Rapid Nondestructive Analysis of Intact Canola Seeds Using a Handheld Near-Infrared Spectrometer

Véronique J. Barthet¹ · Michael W.P. Petryk¹ · Bert Siemens¹

Abstract The near-infrared (NIR) models for canola quality were developed with samples from Canadian canola seeds harvested in 2016 and 2017. All calibration models were first tested on a 2017 external validation sample set. The handheld NIR spectrometer used in this study has a limited wavelength range 908.1–1676.2 nm; however, the validation results showed that it could be used to predict several important parameters that defined canola seed quality. Final testing was performed using calibration models with the least number of factors on a second external canola validation sample set (2018 harvest). Some calibration models showed excellent stability and predictive powers with $R^2_{\text{val}}$ values of 0.94–0.99 (i.e., oil, protein, oleic acid and iodine value) and low SEPs for both external validation sample sets. The α-linolenic acid model had an $R^2_{\text{val}}$ of 0.93 when applied to the 2017 external validation set, the correlation fell slightly to 0.88 when applied to the 2018 external validation sample set, potentially indicating a slight instability in the model. The prediction model for total glucosinolate was not very good, but still could be used to segregate the samples into low and high glucosinolate samples. Finally, the predictive models for chlorophyll and total saturates were unusable. The chlorophyll model was very unstable, likely due to the instrument’s limited wavelength range.

Keywords NIR spectroscopy · Intact seed · Handheld · NIR spectrometer · Protein · Oil · Total glucosinolate · Chlorophyll · Oleic acid · α-Linolenic acid · Total saturates · Iodine value

Introduction

In 2018, Canadian producers seeded over 9 million hectares of canola, producing over 20 million metric tonnes of canola seeds (Statistic Canada: Estimated areas, yield, production, average farm price and total farm value of principal field

Abbreviations

AOCS American Oil Chemists’ Society
ISO International Organization for Standardization
NF number of factors
NIR near-infrared
PLS partial least-square
$R^2_{\text{cross-val}}$ coefficient of determination for the cross-validation set
$R^2_{\text{cal}}$ coefficient of determination for the calibration set
$R^2_{\text{val}}$ coefficient of determination for the validation set
RPD residual prediction deviation
SE standard error
SEC standard error of calibration
SECV standard error of cross validation
SEP standard error of prediction
SNV standard normal variate

Received: 22 October 2019 / Revised: 10 January 2020 / Accepted: 10 January 2020
© 2020 Her Majesty the Queen in Right of Canada Journal of the American Oil Chemists’ Society © 2020 AOCS

Published online: 11 February 2020
J Am Oil Chem Soc (2020) 97: 577–589
crops, in metric and imperial units, 2019). Canada is the world’s largest exporter of canola seed; since 2016, over 10 million metric tonnes have been exported annually, representing about half of the annual production (Canola Council of Canada, 2019). Canola seeds are crushed to produce protein-rich meal and oil. Oil content is the main parameter to assess canola seed quality for market, but it is important to measure all seed components that will affect the overall quality of the canola oil. Once the crude canola oil is obtained, it is then refined to remove all co-extracted products that negatively affect oil quality. Due to the short growing season, Canadian canola samples can contain immature canola seeds showing high levels of chlorophyll. Chlorophyll and its breakdown products produce an undesirable color and lead to oil oxidation; it has to be removed from the crude oil during the bleaching step of the refining process (Daun and Unger, 2016). High chlorophyll content in canola seeds also increases refining costs; measuring chlorophyll content rapidly and accurately is important for crushers to be able to optimize their crushing and oil refining processes. Fatty acid composition is also an important parameter to qualify canola seed quality as some canola varieties have been developed to produce an oil with high thermal stability, ideal for frying applications. To that end, iodine value has historically been used to classify canola seed oil as it is a measure of the total degree of unsaturation of the oil. Usually, conventional canola oils have iodine values higher than 108 units, whereas high thermal stability canola seeds produce an oil with iodine values ranging from 92 to 107 units. Total saturated fatty acid content is another quality parameter that defines canola oil. The canola industry has been breeding canola to decrease its saturated fatty acid content and increase its health benefits—canola oil has the lowest saturated fatty acid content of any vegetable oil. When used as feedstock for animals, the commercial value of canola meal is mainly defined by its protein content, the higher the better, and the total glucosinolate content, the lower the better, due to the latter’s negative effect on animal health. Therefore, to define canola seed quality and its value and marketability, one must quantify its oil, protein, total glucosinolate, and chlorophyll contents as well as select fatty acid and total saturated fat content and its iodine value.

Near-infrared (NIR) spectrophotometry is commonly used by the agriculture and food industry to predict quality parameters of major and minor components of various food products, especially grains such as wheat and oilseeds (Daun et al., 1994; Liu et al., 2013; Tkachuk, 1981; Williams and Norris, 2001). Handheld NIR instruments have been successfully used to predict plum quality (Pérez-Marín et al., 2010), fish filet patty authenticity of (Grassi et al., 2018) and the protein, oil and moisture content in soybean, and corn seed and meal (Biller et al., 2014). Handheld NIR instruments often have narrower wavelength ranges than laboratory NIRs, e.g., 950–1650 nm vs. 400–2500 nm for a DS2500 (Foss) and the NIR Systems 6500. The spectral resolution is often not as good for handheld NIR instruments when compared to benchtop ones (6.2 vs. 0.5 nm for the DS2500 and 2.0 mm for the NIRSystems 6500). However, having a portable instrument that could be carried in the field or to any delivery point (e.g., processing plants, primary elevators, or terminal elevators) has numerous advantages, the most important being a cost-saving factor where seeds could be analyzed immediately upon delivery to allow their segregation for use in blending to get the desired end products.

The purpose of this research project was to develop NIR spectroscopy (NIRS) calibrations for a handheld NIR spectrometer so as to be able to rapidly predict the quality of intact canola seeds by determining seed composition in terms of oil, protein, chlorophyll, total glucosinolates, oleic acid, α-linolenic acid, total saturated fats, and the iodine value of the oil.

Materials and Methods

Samples

Canola seeds grown in Canada in 2016, 2017, and 2018 were used to develop and validate the NIR calibration models. Composite canola seed samples (181, intact whole canola seeds) from the 2016 and the 2017 harvest surveys were used for NIR model calibration development; 61 samples were 2016 individual producers samples, 11 were 2017 eastern Canada county composite samples, 56 were 2017 western Canada crop district composite samples, 46 were 2017 western Canada variety composite samples, and 7 were registered seed grower samples. All samples were prepared from seeds graded Canada No.1 Canola. Individual 2017 and 2018 producer samples (intact canola seeds) served as external verification samples to control/verify the NIRS calibration of our handheld NIR instrument; their quality parameters covered the whole range of quality parameters observed that year.

While it would be desirable for the ranges of all of the assayed values (oil, protein, etc.) to be similar in the calibration and validation dataset, this cannot be guaranteed for all constituents as validation set samples were randomly taken during harvest. Canola samples were stored in sealed containers or plastic bags at 4 °C for long-term storage or at room temperature for short-term storage (e.g., prior to acquiring the NIR spectra).

Reference Analytical Methods

Moisture Content

Moisture was determined on ground seed (2 g) dried for 3 hours in a forced air oven at 105 °C.
**Oil Content**

Crude oil content, reported as a percentage, was determined using a modified ISO 659, 2009 reference method, using only 4 g of ground sample with an automatic immersion instrument with petroleum ether as extracting solvent.

**Protein Content**

Crude protein content was determined by the combustion method according to AOCS Official Method Ba 4e-93, 2017 (Generic Combustion Method for Crude Protein). Total nitrogen content was determined and crude protein results were calculated as percentages using N × 6.25.

**Total Glucosinolate Content**

Total glucosinolate content, reported as μmol/g seed, was determined by measuring total glucose upon hydrolysis of the glucosinolates by a myrosinase according to Heaney et al. (1988).

**Chlorophyll Content**

Total chlorophyll content was determined using the ISO 10519, 2015 method (Determination of chlorophyll content—Spectrometric method), with results expressed as mg/kg, seed basis.

**Fatty Acid Composition**

Relative fatty acid composition of the oil was determined by ISO 12966-2, 2017 (Gas chromatography of fatty acid methyl esters—Part 2: Preparation of methyl esters of fatty acids) followed by ISO 12966-1, 2014 (Gas chromatography of fatty acid methyl esters - Part 1: Guidelines on modern gas chromatography of fatty acid methyl esters) after the extraction of the oil.

**Iodine Value**

The iodine value of the oil (expressed as units), which is related to the total unsaturation of the oil, was calculated from the fatty acid composition according to AOCS Recommended Practice Cd 1c-85, 2017.

Seed composition was reported on dry basis for all parameters except chlorophyll, which was reported as is. Some of the samples of the calibration data set lacked reference data for protein, chlorophyll, and total glucosinolate contents; the calibration dataset for these quality parameters contained less than 181 samples. The model validation datasets contain 64 samples for all parameters. The 2018 ultimate/final validation sample set data contained 69 samples.

**Instrument and Spectral Data Acquisition**

All samples (calibration and validation) were treated identically to record their spectra. For each sample, five replicate NIR spectra were recorded in reflectance mode and transformed into absorbance (log 1/R) using a MicroNIR OnSite-W (Viavi Solutions Inc.) hand-held NIR spectrometer with an indium-gallium-arsenide (InGaAs) photodiode array detector and a wavelength range of 908.1–1676.2 nm in increments of approximately 6.2 nm. Each replicate spectrum was acquired using the manufacturer’s default settings of 9.4 ms integration time with 100 scans averaged into one spectrum. These replicate spectra were further averaged to produce one NIR spectrum per sample.

**Chemometric and Statistical Analysis and Software**

Calibration models were developed using The Unscrambler X software, version 10.5.1 (CAMO Analytics, Magnolia, TX, USA). A statistical analysis of the performance of the calibration on the validation dataset (e.g., predicted vs. reference values) was carried out in The Unscrambler X and expanded upon using SAS Enterprise 7.15 HF7 (SAS Institute Inc., Cary, NC, USA). Graphic representations were created using Origin 9.1.0 (OriginLab Corporation, Northampton, MA, USA).

Some reported statistical parameters such as the standard error of calibration (SEC), the coefficient of determination for the calibration set (R²_cal), the standard error of cross validation (SECV), the coefficient of determination for the cross-validation set (R²_cross-val), the standard error of prediction (SEP), and the coefficient of determination for the validation set (R²_val) were calculated within The Unscrambler X. Others parameters, such as the residual prediction deviation (RPD), were calculated using Microsoft Excel as described by Williams and Norris (2001).

**Development and Validation of Calibration Models**

Spectra were processed in their full data range (908.1–1676.2 nm), first with a standard normal variate (SNV) transform (Barnes et al., 1989) to remove multiplicative interferences such as baseline shift then with a Savitzky–Golay (S-G) second derivative transformation (Savitzky and Golay, 1964) to ensure that: (1) peak positions were maintained at the same place as in the original spectra, (2) scattering effects were removed from the background, and (3) to improve spectral resolution and therefore assist in resolving overlapping peaks. During the S-G transform, data were smoothed over four data points on either side of a symmetric kernel and a second-degree polynomial was fitted to the data points. This data treatment was consistent with recommended best practices outlined by other groups (Azzouz et al., 2003; Martens and Næs, 1991; Næs et al., 2002). Calibrations were carried out using the...
Kernel Partial Least Squares (Kernel PLS, a routine method in The Unscrambler X) (Dayal and MacGregor, 1997; Lindgren et al., 1993) method using a 20-fold (internal) cross-validation (with approximately nine samples in each of the 20 segments) in order to develop models with the number of partial least squares (PLS) factors (NF) ranging from 1 to 20.

**Graphical Representations**

All graphical representations were made using Origin 2019b (OriginLab Corporation).

**Results and Discussion**

**Sample Set Description**

Descriptive statistics for whole canola seed composition (oil, protein, total glucosinolate, and chlorophyll contents), main fatty acid composition (oleic acid, α-linolenic acid and total saturates), and iodine value of the oils are presented in Table 1 for all the canola sample sets (calibration and both external validation sets). Descriptive values were mean, standard deviation, median, minimum, and maximum, expressed on dry seed matter, except for chlorophyll content, which was expressed in mg/kg as is. The coefficient of variation (CV) indicated the variability of the sample set for the analyte of interest; chlorophyll and glucosinolate contents where the analytes with the higher CV as suggesting the largest data ranges compared to other analytes (Table 1).

Validation data sets contain individual canola samples randomly taken during the harvest, the goals being to take seeds representative of the quality of the harvest year covering the range of values observed in that year. As environmental factors influenced the quality of the harvest, it could happen that the quality parameters of a validation set in a particular year could be outside the range of the calibration dataset. For oil, protein, chlorophyll, and oleic acid, the values in the validation dataset (lower minimum and higher

**Table 1** Sample description—reference method summary for canola calibration and verification sets

| Units | Oil content | Protein content | Chlorophyll content | Total glucosinolate content | Oleic acid content | α-Linolenic acid content | Total saturates content | Iodine value of the oil |
|-------|-------------|-----------------|---------------------|-----------------------------|-------------------|------------------------|------------------------|------------------------|
| %, dry basis | %, dry basis | mg/kg, as is | μmol/g of seed, dry basis | % in oil | %, in oil | %, in oil | Units, in oil |
| Calibration set (2016, 2017 location and variety composites, 2016 individual producer samples) | | | | | | | | |
| Mean | 48.7 | 22.2 | 13.8 | 11.3 | 63.9 | 8.52 | 6.66 | 111.1 |
| Standard deviation | 2.5 | 2.2 | 11.9 | 3.6 | 3.8 | 2.66 | 0.29 | 5.5 |
| Median | 48.8 | 22.1 | 11.1 | 10.5 | 62.8 | 9.35 | 6.63 | 112.8 |
| Minimum | 41.9 | 16.5 | 1.0 | 4.9 | 56.6 | 1.49 | 5.41 | 90.8 |
| Maximum | 54.9 | 28.5 | 96.8 | 23.7 | 79.4 | 12.27 | 8.01 | 121.9 |
| N | 181 | 168 | 167 | 77 | 181 | 181 | 181 | 181 |
| External validation set (2017 individual producer samples) | | | | | | | | |
| Mean | 48.2 | 23.5 | 19.1 | 11.1 | 64.2 | 8.22 | 6.69 | 110.3 |
| Standard deviation | 3.8 | 3.8 | 24.0 | 4.2 | 4.8 | 2.91 | 0.30 | 6.4 |
| Median | 48.1 | 22.9 | 11.4 | 10.3 | 63.0 | 9.12 | 6.76 | 112.2 |
| Minimum | 40.3 | 16.4 | 1.5 | 4.4 | 54.3 | 1.68 | 5.97 | 94.2 |
| Maximum | 57.0 | 30.9 | 132.7 | 22.1 | 79.5 | 12.10 | 7.39 | 118.3 |
| N | 64 | 64 | 64 | 64 | 64 | 64 | 64 | 64 |
| Final external validation set (2018 individual producer samples) | | | | | | | | |
| Mean | 48.54 | 22.80 | 24.80 | 9.70 | 64.5 | 8.00 | 6.67 | 111.0 |
| Standard deviation | 3.20 | 3.51 | 39.78 | 3.66 | 4.49 | 2.85 | 0.27 | 6.19 |
| Median | 48.37 | 22.46 | 9.74 | 8.86 | 64.06 | 8.49 | 6.69 | 111.01 |
| Minimum | 41.6 | 16.50 | 1.3 | 3.6 | 56.7 | 1.49 | 5.90 | 92.7 |
| Maximum | 55.1 | 31.50 | 210.9 | 21.1 | 77.9 | 12.40 | 7.30 | 119.38 |
| N | 69 | 69 | 69 | 69 | 69 | 69 | 69 | 69 |
maximum) had a higher range than the value in the calibration dataset (Table 1). This led to situations where the calibration model had to extrapolate the prediction for external validation samples, which is often less favorable when testing a NIR prediction model. However, in our work, the robustness of the calibration models is such that they are able to predict quality factors for samples that are outside of the “normal calibration range.” To test the prediction ability of the various models, two verification sets were used; one made of individual producers samples harvested in 2017 and one made of individual producer samples harvested in 2018. The latter verification was crucial as no 2018 harvest samples were used in the calibration sets to develop NIR models. This 2018 sample test was to give an insight into the robustness and applicability of the various models to future harvest years.

There are some relationships between various canola quality parameters: (1) the higher the oil content, the lower the protein content (and vice versa), (2) fatty acid composition is reported in relative amounts, therefore high levels of some acids will correspond to lower levels of other acids, and (3) there is a linear relationship between iodine value and α-linolenic acid content. Iodine value is an indication of the total unsaturation of the oil; α-linolenic acid is the most unsaturated fatty acid found in canola oil, and therefore has the most pronounced effect on the iodine value. The iodine value gives a good indication of the α-linolenic acid content of the canola oil.

**Spectra and Processing**

Typical unprocessed averaged scans of canola samples from the calibration set are presented in Fig. 1a. Fig. 1c presents the same scans after an SNV transform followed by an S-G second derivative with four-point smoothing on either side of the data point (a total of nine smoothing points); these were the ones used for the various calibration developments.

**General Characteristics of the Scans and Calibrations**

In order to assist in visualization, calibration models were developed for NIR spectra that were not SNV corrected. Referred as visualization calibrations, these models did not perform as well as the calibration models reported in this study, which used SNV-transformed NIR spectra. However, the visualization calibration functions could help to illustrate the potential important spectral regions that could affect and influence the models and clarify why a prediction model failed. Fig. 2 presents the visualization calibration functions for oleic and α-linolenic acids. The region

---

**Fig. 1** Typical scans of the canola samples from the calibration set with no spectral processing (a) and after an SNV transform (b) followed by a S-G second derivative with four-point smoothing (c)

**Fig. 2** Output of the calibration/model development—Regression coefficients as a function of the wavelengths for oleic and α-linolenic acids for 6-NF prediction models
below 1110 nm showed very little contribution to both NIR visualization calibrations while the region above the instrument limit of 1676 nm could have a significant impact on the models and the lack of data in this range could lead to an under-performance in the calibration as compared to NIR spectrometers whose wavelength range extends above 1676 nm.

**Calibration Building and Testing**

**Optimum Number of Factors in Calibration Models**

Parameters were analyzed one at time using a partial least-squares regression (PLSR). Two approaches are possible to determine the optimal number of factors (NF) value to use for a model. The first one would be to select that NF which minimizes the standard error (SE) of cross-validation (SECV) and maximizes the coefficient of determination for the cross-validation set ($R^2_{\text{cross-val}}$). This approach could potentially lead to a very large NF value, with little correspondence to the physical meaning of the predicted parameter, leading toward model instability and lack of robustness (Zeaiter et al., 2004). The second approach would be to minimize NF by plotting the explained variance of the calibration as a function of NF. Once the plot plateaued, the optimal NF value is obtained; after this point, additional factors will increase the variance by 6% or more, then we had reached the plateau and the physical meaning of the predicted parameter, leading toward model instability and lack of robustness (Zeaiter et al., 2004). The second approach would be to minimize NF by plotting the explained variance of the calibration as a function of NF. Once the plot plateaued, the optimal NF value is obtained; after this point, additional factors will increase the risk of adding noise to the model (Faber and Rajko, 2007). To limit the NF, it was decided that if the addition of another factor in the calibration model did not increase the explained variance by 6% or more, then we had reached the plateau and no additional factors were included in the model. Both approaches (the 6% limit approach and the one maximizing $R^2_{\text{cross-val}}$/minimizing SECV) were used to develop the models for the various quality parameters, which were then applied to the independent validation dataset with the predicted values being compared to the reference values using a linear regression analysis. Minimizing the NF in a calibration model has been shown to be the best method to develop calibration models as it tends to increase model robustness. The final model testing and most of the discussion will center around these minimalist NF models.

The first step of the model development was to test prediction models with PLS model development with a 20-NF for internal cross-validation. This is supposed to give the best statistical results and therefore help to assess if the model could indeed be developed. The best statistical results were obtained when using a SNG second derivative of the spectra with four-point smoothing. Then, two types of models were developed; models with the minimal NF value, where the additional factor must contributes by at least 6% to a decrease the model RMSE while giving satisfactory prediction results (SEP, SECV, $R^2_{\text{cross-val}}$; and, $R^2_{\text{val}}$) and models with a higher NF but the best/optimized prediction results (SEP, SECV, $R^2_{\text{cross-val}}$, and $R^2_{\text{val}}$). Prediction results for all models are presented in Table 2.

**Oil Content**

Calibration model was developed using 181 intact canola samples (Table 2). The oil content predictions for the 2017 external validation had to be extrapolated, however, the models still performed well as shown in Table 2. The lowest NF for the model developed for the oil content calibration was four, whereas the best statistical results were obtained with a model using 13-NF in the model (Table 2). The standard error of performance (SEP) and (SECV) were similar for both models, at 0.49% and 0.44%, respectively, for the 13-NF model and 0.45% and 0.49%, respectively, for the 4-NF model (Table 2). The coefficients of determination of the 2017 validation set ($R^2_{\text{val}}$) were identical for both models, $R^2_{\text{val}} = 0.986$ (Table 2), Fig. 3a presenting the scatter plot of 2017 verification set for the 4-NF model. Both models also had identical RPD results (7.7); bias was −0.06 and –0.20 for the 13-NF and the 4-NF models, respectively. All of these results indicated that both oil calibration models were very good and both models could be used for any applications (Williams and Norris, 2001).

These results agreed with previously reported results using benchtop instruments. Daun et al. (1994) reported SEPs ranging from 0.43 to 0.55% and $R^2_{\text{val}}$ ranging from 0.963 to 0.979 for canola (Brassica napus and Brassica rapa) seeds with oil contents ranging from 39.8 to 49.9%, whereas Petisco et al. (2010) reported an SEP of 0.54% and an $R^2_{\text{val}}$ 0.98 for Brassica seeds (B. napus and Brassica carinata) with a higher oil content range (34.1–48.4%). Oblath et al. (2016) reported only cross-validation results; the SEC was 0.90% and $R^2_{\text{cross-val}}$ 0.98 for the oil prediction results of a model developed using various seed samples of the Brassicaceae family with an oil content range of 17.6–51.0%. The NF used in the two models also compared positively to what has been previously reported, as Daun et al. (1994) reported using 13-NF, whereas Petisco et al. (2010) used only six NF for their model. Increasing the NF for oil content prediction did not really improve the prediction results (Table 2), the 4-NF model was the chosen/recommended calibration model for further application as it is likely more robust than the 13-NF model, due do the lower NF in the equation.

Finally, the 4-NF oil NIR model was tested on individual canola samples from the 2018 harvest. The results (Table 2) indicated that the NIR model for oil was stable with excellent performance —SEP = 0.59%, $R^2_{\text{val}} = 0.985$, RPD = 5.41 & bias = 0.20— these values compared well with those obtained from the external validation 2017 producer samples (Table 2). This verified that the handheld
The models were developed with 168 calibration samples; the statistics using 64 samples from the 2017 harvest are presented in Table 2. The best SEP (0.50%) and $R^2_{\text{cal}}$ of 0.985 were obtained for a model using 7-NF. They were, however, only marginally better than the SEP and the $R^2_{\text{cal}}$ obtained with the (lowest NF) 5-NF model, 0.55% and 0.984, respectively (Table 2 and Fig. 3b). The RPD value for this 5-NF protein model was 6.8 (6.6 for the 7-NF model), confirming the $R^2_{\text{cal}}$ result. These results were very satisfactory, suggesting that the 5-NF model could be used for most applications, including quality assurance programs according to Williams and Norris (2001).

Our results compared well with previously reported results; Daun et al. (1994) reported an SEP of 0.43% and $R^2_{\text{cal}}$ of 0.99 for protein contents ranging from 19.7 to 29.8%, while Petisco et al. (2010) reported an SEP 0.57% and $R^2_{\text{cal}}$ 0.96 for a protein content range of 14.5 to 31.2%; both used benchtop instruments. Our RPD of 6.6 also compared favorably to the value of 5.0 obtained by Petisco et al. (2010) and was similar to the RPD of 6.5 reported by Daun et al. (1994). These results were obtained with a model using only 5-NFs as compared to Daun et al. (1994) who used eight factors in their protein model.
prediction model while Petisco et al. (2010) used six factors in their model.

The stability and robustness of the 5-NF protein model were tested by analyzing 69 samples from the 2018 Canadian canola harvest survey. The statistical results are presented in Table 2. This 5-NF protein model was stable with excellent performance, $R^2_{\text{val}} = 0.980$ & SEP = 0.50%, these values were nearly identical to the ones obtained with the 2017 canola verification set ($R^2_{\text{val}} = 0.984$ & SEP = 0.55%).

**Total Glucosinolates Content**

Two prediction models were developed and tested on a 64 canola verification set. The lowest NF for the glucosinolates prediction model was three, whereas the best statistical results were obtained with a 13-NF model (Table 2). The statistical results for 3-NF model were not very good; SECV (2.9 μmol/g) and SEP (3.5 μmol/g) were quite large, while $R^2_{\text{cross-val}}$ (0.359) and $R^2_{\text{val}}$ (0.309, Fig. 3c) were low. The RPD for the 3-NF total glucosinolate prediction model were 1.2, also indicating that the NIR calibration accuracy was indeed very poor and its use not recommended (Williams and Norris, 2001). The 13-NF model developed only based on statistical analysis led to an $R^2_{\text{val}} = 0.684$, which according to Williams and Norris (2001) could still be used for material screening; however, the RPD of 1.4 classified it as a very poor model that should not be used.

Daun et al. (1994) used 11 and 13 factors in their models with benchtop instruments for a total glucosinolates content range of 9.7–30 μmol/g (B. napus and B. rapa); their SEP

**Fig. 3** Scatter plots of reference data and NIR predicted values form the 2017 external validation for the estimation of the oil, protein glucosinolate, chlorophyll contents plus oleic acid, α-linolenic acid content, and iodine value of intact canola (B. napus) seeds
and $R^2_{\text{val}}$ ranged from 1.7 to 3.8 μmol/g and 0.812 to 0.899, respectively, depending on the instrument model. Oblath et al. (2016) used various Brassica species with a much larger total glucosinolates range (5.5–117.4 μmol/g); their SECV of 8.19 μmol/g was higher than the one obtained with our handheld instrument. Petisco et al. (2010) used $B. napus$ and $B. carinata$ to develop a global total glucosinolates model (range 15.8–97.9 μmol/g) using only seven factors. They further developed a prediction model for only $B. napus$ (range 16.8–39.0 μmol/g) obtaining an SEP of 2.07 μmol/g, an RPD of 2.1 and an $R^2_{\text{val}}$ of 0.83. The reported SEP by these various authors were similar to the ones we obtained using both models (2.3 and 3.5 μmol/g, Table 2). However, our $R^2_{\text{val}}$ were lower than the ones reported above; the 3-NF calibration model being poor (0.243) when compared to the 13-NF calibration model (0.684). It seemed that our 13-NF model compared well with Daun et al. (1994) and even Petisco et al. (2010) models. The low $R^2_{\text{val}}$ was likely due to the narrower total glucosinolates range of the samples used for this project (4.4–23.7 μmol/g). This suggested that for our total glucosinolates model developing a prediction model with the lowest NF as possible (the subsequent factor must increase the explained variance by 6% or more) was restrictive leading to a poor model, this assumption was not valid in that case and would benefit from further study.

The stability of the glucosinolate NIR model was tested by applying it to 2018 canola harvest samples (Table 2). The 3-NF prediction model for glucosinolates performed poorly, SEP = 2.52 μmol/g and $R^2_{\text{val}}$ = 0.201, these results were very similar to the ones obtained with the 2017 external validation sample set values (SEP = 3.52.52 μmol/g, $R^2_{\text{val}}$ = 0.309). The 13-NF total glucosinolates model was also tested but its performance was worse than the 3-NF calibration model and therefore not reported.

Kumar et al. (2010) showed that the optimum spectral region for total glucosinolates prediction was 7502.1–5444.6 cm⁻¹ (1333–1837 nm) and 4601.6–4246.7 cm⁻¹ (2173–2355 nm). Daun et al. (1994) used 890, 1016, 1014, 1776, 2139, and 2230 nm wavelengths for the NIR Systems 6500 and 696, 1360, 1644, 1680, 1759, and 1778 nm wavelengths for the PerCon Infracram 8144. The handheld NIR spectrometer with its 908.1–1676.2 nm range, only overlapped with those wavelengths in the 1333–1676.2 nm region; this also could explain the poor results when compared to other reported ones.

Although the glucosinolate NIR calibration could be qualified as “poor” or “very poor” according to Williams and Norris (2001), it is important to take into consideration the context in which a handheld NIR device could be used in the field for total glucosinolate determination. By definition, canola can contain no more than 30 μmol/g seed of total glucosinolates (8.5% moisture). It is relevant to a buyer to know if the canola seeds, they are purchasing, are really canola seeds or other rapeseeds. The calibration curve in Fig. 3c, representing the 3-NF calibration model, can still be deemed “fit for purpose” if it is used to segregate seeds into those at the upper range of the 30 μmol/g limit vs. seeds that are much lower in glucosinolates content.

**Chlorophyll Content**

The calibration model had to extrapolate the prediction results for samples: the maximum chlorophyll content was 133 mg/kg for the validation set vs. 97 mg/kg for the calibration set (Table 1). Unfortunately, chlorophyll content of Canadian canola seed is highly variable from 1 year to another (Barthet, 2018); it is very difficult to predict the chlorophyll content range of the verification sample set from 1 year to another, as canola samples selected for a verification/calibration dataset during one growing season may not be representative of subsequent seasons due to the growing conditions. A 6-NF model was developed, but its performance was poor (Table 2 and Fig. 3d); the SEP (21 mg/kg) was significantly higher than both SECV (2.9 mg/kg) and SEC (0.82 mg/kg). With an $R^2_{\text{cross-val}}$ of 0.144 and an $R^2_{\text{val}}$ of 0.271 and an RPD value of 0.9, the 6-NF-developed chlorophyll model would not be recommended for any application (Williams and Norris, 2001). The CV was 96%; this very high CV value was worsened due to the very broad range of chlorophyll values and the low chlorophyll mean of the validation set. Slightly better statistical results were obtained by increasing to 12-NF of the prediction model (Table 2). The $R^2_{\text{val}}$ increased to 0.525 (vs. 0.232) while the SEP decreased slightly (16 mg/kg vs. 21 mg/kg) and the RPD increased to 1.4 (vs. 0.9). However, these results still showed that this handheld instrument was not suitable to the accurately prediction of chlorophyll content of intact canola seeds.

Tkachuk and Kuzina (1982) suggested wavelengths ranging from 630 to 754 nm in the visible region and 1640 to 2116 nm in the NIR region were essential to predict chlorophyll content in intact canola seeds using reflectance instruments. Daun et al. (1994) used four different NIR systems to predict canola quality parameters. Their results obtained with a Tecator instrument with a wavelength range of 850–1050 nm range were poor for the chlorophyll prediction model; the $R^2_{\text{val}}$ was 0.576 and the SEP was 9 mg/kg, similar to results presented in Table 2. In comparison, their results obtained with two different 6500 NIR Systems with a wavelength range of 400–2500 nm, presented very satisfactory statistical results, $R^2_{\text{val}}$ = 0.942 and 0.951 and SEP = 3.9 and 3.2 mg/kg; the wavelengths which used were 662, 686, 1534, and 1753 nm, confirming
that a broader range of wavelengths, especially in the visible part of the spectra (662 & 686 nm), was necessary for the accurate prediction of chlorophyll levels. In this study, the instrument wavelength range was 908–1676 nm which is likely not sufficient to develop good models to accurately predict canola chlorophyll content. Additionally, in Canada canola chlorophyll is nonhomogeneously distributed within harvest samples; individual seeds can contain high levels of chlorophyll (>700 mg/kg) among a multitude of seeds containing zero or near-zero levels of chlorophyll (Daun, 2003; Daun and Symons, 2000). This could affect the repeatability and reproducibility of any determination method.

The statistical results for the NIR analysis of the canola chlorophyll content using the 6-NF chlorophyll calibration model to predict the 2018 harvest canola verification set, confirmed that this model performed poorly, $R^2_{\text{val}}$ was 0.116 and SEP of 7.33 mg/kg. Some of the samples in the 2018 external validation set with chlorophyll contents higher than 80 mg/kg had their predicted values ranging from 6 to 33 mg/kg, which showed that in 2018 the 6-NF model could not even be used to segregate the seeds into low or high chlorophyll samples. The 12-NF chlorophyll model was also tested but performances were worse than the 6-NF calibration model and therefore not reported. This also confirmed that wavelengths in visible range are essential to develop good calibration models for chlorophyll prediction.

### Oleic Acid Content of the Oil

The calibration function of oleic acid is shown in Fig. 2. The region below 1110 nm contributed very little to the NIR visualization calibrations, indicating that the 900–1110 nm region had little influence in the NIR oleic acid content model. In contrast, the instrument’s cutoff at 1676 nm appeared to truncate a significant absorption feature that could have a significant and negative impact on model performance.

A 6-NF model was developed to predict oleic acid, the scatter plot of the predicted value vs. the reference data for the canola harvest in 2017 is presented in Fig. 3e. It shows that an outlier was present in the validation dataset; however, this sample spectrum was not classified as an outlier during the verification testing. The reference oleic acid content of this sample (CN17-51) was 54% whereas the NIR predicted oleic acid content was 72% (Fig. 3e). The complete fatty acid composition of that sample (by reference method) showed that it contained over 10% erucic acid. To be identified as canola, a rapeseed (Brassica napus, B. rapa or B. juncea) cannot contain more than 2% erucic acid, this sample was a canola sample contaminated by rapeseed and should not have been included in the verification set as none of our canola samples contain more than 0.1% erucic acid in the calibration set.

To test the real performance of the NIR oleic content calibration, this outlier rapeseed sample was removed from the 2017 canola verifications set. The performance of the 6-NF oleic content calibration was greatly improved, $R^2_{\text{val}}$ increased to 0.935 from 0.854 (Fig. 3f). This situation highlighted a limitation of NIRS; the predicted samples must be similar to the samples used in the calibration set. If a producer sample of canola is contaminated or commingled with other rapeseeds, the NIR-predicted oleic acid content can contain a significant error, which will not be detected unless a prediction model is developed and implemented to predict contamination by high erucic acid seeds.

The performance results of the 6-NF oleic acid model on the reduced 2017 external validation showed a low SECV of 1.05 and good $R^2_{\text{cross-val}}$ of 0.926, both obtained during cross-validation (Table 2). The SEP, $R^2_{\text{val}}$ and RPD were 1.22%, 0.935 and 3.7, respectively. This indicated a fair model that could be used for screening canola seed (Williams and Norris, 2001). The best results for the 2017 external validation were obtained with a 10-NF calibration model (Table 2); however, the performance results were very similar to the ones obtained with the 6-NF model suggesting no improvement in the model going from 6-NF to 10-NF.

There were two groups of samples based on the oleic acid content (Fig. 3f), a first large group with oleic acid content ranging from 58 to 68% and a second small group of high oleic acid samples (74–76%). The RPD values underestimated the utility of the oleic acid model as the standard deviation for the oleic acid dataset was biased due to the small number of extreme samples (high oleic acid) in the bimodal distribution population of the external validation dataset (Starr et al., 1981). The CV was 1.9%, which suggested that in reality this 6-NF calibration model could be considered a good NIR model to predict oleic acid content (Williams and Norris, 2001).

Our results agreed with and compared favorably to previously reported results. Our SEC (1.05%) was lower than the one obtained by Oblath et al. (2016) who reported an SEC of 3.1% for oleic acid with a sample range of 7.1–65.5%. In 1994, Daun et al. reported an SEP values of 4.7% and 4.8% for an oleic acid range of 55.6–65.6% for a model obtained with 13 factors ($R^2_{\text{val}} = 0.903$ and 0.909, RPD = 2.54 and 2.62). This was further improved, as Siemens and Daun (2005) reported an SEP 0.77% and an $R^2_{\text{val}}$ of 0.919 for a prediction set containing 997 samples with an oleic acid content of 54.1–75.5%. Velasco et al. (1999) reported an SEP value of 3.8%, for an $R^2_{\text{val}}$ of 0.93; however, the population was very different that the one used in this study. Velasco et al. (1999) used various Brassica
varieties where the oleic acid mean was 22.1% and the range was from 5.4–80.8%, with most of the seeds being between 5.4% to just over 40%; the larger range and the sample distribution improved the reported $R^2_{val}$ value.

The stability of the 6-NF oleic acid calibration model was tested against our second external validation set (2018 harvest canola). The results (Table 2) indicated that the 6-NF oleic acid NIR calibration model was stable with excellent performance, $R^2_{val} = 0.937$ and $SEP = 1.14%$; these values are nearly identical to those obtained using the 2017 external validation set of (0.926 and 1.2%, respectively).

**α-Linolenic Acid Content of the Oil**

The performances of the selected 6-NF calibration model for α-linolenic acid content are presented in Table 2 and Fig. 3g. The SECV of 0.976% was promising, the $R^2_{cross-val}$ was 0.867 and the $R^2_{val}$ 0.929. The SEP and RPD values for the 6-NF α-linolenic acid calibration model were 0.773% and 3.8, respectively. These prediction results suggested that this is a fair model that can be used for screening (Williams and Norris, 2001). It was interesting to notice that the $R^2_{val}$ was higher than $R^2_{cross-val}$, likely due to the random assembling of the samples in both sample sets (calibration and external validation). Fig. 3g shows that there were a large set of samples in the 8–12% range and another smaller set of samples in the 2% range. The CV was 9.3%, which could indicate that the prediction model was poor (Williams and Norris, 2001), however, the CV is dependent of the standard deviation of the validation dataset, which in this case was unusually high because of the bimodal nature of the distribution.

The results compared well with previously reported results for benchtop instruments. Oblath et al. (2016) reported a SEC value of 0.59%, for a α-linolenic acid range of 4.7–38.6%. Velasco et al. (1999) reported an SEP value of 3.44% with an α-linolenic acid range of 1.0–69.8%, which is much larger than the range used in this study. In 1994, Daun et al. reported SEP values of 1.2–1.3 with an α-linolenic acid range of 6.9–12.9% (which is narrower than ours) with a 12-NF NIR prediction model. Siemens and Daun (2005) reported an SEP of 0.4% for an α-linolenic acid range of 1.4–14.4% (slightly larger than ours) but using a validation set containing 997 samples.

Daun et al. (1994), Velasco et al. (1999), Siemens and Daun (2005) and Oblath et al. (2016), all had better $R^2_{val}$ values (all were in excess of 0.95) than the one we obtained with our prediction model, but all used benchtop NIR spectrometer with wider wavelength range than the handheld NIR used in this project. In this study, the α-linolenic acid calibration model used only six factors. One explanation for the lower performance of our calibrations is that, of the nine principal wavelengths used by Daun et al. (1994) only four were accessible by the handheld spectrometer. The limitations of the wavelength range could also be seen in Fig. 2 where the visualization calibration function for α-linolenic acid showed that the instrument cutoff above 1676 nm truncated a significant absorption feature in α-linolenic acid, likely degrading the quality of the NIR calibration.

The stability of the 6-NF α-linolenic acid calibration model developed with 2016–2017 canola samples was tested on the second external validation canola sample set (2018 samples) and results are presented in Table 2. This final testing indicated that this NIR model was stable and robust with very good performance, $R^2_{val}$ was 0.883 and SEP was 0.98%. These values compare well with the $R^2_{val}$ and SEP obtained with the 2016–2017 external validation set (SEP = 0.773% and $R^2_{val}$ = 0.929).

**Total Saturates of the Oil**

The total saturates contents of the validation sample set were well within the range of the total saturates contents of the calibration sample set (Table 1), which was a favorable condition for model testing. Unfortunately, model performance was very poor; both in internal cross-validation and in external validation models performed poorly irrespective of the NF in the models (Table 2 and Fig. 3h). Even in model development, the $R^2_{dev}$ was low (i.e., 0.809) despite having 20 NFs. The best model was determined to have 8-NFs with an attendant SECV of 0.22 and an $R^2_{cross-val}$ of 0.399. This gave a model validation with an SEP of 0.26% and an $R^2_{val}$ of 0.290 for the 2017 canola verification set. The NIR models developed for total saturates in canola seed for this handheld NIR spectrometer were not usable.

Siemens and Daun (2005) reported an SEP of 0.23%, the total saturates range was 5.2–8.8%, similar to our range 5.4–8.0%, however they had a very large number of samples in their validation sample set ($N = 997$). In a previous attempt, Daun et al. (1994) reported an SEP of 0.61% (total saturated content ranging from 4.5 to 6.3%) for a model also using nine factors. Of the nine principal wavelengths reported by Daun et al. (1994) to have been used and found to be important in their total saturates prediction model, only three are accessible by the handheld spectrometer due to its limited wavelength range.

As expected, when tested on the second external validation sample set (2018 canola harvest), the 8-NF calibration model performed poorly, $R^2_{val}$ = 0.202 and SEP = 0.25% and was considered unusable. In this study, the handheld NIR with the developed total saturate model cannot be used to accurately predict the total saturates content of intact canola seeds.
Iodine Value of the Oil

The model performances are presented in Table 2 and Fig. 3i shows the scatter plot with the linear regression of the NIR predicted values vs. the reference iodine values. The lowest NF that could be used in a calibration model was 6, whereas a 9-NF calibration model gave the best results for our 2017 external validation set. The SECV and SEP of the 6-NF iodine value calibration model were both low, 0.897 and 1.14 units, respectively, with very good R-squares; R$^2_{\text{cross-val}} = 0.974$ and R$^2_{\text{val}} = 0.968$. The RPD of 6.6 suggested a very good NIR prediction model that could be used for quality and process control (Williams and Norris, 2001).

The model prediction results obtained with the handheld NIR compared favorably with the results reported by Siemens and Daun (2004). Their SEP was 3.0 units and R$^2_{\text{val}}$ 0.983 for and RPD of 7.3 using a data set of 997 samples for a total iodine value range of 31 units (95.5–126.5 units).

Iodine value is a parameter that takes into account to total unsaturation of the oil; it is widely used by the industry as an oil quality parameter. Iodine value by AOCS Cd 1c-85 is a calculated value taking into account the entire fatty acid composition of the oil (AOCS Cd 1C-85:2017), especially the polyunsaturated fatty acids. For canola oil, the fatty acids with the most influence on the iodine value are α-linolenic acid (C18:3) and linolenic acid (C18:2). In a way, it is surprising that iodine value prediction by NIRS generated very good statistical results when individual fatty acids prediction gave less than satisfactory results for the individual fatty acids.

The stability of the 6-NF NIR iodine value model was tested by analyzing the 2018-canola samples and the results presented in Table 2. The NIR model was stable with excellent performance—the R$^2_{\text{val}}$ was 0.987 and SEP was 0.98 units. The model appeared to be stable and robust as the new R$^2_{\text{val}}$ and SEP values were very similar to those values obtained from the 2017 external validation set (R$^2_{\text{val}} = 0.966$ and SEP = 1.14 units).

Conclusion

The handheld NIR spectrometer used in this study has a limited wavelength range (908.1–1676.2 nm). However, it could be used to predict several of important parameters that defined canola seed quality, e.g., oil and protein contents as well as oleic acid content and the iodine value of the oil. The α-linolenic acid model was usable, especially when the interpretation of the prediction results were associated with oleic acid and iodine value results. This instrument could also be used to monitor total glucosinolate content and could help to segregate samples based on their total glucosinolates. Chlorophyll content could not be accurately predicted using the handheld instrument; it could not even be used to segregate samples as shown with the 2018 external validation set as no visible wavelengths were accessible. This might be a limiting factor as chlorophyll content is an important parameter to define good seed quality in Canada.

Acknowledgments The authors thank Pamela Lavallée for her help scanning the samples. The project was fully funded by the Canadian Grain Commission (#1309).

References

AOCS Official Method Ba 4e-93.: Reapproved 2017. Generic combustion method for crude protein.
AOCS Recommended Practice Cd 1c-85: Reapproved 2017. Calculated iodine value.
Azzouz, T., Puigdomenech, A., Aragay, M., & Tauler, R. (2003) Comparison between different data pre-treatment methods in the analysis of forage samples using near-infrared diffuse reflectance spectroscopy and partial least-squares multivariate calibration method. Analytica Chimica Acta, 484:121–134.
Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989) Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy, 43:772–777.
Barthet, V.J. (2018).
Bill, C., Hurburgh, C., Cao, N., & Rippke, G. (2014) Calibration of the JDSU MicroNir 1700 for agricultural product analysis. NIRD News, 25:16–18.
Canola Council of Canada. (2019) Winnipeg, MB, Canada: Canadian Grain Commission. https://grainscanada.gc.ca/en/grain-research/export-quality/oilseeds/canola/ (access January 23, 2020).
Biller, C., Hurburgh, C., Cao, N., & Rippke, G. (2014) Calibration of the JDSU MicroNir 1700 for agricultural product analysis. NIRD News, 25:16–18.
Canola Council of Canada. (2019) Winnipeg, MB, Canada: Historic seed exports. https://www.canolacouncil.org/markets-stats/historic-seed-exports (accessed November 6, 2019).
Daun, J. K. (2003) How green is green. Journal of the American Oil Chemists’ Society, 80:119–122.
Daun, J. K., Clear, K. M., & Williams, P. K. (1994) Comparison of three whole seed near-infrared analyzers for measuring quality components of canola seeds. Journal of the American Oil Chemists’ Society, 71:1063–1068.
Daun, J. K., & Symons, S. (2000) How green is green. Journal of the American Oil Chemists’ Society, 77:1209–1213.
Daun, J. K., & Unger, E. H. (2016) In Canola processing. (Ed.), Encyclopedia of food and grains (Vol. 3, 2nd ed., pp. 497–502). Amsterdam, Tokyo: Elsevier-Academic Press.
Dayal, B. S., & MacGregor, J. F. (1997) Improved PLS algorithms. Journal of Chemometrics, 11:73–85.
Faber, M. N., & Rajko, R. (2007) How to avoid over-fitting in multivariate calibration—The conventional validation approach and an alternative. Analytica Chimica Acta, 595:98–106.
Grassi, S., Casiraghi, E., & Alamprese, C. (2018) A non-targeted approach to assess authenticity of fish fillets and patties. Food Chemistry, 243:382–388.
Heaney, R. K., Spinks, E. A., & Fenwick, G. R. (1988) Improved method for the determination of the total glucosinolate content of rapeseed by determination of enzymatically released glucose. Analyst, 113:1515–1518.
International Organization for Standardization (ISO). ISO 10519.: 2015. Rapeseed—Determination of chlorophyll content—
Spectrometric method. Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardization (ISO). ISO 12966-1: 2014. Animal and vegetable fats and oils—Gas chromatography of fatty acid methyl esters—Part 1: Guidelines on modern gas chromatography of fatty acid methyl esters. Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardization (ISO). ISO 12966-2: 2017. Animal and vegetable fats and oils—Gas chromatography of fatty acid methyl esters—Part 2: Preparation of methyl esters of fatty acids. Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardization (ISO). ISO 659:2009. Reviewed and confirmed in 2014. Oilseeds—Determination of oil content (Reference method). Geneva, Switzerland: International Organization for Standardization.

Kumar, S., Chauhan, J. S., & Kumar, A. (2010) Screening for erucic acid and glucosinolate content in rapeseed-mustard seeds using near infrared reflectance spectroscopy. Journal of Food Science and Technology, 47:690–692.

Lindgren, F., Gelati, P., & Wold, S. (1993) The kernel algorithm for PLS. Journal of Chemometrics, 7:45–59.

Liu, D., Zeng, X. A., & Sun, D. W. (2013) NIR spectroscopy and imaging techniques for evaluation of fish quality—A review. Applied Spectroscopy Reviews, 48:609–628.

Martens, H., & Næs, T. (1991). Multivariate Calibration, JohnWiley & Sons, Chichester, New-York, Brisbane, Toronto, Singapore. Copyright 1989, reprinted with correction in 1991, reprinted 1993, 1994, 1998, 2001 & 2002.

Næs, T., Issakson, T., Fearn, T., & Davies, T. (2002) A user-friendly guide to multivariate calibration and classification (Vol. 2002). Chichester, England: NIR Publications.

Oblath, E. A., Isbell, T. A., Berhow, M. A., Allen, B., Archer, D., Brown, J., ... Longh, D. S. (2016) Development of near-infrared spectroscopy calibrations to measure quality characteristics in intact Brassicaceae germplasm. Industrial Crops and Products, 89:52–58.

Pérez-Marin, D., Paz, P., Guerrero, J. E., Garrido-Varo, A., & Sánchez, M. T. (2010) Miniature handheld NIR sensor for the onsite non-destructive assessment of post-harvest quality and refrigerated storage behavior in plums. Journal of Food Engineering, 99:294–302.

Petisco, C., García-Criado, B., Vázquez-de-Aldana, B. R., de Haro, A., & García-Ciudad, A. (2010) Measurement of quality parameters in intact seeds of brassica species using visible and near-infrared spectroscopy. Industrial Crops and Products, 32:39–46.

Savitzky, A., & Golay, M. J. E. (1964) Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry, 36:1627–1639R.

Siemens, B. J., & Daun, J. K. (2005) Determination of the fatty acid composition of canola, flax, and Solin by near-infrared spectroscopy. Journal of the American Oil Chemists’ Society, 82:153–157.

Starr, C., Morgan, A. G., & Smith, D. B. (1981) An evaluation of near infra-red reflectance analysis in some plant breeding programs. Journal of Agricultural Science, 97:107–118.

Statistic Canada.: Estimated areas, yield, production, average farm price and total farm value of principal field crops, in metric and imperial units, 2019. Ottawa, ON: Statistics Canada. https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=3210035901#timeframe (accessed November 6, 2019)

Tkachuk, R. (1981) Oil and protein analysis of whole rapeseed kernels by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists’ Society, 58:817–822.

Tkachuk, R., & Kuzina, F. D. (1982) Chlorophyll analysis of whole rapeseed kernels by near-infrared reflectance. Canadian Journal of Plant Science, 62:875–884.

Velasco, L., Goffman, F. E., & Becker, H. C. (1999) Development of calibration equations to predict oil content and fatty acid composition in Brassicaceae Germplasm by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists’ Society, 76:25–30.

Williams, P. K., & Norris, K. (2001) Near-infrared technology in the agricultural and food industries (2nd ed.). Minneapolis, MN: American Association of Cereal Chemists, Inc.

Zeaiter, M., Roger, J. M., Bellon-Maurel, V., & Rutledge, D. N. (2004) Robustness of models developed by multivariate calibration. Part I: The assessment of robustness. Trends in Analytical Chemistry, 23:157–170.