

L'Institut Agro Rennes-Angers

☐ Site d'Angers ☒ Site de Rennes

Année universitaire: 2024-2025

Spécialité: Master TELENVI

Mémoire de fin d'études

☐ d'ingénieur de l'Institut Agro Rennes-Angers (Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement)

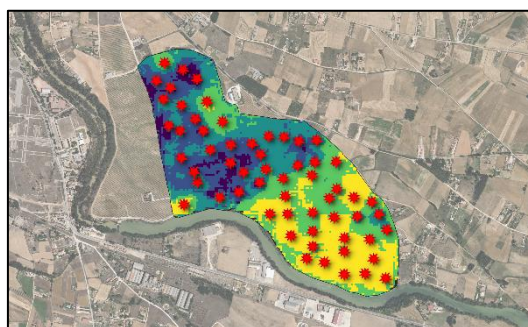
☒ de master de l'Institut Agro Rennes-Angers (Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement)

☐ de l'Institut Agro Montpellier (étudiant arrivé en M2)

☐ d'un autre établissement (étudiant arrivé en M2)

Couplage de séries temporelles Sentinel-2 avec des spectres VPIR laboratoire pour cartographier les teneurs en carbone organique et en oxydes de fer des sols d'un vignoble espagnol

Colin FABRE



Soutenu à Rennes, le vendredi 19 septembre 2025

Devant le jury composé de:

Laurence HUBERT-MOY, présidente

Youssef FOUAD, encadrant de stage et enseignant référent

Hayfa ZAYANI, co-encadrante

Emmanuelle VAUDOUR, examinatrice

Baptiste MARTEAU, examinateur

Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celle de l'Institut Agro Rennes-Angers

Remerciements

Ces six mois de stage de fin d'études furent six excellents mois grâce à plusieurs personnes, pour qui je formule ces quelques lignes de remerciements.

Un remerciement tout particulier à Youssef, qui n'a pas hésité à me proposer ce sujet de stage dès mes premières semaines de recherche, malgré le fait que le sujet était encore en début de construction et qu'il y avait beaucoup à faire sur le plan conceptuel comme technique en seulement six mois. Merci pour ton amabilité immuable, ton soutien dans mes propositions et pour toutes les galères administratives qu'il aura fallu endurer. Je te dois ces six beaux mois de stage, aussi studieux que plaisants.

Un merci tout aussi chaleureux à Hayfa, avec qui ce fût un plaisir de travailler sur le projet SANCHOSTHIRST pendant toute la durée du stage. Cela a commencé par un voyage enrichissant à Madrid pour le *mid-term meeting* de Sancho, pour se terminer dans la sueur des posters Eurosoil.

Merci à l'homme-à-tout-faire du labo, j'ai bien nommé Pascal. Merci pour ton aide à chaque fois que j'ai bloqué dans mes codes, pour transformer mes scripts de 127 lignes en 4, mon boost d'ordinateur et tes remarques toujours pertinentes. C'était un vrai plaisir de travailler avec toi sur ce projet, et de pouvoir venir discuter dans ton bureau à la débottée sur telle ou telle manière de procéder en traitement de données.

Merci à Didier, pour toutes les belles connaissances pédologiques que tu m'as transmises, ton humour bourguignon sans faille et ton cardio à toute épreuve. Avec toi, c'était la bonne humeur assurée sur le terrain.

Merci à tous mes autres collègues de SANCHOSTHIRST pour vos remarques et conseils lors des réunions de projet pour m'aider à avancer: Emmanuelle, Nicolas, María José.

Merci à Hugues et Syrine, mes collègues de bureau, pour la bonne ambiance de travail et la cohésion face aux galères administratives. Merci Hugues pour ces kilomètres avalés de course à pied, le sourire aux lèvres et le souffle long, alors que nous autres avions de la peine à garder le rythme.

Merci à tous mes autres collègues de PSN pour nos cafés-goûters et discussions de qualité à table: Zahra, Nelly, Gaëlle, Nicolas, François, Stéphane, Élodie, Mohammad, Claire, Guilhem.

Merci à Séverine, pour ton aide précieuse et ta lumière dans la tempête administrative.

Des remerciements chaleureux à ma famille et mes amis, qui m'ont toujours soutenu dans mes choix et mes travaux. Merci. Vous savez tout ce que je vous dois.



Glossaire

Tous les sigles et noms complets d'indices spectraux utilisés dans ce travail sont présentés dans l'Annexe 2.

(N)RMSE	<i>Normalised Root Mean-Square Error</i>
R_{CV}^2	R ² obtenu par <i>Cross Validation</i>
Abs	Absorbance
ACP	Analyse en Composantes Principales
AFM	Analyse Factorielle Multiple
ANN	<i>Artificial Neural Network</i>
API	<i>Application Programming Interface</i>
AUC	<i>Area Under Curve</i>
BV	Bassin Versant
CC	<i>Covered Crops</i>
CEC	<i>Cation-Exchange Capacity</i>
CNES	Centre National d'Études Spatiales
CNS	Cartographie Numérique des Sols
CR	<i>Continuum-Removal</i>
CV	<i>Cross Validation</i>
DL	<i>Deep Learning</i>
DNN	<i>Deep Neural Network</i>
FeOx	<i>Ferrous Oxides</i>
FWHM	<i>Full Width at Half Maximum</i>
GTB	<i>Gradient Tree Boosting</i>
LOOCV	<i>Leave-One-Out Cross Validation</i>
MBE	<i>Mean Bias Error</i>
ML	<i>Machine Learning</i>
MLR	<i>Multiple Linear Regression</i>
MO	Matière Organique
MSC	<i>Multiplicative Scatter Correction</i>

NIR	<i>Near-Infrared</i>
PLSR	<i>Partial Least Squares Regression</i>
ReLU	<i>Rectified Linear Unit</i>
RF	<i>Random Forest</i>
RPD	<i>Relative Percent Difference</i>
RPIQ	<i>Ratio of Performance to InterQuartile distance</i>
RVI	<i>Relative Variable Importance</i>
S-1C	Sentinel-1C
S-2A/B	Sentinel-2A/B
SG	SAVITZKY-GOLAY
SNV	<i>Standard Normal Variate</i>
SOC	<i>Soil Organic Carbon</i>
VNIR	<i>Visible and Near-InfraRed</i>
VPIR	Visible et Proche-InfraRouge

Table des matières

Introduction.....	10
Chapitre 1 – Revue de la littérature.....	11
Carbone organique et oxydes de fer dans les sols agricoles	11
Télédétection optique pour prédire la teneur en SOC	12
Revue antérieures et lacunes.....	13
Chapitre 2 – Données & méthodes.....	15
Site d'étude	15
Contexte général et localisation	15
Données terrain et caractéristiques pédologiques.....	16
Données de télédétection.....	16
Séries temporelles Sentinel-2	16
Spectres laboratoire d'échantillons terrain	17
Réduction de la dimensionnalité des variables prédictives	18
Création de combinaisons prédictives	20
Calibration de modèles prédictifs.....	21
Chapitre 3 – Résultats	22
Statistiques descriptives des échantillons terrain	22
Statistiques descriptives des teneurs en SOC et FeOx dans les trois tirages	22
Statistiques de performance des modèles prédictifs pour le SOC	23
Statistiques de performance des modèles prédictifs pour l'oxyde de fer (III)	23
Résultats des prédictions du SOC à l'aide de toute la série temporelle S-2	24
Chapitre 4 – Perspectives	25
Intégration de l'information spatialisée de la charge en éléments grossiers.....	25
Apport de l'imagerie hyperspectrale	25
Optimisation du plan d'échantillonnage et de l'incrémentation des jeux de données	26
Approches hybrides.....	26
Conclusion	27
Références.....	28
Annexes.....	34

Table des illustrations

Figure 1 – Carte de localisation du site d'étude et des points d'échantillonnage terrain.....	15
Figure 2 – Frise chronologique des campagnes de prélèvements terrain sur le site d'étude, emprunté à ZAYANI et al., (travaux non-publiés); les échantillons d'Aranda ont été prélevés lors des campagnes de janvier et mars 2024.....	16
Figure 3 – Courbe du NDVI de la vigne en climat méditerranéen, LI et al. (2021).....	16
Figure 4 – Diagramme de traitements des données pédologiques et de télédétection.....	19
Figure 5 – Boxplots de la distribution des teneurs en carbone organique, fer total libre et oxydes de fer (Fe^{2+} et Fe^{3+}) sur le site d'Aranda de Duero.....	22
Figure 6 – Histogrammes de la distribution de la teneur en SOC.....	22
Figure 7 – Histogrammes de la distribution de la teneur en Fe^{3+}	22
Figure 8 – Statistiques de performance des modèles SVR pour le SOC.....	23
Figure 9 – Statistiques de performance des modèles SVR pour le Fe^{3+}	23
Figure 10 – Carte des prédictions de la teneur moyenne en SOC sur le vignoble d'Aranda à l'aide de la série temporelle S-2.....	24
Figure 11 – Carte de l'écart-type des prédictions de la teneur moyenne en SOC sur le vignoble d'Aranda à l'aide de la série temporelle S-2.....	24

Table des annexes

Annexe 1 – Diagramme ombrothermique d’Aranda de Duero	34
Annexe 2 – Carte pédologique d’Aranda de Duero	34
Annexe 3 – Répartition des images par mois de la banque finale d’images de la série temporelle Sentinel-2	35
Annexe 4 – Indices spectraux calculés à partir des images Sentinel-2.....	35
Annexe 5 – Indices spectraux hyperspectraux adaptés à Sentinel-2	37
Annexe 6 – Montage du spectroradiomètre pour l’acquisition des spectres de laboratoire des échantillons terrain	38
Annexe 7 – Paramètres géométriques de la courbe spectrale corrigée calculés au droit des bandes d’absorption de l’oxyde de fer (III).....	38
Annexe 8 – Tableau de composition des jeux de données.....	39
Annexe 9 – Statistiques de calibration des modèles prédictifs pour le SOC.....	40
Annexe 10 – Statistiques de calibration des modèles prédictifs pour les FeOx	42
Annexe 11 – Carte de l’indice de confiance au seuil de 95% des prédictions de la teneur moyenne en SOC à l’aide de la série temporelle S-2	44
Annexe 12 – Carte du nombre de dates en sol par pixel utilisées pour les prédictions de la teneur moyenne en SOC à l’aide de la série temporelle S-2	44
Annexe 13 – Carte des prédictions de la teneur moyenne en SOC sur le vignoble d’Aranda à l’aide de la date la plus en sol nu	45
Annexe 14 – Carte de l’écart-type des prédictions de la teneur moyenne en SOC sur le vignoble d’Aranda à l’aide de la date la plus en sol nu	45
Annexe 15 – Carte du NSSI	46

Introduction

Les sols agricoles jouent un rôle clé dans la régulation du climat, la sécurité alimentaire et la préservation de la biodiversité. En tant que premier réservoir de carbone organique terrestre, ils stockent plus de carbone que l'atmosphère et la végétation réunies (LAL et al., 2018). Pourtant, leur dégradation, accentuée par les pratiques culturales intensives et les aléas climatiques, menace cette fonction essentielle. À l'échelle mondiale, des initiatives comme « l'Initiative 4 pour 1000 » (Ministère de l'Agriculture, FRANCE 2015) visent à augmenter les stocks de carbone organique des sols (SOC) de 0.4 % par an pour atténuer les impacts du changement climatique et améliorer la fertilité des sols. En Europe, cette problématique est particulièrement critique dans les régions méditerranéennes semi-arides, où les sols, soumis à une érosion accrue et à des stress hydriques intenses, voient leur teneur en SOC diminuer, compromettant leur qualité et leur résilience. Dans ce contexte, la cartographie précise du SOC est un enjeu majeur pour évaluer l'état des sols et orienter les pratiques agricoles vers une gestion durable. Les oxydes de fer (FeOx), abondants dans les sols méditerranéens, interviennent à la fois dans la pédogenèse (via des processus d'oxydo-réduction) et dans la stabilisation du SOC (WIESMEIER et al., 2012). Leur présence influence également la réponse spectrale des sols, ce qui en fait des co-variables potentielles pour améliorer les modèles prédictifs du SOC.

Ce mémoire s'inscrit dans le cadre du projet européen SANCHOSTHIRST¹ (2023–2026), qui vise à évaluer l'impact des modes de conduite des sols (labour traditionnel vs couvert végétal maintenu) sur leur santé et leur capacité à stocker du carbone. Plus spécifiquement, ce travail se concentre sur l'amélioration des modèles de prédiction du SOC en couplant des données satellitaires optiques (Sentinel-2) et des spectres de laboratoire (Vis-PIR). Les hypothèses de recherche sont que l'utilisation de combinaisons incrémentielles de variables (indices spectraux, bandes Sentinel-2, données pédologiques), ainsi que l'intégration des FeOx comme prédicteurs, permettra d'affiner la précision des cartographies du SOC. Pour tester ces hypothèses, différentes approches modélisatrices sont comparées: des modèles linéaires (PLSR), des méthodes de machine learning (SVR, RF) et des réseaux de neurones profonds (DNN). La méthodologie employée repose sur trois axes: (1) la construction de jeux de données incrémentiels, intégrant progressivement des co-variables spectrales et pédologiques pour évaluer leur apport respectif; (2) la comparaison systématique des performances des modèles, en fonction de leur complexité et de leur capacité à capturer les interactions entre SOC et FeOx; (3) l'analyse spatiale des résidus de prédiction, pour identifier si les zones mal estimées coïncident avec des teneurs élevées en FeOx, et ainsi affiner les stratégies de modélisation.

Ce mémoire débutera par un état de l'art des méthodes de cartographie numérique du SOC par télédétection, avant de présenter le site d'étude, les données utilisées (terrain, satellite, laboratoire), et les modèles prédictifs mis en œuvre. Les résultats seront enfin discutés au regard de leur potentiel pour une gestion adaptée des sols viticoles, en vue d'optimiser leur durabilité et leur contribution à la séquestration du carbone.

¹ Cover cropS (CC) ANd soil health and climAte CHaNge adaptatiOn in Semiarid woody crops. THe RemOte SensIng and furTHer scenaRIoS projecTions

Chapitre 1 – Revue de la littérature

L'étude des interactions entre la végétation et les processus physico-chimiques dans le sol connaît un certain historique, principalement autour de la question de la séquestration du carbone organique. En effet, cette thématique a suscité un intérêt croissant au sein de la communauté scientifique en raison du rôle crucial du SOC dans la régulation du climat et la mitigation des effets du changement climatique, ainsi que de son importance en tant qu'indicateur de la qualité des sols. Le SOC est impliqué dans de nombreux services écosystémiques en participant au cycle global du carbone, en contribuant à lutter contre la dégradation de sols et en renforçant la sécurité alimentaire. Les études réalisées dans ce domaine ont mis en lumière les mécanismes par lesquels les plantes capturent le dioxyde de carbone atmosphérique et le transfèrent vers le sol, où il est stocké sous forme de matière organique. Ce processus, connu sous le nom de séquestration du carbone, est influencé par divers facteurs, notamment les types de végétation, les propriétés du sol et les conditions climatiques.

Carbone organique et oxydes de fer dans les sols agricoles

Le SOC et les FeOx jouent un rôle central dans la fertilité, la structure et la résilience des sols viticoles. Le fer s'y présente sous deux formes principales: le fer libre (oxydes et hydroxydes cristallins ou amorphes, comme la goethite, l'hématite ou la ferrihydrite) et le fer non-libre (intégré dans les structures cristallines des minéraux primaires ou argiles). Le fer libre, extrait par des réactifs comme le dithionite-citrate-bicarbonate (DCB), joue un rôle clé dans la stabilisation du SOC via des mécanismes d'adsorption, de co-précipitation et de complexation. Les oxydes de fer amorphes (comme la ferrihydrite) sont particulièrement réactifs en raison de leur surface spécifique élevée (jusqu'à 600m²/g), qui favorise la formation de complexes organo-minéraux avec les acides humiques et fulviques, limitant ainsi la minéralisation du SOC (YE et al., 2022, JIA et al., 2024).

La valence du fer (Fe²⁺ vs Fe³⁺) influence également ces interactions. En conditions anoxiques ou humides (dans des sols engorgés temporairement), le Fe²⁺ domine et peut former des complexes solubles avec la matière organique, augmentant sa mobilité mais réduisant sa stabilité à long terme. À l'inverse, en conditions oxydantes (sols bien drainés), le Fe³⁺ précipite sous forme d'oxydes (goethite, hématite), qui adsorbent fortement le SOC et le protègent de la décomposition microbienne. Une étude récente (WU et al., 2024) a montré que jusqu'à 33% du SOC total dans les horizons de surface est associé aux oxydes de fer, avec une affinité particulière pour les fractions aromatiques et hydrophobes du SOC, stabilisées par des liaisons de type ligand exchange ou ponts cationiques (KIRSTEN et al., 2021). Les mécanismes de stabilisation du SOC par les FeOx dépendent du pH et des conditions redox du sol. En milieu acide et humide, les FeOx amorphes présentent une forte affinité pour le SOC, tandis qu'en conditions alcalines ou sèches, leur rôle est moins marqué en raison de la cristallisation des phases ferriques et de la compétition avec les argiles. Aussi, la dynamique redox des sols viticoles (alternance de périodes humides/sèches), leur pauvreté en MO et une forte érosion déstabilise les complexes SOC-Fe. Cette vulnérabilité est accentuée dans les vignobles où les pratiques culturales (travail du sol, apports de cuivre) modifient la spéciation du fer. Le cuivre, utilisé comme fongicide, s'adsorbe préférentiellement sur les FeOx amorphes, réduisant leur capacité à lier le SOC et favorisant la formation de complexes Cu-Fe-MO, moins stables que les liaisons Fe-SOC pures. Par exemple (PARAT et al., 2002), dans les sols calcaires des vignobles méditerranéens, jusqu'à 61% du cuivre accumulé (issu des traitements fongicides) est lié aux FeOx amorphes, ce qui modifie la répartition du SOC entre les fractions granulométriques et renforce sa persistance. Cependant, sous des conditions d'anaérobiose intermittente (ex: sols temporairement engorgés), la réduction microbienne des FeOx peut libérer le SOC précédemment protégé, augmentant ainsi sa minéralisation et les émissions de CO₂ (CHEN et al., 2020).

L'utilisation de la spectroscopie VPIR pour caractériser les propriétés des sols, en particulier les FeOx et le SOC, a fait l'objet de nombreuses avancées, bien que les résultats varient selon les contextes pédologiques et les méthodes employées. SUMMERS et al. (2011), travaillant sur des sols à faible fertilité en Australie méridionale, ont montré que cette approche permet de prédire avec une bonne précision les teneurs en argile et en carbonates (RPD \approx 2.1, R² \approx 0.68) à partir de spectres de terrain (400-2'500nm), mais se heurte à des limites pour le SOC et les FeOx. Leurs analyses révèlent des longueurs d'onde clés associées à ces propriétés: 1'900nm et 2'300nm pour le SOC (liées aux vibrations des groupes fonctionnels organiques), et 550nm et 900nm pour les FeOx, correspondant aux bandes d'absorption de la goethite. Cependant, les auteurs soulignent un manque de transférabilité spatiale des modèles, imputable à la variabilité minéralogique entre sites, et proposent de coupler les données spectroradiométriques de laboratoire avec des images hyperspectrales pour réduire les erreurs introduites par le krigeage des spectres. Cette difficulté à prédire les FeOx et le SOC par spectroscopie seule est également mise en évidence par SAHWAN et al. (2021), qui montrent que, dans des sols alcalins du nord de la Jordanie, les FeOx dominent la signature spectrale (notamment dans les gammes 450-600nm et 600-750nm), masquant souvent l'influence du SOC, dont les teneurs restent faibles (<1.9%). Leurs résultats indiquent une corrélation exponentielle forte (R² = 0.86-0.95) entre l'indice de rougeur RI_{LAB-BT} (dérivé des valeurs CIELAB) et la raideur des courbes sigmoïdales des spectres, reflétant l'impact des FeOx sur la réflectance, tandis que le SOC

et les carbonates jouent un rôle secondaire. Ces observations confirment que les FeOx, en raison de leur absorption caractéristique (notamment dans le visible), peuvent éclipser le signal du SOC, surtout lorsque ce dernier est présent en faibles concentrations.

Pour pallier cette limitation, WIESMEIER et al. (2012) préconisent d'intégrer les teneurs en FeOx comme co-variables dans les modèles prédictifs du SOC, plutôt que de se fier uniquement aux données spectrales. Leur étude démontre que les FeOx, qu'ils soient cristallins ou amorphes, stabilisent mécaniquement et chimiquement le SOC au sein des agrégats, via des mécanismes de complexation et de protection physique. En combinant les FeOx avec d'autres propriétés pédologiques (texture, pH, topographie), ils obtiennent des modèles plus robustes et spatialement cohérents, avec une réduction significative des erreurs de prédiction du SOC. Cette approche met en lumière le lien fonctionnel entre FeOx et SOC: les oxydes de fer ne sont pas seulement des marqueurs spectraux, mais des acteurs clés de la dynamique du carbone, dont l'inclusion dans les modèles renforce leur pertinence écologique. Ainsi, bien que la spectroscopie seule puisse peiner à distinguer le SOC des FeOx dans certains contextes (notamment en présence de carbonates ou de faibles teneurs en matière organique), son couplage avec des données géochimiques et des co-variables pédologiques permet de mieux capturer les interactions complexes entre ces composants. Ces travaux soulignent l'importance d'une approche multimodale, associant spectroscopie, analyses géochimiques et méthodes géostatistiques, pour une cartographie précise du SOC dans des environnements agricoles hétérogènes.

Télédétection optique pour prédire la teneur en SOC

La télédétection optique, qu'elle soit multispectrale ou hyperspectrale, offre un potentiel considérable pour la cartographie du SOC, mais son efficacité dépend fortement des stratégies d'échantillonnage et de la représentativité spectrale des données utilisées. CASTALDI et al. (2019), travaillant sur des sols agricoles en Allemagne, ont comparé plusieurs algorithmes de sélection d'échantillons (cLHS, Kennard-Stones, Puchwein, SELECT) appliqués à des données S-2A/B et EnMAP (simulé). Leurs résultats révèlent que les algorithmes Puchwein et Kennard-Stones, appliqués aux données S2-A, offrent les meilleures performances pour prédire le SOC, avec des NRMSE inférieures à 9%. Cependant, malgré la résolution spectrale supérieure d'EnMAP, sa résolution spatiale plus grossière (30m contre 10m pour S2-A) limite sa capacité à capturer la variabilité fine du SOC, surtout dans des paysages fragmentés. Les auteurs soulignent que la distance spectrale entre échantillons est fortement corrélée aux différences de teneur en SOC ($R^2 = 0.92$), confirmant que les données spectrales peuvent guider la sélection d'échantillons couvrant une large gamme de valeurs. Cette étude met en lumière l'importance de combiner résolution spatiale et spectrale pour optimiser les campagnes de terrain, tout en évitant les biais liés à une répartition non représentative des points d'échantillonnage. Dans une approche complémentaire, URBINA-SALAZAR et al. (2023) ont exploité des mosaïques temporelles d'images Sentinel-2 (2016-2021) pour spatialiser le SOC dans la Beauce, en filtrant les sols humides ($\leq 20\%$ vol.) et les zones végétalisées ($NDVI < 0.3$, $NBR2 < 0.175$). Leurs résultats, bien que modestes ($R^2 \leq 0.33$), montrent que les mosaïques construites à partir de conditions de sol nu optimales (périodes sèches, absence de résidus) améliorent la prédiction, surtout lorsque des co-variables environnementales (topographie, thorium) sont intégrées. Les auteurs recommandent d'ajuster les seuils spectraux de détection du sol nu en fonction de l'humidité et de la rugosité, et suggèrent que l'intégration de données hyperspectrales pourrait affiner la discrimination des sols et la prédiction du SOC.

Les images satellitaires S2-A seules offrent également des perspectives prometteuses pour l'estimation des propriétés physico-chimiques des sols à l'échelle des agrosystèmes. VAUDOUR et al. (2019) ont exploité ces données pour prédire la teneur en SOC, le pH, la teneur en carbonate de calcium $CaCO_3$, la CEC, ainsi que les teneurs en argile et en fer des horizons de surface dans des contextes contrastés: la plaine de Versailles (climat tempéré) et le bassin versant de la Peyne (climat méditerranéen). Leur méthodologie combine des indices de végétation (NDVI, SAVI, OSAVI) et des indices liés au sol (BI, CI, RI), ainsi que des échantillons de terrain analysés en laboratoire pour la calibration et la validation des modèles. Grâce à une PLSR, les résultats montrent que les propriétés du sol sont bien estimées dans la plaine de Versailles, avec des R^2_{CV} supérieurs ou égaux à 0.5 et des RDP supérieurs ou égaux à 1.4 pour le SOC, le pH, le $CaCO_3$ et la CEC. En revanche, la teneur en argile présente des performances légèrement inférieures ($0.5 > R^2_{CV} > 0.39$; $1.4 > RDP > 1.3$). Dans le bassin versant de la Peyne, les estimations sont plus incertaines, notamment pour la teneur en argile, en fer et la CEC. Ces résultats soulignent l'efficacité des images Sentinel-2 pour la CNS, tout en mettant en évidence les défis liés à la variabilité des conditions pédoclimatiques. L'étude d'URBINA-SALAZAR et al. (2021) dans le Sud-Ouest de la France vise à évaluer l'impact des cartes d'humidité dérivées des données S1/2, ainsi que des décalages temporels entre les dates d'acquisition des images et les dates d'échantillonnage, sur la précision des estimations du SOC. En utilisant PLSR, les auteurs ont spatialisé la teneur en SOC à partir d'échantillons de terrain et du calcul d'indices spectraux, tels que le NDVI et le NBR2, pour discriminer les sols nus. Leurs résultats montrent que l'intégration de cartes d'humidité dérivées des données S1/2 améliore significativement la précision des estimations de SOC, en particulier lorsque les conditions d'humidité varient entre les dates d'échantillonnage et d'acquisition des images.

Cependant, ils soulignent également que plus le décalage temporel entre les dates d'échantillonnage et l'acquisition des images est important, plus l'erreur d'estimation du SOC augmente. Ces observations mettent en lumière l'importance de la synchronisation des données terrain et satellitaires pour optimiser la précision des modèles prédictifs. Pour optimiser la cartographie du SOC dans les terres arables, VAUDOUR et al. (2021) ont exploré différentes méthodes de mosaïquage temporel d'images S2-A. Leur objectif était d'évaluer comment la combinaison de multiples acquisitions à différentes dates peut affiner les estimations du SOC, en intégrant les variations saisonnières et les conditions de surface, tout en surmontant les contraintes liées à la nébulosité. Les auteurs ont travaillé sur les mêmes sites que dans leurs travaux précédents, à savoir la plaine de Versailles et le bassin versant de la Payne, en calculant des indices spectraux tels que le NDVI et le NBR2 pour plusieurs années, saisons ou seuils, afin de discriminer les sols nus. La spatialisation de la teneur en SOC a été réalisée par régression par PLSR, en établissant des relations entre les signatures spectrales des mosaïques et les mesures de SOC issues des échantillons terrain. Leurs résultats montrent que les mosaïques temporelles permettent d'augmenter significativement la surface de sol nu détectée, passant de 30-40% avec des images individuelles à plus de 80% avec des mosaïques annuelles. De plus, l'estimation du SOC s'en trouve améliorée, avec un R^2 de 0.65 et un RPD supérieur à 1.5. Ces résultats soulignent l'efficacité des approches de mosaïquage temporel pour optimiser la cartographie des propriétés des sols à partir de données satellitaires.

L'intégration de données multisources – combinant imagerie satellitaire, spectres de laboratoire et co-variables pédologiques – émerge comme une stratégie prometteuse pour améliorer la précision des modèles de SOC. ZAYANI 2023 a évalué l'apport successif de données S2-A, S-1C et de spectres de laboratoire pour prédire le SOC à deux profondeurs (0-15cm et 15-25cm). Ses résultats montrent que la PLSR appliquée aux spectres de laboratoire atteint des performances élevées (RMSE = 4.1g/kg), mais que l'intégration des données S2-A (via des modèles RF ou DNN) améliore significativement les prédictions, avec un RPIQ passant de 1.36 (S2-A seul) à 2.01 (fusion S2-A + labo). Le DNN surpasse systématiquement la PLSR et le RF, surtout lorsque les données sont fusionnées, soulignant l'intérêt des approches non-linéaires pour capturer les interactions complexes entre SOC, humidité et propriétés spectrales. FOUAD et al. (2023) confirment cette tendance en montrant que, bien que la combinaison des bandes S2-A, indices spectraux (NDVI, NBR2) et humidité (S1-C) optimise la calibration ($R^2 = 0.9$), la validation révèle que les modèles les plus simples (bandes S2-A seules) peuvent parfois performer davantage ($R^2 = 0.65$), en raison de biais d'échantillonnage ou de conditions non représentatives dans les séries temporelles. Ces études soulignent l'importance de diversifier les sources de données et de valider les modèles sur des jeux de test indépendants pour éviter le surapprentissage. CASTALDI et al. (2023) ont comparé l'efficacité de différents mosaïquages temporels (maxBSI, minS2WI, médian, R90) et modèles (RF à quantiles, LMEM) pour prédire le SOC et l'argile dans des terres cultivées européennes. Leurs résultats indiquent que les mosaïques "médianes" et R90 (excluant les valeurs extrêmes de réflectance) offrent les meilleures performances, surtout lorsque les modèles intègrent une composante spatiale (QRFxy, LMEM), avec des R^2 atteignant 0.96 pour certains sites. Cette étude renforce l'idée que la construction rigoureuse des séries temporelles (filtres de couverture nuageuse, sélection des périodes de sol nu) et l'utilisation de modèles spatio-temporels sont essentielles pour maximiser la précision des prédictions.

Malgré ces avancées, plusieurs défis persistent, notamment en ce qui concerne la transférabilité des modèles et la variabilité des conditions pédoclimatiques. WANG et al. (2024), dans une revue sur l'utilisation de l'hyperspectral (HRS) pour le suivi des métaux lourds, soulignent que les performances des modèles ($R^2 = 0.52-0.94$) dépendent fortement des prétraitements spectraux (CR, SNV, MSC) et des algorithmes (PLSR, RF, DNN). Ils notent que la fusion de données multi-sources (laboratoire, terrain, satellite) améliore la robustesse, mais que la complexité des interactions sol-métaux-végétation limite la généralisation des résultats. Cette observation est corroborée par CHLOUVERAKI et al. (2025), qui montrent que, bien que l'HRS (470-900nm) prédise efficacement les FeOx ($R^2 = 0.57$) et les carbonates ($R^2 = 0.74$), le SOC reste difficile à estimer en raison de faibles teneurs et d'interférences avec d'autres composants (carbonates). Les auteurs recommandent d'explorer des méthodes avancées comme l'AutoML et les auto-encodeurs pour optimiser la sélection de caractéristiques et réduire le surapprentissage. Ces études convergent vers plusieurs pistes d'amélioration, dont l'affinage de la sélection des périodes d'acquisition (sol nu, conditions hydriques optimales), la combinaison de données spectrales et co-variables pédologiques (FeOx, texture, pH) pour capturer les interactions complexes, l'utilisation de modèles non-linéaires (DNN, RF) et de techniques de réduction de dimensionnalité (PCA, UMAP) pour exploiter pleinement le potentiel des données hyperspectrales, ainsi que la validation des modèles sur des jeux de données indépendants et multi-sites pour évaluer leur transférabilité.

Revue antérieures et lacunes

Les recherches sur la gestion des sols viticoles, notamment l'estimation du SOC et des FeOx par télédétection et apprentissage automatique, ont connu des avancées significatives, mais des lacunes majeures persistent, limitant leur application opérationnelle. Les études antérieures se sont principalement concentrées sur l'utilisation

d'imagerie satellitaire multispectrale (Sentinel-2, Landsat) et de données hyperspectrales pour cartographier le SOC, en combinant des algorithmes de régression (PLSR, RF, SVR) et des techniques de réduction de dimension (PCA, UMAP). VAUDOUR et al. (2022), dans leur synthèse des approches spectrales appliquées aux zones cultivées, soulignent que les études mobilisent principalement des capteurs multispectraux (Landsat, Sentinel-2, SPOT) et hyperspectraux (Hyperion, EnMAP), ainsi que des données radar (Sentinel-1) ou complémentaires (MODIS, Sentinel-3). Bien que ces capteurs offrent une couverture spatio-temporelle inégale, leur utilisation se heurte à des biais méthodologiques persistants. Par exemple, l'extraction des zones de sol nu repose presque exclusivement sur des seuils d'indices végétaux (NDVI, NBR2), une approche qui peut sous-estimer la variabilité intra-parcellaire en raison de la sensibilité de ces indices aux résidus de culture ou à l'humidité résiduelle. Les méthodes analytiques de référence pour mesurer le SOC (combustion sèche dans 50 % des cas, oxydation humide dans 35 %) introduisent également des incertitudes, notamment lorsque les protocoles ne sont pas harmonisés entre études.

De manière générale, ces approches se heurtent à plusieurs défis: la variabilité spatio-temporelle des signaux spectraux, influencée par l'humidité, la texture du sol et la couverture végétale, qui introduit des biais dans les modèles prédictifs (LI et al., 2024; LIMA et al., 2025); le manque de données de référence à haute résolution, notamment pour les sols viticoles, où les variations locales de SOC et de FeOx sont mal capturées par les capteurs satellitaires existants qui ont une résolution spatiale supérieure ou égale à 10m (DVORAKOVA et al., 2020); l'absence de protocoles standardisés pour l'intégration des données multi-sources (laboratoire, drone, satellite), ce qui limite la transférabilité des modèles entre régions et contextes pédoclimatiques (LI et al., 2024). Les algorithmes de modélisation les plus employés – PLSR, RF, SVM ou GTB – montrent des performances variables selon les contextes pédoclimatiques, avec des R^2 généralement compris entre 0.5 et 0.8 pour les régions tempérées (Europe, Amérique du Nord, Chine). Cependant, ces modèles se concentrent majoritairement sur des sols arables tempérés (cambisols, luvisols), où les teneurs médianes en SOC ($\approx 15\text{g/kg}$) masquent souvent la variabilité locale critique pour les agrosystèmes spécialisés comme les vignobles. Par ailleurs, les interactions entre le SOC et les FeOx, bien documentées en laboratoire, restent difficiles à quantifier à l'échelle du paysage en raison du chevauchement spectral entre les signatures du SOC, des FeOx et de l'eau dans le VNIR-SWIR (PEARLSHTIEN et al., 2020). Les études récentes soulignent aussi le besoin de méthodes robustes pour distinguer les fractions de SOC stabilisées par les FeOx (ex: SOC lié aux ferrihydrites) des fractions libres, afin d'améliorer les estimations de stock de carbone et de prédire leur vulnérabilité face aux changements climatiques (ODEBIRI et al., 2024; SÁNCHEZ-MORENO et al., 2025). Peu d'études intègrent des modèles de *deep learning*, explorent la variabilité saisonnière du SOC, ou évaluent l'impact des pratiques agricoles (labour, couverts végétaux) sur sa dynamique. Pourtant, ces facteurs sont cruciaux pour comprendre comment les changements climatiques (sécheresses, canicules, pluies diluviennes) affectent la résilience des sols et leur capacité à stocker du carbone. Enfin, la fusion de données multi-capteurs (SAR/hyperspectral ou LiDAR/spectral) reste marginalement exploitée, bien qu'elle puisse potentiellement améliorer la résolution spatiale et la précision des prédictions en capturant des propriétés complémentaires du sol (rugosité, humidité, texture).

En France, les recherches en CNS reflètent ces mêmes défis, tout en mettant en avant des spécificités locales liées à la diversité des contextes pédologiques et agricoles. RICHER-de-FORGES et al. (2023) soulignent que les données de télédétection y sont principalement utilisées comme substitut aux mesures terrain, permettant une couverture géographique étendue et une fréquence temporelle accrue par rapport aux campagnes pédologiques classiques. Les propriétés les plus étudiées – humidité, rugosité, SOC, texture, couleur des sols – répondent à des enjeux majeurs comme la perte de carbone organique, l'imperméabilisation, l'érosion, ou la salinisation due à une irrigation excessive. Cependant, malgré ces avancées, plusieurs obstacles persistent. D'une part, le manque de standardisation dans l'acquisition et le prétraitement des données limite la comparabilité et la transférabilité des modèles entre régions. Par exemple, les indices spectraux (NDVI, NBR2) ou les méthodes de sélection des pixels de sol nu varient d'une étude à l'autre, introduisant des biais dans les estimations de SOC. D'autre part, les interactions complexes entre le SOC, les FeOx et d'autres propriétés du sol (argiles, carbonates) sont rarement prises en compte de manière intégrée dans les modèles prédictifs. Pourtant, ces interactions influencent directement la stabilité du SOC et sa détectabilité par télédétection, notamment dans les sols viticoles où les oxydes de fer jouent un rôle clé dans sa stabilisation (WIESMEIER et al., 2012). Enfin, bien que les données multi-sources (satellite, drone, laboratoire) offrent un potentiel considérable pour affiner les prédictions, leur intégration reste limitée par des contraintes techniques (résolutions spatiales et spectrales incompatibles) et méthodologiques (absence de protocoles unifiés pour la fusion de données). Ces lacunes soulignent la nécessité de développer des approches hybrides, combinant télédétection, analyses de laboratoire et modélisation avancée (*deep learning*, fusion SAR/hyperspectral), pour surmonter les défis actuels et produire des cartes de SOC robustes et transférables à différentes échelles.

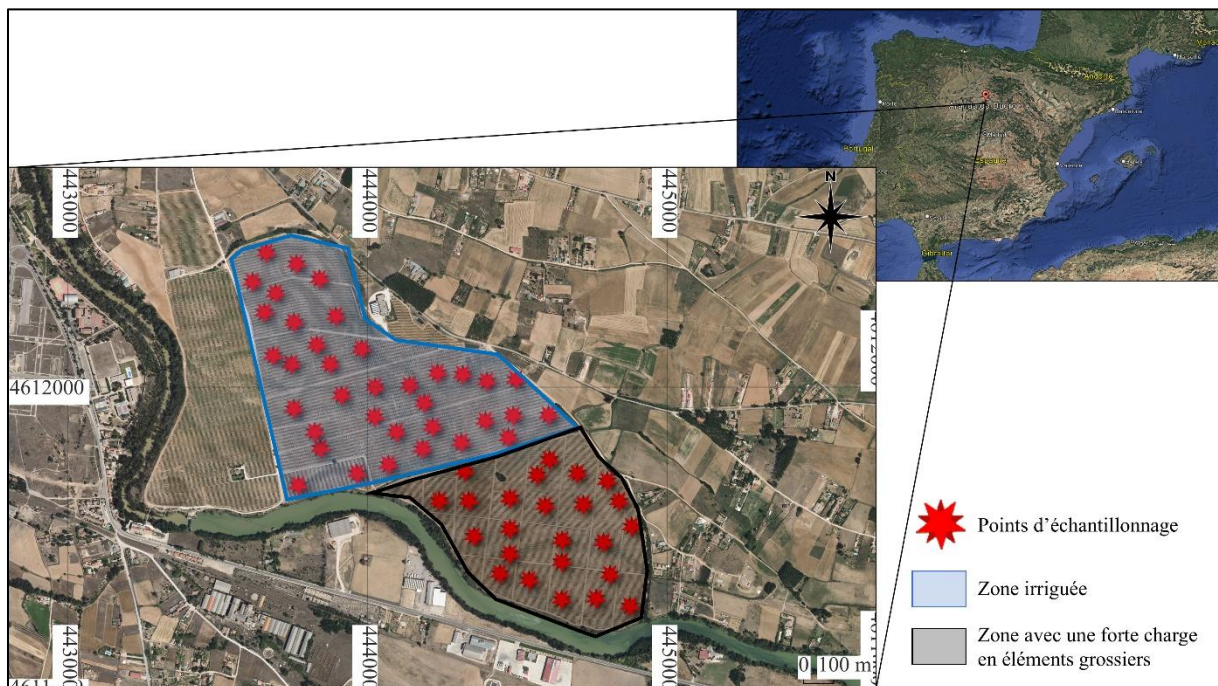
Chapitre 2 – Données & méthodes

Site d'étude

Contexte général et localisation

Le centre de l'Espagne présente un climat méditerranéen à semi-aride, avec des températures moyennes estivales de 25-30°C et un cumul annuel de précipitations ne dépassant pas 350 à 400mm. Les sols agricoles, notamment ceux des vignobles et oliveraies, sont fortement impactés par les phénomènes d'érosion et sont en conséquence très dégradés (STRAFFELINI et al., 2024) par ces faibles pluies. La compétition en eau induite pour les végétaux, et par conséquent la productivité et la durabilité des exploitations viticoles et oléicoles, est un enjeu central des systèmes agricoles en environnement semi-aride. Compte tenu de ce climat peu favorable à l'agriculture, l'irrigation des vignobles et oliveraies est autorisée, à l'inverse de la France. Le travail du sol dans les inter-rangs y est donc très important – tant en fréquence qu'en intensité – de sorte à limiter la croissance des adventices et à éviter la compétition hydrique au droit de la vigne et de l'olivier. Cette approche, largement partagée par les agriculteurs locaux, vise à garantir une production continue malgré un contexte défavorable, mais repose sur un travail du sol intensif. Labourage traditionnel, scarification, passage du rouleau, taille mécanisée: autant d'opérations qui ont pourtant des impacts négatifs démontrés sur les dynamiques physico-chimiques du système pédologique. Il y a un réel enjeu à bien comprendre et modéliser ces processus du sol afin de pouvoir faire le lien avec les différentes conduites, et de faire évoluer les pratiques en faveur d'une agriculture durable, respectueuse de l'écosystème pédologique – mais qui demeure économiquement viable pour les agriculteurs.

Le site d'étude est un vignoble situé près d'Aranda de Duero, dans la province de Burgos (Castille-et-León, Espagne). Ce secteur se caractérise par un climat méditerranéen semi-aride avec moins de 400mm de cumul annuel de précipitations et une température moyenne estivale supérieure à 20°C (cf. Annexe 1), classé selon KÖPPEN entre Csb (tempéré avec été sec) et Bsh (semi-aride chaud). Il y a une influence topographique marquée due à la proximité du rio Arlanzón. Le vignoble s'étend sur 82 hectares, planté en 2004, et présente des vignes d'une hauteur de 1.5 à 1.8m. Depuis 2006, les inter-rangs sont maintenus avec un couvert végétal, et la partie Nord du site est irriguée. Une carte du site (cf. Figure 1) illustre la répartition des 56 points d'échantillonnage utilisés pour les mesures terrain. L'altitude varie entre 787 et 813m (altitude moyenne de 800m), avec des pentes comprises entre 0 et 14° (moyenne de 3.3°). Les sols, principalement des Calcisols à Fluvisols (cf. Annexe 2), sont représentatifs des agroécosystèmes méditerranéens semi-arides, avec des teneurs notables en oxydes de fer.



Données terrain et caractéristiques pédologiques

Les échantillons de sol ont été constitués à partir de composites de 10 prélèvements à une profondeur de 0-5cm, sur différentes campagnes de prélèvements (cf. Figure 2). Les analyses physico-chimiques incluent la mesure de la teneur en SOC, en fer libre total, en FeOx (Fe²⁺ et Fe³⁺), la granulométrie (argile, limon, sable) et la charge en éléments grossiers. La teneur en SOC est mesurée par la méthode WALKLEY-BLACK. La teneur en Fe³⁺ est calculée par soustraction de la teneur en Fe²⁺ (mesurée par la méthode d'extraction à la 1,10-phénanthroline de TARAFDER-THAKUR) à la teneur en fer total libre (mesurée par la méthode d'extraction au CBD de MEHRA-JACKSON). Les pratiques agricoles (dates, types d'interventions) sont systématiquement enregistrées et cartographiées, permettant une analyse spatio-temporelle fine des interactions entre gestion des sols et propriétés pédologiques (ZAYANI et al., travaux non-publiés).

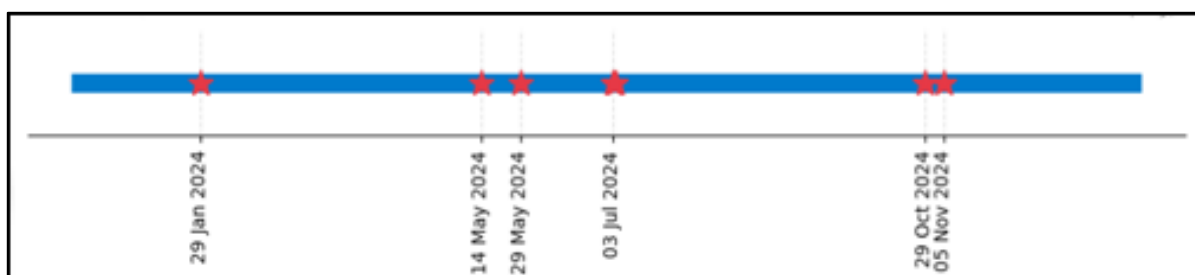


Figure 2 – Frise chronologique des campagnes de prélèvements terrain sur le site d'étude, emprunté à ZAYANI et al., (travaux non-publiés); les échantillons d'Aranda ont été prélevés lors des campagnes de janvier et mars 2024

Données de télédétection

Séries temporelles Sentinel-2

Une série temporelle Sentinel-2 est constitué sur la période 2022-2024 à partir des images des satellites S-2A et S-2B couvrant le site d'étude (tuile T30TVM). Pour se faire, toutes les images disponibles du 01/01/2022 au 31/12/2024 sont téléchargées depuis les serveurs du CNES via la nouvelle API GEODES, qui remplace l'ancien service PEPS. Comme l'ensemble de la chaîne de traitement des données satellitaires a été codée en langage Python, l'API GEODES est intégrée dans la chaîne via son interface Python *pygeodes*.

Cela constitue ainsi une bande de 238 images, qu'il faut ensuite filtrer selon le taux de nébulosité sur le vignoble. Pour limiter les effets de bords sur les pourtours du vignoble, un nouveau périmètre d'étude est créé à partir des limites du vignobles, auxquelles est ajoutée une zone tampon de 100m (i.e. 10 pixels de 10m, résolution native des bandes visibles de Sentinel-2). Les images et leur masque de nuages sont découpés avec ce périmètre d'étude, afin de ne couvrir que le vignoble et non plus l'entièreté de la tuile T30TVM. Cela réduit la banque de 238 images à 109 images. Durant cette même opération, les bandes B11 et B12 sont ré-échantillonnées à 10m (même résolution que les bandes VNIR) par méthode des plus proches voisins.

L'activité biologique de la vigne, maximale pendant l'été et jusqu'en septembre (cf. Figure 3), perturbe fortement le signal spectral enregistré par le satellite, à cause d'une couverture foliaire très dense. Il est donc nécessaire de filtrer la série temporelle pour ne garder que les images pendant la période de dormance de la vigne, du 31 octobre au 28 février de l'année suivante. Cela permet de mieux isoler la contribution spectrale du sol et ainsi limiter les biais liés à la variabilité phénologique de la canopée dans l'estimation des propriétés pédologiques. Cela réduit la bande de 109 images à 31 images.

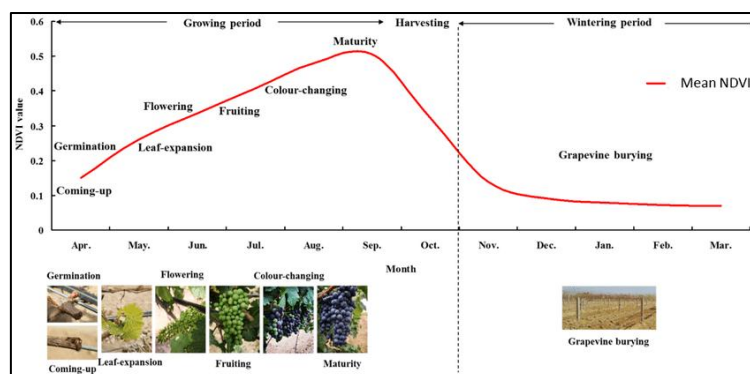


Figure 3 – Courbe du NDVI de la vigne en climat méditerranéen, LI et al. (2021)

La répartition des images par mois dans la banque finale est donnée en Annexe 3.

Comme montré par la revue de la littérature (cf. Chapitre 1), le maintien de couverts végétaux dans les inter-rangs du vignoble depuis 2006 est bénéfique à bien des égards; mais cela brouille la signature du sol, et donc de nos variables d'intérêt, dans la réponse spectrale capturée par le satellite. Il convient donc de filtrer les pixels en sol nu du total de pixels de chaque image. Pour cela, plusieurs méthodes ont déjà été testées dans la littérature, avec deux approches dominantes. Une première approche consiste en un seuillage d'indices ou de couplages d'indices, dont le couplage NDVI+NBR2 largement utilisé dans la littérature, notamment par CASTALDI et al. (2019) et dans tous les travaux du tandem URBINA-SALAZAR et VAUDOUR et al. (2021, 2022, 2023). Le seuillage d'indices a l'avantage d'être léger en termes d'équations et de ressources de calcul. Le BSI (DIEK et al., 2017) est aussi un indice intéressant car spécifiquement calibré pour la discrimination des pixels en sol nu, mais a déjà montré des limites par le passé en fonction du contexte pédoclimatique et se retrouve souvent dépassé par le couplage NDVI+NBR2, comme montré par CASTALDI et al. (2019). Une deuxième approche se fait par série temporelle d'indice et surveillance des pixels passant sous un seuil mobile sur la série temporelle, dont les indices PV+Blue de ROGGE et al. (2018) et PV+IR2 de ZEPP et al. (2023). L'utilisation d'un seuil mobile sur une série temporelle permet de détecter dynamiquement les pixels en sol nu en s'affranchissant des variations saisonnières et interannuelles de la réflectance, ce qui améliore la robustesse de la discrimination en intégrant la variabilité phénologique et les conditions locales de surface. En tenant compte de la variabilité phénologique, la distribution temporelle d'un pixel en sol nu est aussi plus constante dans le temps.

Ici, deux méthodes ont été testées: (1) le seuillage strict par couplage NDVI+NBR2, et (2) le seuillage mobile par PV+Blue et PV+IR2. Si le nombre de dates en sol nu pour chaque point d'échantillonnage augmente avec la méthode 2, la définition du seuillage mobile (d'abord reprise *stricto sensu* de ZEPP et al., 2023) reste floue et difficilement adaptable au contexte pédoclimatique de ce vignoble. Une première approche fût de collecter la valeur minimale de l'indice (PV+Blue et PV+IR2) pour chaque date d'acquisition Sentinel-2, puis de considérer le 5^e percentile des minima d'une période de dormance comme valeur de seuillage des pixels en sol nu pour cette même période de dormance. Ce seuil s'avéra beaucoup trop restrictif, avec une majorité de pixels réellement en sol nu classés comme étant couverts. Une deuxième approche fût d'utiliser la valeur médiane de l'indice pour une période de dormance comme valeur de seuillage pour cette même période de dormance. Les résultats furent beaucoup plus exhaustifs qu'avec la première approche, mais restaient en-dessous de ceux obtenus avec la méthode 1. C'est donc la méthode 1, le seuillage strict par couplage NDVI+NBR2 qui est retenue. Après plusieurs tests, et en lien avec ceux recensés dans la littérature (CASTALDI et al., 2019), les pixels en sol nu sont filtrés lorsqu'ils respectent conjointement les deux valeurs de seuil, à savoir un NDVI inférieur à 0.3 et un NBR2 inférieur à 0.175.

À partir de cette banque de 31 images en période de dormance de la vigne de 2022 à 2024, et aux pixels en sol nu filtrés, sont calculés une batterie d'indices spectraux (cf. Annexe 4). Ils sont pour la majorité repris de la thèse de ZAYANI 2023 car particulièrement adaptés aux sols agricoles en climat méditerranéen semi-aride. D'autres sont sourcés depuis des articles de chercheurs ayant travaillé dans le même contexte pédoclimatique, ou avec des méthodes similaires à celles utilisées ici pour la cartographie du SOC et des FeOx. C'est le cas notamment des travaux d'ESCADAFAL (1989), MADEIRA et al. (1997), MATHIEU et al. (1998), BULLARD et al. (2002), BARTHOLOMEUS et al. (2007), HAUBROCK et al. (2008), PENG et al. (2015) et SAHWAN et al. (2020). Certains indices, habituellement utilisés en imagerie hyperspectrale (comme EnMAP), sont adaptés en fonction de la partie intégrée du spectre pour chaque bande des capteurs Sentinel-2 (cf. Annexe 5).

Spectres laboratoire d'échantillons terrain

Les échantillons terrain sont tamisés à 2mm, séchés à l'étuve pendant 48h, puis préparés en boîtes de Pétri. Leurs spectres sont acquis à l'aide du spectroradiomètre de terrain ASD FieldSpec3® équipé d'une *ContactProbe* contenant une source de lumière connue. Le spectre couvre la plage de 350 à 2'500nm selon un pas de 1nm, et est calculé par moyenne de 25 primo-spectres acquis chacun sur les 4 points cardinaux de la boîte de Pétri de chaque échantillon. Une référence blanche pour le spectroradiomètre est acquise sur Spectralon® tous les 3 échantillons à mesurer. Tout ceci est effectué en conditions contrôlées de laboratoire: chambre noire à température et hygrométrie constante, conditions d'éclairage fixes et connues, position fixe et connue de la *ContactProbe*. Des photos du montage est disponible dans l'Annexe 6.

Les spectres ainsi acquis sont appelés par la suite « spectres labo ».

Conformément à la littérature, notamment les articles de PINHEIRO et al. (2017), XU et al. (2021) et MIRZAEI et al. (2022), l'application de pré-traitements aux spectres labo permet de mieux mettre en évidence la relation entre la réflectance du sol et la teneur en SOC et FeOx. Cela permet entre autres de réduire l'influence de l'humidité ou de résidus végétaux sur la réponse spectrale de l'échantillon. Aussi, les spectres labo subissent les pré-traitements suivants:

- *Raw*, version brute de la réflectance relative (R) du spectre labo
- *Abs*, conversion du spectre labo en pseudo-absorbance (A) avec $A = \log_{10}(1/R)$
- *SG*, lissage du spectre avec le filtre de SAVITZKY-GOLAY 1964, soit un noyau polynomial pour réduire le bruit et accentuer les extrema locaux
- *SNV*, centrage-réduction du spectre pour supprimer les effets de diffusion et de gain multiplicatif
- *MSC*, alignement du spectre sur une moyenne de référence pour réduire les effets de diffusion multiplicative
- *CR*, application du continuum-removal selon CLARK-ROUSH 1984, soit une normalisation du spectre par enveloppe convexe interpolée sur les maxima locaux pour isoler la forme des bandes d'absorption indépendamment de l'albédo global

Dans le cadre de la prédiction de la teneur en FeOx, différentes études (BARTHOLOMEUS et al., 2008; ZHAO et al., 2023) ont démontré que l'ajout d'informations issues des caractéristiques spectrales, notamment les paramètres géométriques des bandes d'absorption, permet d'améliorer significativement la précision des modèles. Les paramètres tels que la profondeur, la largeur à mi-hauteur ou l'asymétrie des bandes d'absorption, en particulier celles associées au Fe^{3+} , apportent une information complémentaire essentielle pour mieux prédire la teneur en FeOx dans les sols. Ces paramètres, calculés à partir du signal spectral corrigé par CR, reflètent la forme et l'intensité des absorptions spécifiques liées aux oxydes de fer. Ainsi, les valeurs de paramètres géométriques (cf. Annexe 7) de la courbe spectrale corrigée sont calculés au droit des bandes d'absorption de l'oxyde de fer (III), pour être utilisées comme co-variables additionnelles car bien corrélés avec la teneur en FeOx

Réduction de la dimensionnalité des variables prédictives

Les jeux de données spectraux, qu'ils proviennent de mesures de laboratoire ou de séries temporelles satellitaires, se caractérisent par un grand nombre de variables (longueurs d'onde et/ou indices spectraux) souvent hautement corrélées entre elles. Cette multicollinéarité peut non seulement biaiser les estimations des coefficients dans les modèles prédictifs, mais aussi dégrader leurs performances en introduisant du bruit et en augmentant la complexité computationnelle. Par ailleurs, la haute dimensionnalité des séries temporelles (*nombre d'indices* \times *nombre de dates*) et des spectres labo rend les modèles plus sensibles au sur-apprentissage, tout en compliquant leur interprétabilité. Dans ce contexte, la réduction de la dimensionnalité est une étape importante pour extraire l'information pertinente tout en limitant la redondance. Cela permet de conserver les structures latentes liées au SOC et aux FeOx, tout en améliorant l'efficacité et la généralisation des modèles. L'application d'une AFM suivie d'une ACP (cf. Figure 4) transforment les variables initiales en un ensemble réduit de « composantes principales » orthogonales, facilitant ainsi l'identification des signaux les plus discriminants.

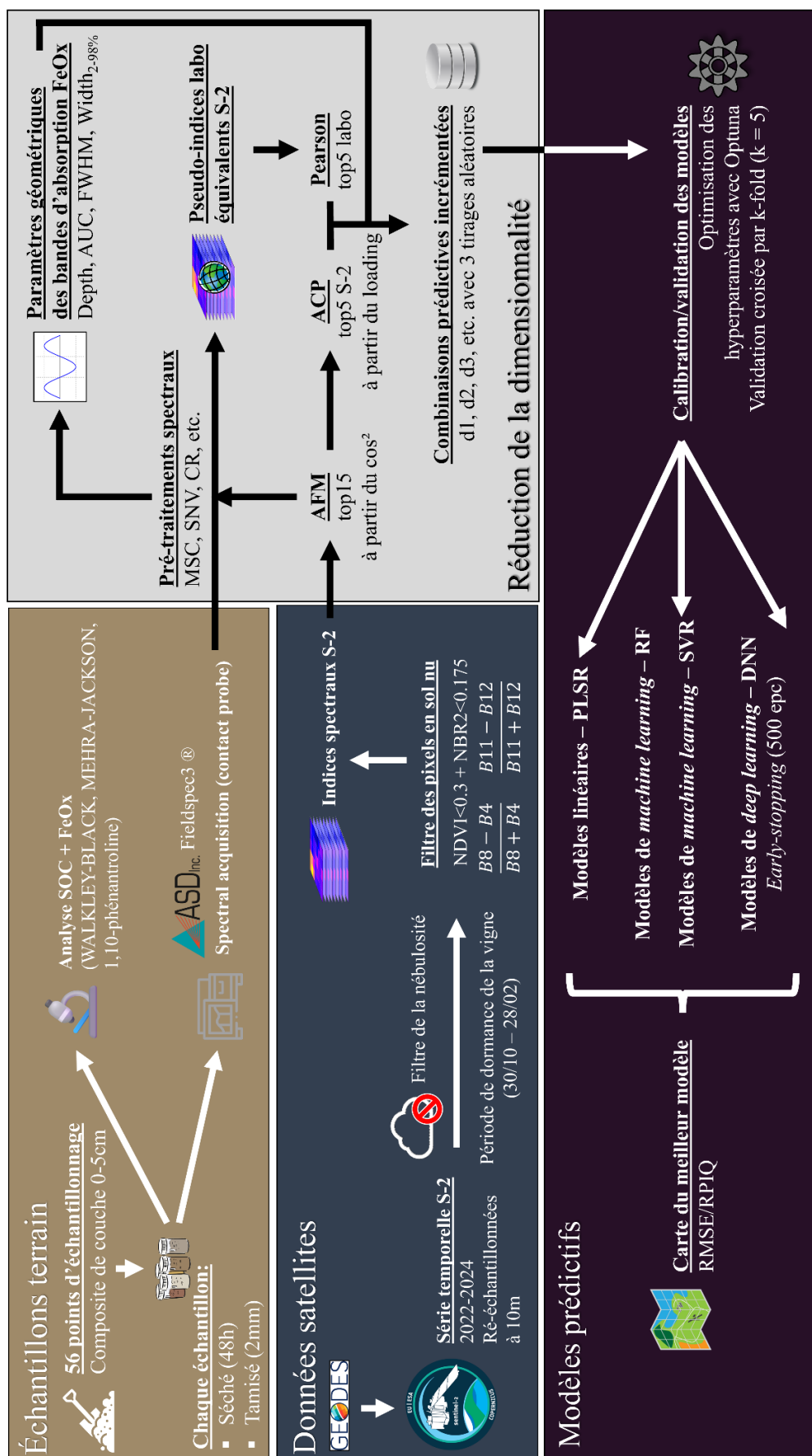


Figure 4 – Diagramme de traitements des données pédologiques et de télédétection

Calcul et spatialisation des pseudo-indices de laboratoire équivalents S-2

Afin d'enrichir les modèles prédictifs, des pseudo-indices spectraux équivalents aux indices Sentinel-2 ont été calculés à partir des spectres labo, les indices à calculer correspondant à ceux du top15 en sortie d'AFM. Comme les spectres de laboratoire ont été pré-traités au préalable, ces indices sont qualifiés de « pseudo-indices » : ils ne sont pas calculés directement à partir des valeurs de réflectance brutes, mais à partir de spectres transformés pour maximiser leur pertinence prédictive (sauf pour les spectres *Raw*). Bien que l'information spectrale brute des échantillons de sol contienne déjà l'ensemble des signatures spectrales et que les indices n'en soient que des dérivés, le calcul de ces pseudo-indices offre une représentation synthétique et standardisée des caractéristiques spectrales, souvent plus robuste pour la prédiction des propriétés du sol que les spectres bruts, notamment en réduisant la redondance et le bruit (STEVENS et al., 2010). Contrairement aux spectres Sentinel-2, non-homogènes en raison de la mixité des signatures spectrales au sein d'un pixel, les spectres de laboratoire sont acquis dans des conditions contrôlées, ce qui permet une analyse plus robuste de leur relation avec le SOC et les FeOx. Ainsi, une sélection supplémentaire a été effectuée en triant les 15 indices selon leur p_{value} issue d'un test de corrélation linéaire de Pearson avec la variable cible (SOC ou FeOx). Seuls les 5 indices présentant les corrélations les plus significatives ($p_{value} \leq 0.05$) ont été retenus, garantissant que seuls les indices labo linéairement liés à la propriété du sol étudiée sont conservés. Cette approche est rendue possible grâce à la qualité et à la précision des mesures en laboratoire, où les interférences (couverture végétale, ombre, humidité variable) sont absentes.

Pour spatialiser ces pseudo-indices à l'échelle du site d'étude, une approche géostatistique a été adoptée. Le krigeage ordinaire a été privilégié, car il permet d'estimer les valeurs des indices en tout point non échantillonné tout en tenant compte de la structure spatiale des données (GOOVAERTS, 1997). Cette méthode est particulièrement adaptée aux propriétés du sol, qui présentent souvent une autocorrélation spatiale marquée (WEBSTER & OLIVER, 2007). Le krigeage ordinaire, en s'appuyant sur les semi-variogrammes expérimentaux, minimise la variance d'estimation et fournit une cartographie continue des pseudo-indices, nécessaire pour intégrer ces variables dans les modèles prédictifs. Les 5 pseudo-indices labo sont spatialisés sur le site d'étude et co-registrés sur les indices S-2 (même emprise et résolution spatiale).

Création de combinaisons prédictives

Pour évaluer l'impact de l'ajout progressif de co-variables sur les performances des modèles prédictifs, une série de jeux de données incrémentiels a été construite pour chaque pré-traitement appliqué aux spectres labo. Cette approche permet de tester systématiquement l'apport de chaque type de variable (bandes Sentinel-2, indices Sentinel-2, pseudo-indices labo) dans la prédiction des teneurs en SOC et FeOx. Pour chaque pré-traitement, les 5 indices Sentinel-2 et les 5 pseudo-indices labo les plus pertinents ont été combinés de manière progressive. Ainsi, les jeux de données ont été construits en ajoutant un à un les indices Sentinel-2, suivis des pseudo-indices labo. Par exemple, le jeu de données d1 ne contient que les bandes spectrales Sentinel-2, tandis que d7 intègre les bandes spectrales, les 5 indices Sentinel-2 et le premier pseudo-indice labo. Le jeu de données d12 ne comporte que les bandes spectrales Sentinel-2 et le premier indice labo, afin d'étudier son apport sans les indices Sentinel-2. Son apport seul est aussi étudié avec le jeu de données d20, qui ne contient rien d'autre que ce premier indice labo. Dans le cadre du SOC, quatre jeux de données supplémentaires sont créés, de d30 à d32, afin d'intégrer progressivement les données de fer, ainsi que le jeu d33 pour étudier l'apport des données d'oxyde de fer (III) seules. Enfin, dans le cadre des FeOx, un jeu de données supplémentaire d50 est créé pour intégrer les valeurs des paramètres géométriques des courbes d'absorption des FeOx et étudier leur apport. Cette approche incrémentielle permet d'identifier quelles co-variables apportent une information complémentaire significative pour la prédiction du SOC et des FeOx, tout en évitant la redondance. Un tableau récapitulatif des jeux de données et de leur composition est donné en Annexe 8.

Afin de s'affranchir des biais liés à une répartition aléatoire défavorable des points d'échantillonnage, trois tirages indépendants de partitions train/test ont été réalisés pour chaque combinaison de variables, en utilisant des graines aléatoires distinctes fixes. Plusieurs ratios de partitionnement ont été évalués, sans impact majeur relevé, c'est donc le ratio 70/30 qui est retenu. Cette répétition permet de moyenner les performances des modèles et de vérifier leur robustesse face à la variabilité des sous-échantillons. Les statistiques descriptives (moyenne, médiane, écart-type) des teneurs en SOC et FeOx ont été calculées pour chaque partition, afin de s'assurer que les ensembles d'entraînement et de test étaient représentatifs de la distribution globale des données. Cette précaution est essentielle pour éviter les conclusions biaisées par une répartition particulièrement favorable ou défavorable des points entre les ensembles.

Une attention particulière a été portée sur l'utilisation des dates en sol nu dans les séries temporelles S-2. En retenant un maximum de dates où le sol est visible (sans couverture végétale), les modèles sont exposés à une diversité de conditions de surface (humidité, rugosité, état de la croûte superficielle) pour des mêmes valeurs de

SOC et de FeOx, considérées comme stables à l'échelle de temps de l'étude. Cette stratégie vise à renforcer la généralisation des modèles en leur permettant d'apprendre à discriminer le signal lié au SOC et aux FeOx de celui lié aux variations superficielles du sol. En effet, bien que le SOC et les FeOx soient des propriétés relativement stables dans le temps, les signatures spectrales du sol nu peuvent varier en fonction des conditions environnementales (humidité résiduelle, état de la surface). En intégrant ces variations dans les données d'entraînement, les modèles deviennent plus robustes aux artefacts liés à l'acquisition et mieux capables de prédire le SOC et les FeOx dans des contextes variés. Cette approche est pertinente pour des modèles basés sur des données satellitaires, où chaque pixel intègre des signatures spectrales potentiellement bruitées par des facteurs externes (résidus de végétation, ombres). En contrastant ces données avec les pseudo-indices labo, acquis en conditions contrôlées, les modèles peuvent mieux isoler le signal lié au SOC et aux FeOx. Cela permet donc d'identifier les combinaisons optimales de co-variables pour prédire le SOC et les FeOx, en équilibrant précision et parcimonie, et d'évaluer la complémentarité entre les indices dérivés des séries temporelles S-2 et les pseudo-indices labo, qui capturent des informations spectrales à des résolutions et des conditions d'acquisition différentes. En croisant ces approches, il devient possible de déterminer si les pseudo-indices labo apportent une information supplémentaire significative par rapport aux indices S-2, et si l'ajout des teneurs en fer améliore la prédiction en capturant des processus pédogénétiques liés à la minéralogie du sol. Les résultats de cette étape guideront le choix des variables pour les modèles prédictifs ultérieurs (PLSR, SVR, RF, DNN).

Calibration de modèles prédictifs

La calibration des modèles prédictifs a été principalement réalisée à l'aide des bibliothèques « scikit-learn » (PEDREGOSA et al., 2011) pour la construction des modèles et « Optuna » (AKIBA et al., 2019) pour l'optimisation des hyperparamètres² par recherche bayésienne. Cette approche permet d'éviter à la fois le sous-apprentissage et le sur-apprentissage, en retenant les valeurs d'hyperparamètres qui maximisent la variance expliquée tout en limitant la complexité du modèle. La calibration de chaque modèle a également été réalisée avec une validation croisée k-fold ($k = 5$) pour évaluer la robustesse des prédictions. Les modèles PLSR, SVR et DNN ont nécessité la standardisation des variables prédictives en entrée, pour garantir que chaque variable contribue équitablement à la construction des composantes latentes, indépendamment de leur échelle initiale de valeurs. Le RF est lui intrinsèquement robuste aux différences d'échelle entre les prédicteurs. La calibration du modèle DNN a été réalisée en utilisant une architecture personnalisée construite avec TensorFlow/Keras (ABADI et al., 2015/CHOLLET et al., 2015), optimisée par Optuna pour identifier la configuration la plus performante. La validation croisée est complétée par un *early stopping*, basé sur la perte de validation pour arrêter l'entraînement dès que les performances cessent de s'améliorer (sur 500 époques). Les modèles sont évalués en fin d'entraînement à l'aide du R^2 , la RMSE, le RPD et la MBE. Ces indicateurs permettent d'assurer que les modèles capturent efficacement la relation entre les variables spectrales (indices Sentinel-2, pseudo-indices labo) et la teneur en SOC et en FeOx, tout en restant généralisables à de nouveaux échantillons.

² Exemples d'hyperparamètres optimisés: nombre de composantes latentes pour la PLSR, nombre et profondeur des arbres pour le RF, fonction noyau et paramètres C et γ pour le SVR, taux d'apprentissage et régularisation pour le DNN

Chapitre 3 – Résultats

Statistiques descriptives des échantillons terrain

Les boxplots ci-dessous (cf. Figure 5) présentent la distribution des teneurs en SOC et FeOx des échantillons terrain prélevés sur le site d'étude. Le carbone organique affiche une valeur médiane de 10.64g/kg et une variabilité marquée (étendue de 2.97 à 22.97g/kg). Le fer total libre affiche valeur médiane de 0.07g/kg et une variabilité peu marquée. L'oxyde ferreux (oxyde de fer (II)) affiche une valeur médiane de 0.05g/kg et une variabilité marquée (étendue de 0.02 à 0.13g/kg) reflétant la pédogenèse spécifique des sols méditerranéens. L'oxyde ferrique (oxyde de fer (III)) affiche une valeur médiane de 0.05g/kg et une variabilité marquée (étendue de 0.005 à 0.165g/kg) à l'image de l'oxyde ferreux.

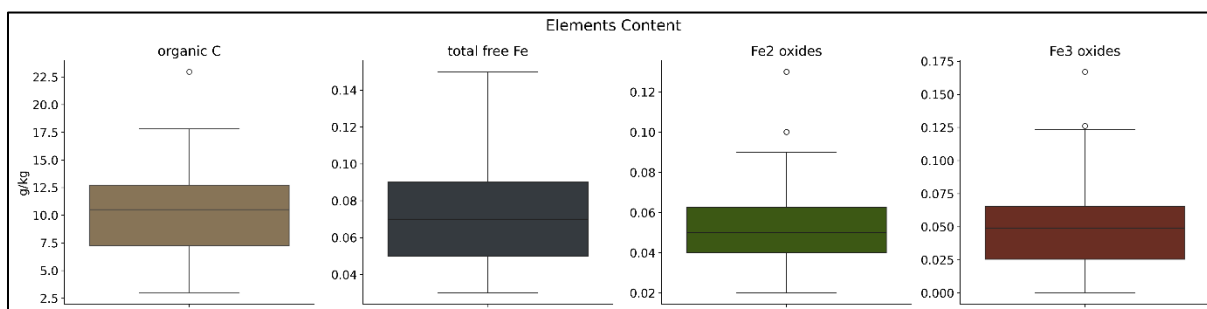


Figure 5 – Boxplots de la distribution des teneurs en carbone organique, fer total libre et oxydes de fer (Fe^{2+} et Fe^{3+}) sur le site d'Aranda de Duero

Statistiques descriptives des teneurs en SOC et FeOx dans les trois tirages

Les histogrammes ci-dessous présentent la distribution des teneurs moyennes en SOC (cf. Figure 6) et FeOx (cf. Figure 7) à travers les trois tirages pour les jeux de données d'entraînement et de test des modèles prédictifs. Le carbone organique affiche une valeur moyenne de 10g/kg et une variabilité marquée (écart-type de 8.3g/kg). L'oxyde ferrique (oxyde de fer (III)) affiche une valeur moyenne de 0.05g/kg et une variabilité marquée (écart-type de 0.06g/kg).

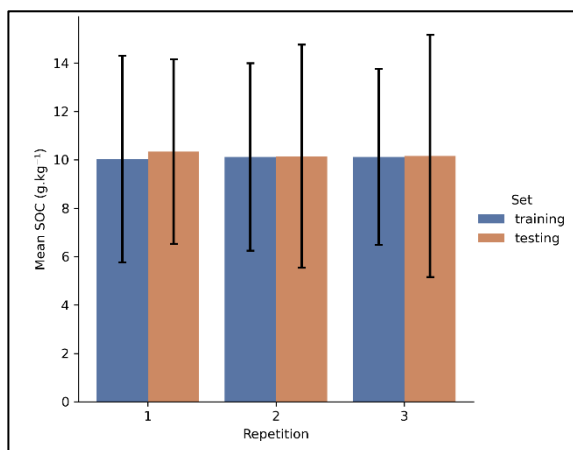


Figure 6 – Histogrammes de la distribution de la teneur en SOC

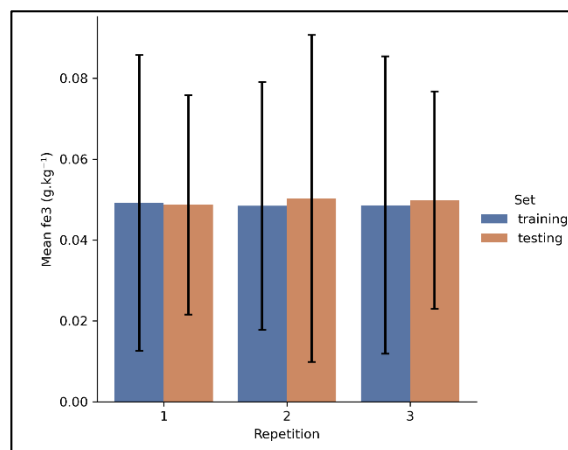


Figure 7 – Histogrammes de la distribution de la teneur en Fe^{3+}

Statistiques de performance des modèles prédictifs pour le SOC

Pour le SOC comme les FeOx, le meilleur type de modèle s'avère être le SVR. Les statistiques de calibration des modèles PLSR, RF, SVR et DNN calibrés pour prédire le SOC et les FeOx sont présentées dans les Annexes 9 et 10.

De tous les pré-traitements labo appliqués, c'est l'application du CR qui performe le mieux pour la prédiction du SOC; et de toutes les combinaisons prédictives testées, c'est le dataset d12 qui performe le mieux. Aussi, c'est le modèle SVR entraîné sur les 3 tirages de cette combinaison d12-CR qui est utilisé pour la prédiction du SOC. La figure 8 montre ses statistiques de performance, avec une RMSE de 3.2g/kg et un RPIQ de 1.83. Ces scores sont jugés bons au regard des standards de la télédétection pédologique, où un RPIQ supérieur à 1.7 est généralement jugé très satisfaisant (BELLON-MAUREL & McBRATNEY, 2011). Le modèle présente une bonne stabilité dans ses 3 répétitions, mais une tendance à la surestimation des valeurs faibles de SOC et une sous-estimation des valeurs élevées de SOC. Cette distribution des résidus suggère une difficulté à modéliser les extrêmes de la gamme de teneur du SOC dans ces sols.

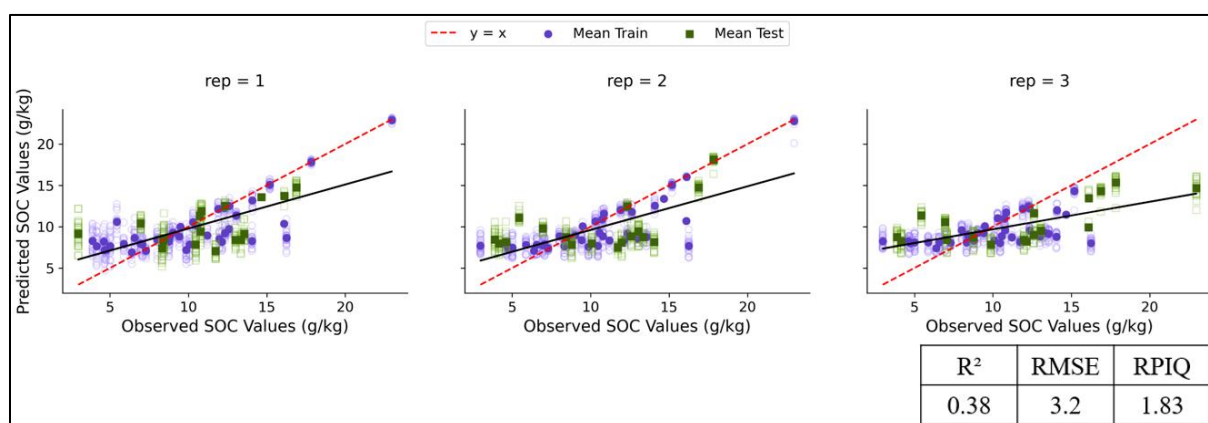


Figure 8 – Statistiques de performance des modèles SVR pour le SOC

Statistiques de performance des modèles prédictifs pour l'oxyde de fer (III)

De tous les pré-traitements labo appliqués, c'est l'application du SNV qui performe le mieux pour la prédiction du Fe³⁺, confirmant son utilité pour réduire les effets de diffusion et améliorer la cohérence des spectres. Parmi les combinaisons prédictives testées, c'est le dataset d7 qui performe le mieux. Aussi, c'est le modèle SVR entraîné sur les 3 tirages de cette combinaison d7-SNV qui est utilisé pour la prédiction du Fe³⁺. La figure 9 montre ses statistiques de performance, avec une RMSE de 0.03g/kg et un RPIQ de 1.3, jugé faible car en deçà des attentes pour la télédétection pédologique. Le modèle présente une stabilité relative dans ses 3 répétitions, avec une tendance à la surestimation des valeurs faibles de Fe³⁺ et une sous-estimation des valeurs élevées de Fe³⁺.

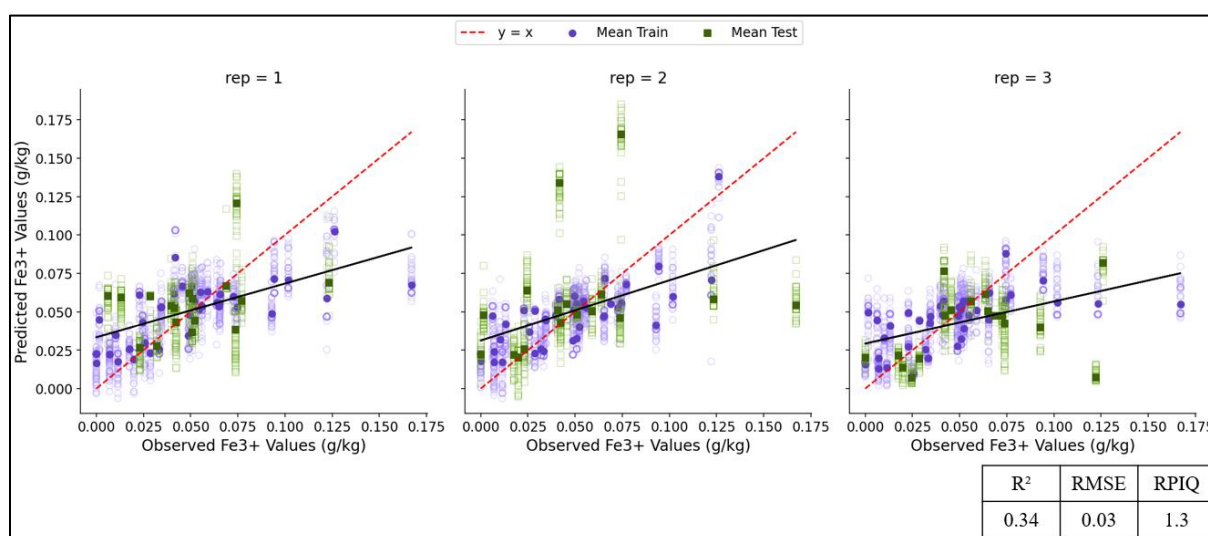


Figure 9 – Statistiques de performance des modèles SVR pour le Fe³⁺

Résultats des prédictions du SOC à l'aide de toute la série temporelle S-2

Cette carte illustre les prédictions spatiales de la teneur en SOC sur le vignoble d'Aranda (cf. Figure 10), comparées aux valeurs mesurées en laboratoire. Les couleurs, allant du violet (<8g/kg) au jaune (>10.5g/kg), représentent les classes de quantiles des teneurs moyennes en SOC prédites. Les disques représentent les points d'échantillonnage terrain, dont le rayon est proportionnel à la teneur en SOC réelle, et dont le remplissage bleu ou rouge indique respectivement une sur-estimation ou une sous-estimation de la teneur en SOC prédite par le modèle. L'analyse montre une variabilité spatiale marquée du SOC et une fracture Nord-Sud bien visible en termes de répartition. La zone Sud du vignoble, à forte charge en éléments grossiers, présente les teneurs observées les plus élevées. Les teneurs prédites sont concordantes, bien que le modèle ait tendance à sous-estimer ces fortes teneurs en SOC. À l'inverse de la zone Sud, la zone Nord du vignoble montre des teneurs observées et prédites relativement plus faibles. De manière générale et au droit des points extrêmes, le modèle a tendance à sous-estimer les teneurs observées élevées en SOC et à sur-estimer les teneurs observées faibles en SOC. La distribution en violine dans le carton confirme une distribution des valeurs prédites majoritairement centrée autour de 10g/kg, avec un minimum de 6.9g/kg et un maximum de 16.7g/kg.

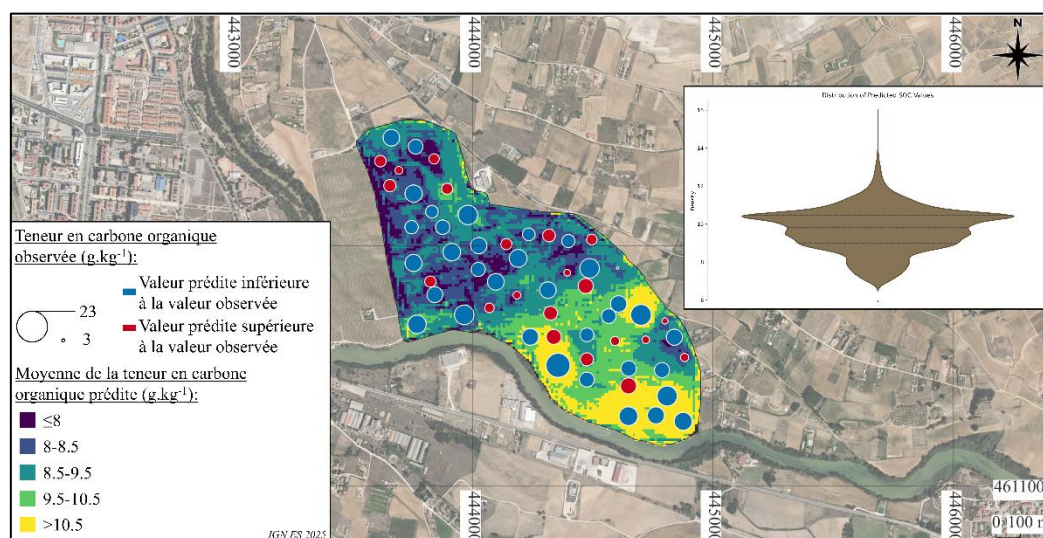


Figure 10 – Carte des prédictions de la teneur moyenne en SOC sur le vignoble d'Aranda à l'aide de la série temporelle S-2

La carte de l'écart-type des prédictions (cf. Figure 11) confirme la répartition spatiale décrite ci-avant: la variabilité des prédictions est très faible (<0.5g/kg) dans la zone Sud du vignoble (à l'exception de l'Est) où les teneurs observées et prédites y sont relativement élevées, tandis qu'elle est très marquée dans la zone Nord où les teneurs observées et prédites y sont relativement faibles.

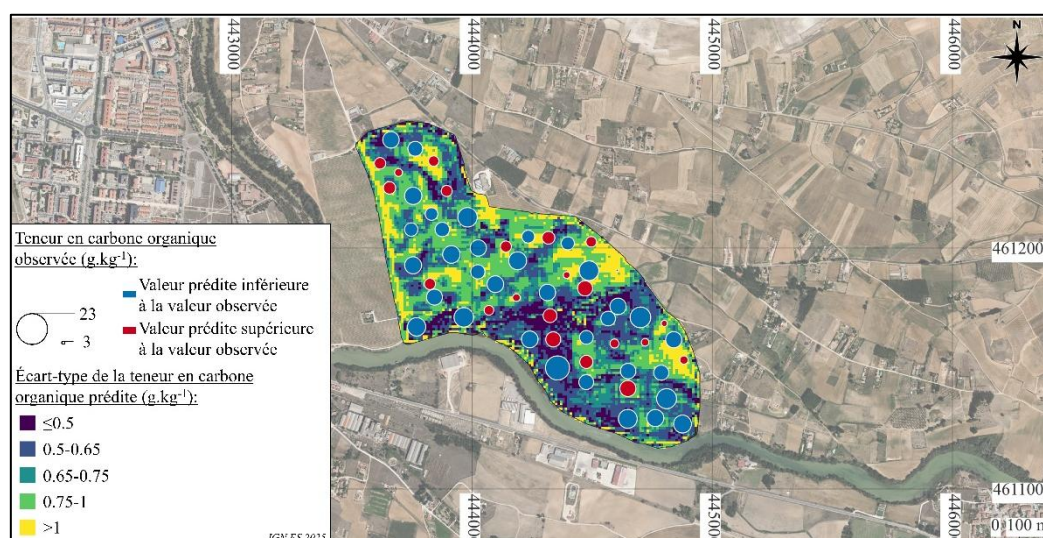


Figure 11 – Carte de l'écart-type des prédictions de la teneur moyenne en SOC sur le vignoble d'Aranda à l'aide de la série temporelle S-2

La carte de l'indice de confiance au seuil de 95% et du nombre de dates utilisées par prédiction sont présentées dans les Annexes 11 et 12. À titre de comparaison, celles des prédictions (moyenne et écart-type) à l'aide de la date la plus en sol nu sont présentées dans les Annexes 13 et 14.

Chapitre 4 – Perspectives

La cartographie du SOC reste cependant un défi, surtout si elle repose sur la spatialisation de co-variables par géostatistique. En effet, LAGACHERIE et al. (2017) montrent que 60 % de la variabilité du SOC s'observe dans les 100 premiers mètres autour d'un point d'échantillonnage. Cette forte variabilité locale, souvent liée à des facteurs pédologiques, topographiques ou de gestion parcellaire, limite la capacité des méthodes géostatistiques classiques à capturer fidèlement les gradients fins de cette propriété à travers la réponse spectrale enregistrée dans une image satellitaire. Bien que les pseudo-indices spectraux dérivés des spectres de laboratoire soient moins sensibles aux bruits environnementaux que les données satellitaires, leur spatialisation par krigeage ordinaire peut également être affectée par cette variabilité à courte portée. En effet, si les indices labo reflètent des signatures spectrales plus stables et contrôlées, leur distribution spatiale reste influencée par les mêmes facteurs que le SOC (pratiques culturales, microtopographie). Ainsi, bien que le krigeage permette d'obtenir une estimation continue, la résolution et la précision de la cartographie restent contraintes par la densité d'échantillonnage et la structure spatiale complexe du SOC. Cette limite justifie l'intégration complémentaire de co-variables environnementales (indices spectraux, topographie, texture) pour affiner les prédictions spatiales.

Intégration de l'information spatialisée de la charge en éléments grossiers

Les résultats de cette étude révèlent que les prédictions de teneur en SOC coïncident spatialement avec la répartition de la charge en éléments grossiers, suggérant un lien fonctionnel entre ces deux propriétés du sol. Cette observation est soutenue par les travaux récents de ZAYANI et al. (2025, non-publiés), qui mettent en évidence une corrélation fortement significative entre la charge en éléments grossiers et l'indice spectral NSSI dérivé des images Sentinel-2 sur le même site d'étude. Le NSSI, initialement développé pour distinguer les sols nus de la végétation non photosynthétique, exploite les différences de réflectance dans le proche infrarouge (750-900nm) et s'avère particulièrement sensible à la présence d'éléments grossiers, dont la signature spectrale diffère de celle des fractions fines du sol. En effet, la littérature confirme que les éléments grossiers influencent la distribution et la stabilité du SOC en modifiant la porosité, l'aération et la rétention hydrique du sol. Par exemple, des études (URBINA-SALAZAR et al., 2021) ont montré que les zones riches en éléments grossiers peuvent présenter des teneurs en SOC plus faibles en raison d'une surface spécifique réduite pour l'adsorption de matière organique, mais aussi des dynamiques de minéralisation distinctes liées à des microclimats locaux plus secs ou mieux drainés. La variabilité spatiale du NSSI sur le vignoble d'Aranda est concordante avec la variabilité spatiale du SOC prédit (cf. Annexe 15). L'intégration du NSSI comme co-variable dans les modèles de prédiction du SOC permettrait donc de mieux capturer cette variabilité, surtout dans des contextes viticoles où les pratiques culturales (labour, scarification) et l'érosion exposent fréquemment ces éléments en surface. De plus, l'utilisation d'indices spectraux comme le NSSI, combinée à des données multi-sources (S-1 pour l'humidité et S-2 pour la réflectance), pourrait améliorer la résolution spatiale et la précision des cartographies, en distinguant les zones où la charge en éléments grossiers masque ou altère le signal du SOC.

Apport de l'imagerie hyperspectrale

L'imagerie hyperspectrale émerge comme un outil prometteur pour surmonter les limites des méthodes traditionnelles d'analyse des sols viticoles, en offrant une caractérisation rapide et à haute résolution spectrale des propriétés clés comme le SOC et les FeOx. Les capteurs hyperspectraux couvrant le VNIR-SWIR (400-2'500nm) permettent de détecter des signatures spectrales spécifiques liées aux liaisons chimiques du SOC (bandes d'absorption à 1'700-2'500 nm pour les groupes C-H, O-H) et aux FeOx (pics à 480, 670 et 900nm pour la goéthite et l'hématite). Dès le début des années 2000, des études (GOUDA et al., 2024) ont montré que l'usage de l'imagerie hyperspectrale comme entrée dans des algorithmes de ML permet d'estimer le SOC avec une précision accrue (R^2 jusqu'à 0.85) par rapport aux méthodes conventionnelles, tout en réduisant les coûts et le temps d'analyse. Les drones équipés de capteurs comme la Mjöltnir VS620 de Hypspec, et les satellites EnMAP ou PRISMA, ouvrent des perspectives nouvelles pour une cartographie fine et dynamique des propriétés des sols à l'échelle parcellaire. Par exemple, l'utilisation d'un capteur hyperspectral sur drone a permis de cartographier la variabilité intra-parcellaire du SOC et des FeOx avec une résolution centimétrique, révélant des zones de stress hydrique ou de compaction non détectables par télédétection multispectrale classique.

Les avancées récentes en imagerie hyperspectrale confirment son potentiel transformateur pour la cartographie fine des propriétés du sol, mais révèlent aussi des défis méthodologiques persistants, notamment en matière de généralisation et de résolution spatiale. Les travaux de SELIGE et al. (2006) illustrent cette précision: en exploitant des images aéroportées HyMap (128 bandes, 420-2'480nm), ils obtiennent des R^2 supérieurs à 0.85 pour la prédiction du SOC, de l'azote et de la texture (sable), grâce à des modèles de MLR et surtout de PLSR ($R^2 = 0.9$

pour le SOC). Ces performances, bien que remarquables, soulignent un préalable critique: la nécessité de disposer d'un jeu de calibration robuste et représentatif, sans quoi les modèles peinent à se transférer à d'autres contextes pédologiques ou climatiques. Cette dépendance aux données terrain est également mise en évidence par HIVELY et al. (2011), qui montrent que la PLSR, appliquée à des images hyperspectrales pré-traitées (filtrage passe-bas, dérivées de SAVITZKY-GOLAY), permet d'estimer le SOC avec un R^2 de 0.65 – une précision stable même sur des champs indépendants. Leurs résultats suggèrent cependant que l'intégration de co-variables topographiques (pente, indice de rugosité) pourrait affiner les prédictions, en capturant des gradients de matière organique liés à l'érosion ou aux pratiques culturales.

Pourtant, la résolution spatiale reste un facteur limitant, surtout lorsque l'on passe des capteurs aéroportés ou drone aux plateformes satellitaires. LAGACHERIE et al. (2012), travaillant dans les vignobles du Languedoc, démontrent que l'utilisation d'images HyMap couplées à un co-krigeage améliore significativement l'estimation de la teneur en argile, avec une RMSE d'autant plus faible que la résolution est fine (de 50 à 500m). Leur approche, basée sur l'indice CR_{2206} ciblant les bandes d'absorption spécifiques de l'argile (2'206nm), révèle que les données hyperspectrales complètent avantageusement les méthodes géostatistiques classiques, mais que leur plein potentiel nécessite des techniques de démixage spectral pour séparer les signatures des minéraux, de la matière organique et des résidus végétaux. Cette complexité est encore accentuée lorsque l'on considère des capteurs satellitaires comme EnMAP ou PRISMA, dont la résolution spatiale plus grossière (30m) introduit des effets de mélange difficiles à corriger. WARD et al. (2020) le confirment: bien que leurs modèles PLSR atteignent un R^2 de 0.86 avec des spectres de laboratoire, les performances chutent à 0.48 avec les données EnMAP simulées, en raison des interférences liées à l'humidité résiduelle et à la couverture partielle du sol. Cette perte de précision souligne l'importance de calibrer les modèles sur des zones de sol nu bien caractérisées, et de combiner les données hyperspectrales avec des cartes de référence haute résolution pour guider le démixage des pixels satellitaires. De tous les défis qui persistent, sont à noter la nécessité de calibrer les modèles avec des données de terrain précises et la gestion des interférences spectrales (humidité, couverture végétale), qui peuvent masquer les signaux du SOC et des minéraux. Dans le cas de l'imagerie hyperspectrale satellite, il est donc nécessaire de considérer le démixage spectral des signatures en raison de la résolution spatiale limitée des capteurs. L'utilisation de cartes de référence construites à partir d'images aériennes ou drone, à plus haute résolution, offrent une solution prometteuse pour affiner la décomposition des spectres et améliorer la précision des estimations de variables pédologiques.

Face à ces limites, les approches non linéaires émergent également comme une piste sérieuse. ANGELOPOULOU et al. (2023) comparent ainsi les performances des modèles PLSR, Cubist, RF et SVM pour prédire le SOC et les carbonates à partir de capteurs aéroportés (HySpex, 1.9m) et satellitaires (PRISMA, 30m). Leurs résultats montrent que le Cubist surpasse systématiquement la PLSR, avec des R^2 de 0.79 (HySpex) et 0.76 (PRISMA) pour le SOC, malgré la résolution très différente des capteurs. Cependant, la prédiction des carbonates reste beaucoup plus incertaine avec PRISMA ($R^2 = 0.36$), en raison de la sensibilité des signatures spectrales à la rugosité de surface et à l'humidité – deux paramètres rarement intégrés dans les modèles actuels. Les auteurs proposent d'explorer des architectures de *deep learning*, capables de capturer les relations non linéaires entre les variables spectrales et pédologiques, et d'augmenter la taille des échantillons pour améliorer la généralisation. Cette transition vers des méthodes plus avancées pourrait également permettre de fusionner des données multi-sources (hyperspectral, LiDAR, topographie), afin de compenser les limites intrinsèques de chaque capteur et d'aboutir à des cartographies à la fois précises et spatialement cohérentes.

Optimisation du plan d'échantillonnage et de l'incrémentation des jeux de données

La représentativité des échantillons de sol est cruciale pour établir des modèles robustes de prédiction des teneurs en SOC et FeOx. Un plan d'échantillonnage adapté, combinant une densité suffisante de points et une stratification selon les types de sol (rugosité), couverture végétale ou travaux du sol, permettrait de mieux capturer la variabilité spatiale à petite échelle. Une autre piste envisageable consisterait à optimiser la sélection des dates d'acquisition en réduisant leur nombre et en ne retenant que celles contribuant le plus à la précision des estimations. Une approche itérative, basée sur l'ajout progressif des dates minimisant la RMSE, pourrait permettre d'identifier un sous-ensemble optimal de données temporelles, limitant ainsi les risques de redondance ou de sur-apprentissage. Des tests préliminaires de réduction ciblée du nombre de dates montrent une réduction de 15% de la RMSE.

Approches hybrides

L'intégration de modèles mécanistes (modèles de transfert radiatif comme PROSAIL ou modèles de bilan carbone) avec les données d'imagerie pourrait améliorer la compréhension des interactions entre couverture végétale, propriétés du sol et signal spectral. Cela permettrait notamment de mieux isoler l'effet des pratiques culturales (enherbement, travail du sol) sur les teneurs en SOC et FeOx.

Conclusion

Les sols agricoles du bassin méditerranéen, soumis à des pressions climatiques et anthropiques croissantes, représentent un enjeu majeur pour la séquestration du carbone et la durabilité des systèmes de production. Comme souligné en introduction, leur capacité à stocker du SOC et à maintenir des processus pédogéniques équilibrés, notamment via leur teneur élevée en FeOx, est essentielle pour atténuer le changement climatique et préserver la fertilité des sols. Ce mémoire s'est inscrit dans cette problématique en explorant le potentiel de la télédétection optique, couplée à des spectres de laboratoire, pour prédire spatialement les teneurs en SOC et en FeOx dans des vignobles espagnols. Les résultats obtenus confirment partiellement les hypothèses initiales, tout en révélant des pistes d'amélioration pour les approches futures.

L'utilisation de séries temporelles S-2, combinée à des spectres Vis-PIR acquis en laboratoire, a permis de tester différentes stratégies de modélisation pour prédire le SOC et les FeOx. Parmi les approches évaluées, les modèles de *machine learning* (SVR) se sont avérés les plus performants, tant pour le SOC que pour les FeOx, avec une nette supériorité en termes de compromis précision/temps de calcul face aux modèles de *deep learning* (DNN). Pour le SOC, la configuration optimale a été obtenue avec un pré-traitement des spectres labo par CR et le dataset d12 (bandes S-2 + premier pseudo-indice labo uniquement), tandis que pour les FeOx, un pré-traitement SNV et le dataset d7 (tout S-2 + premier pseudo-indice labo) ont donné les meilleurs résultats. Ces observations soulignent l'importance du choix des pré-traitements spectraux et de la composition du dataset d'entrée, qui doivent être adaptés aux propriétés physico-chimiques des variables cibles. L'analyse de la variabilité spatiale du SOC a par ailleurs révélé une meilleure cohérence des prédictions lorsque l'ensemble de la série temporelle était utilisé, plutôt que la date la plus en sol nu. Ce résultat met en lumière l'apport des dynamiques temporelles pour capturer les variations de réflectance liées aux changements d'état de surface, même si cela introduit un biais potentiel par l'attribution d'une valeur fixe de SOC à des points présentant une forte variabilité spectrale au cours du temps. L'intégration des FeOx comme co-variables a bien amélioré les prédictions du SOC, bien que dans une moindre mesure que prévu, suggérant que leur rôle dans la stabilisation du SOC ou dans la réponse spectrale des sols pourrait être plus complexe qu'envisagé initialement. Enfin, les prédictions du SOC sont spatialement cohérentes compte tenu de la connaissance du vignoble et des caractéristiques pédologiques (humidité, rugosité). Elles ont toutefois tendance à être sous-estimée dans la partie Sud, où la charge en éléments grossiers est élevée. Cet aspect met en évidence la nécessité d'évaluer le stock de carbone (g/m^2) plutôt que la teneur (g/kg), en particulier dans des sols hétérogènes comme le sont ces sols méditerranéens en climat semi-aride.

Malgré ces avancées, plusieurs limites méritent d'être soulignées. L'apport des indices spectraux dérivés des images S-2 s'est avéré marginal, voire négligeable, comparé à celui des spectres de laboratoire. Ce constat rejoint les observations de la littérature, où les bandes spectrales larges des capteurs multispectraux peinent souvent à capturer les signatures fines des constituants du sol. De même, l'ajout de co-variables pédologiques ou d'indices supplémentaires n'a pas systématiquement amélioré les performances des modèles, confirmant que la redondance d'information peut, dans certains cas, nuire à la robustesse des prédictions, et ce malgré l'étape de réduction de la dimensionnalité.

Les résultats de ce mémoire ouvrent plusieurs pistes pour affiner les stratégies de cartographie du SOC et des FeOx. L'une des perspectives les plus prometteuses réside dans l'utilisation de l'imagerie hyperspectrale, notamment par drone, qui permettrait de capturer des signatures spectrales plus fines et d'améliorer la résolution spatiale des prédictions. Une telle approche pourrait également faciliter l'intégration de données sur la charge en éléments grossiers, souvent négligée dans les modèles actuels mais qui semble apporter des informations importantes pour la prédiction du SOC sur ce site. Au-delà des aspects méthodologiques, ce travail souligne l'intérêt opérationnel de disposer de cartographies précises du SOC et des FeOx pour la gestion des sols viticoles. En identifiant les zones à faible teneur en SOC ou à forte concentration en FeOx, les viticulteurs pourraient adapter leurs pratiques culturales (par exemple, en favorisant le maintien d'un couvert végétal) pour améliorer la résilience des sols et leur capacité à séquestrer le carbone. Ces données pourraient également alimenter des outils d'aide à la décision, dans le cadre de projets comme SANCHOSTHIRST, pour évaluer l'impact des modes de conduite des sols sur leur santé à long terme.

En conclusion, ce mémoire a permis de progresser dans la compréhension des interactions entre télédétection, propriétés pédologiques et modélisation prédictive. Si les défis méthodologiques persistent, notamment en matière de résolution spectrale, spatiale et d'échantillonnage, les résultats obtenus confirment le potentiel des approches hybrides pour affiner la prédiction spatiale du SOC et des FeOx. À l'heure où les enjeux de durabilité agricole et de lutte contre le changement climatique s'intensifient, ces avancées offrent des outils concrets pour concilier productivité viticole et préservation des sols, en ligne avec les objectifs de l'Initiative 4 pour 1000 et des politiques européennes de gestion durable des terres.

Références

- A. Akiba et al., « Optuna: A Next-generation Hyperparameter Optimization Framework », *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 2623–2631, juill. 2019, doi: 10.1145/3292500.3330701.
- F. Al-Khaier, « Soil Salinity Detection Using Satellite Remote Sensing », 2003, URL: https://www.researchgate.net/publication/252252892_Soil_Salinity_Detection_Using_Satellite_Remote_Sensing
- T. Angelopoulou et al., « Evaluation of Airborne HySpex and Spaceborne PRISMA Hyperspectral Remote Sensing Data for Soil Organic Matter and Carbonates Estimation », *Remote Sensing*, vol. 15, no 4, p. 1106, févr. 2023, doi: 10.3390/RS15041106.
- C. Anchisi et al., « Stability studies of new cosmetic formulations with vegetable extracts as functional agents », *Farmaco*, vol. 56, no 5-7, p. 427–431, juill. 2001, doi: 10.1016/S0014-827X(01)01055-2.
- D. Arrouays et al., « Digital soil mapping across the globe », *Geoderma Regional*, vol. 9, p. 1–4, juin 2017, doi: 10.1016/J.GEODRS.2017.03.002.
- E. Barnes et al., « Coincident detection of crop water stress, nitrogen status, and canopy density using ground based multispectral data », 2000, URL: https://www.researchgate.net/publication/43256762_Coincident_detection_of_crop_water_stress_nitrogen_status_and_canopy_density_using_ground_based_multispectral_data.
- H. Bartholomeus et al., « Determining iron content in Mediterranean soils in partly vegetated areas, using spectral reflectance and imaging spectroscopy », *International Journal of Applied Earth Observation and Geoinformation*, vol. 9, no 2, p. 194–203, mai 2007, doi: 10.1016/J.JAG.2006.09.001.
- H. Bartholomeus et al., « Spectral reflectance based indices for soil organic carbon quantification », *Geoderma*, vol. 145, no 1–2, p. 28–36, mai 2008, doi: 10.1016/j.geoderma.2008.01.010.
- V. Bellon-Maurel, A. McBratney. « Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives », *Soil Biology and Biochemistry*, vol. 43, no 7, p. 1398–1410, Juillet 2011, doi: 10.1016/j.soilbio.2011.02.019.
- J. E. Bullard et al., « Quantifying iron oxide coatings on dune sands using spectrometric measurements: An example from the Simpson-Strzelecki Desert, Australia », *Journal of Geophysical Research: Solid Earth*, vol. 107, no B6, p. ECV 5-1, juin 2002, doi: 10.1029/2001JB000454.
- K. Cartwright, « Simpson’s Rule Cumulative Integration with MS Excel and Irregularly-spaced Data », *Journal of Mathematical Sciences and Mathematics Education*, vol. 12, no. 2, pp. 1–9, 2017. URL: https://www.researchgate.net/publication/319968921_Simpson’s_Rule_Cumulative_Integration_with_MS_Excel_and_Irregularly-spaced_Data
- F. Castaldi et al., « Assessing the capability of Sentinel-2 time-series to estimate soil organic carbon and clay content at local scale in croplands », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 199, p. 40–60, mai 2023, doi: 10.1016/J.ISPRSJPRS.2023.03.016.
- F. Castaldi et al., « Soil Organic Carbon Mapping Using LUCAS Topsoil Database and Sentinel-2 Data: An Approach to Reduce Soil Moisture and Crop Residue Effects », *Remote Sensing*, vol. 11, no 18, p. 2121, sept. 2019, doi: 10.3390/RS11182121.
- F. Castaldi et al., « Sampling Strategies for Soil Property Mapping Using Multispectral Sentinel-2 and Hyperspectral EnMAP Satellite Data », *Remote Sensing*, vol. 11, no 3, p. 309, févr. 2019, doi: 10.3390/RS11030309.
- P. Ceccato et al., « Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1: Theoretical approach », *Remote Sensing of Environment*, vol. 82, no 2-3, p. 188–197, oct. 2002, doi: 10.1016/S0034-4257(02)00037-8.
- C. Chen et al., « Iron-mediated organic matter decomposition in humid soils can counteract protection », *Nature Communications*, vol. 11, no 1, p. 1–13, mai 2020, doi: 10.1038/s41467-020-16071-5.
- E. Chlouveraki et al., « Estimation of soil properties using Hyperspectral imaging and Machine learning », *Smart Agricultural Technology*, vol. 10, p. 100790, mars 2025, doi: 10.1016/J.ATECH.2025.100790.
- R. N. Clark et al., « Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications », *Journal of Geophysical Research*, vol. 89, no B7, p. 6329–6340, 1984, doi: 10.1029/JB089iB07p06329.

- B. Delaney et al., « Satellite Remote Sensing Techniques and Limitations for Identifying Bare Soil », *Remote Sensing*, vol. 17, no 4, p. 630, févr. 2025, doi: 10.3390/RS17040630.
- S. Diek et al., « Barest Pixel Composite for Agricultural Areas Using Landsat Time Series », *Remote Sensing*, vol. 9, no 12, p. 1245, déc. 2017, doi: 10.3390/RS9121245.
- K. Dvorakova et al., « Soil Organic Carbon Mapping from Remote Sensing: The Effect of Crop Residues », *Remote Sensing*, vol. 12, no 12, p. 1913, juin 2020, doi: 10.3390/RS12121913.
- R. Escadafal, « Caractérisation de la surface des sols arides par observations de terrain et par télédétection : applications : exemple de la région de Tataouine (Tunisie) », 1989, URL: <https://www.documentation.ird.fr/hor/fdi:26944>.
- R. Escadafal, « Indices spectraux pour la télédétection de la dégradation des milieux naturels en Tunisie aride », 1994, URL: <https://fr.scribd.com/document/458781617/indices-spectraux-pour-la-teledetection-de-la-degradation-des-milieux-naturels-en-tunisie-aride-r-escadafal-a-belghith-et-h-ben-moussa-pdf>.
- G. L. Feyisa et al., « Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery », *Remote Sensing of Environment*, vol. 140, p. 23–35, janv. 2014, doi: 10.1016/J.RSE.2013.08.029.
- M. Gouda et al., « Integration of remote sensing and artificial neural networks for prediction of soil organic carbon in arid zones », *Frontiers in Environmental Science*, vol. 12, p. 1448601, sept. 2024, doi: 10.3389/FENVS.2024.1448601.
- A. A. Gitelson, « Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation », *Journal of Plant Physiology*, vol. 161, no 2, p. 165–173, janv. 2004, doi: 10.1078/0176-1617-01176.
- A. A. Gitelson et al., « Novel algorithms for remote estimation of vegetation fraction », *Remote Sensing of Environment*, vol. 80, no 1, p. 76–87, avr. 2002, doi: 10.1016/S0034-4257(01)00289-9.
- A. A. Gitelson et al., « Use of a green channel in remote sensing of global vegetation from EOS-MODIS », *Remote Sensing of Environment*, vol. 58, no 3, p. 289–298, déc. 1996, doi: 10.1016/S0034-4257(96)00072-7.
- A. A. Gitelson et al., « Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves », *Photochemistry and Photobiology*, vol. 74, no 1, p. 38–45, juill. 2001, doi: 10.1562/0031-8655(2001)074<0038:OPANEO>2.0.CO;2.
- A. A. Gitelson et al., « Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves », *Journal of Plant Physiology*, vol. 160, no 3, p. 271–282, mars 2003, doi: 10.1078/0176-1617-00887.
- U. Heiden et al., « Soil Reflectance Composites—Improved Thresholding and Performance Evaluation », *Remote Sensing*, vol. 14, no 18, p. 4526, sept. 2022, doi: 10.3390/RS14184526.
- R. D. Hewson et al., « Geologic and alteration mapping at Mt fitton, South Australia, using ASTER satellite-borne data », *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2, p. 724–726, 2001, doi: 10.1109/IGARSS.2001.976615.
- W. D. Hively et al., « Use of Airborne Hyperspectral Imagery to Map Soil Properties in Tilled Agricultural Fields », *Applied and Environmental Soil Science*, vol. 2011, p. 358193, janv. 2011, doi: 10.1155/2011/358193.
- A. R. Huete, « A soil-adjusted vegetation index (SAVI) », *Remote Sensing of Environment*, vol. 25, no 3, p. 295–309, août 1988, doi: 10.1016/0034-4257(88)90106-X.
- N. Jia et al., « Important role of Fe oxides in global soil carbon stabilization and stocks », *Nature Communications*, vol. 15, no 1, p. 1–11, nov. 2024, doi: 10.1038/s41467-024-54832-8.
- Z. Jiang et al., « Development of a two-band enhanced vegetation index without a blue band », *Remote Sensing of Environment*, vol. 112, no 10, p. 3833–3845, oct. 2008, doi: 10.1016/J.RSE.2008.06.006.
- H. Jin et al., « Comparing Laboratory and Satellite Hyperspectral Predictions of Soil Organic Carbon in Farmland », *Agronomy*, vol. 14, no 1, p. 175, janv. 2024, doi: 10.3390/AGRONOMY14010175.
- Y. J. Kaufman et al., « Atmospherically Resistant Vegetation Index (ARVI) for EOS-MODIS », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no 2, p. 261–270, 1992, doi: 10.1109/36.134076.
- M. Khosravi et al., « On the impact of soil texture on local scale organic carbon quantification: From airborne to spaceborne sensing domains », *Soil and Tillage Research*, vol. 241, p. 106125, sept. 2024, doi: 10.1016/J.STILL.2024.106125.

- Y. Koerting et al., « VNIR-SWIR Imaging Spectroscopy for Mining: Insights for Hyperspectral Drone Applications », *Mining*, vol. 4, no 4, p. 1013–1057, nov. 2024, doi: 10.3390/MINING4040057.
- P. Lagacherie et al., « Using scattered hyperspectral imagery data to map the soil properties of a region », *European Journal of Soil Science*, vol. 63, no 1, p. 110–119, févr. 2012, doi: 10.1111/J.1365-2389.2011.01409.X.
- R. Lal, « Soil carbon sequestration to mitigate climate change », *Geoderma*, vol. 123, no 1-2, p. 1–22, nov. 2004, doi: 10.1016/J.GEODERMA.2004.01.032.
- W. Li et al., « A Practical Remote Sensing Monitoring Framework for Late Frost Damage in Wine Grapes Using Multi-Source Satellite Data », *Remote Sensing*, vol. 13, no 16, p. 3231, août 2021, doi: 10.3390/RS13163231.
- A. A. Lima et al., « Soil Organic Carbon Assessment Using Remote-Sensing Data and Machine Learning: A Systematic Literature Review », *Remote Sensing*, vol. 17, no 5, p. 882, mars 2025, doi: 10.3390/RS17050882.
- S. K. McFeeters, « The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features », *International Journal of Remote Sensing*, vol. 17, no 7, p. 1425–1432, 1996, doi: 10.1080/01431169608948714.
- F. van der Meer, « Remote sensing image analysis and geostatistics », *International Journal of Remote Sensing*, vol. 33, no 18, p. 5644–5676, 2012, doi: 10.1080/01431161.2012.666363.
- O. P. Mehra et al., « Iron oxide removal from soils and clays by a dithionite-citrate system buffered with sodium bicarbonate », *Clays and Clay Minerals*, p. 317–327, 2013, doi: 10.1016/B978-0-08-009235-5.50026-7.
- K. M. Oliveira et al., « Predicting Particle Size and Soil Organic Carbon of Soil Profiles Using VIS-NIR-SWIR Hyperspectral Imaging and Machine Learning Models », *Agriculture*, vol. 14, no 16, p. 2869, août 2024, doi: 10.3390/AGRICULTURE14162869.
- C. Parat et al., « The relationship between copper accumulated in vineyard calcareous soils and soil organic matter and iron », *European Journal of Soil Science*, vol. 53, no 4, p. 663–670, déc. 2002, doi: 10.1046/J.1365-2389.2002.00478.X.
- M. Pavlovic et al., « A Deep Learning Approach to Estimate Soil Organic Carbon from Remote Sensing », *Remote Sensing*, vol. 16, no 4, p. 655, févr. 2024, doi: 10.3390/RS16040655.
- F. T. Payen et al., « Soil organic carbon sequestration rates in vineyard agroecosystems under different soil management practices: A meta-analysis », *Journal of Cleaner Production*, vol. 290, p. 125736, mars 2021, doi: 10.1016/J.JCLEPRO.2020.125736.
- D. H. Pearlshtien et al., « Effect of Organic Matter Content on the Spectral Signature of Iron Oxides across the VIS–NIR Spectral Region in Artificial Mixtures: An Example from a Red Soil from Israel », *Remote Sensing*, vol. 12, no 12, p. 1960, juin 2020, doi: 10.3390/RS12121960.
- Y. Peng et al., « Modeling Soil Organic Carbon at Regional Scale by Combining Multi-Spectral Images with Laboratory Spectra », *PLOS ONE*, vol. 10, no 11, p. e0142295, nov. 2015, doi: 10.1371/JOURNAL.PONE.0142295.
- M. Pinheiro et al., « Prediction of Soil Physical and Chemical Properties by Visible and Near-Infrared Diffuse Reflectance Spectroscopy in the Central Amazon », *Remote Sensing*, vol. 9, no 4, p. 293, mars 2017, doi: 10.3390/RS9040293.
- D. Rogge et al., « Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014) », *Remote Sensing of Environment*, vol. 205, p. 1–17, févr. 2018, doi: 10.1016/J.RSE.2017.11.004.
- G. Rondeaux et al., « Optimization of soil-adjusted vegetation indices », *Remote Sensing of Environment*, vol. 55, no 2, p. 95–107, févr. 1996, doi: 10.1016/0034-4257(95)00186-7.
- L. C. Rowan et al., « Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data », *Remote Sensing of Environment*, vol. 84, no 3, p. 350–366, mars 2003, doi: 10.1016/S0034-4257(02)00127-X.
- W. Sahwan et al., « Relationships between spectral features, iron oxides and colours of surface soils in northern Jordan », *European Journal of Soil Science*, vol. 72, no 1, p. 80–97, janv. 2021, doi: 10.1111/EJSS.12986.
- G. M. Salani et al., « Soil Organic Carbon Estimation in Ferrara (Northern Italy) Combining In Situ Geochemical Analyses and Hyperspectral Remote Sensing », *Environments*, vol. 10, no 10, p. 173, oct. 2023, doi: 10.3390/ENVIRONMENTS10100173.

- T. Selige et al., « High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures », *Geoderma*, vol. 136, no 1-2, p. 235–244, déc. 2006, doi: 10.1016/J.GEODERMA.2006.03.050.
- A. Stevens et al., « Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils », *Geoderma*, vol. 144, no 1-2, p. 395–404, mars 2008, doi: 10.1016/J.GEODERMA.2007.12.009.
- E. Straffelini et al., « European vineyards and their cultural landscapes exposed to record drought and heat », *Agricultural Systems*, vol. 219, p. 104034, août 2024, doi: 10.1016/J.AGSY.2024.104034.
- Y. Sui et al., « Improving the Spatiotemporal Transferability of Hyperspectral Remote Sensing for Estimating Soil Organic Matter by Minimizing the Coupling Effect of Soil Physical Properties on the Spectrum: A Case Study in Northeast China », *Agronomy*, vol. 14, no 5, p. 1067, mai 2024, doi: 10.3390/AGRONOMY14051067.
- D. A. Summers et al., « Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties », *Ecological Indicators*, vol. 11, no 1, p. 123–131, janv. 2011, doi: 10.1016/J.ECOLIND.2009.05.001.
- P. K. Tarafder et al., « An Optimised 1,10-Phenanthroline Method for the Determination of Ferrous and Ferric Oxides in Silicate Rocks, Soils and Minerals », *Geostandards and Geoanalytical Research*, vol. 37, no 2, p. 155–168, juin 2013, doi: 10.1111/J.1751-908X.2012.00183.X.
- E. Tian et al., « Spatiotemporal prediction of soil organic carbon density for Europe (2000–2022) in 3D+T based on Landsat-based spectral indices time-series », *Research Square*, sept. 2024, doi: 10.21203/RS.3.RS-5128244/V1.
- N. Tian et al., « Development of a two-band enhanced vegetation index without a blue band », *Remote Sensing of Environment*, vol. 112, no 10, p. 3833–3845, oct. 2008, doi: 10.1016/J.RSE.2008.06.006.
- W. Tian et al., « Overview of the radiometric and biophysical performance of the MODIS vegetation indices », *Remote Sensing of Environment*, vol. 83, no 1-2, p. 195–213, nov. 2002, doi: 10.1016/S0034-4257(02)00096-2.
- D. Urbina-Salazar et al., « Using Sentinel-2 Images for Soil Organic Carbon Content Mapping in Croplands of Southwestern France. The Usefulness of Sentinel-1/2 Derived Moisture Maps and Mismatches between Sentinel Images and Sampling Dates », *Remote Sensing*, vol. 13, no 24, p. 5115, déc. 2021, doi: 10.3390/RS13245115.
- D. Urbina-Salazar et al., « Sentinel-2 and Sentinel-1 Bare Soil Temporal Mosaics of 6-Year Periods for Soil Organic Carbon Content Mapping in Central France », *Remote Sensing*, vol. 15, no 9, p. 2410, mai 2023, doi: 10.3390/RS15092410.
- D. Urbina-Salazar et al., « Combined use of Sentinel-2 images and Sentinel-1-derived moisture maps for soil organic carbon content mapping in croplands, South-western France », *EGUsphere*, mars 2021, doi: 10.5194/EGUSPHERE-EGU21-8836.
- E. Vaudour et al., « Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems », *Remote Sensing of Environment*, vol. 223, p. 21–33, mars 2019, doi: 10.1016/J.RSE.2019.01.006.
- E. Vaudour et al., « Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands », *International Journal of Applied Earth Observation and Geoinformation*, vol. 96, p. 102277, avr. 2021, doi: 10.1016/J.JAG.2020.102277.
- E. Vaudour et al., « Satellite Imagery to Map Topsoil Organic Carbon Content over Cultivated Areas: An Overview », *Remote Sensing*, vol. 14, no 12, p. 2917, juin 2022, doi: 10.3390/RS14122917.
- G. Vivone et al., « Deep Learning in Remote Sensing Image Fusion: Methods, protocols, data, and future perspectives », *IEEE Geoscience and Remote Sensing Magazine*, 2024, doi: 10.1109/MGRS.2024.3495516.
- A. Walkley and A. Black, « An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method », *Soil Science*, vol. 37, pp. 29–38, 1934, URL: https://journals.lww.com/soilsci/citation/1934/01000/an_examination_of_the_degtjareff_method_for.3.aspx
- K. J. Ward et al., « Mapping Soil Organic Carbon for Airborne and Simulated EnMAP Imagery Using the LUCAS Soil Database and a Local PLSR », *Remote Sensing*, vol. 12, no 20, p. 3451, oct. 2020, doi: 10.3390/RS12203451.
- R. Webster et al., « Geostatistics for Environmental Scientists: Second Edition », John Wiley and Sons, janv. 2008, doi: 10.1002/9780470517277.

M. Wiesmeier et al., « Soil organic carbon stocks in southeast Germany (Bavaria) as affected by land use, soil type and sampling depth », *Global Change Biology*, vol. 18, no 7, p. 2233–2245, juill. 2012, doi: 10.1111/J.1365-2486.2012.02699.X.

E. H. Wilson et al., « Detection of forest harvest type using multiple dates of Landsat TM imagery », *Remote Sensing of Environment*, vol. 80, no 3, p. 385–396, juin 2002, doi: 10.1016/S0034-4257(01)00318-2.

M. Xu et al., « Improving the accuracy of soil organic carbon content prediction based on visible and near-infrared spectroscopy and machine learning », *Environmental Earth Sciences*, vol. 80, no 8, p. 1–10, avr. 2021, doi: 10.1007/S12665-021-09582-X.

W. Zayani, « Estimations spatiale et temporelle des teneurs en carbone organique des sols agricoles par proxy-détection et télédétection satellitaire : Application à deux sites d'étude en Bretagne et en Tunisie centrale », Thèse, Institut Agro - Rennes Angers ; Université de Carthage (Tunisie) ; Institut national de la recherche agronomique (Tunisie), 2023, URL: <https://theses.hal.science/tel-04503499>.

H. Zayani et al., « Potential of C-Band Sentinel-1 Data for Estimating Soil Moisture and Surface Roughness in a Watershed in Western France », *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2022-July, p. 6104–6107, 2022, doi: 10.1109/IGARSS46834.2022.9883957.

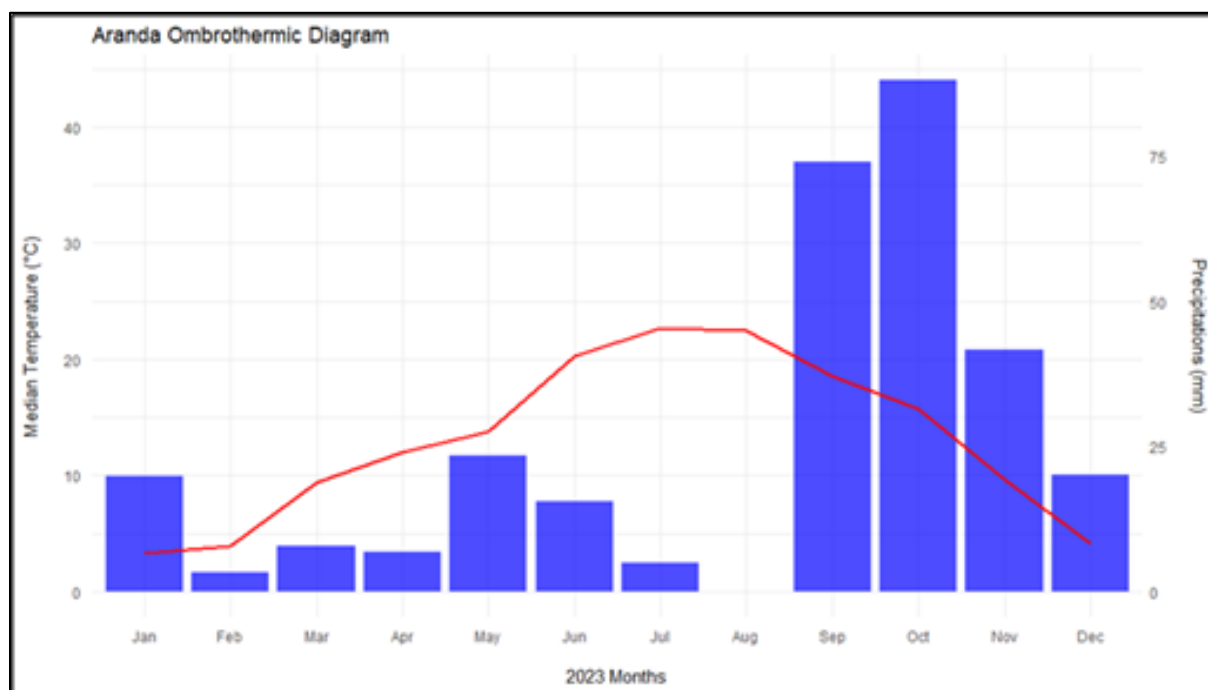
H. Zayani et al., « Using Machine-Learning Algorithms to Predict Soil Organic Carbon Content from Combined Remote Sensing Imagery and Laboratory Vis-NIR Spectral Datasets », *Remote Sensing*, vol. 15, no 17, p. 4264, août 2023, doi: 10.3390/RS15174264.

N. Zepp et al., « Optimized bare soil compositing for soil organic carbon prediction of topsoil croplands in Bavaria using Landsat », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 202, p. 287–302, août 2023, doi: 10.1016/J.ISPRSJPRS.2023.06.003.

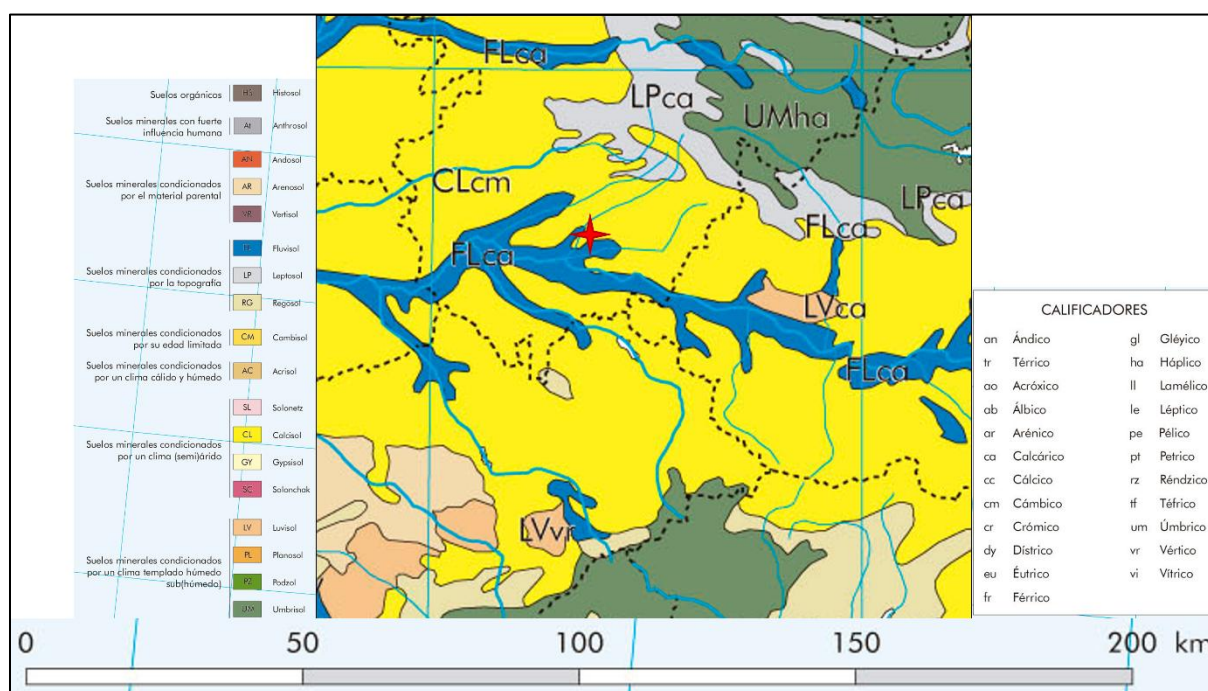
P. Zhao et al., « Quantifying soil properties relevant to soil organic carbon biogeochemical cycles by infrared spectroscopy: The importance of compositional data analysis », *Soil and Tillage Research*, vol. 231, juillet 2023, doi: 10.1016/j.still.2023.105718

Annexes

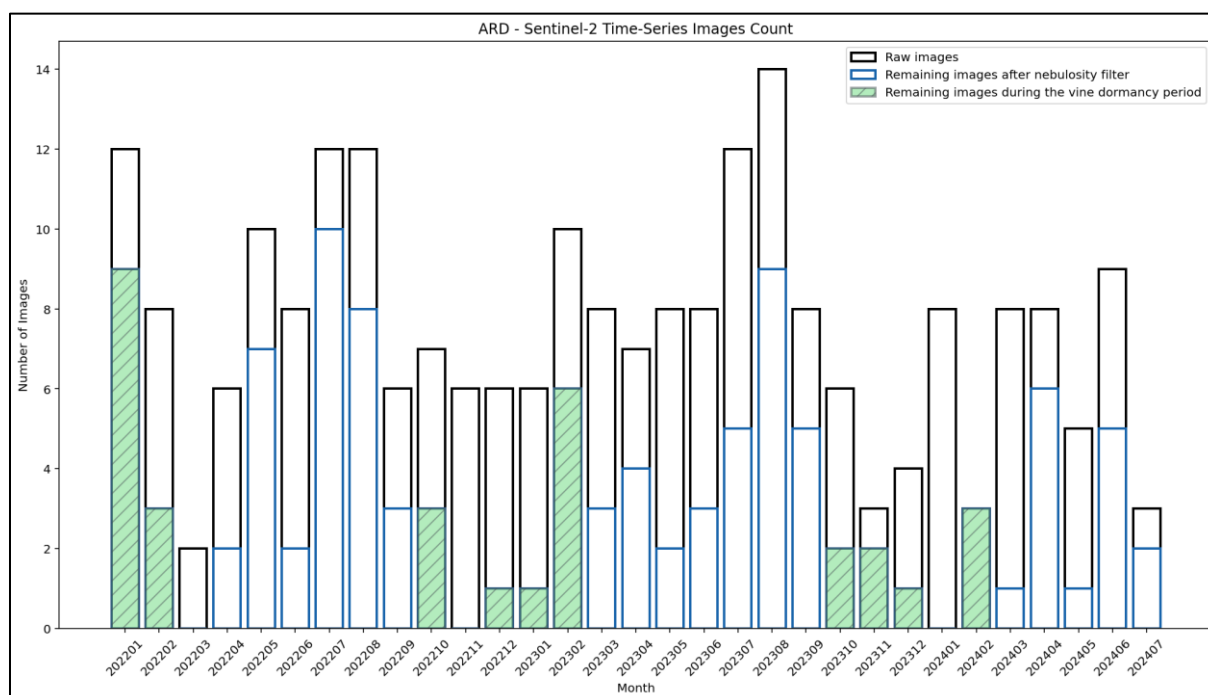
Annexe 1 – Diagramme ombrothermique d'Aranda de Duero



Annexe 2 – Carte pédologique d'Aranda de Duero



Annexe 3 – Répartition des images par mois de la banque finale d'images de la série temporelle Sentinel-2



Annexe 4 – Indices spectraux calculés à partir des images Sentinel-2

AFRI16	<i>Adjusted Forest Resistance Index 16</i>	$(B8A - 0.66 * B11)/(B8A + 0.66 * B11)$	KAUFMAN & TANRÉ 1992
AFRI21	<i>Adjusted Forest Resistance Index 21</i>	$(B8A - 0.5 * B12)/(B8A + 0.5 * B12)$	KAUFMAN & TANRÉ 1992
ARI	<i>Anthocyanin Reflectance Index</i>	$(1/B3) - (1/B5)$	GITELSON et al., 2001
ARVI2	<i>Atmospherically Resistant Vegetation Index 2</i>		KAUFMAN & TANRÉ 1992
AWEI	<i>Automated Water Extraction Index</i>	$4*(B3-B11) - (0.25*B8 + 2.75*B12)$	FEYISA et al., 2014
BI	<i>Brightness Index</i>	$\sqrt{((B2^2 + B3^2 + B4^2)/3)}$	MATHIEU et al., 1998
BRI	<i>Brown Reflectance Index</i>	$((1/B3) - (1/B5))/B6$	BROGE & LEBLANC 2000
BSI	<i>Bare Soil Index</i>	$((B12 + B4) - (B8 + B2))/((B12 + B4) + (B8 + B2))$	RIKIMARU et al., 2002
BWDRVI	<i>Broad-Wavelength Difference Vegetation Index</i>	$(0.1*B7 - B2)/(0.1*B7 + B2)$	GITELSON 2004
CCCI	<i>Canopy Chlorophyll Content Index</i>	$((B8-B5)/(B8+B5))/((B8-B4)/(B8+B4))$	BARNES et al., 2000
CI	<i>Chlorophyll Index</i>	$(B4-B3)/(B4+B3)$	GITELSON et al., 2003
CRCI	<i>Canopy Resistance Chlorophyll Index</i>	$(B11-B3)/(B11+B3)$	Index Database
EVI	<i>Enhanced Vegetation Index</i>	$2.5*(B8-B4)/(B8 + 6*B4 - 7.5*B2 + 1)$	HUETE et al., 2002
EVI2	<i>Enhanced Vegetation Index 2</i>	$2.4*(B8-B4)/(B8 + B4 + 1)$	JIANG et al., 2008
Fe2	<i>Ferrous Iron Index</i>	$(B12/B8) - (B3/B4)$	Index Database

Fe3	<i>Ferric Iron Index</i>	B3/B4	Index Database
Fe2+	<i>Ferrous Iron</i>	$(B12/B8) + (B3/B4)$	Annexe 3
Fe3+	<i>Ferric Iron</i>	B4/B3	Annexe 3
FeOx	<i>Iron Oxides Index</i>	B11/B8	Annexe 3
FeSi	<i>Ferrous-Silicate Index</i>	B12/B11	Annexe 3
FI	<i>Flavonoid Index</i>	$(2*B4 - B3 - B2)/(B3 - B2)$	PEÑUELAS et al., 1995
FO	<i>Ferrous Oxides Index</i>	B11/B8	Index Database
FS	<i>Ferrous Silicate Index</i>	B12/B11	Index Database
GARI	<i>Green Atmospherically Resistant Index</i>	$(B8 - (B3 - (B2 - B4)))/(B8 + (B3 + (B2 - B4)))$	GITELSON et al., 2002
GLI	<i>Green Leaf Index</i>	$(2*B3 - B5 - B2)/(2*B3 + B5 + B2)$	LOUHAICHI et al., 2001
GNDVI	<i>Green Normalized Difference Vegetation Index</i>	$(B8-B3)/(B8+B3)$	GITELSON et al., 1996
Gossan	<i>Gossan Index</i>	B11/B4	Index Database
GVMi	<i>Green Vegetation Moisture Index</i>	$((B8 + 0.1) - (B12 + 0.02))/((B8 + 0.1) + (B12 + 0.02))$	CECCATO et al., 2002
Hue	<i>Hue Index</i>	$\arctan((2*B5 - B3 - B2)/(30.5*(B3 - B5)))$	ESCADAFAL 1994
Maccioni	<i>Maccioni Index</i>	$(B7-B5)/(B7-B4)$	MACCIONI et al., 2001
MBI	<i>Mineralogical Bare Index</i>	$(B11 - B12 - B8)/(B11 + B12 + B8) + 0.5$	Index Database
MNDWI	<i>Modified Normalized Difference Water Index</i>	$(B3-B11)/(B3+B11)$	XU 2006
NBR	<i>Normalized Burn Ratio</i>	$(B8-B12)/(B8+B12)$	KEY & BENSON 2006
NBR2	<i>Normalized Burn Ratio 2</i>	$(B11-B12)/(B11+B12)$	CASTALDI et al., 2019
NDI5	<i>Normalized Difference Index 5</i>	$(B8-B11)/(B8+B11)$	Index Database
NDI7	<i>Normalized Difference Index 7</i>	$(B8-B12)/(B8+B12)$	Index Database
NDMI	<i>Normalized Difference Moisture Index</i>	$(B8-B11)/(B8+B11)$	WILSON & SADER 2002
NDRI	<i>Normalized Difference Red-edge Index</i>	$(B12-B4)/(B12+B4)$	Index Database
NDTI	<i>Normalized Difference Till Index</i>	$(B11-B12)/(B11+B12)$	VAN der MEER et al., 2012
NDVI	<i>Normalized Difference Vegetation Index</i>	$(B8-B4)/(B8+B4)$	ROUSE et al., 1973
NDWI	<i>Normalized Difference Water Index</i>	$(B3-B8)/(B3+B8)$	McFEETERS 1996
NSSI	<i>Normalized Soil Senescence Index</i>	$(B8A-B7)/(B8A*B7)$	Index Database
OSAVI	<i>Optimized Soil-Adjusted Vegetation Index</i>	$(1 + 0.16)*((B8-B4)/(B8+B4+0.16))$	RONDEAUX et al., 1996
OSAVI2	<i>Optimized Soil-Adjusted Vegetation Index 2</i>	$(B8-B4)/(B8+B4+0.16)$	RONDEAUX et al., 1996
PANDVI	<i>Panchromatic Normalized Difference Vegetation Index</i>	$(B8 - (B3 + B4 + B2))/(B8 + (B3 + B4 + B2))$	Index Database
PV+Blue	<i>Photochemical Reflectance Index + Blue</i>	$(B8-B4)/(B8+B4) + (B8-B2)/(B8+B2)$	ROGGE et al., 2018

PV+IR2	<i>Photochemical Reflectance Index + Infrared 2</i>	$(B8-B4)/(B8+B4) + (B8-B12)/(B8+B12)$	ZEPP et al., 2023
RedI	<i>Redness Index</i>	$(B4^2)/(B3^3)$	Index Database
RI	<i>Redness Index</i>	$(B4^2)/(B2*(B3^3))$	Annexe 3
RI2	<i>Redness Index 2</i>	$B4/(B4 + B3 + B2)$	Index Database
S2WI	<i>Sentinel-2 Water Index</i>	$(B8-B11-B12)/(B8+B11+B12)$	Index Database
SAVI	<i>Soil-Adjusted Vegetation Index</i>	$1.5*(B8-B4)/(B8+B4+0.5)$	HUETE 1988
SCI	<i>Soil Color Index</i>	$(B11-B8)/(B11+B8)$	Annexe 3
SI	<i>Structure Intensive Index</i>	$(B4-B2)/B4$	RONDEAUX et al., 1996
SIWSI	<i>Shortwave Infrared Water Stress Index</i>	$(B8A-B11)/(B8A+B11)$	Index Database
SNDVI	<i>Soil-Adjusted Normalized Difference Vegetation Index</i>	$(B8-B4)/(B8+B4+0.16)$	RONDEAUX et al., 1996
SOCI	<i>Soil Organic Carbon Index</i>	$B2/(B3*B4)$	Annexe 3
SRRBIO	<i>Simple Ratio Red/Bio</i>	$((B4 + B5 + B6)/3)/B2$	Annexe 3
STI	<i>Soil Till Index</i>	$B11/B12$	Index Database
TSAVI	<i>Transformed Soil-Adjusted Vegetation Index</i>	$1.22*((B8 - 1.22)*(B4 - 0.03))/(1.22*((B8 + B4 - 1.22)*0.03 + 0.08*(1 + 1.22^2)))$	BARET et al., 1989

Annexe 5 – Indices spectraux hyperspectraux adaptés à Sentinel-2

Fe2+	<i>Ferrous iron</i>	$\frac{r_{2145:2185}}{r_{760:860}} + \frac{r_{520:600}}{r_{630:690}}$	$\frac{B12}{B8} + \frac{B3}{B4}$	ROWAN et al., 2003
Fe3+	<i>Ferric iron</i>	$\frac{r_{630:690}}{r_{520:600}}$	$\frac{B4}{B3}$	ROWAN et al., 2003
FeOx	<i>Ferric oxides</i>	$\frac{r_{1600:1700}}{r_{760:860}}$	$\frac{B11}{B8}$	Index Database
FeSi	<i>Ferrous Silicates</i>	$\frac{r_{2145:2185}}{r_{1600:1700}}$	$\frac{B12}{B11}$	Index Database
IFe	<i>Ferric Index</i>	$\frac{(r_{630:690} - r_{520:600})}{(r_{630:690} + r_{520:600})}$	$\frac{(B4 - B3)}{(B4 + B3)}$	MADEIRA et al., 2010
IHm	<i>Hematite Index</i>	$10000 \times \frac{r_{630:690}^2}{r_{450:520} \times r_{520:600}^3}$	$10000 \times \frac{B4^2}{B2 \times B3^3}$	MADEIRA et al., 2010
IORI	<i>Iron Oxide Redness Index</i>	$\frac{r_{693}^2}{r_{447} \times r_{556}^3}$	$\frac{B4^2}{B1 \times B3^3}$	MADEIRA et al., 2010
NSMI	<i>Normalized difference Soil Moisture Index</i>	$\frac{(r_{1800} - r_{2119})}{(r_{1800} + r_{2119})}$	$\frac{(B11 - B12)}{(B11 + B12)}$	HAUBROCK et al., 2010
RI	<i>Redness Index</i>	$\frac{r_{600:700}}{(r_{600:700} + r_{500:600} + r_{400:500})}$	$\frac{B4}{(B4 + B3 + B2)}$	MADEIRA et al., 2010

SCI	<i>Soil Composition Index</i>	$\frac{(r_{1600:1700} - r_{760:860})}{(r_{1600:1700} + r_{760:860})}$	$\frac{(B11 - B8)}{(B11 + B8)}$	AL-KHAIER et al., 2003
SOCI	<i>Soil Organic Carbon Index</i>	$\frac{r_{478}}{(r_{546} \times r_{659})}$	$\frac{B2}{B3 \times B4}$	BARTHOLOMEUS et al., 2007
SRRBIO	<i>Simple Ratio Red/Blue Iron Oxide</i>	$\frac{r_{640:760}}{r_{420:480}}$	$\frac{B5}{B1}$	HEWSON et al., 2001
SRSNFM	<i>Simple Ratio SWIR1/NIR Ferrous Minerals</i>	$\frac{r_{1570:1650}}{r_{760:860}}$	$\frac{B11}{B9}$	HEWSON et al., 2001

Annexe 6 – Montage du spectroradiomètre pour l’acquisition des spectres de laboratoire des échantillons terrain



Annexe 7 – Paramètres géométriques de la courbe spectrale corrigée calculés au droit des bandes d’absorption de l’oxyde de fer (III)

La courbe spectrale des spectres labo présente des « sauts » dans le spectre au droit des changements de capteurs du spectroradiomètre ASD FieldSpec 3, entre la partie visible et proche-infrarouge (à 1'008nm) et entre l'infrarouge à ondes courtes 1 et 2 (à 1'812nm). Le spectre est donc préalablement corrigé par méthode de *splice correction*: cela consiste à interpoler linéairement le spectre brut sur un intervalle de $\pm 50\text{nm}$ (50 bandes ici) autour de chaque point de suture, afin d'harmoniser les niveaux de réflectance et d'éviter les artefacts pour le calcul des paramètres géométriques qui nécessite un spectre sans discontinuités. Le spectre ainsi lissé est ensuite traité par CR.

- Abs_{wl} : la longueur d'onde de la bande d'absorption maximale est calculée en identifiant le minimum de réflectance dans le voisinage spectral passé en entrée (un voisinage spectral englobant une longueur d'onde d'absorption relativement significative connue pour les FeOx).
- *Depth*: la profondeur d'absorption est calculée en soustrayant la réflectance minimale (au droit de Abs_{wl}) de la réflectance du continuum (égale à 1).
- *AUC*: l'aire sous la courbe d'absorption est calculée en utilisant la méthode de Simpson, qui intègre numériquement la réflectance sur le voisinage spectral passé en entrée.
- *FWHM*: la largeur de la bande d'absorption à mi-hauteur est estimée en modélisant la courbe d'absorption par une distribution normale centrée sur Abs_{wl} . La largeur est définie comme l'écart entre les points de part et d'autre de la longueur d'onde d' Abs_{wl} où la courbe atteint 50% de l'AUC.
- *Width_{2-98%}*: à l'image de la FWHM, la largeur basale de la bande d'absorption est définie comme l'écart entre les points de part et d'autre de la longueur d'onde d' Abs_{wl} où la courbe atteint 2% et 98% de l'AUC.

Annexe 8 – Tableau de composition des jeux de données

		Sentinel-2						Spectres labo					FeOx			
		Bandes	Indice 1	Indices 1-2	Indices 1-3	Indices 1-4	Indices 1-5	Indice 1	Indices 1-2	Indices 1-3	Indices 1-4	Indices 1-5	Fe _{tot}	Fe _{tot} + Fe ²⁺	Fe _{tot} + Fe ³⁺	Fe ³⁺
	d1															
	d2															
	d3															
	d4															
	d5															
	d6															
	d7															
	d8															
	d9															
	d10															
	d11															
	d12															
	d20															
	d30															
	d31															
	d32															
	d33															
Fe ³⁺ uniquement	d50												paramètres géométriques			

Annexe 9 – Statistiques de calibration des modèles prédictifs pour le SOC

Les statistiques de calibration des modèles PLSR pour le SOC (cf. Figure 11) révèlent une variabilité significative selon les prétraitements spectraux appliqués et les jeux de données utilisés. Les pré-traitements de MSC et CR se distinguent par une réduction notable du RMSE et une augmentation du RPIQ, suggérant une amélioration de la robustesse et de la précision des prédictions. La RMSE varie de 2.7 à 4.8g/kg selon les pré-traitements et les valeurs *train/test*, et le RPIQ de 1.3 à 2.15. Les meilleurs résultats, en valeur absolue et en stabilité entre *train/test* et au cours de l'incrémentation des jeux de données, sont obtenus à l'aide du pré-traitement CR. C'est celui qui sera particulièrement considéré par la suite, et c'est le seul pré-traitement appliqué pour la construction des jeux d30 à d33 avec l'intégration progressive des données de FeOx. On note également l'absence d'amélioration des performances au fur et à mesure du rajout d'indices spectraux S-2; c'est à l'ajout du premier pseudo-indice labo que la RMSE baisse et que le RPIQ augmente de manière significative. L'ajout de pseudos-indices supplémentaires ne semble pas améliorer les performances. L'ajout des données FeOx permet une légère amélioration supplémentaire des scores de prédictions, montrant l'intérêt d'ajouter ces éléments aux prédicteurs, mais aussi la complexité de la réponse spectrale mélangée induite.

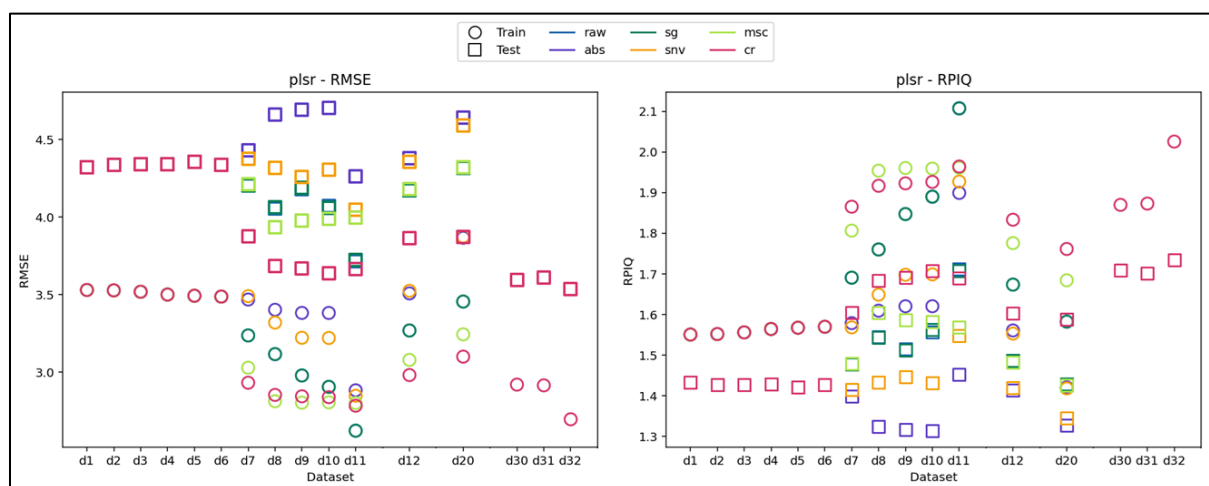


Figure 11 – Statistiques de calibration lors de la validation des modèles PLSR SOC (moyennes des performances sur les trois répétitions)

Les statistiques de calibration des modèles RF pour le SOC (cf. Figure 12) mettent en évidence une variabilité moins marquée selon les prétraitements spectraux. La RMSE, comprise entre 1 et 4.9g/kg, et le RPIQ, compris entre 1.3 et 7.8, montrent des niveaux de précision plus élevés avec cette approche non-linéaire par rapport à la PLSR. On observe toujours l'absence d'amélioration des performances pendant l'intégration progressive des indices S-2 mais le bond à l'ajout du premier pseudo-indice labo, confirmant la complexité de la réponse spectrale et l'importance des pré-traitements appliqués aux spectres labo. Le dataset d20, ne comportant que le premier pseudo-indice labo, montre des scores similaires à d12, montrant le peu d'apport que représentent les bandes S-2. Enfin, l'ajout des données de FeOx semble apporter plus de bruit au modèle qu'une réelle amélioration des scores.

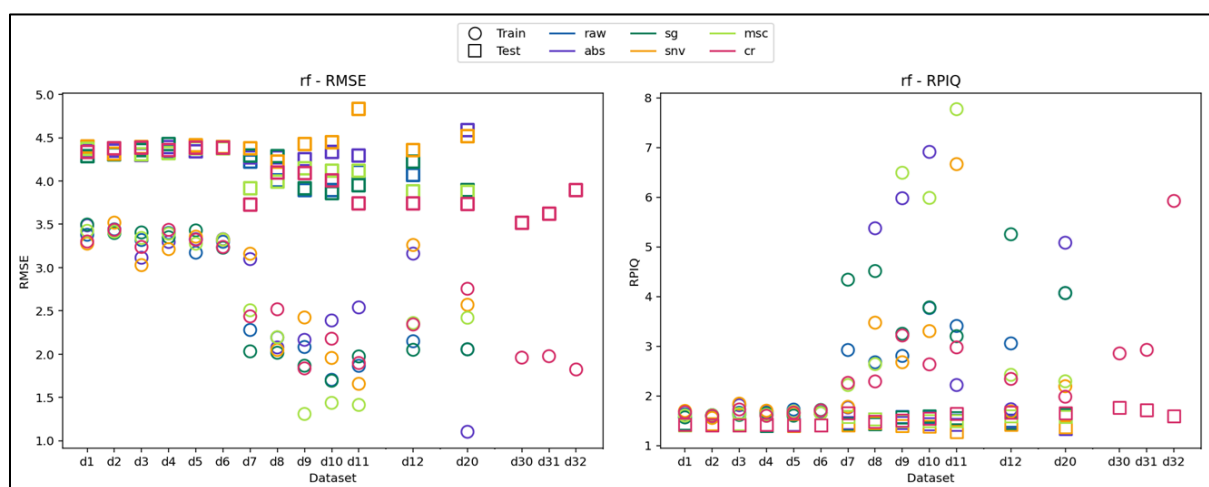


Figure 12 – Statistiques de calibration lors de la validation des modèles RF SOC (moyennes des performances sur les trois répétitions)

Les statistiques de calibration des modèles SVR pour le SOC (cf. Figure 13) révèlent une variabilité modérée selon les prétraitements spectraux appliqués, avec des valeurs de RMSE comprises entre 2.3 et 4.7g/kg et des RPIQ allant de 1.2 à 2.5. Comme observé précédemment avec les modèles RF et PLSR, l'intégration progressive des indices S-2 ne se traduit pas par une amélioration des performances, tandis que l'ajout du premier pseudo-indice de laboratoire induit une réduction notable du RMSE et une augmentation du RPIQ, confirmant l'apport crucial des spectres labo. Cependant, l'incorporation des données relatives aux oxydes de fer, en particulier pour le jeu de données d33 centré sur le Fe^{3+} , ne semble pas améliorer significativement les scores et introduit même une variabilité supplémentaire, suggérant que la relation entre Fe^{3+} et SOC pourrait être plus complexe que prévue ou masquée par d'autres facteurs.

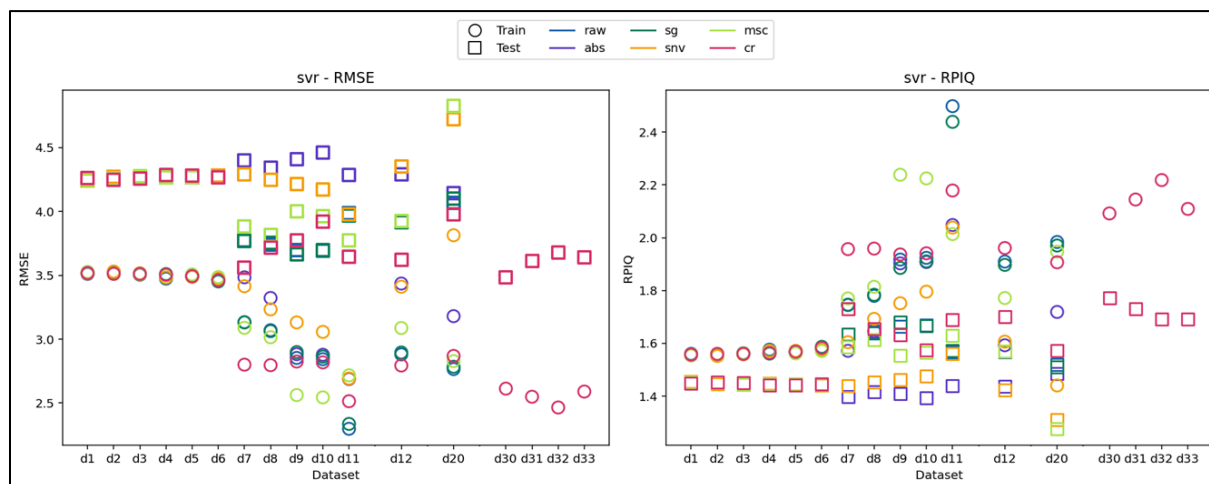


Figure 13 – Statistiques de calibration lors de la validation des modèles SVR SOC (moyennes des performances sur les trois répétitions)

Les statistiques de calibration des modèles DNN pour le SOC (cf. Figure 14) démontrent une stabilité remarquable des performances, avec des valeurs de RMSE comprises entre 3.25 et 3.42g/kg et des RPIQ oscillant entre 1.68 et 1.82, indépendamment des prétraitements spectraux appliqués. Contrairement aux autres approches testées, l'intégration progressive des indices S-2 ou l'ajout de pseudo-indices de laboratoire ne génère qu'une faible variation des scores, indiquant une robustesse intrinsèque du modèle face à la diversité des données d'entrée. Aussi, le jeu de données d33, qui intègre exclusivement les données de Fe^{3+} , ne permet pas d'amélioration notable des performances et confirme que cette variable, bien que théoriquement pertinente pour expliquer la variabilité du SOC, ne suffit pas à elle seule à renforcer la précision prédictive du modèle. L'écart stable des performances entre *train/test* suggère un sur-apprentissage du modèle, malgré la validation croisée par k-fold et la fonction d'*early stopping*.

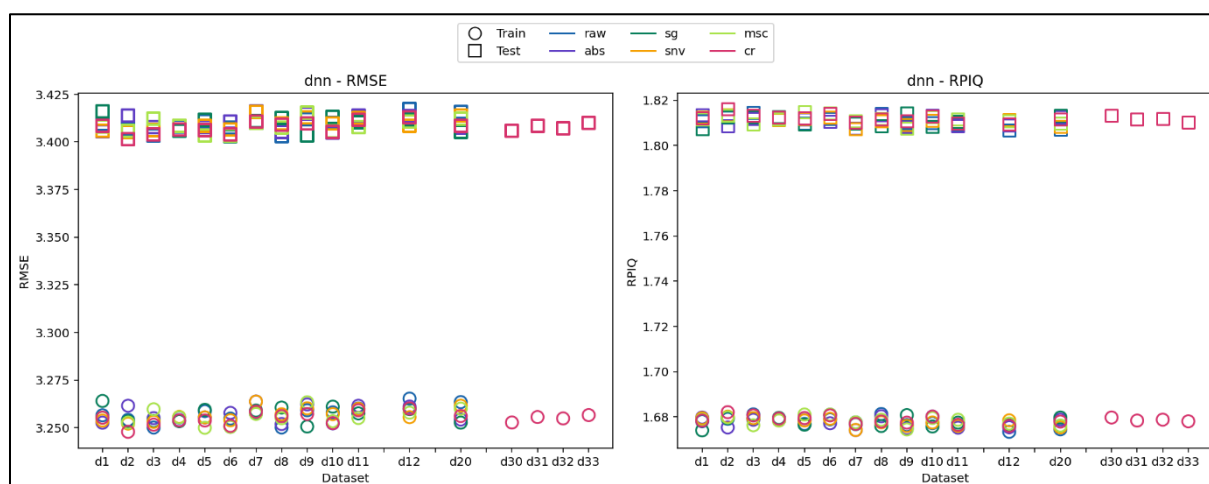


Figure 14 – Statistiques de calibration lors de la validation des modèles DNN SOC (moyennes des performances sur les trois répétitions)

Annexe 10 – Statistiques de calibration des modèles prédictifs pour les FeOx

Les statistiques de calibration des modèles PLSR pour la prédiction du Fe^{3+} (cf. Figure 15) révèlent une variabilité modérée des résultats selon les prétraitements spectraux appliqués, avec des valeurs de RMSE comprises entre 0.028 et 0.036 et des RPIQ entre 1.15 et 1.35. Contrairement aux observations réalisées pour le SOC, le prétraitement SNV semble offrir une meilleure stabilité des performances, notamment pour les jeux de données d12 et d20. On note cependant une absence d'amélioration significative des scores avec l'intégration progressive des indices S-2, tandis que l'ajout des pseudo-indices labo ne se traduit pas par une réduction marquée du RMSE ni par une augmentation du RPIQ.

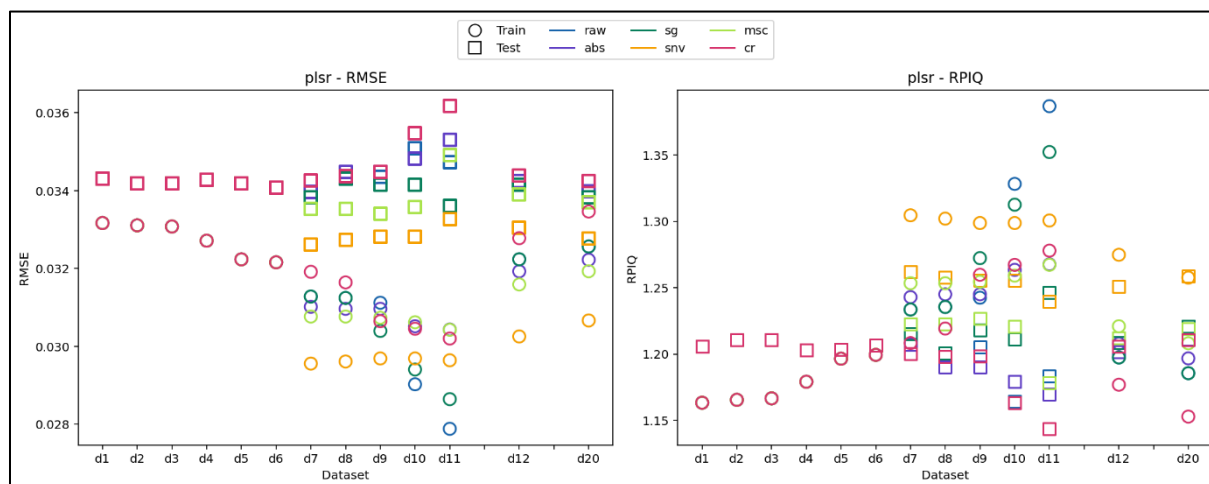


Figure 15 – Statistiques de calibration lors de la validation des modèles PLSR FeOx (moyennes des performances sur les trois répétitions)

Les statistiques de calibration des modèles RF pour la prédiction du Fe^{3+} (cf. Figure 16) montrent une variabilité limitée des résultats selon les prétraitements spectraux, avec des valeurs de RMSE comprises entre 0.02 et 0.04 et des RPIQ allant de 1 à 2. Le SNV permet d'atteindre des performances légèrement supérieures, notamment pour les jeux de données d12 et d20, où le RPIQ dépasse 1.8. Cependant, à l'instar des observations réalisées avec les modèles PLSR, l'intégration progressive des indices S-2 ne se traduit pas par une amélioration des performances. L'ajout de pseudo-indices de laboratoire ne semble pas non plus induire de gain significatif en termes de RMSE ou de RPIQ, confirmant la difficulté à améliorer la prédiction du Fe^{3+} à partir des données spectrales disponibles.

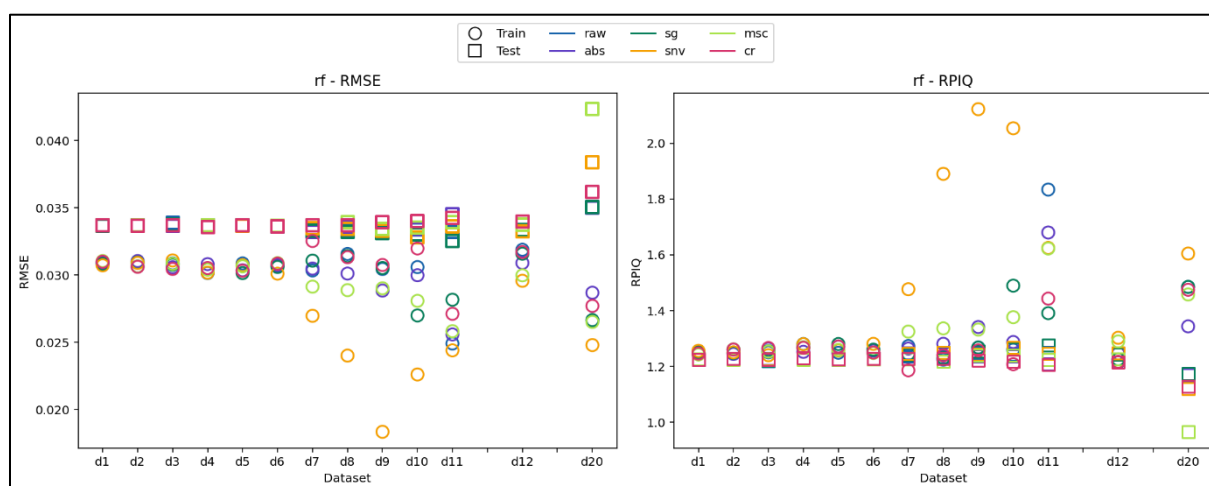


Figure 16 – Statistiques de calibration lors de la validation des modèles RF FeOx (moyennes des performances sur les trois répétitions)

Les statistiques de calibration des modèles SVR pour la prédiction du Fe^{3+} (cf. Figure 17) montrent une variabilité limitée des résultats selon les prétraitements spectraux, avec des valeurs de RMSE comprises entre 0.02 et 0.04 et des RPIQ allant de 1 à 2. Le SNV permet d'atteindre des performances légèrement supérieures, notamment pour les jeux de données d12 et d20, où le RPIQ dépasse 1.8. Cependant, à l'instar des observations réalisées avec les modèles PLSR, l'intégration progressive des indices S-2 ne se traduit pas par une amélioration des performances. L'ajout de pseudo-indices de laboratoire ne semble pas non plus induire de gain significatif en termes de RMSE ou de RPIQ, confirmant la difficulté à améliorer la prédiction du Fe^{3+} à partir des données spectrales disponibles.

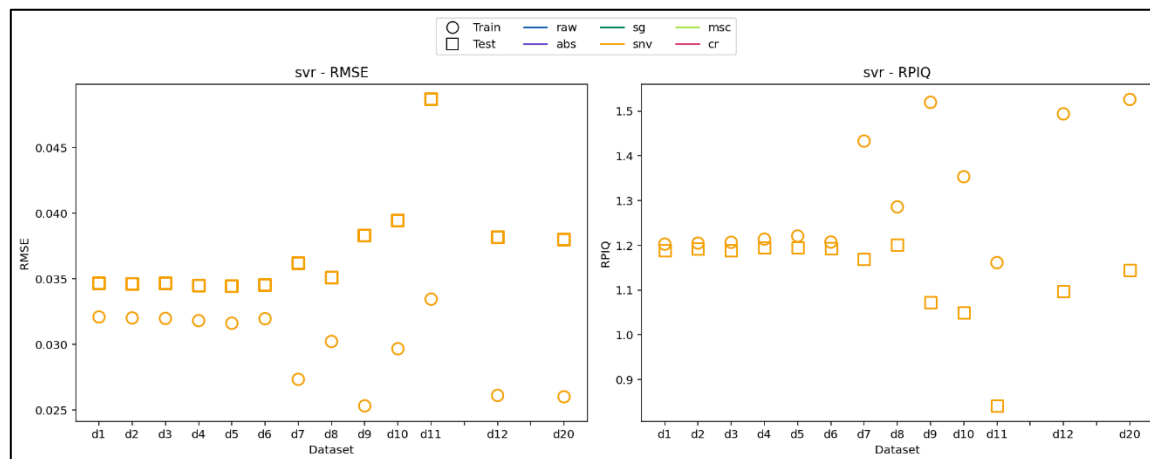


Figure 17 – Statistiques de calibration lors de la validation des modèles SVR FeOx (moyennes des performances sur les trois répétitions)

Les statistiques de calibration des modèles DNN pour la prédiction du Fe^{3+} (cf. Figure 18) révèlent une variabilité quasi-nulle avec une RMSE oscillant entre 0.01 et 0.025g/kg et un RPIQ entre 1.8 et 3.7. Contrairement aux autres approches testées, l'intégration progressive des indices S-2 ou l'ajout de pseudo-indices de laboratoire ne génère qu'une faible variation des scores, indiquant une robustesse intrinsèque du modèle face à la diversité des données d'entrée. Le SNV permet d'atteindre des performances légèrement supérieures, notamment pour les jeux de données d12 et d20, où le RPIQ dépasse 2.5. Cependant, à l'instar des observations réalisées avec les modèles PLSR, l'intégration progressive des indices S-2 ne se traduit pas par une amélioration des performances. L'ajout de pseudo-indices de laboratoire ne semble pas non plus induire de gain significatif en termes de RMSE ou de RPIQ, confirmant la difficulté à améliorer la prédiction du Fe^{3+} à partir des données spectrales disponibles.

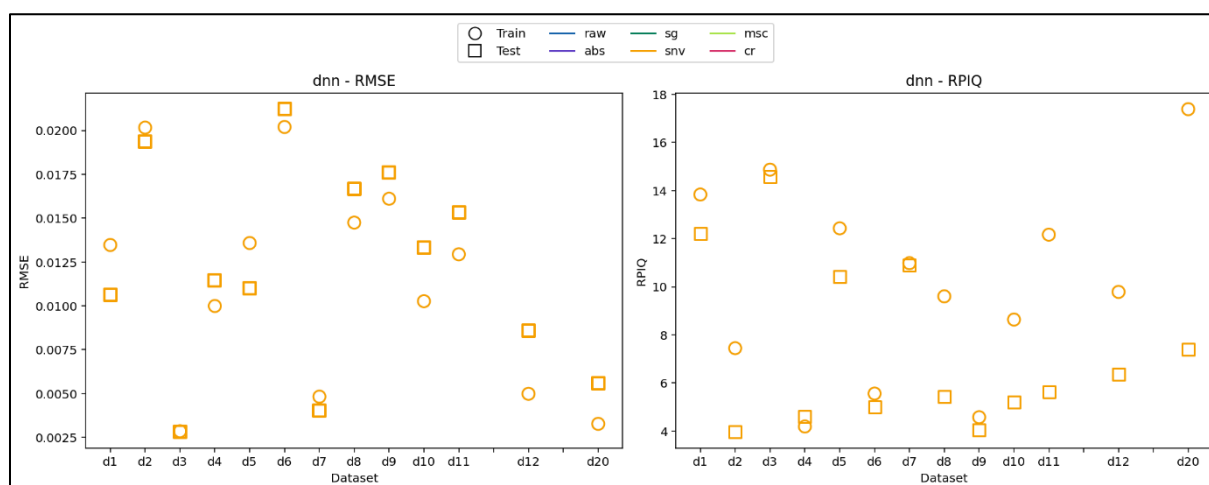
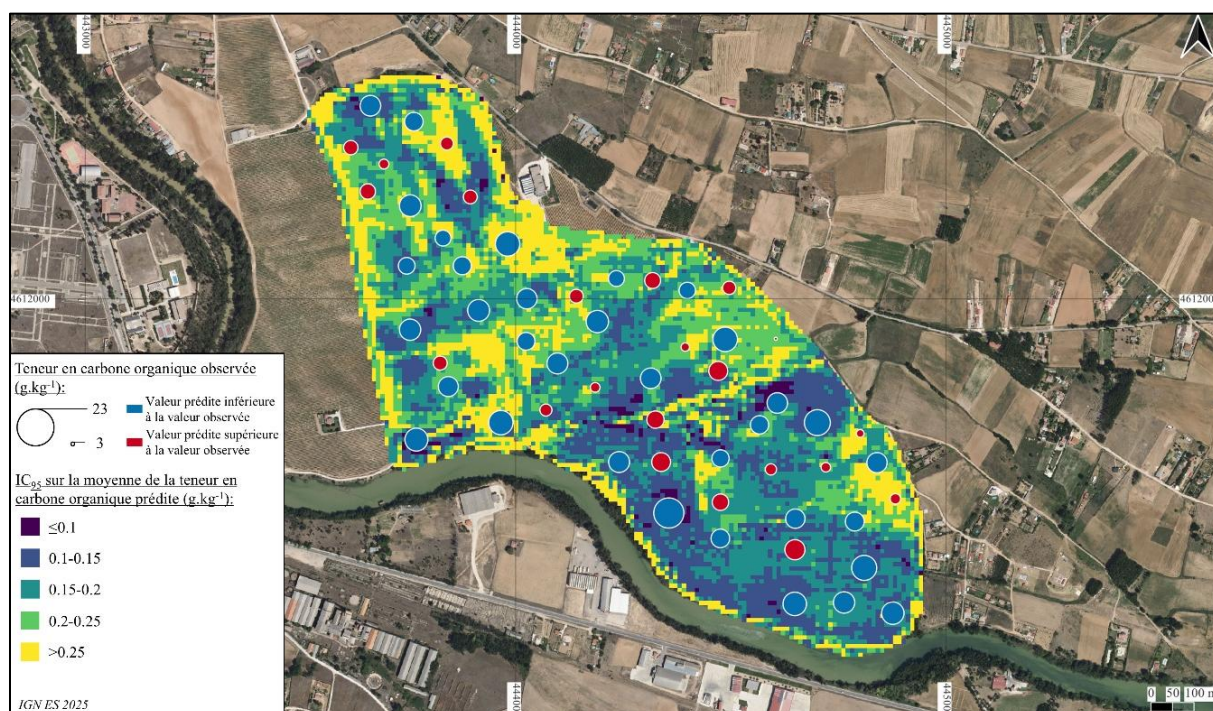
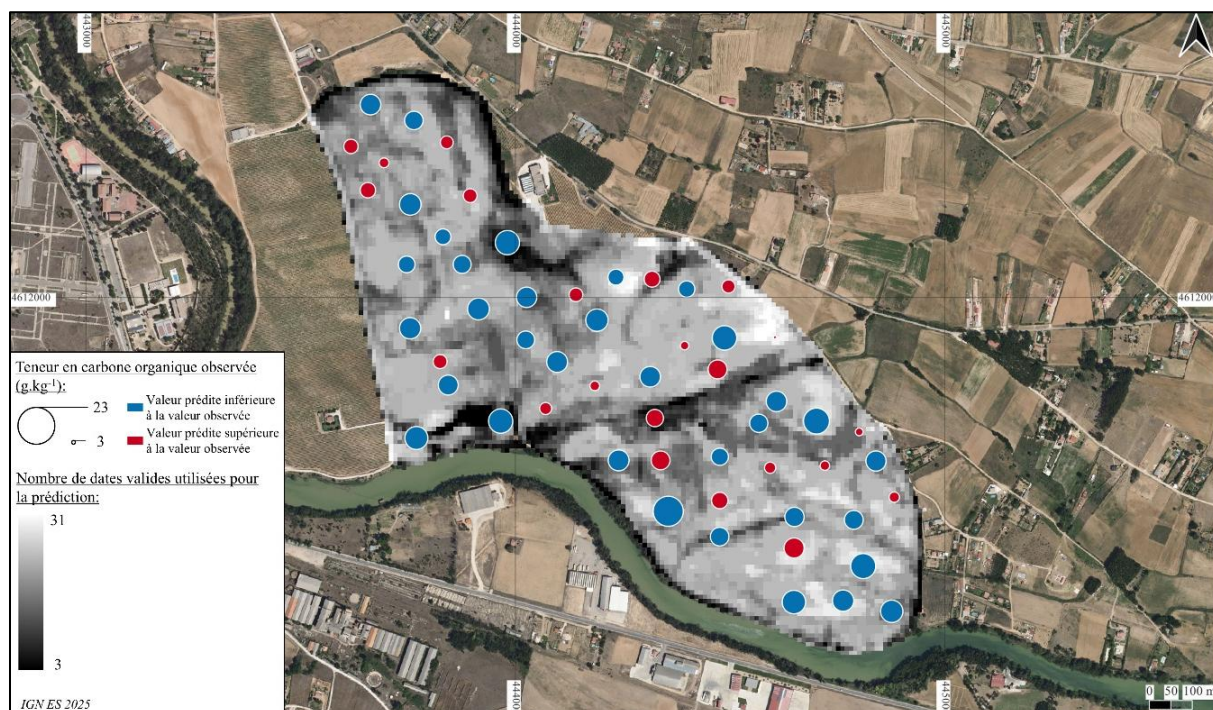


Figure 18 – Statistiques de calibration lors de la validation des modèles DNN FeOx (moyennes des performances sur les trois répétitions)

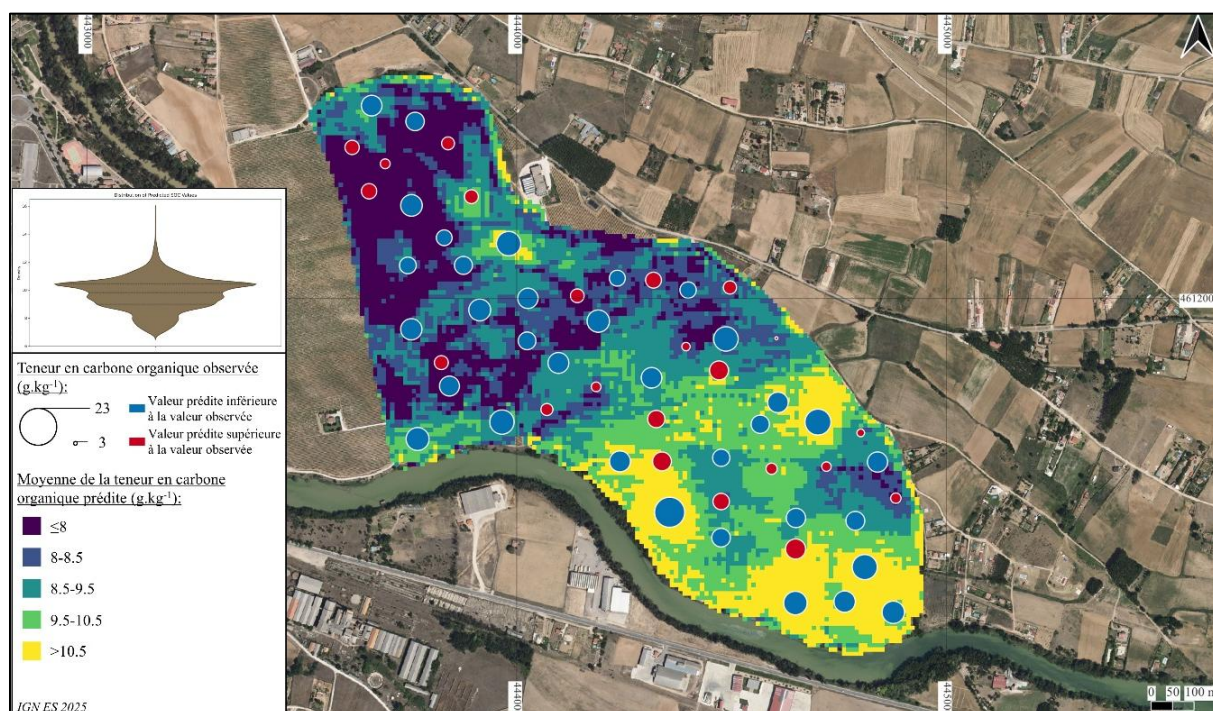
Annexe 11 – Carte de l'indice de confiance au seuil de 95% des prédictions de la teneur moyenne en SOC à l'aide de la série temporelle S-2



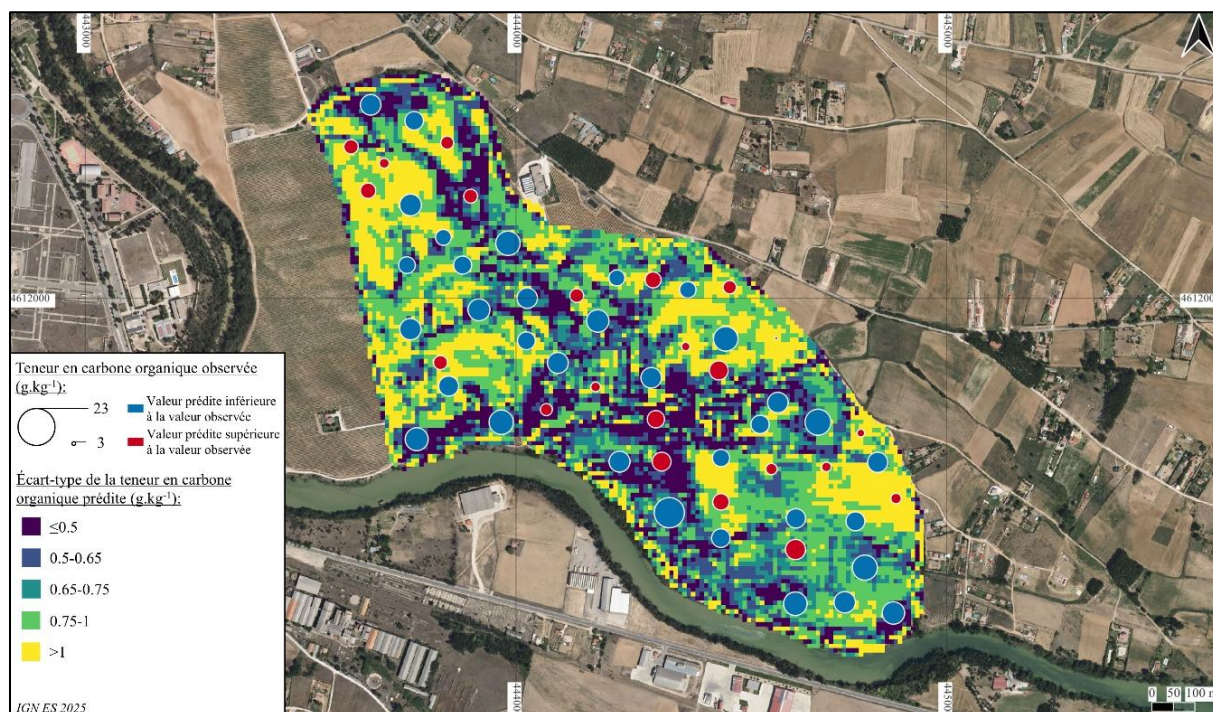
Annexe 12 – Carte du nombre de dates en sol par pixel utilisées pour les prédictions de la teneur moyenne en SOC à l'aide de la série temporelle S-2



Annexe 13 – Carte des prédictions de la teneur moyenne en SOC sur le vignoble d'Aranda à l'aide de la date la plus en sol nu



Annexe 14 – Carte de l'écart-type des prédictions de la teneur moyenne en SOC sur le vignoble d'Aranda à l'aide de la date la plus en sol nu



Annexe 15 – Carte du NSSI

