Computers and Electronics in Agriculture 199 (2022) 107169

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

# High-throughput screening of soybean di-nitrogen fixation and seed nitrogen content using spectral sensing

Johann Vollmann [a,*], Pablo Rischbeck [b], Martin Pachner [a], Vuk Đorđević [c], Ahmad M. Manschadi [d]

[a] University of Natural Resources and Life Sciences, Vienna, Department of Crop Sciences, Institute of Plant Breeding, Konrad Lorenz Str. 24, 3430 Tulln an der Donau, Austria
[b] University of Natural Resources and Life Sciences, Vienna, Department of Crop Sciences, Konrad Lorenz Str. 24, 3430 Tulln an der Donau, Austria
[c] Institute of Field and Vegetable Crops, Maksima Gorkog 30, 21101 Novi Sad, Serbia
[d] University of Natural Resources and Life Sciences, Vienna, Department of Crop Sciences, Institute of Agronomy, Konrad Lorenz Str. 24, 3430 Tulln an der Donau, Austria

ARTICLE INFO

*Keywords:*
Soybean
Phenotyping
Di-nitrogen fixation
Seed protein content
Spectral reflectance index

ABSTRACT

Symbiotic di-nitrogen fixation of grain legumes has a substantial impact on crop performance, harvest product quality, and nitrogen (N) balance of crop rotations, particularly under organic management regimes. In soybean breeding, selection for increased nitrogen fixation is desirable for improving seed protein content and N balance of cropping systems. However, the lack of high-throughput screening methods for direct measurement of $N_2$ fixation rates prohibits practical breeding efforts. Therefore, hyperspectral canopy reflectance measurement as a field-based phenotyping method was evaluated in three environments for indirect estimation of N fixation and uptake of soil nitrogen in a set of early maturity soybean genotypes exhibiting a wide range in seed protein content. Reflectance spectra were collected in repeated measurements during flowering and early seed filling stages. Subsequently, various spectral reflectance indices (SRIs) were calculated for characterizing nitrogen accumulation of individual genotypes. Moreover, prediction models for seed protein content as an end-of-season target trait were developed utilizing full spectral information in partial-least-square regression (PLSR) models. A number of N-related SRIs calculated from spectral reflectance data recorded at the beginning of the seed filling stage were significantly correlated to seed protein content. The best prediction of seed protein content, however, was achieved in PLSR models (validation $R^2 = 0.805$ across all three environments). Environments lower in initial soil mineral N content appeared as more favorable selection sites in terms of prediction accuracy, because N fixation is not masked by soil N uptake in such environments. Hyperspectral reflectance data proved to be a valuable method for determining genetic variation in crop N accumulation, which might be implemented in high-throughput screening protocols for N fixation in plant breeding programs.

## 1. Introduction

Fixation of atmospheric di-nitrogen in hypoxic root nodules is a unique feature of the endosymbiosis between rhizobial bacteria and plants of the *Fabaceae* family. Basic signaling processes between host plant root hairs and symbiotic bacteria as well as the regulatory cascade of nodule development involving lateral root growth are understood only recently (Bosse et al., 2021; Szczyglowski and Ross, 2021). Nevertheless, the symbiosis-mediated production of ureides/ammonia has a tremendous impact on agricultural systems utilizing legumes in crop rotations. Estimates of $N_2$ fixation range between 30 and 120 kg/ha/year for grain legumes from global areas, but fixation rates of individual crops might be considerably higher (Herridge et al., 2008). In soybean (*Glycine max* [L.] Merr.), experimental results on N fixation indicate that the percentage of nitrogen derived from the atmosphere (% Ndfa) is 68% on average (Herridge et al., 2008), whereas under growing conditions of Central Europe the % Ndfa was lower ranging between 24 and 57% (Oberson et al., 2007; Schweiger et al., 2012; Zimmer et al., 2016). This might be due to cooler soil temperatures, higher rates of soil mineral N as well as soybean genotype and bacterial inoculant strain effects. Thus, due to the high seed protein content of soybean, particularly in high yield environments, the balance between nitrogen fixation

* Corresponding author.
*E-mail address:* johann.vollmann@boku.ac.at (J. Vollmann).

https://doi.org/10.1016/j.compag.2022.107169
Received 25 February 2022; Received in revised form 18 June 2022; Accepted 21 June 2022
Available online 30 June 2022

per season and nitrogen export by the crop harvest might be negative (Salvagiotti et al., 2008). As nitrogen is the most important plant macronutrient determining growth and crop yield, this is particularly unfavorable under organic farming conditions, in which N fixation of legume crops is the key element of N supply within the cropping system. Thus, improving nitrogen fixation is considered as a paramount goal for grain legumes positively contributing to N balance under organic management (Reckling et al., 2020).

Although an improvement in N fixation capacity of soybean would be highly desirable for both increased yield performance and positive crop rotation effects, a direct measurement of N fixation utilizing methods such as $N^{15}$ isotope-based protocols, petiole ureide concentration or nitrogenase activity / acetylene reduction (Herridge et al., 2008; Schweiger et al., 2012) is not suitable for screening a large number of genotypes in plant breeding programs. Furthermore, sampling based on individual plants at a given time point would contribute to high variability in estimates of the nitrogen fixation rate. Therefore, instead of directly measuring symbiotic nitrogen fixation, high-throughput phenotyping methods (e.g. Araus et al., 2018) for predicting yield, above-ground biomass or leaf area index as well as chlorophyll content and N accumulation (Xie and Yang, 2020) could be adopted for indirect estimation of N fixation between genotypes differing in nodulation characteristics. In agronomic and plant breeding research with cereal crops, canopy N content and N-use efficiency were estimated effectively utilizing spectral reflectance data and vegetation indices calculated thereof (Banerjee et al., 2020; Feng et al., 2016; Lausch et al., 2015; Li et al., 2014; Palka et al., 2021; Prey et al., 2020). In soybean, spectral reflectance based prediction of total biomass, grain yield, time to maturity or disease incidence has been reported (Christenson et al., 2016; Crusiol et al., 2021; Herrmann et al., 2018; Moreira et al., 2021; Zhang et al., 2019). More specifically and in relation to nitrogen, Chen et al. (2019) demonstrated the estimation of soybean leaf nitrogen content under various shading conditions utilizing hyperspectral reflectance data. Chiozza et al. (2021) attempted to predict grain yield and seed protein content from spectral data collected at the R4 (Fehr and Caviness, 1977) developmental stage. They obtained better precision of predictions in protein content than in grain yield, oil content or other traits. Bi et al. (2018) presented the spectral measurement of soybean leaf N content depending on drought conditions and mycorrhizal inoculation. Earlier, Krezhova and Kirova (2011) had reported hyperspectral differences in N fixation status measured in soybean plants subjected to different salinity or UV-radiation stress treatments, whereas chlorophyll metering (SPAD meter readings) and leaf imaging protocols revealed varying predictability of leaf N, chlorophyll and seed protein content (Fritschi and Ray, 2007; Vollmann et al., 2011). In contrast to spectral reflectance indices (SRIs) which are calculated from reflectance data at a few given wavelength points only, partial-least-square regression (PLSR) models utilizing the whole range of hyperspectral information available have recently been utilized for the direct prediction of above-ground biomass, canopy and seed nitrogen concentration or grain yield (Chiozza et al., 2021; Crusiol et al., 2021; Li et al., 2014; Moreira et al., 2021; Prey et al., 2020).

Given the importance of legume symbiotic dinitrogen fixation, and as an efficient direct measurement of nitrogen fixation is not applicable for plant breeding set-ups with higher numbers of genotypes, it was hypothesized that screening of crop N status through hyperspectral reflectance measurement at the onset of the seed-filling stage might indicate nitrogen available for seed protein biosynthesis. Thus, the canopy N content acquired through both symbiotic $N_2$ fixation and soil-N uptake until the stage of measurement would reflect nitrogen accumulation over time, which might later be mirrored by seed protein content as an end-of-season trait. Therefore, the main objectives of the present research were to (i) evaluate spectral reflectance indices (SRIs) collected in small breeding plots for differentiating between genotypes in replicated field experiments, and (ii) investigate the correlation between spectral data/SRIs and seed protein for predicting seed protein content at maturity as a target trait. This would indicate genetic differences in N fixation which could subsequently be utilized for selecting soybean genotypes with improved $N_2$ fixation.

## 2. Materials and methods

### 2.1. Environments and experimental setup

Field experiments were carried out in three macro-environments located in the east of Austria, i.e. Tulln 2019 (TU 2019), Tulln 2020 (TU 2020) and Gross Enzersdorf 2020 (GE 2020). At Tulln (located 26 km north-west of Vienna), maize was the preceding crop in both years, and fields were under low-input management receiving no fertilizer for the soybean crop. At Gross Enzersdorf (Raasdorf site of Experimental Farm Gross Enzersdorf, 17 km east of Vienna), winter wheat was the preceding crop, and fields were under full organic management. Experiments were arranged in single-row plots of 2 m length and 0.5 m row spacing (1 $m^2$) at a planting density of 50–70 plants per plot. Soybeans were planted in the last week of April in each year, and combine harvesting took place when all plots had reached full maturity, i.e. in the first week of October and the last week of September for the 2019 and 2020 seasons, respectively. Due to dry conditions, the Gross Enzersdorf experiment was irrigated after planting as well as at two later dates in July and August. Experiments were arranged in randomized complete block (RCB) designs with two (2019) and two or three replications (2020) depending on genotype subset. In year 2019, 160 plots were established, whereas in 2020 235 plots were planted in each of the two locations; thus, a final number of 630 plots was subject to data collection and all further analyses.

### 2.2. Soybean genotypes

A total set of 95 soybean genotypes of maturity groups 00 and 000 were utilized in the present experiments. In order to be able to observe a wide range of variation in nitrogen metabolism and finally seed protein content, soybean genotypes of five different subsets (genotype classes) were selected: Subsets 1 and 2 were breeding lines derived from a cross between a non-nodulating soybean ($rj_1$) and a nodulating ($Rj_1$) parent, with lines of subset 1 expressing the non-nodulating phenotype, whereas lines of subset 2 were regularly nodulating. Subsets 3 and 4 were breeding lines previously selected for either high seed protein content or high pod set, whereas subset 5 was comprised of standard soybean cultivars. Subsets 3, 4 and 5 were each planted in 2 replications in each of the three environments. Subsets 1 and 2 were planted in 1–2 replications in the Tulln 2019 environment due to limited seed availability and in 3 replications each in the Tulln 2020 and Gross Enzersdorf 2020 environments.

### 2.3. Data collection

#### 2.3.1. Field data, seed quality determination, SPAD- and PolyPen readings

Phenotypic characters such as time to maturity (expressed in the number of days after July 31 until full maturity (R8) stage) and plant height (main stem length in cm) were determined as one reading per plot. After harvest, seed samples were gently dried at ambient conditions for several weeks reaching an average moisture content of approx. 8% for optimum storage and further analysis. Samples of 8–10 g were then finely ground (2-mm sieve, Cyclotec 1093 mill, Foss Tecator, Höganäs, Sweden) for near-infrared reflectance spectroscopy (NIRS) based seed quality analysis using a Bruker Matrix-I Fourier transform NIRS instrument (Bruker, Ettlingen, Germany). Seed oil, protein and sucrose contents were calculated from respective calibrations, and results were all expressed on a dry matter basis (0% moisture). In the Tulln 2019 experiment, leaf chlorophyll content was estimated at three developmental stages (R2-R3: full bloom – beginning pod; R4: full pod; R5-R6: beginning – full seed; Fehr and Caviness, 1977) using the SPAD-502

**Table 1**
Summary of comparative ANOVA results for SPAD meter measurements of chlorophyll content and spectral reflectance indices NRI, PRI570 and MA1_R from the same plots at three measuring dates during the TU 2019 season (for all data: standardization through z-transformation) as well as their respective correlation to seed protein content.

| | R2-R3: Full bloom - Beginning pod (17/18 Jul 2019) | | | | R4: Full pod (31 Jul/1 Aug 2019) | | | | R5-R6: Beginning - Full seed (15/16 Aug 2019) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ANOVA**[a] | **SPAD** | **NRI** | **PRI570** | **MA1_R** | **SPAD** | **NRI** | **PRI570** | **MA1_R** | **SPAD** | **NRI** | **PRI570** | **MA1_R** |
| Model F ratio | 5.417 | 4.368 | 7.926 | 7.785 | 6.652 | 12.722 | 5.661 | 19.392 | 9.361 | 12.626 | 7.27 | 34.747 |
| Genotype F ratio | 5.488 | 4.388 | 8.039 | 7.941 | 6.696 | 12.987 | 5.764 | 19.803 | 9.553 | 12.777 | 7.274 | 35.13 |
| Error mean squ. | 0.274 | 0.331 | 0.194 | 0.197 | 0.228 | 0.124 | 0.263 | 0.083 | 0.166 | 0.125 | 0.21 | 0.047 |
| **Coefficient of correlation** (r) to seed protein content (n = 160) | | | | | | | | | | | | |
| | 0.404 | −0.558 | 0.803 | −0.715 | 0.503 | −0.834 | 0.838 | −0.872 | 0.799 | −0.833 | 0.803 | −0.880 |

[a] all F values are significant at the $p < 0.0001$ level.

chlorophyll meter device (Konica Minolta Sensing, Osaka, Japan). For each plot, 30 chlorophyll-meter readings were taken from the uppermost fully developed leaves to obtain one average SPAD value. In the Tulln 2020 experiment, a PolyPen RP 410 NIR device (PSI, Drasov, Czech Republic; 640–1050 nm wavelength range) was used at the R4 (full pod) stage on 98 plots for collecting leaf reflectance spectra. From fully developed uppermost leaves, 25 readings per plot were taken with 3 internal replications per reading, and reflectance indices were calculated using SpectraPen software (Vers. 1.1; PSI, Drasov, Czech Republic).

### 2.3.2. *Hyperspectral data acquisition*

Soybean spectral reflectance was collected using a FieldSpec Hand-Held 2 (ASD Inc., Boulder, CO, USA) VNIR spectroradiometer in the 325–1075 nm wavelength range. For data acquisition, the device was held about 30 cm above the soybean canopy resulting in a 17.5 cm diameter of the measurement area based on a field of view of 25° angle with the standard optical input device. The instrument was then gently moved (0.05–0.1 m/sec) along each 2 m soybean row for about 20–30 s at a setting of 34 ms integration time in order to obtain a total of 400 (10 × 40) reflectance spectra. All measurements were taken within ± 2 h of solar noon under clear sky, and a white reference standard adjustment was carried out every 10 min using the polytetrafluoroethylene based Spectralon (LabSphere Inc., North Sutton, NH, USA) panel. In each environment, measurements were repeated at several time intervals between mid of July (R1 stage – beginning of flowering) and mid of August (R6 stage – full seed development). Measurements were stopped before the onset of senescence in the earliest maturity plots.

## 2.4. *Data processing, spectral indices and statistical analysis*

The collected spectral information was exported as reflectance data to a spreadsheet format without any pre-processing using ViewSpec Pro Vers. 6.2 software (ASD Inc.). Data from multiple measurements per plot were averaged for further processing. Hyperspectral data were then directly utilized to calculate correlations to end-of-season traits such as seed protein content. Furthermore, a total of 43 spectral reflectance indices (SRIs) describing either the biomass/yield/photosynthesis complex, nitrogen- or water-related traits and suitable for the wavelength range covered by the instrument used (Suppl. Table 1 for description of SRIs and the respective wavelengths used) were calculated from spectral data, and their correlation to other phenotypic traits was studied. As the highest correlations to seed protein content were found for spectral data collected at the R4-R5 stage of development (full pod - beginning seed), data from that stage were utilized for joint analysis of spectral and seed data across all environments, as described below. In addition to SRIs, the whole range of spectral data was also utilized to predict seed protein content based on partial least square regression (PLSR) models using Unscrambler Vers. 11 software (Camo Software, Oslo, Norway). Both spectral indices and phenotypic data were subject to analysis of variance (ANOVA) in order to test the significance of genotype classes, individual genotypes and environments using Proc MIXED of the SAS Vers. 9.4 package (SAS Inst., Cary, NC, USA). Graphical data visualization and additional statistical analysis were carried out using OriginPro Vers. 2021b software (OriginLab Corp., Northampton, MA, USA).

## 3. Results

### 3.1. *Seed protein content and comparison of phenotyping devices*

A wide range of soybean seed protein content (290–490 g/kg) was determined in harvest samples from different environments and genotype classes (Fig. 1) pointing to differences in nitrogen metabolism. Non-nodulating soybeans (subset 1) were clearly lower in seed protein content than their nodulating counterparts in all environments except for

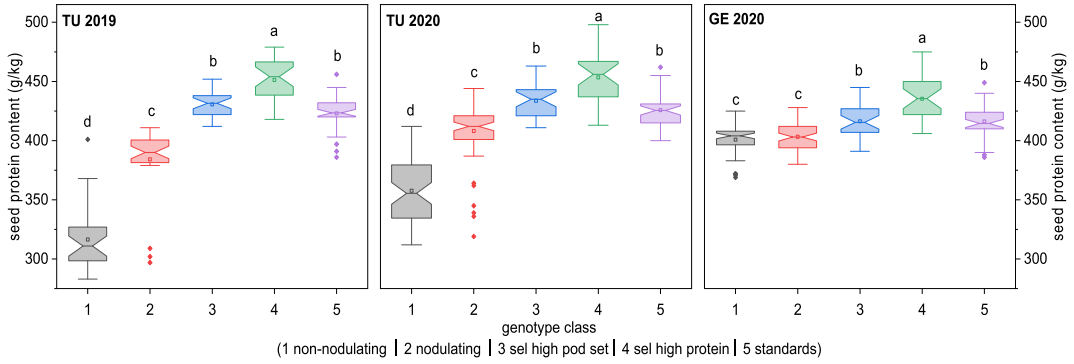



**Fig. 1.** Variation of seed protein content (g/kg) for five genotype classes (subsets) across each of the three environments TU 2019, TU 2020, and GE 2020. Notched boxes with 25–75% range, bar marking 1.5 interquartile range, median line (notches indicating 95% lower and upper confidence intervals of median value), arithmetic mean (dot) and outliers (filled dots); letters above boxes indicating significant differences between genotype classes within environment (Tukey-Kramer multiple comparison at $p = 0.05$ level).

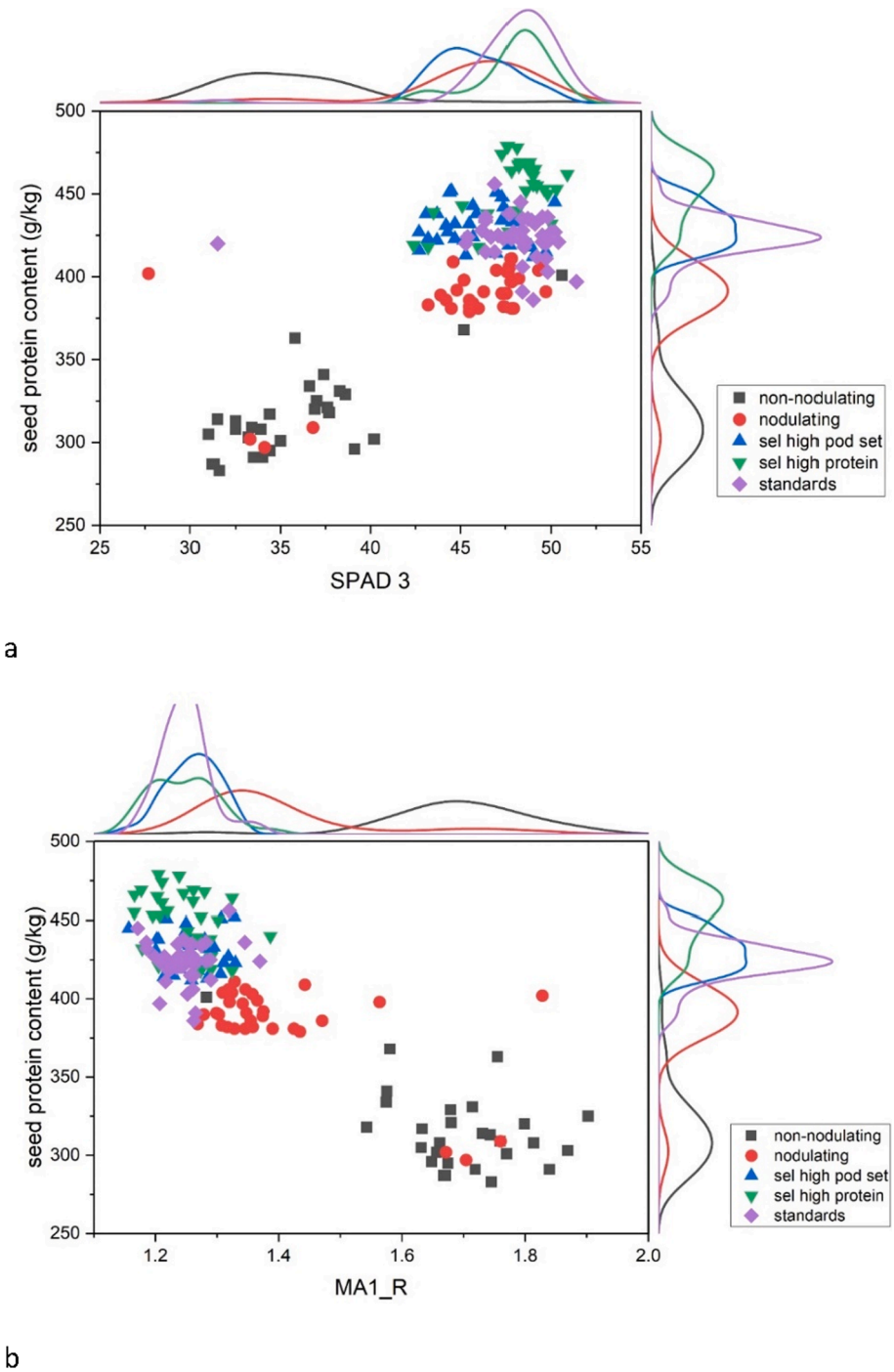



**Fig. 2.** Seed protein content (TU 2019) as related to (a) SPAD measurement of chlorophyll content (r = 0.799; n = 160) and (b) spectral index MA1_R (r = −0.872; n = 160) both determined at R5-R6 stage (15 Aug 2019). Overlay lines: Distributions of genotype classes (Kernel smooth density function) with line color matching the color of genotype classes in legends.

GE 2020. Standard cultivars and genotypes selected for high pod set were similar in protein content, whereas the highest protein content expectedly was present in lines previously selected for high seed protein content.

Determination of chlorophyll content by SPAD-metering (TU 2019 environment) as an indirect measure of crop N status could differentiate between soybean genotypes. SPAD-measurements at later stages of development resulted in increased F-ratios and reduced error mean squares (Table 1). For comparison, three SRIs calculated from FieldSpec hyperspectral data and related to N uptake (NRI, PRI570, MA1_R) had a similar or better analytical power in terms of ANOVA parameters than SPAD values. Similarly, correlations between SPAD or SRIs and final seed protein content also increased with developmental stage, and the spectral index MA1_R revealed the best overall performance both in ANOVA and correlation results (Table 1). The relationship between SPAD measurement and seed protein content (r = 0.799) is illustrated in Fig. 2a; overlay distributions for SPAD values indicate differences between non-nodulating, high pod-set and the other genotype subsets. For comparison, the correlation between spectral index MA1_R and seed protein content (r = −0.872) is higher, and the differentiation between genotype classes (Fig. 2b, MA1_R overlay distributions) is better than for the SPAD parameter. In addition to the SPAD metering device, a PolyPen

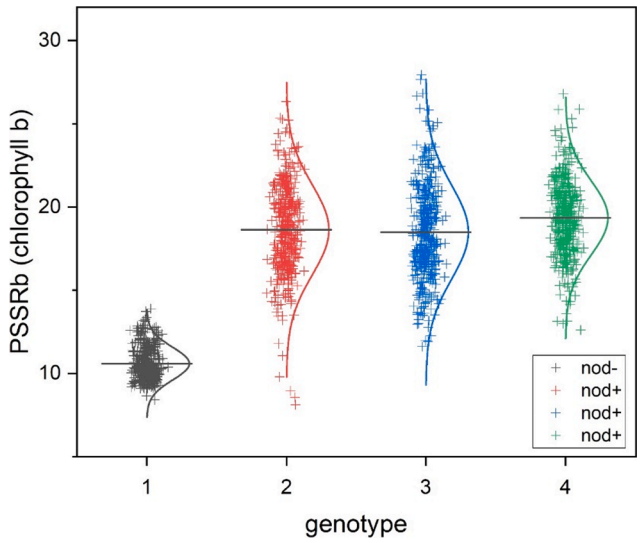



**Fig. 3.** Index PSSRb indicating chlorophyll *b* content of 4 genotypes differing in nodulation status (1 = non-nodulating, 2, 3, 4: nodulating) measured in a total of 400 replications each.

RP410 was compared to the ASD hyperspectral device in the TU 2020 environment. Performance of the PolyPen device was similar to the SPAD device both in terms of correlation to seed protein content (Suppl. Fig. 1a) as well as in ANOVA results (Suppl. Table 2). Several SRIs calculated from the hyperspectral data collected with the ASD device out-performed the PolyPen results (index REIP in Suppl. Fig. 1b and Suppl. Table 2).

### 3.2. Hyperspectral measurement procedure

Instead of single point measurements commonly implemented, the ASD hyperspectral device was moved over the canopy rows (see 2.3.2. for details) to collect a maximum of spectral information possible by utilizing 400 spectral replications. As an example, distributions in chlorophyll *b* content from individual spectral replications are presented for 4 different genotypes (Fig. 3); while the non-nodulating genotype is clearly lower in chlorophyll *b* content, smaller differences between genotypes 2, 3, and 4 are visible as well. The power of the high number of spectral replications is evident from Fig. 4 which illustrates a stable ranking of genotypes after about 100 replications, and small differences between genotypes appear to be detectable. This applies to other indices as well, as outlined in Suppl. Figs. 2 and 3 for carotenoid content (index PSSRc) as well as in Suppl. Fig. 4 for the double-peak canopy nitrogen index (DCNI) shown here for example.

### 3.3. Spectral correlations and spectral indices

Correlations between spectral reflectance at individual wavelengths and seed protein content followed clear patterns along the wavelength range from 325 to 1075 nm (Fig. 5) for the nodulating and non-nodulating subsets of genotypes at different measurement times of the TU 2019 experiment; patterns were less clear for the other subsets with less variation in protein content (Suppl. Fig. 5), but the red edge correlation peak at about 700 nm is evident here as well. Utilizing all genotype classes in one analysis, correlations across wavelengths between spectral reflectance and either seed protein or sucrose content revealed symmetrically diverging patterns (Fig. 6) which is an indication of the negative correlation between protein and sucrose content. In contrast, the correlogram for time to maturity is different from the ones for protein and sucrose.

The correlations along wavelengths between spectral reflectance and individual traits encouraged the calculation of dedicated spectral indices for a meaningful characterization of genotype classes and individual genotypes. Consequently, ANOVA results revealed highly significant differences between genotype classes for both phenotypic traits and SRIs (Table 2).

Differences between genotypes within genotype classes were significant for most of the SRIs related to nitrogen metabolism, whereas genotypes were not significant for most of the water- related indices. In addition, strong environment effects and significant genotype by environment interactions were found for most SRIs as well, indicating that SRIs can be considered as regular quantitative traits (Table 2). The descriptive power of selected SRIs for differentiating genotypes in N-related traits is also illustrated by their significant multiple mean separation potential as compared to the seed protein content trait (Suppl. Fig. 6).

### 3.4. Correlations between spectral predictors and seed protein content

Individual SRIs were correlated to phenotypic traits such as time to maturity, plant height, oil, protein or sucrose content, and 1000-seed weight (Suppl. Table 3). Eight SRIs highly correlated to seed protein content and their correlations to other seed traits as well as to each other SRI are presented in Fig. 7. As already indicated in Table 1 for the TU 2019 environment only, the spectral index MA1_R also revealed the highest correlation ($r = -0.75$) to seed protein content across all environments. While this correlation is based on individual plots ($n = 630$) across all environments, the high predictive power of MA1_R for seed protein content is evident from a correlation based on genotype means within environment of $r = -0.903$ (Suppl. Fig. 7a vs. 7b). Alternatively to single correlation/regression models for predicting seed protein content from particular SRIs, PLSR-based models were developed to utilize the whole spectroscopic information available. In models utilizing the samples from all three environments for calibration, $R^2$ values of

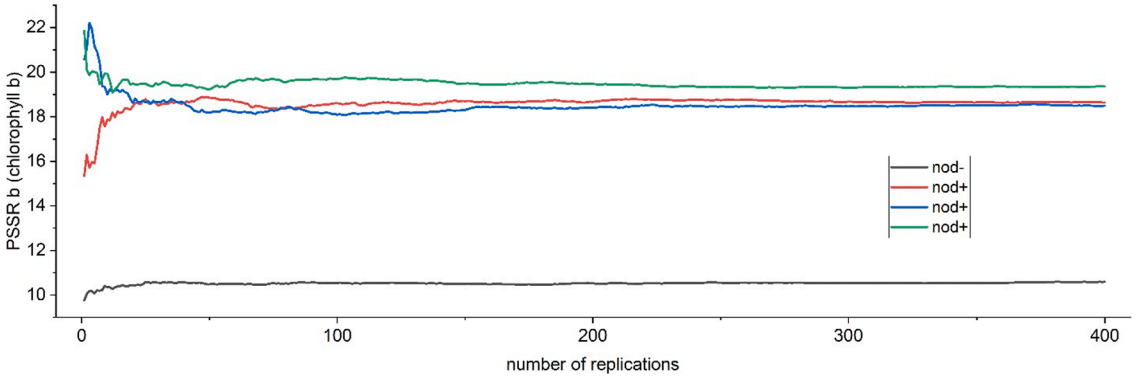



**Fig. 4.** Effect of the number of replications for measuring chlorophyll *b* (PSSR b) in 4 genotypes differing in nodulation status.

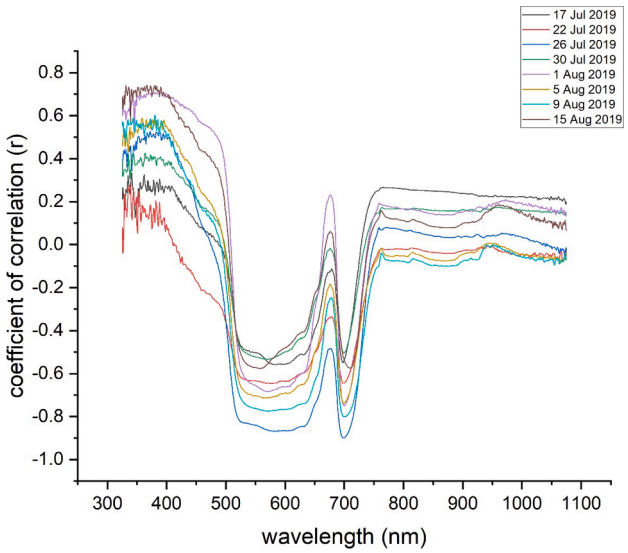



**Fig. 5.** Relationship between hyperspectral reflection at eight different data collection dates during the soybean seed filling period and seed protein content of the harvest product: Correlograms describing correlations between reflectance at given wavelengths (1 nm increment) and seed protein content for the Tulln 2019 subsets 1 and 2 (nodulating and non-nodulating genotype classes).

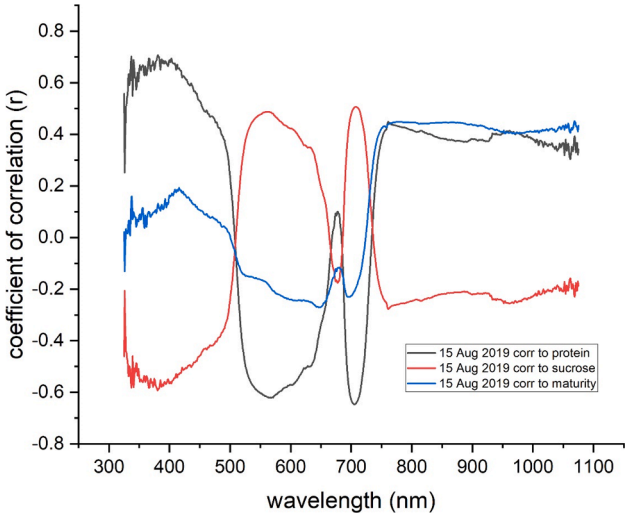



**Fig. 6.** Relationship between hyperspectral reflection at the 15 Aug 2019 measurement date and different crop traits: Correlograms describing correlations between reflectance at given wavelengths (1 nm increment) and either seed protein content, seed sucrose content or time to maturity. The negative correlation between seed protein and seed sucrose content is evident from diverging correlations, whereas the correlogram for time to maturity is revealing a different pattern with highest correlations in the near-infrared region from 750 nm on.

0.844 and 0.805 were achieved for calibration and cross validation, respectively (Fig. 8). Higher $R^2$ values were found in models for individual environments (Fig. 9), and modelling results either separate for each single environment or combined across all environments are summarized in Table 3. The least precise model was found for the GE 2020 environment, in which variation in seed protein content was lowest. As PLSR models exhibit the highest precision in predicting seed protein content from spectral information, the most important wavelength regions contributing to the models are indicated in terms of their weighted regression coefficients for the four main regression factors (Suppl. Fig. 8); green, red, red edge and the infrared shoulder at 770–900 nm are of particular importance in the first and second factor. Regression factor weights along wavelengths were similar both in models covering all environments or particular environments only (Suppl. Fig. 9). In addition, regression factor weights are in partial agreement with direct correlations between spectral reflectance and seed protein content as shown above (Fig. 5).

## 4. Discussion

This study aimed at evaluating the performance of non-destructive spectral-based methods for estimating genotypic variation in spectral indices related to N accumulation and subsequently seed protein content among soybean genotypes in multi-environment plant breeding trials, which might serve as indirect screening procedures for N fixation. In general, soybean seed protein concentration is in the range of 40–42% depending on genotype, environmental conditions and management practice (Bellaloui et al., 2011; Medic et al., 2014). However, as symbiotic N fixation plays a key role in both yield formation and seed quality of soybean, variation in seed protein content may occur in a much wider range due to soybean genotype and rhizobial inoculant effects (Zimmer et al., 2016). This applies to the present experiments as well, revealing a seed protein content from 290 to 490 g/kg (Fig. 1). In the context of this research, a wider range in protein content as an end-of-season target trait was desirable for developing models to predict seed protein content from mid-season spectral traits, which can be considered as an indirect method for quantifying N fixation and plant N uptake.

As there is a close relationship between N accumulation and photosynthetic rate during vegetative development (Sinclair, 2004), chlorophyll metering (SPAD-meter) represents a further option for quantifying N uptake, which is also related to seed protein content (Vollmann et al., 2011). While the merit of SPAD metering was confirmed in the present research (Table 1), spectral indices derived from reflectance measurements performed better (index MA1_R and other SRIs in Table 1 and Fig. 2) than SPAD metering. This appears to be mainly due to the benefit of much larger sampling area covered by hyperspectral reflectance, whereas SPAD metering and similar devices such as the PolyPen spectrometric instrument (Suppl. Fig. 1 and Suppl. Table 2) are based on relatively small sampling spots requiring replicated measurements, which limits their practical application for screening larger numbers of genotypes.

In agronomic research, reflectance measurement with hand-held spectroscopic devices is often implemented as a point measurement with several replications along a larger plot (Christenson et al., 2016; Duan et al., 2019; Palka et al., 2021). As single-row soybean plots are expected to be less homogenous in canopy reflection properties than larger agronomic plots of cereal crops, hyperspectral data collection was implemented here utilizing a continuous spectra acquisition mode (similar to the implementation described in Li et al., 2014) with 400 replications per single plot. Thus, although vegetation indices calculated from individual spectra exhibit a wide variation (Fig. 3, Suppl. Figs. 2–4), a clear differentiation of genotypes in characters such as chlorophyll *b* (PSSRb), carotenoids (PSSRc) or canopy nitrogen (DCNI) was feasible based on averaging across a larger number of spectral replications. This suggests that the spectral data collection method is of particular importance to plant breeding-related crop physiology research. Consequently, spectral reflectance indices were statistically significant for the different genotype classes (Table 2). Most of the indices related to biomass accumulation or N metabolism (see Suppl. Table 1 for description of SRIs) were also significant on the genotype level (Table 2), which allows to characterize and classify genotypes according to their N metabolism similar to seed protein content (Suppl. Fig. 6) or other quantitative traits. While nitrogen related SRIs often derived from red-edge wavelengths in the 690–730 nm range were significant, water-related indices centering around the 970 nm water absorption band (Christenson et al., 2016; Prey et al., 2020) were not significant for genotypes within classes (Table 2). This indicates that

**Table 2**
Mixed-model ANOVA results (F-ratios) for soybean phenotypic traits and SRIs (R4; full pod developmental stage) across 3 environments.

| trait | g_class | | geno (g_class) | | rep (env) | | env | | e × g (g_class) | | residual error MS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | fixed | | fixed | | random | | random | | random | | |
| time to maturity | 192.57 | *** | 9.73 | *** | 2.05 | ns | 80.69 | *** | 2.18 | *** | 4.21 |
| plant height | 43.17 | *** | 4.83 | *** | 1.96 | ns | 69.81 | *** | 2.21 | *** | 85.46 |
| seed oil content | 239.69 | *** | 1.65 | ** | 2.52 | * | 9.67 | ** | 4.42 | *** | 81.58 |
| seed protein cont | 117.45 | *** | 1.39 | * | 2.71 | * | 12.52 | *** | 12.23 | *** | 84.55 |
| seed sucrose cont | 77.19 | *** | 3.85 | *** | 0.41 | ns | 28.59 | *** | 6.84 | *** | 0.034 |
| 1000-seed weight | 85.47 | *** | 4.6 | ** | 1.66 | ns | 1.55 | ns | 4.23 | *** | 85.52 |
| NDVI | 16.26 | *** | 0.97 | ns | 0.28 | ns | 18.94 | *** | 2.55 | *** | 0.00012 |
| SIPI | 10.21 | *** | 1.22 | ns | 0.15 | ns | 16.54 | *** | 2.1 | *** | 0.00008 |
| PSSRa | 18.55 | *** | 1.32 | ns | 0.36 | ns | 16.49 | *** | 2.98 | *** | 4.21 |
| PSSRb | 49.21 | *** | 1.47 | * | 0.4 | ns | 14.73 | *** | 3.82 | *** | 2.44 |
| CRI | 41.5 | *** | 2.28 | *** | 1.63 | ns | 19.41 | ** | 1.24 | * | 31.81 |
| ARI | 93.69 | *** | 2.66 | *** | 1.97 | ns | 5.57 | * | 1.36 | ** | 9.78 |
| PSSRc | 6.45 | *** | 1.63 | ** | 0.34 | ns | 17.55 | *** | 2.3 | *** | 5.74 |
| NRI | 62.12 | *** | 2.96 | *** | 0.69 | ns | 14.9 | *** | 2.35 | *** | 0.00045 |
| RVSI | 84.8 | *** | 1.57 | ** | 3.09 | ** | 16.03 | ** | 1.87 | *** | 0.00001 |
| PSRI | 24.07 | *** | 0.78 | ns | 1.91 | ns | 21.17 | *** | 2.43 | *** | 5.78E-06 |
| CI | 93.43 | *** | 1.53 | ** | 1.12 | ns | 25.34 | *** | 4.53 | *** | 0.051 |
| PRI570 | 81.84 | *** | 1.56 | ** | 3.38 | ** | 43.61 | *** | 5.13 | *** | 2E-05 |
| DCNI | 137.92 | *** | 3.01 | *** | 1.97 | ns | 2.13 | ns | 2.11 | *** | 2.44 |
| GI | 52.19 | *** | 2.61 | *** | 0.74 | ns | 15.24 | *** | 2.36 | *** | 0.028 |
| GNDVI | 54.11 | *** | 0.99 | ns | 0.85 | ns | 13.92 | *** | 5.83 | *** | 0.00025 |
| VOG1 | 93.92 | *** | 1.35 | * | 1.35 | ns | 30.28 | *** | 5.11 | *** | 0.00304 |
| VOG2 | 109.23 | *** | 1.41 | * | 1.36 | ns | 39.31 | *** | 4.66 | *** | 0.00027 |
| VOG3 | 110.97 | *** | 1.56 | ** | 1.28 | ns | 35.78 | *** | 4.32 | *** | 0.00083 |
| R705 | 63.83 | *** | 1.19 | ns | 0.92 | ns | 21.42 | *** | 7.89 | *** | 0.00011 |
| R780/740 | 85.49 | *** | 1.2 | ns | 1.36 | ns | 36.34 | *** | 4.6 | *** | 0.00024 |
| MSR705_445 | 108.42 | *** | 1.74 | ** | 1.2 | ns | 27.07 | *** | 4.98 | *** | 0.1437 |
| REIP | 71.6 | *** | 0.97 | ns | 1.37 | ns | 26.74 | *** | 7 | *** | 0.2895 |
| WI | 56.22 | *** | 1.01 | ns | 1.87 | ns | 98.65 | *** | 2.63 | *** | 7.4E-05 |
| NWI-1 | 56.17 | *** | 1 | ns | 1.86 | ns | 98.15 | *** | 2.6 | *** | 1.6E-05 |
| NWI-2 | 72.87 | *** | 1.2 | ns | 2.79 | * | 74.15 | *** | 2.46 | *** | 2.4E-05 |
| NWI-3 | 47.79 | *** | 1.05 | ns | 2.47 | * | 87.28 | *** | 2.5 | *** | 1.2E-05 |
| NWI-4 | 64.17 | *** | 1.13 | ns | 2.63 | * | 73.23 | *** | 2.44 | *** | 2.2E-05 |
| NWI-5 | 19.63 | *** | 1.03 | ns | 0.97 | ns | 196.22 | *** | 2.03 | *** | 1.5E-05 |
| WI_1 | 21.04 | *** | 0.96 | ns | 5.45 | *** | 6.71 | * | 1.79 | *** | 6.3E-05 |
| WI_2 | 44.45 | *** | 0.95 | ns | 2.81 | * | 79.18 | *** | 2.27 | *** | 4.9E-05 |
| WI_3 | 2.82 | * | 0.8 | ns | 4.9 | ** | 25.23 | ** | 1.94 | *** | 0.00011 |
| RNDVI_1 | 14.42 | *** | 0.98 | ns | 0.26 | ns | 17.06 | *** | 2.52 | *** | 0.00011 |
| RNDVI_2 | 13.62 | *** | 1 | ns | 0.18 | ns | 15.41 | *** | 2.56 | *** | 0.00011 |
| RNDVI_3 | 13.67 | *** | 1.02 | ns | 0.33 | ns | 15.24 | *** | 2.56 | *** | 0.00011 |
| RENDVI_1 | 69.1 | *** | 1.11 | ns | 0.9 | ns | 26.68 | *** | 5.64 | *** | 0.00042 |
| RENDVI_2 | 68.23 | *** | 1.18 | ns | 0.63 | ns | 24.77 | *** | 5.59 | *** | 0.00041 |
| RENDVI_3 | 70.85 | *** | 1.23 | ns | 1.04 | ns | 26.58 | *** | 5.86 | *** | 0.00038 |
| WI / NDVI | 20.03 | *** | 1.5 | ** | 1.25 | ns | 55.36 | *** | 2.35 | *** | 0.00014 |
| WBI | 56.08 | *** | 0.99 | ns | 1.84 | ns | 98.01 | *** | 2.58 | *** | 5E-05 |
| MA1_N | 90.94 | *** | 2.12 | *** | 0.38 | ns | 20.2 | *** | 6.74 | *** | 0.00015 |
| MA1_R | 84.89 | *** | 1.91 | *** | 0.32 | ns | 20.4 | *** | 7.03 | *** | 0.0013 |
| MB1_N | 85.23 | *** | 2.04 | *** | 0.9 | ns | 15.54 | *** | 6.8 | *** | 0.00017 |
| MB1_R | 79.79 | *** | 1.82 | ** | 0.74 | ns | 16.33 | *** | 7.18 | *** | 0.0015 |

Significance levels of F-ratios: ***: $p < 0.0001$; **: $p < 0.01$; *: $p < 0.05$; ns: not significant.

under the prevailing environmental conditions in the field trials, relevant differences in water-related traits (i.e. leaf water content variation due to water stress, transpiration, degree of maturity) could not be detected among individual genotypes at the time of spectral measurement.

Each of the spectral indices highly correlated to soybean seed protein content (Fig. 7) was inversely correlated to both oil and sucrose content as well. This is due to the frequently described negative relationships between both protein and oil as well as protein and sucrose content in the soybean seed (Sato et al., 2014; Wilcox and Shibles, 2001; Wilson, 2004). Most of the indices related to protein content in this research were previously described because of their relationship to canopy N status in wheat, soybean or other crops (Duan et al., 2019; Feng et al., 2016; Li et al., 2014; Prey et al., 2020). Thus, they detect N accumulated in the green plant biomass which is later mobilized and transferred to the seed for protein biosynthesis. From a recently published set of index terms in soybean, index MA1_R clearly revealed the best performance in terms of high correlation to seed protein content and low error variance throughout all present experiments (Table 1, Suppl. Table 3, Fig. 7, Suppl. Fig. 7). MA1_R is a ratio vegetation index based on wavelengths of $\lambda = 638$ and 674 nm, which had originally been derived from wavelength optimization for soybean yield prediction at R5 developmental stage with aerial remote sensing platforms (Zhang et al., 2019); it is based on a red wavelength similar to NRI and DCNI nitrogen indices which explains its suitability for predicting protein content as well. This suggests that protein yield, which is correlated to both grain yield and seed protein content, could be predicted as well as an end-of-season trait indicating the total amount of N harvested.

While SRIs appear as suitable parameters which could be implemented in multispectral imaging devices for high-throughput prediction (e.g. drones with dedicated multiband camera sensors), partial least square regression (PLSR) models can utilize information along the full reflectance spectrum. Thus, the best relationship to seed protein content across all environments was $r^2 = 0.56$ for the spectral index MA1_R (see Fig. 7; $r = -0.75$ for simple correlation between protein and MA1_R), whereas the PLSR prediction model on the same data set reached $R^2$

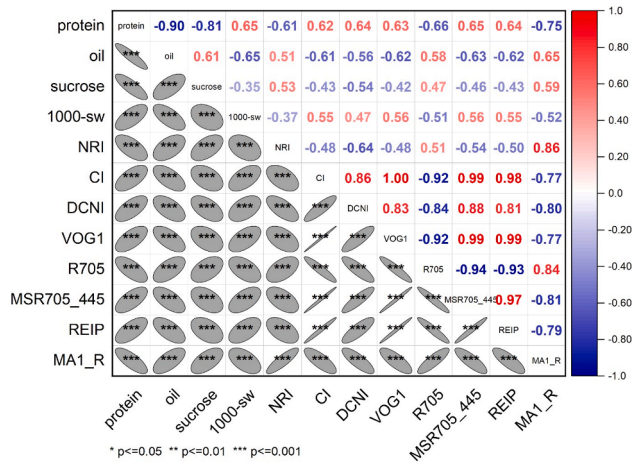
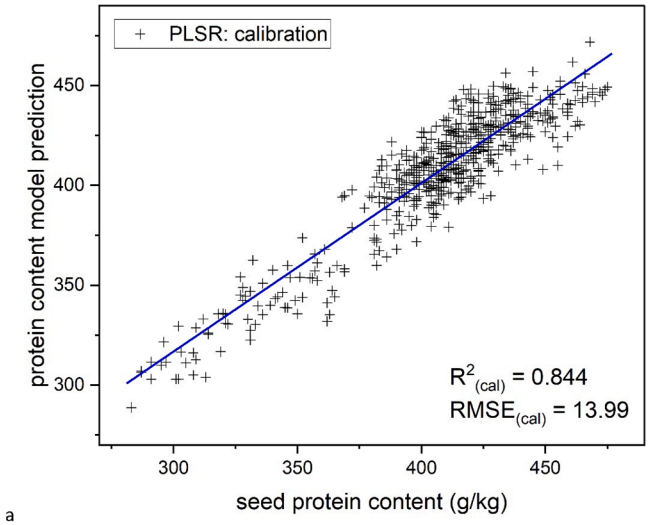



**Fig. 7.** Pearson coefficient of correlation (r) between soybean seed traits and nitrogen/protein-related SRIs (n = 630, individual plots basis, 3 environments).



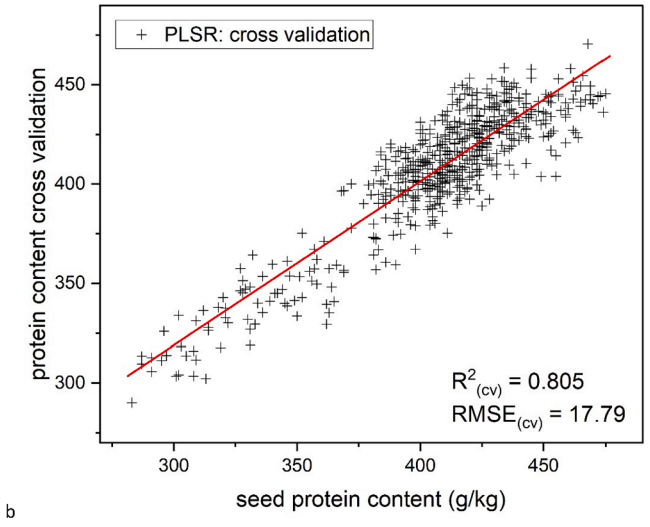
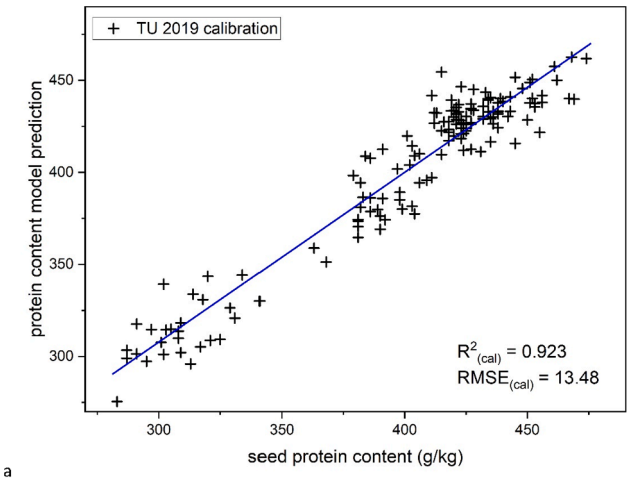




**Fig. 8.** Relationship between seed protein content and PLSR calibration model (a) and cross validation (b) based on spectral data (n = 589 plots) from all three environments.



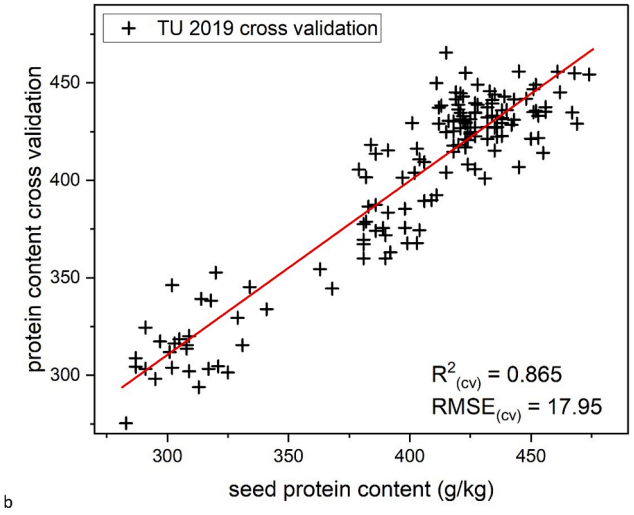




**Fig. 9.** Relationship between seed protein content and PLSR calibration model (a) and cross validation (b) based on spectral data (n = 149 plots) from TU 2019 environment only.

values of 0.84 and 0.81 for calibration and validation, respectively (Table 3; all environments, outliers removed). A similarly clear improvement of prediction through PLSR modelling as compared to SRIs was reported for canopy N content in wheat (Li et al., 2014). However, while Li et al. (2014) were predicting canopy N content for the actual timepoint of spectral data collection, in the present research canopy spectral properties at early seed filling stage were used to predict "future" seed protein content as the trait of interest at the end of season. In soybean, Chiozza et al. (2021) reported PLSR-based predictions for protein, grain yield, leaf area index (LAI) and other traits from hyperspectral reflectance data sampled at R5/R6 stage. While their percentage of variance explained was highest for LAI, it was 39% for a protein prediction model. This lower response might have been caused by a smaller range in seed protein content in their calibration data set as well as by later sampling stages and a comparatively low number of spectral scans (replications) per experimental unit.

The crucial role of timing in spectral measurements has been illustrated in soybean yield prediction models which were most accurate for spectral readings taken at the R5 stage (Crusiol et al., 2021). For seed protein prediction, earlier stages such as R4-R5 might be most indicative, as N translocation for seed filling would reduce leaf N at later stages thus reducing differences between genotypes. This is also indicated by peak correlations of NRI and PRI570 indices to seed protein content at R4 developmental stage (Table 1). As different traits of interest might have prediction optima at different timepoints, temporal dynamics of

**Table 3**
Summary statistics for PLSR modelling results for prediction of seed protein content from hyperspectral reflectance data based on either individual environments or combined across all three environments.

| | | TU 2019 | | TU 2020 | | GE 2020 | | all environments | |
|---|---|---|---|---|---|---|---|---|---|
| Parameter | | all[a] | rem.outl.[b] | all | rem.outl. | all | rem.outl. | all | rem.outl. |
| No. of factors | | 12 | 12 | 12 | 12 | 12 | 12 | 15 | 15 |
| RMSE[c] | - calibration | 16.26 | 13.48 | 15.97 | 13.51 | 12.06 | 10.74 | 16.03 | 13.99 |
| | - validation | 20.68 | 17.95 | 19.32 | 16.39 | 14.26 | 12.63 | 17.79 | 15.64 |
| R-square | - calibration | 0.892 | 0.923 | 0.8269 | 0.8735 | 0.5073 | 0.6256 | 0.8008 | 0.8441 |
| | - validation | 0.8275 | 0.8653 | 0.7502 | 0.8158 | 0.3187 | 0.4878 | 0.7548 | 0.8052 |

[a] all plots/samples utilized in model.
[b] outliers removed.
[c] residual mean square error.

developmental processes such as N fixation, pod development or seed filling as well as genotype differences therein could probably not be modelled efficiently in single spectral readings. Therefore, a temporal approach based on repeated measurements over time has recently been proposed utilizing random regression models for predicting above-ground biomass in soybean (Moreira et al., 2021); such approaches might be efficient for other longitudinal traits with high temporal dynamics such as N accumulation and grain yield formation as well.

While organic farming and other low-input cropping systems rely on di-nitrogen fixation as the most important source of N (Reckling et al., 2020), soybean nodulation and N fixation are variable depending on environmental factors as well as on genetic and biological interactions (Zimmer et al., 2016; Brar and Lawley, 2020). This complexity is also mirrored in the pattern of seed protein contents given in Fig. 1: In contrast to other environments, the protein content of the non-nodulating subset 1 is similar to other genotype classes in the GE 2020 environment, which is probably due to enhanced and continuing N mineralization throughout the soybean growing period under irrigation after a previous longer period of severe drought. In case of a single environment calibration, such conditions led to reduced predictive power of PLSR modelling (Table 3, GE 2020), whereas it practically did not affect the quality of prediction models across all environments. Similar findings (Crusiol et al., 2021) demonstrate the potential of multi-environment PLSR modelling for increasing the accuracy of soybean yield prediction and applicability of models across environments. Thus, chemometric strategies of sample selection from different growing seasons as widely used in NIRS calibration development (e.g. Zhao et al., 2021) might be applied in hyperspectral modelling across environments as well. While breeding for increased N fixation would be an important agronomic component of soybean competitiveness in organic farming (Vollmann and Menken, 2012), the finding of a lower seed protein prediction quality of PLSR models in the high soil N environment GE 2020 (Table 3) also indicates the need for suitable selection sites low in soil-N to avoid masking genetic differences in N fixation by excessive N uptake from soil. Likewise and in addition to the selection environment, the degree of genetic variation or heritability of N fixation in breeding populations would be equally important for significant responses to selection.

## 5. Conclusions

The present paper has addressed the options for predicting soybean seed protein content as a trait of interest from hyperspectral reflectance data collected during the early seed filling stages. As canopy N content is the result of both symbiotic di-nitrogen fixation and N uptake from soil, this would indirectly quantify total N accumulation as well, and thus spectral indices related to N metabolism could be used to determine N uptake of individual genotypes. As demonstrated in the results, the selected SRIs were effective not only in differentiating between non-nodulating and nodulating soybeans, but also could identify significant quantitative differences between conventionally nodulating genotypes, thus indicating quantitative genetic differences in N accumulation. For prediction of seed protein content as an end-of-season trait, coefficients of determination ($R^2$) of 0.84 and 0.81 for calibration and validation, respectively, were found using PLSR models of hyperspectral data. These models apparently are more powerful than SRIs with simple correlations to seed protein content based on few wavelength points along the spectrum only. However, spectral index-based approaches could more easily be implemented in robust high-throughput procedures for screening large numbers of genotypes in soybean breeding programs.

## Funding

The research leading to these results has partly received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 771367, ECOBREED. The support from BOKU project "Phenotyping Across Experimental Scales" (project no. IA 13460) is also acknowledged. Open access funding of the publication is provided by University of Natural Resources and Life Sciences, Vienna (BOKU).

## Data availability statement

Spectral and experimental raw data are available at the Zenodo data repository at https://doi.org/10.5281/zenodo.6668497.

### *CRediT authorship contribution statement*

**Johann Vollmann:** Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft. **Pablo Rischbeck:** Methodology, Investigation, Writing – review & editing. **Martin Pachner:** Investigation, Data curation. **Vuk Đorđević:** Methodology, Supervision. **Ahmad M. Manschadi:** Conceptualization, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2022.107169.

## References

Araus, J.L., Kefauver, S.C., Zaman-Allah, M., Olsen, M.S., Cairns, J.E., 2018. Translating high-throughput phenotyping into genetic gain. Trends Plant Sci. 23 (5), 451–466. https://doi.org/10.1016/j.tplants.2018.02.001.

Banerjee, B.P., Joshi, S., Thoday-Kennedy, E., Pasam, R.K., Tibbits, J., Hayden, M., Spangenberg, G., Kant, S., 2020. High-throughput phenotyping using digital and

hyperspectral imaging-derived biomarkers for genotypic nitrogen response. Journal of Experimental Botany 71, 4604-4615. https://doi.org/10.1093/jxb/eraa143.
Bellaloui, N., Reddy, K.N., Bruns, H.A., Gillen, A.M., Mengistu, A., Zobiole, L.H.S., Fisher, D.K., Abbas, H.K., Zablotowicz, R.M., Kremer, R.J., 2011. Soybean seed composition and quality: Interactions of environment, genotype, and management practices. In: Maxwell, J.E. (Ed.), Soybeans: Cultivation, Uses and Nutrition. Nova Science Publishers, Inc., New York, USA, pp. 1–42.
Bi, Y., Kong, W., Huang, W., 2018. Hyperspectral diagnosis of nitrogen status in arbuscular mycorrhizal inoculated soybean leaves under three drought conditions. Int. J. Agric. Biol. Eng. 11, 126–131. https://doi.org/10.25165/j.ijabe.20181106.4019.
Bosse, M.A., da Silva, M.B., de Oliveira, N.G.R.M., de Araujo, M.A., Rodrigues, C., de Azevedo, J.P., dos Reis, A.R., 2021. Physiological impact of flavonoids on nodulation and ureide metabolism in legume plants. Plant Physiol. Biochem. 166, 512–521. https://doi.org/10.1016/j.plaphy.2021.06.007.
Brar, N., Lawley, Y., 2020. Short-season soybean yield and protein unresponsive to starter nitrogen fertilizer. Agron. J. 112 (6), 5012–5023. https://doi.org/10.1002/agj2.20378.
Chen, J., Li, F., Wang, R., Fan, Y., Raza, M.A., Liu, Q., Wang, Z., Cheng, Y., Wu, X., Yang, F., Yang, W., 2019. Estimation of nitrogen and carbon content from soybean leaf reflectance spectra using wavelet analysis under shade stress. Comput. Electron. Agric. 156, 482–489. https://doi.org/10.1016/j.compag.2018.12.003.
Chiozza, M.V., Parmley, K.A., Higgins, R.H., Singh, A.K., Miguez, F.E., 2021. Comparative prediction accuracy of hyperspectral bands for different soybean crop variables: From leaf area to seed composition. Field Crops Res. 271, 108260. https://doi.org/10.1016/j.fcr.2021.108260.
Christenson, B.S., Schapaugh, W.T., An, N., Price, K.P., Prasad, V., Fritz, A.K., 2016. Predicting soybean relative maturity and seed yield using canopy reflectance. Crop Sci. 56 (2), 625–643. https://doi.org/10.2135/cropsci2015.04.0237.
Crusiol, L.G.T., Nanni, M.R., Furlanetto, R.H., Sibaldelli, R.N.R., Cezar, E., Sun, L., Foloni, J.S.S., Mertz-Henning, L.M., Nepomuceno, A.L., Neumaier, N., Farias, J.R.B., 2021. Yield prediction in soybean crop grown under different levels of water availability using reflectance spectroscopy and partial least squares regression. Rem. Sens. 13, 977. https://doi.org/10.3390/rs13050977.
Duan, D.D., Zhao, C.J., Li, Z.H., Yang, G.J., Zhao, Y., Qiao, X.J., Zhang, Y.H., Zhang, L.X., Yang, W.D., 2019. Estimating total leaf nitrogen concentration in winter wheat by canopy hyperspectral data and nitrogen vertical distribution. J. Integrat. Agric. 18, 1562–1570. https://doi.org/10.1016/S2095-3119(19)62686-9.
Fehr, W.R., Caviness, C.E., 1977. Stages of soybean development. Iowa State University, Special Report No. 80, Cooperative Extension Service.
Feng, W., Zhang, H.Y., Zhang, Y.S., Qi, S.L., Heng, Y.R., Guo, B.B., Ma, D.Y., Guo, T.C., 2016. Remote detection of canopy leaf nitrogen concentration in winter wheat by using water resistance vegetation indices from in-situ hyperspectral data. Field Crops Res. 198, 238–246. https://doi.org/10.1016/j.fcr.2016.08.023.
Fritschi, F.B., Ray, J.D., 2007. Soybean leaf nitrogen, chlorophyll content, and chlorophyll a/b ratio. Photosynthetica 45 (1), 92–98. https://doi.org/10.1007/s11099-007-0014-4.
Herridge, D.F., Peoples, M.B., Boddey, R.M., 2008. Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil 311 (1-2), 1–18. https://doi.org/10.1007/s11104-008-9668-3.
Herrmann, I., Vosberg, S.K., Ravindran, P., Singh, A., Chang, H.X., Chilvers, M.I., Conley, S.P., Townsend, P.A., 2018. Leaf and canopy level detection of *Fusarium virguliforme* (sudden death syndrome) in soybean. Rem. Sens. 10, 426. https://doi.org/10.3390/rs10030426.
Krezhova, D., Kirova, E., 2011. Hyperspectral remote sensing of the impact of environmental stresses on nitrogen fixing soybean plants (*Glycine max* L.). In: Ilarslan, M., Ince, F., Kaynak, O., Basturk, S. (Eds.), Proceedings of 5th International Conference on Recent Advances in Space Technologies 5966816 - RAST 2011, Istanbul, 9-11 June 2011, pp. 172-177, IEEE Catalog Number CFP11819, Istanbul, Turkey, ISBN 978-1-4244-9614-3.
Lausch, A., Salbach, C., Schmidt, A., Doktor, D., Merbach, I., Pause, M., 2015. Deriving phenology of barley with imaging hyperspectral remote sensing. Ecol. Model. 295, 123–135. https://doi.org/10.1016/j.ecolmodel.2014.10.001.
Li, F., Mistele, B., Hu, Y., Chen, X., Schmidhalter, U., 2014. Reflectance estimation of canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices and partial least squares regression. Eur. J. Agron. 52, 198–209. https://doi.org/10.1016/j.eja.2013.09.006.
Medic, J., Atkinson, C., Hurburgh, C.R., 2014. Current knowledge in soybean composition. J. Am. Oil. Chem. Soc. 91 (3), 363–384. https://doi.org/10.1007/s11746-013-2407-9.
Moreira, F.F., de Oliveira, H.R., Lopez, M.A., Abughali, B.J., Gomes, G., Cherkauer, K.A., Brito, L.F., Rainey, K.M., 2021. High-throughput phenotyping and random regression models reveal temporal genetic control of soybean biomass production. Front. Plant Sci. 12, 715983. https://doi.org/10.3389/fpls.2021.715983.
Oberson, A., Nanzer, S., Bosshard, C., Dubois, D., Mäder, P., Frossard, E., 2007. Symbiotic $N_2$ fixation by soybean in organic and conventional cropping systems estimated by $^{15}N$ dilution and $^{15}N$ natural abundance. Plant Soil 290 (1-2), 69–83. https://doi.org/10.1007/s11104-006-9122-3.
Palka, M., Manschadi, A.M., Koppensteiner, L., Neubauer, T., Fitzgerald, G.J., 2021. Evaluating the performance of the CCCI-CNI index for estimating N status of winter wheat. Eur. J. Agron. 130, 126346. https://doi.org/10.1016/j.eja.2021.126346.
Prey, L., Hu, Y., Schmidhalter, U., 2020. High-throughput field phenotyping traits of grain yield formation and nitrogen use efficiency: Optimizing the selection of vegetation indices and growth stages. Front. Plant Sci. 10, 1672. https://doi.org/10.3389/fpls.2019.01672.
Reckling, M., Bergkvist, G., Watson, C.A., Stoddard, F.L., Bachinger, J., 2020. Re-designing organic grain legume cropping systems using systems agronomy. Eur. J. Agron. 112, 125951. https://doi.org/10.1016/j.eja.2019.125951.
Salvagiotti, F., Cassman, K.G., Specht, J.E., Walters, D.T., Weiss, A., Dobermann, A., 2008. Nitrogen uptake, fixation and response to fertilizer N in soybeans: A review. Field Crops Res. 108 (1), 1–13. https://doi.org/10.1016/j.fcr.2008.03.001.
Sato, T., Van Schoote, M., Wagentristl, H., Vollmann, J., Singh, R., 2014. Effects of divergent selection for seed protein content in high-protein vs. food-grade populations of early maturity soybean. Plant Breeding 133 (1), 74–79. https://doi.org/10.1111/pbr.12138.
Schweiger, P., Hofer, M., Hartl, W., Wanek, W., Vollmann, J., 2012. $N_2$ fixation by organically grown soybean in Central Europe: Method of quantification and agronomic effects. Eur. J. Agron. 41, 11–17. https://doi.org/10.1016/j.eja.2012.01.011.
Sinclair, T.R., 2004. Improved carbon and nitrogen assimilation for increased yield. In: Boerma, H.R., Specht, J.E. (Eds.), Soybeans: Improvement, Production, and Uses, Series Agronomy, No. 16, (third ed.), American Society of Agronomy, Madison, WI, USA, pp. 537-568. https://doi.org/10.2134/agronmonogr16.3ed.c11.
Szczyglowski, K., Ross, L., 2021. Baring the roots of nodulation. Nature Plants 7, 244-245. https://doi.org/10.1038/s41477-021-00886-1.
Vollmann, J., Walter, H., Sato, T., Schweiger, P., 2011. Digital image analysis and chlorophyll metering for phenotyping the effects of nodulation in soybean. Comput. Electron. Agric. 75 (1), 190–195. https://doi.org/10.1016/j.compag.2010.11.003.
Vollmann, J., Menken, M., 2012. Soybean: breeding for organic farming systems. In: Lammerts van Bueren, E.T., Myers, J.R. (Eds.), Organic Crop Breeding. Wiley, pp. 203–214. https://doi.org/10.1002/9781119945932.ch12.
Wilcox, J.R., Shibles, R.M., 2001. Interrelationships among seed quality attributes in soybean. Crop Sci. 41 (1), 11–14. https://doi.org/10.2135/cropsci2001.41111x.
Wilson, R.F., 2004. Seed composition. In: Boerma, H.R., Specht, J.E. (Eds.), Soybeans: Improvement, Production, and Uses, Series Agronomy, No. 16, (third ed.), American Society of Agronomy, Madison, WI, USA, pp. 621-677. https://doi.org/10.2134/agronmonogr16.3ed.c13.
Xie, C., Yang, C., 2020. A review on plant high-throughput phenotyping traits using UAV-based sensors. Comput. Electron. Agric. 178, 105731. https://doi.org/10.1016/j.compag.2020.105731.
Zhang, X., Zhao, J., Yang, G., Liu, J., Cao, J., Li, C., Zhao, X., Gai, J., 2019. Establishment of plot-yield prediction models in soybean breeding programs using UAV-based hyperspectral remote sensing. Remote Sensing 11, 2752. https://doi.org/10.3390/rs11232752.
Zhao, X., Zhao, X., Huang, M., Zhu, Q., 2021. An uncertainty sampling strategy based model updating method for soluble solid content and firmness prediction of apples from different years. Chemomet. Intellig. Laborato. Syst. 217, 104426. https://doi.org/10.1016/j.chemolab.2021.104426.
Zimmer, S., Messmer, M., Haase, T., Piepho, H.P., Mindermann, A., Schulz, H., Habekuß, A., Ordon, F., Wilbois, K.P., Heß, J., 2016. Effects of soybean variety and *Bradyrhizobium* strains on yield, protein content and biological nitrogen fixation under cool growing conditions in Germany. Eur. J. Agron. 72, 38–46. https://doi.org/10.1016/j.eja.2015.09.008.