Faculté des bioingénieurs

Multispectral and Hyperspectral Remote Sensing for Winter Cover Crop Identification, Mixture and Species Characterisation in Wallonia, Belgium

Acteur: Hannah Firl
Promoteurs: Dr. Pierre Defourny
Louis Descarpentries
Lecteurs: Dr. Sophie Bontemps
Dr. Xavier Draye

Année académique 2025-2026
Mémoire de fin d'études présenté en vue de l'obtention du diplôme de Master : SAIV - MSc Erasmus+ GEM : Geo-Information Science and Earth Observation for Environmental Modelling and Management

Acknowledgements

First and foremost, I would like to especially thank Dr. Pierre Defourney and Louis Descarpentries for their invaluable guidance, support and constructive feedback throughout the development of this thesis at UCLouvain in Belgium. I am also grateful to the PhD students of the Earth and Life department for their help and for creating a supportive and motivating research environment. I also extend my gratitude to Dr. Sophie Bontemps and Dr. Xavier Draye for dedicating their time and effort to review and evaluate this work. Furthermore, I would like to thank Lund University for all that I learned during the first year of my Master's programme, which provided an essential foundation for this thesis. Finally, I would like to thank my family and friends for their constant encouragement, patience and understanding, which have been essential to the completion of this work.

Contributions to this work

Stage	Main contributor	Other contributor(s)	Artificial Intelligence
Conceptualisation	Hannah Firl	Main research idea, scientific guidance and feedback, provision of academic papers: Pierre Defourny, Louis Descarpentries	Bibliographic research: Perplexity & Research Rabbit
Methodology	Hannah Firl	Methodological advice: Pierre Defourny, Louis Descarpentries, Juan Garcia Peña Pre-processing of the Sentinel-2 tile: Louis Descarpentries	Generation and correction of code: Perplexity and GitHub Copilot
Investigation	Conduction of research process and part of data collection: Hannah Firl	Data collection: Louis Descarpentries	
Formal analysis	Hannah Firl	Discussion and interpretation support: Louis Descarpentries	
Validation	Hannah Firl	Pierre Defourny, Louis Descarpentries	Generation and correction of code: Perplexity and GitHub Copilot
Writing – original draft	Hannah Firl		
Writing – review & editing	Hannah Firl	Revision, structuring, and correction of intellectual content: Pierre Defourny, Louis Descarpentries	Writing and formulation: Perplexity, DeepL Write
Visualisation	Hannah Firl	Pierre Defourny, Louis Descarpentries	
Supervision		Pierre Defourny, Louis Descarpentries	

The work presented in this thesis was primarily carried out by Hannah Firl, the student, who was responsible for the overall conceptualisation, methodological design, investigation, formal analysis, validation, visualisation and writing of the manuscript. The supervisors, Pierre Defourny and Louis Descarpentries, provided the main research idea, scientific guidance, methodological advice, discussion and interpretation of results, and critical revision of the intellectual content, with additional methodological input from Juan Garcia Peña. Louis Descarpentries also made a substantial contribution to the collection of in-situ data. AI-based tools (Perplexity, Research Rabbit, GitHub Copilot and DeepL Write) were employed to support literature searches, code generation and correction, and language editing. All scientific decisions, methodological choices and interpretations remain the responsibility of the student and the other named contributors.

Table of contents

List of figures

List of tables

List of acronyms

ANN	Artificial Neural Networks
AOI	Area of Interest
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
BCAE	Bonnes Conditions Agricoles et Environnementales
BELCAM	BELgian Collaborative Agriculture Monitoring
BOA	Bottom-of-Atmosphere
BSI	Bare Soil Index
BV	Biophysical Variable
CAP	Common Agricultural Policy
CC	Cover Crop
CIPAN	Catch Crop for Nitrate Trapping
CNN	Convolutional Neural Network
CSD-SI	Cloud/Shadow Detection based on Spectral Indices
DL	Deep Learning
EM	Electromagnetic
EnMAP	Environmental Mapping and Analysis Program
EO	Earth Observation
EVI	Enhanced Vegetation Index
fAPAR	Fraction of Absorbed Photosynthetically Active Radiation
fCover	Fraction of Green Vegetation Cover
GEE	Google Earth Engine
HLS	Harmonised Landsat-8 and Sentinel-2
HS	Hyperspectral
HSI	Hyperspectral Imagery
HVI	Hyperspectral Vegetation Index
LAI	Leaf Area Index
LPIS	Land Parcel Identification System
MDA	Mean Decrease in Accuracy
MDG	Mean Decrease in Gini
ML	Machine Learning
MODIS	Moderate Resolution Imaging Spectroradiometer
MS	Multispectral
MSI	Multispectral Imagery
NDRE	Normalized Difference Red Edge Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near-Infrared
NVZ	Nitrate Vulnerable Zones
NWCC	Non-Winter Cover Crops
OA	Overall Accuracy
OOB	Out-of-Bag
P	Class Probability
PCA	Principal Component Analysis
PGDA	Programme de Gestion Durable de l'Azote
PGML	Process-Guided Machine Learning
RE	Red-Edge
RF	Random Forest
RMSE	Root Mean Square Error
RS	Remote sensing
S1	Sentinel-1
S2	Sentinel-2
SAR	Synthetic Aperature Radar
SAVI	Soil-Adjusted Vegetation Index
SIGEC	Système Intégré de Gestion et de Contrôle

SNR .... Signal-to-Noise Ratio
SRF .... Spectral Response Function
SVM .... Support Vector Machine
SWIR .... Shortwave Infrared
TOA .... Top-of-Atmosphere
UAV .... Unmanned Aerial Vehicle
VI .... Vegetation Index
VIS .... Visible
VNIR .... Visible-Near-Infrared
WCC .... Winter Cover Crop
κ .... Cohen's Kappa

INTRODUCTION

Timely and spatially explicit information on agricultural practices, such as crop rotation, tillage, fertiliser application and winter cover crops (WCCs) is essential for sustaining food production while reducing environmental impact (Ghazaryan et al., 2018; Alami Machichi et al., 2023). WCCs are one such management option that has received increasing attention, as they protect soils during the fallow period in autumn/winter, reduce nitrogen leaching, contribute to carbon sequestration, and enhance the overall resilience of the agro-ecosystem (e.g. by controlling erosion, suppressing weeds, and improving water infiltration). WCCs are therefore increasingly recognised as a cornerstone of regenerative agriculture, an approach that maintains permanent soil cover to restore soil health, enhance biodiversity and sequester carbon (Khangura et al., 2023). However, the benefits of WCCs depend strongly on the species composition, since legumes, grasses and brassicas differ in traits such as nitrogen fixation and biomass production. Accurately identifying and systematically mapping WCCs on a regional scale is therefore crucial for quantifying their environmental benefits, and for supporting agricultural producers and policymakers in designing and evaluating agro-environmental schemes. Nevertheless, institutional data on the spatial distribution and species composition of WCCs at a regional scale remain scarce, which limits the evaluation of their realised agri-environmental benefits. Remote sensing (RS) is a powerful means of consistently monitoring such practices across large areas, offering a scalable alternative to costly, labour-intensive field surveys with limited temporal coverage.

RS monitoring has already proven to be an effective method for WCCs applications, particularly when multispectral (MS) sensors, such as Sentinel-2 (S2) and Landsat, are combined with vegetation indices (VIs) and multi-temporal classification approaches. Current research has focused primarily on the binary detection of the presence or absence of WCCs, or the quantification of biomass, often using Normalized Difference Vegetation Index (NDVI) time series to characterise establishment and growth. In contrast, differentiation between mono- and multi-species mixtures, and WCC species identification or functional groups, have received little attention.

From an RS perspective, the limited species-level information obtained so far can partly be attributed to the spectral constraints of MS imagery (MSI). The broad spectral bands and strong correlations between reflectance in neighbouring wavelengths reduce biochemical sensitivity, which makes identifying species with similar phenology or canopy structure challenging, particularly in heterogeneous mixtures. Hyperspectral (HS) imaging (HSI), as provided by the Environmental Mapping and Analysis Program (EnMAP), samples the visible (VIS) to shortwave infrared (SWIR) spectrum in hundreds of narrow, contiguous bands. This high spectral resolution enables specific absorption features related to pigments, water, nitrogen and structural compounds to be targeted. It has also been demonstrated that this improves the discrimination of crops and species in various ecosystems when used alongside appropriate feature selection and classification methods. However, applications of HSI to WCCs are still rare, being mostly limited to the estimation of biomass and canopy nitrogen from airborne campaigns over relatively small experimental areas rather than the mapping of species and mixtures on a regional scale.

These developments raise two key questions about monitoring WCCs in temperate European cropping systems. Firstly, how can S2 time series support the detection and the species characterization of WCCs? Secondly, does spaceborne HSI, such as that provided by EnMAP, offer any measurable added value over S2 for these two specific applications? If so, in which spectral regions and for which classification tasks? Addressing these questions is important for improving the monitoring of WCC adoption and composition within agri-environmental policies, as well as for evaluating the potential of future HSI missions for agricultural applications.

This thesis investigates the use of S2 MSI and EnMAP HSI for identifying and characterising WCCs in Wallonia, Belgium. Chapter 1 provides a literature review of the agro-environmental functions of WCCs and summarises the key principles of RS including vegetation spectroscopy. It also discusses current MS and emerging HS approaches for WCC detection and mixture characterisation, together with relevant Machine Learning (ML) methods. Chapter 2 sets out the specific objectives: S2 time series classification; a comparison of single-date classifications between EnMAP and S2; and EnMAP single-date classification using multiple bands to classify the presence of mixtures and detect dominant species. Chapter 3 describes the study area, the satellite datasets and in-situ data used, the pre-processing and feature construction methods, and the Random Forest (RF) modelling framework. Chapter 4 presents classification results for both sensors, including mean spectra and feature importance analyses. Chapter 5 discusses the findings in the context of existing literature and in situ data, as well as the trade-offs between sensors and the implications for the operational monitoring of WCCs. Finally, Chapter 6 summarises the main conclusions and outlines how MS and HS information could be integrated into future agricultural and environmental observation systems.

CHAPTER
ONE

LITERATURE REVIEW

This chapter reviews the current knowledge of WCCs and their monitoring using RS, in order to contextualise the present study within the broader research and policy landscape. First, it summarises how different WCC species and mixtures contribute to agri-environmental functions, and explains why reliable information on their spatial distribution is increasingly necessary under the Common Agricultural Policy (CAP) and related schemes. It then discusses approaches using MSI and HSI approaches for detecting WCC presence, estimating biomass, and, more recently, distinguishing between functional types and species compositions. The chapter highlights the progress that has been made in this area, as well as the remaining gaps that motivate the research questions addressed in the following chapters.

1.1. Winter cover crops specific composition influence their agro-environmental functions

Cover Crops (CCs) are close-growing plants that are established between the harvest of a main crop and the next sowing, also called the intercropping or fallow period. Establishment during this period, often characterised by bare soil, maintains continuous soil cover and enhances the physical, chemical and biological properties of the soil (Figure 1) (Sharma et al., 2018; Chapagain et al., 2020).

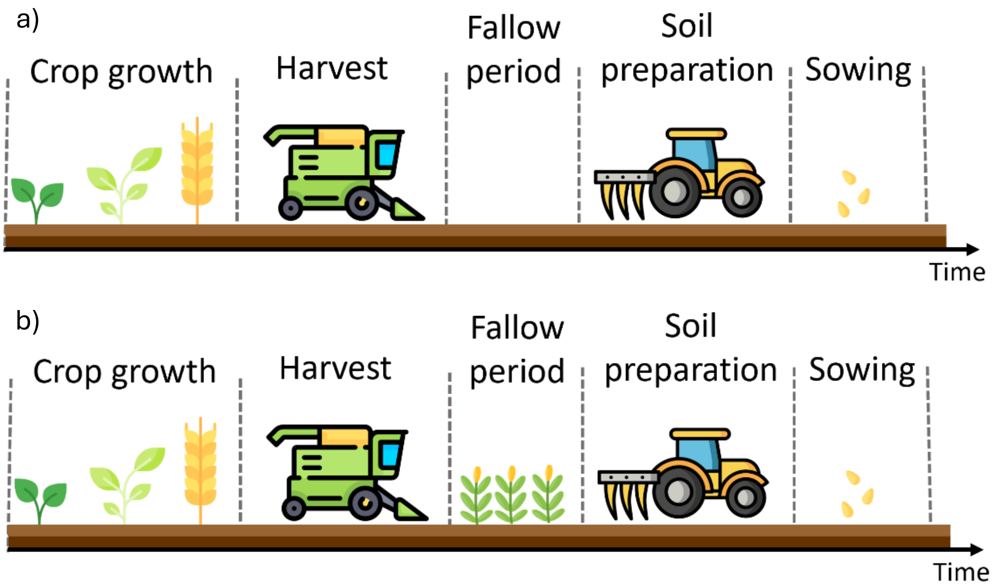


Figure 1: Schematic timeline of an agricultural cropping cycle. **a)** A conventional cycle with a bare fallow period following the main crop harvest. **b)** An adapted cycle where WCCs are established during the fallow period to maintain continuous soil cover before the next sowing.

CC establishment time, whether short or long-term, depends both on the crop rotation system and on the regional or national legal requirements (Kc et al., 2021). Management and termination strategies also vary according to species traits and climatic conditions. CCs can include a wide range of species spanning three common functional groups, including legumes, non-legume grasses and brassicas. Examples include red clover (*Trifolium pratense* L.), oats (*Avena sativa* L.), and cereal rye (*Secale cereale* L.) (Adetunji et al., 2020; Kc et al., 2021; Jennewein et al., 2024). Some species are winter-hardy and resume growth in spring, whereas others are killed by frost and low temperatures in winter (winter-kill). CCs can be terminated mechanically (e.g. by mowing or rolling) or chemically using herbicides (Chapagain et al., 2020; Kc et al., 2021). CCs are planted at different times of the year, leading to distinctions such as winter, spring, summer and autumn CCs. WCCs are typically sown in late summer after a winter crop, or in late autumn after spring or summer crops (e.g. beets, potato, maize) have been harvested. They are either destroyed before the end of the calendar year (e.g. catch crops) or they continue to grow through winter, showing lush green growth by the following spring (Do Nascimento Bendini et al., 2024). This study focuses on WCCs.

WCCs play a multifunctional role in providing a wide range of agro-environmental benefits that enhance both soil quality and overall ecosystem functioning. They increase the levels of soil organic carbon and soil organic nitrogen, enhancing aggregate stability, and supporting microbial biodiversity. By providing continuous soil cover, they reduce erosion, conserve soil moisture, and improve water infiltration and availability for subsequent crops (Sharma et al., 2018; Adetunji et al., 2020). WCCs also mitigate weeds and pests, enhance nutrient and water use efficiency, and mitigate greenhouse gas emissions through carbon sequestration (Adetunji et al., 2020; Chapagain et al., 2020). Notably, WCCs also reduce nitrogen leaching by absorbing residual soil nitrate after the main crop harvest and temporarily storing it in their biomass during the non-growing season. This nitrogen is then gradually released back into the soil as the WCCs decompose, thereby improving nitrogen retention and synchronising nutrient availability with subsequent crop demand (Couëdel et al., 2018; Fan et al., 2020).

While individual WCC species offer many benefits, mixtures are increasingly used to combine complementary functions, such as nitrogen fixation by legumes and weed suppression or erosion control by grasses. Diverse species compositions in WCC mixtures enhance resource-use complementarity, promote soil carbon accumulation, stimulate microbial abundance and diversity, and are therefore crucial for maximising agro-ecological functions and supporting the sustainability and resilience of cropping systems (Adetunji et al., 2020; Chapagain et al., 2020; Fiorini et al., 2022; Kharel et al., 2023).

Despite these advantages, adoption remains limited in many areas due to higher management costs, labour requirements, and uncertainties regarding species selection, seeding rates, and termination timing (Schmidt et al., 2018; Sharma et al., 2018; Chapagain et al., 2020). At the same time, renewed agronomic and policy interest under CAP and related regional schemes is gradually expanding the use of WCCs in Europe. In several European regions, including parts of Belgium, increased incentive payments and agri-environmental measures have contributed to a growing uptake of WCCs, with adoption reaching early- to late-majority levels in some contexts (Figure 2) (European Commission. Joint Research Centre., 2019; Kharel et al., 2023; Heller et al., 2024). This ongoing, albeit spatially heterogeneous, expansion means that robust information on the acreage under different WCC species and mixtures is required to quantify realised benefits and support the targeting and evaluation of incentive schemes. Traditional ground-based surveys are time-consuming, logistically challenging, and often lack the spatial and temporal resolution needed for regional monitoring (Kc et al., 2021). However, RS allows for a scalable and cost-effective way to monitor WCC distribution and species composition across landscapes. Nevertheless, WCCs have not yet been extensively analysed in RS research (M. L. Barnes et al., 2021; Kc et al., 2021; Vanpoucke et al., 2024).

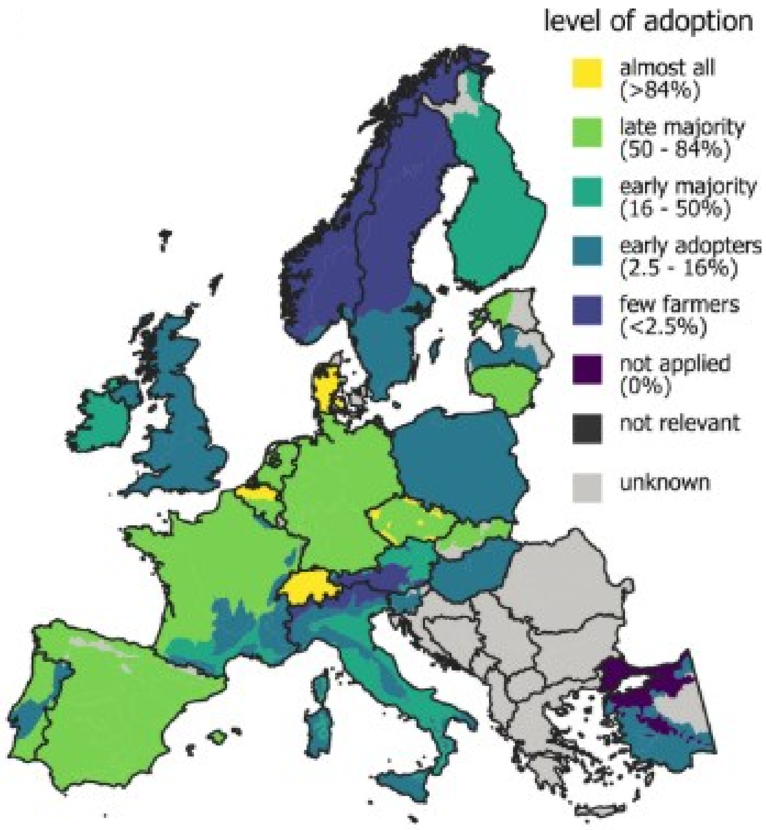


Figure 2: Level of adoption of CCs in Europe (Heller et al., 2024).

1.2. Remote sensing techniques applied to winter cover crops detection and characterization

1.2.1. Fundamentals of remote sensing and vegetation spectroscopy

RS can be defined as the art and science of obtaining information about an object without being in direct physical contact with it (Lillesand et al., 2015). This study focuses on optical RS, which uses reflected and emitted electromagnetic (EM) radiation in the VIS, near-infrared (NIR) and SWIR regions to image the Earth's surface. The physical basis of RS lies in the behaviour of EM radiation, which transfers energy in the form of waves travelling at the speed of light (De Lange, 2020). The wavelength and frequency of each EM wave determine the radiation's physical properties (Lillesand et al., 2015). The full range of these wavelengths constitutes the EM spectrum.

RS exploits physical radiation processes in the atmosphere to derive information about objects on the Earth's surface. Under daylight conditions, solar radiation acts as the main source of illumination and predominantly emits energy within the shortwave region of the spectrum (De Lange, 2020). When radiation interacts with surface materials, it can be reflected, transmitted, or absorbed, depending on the object's physical and chemical properties as well as its state (e.g. surface temperature or vegetation growth stage) (Albertz, 2009). The reflected component leaves the target as spectrally modified radiance, which depends on these intrinsic properties, illumination, viewing geometry, and surface roughness. Ultimately, the intensity and spectral characteristics of the radiation received by the sensor therefore result from the combined effects of atmospheric scattering and absorption, surface interactions, and the solar elevation angle and atmospheric conditions along the path (Lillesand et al., 2015; Brosinsky et al., 2022).

The graph of a function that represents the proportion of reflected or emitted energy across different wavelengths is referred to as a spectral signature. These signatures arise from wavelength-dependent absorption, transmission and reflection processes. These processes depend on the target's intrinsic physical, biochemical and structural properties, and therefore provide comprehensive insights into factors such as vegetation condition, water content and mineral composition. By comparing spectral signatures of materials such as green vegetation, senescent vegetation and bare soil (Figure 3), spectroscopy allows specific absorption features to be analysed and used for the qualitative and quantitative characterisation of different surfaces (Kumar et al., 2001; Albertz, 2009; Kaufmann et al., 2012).

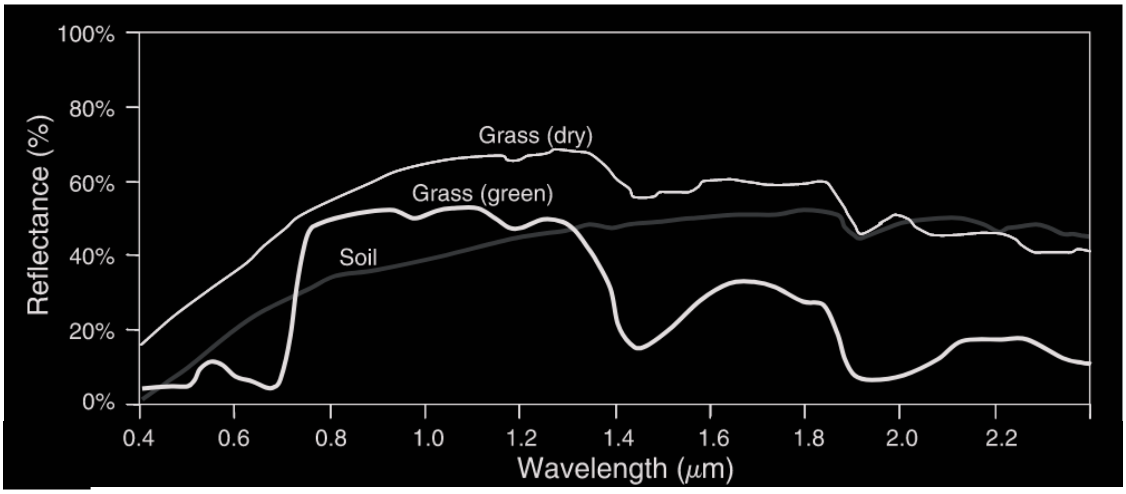


Figure 3: Spectral reflectance curves for various feature types (green grass, dry grass and soil) (Lillesand et al. 2015).

1.2.2. Spectral behaviour of vegetation and implications for species discrimination

Spectral reflectance is primarily used to identify different crops in optical satellite imagery because their leaves and canopies vary in pigment composition, internal structure, water content and biochemistry (Tan et al., 2025). Regarding vegetation Chlorophyll a and b strongly absorb energy in the VIS (400-700 nm), particularly around 450 nm and 670 nm, which corresponds to the blue and red spectral domains. Conversely, vegetation reflects more energy in the green region, commonly referred to as the ‘chlorophyll reflectance maximum’, which gives healthy plants their green appearance to the human eye (Lillesand et al., 2015). Changes in chlorophyll concentration cause a steep increase in reflectance at the transition from the VIS to the NIR region, known as the red edge (RE), typically located between approximately 680 and 750 nm (van der Meer, 2001; Lillesand et al., 2015). In the NIR region between 750 and 1,300 nm, vegetation reflects approximately 40–50% of incoming radiation; however, only a small fraction (typically less than 5%) is absorbed due to the internal leaf structure (van der Meer, 2001). This reflectance behaviour often characterises species-specific canopy architecture (Lillesand et al., 2015). While red reflectance is highly correlated with chlorophyll concentration, NIR reflectance is mainly controlled by multiple scattering within the canopy and therefore depends on canopy structure, leaf water content and the density of green vegetation, often summarised by the Leaf Area Index (LAI) (Section 1.2.5.) (Bannari et al., 1995). In the SWIR region (1,500–2,500 nm), the strong absorption of radiation by liquid water and biochemical constituents, such as cellulose and lignin, means that most of the radiation entering the canopy is either absorbed or reflected at the surface, with very little transmitted through it (Lillesand et al., 2015). Reflectance decreases sharply at wavelengths around 970, 1,200, 1,450, 1,930 and 2,500 nm because water strongly absorbs radiation at these spectral ranges, known as the water absorption bands (Zhang et al., 2010). Meanwhile, cellulose, lignin and starch exhibit additional absorption features throughout the 1,200-2,500 nm range (van der Meer, 2001).

The spectral signatures of vegetation are strongly influenced by phenological development, i.e. the timing and sequence of growth stages, and therefore vary between crop types, management practices (e.g. sowing dates, seeding rate, row spacing, termination timing) and climates. Because phenology drives changes in chlorophyll concentration, leaf structure and water content, it produces characteristic temporal variations in reflectance, particularly in the RE, NIR and SWIR regions, which are crucial for species classification. When species share similar biophysical traits, spectral separability can be limited, so accounting for phenological and management-induced differences is particularly beneficial for discriminating between them, especially in WCC mixtures with heterogeneous stand structures (Dudley et al., 2015; Gerstmann et al., 2018; Ghazaryan et al., 2018; Galvão et al., 2019; Basinger et al., 2020).

1.2.3. Spectral, spatial, temporal and radiometric resolution in optical remote sensing

The extent to which optical RS can capture the physical, biochemical, and structural properties of a surface depends heavily on the spectral resolution of the sensor. Spectral resolution is defined as the number and width of spectral bands recorded within specific regions of the EM spectrum (Lillesand et al., 2015). Many RS systems are MS, capturing reflectance across a limited number of broad wavelength bands within the EM spectrum, typically 4–13, as in S2 (CDSE, n.d.). HS sensors, by contrast, measure hundreds of contiguous narrow bands (5–10 nm wide) as in EnMAP. This provides detailed spectral information about plant biochemical and structural properties, allowing for the detection of subtle differences between species and improved discrimination of mixed vegetation signals (Figure 4) (Thenkabail et al., 2004).

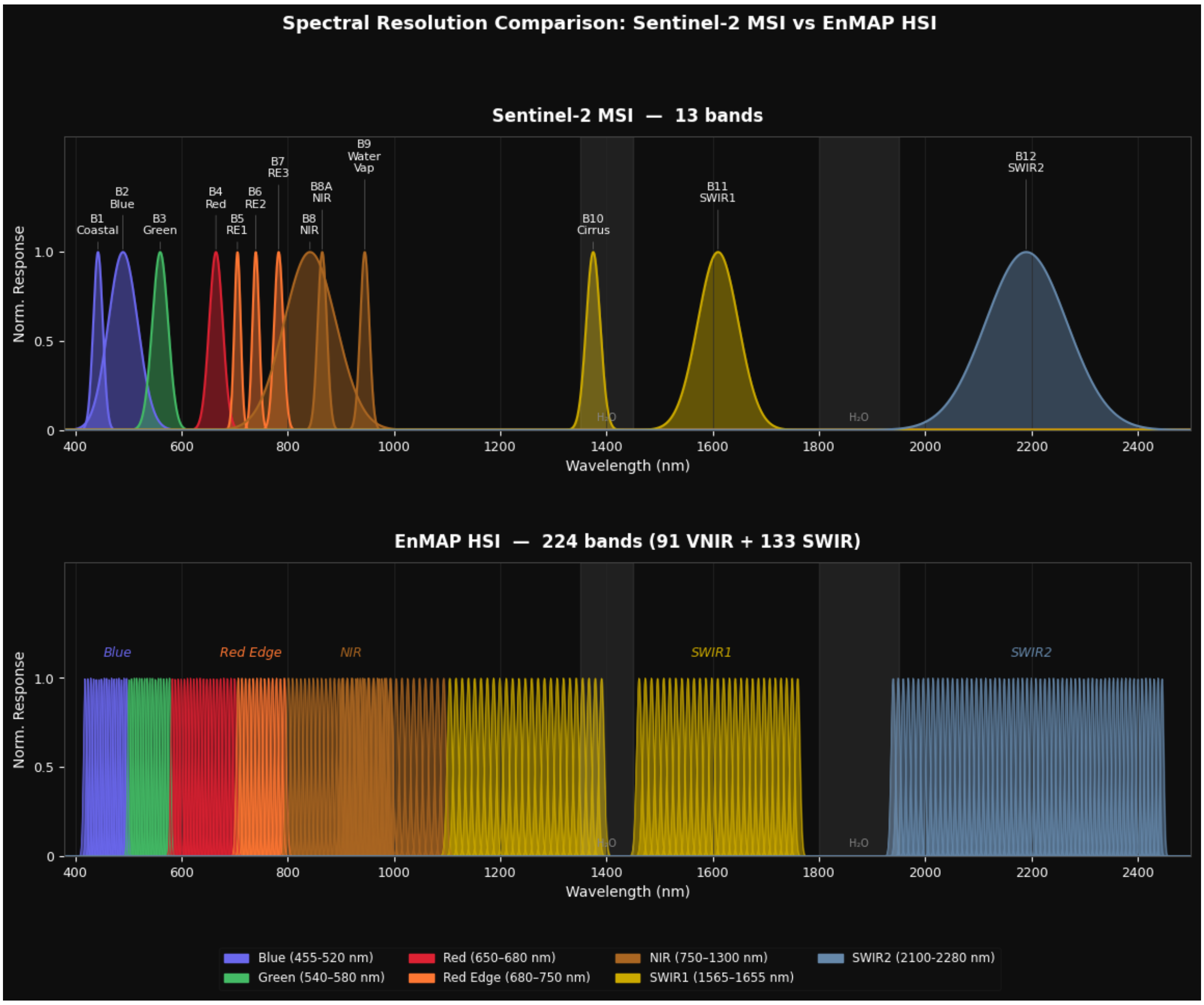


Figure 4: Normalised spectral response function of S2 MSI (13 bands) and EnMAP HSI (224 bands) across the 420-2450 nm range.

RS sensors also differ in terms of their spatial, temporal and radiometric resolutions. These resolutions determine the size of the smallest ground feature that can be detected independently, the revisit frequency of observations, and the ability to distinguish subtle differences in signal intensity (Lillesand et al., 2015). Increasing spectral resolution generally reduces spatial resolution or the signal-to-noise ratio (SNR), so sensor design must strike a balance between these dimensions to achieve an appropriate compromise for the intended application (Brosinsky et al., 2022). Similarly, a compromise exists between spatial and temporal resolution: sensors with smaller pixels typically cover a narrower swath and therefore revisit a given location less frequently. In contrast, missions with short revisit times often use coarser pixels to achieve this. This trade-off can be partially mitigated by flying constellations of identical platforms (Brosinsky et al., 2022; CDSE, n.d.).

1.2.4. Vegetation monitoring with multispectral data

Research in optical RS has shown that crop and species discrimination can be achieved by exploiting differences in spectral reflectance patterns, phenological development and multi-temporal VI trajectories (Li et al., 2014; Gerstmann et al., 2018; Ghazaryan et al., 2018; Misra et al., 2020). Extensive research has been conducted on the mapping of main crop types using MSI (Ghazaryan et al., 2018; Alami Machichi et al., 2023).

Most studies use pixel- or object-based image analysis, applying single- or multi-temporal satellite imagery alongside feature engineering techniques to derive covariates for data-driven models. In recent work, crop mapping has increasingly relied on multi-temporal imagery rather than single-date scenes - as it considers the full phenological profiles allowing crops differentiation based on their growing patterns rather than on spectral contrast at a single point in time (Section 1.2.2.) (Kc et al., 2021; Ahmed et al., 2023; Vanongeval et al., 2026). Classification models (Section 1.2.9.) typically use input features comprising spectral bands, VIs and image textural features derived from grey-level co-occurrence matrices (e.g. variance, contrast, and entropy) (Alami Machichi et al., 2023; Vanongeval et al., 2026). VIs summarise spectral information into single numerical values that reflect key plant properties, such as greenness, pigment content and water status (Roberts et al., 2019). Due to their computational simplicity, relative independence from site-specific factors, and operational implementability, VIs are widely used to derive quantitative estimates of vegetation biophysical parameters from RS data. This provides one of the most practical ways to obtain such variables over large areas with high temporal coverage (Frampton et al., 2013). Examples include the NDVI, which quantifies vegetation greenness and vigour based on the contrast between red and NIR reflectance. Other examples include harmonic modelling of seasonal VI trajectories and rule-based phenological thresholds, which help distinguish crop types even when field data are limited. However, despite its widespread use, the NDVI saturates in dense canopies and is affected by soil background reflectance, limiting its ability to discriminate between species or functional types at higher biomass levels (Lillesand et al., 2015; Ghazaryan et al., 2018; Alami Machichi et al., 2023). RE-based indices, soil-adjusted indices and SWIR-based water indices are increasingly being used to reduce NDVI saturation and soil background effects, capturing variations in chlorophyll and water content that are important for crop monitoring (Huete, 1988; Gao, 1996; E. M. Barnes et al., 2000; Frampton et al., 2013).

In addition to traditional VIs, many studies use or advocate the use of biophysical variables (BVs) derived from MS data, particularly LAI, fraction of green vegetation cover (fCover) and fraction of absorbed photosynthetically active radiation (fAPAR) (Waldner et al., 2015; Goffart et al., 2021; Gobin et al., 2023; Do Nascimento Bendini et al., 2024; Vanpoucke et al., 2024). LAI is defined as half the developed area of photosynthetically active elements of the vegetation per unit horizontal ground area (Wolf et al., 1972; Weiss et al., 2020). It is a primary canopy property that should be independent of viewing conditions, yet it is strongly and non-linearly related to reflectance, which makes its RS retrieval scale dependent. Remotely sensed LAI effectively integrates all green layers (including the understorey), as canopy foliage is often clumped rather than randomly distributed (Weiss et al., 2020). fAPAR is absorbed by the green canopy and can be computed directly from radiative transfer within the canopy at a given moment. fCover is defined as the fraction of ground area covered by green canopy elements when viewed from above, i.e. the gap fraction in the nadir viewing direction (Waldner et al., 2015; Weiss et al., 2020). These variables offer valuable insights into crop phenology and provide a more precise description of canopy structure and density. They are also less susceptible to saturation at high biomass levels than the NDVI and are not sensor-specific. This makes them ideal for monitoring green vegetation and separating vegetation from soil (Frampton et al., 2013; Waldner et al., 2015; Gobin et al., 2023).

1.2.5. Multispectral approaches for winter cover crop detection and species identification

Building on the widespread use of VIs and BVs for monitoring crops, MS studies of WCC typically exploit the time series of these indicators to characterise the status of winter soil cover and the development of WCC. In this context, most approaches rely on Landsat, S2 or PlanetScope time series, due to their high spatial and temporal resolution. These time series are used to derive VIs or BVs and temporal metrics to describe WCC phenological stages characteristics such as establishment time, crop-specific greening, peak biomass and senescence (Goffart et al., 2021; Kc et al., 2021; Ahmed et al., 2023; Q. Zhou et al., 2026). Most MS studies have focused on detecting the presence of WCC (Cruz-Ramírez et al., 2012; Seifert et al., 2019; M. L. Barnes et al., 2021; Kc et al., 2021; Ahmed et al., 2023; Najem et al., 2024; Vanpoucke et al., 2024; Vanongeval et al., 2026; Zhou et al., 2026) or estimating its biomass (Goffart et al., 2021; Xia et al., 2021; Jennewein et al., 2022; Do Nascimento Bendini et al., 2024) using both single- and multi-temporal MSI. However, studies identifying WCC species or mixtures remain limited due to the spectral and phenological similarities between WCC types (Cruz-Ramírez et al., 2012; Goffart et al., 2021; Schulz et al., 2021; Xia et al., 2021; Do Nascimento Bendini et al., 2024; Vanpoucke et al., 2024; Vanongeval et al., 2026).

Xia et al. (2021) showed that S2-based biomass estimation improved substantially for single-species WCCs compared to mixed canopies (adjusted $R^2 = 0.52$ pooled vs. 0.89 for red clover alone), highlighting the added difficulty of functional diversity.

In Maryland, Jennewein et al. (2022) evaluated the use of S2 and Sentinel-1 (S1) Synthetic Aperture Radar (SAR) satellite imagery to estimate the biomass of WCC across 27 fields over multiple seasons. They emphasised the importance of RE bands in overcoming saturation issues in biomass estimation and also identified some limitations relating to the effects of different species.

In France, Do Nascimento Bendini et al. (2024) showed that RF trained on dense S2 time series substantially improved biomass estimation across diverse mixtures ($R^2 \approx 0.75$), confirming the value of ML over empirical VI-based approaches.

In Belgium, Goffart et al. (2021) used S2 data to estimate the aboveground biomass of mixed and pure (mustard, phacelia and oat) WCCs. Biomass samples from 2016-2017 were related to 73 VIs derived from S2 bands, and over 450 empirical models were tested using exponential and power regressions with 10-fold cross-validation. The best models were achieved by using the Chlorophyll Vegetation Index for mustard and MERIS Terrestrial Chlorophyll Index for phacelia. They achieved Root Mean Square Errors (RMSEs) of 0.36 and 0.30 Mg DM ha^{-1} ($R^2 = 0.92$ and 0.78), while mixtures showed higher errors (≈ 0.61 Mg DM ha^{-1}) due to spectral variability. The study confirmed S2's potential for objective biomass estimation, though species mixtures and cloud cover remain challenges.

Seifert et al. (2019) trends in the adoption of CCs and their effects on yields across maize and soybean fields in eight Midwestern U.S. states between 2008 and 2016. Using Landsat composites and RF classification, they achieved an accuracy rate of 91.2% (Cohen's kappa (κ) = 0.74) and established a link between the presence of CC and yield outcomes. This demonstrates how RS can support management and policy evaluation.

In the US state of Indiana, Barnes et al. (2021) advanced MS approaches by combining harmonised Landsat–Sentinel reflectance with SWIR and thermal surface temperature data in RF models. Using two- and three-class classifiers, they predictes winter land use (WCC vs. non-WCC (NWCC)) and distinguished WCCs, crop residues and bare soil. They systematically increased model complexity, progressing from inputs of only NDVI to VIS–NIR indices, SWIR-based tillage indices, and finally thermal surface temperature metrics. While models using only NDVI showed poor accuracy ($\kappa \leq 0.2$), the most complex configurations, which included NDVI, VIS-NIR, SWIR and thermal data, achieved accuracies close to 90% for the presence/absence of WCCs ($\kappa \approx 0.72$) and around 80% for the three-class scheme ($\kappa \approx 0.69$). Surface temperature emerged as the most important predictor, while SWIR tillage indices provided additional discriminatory power between crop residues and bare soil.

As well in Indiana, Zhou et al. (2026) used phenology-based NDVI time series from multiple satellite sensors (harmonised Landsat-8 and S2 data (HLS), Moderate Resolution Imaging Spectroradiometer (MODIS) and MODIS-calibrated PlanetScope) to detect WCCs. They found that HLS achieved the highest accuracy (≈76%). The detection performance increased with larger field sizes and higher regional adoption rates and varied strongly among species. Wheat and winter grains were the easiest to detect, while cereal rye performed worst. This highlights how management success and species traits can influence the strength of remotely sensed WCC signals.

In the Maumee River watershed in Ohio, Kc et al. (2021) developed a scalable approach for mapping WCCs using Landsat imagery (2008–2019) and field observations in Google Earth Engine (GEE). An RF classifier was used to distinguish between four categories of WCC: Winter Hardy, Winter Kill, Spring Emergent and Not Covered. This was achieved using seasonal NDVI composites for autumn and spring, with an accuracy of 75% (κ = 0.63). The study also examined the relationship between WCC dynamics and temperature and growing degree days, revealing the key climatic factors influencing growth.

Ahmed et al. (2023) developed an RS framework for detecting WCC in Arkansas. Using Landsat 8 imagery from November to March between 2013 and 2019, they generated single-season median composites that minimised cloud contamination while capturing the WCC growth period. They then trained a pixel-based RF classifier in GEE using ground-truthed WCC versus NWCC fields as the training data. They filtered their training data with an NDVI threshold of 0.3 to ensure that only sufficiently green CC pixels were included and they performed calibration–validation splits at field level to reduce spatial autocorrelation. The resulting binary WCC/NWCC maps achieved very high accuracy (overall accuracy (OA) ≈ 98%, κ ≈ 0.94), and feature importance analysis showed that VIs, especially NDBI, BSI and NDVI, were more informative than individual bands. This underlines the utility of multi-index, multi-temporal composites for large-area WCC mapping.

Similar to Jennewein et al. (2022), who combined optical and SAR inputs, Najem et al. (2024) demonstrated that a S1 SAR time series can be used to map WCCs by exploiting plot-level backscatter dynamics within a decision tree framework. Their approach, which uses filters to exclude winter crops with overlapping phenology, such as wheat and rapeseed. This achieved high performance (recall 83.5–95.0%, precision 81.5–89.2%, κ = 0.72–0.89), outperforming a RF baseline and highlighting both the robustness of SAR to cloud cover during the WCC season, as well as the remaining limitations for sparse WCC parcels and those with crop residues.

In Germany, Schmidt et al. (2021) developed an ML approach for monitoring WCC at the parcel level. Using S2 NDVI time series from July to April, 19 phenology-based features were extracted and used to train RF classifiers across four federal states and four years. This achieved mean accuracies of around 84% for WCC versus NWCC and of up to 90% in years with favourable weather conditions. However, different seed mixtures could not be discriminated, and model performance decreased in years with anomalous heat or frost.

In northern Belgium, Vanpoucke et al. (2024) used an parcel-based approach, aggregating S2 time series at Land Parcel Identificarion System (LPIS) parcel (Section 3.2.3.) level, to perform field-level winter crop classification in a two-level hierarchy. First, they separated winter-crop presence or absence, and second, they distinguished three WCC groups (leafy, grass-like and mixtures) from three NWCC groups (legumes, winter cereals and other winter crops). For the WCC specifically, the start of the season, the rate of spring greening and the rate of autumn senescence, derived from VIs (e.g. NDVI, Enhanced Vegetation Index (EVI) and Soil-Adjusted Vegetation Index (SAVI)) and BVs (fAPAR and fCover) time series, emerged as key features, with the highest median F1-scores around 86–89% for coarse labels. RF provided the most robust performance across different temporal (monthly, decadal, daily) and spectral (NDVI-only, spectral bands and VIs + BVs) inputs. Despite the high scores for coarse labels, fine label classification remained challenging due to substantial confusion among the leafy, grass-like, and mixture WCC groups, which have similar temporal profiles and high within-group variability in mixed canopies.

Similarly, in northern Belgium, Vanongeval et al. (2026) demonstrated that dense S2 time series facilitate the operational mapping of winter soil cover types, including WCCs, crop residues, and bare soil. This approach achieved an OA of approximately 0.85 (κ = 0.73) at the broadest classification level. The study compared S2-derived monthly composites with parameters from double-sigmoid fits to VI time series. This showed that curve-fitted phenological metrics, particularly start-of-season metrics and rates of spring greening and autumn senescence, substantially improved the discrimination of early-, medium- and late-sown WCCs. However, detailed species-level classes remained difficult to separate, achieving an OA of 0.79 (κ = 0.49).

Overall, these studies demonstrate that combining MS time series with related VIs, and BVs provides a robust framework for WCC detection and characterization. However, mixed WCCs remain more difficult to detect and characterize using MS data than monospecific stands.

1.2.6. Limitations of multispectral approaches for winter cover crop detection and monitoring

The main limitation of MSI for species-level WCC mapping arises from its coarse spectral sampling. Broad bands and a small number of spectral channels reduce sensitivity to narrow absorption features associated with pigments, water and structural biochemistry, restricting the ability to discriminate between species with similar overall reflectance but differing biochemical composition (Ghazaryan et al., 2018; Jennewein et al., 2024). Similar patterns have been reported in WCC mapping, where S2 classifications achieved high accuracies for broad winter soil cover classes (e.g. WCC versus winter cereals and fallow), but substantially lower accuracies at the most detailed levels representing individual WCC species or mixtures with overlapping phenology and highly similar spectral responses (Vanpoucke et al., 2024). In addition, many MS missions either lack SWIR bands or include them only coarsely, even though SWIR wavelengths are particularly informative for species discrimination (Basinger et al., 2020). As many species-relevant biochemical traits (e.g. lignin and cellulose content) are expressed through narrow VIS-SWIR absorption features, coarse MS sampling further constrains species-level detection (Xia et al., 2021; Jennewein et al., 2024).

Overall, these constraints imply that while MS RS is well suited to mapping WCC presence, broad functional types and, to some extent, biomass at large scales, it is less suitable for reliably characterising species-level mixtures. By contrast, HS satellites such as PRISMA and EnMAP provide dense sampling across the VNIR-SWIR spectrum, enabling subtle biochemical and biophysical differences related to pigments, water content, biomass and leaf structure to be captured more effectively, even though their spatial resolution (≈30 m) is coarser than that of S2 (10 m) (Clark et al., 2005; Galvão et al., 2019; Basinger et al., 2020; Kharel et al., 2023; Mirzaei et al., 2024). Several studies have shown that, for crop-type mapping, narrow HS bands can yield markedly higher classification accuracies than broadband MS configurations when SNR are sufficient, particularly in the SWIR region (Clark et al., 2005; Bostan et al., 2016; Galvão et al., 2019; Yao et al., 2019; Mirzaei et al., 2024). This indicates that, for certain tasks, spectral richness is a stronger limiting factor than spatial resolution. Successful species discrimination requires interspecific spectral variability to exceed intraspecific variability (Clark et al., 2005), a condition that is more likely to be met when HS data are available.

1.2.7. Hyperspectral remote sensing data applied to vegetation monitoring and species discrimination

HSIs can be viewed as three-dimensional data cubes, with two spatial dimensions (along-track and across-track) and one spectral dimension. For each pixel in the spatial grid, the sensor records a full reflectance spectrum across hundreds of contiguous bands, so every pixel location (x, y) is associated with a detailed spectral signature of the surface. (Ortenberg, 2019; Brosinsky et al., 2022).

Imaging spectroscopy, the term now often used for HSI, has been a well-established RS approach since the 1980s. It is primarily based on airborne and field spectrometers and enables the identification and quantification of bio- and geochemical surface properties through characteristic spectral features in the visible near-infrared (VNIR) and SWIR regions (Brosinsky et al., 2022). HSI data for crop and species discrimination have predominantly been obtained from proximal field spectrometers and airborne platforms, such as unmanned aerial vehicles (UAVs) and aircraft. Sensors like the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) have been widely used in agricultural campaigns (Clark et al., 2005; Papp et al., 2021; Qian, 2021; Ram et al., 2024). Although these platforms offer very high spatial and temporal resolution and are less affected by atmospheric disturbances, the cost and limited spatial coverage of acquisitions hinders larger-scale applications (Lillesand et al., 2015).

In recent decades, spaceborne HS sensors such as Hyperion, PRISMA and EnMAP have begun to bridge this gap by providing HS data at regional scales. Although they have coarser spatial resolution than many airborne systems and are less widely available than MSIs, they provide rich spectral information over large areas in a more practical and cost-effective way than repeated airborne campaigns, making them increasingly well suited to regional agricultural monitoring (Ortenberg, 2019; Ram et al., 2024; Bourriz et al., 2025).

The substantially richer spectral information and enhanced sensitivity of HSI support a wide range of applications, including the quantitative assessment of crop monitoring, the detection of diseases and weeds, the measurement of chlorophyll content, the early detection of stress, the differentiation of vegetation at the species and crop type levels, and the mapping of fine-scale land cover (Thenkabail, 2002; Thenkabail et al., 2004; Dudley et al., 2015; Tagliabue et al., 2022; Chabrillat et al., 2024; Ram et al., 2024).

However, the high dimensionality of HS data also introduces several challenges. These include large storage requirements, increased data and transmission rates, and substantial computational burdens for data handling and analysis, which together call for new methods tailored to high-dimensional datasets (Thenkabail et al., 2019). HSI is rich in information, but the large number of bands can degrade classification performance when training data are limited. This is a manifestation of the Hughes phenomenon, whereby adding features leads to unstable accuracies and overfitting because model parameters are estimated from too few samples in the presence of highly collinear bands (Galvão et al., 2019; Tan et al., 2025). Using raw HSI spectra directly in ML models, often reduces the accuracy of biochemical and biophysical trait retrieval, underlining the importance of appropriate pre-processing steps such as noise reduction, reflectance standardisation and feature engineering (Ram et al., 2024; Tan et al., 2025). Tan et al. (2025) demonstrated that a band selection method employing constrained sampling with affinity propagation could enhance mapping accuracy by 3–4 percentage points on the

Salinas and Indian Pines crop datasets compared to full-band and other band-selection baselines, while utilising fewer, more discriminative bands. Furthermore, strong inter-band correlations generate large amounts of redundant spectral information. In high-resolution scenes, fine-scale variability in crop chemical composition can reduce inter-class separability, making it more difficult to distinguish between land cover or crop classes based solely on spectral information.

Dimensionality reduction, which eliminates redundant bands while preserving or enhancing the separability of different classes, is a critical pre-processing stage in HSI workflows. Band selection is one of the most widely used approaches to reducing the number of channels while maintaining classification performance (Thenkabail et al., 2019; Tan et al., 2025). Other techniques such as lambda-lambda plots, which identify redundant versus unique spectral bands, and Principal Component Analysis (PCA), which condenses highly collinear bands into a smaller set of uncorrelated components, are commonly used to extract the most informative regions of the spectrum (Thenkabail et al., 2019). Galvão et al. (2019) also emphasise the need for feature selection and for combining the selected narrow bands and HS VIs (HVIs) – which follow the same general logic as conventional VIs, but exploit very narrow spectral bands – and spectral absorption features as inputs to ML classifiers (Section 1.2.9.).

Early work into the optimal placement of HS bands for vegetation studies showed that only a few narrow bands carry most of the information needed to quantify crop and canopy traits. However, many of these studies covered a limited number of species and a restricted spectral range (400-1,050 nm). This motivated a broader assessment across the 400-2,500 nm range, particularly the systematic evaluation of SWIR bands. Subsequent analyses confirmed the importance of SWIR for agricultural and vegetation studies. It was noted that, although several new-generation sensors include SWIR bands, these are still relatively broad. This limits the exploitation of narrow absorption features (Thenkabail et al., 2004).

Using Hyperion data, Thenkabail et al. (2004) showed that many optimal bands recur at similar centres across vegetation types. They also showed that reducing hundreds of HS bands to the first five principal components, which still account for around 95% of the variance, can reduce the volume of data by about 97%. In related work focusing on six agricultural crops, Thenkabail (2002) found that using the twelve best narrow bands produced a rapid increase in spectral discrimination. This raised OA from 73% ($\kappa \approx 0.71$) when using six bands to 84% ($\kappa \approx 0.79$) when using twelve bands. However, additional bands only marginally improved discrimination, and accuracies plateaued around thirty bands, with peak OA reaching approximately 94% ($\kappa \approx 0.92$) (Thenkabail, 2002). Detailed error analyses indicated that adding bands reduces both omission and commission errors. However, beyond approximately 30 narrow bands, the incremental gains are minimal and can be more effectively achieved by adding or substituting carefully chosen mid-infrared bands. These are particularly useful for crops with strong soil background influences (Thenkabail, 2002). Based on these findings, Thenkabail (2002) and Thenkabail et al. (2004, 2019) concluded that many of the most informative narrow bands cluster around consistent centres within the VIS–NIR region. They also found that a set of approximately 12-14 bands, each with a nominal bandwidth of around 10-20 nm, is typically sufficient for capturing most crop-relevant information. Very narrow bands (≤5 nm) tend to have a low SNR, whereas very broad bands (≥30 nm) can lose the specificity of key absorption features. Based on an extensive review of modelling, statistical feature selection, and classification experiments, Thenkabail et al. (2019) proposed an 'optimal' set of HS narrow bands for studying vegetation and croplands in the 400-2,500 nm range. They argued that focusing on these bands reduces data volumes and alleviates the curse of dimensionality. This enables conventional classification methods to effectively operate on a compact subset of channels that captures most of the variability in vegetation traits and crop types.

Spectral pre-processing involving smoothing and derivative transformation is an additional step in feature engineering that can enhance the discriminative power of HSI data. Atmospheric correction can introduce high-frequency noise into reflectance spectra, so a smoothing step is often necessary to stabilise the spectral profiles prior to further analysis (Tsai & Philpot, 1998; Aneece et al., 2019). The spectral smoothing reduces random noise while preserving the main shape of the vegetation spectra. Previous work has shown that spectral averaging after atmospheric correction can improve the robustness of biophysical retrievals and classifications (Aneece et al., 2019). Of the available smoothing methods, the moving-average filter is the most computationally straightforward. It averages the reflectance within a local spectral window, thereby reducing random noise. However, it offers slightly weaker spectral detail preservation than polynomial-based approaches, such as the Savitzky–Golay filter (Tsai & Philpot, 1998; Wang et al., 2019). Following smoothing, computing the first-order spectral derivative — the finite difference between adjacent bands — amplifies local slope changes in the reflectance curve. This enhances narrow absorption features that would otherwise be obscured in the raw reflectance data (Tsai & Philpot, 1998). Derivative spectra are particularly valuable for detecting RE shifts associated with chlorophyll content, as well as for suppressing additive background effects such as illumination variation and soil brightness. These effects can reduce inter-class separability (Tsai & Philpot, 1998; Roberto et al., 2019; Wang et al., 2019). However, because derivatives are sensitive to noise, smoothing must precede differentiation. If smoothing is inaccurate, the differentiation step propagates rather than removes error (Wang et al., 2019). Derivative spectroscopy has been applied to various vegetation parameters, such as leaf and canopy water content, chlorophyll concentration, and stress detection. It has consistently demonstrated added value over raw reflectance when narrow absorption features are the target (Roberto et al., 2019; Wang et al., 2019).

Together, these studies emphasise that achieving the full potential of imaging spectroscopy for monitoring vegetation and discriminating between species or crops depends on more than just spectral richness. Rigorous pre-processing, dimensionality reduction and band optimisation strategies that target the most informative parts of the spectrum are also required, as are strategies that control data volume and model complexity.

1.2.8. Emerging hyperspectral remote sensing applications for winter cover crops detection and characterization

Although HSI has been widely applied to vegetation monitoring and characterisation, only a few studies have explicitly targeted WCC broad functional types and biomass estimation.

Wang et al. (2023) conducted the first airborne HS study to quatify WCC growth traits in the US Corn Belt. Using imaging spectroscopy (400–2400 nm, 0.5 m resolution) over 23 cereal-rye fields, they developed a Process-Guided Machine Learning (PGML) framework that integrates radiative-transfer models with ML. PGML achieved high predictive performance for above-ground biomass ($R^2 = 0.72$, relative RMSE = 15.2 %) and canopy nitrogen content ($R^2 = 0.69$, RMSE = 16.6 %), demonstrating the ability HS data to capture biochemical variability. It also shows that PGML can reduce the need for field data while improving transferability across heterogeneous soil and vegetation conditions.

Kharel et al. (2023) evaluated the estimation of biomass in mixed-species WCCs by combining high-resolution Planet imagery with HS canopy reflectance data collected using a FieldSpec spectroradiometer. They tested whether biomass in WCC mixtures could be predicted using ML models and investigated how strongly prediction performance depended on growth stage and species/mix information. The models performed better when species/mix information was provided, and overall, biomass prediction was more accurate using HS data than Planet imagery.

To date, no study has systematically evaluated the ability of spaceborne HSI to detect WCCs or distinguish individual WCC species and dominant species within mixtures at regional scale, particularly under European policy and management conditions. However, the existing work by Kharel et al. (2023) clearly shows that HS sensors can capture fine spectral variations related to plant structure and chemistry of WCCs.

1.2.9. Machine learning approaches to classify winter cover crops using earth observation data

Many recent studies using MSI and HSI rely on supervised ML algorithms to classify crop types and retrieve crop traits (Basinger et al., 2020; Vanpoucke et al., 2024). Supervised classifiers are favoured over model-based approaches because they are more robust. These classifiers learn the spectral-temporal characteristics of target classes from labelled samples and can then identify these patterns in unlabelled data. However, several challenges must be addressed to develop an efficient supervised classifier, including the Hughes phenomenon, nonlinear relationships between variables, imbalanced and noisy training data, and ensuring computational efficiency (Belgiu & Drăguţ, 2016).

Following feature extraction or selection, a variety of classifiers can be applied, including support vector machines (SVM), RF, stochastic gradient boosting and linear discriminant analysis, as well as deep-learning (DL) approaches such as artificial neural networks (ANN), convolutional neural networks (CNNs) and multilayer perceptrons. These data-driven methods are well suited to high-dimensional spectral data and can capture complex, nonlinear relationships between reflectance features and crop classes (Belgiu & Drăguţ, 2016).

Among various supervised ML algorithms, RF has consistently demonstrated strong performance in classification using HSI (Belgiu & Drăguţ, 2016), as well as for WCC classifications, being robust to noise, overfitting, and imbalanced training datasets (Seifert et al., 2019; Barnes et al., 2021; Kc et al., 2021; Ahmed et al., 2023; Wu et al., 2023; Vanpoucke et al., 2024). RF is an ensemble method that builds many unpruned Classification and Regression Trees from bootstrapped samples of the training data. Each tree is grown using a different random subset of observations and predictor variables at each split. Roughly two-thirds of the samples (the in-bag set) are used to train each tree, while the remaining one-third (the out-of-bag samples) are used for internal cross-validation to estimate prediction error (the out-of-bag (OOB) error). By aggregating many high-variance, low-bias trees, RF forms a stable ensemble. Two key parameters must be specified when growing the forest: the number of decision trees and the number of predictor variables considered at each split (Belgiu & Drăguţ, 2016). The final prediction is obtained by majority voting across all trees, which stabilises the model and reduces overfitting relative to a single decision tree. In addition, RF yields variable importance measures that help to identify the most informative spectral bands and VIs for discrimination. This makes it particularly attractive for earth observation (EO) applications where the number of predictors is high, they are correlated and noisy, and training data are limited (Ahmed et al., 2023; Wu et al., 2023; Vanpoucke et al., 2024). In RF, variable importance can be computed in different ways, for instance as Mean Decrease in Gini (MDG) or Mean Decrease in Accuracy (MDA). MDG quantifies the reduction in the Gini impurity metric for a given class. It accumulates the weighted reduction in node impurity at each split where a variable is used, averaged across all trees in the forest. MDA is based on the increase in OOB error after permuting the values of a given variable relative to the original dataset (Louppe et al., 2013; Belgiu & Drăguţ, 2016).

Recently, DL algorithms have gained attention in crop-mapping studies. Several studies report that CNN- or ANN-based models achieve higher classification accuracy than classical ML algorithms, particularly when spectral features are combined with spatial texture and contextual information (Alami Machichi et al., 2023; Vanongeval et al., 2026). DL architectures can learn joint spectral-spatial representations and are highly scalable with modern parallel computing. However, they typically require large, well-labelled training datasets and extensive hyperparameter tuning. They may also overfit or fail to generalise when applied to relatively small, imbalanced agricultural datasets. Moreover, their internal decision processes are often opaque, which limits interpretability (Alami Machichi et al., 2023). In contrast, ML methods an especially tree-based ensemble methods such as RF generally perform competitively on modest-sized datasets. They have lower computational requirements and more transparent variable importance diagnostics and are therefore frequently preferred for operational WCC mapping applications (Vanpoucke et al., 2024; Vanongeval et al., 2026).

CHAPTER
TWO

OBJECTIVES

Building on the knowledge gaps identified in the literature review, this chapter defines the research objectives of the thesis. The focus is on quantifying how far current MS and HS satellite data can support operational monitoring of WCCs in Wallonia at different levels of detail, from simple presence detection to mixture type and dominant species composition. The chapter first summarises the overarching research question and then specifies three concrete objectives, each corresponding to a particular data configuration and classification task.

RS of WCC faces several constraints related to sensors and data. Persistent cloud cover in autumn and winter and low sun angles reduce the availability and quality of optical observations during the key establishment and early-spring periods. Additionally, within-pixel heterogeneity and class mixing are limitations: individual pixels often integrate signals from multiple surface types, complicating species-level interpretation (Cruz-Ramírez et al., 2012). Temporal and spatial coverage are particularly limited for spaceborne HSI, with only a few usable scenes and sparse time series available over most regions.

Further methodological challenges restrict the mapping of WCCs. Spectral confusion between WCCs and crop residues is common, particularly for weak or low biomass WCCs, whose signal blends soil background. Species with similar canopy architecture are inherently more difficult to distinguish spectrally. MS studies on WCCs report decreased classification performance when attempting to distinguish between different species. Classification performance also depends strongly on phenological stage, and model transfer across years, regions and management systems remains challenging. Systematic work on identifying optimal acquisition windows for WCC species and mixture discrimination in real cropping systems is still scarce (Basinger et al., 2020; Vanpoucke et al., 2024).

Finally, knowledge gaps remain in our understanding of the relative benefits of MS and HS data. MSI is well established and successfully used for mapping the presence of WCC and estimating biomass, mostly using seasonal NDVI time series or related VIs. However, classifying WCC species remains difficult. There are few harmonised comparisons of HS versus MS data for discriminating between crop types or species at a satellite scale, despite HS sensors theoretically offering improved biochemical and structural separability. To date, no studies have examined the use of spaceborne HSI for determining the presence or absence of WCCs and for distinguishing between individual and dominant species within mixtures at regional scales.

Building on these gaps, the aim of this thesis is to evaluate the extent to which MS time series, single-date MS and HS data can be used to detect the presence of WCC, differentiate between mixture types, and infer species composition in the Walloon region of Belgium. Performance across all objectives is quantified using OA, $F1_{macro}$ and class-wise F1 scores. The specific objectives are to:

I. Quantify the added value of EnMAP HS data for WCC monitoring by analysing spectral separability and RF feature importance, evaluating whether additional narrow band information improves WCC presence detection, mixture identification, and species composition inference.
II. Compare EnMAP HSI and S2 MSI for WCC identification, mixture differentiation, and dominant species detection, under harmonised pre-processing, feature sets, calibration–validation splits, and RF settings, to isolate the effect of spectral resolution.
III. Assess the classification performance of monthly S2 composites from August 2025 to March 2026 for detecting WCC presence and separating mono- from multi-species mixtures in Wallonia (tile 31UFS), using RF classifiers with spectral bands, VIs and LAI.

CHAPTER
THREE

MATERIAL & METHODS

This section describes the study area, datasets and methods used to quantify the potential of EnMAP HSI and S2 MSI for monitoring WCC in Wallonia. First, the agronomic and policy context of the study area is presented. This is followed by a description of the satellite-based data sources (EnMAP and S2), LAI products, LPIS parcels and in-situ field observations used as ground-truth data. The pre-processing of each dataset, the construction of spectral and temporal features, and the RF classification framework designed to address the three objectives formulated in Section 2 are then detailed: (i) the presence or absence of WCC; (ii) the WCC mixture type (pure or mixed); and (iii) the dominant WCC species composition.

3.1. Study area

3.1.1. Proposition: Abiotic factors and agricultural context

The study area is located in the Hesbaye region of Wallonia (16,901 km²) in southern Belgium (see Figure 5). Within Wallonia, nine agro-geographical regions are defined based on biophysical conditions and agricultural systems, one of which is Hesbaye (Zhou et al., 2025). Situated in the north-east of Wallonia, Hesbaye lies on the border between the provinces of Liège, Namur, and Walloon Brabant. This study focuses on two areas of interest (AOIs) within the Hesbaye region. The first AOI corresponds to the Walloon part of S2 tile 31UFS, while the second AOI is defined by the spatial intersection between the first AOI and the extent of the available EnMAP acquisition. As the S2 and EnMAP satellites have different swath widths and orbital paths (Section 3.2.1.), their spatial footprints do not align perfectly. The S2 AOI covers 3,691.6 km², whereas the common S2–EnMAP AOI covers 644.5 km² of Wallonia.

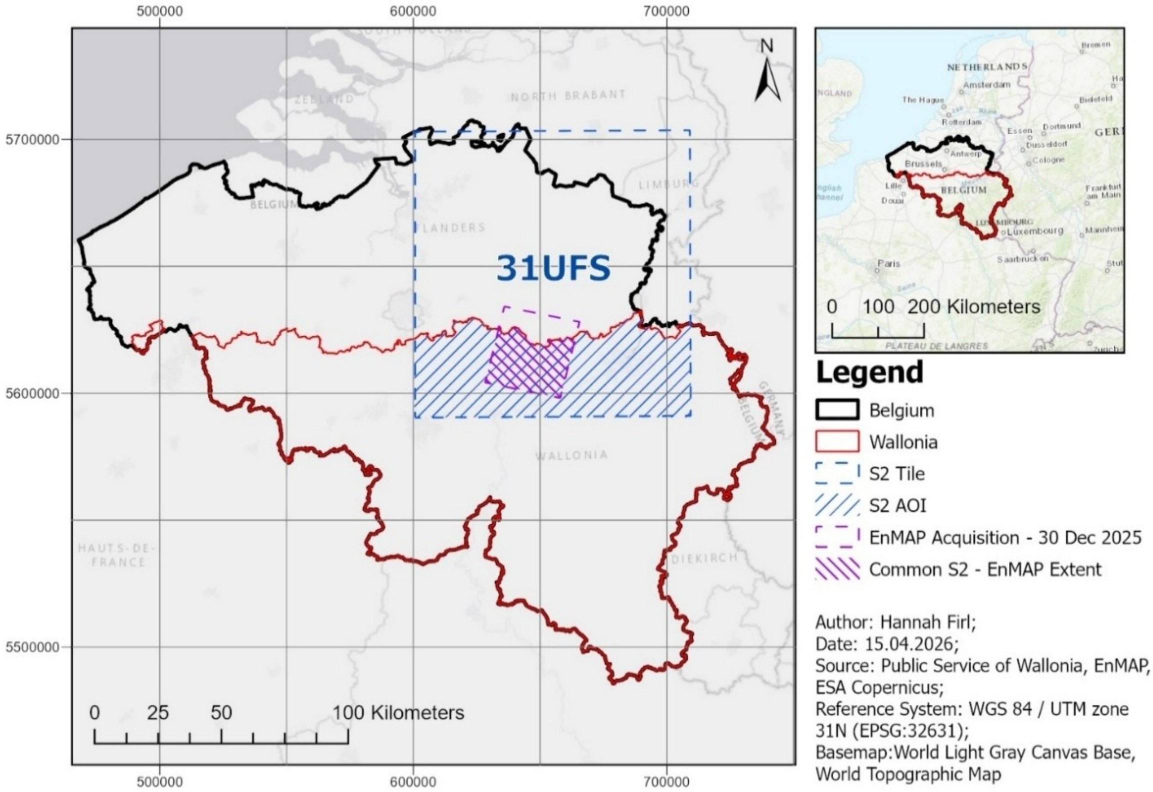


Figure 5: Overview of the study area encompassing Wallonia in southern Belgium within north-western Europe and analysed EnMAP and S2 tiles.

In the Hesbaye region, altitude increases progressively from 50 metres above sea level in the north to a maximum of 200 metres in the south. The annual total precipitation ranges from around 800 to 1,000 mm, and the mean annual temperature ranges from around 9 to 10.5 °C (Chartin et al., 2017). The region has a warm temperate, permanently humid climate and can be classified as Cfb according to Köppen (1936). It exhibits a temperate oceanic climate influenced by Atlantic air masses, characterised by relatively mild winters and cool summers with evenly distributed precipitation (IRM, 2026).

The Hesbaye region is characterised by intensive agriculture and fertile silt loam soils. Major crops grown here include permanent grassland, winter wheat, potatoes, sugar beet, maize and flax. The mean parcel size in the S2 AOI is 2.9 ha, which is slightly above the Walloon average of 2.55 ha. This reflects the open-field arable nature of Hesbaye (Géoportail de la Wallonie, 2025; Zhou et al., 2025).

Overall, the combination of deep, fertile soils, open-field, intensive arable systems, and regional climate and topographical gradients provides favourable conditions for WCC establishment.

3.1.2. Suitability of Wallonia for hyperspectral and multispectral winter cover crop analysis

The adoption of WCCs varies geographically. In many European countries, including Belgium, legal requirements and improved communication have led to a significant increase in WCC use over the past thirteen years. In Belgium, the establishment of WCCs has more than doubled (Figure 6).

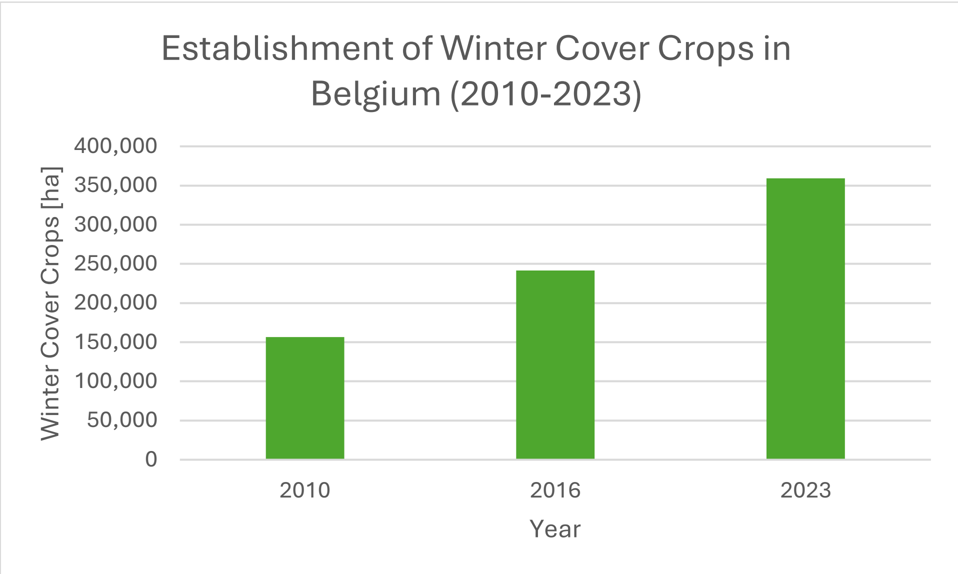


Figure 6: Evolution of WCC establishment in Belgium between 2010 and 2023 (Data: eurostat.eu).

Similar trends are observed at the provincial scale in Namur, Liège and Walloon Brabant, where the area under WCCs has steadily increased between 2010 and 2023 (Appendix 1). Specifically, the Hesbaye region is entirely located within the Nitrate Vulnerable Zones (NVZ), as designated by the European Union’s Nitrate Directive. This regulation mandates that farmers in the NVZ region are obliged to cover the soil with WCCs to reduce nitrate leaching and protect water quality (Zhou et al., 2025). Further growth is anticipated under the CAP Strategic Plan for the period 2023–2027 (European Comission, 2022, 2025; Eurostat, 2025). In Belgium, WCCs are typically established between August/September and March/April. The choice of WCC species is strongly influenced by crop rotations and regional regulations. Common WCCs are often sown as mixtures, frequently including fast-growing species such as mustard (*Sinapis alba* sp.), phacelia (*Phacelia tanacetifolia* sp.) and oats (*Avena sativa* L.), often with a range of mixed cover (two to four different species), to take advantage of the benefits of multispecies WCCs (Section 1.1.) (Goffart et al., 2021; Vanongeval et al., 2026).

In addition to agronomic motivations, the use of WCC in Wallonia is heavily influenced by policy frameworks. Two regulatory pillars are particularly relevant:

(1) the conditionality obligations of the CAP, introduced in January 2023,
(2) the Programme de Gestion Durable de l'Azote (PGDA), Wallonia's nitrate directive, which was updated in April 2023 (European Commission, 2025).

These policies require farmers, under certain conditions, to establish WCCs and catch crops on specific plots to enhance soil protection and water quality. In the context of Wallonia and EU regulations, catch crops are a specific type of WCCs established primarily to catch residual nitrogen and reduce nitrate leaching (*Cultures Intermédiaires Pièges à Nitrates* (CIPAN), engl. Catch Crop for Nitrate Trapping). These obligations apply to farmers with holdings of more than 10 ha of arable land, where less than 75% of the area is covered by grass or other herbaceous forage crops and less than 75% of the total utilised agricultural area is permanent grassland. The three main measures affecting WCC management are BCAE (*Bonnes Conditions Agricoles et Environnementales*) 8 (*Surfaces & Éléments Non Productifs*), BCAE 6 (*Couverture minimale des sols*), and the PGDA (European Commission, 2025).

BCAE 8 sets out the maintenance and protection of non-productive landscape features. When farmers choose to use *dérobées* (rapidly growing catch crops or WCCs) as non-productive surfaces, they must adhere to specific management rules. The cover must be sown between 1st of July and 30th of September and maintained for at least three months. Destruction is permitted only mechanically or by natural frost until 15th of February, after which chemical destruction is allowed. BCAE 8 does not impose requirements regarding species mix and does not define CIPAN (European Commission, 2025).

The PGDA establishes obligations relating to the management of nitrogen, particularly in NVZs. Under the PGDA, farmers must sow a CIPAN on specific plots before 15 September and maintain it until at least 15 November (European Commission, 2025). CIPAN mixtures must adhere to specific composition rules, such as restrictions on legume proportions, and ensure at least 75% soil coverage by early November (European Commission, 2025).

BCAE 6 aims to guarantee a minimum level of soil cover and avoid bare soil during sensitive periods. Between the 15th of September and the 15th of November, farmers are required to maintain vegetative cover on at least 80% of their arable land that will be sown with spring crops the following year. Acceptable covers include crop residues, regrowth of cereals or oilseeds, intercrops or winter crops. On plots with a high or very high risk of erosion, this vegetative cover must be maintained until 31st of December (SPW, 2025).

Overall, the study area provides an appropriate environment in which to analyse the detection and characterisation of WCCs using MSI and HSI, given the intensive agricultural land use and widespread adoption of WCCs in this region. Furthermore, this region benefits from the frequent revisit capabilities of satellite constellations at this latitude, which helps mitigate persistent cloud cover (Section 3.2.1.).

3.2. Data description

3.2.1. Satellite data

For the HSI and MSI analysis, Level-2A surface reflectance products from respectively the German EnMAP mission and the Copernicus S2 mission were used. Level-2A products provide atmospherically corrected bottom-of-atmosphere (BOA) surface reflectance derived from Level-1C top-of-atmosphere (TOA) data (CDSE, n.d.). Key sensor characteristics are summarised in Table 1.

Table 1: Technical specifications of EnMAP and S2 (Qian, 2021; Scheffler et al., 2023; Chabrillat et al., 2024; CDSE, n.d.; Copernicus, n.d.; EnMAP, n.d.).

	EnMAP	**S2**
Number of spectral bands	213	13
Spectral range [nm]	420-1,000 (VNIR) 900-2,450 (SWIR)	443-2,190
Bandwidth [nm]	6.5 (VNIR) – 10 (SWIR)	15-180
Ground sampling distance [m]	30	10 (VNIR), 20 (VNIR in RE, SWIR), 60 (atmospheric bands)
Swath width [km]	30	290
Revisit time [days]	27 (can be reduced to ~4 through ±30° off-nadir pointing)	5 (in Belgium 2-3)
SNR	> 500 (at 495 nm; VNIR) > 170 (at 220 nm;SWIR)	~70-170 (VNIR) ~50-100 (SWIR)
Radiometric Resolution	≥ 14 bits	12 bits
Radiometric accuracy / stability	5% / 2.5 % between calibrations	≤ 5 % absolute; 1 % multi-temporal relative

EnMAP is a German HS satellite mission launched in 2022, designed to provide high-quality imaging spectroscopy data for monitoring terrestrial and aquatic ecosystems (Earth Observation Center of DLR, n.d.; Kaufmann et al., 2012). It carries a prism-based push-broom dual spectrometer with on-board calibration and stringent data quality requirements, resulting in high SNRs in both VNIR and SWIR regions. Since November 2022, acquisitions have been planned via a user-driven observation request system. In this system, users submit specific imaging requests, which are then scheduled by the mission planning system (Scheffler et al., 2023; Chabrillat et al., 2022, 2024). By contrast, S2 systematically collects data over all land surfaces, providing freely and consistently accessible multispectral imagery that can be downloaded without prior acquisition request (CDSE, n.d.).

The S2 mission comprises a series of European Earth observation satellites launched between 2015 and 2024. S2B and S2C operate in the same sun-synchronous orbit, 180° apart, to achieve global coverage. S2A currently operates 36° away from S2B. The MS instrument uses a push-broom sensor. In temperate European croplands, dense S2 time series are particularly well suited to monitoring winter soil cover under the European Union's CAP, which links financial support to sustainable soil management (Copernicus, n.d.; Vanongeval et al., 2026).

The HSI was acquired over the Walloon study area on 30 December 2025. Although an acquisition synchronous with the with the primary field campaign on 7 November (Section 3.2.4.) was initially planned to capture peak green above ground vegetation, persistent cloud cover rendered the November imagery unusable. Consequently, the first clear-sky acquisition available in late December was selected to represent WCC conditions. For the analysis, a single spatial tile encompassing the core study region (Tile 002, derived from continuous data acquisition ID 0000173176, captured at 11:22:29 UTC) was selected. The Level 2A product was radiometrically calibrated and atmospherically corrected to BOA reflectance using the standard EnMAP Ground Segment processing chain (version 01.05.05), which converts at-sensor radiance to surface reflectance (Kaufmann et al., 2012).

To cover the entire WCC growing season in Wallonia, from emergence to termination, S2 data was acquired for the 31UFS tile between August 2025 and March 2026. After accounting for acquisition frequency and cloud screening, approximately 8-14 usable scenes per month were obtained (Appendix 2). Only the 10 m (B2, B3, B4 and B8) and 20 m (B5, B6, B7, B8A, B11 and B12) bands of the 13 S2 bands were used (Appendix 3) (CDSE, n.d.). In addition to the multitemporal time series, a single S2 image from 30 December 2025 was extracted for direct comparison, as this was the image closest in time to the EnMAP acquisition. The same 10 m and 20 m bands were used for this analysis.

3.2.2. Biophysical variable: Leaf Area Index

LAI time series were obtained from the BELCAM (BELgian Collaborative Agriculture Monitoring) platform, which provides operational crop biophysical indicators derived from S2 imagery for Belgium at a spatial resolution of 10–20 m, updated approximately every five days. LAI is retrieved by applying calibrated empirical relationships between S2 reflectance (and derived VIs) and field-measured canopy properties, in line with standard approaches for LAI estimation from optical RS. For this study, BELCAM LAI values were extracted for all LPIS parcels within the S2 tile 31UFS from August 2025 to March 2026 (Belcam, 2026).

3.2.3. Land Parcel Identification System

To extract satellite spectral signatures at the level of individual parcels, the LPIS has been used as the reference layer for polygon geometries. The LPIS is a geospatial administrative database used by all European Union Member States to support the management and control of agricultural subsidies under the CAP. It provides a harmonised, plot-based representation of agricultural land and serves as the reference system (Grandgirard & Zielinski, 2008; Leonhardt et al., 2024).

Although LPIS types vary across Europe, they generally comprise vector polygon data defining agricultural boundaries, as well as parcel-specific identifiers and crop attributes. The reference parcel system is typically derived from high-resolution orthophotos and is often structured as agricultural parcels. As landscapes are subject to constant anthropogenic and environmental modifications, LPIS databases are regularly updated to reflect changes in land use, such as the splitting or merging of fields or the construction of new features (Grandgirard & Zielinski, 2008; Leonhardt et al., 2024).

For this study, LPIS data from 2025 was used. In Wallonia, LPIS is implemented via the *Système Intégré de Gestion et de Contrôle* (SIGEC) system. The public SIGEC dataset used here is an anonymised version that does not contain any personal information linking parcels to individual farmers, and it is disseminated annually. Discrepancies can occur between declared LPIS parcels and actual field conditions, particularly for short-lived or partially established land covers (Grandgirard & Zielinski, 2008; Géoportal of Wallonie, 2025). As SIGEC/LPIS also served as the primary instrument for planning in-situ data acquisition, field validation was conducted during the campaigns to verify that the declared LPIS attributes accurately reflected the presence and type of WCCs.

3.2.4. Ground truth data

In-situ field data were collected during targeted campaigns conducted within the study area in the Hesbaye region of Wallonia in autumn/winter 2025/26. The study area was selected based on practical considerations and RS criteria. Its proximity to Louvain-la-Neuve (the starting point of the study) reduced logistical constraints, and the 2025/26 agricultural year coincided precisely with the overlap of the planned S2 acquisitions and the available EnMAP HSI. The field observations provide reference information on the presence of WCC, its mixture, species composition, and structural characteristics at the level of individual agricultural parcels. This information serves as the primary ground truth for training and validating satellite-derived classification models (Section 3.5.2.). The dataset is stored as a spatial point layer, with each observation linked to a unique agricultural parcel ID. This enables direct association with LPIS parcels and satellite images.

Field visits were scheduled to coincide with cloud-free satellite overpasses during the vegetative phase, as closely as possible. Three main field campaigns were conducted. The first took place on 7 November 2025 in the Hannut region, coinciding with an S2 overpass and one day after an EnMAP acquisition on 6 November. However, the corresponding EnMAP image was unusable due to heavy cloud cover. A second campaign was carried out on 28 November 2025 in the region around Louvain-la-Neuve/Corbais. The final campaign took place on 13 January 2026, providing ground truth for a clear EnMAP image acquired on 30 December 2025. To account for the delayed EnMAP acquisition and ensure that ground observations matched the land cover state at the satellite overpass as closely as possible, previously sampled plots were revisited and re-evaluated. Heavy snowfall from 6 to 11 January prevented earlier fieldwork, so the field observations may differ from conditions at the time of the EnMAP acquisition due to some frost damage and canopy degradation. However, no WCCs were destroyed or terminated between the acquisition and the field visit, and all sampled plots still contained living WCC stands.

Candidate parcels were pre-selected using recent S2 imagery and road-access constraints, with the aim of maximising the number of parcels visited per campaign while ensuring spatial coverage across the study area. Only arable parcels (excluding permanent and temporary grassland) with an area of at least 0.25 hectares and clearly delineated boundaries were retained to ensure robust spectral extraction. The size limit was chosen with EnMAP's 30 m native spatial resolution in mind, as it is the coarsest resolution among the datasets used. In combination with the 20 m inward buffer applied, this ensured that each retained parcel still contained at least a few interior pixels suitable for spectral analysis. During fieldwork, geolocated observations and photographs were collected using the QField mobile application on a smartphone to ensure direct synchronisation with the GIS database (QField, n.d.).

As summarised in Table 2, the recorded attributes included WCC presence or absence, mixture status (pure or mixed), species composition, intra-parcel canopy heterogeneity, estimated soil cover, and vegetation height. Species composition was documented qualitatively. During the follow-up visits, parcels were re-evaluated to document whether they still hosted active WCCs, had been terminated, or had transitioned into winter cereals or other crops. This transition status was recorded as a separate attribute ('Class_old'), thereby improving the interpretability of the EO-based classifications.

Table 2: Field Attributes Recorded in the Data Acquisition.

Attribute	Description	Units / Format
Parcel ID ('ID')	Unique identifier for each agricultural parcel	Text / Numeric
Date of visit ('date_visit')	Date of field observation	YYYY-MM-DD
WCC presence ('CC_presenc')	Presence or absence of WCC	Binary: Yes / No
'Class'	Simplified binary classification of parcel land cover	Categorical: 1 = WCC / 2 = NWCC
'Class_old'	Follow-up categorization to document WCC persistence or transition	Categorical: 1= still WCC / 2= winter cereal / 3 = other crops
WCC mixture ('CC_mixed')	Classification of WCC as pure or mixed	Categorical: Pure / Mixed
Species composition ('CC_comment')	Qualitative description of species present	Text (e.g., mustard, phacelia, ryegrass, clovers)
Vegetation height ('Height')	Height of dominant species	cm, recorded in 10 cm intervals
Percentage cover ('perc_cover')	Estimated canopy cover of WCC	%, recorded in 20% intervals
Visual intra-parcel canopy heterogeneity ('Heterogene')	Degree of spatial variability within parcel	Categorical: 2 = High / 1 = Intermediate/ 0 = Homogeneous
Georeferenced photographs ('photo1-photo3')	Visual documentation of vegetation conditions	Image files with geolocation (if available)

The most observed species were mustard, phacelia, nyger, ryegrass, clover, vetch, field beans, fodder radish, sunflowers, cabbage, cereals and alfalfa (Table 3). Throughout the field campaigns, we also encountered several NWCC classes, including winter cash crops, bare soil, crop residues, and other NWCC uses. The field campaigns revealed considerable diversity in the appearance of both WCC and NWCC plots. Among WCC parcels, there was substantial variation in canopy density, species dominance, height and soil cover: dense, homogeneous WCCs (e.g. pure phacelia or pure mustard) contrasted markedly with sparse or mixed stands where significant bare soil was visible between plant rows. Mixed stands often combined species with contrasting leaf architectures, such as broad-leaved mustard alongside fine-leaved ryegrass or clover. This produced highly heterogeneous canopies at the sub-parcel scale. NWCC parcels were equally diverse and included winter rapeseed (which is botanically related to mustard but is a cash crop rather than a WCC), early-stage winter wheat on bare soil and flax residues. Table 4 provides examples of this diversity and illustrates how visually similar some pure and mixed WCC and NWCC parcels can appear, as well as showing how stand appearance can vary within a single species class. Additional examples are provided in Appendix 4.

Table 3: Overview of all used WCC species with their common name, scientific name and group (Sharma et al., 2018; Adetunji et al., 2020; Vanpoucke et al., 2024).

Common Name	Scientific Name	Classification Group
Ryegrass	*Lolium perenne*	Grass Group (Non-legume)
Cereal Rye	*Secale cereale*	Grass Group (Non-legume)
Mustard	*Sinapis alba*	Brassica (Non-legume)
Cabbage	*Brassica oleracea*	Brassica (Non-legume)
Fodder Radish	*Raphanus sativus subsp. oleiferus*	Brassica (Non-legume)
Phacelia	*Phacelia tanacetifolia*	Other Broadleaf (Non-legume)
Nyger	*Guizotia abyssinica*	Other Broadleaf (Non-legume)
Sunflower	*Helianthus annuus*	Other Broadleaf (Non-legume)
Ravanelle (Wild Radish)	*Raphanus raphanistrum*	Other Broadleaf (Non-legume)
Vetch	*Vicia villosa*	Legume
Pea	*Pisum sativum*	Legume
Field Bean (Féverolle)	*Vicia faba*	Legume
Alfalfa	*Medicago sativa*	Legume
Clover	*Trifolium* spp.	Legume

Table 4: Representative field photographs from the November 2025 campaign illustrating the diversity of WCC and NWCC parcels encountered during in-situ data collection.

WCC

Pure:

Phacelia

Mustard

Mixed:

Phacelia + Mustard + Ray grass

Phacelia + Clover + Nyger

NWCC

Winter rapeseed

Flax residues

Sed beed with crop resdiues

Winter wheat emergence after deep plowing

The final dataset comprised 289 WCC parcels and 293 NWCC parcels across the full S2 AOI. Figure 7 shows the spatial distribution of the sampled parcels across the agricultural landscape, distinguishing between the full S2 study area (nWCC = 289, nNWCC = 293) and the smaller common S2–EnMAP extent centred on the Hannut region (nWCC = 179, nNWCC = 92). Some observation points are flagged as 'detection only', indicating that only binary land cover information (WCC vs. NWCC) was recorded, with no details of crop mixture/species or structural attributes. Including NWCC parcels was essential for assessing the classification performance of the model, since the early phenological stages of NWCCs can appear similar to WCCs in optical imagery (Ahmed et al., 2023). Class distribution in the in-situ dataset was set to approximately 39% WCC and 61% NWCC parcels to reflect the multi-annual mean proportion of WCC within the S2 31UFS tile (mean over 2017–2023: ~39%) (Descarpentries, 2025, unpublished work).

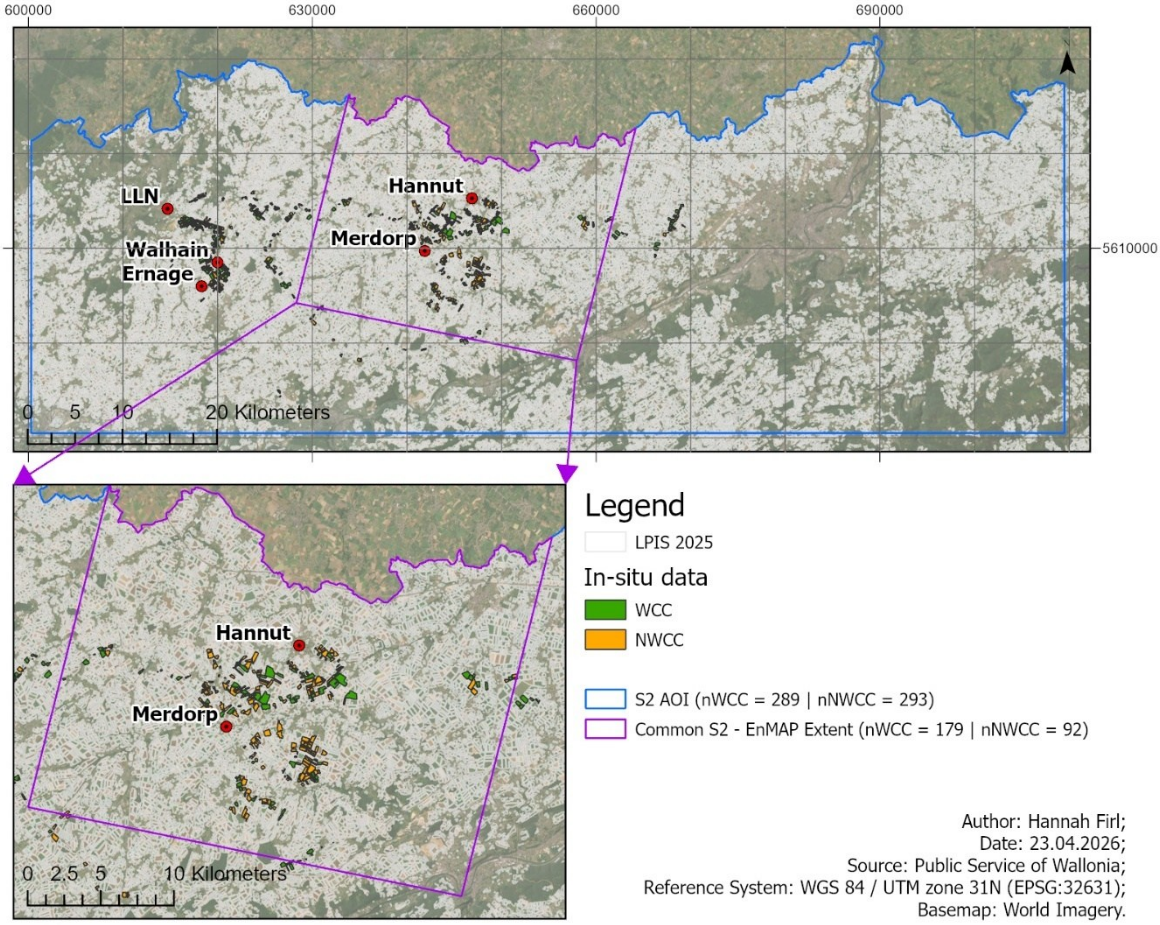


Figure 7: Spatial distribution of in-situ WCC and NWCC sampling parcels within the S2 AOI and the common S2–EnMAP extent, Hesbaye region, Wallonia (2025/26 growing season).

3.3. Overview of methodological framework

Figure 8 summarises the general workflow applied in this study. Detailed, step-by-step workflows for the three experimental setups (EnMAP single date, EnMAP–S2 comparison and S2 monthly composites) are presented in Figure 9 and provided in full size in Appendices 5-7.

EO data (S2 or EnMAP) are first pre-processed using atmospheric correction, cloud masking and clipping to the AOI, as well as imagery-specific signal analysis, including spectral pre-processing steps (Sections 3.4.1.–3.4.2.). LPIS parcel geometries and in situ observations are harmonised by buffering and filtering parcels by size, aggregating field observations spatially to parcel-level classes and rasterising the resulting labels (Sections 3.4.3.–3.4.4.). Finally, RF models are trained using repeated calibration-validation splits on pixel-level predictors, followed by feature-importance analysis, parcel-level majority filtering and accuracy assessment (Sections 3.5.2.–3.5.4.). In this study, three RF models are used: RF-1 (WCC vs. NWCC), RF-2 (pure vs. mixed WCC) and RF-3 (WCC species composition).

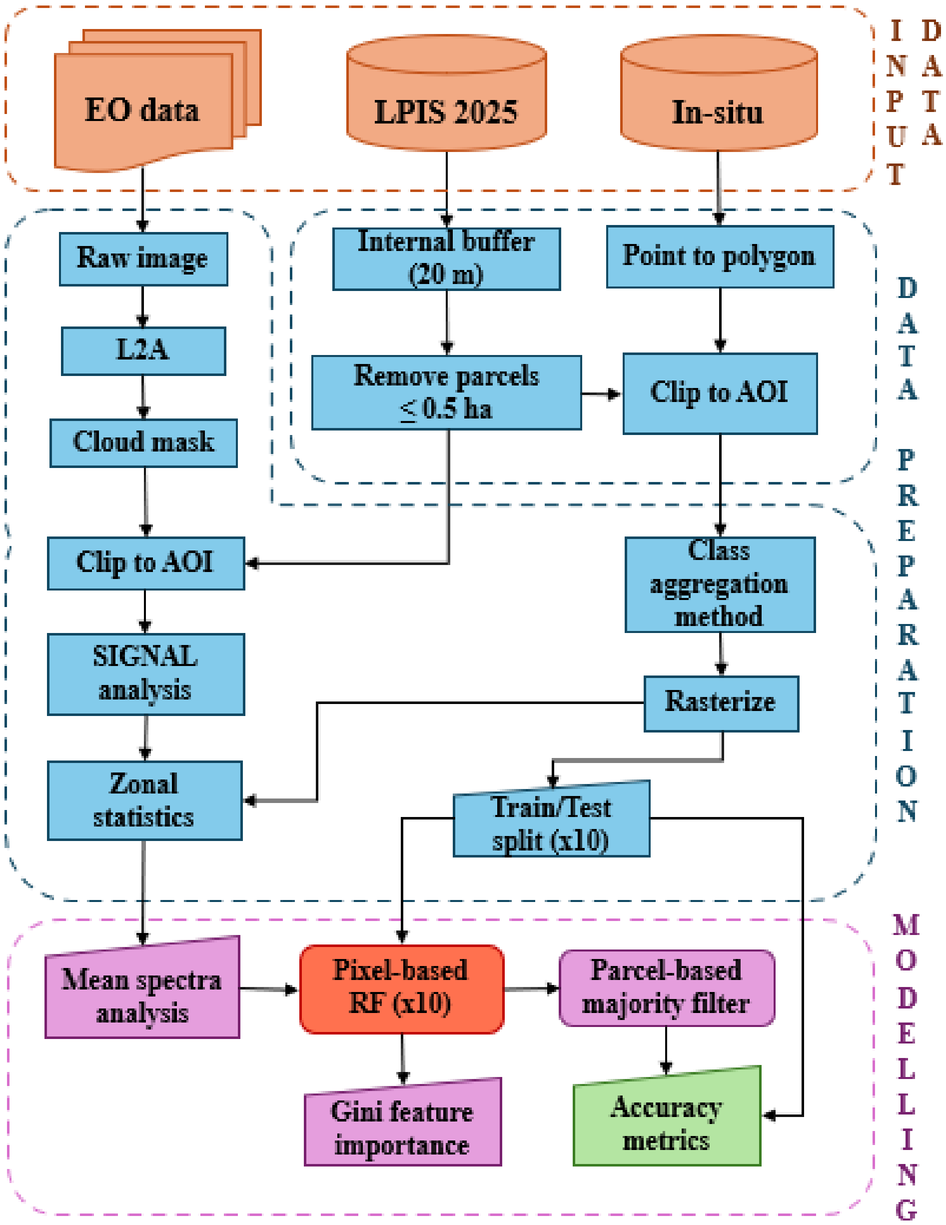


Figure 8: General workflow for EO-based WCC presence, mixture and species mapping.

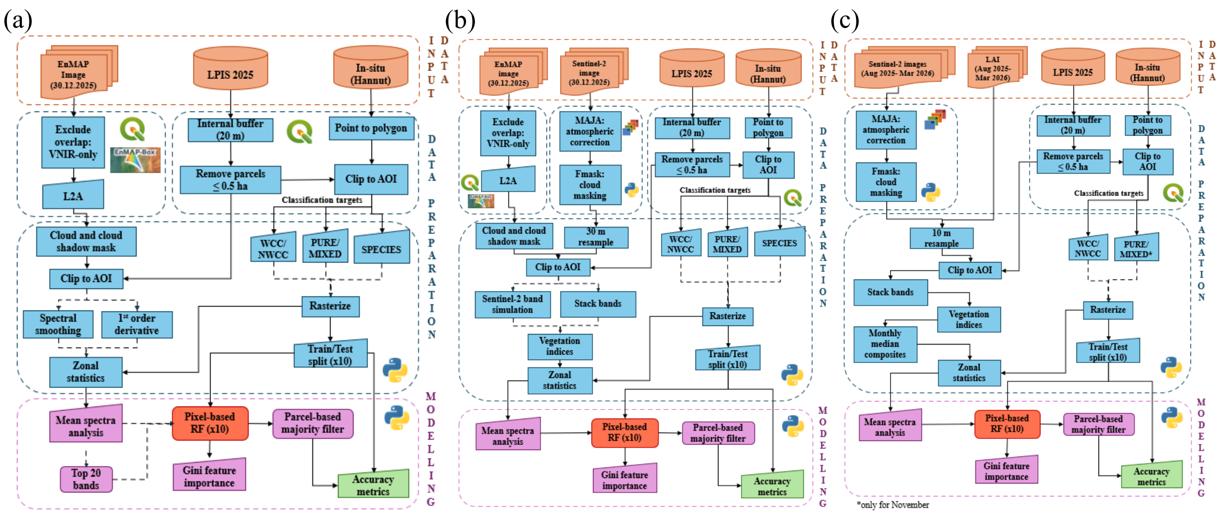


Figure 9: Detailed workflows for EO-based WCC presence, mixture and species mapping for (a) Setup 1, (b) Setup 2 and (c) Setup 3.

Figure 10 illustrates the overall sampling and modelling workflow (Sections 3.5.2.-3.5.4.). It shows the ten randomly generated parcel-wise calibration-validation splits, the pixel-based RF training and prediction based on these datasets, and the final parcel-level majority filtering. This is used to derive spatially consistent classifications for RF-1 to RF-3. This is followed by the computation of accuracy metrics.

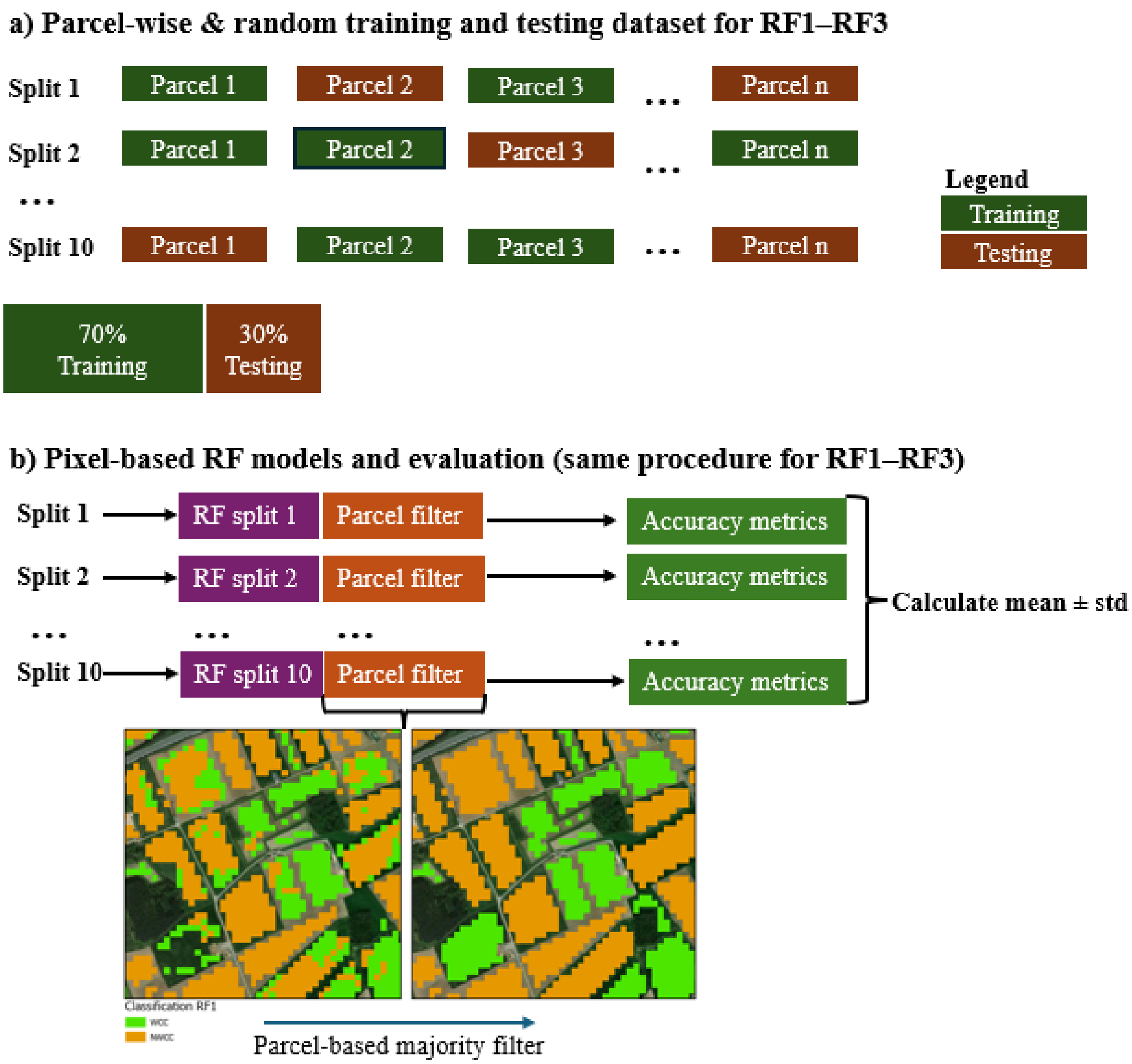


Figure 10: Workflow for (a) ten independent parcel-wise train-test splits generated by repeated random sampling and (b) pixel-based RF classification and parcel-level post-processing.

3.4. Pre-processing of satellite, parcel data and ground truth

Prior to classification, a series of pre-treatment steps were applied to the EnMAP and S2 images, the LPIS parcel geometries and the ground-truth datasets. All pre-processing were implemented in Python (v3.12.6) (Python Software Foundation, 2026).

3.4.1. Pre-processing of EnMAP images

The EnMAP acquisitions were visually inspected in QGIS (v3.34.10-Prizren) (QGIS Development Team, 2023). Level-2A products were loaded as HS raster layers using the EnMAP-Box 3 Toolbox (v3.15.3) in QGIS, which reads the product metadata and applies the recommended band scaling. To address the spectral overlap between the VNIR and SWIR spectrometers in the 900-1,000 nm region, only the SWIR portion of the overlapping bands was retained. This resulted in the processed images containing 208 spectral bands.

Cloud and cloud-shadow masking

The masks delivered with the HSI, which included haze, clouds, cloud shadows and cirrus, were then applied to the EnMAP imagery. However, visual inspection revealed inadequate cloud shadows flagging in the native mask. To address this issue, the Cloud/Shadow Detection based on Spectral Indices (CSD-SI) algorithm proposed by Zhai et al. (2018) was adapted to the EnMAP HSI to detect residual cloud shadows. The CSD-SI algorithm calculates specific Cloud Indices and Cloud Shadow Indices from the VNIR bands, using spatial matching and median filtering to distinguish genuine shadows from dark land cover features (Zhai et al., 2018). Despite threshold optimization, detection of diffuse cloud edges was not satisfactory for the study, especially at the border of the Walloon region. Consequently, additional spatial exclusion rules were applied to remove remaining artefacts. A 1.5 km internal buffer was applied along the border between Wallonia and Flanders to eliminate edge-of-scene clouds, and a custom geometric polygon was manually digitised to permanently mask a complex, persistent cloud structure in the lower right of the EnMAP tile. These masked areas did not overlap the in-situ parcels and had only limited impact on predictions, as most parcels in this zone were excluded from the analysis. Finally, the raster was clipped to the buffered LPIS parcels (Section 3.4.3.).

Spectral pre-processing: smoothing, derivatives and band selection

Additional spectral pre-processing was applied to the atmospherically corrected EnMAP VNIR image to reduce high-frequency noise and derive alternative spectral features for EnMAP-based classification.

First, a one-dimensional moving-average filter with a window size of three bands was applied along the spectral axis to obtain a smoothed reflectance cube, whereby each band's reflectance was replaced by the mean of itself and its two adjacent bands. The mean filter smoothing is computed by Eq. (1):

$$\hat{s}(\lambda_j) \frac{\sum s(\lambda_j)}{n} \qquad (1)$$

where n (number of sampling points) is the filter size and j is the index of the middle point of the filter (Tsai & Philpot, 1998).

Secondly, the first-order spectral derivative was computed from the smoothed cube by taking the finite difference between adjacent bands. In this study the first derivative is estimated by Eq. (2):

$$\frac{ds}{d\lambda}\bigg|_i \approx \frac{s(\lambda_i) - s(\lambda_j)}{\Delta\lambda} \qquad (2)$$

where $\Delta\lambda$ is the separation between adjacent bands, $\Delta\lambda = \lambda_j - \lambda_i$ and $\lambda_j > \lambda_i$, and the interval between bands is assumed to be constant (Tsai & Philpot, 1998).

Both the smoothed reflectance and the first-order derivative were stored as separate datasets and subsequently used as alternative predictor sets in the RF classification experiments.

EnMAP-Sentinel-2 band simulation (single-date comparison)

For the EnMAP–S2 comparison on 30 December 2025, the EnMAP data were spectrally resampled to match the S2 MSI bands. Using the official EnMAP spectral response information (EnMAP, 2023), the central wavelength and bandwidth of each S2 band (Appendix 3) were matched to contiguous groups of EnMAP bands. For instance, the S2 B2 (blue) band was simulated by averaging EnMAP bands 9–24, which cover a similar wavelength range. The same approach was used to simulate bands B3, B4, B5, B6, B7, B8, B8A, B11 and B12. The mean reflectance across the corresponding EnMAP bands was computed to obtain one simulated S2 band. All simulated bands were stacked into a single 10-band raster. NDVI, NDRE, NDWI and BSI (Table 5) were then calculated from these bands and appended as additional layers. This harmonisation was designed to isolate the effect of sensor-intrinsic characteristics (radiometric resolution, SNR, spectral response function (SRF), calibration) from differences that arise purely from band placement and spatial resolution, so that any residual performance gap between the two datasets under identical feature sets could be attributed to these instrument-level properties rather than to differences in spectral sampling alone.

3.4.2. Pre-processing of Sentinel-2 images

Atmospheric correction, cloud and cloud-shadow masking

S2 imagery was pre-processed using the MAJA (MACCS-ATCOR Joint Algorithm) atmospheric correction chain in combination with the Fmask cloud masking algorithm. Level-2A products were generated from S2 Level-1C images with MAJA (version 4.10.0). MAJA performs radiometric, geometric, and atmospheric corrections to derive surface reflectance products (Hagolle et al., 2017).

After atmospheric correction, Fmask (version 4.6) was applied using a highly conservative filtering approach (Qiu et al., 2019). Only pixels with a cloud probability of 10% or less were retained. Pixels affected by clouds, cloud shadows, snow, water or other invalid surface conditions were excluded from further analysis.

Sentinel-2 monthly composites

All S2 raster files were organised systematically by acquisition date, and the consistency of their band structure and metadata was verified. To enable multi-band analyses at a common spatial resolution, the 20 m bands were resampled to 10 m using bilinear interpolation and all the images were cropped to the AOI. The resampled 20 m bands and the original 10 m bands were then combined into a single multi-band raster for each date. Reflectance values outside the expected S2 surface reflectance range (0–10,000) were set to 'nodata', and valid values were scaled to the 0-1 range. For each date, a set of VIs (NDVI, NDRE, NDWI and BSI; Table 5) was computed and stacked with the reflectance bands. This yielded a cloud-masked, multi-band product for each acquisition. To generate cloud-reduced monthly composites, all available stacks within each month were merged using a median compositing approach.

To integrate a BV into these monthly composites, BELCAM LAI images available for the same months were resampled to 10 m and aggregated into monthly median composites using the same method. These monthly LAI composites were then added to the reflectance-VI stacks, producing a 15-band monthly composite for each pixel. This was then clipped to the buffered LPIS parcels (Section 3.4.3.).

Table 5: Spectral VIs used in this study, including their mathematical formulation (S2 band notation) and their original author.

Original acronym	Index	(Derived) S2 Formula	Original author
NDVI	Normalized Difference Vegetation Index	$\frac{(B8 - B4)}{(B8 + B4)}$	(Rouse. et al., 1974)
NDWI	Normalized Difference Water Index	$\frac{(B8 - B11)}{(B8 + B11)}$	(Gao, 1996)
NDRE	Normalized Difference Red Edge Index	$\frac{(B8 - B5)}{(B8 + B5)}$	(Barnes et al., 2000)
BSI	Bare Soil Index	$\frac{(B11 + B4) - (B8 + B2)}{(B11 + B4) + (B11 + B2)}$	(Chen et al., 2004)

Sentinel-2 image for EnMAP comparison (single-date comparison)

For the single-date comparison with EnMAP on 30 December 2025, the corresponding S2 L2A image was clipped to match the exact EnMAP footprint (Section 3.4.1.). To match the EnMAP pixel size, all relevant 10 m and 20 m bands were resampled to 30 m using bilinear interpolation. The NDVI, NDRE, NDWI and BSI were then calculated at 30 m and appended to the band stack (Table 5). Finally, this was then clipped to the buffered LPIS parcels (Section 3.4.3.).

3.4.3. Pre-processing of LPIS data

To minimise edge effects in the pixel-based analysis, a 20 m inward buffer was applied to each LPIS parcel to exclude spectrally mixed boundary pixels and retain pixels that predominantly represent the target land-cover within each field (i.e. spectrally more uniform interior pixels), following Vanpoucke et al. (2024). After buffering, parcels with a remaining core area below 0.5 ha were discarded. This threshold was chosen empirically by inspecting the buffered polygons in QGIS: smaller cores often degenerated into small and narrow strips or irregular fragments rather than compact field interiors, which seem less suitable for reliable pixel extraction and pixel-based RF modelling.

3.4.4. Pre-processing of ground-truth data

First, in-situ observations were harmonised with the buffered LPIS parcels by spatially joining field points to the corresponding agricultural polygons, with the parcel geometry assigned as the reference unit. For each parcel, we distinguished two broad land-use categories: WCC parcels, where a WCC had been established during the fallow period, and NWCC parcels, where no WCC was present.

To support the classification of species composition (RF-3, Section 3.5.), individual species records in WCC parcels were aggregated into dominant canopy composition classes based on the species present in the upper canopy, which are primarily visible to the sensors. This aggregation followed a set of hierarchical rules that group pure phacelia and mustard stands, their main mixtures, and all other combinations into a limited set of classes (Table 6).

Table 6: Dominant canopy composition classes used for RF-3 species-composition classification.

Class name	Type	Description
NWCC	NWCC	No recorded WCC species; parcel under bare soil, crop residues, winter cash crops or other NWCC.
Unknown WCC (detection only)	WCC	WCC present but no species specified in the field data; used only for RF-1 presence detection.
Pure Phacelia	WCC	Single-species WCC stands dominated by phacelia.
Pure Mustard	WCC	Single-species WCC stands dominated by mustard.
Phacelia mix	WCC	Phacelia with one or more other WCC species, excluding mustard.
Mustard mix	WCC	Mustard with one or more other WCC species, excluding phacelia.
Phacelia+Mustard	WCC	Mixtures containing phacelia and mustard only (no additional WCC species).
Other mixes/stands	WCC	All remaining WCC stands pure stands of other species and mixtures not captured by classes above.

3.5. Classification Workflow

The classification framework consisted of three sequential RF models applied at pixel level: (i) WCC vs. NWCC (binary level) (RF-1), (ii) WCC mixture type (pure vs. mixed, binary level, RF-2) and (iii) WCC species composition (RF-3) (Table 7). RF-1 discriminates WCC from NWCC using all available in-situ parcels. RF-2 as well as RF-3 were trained only on in-situ parcels labelled as WCC with reliable mixture and detailed species-level information. Predictions from RF-2 and RF-3 were applied only to pixels classified as WCC by RF-1, so that mixture and species classification were conditional on WCC presence pixels to reduce noise. Each model was run 10 times using independent calibration-validation splits to reduce the influence of a single partition on reported accuracies.

The three RF models were applied in three experimental setups designed to isolate the effect of sensor characteristics and temporal aggregation and to address the research questions (Table 7). Setup 1 (EnMAP-only) used the 20 most informative EnMAP bands on the 30 December 2025 EnMAP acquisition to evaluate the added value of HS narrow bands for WCC mapping (Section 3.4.1.). Setup 2 (EnMAP-S2 comparison) used a spectrally matched subset of EnMAP bands and S2 bands (plus the same set of VIs) from 30 December 2025 to directly compare EnMAP HSI and S2 MSI under otherwise harmonised pre-processing, features and RF settings (Sections 3.4.1.-3.4.2.). Setup 3 used S2 monthly median composites from August 2025 to March 2026, combining 10 and 20 m bands, VIs and LAI, to assess the influence of seasonal information on classification performance (Section 3.4.2.). RF-3 was only applied in Setups 1 and 2. RF-2 in Setup 3 was restricted to the November composite, which coincided with the field campaign and showed among the highest RF-1 accuracies for WCC presence detection.

Table 7: Overview of the three setups and of the three RF models and their corresponding target labels and class definitions used for WCC classification.

Setup	Sensor data	Date / period	Purpose	Models used	Target labels	Classes
1	EnMAP HSI	30 Dec 2025	Added value of HS narrow bands	RF-1	WCC presence	WCC (1), NWCC (2)
				RF-2	Mixture type	Pure (1), mixed (2)
				RF-3	Species composition	Pure Phacelia (1), Phacelia mix (2), Pure Mustard (3), Mustard mix (4), Phacelia + Mustard (5), Other mixes/stands (6)
2	S2-harmonized EnMAP HSI, S2 MSI resampled to 30 m	30 Dec 2025	Direct HIS EnMAP-MSI S2 comparison under harmonised setup	RF-1	WCC presence	WCC (1), NWCC (2)
				RF-2	Mixture type	Pure (1), mixed (2)
				RF-3	Species composition	Pure Phacelia (1), Phacelia mix (2), Pure Mustard (3), Mustard mix (4), Phacelia + Mustard (5), Other mixes/stands (6)
3	S2 monthly composites	Aug 2025 - Mar 2026 (Nov for RF-2)	Potential of MSI; seasonal information and timing for WCC mapping	RF-1	WCC presence	WCC (1), NWCC (2)
				RF-2	Mixture type	Pure (1), mixed (2)

3.5.1. Zonal statistics, parcel-mean spectra analysis and EnMAP band selection

For exploratory spectral analysis, parcel-mean spectra were computed from the buffered LPIS-clipped satellite datasets (Setup 1, 2 and 3) for all parcels with in-situ observations, by averaging reflectance over all valid pixels within each parcel. For each Setup, class-mean spectra for WCC and NWCC, for pure versus mixed WCCs, and for the species composition groups were then obtained by averaging these parcel-level spectra across all parcels in a given class. These spectra were used to visualise class-specific spectral differences and to support the interpretation of band-selection and feature-importance results, but they were not used directly for training the pixel-based RF models.

To identify the 20 most informative spectral bands for the Setup 1 classifications, a simple standardized mean difference separability criterion was computed for each EnMAP band using the parcel-level mean reflectance. For each band b, the separability score between two classes c_1 and c_2 is defined by Eq. (3) and (4):

$$e_b = \frac{\mu_{c1,b} - \mu_{c2,b}}{\sqrt{\sigma^2_{c1,b} + \sigma^2_{c2,b}}} \qquad (3)$$

$$score_b = | e_b | \qquad (4)$$

where μ and σ denote the class-specific mean and standard deviation of pixel-mean reflectance across all in-situ parcels for class c. The absolute value $|e_b|$ reflects how strongly the two classes are separated at a given wavelength relative to within-class variability.

For RF-1, $|e_b|$ was computed between WCC and NWCC parcels using the smoothed EnMAP reflectance and the first-order derivative. Bands were ranked in descending order of their separability score, so that bands with higher $score_b$ appeared at the top of the list. Rather than simply taking the top 20 ranked bands, a greedy selection with a minimum spacing of three band indices was used, to avoid selecting many highly correlated adjacent bands and to obtain a more spectrally distributed subset. For RF-2 and RF-3, the same ranking procedure was applied to the respective EnMAP smoothed reflectance and derivative representations, and the 20 most informative bands were selected using the same minimum spacing of three band indices between selected bands. This ensured a spectrally distributed subset and avoided selecting many highly correlated adjacent bands across all three RF models. To reduce data volume, mitigate the curse of dimensionality while conserving discriminant features for crop type differentiation, a subset of 20 HS bands has been selected. This is supported by several studies showing that with a maximum of 30 bands, detection performance is maintained whilst minimising computational costs (Thenkabail, 2002; Thenkabail et al., 2004, 2019; Section 1.2.7.). This band-selection step was not applied to Setup 2 or 3, for which all available bands were used.

3.5.2. Calibration-validation datasets

Separate calibration-validation datasets were prepared for the different Setups (Table 7, Figures 5 & 7), due to different scale of acquisitions between S2 and EnMAP (Section 3.1.1.).

To obtain robust and unbiased accuracy estimates, a 70/30 parcel-wise train–test split was adopted in each setup and repeated ten times with independently drawn random seeds (repeated random holdout). Entire parcels were assigned either to calibration or validation, ensuring that no parcel contributed pixels to both sets, in line with recommendations to limit spatial autocorrelation and optimistic accuracy estimates in pixel-based crop-type mapping (Ahmed et al., 2023). Unlike k-fold cross-validation, in which each sample appears in the test set exactly once, the repeated holdout design allows a parcel to be assigned to calibration in some runs and to validation in others, depending on the random shuffle. Across all runs, the relative proportions of WCC and NWCC parcels were kept close to their landscape prevalence so that the models were trained and evaluated under realistic class distributions (Section 3.2.4.). This design provides a more realistic assessment of the expected performance of the wall-to-wall maps derived from the trained RF models.

For Setup 1 and 2, only parcels located within the EnMAP swath in the Hannut region were considered. Figure 11 summarises the total number of in-situ parcels and pixels available per class for RF-1–RF-3 in this EnMAP-based setup. After the internal buffering, the in-situ parcels were rasterised to the EnMAP grid and masks were applied (e.g. clouds, 'nodata', sensor overlap). At this stage, some parcels that initially met the 0.5 ha core-area criterion (Section 3.4.3.) contained only a few valid interior pixels. To avoid calibrating and validating on such extremely small and potentially unrepresentative samples, parcels with fewer than three valid interior pixels were removed, leaving 271 parcels that were used as the sampling frame for the RF models. Across the 10 RF-1 runs, between 184 and 203 parcels were assigned to calibration and 68-87 parcels to validation. Within the WCC subset, RF-2 reused the same parcel-level splits but was restricted to WCC parcels only, resulting in 93-105 calibration parcels and 35-44 validation parcels per run. RF-3 further split the WCC parcels into species-composition classes, with 93-108 calibration parcels and 32-47 validation parcels per run. For RF-2 and RF-3, only WCC parcels with reliable mixture or species information were retained; parcels flagged as detection only were excluded from these two tasks because they did not contain the class information required for mixture or species-level classification.

For Setup 3, the full in-situ dataset was used (Figure 12). The S2 study area covers approximately six times the spatial extent of the EnMAP swath, which explains the substantially larger number of available parcels compared to Setups 1 and 2. After applying the internal buffering, the rasterization to the S2 grid, and cloud/'nodata' masking, in-situ parcels with fewer than three valid interior pixels were again excluded, resulting in 582 parcels that were used as the sampling frame for the RF models. Across the 10 RF-1 runs, between 401 and 431 parcels were used for calibration and 151-181 parcels for validation. For the 10 RF-2 runs, the same parcel-level splits were applied but restricted to WCC parcels with reliable mixture information; detection-only WCC parcels were excluded, resulting in 133-149 calibration parcels and 47-63 validation parcels per run. RF-3 was not applied to the S2 composites, because the detailed species-composition analysis was focused on the EnMAP-based setups.

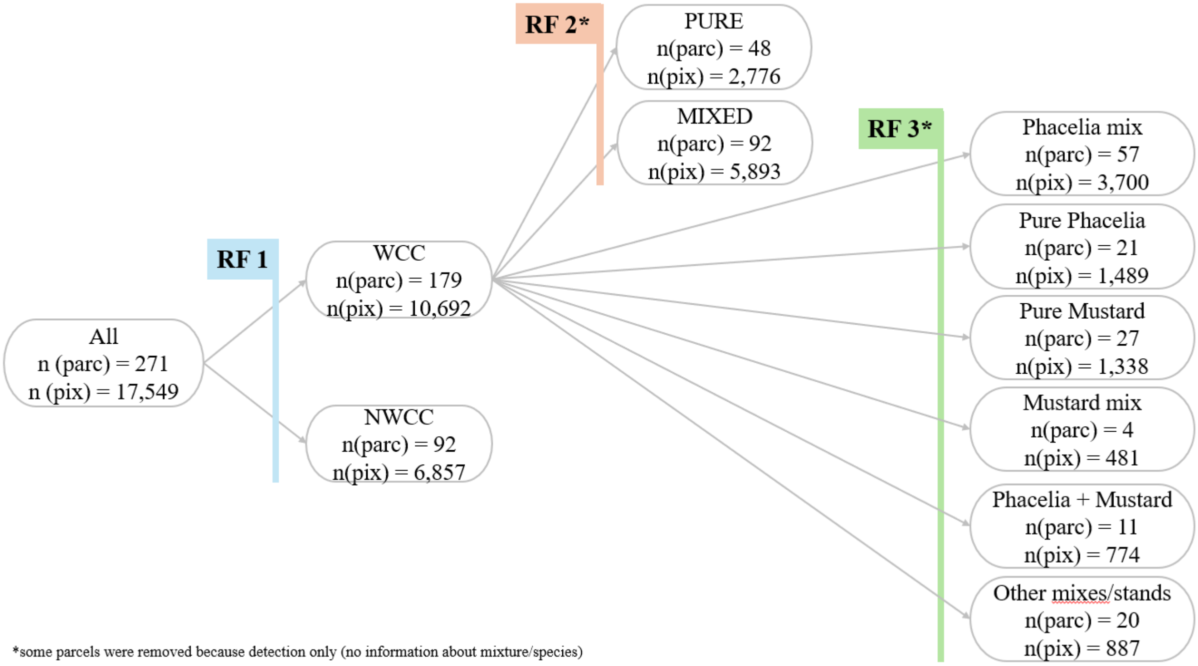


Figure 11: Summary of the number of parcels and pixels per class used for RF-1–RF-3 in Setup 1 and 2.

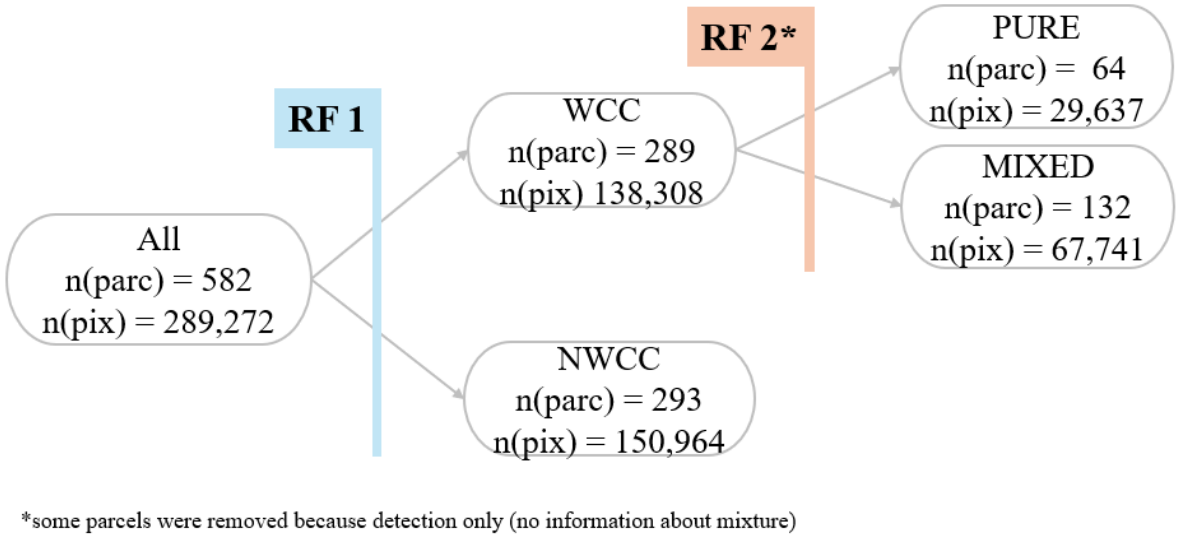


Figure 12: Summary of the number of parcels and pixels per class used for RF-1–RF-3 in the Setup 3.

3.5.3. Random forest models

Because the spectral heterogeneity within agricultural plots significantly impacts species distribution analysis, a pixel-based classification approach was selected over an object-based approach.

All classifications were performed with RandomForestClassifier from scikit-learn (version 1.8.0). An initial hyperparameter tuning was conducted on one representative calibration split using OOB accuracy as criterion (Table 8). The best configuration was then fixed and applied to all 10 runs for the corresponding experiment.

Table 8: Tested hyperparameters set for each RF-model for each setup.

Hyperparameter	Values
Number of estimators	100, 300, 500
Max depth	None, 10, 20
Min samples per split	2, 5, 10
Min samples per leaf	1, 2, 5
Max features	log2, sqrt, 0.5

To identify which features (e.g. spectral bands, VIs) contributed most to classification, MDG feature importance (Section 1.2.9.) was extracted from each of the 10 trained RF models and averaged across runs to obtain a stable ranking of the most informative predictors. MDG quantifies the importance of a variable X_m as the total weighted reduction in Gini impurity accumulated at all nodes where X_m is used to split, averaged over all N_T trees in the forest (Breiman, 2001; Louppe et al., 2013). MDG is computed by Eq. (5):

$$MDG(X_m) = \frac{1}{N_T} \sum_T \sum_{\substack{t \epsilon T \\ v(s_t)=X_m}} p(t) \Delta i(s_t, t) \qquad (5)$$

where $p(t) = N_t/N$ is the proportion of training samples reaching node t, $v(s_t)$ is the variable used at split s_t, and $\Delta i(s_t, t) = i(t) - p_L\, i(t_L) - p_R\, i(t_R)$ is the decrease in Gini impurity at that node, with p_L and p_R the proportions of samples going to the left and right child nodes, respectively. A higher MDG value indicates that a feature more consistently reduces class impurity across the ensemble and thus contributes more to discrimination (Breiman, 2001; Louppe et al., 2013).

Although the RF models operated at pixel level, the final classifications were post-processed at parcel level using the buffered LPIS polygons. For each classification task, a majority vote was applied within every buffered LPIS parcel, assigning the most frequent class among valid pixels as the parcel-level label. This filter was applied both during accuracy assessment on validation parcels and for the generation of the final wall-to-wall classification maps (Section 3.5.4.), enforcing spatial consistency and reducing salt-and-pepper noise. This simple parcel-level majority filter follows the same rationale as object-based image analysis approaches that reduce speckle by aggregating spatially connected pixels into meaningful objects before classification (Ahmed et al., 2023). Pixels flagged as 'nodata' were excluded from the majority vote; however, if a parcel contained only a single valid pixel, the class of that pixel fully determined the parcel-level label, so that parcels with very limited valid observations could still receive a classification.

3.5.4. Accuracy Assessment and Evaluation Metrics

The training and accuracy assessment were carried out separately for each of the 10 calibration-validation splits, and performance metrics were summarised as mean ± standard deviation across runs to quantify the variability due to random partitioning (Figure 11). RF predictions were generated at pixel level for all valid pixels within validation parcels. Because in-situ labels were assigned at parcel level and within-parcel spectral homogeneity was assumed, supported by the inward buffer applied during pixel extraction (Section 3.4.1.), the parcel-level label was considered representative of all pixels within the buffered polygon. A majority vote was subsequently applied to the pixel-level predictions within each parcel, assigning the most frequently predicted class as the parcel-level output. Accuracy metrics were then computed by comparing this majority-voted parcel label against the field-observed reference label. This ensures internal consistency between the parcel-level split design, the granularity of the in-situ reference data, and the reported accuracy metrics. For all experiments, the primary evaluation metrics were OA, macro-averaged F1 ($F1_{macro}$) and class-wise F1 scores.

OA was computed as the proportion of correctly classified pixels over all validation pixels, providing an intuitive single measure of overall performance. OA is is computed by Eq. (6):

$$OA = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

where TP is equal to the number of true positives, TN is the number of true negatives, FN is the number of false negatives, and FP is the number of false positives.

Because OA can conceal class-specific errors in imbalanced datasets, $F1_{macro}$ was additionally reported as the unweighted mean of the per-class F1 scores, giving equal weight to WCC and NWCC, to pure and mixed WCCs, and to each species class in RF-3 irrespective of their support. $F1_{macro}$ is computed by Eq. (7) and (8):

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (7)$$

$$F_{1macro} = \frac{1}{C} \sum_{i=1}^{C} F_{1i} \quad (8)$$

where C is the number of classes.

For the binary RF-1 and RF-2 tasks, separate F1 scores were also reported for the classes (WCC vs. NWCC; pure vs. mixed) to explicitly quantify model performance for WCC detection and mixture discrimination, where the per-class F1 score follows directly from Eq. (7) applied to a single class (C = 1). For RF-3, F1 scores were computed for each of the six dominant species or mixture classes.

To assess whether differences in classification performance between sensor configurations or feature sets were statistically meaningful, paired t-tests were applied to the accuracy metrics across the 10 calibration/validation runs. For each metric (OA, $F1_{macro}$, class-wise F1), the t-test compares the scores of two configurations (e.g. for Setup 2) while accounting for the fact that scores are paired by run, i.e. computed on the same underlying calibration/validation split. The null hypothesis is that the mean difference in performance between the two configurations is zero, and the alternative hypothesis is that this mean difference is non-zero. The t-statistic quantifies the size of the observed mean difference relative to the variability of that difference across the 10 runs; a large absolute t-value indicates that the observed gap is large compared to run-to-run variability. The associated p-value is the probability of observing a difference at least as extreme as the one obtained, assuming the null hypothesis of no difference is true. Under the null hypothesis of zero mean difference, the t-statistic follows a t-distribution with $n - 1$ degrees of freedom. The two-sided p-value is the probability of observing a t-value at least as extreme as the one obtained, computed from this t-distribution. If the p-value is below the predefined significance level $\alpha = 0.05$, the null hypothesis is rejected and the difference in performance is considered statistically significant; otherwise, the observed difference is attributed to random variation in the splits rather than to a systematic advantage of one configuration over the other. In line with common practice, p-values were interpreted as follows: results with $p \leq 0.05$ were considered statistically significant (*), $p \leq 0.01$ highly significant (**), and $p \leq 0.001$ very highly significant (***) (Rainio et al., 2024). The t-statistic was computed according to Eq. (9):

$$t = \frac{d_{mean}}{s_d/\sqrt{n}} \tag{9}$$

where d_{mean} is the mean difference across the n = 10 runs and s_d is the standard deviaton of these differences.

CHAPTER
FOUR

RESULTS

This chapter presents the performance and outputs of the three RF models across the different experimental setups. First, Sections 4.1.–4.4. report results for RF-1 for Setup 1 (EnMAP), Setup 2 (EnMAP–S2 comparison) and Setup 3 (S2 monthly composites), respectively, including mean spectra analysis, feature importance patterns and accuracy metrics. Sections 4.5.–4.8. then summarise the corresponding RF-2 results for the three setups, with a focus on how mixture-type discrimination varies across sensors and acquisition dates. Finally, Sections 4.9.-4.11. present the RF-3 results on WCC species composition for Setup 1 and 2, highlighting class-specific accuracies and spatial patterns of dominant canopy compositions. Section 4.12. analyses how the parcel-based majority filter affects all classification outcomes. While this chapter raises some minor discussion points for clarity, the majority of the results discussion, comparisons and critical analyses will be addressed in Chapter 5.

4.1. Winter cover crop detection using EnMAP (Setup 1, RF-1)

Prior to comparing classification performance, Figure 13 provides an overview of the EnMAP band configuration, showing band indices versus centre wavelength together with the main spectral regions (VIS-SWIR). This contextual plot is used as a reference in the following sections when discussing selected bands and feature-importance patterns (Section 4.1., 4.5., 4.9.).

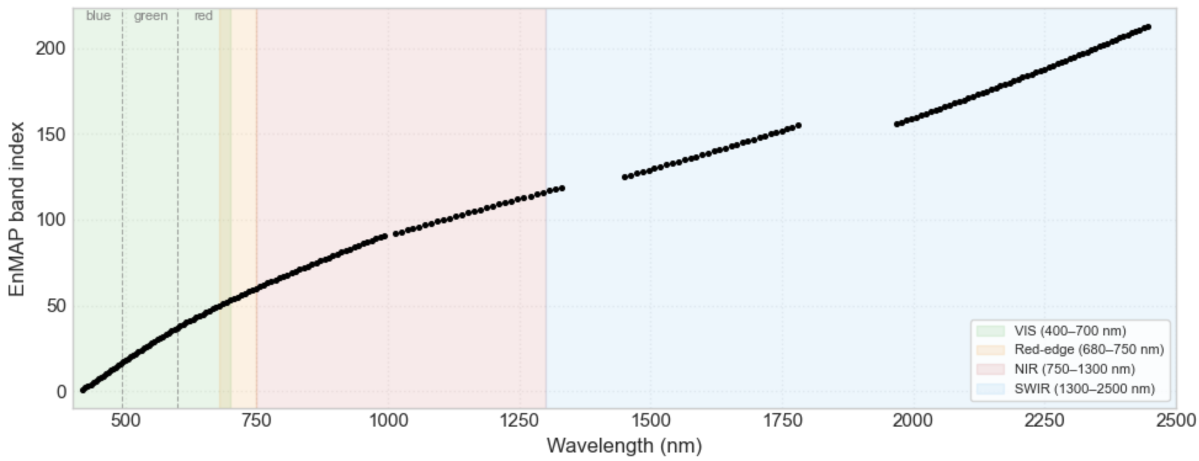


Figure 13: EnMAP band indices and centre wavelengths with main spectral regions (VIS, RE, NIR, SWIR) indicated in the background.

Figure 14 shows the mean EnMAP spectra for WCC and NWCC pixels using all available bands, based on the smoothest reflectance (a) and the first-order derivative of the smoothed reflectance (b). The smoothed reflectance curves reveal very similar spectral shapes for both classes over the full VNIR-SWIR range, with only slight differences in magnitude and strongly overlapping range of variation, whereas the first-derivative spectra highlight subtle differences in the position and strength of several absorption and RE features that remain small relative to the within-class variability. For the classifications using a band selection (Section 3.5.1.), the final feature sets consisted of 20 EnMAP bands each: for smoothed reflectance, bands 3, 6, 9, 12, 15, 18, 22, 25, 28, 31, 34, 37, 54, 130, 133, 136, 139, 143, 146 and 149 were retained, whereas for the first-order derivative input, bands 2, 13, 26, 38, 65, 68, 72, 75, 79, 83, 94, 97, 103, 111, 151, 166, 169, 185, 189 and 193 were selected.

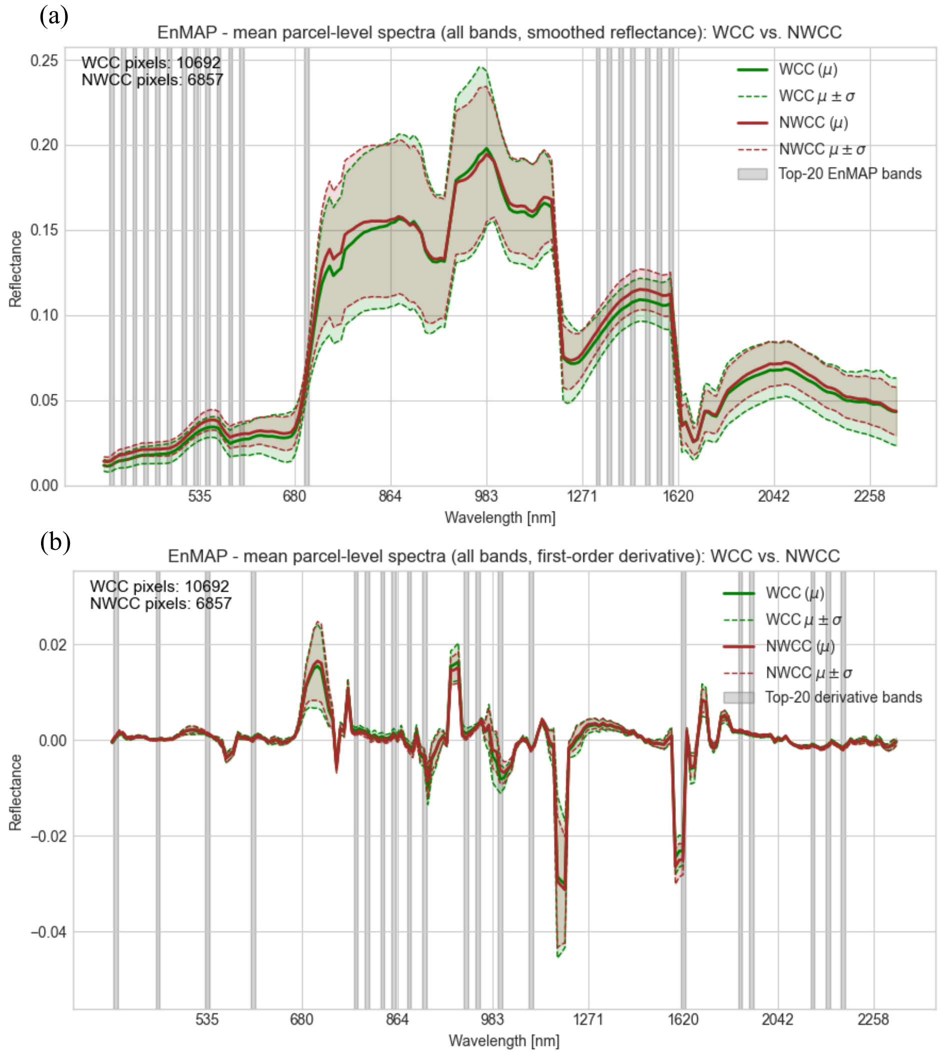


Figure 14: Mean EnMAP parcel-level spectra for WCC and NWCC within the Hannut region. (a) Smoothed reflectance spectra across all bands. (b) First-order derivative spectra across all bands.

For the classifications using the top 20 band selection, both feature importance plots summarise the most influential bands within the band selection in the RF models, based on mean MDG across 10 runs (Figure 15). In the model that uses smoothed EnMAP reflectance (a), the highest-ranked predictors are mainly located in the VIS region, especially the blue, with fewer relevant bands in the VNIR and SWIR. By contrast, the model based on first-order derivatives (b) highlights RE and NIR bands as the strongest contributors, while still retaining several VIS and SWIR bands among the top-ranked features.

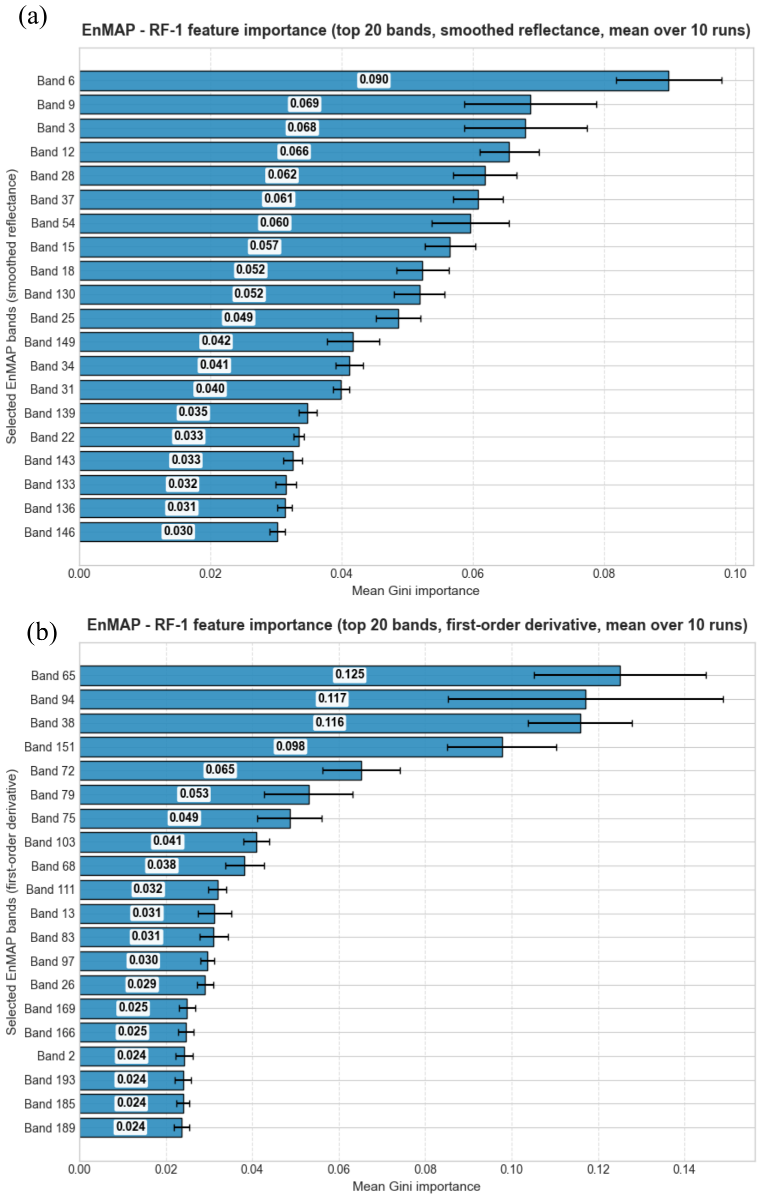


Figure 15: RF-1 feature importance of the 20 selected EnMAP bands in the Hannut region. (a) MDG for smoothed reflectance bands. (b) MDG for first-order derivative bands.

Figure 16 summarises the effect of using smoothed versus first-order derivative EnMAP spectra (each restricted to the 20 selected bands) on RF-1 classification performance across 10 validation runs. Across runs, OA range from 82.39-91.2% for the smoothed-spectra configuration and from 77.56-92.55% for the first-order derivatives, while $F1_{macro}$ range from 79.38-90.83% and 76.25-92.3%, respectively. Across both configurations, the F1 score of the WCC class is consistently higher than that of the NWCC class, indicating slightly better performance for WCC detection than for correctly identifying NWCC pixels. Mean OA, $F1_{macro}$ and class-wise F1 for WCC and NWCC are very similar between the two variables included in the classification, with average differences below 1 percentage point for all metrics. None of these differences are statistically significant at $\alpha = 0.05$ (all $p > 0.49$), indicating that, for RF-1, smoothed and first-order derivative EnMAP spectra perform equivalently when using a reduced set of 20 bands.

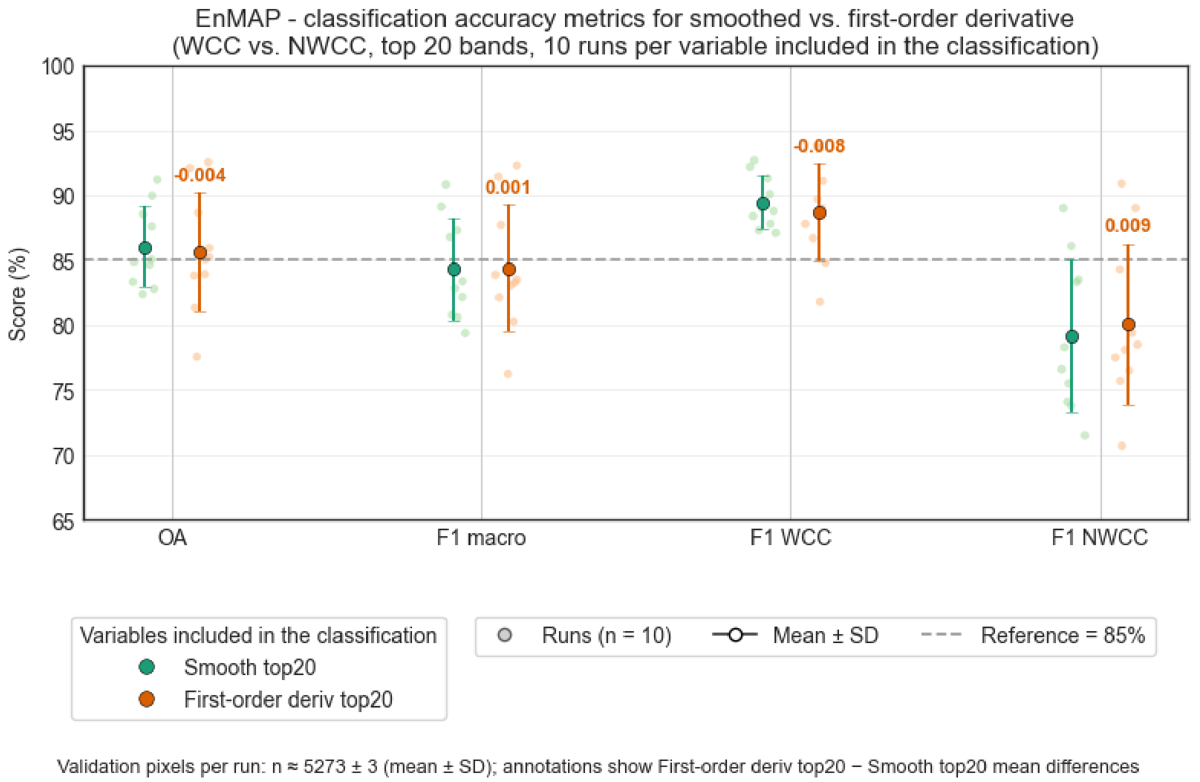


Figure 16: Classification accuracy of RF-1 with EnMAP input when using the 20 most informative smoothed reflectance bands (green) versus the 20 most informative first-order derivative bands (orange).

Appendix 8 extends this comparison showing the accuracy metrics of the classifications of smoothed reflectance using only top 20 bands, smoothed reflectance using all bands, first-order derivative using top 20 bands and first-order derivative using all bands, to assess whether band selection causes any loss of classification performance relative to using the full spectral input. Across all four metrics (OA, $F1_{macro}$, $F1_{WCC}$ and $F1_{NWCC}$), the differences between the top-20 and all-band configurations are consistently small, ranging from −0.003 to −0.029 percentage points. Paired t-tests confirmed that none of the top-20 vs. all-bands differences are statistically significant at $\alpha = 0.05$ for either the smoothed or the derivative input.

4.2. Winter cover crop detection using EnMAP & Sentinel-2 (Setup 2, RF-1)

Prior to comparing classification performance, the pixel-wise agreement between the spectrally resampled EnMAP bands and the corresponding S2 bands and VIs was assessed using Pearson correlation (Appendix 9). Correlations were strong across most band pairs and VIs — particularly for the red (B4, r = 0.76), RE3 (B7, r = 0.74), NIR (B8, r = 0.73), SWIR2 (B12, r = 0.7) and the VIs NDVI (r = 0.82), NDRE (r = 0.80), NDWI (r = 0.81) and BSI (r = 0.81). The weakest agreement was observed for B5 RE1 (r = 0.35) and B11 SWIR1 (r = 0.48).

Beyond pixel-wise correlation per band, the shape of the SRF of each band determines how energy is weighted across the band's spectral range, and differences in SRF shape can affect the effective signal recorded even when band centres are nominally aligned. Figure 17 illustrates these differences between native S2 and EnMAP-simulated SRFs. In the VNIR domain, the native S2 SRFs exhibit broad, smoothly tapering shapes with gradual flanks on both sides of the band centre, whereas the EnMAP-simulated SRFs show narrower, more sharply bounded profiles with steeper edges at the spectral limits of each band. In the SWIR domain, both sensors show broader SRFs, but the EnMAP-simulated profiles remain more rectangular in shape; the number of EnMAP channels contributing to each simulated SWIR band is also visibly smaller than in the VNIR.

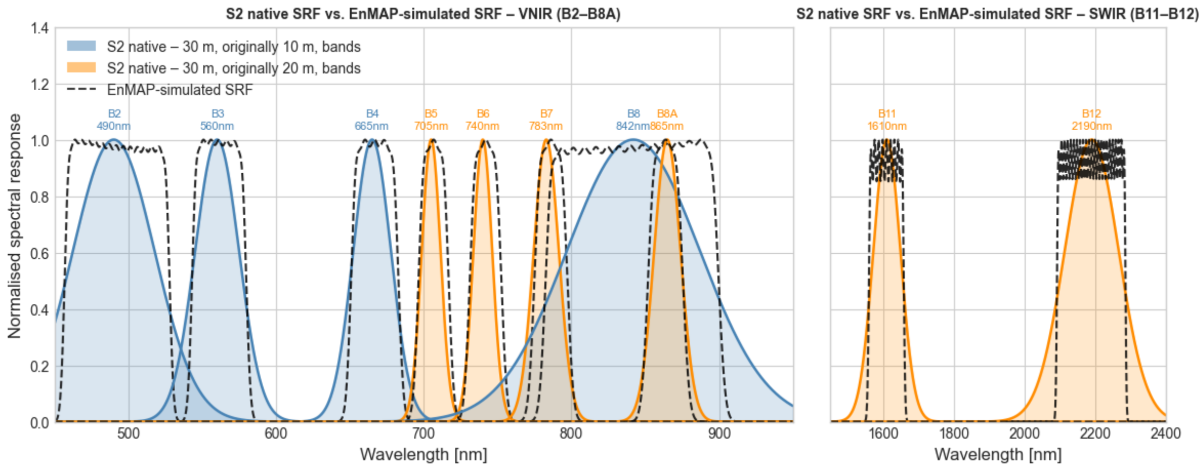


Figure 17: Normalised SRF of native S2 MSI bands and EnMAP-simulated S2-equivalent bands.

Figure 18 compares mean EnMAP (a) and S2 (b) spectral signatures for WCC and NWCC pixels. Both sensors show very similar overall spectral shapes, with low reflectance in VIS, a marked RE rise between 700-783 nm, and highest values around 865 nm representing the NIR. For EnMAP, WCC and NWCC exhibit very similar spectral shapes, with only small differences in reflectance magnitude across the bands. The mean curves and ranges of variation overlap strongly for all wavelengths. The corresponding S2 curves exhibit a similar pattern, but with generally higher absolute reflectance and greater variability, particularly in the NIR and SWIR bands. Across both sensors, the largest separability between WCC and NWCC is observed around the RE/NIR plateau (approximately 740-865 nm).

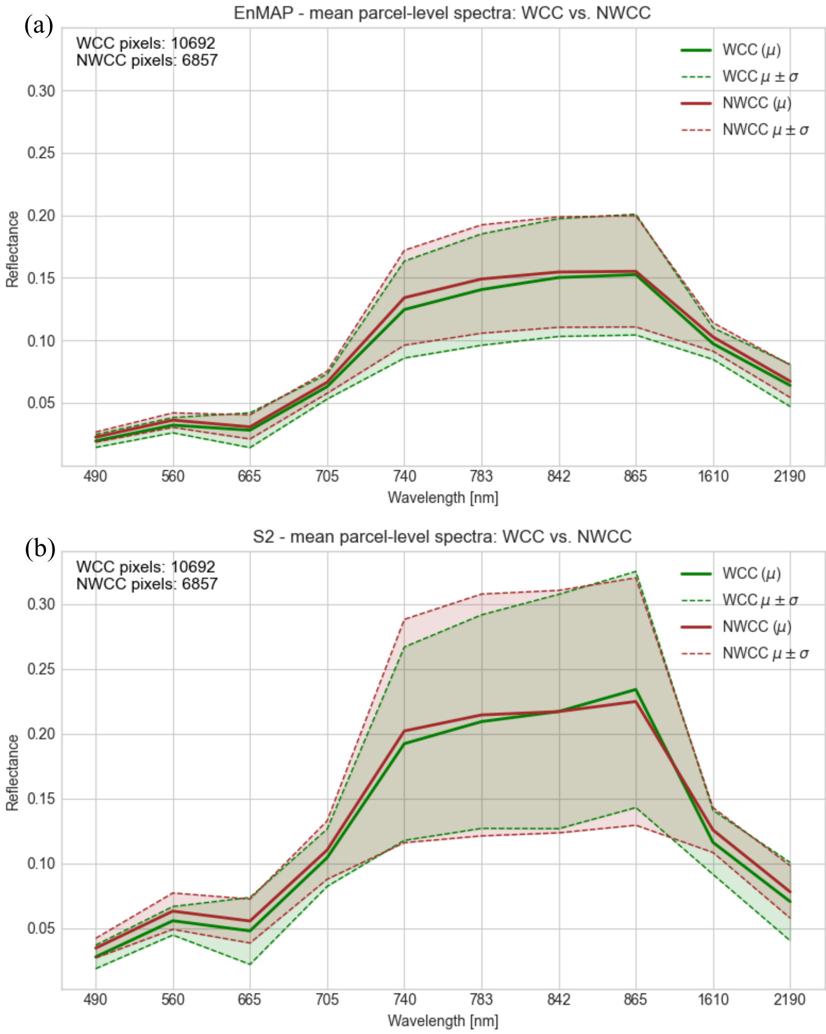


Figure 18: Mean EnMAP-simulated S2 and S2 parcel-level spectra for WCC and NWCC within the Hannut region. (a) EnMAP-simulated S2 reflectance spectra. (b) S2 reflectance spectra.

For both EnMAP-simulated (a) and S2 (b) predictors, the ranking of important variables was very stable across the 10 runs, with the blue band (B2) consistently dominating the MDG scores, followed by NDVI, NDRE and BSI (Figure 19). Bands in the VIS and VIs in the VIS and RE/NIR range contributed most strongly to WCC–NWCC separation, whereas single SWIR and RE/NIR bands showed lower but still non-negligible importance.

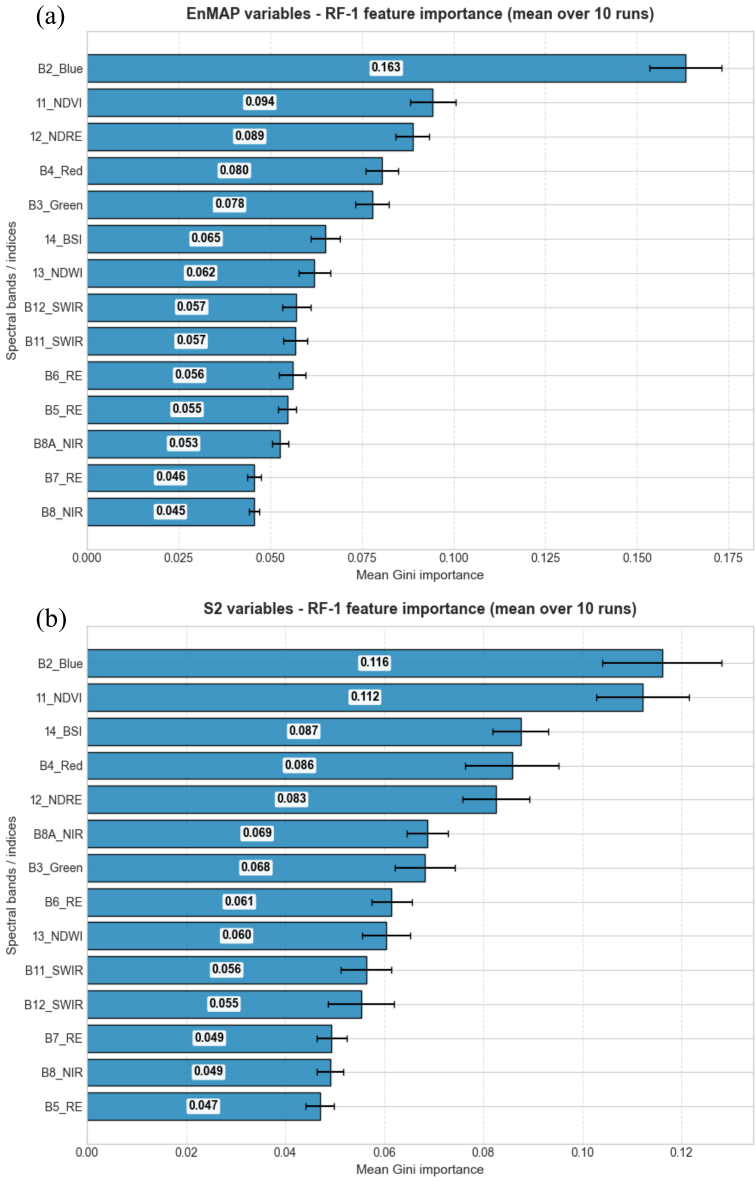


Figure 19: RF-1 feature importance of the Setup 2 variables in the Hannut region. (a) MDG for the EnMAP-simulated S2 bands and VIs. (b) MDG for the corresponding S2 bands and VIs.

Using the same calibration-validation splits, EnMAP systematically outperforms S2 for all considered accuracy metrics (Figure 20). Across the 10 runs, EnMAP achieves a mean OA of 86.84% ± 3.99 compared to 80.72% ± 6.07 for S2, and a mean $F1_{macro}$ of 85.5% ± 4.68 versus 78.6% ± 6.74. The class-specific F1 scores show the same pattern: for WCC, EnMAP reached 89.9% compared with 85.2% for S2, and for NWCC 81.1% versus 72%. For both sensors, the WCC class attains higher F1 scores than the NWCC class, indicating that WCC pixels are detected more reliably than NWCC pixels. EnMAP outperforms S2 in all accuracy metrics (paired t-test, $p < 0.05$), with highly significant differences ($p < 0.01$) for OA, $F1_{macro}$, and F1 NWCC (Appendix 10).

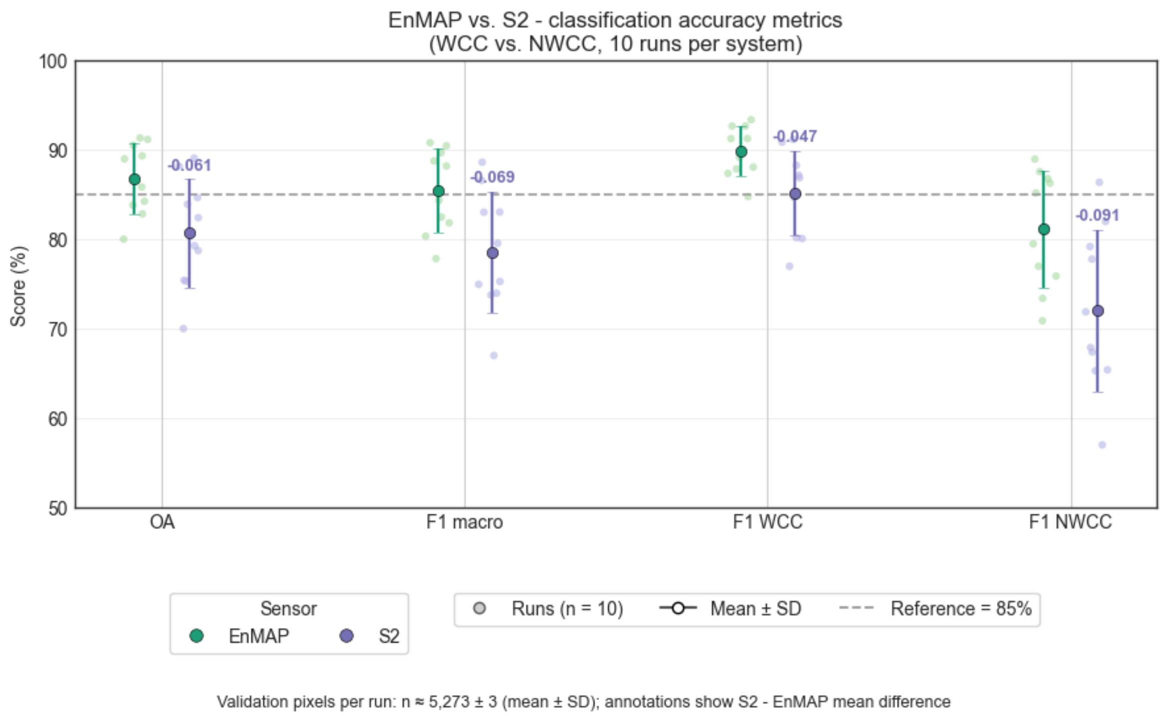


Figure 20: Classification accuracy of Setup 2 RF-1 when using the EnMAP simulated S2-bands (green) versus the corresponding S2 bands (purple).

4.3. Winter cover crop detection using Sentinel-2 monthly composites (Setup 3, RF-1)

Figure 21 shows the mean S2 spectral signatures of WCC and NWCC pixels for the November composite. Both classes follow a similar overall shape, with low reflectance in VIS, a strong increase around the RE and NIR bands, and a decrease towards the SWIR region. Across wavelengths in RE and NIR, WCC exhibits clearly higher reflectance than NWCC, and the ranges of variation are largely separated.

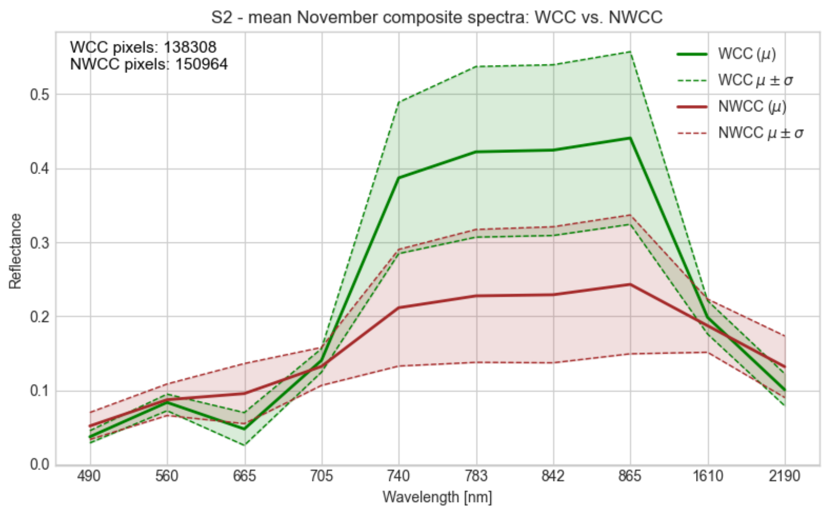


Figure 21: Mean S2 November median composite parcel-level spectra for WCC and NWCC within the full in-situ AOI.

For the November S2 composite, the RF feature importance analysis shows that the most informative predictors for RF-1 are the RE and NIR bands, followed by SWIR bands and the derived VIs NDVI and

NDWI. Together, these variables clearly dominate the MDG, whereas VIS bands and the LAI contribute substantially less (Figure 22).

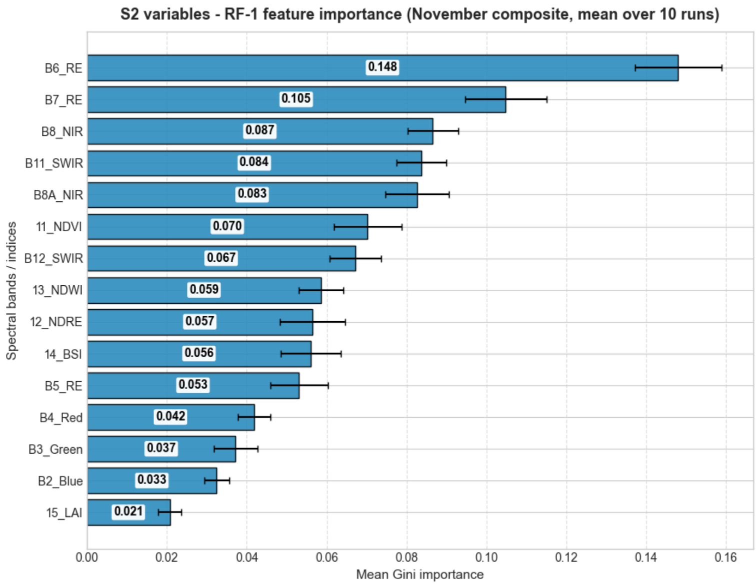


Figure 22: RF-1 feature importance of the S2 November monthly composite bands within the full in-situ AOI.

The RF-1 for Setup 3 reaches mean OA and F1 between about 76-86 % across months, with standard deviations around 3.2-4.6 %. Classswise F1 scores show comparable values for WCC and NWCC. The highest OAs are observed for March, November, January, October and February (Figure 23).

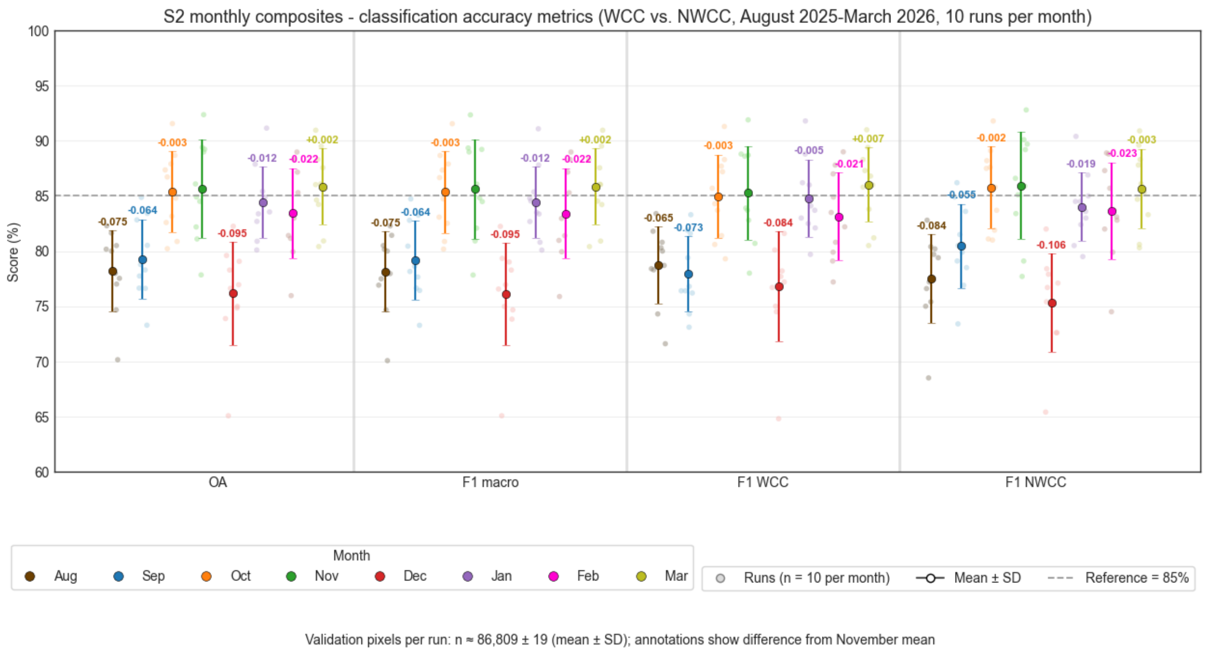


Figure 23: Classification accuracy of RF-1 in Setup 3 when using monthly median S2 composites from August 2025 to March 2026.

To contextualise the month-to-month variation in classification performance, the proportion of 'nodata' pixels inside LPIS parcels was calculated for each monthly RF-1 classification after applying the parcel-based majority filter (Appendix 11). Across the study period, the fraction of 'nodata' pixels remained very low for most months, with values of 0 % for August 2025, November 2025 and March

2026, below 0.01 % for September 2025 and February 2026, and around 1.0–1.7 % for October and December 2025. In contrast, January 2026 showed a higher proportion of 'nodata' within LPIS parcels (10.9 %). Furthermore, the parcel-level NDVI trajectories of WCC and NWCC parcels were examined across the full study period (Figure 24). In August, WCC parcels show a low mean NDVI value (~0.2), while NWCC parcels retain higher NDVI. WCC parcels show a rapid increase in NDVI from September onwards. The spectral separation between the two classes is greatest in November, coinciding with the peak mean NDVI for WCC. From December onwards, the WCC mean NDVI declines, while NWCC mean NDVI remains relatively stable. From January to March, the trajectories diverge again, with WCC maintaining a moderate greenness signal (mean NDVI ~0.4) while NWCC parcels show more variable behaviour but generally an increasing mean NDVI.

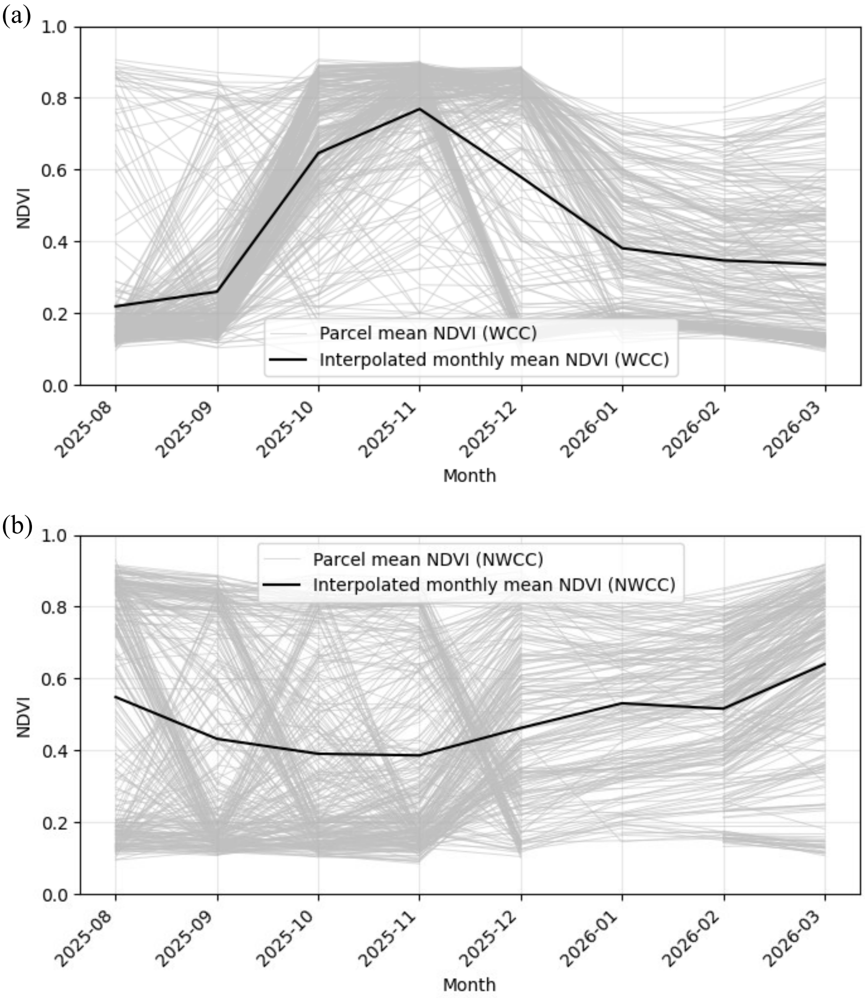
(a)
(b)


Figure 24: Parcel-level NDVI trajectories from S2 monthly composites in Setup 3 for (a) WCC and (b) NWCC parcels.

To assess whether there were differences in classification performance between months, paired t-tests were applied to the ten repeated runs per month, separately for OA, $F1_{macro}$ and the class-specific F1 scores for WCC and NWCC. Comparing August/September and November, November showed significantly higher performance for all metrics (all $p \leq 0.006$). Comparing November and December revealed consistent and significant degradation in December (all $p < 0.001$), confirming that December is less suitable for WCC mapping. Conversely, no statistically significant differences were identified between November and October, January, February or March for any metric (all $p \geq 0.05$), suggesting that these months exhibited performance comparable to that of November. Although March achieved slightly higher OA, $F1_{macro}$ and $F1_{WCC}$, November was selected as the reference month for subsequent RF-2 because its performance was statistically indistinguishable from March, it coincided with the main field campaign (Section 3.2.4.), and it was available from the outset of the study, whereas the March composite became available only later.

4.4. Cross-setup overview of winter cover crop detection (RF-1)

Across all three setups, WCC presence was mapped with high and stable accuracy. In Setup 1, EnMAP with 20 selected bands reached mean OA and $F1_{macro}$ consistently above 80%, with no statistically significant difference between smoothed reflectance and first-order derivative inputs (all $p > 0.49$), indicating that both spectral representations are equivalent for this binary task. In Setup 2, EnMAP systematically outperformed S2 under harmonised conditions for all accuracy metrics, with gains most pronounced for OA, $F1_{macro}$ and F1 NWCC (all $p < 0.05$); across both sensors. For Setups 1 and 2 WCC pixels were detected more reliably than NWCC pixels. In Setup 3, S2 monthly composites achieved OA and F1 between approximately 76% and 86%, with November, October, January, February and March performing at statistically comparable levels, while August, September and December showed significant degradation.

4.5. Winter cover crop mixture type detection using EnMAP (Setup 1, RF-2)

Figure 25 shows the mean EnMAP spectra for pure and mixed WCC pixels using all available bands, visualised as smoothed reflectance (a) and its first-oder derivative (b). The smoothed reflectance curves indicate broadly similar spectral shapes for both mixture types across the VNIR-SWIR range, but mixed stands tend to exhibit higher reflectance, especially around the RE and NIR plateau, and their range of variation are slightly wider. The first-derivative spectra emphasise subtle differences in the strength and position of several absorption and RE features between pure and mixed canopies, although these differences remain small relative to the within-class variability. For the RF-2 using band selection only, the final feature sets again consisted of 20 EnMAP bands: for smoothed reflectance, bands 44, 47, 50, 59, 63, 68, 71, 74, 77, 82, 87, 94, 97, 101, 125, 152, 156, 159, 163 and 168 were retained, whereas for the first-order derivative input, bands 8, 12, 29, 33, 36, 39, 42, 47, 52, 55, 61, 79, 83, 93, 103, 107, 115, 145, 159 and 199 were selected.

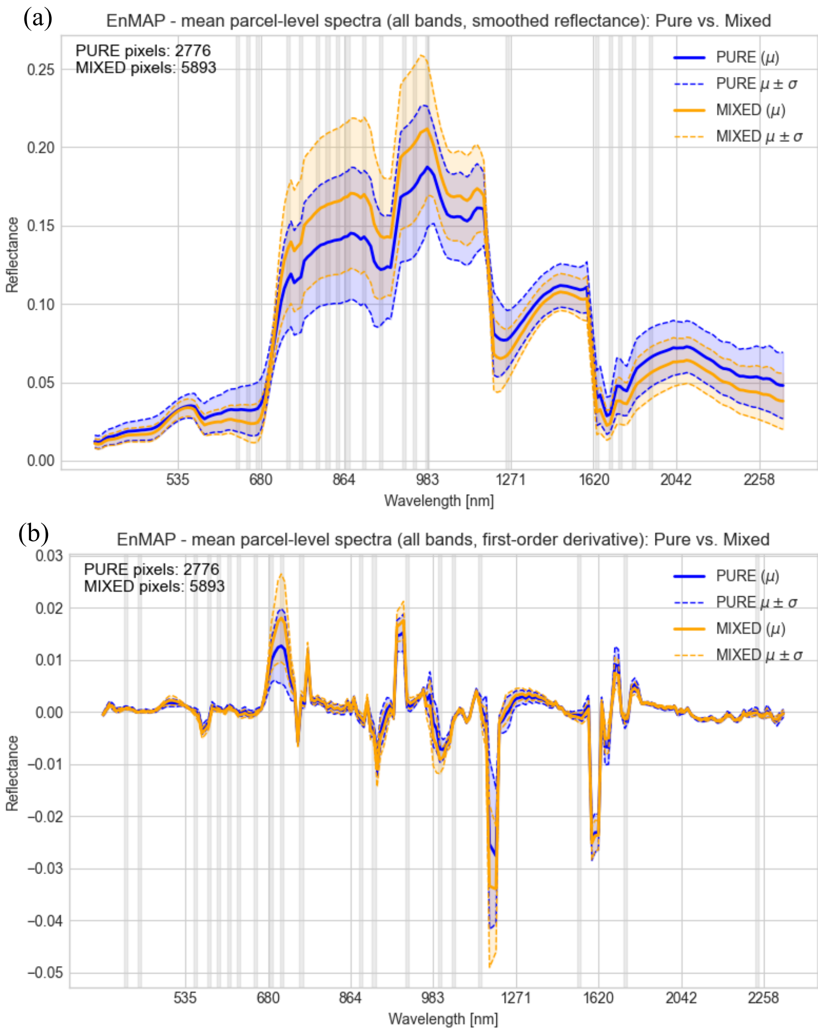


Figure 25: Mean EnMAP parcel-level spectra for pure and mixed within the Hannut region. (a) Smoothed reflectance spectra across all bands. (b) First-order derivative spectra across all bands.

For the top 20 band selection classifications, the RF-2 feature importance plots summarise the most influential EnMAP bands for separating pure and mixed WCC pixels, based on mean MDG across 10 runs (Figure 26). In the model using smoothed reflectance (a), the highest contributions are concentrated in the red and early RE/NIR range (e.g. bands 50, 47 and 44), with slight contribution of several SWIR bands (e.g. band 159). In contrast, the model based on first-order derivatives (b) emphasises a broader set of VIS, RE and SWIR bands among the top-ranked features (e.g. bands 47, 55, 52 and 93).

(a)

(b)

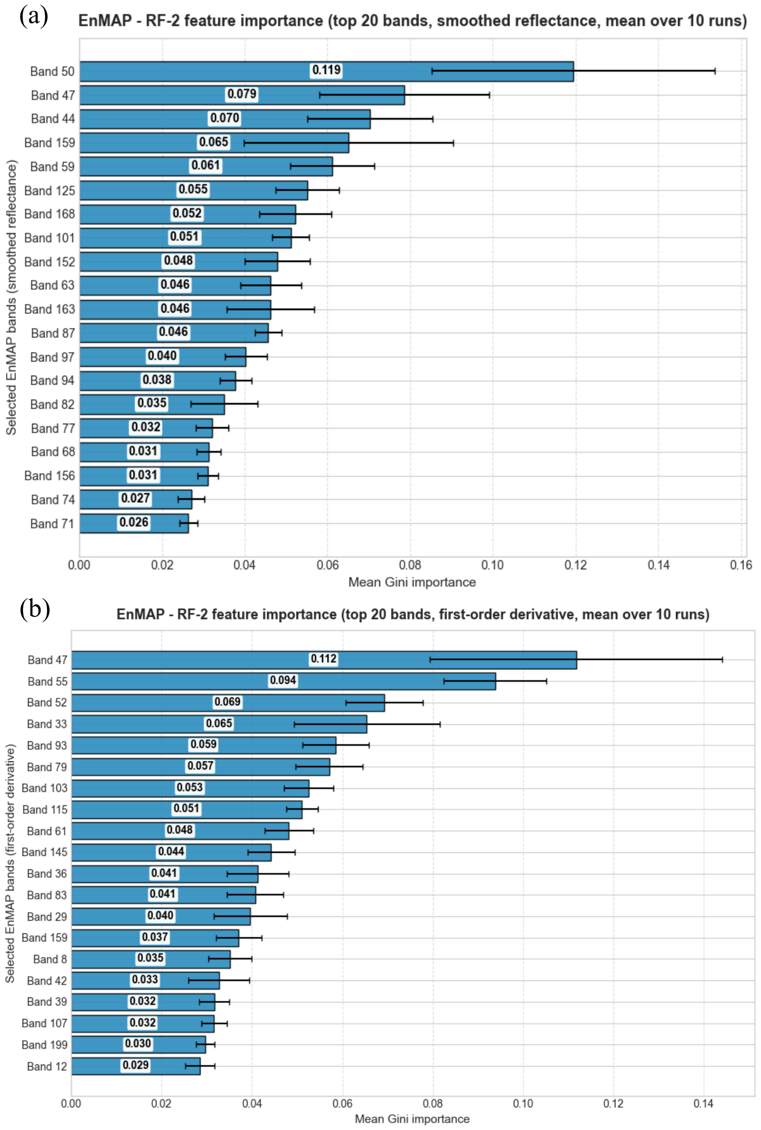


Figure 26: RF feature importance of the 20 selected EnMAP bands for RF-2 in the Hannut region. (a) MDG for smoothed reflectance bands. (b) MDG for first-order derivative bands.

Figure 27 summarises the effect of using 20 bands of the smoothed versus the first-order derivative EnMAP spectra on RF-2 classification performance for discriminating pure and mixed WCC pixels across 10 validation runs. Across runs, OA ranged from 63.5-75.14% for the smoothed spectra and from 58.66-82.96% for the first-order derivatives, while $F1_{macro}$ ranged from 51.0-68.76% and 50.91-78.62%, respectively. Mean OA and $F1_{macro}$ were slightly higher for the first-order derivatives, whereas the class-wise F1 scores revealed a trade-off: first-order derivatives improved F1 for pure pixels but tended to reduce F1 for mixed pixels compared to smoothed reflectance. For both smoothed reflectance and first-order derivative, however, F1 for pure pixels was substantially lower than F1 for mixed pixels. None of the differences between the smoothed and first-order derivative configurations were statistically significant at α = 0.05 (all p ≥ 0.20), indicating that, for RF-2, both spectral representations achieved comparable OA, $F1_{macro}$ and class-wise F1 scores for pure and mixed pixels.

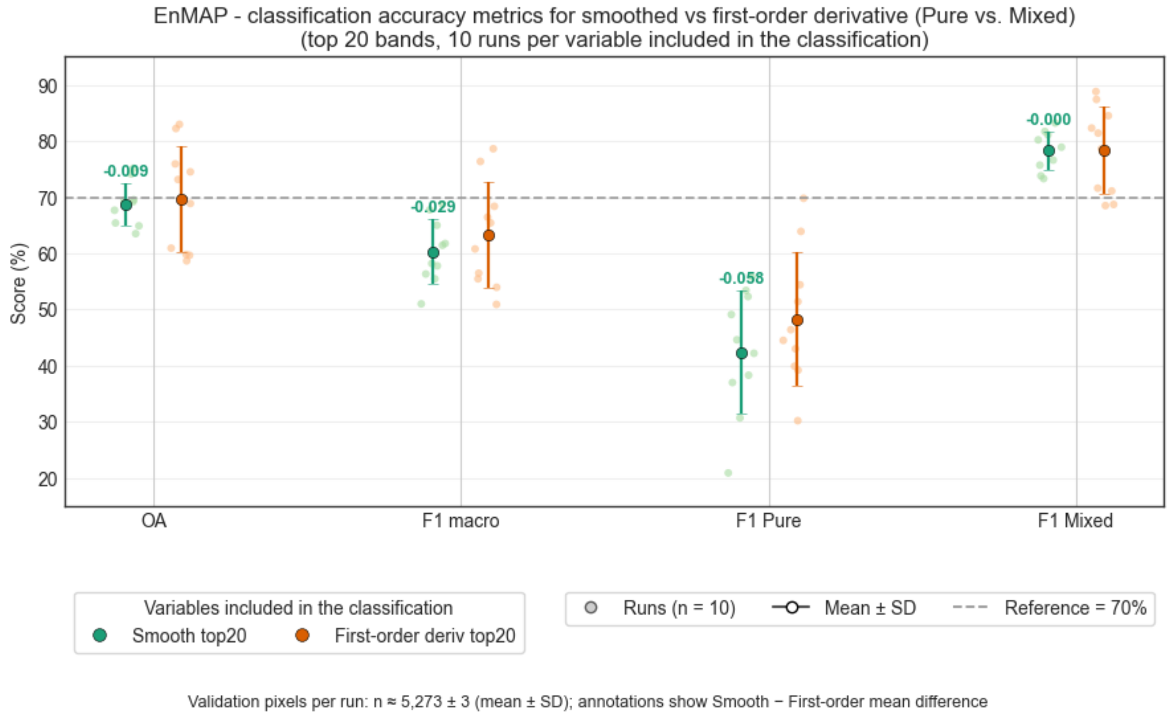


Figure 27: Classification accuracy of EnMAP RF-2 when using the 20 most informative smoothed reflectance bands (green) versus first-order derivative bands (orange).

4.6. Winter cover crop mixture type detection using EnMAP & Sentinel-2 (Setup 2, RF-2)

Figure 28 compares mean EnMAP (a) and S2 (b) parcel-level spectral signatures for pure and mixed WCC parcels. Across both sensors, pure and mixed parcels share very similar spectral shapes, but mixed parcels consistently display higher reflectance than pure parcels from the RE into the NIR and SWIR, together with slightly wider ranges of variation. The EnMAP and S2 curves follow the same trend with smoother mean spectra, smaller range of variation around the mean in the NIR and SWIR domains for S2.

(a)

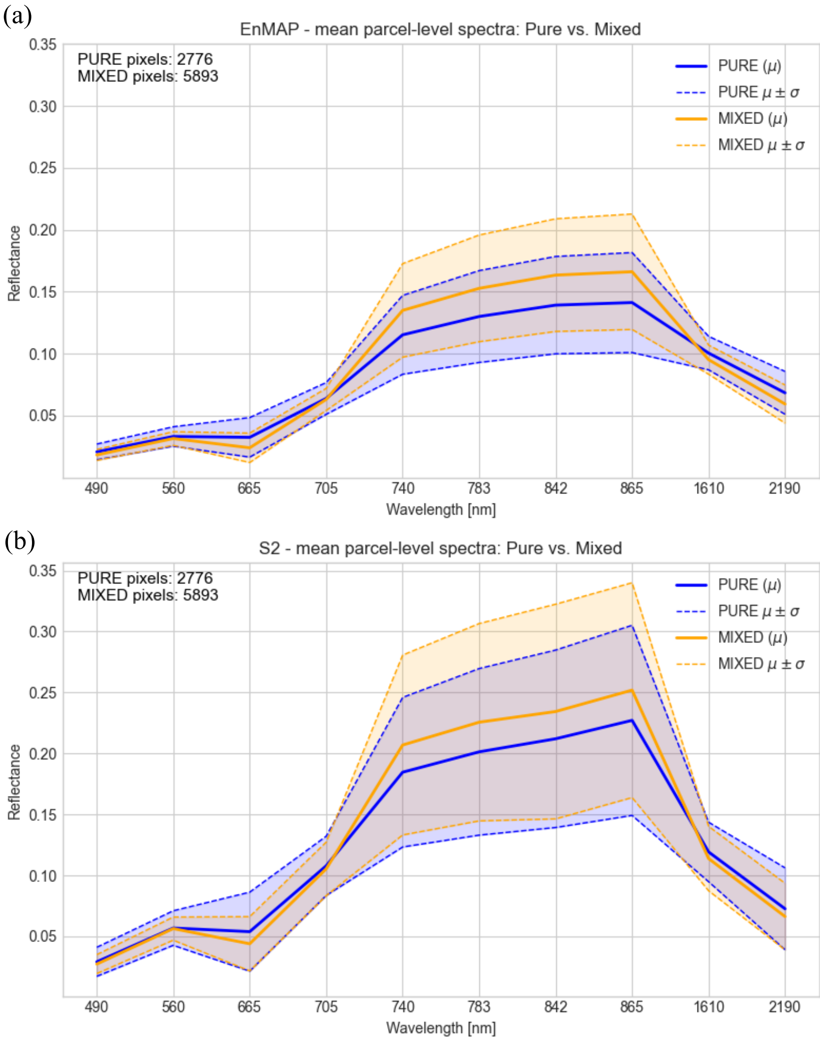


Figure 28: Mean EnMAP-simulated S2 and S2 parcel-level spectra for pure and mixed within the Hannut region. (a) EnMAP-simulated S2 reflectance spectra. (b) S2 reflectance spectra.

For both EnMAP- (a) and S2-based (b) predictors, the ranking of important variables in RF-2 are highly consistent across the 10 runs, with the red band clearly dominating the MDG scores, followed by NDRE, band 12 (SWIR) and NDVI (Figure 29). In both systems, VIS bands and VIs in the VIS-RE/NIR region contribute most strongly to separating pure and mixed WCC parcels. In contrast, individual RE, NIR and SWIR bands show slightly lower but still non-negligible importance.

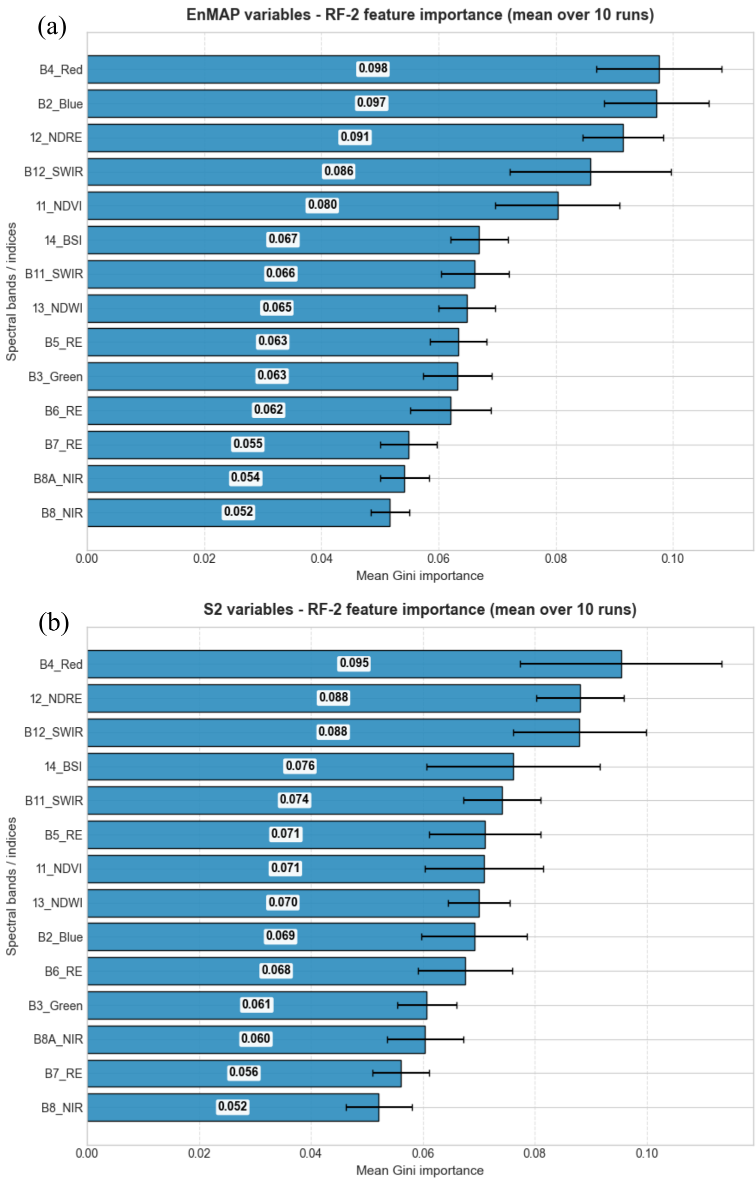


Figure 29: RF-2 feature importance of the Setup 2 variables in the Hannut region. (a) MDG for the EnMAP-simulated S2 bands and VIs. (b) MDG for the corresponding S2 bands and VIs.

Using identical calibration-validation splits, EnMAP and S2 achieve broadly similar classification performance for discriminating pure and mixed WCC parcels (Figure 30). Across the 10 runs, mean OA and $F1_{macro}$ are slightly higher for EnMAP (OA: 70.23% ± 9.61, $F1_{macro}$: 65.55% ± 10.41) than for S2 (OA: 67.8% ± 7.53, $F1_{macro}$: 62.81% ± 8.35), and the class-wise F1 scores show the same tendency, with marginally better F1 for both pure and mixed parcels under the EnMAP classification. For both sensor classifications $F1_{Pure}$ is markedly lower than $F1_{Mixed}$. Paired t-tests indicate that none of the accuracy metrics differences between EnMAP and S2 were statistically significant at $\alpha = 0.05$ (all $p \geq 0.23$) (Appendix 12), suggesting that, for RF-2 in this setup, both sensors perform comparably when using the selected bands and VIs.

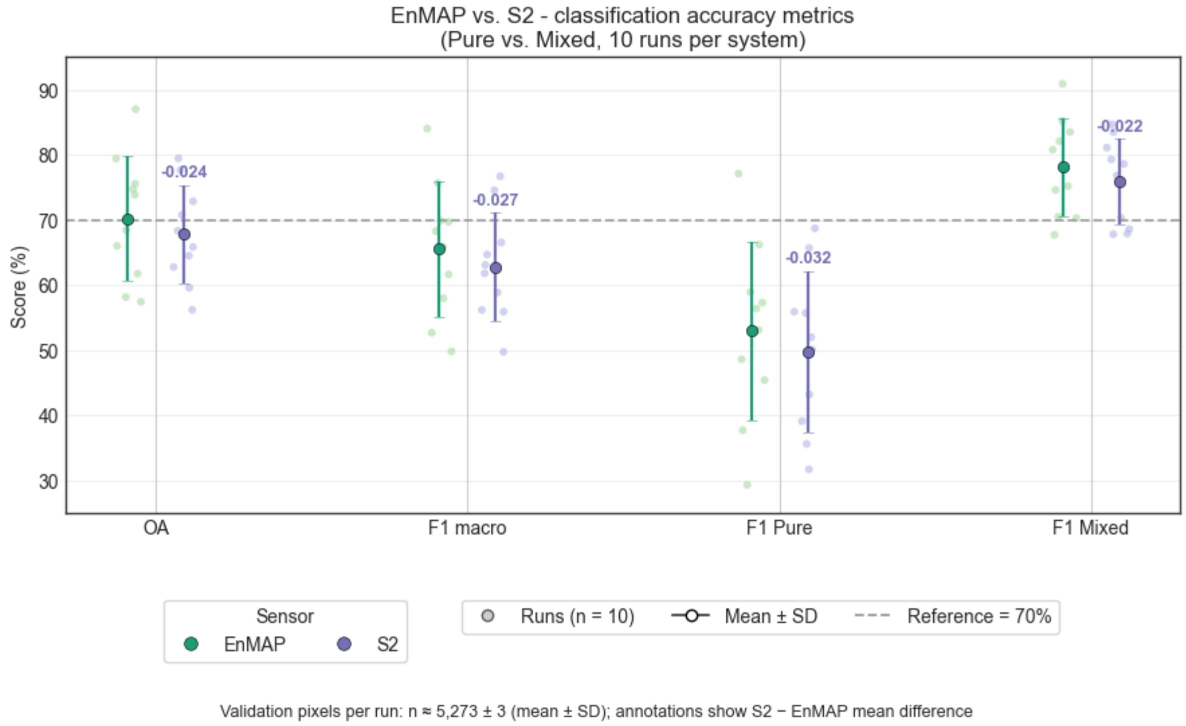


Figure 30: Classification accuracy of RF-2 in Setup 2 when using EnMAP-simulated S2 bands (green) versus the corresponding S2 bands (purple).

In addition to hard class predictions, the RF classifier outputs a predicted class probability (P) for each pixel, defined as the proportion of decision trees in the ensemble that voted for a given class (ranging from 0 to 1). A value close to 1 indicates high model confidence in the assigned class, while values near 0.5 reflect maximum uncertainty. To complement the lower F1 class-wise values for pure than for mixed in RF-2 in all setups, the pixel-level class probability distributions were examined for both RF-1 and RF-2 in Setup 2 as an illustrative example (Figure 31). For RF-1 (WCC vs. NWCC), the distributions of P(WCC) and P(NWCC) are strongly polarised towards 0 and 1, with most pixels receiving a predicted probability close to the class boundaries. For RF-2 (pure vs. mixed), the distributions show a markedly different pattern: both P(Pure) and P(Mixed) are broadly spread across the full probability range, with substantially fewer pixels assigned near-certain probabilities.

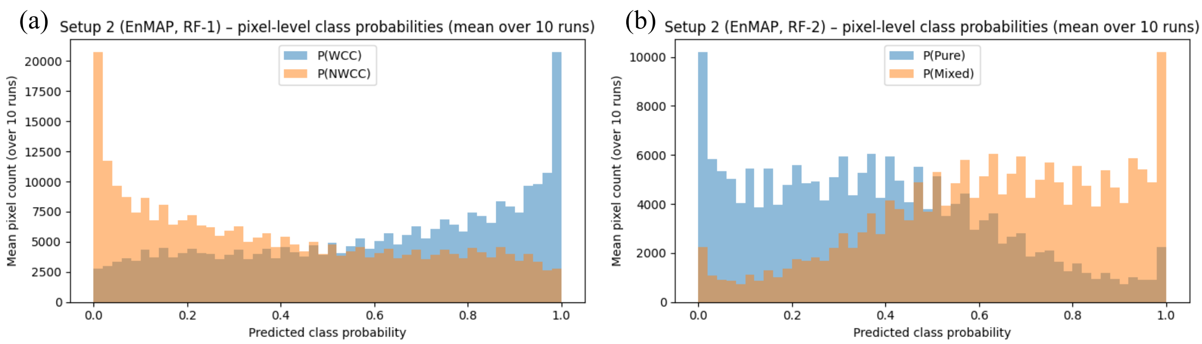


Figure 31: Pixel-level class probability distributions for the EnMAP RF-1 and RF-2 models in Setup 2 (mean over 10 runs). (a) RF-1: WCC vs. NWCC classification. (b) RF-2: pure vs. mixed classification.

4.7. Winter cover crop mixture type detection using November Sentinel-2 monthly composite (Setup 3, RF-2)

Figure 32 shows the mean S2 composite spectra for pure and mixed WCC pixels in November. Both classes follow a very similar vegetation-type shape. Across the RE and NIR wavelengths, the pure parcels exhibit clearly higher reflectance than mixed parcels, as well as a greater range of variation of both. The two classes show comparable reflectance in the SWIR region.

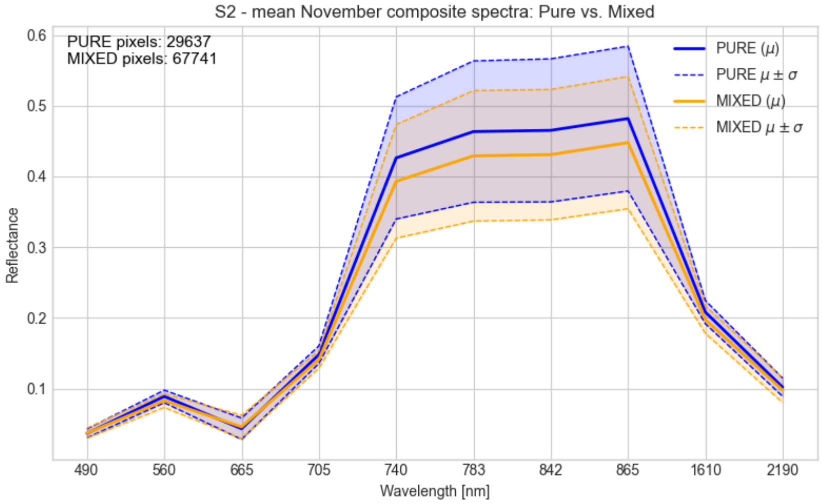


Figure 32: Mean S2 November median composite parcel-level spectra for pure and mixed within the full in-situ AOI.

For the November S2 composite, the RF-2 feature importance analysis indicates that the green band (B3) is by far the most informative predictor, followed by the SWIR bands B12 and B11 and the NDRE (Figure 33). RE and NIR bands, as well as the VIs NDVI and NDWI also contribute substantially to separating pure and mixed parcels, whereas the other VIS bands (B2 & B4), the BSI and the LAI have comparatively lower MDG values and play only a secondary role in the RF-2 model.

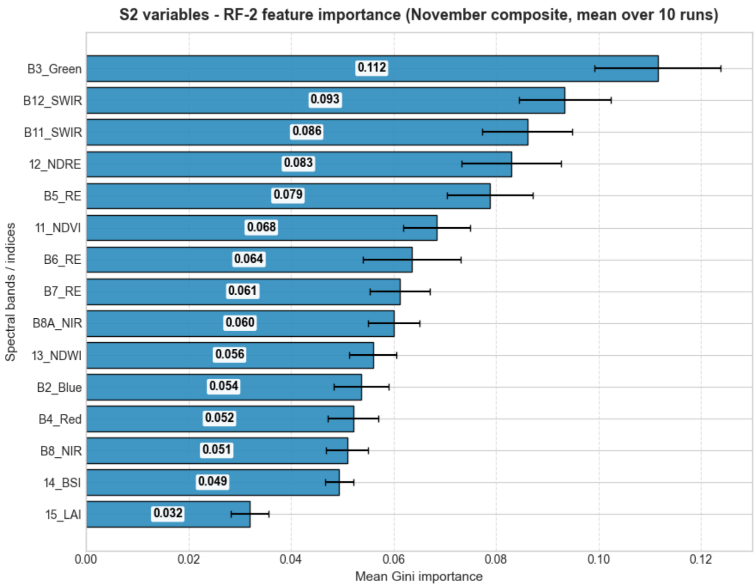


Figure 33: RF-2 feature importance of the S2 November monthly composite bands within the full in-situ AOI.

Using the November S2 median composite, RF-2 achieves a mean OA of about 72.76 %, with relatively low variability (± 4.05) across the ten runs (Figure 34). The mean $F1_{macro}$ score is lower at 64.41 %, reflecting an imbalance in class-wise performance (± 7.11). While the classifier attains only moderate F1 scores for the pure class (mean ≈ 47.4 %), its performance for the mixed class is consistently high, with mean F1 scores around 81.5 % and little run-to-run variation.

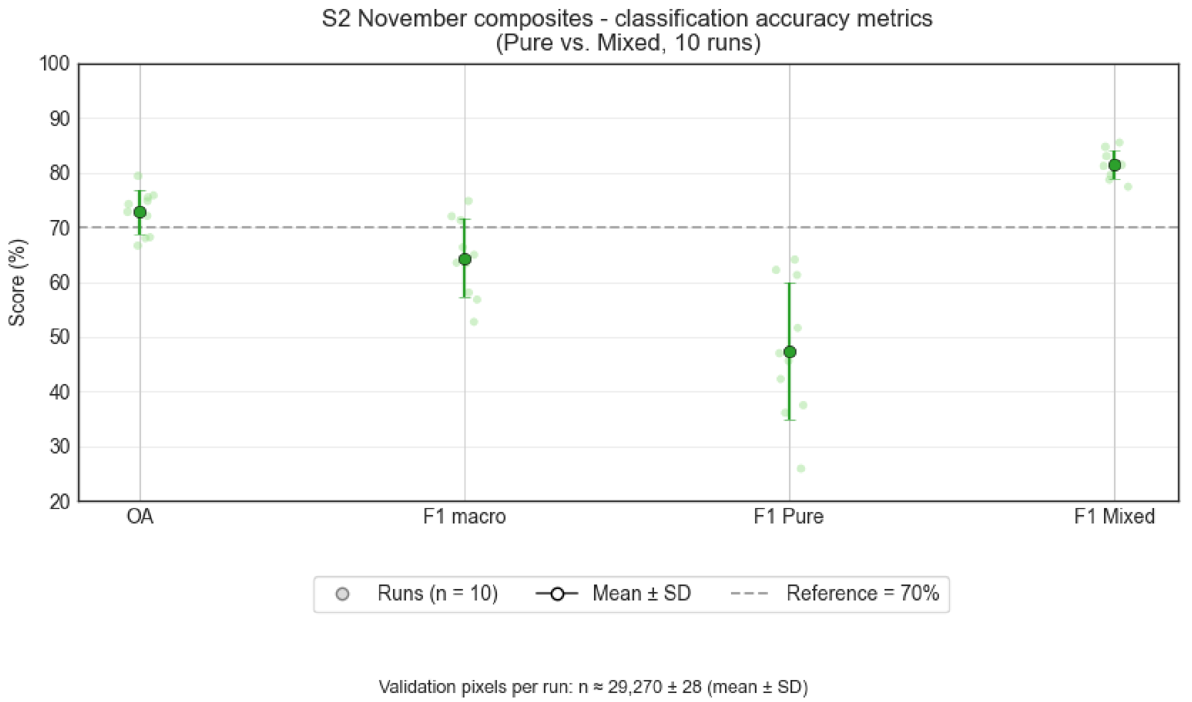


Figure 34: Classification accuracy of RF-2 in Setup 3 when using November median S2 composite.

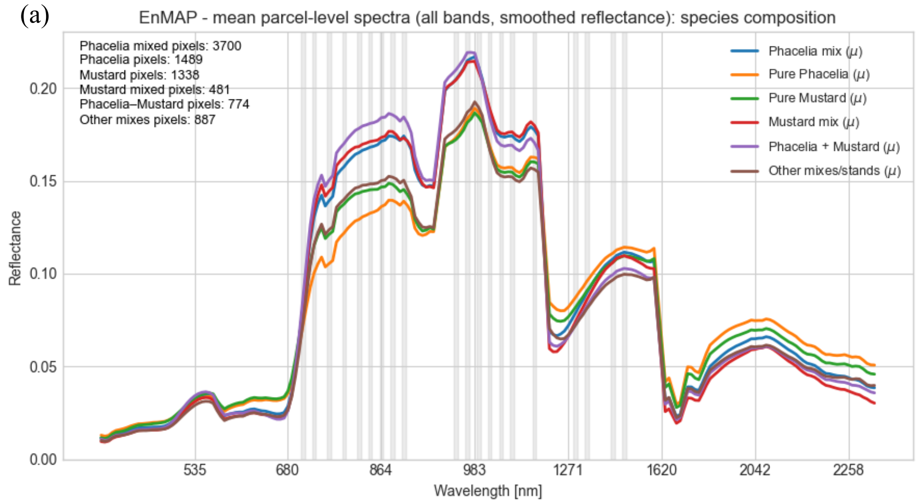
4.8. Cross-setup overview of winter cover crop mixture type detection (RF-2)

Across all three setups, distinguishing pure from mixed WCC stands proved consistently more challenging than WCC presence detection, with OA ranging from approximately 60–80% and $F1_{macro}$ values remaining moderate. A recurrent pattern across all setups and both sensors is that mixed stands were classified more reliably than pure stands, as reflected by systematically higher F1 for the mixed class. In Setup 1, first-order derivatives yielded slightly higher mean OA and $F1_{macro}$ than smoothed reflectance, but shifted the class-wise balance by improving F1 for pure pixels at the expense of mixed pixels; none of these differences were statistically significant (all $p \geq 0.20$), indicating that both spectral representations are equivalent for this task under the 20-band EnMAP configuration. In Setup 2, EnMAP and S2 performed comparably under harmonised conditions, with EnMAP achieving marginally higher OA, $F1_{macro}$ and class-wise F1 scores that did not reach statistical significance (all $p \geq 0.23$). In Setup 3, the November S2 composite achieved a mean OA of approximately 73% and a mean $F1_{macro}$ of 64%, with consistently high F1 for mixed stands (≈ 82%) but only moderate performance for pure stands (≈ 47%).

4.9. Winter cover crop species composition using EnMAP (Setup 1, RF-3)

Figure 35 shows the mean EnMAP spectra for the different WCC canopy compositions using all available bands, visualised as smoothed reflectance (a) and its first-order derivative (b). The smoothed reflectance curves reveal very similar spectral shapes across all mixtures over the VNIR–SWIR range, with systematic but moderate differences in reflectance magnitude: ‘Phacelia mix’ and ‘Mustard mix’, as well as ‘Phacelia + Mustard’ tend to exhibit the highest reflectance on the RE/NIR plateau and slightly lower reflectance in the SWIR, whereas ‘Pure Mustard’ and ‘Pure Phacelia’ and ‘Other mixes’ show somewhat reduced NIR reflectance and slightly higher SWIR levels. The corresponding first-derivative spectra emphasise narrow absorption and RE features around 680–740 nm and in the NIR but again show only subtle differences between canopy types that remain small relative to the within-class variability. For the classifications based on band selection only, global feature sets of 20 EnMAP bands were derived by combining class-specific separability rankings and enforcing a minimum spacing in band index: for smoothed reflectance, bands 55, 58, 62, 66, 70, 73, 76, 79, 82, 96, 99, 102, 105, 108, 111, 117, 128, 131, 138 and 141 were retained, whereas for the first-order derivative input, bands 8, 12, 19, 31, 44, 47, 50, 53, 62, 79, 82, 98, 103, 114, 138, 158, 170, 181, 184 and 199 were selected.

(a)



(b)

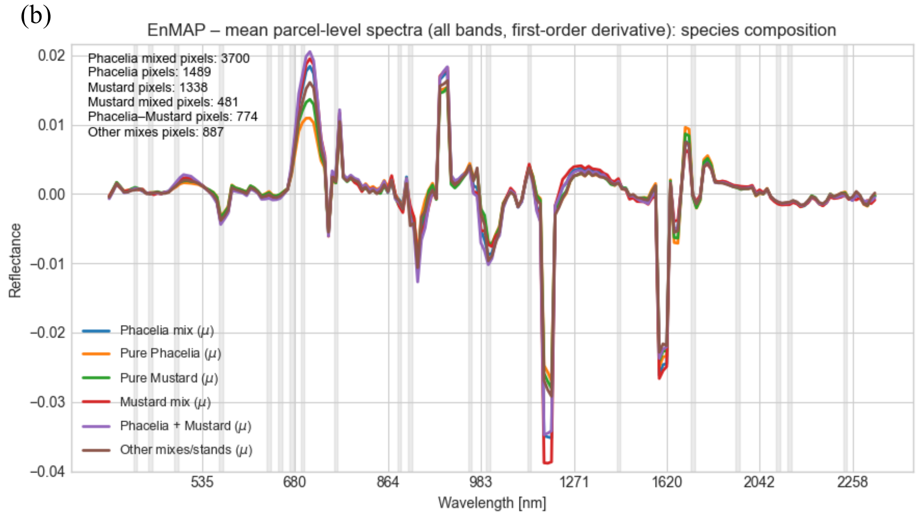
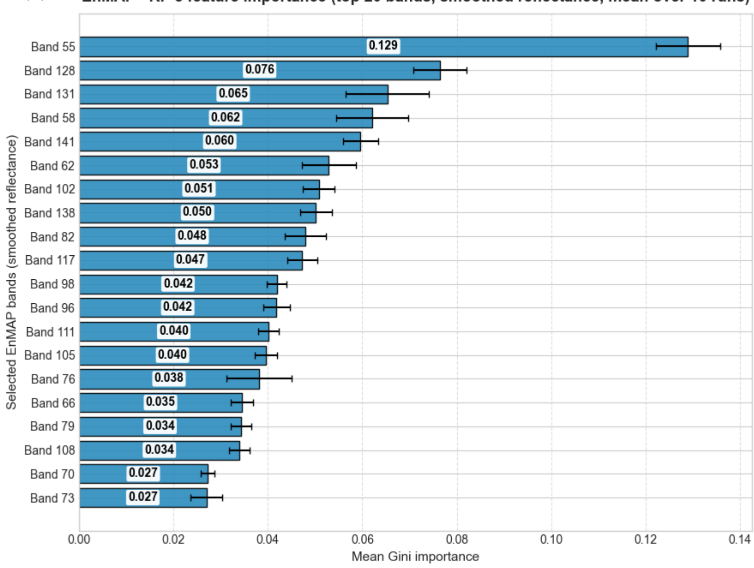


Figure 35: Mean EnMAP parcel-level spectra for species compositions within the Hannut region. (a) Smoothed reflectance spectra across all bands. (b) First-order derivative spectra across all bands.

For the top-20 band selection classifications, the RF-3 feature importance plots summarise the most influential EnMAP bands for separating WCC species composition pixels, based on mean MDG across 10 runs (Figure 36). In the model using smoothed reflectance (a), importance is strongly concentrated in early RE/NIR bands, with band 55 clearly dominating and bands 128, 131 (both SWIR) and 58 forming a second tier of predictors. In contrast, the model based on first-order derivatives (b) distributes importance more evenly across a broader set of VIS, RE and SWIR bands, with bands 50, 53 and 103 showing the highest contributions, followed by several additional bands in the 30-80 and 100-140 regions that still attain moderate MDG values.

(a)



(b)

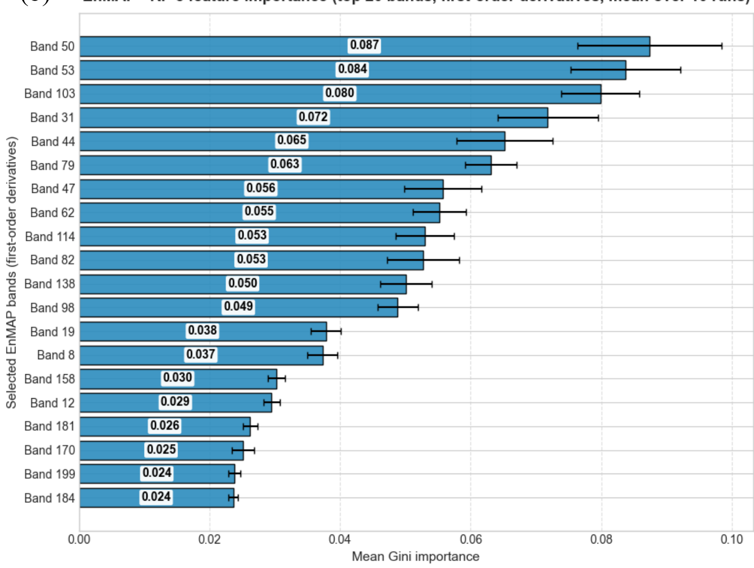


Figure 36: RF-3 feature importance of the 20 selected EnMAP bands in the Hannut region. (a) MDG for smoothed reflectance bands. (b) MDG for first-order derivative bands.

Figure 37 summarises the effect of using the 20 most important smoothed versus first-order derivative EnMAP bands on RF-3 classification performance for distinguishing WCC species compositions across 10 validation runs. Across runs, OA range from 26.55-53.15 % for the first-order derivatives and from 29.7-51.41 % for the smoothed bands, with very similar mean OA around 39.48-39.64 %. $F1_{macro}$ cover a wider range, from 11.14-35.65 % for the first-order derivatives and from 12.17-28.37 % for the smoothed configuration, with mean values of 22.65 % and 18.58 %, respectively. Class-wise F1 scores show that both configurations reach their highest values for phacelia-dominated mixtures and pure mustard, whereas F1 for pure phacelia and other mixtures remains substantially lower and more variable, and no correct predictions were obtained for the mustard mix and phacelia-mustard classes under either classification. None none of the differences between the smoothed and first-order derivative configurations were statistically significant at $\alpha = 0.05$ for OA, $F1_{macro}$, phacelia mixes, pure mustard and other mixes (all $p \geq 0.06$), while F1 for pure phacelia was significantly higher under the first-order derivative setup ($p = 0.048$).

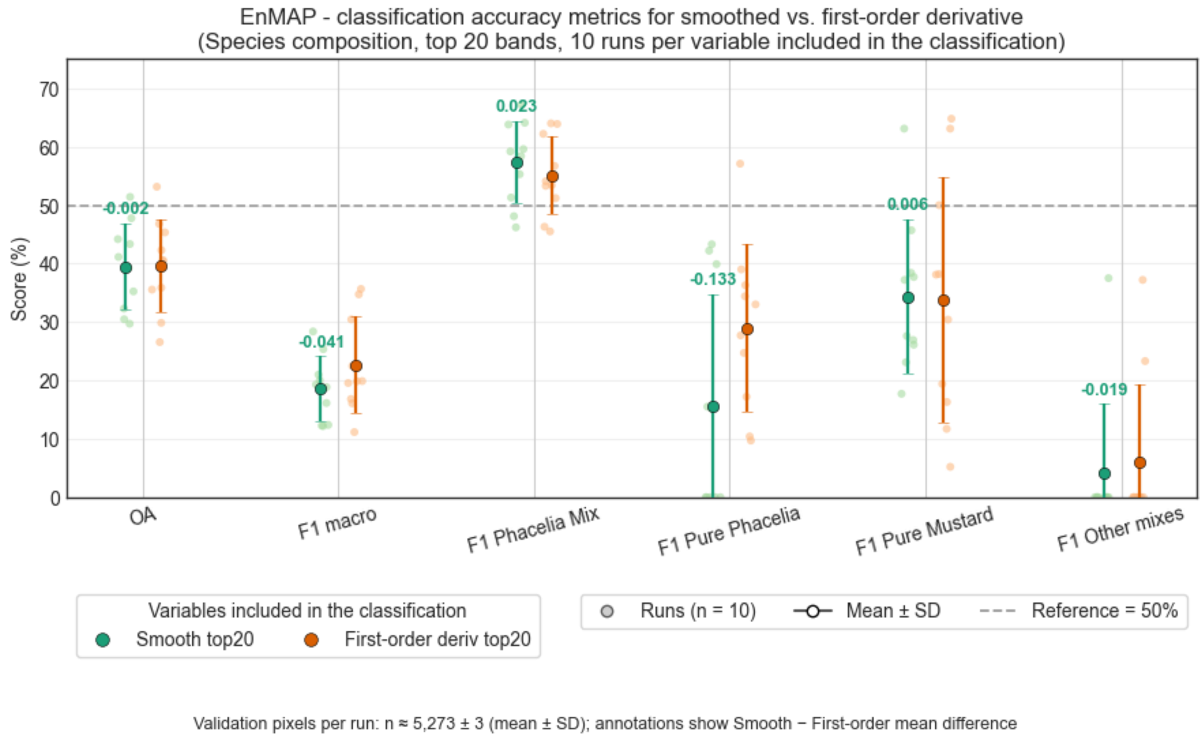
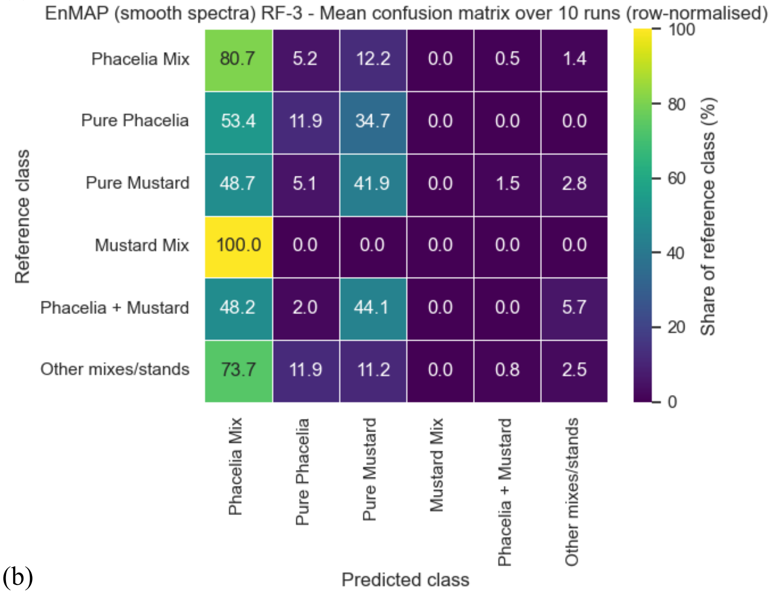


Figure 37: Classification accuracy of RF-3 with EnMAP input when using the 20 most informative smoothed reflectance bands (green) versus the 20 most informative first-order derivative bands (orange).

Figure 38 visualises the mean confusion patterns of RF-3 for the top 20 EnMAP bands, comparing smoothed reflectance (a) and first-order derivative (b) configurations, averaged over ten validation runs and row-normalised. In both cases, the classifier reliably identifies phacelia-dominated mixtures and pure mustard, with around 35.9-80.7 % of the respective reference pixels mapped to the correct class, but a substantial share of these plots is still confused within the phacelia-mustard group, most often being interchanged with pure phacelia or pure mustard. The mustard mix and phacelia–mustard classes are never or only rarely mapped to their correct label and instead collapse almost entirely into the dominant phacelia-mix and pure mustard classes, which is consistent with the near-zero F1 accuracies reported for these classes. The heterogeneous ‘Other mixes/stands’ class remains particularly problematic under both band configurations, with most of its reference pixels being assigned to phacelia-mix or pure mustard and only a small fraction correctly identified, again mirroring the low class-wise F1 scores.

(a)



(b)

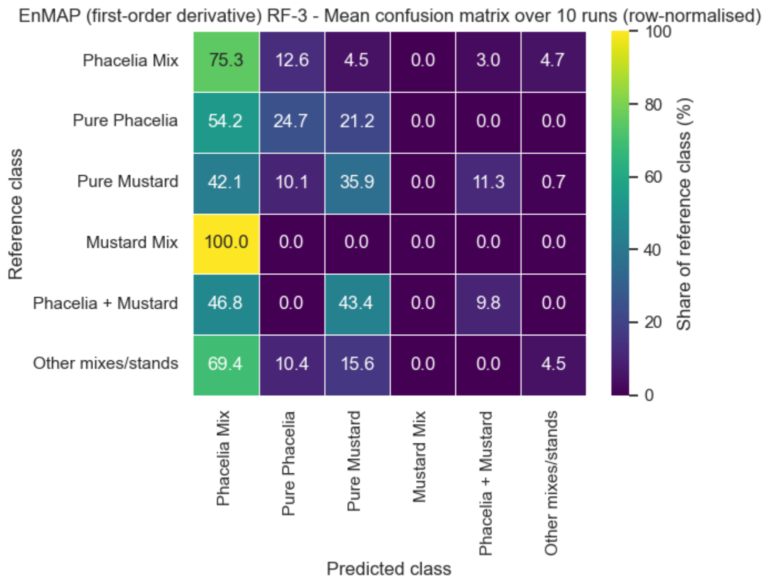


Figure 38: Mean confusion matrix over 10 runs for RF-3 using EnMAP (a) smoothed reflectance and (b) first-order derivative.

4.10. Winter cover crop species composition using EnMAP & Sentinel-2 (Setup 2, RF-3)

Figure 39 compares mean EnMAP (a) and S2 (b) parcel-level spectral signatures for different species compositions of WCC parcels. For both sensors, across the spectrum, all species compositions follow a similar vegetation-type shape. 'Phacelia + Mustard', 'Phacelia mix' and 'Mustard mix' tend to exhibit the highest reflectance on the RE/NIR plateau and slightly lower reflectance in the SWIR, whereas 'Pure Mustard' and 'Pure Phacelia' and 'Other mixes/stands' show somewhat reduced NIR reflectance and slightly higher SWIR levels, like in Setup 1 as well. Compared to S2, the EnMAP spectra appear smoother and slightly compressed, with somewhat lower reflectance levels in the NIR and SWIR but otherwise very consistent shapes and separations between the different species compositions.

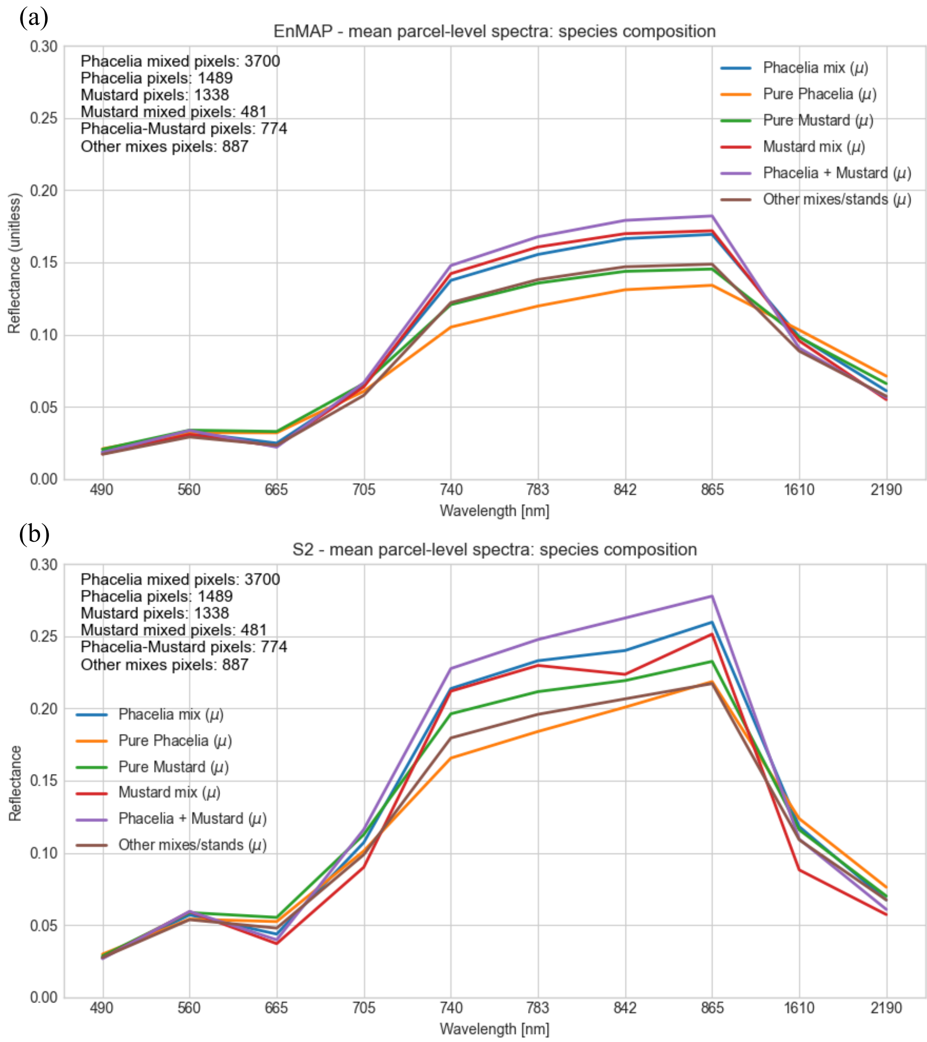


Figure 39: Mean EnMAP-simulated S2 and S2 parcel-level spectra for species composition within the Hannut region. (a) EnMAP-simulated S2 reflectance spectra. (b) S2 reflectance spectra.

For RF-3 in Setup 2, the feature importance patterns for EnMAP- (a) and S2-simulated (b) predictors in the Hannut region are broadly similar, but with distinct leading variables (Figure 40). In the S2-based model, the SWIR band (B11) shows the highest MDG scores, closely followed by the RE band (B5), NDRE and NDWI. In contrast, the EnMAP-based model assigns the greatest importance to the blue band (B2), NDRE and the green band (B3), followed by the SWIR bands (B12 and B11) and the RE band (B5).

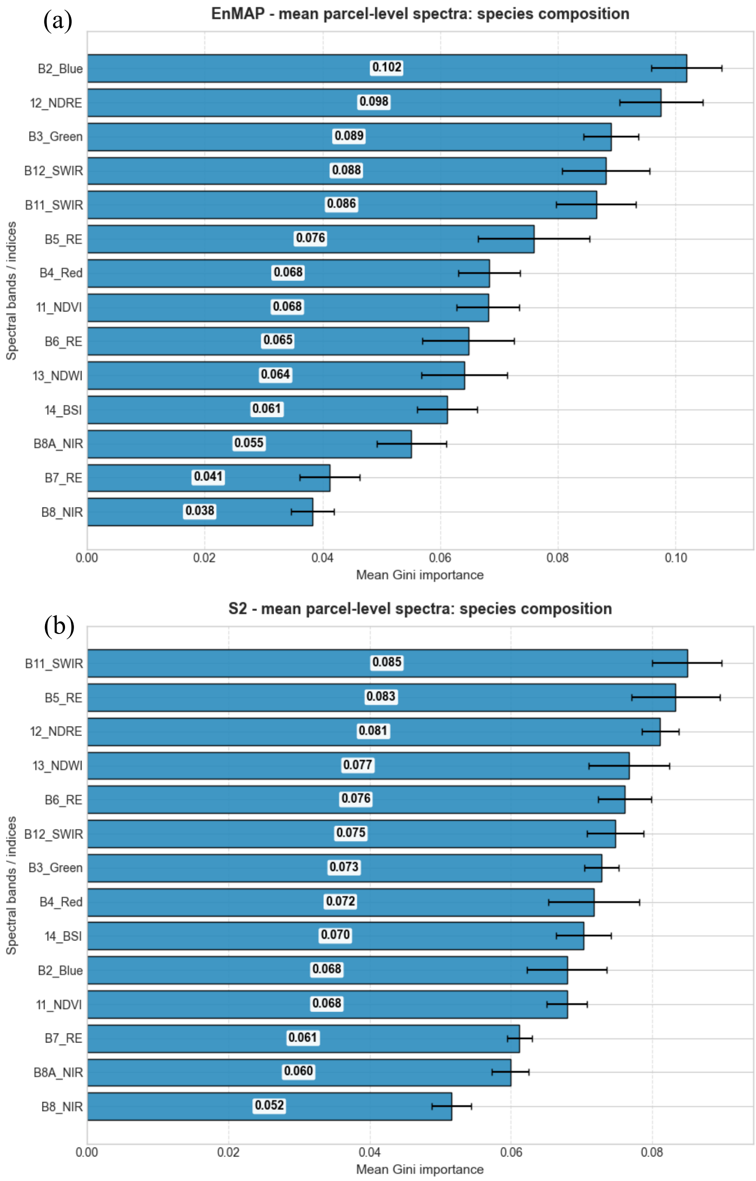


Figure 40: RF-3 feature importance of the Setup 2 variables in the Hannut region. (a) MDG for the EnMAP-simulated S2 bands and VIs. (b) MDG for the corresponding S2 bands and VIs.

Figure 41 compares the classification accuracy of RF-3 for species composition between EnMAP and S2 across ten validation runs. OAs are moderate and similar for both sensors, with mean OA around 43–45 % and no significant difference between EnMAP and S2 (mean OA difference ≈ 1.2 percentage points, $p = 0.47$). Mean $F1_{macro}$ scores are clearly higher for S2 than for EnMAP (about 30–32 % vs. 22–29 %), and this difference of roughly seven percentage points is statistically significant ($p = 0.008$), indicating a more balanced class-wise performance for S2 (Appendix 13). At the class level, both sensors achieve the highest F1 scores for phacelia-dominated mixtures and pure mustard, while F1 for pure phacelia remains much lower and highly variable. For 'Other mixes' S2 outperforms EnMAP by a wide margin, with mean F1 differences of about 35 percentage points (mean F1 ≈ 0.46 vs. 0.09; $p = 0.002$), whereas for the remaining classes (phacelia mix, pure phacelia, pure mustard, phacelia–mustard) the average S2-EnMAP differences in F1 are small and not statistically significant. For the mustard mix class, both sensors achieved F1 scores frequently close to 0 % across runs.

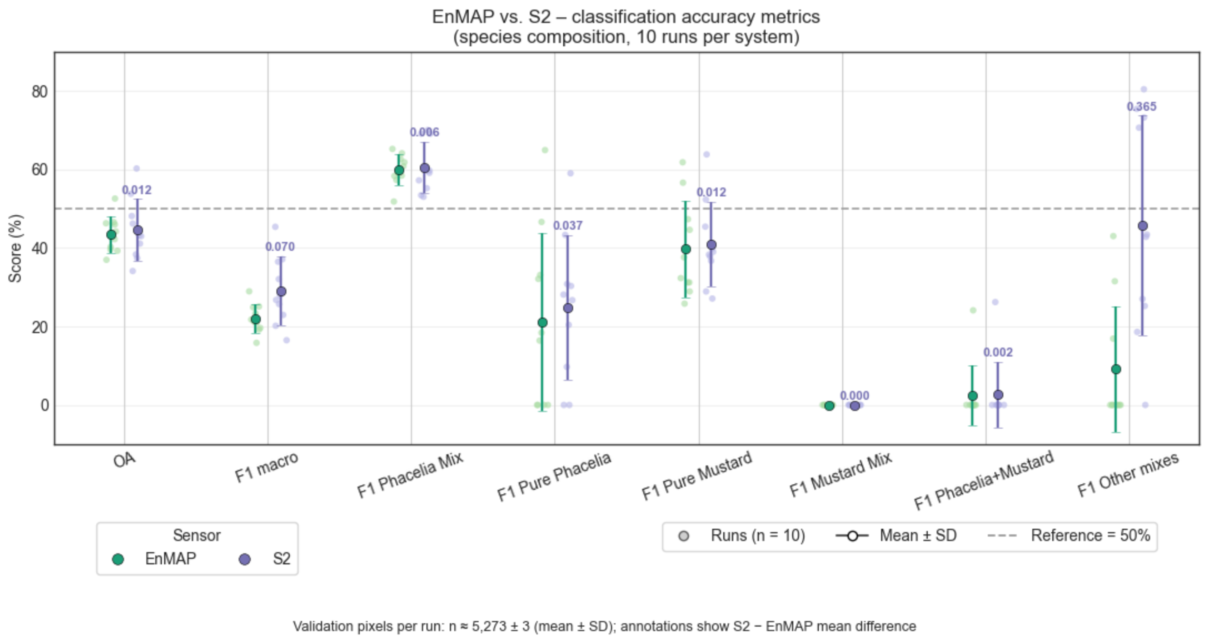


Figure 41: Classification accuracy of RF-3 in Setup 2 when using EnMAP-simulated S2 bands (green) versus the corresponding S2 bands (purple).

Figure 42 illustrates the mean confusion structure of RF-3, averaged over ten runs, separately for EnMAP (a) and S2 (b). Both sensors correctly identify a large share of phacelia-dominated mixtures and pure mustard, with 45.3-85.9 % of the respective reference pixels mapped to the correct class (diagonal cells). Substantial confusion remains within the phacelia-dominated group: a notable fraction of phacelia mixes is predicted as pure phacelia and vice versa. The matrices further show that the mustard mix class is never reliably recognised (100 % of its pixels are assigned to phacelia mix for both sensors), confirming the near-zero F1 scores for this category. For the phacelia-mustard mixtures, both sensors mainly confuse them with pure mustard. A clear difference between sensors appears for the heterogeneous 'Other mixes/stands' class: EnMAP tends to misclassify these plots as phacelia-dominated or pure mustard, whereas S2 correctly assigns 37.5% of this class and thus retains a substantially higher F1 score.

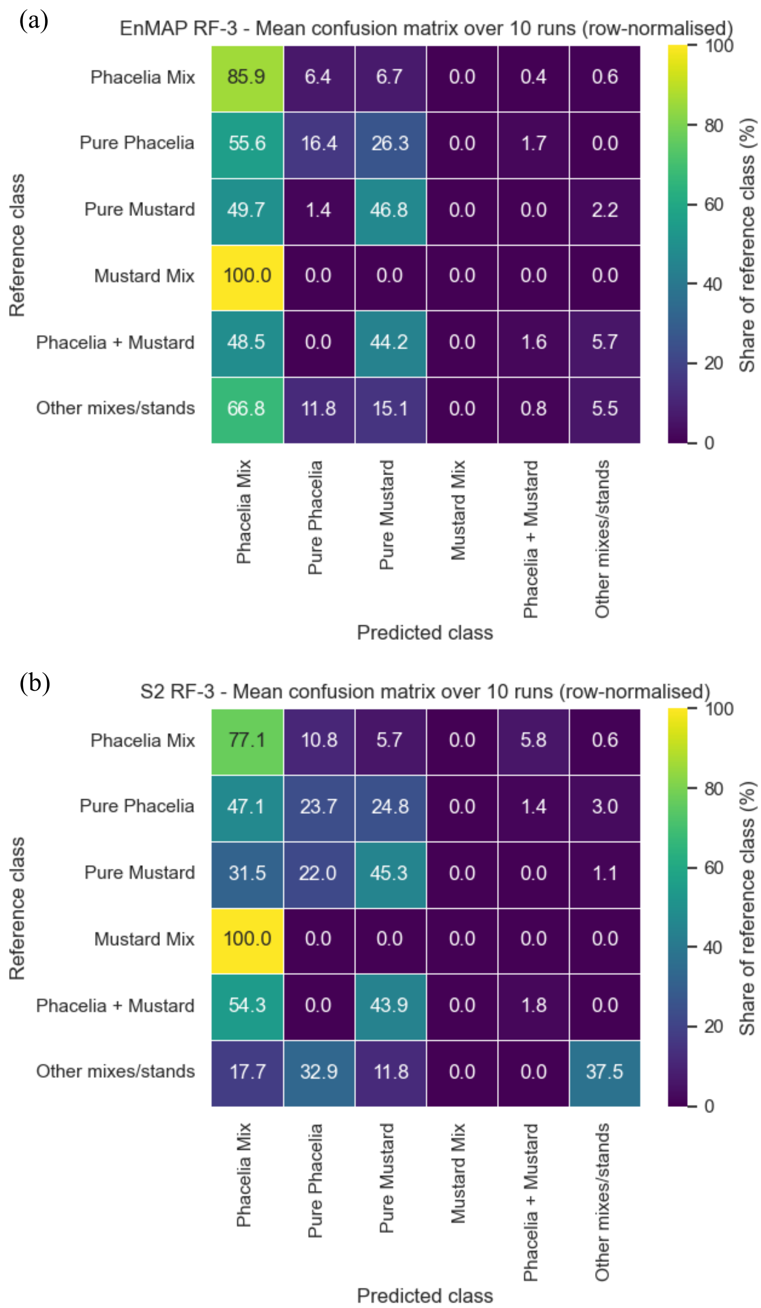


Figure 42: Mean confusion matrix over 10 runs for RF-3 using (a) EnMAP and (b) S2.

4.11. Cross-setup overview of winter cover crop species composition (RF-3)

Across both setups, mapping detailed WCC species composition remained the most challenging classification task, with mean OA of approximately 39–45% and $F1_{macro}$ values between 19–32%. A consistent pattern across setups and sensors is that phacelia-dominated mixtures and pure mustard were classified most reliably, whereas mustard mix and phacelia-mustard classes were rarely or never correctly identified and effectively collapsed into the dominant classes; this is confirmed by both the F1 scores and the confusion matrices. In Setup 1, smoothed reflectance and first-order derivatives yielded very similar overall performance under the 20-band EnMAP configuration, with only a marginal and class-specific advantage of derivatives for pure phacelia ($p = 0.048$); the error structure was dominated by within-group confusions among compositionally similar mixtures under both spectral representations. In Setup 2, EnMAP and S2 achieved comparable OA (mean difference ≈ 1.2 percentage points, $p = 0.47$), but S2 yielded a significantly higher $F1_{macro}$ ($p = 0.008$) and a substantially better F1 for the heterogeneous 'Other mixes' class (mean F1 ≈ 0.46 vs. 0.09; $p = 0.002$), while class-wise differences for the remaining categories were not statistically significant.

4.12. Impact of parcel-based filtering on the classification performance

All results presented so far were obtained from classifications that applied the parcel-based majority filter. In the following, filtered and unfiltered runs are compared to quantify how this post-processing step affects OA, $F1_{macro}$ and class-wise F1 scores across the different setups.

For RF-1 in Setup 1, parcel-based filtering strongly improve EnMAP classification based on the top-20 bands (Appendix 14). For both smoothed reflectance and first-order derivatives, applying the filter increase mean OA and $F1_{macro}$ by about 7–9 percentage points, and class-wise F1 scores for WCC and NWCC showed gains of a similar magnitude. All improvements are highly significant in paired t-tests (all $p < 0.001$).

For RF-1 in Setup 2, filtering again has a strong positive effect on EnMAP but only a minor effect on S2 (Appendix 15). For EnMAP, OA and $F1_{macro}$ increase by about 10 percentage points and all determined accuracy metrics improve significantly (all $p < 0.001$). In contrast, for the corresponding S2 model, differences between filtered and unfiltered configurations remain small (OA and $F1_{macro}$ changes below about 2 percentage points, all $p \geq 0.10$).

For RF-1 in Setup 3, filtering consistently improves the S2 November composite classification (Appendix 16). Across the 10 runs, applying the filter increases mean OA, $F1_{macro}$ and the class-wise F1 scores by roughly 2 percentage points and all differences were statistically significant ($p < 0.01$).

For RF-2 in Setup 1, filtering improves EnMAP top-20 classification performance for both smoothed and first-order spectra (Appendix 17). For smoothed reflectance, OA and $F1_{macro}$ increase by about 4-5 percentage points and F1 for the mixed class by about 4 percentage points; all three gains are significant, whereas the smaller increase for the pure class (≈ 2.3 percentage points) is not. For the first-order derivative, filtering produces slightly larger and consistently significant gains, raising OA and $F1_{macro}$ by about 5 percentage points, F1 for the pure class by almost 6 percentage points, and F1 for the mixed class by about 4 percentage points.

For RF-2 in Setup 2, filtering also tends to improve EnMAP classification performance, but the effect is more moderate than for RF-1 (Appendix 18). Mean OA increases by about 5.6 percentage points ($p \approx 0.03$), while $F1_{macro}$ and the F1 score of the pure class show similar average improvements but are not significant. The F1 score for the mixed class improves significantly for EnMAP (mean +4.6 percentage points, $p \approx 0.01$). For S2, differences between filtered and unfiltered RF-2 configurations remain small for all metrics (mean changes ≤ 3 percentage points, all $p \geq 0.10$).

For RF-2 in Setup 3, filtering again improves the S2 November composite but with a more selective effect than for RF-1 (Appendix 19). Mean OA increases from about 69.9 % to 72.8 % ($p \approx 0.01$). F1 for the mixed class also benefits, rising from 79.2 % to 81.5 % ($p \approx 0.005$), whereas the corresponding increases in $F1_{macro}$ and in F1 for the pure class are smaller (≈ +2.8 and +3.4 percentage points, respectively) and not statistically significant ($p \approx 0.05$–0.12).

For RF-3 in Setup 1, the impact of filtering also depends on the spectral pre-processing (Appendix 20). For the smoothed spectra, mean OA increases from about 33.38 % to 39.48 % ($p \approx 0.002$), but reduced $F1_{macro}$ from 20.7 % to 18.58 %. For the corresponding first-order derivative, mean OA likewise increases from roughly 35.4 % to 39.6 % ($p \approx 0.02$), while the difference in $F1_{macro}$ between filtered and unfiltered runs remain very small and non-significant (23.4 % vs. 22.7 %). Class-wise tests show that in both configurations filtering significantly improves the class-wise F1 for the dominant 'Phacelia mix' and, for smoothed spectra, also for 'Pure Mustard', but leads to significant reductions for 'Mustard mix' and 'Phacelia+Mustard' and, in the derivative case, for the heterogeneous 'Other mixes/stands' class. At the class level, filtering significantly improves F1 for phacelia mix and pure mustard significantly, whereas F1 for mustard mix and phacelia-mustard dropped to zero from low but non-zero values in the unfiltered case; these decreases were significant.

For RF-3 in Setup 2, filtering again increases OA for both sensors but had a more nuanced effect on class-wise F1 scores (Appendix 21). For EnMAP, OA increases by about 6.9 percentage points (from 36.5 % to 43.4 %; $p \approx 0.0001$), while $F1_{macro}$ decreases slightly by about 1.5 percentage points and this reduction is not significant ($p \approx 0.07$). Class-wise, filtering significantly increases F1 for the phacelia-mix and pure mustard classes (mean gains of roughly 8.3 and 6.4 percentage points, respectively), but leads to significant decreases for the mustard mix and phacelia-mustard mixtures. For S2, the filter similarly produces a significant OA increase of about 5.0 percentage points (39.6 % to 44.5 %), while changes in $F1_{macro}$ remained small and non-significant. Class-wise tests indicate a clear benefit only for phacelia mix, whereas effects on the remaining species-composition classes are either small or associated with very low absolute F1 values.

Overall, these experiments show that parcel-based filtering systematically increases OA and often boosts F1 for dominant classes, but its effect on minority and mixture classes is configuration-dependent and can entail trade-offs in class-wise balance.

CHAPTER
FIVE

DISCUSSION

This study evaluates the potential of MSI and HSI satellite data for monitoring WCCs in the Hesbaye region of Wallonia across three levels of classification complexity: WCC presence versus absence (RF-1), mixture type (RF-2), and species composition (RF-3). Addressing each level requires balancing sensor and temporal limitations against the increasing spectral similarity of finer WCC classes — a challenge that has been only partially resolved in the existing literature. The following sections discuss how performance varies across setups and RF levels, what methodological and data constraints shape these outcomes, and what the results imply for operational monitoring and future MSI and HSI missions.

5.1. Overall performance of MSI and HSI approaches for WCC monitoring

Across all sensors and experimental setups, classification performance decreases systematically as classification complexity increases.

At the coarsest level (RF-1), both EnMAP single-date imagery and the highest-performing S2 monthly composites achieve stable, high accuracies (OA > 85%, $F1_{macro}$ > 84%) when distinguishing between WCC and NWCC pixels. This confirms that the detection of WCC presence using satellites is robust under the conditions evaluated here. Under harmonised conditions, the EnMAP-simulated S2 predictors significantly outperform the corresponding S2 observations for RF-1 (OA = 86.84 vs. 80.72%, $F1_{macro}$ = 85.5 vs. 78.6%). Within Setup 3, the S2 time series yield broadly similar F1 scores for WCC and NWCC. However, in Setups 1 and 2, $F1_{NWCC}$ is consistently lower than for $F1_{WCC}$. Although binary WCC detection is the simplest task considered here, it should not be underestimated; signals from WCC fields are often contaminated by background contributions from soil and crop residues (Ahmed et al., 2023; Wang et al., 2023).

For RF-2, the accuracy level decreased but remained within a range that is still useful for regional analyses. OA and $F1_{macro}$ generally fall within the 60–73% range. This behaviour is consistent with other MSI-based studies that have attempted to go beyond simple presence mapping, reporting similar accuracy levels for broad WCC type classes (Kc et al., 2021; Vanpoucke et al., 2024; Vanongeval et al., 2026).

Performance drops markedly at the most detailed level (RF-3): mean OA falls to approximately 40–45% and $F1_{macro}$ remains below 30% across Setups 1 and 2. Moderate F1 scores are largely confined to phacelia-dominated mixtures and pure mustard, while minority classes — including pure phacelia, mustard mixes, and phacelia-mustard combinations — are poorly detected or not detected at all. These findings mirror those of several recent studies which report that mixed or structurally heterogeneous WCC categories can be mapped reliably at coarse levels, but that resolving species-level composition remains difficult due to spectrally and phenologically overlapping functional groups (Do Nascimento Bendini et al., 2024; Mirzaei et al., 2024; Vanpoucke et al., 2024; Vanongeval et al., 2026).

Overall, these results suggest that the current MS and HS configurations are well suited to operational presence mapping and, to a certain extent, identifying broad mixture types. However, reliable species-level composition mapping remains unattainable under the current sensor, temporal and training data conditions. The following sections examine performance at each RF level in detail (Sections 5.2.–5.6.) before addressing the methodological and data constraints that shaped these outcomes (Section 5.7.) and their implications for operational monitoring and future research (Section 5.8.).

5.2. Added value of EnMAP HS data

5.2.1. Single-date EnMAP performance (Setup 1 — Objective I)

Using a single EnMAP acquisition on 30 December 2025, RF-1 achieves high and stable accuracies (OA > 85%, $F1_{macro}$ > 84%) across all tested configurations, confirming that a single late-season HS image acquired at this stage of the WCC vegetative cycle is sufficient for robust binary WCC vs. NWCC classification. However, it should be noted that this result only applies to the specific late-December phenological conditions, when WCC canopy growth is almost complete. Acquisitions at the top of the WCC's vegetative phase, coinciding with peak biomass and maximum vegetation–soil contrast, would likely produce equal or stronger discrimination, as discussed in Section 5.7.2. Under these acquisition conditions, restricting the input from the full EnMAP spectrum to the 20 most informative bands has a negligible effect on the outcome. OA and $F1_{macro}$ are preserved, with mean differences below one percentage point. Paired t-tests show no significant differences between the top-20 and full-spectrum configurations (Appendix 8). This is consistent with earlier HS work demonstrating that discriminative power for crop-relevant classes is concentrated in a narrow portion of the spectrum, and that a compact set of well-placed narrow bands captures most of the available information with only marginal gains beyond approximately 12–30 bands (Thenkabail, 2002; Thenkabail et al., 2004, 2019). The most informative predictors for RF-1 are concentrated in the VIS, early RE/NIR and SWIR regions. For a December acquisition, the importance of blue and red bands is physically plausible: these bands capture the strong brightness contrast between bright, bare or residue-covered soil and the darker WCC canopies that absorb chlorophyll. Meanwhile, the RE and NIR bands reflect the vegetation-soil contrast that underpins most greenness-based discrimination. The systematic pattern of $F1_{WCC}$ exceeding $F1_{NWCC}$ across Setups 1 and 2 reflects the greater heterogeneity of the NWCC class. Bare soil, crop residues and winter cash crops are all incorporated into this class under a single label. This produces a wider range of spectral signatures that are more difficult to classify. This heterogeneity is further amplified in December when WCC canopies span a wide range of conditions from healthy green stands to partially senescent or frost-affected vegetation. The spectral overlap with certain NWCC surface types is increased by this.

For RF-2, performance decreases (with OA at ~69% and $F1_{macro}$ at 60-63%, which is somewhat lower due to class imbalance), with feature importance shifting towards the early RE/NIR region and secondary SWIR contributions. Mixed stands are consistently classified more accurately than pure stands, as discussed in detail in Section 5.4. The modest accuracy of the RF-2 model based on a single acquisition in December is consistent with the idea that the separability of mixture types is primarily driven by differences in overall canopy density and residue exposure rather than by fine biochemical differences. This suggests that these broad structural signals are captured by a small subset of RE/NIR and SWIR bands.

RF-3 yields the poorest classification performance: the mean OA falls to approximately 39% and $F1_{macro}$ remains between approximately 18.6-22.7%. Moderate F1 scores are largely confined to phacelia-dominated mixtures and pure mustard. Minority classes, including mustard mixes, phacelia-mustard combinations and other mixes, are either rarely detected or collapse entirely into the dominant categories. This pattern reflects two compounding constraints. Firstly, the December acquisition captures a convergent phenological window in which many species and mixtures have similar mature or partially senescing canopies. This reduces the spectral variability between species that would otherwise aid discrimination. Secondly, the limited number of training parcels for minority classes means that the RF has not learned robust decision boundaries for rare compositions. The feature importance for RF-3 emphasises the RE/NIR and SWIR regions more strongly than for RF-1 and RF-2. This is consistent with the increasing role of structural and moisture-related SWIR signals in mature canopies, as reported by Basinger et al. (2020). However, even these additional spectral dimensions are insufficient to overcome the fundamental similarity of functional groups that share canopy architecture, such as erect broadleaf species, at this phenological stage. Section 5.5. discusses the role of class design in driving these errors in detail.

Spectral pre-processing only has a subtle influence across all RF levels. There are no systematic or statistically significant differences in OA or $F1_{macro}$ for RF-1 and RF-2, whether smoothed reflectance or first-order derivatives are used. This is consistent with the mean spectra, which reveal similar shapes for WCC and NWCC, with only slight variations in magnitude and absorption features (Figure 14). Both representations essentially capture the same discriminative information within the top-20 band configuration. For RF-3, first-order derivatives slightly enhance class-wise F1 for some pure stands, most notably pure phacelia, while smoothed reflectance occasionally favours more heterogeneous mixed classes. However, these shifts do not translate into consistent gains in OA or $F1_{macro}$. This behaviour reflects the expected trade-off: smoothing reduces random noise that would otherwise be amplified by differentiation, while derivatives emphasise spectral slopes, absorption edges and subtle feature positions, thereby suppressing broad illumination and soil background effects. Under the present noise level and single-date design, however, these advantages are insufficient to overcome the fundamental spectral similarity and within-class variability of the WCC mixtures. Overall, the choice of pre-processing primarily redistributes performance among individual classes in RF-3 rather than improving aggregate accuracy.

5.2.2. EnMAP versus S2 under harmonised conditions (Setup 2 — Objective II)

To isolate the effect of spectral resolution from other sensor-related differences, the reflectances of the EnMAP sensor were resampled to resemble the bands of the S2 sensor. Both datasets were then co-registered to the same 30 m grid. The harmonised EnMAP and S2 inputs demonstrate significant agreement at a pixel level across most band pairs and VIs. This confirms that the two sensors are spectrally and spatially comparable under controlled conditions. Any residual disagreement between the sensors is likely to reflect genuine differences at the instrument level, such as distinct atmospheric correction algorithms, SRFs and minor shifts in the centre of the bands, rather than being a limitation of the resampling approach. It also suggests that the observed differences in classification performance are due to subtle instrument-level characteristics rather than variations in band placement or spatial resolution.

For RF-1, EnMAP consistently and significantly outperforms S2 in terms of OA, $F1_{macro}$, and class-wise F1, with the advantage being most pronounced for the NWCC class. Since both inputs use the same S2-like bands at 30 m, this result cannot be attributed to finer spectral sampling alone. Instead, it points to more subtle differences at the instrument level. Several non-exclusive explanations are plausible. Firstly, EnMAP's higher radiometric resolution (≥14 bits versus 12 bits for S2) and superior SNR reduce quantisation noise (Table 1), yielding a cleaner effective signal particularly in the NIR and SWIR bands. Secondly, the shape of the SRF determines the spectral purity of the signal within each band. The EnMAP-simulated S2-like bands are constructed by averaging only the narrow EnMAP channels whose centres fall within each S2 band's spectral range. This produces a more rectangular effective SRF with sharper band boundaries and more uniform spectral weighting (Figure 17). The practical consequence of this is that there is less spectral mixing within the bands: fewer wavelengths from outside the target spectral region contribute to the aggregated reflectance value, yielding a cleaner signal for the RF classifier. This band-sharpening effect is unequal across the spectrum. In the VNIR, EnMAP's densely packed narrow channels closely approximate a rectangular window for each simulated band, maximising spectral purity. In the SWIR, however, the wider native EnMAP band spacing means that fewer channels contribute to each simulated band. This results in a smaller sharpening advantage over native S2, which may partly explain why the performance gap between EnMAP and S2 is less pronounced for features that rely heavily on SWIR information. Thirdly, the stringent in-orbit calibration of EnMAP and its optimised gain settings for vegetation reflectance levels may further reduce systematic radiometric error, though this is difficult to verify directly from classification outputs. Fourthly, acquisition-specific factors, such as the low solar elevation angle typical of a December scene in Belgium, potential differences in atmospheric correction performance and residual aerosol or adjacency effects, may amplify radiometric differences between the two stacks.

Analysis of within-class variance confirms systematically lower spectral variability in the EnMAP stack, particularly in the NIR and SWIR bands, and for heterogeneous NWCC parcels. This suggests that EnMAP's measurement characteristics yield a cleaner effective signal for the binary NWCC vs. WCC task, even when restricted to S2-like bands (Figure 18). This finding is consistent with previous studies demonstrating that narrow HS bands can achieve markedly higher classification accuracies than broadband MS configurations when SNR is sufficient, particularly in the SWIR region (Clark et al., 2005; Bostan et al., 2016; Galvão et al., 2019; Yao et al., 2019; Mirzaei et al., 2024).

As in Setup 1, $F1_{WCC}$ remains higher than $F1_{NWCC}$ for both sensors. This is probably because of the natural variation within the NWCC group and the limitations of using a single winter date without specific information about the timing of the growing season.

For RF-2, EnMAP achieves slightly higher OA (= 70.23%), $F1_{macro}$ (= 65.55%), and class-wise F1 scores than S2 across most runs (OA = 67.8%, $F1_{macro}$ = 62.81%), but none of these differences are statistically significant at the 0.05 level based on paired t-tests. This marginal, non-significant advantage suggests that the additional radiometric quality of EnMAP is insufficient to overcome the structural ambiguity of the pure vs. mixed distinction for mixture type classification under the present feature set and sample size (Section 5.4.). This is physically plausible: mixture type separability at a single date in December is largely driven by broad differences in canopy density and residue exposure — signals that S2 already captures adequately through its RE, NIR and SWIR bands. Finer spectral detail and a cleaner signal only contribute marginally once these dominant structural contrasts are resolved.

For RF-3, no systematic advantage is observed for EnMAP (OA = 43.38%, $F1_{macro}$ = 22.06%), whereas S2 occasionally yields more balanced class-wise performance with a significantly higher $F1_{macro}$ (OA = 44.54%, $F1_{macro}$ = 29.09%). This counterintuitive result highlights that the main bottleneck at the species composition level is class design, training sample size and temporal sampling rather than spectral resolution (Section 5.5.).

5.2.3. Spectral regions driving WCC discrimination

Across both setups (1 and 2) and all three RF levels, patterns of feature importance converge on a consistent set of spectral regions. The RE and NIR are informative predictors at all classification levels, followed by selected SWIR bands, the relative contribution of which increases progressively from RF-1 to RF-3. VIS bands, particularly those in the blue spectrum, play a strong role at RF-1, but become less important at finer levels, where canopy biochemistry and structure, rather than simple vegetation-soil contrast, become the dominant discriminating signal.

This spectral hierarchy is physically interpretable. In Setup 1, where only EnMAP bands were used, the top 20 RF-1 predictors are concentrated in the VIS and SWIR region for smoothed reflectance and particularly in the early RE/NIR region for first order derivatives, with additional, but smaller, contributions from a few SWIR bands. In Setup 2, where spectral bands were combined with VIs, the most important RF-1 predictors were again the blue and red bands, as well as the NDVI, NDRE and BSI. Meanwhile, the green, RE and SWIR bands played secondary roles. The RE and NIR bands, and their derivatives in VIs such as NDVI and NDRE, capture the sharp contrast between chlorophyll-rich WCC canopies and brighter, less vegetated NWCC backgrounds. This is the same mechanism that underpins NDVI-based WCC detection in the broader literature (Ahmed et al., 2023; Vanpoucke et al., 2024). The high importance of the blue and red bands in both setups reflects the strong brightness contrast between bare or residue-covered soil and vegetated WCC fields in winter. It was considered that the high importance of the blue region bands might be an artefact of the imperfect cloud and shadow masking of the EnMAP image used in Setups 1 and 2, or of the radiometric pre-processing. However, several checks did not support this explanation. Similarly high importance was observed for the blue region bands in Setup 2 for S2 and removing the blue band merely shifted importance to the green and red bands, slightly reducing OA. This suggests that all VIS bands consistently distinguish between bare soil and vegetation, rather than the pattern being driven by residual cloud or pre-processing artefacts.

As the complexity of classification increases towards RF-2 and RF-3, the most informative bands shift further into the RE/NIR and SWIR. This is consistent with the increasing importance of canopy density, water content and structural biochemistry — including lignin and cellulose — in distinguishing between compositionally similar WCC types. Basinger et al. (2020) specifically demonstrated that SWIR1 (1,565-1,655 nm) and SWIR2 (2,100-2,280 nm) wavelengths offer greater differentiation capabilities for mature canopies, where early-season differences in VIS and NIR spectrum have converged and species discrimination increasingly relies on structural absorption features.

The EnMAP bands identified as most important across setups align well with the optimal narrow-band placement recommended by Thenkabail et al. (2004, 2019) from their systematic analysis of vegetation discrimination across the 400–2,500 nm range. Their work showed that a compact set of approximately 12–30 bands — centred on consistent wavelength positions in the VIS chlorophyll absorption region, the RE inflection zone, the NIR plateau, and key SWIR absorption features — captures most of the crop-relevant spectral variability, with only marginal gains from additional bands. The concentration of importance in the top-20 EnMAP bands observed in this study, and the plateau in accuracy beyond that subset, is consistent with this finding. However, a direct band-by-band comparison between the optimal Thenkabail positions and the specific EnMAP bands selected here would be a valuable further analysis to determine whether the most informative bands in this WCC context align with the universally recommended set or reflect WCC-specific spectral characteristics not anticipated in broader crop discrimination studies.

5.3. Sentinel-2 monthly composites for WCC monitoring (Setup 3 — Objective III)

The S2 monthly composites produce RF-1 mean OA, $F1_{macro}$ and class-wise F1 values ranging from approximately 76% to 86% across the months. November, October, January, February and March perform similarly well, while the accuracy in August, September and December is substantially lower. The magnitude of $F1_{WCC}$ and $F1_{NWCC}$ is similar across these months, indicating that WCC and NWCC parcels are detected with comparable reliability rather than the model being biased towards one class.

The month-to-month variation in RF-1 accuracy closely follows the phenological dynamics captured in the NDVI trajectories (Figure 24). In August, the low WCC NDVI reflects the absence of an established WCC canopy; the fields have recently been harvested or tilled, and WCC sowing has either not yet occurred or has produced negligible biomass. At this stage, the spectral signal of WCC parcels is dominated by bare soil, which lacks class specificity, thereby limiting the RF's ability to distinguish between future WCC and other bare or harvested NWCC fields. The recovery in September and strong performance from October onwards coincide with progressive WCC canopy closure and increasing inter-class NDVI contrast. The peak performance in November reflects the moment of maximum spectral separability, when WCC biomass and greenness are at their highest, while NWCC fields are either harvested or tilled or hosting early-stage winter crops, which have a substantially lower NDVI. This supports the selection of November as the reference month for subsequent analyses. The mid-winter degradation in December is consistent with two concurrent processes: WCC senescence reduces the greenness advantage of WCC parcels while the presence of still-green winter cereals or volunteer residues on NWCC fields increases their NDVI, thereby narrowing the inter-class contrast. From January to March, accuracies recover and remain high, even though the mean NDVI of NWCC exceeds that of WCC in some parcels. This suggests that the RF exploits persistent structural or textural differences between the classes rather than relying on the WCC always being uniformly greener than the NWCC. However, the good performance in March should be interpreted with caution, since the composite was based on acquisitions up to approximately 20 March only, and data collection and analysis were conducted before the end of the month. Consequently, the composite does not capture the full phenological state of March and may not be representative of the entire month.

The RF-1 OA levels obtained here (76–86%) are at the lower end of the range of accuracies reported by regional WCC mapping studies that use denser VI or BV time series with explicit phenological metrics (Schmidt et al., 2021; Seifert et al., 2019; Barnes et al., 2021; Ahmed et al., 2023). These studies usually address binary or few-class issues in large, relatively homogeneous regions with multi-year ground truth datasets comprising several thousand fields, which helps to explain their high levels of accuracy (often ≥90%). In northern Belgium, Vanongeval et al. (2026) achieved OA = 0.85 and κ = 0.73 at the broadest classification level using S2 time series and a hierarchical RF, while Vanpoucke et al. (2024) reported median F1 scores of 0.88–0.89 for coarse WCC vs. NWCC labels from a parcel-based approach. The present results fall somewhat below these benchmarks, which is consistent with the more limited number of in situ parcels, the higher functional diversity of Walloon WCC mixtures, and the use of monthly composites rather than optimised phenological metrics in this study. A related limitation concerns the use of temporal information within Setup 3: each monthly composite was classified independently, meaning the RF only had access to the spectral state of a single time window and had no knowledge of how parcels evolved across the season. Combining all monthly composites into a single multi-temporal feature space would have enabled the classifier to exploit the full phenological trajectory, allowing feature importance to reflect the most discriminating phenological stage for each variable. This represents a tractable, high-priority extension to the current Setup 3 framework.

The feature importance for the November composite is dominated by RE and NIR bands, along with NDVI and NDRE, while VIS bands and LAI contribute comparatively little. This reflects the strong contrast between green WCC canopies and the sparser vegetation or residues in NWCC parcels in early winter. This is consistent with previous findings (Barnes et al., 2021; Schmidt et al., 2021; Ahmed et al., 2023; Vanongeval et al., 2026). LAI derived from S2 showed consistently low Gini importance across months, likely due to sparse and heterogeneous winter vegetation cover, substantial proportions of masked or low-quality LAI retrievals, and the possibility that simple NIR-based indices already capture most of the separability between WCC and NWCC under these conditions.

For RF-2, the November composite yields a mean OA of 72.76% and $F1_{macro}$ of 64.41%. Mixed stands are classified substantially more accurately than pure stands, which is consistent with the pattern observed in Setups 1 and 2, and discussed in detail in Section 5.4. The feature importance analysis emphasises the green and SWIR bands (B11 and B12), as well as the NDRE band, with secondary contributions from the NDVI and NDWI bands. Meanwhile, the VIS bands, LAI and BSI bands play a minor role. The greater prominence of the SWIR bands at RF-2 compared to RF-1 is consistent with the shift towards canopy structure and moisture signals as the classification problem becomes more complex (Barnes et al., 2021), mirroring the spectral hierarchy observed in Setups 1 and 2 (Section 5.2.3).

5.4. Spectral blending impacts on the pure vs mixed classification

A striking and counterintuitive pattern across all setups and sensors is that mixed WCC stands are consistently classified more accurately than pure stands. This can be best explained as a labelling and management artefact rather than a genuine spectral phenomenon. In this study, 'pure' parcels are identified based on field information rather than a strict spectral purity threshold. They may contain establishment gaps, uneven sowing density, volunteer species, and weeds. These factors increase within-parcel spectral heterogeneity. By contrast, mixed stands are often dominated by one or two vigorous species, typically mustard or phacelia, which form dense, closed canopies with more homogeneous spectral signatures than many nominally pure stands. Probability-threshold sensitivity tests from the Setup 2 EnMAP RF-2 model support this interpretation: tightening the decision boundary for the 'pure' class substantially increases the F1 score for 'pure' pixels (from 53.0% to 58.6%, $p < 0.01$), while reducing it for 'mixed' pixels (from 78.1% to 73.4%, $p < 0.01$). There is no significant change in OA or $F1_{macro}$. Probability histograms further support this: while RF-1 probabilities for WCC vs. NWCC are strongly bimodal and concentrated near 0 and 1 (indicating confident decisions), the pure–mixed RF-2 probabilities show much heavier density around intermediate values (Figure 31). This suggests that many pixels are close to the decision boundary. This is consistent with the idea that the current pure versus mixed labels represent a spectrally ambiguous distinction rather than two distinct classes.

The high run-to-run variability in RF-2 pure-class F1 scores warrants a separate explanation from the average underperformance discussed above. Two mechanisms operate simultaneously, but they should not be conflated. Firstly, the small number of validation parcels per run — 35–44 parcels for the full WCC subset in Setups 1–2, distributed across three or four species-composition classes — means that the classification of a small cluster of spectrally challenging pure parcels in the calibration or validation set can substantially affect the class-wise F1 score. This is a statistical artefact of the sample size rather than a property of the sensor or classifier. Secondly, pure WCC parcels do not form a spectrally homogeneous group. Pure grasses, pure legumes and pure brassicas have different spectral signatures, and these differences within classes increase the structural complexity of the pure class independently of sample size. Disentangling these two effects would require a dedicated experiment that treats pure grass, pure legume and mixed stands as separate target classes. This analysis was beyond the scope of this study, but represents a natural next step for species-level WCC monitoring.

Therefore, improving RF-2 performance will require cleaner label definitions, such as using continuous mixture proportions derived from field spectroscopy or high-resolution imagery, rather than simply improving sensor characteristics or temporal sampling.

5.5. Species composition mapping

In both Setup 1 and Setup 2, the RF-3 mean OA falls to approximately 40–45% and $F1_{macro}$ remains below 30%, with moderate performance largely confined to phacelia-dominated mixtures and pure mustard. Minority classes, such as mustard mixes, phacelia-mustard combinations, pure phacelia and other mixes, are either poorly detected or almost entirely collapsed into the dominant phacelia and mustard categories. This pattern remains consistent when considering different sensors and pre-processing choices. This confirms that the bottleneck at RF-3 is not sensor-specific, but rather reflects deeper constraints related to spectral convergence, class design and the availability of training data.

The primary constraint is canopy dominance. In mixed WCC stands, one or two vigorous species — typically mustard and/or phacelia — produce the dominant canopy signal. Subordinate species, such as clover, nyger and minor grass components, contribute little to pixel-level reflectance. This means that mixtures with different compositions can appear spectrally identical when their dominant species are the same, regardless of what else is sown beneath the main canopy. Furthermore, the December acquisition date captures a convergent phenological window in which many WCC species have reached a similar mature or partially senescing canopy state. Earlier in the growing season, typically two to five weeks after planting, species occupy distinct phenological stages and canopy architectures diverge, resulting in the greatest spectral differences between functional groups (Basinger et al., 2020). By late December, these differences have largely disappeared. Sowing date strongly modulates the phenological stage at any given acquisition, but it was not available as a covariate in this study. Its absence is meaningful, as it could have helped account for management-driven variability in canopy development that blurs spectral class boundaries (Vanongeval et al., 2026).

The second constraint is that the class aggregation scheme itself may have contributed to the poor performance. The variance within classes could be as large as the variance between them. The defined categories grouped species mixtures based on compositional identity rather than spectral coherence. From a functional group perspective, these findings closely align with the observation that species with similar canopy architecture — erect broadleaf, prostrate or grass-like — are inherently difficult to distinguish spectrally, regardless of sensor (Basinger et al., 2020). In this dataset, for example, both mustard and phacelia are erect broadleaf species with similar canopy structures at peak biomass, producing only subtle differences in reflectance magnitude and minor shifts in RE absorption features. 'Other mixes' combine species from several morphological and phenological groups, generating especially heterogeneous and overlapping signatures. Additional agronomic factors, such as sowing rate, row spacing, termination timing, soil type and local weather conditions, likely have a stronger influence on canopy composition and structure in December than the nominal seed-mix ratio. This further blurs the spectral distinctions between classes (Ahmed et al., 2023). Similar conclusions were drawn by Vanpoucke et al. (2024), who reported significant overlap among their leafy, grass-like and mixture catch-crop groups, and by Mirzaei et al. (2024), who identified canopy architecture convergence as a key limitation for species-level HS classification. Whether a redesign of class aggregations would yield meaningful accuracy gains is an open question that requires explicit testing in future studies.

The confusion structure across setups confirms that residual errors predominantly arise from within-group confusions among compositionally similar mixtures, rather than from mix-ups between functionally unrelated classes. Phacelia-dominated mixtures and pure mustard are frequently interchanged, and minority classes — such as mustard mix and phacelia-mustard — almost entirely collapse into these dominant categories (Figure 38, 42). This pattern of small functional groups being absorbed into larger dominant classes is a recurring theme in RF-based WCC mapping. It reflects the limited sample size for rare classes, as well as genuine spectral similarity driven by overlapping management calendars and growth habits (Vanpoucke et al., 2024; Mirzaei et al., 2024). Applying the parcel-based majority filter amplified this effect, increasing OA and F1 for dominant classes by enforcing within-parcel consistency, but further suppressing the already low F1 scores of minority classes, whose sparse pixel-level detections were overridden by majority voting.

The occasional advantage of S2 over EnMAP in terms of $F1_{macro}$ at the RF-3 level, observed in Setup 2, reinforces this interpretation. Since both sensors used the same S2-like bands under harmonised conditions, the difference cannot be due to spectral resolution. Instead, S2's slightly more balanced treatment of heterogeneous classes likely results in smoother spectral averaging over mixed pixels. This reduces within-class variance for complex canopies and helps the classifier to produce more consistent predictions across classes. However, this does not represent a systematic advantage of S2 over EnMAP. For species discrimination, no sensor can learn reliable boundaries for rare minority classes due to the present class design and sample sizes. Therefore, the performance differences between sensors are dominated by stochastic variation in the training splits rather than systematic spectral differences.

To move beyond the limited RF-3 performance demonstrated here, changes are needed in the timing of acquisitions, the design of classes and the availability of reference data. These constraints are common to both setups and are addressed in the context of future research directions in Section 5.8.3.

5.6. Temporal richness vs. spectral richness: a cross-setup synthesis

The results from Setup 3 cannot be directly compared to those from Setups 1 and 2 because the setups differ in terms of geographic coverage, feature design and temporal sampling. Nevertheless, it is noteworthy that similar RF-1 accuracy ranges are achieved across all three configurations, albeit through different mechanisms. Temporal composites exploit month-to-month phenological trajectories implicitly, while single-date imagery relies on spectral contrast within a single acquisition. Furthermore, a moderate accuracy ceiling for RF-2 is consistently observed across all setups, regardless of whether the input was spectrally rich or temporally dense. This suggests that the dominant limiting factors at finer classification levels — spectral similarity between functional groups, label ambiguity and limited training data — cannot be resolved by either approach alone. It also indicates the need for strategies that combine denser phenological metrics with well-timed HS acquisitions (Vanongeval et al., 2026; Vanpoucke et al., 2024). Testing this hypothesis directly is an important area for future research (Section 5.8.3.).

5.7. Drivers and limitations of classification performance

The classification performance across all setups and RF levels emerges from the combined influence of methodological choices, sensor and acquisition characteristics, and reference data constraints, rather than from any single limiting factor. The following subsections address these factors in turn, starting with methodological choices (Section 5.7.1.) and moving on to data and acquisition constraints (Section 5.7.2.).

5.7.1. Methodological choices

The study relies on a RF algorithm with minor hyperparameter tuning and band selection across all setups. RF was chosen as it is used in most MSI-based WCC studies, delivers robust and consistent results without extensive tuning, and is inherently resistant to overfitting (Belgiu & Drăguţ, 2016; Ahmed et al., 2023; Vanpoucke et al., 2024). This is a critical advantage when working with small in situ datasets and when input data is highly variable — both of which are conditions characteristic of the present study. However, RF's performance may be limited for complex, multi-class problems such as RF-3, where more flexible architectures — including gradient boosting or deep learning approaches — might better capture the non-linear decision boundaries between spectrally similar classes. Although DL models could improve accuracy further, particularly when rich temporal information is available, they typically require substantially larger labelled datasets than were available here. This makes RF a pragmatic compromise under the present data conditions (Vanpoucke et al., 2024).

A 'flat' classification strategy was adopted at each RF level. RF-2 and RF-3 were applied sequentially within an RF-1 WCC mask. This propagates RF-1 errors into subsequent levels and does not constrain predictions to be consistent across the class hierarchy. For example, a parcel near the WCC/NWCC decision boundary may be assigned a detailed species composition class, even if the WCC label itself is uncertain. By contrast, hierarchical frameworks decompose the classification problem into a sequence of linked decisions that explicitly follow the class tree. This reduces error propagation and improves interpretability. Recent work has demonstrated that hierarchical approaches can substantially enhance overall and class-wise accuracy for WCC mapping compared to flat models, particularly for minority classes. Consistent with this, Vanongeval et al. (2026) reported superior performance for a hierarchical approach (OA = 0.64; κ = 0.59) versus a flat classification approach (OA = 0.59; κ = 0.53), with notably greater F1 scores for late-sown grass and winter rye. Within their hierarchical framework, the choice of covariates also influenced model performance: while monthly spectral composites effectively distinguished broad classes, parameters derived from curves fitted to VI time series improved classification at the mixture level (OA +21%; κ +47%), highlighting the importance of selecting a feature engineering approach tailored to the specific classification level. While hierarchical designs improve interpretability and flexibility, they typically require larger, more balanced training datasets and more sophisticated temporal feature engineering than are available in this study.

Combining pixel-based classification with a parcel-based majority filter produces coherent parcel maps while enabling a detailed representation of variability within fields. RF is known to be sensitive to class imbalance, spatial autocorrelation and sampling design (Belgiu & Drăguţ, 2016), and these factors also influence the present results. Compared with a purely parcel-based approach, using a pixel-level classification increased the effective number of samples used for training and validation, which is advantageous given the limited in-situ data available. Furthermore, parcel-wise splitting ensured that all pixels from a given field were assigned to either the training or validation set. This limited overly optimistic accuracies due to spatial autocorrelation. At the same time, the parcel-based majority filter reduced within-parcel variability by enforcing a single class label per field. This is consistent with how in-situ labels were collected (i.e. parcel-level descriptions), and therefore tends to increase OA and F1 scores for dominant classes, as seen across RF-1 and RF-2. However, this aggregation accentuates class imbalance, as minority and mixture classes, which already occur in only a few pixels, are more easily absorbed into larger neighbouring areas. This results in unstable and often near-zero F1 scores in RF-3, highlighting the trade-off between spatial coherence, 'nodata' reduction and class-wise balance. Additionally, the introduction of false positives and negatives in the classification must be considered, particularly for parcels with a high proportion of 'nodata' pixels, since filtering is applied to all parcels regardless of their internal 'nodata' ratio.

Repeated calibration and validation splits revealed substantial variation in classification performance between runs, with differences of up to 20 percentage points in some accuracy metrics. This variability can be attributed to several factors: the random allocation of a small number of minority-class parcels between the training and validation sets; the internal stochasticity of the RF algorithm through bootstrap sampling and random feature selection; the potential spatial clustering of spectrally similar parcels; and the occasional influence of outlier parcels with atypical spectral signatures. Parcel-level attributes, such as canopy height, percentage cover and within-field heterogeneity, could also affect the spectral signatures of WCC parcels and thus influence the representativeness of a given split for each class. A detailed analysis of parcel attributes for each split was beyond the scope of this study.

To ascertain whether this variability is primarily attributable to random partitioning noise or genuine sensitivity to the training data, an additional experiment was conducted for the S2 RF-1 model from Setup 2. This involved 30 independent parcel-wise calibration/validation splits, averaged over 10 RF realisations. The resulting distributions of OA, $F1_{macro}$ and class-wise F1 scores for WCC and NWCC show little variability (with standard deviations of around 0.04–0.06 percentage points and coefficients of variation below 0.1%), exhibiting narrow 95% confidence intervals and no systematic drift across repetitions (Appendix 22). This indicates that global and class-wise metrics for the binary WCC vs. NWCC task are highly stable with respect to the specific random split and RF initialisation. The test was not repeated for RF-2 and RF-3. However, the much stronger run-to-run fluctuations observed for certain mixture types and species compositions, together with the small number of parcels available for some classes, suggest that limited and imbalanced training data likely contributed to their instability.

5.7.2. Data and acquisition constraints

The single-date design of Setups 1 and 2 poses the greatest challenge in terms of acquiring data for RF-2 and RF-3. One December image cannot capture the phenological trajectories that distinguish WCC species and mixture types, as it only samples one point on class-specific growth curves which diverge most strongly in the weeks immediately after sowing, before converging again by mid-winter. However, the optimal acquisition timing likely differs between classification tasks. For RF-1 and RF-2, an EnMAP acquisition in November, when WCC canopies are at peak biomass and the divergence in phenology between pure and mixed parcels is still partially preserved, would have provided improved spectral separability compared to the December date used here. For RF-3, an acquisition prior to canopy closure may be more valuable: in the weeks after emergence, individual species within a mixture have not formed overlapping canopies yet, so their distinct spectral signatures are more likely to be detectable before canopy dominance homogenises the mixed signal. However, usable EnMAP acquisitions over the study area are limited by the sensor's narrow swath and modest revisit frequency, so this could not be guaranteed within the study period. This highlights a fundamental constraint of current spaceborne HS missions: the combination of limited spatial coverage (~5,000 km²/day) and limited revisit time means that obtaining multiple well-timed acquisitions over a specific agricultural region within a single growing season is not yet reliably achievable (Chabrillat et al., 2024).

For Setup 3, cloud cover is the main factor affecting the quality of the data. The number of cloud-free S2 observations is markedly lower in December and January than in other months, which reduces the quality of the composite during a potentially informative phenological window. For practical reasons, the time series was truncated around 20 March, which limits the representativeness of the March composite. While the monthly composite strategy effectively mitigated random cloud contamination and reduced noise for most months, it also smoothed short-lived phenological signals, such as differences in establishment timing driven by sowing dates, that could otherwise help to distinguish WCC species and sowing strategies. More sophisticated temporal aggregation approaches, such as decadal composites or curve-fitted phenological metrics derived from daily or near-daily time series, have been shown to preserve these transient signals more effectively and improve classification at finer levels (Vanpoucke et al., 2024; Vanongeval et al., 2026). Inter-annual variability presents an additional constraint that was not assessed in this single-season study. Vanongeval et al. (2026) found that model performance decreased substantially when applied to an omitted season, likely due to variation in species distributions and weather-driven phenology from one year to the next. This highlights the importance of using multi-year training data when building temporally robust WCC classifiers that can be used across seasons.

Another spatial constraint relates to the resolution mismatch between sensors. EnMAP's 30 m pixel integrates contributions from multiple surface types within a single parcel, such as WCC canopies, soil gaps, crop residues and inter-row vegetation, particularly in parcels with uneven establishment. Although S2's 10 m bands partially reduce this mixing effect for smaller spatial features, 20 m SWIR bands still suffer from sub-pixel mixing in heterogeneous parcels. Although an internal LPIS buffer was applied to reduce edge effects, within-parcel heterogeneity remains a source of spectral noise, particularly for nominally pure WCC stands with variable establishment density. Compared to studies using PlanetScope (3 m) or airborne imagery, the 10–30 m resolution of S2 and EnMAP inevitably limits the precision with which sub-parcel composition can be determined.

Classification performance is strongly influenced by the limited number of parcels available for several WCC classes and pronounced class imbalance, particularly in RF-2 and RF-3, across all setups. RF-1 benefits from a relatively large and balanced pool of WCC and NWCC parcels. However, the pure and mixed classes in RF-2 and the species composition classes in RF-3 are represented by small numbers of parcels. This results in unstable F1 scores and frequent runs with near-zero values for minority classes, such as mustard mixes and phacelia-mustard. The scarcity of training examples for rare classes likely constrains the ability of RF models to learn robust decision boundaries, and the uneven class distribution may have caused the classifier to favour dominant classes — a well-documented behaviour of RF under class imbalance (Belgiu & Drăguţ, 2016). Future work could address this issue by using oversampling, generating synthetic samples, or weighting the loss function to better represent underrepresented categories (Vanongeval et al., 2026).

A second key driver of performance variability is label uncertainty. The recorded label of in-situ data did not always accurately reflect the visible canopy at the time of satellite acquisition — particularly where sowing failures, patchy emergence or late-season management interventions had altered the composition of the canopy between the field campaign and the November/December acquisition. For the S2 monthly composites used in Setup 3, label validity could be confirmed only up to mid-January 2026. Parcels that ended before the end of the composite period in March 2026 could not be verified, which could introduce potential label-date mismatches in later monthly composites. Furthermore, many WCC fields exhibited significant variability within parcels in terms of stand density, species dominance and soil exposure. Consequently, parcel mean spectra or pixel samples did not always accurately represent a single homogeneous canopy type. Future field campaigns would benefit from collecting reference data at the spatial scale of the target sensor pixel rather than at the parcel level. This would provide ground truth at the pixel level that is directly matched to the satellite acquisition geometry. It would also avoid the implicit assumption of within-parcel homogeneity that currently propagates label noise into the model.

A final concern relates to the use of the LPIS database as the spatial reference framework. LPIS parcels may contain small non-crop elements, infrastructure, or management differences that are not reflected in the administrative polygons, which introduces additional spectral noise in parcel-based samples. Furthermore, the LPIS is subject to administrative restrictions that limit the direct transferability of this approach to regions where comparable datasets are unavailable (Leonhardt et al., 2024). Future work could explore using open-access reference datasets or RS parcel delineation methods to make the methodology more widely applicable (Vanongeval et al., 2026).

5.8. Implications and future directions

5.8.1. Operational relevance for CAP and PGDA monitoring in Wallonia

The results of this study have direct implications for the operational use of satellite-based WCC monitoring within Walloon agricultural policy frameworks, particularly for BCAE 8, PGDA and BCAE 6. The RF-1's consistent performance across all setups shows that S2-based monthly composites can effectively document the presence of winter soil cover at the parcel level across large areas at a low cost and without the need for field intervention. These products could complement farmer declarations and targeted field inspections by providing exhaustive spatial coverage. For example, they could be used to screen parcels with systematically low or absent winter cover, identify spatial clusters of potential non-compliance and evaluate regional adoption patterns over time. The November composite, which captures peak WCC biomass in the 2025–2026 growing season and produces the highest classification accuracy, suggests that this acquisition window aligns well with the regulatory requirement for WCC establishment before the end of autumn and is a practically promising candidate. However, the optimal composite month is likely to vary between years, depending on prevailing weather conditions, sowing dates and the pace of canopy development. Without a multi-year analysis, it is not possible to conclude that November consistently represents the optimal acquisition timing across growing seasons.

RF-2 shows moderate but consistent accuracy in distinguishing between pure and mixed classes. This suggests that S2 time series could also be used to roughly categorise WCC structural types, for example by identifying fields dominated by single-species monocultures versus more diverse mixed stands. This is important for evaluating the potential agro-environmental benefits of different WCC compositions. However, persistent difficulties in resolving detailed species composition (RF-3), combined with the strong dependence of F1 minority class scores on prevalence and management success, suggest that operational indicators should initially focus on binary presence/absence and a small number of robust mixture categories, rather than enforcing species-specific prescriptions based solely on satellite data. Fine-scale species composition will continue to require targeted in situ verification in the near term.

5.8.2. Added value of hyperspectral missions for future WCC monitoring

Experiments based on EnMAP demonstrate that single-date HSI can match or exceed S2 for the detection of WCC presence and, under harmonised conditions, provide at least comparable performance for classification of mixture types despite having a coarser spatial resolution. Therefore, the most important near-term contribution of HSI missions to WCC monitoring is not operational species-level mapping, but rather diagnosis: identifying which spectral regions carry the most information for discriminating WCC functional groups; how this information is distributed across the VNIR-SWIR range; and how it complements the temporal signals already accessible from S2 time series. These insights can inform the design of future multispectral band configurations and guide the selection of optimal acquisition windows for upcoming missions.

The main operational constraints of current spaceborne HS missions — limited spatial coverage and infrequent revisits — mean that obtaining multiple well-timed acquisitions over a specific agricultural region within a single growing season is not yet reliably achievable with EnMAP alone. Multi-mission HS fusion strategies that combine EnMAP with PRISMA, DESIS or EMIT partially mitigate this limitation, and have already been shown to be consistent and complementary for agricultural monitoring (Bourriz et al., 2025). As upcoming missions such as CHIME and PRISMA-SG will substantially increase the density of spaceborne HS data over time, combined MS-HS strategies could gradually move from experimental studies towards operational products (Mirzaei et al., 2024). For example, well-timed HS acquisitions could be used to characterise species-specific spectral signatures, which could then be transferred to denser S2 time series through fusion or transfer learning. These products would provide reliable indicators of dominant species composition and, ultimately, species-specific ecosystem services at a regional scale.

5.8.3. Priority directions for future research

Several concrete directions emerge from the constraints identified in this study:

- **Earlier and better-timed acquisitions.** Obtaining HSI and MSI two to five weeks after sowing — when phenological divergence between species is greatest — is the single most impactful improvement available for RF-2 and RF-3. A cloud-free EnMAP acquisition earlier in the growing season over the study area would be a first test of this hypothesis (Basinger et al., 2020).
- **Explicit phenological metrics.** As a first step, stacking all monthly composites into a single multi-temporal feature space — so that the classifier jointly exploits the full growing-season trajectory rather than each month in isolation — would provide a more direct test of whether the temporal signal is sufficient to close the gap with phenological-metric approaches. Replacing or supplementing monthly composites with curve-fitted phenological parameters —

such as start-of-season timing, peak NDVI date, and rates of spring greening and autumn senescence — has been shown improve mixture-level classification substantially and should be prioritised in future S2-based setups (Vanongeval et al., 2026; Vanpoucke et al., 2024).

- **Improved reference data and class design.** Redesigning the classification hierarchy around morphologically or agronomically meaningful functional groups would produce more spectrally coherent classes and reduce within-group variability. Replacing coarse categorical labels with continuous mixture proportions derived from field spectroscopy or high-resolution imagery would provide cleaner training signals across all three classification levels. Simultaneously, oversampling, synthetic sample generation, or loss-function weighting for underrepresented classes, combined with pixel-level labelling to increase effective sample sizes without additional field effort, would directly address the class imbalance that dominated RF-3 errors (Vanpoucke et al., 2024; Vanongeval et al., 2026).
- **Multi-year training data.** Expanding the reference dataset across multiple growing seasons is essential for building temporally robust classifiers that generalise across year-to-year variation in weather and management practices.
- **Spatial transferability.** All models were trained and validated within a single agricultural region. It remains unknown how well they would generalise to other areas with different soil types, topography, local cultivar choices, or agronomic practices. Testing the trained models on an independent AOI in Belgium, northern France, or Germany without retraining would provide a direct measure of spatial transferability.
- **Complementary data sources.** Integration of SAR data (e.g. S1) would improve temporal coverage under cloudy winter conditions and provide structural backscatter information complementary to optical reflectance (Jennewein et al., 2022). Thermal data have also shown promise for distinguishing bare soil from residue-covered fields in winter (Barnes et al., 2021). Sowing date records from farm management systems, where accessible, could substantially reduce the management-driven spectral variability that currently limits species-level classification (Vanongeval et al., 2026).
- **Advanced model architectures.** As larger labelled datasets become available, DL approaches could be tested against RF baselines (Vanpoucke et al., 2024).

CHAPTER

SIX

CONCLUSION

This thesis evaluated the potential of S2 MSI and EnMAP HSI for monitoring WCC in the Hesbaye region of Wallonia at three levels of detail: WCC presence; pure versus mixed WCC stands; and WCC species composition. The study benefited from a rich and diverse reference dataset collected synchronously with the satellite acquisitions, encompassing a wide range of WCC types. Three setups were implemented using a consistent RF framework: single-date EnMAP classification; a harmonised, single-date EnMAP–S2 comparison; and S2 monthly composites spanning the WCC growing season.

At the coarsest level, WCC presence was mapped with high and stable accuracy across all setups (OA > 80%, $F1_{macro}$ > 78%), confirming that both sensors provide a solid foundation for operational monitoring of winter soil cover. Restricting EnMAP input to the 20 most informative bands preserved RF-1 accuracy almost entirely. After resampling to S2-like bands, EnMAP outperformed S2 (OA = 86.84% vs. 80.72%, $F1_{macro}$ = 85.5% vs. 78.6%) for RF-1, suggesting that radiometric quality and acquisition conditions contribute as much as spectral sampling. S2 monthly composites achieved OA of 76–86%, and the November composite performed best, coinciding with peak WCC biomass. Together, these results demonstrate that WCC maps from current satellite missions can support CAP and PGDA compliance monitoring.

Distinguishing between pure and mixed WCC stands was more challenging. In all setups, RF-2 achieved moderate OA (67.8-72.76%) and significantly higher F1 scores for mixed stands than for pure stands, reflecting the spectral variability and limited sample size of pure parcels. These findings imply that, in an operational context, information on mixtures can be provided at a small number of robust levels (e.g. 'pure-like' versus 'mixed-like'), whereas more detailed mixture compositions should be interpreted with caution. EnMAP achieved slightly but non-significantly higher accuracy than S2 for RF-2.

Identifying the dominant species from the WCC canopy perspective remained the most challenging task for both sensors (OA: 39.48-44.52%, $F1_{macro}$: 18.58-29.09%). Although single-date EnMAP achieved moderate F1 scores for phacelia-dominated mixtures and pure mustard, it performed poorly for minority or heterogeneous classes. No systematic sensor advantage in Setup 2 was observed for RF-3; performance differences between EnMAP and S2 were dominated by class design and sample size rather than spectral characteristics.

Several limitations qualify the interpretation of these findings and highlight areas for future research. Class imbalance and the small number of parcels in several WCC categories (especially RF-2 and RF-3) led to unstable class-wise F1 scores and frequent classification errors for rare mixtures. Parcel-level attributes — including within-field heterogeneity, percentage cover, canopy height, species dominance, and management practices such as sowing date and termination timing — vary considerably within a single class, meaning that parcels sharing the same label can produce markedly different spectral signatures depending on their agronomic state and management history at the time of acquisition. Reliance on a single EnMAP acquisition and one S2 acquisition in Setups 1 and 2 restricted the range of phenological conditions captured and is arguably the most consequential limitation for RF-2 and RF-3: a November acquisition — coinciding with peak WCC biomass — would likely improve mixture-type discrimination, while an even earlier acquisition, prior to canopy closure, may be more valuable for species composition mapping, as individual species signatures are most distinct before canopy dominance homogenises the mixed signal. Monthly compositing in Setup 3 smoothed short-lived dynamics that could otherwise help distinguish sowing strategies or closely related mixtures, and a time-series approach based on explicit phenological metrics rather than composited reflectance would better capture the seasonal trajectories that differentiate WCC types.

Despite these constraints, the thesis presents a concrete prototype for regional WCC monitoring under the CAP and the PGDA. It demonstrates that S2 time series can deliver reliable maps of WCC presence and broad mixture types in Wallonia and that EnMAP-like imaging spectroscopy can improve the detection of WCC presence on a single date and assist in identifying spectrally informative regions for functional group separation. The results also clarify the limitations of current satellite-based products, particularly regarding detailed mixture and species composition. At the species composition level specifically, redesigning the classification hierarchy around spectrally coherent functional groups rather than compositional identity, combined with earlier acquisition timing prior to canopy closure, represents the most tractable path towards meaningful accuracy gains.

Future research should concentrate on extending the analysis to multi-year, multi-date S2 time series and additional EnMAP-like acquisitions. This could allow the derivation of explicit phenological metrics and biomass-related variables that better capture differences in sowing strategies, growth trajectories, and termination behaviour. Improving the reference data in terms of sample size, class balance, parcel delineation, and within-parcel descriptors would stabilise mixture and species-level classifications, enabling more sophisticated hierarchical models. Finally, rather than using MS and HS observations independently as in this study, combining them within an integrated workflow — for example by using well-timed HS acquisitions to characterise species-specific spectral signatures that are then transferred to denser S2 time series through fusion or transfer learning — offers a realistic path towards operational monitoring systems that provide reliable information on winter soil cover and WCC management in Wallonia and comparable regions, thereby supporting the design and evaluation of agro-environmental schemes.

References

Adetunji, A. T., Ncube, B., Mulidzi, R., & Lewu, F. B. (2020). Management impact and benefit of cover crops on soil quality: A review. *Soil and Tillage Research*, *204*, 104717. https://doi.org/10.1016/j.still.2020.104717

Ahmed, Z., Nalley, L., Brye, K., Steven Green, V., Popp, M., Shew, A. M., & Connor, L. (2023). Winter-time cover crop identification: A remote sensing-based methodological framework for new and rapid data generation. *International Journal of Applied Earth Observation and Geoinformation*, *125*, 103564. https://doi.org/10.1016/j.jag.2023.103564

Alami Machichi, M., Mansouri, L. E., Imani, Y., Bourja, O., Lahlou, O., Zennayi, Y., Bourzeix, F., Hanadé Houmma, I., & Hadria, R. (2023). Crop mapping using supervised machine learning and deep learning: A systematic literature review. *International Journal of Remote Sensing*, *44*(8), 2717–2753. https://doi.org/10.1080/01431161.2023.2205984

Albertz, J. (2009). *Einführung in die Fernerkundung: Grundlagen der Interpretation von Luft- und Satellitenbildern* (4., aktualisierte Aufl). WBG (Wissenschaftliche Buchgesellschaft).

Aneece, I. P., Thenkabail, P. S., Lyon, J. G., Huete, A., & Slonecker, T. (2019). Spaceborne hyperspectral EO-1 Hyperion data pre-processing: Methods, approaches, and algorithms. In P. S. Thenkabail, J. G. Lyon, & A. Huete (Eds.), *Hyperspectral remote sensing of vegetation, Volume I: Fundamentals, sensor systems, spectral libraries, and data mining for vegetation* (2nd ed., pp. 251–272). CRC Press, Taylor & Francis Group.

Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, *13*(1–2), 95–120. https://doi.org/10.1080/02757259509532298

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R. J., Li, H., & Moran, M. S. (2000). COINCIDENT DETECTION OF CROP WATER STRESS, NITROGEN STATUS AND CANOPY DENSITY USING GROUND-BASED MULTISPECTRAL DATA. *Proceedings of the Fifth International Conference on Precision Agriculture*, *1619*.

Barnes, M. L., Yoder, L., & Khodaee, M. (2021). Detecting Winter Cover Crops and Crop Residues in the Midwest US Using Machine Learning Classification of Thermal and Optical Imagery. *Remote Sensing*, *13*(10), 1998. https://doi.org/10.3390/rs13101998

Basinger, N. T., Jennings, K. M., Hestir, E. L., Monks, D. W., Jordan, D. L., & Everman, W. J. (2020). Phenology affects differentiation of crop and weed species using hyperspectral remote sensing. *Weed Technology*, *34*(6), 897–908. https://doi.org/10.1017/wet.2020.92

Belcam. (2026). *Belcam—BELgian Collaborative Agriculture Monitoring*. The Project. https://belcam.info/

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bostan, S., Ortak, M. A., Tuna, C., Akoguz, A., Sertel, E., & Berk Ustundag, B. (2016). Comparison of classification accuracy of co-located hyperspectral & multispectral images for agricultural purposes. *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–4. https://doi.org/10.1109/Agro-Geoinformatics.2016.7577671

Bourriz, M., Laamrani, A., El-Battay, A., Hajji, H., Elbouanani, N., Ait Abdelali, H., Bourzeix, F., Amazirh, A., & Chehbouni, A. (2025, March 15). *An Intercomparison of Two Satellite-Based Hyperspectral Imagery (PRISMA & EnMAP) for Agricultural Mapping: Potential of these sensors to produce hyperspectral time-series essential for tracking crop phenology and enhancing crop type mapping*. https://doi.org/10.5194/egusphere-egu25-18417

Breiman, L. (2001). Random Forests. *Machine Learning*, (45), 5–32. https://doi.org/10.1023/A:1010933404324

Brosinsky, A., Wilczok, C., & Foerster, S. (2022). *Beyond the Visible – Introduction to Hyperspectral Remote Sensing. Offline document accompanying the MOOC* (p. 131 pages, 6 MB) [Application/pdf]. GFZ Data Services. https://doi.org/10.48440/ENMAP.2022.003

CDSE, C. D. S. (n.d.). *Sentinel-2 | Copernicus Data Space Ecosystem*. https://dataspace.copernicus.eu/data-collections/copernicus-sentinel-data/sentinel-2

Chabrillat, S., Foerster, S., Segl, K., Beamish, A., Brell, M., Asadzadeh, S., Milewski, R., Ward, K. J., Brosinsky, A., Koch, K., Scheffler, D., Guillaso, S., Kokhanovsky, A., Roessner, S., Guanter, L., Kaufmann, H., Pinnel, N., Carmona, E., Storch, T., … Fischer, S. (2024). The EnMAP spaceborne imaging spectroscopy mission: Initial scientific results two years after launch. *Remote Sensing of Environment*, *315*, 114379. https://doi.org/10.1016/j.rse.2024.114379

Chabrillat, S., Guanter, L., Kaufmann, H., Foerster, S., Beamish, A., Brosinsky, A., Wulf, H., Asadzadeh, S., Bochow, M., Bohn, N., Boesche, N., Bracher, A., Brell, M., Buddenbaum, H., Cerra, D., Fischer, S., Hank, T., Heiden, U., Heim, B., … Segl, K. (2022). EnMAP Science Plan. In *EnMAP Technical Report;* (p. 87 pages, 2 MB) [Pdf]. GFZ Data Services. https://doi.org/10.48440/ENMAP.2022.001

Chapagain, T., Lee, E. A., & Raizada, M. N. (2020). The Potential of Multi-Species Mixtures to Diversify Cover Crop Benefits. *Sustainability*, *12*(5), 2058. https://doi.org/10.3390/su12052058

Chartin, C., Stevens, A., Goidts, E., Krüger, I., Carnol, M., & Van Wesemael, B. (2017). Mapping Soil Organic Carbon stocks and estimating uncertainties at the regional scale following a legacy sampling strategy (Southern Belgium, Wallonia). *Geoderma Regional*, *9*, 73–86. https://doi.org/10.1016/j.geodrs.2016.12.006

Chen, W., Liu, L., Zhang, C., Wang, J., Wang, J., & Pan, Y. (2004). Monitoring the seasonal bare soil areas in beijing using multi-temporal TM images. *IEEE International IEEE International IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004*, *5*, 3379–3382. https://doi.org/10.1109/IGARSS.2004.1370429

Clark, M. L., Roberts, D. A., & Clark, D. B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, *96*(3–4), 375–398. https://doi.org/10.1016/j.rse.2005.03.009

Copernicus. (n.d.). *S2 Mission*. sentiwiki. https://sentiwiki.copernicus.eu/web/s2-mission

Couëdel, A., Alletto, L., Tribouillois, H., & Justes, É. (2018). Cover crop crucifer-legume mixtures provide effective nitrate catch crop and nitrogen green manure ecosystem services. *Agriculture, Ecosystems & Environment*, *254*, 50–59. https://doi.org/10.1016/j.agee.2017.11.017

Cruz-Ramírez, M., Hervás-Martínez, C., Jurado-Expósito, M., & López-Granados, F. (2012). A multi-objective neural network based method for cover crop identification from remote sensed data. *Expert Systems with Applications*, *39*(11), 10038–10048. https://doi.org/10.1016/j.eswa.2012.02.046

De Lange, N. (2020). *Geoinformatik in Theorie und Praxis: Grundlagen von Geoinformationssystemen, Fernerkundung und digitaler Bildverarbeitung*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-60709-1

Do Nascimento Bendini, H., Fieuzal, R., Carrere, P., Clenet, H., Galvani, A., Allies, A., & Ceschia, É. (2024). Estimating Winter Cover Crop Biomass in France Using Optical Sentinel-2 Dense Image Time Series and Machine Learning. *Remote Sensing*, *16*(5), 834. https://doi.org/10.3390/rs16050834

Dudley, K. L., Dennison, P. E., Roth, K. L., Roberts, D. A., & Coates, A. R. (2015). A multi-temporal spectral library approach for mapping vegetation species across spatial and temporal phenological gradients. *Remote Sensing of Environment*, *167*, 121–134. https://doi.org/10.1016/j.rse.2015.05.004

Earth Observation Center of DLR. (n.d.). *Mission—EnMAP*. https://www.enmap.org/mission/

EnMAP. (2023). *EnMAP_Spectral_Bands_update.xlsx*. EnMAP Spectral Bands. https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww.enmap.org%2Fdata%2Fdoc%2FEnMAP_Spectral_Bands_update.xlsx&wdOrigin=BROWSELINK

EnMAP. (n.d.). *EnMAP HSI Instrument Specification*. EnMAP. EnMAP_Specs. https://www.enmap.org/data/doc/EnMAP_Specs.pdf

European Comission. (2022). *New Common Agricultural Policy: Set for 1 January 2023*. European Commission - European Commission. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7639

European Comission. (2025, November 10). *Belgium (Wallonia)—Agriculture and rural development*. https://agriculture.ec.europa.eu/cap-my-country/cap-strategic-plans/belgium-wallonia_en

European Commission. Joint Research Centre. (2019). *Adoption of cover crops for climate change mitigation in the EU.* Publications Office. https://doi.org/10.2760/638382

Eurostat. (2025). *Soil cover by arable land area, farmtype, crop rotation and NUTS 2 region* [Data set]. Eurostat. https://doi.org/10.2908/EF_MP_SOIL

Fan, X., Vrieling, A., Muller, B., & Nelson, A. (2020). Winter cover crops in Dutch maize fields: Variability in quality and its drivers assessed from multi-temporal Sentinel-2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, *91*, 102139. https://doi.org/10.1016/j.jag.2020.102139

Fiorini, A., Remelli, S., Boselli, R., Mantovi, P., Ardenti, F., Trevisan, M., Menta, C., & Tabaglio, V. (2022). Driving crop yield, soil organic C pools, and soil biodiversity with selected winter cover crops under no-till. *Soil and Tillage Research*, *217*, 105283. https://doi.org/10.1016/j.still.2021.105283

Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *82*, 83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007

Galvão, L. S., Epiphanio, J. C. N., Breunig, F. M., & Formaggio, A. R. (2019). Crop type discrimination using hyperspectral data: Advances and perspectives. In P. S. Thenkabail, J. G. Lyon, & A. Huete (Eds.), *Hyperspectral remote sensing of vegetation, Volume III: Biophysical and biochemical characterization and plant species studies* (2nd ed., pp. 183–210). CRC Press, Taylor & Francis Group.

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Géoportail de la Wallonie. (2025). *Parcellaire agricole issu du SIGEC - dernière situation disponible*. Géoportail de La Wallonie - Le Site de l'information Géographique Wallonne. https://geoportail.wallonie.be/catalogue/342e11da-e4bf-4977-b1bc-8e80211e068d.html

Gerstmann, H., Gläßer, C., Thürkow, D., & Möller, M. (2018). Detection of Phenology-Defined Data Acquisition Time Frames For Crop Type Mapping. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, *86*(1), 15–27. https://doi.org/10.1007/s41064-018-0043-6

Ghazaryan, G., Dubovyk, O., Löw, F., Lavreniuk, M., Kolotii, A., Schellberg, J., & Kussul, N. (2018). A rule-based approach for crop identification using multi-temporal and multi-sensor phenological metrics. *European Journal of Remote Sensing*, *51*(1), 511–524. https://doi.org/10.1080/22797254.2018.1455540

Gobin, A., Sallah, A.-H. M., Curnel, Y., Delvoye, C., Weiss, M., Wellens, J., Piccard, I., Planchon, V., Tychon, B., Goffart, J.-P., & Defourny, P. (2023). Crop Phenology Modelling Using Proximal and Satellite Sensor Data. *Remote Sensing*, *15*(8), 2090. https://doi.org/10.3390/rs15082090

Goffart, D., Curnel, Y., Planchon, V., Goffart, J.-P., & Defourny, P. (2021). Field-scale assessment of Belgian winter cover crops biomass based on Sentinel-2 data. *European Journal of Agronomy*, *126*, 126278. https://doi.org/10.1016/j.eja.2021.126278

Grandgirard, D., & Zielinski, R. (2008). Land Parcel Identification System (LPIS) anomalies' sampling and spatial pattern: Towards convergence of ecological methodologies and GIS technologies. *JRC European Commission*. https://doi.org/10.2788/91912

Hagolle, O., Huc, M., Desjardins, C., Auer, S., & Richter, R. (2017). *MAJA ATBD Algorithm Theoretical Basis Document*. CNES, CESBIO, DLR.

Heller, O., Bene, C. D., Nino, P., Huyghebaert, B., Arlauskienė, A., Castanheira, N. L., Higgins, S., Horel, A., Kir, A., Kizeková, M., Lacoste, M., Munkholm, L. J., O'Sullivan, L., Radzikowski, P., Rodríguez-Cruz, M. S., Sandén, T., Šarūnaitė, L., Seidel, F., Spiegel, H., … Vanwindekens, F. (2024). Towards enhanced adoption of soil-improving management practices in Europe. *European Journal of Soil Science*, *75*(2), e13483. https://doi.org/10.1111/ejss.13483

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

IRM. (2026). *Climat de la Belgique*. KMI. https://www.meteo.be/fr/climat/climat-de-la-belgique/mois-apres-mois/contexte-general

Jennewein, J. S., Hively, W., Lamb, B. T., Daughtry, C. S. T., Thapa, R., Thieme, A., Reberg-Horton, C., & Mirsky, S. (2024). Spaceborne imaging spectroscopy enables carbon trait estimation in cover crop and cash crop residues. *Precision Agriculture*, *25*(5), 2165–2197. https://doi.org/10.1007/s11119-024-10159-4

Jennewein, J. S., Lamb, B. T., Hively, W. D., Thieme, A., Thapa, R., Goldsmith, A., & Mirsky, S. B. (2022). Integration of Satellite-Based Optical and Synthetic Aperture Radar Imagery to Estimate Winter Cover Crop Performance in Cereal Grasses. *Remote Sensing*, *14*(9), 2077. https://doi.org/10.3390/rs14092077

Kaufmann, H., Förster, S., Wulf, H., Segl, K., Guanter, L., Bochow, M., Heiden, U., Müller, A., Heldens, W., Schneiderhan, T., Leitão, P. J., Van Der Linden, S., Hostert, P., Hill, J., Buddenbaum, H., Mauser, W., Hank, T., Krasemann, H., Röttgers, R., … EnMAP Consortium. (2012). Science Plan 2012 of the Environmental Mapping and Analysis Program (EnMAP). *EnMAP Technical Report*. https://doi.org/10.2312/ENMAP.2015.013

Kc, K., Zhao, K., Romanko, M., & Khanal, S. (2021). Assessment of the Spatial and Temporal Patterns of Cover Crops Using Remote Sensing. *Remote Sensing*, *13*(14), 2689. https://doi.org/10.3390/rs13142689

Khangura, R., Ferris, D., Wagg, C., & Bowyer, J. (2023). Regenerative Agriculture—A Literature Review on the Practices and Mechanisms Used to Improve Soil Health. *Sustainability*, *15*(3), 2338. https://doi.org/10.3390/su15032338

Kharel, T. P., Bhandari, A. B., Mubvumba, P., Tyler, H. L., Fletcher, R. S., & Reddy, K. N. (2023). Mixed-Species Cover Crop Biomass Estimation Using Planet Imagery. *Sensors*, *23*(3), 1541. https://doi.org/10.3390/s23031541

Köppen, W. (1936). Das geographische System der Klimate. In: Köppen, W., Geiger, R., (Hrsg): *Handbuch der Klimatologie* (Bd. 1: Teil C). Bornträger.

Kumar, L., Schmidt, K., Dury, S., & Skidmore, A. (2001). Imaging spectrometry and vegetation science. In F. D. Van der Meer & S. M. De Jong (Eds.), *Imaging spectrometry: Basic principles and prospective applications* (pp. 111–155). Springer.

Leonhardt, H., Wesemeyer, M., Eder, A., Hüttel, S., Lakes, T., Schaak, H., Seifert, S., & Wolff, S. (2024). Use cases and scientific potential of land use data from the EU's Integrated Administration and Control System: A systematic mapping review. *Ecological Indicators*, *167*, 112709. https://doi.org/10.1016/j.ecolind.2024.112709

Li, Q., Cao, X., Jia, K., Zhang, M., & Dong, Q. (2014). Crop type identification by integration of high-spatial resolution multispectral data with features extracted from coarse-resolution time-series vegetation index data. *International Journal of Remote Sensing*, *35*(16), 6076–6088. https://doi.org/10.1080/01431161.2014.943325

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2015). *Remote sensing and image interpretation* (Seventh edition). John Wiley & Sons, Inc.

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. *Advances in Neural Information Processing Systems*, (26).

Mirzaei, S., Pascucci, S., Carfora, M. F., Casa, R., Rossi, F., Santini, F., Palombo, A., Laneve, G., & Pignatti, S. (2024). Early-Season Crop Mapping by PRISMA Images Using Machine/Deep Learning Approaches: Italy and Iran Test Cases. *Remote Sensing*, *16*(13), 2431. https://doi.org/10.3390/rs16132431

Misra, G., Cawkwell, F., & Wingler, A. (2020). Status of Phenological Research Using Sentinel-2 Data: A Review. *Remote Sensing*, *12*(17), 2760. https://doi.org/10.3390/rs12172760

Najem, S., Baghdadi, N., Bazzi, H., Lalande, N., & Bouchet, L. (2024). Detection and Mapping of Cover Crops Using Sentinel-1 SAR Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 1446–1461. https://doi.org/10.1109/JSTARS.2023.3337989

Papp, L., Van Leeuwen, B., Szilassi, P., Tobak, Z., Szatmári, J., Árvai, M., Mészáros, J., & Pásztor, L. (2021). Monitoring Invasive Plant Species Using Hyperspectral Remote Sensing Data. *Land*, *10*(1), 29. https://doi.org/10.3390/land10010029

Python Software Foundation. (2026). *Welcome to Python.org*. Python.Org. https://www.python.org/psf-landing/

QField. (n.d.). *QField for QGIS*. QField - Efficient Field Work Built for QGIS. https://qfield.org/

Qian, S.-E. (2021). Hyperspectral Satellites, Evolution, and Development History. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 7032–7056. https://doi.org/10.1109/JSTARS.2021.3090256

Qiu, S., Zhu, Z., & He, B. (2019). Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sensing of Environment*, *231*, 111205. https://doi.org/10.1016/j.rse.2019.05.024

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, *14*(1), 6086. https://doi.org/10.1038/s41598-024-56706-x

Ram, B. G., Oduor, P., Igathinathane, C., Howatt, K., & Sun, X. (2024). A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Computers and Electronics in Agriculture*, *222*, 109037. https://doi.org/10.1016/j.compag.2024.109037

Roberto, C., Lorenzo, B., Michele, M., Micol, R., Cinzia, P. (2019). Optical remote sensing of vegetation water content. In P. S. Thenkabail, J. G. Lyon, & A. Huete (Eds.), *Hyperspectral remote sensing of vegetation, Volume II: Hyperspectral indices and image classifications for agriculture and vegetation* (2nd ed., pp. 183–200). CRC Press, Taylor & Francis Group.

Roberts, D. A., Roth, K. L., Wetherley, E. B., Meerdink, S. K., & Perroy, R. L. (2019). Hyperspectral vegetation indices. In P. S. Thenkabail, J. G. Lyon, & A. Huete (Eds.), *Hyperspectral remote sensing of vegetation, Volume II: Hyperspectral indices and image classifications for agriculture and vegetation* (2nd ed., pp. 3–26). CRC Press, Taylor & Francis Group.

Rouse., J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). *MONITORING VEGETATION SYSTEMS IN THE GREAT PLAINS WITH ERTS*.

Scheffler, D., Brell, M., Bohn, N., Alvarado, L., Soppa, M. A., Segl, K., Bracher, A., & Chabrillat, S. (2023). EnPT – an Alternative Pre-Processing Chain for Hyperspectral EnMAP Data. *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 7416–7418. https://doi.org/10.1109/IGARSS52108.2023.10281805

Schmidt, R., Gravuer, K., Bossange, A. V., Mitchell, J., & Scow, K. (2018). Long-term use of cover crops and no-till shift soil microbial community life strategies in agricultural soil. *PLOS ONE*, *13*(2), e0192953. https://doi.org/10.1371/journal.pone.0192953

Schulz, C., Holtgrave, A.-K., & Kleinschmit, B. (2021). Large-scale winter catch crop monitoring with Sentinel-2 time series and machine learning–An alternative to on-site controls? *Computers and Electronics in Agriculture*, *186*, 106173. https://doi.org/10.1016/j.compag.2021.106173

Seifert, C. A., Azzari, G., & Lobell, D. B. (2019). Corrigendum: Satellite detection of cover crops and their effects on crop yield in the Midwestern United States (2018 *Environ. Res. Let.* 13 064033). *Environmental Research Letters*, *14*(3), 039501. https://doi.org/10.1088/1748-9326/aaf933

Sharma, P., Singh, A., Kahlon, C. S., Brar, A. S., Grover, K. K., Dia, M., & Steiner, R. L. (2018). The Role of Cover Crops towards Sustainable Soil Health and Agriculture—A Review Paper. *American Journal of Plant Sciences*, *09*(09), 1935–1951. https://doi.org/10.4236/ajps.2018.99140

SPW. (2025). *BCAE 6 Couverture des sols minimale en vue d'éviter les sols nus dans les périodes les plus sensibles (Nouveauté 2025)—Portail de l'agriculture wallonne*. Agriculture en Wallonie. https://agriculture.wallonie.be/home/aides/pac-2023-2027-description-des-interventions/conditionnalite-renforcee-nouveaute-2025/bcae-6-couverture-des-sols-minimale-en-vue-d-eviter-les-sols-nus-dans-les-periodes-les-plus-sensibles.html

Tagliabue, G., Boschetti, M., Bramati, G., Candiani, G., Colombo, R., Nutini, F., Pompilio, L., Rivera-Caicedo, J. P., Rossi, M., Rossini, M., Verrelst, J., & Panigada, C. (2022). Hybrid retrieval of crop traits from multi-temporal PRISMA hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *187*, 362–377. https://doi.org/10.1016/j.isprsjprs.2022.03.014

Tan, Y., Gu, J., Lu, L., Zhang, L., Huang, J., Pan, L., Lv, Y., Wang, Y., & Chen, Y. (2025). Hyperspectral Band Selection for Crop Identification and Mapping of Agriculture. *Remote Sensing*, *17*(4), 663. https://doi.org/10.3390/rs17040663

Thenkabail, P. S. (2002). Optimal hyperspectral narrowbands for discriminating agricultural crops. *Remote Sensing Reviews*, *20*(4), 257–291. https://doi.org/10.1080/02757250109532439

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, *91*(3–4), 354–376. https://doi.org/10.1016/j.rse.2004.03.013

Thenkabail, P. S., Lyon, J. G., & Huete, A. (2019). Advances in hyperspectral remote sensing of vegetation and agricultural crops. In P. S. Thenkabail, J. G. Lyon, & A. Huete (Eds.), *Hyperspectral remote sensing of vegetation, Volume I: Fundamentals, sensor systems, spectral libraries, and data mining for vegetation* (2nd ed., pp. 3–40). CRC Press, Taylor & Francis Group.

Tsai, F., & Philpot, W. (1998). *Derivative Analysis of Hyperspectral Data*. Remote Sensing of Environment, 66, 41-51. http://dx.doi.org/10.1016/S0034-4257(98)00032-7

Van der Meer, F. D. (2001). Basic physics of spectrometry. In F. D. Van der Meer & S. M. De Jong (Eds.), *Imaging spectrometry: Basic principles and prospective applications* (pp. 3-16). Springer.

Vanongeval, F., Orshoven, J. V., & Gobin, A. (2026). Mapping of cropland winter soil cover in northern Belgium using Sentinel-2 time series. *Computers and Electronics in Agriculture*, *241*, 111254. https://doi.org/10.1016/j.compag.2025.111254

Vanpoucke, K., Heremans, S., Buls, E., & Somers, B. (2024). Field-Level Classification of Winter Catch Crops Using Sentinel-2 Time Series: Model Comparison and Transferability. *Remote Sensing*, *16*(24), 4620. https://doi.org/10.3390/rs16244620

Waldner, F., Lambert, M.-J., Li, W., Weiss, M., Demarez, V., Morin, D., Marais-Sicre, C., Hagolle, O., Baret, F., & Defourny, P. (2015). Land Cover and Crop Type Classification along the Season Based on Biophysical Variables Retrieved from Multi-Sensor High-Resolution Time Series. *Remote Sensing*, *7*(8), 10400–10424. https://doi.org/10.3390/rs70810400

Wang, Q., Jin, J., Sonobe, R., & Chen, J. M. (2019). Derivative hyperspectral vegetation indices in characterizing forest biophysical and biochemical quantities. In P. S. Thenkabail, J. G. Lyon, & A. Huete (Eds.), *Hyperspectral remote sensing of vegetation, Volume II: Hyperspectral indices and image classifications for agriculture and vegetation* (2nd ed., pp. 27–66). CRC Press, Taylor & Francis Group.

Wang, S., Guan, K., Zhang, C., Jiang, C., Zhou, Q., Li, K., Qin, Z., Ainsworth, E. A., He, J., Wu, J., Schaefer, D., Gentry, L. E., Margenot, A. J., & Herzberger, L. (2023). Airborne hyperspectral imaging of cover crops through radiative transfer process-guided machine learning. *Remote Sensing of Environment*, *285*, 113386. https://doi.org/10.1016/j.rse.2022.113386

Weiss, M., Baret, F., & Jay, S. (2020). *S2ToolBox Level 2 products: LAI, FAPAR, FCOVER*. (V2.1).

Wolf, D. D., Carson, E. W., & Brown, R. H. (1972). Leaf Area Index and Specific Leaf Area Determinations. *Jour. of Agron. Educ.*, *1*.

Wu, F., Maleki, R., Oubara, A., Gómez, D., Eftekhari, A., & Yang, G. (2023). Machine Learning Approaches for Crop Identification from Remote Sensing Imagery: A Review. In A. Abraham, T. Hanne, N. Gandhi, P. Manghirmalani Mishra, A. Bajaj, & P. Siarry (Eds), *Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022)* (Vol. 648, pp. 325–336). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-27524-1_31

Xia, Y., Guan, K., Copenhaver, K., & Wander, M. (2021). Estimating cover crop biomass nitrogen credits with Sentinel-2 imagery and sites covariates. *Agronomy Journal*, *113*(2), 1084–1101. https://doi.org/10.1002/agj2.20525

Yao, H., Huang, Y., Tang, L., Tian, L., Bhatnagar, D., & Cleveland, T. E. (2019). Using hyperspectral data in precision farming applications. In P. S. Thenkabail, J. G. Lyon, & A. Huete (Eds.), *Hyperspectral remote sensing of vegetation, Volume IV: Advanced applications in remote sensing of agricultural crops and natural vegetation* (2nd ed., pp. 3–36). CRC Press, Taylor & Francis Group.

Zhai, H., Zhang, H., Zhang, L., & Li, P. (2018). Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *144*, 235–253. https://doi.org/10.1016/j.isprsjprs.2018.07.006

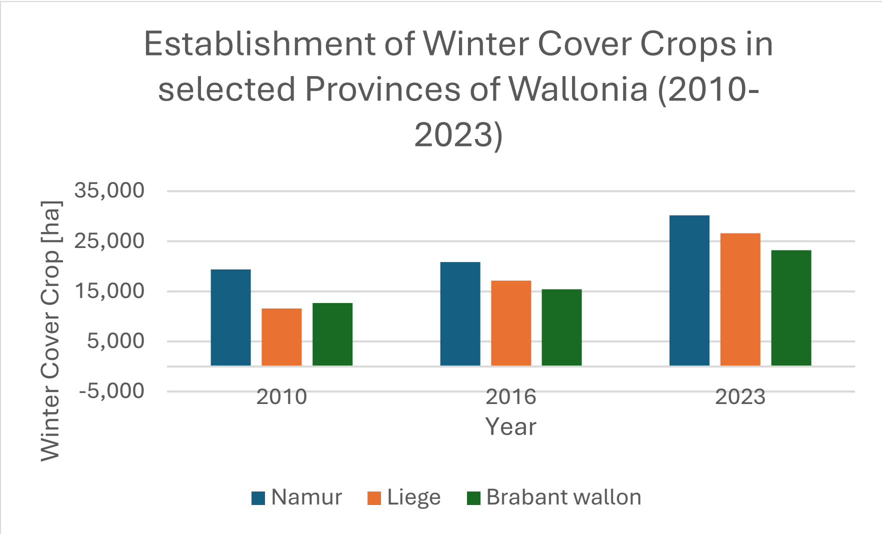
Zhang, J., Xu, Y., Yao, F., Wang, P., Guo, W., Li, L., & Yang, L. (2010). Advances in estimation methods of vegetation water content based on optical remote sensing techniques. *Science China Technological Sciences*, *53*(5), 1159–1167. https://doi.org/10.1007/s11431-010-0131-3

Zhou, Q., Guan, K., Wang, S., Wu, X., Stroebel, S., & Hipple, J. (2026). Multi-sensor assessment of phenology-based field-level cover cropping detection using satellite vegetation time series from Harmonized Landsat-8 and Sentinel-2, MODIS, and PlanetScope. *International Journal of Applied Earth Observation and Geoinformation*, *146*, 105004. https://doi.org/10.1016/j.jag.2025.105004

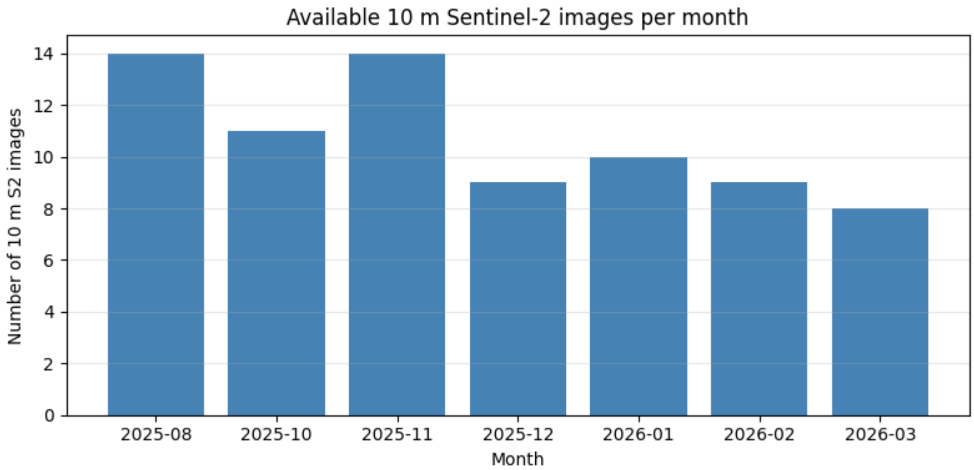
Zhou, Y., Ferdinand, M. S., Van Wesemael, J., Dvorakova, K., Baret, P. V., Van Oost, K., & Van Wesemael, B. (2025). A framework for mapping conservation agricultural fields using optical and radar time series imagery. *Remote Sensing of Environment*, *328*, 114858. https://doi.org/10.1016/j.rse.2025.114858

Appendix

Appendix 1. Evolution of WCC establishment in selected provinces of Wallonia between 2010 and 2023 (Data: eurostat.eu).



Appendix 2. Available S2 images per month (August 2025-March 2026).



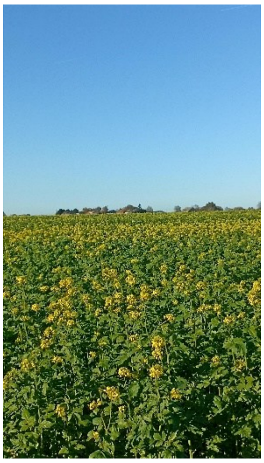
Appendix 3. S2 MSI spectral bands at 10 and 20 m spatial resolution (CDSE, n.d.).

Band Name	Band ID	Bandwidth [nm]	Spatial Resolution [m]
Blue	B2	457.5-522.5	10
Green	B3	542.5-577.5	10
Red	B4	650-680	10
Red-edge 1	B5	698.5-712.5	20
Red-edge 2	B6	732.5-747.5	20
Red-edge 3	B7	773-793	20
Near-Infrared	B8	784.5-899.5	10
Near-Infrared	B8A	855-875	20
Shortwave Infrared (SWR1)	B11	1,565-1,655	20
Shortwave Infrared (SWR2)	B12	2,100-2,280	20

Appendix 4. Additional Field Photographs from In-Situ Campaigns (November 2025)

WCC

Mustard



Phacelia + Mustard + Fodder radish, + Ravanelle + Clover + Vetch

Mixed:

Phacelia + Mustard

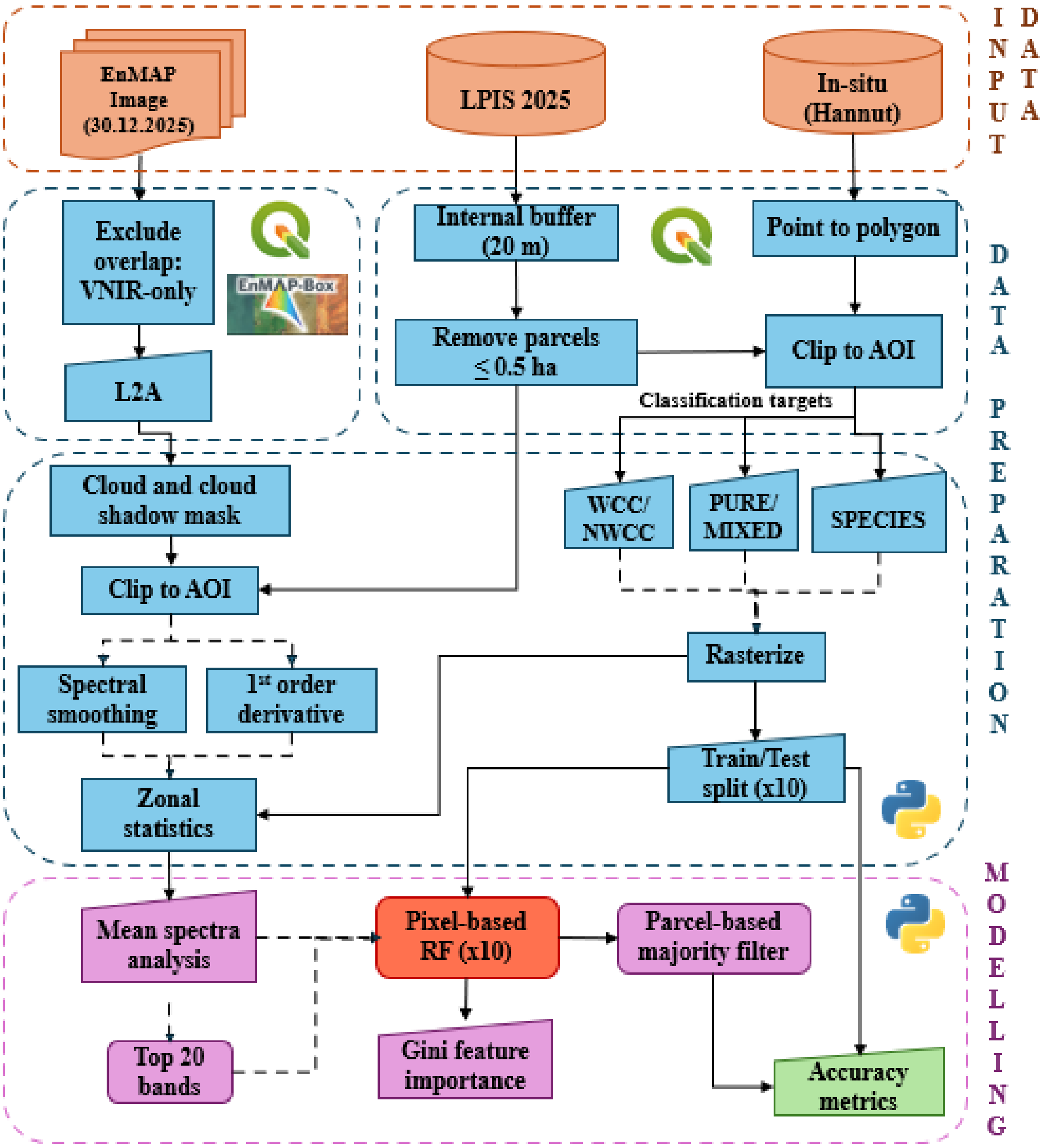
Phacelia + Cabbage + Nyger

NWCC

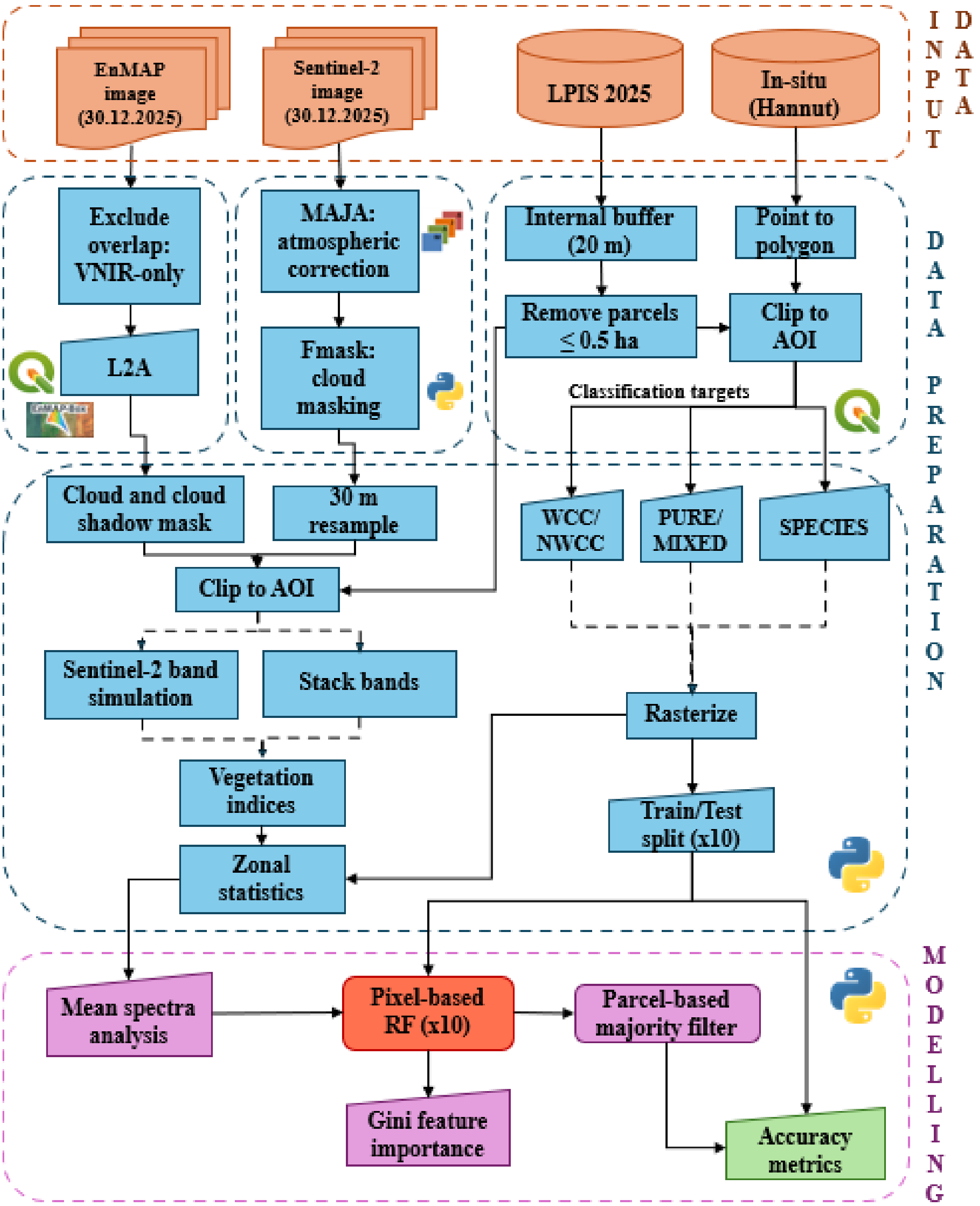
Winter wheat emergence with important maize residues

Winter wheat sowing with previous deep plowing

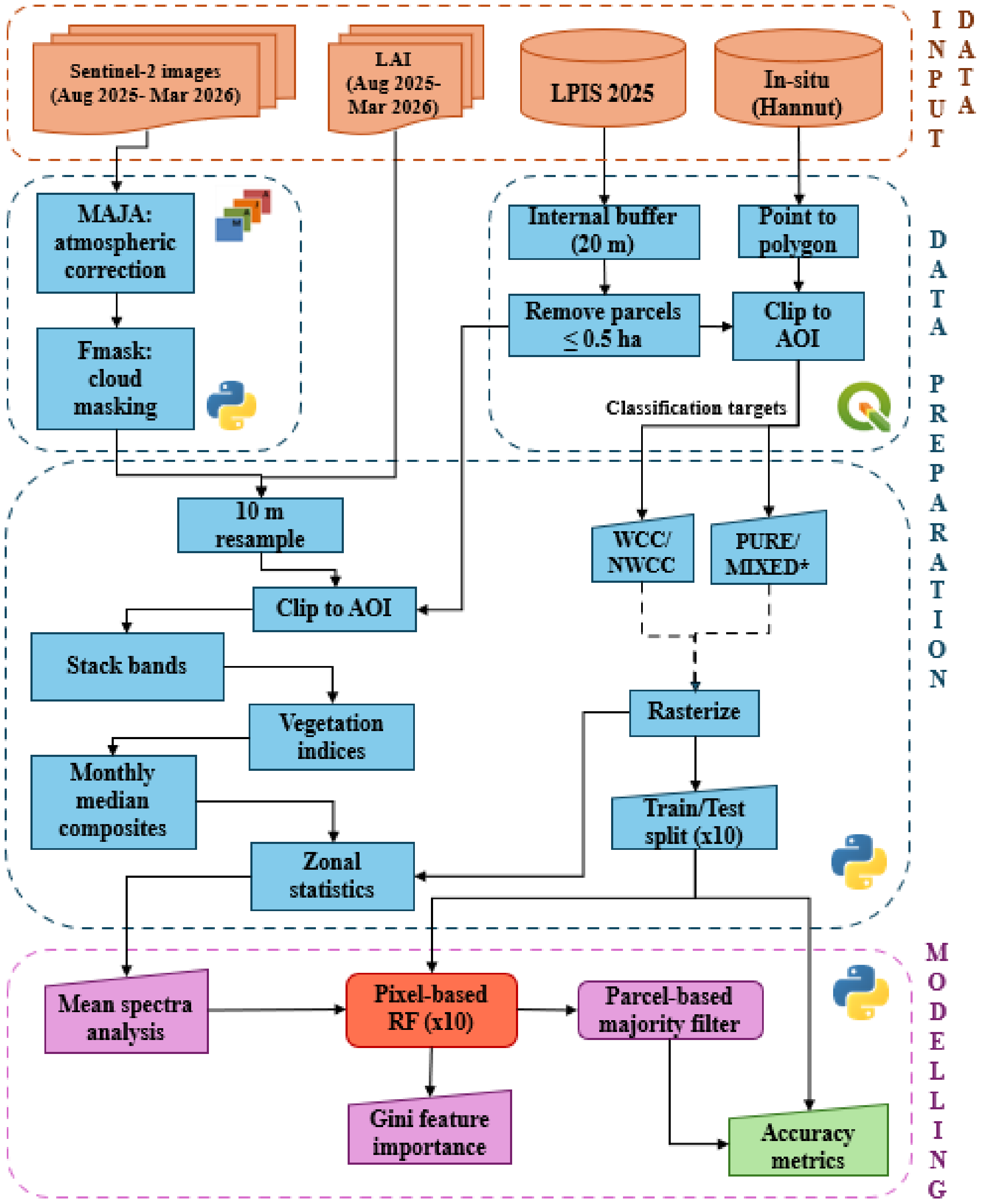
Appendix 5. Workflow for Setup 1: Monitoring WCC presence, mixture type and species composition.



Appendix 6. Workflow for Setup 2: Monitoring WCC presence, mixture type and species composition.

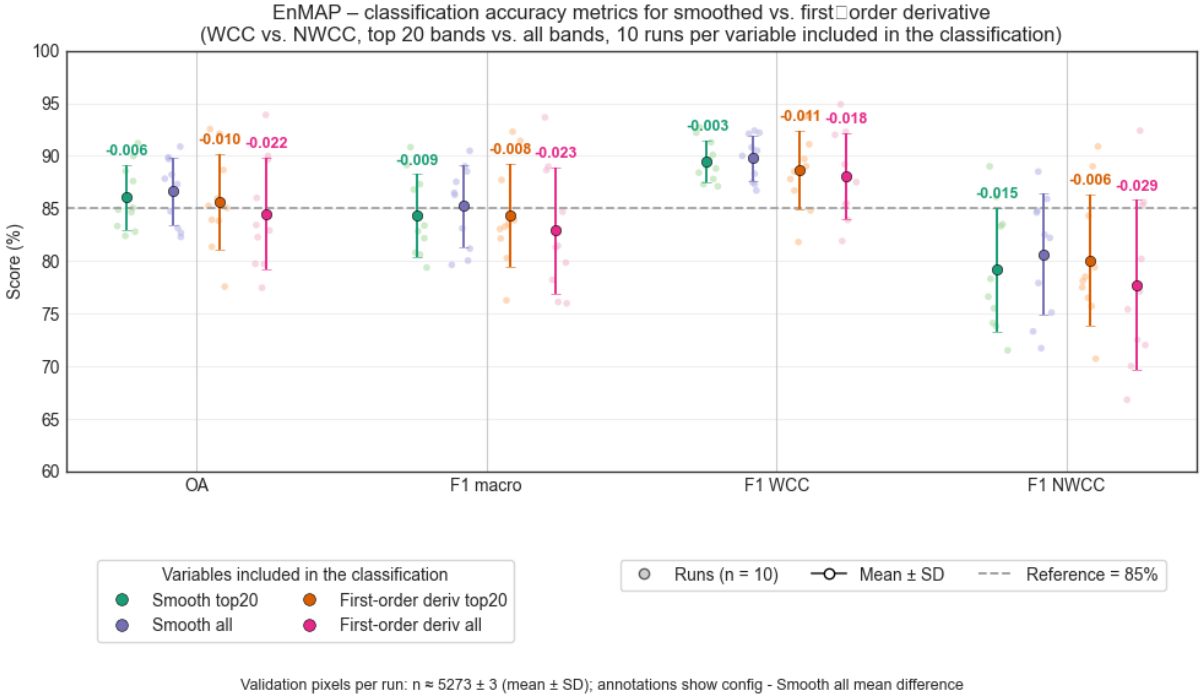


Appendix 7. Workflow for Setup 3: Monitoring WCC presence, mixture type and species composition.

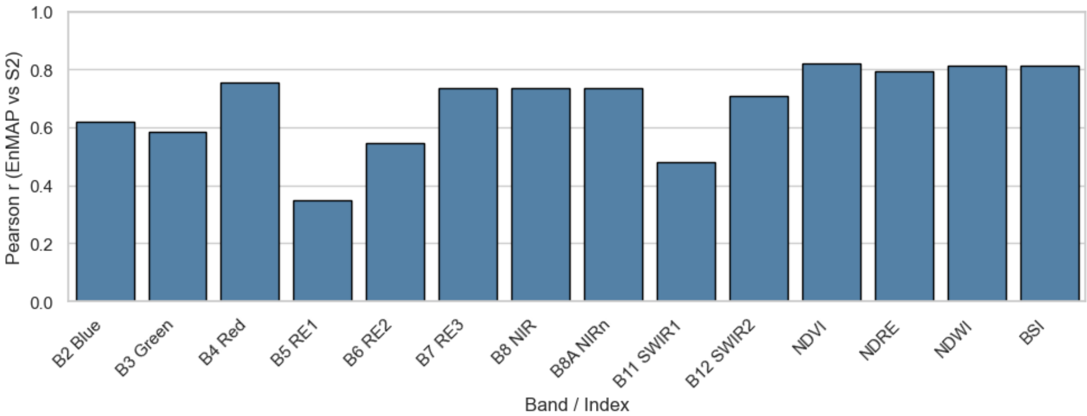


*only for November

Appendix 8. Classification accuracy of RF-1 with EnMAP input when using all vs. the 20 most informative smoothed reflectance bands (purple and green) versus all vs. the 20 most informative first-order derivative bands (red & orange).



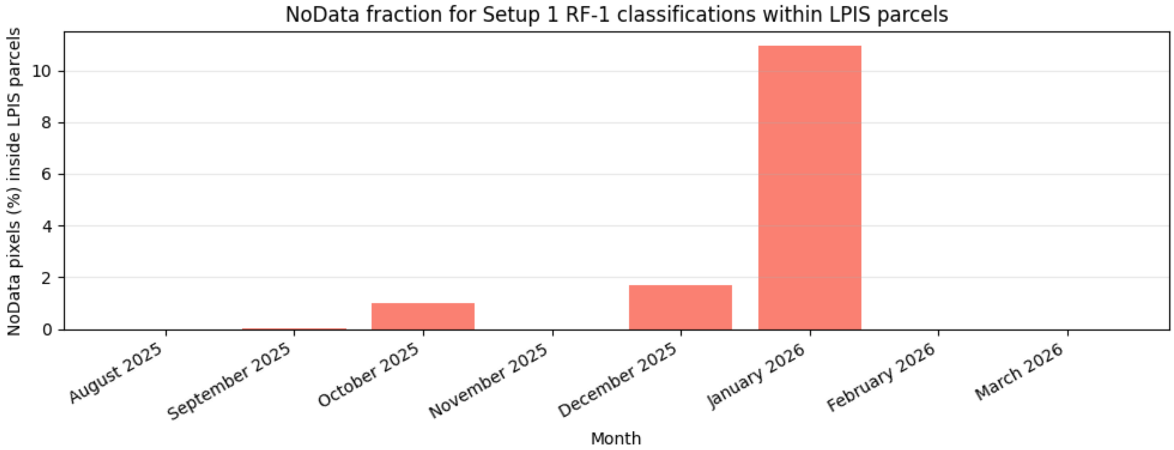
Appendix 9. Pixel wise Pearson correlation between spectrally resampled EnMAP reflectances and corresponding S2 bands and VIs for the harmonised single date comparison (Setup 2).



Appendix 10. Results of paired t tests comparing RF-1 accuracy metrics between EnMAP and S2 across 10 validation runs.

Accuracy Metric	t	P (*: α <= 0.5, **: α <= 0.01)
OA	3.416	0.0077**
F1_macro	3.832	0.004**
F1 class 1 (WCC)	3.085	0.013*
F1 class 2 (NWCC)	4.236	0.002**

Appendix 11. Fraction of 'nodata' pixels within LPIS parcels for the monthly RF 1 classifications in Setup 3, after applying the parcel-based majority filter, expressed as percentage of all pixels inside the LPIS mask for each month.



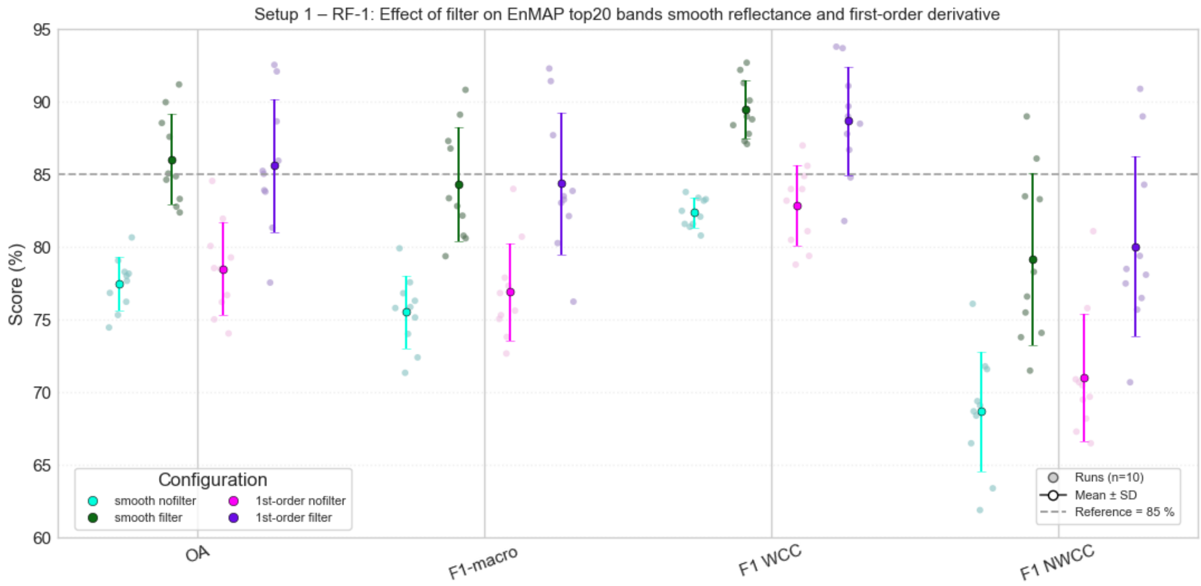
Appendix 12. Results of paired t tests comparing RF-2 accuracy metrics between EnMAP and S2 across 10 validation runs.

Accuracy Metric	t	P (*: α <= 0.5, **: α <= 0.01)
OA	1.042	0.325
F1_macro	0.921	0.381
F1 class 1 (PURE)	0.71	0.495
F1 class 2 (MIXED)	1.286	0.231

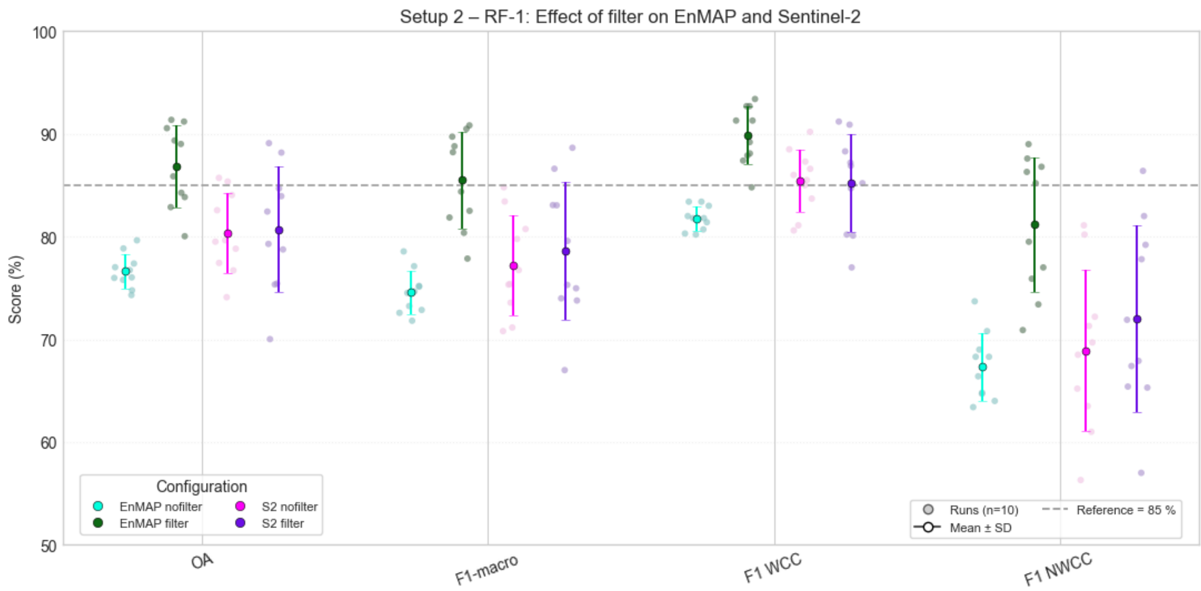
Appendix 13. Results of paired t tests comparing RF-3 accuracy metrics between EnMAP and S2 across 10 validation runs.

Accuracy Metric	t	P (*: α <= 0.5, **: α <= 0.01)
OA	0.763	0.465
F1_macro	3.387	0.008**
F1 class 1 (PhMix)	0.351	0.734
F1 class 2 (PurePh)	0.533	0.607
F1 class 3 (PureMu)	0.474	0.647
F1 class 4 (MuMix)	-	-
F1 class 5 (Ph+Mu)	1	0.343
F1 class 6 (Other mix)	4.322	0.002**

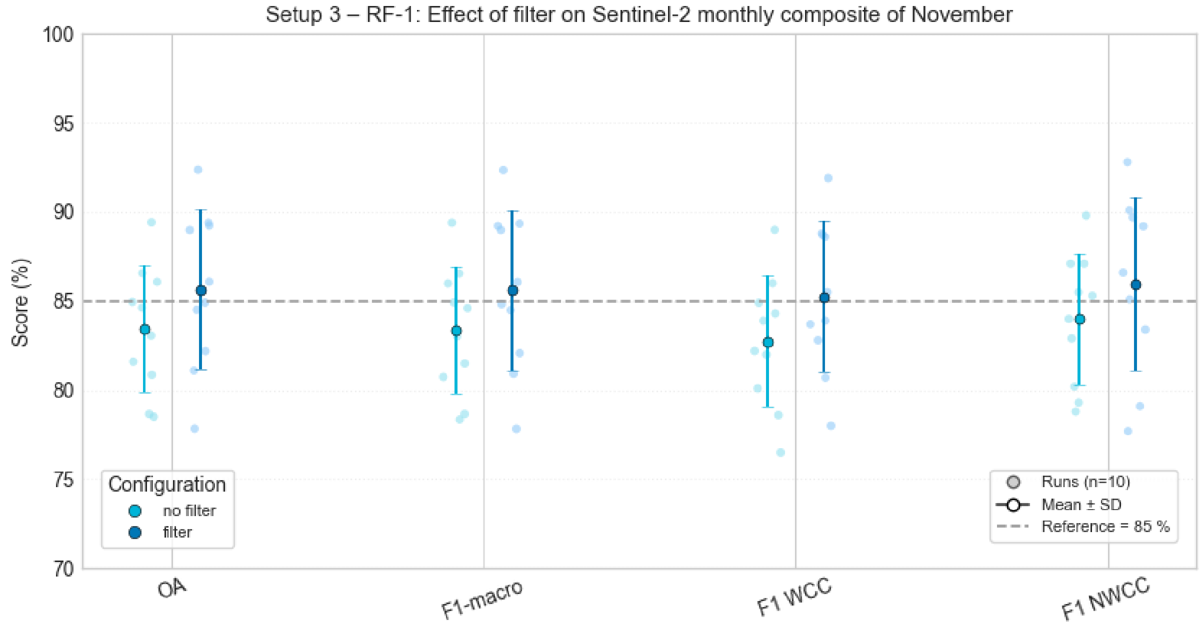
Appendix 14. Effect of parcel-based majority filtering on EnMAP RF-1 performance (Setup 1)



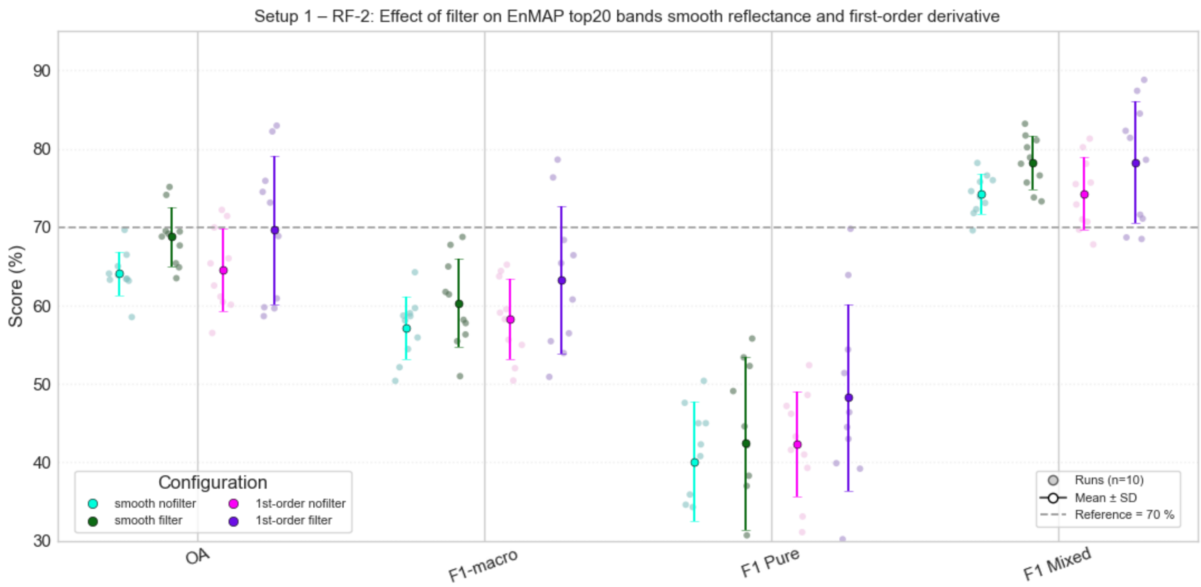
Appendix 15. Effect of parcel-based majority filtering on EnMAP and S2 RF-1 performance (Setup 2)



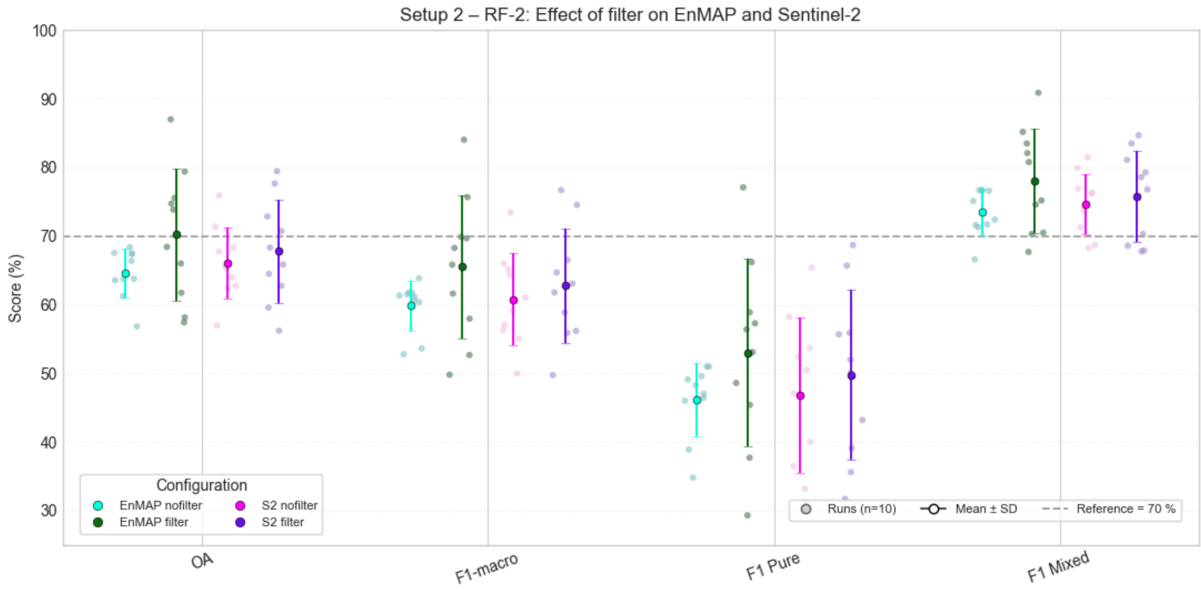
Appendix 16. Effect of parcel-based majority filtering on S2 monthly composite of November RF-1 performance (Setup 3)



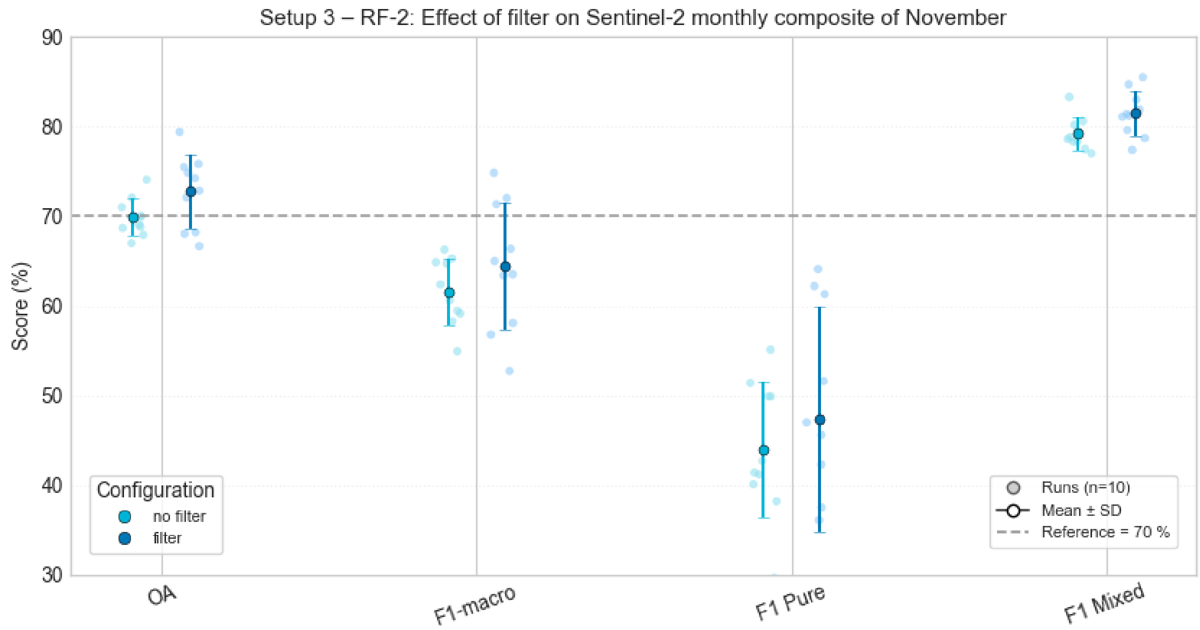
Appendix 17. Effect of parcel-based majority filtering on EnMAP RF-2 performance (Setup 1)



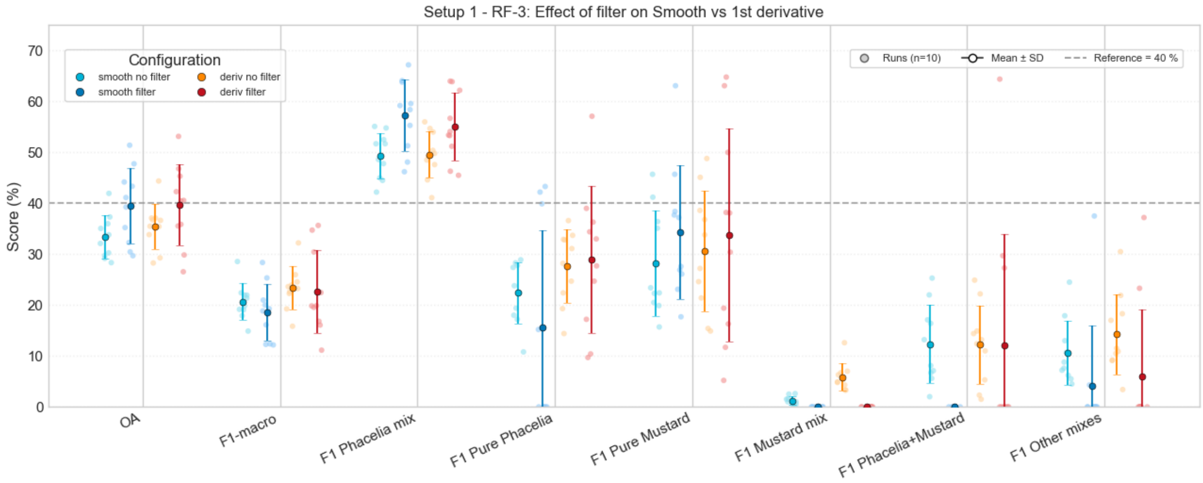
Appendix 18. Effect of parcel-based majority filtering on EnMAP and S2 RF-2 performance (Setup 2)



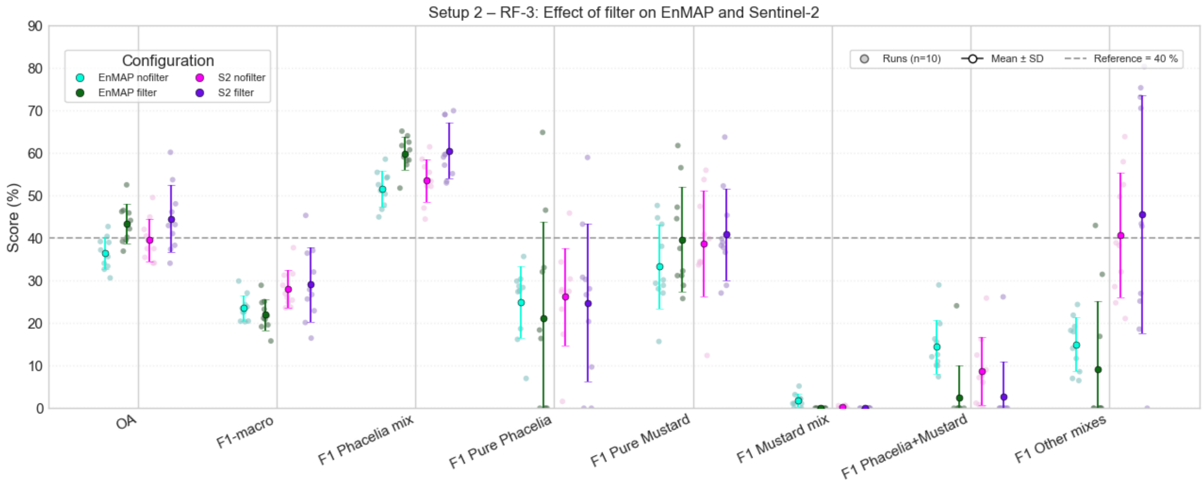
Appendix 19. Effect of parcel-based majority filtering on S2 monthly composite of November RF-2 performance (Setup 3)



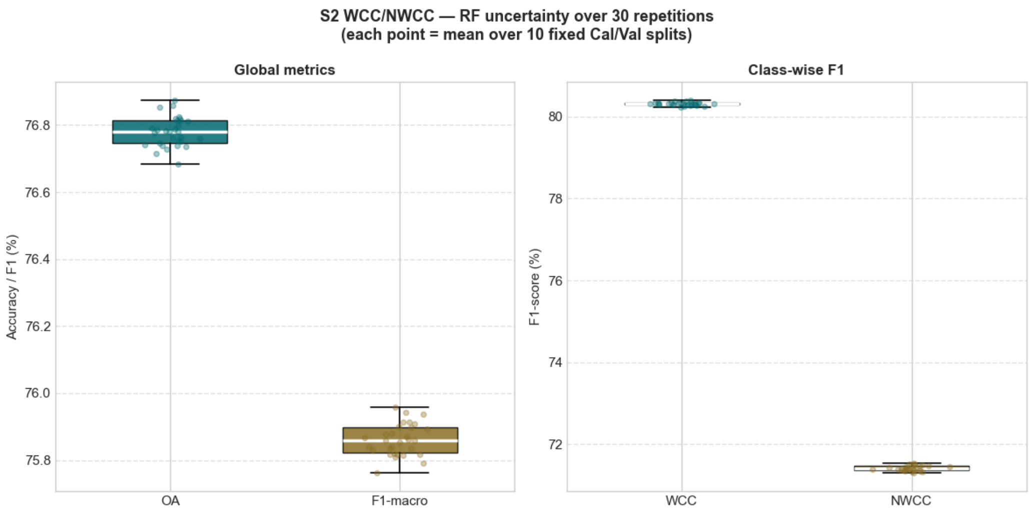
Appendix 20. Effect of parcel-based majority filtering on EnMAP RF-3 performance (Setup 1)



Appendix 21. Effect of parcel-based majority filtering on EnMAP and S2 RF-3 performance (Setup 2)



Appendix 22. Variability of S2 RF 1 WCC/NWCC classification over 30 independent parcel wise calibration/validation splits for S2 in Setup 2.



Multispectral and Hyperspectral Remote Sensing for Winter Cover Crop Identification, Mixture and Species Characterisation in Wallonia, Belgium

Hannah Firl

Winter cover crops (WCCs) deliver key agro-environmental benefits that depend strongly on the composition of the species and mixtures used. Despite growing policy interest under the Common Agricultural Policy (CAP), spatially explicit information on WCC composition at a regional scale remains scarce. This thesis evaluates the potential of Sentinel-2 (S2) multispectral (MSI) and EnMAP hyperspectral (HSI) imagery for monitoring WCC in the Hesbaye region of Wallonia, Belgium, across three hierarchical classification tasks: WCC presence versus absence; pure versus mixed WCC stands; and dominant species composition. Using a reference dataset encompassing the agronomic diversity of WCCs in the region — pure stands, binary mixtures, and complex multi-species combinations dominated by mustard, phacelia, and companion species — three experimental setups were implemented within a consistent Random Forest (RF) framework: single-date EnMAP classification; a harmonised single-date EnMAP–S2 comparison; and S2 monthly composites spanning the WCC growing season (August 2025–March 2026). WCCs presence was mapped with high accuracy across all setups (OA: 80.72-86.04%, $F1_{macro}$: 78.6-85.61%), confirming the viability of both sensors for operational soil cover monitoring. EnMAP outperformed S2 significantly for RF-1, suggesting that radiometric quality is as important as spectral sampling density. Distinguishing pure from mixed WCC stands remained moderately achievable (OA: 67.8-72.76%, $F1_{macro}$: 62.81-65.55%), with consistently higher F1 scores for mixed than for pure classes across all setups. Species composition mapping was the most challenging task (OA: 39.48-44.54%, $F1_{macro}$: 18.58-29.09%). Single-date EnMAP achieved moderate F1 scores only for phacelia-dominated and pure mustard classes. No systematic sensor advantage over S2 was observed at this level as classification errors were dominated by class imbalance and within-class spectral variability rather than spectral resolution. These results demonstrate that S2 time series can reliably support the monitoring of broad WCC, while EnMAP-like HSI adds measurable value for WCC detection and spectral band diagnostics, but not yet for reliable species-level mapping under current data and acquisition constraints.

Keywords: winter cover crops · multispectral satellite imagery (Sentinel-2) · hyperspectral satellite imagery (EnMAP) · Random Forest classification.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des bioingénieurs
Croix du Sud, 2 bte L7.05.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/agro**