NON-DESTRUCTIVE ANALYSIS OF SINGLE PLANT CANOLA *(BRASSICA NAPUS)* SEED USING NEAR INFRARED SPECTROSCOPY

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Harjot Sidhu

In Partial Fulfillment of the Requirements
For the Degree of
MASTER OF SCIENCE

Major Department:
Agricultural and Biosystems Engineering

November 2012
Fargo, North Dakota

North Dakota State University
Graduate School

Title

NON-DESTRUCTIVE ANALYSIS OF SINGLE PLANT CANOLA (*BRASSICA NAPUS*) SEED USING NEAR INFRARED SPECTROSCOPY

By

Harjot Sidhu

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Dennis Wiesenborn
Chair

Dr. Scott Pryor

Dr. Rhonda Magel

Dr. Clifford Hall

Approved:

06-18-2013 — Date

Dr. Sreekala Bajwa — Department Chair

ABSTRACT

Near Infrared Spectroscopy (NIRS) is widely used for quantitative analysis of oilseeds in a non-destructive manner. Canola (*Brassica napus*) is a popular oilseed crop that is used for food and biofuel markets. Due to limited seed availability in plant breeding programs, single plant analysis is often preferred. An NIRS commercial calibration model was evaluated to predict single plant canola seed, but the results showed the need for new NIRS calibration models to predict moisture content, oil content, and fatty acid content for single plant canola seed (3 g) with minimal sample preparation. A separate NIRS calibration model was developed for glucosinolates content utilizing 20 g seed. The resulting NIRS calibration models for moisture and oil content were acceptable. However, suitable NIRS calibration models were not obtained for fatty acids and glucosinolates content due to limited constituent variability and the narrow wavelength range used to collect spectra.

ACKNOWLEDGMENTS

I am heartily thankful to my major advisor, Dr. Dennis P. Wiesenborn, for continuously encouraging me and pushing me to my limits to complete my thesis, and for all the support and guidance he gave me throughout my research work. Working with him gave me the most valuable experience of my life.

I would also like to thank the members of my advisory committee: Dr. Scott Pryor, Dr. Clifford Hall, and Dr. Rhonda Magel for their valuable guidance and suggestions during my research proposal presentation.

I am most grateful to Dr. Darrin Haagenson, research specialist in the Department of Agricultural and Biosystems Engineering, to provide me technical help and guidance from the initial to the final level that enabled me to develop an understanding of the subject. I extend my appreciation to undergraduate students Autmn McDougall, and Laura Chavez for their assistance in experimentation. Nonetheless, special thanks go to Dr. Mukhlesur Rahman to provide a key idea for this research project.

I thank Dr. Bhavnita Sandhu to be my mentor and for her endless support and encouragement. Last, but by no means least, I offer my regards and blessings to all of my family, friends and graduate students in the Department of Agricultural and Biosystems Engineering who provided me moral support in any respect during the completion of this research project.

Sincere thanks to North Dakota Center of Excellence for Oilseeds Development and North Dakota Agricultural Experiment Station for providing funding for this project.

DEDICATION

To my mother, Joginder Sidhu.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGMENTS iv

DEDICATION v

LIST OF TABLES xi

LIST OF FIGURES xii

LIST OF APPENDIX TABLES xiii

GENERAL INTRODUCTION 1

Statement of Objectives 3

Thesis Organization 4

References 5

LITERATURE REVIEW 7

Rapeseed and Canola 7

Canola Production 8

Issues with Canola Quality 9

Recommended Canola Storage 10

Canola as Biodiesel Feedstock…………………………………………………………...11

Canola Seed Quality Analysis Method...12

Standard Wet Chemistry Analysis...12

Spectroscopic Analysis...13

Spectroscopic Methods Involved in Determining Canola Seed Quality....................14

NIR Basic Concepts 14

NIR Instrumentation...16

Sampling and Model Development..19

Effect of Canola Seed sample size on NIR analysis 22

References 23

PAPER 1: DIODE ARRAY NEAR INFRARED SPECTROMETER CALIBRATIONS FOR COMPOSITION ANALYSIS OF SINGLE PLANT CANOLA *(BRASSICA NAPUS)* SEED...32

Abstract...32

Introduction...33

Materials and Methods..36

Materials 36

Collection of Reference Analysis Data 36

Collection of Spectra for Single Plant Canola Seed (3 g) 38

Experimental Design and Statistical Analysis 39

Evaluation of NIRS Commercial Calibration Model to predict 3 g Canola Seed 39

NIRS Multivariate Calibration Method Development for 3 g Canola Samples 40

Results and Discussion 41

Evaluation of Commercial NIRS Calibration Model for Single Plant (3 g) Canola Seed 41

NIRS Multivariate Calibration Method Development for 3 g canola portions 42

Reference Analysis 42

NIRS Calibration Models 43

Moisture Content 45

Oil Content 46

Fatty Acid Content 47

Validation of the Calibration Models with Single Plant Canola Seed 50

Conclusions 53

Acknowledgements 53

References 53

PAPER 2: NON DESTRUCTIVE ANALYSIS OF GLUCOSINOLATES CONTENT OF CANOLA (*BRASSICA NAPUS*) SEED USING A DIODE ARRAY NEAR INFARRED SPECTROSCOPY 58

Abstract 58

Introduction 59

Materials and Methods 61

Materials 61

Methods 62

Reference Chemistry 62

Collection of Spectra for Canola Seed (20g) 62

Experimental Design and Statistical Analysis 63

Results and Discussion 63

Reference Analysis 63

NIRS Model Evaluation 66

Conclusions 69

References 70

GENERAL CONCLUSIONS AND RECOMMENDATIONS...73

General Conclusions 73

Recommendations for Future Work..74

APPENDIX A: STUDY ON PERFORAMCE OF MYLAR BAGS IN STORING CANOLA SEED 76

APPENDIX B: STATISTICAL DATA OF DIFFERENT NIRS CALIBARTION MODELS....77

APPENDIX C: SUMMARY OF LITERATURE REPORTS OF PREDICTION OF OIL CONTENT, FATTY ACIDS, AND TOTAL GLUCOSINOLATES CONTENT BY NIR...82

References 86

LIST OF TABLES

Table Page

1. Fatty acid profile (wt. %) of common biodiesel feedstock oilseeds....................................12

2. Standard methods used to measure Oilseeds constituents by the Canadian Grain Commission..13

3. Absorption frequencies of IR spectra...14

4. Common NIR band of organic compounds..15

5. Guidelines for interpretation of R^2 and RPD statistics...22

6. Predicted compositions of 2500 canola samples obtained from 5 different ND locations in 2011..37

7. Statistics of validation of NIRS commercial calibration model with 3 g canola portions (n = 85)... 42

8. Calibration and validation statistics for the PLS-1 (partial least squares algorithm) calibration equations obtained for the prediction of 3 g canola seed................................49

9. Validation Statistics of developed NIRS models to predict seed from a single plant 51

10. Summary of statistics of reference chemistry data of canola seed total glucosinolates content...64

11. Calibration and validation statistics for the NIRS calibration equations obtained for the prediction of glucosinolates content (µmol/g seed) of 20 g canola seeds..............67

LIST OF FIGURES

Figure | Page

1. *Brassica* species relationships shown by the triangle of U 7

2. Comparative analysis of world production of major oilseeds...9

3. Near IR absorption spectrum of intact rapeseed samples in the whole NIR range 16

4. Basic NIR Instrumentation design 18

5. Steps Involved in building, Calibrating and Validating of the model..................................19

6. Micro Mirror Module (Perten Instruments) with 3 g canola seed 39

7. Distribution plots for moisture content, oil content, palmitic and stearic acid content, and oleic acid content based on (A) predicted values from a commercial calibration model, and (B) reference chemistry values..44

8. Scatter plot of predicted versus measured values of single plant canola seed samples for (A) moisture content, (B) oil content, and (C) oleic acid content..................................52

9. General glucosinolate structure 59

10. Breeder's sample cup developed by Perten Instruments..62

11. Distribution plot of calibration and validation sample sets for glucosinolates content......65

12. Scatterplot of NIRS predicted values vs. reference value for glucosinolates content (µmol/g) of canola seed 66

LIST OF APPENDIX TABLES

Table Page

A1. Moisture content (%) of canola seed stored in Mylar bags over the period of 9 months....76

B1. Calibration and validation statistics for the equations obtained for the prediction of moisture content (%) of 3 g canola seeds....78

B2. Calibration and validation statistics for the equations obtained for the prediction of oil content (dry basis, %) of 3 g canola seeds....79

B3. Calibration and validation statistics for the equations obtained for the prediction of oleic acid content (%) of 3 g canola seeds....80

B4. Calibration and validation statistics for the equations obtained for the prediction of saturated fatty acid (C16:0 + C18:0) content (%) of 3 g canola seeds....81

C1. Summary of literature reports of canola seeds prediction by NIR....83

GENERAL INTRODUCTION

Rapeseed is a traditional oilseed crop which belongs to the family *Brassicaceae*, also known as the mustard family or cabbage family. High contents of erucic acid and glucosinolates content make rapeseed unfit for human and animal consumption. Canadian varieties have been genetically modified to correct these problems and were named as "Canola" in 1979. Therefore, by the most recent definition, canola refers to rapeseed cultivars that contain less than 2% erucic acid in the oil and less than 30 μmol/g of glucosinolates content in the meal. In 1985 the US Food and Drug Administration (FDA) recognized that rapeseed and canola were different species and thus granted GRAS (generally recognized as safe) status to canola (Niewiadomski 1990).

Canola plays an important role in the world due to its use as an edible and industrial oil and animal feed. Because of the high oil content and desirable fatty acid profile, its utilization as a biodiesel feedstock is still growing. The search for higher oil content, low levels of saturated fatty acids, low levels of glucosinolates content and other traits requires extensive plant breeding research and wet chemistry methods. Wet chemistry methods for determining canola composition are time consuming and are destructive in nature. For canola breeding programs, nondestructive and rapid methods are needed for analyzing canola seed composition from a single plant.

Near Infrared spectroscopy (NIRS) technology was introduced over the last decades for wide-scale, nondestructive, inexpensive chemical analysis of food and agricultural commodities. The nondestructive nature of the method is a major advantage of analyzing

composition by NIRS, along with rapid measurement and minimal sample preparation. NIRS measurements obey Lambert and Beer's law, and quantitative measurements can be successfully made with high speed and ease of operation. For quantitative measurements, NIR instruments need a calibration model to predict composition from the sample spectrum. NIRS calibration models have been developed to predict canola composition (Petisco et al. 2010; Hom et al. 2007; Velasco and Becker 1998).

Developing accurate and precise NIRS calibration models is expensive and time consuming; therefore much care is required for the sample preparation prior to spectra collection. In literature, studies have been found where rapeseed samples were oven-dried prior to spectra collection (Hom et al., 2007, Petisco et al., 2010, Mika et al., 2003). The ability of NIRS to predict sample composition with minimal sample preparation is highly desired as it reduces analysis time. Reliable NIRS calibration models should be developed without this additional seed drying step.

Accurate and precise NIRS calibration models are important and require careful selection of reference samples leading up to calibration. The composition of samples to be analyzed by NIRS should be within the composition range of samples used for the development of the calibration model. Therefore, a wide range of composition in reference samples is essential to build an accurate and reliable calibration model.

NIRS calibration models can show variance depending on the type of instrument used for spectra collection. In recent years, the development of an advanced dispersive spectrophotometer Diode Array (DA) has been possible due to the availability of silicon-based sensors in linear arrays. Monono et al. (2012) and Hall (2001) discussed the features

and advantages of using a DA dispersive NIR instrument. A DA NIRS instrument has been used to develop a successful calibration for soybean seeds collected from a single plant (Naeve et al. 2008).

To develop NIRS models for small seed samples, the sample is placed in an adapter inserted in a ring cup. Adapters used in previous studies were made of a variety of materials such as Teflon and PVC (Hom et al., 2007), quartz glass and anodized aluminum (Font et al., 2006), optical grade quartz glass cover (Petisco et al., 2010), and standard polyvinyl sheets with a cover of aluminum foil (Velasco et al., 1999). In contrast to standard sampling accessories, a mirrored cup enhances the reflectance signal from each seed, and restricts stray light from entering the spectrophotometer detector. Naeve and coauthors (2009) developed NIRS calibration models for soybean seeds using a mirrored cup and acknowledged that the mirrored cup provided a confidence interval of about 2 times that of the standard small sampling ring cup. In the literature, no study has been reported to build a calibration model for canola seeds using a mirrored cup and DA NIRS instrument.

Statement of Objectives

The main objectives of this study were:

(1) Evaluate a commercial NIRS calibration model to predict canola composition of single plant canola seed.

(2) Minimize the sample preparation steps for developing NIRS calibration models.

(3) Develop new NIRS calibration models for predicting canola composition from single plant seed to predict moisture content, oil content, and fatty acid content utilizing a mirrored cup.

(4) Develop NIRS calibration models for predicting glucosinolates content using 20 g canola seed utilizing breeder's cup.

Thesis Organization

The thesis consists of a literature review and two research papers. The literature provided the background and issues related to canola seed quality, and NIR concepts.

Paper 1, entitled "Nondestructive analysis of single plant canola (*Brassica napus*) seed using Near Infrared Spectroscopy", evaluated the performance of NIRS calibration model to predict canola composition of single plant canola seed (Objective 1). It also discussed the development and evaluation of NIRS calibration models to predict the single plant canola seed for moisture content, oil content, and fatty acid composition by minimizing the sample preparation steps (Objective 2 and 3). Paper 2, entitled "Nondestructive analysis of total glucosinolates content of canola (*Brassica napus*) seed using Near Infrared Spectroscopy", explored the ability of NIRS to predict glucosinolates content in canola seed (Objective 4).

Subsequent to the papers, general conclusions and recommendations section summarizing the results from Paper 1 and Paper 2 are given. Recommendations for future research are also included in this section. The appendices of the thesis present a performance of Mylar bags in storing canola seed.

References

Font R, del Río-Celestino M, de Haro-Bailón A (2006) The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. Ind Crop Prod 24:307–313

Hom N H, Becker H C, Mollers C (2007) Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. Euphytica 153:27–34

Naeve S L, Orf J, Killam A, Shadow W, Honigs D (2009) Micro Mirror Module analysis of small seed samples for soybean seed quality. http://www.perten.com/Publications/Articles/Micro-Mirror-Module-analysis-of-small-seed-samples-for-soybean-seed-quality/. Accessed September 18 2012

Naeve S L, Proulx R A, Hulke B S, O'Neill T A (2008) Sample size and heterogeneity effects on the analysis of whole soybean seed using near infrared spectroscopy. Agron J 100 (1):231-234

Petisco C, Garcia-Criado B, Vazquez-de-Aldana B R, de Haro A, Garcia-Ciudad A (2010) Measurement of quality parameters in intact seeds of Brassica species using visible and near-infrared spectroscopy. Ind Crop Prod 32 (2):139–146

Velasco L, Becker H C (1998) Estimating the fatty acid composition of the oil in intact-seed rapeseed (*Brassica napus* L.) by near- infrared reflectance spectroscopy. Euphytica 101:221–230

Velasco L, Mollers C, Becker H C (1999) Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Euphytica 106:79–85

Monono E M, Haagenson D M, Pryor S W (2012) Developing and Evaluating NIR Calibration Models for Multi-Species Herbaceous Perennials. Ind Biotechnol 8(5):285-292

Mika V, Tillmann P, Koprna R, Nerusil P, Kucera V (2003) Fast prediction of quality parameters in whole seeds of oilseed rape (*Brassica napus* L.). Plant Soil Environ 49 (4):141–145

Hall J W (2001) Variables Affecting Near-Infrared Spectroscopic Analysis. In: Williams P, Norris K (eds) Near-Infrared Technology in the Agricultural and Food Industries, 2nd edn. AACC, Inc., St. Paul, Minnesota, pp 171–198

Niewiadomski H (1990) Rapeseed Varieties. In: Rapeseed Chemistry and Technology. ELSEVIER, PWN-Polish Scientific Publishers, Poland, pp13–31

LITERATURE REVIEW

Rapeseed and Canola

Rapeseed is a traditional oilseed crop which belongs to the family *Brassicaceae* and is closely related to other Brassica species. The close relationship of rapeseed to other *Brassica* species is very well demonstrated by the "U triangle" (Figure 1) proposed by the Japanese scientist U in 1935. Traditional rapeseed varieties are high in erucic acid, which makes the oil unfit for human consumption (Iqbal et al. 2008). Rapeseed is also high in glucosinolates content, which inhibits growth in livestock and poultry (Griffiths et al. 1998). In order to lower the erucic acid and glucosinolates content, plant breeders artificially synthesized three rapeseed varieties: *B. carinata*, *B. juncea*, and *B. napus,* by hybridization and chromosome doubling of the three *Brassica* species: *B. nigra, B. rapa*, and *B. oleracea* (Raymer 2002).

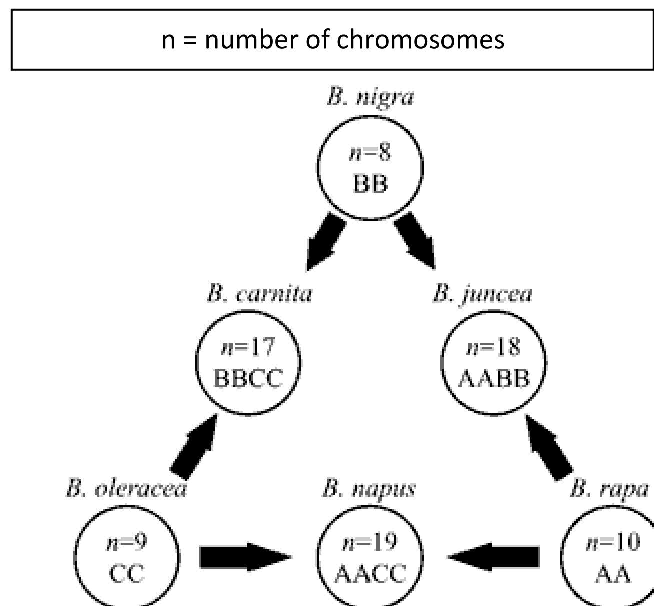


Figure 1. *Brassica* species relationships shown by the triangle of U (Raymer 2002).

Canola was a name given to modified forms of rapeseed that are low in erucic acid and glucosinolates levels, and includes varieties of three botanical species: *Brassica napus*, *B. rapa,* and *B. juncea*. In order to call modified rapeseed “canola”, the oil must not contain more than 2% erucic acid and the solid component of the seed must not have more than 30 μmol/g of glucosinolates content (Booth and Gunstone 2004). Canola oil's original name was "LEAR" oil (Low Erucic Acid Rape). In North America the word “canola” is also referred to as “edible rapeseed” (Raymer 2002).

Canola seed is crushed for oil and animal feed meal. Canola oil is used for cooking purposes and achieved the status of GRAS in January 1985 by the U.S. Food and Drug Administration. Canola is now widely grown for edible and industrial oil. Canola meal serves as an important animal feed, as it is a rich source of proteins.

The fatty acid profile of edible oil determines its end use. Canola oil has lower levels of saturated fatty acids than any other widely used edible oils, which make it a healthy choice for consumers—as saturated fatty acids are linked to cardiovascular diseases and type 2 diabetes. It is high in monounsaturated fatty acids and has moderate levels of polyunsaturated fatty acids (PUFA), which have a number of health benefits such as the reduction of cholesterol levels and protection against heart diseases. Canola oil is in demand by food processing and fast food industries because of its desirable properties.

Canola Production

Over the past few years, global canola production has grown rapidly. Canola is the second largest oilseed crop grown worldwide, second only to soybean (USDA 2011b). A comparison of the world’s major oilseed global production over the last 4 years is shown in Figure 2. Most canola production in the United States takes place in North Dakota, Idaho,

Montana, Oklahoma and Minnesota. North Dakota alone produced 90 percent of the nation's canola crop in 2009 and 2010 (USDA 2011a). In the United States most of the canola produced is *B. napus.*

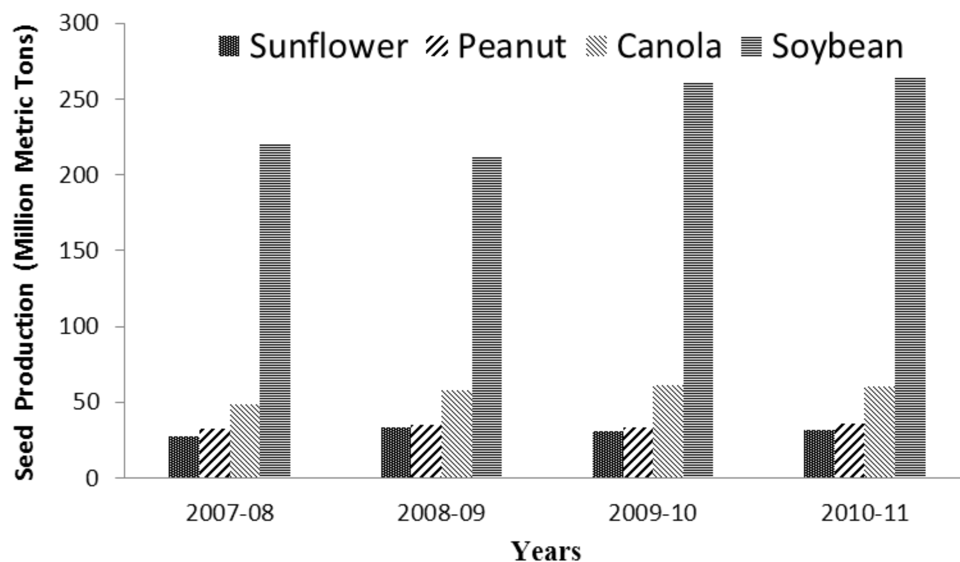


Figure 2. Comparative analysis of world production of major oilseeds (USDA 2011b).

Issues with Canola Quality

The composition of different fatty acids, glucosinolates content, erucic acid, and oil content present in the canola determines its quality and economic value. Although canola is bred to contain less than 30 µmol/g of glucosinolates and less than 2% of erucic acid, the development of new canola breeding lines with lower levels of glucosinolates content and erucic acid is of significant interest to canola breeders. This is due to the toxic and anti-nutritive effects associated with these compounds (Velasco and Becker 1997).

Saturated fatty acids are linked to the risk of coronary heart diseases (Siri-Tarino 2010). *Brassica napus* contains about 7% saturated fatty acids. Although canola oil contains the lowest amount of saturated fatty acids among all vegetable oils, canola breeders are trying to reduce the major saturates further for additional human health benefits (Beaith et al. 2005). Many researchers are trying to reduce the saturated fatty acid content of canola, while some research programs associated with the industrial sector are interested in increasing the saturated fatty acid content for the production of cosmetics, softeners, lubricants, and other oil-based materials (Stoll et al. 2005).

Recommended Canola Storage

Storing canola seeds in zipper bags made of low density polyethylene (LDPE) material is becoming very common (Williams 2008). LPDE material provides Moisture Vapor Transmission Rates (MVTR) in the range between 15 - 22.5 $g/m^2/d$ (Zhang and Zhou, 2009), and is not an adequate barrier against moisture loss (Adom et al. 1996). This can result in changes in the moisture content of stored samples over time. Mylar bags, made of polyester resin and laminated to aluminum foil layer, provide moisture barrier properties with a Moisture Vapor Transmission Rate (MVTR) of less than 0.078 $g/m^2/d$.

Spoilage of canola is closely related to the seed moisture content as well as the moisture and temperature of the storage atmosphere. For safe, long-term storage of canola seed, moisture below 8% with a temperature below 15 °C is recommended (Hammond 2011). At higher temperatures and higher moisture, canola seeds can deteriorate due to one of the following changes: metabolic processes, oxidative reactions, action of insects and fungi, accumulation of carbon monoxide and carbon dioxide (Reuss and Pratt 2001).

Canola as Biodiesel Feedstock

Throughout the world there is an emerging problem of overuse of petroleum-based fuels, due to threats to the supply by global political instability and the serious environmental concerns such as global climate change associated with petroleum usage (Monteiro et al. 2008). This raised the interest in biofuels because of their potential for reduced greenhouse gas emissions and renewability. Biofuels are alternative fuels produced from diverse bio-feedstock. Currently biodiesel is a viable transportation fuel option compared to other biofuels such as bioethanol, biomethanol, and biohydrogen (Yusuf et al. 2011).

Biodiesel is defined as mono-alkyl esters of long chain fatty acids derived from vegetable oil or animal fat (Silveira et al. 2011). Feedstock choice impacts biodiesel quality and its production cost. Canola has 40% oil content, which is highest among potential biodiesel feedstocks (soybean and palm both contain 20% oil content). Oils with high saturation have poor cold flow characteristics and can become solid at low temperatures due to high freezing points. High levels of polyunsaturation lead to oxidation and polymerization. Based on this criterion, the ideal feedstock for biodiesel production exhibits low saturation and low polyunsaturation (Walker 2004). It is clear from Table 1 that canola oil meets these requirements.

Currently the European Union produces 60% of the global biodiesel production and 50% of this global biodiesel production is canola biodiesel production (Aukema and Campbell 2011). In the United States, canola accounts for 20% of biodiesel production.

Overall, it is estimated that, out of the total global biodiesel production, 30% is canola biodiesel.

Table 1. Fatty acid profile (wt. %) of common biodiesel feedstock oilseeds (Moser 2009).

Fatty Acid	Canola	Palm	Soybean	Corn
Palmitic (C16:0)	4	45	11	11
Stearic (C18:0)	2	4	4	2
Oleic (C18:1)	61	39	23	28
Linoleic (C18:2)	22	11	54	58
Linolenic (C18:3)	10	-	8	1

Canola Seed Quality Analysis Methods

Standard Wet Chemistry Analysis

For analysis of seed constituents using reference lab methods, it is of utmost importance that results are precise and repeatable. Reference methods developed by internationally recognized standard writing agencies such as the American Oil Chemists' Society (AOCS) and International Organization for Standardization (ISO) have been accepted worldwide for oilseed testing. For canola quality testing, the Canadian Grain Commission, Winnipeg, MB, Canada, provides reference methods based on or calibrated against the methods developed by AOCS or ISO. Reference methods to analyze oil content, fatty acid composition, and glucosinolates content, used by the Canadian Grain Commission, are presented in Table 2.

While performing the laboratory tests or using standard reference methods, rigorous controls must be maintained to ensure the accuracy and precision of the results. Also, many reference analyses are labor-intensive, destructive, time-consuming and expensive. These

factors contribute to the necessity of identifying reliable alternative methods that can be carried out nondestructively with time and cost savings (Font et al. 2005).

Table 2. Standard methods used to measure oilseeds constituents by the Canadian Grain Commission.

Constituents	Standard Method	Method ID
Oil Content	Nuclear magnetic resonance (NMR)	ISO 10565:1992(E)
Fatty Acid Composition	Gas Chromatography	ISO 5508:1990(E)
Glucosinolate content	Spectrometric method for total glucosinolates by glucose release	ISO 9167–3:2007(E)

Spectroscopic Analysis

Spectroscopy is the interaction of light with matter. It has captured the interest of many scientists to predict oilseed composition in recent years. Among different kinds of spectroscopic tools available (UV-visible, Infrared (IR), Raman, NMR), NIR spectroscopy has gained particular interest of many scientists to predict oilseed composition. Many authors have used the near-infrared (NIR) spectrum for predicting content of oilseed constituents. NIRS has been used to predict chemical composition of oilseeds such as sunflower (Fassio and Cozzolino 2004), olive (Morales-Sillero et al. 2011), soybean (Lee and Choung 2011), and canola (Petisco et al. 2010; Velasco et al. 1999; Velasco et al. 2002). Blanco et al. (2002) discussed the fundamental aspects of NIR spectroscopy and its advantages over other analytical tools. NIR spectroscopy has also been used in the

determination of seed quality for plant breeding because it is non-destructive and analyses of several traits can be conducted simultaneously (Font et al. 2006).

Spectroscopic Methods Involved in Determining Canola Seed Quality

NIR Basic Concepts

Infrared spectroscopy (IR spectroscopy) refers to the infrared region of the electromagnetic spectrum, which is light with a longer wavelength and lower frequency than visible light. IR spectroscopy is used to identify the type of bonds present and structural information. Different chemical bonds absorb radiation at different frequencies and examination of the absorption spectrum may provide clues on substance composition and concentration. Infrared radiation is a region of the electromagnetic spectrum with a wave number ranging from 13,000 to 10 cm^{-1}, or wavelength ranging from 0.78 to 1000 µm (Hsu 1997). Furthermore, the IR region is sub divided into three areas: Near IR, Mid IR, and Far IR. Information on wavelength and wave number of IR sub areas is presented in Table 3.

Table 3. Absorption frequencies of IR spectra (Hsu 1997).

	Near IR	**Mid IR**	**Far IR**
Wave number	13,000-4,000 cm^{-1}	4,000-200 cm^{-1}	200-10cm^{-1}
Wavelength	0.78-2.5 µm	2.5-50 µm	50-1,000 µm

Different substances, based on their electron configuration, absorb different wavelengths of light. Absorption in the NIR region results from molecular vibrations (mainly stretching and bending) within a compound. These absorptions are quantum mechanical in nature and can only absorb discrete packets of energy called photons. The absorption of one photon of energy by the molecule in its ground state is called a fundamental absorption. The absorption of two such photons is called first overtone—and so on for higher overtones (Hsu 1997). Table 4 shows NIR-absorption bands for common organic bonds.

Table 4. Common NIR band of organic compounds (Stuart, 2004).

Wavelength (nm)	Assignments
2,200-2,450	Combination of C-H stretching
2,000-2,200	Combination of N-H and O-H stretching
1,650-1,800	First overtone C-H stretching
1,400-1,500	First overtone N-H and O-H stretching
1,300-1,420	Combination C-H stretching
1,100-1,225	Second overtone C-H stretching
950-1,100	Second overtone N-H and O-H stretching
850-950	Third overtone C-H stretching
775-850	Third overtone N-H and O-H stretching

NIR spectra consists of overtones and combination bands of the fundamental absorptions in the mid infra-red region. The number of photons absorbed is dependent on the type of the chemical bonds present in the sample. There is no direct way to measure the number of photons absorbed as they disappear one-by-one. However, absorbance can be calculated from transmittance (ratio of the radiation transmitted by the sample to the radiant power incident on the sample) using equation 1.

$$Absorbance = log\left(\frac{1}{Transmittance}\right) = \alpha\lambda lc \qquad (1)$$

where: $\alpha\lambda$ = the molar absorption coefficient

l = path length of light source

c = sample concentration

NIR spectra are traditionally displayed as wavenumber or wavelength on x-axis, and absorption intensity or percent transmittance on y axis (Figure 3). Wavenumber and wavelength are inversely related to each other through equation 2.

$$\text{Wave number } (cm^{-1}) = \frac{1}{\text{Wave length } (\mu m)} \text{ x } 10^4 \qquad (2)$$

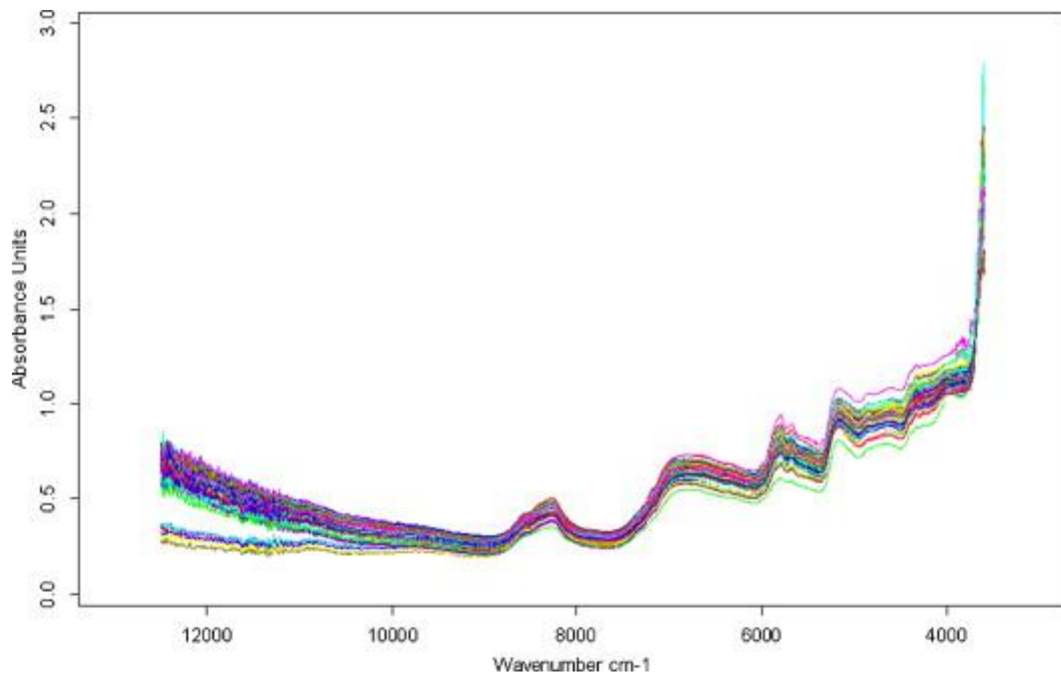


Figure 3. Near IR absorption spectra of intact rapeseed samples in the whole NIR range (Bala and Singh, 2013).

NIR Instrumentation

NIR instruments consist mainly of a light source, an optical splitter, and a detector. Detailed information of the different types of NIR instruments commercially available is

given by Workman and Burns (2008). Most of the NIR instruments available commercially use either dispersive or Fourier transform (FT) spectrometric techniques, which differ depending on the optical splitter used. The FT-IR spectrometer uses the optical splitter “interferometer” and the dispersive spectrometers use a “monochromator” as an optical splitter. Comparison of dispersive and FT NIR instruments was done by Armstrong et al. (2006) for measuring grain and flour attributes. They reported that the FT-NIR and dispersive NIR instruments were essentially equal in measurement accuracy and concluded that there are no apparent advantages of one over the other. In literature, dispersive NIR instrumentation has been used widely to develop NIR calibration models for rapeseed (Hom et al., 2007, Velasco and Becker 1998, Velasco et al., 1999, Petisco 2010).

Diode array spectrometers have been recently introduced and are classified as an advanced dispersive spectrophotometer. This technology does not use any moving optical parts, so there is less chance of mechanical misalignment and wear of critical moving parts over time compared to conventional dispersive and FT NIR instruments (Jerome and Jones 1995). Diode array NIR spectrometers have hundreds of photo diode detectors integrated on a single silicon chip, with each photo diode measuring a different portion of the spectrum. This technology enables all wavelengths to be measured simultaneously on the diode array, increasing the speed of the spectra collection (McClure 2001). The capability of measuring all of the wavelengths at the same time gives diode array spectrometers an edge over conventional dispersive and Fourier transform spectrometers, where there is a time difference between each spectral measurement, thus increasing the analysis time required per sample. Therefore, more scans can be performed with diode array spectrometers to ensure greater accuracy in a short time.

The NIR spectrum can be created with either a reflectance or transmittance measurement (Figure 4) and is sample dependent. Reflectance measurements are used for fine-ground samples as the radiation can only penetrate from 1 to 4 mm into the particulate sample surface; whereas, samples used for transmittance measurements are not ground, because the radiation penetrates entirely into the sample. Ideal NIR instruments should have both reflectance and transmittance capabilities.

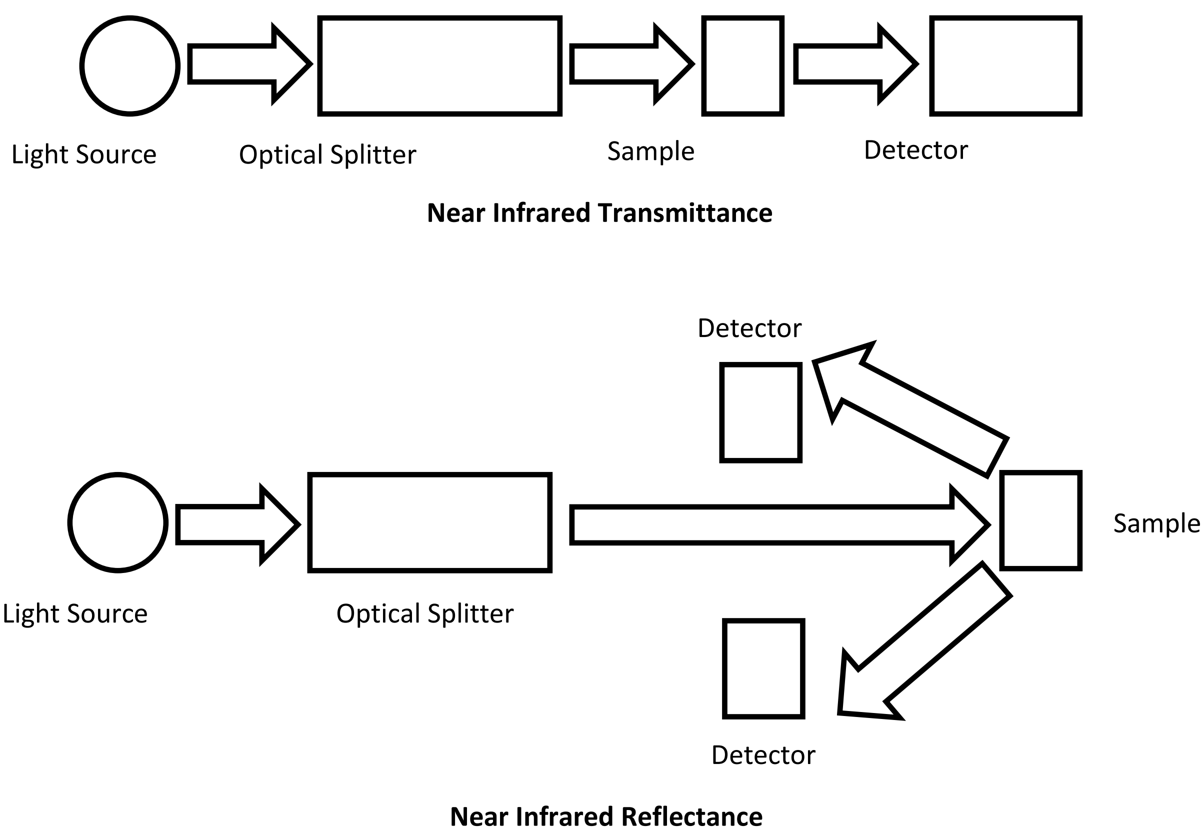


Figure 4. Basic NIR Instrumentation design (Workman and Burns 2008).

Sampling and Model Development

An NIRS instrument uses a mathematical model to predict sample composition from the spectrum. The process of building a mathematical model is called calibration and is built using the spectral data and the wet chemistry reference data. The goal of the calibration is to predict the composition of an unknown sample using the spectral absorbance collected on the NIR instrument, calibrated with the developed mathematical model (Thomas and Ge 2000). A mathematical model is developed for the analysis of each constituent. Basic steps in developing mathematical models are shown in Figure 5.

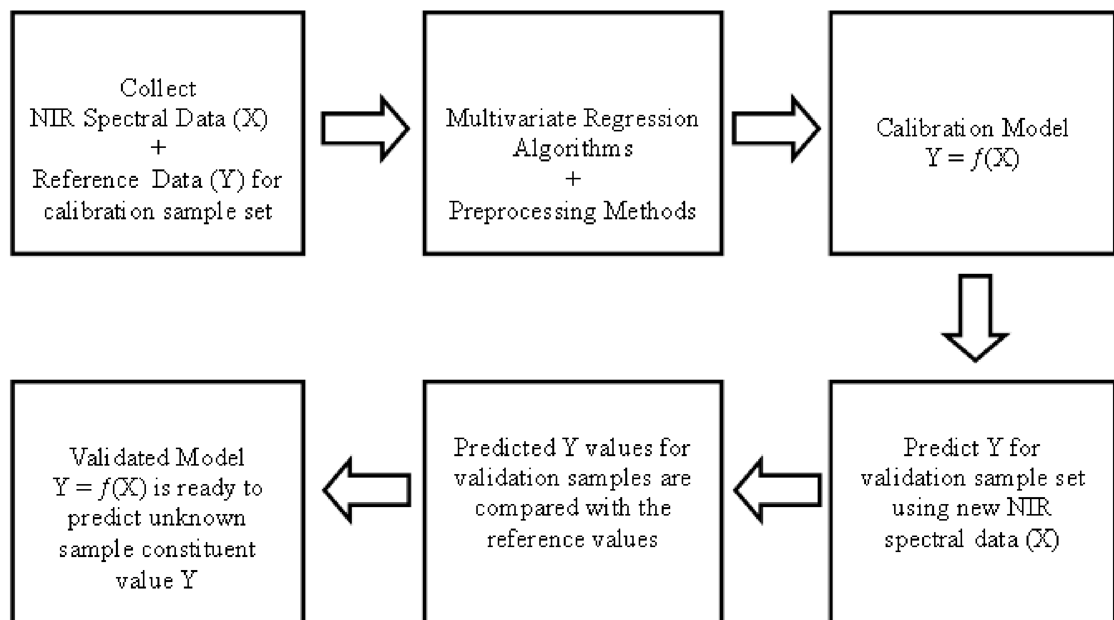


Figure 5. Steps involved in building, calibrating and validating the model.

The first and key step to building a calibration is the selection of samples. Reliability of the model greatly depends on the samples used for building the calibration. Including samples with wide variability in constituent values enables more robust

prediction capabilities (Peirs et al. 2003). The more variability in the samples, the better the model will be (Peirs et al. 2003). Sample sets having a bell-shaped distribution of composition about the mean may bias predictions towards the mean composition of the sample set. Sample sets having uniform distribution of the composition across the anticipated range is ideal for NIR calibration (Williams 2001).

After selecting samples, the next step is to classify calibration and validation sample sets. Validation samples are used to validate the developed calibration model. A minimum of 100 samples is recommended for building the calibration and a 3:1 ratio for calibration and validation samples sets is recommended (Williams and Norris 2001; Williams 2001).

Sample selection is followed by sample preparation which includes sub-sampling, removal of foreign materials, grinding, and storage. Preparation of samples is extremely important, as this alone accounts for 60 to 70% of the overall testing error (Williams 2001).

After sample preparation, NIR spectra are collected by the instrument, followed by the collection of reference data using standard methods. The accuracy and the reproducibility of the standard methods is crucial. A calibration developed from inaccurate reference data is useless. One of the biggest factors impacting a calibration is the accuracy of the moisture content analysis because it fluctuates with time, if the samples are not stored properly. An error of 2% in determination of the moisture content can lead to errors of over 1% in erroneous measurement of oil, protein, and fat content (Williams 2008).

After collecting the reference and spectral data, a mathematical model is developed using statistical processes incorporating chemometrics software. Chemometrics is the method of developing calibrations by relating spectral absorbance to chemical properties. A mathematical relationship to relate the spectral absorbance to the desired quantity must be established, and is done through regression analysis. For building NIR calibration models, regression analysis — including multilinear regression (MLR) to full spectrum methods, such as partial least squares (PLS) or principal component regression (PCR) — can be used (Brimmer and Hall 2001). PLS modeling uses the entire spectrum and is capable of solving problems associated with the sample matrix, and the physical variations in the sample that affects the spectra (Brimmer and Hall 2001). Out of the different types of regression analysis methods, PLS and PCR have been most cited for building canola NIR calibrations (Thomas and Ge 2000).

The developed calibration model is verified for accuracy and reproducibility by performing different statistical tests on the predictions from a validation sample set. The statistical tests that can be performed include: coefficient of determination (R^2), coefficient of correlation (r), F-test, standard error of prediction (SEP), root mean squared error of prediction (RMSEP), standard error of cross-validation, root mean squared error of cross validation (RMSECV), prediction sum of squares (PRESS), bias, and ratio performance deviation (RPD = ratio of the SEP to the standard deviation). A calibration with a high RPD statistic along with a high coefficient of correlation (R^2) is greatly desired (Williams 2001).

Table 5. Guidelines for interpretation of validation R^2 and RPD statistics (Williams 2001).

Values	Interpretation
Coefficient of determination (R^2)	
0.00 – 0.25	Very Poor - Not useable
0.26 – 0.49	Poor - Poor correlation
0.50 – 0.64	Fair - Ok for rough screening
0.65 – 0.81	Fair - Ok for screening
0.82 – 0.90	Good - Useable with caution for most applications
0.91 – 0.98	Very Good - Useable for most applications
0.98 – 1.00	Excellent- Useable in any application
Ratio performance deviation (RPD)	
0.0 – 2.3	Very Poor – Not useable
2.4 – 3.0	Poor – Ok for very rough screening
3.1 – 4.9	Fair – Ok for screening process
5.0 – 6.4	Good – Useable for quality control
6.5 – 8.0	Very Good – Useable for process control
8.1 +	Excellent – Useable in any application

Effect of Canola Seed Sample Size on NIR Analysis

Sample size is extremely important in building a calibration. The collection of samples from many plants results in more constituent heterogeneity as compared to sample from only a few plants (Naeve et al. 2008). Due to the importance of canola breeding, analysis from a single plant or a single seed is of significant interest. Single canola seed NIR calibrations have been developed for seed constituents such as: weight, oil content, protein content, fatty acid composition, and total glucosinolates content (Velasco et al. 1999; Velasco and Möllers 2002; Hom et al. 2007). Sample holder design plays an important role in building NIR calibration. Sample holders from Perten Instruments, Inc. can be used for a wide range of seed sample sizes. For example, a breeder's cup can hold 20 g of seed, and the micro mirror module cup (mirrored cup) can hold a few seeds to a few grams. The mirrored cup is unique in a way that enhances the signal from every single seed for accurate results at a small scale. Naeve et al. (2009) developed calibrations for

analyzing protein, oil, fiber, ash, and fatty acids for small sized soybean samples (8–14 seeds) using a mirrored cup. Those authors found that the calibration developed using this cup predicted constituents in small soybean samples (8–14 seed) with nearly the same level of precision as reference methods. However, the same authors acknowledged that the mirrored cup provided a confidence interval of about 2 times that of the standard small sampling ring cup. A summary of literature reports of prediction of oil content, saturated fatty acids, and total glucosinolate content of rapeseed by NIR is presented in Appendix C (Table C1).

References

Ackman R G, Hooper S N, Hooper D L (1974) Linolenic acid artifacts from the deodorization of oils. J Am Oil Chem Soc 51 (3):42-49

Adom K K, Dzogbefia V P, Ellis W O, Simpson B K (1996) Solar drying of okra - Effects of selected package materials on storage stability. Food Res Int 29 (7):589-593

Armstrong P R, Maghirang E B, Xie F, Dowell F E (2006) Comparison of Dispersive and Fourier-Transform NIR Instruments for Measuring Grain and Flour Attributes.. Appl Eng Agric 22(3): 453-457

Aukema H, Campbell L (2011) Oil nutrition and Utilization. In Daun J K, Eskin N A M, Hickling D (eds) Canola Chemistry, Production, Processing, and Utilization , AOCS Press, pp 245-280

Bala M, Singh M (2013) Non destructive estimation of total phenol and crude fiber content in intact seeds of rapeseed-mustard using FTNI. Ind Crop Product 42: 357-362

Beaith M E, Fletcher R S, Kott L S (2005) Reduction of saturated fats by mutagenesis and heat selection in Brassica napus L. Euphytica 144 (1-2):1-9

Blanco M, Villarroya I (2002) NIR spectroscopy: a rapid-response analytical tool. Trac-Trend Anal Chem 21 (4):240-250.

Brimmer P J, Hall J W (2001) Method Development and Implementation of Near-Infrared Spectroscopy in Industrial Manufacturing Support Laboratories. In Williams P, Norris K (eds) Near-Infarred Technology in the Agricultural and Food Industries, St. Paul, Minnesota, USA: American Association of Cereal Chemists, Inc. pp 199-214

Booth E J, Gunstone F D (2004) Rapeseed and rapeseed oil: agronomy, production and trade. In Gunstone F D (eds) Rapeseed and Canola Oil Production, Processing, Properties and Uses, CRC Press: Blackwell Publishing Ltd. pp 1-16

Canola, National Statistics. 2011.National Ag Statistics Service (NASS)

Crop Production 2010 Summary. 2011. NASS (USDA)

Espinoza-Pérez J D, Haagenson D M, Pryor S W, Ulven C A, Wiesenborn D P (2009) Production and characterization of epoxidized canola oil. Trans ASABE 52:1–9

Fassio A, Cozzolino D (2004) Non-destructive prediction of chemical composition in sunflower seeds by near infrared spectroscopy. Ind Crop Prod 20 (3):321-329

Font R, del Rio M, Fernandez-Martinez J M, de Haro-Bailon A (2004) Use of near-infrared spectroscopy for screening the individual and total glucosinolate contents in Indian mustard seed (*Brassica juncea* L. Czern. & Coss.). J Agric Food Chem 52 (11):3563-3569

Font R., del Rio-Celestino M, Cartea E, de Haro-Bailon A. 2005. The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. Ind Crop Prod 24 (3):307-313

Font R, del Rio-Celestino M, de Haro-Bailon A (2006) Quantification of glucosinolates in leaves of leaf rape (*Brassica napus* ssp. *pabularia*) by near-infrared spectroscopy. Phytochem 66 (2):175-185

Golebiowski T (2004) Near infrared reflectance spectroscopy of oil in intact canola seed (*Brassica napus*, L.). I. Attributes of the intact seed spectrum. J Near Infrared Spectrosc 12 (5):325-330

Golebiowski T, Leong A S, Panozzo J F (2005) Near infrared reflectance spectroscopy of oil in intact canola seed (*Brassica napus* L.). II. Association between principal components and oil content. J Near Infrared Spectrosc 13 (5):255-264

Lennoxd G, Beckman C (2011) Industry Trade and Economics. In Daun J K, Eskin N A M, Hickling D (eds) Canola Chemistry, Production, Processing, and Utilization, AOCS Press pp 313-341

Hammond D (2011) Agronomy. In Daun J K, Eskin N A M, Hickling D (eds) Canola Chemistry, Production, Processing, and Utilization, AOCS Press pp 93-118

Haagenson, D M, Brudvick R L, Lin H, Wiesenborn D P (2011) Implementing an In Situ Alkaline Transesterification Method for Canola Biodiesel Quality Screening. J Am Oil Chem Soc 87(11): 1351-1358

Henon G, Kemeny Z, Recseg K, Zwobada F, Kovari K (1999) Deodorization of vegetable oils. Part I: Modeling the geometrical isomerization of polyunsaturated fatty acids. J Am Oil Chem Soc 76 (1):73-81

Hom, N. H., H. C. Becker, and C. Mollers. 2007. Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. Euphytica 153 (1-2):27-34

Hsu, Sherman C P (1997) Infrared Spectroscopy. In Settle F A (eds) Handbook of Instrumental Techniques for Analytical Chemistry. Englewood Cliffs, NJ: Prentice-Hall pp 247-283

Workman Jr J (1995) Diode array process instrumentation for the near IR. In Anal Spec Lib 6:369-384

Kemeny Z, Recseg K, Henon G, Kovari K, Zwobada F (2001) Deodorization of vegetable oils: Prediction of trans polyunsaturated fatty acid content. J Am Oil Chem Soc 78 (9):973-979

Lee J H, Choung M G (2011) Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. Food Chem 126 (1):368-373

Major Oilseeds: World Supply and Distribution (Commodity View). 2011. FAS (USDA).

McCarthy W J, Kemeny G J (2008) Fourier Transform spesctrophotometers in the near-infrared. In Burns D A, Ciurczak E W (eds) Handbook of Near-Infrared Analysis, Boca Raton, Fl: CRC Press pp 79-92

McClure W F (2001) Near-Infrared Instrumentation. In Williams P, Norris K (eds) Near-Infrared Technology in the Agricultural and Food Industries, St. Paul, Minnesota: American Association of Cereal Chemists pp 109-128

Mika V, Tillmann P, Koprna R, Nerusil P, Kucera V (2003) Fast prediction of quality parameters in whole seeds of oilseed rape (*Brassica napus* L.). Plant Soil Environ 49 (4):141–145

Monteiro M R, Ambrozin A R P, Liao L M, Ferreira A G (2008) Critical review on analytical methods for biodiesel characterization. Talanta 77 (2):593-605

Morales-Sillero A, Fernandez-Cabanas V M, Casanova L, Jimenez M R, Suarez M P, Rallo P (2011) Feasibility of NIR spectroscopy for non-destructive characterization of table olive traits. J Food Eng 107 (1):99-106

Moser B R (2009) Biodiesel production, properties, and feedstocks. In Vitro Cell Dev Biol -Plant 45 (3):229-266

Moss J (2006) Labeling of trans fatty acid content in food, regulations and limits - The FDA view. Atherosclerosis Supp 7(2):57-59

Naeve S L, Proulx R A, Hulke B S, O'Neill T A (2008) Sample size and heterogeneity effects on the analysis of whole soybean seed using near infrared spectroscopy. Agron J 100(1):231-234

Naeve S L, Orf J, Killam A, Shadow W, Honigs D (2009) Micro Mirror Module analysis of small seed samples for soybean seed quality. http://www.perten.com/Publications/Articles/Micro-Mirror-Module-analysis-of-small-seed-samples-for-soybean-seed-quality/. Accessed September 18 2012

Oil Crops Yearbook. 2011. ERS(USDA)

Peirs A, Tirry J, Verlinden B, Darius P, Nicolai B M (2003) Effect of biological variability on the robustness of NIR models for soluble solids content of apples. Postharvest Biol Tech 28(2):269-280

Petisco C, Garcia-Criado B, Vazquez-de-Aldana BR, de Haro A, Garcia-Ciudad A (2010) Measurement of quality parameters in intact seeds of Brassica species using visible and near-infrared spectroscopy. Ind Crop Prod 32 (2):139–146

Raymer P L (2002) Canola: An Emerging Oilseed Crop. In Janick J, Whipkey A (eds) Trends in new crops and new uses. Alexandria, VA: ASHS Pres.

Sato T (2008) Nondestructive Measurements of Lipid Content and Fatty Acid Composition in Rapeseeds (*Brassica napus* L.) by Near Infrared Spectroscopy. Plant Prod Sci 11:146–150

Sato T, Uezono I, Morishita T, Tetsuka T (1998) Nondestructive Estimation of Fatty Acid Composition in Seeds of *Brassica napus* L by Near-Infrared Spectroscopy. J Am Oil Chem Soc 75:1877–1881

Sebedio J L, Catte M, Boudier M A, Prevost J, Grandgirard A (1996) Formation of fatty acid geometrical isomers and of cyclic fatty acid monomers during the finish frying of frozen prefried potatoes. Food Res Int 29(2):109-116

Silveira E L C, de Caland L B, Tubino M (2011) Molecular absorption spectrophotometric method for the determination of phosphorus in biodiesel. Fuel 19(11): 3485-3488

Siemens B J, Daun J K (2005) Determination of the Fatty Acid Composition of Canola, Flax, and Solin by Near-Infrared Spectroscopy. J Am Oil Chem Soc 82(3):153–157

Stoll C, Luhs W, Zarhloul M K, Friedt W (2005) Genetic modification of saturated fatty acids in oilseed rape (*Brassica napus*). Eur J Lipid Sci Tech 107 (4):244-248

Stuart B H (2004) Infrared spectroscopy: Fundametals and Applications. Analytical techniques in science. Johh wileys & Sons Ltd, England.

Thomas E V, Nanxiang Ge. 2000. Development of Robust Multivariate Calibration Models. Technometrics 42 (2):168-177

Thomas P (2003) Canola Varieties. In Canola Growers Manual: Canola Council of Canada

Velasco L, Becker H C (1997) Analysis of total glucosinolate content and individual glucosinolates in Brassica spp. by near-infrared reflectance spectroscopy. Plant Breeding 117:97-102

Velasco L, Becker HC (1998) Estimating the fatty acid composition of the oil in intact-seed rapeseed (*Brassica napus* L.) by near- infrared reflectance spectroscopy. Euphytica 101:221–230

Velasco L, Möllers C, Becker HC (1999) Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Euphytica 106:79–85

Velasco L, Möllers C (2002) Nondestructive assessment of protein content in single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. *Euphytica* 123 (1):89-93

Walker K (2004) Non-Food Uses. In Gunstone F D (eds) Rapeseed And Canola Oil Production, Processing, Properties and Uses, CRC Press, pp 154-185

Williams PC (2001) Implementation of Near-Infrared Technology. In: Williams P, Norris K (eds) Near-Infrared Technology in the Agricultural and Food Industries, 2nd edn. AACC, Inc., St. Paul, Minnesota, pp 145–170

Williams PC, Norris K (2001) Variables Affecting Near-Infrared Spectroscopic Analysis. In: Williams P, Norris K (eds) Near-Infrared Technology in the Agricultural and Food Industries, 2nd edn. AACC, Inc., St. Paul, Minnesota, pp 171–198

Williams, P (2008) Sampling, Sample Preparation, and Sample Selection. In: Burns DA, Ciurczak EW (eds) Handbook of Near-Infrared Analysis, 3rd edn. CRC Press, Boca Raton, Florida, pp 267–296

Yusuf N, Kamarudin S K, Yaakub Z (2011) Overview on the current trends in biodiesel production. Energ Convers Manage 52 (7):2741-2751

Zhang G G Z, Zhou D (2009) Crystalline and Amorphous Solids. In Liu L, Porler W R (eds), Developing Solid Oral Dosage Forms: Pharmaceutical Theory & Practice, Elsevier Inc. London pp: 25-60

PAPER 1: DIODE ARRAY NEAR INFRARED SPECTROMETER CALIBRATIONS FOR COMPOSITION ANALYSIS OF SINGLE PLANT CANOLA (BRASSICA NAPUS) SEED

Abstract

A canola breeder needs an accurate, rapid, non-destructive method for analyzing seeds from a single plant to select the most promising samples for further breeding trials. The introduction of diode array (DA) spectrometers has improved the speed, sensitivity and stability of Near Infrared Spectroscopy (NIRS) instrumentation, and can be helpful to develop multivariate prediction models for canola seed. Analyzing single plant sample sizes with a mirrored cup allows for precise measurement of every single seed. This research was aimed at assessing the potential of DA-NIRS (950-1650 nm wavelength range) in the prediction of single plant canola seed constituents. Eighteen different NIRS Calibration models were developed using 100 samples for each constituent with different pre-processing techniques (mean center, derivatives, variates) and models (PLS, PCR). The relative performance of different calibration models for each constituent was compared using R^2 and RPD values obtained from the validation set of 30 samples. NIRS models developed using the PLS regression algorithm for moisture content ($R^2 = 0.97$, RPD = 6.13) and oil content ($R^2 = 0.84$, RPD = 4.16) were successful. However, acceptable NIRS models were not obtained for fatty acid and glucosinolates content likely due to limited

variability and low levels of the constituent and a narrow wavelength range of the DA-NIR instrument.

Paper 1 is an extensively revised version of a paper presented at the 2012 *ASABE* Annual International *Meeting* in Dallas, TX, July 29 - August 1 2012. Authors: Harjot Sidhu, Darrin Haagenson, Dennis Wiesenborn. Title: Nondestructive analysis of single plant canola (*Brassica napus*) seed using Near Infrared Spectroscopy. Paper number: 12-1337308. Harjot Sidhu, the author of this thesis, is the first author of Paper 1. She designed and conducted the experiments in this paper. Co-authors assisted in the editing of Paper 1.

Introduction

Crop breeders continually strive to develop more productive varieties that require fewer inputs or which are resistant to drought, insects and disease. Increased content of valuable components, such as lipids and proteins, and decreased levels of undesirable components, are also frequently sought. Breeding programs benefit from enhanced tools that permit rapid, non-destructive screening of seed composition from a single plant. Some of the promising tools which were recently introduced for this purpose are diode-array near infrared spectroscopy (DA-NIRS) and a mirrored cup for analysis of seed samples from a single plant.

Canola has undergone significant improvements through breeding, but efforts remain to meet the ever increasing demand for vegetable oil. Canola is a form of rapeseed modified by breeders to have little or no erucic acid (< 2%) and low glucosinolates (<30 μmol g^{-1}). Canola typically contains 40–45% oil (dry basis), and the content of oleic acid and saturated fatty acids is 60% and 6–7 % of total fatty acids, respectively. Composition

varies widely according to the growing conditions and variety (Ratnayake and Daun, 2004). High oil and oleic acid content, in addition to low saturated fatty acid content, make canola seed excellent for edible oil as well as a suitable feedstock for biodiesel production. Canola breeders are striving to improve these attributes, and rapid, non-destructive analysis of seed samples from single plants in a breeding line development program is a valuable aid. In breeding for high oil content and desired fatty acids, the selection of individual plants with the preferred traits are the most important breeding objectives. Each line (single plant) should be characterized for its respective traits.

Near Infrared Spectroscopy (NIRS) is routinely used for the rapid, nondestructive screening of large sample numbers. NIRS has been used for the analysis of oil, protein, fatty acid, and total glucosinolates content of canola/rapeseed (Velasco and Becker 1998; Sato et al. 1998; Velasco et al., 1999; Míka et al., 2003; Hom et al., 2007; Sato 2008., Niewitetzki et al. 2010). Published NIRS research has been based on conventional dispersive or Fourier transform NIR spectrometer. Recently, diode array (DA) spectrometers have been introduced and offer improved features in terms of speed, sensitivity and stability (McClure, 2001). The DA-NIRS incorporates a diode array detector, as well as fiber optics that improve the energy throughput of the instrument. The diode array effectively contains hundreds of detectors that acquire a complete spectrum simultaneously in a fraction of second. The fiber optics collect most of the reflectance spectra from the sample directly to a fixed grating in a monochromator, hence reducing the mechanical wear and tear (Workman Jr, 1995). Therefore, DA-NIRS ensures greater accuracy and a short analysis time compared to conventional spectrometers. Although with these benefits, the scientific literature documenting the use of DA-NIRS for the

compositional analysis of agricultural crops is limited (Naeve et al. 2008; Niewitetzki et al. 2010; Welle et al. 2005; Welle et al. 2007).

For analyzing small seed samples, the sample is placed in an adapter inserted in a ring cup. Adapter made of different materials were used in previous studies such as Teflon and PVC (Hom et al., 2007), quartz glass and anodized aluminum (Font et al., 2006), optical grade quartz glass cover (Petisco et al., 2010), and standard polyvinyl sheets and cover of aluminum foil (Velasco et al., 1999). The above mentioned adapters were used to scan small samples ranging from single seed to 4 g seed. Mirrored cup was used to develop NIRS calibrations for soybean seeds by Naeve et al. (2009). Mirrored cup is unique in a way that it enhances the reflectance signal from each seed, and restricts stray light from entering the spectrophotometer detector. Those authors found that the calibration developed using this cup predicted constituents in small soybean samples (8–14 seeds) with nearly the same level of precision as reference methods utilizing large sample size. However, the same authors acknowledged that the mirrored cup provided a confidence interval of about 2 times that of the standard small sampling ring cup. It would be helpful to canola breeders if use of the mirrored cup could be extended to samples of canola seed from a single plant.

This paper describes the development and evaluation of new NIRS calibration models for predicting moisture, oil, and fatty acid content from single canola plant seed samples. Analysis was completed utilizing a DA-NIRS with a mirrored cup.

Materials and Methods

Materials

Twenty five hundred canola (*Brassica napus*) samples grown in 2011 breeding variety trials were collected from five North Dakota locations: Carrington (n = 606), Drake (n = 742), Hettinger (n = 149), Langdon (n = 745), and Minot (n =258), to provide seed samples with diverse constituent ranges. Upon receiving these harvested samples, seed were cleaned using a Carter Day Dockage Tester to remove foreign material according to the methods of USDA-GIPSA (2004). Cleaned seed were then packed into reclosable polyethylene bags.

The constituent values were predicted using a breeder's cup (22 mL small sample dish) developed by Perten Instruments for use with bulk seed with a DA7200 NIR spectrometer (Perten Instruments, Springfield, IL). Perten's commercial calibration model released in 2008 was used for these predictions. The predicted seed moisture content, oil content (dry basis), and fatty acid composition are summarized by growing location in Table 1. This study sample selection criteria is discussed in the Experimental Design section. All selected samples were stored in Mylar® (PAKDRY1500) bags to prevent moisture loss.

Collection of Reference Analysis Data

Seed moisture content (% dry basis) was determined by drying 4 g samples in duplicate at 103°C for 5 h using a gravity convection oven (Precision Scientific Inc;

Winchester, IL, USA) according to the ISO 665 method for oilseeds. The fatty acid composition of canola samples was quantified by gas chromatography (GC) according to the method of Espinoza-Peréz et al (2009). The results are reported as % of total fatty acids.

Table 6. Predicted compositions of 2500 canola samples obtained from 5 different ND locations in 2011.

Location	Min.[a]	Max.[b]	Mean[c]	SD[d]
	Moisture Content %			
Carrington	3.9	11.9	8.2	1.4
Drake	3.4	5.2	4.2	0.3
Hettinger	2.4	4.5	3.7	0.5
Langdon	2.4	14.5	6.3	2.8
Minot	3.3	4.9	4.2	0.3
	Oil Content %			
Carrington	43.0	50.5	46.0	1.3
Drake	43.0	50.3	47.3	1.0
Hettinger	41.2	49.6	44.7	1.7
Langdon	43.6	56.7	50.1	1.9
Minot	42.8	49.9	46.6	1.3
	Palmitic and Stearic Acid Content (C16:0 + C18:0) %			
Carrington	3.7	6.8	6.0	0.4
Drake	5.2	6.6	5.9	0.2
Hettinger	5.5	7.0	6.34	0.3
Langdon	1.4	7.3	6.2	0.7
Minot	5.0	6.7	5.8	0.3
	Oleic Acid Content (C18:1) %			
Carrington	55.4	69.7	61.5	3.4
Drake	54.3	68.8	62.6	2.7
Hettinger	54.3	64.8	59.0	2.9
Langdon	55.1	74.9	66.4	4.6
Minot	54.5	69.5	62.2	2.8

[a] minimum. [b] maximum. [c] standard deviation.

Oil content was determined using an accelerated solvent extraction unit (ASE 200, Dionex Corp, Sunnyvale, CA) according to the methods of Haagenson et al. (2010). Canola samples were oven dried at 103°C for 5 h or until no change in mass was observed. Dried

canola samples were ground in a coffee grinder with 2.5 g diatomaceous earth (DE) and samples were then loaded into 11 mL stainless steel cells, and void volume was filled with DE. ASE oil extractions were carried out with n-hexane at 100°C, 6.7MPa with a five min equilibration time. Three static cycles (10 min each) having a 100 % flush volume and 60 s purge time were used for the oil extraction. Desolventization of extracted oil was performed using a stream of dry air (-70°C dew point). Extracted samples were reground with DE for a second extraction, and the total oil recovery from both extractions was recorded. Oil content is reported as % dry basis.

To check for the sampling error, SEL (standard error of laboratory) was calculated. An SEL value of less than 1.3 is acceptable in reference chemistry methods for building calibration models according to NIRS technology guidelines (Shenk et al., 2008). Furthermore, it is recommended for NIRS calibration models that SEP (standard error of prediction) values should be less than twice the value of SEL.

Collection of Spectra for Single Plant Canola Seed (3 g)

Spectra were collected at room temperature (24 ± 1 °C) on a DA 7200 NIR spectrometer using the Micro Mirror Module™ from Perten Instruments (Figure 6) with a 950 to 1,650 nm wavelength range and a scan resolution of 5 nm. Each 3 g canola sample was scanned twice, repacked and again scanned twice. The average of four scans was used in chemometric analysis.

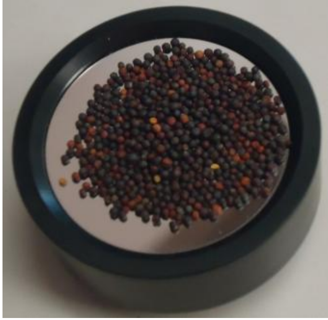


Figure 6. Micro Mirror Module (Perten Instruments) with 3 g canola seed.

Experimental Design and Statistical Analysis

Evaluation of NIRS Commercial Calibration Model to predict 3 g Canola Seed

NIRS predictions for 3 g canola seed samples were evaluated against reference analysis data. Predictions were collected using a commercial NIRS model developed for use with 20 g bulk canola samples (Perten calibration model released in 2008). For this experiment, 85 samples were randomly selected from the pool of 2500 samples (Table 1). The constituents that were analyzed were moisture, oil, and fatty acids (C16:0 + C18:0, C18:1). Spectra were collected in duplicate, and the final spectral value used at each wavelength was the average of the two scans. The commercial NIRS model was evaluated by comparing the predicted values with reference values using the coefficient of determination (R^2) and ratio performance deviation (RPD) values. RPD is calculated by taking the ratio of standard deviation of reference samples to the standard error of prediction.

NIRS Multivariate Calibration Method Development for 3 g Canola Samples

NIRS calibration models were built to predict moisture content, oil content, and fatty acids (C16:0 + C18:0, C18:1). To minimize the tendency of predictions to regress towards the mean, 100 samples for calibration development were selected to have uniform distribution across the range of each constituent (Williams, 2001). Figure 7A shows the distribution plots for selected calibration samples from the pool of 2500 samples. An additional 30 samples were selected from the pool of 2500 samples for validation of the developed NIRS models for each constituent. Validation samples covered the range of constituent variability used in the calibration development for respective constituents (Figure 7A). Utilizing the preexisting calibration to select calibration and validation sample sets saved much time and expense compared to testing all 2500 samples using reference chemistry methods.

Spectral and reference data were exported to the GRAMS Suite v9.0 statistical software package (Thermo Fisher Scientific; Waltham, MA, USA). Eighteen different calibration models were built for each constituent: moisture content, oil content, saturated (palmitic + stearic) fatty acid content, and oleic acid content; these models included combinations of four pre-processing techniques (mean center, Savitzky-Golay 1st derivative, Savitzky-Golay 2nd derivative, standard normal variate) and three regression models (PLS-1, PLS-2, PCR).

Developed NIRS models were evaluated against the above-mentioned 30 validation samples to choose the best calibration model for each constituent. To determine the accuracy of the developed NIRS models, the coefficient of determination (R^2), standard

error of prediction (SEP), and ratio performance deviation (RPD) were computed. Out of all the developed calibrations models, the best calibration model for each constituent was selected on the basis of the higher R^2 and RPD values and lower SEP values of the validation set. The best calibration model of each constituent was further validated with 30 single plant canola samples to check the accuracy of the NIRS models to predict actual canola seed composition from a single seed. These 30 single plant canola seed samples ranged from 2.1–4.3 g.

Results and Discussion

Evaluation of Commercial NIRS Calibration Model for Single Plant (3 g) Canola Seed

A Perten NIRS model developed to predict composition of 20 g portions of canola seed was evaluated for use with 3 g portions of canola seed. Reference chemistry data for moisture content, oil content, and fatty acids (palmitic + stearic, oleic) content was collected on the same 85 samples analyzed with NIRS. The predicted values were compared with reference values and evaluated statistically for $R^{2,}$ and RPD values (Table 7). An R^2 value of at least 0.64 and RPD value of at least 3 are recommended for screening purposes (Williams, 2001; Williams and Sobering, 1993). The low R^2 (0.11 to 0.42) and RPD values (0.4 to 1.3) for all these constituents indicate that this commercial calibration model is not adequate for 3 g canola seed samples. This supports the need for model development appropriate to the sample size and cup.

Table 7. Statistics of validation of NIRS commercial calibration model with 3 g canola portions (n = 85).

Constituents	**Range**	**Mean**	**SD**[a]	R^{2b}	**RPD**[c]
Moisture Content %	3.5-15.5	6.8	2.3	0.42	1.3
Oil Content %	27.6-59.7	42.0	5.1	0.44	1.3
Saturated Fatty Acids[d] %	6.4-10.3	7.6	0.6	0.22	1.1
Oleic Acid %	60.4-69.4	63.8	1.6	0.11	0.4

[a] standard deviation. [b] coefficient of determination. [c] ratio performance deviation. [d] C16:0 + C18.0.

NIRS Multivariate Calibration Method Development for 3 g Canola Portions

Reference Analysis

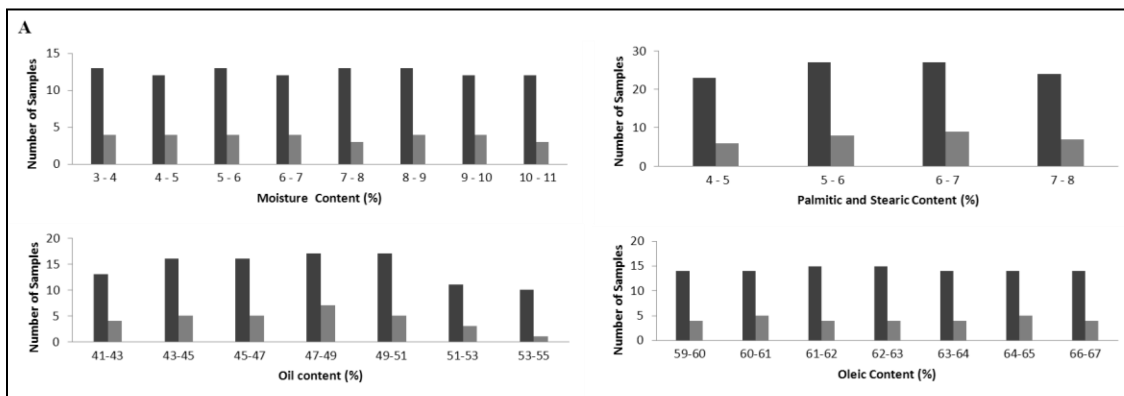
For each constituent, 100 canola samples for calibration development, and an additional 30 validation samples were selected. Figure 7B shows the distribution plots for calibration and validation sample sets based on reference chemistry values. The constituent values varied between 3–11% moisture, 41–55 % oil, 4–8 % saturated fatty acids, and 59–67 % oleic acid content. The reference values generally matched the values predicted by NIRS at the 20 g scale except for moisture content which differed by 1–3 % for the 30 validation samples.

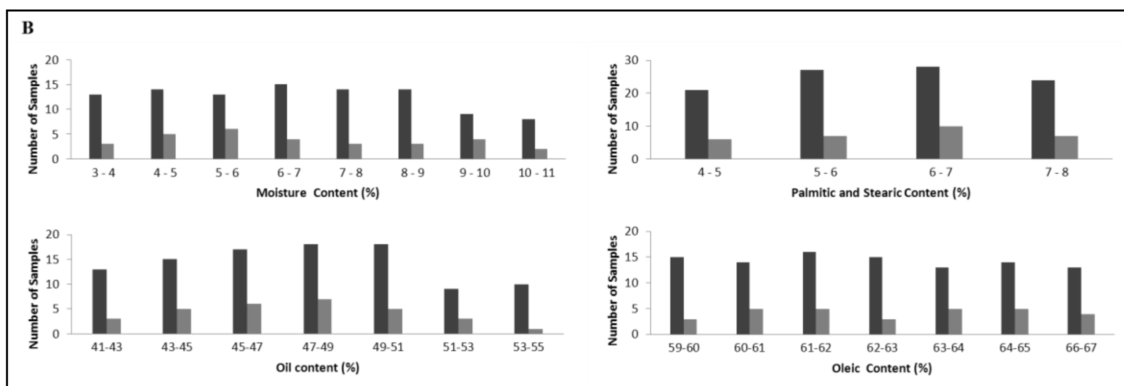
The range of moisture content for this study is narrower than the 4–14 % range reported by Mika et al. (2003) for *B. napus*. Seed samples with moisture greater than 11 % were not included because of their proneness to spoilage due to high moisture content. The oil content range obtained (41–55 % dry basis) is similar to that used in other *B. napus* NIRS models: 36.5–48.4 % dry basis (Petisco et. al. 2010), 28.5–54.9 % as is basis (Velasco et al. 1999), and 26.2–61.1 % dry basis (Hom et al. 2007). The oleic acid range (59–67 %) is narrower than what other authors have reported. Siemens and Daun (2005)

reported an oleic acid content range of 55.1–76.5%, and Sato (2008) reported an oleic acid content range of 15.4–65.9% for building NIRS calibrations. For total saturated fatty acids (palmitic + stearic content), the obtained range is similar to the range reported by other authors to develop NIRS models for *B. napus*: 4.3–7.5 % (Siemens and Daun, 2005), 4.12–8.36 % (Sato 2008), 3.3–8.9 % (Velasco and Becker, 1998).

NIRS Calibration Models

Eighteen different NIRS models for each constituent were developed using three different regression algorithms and four different preprocessing techniques (Appendix B). The relative performance of different calibration models for each constituent were compared using R^2 and RPD values obtained from the validation sample set. As noted previously, an R^2 value of at least 0.64 and RPD value of at least 3 is recommended for NIRS models.





Calibartion Set   Validation Set

Figure 7. Distribution plots for moisture content, oil content, palmitic and stearic acid content, and oleic acid content based on (A) predicted values from a commercial calibration model, and (B) reference chemistry values.

Moisture Content

Fluctuating moisture makes it difficult to develop robust calibration models, because errors in determination of moisture content can cause further problems in predicting other constituents (Baker et al. 1994, Williams 2008). Therefore, a separate moisture NIRS calibration model was developed using PLS and PCR regression algorithms. Using the PLS-2 regression model with different preprocessing methods gave validation R^2 values between 0.80–0.92; SEP values between 0.43–0.47, and RPD values between 4.1–4.6. The poorest validation statistics were obtained in the case of NIRS models built with the PCR regression algorithm where validation R^2, SEP and RPD values were less than 0.87, 0.69, and 3.1, respectively. NIRS models obtained from the PLS-1 regression algorithm were better than PLS-2 and PCR regression models. Table 3 summarizes the calibration and validation statistics for all combinations of PLS-1 regression algorithm with different preprocessing methods. The best validation results were obtained with the PLS-1 algorithm and three processing methods: mean centering, standard normal variate, and Savitzky-Golay 2nd derivative (Table 3). For this NIRS model, the SEP value of 0.32 is consistent with the Standard Error of Laboratory (SEL) of the reference method used for the determination of moisture content (0.25). An SEL value for moisture reference data set is calculated by taking the average of the SD of duplicates of each sample. The validation results (R^2 = 0.97 and RPD = 6.13) indicates that this NIRS calibration model can be used for quality control purposes (Williams, 2001).

Oil Content

The PLS-1 regression algorithm provided better validation results compared to the PLS-2 and PCR regression algorithms. Validation and calibration statistics for the NIRS models developed with the PLS-1 regression algorithm and different pre-processing techniques are summarized in Table 8. Validation R^2 values varied between 0.75–0.84, 0.75–0.82, and 0.72–0.81 for NIRS calibration models developed with PLS-1, PLS-2, and PCR regression algorithms, respectively. RPD values varied from 3.5–4.2, 2.9–3.2, and 2.8–3.1 for PLS-1, PLS-2, and PCR NIRS models, respectively. The best statistical results were obtained by the PLS-1 regression algorithm combined with mean centered preprocessing (Table 8). An SEP value of this NIRS model (0.61) is in accordance with the SEL value of 0.58 calculated for the total reference data set for calculating oil content as the average of the SD of the duplicates of each sample. The validation of this calibration model gave an R^2 value of 0.84 and RPD value of 4.16, indicating that this calibration model is recommended for screening purposes according to the guidelines by Williams (2001).

An identical R^2 value (0.85) and poor SEP value (1.87) was obtained for an NIRS calibration model developed for a single seed of *Brassica* species (Velasco et al., 1999). That model was based on a wider range of oil content (28.5–54.9), and wider wavelength range (400–2500 nm) compared to the oil content and wavelength range used in this study. Poor validation statistics in terms of R^2 (0.71) and SEP (0.80) were also reported for another NIRS calibration model developed for a single seed of *Brassica* species (Mika et al., 2003).

Fatty Acid Content

Calibration models were built for oleic acid and major saturated fatty acids – the sum of palmitic and stearic acid. For oleic acid content, best results were obtained by the PLS-1 regression algorithm and a combination of mean centering and Savitzky-Golay 2nd preprocessing techniques, where validation R^2, SEP, and RPD were 0.86, 0.46, and 2.6, respectively (Table 8).

In contrast to the oleic acid results obtained in our study (calibration R^2 value = 0.81), improved calibration R^2 (0.99) was reported by Velasco and Becker (1998). These authors incorporated high oleic acid mutants to obtain a wider range of oleic acid content (9.6–81.5 %) compared to 59–67 % used in our work. The same authors reported that oleic acid is best analyzed at wavelengths: 2316, 1752, 1800, and 1734 nm; the narrower wavelength range (950–1650 nm) of the NIR instrument used in our study may have resulted in poorer calibration models. However, our validation R^2 value (0.86) is in accordance with the validation R^2 (0.86) reported by Velasco et al. 1999. However, the reported SEP of 8.94 was much higher than our SEP value (0.46). The standard error of the estimate is a measure of the accuracy of predictions. An R^2 value of 0.86 and RPD value of 2.6 indicates our NIRS model is suitable for rough screening purposes (Williams, 2001).

No combination of algorithm and preprocessing methods provided an acceptable calibration model for saturated fat content (sum of palmitic and stearic acid). This may not be surprising as low calibration R^2 values for palmitic acid (0.38) and stearic acid (0.54) content were reported by Sato (2008). Velasco and Becker (1998) reported 0.68 R^2 for both palmitic and stearic acid content. These authors used a wavelength range of 400–2500 nm

to collect the spectra. There may be multiple explanations for a poor palmitic and stearic acid NIRS model. Functional groups present in canola fatty acids absorb NIR wavelengths up to 2470 nm (Westad et al. 2008), and our study included a wavelength range of 950–1650 nm. Spectra were collected from a 3 g portion of canola seed, whereas GC reference analysis was done on only on a few seeds of that 3 g portion;   however, the composition of the reference sample may not be exactly the same as that of the 3 g sample. The variation for palmitic and stearic acids was considerably low (4–8 %). Siemens and Daun (2005) reported that NIRS calibrations based on absolute fatty acid content would give better results than NIRS calibrations based on relative fatty acid content. This is because of Beer-Lambert's law which states that the absorbance varies linearly with the analyte concentration. This may have also contributed to our poor calibration models for oleic, palmitic, and stearic acid.

Table 8. Calibration and validation statistics for the PLS-1 (partial least squares algorithm) calibration equations obtained for the prediction of 3 g canola seed.

Preprocessing	**Calibration**		**Validation**		
	R^{2a}	**SEC[b]**	**R^{2a}**	**SEP[c]**	**RPD[d]**
	Moisture Content (%)				
MC[e]	0.93	0.38	0.92	0.35	5.8
MC + SNV[f]	0.93	0.38	0.93	0.38	5.1
MC + SG1st[g]	0.94	0.39	0.93	0.40	4.8
MC + SNV+ SG1st	0.94	0.40	0.93	0.45	4.3
MC + SG2nd[h]	0.95	0.37	0.94	0.34	5.8
MC + SNV+ SG2nd*	0.96	0.40	0.97	0.32	6.1
	Oil Content (%)				
MC*	0.82	0.81	0.84	0.61	4.2
MC + SNV	0.79	0.87	0.75	0.67	3.8
MC + SG1st	0.80	0.85	0.81	0.71	3.6
MC + SNV+ SG1st	0.80	0.78	0.78	0.73	3.5
MC + SG2nd	0.80	0.80	0.76	0.71	3.6
MC + SNV+ SG2nd	0.81	0.87	0.80	0.62	4.1
	Oleic acid Content (%)				
MC	0.78	0.72	0.75	0.79	1.5
MC + SNV	0.79	0.72	0.78	0.56	2.1
MC + SG1st	0.79	0.75	0.78	0.60	1.9
MC + SNV+ SG1st	0.80	0.75	0.76	0.67	1.8
MC + SG2nd*	0.81	0.69	0.86	0.46	2.6
MC + SNV+ SG2nd	0.81	0.70	0.85	0.49	2.4

[a] coefficient of determination. [b] standard error of calibration. [c] standard error of prediction. [d] ratio performance deviation. [e] mean center. [f] standard normal variate. [g] Savitzky-Golay 1st derivative. [h] Savitzky-Golay 2nd derivative. * Selected NIRS calibration models for moisture, oil, and oleic acid content.

Validation of the Calibration Models with Single Plant Canola Seed

NIRS Calibration models developed in Experiment 2 were from 3 g portions of seed blended from multiple plants. The robustness of selected calibration models to predict canola seed constituents from a single plant was further tested. Thirty single plant canola seed samples were analyzed for moisture, oil, and oleic acid content using the selected calibration models (Table 8) for each constituent. The predicted values were then compared with the reference values (Table 9). The scatterplots of reference values vs. NIRS values for moisture, oil, and oleic acid content are shown in Figure 8.

The range of moisture content, and oleic acid content of these single plant samples was similar to the constituent variability contained in the calibration models, except for oil content. For oil content, the variability (30.1–50.9 %) of single plant canola seed samples was wider than the variability (41–55 %) used to develop the NIRS model. This may account for the reduced validation RPD and R^2 values measured for the single plant oil content predictions versus those from 3 g mixed-plant samples.

For moisture content, the validation statistics were in close agreement with the validation statistics obtained from the 3 g portion of canola seed. In contrast, poor validation statistics were obtained for predicting oleic acid content of single plant seed. Again, this may be explained by the narrow wavelength range used to develop the NIRS calibration model for this constituent.

Variation in sample sizes can affect the NIRS predictions (Williams, 2008). The calibration models were developed specifically for 3 g canola seed, whereas the single

plant seed samples ranged between 2.1–4.3 g. This might be one factor for obtaining different validation statistics for the two sets of samples.

Table 9. Validation Statistics of developed NIRS models to predict seed from a single plant.

Constituents	Min.[a]	Max.[b]	Mean	SD[c]	R^{2}[d]	SEP[e]	RPD[f]
Moisture content %	6.2	11.0	8.5	1.6	0.96	0.32	4.9
Oil content %	30.1	50.9	41.7	5.5	0.91	1.60	3.4
Oleic acid content %	59.6	69.7	65.3	3.0	0.60	2.00	1.5

[a] minimum. [b] maximum. [c] standard deviation. [d] coefficient of determination. [e] standard error of prediction. [f] ratio performance deviation.

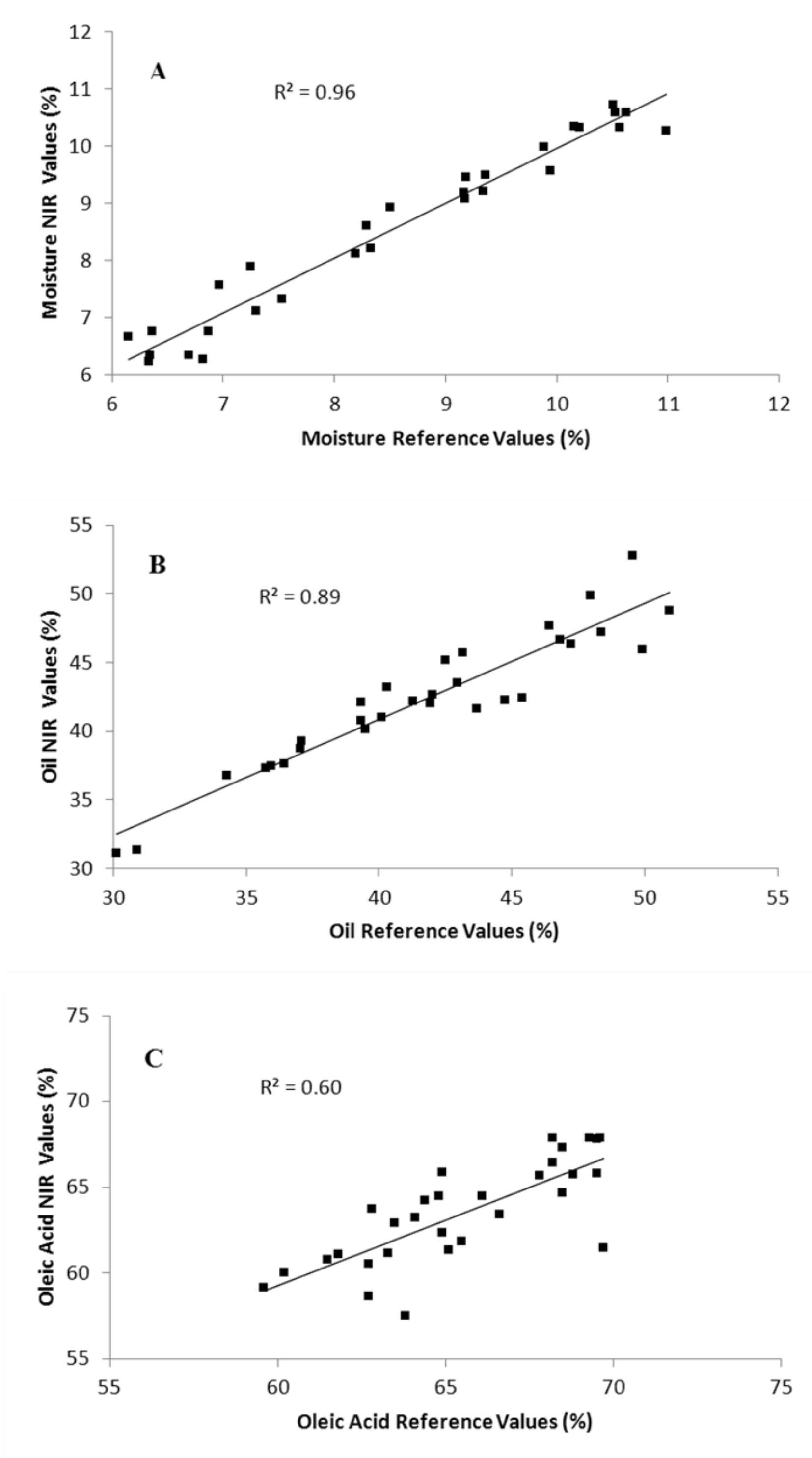


Figure 8. Scatter plot of predicted versus measured values of single plant canola seed samples for (A) moisture content, (B) oil content, and (C) oleic acid content.

Conclusions

In conclusion, the commercial NIRS calibration models were not adequate for small (3 g) canola seed samples. The mirrored cup permitted the development of suitable NIRS calibration models to predict moisture and oil content of single plant canola seeds (3 g). However, satisfactory results for fatty acids were not obtained due to the limited variability of fatty acid constituents and the relatively narrow wavelength range of the spectrophotometer.

Acknowledgements

The authors thank North Dakota Center of Excellence for Oilseeds Development and North Dakota Agricultural Experiment Station for providing funding for this project. The authors also thank Laura Chavez and Leonard Cook for technical assistance in oil reference analysis.

References

Baker, C.W., D.I. Givens, and E.R. Deaville. 1994. Prediction of organic matter digestibility in vivo of grass silage by near infrared reflectance spectroscopy: effect of calibartion method, residual moisture and particle size. *Anim. Feed Sci. Technol.* 50(1-2):17–26.

Espinoza-Pérez, J.D., D.M. Haagenson, S.W. Pryor, C.W. Ulven, and D.P. Wiesenborn. 2009. Production and characterization of epoxidized canola oil. *Trans. ASABE* 52:1–9.

Font, R., M. Del Río, J.M. Fernández-Martínez, and A. De Haro-Bailón. 2004. Use of near-infrared spectroscopy for screening the individual and total glucosinolate contents in Indian mustard seed (*Brassica juncea* L. Czern. & Coss.). *J. Agric. Food Chem.* 52(11):3563–3569.

Font, R., M. del Río-Celestino, and A. de Haro-Bailón. 2006. The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. *Ind. Crop Prod.* 24(3):307–313.

Haagenson, D.M., R.L.Brudvick, H. Lin, and D.P. Wiesenborn. 2010. Implementing an In Situ Alkaline Transesterification Method for Canola Biodiesel Quality Screening**.** *J. Am. Oil Chem. Soc.* 87(11):135–1358.

Hom, N.H., H.C. Becker, and C. Mollers. 2007. Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. *Euphytica* 153(1-2):27–34.

McClure, W. F. 2001. Near-Infrared Instrumentation. In: *Near-Infrared Technology in the Agricultural and Food Industries*, 109-128. P. Williams, and K. Norris, eds. AACC, Inc., St. Paul, Minnesota.

Mika, V., P. Tillmann, R. Koprna, P. Nerusil, and V. Kucera. 2003. Fast prediction of quality parameters in whole seeds of oilseed rape (*Brassica napus* L.). *Plant Soil. Environ.* 49 (4):141–145.

Naeve, S. L., R. A. Proulx, B. S. Hulke, and T. A. O'Neill. 2008. Sample size and heterogeneity effects on the analysis of whole soybean seed using near infrared spectroscopy. *Agron. J.* 100 (1):231-234.

Naeve, S.L., J. Orf, A. Killam, W. Shadow, and D. Honigs. 2009. Micro Mirror Module analysis of small seed samples for soybean seed quality. http://www.perten.com/Publications/Articles/Micro-Mirror-Module-analysis-of-small-seed-samples-for-soybean-seed-quality/. Accessed 18 September 2012.

Niewitetzki, O., P. Tillman, H.C. Becker, and C. Mollers. 2010. A New Near-Infrared Reflectance Spectroscopy for High-Throughput of Oleic Acid and Linolenic Acid Content of Single Seeds in Oilseed Rape (*Brassica napus* L.). *J. Agric. Food Chem.* 58(1):94-100.

Petisco, C., B. Garcia-Criado, B.R. Vazquez-de-Aldana, A. de Haro, and A. Garcia-Ciudad. 2010. Measurement of quality parameters in intact seeds of Brassica species using visible and near-infrared spectroscopy. *Ind. Crop Prod.* 32 (2):139–146.

Ratnayake, W.M.N., and J.K. Daun. 2004. Chapter 3: Chemical composition of canola and rapeseed oils. In: *Rapeseed and Canola Oil Production, Processing, Properties and Uses*, 37-78. F.D. Gunstone, ed. Blackwell Publising Ltd.

Sato, T., I. Uezono, T. Morishita, and T. Tetsuka. 1998. Nondestructive Estimation of Fatty Acid Composition in Seeds of *Brassica napus* L by Near-Infrared Spectroscopy. *J. Am. Oil Chem. Soc.* 75(12):1877–1881.

Sato, T. 2008. Nondestructive Measurements of Lipid Content and Fatty Acid Composition in Rapeseeds (*Brassica napus* L.) by Near Infrared Spectroscopy. *Plant Prod. Sci.*11(1):146–150.

Shenk, J.S., J.J. Workman Jr, M.O. Westerhaus. 2008. Chapter 16: Application of NIR Spectroscopy to Agricultural Products. In: *Handbook of Near-Infrared Analysis,* 347-386. D.A.Burns, and E.W. Ciurczak, eds. CRC Press, Boca Raton, Florida.

Siemens, B.J., and J.K. Daun. 2005. Determination of the Fatty Acid Composition of Canola, Flax, and Solin by Near-Infrared Spectroscopy. *J. Am. Oil Chem. Soc.* 82(3):153–157.

Velasco, L., and H.C. Becker. 1998. Estimating the fatty acid composition of the oil in intact-seed rapeseed (*Brassica napus* L.) by near- infrared reflectance spectroscopy. *Euphytica* 101(2):221–230.

Velasco, L., C. Mollers, and H.C. Becker. 1999. Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. *Euphytica* 106(1):79–85.

Welle, R., W. Greten, T. Muller, G. Weber, and H. Wehrmann. 2005. Application of near infrared spectroscopy on-combine in corn grain breeding. *J. Near Infrared Spectrosc.* 13(2):69-75.

Welle, R., K. Zähle, C. Hildebrand, K. Kräling, and W. Greten. 2007. Application of near infrared spectroscopy on-combine for canola breeding. *J. Near Infrared Spectrosc.* 15(5):317–325.

Westad, F., A. Schmidt, and M. Kermit. 2008. Incorporating chemical band-assignment in near infrared spectroscopy regression models. *J. Near Infrared Spectrosc*. 16:265–273.

Williams, P.C., and D.C. Sobering. 1993. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J. Near Infrared Spectrosc.* 1(1):25–32.

Williams, P.C. 2001. Implementation of Near-Infrared Technology. In: *Near-Infrared Technology in the Agricultural and Food Industries*, 145-170. P. Williams, and K. Norris, eds. AACC, Inc., St. Paul, Minnesota.

Williams, P.C., and K. Norris. 2001. Variables Affecting Near-Infrared Spectroscopic Analysis. In: *Near-Infrared Technology in the Agricultural and Food Industries*, 171–198. P. Williams, and K. Norris, eds. AACC, Inc., St. Paul, Minnesota.

Williams, P. 2008. Chapter 12: Sampling, Sample Preparation, and Sample Selection. In: *Handbook of Near-Infrared Analysis,* 267–296. D.A.Burns, and E.W. Ciurczak, eds. CRC Press, Boca Raton, Florida.

Workman Jr, J. 1995. Diode array process instrumentation for the near IR. *Anal. Spectro. Lib.* 6:369-384.

PAPER 2: NON-DESTRUCTIVE ANALYSIS OF GLUCOSINOLATES CONTENT OF CANOLA (*BRASSICA NAPUS*) SEED USING A DIODE ARRAY NEAR INFRARED SPECTROSCOPY

Abstract

The use of canola meal in animal feed rations has made a low content of glucosinolate an absolute necessity. Currently, the maximum acceptable glucosinolate concentration in canola seed is 30 µmol/g; whereas, plant breeders are trying to reduce it further. To accomplish this, a non-destructive and rapid method of determining glucosinolates content will be beneficial to canola breeders. The objective of this study was to develop and evaluate an NIRS calibration model for analyzing total glucosinolates content in intact seed of canola (*Brassica napus)* specifically for 20 g seed. NIR spectra were collected from 100 canola seed samples on a diode array spectrometer using a breeder's cup. The calibration models were evaluated for coefficient of determination (R^2), standard error of prediction (SEP), and ratio performance deviation (RPD). However, a robust NIRS calibration model was not obtained for glucosinolates content, probably due to limited variability and low levels of the constituent together with narrow NIR wavelength range of the NIR instrument.

Introduction

Glucosinolates are sulfur-containing secondary metabolites synthesized by plants, which occur in all economically important varieties of Brassica (Tripathi and Mishra 2007). More than 120 different glucosinolates have been identified, all sharing a common structure comprising of a cyano group, a sulphate, a β-D- glucopyranosyl, and a variable side chain derived from methionine, tryptophan or phenylalanine (Figure 9). Content and composition of glucosinolates differs for every plant species. In *B. napus*, 30 different glucosinolates have been identified and their detailed information is reviewed by Shahidi (1990).

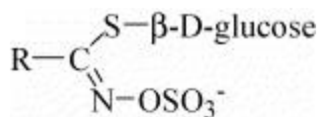


Figure 9. General glucosinolate structure (Tripathi and Mishra, 2007).

Glucosinolates have been shown to have negative health risks on livestock when consumed in high concentrations (Griffiths et al., 1998). A detailed review of negative effects of glucosinolates has been discussed in detail by Mawson et al. (1993) and Tripathi and Mishra (2007). High levels of glucosinolates found in rapeseed meal have restricted the use of this seed as a source of protein in animal feeds. Plant breeding to reduce the level of glucosinolates in rapeseed resulted in the varieties now known as "canola". The current definition of canola requires a total glucosinolates content of less than 30 μmol/g. The target pursued by the Canadian Grain Commission is a further reduction of the generally acceptable maximum glucosinolate concentration in canola to ≤18 μmol/g.

Chemical methods to analyze total glucosinolates content such as HPLC, GLC, glucose release, palladium and thymol tests are destructive to the seed and time consuming (Biston et al. 1988). Near infrared spectroscopy (NIRS), a fast and non-destructive alternative method has been shown to provide an alternative method to traditional techniques for analyzing rapeseed constituents (Petisco et al. 2010, Hom et al. 2007, Velasco et al. 1999, Siemens and Daun 2005, Sato 2008, Velasco and Becker 1998). Much previous successful NIR work has dealt with determining glucosinolates over a wide range (0.6-196.7 μmol/g) for rapeseed using monochromator detectors (Petisco et al, 2010, Hom et al. 2007, Font et al. 2004). Recent photodiode array detectors provide greater accuracy and a short time analysis compared to conventional monochromator detectors (Workman Jr, 1995).

NIR calibration model developed with *B. juncea* variety is currently being used by Canadian Grain Commission for proficiency tests for glucosinolates content in canola. In literature, very little importance is given to NIR work specifically for canola seed. A recent commercial calibration model for glucosinolates content (Perten Instruments, released in 2008) was not suitable to predict glucosinolates content for canola (Personal communication with Dr. Darrin Haagenson, 2012).

A preliminary NIRS calibration for glucosinolates in canola seed on a single plant scale (3 g), failed. The sample size was probably too small, given the low levels of glucosinolates in canola. Quantity of sample is an important factor when it comes to NIR calibration development (Williams 2008). Therefore, increasing the sample size can result in a better calibration model for the minor yet important canola constituent, glucosinolates.

Recognizing the importance of monitoring the level of glucosinolates by canola plant breeders in *B. napus* varieties alone, the objective of this work was to determine if Diode-array NIRS can be useful to analyze total glucosinolates content specifically for *B. napus* varieties on a 20 g scale.

Materials and Methods

Materials

One hundred and thirty canola seed samples were randomly selected from the pool of 2500 as described in Paper 1 (Table 6). Of the 130 samples selected, 100 were used in the calibration and 30 were used for the validation. All selected canola seed samples were cleaned using a Carter Day Dockage Tester to remove foreign material according to the methods of USDA-GIPSA (2004). Cleaned seed were then packed into Mylar bags for long-term storage.

For determination of glucosinolates content, DEAE Sephadex A-25 was obtained from Pharmacia Corporation, USA. Glucose oxidase (50,000 units/1100 mg), peroxidase, 4-aminoantipyrine, and Trizma base were obtained from Sigma-Aldrich Corporation. Myrosinase was obtained from Biocataysts Ltd., USA.

Methods

Reference Chemistry

Total glucosinolates content was determined on 200 mg samples in duplicate by the glucose release method described by Gallaher et al. (2012) with some modifications. These modifications were in accordance to the International Organization for Standardization method reference number ISO 9167–3: 2007 (E) Rapeseed—Determination of glucosinolate content—Part 3: Spectrometric method for total glucosinolates by glucose release.

Collection of Spectra for Canola Seed (20 g)

Spectra were collected at room temperature (24 ± 1 °C) on a DA 7200 NIR spectrometer using a breeder's cup (22 mL small sample dish, Figure 10) developed by Perten Instruments with a 950-1,650 nm wavelength range and a scan resolution of 5 nm. Each 20 g canola sample was scanned twice, repacked and again scanned twice. The average of four scans was used in chemometric analysis.

Figure 10. Breeder's sample cup developed by Perten Instruments.

Experimental Design and Statistical Analysis

NIRS calibration models were built to predict total glucosinolates content of canola seed. Out of 130 canola seed samples, a calibration set and validation set was chosen in such a way that both sets showed the same variance dimensions (Williams, 2001).

Spectral and reference data were exported to the GRAMS Suite statistical software package (Thermo Fisher Scientific; Waltham, MA, USA). Eighteen different calibration models were built for total glucosinolates content; these models included combinations of four pre-processing techniques (mean center, Savitzky-Golay 1st derivative, Savitzky-Golay 2nd derivative, standard normal variate) and three regression models (PLS-1, PLS-2, PCR).

Each NIRS model was evaluated against the 30 validation samples to choose the best calibration model. To determine the accuracy of the developed NIRS model, the coefficient of determination (R^2), standard error of prediction (SEP), and ratio performance deviation (RPD), were computed. Out of all the developed calibrations models, the best calibration model for each constituent was selected on the basis of the highest R^2 and RPD values and lowest SEP values of the validation set.

Results and Discussion

Reference Analysis

For 130 canola seed samples, total glucosinolates content was determined. The statistics of glucosinolates content of canola seed is presented in Table 10 with further

distribution of calibration and validation set statistics. The glucosinolates content varied between 1.5 – 49.0 µmol/ g seed; however, only2 samples had a glucosinolates content > 30 µmol/ g seed (Figure 11). According to the definition of canola by the Canadian Grain Commission, the glucosinolates content of canola seed must not exceed 30 µmol/ g seed. Glucosinolate content of 15.8 – 39.0 µmol/ g seed was reported for *B. napus* varieties by Petisco et al. (2010). Very high variation (0.6 – 118.9 µmol/ g seed) in glucosinolate content of *B. napus* seed was reported by Hom et al. (2007). However, in that study highly inbred *B. napus* seed were included to increase the variation in glucosinolates content.

Table 10. Summary of reference chemistry data for total glucosinolates content in canola seed**.**

No. of samples (Sample set)	**Range (µmol/g)**	**Mean (µmol/g)**	**SD[a] (µmol/g)**
130 (All)	1.5 – 49.0	9.7	5.6
100 (Calibration set)	1.5 – 49.0	9.8	6.0
30 (Validation set)	1.8 – 19.2	10.5	4.8

[a]standard deviation

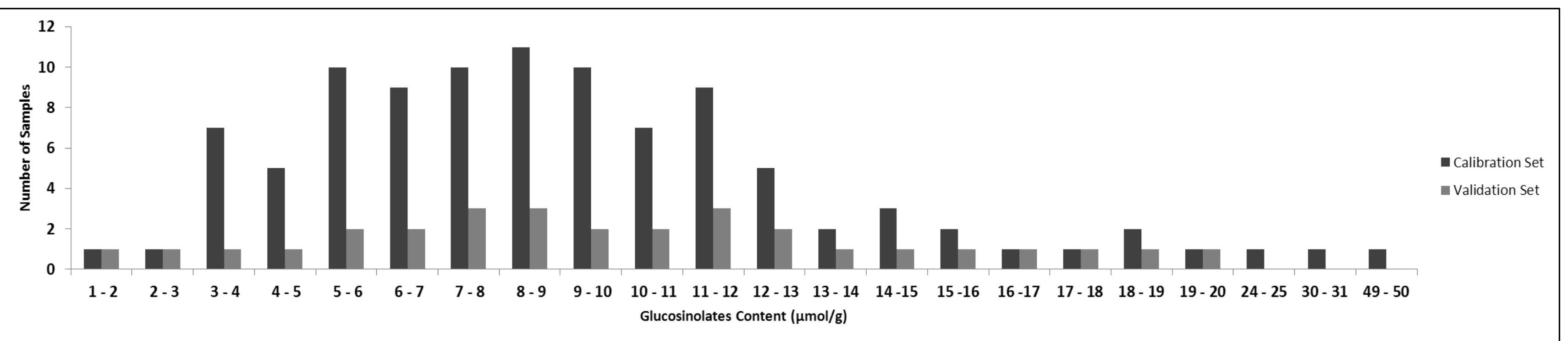


Figure 11. Distribution plot of calibration and validation sample sets for glucosinolates content.

NIRS Model Evaluation

NIRS calibration models were developed to predict glucosinolates content in canola seed. No combination of algorithm and preprocessing methods provided an acceptable calibration model for glucosinolates content in this study according to the guidelines on acceptability of NIRS models (Table 11). Validation RPD ranged between 0.8 – 1.5, and validation R^2 ranged between 0.40 – 0.55 for this study, whereas an RPD of at least 3, and R^2 of at least 0.66 is recommended (Williams, 2001). Figure 12 shows the scatterplot of predicted glucosinolates content (using the best calibration model obtained in this study) versus the reference glucosinolates content. For all the calibration models, calibration R^2 ranged between 0.40 – 0.49.

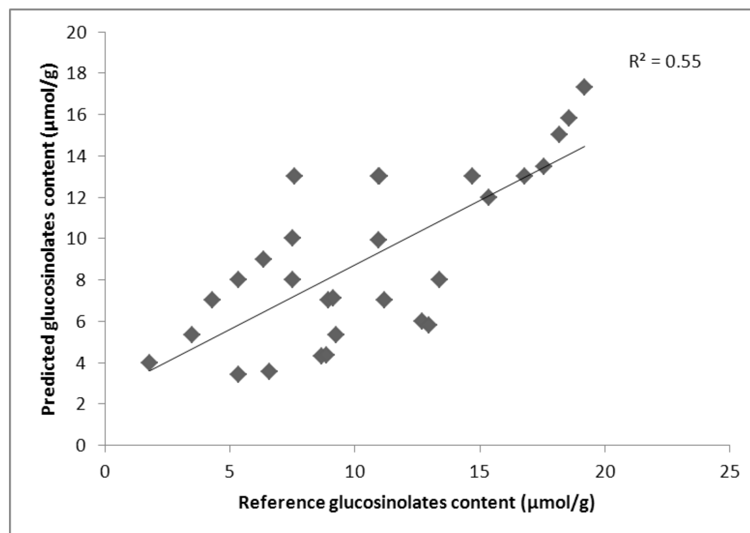


Figure 12. Scatterplot of NIRS predicted values vs. reference values for glucosinolates content (µmol/g) of canola seed (values were predicted using calibration model developed with PCR regression algorithm and mean center preprocessing method).

Table 11. Calibration and validation statistics for the NIRS calibration equations obtained for the prediction of glucosinolates content (µmol/g seed) of 20 g canola seeds.

Calibration Type	**Preprocessing**	**Calibration**		**Validation**		
		R^{2}[a]	**SEC[b]**	**R^{2}[a]**	**SEP[c]**	**RPD[d]**
PLS[e]-1	MC[g]	0.41	4.10	0.46	3.89	1.2
	MC + SNV[h]	0.43	3.98	0.42	3.75	1.3
	MC + SG1st[i]	0.45	4.56	0.47	3.49	1.4
	MC + SNV+ SG1st	0.41	4.07	0.41	4.18	1.1
	MC + SG2nd[j]	0.47	4.45	0.45	4.48	1.1
	MC + SNV+ SG2nd	0.44	3.25	0.49	3.86	1.2
PLS-2	MC	0.49	3.41	0.47	3.43	1.4
	MC + SNV	0.40	3.49	0.49	3.51	1.4
	MC + SG1st	0.47	3.39	0.40	3.44	1.4
	MC + SNV+ SG1st	0.41	4.41	0.48	3.43	1.4
	MC + SG2nd	0.45	3.43	0.41	5.47	0.8
	MC + SNV+ SG2nd	0.45	3.42	0.42	4.45	1.1
PCR[f]	MC*	0.48	3.59	0.55	3.18	1.5
	MC + SNV	0.47	4.28	0.47	3.69	1.3
	MC + SG1st	0.42	4.61	0.45	3.68	1.3
	MC + SNV+ SG1st	0.44	4.53	0.49	3.63	1.3
	MC + SG2nd	0.46	4.58	0.40	3.68	1.3
	MC + SNV+ SG2nd	0.42	3.68	0.48	3.67	1.3

[a]coefficient of determination, [b]standard error of calibration, [c]standard error of prediction, [d]ratio performance deviation, [e]partial least squares algorithm, [f]principal component regression, [g]mean center, [h]standard normal variate, [i]Savitzky-Golay 1st derivative, [j]Savitzky-Golay 2nd derivative, * best calibration model

Low correlation obtained in this study between the reference method and NIR might be due to the low levels of glucosinolates content and the narrow wavelength range of 950-1650 nm for the diode-array NIR used in this study. For rapeseed glucosinolates determination by NIR, the inclusion ofwavelengths beyond 1650 nm has been recommended: 1680, 1734, 1759, 1776, 1778, 1982, 2139, 2190, 2208 and 2230 nm (Starr et al. 1985, Biston et al., 1988, Daun et al., 1994). Use of a Foss NIR spectrophotometer (NIR Systems model 6500, NIR Systems, Inc., Silver Springs, MD), covering the range of

400 – 2500 nm, has been successfully used to collect spectra for rapeseed (Velasco et al. 1998, Velasco et al. 1999, Hom et al. 2007).

Tkachuk (1981) investigated the measurement of glucosinolates in rapeseed (*B. napus* and *B. campestris*) and obtained an R^2 value of 0.50, which is very similar to this work. This author concluded that glucosinolates content of rapeseed cannot be measured satisfactorily with the near-infrared region alone, and recommended combining the visible region along with NIR region to develop calibration model for glucosinolates in rapeseed. Similar conclusion of inability of NIRS to predict glucosinolates content in *B. napus* varieties was demonstrated by Starr et al. (1985).

Some of the latest NIR work on determination of glucosinolates has supported that glucosinolates in rapeseed can be determined using theNIR region (Font et al. 2004, Font et al. 2006, Hom et al. 2007, Petisco et al. 2010). However, all of this previous work was done with a wide NIR range of 1100-2500 nm, and included sample sets of other *Brassica* species different to those studied in the present work to increase the variability in glucosinolates content. Glucosinolates range of 15.8 – 97.9 µmol/g seed has been used to develop global NIRS calibration model for rapeseed by including seed from *B. napus* and *B. carinata* species (Petisco et al. 2010). This author obtained interesting results as RPD value declined from 10 to 2.34 when only *B. napus* species(16.8 – 39 µmol/g) was used for validation compared to both *B. napus* and *B. carinata* species. In a different study, glucosinolates variation of 0.6 – 118.9 µmol/g in *B. napus* seed was used to develop NIRS calibration model. However, the authors used highly inbred varieties of *B. napus* to

increase the glucosinolates variation to develop better NIRS calibration models (Hom et al. 2007).

An NIR calibration model specifically for low glucosinolates range for canola seed (9.7–30.3 μmol/g) was developed by Daun et al. (1994). This author used an NIR wavelength range up to 2230 nm and achieved better R^2 value of 0.82 compared to R^2 of 0.55 obtained in this present work. However, RPD value (1.4) obtained by Daun et al. is consistent with the range of RPD values (0.8–1.5) obtained in the present work.

Conclusions

In conclusion, a satisfactory calibration model for prediction of glucosinolates content in canola seed was not obtained due to the limited variability and low levels of glucosinolate content and a narrow wavelength range of the DA-NIR instrument. However, in the future, robust NIRS models can be developed by increasing constituent variability. To do that, a similar approach can be adapted as Petisco et al. (2010) by developing a global NIR calibration model by including other Brassica species (*B. carinata, B. juncea*) to increase variability of glucosinolate content. This global calibration model can be used to monitor glucosinolate content specifically in canola seed. Inbred varieties can also be used to develop NIRS calibration model for glucosinolate content in accordance to the work done by Hom et al. (2007).

References

Biston R (1988) Fast analysis of Rapeseed Glucosinolates by Near Infrared Reflectance Spectroscopy. J Am Oil Chem Soc 65:1599-1600

Daun J M, Clear K M, Williams P (1994) Comparison of Three Seed Near-Infarred Analyzers for Measuring Quality Components of Canola Seed. J Am Oil Chem Soc 71:1063-1068

Font R, Del Río M, Fernández-Martínez J M, De Haro-Bailón A (2004) Use of near-infrared spectroscopy for screening the individual and total glucosinolate contents in Indian mustard seed (*Brassica juncea* L. Czern. & Coss.). J Agric Food Chem 52(11):3563–3569

Font R, del Río-Celestino M, de Haro-Bailón A (2006) The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. Ind Crop Prod 24:307–313

Gallaher C M, Gallaher D D, Peterson S (2012) 1. Development and Validation of a Spectrophotometric Method for Quantification of Total Glucosinolates in Cruciferous Vegetables. J Agr Food Chem 60 (6):1358-1362

Griffithsa D W, Deightona N, Nicholas E Bircha A, Patrianb B, Baurb R, Städlerb E (2001) Identification of glucosinolates on the leaf surface of plants from the Cruciferae and other closely related species. Phytochem 57(2):131-140

Hom NH, Becker HC, Mollers C (2007) Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. Euphytica 153:27–34

Mawson R, Heaney R K, Piskula M, Kozlowska H (1993) Rapeseed meal-glucosinolates and their antinutritional effects Part 1. Rapeseed production and chemistry of glucosinolates. Mol Nutr Food Res 37(2):131-140

Petisco C, Garcia-Criado B, Vazquez-de-Aldana BR, de Haro A, Garcia-Ciudad A (2010) Measurement of quality parameters in intact seeds of Brassica species using visible and near-infrared spectroscopy. Ind Crop Prod 32 (2):139–146

Sato T (2008) Nondestructive Measurements of Lipid Content and Fatty Acid Composition in Rapeseeds (*Brassica napus* L.) by Near Infrared Spectroscopy. Plant Prod Sci11:146–150

Shahidi F (1990) Glucosinolates: Structure-Properties-Fuction. In Shahidi F (eds) Canola and Rapeseed: Production, Chemistry, Nutrition, and Processing Technology, Von Nostrand Reinhold, New York, NY, pp 149-231

Siemens B J, Daun J K (2005) Determination of the Fatty Acid Composition of Canola, Flax, and Solin by Near-Infrared Spectroscopy. J Am Oil Chem Soc 82(3):153–157

Starr C, Suttle J, Morgan A G, Smith D B (1985) A comparison of sample preparartion and calibartion techniques for the estimation of nitrogen, oil and glucosinolate content of rapeseed by near infarred spectrscopy. J Agric Sci 104:317-323

Tkachuk R (1981) Oil and Protein Analysis of Whole Rapseed Kernels by Near Infrared Reflectance Spectroscopy. J Am Oil Chem Soc 58:819-822

Tripathi M K, Mishra A S (2007) Glucosinolates in animal nutrition: A review. Anim Feed Sci Technol 132:1–27

Velasco L, Becker H C (1998) Estimating the fatty acid composition of the oil in intact-seed rapeseed (*Brassica napus* L.) by near- infrared reflectance spectroscopy. Euphytica 101:221–230

Velasco L, Mollers C, Becker H C (1999) Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Euphytica 106:79–85

Williams, P (2008) Sampling, Sample Preparation, and Sample Selection. In: Burns D A, Ciurczak E W (eds) Handbook of Near-Infrared Analysis, 3rd edn. CRC Press, Boca Raton, Florida, pp 267–296

GENERAL CONCLUSIONS AND RECOMMENDATIONS

General Conclusions

One of the primary objectives of the study was to evaluate the performance of NIRS commercial calibration model to predict composition of single plant canola seed. This commercial calibration model was developed for use with 20g canola seed, whereas; spectra were collected utilizing 3g canola seed. Reducing the sample size from 20 to 3g on a commercial calibration model reduced the accuracy of the predictions. Therefore, this study focused on the development of NIRS calibration models on a single plant scale for canola seed.

The mirrored cup permitted the development of NIRS calibration models to predict moisture and oil content on a single plant scale. These NIRS models were robust enough for quality assurance purposes based on their R^2 and RPD. For oleic acid content, the NIRS calibration model obtained is suitable for rough screening. There is a lot of room for improvement to develop a better calibration model for oleic acid. For all these constituents, PLS (Partial Least Square) algorithm worked better than the PCR (Principal Component Regression) algorithm.

The results of the present work also demonstrated the good performance of NIRS with minimal sample preparation. The developed calibration models can be used without a drying step, which will allow high throughput prediction of canola seeds for breeding purposes. Furthermore; for storage purposes, Mylar bags showed good performance in maintaining moisture content of the storage canola seed (Appendix A).

Satisfactory results were not obtained for saturated fatty acid and glucosinolates content at a single plant scale due to the limited variability in constituents and the relatively narrow wavelength range of the spectrophotometer. A separate calibration model was developed specifically for glucosinolates content at 20g scale, which showed some promise for the rough screening application.

In general conclusion, NIRS calibration models developed for moisture content, oil content, and oleic acid content at the single plant scale in this study provided rapid, inexpensive, non-destructive, and reliable predictions. These NIRS calibration models developed at a single plant scale is extremely important to plant breeders and will provide the benefit of analyzing seed obtained from the single plant without the need of mixing it with seed from different plants just to increase the size. This application in plant breeding will enable high throughput screenings of canola seed for the above mentioned constituents at a single plant scale.

Recommendations for Future Work

The NIRS method showed a promising alternative to wet chemistry methods to predict moisture content and oil content. However, further study is needed to improve NIRS calibration models for predicting fatty acids and glucosinolates content. These improvements can be accomplished by following:

1) Utilizing an NIR instrument that can collect spectra up to 2,500 nm in which the combination bands from 1800 to 2500 nm are included.

2) Increasing the variability in constituents by including samples from different Brassica species.
3) Increasing the variability in constituents by including inbred varieties (obtained by crossing canola and traditional rapeseed).

APPENDIX A: STUDY ON PERFORAMCE OF MYLAR BAGS IN STORING CANOLA SEED

A study was conducted to check the performance of Mylar bags to retain the moisture of canola seed during storage. Mylar bags, made of polyester resin and laminated to aluminum foil layer, provide moisture barrier properties with a Moisture Vapor Transmission Rate (MVTR) of less than 0.078 $g/m^2/d$.

Twenty canola seed samples were tested for moisture content (% dry basis) before storing into Mylar bags. Mylar bags were stored in a cold room. Seed moisture content was determined after every 3 months for 9 months.

Moisture content was determined by drying 4 g samples in duplicate at 103°C for 5 h using a gravity convection oven (Precision Scientific Inc; Winchester, IL, USA) according to the ISO 665 method for oilseeds.

Results showed that the Mylar bags were very effective in retaining moisture content of canola seed (Table A1). The mean of moisture contents obtained for every time interval of 3 months did not differ significantly at $P \leq 0.05$.

Table A1. Moisture content (%) of canola seed stored in Mylar bags over the period of 9 months.

Time (months)	Range (%)	Mean (%)	S D (%)
0	3.5-8	6.5	1.3
3	3.3-7.9	6.4	1.3
6	3.2-7.9	6.9	1.4
9	2.9-7.8	6.4	1.4

APPENDIX B: STATISTICAL DATA OF DIFFERENT NIRS CALIBARTION MODELS

Different NIRS models developed for moisture content, oil content, oleic acid content, and saturated fatty acid content using three different regression algorithms and four different preprocessing techniques (Paper 2). For every constituent, 18 different models were developed using these combinations (Table13, 14, 15, and 16).

Table B1. Calibration and validation statistics for the equations obtained for the prediction of moisture content (%) of 3 g canola seeds.

Calibration Type	Preprocessing	Calibration		Validation		
		R^2	SEC	R^2	SEP	RPD
PLS-1	MC	0.93	0.38	0.92	0.35	5.8
	MC + SNV	0.93	0.38	0.93	0.38	5.1
	MC + SG1st	0.94	0.39	0.93	0.40	4.8
	MC + SNV+ SG1st	0.94	0.40	0.93	0.45	4.3
	MC + SG2nd	0.95	0.37	0.94	0.34	5.8
	MC + SNV+ SG2nd	0.96	0.40	0.97	0.32	6.1
PLS-2	MC	0.89	0.41	0.87	0.43	4.6
	MC + SNV	0.89	0.41	0.89	0.43	4.6
	MC + SG1st	0.91	0.42	0.80	0.44	4.5
	MC + SNV+ SG1st	0.91	0.41	0.88	0.43	4.6
	MC + SG2nd	0.94	0.43	0.91	0.47	4.1
	MC + SNV+ SG2nd	0.93	0.42	0.92	0.45	4.4
PCR	MC	0.79	0.59	0.76	0.62	3.1
	MC + SNV	0.85	0.58	0.82	0.69	2.8
	MC + SG1st	0.86	0.61	0.87	0.68	2.9
	MC + SNV+ SG1st	0.84	0.53	0.81	0.63	3.1
	MC + SG2nd	0.86	0.58	0.85	0.68	2.9
	MC + SNV+ SG2nd	0.82	0.68	0.79	0.67	2.9

R^2 = coefficient of determination, SEC = standard error of calibration, SEP = standard error of prediction, RPD = ratio performance deviation, PLS = partial least squares algorithm, PCR = principal component regression, MC = mean center, SNV = standard normal variate, SG1st = Savitzky-golay 1st derivative, SG2nd = Savitzky-golay 2nd derivative

Table B2. Calibration and validation statistics for the equations obtained for the prediction of oil content (dry basis, %) of 3 g canola seeds.

Calibration Type	Preprocessing	Calibration		Validation		
		R^2	SEC	R^2	SEP	RPD
PLS-1	MC	0.82	0.81	0.84	0.61	4.2
	MC + SNV	0.79	0.87	0.75	0.67	3.8
	MC + SG1st	0.80	0.85	0.81	0.71	3.6
	MC + SNV+ SG1st	0.80	0.78	0.78	0.73	3.5
	MC + SG2nd	0.80	0.80	0.76	0.71	3.6
	MC + SNV+ SG2nd	0.81	0.87	0.80	0.62	4.1
PLS-2	MC	0.79	0.85	0.75	0.87	2.9
	MC + SNV	0.74	0.87	0.76	0.85	3.0
	MC + SG1st	0.78	0.84	0.78	0.80	3.2
	MC + SNV+ SG1st	0.78	0.85	0.79	0.87	2.9
	MC + SG2nd	0.81	0.87	0.80	0.84	3.1
	MC + SNV+ SG2nd	0.80	0.82	0.82	0.82	3.1
PCR	MC	0.75	0.89	0.72	0.90	2.8
	MC + SNV	0.72	0.88	0.78	0.85	3.0
	MC + SG1st	0.73	0.88	0.76	0.87	2.9
	MC + SNV+ SG1st	0.75	0.85	0.74	0.84	3.1
	MC + SG2nd	0.79	0.84	0.76	0.87	2.9
	MC + SNV+ SG2nd	0.80	0.81	0.81	0.85	3.0

R^2 = coefficient of determination, SEC = standard error of calibration, SEP = standard error of prediction, RPD = ratio performance deviation, PLS = partial least squares algorithm, PCR = principal component regression, MC = mean center, SNV = standard normal variate, SG1st = Savitzky-golay 1st derivative, SG2nd = Savitzky-golay 2nd derivative

Table B3. Calibration and validation statistics for the equations obtained for the prediction of oleic acid content (%) of 3 g canola seeds.

Calibration Type	Preprocessing	Calibration		Validation		
		R^2	SEC	R^2	SEP	RPD
PLS-1	MC	0.78	0.72	0.75	0.79	1.5
	MC + SNV	0.79	0.72	0.78	0.56	2.1
	MC + SG1st	0.79	0.75	0.78	0.60	1.9
	MC + SNV+ SG1st	0.80	0.75	0.76	0.67	1.8
	MC + SG2nd	0.81	0.69	0.86	0.46	2.6
	MC + SNV+ SG2nd	0.81	0.70	0.85	0.49	2.4
PLS-2	MC	0.75	0.78	0.79	0.62	1.9
	MC + SNV	0.78	0.75	0.75	0.67	1.8
	MC + SG1st	0.79	0.76	0.73	0.68	1.7
	MC + SNV+ SG1st	0.82	0.79	0.78	0.70	1.7
	MC + SG2nd	0.81	0.73	0.79	0.71	1.7
	MC + SNV+ SG2nd	0.79	0.75	0.81	0.70	1.7
PCR	MC	0.78	0.81	0.75	0.83	1.4
	MC + SNV	0.76	0.84	0.75	0.76	1.6
	MC + SG1st	0.71	0.87	0.73	0.83	1.4
	MC + SNV+ SG1st	0.73	0.81	0.81	0.78	1.5
	MC + SG2nd	0.74	0.79	0.78	0.49	2.4
	MC + SNV+ SG2nd	0.79	0.87	0.81	0.79	1.5

R^2 = coefficient of determination, SEC = standard error of calibration, SEP = standard error of prediction, RPD = ratio performance deviation, PLS = partial least squares algorithm, PCR = principal component regression, MC = mean center, SNV = standard normal variate, SG1st = Savitzky-golay 1st derivative, SG2nd = Savitzky-golay 2nd derivative

Table B4. Calibration and validation statistics for the equations obtained for the prediction of saturated fatty acid (C16:0 + C18:0) content (%) of 3 g canola seeds.

Calibration Type	Preprocessing	Calibration		Validation		
		R^2	SEC	R^2	SEP	RPD
PLS-1	MC	0.10	1.10	0.18	0.89	0.6
	MC + SNV	0.12	0.48	0.17	0.75	0.7
	MC + SG1st	0.17	0.48	0.15	0.49	1.1
	MC + SNV+ SG1st	0.16	1.13	0.09	1.18	0.4
	MC + SG2nd	0.17	0.45	0.19	0.48	1.1
	MC + SNV+ SG2nd	0.17	1.25	0.18	0.86	0.6
PLS-2	MC	0.15	0.89	0.15	0.49	1.1
	MC + SNV	0.10	0.87	0.18	0.58	0.9
	MC + SG1st	0.14	0.84	0.10	0.89	0.6
	MC + SNV+ SG1st	0.13	1.09	0.17	1.16	0.5
	MC + SG2nd	0.12	0.95	0.08	0.86	0.6
	MC + SNV+ SG2nd	0.16	0.87	0.17	0.76	0.7
PCR	MC	0.12	0.95	0.11	0.52	1.0
	MC + SNV	0.16	0.86	0.18	0.86	0.6
	MC + SG1st	0.19	0.99	0.09	0.87	0.6
	MC + SNV+ SG1st	0.09	1.50	0.19	0.76	0.7
	MC + SG2nd	0.13	0.97	0.18	0.59	0.9
	MC + SNV+ SG2nd	0.15	0.75	0.08	1.06	0.5

R^2 = coefficient of determination, SEC = standard error of calibration, SEP = standard error of prediction, RPD = ratio performance deviation, PLS = partial least squares algorithm, PCR = principal component regression, MC = mean center, SNV = standard normal variate, SG1st = Savitzky-golay 1st derivative, SG2nd = Savitzky-golay 2nd derivative

APPENDIX C: SUMMARY OF LITERATURE REPORTS OF PREDICTION OF OIL CONTENT, FATTY ACIDS, AND TOTAL GLUCOSINOLATES CONTENT BY NIR

Table C1. Summary of literature reports of canola seeds prediction by NIR.

Constituent	Variety	Reference Chemistry	Sample Size	Calibration n	Range	R^2	SEC
Oil content (%)							
Oil	*B. napus* & *B. carinata*	NMR	4 g	86	34.1 – 48.4	0.98	0.51
Oil	*B. napus*	Gravimetric method	single seed	206	26.2 – 61.1	0.98	0.98
Oil (as%)[a]	*B. napus*	-	single seed	125	28.5 – 54.9		
Total Glucosinolates Content (µmol/g)							
TGC	-	HPLC	12 g	74	4 – 100	0.99	2.84
TGC	*B. napus*	HPLC	120 g	-	9.7 – 30.3	0.90	2.10
TGC	*B. napus* & *B. carinata*	HPLC	4 g	84	15.8 – 97.9	0.99	2.57
TGC	*B. napus*	HPLC	single seed	111	0.6 – 118.9	0.97	5.01
TGC	*B.juncea*	HPLC	-	139	16.1– 196.7	-	-
Fatty Acid (%)							
C 18:1	*B. napus*	GC	2.7 g	30	15.4 – 65.9	0.86	8.31
C 18:1	*B. napus*	GC	single seed	219	50.8 – 84.5	0.76	3.39
C 18:1	*B. napus*	GC	120 g	704	55.1 – 76.5	-	-
Total SFA[b]	*B. rapa* & *B. napus*	GC	120 g	707	5.2 – 8.8	-	-
C 16:0	*B. rapa* & *B. napus*	GC	120 g	610	2.9 – 4.7	-	-
C 16:0	*B. napus*	GC	2.7 g	30	2.9 – 4.8	0.42	0.43
C 16:0	*B. napus*	GC	single seed	30	2.9 – 4.8	0.38	0.45
C 18:0	*B. rapa* & *B. napus*	GC	120 g	605	1.4 – 2.8		
C 18:0	*B. napus*	GC	2.7 g	30	1.2 – 3.5	0.74	0.26
C 18:0	*B. napus*	GC	Single seed	30	1.2 – 3.5	0.54	0.35
C 16:0	*B. napus*	GLC	3 g	220	2.7 – 6.1	0.85	0.26
C 16:0	*B. napus*	GLC	300 mg	220	2.7 – 6.1	0.79	0.31
C 18:0	*B. napus*	GLC	3 g	220	0.6 –2.8	0.67	0.2
C 18:0	*B. napus*	GLC	300 mg	220	0.6 –2.8	0.74	0.18
C 16:0	*B. napus*	GLC	60 mg	220	2.7 – 6.1	0.68	0.37
C 18:0	*B. napus*	GLC	60 mg	220	0.6 – 2.8	0.68	0.20
C 20:0	*B. napus*	GC	2.7 g	30	0.3 – 1.3	0.73	0.13
C 20:0	*B. napus*	GC	Single seed	30	0.3 – 1.3	0.38	0.45

R2 = coefficient of determination, SEC = standard error of calibration, SECV = standard error of cross validation, SEP = standard error of prediction, RPD = ratio performance deviation. [a] Oil content expressed on an 'asis' basis. [b] Sum saturated FA (C14:0, C16:0, C18:0, C20:0, C22:0, C24:0)

Table C1. Summary of literature reports of canola seeds prediction by NIR (contd.).

Constituent	Cross Validation		Validation			
	R^2	SECV	n	R^2	SEP	RPD
Oil content (%)						
Oil	-	-	47	0.98	0.54	6.50
Oil	0.97	1.14	-	-	-	-
Oil (as%)[a]	0.88	1.98	35	0.85	1.87	-
Total Glucosinolates Content (µmol/g)						
TGC	-	-		0.82	2.5	1.36
TGC	-	-	49	0.99	2.6	10.0
TGC	-	-	-		-	-
TGC	0.86	10.3	-		-	-
TGC	-	-	69	0.82	-	2.18
Fatty Acid (%)						
C 18:1	-	-	-	-	-	-
C 18:1	-	-	29	0.83	2.7	2.4
C 18:1	-	0.62	-	-	-	-
Total SFA[b]	-	0.19	-	-	-	-
C 16:0	-	0.13	-	-	-	-
C 16:0	-	-	30	-	-	-
C 16:0	-	-	30	-	-	-
C 18:0	-	0.11	-	-	-	-
C 18:0	-	-	30	-	-	-
C 18:0	-	-	30	-	-	-
C 16:0	0.76	0.33	-	-	-	-
C 16:0	0.72	0.35	-	-	-	-
C 18:0	0.62	0.22	-	-	-	-
C 18:0	0.67	0.20	-	-	-	-
C 16:0	0.64	0.39	-	-	-	-
C 18:0	0.60	0.22	-	-	-	-
C 20:0	-	-	30	-	-	-
C 20:0	-	-	30	-	-	-

R2 = coefficient of determination, SEC = standard error of calibration, SECV = standard error of cross validation, SEP = standard error of prediction, RPD = ratio performance deviation. [a] Oil content expressed on an 'asis' basis. [b] Sum saturated FA (C14:0, C16:0, C18:0, C20:0, C22:0, C24:0)

Table C1. Summary of literature reports of canola seeds prediction by NIR (contd.).

Constituent	Prediction					References
	n	Range	R^2	SEP	RPD	
			Oil content (%)			
Oil	-	-	-	-	-	Petisco et al. 2010
Oil	-	-	-	-	-	Hom et al. 2007
Oil (as%)[a]	-	-	-	-	-	Velasco et al. 1999
			Total Glucosinolates Content (µmol/g)			
TGC	20	4 - 87	0.99	-	-	Biston et al. 1988
TGC	-	-	-	-	-	Daun et al. 1994
TGC	-	-	-	-	-	Petisco et al. 2010
TGC	-	-	-	15.65	-	Hom et al. 2007
TGC	-	-	-	-	-	Font et al. 2004
			Fatty Acid (%)			
C 18:1	-	-	-	-	-	Sato 2008
C 18:1	-	-	-	-	-	Velasco et al. 1999
C 18:1	997	54.1 – 75.5	0.91	0.77	3.5	Siemens and Daun 2005
Total SFA[b]	997	5.1-9.3	0.87	0.23	2.8	Siemens and Daun 2005
C 16:0	997	2.9 - 4.9	0.82	0.13	2.3	Siemens and Daun 2005
C 16:0	-	-	-	-	-	Sato 2008
C 16:0	-	-	-	-	-	Sato 2008
C 18:0	997	1.3 - 2.7	0.73	0.13	1.9	Siemens and Daun 2005
C 18:0	-	-	-	-	-	Sato 2008
C 18:0	-	-	-	-	-	Sato 2008
C 16:0	-	-	-	-	-	Velasco and Becker 1998
C 16:0	-	-	-	-	-	Velasco and Becker 1998
C 18:0	-	-	-	-		Velasco and Becker 1998
C 18:0	-	-	-	-	-	Velasco and Becker 1998
C 16:0	-	-	-	-	-	Velasco and Becker 1998
C 18:0	-	-	-		-	Velasco and Becker 1998
C 20:0	-	-	-	-	-	Sato 2008
C 20:0	-	-	-	-	-	Sato 2008

R2 = coefficient of determination, SEC = standard error of calibration, SECV = standard error of cross validation, SEP = standard error of prediction, RPD = ratio performance deviation. [a] Oil content expressed on an 'asis' basis. [b] Sum saturated FA (C14:0, C16:0, C18:0, C20:0, C22:0, C24:0)

References

Biston R (1988) Fast analysis of Rapeseed Glucosinolates by Near Infrared Reflectance Spectroscopy. J Am Oil Chem Soc 65:1599-1600

Daun J M, Clear K M, Williams P (1994) Comparison of Three Seed Near-Infarred Analyzers for Measuring Quality Components of Canola Seed. J Am Oil Chem Soc 71:1063-1068

Font R, Del Río M, Fernández-Martínez J M, De Haro-Bailón A (2004) Use of near-infrared spectroscopy for screening the individual and total glucosinolate contents in Indian mustard seed (*Brassica juncea* L. Czern. & Coss.). J Agric Food Chem 52(11):3563–3569

Hom N H, Becker H C, Mollers C (2007) Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. Euphytica 153:27–34

Petisco C, Garcia-Criado B, Vazquez-de-Aldana BR, de Haro A, Garcia-Ciudad A (2010) Measurement of quality parameters in intact seeds of Brassica species using visible and near-infrared spectroscopy. Ind Crop Prod 32 (2):139–146

Sato T (2008) Nondestructive Measurements of Lipid Content and Fatty Acid Composition in Rapeseeds (*Brassica napus* L.) by Near Infrared Spectroscopy. Plant Prod Sci11:146–150

Siemens B J, Daun J K (2005) Determination of the Fatty Acid Composition of Canola, Flax, and Solin by Near-Infrared Spectroscopy. J Am Oil Chem Soc 82(3):153–157

Velasco L, Becker H C (1998) Estimating the fatty acid composition of the oil in intact-seed rapeseed (*Brassica napus* L.) by near- infrared reflectance spectroscopy. Euphytica 101:221–230

Velasco L, Mollers C, Becker H C (1999) Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Euphytica 106:79–85