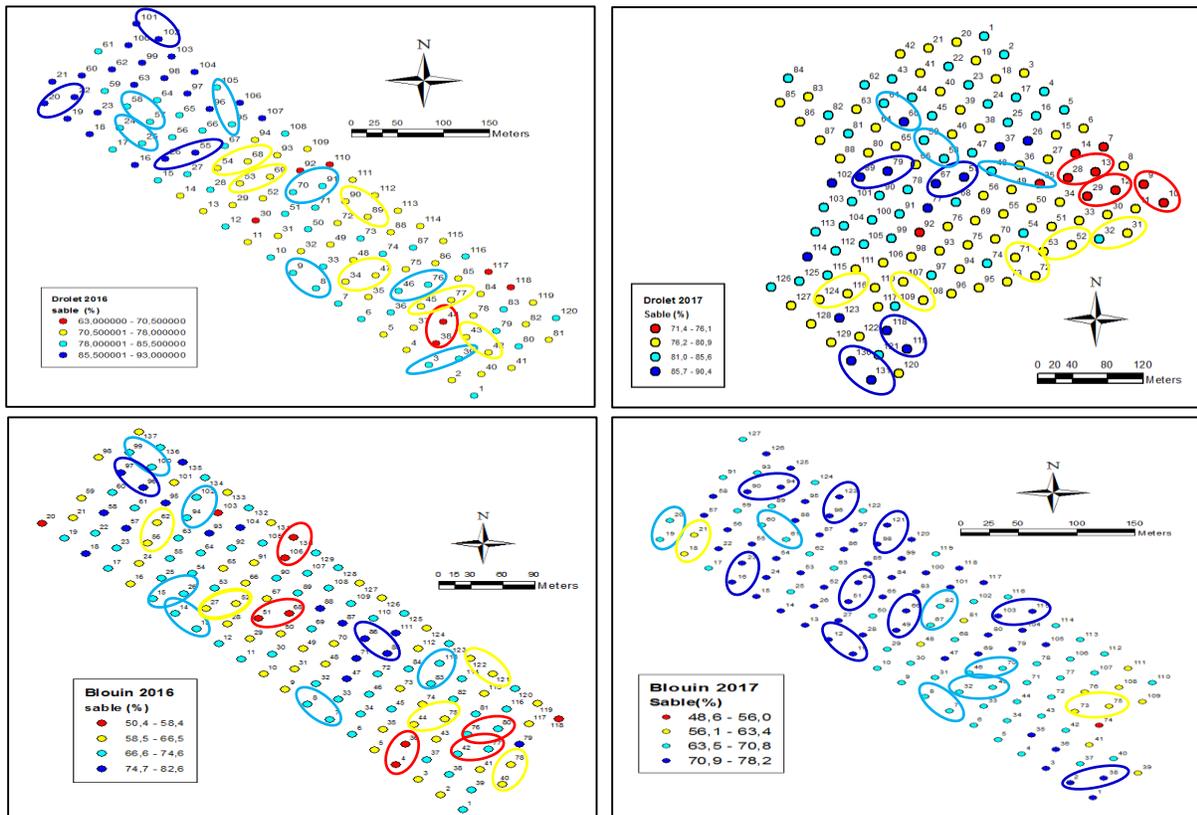
Xu, Q. S., Liang, Y. Z. et Du, Y. P. (2004) 'Monte Carlo cross-validation for selecting a model and estimating the prediction error in multivariate calibration', *Journal of Chemometrics*, 18(2), pp. 112–120. doi: 10.1002/cem.858.

Yeasmin, S., Singh, B. Johnson, C. et Sparks, D. (2017) 'Evaluation of pre-treatment procedures for improved interpretation of mid infrared spectra of soil organic matter', *Geoderma*. Elsevier B.V., 304, pp. 83–92. doi: 10.1016/j.geoderma.2016.04.008.

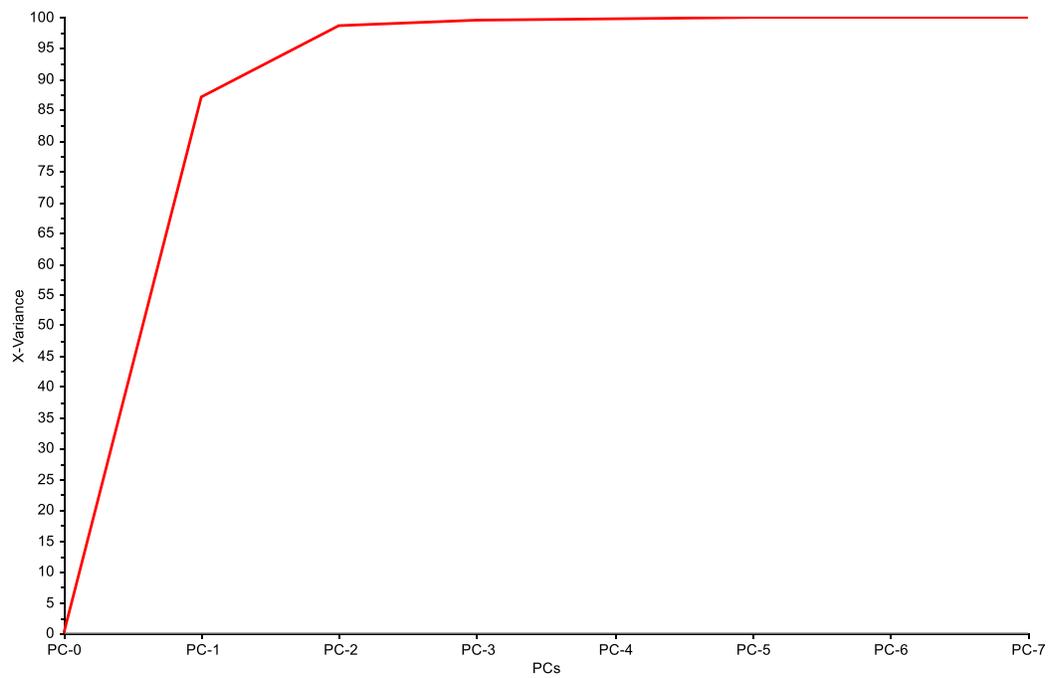
Yegül, U., Türker, U. et Talebpour, B. (2011) 'Determination of Some Soil Parameters with Electromagnetic Induction Sensor', *Journal of Agriculture Machinery Science*, 7(1), pp. 19–25.

Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K.M., Arcenegui, V. et Mataix-Beneyto, J. (2008) 'Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils', *Soil Biology & Biochemistry*, 40, pp. 1923–1930. doi: 10.1016/j.soilbio.2008.04.003.

7 ANNEXE



Annexe 1 : Échantillons sélectionnés pour les quatre champs étudiés.



Annexe 2 : Évolution de la variance expliquée par les composantes principales.

Annexe 3 : Analyse descriptive des valeurs observées de teneur en eau, argile, sable et matière organique dans le sous-ensemble de calibration pour les classes d'humidité du sol C1 (0-10 %), C2 (10-19 %) et C3 (>19 %).

Propriété	Classe	Min	Max	Moyenne	Écart type	Éléments
Teneur en eau (%)	C1	0	9.52	2.96	2.91	132
	C2	10.01	18.77	14.26	2.43	49
	C3	19.08	33.84	23.88	3.65	35
Teneur en argile (%)	C1	2.71	13.23	7.15	2.38	132
	C2	2.85	13.23	7.28	2.36	49
	C3	3.37	12.45	6.78	2.35	35
Teneur en sable (%)	C1	54.24	91.09	74.28	9.23	132
	C2	54.24	88.92	72.74	8.34	49
	C3	54.55	91.09	75.98	9.34	35
Teneur en matière organique	C1	1.89	7.19	4.30	1.38	132
	C2	1.89	6.94	4.12	1.22	49
	C3	2.54	7.05	4.77	1.35	35

Annexe 4 : Analyse descriptive des valeurs observées de teneur en eau, argile, sable et matière organique dans le sous-ensemble de validation pour les classes d'humidité du sol C1 (0-10 %), C2 (10-19 %) et C3 (>19 %).

Propriété	Classe	Min	Max	Moyenne	Écart type	Éléments
Teneur en eau (%)	C1	0	9.71	3.06	3.00	44
	C2	10.38	18.77	14.45	2.51	16
	C3	19.51	34.33	24.36	4.28	12
Teneur en argile (%)	C1	3.21	13.23	7.38	2.56	44
	C2	3.56	10.71	7.29	2.12	16
	C3	2.85	13.23	6.14	2.95	12
Teneur en sable (%)	C1	54.24	88.92	74.06	10.11	44
	C2	56.31	91.09	74.63	9.39	16
	C3	54.91	91.09	77.49	9.74	12
Teneur en matière organique	C1	2.54	6.94	4.10	1.19	44
	C2	2.99	7.15	4.52	1.34	16
	C3	1.89	7.15	4.22	1.51	12