

Département de géomatique appliquée  
Faculté des lettres et sciences humaines  
Université de Sherbrooke

**Contribution au développement d'un système d'aide au diagnostic  
en agriculture de précision dans la région de Chaudière-Appalaches.**

Valérie Duchesne

Essai présenté pour l'obtention du grade de  
Maître ès en géomatique appliquée et télédétection (M.Sc.),  
cheminement de type cours

Septembre 2024

© Valérie Duchesne, 2024

Directeur d'essai : Yacine Bouroubi (Département de géomatique appliquée, Université de Sherbrooke)

Codirecteurs :

Mickaël Germain (Département de géomatique appliquée, Université de Sherbrooke)

Samuel Foucher (Département de géomatique appliquée, Université de Sherbrooke)

Collaborateur : Victor Morin (Président, entreprise XLKey)

Membre du jury :

Ramata Magagi (Département de géomatique appliquée, Université de Sherbrooke)

## Résumé

En raison de la dégradation des ressources naturelles, l'agriculture évolue vers une approche plus respectueuse de l'environnement, soutenue par des technologies de pointe pour optimiser les rendements, tout en minimisant les impacts environnementaux et les coûts de production. L'objectif principal de ce projet consiste à contribuer au développement des bases de l'infrastructure d'un système d'aide à la décision au diagnostic en agriculture pour des champs analysés dans la région de Chaudière-Appalaches. Les objectifs spécifiques de cette étude sont la valorisation des données, la définition de produits dérivés, l'identification des zones de gestion ainsi que l'aide au diagnostic des rendements au champ. Les données utilisées dans ce projet proviennent de diverses sources, mais principalement de l'entreprise XLKey, et servent à analyser des caractéristiques agricoles spécifiques. Elles sont divisées en trois catégories : (1) les caractéristiques du sol, (2) les données d'opération et (3) les cultures. La méthodologie développée a été appliquée sur un ensemble de données de champs provenant de trois exploitations agricoles, clients de l'entreprise XLKey, en utilisant la programmation Python 3.6 sur Google Colaboratory et QGIS pour les cartographies. Des cartes d'échantillonnage par zone et de calibration (CE-MO) ont été établies en vue de déterminer les emplacements optimaux pour les prélèvements d'échantillons de sol et elles ont été conçues pour être automatisées via des scripts Python. Une carte basée sur l'indice de position topographique (TPI) a permis d'identifier des zones susceptibles d'accumulation d'eau ou au contraire, celles à risque de stress hydrique, en fonction de la topographie. Les analyses géospatiales basées sur les zones de gestion et un modèle de régression (*Random Forest*) ont été élaborés pour prédire le rendement en se basant sur les caractéristiques du sol et la topographie.

Les résultats de ces analyses ont été exportés vers un fichier texte et des graphiques illustrant l'importance des différentes variables explicatives pour les rendements agricoles ont été produits. L'application de la méthodologie a permis d'obtenir des résultats qui mettent en évidence la corrélation entre différentes données agricoles et d'offrir une compréhension approfondie des facteurs influençant les rendements agricoles dans certaines zones de champs appuyée par des analyses statistiques. Les analyses ont identifié des ajustements essentiels pour améliorer la précision des prévisions de rendement agricole en évitant l'inclusion de données de zones non cultivées, afin d'optimiser les décisions et maximiser les rendements.

# Table des matières

|   |     |
|---|-----|
| Résumé.....   | i   |
| Table des matières.....                                   | ii  |
| Liste des figures.....                                    | iv  |
| Liste des tableaux.....                                   | v   |
| Liste d'acronymes.....                                    | vi  |
| Remerciements.....  | vii |
| 1 Introduction.....                                       | 1   |
| 1.1 Mise en Contexte et problématique.....                | 1   |
| 1.2 Objectifs.....  | 2   |
| 2 Cadre Théorique.....                                    | 3   |
| 2.1 Agriculture de précision.....                         | 3   |
| 2.2 Valorisation des données acquises sur les champs..... | 4   |
| 2.2.1 Acquisition des données.....                        | 4   |
| 2.3 Définition des produits dérivés.....                  | 7   |
| 2.3.1 Carte de gestion des sols.....                      | 8   |
| 2.3.2 Cartes de rendement.....                            | 8   |
| 2.4 Identification des zones de gestion.....              | 9   |
| 2.5 Diagnostic des rendements au champ.....               | 10  |
| 2.5.1 Évaluation des rendements.....                      | 10  |
| 2.5.2 Analyse des pratiques agricoles.....                | 11  |
| 2.5.3 Analyse des facteurs environnementaux.....          | 12  |
| 2.5.4 Évaluation économique.....                          | 12  |
| 3 Matériel et méthodes.....                               | 13  |
| 3.1 Présentation des sites d'étude.....                   | 13  |
| 3.2 Collecte des données.....                             | 18  |
| 3.2.1 Données sur les propriétés du sol.....              | 18  |
| 3.2.2 Données sur les opérations.....                     | 19  |
| 3.2.3 Données sur les cultures.....                       | 19  |
| 3.3 Méthodologie et traitement des données.....           | 20  |
| 3.3.1 Prétraitement et interpolations des données.....    | 21  |
| 3.3.2 Carte d'échantillonnage par zone.....               | 22  |

|       |   |    |
|-------|---|----|
| 3.3.3 | Carte d'échantillonnage de calibration CE-MO.....                     | 23 |
| 3.3.4 | Indice TPI (Indice de position topographique).....                    | 24 |
| 3.3.5 | Analyses géospatiales et corrélations des données agricoles .....     | 25 |
| 3.3.6 | Modélisation du rendement .....                                       | 25 |
| 4     | Résultats.....  | 27 |
| 4.1   | Carte d'échantillonnage par zone .....                                | 27 |
| 4.2   | Carte d'échantillonnage de calibration CE-MO.....                     | 29 |
| 4.3   | Carte illustrant le TPI.....  | 31 |
| 4.4   | Modélisation du rendement .....                                       | 34 |
| 4.4.1 | Diagnostic expliqué par les données LiDAR .....                       | 37 |
| 4.4.2 | Variations du rendement reliées à la matière organique .....          | 40 |
| 4.4.3 | Variation du rendement expliquée par la conductivité électrique ..... | 43 |
| 5     | Interprétation et discussion des résultats .....                      | 45 |
| 6     | Conclusion et Perspectives.....                                       | 48 |
| 7     | Références.....   | 50 |

## Liste des figures

|  |    |
|--|----|
| Figure 1. Localisation des 5 parcelles agricoles à l'étude de la ferme « Forfait André et Frédéric Roy à Honfleur » .....  | 14 |
| Figure 2. Localisation des 9 parcelles agricoles à l'étude de la ferme « Marcel Nadeau » dans les municipalités de Saint-Flavien, Saint-Gilles, Saint-Bernard et Saint-Patrice-de-Beaurivage.....  | 14 |
| Figure 3. Localisation des 26 parcelles agricoles à l'étude de la ferme « 9077-9729 inc. » dans les municipalités de Saint-Bernard, Saint-Narcisse de Beaurivage et Saint-Antoine de Tilly.....  | 15 |
| Figure 4. Capteur Veris MSP3 qui permet de calculer des données sur la conductivité électrique (EC), d'obtenir la mesure de pH du sol ainsi que des données sur les réflectances (PIR et R) pour obtenir des valeurs de matière organique du sol (OM).....   | 18 |
| Figure 5. Organigramme méthodologique pour ce projet. ....   | 21 |
| Figure 6. Organigramme des étapes réalisées afin d'obtenir les cartes d'échantillonnage de chacun des champs. ....   | 23 |
| Figure 7. Organigramme des étapes effectués par le modèle d'apprentissage automatique (« Random Forest ») (Belgiu et Drăguț, 2016). ....   | 26 |
| Figure 8. Carte d'échantillonnage selon la texture du sol des champs Be01, Be02, Be03 et Be04. ....  | 28 |
| Figure 9. Graphiques sur les relations entre les variables explicatives analysées dans ce projet sur des données compilées remises par l'entreprise XLKey a) Relation du rendement (MS) en fonction de la CE avec un coefficient de détermination de 0,16 ; b) Corrélacion entre la réflectance du R en fonction de la PIR avec un $r = 0,77$ ; c) Relation entre la CE et la MO ( $r = 0,12$ ) ; d) Graphique de la classification par la méthode du « k-means » afin de déterminer 3 classes de regroupement. .... | 30 |
| Figure 10. Identification des 3 points d'échantillonnage les plus représentatif du champ selon la répartition des 3 classes réalisées lors de la classification non supervisée (K-means) : classe de la très forte teneur en MO et CE (-71.3342, 46.4582), classe de la très faible teneur en MO et CE (-71.3373, 46.4570) et la classe de la teneur moyenne en MO et CE (-71.3360, 46.4575). ....   | 31 |
| Figure 11. Carte illustrant les valeurs de TPI des champs BS02 et B04 de la ferme 9077-9729 Qc inc.....  | 32 |
| Figure 12. Cartes illustrant les valeurs de TPI et de la topographie du champ BS02 de la ferme 9077-9729 Qc inc. L'échelle de couleur appliquée à la carte indique que le vert correspond à des valeurs de hautes altitudes et le rouge à des valeurs basse altitudes .....  | 33 |
| Figure 13. Cartes illustrant les valeurs de TPI et de la topographie du champ B04 de la ferme 9077-9729 Qc inc.....  | 34 |
| Figure 14. Aperçu d'un fichier texte obtenu suite à l'exécution du code Random Forest. L'exemple correspond au résultat du champ B04 de la ferme 9077-9729 Qc inc. ....  | 35 |
| Figure 15. Graphique de l'importance des caractéristiques explicatives pour le rendement de matière humide (Moisture) pour la parcelle B04 de la ferme 9077-9729 situé à Saint-Bernard. ....   | 35 |
| Figure 16. Graphique de l'importance des caractéristiques explicatives pour le rendement de matière sèche (Wet) pour la parcelle B04 de la ferme 9077-9729 situé à Saint-Bernard. ....   | 35 |
| Figure 17. Carte des analyses de la parcelle B03 de la ferme 9077-9729 Qc inc. située à Saint-Bernard de la MS et MH en considérant les données LIDAR (topographie). ....  | 38 |
| Figure 18. Carte des analyses de la parcelle B04 de la ferme 9077-9729 Qc inc. située à Saint-Bernard de la MS et MH en considérant les données LIDAR (topographie). ....  | 39 |

|   |    |
|---|----|
| Figure 19. Carte des analyses de la parcelle BS02 de la ferme 9077-9729 Qc inc. située à Saint-Bernard de la MS et MH en considérant les données LIDAR (topographie). .....               | 40 |
| Figure 20. Carte des analyses de la parcelle N01 de la ferme 9077-9729 Qc inc situé à Saint-Narcisse-de-Beaurivage de la MS et MH en considérant les données de la réflectance du R. .... | 42 |
| Figure 21. Carte des analyses de la parcelle FR02 de la ferme Forfait André et Frédéric Roy située à Honfleur de la MS et MH en considérant les données de la réflectance du R. ....      | 43 |
| Figure 22. Carte des analyses de la parcelle GIL01 de la ferme Marcel Nadeau situé à Saint-Gilles de la MS et MH en considérant les données de la CE.....                                 | 44 |

## Liste des tableaux

|   |    |
|---|----|
| Tableau 1. Source des données utilisées dans le cadre de ce projet. ....                              | 16 |
| Tableau 2. Description des données qui ont été sélectionnée pour les analyses. ....                   | 17 |
| Tableau 3. Sommaire des résultats obtenus lors de la modélisation « Random Forest » (Régression)..... | 36 |

## Liste d'acronymes

|           |  |
|-----------|--|
| AAC :     | Agriculture et Agroalimentaire Canada                                |
| BD :      | Base de données  |
| CE (EC) : | Conductivité électrique  |
| CPVQ :    | Conseil des productions végétales du Québec inc.                     |
| CRAAQ :   | Centre de référence en agriculture et agroalimentaire du Québec      |
| BDPPAD :  | La Base des données des parcelles et productions agricoles déclarées |
| GES :     | Gas à effet de serre   |
| GPS :     | Système de Positionnement Global                                     |
| ha :      | Hectare (unité de mesure)  |
| LiDAR :   | Light Detection and Ranging  |
| LVC-1 :   | La validation croisée « en omettre un »                              |
| MH :      | Matière humide   |
| MNT :     | Modèle numérique de terrain  |
| MO :      | Matière organique  |
| MRC :     | Municipalité régionale de comté                                      |
| MRNF :    | Ministère des Ressources naturelles et de la Faune                   |
| MS :      | Matière sèche  |
| PIR :     | Proche infrarouge  |
| SIG :     | Système d'information géographique                                   |
| TPI :     | Indice de position topographique                                     |
| ZG :      | Zones de gestion   |

## Remerciements

Merci à mon directeur, Yacine Bouroubi, pour m’ avoir offert l’ opportunité d’ explorer un sujet qui me passionne dans le cadre de mon essai de maîtrise. Les commentaires constructifs m’ ont permis de progresser et de me dépasser et je vous en suis très reconnaissante.

Merci à mes deux codirecteurs, Mickaël Germain et Samuel Foucher, pour leur encadrement attentif et leurs précieuses expertises qui ont enrichi mon parcours dans ce projet.

Merci à Victor Morin, président de XLKey, pour sa collaboration précieuse qui a grandement facilité la compréhension de ce projet d' envergure. Je souhaite également exprimer ma gratitude à Étienne Patry, agronome au sein de l’ équipe XLKey, pour son expertise dans le domaine agricole et ses précieuses contributions.

Merci à Yannick Duguay du CGQ pour son aide et son expertise dans le domaine de la télédétection.

Merci à mon conjoint Kevin et à mes enfants William, Abygaël et Émylou pour leur soutien et leur encouragement constants.

Merci également à Ramata Magagi, membre du jury, pour la révision de mon essai.

# 1 Introduction

## 1.1 Mise en Contexte et problématique

Depuis quelques années, les pratiques reliées à l'agriculture ont été remises en question afin de mettre en place une agriculture qui respecte les contraintes environnementales, économiques et sociales (Ordre des agronomes du Québec, s.d.). Ce changement important est survenu suite à la prise de conscience de la dégradation des ressources naturelles, ainsi que la pollution des sols, de l'air et de l'eau, dus principalement aux pratiques de l'agriculture intensive. Pour ces raisons, le domaine a connu une grande évolution afin de mettre en place de nouvelles technologies qui permettent de suivre les caractéristiques des sols et des cultures de manière à faciliter les diagnostics et les analyses des rendements agricoles. On peut alors acquérir des données par divers capteurs (caméras, drones, etc.) afin d'en extraire des informations sur l'état des cultures, comme la densité de végétation, la santé des plantes, ou même la présence de parasites. Également, des données récoltées par ces capteurs peuvent fournir des observations continues sur des paramètres environnementaux comme la température, l'humidité du sol, l'acidité du sol, ou encore sur la qualité de l'air. Cependant, ces données brutes nécessitent une transformation et un traitement pour être utiles aux agriculteurs (McDonald, 2015 ; Toupin, 2022). D'ailleurs, de vastes collections de données provenant de plusieurs sources sont disponibles en ligne et elles peuvent servir de point de départ pour l'analyse de données et la création de nouvelles perspectives de services pour les producteurs agricoles (Lee & Park, 2021). La valorisation des données devient donc un élément important étant donné qu'elle permet d'exploiter ces informations pour une meilleure gestion des fermes.

L'analyse d'image ou d'autres types de données permet d'adapter les pratiques afin d'aider les agriculteurs à améliorer les rendements de leurs cultures, mais aide également à réduire les coûts de production, favorise la préservation de l'environnement et permet des économies de ressources en limitant l'utilisation d'intrants agricoles là où ils ne sont pas nécessaires (Cheng, 2023). Il est également possible de réaliser une analyse historique des données afin de repérer des changements dans les rendements agricoles, ce qui permet de mieux comprendre l'impact des pratiques agricoles employées (Ohana-Levi *et al.*, 2022). D'ailleurs, ces informations deviennent indispensables pour la planification à long terme et pour l'élaboration de pratiques agricoles durables, ce qui permet d'évoluer vers une agriculture plus respectueuse de

l'environnement. De plus, en exploitant pleinement ces données, les agriculteurs peuvent prendre des décisions éclairées, optimiser leurs pratiques et contribuer à la préservation des ressources naturelles.

Ainsi, c'est dans cette optique que l'entreprise XLKey (<https://xlkey.ca/>) a pris l'initiative de recueillir sur plusieurs parcelles des données de haute qualité avec des outils à la fine pointe de la technologie. Cette collecte vise à permettre aux agriculteurs clients de comprendre les performances de leurs cultures et de leurs opérations en comparaison avec d'autres milieux agricoles similaires (XLKey, 2023). Les données seront stockées dans une base de données (BD), permettant ainsi l'analyse afin d'identifier des problèmes rapidement, parfois avant même qu'ils ne causent des dommages aux cultures. Ainsi, cette valorisation des données est réalisée dans le but de permettre aux agriculteurs clients d'augmenter leurs rendements, de diminuer leurs coûts de production ainsi que de minimiser les émissions de Gas à effet de serre (GES).

## 1.2 Objectifs

L'objectif principal de ce projet consiste à contribuer au développement d'un système d'aide au diagnostic en agriculture de précision, en utilisant des données provenant des producteurs agricoles de la région de Chaudière-Appalaches. Les traitements et les analyses visent à accroître les rendements agricoles tout en optimisant les coûts de production à travers la création de cartes de gestion des sols et des cartes de rendement.

Objectifs spécifiques :

1. Réalisation de cartes de zones de gestion par analyse des données de texture du sol;
2. Cartographies des variations des propriétés du sol et du relief des champs agricoles pour identifier les zones aux caractéristiques similaires, observer l'écoulement de l'eau et l'humidité du sol, puis automatisation des méthodes avec le langage Python;
3. Analyses géospatiales, statistiques en fonction des zones de gestion et modélisation du rendement (*Random Forest*) en fonction des propriétés du sol et du relief du terrain et automatisation des méthodes avec le langage Python.

## 2 Cadre Théorique

### 2.1 Agriculture de précision

L'agriculture de précision est une pratique de plus en plus courante qui utilise des technologies de pointe afin de récolter des données sur les sols et les cultures dans le but de leur appliquer des analyses géospatiales. Elle permet, entre autres, d'améliorer la gestion des exploitations agricoles afin d'optimiser les rendements. Par le fait même, comme les avancées technologiques sont en amélioration continue, on arrive à identifier de meilleures pratiques de gestion, ce qui permet ainsi de réduire les intrants, d'augmenter le profit des agriculteurs et de respecter l'environnement (Say *et al.*, 2018). Selon l'entreprise XLKey (2023), l'agriculture de précision est définie par quatre étapes importantes, soit : l'acquisition des données agricoles, la gestion de ces données, la prise de décision et finalement par la réalisation des opérations suite aux décisions rendues. L'ensemble de ces éléments permet de détecter rapidement et précisément des problématiques en ciblant des zones précises dans un champ et également, d'améliorer la gestion des intrants de ces zones. Donc, l'objectif ici est d'obtenir le rendement maximal de chaque zone d'un champ, que ce soit par l'analyse de sa variabilité intra-parcellaire ou par la comparaison entre les parcelles d'un même milieu. Pour arriver à bien définir les différentes zones de gestion, qui sont déterminées par la performance, la profitabilité, ou encore en lien avec des problématiques spécifiques, il est nécessaire d'utiliser des cartes de sol et de rendement, des images ou photographies acquises par drone ou des imageries satellites (XLKey, 2023). Ces éléments permettent ensuite de réaliser une division des différentes sections d'un champ afin d'identifier les zones rentables ou celles comportant des problèmes présents ou anticipés (Lherbier, 2005). Cette carte de zones de gestion est la base des analyses, car, par exemple, elle permet d'observer la variabilité dans un champs ou encore la variabilité inter-parcellaire. Ces observations nous permettent, par la suite, de réaliser des diagnostics par rapport au rendement observé dans le but d'améliorer les performances futures. Donc, la connaissance précise des données dans un champ agricole, notamment sur le sol et les cultures, permet de cibler de façon efficace les enjeux et de poser rapidement des actions pour augmenter la productivité, améliorer la rentabilité et diminuer les coûts de production, tout en minimisant l'impact environnemental.

## 2.2 Valorisation des données acquises sur les champs

La valorisation des données en agriculture de précision a pour objectif, entre autres, d'améliorer la gestion des exploitations agricoles, d'augmenter la productivité, de réduire les coûts, de prendre des décisions de plus en plus écologiques et durables. D'ailleurs, l'agriculture est un domaine qui utilise depuis des décennies des données afin d'étudier la santé des cultures, les propriétés du sol ainsi que pour obtenir de l'information sur les opérations agricoles. L'ensemble de ces données permet aux agriculteurs d'optimiser leur rendement et d'obtenir de meilleures récoltes dans une superficie donnée, en plus d'être avantageuse pour les distributeurs et détaillants qui peuvent suivre les prévisions et la gestion des stocks pour mieux répondre à la demande (Maycock, 2020). En d'autres mots, la valorisation des données agricoles permet de faciliter la prise de décision (Lherbier, 2005). Également, elle permet d'apporter une amélioration au niveau de la gestion des opérations en fonction des caractéristiques de la ferme, et ce grâce à l'accumulation de données précises permettant de mieux cibler les problématiques. On peut d'ailleurs noter quatre grandes étapes au niveau de la gestion de production agricole, soit : (1) la collecte et le stockage des données, (2) les prétraitements de ces données, (3) l'analyse des données à l'échelle intra- et inter-parcellaire et la production de l'information utile, et (4) l'information transmise au producteur afin d'optimiser leurs pratiques agricoles (XLKey, 2023). Ainsi, l'exploitation de ces données permet de contribuer à une agriculture plus durable qui contribue à l'amélioration de la santé globale des écosystèmes agricoles (Lherbier, 2005).

### 2.2.1 Acquisition des données

La collecte des données constitue la première étape dans la chaîne de valorisation des données. Cette étape est très importante, car la précision des analyses et la fiabilité des diagnostics émis sont tributaires de la qualité de ces données. Ceci explique, entre autres, pourquoi il y a eu une grande évolution dans le domaine de l'acquisition des données et des technologies des capteurs. Par exemple, dans certaines études, on exploite différents systèmes qui permettent de collecter des données sur les cultures et les variables environnementales sous forme de série chronologique ou encore, sur les conditions météorologiques (vitesse du vent, humidité, température, rayonnement solaire, précipitations) et également sur le suivi des opérations agricoles (géopositionnement du tracteur, taux d'application d'intrants, vitesse du tracteur, etc.) (Aydin & Aydin, 2020 ; Castro et al., 2017 ; Lee & Park, 2021 ; Reyes et al., 2012). En

agriculture de précision, l'évolution dans l'acquisition des données a permis de suivre les caractéristiques des parcelles de façon très précise, ce qui facilite les diagnostics et les analyses des rendements agricoles. On peut alors obtenir des données par caméras de drones ou encore à l'aide de capteurs embarqués sur la machinerie pour ensuite les traiter et les analyser afin d'aider à la prise de décision future. Il est également possible avec les avancées récentes d'obtenir des collectes de grands volumes de données sur des conditions environnementales et des pratiques agricoles *in situ* (Lee & Park, 2021 ; McDonald, 2015 ; Toupin, 2022). Dans le but d'obtenir des données de qualité et précises, l'entreprise XLKey s'est dotée des technologies les plus performantes à ce jour pour recueillir les données sur le sol et le terrain, les cultures ainsi que sur les opérations agricoles pour ses clients agriculteurs, ce qui incite l'entreprise à développer des méthodes d'analyses et de diagnostics de plus en plus précises.

#### 2.2.1.1 Données sur les propriétés du sol

Afin d'obtenir des caractéristiques du sol, l'entreprise XLKey a réalisé des prises de mesures via le capteur Veris MSP3 qui est installé derrière un tracteur agricole. Cet outil sophistiqué permet d'évaluer et de cartographier différentes propriétés du sol en récoltant des mesures géoréférencées, de manière simultanée. Il mesure notamment la conductivité électrique du sol (CE), l'acidité (pH), ainsi que la réflectance du sol dans les spectres du rouge (R) et du proche infrarouge (PIR). Cette dernière caractéristique est particulièrement utile pour obtenir des données sur la teneur en MO du sol (Veris Technologies, 2012).

Les valeurs récoltées de réflectance dans le PIR et R sont des données importantes dans l'analyse des propriétés du sol. Ces mesures permettent de caractériser la composition et la texture, ainsi que d'évaluer la santé des cultures. Les réflectances dans le spectre du PIR et du R fournissent des informations sur la structure et la composition du sol, c'est pourquoi elles sont souvent utilisées pour estimer la teneur en MO du sol (XLKey, 2023). De plus, il est important de noter que les sols nus possèdent une signature spectrale spécifique qui permet de les distinguer des autres types d'occupation du sol. Dans la région visible et proche infrarouge, la réflectance augmente proportionnellement avec la longueur d'onde, mais le taux d'augmentation dépend de nombreux facteurs, notamment la texture, la teneur en eau et la structure du sol, qui influencent la réflexion. Les sols avec une teneur en eau et en MO élevées absorbent plus et auront généralement une réflectance plus faible. Ces données sont ensuite traitées et analysées pour

générer des cartes de distribution spatiale, aidant ainsi les agriculteurs à prendre des décisions éclairées pour optimiser les pratiques de gestion des terres et des cultures (SEOS, s.d. ; Veris Technologies, 2012.).

Également, des données de LiDAR aéroporté provenant du ministère des ressources naturelles et des forêts (MRNF) ont été utilisées dans cette étude (MRNF, 2016). Elles permettent d'obtenir des informations topographiques détaillées sur les champs agricoles. En fait, cette technologie de télédétection aéroportée permet aux agriculteurs d'obtenir des modèles numériques de terrain (MNT) précis, permettant une analyse approfondie de la topographie des terres agricoles, notamment sur les écoulements des eaux. Les scanners LiDAR émettent des faisceaux laser vers le sol et enregistrent le temps qu'il faut pour que ces faisceaux rebondissent. Cette information est utilisée conjointement avec les mesures topographiques pour cartographier avec une grande précision les altitudes du terrain (ArcGIS, 2023).

Les données LiDAR aéroportées fournissent des informations essentielles pour la planification et la gestion agricole. Elles permettent d'identifier avec précision les zones à forte pente, les dépressions et les zones sujettes à l'accumulation d'eau, facilitant ainsi la prise de décisions éclairées en matière d'irrigation, de drainage et d'aménagement des cultures (FlyGuys, 2021). De plus, les données LiDAR peuvent être utilisées pour évaluer l'érosion des sols, optimiser le tracé des routes et des chemins agricoles, ainsi que pour la modélisation des bassins versants. En outre, elles permettent d'effectuer d'autres analyses géospatiales telles que l'indice de position topographique (TPI), qui fournit des informations supplémentaires sur les dépressions et les reliefs dans un champ, enrichissant ainsi la compréhension de la topographie et des caractéristiques du terrain (De Reu *et al.*, 2013).

#### 2.2.1.2 *Données sur les opérations*

Dans cette étude, certaines données issues de la machinerie agricole, notamment les rendements de cultures, ont été collectées en vue d'une analyse approfondie. Les données relatives aux rendements des cultures, qu'elles concernent la matière sèche (MS) ou la matière humide (MH) des récoltes, permettent d'évaluer les performances des cultures et améliorer les pratiques agricoles en vue de maximiser les rendements, tout en réduisant les intrants au minimum nécessaire.

La collecte des données de rendement s'effectue généralement à l'aide de moissonneuses-batteuses équipées de systèmes de positionnement global (GPS) intégrés (XLKey, 2023). Ces systèmes permettent de capturer précisément la quantité récoltée à différents endroits du champ, en émettant des points géolocalisés tout au long de l'opération de récolte. D'ailleurs, au cours des dernières années, les moissonneuses-batteuses ont été de plus en plus équipées d'options technologiques avancées visant à maximiser les rendements, tout en préservant la qualité des récoltes (Gagnon, 2023). On peut également obtenir d'autres types d'informations agronomiques avec cet outil telles que les données sur l'irrigation du sol et les dates de semis (XLKey, 2023). Cette approche globale permet de rassembler un ensemble complet de données, offrant ainsi une vision détaillée et précieuse de la dynamique des cultures et des pratiques agricoles.

De ce fait, l'acquisition des données sur le rendement des cultures constitue une étape fondamentale dans le processus d'évaluation et de gestion des exploitations agricoles. En combinant les technologies de géolocalisation avec des méthodes de traitement et d'analyse de données avancées, cette approche fournit aux agriculteurs et aux chercheurs des informations précieuses pour améliorer la productivité et la durabilité des systèmes agricoles.

#### *2.2.1.3 Données sur les cultures*

Les grandes cultures au Québec sont principalement le maïs, le soya, les céréales à paille (blé, orge et avoine) ainsi que le canola (Gouv. Québec, 2024). Chaque culture, qu'il s'agisse du soja ou du maïs, se développe de manière optimale dans un environnement présentant des caractéristiques spécifiques adaptées à ses besoins. De ce fait, la comparaison intra- ou inter-parcelle des cultures permet de faire le diagnostic des cultures, notamment en fonction des caractéristiques des parcelles. Les variations du rendement peuvent être évaluées selon les différentes cultures et ce, au sein d'une même parcelle ou entre différentes parcelles avec des conditions environnementales semblables. Finalement, en divisant les caractéristiques analysées en fonction des types de culture, il est possible d'optimiser les rendements agricoles et la gestion durable des ressources.

### 2.3 Définition des produits dérivés

Pour élaborer un système d'aide au diagnostic, deux types de produits dérivés sont créés dans ce projet : des cartes de gestion du sol et des cartes de rendement.

### 2.3.1 Carte de gestion des sols

Les cartes de gestion du sol représentent des outils indispensables dans le domaine de l'agriculture de précision. Elles permettent une caractérisation approfondie de la variabilité spatiale des propriétés du sol, incluant la texture, la structure, le pH ainsi que la teneur en matière organique et en nutriments (Oliveira *et al.*, 2022). Ces cartes sont générées à partir de données collectées sur le terrain, telles que des échantillons de sol, des mesures directes, ou des données satellitaires et des informations provenant de capteurs embarqués sur des équipements agricoles (Heung *et al.*, 2021). Dans Brevik *et al.* (2006), des cartes de gestion des sols ont été produites à partir de données de CE. La CE permet d'observer la variabilité des sols au sein d'une parcelle. Les données de CE permettront d'obtenir de l'information sur la texture du sol. Dans le cadre de ce projet, l'utilisation du capteur Veris MSP3 a permis d'obtenir des informations sur la CE, le pH, et la réflectance des spectres dans les bandes du R et du PIR. Ces données sont essentielles pour caractériser les propriétés du sol de chacun des champs étudiés (XLKey, 2023).

Grâce à des techniques d'analyse géospatiale et statistique, les informations recueillies sont interprétées pour produire des cartes décrivant la distribution des propriétés du sol dans un champ. Ces cartes fournissent aux agriculteurs des informations essentielles pour adapter leurs pratiques de gestion, telles que la fertilisation, l'irrigation et le choix des cultures, en fonction des besoins spécifiques de chaque zone du champ (XLKey, 2023).

### 2.3.2 Cartes de rendement

Les cartes de rendement constituent des outils essentiels dans le domaine de l'agriculture de précision. Elles offrent une représentation spatiale détaillée des performances des cultures à travers une parcelle agricole, car elle définit des zones de gestion (ZG) en fonction du rendement. Ces cartes sont générées à partir d'une multitude de données, comprenant des mesures de rendement récoltées par des capteurs embarqués sur des machines agricoles, des observations visuelles, des relevés manuels ou encore, des analyses géospatiales (Martel *et al.*, 2015 ; XLKey, 2023). Selon XLKey (2023), les données de rendement peuvent provenir d'une moissonneuse-batteuse équipée de capteurs de pointe. Ces données recueillies peuvent ensuite permettre de cartographier les variations de rendement à travers le champ (XLKey, 2023). Par la suite, en analysant ces cartes, il est possible d'identifier les zones de haute ou de faible

performance et d'adapter de manière stratégique les pratiques de gestion agricole afin d'optimiser les rendements et la rentabilité.

De plus, si ces cartes de rendement sont réalisées chaque année, il est alors possible de suivre l'évolution de l'efficacité d'un champ au fil des années analysées, en fonction des changements apportés dans les différentes zones de gestion (Martel *et al.*, 2015). On peut alors observer si les changements suggérés effectués sont efficaces par rapport aux objectifs de performances.

## 2.4 Identification des zones de gestion

Les zones de gestion (ZG) en agriculture permettent de caractériser la variabilité spatiale intra-parcellaire au sein d'un champ agricole. Ces ZG peuvent être délimitées grâce à l'analyse de diverses données portant sur les propriétés chimiques et physiques du sol, ainsi que sur la conductivité électrique du sol et les données de rendements historiques (Yuan *et al.*, 2022). Les sous-régions du champ ainsi créées permettent par la suite de classer les milieux en fonction de leur performance et de leur rentabilité, ou encore en réponse à une problématique spécifique, telle que l'optimisation de l'irrigation en fonction des niveaux d'humidité identifiés grâce à la CE du sol, ou la gestion différenciée de la fertilisation en tenant compte des variations de texture du sol et du taux de MO (IRDA, 2016). Ces stratégies personnalisées, basées sur l'analyse des données de chaque zone, permettent une meilleure allocation des ressources et une gestion plus efficace des cultures. Les ZG peuvent être classifiées selon différentes méthodes, telles que les techniques d'analyse groupée qui segmentent les données en groupes homogènes, favorisant la similarité intra-groupe tout en réduisant la similarité inter-groupe. D'autres approches incluent l'utilisation d'algorithmes de détection d'anomalies pour identifier les zones inhabituelles ou présentant des comportements atypiques. En outre, les méthodes de classification hiérarchique peuvent être utilisées pour construire une hiérarchie de regroupements, ce qui permet une exploration plus détaillée des relations entre les différentes zones de gestion. La classification non-supervisée « *clustering* » par k-means est également un outil d'exploration de données et permet de regrouper et sélectionner différents traits agronomiques des ZG (Han *et al.*, 2018). Cette méthode qui permet d'extraire des modèles et des connaissances à partir d'un grand volume de données (Heka.AI, 2022). De même, la classification non supervisée peut être appliquée à des séries temporelles, offrant ainsi une dimension évolutive qui permet d'analyser les données en

fonction du temps. Cette adaptabilité rend ces analyses flexibles et ajustables aux nouvelles données collectées et aux nouveaux sites d'étude créés au fil du temps.

## 2.5 Diagnostic des rendements au champ

Le diagnostic des rendements agricoles est une pratique qui permet d'évaluer systématiquement la performance d'une exploitation agricole. C'est un processus qui vise à évaluer l'efficacité de la gestion agricole, à identifier les problèmes potentiels et à proposer des solutions pour améliorer la productivité, la rentabilité et la durabilité de la ferme (Sharma *et al.*, 2023). La variabilité des rendements à l'intérieur des parcelles peut être attribuée à plusieurs facteurs. Selon la répartition des rendements dans un champ, qu'elle soit uniforme ou non, il est possible d'identifier et de diagnostiquer diverses causes potentielles, qu'elles soient naturelles ou résultant des pratiques culturales, à l'origine de ces variations de rendement (Martel *et al.*, 2015 ; XLKey, 2023). La superposition des différentes cartes telles que (1) le rendement, (2) les textures du sol, (3) la topographie et (4) l'humidité du sol permet de clarifier comment le type de sol et la pente influent sur le drainage des parcelles et les rendements agricoles (Martel *et al.*, 2015). Afin d'obtenir une meilleure précision quant au diagnostic posé, les données seront éventuellement accumulées sur plusieurs années, ce qui permettra d'obtenir un historique de rendement des cultures.

### 2.5.1 Évaluation des rendements

Nous avons vu précédemment qu'avec les technologies récentes, il est possible d'estimer les rendements en se basant sur les données collectées aux champs. Cette méthode permet alors d'optimiser la gestion des terres agricoles tout en améliorant la productivité et la rentabilité de l'entreprise et en réduisant la quantité d'intrants. On peut d'ailleurs observer que l'ajout d'intrants, par exemple de nutriments ou de MO, permet parfois d'améliorer la production au champ, mais cet ajout comporte un coût économique et a également un effet sur l'écosystème. C'est pourquoi il est essentiel de s'en tenir à une quantité contrôlée (Timsina *et al.*, 2021). De plus, les taux d'engrais qui sont recommandés au Québec, pour un type de culture, correspondent à des valeurs générales et donc, ne tiennent pas compte des éléments tels que les conditions pédologiques et météorologiques spécifiques à un champ (CPVQ, 1996 ; CRAAQ, 2010 ; Timsina *et al.*, 2021). De ce fait, le suivi des données correspondant à l'écosystème de chacun

des champs afin d'obtenir les bonnes doses d'intrants à ajouter est nécessaire afin d'optimiser les rendements.

Dans ce contexte, des outils d'analyse comme l'algorithme *Random Forest* permettent d'évaluer les rendements agricoles en réalisant une prédiction. Il s'agit de modéliser des relations complexes entre divers facteurs influents, tels que les caractéristiques du sol, les données climatiques et les pratiques culturales, entre autres. En intégrant ces données, il est possible d'établir des prévisions de rendement qui tiennent compte des spécificités de chaque parcelle. L'algorithme de modélisation *Random Forest* se distingue également par sa capacité à gérer des données hétérogènes et volumineuses, ainsi que des données manquantes, courantes en milieu agricole (Marques Ramos *et al.*, 2020).

Le rendement est calculé en quantifiant la masse des cultures récoltées. Le résultat de cette étape peut être obtenu sous forme de carte de rendement qui permet de visualiser la variation des rendements de façon intra- et inter-parcellaire (XLKey, 2023). D'ailleurs, les modèles de prédiction comme le *Random Forest* permettent d'identifier avec plus de précision les différentes zones de rendement dans un champ afin de fournir des recommandations spécifiques à chaque parcelle (Marques Ramos *et al.*, 2020). Ceci permet par la suite d'ajuster au besoin la quantité d'intrants, tels que les engrais ou les pesticides, selon les zones de rendement dans un champs, tout en améliorant la précision de la gestion agricole ainsi qu'en réduisant les impacts environnementaux.

### 2.5.2 Analyse des pratiques agricoles

Les pratiques des agriculteurs sont analysées afin d'identifier les méthodes et techniques utilisés pour la gestion de leurs fermes. Cette analyse permet d'évaluer l'efficacité, la durabilité ainsi que la rentabilité de l'exploitation agricole en analysant leur manière de faire, la quantité d'intrants utilisés ou autres éléments pertinents tels que la gestion de l'eau, la qualité du sol, les semences, etc. (XLKey, 2023). Cette étape devient un outil essentiel afin d'aider les agriculteurs à améliorer leurs opérations et à faciliter l'adaptabilité face à des imprévus. Dans Mairura *et al.* (2007b), il a été observé que les parcelles ayant reçu le moins de MO, donc un déficit en éléments nutritifs, étaient significativement moins fertiles que les champs qui avaient obtenu des éléments nutritifs en abondance et ce, de façon constante. Donc, une corrélation positive a été observée entre les indicateurs de fertilité du sol et les concentrations en nutriments (Tittone, *et*

*al.*, 2005). Les pratiques agricoles ont un rôle significatif pour les agriculteurs, car elles permettent d'établir une agriculture durable en cultivant des agroécosystèmes qui se restaurent afin de conserver une fertilité du sol dans le temps. En d'autres mots, une bonne gestion des pratiques agricoles permet une agriculture durable et donc, un bon rendement et ce, sans causer de dommages irréversibles sur la santé des écosystèmes agricoles (Mairura *et al.*, 2007b).

### 2.5.3 Analyse des facteurs environnementaux

Les conditions environnementales influencent grandement les rendements aux champs. Parmi ces facteurs environnementaux, notons la gestion de l'eau et le nivellement des sols, les types de sol ou encore, le climat (Sharma *et al.*, 2023). Ceux-ci représentent des variables essentielles afin de maximiser le potentiel de rendements agricoles et pour la réduction des coûts et des intrants (Pradipta *et al.*, 2022). En agriculture, plusieurs éléments sont interconnectés, ce qui en fait un système complexe à analyser. On peut observer que les sols plus riches en MO obtiennent de meilleur rendement et se retrouvent donc plus fertile. Tandis que dans les types de sol comportant plus de limon, on obtient plus de risque d'érosion des sols, ce qui amène une contrainte pour la production agricole (Mairura *et al.*, 2007a). De ce fait, la morphologie des champs, qu'elle soit en pente ou avec des dénivelés importants influencera grandement la production agricole.

### 2.5.4 Évaluation économique

Nous avons vu précédemment que l'agriculture de précision était une pratique moderne qui permettait d'améliorer les rendements agricoles en utilisant des technologies avancées. Cependant, cette pratique a certains coûts qui y sont associés. Elle permet d'améliorer les rendements, mais, parfois, les changements apportés peuvent comporter un coût, par exemple, l'achat de nouveaux outils. Tandis que certains changements vont plutôt diminuer les coûts, par exemple, la diminution d'intrants. C'est pourquoi il demeure important de déterminer si une exploitation agricole génère un bénéfice supérieur à ses coûts associés en réalisant ces changements. Pour ce faire, il devient important d'analyser les coûts et les avantages économiques liés à l'adoption des technologies de précision (Jacques, 2022). Ainsi, il s'agit d'évaluer si l'intégration de ces technologies récentes devient une option rentable pour les agriculteurs. Autrement dit, est-ce que les coûts des modifications apportés sont inférieurs au coût de production pour que l'exploitation agricole demeure rentable suite au changement

apporté. Le changement qui permet d'apporter une agriculture durable ne peut qu'être bénéfique pour l'agroenvironnement, mais la motivation principale pour les agriculteurs demeure le bénéfice plus élevé, à long terme, rendu possible par ces changements. En d'autres mots, est-ce que les changements apportés couvrent l'investissement nécessaire ? Donc, il devient essentiel d'évaluer si l'agriculture de précision est rentable, efficace et également stratégique pour les agriculteurs (XLKey, 2023).

## 3 Matériel et méthodes

### 3.1 Présentation des sites d'étude

Les données de cette étude ont été collectées dans trois exploitations agricoles de la région de Chaudière-Appalaches : la ferme « Forfait André et Frédéric Roy » (fig.1), la ferme « Marcel Nadeau » (fig.2) et la ferme « 9077-9729 Qc. » (fig.3). Ces producteurs se spécialisent dans les grandes cultures, et pour ce projet, seules les cultures de maïs et de soya ont été prélevées. Les champs analysés dans cette étude sont répartis dans trois municipalités régionales de comté (MRC), soit dans la MRC de Lotbinière, de la Nouvelle-Beauce et de Bellechasse.

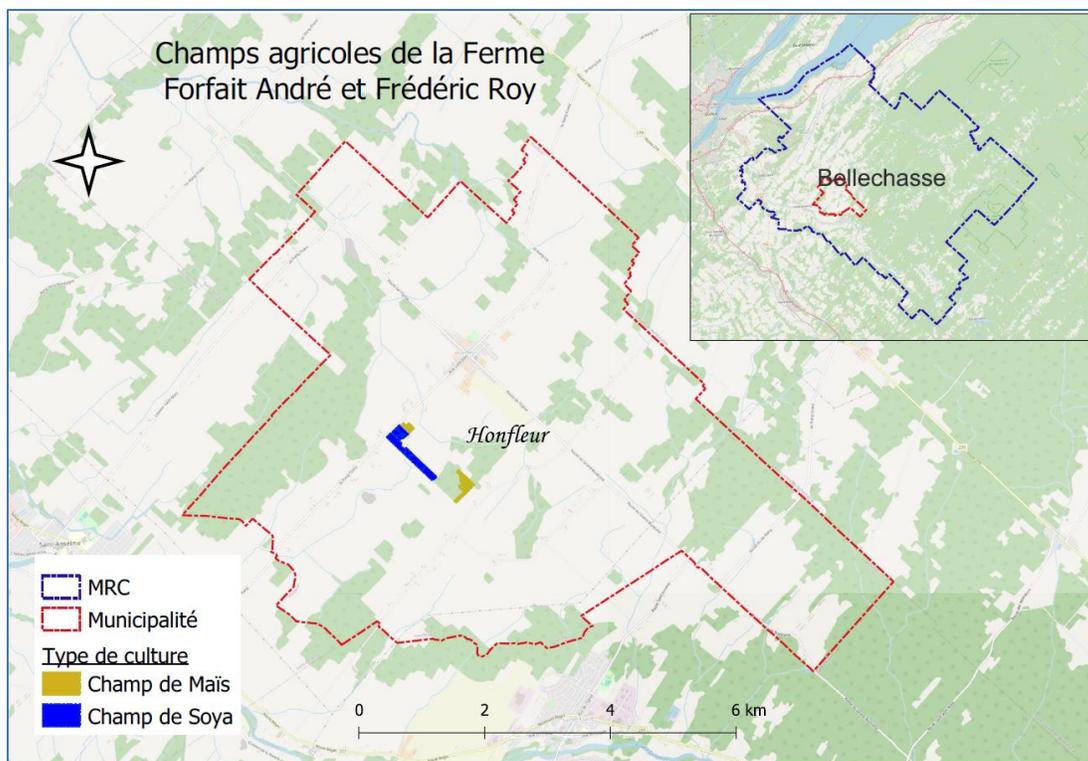


Figure 1. Localisation des 5 parcelles agricoles à l'étude de la ferme « Forfait André et Frédéric Roy à Honfleur »

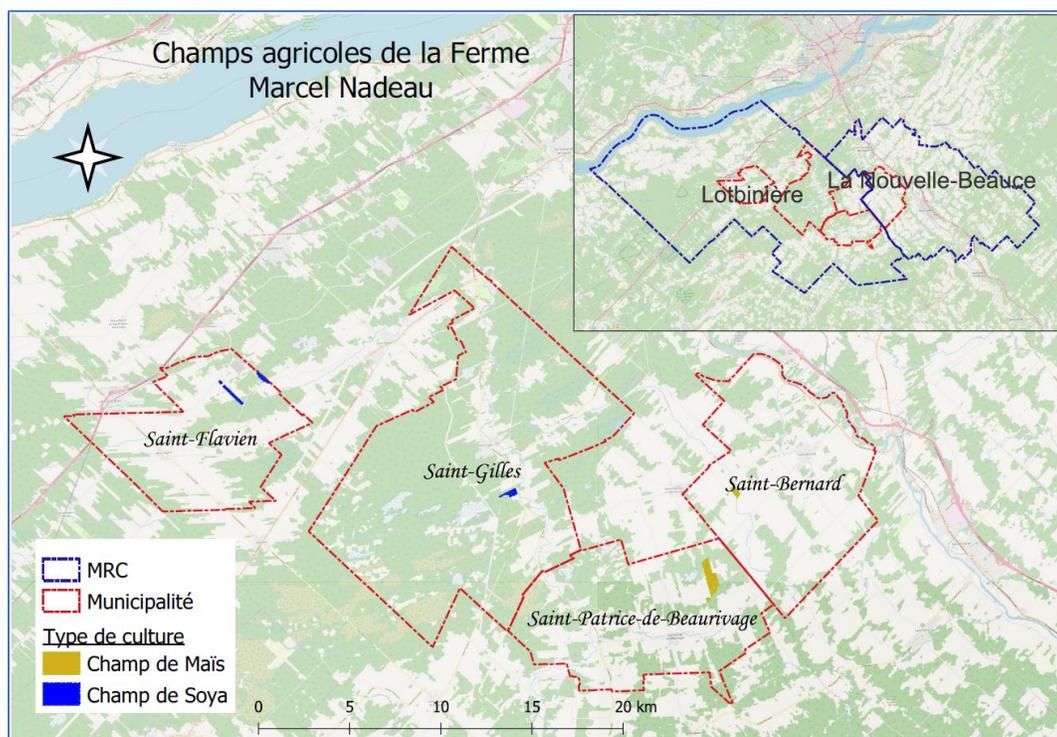


Figure 2. Localisation des 9 parcelles agricoles à l'étude de la ferme « Marcel Nadeau » dans les municipalités de Saint-Flavien, Saint-Gilles, Saint-Bernard et Saint-Patrice-de-Beaurivage.

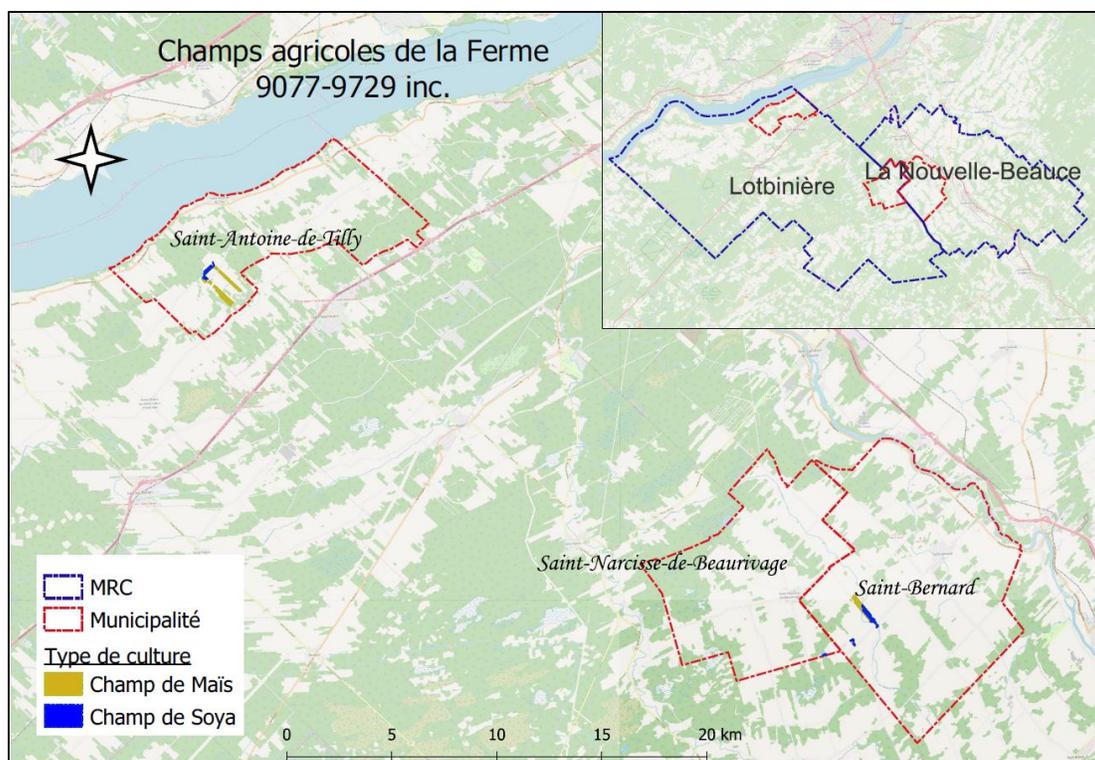


Figure 3. Localisation des 26 parcelles agricoles à l'étude de la ferme « 9077-9729 inc. » dans les municipalités de Saint-Bernard, Saint-Narcisse de Beaurivage et Saint-Antoine de Tilly.

Sur l'ensemble des parcelles agricoles, nous retrouvons plusieurs types de textures de sol, allant d'un sol à texture très fine à un sol à texture très grossière. Cette variabilité de texture de sol a été mesurée à l'aide de la sonde de sol Veris dans l'ensemble des champs (Veris MSP3, 2012). De ce fait, nous observons une diversité significative dans les propriétés physiques du sol, ce qui peut influencer la capacité de rétention d'eau, la perméabilité, et la disponibilité des nutriments pour les cultures. De plus, les différentes parcelles agricoles présentent des variations d'altitude tant à l'intérieur qu'entre les parcelles, ce qui crée des disparités de niveau du terrain à travers le champ. Par conséquent, certaines zones du champ présentent des caractéristiques topographiques telles que des pentes abruptes, des zones plus humides ou plus sèches, ainsi qu'une répartition inégale de l'eau et des nutriments. Ces éléments ont une influence directe sur le rendement des cultures.

Dans ce projet, les données provenaient à la fois de données brutes et de données prétraitées et nettoyées par l'entreprise XLKey. L'ensemble des données brutes est décrit dans le tableau 1. Les données ont été définies en trois types, soit les données sur les caractéristiques du sol, les opérations ainsi que les cultures.

Tableau 1. Source des données utilisées dans le cadre de ce projet.

| CATÉGORIE DE DONNÉE  | INFORMATION QUE CONTIENT LA DONNÉE                      | DE QUEL FAÇON EST PRISE LA DONNÉE               | FORMAT DE LA DONNÉE | SOURCE          | INFORMATION SUPPLÉMENTAIRE                                   |
|----------------------|---|---|---------------------|-----------------|--|
| <b>SOL / TERRAIN</b> | Limite de contour du champ                              | Transmise par les agriculteurs                  | .shp                | BDPPAD          |  |
|                      | Conductivité électrique (CE)                            | Capteur Veris MSP3 (clé usb)                    | .DAT, .txt          | XLKey           |  |
|                      | Matière organique (Mo)                                  | Capteur Veris MSP3 (clé usb)                    | .DAT, .txt          | XLKey           | Réflectance de la bande spectrale rouge et proche infrarouge |
|                      | pH  | Capteur Veris MSP3 (clé usb)                    | .DAT, .txt          | XLKey           |  |
|                      | Élévation de terrain                                    | LiDAR   | Tiff                | MRNF            |  |
|                      | Surfaces naturelles ou artificielles de l'environnement | Imagerie Drone                                  | Tiff ou Geotiff     | XLKey           | PIX4DMapper (photogrammétrie)                                |
| <b>CULTURE</b>       | Type de culture   | Transmise par les agriculteurs ou via la BDPPAD | .csv                | XLKey et BDPPAD |  |
| <b>OPÉRATION</b>     | Poids collecté / Superficie                             | Moissonneuse-batteuse                           | .csv                | XLKey           | Conversion nécessaire de la machinerie au format exploitable |

Certains champs spécifiques ont été choisis par le collaborateur XLkey pour une analyse approfondie, notamment en raison de la complétude des données disponibles. Les données relatives au rendement de MS ou MH, à la CE et à la réflectance dans les bandes du R et du PIR ont été prétraitées et nettoyées par l'équipe XLKey afin de générer des données matricielles (image .tiff). Également, des données LiDAR provenant du MRNF ont été utilisées. Une partie de l'ensemble de ces données est présentée dans le tableau 2. Les champs Be01, Be02, Be03 et Be04, appartenant à la ferme Marcel Nadeau, ont été fournis pour la réalisation des cartes d'échantillonnage par zone. Cependant, seules les données de la CE ont été fournies pour ces champs. Par conséquent, ils ne figurent pas dans le tableau 2.

Tableau 2. Description des données qui ont été sélectionnée pour les analyses.

|  | TYPE    | NOM DE LA PARCELLE | YIELD(RENDEMENT) | VERIS | LIDAR |     |      |      |      |
|--|---------|--------------------|------------------|-------|-------|-----|------|------|------|
| <b>FORFAIT<br/>ANDRÉ<br/>ET FRÉDÉRIC<br/>ROY</b> | Soya    | 01                 | 2022             | 2022  | 2018  |     |      |      |      |
|  |         | 02                 |                  |       |       |     |      |      |      |
|  | Maïs    | 05                 | 2021             | 2022  | 2018  |     |      |      |      |
|  |         | FR01               |                  |       |       |     |      |      |      |
|  |         | FR02               |                  |       |       |     |      |      |      |
| <b>FERME<br/>MARCEL<br/>NADEAU</b>               | Soya    | AGA01              | 2022             | 2022  | 2012  |     |      |      |      |
|  |         | LAB01              |                  |       |       |     |      |      |      |
|  |         | LAB02              |                  |       |       |     |      |      |      |
|  | Maïs    | GIL01              | -                | 2022  | 2016  |     |      |      |      |
|  |         | MEU01              | -                | 2022  | 2016  |     |      |      |      |
| <b>FERME 9077-9729 QC INC</b>                    | Maïs    | B01                | 2022             | 2022  | 2016  |     |      |      |      |
|  |         | BS01               |                  |       |       |     |      |      |      |
|  |         | BS02               |                  |       |       |     |      |      |      |
|  |         | BS03               |                  |       |       |     |      |      |      |
|  |         | A05                |                  |       |       |     |      |      |      |
|  | Maïs    | A06                | 2021             | 2021  | 2016  |     |      |      |      |
|  |         | A07                |                  |       |       |     |      |      |      |
|  |         | A01                |                  |       |       |     |      |      |      |
|  |         | A09                |                  |       |       |     |      |      |      |
|  |         | A12                |                  |       |       |     |      |      |      |
|  |         | A13                |                  |       |       |     |      |      |      |
|  |         | A14                |                  |       |       |     |      |      |      |
|  |         | A15                |                  |       |       |     |      |      |      |
|  |         | Soya               |                  |       |       | N01 | 2022 | 2022 | 2016 |
|  |         |                    |                  |       |       | A02 |      |      |      |
|  | A03-A04 |                    |                  |       |       |     |      |      |      |
|  | B02     |                    |                  |       |       |     |      |      |      |
|  | B03     |                    |                  |       |       |     |      |      |      |
|  | B04     |                    |                  |       |       |     |      |      |      |
|  | B05     |                    |                  |       |       |     |      |      |      |
|  | B06     |                    |                  |       |       |     |      |      |      |
|  | B07     |                    |                  |       |       |     |      |      |      |
|  | B08     |                    |                  |       |       |     |      |      |      |
|  | B09     |                    |                  |       |       |     |      |      |      |
|  | BS04    |                    |                  |       |       |     |      |      |      |
| BS05   |         |                    |                  |       |       |     |      |      |      |

## 3.2 Collecte des données

### 3.2.1 Données sur les propriétés du sol

Les données nécessaires aux analyses du sol proviennent de trois sources distinctes. Le contour du champ est extrait de la base de données des parcelles et des productions agricoles déclarées (BDPPAD), tandis que les données LiDAR sur les élévations intra-parcellaire proviennent du MRNF. Les mesures de conductivité électrique (CE), de pH, ainsi que les données de réflectance dans les spectres du PIR et du R sont produites par l'entreprise XLKey, collectées à l'aide du capteur Veris MSP3, une sonde de sol montée sur un tracteur agricole (fig.4) (Veris Technologies, 2012).



Figure 4. Capteur Veris MSP3 qui permet de calculer des données sur la conductivité électrique (EC), d'obtenir la mesure de pH du sol ainsi que des données sur les réflectances (PIR et R) pour obtenir des valeurs de matière organique du sol (OM).

Le capteur Veris MSP3 recueille des données ponctuelles sur la conductivité électrique, le pH du sol et les réflectances PIR et R qui sont liées à la teneur en MO. Fixé à l'arrière d'un tracteur, il effectue des mesures sur toute la zone d'étude. Afin d'assurer la précision des mesures de CE, le capteur Veris MSP3 vérifie la continuité de son signal électrique et l'isolation de ses électrodes avant chaque session de collecte de données, avec une répétition de cette vérification toutes les 10 heures de collecte en continu. De plus, un suivi de vérification est effectué à chaque nouvelle utilisation sur le terrain à l'aide du système de cartographie optique (OM - *optimap*) pour garantir le bon fonctionnement du capteur optique (donnée sur la réflectance) et du contrôleur

mesurant la conductivité électrique. La profondeur de la sonde est également enregistrée pour garantir la constance des mesures, exprimées en millisiemens par mètre (mS/m), avec une profondeur idéale comprise entre 1 et 2 pouces (Veris Technologies, 2012). L'ensemble de ces vérifications faites à chaque prise de mesure permet d'utiliser la CE afin de déterminer la texture du sol étant donné que les conditions d'humidités et de topographie sont relativement constante lors des prises de mesure par le Veris MSP3.

Le capteur de pH du Veris est calibré avant chaque mesure avec un tampon à pH 4. Deux mesures sont prélevées lors de la collecte de données sur le terrain, et une moyenne est calculée à l'aide d'un logiciel dédié. Les valeurs de pH sont exprimées sur l'échelle logarithmique de 0 à 14 unités (Veris Technologies, 2012). Toutes ces données sont ensuite exportées au format .csv puis converties en shapefile pour être utilisées dans le logiciel QGIS, simplifiant ainsi leur analyse (QGIS, 2022).

### 3.2.2 Données sur les opérations

Les données relatives au rendement des cultures, qu'il s'agisse de la quantité de MS ou MH, sont recueillies à partir des résultats de la quantité récoltée par une moissonneuse-batteuse. Ces données sont captées par les systèmes de positionnement global (GPS) intégrés à la moissonneuse-batteuse, qui émettent des points géolocalisés. Par la suite, ces données sont décryptées manuellement et converties en fichiers shapefile afin de faciliter leur analyse. Ce processus est également utilisé pour recueillir d'autres types d'informations, notamment sur l'irrigation du sol et les dates de semis. L'acquisition des données sur le rendement des cultures a été réalisé en fin de saison, soit au mois d'octobre pour la culture du maïs et à la fin du mois de septembre pour la culture du soya (XLKey, 2023).

### 3.2.3 Données sur les cultures

Les données sur les types de cultures dans les champs ont été transmises par les agriculteurs ainsi que par le biais de la BDPPAD. De plus, l'entreprise XLKey a mis en place un système permettant aux agriculteurs d'ajouter ces informations directement dans une BD. Ce système permet également de spécifier si un champ comporte une zone cultivée d'un type spécifique de culture tandis qu'à une extrémité, il présente un autre type de culture. Cette fonctionnalité facilite la gestion et l'exploitation de ces données (XLKey, 2023).

### 3.3 Méthodologie et traitement des données

Le schéma méthodologique du projet, qui décrit les traitements effectués, est présenté à la figure 5. La démarche se compose principalement de 6 étapes :

1. Acquisition et prétraitement des données
2. Réalisation d'une carte d'échantillonnage par zone
3. Présentation d'une carte d'échantillonnage de calibration « CE-MO »
4. Réalisation d'une carte de l'indice de position topographique (TPI)
5. Analyses géospatiales et calcul des corrélations des données agricoles
6. Modélisation du rendement

Ainsi, plusieurs cartes ont été produites (fig.5). Pour commencer des cartes d'échantillonnage par zone, basées sur la texture du sol déterminée à partir des données de CE (Brevik *et al.*, 2006). Ensuite, des cartes d'échantillonnage pour la calibration « CE-OM » ont été générées en superposant les données de conductivité électrique et de matière organique par méthode de classification par regroupement (*clustering*). Finalement, des cartes décrivant l'indice de position topographique ont été produites en utilisant des données LiDAR. L'ensemble de ces cartes de gestion du sol contribuent à optimiser l'utilisation des ressources, à améliorer les rendements et à promouvoir une agriculture durable et respectueuse de l'environnement, en fournissant un support décisionnel précieux pour les agriculteurs.

Tous les traitements de programmation ont été effectués en utilisant Python 3.6 sur la plateforme *Google Colaboratory* (Google, 2023), en utilisant les bibliothèques suivantes : *numpy*, *matplotlib.pyplot*, *gdal*, *pandas*, ainsi que les modules *train\_test\_split*, *RandomForestRegressor*, *mean\_absolute\_error*, *r2\_score*, *mean\_squared\_error* de *sklearn.model\_selection* et *sklearn.ensemble*. De plus, les modules *k-means* de *sklearn.cluster* et *distance* de *scipy.spatial* ont été utilisés pour effectuer des opérations sur la classification non-supervisée de type « *clustering* ». Les cartographies ont été réalisées dans QGIS 3.16.11 (QGIS, 2022).

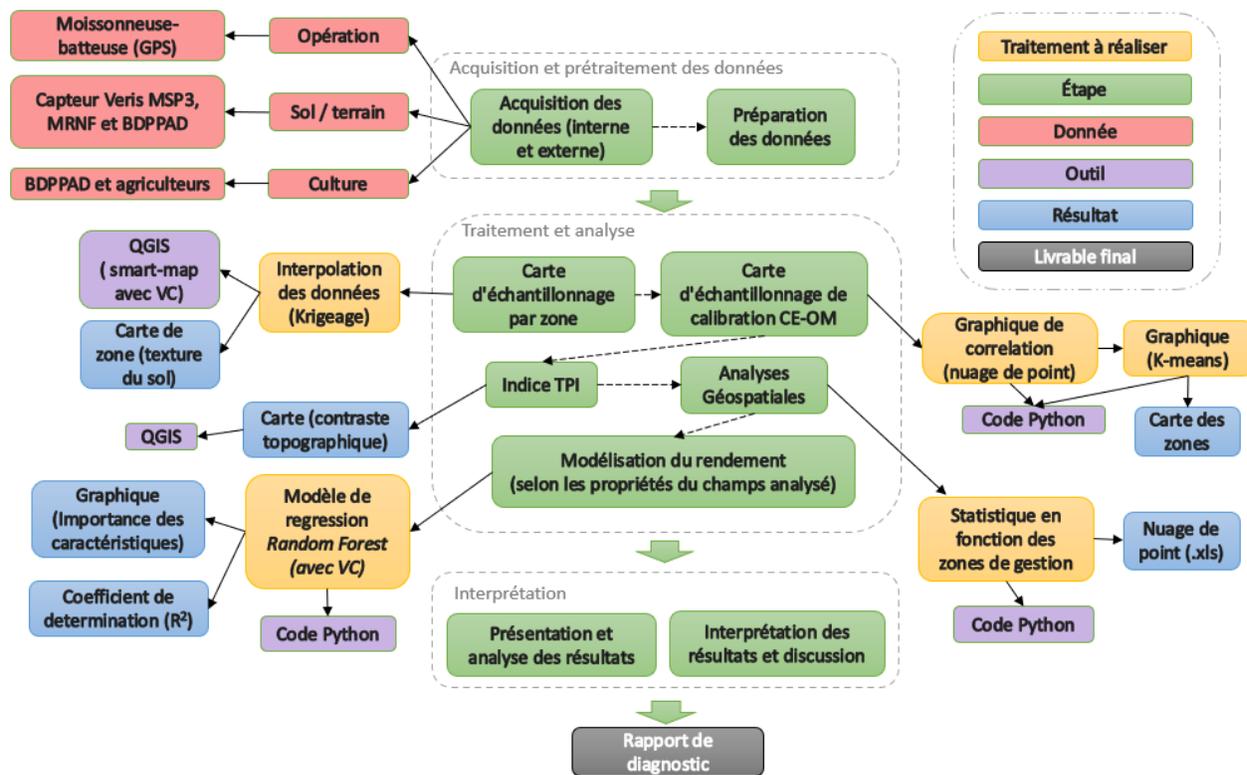


Figure 5. Organigramme méthodologique pour ce projet.

### 3.3.1 Prétraitement et interpolations des données

Comme il a été mentionné précédemment, les données utilisées dans ce projet proviennent à la fois de données brutes et de données prétraitées et nettoyées par l'entreprise XLKey. Lors d'analyse avec les données brutes, qui peuvent parfois comporter des imperfections telles que des erreurs de mesure, des valeurs aberrantes ou des données manquantes, il est nécessaire de réaliser un prétraitement sur ces données afin de garantir une qualité, une cohérence et une fiabilité dans les informations utilisées dans les analyses quantitatives.

Dans le cadre de ce projet, seul le jeu de données de rendement nécessitait un prétraitement. Après avoir analysé l'histogramme de distribution des données, il a été convenu de ne conserver que les données situées entre les valeurs 2 et 20 t/ha. Toutes les autres données n'ont pas nécessité de prétraitement pour les analyses ultérieures.

Après la préparation des données qui consiste à éliminer les données aberrantes, une méthode par *krigeage* et d'apprentissage automatique a permis de créer un modèle numérique de terrain (MNT) à l'aide de l'extension « *smart-map* » dans le logiciel QGIS (QGIS, 2022). Ceci a permis

d'utiliser toutes les données de type point, soit les données sur la conductivité électrique, le pH du sol ainsi que les données de réflectance dans le PIR et le R afin de déterminer la teneur en MO du sol et d'en extraire une couche matricielle sur l'ensemble de la superficie du champ. Les données sur le rendement, qui est représenté par la masse sèche des résidus de culture, étaient de type polygone.

La validation croisée « en omettre un » (LVC-1) a été utilisée dans l'extension « *smart-map* » afin de déterminer dans quelle mesure le modèle d'interpolation utilisé correspondait aux données. Cette validation croisée consiste à utiliser toutes les données, sauf une, qui est omise volontairement, pour ensuite interpoler cette donnée par la méthode d'interpolation choisie afin de déterminer la précision du modèle (Pereira et al., 2022).

### 3.3.2 Carte d'échantillonnage par zone

Les différentes étapes pour générer une carte d'échantillonnage pour chaque champ sont illustrées dans la figure 6. Tout d'abord, des analyses ont été effectuées sur la texture des sols suite à l'interpolation des données (*krigeage*) de la CE dans le but de définir quatre classes de types de sols. La redistribution des données dans chacune des classes a été possible en suivant les étapes de la figure 6. Le seuil minimal attribué par champ a été déterminé par le collaborateur XLKey en fonction du nombre de zones souhaitées par champ. Les différentes classes ont été attribuées telles que les valeurs élevées de CE correspondent à une texture très fine, tandis que les valeurs faibles de CE indiquent une texture très grossière, ce qui a permis de générer des cartes sur les zones selon la texture du sol (XIKey, 2023). Ensuite, une méthode d'échantillonnage aléatoire a été mise en œuvre pour prélever des échantillons dans chaque zone, en fonction de la texture du sol, en vue d'une analyse en laboratoire afin d'obtenir des informations sur la composition du sol dans ces environnements.

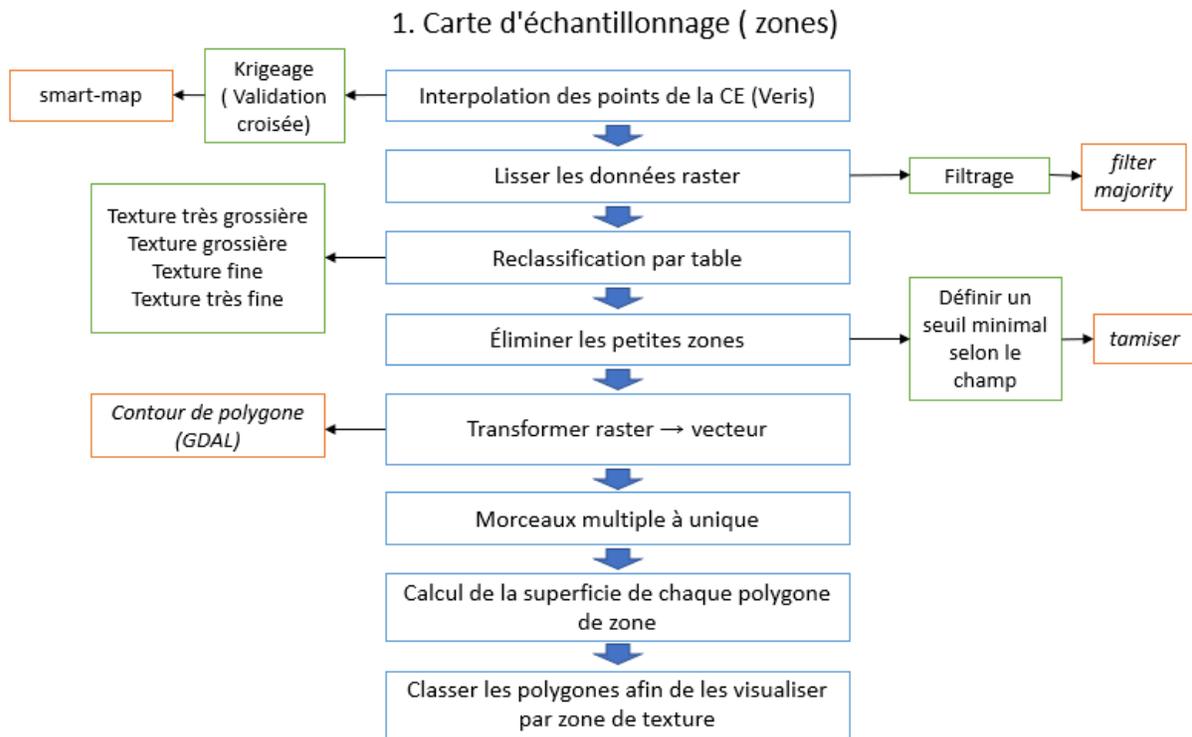


Figure 6. Organigramme des étapes réalisées afin d'obtenir les cartes d'échantillonnage de chacun des champs.

### 3.3.3 Carte d'échantillonnage de calibration CE-MO

Dans le but de calibrer des modèles ou des méthodes d'estimation de la MO en fonction de la CE du sol, une carte d'échantillonnage de calibration CE-MO a été réalisée. Pour permettre la réplication de ces analyses sur plusieurs champs, un code python a été développé. Il consiste à lire et extraire des données à partir de fichier Excel fournis par l'entreprise XLKey. Il effectue également des graphiques pour évaluer les relations : (1) entre la CE et le rendement de matière humide, (2) la réflectance dans les bandes du R et du PIR ainsi que (3) l'humidité du sol et la CE.

Par la suite, une classification non supervisée, de type « *clustering* » (*k-means*), est réalisée afin de produire trois classes à partir des variables de la CE et de la réflectance dans les bandes du R, qui permettent de décrire la teneur en MO. Pour cette analyse, la réflectance dans les bandes du PIR n'a pas été retenue, car la réflectance dans les bandes du R ont été jugées suffisantes pour décrire la teneur en MO. Cette méthode de classification, fréquemment utilisée en apprentissage automatique, a permis de regrouper des caractéristiques similaires de texture du sol. Cela a

conduit à l'élaboration d'une carte d'échantillonnage de calibration de la CE et de la MO, mettant en évidence trois zones distinctes : (1) à très faible teneur en MO et en CE, (2) à teneur moyenne en MO et en CE et (3) à très forte teneur en MO et en CE. Les trois zones précédemment nommées ont été déterminé par l'équipe du collaborateur XLKey.

Suite à la classification non supervisée, des graphiques de nuage de points ont été générés et affichés avec une référence géographique, utilisant les coordonnées (x, y) pour représenter la distribution spatiale de ces points sur une carte géographique

Cette représentation spatiale permet ensuite de délimiter ces zones de gestion pour prélever des échantillons de sol qui seront analysés en laboratoire, contribuant ainsi à la prise de décisions en matière de gestion des cultures. Ces données peuvent notamment être utilisées pour optimiser les apports d'engrais ou aider à la gestion de l'irrigation, entre autres (TechVidvan, 2024).

#### 3.3.4 Indice TPI (Indice de position topographique)

Le TPI est produit par une analyse algorithmique utilisant un système d'information géographique (SIG) pour l'analyse du relief afin d'évaluer les caractéristiques topographiques d'un paysage, en déterminant la position d'un point par rapport à ses voisins en fonction de l'altitude du terrain (Muddarisna *et al.*, 2020). Ainsi, il permet de quantifier la pente et la situation d'un point par rapport à son environnement.

$$TPI = Z_0 - Z$$

Où  $Z_0$  représente un point central qui exprime une position verticale en longueur au-dessous ou en dessous du point d'élévation comparatif (altitude du point d'intérêt).

$Z$  représente l'altitude moyenne des points environnants

Cet indice, basé sur le concept des plus proches voisins, offre des informations cruciales sur la présence de dépressions ou de crêtes dans le paysage, ainsi que sur la configuration du drainage. Ces données sont essentielles pour comprendre les variations de rendement observées dans une parcelle agricole, par exemple, elles peuvent révéler des zones sujettes à des problèmes d'accumulation d'eau ou de mauvaise circulation, pouvant entraîner des pertes de rendement.

Dans cette étude, la couche de TPI résultante a été réalisé sur QGIS avec l'outil *Index de Position Topographique (GDAL)*, où la couche LiDAR a été utilisée comme source de données d'entrée. Cette analyse *raster* a permis d'obtenir une représentation visuelle et quantitative des

caractéristiques topographiques du paysage, fournissant ainsi des informations précieuses pour la gestion agricole et la prise de décisions en matière d'aménagement du territoire.

### 3.3.5 Analyses géospatiales et corrélations des données agricoles

Un code de programmation Python a été développé dans le but d'analyser les données agricoles provenant de sources géospatiales telles que les rasters LiDAR et les données Veris, tout en prenant en compte les données de rendement. Le collaborateur XLKey a fourni les données de ces champs pour les analyses géospatiales subséquentes. Le choix des champs analysés a été déterminé selon la quantité et la qualité des données collectées par champ. En combinant ces données agricoles, le code Python a pour objectif de réaliser des analyses détaillées et à calculer les corrélations entre différentes variables clés. Le code sera ensuite utilisé pour l'ensemble des champs des clients du collaborateur XLKey lorsque les données seront collectées.

Tout d'abord, une routine est lancée afin de récupérer les données matricielles de chacune des variables analysées. Ensuite, chaque pixel est filtré pour éliminer les valeurs non valides, puis converti en points. Après avoir transformé les pixels en points pour les apparier géométriquement, une moyenne pondérée (par l'inverse de la distance) des 50 plus proches voisins est effectuée, ce qui permet de réduire le bruit des fluctuations observées dans ces données à 1 m de résolution. Les résultats sont ensuite sauvegardés en format de fichier Excel pour être utilisés dans le programme de modélisation du rendement à l'étape suivante, notamment comme données d'entrée pour le code de programmation du modèle de *Random Forest*. Le fichier résultant contient des données organisées en colonnes, où chaque colonne représente une variable explicative associée au rendement. Ainsi, pour chaque point de rendement, une valeur moyenne pondérée est calculée pour chaque variable explicatives soit sur les points LiDAR, les points de réflectance dans la bande rouge et du proche infrarouge, ainsi que les points de conductivité électrique. Le résultat de ces analyses ont permis de réaliser le modèle de prédiction du rendement.

### 3.3.6 Modélisation du rendement

Avec les fichiers Excel générés par le code précédent, un modèle d'apprentissage automatique (*Random Forest* de régression) a été développé pour réaliser des prédictions robustes et précises. L'analyse réalisée dans le code est une régression afin de prédire les niveaux de rendement de MS ou MH en fonctions des caractéristiques explicatives, soit les données matricielles

suivantes : LiDAR (élévation), RED (réflexion du rouge), NIR (réflexion du proche infra-rouge), ainsi que le EC\_DP (conductivité électrique profonde). La première représente la topographie de la parcelle, tandis que les trois autres variables sont liées à la texture du sol.

Dans ce code, une fonction est définie pour importer les fichiers Excel contenant les données ponctuelles issues des analyses géospatiales effectuées à l'étape précédente. Par la suite, le programme *Random Forest* de régression analyse ces données tel que décrit à la figure 7. Tout d'abord, on observe à l'étape 1, qu'il y a une sélection aléatoire (*bootstap*) qui permet ensuite de construire l'arbre de décision. À l'étape 2, il y a une agrégation des prédictions. De ce fait, chaque arbre fait une prédiction pour ensuite être agrégée et cela permet de donner un résultat final soit les données de la moyenne des prédictions (Belgiu et Drăguț, 2016).

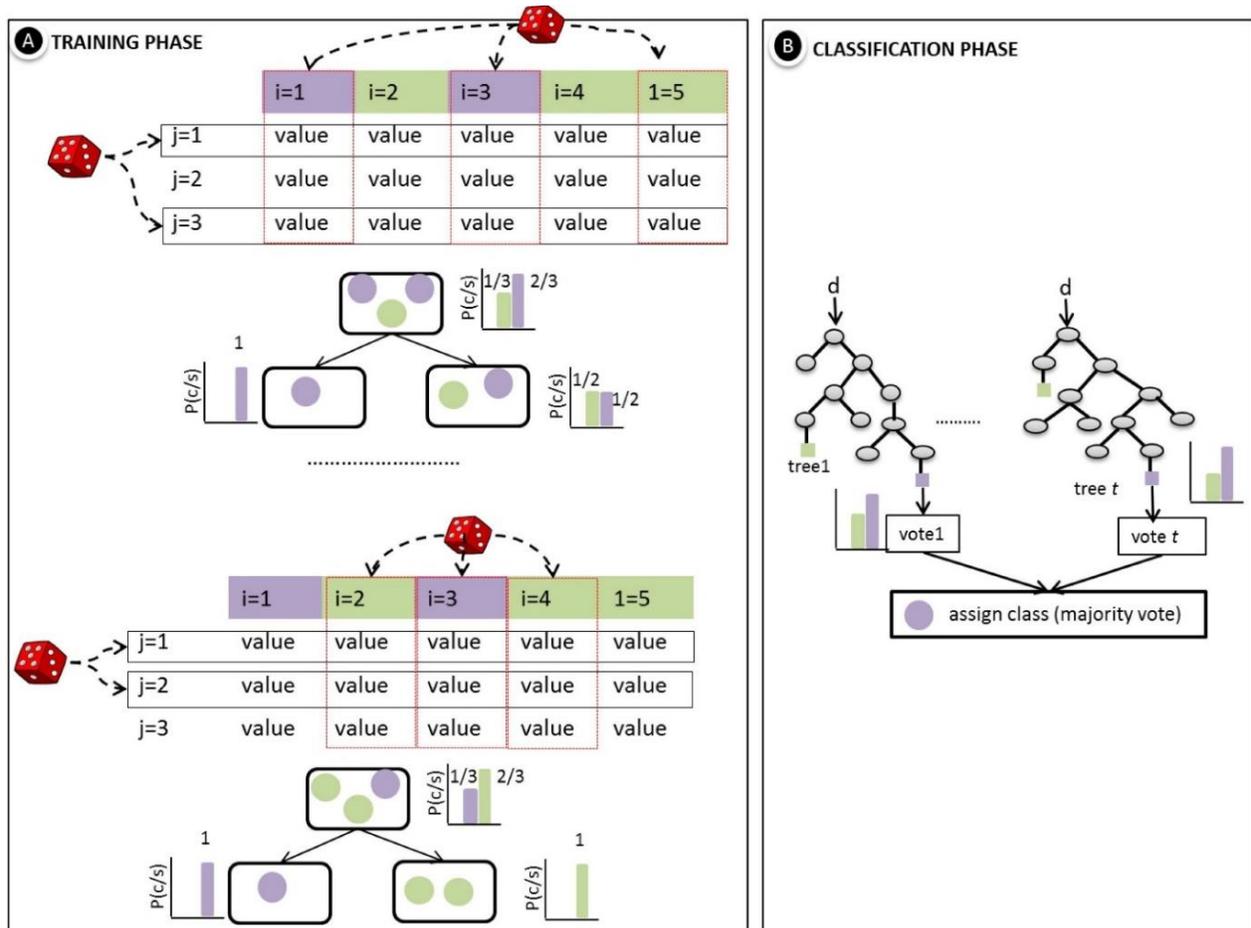


Figure 7. Organigramme des étapes effectués par le modèle d'apprentissage automatique (« Random Forest ») (Belgiu et Drăguț, 2016).

Ensuite, une validation croisée est réalisée, avec l'ensemble des données utilisées dans le *Random Forest*, pour évaluer la performance du modèle, comprenant le calcul des métriques de performance telles que le RMSE (Root Mean Square Error) et le  $R^2$  (coefficient de détermination). Finalement, suite à l'analyse du modèle, des graphiques sont générés pour illustrer l'importance des variables explicatives dans la parcelle étudiée par rapport aux rendements obtenus, dans le but de diagnostiquer l'une des causes possibles de la perte de rendement dans certaines zones d'un champ agricole.

## 4 Résultats

Plusieurs analyses ont été menées dans le cadre de ce projet dans le but de permettre aux agriculteurs client de XLKey de prendre des décisions plus précises et ciblées pour améliorer la productivité, réduire les coûts et minimiser les impacts environnementaux. Les traitements réalisés ont permis d'obtenir cinq résultats principaux. Tout d'abord, des cartes ont été créées afin d'échantillonner le sol selon la texture ou en fonction de la calibration de la CE en relation avec la Mo. Ensuite, une cartographie des reliefs topographiques a été réalisé sur quelques champs à l'étude. Finalement, des programmes en Python ont été exécutés afin d'obtenir une modélisation des prédictions de rendement selon les variables de texture du sol ou de la topographie du terrain.

Dans le présent rapport, seules quelques parcelles sont représentées dans les résultats, alors que les analyses ont été effectuées sur l'ensemble des données de ferme fournies par le collaborateur XLKey. Cette approche vise à améliorer la visualisation et à synthétiser les résultats sur une partie restreinte, essentielle pour les diagnostics, c'est-à-dire les analyses et évaluations effectuées pour identifier des problèmes ou des opportunités spécifiques dans la gestion des cultures. Les éléments décrits dans ces parcelles pourront ensuite être utilisé pour aider à l'analyse de l'ensemble de tous les parcelles présentes et futures.

### 4.1 Carte d'échantillonnage par zone

Une cartographie des champs Be01, Be02, Be03 et Be04 du site Belfast (fig. 8) a été élaborée dans le but de caractériser les textures du sol. La classification des textures du sol a été réalisée en se basant sur les données de la CE collectées par le capteur Veris MSP3. Les données de la

CE ont été réparties selon 4 classes de texture, allant de la texture très grossière, correspondant à des valeurs faibles de CE, à la texture très fine, associée à des valeurs élevées de CE. Cependant, en raison de la petite superficie des champs Be01 et Be02, seules trois classes ont été définies pour ces derniers, comme indiqué dans le tableau de la figure 8.

Cette analyse a permis de procéder par la suite à un échantillonnage aléatoire par zone de texture de sol afin d'identifier de manière précise les constituants du sol et leurs niveaux de concentration. Par conséquent, un échantillon sera prélevé dans chaque zone pour être analysé en laboratoire afin d'obtenir une représentation complète et fiable de la constitution du sol de chaque zone.

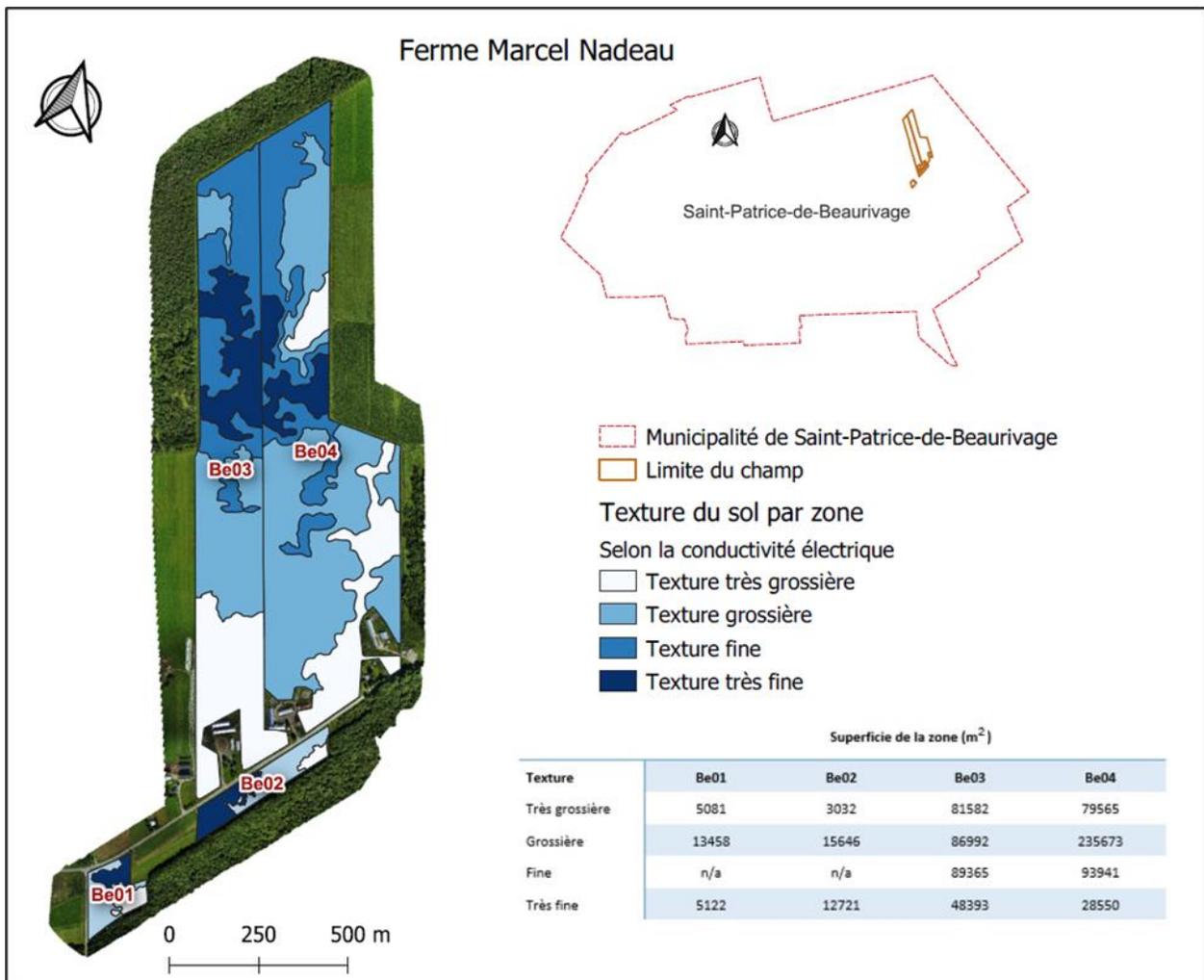


Figure 8. Carte d'échantillonnage selon la texture du sol des champs Be01, Be02, Be03 et Be04.

## 4.2 Carte d'échantillonnage de calibration CE-MO

Les résultats obtenus dans cette section permettent de visualiser les corrélations entre différentes données agricoles. Des données compilées remises par le collaborateur XLKey ont été utilisées dans cette analyse. En tant que livrable, plusieurs graphiques sont produits, comprenant (1) la relation entre les données de rendement et la CE (fig.9.a), (2) la réflectance dans le spectre du R et du PIR (fig.9.b), ainsi que (3) l'humidité du sol et la CE (fig.9.c). En superposant les données sur le rendement et la CE pour observer les relations, il a été possible également d'effectuer un *clustering (k-means)*, ce qui a donné le graphique de corrélation représenté à la figure 9.d, soit le graphique de la classification par la méthode du *k-means* afin de déterminer trois classes de regroupement. Ceci permettra, par la suite, de cartographier ces éléments en coordonnées géographiques afin d'extraire des points d'échantillonnage (fig.10). Dans le graphique 9a, qui correspond à un graphique présentant la relation du rendement (MS) en fonction de la CE, on constate une dispersion spatiale importante ( $r=0,16$ ). Le graphique illustrant la relation entre la réflectance dans la bande spectrale du R et du PIR démontre une corrélation forte entre ces deux variables ( $r=0,77$ ) (fig.9b). En revanche, le graphique c montre une corrélation faible ( $r=0,12$ ) entre l'humidité du sol et la CE (fig.9c). Enfin, le graphique de la figure 9d classe les points en trois groupes, soit un regroupement de points qui correspondent à une très forte teneur en MO et en CE, un regroupement de points ayant une faible teneur en MO et en CE et le dernier groupe correspond à des valeurs de point ayant une teneur moyenne en MO et en CE.

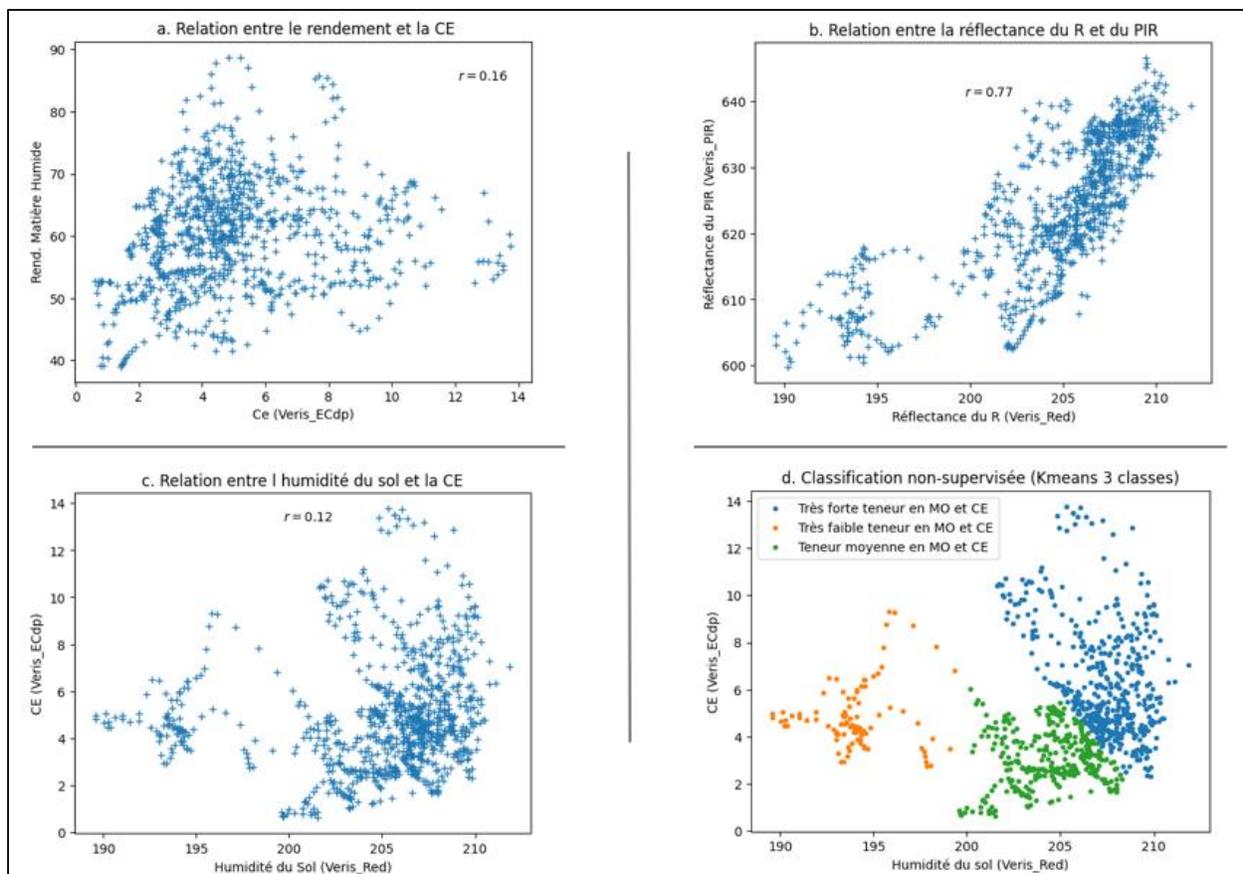


Figure 9. Graphiques sur les relations entre les variables explicatives analysées dans ce projet sur des données compilées remises par l'entreprise XLKey a) Relation du rendement (MS) en fonction de la CE avec un coefficient de détermination de 0,16 ; b) Corrélation entre la réflectance du R en fonction de la PIR avec un  $r = 0,77$  ; c) Relation entre la CE et la MO ( $r = 0,12$ ) ; d) Graphique de la classification par la méthode du « k-means » afin de déterminer 3 classes de regroupement.

Le graphique de la figure 10 permet de visualiser la distribution des points, suite à la classification non supervisée (*k-means*) selon leurs coordonnées spatiales (x,y). Cette classification a permis d'obtenir trois groupes distincts : ceux caractérisés par une forte teneur en MO et CE, ceux présentant une teneur moyenne en MO et CE, et enfin ceux affichant une faible teneur en MO et CE. Pour chacun de ces groupes, un point central a été généré et ses coordonnées géospatiales ont été enregistrées dans le but de réaliser un échantillonnage précisément à cet endroit pour analyser les éléments du sol en laboratoire afin de représenter chacun des trois groupes de type de sol.

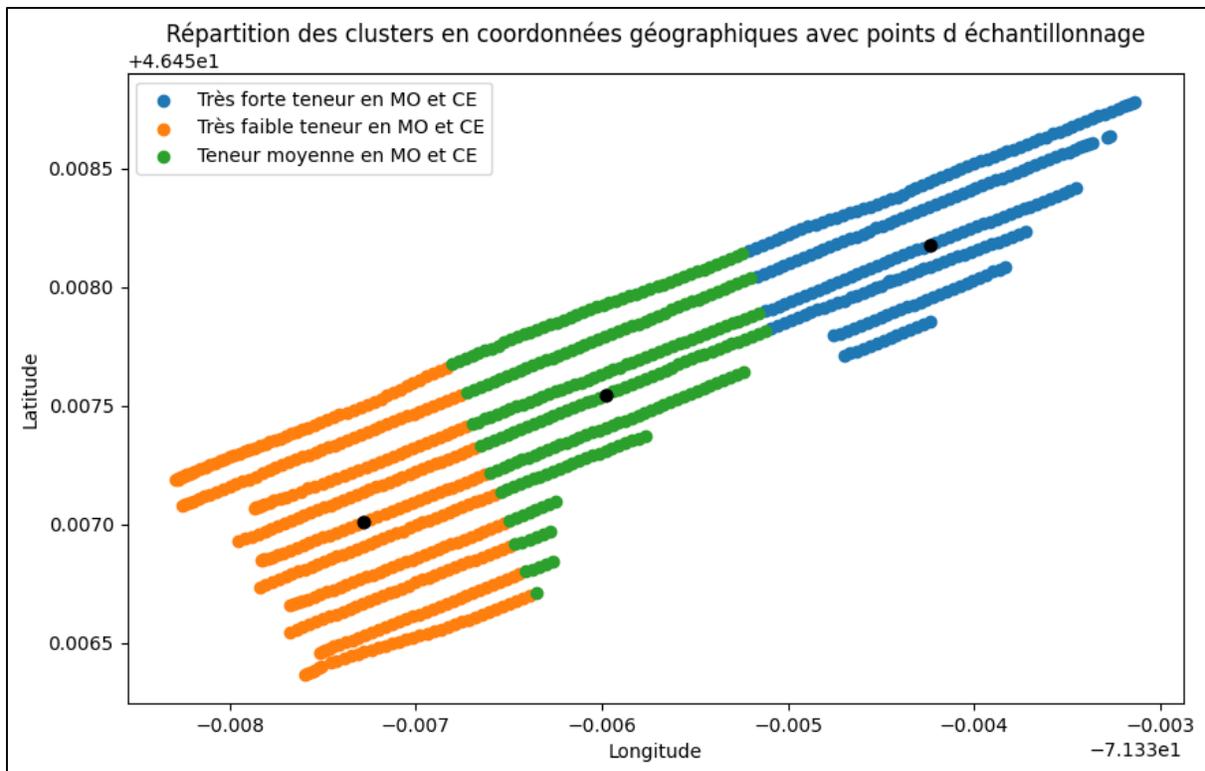


Figure 10. Identification des 3 points d'échantillonnage les plus représentatif du champ selon la répartition des 3 classes réalisées lors de la classification non supervisée (K-means) : classe de la très forte teneur en MO et CE (-71.3342, 46.4582), classe de la très faible teneur en MO et CE (-71.3373, 46.4570) et la classe de la teneur moyenne en MO et CE (-71.3360, 46.4575).

### 4.3 Carte illustrant le TPI

Une cartographie des valeurs de TPI des champs BS02 et B04 a été réalisé afin d'identifier plusieurs zones caractérisées par des dépressions ou des pentes (fig.11). Les valeurs de TPI positives (pixels verts), représentent les emplacements au-dessus de la moyenne de leur environnement, tandis que les valeurs de TPI négatives (pixels rouges) permettent de visualiser les zones plus basses que leur environnement. Les valeurs de TPI proches de zéro (pixels jaunes) représentent des zones planes ou des zones de pente constante (où la pente est presque nulle) (Marchese *et al.*, 2020).

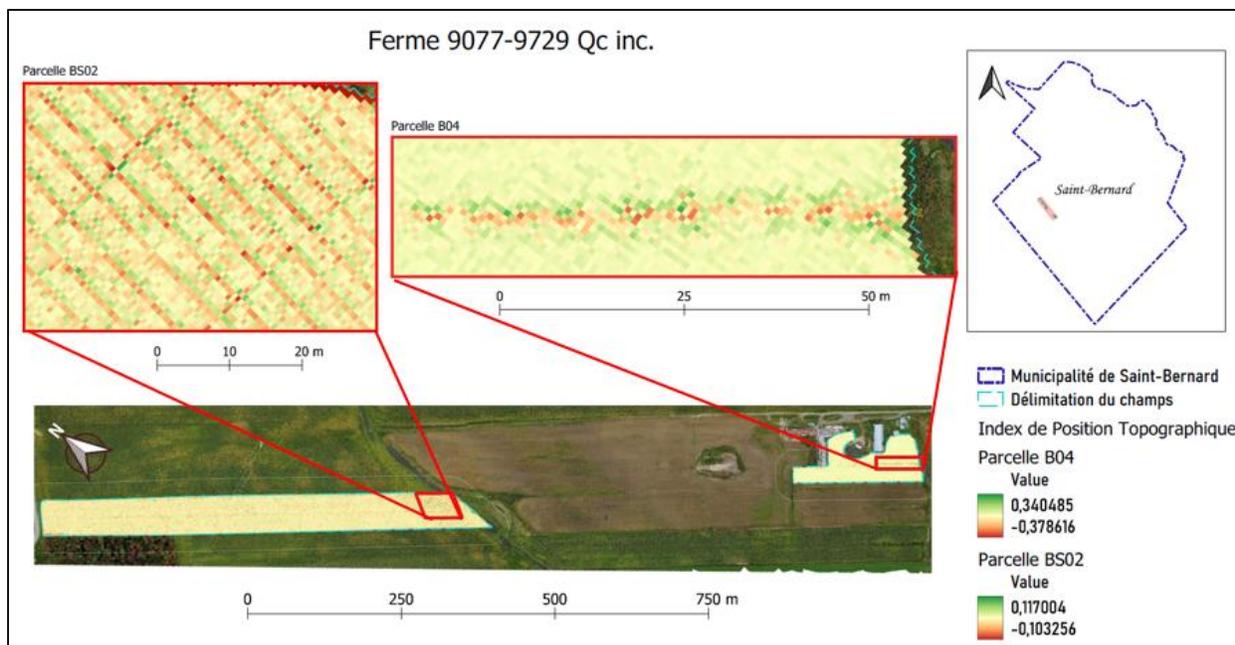


Figure 11. Carte illustrant les valeurs de TPI des champs BS02 et B04 de la ferme 9077-9729 Qc inc.

Avec les données du TPI de la parcelle BS02 (fig.11), on observe qu'une zone à l'extrémité du champs, situé au sud-est, présente plusieurs alternances de pixels rouges et verts. Afin d'analyser le relief de cette zone spécifique, la carte de la figure 12 a été réalisée. Cette carte présente des données topographiques ainsi qu'une orthophotographie permettant de visualiser un cours d'eau à proximité des valeurs de TPI fluctuantes. De plus, l'image LiDAR révèle une dénivellation significative dans cette région.

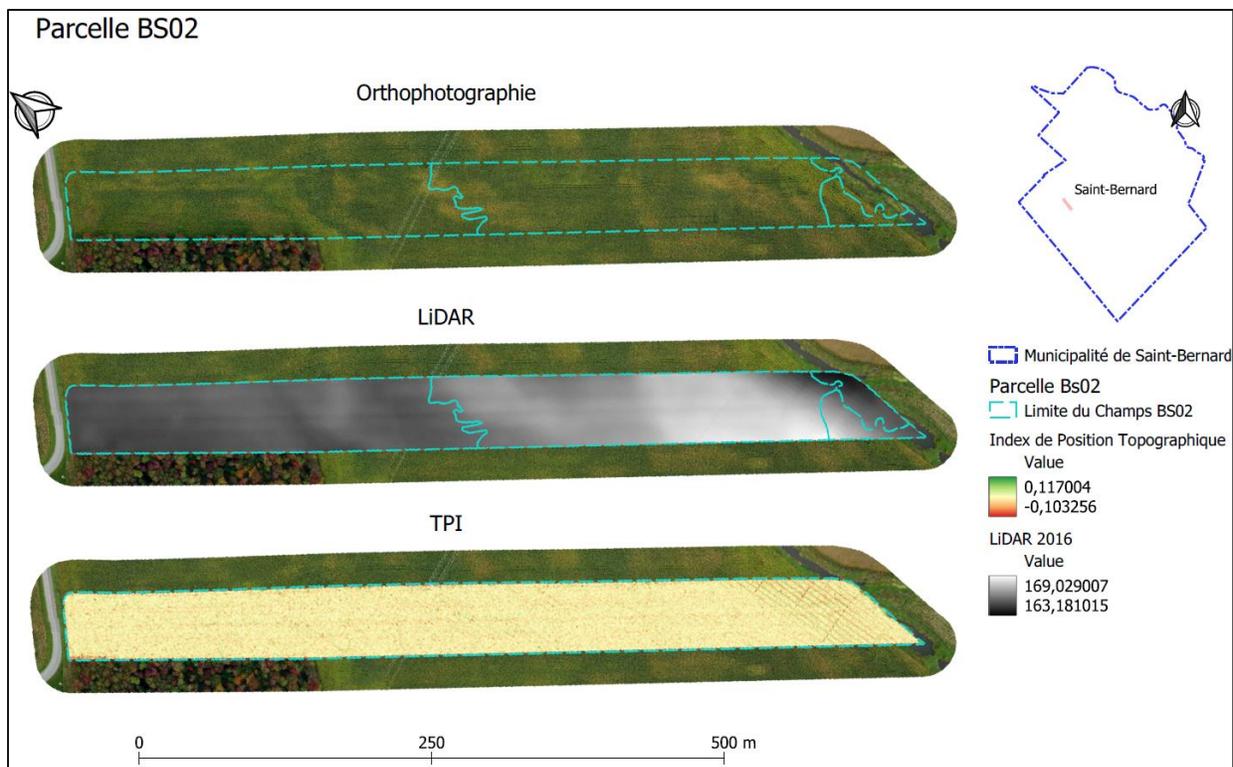


Figure 12. Cartes illustrant les valeurs de TPI et de la topographie du champ BS02 de la ferme 9077-9729 Qc inc. L'échelle de couleur appliquée à la carte indique que le vert correspond à des valeurs de hautes altitudes et le rouge à des valeurs basse altitudes

Les valeurs de TPI observées pour le champ B04, telles qu'illustrées à la figure 11, révèlent une zone linéaire de fluctuation des valeurs entre les pixels rouges et verts. La carte présentée à la figure 13 montre que cette zone linéaire correspond à une barrière physique, ce qui pourrait peut-être correspondre à un fossé, étant donné les valeurs plus basses du LiDAR à cet endroit.

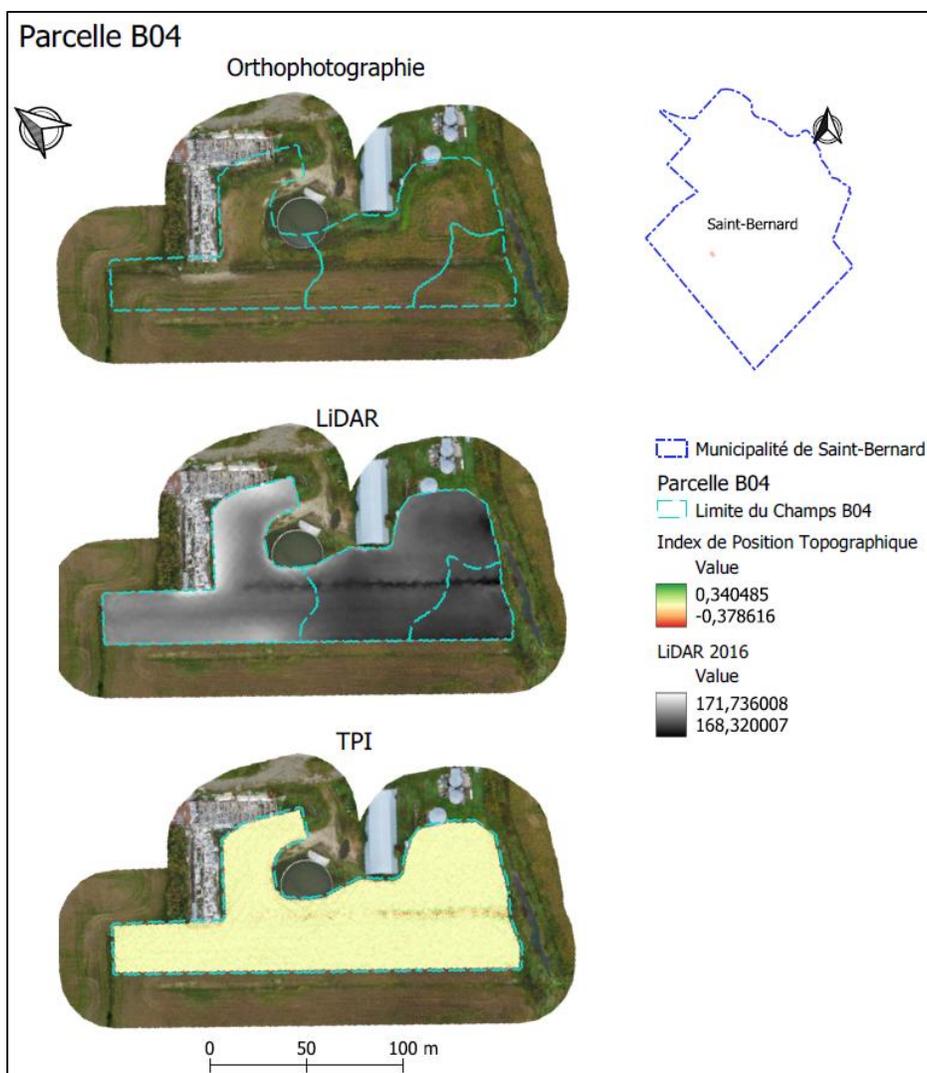


Figure 13. Cartes illustrant les valeurs de TPI et de la topographie du champ B04 de la ferme 9077-9729 Qc inc.

#### 4.4 Modélisation du rendement

La mise en œuvre du modèle d'apprentissage automatique (*Random Forest*) a permis de générer des prédictions sur les variables pouvant expliquer les écarts de rendement dans un champ. L'ensemble des résultats obtenus est présenté dans le tableau 3 pour les champs qui ont été inclus dans le code Python. Il est à noter que certaines valeurs sont manquantes dans le tableau des résultats pour la ferme « Forfait André et Frédéric Roy », car il n'y avait pas de données pour les rendements de MS dans la BD de l'entreprise XLKey.

Les résultats ont tout d'abord été exportés vers un fichier texte afin d'observer les valeurs du coefficient de détermination pour le rendement de la MH et de la MS (fig. 14). De plus, des

graphiques sur les importances relatives des caractéristiques explicatives ont été exportés au format d'image (figs. 15 et 16), permettant ainsi de visualiser les variables potentiellement significatives pour expliquer les rendements dans certaines zones du champ.

```

Le nom de la parcelle analysée (SOYA): Ferme 9077-9729 Qc inc_Saint-Bernard_B04_yield_2022_SOYBEANS.xlsx

Résultats pour Wet :
RMSE avec Wet: 3794.784357565726
R2 avec Wet: 0.9679006573010537
RMSE avec Wet (Validation croisée): 3937.3359488558185
R2 avec Wet (Validation croisée): 0.9625129733683563

Résultats pour Moisture :
RMSE avec Moisture: 0.34258239394734596
R2 avec Moisture: 0.7601642134725745
RMSE avec Moisture (Validation croisée): 0.3372383964254091
R2 avec Moisture (Validation croisée): 0.7877017857895978
    
```

Figure 14. Aperçu d'un fichier texte obtenu suite à l'exécution du code Random Forest. L'exemple correspond au résultat du champ B04 de la ferme 9077-9729 Qc inc.

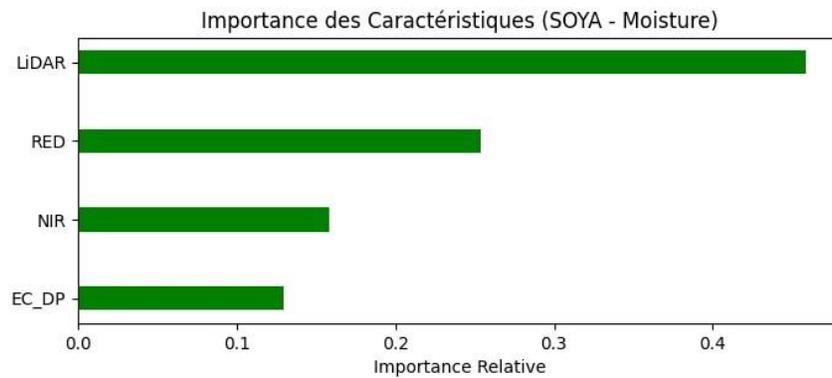


Figure 15. Graphique de l'importance des caractéristiques explicatives pour le rendement de matière humide (Moisture) pour la parcelle B04 de la ferme 9077-9729 situé à Saint-Bernard.

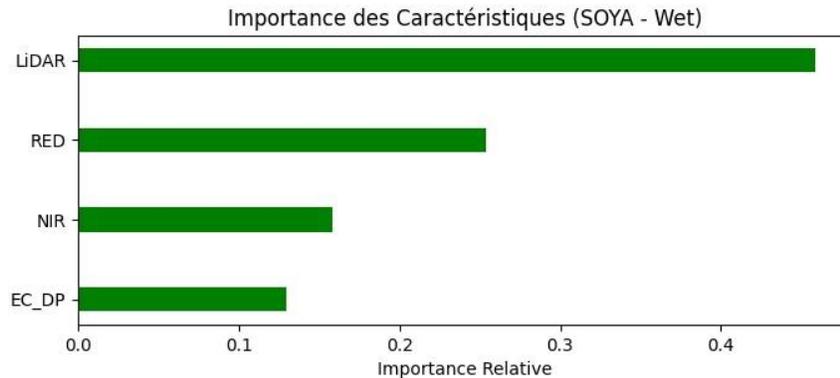


Figure 16. Graphique de l'importance des caractéristiques explicatives pour le rendement de matière sèche (Wet) pour la parcelle B04 de la ferme 9077-9729 situé à Saint-Bernard.

Tableau 3. Sommaire des résultats obtenus lors de la modélisation par « Random Forest » (Régression).

|                                      | Type | Nom de la Parcelle | Résultats         |                   |                           | Année image .tif |       |       |           |      |      |      |
|--------------------------------------|------|--------------------|-------------------|-------------------|---------------------------|------------------|-------|-------|-----------|------|------|------|
|                                      |      |                    | R <sup>2</sup> MS | R <sup>2</sup> MH | Caractéristique incidente | Yield            | Veris | LiDAR |           |      |      |      |
| <b>Forfait André et Frédéric Roy</b> | Soya | 01                 | -                 | 0,25              | Veris_RED                 | 2022             | 2022  | 2018  |           |      |      |      |
|                                      |      | 02                 | -                 | 0,37              | Veris_RED                 |                  |       |       |           |      |      |      |
|                                      | Maïs | 05                 | 0,54              | 0,54              | LiDAR                     | 2021             | 2022  | 2018  |           |      |      |      |
|                                      |      | FR01               | 0,63              | 0,63              | Veris_EC_DP               |                  |       |       |           |      |      |      |
|                                      |      | <b>FR02</b>        | -                 | <b>0,59</b>       | Veris_NIR                 |                  |       |       |           |      |      |      |
| <b>Ferme Marcel Nadeau</b>           | Soya | AGA01              | 0,15              | 0,23              | Veris_EC_DP               | 2022             | 2022  | 2012  |           |      |      |      |
|                                      |      | LAB01              | 0,71              | 0,72              | LiDAR                     |                  |       |       |           |      |      |      |
|                                      |      | LAB02              | 0,50              | 0,45              | LiDAR                     |                  |       |       |           |      |      |      |
|                                      |      | <b>GIL01</b>       | <b>0,59</b>       | <b>0,52</b>       | Veris_EC_DP               |                  |       |       |           |      |      |      |
|                                      | Maïs | MEU01              | 0,15              | 0,17              | LiDAR                     | -                | 2022  | 2016  |           |      |      |      |
| <b>Ferme 9077-9729 Qc inc</b>        | Maïs | B01                | 0,42              | 0,30              | Veris_EC_DP               | 2022             | 2022  |       |           |      |      |      |
|                                      |      | BS01               | 0,71              | 0,60              | LiDAR                     |                  |       |       |           |      |      |      |
|                                      |      | <b>BS02</b>        | <b>0,67</b>       | <b>0,74</b>       | LiDAR                     |                  |       |       |           |      |      |      |
|                                      |      | BS03               | 0,61              | 0,49              | LiDAR                     |                  |       |       |           |      |      |      |
|                                      | Maïs | A05                | 0,49              | 0,49              | LiDAR                     | 2021             | 2021  | 2016  |           |      |      |      |
|                                      |      | A06                | 0,41              | 0,46              | LiDAR                     |                  |       |       |           |      |      |      |
|                                      |      | A07                | 0,23              | 0,27              | LiDAR                     |                  |       |       |           |      |      |      |
|                                      |      | A01                | 0,35              | 0,19              | LiDAR                     |                  |       |       |           |      |      |      |
|                                      |      | A09                | 0,69              | 0,47              | LiDAR                     |                  |       |       |           |      |      |      |
|                                      |      | A12                | 0,53              | 0,38              | Veris_NIR                 |                  |       |       |           |      |      |      |
|                                      |      | A13                | 0,41              | 0,48              | Veris_NIR                 |                  |       |       |           |      |      |      |
|                                      |      | A14                | 0,50              | 0,34              | Veris_EC_DP               |                  |       |       |           |      |      |      |
|                                      |      | A15                | 0,62              | 0,41              | Veris_EC_DP               |                  |       |       |           |      |      |      |
|                                      |      | Soya               | <b>N01</b>        | <b>0,74</b>       | <b>0,73</b>               |                  |       |       | Veris_RED | 2022 | 2022 | 2016 |
|                                      |      |                    | A02               | 0,58              | 0,44                      |                  |       |       | Veris_RED |      |      |      |
|                                      |      |                    | A03-A04           | 0,48              | 0,53                      |                  |       |       | Veris_RED |      |      |      |
|                                      |      |                    | B02               | 0,36              | 0,47                      |                  |       |       | LiDAR     |      |      |      |
|                                      |      |                    | <b>B03</b>        | <b>0,94</b>       | <b>0,55</b>               |                  |       |       | LiDAR     |      |      |      |
|                                      |      |                    | <b>B04</b>        | <b>0,97</b>       | <b>0,76</b>               |                  |       |       | LiDAR     |      |      |      |
|                                      |      |                    | B05               | 0,41              | 0,39                      |                  |       |       | LiDAR     |      |      |      |
|                                      | B06  |                    | 0,75              | 0,80              | LiDAR                     |                  |       |       |           |      |      |      |
|                                      | B07  |                    | 0,46              | 0,43              | Veris_RED                 |                  |       |       |           |      |      |      |
|                                      | B08  |                    | 0,62              | 0,60              | Veris_RED                 |                  |       |       |           |      |      |      |
|                                      | B09  | 0,42               | 0,72              | LiDAR             |                           |                  |       |       |           |      |      |      |
|                                      | BS04 | 0,50               | 0,75              | LiDAR             |                           |                  |       |       |           |      |      |      |
| BS05                                 | 0,50 | 0,75               | LiDAR             |                   |                           |                  |       |       |           |      |      |      |

Après avoir examiné ces résultats, six parcelles ont été sélectionnées en fonction des coefficients de détermination et de la caractéristique pertinente obtenue. Cette sélection permettra de mener un diagnostic sur les variables influençant les écarts de rendement, dans le but d'optimiser et d'améliorer les performances agricoles. Les parcelles suivantes ont ainsi été choisies : la parcelle FR02 de la ferme « Forfait André et Frédéric Roy », la parcelle GIL01 de la ferme « Marcel Nadeau », ainsi que les champs BS02, B03, B04 et N01 de la ferme « 9077-9729 Qc inc ». Dans la suite de ce rapport, seule l'analyse de ces champs sera réalisée, en se basant sur la caractéristique pertinente obtenue lors de la modélisation, soit le LiDAR, la MO ou la CE.

#### 4.4.1 Diagnostic expliqué par les données LiDAR

Selon les résultats de la modélisation, les parcelles B03 et B04, cultivées en soya, ainsi que la parcelle BS02, cultivée en maïs, révèlent une caractéristique déterminante pour le rendement : la topographie du terrain, représentée par la variable LiDAR. Ce sont trois parcelles cultivées par la ferme « 9077-9729 Qc inc. », située dans la municipalité de Saint-Bernard. Ainsi, les cartes des figures 17 à 19 illustrent les divers éléments analysés afin d'expliquer les zones de perte de rendement en fonction de la topographie.

Pour le champ B03 (tableau 3), la topographie a été le critère qui a influencé le rendement. On note un coefficient de détermination ( $R^2$ ) de 0,94 pour le rendement de MS et de 0,55 pour le rendement de MH. Ces résultats suggèrent que la topographie a une incidence significative sur la quantité de MS récoltée, avec une corrélation très forte, tandis que son impact sur le rendement de MH semble être moins marqué, avec une corrélation moyenne.

Dans la figure 17, qui représente la parcelle B03, on peut observer des similarités entre les rendements de MH et MS dans les zones situées à l'extrémité sud-sud-est du champ (fig.17a et 17b). C'est d'ailleurs la zone où l'on observe un dénivelé plus bas, soit une zone de dépression (fig. 17c). Dans la zone nord-ouest, à l'autre extrémité du champ, on observe également une zone de faible rendement (fig. 17a et 17b), tandis que dans la figure 17c, on peut observer que cette zone comporte une pente.

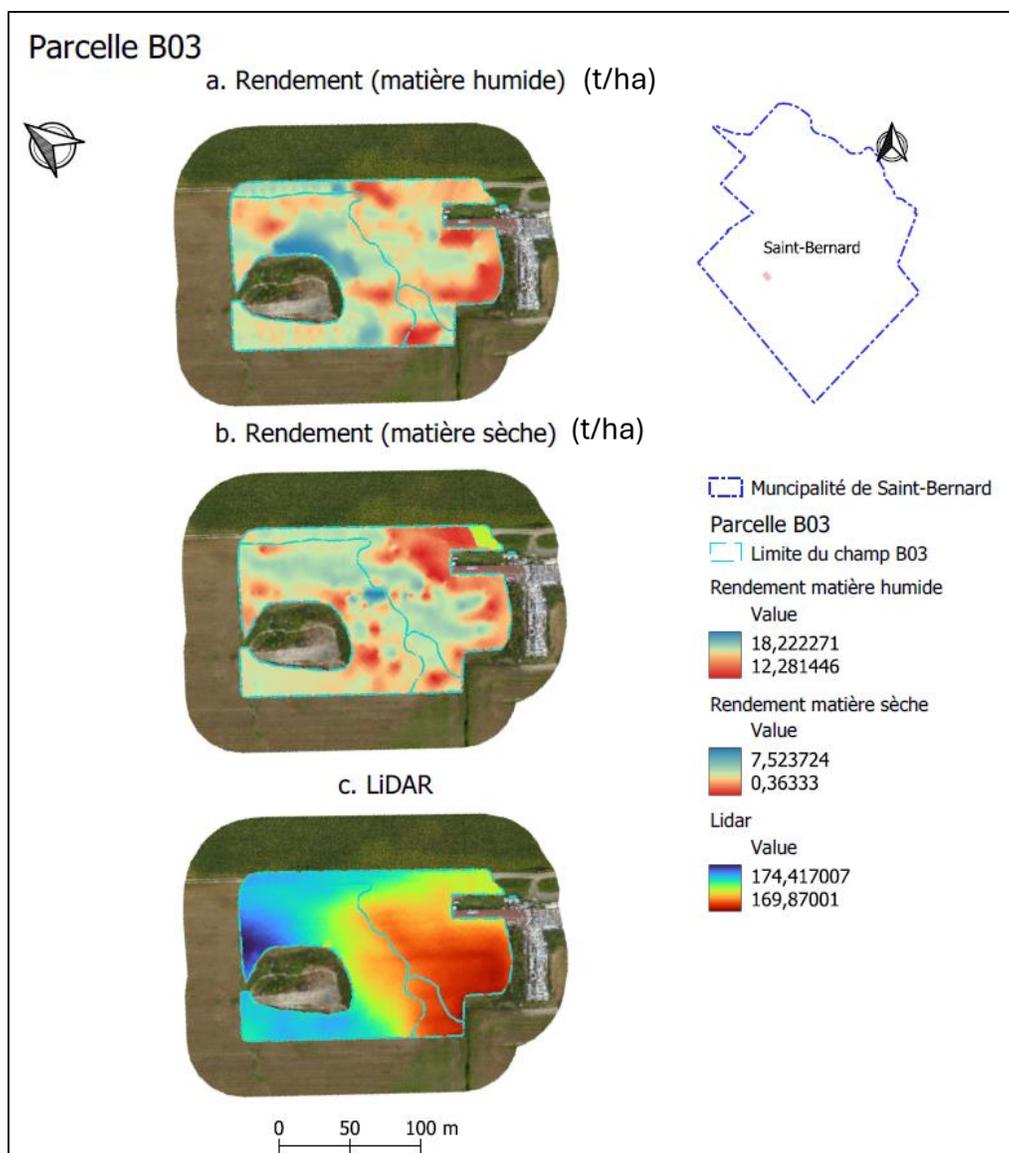


Figure 17. Carte des analyses de la parcelle B03 de la ferme 9077-9729 Qc inc. située à Saint-Bernard de la MS et MH en considérant les données LIDAR (topographie).

Les résultats de la modélisation de la parcelle B04 mettent en évidence l'influence des données topographiques sur le rendement (tableau 3). Le coefficient de détermination ( $R^2$ ) obtenu pour le rendement en matière sèche est de 0,97, ce qui démontre une corrélation forte. De même, le coefficient de détermination pour le rendement en matière humide atteint une valeur notable de 0,76. Ces observations suggèrent que la topographie joue un rôle prépondérant dans la quantité de MS récoltée, tandis que son impact sur le rendement de MS est également notable mais légèrement moindre.

Dans la parcelle B04 (fig.18), la carte du rendement de MH (fig. 18a) démontre qu'il y a une perte de rendement linéaire qui se situe, sensiblement, au même endroit que les pixels rouges dans la figure 18c représentant des points de plus basse altitude. Également, on peut observer que dans la zone située au nord-nord-ouest, là où l'on retrouve les pixels de plus hautes altitudes (fig. 18c), il y a un rendement moyen plus faible (fig. 18a).

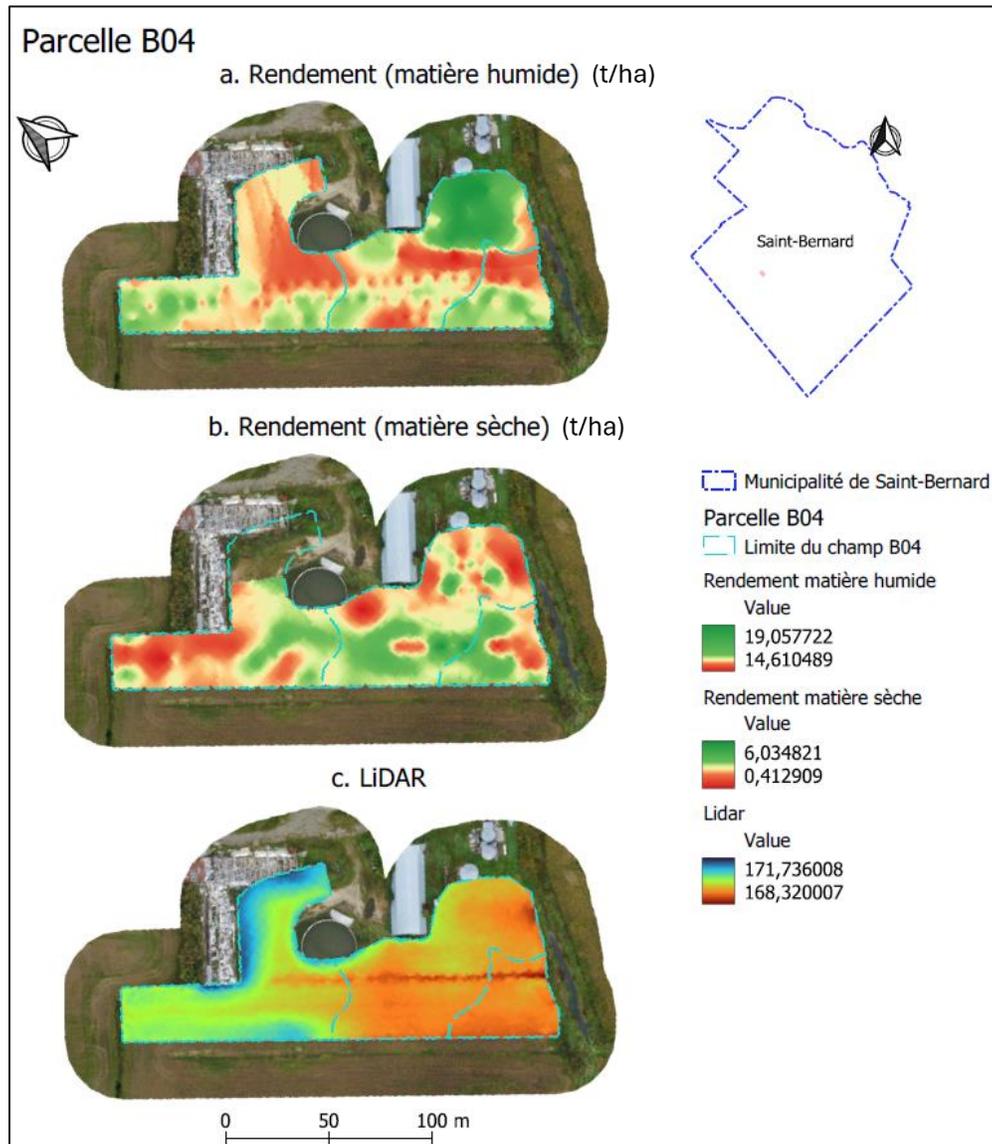


Figure 18. Carte des analyses de la parcelle B04 de la ferme 9077-9729 Qc inc. située à Saint-Bernard de la MS et MH en considérant les données LIDAR (topographie).

Les résultats de la modélisation pour le champ BS02 (tableau 3) ont mis en évidence que les données LiDAR étaient la caractéristique prédominante influençant le rendement, avec des valeurs de coefficient de détermination ( $R^2$ ) de 0,74 pour le rendement en matière humide et de

0,67 pour le rendement en matière sèche. Ces observations soulignent l'importance de la topographie dans la prédiction du rendement, avec une corrélation forte tant pour le rendement en matière humide que pour le rendement en matière sèche.

En observant la figure 19a de la parcelle BS02, on remarque une concentration de perte de rendement près de l'extrémité sud-sud-est du champ. Cette région correspond également à une zone distincte dans la figure 19c où les valeurs se trouvent aux extrêmes des données d'altitude élevée qui peuvent être caractérisées par une pente dans cette zone. Par ailleurs, la zone ayant enregistré le plus grand rendement en matière humide correspond aux valeurs centrales (moyennes) des données topographiques. Cette observation suggère que le relief peut jouer un rôle crucial dans la distribution spatiale du rendement, avec d'importantes variations en fonction de l'altitude et de la pente du terrain.

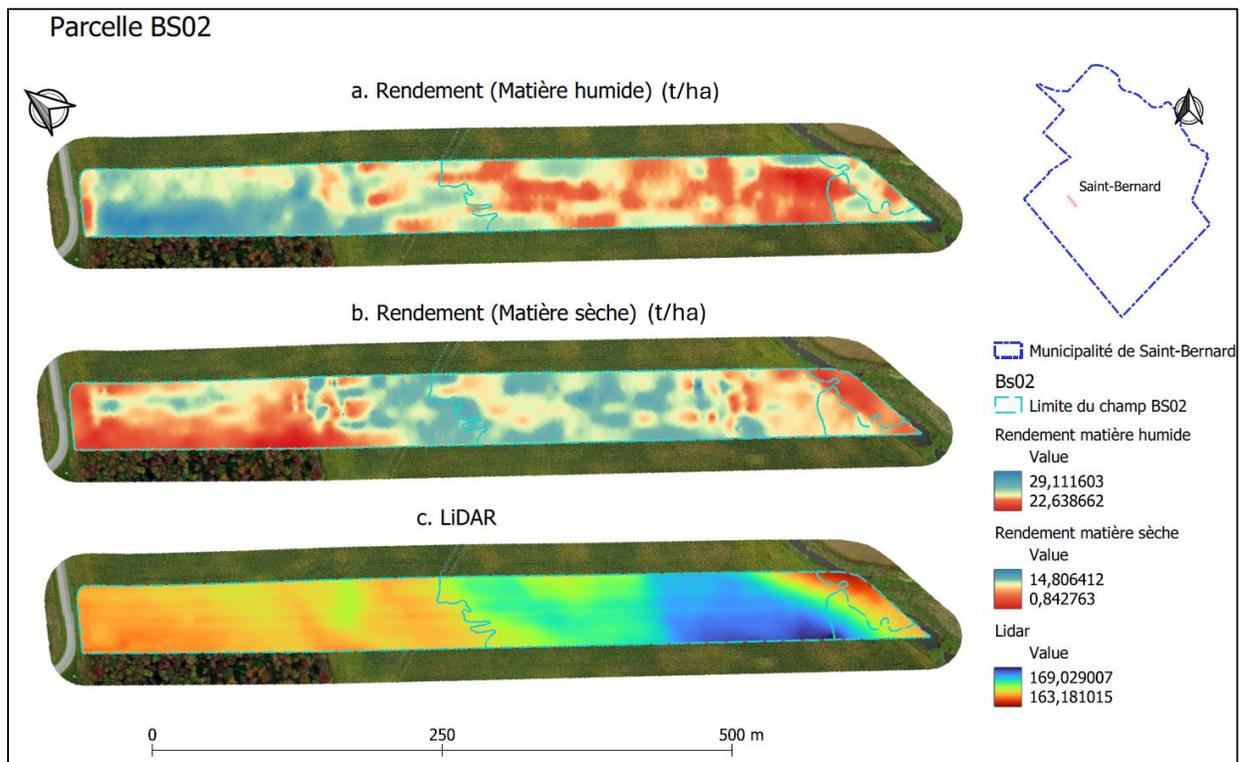


Figure 19. Carte des analyses de la parcelle BS02 de la ferme 9077-9729 Qc inc. située à Saint-Bernard de la MS et MH en considérant les données LIDAR (topographie).

#### 4.4.2 Variations du rendement reliées à la matière organique

Selon les résultats de la modélisation observés dans le tableau 3, la parcelle N01, issue de la ferme « 9077-9729 Qc inc. », située à Saint-Narcisse-de-Beaurivage, ainsi que la parcelle FR02

de la ferme « Forfait André et Frédéric Roy », localisée à Honfleur, mettent en évidence une caractéristique qui influence le rendement : la teneur en matière organique, déterminée par les variables de réflectance du R ou du PIR. La parcelle N01 est cultivée en soya, tandis que la parcelle FR02 est cultivée en maïs.

Dans la parcelle N01 (fig. 20), la variable de réflectance de la bande du rouge s'est avérée être la caractéristique ayant eu l'influence la plus marquée sur le rendement. Le coefficient de détermination ( $R^2$ ) obtenu lors de la modélisation est de 0,74 pour le rendement en MS et de 0,73 pour le rendement en MH, ce qui indique une corrélation forte entre la teneur en matière organique et le rendement.

De plus, on remarque qu'il y a une hétérogénéité dans la texture de sol, ou de la teneur en MO, sur l'ensemble du champ (fig. 20c). Par exemple, la carte 20c permet de visualiser une zone au Sud-sud-ouest du champ où la réflectance de la bande spectrale rouge est légèrement plus élevée (zone en bleu foncé). Cette zone correspond également à une zone de perte de rendement observée à la figure 20a. Également, c'est majoritairement dans la zone où la réflectance de la bande du R est plutôt neutre que l'on retrouve les plus hautes valeurs de rendement. Cette observation suggère une relation complexe entre la composition du sol, la réflectance spectrale et le rendement, soulignant l'importance de considérer ces facteurs dans l'analyse et la gestion agronomique de la parcelle.

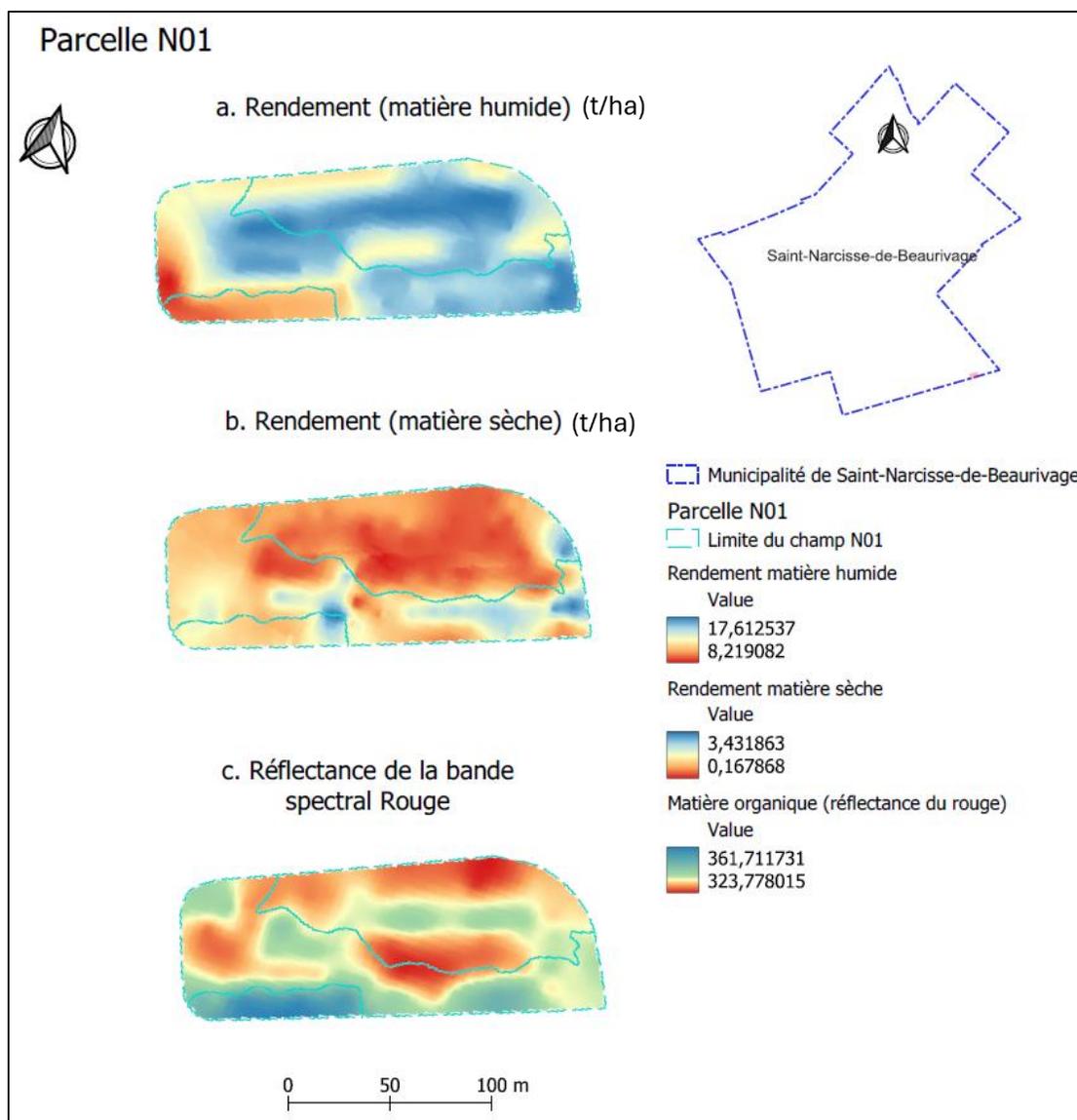


Figure 20. Carte des analyses de la parcelle N01 de la ferme 9077-9729 Qc inc situé à Saint-Narcisse-de-Beaurivage de la MS et MH en considérant les données de la réflectance du R.

Les résultats de la modélisation pour la parcelle FR02 (tableau 3) ont révélé que les données sur la teneur en matière organique, obtenue avec la réflectance de la bande du PIR étaient la caractéristique prédominante influençant le rendement. Le coefficient de détermination ( $R^2$ ) pour le rendement de matière humide correspond à une valeur de 0,59. Ce résultat souligne l'importance de la teneur en matière organique dans la prédiction du rendement, avec une corrélation faible. Cependant, il est à noter que les données sur le rendement de matière sèche étaient manquantes pour cette parcelle, ce qui peut influencer la prédiction du modèle de

*Random Forest* étant donné que le modèle prend en considération le nombre de variables explicatives analysées alors que pour ce champ, des données sur l'une des variables n'étaient pas comptabilisées.

En observant la figure 21a, on remarque un patron de variation linéaire pour les données sur le rendement. Tandis que les données de réflectance dans la bande spectrale du PIR (fig. 21b) comportent plusieurs zones circulaires avec de faibles ou de fortes valeurs, ce qui ne s'apparient pas avec les zones de perte de rendement observées.

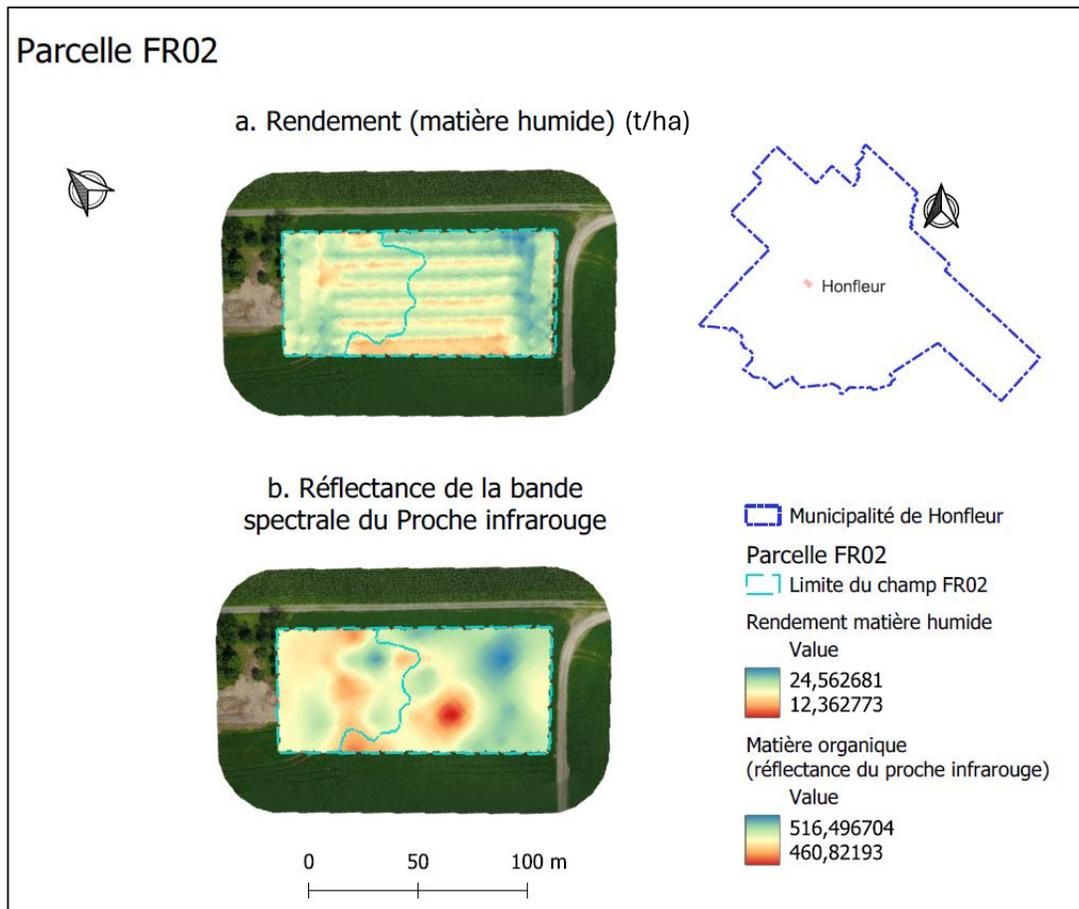


Figure 21. Carte des analyses de la parcelle FR02 de la ferme Forfait André et Frédéric Roy située à Honfleur de la MS et MH en considérant les données de la réflectance du R.

#### 4.4.3 Variation du rendement expliquée par la conductivité électrique

Dans la parcelle GIL01 (fig. 22), la CE s'est avérée être le principal facteur influençant le rendement (tableau 3). Le coefficient de détermination ( $R^2$ ) obtenu lors de la modélisation est de

0,59 pour le rendement en matière sèche et de 0,52 pour le rendement en matière humide, ce qui indique une corrélation importante entre la CE du sol et les rendements agricoles obtenus.

Dans la figure 22, on remarque des zones où des pertes de rendement de matière sèche sont observées, soit les pixels rouges de la figure 22b présentent également des valeurs plus élevées de CE. Cependant, dans l'ensemble, des données sont manquantes pour la zone au nord-ouest du champ, ce qui constitue une lacune significative susceptible d'influencer le modèle de prédiction.

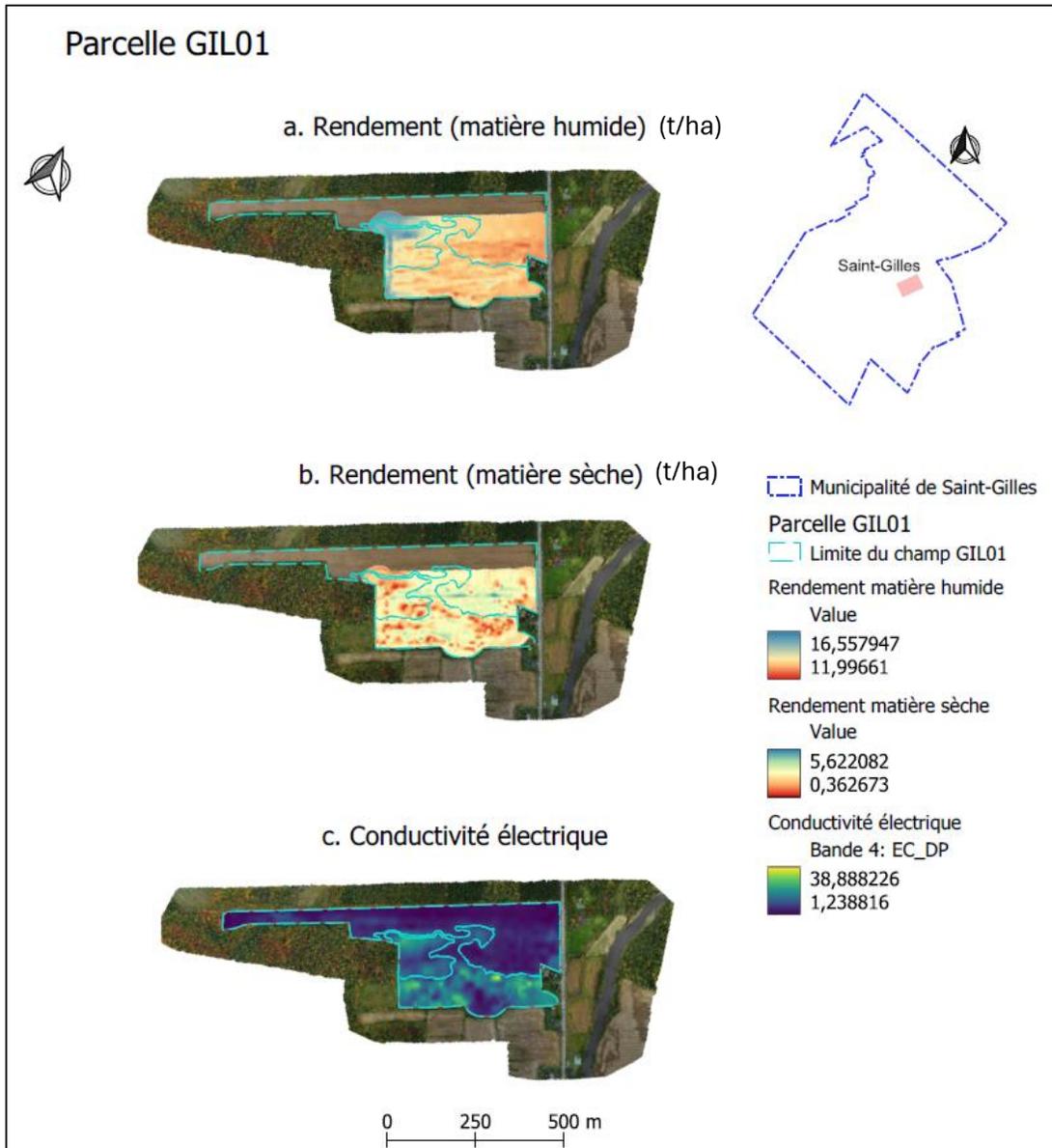


Figure 22. Carte des analyses de la parcelle GIL01 de la ferme Marcel Nadeau situé à Saint-Gilles de la MS et MH en considérant les données de la CE.

## 5 Interprétation et discussion des résultats

Les objectifs de cet essai étaient de contribuer à élaborer une méthodologie pour établir les fondations d'une infrastructure d'un système d'aide à la décision, visant à diagnostiquer et à remédier aux pertes de rendement en exploitant les données provenant d'agriculteurs. Afin d'y parvenir, des cartes d'échantillonnage par zone et de calibration CE-MO ont été réalisées dans ce projet. Elles ont ensuite permis de déterminer les emplacements optimaux pour les échantillons de sol recueillis pour les analyses en laboratoire en fonction des zones de variabilité des textures du sol dans le but de pouvoir interpréter les variations dans les rendements selon ces caractéristiques de sol. De plus, la carte de calibration CE-MO devait être automatisée à l'aide de script Python afin de pouvoir visualiser l'évolution de ces propriétés du sol. La cartographie basée sur le TPI devait permettre de cartographier les reliefs topographiques afin d'identifier des zones sujettes à l'accumulation d'eau ou à l'inverse, celles présentant un risque de sécheresse ou tout autre élément nuisant au rendement des cultures en lien avec la topographie. Finalement, les analyses géospatiales en fonction des zones de gestion et la réalisation d'un modèle de régression (*Random Forest*) afin de prédire le rendement en fonction des caractéristiques du sol et de la topographie ont été produits. L'ensemble de ces éléments a été réalisé.

Pour commencer, une carte d'échantillonnage par zone basée sur les différentes textures du sol attribués par la CE a été obtenue. Le partitionnement des champs agricoles en sous-parcelles homogènes, également appelées zones de gestion (ZG), est une pratique courante pour définir des zones en fonction des propriétés spatialement variables du champ. Cette approche permet d'identifier les zones selon la texture du sol, ce qui représente une stratégie clé en agriculture de précision pour soutenir les décisions et les actions visant à accroître l'efficacité, la productivité et la durabilité des processus agricoles (Ohana-Levi *et al.*, 2022). Les champs Be01, Be02, Be03 et Be04 de la ferme « Marcel Nadeau » ont été répartis en quatre classes : texture très grossière, texture grossière, texture fine et texture très fine.

Les zones délimitées sur la carte d'échantillonnage ont permis de mettre en évidence les variations de texture du sol en fonction de la CE, facilitant ainsi la collecte d'échantillons de sol. Ces échantillons seront prélevés de manière aléatoire dans chaque zone, garantissant ainsi la représentativité de chaque type de texture. Chaque zone disposera alors d'un échantillon destiné à une analyse en laboratoire pour obtenir plus de précisions sur les composantes du sol. Les

différentes caractéristiques du sol, réparties en zones distinctes, ouvrent la voie à une variété d'applications de gestion spécifiques à chaque sous-parcelle. Ces applications englobent des aspects tels que l'irrigation, la fertilisation et la lutte contre les éléments indésirables (Ohana-Levi *et al.*, 2022).

Ensuite, une carte des variations spatiales des propriétés du sol, en mettant en relation la CE et la MO, a été établie à l'aide de la méthode de classification non supervisée (*k-means*). Cette approche a permis de regrouper les données en « *clusters* » en fonction de la texture du sol, en combinant la CE et la teneur en MO. D'une part, la CE permet de définir la variabilité de la texture du sol, étant donné que certains éléments, comme l'argile, présentent une meilleure conductivité électrique que le limon et le sable, ce qui est expliqué par leur comportement hydrique. D'autre part, la MO améliore simultanément les caractéristiques physiques et chimiques du sol, entraînant des bénéfices pour sa qualité globale, ce qui contribue également à la description de la texture du sol dans un champ donné. En d'autres mots, la MO, une composante essentielle du sol, influence de nombreux aspects du sol, tels que sa stabilité, sa capacité de rétention d'eau, sa structure et sa fertilité. D'ailleurs, elle joue un rôle important dans les cycles biogéochimiques et est indispensable au bon fonctionnement des écosystèmes terrestres (Quideau *et al.*, 2021). C'est pourquoi ces deux éléments sont comparés dans les graphiques caractérisant les textures du sol.

Les analyses statistiques réalisées ont conduit à diviser la parcelle en trois classes : une zone caractérisée par une forte teneur en MO et en CE, une autre par une faible teneur en MO et en CE et enfin, une zone présentant une teneur moyenne en MO et en CE. Dans le code, des points centraux ont été ajoutés afin d'identifier l'endroit où l'échantillon de sol sera prélevé. Ainsi, chaque zone sera représentée par un échantillon pour une analyse en laboratoire des différentes composantes du sol, tel que demandé par l'entreprise XLKey.

Par la suite, une représentation cartographique identifiant les variations topographiques a été réalisée sur les champs B04 et BS02. Par ailleurs, les valeurs de TPI sont calculées en comparant l'altitude de chaque pixel avec l'altitude moyenne des pixels environnants. Les valeurs positives indiquent des crêtes, tandis que les valeurs négatives représentent des dépressions ou des creux. Les valeurs neutres correspondent à des zones planes ou à une pente constante (Marchese *et al.*, 2020).

Dans le champ BS02, il a été possible d'observer que dans une zone distincte, il y avait une variation dans les valeurs de TPI passant du rouge au vert en alternance. Donc, le contraste obtenu avec les valeurs de TPI a permis d'identifier certaines zones propices à l'accumulation d'eau, ou encore, des pentes abruptes qui pourraient augmenter l'érosion du sol à certains endroits. De plus, cette parcelle a une légère pente constante, et la zone marquée par la variation des valeurs de TPI correspond à la zone en plus haute altitude du champ. La pente constante permet, entre autres, de modéliser l'écoulement de l'eau. Cependant, ces deux éléments observés devront être combinés à d'autres facteurs afin d'identifier si cette zone est problématique pour le rendement des cultures. En effet, pris individuellement ou conjointement, ils ne suffisent pas à confirmer le diagnostic du rendement dans ce champ.

Dans le champ B04, une zone linéaire présentant un contraste marqué des valeurs de TPI est observée. Cette irrégularité est également visible dans les données LiDAR et sur l'orthophotographie. L'observation de ces éléments permet de constater que ce milieu peut correspondre à une barrière physique dans le champ, c'est-à-dire une zone où il n'y a pas de culture. Afin d'évaluer précisément le rendement de cette parcelle, il serait nécessaire de recréer les limites du contour du champ afin d'exclure cette zone dépourvue de culture. D'ailleurs, dans ce projet, plusieurs données géospatiales ont été interpolées afin d'être ensuite analysées. Ainsi, l'inclusion de cette zone sans culture dans les données pourrait directement influencer les résultats des analyses. Cela pourrait entraîner un biais dans les résultats et une interprétation incorrecte des données, ce qui pourrait conduire à des décisions et des actions inappropriées, car elles seraient basées sur des analyses contenant des données incorrectes.

Enfin, des analyses statistiques ont été réalisées et automatisées. L'ensemble de tous les champs transmis par XLKey a été examiné dans le cadre de la modélisation du rendement en fonction des propriétés du sol (CE et MO) et de la topographie (LiDAR). Les parcelles suivantes ont été analysées en fonction des variables incidentes obtenues, soit la caractéristique la plus pertinente obtenue dans la prédiction du modèle : FR02, GIL01, BS02, B03, B04 et N01. Pour chaque valeur incidente associée à une parcelle, il a été possible d'établir des liens avec le rendement. En ce qui concerne les parcelles présentant une corrélation forte avec les données du LiDAR, il a été observé que certaines zones de rendement pouvaient être corrélées avec la topographie du

champ. Il en va de même avec les parcelles pour lesquelles on observe une corrélation du rendement en fonction de la teneur en MO ou la CE.

Dans l'ensemble, le projet a permis d'identifier certaines zones présentant des problématiques. Toutefois, ce projet devra être davantage élaboré afin d'obtenir des résultats plus concrets. D'ailleurs, l'intégration des données de précipitations, des heures d'ensoleillement, de la température (minimale ou maximale) et de l'humidité relative est essentielle pour élaborer un diagnostic sur les pertes de rendement (Sharma *et al.*, 2023). De même, les informations sur l'irrigation et la quantité d'intrants devront être ajoutées. Ainsi, les variables explicatives de la topographie et des textures du sol qui ont été utilisées dans ce rapport ne suffisent pas à elles seules pour confirmer les problématiques liées au rendement dans un milieu agricole. En somme, l'ajout de données, que ce soit des données de sol, de culture ou d'opération, ne peut qu'être bénéfique pour l'aide à la décision. De plus, l'ajout de données sur les champs à l'étude au cours des années à venir permettra de réaliser des analyses évolutives sur les rendements, ce qui facilitera les analyses et les décisions à venir.

## 6 Conclusion et Perspectives

L'ensemble des analyses ont permis d'obtenir certaines informations pour l'aide à la prise de décision afin d'augmenter les rendements agricoles. Cependant, quelques modifications sont à prévoir pour améliorer l'exactitude et la fiabilité des prévisions de rendement des cultures, car la précision est très importante afin de bien évaluer les modifications à apporter afin de réduire les coûts et augmenter les rendements des cultures (Sharma *et al.*, 2023).

D'ailleurs, malgré le nettoyage des données aberrante, certaines données devraient être exclues des analyses. En effet, les limites des champs comportent parfois des superficies sans culture et les données sont interpolées sur l'ensemble du champ prédéfini par ces limites. Par conséquent, l'analyse utilise également les données sur ces zones sans cultures, ce qui émet un biais sur les résultats obtenus, car les données interpolées ne représentent pas la réalité des rendements agricoles dans les zones cultivées. D'un autre côté, certaines données étaient manquantes pour une variable explicative tel que pour le champ FR02, ce qui modifiait également les résultats obtenus lors de la modélisation.

De plus, la prise de donnée a été effectuée par plusieurs personnes dans un cadre expérimental avec des conditions environnementales changeantes, mais ces variables abiotiques n'ont pas été prises en compte dans les analyses, ce qui constitue une limite significative pour le projet. Ainsi, le manque de rigueur dans l'acquisition des données et la non-prise en compte de l'effet des variables environnementales limitent la qualité des données utilisées dans cette étude.

Enfin, l'entreprise avait collecté des données à l'aide du capteur Veris MSP3 en 2021 et 2022, mais la prise de données a été considérablement limitée en 2023 en raison des conditions peu propices à l'utilisation de cet outil. Comme mentionné précédemment, les conditions d'humidité et de topographie doivent demeurer constante afin de considérer la texture du sol pour déterminer la CE ainsi que la température optimale et l'absence de précipitation au moment des mesures demeurent nécessaires pour pouvoir utiliser les données lors d'analyse. De ce fait, l'entreprise a décidé d'opter pour un autre outil de collecte de données pour l'année 2024 et les années à venir. Étant donné l'évolution rapide de la technologie, d'autres changements seront également à prévoir afin d'augmenter la qualité de la donnée et donc, l'automatisation des méthodes par script en Python réalisée permettra facilement de modifier les éléments qui évoluent dans ce projet.

À la lumière de ce qui a été présenté dans cette étude, l'objectif principal de contribuer au développement des bases de l'infrastructure d'un système d'aide à la décision au diagnostic en agriculture a été atteint. Des résultats d'analyse de sol seront également ajoutés au projet afin d'évaluer plus efficacement les problématiques de rendement. Ce projet de maîtrise était une des premières étapes d'un projet d'envergure de l'entreprise XLKey afin de permettre aux agriculteurs clients de situer la performance de leurs cultures et de leurs opérations en comparaison à d'autres producteurs partageant des conditions semblables. Quelques modifications sont à prévoir, cependant en rassemblant des données sur le sol, les plantes et les opérations sur plusieurs années et en réalisant les analyses présentées dans ce présent rapport, l'entreprise pourra mieux cibler les problématiques pour améliorer le rendement, réduire les coûts de production et également, diminuer les émissions de GES.

## 7 Références

ArcGIS Pro, 2023. À quoi correspondent les données LiDAR? Archive de l'aide. [site Web].

Consulté le 11 novembre 2023. <https://pro.arcgis.com/fr/pro-app/latest/help/data/las-dataset/what-is-lidar-.htm>

Aydin, S. et Aydin, M. N. (2020) Ontology-based data acquisition model development for agricultural open data platforms and implementation of OWL2MVC tool. *Computers and Electronics in Agriculture*, vol. 175, p. 105589.

Belgiu, M. et Drăguț, L. (2016) Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, p. 24-31.

Brevik E.C, Thomas, E. F et A. Lazari, 2006. Soil electrical conductivity as a function of soil water content and implications for soil mapping. *Precision Agriculture*, Volume 7, Issue 6, pp. 393-404.

Castro, A., Martínez-Osuna, J. F., Michel, R., Escoto-Rodríguez, M., Bullock, S. H., Cueva, A., López-Reyes, E., Reimer, J., Salazar, M., Villarreal, S. et Vargas, R. (2017) A low-cost modular data-acquisition system for monitoring biometeorological variables. *Computers and Electronics in Agriculture*, vol. 141, p. 357-371.

Centre de Référence en Agriculture et Agroalimentaire du Québec (CRAAQ) (2010). Guide de référence en fertilisation. 473 – 2<sup>e</sup> édition.

Cheng, X., Guo, T., Mo, X., Zhang, L., Wang, R., Xue, Y., Fan, X. et Sun, S. (2023) Reduced nutrient release and greenhouse gas emissions of lignin-based coated urea by synergy of carbon black and polysiloxane. *International Journal of Biological Macromolecules*, vol. 231, p. 123334.

Conseil des productions végétales du Québec inc (CPVQ)(1996). Grilles de référence en Fertilisation. 128 2<sup>e</sup> édition. Agdex 540. Publication 02-9605

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., Chu, W., Antrop, M., De Maeyer, P., Finke, P., Van Meirvenne, M., Verniers, J. et Crombé, P. (2013) Application of the topographic position index to heterogeneous landscapes. *Geomorphology*, vol. 186, p. 39-49.

FlyGuys (2021). 7 Uses for LiDAR in Agriculture | Maximize Profits, Minimize Costs. FlyGuys. [site web]. Consulté le 8 mai 2024. <https://flyguys.com/uses-for-lidar-in-agriculture/>

- Gagnon, M. (2023). Les moissonneuses-batteuse à l'heure de l'intelligence artificielle. La Terre de chez nous. [site web]. Consulté le 29 février 2024. <https://www.laterre.ca/productions/grandescultures/les-moissonneuses-batteuses-a-lheure-de-lintelligence-artificielle/>
- Google. (2023). Google Colaboratory. Consulté le 10 octobre 2023. <https://colab.research.google.com/>
- Gouvernement du Québec (2024). Culture des grains (céréales et oléagineux). Gouv. Québec, Agriculture, environnement et ressources naturelles. [site web]. Consulté le 13 septembre 2024. <https://www.quebec.ca/agriculture-environnement-et-ressources-naturelles/agriculture/industrie-agricole-au-quebec/productions-agricoles/culture-grains-cereales-oleagineux>
- Han, L., Yang, G., Yang, H., Xu, B., Li, Z. et Yang, X. (2018) Clustering field-based maize phenotyping of plant-height growth and canopy spectral dynamics using a UAV remote-sensing approach. *Frontiers in Plant Science*, vol. 871.
- Heka.AI (2022). Clustering de séries temporelles, Marketing & Sales. [site web]. Consulté le 14 janvier 2023. <https://www.heka.ai/fr/nos-publications/clustering-de-series-temporelles>
- Heung, B., Saurette, D. et Bulmer, C. (2021) Cartographie numérique des sols. University of Saskatchewan. [site officiel]. Consulté le 14 janvier 2024. <https://openpress.usask.ca/soilsciencefrench/chapter/digital-soil-mapping/>
- Institut de recherche et de développement en agroenvironnement (IRDA) (2016). Optimisation de l'irrigation par une caractéristique physique précise des sols-II. Rapport final. Gouvernement du Québec. [Fichier pdf]. 54p. <https://irda.blob.core.windows.net/media/5455/boivin-et-al-2016-optimisation-de-lirrigation-par-une-caracterisation-physique-precise-des-sols-ii-rapport.pdf>
- Jacques, D. 2022. Le quoi et le pourquoi du coût de production. Économie et gestion, Agri-Réseau. Blogue. Consulté le 8 mai 2024. <https://www.agrireseau.net/economie-et-gestion/blogue/108761/le-quoi-et-le-pourquoi-du-co%C3%BBt-de-production>
- Lee, S. et Park, J. (2021) A spatiotemporal data acquisition toolkit for volume estimation tools in precision agriculture. *Concurrency and Computation: Practice and Experience*, vol. 33, n°3.
- Lherbier, J. (2005) Valorisation de l'information géographique en agriculture de précision (phdthesis). Thèse de doctorat, Université Joseph-Fourier - Grenoble I.

- Martel, S., Fournier-Marcotte C. et H. Moore, 2015. Guide pour la production et l'utilisation des cartes de rendement, 23 pages + 1 annexe. [fichier pdf]. Consulté le 10 janvier 2014. <https://www.agrinova.qc.ca/wp-content/uploads/2016/09/PAF-1804-1-Guide-version-finale-1.pdf>
- Mairura, F. S., Mugendi, D. N., Mwanje, J. I., Ramisch, J. J., Mbugua, P. K. et Chianu, J. N. (2007a) Integrating scientific and farmers' evaluation of soil quality indicators in Central Kenya. *Geoderma*, vol. 139, n°1, p. 134-143.
- Mairura, F. S., Mugendi, D. N., Mwanje, J. I., Ramisch, J. J., Mbugua, P. K. et Chianu, J. N. (2007b) Scientific evaluation of smallholder land use knowledge in central Kenya. *Land Degradation and development, Tropical Soil Biology and Fertility*. Nairobi, Kenya. p. 77-90.
- Marchese, F., Bracchi, V., Lisi, G., Basso, D., Corselli, C. et Savini, A. (2020) Assessing Fine-Scale Distribution and Volume of Mediterranean Algal Reefs through Terrain Analysis of Multibeam Bathymetric Data. A Case Study in the Southern Adriatic Continental Shelf. *Water*, vol. 12, p. 157.
- Marques Ramos, A. P., Prado Osco, L., Elis Garcia Furuya, D., Nunes Gonçalves, W., Cordeiro Santana, D., Pereira Ribeiro Teodoro, L., Antonio da Silva Junior, C., Fernando Capristo-Silva, G., Li, J., Henrique Rojo Baio, F., Marcato Junior, J., Eduardo Teodoro, P. et Pistori, H. (2020) A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, vol. 178, p. 105791.
- Maycock, D. (2020). The new data wave in agriculture. *Forbes Technology council*. [site Web]. Consulté le 16 janvier 2023. <https://www.forbes.com/sites/forbestechcouncil/2020/12/11/the-new-data-wave-in-agriculture/?sh=24b32deb79d5>
- McDonald, I. (2015). Gestion spécifique au site en agriculture de précision. Ministère de l'agriculture, de l'alimentation et des affaires rurales, Ontario [En ligne]. Consulté le 2 novembre 2022. <http://www.omafr.gov.on.ca/french/crops/field/news/croptalk/2015/ct-0615a2.htm>
- Ordre des agronomes du Québec (OAQ) (s.d.) Agriculture Durable. Ordre des agronomes du Québec. [Site web]. Consulté le 2 novembre 2022. <https://oaq.qc.ca/communications/agro-express/agriculture-durable/>
- Ministère des ressources naturelles et des forêts (MRNF), 2016. LiDAR - Modèles numériques

- (terrain, canopée, pente), [Jeu de données], dans Données Québec, 2016, mis à jour le 15 janvier 2024. <https://www.donneesquebec.ca/recherche/dataset/produits-derives-de-base-du-lidar>. consulté le 18 mars 2023.
- Muddarisna, N., Yuniwati, E. D., Masruroh, H. et Oktaviansyah, A. R. (2020) An Automated Approach Using Topographic Position Index (TPI) for Landform Mapping (Case Study: Gede Watershed, Malang Regency, East Java, Indonesia). IOP Conference Series: Earth and Environmental Science, vol. 412, n°1, p. 012027.
- Ohana-Levi, N., Gao, F., Knipper, K., Kustas, W. P., Anderson, M. C., del Mar Alsina, M., Sanchez, L. A. et Karnieli, A. (2022) Time-series clustering of remote sensing retrievals for defining management zones in a vineyard. Irrigation Science, vol. 40, n°4, p. 801-815
- Oliveira, M. F. D., Ortiz, B. V., Morata, G. T., Jiménez, A.-F., Rolim, G. D. S. et Silva, R. P. D. (2022) Training Machine Learning Algorithms Using Remote Sensing and Topographic Indices for Corn Yield Prediction. Remote Sensing, vol. 14, n°23, p. 6171.
- Pereira, G. W., Valente, D. S. M., Queiroz, D. M. de, Coelho, A. L. de F., Costa, M. M. et Grift, T. (2022) Smart-Map: An Open-Source QGIS Plugin for Digital Mapping Using Machine Learning Techniques and Ordinary Kriging. Agronomy, vol. 12, n°6, p. 1350.
- Pradipta, A., Soupios, P., Kourgialas, N., Doula, M., Dokou, Z., Makkawi, M., Alfarhan, M., Tawabini, B., Kirmizakis, P. et Yassin, M. (2022). Remote Sensing, Geophysics, and Modeling to Support Precision Agriculture—Part 2: Irrigation Management. Water (Switzerland), vol. 14, n°7.
- QGIS.org, 2022. QGIS Geographic Information System. QGIS Association. [logiciel gratuit].  
Télécharger le 10 janvier 2022. <http://www.qgis.org>
- Quideau, S., Simpson, M. et Gillespie, A. (2021) Matière organique du sol. [site Web]. Consulté le 16 avril 2024
- Reyes, J. F., Correa, C., Esquivel, W. et Ortega, R. (2012) Development and field testing of a data acquisition system to assess the quality of spraying in fruit orchards. Computers and Electronics in Agriculture, vol. 84, p. 62-67.
- Say S.M., Keskin M., Sehrli M., Y.E. Sekerli (2018). Adoption of precision agriculture technologies in developed and developing countries. The Online Journal of Science and Technology (TOJSAT). 8(1):7 15.
- Science education through earth observation for high schools (SEOS eLearning Tutorials) (s.d.).  
Télétection et SIG en agriculture – Supplément – Indices de végétation. Université Carl Von Ossietzky d’Oldenburg Allemagne. Site officiel [site Web]. Consulté le 8 février 2024. <https://seos-project.eu/agriculture/agriculture-c01-s04.fr.html>

- Sharma, S., Sharma et Gaur, K. (2023). Crop yield predictions and recommendations using random forest regression in 3A agroclimatic zone, Rajasthan. *Shu Ju Cai Ji Yu Chu Li/Journal of Data Acquisition and Processing*, vol. 38, p. 1635-1651.
- TechVidvan. (2024). Cluster analysis in R – Complete Guide on Clustering in R. R- Tutorials. [site Web]. Consulté le 12 février 2024. <https://techvidvan.com/tutorials/cluster-analysis-in-r/>
- Timsina J., Dutta S., Devkota K. P, Chakraborty S., Neupane R. K., Bishta S., Amgain L. P., Singh V. K., Islam S. & K. Majumdar. (2021). Improved nutrient management in cereals using Nutrient Expert and machine learning tools: Productivity, profitability and nutrient use efficiency – ScienceDirect. 192 – 103181.
- Tittonell P, Vanlauwe B, Leffelaar PA, Rowe E, Giller KE. (2005). Exploring diversity in soil fertility management of smallholder farms in Western Kenya I. Heterogeneity at region and farm scale. *Agriculture, Ecosystems and Environment* 110: 149–165.
- Toupin M. (2022). Intelligence artificielle en agriculture : pour une récolte de données précises et Précieuses. *La Nouvelle Union et L’Avenir de l’Érable*. [En ligne]. consulté le 2 novembre 2022. <https://www.lanouvelle.net/actualite/intelligence-artificielle-en-agriculture-pour-une-recolte-de-donnees-precises-et-precieuses/>
- Veris Technologies – Operating instructions. (2012). Broadway, Salina Kansas. 106p. [fichier pdf]. Consulté le 15 janvier 2023. [https://3d58a414-6c87-4db3-9429-c84a475a79bf.filesusr.com/ugd/1fc344\\_44e9e48518704472a4bd584f7d85fce0.pdf](https://3d58a414-6c87-4db3-9429-c84a475a79bf.filesusr.com/ugd/1fc344_44e9e48518704472a4bd584f7d85fce0.pdf)
- XLKey (2023). *Entreprise XLKey – Agriculture de précision*, Site officiel [site web]. Consulté le 9 janvier. <https://xlkey.ca/>
- Yuan, Y., Shi, B., Yost, R., Liu, X., Tian, Y., Zhu, Y., Cao, W. et Cao, Q. (2022) Optimization of Management Zone Delineation for Precision Crop Management in an Intensive Farming System. *Plants*, vol. 11, n°19, p. 2611.