

Faculté des bioingénieurs

Mapping perennial crops from Sentinel-2 image time series: a use case in Belgium

Auteur : Boris Norgaard

Promoteurs : Pierre Defourny
Sophie Bontemps

Lecteurs : Xavier Draye
Julien Radoux

Année Académique : 2021 - 2022

Mémoire de fin d'études présenté en vue de l'obtention du diplôme de

Bioingénieur : Sciences et technologies de l'environnement

Acknowledgments

I would first like to thank my supervisor, Pierre Defourny, as well as my co-supervisor Sophie Bontemps for their guidance which greatly contributed to the realization of this master's thesis. I would also like to thank Nicolas for the many meetings we had and the advice he gave me throughout the year. I also thank my friends Thibault and Romain for the many discussions in the openspace and the cafeteria. Finally I would like to thank my family for their support and the motivation they gave me.

I wish the readers, Mr. Draye and Mr. Radoux, as much pleasure in reading this master's thesis as I had in writing it.

Contents

Acknowledgments	i
List of Figures	iv
List of Tables	v
List of Tables	vii
1 Introduction	1
2 Literature review	2
2.1 Perennial crop definition	2
2.2 Remote sensing data for perennial crop mapping	3
2.2.1 Satellite images time series	3
2.2.2 Very high spatial resolution imagery and texture metrics	4
2.2.3 Spectral indices time series	5
2.2.4 Feature construction for classifier inputs	6
2.2.5 Gapfilling in optical time series	9
2.3 Crop classification in remote sensing	10
2.3.1 Supervised classification	11
2.3.2 Unsupervised crop classification in remote sensing	14
2.3.3 Performance metrics	15
2.4 Object-based image analysis and segmentation	16
2.5 Literature review conclusion	17
3 Objectives	18
3.1 General context	18
3.2 Objectives	18
4 Materials and method	20
4.1 Area of interest	21
4.2 <i>In-situ</i> data preparation	21
4.2.1 Class aggregation	23
4.2.2 Data set split	23
4.2.3 <i>In-situ</i> data screening	23
4.3 Sentinel-2 data pre-processing	25
4.3.1 Time series composition and cloud masking	25
4.3.2 Temporal resampling and gapfilling	26
4.4 Spectral indices	28

4.5	Crop-specific temporal analysis and feature design	28
4.5.1	Metrics relative to growing season	28
4.6	Calibration set sampling design	41
4.7	Feature selection scenarios, algorithm training and model comparison . . .	42
4.7.1	Hyper-parameter tuning with halving random search cross-validation	43
4.7.2	Prediction	44
4.8	Post-classification processing	45
4.8.1	Pixel-based analysis	45
4.8.2	Object-based analysis	46
4.8.3	Results evaluation	46
4.9	Overall approach	46
5	Results and discussion	48
5.1	Crop group binary classification results	48
5.1.1	Quantitative analysis of the pixel-based classifications	49
5.1.2	Object-based classification	53
5.1.3	Qualitative analysis of the object-based classifications	55
5.2	Crop type classification results	59
5.2.1	Quantitative analysis of the pixel-based crop type classifications . .	59
5.2.2	Quantitative analysis of the object-based crop type classification . .	64
5.2.3	Qualitative analysis of the object-based crop type classification . . .	66
5.3	Discussion and perspectives	68
6	Conclusion	70
	Bibliography	71
A	NDVI time series of each crop type in the <i>in-situ</i> data	75
B	Results of the object-based RF crop group classification	83

List of Figures

2.1	NDVI time-serie of 6 different crops: Corn, soybean, Winter wheat, Winter wheat and soybean in double cropping systems (WWsoybean), Warm Season Grass (WSG) and Cold Season Grass (CSG), Li et al. (2018)	7
2.2	Smoothing of an NDVI temporal profile with three methods (RMMEH, MVI and 4235EHtwice) (Jin and Xu, 2013)	10
3.1	Overview of the main steps of the overall approach	19
4.1	Location of the Area Of Interest (AOI) (Sentinel-2 T31UFR tile) in Belgium.	21
4.2	Distribution of the pixel count of each perennial class on the T31UFR Sentinel-2 tile	22
4.3	Overview of the preparation of the <i>in-situ</i> data coming from the SIGEC . .	25
4.4	Comparison of Sen2Cor and Fmask on the image of April 7, 2020.	27
4.5	Overview of the pre-processing steps applied to the Sentinel-2 images . . .	27
4.6	NDVI time series of 3 perennial (left column) and 3 annual crops (right column). Solid lines correspond to the mean interpolated NDVI time series of and dashed lines to the corresponding standard deviation (above and under the mean curve). The number of parcels is specified under the name of each crop in the top-left corner of each plot.	29
4.7	Comparison of the Savitsky-Golay smoothing of the NDVI time series of a vineyard and a winter wheat parcel at pixel scale	31
4.8	Amplitude of the smoothed NDVI time series over the North-West part of the Area Of Interest (AOI) (bottom image), and high-resolution image from Google maps (top image)	33
4.9	Values and illustration of the $NDVI_{GS}$ computed on mean NDVI time series at parcel level for 4 crops	35
4.10	Area under the growing season of the NDVI time series of each pixel over the North-West part of the Area Of Interest (AOI) (bottom image), and high-resolution image from Google maps (top image).	36
4.11	Values of the $NDVI_{ABS(der)}$ and profiles of the derivative obtained from the mean NDVI time series at single parcel level	38
4.12	Area under the absolute value of the derivative of the NDVI time series of each pixel over the North-West part of the Area Of Interest (AOI) (bottom image), and high-resolution image from Google maps (top image).	39
4.13	Maximum value of the BSI time series of each pixel over the South-West part of the Area Of Interest (AOI)	40
4.14	Initial raster stack	42
4.15	Sample matrix structure	42
4.16	Results of the majority filter on a pixel-based classification	46

4.17	Schematic overview of the method applied through this master's thesis . . .	47
5.1	Gini feature importance of the RF using the NDVI time series and the crop group-specific metrics as features	52
5.2	Gini feature importance of the EGB using the NDVI time series and the crop group-specific metrics as features	53
5.3	Misclassified parcels by the EGB classifier with NDVI time series and crop group-specific metrics as features	55
5.4	Misclassified parcels by the EGB classifier with NDVI time series as features	56
5.5	Probability map for the misclassified perennial crop parcels, with the probability being given by the number of trees having voted for the perennial crop class in the ensemble	57
5.6	NDVI time series of two misclassified perennial crop parcels: A vineyard (left) and a miscanthus parcel (right)	57
5.7	Repartition of the misclassified parcels of both models on a true color composite of the image of September 14,2020	58
5.8	Gini feature importance of the EGB using crop type-specific features . . .	62
5.9	Gini feature importance of the EGB using crop type-specific and crop group-specific metrics as features	63
5.10	Parcels containing partially correct prediction areas overlaying the high resolution map of Google maps. The top left is a miscanthus parcel, upper middle, top right and lower left parcels are fruit trees and lower middle is a vineyard.	66
5.12	Parcel potentially containing multiple crops, highlighted by NDVI corrected amplitude and a high resolution image from Google maps (a)	67

List of Tables

2.1	Sentinel-2 spectral bands physical characteristics	6
2.2	Temporal remote sensing crop cycle stages and their corresponding phenological stages defined by Matton et al. (2015)	7
2.3	Temporal NDVI features computed over a year-long NDVI time series by Valero et al. (2016) to distinguish annual cropland from other land cover classes	8
2.4	Temporal metrics created by Li et al. (2018) using NDVI values at different dates	9
4.1	Aggregated classes of annual and perennial crops according to JECAM guidelines	23
4.2	Ambiguous crop type classes removed because of unclear composition or unclear temporal system (annual or perennial)	24
4.3	Feature composition scenarios used for model calibration	43
4.4	Tested values for parameters in the RF optimization process	44
4.5	Tested values for parameters in the EGB optimization process	44
5.1	Overall accuracy results of the two models tested with different feature composition scenarios for the binary annual/perennial crop classification .	48
5.2	Class specific metrics obtained for each feature composition scenario for unfiltered pixel-based RF classifications	49
5.3	Class specific metrics obtained for the filtered RF classification with the NDVI time series and the crop group-specific metrics as features	50
5.4	Confusion matrix of the pixel-based filtered RF classification with NDVI time series and crop group-specific metrics as features	50
5.5	Class specific metrics obtained for each feature composition scenario for unfiltered pixel-based EGB classifications	51
5.6	Confusion matrix of the pixel-based filtered EGB classification with NDVI time series and crop group-specific metrics as features	51
5.7	Class specific metrics obtained for each feature composition scenario for the object-based EGB classifications	54
5.8	Confusion matrix of the object-based EGB classification with NDVI time series as features	54
5.9	Confusion matrix of the object-based EGB classification with NDVI time series and crop group-specific metrics as features	55
5.10	Overall accuracy metrics obtained for each feature composition scenario for the pixel-based and object-based approaches using an EGB classifier	59
5.11	Class specific metrics obtained for each feature composition scenario for the pixel-based EGB classification	60

5.12	Confusion matrix of the EGB crop type classification using crop type-specific features	61
5.13	Confusion matrix of the EGB crop type classification using crop type specific features and crop group-specific metrics as features	61
5.14	Class specific metrics obtained for each feature composition scenario for the object-based EGB classification	64
5.15	Confusion matrix of the object-based EGB crop type classification using crop type-specific features	65
5.16	Confusion matrix of the object-based EGB crop type classification using crop type-specific and crop group-specific metrics as features	65
B.1	Class specific metrics obtained for each feature composition scenario for the object-based RF classifications	83

Acronyms

AOI	Area Of Interest
BSI	Bare Soil Index
DT	Decision Trees
EGB	Extreme Gradient Boosting
ESA	European Space Agency
FN	False Negatives
FP	False Positives
LDA	Linear Discriminant Analysis
NDVI	Normalized Difference Vegetation Index
NDVWI	Normalized Difference Water Index
OA	Overall Accuracy
OBIA	Object Based Image Analysis
RF	Random Forests
SITS	Satellite Image Time Series
SVM	Support Vector Machines
TN	True Negatives
TP	True Positives
VHSR	Very High Spatial Resolution

Chapter 1 : Introduction

Agriculture is one of the cornerstones of modern society as it is responsible for the world's food production. However, pressures on agricultural systems due to population growth and climate change are forcing this sector to undergo a transition. In some parts of the world, measures are taken to make agriculture more sustainable, but in others, extensive and intensive policies are adopted and threaten soil quality as well as ecosystem balance. In this context, crop monitoring and mapping systems become crucially needed as decision support tools for smallholders and farmers but also as global land-use evolution assessment products.

Regarding the implementation of such systems, remote sensing data has emerged as a particularly relevant tool. Indeed, the access to a wide variety of sensors, each having particular characteristics in terms of image resolution and frequency of acquisition dates, makes these solutions globally implementable. In parallel, advances in machine learning, artificial intelligence, computing power and storage size of computers have made the use of algorithms on remote sensing data faster and more efficient.

The conjunction of these two fields allowed the creation of a large number of classification frameworks, among which are land-cover mapping and crop type classifications. However, it has been noted that a particular focus was given to annual crops and much less to perennial crops, even though these represent a significant proportion of the agricultural production of some countries. Therefore, this master's thesis had the objective to create a classification model able to distinguish perennial from annual crops based on remote sensing imagery.

The next section of this master's thesis will be dedicated to the state of the art of scientific knowledge related to remote sensing-based crop classification, with a particular focus on perennial crops. The objectives will then be stated following the conclusion of the state of the art. Afterwards, the methodology applied and the data used to classify perennial crops will be detailed. Finally, the results and their interpretation will be shared in the last two sections.

Chapter 2 : Literature review

This chapter will focus on the state of the art in terms of scientific knowledge related to remote sensing-based crop mapping and more precisely perennial crop classification. This will allow to precisely define the thesis's scientific contribution, to develop an experimental protocol inspired by the existing works and to justify the choices regarding methods and hypotheses. Finally, it will also stand as a comparison base for the results obtained in the framework of the thesis with the results reported in the literature.

Remote sensing-based crop mapping and perennial crop characterization in literature will first be detailed. Then, different types of satellite images used in the context of crop classification will be listed. At last, algorithms, as well as classification methods (object-based and pixel-based) used in scientific articles will be compared.

2.1 Perennial crop definition

Perennial crop characterization is a subject around which no consensus has been clearly established. Multiple definitions have been stated by a series of institutions and they vary depending on the context. Perennial crops generally are characterized by an absence of rotation and a multi-annual cover of the same crop. This can be put in opposition to annual croplands, which are defined by the Joint Experiment for Crop Assessment and Monitoring (JECAM) as "a piece of arable land that is sowed or planted at least once within a 12 months period". However, this definition is contextual and specific to a remote sensing perspective. The Food and Agriculture Organization (FAO) defines perennial crops as "plants that last for more than two growing seasons, either dying back after each season or growing continuously", separating these crops into nine international standard industrial classification classes. This definition stands in the context of agricultural goods production and therefore excludes permanent grasslands, unlike the previous definition. Hatfield and Walthall (2014) gave an even narrower definition of perennial crops in the Encyclopedia of Agriculture and Food Systems (2014), stating that "perennial crops are typically considered as those that are more permanent, requiring a number of growth cycles before fruit is produced", which excludes several crops included in the ones cited above. From a more geographical point of view, Eurostat defines perennials as ligneous

crops grown outside the crop rotation system, occupying the land and providing crops for several consecutive years (usually more than five)(Eurostat, 2020). This definition, therefore, neglects herbaceous plants and permanent meadows.

The lack of uniformity regarding the definition of perennial crops reflects the variety of economic and agro-ecological contexts in which they are embedded. In addition, perennial crops have different impacts depending on where they are grown, which partially explains the diversity of their definition. For instance, perennials grown in temperate and continental regions mainly include fruit tree orchards, species of nut trees, herbaceous perennials, berry shrubs, and some cultures related to the production of alcohols (vineyards and hop plantations, for example). Mediterranean climate, on the other hand, is particularly well suited for the development of citrus fruit trees, vineyards, and olive trees which are the majority perennial crops in Europe and have significant economic implications for cultivating countries, which is also observed in arid and harsh environments where perennials like date palms are grown. Finally, equatorial and tropical perennials include oilseeds and palm trees, fruit trees growing in hot and humid environments such as coffee or cocoa, or cultures of non-alimentary goods such as rubber. These are currently subject to discussions because of their impact due to intensive mono-cultures resulting in consequent production but also in a biodiversity loss and raising ecological concerns.

2.2 Remote sensing data for perennial crop mapping

So far, research on remote sensing-based classification of crops has mainly been focused on annual cropping systems and permanent grasslands (Peña and Brenning, 2015 ; Brinkhoff et al., 2019). In addition, perennial crop species sometimes show similar temporal phenological profiles, further complicating their discrimination. Different methods and data types have been used across literature to overcome this challenge, differentiate between annual and perennial crops, and classify species across perennial crops.

2.2.1 Satellite images time series

Satellite Image Time Series (SITS) capture the same region of interest at different times. Regarding crop classification, SITS allow to distinguish between agricultural practices and phenological characteristics specific to species and to consider the evolution of spectral indices. Multiple SITS types such as Sentinel-2 multi-spectral images and Sentinel-1 Synthetic Aperture Radar (SAR) can be used jointly. For instance, the combination of the features of radar and optical images seems to show better accuracy in classifications in cloudy areas than when used separately in single-sensor procedures (Van Tricht et al., 2018).

Brinkhoff et al., (2019) used radar and multi-spectral optical images from May 2018

to April 2019, aiming to classify nine perennial crops in Australia. As their Area Of Interest (AOI) spanned over 4 Sentinel-2 they computed high-quality monthly mosaics for each of the 11 Sentinel-2 and 2 Sentinel-1 bands as well as three monthly composites, resulting in an image with $(12 + 3) \times (11 + 2) = 195$ bands. These were then used through pixel-classification and object-based image analysis. These methods will be detailed later.

Peña and Brenning (2015) used Landsat SITS to classify four fruit tree species in Chile. Their study also aimed to compare the classification accuracy of four different sets of features: the SITS of 6 bands (the 6 bands of 8 acquisition dates), the Normalized Difference Vegetation Index (NDVI) temporal profile (8 images), the Normalized Difference Water Index (NDVWI) temporal profile (8 images) and a combination of the three aforementioned. The best accuracy was obtained using the 6 bands SITS and a Linear Discriminant Analysis (LDA) classifier, with a misclassification error rate ranging from 0.05 to 0.08 depending on the size of the training set. The misclassification error rate is defined as the proportion of misclassified samples in a data set.

Stoian et al. (2019) used 11 Sentinel-2 tiles spread over the French territory to classify pixels within 17 classes, among which were orchards and vineyards. They used neural networks, and, more specifically, U-net architecture and compared it to a Random Forests (RF) approach, also adding NDVI and NDVWI as features in addition to the 33 dates $\times$ 10 bands of the images. They showed that the U-net neural network performed better than the RF for all tiles except for one, which, according to them, was due to the difference in the distribution of classes in this tile.

2.2.2 Very high spatial resolution imagery and texture metrics

Although SITS are currently widely used for perennial crop classification research, other techniques and data types have been used through literature. Very high spatial resolution imagery (Very High Spatial Resolution (VHSR)) are images whose pixel resolution is inferior to one meter. This type of image enables the highlight of specific patterns, which are detected as recurring gray tones with a method called texture metrics analysis. Tones are based on the different values of gray that can be found in a given image, whereas texture is a spatial interpretation of tone distribution (Haralick et al., 1973).

Delenne et al. (2008) used texture metrics and grid and line patterns to precisely map vineyards in France using orthophotography with 50cm of resolution. They used two texture metrics using a sliding window. The first one is co-occurrence analysis and aims to detect changes in values from pixel to pixel following particular angles. This method allowed them to detect vineyards arranged in rows, the contrast between pixels being high following a certain angle and low following the perpendicular of this angle. Co-occurrence analysis is based on co-occurrence matrices. This matrix type is defined as an $N \times N$ array with N the number of grey tones composing the image. It is computed using the relative

distance between pixels d and the relative orientation of pixel pairs θ as parameters. Each matrix element with coordinates i, j (line and column index) is then equal to the number of transitions in pixel values from i to j between pixels with distance d and orientation θ . It can be summarized as a two-dimensional histogram of value transition within an image with respect to a specific orientation and distance. Haralick et al. (1973) derived 14 textural metrics relative to images' co-occurrence matrices and gray tones. These metrics include contrast reflecting local variations, correlation which is a measure of gray-tones linear dependency, and angular second momentum, which is a measure of homogeneity in an image.

The second method used by Delenne et al. (2008) is based on frequency analysis and Fourier transform. The pixels are converted into a sum of sinusoidal waves having different frequencies using the fast-Fourier transform algorithm, and the amplitude is represented on an image of the same size as the initial one. The patterns observed on this amplitude image can then be interpreted as vineyard or non-vineyard. They obtained an accuracy of 70.8% with the co-occurrence analysis and 86% with frequency analysis.

2.2.3 Spectral indices time series

Optical images, for instance provided by the sentinel-2 satellite, are composed of 13 spectral bands detailed in Table 2.1. These bands can be used to display true or false-color images and to calculate biophysical variables and spectral indices aiming to detect specific patterns. In addition, the surface reflectances values for all bands and indices derived from them can be used as classifier features which uses them as explanatory variables for certain classes. Commonly used indices for crop classification and their use in time series are further detailed below.

The Normalized Difference Vegetation Index (NDVI) is an index primarily used to monitor green vegetation using the reflectance in the NIR (near infra-red)(ρ_{NIR}) and red bands (ρ_{Red}). This index is a representative indicator of vegetation physiology since chlorophyll absorbs radiation in the red band and the foliar mesophyll reflects in the infra-red band (Rouse and Haas, 1974). It is calculated as follows and spans from -1 to 1 :

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (2.1)$$

The Normalized Difference Water Index (NDWI) is used to highlight water bodies and water content in observed areas (Gao, 1996). It is calculated using the NIR and SWIR (short wave infra-red) bands and spans from -1 to 1.

$$NDWI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}} \quad (2.2)$$

	Sentinel-2A		Sentinel-2B		
Sentinel-2 bands	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
Band 1 – Coastal aerosol	442.7	21	442.2	21	60
Band 2 – Blue	492.4	66	492.1	66	10
Band 3 – Green	559.8	36	559.0	36	10
Band 4 – Red	664.6	31	664.9	31	10
Band 5 – Vegetation red edge	704.1	15	703.8	16	20
Band 6 – Vegetation red edge	740.5	15	739.1	15	20
Band 7 – Vegetation red edge	782.8	20	779.7	20	20
Band 8 – NIR	832.8	106	832.9	106	10
Band 8A – Narrow NIR	864.7	21	864.0	22	20
Band 9 – Water vapour	945.1	20	943.2	21	60
Band 10 – SWIR – Cirrus	1373.5	31	1376.9	30	60
Band 11 – SWIR	1613.7	91	1610.4	94	20
Band 12 – SWIR	2202.4	175	2185.7	185	20

Table 2.1: Sentinel-2 spectral bands physical characteristics

Bare Soil Index (BSI) is an index whose components vary depending on the climatic context in which its calculation is embedded but generally combines blue, red, near-infrared and short wave infrared spectral bands to capture soil variations (Diek et al., 2017).

$$BSI = \frac{(\rho_{Red} + \rho_{SWIR}) - (\rho_{NIR} + \rho_{Blue})}{(\rho_{Red} + \rho_{SWIR}) + (\rho_{NIR} + \rho_{Blue})} \quad (2.3)$$

A particularly relevant biophysical variable is the leaf area index (LAI) which is given by the proportion of leaf area per ground unit area (Boegh, et al., 2002). LAI variation overtimes are highly correlated to phenological stages of crop growth, and LAI time series are therefore instrumental to differentiate crop types. For example, perennial crops' LAI cycles will theoretically be broader and of lesser amplitude than annual crops.

2.2.4 Feature construction for classifier inputs

Using spectral indices and reflectance values as time series allows the creation of crop-specific temporal metrics. This section illustrates different approaches to design features specific to cropland.

Figure 2.1 highlights the difference in temporal profile of NDVI accross different crops. The values displayed in Figure 2.1 were computed using the mean NDVI of polygons of each crop type and for each date as part of an object-based analysis.

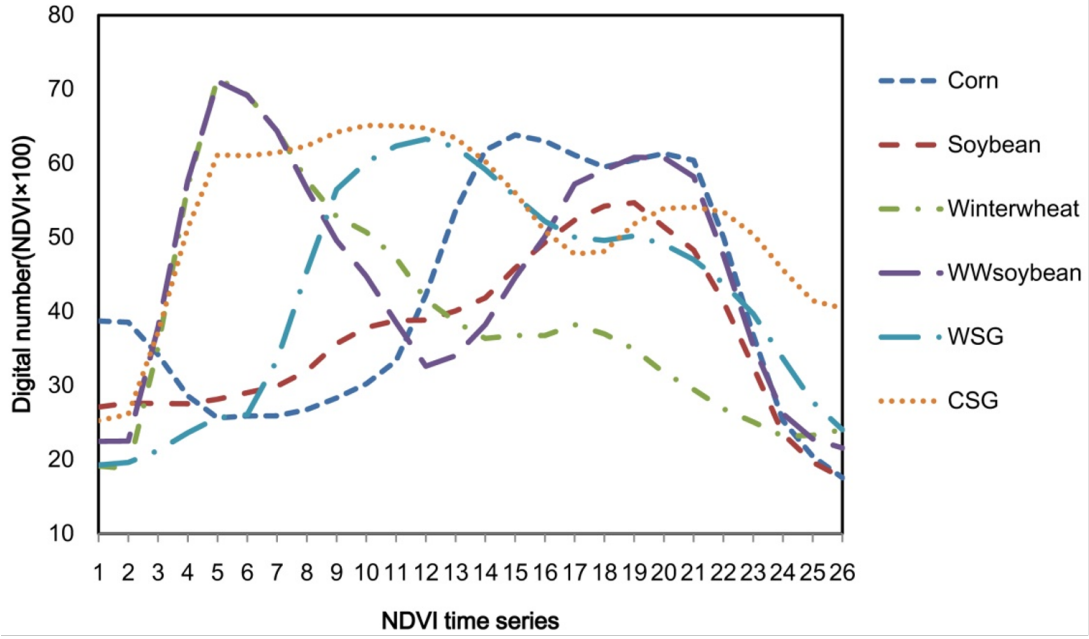


Figure 2.1: NDVI time-series of 6 different crops: Corn, soybean, Winter wheat, Winter wheat and soybean in double cropping systems (WWsoybean), Warm Season Grass (WSG) and Cold Season Grass (CSG), Li et al. (2018)

2.2.4.1 Annual crop features

Matton et al.(2015) designed crop-specific temporal features defined using crop growth characteristics. These are based on key elements of annul crop growth cycle and allow for species distinction. The key remote sensing stages in the crop cycle and their corresponding phenological stages used in this paper are summarized in table 2.2:

Table 2.2: Temporal remote sensing crop cycle stages and their corresponding phenological stages defined by Matton et al. (2015)

Timing in the red reflectance and NDVI profile	Crop cycle stage
maximum reflectance of the red band	minimal chlorophyll content - bare soil
maximum positive slope of the NDVI time serie	growth phase
maximum value of NDVI	peak of green vegetation
maximum negative slope of the NDVI time series	harvest and/or senescence
minimum value of NDVI	end of harvest - bare soil

The values of 4 spectral bands at these respective stages were computed for each pixel, resulting in 20 features. A selection of the 5 features giving the best mean Overall Accuracy (OA) over the 8 areas of interest was retained.

2.2.4.2 Spectral indices features

Valero et al. (2016) aimed to produce a binary “annual-cropland/no-annual-cropland” map using SPOT4 and Landsat-8 images benchmarked over 12 test sites. Their goal was to produce a dynamic processing chain, meaning the model would increase in accuracy as new data would be progressively integrated into the model. They used a RF classifier for pixel-based binary classification. This was followed by two object-based methods, the first one based on feature extraction using a principal component analysis, segmentation, and homogenization of pixel values in a close neighborhood at object scale, and the second one on the filtering of salt and pepper effect of the pixel-based classification.

The features used for the pixel-based classification were based on NDVI, NDWI, and Brightness time series. However, as the objective was to create a dynamic model in time, NDVI values of pixels at a particular date were not deemed relevant.

A set of 17 features were extracted from the NDVI time series and aimed to distinguish annual cropland from the rest of the images and are detailed in table 2.3. In addition, the mean, median, standard deviation, maximum, and minimum of the Brightness and NDWI were computed.

Table 2.3: Temporal NDVI features computed over a year-long NDVI time series by Valero et al. (2016) to distinguish annual cropland from other land cover classes

Notation	Description
$NDVI_{max}$	Maximum value
$NDVI_{mean}$	Mean value
$NDVI_{std}$	Standard deviation value
$NDVI_{Dif-max}$	Maximum difference in a sliding temporal neighborhood having a size w (default value $w=2$)
$NDVI_{Dif-min}$	Minimum difference in a sliding temporal neighborhood having a size w (default value $w=2$)
$NDVI_{Dif-Dif}$	Difference between $NDVI_{Dif-max}$ and $NDVI_{Dif-min}$ value estimating the transition jump
$NDVI_{max}$	Maximum mean value in a sliding temporal neighborhood having a size w (default value $w=2$)
$NDVI_{max-Lg}$	Length of the flat interval containing the peak area associated to $NDVI_{max}$ value.
$NDVI_{max-Sur}$	Surface of the flat interval area containing the peak value associated to $NDVI_{max}$
$NDVI_{Pos-Sur}$	Maximum surface of the first positive derivative interval
$NDVI_{Pos-Lg}$	Length of the first positive derivative interval associated to $NDVI_{Pos-Sur}$ value
$NDVI_{Pos-Rt}$	Rate of the first positive derivative interval associated to $NDVI_{Pos-Sur}$
$NDVI_{Neg-Sur}$	Maximum surface of the first negative derivative interval
$NDVI_{Neg-Lg}$	Length of the first negative derivative interval associated to $NDVI_{Neg-Sur}$ value.
$NDVI_{Neg-Rt}$	Rate of senescence of the first negative derivative interval associated to $NDVI_{Neg-Sur}$
$NDVI_{Pos-Tr}$	Flag detecting if there is a bare soil transition before the positive derivative
$NDVI_{Neg-Tr}$	Flag detecting if there is a bare soil transition after the negative derivative

2.2.4.3 Other features

Temporal metrics computed exclusively using single indices time series have also been used in previous work. Li et al. (2018) created time series-specific metrics by applying operations on NDVI values at different dates to detect change patterns of specific crops. The initial data consisted of all the Landsat-MODIS pairs of images (with the same acquisition date) available over the AOI located in Southwest Missouri, United States. After pre-processing, the time series contained 26 dates from February to June 2007. The NDVI values during those dates were used to build crop-specific metrics, which are detailed in Table 2.4.

Table 2.4: Temporal metrics created by Li et al. (2018) using NDVI values at different dates

Metric	Description
$(NDVI_5 - NDVI_2) + NDVI_6$	reflects the change patterns of crops with early NDVI peak
$(NDVI_{11} - NDVI_6) + (NDVI_{12} - NDVI_8)$	reflects the change patterns of warm-season grass
$(NDVI_{15} - NDVI_{10}) + (NDVI_{15} - NDVI_{17})$	reflects the change patterns of corn
$(NDVI_{12} - NDVI_5) + NDVI_{24}$	reflects the change patterns of cool-season grass
$(NDVI_{20} - NDVI_{12}) + (NDVI_{19} + NDVI_{20})/2$	reflects the change patterns of winter wheat/soybeen double cropping systems

2.2.5 Gapfilling in optical time series

The usability of optical remote sensing images is subject to atmospheric conditions and, more particularly, to cloud screening. Some places on the planet have a frequent cloud cover resulting in data gaps when building time series. Gaps can also be caused by sensor failures or processing errors. These potential lacks or noise in data must be handled cautiously so as not to be led to incorrect interpretations. For instance, linear interpolation between values should only be applied when the time interval between missing data is short enough or that there is no risk of consequent modification of the time series.

Multiple smoothing techniques of varying levels of complexity can be used depending on the use case. Li et al. (2018) used a second-degree Savitsky-Golay filter to smooth an NDVI time series and prevent atmospheric and cloud effects. This algorithm operates by splitting the signal into intervals of equal length on which a polynomial regression minimizing the least squared error on the interval is applied. The degree of the polynomial regression can vary, a 0 degree being a sliding mean smoothing.

Other algorithms such as a Whittaker-smoother or Fourier decompositions serve the same gap-filling and smoothing purposes but are more contextually restrained, particularly for the latter, which requires a sinusoidal signal to begin with.

Finally, an algorithm called the running median maximum operation endpoint processing hanning (RMMEH) has been specifically designed for NDVI time series processing. It is based on the assumptions that temporal NDVI profiles change smoothly over time following seasonal growth cycles and that external effects (atmospheric, clouds, and bidirectional effects) tend to diminish NDVI values. This algorithm works by simultaneously running sliding means and medians through the temporal profile. The maximum of the set composed by the original NDVI value, the sliding mean and median is then selected for each date. The last steps consist of an endpoint processing to ensure the first and last points of the time series are correctly assessed and a weighted moving average of the computed temporal profile. Figure 2.2 displays the result of the algorithm on an NDVI temporal profile compared to the original signal and two other smoothing methods, the mean value iteration (MVI) and a non-linear smoother named 4235EHtwice (Jin and Xu, 2013).

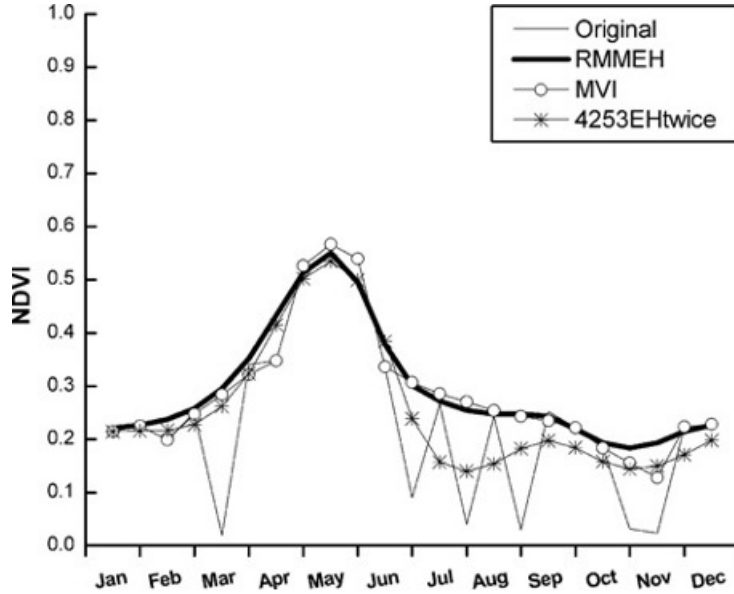


Figure 2.2: Smoothing of an NDVI temporal profile with three methods (RMMEH, MVI and 4235EHtwice) (Jin and Xu, 2013)

2.3 Crop classification in remote sensing

Classification algorithms can be separated into two categories depending on the use of labeled training data or if the classification occurs without training data set. When training samples representative of specific classes are created and given as input to a classification algorithm, it is referred to as supervised. In addition, machine learning is a field in constant evolution, and the classification algorithms currently considered as state of the art might not be the same in a few years. Furthermore, these algorithms do not perform equally when given different data types. Widely used classifiers across literature related to remote sensing will be further detailed in the following sections.

2.3.1 Supervised classification

2.3.1.1 Decision trees and random forest

Decision Trees (DT) and RF are algorithms which can be used as regression or classification tools. DT are composed of nodes representing variables, branches representing value choices, and leafs representing the probability of belonging to a certain class.

RF is an ensemble method, meaning multiple models are combined to improve accuracy. RF are constructed with a multitude of DT whose decisions are aggregated to assign a class according to the majority result of the trees (Breiman, 2001). RF have been widely used in land cover classification and often presents the best results compared to other algorithms when using large data sets with lots of training features (Lebourgeois et al., 2017; Van Tricht et al., 2018).

RF rely on a method called bagging, which determines how the weak learners (i.e., trees) are sampled. Reduced data sets are created in which all observations have equal chances of appearing. Trees are then built by randomly sampling observations. The benefit of random sampling methods is that variance is reduced through the use of multiple learners.

Almost all the studies cited in this literature review have used this algorithm or a variation of it with varying levels of success. The use of RF is generally quite similar, although the statistical parameters (numbers and depth of trees, for example) are fundamental and vary depending on the use case.

2.3.1.2 Gradient boosting

Gradient boosting is another ensemble algorithm gaining popularity in data sciences and machine-learning related fields. This ensemble method differs from RF in the sampling process, during which boosting is adopted instead of bagging. Boosting is a sequential process during which weak learners (trees that misclassify samples) are taken into account to weight important observations contrarily to bagging, which processes in parallel and does not weight samples in the training phase. The classification result is then given by the weighted average of the decisions of each tree (Hastie et al., 2009).

Gradient boosting has the advantage of performing well on imbalanced data sets as it uses a gradient descent of a loss function and weights weak learners, but it is more prone to overfitting than RF, as pointed out by Dimov et al. (2019), who used a gradient boosting algorithm to classify crops using a combination of optical and SAR data.

Extreme Gradient Boosting (EGB) is a specific decision tree-based variation of gradient boosting which is also based on the minimization of a loss function. The minimization is applied using second order gradients as a proxy for the gradient descent, which tends to

converge much faster than when minimizing the regular loss function of the classifier. EGB also uses shrinkage during its regularization process which tends to diminish overfitting occurrence. One of the main advantages of the algorithm is the possibility to parallelize the computations on several CPU's, drastically decreasing computation time.

2.3.1.3 Support vector machines

Support Vector Machines (SVM) are machine learning models that can be used for classification or regression. In the case of classification, data is transformed through a linear application into a higher dimension space where it can be discriminated by a hyperplane (Cortes and Vapnik, 1975). The original model is designed for linear binary classifications, but modern variations can split multi-class problems into multiple two-classes classifications and use kernel functions which make SVM's non-linear classifiers. SVM have the advantage of being tolerant to outliers and handling overlapping class problems relatively well.

Kernel functions systematically find support vector classifiers (i.e., hyperplanes discriminating samples in high dimension spaces). Widely used kernel functions include the polynomial kernel, which has a parameter d standing for the degree of the polynomial transformation, the radial basis function kernel, which finds a support vector classifier in infinite dimensions, and the sigmoid kernel.

SVM and RF often are compared and a trade-off is most of the time to be found between higher accuracy of SVM and reduced computation time of RF (Sitokonstantinou et al. 2018; Brinkhoff et al. 2019).

2.3.1.4 Deep learning

Deep learning is a branch of machine learning based on neural networks inspired by the way human neurons process information. These models have seen a significant improvement over the last decade and have sometimes outperformed state-of-the-art classifiers and regression algorithms in some research areas, including remote sensing (Yapici et al., 2019). Neural networks are characterized by their architecture (i.e., the structure of the layers within the network), which always begins with an input layer and ends with an output layer. Many types of architectures meeting various objectives have been developed, among which are convolutional neural networks, which are particularly efficient at pattern detection in two-dimensional spaces (Chen et al., 2014). Another frequently used architecture is the recurrent neural network, which is used in remote sensing research fields to process sequential data and time series.

In the context of crop classification, Zhong et al. (2019) compared two deep learning models to state-of-the-art supervised classifiers (RF, EGB and SVM). Their work aimed to develop a deep-learning-based classification framework based on SITS. They used

Landsat 7 and 8 images and trained their models to classify crop types, among which some were perennials (vineyards and orchards). In order to do so, they compared a Long Short-term memory neural network to a one-dimensional convolutional network. As a result, they obtained the best accuracy using the one-dimensional architecture, which slightly outperformed EGB by 1% and RF by 2% with an OA of 85.53%.

2.3.1.5 Other classifiers

Other classifiers worth mentioning will be briefly detailed in this section but will not be extensively covered given their more limited presence in the literature.

The first one is LDA which consists of searching for the best linear combination of features to discriminate two or more classes.

Another algorithm is the maximum likelihood classifier which supposes the samples are normally distributed and includes prior pieces of information with regard to the class repartition in the form of prior probabilities to determine the class of a given sample.

2.3.1.6 Importance of the distribution of the training data and resampling methods

Although the classifier choice and parameters are crucial for a given use case, it is important to note that the statistical distribution of each class among the data and their respective repartition in training and validation sets is equally important.

Unequal class repartition in training data is globally referred to as the imbalanced learning problem and is a field of study in itself (Chawla et al., 2013). This situation occurs when instances of classes are less represented than others in data sets which is intrinsic to land cover classification since classes often are naturally unbalanced, but can also happen due to the data collection process. The imbalanced learning problem can be tackled at the data level, and classes can be resampled to correct the class distribution. This approach is the most widely used in land cover classification. It has multiple benefits due to the preventive nature of the strategy. It can indeed be used independently from the classifier choice and thus must only be applied once.

The resampling method can be uninformed, randomly re-balancing the data set, or informed, purposely removing or adding training samples based on local or global information. The main approaches regarding resampling are oversampling, which adds training data to minority classes, undersampling, which removes data from majority classes, and hybrid approaches combining over and under-sampling (Douzas et al., 2019). The most widely used oversampling algorithm in land-cover classification is SMOTE which stands for Synthetic Minority Oversampling Technique. It creates synthetic samples using the k -nearest neighbors of a given point part of a minority class, selecting one random point

among the k -neighbors, drawing a line between the original point and the selected neighbor, and randomly generating samples across that line (Chawla et al., 2002).

The skewed distribution can also be tackled at an algorithmic level. For instance, classifiers can be tuned to account for the uneven distribution of classes in the training data.

Finally, cost-sensitive approaches can be set up to account for imbalanced data sets, such as cost-matrices establishing higher misclassification costs for less represented classes.

2.3.1.7 Classifier performance assessment and parameters tuning with k -fold cross-validation

Classifier parameters cannot be randomly selected as the model's performance for a given use case is directly dependent on them. However, machine learning and deep learning algorithms sometimes can be challenging to understand due to their "black-box" behavior. K -fold cross-validation is a method aiming to assess a model's performance and fine tune its parameters.

The first step of this method is to split the data into k samples of equal size (traditionally 10). Each sample is then consecutively used as a validation set and the $k-1$ other as training sets fitting the model. The mean performance is then computed between the k results obtained. This type of method is robust because it prevents the distribution of the validation set from being influential when evaluating the model's performance.

Once a classifier's performance is assessed, cross-validations can be consecutively performed, iterating through new parameter values. The parameters causing the best performances can then be selected. Multiple methods exist regarding hyper-parameter tuning, and the best choice generally depends on the size of the data set. Some methods select the best parameter combination among all the possible parameter configurations, which can be hugely time-consuming. For instance, fitting 4 parameters with three possible values generates 27 parameter combinations which have to be cross-validated k times, meaning the model has to be trained and tested $27 \times k$ times. Other methods converge much faster and imply the random search of optimal parameters or successive halving of all the possible combinations of parameters but have the disadvantage of not always finding the most accurate parameter fit.

2.3.2 Unsupervised crop classification in remote sensing

Unsupervised classification does not require labeled training samples and is a powerful tool to cluster and label large quantities of raw data. Therefore, the classifiers only rely on intrinsic statistical features to predict a sample's class. Furthermore, unsupervised

methods can be combined with supervised algorithms to achieve better accuracy resulting in semi-supervised frameworks.

The features used in unsupervised machine learning algorithms must be cautiously selected. Indeed, the accuracy and time cost of these algorithms increase drastically when fed with very high dimensional data (Friedman et al., 2001)

K-means, a popular unsupervised algorithm in remote-sensing image classification, is further detailed in the next section.

2.3.2.1 K-means clustering

K-means separates data into a predetermined number of clusters k . The clusters are computed with iterative steps by minimizing the euclidian distance between all samples within a cluster and the centroids of said cluster. It is important to note that the clustering quality is sensitive to the initial samples and centroids and that unsupervised classification algorithms are sensitive to outliers. Since the clustering quality relies on randomness, it is advised to use cross-validation to ensure the best clustering is found.

Brinkhoff et al. (2019) found k -means useful as a complement to supervised classification and Object Based Image Analysis (OBIA), adding unsupervised clustering having raised the accuracy of their method by 8%.

2.3.3 Performance metrics

Classification quality can be assessed through multiple metrics. Most of those are derived from confusion matrices which allow the visualization of a classifier's performance. Rows of the matrix represent the actual class while columns depict the predicted classes of samples.

For each class, True Negatives (TN), False Negatives (FN), True Positives (TP) and False Positives (FP) can be used to compute performance metrics such as accuracy, which describes the number of correctly predicted samples divided by the total number of predicted samples for a certain class, recall which describes the number of correctly predicted samples divided by number of samples truly belonging to a class, and specificity which gives the proportion of correct negative labelling for each class. They are respectively given by equations 2.4, 2.5 and 2.6.

$$Accuracy = \frac{\sum_c TP_c}{\sum_c (TP_c + FP_c)} \quad (2.4)$$

$$Recall = \frac{\sum_c TP_c}{\sum_c (TP_c + FN_c)} \quad (2.5)$$

$$Specificity = \frac{\sum_c TN_c}{\sum_c (TN_c + FP_c)} \quad (2.6)$$

Metrics can also be derived from recall and accuracy, such as the F-score or the G-mean, which are particularly useful indicators in case of an imbalanced data problem and are given by equations 2.7 and 2.8.

$$F - score = 2 * \frac{Recall * Accuracy}{Recall + Accuracy} \quad (2.7)$$

$$G - mean = \sqrt{Specificity * Recall} \quad (2.8)$$

Finally, the OA of the model is given by equation 2.9.

$$OA = \frac{\sum_c (TP_c + TN_c)}{\sum_c (TP_c + TN_c + FP_c + FN_c)} \quad (2.9)$$

2.3.3.1 Classifier specific metrics

The performance of binary classifiers can also be assessed through Cohen's Kappa (κ). This metric is a measure of concordance between a classifier and validation data and is given by :

$$\kappa = \frac{\alpha - P(\beta)}{1 - P(\beta)} \quad (2.10)$$

- α : The proportion of agreement between the classifier and validation data, which can be calculated by dividing the sum of the diagonal of the confusion matrix by the sum of all it's elements
- $P(\beta)$: The probability of a random correct classification

The importance of each feature in the classification process can also be computed for some classifiers such as RF. After training the model, each feature is given a coefficient ranging from 0 to 1 that can be interpreted as the weighted importance of features in the model, with the sum of all coefficients equal to 1.

2.4 Object-based image analysis and segmentation

OBIA is an object-oriented method aiming to spatially group pixels belonging to a particular class as homogeneous objects. This method prevents pixel-level spatial heterogeneity from causing a salt and pepper effect by working at object scale and not pixel scale (Vieira et al., 2012; Li et al., 2015). OBIA is particularly relevant for classifying perennial crops

where trees can have a similar scale as pixels, resulting in spatial heterogeneity between adjacent pixels (Brinkhoff et al., 2019).

If the objects are not already available and must be computer-generated, their delimitation is generally based on image segmentation. For instance, Brinkhoff et al. (2019) used a method called the superpixel non-iterative clustering to cluster pixels into objects. Other studies about OBIA have opted for the segmentation algorithm proposed by the commercial software Definiens (Li et al., 2012; Vieira et al., 2012). Li et al. (2015) also used the eCognition technique in another paper, which merges similar pixels in iterative steps starting with single pixels.

Methods used to classify segments in OBIA in literature differ depending on the context. Objects can be classified based on object-based metrics by aggregating features at polygon scale. The classification can also initially be pixel-based and then refined at object-scale. This can be applied by defining a threshold proportion of pixels that must belong to the same class within an object. If the proportion of pixels of the majority class is higher than the threshold, this class is assigned to the object, otherwise the segment is labelled as "other". The threshold method allows to assess the degree of certainty of an object's predicted class. As mentioned by Brinkhoff et al. (2019), a high threshold will ensure high accuracy for each class, while a low threshold will lower the accuracy but raise the recall. The threshold value must therefore be determined according to the objective of the classification. A hybrid method combining the two aforementioned can also be used, further refining classification accuracy.

2.5 Literature review conclusion

The literature review showed that a wide variety of methods could be used to produce classifications using remote sensing data. It was also retained that good results could be obtained in many ways, both through the multitude of algorithms and a large number of data types.

Within the framework of crop type classification, particular attention was paid to methods of computation of specific features allowing for discrimination between crop types.

It was also noted that very few literature articles focused on perennial crop classification as the sole subject, even though perennials were often mentioned.

Based on existing algorithms, two classification methods were selected, the first using RF and the second using EGB.

Chapter 3 : Objectives

3.1 General context

Copernicus is the European Union’s earth observation program. It was launched in 1998 and aims to offer information services based on satellite data and *in-situ* ground measurements. As part of this program, the European Space Agency (ESA) has developed 7 Sentinel missions to observe the earth using multiple types of sensors with various purposes such as atmospheric composition monitoring, ocean monitoring, or land monitoring. Sentinel-2 is one of these missions and provides high-resolution, operational, free, and long-term optical data. These characteristics are why the Sentinel-2 mission provides an ideal data source for designing crop mapping tools whose methodological results can be generalized to other regions.

In Belgium, as part of measures mandated by the Common Agricultural Policy (CAP), farmers are asked to declare their parcels’ administrative limits and what they contain each year in the *Système Intégré de Gestion et de Contrôle* (SIGEC), generating high-quality *in-situ* data in the process.

In this context, this master’s thesis will focus on Belgian cropping systems to implement the objectives detailed below.

3.2 Objectives

The overall objective of this master’s thesis is to develop a classification approach aiming to discriminate perennial crops from annual crops using a time series of Sentinel-2 images. A second phase will be dedicated to the design and test of a classification model aiming to classify perennial crop types.

Three main parts compose the overall approach to achieve the objectives. First, the *in-situ* and the Sentinel-2 data have to be analyzed and preprocessed according to state-of-the-art standards. Secondly, the discrimination of perennial crops requires to leverage existing features and to develop metrics defined specifically for perennials. Finally, the classification process includes the choice of a classifier and of associated methods to define

their respective optimal parameter values, which is followed by the accuracy assessment of said classifiers. The main steps of the approach are summarized in Figure 3.1, each box of the diagram will be detailed in the next chapter.

The specific objectives that are derived from the overall approach are:

- The comparison of two classification algorithms, the first one considered as the baseline is the random forest, and the second one, which currently emerges as a very efficient model with unbalanced data sets, is extreme gradient boosting.
- The design of crop group-specific features aiming to discriminate perennials from annuals based on time series analysis.
- The assessment of the relevance of successive classifications (binary then by crop type) aiming to identify perennial crop types after having masked annual crops using the created models and metrics.

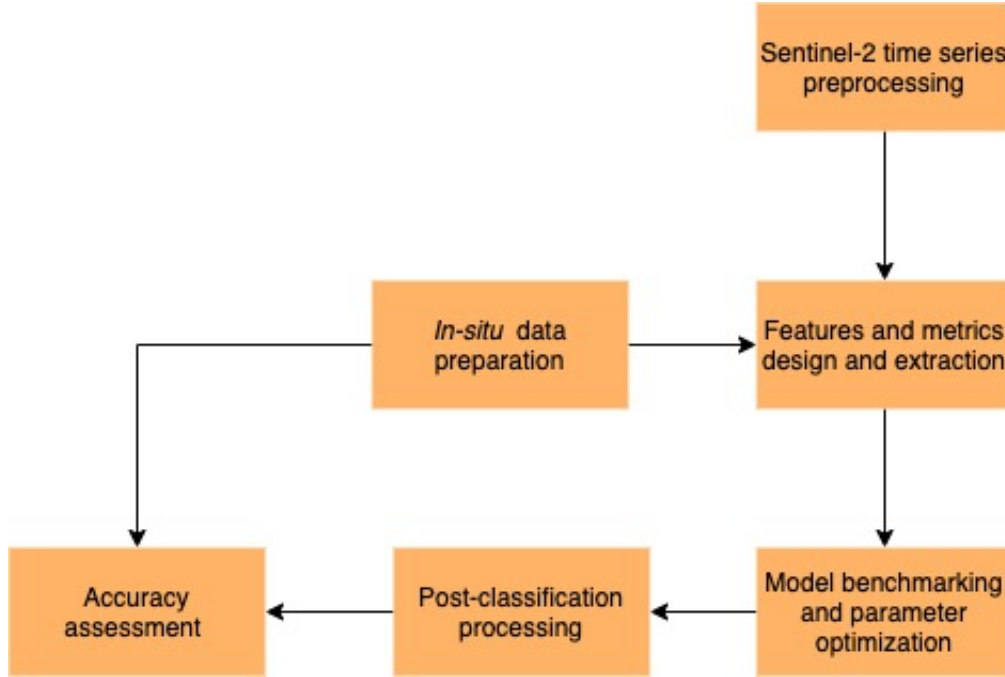


Figure 3.1: Overview of the main steps of the overall approach

Chapter 4 : Materials and method

This chapter describes the protocol followed throughout this master's thesis. The data (images and polygons) as well as all the pre-processing and processing steps will be explained.

The AOI will first be described with a brief description of the context. Then, an explanation of how the classes of the polygons were aggregated and grouped in macro-classes will be given. Through the rest of this master's thesis, The temporal system of crops (annual or perennial) will be referred to as the crop group.

The composition of the spectral bands' time series and their spatial and temporal processing will then be addressed. Afterwards, a qualitative analysis of crop-specific profiles of certain bands, indices, and metrics will be discussed.

Feature design and explanation will follow the qualitative analysis and detail the rationale behind computed metrics.

Furthermore, methods of parameters optimization, sampling design, and selection of input features of the algorithms will be detailed.

With regard to the software tools used, the programming language Python was used, more specifically the packages of vector data analysis *geopandas*, the package of raster data analysis *rasterio*, the geo-referenced data analysis package *gdal* and the machine learning packages *Scikit-learn* and *XGboost*. *Orfeo toolbox* which is an open-source remote sensing related software, was also used for some analysis. Finally, *Qgis* was used for visual inspections mapping purpose.

4.1 Area of interest

The AOI is located in Wallonia in the south of Belgium. More precisely, it is the Sentinel-2 T31UFR tile, which is a 100 kilometers of side square (Figure 4.1). Wallonia is not a region with massive agricultural production of perennial crops. In the north of the region, they mainly consist of orchards and other fruit crops, whereas the south has a greater diversity of multi-annual crops (which will be detailed in the next section). Therefore, the whole tile was used as region of interest to maximize perennial crop diversity in data.

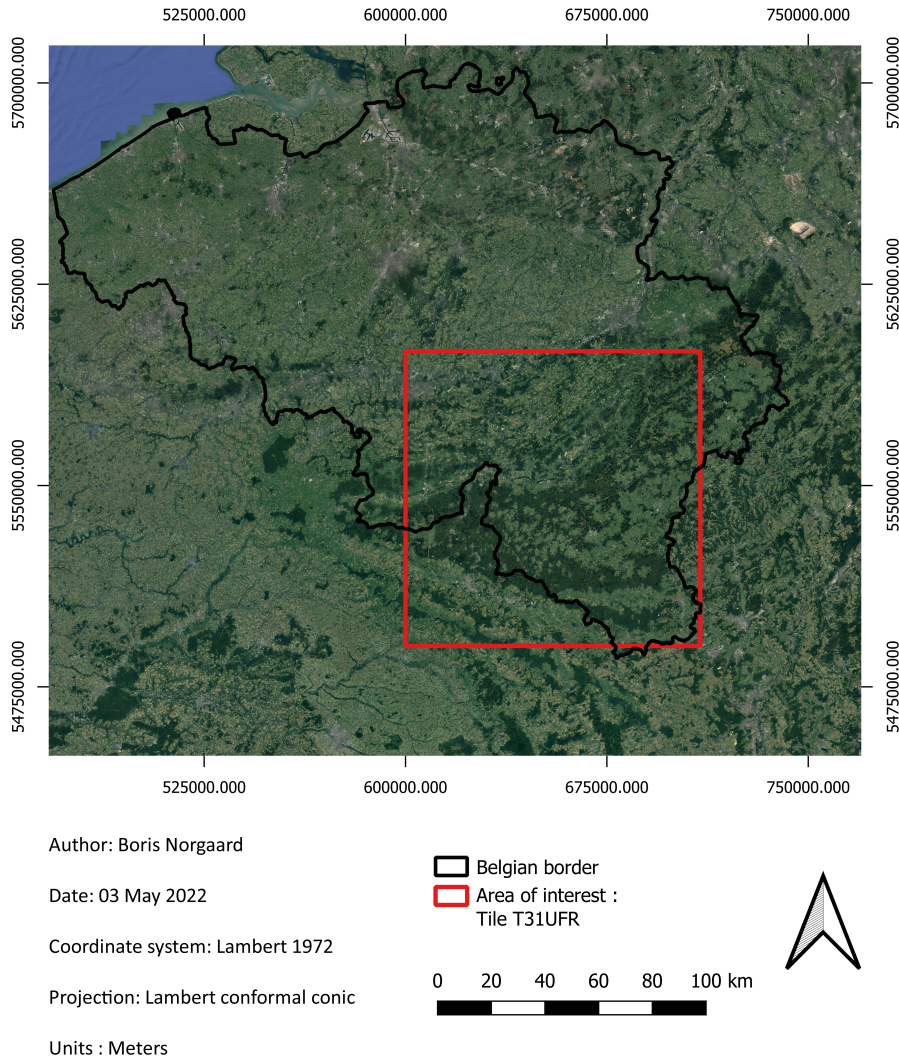


Figure 4.1: Location of the AOI (Sentinel-2 T31UFR tile) in Belgium.

4.2 *In-situ* data preparation

The polygons used as reference data come from the 2020 version of the *Système Intégré de Gestion et de Contrôle* (SIGEC) which is the Belgian parcel map of agricultural land

use. This set of polygons is updated yearly based on farmers' administrative declarations. A first screening step was applied by removing all the polygons which were not wholly included in the AOI and by removing all the parcels which were not labeled as perennial or annual crops (grassland, build-up areas, and fallows). This resulted in a shapefile containing 21134 parcels divided into respectively 20605 and 529 annual and perennial crops polygons. The perennial crop parcels in the AOI are composed of 12 distinct classes:

- Silvergrass (miscanthus)
- Fruit crops (multi-year) (high stem)
- Christmas tree
- Cultivation can not be determined
- Fruit crops (multi-year) (low stem)
- Plant nursery of fruit plants or ornamental plants
- Grapes
- Tree cultivation (forest plants)
- Hazelnuts
- Ornamental plants (perennial)
- Walnut tree
- Rhubarb (fresh consumption)

The count in terms of number of 10 meters by 10 meters pixels of each class on the tile is shown in Figure 4.2.

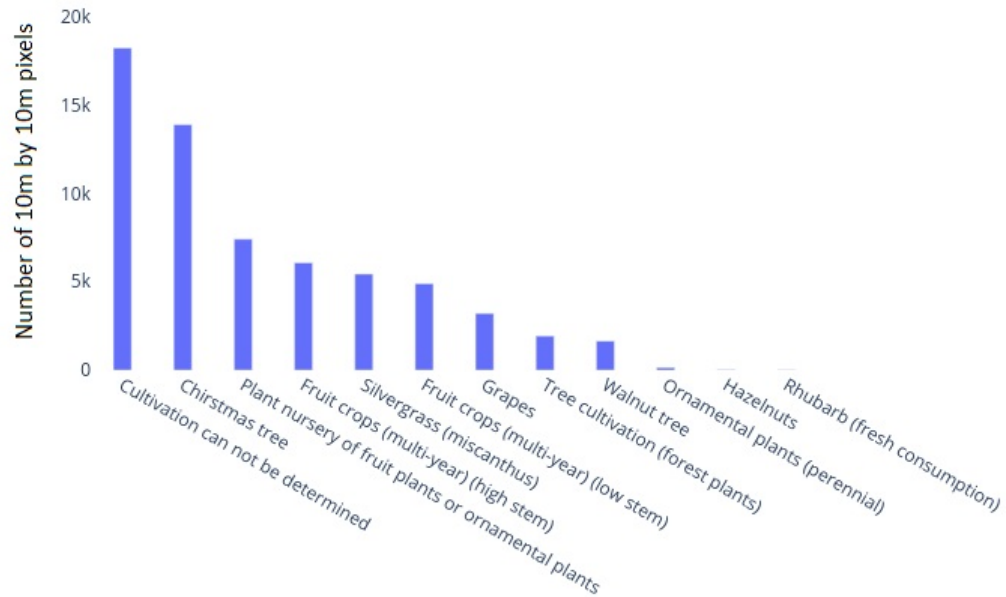


Figure 4.2: Distribution of the pixel count of each perennial class on the T31UFR Sentinel-2 tile

4.2.1 Class aggregation

As some crop types (especially annual crops and some perennial crops) show analog phenological characteristics and similar temporal reflectance profiles, multiple aggregation levels will be considered to ensure the algorithms can correctly discriminate the samples.

Table 4.1 shows the different macro-classes composed for the annual crops as well as for the perennial crops. JECAM guidelines were followed to group crops.

Table 4.1: Aggregated classes of annual and perennial crops according to JECAM guidelines

Macro-classes	Group
Cereals	Annuals
Vegetables and melons	Annuals
Oilseed crops	Annuals
Root/tuber crops with high starch or insulin content	Annuals
Beverage and spice crops	Annuals
Leguminous crops	Annuals
Sugar crops	Annuals
Other annual crops	Annuals
Fruits trees	Perennials
Vineyards	Perennials
Olive groves	Perennials
Trees	Perennials
Other perennial crops	Perennials

4.2.2 Data set split

Once the classes were aggregated, and the look-up table could be used to create a final shapefile, the data had to be separated into calibration and a validation set. To do so, a fixed proportion of each crop type in the data set was assigned to the calibration set and the rest to the validation set. The chosen proportion was set at 75% for calibration and 25% for validation based on what was used in other classification frameworks mentioned in the literature review.

4.2.3 *In-situ* data screening

In order to enhance the representativity of the training pixels, an internal buffer of 10 m was applied to all calibration and validation polygons to avoid the contaminated border

pixels. This ensures the reflectance values are not skewed by multiple vegetation types such as hedges, grass strips, or trees, which often border parcels.

Polygons too small to have a buffer applied were removed from the training and validation sets. This resulted in the removal of 294 polygons for the training set and 85 for the validation set.

To avoid training the classifiers with classes that are not well represented, some classes were removed from the calibration data.

Unwanted parcels were removed based on one of these three criteria:

- Below a minimal threshold of number of parcels of a given crop type in the whole data set (calibration and validation combined)
- When the SIGEC crop type category is ambiguous as it might contain multiple crop types
- For certain crops that can be grown in perennial and annual ways.

Therefore, all crop types not reported at least 6 times in the data set were removed. In addition, since they might contain multiple crop types or might be grown either annually or in a perennial way, the SIGEC categories listed in table 4.2 were also removed :

Table 4.2: Ambiguous crop type classes removed because of unclear composition or unclear temporal system (annual or perennial)

Removed crop type
Maize (for livestock)
Other fodder
Cereals and Legumes
Lucerne (alfalfa)
Other vegetables
Clover
Mixed protein crops cereals or others (winter)
Mixed protein crops cereals or others (spring)
Mix fodder leguminous cereals and other
Cultivation can not be determined
Ornamental plants
Rhubarb
Asparagus

Figure 4.3 shows a schematic overview of the pre-processing steps applied to the in-situ

data.

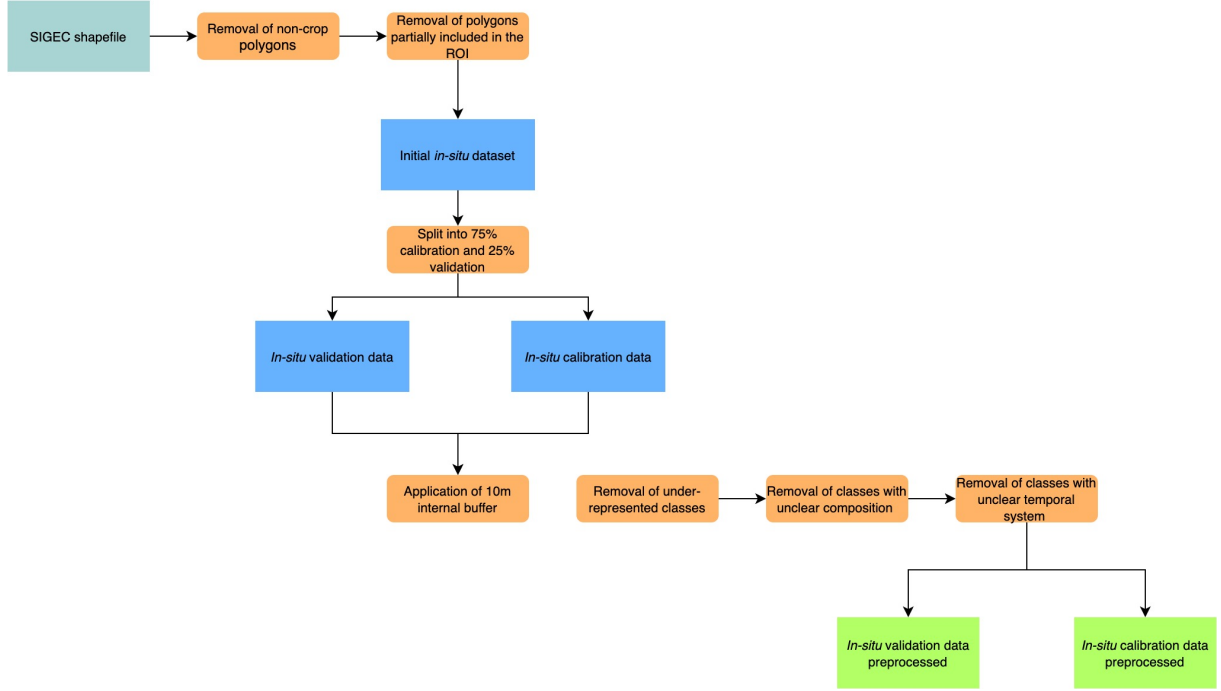


Figure 4.3: Overview of the preparation of the *in-situ* data coming from the SIGEC

4.3 Sentinel-2 data pre-processing

As the input data for the algorithms are made up of large time series, it is crucial that they are efficiently pre-processed to maximize their quality and enhance the accuracy of the classifiers. Multiple problems intrinsic to raw satellite data, such as the presence of clouds or residual atmospheric effects, must be corrected for the data to be usable.

The initial optical data consist in L2A images, meaning they have already been atmospherically corrected using Sen2Cor and the pixel values correspond to bottom of atmosphere calibrated reflectances.

Considering the diversity of spectral signatures observed through the selected crop classes but also through their phenological characteristics, the whole year of 2020 was used. Indeed, restricting the temporal window would have resulted in a loss of information for certain classes.

Finally, all the bands at 10 and 20 meters of resolution were selected for further pre-processing.

4.3.1 Time series composition and cloud masking

The total amount of acquisition dates of images on the AOI covering the year 2020 is 71, spanning from January first, 2020, to December 26th, 2020.

In order for all the images used to have the same spatial resolution, the 20m resolution bands were resampled to 10m of pixel resolution. The nearest neighbor interpolation method was used, meaning the value of newly generated pixels was determined by the value of the nearest pixel center on the input image.

These images often included clouds that had to be masked. To do so, the Sen2cor and Fmasks cloud masks were compared. Sen2cor is an image processing algorithm selected by the European spatial agency (ESA), which includes, a radiometric calibration, an atmospheric correction and a method of cloud detection providing three levels of probability (low, medium, and high) reported in a coarse scene classification which also detects snow, water, and invalid pixels. Fmask, on the other hand, is a code used for detecting clouds, cloud shadows, and snow for Landsats 4-8 and Sentinel 2 images, also providing several levels of detection confidence.

The visual comparison of several masks shows that Fmask performs better overall, which is confirmed by benchmarks issued in the framework of the Sen4Stat project.

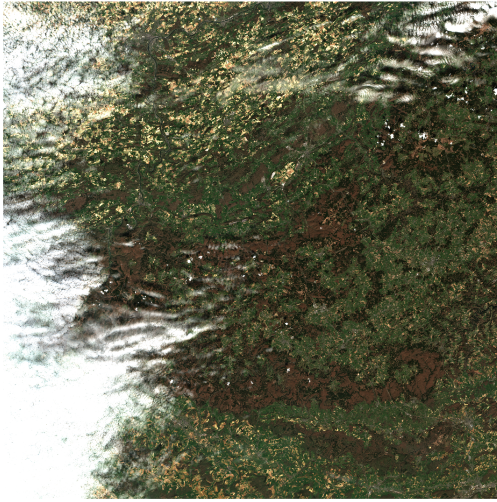
Figure 4.4 shows the difference between Fmask and Sen2cor when applying a mask on the image of April 7, 2020. Clearly, Fmask detects much more clouds, especially in the north-east part of the AOI where the cloud detection is much more homogeneous. Sen2cor also tends to detect water as clouds, which would not significantly impact perennial crop classification, but further justifies the choice of Fmask as cloud masks for the time series.

The Fmask cloud and cloud shadows masks have therefore been applied to all the corresponding images of the time series. The cloud-masked images were then stacked for each band as separate rasters, resulting in 10 images containing 71 bands (one for each date) in which cloudy pixels were replaced by no-data values.

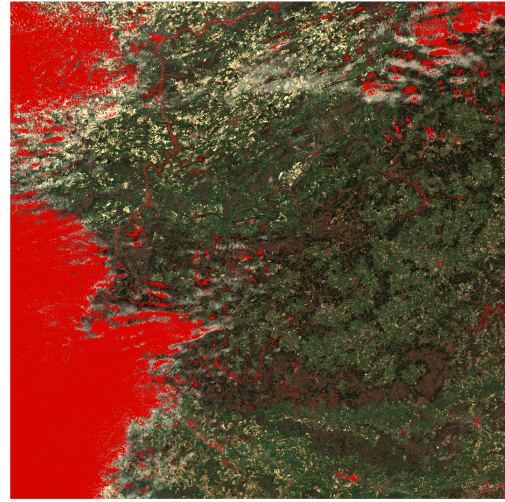
4.3.2 Temporal resampling and gapfilling

Once the resolution of all images was harmonized at 10m and the clouds masked out, the cloud-masked pixels needed to be addressed. To do so, temporal resampling at pixel scale was applied. First, a new set of artificial acquisition dates was created, beginning January 1, 2020, and ending December 31, 2020, with 10-day intervals between each date, and resulting in 36 dates. Then, for each pixel of the AOI, the time series of reflectance was computed for each band, and the cloud-masked dates were replaced by linear interpolation between the two closest cloud-free dates before and after the masked date.

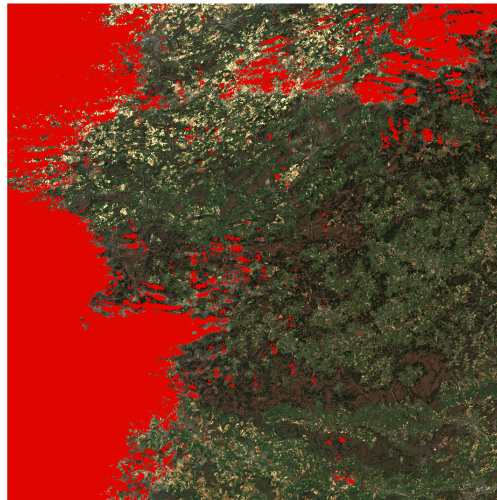
Figure 4.5 shows the pre-processing steps as a schematic overview.



(a) True color image



(b) Sen2Cor cloud mask



(c) Fmask cloud and cloud shadow mask

Figure 4.4: Comparison of Sen2Cor and Fmask on the image of April 7, 2020.

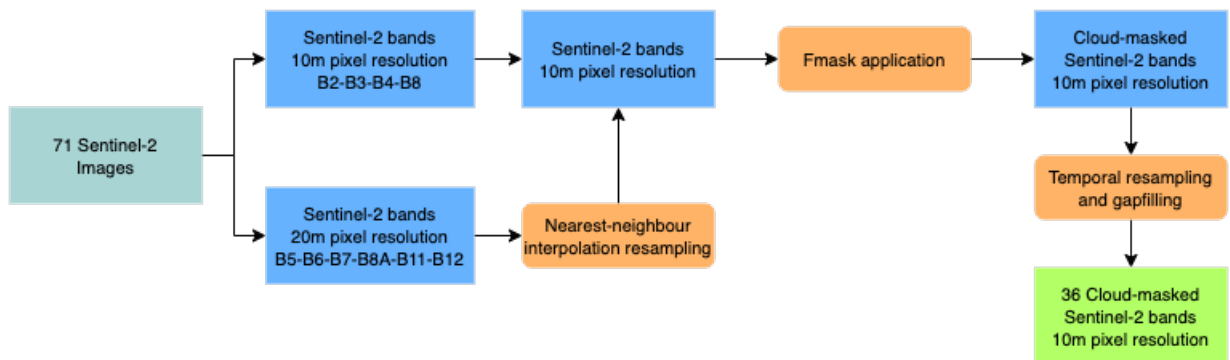


Figure 4.5: Overview of the pre-processing steps applied to the Sentinel-2 images

4.4 Spectral indices

Once the pre-processing was completed, spectral indices time series were computed. These time series can be used as features for the classifiers but can also be used to compute crop-specific metrics, which will be detailed in further sections. As all the computed indices were explained in chapter 1, their utility will not be further detailed in this section. The following indices were computed for the 36 images:

- NDVI
- NDWI
- Brightness
- BSI

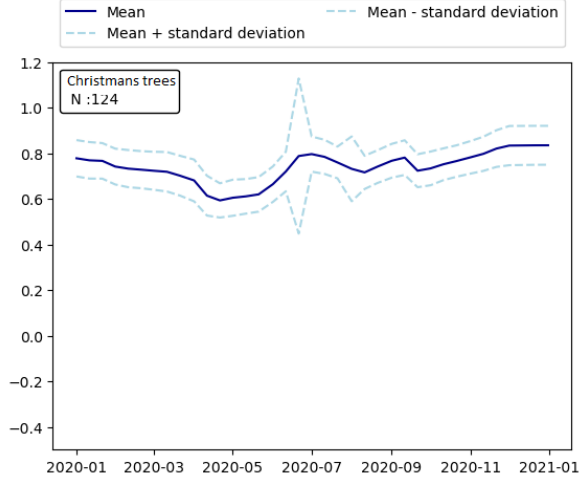
4.5 Crop-specific temporal analysis and feature design

The first mandatory step to determine the features that will allow to discriminate perennial crops from annuals and then differentiate perennial crop types is to analyze the reflectance and spectral indices time series of these crops. This allows to qualitatively understand the intrinsic differences that could be captured by the designed metrics.

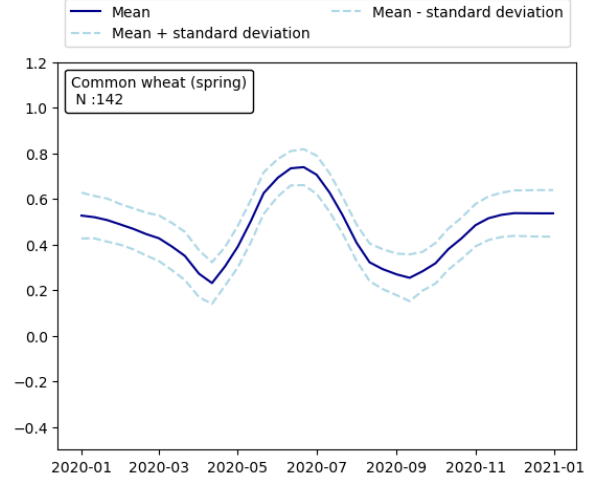
Therefore, the reflectance time series for each crop type was calculated by averaging the reflectance of a given band at parcel level for each crop type successively and then averaging for each date, resulting in mean temporal profiles for each crop type. This process was repeated for all the spectral bands and the previously computed spectral indices. This allowed to visually assess the principal differences between perennial and annual crop temporal profiles.

4.5.1 Metrics relative to growing season

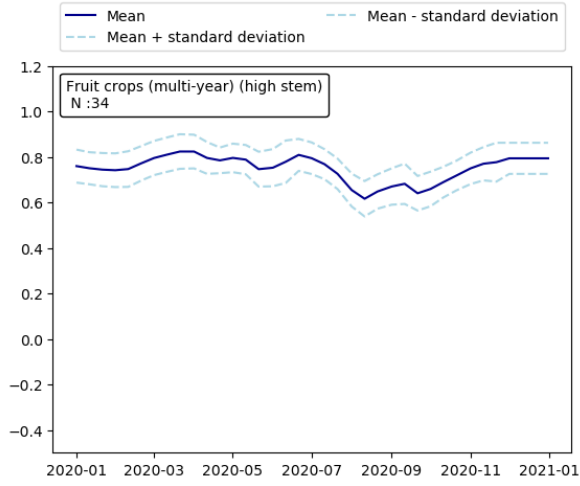
Several key differences stand out when comparing the time series of perennial plants to those of annual plants. Among these, a sustained growth period at the beginning of the season and a harvest characterized by bare soil at the end and/or the beginning of the season are the main stages of crop cycle that annuals do not share with perennial crops. Figure 4.6 shows the NDVI time series of a subset of annual and perennial crops where these stages can be clearly pointed out. The NDVI time series of every crop type can be found in appendix A.



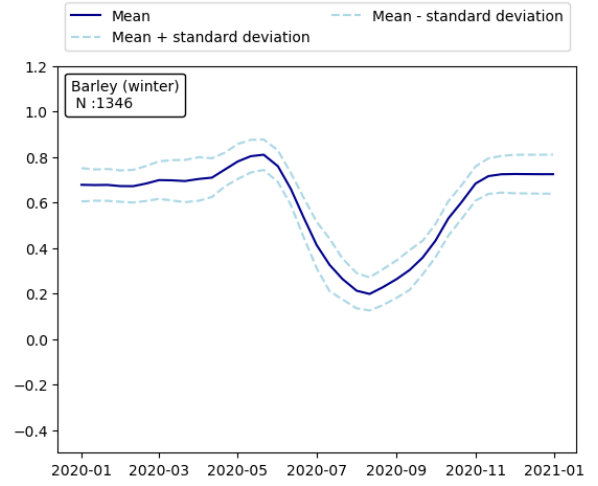
(a) NDVI time series of Christmas trees



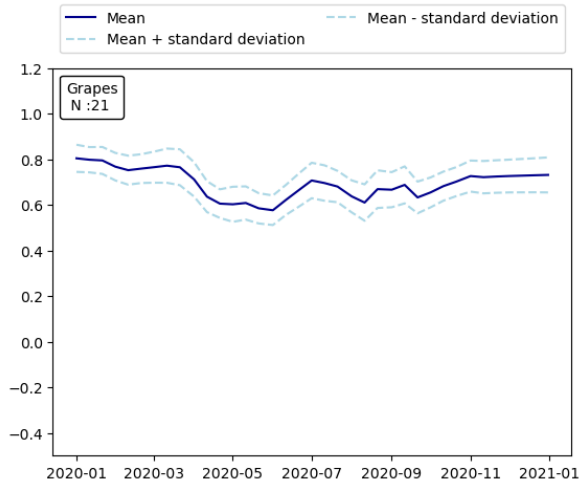
(b) NDVI time series of spring common wheat



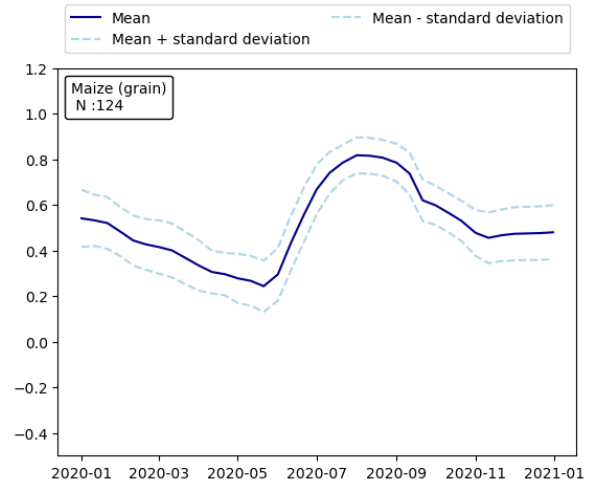
(c) NDVI time series of high-stemmed fruit crops



(d) NDVI time series of winter barley



(e) NDVI time series of vineyards



(f) NDVI time series of maize grown for grain

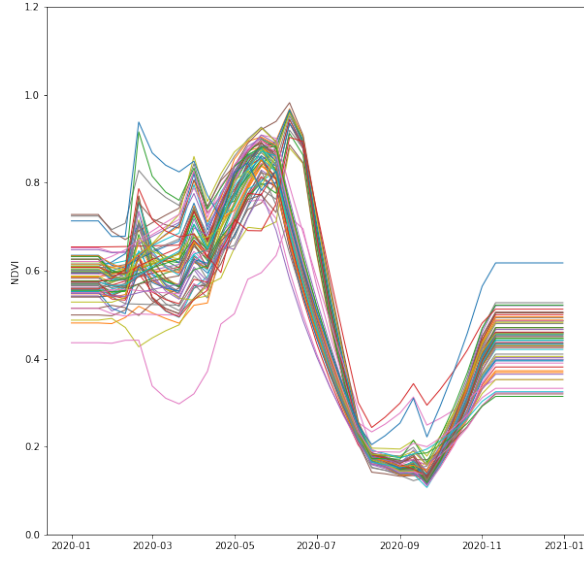
Figure 4.6: NDVI time series of 3 perennial (left column) and 3 annual crops (right column). Solid lines correspond to the mean interpolated NDVI time series of and dashed lines to the corresponding standard deviation (above and under the mean curve). The number of parcels is specified under the name of each crop in the top-left corner of each plot.

To exploit these characteristics, three metrics were designed based on the NDVI time series and two based on the BSI time series. Two NDVI metrics were based on the amplitude of the growing season, and the third took advantage of the sustained growing stage and the senescence and harvest stage of the annual crop cycle. It is worth mentioning that 12-months NDVI time series of annual crops might include more than one growing cycle, or the combined parts of two different ones, as shown on the barley NDVI time series in Figure 4.6.

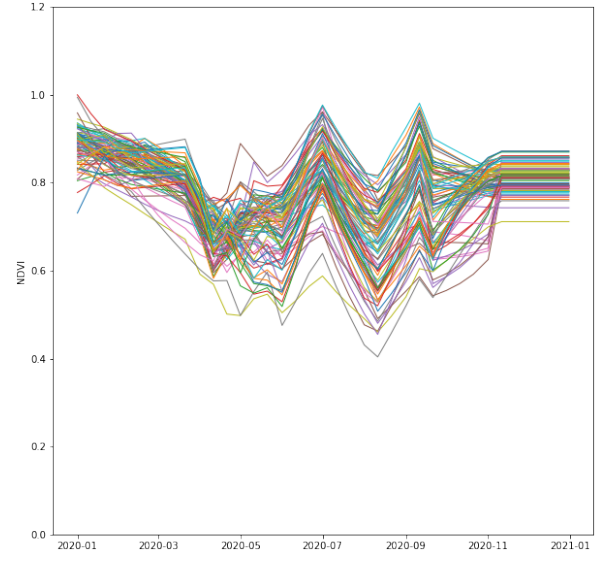
4.5.1.1 Savitsky-Golay filtering for NDVI metrics

It is necessary to underline that although the NDVI temporal profiles averaged by crop type are smooth, it would be unrealistic to consider that the same profiles at pixel level might show the same behavior. Therefore, for all the metrics relative to the NDVI, a Savitsky-Golay filter was applied to the NDVI time series to remove noise that could potentially lower the discriminating ability of the metrics. Indeed, both the amplitude, the changes of convexity, and the sign of the derivative of the time series can be impacted by noise, and as the metrics strongly rely on those properties, filtering was applied as it seemed to be the most effective method at preventing the impact from potential noise. The smoothing process is shown in Figure 4.7 where the NDVI time series at pixel scale of a vineyard and a winter wheat parcel were smoothed.

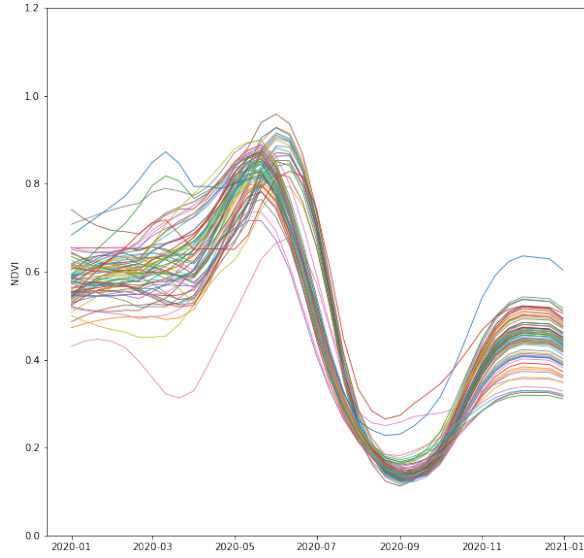
The polynomial order of the Savitsky-Golay filter was purposefully set at a high value (9) to ensure pronounced smoothing. As shown in Figure 4.7 and as will be explained in the sections specific to metrics, it is preferable to obtain a highly smoothed time series since this maintains the general shape of the time series for annuals while it slightly flattens the general shape of the time series for perennial crops in addition to removing the noise. Here, the metrics correspond to features designed on purpose for a specific use of discrimination.



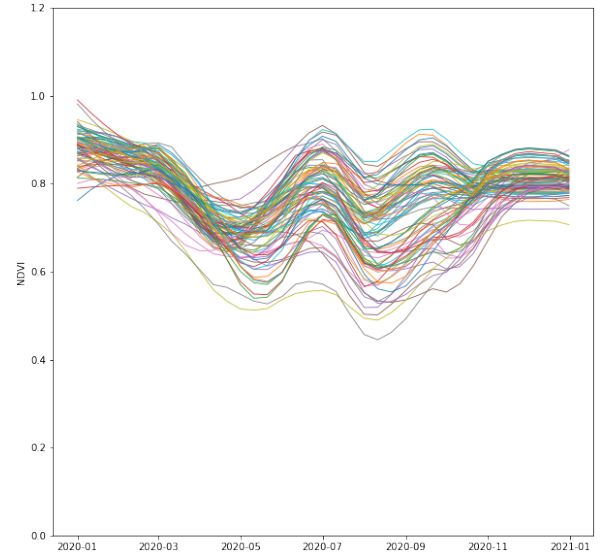
(a) NDVI time series of pixels of a winter wheat parcel



(b) NDVI time series of pixels of a vineyard



(c) NDVI time series of pixels of a winter wheat parcel smoothed with a Savitsky-Golay filter



(d) NDVI time series of pixels of a vineyard smoothed with a Savitsky-Golay filter

Figure 4.7: Comparison of the Savitsky-Golay smoothing of the NDVI time series of a vineyard and a winter wheat parcel at pixel scale

4.5.1.2 NDVI amplitude

Based on the qualitative analysis of the NDVI profiles, the NDVI profile variation along the year seems obviously different between annual and perennial crops. Therefore, the first computed metric is the corrected amplitude of the NDVI time series, given by:

$$Amplitude = \frac{\sum_{n=1}^3 \max(NDVI_t, n)}{3} - \frac{\sum_{n=1}^3 \min(NDVI_t, n)}{3} \quad (4.1)$$

with $\min(NDVI_t, n)$ the n -th smallest value of the time series $NDVI_t$ and $\max(NDVI_t, n)$ the n -th biggest values.

The amplitude is a good metric to discriminate a vast majority of annual crops from perennials when computed from the average of the three minimums and average of the three maximums to reduce the impact of outliers and local extremas . Figure 4.8 shows a map of the NDVI amplitude over 12 months and highlights the low values of the metric of perennials (black polygons) against the higher values of annual crops (green polygons). However, some crop types cannot be identified as perennials or annuals solely based on their NDVI time series amplitude.

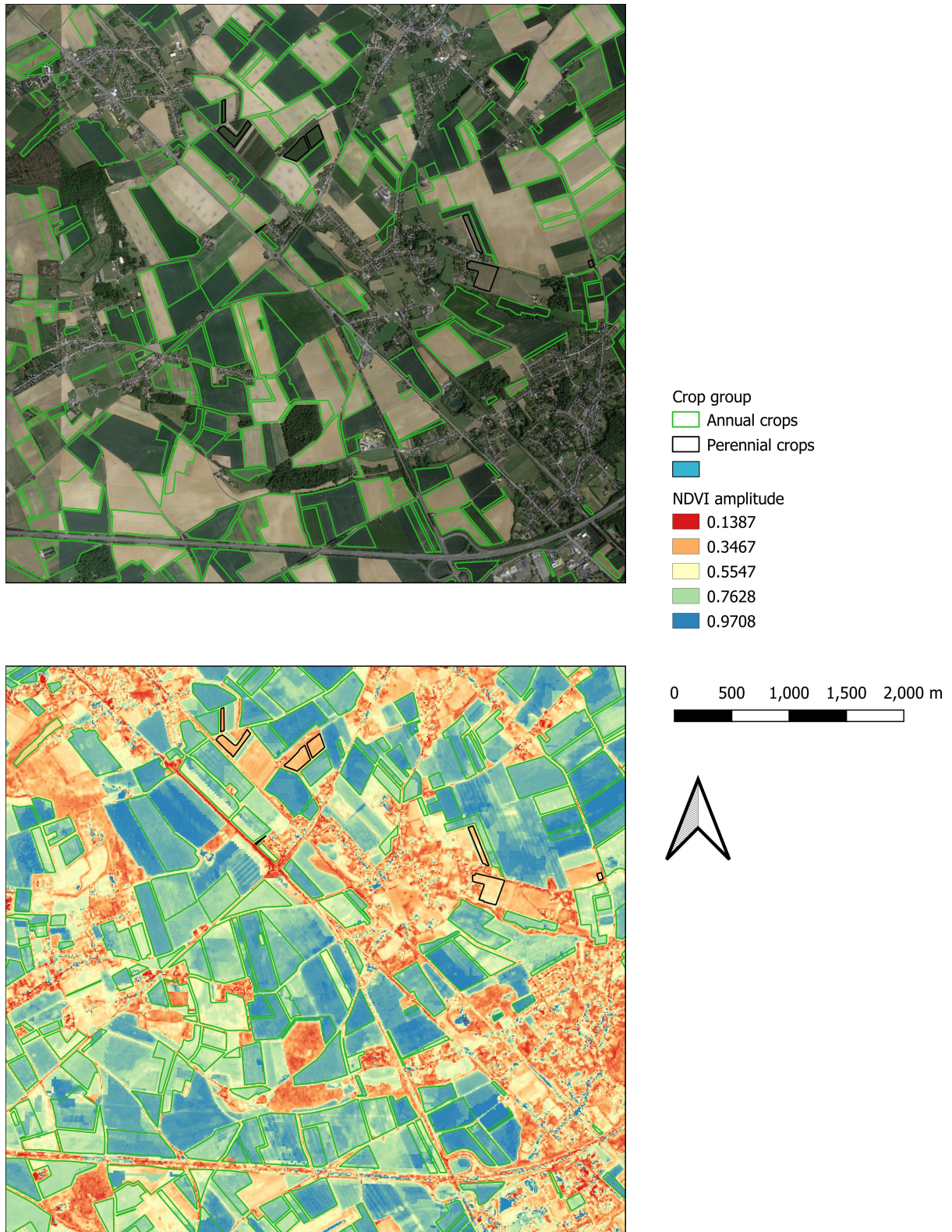


Figure 4.8: Amplitude of the smoothed NDVI time series over the North-West part of the AOI (bottom image), and high-resolution image from Google maps (top image)

4.5.1.3 Area under NDVI curve between growth and harvest

The goal of the second metric computed was to determine the area under the NDVI curve during the growing season and before harvest. This metric takes advantage of annual crops' high NDVI amplitude and the relatively stable NDVI profiles of perennials. High values would be expected for crops with high NDVI amplitudes and a prolonged growing cycle between the beginning of growth and harvest. Conversely, low values would be expected for more stable NDVI profiles.

The metric is computed in two steps, first by integrating the NDVI time series and second by removing the surface relative to the latent NDVI. The surface of latent NDVI is the surface below the horizontal line whose equation is given by 4.2 and represents the mean of the three minimal values of the time series:

$$y = \frac{\sum_{n=1}^3 \min(NDVI_t, n)}{3} \quad (4.2)$$

with $\min(NDVI_t, n)$ the $n - th$ smallest value of the time series $NDVI_t$.

The metric is then given by:

$$NDVI_{GS} = \int_a^b NDVI_{ts} dt - \int_a^b y dt \quad (4.3)$$

With $NDVI_t$ the NDVI time series as a function of time and a, b the beginning and end of the time series.

This produces a metric with very high values for annual winter crops such as winter barley or wheat, high values for other annual crops, and lower for perennial crops, as illustrated by the graphic examples in Figure 4.9 and the visual result of the metric in Figure 4.10. It should be noted that for some annual crops (winter crops in particular and all crops having a high NDVI value in winter) the produced surface under the curve is not representative of the growing season but does still differentiate the two crop groups.

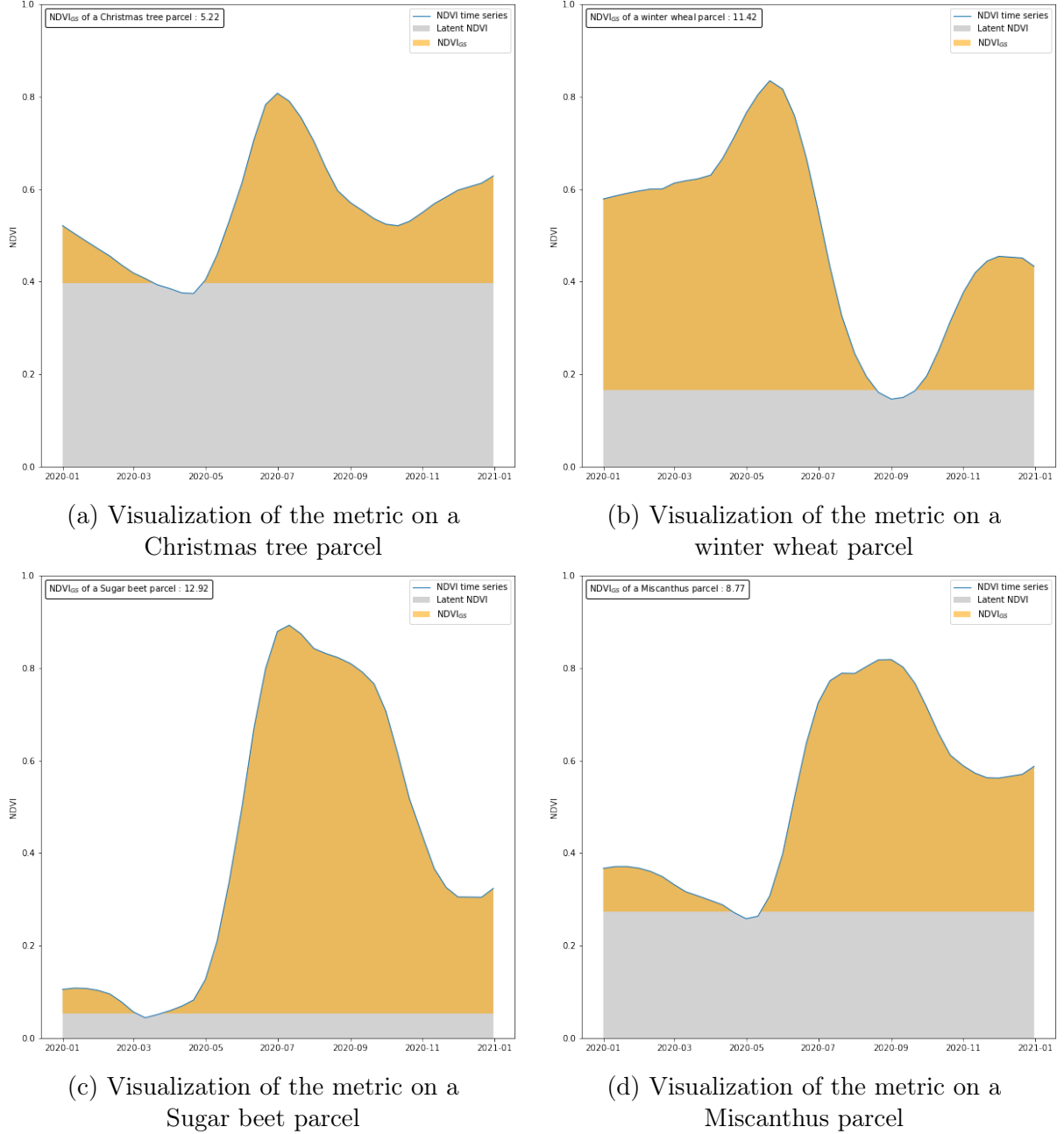


Figure 4.9: Values and illustration of the $NDVI_{GS}$ computed on mean NDVI time series at parcel level for 4 crops

As shown in Figure 4.9 and in Appendix A, some perennial crops such as miscanthus do not show stable or low amplitude NDVI profiles, with a high NDVI through the second part of the year as the miscanthus is a perennial plant following the overall greening and senescence phases of most spring crops. For these, the discriminating capacity of the metric was found to be more limited.

This metric was also computed for the red band time series to detect potential differences in chlorophyll absorption of red light between perennial and annual crops.

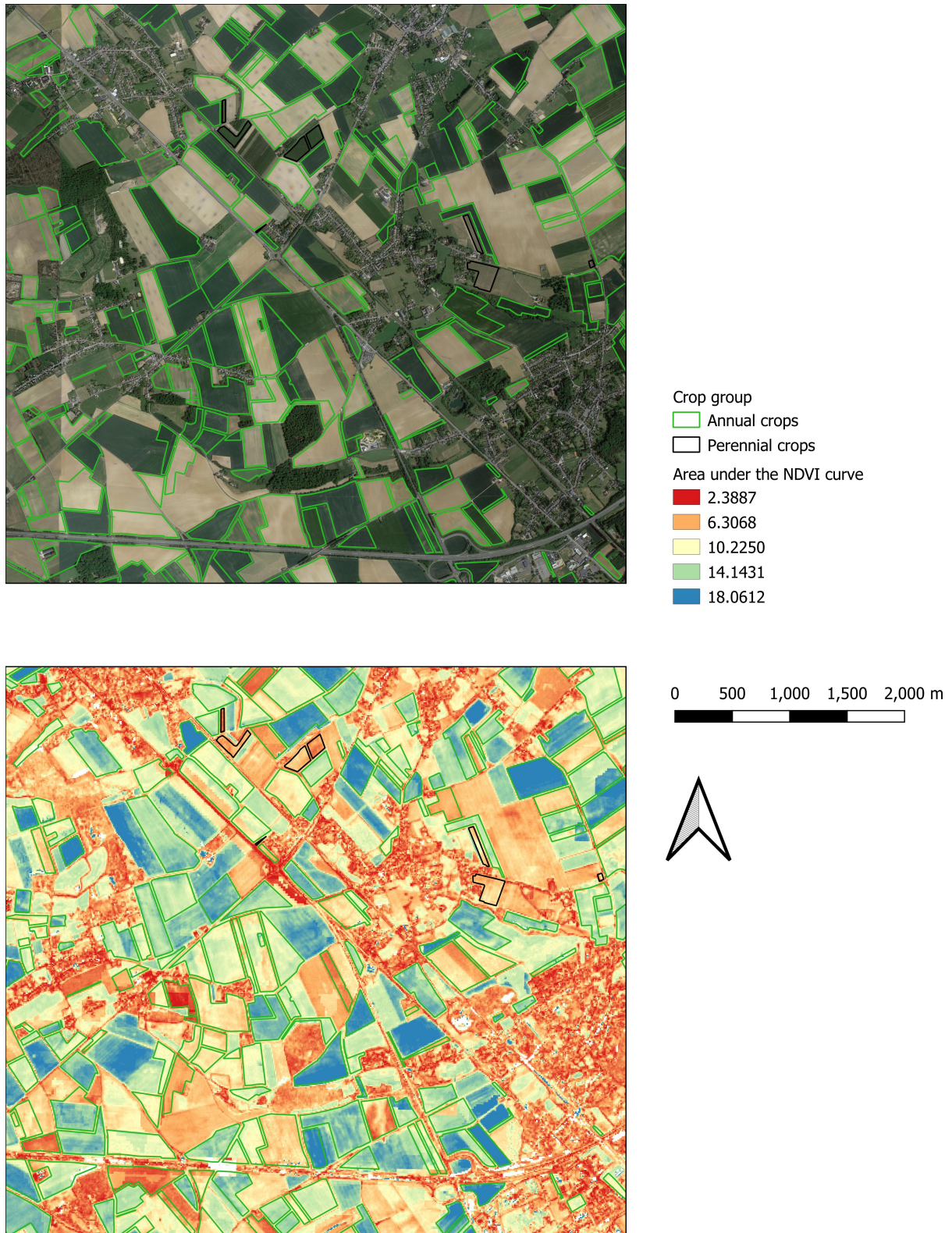


Figure 4.10: Area under the growing season of the NDVI time series of each pixel over the North-West part of the AOI (bottom image), and high-resolution image from Google maps (top image).

4.5.1.4 Area under the absolute value of the derivative of the NDVI time series

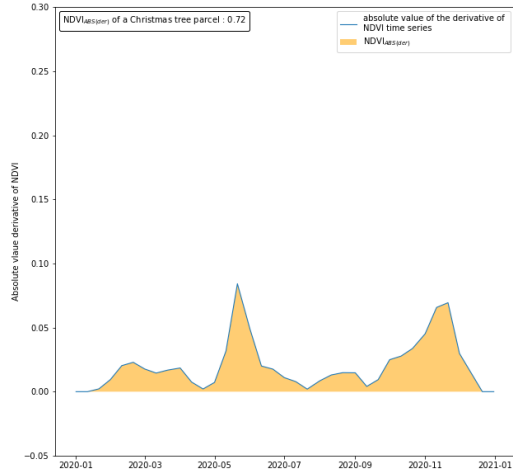
The area under the absolute value of the derivative of the NDVI time series aims to take advantage of the crop cycle stages during which the NDVI varies greatly as much positively as negatively, characterized in annual plants by the onset of the growth phase but also the senescence and harvest phase. Indeed, the area under the derivative of the time series will be higher for crops with high amplitude and cyclical NDVI. This metric was designed for discerning perennial crops such as miscanthus, whose NDVI time series amplitude is comparable to those of annual crops but does not show a substantial decrease or increase in NDVI through their crop cycle.

The metric is given by:

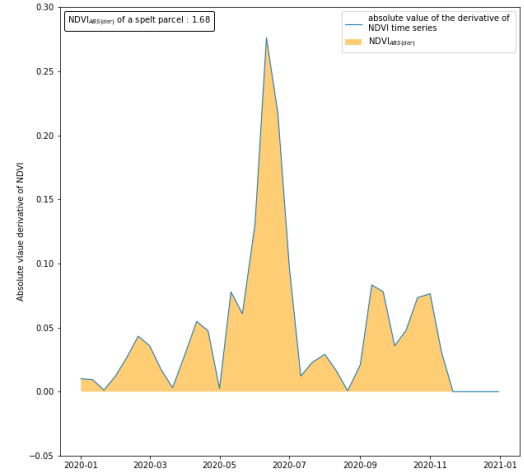
$$NDVI_{ABS(der)} = \int_a^b \left| \frac{dNDVI_{ts}}{dt} \right| dt \quad (4.4)$$

With $\frac{dNDVI_{ts}}{dt}$ the derivative of the NDVI time series as a function of time and a, b the beginning and end of the time series.

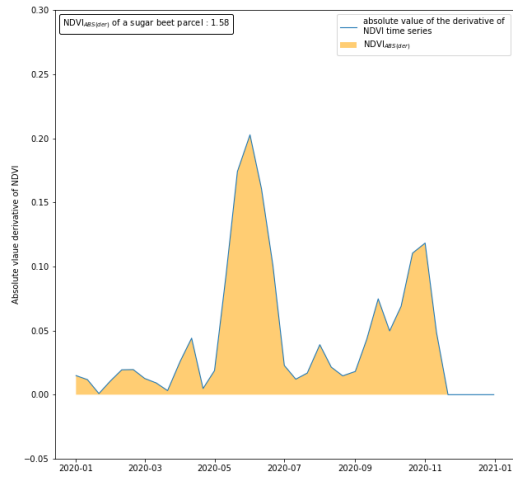
Figure 4.11 shows examples of the metrics averaged at parcel level for the same 4 crop types as for the area under NDVI curve between growth and harvest, and Figure 4.12 shows the metric as a composite over the North-West part of the AOI.



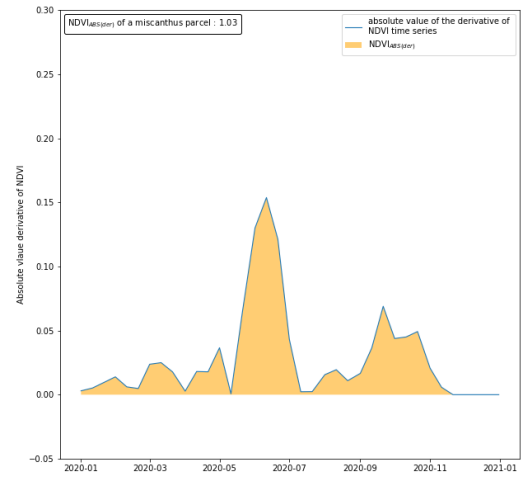
(a) Visualization of the metric on a Christmas tree parcel



(b) Visualization of the metric on a spelt parcel



(c) Visualization of the metric on a Sugar beet parcel



(d) Visualization of the metric on a Miscanthus parcel

Figure 4.11: Values of the $NDVI_{ABS(der)}$ and profiles of the derivative obtained from the mean NDVI time series at single parcel level

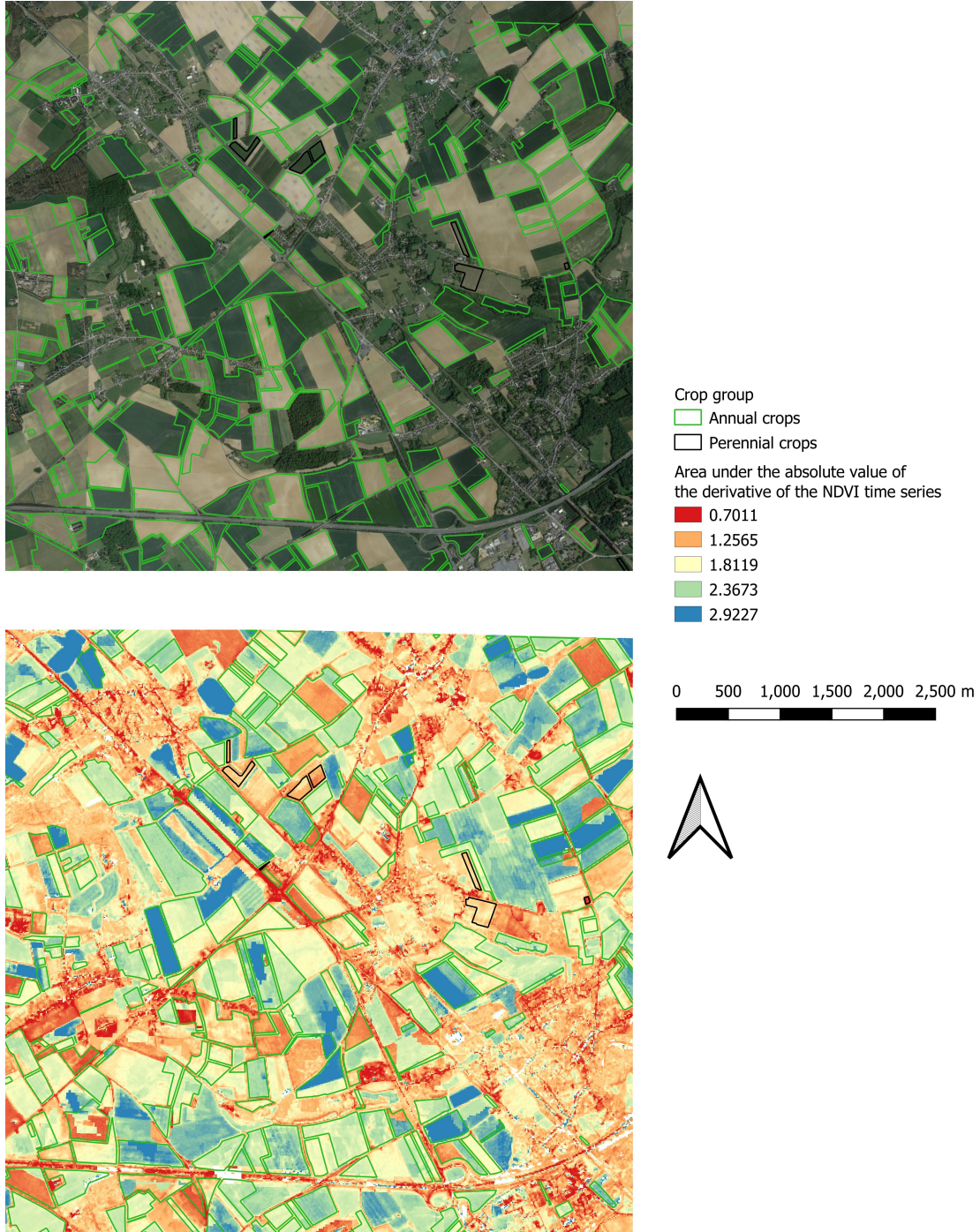


Figure 4.12: Area under the absolute value of the derivative of the NDVI time series of each pixel over the North-West part of the AOI (bottom image), and high-resolution image from Google maps (top image).

4.5.1.5 Metrics relative to the identification of a bare soil phase

One key difference between annuals and perennials is the fact that annuals are sowed harvested, and parcels go through a bare-soil phase. To exploit this information, two BSI statistics were computed in order to flag this bare soil phase.

As higher values of BSI are expected when bare soil is observed, the first metric com-

puted was the maximal value of the BSI time series. Figure 4.13 shows the resulting composite using the metric as pixel value. As for the NDVI-related metrics, a preliminary visual distinction can quickly be established, but some crop types still appear to have a high maximal BSI value among perennials (vineyards and orchards) and conversely low for some annuals.

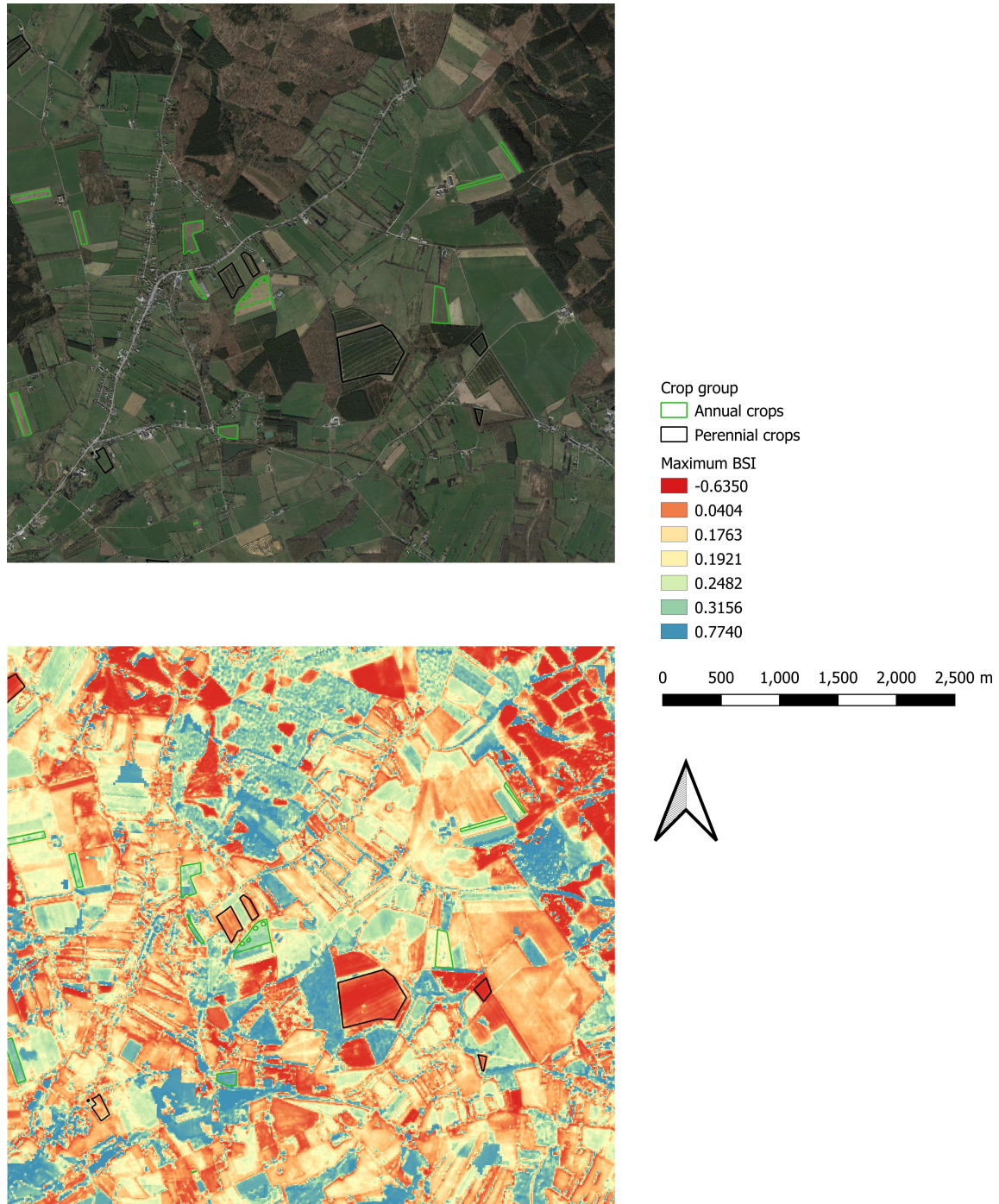


Figure 4.13: Maximum value of the BSI time series of each pixel over the South-West part of the AOI

The second metric relative to BSI is the amplitude of the time series, which was added

with the same idea as the NDVI amplitude to discern changes in BSI after the harvesting period that would not be flagged for perennial crops.

4.5.1.6 Existing features

From the literature review, temporal features (Matton et al., 2015; Valero et al., 2016) (Table 2.3 and Table 2.2) were also computed as they aimed to discern annual cropland for the first and crop types according to phenological stages for the latter. With the aforementioned metrics and existing features, the classifiers should be able to extract sufficient statistical information to separate perennial from annual crops and crop types among perennials.

4.6 Calibration set sampling design

The *in-situ* data being heavily imbalanced regarding the two crop groups, it was necessary to have an appropriate sampling method. Furthermore, to avoid spatial auto-correlation as much as possible, the sample also ideally needed to be evenly distributed through the AOI.

To obtain such samples, the *in-situ* data was first transformed from polygonal data to a raster whose pixel values were given by the factored crop type attribute of the polygons, transforming the crop types classes to numbers. Then, all the features and the raster created previously were added to a 3D matrix. Each pixel could then be thought of as a vector whose first element corresponds to the crop type of the sample, and all the others correspond to the values of the features for this pixel. Each sample was then placed in a 2D matrix whose rows were shuffled to ensure the selection would be spatially random. Figure 4.14 and 4.15 summarize the matrix transformation from the initial features to the 2D matrix of all the samples.

As the number of perennial crop samples was largely inferior to the number of annual crop samples, all pixels whose crop group were perennial were sampled from the matrix.

The annual samples were drawn from the sample matrix using a fixed proportion of the total number of pixels of each crop type. In order to do so, the proportion of the majority crop type class of annuals (wheat) being equal to 1000 pixels was first computed. Then, the same proportion of each annual crop type was drawn, under the condition that it resulted in a number of samples superior to 100 pixels (arbitrarily chosen).

The same procedure was followed for the sampling design of the perennial crop type classification.

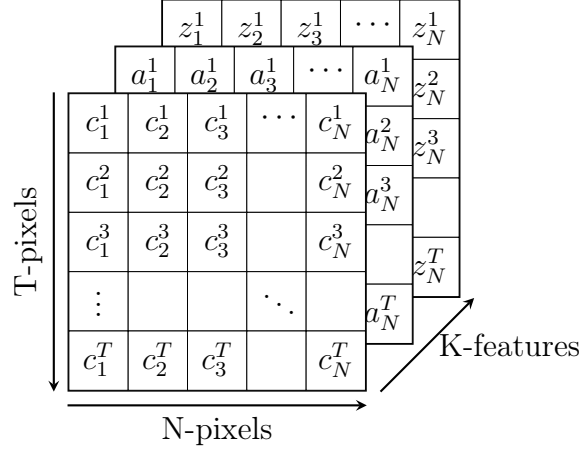


Figure 4.14: Initial raster stack

$$A_{T*N*K} = \begin{bmatrix} c_{11} & a_{11} & \cdots & z_{11} \\ c_{21} & a_{21} & \cdots & z_{21} \\ \vdots & \vdots & \ddots & \vdots \\ c_{t1} & a_{t1} & \cdots & z_{t1} \\ c_{12} & a_{12} & \cdots & z_{12} \\ c_{22} & a_{22} & \cdots & z_{22} \\ \vdots & \vdots & \ddots & \vdots \\ c_{tn} & a_{tn} & \cdots & z_{tn} \end{bmatrix}$$

Figure 4.15: Sample matrix structure

The c cells represent the classes the pixels belong to, and the a and z cells the values of the computed features. In the real class raster, the pixel values around the extent of the polygon are no-data values which are therefore not added to the sample matrix.

4.7 Feature selection scenarios, algorithm training and model comparison

To assess the impact of each feature set on the classification accuracy, different scenarios were considered regarding the input features of the classifiers. The 4 sets of features used were:

- The NDVI time series : 36 bands
- The crop group-specific metrics : 6 bands

- NDVI features (Valero et al., 2016) : 27 bands
- Crop type-specific features (Matton et al., 2015) for all bands : 56 bands

Table 4.3 summarizes the different scenario that were carried out regarding input feature composition. These were tested for the crop group prediction as well as the perennial crop type prediction (with a slight variation explained in the results chapter).

Table 4.3: Feature composition scenarios used for model calibration

Scenario	Content	Dimension of features
Scenario A	NDVI time series	$36 \times 10980 \times 10980$
Scenario B	Computed metrics	$6 \times 10980 \times 10980$
Scenario C	NDVI time series + computed metrics	$42 \times 10980 \times 10980$
Scenario D	NDVI and crop type-specific features	$83 \times 10980 \times 10980$
Scenario E	NDVI and crop type-specific features + computed metrics	$89 \times 10980 \times 10980$

4.7.1 Hyper-parameter tuning with halving random search cross-validation

As explained in the literature review, model calibration can become computationally intensive in terms of time and CPU resources when dealing with large data sets. This poses a significant problem in view of the need to find optimal hyperparameters and to perform cross-validation. Indeed, while models can be accurate with default hyperparameters given as input, optimizing them can lead to a consequent increase in accuracy. Multiple methods exist regarding hyperparameter optimum research, but the halving random search cross-validation method was adopted due to the size of the calibration set.

This method proceeds in iterative steps to test all the possible combinations of hyperparameters combined with a k -fold cross-validation. To do so, the values that must be tested are specified for each parameter, and all possible parameter combinations are computed. Then, the model is trained on $\frac{k-1}{k} * 100\%$ of a small subset of the training data for all these combinations and assessed on the remaining $\frac{1}{k} * 100\%$ of the small subset, the process being repeated k times for cross-validation. Then, a fixed proportion of the parameter combination is selected based on the cross-validated accuracy obtained on the small subset. This proportion was set at 0.5, meaning half the parameters combination were rejected based on the worst accuracy on small validation samples. The same process is repeated but with an increasing number of calibration samples until the last iteration, where all the available samples are used for calibration and k -fold cross-validation of the two last parameters combinations.

While this method does not automatically find the absolute optimum of hyperparameters, it is the best trade-off between computation time and accuracy. Indeed, compared to a method such as GridSearch cross-validation that assesses the accuracy of a model using all available samples for all the parameters combinations and k -times, halving random search cross-validation was found to be more than 10 times faster, with the first taking more than two days to test all the combinations. In contrast, the latter only took 4 hours and performed as well in terms of OA.

4.7.1.1 Random Forest

Although the method mentioned above was found to be fast and practical, the parameters and the tested values had to be carefully selected to restrain the number of possible parameter combinations and avoid lengthy calibration procedures.

The parameters, their importance in the model, and the values tested for each of them are summarized in Table 4.4

Table 4.4: Tested values for parameters in the RF optimization process

Parameter	Signification	Tested values
n_estimators	The number of trees in the forest	100,200, 500,1500,2000,3000
max_features	The number of features to consider when looking for the best split	all features, square root, log2
max_depth	The maximum depth of the tree	5,7,10,12,15
min_samples_split	The minimum number of samples required to split a node in a tree	2,3,4,5
min_samples_leaf	The minimum number of samples required to be at a leaf node	1,2,3,4,5

4.7.1.2 Extreme gradient boosting

As EGB calibration time gets more consequent with the number of input parameters, some parameters were not tuned, and their value was left at default.

The optimization parameters of the EGB model are displayed in Table 4.5.

Table 4.5: Tested values for parameters in the EGB optimization process

Parameter	Signification	Tested values
learning_rate	Weighting factor for the corrections by new trees when added to the model	0.01,0.05,0.1,0.2
max_depth	Maximum depth of the trees	4,6,8,12
min_child_weight	Number allowing nodes to split on above certain probabilities	1,2,3,4

4.7.2 Prediction

Once the models were calibrated, the crop group of the pixels of the whole AOI was predicted, followed by the perennial crop type of the perennial crop validation pixels,

resulting in two pixel-based classifications. First, all the pixels of the AOI were classified, including those belonging to non-crop land-cover. Then, this procedure was repeated for the two models and 4 feature composition scenarios.

4.8 Post-classification processing

Once the classification maps were generated, post-classification processing was needed to refine classification results further and isolate the annual and perennial crop pixels used for calibration and validation from the other pixels that are not relevant to this work.

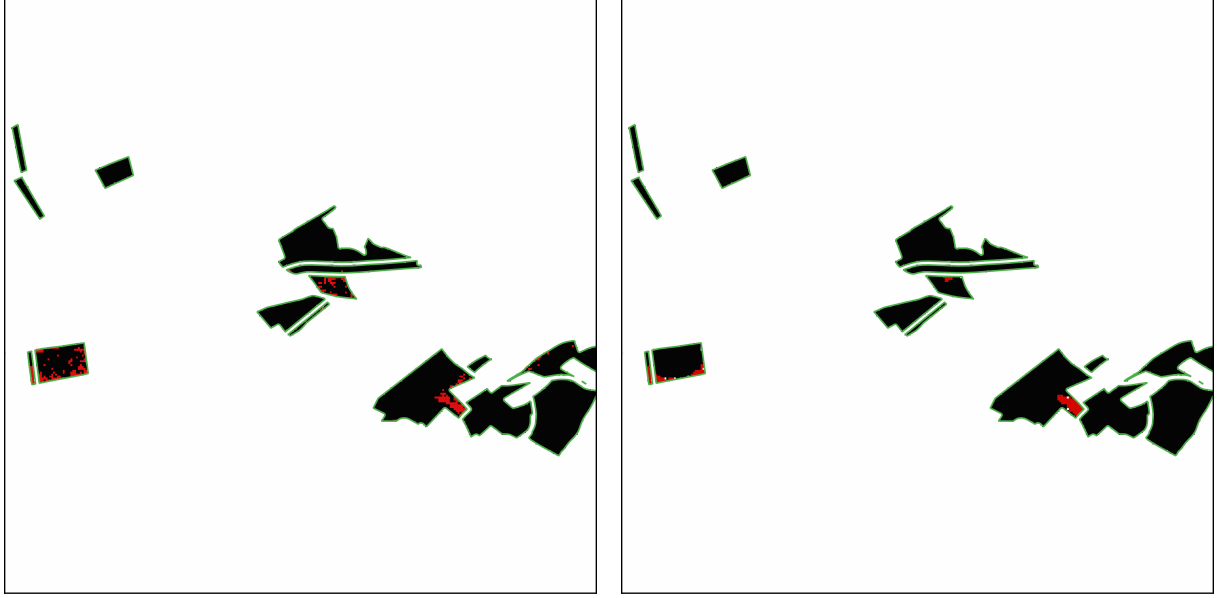
4.8.1 Pixel-based analysis

The first post-classification step consisted of applying a mask to the classifications to remove pixels not concerned by the model, like forests, built-up areas, or water. At first, a binary crop/non-crop mask was applied to the classification aiming to discriminate perennials from annuals using the polygons of the calibration and validation set as masks. It is worth mentioning that this step is only possible thanks to the high-quality *in-situ* data of the SIGEC.

A similar process was repeated for the perennial crop type classification, except only perennial polygons were selected as masks.

As the parcels whose class is to be predicted have a unique class, it seemed relevant to apply a majority filter to the predicted pixels. This method uses a sliding window to re-assign classes of isolated pixels of a given class surrounded by a majority of pixels of another class. The degree of intensity of the filtering varies depending on the window size in terms of the number of pixels and the threshold considered for a pixel to be labeled as isolated. Figure 4.16 shows an example of majority filtering having a noticeable impact on reducing the salt and pepper effect.

The majority filter was applied using a circular window with a radius of three pixels. The maximum number of neighbors with the same label as the center pixel threshold was set at 2 pixels. If a tie occurred in the count of pixels of each class in the window, the original pixel value was maintained.



(a) Unfiltered classification with noticeable salt and pepper effect

(b) Filtered classification with reduced salt and pepper effect

Figure 4.16: Results of the majority filter on a pixel-based classification

4.8.2 Object-based analysis

In addition to the pixel-based product mentioned above, an object-based aggregation of the results was also computed. The method relied on the pixel-based classification, which is transformed by aggregating the majority of predicted classes in polygons and comparing it to the actual polygon label. This method allows highlighting the classifier's errors at the parcel level, which means problematic crop types for the two models can be identified.

4.8.3 Results evaluation

The assessment of the performance of the classifiers was conducted using confusion matrices and the metrics derived from them. The overall accuracy served as a base comparison, and each class's f1-score, recall, and accuracy were used to analyze intra and inter-class prediction performance. The pixel-based approach was compared to the object-based approach using the metrics mentioned above.

4.9 Overall approach

Figure 4.17 shows a schematic overview of the method used. The part of the diagram related to the calibration and validation of models shows a generic process of model application but was repeated multiple times. Firstly for binary and crop type classifications but also to compare models and feature selection scenarios.

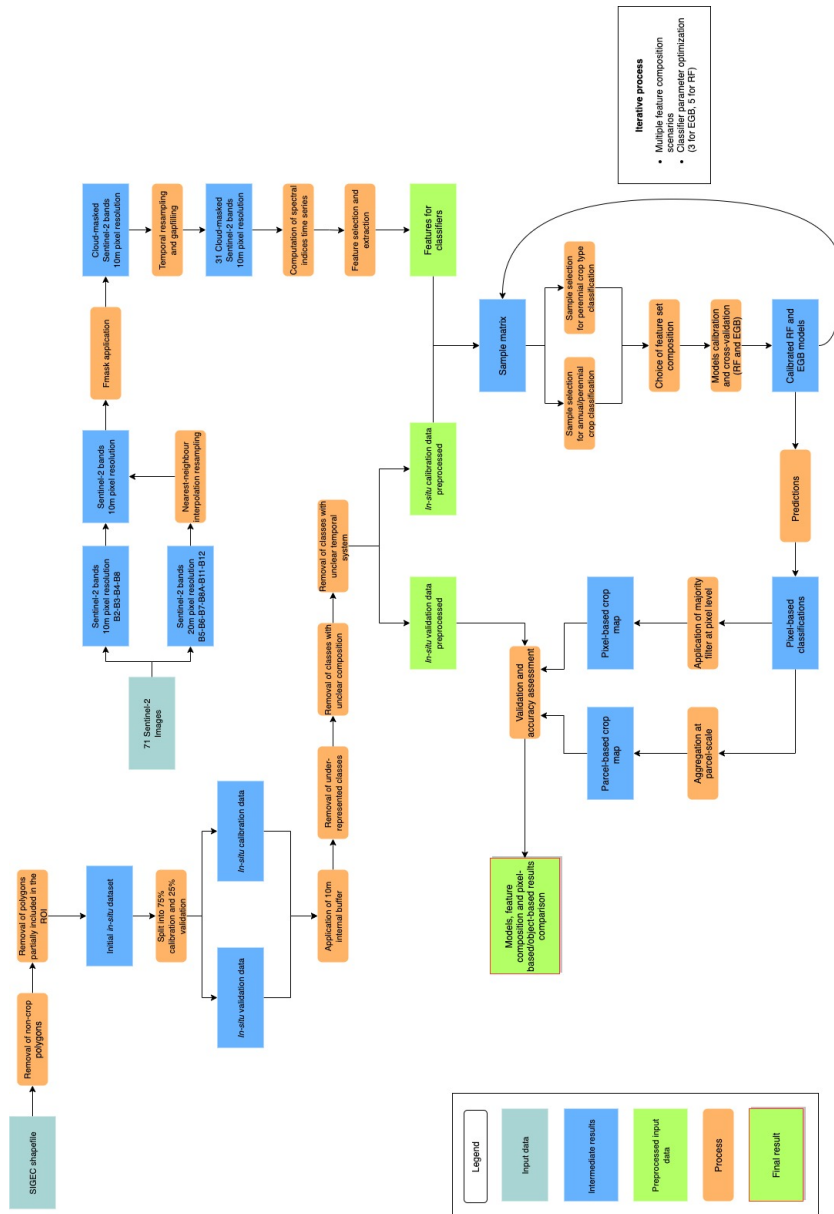


Figure 4.17: Schematic overview of the method applied through this master's thesis

Chapter 5 : Results and discussion

This chapter will report the results obtained from the implementation of the method explained in the previous section. The binary classification results aiming to discern perennial crops from annual crops will first be detailed. Then, the crop type classification results of perennial crops will be assessed. Finally, reflections emerging from these results and a discussion will follow, and the chapter will conclude with perspectives regarding potential method improvements and changes.

5.1 Crop group binary classification results

The overall accuracy results of the two models for each feature composition scenario are displayed in Table 5.1. It is essential to mention that while the overall accuracy of the models might be a good metric regarding the comparison of their global performances, as annual crops are substantially predominant in the data, this evaluation metric has to be interpreted cautiously. Indeed, with such an imbalanced in-situ data set, the models could have misclassified every perennial crop pixel and still obtained an overall accuracy of over 90%. Therefore, confusion matrices and class-specific metrics need to be analyzed to withdraw relevant conclusions about the models.

Table 5.1: Overall accuracy results of the two models tested with different feature composition scenarios for the binary annual/perennial crop classification

Classification type		pixel-based OA		filtered pixel-based OA		Object-based OA	
Model	Features	RF	EGB	RF	EGB	RF	EGB
	NDVI time series	0.984	0.992	0.987	0.992	0.974	0.983
	NDVI time series and crop group-specific metrics	0.988	0.993	0.989	0.993	0.977	0.984
	crop group-specific metrics	0.940	0.947	0.940	0.947	0.910	0.934
	crop type-specific and NDVI features	0.984	0.988	0.985	0.986	0.970	0.976
	crop type-specific features, NDVI features and crop group-specific metrics	0.986	0.990	0.987	0.990	0.974	0.982

The best overall accuracy was obtained using EGB and the NDVI time series combined with the crop group-specific metrics as input features. Overall, EGB showed better overall

accuracy results for every feature composition scenario.

The NDVI time series used as the only feature gave the second-best overall accuracy by a margin of 0.5%. The crop group-specific metrics as only input features, on the other hand, gave the worst results in terms of overall accuracy.

5.1.1 Quantitative analysis of the pixel-based classifications

To highlight the performance of the models without being biased by the unbalanced validation set, the models were quantitatively analyzed using precision, recall, and the f1-score as accuracy metrics. This allowed to compare the ability of the models to detect perennial crops among the vast majority of annual crops in the *in-situ* data.

5.1.1.1 Random forest classification results

Tables 5.2 shows the class-specific accuracy metrics resulting from the RF classification using multiple feature composition scenarios.

Table 5.2: Class specific metrics obtained for each feature composition scenario for unfiltered pixel-based RF classifications

Crop type Metric Features	Annual crops			Perennial crops		
	Precision	Recall	f1-score	Precision	Recall	f1-score
NDVI time series	0.999	0.987	0.993	0.552	0.942	0.696
NDVI time series and crop group-specific metrics	0.999	0.990	0.994	0.604	0.941	0.736
crop group-specific metrics	0.998	0.941	0.969	0.204	0.914	0.333
crop type-specific and NDVI features	0.997	0.987	0.992	0.513	0.816	0.63
crop type-specific, NDVI features and crop specific metrics	0.997	0.989	0.993	0.558	0.82	0.664
Number of validation samples		1.09*10 ⁶		1.8*10 ⁴		

It can be noted that the precision of the annual crop class is very high for each feature composition scenario, meaning the proportion of correctly labeled annual crops among all the predicted ones is high. However, as the number of annual crop samples is much higher than that of perennial samples, this metric is difficult to interpret as good or bad without putting the classification in a specific context. Moreover, the two feature composition scenarios, including the NDVI time series, showed a high recall for perennial crops meaning that almost all the pixels that should be labeled as perennials were correctly classified.

Table 5.3 shows the results obtained when applying the majority filter. In order to keep the results presented succinctly, the filtering results will only be detailed when applied to the unfiltered classification with the highest class-specific metrics. For the RF classifications, the filtering was noticed to be more effective on the models that did not include

the NDVI time series in their features and were more prone to show a salt and pepper effect in the prediction. In addition, for the NDVI time series and crop group-specific metrics feature composition scenario, an increase in precision of only 0.1% was noted for the precision of the perennial crop class. The filtering step did not result in a significant increase in model performance.

Table 5.3: Class specific metrics obtained for the filtered RF classification with the NDVI time series and the crop group-specific metrics as features

Crop type		Annual crops			Perennial crops		
Features	Metric	Precision	Recall	f1-score	Precision	Recall	f1-score
NDVI time series and crop group-specific metrics		0.999	0.990	0.994	0.605	0.941	0.736
Number of validation samples		1.09*10 ⁶			1.8*10 ⁴		

The confusion matrix of the filtered classification showing the best results is displayed in Table 5.4. Regarding perennial crop classification, this confusion matrix can be seen as problematic since slightly less than half of the pixels predicted to be perennial crops are annual crops. In this context, a high precision would have been more desirable than a high recall.

Table 5.4: Confusion matrix of the pixel-based filtered RF classification with NDVI time series and crop group-specific metrics as features

True class \ Predicted class	Predicted class	
	Annual crops	Perennial crops
Annual crops	1 082 893	11 152
Perennial crops	1065	17 063

5.1.1.2 Extreme gradient boosting classification results

Tables 5.2 shows the class-specific accuracy metrics resulting from the EGB classification using the same feature composition scenarios and optimization method as for the RF classification.

Table 5.5: Class specific metrics obtained for each feature composition scenario for unfiltered pixel-based EGB classifications

Crop group Metric Features	Annual crops			Perennial crops		
	Precision	Recall	f1-score	Precision	Recall	f1-score
NDVI time series	0.999	0.993	0.996	0.700	0.932	0.800
NDVI time series and crop group-specific metrics	0.999	0.989	0.994	0.711	0.944	0.811
crop group-specific metrics	0.995	0.937	0.965	0.296	0.853	0.439
crop type-specific and NDVI features	0.999	0.992	0.996	0.658	0.939	0.777
crop type-specific, NDVI features and crop group-specific metrics	0.999	0.993	0.996	0.695	0.946	0.801
Number of validation samples (pixels)	1.09*10 ⁶			1.8*10 ⁴		

The feature composition resulting in the best classification was found to be the same as for the RF classification. In addition, EGB showed better results for all the other scenarios. A significant increase in the accuracy metrics of the perennial crops class can be noticed when compared to RF, with a 10.7% increase in precision and a 0.3% increase in recall which result in an increase of 7.5% in f1-score for the EGB classification using the NDVI time series and the crop group-specific metrics as features. On the other hand, as for the RF classification, the filtering step primarily benefited the least accurate classifications and did not result in a significant increase in accuracy for the EGB classifications using NDVI time series and/or crop group-specific metrics as features.

Table 5.6 shows the confusion matrix of the EGB classification with NDVI time series and crop group-specific metrics as features. A decrease of 4209 misclassified annual crop pixels was observed compared to the RF version, while the number of correctly labeled perennial crop pixels slightly increased. Since the sampling design, optimization method, and feature composition scenarios were similar; it can be concluded that the EGB classifier performed significantly better. However, 28.9% of the perennial labeled predictions were still misclassified and belonged to the annual class.

Table 5.6: Confusion matrix of the pixel-based filtered EGB classification with NDVI time series and crop group-specific metrics as features

True class \ Predicted class	Annual crops	Perennial crops
Annual crops	1 087 102	6943
Perennial crops	998	17 130

5.1.1.3 Comparison of the Gini feature importances of the RF and EGB classifiers

Since the two classifiers were found to give the best results using the same feature composition, their Gini importance was computed to compare the most impactful features of the two models. The Gini feature importance of the RF model is shown in Figure 5.1 and of the EGB model in Figure 5.2. A high relative value means the feature had a significant weight in the classification process.

This metric is also called the mean decrease in impurity (MDI), with impurity being a measure of diversity in a set of samples. In essence, MDI captures the decrease of diversity in samples caused by using a variable to split nodes in a decision tree. In practice, MDI is computed for a given feature by summing the decrease in node impurity in a tree, weighted by the probability for a sample to reach that node, and then by averaging the results for each decision tree in the RF or ensemble of trees of the EGB classifier (Han et al., 2016).

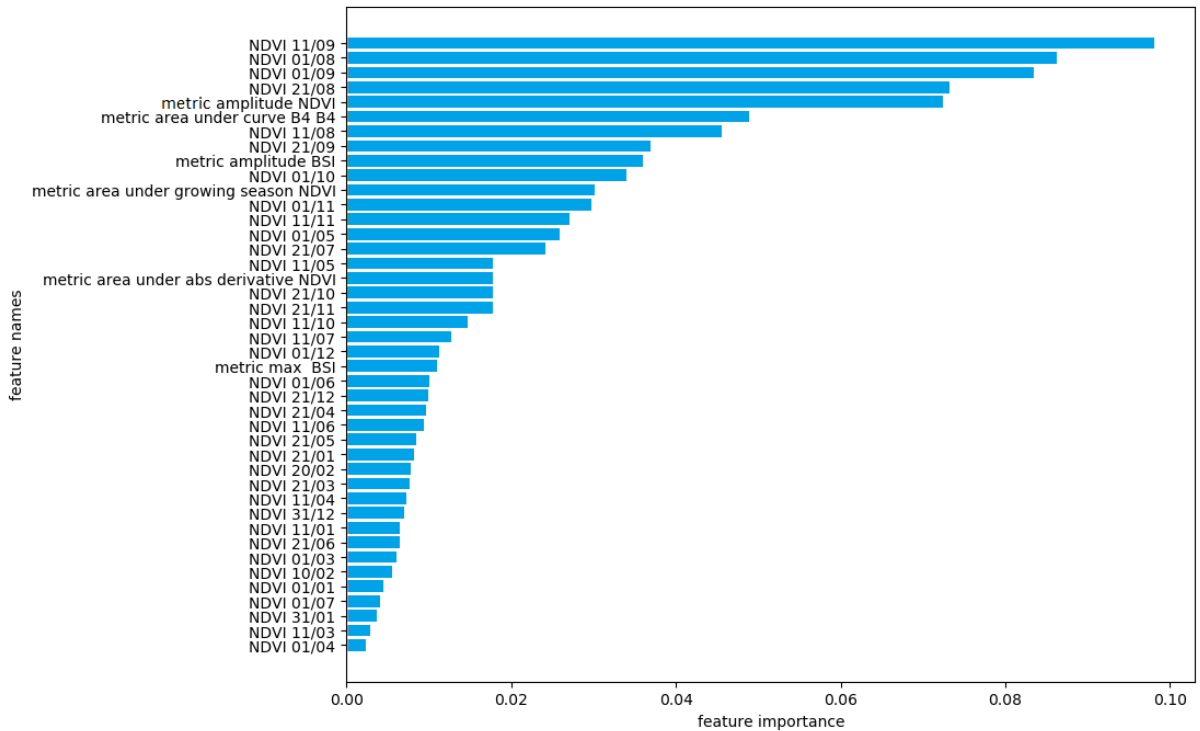


Figure 5.1: Gini feature importance of the RF using the NDVI time series and the crop group-specific metrics as features

The values of the MDI of the RF classifier were found to be relatively uniform, the biggest MDI having a value close to 0.10, gradually decreasing with the other features. An interesting result is the fact that the NDVI dates have more impact on the classification process than the crop group-specific metrics and that the NDVI values of August,

September, and October are much more impacting than the NDVI values of the rest of the year. As expected, the most impacting crop group-specific metric was the NDVI corrected amplitude.

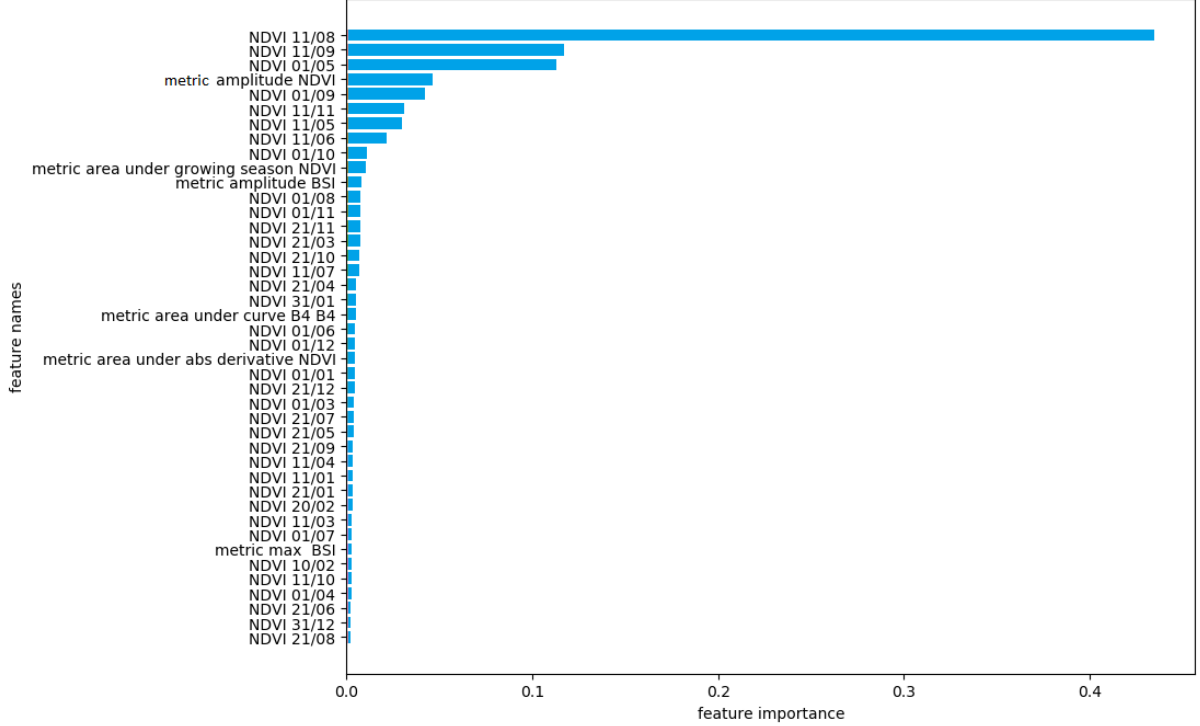


Figure 5.2: Gini feature importance of the EGB using the NDVI time series and the crop group-specific metrics as features

The computed values of MDI and the order of features in terms of the importance of the EGB classifier were significantly different from the RF classifier, with a few features having significantly higher values. The NDVI values on August 11, 2020, was the most impacting variable by a consequent margin, followed by the NDVI values on September 11, 2020. The NDVI amplitude was the most impacting crop group-specific metric, the other having a very low relative importance.

In general, NDVI values between May and October (particularly between the beginning of August and the end of September) seemed to be more prone to be impacting the models. This might be correlated with the harvest period of the majority of annual crops in the *in-situ* data (barley, wheat, or maize, for instance), which takes place between July and September.

5.1.2 Object-based classification

As it was shown, EGB classifiers performed significantly better than RF. Therefore, the object-based results will only be exposed for the EGB models for the sake of clarity and

brevity. However, confusion matrices and accuracy metrics of the object-based analysis of each feature composition scenario of the RF classifications can be found in appendix B.

Aggregating the pixel-based results to object-based was done by considering that the class of an object was given by the majority class of the pixels whose centroids were included in the polygon. Table 5.7 exposes the results obtained following the transition from pixel scale to object scale.

Table 5.7: Class specific metrics obtained for each feature composition scenario for the object-based EGB classifications

Crop group Metric Features	Annual crops			Perennial crops		
	Precision	Recall	f1-score	Precision	Recall	f1-score
NDVI time series	0.999	0.983	0.991	0.645	0.961	0.772
NDVI time series and crop group-specific metrics	0.998	0.986	0.992	0.671	0.941	0.784
crop group-specific metrics	0.995	0.937	0.965	0.296	0.853	0.439
crop type-specific and NDVI features	0.996	0.983	0.989	0.615	0.863	0.718
crop type-specific, NDVI features and crop group-specific metrics	0.995	0.981	0.988	0.578	0.833	0.683
Number of validation samples (polygons)		3267		102		

Without surprise, in light of the pixel-based results, the NDVI time series alone and the NDVI time series coupled with the crop group-specific metrics remained the features resulting in the best results. The NDVI time series alone generated the highest recall for the perennial crop class, whereas coupled with the crop group-specific metrics, it produced the highest precision and f1-score. The confusion matrices of the two feature composition scenarios are displayed in Tables 5.8 and 5.9

Table 5.8: Confusion matrix of the object-based EGB classification with NDVI time series as features

True class \ Predicted class		
	Annual crops	Perennial crops
Annual crops	3213	54
Perennial crops	4	98

Table 5.9: Confusion matrix of the object-based EGB classification with NDVI time series and crop group-specific metrics as features

True class \ Predicted class	Predicted class	
	Annual crops	Perennial crops
Annual crops	3220	47
Perennial crops	6	96

The model using the NDVI time series alone as features correctly labeled two more parcels but misclassified (as false positives) 7 more as perennials than when coupled with the crop group-specific metrics. The best model therefore depends on the context of the classification. If the context is favorable to classify as many perennial crop parcels as possible with a low impact for false positives, then the NDVI time series alone would be chosen for its higher recall. However, precision would prevail over recall in the context of successive classifications or crop mask creation, and the coupled feature sets would be used.

5.1.3 Qualitative analysis of the object-based classifications

The misclassified parcels were identified and stored in a shapefile during the object-based analysis. This was done to obtain insight into potential reasons for model failure. The first step was to identify the crop types that were not correctly labeled as perennial or annual. Figures 5.3 and 5.4 summarize the census of the misclassified crop types for the two feature composition scenarios.

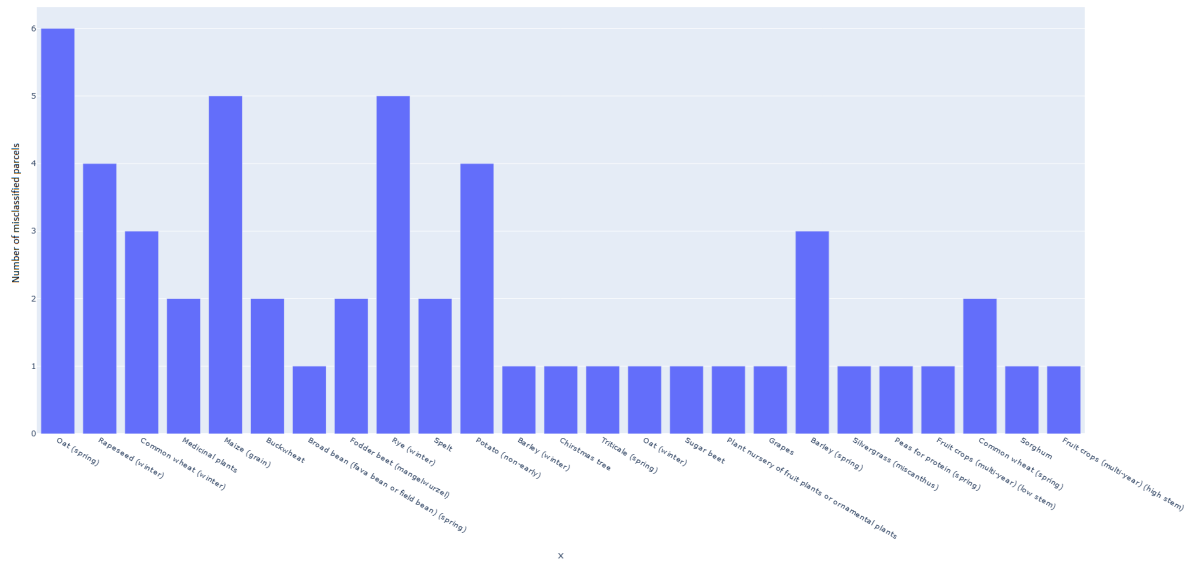


Figure 5.3: Misclassified parcels by the EGB classifier with NDVI time series and crop group-specific metrics as features

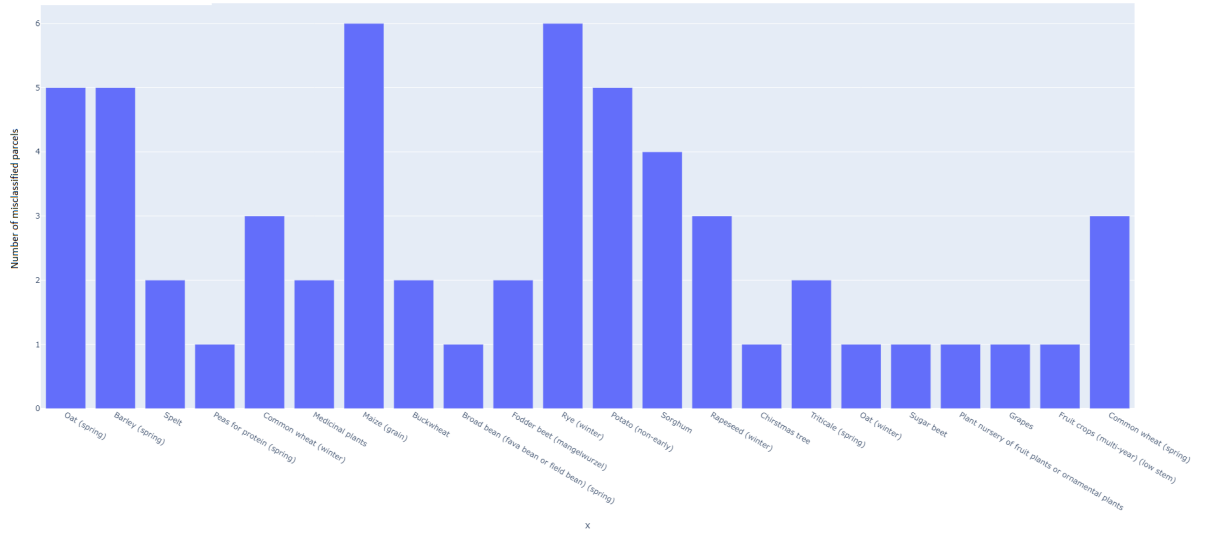


Figure 5.4: Misclassified parcels by the EGB classifier with NDVI time series as features

The census showed that the most misclassified crop types were cereals (also predominant in the *in-situ* data). In addition, the misclassified perennial crop parcels were examined in more detail as there were only 6 in the classification with the lowest recall, and the 4 misclassifications in the other classification (and were common to both of them). To do so, the probability map of classification confidence was generated by assigning the proportion of trees having voted for the perennial crop class in the ensemble to pixel values (Figure 5.5).

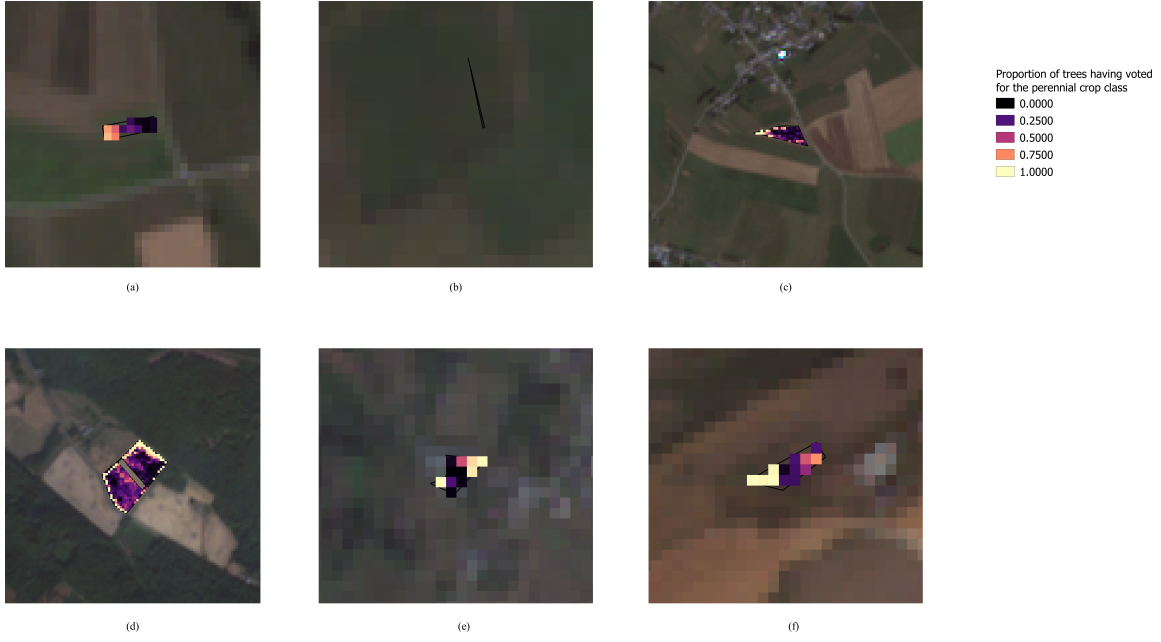


Figure 5.5: Probability map for the misclassified perennial crop parcels, with the probability being given by the number of trees having voted for the perennial crop class in the ensemble

Two parcels were detected as irrelevant, namely, parcel *b*, which does not contain any pixel's centroid, and parcel *e*, which turned out to be under a greenhouse after looking at high-resolution images. Parcels *a* and *f* which were respectively a Christmas tree and a high-stem fruit crops parcel had the left part of their surface detected as perennial with a high confidence but the rest as annual. Parcels *c* and *d* (respectively miscanthus and a vineyard) were largely misclassified as only the border pixels were labeled as perennial by a high proportion of trees in the ensemble. The mean NDVI time series of these parcels were computed and are displayed in Figure 5.6.

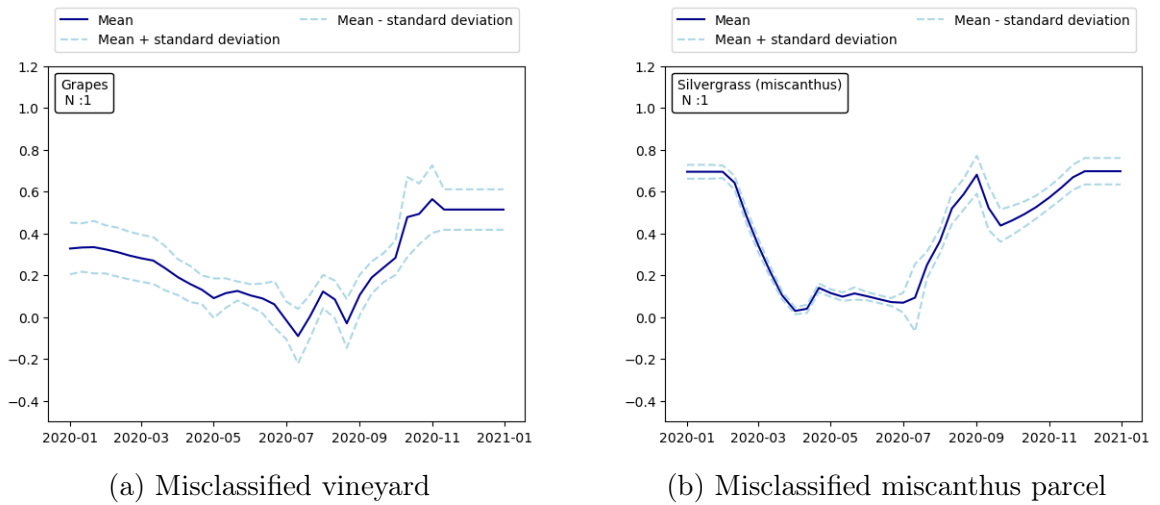


Figure 5.6: NDVI time series of two misclassified perennial crop parcels: A vineyard (left) and a miscanthus parcel (right)

The mean NDVI temporal profile of the misclassified vineyard in winter and spring was much lower than for the other vineyards in the AOI (Figure 4.6), which suggests the vines could have been planted in 2020 or the year prior. The mean NDVI temporal profile of the miscanthus shows a substantial decrease in NDVI around February. This might be caused by the crushing stage that this crop undergoes in its second year of exploit. Indeed, the harvest typically happens from March to April.

5.1.3.1 Spatial repartition of misclassifications

The repartition of the misclassified parcels was analyzed to determine whether a particular area showed a high density of incorrect predictions. As presented in Figure 5.7, no specific spatial pattern seemed to emerge beside a slightly higher concentration of misclassifications in the eastern part of the AOI.

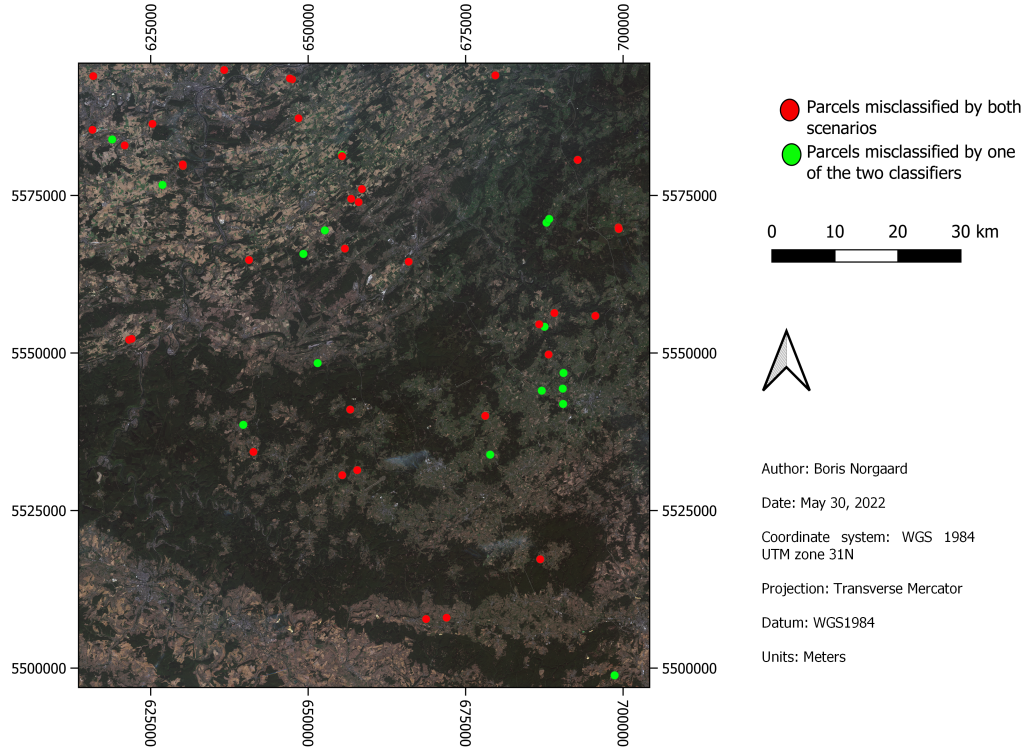


Figure 5.7: Repartition of the misclassified parcels of both models on a true color composite of the image of September 14,2020

5.2 Crop type classification results

Given the substantially better results shown by EGB for the binary classification, RF was not used for the perennial crop type classification. In addition, the crop types of fruit trees were aggregated (low-stem and high stem), resulting in 4 classes. Since NDVI features (Valero et al., 2016) and crop type-specific features (Matton et al., 2015) were specifically designed to differentiate annual crops and crop types, they were separately tested as features for a crop type classification. As the crop type-specific features performed significantly better, NDVI features were not included in the feature scenarios on their own. The composition scenarios and approaches were then compared and their results are summarized in Tables 5.10 and 5.11.

Table 5.10: Overall accuracy metrics obtained for each feature composition scenario for the pixel-based and object-based approaches using an EGB classifier

	Pixel-based	Object-based
NDVI time series	0.839	0.745
NDVI time series and crop group-specific metrics	0.847	0.755
crop group-specific metrics	0.767	0.657
crop type-specific features	0.912	0.794
crop type-specific and NDVI features	0.861	0.775
crop type-specific and crop group-specific metrics	0.910	0.784

Once again, the OA of the classifiers has to be interpreted cautiously since the classes are imbalanced. Nonetheless, the features including crop type-specific features show better results than those including the NDVI time series. The best results in terms of overall accuracy were obtained using crop type-specific features.

5.2.1 Quantitative analysis of the pixel-based crop type classifications

Table 5.11 shows the class specific accuracy metrics computed for each feature composition scenario of the pixel-based classification.

Table 5.11: Class specific metrics obtained for each feature composition scenario for the pixel-based EGB classification

Crop type		Fruit trees			Vineyards		
Features	Metric	Precision	Recall	f1-score	Precision	Recall	f1-score
NDVI time series		0.875	0.716	0.788	0	0	0
NDVI time series and crop group-specific metrics		0.891	0.781	0.834	0	0	0
crop group-specific metrics		0.483	0.424	0.451	0.023	0.007	0.011
crop type-specific features		0.844	0.780	0.811	0.715	0.211	0.325
crop type specific and NDVI features		0.673	0.659	0.666	0.049	0.003	0.005
crop type specific and crop specific metrics		0.876	0.786	0.828	0.580	0.248	0.347
Number of validation samples		2091			703		

Crop type		Christmas trees			Miscanthus		
Features	Metric	Precision	Recall	f1-score	Precision	Recall	f1-score
NDVI time series		0.897	0.928	0.913	0.351	0.512	0.416
NDVI time series and crop group-specific metrics		0.902	0.930	0.915	0.363	0.510	0.424
crop group-specific metrics		0.833	0.891	0.861	0.468	0.379	0.419
crop type-specific features		0.928	0.981	0.953	0.807	0.635	0.711
crop type specific and NDVI features		0.889	0.957	0.922	0.8	0.395	0.529
crop type specific and crop specific metrics		0.931	0.976	0.953	0.700	0.623	0.659
Number of validation samples		14517			817		

The OA shows that crop type-specific features with or without crop group-specific metrics generated the best results. The fruit trees and Christmas trees were the best-predicted classes with f1-scores respectively reaching 0.828 and 0.953, followed by miscanthus reaching 0.711, and lastly vineyards with an f1-score of 0.347.

Globally, vineyards showed inferior classification results, with a recall of 0.248 and a precision of 0.580 for the classification with the highest f1-score for that class, meaning the model labeled very few pixels as vines, but that the prediction was still correct for approximately two-thirds of them. The results regarding miscanthus were better than those of vineyards, but worst than those of fruit and Christmas trees, with a precision of 0.807 and recall of 0.635 for the model with the highest f1-score for miscanthus. Finally, fruit trees and Christmas trees showed high class-specific metrics with a recall of respectively 0.786 and 0.981 and a precision of 0.876 for the first and 0.928 for the latter.

The model only using crop type-specific features seemed to perform better at classifying Christmas trees and miscanthus parcels. Coupling the crop group-specific metrics resulted in better classification results for fruit trees and vineyards. Tables 5.12 and 5.13 show the

confusion matrices of those two pixel-based classifications.

Table 5.12: Confusion matrix of the EGB crop type classification using crop type-specific features

	Fruit trees	Vineyards	Christmas trees	Miscanthus
Fruit trees	1631	44	413	3
Vineyards	28	148	495	32
Christmas trees	192	0	14326	89
Miscanthus	81	15	202	519

Table 5.13: Confusion matrix of the EGB crop type classification using crop type specific features and crop group-specific metrics as features

	Fruit trees	Vineyards	Christmas trees	Miscanthus
Fruit trees	1643	66	378	4
Vineyards	30	174	488	11
Christmas trees	145	0	14169	203
Miscanthus	58	60	190	509

As the confusion matrices show, both models tended to classify vineyards and miscanthus pixels as Christmas trees. The confusion matrices confirm that the model using crop type-specific and crop group-specific metrics as features obtained better results for fruit trees and vineyards only by a slim margin, while the model using only crop type-specific features performed better for miscanthus and Christmas trees.

5.2.1.1 Analysis of the Gini feature importance of EGB classifier

Feature importances in the classification models were once again compared in order to identify the most statistically valuable variables in the classification process. Figures 5.8 and 5.9 expose the importances for EGB crop type classifications with crop type-specific features alone then coupled with crop group-specific metrics as features. The two feature importance graphs showed a similar trend in values, with the most essential variable having an MDI close to 0.2, followed by three decreasing values and a uniformly low MDI for all the remaining features. The 4 crop type-specific features having the most impact in the model were also similar and were:

- The value of Band 6 (vegetation red-edge with 740.5 *nm* of central wavelength) at the moment of maximum red of each pixel
- The value of Band 7 (vegetation red-edge with 782.2 *nm* of central wavelength) at the moment of minimal NDVI slope in the time series
- The value of Band 5 (vegetation red-edge with 704.1 *nm* of central wavelength) at the moment of maximum red of each pixel
- The value of Band 4 (Red) at the moment of maximum red of each pixel

The only crop group-specific metric that seemed to have a significant impact on the classification process was the NDVI corrected amplitude metric.

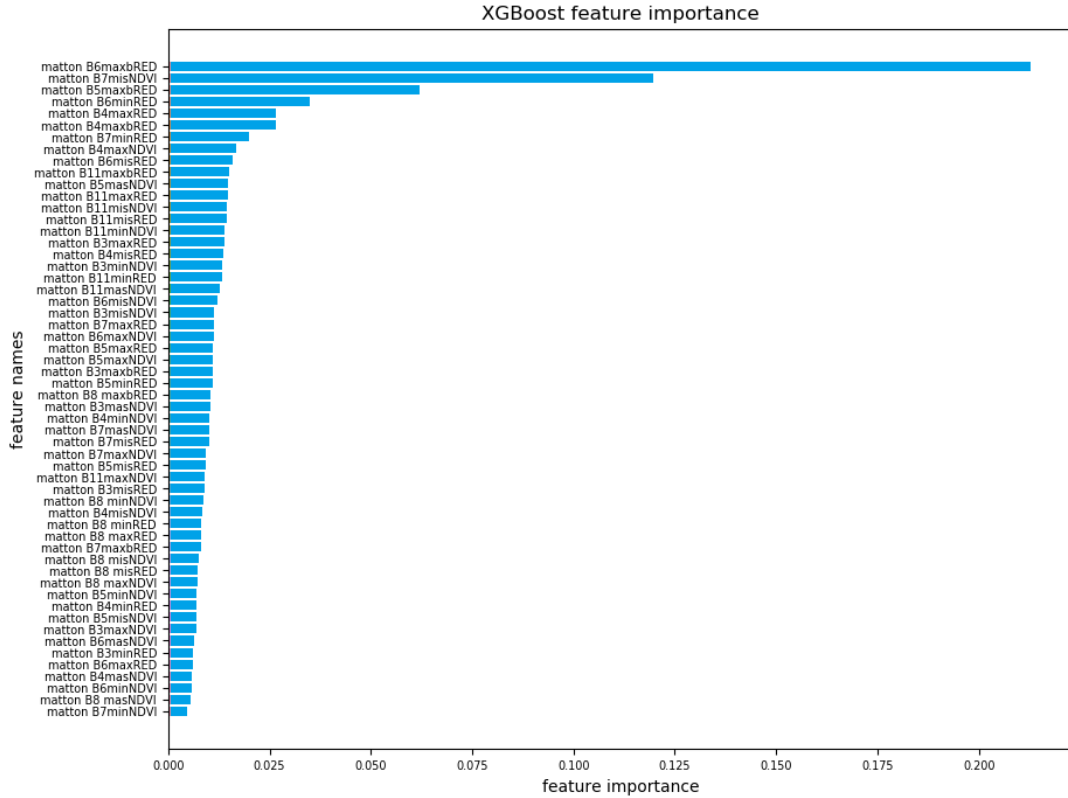


Figure 5.8: Gini feature importance of the EGB using crop type-specific features

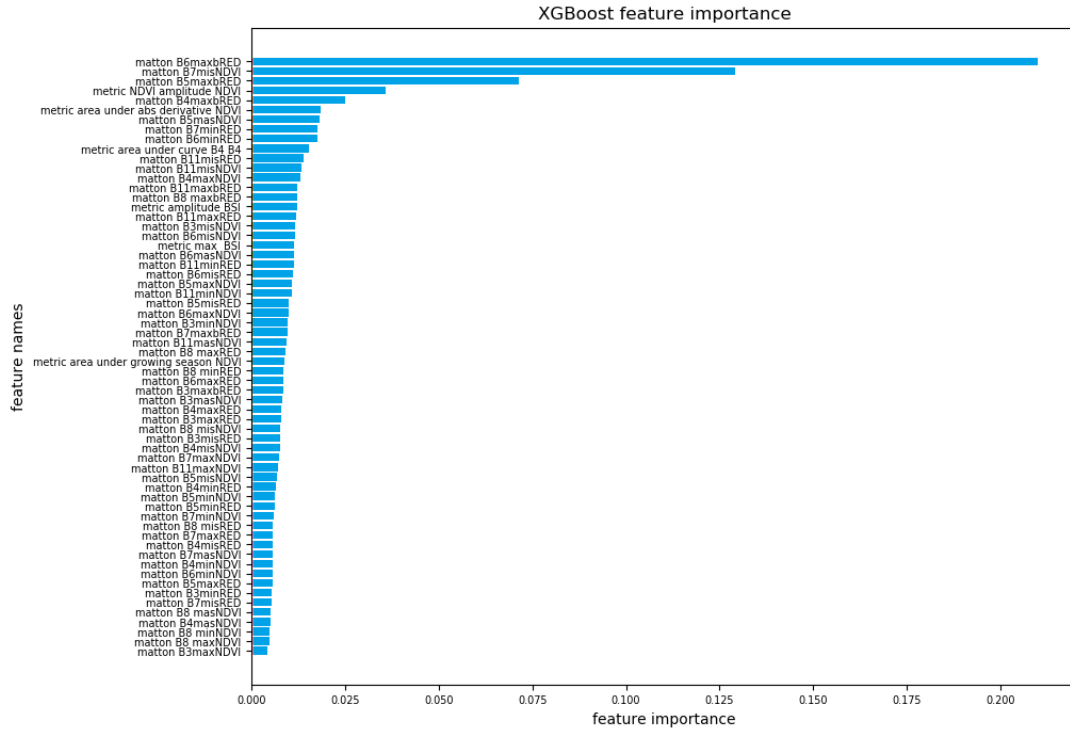


Figure 5.9: Gini feature importance of the EGB using crop type-specific and crop group-specific metrics as features

5.2.2 Quantitative analysis of the object-based crop type classification

Similar to the binary classification, the object-based crop type classification of perennial crops was obtained by assigning the majority class prediction to each polygon. The class-specific accuracy metrics resulting from the aggregation at object-scale for each feature composition scenario are shown in Table 5.14.

Table 5.14: Class specific metrics obtained for each feature composition scenario for the object-based EGB classification

Crop type		Fruit trees			Vineyards		
Metric	Features	Precision	Recall	f1-score	Precision	Recall	f1-score
	NDVI time series	0.778	0.667	0.718	0	0	0
	NDVI time series and crop group-specific metrics	0.800	0.762	0.78	0	0	0
	crop group-specific metrics	0.667	0.476	0.556	0	0	0
	crop type-specific features	0.765	0.619	0.684	1.000	0.167	0.286
	crop type-specific and NDVI features	0.750	0.571	0.649	0	0	0
	crop type-specific and crop specific metrics	0.684	0.619	0.650	1.000	0.167	0.286
Number of validation samples		21			6		

Crop type		Christmas trees			Miscanthus		
Metric	Features	Precision	Recall	f1-score	Precision	Recall	f1-score
	NDVI time series	0.754	0.929	0.832	0.667	0.526	0.588
	NDVI time series and crop group-specific metrics	0.773	0.911	0.836	0.667	0.526	0.588
	crop group-specific metrics	0.649	0.893	0.752	0.778	0.368	0.5
	crop type-specific features	0.784	0.967	0.866	0.900	0.600	0.720
	crop type-specific and NDVI features	0.753	0.967	0.847	1.000	0.600	0.750
	crop type-specific and crop specific metrics	0.781	0.950	0.857	1.000	0.600	0.750
Number of validation samples		60			15		

The two feature composition scenarios giving the best results in pixel-based classifications also led to the best results when analyzing at the object scale. The crop type-specific features alone and coupled with crop group-specific metrics as features resulted in the only two classifications able to identify vineyards. However, the number of detected parcels was very low (identical for both) compared to the actual number of vineyards, as conveyed by the recall of 0.167. Also, using the crop type-specific features, the precision of the miscanthus class significantly increased, reaching 1.0 after the transition from pixel scale to object scale. In addition, the recall was found to be lower, reaching 0.600, which meant

a little over half the miscanthus parcels were identified by the classifier. Concerning fruit trees and Christmas trees, the f1-scores decreased, but the results were still high, with almost all the Christmas tree parcels identified using crop type-specific features, leading to a recall of 0.967.

Tables 5.15 and 5.16 show the confusion matrices obtained for the object-based classifications using crop type-specific features alone and coupled with crop group-specific metrics. The classification results at object scale are almost identical, with only one more Christmas tree parcel identified by the first model and, respectively, 22 and 23 misclassified parcels over a total of 102.

Table 5.15: Confusion matrix of the object-based EGB crop type classification using crop type-specific features

	Fruit trees	Vineyards	Christmas trees	Miscanthus
Fruit trees	13	0	8	0
Vineyards	2	1	3	0
Christmas trees	1	0	58	1
Miscanthus	1	0	5	9

Table 5.16: Confusion matrix of the object-based EGB crop type classification using crop type-specific and crop group-specific metrics as features

	Fruit trees	Vineyards	Christmas trees	Miscanthus
Fruit trees	13	0	8	0
Vineyards	2	1	3	0
Christmas trees	3	0	57	0
Miscanthus	1	0	5	9

5.2.3 Qualitative analysis of the object-based crop type classification

Similar to the binary classification, special attention was paid to analyze the parcels that were not correctly classified. A few characteristic patterns seemed to emerge from this analysis. Some parcels were partially correctly classified, but the number of correctly labeled pixels was insufficient to attribute the correct class to the parcel. Some examples are shown in Figure 5.10. This was particularly observed for fruit tree parcels.

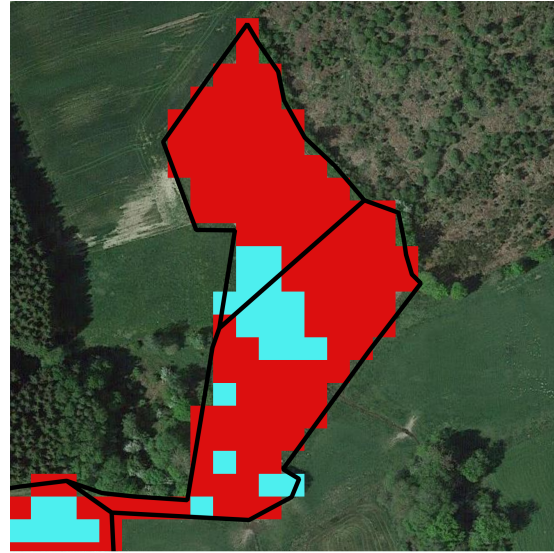


Figure 5.10: Parcels containing partially correct prediction areas overlaying the high resolution map of Google maps. The top left is a miscanthus parcel, upper middle, top right and lower left parcels are fruit trees and lower middle is a vineyard.

Two parcels were deemed irrelevant as the first one was a greenhouse and the second one seemed to show two distinct crop types as shown in Figure 5.12 with the high-resolution image and NDVI corrected amplitude metric.



(a) High resolution image of the parcel



(b) Classification results

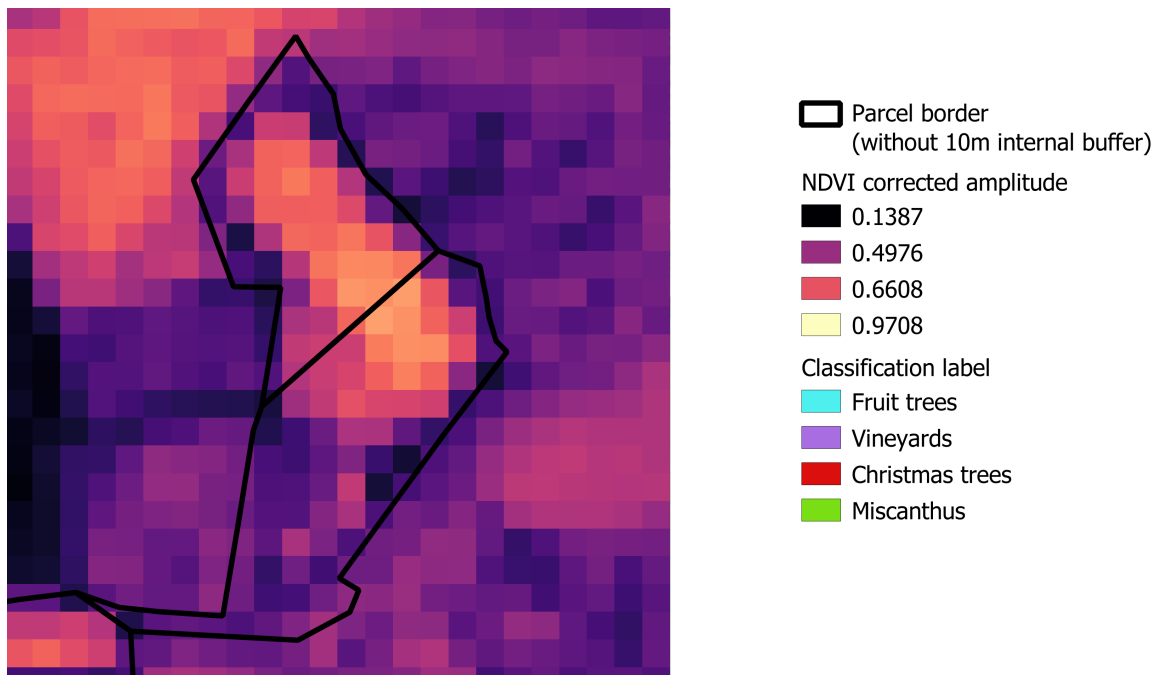


Figure 5.12: Parcel potentially containing multiple crops, highlighted by NDVI corrected amplitude and a high resolution image from Google maps (a)

Most of the rest of the parcels were misclassified as Christmas tree parcels.

5.3 Discussion and perspectives

The results led to the conclusion that the NDVI time series coupled with crop group-specific metrics was the most effective set of features for the binary crop group classification. These features resulted in a very high recall, with 96 out of 102 correctly identified parcels. However, the precision of the annual crop class was found to be too low for the model to be used in the framework of hierarchical classifications or to create a perennial crop mask. Indeed 47 parcels ended up as false positives. The NDVI time series alone showed similar results except for a slight increase in perennial crop recall and a proportional decrease in annual crop precision. As one of the objectives was to assess the feasibility of hierarchical classifications which require high precision, the NDVI time series alone was deemed less effective than when coupled with crop group-specific metrics in this context.

The crop group-specific metrics specifically designed to differentiate perennial from annual crops did not show the expected results as NDVI values at specific dates were much more impacting in the decision process of the classifiers. Nonetheless, these metrics decreased false positives, which suggests other metrics could further increase annual crop precision. In this respect, analyzing the correlations between the designed metrics could help identify redundancies in terms of statistical information between metrics, which can hardly be done by visual assessment. Furthermore, as the NDVI values at specific dates conveyed the most significant weight in the decision process, temporal metrics computed exclusively using single indices time series such as used by Li et al. (2018) could also improve the results.

Another source of error in the model might have been due to the imbalanced dataset and the small number of samples of some perennial crop types. Tests showed that the sampling method used in the binary classification led to better results than when over-sampling minority classes. However, an additional source of perennial crop samples would probably have helped to further increase the model's performance, for instance, by using the polygons and images of the Sentinel-2 tile north of the AOI. In this regard, following the same method as the one applied for this thesis to a country with a higher proportion of perennial crops would have been interesting.

Regarding the algorithm comparison, EGB showed better results than RF for all binary classification feature composition scenarios. As the optimization process was similar for both algorithms, it can be concluded that EGB was the best model in this context. It would have been interesting to test neural networks as classifiers, but the optimization process and complexity of the existing variations of those algorithms would probably require another master's thesis to obtain exhaustive results. Regarding the classification approach, the transition from pixel-scale to object-scale for both classifiers did not signif-

icantly improve classification results but allowed the examination of the misclassifications in a more detailed manner.

The perennial crop type classification showed encouraging results, with two feature composition scenarios, including crop type-specific features and the crop group-specific metrics as features giving the best results. It was noted that fruit tree orchards and vineyards showed low recalls, which could potentially be increased by computing features able to detect regular spatial patterns such as texture metrics developed by Haralick et al. (1973). This would be an opportunity to use radar data such as Sentinel-1 images that could be used for texture but also to compute metrics such as coherence.

Concerning the last specific objective, which was to determine whether the models created could be used in a framework of successive classifications, the limiting factor appears to lie in the precision of the annual crop class after the first binary classification. Indeed, while the results of the crop type classification seem promising, it would not be relevant to classify a set of parcels among which almost a third belong to the annual class. In this regard, an aggregated crop type (Table 4.1) classification without prior binary prediction would have been an interesting comparison.

Another improvement opportunity would have been to work with image time series covering a more extended period. This could allow for computing metrics based on signal periodicity or comparing spectral indices values on the same date of different years. Of course, this demands higher computing and storage capacities which are not infinite and therefore set a limit on the amount of data that could be used.

Finally, it would have been helpful to use the information residing in the undetermined SIGEC classes to train the algorithms. However, since they could contain multiple crop types or be grown annually or perennially, these classes were not relevant, which leads one to think that a stricter crop type typology of the SIGEC, in which the detailed occupation of each parcel would replace mixed crop types, could be useful.

Chapter 6 : Conclusion

This master’s thesis aimed to develop a classification framework to distinguish perennial crops from annual crops. In addition, a secondary objective aimed to develop a model to classify crop types among perennial crops and assess hierarchical classifications’ relevance.

In order to meet these objectives, two machine learning classification models were trained using features derived from Sentinel-2 image time series of the year 2020 and vector data from the SIGEC of the same year. More precisely, 6 crop-group-specific metrics were computed based on temporal and phenological characteristics being different between perennial and annual crops. Those metrics, coupled with the NDVI time series and already existing metrics, were used as classification model features in multiple feature composition scenarios.

The EGB classifier obtained significantly better results than the RF for the crop group binary classification. Moreover, the best results were obtained using the NDVI time series coupled with the crop-group-specific metrics as features. Regarding the crop type classification, the best results were obtained using crop type-specific metrics as features.

Overall, the crop group classification results were mixed, as almost all the perennial crop parcels were correctly identified, but a non-negligible number of annual crop parcels were classified as perennials. Thenceforth, it was concluded that this model could not be used in a successive classifications framework.

The crop type classification results were promising, but vineyards and orchards were found to be difficult to correctly label at parcel scale.

These conclusions leave many improvement opportunities for a potential continuation of experiments in the context of perennial crop classifications.

Bibliography

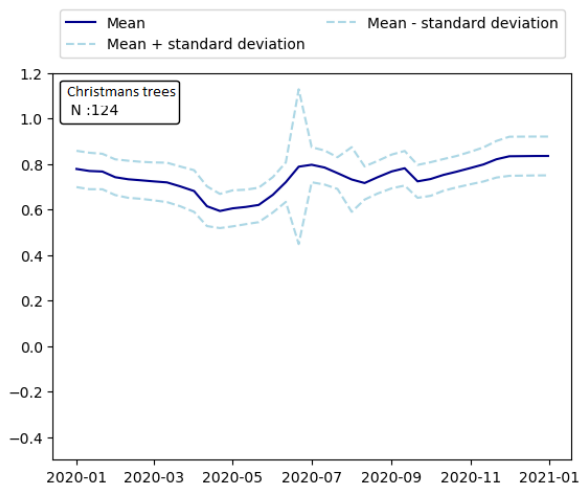
- Alganci, U., E. Sertel, and S. Kaya (July 2018). “Determination of the Olive Trees with Object Based Classification of Pleiades Satellite Image”. In: *International Journal of Environment and Geoinformatics* 5, pp. 132–139. DOI: 10.30897/ijegeo.396713.
- Azizan, F. et al. (Jan. 2021). “Application of Optical Remote Sensing in Rubber Plantations: A Systematic Review”. In: *Remote Sensing* 13, p. 429. DOI: 10.3390/rs13030429.
- Bentéjac, C., A. Csörgő, and G. Martínez-Muñoz (Nov. 2019). *A Comparative Analysis of XGBoost*.
- Beriaux, E. et al. (2021). “Sentinel-1 Time Series for Crop Identification in the Framework of the Future CAP Monitoring”. In: *Remote Sensing* 13.14. ISSN: 2072-4292. URL: <https://www.mdpi.com/2072-4292/13/14/2785>.
- Blaschke, T. (2010). “Object based image analysis for remote sensing”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.1, pp. 2–16. ISSN: 0924-2716. DOI: <https://doi.org/10.1016/j.isprsjprs.2009.06.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0924271609000884>.
- Bogner, C. et al. (Jan. 2017). “Classification of rare land cover types: Distinguishing annual and perennial crops in an agricultural catchment in South Korea”. In:
- Breiman, L. (Oct. 2001). “Random Forests”. In: *Machine Learning* 45, pp. 5–32. DOI: 10.1023/A:1010950718922.
- Brinkhoff, J., J. Vardanega, and A. Robson (Dec. 2019). “Land Cover Classification of Nine Perennial Crops Using Sentinel-1 and 2 Data”. In: *Remote Sensing* 12, p. 96. DOI: 10.3390/rs12010096.
- Chawla, N., N. Japkowicz, and A. Kołcz (June 2004). “Editorial: Special Issue on Learning from Imbalanced Data Sets”. In: *SIGKDD Explorations* 6, pp. 1–6. DOI: 10.1145/1007730.1007733.
- Chawla, N. et al. (June 2002). “SMOTE: Synthetic Minority Over-sampling Technique”. In: *J. Artif. Intell. Res. (JAIR)* 16, pp. 321–357. DOI: 10.1613/jair.953.

- contributors, J. (2014). *JECAM guidelines for cropland and crop type definition and field data collection*. Accessed on year-month-day. URL: http://jecam.org/wp-content/uploads/2018/10/JECAM_Guidelines_for_Field_Data_Collection_v1_0.pdf.
- Cortes, C. and V. Vapnik (Jan. 2009). “Support-vector networks”. In: *Chem. Biol. Drug Des.* 297, pp. 273–297. DOI: 10.1007/%2FBF00994018.
- Defourny, P. et al. (2019). “Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world”. In: *Remote Sensing of Environment* 221, pp. 551–568. ISSN: 0034-4257. DOI: <https://doi.org/10.1016/j.rse.2018.11.007>. URL: <https://www.sciencedirect.com/science/article/pii/S0034425718305145>.
- Dimov, D. et al. (July 2017). “SAR and optical time series for crop classification”. In: DOI: 10.1109/IGARSS.2017.8127076.
- Douzas, G. et al. (Dec. 2019). “Imbalanced Learning in Land Cover Classification: Improving Minority Classes’ Prediction Accuracy Using the Geometric SMOTE Algorithm”. In: *Remote Sensing* 11, p. 3040. DOI: 10.3390/rs11243040.
- Eilers, P., V. Pesendorfer, and R. Bonifacio (June 2017). “Automatic smoothing of remote sensing data”. In: pp. 1–3. DOI: 10.1109/Multi-Temp.2017.8076705.
- Eurostat (2020). *Annual crop statistics, Handbook 2020 edition*. Eurostat, p. 82.
- Fernández, A. et al. (Apr. 2013). “Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches”. In: *Knowledge-Based Systems* 42, pp. 97–110. DOI: 10.1016/j.knosys.2013.01.018.
- Formaggio, A. et al. (Jan. 2010). “Object-Based Image Analysis and Data Mining for mapping sugarcane with Landsat imagery in Brazil”. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 38.
- Han, H., X. Guo, and H. Yu (2016). “Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest”. In: *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pp. 219–224. DOI: 10.1109/ICSESS.2016.7883053.
- Haralick, R., K. Shanmugam, and I. Dinstein (Jan. 1973). “Textural Features for Image Classification”. In: *IEEE Trans Syst Man Cybern SMC-3*, pp. 610–621.
- Hatfield, J. and C. Walthall (2014). “Climate Change: Cropping System Changes and Adaptations”. In: *Encyclopedia of Agriculture and Food Systems*. Ed. by N. K. Van

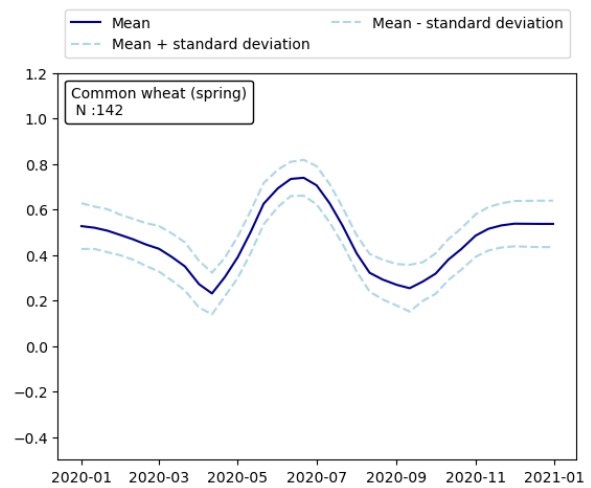
- Alfen. Oxford: Academic Press, pp. 256–265. ISBN: 978-0-08-093139-5. DOI: <https://doi.org/10.1016/B978-0-444-52512-3.00003-6>. URL: <https://www.sciencedirect.com/science/article/pii/B9780444525123000036>.
- Inglada, J. et al. (Sept. 2015). “Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery”. In: *Remote Sensing* 7, pp. 12356–12379. DOI: 10.3390/rs70912356.
- Jin, Z. and B. Xu (July 2013). “A novel compound smoother - RMMEH to reconstruct MODIS NDVI time series”. In: *IEEE Geoscience and Remote Sensing Letters* 10, pp. 942–946. DOI: 10.1109/LGRS.2013.2253760.
- Lambert, M.-J. et al. (2018). “Estimating smallholder crops production at village level from Sentinel-2 time series in Mali’s cotton belt”. In: *Remote Sensing of Environment* 216, pp. 647–657. ISSN: 0034-4257. DOI: <https://doi.org/10.1016/j.rse.2018.06.036>. URL: <https://www.sciencedirect.com/science/article/pii/S0034425718303146>.
- Lawrence, R. et al. (Apr. 2004). “Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis”. In: *Remote Sensing of Environment* 90, pp. 331–336. DOI: 10.1016/j.rse.2004.01.007.
- Lebourgeois, V. et al. (Mar. 2017). “A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM)”. In: *Remote Sensing* 9, p. 259. DOI: 10.3390/rs9030259.
- Maschler, J., C. Atzberger, and M. Immitzer (Aug. 2018). “Individual Tree Crown Segmentation and Classification of 13 Tree Species Using Airborne Hyperspectral Data”. In: *Remote Sensing* 10, p. 1218. DOI: 10.3390/rs10081218.
- Matton, N. et al. (Oct. 2015). “An Automated Method for Annual Cropland Mapping along the Season for Various Globally-Distributed Agrosystems Using High Spatial and Temporal Resolution Time Series”. In: *Remote Sensing* 7, pp. 13208–13232. DOI: 10.3390/rs71013208.
- Nanni, L. et al. (Dec. 2013). “Different Approaches for Extracting Information from the Co-Occurrence Matrix”. In: *PloS one* 8, e83554. DOI: 10.1371/journal.pone.0083554.
- Peña, M. and A. Brenning (2015). “Assessing fruit-tree crop classification from Landsat-8 time series for the Maipo Valley, Chile”. In: *Remote Sensing of Environment* 171,

- pp. 234–244. ISSN: 0034-4257. DOI: <https://doi.org/10.1016/j.rse.2015.10.029>.
URL: <https://www.sciencedirect.com/science/article/pii/S0034425715301814>.
- Raiyani, K. et al. (Jan. 2021). “Sentinel-2 Image Scene Classification: A Comparison between Sen2Cor and a Machine Learning Approach”. In: *Remote Sensing* 13, p. 300. DOI: 10.3390/rs13020300.
- Serra, P. and X. Pons (Apr. 2008). “Monitoring farmers’ decisions on Mediterranean irrigated crops using satellite image time series”. In: *International Journal of Remote Sensing - INT J REMOTE SENS* 29, pp. 2293–2316. DOI: 10.1080/01431160701408444.
- Van Tricht, K. et al. (Oct. 2018). “Synergistic Use of Radar Sentinel-1 and Optical Sentinel-2 Imagery for Crop Mapping: A Case Study for Belgium”. In: *Remote Sensing* 10, p. 1642. DOI: 10.3390/rs10101642.
- Wang, S., G. Azzari, and D. B. Lobell (2019). “Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques”. In: *Remote Sensing of Environment* 222, pp. 303–317. ISSN: 0034-4257. DOI: <https://doi.org/10.1016/j.rse.2018.12.026>. URL: <https://www.sciencedirect.com/science/article/pii/S0034425718305790>.
- Zekoll, V. et al. (Jan. 2021). “Comparison of Masking Algorithms for Sentinel-2 Imagery”. In: *Remote Sensing* 13, p. 137. DOI: 10.3390/rs13010137.
- Zhong, L., L. Hu, and H. Zhou (2019). “Deep learning based multi-temporal crop classification”. In: *Remote Sensing of Environment* 221, pp. 430–443. ISSN: 0034-4257. DOI: <https://doi.org/10.1016/j.rse.2018.11.032>. URL: <https://www.sciencedirect.com/science/article/pii/S0034425718305418>.
- Zhong, L. et al. (Nov. 2011a). “A phenology-based approach to map crop types in the San Joaquin Valley, California”. In: *International Journal of Remote Sensing* 32, pp. 7777–7804. DOI: 10.1080/01431161.2010.527397.
- (Nov. 2011b). “A phenology-based approach to map crop types in the San Joaquin Valley, California”. In: *International Journal of Remote Sensing* 32, pp. 7777–7804. DOI: 10.1080/01431161.2010.527397.

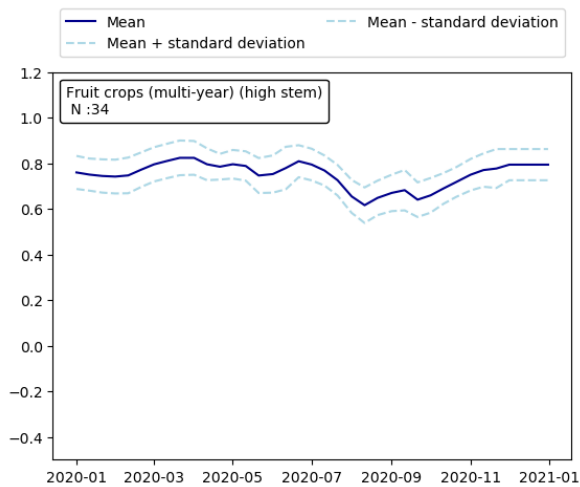
Appendix A : NDVI time series of each crop type in the *in-situ* data



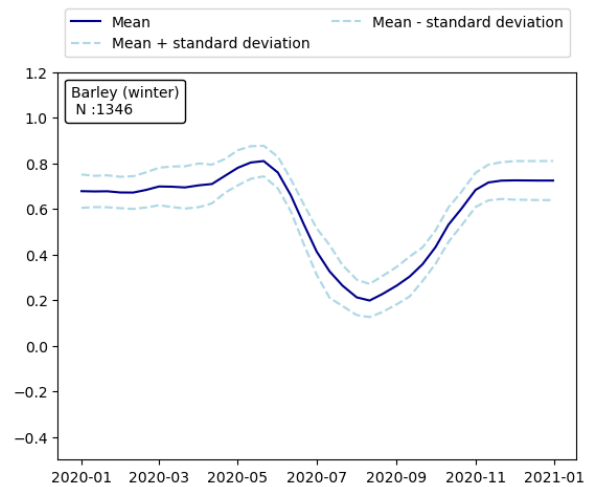
(a) NDVI time series of Christmas trees



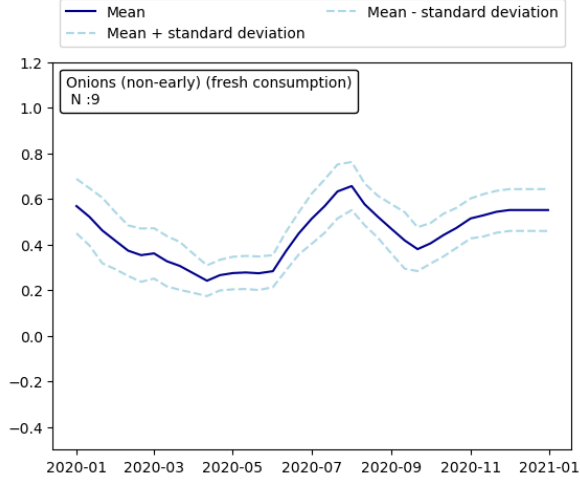
(b) NDVI time series of spring common wheat



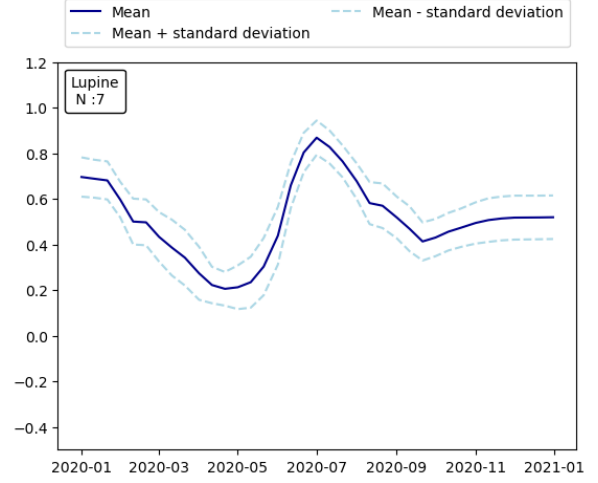
(c) NDVI time series of high-stemmed fruit crops



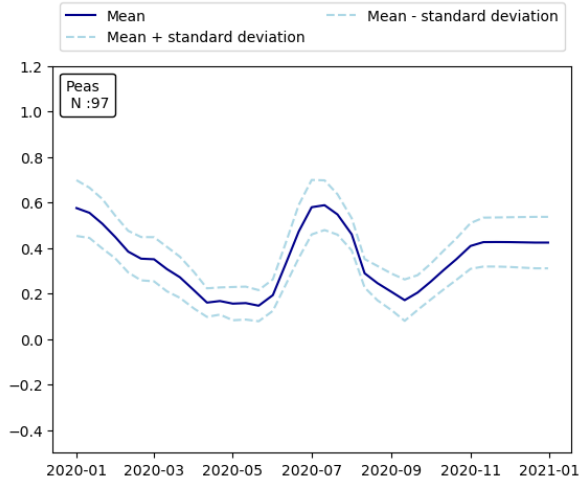
(d) NDVI time series of winter barley



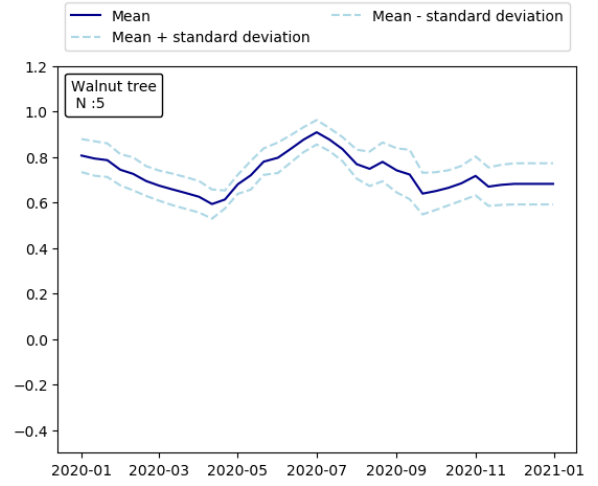
(a) NDVI time series of onions



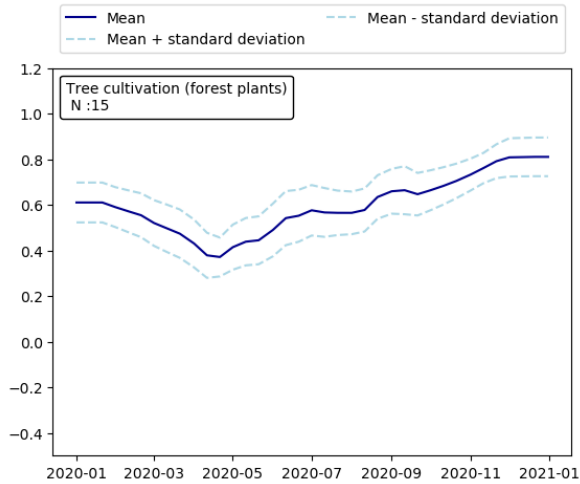
(b) NDVI time series of lupine



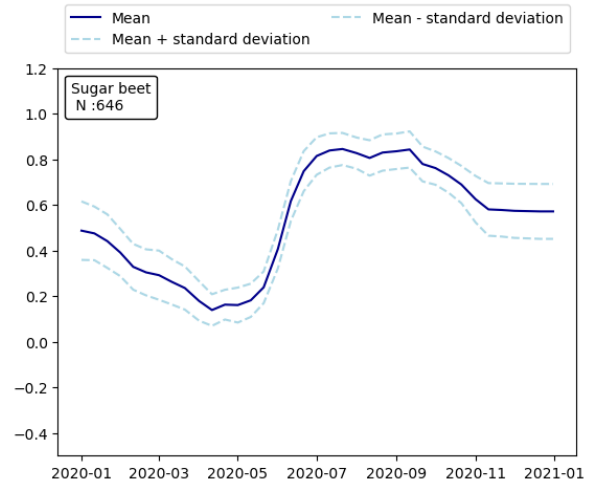
(c) NDVI time series of peas



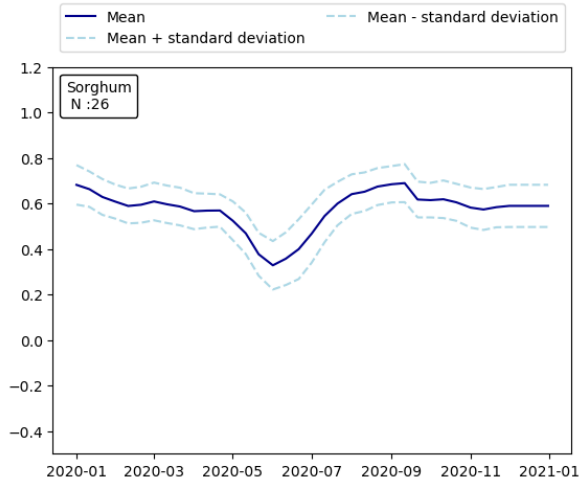
(d) NDVI time series of walnut trees



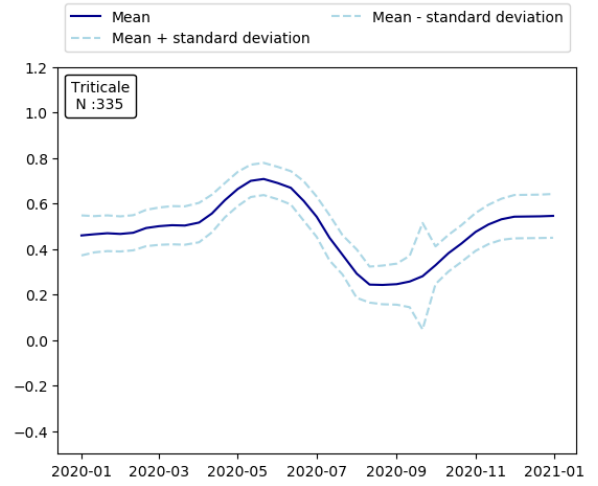
(e) NDVI time series of Tree cultivation (forest plants)



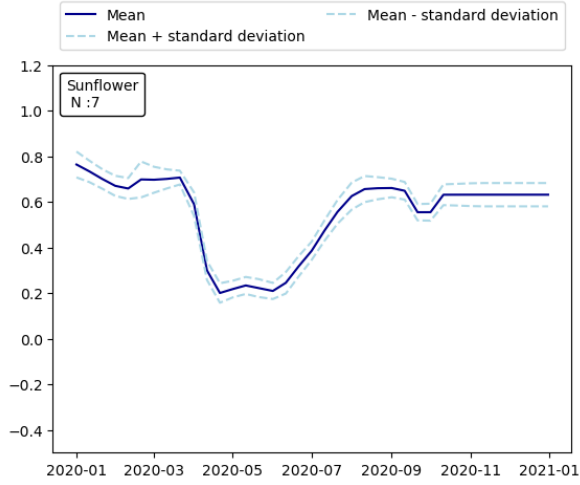
(f) NDVI time series of sugar beet



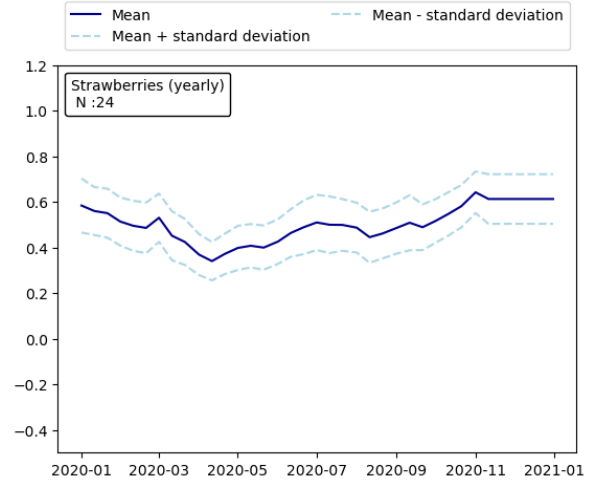
(a) NDVI time series of sorghum



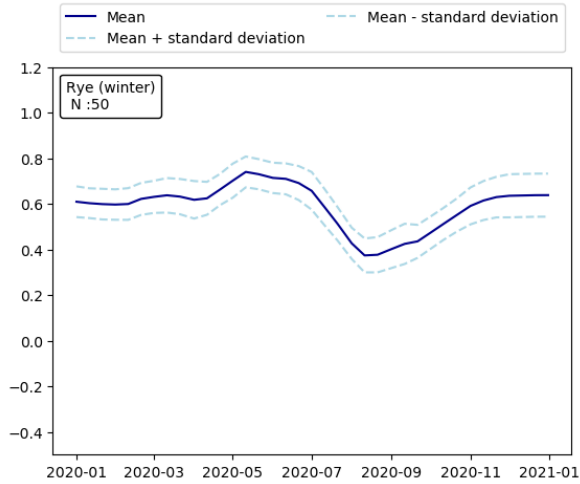
(b) NDVI time series of triticale



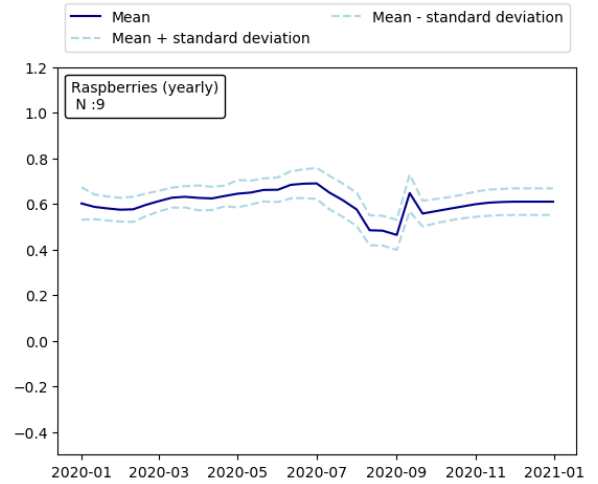
(c) NDVI time series of sunflowers



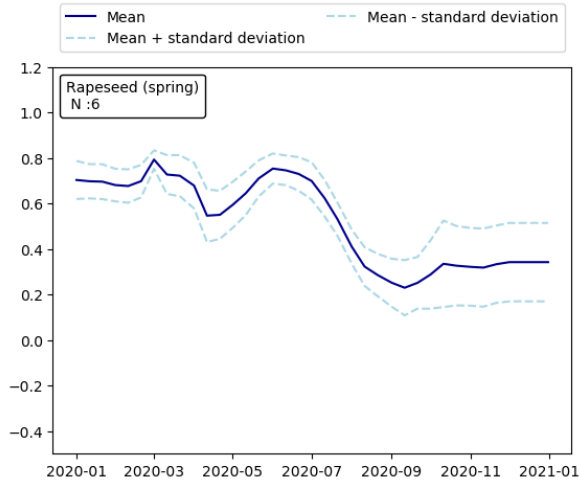
(d) NDVI time series of strawberries



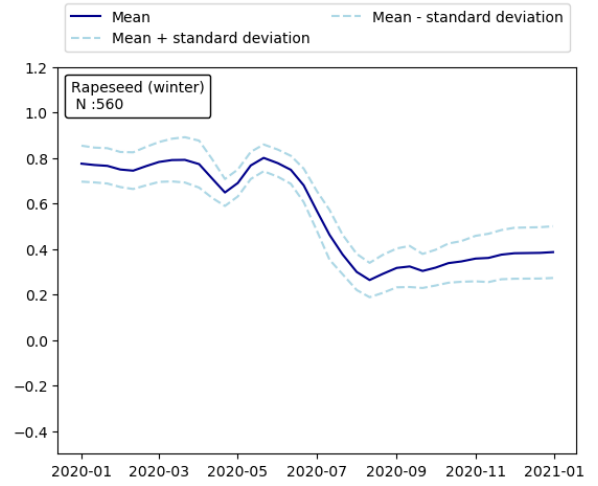
(e) NDVI time series of Rye



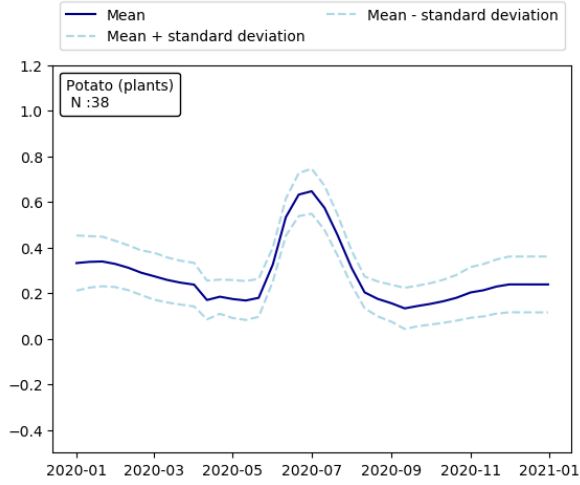
(f) NDVI time series of raspberries



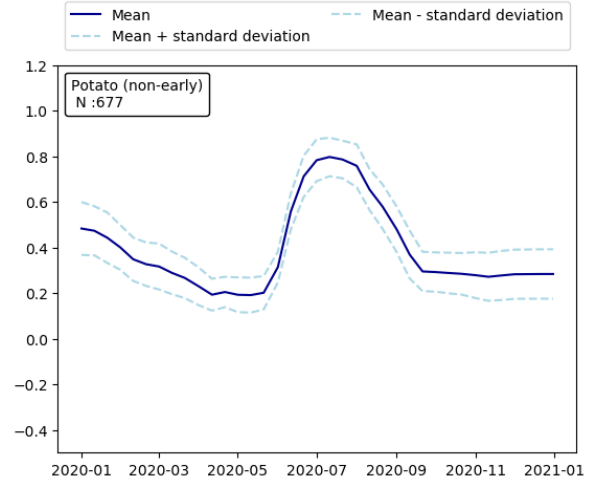
(a) NDVI time series of rapeseed (spring)



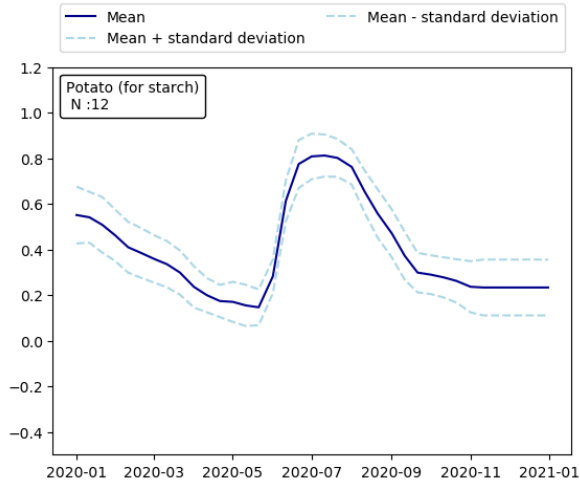
(b) NDVI time series of rapeseed (winter)



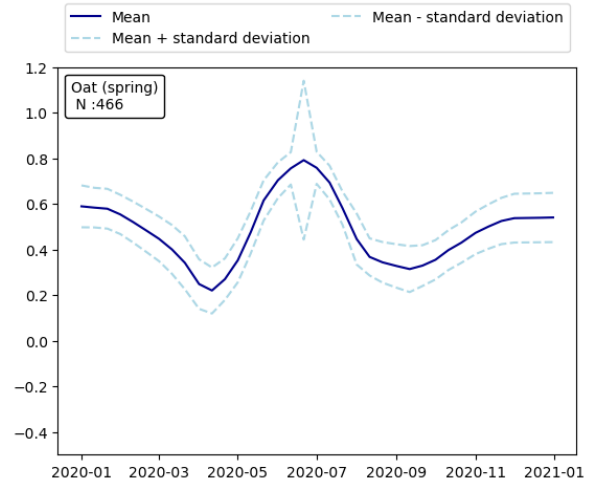
(c) NDVI time series of potato (plants)



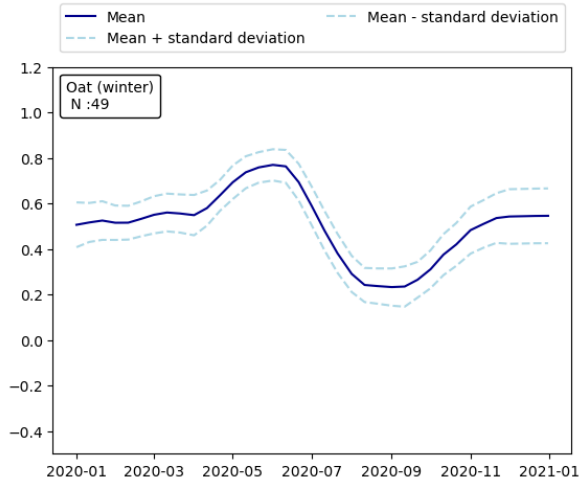
(d) NDVI time series of potato (non-early)



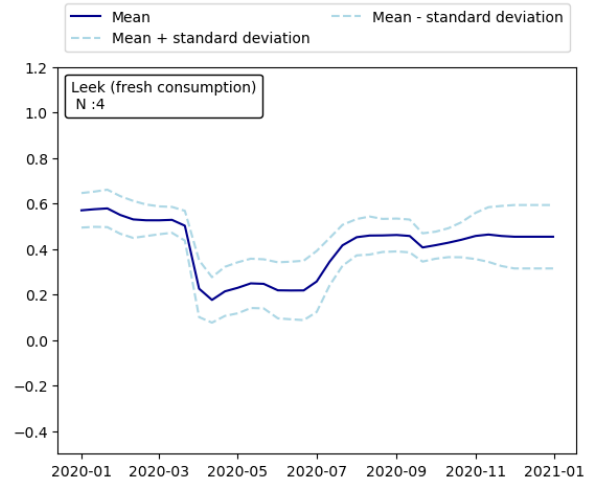
(e) NDVI time series of potato for starch



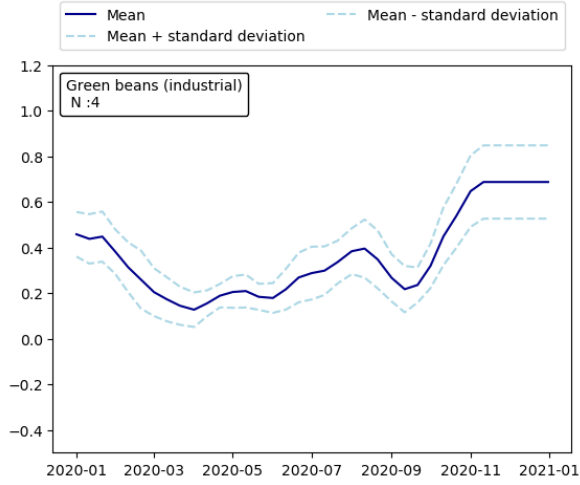
(f) NDVI time series of oat (spring)



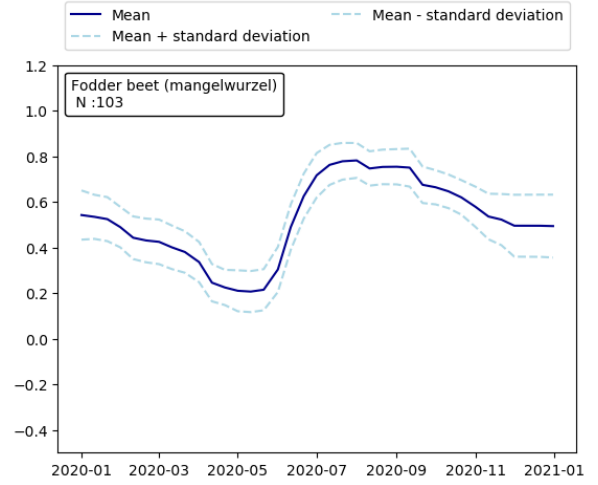
(a) NDVI time series of oat (winter)



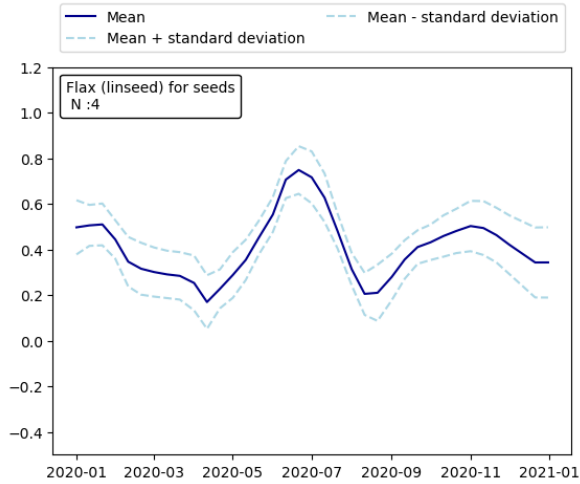
(b) NDVI time series of leek



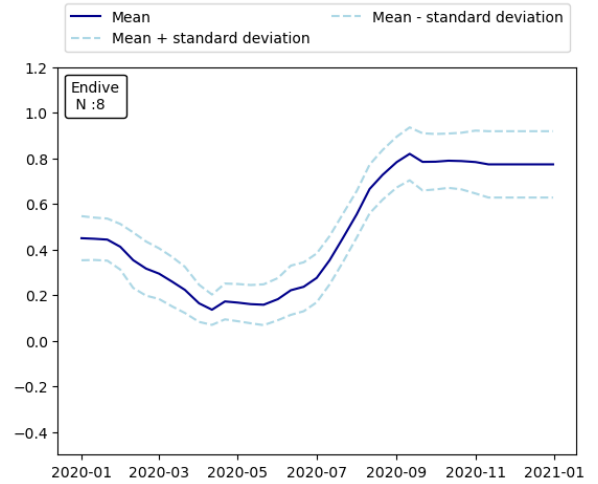
(c) NDVI time series of green beans



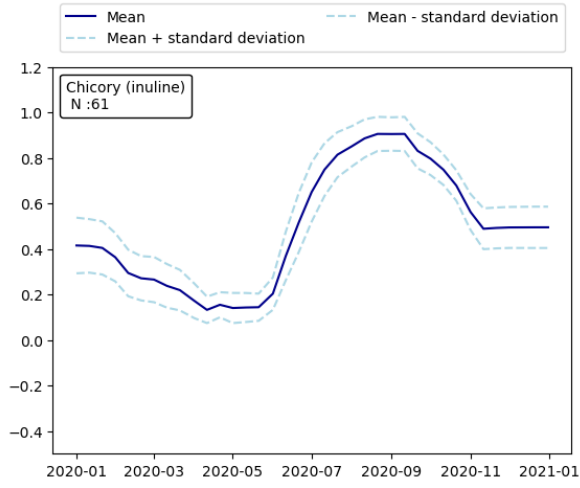
(d) NDVI time series of fodder beet



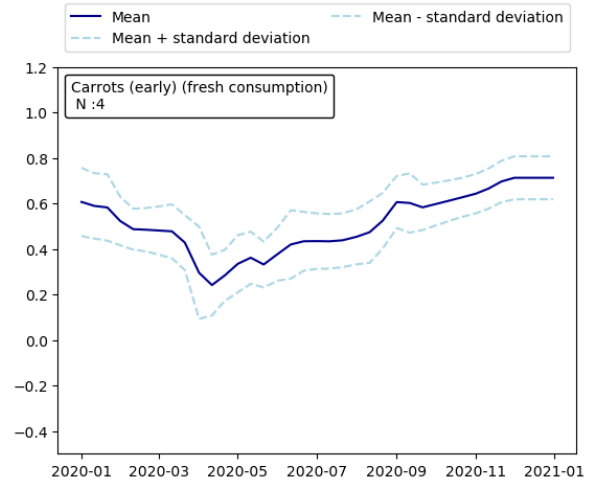
(e) NDVI time series of flax



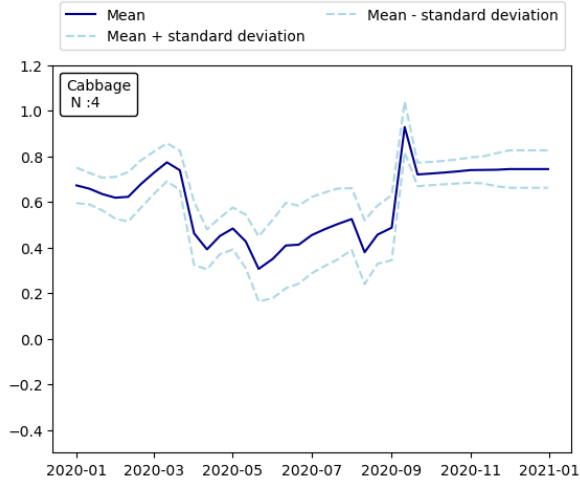
(f) NDVI time series of endives



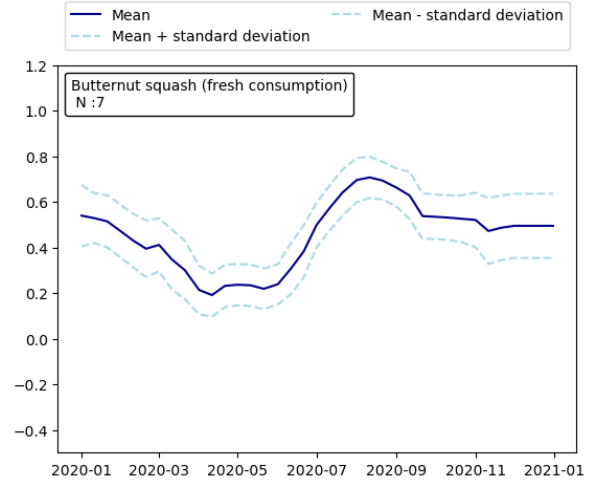
(a) NDVI time series of chicory



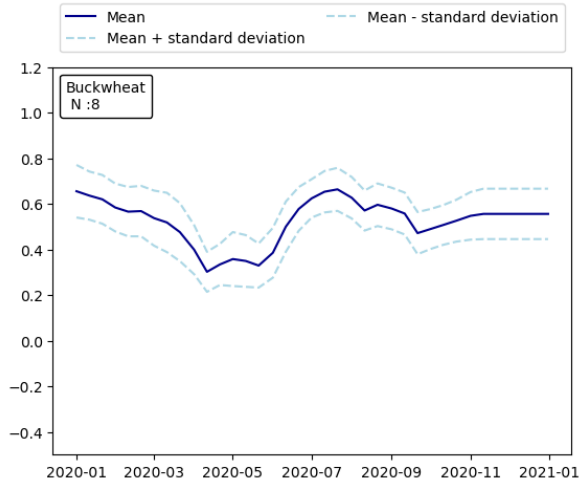
(b) NDVI time series of carrots (early)



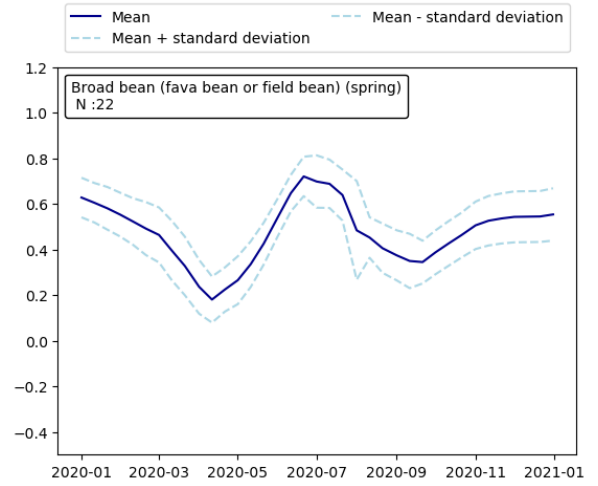
(c) NDVI time series of cabbage



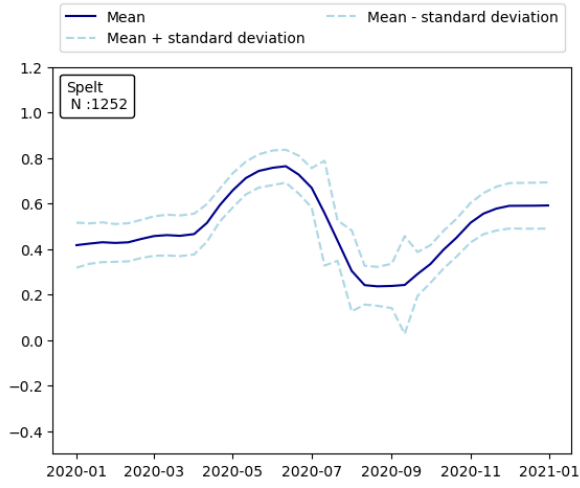
(d) NDVI time series of butternut squash



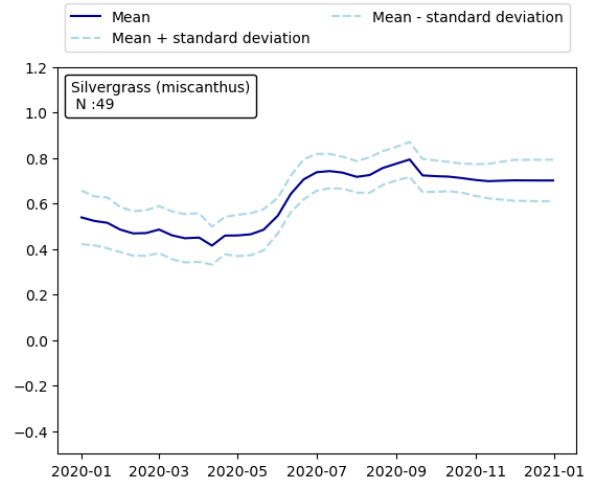
(e) NDVI time series of buckwheat



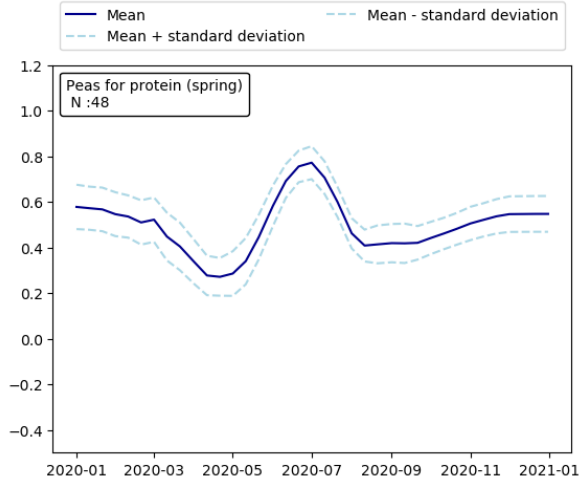
(f) NDVI time series of broad beans



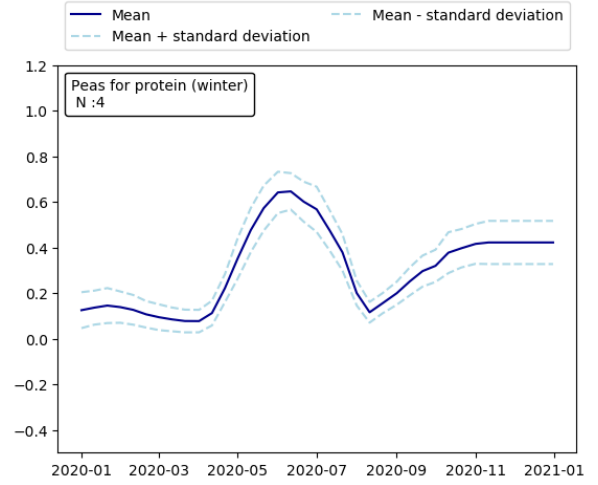
(a) NDVI time series of spelt



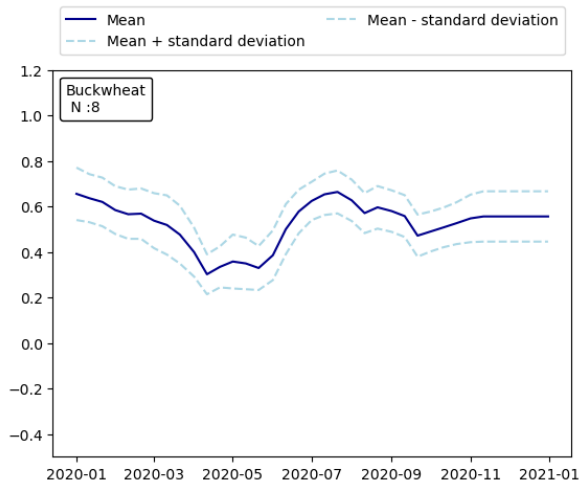
(b) NDVI time series of miscanthus



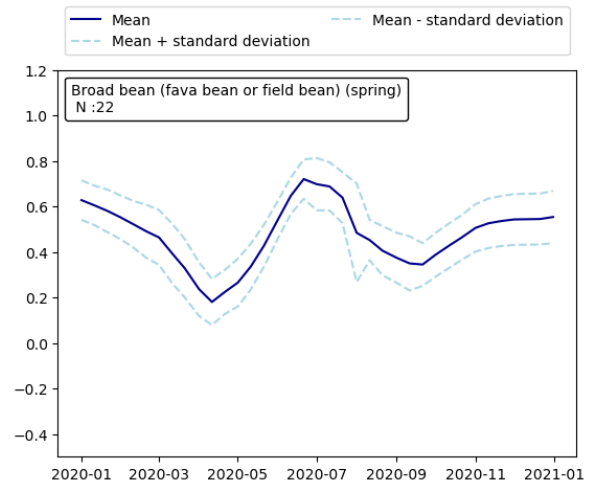
(c) NDVI time series of peas for protein (spring)



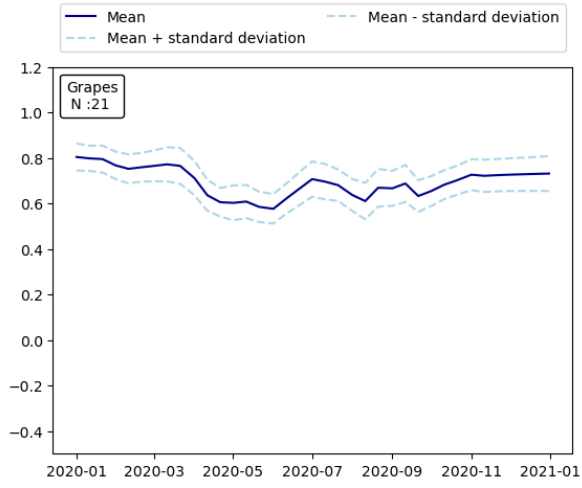
(d) NDVI time series of peas for protein (winter)



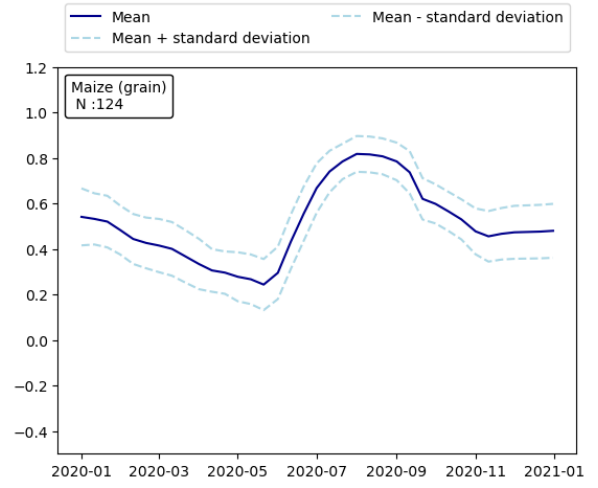
(e) NDVI time series of buckwheat



(f) NDVI time series of broad beans



(a) NDVI time series of vineyards



(b) NDVI time series of maize grown for grain

Appendix B : Results of the object-based RF crop group classification

Table B.1: Class specific metrics obtained for each feature composition scenario for the object-based RF classifications

Crop group		Annual crops			Perennial crops		
Features	Metric	Precision	Recall	f1-score	Precision	Recall	f1-score
NDVI time series		0.999	0.975	0.987	0.541	0.961	0.693
NDVI time series and crop group specific metrics		0.998	0.978	0.988	0.571	0.951	0.713
Crop group specific metrics		0.994	0.908	0.949	0.219	0.824	0.346
Matton and Valero metrics		0.994	0.979	0.987	0.55	0.814	0.656
Matton, Valero and crop group specific metrics		0.994	0.976	0.985	0.574	0.801	0.669
Number of validation samples (polygons)		3267			102		

Mapping perennial crops from Sentinel-2 image time series: a use case in Belgium

Boris Norgaard

In recent years, climate change and population growth have resulted in pressures on agricultural systems. In this context, crop monitoring and mapping systems have become increasingly needed as decision support tools for smallholders and farmers and as global land-use evolution assessment products.

So far, research aiming to develop such systems has mainly been focused on annual crops. Therefore, the objective of this master's thesis is to design a classification approach using Sentinel-2 image time series, aiming to discriminate perennial crops from annual crops, then to classify crop types among perennial crops.

In order to meet these objectives, metrics specific to crops' temporal systems were designed, and the performance of two classification models was assessed using multiple feature composition scenarios. The first classifier considered as the baseline is a random forest, and the second one is an extreme gradient boosting classifier.

Concerning the binary crop group classification, the designed metrics lead to the worst results when used alone as features but to the best results when coupled with the NDVI time series using the extreme gradient boosting classifier. The best crop type classification results were obtained using existing crop type-specific metrics.

Overall, the results were mixed but promising, and improvement opportunities, as well as ideas, were proposed as perspectives for the continuation of the development of the approach.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Faculté des bioingénieurs

Croix du Sud, 2bte L7.05.01, 1348 Louvain-La-Neuve, Belgique | www.uclouvain.be/agro