University of Zurich[UZH]

Department of Geography

Detecting intraspecific genetic variation in the Laegern temperate forest using airborne imaging spectroscopy time series.

GEO 610
Master's Thesis

Ewa Czyż
16-717-167

Supervised by
Carla Guillén-Escribà
Dr. Hendrik Wulf
Faculty representative
Prof. Dr. Michael Schaepman

30.07.2018
RSL, Department of Geography, University of Zurich

Abstract

A key contributing factor in biodiversity loss is the lowering of intraspecific genetic variation within species. Lowering genetic diversity of a population causes increased susceptibility for diseases, reduced evolutionary potential and lower fitness of the next generation. In light of of changing environmental conditions, habitat degradation and more frequent stochastic climatic events monitoring the genetic variation of populations is paramount. Assessing the genetic variation of organisms is commonly based on measuring its allele pool, which requires costly physical sampling with low spatial coverage. Alternatively, remotely sensed imaging spectroscopy data may hold the potential to distinguish genetic clusters, because higher allelic variation within the populations results in higher variation of phenotypic response that is reflected in the spectral information. Based on that, we investigate in this study the potential of remote sensing time series to reveal the intraspecific genetic variation under natural conditions.

We establish a direct connection between spectral data and intraspecific genetic variation of individual trees in a temperate forest system (Laegern, 47°28'N, 8°21'E), located in the Swiss midlands. Our dataset includes genetic and spectral information. Genetic data contain microsatellite analyses of 77 *F. sylvatica* individuals. Remote sensing data encompass annual acquisitions of the APEX (Airborne Prism Experiment) imaging spectrometer data with 2 m spatial resolution and 284 spectral bands from 2009 to 2016. We used Patial Least Squares Discriminant Analysis (PLS–DA) to classify genetic clusters based on their spectral information. To select the most relevant input data for the PLS-DA models for genetic cluster discrimination, we tested and compared various spectral bands subsets. The selection was based on specific spectral regions, Analysis of Variance (ANOVA), Principal Component Analysis (PCA), geometry of data acquisitions and moving window approaches. All analyses were based on two types of data transformation comparing reflectance and derivative-based data.

The different model performances reveal that spectral subsets from the near infrared (NIR) and shortwave infrared (SWIR) wavelength regions, rather than the full spectrum, show better consistency reflectance patterns differing among genetic clusters of sampled trees throughout the years. The spectral subsets selected based on ANOVA, PCA and the variation in reflectance based on geometry of the acquisition does not have a significant influence on the models performance. In addition, we reveal that the genetic cluster of sampled *F. sylvatica* individuals differs in phenols composition related to the spectral region around 1.7 µm wavelength.

We conclude that identifying the specific wavelength regions of the electromagnetic spectrum that are related to genotypic variation and using these data as input for PLS-DA models may enable intraspecific discrimination based on in-situ spectral information. Furthermore, the distinction of trees from different genetic clusters based on remote sensing time-series have a high potential to trace back genetic diversity loss in time and could contribute to the conservation of the species and habitats.

Contents

List of Figures

List of Tables

1. Introduction

Already in 80's of the 20th century it was recognized that the lowering of the genetic variation within species is a key factor in biodiversity loss (Wilson and Peter 1988). Both inter- and intraspecific genetic variety are crucial components influencing ecosystem structure and functioning (Hughes et al. 2008). Minimizing the genetic of a population increases its susceptibility of diseases (Schmid 1994), reduces the evolutional potential, and lowers fitness of the next generation (Ellstrand and Antonovics 1985). Therefore, maintaining highly genetically variable populations and their wider range of adaptive responses could be crucial in the context of changing environmental conditions (Szathmáry, Jordán, and Pál 2001; Gienapp et al. 2008) and more frequent stochastic climatic events (Hartmann et al. 2013). Considering the pace of global climate change and habitat degradation in comparison to evolutionary processes, the ability of organisms to adapt to changes is more important than creating a new variations (Frankham 2010). The short-term evolution and phenotypic plasticity potential is positively correlated with the allelic variation of the population (Gratani 2014). Therefore, the number and frequency of alleles changing in time and space is perceived as a suitable measurement of change in genetic diversity (Hoban et al. 2014).

Measurement of the allele pool, together with DNA and RNA sequencing-based techniques (Michael W Bruford et al. 2017; Yamasaki et al. 2017), are direct estimation of genetic composition, which is defined as one of the six Essential Biodiversity Variables (EBV) that aim to monitor world-wide biodiversity status (Henrique Miguel Pereira et al. 2013; BON 2015). These techniques require physical sampling (Davies et al. 2012), so are costly and time-consuming. Moreover, physical sampling results in individual measurements, which lack the spatial and temporal context relevant for biodiversity monitoring. Alternatively, continuous spatial data based on remote sensing technologies can overcome these limitations and are therefore getting more attention in the community (Rocchini et al. 2010; Turner 2014; O'Connor et al. 2015).

Both passive and active remote sensing technologies have been used to estimate

the functional and taxonomic diversity of different ecosystems. Based on vegetation traits gained from imaging spectroscopy and Light Detection and Ranging (LiDAR) data, Schneider et al. 2017 evaluated functional diversity in temperate forest. Taxonomic diversity assessment has been done by using both trait and spectra based techniques (Asner and R. E. Martin 2009; M. Martin et al. 1998), however more detailed evaluation of diversity than species richness has not yet been widely considered in remote sensing studies.

Similar to species, individuals with different genotype can express different morphological and physiological traits which shape their reflectance responses. However, these differences are less pronounced in lower taxonomic ranks than at species level (Hulshof and Swenson 2010). This means, that the recognition of individuals with different genotypes based only on spectral information gained in non-experimental conditions is limited and therefore not fairly used. However, growing importance of remote sensing in ecological assessments and high interdisciplinary of research groups show the potential link between genetic and spectral variations. For example, the studies of Madritch et al. (2014), which focus on above and below ground processes of trees, are showing the link of foliar reflectance with the genotype. Cavender-Bares et al. (2016) found a correlation between phylogenetic and spectral information. Finally, experts on genetics, where spectroscopic methods have been used for genotype recognition in laboratory conditions (Matsuda et al. 2012), suggest that there is a potential for remote sensing techniques to be used to examine genetic composition in natural habitats (Yamasaki et al. 2017).

Given the need for novel approaches to detect intraspecific genetic variation and the potential of using imaging spectroscopy data (Vihervaara et al. 2017; Geijzendorffer et al. 2016; Navarro et al. 2017), we attempt to find a direct connection between the spectral and genetic information of individual trees within temperate forest. We aim to demonstrate that genetic specific phenotypic responses are reproduced in reflectance information, and that use of spectral information could be an efficient tool for finding the genetic composition in non-experimental conditions. Furthermore, we aim to find evidence that the discrimination of the different genetic clusters, can be based on spectral information without prior knowledge on physiological and morphological characteristics. In our study, we define genetic cluster

(GC) as a group of individual trees with a similar genome, where the similarity was derived based on microsatellite analyses.

We base our study on the hypothesis that foliar reflectance changes on an annual and seasonal basis, which contrasts the expression of genotype specific phenotypic responses of individual trees that is maintained over years. To test this hypothesis, we used multi-annual airborne imaging spectroscopy data of a temperate forest in Switzerland, along with genetic information derived from microsatellite analyses of the individual trees. We combine both datasets using the Partial Least Square (PLS) method to investigate the explanatory power of distinct regions of the electromagnetic spectrum. By merging the interdisciplinary approaches, we underline the potential of imaging spectroscopy data in biodiversity assessments on a intra-specific genetic level.

2. Materials and methods

2.1 Study area

The study area covers 12.6 ha of semi-natural temperate mixed forest located in the Laegern mountain area on the north boundary of Swiss Plateau (47°28'N, 8°21'E) (Fig. 2.1). The climate of the area is characterized by the mean annual temperature of 7.4°C and the mean annual precipitation of 1000 mm (Etzold et al. 2011). The site has an altitudinal gradient of 620 and 810 m a.s.l and lies on a south-facing slope, with a gradient up to 60° steep (Guillén-Escribà et al. in press). According to the The United Nations Environment World Conservation Monitoring Centre (UNEP-WCMC), vegetation cover of the site is a Temperate Deciduous Broadleaf Forest, with 13 tree species consisting of 3 conifers and 10 angiosperms with common beach (*Fagus sylvatica*) being the dominant species. The age of trees spans between 53 – 185 years with the mean height of 30.6 m and the diameter up to 150 cm (Eugster et al. 2007), which creates a complex vertical structure of the mainly closed canopy (Schneider et al. 2017). The study area is a non-managed part of the forest and has been a forest ecosystem research site for the last four decades (Kloeti, Keller, and Guecheva 1989).

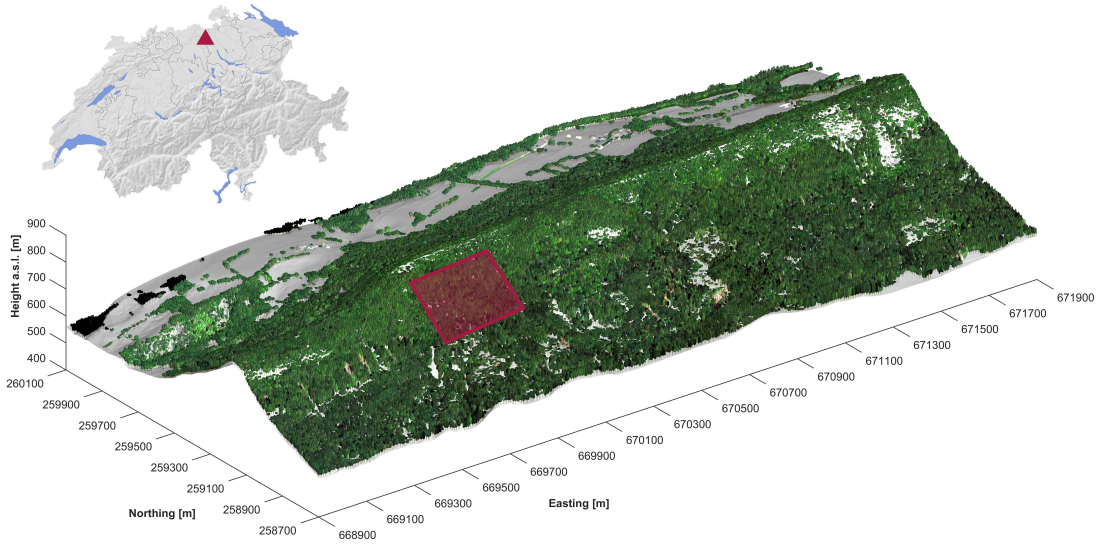


Figure 2.1: Location of the Laegern mountain in Switzerland (red triangle) highlighting the study site (red rectangle). Map projection indicates the Swiss National Grid CH1903 LV03.

2.2 Materials

2.2.1 Genetic data

The genetic data consists of a product of microsatellite analyses reflecting genome for 77 *F. sylvatica* individuals from the stand. The material for the analyses were leaf discs with diameter of 1.15 cm from each of tree sampled in September 2013 and georeferenced using a tachymeter in April 2013 (Leiterer et al. 2015). The DNA from the sampled material was extracted with Cetyl Trimethylammonium Bromide (CTAB) method following "Operationen- und Prozedurenschlüssel" (OPS) Diagnostic procedure. From the extracted DNA, five highly variable microsatellite loci (FS1-03, FS1-15, FS3-04, FS4-46, FCM5) (Pastorelli et al. 2003) were amplified using Polymerase Chain Reaction (PCR) technique. To assess the polymorphism at each microsatellite loci, capillary electrophoresis was performed on ABI-3720 sequencer. For determining the length of the analyzed microsatellites for each sampled tree, the GeneMapper software was used (Tab. A.1). We performed the microsatellite analysis with Bayesian methodology in TESS2 software, to find out the genetic clusters among sampled *F. sylvatica* individuals. Based on an average cross-entropy of admixture model, we determined five genetic clusters, to which each tree was assigned its membership probability (Fig. 2.2).

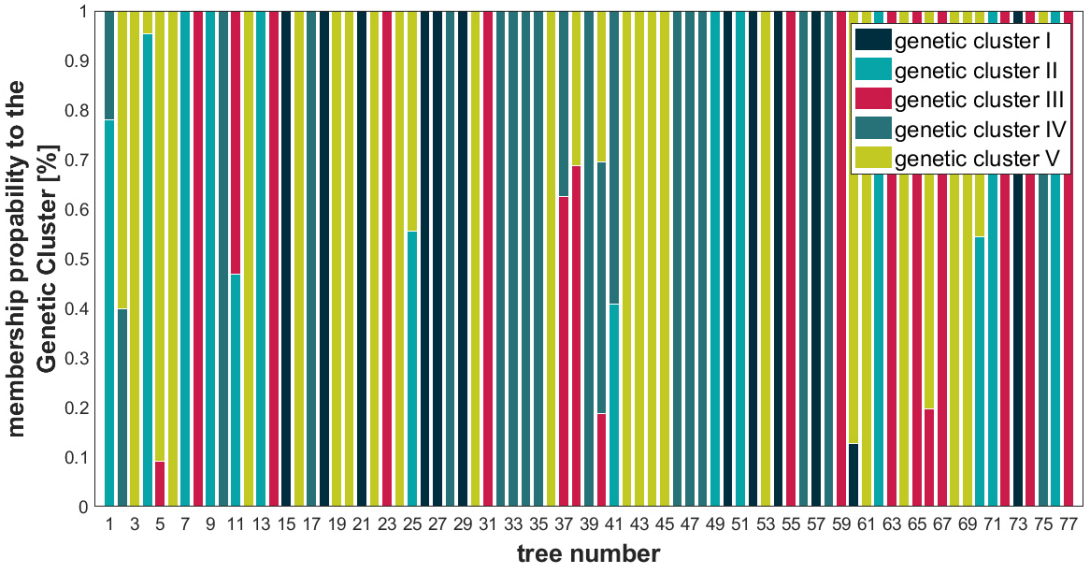


Figure 2.2: Results of the microsatellite analysis showing the membership probability of 77 sampled *F. sylvatica* individuals to the five detected genetic clusters.

2.2.2 Optical data

The spectral dataset contains seven acquisitions of the Airborne Prism Experiment (APEX) Airborne Imaging Spectrometer (AIS) (Schaepman et al. 2015) acquired between 2009 and 2016. To compare trees from similar development stages, images from the range of 300 - 700 Growing Degree Days (GDD) (corresponding to June - July) within each year were selected (Fig. 2.3, Tab. 2.1). The raw APEX data pre-processing chain of calibration and correction for spectral shift with smile effect were performed in the APEX Processing and Archiving Facility and ATCOR smile module, respectively (Hueni, Biesemans, et al. 2009; Hueni, Sterckx, et al. 2012; Richter, Schlapfer, and Muller 2011). Calibrated and corrected radiance data were atmospherically corrected to surface reflectances in ATCOR (Schläpfer and Richter 2002; Hueni, Damm, et al. 2017) resulting in imaging spectroscopy datasets of 284 spectral bands each in the range of 372-2540 nm with 2 m spatial resolution. Each dataset was vicariously calibrated with repeated ASD ground measurements of 30 ground targets to ensure consistent data quality standards and intercompatibility of these datasets.

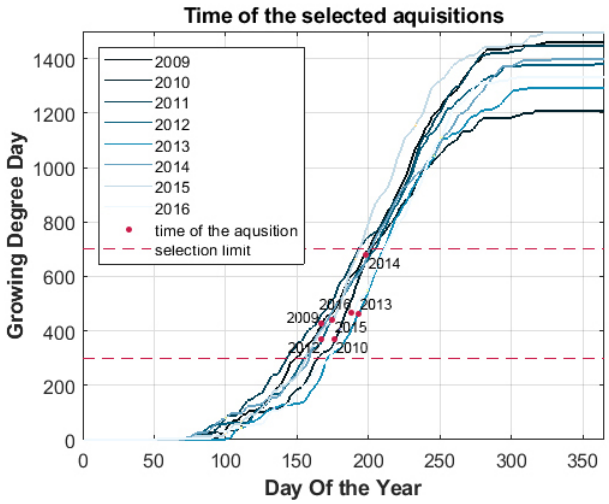


Figure 2.3: Growing Degree Days (GDD) within the years 2009-2016, with marked time of acquiring the images used in the studies. Additionaly, the range of 300-700 GDD, being a time limitation for the studies, is shown.

Table 2.1: Exact date, Day of the Year (DOY), Growing Degree Day (GDD) and solar zenith angle of the acquisition of the seven used hyperspectral Airborne Prism Experiment (APEX) Airborne Imaging Spectrometer (AIS) images.

Date	Day of the year	Growing Degree Day	Solar zenith angle
17.06.2009	168	429	30.4
26.06.2010	177	368	48.1
16.06.2012	168	372	27.1
12.07.2013	193	461	48.1
18.07.2014	199	674	26.5
24.06.2015	175	438	33.7
07.07.2016	189	461	na

Co-registration, Shadow exclusion, Pixel extraction

Multi-temporal images were geographically co-registered using Spectral Angle Mapper (SAM) method, where the object, being a Flux tower, with a spectral signature different from the vegetation was used as a reference point. Due to shadows creating high spatial heterogeneity, and therefore high noise, in a high spatial resolution data (Nagendra and Rocchini 2008; Stickler and Southworth 2008), pixels with total reflectance below 30% of mean total reflectance for whole image pixels were eliminated from the analysis. The high spatial resolution resulting in a high noise was also a reason to adopt object-based rather than pixel-based analyses (Karl and Maurer 2010). The object was defined as a tree crown for each sampled *F. sylvatica* individual derived from Light Detection and Ranging (LiDAR) measurements resulted in a CrownMap of the site (Guillén-Escribà et al. in press). The spectral signature for the object was a mean of unshaded pixels' reflectance from the crown (Fig. 2.4). The number of pixels averaged in a crown had an average of 17 with the standard deviation of 9 pixels.

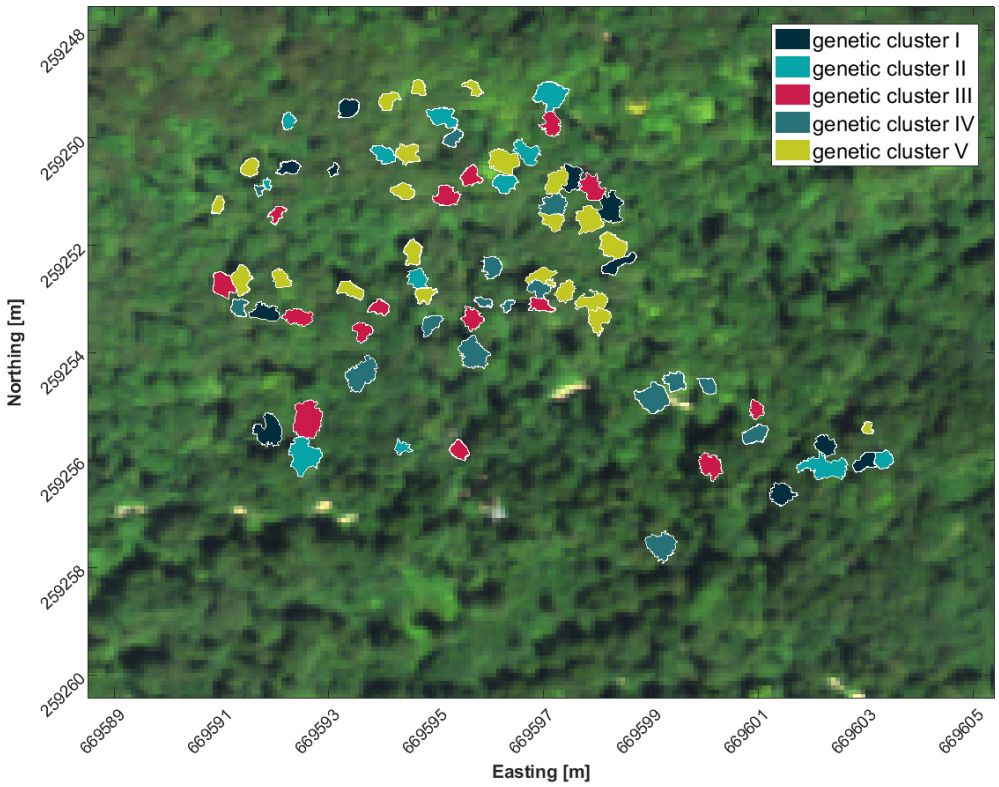


Figure 2.4: Crowns of the 77 *F. sylvatica* individuals sampled in the test site and its assignation to the five detected, based on the microsatellite analyses, genetic clusters. Map projection indicates the Swiss National Grid CH1903 LV03.

Signal transformation

Our analysis for this study is based on two datasets: (a) the averaged reflectance data per tree crown and (b) the 1st derivative derived from z-score of 'a'. The z-score of a mean tree crown reflectance was calculated separately for each spectral band with yearly reference with the formula $\frac{x_b - \mu_b}{\sigma_b}$, where b stands for spectral band (Fig. 2.5). The reasoning to include the 1st derivative in the analysis is to reduce the impact of multi-temporal variations in the reflectance magnitudes and simultaneously emphasize the relative differences between reflectance and absorption/transmittance influenced by structure, water content and organic compounds (Huesca et al. 2016).

The aforementioned co-registration, shadow elimination, spectrum per tree extraction and signal transformations were performed in MatLab R2017b.

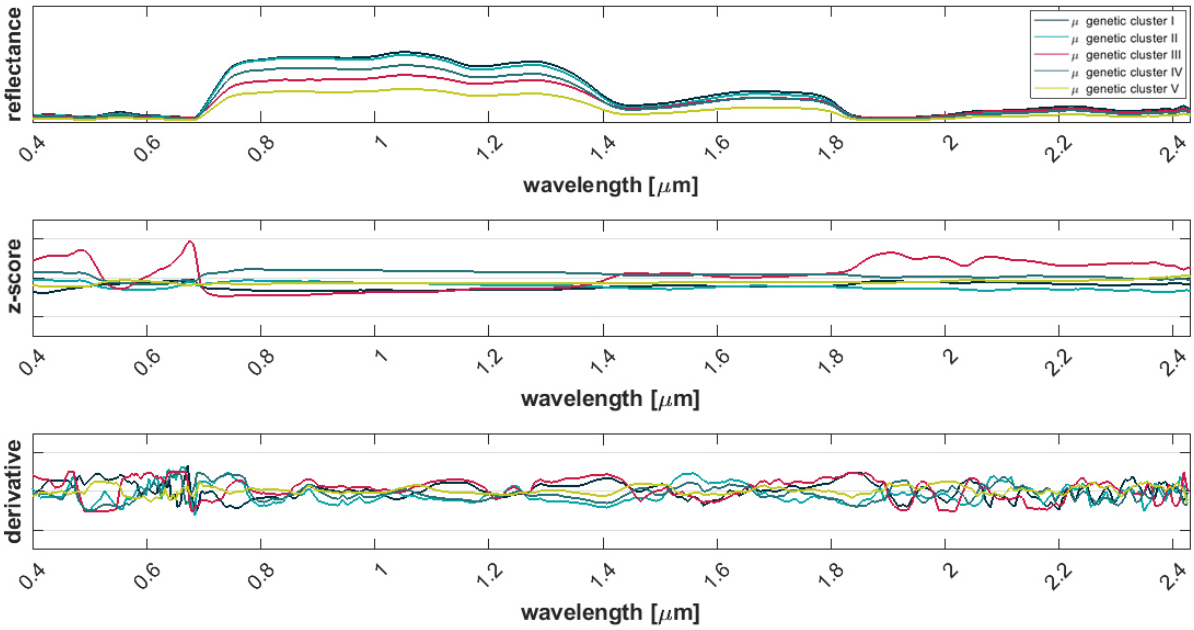


Figure 2.5: Average reflectance, z-score of the reflectance and derivative of the z-score for each detected genetic cluster. Reflectance data is averaged over all years (2009-2016) and all tree crowns within a genetic cluster.

2.3 Methods

2.3.1 Partial Least Square (PLS) analysis

The yearly consistency of the spectra and its location over the full spectrum were investigated by the Partial Least Square (PLS) method. The method is a basic tool in chemometrics (Wold, Sjöström, and Eriksson 2001) and is used in analyzing datasets with numerous bands where the collinearity and noise are problems in an analyses based on regression models (Wold, Ruhe, et al. 1984). Compared to the multiple linear regression (MLR), PLS models not only the X-variables, but also the Y-responses, based on the X-variable structure (Wold, Eriksson, et al. 2004). In our studies X-variables and Y-responses correspond to the spectral and genetic data, respectively. To investigate the influence of different wavelength regions of the electromagnetic spectrum and its potential in genetic clusters discrimination, we compared different spectral band subsets (see 2.3.2. Band subsets selection). Finally, the classification of the trees to the genetic clusters was carried out using the PLS models based Discriminant Analysis (PLS-DA). The classification was done for the spectral subset, which performed best. The workflow of our analyses is presented in the Figure 2.6.

To prepare the datasets for the PLS models, we first assigned the sampled trees to the five genetic clusters based on their maximum genetic membership probability derived from the microsatellite analyses (see 2.2.1. Genetic Data) (Fig. 2.4). To enable a balanced comparison between the genetic clusters, the minimum number of eight trees were selected from each genetic cluster to create the PLS models. This selection results in a 5x40 binary Y-responses matrix, where the columns represent genetic clusters and the rows represent individual trees with units standing for the assignment to the genetic cluster.

Secondly, we created the PLS models for each genetic cluster on a training set of 39 tree-crown spectra from the one year (X-predictors) and its corresponding genetic cluster 39 binary values (Y-responses). The excluded spectra from the spectral training set were fitted to the five created PLS models individually. The fit was done repeatedly for the tree spectrum of each year, and the mean of the predicted responses over all years was taken. In summary, this process was repeated iteratively

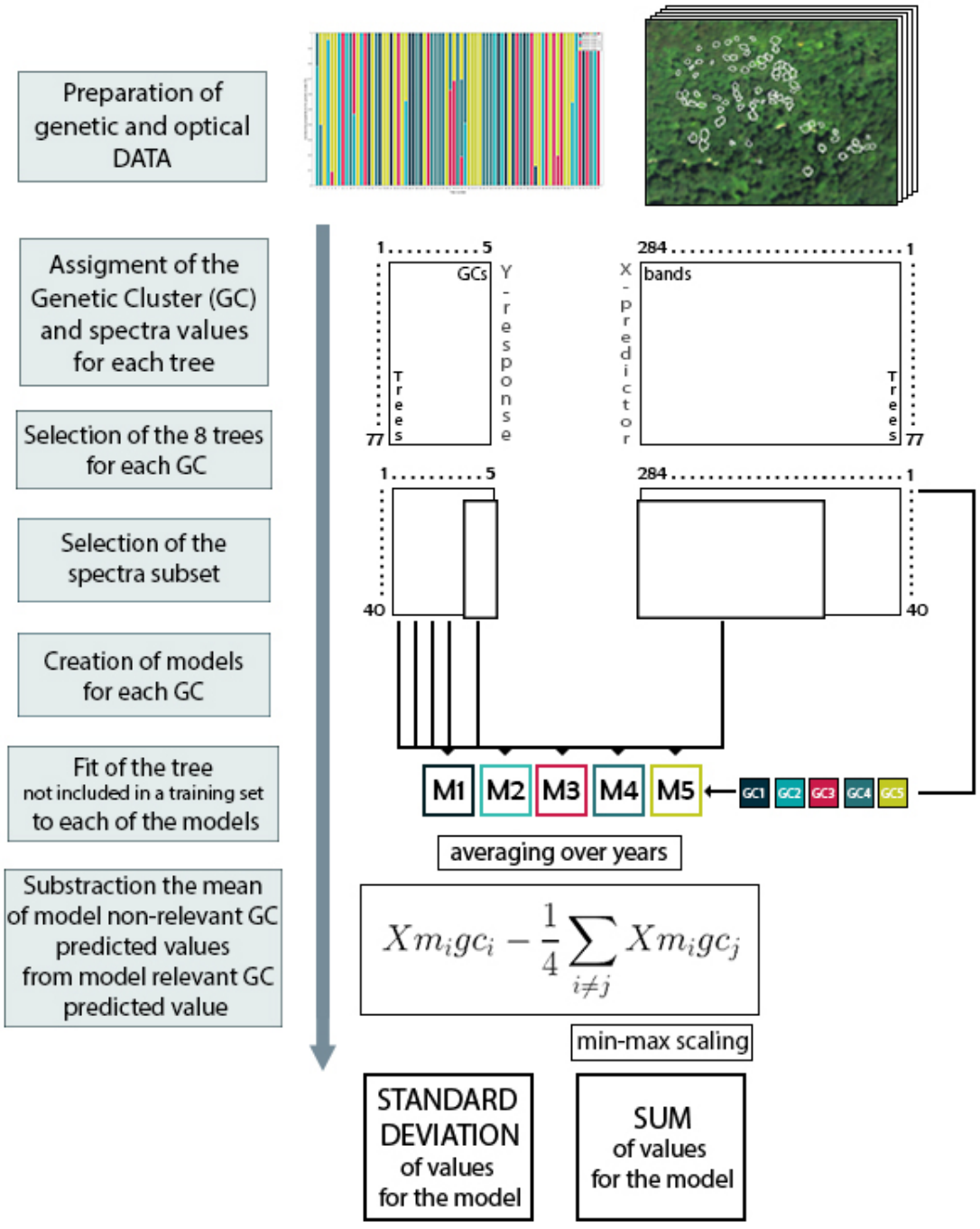


Figure 2.6: Workflow representing the individual steps to evaluate the discrimination power of various spectral subsets. These processing steps have been applied to each spectral subset separately. m - model, gc - genetic cluster.

for each of the 40 trees and for each of the seven years. The overall mean of the outcomes resulted in one predicted response for each genetic cluster to each of five PLS models. According to our hypothesis, given tree fits best to PLS model created for its particular genetic cluster, assuming a spectral consistency over the years within each genetic cluster. The mean predicted response of non-relevant genetic clusters to the genetic specific PLS models was extracted from the overall predicted value of the genetic cluster to its PLS model. That was done due to overfitting problem. The resulting values for each genetic clusters were min-max scaled. We sum up the scaled values for each PLS model and we refer to these sums as scores ranging from 1 to 5. The scaling is needed to compare the outcome of different bands subsets. Additionally, we calculated the standard deviation of genetic clusters values, derived from PLS models, for each spectral subset input. This was done for getting the level of correspondence of bands subsets to a respective genetic cluster.

2.3.2 Band subsets selection

Peerbhay, Mutanga, and Ismail (2013) and Cavender-Bares et al. (2016) demonstrated that spectral subsets result in better PLS-based model performances, than the analyses based on the full spectrum. Therefore, next to the full spectrum (284 spectral bands), we tested spectral subsets as a PLS model input and compared the model performances. The selection of different spectral bands for each subset was based on (a) specific spectral regions, (b) quality criteria and on a (c) moving window approach (Tab. 2.2). The selection was done separately for each of the spectra transformation (see 2.2.2 Signal transformation).

Table 2.2: Overview of spectral subsets used as inputs (X-predictors) for the Partial Least Square analyses. In addition, subset characteristics and number of bands within each subset are provided. This spectral sub-setting was applied to both, reflectance and derivative data. *In case of PCA based selection, the subsets consists of PCs.

	specific spectral regions			**quality criteria**			**moving window**
	VIS	**NIR**	**SWIR**	**PCA**	**ANOVA**	**geometry**	
	Range [nm]						
	400 - 700	700 - 1400	1400 - 2430	Principal Components (PCs) derived from the analyzed tree crowns	Selection based on the highest value of (F-statistic/p-value) ratio	Selection based on the lowest standard deviation for each band between acquisitions with various viewing positions	Number of neighboring bands moved over the full spectrum
number of bands*	59	94	130	5* and 50*	5 and 50	5 and 50	(280x)5 and (235x)50

The three regions of the electromagnetic spectrum were defined as a Visible (VIS: 400-700 nm), Near-infrared (NIR: 700-1400 nm) and Short-wavelength infrared (SWIR: 1400-2430 nm) frequency ranges. Based on this designation three subsets of bands were created with 59, 94 and 130 bands for VIS region, NIR region and SWIR region, respectively. The quality selected inputs consist of subsets of (i) Principal Components (PCs) derived from Principal Component Analysis (PCA), (ii) subsets of bands selected based on Analysis of Variance (ANOVA) and (iii) subsets of bands based on acquire geometry criteria. To compare different model performances, the number of spectral bands, dictated with variation explained by the PCs, was kept constant for all the created subsets.

Based on the variation explained by the PCs, 5 and 50 first PCs were chosen as a subset for the PLS analysis (Fig. A.1). The bands selected by an ANOVA criterion were derived based on the yearly reflectance mean of the GC. The bands with the highest F-statistic/p-value ratio were chosen for the analysis. The F value and p-value (Hogg and Ledolter 1987) were calculated with a MatLab function anova1 conducted on 7-years means of GCs (Fig. A.2). For the subsets of bands based on acquire geometry criteria, we selected the bands with the smallest reflectance variance between different acquisitions of different flight stripes from the same day of gaining (Fig. A.3). A moving window with the width of 5 and 50 bands over the whole spectrum with the interval of 1 band was performed and additional spectrum regions input subsets were created. For comparison reasons, we additionally randomly choose the 5 and 50 bands subsets from the whole spectrum. We have done the selections separately for each transformation (see 2.2.2. Signal transformation)

2.3.3 Partial Least Square Discrimination Analysis (PLS-DA)

In a last step, PLS - Discrimination Analysis (PLS-DA) based on the five generated PLS models was carried out. In this step, each of the 77 sampled trees were classified based on their maximum predicted response to one of the genetic clusters within the seven years. The PLS-DA was conducted for the best PLS model performance based on (a) specific spectral regions (b) quality criteria and (c) for all moving windows.

3. Results

3.1 Detection of GC in different spectral regions

There are noticeable differences in relative genetic cluster recognition in different regions of the spectra (Fig. 3.1). In general, analyses made on pure reflectance data provide higher scores as the 1st derivative of the reflectance data yields lower scores and thus has a lower explanatory power on average. In analyses made for reflectance and derivative, the use of the full spectrum (284 bands) did not perform the best. Based on reflectance data, the SWIR wavelength region performs the best and is outperforming the NIR region. This relation contrasts the results from the derivative data analysis, where the NIR outperforms the SWIR region. In both analysis, the VIS wavelength region has the smallest explanatory power to define the genetic clusters.

There is a pronounced variation of recognition power for different genetic cluster in different regions of the spectra and in different domains. The variations are stronger in the reflectance domain, where GC1 has the biggest yearly consistency in NIR region, GC2 in VIS and NIR regions, GC3 in VIS region, GC4 in NIR region and SWIR region and GC5 in SWIR region. There is also a distinct small yearly consistency of GC3 in NIR region in the reflectance domain. In the derivatives domain, the yearly consistency of the spectral signal for genetic clusters is similar in all spectral regions with the highest for GC4 and GC5 in NIR region.

The classification based on PLS-DA for the highest overall performance of SWIR region in the reflectance domain results in 25 correctly classified trees (accuracy: 32%, F1: 0.2984, kappa: 0.0680) (Fig. 3.2a). In comparison, the classification based on the full spectrum in a reflectance domain results in 13 trees correctly assigned to their genetic clusters (accuracy: 17%, F1: 0.1486, kappa: -0.0467) (Fig. 3.2b).

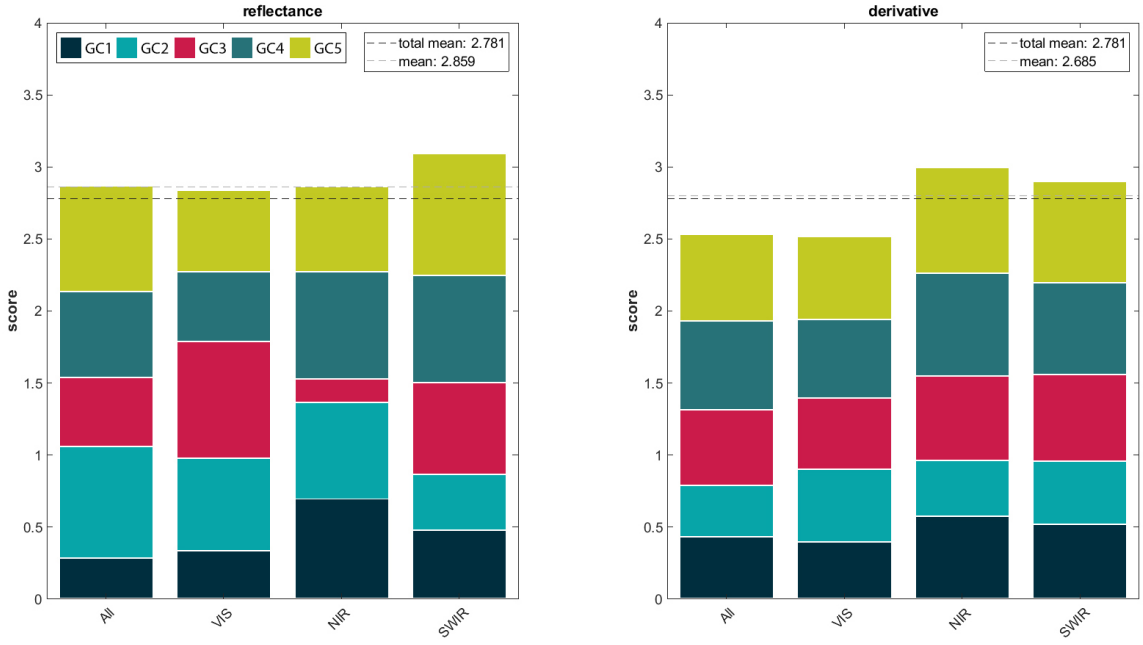
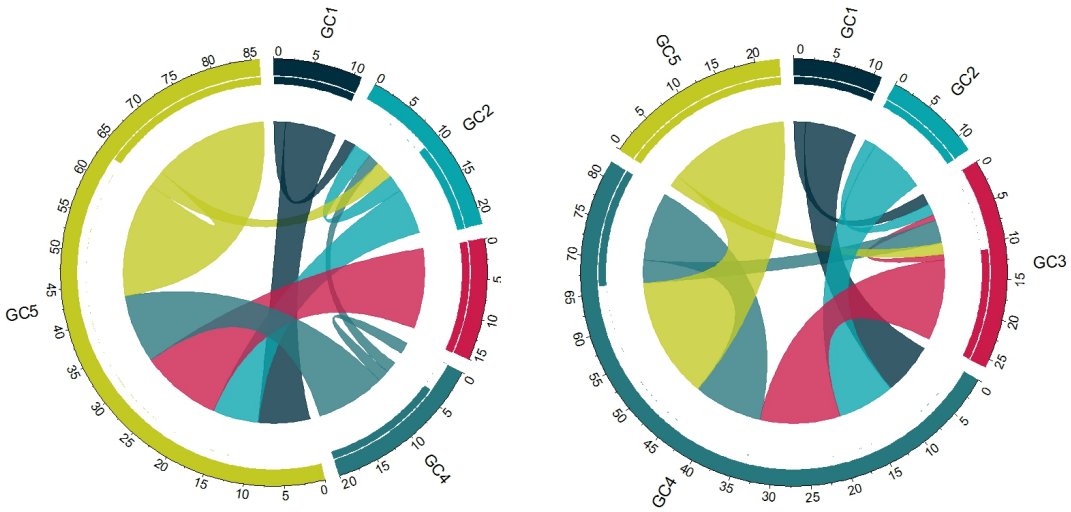


Figure 3.1: The sum of the genetic cluster (GC) scores derived from the PLS analyses conducted on the whole spectrum and specific spectral regions. Results indicate the average value for the time period 2009 to 2016 for the reflectance and derivatives of z-score of the reflectance (derivative)



(a) SWIR region
accuracy: 32%

(b) full spectrum
accuracy: 17%

Figure 3.2: Results of the PLS-DA classification of each genetic clusters based on reflectance for the SWIR spectral region (a) and on the full spectrum (b).

3.2 Detection of GC in chosen quality spectral subsets

The recognition of the GC based on spectral subsets selected based on specific quality criteria perform differently (Fig. 3.3). For both, reflectance and derivative data, use of Principal Components has the highest positive influence on the ability of GC recognition. Overall, the recognition is better achieved with the use of only 5 rather than 50 representative bands based on the same criterion. However, only in the derivative domain, randomly selected bands perform worse than all the quality selection-based subsets. In reflectance domain, subset based on ANOVA criterion performs the worst and geometry-based subsets perform similarly to the randomly selected subset.

In the reflectance domain, there are pronounced differences in a GC yearly consistency of the spectral subsets. Accordingly, GC2 reflectance varies the most in ANOVA and geometry criterion-based subsets of 50 and 5 bands, respectively, and GC3 has the biggest inter-annual variations in spectral subsets of 5 bands based on ANOVA criterion.

Overall, the subset of first 50 PCs derived from reflectance of sampled trees performs the best in the analysis. The PLS-DA based on this subset results in 20 out of 77 trees identified as a correct GC (accuracy: 26%, F1: 0.3603, kappa: 0.0932) (Fig. 3.4a), whereas GC classification with the use of 50 randomly selected bands from full spectra results in 13 trees correctly classified (accuracy: 17%, F1: 0.1744, kappa: -0.0366,)(Fig. 3.4b).

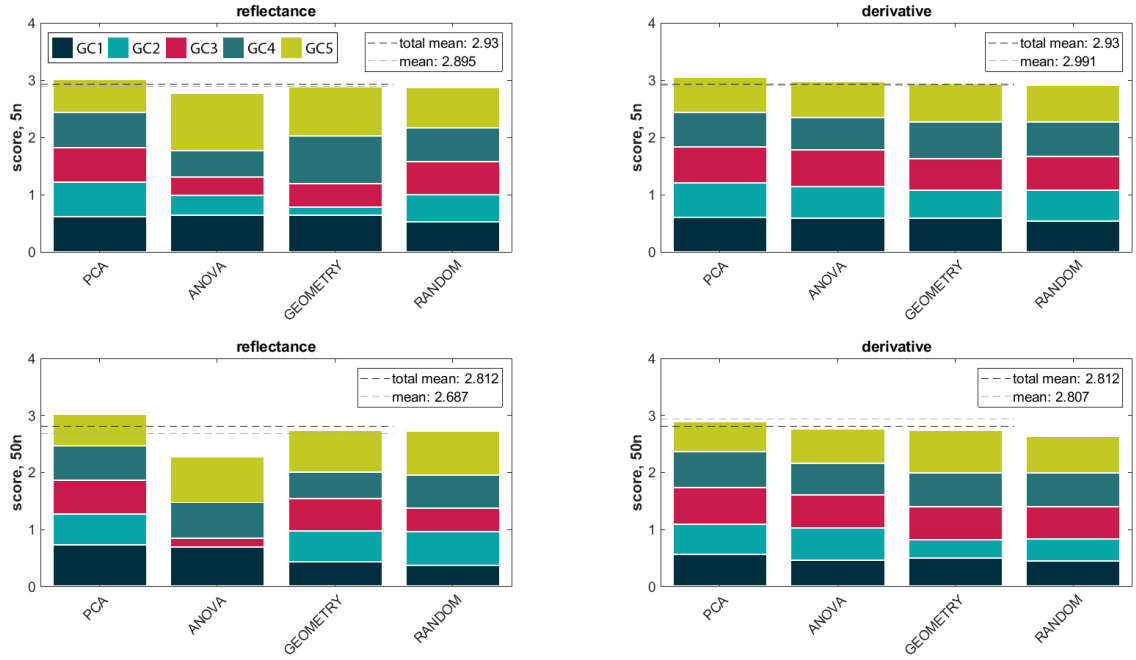
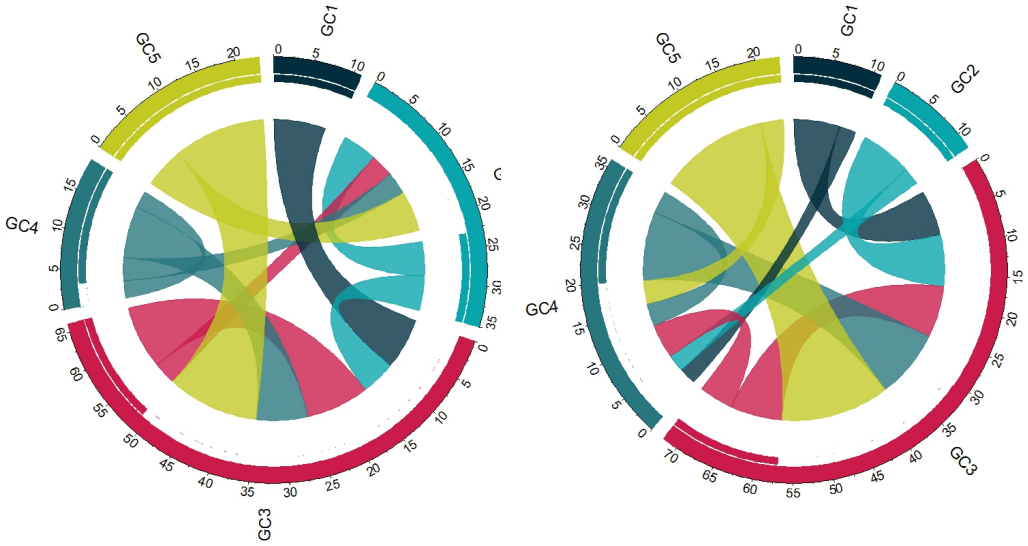


Figure 3.3: Scores of the PLS analysis for four spectral subset criteria (i.e. PCA, ANOVA, geometry, random). Results indicate different subset sizes (i.e. 5 and 50 bands) and different data type (reflectance and derivatives) for the time period 2009 to 2016.



(a) 50th first Principal Components accuracy: 26%

(b) 50 randomly chosen bands accuracy: 17%

Figure 3.4: Results of PLS-DA classification of the genetic clusters based on first 50th Principal Components, derived from reflectance of all sampled *F.sylvatica* individuals (a), and for 50 randomly chosen bands subset(b).

3.3 GC detection accuracy for different windows width over the full spectrum

The analyses made on the full spectrum for different window sizes unravel the regions of the spectrum where identification of a GC has higher accuracy (Fig. 3.5). Overall, using reflectance performs better than using derivatives, especially with use of windows size of 50 bands. Similarly, performance in derivative domain was better for window size of 50 bands. Overall standard deviations of GC scores are higher for wider windows and higher in reflectance domain compared to derivative domain. The highest accuracyof 35% (27 correctly classified trees), was observed in reflectance domain for 50 bands window centered at:

- 0.8324 µm (F1: 0.3057, kappa: 0.1156, stdv: 0.0048),
- 0.8439 µm (F1: 0.3043, kappa: 0.1047, stdv: 0.0046),
- 2.1373 µm (F1: 0.4004, kappa: 0.1151, stdv: 0.0116).

In derivative domain the maximum observed accuracy of 30% (23 correctly classified trees), was for 50 bands window width centered at:

- 0.8938 µm (F1: 0.3011, kappa: 0.1263, stdv: 0.0043).

The best performance of 5 bands window width in reflectance domain and out of the interpolated region of spectra, was equal to 25 correctly classified trees (accuracy: 32%) and was derived from the windows centered at:

- 0.6745 µm (F1: 0.3347, kappa: 0.0403, stdv: 0.0033),
- 1.6840 µm (F1: 0.3516, kappa: 0.0827, stdv: 0.0043),
- 1.7622 µm (F1: 0.3246, kappa: 0.0507, stdv: 0.0024),
- 2.2630 µm (F1: 0.3438, kappa: 0.0444, stdv: 0.0030).

In derivative domain the highest accuracy of 25% (19 correctly classified trees), with use of 5 bands window width being out of the interpolated region of the spectra, was achieved with the window centered at:

- 1.6398 µm (F1: 0.2011, kappa: 0.0327, stdv: 0.0017).

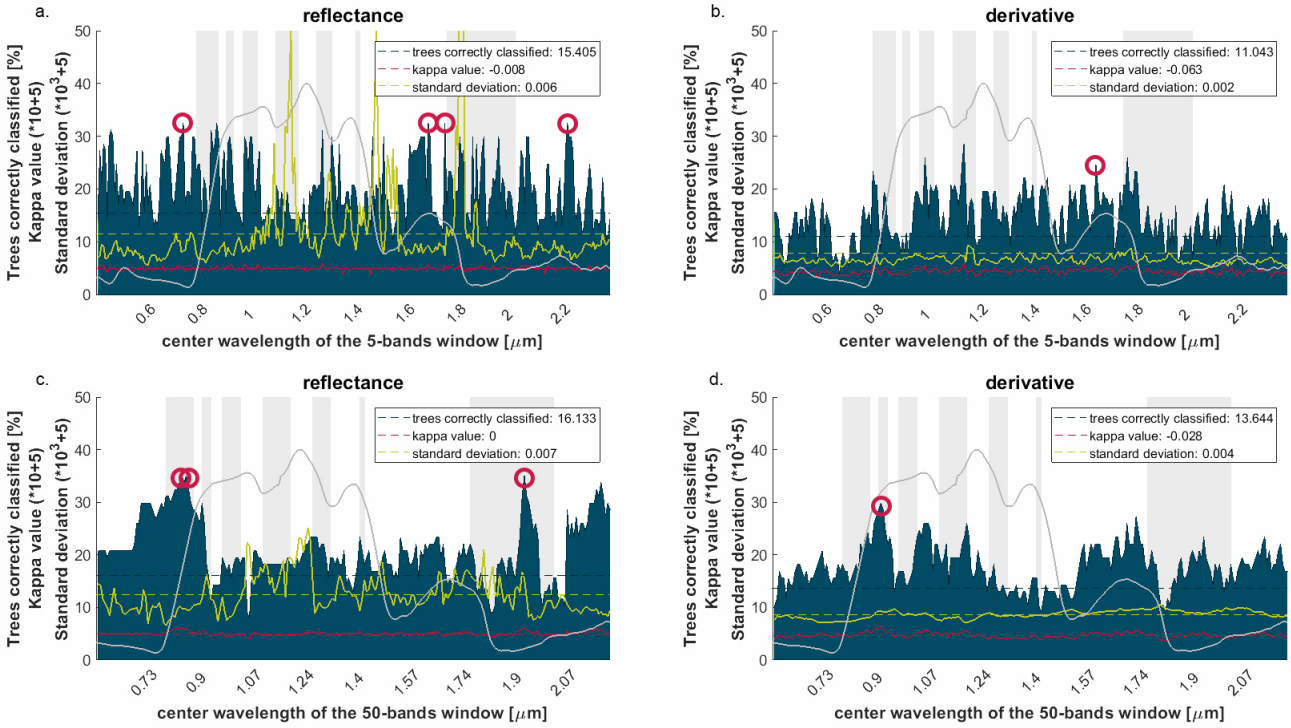


Figure 3.5: Assessment of PLS-DA genetic cluster classification of sampled *F. sylvatica* individuals, based on each moving window over the full spectrum. The presentation of results for 5 (a, b) and 50 (c, d) bands window widths, and for reflectance (a, c) and derivatives of z-score of the reflectance (derivative)(b, d) mean values from 2009-2016. Additionally, the kappa value of the classification and the standard deviation of different GC recognition is shown. The red circles show the highest accuracy results the cases. The light grey colour bars the interpolated wavelengths.

4. Discussion

4.1 Detection of GC in different spectral regions

This study shows the ability of detecting intraspecific genetic variation of *F. sylvatica* based on a multi-annual spectral information. Full spectrum analyses results in a 17% accuracy for defined genetic cluster classification. The relatively poor classification confirms that the spectrum of each sampled tree is strongly affected by phenological and environmental variations. Therefore, the different regions of the spectrum have been used in the analyses and the spectral regions reflecting genetic specific phenotype responses were identified.

In reflectance domain, analyses based on a band subset covering SWIR region results in 50% improvement of genetic cluster identification. Since this region of the spectrum is influenced mostly by the water content of leaves (Tucker 1980; H. W. Gausman and Weidner 1985), this result suggests that there is intraspecific variation in this property for *F. sylvatica*. The outcome is consistent with physiological and genetic studies, where evidence of the origin of the *F. sylvatica* influencing its water management (Peuke et al. 2002) and genetic structure (Demesure, Comps, and Petit 1996) was found. The reflectance in the SWIR region is also related to the internal structure of the leaf and its dry matter concentration (Ceccato et al. 2001). However, these two leaf properties are most pronounced in the NIR region (H. Gausman et al. 1970), which according to our study is relatively the best region of a derivative transformed spectrum for identifying different genetic clusters. This suggests that the internal structure of the leaf together with the dry matter concentration could be specific and yearly consistent for genetic clusters of *F. sylvatica*. This dry matter concentration and internal structure have an impact on mesophyll conductance (Muir et al. 2014) which affects the photosynthetic performance (Flexas, Ribas-Carbo, et al. 2008; Flexas, Barbour, et al. 2012). This performance is related to the physiological adaptation of the organism, therefore is highly susceptible for a natural selection (Arntz and Delph 2001). That could support the variation of the inner leaf architecture between detected genetic clusters.

Difference between spectral performance in reflectance and derivative domain could be argued with atmospherically affection of electromagnetic waves. The use of derivative transformed spectrum reduces the illumination intensity variations and background of the signal (Tsai and Philpot 1998; Otto 2016). Since shorter wavelengths are more affected by the atmosphere than longer wavelengths, the consistency of features through time from NIR spectral region are more pronounced in derivative transformed than in a pure reflectance domain.

4.2 Detection of GC in chosen quality spectral subsets

The 26% accuracy of genetic clusters recognition by using analyses based on PCs derived from a full spectrum shows that the PCA reduction of hyperspectral collinearity is improving the genetic cluster recognition. However, the PCA method is reducing the collinearity of spectra based only on the maximal variances (Maitra and Yan 2008). The spectrum features that reflect the phenotypic response of a genome are expected to be maintained for one generation of *F. sylvatica,* therefore stable in their reflectance at the same development stages. Those regions are not necessarily the most variable ones, and they could be therefore omitted through the PC transformations in favor of more variable features which are not reflecting the intraspecific genetic variation of *F. sylvatica*. Even though the PC transformation of hyperspectral images was commonly used for tree species classification (Lee et al. 2016; Carleer and Wolff 2004), our study shows that the transformation is not correctly preserving the information needed for genetic cluster identification. Moreover, PCs are scene dependent(Ashbindu Singh and Harrison 1985), therefore they are not appropriate in multi- temporal data analyses that were used in our studies.

Poor performance of analyses based on subsets derived from ANOVA suggest that this statistical tool applied on an overall yearly reflectance mean of genetic clusters is not sensitive enough for finding the specific spectral features. Nevertheless, it should be noticed that the recognition of the spectral features based on this criterion, was better achieved with the use of derivatives, what could be related to the more pronouncing absorption features (Huesca et al. 2016) in this domain.

The lack of difference of performance between geometrically and randomly selected subsets in reflectance-based analyses suggests that the differences between

the view angles of used acquisitions are not an important factor in identifying the genetic clusters. However, the use of bands varying the less in between acquisitions caused relative improvement in the derivative domain (especially pronounced in subsets of 50 bands). The derivative-based analyses are utilized in analytical chemistry in cases where elimination of background of the signal is needed (Otto 2016). There is also evidence that the derivative-based indices are estimating the amount of a chemical compound in a plant more accurately, when the background reflectance play a big role (Kochubey and Kazantsev 2012). Therefore, it could be stated that identifying the variabilities in reflectance caused by different view angles is more accurate in derivative domain. The neglection of the wavelengths that are highly related with viewing angle allow us to compare spectral features that vary due to genetic specific phenotypic reaction and that are not influenced by geometry of the acquisition.

4.3 GC detection accuracy for different windows width over the full spectrum

The overall better genetic cluster recognition by reflectance suggests that use of untransformed spectrum is an appropriate approach for detecting intraspecific genetic variation of *F. sylvatica*. However, relatively high standard deviation of accuracy for each genetic cluster can be observed. This accuracy could be affected by the group size of each defined genetic cluster, meaning that the recognition in this domain might be more cluster-specific. On the other side, derivativ-based recognition of genetic clusters is in general more universal – the recognition power of the spectral regions is similar for each genetic cluster. Since intraspecific phenotypic responses variation could be very minor, precise change detection is needed for genetic cluster identification. Therefore, the shape of the spectrum directly linked with the absorption features rather than its absolute amplitude, has greater potential for distinguishing detected genetic clusters of *F. sylvatica*.

The method developed in this study could be also utilized for recognizing which region of the spectra is related to genetic cluster specific phenotypic responses maintained during the years. Even though there are noticeable high informative regions

of spectra with a width of around 5 nm the relatively small kappa value shows that the information from wider spectrum range is recommended for genetic cluster recognition. However, the accuracy values for window size of 50 nm could be highly influenced by interpolated regions of the spectra. The process of interpolation leads to the reduction of intra-annual variation in the spectra, therefore the consistency of the spectral regions connected with genome specific response could be overestimated. Based on that, rather narrow than wide spectral regions should be investigated. The high accuracy together with the relative small standard deviation was derived from analyses based on wavelengths ranging 1.630 - 1.667 µm in derivative domain. That suggest that this region could be the most conspicuous in identification of detected genetic clusters of *F. sylvatica*. The spectral feature located at this wavelengths is caused by C-H bond absorption of phenolic compounds (Kokaly and Skidmore 2015). These compounds are related with many physiological reaction of plants including: protection against UV radiation (Close and McArthur 2002), microbial (Scalbert 1991), fungal (Telles, Kupski, and Furlong 2017) and herbivorous (War et al. 2012) defense as well as pollution (Pasqualini et al. 2003) and climatic responses (Stark, 2015). The variation of phenols between species is already used for trees taxa identification (Asner, R. E. Martin, and Suhaili 2012). Moreover, it was also shown that the amount of phenols varies on the intraspecific genetic basis (J. A. Pereira et al. 2007), which could support the outcomes of this study suggesting that the spectral feature reflecting the phenols composition could be specific for each detected genetic cluster of *F. sylvatica*.

4.4 Outlook and limitations

Despite the relatively poor classification accuracy that we achieved (30-35%), there is evidence that the electromagnetic spectrum of the canopy surface contains information on intraspecific genetic diversity. This conclusion is supported by the analyses of different wavelength regions with varying performances in the genetic clusters classification. The improved performance of the spectral subset-based approach was consistent with Cavender-Bares et al. (2016), where the the bands with highest Variable Importance in Projection (VIP) scores were selected for the analyses. This study also demonstrated the use of spectrum-based approach and time

series analyses can unravel the regions of spectra reflecting genotype-specific phenotypic reactions that are maintained over a seven year period. Based on this finding, we find that the relative genetic classification of temperate trees could be achieved without prior knowledge about intraspecific variations.

It should be emphasized, that the approach we used takes advantage of high spectral, spatial and temporal resolution datasets available. Based on other datasets lacking the spectral or spatial detail information on genetic cluster-specific responses of individual trees could be lost. Lower spectral resolution could be not sufficient to recognise genotype specific phenotypic responses, whereas RS-based studies of genetic composition where spatial resolution is not high enough to recognize individual tree, could be done only in case where the genetic of the area is highly homogenous, like in the studies of Madritch et al. (2014). It should also be noted, that our results are limited to species and location specific information about the genetic intraspecific variation. However, our methodology, in particular the derivative-based approach and PLS analysis, can be applied to similar studies.

5. Conclusions

This study highlights the potential for multi-temporal imaging spectroscopy data to detect intra-specific genetic variations of trees from a temperate forest. We investigated the use of derivative based analyses and PLS based methods to overcome the discrimination challenges caused by multi-factorial influence on the spectral canopy signature acquired under natural conditions. Moreover, our methodology successfully detected the spectral regions most indicative for the genotype recognition. Therefore, we propose that the discrimination of intra-specific genetic variations could be possible without prior knowledge about the genotype specific phenotypic responses.

In this study we create a baseline for further studies focusing on genetic diversity with use of remote sensing techniques. We demonstrated that accessing the intra-specific genetic diversity should be conducted from spectral subsets, rather than full spectrum. Moreover, we revealed that limitations related to multi-temporal data analyses could be overcome with a data derivative transformation approach.

Based on our results we could assume that the analyses where the genetic resolution is low enough for remote sensing ability and high enough for practical purposes, are promising tool for tracing the landscape history of genetic variation (Rocchini et al. 2010) in a direct, efficient and globally consistent way.

6. Acknowledgements

Firstly, I would like to thank my supervisors Carla and Hendrik. They have always been open for discussion, full of meaningful suggestions, patient with corrections, trustfully open-minded and ready to digest new ideas with the most appropriate attitude. Additionally, they always have been there whenever I ran into any trouble spot during the work. Furthermore, I would like to express my gratitude to Michael who entrust me the topic and keep motivated. Additionally, I like to thank Andrew Tedder, Quingyang, Eri Yamasaki and Bernhard Schmidt for preparing and helping with understanding the genetic part of the research.
Great thanks also go to Fabian - for constant readiness of helping and fruitful positive discussions, Chengxiu - for putting a new light on any doubts, Gillian - for questioning to improve the outcomes, officemates - for sharing a cup of a tea, and to all the colleagues from RSL, for simply being around.
Lastly, I would like to thank my friends and family, for proper amount of forbearance and never-ending support.

Bibliography

Arntz, Michele A and Lynda F Delph (2001). "Pattern and process: evidence for the evolution of photosynthetic traits in natural populations". In: *Oecologia* 127.4, pp. 455–467.

Asner, Gregory P and Roberta E Martin (2009). "Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests". In: *Frontiers in Ecology and the Environment* 7.5, pp. 269–276.

Asner, Gregory P, Roberta E Martin, and Affendi Bin Suhaili (2012). "Sources of canopy chemical and spectral diversity in lowland Bornean forest". In: *Ecosystems* 15.3, pp. 504–517.

BON, GEO (2015). *An Essential Biodiversity Variable Approach to Monitoring Biological Invasions: Guide for Countries*.

Bruford, Michael W et al. (2017). "Monitoring changes in genetic diversity". In: *The GEO Handbook on Biodiversity Observation Networks*. Springer, pp. 107–128.

Carleer, Alexandre and Eléonore Wolff (2004). "Exploitation of very high resolution satellite data for tree species identification". In: *Photogrammetric Engineering & Remote Sensing* 70.1, pp. 135–140.

Cavender-Bares, Jeannine et al. (2016). "Associations of leaf spectra with genetic and phylogenetic variation in oaks: Prospects for remote detection of biodiversity". In: *Remote Sensing* 8.3, p. 221.

Ceccato, Pietro et al. (2001). "Detecting vegetation leaf water content using reflectance in the optical domain". In: *Remote sensing of environment* 77.1, pp. 22–33.

Close, Dugald C and Clare McArthur (2002). "Rethinking the role of many plant phenolics–protection from photodamage not herbivores?" In: *Oikos* 99.1, pp. 166–172.

Davies, Neil et al. (2012). "A call for an international network of genomic observatories (GOs)". In: *GigaScience* 1.1, p. 5.

Demesure, Brigitte, Bernard Comps, and Rimy J Petit (1996). "Chloroplast DNA phylogeography of the common beech (Fagus sylvatica L.) in Europe". In: *Evolution* 50.6, pp. 2515–2520.

Ellstrand, Norman C and Janis Antonovics (1985). "EXPERIMENTAL STUDIES OF THE EVOLUTIONARY SIGNIFICANCE OF SEXUAL REPRODUCTION II. A TEST OF THE DENSITY-DEPENDENT SELECTION HYPOTHESIS". In: *Evolution* 39.3, pp. 657–666.

Etzold, Sophia et al. (2011). "The carbon balance of two contrasting mountain forest ecosystems in Switzerland: similar annual trends, but seasonal differences". In: *Ecosystems* 14.8, pp. 1289–1309.

Eugster, Werner et al. (2007). "Methodical study of nitrous oxide eddy covariance measurements using quantum cascade laser spectrometery over a Swiss forest". In: *Biogeosciences* 4.5, pp. 927–939.

Flexas, Jaume, Margaret M Barbour, et al. (2012). "Mesophyll diffusion conductance to CO2: an unappreciated central player in photosynthesis". In: *Plant Science* 193, pp. 70–84.

Flexas, Jaume, MIQUEL Ribas-Carbo, et al. (2008). "Mesophyll conductance to CO2: current knowledge and future prospects". In: *Plant, Cell & Environment* 31.5, pp. 602–621.

Frankham, Richard (2010). "Challenges and opportunities of genetic approaches to biological conservation". In: *Biological conservation* 143.9, pp. 1919–1927.

Gausman, Harold W and Victor R Weidner (1985). "Plant leaf optical properties in visible and near-infrared light". In: *Applied Optics* 24, p. 3305.

Gausman, HW et al. (1970). "Relation of light reflectance to histological and physical evaluations of cotton leaf maturity". In: *Applied Optics* 9.3, pp. 545–552.

Geijzendorffer, Ilse R et al. (2016). "Bridging the gap between biodiversity data and policy reporting needs: An Essential Biodiversity Variables perspective". In: *Journal of Applied Ecology* 53.5, pp. 1341–1350.

Gienapp, Phillip et al. (2008). "Climate change and evolution: disentangling environmental and genetic responses". In: *Molecular ecology* 17.1, pp. 167–178.

Gratani, Loretta (2014). "Plant phenotypic plasticity in response to environmental factors". In: *Advances in botany* 2014.

Hartmann, Dennis L et al. (2013). "Observations: atmosphere and surface". In: *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Hoban, Sean et al. (2014). "Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion". In: *Evolutionary Applications* 7.9, pp. 984–998.

Hogg, Robert V and Johannes Ledolter (1987). *Engineering statistics*. Macmillan Pub Co.

Hueni, Andreas, Jan Biesemans, et al. (2009). "Structure, components, and interfaces of the airborne prism experiment (APEX) processing and archiving facility". In: *IEEE Transactions on Geoscience and Remote Sensing* 47.1, pp. 29–43.

Hueni, Andreas, Alexander Damm, et al. (2017). "Field and Airborne Spectroscopy Cross Validation—Some Considerations". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.3, pp. 1117–1135.

Hueni, Andreas, Sindy Sterckx, et al. (2012). "Operational status of APEX and characteristics of the APEX open science data set". In: *Geoscience and remote sensing symposium (IGARSS), 2012 IEEE International*. IEEE, pp. 5009–5012.

Huesca, Margarita et al. (2016). "Canopy structural attributes derived from AVIRIS imaging spectroscopy data in a mixed broadleaf/conifer forest". In: *Remote Sensing of Environment* 182, pp. 208–226.

Hughes, A Randall et al. (2008). "Ecological consequences of genetic diversity". In: *Ecology letters* 11.6, pp. 609–623.

Hulshof, Catherine M and Nathan G Swenson (2010). "Variation in leaf functional trait values within and across individuals and species: an example from a Costa Rican dry forest". In: *Functional Ecology* 24.1, pp. 217–223.

Karl, Jason W and Brian A Maurer (2010). "Multivariate correlations between imagery and field measurements across scales: comparing pixel aggregation and image segmentation". In: *Landscape Ecology* 25.4, pp. 591–605.

Kloeti, P, HM Keller, and M Guecheva (1989). "Effects of forest canopy on throughfall precipitation chemistry". In: *Baltimore Symposium, Baltimore*. Vol. 179, pp. 203–209.

Kochubey, Svetlana M and Taras A Kazantsev (2012). "Derivative vegetation indices as a new approach in remote sensing of vegetation". In: *Frontiers of Earth Science* 6.2, pp. 188–195.

Kokaly, Raymond F and Andrew K Skidmore (2015). "Plant phenolics and absorption features in vegetation reflectance spectra near 1.66 μm". In: *International journal of applied earth observation and geoinformation* 43, pp. 55–83.

Lee, Juheon et al. (2016). "Individual tree species classification from airborne multisensor imagery using robust PCA". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.6, pp. 2554–2567.

Leiterer, Reik et al. (2015). "Forest canopy-structure characterization: A data-driven approach". In: *Forest Ecology and Management* 358, pp. 48–61.

Madritch, Michael D et al. (2014). "Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1643, p. 20130194.

Maitra, Saikat and Jun Yan (2008). "Principle component analysis and partial least squares: Two dimension reduction techniques for regression". In: *Applying Multivariate Statistical Models* 79, pp. 79–90.

Martin, ME et al. (1998). "Determining forest species composition using high spectral resolution remote sensing data". In: *Remote Sensing of Environment* 65.3, pp. 249–254.

Matsuda, Osamu et al. (2012). "Hyperspectral imaging techniques for rapid identification of Arabidopsis mutants with altered leaf pigment status". In: *Plant and Cell Physiology* 53.6, pp. 1154–1170.

Muir, Christopher D et al. (2014). "Morphological and anatomical determinants of mesophyll conductance in wild relatives of tomato (Solanum sect. Lycopersicon, sect. Lycopersicoides; Solanaceae)". In: *Plant, cell & environment* 37.6, pp. 1415–1426.

Nagendra, Harini and Duccio Rocchini (2008). "High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail". In: *Biodiversity and conservation* 17.14, p. 3431.

Navarro, Laetitia M et al. (2017). "Monitoring biodiversity change through effective global coordination". In: *Current opinion in environmental sustainability* 29, pp. 158–169.

O'Connor, Brian et al. (2015). "Earth observation as a tool for tracking progress towards the Aichi Biodiversity Targets". In: *Remote sensing in ecology and conservation* 1.1, pp. 19–28.

Otto, Matthias (2016). *Chemometrics: statistics and computer application in analytical chemistry*. John Wiley & Sons.

Pasqualini, V et al. (2003). "Phenolic compounds content in Pinus halepensis Mill. needles: a bioindicator of air pollution". In: *Chemosphere* 52.1, pp. 239–248.

Pastorelli, R et al. (2003). "Characterization of microsatellite markers in Fagus sylvatica L. and Fagus orientalis Lipsky". In: *Molecular Ecology Resources* 3.1, pp. 76–78.

Peerbhay, Kabir Yunus, Onisimo Mutanga, and Riyad Ismail (2013). "Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and

partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 79, pp. 19–28.

Pereira, Henrique Miguel et al. (2013). "Essential biodiversity variables". In: *Science* 339.6117, pp. 277–278.

Pereira, José Alberto et al. (2007). "Walnut (Juglans regia L.) leaves: phenolic compounds, antibacterial activity and antioxidant potential of different cultivars". In: *Food and Chemical Toxicology* 45.11, pp. 2287–2295.

Peuke, AD et al. (2002). "Identification of drought-sensitive beech ecotypes by physiological parameters". In: *New phytologist* 154.2, pp. 373–387.

Richter, Rudolf, Daniel Schlapfer, and Andreas Muller (2011). "Operational atmospheric correction for imaging spectrometers accounting for the smile effect". In: *IEEE Transactions on Geoscience and Remote sensing* 49.5, pp. 1772–1780.

Rocchini, Duccio et al. (2010). "Remotely sensed spectral heterogeneity as a proxy of species diversity: recent advances and open challenges". In: *Ecological Informatics* 5.5, pp. 318–329.

Scalbert, Augustin (1991). "Antimicrobial properties of tannins". In: *Phytochemistry* 30.12, pp. 3875–3883.

Schaepman, Michael E et al. (2015). "Advanced radiometry measurements and Earth science applications with the Airborne Prism Experiment (APEX)". In: *Remote Sensing of Environment* 158, pp. 207–219.

Schläpfer, Daniel and Rudolf Richter (2002). "Geo-atmospheric processing of airborne imaging spectrometry data. Part 1: parametric orthorectification". In: *International Journal of Remote Sensing* 23.13, pp. 2609–2630.

Schmid, Bernhard (1994). "Effects of genetic diversity in experimental stands of Solidago altissima–evidence for the potential role of pathogens as selective agents in plant populations". In: *Journal of Ecology*, pp. 165–175.

Schneider, Fabian D et al. (2017). "Mapping functional diversity from remotely sensed morphological and physiological forest traits". In: *Nature communications* 8.1, p. 1441.

Singh, Ashbindu and Andrew Harrison (1985). "Standardized principal components". In: *International journal of remote sensing* 6.6, pp. 883–896.

Stickler, Claudia M and Jane Southworth (2008). "Application of multi-scale spatial and spectral analysis for predicting primate occurrence and habitat associations in Kibale National Park, Uganda". In: *Remote sensing of environment* 112.5, pp. 2170–2186.

Szathmáry, Eörs, Ferenc Jordán, and Csaba Pál (2001). "Can genes explain biological complexity?" In: *Science* 292.5520, pp. 1315–1316.

Telles, Annie Campello, Larine Kupski, and Eliana Badiale Furlong (2017). "Phenolic compound in beans as protection against mycotoxins". In: *Food chemistry* 214, pp. 293–299.

Tsai, Fuan and William Philpot (1998). "Derivative analysis of hyperspectral data". In: *Remote Sensing of Environment* 66.1, pp. 41–51.

Tucker, Compton J (1980). "Remote sensing of leaf water content in the near infrared". In: *Remote sensing of Environment* 10.1, pp. 23–32.

Turner, Woody (2014). "Sensing biodiversity". In: *Science* 346.6207, pp. 301–302.

Vihervaara, Petteri et al. (2017). "How essential biodiversity variables and remote sensing can help national biodiversity monitoring". In: *Global Ecology and Conservation* 10, pp. 43–59.

War, Abdul Rashid et al. (2012). "Mechanisms of plant defense against insect herbivores". In: *Plant signaling & behavior* 7.10, pp. 1306–1320.

Wilson, EO and Frances M Peter (1988). "The Loss of Diversity Causes and Consequences". In:

Wold, Svante, Lennart Eriksson, et al. (2004). "The PLS method–partial least squares projections to latent structures–and its applications in industrial RDP (research, development, and production)". In: *Unea University*.

Wold, Svante, Arnold Ruhe, et al. (1984). "The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses". In: *SIAM Journal on Scientific and Statistical Computing* 5.3, pp. 735–743.

Wold, Svante, Michael Sjöström, and Lennart Eriksson (2001). "PLS-regression: a basic tool of chemometrics". In: *Chemometrics and intelligent laboratory systems* 58.2, pp. 109–130.

Yamasaki, Eri et al. (2017). "Genomics meets remote sensing in global change studies: monitoring and predicting phenology, evolution and biodiversity". In: *Current Opinion in Environmental Sustainability* 29, pp. 177–186.

A. Appendix

Table A.1: The statistics of microsatellites lengths of sampled *F.sylvatica* individuals derived from capillary electrophoresis with use of GeneMapper software.

length statistics	FS1-03	FS1-15	FS3-04	FS4-46	FCM5
standard deviation	4.63	8.07	1.73	22.74	12.51
mean	91.55	111.99	200.95	251.75	299.87
variance	21.39	65.09	2.99	517.23	156.45
maximum	108	137	206	328	322
minimum	83	93	194	221	280

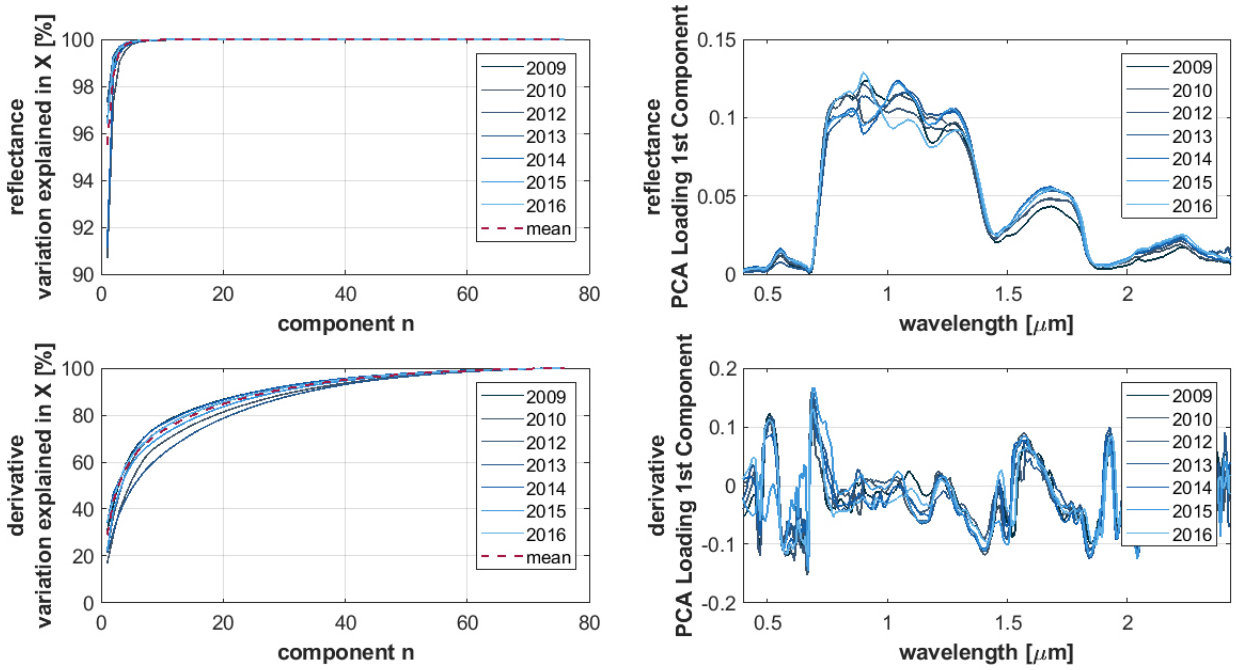


Figure A.1: The cumulative variation explained by Principal Components (PCs) and spectral bands loading to the 1st PC for reflectance and derivatives of z-score of the reflectance (derivative). Both derived from Principal Component Analisis (PCA) conducted on all the pixels included in the crowns of sampled *F.sylvatica* individuals for the considered years separately.

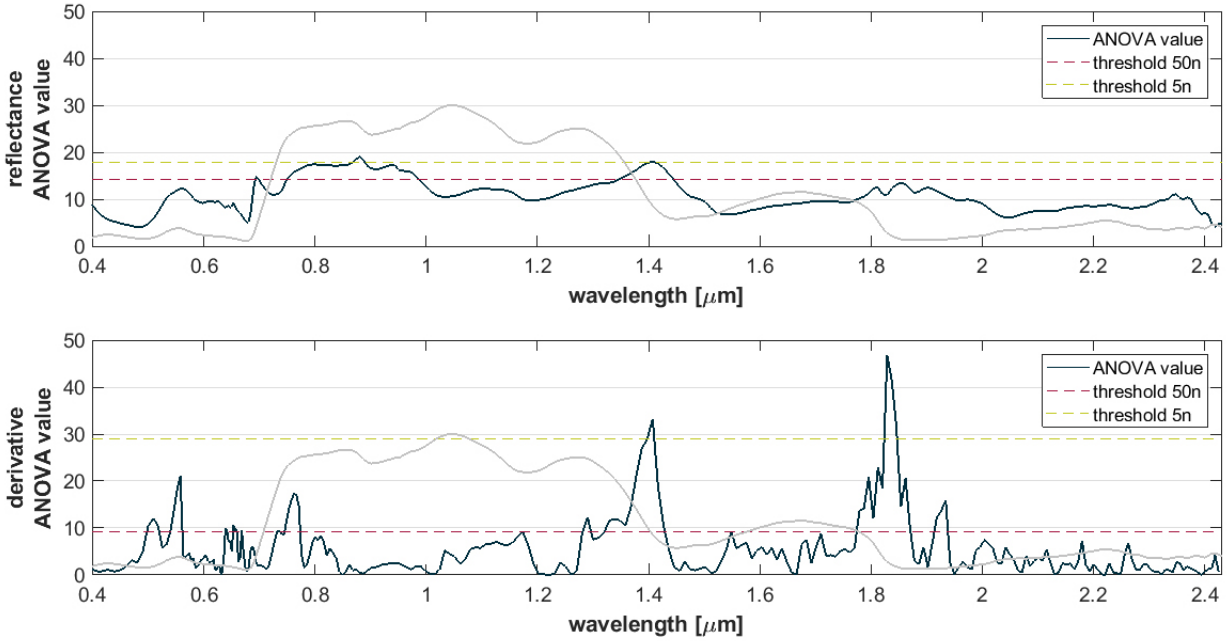


Figure A.2: The ANOVA value, calculated as: (p-value)/*F-statistic* ratio for each spectral band for reflectance and derivatives of z-score of the reflectance (derivative). Additionally, the thresholds for the selections, dictated by 5 and 50 bands with the highest ANOVA value are shown.

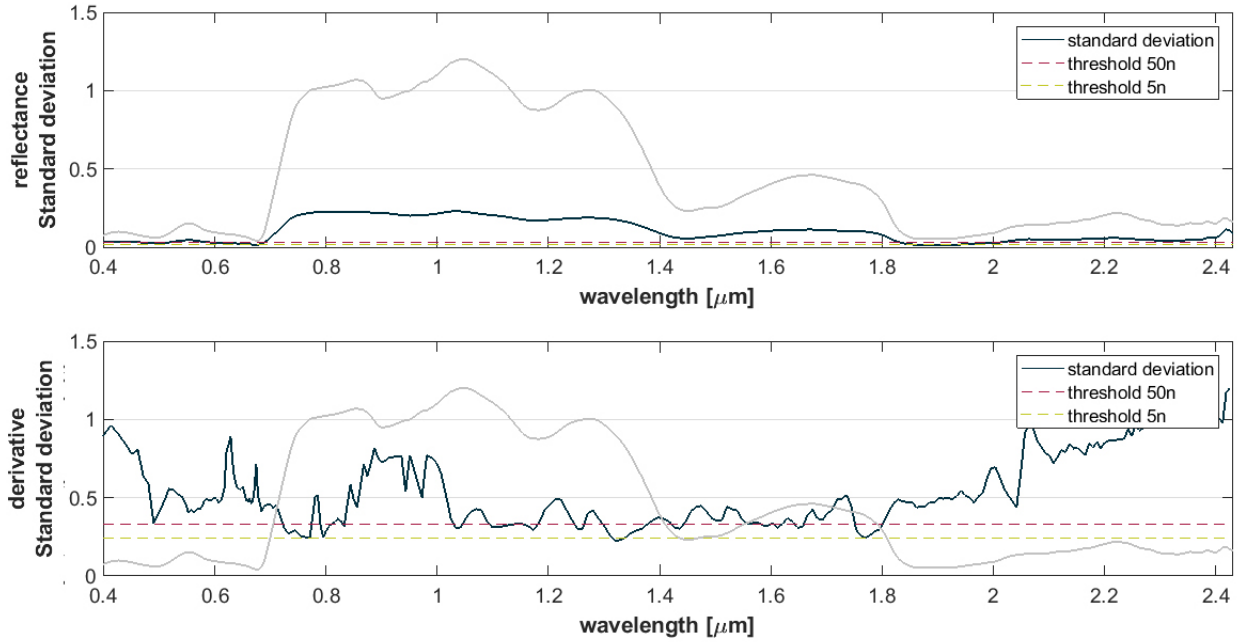
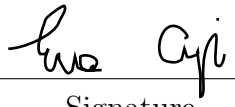


Figure A.3: The standard deviation for each spectral band from the reflectance and derivatives of z-score of the reflectance (derivative) values between the acquisitions taken at the same date, but with different points of viewing*. Additionally, the thresholds for the selections, dictated by 5 and 50 bands with the lowest standard deviation are shown. *Additional images, not considered in the analyses of genetic cluster discrimination, were used.

Personal declaration

I hereby declare that the submitted thesis is the result of my own, independent, work. All external sources are explicitly acknowledged in the thesis.

Zurich, 30.07.2018 

Place, Date Signature