

## Article

# Comparison and Application of Non-Destructive NIR Evaluations of Seed Protein and Oil Content in Soybean Breeding

Guo-Liang Jiang

Agricultural Research Station, Virginia State University, PO Box 9061, Petersburg, VA 23806, USA;  
gjiang@vsu.edu

Received: 24 December 2019; Accepted: 2 January 2020; Published: 5 January 2020



**Abstract:** A plant breeding program needs to evaluate a large number of materials for different traits within a limited time. Near-infrared (NIR) spectroscopy has been used to quickly determine seed composition in various crop species. In this study, we compared whole-seed evaluations of protein and oil content by NIR methods in soybean [*Glycine max* (L.) Merr.], and then discussed the application to plant breeding. The differences among the entries tested were highly significant in both traits for each method used. No significant difference but high correlation and consistency existed between DA 7250 and wet-chemistry methods. Compared with DA 7250, ZX-50 exhibited, to some extent, differences or errors. The differences of ZX-50 methods were found to be correlated with seed sizes and could be corrected using regression equations formulated for bias calculation. After correction, the differences in the predictions between DA 7250 and ZX-50 methods were insignificant. Similar to DA 7250, ZX-50 methods exhibited a high repeatability (> 98%) of the predictions. By validation with 760 bulk samples of different seed types and 240 single-plant samples, it further demonstrated that as a non-destructive, fast and cost-efficient method, ZX-50 NIR analysis with an appropriate bias correction could be used in soybean breeding, specifically suitable for single plant selection based on whole seeds.

**Keywords:** soybean; seed composition; NIR analysis; plant breeding; single plant selection

## 1. Introduction

In modern plant breeding, a breeder needs to assess large numbers (thousands and/or tens of thousands) of breeding materials for multiple traits within a limited period of time. Rapid and robust phenotypic evaluation is still a great challenge and a realistic demand in practical breeding for many quantitative traits of importance including nutrient components of crops [1]. High throughput phenotyping helps breeders to efficiently perform evaluations and timely select the desired genotypes in the breeding populations with complicated variations.

Soybean [*Glycine max* (L.) Merr.] is a major crop grown worldwide and plays an important part in the agricultural production, human food security and international trade. Soybean seed consists mainly of protein, oil, carbohydrates, minerals and water. On average, approximately 40% and 20% of dry seed weight in soybean are protein and oil, respectively. The usages of soybean are considerably dependent on the seed composition [2]. For instance, high-oil cultivars are expected for vegetable-oil processing and/or industrial uses such as biodiesel, while high protein content is usually preferred in human diet and soy-based food industries. Moreover, the physiochemical characteristics of seed may also affect soybean price [3].

Seed protein and oil content are two of the most important seed composition and nutritional quality traits in soybean. Traditionally, concentrations of protein and oil are determined by lab wet

chemical techniques, such as the Kjeldahl procedure and combustion nitrogen analysis for protein and Soxhlet method for oil [4,5]. These conventional methods provide accurate and precise measurements of protein and oil content. However, they are time- and labor-consuming, cost-inefficient and high-lab input [1,6,7], and also generate chemical residues [8]. These techniques need a trained technician to perform and cannot achieve multiple-constituent measurements for a single sample at the same time [9]. More importantly, they are seed-destructive and after analysis, no seed can be used for selection and planting. They are not suitable for and cannot satisfy a large-scale analysis and timely selection of thousands and/or tens of thousands of breeding materials. Thus, these techniques are undesirable for plant breeding, especially when a limited amount of seed is available [6,7,10].

Near-infrared (NIR) spectroscopy is a fast, simple and non-destructive technology for analysis of chemical materials/nutrients in food and crops with little sample preparation [11,12]. It offers fast response, automatic reading and recording, easy operation and limited space required, and in particular, simultaneous analysis of multiple parameters [13,14]. NIR spectroscopy is such a technology that may meet the demand of plant breeding for a timely and large-scale evaluation, because it can measure seed composition traits relatively accurately, particularly protein and oil, much faster than other measurement techniques [11].

Since the 1970s, NIR spectroscopy has been extensively used to determine the protein, oil and other chemical compounds in various crop species [11], including soybean [1,7–9,15–18], corn [19,20], wheat [21], rapeseed and chickpea [6], sunflower [10], rice and pigeon pea [9]. AACC International [22] has also recommended NIR method for the evaluation of protein, oil and moisture in soybean based on whole seed. An overview of NIR analysis for the characterization of plant varieties and cultivars has been presented by Biancolillo and Marini [23]. The seed samples used are mostly taken from combined or bulk harvested seed. Since the 2000s, researchers have also discussed the use of NIR spectroscopy to analyze nutrient content of single kernel instead of multiple-seed samples [13,19,20,24–26]. However, single-seed selection is not so often conducted in plant breeding except for differentiating quantitative and qualitative seed composition mutants from normal seeds. There are still some limitations in using NIR spectroscopy for single seed analysis [11,27]. For instance, seed size and morphological characteristics may cause the spectral variance within seeds of a same cultivar [27]. In addition, the percentage of seed nutrient components within a plant may vary and is associated with seed position where the seed developed [28–30]. The difference in seed composition attributed to seed position within a plant is not inheritable. In practical breeding, selection is typically based on individual plants and/or equivalents, and then the families or lines derived from the selected plants. Therefore, seed composition based on the bulked seeds from a single plant is more meaningful than that of a single seed and more useful for selection. However, use of NIR spectroscopy in the measurements of nutrient components of single plant has been rarely reported.

Perten DA 7250 manufactured by Perten Instruments AB, Sweden is a diffuse reflectance NIR spectroscopy [31]. It has been extensively used in seed protein and oil analysis, and the data generated has been commonly accepted as equivalent to the results of traditional chemical analysis in soybean breeding and research [17,24,32–35]. Relatively speaking, DA 7250 is a high-quality and high-cost machine that allows as many as over 40 parameters to be analyzed simultaneously at a rate of 50–60 samples per minute. It is a good option for lab uses but it is not suitable to use in mobile status or field selection. Different from DA 7250, ZX-50 grain analyzer manufactured by the Zeltex Inc. in the USA [36] has a transmittance module and is a low-cost, affordable and small portable whole seed NIR analyzer that can determine moisture, protein and oil content at a rate of 40–50 samples per hour. It can be used in the field and a moving vehicle when it runs with battery. Similar to DA 7250, ZX-50 can be used for different crop species, such as corn, wheat, soybean and oilseeds. Therefore, we presumed that ZX-50 could be used in soybean breeding and in particular for single-plant selection. The aim of this study was, by analyzing seed protein and oil content using these two different NIR analyzers (DA 7250 and ZX-50), to explore the application of ZX-50 as a rapid, simple and affordable method to single plant evaluation and selection for seed protein and oil content in soybean breeding.

## 2. Materials and Methods

### 2.1. NIR Evaluations and Prediction Correcting of ZX-50 Methods

In total, 20 soybean entries with a large range of variation in 100-seed weight determined based on three subsamples were used in the study. Whole seed samples of 16 genotypes, mostly of maturity group V and VI except one from group 0 and one from group II, were taken from the bulk-harvested seeds of yield trials or crossing blocks in the same field at Virginia State University Randolph Research Farm in Ettrick, Virginia in 2016. For four of the 16 genotypes, additional whole seed samples designated as name \_(a) were also taken from a different year and/or location to examine the consistency between different seed sources. Together, these 20 seed samples were treated as individual entries in analysis.

The protein and oil content of whole seed were determined using lab wet-chemistry methods [4,5], a DA 7250 NIR analyzer (Pertin Instruments, Hägersten, Sweden) [31] and ZX-50 portable grain NIR analyzer (Zeltex, Inc., Hagerstown, Maryland) [36]. For the DA 7250, the calibrations used in this study were developed and updated by the manufacturer for both whole seed and ground samples, having a determination coefficient of 0.881 for protein and 0.824 for oil based on 3538 and 3276 samples, respectively. In the ZX-50, the original calibrations were installed by the manufacturer, with a suggestion for further calibration or correction of predictions to be needed in practical uses. In the meantime, another 25-mm sample cup of ZX-50 NIR analyzer was also modified by placing a piece of 1.2–1.5 cm thick foam material on the bottom and both insides of the cup to reduce the sample-holding space to a desired size, in order that single-plant samples (small as about 10–15 g of seeds) could be analyzed. The method using the modified sample cup in ZX-50 NIR analyzer was named as ZX-50 MC, relative to ZX-50, DA 7250, and wet-chemistry. For each entry, about 40 g whole seeds for DA 7250 and ZX-50 but 15 g used for ZX-50 MC and wet-chemistry were randomly taken from a single plot and analyzed. The analysis was repeated three times by each of the four methods. The seed moisture was also estimated at the same time, averaging 6–9%, and thus the seed protein and oil content were presented as g 100 g<sup>-1</sup> on a dry weight basis (i.e., 0% moisture).

A complete randomized design with three replications was adopted in the statistical analysis. Entries and methods were treated as the main and fix effects, and replications were treated as random effects. Data processing was conducted using Microsoft Excel 2013. Analysis of variance (ANOVA) [37] was performed using PROC GLM in SAS version 9.4 (SAS Institute Inc., Cary, NC, USA). A combined ANOVA including all data was performed to compare the results of the methods, but single ANOVA for individual methods was conducted separately to compare the differences among genotypes and calculate repeatability. The repeatability was estimated as

$$R = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2 / r) \quad (1)$$

where  $\sigma_g^2$  = entry variance,  $\sigma_e^2$  = error variance, and  $r$  = number of subsamples or replications.

There was no significant difference observed between DA 7250 and wet-chemistry, both exhibiting a high consistency, and thus DA 7250 was regarded as the reference for further analysis in this study. The differences between ZX-50 or ZX-50 MC and DA 7250 were calculated for calibrating the results of ZX-50 methods. Since the differences in either protein or oil content between DA 7250 and the ZX-50 methods were found to vary with genotypes and were associated with seed sizes, a linear regression analysis was performed [38]. The linear regression equations computed between the difference and 100-seed weight were used to determine the biases for each entry. Then the corrected values of protein and oil content for ZX-50 and ZX-50 MC were calculated using the original readings and the biases given by the linear regression equations.

### 2.2. Validation and Application of ZX-50 Methods

A total of 760 seed samples were analyzed for protein and oil content as described above to validate the reliability and explore the application of ZX-50 MC methods in breeding. The samples

evaluated were taken from four batches of different soybean seed types and sources of maturity group V and VI. Two of them were the mature seeds harvested after full maturity (R8 stage): 117 lines each with two replications were sampled from 2015 soybean yield trials and 147 lines without replication from 2016 trials. The seeds were stored in a cold storage room and the moisture of seed samples was mostly 6–9% when the analyses were performed. The other two sets of samples were the dried immature or edamame seeds collected from the same trials: 88 lines, 50 of which had two replications for 2015, and 127 lines, 114 of which had two replications for 2016. For the edamame samples, the fresh soybean pods were harvested at R6 stage when the pods were still green [33]. Then the samples were fully dried at 65 °C for 2–3 weeks and threshed to get the dried edamame seeds for analysis. The dried edamame seeds were stored under the room temperature and the moisture was mostly 6–8% when the samples were analyzed. Each sample of 25–30 g seeds was analyzed twice for protein and oil content using AD 7250 and ZX-50 MC, as described above, and 100-seed weight was determined on a scale as well. The original data of ZX-50 MC was corrected using the established regression equation based on the results of calibration with 20 entries.

Data processing was carried out in Microsoft Excel 2013, and ANOVA [37] was performed using PROC GLM in SAS version 9.4. Pearson correlation between DA 7250 and ZX-50 MC was computed based on genotype means for individual sample sets but on sample basis for a combined analysis of all 760 samples. Paired t-test was also computed to determine the significance of difference between DA 7250 and ZX-50 methods.

Furthermore, ZX-50 MC was also used to determine the seed protein and oil content in 240 single-plant samples to confirm its suitability to single plant selection. These single plants were selected from 2016 and 2017 breeding materials of maturity group IV, V and VI and then planted in plant rows in the following crop season. After planting, the remaining seeds were stored in a cold storage. For each of the 240 single plants, a whole-seed sample of 10–20 g was taken from the remaining seed, and the seed moisture was 7–9% when analyzed. Protein and oil content were determined twice for a same sample using ZX-50 MC and AD 7250 as described above. Hundred-seed weight was also determined. The corrected data of ZX-50 MC was calculated using the established regression equation as described above. Data processing, paired t-test and Pearson correlation analysis were performed in Microsoft Excel 2013.

### 3. Results and Discussion

#### 3.1. NIR Predictions and Comparisons of Protein and Oil Content

With all the original data combined together, ANOVA showed that there were highly significant differences ( $p < 0.01$ ) among the entries in both seed protein and oil content (Table 1). The method  $\times$  entry interaction was also significant. ANOVA with the original predictions indicated that the differences among methods were highly significant ( $p < 0.01$ ) for both protein and oil content (Table 1), but no significant difference existed between DA 7250 and wet-chemistry. The wet-chemistry and DA 7250 methods exhibited a high correlation ( $r = 0.977$  for protein and  $0.960$  for oil) (Table 2). The absolute values of differences between these two methods were mostly less than  $1.0 \text{ g } 100 \text{ g}^{-1}$  for protein content and  $0.5 \text{ g } 100 \text{ g}^{-1}$  for oil content, respectively. Paired t-test also indicated that there was no significant difference between DA 7250 and the wet-chemistry. By analyzing separately for different methods, Table 2 presents the differences in prediction of protein and oil content by DA 7250 and ZX-50 methods. For DA 7250, overall the 20 entries, protein and oil content averaged  $42.0$  and  $19.4 \text{ g } 100 \text{ g}^{-1}$  (compared to  $41.8$  and  $19.5 \text{ g } 100 \text{ g}^{-1}$  of the wet-chemistry measurement), respectively, ranging from  $32.3$  to  $47.3 \text{ g } 100 \text{ g}^{-1}$  for protein and from  $16.6$  to  $23.1 \text{ g } 100 \text{ g}^{-1}$  for oil content (Table 2). The ranges of variation in the two traits evaluated with DA 7205 also exhibited a high similarity to those by the wet-chemistry method ( $33.4$ – $47.7 \text{ g } 100 \text{ g}^{-1}$  for protein and  $17.0$ – $23.1 \text{ g } 100 \text{ g}^{-1}$  for oil). The results evidently demonstrated that there was no significant difference but a high consistency between DA 7250 NIR analysis and the wet-chemistry measurement in evaluation of seed protein

and oil content in soybean. Therefore, DA 7250 was then used as the reference for comparison and correction of the data predicted by ZX-50 and ZX-50 MC for further analysis.

**Table 1.** Means, range of variation and F value in ANOVA of seed protein and oil content in 20 soybean entries analyzed by wet-chemistry and three NIR methods (DA 7250, ZX-50 and ZX-50 MC).

Trait	Data	Mean	Range	F <sub>M</sub> <sup>a</sup>	F <sub>G</sub> <sup>a</sup>	F <sub>MG</sub> <sup>a</sup>
Protein	Original	38.7 ± 4.4	29.8–48.0	2819.34 **	325.03 **	16.97 **
	Corrected ZX-50 and ZX-50 MC	42.0 ± 3.5	31.7–48.1	2.14	485.37 **	8.81 **
Oil	Original	20.6 ± 2.2	16.3–25.5	774.10 **	137.61 **	5.97 **
	Corrected ZX-50 and ZX-50 MC	19.4 ± 1.8	16.3–23.4	0.50	184.20 **	3.77 **

\*\* Significant at  $p < 0.01$ . <sup>a</sup> F<sub>M</sub>, F<sub>G</sub>, and F<sub>MG</sub> are F values of significance tests for method, entry, and method x entry interaction effects, respectively.

**Table 2.** Means, range of variation and repeatability of the wet-chemistry and Near-infrared (NIR) analysis methods for soybean seed protein and oil content based on 20 entries and the coefficient of correlation with DA 7250 method.

Method	Protein Content (g 100 g <sup>-1</sup> )				Oil Content (g 100 g <sup>-1</sup> )			
	Mean	Range	Repeatability (%)	Correlation with DA 7250	Mean	Range	Repeatability (%)	Correlation with DA 7250
Wet-chemistry	41.8 ± 3.4	33.4–47.7	99.02	0.977 **	19.5 ± 1.9	17.0–23.1	96.31	0.960 **
DA 7250	42.0 ± 4.0	32.3–47.3	99.36		19.4 ± 2.0	16.6–23.1	99.25	
ZX-50	36.3 ± 2.5	30.0–39.7	99.72	0.942 **	20.8 ± 1.5	18.4–23.9	99.35	0.908 **
ZX-50 MC	34.8 ± 1.6	31.1–37.0	94.71	0.889 **	22.9 ± 1.0	21.3–24.8	89.13	0.894 **
Corrected ZX-50	42.0 ± 3.5	33.7–47.0	99.86	0.970 **	19.5 ± 1.9	16.9–23.2	99.58	0.946 **
Corrected ZX-50 MC	42.0 ± 3.1	35.7–47.0	98.66	0.918 **	19.4 ± 1.6	17.1–22.1	95.50	0.937 **

\*\* Significant at  $p < 0.01$ .

To have an overall comparison between the ZX-50 methods and DA 7250, the 20 entries analyzed were averaged for individual method. The averages of 20 entries in the predicted seed protein content by ZX-50 and ZX-50 MC were 36.3 and 34.8 g 100 g<sup>-1</sup>, which were 5.7 and 7.2 g 100 g<sup>-1</sup> less than that of DA 7250, respectively (Table 2). The predicted values of oil content by ZX-50 and ZX-50 MC averaged 20.8 and 22.9 g 100 g<sup>-1</sup>, respectively, and 1.4 and 3.5 g 100 g<sup>-1</sup> higher than that of DA 7250. It seems that the original predictions by ZX-50 methods could not reflect the exact contents of protein and oil in the seeds. Moreover, the differences between the predictions of ZX-50 or ZX-50 MC and DA 7250 varied considerably with genotypes. This indicated that the predictions must be corrected and a single bias could not satisfy the need of correcting all the predictions in different genotypes.

### 3.2. Seed Size and Correction of Predictions by ZX-50 and ZX-50 MC

There was a large range of variation in seed size among the 20 entries used in this study. The 100-seed weight varied from 12.0 to 33.7 g, with an average of 20.0 g (Table 3). The variation was larger than most of the ranges of 100-seed weight reported previously [39–42]. Past research on selection for high seed protein have mostly tested genotypes with larger seed size but rarely included small seed genotypes with high seed protein [41]. Filho et al. [39] reported significant correlations between 100-seed weight and protein or oil content. However, no significant correlation between 100-seed weight and protein or oil content was also reported [40]. Poeta et al. [41] suggested two contrasting strategies based on seed size to increase protein: seeds with increased seed protein content in large seed genotypes, and seeds with reduced oil and carbohydrate contents in small seed genotypes. In general, large-seeded edamame or vegetable soybeans have higher protein but lower oil content than commercial grain-type soybeans [32]. In the present study, there was significant correlation observed between 100-seed weight and the predictions of protein or oil content in the 20 entries evaluated



with DA 7250 and ZX-50 or ZX-50 MC. Therefore, we supposed that the seed size might affect the predictions [35].

**Table 3.** Hundred-seed weight, and means and variations of protein and oil content determined using DA 7250 NIR analyzer in 20 soybean genotypes/entries.

Genotype	100-Seed Weight (g)	Protein Content (g 100 g <sup>-1</sup> )			Oil Content (g 100 g <sup>-1</sup> )		
		Mean	Range	CV (%)	Mean	Range	CV (%)
PI 549058	20.3	45.4 ± 0.5	44.7–46.2	1.01	16.9 ± 0.3	16.4–17.4	1.90
AGS 346	29.8	43.3 ± 0.7	41.6–44.7	1.71	18.2 ± 0.5	17.0–19.2	2.81
IA 1008LF	14.0	40.2 ± 0.3	39.4–40.7	0.78	18.5 ± 0.3	18.0–19.1	1.84
NC 346	21.4	42.6 ± 0.8	41.8–44.5	1.77	19.1 ± 0.3	18.5–19.8	1.64
VS12-0203	13.9	37.9 ± 0.7	37.3–39.2	1.72	23.1 ± 0.5	22.1–24.2	2.23
VS12-0205	12.0	32.3 ± 0.7	31.5–34.1	2.17	21.0 ± 0.4	20.4–21.7	1.98
VS12-0069	21.9	47.1 ± 0.9	45.8–48.4	1.89	16.6 ± 0.6	15.4–17.4	3.50
VS11-0022	30.7	46.5 ± 0.8	45.2–48.0	1.80	17.3 ± 0.4	16.8–18.1	2.05
NC Green	33.7	47.3 ± 0.8	45.6–48.6	1.78	17.8 ± 0.4	17.2–18.6	2.11
Osage	15.8	44.2 ± 0.8	43.0–45.5	1.80	19.7 ± 0.4	19.1–20.2	1.96
NC Raleigh	15.0	36.0 ± 0.6	35.3–37.2	1.79	22.9 ± 0.3	22.4–23.5	1.45
N6202-8	23.9	46.8 ± 0.7	45.5–47.6	1.51	17.0 ± 0.5	16.1–17.7	3.02
Asmara	23.0	40.8 ± 0.9	39.6–42.3	2.33	17.7 ± 0.7	16.3–18.6	3.75
Ellis	12.7	38.1 ± 0.8	36.9–39.7	2.08	22.3 ± 0.4	21.7–23.3	1.76
VS22-450	13.6	41.3 ± 0.6	40.6–43.1	1.56	21.7 ± 0.3	20.8–22.0	1.52
VS12-0128	24.9	38.8 ± 1.0	37.5–40.5	2.51	19.9 ± 0.6	18.5–20.5	2.85
NC 346_(a)	22.0	43.1 ± 1.0	42.0–44.9	2.39	19.9 ± 0.3	19.3–20.3	1.34
Osage_(a)	14.3	42.8 ± 0.7	41.9–44.1	1.69	20.0 ± 0.3	19.4–20.5	1.66
N6202-8_(a)	24.1	45.4 ± 1.2	43.9–47.1	2.58	18.0 ± 0.6	16.6–18.6	3.41
Ellis_(a)	13.3	39.9 ± 0.6	39.1–41.5	1.39	20.3 ± 0.3	19.9–20.7	1.45
LSD <sub>0.05</sub>		0.92			0.50		

Further analysis revealed that the differences in predictions of protein and oil content between ZX-50 or ZX-50 MC and DA 7250 were significantly correlated with seed sizes. For ZX-50, correlation coefficients between the differences and 100-seed weight were 0.827 in protein and 0.689 in oil, and for ZX-50 MC, the correlation coefficients were 0.790 in protein and 0.745 in oil. Different sizes of seed may lead to a changed interspace within the sample and thus affect the NIR detecting and collection of spectra [27]. Inclusion of genotypes covering a larger range of variation in seed size should help in the bias formulating. Therefore, the linear regression analysis was conducted based on the 100-seed weight and the differences of ZX-50 methods from DA 7250. As a result, the linear regression equations were formulated as:

$$y = 0.240x + 0.898 \quad (2)$$

for protein and

$$y = 0.094x + 0.498 \quad (3)$$

for oil for ZX-50, and

$$y = 0.329x + 0.576 \quad (4)$$

for protein and

$$y = 0.131x + 0.913 \quad (5)$$

for oil for ZX-50 MC,

where  $y$  = the difference between the predictions of ZX-50 methods and DA 7250, and  $x$  = the 100-seed weight.

Then, the linear regression equations were used to calculate the biases for individual genotypes and correct the predictions of protein and oil content accordingly. Based on the results of calibration

with 20 entries, the regression equations were developed to correct the original data of predictions as follows:

$$y = OV + (0.24x + 0.90) \quad (6)$$

for protein content, and

$$y = OV - (0.09x + 0.50) \quad (7)$$

for oil content with ZX-50; and

$$y = OV + (0.33x + 0.6) \quad (8)$$

for protein content, and

$$y = OV - (0.13x + 0.9) \quad (9)$$

for oil content with ZX-50 MC,

where  $y$  = the calibrated value,  $OV$  = the original value, and  $x$  = the 100-seed weight.

Subsequently, a joint ANOVA by combining all DA 7250 data and the corrected ZX-50 and ZX-50 MC data was performed. The results indicated that the difference between the methods were insignificant for both protein and oil content after data correcting, with  $F = 2.14$  ( $p = 0.097$ ) and  $0.50$  ( $p = 0.682$ ), respectively (Table 1). Tables 4 and 5 present the means and variations of the corrected values of protein and oil content predicted by ZX-50 and ZX-50 MC, respectively. After correcting, the predicted values by ZX-50 and/or ZX-50 MC methods were very similar or close to the data of DA 7250 for both protein and oil in most of the genotypes (Tables 3–5). It suggested that the calibrated predictions by ZX-50 methods could be regarded as an equivalent to that of DA 7250. By DA 7250 prediction, the coefficients of variation within a single genotype for protein content varied from 0.78% to 2.58%, with an average of 1.81%, and for oil content they averaged 2.21% with a range of 1.34–3.75% (Table 3). Differences between two extreme observations for a single genotype averaged  $2.50 \text{ g } 100 \text{ g}^{-1}$  for seed protein content and  $1.48 \text{ g } 100 \text{ g}^{-1}$  for oil content. In addition, the differences in protein content between two extreme observations for a single genotype averaged  $1.66 \text{ g } 100 \text{ g}^{-1}$  for ZX-50 and  $1.30 \text{ g } 100 \text{ g}^{-1}$  for ZX-50 MC, which were smaller than that with DA 7250 ( $2.50 \text{ g } 100 \text{ g}^{-1}$ ). For oil content, the differences between two extreme observations with a single genotype averaged  $1.71 \text{ g } 100 \text{ g}^{-1}$  for ZX-50 and  $1.10 \text{ g } 100 \text{ g}^{-1}$  for ZX-50 MC, which were similar to that with DA 7250 ( $1.48 \text{ g } 100 \text{ g}^{-1}$ ). Overall, the coefficients of variation within a single genotype for protein content averaged 1.00%, ranging from 0.54% to 2.68% for ZX-50, and 1.26%, ranging from 0.41% to 2.11% for ZX-50 MC (Table 4). For oil content, the coefficients of variation within a single genotype averaged 2.32% for ZX-50 and 2.24% for ZX-50 MC (Table 5), which were similar to that with DA 7250 in spite of a little larger coefficient of variation for a few genotypes. These results suggested that for the corrected predictions of protein and oil content by either ZX-50 or ZX-50 MC, the coefficients of variation within individual genotypes were also comparable to those of DA 7250.

### 3.3. Repeatability and Correlation

To further compare the consistency between ZX-50 or ZX-50 MC and DA 7250 in predicting protein and oil content, the repeatability of traits was computed based on the results of ANOVA. As shown in Table 2, the repeatability of traits was very high for all the analysis methods used. For the corrected predictions by ZX-50 and ZX50-MC, it was 98.19–99.87%, highly comparable to that of DA 7250. Correlations between ZX-50 or ZX-50 MC and DA 7250 in predicting protein and oil content were also very high. After correcting the predictions, the correlation coefficients were 0.918–0.970 (Table 2).

**Table 4.** Mean, range and coefficient of variation of corrected protein content (g 100 g<sup>−1</sup>) determined using ZX-50 NIR analyzer in 20 soybean genotypes/entries and the difference from the values with DA 7250.

Genotype	ZX-50				ZX-50 MC			
	Mean	Range	CV (%)	Difference	Mean	Range	CV (%)	Difference
PI 549058	43.4 ± 0.4	42.6–44.1	0.94	2.0	43.5 ± 0.5	43.1–44.5	1.25	1.8
AGS 346	44.2 ± 1.2	40.3–45.4	2.68	−0.9	45.5 ± 0.6	44.9–46.3	1.32	−2.3
IA 1008LF	40.2 ± 0.3	39.9–41.0	0.72	−0.1	40.4 ± 0.3	40.0–40.7	0.76	−0.3
NC 346	43.7 ± 0.3	43.1–44.1	0.67	−1.0	43.1 ± 0.5	42.5–43.8	1.12	−0.4
VS12-0203	37.7 ± 0.2	37.2–38.1	0.65	0.3	38.3 ± 0.6	37.2–38.7	1.64	−0.3
VS12-0205	33.7 ± 0.4	32.9–34.6	1.13	−1.4	35.6 ± 0.2	35.3–35.9	0.61	−3.3
VS12-0069	45.8 ± 0.4	45.0–46.4	0.97	1.3	44.4 ± 0.6	43.4–44.8	1.26	2.7
VS11-0022	45.4 ± 0.7	43.8–46.3	1.47	1.1	46.1 ± 0.3	45.8–46.6	0.71	0.4
NC Green	47.0 ± 0.8	44.6–47.9	1.77	0.3	46.9 ± 0.8	46.1–48.1	1.72	0.4
Osage	42.8 ± 0.3	42.1–43.5	0.81	1.4	42.0 ± 0.6	41.2–42.7	1.38	2.1
NC Raleigh	37.1 ± 0.2	36.8–37.4	0.54	−1.1	38.1 ± 0.7	37.1–38.7	1.73	−2.1
N6202-8	46.1 ± 0.6	44.3–46.7	1.21	0.7	45.4 ± 0.3	45.1–45.7	0.67	1.4
Asmara	41.2 ± 0.3	40.5–41.7	0.74	−0.4	41.4 ± 0.9	40.0–42.2	2.11	−0.6
Ellis	38.5 ± 0.2	38.0–38.9	0.63	−0.3	38.5 ± 0.2	38.3–38.7	0.41	−0.4
VS22-450	41.2 ± 0.4	40.2–41.9	1.06	0.1	40.2 ± 0.7	39.6–41.3	1.69	1.1
VS12-0128	41.0 ± 0.3	40.5–41.5	0.65	−2.2	41.7 ± 0.8	40.8–42.5	1.81	−3.0
NC 346_(a)	43.9 ± 0.3	43.1–44.4	0.68	−0.8	44.0 ± 0.4	43.6–44.5	0.80	−0.9
Osage_(a)	41.5 ± 0.4	40.7–42.1	0.86	1.3	40.8 ± 0.6	39.7–41.3	1.50	2.0
N6202-8_(a)	45.7 ± 0.4	45.0–46.1	0.94	−0.3	44.8 ± 0.7	43.9–45.5	1.59	0.6
Ellis_(a)	39.8 ± 0.3	39.1–40.6	0.81	0.2	40.0 ± 0.4	39.4–40.5	1.07	0.0
LSD <sub>0.05</sub>	0.37				1.03			

**Table 5.** Mean, range and coefficient of variation of corrected oil content (g 100 g<sup>−1</sup>) determined using ZX-50 NIR analyzer in 20 soybean genotypes/entries and the difference from the values with DA 7250.

Genotype	ZX-50				ZX-50 MC			
	Mean	Range	CV (%)	Difference	Mean	Range	CV (%)	Difference
PI 549058	18.1 ± 0.4	17.4–18.9	2.40	1.1	18.4 ± 0.2	18.1–18.7	1.25	1.5
AGS 346	18.7 ± 0.9	16.0–19.4	4.96	0.5	18.4 ± 1.1	17.4–20.2	5.81	0.2
IA 1008LF	18.8 ± 0.3	18.2–19.2	1.41	0.2	19.4 ± 0.4	18.8–19.9	2.23	0.9
NC 346	18.0 ± 0.3	17.5–18.7	1.94	−1.1	18.8 ± 0.3	18.5–19.2	1.53	−0.4
VS12-0203	23.1 ± 0.4	22.5–24.2	1.78	0.0	22.2 ± 0.7	21.0–22.8	3.20	−1.0
VS12-0205	21.6 ± 0.2	21.3–22.0	0.91	0.6	21.9 ± 0.2	21.6–22.1	1.04	0.9
VS12-0069	16.9 ± 0.5	15.9–18.0	2.93	0.3	17.5 ± 0.3	17.2–18.0	1.83	0.9
VS11-0022	18.5 ± 0.7	17.3–19.5	3.69	1.1	17.6 ± 0.4	17.1–18.0	2.08	0.2
NC Green	17.3 ± 0.8	15.2–18.6	4.80	−0.5	16.8 ± 0.4	16.1–17.1	2.43	−0.9
Osage	20.1 ± 0.3	19.7–20.6	1.36	0.3	19.5 ± 0.8	18.4–20.5	4.31	−0.2
NC Raleigh	22.5 ± 0.2	22.2–23.1	0.96	−0.4	21.8 ± 0.3	21.3–22.2	1.54	−1.1
N6202-8	17.0 ± 0.8	15.8–19.5	4.63	0.0	17.6 ± 0.2	17.3–17.8	1.18	0.5
Asmara	18.9 ± 0.4	18.3–19.6	1.92	1.2	18.9 ± 0.7	18.4–20.1	3.67	1.2
Ellis	21.8 ± 0.4	21.3–22.9	1.82	−0.5	21.7 ± 0.3	21.5–22.2	1.20	−0.6
VS22-450	21.3 ± 0.5	19.9–22.0	2.33	−0.4	21.2 ± 0.4	20.5–21.4	1.81	−0.5
VS12-0128	19.8 ± 0.3	19.3–20.3	1.44	−0.1	19.7 ± 0.6	18.9–20.5	2.90	−0.2
NC 346_(a)	19.0 ± 0.3	18.4–19.6	1.67	−0.9	18.9 ± 0.4	18.3–19.4	2.20	−1.0
Osage_(a)	20.6 ± 0.3	20.1–21.4	1.48	0.6	20.4 ± 0.2	20.0–20.5	1.06	0.4
N6202-8_(a)	17.4 ± 0.5	16.4–18.3	2.71	−0.6	17.8 ± 0.3	17.4–18.1	1.62	−0.2
Ellis_(a)	20.5 ± 0.2	19.7–20.7	1.20	0.2	20.3 ± 0.4	19.7–20.7	2.00	0.0
LSD <sub>0.05</sub>	0.35				0.99			

In addition, different sources of seed also exhibited a high consistency in the trend of predictions for the same genotypes. Consistent with DA7250, ZX-50 and ZX-50 MC after data correcting showed a lower protein content in Ellis, NC 346, Osage\_(a) and N6206-8\_(a) than that in the same genotypes with a different sample source Ellis\_(a), NC 346\_(a), Osage and N6206-8 (Tables 3–5). Compared to Ellis, a 1.3–2.0% lower oil content in Ellis\_(a) was consistently predicted by all the three NIR methods. It suggested that the corrected predictions of ZX-50 methods could not only discriminate the difference



between genotypes but also could detect the small difference within a same genotype between different seed sources.

### 3.4. Validation and Application of ZX-50 MC with Data Correcting in Bulk-Seed Samples

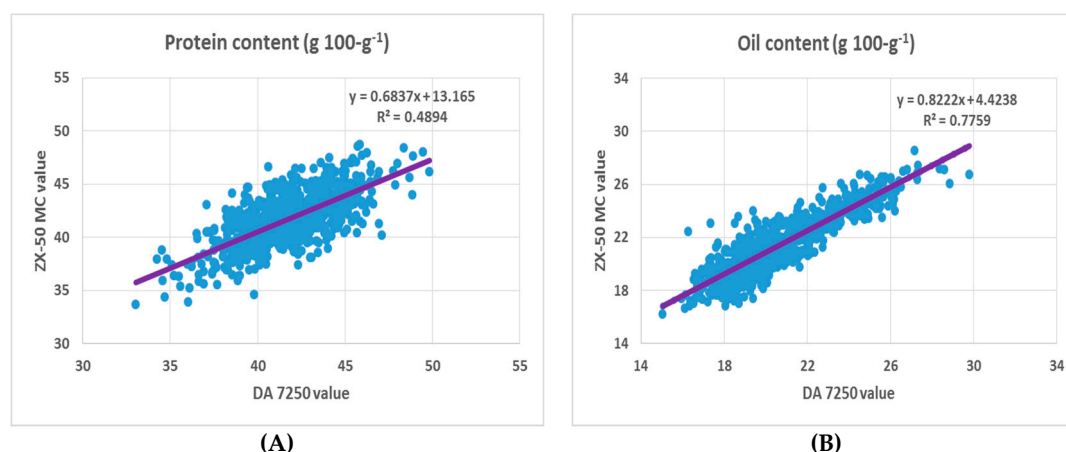
Using ZX-50 MC with the established equations for biases as described above, we evaluated a total of 760 samples in four batches of different bulk-seed samples for seed protein and oil content and compared the results with DA 7250, to validate the correctness and reliability and to explore the application of ZX-50 MC in breeding. Each set of the samples had a large variation in seed size (Table 6). For the mature seed samples, 100-seed weights averaged 20.1 g with a range of 9.5–27.5 g for 2015, and 20.3 g ranging from 10.7 to 30.3 g for 2016. For the dried edamame seed samples, ranges of variation in 100-seed weight were 7.0–20.5 g for 2015 and 7.5–23.5 g for 2016, with an average of 12.8 and 12.6 g, respectively.

**Table 6.** Application of the ZX-50 MC method to evaluate soybean seed protein and oil content in different soybean seed samples and comparison with DA 7250.

Sample	100-Seed Weight (g)		Method	Protein Content (g 100 g <sup>-1</sup> )			Oil Content (g 100 g <sup>-1</sup> )		
	Mean	Range		Mean	Range	Correlation with DA 7250	Mean	Range	Correlation with DA 7250
Mature seed 2015 (117 lines, <i>n</i> = 234)	20.1 ± 3.6	9.5–27.5	DA 7250	43.5 ± 1.8	38.6–48.9		19.3 ± 1.3	16.8–22.5	
			Corrected ZX-50 MC	43.2 ± 1.5	39.9–48.2	0.732 **	20.4 ± 1.3	18.2–24.3	0.846 **
Mature seed 2016 (147 lines, <i>n</i> = 147)	20.3 ± 4.2	10.7–30.3	DA 7250	41.9 ± 2.2	33.0–48.7		19.1 ± 1.5	15.1–22.8	
			Corrected ZX-50 MC	41.6 ± 2.1	33.7–47.1	0.823 **	18.9 ± 1.2	16.2–21.8	0.873 **
Edamame seed 2015 (88 lines, <i>n</i> = 138)	12.8 ± 2.2	7.0–20.5	DA 7250	41.1 ± 2.1	34.5–46.0		20.2 ± 1.5	17.4–24.5	
			Corrected ZX-50 MC	43.5 ± 1.9	38.0–48.7	0.830 **	21.6 ± 1.1	19.4–25.8	0.791 **
Edamame seed 2016 (127 lines, <i>n</i> = 241)	12.6 ± 2.4	7.5–23.5	DA 7250	40.6 ± 2.3	34.6–47.1		23.2 ± 1.9	17.8–29.8	
			Corrected ZX-50 MC	39.5 ± 1.6	34.8–44.6	0.765 **	23.8 ± 1.4	18.5–26.7	0.870 **
Combined samples ( <i>n</i> = 760)	16.4 ± 5.0	6.0–30.3	DA 7250	41.8 ± 2.5	33.0–49.8		20.6 ± 2.5	15.1–29.8	
			Corrected ZX-50 MC	41.8 ± 2.5	33.7–48.7	0.700 **	21.4 ± 2.3	16.2–28.5	0.881 **

\*\* Significant at *p* < 0.01.

ANOVA and/or paired t-tests indicated that the differences between ZX-50 MC and DA 7250 were mostly insignificant in the prediction of seed protein and oil content. The averages of predictions of seed protein and oil content by both ZX-50 MC with data correcting and DA 7250 are presented in Table 6. The averages of predicted values by ZX-50 MC were similar or comparable to those of DA 7250 in most cases. With all the samples combined together, the ranges of variation for the corrected predictions by ZX-50 MC were comparable to those of DA 7250 as well (Table 6). The correlation coefficients between ZX50 MC and DA 7250 were 0.732–0.873 for individual sample sets (Table 6). With all the samples being combined together, the ranges of variation for the corrected predictions by ZX-50 MC were also comparable to those of DA 7250, and the correlations between ZX-50 MC and DA 7250 were high for both predictions (*r* = 0.700 for protein and 0.881 for oil) (Figure 1). The results confirmed that ZX-50 MC could produce a good prediction of protein and oil content that was similar or comparable to those of DA 7250 in most cases.



**Figure 1.** Correlation between near-infrared (NIR) analysis methods DA 7250 and ZX-50 MC in predicting soybean seed protein (A) and oil (B) content.

Comparatively, a little less accurate prediction was observed for the dried edamame seed. It is understandable because no dried edamame sample was included in formulating the bias equations and this seed type with different shape might have an impact to some extent [27]. It might be also partially due to some samples with a very low 100-seed weight that was out of the range of variation among the 20 entries used to establish the bias equations. It suggested that more samples with broader variation of seed size and shape would help in developing new calibrations for accurate predictions.

### 3.5. Validation and Use of ZX-50 MC in Single Plant Evaluation

In soybean, NIR measurements of protein and oil have been extensively used in basic research [17,28,32,43]. These assessments have also been employed in the USDA Uniform Soybean Tests and Regional Quality Testes [44–46]. However, uses of NIR analysis of protein and oil for single plant selection have been rarely reported in practical breeding. Additionally, many NIR analyzers, such as Perten DA 7200 series and FOSS Infratec Grain Analyzers widely used in soybean research [17,32,45,46], need a lab-space and cannot be used directly in a non-lab place like seed threshing and processing room or the field. It is of interest to explore a fast, reliable, easy and portable non-destructive method that is particularly suitable for selection of single plants in breeding.

To explore the suitability and practical application of the ZX-50 MC method to single plant selection in soybean breeding, we analyzed 240 whole seed samples of single plants selected from breeding materials of maturity group IV, V and VI for protein and oil content using ZX-50 MC and DA 7250. Of these plants, the 100-seed weight averaged 20.3 g, varying from 8.9 to 37.8 g. The original predictions by ZX-50 MC averaged 35.2 and 23.6 g 100 g<sup>-1</sup> for protein and oil content, respectively, which presented a difference of 7.5 and 3.3 g 100 g<sup>-1</sup> from that of DA 7250 (Table 7). However, the predictions of ZX-50 MC were substantially improved using the established regression equation as described above. The calibrated values of ZX-50 MC were very similar to DA 7250, with an average of 42.5 g 100 g<sup>-1</sup> for protein and 20.1 g 100 g<sup>-1</sup> for oil content, compared to 42.8 and 20.3 g 100 g<sup>-1</sup> of DA 7250. No significant difference was revealed by t-test in pairwise comparison between DA 7250 and the corrected predictions of ZX-50 MC. Moreover, the calibration also increased the range of variation and the correlation coefficients between ZX-50 MC and AD 7250 predictions (Table 7). The range or difference between the extreme plants was increased from 8.1 to 12.6 g 100 g<sup>-1</sup> for protein and 5.6 to 6.2 g 100 g<sup>-1</sup> for oil content, compared to that of DA 7250 (13.1 and 7.5 g 100 g<sup>-1</sup>). The correlation coefficients between ZX 50-MC and AD 7250 were 0.777 for protein and 0.756 for oil content, respectively, compared to 0.632 and 0.653 with the uncorrected data.

**Table 7.** Application of the ZX-50 MC NIR method to single-plant evaluation of seed protein and oil content in 240 whole seed samples of soybean.

Method	Protein Content (g 100 g <sup>-1</sup> )			Oil Content (g 100 g <sup>-1</sup> )		
	Mean	Range	Correlation with DA 7250	Mean	Range	Correlation with DA 7250
DA 7250	42.8 ± 2.3	37.3–50.4		20.3 ± 1.2	16.5–24.0	
ZX-50 MC	35.2 ± 1.3	32.5–40.6	0.632 **	23.6 ± 0.9	20.4–26.0	0.635 **
Corrected ZX-50 MC	42.5 ± 2.2	37.6–50.2	0.777 **	20.1 ± 1.2	16.5–22.7	0.756 **

\*\* Significant at  $p < 0.01$ .

We also noticed that the range of variation (8.9–37.8 g) in 100-seed weight among the 240 single plants was larger than those of all other sample sets used in this study, and was beyond the ranges reported previously in most other studies with few exceptions [39–41]. That said, the ZX-50 MC with established calibration could still provide reliable and accurate predictions of seed protein and oil content. Interestingly, no significant correlations were found between 100-seed weight and the original (uncorrected) predictions of protein ( $r = -0.054$ ) or oil content ( $r = -0.020$ ) in these single plants. However, after calibrating the predictions by ZX-50 MC, the correlations between 100-seed weight and protein and oil content were significant,  $r = 0.810$  and  $-0.626$ , respectively, consistent with the results of DA 7250 and most other reports [2,39,41]. It indicated that the calibrations developed for the ZX-50 analyzer were appropriate and helped to reveal the real relationships between seed size and protein and/or oil content in soybean. These results further demonstrated that as a fast and simple method, the estimation of seed protein and oil content with ZX-50 NIR analyzer would be comparable to DA 7250 but more suitable and usable for single plant selection in practical breeding.

### 3.6. Other Issues

For plant breeding application, whether a method is suitable depends on if it can detect the differences in phenotypic performance among the genotypes. It appears that DA 7250 and ZX-50 NIR analyzer would be suitable for evaluation and selection of seed protein and oil content since both can reveal the differences among the genotypes. However, accurate and precise measurements are preferred to obtain information that can be referred, particularly for a comparison among studies by different researchers. With the original calibration installed by the manufacturer, DA 7250 provides an accurate and precise prediction of seed protein and oil similar to the wet-chemistry measurement. Thus DA 7250 could be used as an alternate of the wet-chemistry technique in evaluation of seed protein and oil content. However, the original predictions of ZX-50 exhibited significant difference or errors, implying that it should be further calibrated in practical uses. This study was an attempt to explore and demonstrated the applicability of ZX-50 NIR analyzer in evaluation and selection of single plants for seed protein and oil content in soybean. Importantly, it reported, for the first time, that the errors of ZX-50 NIR analyzer in prediction of seed protein and oil content in soybean varied with genotypes, and the differences were associated with 100-seed weight, which was not observed with DA 7250 compared to the wet-chemistry method. Consequently, the bias corrections of predictions were developed based on 20 genotypes, which presented a range of 32.3–47.3 g 100 g<sup>-1</sup> in protein content and 16.6–23.1 g 100 g<sup>-1</sup> in oil content, and 12.0–33.7 g for 100-seed weight. The calibrations were validated in a total of 760 bulk samples from different seed types (mature seeds and dried edamame seeds) and environments or sources. These samples covered a larger variation of the traits: 33.0–49.8 g 100 g<sup>-1</sup> for protein and 15.1–29.8 g 100 g<sup>-1</sup> for oil content. Moreover, their suitability for single plant analysis was directly confirmed in 240 single plants of three maturity groups and with a large range of variation in protein and oil content as well as in seed size. Therefore, this work should be referred to other materials and/or studies. The proposed approach appears to be applicable to soybeans harvested in different years and/or locations as multiple-source samples were included. We would also suggest that the manufacturer should take the impact of seed size on the predictions into consideration when new calibrations are developed for the equipment.

In the present study, a difference between the two different sample sizes (i.e., the regular and modified sample cups) used in the same NIR analyzer ZX-50 was observed (Table 2), suggesting that sample sizes might have an influence on the NIR predictions of seed composition [27]. We also noticed that there was a small or slight difference between the different sample cups used in DA 7250 with the same calibration for the same source of seeds, in particular for the micro mirror module that allows analysis of a very small sample as a few and even only a single seed [31]. These observations implied that it would be helpful for more accurate predictions to establish a separate calibration for different sizes of sample cups.

Due to high-cost and relatively clean operational conditions needed, it seems not appropriate to use DA 7250 in a seed threshing and processing area. Instead, ZX-50 is a better option that can allow immediate evaluation and selection for seed protein and oil composition. As discussed above, data of seed sizes or 100-seed weight is needed in the calibration of predictions by ZX-50 MC. In practice, 100-seed weight might not be so easily determined directly in the field. This restriction would preclude its use in the field selection. However, if there are a single-plant thresher and a scale available in the field, it would not be impossible to use a ZX-50 analyzer powered with battery in the field or in a moving truck.

In addition, the influence of seed colors on prediction of seed composition should also be considered in use of NIR measurements because the seed colors may interfere or prevent the light transmittance and/or the spectra capture [2]. In our experience, the prediction of DA 7250 was not affected by the seed colors. However, ZX-50 analyzer did not apply to deep-color seed like black and brown seeds, but it could analyze the seeds of yellow, green and other light colors. Therefore, deep color seeds were excluded in the present study.

#### 4. Conclusions

There were highly significant differences among the genotypes in both seed protein and oil content in soybean for each of the methods used in this study, and the differences between wet-chemistry or DA 7250 and uncalibrated ZX-50 NIR methods were highly significant for both traits. DA 7250 and the wet-chemistry method exhibited no significant difference but a high correlation and consistency in evaluation of the seed protein and oil composition. DA 7250 may serve as an alternate of the wet-chemistry method in evaluation of seed protein and oil content. The differences or errors incurred with ZX-50 and ZX-50 MC methods were associated with seed sizes and could be satisfactorily corrected using the calibration formula developed accordingly. ZX-50 or ZX-50 MC with appropriate biases and correction could produce relatively accurate and reliable results of predicting seed protein and oil content, which are compared to that of DA 7250.

The established calibration for ZX-50 MC was validated in different types of seed samples and its application to single plants was further confirmed with the whole-seed samples of 240 single plants. Therefore, as a fast, simple and low-cost method, ZX-50 NIR analysis of seed protein and oil content would be applicable to soybean breeding and research, in particular suitable for single plant selection of whole seeds without damaging seeds and with no need of sample preparation. Nonetheless, one should keep in mind that to obtain accurate predictions, an appropriate bias formulation should be established in advance using a set of cultivars with divergent seed sizes and a large range of variation in the target traits. It is especially important when the materials to be analyzed have seed sizes different from those the manufacturer used in the development of the original calibrations.

**Funding:** This research was supported in part by the USDA-NIFA Evans-Allen Research Program, and the USDA-NIFA Capacity Building Grant (CBG) Program (Award Number: 2017-38821-26413).

**Acknowledgments:** Haley Berry, Kyle Epps, Rameka Tylor and Zhane Slade, the undergraduate students of Virginia State University, participated in the projects. Ronald Bowen, former Plant Science Manager retired from Virginia State University, provided technical support and help in field trials. Zaid Abraha-Eyob performed the lab wet-chemistry analysis. This article is a contribution of the Virginia State University, Agricultural Research Station (Journal Series No. 364).

**Conflicts of Interest:** The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Lee, H.; Cho, B.K.; Kim, M.S.; Lee, W.H.; Tewari, J.; Bae, H.; Sohn, S.I.; Chi, H.Y. Prediction of crude protein and oil content of soybeans using Raman spectroscopy. *Sens. Actuators B Chem.* **2013**, *185*, 694–700. [CrossRef]
2. Liu, K. *Soybeans: Chemistry, Technology, and Utilization*; Aspen Publishers, Inc.: Gaithersburg, MD, USA, 1999.
3. Huang, H.; Yu, H.; Xu, H.; Ying, Y. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *J. Food Eng.* **2008**, *87*, 303–313. [CrossRef]
4. American Association of Cereal Chemists (AACC). *Approved Methods of the AACC, Methods 46–30*; American Association of Cereal Chemists: St. Paul, MN, USA, 2000.
5. AOAC. *Official Methods of Analysis*; Association of Official Analytical Chemists: Arlington, VA, USA, 1990.
6. Font, R.; del Rio-Celestino, M.; de Haro-Bailon, A. The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. *Ind. Crop. Prod.* **2006**, *24*, 307–313. [CrossRef]
7. Weir, A.D.; Omielan, J.; Lee, E.A.; Rajcan, I. Use of NMR for predicting protein concentration in soybean seeds based on oil measurements. *JAOCS* **2005**, *82*, 87–91. [CrossRef]
8. Santos, L.R.; Zangirolami, M.D.S.; Silva, N.O.; Valderrama, P.; Março, P.H. Rapid non-invasive assessment of quality parameters in ground soybean using near-infrared spectroscopy. *Pesq. Agropec. Bras.* **2018**, *53*, 97–104. [CrossRef]
9. Singh, S.; Patel, S.; Litoria, N.; Gandhi, K.; Faldu, P.; Patel, K.G. Comparative efficiency of conventional and NIR based technique for proximate composition of pigeon pea, soybean and rice cultivars. *Int. J. Curr. Microbiol. Appl. Sci.* **2018**, *7*, 773–782. [CrossRef]
10. Grunvald, A.K.; Carvalho, C.G.P.; Leite, R.S.; Mandarino, J.M.G.; Andrade, C.A.B.; Scapim, C.A. Predicting the oil contents in sunflower genotype seeds using near-infrared reflectance (NIR) spectroscopy. *Acta Scientiarum. Agron.* **2014**, *36*, 233–237. [CrossRef]
11. Cen, H.; He, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.* **2007**, *18*, 72–83. [CrossRef]
12. Stuth, J.; Jama, A.; Tolleson, D. Direct and indirect means of predicting forage quality through near infrared reflectance spectroscopy. *Field Crop. Res.* **2003**, *84*, 45–56. [CrossRef]
13. Osborne, B.G. Applications of near infrared spectroscopy in quality screening of early-generation material in cereal breeding programmes. *J. Near Infrared Spectrosc.* **2006**, *14*, 93–101. [CrossRef]
14. Pathmell, C.A. Rapid NIR Measurement of Oil and Protein Content in Soybean. 2016. Available online: <https://oceanoptics.com/rapid-soybean-oil-protein-nir-measurement/> (accessed on 15 January 2019).
15. Patil, A.G.; Oak, M.D.; Taware, S.P.; Tamhankar, S.A.; Rao, V.S. Nondestructive estimation of fatty acid composition in soybean [*Glycine max* (L.) Merrill] seeds using near-infrared transmittance spectroscopy. *Food Chem.* **2010**, *120*, 1210–1217. [CrossRef]
16. Pazdernik, D.; Killam, A.S.; Orf, J.H. Analysis of amino and fatty acid composition in soybean seed, using near infrared reflectance spectroscopy. *Agron. J.* **1997**, *89*, 679–685. [CrossRef]
17. Wang, X.; Jiang, G.L.; Green, M.; Scott, R.A.; Song, Q.; Hyten, D.L.; Cregan, P.B. Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred populations of soybean. *Mol. Genet. Genom.* **2014**, *289*, 935–949. [CrossRef] [PubMed]
18. Zhu, Z.; Chen, S.; Wu, X.; Xing, C.; Yuan, J. Determination of soybean routine quality parameters using near-infrared spectroscopy. *Food Sci. Nutr.* **2018**, *6*, 1109–1118. [CrossRef]
19. Jiang, H.Y.; Zhu, J.; Wei, L.M.; Dai, J.R.; Song, T.M.; Yan, Y.L.; Chen, S.J. Analysis of protein, starch and oil content of single intact kernels by near infrared reflectance spectroscopy (NIRS) in maize (*Zea mays* L.). *Plant Breed.* **2007**, *126*, 492–497. [CrossRef]
20. Tallada, J.G.; Palacios-Rojas, N.; Armstrong, P.R. Prediction of maize seed attributes using a rapid single kernel near infrared instrument. *J. Cereal Sci.* **2009**, *50*, 381–387. [CrossRef]
21. Long, D.S.; Engel, R.E.; Siemens, M.C. Measuring grain protein concentration with in-line near infrared reflectance spectroscopy. *Agron. J.* **2008**, *100*, 247–252. [CrossRef]



22. AACC International. Near-infrared reflectance method for whole-grain analysis in soybeans. In *AACC Approved Methods of Analysis*, 11th ed.; Method 39-21.01; Cereals & Grains Association: St. Paul, MN, USA, 2010.
23. Biancolillo, A.; Marini, F. Chapter Four—Chemometrics Applied to Plant Spectral Analysis. In *Vibrational Spectroscopy for Plant Varieties and Cultivars Characterization, Comprehensive, Analytical Chemistry*; Lopes, J., Sousa, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2018; Volume 80, pp. 69–104. [\[CrossRef\]](#)
24. Agelet, L.E.; Ellis, D.D.; Duvick, S.; Goggi, A.S.; Hurburgh, C.R.; Gardner, A.C. Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels. *J. Cereal Sci.* **2012**, *55*, 160–165. [\[CrossRef\]](#)
25. Armstrong, P.R. Rapid single-kernel NIR measurement of grain and oil-seed attributes. *Appl. Eng. Agric.* **2006**, *22*, 767–772. [\[CrossRef\]](#)
26. Spielbauer, G.; Armstrong, P.; Baier, J.W.; Allen, W.B.; Richardson, K.; Shen, B.; Settles, A.M. High-throughput near-infrared reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels. *Cereal Chem.* **2009**, *86*, 556–564. [\[CrossRef\]](#)
27. Agelet, L.E.; Hurburgh, C.R., Jr. Limitations and current applications of near infrared spectroscopy for single seed analysis. *Talanta* **2014**, *121*, 288–299. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Huber, S.C.; Li, K.; Nelson, R.; Ulanov, A.; DeMuro, C.M.; Baxter, I. Canopy position has a profound effect on soybean seed composition. *PeerJ* **2016**, *4*, e2452. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Huskey, L.L.; Snyder, H.E.; Gbur, E.E. Analysis of single soybean seeds for oil and protein. *JAOCS* **1990**, *67*, 686–688. [\[CrossRef\]](#)
30. Mosjidis, J.A.; Yermanos, D.M. Plant position effect on seed weight, oil content, and oil composition in sesame. *Euphytica* **1985**, *34*, 193–199. [\[CrossRef\]](#)
31. Perten Instruments AB. *DA 7250 Installation and Operation Manual*; Perten Instruments AB: Hagersten, Sweden, 2016.
32. Jiang, G.L.; Chen, P.; Zhang, J.; Florez-Palacios, L.; Zeng, A.; Wang, X.; Bowen, R.A.; Miller, A.; Berry, H. Genetic analysis of sugar composition and its relationship with protein, oil, and fiber in soybean. *Crop Sci.* **2018**, *58*, 2413–2421. [\[CrossRef\]](#)
33. Jiang, G.L.; Rutto, L.K.; Ren, S.; Bowen, R.A.; Berry, H.; Epps, K. Genetic analysis of edamame seed composition and trait relationships in soybean lines. *Euphytica* **2018**, *214*, 158. [\[CrossRef\]](#)
34. Smallwood, C.J.; Gillman, J.D.; Saxton, A.M.; Bhandari, H.S.; Wadl, P.A.; Fallen, B.D.; Hyten, D.L.; Song, Q.; Pantalone, V.R. Identifying and exploring significant genomic regions associated with soybean yield, seed fatty acids, protein and oil. *J. Crop Sci. Biotech.* **2017**, *20*, 243–253. [\[CrossRef\]](#)
35. Zhang, J.; Wang, X.; Lu, Y.; Bhusal, S.; Song, Q.; Cregan, P.B.; Yen, Y.; Brown, M.; Jiang, G.L. Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. *Mol. Plant* **2018**, *11*, 460–472. [\[CrossRef\]](#)
36. Zeltex, Inc. *ZX-50 Portable Grain Analyzer with AutoBIAS*; User's Manual Version 4.1.; Zeltex, Inc.: Hagerstown, MD, USA, 2007.
37. Scheffé, H. *The Analysis of Variance*; Wiley: New York, NY, USA, 1959. [\[CrossRef\]](#)
38. Sokal, R.R.; Rohlf, F.J. *Biometry*, 3rd ed.; W.H. Freeman and Company: New York, NY, USA, 1995.
39. Filho, M.M.; Destro, D.; Miranda, L.A.; Spinosa, W.A.; Carrao-Panizzi, M.C.; Montalvan, R. Relationships among oil content, protein content and seed size in soybeans. *Braz. Arch. Biol. Technol.* **2001**, *44*, 23–32. [\[CrossRef\]](#)
40. Maestri, D.M.; Labuckas, D.O.; Guzman, C.A.; Giorda, L.M. Correlation between seed size, protein and oil contents, and fatty acid composition in soybean genotypes. *Grasas y Aceites* **1998**, *49*, 450–453. [\[CrossRef\]](#)
41. Poeta, F.; Borrás, L.; Rotundo, J.L. Variation in seed protein concentration and seed size affects soybean crop growth and development. *Crop Sci.* **2016**, *56*, 3196–3208. [\[CrossRef\]](#)
42. Zhang, J.; Song, Q.; Cregan, P.B.; Jiang, G.L. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.* **2016**, *16*, 217. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Wee, C.D.; Hashiguchi, M.; Ishigaki, G.; Muguerza, M.; Oba, C.; Abe, J.; Harada, K.; Akashi, R. Evaluation of seed components of wild soybean (*Glycine soja*) collected in Japan using near-infrared reflectance spectroscopy. *Plant Genet. Resour.* **2017**, *16*, 94–102. [\[CrossRef\]](#)

44. Crochet, W.D. *Uniform Soybean Tests Northern States 2012*; USDA-ARS and Purdue University: West Lafayette, IN, USA, 2013.
45. Gillen, A.M.; Shelton, G.W. *Uniform Soybean Tests Southern States 2013*; USDA-Agricultural Research Service: Stoneville, MS, USA, 2014.
46. O'Brien, T.; Graef, G. *Regional Quality Traits Test 2013: Group 0-V*; University of Nebraska: Lincoln, NE, USA, 2014.



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).