

Filière Systèmes industriels

Orientation Infotronics

Travail de bachelor Diplôme 2025

Mathieu Berrut

*Imagerie hyperspectrale pour l'analyse de
traits fonctionnels de plantes*



Professeur
Joseph Moerschell



Expert
Christophe Kinnard



Date de la remise du rapport
14.08.2025

Filière / Studiengang SYND	Année académique / <i>Studienjahr</i> 2024-25	No TB / Nr. BA IT/2025/10
Mandant / Auftraggeber	Etudiant·e / <i>Student/in</i> Mathieu Berrut	Lieu d'exécution / <i>Ausführungsort</i>
<input type="checkbox"/> HES—SO Valais-Wallis <input type="checkbox"/> Industrie <input checked="" type="checkbox"/> Etablissement partenaire <i>Partnerinstitution</i>		<input type="checkbox"/> HES—SO Valais-Wallis <input type="checkbox"/> Industrie <input checked="" type="checkbox"/> Etablissement partenaire <i>Partnerinstitution</i>
Travail confidentiel / <i>vertrauliche Arbeit</i>	Expert·e / <i>Experte/Expertin</i> (nom, prénom, E-Mail / Name, Vorname, E-Mail) Christophe Kinnard, christophe.kinnard@uqtr.ca	
<input type="checkbox"/> oui / ja <input checked="" type="checkbox"/> non / nein		

Titre / Titel**Imagerie hyperspectrale pour l'analyse de traits fonctionnels de plantes****Description / Beschreibung**

Les traits fonctionnels de plantes caractérisent leur état de santé et de croissance. Il est important de les connaître pour optimiser des cultures, ou encore déterminer l'influence de l'environnement, p.ex. du changement climatique sur la végétation.

L'observation de traits fonctionnels tels que la masse végétale, la hauteur des plantes, le contenu d'eau et de chlorophylle peut être fait à partir d'images hyperspectrales, où chaque pixel est un spectre de lumière.

Le but de ce projet est de réaliser une caméra hyperspectrale de faible coût pour des observations dans le visible et l'infrarouge proche, et qui peut potentiellement être embarquée sur un drone.

Des observations de cultures seront faites, et les résultats analysés, pour valider la méthode de mesure.

Objectifs du projet individuel (Pr4) / Ziele des individuellen Projekts (Pr4)

- Familiarisation avec les mesures de phénotype de plantes
- Réalisation d'une caméra hyperspectrale basée sur un Raspberry Pi
- Intégration et tests de la caméra en laboratoire

Objectifs du travail de bachelor (TB) / Ziele der Bachelorarbeit (TB)

- Observation de cultures avec la caméra hyperspectrale
- Conception et réalisation d'une application d'extraction de traits fonctionnels
- Observation de surfaces d'eau avec la caméra hyperspectrale
- Etude et développement d'une application avec caméra hyperspectrale embarquée sur drone.

Délais / Termine

Démarrage du projet individuel (Pr4)
Start des individuellen Projekts (Pr4)

17.02.2025

Présentation du projet individuel (Pr4)
Präsentation:des individuellen Projekts (Pr4)

02.05.2025

Démarrage du travail de bachelor
Start der Bachelorarbeit:

19.05.2025

Remise du rapport final / *Abgabe des Schlussberichts:*
14.08.2025, 12:00

Expositions et Pitch / *Ausstellungen und Pitch der Bachelorarbeiten:*

22.08.2025 – HEI

25.08.2025 – Monthey

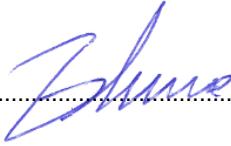
28.08.2025 – Visp

Défense orale / *Mündliche Verfechtung:*

Semaines/Wochen 36-37 (01-12.09.2025)

Signature ou visa / Unterschrift oder Visum

Responsable de l'orientation /
Leiter/in der Vertiefungsrichtung:



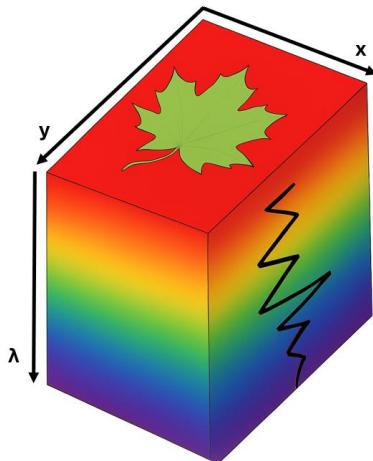
.....

¹ Etudiant·e / Student/in:



.....

¹ Par sa signature, l'étudiant·e s'engage à respecter strictement la directive DI.1.2.02.07 « Travail de bachelor ». Durch seine Unterschrift verpflichtet sich der/die Student/in, sich an die Richtlinie DI.1.2.02.07 „Bachelorarbeit“ zu halten.



Travail de diplôme | édition 2025 |

Filière
Systèmes industriels

Domaine d'application
Infotonics

Professeur responsable
Joseph Moerschell
Joseph.moerschell@hevs.ch

Partenaire
Université du Québec à Trois-Rivières (UQTR), Canada

Imagerie hyperspectrale pour l'analyse de traits fonctionnels de plantes

 Diplômant/e Mathieu Berrut

Objectif du projet

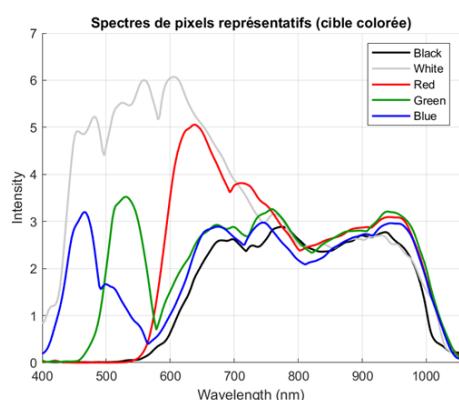
Évaluer le potentiel de l'imagerie hyperspectrale sur les végétaux à l'aide d'une caméra low-cost basée sur Raspberry Pi, par le traitement des spectres de réflectance, la modélisation statistique et l'apprentissage automatique.

Méthodes | Expériences | Résultats

Ce projet, réalisé au sein d'un laboratoire de recherche à l'Université du Québec à Trois-Rivières (Canada), a consisté à mettre en œuvre une caméra hyperspectrale low-cost basée sur un Raspberry Pi et un réseau de diffraction, puis à l'utiliser pour collecter près de 500 images hyperspectrales de feuilles d'érable. Les acquisitions ont été réalisées dans un dispositif contrôlé, puis les spectres ont été traités pour calculer des indices spectraux reconnus (NDVI, PRI, etc.). Toutes les données ont été intégrées dans une base relationnelle SQL, structurée pour permettre l'analyse statistique ainsi que la modélisation prédictive (ML). Des méthodes de réduction de dimension (PCA, PLS) et des algorithmes de machine learning (RF, SVM, GPR) ont été appliqués. Les résultats mettent en évidence une variabilité spectrale inter- et intraspécifique. Ils montrent également une capacité à modéliser certains traits avec des R^2 supérieurs à 50 %. L'apprentissage automatique a aussi montré de bonnes performances, avec des erreurs de prédiction inférieures à 30 %. Ce travail apporte des résultats encourageants, soulignant la faisabilité de l'imagerie hyperspectrale économique pour l'étude des végétaux.



Caméra hyperspectrale low-cost construite à partir d'un Raspberry Pi et d'un réseau de diffraction, utilisée pour l'acquisition de données végétales.



Spectres extraits d'un hypercube capturé sur une cible colorée (noir, blanc, rouge, vert, bleu) utilisés pour la validation spectrale du système.

Intégration de la durabilité

Ce travail de Bachelor répond à des défis techniques tout en s'intégrant dans une démarche durable. Il satisfait plusieurs objectifs de développement durable définis par l'Organisation des Nations Unies (ONU). [1]

Objectif 2 - Faim "zéro"

Dans certaines régions du monde l'accès à une agriculture productive est limité. L'imagerie hyperspectrale appliquée à la végétation offre une solution prometteuse pour surveiller rapidement l'état de santé des cultures en identifiant certains traits fonctionnels. Ces données permettent d'optimiser les pratiques agricoles comme l'irrigation, la détection de stress ou les apports nutritionnels, contribuant ainsi à réduire les pertes et améliorer les rendements.

Objectif 9 - Industrie, innovation et infrastructure

Ce projet repose sur la mise en œuvre d'un dispositif open-source à faible coût, basé sur des composants accessibles tels qu'un Raspberry et l'impression 3D. Cette approche permet de démocratiser l'innovation technologique et de rendre ces outils accessibles à une large communauté académique. L'accessibilité du système favorise l'égalité d'accès aux outils de recherche.

Objectif 12 - Consommation et production responsables

L'imagerie hyperspectrale est une méthode non destructive d'analyse de la végétation. Les méthodes classiques nécessitent parfois des prélèvements ou l'utilisation de réactifs chimiques. La caméra hyperspectrale permet de collecter des données sans altérer les écosystèmes. Elle contribue ainsi à réduire l'impact environnemental et à limiter la production de déchets.

Objectif 13 - Lutte contre les changements climatiques

En facilitant la détection de stress environnementaux (sécheresse, carences, maladies) sur de vastes zones de végétation, l'imagerie hyperspectrale permet d'anticiper les impacts du changement climatique. Ces données peuvent être utilisées pour guider les actions environnementales. L'approche contribue ainsi à une meilleure préservation de l'environnement face aux enjeux climatiques.

Table des matières

1	Introduction	3
1.1	Contexte et enjeux	3
1.2	Problématique.....	3
1.3	Objectifs du projet	3
1.4	Contexte institutionnel.....	3
1.5	Organisation du rapport.....	4
2	Etat de l'art	5
2.1	Méthodes d'analyse	5
2.2	Matériel hyperspectral.....	5
3	Matériel.....	6
3.1	Dispositif d'acquisition	6
3.2	Système hyperspectral	6
3.2.1	Architecture matérielle	7
3.2.2	Architecture logicielle	7
3.2.3	Composants et budget	8
3.2.4	Validation du fonctionnement de la caméra.....	8
3.2.5	Principe de fonctionnement de la caméra.....	10
4	Séquence de traitement des données hyperspectrales.....	12
4.1	Acquisition de l'image hyperspectrale	12
4.2	Extraction d'un spectre caractéristique	12
4.3	Correction des effets d'éclairage (normalisation)	13
4.4	Calcul des indices spectraux.....	14
4.5	Envoi vers une base de données	14
5	Contexte technique et organisationnel.....	16
5.1	Organisation générale du projet	16
5.2	Organisation par session de données	17
5.3	Base de données	19
5.3.1	Technologie	19
5.3.2	Description du schéma relationnel	19
5.3.3	Détail des tables et des champs.....	21
5.3.4	Sources et alimentation de la base de données	22
6	Analyse des données hyperspectrales	23
6.1	Introduction	23
6.2	Approche 1 - Analyse de la variabilité inter- et intra-spécifique.....	23

6.2.1	Objectif	23
6.2.2	Jeu de données	23
6.2.3	Méthodes	24
6.3	Approche 2 - Corrélation et modélisation des traits fonctionnels.....	27
6.3.1	Objectif.....	27
6.3.2	Jeu de données	27
6.3.3	Méthodes	28
6.4	Approche 3 - Corrélations et modélisation de traits fonctionnels archivés	32
6.4.1	Objectif.....	32
6.4.2	Jeu de données	32
6.4.3	Méthodes	32
7	Résultats	33
7.1	Résultats - Approche 1	33
7.1.1	ACP - Analyse en composante principales	33
7.1.2	Visualisation de PC1 selon différents niveaux.....	34
7.1.3	Analyse de la variance.....	35
7.1.4	Apprentissage automatique supervisé (classification).....	36
7.2	Résultats - Approche 2	39
7.2.1	Corrélation de Pearson.....	39
7.2.2	Modélisation linéaire multivariée des traits	41
7.2.3	Apprentissage automatique supervisé (régression)	43
7.3	Résultats - Approche 3	47
7.3.1	Corrélation de Pearson.....	47
7.3.2	Modélisation linéaire multivariée des traits	48
8	Discussion.....	50
8.1	Interprétations des principaux résultats	50
8.1.1	Approche 1 - Analyse de la variabilité inter- et intra-spécifique.....	50
8.1.2	Approche 2 - Corrélations et modélisation des traits fonctionnels	50
8.1.3	Approche 3 - Corrélations et modélisation de traits fonctionnels archivés	52
8.2	Limites du projet	52
8.3	Perspectives	52
9	Conclusion	54
10	Références.....	55
11	Annexes	57

1 Introduction

1.1 Contexte et enjeux

L'imagerie hyperspectrale est une technologie qui permet de capturer la lumière réfléchie par un objet à travers plusieurs centaines de bandes spectrales. Contrairement aux caméras classiques RGB ou multispectrales, elle génère un spectre complet pour chaque pixel de l'image capturée et contient ainsi une information plus riche. L'élément capturé est un hypercube qui est une matrice 3D ayant à la fois une dimension spatiale (x et y) et une dimension spectrale (λ). Un exemple sur une feuille végétale est présenté dans la Figure 1.1-1.

Cette technologie suscite un intérêt croissant dans le domaine de l'environnement car elle permet d'accéder à des caractéristiques internes des plantes sans les détruire.

Dans un contexte marqué par les changements climatiques et la pression sur les ressources naturelles, il est crucial de comprendre le fonctionnement des plantes afin de préserver au mieux l'environnement. Pouvoir mesurer rapidement et à grande échelle l'état physiologique des végétaux devient un réel enjeu.

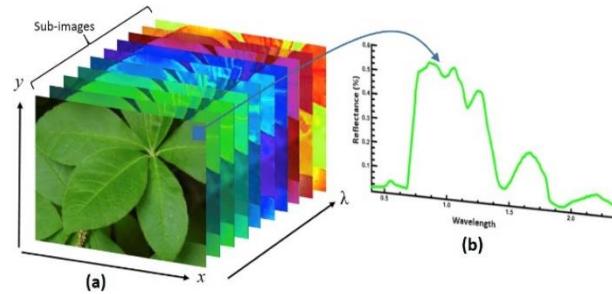


Figure 1.1-1 : Exemple d'un hypercube acquis à partir d'une feuille (a). Spectre de réflectance d'un pixel particulier (b). (Source : Mishra et al. (2017) [2])

1.2 Problématique

Les méthodes classiques de mesure des traits fonctionnels végétaux sont parfois longues, coûteuses ou limitées à un petit nombre d'échantillons. De plus, elles impliquent souvent un prélèvement manuel.

L'imagerie hyperspectrale pourrait surpasser ces limitations en fournissant un grand volume de données en peu de temps à l'aide d'un outil unique. À terme, une caméra hyperspectrale embarquée sur un drone pourrait permettre une acquisition à large échelle. Cependant, avant d'envisager ce type d'application, il est nécessaire d'en évaluer le potentiel.

1.3 Objectifs du projet

L'objectif principal de ce travail de Bachelor est d'explorer le potentiel de l'imagerie hyperspectrale pour l'analyse de feuilles d'érable à l'aide d'une caméra hyperspectrale *low-cost*. Le projet repose sur la mise en œuvre d'un dispositif open-source basé sur un Raspberry Pi et un réseau de diffraction. Des acquisitions d'images hyperspectrales ont été réalisées en conditions contrôlées sur des feuilles d'érable. Les données ont été organisées dans une base relationnelle, puis analysées à l'aide de méthodes statistiques et d'apprentissage automatique pour évaluer leur capacité à refléter certains traits physiologiques.

1.4 Contexte institutionnel

Ce travail de Bachelor a été réalisé dans le cadre d'un échange international en collaboration avec l'Université du Québec à Trois-Rivières (UQTR) au Canada. Le projet a été mené de mi-mai à mi-août 2025 au sein d'un laboratoire spécialisé en sciences de l'environnement à l'UQTR. Une équipe

locale réalisait en parallèle des mesures de traits fonctionnels sur des feuilles d'érables, offrant un contexte idéal pour évaluer l'intérêt des données hyperspectrales vis-à-vis des végétaux.

Il convient de noter que la donnée initiale du projet diverge partiellement des objectifs présentés ici. Dans le contexte d'un travail à l'étranger, les possibilités réelles étaient encore floues au moment de la rédaction initiale. Ce projet s'inscrit dans une démarche exploratoire : les objectifs se sont ajustés au fil des premières expérimentations.

1.5 Organisation du rapport

Le rapport est structuré de manière à suivre la démarche logique réalisée tout au long du projet, depuis la mise en œuvre de la caméra hyperspectrale jusqu'à l'analyse des résultats.

Le chapitre 2 - Etat de l'art présente une base théorique et technique de l'imagerie hyperspectrale, les méthodes d'analyse courantes ainsi qu'un aperçu du matériel disponible.

Le chapitre 3 - Matériel décrit en détail le dispositif d'acquisition et la caméra hyperspectrale utilisée (caractéristiques, budget, etc.), ainsi qu'un test de validation accompagné d'une brève description de son fonctionnement.

Le chapitre 4 - Séquence de traitement des données hyperspectrales détaille le processus complet de traitement des hypercubes : extraction des spectres, normalisation, calcul des indices spectraux et organisation des données.

Le chapitre 5 - Contexte technique et organisationnel explique la structuration du projet et sa mise en œuvre en précisant les outils utilisés, ainsi que le fonctionnement de la base de données relationnelle.

Le chapitre 6 - Analyse des données hyperspectrales expose les différentes approches analytiques utilisées pour interpréter les données, ainsi que les méthodes statistiques et d'apprentissage automatique utilisées à cette fin (analyse de variance, corrélations, régression multivariée, prédiction supervisée, etc.).

Le chapitre 7 - Résultats présente les résultats obtenus pour chaque approche d'analyse, accompagnés de visualisations.

Le chapitre 8 - Discussion propose une analyse critique des résultats, identifie les limites du projet et suggère des pistes d'amélioration ou d'extensions possibles.

Le chapitre 9 - Conclusion résume le travail réalisé durant le projet, les résultats clés obtenus et conclut avec une mise en perspective du potentiel de l'imagerie hyperspectrale low-cost.

Les chapitres 10 - Références et 11 - Annexes regroupent respectivement les références bibliographiques utilisées et décrivent les annexes complémentaires au projet (code, base de données, fichiers nécessaires à la mise en œuvre de la caméra, etc.).

2 Etat de l'art

2.1 Méthodes d'analyse

Pour tirer des informations utiles des images hyperspectrales des végétaux afin d'en déterminer les traits fonctionnels, il est courant de calculer des indices spectraux comme le PRI, le WBI ou le NDVI qui sont des rapports entre l'intensité de différentes bandes spectrales et qui permettent d'estimer rapidement des paramètres comme la chlorophylle, l'eau ou la biomasse. Ces indices sont faciles d'utilisation mais ils montrent leurs limites dans certains cas, par exemple quand les conditions lumineuses changent ou dans des milieux très variés. [3], [4]

Cependant, au-delà des indices spectraux, certaines études montrent que l'on peut utiliser directement les signatures hyperspectrales pour estimer certains traits fonctionnels comme la capacité photosynthétique des feuilles (V_{cmax}), un indicateur important de l'état de santé des plantes. Pour cela, différents modèles sont utilisés comme des méthodes statistiques (PCA, PLSR) ou des approches plus avancées d'apprentissage automatique (RFR, GPR) qui permettent d'exploiter toutes les longueurs d'onde mesurées et d'obtenir des estimations plus précises qu'avec des indices classiques. [5], [6]

Toutefois ces modèles sont limités en termes de précision. C'est pourquoi certaines approches basées sur le Deep Learning, comme les CNN, ont montré de meilleures performances. Ces méthodes sont capables d'apprendre automatiquement des schémas dans les spectres, mais demandent beaucoup de données d'entraînement pour être mises en œuvre. [7]

2.2 Matériel hyperspectral

Plusieurs types d'instruments permettent d'acquérir des données spectrales, chacun avec ses avantages et ses limites. Les spectroradiomètres portables sont conçus pour mesurer la réflectance ponctuelle d'un matériau avec une grande précision spectrale. Ils fournissent un spectre pour une seule zone à la fois, ce qui n'est pas l'idéal pour profiter pleinement de la réflectance d'une feuille et de ses structures.

À l'inverse, les caméras hyperspectrales sont capables de produire des hypercubes, c'est-à-dire des images où chaque pixel contient un spectre complet. Cette capacité en fait un outil puissant pour analyser la variabilité spectrale et spatiale au sein d'un échantillon. Grâce à cette spécificité, il est possible d'identifier des structures internes. Ces caméras sont donc plus adaptées à l'analyse d'objets complexes comme des feuilles.

Cependant, les dispositifs professionnels sont très limitants en termes de coût (plusieurs milliers à dizaines de milliers de francs suisses). Ils sont difficilement accessibles pour des projets étudiants.

Pour répondre à cette problématique, plusieurs travaux récents ont exploré des alternatives *low-cost* en utilisant des composants et des méthodes standards comme l'impression 3D. Le système proposé par Salazar-Vazquez et Mendez-Vazquez (2020) [8] permet de construire une caméra hyperspectrale fonctionnelle pour environ 475 CHF, avec un spectre allant de 400 à 1050 nm et une résolution spectrale de 2 nm. De son côté, Pechlivani et al. (2023) [9] ont développé un système similaire couvrant une gamme spectrale de 379 à 937 nm avec une résolution de 1.9 nm. Ces approches sont basées sur des éléments optiques simples (lentilles et réseau de diffraction) et un micro-ordinateur comme un Raspberry Pi. Il est donc aujourd'hui possible de concevoir des dispositifs hyperspectraux économiques permettant ainsi une démocratisation de cette technologie dans des domaines variés comme l'agriculture.

3 Matériel

Ce chapitre présente le matériel nécessaire à la réalisation du projet.

3.1 Dispositif d'acquisition

Dans le cadre d'une étude exploratoire concernant l'analyse de traits fonctionnels de feuilles d'érable à l'aide d'une caméra hyperspectrale *low-cost*, nous mesurons les réflectances des feuilles dans des conditions d'éclairage maîtrisées.

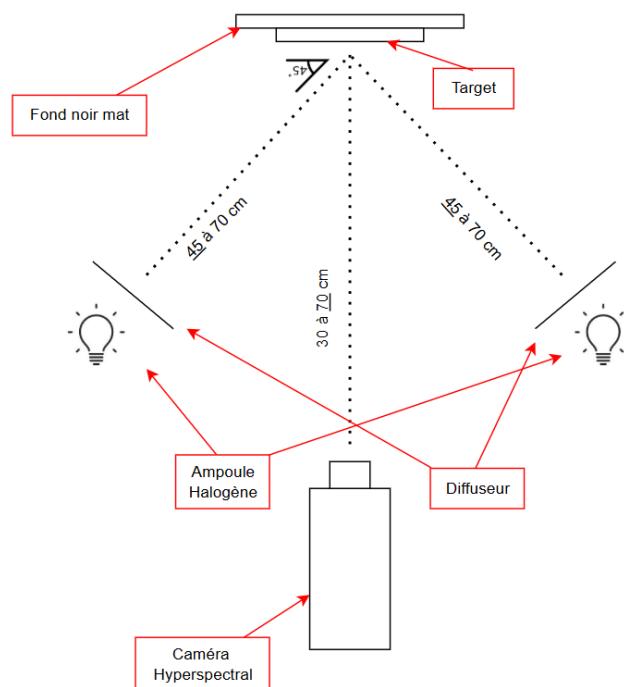


Figure 3.1-1 : Schéma du dispositif utilisé pour capturer un hypercube (Vue de dessus)

Le dispositif de capture est situé dans une pièce qui n'a pour éclairage que les lampes halogènes.

Les ampoules halogènes sont utilisées en raison de leur spectre lumineux continu, ce qui est essentiel en hyperspectral, car il permet de mesurer la réflectance de la cible à chaque bande spectrale.

Deux sources sont disposées à 45° de part et d'autre de la cible afin d'éliminer les ombres et de garantir une illumination homogène sur toute la surface.

Des diffuseurs sont ajoutés devant chaque source pour adoucir la lumière et éviter les pics d'intensité localisés qui pourraient fausser les mesures.

La cible est placée sur un fond noir mat pour empêcher toute réflexion parasite autour de la zone utile.

Toutes les captures d'hypercube durant le projet ont été réalisées selon ce dispositif.

3.2 Système hyperspectral

Après analyse des caméras hyperspectrales *low-cost* existantes, nous avons retenu la solution issue du papier scientifique « *A plug-and-play Hyperspectral Imaging Sensor using low-cost equipment* » [8]. Ce choix est motivé par les raisons suivantes :

- Projet Open-source (les fichiers STL et logiciels associés sont disponibles en ligne)
- Faible coût par rapport à une caméra hyperspectrale ordinaire
- Un large spectre couvert

Le tableau ci-dessous résume les caractéristiques principales de la caméra :

Tableau récapitulatif du système de Jairo Salazar-Vazquez et Andres Mendez-Vazquez	
Coût Global	~ 475 CHF
Poids Total	300g
Dimensions	18 × 9 × 7 cm
Spectre Couvert	De 400 à 1050 nm
Nombre de Bandes Spectrales	315 bandes (résolution spectrale ~2,06 nm)
Résolution Spatiale	116 × 110 pixels
Matériaux	PLA
Alimentation	5V
Software	HypRaspCam (serveur d'acquisition) et squareHSI (client)

Tableau 1 : Description de la caméra hyperspectrale

3.2.1 Architecture matérielle

La caméra hyperspectrale a été entièrement montée avec succès. Le boîtier, imprimé en PLA à la HEI, intègre un Raspberry Pi 4, une caméra Pi NoIR, une lentille téléphoto, une lentille macro et un réseau de diffraction, selon la conception open-source de Salazar-Vazquez (2020) [8]. La face arrière de la caméra a été redessinée pour y intégrer un connecteur USB-C.



Figure 3.2-1 : Caméra hyperspectrale montée

3.2.2 Architecture logicielle

Le système repose sur une architecture client/serveur : le Raspberry Pi joue le rôle de serveur, via le programme *HypRaspCam*, qui reçoit des commandes réseau et utilise la caméra pour capturer des images brutes (via la librairie *raspistill*). Le PC est le client qui fonctionne grâce à l'application *SquareHSI* qui se connecte au Raspberry, envoie les commandes, reçoit les images capturées et construit localement l'hypercube qui peut ensuite être exporté.

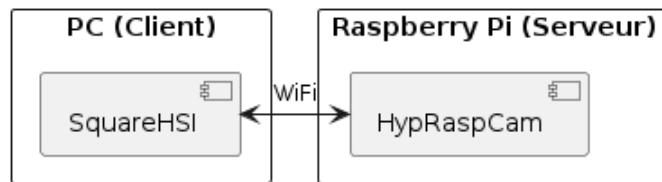


Figure 3.2-2 : Architecture logicielle

Du côté Raspberry Pi, le principal problème rencontré est la disparition de la commande *raspistill* sur les dernières versions de *Raspberry Pi OS*. Cela empêchait le fonctionnement de *HypRaspCam*. Après plusieurs essais avec différents systèmes d'exploitation, *RaspiOS Buster* de 2021 s'est révélée être le seul utilisant encore la librairie *raspistill* et étant compatible avec un Raspberry Pi 4.

Côté PC, l'application *SquareHSI* est distribuée au format .ApplImage (Linux), mais n'était pas stable avec les versions récentes d'*Ubuntu*. Le système a donc été installé sur *Ubuntu 18.04*, seul environnement compatible.

Après ces ajustements, l'ensemble de la chaîne logicielle fonctionne désormais correctement : la connexion client-serveur est stable, les acquisitions d'images sont fiables et les hypercubes générés sont exploitables dans des environnements comme MATLAB ou Python.

3.2.3 Composants et budget

Le Tableau 2 présente le récapitulatif des composants utilisés, leur provenance et leur coût approximatif.

Tableau des composants			
Dénomination	Fournisseur	N° Référence	Prix CHF (sans TVA)
Raspberry Pi 4	DigiKey	2648-SC0195(9)-ND	67.50
Pi NoIR Camera	DigiKey	2648-SC0024-ND	22.50
Micro-SD	DigiKey	3247-USDCOEM-32GB-ND	10.59
Front Camera	EdmundOptics	59-872	280.00
Macro Lens	Galaxus	22453538	51.90
Réseau de diffraction	Ebay	313366264746	15.00
			Total : 447.49

Tableau 2 : Composants utilisés pour la caméra hyperspectrale, fournisseurs et prix associés

Ce tableau présente un budget total d'environ 450 CHF (hors TVA). Le matériel a été commandé en mars 2025.

3.2.4 Validation du fonctionnement de la caméra

Afin de valider le fonctionnement de la caméra hyperspectrale, une acquisition a été réalisée sur une cible composée de cinq pastilles colorées : rouge, vert, bleu, blanc et noir (Figure 3.2-3). Cette cible permet d'évaluer la capacité du système à distinguer différentes signatures spectrales.

La capture hyperspectrale brute capturée par le système (Figure 3.2-4) montre une structure de diffraction conforme avec une distribution spatiale et spectrale correcte, confirmant le bon alignement de l'optique.

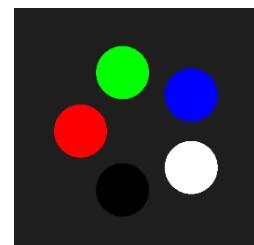


Figure 3.2-3 : Cible colorée

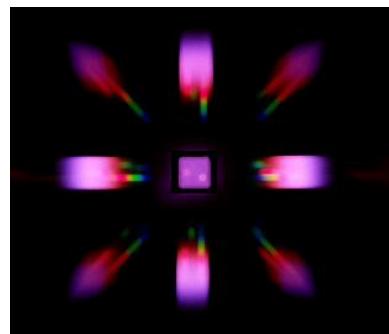


Figure 3.2-4 : Image brute issue de HypRaspCam

Les spectres ont été extraits à partir de cinq pixels représentatifs, chacun situé au centre d'une des cinq zones colorées. Les courbes obtenues (Figure 3.2-5) présentent des caractéristiques attendues :

- Un pic dans le bleu (~460 nm) pour la pastille bleue
- Un pic dans le vert (~550 nm) pour la pastille verte
- Un pic dans le rouge (~650 nm) pour la pastille rouge
- Un spectre plus large et plus intense pour le blanc
- Une intensité plus faible pour le noir

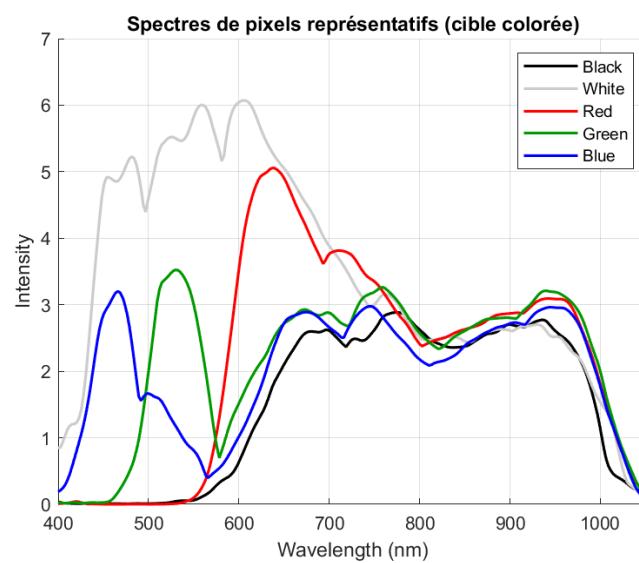


Figure 3.2-5 : Spectres des pixels colorés extraits de l'hypercube capturé de la cible

Une composition RGB reconstruite à partir des bandes hyperspectrales à 450, 550 et 650 nm (Figure 3.2-6) permet de mettre visuellement en évidence les zones colorées de la cible.

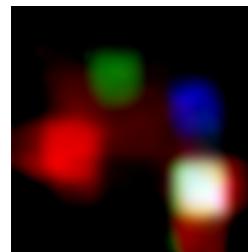


Figure 3.2-6 : Composition RGB reconstruite à partir de l'hypercube brut (450, 550, 650 nm)

Ces résultats démontrent que la caméra est capable de capturer des spectres cohérents spatialement et spectralement et que les reconstructions RGB sont visuellement fidèles. Ce test valide les principales fonctionnalités attendues du système.

Remarque : alors que l'auteur de la caméra annonçait 315 bandes spectrales dans son article [8], la calibration effectuée avec le logiciel dédié n'a permis d'en exploiter que 255.

3.2.5 Principe de fonctionnement de la caméra

La caméra hyperspectrale utilisée dans ce projet repose sur un système optique combinant une caméra Raspberry Pi NoIR, un objectif, une lentille, et un réseau de diffraction. L'ensemble du montage est représenté sur la Figure 3.2-7 ci-dessous.

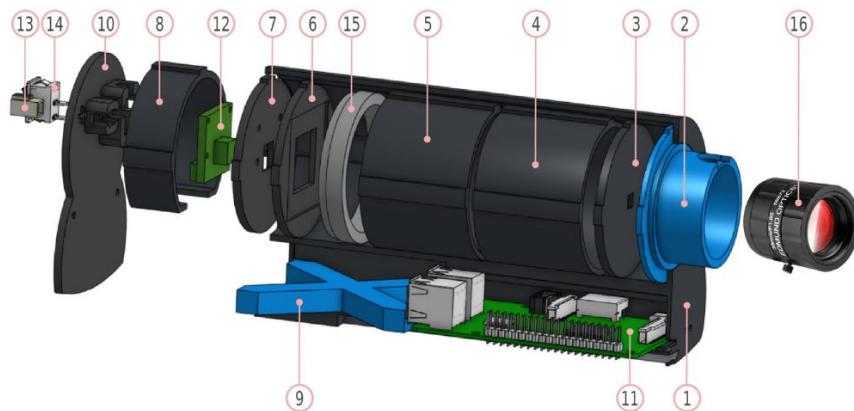


Figure 3.2-7 : Vue éclatée du système HSI proposé par Salazar-Vazquez et Mendez-Vazquez [8]

La lumière provenant de la scène entre par l'objectif principal (*pièce 16*). Elle passe ensuite à travers une ouverture carrée (*pièce 3*), qui définit le champ de vision effectif de la caméra. Cette ouverture permet de "filtrer" une portion précise de la scène observée.

La lumière continue son chemin à travers une lentille macro +10 (*pièce 15*). Celle-ci sert à agrandir l'image de la scène pour capturer des détails sur de petites zones, comme les nervures d'une feuille.

Vient ensuite un réseau de diffraction à double axe (*monté sur pièce 6*) qui décompose la lumière provenant de la scène.

Enfin, la lumière diffractée est captée par la caméra Raspberry Pi NoIR (*pièce 12*), qui enregistre une image 2D souvent appelée *SP* (*Spectral Photograph*). Un exemple d'image brute acquise sur une feuille végétale est montré ci-dessous :

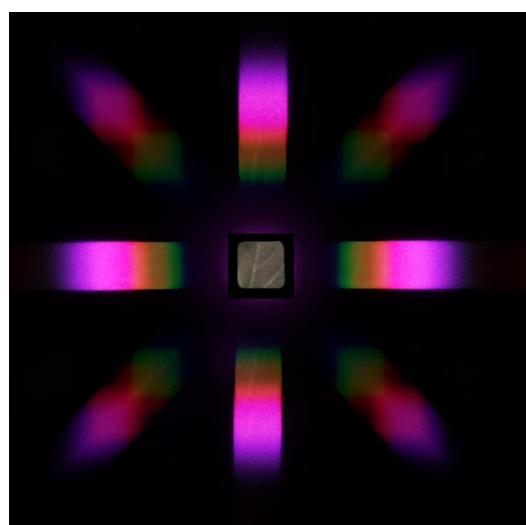


Figure 3.2-8 : Image SP brute d'une feuille végétale

Au centre de cette image, on distingue la scène originale non diffractée (ordre 0) également appelée *ZM (Zero-order mode)*. Autour, plusieurs copies colorées de la scène apparaissent : ce sont les lobes de diffraction (ordres $\pm 1, \pm 2$). Chaque lobe représente une version spectrale de la scène.

L'image brute ne constitue pas directement un hypercube. Pour reconstruire un hypercube 3D (dimensions : X, Y, λ) à partir de l'image, SquareHSI utilise un algorithme d'Expectation-Maximization (EM) [10]. C'est un processus itératif qui permet d'estimer la contribution spectrale de chaque pixel en fonction de sa position dans l'image diffractée.

Concrètement, l'algorithme utilise une matrice de projection qui relie les coordonnées de l'hypercube (X, Y, λ) aux positions des pixels dans l'image diffractée (*SP*). Grâce à une calibration préalable effectuée à partir d'une source lumineuse ponctuelle, on connaît la position spatiale des différentes longueurs d'onde vues par le capteur. L'algorithme Expectation-Maximization (EM) itère alors pour estimer le contenu spectral de chaque pixel du cube en ajustant progressivement l'hypercube jusqu'à correspondre le mieux possible à l'image diffractée observée (*SP*). Une centaine d'itérations est nécessaire pour converger vers une solution cohérente.

4 Séquence de traitement des données hyperspectrales

La séquence de traitement des données hyperspectrales comprend plusieurs étapes nécessaires pour rendre les données brutes exploitables à des fins d'analyse. Comme illustré dans la Figure 3.2-1, le processus débute par l'acquisition d'une image hyperspectrale (obtention de l'hypercube), suivie de l'extraction d'un spectre caractéristique. Ce spectre est ensuite normalisé pour corriger l'effet de l'éclairage. Des indices spectraux sont alors calculés à partir des signatures spectrales extraites, avant que les données finales ne soient envoyées vers une base de données pour analyse ultérieure.

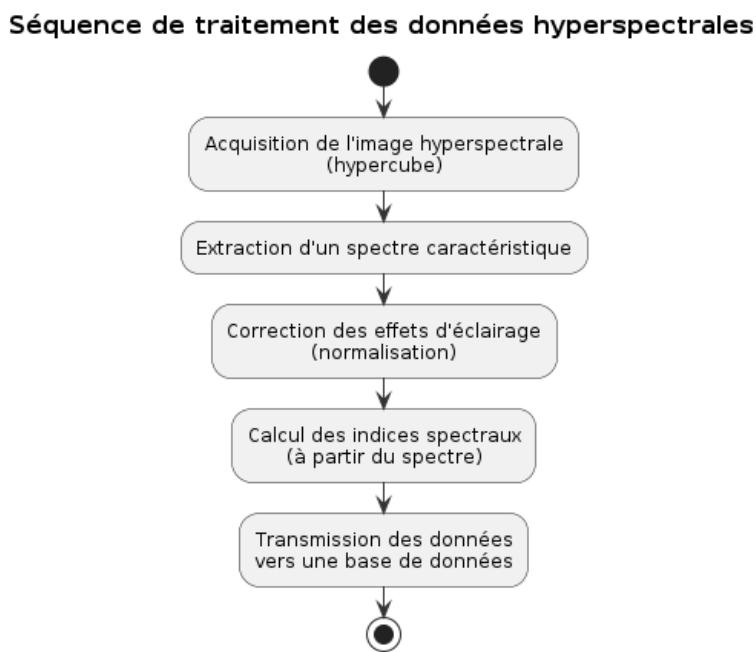


Figure 3.2-1 : Chaîne de traitement des données hyperspectrales

4.1 Acquisition de l'image hyperspectrale

L'acquisition est réalisée à l'aide du système hyperspectral et du dispositif d'acquisition décrits au chapitre Matériel. Le fichier qui résulte de cette étape est une matrice 3D (hypercube) qui contient un spectre pour chaque pixel dans une plage de longueurs d'onde de 400 nm à 1050 nm.

Remarque : Les acquisitions ont été réalisées sur des plantes observées pendant une période favorable avec peu de signes de stress hydrique ou physiologique.

4.2 Extraction d'un spectre caractéristique

Une fois l'hypercube capturé, un spectre caractéristique est extrait pour chaque échantillon. Nous sélectionnons, via une interface graphique qui affiche l'image RGB recomposée de la feuille, une zone (la plus grande possible) qui ne contient pas de nervures. Nous faisons ensuite la moyenne des spectres de chaque pixel qui compose cette zone pour produire un spectre caractéristique de l'échantillon. Le spectre qui en résulte est ensuite stocké dans un tableau 2D.

Cette étape permet de faciliter l'analyse en réduisant la taille des données tout en conservant une information spectrale représentative.

4.3 Correction des effets d'éclairage (normalisation)

Les spectres bruts sont ensuite soumis à une normalisation radiométrique afin de corriger les effets liés à la source de lumière et au capteur. Cette étape permet de transformer les intensités mesurées (dépendantes du temps d'exposition réglé sur la caméra et de la source de lumière) en valeurs de réflectance relatives dans une plage comprise entre 0 et 1.

Pour normaliser, nous utilisons deux spectres de référence :

- **La référence noire (dark)**, capturée sans lumière, permet de compenser le bruit de fond du système.
- **La référence blanche (white)**, mesurée sur une surface à réflectance maximale (comme un Spectralon avec 99% de réflectance), permet de compenser les variations d'intensité lumineuse.

Le calcul de la réflectance R est réalisé selon la formule suivante :

$$R = \frac{S - D}{W - D}$$

Où :

- S est le spectre mesuré sur l'échantillon
- D est le spectre dark
- W est le spectre white

Cette méthode, couramment utilisée en imagerie hyperspectrale, est décrite par *ElMasry et al., 2007* [11]. Elle garantit des données comparables d'un échantillon à l'autre si les conditions d'éclairage changent.

Pour illustrer l'effet de cette opération, un exemple se trouve ci-après. Le premier graphique (Figure 4.3-1) présente des spectres bruts mesurés lors d'une observation sur un échantillon ; le spectre mesuré sur une feuille d'érable quelconque (S), le spectre mesuré sur une référence blanche (W) et le spectre mesuré sur une référence noire (D).

Sur le second graphique (Figure 4.3-2) se trouve le spectre de réflectance de la feuille d'érable (R) qui résulte de l'opération de normalisation. Ce spectre est indépendant des conditions d'éclairage.

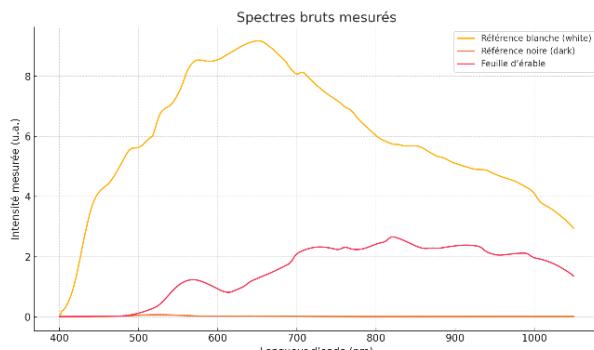


Figure 4.3-1 : Exemple de spectres bruts mesurés S, D, W

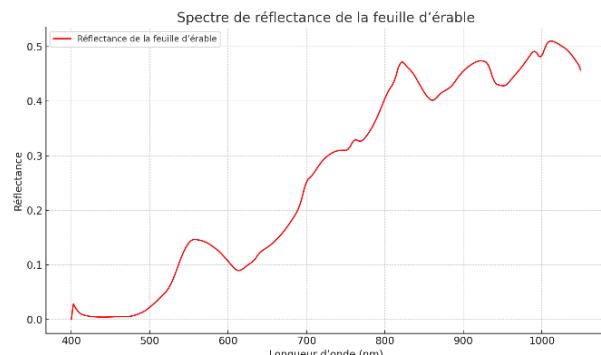


Figure 4.3-2 : Exemple d'un spectre de réflectance

4.4 Calcul des indices spectraux

Plusieurs indices spectraux couramment utilisés dans l'analyse de la végétation ont été calculés à partir des spectres normalisés (réflectance). Ces indices permettent d'évaluer des propriétés physiologiques ou chimiques telles que l'activité photosynthétique ou le stress hydrique.

Le tableau ci-dessous (Tableau 3) présente une sélection d'indices choisis pour leur reconnaissance dans la littérature et leur capacité à refléter des traits fonctionnels.

Ces formules sont issues de la synthèse proposée par Sellami et al. (2022) [3].

Tous ces indices sont calculés pour chaque échantillon. Les résultats sont ensuite enregistrés dans un fichier CSV de session (plus de détails au chapitre Contexte technique et organisationnel).

Tableau des indices spectraux calculés

Acronyme	Nom complet	Formule
NDVI	Normalized Difference Vegetation Index	$(R860 - R650) / (R860 + R650)$
PRI	Photochemical Reflectance Index	$(R531 - R570) / (R531 + R570)$
WBI	Water Band Index	$R970 / R900$
CRI	Carotenoid Reflectance Index	$(1 / R510) - (1 / R550)$
PSRI	Plant Senescence Reflectance Index	$(R680 - R500) / R750$
NDRE	Normalized Difference Red-Edge Index	$(R790 - R720) / (R790 + R720)$
SICI	Structure Insensitive Pigment Index	$(R800 - R445) / (R680 - R445)$
CARI	Chlorophyll Absorption Reflectance Index	$(R700 - R670) - 0.2 \times (R700 - R550)$
RVSI	Red-Edge Vegetation Stress Index	$((R714 + R752)/2) - R733$

Tableau 3 : Liste des indices spectraux calculés et leurs formules associées

4.5 Envoi vers une base de données

A la suite du calcul des indices spectraux, toutes les données (réflectance, indices spectraux, métadonnées associées aux feuilles et aux individus, traits fonctionnels en relation avec les

feuilles, etc.) sont intégrées dans une base de données relationnelle. Ce choix s'explique par la structure interconnectée des données du projet, notamment :

- La relation un-à-plusieurs entre les individus, les feuilles et leurs mesures.
- Le lien entre les mesures hyperspectrales, les instruments utilisés et les traits mesurés.
- La possibilité d'indexer facilement les données par individu, session, site, ou encore type d'instrument.

L'utilisation d'un modèle relationnel basé sur le langage SQL permet une structuration normalisée des données, tout en facilitant leur accès via des requêtes. Le modèle relationnel utilisé est présenté en détail au chapitre Base de données.

Ce format présente également l'avantage de regrouper l'ensemble des données dans un fichier unique, ce qui le rend portable, facilement transmissible et indépendant des fichiers locaux du projet. Ainsi, les données peuvent être partagées et exploitées facilement par d'autres utilisateurs sans dépendre de la structure du projet initial.

5 Contexte technique et organisationnel

Ce chapitre décrit comment est organisé le projet et quels outils ont été utilisés. Une structuration claire et rigoureuse est essentielle dans un projet manipulant un volume important de données.

La structure du projet a été conçue pour répondre à plusieurs objectifs :

- Assurer une bonne correspondance avec la séquence de traitement présentée au chapitre précédent.
- Automatiser les étapes du flux de traitement, de la normalisation au calcul d'indices spectraux.
- Faciliter la traçabilité des données.
- Rendre le projet modulaire et évolutif.

L'ensemble du projet a été développé en Python dans l'environnement de PyCharm, un langage particulièrement adapté pour ce type de traitement. En effet, Python offre :

- Une large gamme de bibliothèques (NumPy, pandas) pour le traitement de données numériques.
- Des outils de visualisation (matplotlib).
- Ainsi que des modules avancés pour l'analyse statistique et l'apprentissage automatique.

Ce choix permet également une bonne intégration avec des notebooks Jupyter, ce qui facilite la présentation de la démarche.

5.1 Organisation générale du projet

La Figure 5.1-1 présente comment sont organisés les principaux répertoires du projet :

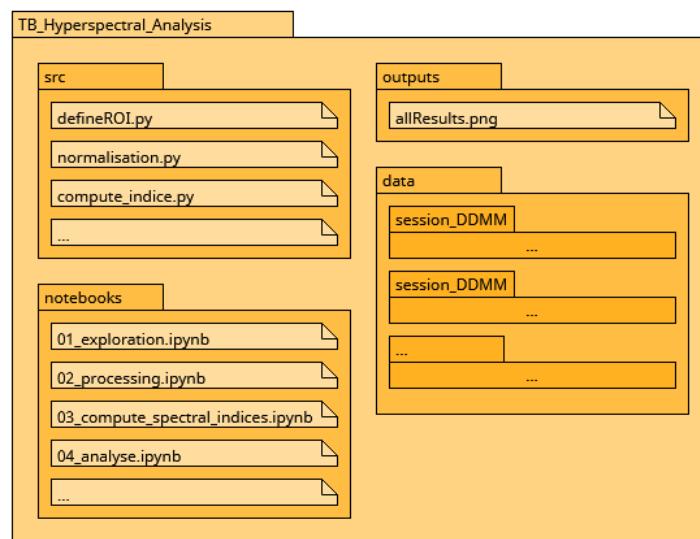


Figure 5.1-1 : Organisation générale des répertoires du projet

- **src/** : contient les scripts Python correspondant aux différentes étapes de traitement (extraction du spectre, normalisation, calcul d'indices, etc.).
- **notebooks/** : regroupe les notebooks Jupyter utilisés pour exécuter, documenter et visualiser la chaîne de traitement. Chaque notebook correspond à une étape clé de l'analyse.

- **data/** : structure hiérarchique contenant les données brutes et traitées, organisées par session d'acquisition.
- **outputs/** : dossiers de sortie pour les graphiques et résultats globaux sous forme d'images.

5.2 Organisation par session de données

Chaque sous-dossier de *data/* correspond à une session d'acquisition hyperspectrale, identifiée par la date de l'acquisition en portant le nom « *session_DDMM* ». La Figure 5.2-1 présente l'organisation interne de chacun des répertoires représentant une session d'acquisition.

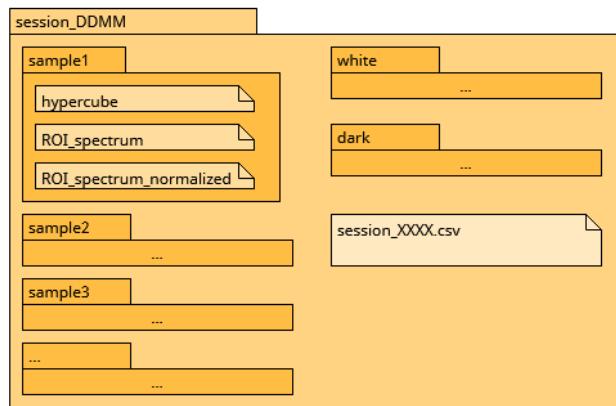


Figure 5.2-1 : Organisation interne d'une session d'acquisition

A l'intérieur de chaque dossier de session se trouve :

- Chaque échantillon (feuille) représenté par un dossier qui porte le nom de l'échantillon, ici *sampleX/* pour l'exemple. Ce dossier contient les fichiers suivants :
 - Le cube hyperspectral brut
 - Le spectre caractéristique brut de l'échantillon
 - Le spectre de réflectance après normalisation
- Les dossiers *white/* et *dark/* contiennent respectivement les hypercubes de référence nécessaire à la normalisation des spectres. Le choix de disposer d'une référence blanche et noire par session garantit une correction cohérente pour l'ensemble des échantillons car le Dispositif d'acquisition reste fixe tout au long d'une session d'acquisition, ce qui rend les conditions de capture suffisamment homogènes pour que les mêmes références soient valables pour tous les échantillons. En revanche, le dispositif peut être repositionné ou légèrement modifié entre deux sessions, justifiant l'usage d'un nouveau jeu de références pour chaque session
- Le fichier *session_DDMM.csv* qui joue un rôle important dans l'automatisation du code et la traçabilité des fichiers.

Le fichier *session_DDMM.csv* lie les identifiants uniques des échantillons avec les fichiers et les indices spectraux calculés correspondants.

Ce fichier joue un rôle essentiel dans l'organisation du code. Il permet de piloter les fonctions de traitement sans avoir à manipuler les fichiers manuellement. Son contenu est décrit dans la Figure 5.2-2.

Colonnes	
A	ID
B	path_raw_mat
B	path_spectre_ROI
B	path_spectre_ROI_normalized
C	WBI
C	NDVI
C	PRI
C	CRI
C	PSRI
C	NDRE
C	SIPI
C	CARI
C	RVSI

Figure 5.2-2 : Contenu d'un fichier de session

Les colonnes sont regroupées en trois catégories : A (identifiant unique), B (chemins vers les données), C (indices spectraux).

A) Identifiant unique

La colonne ID contient un identifiant unique pour chaque feuille. Il suit la structure :

$$fX_rY_iZ$$

où :

- Z = numéro de l'individu
- Y = numéro du rameau sur cet individu
- X = numéro de la feuille sur ce rameau

Cet identifiant est utilisé :

- Pour faire le lien entre les données hyperspectrales et les données de référence (traits mesurés) via une jointure sur l'identifiant.
- Comme clé dans les fonctions du code Python, par exemple :

```
plot_band(session_name: str, ID: str, band: int)
plot_spectrum(session_name: str, ID: str, type: str)
normalize_roi_spectrum(session_name: str, ID: str)
compute_indices(session_name: str, ID: str, type: str)
```

Ces fonctions exploitent le nom de la session d'acquisition et l'ID pour accéder automatiquement aux données via les chemins présents dans les colonnes B du fichier de session. Cela nous évite de devoir exploiter directement les chemins bruts et standardise les appels de fonction.

B) Chemin relatif des données

Les colonnes suivantes stockent les chemins relatifs vers les fichiers générés durant la chaîne de traitement :

- **path_raw_mat** : chemin vers l'hypercube brut, au format **.mat**. Ce fichier contient une matrice de dimension (x, y, λ) .
- **path_spectre_ROI** : chemin vers le spectre caractéristique de l'échantillon au format **.npz** (NumPy compressed array).
- **path_spectre_ROI_normalized** : chemin vers le spectre de réflectance également stocké au format **.npz**.

C) Indices spectraux

Les colonnes suivantes contiennent les indices spectraux calculés à partir du spectre de réflectance, ils sont définis au chapitre « Calcul des indices spectraux ».

L'ensemble du projet a été conçu pour être automatisé au maximum, avec une génération systématique des fichiers aux bons formats, enregistrés dans les emplacements appropriés avec des noms cohérents. Les fichiers CSV de session sont mis à jour dynamiquement à chaque étape du traitement.

5.3 Base de données

5.3.1 Technologie

La base de données est implémentée à l'aide du moteur SQLite intégré dans PyCharm, un système de gestion de base de données léger sans serveur, adapté à des projets de petite à moyenne échelle.

L'interaction avec la base est gérée en Python via la bibliothèque SQLAlchemy [12], qui permet d'exécuter des requêtes SQL et de définir des tables sous forme d'objets Python.

5.3.2 Description du schéma relationnel

Le schéma relationnel de la base de données a été conçu pour représenter les dépendances logiques des données collectées au cours du projet. Il repose sur un ensemble de huit tables principales, reliées entre elles par des clés étrangères. La Figure 5.3-1 ci-dessous présente le modèle relationnel.

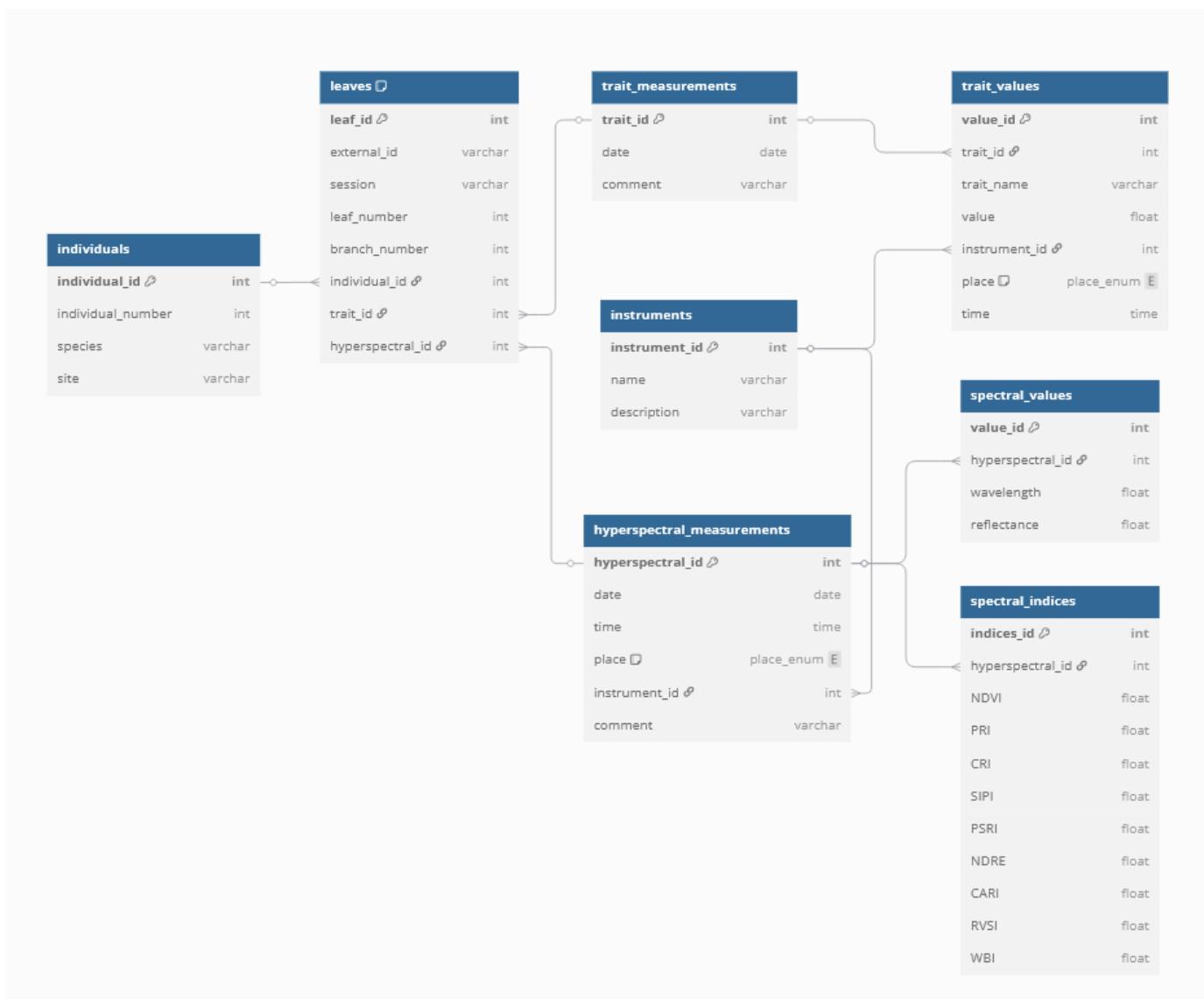


Figure 5.3-1 : Schéma relationnel de la base de données

Les tables peuvent être regroupées selon trois catégories :

Données biologiques et échantillons

- *individuals* : information sur les arbres étudiés
- *leaves* : information sur les feuilles échantillonnées avec liens vers leurs mesures et leur arbre source

Mesures hyperspectrales

- *hyperspectral_measurements* : métadonnées associées aux acquisitions
- *spectral_values* : spectres normalisé (réflectance par longueur d'onde)
- *spectral_indices* : indices spectraux dérivés (NDVI, PRI, etc.)

Traits fonctionnels mesurés manuellement

- *trait_measurements* : sessions de mesures de traits
- *trait_values* : valeurs individuelles des traits

- *instruments* : appareils utilisés pour les mesures des traits ou pour l'acquisition hyperspectrale

5.3.3 Détail des tables et des champs

Table individuals

La table **individuals** contient les informations relatives aux érables desquels ont été échantillonnées les feuilles. Chaque individu est identifié par un identifiant unique *individual_id* et un numéro de terrain *individual_number*. Les colonnes *species* et *site* permettent d'indiquer l'espèce au format texte et le site d'échantillonnage. La base de données comprend un total de 7 espèces réparties sur 4 sites. Cette table est en relation un-à-plusieurs avec la table **leaves**, chaque individu pouvant être lié à plusieurs feuilles mesurées.

Table leaves

La table **leaves** est le nœud central du modèle relationnel, reliant les feuilles aux mesures hyperspectrales et aux mesures de traits ainsi qu'à leur arbre source. Chaque entrée correspond à une feuille unique, identifiée par un identifiant interne *leaf_id*. Le champ *external_id* correspond à l'identifiant utilisé dans les fichiers de session produits par la chaîne de traitement. Le champ *session* contient le nom de la session d'acquisition, identique au nom du fichier de session dont la feuille est issue. Les champs *leaf_number* et *branch_number* précisent respectivement le numéro de la feuille au sein d'un même rameau, et le numéro du rameau au sein d'un même individu. Enfin, la table **leaves** contient des clés étrangères vers **individuals**, **trait_measurements** et **hyperspectral_measurements** pour faire les liens entre ses mesures et sa provenance.

Table hyperspectral_measurements

Cette table regroupe les métadonnées des acquisitions hyperspectrales, telles que la date (*date*), l'heure (*time*), le lieu (*place*) et l'instrument utilisé. Le champ *instrument_id* pointe vers la caméra hyperspectrale développée dans le cadre de ce Travail de Bachelor. Bien qu'un seul instrument ait été utilisé pour les mesures hyperspectrales, la structure permet d'en intégrer d'autres si nécessaire. La table est reliée à **spectral_values**, **spectral_indices** (relation un-à-un), et **leaves**.

Table spectral_values

Cette table contient les spectres normalisés (réflectance) issus des acquisitions hyperspectrales. Chaque ligne représente une longueur d'onde unique et la valeur de réflectance associée pour une feuille donnée. Chaque spectre contient 255 valeurs de réflectance aux différentes longueurs d'onde, allant de 400 nm à 1050 nm, ce qui correspond à une résolution spectrale de 2,55 nm.

Table spectral_indices

Les indices spectraux dérivés des spectres sont stockés dans cette table. Chaque enregistrement correspond à un ensemble d'indices calculés à partir d'un spectre unique (*hyperspectral_id*), avec une relation un-à-un. Les indices stockés sont définis au chapitre 4.4.

Table trait_measurements

La table **trait_measurements** décrit les sessions de mesures des traits fonctionnels. Dans ce projet, toutes les mesures de traits ont été réalisées le même jour pour chaque feuille. Le champ *comment* permet d'ajouter des informations contextuelles. Cette table est liée à **trait_values** (un-à-plusieurs) et à **leaves**.

Table trait_values

Cette table contient les valeurs individuelles des traits fonctionnels mesurés sur chaque feuille. Chaque valeur est associée à une session de mesure (*trait_id*), un nom de trait (*trait_name*), une valeur numérique (*value*), un instrument (*instrument_id*) et un lieu de mesure (*place*) qui est exprimé sous forme d'énumération avec « lab » ou « terrain ». Le champ *time* permet également de renseigner l'heure de la mesure.

Table instruments

Enfin, la table **instruments** permet de préciser avec quel appareil a été fait chaque mesure. Chaque instrument est identifié par un identifiant unique (*instrument_id*) et décrit par un nom (*name*) et une courte description (*description*). Ce découplage permet d'éviter les redondances dans les tables **trait_values** et **hyperspectral_measurements**.

5.3.4 Sources et alimentation de la base de données

Toutes les acquisitions hyperspectrales réalisées au cours du projet ont été systématiquement ajoutées à cette base de données. Les fichiers sources pour remplir les tables sont :

- Les fichiers de session générés par la Séquence de traitement des données hyperspectrales.
- Les fichiers de référence fournis par les membres du laboratoire, comprenant les mesures de traits fonctionnels et les informations sur les individus étudiés.

Toutes les analyses s'appuient exclusivement sur cette base de données.

6 Analyse des données hyperspectrales

6.1 Introduction

L'analyse des données hyperspectrales s'organise autour de trois approches complémentaires afin d'explorer différents aspects de la relation entre les signatures spectrales et les traits fonctionnels des végétaux.

Ces approches permettent d'aborder la richesse du jeu de données sous plusieurs angles :

- La variabilité des spectres selon les niveaux biologiques (espèce, individu, rameau)
- La capacité à expliquer des traits fonctionnels mesurés avec des instruments traditionnels à l'aide des spectres et/ou indices spectraux
- La valorisation de données archivées en testant la corrélation entre les spectres mesurés lors du projet et des traits collectés l'année dernière.

Cette structuration en trois axes va de l'analyse exploratoire à la modélisation prédictive.

6.2 Approche 1 - Analyse de la variabilité inter- et intra-spécifique

6.2.1 Objectif

L'objectif de cette approche est de déterminer dans quelle mesure les données hyperspectrales permettent de différencier les niveaux hiérarchiques suivants :

- Les espèces d'érable
- Les individus au sein d'une même espèce
- Les rameaux au sein d'un même individu

Cette analyse vise à caractériser les sources de variations sans considérer directement des traits fonctionnels. Aucune autre mesure que la capture des images hyperspectrales n'est associée à cette approche.

6.2.2 Jeu de données

Le jeu de données utilisé pour cette approche est constitué de feuilles issues de quatre espèces d'érables :

- Acer platanoides
- Acer rubrum
- Acer pensylvanicum
- Acer saccharum

Pour chaque espèce, quatre individus ont été sélectionnés pour un total de 16 individus.

Sur chaque individu, dix feuilles ont été prélevées, réparties entre deux rameaux (soit 5 feuilles par rameau), ce qui représente un total de **160 échantillons**. Les feuilles ont été récoltées sur le site de l'UQTR.

Chaque échantillon est caractérisé par :

- Son spectre de réflectance hyperspectrale couvrant 400 à 1050 nm.
- Les indices spectraux (décris au chapitre Calcul des indices spectraux) calculés à partir du spectre.

6.2.3 Méthodes

ACP - Analyse en composantes principales

Dans le but de réduire la dimensionnalité des spectres de réflectance qui sont composés de 255 points, une Analyse en Composantes Principales (ACP) a été réalisée [13]. Cette méthode permet de projeter les échantillons dans un espace de plus faible dimension tout en conservant la majorité de l'information.

L'analyse a été réalisée de manière à conserver le nombre de composantes principales expliquant **95%** de la variance totale entre les spectres. Par ailleurs, la projection des données dans le plan PC1 VS PC2 a été utilisé pour évaluer le regroupement potentiel des échantillons selon leur espèce et leur individu d'origine.

Visualisation de PC1 selon différents niveaux

Nous explorons visuellement la répartition de la valeur de la première composante principale (PC1 issue de l'ACP), en fonction des niveaux hiérarchiques de l'échantillonnage, à travers des diagrammes en boîtes.

L'objectif de cette étape est de visualiser comment cette composante, qui explique la majorité de la variance totale, se répartit à travers les différents groupes.

Trois diagrammes en boîte sont générés :

- PC1 selon l'espèce
- PC1 selon les individus (avec indication de l'espèce)
- PC1 selon les branches (groupées par individu et espèce)

Analyse de la variance

Une analyse de la variance (ANOVA) a été appliquée sur les spectres et les indices spectraux dérivés pour explorer la répartition de la variabilité spectrale aux différents niveaux hiérarchiques.

L'ANOVA [14] est une méthode statistique permettant de tester si les moyennes de plusieurs groupes sont significativement différentes. Elle compare la variance expliquée par la structure des groupes (par exemple : espèces) à la variance résiduelle au sein des groupes. Lorsque cette différence est grande, le test retourne une valeur de p faible.

Plusieurs analyses de variances ont été menées notamment pour évaluer à chaque longueur d'onde et pour chaque indice spectral si les différences observées entre groupes sont statistiquement significatives. L'ANOVA a été systématiquement réalisée au niveau des groupes suivants :

- ANOVA inter-espèces
- ANOVA intra-espèces effectuées indépendamment pour chacune des quatre espèces
- ANOVA intra-individuelle effectuées indépendamment pour chaque individu

Les résultats sont exprimés en termes de valeurs de p transformées en $-\log_{10}(p\text{-value})$ pour faciliter la visualisation. Les courbes de significativité ont été tracées en fonction de la longueur d'onde, tandis que les indices spectraux sont comparés sous forme d'histogramme.

Apprentissage automatique supervisé (classification)

L'objectif de cette méthode est d'évaluer dans quelle mesure les données hyperspectrales permettent de prédire l'espèce des feuilles à l'aide de modèles d'apprentissage automatique supervisé.

Trois modèles de classification courants ont été utilisés pour leur performance sur des jeux de données de petite taille :

- **K plus proches voisins (KNN)** : un classificateur simple basé sur la distance entre les échantillons. Son principal avantage est sa facilité d'interprétation. [15]
- **Forêt aléatoire (Random Forest)** : un ensemble d'arbres de décision construit à partir de sous-échantillons aléatoires. Ce modèle est robuste au bruit et gère bien les données de haute dimension. [16]
- **Support Vector Machine (SVM)** : un modèle à noyau performant lorsque le nombre de caractéristiques est plus grand que le nombre d'échantillons. [17]

Afin d'explorer la capacité prédictive des modèles en fonction du type de données utilisées, plusieurs ensembles de variables explicatives (caractéristiques) ont été testés en entrée des modèles de classification :

- **Composantes principales (PC1 à PC7)**

Les 7 premières composantes (95% de la variance totale exprimée) issues de l'Analyse en Composantes Principales (ACP) ont été utilisées, elles résument l'essentiel de la variance avec moins de variable que l'entièreté du spectre, ce qui permet au modèle de mieux généraliser.

- **Indices spectraux**

Les indices calculés pour chaque feuille, car ils sont conçus pour refléter des propriétés physiologiques.

- **Spectre complet (toutes les bandes)**

Toutes les bandes du spectre hyperspectral (255 variables) ont été utilisées directement comme entrée. Ce test permet de voir si les modèles peuvent exploiter l'information brute sans réduction de dimension.

- **Bandes significatives (issues de l'ANOVA)**

Un sous-ensemble de bandes a été sélectionné en fonction de leur *p-value* lors de l'analyse de variance inter-espèces. Toutes les bandes ayant une valeur de $-\log_{10}(p\text{-value}) > 15$ ont été conservées. L'objectif est de réduire le nombre de variables en entrée, tout en conservant les longueurs d'onde statistiquement discriminantes entre espèces afin de visualiser si les modèles sont capables de prédire l'espèce à partir de portions du spectre.

- **Composantes principales (PC1 à PC7) + indices spectraux**

Cette combinaison permet de tester l'intérêt de coupler la réduction de dimension avec des variables interprétables comme les indices spectraux.

Nous avons mis en place un pipeline standard (Figure 6.2-1) pour évaluer objectivement les performances des différents modèles de classification. Les différentes étapes sont décrites ci-dessous.

1. Séparation en jeu d'entraînement et jeu de test

Les données ont d'abord été divisées en deux sous-ensembles : 80 % pour l'entraînement du modèle et 20 % pour le test (évaluation finale).

Cette séparation permet de tester le modèle sur des données qu'il n'a jamais vues, ce qui est important pour estimer sa capacité à généraliser.

2. Mise à l'échelle des variables (scaling)

La mise à l'échelle est indispensable pour les modèles sensibles à l'échelle des données comme KNN et SVM où les distances entre les points influencent directement la classification. [18]

3. Recherche des meilleurs hyperparamètres par validation croisée

Chaque modèle possède des hyperparamètres (comme le nombre de voisins pour KNN ou le nombre d'arbres pour Random Forest).

Pour trouver les combinaisons optimales, une validation croisée à 5 plis a été utilisée.

La validation croisée consiste à diviser le jeu d'entraînement en 5 sous-groupes : le modèle est entraîné sur 4 et testé sur le 5e, en tournant à chaque fois. Cela permet de mieux généraliser et est très utile lorsque le nombre d'échantillons est limité. Le modèle est ensuite entraîné avec les meilleurs hyperparamètres à l'aide de l'entièreté du jeu d'entraînement. [19]

4. Évaluation finale

Les modèles sont évalués sur le jeu de test selon plusieurs métriques [20] :

- Accuracy** : proportion de bonnes prédictions sur l'ensemble du test.
- F1-score** : moyenne harmonique entre la précision (nombre de vrais positifs sur les positifs prédits) et le rappel (nombre de vrais positifs sur les positifs réels), pondérée par le nombre d'exemples dans chaque classe. Elle est plus robuste en cas de classes déséquilibrées.
- Matrice de confusion** : tableau montrant les prédictions correctes et les erreurs, classe par classe. Elle permet de visualiser les confusions entre espèces.

Pipeline de modélisation

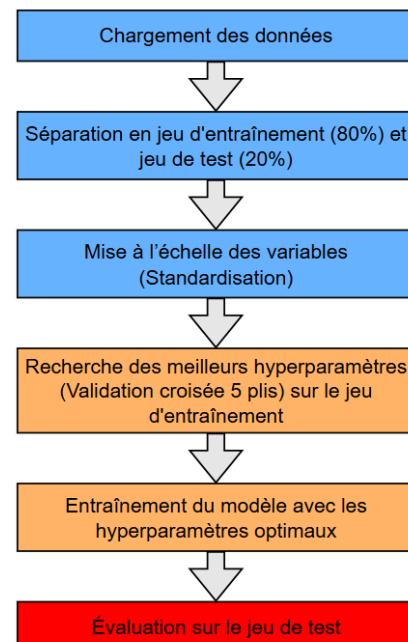


Figure 6.2-1 : Pipeline de modélisation ML approche 1

6.3 Approche 2 - Corrélation et modélisation des traits fonctionnels

6.3.1 Objectif

Cette deuxième approche vise à évaluer le potentiel explicatif et prédictif des données hyperspectrales vis-à-vis des traits fonctionnels des feuilles.

L'objectif principal est d'étudier dans quelle mesure les spectres, ou leurs dérivés (indices spectraux, composantes principales...), sont corrélés à des traits fonctionnels mesurés avec des instruments de laboratoire classiques.

Cette analyse permet d'identifier les bandes spectrales ou les indices spectraux qui capturent le mieux certaines caractéristiques biologiques et d'évaluer la capacité des modèles de régression à prédire ces traits à partir des données hyperspectrales.

6.3.2 Jeu de données

Le jeu de données utilisé dans cette approche provient d'un suivi de terrain réalisé dans le cadre d'un travail de master à l'Université du Québec à Trois-Rivières (UQTR). Il porte sur 28 individus d'érables appartenant à 7 espèces différentes réparties équitablement sur 4 sites. Les espèces incluses sont les suivantes :

- Acer spicatum
- Acer pensylvanicum
- Acer saccharinum
- Acer rubrum
- Acer platanoides
- Acer saccharum
- Acer negundo

Un échantillonnage a lieu toutes les deux semaines pendant l'été. Un rameau est prélevé pour chaque individu puis il est rapporté en laboratoire. Pour chaque rameau, trois feuilles sont sélectionnées pour qu'une image hyperspectrale ainsi que certains traits fonctionnels de référence (Tableau 4) soient mesurés.

Au total, quatre campagnes d'échantillonnage ont été réalisées et ont contribué aux acquisitions hyperspectrales, le jeu de données sur lequel les analyses de l'approche 2 reposent comporte ainsi un total de :

$$nb. d'échantillons = 4 \text{ campagnes} \times 28 \text{ individus} \times 3 \text{ feuilles} = 336 \text{ échantillons}$$

Tableau des traits fonctionnels de référence mesurés

Acronyme	Nom complet	Unité	Type
PhiPS2	Rendement quantique du photosystème II	-	Physiologique
ETR	Taux de transport des électrons de la photosynthèse	µmol m-2 s-1	Physiologique
gsw	Conductance stomatique de l'eau	mmol H ₂ O m-2 s-1	Physiologique
ChlM	Teneur foliaire en chlorophylle	Unit	Chimique
FlvM	Teneur foliaire en flavonoïde	Unit	Chimique
AnthM	Teneur foliaire en anthocyanine	Unit	Chimique
psi_midday	Potentiel hydrique de mi-journée	MPa	Physiologique

Tableau 4 : Traits fonctionnels de référence mesurés. Informations communiquées par l'équipe encadrante.

Les traits sont mesurés au même moment que l'acquisition hyperspectrale (en laboratoire) sauf *psi_midday* qui est mesuré sur le terrain.

Ce jeu de données constitue une base riche pour l'analyse des corrélations entre les signatures spectrales et les traits mesurés avec des acquisitions hyperspectrales réalisées dans des conditions identiques pour chaque échantillon.

6.3.3 Méthodes

Corrélation de Pearson

L'analyse de corrélation est la première étape exploratoire pour évaluer la relation entre les variables spectrales et les traits fonctionnels. L'objectif est de repérer les régions du spectre et les indices spectraux qui présentent des corrélations significatives avec les traits fonctionnels mesurés.

La méthode repose sur le calcul du coefficient de corrélation de Pearson (*r*) [21] qui mesure l'intensité et la direction de la relation linéaire entre deux variables. Le coefficient est calculé :

- Entre chaque bande spectrale (longueurs d'onde) et chaque trait fonctionnel donnant un profil de corrélation spectrale continue
- Entre chaque indice spectral (précalculé à partir du spectre de réflectance) et les traits

Ces deux analyses sont réalisées à deux niveaux :

- Global, toutes espèces confondues
- Par espèce, afin d'évaluer l'éventuelle influence de la variabilité interspécifique

Modélisation linéaire multivariée des traits

L'objectif de cette méthode est d'évaluer dans quelle mesure les traits fonctionnels peuvent être modélisés à l'aide de régressions linéaires multivariées. Pour cela, nous testons plusieurs types de données hyperspectrales en tant que variables explicatives, combinées ou non entre elles :

- **Indices spectraux**
- **Composantes principales (PCA)** (calculées sur les spectres) : réduction de dimension non supervisée permettant de capturer la variance globale du spectre. [13]
- **Composantes latentes (PLS)** (calculées sur les spectres et les traits) : réduction de dimension supervisée et spécifique à chaque trait pour maximiser la covariance entre le spectre et la variable cible. [22]

Remarque : Nous conservons les 4 premières composantes principales (PCA) expliquant **88.26 %** de la variance spectrale, et les 4 premières composantes latentes (PLS).

Nous testons les combinaisons suivantes :

- Indices spectraux seuls (9 variables)
- Composantes principales (PCA) seules (4 variables)
- Indices spectraux + PCA (13 variables)
- Composantes latentes (PLS) seules (4 variables)
- Indices spectraux + PLS (13 variables)

Chaque combinaison permet de tester quelles représentations spectrales sont les plus performantes pour modéliser chaque trait.

Nous testons également l'impact de la variable espèce dans la modélisation, pour ce faire, toutes les combinaisons de variables explicatives sont évaluées à travers un modèle linéaire classique [23] ainsi qu'un modèle linéaire mixte [24] où l'espèce est prise en compte en tant qu'effet aléatoire.

Cette approche nous permet d'estimer la contribution de l'espèce à la modélisation des traits et d'identifier pour chaque trait la meilleure combinaison explicative atteignable dans un cadre linéaire.

Pour évaluer les performances de la modélisation dans le cadre d'une régression, nous comparons les valeurs réelles (ici les valeurs de traits fonctionnels mesurés) et les prédictions du modèle en calculant deux indicateurs :

- **Le coefficient de détermination (R^2)** qui mesure la proportion de la variance totale des valeurs observées expliquée par le modèle. [25]

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Où :

- *y_i est la valeur réelle du i – ème échantillon*
- *ŷ_i est la valeur prédite par le modèle*
- *n est le nombre total d'échantillons*
- *ȳ est la moyenne des valeurs observée*

- **L'erreur quadratique moyenne normalisée (nRMSE)** qui permet de comparer les erreurs de prédiction de manière relative entre les différents traits.

$$nRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} / \bar{y}$$

Le RMSE [26] est normalisé par la moyenne des observations (\bar{y}), ce qui permet une comparaison relative entre variables de différentes unités ou échelles.

Ces deux indicateurs sont complémentaires : R^2 évalue la capacité du modèle à expliquer la structure des données, tandis que nRMSE mesure l'erreur par rapport à la vraie valeur. Leur interprétation conjointe est importante car pour des traits qui varient peu, un modèle peut afficher une faible erreur (nRMSE bas) sans pour autant expliquer la variance observée (R^2 faible).

Remarque : Le spectre entier de réflectance (255 variables) comme variables explicatives n'est pas utilisable à cause du surajustement et de la multi colinéarité, d'où l'intérêt d'approches comme la PCA ou la PLS.

Apprentissage automatique supervisé (régression)

L'objectif de cette approche est d'évaluer dans quelle mesure les données hyperspectrales permettent de prédire des traits fonctionnels à l'aide de modèles d'apprentissage automatique supervisé non-linéaire.

Trois modèles de régression ont été sélectionnés en raison de leur forte présence dans la littérature, notamment dans le domaine de la spectroscopie. [6], [27]

Les modèles étant de nature très différentes (à noyau, à base d'arbres, probabilistes), nous pourrons comparer leur capacité d'ajustement.

- **Random Forest Regression (RFR)** : méthode basée sur la construction de multiples arbres de décision indépendants à partir de sous-échantillons aléatoires du jeu de données. Ce modèle gère bien les relations non linéaires et les interactions complexes entre variables. [28]
- **Support Vector Regression (SVR)** : méthode à noyau qui permet d'estimer des relations complexes dans des espaces de grande dimension. (Utilisation du noyau *rbf*). [29]
- **Gaussian Process Regression (GPR)** : modèle bayésien, qui considère que les observations proviennent d'un processus aléatoire gaussien. [30]

Dans le domaine de l'apprentissage automatique, nous ne parlons plus de variables explicatives mais de prédicteurs. Dans le but d'explorer un large éventail de possibilités, plusieurs jeux de prédicteurs ont été définis et testés.

Cette démarche permet non seulement de comparer les performances entre modèles, mais aussi d'évaluer l'apport informatif de différentes sources de données hyperspectrales.

Les prédicteurs testés incluent :

- Le spectre complet (255 variables)
- Les indices spectraux, calculés à partir des spectres (9 variables)
- Les composantes principales (PCA), calculées à partir des spectres (4 variables)
- L'espèce a été incluse comme variable catégorielle susceptible d'améliorer les prédictions compte tenu d'éventuelles interactions entre les spectres et l'espèce de l'échantillon.
- Des combinaisons de ces sources (par exemple : indices + espèce, ou indices + PCA + espèce).

Remarque : Les composantes latentes issues de la PLS (*Partial Least Squares*) n'ont pas été utilisées ici en raison de la fuite de données (*data leakage*). En effet, ces composantes sont calculées à partir de la variable cible qui doit rester inconnue du modèle au moment de la prédiction.

L'évaluation des performances des modèles de régression repose sur un pipeline de modélisation structuré, illustré à la Figure 6.3-1. Celui-ci suit une logique emboîtée où pour chaque trait fonctionnel, un ensemble de prédicteurs (*features*) est testé sur les différents modèles d'apprentissage.

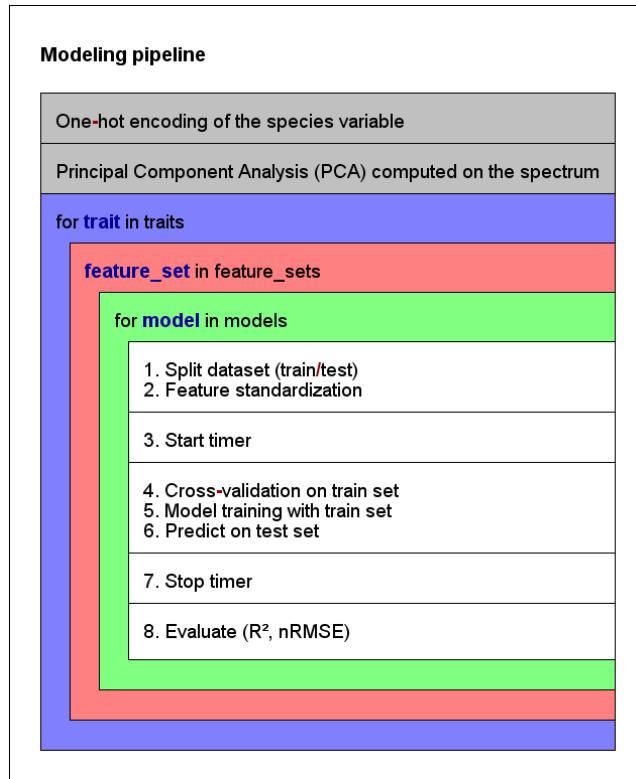


Figure 6.3-1 : Pipeline de modélisation ML approche 2

Deux opérations de prétraitement global sont réalisées en amont des boucles de modélisation :

- **Encodage de la variable "espèce"** : l'espèce est une variable catégorielle. Pour l'inclure comme prédicteur dans les modèles, elle est encodée en binaire via *OneHotEncoder* [31].
- **Calcul des composantes principales (PCA)** réalisé pour les intégrer aux prédicteurs. [13]

Ensuite, le processus de modélisation suit la logique suivante :

1. Une première boucle parcourt chaque trait à prédire.
2. Une seconde boucle teste l'ensemble des combinaisons de prédicteurs.
3. Enfin, une troisième boucle évalue chaque modèle de régression sur les variables d'entrée pour un trait donné.

Pour chaque itération, les étapes suivantes sont réalisées. Elles sont décrites dans le chapitre « Apprentissage automatique supervisé (classification) », et sont reprises ici pour les mêmes raisons :

- Séparation du jeu de données en jeu d'entraînement et jeu de test pour éviter le surajustement (*overfitting*). Essentiel avec un jeu de données de taille limitée.
- Mise à l'échelle des variables d'entrée.
- Validation croisée utilisée pour sélectionner les meilleurs hyperparamètres, définis comme ceux minimisant la RMSE.
- Entraînement du modèle, puis prédiction sur le jeu de test.

Pour évaluer les performances, nous utilisons les mêmes indicateurs que ceux employés dans la Modélisation linéaire multivariée des traits : le coefficient de détermination R^2 et l'erreur quadratique moyenne normalisée (nRMSE).

Dans un contexte prédictif, nous privilégions le nRMSE comme critère de comparaison entre les performances.

Enfin, un chronomètre (*timer*) a été intégré dans le pipeline afin de mesurer le temps d'entraînement de chaque modèle pour donner un aperçu sur leur coût de calcul.

6.4 Approche 3 - Corrélations et modélisation de traits fonctionnels archivés

6.4.1 Objectif

Cette approche vise à valoriser davantage la base de données en testant la corrélation avec des données hyperspectrales mesurées en 2025 et certains traits fonctionnels qui ont été mesurés en 2024.

6.4.2 Jeu de données

Les 28 individus décrits dans le chapitre « Approche 2 - Corrélation » ont également été suivis au cours de l'année 2024. La moyenne des différents traits mesurés au cours de l'été 2024 est disponible pour chacun de ces individus.

Afin de constituer le jeu de données pour cette approche, les spectres et les indices spectraux collectés en 2025 ont été moyennés pour chaque individu en intégrant toutes les feuilles mesurées et l'ensemble des campagnes d'échantillonnage de l'été.

Le jeu de données final comporte donc **28 échantillons**, chacun étant décrit par un spectre hyperspectral moyen mesuré en 2025 et 16 valeurs moyennes de traits mesurés en 2024.

6.4.3 Méthodes

Corrélation de Pearson

Comme dans le chapitre « Corrélation de Pearson » de l'approche 2, une analyse exploratoire par corrélation de Pearson a été réalisée afin d'identifier les longueurs d'onde et les indices spectraux les plus fortement corrélés aux traits fonctionnels mesurés.

Compte tenu de la taille réduite du jeu de données, les corrélations ont été calculées uniquement à l'échelle globale (toutes espèces confondues).

Modélisation linéaire multivariée des traits

La même méthodologie que dans « Modélisation linéaire multivariée des traits » de l'approche 2 a été appliquée afin d'évaluer la capacité de la régression linéaire à modéliser des traits fonctionnels mesurés l'année précédente à l'aide des données hyperspectrales.

Afin de limiter le risque de surajustement à cause de la taille du jeu de données (28 échantillons), seules les deux premières composantes issues de la PCA et de la PLS ont été retenues en tant que variables explicatives dans la régression multivariée.

Contrairement à l'approche 2, les modèles linéaires mixtes n'ont pas été retenus ici car ils ne parviennent pas à converger avec seulement 4 échantillons par espèce. Par conséquent, seuls des modèles linéaires classiques ont été utilisés pour cette approche.

Remarque : Les deux premières composantes principales expliquent **90.06%** de la variance totale des spectres.

7 Résultats

7.1 Résultats - Approche 1

7.1.1 ACP - Analyse en composante principales

L'analyse en composantes principales a été réalisée sur les spectres de réflectance des feuilles (255 variables).

La Figure 7.1-1 ci-dessous présente la part de la variance totale expliquée par les premières composantes principales. Les 3 premières composantes expriment à elles seules environ 85% de la variance totale. Quant aux 7 premières composantes principales, elles expriment 95% de la variance totale.

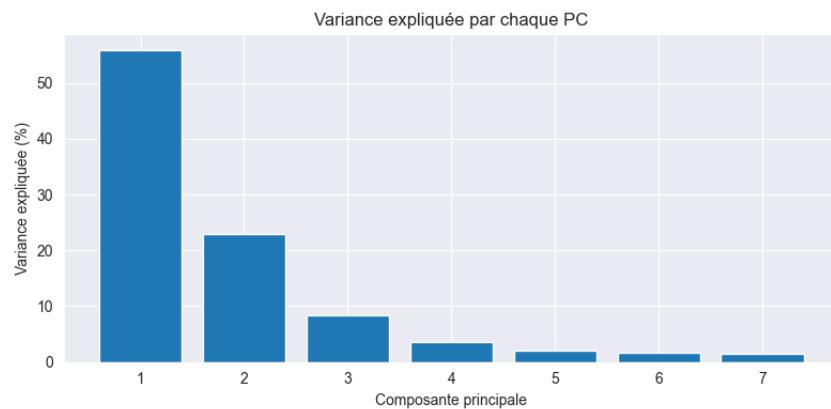


Figure 7.1-1 : Variance expliquée par chaque composante principale

La projection des échantillons dans l'espace défini par les deux premières composantes principales est présentée sur la Figure 7.1-2. Chaque point correspond à une feuille colorée selon l'espèce d'origine et distingués selon l'individu d'origine par différents symboles. Les regroupements des points permettent de visualiser la distribution spectrale des feuilles en fonction des niveaux hiérarchiques (espèces, individus).

(Les détails de la méthode sont présentés dans le chapitre ACP - Analyse en composantes principales)



Figure 7.1-2 : projection des échantillons selon PC1 vs PC2

7.1.2 Visualisation de PC1 selon différents niveaux

La distribution des valeurs de la première composante principale a été visualisée selon les espèces (Figure 7.1-3), les individus (Figure 7.1-4), puis les branches (Figure 7.1-5).

Ces visualisations permettent d'observer la dispersion des spectres projetés sur PC1 au sein de chaque groupe.

Les numéros qui accompagnent les espèces sur les deux derniers graphiques sont les identifiants uniques des individus qui permettent de les différencier.

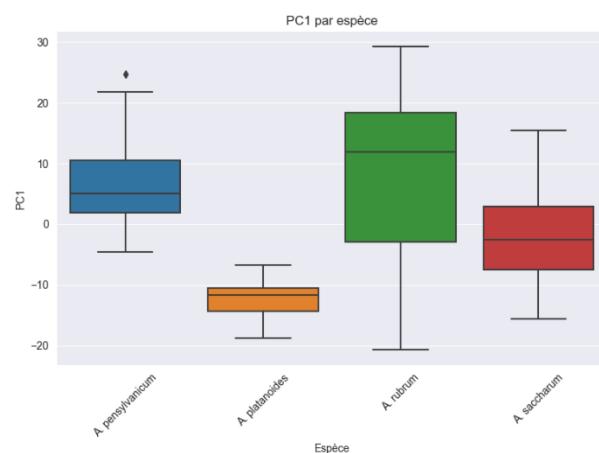


Figure 7.1-3 : Visualisation de PC1 par espèce

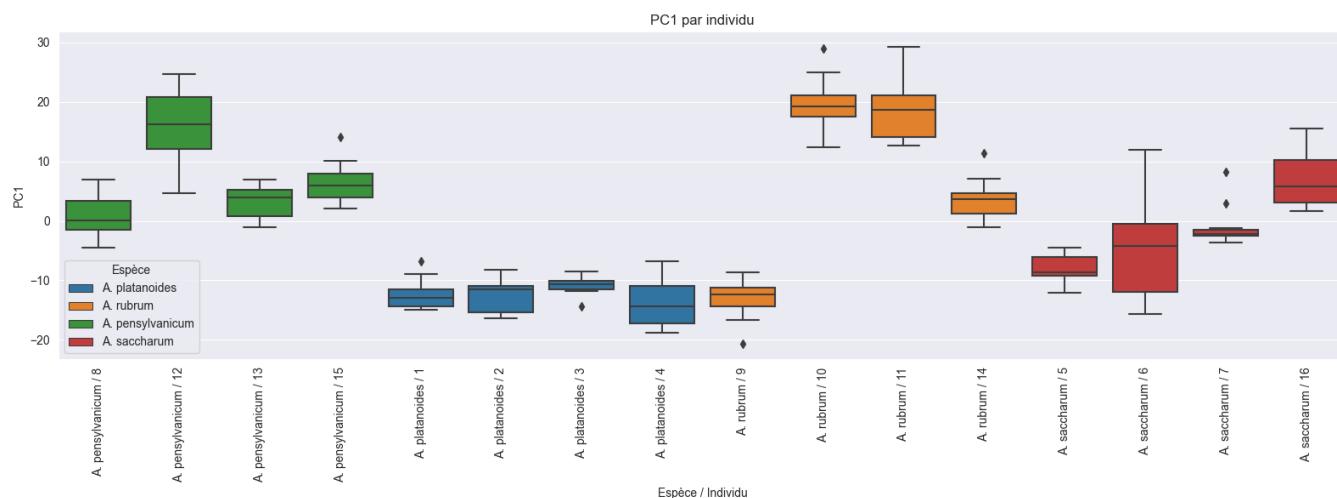


Figure 7.1-4 : Visualisation de PC1 par individu

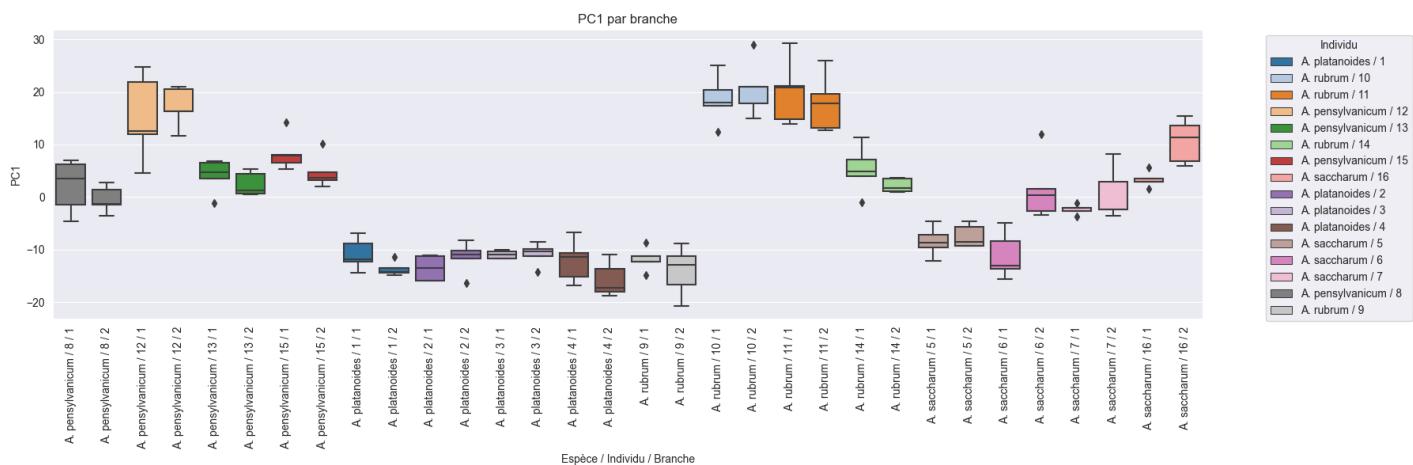


Figure 7.1-5 : Visualisation de PC1 par branche

7.1.3 Analyse de la variance

Les figures ci-dessous présentent les résultats des différentes analyses de variance :

- **Analyse inter-espèce par longueur d'onde** (Figure 7.1-6) : elle met en évidence les plages spectrales pour lesquelles les différences entre espèces sont statistiquement significatives.
- **Analyse intra-espèces par longueur d'onde** (Figure 7.1-7) : cette analyse permet d'observer comment la variabilité spectrale entre individus varie selon les espèces et selon les longueurs d'onde.
- **Analyse intra-individuelle par longueur d'onde** (Figure 7.1-8) : elle met en évidence la variance entre les rameaux au sein d'un même individu. Une courbe par individus a été calculée mais pour améliorer la lisibilité, les résultats ont été moyennés par espèce. (Chaque courbe présentée correspond donc à la moyenne des tests effectués sur les individus d'une même espèce).
- **Analyse inter-espèce par indice spectral** (Figure 7.1-10) : les barplots permettent de visualiser, pour chaque indice spectral, le niveau de significativité des différences entre espèces.
- **Analyse intra-espèces par indice spectral** (Figure 7.1-9) : cette analyse montre pour chaque espèce quels indices spectraux varient le plus au sein d'une même espèce.

Une ligne horizontale correspondant au seuil $p = 0.05$ est tracée sur chaque graphique, soit le seuil de significativité conventionnel [32]. Cela signifie que les valeurs situées au-dessus de cette ligne indiquent une différence suffisamment forte pour qu'elles aient moins de 5 % de chances d'être due au hasard.

(*Les détails de la méthode sont présentés dans le chapitre Analyse de la variance*)

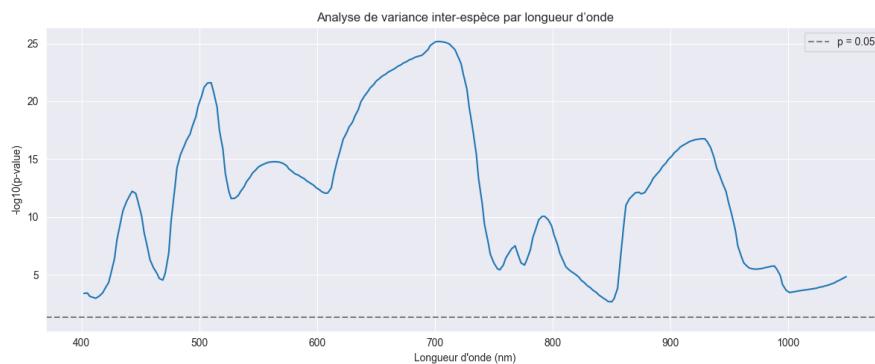


Figure 7.1-6 : Analyse de variance inter-espèce par longueur d'onde

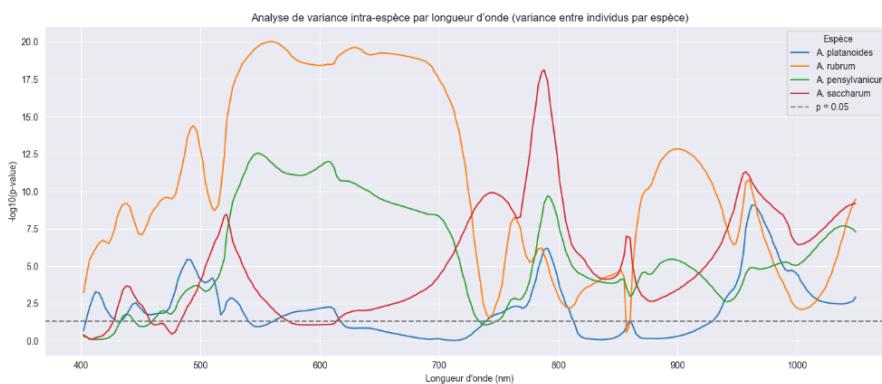


Figure 7.1-7 : Analyse de variance intra-espèces par longueur d'onde



Figure 7.1-8 : Analyse de variance intra-individu par longueur d'onde

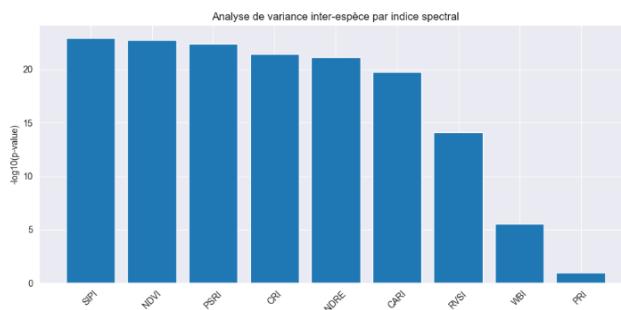


Figure 7.1-10 : Analyse de variance inter-espèce par indice spectral

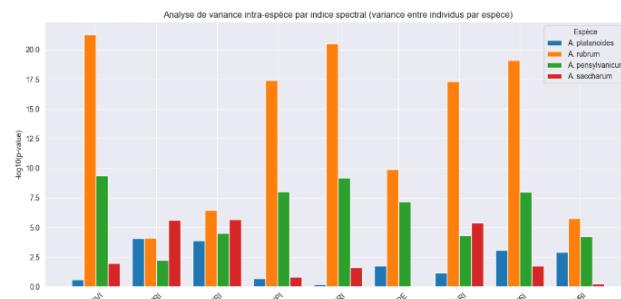


Figure 7.1-9 : Analyse de variance intra-espèces par indice spectral

7.1.4 Apprentissage automatique supervisé (classification)

L'évaluation des performances des trois modèles de classification (KNN, Random Forest et SVM) a été réalisée pour chacun des types de prédicteurs testés. Les performances sont résumées dans le Tableau 5 ci-dessous.

(Les détails de la méthode sont présentés dans le chapitre Apprentissage automatique supervisé (classification))

Tableau de performance de classification de l'espèce

Prédicteurs d'entrée	Modèle	Performances	
		Accuracy	F1
PC1 à PC7	KNN	0.750	0.745
	RandomForest	0.781	0.781
	SVM	0.843	0.841
Toutes les bandes spectrales	KNN	0.750	0.756
	RandomForest	0.875	0.874
	SVM	0.843	0.844
Bandes spectrales les plus significatives	KNN	0.562	0.530
	RandomForest	0.562	0.534
	SVM	0.656	0.625
Indices spectraux	KNN	0.781	0.780
	RandomForest	0.843	0.843
	SVM	0.781	0.783

PC1 à PC7 + Indices spectraux	KNN	0.750	0.745
	RandomForest	0.906	0.905
	SVM	0.750	0.742

Tableau 5 : Résumé des performances de classification d'espèce

Les cinq combinaisons modèle-caractéristiques les plus performantes ont été mises en gras dans le tableau ci-dessus. Pour chacune de ces meilleures configurations, la matrice de confusion ainsi que les meilleurs hyperparamètres utilisés sont décrits ci-dessous.

Remarque : les matrices de confusion sont calculées sur le jeu de test représentant 20 % du jeu de données total, soit 32 échantillons (8 par espèce). Elles permettent de voir quelles espèces sont confondues entre elles.

1. Modèle : RandomForest, Caractéristiques : PC1 à PC7 + Indices spectraux

Le modèle prédit 90.6% des espèces correctement avec un f1-score de 90.5%.

Les meilleurs hyperparamètres utilisés sont :

- Max_depth : none
- Min_sample_split : 5
- N_estimator : 100

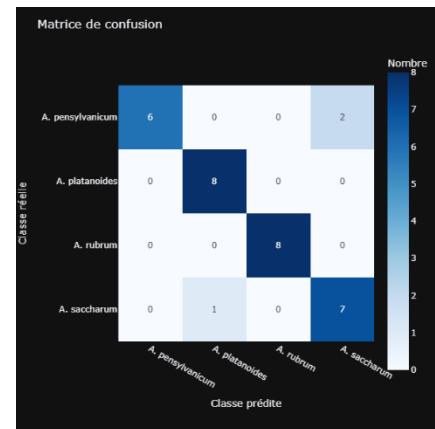


Figure 7.1-11 : Matrice de confusion RF et PC1 à PC7 avec Indices spectraux

2. Modèle : RandomForest, Caractéristiques : Toutes les bandes spectrales

Le modèle prédit 87.5% des espèces correctement avec un f1-score de 87.4%.

Les meilleurs hyperparamètres utilisés sont :

- Max_depth : none
- Min_sample_split : 5
- N_estimator : 100

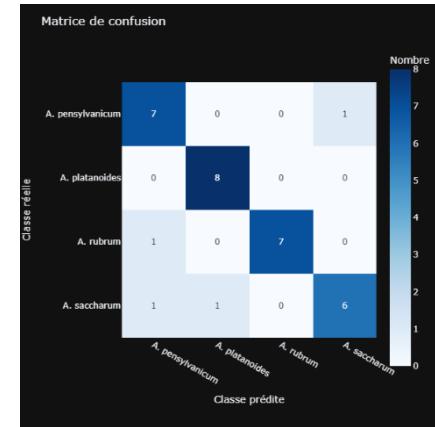


Figure 7.1-12 : Matrice de confusion RF et toutes les bandes spectrales

3. Modèle : SVM, Caractéristiques : Toutes les bandes spectrales

Le modèle prédit 84.3% des espèces correctement avec un f1-score de 84.4%.

Les meilleurs hyperparamètres utilisés sont :

- C : 10
- Gamma : auto
- Kernel : rbf

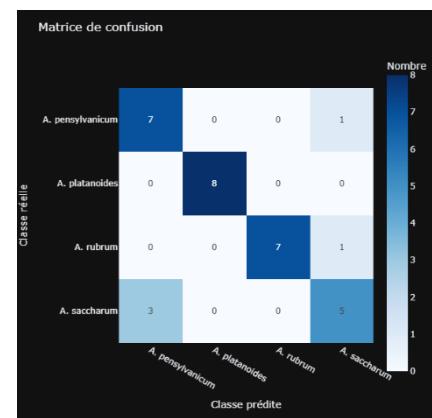


Figure 7.1-13 : Matrice de confusion SVM et toutes les bandes spectrales

4. Modèle : RandomForest, Caractéristiques : Indices spectraux

Le modèle prédit 84.3% des espèces correctement avec un f1-score de 84.3%.

Les meilleurs hyperparamètres utilisés sont :

- Max_depth : 5
- Min_sample_split : 2
- N_estimators : 50

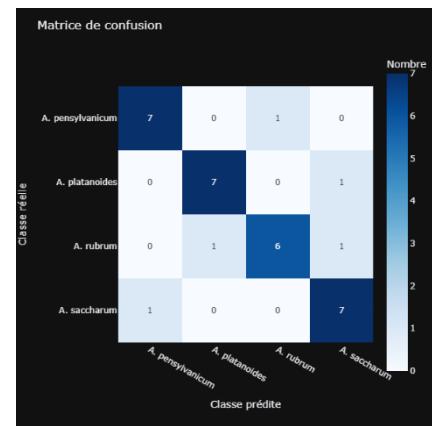


Figure 7.1-14 : Matrice de confusion RF et Indices spectraux

5. Modèle : SVM, Caractéristiques : PC1 à PC7

Le modèle prédit 84.3% des espèces correctement avec un f1-score de 84.1%.

Les meilleurs hyperparamètres utilisés sont :

- C : 10
- Gamma : scale
- Kernel : rbf

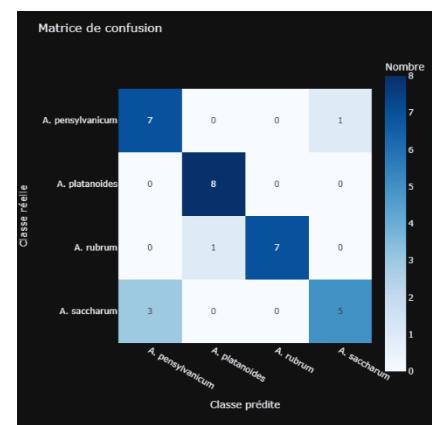


Figure 7.1-15 : Matrice de confusion SVM et PC1 à PC7

7.2 Résultats - Approche 2

7.2.1 Corrélation de Pearson

Comme indiqué au chapitre « Corrélation de Pearson », le coefficient (r) est calculé une fois dans le jeu de données avec toutes les espèces confondues puis séparément pour chaque espèce.

Corrélations toutes espèces confondues

La Figure 7.2-1 ci-dessous présente pour chaque trait fonctionnel, un profil de corrélation continue avec les longueurs d'onde. Les valeurs du coefficient de Pearson varient selon les traits et les plages spectrales.

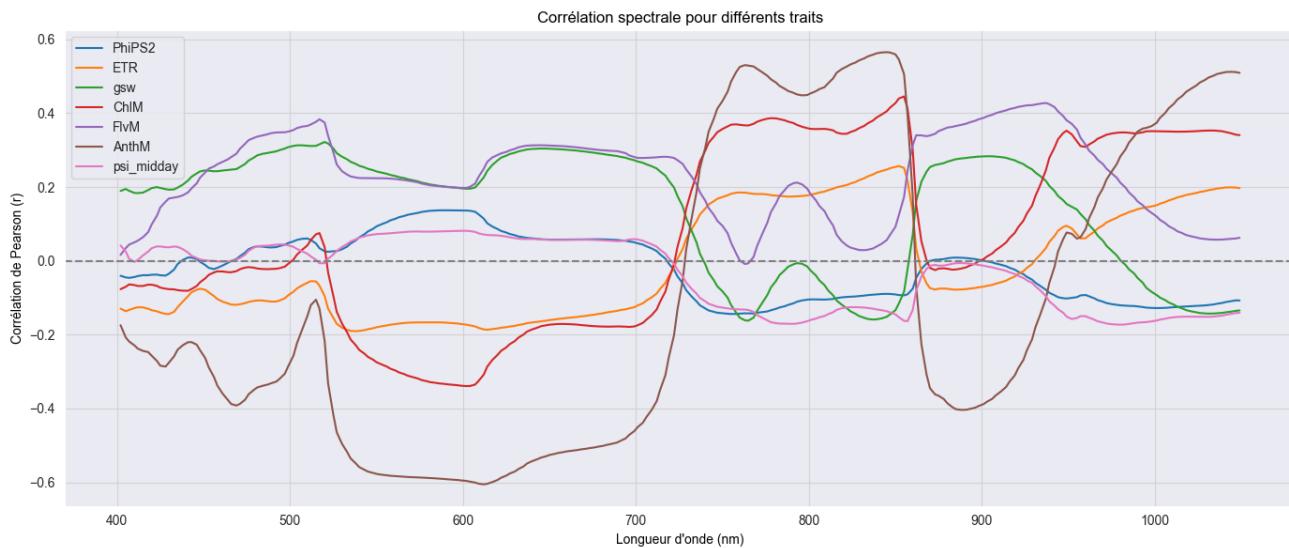


Figure 7.2-1 : Corrélation spectrale (Pearson) entre chaque bande spectrale (400–1050 nm) et les traits fonctionnels, toutes espèces confondues

Une matrice de corrélation a été calculée entre les indices spectraux et les traits fonctionnels (Figure 7.2-2). Chaque cellule de la matrice indique le coefficient de Pearson (r) pour une paire indice-trait.

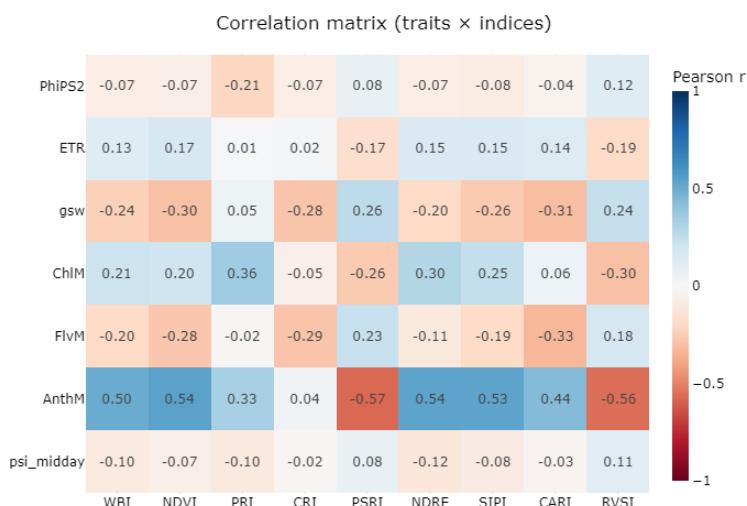


Figure 7.2-2 : Matrice de corrélation (Pearson) entre les indices spectraux et les traits fonctionnels, toutes espèces confondues

Corrélation par espèce

Les spectres de corrélation et les matrices de corrélation ont également été calculées séparément pour chaque espèce.

Les profils de corrélation entre chaque bande spectrale et trait sont regroupées dans la Figure 7.2-3, tandis que les matrices de corrélations indices x traits sont affichées dans la Figure 7.2-4.



Figure 7.2-3 : Corrélation (Pearson) entre les bandes spectrales et traits fonctionnels, calculées séparément pour chaque espèce

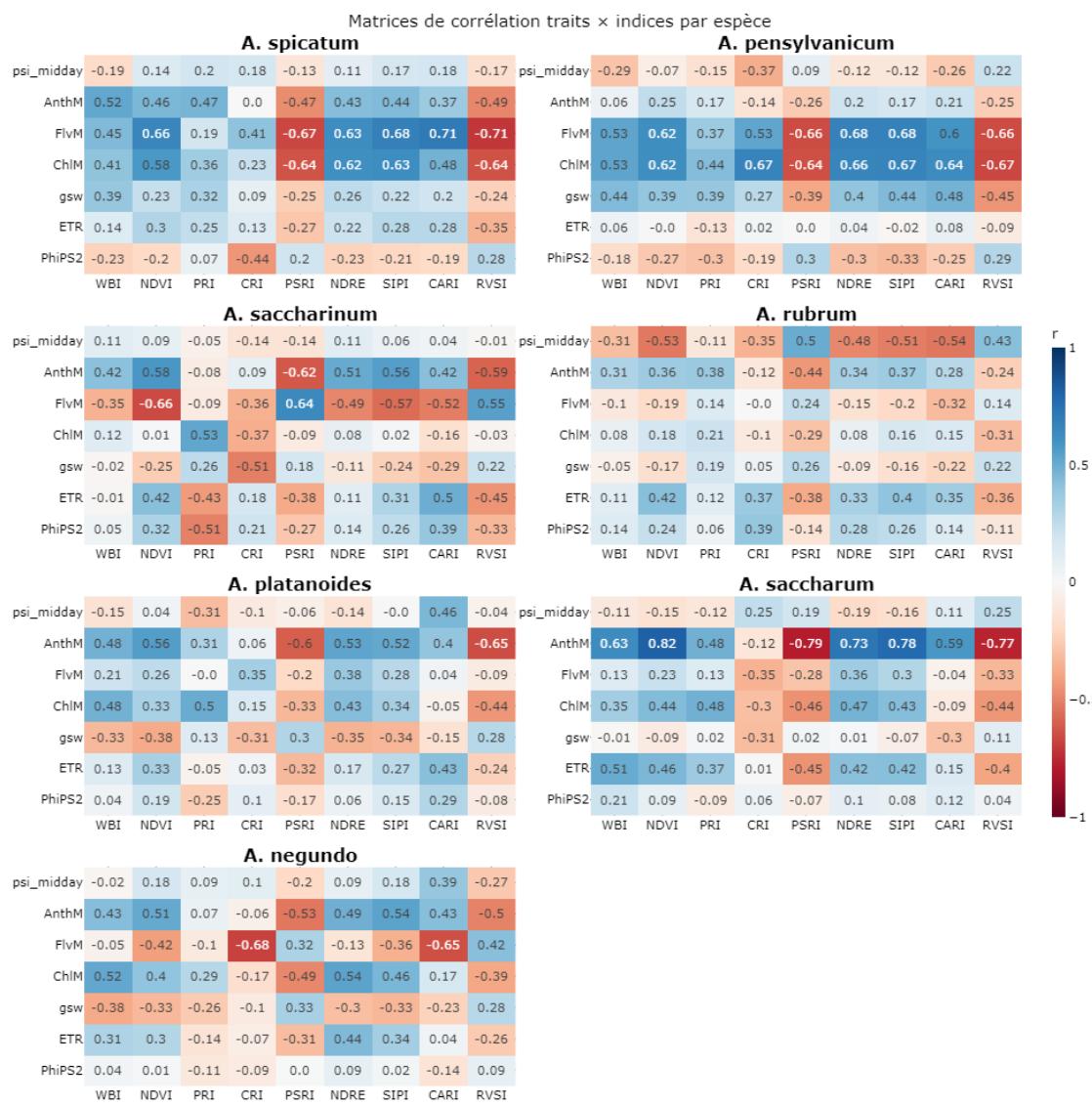


Figure 7.2-4 : Matrices de corrélation (Pearson) entre indices spectraux et traits fonctionnels, calculées séparément pour chaque espèce

7.2.2 Modélisation linéaire multivariée des traits

Les performances de la régression linéaire ont été évaluées pour chaque trait en testant différentes combinaisons de variables explicatives issues des données hyperspectrales. Deux types de modélisations ont été considérés : sans effet aléatoire (modèles linéaires classiques) et avec effet aléatoire lié à l'espèce.

(Les détails de la méthode sont présentés dans le chapitre Modélisation linéaire multivariée des traits)

La Figure 7.2-5 ci-dessous présente les résultats de R^2 obtenus pour chaque trait en fonction des différentes combinaisons de variables testées. Les résultats sont séparés selon la présence ou non de l'effet aléatoire « espèce ».

Comparaison des R^2 selon la présence d'un effet aléatoire et le type de variables explicatives

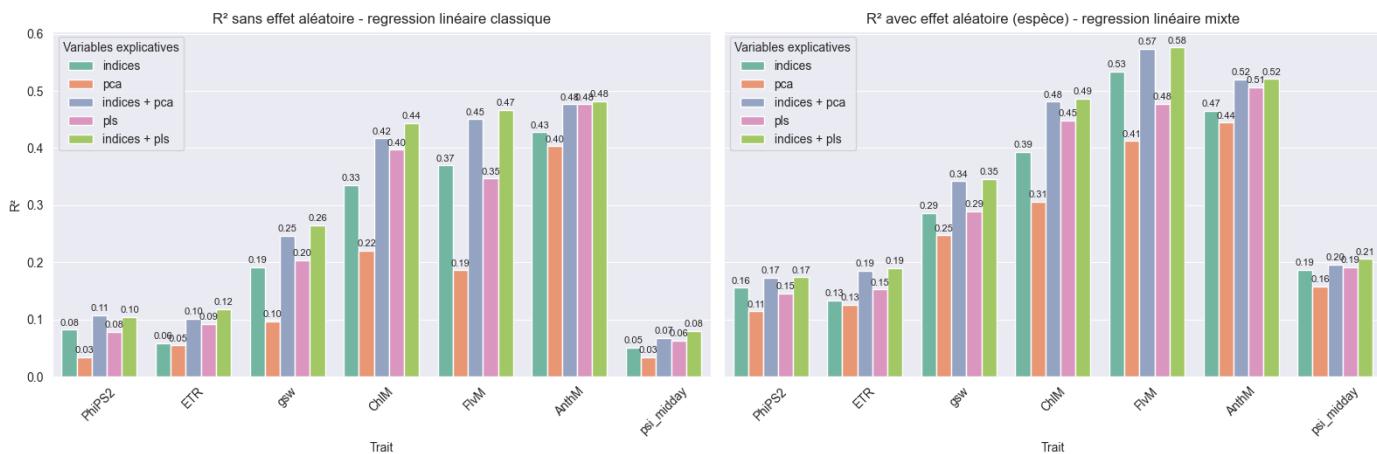


Figure 7.2-5 : Comparaison des R^2 selon la présence d'un effet aléatoire et le type de variables explicatives

La Figure 7.2-6 ci-dessous présente les erreurs (nRMSE en pourcentage) entre les valeurs réelles des traits et les valeurs prédictes par les modèles de régression.

Comparaison des nRMSE (%) selon la présence d'un effet aléatoire et le type de variables explicatives

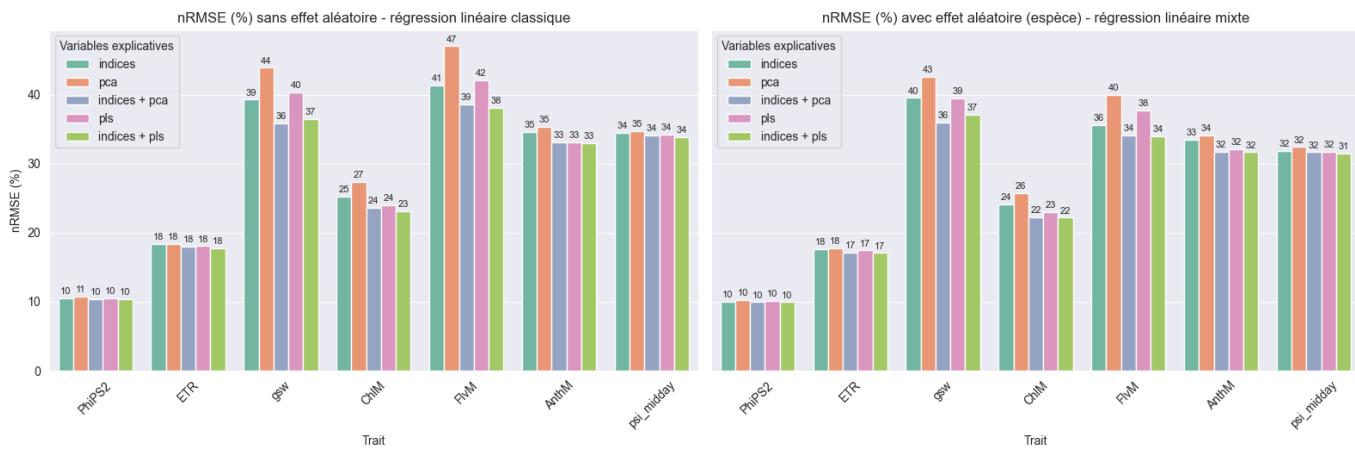


Figure 7.2-6 : Comparaison des nRMSE (%) selon la présence d'un effet aléatoire et le type de variables explicatives

Le Tableau 6 présente, pour chaque trait, le meilleur score de R^2 obtenu à l'aide de la régression linéaire, accompagné de la valeur du nRMSE. Il précise également la configuration associée : présence ou non d'un effet aléatoire lié à l'espèce, ainsi que les variables explicatives utilisées.

Tableau des meilleurs scores par trait

Trait	R^2	nRMSE	Variables explicatives	Effet aléatoire
FvM	0.575	0.339	Indices + PLS	Espèce
AnthM	0.520	0.316	Indices + PLS	Espèce
ChlM	0.486	0.221	Indices + PLS	Espèce
gsw	0.346	0.370	Indices + PLS	Espèce
psi_midday	0.205	0.314	Indices + PLS	Espèce
ETR	0.189	0.170	Indices + PLS	Espèce
PhiPS2	0.173	0.099	Indices + PLS	Espèce

Tableau 6 : Tableau du meilleur score R^2 obtenu pour chaque trait (approche 2)

Afin d'identifier quels types de variables explicatives permettent de mieux modéliser les traits de manière générale, nous calculons la moyenne des R^2 obtenus pour chaque combinaison de variables en regroupant l'ensemble des traits ainsi que les configurations avec ou sans effet aléatoire. La Figure 7.2-7 ci-dessous présente les scores moyens.

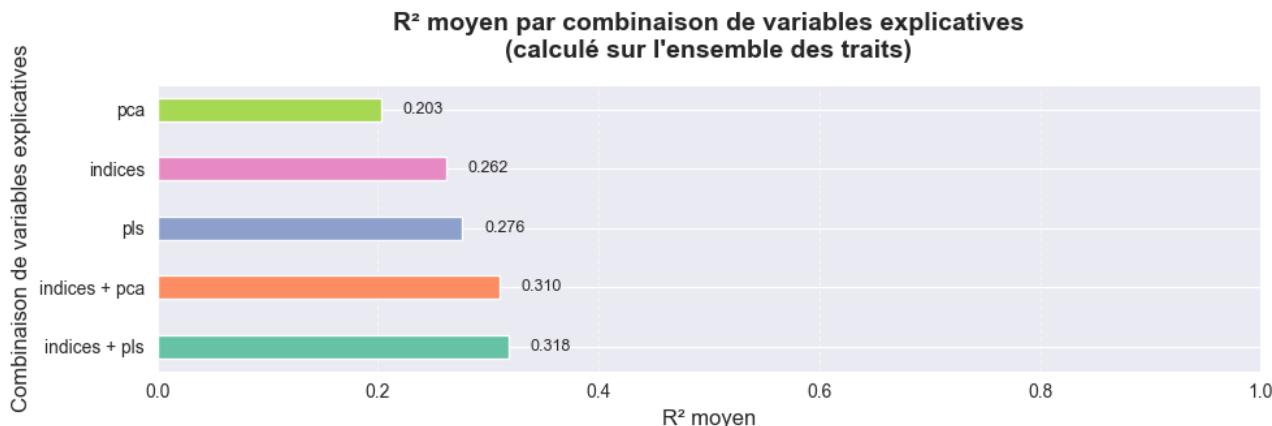


Figure 7.2-7 : R^2 moyen par combinaison de variables explicatives (calculé sur l'ensemble des traits)

Enfin, la moyenne globale des R^2 a été calculée selon la présence ou non de l'effet aléatoire de l'espèce dans les modèles de régression (traits et type de variables explicatives confondues). Sans effet aléatoire, le R^2 moyen observé est de **0.229**. Avec l'effet aléatoire d'espèce inclus, ce R^2 moyen s'élève à **0.318**. Soit une augmentation générale du score R^2 de **38.8%**.

7.2.3 Apprentissage automatique supervisé (régression)

Les performances de différents modèles de régression par apprentissage automatique ont été évaluées pour chaque trait mesuré. Chaque modèle a été testé en combinaison avec plusieurs prédicteurs dérivées des données hyperspectrales.

(Les détails de la méthode sont présentés dans le chapitre Apprentissage automatique supervisé (régression))

Pour chaque modèle RFR (Figure 7.2-9), SVR (Figure 7.2-8) et GPR (Figure 7.2-10), les figures présentent les performances de prédiction pour chaque trait fonctionnel, évaluées sur le jeu de test à l'aide du nRMSE (%) et du R^2 . Chaque barre correspond à un jeu de prédicteurs.

Performance des prédictions de chaque trait pour le modèle : RFR

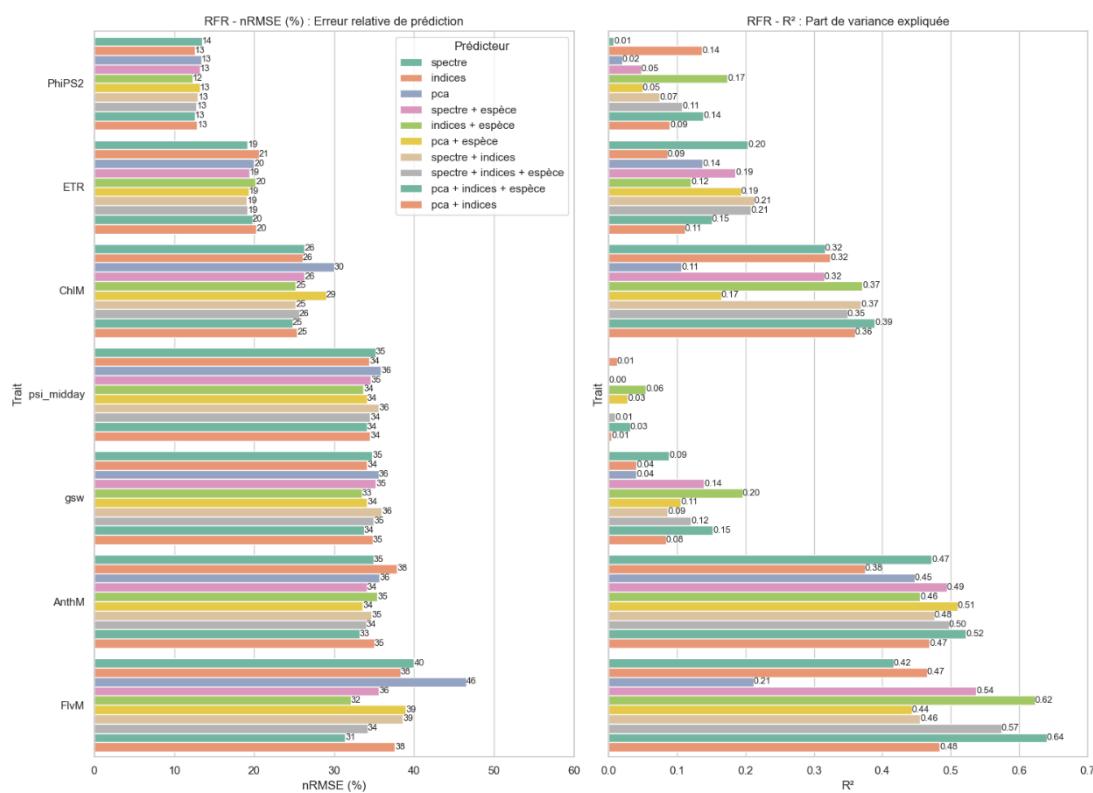


Figure 7.2-9 : Performances de prédiction par trait pour le modèle RFR évalué à l'aide du nRMSE (%) et du R² sur le jeu de test

Performance des prédictions de chaque trait pour le modèle : SVR

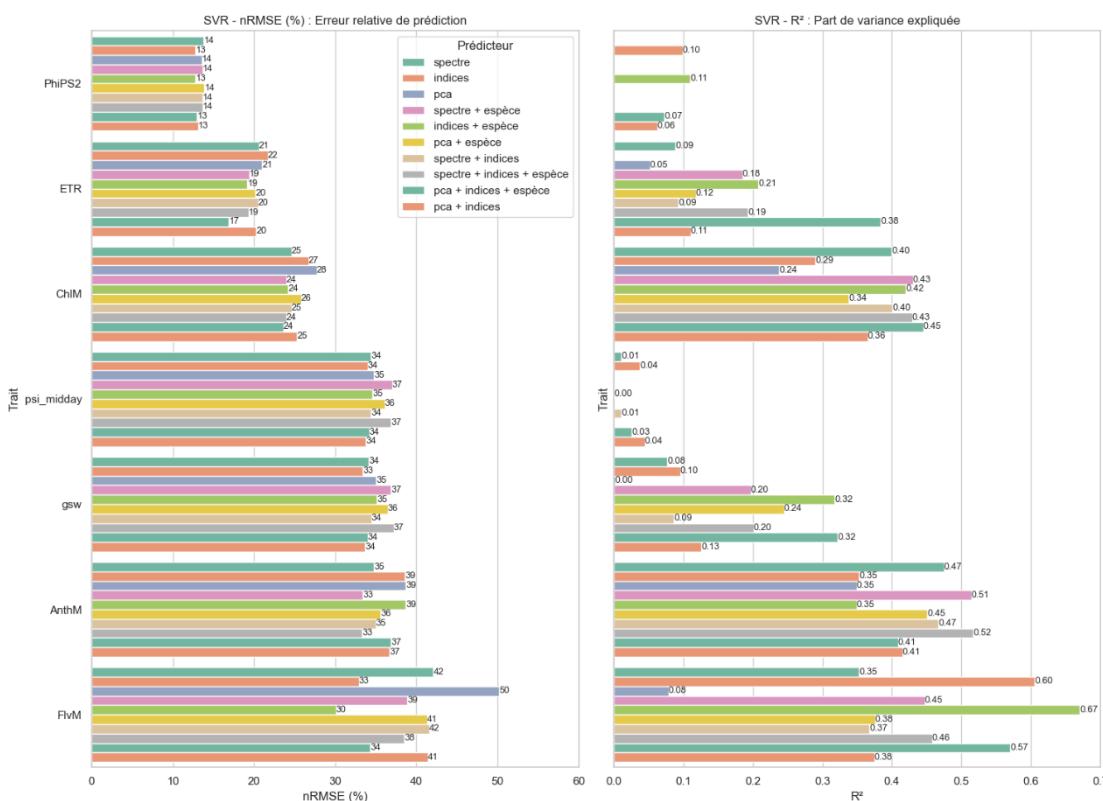


Figure 7.2-8 : Performances de prédiction par trait pour le modèle SVR évalué à l'aide du nRMSE (%) et du R² sur le jeu de test

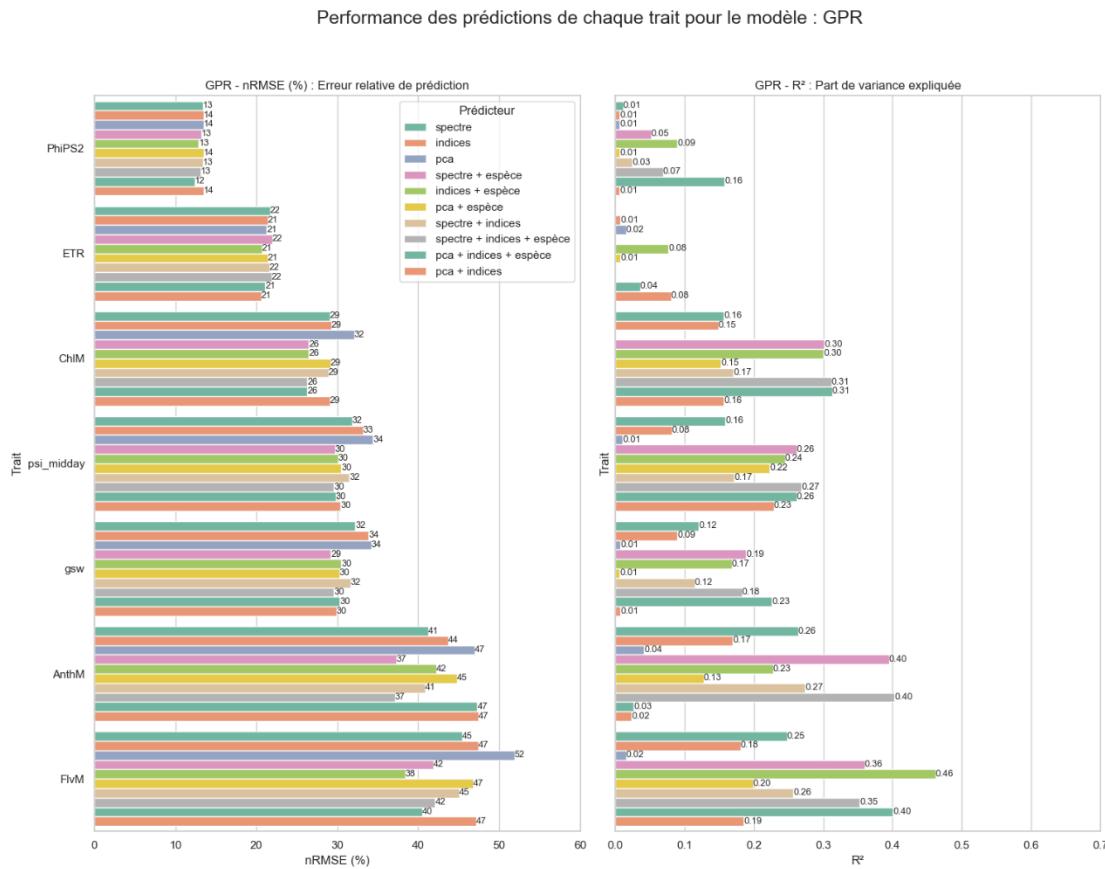


Figure 7.2-10 : Performances de prédiction par trait pour le modèle GPR évalué à l'aide du nRMSE (%) et du R^2 sur le jeu de test

Le Tableau 7 ci-dessous synthétise, pour chaque trait, les meilleures performances de prédiction obtenues : il indique le nRMSE minimal, le modèle et les prédicteurs correspondants, ainsi que la part de variance expliquée (R^2).

Tableau des meilleurs scores par trait

Trait	Modèle	Prédicteurs	nRMSE [%]	R2
PhiPS2	RFR	Indices + Espèce	12.3	0.174
ETR	SVR	PCA + Indices + Espèce	16.9	0.384
ChIM	SVR	PCA + Indices + Espèce	23.6	0.445
gsw	GPR	Spectre + Espèce	29.2	0.189
psi_midday	GPR	Spectre + Indices + Espèce	29.6	0.269
FlvM	SVR	Indices + Espèce	30.1	0.670
AnthM	RFR	PCA + Indices + Espèce	33.2	0.522

Tableau 7 : Tableau du meilleur score nRMSE obtenu pour chaque trait (approche 2)

La moyenne des nRMSE (%) a été calculée pour chaque prédicteur, en regroupant l'ensemble des traits et des modèles. La Figure 7.2-11 présente ces moyennes, qui permettent de comparer globalement la performance prédictive de chaque type de prédicteurs hyperspectraux.

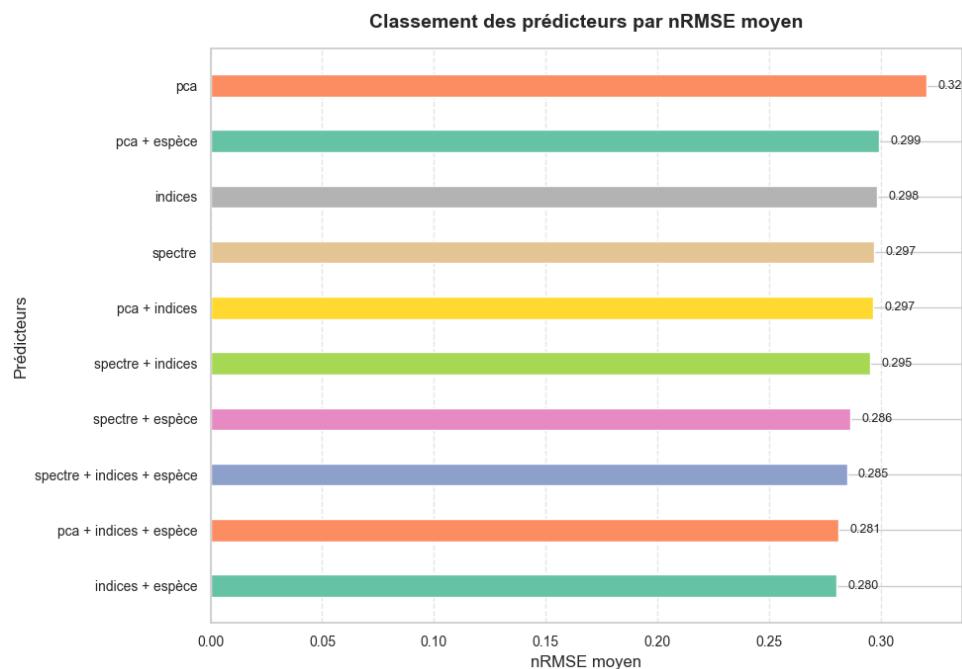


Figure 7.2-11 : Erreur de prédiction moyenne (nRMSE) selon le type de prédicteur hyperspectral

De manière analogue, nous avons calculé la moyenne des erreurs de prédiction obtenues pour chaque algorithme (tous traits et prédicteurs confondues) afin de comparer globalement la performance des modèles (Figure 7.2-13).

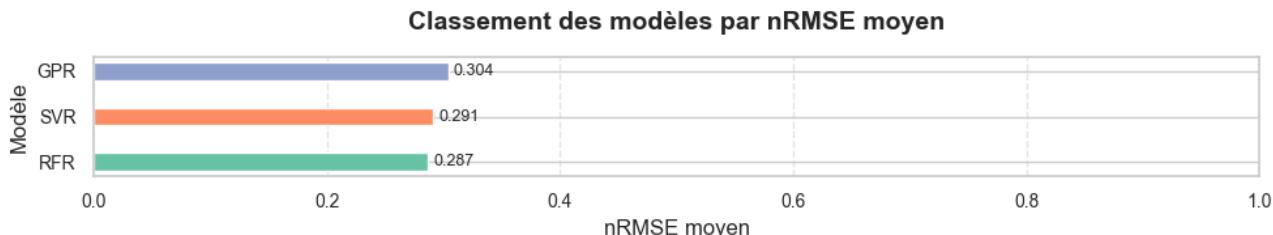


Figure 7.2-13 : Erreur de prédiction moyenne (nRMSE) selon le modèle de régression

Enfin, la Figure 7.2-12 présente le temps moyen d'entraînement de chaque modèle de régression pour comparer leur coût computationnel.

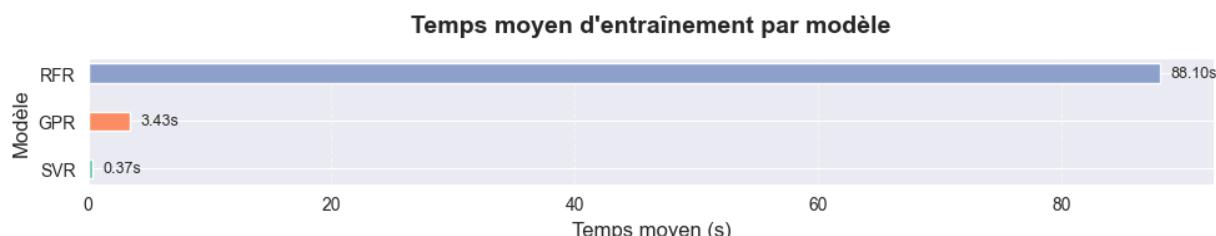


Figure 7.2-12 : Temps moyen d'entraînement par modèle de régression

7.3 Résultats - Approche 3

7.3.1 Corrélation de Pearson

La Figure 7.3-1 ci-dessous présente les valeurs du coefficient de corrélation de Pearson (r) calculées entre chaque bande spectrale mesurée en 2025 et chacun des 16 traits fonctionnels mesurés en 2024.

(Les détails de la méthode sont présentés dans le chapitre Corrélation de Pearson)

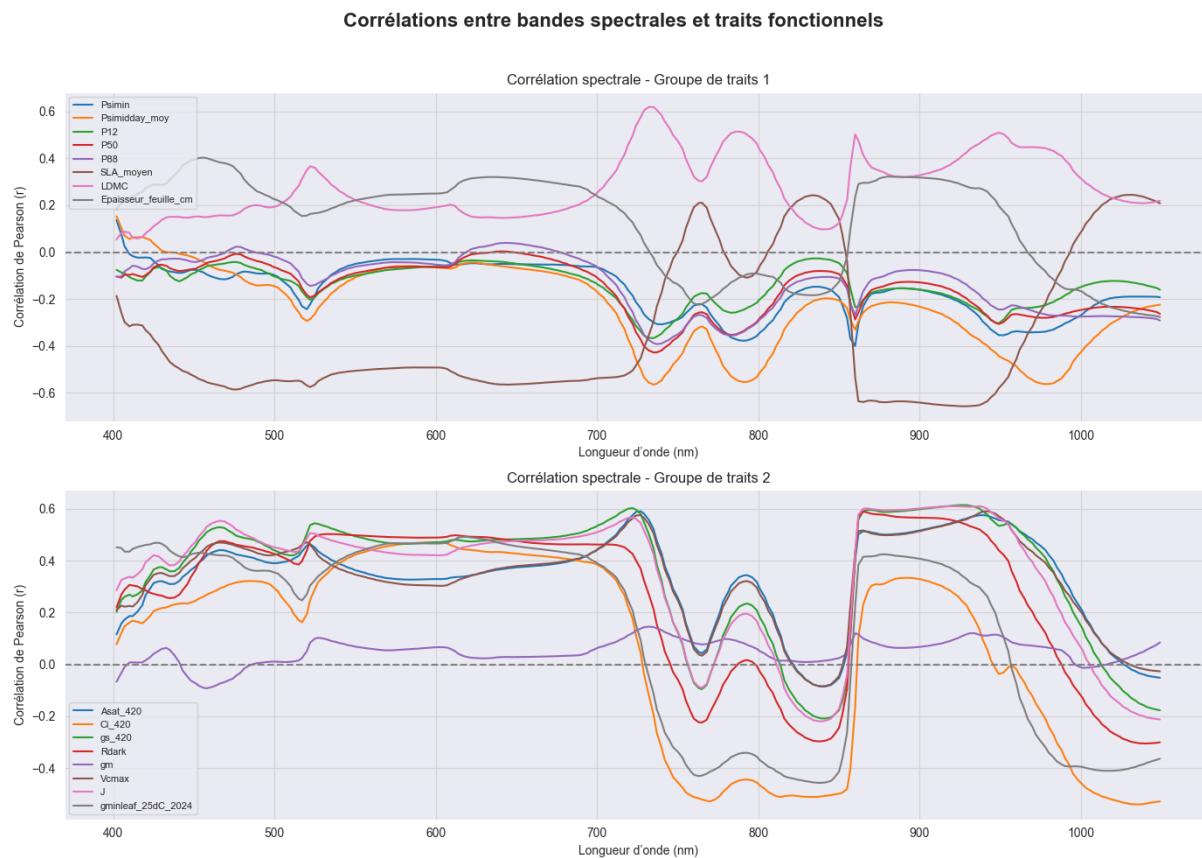


Figure 7.3-1 : Corrélation de Pearson entre chaque bande spectrale et chaque trait mesuré l'année précédente

Les valeurs de coefficient de Pearson, disposées dans une matrice de corrélation (Figure 7.3-2), ont aussi été calculées entre les indices spectraux et les traits fonctionnels.

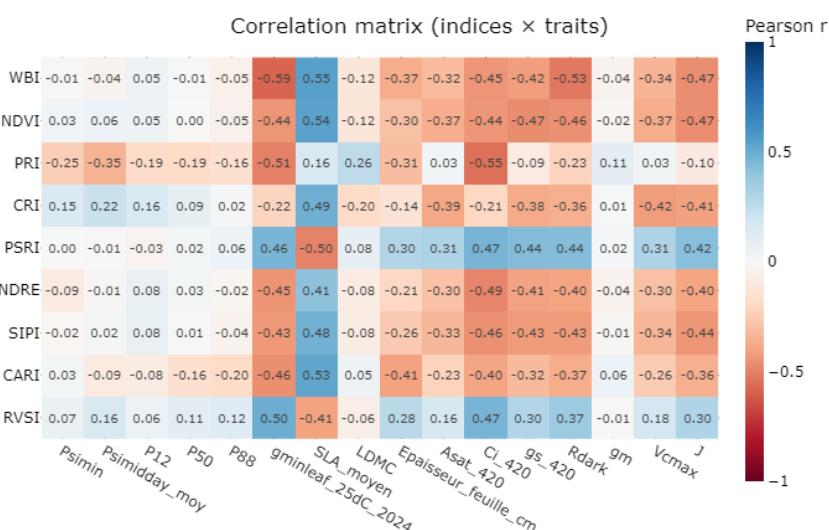


Figure 7.3-2 : Corrélation de Pearson entre chaque indice spectral et chaque trait mesuré l'année précédente

7.3.2 Modélisation linéaire multivariée des traits

Nous évaluons, pour chaque trait fonctionnel, la performance de modèles de régression linéaire construits à partir des composantes principales (PCA) et des composantes latentes (PLS) calculées sur les spectres.

(*Les détails de la méthode sont présentés dans le chapitre Modélisation linéaire multivariée des traits*)

La Figure 7.3-3 ci-dessous présente le score R^2 obtenu pour chaque trait et pour chaque variable explicative.

Comparaison des R^2 selon le type de variables explicatives - regression linéaire classique

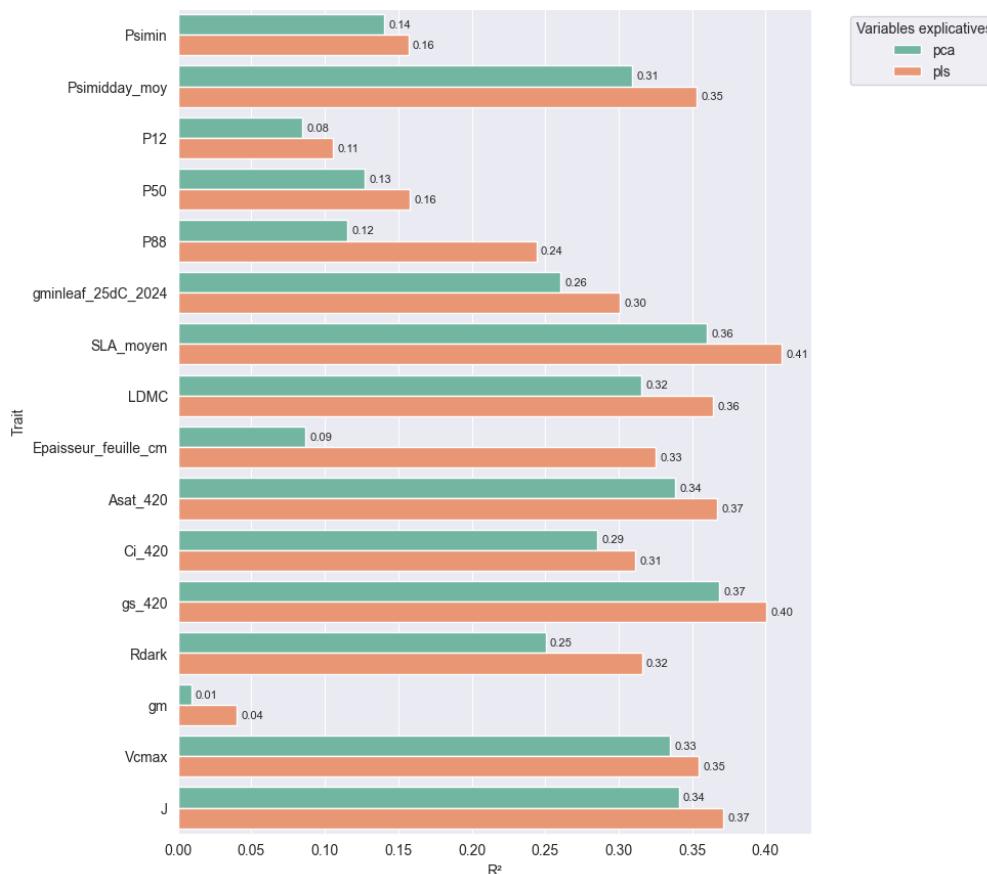


Figure 7.3-3 : Comparaison des R^2 pour chaque trait selon le type de variable explicative

La Figure 7.3-4 présente l'erreur relative des prédictions des modèles linéaires (nRMSE) pour chaque combinaison de trait et de variables explicatives.

Comparaison des nRMSE (%) selon le type de variables explicatives - regression linéaire classique

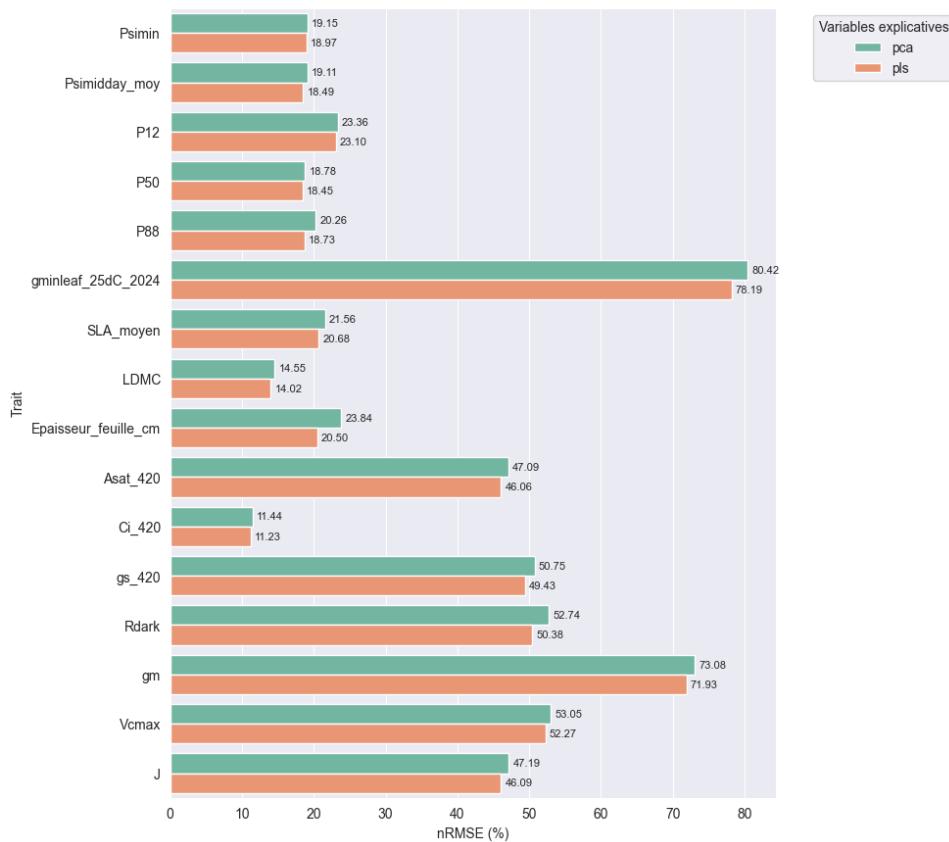


Figure 7.3-4 : Comparaison des nRMSE (%) pour chaque trait selon le type de variable explicative

Enfin, le Tableau 8 récapitule pour chaque trait le meilleur score de R^2 obtenu, ainsi que la valeur de nRMSE correspondante et les variables explicatives ayant permis d'atteindre ce score.

Tableau des meilleurs scores par trait

Trait	R^2	nRMSE	Variables explicatives
SLA_moyen	0.410	0.206	PLS
gs_420	0.400	0.494	PLS
J	0.371	0.460	PLS
Asat_420	0.367	0.460	PLS
LDMC	0.364	0.140	PLS
Vcmax	0.354	0.522	PLS
Psimidday_moy	0.353	0.184	PLS
Epaisseur_feuille_cm	0.325	0.204	PLS
Rdark	0.315	0.503	PLS
Ci_420	0.311	0.112	PLS
gminleaf_25dC_2024	0.300	0.781	PLS
P88	0.244	0.187	PLS
P50	0.157	0.184	PLS
Psimin	0.156	0.189	PLS
P12	0.105	0.230	PLS
gm	0.040	0.719	PLS

 Tableau 8 : Tableau du meilleur score R^2 obtenu pour chaque trait (approche 3)

8 Discussion

8.1 Interprétations des principaux résultats

8.1.1 Approche 1 - Analyse de la variabilité inter- et intra-spécifique

La projection des spectres dans le plan PC1-PC2 (7.1.1) met en évidence des regroupements par espèces ainsi qu'une variabilité différente selon l'espèce.

L'analyse de variance (7.1.3) confirme ces observations. Elle indique une différence significative entre les spectres inter-espèces (*valeurs de p < 5 %*) et met en évidence des variations intra-espèces notables. En revanche, nous n'avons pas détecté une différence significative au niveau intra-individu ($p > 0,5$). Nous avons également observé que les bandes spectrales qui varient le plus diffèrent selon l'espèce considérée, ce qui montre que chaque espèce possède des signatures spectrales caractéristiques.

Les modèles d'apprentissage automatique (7.1.4) confirment le potentiel discriminant des données hyperspectrales vis-à-vis de l'espèce. Malgré un jeu de données limité en taille, nous atteignons jusqu'à 90 % de précision dans la prédiction de l'espèce des feuilles à partir des spectres.

8.1.2 Approche 2 - Corrélations et modélisation des traits fonctionnels

L'analyse des corrélations entre les spectres et les traits fonctionnels (7.2.1) a permis d'identifier des plages spectrales informatives qui dépendent fortement du trait considéré. Les corrélations calculées espèce par espèce sont généralement plus élevées que celles obtenues sur l'ensemble des données (toutes espèces confondues), ce qui suggère que les liens entre spectres et traits fonctionnels dépendent de l'espèce.

Les modèles linéaires classiques (7.2.2) ont permis d'améliorer les performances par rapport à la corrélation simple, en particulier la méthode PLS (*Partial Least Squares*). En intégrant l'espèce comme étant un effet aléatoire via un modèle de régression linéaire mixte, nous atteignons des valeurs de R^2 encore meilleures avec des scores supérieurs à 50 % pour certains traits. Cela confirme à nouveau le rôle clé de l'espèce dans les relations entre spectres et traits.

Les modèles d'apprentissage automatique testés (*Random Forest, SVR, GPR*) (7.2.3) prédisent la valeur des traits avec une erreur relative comprise entre 12 % et 33 % selon le trait considéré. L'intégration de la variable espèce dans les prédicteurs améliore systématiquement les performances. En termes d'algorithmes, le *Random Forest* est le plus performant de façon générale mais son temps d'entraînement est relativement élevé. Le *SVR* est un bon compromis entre précision et temps d'entraînement plus rapide.

Afin de comparer de manière synthétique les performances des différentes approches :

- Le Tableau 9 présente pour chaque trait fonctionnel, la valeur maximale de R^2 obtenue selon cinq méthodes : corrélation avec une bande spectrale, corrélation avec un indice spectral, régression linéaire classique, régression linéaire mixte, et modèles d'apprentissage automatique.
- Le Tableau 10 indique pour chaque trait, la plus faible erreur normalisée (nRMSE) obtenue à l'aide des modèles de régression ou d'apprentissage automatique.

Meilleur R² par trait fonctionnel selon la méthode

Trait	Bandé (R ² , λ)	Indice spectral (R ² , nom)	Régr. lin. classique (R ²)	Régr. lin. mixte (R ²)	Modèle ML (R ² , modèle)
ChlM	0.198 (855 nm)	0.128 (PRI)	0.442	0.486	0.445 (SVR)
gsw	0.104 (520 nm)	0.094 (CARI)	0.259	0.346	0.189 (GPR)
ETR	0.066 (852 nm)	0.035 (RVSI)	0.124	0.190	0.384 (SVR)
PhiPS2	0.021 (755 nm)	0.046 (PRI)	0.112	0.174	0.174 (RFR)
AnthM	0.366 (612 nm)	0.329 (PSRI)	0.484	0.521	0.522 (RFR)
FlvM	0.183 (937 nm)	0.110 (CARI)	0.473	0.576	0.670 (SVR)
psi_midday	0.030 (980 nm)	0.015 (NDRE)	0.084	0.206	0.269 (GPR)
Moyenne	0.138	0.108	0.283	0.357	0.379

Tableau 9 : Meilleur R² par trait fonctionnel selon la méthode

Nous observons que les corrélations simples (bande ou indice) permettent d'expliquer qu'une faible part de la variance ($R^2 < 0.15$ en moyenne). À l'inverse, les méthodes de régression atteignent des R^2 plus élevés, avec un gain important pour les modèles mixtes (avec l'espèce en effet aléatoire) qui permettent d'expliquer jusqu'à 57 % de la variance pour FlvM.

La comparaison directe entre les modèles d'apprentissage automatique et les modèles de régression doit être nuancée car ils ne sont pas évalués dans les mêmes conditions : les modèles de régression en vue de caractériser les traits sont appliqués sur l'ensemble des données, tandis que les modèles de prédiction, dans le but de prédire, sont évalués sur un jeu de test indépendant, non vu à l'entraînement (dans une approche 80/20 classique).

Néanmoins, nous pouvons calculer un R^2 sur les prédictions issues de l'apprentissage automatique afin de quantifier leur pouvoir explicatif. Nous observons que pour certains traits, ces derniers atteignent des R^2 légèrement supérieurs, ce qui souligne la pertinence du *machine learning (ML)* malgré une interprétabilité limitée.

Meilleur nRMSE (%) par trait fonctionnel selon la méthode

Trait	Régr. lin. classique (nRMSE [%])	Régr. lin. mixte (nRMSE [%])	Modèle ML (nRMSE [%])
ChlM	24.1	22.3	23.6
gsw	37.3	36.8	29.2
ETR	18.1	17.2	16.9
PhiPS2	10.2	10.1	12.3
AnthM	33.0	32.3	33.2
FlvM	38.9	34.2	30.1
psi_midday	34.2	31.8	29.6
Moyenne	28.0	26.4	25.0

Tableau 10 : Meilleur nRMSE (%) par trait fonctionnel selon la méthode

L'indicateur nRMSE permet de s'intéresser directement à la précision des modèles, contrairement au R^2 qui s'intéresse plutôt à la part de variance expliquée. Pour rappel, les modèles d'apprentissage automatique sont évalués sur un jeu de test indépendant, ce qui constitue un cadre d'évaluation plus exigeant et moins favorable.

Malgré cela, l'erreur moyenne de prédiction est légèrement plus faible, ce qui confirme l'intérêt du *ML* dans ce contexte.

8.1.3 Approche 3 - Corrélations et modélisation de traits fonctionnels archivés

Malgré l'écart temporel, le calcul des corrélations entre les traits et les indices spectraux / bandes spectrales (7.3.1) montre des résultats modérés avec certaines corrélations atteignant $r = 0.6$. Toutefois, la qualité des corrélations reste variable selon les traits considérés.

La modélisation linéaire multivariée (7.3.2) permet d'améliorer les résultats avec des scores de R^2 atteignant 0.41 pour le SLA. Les performances sont, logiquement, inférieures à celles de l'approche 2 mais ils suggèrent la présence résiduelle d'une information fonctionnelle dans les spectres.

8.2 Limites du projet

Le principal frein rencontré dans ce travail concerne la taille limitée du jeu de données et la faible variabilité des traits fonctionnels mesurés. Ces limitations restreignent la capacité de généralisation des modèles d'apprentissage automatique. Bien qu'ils aient montré des résultats encourageants, l'idéal serait de disposer d'au moins 1000 échantillons. Plus de volume de données rendrait les modèles plus robustes et performants.

L'équipement présente également une contrainte pour aller plus loin dans le déploiement. Nous avons constaté que le capteur Raspberry Pi NoIR (sans filtre infrarouge) de la caméra hyperspectrale low-cost utilisée, bien que fonctionnelle, présente une faible sensibilité et nécessite des temps d'exposition longs (quelques secondes). Cette limitation pourrait poser des problèmes lors d'acquisitions dans des environnements dynamiques, comme l'embarcation sur drone. Toutefois, ces limitations n'ont pas compromis la qualité des mesures dans le cadre de ce projet, réalisé en environnement contrôlé.

8.3 Perspectives

Plusieurs axes d'amélioration peuvent être envisagés. D'une part, l'enrichissement du jeu de données en nombre et en diversité en réalisant des mesures dans des contextes environnementaux contrastés, y compris en conditions de stress hydrique afin d'obtenir plus de variabilité dans les traits observés et de mieux évaluer la capacité de généralisation des modèles.

Compte tenu de l'influence de l'espèce dans la relation spectre-trait, une idée pertinente consisterait à adopter une architecture hiérarchique pour la prédiction des traits fonctionnels (illustré à la Figure 8.3-1).

Dans cette approche, un modèle de classification initial permettrait d'identifier automatiquement l'espèce à partir du spectre hyperspectral (facilement réalisable considérant les résultats obtenus dans l'approche 1 (7.1.4)). En fonction de l'espèce prédictive, le système dirigerait alors l'échantillon vers un modèle de régression spécifique, entraîné uniquement sur des données issues de cette même espèce. Ce modèle devrait mieux performer qu'un modèle généraliste, car il n'a pas à gérer la variabilité inter-espèces qui peut causer de la confusion.

Architecture de prédiction des traits fonctionnels

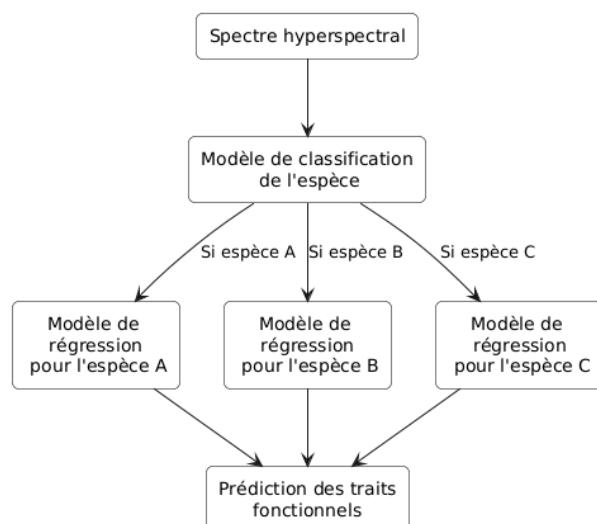


Figure 8.3-1 : Schéma conceptuel d'une architecture hiérarchique pour la prédiction des traits fonctionnels à partir de spectres hyperspectraux

Sur le plan méthodologique, l'utilisation de modèles plus complexes issus du *deep learning*, tels que les réseaux de neurones, représenterait une avancée intéressante. Plusieurs travaux récents [7], [27], [33], [34] dans le domaine de l'analyse hyperspectrale montrent que les approches de *deep learning* permettent d'améliorer significativement les performances prédictives en surpassant à la fois les méthodes statistiques classiques et les algorithmes d'apprentissage automatique traditionnels.

Un modèle du type réseaux de neurones convolutifs (CNN), couramment utilisée pour traiter des images, permettrait de tirer parti de l'hypercube hyperspectral dans son ensemble, sans passer par un spectre représentatif moyen extrait de l'hypercube. Toutefois, ce type de modèle est complexe à mettre en œuvre car il nécessite un volume de données conséquent (plusieurs dizaines de milliers d'échantillons).

Sur le plan matériel, l'utilisation d'une caméra hyperspectrale professionnelle pourrait améliorer les acquisitions hyperspectrales avec une meilleure résolution spatiale et spectrale et une plage spectrale au-delà de 1050 nm. Les caméras professionnelles offrent aussi des temps d'exposition plus rapide (de l'ordre de quelques dizaines de millisecondes), ce qui serait plus adapté à une utilisation embarquée sur un drone. Ce type de matériel représente un coût conséquent (plusieurs dizaines de millier de CHF) mais il permettrait des acquisitions plus fiables, surtout en extérieur.

Enfin, un prolongement naturel de ce travail consisterait à tester le dispositif dans des conditions d'éclairage non contrôlées, notamment en extérieur. Cela permettrait d'évaluer la robustesse des acquisitions hyperspectrales face aux variations de lumière.

9 Conclusion

Ce travail de Bachelor avait pour objectif d'évaluer le potentiel de l'imagerie hyperspectrale low-cost dans l'analyse des traits fonctionnels, à travers une application concrète sur des feuilles d'érable en environnement contrôlé. Pour cela, un système d'acquisition basé sur un Raspberry Pi a été mis en œuvre avec succès. La caméra hyperspectrale a permis de réaliser près de 500 acquisitions exploitables, démontrant la faisabilité d'un tel dispositif à faible coût.

Les données ont été traitées de manière rigoureuse puis structurées dans une base de données relationnelle, facilitant leur analyse et leur réutilisation. L'analyse statistique et les méthodes d'apprentissage automatique ont mis en évidence une variabilité spectrale significative entre espèces. Elles ont également montré des résultats prometteurs, révélant une capacité réelle à relier certains traits fonctionnels de feuilles végétales à leur signature spectrale.

Ce projet montre ainsi que l'imagerie hyperspectrale low-cost est capable de capturer des signatures biologiques exploitables, constituant une alternative aux systèmes professionnels, bien plus coûteux. Malgré les limites techniques de l'équipement et un jeu de données encore restreint, les résultats sont encourageants.

Cette mobilité au sein d'une équipe de recherche, dans un domaine différent de ma formation initiale, a exigé une grande autonomie et une adaptation rapide aux méthodes propres aux sciences environnementales, offrant ainsi une expérience particulièrement enrichissante.

Ce travail constitue une base expérimentale et méthodologique réutilisable pour des applications futures sur le terrain ou embarquées sur drone, ainsi que pour des méthodes d'analyse plus avancées.

10 Références

- [1] « THE 17 GOALS | Sustainable Development ». Consulté le: 18 juillet 2025. [En ligne]. Disponible sur: <https://sdgs.un.org/goals>
- [2] P. Mishra, M. S. M. Asaari, A. Herrero-Langreo, S. Lohumi, B. Diezma, et P. Scheunders, « Close range hyperspectral imaging of plants: A review », *Biosyst. Eng.*, vol. 164, p. 49-67, déc. 2017, doi: 10.1016/j.biosystemseng.2017.09.009.
- [3] M. H. Sellami *et al.*, « Selection of Hyperspectral Vegetation Indices for Monitoring Yield and Physiological Response in Sweet Maize under Different Water and Nitrogen Availability », *Agronomy*, vol. 12, n° 2, Art. n° 2, févr. 2022, doi: 10.3390/agronomy12020489.
- [4] P. S. Thenkabail, R. B. Smith, et E. De Pauw, « Hyperspectral Vegetation Indices and Their Relationships with Agricultural Crop Characteristics », *Remote Sens. Environ.*, vol. 71, n° 2, p. 158-182, févr. 2000, doi: 10.1016/S0034-4257(99)00067-X.
- [5] J. Wu *et al.*, « Leaf reflectance spectroscopy captures variation in carboxylation capacity across species, canopy environment and leaf age in lowland moist tropical forests », *New Phytol.*, vol. 224, n° 2, p. 663-674, 2019, doi: 10.1111/nph.16029.
- [6] L. Wan et F. Ma, « Estimating leaf photosynthetic capacity using hyperspectral reflectance: Model variability and transferability », *Comput. Electron. Agric.*, vol. 220, p. 108837, mai 2024, doi: 10.1016/j.compag.2024.108837.
- [7] M. F. Guerri, C. Distante, P. Spagnolo, F. Bougourzi, et A. Taleb-Ahmed, « Deep learning techniques for hyperspectral image analysis in agriculture: A review », *ISPRS Open J. Photogramm. Remote Sens.*, vol. 12, p. 100062, avr. 2024, doi: 10.1016/j.jophoto.2024.100062.
- [8] J. Salazar-Vazquez et A. Mendez-Vazquez, « A plug-and-play Hyperspectral Imaging Sensor using low-cost equipment », *HardwareX*, vol. 7, p. e00087, avr. 2020, doi: 10.1016/j.ohx.2019.e00087.
- [9] E. M. Pechlivani, A. Papadimitriou, S. Pemas, N. Giakoumoglou, et D. Tzovaras, « Low-Cost Hyperspectral Imaging Device for Portable Remote Sensing », *Instruments*, vol. 7, n° 4, Art. n° 4, déc. 2023, doi: 10.3390/instruments7040032.
- [10] R. Habel, M. Kudenov, et M. Wimmer, « Practical Spectral Photography », *Comput. Graph. Forum*, vol. 31, n° 2pt2, p. 449-458, mai 2012, doi: 10.1111/j.1467-8659.2012.03024.x.
- [11] G. ElMasry, N. Wang, A. ElSayed, et M. Ngadi, « Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry », *J. Food Eng.*, vol. 81, n° 1, p. 98-107, juill. 2007, doi: 10.1016/j.jfoodeng.2006.10.016.
- [12] « SQLAlchemy ». Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: <https://www.sqlalchemy.org>
- [13] « PCA », scikit-learn. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: <https://scikit-learn/stable/modules/generated/sklearn.decomposition.PCA.html>
- [14] « How to Perform a One-Way ANOVA in Python », GeeksforGeeks. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: <https://www.geeksforgeeks.org/python/how-to-perform-a-one-way-anova-in-python/>
- [15] « KNeighborsClassifier », scikit-learn. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: <https://scikit-learn/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>
- [16] « RandomForestClassifier », scikit-learn. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: <https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- [17] « SVC », scikit-learn. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: <https://scikit-learn/stable/modules/generated/sklearn.svm.SVC.html>

- [18] « RobustScaler », scikit-learn. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>
- [19] « 3.1. Cross-validation: evaluating estimator performance », scikit-learn. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: https://scikit-learn/stable/modules/cross_validation.html
- [20] « sklearn.metrics », scikit-learn. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: <https://scikit-learn/stable/api/sklearn.metrics.html>
- [21] « pearsonr — SciPy v1.16.0 Manual ». Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mstats.pearsonr.html>
- [22] « PLSRegression », scikit-learn. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: https://scikit-learn/stable/modules/generated/sklearn.cross_decomposition.PLSRegression.html
- [23] « statsmodels.regression.linear_model.OLS ». statsmodels.org. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: https://www.statsmodels.org/dev/generated/statsmodels.regression.linear_model.OLS.html
- [24] « Linear Mixed Effects Models - statsmodels ». statsmodels.org. Consulté le: 21 juillet 2025. [En ligne]. Disponible sur: https://www.statsmodels.org/dev/examples/notebooks/generated/mixed_lm_example.html
- [25] « r2_score », scikit-learn. Consulté le: 22 juillet 2025. [En ligne]. Disponible sur: https://scikit-learn/stable/modules/generated/sklearn.metrics.r2_score.html
- [26] J. Frost, « Root Mean Square Error (RMSE) », Statistics By Jim. Consulté le: 22 juillet 2025. [En ligne]. Disponible sur: <https://statisticsbyjim.com/regression/root-mean-square-error-rmse/>
- [27] P. D. Dao, Y. He, et C. Proctor, « Plant drought impact detection using ultra-high spatial resolution hyperspectral images and machine learning », *Int. J. Appl. Earth Obs. Geoinformation*, vol. 102, p. 102364, oct. 2021, doi: 10.1016/j.jag.2021.102364.
- [28] « RandomForestRegressor », scikit-learn. Consulté le: 22 juillet 2025. [En ligne]. Disponible sur: <https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>
- [29] « SVR », scikit-learn. Consulté le: 22 juillet 2025. [En ligne]. Disponible sur: <https://scikit-learn/stable/modules/generated/sklearn.svm.SVR.html>
- [30] « GaussianProcessRegressor », scikit-learn. Consulté le: 22 juillet 2025. [En ligne]. Disponible sur: https://scikit-learn/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html
- [31] « OneHotEncoder », scikit-learn. Consulté le: 22 juillet 2025. [En ligne]. Disponible sur: <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>
- [32] « Niveau de signification Determination de la pertinence statistique de l ANOVA », FasterCapital. Consulté le: 22 juillet 2025. [En ligne]. Disponible sur: <https://fastercapital.com/fr/contenu/Niveau-de-signification---Determination-de-la-pertinence-statistique-de-l-ANOVA.html>
- [33] S. Yu *et al.*, « Deep learning models based on hyperspectral data and time-series phenotypes for predicting quality attributes in lettuces under water stress », *Comput. Electron. Agric.*, vol. 211, p. 108034, août 2023, doi: 10.1016/j.compag.2023.108034.
- [34] E. Cherif *et al.*, « From spectra to plant functional traits: Transferable multi-trait models from heterogeneous and sparse data », *Remote Sens. Environ.*, vol. 292, p. 113580, juill. 2023, doi: 10.1016/j.rse.2023.113580.

OpenAI. ChatGPT [Outil d'intelligence artificielle]. Utilisé ponctuellement pour des suggestions de formulation, de relecture, et pour assister à la rédaction de portions de code dans le cadre de ce travail. Disponible à l'adresse : <https://chat.openai.com/>

11 Annexes

Le Tableau 11 regroupe les éléments complémentaires et nécessaires à la reproductibilité du projet. Les annexes mentionnées ci-dessous sont fournies sous la forme d'un dossier intitulé *Annexes/*, accompagné du présent rapport.

Contenu du dossier *Annexes/*

Dossier	Description
Application_SquareHSI/	Contient l'application client SquareHSI (format <i>.AppImage</i> , version PC), utilisée pour l'acquisition hyperspectrale. Logiciel open-source développé par Salazar-Vazquez et al. [8].
Code_HypRaspCam/	Code source du serveur d'acquisition tournant sur le Raspberry Pi. Code open-source développé par Salazar-Vazquez et al. [8].
STL_Camera/	Modèles 3D (<i>.stl</i>) des pièces de la caméra hyperspectrale. Ces fichiers proviennent du projet de Salazar-Vazquez [8], à l'exception d'une pièce modifiée (adaptation du port d'alimentation).
Calibration/	Fichier <i>.xml</i> généré par SquareHSI après avoir effectué la calibration de la caméra, nécessaire à la construction de l'hypercube.
Reglages/	Fichier de réglage de la caméra au format <i>.xml</i> contenant l'ensemble des paramètres d'acquisition (AWB, temps d'exposition, etc.), utilisés de manière constante pour toutes les observations durant le projet.
Database/	Base de données relationnelle au format <i>.db (SQLite)</i> , contenant toutes les mesures acquises durant le projet. Structure décrite dans le chapitre Base de données. Contient environ 500 acquisitions.
Code_Analyse/	Code développé durant ce projet : scripts Python (/src) et notebooks Jupyter (/notebooks) utilisés pour le prétraitement (normalisation, calcul des indices, etc...), les analyses statistiques et l'apprentissage automatique.

Tableau 11 : Contenu du dossier des Annexes