



Analyse de la variabilité pédologique intra-parcellaire à partir de données spatiales de télédétection, de résistivité et de pédologie

Rémy Bouyrat

► To cite this version:

Rémy Bouyrat. Analyse de la variabilité pédologique intra-parcellaire à partir de données spatiales de télédétection, de résistivité et de pédologie. Science des sols. 2023. hal-04196767

HAL Id: hal-04196767

<https://hal.inrae.fr/hal-04196767v1>

Submitted on 5 Sep 2023

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

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



Master Sciences de l'Eau
Hydrosystèmes et Bassins Versants - HBV

*Rapport de stage de fin d'études pour l'obtention
de la 2nde année de Master*

**Analyse de la variabilité pédologique
intra-parcellaire à partir de données
spatiales de télédétection, de résistivité et
de pédologie**



Année universitaire : 2022 – 2023

Encadrement : Manuel MARTIN et Hocine BOURENNANE

Organisme d'accueil : INRAe

Avenue de la Pomme de Pin Orléans 45075

Remerciements

Je tiens à remercier en premier Hocine Bourennane et Manuel Martin, mes deux encadrants qui ont été excellents dans mon encadrement. Je les remercie pour leur disponibilité, leurs réponses simples, claires et surtout sans jugement. J'ai pu m'épanouir pleinement dans un sujet qui n'était pas facile pour moi n'ayant aucune connaissance dans le domaine. Les conseils qu'ils ont pu me prodiguer ont permis de m'améliorer scientifiquement. Mais je les remercie aussi pour leur sens de l'humour qui permet de discuter aussi d'autres sujets et je pense à ce titre que cela est primordial pour le bon déroulement d'un stage. Je tiens aussi à remercier Emmanuelle Vaudour pour son aide précieuse dans ce stage et qui a pu m'aiguiller dans les démarches à faire avec la télédétection. Merci aussi à Maxence Dodin qui grâce à son travail de thèse m'a permis d'obtenir des données pour mon travail de manière rapide.

Merci aussi aux partenaires du projet qui m'ont fourni ces données dans ce stage, Maud Seger et Ghislain Girot (j'ai adoré la sortie terrain) ainsi que Lucie Martin et Antoine Savoie qui m'ont aidé à mieux comprendre le site d'étude et donc à améliorer mon travail.

Je remercie aussi les « CDD » et stagiaires présents avec moi durant ces 6 mois qui m'ont permis de me sentir bien dans cette ville monstrueuse qu'est Orléans (Vive Tours). Une pensée surtout à mes collègues de bureau, Oscar qui m'a beaucoup appris sur R, Lucien pareillement et puis un fabuleux partenaire de sport. Mais surtout une grosse pensée pour LouannE, mon pilier dans cet univers de travail avec qui les discussions ont abordé à peu près tous les thèmes imaginables. A Samir avec qui j'ai pu parler de sport auto et ça c'est cool. Merci à Eoghann pour nos petits trajets ensemble parce qu'il ne se lève jamais à l'heure. Claire aussi qui a été voisine de bureau et qui entendait régulièrement nos rigolades. Une pensée pour les thésards que sont Maëlys et Emile (cet homme plein de bons conseils sur la vie) qui vont avoir besoin de courage pour la fin de thèse. Bref, merci tout ce petit monde.

Puis je remercie le reste de l'équipe d'Info&Sols avec qui j'ai pu rigoler et passer de bons moments ainsi que pour leurs conseils scientifiques (pensée spéciale Eugénie).

Vu que c'est la dernière fois que j'écis pour l'université (j'espère), je remercie ma mère de m'avoir supporté tout ce temps avec les hauts et les bas de la vie étudiante. Je remercie mon père pour ces dimanches ensemble qui font du bien.

Et enfin, merci à Morgane de m'avoir aidé, et d'avoir enduré mon stress durant maintenant deux ans où j'ai pu partager un petit bout de vie avec elle. Elle a autant contribué à ma réussite scolaire que personnelle de par sa façon d'être et d'aimer. Avoir une Morgane dans la vie c'est un code de triche pour être heureux.

Table des matières

Introduction	1
I. Contexte	1
II. Matériels et méthodes.....	3
A. Site d'étude	3
B. Acquisition des sources de données	4
i. Télédétection	4
ii. Résistivité.....	5
iii. Pédologie.....	6
iv. L'itinéraire culturel	7
C. Démarche globale	7
III. Résultats	10
A. Analyse descriptive des données	10
B. Spatialisation des données	13
i. Pédologie	13
ii. Résistivité.....	16
C. Classification des données.....	16
D. Evaluation statistique des classifications sur les données sources	20
i. Télédétection	20
ii. Pédologie.....	22
iii. Résistivité.....	22
E. Signification pédologique des classifications sur les données de télédétection et de résistivité.....	22
IV. Discussion	25
A. Incertitude sur les données	25
i. Incertitudes sur les mesures	25
ii. Incertitudes induites par le prétraitement.....	25
B. La classification des données	26
i. Méthode de classification	26
ii. Classifications	26
C. Signification de la comparaison télédétection/pédologie et résistivité/pédologie	27
V. Conclusion.....	28
Bibliographie	30
Liste figures/tableaux	32
Annexes	34
Résumé	37

Abstract.....	37
---------------	----

Glossaire

Clustering : méthode de partitionnement des données selon des critères de distance ou de similarité.

Classification : équivalent à clustering

Introduction

L'agriculture a suscité l'intérêt de nombreuses recherches dans des domaines variés dont la télédétection. Les techniques propres à ce domaine sont accessibles et, une fois les satellites mis en orbite, économiques (Khaliq et al., 2018). Elles offrent en outre l'opportunité d'étudier les sols de manière non-destructive. Des auteurs ont aussi démontré que la télédétection peut permettre de quantifier des propriétés des horizons de surface des sols, comme la teneur en argile ou encore la matière organique (eg. Chabrillat et al., 2019 ; Santaga et al., 2021 ; Vaudour et al., 2022 ; Richer-de-Forges et al., 2023). D'autres propriétés pédologiques comme l'humidité (Post et al., 2000 ; Galvão et al., 2001), la salinité (Mohamed et al., 2018), ou encore l'azote et le phosphore (Hatfield et al., 2008) des horizons de surface des sols peuvent être quantifiés.

La télédétection est un ensemble de techniques qui consistent à recueillir des informations sur les propriétés d'objets, comme le sol, à partir de la réflectance de ces derniers afin d'en analyser les caractéristiques à distance. Les bandes spectrales déterminent l'intervalle de capture des longueurs d'ondes. Le programme Sentinel-2, lancé par l'European Space Agency (ESA) en 2015, couple deux satellites jumeaux. Ces derniers sont équipés de 13 bandes captant des longueurs d'ondes allant de 440 nm à 2200 nm. Le résultat obtenu à partir de ces satellites sont des images multi-spectrales à haute résolution spatiale (de 10 m à 60 m en fonction des bandes) (Khaliq et al., 2018).

En fonction de la cible, soit les bandes spectrales, soit les indices spectraux peuvent être utilisés. Les indices sont le résultat de combinaison de bandes spectrales, comme le *Normalised Difference Vegetation Index* (NDVI) (Rouse & Haas, 1973), l'indice le plus renseigné en télédétection. Cela étant, des études récentes (e.g., Vaudour et al., 2021) suggèrent d'utiliser conjointement plusieurs indices dans l'étude des propriétés des horizons de surface des sols.

Ce stage est une suite directe du projet DUGIATOE, qui rassemblait INRAe (unités Info&Sols, Orléans et EcoSys Versailles-Saclay) et le laboratoire d'analyses de sols Auréa. Ce projet, initié fin 2021, visait à établir des méthodes de localisation des analyses de terre réalisées par Auréa à la demande des gestionnaires des terres, notamment des agriculteurs, en se basant sur les données de télédétection Sentinel 2. Courant 2022, un stage M1 avait été réalisé (Pean, 2022) et avait permis d'établir des pistes de traitement des données Sentinel 2 afin de caractériser l'hétérogénéité pédologique intra-parcellaire. Il s'agissait d'un préliminaire indispensable à la localisation des zones de prélèvements par Auréa, sachant que le nombre maximal de prélèvements par Auréa est de 2 à 3 prélèvements composites par parcelle. J'ai pris la suite de ce travail, avec des objectifs plus orientés recherche et moins directement appliqués qu'initialement dans le projet DUGIATOE (Auréa s'étant retiré du projet avant mon début de stage), en me focalisant sur la caractérisation de l'hétérogénéité intra-parcellaire à l'aide de différentes sources de données, y compris et surtout de télédétection. L'objectif de ce stage est d'utiliser la télédétection satellitaire, spécifiquement les séries satellitaires Sentinel-2, afin de déterminer des zones avec des réponses spectrales homogènes au sein de parcelles agricoles. Il s'agit ensuite de valider le zonage spectral obtenu au moyen de données pédologiques et de résistivité électrique des horizons des sols.

I. Contexte

L'objectif de ce stage est de délimiter des zones aux propriétés pédologiques homogènes en vue de proposer des plans d'échantillonnage optimisés pour des campagnes d'acquisition de données pédologiques.

Ce stage s'inscrit dans un contexte international de recherche de méthodes peu onéreuses et permettant d'étudier le sol précisément. Une des solutions est l'utilisation de la télédétection car peu coûteuse, exhaustive et non destructive. Cette source de données peut être utilisée pour des objectifs appliqués à l'agriculture comme mieux comprendre la variabilité pédologique d'une parcelle pour les pratiques agricoles. Elle peut aussi être utilisée pour des objectifs de recherche comme la compréhension de la distribution des propriétés pédologiques dans l'espace. La télédétection a l'avantage d'être de plus en plus accessible à tous, néanmoins l'exploitation des données qu'elle produit demeure complexe. Des chercheurs ont mis en place des indices spectraux afin de mieux appréhender diverses propriétés des horizons des sols. Les indices spectraux sont les résultats de combinaisons particulières des différentes bandes spectrales. Chaque indice a une signification propre et exprime une propriété ou un état particulier des horizons de surface des sols. Par exemple, le NDVI (Rouse and Haas, 1973) est l'un des indices les plus utilisés dans le domaine de l'étude des sols pour sa capacité à différencier un sol nu d'un sol recouvert. D'autres indices présentent des corrélations avec des variables pédologiques comme le *Soil Adjusted Total Vegetation Index* (SATVI) avec le carbone organique du sol (Gholizadeh et al., 2018) dans des conditions précises uniquement. Cette corrélation est peut-être due à la sensibilité des longueurs d'ondes des indices spectraux de végétation aux variations de carbone organique du sol (Jin et al., 2016). Ces indices ont ainsi été utilisés pour quantifier des propriétés pédologiques comme le carbone organique du sol à partir du NDVI par exemple (Bhunja et al., 2019). En effet, certains auteurs ont constaté des corrélations entre des propriétés des sols et des indices spectraux. Ils ont ensuite établi des fonctions de pédo-transfert qui permettent d'estimer spatialement la variable d'intérêt (e.g. Chen et al., 2018 ; Guo et al., 2021). Certains indices peuvent détecter la présence d'un type de végétation en particulier, comme les plantes halophytes, permettant d'en déduire des propriétés pédologiques comme la salinité (Mohamed et al., 2018). Ils sont aussi utilisés lors de l'estimation de résidus de cultures (Serbin et al., 2010) ou pour différencier les types de sol à partir des corrélations entre les indices et les paramètres pédologiques et par l'utilisation de modèles linéaires (Huete, 1988).

Ce stage a été réalisé au sein de l'unité de recherche Info&Sols de l'INRAe d'Orléans. L'UR Info&Sols est une unité propre de recherches du département Agroécosystèmes de l'INRAe. La mission principale confiée à l'UR Info&Sols est de développer des travaux de recherche en Science du Sol, sur l'évaluation quantitative du fonctionnement des sols en vue notamment de proposer divers paramètres mesurables ou quantifiables par des modèles et de définir des indices/indicateurs dans l'objectif de les positionner dans un référentiel. Ce besoin s'inscrit pleinement dans une demande générale des politiques publiques agricoles, environnementales et d'aménagement du territoire car tous les acteurs des sols réclament des indicateurs simples et accessibles, accompagnés d'une échelle de valeur pour appuyer leurs décisions dans la gestion des sols. Dans le cadre du Groupement d'intérêt scientifique Sol (GIS Sol) et en lien avec des partenaires en région, l'unité Info&Sols organise la surveillance et l'inventaire des sols de France afin de constituer le système national d'information sur les sols. L'unité coordonne plusieurs grands programmes nationaux comme le programme multi-échelle de cartographie des sols "Inventaire, gestion et conservation des sols (IGCS)" qui vise à décrire et à cartographier les sols à différentes échelles. Cette unité pilote également des programmes de surveillance parmi lesquels le réseau de mesure de la qualité des sols (RMQS), la base de données des analyses de terre (BDAT) et la base de données des éléments traces métalliques (BDETM).

L'objectif global de mon stage est d'utiliser une source de données exhaustive, la télédétection, pour délimiter des zones homogènes en matière de réflectance seule et/ou

d'indices spectraux. La spécificité de mon stage consiste à intégrer une composante temporelle dans la démarche en considérant une série temporelle d'images de télédétection du site d'étude. Cette dimension temporelle permet d'étudier l'évolution de la réflectance du sol à différents stades de couverture de la parcelle (sol nu et sol couvert par la végétation) afin de mettre en évidence de zones aux propriétés pédologiques homogènes. Ce stage mobilise aussi des données pédologiques éparées acquises dans deux autres projets de recherche (STEROPES et MELICERTES) et des données issues d'une campagne de résistivité électrique des sols du site d'étude. Les parcelles mobilisées dans le cadre de ce stage sont sur un site INRAE (Centre Val de Loire), et elles sont le support de plusieurs projets scientifiques (notamment STEROPES, MELICERTES, MARVIC, etc.) ainsi que d'un dispositif d'observation de longue durée (Méta-Métha, (Dodin et al., 2023)).

II. Matériels et méthodes

A. Site d'étude

Le site d'étude est situé dans la commune de Nouzilly, en Indre-et-Loire (Figure 1A). Le site d'étude est composé de deux parcelles agricoles, nommées dans la suite de ce rapport parcelle 29 et parcelle 30 (Figure 1B).

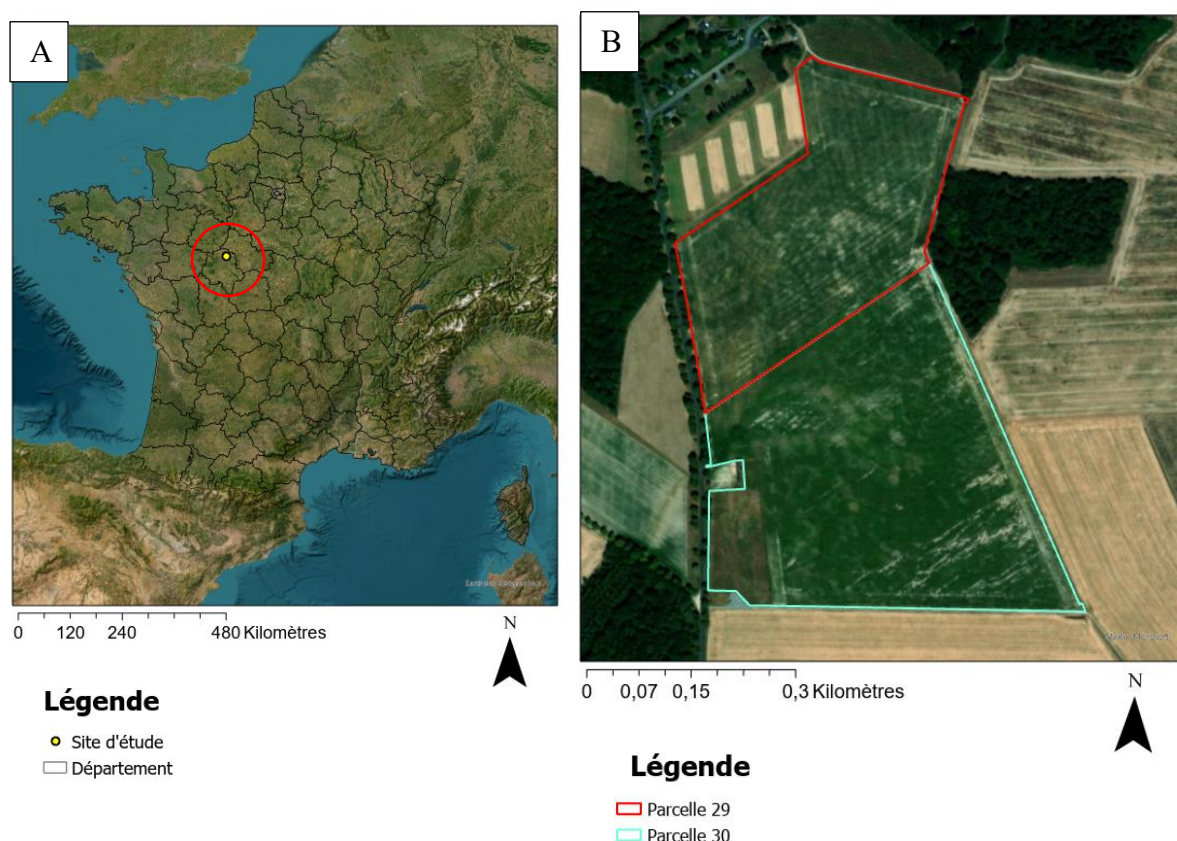


Figure 1 : A- Localisation du site d'étude. B- Composition du site d'étude.

Les deux parcelles se distinguent principalement par leur itinéraire cultural différent, amenant des différences au sein de l'évolution de la végétation au cours du temps. Ces parcelles couvrent une superficie de 11,34 ha et 16,71 ha respectivement et sont situées majoritairement sur des limons des plateaux (LP), avec une partie à l'ouest d'argile blanche ou beige à silex (Rs1) (Geoportail). Les types de sols (Figure 2) ont été déterminés à partir d'une prospection cartographique conventionnelle incluant des descriptions des horizons pédologiques sur

sondage à la tarière à main et sur deux fosses pédologiques. Dans la partie Nord du site d'étude se trouvent en majorité des sols lessivés hydromorphes profonds. Les planosols sont situés à l'extrémité Sud ainsi que vers le centre du site d'étude en deux lentilles distinctes. Enfin, les brunisols sont juxtaposés aux planosols situés au centre. C'est en majorité des sols lessivés qui occupent le site d'étude. Une quantification de la granulométrie des horizons, de leurs capacité d'échange cationique (CEC) et de leurs teneurs en carbone a été également réalisée pour 56 sondages (Giro, 2023).

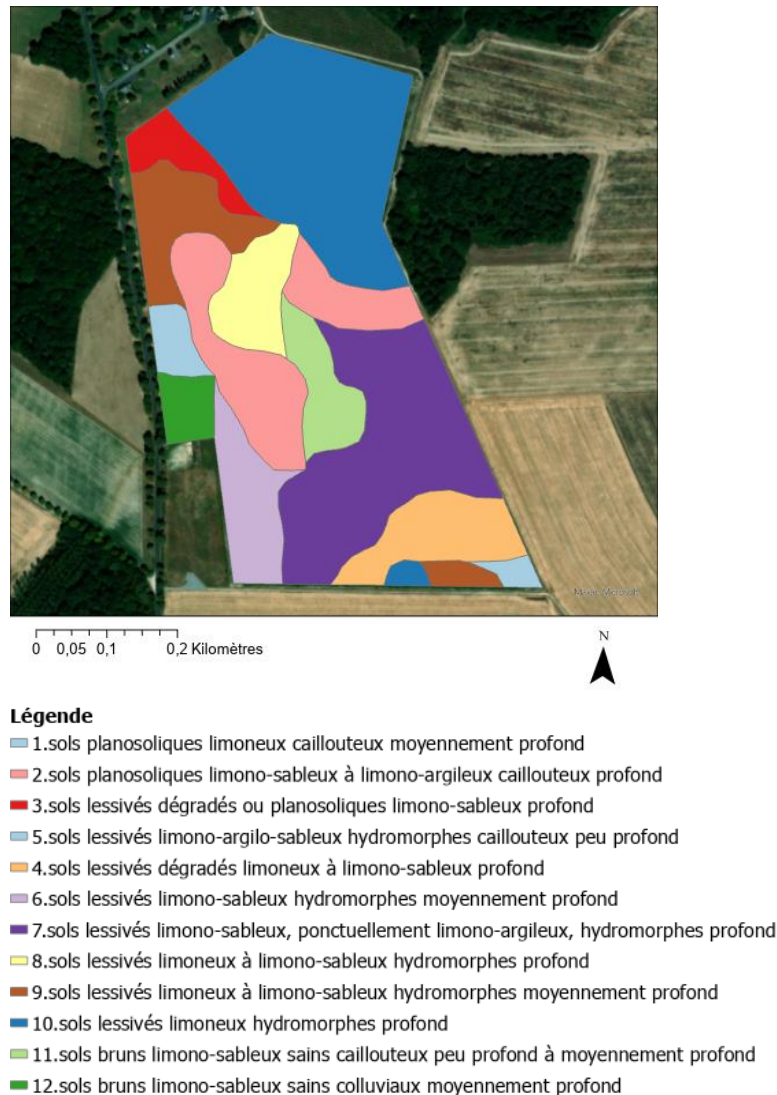


Figure 2 : Carte détaillée des types de sol au sein des deux parcelles d'études (Giro, 2023)

B. Acquisition des sources de données

i. Télédétection

La télédétection est la principale source de données utilisée. Elle se présente sous la forme d'une base de données composée d'images issues de la mission Sentinel-2. Cette mission fait partie du programme Copernicus, géré par l'Union Européenne, qui vise à collecter et mettre à disposition des données concernant l'environnement et la sécurité relative aux événements naturels (copernicus.eu). Elle a pour objectif de surveiller la variabilité des conditions du sol. Elle possède une large fenêtre d'observation (290 km) ainsi qu'un temps de revisite court (environ 5 jours).

Les images fournies portent sur la totalité du site d'étude, englobant les parcelles 29 et 30. Les images utilisées ont été prises depuis la mise en service du satellite sur une période de 5 ans, du 26/01/2017 au 06/08/2022. Le jeu de données d'images satellites utilisées vient de la plateforme Théia et plus précisément du produit Sentinel-2 corrigé en réflectance. L'extraction d'une partie des images a été réalisée à partir de la plateforme Théia (thèse en cours, Dodin) et a mobilisé des scripts python (O. Hagolle) et des scripts shell (M. Martin). L'autre partie a été fournie par un autre projet en cours sur les parcelles d'études. La base de données finale est composée de 159 images de télédétection.

Toutes les images ont été traitées avec le logiciel R (10.0) et plus précisément à l'aide du package terra (1.7) pour les requêtes en lien avec le système d'information géographique (SIG).

ii. Résistivité

Les données de résistivité ont été acquises à l'aide de la méthode ARP (Automatic Resistivity Profiling) le 17 et 18 octobre 2016 (Figure 3).

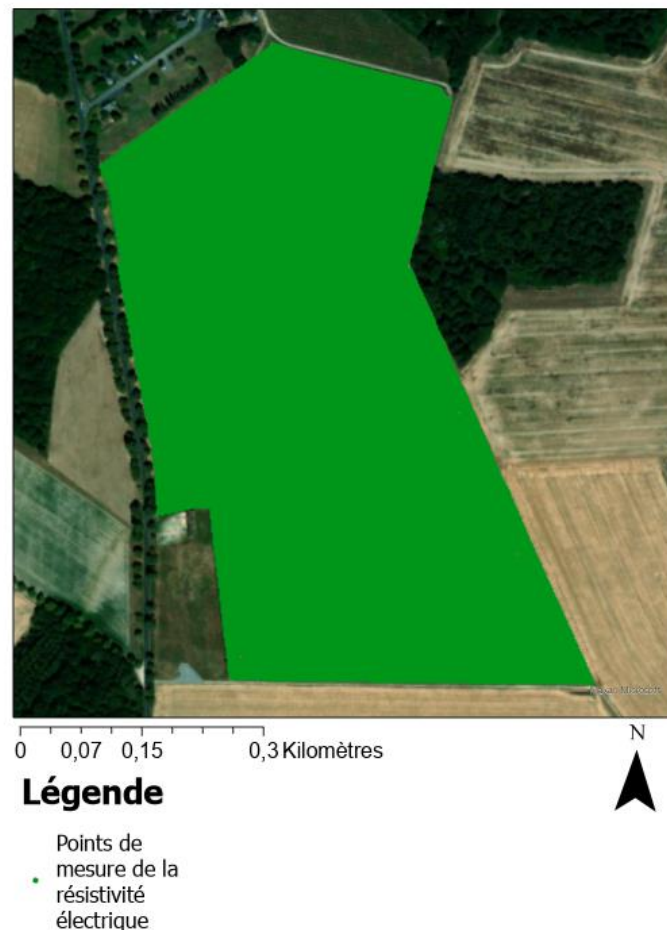


Figure 3 : Points de mesure de la résistivité électrique

L'espacement entre les profils électriques est de 6 mètres et le long du profil, une acquisition est enregistrée tous les 10 à 15 cm. Finalement pour chaque point d'acquisition, 3 valeurs de résistivité électrique sont enregistrées. Elles correspondent à 3 différentes profondeurs : résistivité de 0-50 cm (voie 1), résistivité de 50-100 cm (voie 2) et résistivité de 100-150 cm (voie 3). Les conditions météorologiques, ayant un impact direct sur la résistivité, sont : aucune précipitation durant les 2 jours de prospection, avec de faibles précipitations dans les semaines précédentes.

iii. Pédologie

Les données de pédologie sont issues de deux jeux de données provenant de deux études distinctes (Ecosys (L. Martin et E. Vaudour) et l'unité PAE (A. Savoie) dans le cadre du projet STEROPES) et les variables quantifiées ont été réalisées par deux laboratoires d'analyses. Le premier jeu de données est composé de 28 individus qui font partie d'un échantillonnage en

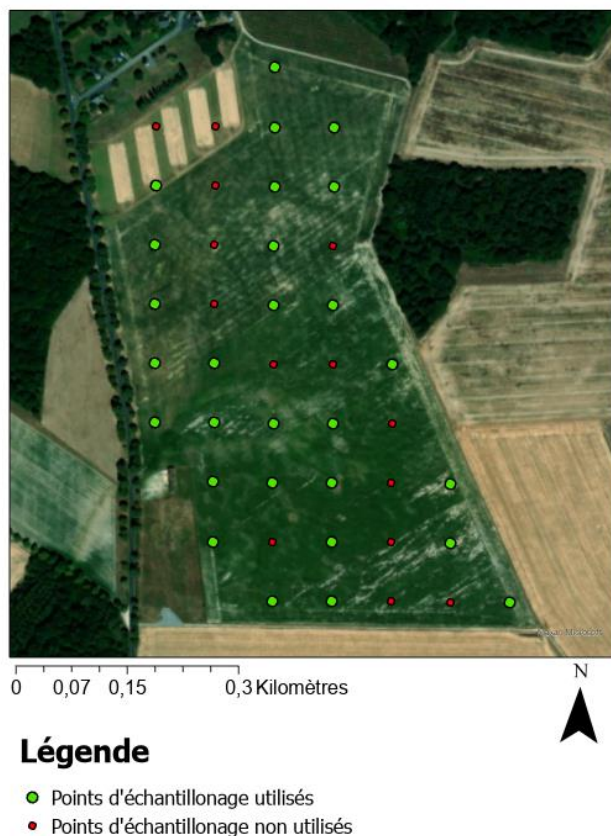


Figure 4 : Points d'échantillonnage retenus/non retenus pour analyse issus du premier jeu de données pédologiques

grille régulière de 40 points (individus) sur le site d'étude. Sur ces 40 individus, on dispose des analyses pédologiques pour un sous-ensemble de 28 individus (Figure 4), car les points de prélèvement supposés appartenir à une même unité pédologique n'ont pas tous été analysés. Ces analyses portent sur 13 variables pédologiques et sur l'ensemble des horizons pédologiques des sols.

Le second jeu de données est composé également de 28 individus issus d'un échantillonnage mixte (aléatoire et orienté), afin d'échantillonner une zone en particulier (adventices, zone caillouteuse, etc.). Les variables analysées dans ce jeu de données sont au nombre de 9.

Le jeu de données pédologiques final considéré dans cette étude est un assemblage des deux jeux de données présentés ci-dessus, portant donc sur 56 individus (Figure 5) et nous ne considérerons par la suite que les variables pédologiques mesurées dans les deux jeux de données, c'est-à-dire le carbone organique (désigné par le terme « carbone » par la suite), la CEC (cobaltihexamine), et les 5 fractions granulométriques pour l'horizon de surface.



Figure 5 : Représentation des points d'échantillonnage issus des deux jeux de données pédologiques

iv. L'itinéraire cultural

L'itinéraire cultural des parcelles sur la période 2016 à 2022 a été fourni par la personne en charge des parcelles (A. Savoie) sur la période 2016 à 2022. Ces informations contiennent les dates de semis et de récoltes ainsi que les actions effectuées sur les parcelles comme l'épandage (type, quantité, date, etc) ou les actions de travail du sol (superficiel, enfouissement, etc.).

C. Démarche globale

La figure 6 présente la démarche adoptée pour répondre à l'objectif de ce travail. Le texte de cette section décrit les principales étapes sans pour autant rentrer dans le formalisme mathématique sous-jacent aux différentes méthodes statistiques, géostatistiques et de système d'information géographique mobilisées dans ce travail.

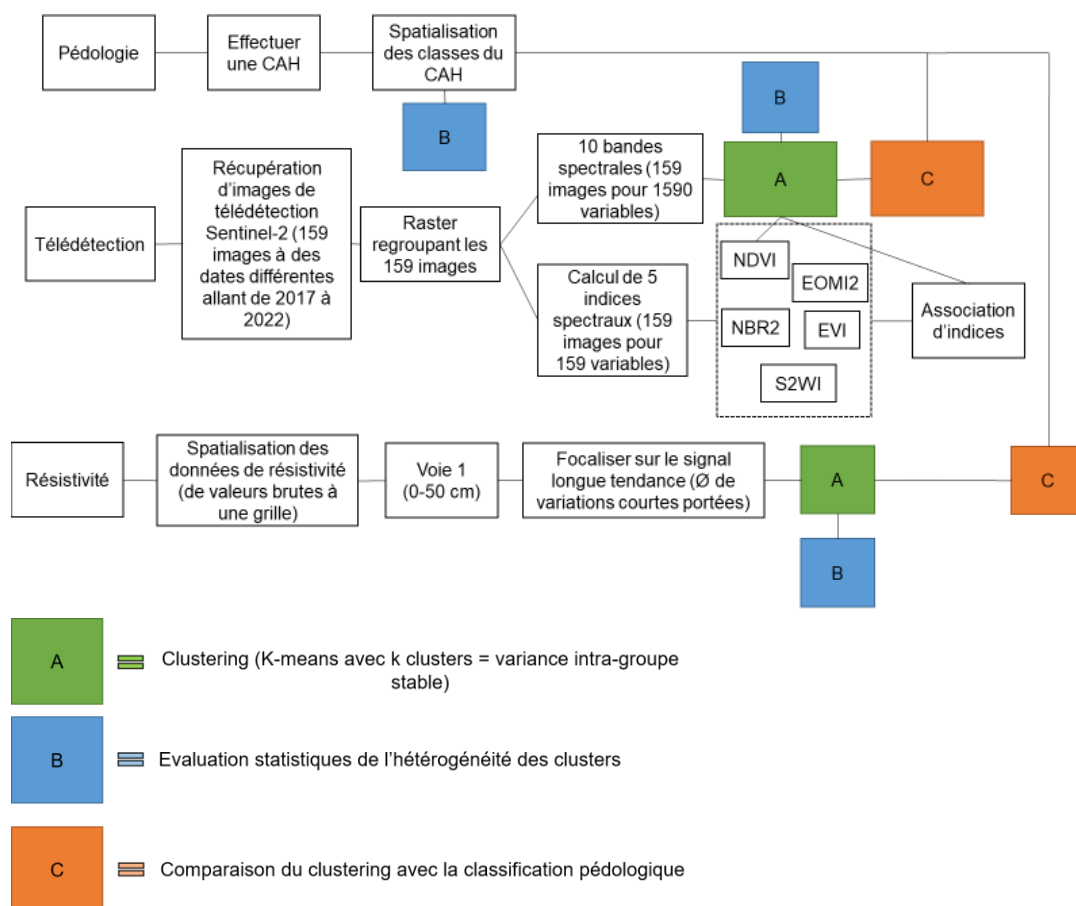


Figure 6 : Démarche globale utilisée dans cette étude. (« A » représente le clustering effectué par k-means où k clusters varie entre les deux sources de données (télédétection et résistivité). « B » représente une évaluation statistique des clusters qui a été faite soit par un test de Wilcoxon ou par une ANOVA. « C » représente la comparaison des clusterings de télédétection et de résistivité avec le résultat de la spatialisation des classes du CAH par un tableau de contingence et un test statistique appliqué si nécessaire.

Les individus du jeu de données pédologiques ont été classés grâce à une CAH : Classification Ascendante Hiérarchique (Szekely and Rizzo, 2005). La CAH est initiée en posant que le nombre de classes est égal au nombre n d'individus soit $n = n_{classes}$. A chaque itération, $n_{classes}$ doit diminuer afin d'atteindre une valeur de dissimilarité interclasse optimale. Si une classe est similaire à une autre alors elle y sera associée et ainsi de suite. Le résultat de cette classification a été spatialisé grâce à un krigeage d'indicatrice (Abzalov, 2016).

Les images issues de la télédétection ont été organisées en 3 rasters distincts. Un raster multidade des 10 bandes spectrales (soit 1590 variables, « bandes X date »), un raster multidade présentant les valeurs de l'indice NDVI (soit 159 variables, « indice X date ») et enfin un raster multidade présentant les valeurs de 5 indices spectraux que sont le NDVI, l'EOMI2¹, le NBR2², l'EVI³ et le S2WI⁴ (soit 795 variables, « indices X date ») (Tableau 1). Par la suite chacun de ces rasters sera dénommé respectivement raster « multi-bandes », raster « NDVI » et raster « multi-indices ».

¹ Exogenous Organic Matter Index 2 (Dodin et al., 2021)

² Normalized Burn Ratio 2 (Miller and Thode, 2007)

³ Enhanced Vegetation Index (Huete et al., 2002)

⁴ Soil Surface Moisture Index (Vaudour et al., 2019)

Tableau 1 : Equation des cinq indices spectraux utilisés et leur nom associé.

<i>Nom de l'indice</i>	<i>Equation de l'indice</i>
NDVI	$\frac{B8 - B4}{B8 + B4}$
NBR2	$\frac{B11 - B12}{B11 + B12}$
S2WI	$\frac{B8A - B11 - B12}{B8A + B11 + B12}$
EVI	$G \times \frac{B8 - B4}{B8 + C1 \times B4 - C2 \times B2 + L}$
Où G est un coefficient pour normaliser l'indice, C1 et C2 sont des coefficients de correction atmosphérique et L est un coefficient compensant les variations saisonnières	
EOMI2	$\frac{B12 - B4}{B12 + B4}$

Les bandes utilisées sont : B2 (réflexion de la lumière bleue) ; B4 (réflexion lumière rouge) ; B8 (réflexion infrarouge moyen) ; B8A (réflexion infrarouge étroit) ; B11 (réflexion infrarouge courte longueur d'onde) et B12 (réflexion infrarouge longue longueur d'onde). Ces bandes sont régulièrement utilisées dans la littérature pour décrire la végétation, les sols ou les paramètres pédologiques.

Les données de résistivité ont été organisées sur une grille de même résolution (10 m X 10 m) que les images issues de Sentinel-2. Cette grille correspond à la résistivité des 50 premiers centimètres du sol (voie 1). Le signal global de cette grille correspond aux bruits, les variations à courtes distances et les variations à longues distances. Lors de la confrontation aux données Sentinel-2 et aux données pédologiques, nous nous sommes focalisés sur les variations à longues distances.

La classification k-means a été introduite par Macqueen (1967). D'autres algorithmes similaires ont été développés par Forgy (1965) (centres mobiles) ainsi que Friedman et Rubin (1967). La classification k-means est une méthode itérative qui, quel que soit son point de départ, converge vers une solution. La solution obtenue n'est pas nécessairement la même quel que soit le point de départ. Pour cette raison, on répète en général plusieurs fois les calculs pour ne retenir que la solution la plus optimale pour le critère choisi. Pour la première itération on choisit un point de départ qui consiste à associer le centre des K classes à K objets (ici des pixels). On calcule ensuite la distance euclidienne entre les objets et les K centres et on affecte les objets aux centres dont ils sont les plus proches. Puis on redéfinit les centres à partir des objets qui ont été affectés aux différentes classes. Puis on affecte à nouveau les objets en fonction de leur distance aux nouveaux centres. Et ainsi de suite jusqu'à ce que la convergence soit atteinte.

Plusieurs critères de classification peuvent être utilisés pour parvenir à une solution. Le déterminant de la matrice de covariance intra-classe commune est un critère nettement moins sensible aux effets d'échelle que le critère trace de la matrice. Par ailleurs, la taille des groupes peut être moins homogène qu'avec le critère de la trace. Afin de déterminer k le nombre de

clusters optimal, un calcul de l'évolution de la variance intra-groupe a été effectué pour chacun des rasters. Le k retenu est 10 pour plusieurs raisons. La variance intra-groupe se stabilise aux alentours de k égal à 10 (Annexe 1). Puis, pour garder une cohérence dans la cartographie des types de sol qui sont estimés à 12 (cf. II.A) et enfin pour qu'un nombre minimum d'individus soient présents au sein d'un cluster.

Les zones homogènes issues des 3 types de données ont dans un premier temps été validées statistiquement par rapport aux données supports de la classification (télédétection, résistivité, pédologie). Dans un second temps, les zones homogènes issues de la télédétection et de la résistivité sont mises en regard de celles issues de la pédologie pour tenter d'expliquer notamment les divergences entre les zones homogènes obtenues à partir de ces deux sources de données auxiliaires. Après avoir obtenu une classification de la réflectance sur les parcelles 29 et 30 simultanément, l'objectif est de comparer cette classification avec la carte des classes des propriétés pédologiques obtenue par krigeage d'indicatrice du résultat de la CAH (cf. Figure 6) afin d'évaluer si les données de télédétection peuvent être utilisées pour la cartographie des propriétés de surface des sols. Cette évaluation se base sur un tableau de contingence qui permet d'obtenir la répartition, pour chaque individu (pixel) des clusters de télédétection, dans un des quatre clusters pédologiques. Des « signatures pédologiques » (répartitions des pixels dans les quatre clusters pédologiques), des clusters issus de la télédétection, significativement différentes les unes des autres indiqueraient que la télédétection permet de discriminer les différentes zones pédologiques. Le raisonnement est le même pour l'évaluation des classifications basées sur les données de résistivité.

III. Résultats

A. Analyse descriptive des données

L'analyse des données de pédologie a mis en évidence des valeurs extrêmes pour l'argile, le carbone, la CEC et les limons grossiers (Tableau 2).

Tableau 2 : Analyse descriptive des données pédologiques

	Minimum	Maximum	Moyenne	Médiane	Coef. d'asymétrie	Nombre d'individus	Unité
Argile	124	279	171,13	171	1,22	56	g/kg
Limons fins	169	443	276,27	264,5	0,56	56	g/kg
Limons grossiers	194	478	358,14	357,5	-0,25	56	g/kg
Sables fins	21	110	77,91	78	-0,43	56	g/kg
Sables grossiers	39	216	103,57	95,5	0,6	56	g/kg
Carbone	8,51	23	14,69	14,05	0,83	56	g/kg
CEC	4,8	13,9	8,31	8,02	0,6	56	cmol+/kg

La médiane et la moyenne ne semblent pas différer fortement or le coefficient d'asymétrie inférieur à 0 indique que pour les limons grossiers et les sables fins, la distribution est étirée vers les valeurs les plus faibles. Pour les autres variables, le coefficient d'asymétrie est supérieur à 0, ce qui peut représenter une distribution étirée vers les valeurs les plus élevées, notamment pour l'argile.

Les valeurs de limons fins et limons grossiers sont situées parmi les valeurs les plus élevées à l'échelle de la France (GisSol).

La matrice de corrélation (Figure 7) révèle des liens linéaires modérés à faibles entre les variables. Toutefois, certaines ne sont pas statistiquement significatives (ex : carbone, sables grossiers).

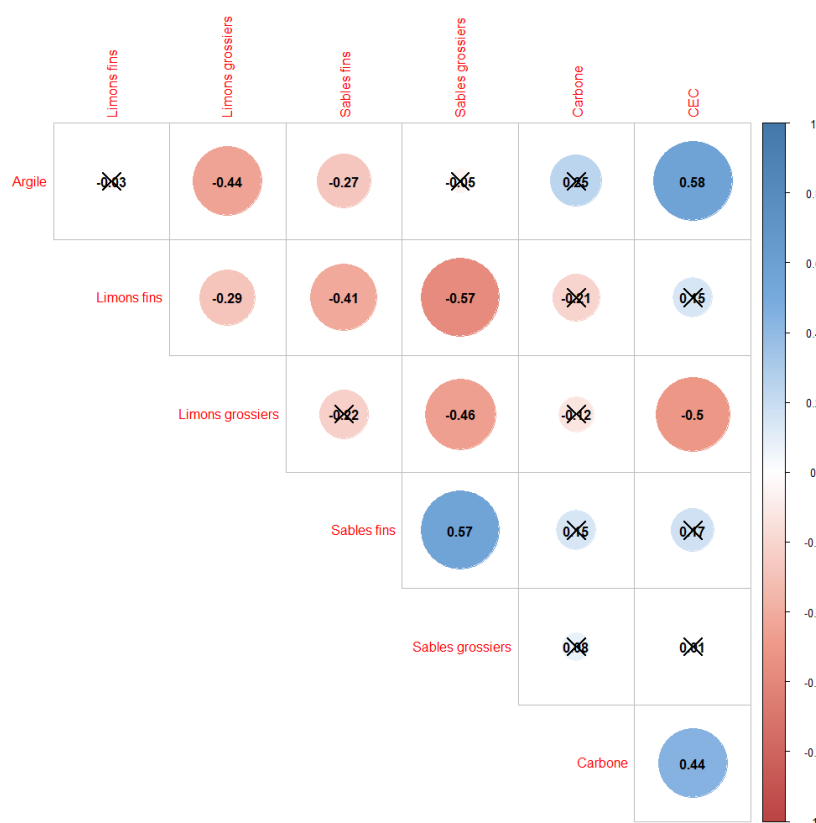


Figure 7 : Matrice de corrélation des variables pédologiques. (Les croix représentent les corrélations qui ne sont pas significatives à l'intervalle de confiance 95%).

Les données de résistivité, soient les 3 voies de résistivité, présentent des valeurs extrêmes (Tableau 3).

Tableau 3 : Analyse descriptive de la résistivité sur l'ensemble du site d'étude sur 3 intervalles de profondeur distincts

	Minimum	Maximum	Moyenne	Médiane	Coefficient d'asymétrie	Nombre d'individus	Unité
Voie 1 (0-50 cm)	1,17	739,20	129,51	118,40	0,71	605165	ohm-mètre
Voie 2 (50-100 cm)	1,68	1314,20	86,09	79,36	1,01	605165	ohm-mètre
Voie 3 (100-150 cm)	1,13	1025,60	53,02	46,70	2,83	605165	ohm-mètre

Le tableau 3 indique que la résistivité diminue en moyenne avec la profondeur. Ceci laisse supposer la présence d'horizons conducteurs en profondeur, possiblement des horizons argileux et donc potentiellement plus humides en profondeur. Ceci est cohérent avec la présence de sols de type planosols sur le site d'étude. Les valeurs fortes de résistivité peuvent être associées à des fragments de roche. En effet, des sols caillouteux ont aussi été identifiés sur le site d'étude.

L'étude statistique des données de télédétection se concentre sur les 10 bandes disponibles pour les 159 images sur l'ensemble du site d'étude (les parcelles 29 et 30). Les intervalles de valeurs pour les 10 bandes ainsi que pour les 5 indices (Tableau 4) permettent de vérifier que les valeurs sont dans les gammes indiquées dans la littérature.

Tableau 4 : Intervalles de valeur pour les 10 bandes spectrales et les 5 indices spectraux.

	B2	B3	B4	B5	B6	B7	B8	B8A	B11	B12	NDVI	NBR2	S2WI	EOMI2	EVI
Min.	0	0	0	0	0	0	0	0	0	0	0	-0,1	-0,6	-0,3	0
Max.	0,2	0,3	0,4	0,4	0,5	0,6	0,6	0,6	0,6	0,5	1	0,5	0,6	1	0,9

Le profil spectral moyen des 10 bandes sur les 159 images (Figure 8) permet d'observer une augmentation du pourcentage de réflectance de 5 (B2) jusqu'à 30 (B8A) puis une diminution jusqu'à 16 (B12). Cela correspond à un profil dominé par la végétation.

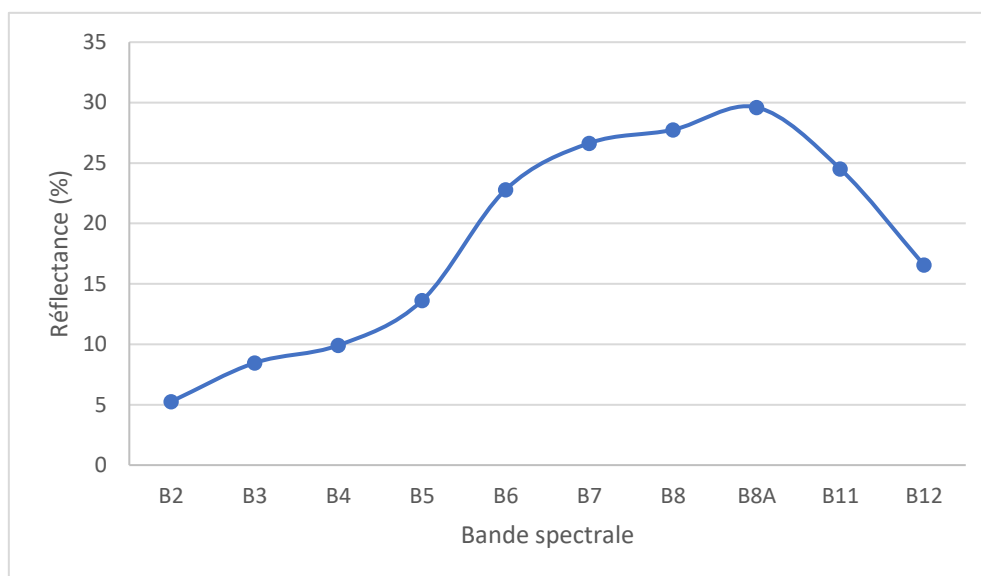


Figure 8 : Profil spectral des 159 images de télédétection à partir des 10 bandes étudiées

L'évolution des valeurs du NDVI (Figure 9) associée aux périodes de cultures renseigne sur les fluctuations de cet indice en fonction de la saison et du type de cultures et indique que la végétation est majoritairement présente au long des 159 dates (images). Le maximum de NDVI est atteint lors de la culture d'orge d'hiver (04/2018) quant au minimum, il est atteint lors de la culture de soja (05/2022). Pour chaque culture le NDVI présente une évolution croissante entre le semis et le pic puis une décroissance entre le pic et la récolte. Les résultats représentant les deux parcelles simultanément, il est important de rappeler qu'à partir du 12/07/2020, les itinéraires culturels des parcelles sont différents. Les fluctuations de NDVI hors culture sont dues aux cultures intermédiaires.

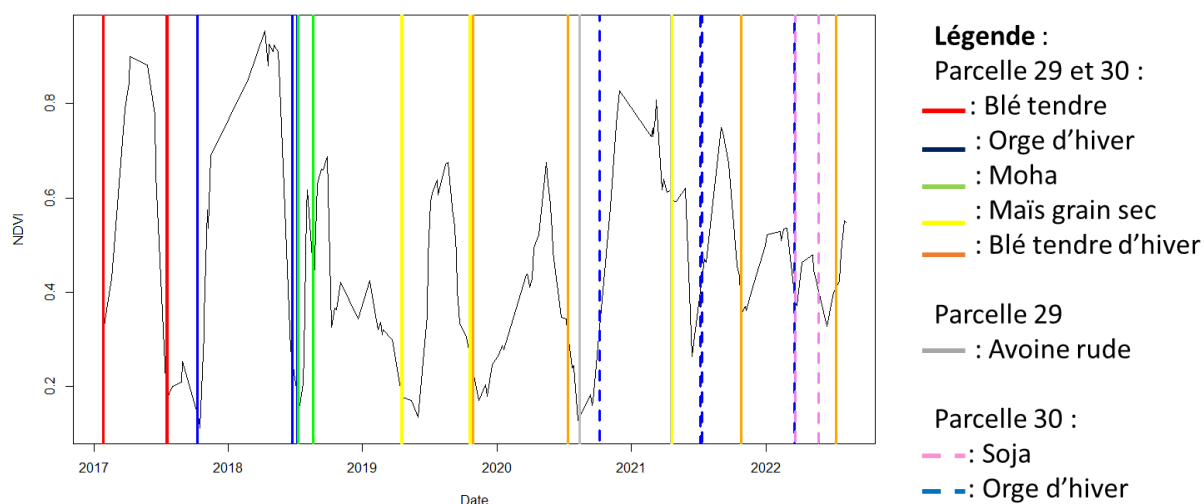


Figure 9 : Evolution temporelle de l'indice spectral NDVI moyen sur l'ensemble de la zone (parcelle 29 + parcelle 30) avec les cultures mises en place sur les deux parcelles. Les résultats représentant les deux parcelles simultanément, il est important de rappeler qu'à partir du 12/07/2020, les itinéraires culturaux des parcelles sont différents.

B. Spatialisation des données

i. Pédologie

La CAH conduite avec les variables pédologiques a révélé 4 groupes composés de 27, 16, 11 et 2 individus (clusters 1, 3, 2 et 4 respectivement) (Figure 10).

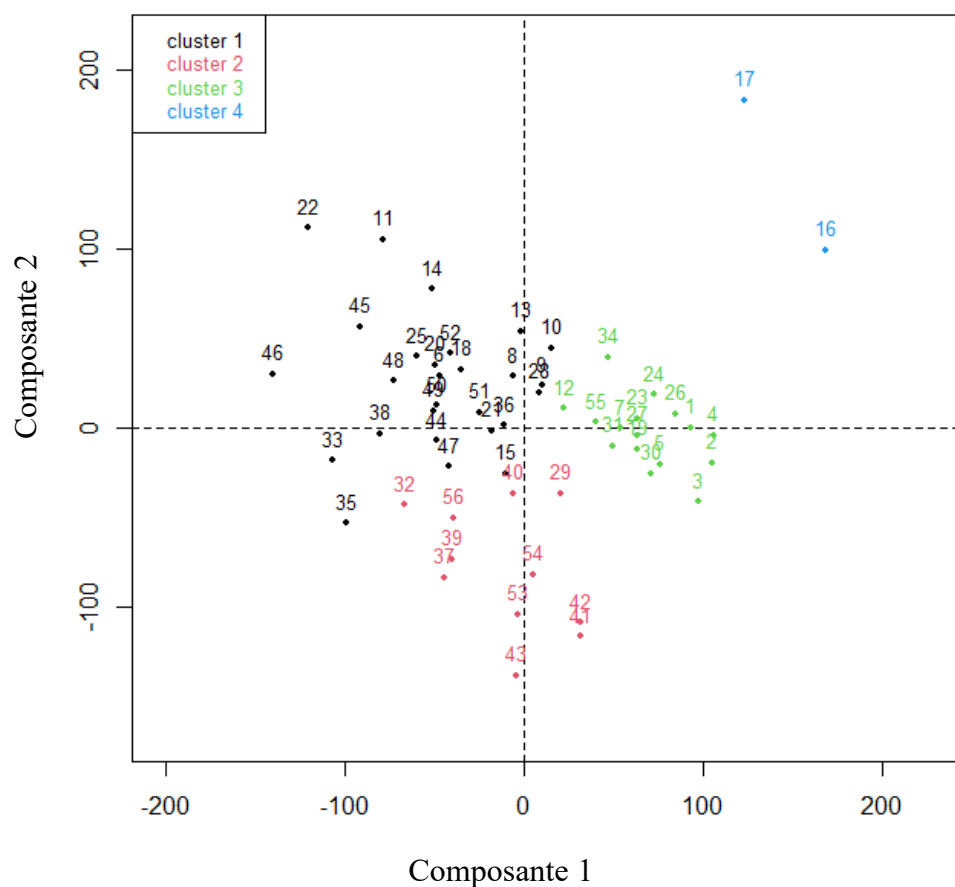


Figure 10 : Agrégation issue de la Classification Ascendante Hiérarchique avec 56 points d'échantillonnage (individus) et 7 variables

Chaque cluster a une signature pédologique propre qui le caractérise et le différencie des autres. La moyenne des 7 variables pour chaque cluster permet de visualiser cette signature pédologique (Figure 11).

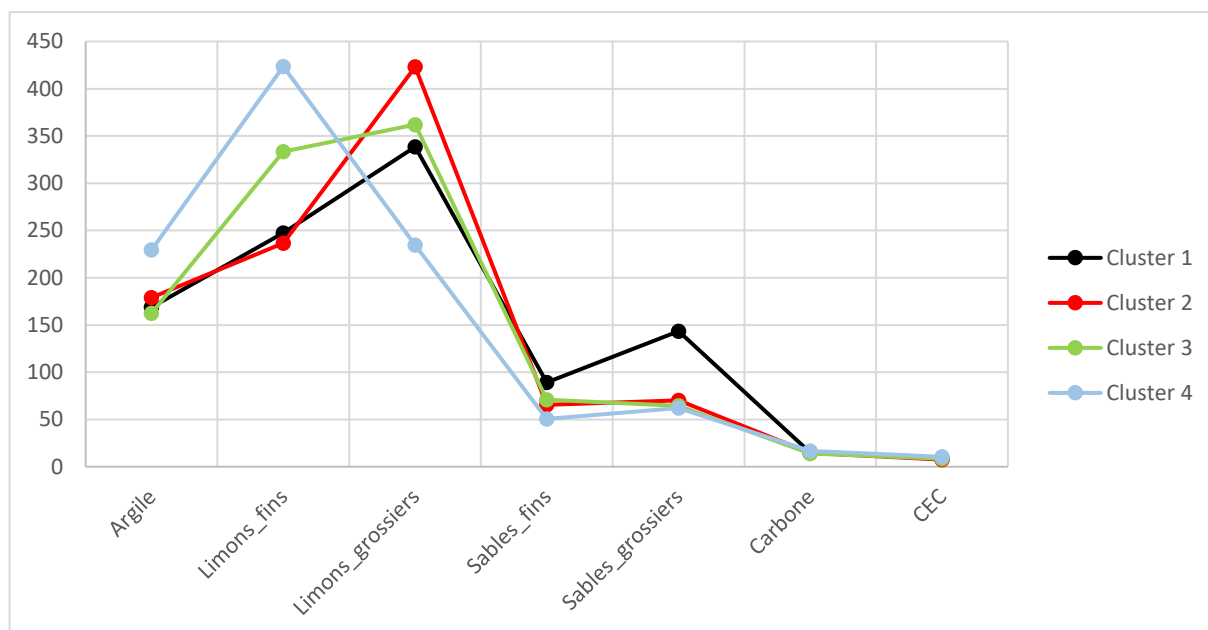


Figure 11 : Profil d'analyse pédologique des 4 clusters pédologiques

Le cluster 1 est caractérisé par une teneur en sables grossiers la plus élevée ainsi qu'une teneur en limons grossiers forte. Le cluster 2 est lui caractérisé par une teneur en limons grossiers la plus élevée. Le cluster 3 est quant à lui caractérisé par des fortes teneurs en limons fins et limons grossiers. Enfin, le cluster 4 est caractérisé par des teneurs en limons fins et en argile les plus élevées. Par conséquent, les clusters se distinguent entre eux par les variables granulométriques en priorité, néanmoins la CEC et le carbone peuvent jouer un rôle eux aussi à la différenciation.

La variable qualitative nominale issue de la CAH a été spatialisée par krigeage d'indicatrice (Figure 12). Ce krigeage nécessite particulièrement l'estimation des variogrammes simples et croisés des 4 classes et leur modélisation par une même combinaison linéaire de variogrammes théoriques.

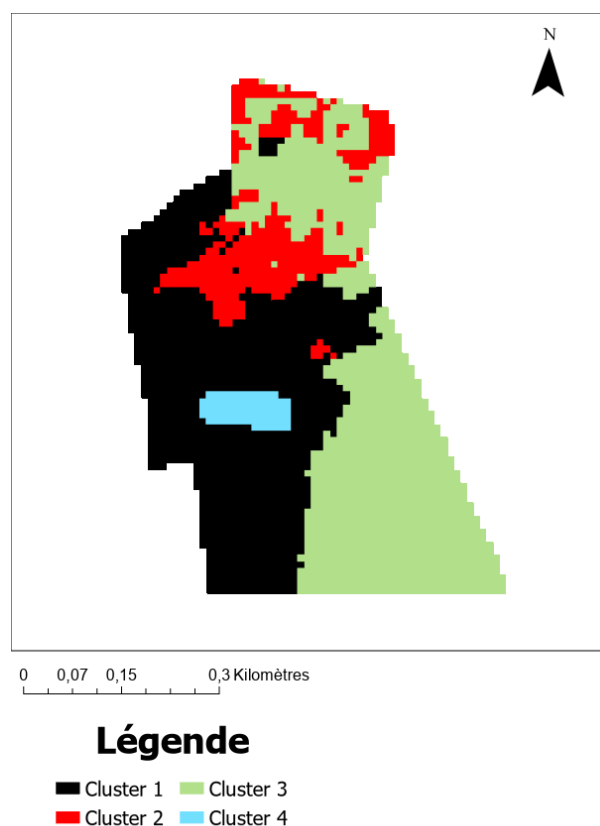


Figure 12 : Spatialisation par krigeage d'indicatrice des classes issues de la CAH menée sur les variables pédologiques

La spatialisation des classes issues de la CAH sur les données pédologiques permet d'observer que le site d'étude est majoritairement composé de deux classes : la classe 1 (sables grossiers) et la classe 3 (limons fins et limons grossiers). La classe 2 (limons grossiers) est plus faiblement représentée et enfin la classe 4 (limons fins et argile) est la moins représentée (Tableau 5).

Tableau 5 : Superficie occupée par les clusters issus de la classification pédologique sur le site d'étude et le pourcentage représenté dans le résultat de la CAH.

	1	2	3	4
Superficie des clusters issues des variables pédologiques (%)	42,8	12,8	41,9	2,5
Pourcentage représenté dans la CAH (%)	48,2	19,6	28,6	3,6

Seul le cluster 3 a été plus représenté lors de la spatialisation par rapport au pourcentage qu'il représentait dans le résultat de la CAH.

ii. Résistivité

L'estimation et la modélisation du variogramme de la résistivité électrique, organisée sur une grille régulière de 10 m de résolution, a permis de décomposer (grâce au krigeage des

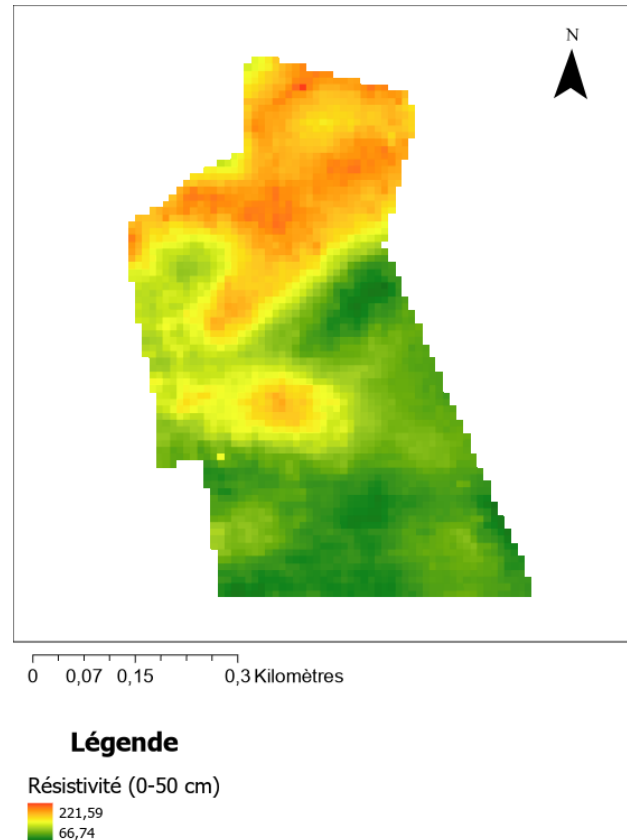


Figure 13 : Spatialisation de la composante de longue portée de la résistivité (0-50 cm) en ohm.mètre

composantes), cette variable en une composante pépétique, une composante représentant les hautes fréquences et une composante qui représente les basses fréquences. Cette dernière (Figure 13) qui indique les grandes tendances de la résistivité électrique sur le site d'étude est retenue pour l'étape de comparaison aux autres types de données disponibles.

C. Classification des données

Les résultats de la classification sur la télédétection sont au nombre de trois, représentant une classification pour le raster « NDVI », le raster « multi-indices » et le raster « multi-bandes » (Figure 14).

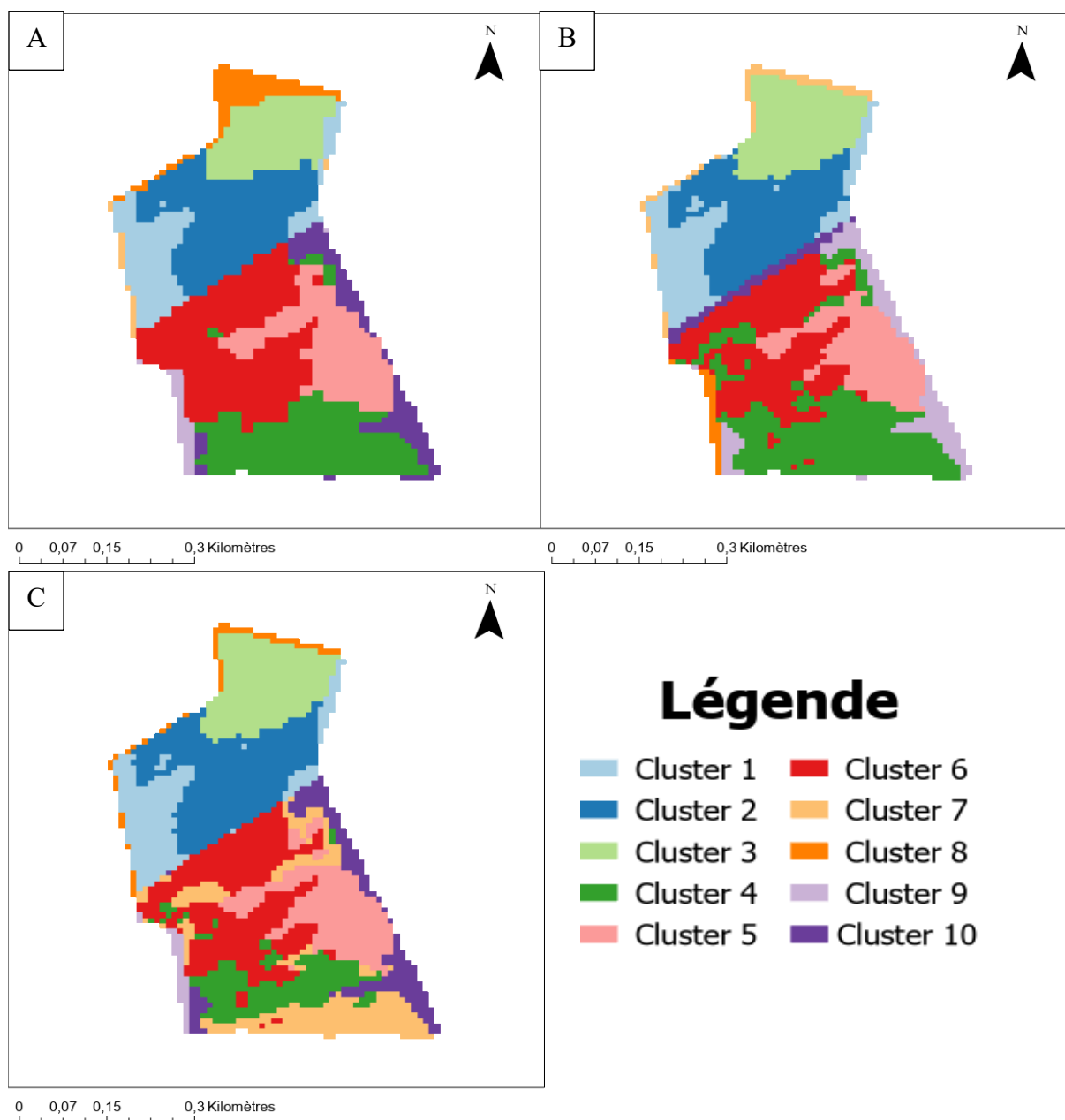


Figure 14 : A- Classification du raster "multi-bandes". B- Classification du raster "multi-indices". C- Classification du raster "NDVI". (La légende est équivalente pour les 3 classifications)

Les dix classes des 3 rasters représentent chacune une surface (Tableau 6) différente. Pour la classification « multi-bandes », les clusters 1, 2, 3, 4, 5 et 6 représentent la majorité du site d'étude, les clusters restants étant peu représentés, à l'exception du cluster 10 qui occupe une surface conséquente en bordure de la parcelle 30. La classification « multi-indices » est aussi représentée majoritairement par 6 clusters (1, 2, 3, 4, 5 et 6). Quant à la classification « NDVI », elle est majoritairement représentée par 7 clusters (1, 2, 3, 4, 5, 6 et 7).

Tableau 6 : Répartition des clusters issus de la classification sur la télédétection sur le site d'étude exprimé en pourcentage de surface

	1	2	3	4	5	6	7	8	9	10
<i>Multi-bandes</i>	9,4	16,9	9,6	16,2	12,4	22,2	0,8	4,4	1,6	6,5
<i>Multi-indices</i>	9,7	16,6	11,7	20,1	11,2	18,4	2,4	1,1	6,5	2,3
<i>NDVI</i>	9,6	17,2	11,9	9,1	11,9	11,4	18,4	2,2	1,2	7,0

Les clusters, significatifs en termes de surface, ont été étudiés par rapport à la variable « NDVI », plus facilement interprétable que les autres variables, sur les 3 rasters (Tableau 7). Ces résultats apportent une précision sur la répartition des clusters au sein des classifications. Pour la classification « multi-bandes » les clusters 4, 5 et 6 représentent les valeurs de NDVI les plus élevées et sont situés sur la même parcelle tandis que les clusters 1, 2 et 3 représentent des valeurs de NDVI plus faibles et sont eux aussi situés sur la même parcelle. Il y a donc une séparation de valeurs entre les deux parcelles pour la classification « multi-bandes ». Le même constat a été fait pour les deux autres classifications qui suivent elles aussi cette même organisation.

Tableau 7 : Médianes des valeurs de l'indice spectral NDVI au sein des différents clusters significatifs des 3 classifications de la télédétection

	1	2	3	4	5	6	7	8	9	10
<i>Multi-bandes</i>	0,36	0,38	0,39	0,43	0,48	0,47	/	/	/	/
<i>Multi-indices</i>	0,36	0,38	0,39	0,43	0,48	0,48	/	/	/	/
<i>NDVI</i>	0,36	0,38	0,39	0,45	0,41	0,49	0,48	/	/	/

Par ailleurs, le k -means appliqué à la composante des basses fréquences de la résistivité (composante longue portée) a conduit à 5 classes (Figure 15) en se basant sur le critère de la variance intra-groupe (Annexe 1).

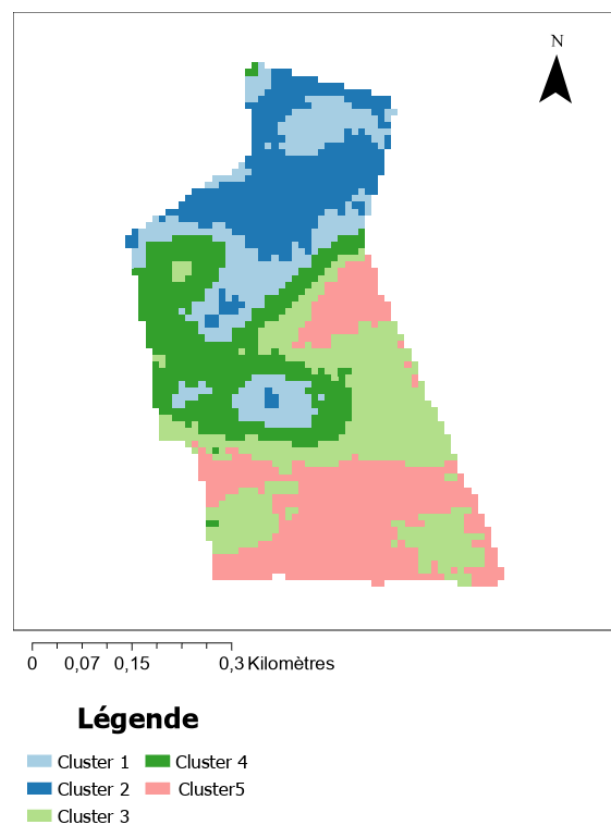


Figure 15 : Clustering de la résistivité électrique (0-50 cm) à partir des données de la composante longue portée

Les clusters de la classification sur la résistivité (0-50 cm) sont de proportion équivalente dans l'ensemble (Tableau 8). Certains clusters ne sont pas concentrés sur un endroit précis et occupent une zone plus étendue en différentes lentilles.

Tableau 8 : Répartition des clusters issus de la classification de la résistivité sur le site d'étude exprimé en pourcentage de surface

	1	2	3	4	5
Résistivité (0-50 cm)	15,6	16,7	23,7	18,7	25,3

Enfin, la même méthode de description que pour les clusters de télédétection a été appliquée sur les clusters de la résistivité (Tableau 9).

Tableau 9 : Médianes des valeurs de résistivité électrique (0-50 cm) au sein des cinq clusters issus de la classification de la résistivité (0-50cm)

	1	2	3	4	5
Résistivité électrique (0-50 cm)	164,89	187,37	101,87	132,33	84,61

La partie Nord du site d'étude (cluster 1 et 2) est caractérisée par les valeurs de résistivité (0-50 cm) les plus fortes. Le cluster 4, qui occupe la zone entre les deux parcelles ainsi que la partie Est, est caractérisé par des valeurs de résistivité plus intermédiaires. Enfin les clusters 3 et 5, qui occupent la majorité Sud du site d'étude, sont en moyenne plus conducteurs.

D. Evaluation statistique des classifications sur les données sources

i. Télédétection

Une ANOVA (analyse de la variance) appliquée aux classes issues de *k-means* menée sur le raster « NDVI » révèle que les classes 6 et 5 sont statistiquement similaires ($p\text{-value} = 0,95$) (Figure 16A). Ces deux classes se situent dans la parcelle 30 et sont proches. Cette proximité peut expliquer les moyennes statistiquement égales. La différenciation est alors peut-être due à une des écarts de NDVI à des dates différentes.

La même analyse (ANOVA) a été appliquée sur la classification du raster « multi-bandes » où les variables étudiées sont les 10 bandes spectrales (Figure 16B). Les clusters 7, 8, 9 et 10 étant situés sur les bords et/ou avec peu d'individus ne sont pas détaillés dans la suite. Pour les 6 clusters restants il existe donc 2 cas de figure. Les clusters sont situés dans la même parcelle ou bien chaque cluster est sur une parcelle différente. Dans le premier cas de figure, il est possible que sur les bandes concernées, une distinction ait été faite sur des dates différentes mais résulte en une moyenne équivalente. Ceci se présente pour les clusters 1-3 et les clusters 5-4, où sur la bande spectrale B5 leur $p\text{-value}$ est de 0,39 et 0,99 respectivement. Dans le second cas de figure, cela est plus complexe en partant du postulat que les clusters ne sont pas situés sur la même parcelle il n'y a pas d'explications limpides pour ce cas de figure. Néanmoins, les moyennes significativement égales n'étant pas pour toutes les bandes, il est possible que sur certaines longueurs d'ondes des clusters issus de parcelles différentes, elles soient identiques. Ce second cas de figure est visible pour les clusters 6-2 et les clusters 4-3, où sur la bande spectrale B6 leur $p\text{-value}$ est égal à 0,58 et 0,99 respectivement. Chaque bande représentant un phénomène précis alors il est possible qu'on retrouve ces phénomènes sur deux clusters n'étant pas à côté mais que d'autres phénomènes les séparent plus fortement.

La classification du raster « multi-indices » a été faite à partir des 5 indices et a été étudiée statistiquement selon la même méthode que les rasters précédents. Les clusters 7, 8, 9 et 10 sont jugés non représentatifs sur les parcelles (peu d'individus et/ou sur les bords des parcelles). Seuls les clusters 3-2 et 5-6 ont présenté des moyennes significativement égales selon l'ANOVA (Figure 16C) tout en étant sur la même parcelle. Pour le premier duo, les clusters ont des moyennes significativement égales pour l'indice NDVI, avec une $p\text{-value}$ égale à 0,99. Le second duo est composé de clusters qui ont des moyennes significativement égales pour l'indice NBR2, avec une $p\text{-value}$ égale aussi à 0,99. Les clusters étant côte à côte, il est possible qu'ils soient similaires pour un indice mais que pour les indices restants ils se différencient significativement. Dans cette classification, aucun cluster situé sur une parcelle différente d'un autre a présenté des moyennes significativement égales sur l'ensemble des variables.

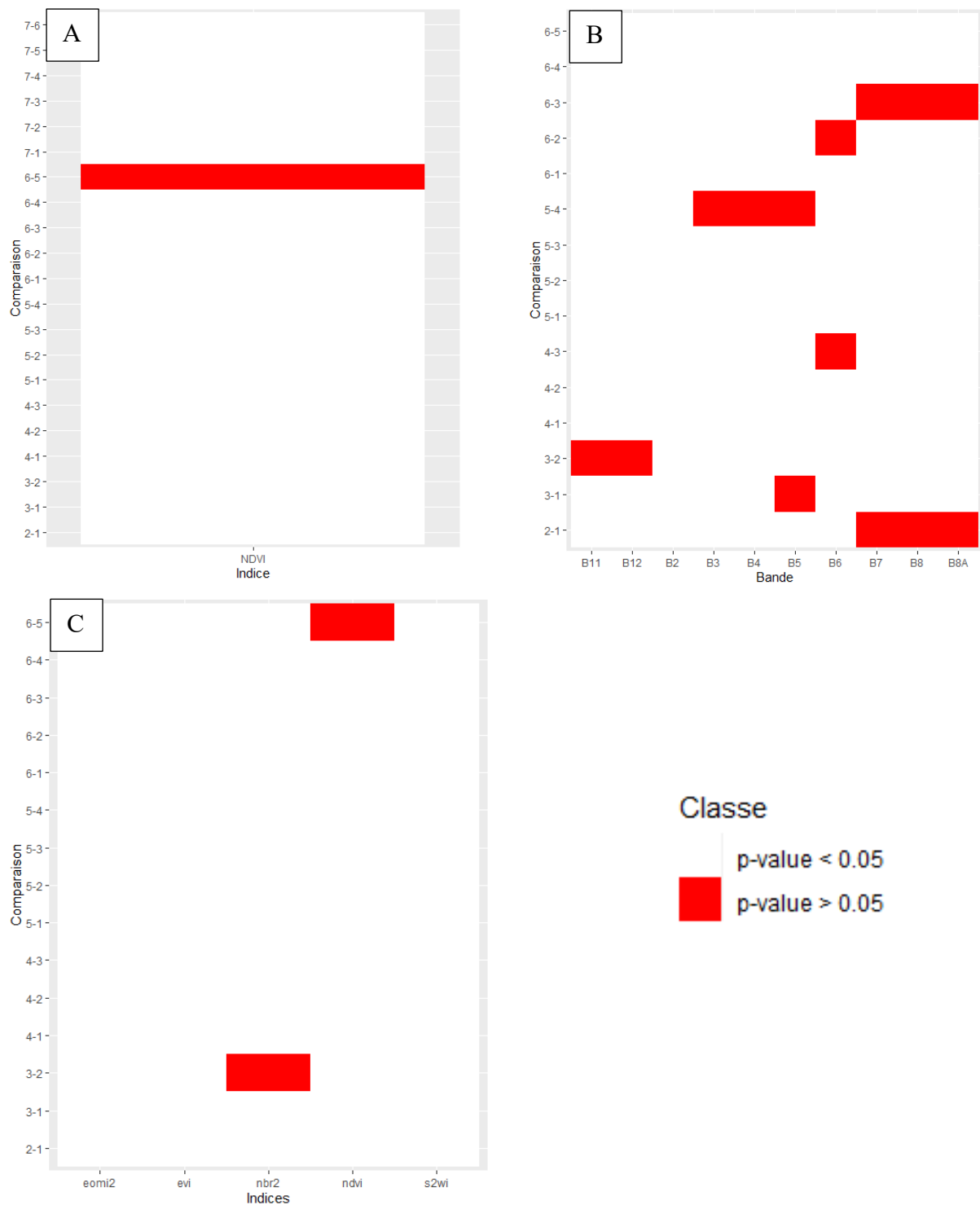


Figure 16 : Représentation de la significativité de l'ANOVA entre les clusters significatifs. A- Classification raster "NDVI". B- Classification sur le raster "multi-bandes". C- Classification sur le raster "multi-indices".

Les classifications effectuées sur les 3 rasters de télédétection sont validées au vu des résultats de l'évaluation statistique, qui confirme que les clusters sont bien différents entre eux selon les variables utilisées.

ii. Pédologie

Les moyennes des 7 variables pédologiques classe par classe ont été comparées (Tableau 10) par le test de Wilcoxon.

Tableau 10 : Comparaison des moyennes des variables pédologiques par paire de classes. Vert : $\mu_1 = \mu_2$. Orange : $\mu_1 \neq \mu_2$.

	Argile	Limons fins	Limons grossiers	Sables fins	Sables grossiers	Carbone	CEC
1-2							
1-3							
1-4							
2-3							
2-4							
3-4							

Seuls les limons grossiers sont significativement différents dans tous les clusters, suivis des limons fins où seul les clusters 1-2 ont des moyennes significativement identiques. Les sables fins sont déterminants pour 4 paires de clusters sur 6 et les sables grossiers pour 3 paires de clusters sur 6 et distinguent principalement le cluster 1 des autres. L'argile, le carbone et la CEC ne sont pas significativement déterminants pour la différenciation des clusters. Par conséquent, en se référant au profil pédologique (cf. III.B.i), le cluster 1 est caractérisé par les sables grossiers et est bien différencié par cette variable lors du test de Wilcoxon par rapport aux autres clusters. Le cluster 2 est lui caractérisé par les limons grossiers et est bien différencié par cette variable aussi. Le troisième cluster se caractérise moins fortement que les autres clusters néanmoins les limons fins ainsi que les limons grossiers sont les variables qui le caractérisent le plus. Il est significativement différent des autres clusters sur ces 2 variables. Enfin, le dernier cluster caractérisé par l'argile et les limons fins se différencie significativement uniquement sur les limons fins. Par conséquent, le test de Wilcoxon confirme l'hypothèse du profil pédologique des clusters.

iii. Résistivité

Selon le test de Wilcoxon, les moyennes de la résistivité de la composante longue portée de la voie 1 (0-50 cm) sont significativement différentes entre tous les clusters (Annexe 3).

E. Signification pédologique des classifications sur les données de télédétection et de résistivité

L'évaluation a été faite pour les 3 rasters issus des *k-means* sur les données de télédétection (Tableau 11 à 13) ainsi que pour le raster issu du *k-means* sur les données de la résistivité (Tableau 14). Les clusters localisés sur les bords des parcelles et/ou ayant peu d'individus ne sont pas analysés. Si deux clusters présentent des signatures pédologiques proches, alors un test de χ^2 a été effectué pour vérifier si les compositions sont significativement identiques. Ce test a été effectué uniquement lorsque les pourcentages d'individus sur les deux clusters les plus représentés étaient proches de moins de dix pourcents (Annexe 4).

Tableau 11 : Tableau de contingence de la classification du raster "multi-bandes". (* signifie qu'un test Khi² a été effectué.)

	1 (Sables grossiers)	2 (Limons grossiers)	3 (Limons fins et limons grossiers)	4 (Limons fins)
1	15,0	14,2	2,6	0
2	17,0	55,7	5,8	0
3	0,9	13,6	18,1	0
4	11,2	0	27,2*	6,9
5	5,2	0	22,8*	30,6
6	40,8	2,0	6,2	62,5

Pour la classification à partir du raster « multi-bandes », le cluster pédologique 1 (sables grossiers) est représenté en majorité par le cluster de télédétection 6 (NDVI élevé). Le cluster pédologique 2 (limons grossiers) est lui représenté en majorité par le cluster de télédétection 2 (NDVI faible). Le cluster pédologique 3 (limons fins/limons grossiers) est représenté en légère majorité par le cluster de télédétection 4 (NDVI moyen) suivi du cluster de télédétection 5 (). Enfin, le cluster pédologique 4 (limons fins) est représenté en majorité par le cluster de télédétection 6 (NDVI élevé).

Les tests Khi² ont montré que les proportions entre les clusters de télédétection 5 et 4 sont significativement différentes. Par conséquent, bien que composés d'assemblages de propriétés pédologiques de surface communes, les clusters semblent significativement différents d'un point de vue pédologique.

Tableau 12 : Tableau de contingence de la classification du raster "NDVI". (* signifie qu'un test Khi² a été effectué.)

	1 (Sables grossiers)	2 (Limons grossiers)	3 (Limons fins et limons grossiers)	4 (Limons fins)
1	15,7	13,6	2,6	0
2	16,9	55,7	6,6	0
3	1,0	22,0	20,9*	0
4	11,0	0	10,7	0
5	7,6	0	18,6	36,1*
6	4,7	0,6	21,0*	26,4
7	34,4	1,4	5,5	37,5*

La classification à partir du raster « NDVI », le cluster pédologique 1 (sables grossiers) est représenté en majorité par le cluster de télédétection 7 (NDVI élevé). Le cluster pédologique 2 (limons grossiers) est lui représenté en majorité par le cluster de télédétection 2 (NDVI faible). Le cluster pédologique 3 (limons fins/limons grossiers) est représenté en légère majorité par le cluster de télédétection 6 (NDVI le plus élevé) suivi du cluster de télédétection 3 (NDVI faible/moyen). Enfin, le cluster pédologique 4 (limons fins) est représenté en légère majorité par le cluster de télédétection 7 (NDVI élevé) et le cluster de télédétection 5 (NDVI moyen).

Les tests Khi² ont montré que les proportions entre les clusters de télédétection 6 et 3 sont significativement différentes. Par conséquent, bien que composés d'assemblages de propriétés pédologiques de surface communes, les clusters semblent significativement

différents d'un point de vue pédologique. Le même résultat a été trouvé pour le cluster pédologique 4 avec les clusters de télédétection 7 et 5.

Tableau 13 : Tableau de contingence de la classification du raster "multi-indices". (* signifie qu'un test χ^2 a été effectué.)

	1 (Sables grossiers)	2 (Limons grossiers)	3 (Limons fins et limons grossiers)	4 (Limons fins)
1	15,3	13,9	3,1	0
2	16,0	55,4	6,0	0
3	1,0	21,4	20,7	0
4	16,2	0	29,9*	30,6
5	4,0	0,3	21,4*	25,0
6	33,9	1,7	5,5	44,4

La classification à partir du raster « multi-indices », le cluster pédologique 1 (sables grossiers) est représenté en majorité par le cluster de télédétection 6 (NDVI le plus élevé). Le cluster pédologique 2 (limons grossiers) est lui représenté en majorité par le cluster de télédétection 2 (NDVI faible). Le cluster pédologique 3 (limons fins/limons grossiers) est représenté en majorité par le cluster de télédétection 4 (NDVI moyen). Enfin, le cluster pédologique 4 (limons fins) est représenté en majorité par le cluster de télédétection 6 (NDVI le plus élevé).

Les tests χ^2 ont montré que les proportions entre les clusters de télédétection 4 et 5 sont statistiquement différentes. Par conséquent, bien que composés d'assemblages de propriétés pédologiques de surface communes, les clusters semblent significativement différents d'un point de vue pédologique.

Tableau 14 : Tableau de contingence de la classification du raster "résistivité". (* signifie qu'un test χ^2 a été effectué.)

	1 (Sables grossiers)	2 (Limons grossiers)	3 (Limons fins et limons grossiers)	4 (Limons fins)
1	12,3	36,2*	9,4	63,9
2	5,2	44,3*	20,7	6,9
3	23,0	2,9	32,6*	0
4	35,7	15,1	1,2	29,2
5	23,7	1,4	36,1*	0

La classification à partir du raster « résistivité », le cluster pédologique 1 (sables grossiers) est représenté en majorité par le cluster de résistivité 4 (résistivité moyenne). Le cluster pédologique 2 (limons grossiers) est lui représenté en majorité par le cluster de résistivité 2 (résistivité forte). Le cluster pédologique 3 (limons fins/limons grossiers) est représenté en légère majorité par le cluster de résistivité 5 (faible résistivité) suivi du cluster de résistivité 3 (faible résistivité). Enfin, le cluster pédologique 4 (limons fins) est représenté en majorité par le cluster de résistivité 1 (résistivité forte).

Les tests χ^2 ont montré que les proportions entre les clusters de résistivité 1 et 2 sont significativement différentes. Par conséquent, bien que composés d'assemblages de propriétés pédologiques de surface communes, les clusters semblent significativement différents d'un

point de vue pédologique. Les proportions des clusters de résistivité 3 et 5 sont, elles, statistiquement égales. Par conséquent, ces clusters sont composés significativement des mêmes propriétés pédologiques de surface et ne sont pas représentatifs d'un cluster pédologique en particulier.

IV. Discussion

A. Incertitude sur les données

Les trois sources de données utilisées dans cette étude sont des images de télédétection, des données de résistivité électrique de surface (0-50 cm) et des données pédologiques issus de deux laboratoires d'analyses distincts. Ces données présentent des incertitudes sur leur origine ou sur leur traitement. Ces incertitudes ont pu fausser certains résultats présentés ci-dessus et cela à différentes étapes d'analyse.

i. Incertitudes sur les mesures

Les données pédologiques sont issues de deux laboratoires d'analyse différents et sont donc potentiellement biaisées par des erreurs d'analyses différentes entre laboratoire.

Les données de résistivité ont quant à elles été récupérées dans des conditions jugées bonnes mais il est possible que le peu d'humidité présente ait pu fausser légèrement les résultats en diminuant les valeurs de résistivité électrique obtenues.

Quant aux données de télédétection, l'incertitude peut se situer sur des paramètres physiques de l'atmosphère. En effet, l'humidité présent dans l'air peut impacter la prise de vue (Pimstein et al., 2011 ; Liu et al., 2022) ou encore les nuages présents peuvent altérer la qualité de l'image lors de la prise de vue. Somers et al. (2010) ont démontré que l'humidité dans le sol peut influencer fortement sur la réflectance. Par ailleurs, l'état de surface des sols (résidus, labour, etc.) peut aussi impacter la réflectance des sols (Gholizadeh et al., 2018).

De plus, un des biais de la télédétection dans notre étude a été la répartition temporelle des images. En effet, la distribution des images n'est pas équivalente en tout point de l'année due aux nuages ou autres inconvénients atmosphériques. De ce fait, certains mois, ou plus précisément certaines périodes de l'évolution phénologique sont privilégiés. Cela peut alors avoir comme répercussion de concentrer la classification sur des propriétés des sols identifiables à ces périodes mais de négliger d'autres propriétés des sols, faute d'images disponibles durant les périodes clés d'observation de ces propriétés. En d'autres termes, il y a un biais d'observation, puisque nous ne classifions que ce que nous pouvons observer.

ii. Incertitudes induites par le prétraitement

La spatialisation des résultats de la pédologie a été faite à partir d'un krigeage d'indicatrice. Le résultat de la spatialisation peut être impacté par (i) le modèle de variographie estimé et ajusté à partir de seulement 56 individus, (ii) le risque de classification erronée inhérente à la CAH en amont du krigeage.

La spatialisation des données de résistivité a débuté par une réduction de la quantité de d'informations en transformant des points en une grille de même résolution que les images Sentinel-2. Par la suite, uniquement la composante longue portée a été gardée afin de lisser les données pour le clustering. Il faut par conséquent prendre en compte que la classification de la résistivité représente uniquement les variations de longue portée, et que cela a lissé le résultat final de la classification. L'ensemble pourrait être à l'origine d'autres incertitudes pouvant impacter le résultat final.

B. La classification des données

i. Méthode de classification

La méthode de classification des données n'est pas identique pour tous les jeux de données, les données de pédologie étant classées par une CAH contrairement aux données de télédétection et de résistivité qui ont été classées selon la méthode des *k-means*. Les algorithmes ne fonctionnant pas de la même manière, la CAH étant un clustering hiérarchique et la méthode *k-means* étant un clustering partitionné (Saxena et al., 2017). Les méthodes hiérarchiques fonctionnent en structure d'arbre qui regroupent les clusters selon leurs similitudes tandis que les méthodes partitionnées divisent les données en des groupes distincts après plusieurs itérations. La méthode des *k-means* est mieux adaptée pour les grands jeux de données comme la télédétection et la résistivité contrairement à la CAH (Gelb and Apparicio, 2021) et la CAH est très couramment utilisée pour les jeux de données de dimension modeste. La CAH a donc été choisie pour rassembler dans des classes les points aux caractéristiques pédologiques similaires au vu du nombre d'individus. En tout état de cause, les résultats de la CAH sur les données pédologiques ont été consolidés par un *k-means* conduisant au même résultat.

Une des limites du *k-means*, utilisé pour la télédétection et la résistivité, est que la dépendance temporelle d'une image par rapport à une autre n'est pas intégrée. Ceci peut entraîner des biais dans le résultat en exagérant les caractères communs de deux individus. En effet, il y a un effet de saisonnalité dans les variables dû aux pratiques culturales. En d'autres termes, le *k-means* considère les images dans un ordre aléatoire, ce qui supprime l'information relative à l'évolution temporelle de la réflectance pourtant bien présente (Figure 9). De plus, *k-means* a le défaut d'inclure des points au sein d'un cluster même si ils sont éloignés l'un de l'autre (Saxena et al., 2017) ce qui peut amener un biais dans une étude comme celle-ci qui essaye de créer des groupes les plus homogènes possibles tout en étant restreints spatialement.

ii. Classifications

La classification sur le raster « multi-indices » est difficilement interprétable de par sa complexité. De fait, les indices représentent un phénomène propre à chacun et il est par conséquent compliqué d'en expliquer chacun des aspects sur une classification. Certains indices présentent des corrélations entre eux pouvant impacter le résultat final. Néanmoins le résultat n'en reste pas moins intéressant à comparer avec la classification issue des variables pédologiques. Le choix des indices a été fait de manière arbitraire, si un autre quinté d'indices avait été utilisé le résultat aurait été différent, par conséquent cette classification « multi-indices » est perfectible sur le choix des indices afin d'obtenir une meilleure explication de la variabilité pédologique. Enfin, les indices ne représentent pas la réalité du terrain mais uniquement un phénomène de manière brute en utilisant des bandes spectrales (Hatfield et al., 2019)

Les classifications des données ne présentent pas le même nombre de classes (4, 5, 10 pour la pédologie, la résistivité et la télédétection, respectivement). Cela peut s'expliquer par la quantité d'informations disponibles au départ mais à partir de ce postulat la résistivité aurait dû être classifiée en un nombre de classes bien supérieur au vu de la quantité d'informations qu'elle contient. Tout comme la pédologie qui aurait pu présenter une forte variabilité et par conséquent être classée en un nombre de classes supérieur. Le nombre de classe est représentatif de l'hétérogénéité, ou l'homogénéité, des données. En effet, plus le nombre de classes est élevé plus les données sont différentes entre elles selon un critère de distance ou un paramètre précis. Mais, le nombre de classe ne peut pas être estimé à k égal n individus car cela n'aurait pas de réel sens scientifique. Par conséquent, pour trouver le nombre de classe k idéal, il est souvent nécessaire d'avoir recours à un critère de jugement. Dans cette étude, c'est la variance intra-

classe qui a été choisie. Ce paramètre atteint un palier, c'est-à-dire qu'il ne variera plus après un certain k_{palier} et alors $k_{optimal}$ équivaut à k_{palier} . Par conséquent, avec $k_{optimal}$ les données sont censées être regroupées le plus correctement possible. Néanmoins, le nombre de classes optimal est compliqué à déterminer avec certitude et peut toujours être modifié à partir des connaissances sur les données. Pour conclure, le nombre de classes utilisé dans cette étude pour chaque source de données n'est pas fixe et peut être amélioré avec d'autres critères ou par l'ajout d'autres variables.

La résistivité a été classifiée selon 5 classes or la résistivité est une variable qui est dépendante de plusieurs facteurs comme l'humidité ou la teneur en eau dans le sol (Zhang et al., 2022). La présence de planosol peut amener à une teneur en eau élevée (Baize and Girard, 2009) entraînant des baisses de la résistivité à certains endroits localisés. De ce fait la résistivité peut être classée selon un nombre de classes plus élevé, d'où l'utilisation d'une étape de filtrage afin de lisser les données pour réduire le nombre de classes.

La télédétection utilise la réflectance des objets, or cette réflectance ne représente que la surface d'un objet. Néanmoins, dans un sol, les phénomènes de profondeur peuvent avoir une incidence sur les phénomènes de surface et réciproquement. Par conséquent, il est possible que la télédétection nous renseigne certes sur des phénomènes de surface, mais aussi sur des phénomènes plus profonds comme l'humidité et la fertilité des sols en profondeur. L'humidité en profondeur peut impacter la couleur du sol en surface, et la fertilité des sols en profondeur peut impacter la croissance végétale observée. La télédétection peut donc permettre dans une moindre mesure d'obtenir des informations sur des horizons profonds des sols. Il faut tout de même nuancer cette hypothèse, car pour expliquer cela, une analyse rigoureuse et complexe doit être mise en place. Enfin, pour compléter cette hypothèse il aurait été intéressant de se concentrer sur des variables en profondeur ou représentant cette dernière. La pédologie a été classifiée principalement sur la granulométrie accompagnée du carbone et de la CEC. Or si l'objectif était de catégoriser des types de sol, il aurait été judicieux d'ajouter d'autres variables comme la profondeur du sol ainsi que des données sur les horizons profonds qui peuvent permettre d'obtenir une idée plus précise sur ces phénomènes en profondeur. La résistivité, quant à elle, contient des informations jusqu'à 150 cm de profondeur ce qui peut permettre aussi d'étudier ces phénomènes en profondeur.

Enfin, la rampe d'irrigation ainsi que les actions entreprises par les gérants de parcelle (épandage par exemple) ont pu nettement impacter la réflectance du sol à certains moments ce qui peut avoir comme répercussion des diminutions de la réflectance.

C. Signification de la comparaison télédétection/pédologie et résistivité/pédologie

La classification de la télédétection et la classification de la résistivité ont montré des résultats similaires, c'est-à-dire que les clusters pédologiques 1 (sables grossiers), 2 (limons grossiers) et 4 (limons fins) sont bien expliqués par des clusters de télédétection ou de résistivité. Seul le cluster pédologique 3, qui est le seul cluster pédologique à ne pas se différencier nettement des autres par une variable prédominante, a nécessité des tests χ^2 . En effet, dans le cluster pédologique 3, les répartitions sont souvent similaires ce qui ne permet pas de définir un cluster de la télédétection ou de la résistivité comme prédominant en son sein. Ce cluster est étendu sur l'ensemble du site d'étude, ce qui a pu rendre complexe l'identification de cluster de télédétection ou de résistivité s'y rattachant de façon prédominante.

Les clusters pédologiques peuvent être globalement expliqués à 50% minimum avec deux clusters de télédétection ou de résistivité. Ce pourcentage peut aller jusqu'à 95% pour le cluster pédologique 4 (limons fins) avec la classification sur le raster « multi-bandes ». De ce

fait, la télédétection ainsi que la résistivité peuvent être utilisés pour expliciter la variabilité pédologique dans certaines conditions. En effet, le nombre de clusters n'étant pas identiques entre la télédétection (ou la résistivité) et la pédologie, il est logique d'obtenir des pourcentages explicites de cet ordre. Néanmoins, comme il est dit précédemment, deux clusters suffisent pour expliciter au minimum 50%. Ainsi, sur la base de cette hypothèse, en regroupant certains clusters entre eux, par conséquent en diminuant le nombre k de clusters, il est probable que la classification de la télédétection ainsi que de la résistivité puisse mieux représenter la variabilité pédologique. Mais, il faut alors garder en principe fondamental qu'en réduisant le nombre de clusters, on réduit l'explication de l'information. Par conséquent, un équilibre doit être trouvé entre l'explication de la variable de base et l'explication de la variabilité pédologique. À cela, il faut aussi ajouter l'hypothèse de variabilité pédologique qui peut changer. En effet, en rajoutant des variables sur les horizons profonds ou même d'autres variables pédologiques, il est probable d'obtenir une classification plus détaillée et alors la comparaison entre télédétection/résistivité et pédologie s'en trouvera impactée avec une possible convergence du nombre optimal de cluster entre la pédologie et les données externes.

La télédétection comporte une dimension temporelle, représentant 6 ans environ d'itinéraire cultural et plus précisément les successions culturales (comprenant des sols nus ou des intercultures) ainsi que le comportement de la végétation. La pédologie, elle, est représentée selon des analyses faites par échantillonnage à un moment précis. Par conséquent, la télédétection dans cette étude se veut être représentative d'une évolution temporelle de la végétation, évolution liée aux variables pédologiques. Il est alors probable que la classification de la télédétection et la classification de la pédologie ne soient pas représentatives du même phénomène puisqu'elles n'ont pas la même dimension temporelle. Cette hypothèse est à nuancer car les variables pédologiques mesurées sont relativement pérennes (environ dix ans) et sont responsables des variations intra-annuelles. Néanmoins, l'apport de la dimension temporelle au sein de la télédétection est intéressant de par sa capacité lors de la classification à représenter une évolution phénologique. La granulométrie d'un sol ne varie que très peu au cours du temps, et sachant que ce sont les variables granulométriques qui sont prédominantes au sein des clusters pédologiques, la classification pédologique ne permet pas d'observer l'évolution des sols au cours du temps.

La classification de la résistivité et de la télédétection ayant fourni des résultats équivalents sur certains clusters pédologiques, il aurait été intéressant de combiner les deux sources de données afin d'observer si cela avait mené à une amélioration de la classification ou à une dégradation de cette dernière. En effet, la résistivité a permis d'expliquer correctement le cluster pédologique 4 (limons fins), ce que la télédétection n'a pas toujours réussi. Néanmoins, la résistivité reste, au même titre que la télédétection, une source de données exhaustive permettant une première approche d'une représentation de la variabilité pédologique en surface. Caccetta et al. (2000) ont d'ailleurs suggéré d'utiliser des données topographiques afin d'améliorer la qualité de classification des images de télédétection.

V. Conclusion

Des délimitations de zones aux propriétés homogènes ont été réalisées à partir de trois sources de données (159 images de télédétection Sentinel-2, une prospection de résistivité électrique et des variables pédologiques mesurées ponctuellement en 56 sites sur le domaine d'étude). Les zones homogènes obtenues notamment grâce à la méthode *k-means* sont différentes en nombre et dans leur arrangement. Ceci est probablement lié à la source de données et son exhaustivité mais également du fait que les trois sources de données n'expriment pas nécessairement les mêmes propriétés et processus prépondérants sur le site d'étude. Les

clusters issus des données pédologiques ponctuelles et éparées sont en nombre réduit et généralement de grande taille contrairement à ceux obtenus à partir de la prospection de la résistivité électrique et des images Sentinel-2. Toutefois, la concordance entre les clusters issus des trois sources de données n'est pas insignifiante pour trois quarts des clusters issus des données pédologiques.

En perspective de ce travail, il serait intéressant de tenter une délimitation de zones homogènes sur le domaine d'étude en considérant simultanément les trois sources de données (voire plus : par exemple des cartes de rendement de cultures disponibles sur le site d'étude) dans un algorithme de classification. Par ailleurs, la méthodologie développée dans ce travail ainsi que la perspective évoquée ci-dessus peuvent être conduites sur chacune des deux parcelles constituant le domaine d'étude étant donné qu'elles bénéficient d'une conduite agronomique différente au cours d'une année culturale. D'autres méthodologies peuvent aussi être appliquées comme la différenciation de sol nu et sol couvert pour tenter de préciser certains phénomènes phénologiques en réaction aux propriétés pédologiques des sols.

Bibliographie

- Abzalov, M., 2016. Multiple Indicator Kriging, in: *Applied Mining Geology, Modern Approaches in Solid Earth Sciences*. Springer International Publishing, Cham, pp. 291–294. https://doi.org/10.1007/978-3-319-39264-6_21
- Baize, D., Girard, M.-C., 2009. *Référentiel pédologique*, [Éd.] 2008. ed, Savoir-faire. Éd. Quae, Versailles.
- Bhunja, G.S., Kumar Shit, P., Pourghasemi, H.R., 2019. Soil organic carbon mapping using remote sensing techniques and multivariate regression model. *Geocarto Int.* 34, 215–226. <https://doi.org/10.1080/10106049.2017.1381179>
- Caccetta, P., Allen, A., Watson, I., 2000. The Land Monitor Project.
- Chabrillat, S., Ben-Dor, E., Cierniewski, J., Gomez, C., Schmid, T., Van Wesemael, B., 2019. Imaging Spectroscopy for Soil Mapping and Monitoring. *Surv. Geophys.* 40, 361–399. <https://doi.org/10.1007/s10712-019-09524-0>
- Chen, H., Huang, W., Li, W., Niu, Z., Zhang, L., Xing, S., 2018. Estimation of LAI in Winter Wheat from Multi-Angular Hyperspectral VNIR Data: Effects of View Angles and Plant Architecture. *Remote Sens.* 10, 1630. <https://doi.org/10.3390/rs10101630>
- Dodin, M., Levavasseur, F., Savoie, A., Martin, L., Foulon, J., Vaudour, E., 2023. Sentinel-2 satellite images for monitoring cattle slurry and digestate spreading on emerging wheat crop: a field spectroscopy experiment. *Geocarto Int.* 1–17. <https://doi.org/10.1080/10106049.2023.2245371>
- Dodin, M., Smith, H.D., Levavasseur, F., Hadjar, D., Houot, S., Vaudour, E., 2021. Potential of Sentinel-2 Satellite Images for Monitoring Green Waste Compost and Manure Amendments in Temperate Cropland. *Remote Sens.* 13, 1616. <https://doi.org/10.3390/rs13091616>
- Forgy, E., 1965. Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classifications. *Biometrics* 21, 768–780.
- Friedman, H.P., Rubin, J., 1967. On Some Invariant Criteria for Grouping Data. *J. Am. Stat. Assoc.* 62, 1159–1178. <https://doi.org/10.1080/01621459.1967.10500923>
- Galvão, L.S., Pizarro, M.A., Epiphany, J.C.N., 2001. Variations in Reflectance of Tropical Soils. *Remote Sens. Environ.* 75, 245–255. [https://doi.org/10.1016/S0034-4257\(00\)00170-X](https://doi.org/10.1016/S0034-4257(00)00170-X)
- Gelb, J., Apparicio, P., 2021. Apport de la classification floue c-means spatiale en géographie : essai de taxinomie socio-résidentielle et environnementale à Lyon. *Cybergeogeo*. <https://doi.org/10.4000/cybergeogeo.36414>
- Gholizadeh, A., Žizala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens. Environ.* 218, 89–103. <https://doi.org/10.1016/j.rse.2018.09.015>
- Guo, L., Fu, P., Shi, T., Chen, Y., Zeng, C., Zhang, H., Wang, S., 2021. Exploring influence factors in mapping soil organic carbon on low-relief agricultural lands using time series of remote sensing data. *Soil Tillage Res.* 210, 104982. <https://doi.org/10.1016/j.still.2021.104982>
- Hatfield, J.L., Gitelson, A.A., Schepers, J.S., Walthall, C.L., 2008. Application of Spectral Remote Sensing for Agronomic Decisions. *Agron. J.* 100. <https://doi.org/10.2134/agronj2006.0370c>
- Hatfield, J.L., Prueger, J.H., Sauer, T.J., Dold, C., O'Brien, P., Wacha, K., 2019. Applications of Vegetative Indices from Remote Sensing to Agriculture: Past and Future. *Inventions* 4, 71. <https://doi.org/10.3390/inventions4040071>
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)

- Jin, X., Du, J., Liu, H., Wang, Z., Song, K., 2016. Remote estimation of soil organic matter content in the Sanjiang Plain, Northeast China: The optimal band algorithm versus the GRA-ANN model. *Agric. For. Meteorol.* 218–219, 250–260. <https://doi.org/10.1016/j.agrformet.2015.12.062>
- Khaliq, A., Peroni, L., Chiaberge, M., 2018. Land cover and crop classification using multitemporal sentinel-2 images based on crops phenological cycle, in: 2018 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS). Presented at the 2018 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS), IEEE, Salerno, pp. 1–5. <https://doi.org/10.1109/EESMS.2018.8405830>
- Liu, Q., He, L., Guo, L., Wang, M., Deng, D., Lv, P., Wang, R., Jia, Z., Hu, Z., Wu, G., Shi, T., 2022. Digital mapping of soil organic carbon density using newly developed bare soil spectral indices and deep neural network. *CATENA* 219, 106603. <https://doi.org/10.1016/j.catena.2022.106603>
- Macqueen, J., 1967. Some methods for classification and analysis of multivariate observations. *Multivar. Obs.* 5.1, 281–297.
- Miller, J.D., Thode, A.E., 2007. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). *Remote Sens. Environ.* 109, 66–80. <https://doi.org/10.1016/j.rse.2006.12.006>
- Mohamed, E.S., Saleh, A.M., Belal, A.B., Gad, A., 2018. Application of near-infrared reflectance for quantitative assessment of soil properties. *Egypt. J. Remote Sens. Space Sci.* 21, 1–14. <https://doi.org/10.1016/j.ejrs.2017.02.001>
- Pean, L., 2022. Détermination d'unités de gestion Intra-Parcellaires par analyse d'images de télédétection. (Stage). Orléans.
- Pimstein, A., Notesco, G., Ben-Dor, E., 2011. Performance of Three Identical Spectrometers in Retrieving Soil Reflectance under Laboratory Conditions. *Soil Sci. Soc. Am. J.* 75, 746–759. <https://doi.org/10.2136/sssaj2010.0174>
- Post, D.F., Fimbres, A., Matthias, A.D., Sano, E.E., Accioly, L., Batchily, A.K., Ferreira, L.G., 2000. Predicting Soil Albedo from Soil Color and Spectral Reflectance Data. *Soil Sci. Soc. Am. J.* 64, 1027–1034. <https://doi.org/10.2136/sssaj2000.6431027x>
- Richer-de-Forges, A.C., Chen, Q., Baghdadi, N., Chen, S., Gomez, C., Jacquemoud, S., Martelet, G., Mulder, V.L., Urbina-Salazar, D., Vaudour, E., Weiss, M., Wigneron, J.-P., Arrouays, D., 2023. Remote Sensing Data for Digital Soil Mapping in French Research—A Review. *Remote Sens.* 15, 3070. <https://doi.org/10.3390/rs15123070>
- Rouse, W., Haas, R.H., 1973. MONITORING VEGETATION SYSTEMS IN THE GREAT PLAINS WITH ERTS.
- Santaga, F.S., Agnelli, A., Leccese, A., Vizzari, M., 2021. Using Sentinel-2 for Simplifying Soil Sampling and Mapping: Two Case Studies in Umbria, Italy. *Remote Sens.* 13, 3379. <https://doi.org/10.3390/rs13173379>
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W., Lin, C.-T., 2017. A review of clustering techniques and developments. *Neurocomputing* 267, 664–681. <https://doi.org/10.1016/j.neucom.2017.06.053>
- Serbin, G., Hunt, E.R., Daughtry, C.S.T., Brown, D.J., McCarty, G.W., Doraiswamy, P.C., 2010. Assessment of spectral indices for crop residue cover estimation, in: 2010 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2010 - 2010 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Honolulu, HI, USA, pp. 1827–1830. <https://doi.org/10.1109/IGARSS.2010.5654193>
- Somers, B., Gysels, V., Verstraeten, W.W., Delalieux, S., Coppin, P., 2010. Modelling moisture-induced soil reflectance changes in cultivated sandy soils: a case study in citrus orchards. *Eur. J. Soil Sci.* 61, 1091–1105. <https://doi.org/10.1111/j.1365-2389.2010.01305.x>
- Szekely, G.J., Rizzo, M.L., 2005. Hierarchical Clustering via Joint Between-Within Distances: Extending Ward's Minimum Variance Method. *J. Classif.* 22, 151–183. <https://doi.org/10.1007/s00357-005-0012-9>
- Vaudour, E., Gholizadeh, A., Castaldi, F., Saberioon, M., Borůvka, L., Urbina-Salazar, D., Fouad, Y., Arrouays, D., Richer-de-Forges, A.C., Biney, J., Wetterlind, J., Van Wesemael, B., 2022. Satellite Imagery to Map Topsoil Organic Carbon Content over Cultivated Areas: An Overview. *Remote Sens.* 14, 2917. <https://doi.org/10.3390/rs14122917>
- Vaudour, E., Gomez, C., Lagacherie, P., Loiseau, T., Baghdadi, N., Urbina-Salazar, D., Loubet, B., Arrouays, D., 2021. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content

- mapping in croplands. *Int. J. Appl. Earth Obs. Geoinformation* 96, 102277. <https://doi.org/10.1016/j.jag.2020.102277>
- Vaudour, E., Gomez, C., Loiseau, T., Baghdadi, N., Loubet, B., Arrouays, D., Ali, L., Lagacherie, P., 2019. The Impact of Acquisition Date on the Prediction Performance of Topsoil Organic Carbon from Sentinel-2 for Croplands. *Remote Sens.* 11, 2143. <https://doi.org/10.3390/rs11182143>
- Zhang, B., Zhang, M., Sun, P., Liu, H., Feng, L., 2022. Resistivity is used as a tool to evaluate the variability of soil water content. *J. Mt. Sci.* 19, 3533–3547. <https://doi.org/10.1007/s11629-022-7445-z>

Références internet

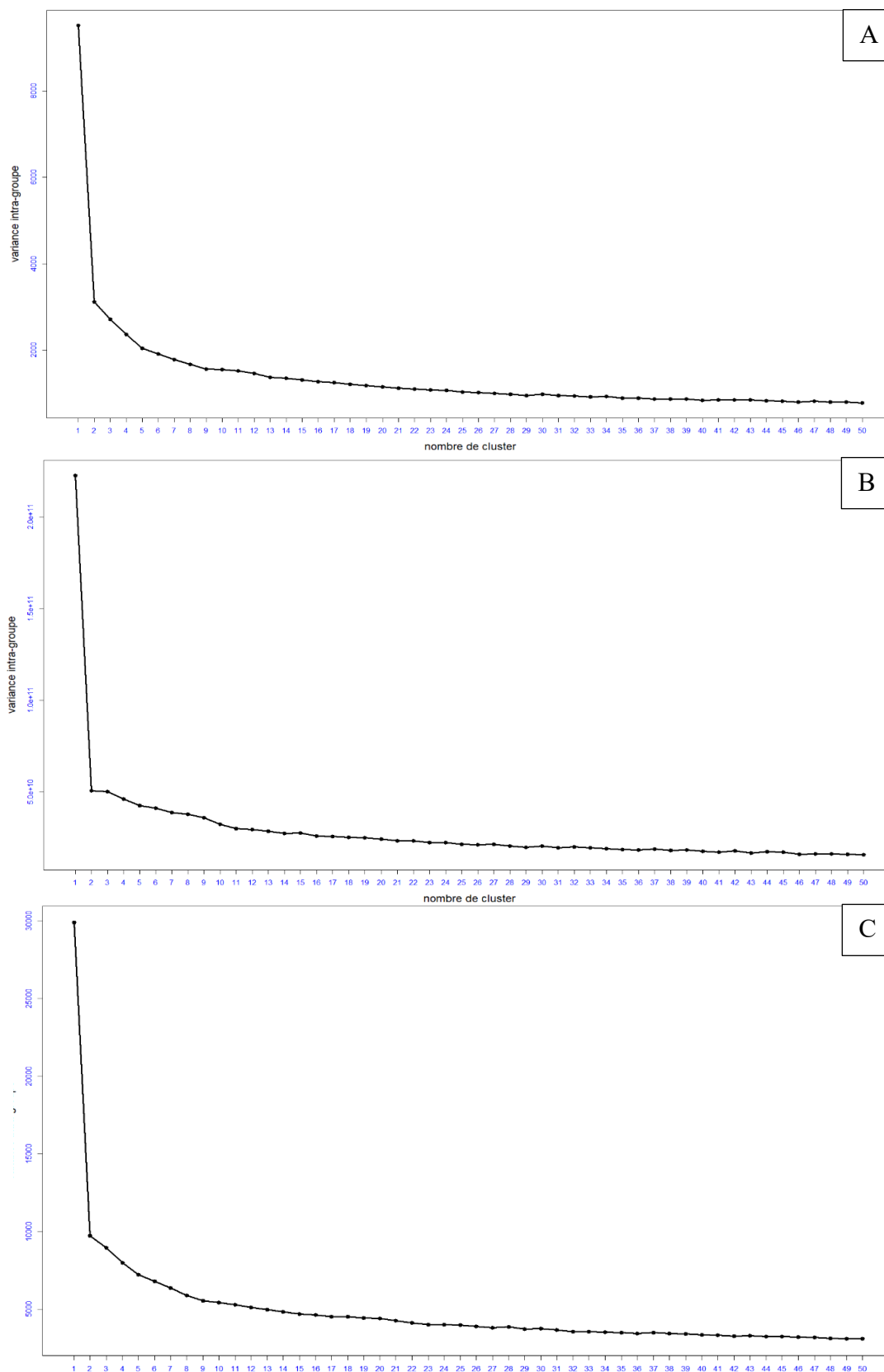
- Copernicus, 2023. European Union Space Program [en ligne]. Disponible sur <https://www.copernicus.eu/en/> (consulté le 09/08/2023).
- GisSol, 2023. Geosol [en ligne]. Disponible sur <https://webapps.gissol.fr/geosol/> (consulté le 11/08/2023).

Liste figures/tableaux

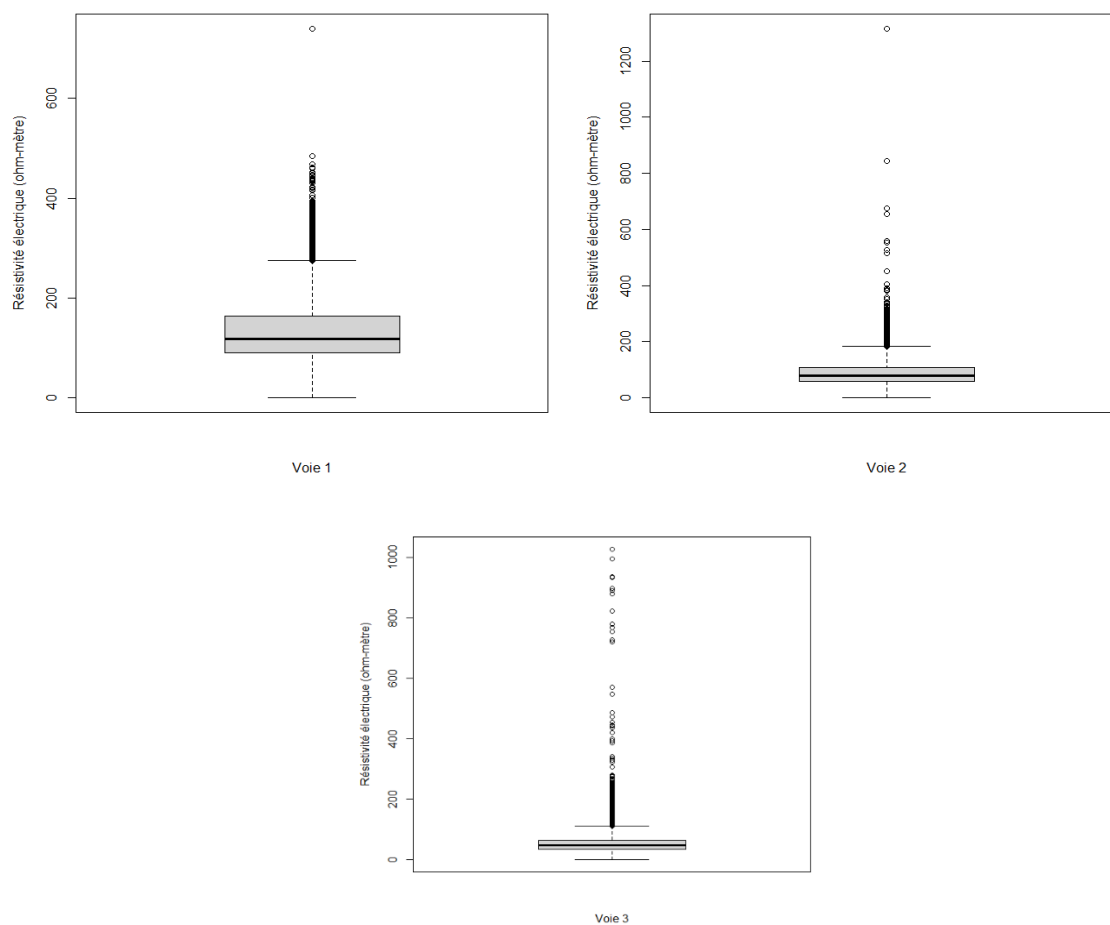
Figure 1 : A- Localisation du site d'étude. B- Composition du site d'étude.	3
Figure 2 : Carte détaillée des types de sol au sein des deux parcelles d'études (Giroto, 2023).....	4
Figure 3 : Points de mesure de la résistivité électrique	5
Figure 4 : Points d'échantillonnage retenus/non retenus pour analyse issus du premier jeu de données pédologiques.....	6
Figure 5 : Représentation des points d'échantillonnage issus des deux jeux de données pédologiques.....	7
Figure 6 : Démarche globale utilisée dans cette étude. (« A » représente le clustering effectué par k-means où k clusters varie entre les deux sources de données (télédétection et résistivité). « B » représente une évaluation statistique des clusters qui a été faite soit par un test de Wilcoxon ou par une ANOVA. « C » représente la comparaison des clusterings de télédétection et de résistivité avec le résultat de la spatialisation des classes du CAH par un tableau de contingence et un test statistique appliqué si nécessaire.....	8
Figure 7 : Matrice de corrélation des variables pédologiques. (Les croix représentent les corrélations qui ne sont pas significatives à l'intervalle de confiance 95%).	11
Figure 8 : Profil spectral des 159 images de télédétection à partir des 10 bandes étudiées	12
Figure 9 : Evolution temporelle de l'indice spectral NDVI moyen sur l'ensemble de la zone (parcelle 29 + parcelle 30) avec les cultures mises en place sur les deux parcelles. Les résultats représentant les deux parcelles simultanément, il est important de rappeler qu'à partir du 12/07/2020, les itinéraires culturels des parcelles sont différents.....	13
Figure 10 : Agrégation issue de la Classification Ascendante Hiérarchique avec 56 points d'échantillonnage (individus) et 7 variables	13
Figure 11 : Profil d'analyse pédologique des 4 clusters pédologiques	14
Figure 12 : Spatialisation par krigeage d'indicatrice des classes issues de la CAH menée sur les variables pédologiques.....	15
Figure 13 : Spatialisation de la composante de longue portée de la résistivité (0-50 cm) en ohm.mètre	16

Figure 14 : A- Classification du raster "multi-bandes". B- Classification du raster "multi-indices". C- Classification du raster "NDVI". (La légende est équivalente pour les 3 classifications).....	17
Figure 15 : Clustering de la résistivité électrique (0-50 cm) à partir des données de la composante longue portée.....	19
Figure 16 : Représentation de la significativité de l'ANOVA entre les clusters significatifs. A- Classification raster "NDVI". B- Classification sur le raster "multi-bandes". C- Classification sur le raster "multi-indices".....	21
Tableau 1 : Equation des cinq indices spectraux utilisés et leur nom associé.	9
Tableau 2 : Analyse descriptive des données pédologiques	10
Tableau 3 : Analyse descriptive de la résistivité sur l'ensemble du site d'étude sur 3 intervalles de profondeur distincts	11
Tableau 4 : Intervalles de valeur pour les 10 bandes spectrales et les 5 indices spectraux.	12
Tableau 5 : Superficie occupée par les clusters issus de la classification pédologique sur le site d'étude et le pourcentage représenté dans le résultat de la CAH.	15
Tableau 6 : Répartition des clusters issus de la classification sur la télédétection sur le site d'étude exprimé en pourcentage de surface	18
Tableau 7 : Médianes des valeurs de l'indice spectral NDVI au sein des différents clusters significatifs des 3 classifications de la télédétection.....	18
Tableau 8 : Répartition des clusters issus de la classification de la résistivité sur le site d'étude exprimé en pourcentage de surface.....	19
Tableau 9 : Médianes des valeurs de résistivité électrique (0-50 cm) au sein des cinq clusters issus de la classification de la résistivité (0-50cm).....	19
Tableau 10 : Comparaison des moyennes des variables pédologiques par paire de classes. Vert : $\mu_1 = \mu_2$. Orange : $\mu_1 \neq \mu_2$	22
Tableau 11 : Tableau de contingence de la classification du raster "multi-bandes". (* signifie qu'un test χ^2 a été effectué.)	23
Tableau 12 : Tableau de contingence de la classification du raster "NDVI". (* signifie qu'un test χ^2 a été effectué.)	23
Tableau 13 : Tableau de contingence de la classification du raster "multi-indices". (* signifie qu'un test χ^2 a été effectué.)	24
Tableau 14 : Tableau de contingence de la classification du raster "résistivité". (* signifie qu'un test χ^2 a été effectué.)	24
Annexe 1 : Evolution de la variance intra-cluster jusqu'à k égal 50. A- Classification du raster "NDVI". B- Classification du raster "multi-bandes". C- Classification du raster "multi-indices".	34
Annexe 2 : Représentation de la répartition des valeurs sur l'ensemble des voies de la résistivité électrique.....	35
Annexe 3 : Significativité des tests de Wilcoxon effectués sur les clusters issus de la classification du raster "résistivité". Les valeurs utilisées lors des tests correspondent à la résistivité électrique (0-50 cm). (***) signifie $p\text{-value} < 2,2 \cdot 10^{-16}$).	35
Annexe 4 : Résultats des tests χ^2 effectués pour distinguer les clusters prédominants. (En gras le duo de clusters où le test a montré que leurs compositions étaient significativement égales).....	36

Annexes



Annexe 1 : Evolution de la variance intra-cluster jusqu'à k égal 50. A- Classification du raster "NDVI". B- Classification du raster "multi-bandes". C- Classification du raster "multi-indices".



Annexe 2 : Représentation de la répartition des valeurs sur l'ensemble des voies de la résistivité électrique

	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
Test de Wilcoxon	***	***	***	***	***	***	***	***	***	***

Annexe 3 : Significativité des tests de Wilcoxon effectués sur les clusters issus de la classification du raster "résistivité". Les valeurs utilisées lors des tests correspondent à la résistivité électrique (0-50 cm). (***) signifie $p\text{-value} < 2,2 \cdot 10^{-16}$.

<i>Classification</i>	<i>Clusters concernés</i>	<i>p-value du test Khi²</i>
<i>Multi-bandes</i>	4-5 (Cluster pédologique 3)	1,135.10 ⁻⁶
<i>Multi-indices</i>	4-5 (Cluster pédologique 3)	1,802.10 ⁻⁸
<i>NDVI</i>	3-6 (Cluster pédologique 3) ; 5-7 (Cluster pédologique 4)	2,2.10 ⁻¹⁶ ; 0,006
<i>Résistivité</i>	1-2 (Cluster pédologique 2) ; 3-5 (Cluster pédologique 3)	2,2.10 ⁻¹⁶ ; 0,286

Annexe 4 : Résultats des tests Khi² effectués pour distinguer les clusters prédominants. (En gras le duo de clusters où le test a montré que leurs compositions étaient significativement égales).

Résumé

L'étude de la variabilité pédologique intra-parcellaire est devenue de plus en plus importante dans un contexte d'agriculture de précision qui permet l'économie et la protection des ressources. La télédétection est, à ce titre, utilisée de nos jours afin de décrire les sols sans les détruire. L'objectif de ce travail est d'utiliser la télédétection afin de déterminer des zones avec des réponses spectrales homogènes au sein de parcelles agricoles. Il s'agit ensuite de valider le zonage spectral obtenu par des données pédologiques et par l'appui de la résistivité électrique des horizons des sols. Le site d'étude est composé de deux parcelles agricoles conjointes ayant des itinéraires culturaux différents. Les données mobilisées dans ce travail sont 159 images du site d'étude fournies par Sentinel-2, une prospection géophysique de la résistivité électrique et des analyses pédologiques issues de 56 sondages ponctuels réalisés sur le site d'étude. L'algorithme de classification *k-means* est le principal outil statistique mis en œuvre afin de délimiter des zones aux propriétés homogènes sur le site d'étude à partir des 3 sources de données citées. Les résultats ont montré un nombre et une structure de clusters conditionnés par le type de donnée considérée et l'exhaustivité de l'information. Toutefois, la concordance entre les clusters issus des trois sources de données n'est pas insignifiante pour trois quarts des clusters issus des données pédologiques considérés comme référence malgré leur faible abondance. Une des perspectives de ce travail serait de considérer une délimitation de zones homogènes sur le domaine d'étude en utilisant simultanément les trois sources de données voire davantage (ex. carte de rendements de cultures) dans un algorithme de classification.

Mots-clés : télédétection ; indice spectral ; clustering ; variabilité pédologique ; agriculture ; résistivité

Abstract

The study of intra-parcel soil variability has become increasingly important in the context of precision agriculture, which helps to save and protect resources. Remote sensing is now being used to describe soils without destroying them. The aim of this work is to use remote sensing to determine areas with homogeneous spectral responses within agricultural plots. The spectral zoning obtained is then validated using pedological data and the electrical resistivity of soil horizons. The study site consists of two agricultural plots with different cropping patterns. The data used in this work consisted of 159 images of the study site provided by Sentinel-2, a geophysical survey of the electrical resistivity and soil analyses from 56 spot drillings carried out on the study site. The *k-means* classification algorithm is the main statistical tool used to delimit zones with homogeneous properties on the study site based on the 3 data sources mentioned above. The results showed a number and structure of clusters conditioned by the type of data considered and the completeness of the information. However, the concordance between the clusters from the three data sources is not insignificant for three quarters of the clusters from the soil data considered as a reference, despite their low abundance. One of the perspectives of this work would be to consider delimiting homogeneous zones in the study area by simultaneously using the three data sources or even more (e.g. crop yield map) in a classification algorithm.

Keywords: remote sensing; spectral index; clustering; soil variability; agriculture; resistivity