

RESEARCH ARTICLE

Making the Genotypic Variation Visible: Hyperspectral Phenotyping in Scots Pine Seedlings

Jan Stejskal^{1*}, Jaroslav Čepl¹, Eva Neuwirthová^{1,3}, Olusegun Olaitan Akinyemi^{1,2}, Jiří Chuchlík¹, Daniel Provazník¹, Markku Keinänen^{2,4}, Petya Campbell^{5,6}, Jana Albrechtová³, Milan Lstibůrek¹, and Zuzana Lhotáková³

¹Department of Genetics and Physiology of Forest Trees, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic. ²Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu, Finland. ³Department of Experimental Plant Biology, Charles University, Prague, Czech Republic. ⁴Center for Photonic Sciences, University of Eastern Finland, Joensuu, Finland. ⁵Department of Geography and Environmental Sciences, University of Maryland Baltimore County, Baltimore, MD, USA. ⁶Biospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA.

*Address correspondence to: stejskalj@fd.czu.cz

Hyperspectral reflectance contains valuable information about leaf functional traits, which can indicate a plant's physiological status. Therefore, using hyperspectral reflectance for high-throughput phenotyping of foliar traits could be a powerful tool for tree breeders and nursery practitioners to distinguish and select seedlings with desired adaptation potential to local environments. We evaluated the use of 2 nondestructive methods (i.e., leaf and proximal/canopy) measuring hyperspectral reflectance in the 350- to 2,500-nm range for phenotyping on 1,788 individual Scots pine seedlings belonging to lowland and upland ecotypes of 3 different local populations from the Czech Republic. Leaf-level measurements were collected using a spectroradiometer and a contact probe with an internal light source to measure the biconical reflectance factor of a sample of needles placed on a black background in the contact probe field of view. The proximal canopy measurements were collected under natural solar light, using the same spectroradiometer with fiber optical cable to collect data on individual seedlings' hemispherical conical reflectance factor. The latter method was highly susceptible to changes in incoming radiation. Both spectral datasets showed statistically significant differences among Scots pine populations in the whole spectral range. Moreover, using random forest and support vector machine learning algorithms, the proximal data obtained from the top of the seedlings offered up to 83% accuracy in predicting 3 different Scots pine populations. We conclude that both approaches are viable for hyperspectral phenotyping to disentangle the phenotypic and the underlying genetic variation within Scots pine seedlings.

Introduction

Changes under global climatic conditions necessitate the selection of more resistant tree genotypes in forestry [1,2]. Exploring intrapopulation variability may be the crucial first step in curating adaptations to the future climate in tree species with long-generation cycles. Scots pine has one of the largest distribution ranges and grows naturally across various climatic conditions, creating different ecotypes [3,4]. Some ecotypes may be more suitable for afforestation of specific sites due to their resistance to various environmental stresses [5,6]. Selection of resistant individuals based on their phenotype may not be reliable [7] because of environmental effect; instead, selection must be based on genetic estimates after a proper design.

High-throughput phenotyping (HTP) has become widely utilized in crop science [8,9]; however, in forestry, it still needs more research and verification to be adopted [10,11]. The use of HTP is driven by the need to understand plant performance and growth response to environmental stresses and pathological agents and to offer solutions to various breeding scenarios [8,12]. Managing large-scale phenotypic data remains one of the major bottlenecks to plant breeding and functional genomics [11,13].

Vegetation 350- to 2,500-nm [visible to shortwave infrared (VSWIR)] reflectance data contain valuable information about the leaf's chemical composition, such as photosynthetic pigments [14–16], water content [17,18], and structural compounds [19–22] that correspond to plant physiological status [23–26],

Citation: Stejskal J, Čepl J, Neuwirthová E, Akinyemi OO, Chuchlík J, Provazník D, Keinänen M, Campbell P, Albrechtová J, Lstibůrek M, et al. Making the Genotypic Variation Visible: Hyperspectral Phenotyping in Scots Pine Seedlings. *Plant Phenomics* 2023;5:Article 0111. <https://doi.org/10.34133/plantphenomics.0111>

Submitted 26 January 2023
Accepted 10 October 2023
Published 14 November 2023

Copyright © 2023 Jan Stejskal et al. Exclusive licensee Nanjing Agricultural University. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY 4.0).

and growth parameters [27] changing with leaf phenology during the vegetative season [28,29]. Therefore, leaf functional traits derived from seedlings' hyperspectral reflectance are potentially a powerful tool in tree nurseries for evaluating the quality of the seedlings. We suppose that some leaf functional traits influencing optical response are provenance-specific; thus, if the seedling provenance is traceable through VSWIR reflectance data, hyperspectral phenotyping can serve to avoid possible maladaptation of seed sources. Even in certified stands where reproduction material is collected, there are still uncertainties about the actual origin and the true ecotypic features due to many decades of artificial regeneration [30,31]. Another benefit of the leaf functional traits detected using hyperspectral reflectance could be the predictions of the seedling's performance under stress, which could help identify seedlings unsuitable for deployment. Using optical leaf data, it is possible to monitor seedling infections already at the nursery [32].

Since the early 1990s, accurate laboratory and field measurements of hyperspectral reflectance have been made with spectroradiometers covering spectra from the visible (VIS) through the near-infrared (NIR) to the SWIR [33]. The resulting reflectance captures the information about the leaf or canopy properties: The reflectance in the VIS part of the spectrum (300 to 700 nm) responds primarily to pigment content [34]; NIR (750 to 1,350 nm) reflectance is influenced by the internal structure of the leaves [35] or canopy structure [36] and water content [37]. The sharp increase in reflectance between VIS and NIR is called the "red edge" (RE), is usually defined by a wavelength range of 680 to 750 nm [38], and is known to be indicative of nitrogen levels and stress, e.g., [24,25,39]. SWIR spectral region could be subdivided into SWIR1 (1,350 to 1,800 nm) and SWIR2 (2,000 to 2,400 nm), and, here, the reflectance is dependent on the content of water and structural compounds of the leaves [40–42].

Multispectral [43] or hyperspectral optical sensors [44] could detect the leaf optical properties and chlorophyll content at the canopy level. This approach is usually applied in remote sensing studies using satellite, airborne, or unmanned automated vehicle sensors. Although hyperspectral cameras with spatial resolution can be used under laboratory conditions, spectroradiometers without spatial resolution are still viable for detecting leaf physiological properties without labor-intensive image segmentation [45], a primary bottleneck in image-based spectrometry. These spectroradiometers can be used under the field or laboratory conditions, using a contact or noncontact approach. Commonly used field spectroradiometers are equipped with a fiber optical cable (OC) with optional attachable equipment, such as a contact probe (CP) with an internal light source [21,42]. Each measurement approach has a different light source and sensor geometry, providing specific advantages regarding speed, laboriousness, and field operability. The reflectance data acquired by these measurement approaches are not fully compatible, and, therefore, they are often compared with each other: OC versus CP [37,46,47].

The CP is designed to measure the reflectance of planar leaves (at the leaf level) that optimally cover its entire field of view (1 cm²). In planar leaves, which evenly cover the field of view, CP measurements are simple and fast. However, in the case of conifers, issues such as needle overlapping and gap fraction between the needles need to be addressed for contact measuring, making it a challenging and time-consuming process [48]. Nevertheless, some studies have measured the needle

optical properties by CP on shoot samples [49], stacked in one layer [46,50], or fixed in a custom-made holder for integration sphere [48,51,52].

A spectroradiometer outfitted with a CP attached to the OC allows us to obtain a biconical reflectance factor (BCRF) [53]. Using the OC to measure the object and a white panel at a distance under an external light source (i.e., solar light or an artificial one), we obtain a hemispherical conical reflectance factor (HCRF). BCRF focuses on a limited range of incident light within a conical solid angle, while HCRF considers light coming from the entire hemisphere, in our case, ambient sky illumination [53,54]. Both reflectance quantities BCRF and HCRF are expressed as relative values related to the reflectance of the white reference panel measured in a given setup.

The leaf's internal structure affects its optical properties measured by a spectroradiometer equipped with the CP [55] or the integration sphere [56]. However, twig, branch, or canopy morphology, which could be ecotype specific, is not influencing the CP spectral signal. The OC alone can be used for proximal/canopy hyperspectral reflectance measurements to yield reproducible results, albeit highly susceptible to ambient light and atmospheric conditions [54]. The main advantages of the OC method under stable light conditions are its relatively higher speed in contrast to the CP method and the larger field of view, depending on the proximity to the measured object.

Moreover, the OC reflectance acquisition enables canopy level measurement of small plants. Thus, the spectral information also involves signals from the canopy structure and other morphological features of the seedlings unattainable by the CP measurements. The prediction and classification of samples using machine learning (ML) algorithms with VSWIR reflectance are known to the scientific community, including various applications in forestry. It is important to note that the reflectance data have been used to classify trees accurately at the species level [57,58]. Recently, emerging studies successfully used reflectance spectra to classify plant groups at the species level, i.e., interspecific hybrids [59,60] and local populations of single species [42]. Nevertheless, a comparative case study in BCRF and HCRF in an ML context to detect various conifer seedling populations/ecotypes is missing.

In the present study, we compare the potential of using the OC versus CP for the fast and efficient manual HTP of Scots pine seedlings originating from 3 local populations corresponding to 3 seed orchards. The specific aims were to (a) demonstrate that both approaches are viable for disentangling the phenotypic and potential genetic variation within the 3 populations of Scots pine seedlings; (b) evaluate which spectral regions (from VIS to SWIR) were the most sensitive in population distinction; (c) predict the given origin of seedlings by 2 selected ML algorithms - [random forest (RF) and support vector machine (SVM)] based on reflectance factor acquired by CP (BCRF) and OC (HCRF).

Materials and Methods

Study site and plant material

The study was conducted using 1-year-old (their second growing season) Scots pine (*Pinus sylvestris* L.) seedlings originating from 3 seed orchards: Plasy (P), Třeboň (T) and Děčín (D) from the Czech Republic. Together, 1,788 seedlings were measured,

specifically 746 seedlings from Plasy, 750 from Třeboň, and 292 from Děčín (Fig. 1). Seedlings were grown in planters with 60 individuals. In a single planter, there were always only seedlings from one of the seed orchards, and single half-sib families shared the same line of cells. The reflectance factor measurements were carried out between 7 and 10 September 2021, at the Sofronka arboretum in Pilsen, Czech Republic, where all seedlings were grown. It is worth mentioning that there were differences among seedling populations in terms of mortality rate during the first winter (2021 season) after germination (Table 1). The upland pine ecotype represented by the population from Děčín exhibited the highest mortality. In addition, seedlings from the Děčín population exhibited lower average height in comparison to seedlings from Plasy and Třeboň populations.

Measurements of the reflectance factors

For the spectral data assessment of 1,788 Scots pine seedlings, we used ASD FieldSpec 4 Wide-Res portable spectroradiometer (Malvern Panalytical Ltd., United Kingdom). The reflectance factors were measured in a span of 350 to 2,500 nm with an integration time of 8 ms for CP and 34 ms for OC, respectively. Three sensors cover the spectroradiometer range, changing in 1,000 and 1,800 nm. The spectroradiometer was equipped with an OC and attachable CP (ASD Plant Probe, Malvern Panalytical Ltd., United Kingdom). In all seedlings, the leaf-level BCRF was measured using the CP, and the whole seedling HCRF was acquired by OC. The seedlings were measured in the same order by both methods to ensure an accurate match in the sequence of measured individuals within 4 consecutive days. In both spectra acquisition approaches, the reflectance

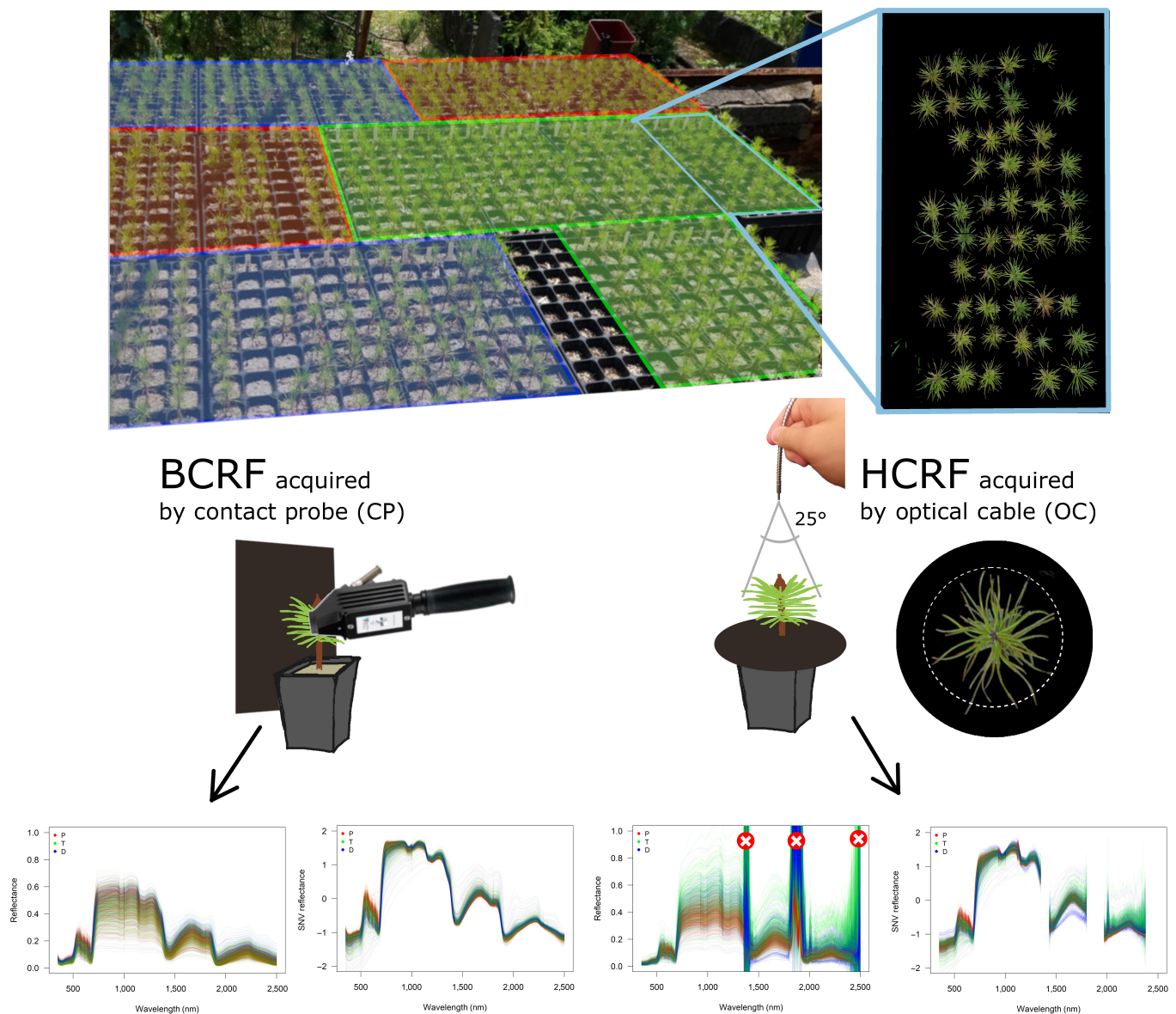


Fig. 1. Experimental and technical design. Planters' photo is a generic illustration of how the ecotypes were replicated; spectral reflectance measurements were done with a spectroradiometer equipped with a CP acquiring BCRF on the left and by OC acquiring HCRF on the right with an additional nadir image of the measured seedling. Data processing demonstrated on graphs of spectral reflectance curves with the x axis corresponding to wavelength (in nanometers) and the y axis corresponding to reflectance (%). Red circles with a white cross above maximum reflectance values in the third graph denote unreliable spectral regions that were removed. The second and fourth graphs illustrate the impact of SNV transformation.

Table 1. Description of the seed origin (3 distinct local populations comprising lowland and upland ecotype) and a common garden site. Děčín plus trees were mostly selected in the elevation of 700 m above the sea level. Climatic variables for 2021 were adopted from <https://www.inmeteo.cz/>. The mean heights are reported with respective standard deviations.

	Seed orchards			Common garden
	Plasy	Třeboň	Děčín	Sofronka
GPS coordinates	49.9087617°N 13.4427394°E	49.0166106°N 14.8247583°E	50.8191364°N 14.1172131°E	49.789113°N 13.386193°E
Elevation (m.a.s.l.)	385	430	465	331
Ecotype	Lowland	Lowland	Upland	–
Mean annual precipitation (mm)	489	623	826	532
Mean annual temperature (°C)	11.1	11.1	8.9	8.5
Mean annual sunshine hours (h)	1,524.3	1,571.5	1,099.9	1,457.6
Mean height of the progeny (mm)	49.6 ± 9.6	51.2 ± 10.4	46.5 ± 10.4	–
Mortality within the progeny (%)	5.7	13.9	28.6	–

factor of the white panel with approximately 99% reflectance was recorded every 15 min and/or after the change in incoming radiation and later used as a white reference. The reflectance of the white spectralon panel (SphereOptics, Herrsching am Ammersee, Germany) used to calibrate the raw data to reflectance reaches 98 to 99% reflectance up to 1,800 nm and further varies between 94% and 98% reflectance.

Leaf level BCRF measurement

BCRF refers to the ratio of the reflected flux collected through an OC coupled with a CP, where the incident flux is limited to a conical solid angle. The incident flux only includes light coming from a particular range of directions, in our case, an integrated light source (a krypton halogen bulb) inside the CP [53].

The contact measurement at the leaf level was not reliant on weather conditions. The needles were pushed to a nonreflective black surface in the entire measured spectral range to minimize the background spectral noise or radiation transmitted [46]. The needles were arranged to cover the field of view. Each measurement was recorded from 25 scans as protection against overheating.

Canopy level HCRF measurement

HCRF refers to the ratio of the reflected flux collected through an OC to the incident flux from the entire hemisphere. In this case, the incident flux encompasses light coming from all directions of ambient sky illumination. The reflected flux is measured within a specific conical solid angle by bare OC [53].

Canopy level reflectance factor measurements were collected under clear sky conditions, rendering stable incoming light.

The white panel was repeatedly measured and used to calibrate to reflectance the closest in time, thus accounting for short-term variation under illumination conditions and the diurnal change in solar illumination. We applied black nonreflective paper stencils to the seedling bases to eliminate soil reflectance from the recorded signal. For each measurement, the OC was pointed at the nadir (i.e., positioned vertically) to the terminal bud of the seedling from approximately 5 cm in height. The OC's field of view angle is 25°, so the entire seedling was included in the field of view, and the background was effectively minimized. Each measurement acquired recorded an average from 50 consecutive ASD scans.

Data processing and statistical analysis

Data preprocessing

Because of prominent noise in HCRF spectra (Fig. 1), we removed the reflectance from the NIR (1,350 to 1,430 nm) and SWIR (1,800 to 1,970 nm and 2,380 to 2,500 nm) regions, strongly affected by atmospheric aerosols and scattering. Similar spectral range removal is a frequent practice when working with field spectra acquired under solar illumination [59,61]. Moreover, we transformed BCRF and HCRF for consecutive analysis by standard normal variate (SNV) [62]. To make all spectra comparable in terms of intensities, the SNV method subtracts each spectrum from its own mean, followed by a division by its own standard deviation.

Statistical analysis

For statistical analysis, native functions of the R software version 4.0.4 (R Core Team 2021) and the ASReml library for R

version 4 [63] were used. A univariate linear mixed model was fitted to evaluate all traits of interest with the following terms:

$$y = 1\mu + X_1\beta_x + X_2\beta_y + X_3\beta_{xy} + Zm + e$$

where y corresponds to the data vector; μ is the overall mean effect; β_x is a fixed effect associated with sites of the origin; β_y is a fixed effect of the measurement day; β_{xy} is a vector of the fixed interaction between origin and measurement day; m is the maternal general combining ability, with $m \sim N(0, \sigma_m^2 Im)$; and e is the random vector of errors, with $e \sim N(0, R)$. The symbol 1 is used for a vector of ones, while X and Z designate incidence matrices associated with fixed and random effects, and Im provides an identity matrix of order m . The R matrix was evaluated assuming independent errors, i.e., $R = \sigma^2 Im$.

In ASReml software, we performed the Wald tests to compare ecotypes and populations within our study. Wald test is a pseudo-analysis of variance using incremental Wald statistics. The incremental Wald tests have an asymptotic χ^2 distribution, with degrees of freedom given by the number of estimable effects [63]. For pairwise comparison, we based our predicted site means of seed origin on the above-described linear mixed model.

ML algorithms and predictions

Two ML algorithms, RF and SVM, were applied to predict the seedling population of origin based on the VSWIR representative reflectance dataset (spectral phenome). R software was used to implement RF and SVM. For RF, we used the randomForest function from the randomForest package (v 4.6-14) [64], with $ntree = 1,000$ and retaining all other default parameters. The importance of individual wavebands for the RF model was extracted using the importance function.

The SVM model was fit using the svm function from the e1071 package (v 1.7-3) [65] with default parameters. The raw and SNV-transformed spectra of BCRF and HCRF were used for ML prediction, and prediction results were compared.

For both OC and CP, data were either treated in whole or separated by the day of measurement to inspect whether the predictions within the light conditions of a single day would be more consistent (it concerns mostly the OC measurement). We used a cross-validation approach with 100 iterations, randomly dividing observations into training (90%) and testing (10%) subsets. We then reported the mean accuracy of predictions for different input data.

Results

Differentiation of population based on leaf and seedling-level reflectance factors BCRF and HCRF

The differences between the mean reflectance factors of seedlings according to their origin (P, T, and D populations) were tested by the mixed linear model for each wavelength of BCRF and HCRF separately, and the concluding P values are depicted in Figs. 2 and 3).

Leaf BCRF measured by the CP

The average BCRF of all seedlings populations was significantly different in pairwise comparison ($\alpha = 0.05$) in the spectral regions measured by CP (Fig. 2). Differences in BCRF were the most pronounced within VIS and RE. In all 3 pairwise comparisons, a discrepancy occurred in the vicinity of 1,000 nm

(sensor switch), where the P value exceeded the 0.05 significance threshold (in Fig. 2, it effectively dropped below the broken line, which indicates the threshold on the logarithmic scale). The pairwise comparison of the T and D populations exhibited smaller general differences in all spectral regions compared to the comparison between P and T (Fig. 2A) or P and D (Fig. 2B). Moreover, their difference was insignificant in a narrow region before RE (Fig. 2C).

HCRF of whole seedlings measured by the OC

In the case of the OC measuring method, the average of HCRF of all seedling populations differed significantly in pairwise comparisons ($\alpha = 0.05$) across the VIS and RE and a substantial portion of the NIR and SWIR (Fig. 3). The reported P values exceeded the 0.05 significance threshold ($\alpha = 0.05$), in the region near 1,000 nm (sensor switch) and 1,500 nm in all 3 comparisons. Pairwise comparisons of wavelengths in SWIR2 (beyond 2,000 nm) exhibited pronounced amplitude variation (Fig. 3A to C). Pairwise comparisons did not detect any significant differences between T and D in SWIR1 (1,430 to 1,970 nm), and these differences in SWIR2 approached the critical level (Fig. 3C).

Leaf and canopy, BCRF and HCRF, detection of different patterns among seedling populations

In general, P values of pairwise comparison of leaf and canopy BCRF and HCRF data of all 3 seedling populations maintained a similar trend at the VSWIR spectral range (350 to 2,500 nm). However, we encountered several discrepancies (Fig. 4) in P value patterns when visualizing these trends. P values of pairwise comparisons of BCRF are closer to zero than HCRF in a large part of the examined spectral interval. For the BCRF, the significance of the difference was most pronounced in the VIS, with a decrease in the red region (near 660 nm). RE is the second-highest peak in terms of significance for BCRF and HCRF. In the NIR, the significance of the comparison decreases toward the sensor switch (1,000 nm) and falls below significance for both BCRF and HCRF. In this region, the patterns are slightly different for BCRF and HCRF, but the relevance is questionable. In the spectral band between 1,100 and 1,400 nm, all pairwise comparisons of BCRF and HCRF based on the 3 populations exhibited significant differences. In the vicinity of 1,500 nm, the pairwise comparison of populations HCRF was insignificant compared to the BCRF. The main difference between the distribution of the P values between BCRF and HCRF in the population comparison occurred in the SWIR (1,430 to 1,970 nm). The P values of the BCRF (1,500 to 1,800 nm) are noticeably closer to zero, in contrast to the P values of HCRF. In the SWIR beyond the 1,970 nm, the P values of the BCRF were also closer to zero, with a lesser amplitude, than HCRF. Despite this observed difference, both HCRF and BCRF P values were significant within SWIR beyond the 1,970 nm.

Using ML algorithms to distinguish seedling populations based on reflectance factors

We evaluated the ability of selected ML algorithms to separate the populations by comparing the predictions' accuracy. We summarized the RF and SVM (Table 2) accuracies for raw and SNV standardized canopy HCRF and leaf BCRF. The accuracy (in percentage) indicates the probability with which the given method correctly assigns the seedling to the respective population

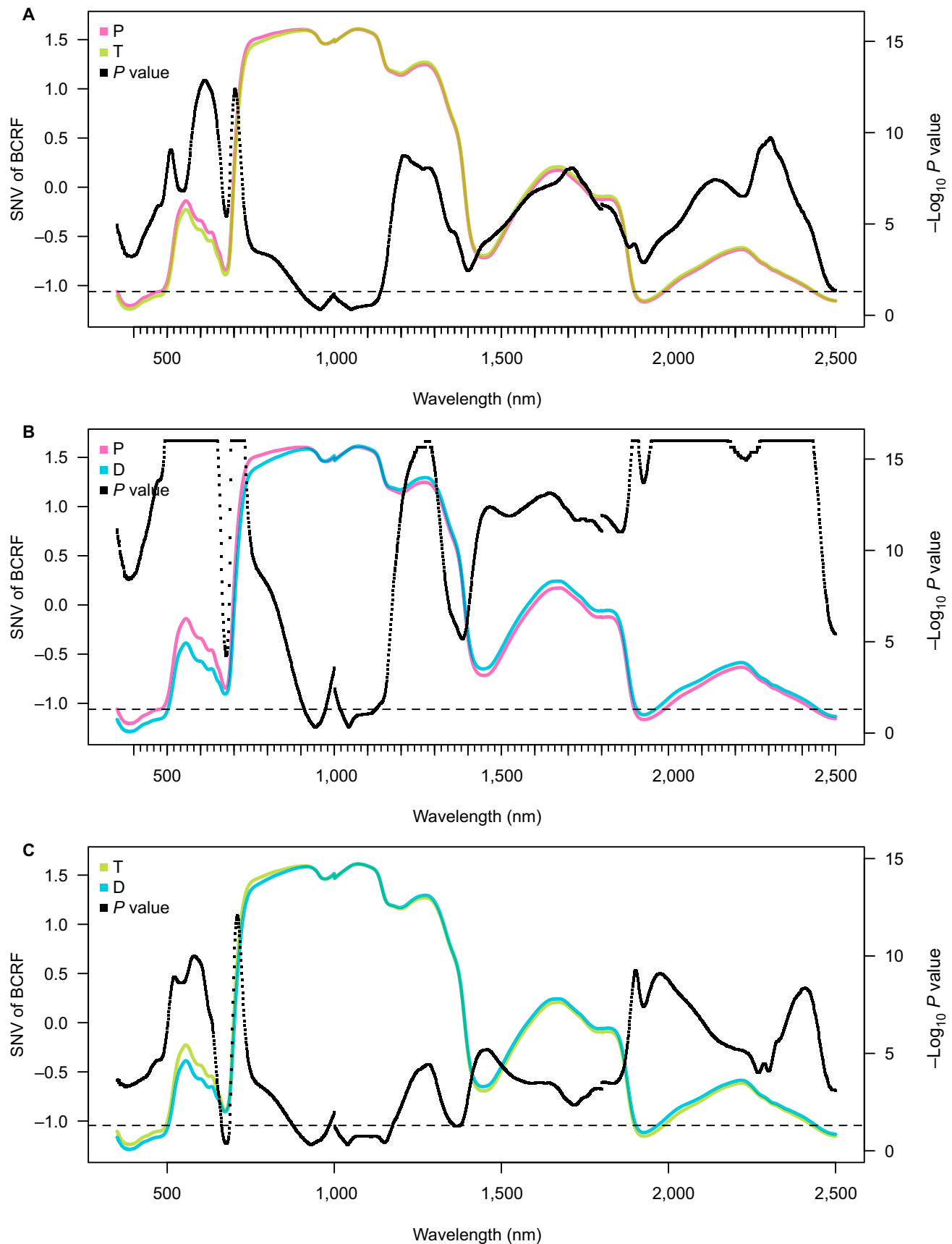


Fig. 2. Average BCRF of seedlings from Plasy and Třeboň (A), Plasy and Děčín (B), and Třeboň and Děčín (C) measured by the CP. The reflectance spectral curve represents the average values from the individual sets of BCRF (SNV-transformed). The black points represent the negative decadic logarithmic P value of the pairwise comparison based on the linear model. P values of the pairwise comparison above the horizontal line represent a significant difference ($\alpha = 0.05$) between the seedling populations from 2 seed orchards.

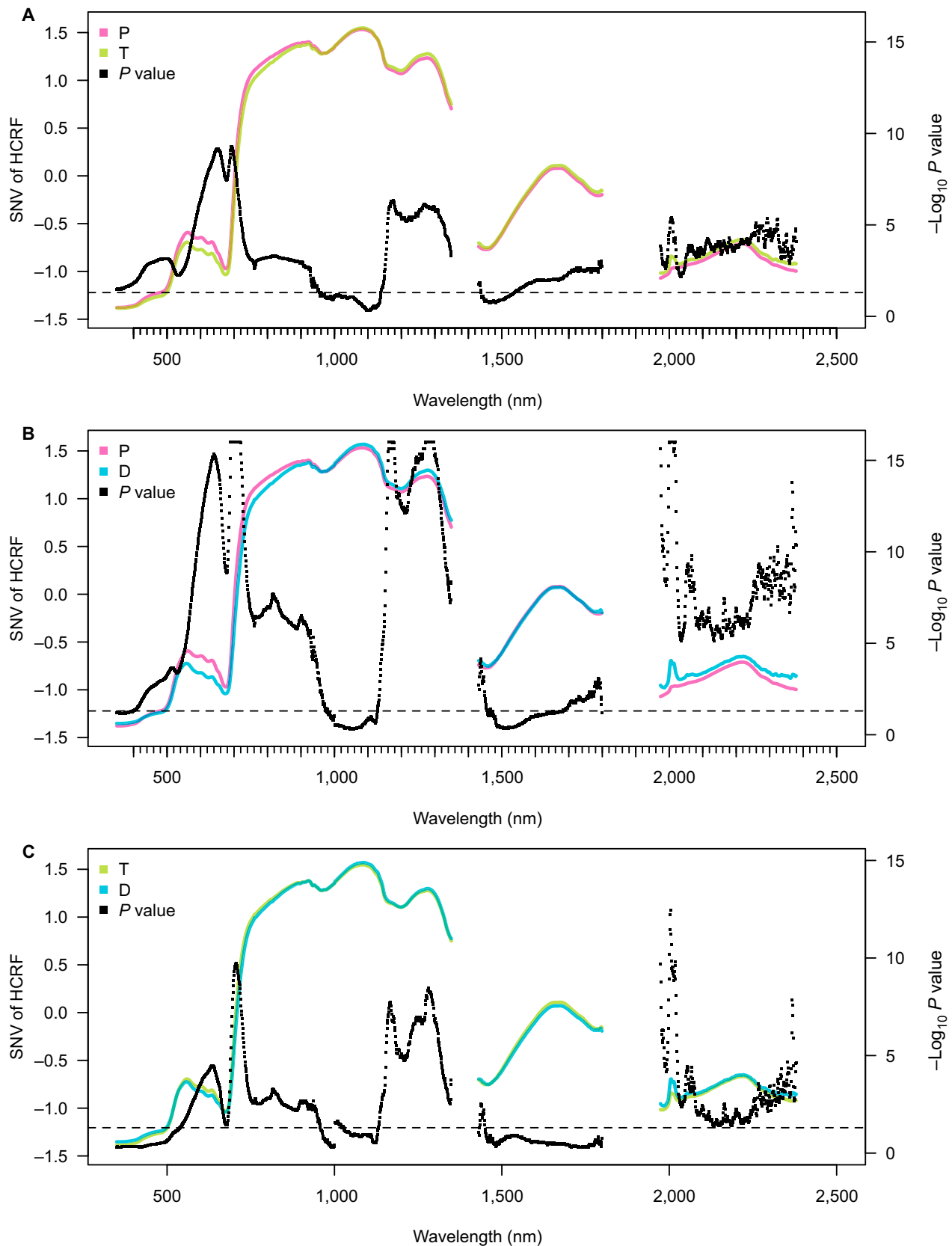


Fig. 3. Average of HCRF of seedlings from Plasy and Třeboň (A), Plasy and Děčín (B), and Třeboň and Děčín (C) measured by the OC. The reflectance spectral curve represents the average values from the individual sets of HCRF (SNV-transformed). The black points represent the negative decadic logarithmic P value of the pairwise comparison based on the linear model. P values of pairwise comparison occurring above the horizontal line represent a significant difference ($\alpha = 0.05$) between the seedling population from 2 different seed orchards. Because of prominent noise in HCRF spectra, we removed the reflectance from the NIR (1,350 to 1,430 nm) and SWIR (1,800 to 1,970 nm and 2,380 to 2,500 nm) regions.

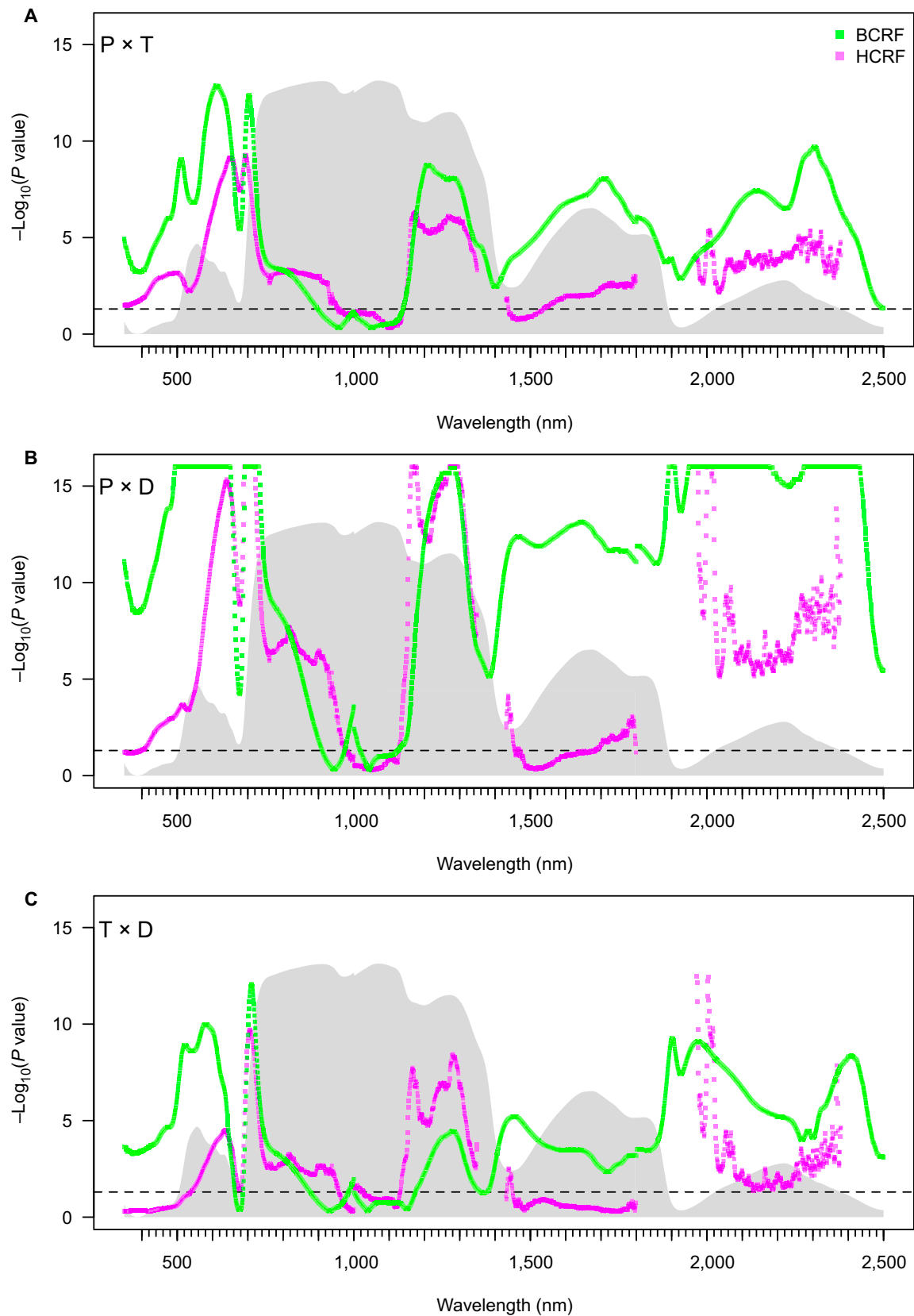


Fig. 4. Two lines represent the negative logarithmic P value of pairwise comparisons of BCRF and HCRF of seedlings (SNV-transformed): Plasy and Třeboň (A), Plasy and Děčín (B), and Třeboň and Děčín (C) measured by CP (needles; CP = green; BCRF) and OC (whole seedling; OC = pink; HCRF). Gray shadow represents the average reflectance factor of the spectral curve. P values of pairwise comparison occurring above the horizontal dashed line represent a significant difference ($\alpha = 0.05$) between the seedling population from 2 different seed orchards. Because of prominent noise in HCRF spectra, we removed the reflectance from the NIR (1,350 to 1,430 nm) and SWIR (1,800 to 1,970 nm and 2,380 to 2,500 nm) regions.

of origin. We obtained the highest accuracy in the model based on raw whole seedling HCRF. The probability that HCRF of the seedling was assigned to the correct population reached 83% (RF) for day 1. The comparably worse results were obtained from the BCRF measured by the CP (54% to 68%, depending on measurement day and transformation). The maximum accuracy of BCRF prediction was 68% on day 2.

The SNV transformation led to an overall increase in accuracy no matter which data or ML algorithm was used. It is worth noting that prediction improvement was more visible in HCRF (6%) compared to BCRF (1% to 2%) when both days were included. This general trend was disrupted by day 1 of

HCRF, where the difference made by SNV amounts to 10%. Similarly, day 1 of BCRF exhibited an increase in prediction accuracy (5%) after SNV transformation. The inconsistencies among measurement days will be further discussed.

Importance of specific spectral regions for RF separation of seedling populations

Another supporting evidence of why the BCRF and HCRF differ in their ability to predict seedling origin by RF was documented using the importance values of each wavelength for an RF model (Fig. 5). In both reflectance quantities, the importance showed similar peaks in VIS and RE. For BCRF, the important wavelengths for seedling population prediction were mostly limited to these wavelengths in VIS and RE. In contrast, HCRF showed several more spectral regions with high importance in NIR (900 to 1,200 nm and 1,100 to 1,200 nm) and SWIR (1,450 to 1,520 nm, 2,000 to 2,020 nm, and 2,280 to 2,380 nm). This difference in the number of spectral intervals with high importance for RF prediction between BCRF and HCRF is probably closely connected to the higher prediction accuracy of RF models based on HCRF data. The SNV transformation increased the prediction power of RF and SVM to a variable degree, which was also reflected in the patterns of importance values (Fig. S1).

Discussion

In the last decade, hyperspectral phenotyping has been frequently applied in the evaluation of seedling susceptibility to drought [10], pollution [66], and early detection of various pathogen-induced diseases [32,67,68]. Spectral reflectance enables the efficient generation of multivariate datasets with high numbers of samples, including time series, ideal for discrimination of seedling conditions, using ML algorithms, which are an indispensable tool for hyperspectral data analysis [69,70].

Table 2. The summarized accuracy values (%) represent the relative number of correctly assigned true origins based on 2 ML algorithms: RF and SVM. ML algorithms modeled BCRF raw values and HCRF raw values and their values normalized by SNV. Algorithms were compared on the first (day 1), the second day (day 2), and for both days together (both days).

	Day 1		Day 2		Both days	
	RF	SVM	RF	SVM	RF	SVM
The summarized accuracy values (%)						
HCRF	73	64	66	67	65	67
HCRF (SNV)	83	81	71	72	71	73
BCRF	56	54	67	64	53	51
BCRF (SNV)	61	60	68	62	55	52

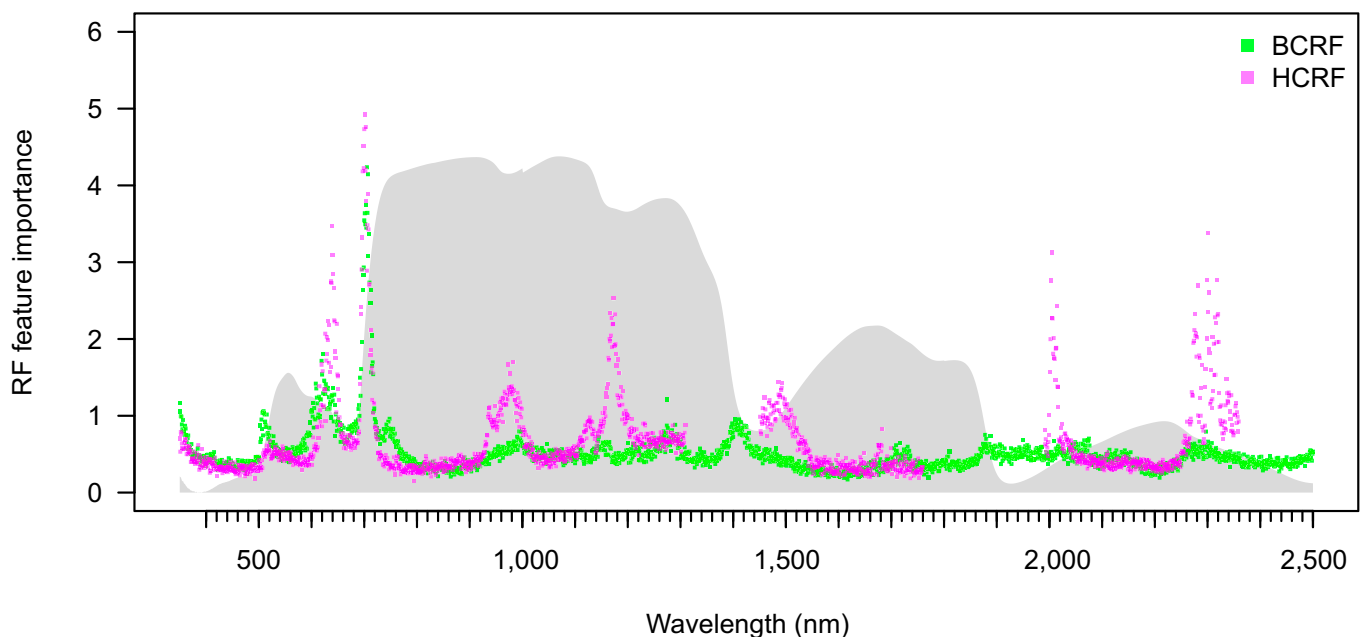


Fig. 5. The points represent the prediction importance of each wavelength (350 to 2,500 nm) to the RF algorithm based on the BCRF (needles; CP = green; BCRF) and HCRF (whole seedling; OC = pink; HCRF). Gray shadow represents the average reflectance factor of the spectral curve. The spectra underwent the SNV transformation.

There is also growing evidence that hyperspectral data contain information about inter- and intraspecific genetic variability [42,60] and the phylogenetic history of plant lineages [71]. In the present study, we demonstrate the ability of hyperspectral reflectance to distinguish individual populations of Scots pine seedlings. Two ways of measuring reflectance factors were compared—leaf/needle—BCRF by CP and whole seedling HCRF measurements by OC. The performance of ML classifiers (RF and SVM) was tested. Our goal was to investigate the benefits and limitations of BCRF and HCRF for the practical deployment of seedlings. We aimed to categorize them into similar groups to facilitate effective afforestation across the sites with varying ecological requirements without the risk of potential maladaptation.

Differentiation of seedling populations based on leaf BCRF and seedling HCRF

The BCRF revealed a consistent, statistically significant ($\alpha = 0.05$) difference among all studied populations across the measured spectral range (Fig. 2). The differences among populations were the greatest in the VIS part of the spectra in the regions related to photosynthetic pigments content. The significance peaked in the RE in all pairwise population comparisons. We interpret the maximum in BCRF interpopulation difference as evidence for the genetic background (local adaptation), supported by the peak of heritable variation in needle spectral reflectance in RE reported by Čepel et al. [72].

The studied populations differed significantly in BCRF also in NIR and SWIR regions. Considering the contact measurements of needle stacks flattened on a spectrally black background, their optical properties in NIR are mainly influenced by their internal anatomical structure. The volume and surface of intercellular spaces in mesophyll determine the multiple scattering and affect the leaf-level reflectance, as shown for red spruce and eastern hemlock [73]. The reason for BCRF detecting the population differences using the NIR may be based on the differences in their mesophyll structures, which should be addressed by needle anatomical analysis in the future. The dominant factor for leaf optical properties in the SWIR spectral region is the water content [74] and structural polymers. Schwanninger et al. [75] summarized spectral wavelengths, or spectral range, corresponding to lignin, cellulose, or other aromatics: 1,170, 1,143, 1,188 to 1,195 nm, 2,280 to 2,340 nm, and 2,330 nm [75], same as [33,76] and [77]. These specific wavelengths also exhibited significant differences among the populations in our study.

Similar to the pigment content variability, plant water management variation is connected to local adaptations under genetic control [78]. Therefore, the interpopulation differences in the SWIR reflectance region could be associated with genetically based differences in plant water management and structural compound accumulation. We have not conducted destructive analyses to measure the needle functional traits because the seedlings were used in a further phenotyping experiment.

The whole seedling HCRF acquired by the OC yielded similar pairwise differences between seedling populations to the leaf BCRF from CP. In the case of the average HCRF, all seedling populations differed significantly in pairwise comparisons ($\alpha = 0.05$), similarly to BCRF, across the VSWIR. HCRF pairwise comparisons of seedling populations in SWIR1 and SWIR2 showed much less significance than BCRF. In longer SWIR2, the HCRF significance of pairwise comparisons ($P \times D$ and

$T \times D$) exhibited pronounced amplitude. However, the P value showed stronger fluctuation in SWIR2 compared to shorter wavelengths, probably due to the relatively low signal intensity as compared to the noise level.

The influence of water vapor still needs to be considered by excluding the noisiest NIR spectral bands [59,61]. As needle water content strongly alters the shape and magnitude of the reflectance curve in NIR and SWIR [74,79,80], the information about water content is lost if the spectral bands were removed from our dataset (especially near the 1,390 and 1,890 nm). It was shown that seedlings from different Scots pine provenances vary in their drought response and water management [81–83]. Thus, excluding noisy water absorption spectral regions may limit the distinction power among provenances. However, other informative water features could separate the ecotypes/populations in the 850- to 1,500-nm region. These limitations of field-measured HCRF may be a constraint for breeding application. Several growth traits (height, diameter, and crown length) in Norway spruce correlate strongly with needle reflectance in NIR and SWIR spectral regions, potentially useful in indirect selection [84].

To summarize, BCRF and HCRF were useful in determining the differences among the represented seedling populations in a common garden experiment. Except for the described differences between the 2 applied reflectance factor acquisition methods, both approaches share some similarities in the interpretation: mean BCRF and HCRF in the VIS spectral region of the D population were lower than T and P. The lower VIS mean reflectance of the D population could be attributed to the higher absorption. The D population might not be well adapted to our trial's conditions corresponding to lowland climatic and environmental conditions, but its BCRF and HCRF did not indicate lower chlorophyll content. The D population's lower average height and higher mortality compared to T and P populations could result from limited acclimation to trial conditions and provenance-specific growth rate (all 3 populations are significantly different in height, despite the relatively high variation within the populations).

In contrast, in the latter part of NIR and within SWIR, the BCRF and HCRF of Děčín seedlings were comparably higher, indicating variable water content that may be connected with the maladaptation. On a European scale, a trade-off was shown between height growth and stress tolerance in Scots pine (stress-resistant southern provenances with slower growth [82]). Thus, it is difficult for breeders to select elite offspring based on height at the juvenile stage, as Seidel and Menzel [81] demonstrated that taller and heavier Scots pine seedlings were more prone to drought-induced mortality. The main reason is the significantly different growth dynamics between ecotypes and generally low correlations between juvenile material and the rotation age of tree species [85].

Impact of canopy biomass on reflectance factor measurement

Seedling biomass, morphology, and needle density or branching are additional features affecting reflectance factors measurements. On the one hand, when measuring BCRF, the needles were measured flattened against the black plate background; therefore, the effect of crown morphology was removed. On the other hand, HCRF was measured from above on the intact seedlings; thus, the leaf angle distribution of the needles and their density affected the reflectance. These influences

are visible in provenance comparison and in ML predictions. Adult conifers have apparent differentiation into sun and shade needles according to the light exposure given by the vertical position within the canopy [86,87]. However, the seedling height was too small for needles to differentiate in this way.

Prediction of the population origin based on the reflectance factor using ML

On the basis of the RF classification performance comparing the BCRF and the HCRF superior results, it can be concluded that the seedling morphology, which influences the HCRF data, contributes remarkably to classification accuracy. RF predictions resulted in higher accuracies from data with high variance and data that included information about seedling morphological differences. When these parameters are controlled or limited, as in the leaf-level CP data, the probability of correct origin prediction decreases. Therefore, the data obtained from whole seedlings by the OC (HCRF) yielded higher accuracy in predicting the 3 Scots pine populations.

Several machine learning algorithms have been previously used in distinguishing plant groups based on genetic variation at the subspecies level. Cavender-Bares et al. [42] applied partial least-squares discriminant analysis on BCRF to distinguish single oak species' local populations. The same algorithm proved efficient in classifying hybrids between 2 cooccurring species of *Dryas* [60]. An SVM algorithm has been used for a similar case study to the present one, distinguishing *Tamarix* interspecific hybrids and pure species [59]. The classification performance of SVM was tested on both BCRF and HCRF acquired on *Tamarix* cuttings. However, the overall producers' classification accuracies differed slightly (92% and 96% for BCRF and HCRF, respectively) [59]. The author also tested another ML algorithm, stochastic gradient boosting, with comparably high classification accuracies (91% and 93% for BCRF and HCRF, respectively) and claimed that stochastic gradient boosting is more suitable for their small sample numbers than the RF classifier.

The BCRF and HCRF differences in their prediction accuracy were described above and were further documented by the "importance values" (i.e., weighting factors) for each wavelength inferred from RF. Unfortunately, the information about wavelengths contributing significantly to the classification model is unavailable for every ML algorithm, including SVM. In RF classification using HCRF, the importance of each wavelength was evaluated and rated relatively highly across all spectral regions (including NIR and SWIR). In contrast, the importance peaks based on BCRF were limited to the VIS and RE regions. Čepel et al. [72] described the importance of the RE in the context of genetic variation, which complies with our findings.

Similarly, Cavender-Bares et al. [42] reported important wavelengths for classifying oak populations using partial least-squares discriminant analysis preferentially in the VIS and RE; the NIR and SWIR spectral regions also contributed to the model but less. This difference mentioned above between BCRF and HCRF may be connected to the variable prediction accuracy of RF models. Specifically, the importance peaks in the vicinity of 1,200 nm are unique for the HCRF data, and they might have been instrumental in the prediction of the D population (Fig. 5). Interestingly, this peak of importance coincides with the increasing *P* values in pairwise comparisons between D and P and between D and T populations, respectively.

The classification of the various tree populations at the subspecies level (populations of a pure species and their interspecific

hybrids) based on reflectance has been recently reported [42,59,88,89]. However, predicting the specific population/provenance could be beneficial for selecting suitable individuals for breeding or reforestation. This classification should be done nondestructively concerning plant functional traits. The benefits of leaf-level spectroscopy are clearly the easy operation and relatively high throughput without sacrificing the measurement precision. Moreover, the proposed methods (OC/CP) do not rely on complex image segmentation and pixel classification connected with image-based spectroscopy. Although genomics in many tree species is still limited [90], genotyping using single-nucleotide polymorphisms has recently become more accessible for plant biologists and tree breeders. Currently, spatial genetics capitalizes on using spectral information about large and narrow scales [91]. Its ecology utilization has been predicted over a decade ago [92]. Genetic evaluation of hyperspectral data for breeding purposes was already efficiently combined with genotyping efforts [93,94]. Combining these 2 methods will accelerate the selection of seedlings with the required adaptation potential to the local environment [95].

Limitations

Limitations inherent to each method may potentially affect the research findings and provide opportunities to improve in future investigations; thus, they are discussed in the following section.

Measuring BCRF by a CP equipped with an internal light source reduced the noise caused by ambient light conditions and atmospheric aerosol variation. The time spent measuring each sample should be adjusted to the plant material as the heat emitted from the light source could damage the younger leaves, especially when repeated measurements are needed. This drawback is particularly critical for studies dealing with induced drought stress. The other limitation of using CP in conifer seedlings is the risk of apical bud damage by the manipulation. The reason for slowing down the whole measurement process is the need to place a sufficient number of needles to cover the field of view, which will determine the quality and strength of the signal obtained. Considering the contact measurements of needle stacks flattened on a spectrally black background, their optical properties in NIR are mainly influenced by their internal anatomical structure. Therefore, the effect of crown morphology was removed. Measuring by CP is less dependent on the current conditions, its white reference is more stable, and measuring is comparable within the time.

HCRF measured by bare cable in the ambient illumination strongly depends on the current weather conditions. In contrast to CP-acquired BCRF, proximal OC measurements are less time-consuming when light conditions are stable. Thus, we can expect more consistent results in a shorter time on clear days and close to solar noon, reducing the variability given by changing light intensity and geometry. Aerosol absorption in the spectral ranges (especially near 1,390 and 1,890 nm) varied depending on the current conditions. These specific ranges were removed at the cost of losing water content information. Noncontact measuring also depends on the cable's geometry, which the CP stabilizes. HCRF was measured from above the level of the seedlings; thus, the reflectance was affected by the angular distribution of needles and their density. Since the proximal measurement of whole seedlings depends on ambient light, the temporal changes under light conditions influence the predictions. To some extent, the effect of differences in light geometry and intensity is

reduced by frequent white panel normalization and timing of the measurements only close to solar noon [59]. Measuring by OC also yields a weaker signal than measuring by CP.

Despite the limitations inherent to each reflectance acquisition method, they both proved viable for the manual HTP of Scots pine seedlings. We can hypothesize that the day-to-day differences in ML predictions were mainly due to the change in incoming illumination. However, the variation between days also affected CP-based ML prediction accuracies. In this case, it could most likely be attributed to the sampling effect—we measured a different subset of seedlings daily.

In the present study, we demonstrated that progeny populations of 3 different parent populations of Scots pine could be distinguished using their reflectance factors (BCRF and HCRF). The limitation of our study lies in the absence of supporting biophysical measurements of needle pigments, dry matter, water content, and morphological (needle density) seedling traits determined by traditional plant physiology approaches. The study was conducted on the seedlings designated for a drought resistance experiment; thus, the seedlings could not be sampled. However, we obtained the precise height of each seedling, and the populations' progenies differ significantly in mean height. However, several similarly designed studies have shown that spectral reflectance may result in better classification than multiple leaf traits [42,89].

Conclusion

The present study supports the emerging evidence that leaf and canopy reflectance spectra are useful to distinguish interpopulation variability and thus represent a valuable tool in high throughput plant phenotyping. Using reflectance factors and ML algorithms to detect various conifer seedling genotypes in a common garden or nursery experiment is a novel approach that could offer a vital tool for future effective nursery practices.

The present study was conducted on Scots pine seedlings, ecologically and economically important forest species with an extremely large spatial distribution and phenotype plasticity. Both reflectance factor measurement approaches proved viable in detecting differences among seedlings from 3 local populations (represented by 3 orchards), separating the mean BCRF and HCRF of the Děčín population that were well differentiated from Třeboň and Plasy in all cases. The classification models based on RF and SVM algorithms using HCRF that also included canopy structural information resulted in a higher probability of correct prediction of the respective population than BCRF. HCRF contactless measurement was faster and included morphological information, leading to higher prediction models' accuracy. The proximal data obtained from the top of the seedlings offered over 80% accuracy in predicting 3 different Scots pine populations using RF and SVM algorithms.

Hyperspectral reflectance is gaining importance for breeding programs as measuring reflectance factors can be used for simple and nondestructive genetic evaluation. The present study's take-home message is that manual hyperspectral phenotyping could provide an efficient tool in tree nurseries when a phenotyping unit is not affordable.

Acknowledgments

Dr. Jan Kaňák, Arboretum Sofronka, Pilsen.

Funding: This research was funded mainly by the Ministry of Education, Youth and Sports of the Czech Republic, scheme

INTER-EXCELLENCE, INTER-ACTION, grant number LTA-USA19113, titled “Genetic variability of hyper-spectral reflectance in Scots pine ecotypes for selection of drought-resistant individuals”. This project was coordinated with U.S. partners: P. Campbell and J. Brawner. This research received funding from the European Union's Horizon Europe Research and Innovation Programme under grant agreement no.: 101081774 — OptFORESTS. P.C.'s contribution was supported by NASA, LCLUC Program NNH17ZDA001N-LCLUC, grant no: 80NSS-C18K0337, titled “Prototyping MuSLI canopy chlorophyll content for assessment of vegetation function and productivity”. M.K. was supported by the Academy of Finland Flagship on Photonics Research and Innovation (PREIN, 320166).

Author contributions: Conceptualization: J.S. with Z.L. Methodology: J.S., Z.L., J.C., D.P., and J.Č. Statistical analysis: J.Č. and J.S. Formal analysis: J.S., J.Č., J.C., and D.P. Investigation: J.S., J.Č., J.C., Z.L., and E.N. Resources: M.L., J.A., and P.C. Writing—original draft preparation: J.S., J.C., J.Č., E.N., Z.L., M.K., and O.O.A. Writing—review and editing: J.S., J.Č., E.N., Z.L., J.A., M.K., and P.C. Visualization, J.C. Supervision: J.S., Z.L., and J.A. Project administration: J.S., Z.L., M.L., M.K., and J.A. Funding acquisition: J.S., M.L., J.A., and P.C. All authors have read and agreed to the published version of the manuscript.

Competing interests: The authors declare that they have no competing interests.

Data Availability

The complete dataset is freely available upon request.

Supplementary Materials

Fig. S1. Change of importance values after SNV transformation in BCRF and HCRF.

References

1. Allen CD, Macalady AK, Chenchouni H, Bachelet D, McDowell N, Vennetier M, Kitzberger T, Rigling A, Breshears DD, Hogg EH, et al. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For Ecol Manag.* 2010;259(4):660–684.
2. Dyderski MK, Paź S, Frelich LE, Jagodziński AM. How much does climate change threaten European forest tree species distributions? *Glob Chang Biol.* 2018;24(3):1150–1163.
3. Ostonen I, Löhmus K, Helmisaari H-S, Truu J, Meel S. Fine root morphological adaptations in Scots pine, Norway spruce and silver birch along a latitudinal gradient in boreal forests. *Tree Physiol.* 2007;27(11):1627–1634.
4. Uriá-Diez J, Pommerening A. Crown plasticity in Scots pine (*Pinus sylvestris* L.) as a strategy of adaptation to competition and environmental factors. *Ecol Model.* 2017;356:117–126.
5. Palmroth S, Berninger F, Nikinmaa E, Lloyd J, Pulkkinen P, Hari P. Structural adaptation rather than water conservation was observed in Scots pine over a range of wet to dry climates. *Oecologia.* 1999;121:302–309.
6. Ranade SS, García-Gil MR. Ecotypic variation in response to light spectra in Scots pine (*Pinus sylvestris* L.). *Tree Physiol.* 2013;33(2):195–201.
7. Poorter L, Castilho CV, Schiatti J, Oliveira RS, Costa FRC. Can traits predict individual growth performance? A test in a hyperdiverse tropical forest. *New Phytol.* 2018;219(1):109–121.

8. Araus JL, Cairns JE. Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* 2014;19(1):52–61.
9. Mir RR, Reynolds M, Pinto F, Khan M, Bhat M. High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci.* 2019;282:60–72.
10. Mazis A, Choudhury SD, Morgan PB, Stoerger V, Hiller J, Ge Y, Awada T. Application of high-throughput plant phenotyping for assessing biophysical traits and drought response in two oak species under controlled environment. *For Ecol Manag.* 2020;465:Article 118101.
11. Bian L, Zhang H, Ge Y, Čepel J, Stejskal J. Closing the gap between phenotyping and genotyping: Review of advanced, image-based phenotyping technologies in forestry. *Ann For Sci.* 2022;79:22.
12. Reynolds M, Chapman S, Crespo-Herrera L, Molero G, Mondal S, Pequeno DNL, Pinto F, Pinera-Chavez FJ, Poland J, Rivera-Amado C, et al. Breeder friendly phenotyping. *Plant Sci.* 2020;295:Article 110396.
13. Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, Xiong L, Yan J. Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol Plant.* 2020;13(2):187–214.
14. Sims DA, Gamon JA. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens Environ.* 2002;81(2–3):337–354.
15. le Maire G, François C, Dufrêne E. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sens Environ.* 2004;89:1–28.
16. Gitelson AA, Solovchenko A. Non-invasive quantification of foliar pigments: Possibilities and limitations of reflectance- and absorbance-based approaches. *J Photochem Photobiol B Biol.* 2018;178(1):537–544.
17. Granlund L, Keski-Saari S, Kumpulainen T, Oksanen E, Keinänen M. Imaging lichen water content with visible to mid-wave infrared (400–5500 nm) spectroscopy. *Remote Sens Environ.* 2018;216:301–310.
18. Li H, Yang W, Lei J, She J, Zhou X. Estimation of leaf water content from hyperspectral data of different plant species by using three new spectral absorption indices. *PLOS ONE.* 2021;16(3):Article e0249351.
19. Martin ME, Aber JD. High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. *Ecol Appl.* 1997;7(2):431–443.
20. Buschmann C, Lenk S, Lichtenthaler HK. Reflectance spectra and images of green leaves with different tissue structure and chlorophyll content. *Isr J Plant Sci.* 2012;60(1–2):49–64.
21. Neuwirthová E, Lhotáková Z, Albrechtová J. The effect of leaf stacking on leaf reflectance and vegetation indices measured by contact probe during the season. *Sensors.* 2017;17(6):1202.
22. Neuwirthová E, Lhotáková Z, Lukeš P, Albrechtová J. Leaf surface reflectance does not affect biophysical traits modelling from VIS-NIR spectra in plants with sparsely distributed trichomes. *Remote Sens.* 2021;13(20):4144.
23. Peñuelas J, Filella I. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends Plant Sci.* 1998;3(4):151–156.
24. Campbell PKE, Rock BN, Martin ME, Neefus CD, Irons JR, Middleton EM, Albrechtová J. Detection of initial damage in Norway spruce canopies using hyperspectral airborne data. *Int J Remote Sens.* 2004;25(24):5557–5584.
25. Campbell PKE, Middleton EM, McMurtrey JE, Corp LA, Chappelle EW. Assessment of vegetation stress using reflectance or fluorescence measurements. *J Environ Qual.* 2007;36(3):832–845.
26. Pérez V, Molero G, Serbin SP, Condon AG, Reynolds MP, Furbank RT, Evans JR. Hyperspectral reflectance as a tool to measure biochemical and physiological traits in wheat. *J Exp Bot.* 2017;69(3):483–496.
27. Tao H, Feng H, Xu L, Miao M, Long H, Yue J, Li Z, Yang G, Yang X, Fan L, et al. Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data. *Sensors.* 2020;20(5):1296.
28. Campbell PKE, Huemmrich KF, Middleton EM, Ward LA, Julitta T, Daughtry CST, Burkart A, Russ AL, Kustas WP. Diurnal and seasonal variations in chlorophyll fluorescence associated with photosynthesis at leaf and canopy scales. *Remote Sens.* 2019;11(5):488.
29. Neuwirthová E, Kuusk A, Lhotáková Z, Kuusk J, Albrechtová J, Hallik L. Leaf age matters in remote sensing: Taking ground truth for spectroscopic studies in hemiboreal deciduous trees with continuous leaf formation. *Remote Sens.* 2021;13(7):1353.
30. Korecký J, Čepel J, Stejskal J, Faltinová Z, Dvořák J, Lstibůrek M, El-Kassaby YA. Genetic diversity of Norway spruce ecotypes assessed by GBS-derived SNPs. *Sci Rep.* 2021;11(1):23119.
31. Jansen S, Konrad H, Geburek T. The extent of historic translocation of Norway spruce forest reproductive material in Europe. *Ann For Sci.* 2017;74:56.
32. Haagsma M, Page GFM, Johnson JS, Still C, Waring KM, Sniezko RA, Selker JS. Using hyperspectral imagery to detect an invasive fungal pathogen and symptom severity in *Pinus strobiformis* seedlings of different genotypes. *Remote Sens.* 2020;12(24):4041.
33. Curran PJ. Remote sensing of foliar chemistry. *Remote Sens Environ.* 1989;30(3):271–278.
34. Gitelson AA, Merzlyak MN. Remote estimation of chlorophyll content in higher plant leaves. *Int J Remote Sens.* 1997;18(12):2691–2697.
35. Gates DM, Keegan HJ, Schleter JC, Weidner VR. Spectral properties of plants. *Appl Opt.* 1965;4(1):11–20.
36. Croft H, Chen JM, Zhang Y. The applicability of empirical vegetation indices for determining leaf chlorophyll content over different leaf and canopy structures. *Ecol Complex.* 2014;17:119–130.
37. Eitel JUH, Gessler PE, Smith AMS, Robberecht R. Suitability of existing and novel spectral indices to remotely detect water stress in *Populus* spp. *For Ecol Manag.* 2006;229(1–3):170–182.
38. Gitelson AA, Merzlyak MN, Lichtenthaler HK. Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *J Plant Physiol.* 1996;148(3–4):501–508.
39. Rock B, Vogelmann J, Williams D, Vogelmann A, Hoshizaki T. Remote detection of forest damage: Plant responses to stress may have spectral “signatures” that could be used to map, monitor, and measure forest damage. *Bioscience.* 1986;36(7):439–445.
40. Martin ME, Plourde LC, Ollinger SV, Smith M-L, McNeil BE. A generalizable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. *Remote Sens Environ.* 2008;112(9):3511–3519.
41. Kokaly RF, Asner GP, Ollinger SV, Martin ME, Wessman CA. Characterizing canopy biochemistry from imaging

- spectroscopy and its application to ecosystem studies. *Remote Sens Environ.* 2009;113(Suppl. 1):S78–S91.
42. Cavender-Bares J, Meireles JE, Couture JJ, Kaproth MA, Kingdon CC, Singh A, Serbin SP, Center A, Zuniga E, Pilz G, et al. Associations of leaf spectra with genetic and phylogenetic variation in oaks: Prospects for remote detection of biodiversity. *Remote Sens.* 2016;8(3):221.
 43. Claverie M, Ju J, Masek JG, Dungan JL, Vermote EF, Roger J-C, Skakun SV, Justice C. The harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens Environ.* 2018;219:145–161.
 44. Vanbrabant Y, Tits L, Delalieux S, Pauly K, Verjans W, Somers B. Multitemporal chlorophyll mapping in pome fruit orchards from remotely piloted aircraft systems. *Remote Sens.* 2019;11(12):1468.
 45. Einzmann K, Atzberger C, Pinnel N, Glas C, Böck S, Seitz R, Immitzer M. Early detection of spruce vitality loss with hyperspectral data: Results of an experimental study in Bavaria, Germany. *Remote Sens Environ.* 2021;266: Article 112676.
 46. Einzmann K, Ng W-T, Immitzer M, Pinnel N, Atzberger C. Method analysis for collecting and processing in-situ hyperspectral needle reflectance data for monitoring Norway spruce & BR> Methodenanalyse zur Erfassung und Prozessierung hyperspektraler in-situ Nadelreflexionsdaten zum monitoring von Fichten. *Photogramm Fernerkun Geoinform.* 2014;2014:423–434.
 47. Martinez NE, Sharp JL, Kuhne WW, Johnson TE, Stafford CT, Duff MC. Assessing the use of reflectance spectroscopy in determining CsCl stress in the model species *Arabidopsis thaliana*. *Int J Remote Sens.* 2015;36(23):5887–5915.
 48. Lhotáková Z, Kopačková-Strnadová V, Oulehle F, Homolová L, Neuwirthová E, Švik M, Janoutová R, Albrechtová J. Foliage biophysical trait prediction from laboratory spectra in Norway spruce is more affected by needle age than by site soil conditions. *Remote Sens.* 2021;13(3):391.
 49. Kupková L, Potůčková M, Buřičová M, Kopačková V, Lhotáková Z, Albrechtová J. Determination of lignin content in Norway spruce foliage using Nir spectroscopy and hyperspectral data. Paper presented at: Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium; 2012 July 22–27; Munich, Germany.
 50. Rajewicz PA, Atherton J, Alonso L, Porcar-Castell A. Leaf-level spectral fluorescence measurements: Comparing methodologies for broadleaves and needles. *Remote Sens.* 2019;11(5):532.
 51. Malenovský Z, Albrechtová J, Lhotáková Z, Zurita-Milla R, Clevers JGPW, Schaepman ME, Cudlín P. Applicability of the PROSPECT model for Norway spruce needles. *Int J Remote Sens.* 2006;27(24):5315–5340.
 52. Hovi A, Möttöus M, Juola J, Manoocheri F, Ikonen E, Rautiainen M. Evaluating the performance of a double integrating sphere in measurement of reflectance, transmittance, and albedo of coniferous needles. *Silva Fenn.* 2020;54(2):10270.
 53. Schaepman-Strub G, Schaepman ME, Painter TH, Dangel S, Martonchik JV. Reflectance quantities in optical remote sensing—Definitions and case studies. *Remote Sens Environ.* 2006;103(1):27–42.
 54. Melendo-Vega J, Martín M, Pacheco-Labrador J, González-Cascón R, Moreno G, Pérez F, Migliavacca M, García M, North P, Riaño D. Improving the performance of 3-D radiative transfer model FLIGHT to simulate optical properties of a tree-grass ecosystem. *Remote Sens.* 2018;10(12):2061.
 55. Slaton MR, Smith WK. Mesophyll architecture and cell exposure to intercellular air space in alpine, desert, and Forest species. *Int J Plant Sci.* 2002;163:937–948.
 56. Lukeš P, Neuwirthová E, Lhotáková Z, Janoutová R, Albrechtová J. Upscaling seasonal phenological course of leaf dorsiventral reflectance in radiative transfer model. *Remote Sens Environ.* 2020;246:Article 111862.
 57. Cao L, Coops NC, Innes JL, Dai J, Ruan H, She G. Tree species classification in subtropical forests using small-footprint full-waveform LiDAR data. *Int J Appl Earth Obs Geoinf.* 2016;49:39–51.
 58. Zhang Y, Wang J, Wu Z, Lian J, Ye W, Yu F. Tree species classification using plant functional traits and leaf spectral properties along the vertical canopy position. *Remote Sens.* 2022;14(24):6227.
 59. Tesfamichael SG, Newete SW, Adam E, Byrne MJ. Discriminating pure *Tamarix* species and their putative hybrids using field spectrometer. *Geocarto Int.* 2021;37(25):7733–7752.
 60. Stasinski L, White DM, Nelson PR, Ree RH, Meireles JE. Reading light: Leaf spectra capture fine-scale diversity of closely related, hybridizing arctic shrubs. *New Phytol.* 2021;232(6):2283–2294.
 61. Sun H, Feng M, Xiao L, Yang W, Wang C, Jia X, Zhao Y, Zhao C, Muhammad SK, Li D. Assessment of plant water status in winter wheat (*Triticum aestivum* L.) based on canopy spectral indices. *PLOS ONE.* 2019;14(6):e0216890.
 62. Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl Spectrosc.* 1989;43:772–777.
 63. Butler DG, Cullis BR, Gilmour AR, Gogel BJ, Thompson R. ASReml estimates variance components under a general linear mixed model by residual maximum likelihood (REML). In: *ASReml-R reference manual* Hemel Hempstead, UK: VSN International Ltd.; 2018.
 64. Liaw A, Wiener M. Classification and regression by random forest. *R news.* 2002;2:18–22.
 65. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C-C, Lin C-C. *Misc functions of the Department of Statistics, Probability Theory Group (formerly: E1071)* TU Wien; 2021.
 66. Sanaeifar A, Zhu F, Sha J, Li X, He Y, Zhan Z. Rapid quantitative characterization of tea seedlings under lead-containing aerosol particles stress using Vis-NIR spectra. *Sci Total Environ.* 2022;802:Article 149824.
 67. Poona NK, van Niekerk A, Nadel RL, Ismail R. Random forest (RF) wrappers for waveband selection and classification of hyperspectral data. *Appl Spectrosc.* 2016;70(2):322–333.
 68. Pandey P, Payn KG, Lu Y, Heine AJ, Walker TD, Acosta JJ, Young S. Hyperspectral imaging combined with machine learning for the detection of fusiform rust disease incidence in loblolly pine seedlings. *Remote Sens.* 2021;13(18):3595.
 69. Verrelst J, Malenovský Z, Van der Tol C, Camps-Valls G, Gastellu-Etchegorry J-P, Lewis P, North P, Moreno J. Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods. *Surv Geophys.* 2019;40:589–629.
 70. Haagsma M, Page GFM, Johnson JS, Still C, Waring KM, Snieszko RA, Selker JS. Model selection and timing of

- acquisition date impacts classification accuracy: A case study using hyperspectral imaging to detect white pine blister rust over time. *Comput Electron Agric.* 2021;191:Article 106555.
71. Meireles JE, Cavender-Bares J, Townsend PA, Ustin S, Gamon JA, Schweiger AK, Schaepman ME, Asner GP, Martin RE, Singh A, et al. Leaf reflectance spectra capture the evolutionary history of seed plants. *New Phytol.* 2020;228(2):485–493.
 72. Čepl J, Stejskal J, Lhotáková Z, Holá D, Korecký J, Lstibůrek M, Tomášková I, Kočová M, Rothová O, Palovská M, et al. Heritable variation in needle spectral reflectance of Scots pine (*Pinus sylvestris* L.) peaks in red edge. *Remote Sens Environ.* 2018;219:89–98.
 73. Rock BN, Williams DL, Moss DM, Lauten GN, Kim M. High-spectral resolution field and laboratory optical reflectance measurements of red spruce and eastern hemlock needles and branches. *Remote Sens Environ.* 1994;47:176–189.
 74. Jacquemoud S, Baret F. PROSPECT: A model of leaf optical properties spectra. *Remote Sens Environ.* 1990;34:75–91.
 75. Schwanninger M, Rodrigues JC, Fackler K. A review of band assignments in near infrared spectra of wood and wood components. *J Near Infrared Spectrosc.* 2011;19:287–308.
 76. Serrano L, Peñuelas J, Ustin SL. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sens Environ.* 2002;81(2–3):355–364.
 77. Soukupova J, Rock BN, Albrechtova J. Spectral characteristics of lignin and soluble phenolics in the near infrared - a comparative study. *Int J Remote Sens.* 2002;23:3039–3055.
 78. Semerci A, Semerci H, Çalişkan B, Çiçek N, Ekmekçi Y, Mencuccini M. Morphological and physiological responses to drought stress of European provenances of Scots pine. *Eur J Forest Res.* 2017;136:91–104.
 79. Carter GA. Primary and secondary effects of water content on the spectral reflectance of leaves. *Am J Bot.* 1991;78:916–924.
 80. Sims DA, Gamon JA. Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: A comparison of indices based on liquid water and chlorophyll absorption features. *Remote Sens Environ.* 2003;84(4):526–537.
 81. Seidel H, Menzel A. Above-ground dimensions and acclimation explain variation in drought mortality of scots pine seedlings from various provenances. *Front Plant Sci.* 2016;7:1014.
 82. Seidel H, Schunk C, Matiu M, Menzel A. Diverging drought resistance of scots pine provenances revealed by infrared thermography. *Front Plant Sci.* 2016;7:1247.
 83. Seidel H, Matiu M, Menzel A. Compensatory growth of Scots pine seedlings mitigates impacts of multiple droughts within and across years. *Front Plant Sci.* 2019;10:519.
 84. Hejtmánek J, Stejskal J, Čepl J, Lhotáková Z, Korecký J, Krejzková A, Dvořák J, Gezan SA. Revealing the complex relationship among hyperspectral reflectance, photosynthetic pigments, and growth in Norway spruce ecotypes. *Front Plant Sci.* 2022;13:Article 721064.
 85. Lambeth CC. Juvenile-mature correlations in Pinaceae and implications for early selection. *For Sci.* 1980;26(4):571–580.
 86. Niinemets Ü. A review of light interception in plant stands from leaf to canopy in different plant functional types and in species with varying shade tolerance. *Ecol Res.* 2010;25:693–714.
 87. Kubínová Z, Janáček J, Lhotáková Z, Šprtová M, Kubínová L, Albrechtová J. Norway spruce needle size and cross section shape variability induced by irradiance on a macro- and microscale and CO₂ concentration. *Trees.* 2018;32: 231–244.
 88. Madritch MD, Kingdon CC, Singh A, Mock KE, Lindroth RL, Townsend PA. Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales. *Philos Trans R Soc Lond B Biol Sci.* 2014;369(1643):20130194–20130194.
 89. Deacon NJ, Grossman JJ, Schweiger AK, Armour I, Cavender-Bares J. Genetic, morphological, and spectral characterization of relictual Niobrara River hybrid aspens (*Populus × smithii*). *Am J Bot.* 2017;104(12):1878–1890.
 90. Grattapaglia D. Breeding forest trees by genomic selection: Current progress and the way forward. In: Tuberosa R, Graner A, Frison E, editors *Genomics of Plant Genetic Resources: Volume 1* Dordrecht, the Netherlands: Managing, Sequencing and Mining Genetic Resources, Springer; 2014. p. 651–682.
 91. Czyż EA, Guillén Escribà C, Wulf H, Tedder A, Schuman MC, Schneider FD, Schaepman ME. Intraspecific genetic variation of a *Fagus sylvatica* population in a temperate forest derived from airborne imaging spectroscopy time series. *Ecol Evol.* 2020;10:7419–7430.
 92. Rocchini D, Balkenhol N, Carter GA, Foody GM, Gillespie TW, He KS, Kark S, Levin N, Lucas K, Luoto M, et al. Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Eco Inform.* 2010;5:318–329.
 93. Rincet R, Charpentier J-P, Faivre-Rampant P, Paux E, Le Gouis J, Bastien C, Segura V. Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. *G3.* 2018;8(12):3961–3972.
 94. Krause MR, González-Pérez L, Crossa J, Pérez-Rodríguez P, Montesinos-López O, Singh RP, Dreisigacker S, Poland J, Rutkoski J, Sorrells M, et al. Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3.* 2019;9(4):1231–1247.
 95. Grubinger S, Coops NC, O'Neill GA. Picturing local adaptation: Spectral and structural traits from drone remote sensing reveal clinal responses to climate transfer in common-garden trials of interior spruce (*Picea engelmannii* × *glauca*). *Glob Chang Biol.* 2023;29(17):4842–4860.

Making the Genotypic Variation Visible: Hyperspectral Phenotyping in Scots Pine Seedlings

Jan Stejskal, Jaroslav Čepl, Eva Neuwirthová, Olusegun Olaitan Akinyemi, Jiří Chuchlík, Daniel Provazník, Markku Keinänen, Petya Campbell, Jana Albrechtová, Milan Lštibůrek, and Zuzana Lhotáková

Citation: Stejskal J, #epl J, Neuwirthová E, Akinyemi O, Chuchlík J, Provazník D, Keinänen M, Campbell P, Albrechtová J, Lstib#rek M, et al. Making the Genotypic Variation Visible: Hyperspectral Phenotyping in Scots Pine Seedlings. *Plant Phenomics*. 2023;5:0111. DOI: 10.34133/plantphenomics.0111

Hyperspectral reflectance contains valuable information about leaf functional traits, which can indicate a plant's physiological status. Therefore, using hyperspectral reflectance for high-throughput phenotyping of foliar traits could be a powerful tool for tree breeders and nursery practitioners to distinguish and select seedlings with desired adaptation potential to local environments. We evaluated the use of 2 nondestructive methods (i.e., leaf and proximal/canopy) measuring hyperspectral reflectance in the 350- to 2,500-nm range for phenotyping on 1,788 individual Scots pine seedlings belonging to lowland and upland ecotypes of 3 different local populations from the Czech Republic. Leaf-level measurements were collected using a spectroradiometer and a contact probe with an internal light source to measure the biconical reflectance factor of a sample of needles placed on a black background in the contact probe field of view. The proximal canopy measurements were collected under natural solar light, using the same spectroradiometer with fiber optical cable to collect data on individual seedlings' hemispherical conical reflectance factor. The latter method was highly susceptible to changes in incoming radiation. Both spectral datasets showed statistically significant differences among Scots pine populations in the whole spectral range. Moreover, using random forest and support vector machine learning algorithms, the proximal data obtained from the top of the seedlings offered up to 83% accuracy in predicting 3 different Scots pine populations. We conclude that both approaches are viable for hyperspectral phenotyping to disentangle the phenotypic and the underlying genetic variation within Scots pine seedlings.

Image

View the article online

<https://spj.science.org/doi/10.34133/plantphenomics.0111>

Use of this article is subject to the [Terms of Service](#)

Plant Phenomics (ISSN 2643-6515) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005.

Copyright © 2023 Jan Stejskal et al.

Exclusive licensee Nanjing Agricultural University. No claim to original U.S. Government Works. Distributed under a [Creative Commons Attribution License 4.0 \(CC BY 4.0\)](#).