DEVELOPMENT OF PREDICTION MODELS FOR RAPID ANALYSIS OF GLUCOSINOLATE CONTENT IN CANOLA MEAL USING NEAR-INFRARED SPECTROSCOPY

By

Andy Q. Che

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
In partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

Department of Animal Science
University of Manitoba
Winnipeg, Manitoba, Canada
R3T 2N2

Copyright © Andy Q. Che

ABSTRACT

Canola meal (CM) is a valuable protein source for monogastric animals. Despite its effective use as a protein supplement, variation in amino acid digestibility, the presence of glucosinolates (GLS), and high dietary fibre content can limit its inclusion levels in poultry and swine diets. The nutritive value of canola meal often varies due to differences in its chemical composition, which is directly related to the seed processing conditions, particularly the heat and moisture treatments. Excessive heat treatment during processing can result in the destruction of heat sensitive amino acids turning them to biologically unavailable derivatives, and at the same time it can contribute to the decomposition of glucosinolates. Therefore, the glucosinolate contents can serve as an indirect indicator of protein damage in canola meal. As conducting direct wet chemistry analysis on a rapid and large scale might not be feasible, near infrared (NIR) spectroscopy can be used as a rapid, reliable, user-friendly, and environmentally friendly technique to predict components of the chemical composition of canola meal. Therefore, the aim of the project was to determine if NIR models could be developed to predict Canadian canola meal quality, including the measurement of total glucosinolates content. Canola meal and expeller-cold pressed canola samples were collected from various crushing plants over multiple years and were used to develop the NIR prediction models to predict glucosinolates, crude protein, and fat contents in pellet and mash canola meal. Reference methods were used to analyse each sample, which was scanned three times on multiple near-infrared spectrometers. The average scan of each sample was associated with the wet-chemistry data to develop the NIR models, and the instrument's software was used to perform the statistical analysis for model development. The total glucosinolates content in canola meal was not uniformly distributed across the concentration range. Majority of

samples from the conventional pre-press solvent extracted process exhibited very low levels of total glucosinolates, whereas a limited number of samples had high total glucosinolates.

Three prediction models were developed for the prediction of glucosinolate contents, crude protein and ether extract. Poor correlation was observed in Model I for glucosinolates prediction in canola meal (R^2 = 0.0005), and for glucosinolates and crude protein prediction in expeller-cold pressed canola (ECPC), respectively. (R^2 = 0.49 and 0.61). Excellent correlation was observed in Model I for ether extract prediction in ECPC (R^2 = 0.95). Adjustments to Model I resulted in greatly improved correlations for glucosinolates, crude protein and ether extract in Models II for prediction of mash canola meal (R^2 = 0.87, 0.91, 0.74), and Models III for predictions of pellet canola meal (R^2 = 0.91, 0.83, 0.82). Given its precise predictions, each of the three models has components which can be utilized for qualitative analysis in the feed industry. Model I can be used for ether extract prediction in ECPC. Model II can be used for glucosinolates and crude protein prediction in mash canola meal. Model III can be used for glucosinolates, crude protein, and ether extract prediction in pellet canola meal. Overall, these studies underscore the significant potential of NIRS technology within the feed industry as an effective and efficient tool for assessing canola meal quality by predicting glucosinolate, crude protein, and ether extract contents.

DEDICATION

I would first like to dedicate this thesis and Master of Science degree to my incredible parents, lawyer Chế Quang Định, and Đặng Thị Hòng Oanh for a lifetime of nurture and support (and not pressuring me into giving them grandbabies yet). Their escape from war-torn South Vietnam and subsequent sacrifices paved the way to a new life which allowed me to grow up in a land of opportunity and strive for accomplishments in life and academia. Dad, your lawyer's degree was unjustifiably and painfully ripped away but your resiliency along with mom's mental toughness laid the groundwork to this journey. I am forever grateful.

Also, I would also dedicate my years of education to my late uncle Đặng Phú Xuân – a very gifted academic – and my precious Nana, both of whom passed before I was born. They have given me strength and watched over me my entire life.

Ba, me, con muốn tặng bằng cao học cửa con cho ba me. Ba me cho con tất cả cơ hội để sống và học. Với bằng cao học này, con cũng muốn tưởng niệm Cậu Tý và Bà Ngoại đã che chở cho con.

ACKNOWLEGEMENTS

The successful completion of this thesis was accomplished through the contribution of several individuals and institutions. First and foremost, I am very thankful to my advisor Dr. Anna Rogiewicz, and other members of my advisory committee, Dr. Veronique J. Barthet and Dr. Bogdan A. Slominski, for their guidance and support throughout my M.Sc. program.

I am very grateful to the funding organizations, the University of Manitoba, Saskatchewan Canola Development Commission, SaskCanola, Canola Council of Canada, and CBS Bio Platforms, for their financial support and resources towards my program and research.

I want to recognize the tremendous support I received from technical and research personnel, graduate students and administrative staff at both the Department of Animal Science and Canadian Grain Commission. Special thanks to Nicole Pogorzelec for her guidance in the NIRS technology and unconditional support, and to Tiffany Chin for her PCA analysis which brought this project to completion.

During my program, I have had the pleasure of working alongside Dr. Rogiewicz's laboratory members, Dr. Yanxing (Stella) Niu, Dr. Joao Ferreira, Lan Shi, Oluwakemi Omotosho, Tetiana Sessingnong, and Agata Marcinow, whom I am thankful for their support. In addition, I received tremendous support from technical personnel at the Department of Animal Science, including Diana Garcia Posada, Atanas Karamanov, and Gurman Grewal, and post-doctoral fellow Debora M. Holanda.

I would also like to acknowledge the support of my fellow colleagues and professors, starting with Fernando Esposito who introduced me to the program and initiated my entire animal science journey. When I fell on tough times in data analysis, I was supported by Dr. Marcos Cordeiro, who

not only gave helpful suggestions but pushed me to persevere. Outside of the Department of Animal Science, I would like to sincerely thank Dr. Jamie Galka (University of Winnipeg) and Dr. Gregg Tomy (University of Manitoba) for their guidance, mentorship, and encouragement throughout the difficult periods of bachelor's degree.

To my fellow classmates, colleagues, and members of the graduate student association, it was a pleasure to have gotten to know you outside the classroom and share this journey with you, even though portions of our programs was made virtual due to the pandemic.

I'd like to recognize the importance of spending time outside the classroom as well, to which I am grateful for my supportive teammates with Puckin' Clowns, Blades of Steel, and the Grant Park High School alumni hockey teams, CanCentral Sportscards, the Grzenda family, and the select friends who have supported my academic journey since we met as naive middle schoolers.

Lastly but most important, to my parents, I want to emphasize that this is not my degree, it's our degree (and unlike your unfortunate circumstance Dad, this one will not be taken away from us).

FOREWARD

Some sections of this thesis were presented as a poster presentation at the Animal Nutrition Conference of Canada in Saskatoon, May 10-12, 2022, the Animal Nutrition Conference of Canada in Montreal, May 9-11, 2023, where it won third place in the graduate student poster competition, and at the 16th International Rapeseed Congress in Sydney, Australia, September 24-27, 2023.

The abstract titles for the presentations were:

Che, A.Q., Rogiewicz, A., Barthet, V.J. Pogorzelec, N., and Slominski, B. A. 2022. Application of near-infrared spectroscopy for rapid analysis of glucosinolate content in canola meal. Proceedings of the 2022 Animal Nutrition of Canada, Saskatoon, May 2022

Che, A.Q., Pogorzelec, N., Rogiewicz, A., Sessingnong, T., Patterson, R., Barthet, V.J., and Slominski, B.A. 2023. Advancements in NIR calibration for estimation of nutritive value of canola meal and expeller/cold pressed canola using glucosinolates content. Proceedings of the 2023 Animal Nutrition Conference of Canada, Montreal, May 2023

Barthet, V.J., Che, A.Q., Pogorzelec, N., Rogiewicz, A. and Slominski, B.A. 2023. Development of NIR calibration for estimating nutritive value of canola meal and cold pressed canola through glucosinolate content analysis. 16th International Rapeseed Congress in Sydney, Australia, September 24-27, 2023.

This thesis was written in manuscript format, and it consists of one manuscript. The manuscript will be submitted for publication in the Journal of Oil Chemist Society:

Che, A.Q., Rogiewicz, A., Barthet, V.J., Pogorzelec, N., and Slominski, B. A. Application of near-infrared spectroscopy for rapid analysis of glucosinolate content in canola meal.

CONTRIBUTION OF AUTHORS

Che A. Q. is the first and presenting author of the manuscript. He was instrumental in sample inventory management, study design, analytical procedures, NIR scanning of canola meal samples, statistical analyses, and interpretation, and he wrote the manuscript. **Rogiewicz, A.** planned the experiments, trained Che A.Q. in analytical procedures, and supervised his research activities. **Rogiewicz, A.** is the corresponding author, contributing to reviewing, editing, and proofreading of manuscript. **Barthet, V. J.** with her advanced expertise in NIRS was instrumental in conducting the NIRS procedures and she trained Che. A.Q. in NIRS technology. She contributed substantially to the design of the analytical work, data interpretation and proofreading of the thesis. **Slominski, B. A.** made a substantial contribution in the conception and design of the research, interpretation of data and editing of manuscript. He acquired funding for this research. **Pogorzelec, N.** provided expertise in NIRS technology, technical support and was instrumental in designing research. **Sessingnong T.** contributed to determination of chemical composition of selected samples of canola meal. **Patterson R.** supported research activities financially and provided his expertise in the subject matter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA .....amino acids

ANF.....anti-nutritive factor

ANN.....artificial neutral networks

CM .....canola meal

CP.....crude protein

DM .....dry matter

DDGS.....distiller's dried grains with solubles

EPC .....expeller-pressed canola

ECPC.....expeller-cold pressed canola

GLS .....glucosinolates

IR.....infrared

Lys.....lysine

MSC .....multiplicative scatter correlation

MIR .....mid-infrared

MMT .....million metric tons

MR .....Maillard reaction

MRP .....Maillard reaction products

ND.....neutral detergent

NDF.....neutral detergent fibre

NIR.....near-infrared

NIRS .....near-infrared spectroscopy

NDICP.....neutral detergent insoluble crude protein

NDF.....neutral detergent fibre

NSP .....non-starch polysaccharides

PC.....principal components

PCA.....principal component analysis

PLS.....partial least square

RMSEC .....root means square of calibration

RMSEP .....root means square of prediction

RPD.....ratio of performance to deviation

RSM .....rapeseed meal

SNV.....standard normal variate

TDF .....total dietary fibre

CHAPTER 1: GENERAL INTRODUCTION

1.1. CANOLA INDUSTRY IN CANADA

The booming canola industry in Canada has fueled the Canola Council of Canada to implement a strategy to meet the global market demand of 26 million metric tonnes of canola seed by 2025 (Canola Council of Canada, 2019b). Canola represents the quality of rapeseed bred for its low content of glucosinolates (GLS) and erucic acid (Feng & Zuo, 2015). Annual canola production in Canada yields three million metric tons (MMT) of oil and five MMT of canola meal (CM) (Canola Council of Canada, 2022). Canola meal is a co-product of the oil industry and is utilized as a protein supplement in the diets of livestock animals worldwide (Rad-Spice et al., 2018). Over 95% of canola meal produced in Canada is derived from *B. napus* seed, and the remaining fraction consists of *Brassica juncea* and *Brassica rapa* (Canola Council of Canada, 2019a). However, the inclusion of canola meal in livestock diets has traditionally been restricted due to its high dietary fibre content, the potential toxicity of glucosinolates, and the variability in protein bioavailability caused by heat treatment during oil extraction. Thus, the nutritive value of canola meal must be determined before canola meal can be integrated into the livestock animal production industry.

Since the national average of glucosinolates in Canadian canola seeds (i.e., less than 10 μmol g^{-1}) is well below the minimum levels required to cause glucosinolates toxicity in animals, current glucosinolate contents in canola meal are of reduced concern in canola meal (Canadian Grain Commission, 2022). In lieu of this, other parameters have risen as the most important limiting factor such as available energy, protein quality and amino acid digestibility (Bell, 1993).

1.2. THE IMPACT OF CONVENTIONAL OIL EXTRACTION METHODS ON THE NUTRITIVE QUALITY OF CANOLA MEAL

Methods of canola oil extraction apply varying degrees of heat and moisture to canola seeds, negatively altering the chemical composition of the residual meal (Canola Council of Canada, 2019a). In specific, pre-press solvent extraction (Figure 1.) – the most common method of oil extraction in Canada and subjects the canola seeds to up to 120°C of heat during toasting and desolventizing. Expeller (or doubled pressed) and expeller-cold pressed canola (ECPC) are two alternatives in oil extraction, which applies little or no heat.

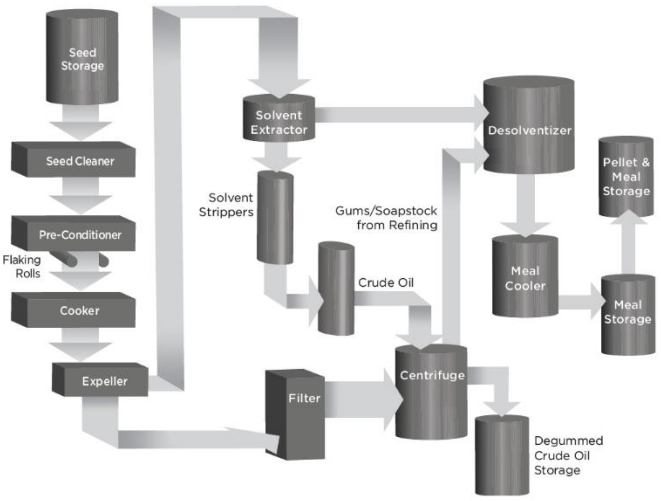

Fig 1. Schematic diagram of prepress solvent extraction of canola oil (Canola Council of Canada, 2019a)

The heat treatment triggers the Maillard reaction (MR) which occurs between amino acids (AA) and reducing sugars and generates Maillard reaction products (MRPs) often referred to as neutral detergent insoluble crude protein (NDICP) or glycoprotein. Variation in the extent of heat and moisture application during canola seed processing will affect the quantities of products of the

Maillard reactions. This non-enzymatic browning reaction produces insoluble nitrogenous polymers from proteins and products of degraded carbohydrates (Karbasi & Madadlou, 2018). Neutral detergent insoluble crude protein is the MRP of AA and reducing sugars, responsible for inflating total dietary fibre content in canola meal. Lysine is the most prone AA to undergo the MR and its bioavailability to animals is greatly reduced in canola meal which has undergone treatment with high heat, resulting in a meal with damaged protein and thus reduced nutritive value. Lysine, within the NDICP, produces reactive or unreactive lysine. Unreactive lysine is the specific fraction that bonds to reducing sugars and is biologically unavailable to the animal. Both Anderson-Hafermann et al., (1993) and Newkirk et al., (2003) have reported decreases in apparent lysine digestibility (85 to 80%) and apparent lysine digestibility coefficient (0.87 to 0.79), respectively, in canola that has undergone desolventizing and toasting during oil extraction. Thus, NDICP is an indicator of protein damage in canola meal. Adewole et al. (2016) calculated a strong correlation ($R^2 = 0.78$) between NDICP and total dietary fibre (TDF) in canola meal. The same authors also calculated a similar correlation ($R^2 = 0.63$) between lysine and TDF in canola meal. While it has been demonstrated that heat treatment in oil extraction results in negative consequences in canola meal, its application has also yielded antitoxic advantages. The levels of glucosinolates in canola meal and their subsequent threats as antinutritive factors to livestock animals have been reduced significantly as a result of the heat and moisture lability of glucosinolates and their associated enzymes in the desolventizer/toaster, on top of the successful conventional breeding that led to the development of canola (Canola Council of Canada, 2019a). Specifically, thioglucosinase (3.2.2.1; myrosinase) is responsible for hydrolyzing glucosinolates into its toxic intermediates (isothiocyanates, oxazolidine-2-thiones, nitriles, thiocyanates) (Brown, 2018). Myrosinase starts

to denature at 65°C, however, it can withstand temperatures of up to 90°C, while glucosinolates have shown complete degradation at 100°C (Aripin & Surugau, 2016; Oerlemans et al., 2006).

Assessing feed ingredient for its nutritive value is a crucial step in animal nutrition to ensure optimal delivery of nutrients to livestock animals. However, there is lack of a direct approach for evaluating fibre-bound lysine and NDICP in canola meal. A series of intense wet chemistry and *in vivo* procedures are required to evaluate protein damages (Noel et al. 2022). In particular, NDICP determination would involve lengthy procedures for NDF and nitrogen analyses. One indirect method indicating protein damages is to assess glucosinolate contents in canola meal. Pahm et al., (2008) and Kim et al. (2012) have utilized the furosine method to analyze unreactive lysine in dried distillers grains with solubles (DDGS), however, this technique, still labour-intensive, is relevant rather to quantify the unreactive lysine content than to provide the nutritive value of protein in the feed ingredient.

Glucosinolates are naturally occurring secondary metabolites in the Brassicaceae family. Like other chemical components in canola meal which show variability based upon the processing condition, glucosinolate contents also exhibit inconsistencies based on heat and moisture treatments. Bell and Keith (1991) demonstrated a 15-77% reduction of glucosinolate content in canola meal, most of the reduction was due to the treatment during desolventizing/toasting. Despite its degradation at high temperatures, glucosinolates are robust, thermostable below room temperature, and can remain intact for years if properly stored.

1.3. THEORY OF NEAR-INFRARED SPECTROSCOPY

Near-infrared spectroscopy (NIRS) is a predictive spectroscopic technique in agriculture science utilized for its numerous benefits such as rapidity, efficiency, reliability, cost-effectiveness, user- and environmental-friendliness (Noel et al., 2022). The NIR absorption region is located between the visible light and mid-infrared (MIR) regions and corresponds to wavelengths between 12500 cm^{-1} (780 nm) and 4000 cm^{-1} (2500 nm). The basic principle of NIRS is that absorption spectra of electromagnetic (light) energy (or radiation) by covalent bonds from molecules within the functional groups (C-H, N-H, O-H, S-H) can extrapolate quantitative information about the target analytes in the sample (Siesler et al., 2001; Noel et al. 2022). The interaction of light with molecules can be described by the harmonic and anharmonic oscillator models provided that the molecules absorb the energy of light (NIR-active) (Workman Jr. & Weyer, 2012). The harmonic oscillator depicts a mass-spring system which two molecules of different masses are connected by a bond (spring) and can be displaced from an equilibrium position (portraying the natural vibrations between the two molecules). Both the natural vibrations from chemical bonds and photons from NIR radiation exist at specific frequencies (vibration per unit time; denoted by the Greek letter ν (nu)). Energy quantized within in a photon (a "packet" of light energy) is proportional to ν in equation 1. Frequency is also equal to the speed of light (3.0 x 10^8 cm sec^{-1}) over wavelength (length of one complete wave cycle; denoted by λ) of the radiation. Variable h is Planck's constant (6.626 x 10^{-34} J sec^{-1}).

$$E = h\nu = h\left(\frac{c}{\lambda}\right) \qquad \text{(Eq 1.1.)}$$

Molecular absorptions of NIR energy occurs if the frequencies of the molecular vibration and the radiation match, resulting in an excitation in energy levels within the molecular system.

Such changes in energy follow the rules of quantum mechanics, in which each level is quantized by a vibrational quantum number (denoted by υ) and can only take integer values (Workman Jr. & Weyer, 2012). The absorption of light by molecules can be described with NIR bands that appear on a NIR spectra. The interaction of light with functional groups ultimately gives rise to fundamental, overtone and combination bands – characteristic absorbances on an NIR spectrum. An excitation from the ground state to the first energy level gives rise to a fundamental band (υ = 0 → υ = 1). The first highest-order energy transition resulting from molecular absorption corresponds to the first overtone band (υ = 0 → υ = 2). Subsequent orders of transition correspond to second, third and fourth overtone bands (υ = 0 → υ = 3,4,5). Two or more transitions can also occur at similar frequencies, giving rise to combination bands. Table 1. Summarizes the functional groups capable of generating overtone and combination bands.

1.3.1. The harmonic oscillator model

The harmonic oscillator illustrates the energy potential of two masses on a spring as a symmetrical parabolic curve (dotted plot; Figure 2). As the bond displacement (dipole) increases, the total energy of the atom approaches a maximum. Energy is defined in equation 1.2., where k is the force constant, and x is the bond's displacement from equilibrium.

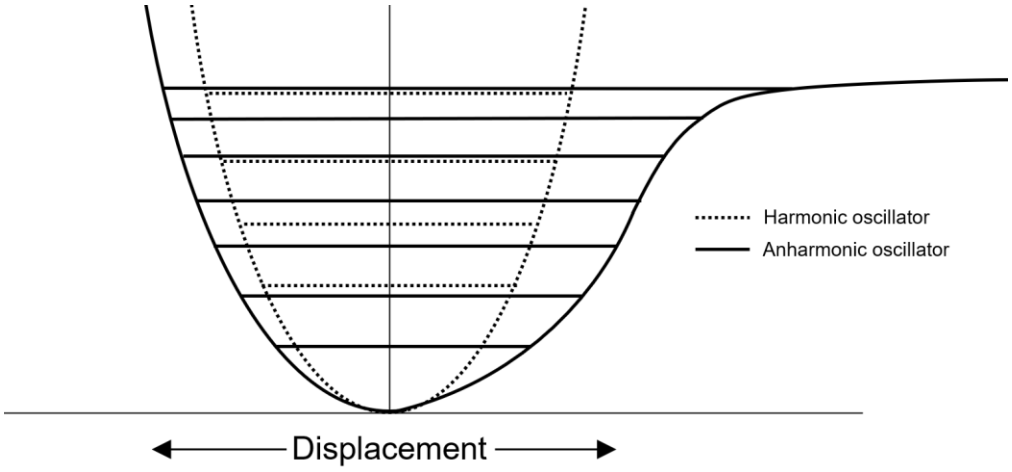

Fig 2. Potential energy curves for harmonic and anharmonic oscillator models (Singh, 2016)

$$E = \frac{1}{2}kx^2 \quad \text{(Eq 1.2.)}$$

$$E = n + \frac{1}{2}hv \quad \text{(Eq 1.3.)}$$

The potential energy curve consists of quantized energy levels that is defined by equation 1.3., where *n* is an integer quantum number representing the oscillating dipole for potential energy levels. The harmonic model is used to predict energy or frequency of fundamental absorption bands associated with asymmetric or symmetric vibrations (stretching, scissoring, bending, wagging, etc.), as each vibration is associated with a fundamental vibration frequency (Workman Jr. & Weyer, 2012). However, the model is too simplistic for predicting band positions specific to overtone vibrations of higher energy when energy absorption takes place, especially since the "masses-and-spring" model does not fully explain the interactions of molecules (Workman Jr. & Weyer, 2012).

1.3.2. The anharmonic oscillator model

It has been well established in spectroscopic theory that the anharmonic model is a better representation of the natural behavior of molecules (Workman Jr. & Weyer, 2012). Specifically, the anharmonic model considers (i) first, second overtone and third (higher-level) vibrations exhibited by molecular absorption of energy (ii) the repulsive forces that occurs when molecules are too close, and (iii) the Morse potential curve. Overtone vibrations take place when molecular energy absorptions occur, causing a transition from the fundamental frequency to the second (first overtone; twice the fundamental frequency) or third frequency (second overtone; three times the fundamental frequency). Compared to its harmonic counterpart, the anharmonic model can deliver more realistic calculations of the common overtone vibrations that exist in NIR spectroscopy. Graphically, the anharmonic model illustrates the Morse curve with its characteristic plateau in energy resulting from bond dissociation due to excessive distance between atoms (solid line in Figure 2.). The quantized energy levels in the anharmonic model also differ from that of the harmonic model, as indicated by the spacing between levels (Figure 2.).

Table 1. List of common functional groups associated with nutrients detected by near-infrared spectroscopy and their major analytical bands

Functional Group	Band	Nutrient/component
C-H	Third overtone stretch	Fat
	First overtone stretch	Glucosinolates
O-H	Second overtone stretch	Moisture
	First overtone stretch	Glucosinolates
N-H	Second and third overtone stretches	Protein
	First overtone stretches	Glucosinolates
S-H	First overtone stretch	Glucosinolates

1.3.3. Franck-Condon Principle

In spectroscopy, the Franck-Condon Principle provides the reasoning for prominence of peaks in a spectrum (i.e., strong, weak, and absent peaks). It states that the change in electronic transitions due to radiation is much faster than that of its atomic nucleus, to the point where nuclear motion essentially remains unchanged. The principle is applied to NIR spectroscopy when radiation triggers a molecule's excitation from a ground to an excited state. The initial and final vibration states that exhibit the most overlap will capture the most intense absorption on the NIR spectra, according to the Frank-Condon principle. Such intense absorptions result in reduced energy transmittance to the detector, generating weak or absent peaks.

1.3.4. Chemometrics

Near-infrared spectroscopy operation is made possible by the use of chemometrics – a chemical discipline that utilizes mathematical, statistical and other methods employing formal logic to design or select optimal measurements, procedure and experiments, and to provide

maximum relevant chemical information by analyzing chemical data (Heberger, 2008). With lack of chemical specificity that often features overlapping bands and ambiguous peaks, NIR spectra are difficult to interpret directly, unlike its related counterparts IR and mid-IR spectroscopy. Thus, common chemometric techniques are associated with NIRS, including principal component analysis (PCA), partial least squares (PLS) regression, multivariate linear regression (MLR) and artificial neutral networks (ANN) (Daszykowski et al., 2008). Although major differences exist in each method in producing a response from predictor variables, the end goal for each method is to generate a viable prediction model for prediction of an analyte of interest utilizing the spectral database. By far, PLS regression is the most dominant method and considered the standard in NIR calibration due to its simplicity and efficiency (Jiao et al., 2020). It is based on modelling of a linear relationship between a dataset of reference (or explanatory) variables (X) and variables of interest or response (y). The PLS algorithm will generate latent variables (LV; denoted by T) that best describe y based on X. Latent variables will maximize the covariance between X (wavelength regions acquired via spectral data) and y (reference values acquired via wet chemistry). High collinearity exists within wavelengths of the dataset and is treated by compressing all relevant wavelengths of the NIR spectra into LVs. Latent variables are continuously determined until all the covariance that exists between X and y can be explained with each passing LV uncorrelated (orthogonal) with the previous LV. In prediction models, the smaller number of LVs is required (Killner et al., 2011) to obtain a robust model. Once all LVs are determined, the X and y variables are regressed linearly, model is considered complete and statistical results for calibration are generated (i.e., R^2, RMSEC, etc.). At this point, the model is ready to be validated.

It is important to recognize that chemometric and spectral preprocessing methods (also called data pretreatment) are pivotal aspects of NIR model development. Preprocessing can

attenuate the technical drawbacks of NIR spectroscopy that generate undesired signals in the spectrum. Overlapping absorption bands and peaks in the NIR spectra can complicate the interpretation of the spectral information, especially within the important peaks (Daszykowski et al., 2008). Savitzky-Golay (SG) first and second derivatives can be applied to reduce peak overlaps. Spectral acquisition can be obstructed by undesired light scattering effects and slope shifting (Daszykowski et al., 2008). Such scattering and shifting effects can be eliminated by multiplicative scatter correlation (MSC) (Jiao et al., 2020). Differences in particle sizes can also result in undesired effects in light scattering and in the resulting spectral data. Such effects can be eliminated with standard normal variate (SNV). Over 100 preprocessing methods exist to treat the raw scans and develop NIR prediction models (Jiao et al., 2020). It is the researcher's best interest to determine which methods and/or combinations best fits their spectral data to develop a prediction model utilizing the most optimal spectral data. Surprisingly, Lee et al., (2017) reports that most NIR prediction models are developed without the application of preprocessing methods. Preprocessing raw spectra is not a prerequisite to calibration development, and thus, neglecting preprocessing can save considerable time and labour. However, sometimes it is at the expense of model robustness as the removal of undesired peaks and the enhancement of spectra can contribute to the accuracy and the precision of the prediction model.

1.3.5. Instrument calibration and validation

As a predictive technique, near-infrared instruments require calibration to predict the quantity of an unknown analyte(s). A calibration of the instrument is the application of a developed prediction model(s) that correlates the spectral data and wet chemistry value of the analyte(s) to be measured (called a reference value). The dataset of a prediction model comprises of many samples where each sample provides a spectrum and reference value(s) the user is calibrating for.

Thus, the quality of a prediction model is determined by the distribution of spectral, chemical, and physical characteristics of samples in the prediction dataset. Specifically, such characteristics need to be of a large range and even distribution to cover the quality of the unknown analyte. The prediction model generates a prediction equation, determined from the correlation of the reference values and instrument-predicted values for each sample in the model.

Evaluating the model includes cross-validation and validation with select statistical parameters as a measure of robustness and quality (see section 4.8. Criteria for evaluation). Cross-validation and validation are performed at different stages of model development with different datasets. A cross-validation is carried out following the development of a prediction model for quality control by removing a select number of samples from the prediction dataset and redeveloping the prediction model. The samples that are removed serve as the validation dataset. Cross-validation generates a prediction equation determined by the correlation of the reference values and instrument-predicted values of the validation dataset. The coefficient of determination (R^2) of the cross validation should generally be high (>0.90) which indicates the model is generating precise predictions.

Validation is the most comprehensive evaluation of a prediction model and is carried out using a set of samples that is completely independent of the set used to develop the prediction model. The validation dataset generates a prediction equation determined by the correlation of the reference values and instrument-predicted values of the validation dataset. The statistics generated from validation indicate the prediction model's robustness, precision, accuracy, and therefore, usability.

1.4. THE AIM OF THE STUDY

Assessing the quality parameters in oilseeds is a crucial step for both the oilseed and animal production industries. In seeds, these parameters typically include oil, protein, glucosinolates, chlorophyll, fatty acids, and total saturated fat (Barthet et al., 2020; Ali Redha et al., 2023). However, canola meal quality assessment using NIR technology is typically limited to dry matter, oil, and protein contents. This research incorporates a correlation between the chemical components of canola meal and its nutritive value to develop advanced NIR prediction models that can accurately estimate the quality traits of canola meal. Considering the importance of predicting protein damage during canola seed processing, the project aimed to develop robust calibrations for glucosinolates content, as well as fat and protein levels for additional quality evaluation in canola meal. Development of a successful prediction models for nutrient analysis of canola meal will play a key role in enhancing the sustainability of both the canola and animal production industries in Canada. Accurate determination of nutritional components in canola meal would ensure the efficient utilization of canola meal in animal nutrition.

CHAPTER 2: LITERATURE REVIEW

2.1 INTRODUCTION

Canola is one of Canada's most valuable oilseeds annually contributing nearly $30 billion to the Canadian economy (LMC International, 2020). Canola meal, the co-product generated from oil extraction, is a valuable protein source for livestock animals and can be complementary to soybean meal in monogastric animal nutrition (Khajali & Slominski, 2012). However, canola meal quality must be assessed before entering the feed industry as a result of the variation of protein damage than can occur during the canola oil extraction process due to Maillard reactions (Adewole et al. 2016). In brief, neutral detergent insoluble crude protein (NDICP, or glycoproteins) are fibre components generated from amino acids bonding to reducing sugars in the presence of excessive heat and moisture. Lysine is especially prone to undergoing the Maillard reaction. Consequently, the NDICP fraction contributes to the inflation of dietary fibre in canola meal, while the bioavailability of crude protein to the animal is reduced and classified as a form of protein damage. The variations in heat and moisture are amplified by the differences in each extraction technique (pre-press solvent-extract, expeller/double-pressed, expeller cold-pressed) used by canola crushing processors and lead to variations in protein damages. Ideally, measuring NDICP in the neutral detergent fibre (NDF) fraction would directly indicate the degree of protein damage in canola meal, and thus, overall canola meal quality. It could be an alternative to amino acid digestibility studies using poultry and swine. This *in vivo* direct measurement of protein damage requires advanced research expertise and is known for being costly and lengthy. However, direct assessment of NDICP would still involve the use of wet chemistry procedures presenting constraints in time, cost, and labour.

Given the need for a reliable and rapid method to assess protein quality at the feed mill level, utilizing the content of other components affected by heat and moisture could potentially serve the purpose. One such option is to utilize total glucosinolates content as an indirect indicator of protein damage and, consequently canola meal quality, through near-infrared (NIR) spectroscopy. It is essential to emphasize that the glucosinolate contents itself serves as a desirable measurement of canola meal quality, and developing a NIR calibration for glucosinolates would hold significant importance for both the canola and feed industries.

Glucosinolates are naturally occurring secondary metabolites in the Brassica genus of plants (broccoli, canola, rapeseed (RS), cauliflower, mustard, etc.). Although non-toxic by itself, glucosinolates were historically considered a major anti-nutritive factor (ANF) in RS crops as their breakdown products triggered thyroid-related toxicity in livestock animals (see 2.4.2.2). A secondary concern is their unfavorable palatability for livestock species, although this is more pronounced for swine and ruminants as their taste and smell senses are more developed compared to poultry (Toghyani et al., 2009). Canadian breeding programs targeting desired properties in RS (higher oil and protein, limited glucosinolates and erucic acid contents) in the 1950s gave rise to improved *Brassica rapa* and *Brassica juncea* varieties (Slominski, 1997; Canola Council of Canada, 2023). These varieties (commonly termed canola for Canada, oil and low erucic acid) are defined by their contents of limited erucic acid (< 2% in the oil) and total glucosinolates (< 30 μmol g^{-1} in the residual meal) (Canola Council of Canada, 2019a). Total glucosinolate contents are comprised of individual compounds based on their chemical structure, including aromatic, aliphatic (non-aromatic) and indole glucosinolates (Wu et al., 2021). Since glucosinolates are heat and moisture labile (breakdown can start to occur at approximately 60°C), their presence in canola

meal can serve as an indirect indicator of protein damage as a result of the heat and moisture conditions during the oil processing.

In agricultural science, NIR spectroscopy (NIRS) is used commonly to predict quality parameters such as fat, protein, and moisture content in most feed ingredients. NIRS is a time-efficient, cost-effective, non-destructive, environmentally friendly, and user-friendly technology capable of performing diffuse reflectance or transmittance spectroscopy (Noel et al., 2022). Its basic principle involves extracting spectral information based on the characteristic absorbance and reflection of light at specific wavelengths by chemical functional groups within a sample. Each sample possesses a unique "fingerprint" with NIR spectral patterns that corresponds to all its various functional groups that allow to identify and quantify the nutrients that are present. The process of transforming spectral data into tangible information, i.e., numerical values, involves the use of chemometrics – the chemical discipline that mathematically analyzes chemical data to achieve optimal measures and outputs maximum chemical information (Varmuza & Filzmoser, 2009). NIR calibrations and routine operations are completely reliant on chemometrics to generate a prediction of the various analyte contents in a sample. A robust prediction model is a fundamental precursor for accurate and precise NIRS analysis and is often recognized as the most painstaking stage of the NIRS operation. A prediction model is simply a mathematical correlation of the spectral data and the true value(s) of the target analyte(s) acquired from wet chemistry. The number of samples to construct a reliable prediction model is often debated, however, a bare minimum of 30 samples is required (Lumex Instruments, 2018). The prediction model functions as a reference dataset when analyzing an unknown sample; a value of the analyte is generated by the instrument based on comparison of the unknown spectra with that of the collection of samples of the same type in the prediction model. While NIR has greatly contributed to advancements in quality control

and assessment of products in the Canadian agriculture industry, more precise assessment is required for Canada to reach the demand of canola meal for food and animal products on a global scale.

2.2. NEAR-INFRARED SPECTROSCOPY TECHNOLOGY

2.2.1. History of near-infrared spectroscopy

The discovery of infrared (IR) light and its neighbouring regions (mid-infrared and near-infrared) in the electromagnetic (EM) spectrum is credited to German-English physicist Frederik William Herschel. His 1800 publication "Experiments on the refrangibility of the invisible rays of the sun" outlined his discovery of invisible "heating rays" beyond the red terminus of the visible light region (Hershel, 1800; Ring, 2000). Crucially, Herschel had detected that these heating rays had emitted the highest temperature out of all the light he analyzed, introducing the scientific community to the possibility of another form of light. In 1835, German scientist Johan Ritter concluded this heating ray possessed a larger wavelength compared to visible light and termed it "infrared light" (Latin for "below red") (Ozaki, 2021). The late 19th century featured noteworthy advancements in spectroscopy, including the first NIR spectra produced by English chemists William Abney and Edward R. Festing, as well as the first NIR spectrometer constructed by William W. Coblentz (Burns & Ciurczak, 2008). Coblentz also innovated the NIRS operation with his interpretation of spectra that remained relevant in contemporary spectra management; he had discovered that compounds produced different spectra, even if they were isomers and that absorption patterns in the specific regions correlated with types of compounds (phenols, aldehydes, etc.). American physicist Karl H. Norris revolutionized NIR technology with his invention of the modern instrument and the introduction of chemometrics during the 1960s, giving rise to his

nickname, “the father of near-infrared spectroscopy” (Roberts et al., 2004, Davies, 2011). Many of his publications remain an industry standard in modern NIRS operation. Wheat analysis in the 1980s by Canadian agricultural scientist Phil Williams marked the first application of NIRS in the commercial industry (Roberts et al., 2004). One of the highlights of Williams’ distinguished career was the integration of NIRS into a survey of grain elevators in 1975 where NIRS technology replaced traditional Kjeldahl method to analyze over two million samples of wheat in a four-year study (Roberts et al., 2004). The success of the collaborative effort between the Neotec Corporation and Canadian Grain Commission resulted in full NIRS utilization in Canadian grain elevators in 1976 (Williams et al., 2019). In the last five decades, there has been extensive growth of NIRS utilization as a reliable predictor of quality in agriculture products and intermediates. Its’ application has branched into fresh cereals and forages, monitoring silage for fermentation, analyzing animal products (meat, milk, eggs, and dairy products), identifying presence of fraudulent materials in feed or feed origins, and analyzing manure to reduce environmental impact (Berzaghi & Riovanto, 2009).

2.2.2. Principles and concepts of near-infrared spectroscopy

Near-infrared spectroscopy is a diffuse reflection spectroscopic technique which involves the absorption of radiation by chemicals possessing C-H, O-H, N-H and S-H covalent bonds (Williams et al, 2019; Noel et al, 2022). The near-infrared (NIR) region of the EM spectrum occurs between 700-2500 nm wavelength region (12500-4000 cm^{-1} in wavenumbers). Absorption of NIR radiation by chemical bonds induces vibrational and electronic excitations which appear as bands (combinations or first, second, and third overtones) on NIR spectra depending on the type of bond (Gunzler & Williams, 2001; Ozaki, 2021). NIR spectroscopy is both an electronic and a vibrational spectroscopic technique. The versatility of NIRS has popularized its usage to investigate many

parameters among many forms of samples (ground samples, intact seed, hulls and whole kernels, liquids) in agriculture science.

A combination of physics, mathematics, and chemistry are required to achieve a comprehensive understanding of NIRS. There are many laws, theories and models that govern NIRS and give possibility to its operation and function. While it is not necessary to grasp this level of insight, a basic introduction to principles and concepts will allow for smoother operation of the instrument (Williams et al., 2019).

Successful spectroscopic interaction depends on absorption of energy of radiation by the molecule. Two conditions must be satisfied for absorption occur (Smith, 1999). Radiation consists of an electric vector component that oscillates between positive and negative charges as it travels in a wave. When interacting with a diatomic molecule, the vector attracts or repels the atoms based on their charges, inducing changes to the bond length, and triggering molecular vibration. If frequency of the molecular vibration and the electric vector are equal, the first condition is satisfied, and energy is transferred from the radiation to the molecule. In addition to its wave-like behavior, radiation also exhibits characteristics of a particle by possessing photons (packets of energy). According to the laws of quantum mechanics, molecules within a bound system (i.e., functional groups) have quantized vibrational energies, represented by vibrational quantum number, ν. Excitation from the ground state ($\nu = 0$) to higher energy levels ($\nu = 1, 2, 3$, etc.) requires a specific amount of energy – provided that the photon carries sufficient energy for molecular excitation, it is absorbed by the molecule and satisfies the second requirement for absorption of radiation. Otherwise, insufficient energy results in the molecule transmitting the energy.

2.3. ADVANTAGES AND DISADVANTAGES OF NEAR-INFRARED SPECTROSCOPY FOR EVAULATION OF NUTRIENTS IN FEED INGREDIENTS

The most significant benefits of NIRS include efficiency, sustainability, and straightforward operations. Analysis of feed ingredients with wet chemistry may involve time-consuming sample preparation (i.e., grinding, defatting, or weighing), destruction of sample, use of harsh chemicals and long durations of incubation or intensive laboratory operation which can be eliminated through NIRS analysis. For example, Kjeldahl analysis of crude protein involves a laborious three-step conversion of nitrogen to ammonia with acids at high temperatures. Furthermore, a titration with more acid is required to determine previously undigested ammonia in the sample (Sáez-Plaza et al., 2013). This entire procedure may last up to two hours. By comparison, a NIR spectrometer can provide results in under one minute. In addition, penetrative ability of NIR radiation offers flexibility for the operator to run many forms of sample (particles, liquids, whole seeds, and kernels, etc.). Benchtop NIR spectrometers are operated with user-friendly software, eliminating the need for following an extensive procedure and manual data recording. This procedure not only reduces the analysis time per sample but also eliminates the need for disposal of harsh chemicals, making it safe for both the user and the environment. In terms of analyzing feed ingredients, the greatest advantage NIRS can offer is its multi-trait ability (Velasco & Becker, 1998). Several parameters or chemical components of interest can be simultaneously predicted with NIRS, reducing time and effort required to run several different wet chemistry procedures for each individual parameter.

Although NIRS provides many advantages, it is essential to acknowledge that it also presents certain drawbacks concerning its results and calibration, which may call its reliability and efficiency into question. The main drawback is that NIRS is a predictive technique and cannot

entirely replace the physical extraction/quantification of analytes by wet chemistry, which might offer higher accuracy (Hall, 2014). Another drawback is that constructing a reliable prediction model for optimal NIRS operation requires significant commitments and efforts. Although wet chemistry is not utilized during the analysis of samples in NIRS, it remains a need to establish the correlations between reference values and spectral data in the prediction model. Therefore, the accuracy and precision of the methods becomes a focal point when acquiring the wet chemistry reference values. Since the NIR predictions of sample chemical components rely on a calibration that correlates spectral and reference values, the wet chemistry methods employed to acquire the reference values must be of very minimal error.

2.4. CHEMICAL COMPOSITION OF FEED INGREDIENTS

2.4.1. The importance of chemical composition

Providing optimal nutrition to livestock animals is a pillar of animal agriculture as it directly influences the health and welfare of the animal, the profitability of the producer, impact on the environment and quality delivered to the consumer (de Jonge & Jackson, 2013). The science that focuses on delivering optimal nutrient to livestock animals for growth, body maintenance, performance, and reproduction is animal nutrition (Cherian, 2020). By analyzing the chemical composition (breakdown of total chemical profile), a nutritive value can be assigned to the feed ingredient. A typical analysis of the chemical composition includes a proximate analysis and any other chemical components related to the feed ingredient (energy, vitamins, minerals, anti-nutritive factors). Animal nutritionists can then use that nutritive value to construct more accurate and precise diet formulations with the goal of meeting the nutrient demands of livestock animals (Cherian, 2020). In addition, chemical composition also plays an important role in maintaining

standards in the agriculture industry by identifying variation within the nutrient profile as many external factors (weather, growing season, soil conditions) influence the quality of nutrients within feed ingredients.

2.4.2. Chemical composition of canola meal

In general, canola meal is recognized for its high protein, low glucosinolates, and high dietary fibre content. Initially a by-product of the canola oil industry, canola meal is transformed into a co-product once integrated into the livestock production industry (Adewole et al., 2016). However, due to various processing techniques, the oil content can range from 1.6 to 2.8% DM in conventional pre-pressed solvent canola meal (Adewole et al., 2016; Radfar et al., 2017). Radfar et al., (2017) observed low range of values for CP (41.5-41.7% DM), total NSP (20.5-21.0% DM), as well as the individual NSP components (individual sugars and uronic acid). Slominski et al., (2012) observed similar ranges for CP (43.3-49.8% DM) and total dietary fibre (24.1-30.1% DM). However, total glucosinolate contents differed between the studies as well as range of values: 9.5-16.7 μmol g^{-1} DM compared to 17.1- 27.1 μmol g^{-1} DM. Likely the wide range within glucosinolate contents was due to the variation in heat and moisture applied during the oil extraction process. Other ANFs in canola meal included phytic acid, sinapine, and tannins.

2.4.2.1. Glucosinolates

Glucosinolates are a class of plant secondary metabolites produced by the *Brassica* genus and derived from many amino acids (Prieto et al., 2019; Wu et al., 2021). They are natural compounds in Brassica species and are involved in various biological functions, including the plant defence mechanism against diseases and pests. When plant tissue is damaged by insects or other pathogens, Glucosinolates are enzymatically hydrolysed into several derivatives, which express

strong antimicrobial and repellent properties. Over 200 different structures of glucosinolates have been identified and their existence is within 3000 different *Brassica species*. Three major categories of individual glucosinolates exist based on their amino acid derivatives, including aliphatic, aromatic, and indole (Wu et al., 2021) (summarized in Table 2.). The distinguishing feature between the three categories is based on the different constituents in the side chain (R) of the glucosinolates structure. Glucosinolates have a common core structure, consisting of a thiohydroximate-*O*-sulfonate group bonded to a β-D-glucose and an aglycone side chain (R) (Prieto et al., 2019) (Figure 3.).

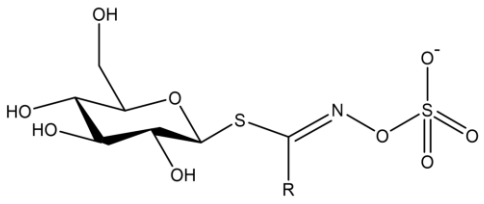

Fig 3. Core chemical structure of glucosinolates (R = aliphatic, aromatic or indole sidechain)

Table 2. Summary of aliphatic, aromatic and indole glucosinolates with examples, and amino acid derivatives (Wu et al., 2021)

Category	Examples	Derived from
Aliphatic	Sinigrin, progoitrin	Alanine, leucine, isoleucine, methionine, valine
Aromatic	Gluconasturtiin	Phenylalanine, tyrosine
Indole	Glucobrassicin, neoglucobrassicin	Tryptophan

The glucosinolates profile of canola meal consists of seven compounds including gluconapin, glucobrassicin, glucobrassicinapin, progoitrin, hydro-glucobrassicin and neoglucobrassicin. The approximate breakdown of aliphatic to indole-glucosinolates found in canola meal is 85-15%, with aromatic-glucosinolates not present in *B. napus*, *B. juncea* or *B. rapa* species. Gluconapin and progoitrin are typically the two most abundant glucosinolates in canola meal.

Glucosinolates content in Canadian canola meal is 3.6 μmol g^{-1} DM, however, has a wide range depending on the processing plant (Canola Council of Canada, 2019a). A survey of eleven plants by Adewole et al., (2016) showed a range of 2.3 to 9.7 μmol g^{-1} DM in which nine plants were under 4.0 μmol g^{-1} DM. The species of canola have also played a role in the total glucosinolate contents of the meal. Conventional, black-seeded *B. napus*, and yellow-seeded *B. napus* and *B. juncea*) have exhibited total glucosinolates of 7.9, 14.6 and 12.7 μmol g^{-1} DM, respectively.

2.4.2.2. Glucosinolates toxicity

Until the 1980s, Canola-derived glucosinolates toxicity was a major concern in animal nutrition. Although glucosinolates are not toxic by themselves, overconsumption of RS meal and canola meal resulted in growth depression, liver damage, and thyroid enlargement in livestock species. Egg production was also depressed in poultry species. The toxic metabolites (called aglucones) of glucosinolates breakdown target various organs following consumption. Isothiocyanates, thiocyanates, and oxazolidine-2-thiones negatively influence thyroid function, while nitriles induce damage to the liver, kidney, and bile ducts (Brown, 2018).

Table 3. Summary of dietary glucosinolates inclusions and effects on livestock species

Species	Dietary glucosinolates (μmol g^{-1})	Effect	Author(s)
Swine	0.16 -0.78	No adverse effects during growth, pregnancy, and lactation	Opalka et al., (2001)
	1.30-2.79	Reduced feed intake and growth	Bell et al., (1991)
	1.30	Reduced weight gain during finishing	Roth-Mailer et al., (2004)
Ruminants	1.20-1.40	No adverse effect on calf liver and thyroid function	Anderssen & Sorensen (1985)
	11.7-24.3	Decline in feed intake and milk production in dairy cows	Waldern (1973)
Poultry	2.3-11.6	No adverse effect on weight gain or intake	Marangos et al., (1974), Leeson et al., (1987)
	7.6-15.3	Severe growth depression	Thomas et al., (1984)

According to Brown (2018), and summarized in Table 3., rumen enzymes are capable of breaking down aglucones into less toxic metabolites. Thus, ruminants can generally tolerate a higher concentration of glucosinolates in their diets. However, in terms of palatability, poultry species can tolerate more glucosinolates despite being more vulnerable toxicity compared to ruminants, due their relative low taste sensitivity. Khajali & Slominski, (2012) proposed the "no-adverse effect" of canola meal-containing diet fed to laying hens where the level of glucosinolates is no higher than 1.5 μmol g^{-1} of the diet. Swine have well developed sense of taste, thus will be more prone to the adverse palatability of diet containing changed levels of canola meal. The most pronounced effect of high levels of canola meal for swine occurs in their impaired body weight, especially in the liver and thyroid (Lee et al., 2020).

2.4.2.3. Crude Protein

The Canola Council of Canada (2019) reported CP to be 42% DM in canola meal, however, it could range from 40.5 to 43.5% DM (Adewole, 2017). Slominski et al. (2012) showed that CP in meals derived from *B. napus* (black), *B. napus* (yellow) and *B. juncea* reached 43.8, 49.8, and 47.4% DM, respectively. However, these CP values may contain a highly indigestible glycoprotein fraction from the Maillard reactions during the oil extraction process (see section 2.4.2.5) (Khajali & Slominski, 2012). The total protein content in canola meal could consist of 10% glycoprotein or more. In general, an increase in glycoprotein resulted in a substantial decrease in protein digestibility which could greatly hinder the utilization of canola meal in livestock diets (Adewole, 2017).

2.4.2.4. Amino Acids (AA)

The AA profile of canola meal is an asset to its utilization and is recommended to be complimentary to SBM in livestock diet formulations (Khajali & Slominski, 2012). Compared to that of SBM, canola meal contains less lysine, arginine, and phenylalanine yet more sulfur-containing AA, i.e., methionine and cysteine. Although lysine (2.04% DM) and arginine (2.19% DM) are generally more abundant than most other AA in canola meal, there are issues associated with both. Typically, in canola meal, lysine digestibility is reduced after the desolventizing/toasting stage (Anderson-Hafermann et al., 1993; Newkirk, 2003). In addition to being the most limiting amino acid in poultry, lysine is needed for growth and maintenance (Khwatenge et al., 2020). Lysine supports development of the breast muscle fibres and feathers, as well as the synthesis of compounds involved in metabolism (Franco Nogueira et al., 2022). Arginine has demonstrated reductions during oil extraction, however, not to the same extent as that of lysine (Salazar-Villanea

et al., 2022). The performance of chickens suffered with low arginine intake, exacerbated by the inability for chickens to sufficiently synthesize endogenous arginine and the unavailability of synthetic arginine (Khajali & Slominski, 2012). Like glucosinolates and protein, the AA profile of canola meal also varies depending on the Canola species. Yellow-seeded *B. napus* contained the highest individual AA content when compared to *B. napus* (black) and *B. juncea* (Slominski et al., 2012, Trindade Neto et al., 2012).

2.4.2.5. Protein damage in canola meal and the subsequent consequences

The Maillard reactions (MR) are spontaneous non-enzymatic chemical reactions resulting from the linkage of the amino group of amino acids and a carbonyl-containing moiety (typically, a reducing sugar) (Aljahdali & Carbonero, 2019). Maillard reaction products (MRPs) come in multiple glycoforms depending on physical factors such as the degree of heat, moisture, and pH exerted on the reactants (Karbasi & Madadlou, 2018; Ruan et al, 2018; Aljahali & Carbonero, 2019). At temperatures as low as 50°C, amino acids (lysine, arginine, histidine) have been demonstrated react to glucose (Ruan et al., 2018). The conventional pre-press solvent extraction of canola oil involves cooking and desolventizing stages, optimal conditions to trigger the MR, where seeds and residual meal are heated between 80-105°C while maintaining 6% in moisture (Canola Council of Canada, 2019a). Maillard reactions can lead to the formation of compounds, which can be measured as a neutral detergent insoluble crude protein (NDICP) – protein within the neutral detergent fibre fraction of total dietary fibre (Adewole et al., 2016). Neutral detergent insoluble crude protein can also be referred to as neutral detergent insoluble nitrogen (NDIN). In this case, the 6.25 conversion factor is not applied to the calculation. The 6.25 conversion factor is only applied to calculate nitrogen content for protein considering that they contain 16% nitrogen (100/16).

Other MRPs include melanoidins, brown nitrogenous polymers, that can be classified as lignin components. Lysine is especially prone to undergo the MR, thus, bioavailability of Lys to livestock animals is greatly reduced in canola meal, negatively influencing protein digestibility of the meal and therefore, diminishing its overall quality as a feed ingredient.

2.4.2.6. Dietary Fibre

Canola is well-known for its high dietary fibre (DF) content which is a major limiting factor for its utilization in livestock diets. Previously, DF in canola meal was considered to be an ANF as high inclusion in diet resulted in decrease of absorption of proteins, energy, and nutrient digestibility (Radfar et al., 2017). The fraction of DF called non-starch polysaccharides (NSP) is responsible for increasing the viscosity of digesta, reducing the available AA for enzymatic digestion in the gastrointestinal tract. Structurally, NSP are carbohydrate components of the cell wall consisting of rhamnose, arabinose, xylose, mannose, galactose, glucose, and uronic acids. They ranged from 20.0 to 22.9% DM in canola meal (Khajali & Slominski, 2012). It has been demonstrated that the use of exogenous enzymes (i.e., carbohydrase supplementation) had not only improved the nutritive value of canola meal but had also promoted development of beneficial bacteria for gut maintenance and health (Niu et al., 2023).

High variations have been observed in total DF with values ranging from 29.8% DM to 39.2 % DM (Rad-Spice et al., 2018; Canola Council of Canada, 2019a). Many factors contribute to total DF as it is highly influenced by the heat and moisture conditions of oil extraction (section 2.4.2.5. Protein damage in canola meal and subsequent consequences). Inflation of DF by NDICP formation increased TDF also contributing to complications in fibre digestibility in the animal.

2.4.3. Canola meal inclusion in livestock diets

Canola meal can be utilized in all livestock diets, however, is largely more popular for monogastric animals than ruminants. High inclusions of canola meal in poultry diets do not negatively effect animal health and performance due to low glucosinolate contents. Furthermore, breeding efforts in canola to obtain favorable protein and fibre contents in the seeds have made canola meal a viable substitute and/or complement to soybean meal. Slominski & Rogiewicz (2019) have demonstrated the effectiveness of 15-20% inclusion of canola meal in the diets of turkeys, broilers, and layers provided diet formulations were prepared respecting the digestible AA and energy. Likewise, Rad-Spice et al., (2018) concluded that a 15% inclusion rate was effective in the diets of broilers during all phases of growth. In swine, inclusion levels were as a high as 40% in piglet starter diets, 25% in hog grower finisher diets, 25% in gestating sow diets, and 35% in lactating sow diets (Canola Council of Canada, 2019a). In the diets of lactating dairy cows, canola meal has demonstrated to increase milk production and essential AA source compared to other protein sources (Martineau et al., 2013; Martineau et al., 2014).

2.5. WET CHEMISTRY USED FOR ANALYZNG NUTRIENT VALUE OF FEED INGREDIENTS

The criteria for assessing the nutritive value in feed are typically moisture (water), protein, fibre, energy value, and anti-nutritive component contents. The presence of moisture can impact other nutrient analyses by inflating the nutrient contents. Moisture is evaluated to calculate the true dry matter (DM) content of a feed.

Protein is evaluated on a crude protein basis as a measure of total nitrogen content (amino acids and non-protein nitrogen). As amino-nitrogen accounts for 16% of protein weight in animal feed, a 6.25 nitrogen to protein conversion factor is applied (Simonne et al., 1997). Protein evaluation can also be reported on an AA basis. Although many methods exist for protein evaluation (Biuret, Lowry, bicinchoninic acid, Bradford assay, ultraviolet (UV) spectroscopy), the complex composition of protein in feeds limits the options of methods to Kjeldahl and Dumas to determine crude protein content of the feed. The Kjeldahl method (AOAC 954.01 and 988.05) and its modified versions are the most widely established and recognized methods (Simonne et al., 1997; Burns & Ciurczak, 2008; Sáez-Plaza et al., 2013). The basic principle is a conversion of nitrogen in the test sample to ammonia (NH_3) through acid digestion with a catalyst (copper, titanium, or potassium) and quantification via titration (AOAC, 2005). Sensitivity of the Kjeldahl method is a major drawback, as it cannot distinguish protein nitrogen from non-protein nitrogen and may inflate true protein content (Sáez-Plaza et al., 2013). The Dumas method is a combustive method to convert all nitrogenous species in the sample to nitrous oxide (NO), and then reduce it to dinitrogen (N_2) for quantification via thermal conductivity (Simonne et al., 1997). The Dumas method has become standard in automated nitrogen analyzers used for measuring of N and assessing the CP content. However, similar to the Kjeldahl method, the Dumas cannot distinguish nitrogen from protein and non-protein contents.

Accurate quantification of AA is complicated and requires advanced analytical methods and technology. The primary method to analyze for AA includes hydrolysis with HCl to release the AA prior to quantification with a chromatographic technique. The performic acid oxidation method (AOAC 994.12) hydrolyzes protein with performic acid and hydrochloric acid to release individual AA. Amino acids are then separated by ion-exchange chromatography and quantified

by a high-temperature post-column reaction with ninhydrin. Automated techniques such as mass spectroscopy, high-pressure liquid chromatograph (HPLC) and AA analyzers are more efficient and can analyze AA that are sensitive to acid hydrolysis.

Energy sources in feed ingredients are divided into fat and carbohydrates (CHO). Crude fat (CF) is commonly referred to as ether extract (EE) since the basic premise involves solubilizing lipid content (triacylglycerols, phospholipids and waxes, sterols) from a sample with an organic extraction solvent (diethyl ether, hexane, dichloromethane, toluene, etc.) at rising temperatures. The solubilities of lipids are different in each solvent due to differences in solvent polarity. For example, polar lipids (glycolipids and phospholipids) are more soluble in polar solvents (diethyl ether), and non-polar lipids (triaclyerols) are soluble in non-polar solvents (hexane). The most widely recognized technique of CF analysis is Soxhlet extraction, or its modified version called the Randall extraction (RE). In Soxhlet extraction the samples are placed a thimble and continuously extracted with an organic solvent, while in RE procedure, samples are completely submerged in boiling solvent to reduce the extraction time (Thiex et al., 2003; Luque de Castro & Priego-Capote, 2010). In modern fat analysis of agricultural products, RE has been adopted into automated and semi-automated instruments such as XT Series Extractor (ANKOM) or Soxtec® 2000 Series (FOSS); further reducing extraction time, amount of extraction solvent used and labour. More complex methods include acid hydrolysis; performed prior to ether extraction to free all fatty acid from glycerides, glycolipids, phospholipids, and sterol ethers.

Dietary CHO content is categorized into water-soluble/non-structural carbohydrates (starches and classes of sugars), and dietary fibre and its components. Standard methods for determination of starch and sugars are typically employed as described in the Official Methods of Analysis (AOAC). Neutral detergent fibre represents the fraction of a feed or ingredient that is

indigestible to the animal (hemicellulose, cellulose and lignin/phenolics). The basic premise of evaluating NDF (AOAC 2002.04) involves removing all soluble dietary components that are digestible (starches, sugars, proteins, fats, etc.) with neutral detergent (ND) solution and α-amylase under high heat to retain the NDF residue. Non-starch polysaccharides (NSP) analysis utilizes lengthy and complex wet chemistry techniques to process the samples, and gas-liquid chromatography to quantify component sugars that build NSP (Englyst and Cummings, 1984; Slominski and Campbell, 1990).

Determination of total dietary fibre involves a series of complex techniques. Dietary fibre is typically analyzed using enzymatic-gravimetric methods. Several TDF methods has been developed, however the most comprehensive TDF measurement was developed by Slominski et al., (1994) and Slominski (2018) and referred to as the "Manitoba method". It is based on digestion and removal of starch and proteins followed by measurement of NSP, lignin with associated polyphenols and glycoproteins, which would be the fibre-protein composite. In this method, measurement of NSP is conducted in the analyzed samples and its NDF residue (residual matter collected following the NDF procedure) to correct for the losses of neutral detergent-soluble NSP. Determination of NSP is especially complex, since they comprise of multiple sugar derivatives, including arabinogalactans, xyloglucans, galactomannans, pectic polysaccharides (Radfar et al., 2017) and their measurements involve the quantification of each individual sugar compound.

Glucosinolates can be determined by gas-liquid chromatography (GC) as a quantification of desulpho-glucosinolates. Samples are subjected to chemical extraction (ethanol or methanol) and enzymatic hydrolysis with a sulfatase. The initial method described by McGregor & Downey (1975) has been modified and improved by Slominski and Campbell (1987). Glucosinolates can also be determined by HPLC-MS, whereby an isotope pattern is detected based on the intact

glucosinolates' intact sulfur atoms. Like the GC method, samples are prepared using a chemical extraction (Matthäus & Luftmann, 2000). Other methods for glucosinolates determination have also been developed, such spectroscopical method (Daxenbichler & Van Etten, 1977), colorimetric assay (Comer, 1956), but they are not employed in research anymore.

2.6. DEVELOPING NEAR-INFRARED PREDICTION MODELS

2.6.1. General overview

Development of a robust prediction model is required to properly calibrate the near-infrared spectrometer for usage. A number of considerations are involved with this process, including sample type, sample size, homogeneity, distribution and range of the reference values, accuracy and precision of the wet chemistry methods, spectral preprocessing, and instrument limitations, among others (Wiedemann et al., 1998). A prediction model is a mathematical equation that correlates a sample's spectral data and the true value(s) associated with that sample. The instrument will apply the equation of prediction model to predict the analyte of interest in an unknown sample. To develop a robust and reliable model, it is critical that reference values acquired from wet chemistry procedures are carried out with high accuracy and precision. The error associated with the reference values may carry over to the prediction model, as reference data are correlated with the spectral data. This will influence all the predicted values of the model as well as the accuracy and the repeatability of the prediction model. In prediction model development and validation, statistical data is required to verify accuracy and precision of the model, including correlation, standard error of calibration and prediction, root square mean (Barthet et al, 2020).

2.6.2. Sampling errors

During the process of developing and validation a prediction model, sample errors may contribute to inaccuracies and reduced precision in prediction. Although it may occur at several steps in the procedure, it is related to the lack of homogeneity in the material being sampled (Williams & Norris, 2001). Different subsamples are analyzed during reference value and spectral data acquisition. Errors from this particular scenario can be reduced by adequate grinding and mixing of the sample before wet chemistry and spectral data acquisition is performed. Errors can also occur from differences in amount of sample used in the analytical procedure as there is an inverse relationship between sample error and the sample size (Williams & Norris, 2001). Sampling error may occur during the loading and reloading stage for spectral acquisition since different subsamples are used, however, this effect is not as influential as the previously mentioned scenarios.

2.6.3. Sample requirements and selection

A reliable and robust calibration model begins with the collection of samples to develop the prediction model. Samples in the prediction dataset must be in the same form and type, and thus, representative of the samples to be measured. Sample homogeneity is also required within the prediction dataset to keep scattering effects to a minimum and to optimally detect the chemical bonds in the functional groups of the analyte of interest. In addition, a large range and even distribution in the prediction dataset is highly recommended in order to cover the range of unknown analyte to be measured. Large sample sizes are preferred as the model needs to account for all the variation of the analyte in the samples to be tested. In the case of canola meal, differences in breed (yellow v. black-seed coated *B. napus*), applied extraction method (conventional meal v.

expeller-cold pressed), and additional processing (mash v. pellet) may complicate the process of selecting samples due to the large variation between the different characteristics. Principal component analysis (PCA) – a mathematical method for data minimization (section 2.6.6. Principal component analysis) – may assist in deciding which samples belong in the prediction dataset based on variation.

The selection of samples to establish a large range and even distribution of the analytes in the prediction dataset is especially important in the case of canola meal. The heat treatment subjected to the canola meal in oil extraction may contribute to higher variation and increase the range in crude protein and glucosinolate contents. Thus, sample selection requires those samples that comprise of very high and very low crude protein and glucosinolate contents to include in the prediction model to cover the potentially a large range of these analytes in the unknown analyte. Reaching an optimal balance between large sample sizes and uniform distribution of the analytes may be difficult as the distribution may deviate as more samples are added depending on the analyte content of the sample (Wiedemann et al., 1998)

2.6.4. Sample preparation

Sample preparation depends on the type of the sample being analyzed. It may not be required for prediction models developed for whole kernel or whole seed analysis. However, scans are affected by the particle size of the samples. If samples to be analyzed underwent preparation (i.e., grinding), the samples in the prediction and verification set must also undergo the same procedure since consistent particle size are required throughout. In the case of meals (i.e., canola and soybean), it is imperative that all samples are homogenous as possible. Leiva et al., 2022 demonstrated the occurrence of financial loses and resources required to compensate for

underestimation of corn crude protein when NIR analysis was performed with samples ground to the improper size. Canola meal samples were ground to pass through with a 1.0 mm sieve to produce a homogenous material to simplify scan interpretation and reduce light scattering.

2.6.5. Spectral data

A typical NIR spectra consists of many overlapping and broad peaks due to its high sensitivity to the physical characteristics of the sample. Certain bonds absorb light more strongly than other bonds. It also contains redundancies for functional group vibrations in the NIR region. For example, C-H is repeated eight times and may allow the user to work with different wavelength regions in model development (Workman Jr., 2014). However, this poses the disadvantage of identifying the wavelengths that are correlated to analytes of interest. Chemometrics makes identifying such wavelengths possible. In developing a prediction model, the wavelengths that are correlated to the analytes of interest are considered relevant to the prediction of the analyte. Thus, they are selected to be included in the model during its development. Other methods called preprocessing techniques or data pretreatment are also used to enhance the spectra (see section 2.6.7. Spectrum preprocessing).

Development of the prediction model library after samples have been selected and prepared is performed with the scanning of samples for spectral data. Depending on the instrument and software, scanning of one sample is taken three to ten times to reduce any variations observed in each individual spectrum. Software can also apply preprocessing techniques instantaneously to the spectral data in the library following scanning. Although this is generally the simplest stage of the developing the prediction model, thorough mixing of samples between scans is needed to maintain homogeneity and adequate cleaning of the sample holder to prevent contamination.

2.6.6. Principal component analysis

Often in the case of canola meal, factors including breed, texture, origin, nutrient content, and variation in certain properties due to oil extraction may give problematic outcomes in NIRS calibration development. This makes robust prediction models difficult to build with samples with high variations across all parameters. Canola meal samples that were subjected to low and high heat resulted in too much variation in texture and chemical composition to build one robust prediction model for all. The presence of samples with a chemical composition very different from the other samples, might be identified as outlier through principal component analysis (PCA). PCA is a data-reduction method applied to the spectral data to reduce the dataset to a smaller number of uncorrelated linear combinations (principal components) that are representative of the dataset (Linting et al., 2007). When applied to the samples, PCA groups samples based on the spectra's commonalities, trends, or patterns and can provide insight to which samples are appropriate to include in the prediction model. The purpose of PCA is to evaluate the spectral differences, categorize the scans into PCs, identify homogeneity among samples, and more conclusively identify outliers for potential removal (Agussabti et al., 2020). PCA can assist in indicating which samples belong to which "family", and therefore, provides an insight as to how many outliers exist in the dataset or prediction models so the operator can construct prediction models based on those commonalities.

2.6.7. Spectrum preprocessing

Raw NIR spectra may contain undesired or unhelpful signals that could potentially result in incorrect prediction models (Jiao et al., 2020). Spectrum preprocessing (or data pre-treatment) is a set of methods implanted to optimize the information acquired from the spectral data to

construct the prediction model (Jiao et al., 2020). While it is not a prerequisite in constructing a prediction model, preprocessing can greatly improve a model's performance – indicated by improvements in the statistics that are outputted (R^2, root means square of prediction or RMSEP, root means square of calibration or RMSEC), validation bias, relative percent difference (RPD). However, a major caveat to preprocessing is only very few specific methods are proven useful to improve a model (Jiao et al., 2020). It is up to the NIR operator to determine which combination of preprocessing methods is most effective to treat the raw scans and develop a prediction model. Hundreds of preprocessing combinations can exist plus the lack of literature available indicating the most appropriate method(s), limits the optimization of the model (Jiao et al., 2020). The most widespread preprocessing methods are smoothing, Savitzky-Golay first and second derivatives (SG1D and SG2D, respectively), straight line subtraction, multiplicative scatter (or signal) correction (MSC), standard normal variate (SNV; also called vector normalization) and orthogonal signal correction.

It is important to recognize that chemometric and spectral preprocessing methods (also called data pretreatment) are pivotal aspects of NIR model development. Preprocessing can attenuate the technical drawbacks of NIR spectroscopy that generate undesired signals in the spectrum. Overlapping absorption bands and peaks in the NIR spectra can complicate the interpretation of the spectral information, especially within the important peaks (Daszykowski et al., 2008). Savitzky-Golay (SG) first and second derivatives can be applied to reduce peak overlaps. Spectral acquisition can be obstructed by undesired light scattering effects and slope shifting (Daszykowski et al., 2008). Such scattering and shifting effects can be eliminated by multiplicative scatter correlation (MSC) (Jiao et al., 2020). Differences in particle sizes can also result in undesired effects in light scattering and in the resulting spectral data. Such effects can be

eliminated with standard normal variate (SNV). Over 100 preprocessing methods exist to treat the raw scans and develop NIR prediction models (Jiao et al., 2020). It is the researcher's best interest to determine which methods and/or combinations best fits their spectral data to develop a prediction model utilizing the most optimal spectral data. Surprisingly, Lee et al., (2017) reports that most NIR prediction models are developed without the application of preprocessing methods. Preprocessing raw spectra is not a prerequisite to calibration development, and thus, neglecting preprocessing can save considerable time and labour. However, sometimes it is at the expense of model robustness as the removal of undesired peaks and the enhancement of spectra can contribute to the accuracy and the precision of the prediction model.

2.6.8. Generating the prediction equation

The correlation of reference values and spectral data is a calibration that forms the prediction model. Each sample in the prediction model library comprises of a reference value correlated with a predicted value generated by chemometrics. Specifically, PLS regression is used to generate the predicted value based on the reference value a corresponding spectrum of the sample. The relationship between reference and predicted values is generated for all samples in the prediction model library and is depicted in a line of best or prediction equation. The model's performance is indicated by the deviation of data points from the prediction equation.

2.6.9. Validation

Validation is a measure of the performance of the developed model. Validation is carried out using a verification dataset – a set of samples homogeneous but independent to the samples in the prediction dataset. Each sample in the verification is analyzed for a reference value and scanned wit the instrument for a predicted value. The relationship between reference value and NIR-

predicted value is generated for all samples in the verification dataset and is depicted in a mathematical equation.

Evaluative statistics on the verification dataset and the prediction equation measures the precision and accuracy of the developed model. In literature, the most common statistics are R^2, bias, standard error of prediction (SEP) and ratio of prediction to deviation (RPD) (see section 4.8. Criteria for evaluation). The accuracy and precision of the model is indicated by how close the model predicts the verification dataset's analyte value with respect to its reference values. Less variation in the verification dataset exists with closer prediction, resulting a higher value of R^2. A reasonable target when developing a model is an R^2 as close to 1.0 as possible. The maximum value than can be achieved for R^2 is 1.0 and is an unrealistic target as this implies the model predicted the exact value for all reference values in the verification set.

2.7. MODEL TRANSFERRING

The transfer of prediction models from one instrument to another is not a straightforward process but it is a done following a very systematic process. It is performed when using multiple instruments for analyzing the same type of samples at different locations, or when replacing an old or broken instrument (Fearn, 2001). When referring to the instruments in a prediction model transfer, literature often labels first and second spectrometers as "old" and "new", "parent" and "child", "source" and "target", or "master" and "slave", respectively (Fearn, 2001; Mark & Workman Jr., 2013; Bin et al., 2023). Ideally, calibration transfer would be performed without the need for recalibration. It would also ideally produce statistically identical results between instruments for same set of samples, however, this is rarely the outcome in reality (Folch-Fortuny et al., 2017; Workman Jr., 2018). For this reason, transfers between different types of instruments

are rarely executed and is easier performed between instruments of the same type, brand, make, or model (Fearn, 2001).

The main obstacles of transferring prediction models are the complications related to the spectral data. Differences exists between the source and target instruments that generate differences and variation between spectra of the same sample (Fearn, 2001). These differences can be on both x-axis (e.g., wavelength shift) and y-axis of the scan (e.g., absorbance scale). Such differences will even exist between instruments of the same make or model. Though these differences in spectra can be minute (e.g., 2 nm band shift) and appear negligible, they can still lead to inaccuracies when they are carried over to the target instrument. To minimize this effect and facilitate NIRS utilization, several strategies have been established, all involving various chemometric approaches. Generally, three typical approaches exist for a prediction model transfer, including (i) developing a robust prediction equation without standardization, (ii) amending a prediction model to make it compatible with a new instrument, and (iii) refine the spectra from one instrument to another redevelop a more robust prediction model. A fourth option is the use a combination of the three approaches. Developing a robust calibration usually involves reducing the scan scattering using various data treatment. This will correct the difference in the absorbance scale between instruments without correcting the wavelength shifts. To reduce the effect of the wavelength shift, the operator can also limit the number of wavelengths used in the models or they can choose to develop the models using scans of the same samples recorded on all the instruments to accounting for their variations in response intensity and wavelength shift. The fastest method is to modify the bias and slope of the prediction models to correct for the instrument differences. This method usually has very limited application as instrument differences must be minimal to be successful. The most commonly used method is to make adjustments to the spectra (approach iii)

so the sample scans from one instrument match the scans from the other instruments for the same samples. When multiple ($n \geq 2$) instruments are involved, the approach is to ensure that the spectra from various instruments are all adjusted to one common instrument, the target or master instrument. This involves running the same samples in exactly the same conditions on both instruments, so the scan differences are only due to the instrument response and not to the sample.

In some cases, despite the user's best efforts to apply all relevant techniques to a model transfer, a well-performing prediction model on a source instrument may consistently generate poor results on a target instrument. Some prediction models could be too specific to the source instrument and could be considered nontransferable. Sometimes, it is easier to keep a sample collection made of a calibration sample set and an independent verification sample set, previously used for the model developments (and testing) and redevelop models by scanning all the calibration samples on the new instruments and test the newly developed model with the independent verification set sample.

CHAPTER 3: OBJECTIVES

The objectives of this thesis were to:

1. Develop primary near-infrared prediction models for glucosinolates contents of canola meal using samples collected from various crushing plants over multiple years.
2. Collect additional canola meal samples, analyze them for the glucosinolates content by gas-liquid chromatography, and generate data to constitute the independent verification sample set.
3. Validate the developed prediction models using the verification sample set and improve the models if necessary.
4. Transfer the developed prediction models to other NIR spectrometer and validate the new models.

CHAPTER 4: METHODS AND MATERIALS

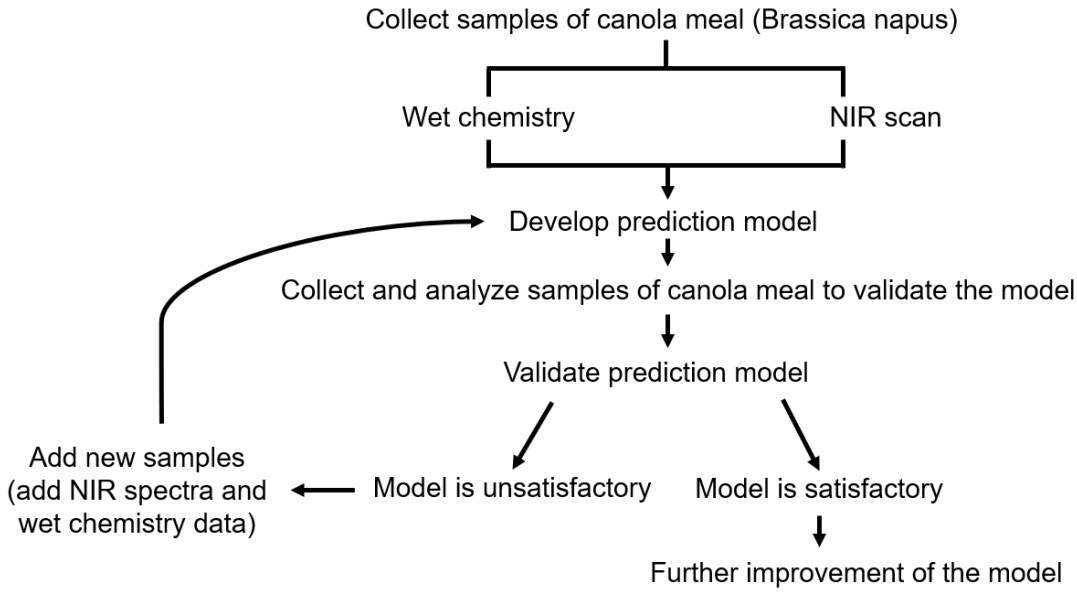

Fig 4. General workflow of developing a near-infrared prediction model

4.1. GENERAL OVERVIEW

Components to be predicted by NIR spectroscopy (glucosinolates, proteins and ether extract) for each sample in the model libraries (calibration sample set and independent verification set) were analyzed by reference methods. The spectral data were acquired with an NIR spectrometer. Values obtained from reference methods were taken as the true/lab predicted values and correlated with preprocessed spectral data to construct the prediction model. Validation was performed with samples that were independent of the prediction dataset and models were evaluated. The process was repeated until model(s) with satisfactory statistical performance were

obtained when tested on the independent verification sample set. Principal component analysis was performed on the raw spectra of the samples in the dataset to detect trends, outliers, and spectral similarities. This provided insight on which samples could be removed or retained in a model's dataset. To develop a model, it is important that the constituent distribution be as uniformly as possible over the range of concentrations that will be predicted by the model. Unfortunately, the components had a binomial distribution in the canola meal sample set, maximum and minimum scans were averaged to fill the gaps and artificially create a library with a uniform distribution of components.

4.2. SAMPLE PREPARATION AND COLLECTION

Samples from two collection periods, obtained from Canadian canola processing plants and from feed ingredients suppliers, were used in this study. A total of 271 canola meal samples were collected between 2012 and 2019, and supplementary samples totalling 161 of different forms (mash, expeller cold-pressed, cake) were collected from Canada, Poland, and China between 2020 and 2023. Sources of canola meal included: Bunge Oilseed Processing (Altona, MB; Harrowby, MB; Nipawin, SK; Fort Saskatchewan, SK and Hamilton, ON); Archer Daniels Midland (ADM) Agri-Industries Ltd. (Windsor, ON; Lloydminster, SK and Yorkton, SK); James Richardson International (JRI) Canola Oil Processing Plant (Yorkton, SK and Lethbridge, AB), and Cargill Canola Processing (Clavet, SK), and ten independent commercial feed processors without disclosing the specific origins (blind sampling). Samples of expeller-cold pressed canola samples were collected from eight different canola seed crushing plants across Western Canada, including: Pleasant Valley Oil Mills; Millford Hutterite Colony; Prairie Home Hutterite Colony; High-Oil Meal; Miltow Hutterite Colony; Mountain View Hutterite Colony; M&C Commodities; Bio Meal Milligan Biofuels. All samples were ground to pass through a 1 mm sieve using a Cycotec tecator

sample mill (FOSS Analytical). Refer to section 4.5. for a summary of sample utilization in each model developed.

4.3. CANOLA MEAL CHEMICAL ANALYSIS

Samples were analyzed for dry matter (DM), ether extract (EE), crude protein (CP), glucosinolates (GLS), total dietary fibre (TDF) and its components, i.e., neutral detergent fibre (NDF) detergent insoluble crude protein (NDICP), and lysine at the Department of Animal Science (University of Manitoba; Winnipeg, MB).

4.3.1. Dry matter

Dry matter (revised AOAC Official Method 930.15) was determined by drying 1.0 g of sample in aluminum dishes for 5 hours (minimum) at 102°C in an oven (Fisher Scientific).

4.3.2. Ether extract

Ether extract in the samples was determined after hexane extraction (AOAC Official Method 920.39) in an ANKOM extraction system (ANKOM Technologies). Each sample was prepared in duplicate by weighing 0.5 g into XT4 filter bags and dried in a drying oven (minimum three hours or overnight at 102°C). Samples are placed inside a desiccator to cool before automatic immersion in hexane in the extraction vessel. Following extraction at 90°C, samples were placed in a drying oven (Fisher Scientific) to remove residual hexane and cooled in a desiccator before final weights were recorded and ether extract content was calculated.

4.3.3. Crude protein

Crude protein was calculated using nitrogen content multiplied by a 6.25. Nitrogen was determined by combustion method (AOAC Official Method 990.04) using a LECO nitrogen analyzer (model FP828; LECO Corporation). Ethylenediaminetetraacetic acid (EDTA) was used to calibrate the instrument. EDTA and duplicate samples (0.150-0.155 g) were weighed into tinfoil and formed into a raindrop shape to load into the autosampler. Combustion and exporting results are automated.

4.3.4. Glucosinolates content

Glucosinolates content was determined by preparing desulfo-glucosinolates extracts for analysis with gas-liquid chromatography, as described by Slominski & Campbell (1987). *Brassica napus*-derived canola meal samples were incubated on a tube rotator for three hours at 25°C in a methanolic solution with 0.2 mL lead barium acetate and 1 mL of internal standard (IS; benzyl glucosinolate or sinigrin). Following incubation, the methanolic supernatant was transferred to DEAE-Sephadex A-25 pyridine acetate form columns post-centrifugation (3000 rpm for 10 minutes) and washed with 67% methanol, distilled water, and 0.02 M pyridine acetate. The supernatant is incubated overnight at 25°C with 0.1 mL purified sulfatase. Following overnight incubation, desulpho-glucosinolates are eluted from columns with 60% methanol, dried completely at 60°C with a reacti-therm heating module and derivatized in a (2:1:0.1:0.05) solution of acetone: bis(trimethylsilyl)acetamide: chlorotrimethyl-silane: 1-methylamidazole. Derivatized glucosinolates were analyzed as individual glucosinolates by gas-liquid chromatography (Varian 430, SCION Instruments).

4.3.5. Dietary fibre and its components

The total dietary fibre procedure includes measuring of content of non-starch polysaccharides, proteins bound within the cell wall, lignin with associated polyphenols, and Maillard reaction products. Total dietary fibre was determined as a sum of neutral detergent fibre and the non-starch polysaccharide fraction of neutral detergent fibre. The content of NSP in both the sample and the residue collected following the NDF analysis. It allows for the correction of losses of neutral detergent soluble NSP. Protein present in the NDF residue is to quantify the protein-containing fibre fraction referred to as glycoprotein or NDICP. This fraction contains cell wall protein and advanced glycation end products of Maillard reaction between AA and reducing sugars.

4.3.5.1. Neutral detergent fibre

Neutral detergent fibre was determined with an ANKOM fibre analyzer by digesting non-fibrous components in neutral detergent (ND) solution (mixture of sodium dodecyl sulfate, ethylenediaminetetraacetic acid, sodium borate, sodium phosphate and tri-ethylene glycol in 1 L distilled H_2O) using agitation. The retained residue represents the neutral detergent fraction (automated AOAC Official Method 2002.04). Ground sample was weighed (0.45-55 g) into F57 filter bags in duplicate and sealed with a heated bag sealer three times to ensure the bag remained sealed during agitation. Sample bags were placed onto circular racks (three per rack) and loaded into the fibre analyzer (Model A200; ANKOM Technology). Along with the samples, a blank bag was included to determine blank bag correction. A total of four washes were performed. The initial wash used 2 L of ND solution, 20 g sodium sulfite and 4 mL alpha-amylase for 75 minutes in an agitated motion. The second and third washes were performed with 2 L of boiling water and 4 mL

alpha-amylase. The fourth wash was performed only with boiling water. The residual NDF was rinsed in acetone, air-dried and oven dried at 102°C before final weights were collected.

4.3.5.2. Non-starch polysaccharides

Quantification of total non-starch polysaccharides was carried out as described by Englyst & Cummings (1984) and Slominski & Campbell (1990). The first part of the procedure disperses the starch content from the sample. Samples are weighed (100 mg for feed, 50 mg for digesta or excreta, 40 mg for NDF residue) in duplicate into 50 mL centrifuge tubes, 2 mL of dimethyl sulfoxide was slowly added, heated, and shaken to disperse sample. Samples are heated at 100°C for 60 minutes. Sodium acetate buffer (pH = 5.2), heated to 50°C, was added to each sample (3 mL) along with 3 mL alpha-amylase. After boiling for another 15 minutes, samples are cooled, 0.1 mL amyloglucosidase was added and incubated in a shaker overnight at 45°C. The following day, 42 mL of 95% ethyl alcohol was added, the samples were mixed and incubated for 1 hour before centrifugation (2500 rpm for 10 minutes). Ethanolic supernatant was discarded, and residue was oven dried. Sulfuric acid (1 mL, 12 M) and a marble was added to the tube to break up the residue at 35°C over a one-hour period. Six mL water and 5 mL myo-inositol were added to each sample which were incubated at 100°C. A 1 mL subsample was taken from each sample and prepared for GC analysis by releasing individual polysaccharides as alditol acetates. To each subsample, a drop of DL-2-octanol, and 12 M ammonium hydroxide (0.26 mL) is added and mixed. Sodium tetrahydroborate was added dropwise and the sample was monitored for hydrogen formation. Samples are left to stand for one hour before adding acetic acid (0.1 mL), N-methyl imidazole and acetic anhydride (2 mL). After another half hour, 5 mL water and 1.0 mL dichloromethane were added, and samples were capped and shaken. The following morning, the organic layer from each

sample was transferred to the GC vials for GC analysis or quantification of rhamnose, arabinose, xylose, mannose, galactose, and glucose.

Uronic acids were determined by spectrophotometry. Distilled water was added to hydrolysate (0.05 mL for canola or flax) and diluted to 0.300 mL. Sodium chloride-boric acid solution (0.3 mL) and concentrated sulfuric acid (5 mL) was then added and vortexed. The sodium was incubated at 70°C for 40 minutes before 3,5-dimethylphenol solution (0.2 mL) was added and mixed. After a 5 minute-incubation period, the absorbance was acquired with a spectrophotometer (Genesys 140). Final concentrations of uronic acids were determined through measuring against a glucuronic acid standard curve.

4.3.5.3 Total Dietary Fibre

Total dietary fibre was determined by a combination of NDF and neutral detergent soluble NSP measurements to address the solubility of NSP in the neutral detergent solution that causes a loss of NSP on NDF assay. Total dietary fibre was calculated as the sum of NDF and ND-soluble NSP (Slominski et al., 1994; Slominski, 2018). Neutral detergent soluble NSP were calculated as total sample NSP minus NSP present in the NDF residue. Neutral detergent insoluble crude protein (NDICP) represented the amount of crude protein present in the NDF residue. The value for lignin and associated polyphenols was calculated by difference [NDF – (NSP +NDICP)].

4.4. SPECTRAL DATA ACQUISITION

Spectral data was acquired in triplicate by a Matrix FT-NIR spectrometer (Bruker Optik GmbH, Germany) with a lead-sulfide (PbS) IR detector and integrating sphere in the 25000 – 4000 cm^{-1} (400 nm – 2500 nm) wavelength number range at 0.5 nm resolution for all samples (calibration sample set and validation sample set). Approximately 20-25 g of each canola meal

sample was carefully transferred into the sample cell after thoroughly mixing for homogeneity before placing it into the spectrometer's rotating holder. Each scan takes approximately 40 seconds, as the NIR spectrometer scanned the sample 64 times to produce the raw spectra. After each spectrum was recorded, sample was returned to its container, mixed, subsampled and transferred to the cell to produce duplicate and triplicate scans of the samples. Between samples, the sample cell was wiped carefully with ethanol and dried with Kimwipe as canola meal dust stayed on the cell due to static interaction.

4.5. SAMPLE DATASETS OF PREDICTION MODELS I, II, AND III

A set of 271 canola meal samples, (canola meal of Canadian origin, plus some high glucosinolates samples of Asian origin) collected from 2012-2019 were used to develop Prediction Model I. Model I were developed with 227 samples for glucosinolates, 258 samples for crude protein and 254 for ether extract. Validation of the various models was carried out with 50 canola meal and 16 expeller cold-pressed canola samples, respectively.

Additionally, 161 samples collected at a later date were added to the first 271 sample set, totalling 432 samples available for model development. This second sample set was reduced to 370 after elimination of expeller cold-pressed canola meal and meals of non-Canadian origin (rapeseed instead of canola) following the PCA analysis. This set of samples was further segregated into mash canola meal samples (n = 261) or pellet canola meal samples (n = 109) for subsequent model development. Models II and III were developed with mash and pellet canola meal, respectively. Additional 128 average spectra were created by scan average increasing the total number mash canola meal samples to 389 to fill the component concentration gap for development and validation of Model II. A set of 87 pellet average spectra were also created for

the same reason (filling the component concentration gap) increasing the total number of pellet canola meal samples to 196 for development and validation of Model III. Model II was developed with 278 samples for glucosinolates, 196 samples for crude protein and 199 samples for ether extract. Validation was performed with 101 samples for glucosinolates, 62 for crude protein and 58 samples for ether extract. Model III was developed with 149 samples for glucosinolates, 148 samples for crude protein and 145 samples for ether extract. Validation was performed with 46 samples for glucosinolates, 47 for crude protein and 47 for ether extract.

4.6. PREPROCESSING

Preprocessing techniques were applied on the raw spectral data to extract optimal spectral information to develop the prediction models. Multiplicative interferences (i.e., baseline slope increase, and baseline offset) were eliminated with a standard normal variate (SNV) treatment. Savitzky-Golay first and second derivatives (SG1D and SG2D, respectively) were applied to amplify the scan information. Specifically, it addressed overlapping peaks and large baseline variations (William & Norris, 2001). Base corrections were made with straight line subtraction. Figure 5. Illustrates the sharpening of the absorbance peeks from raw spectra through SG2D from pellet, mash, and expeller samples. Table 4. summarizes the preprocessing techniques applied to the spectral data.

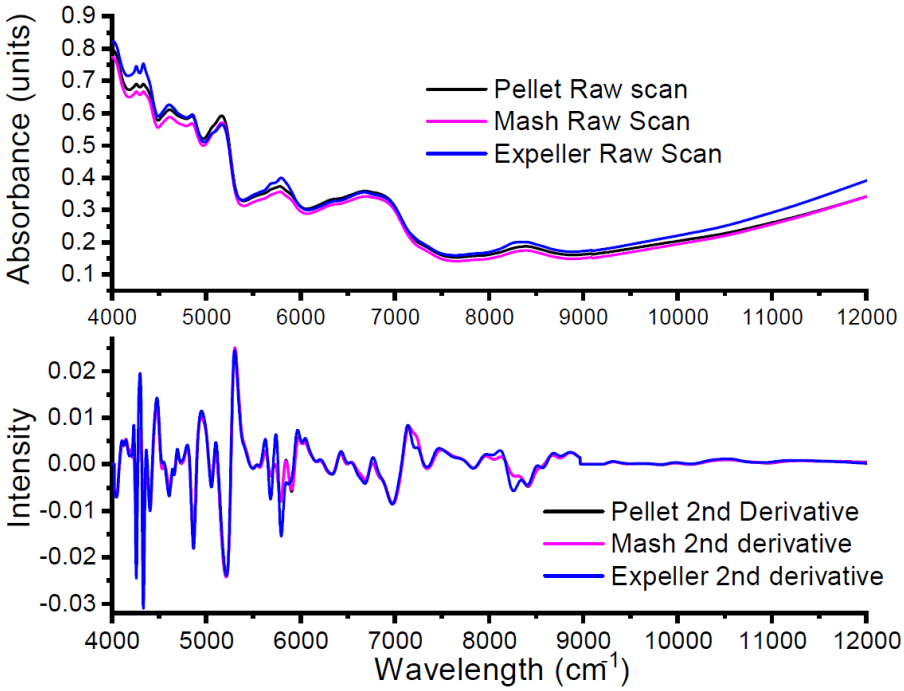

Fig 5. Comparison of raw and second derivative spectra for a pellet, mash, and expeller sample

Table 4. Preprocessing techniques applied to each component in Models II and III

Prediction Model	Component	Derivative applied	SNV[1] applied	Smoothing points	Straight line subtraction	Wavelength selection (cm^{-1})
II (mash canola meal)	Glucosinolates	Second	Yes	8	Yes	7504-5448
	Crude Protein	Second	Yes	9	Yes	9400-5448 and 4600-4248
	Ether Extract	First	Yes	5	No	9400-4600
III (pellet canola meal)	Glucosinolates	Second	Yes	8	Yes	9043.8-4242.9
	Crude Protein	Second	Yes	9	No	9403.8-6094.4 and 5455.1-424.9
	Ether Extract	Second	Yes	8	Yes	9403.8-5446.3 and 4605.5 4242.9

[1]Standard normal variate

4.7. PREDICTION MODEL DEVELOPMENT

The prediction model was developed using OPUS software (version: 8.2 Bruker Optik GmbH, Germany) with partial least squares (PLS) regression. The samples for the calibration and the independent validation sets were scanned during the same time period but the two sets were distinctly categorized (prior scanning) in OPUS as the calibration data set or "test" data set (validation data set) in the data library. Reference values for fat, crude protein and total glucosinolates content (dry matter basis) were entered into the database on a DM basis.

4.8. CRITERIA FOR EVALUATION

All constructed models were evaluated on the following criteria: standard error of calibration and cross-validation (SEC/SECV) with the calibration sample set, coefficient of determination (R^2), ratio of prediction to deviation or residual prediction deviation (RPD), and standard error of prediction (SEP) for the independent validation set (Williams & Norris, 2001; Ali Redha et al., 2023). Secondary criteria include mean, standard deviation, median, minimum, maximum, and bias. Relative percent different was determined following model construction from OPUS. All other criteria were calculated using Microsoft Excel 2021 after exporting calibration and validation reports from OPUS.

Table 5. Common criteria used for evaluating near-infrared prediction models (ISO, 2017)

Criteria	Formula	Evaluation
Coefficient of determination (R^2) (no unit)	$R^2 = 1 - \frac{SS_{error}}{SS_{total}}$	Estimation of the variation between the reference data and the NIR predicted data obtained by the prediction model. Can be calculated using calibration set and the independent verification set data.
Ratio of prediction to deviation/ residual prediction deviation (RPD) (no unit)	$RPD = \frac{SD}{SEP}$	Indicate how the model is able to predict future data based on the variation (SEP) of the model. Can only be calculated using the independent verification set.
Standard error of calibration (SEC) (same unit as the reference value)	$SEC = \sqrt{\frac{\sum(RV-NV)^2}{n-1-p}}$ RV = reference value NV = NIR predicted value p = no. of terms in model	Measure of the precision of the predicted values compared to the reference values during the cross-validation step using the calibration set
Standard error of prediction (SEP) (same unit as the reference value)	$SEC = \sqrt{\frac{\sum(RV-NV)^2}{n}}$ RV = reference value NV = NIR predicted value	Measure of the precision of the predicted values compared to the reference values using the independent validation set data
Bias (same unit as the reference value)	$Bias = \frac{\sum(\hat{Y}-y)}{n}$ $\hat{Y}$ = reference value y = NIR predicted value	Measure of skewness between the predicted dataset and reference value dataset
Root mean standard error of calibration (RMSEC) (same unit as the reference value)	$RMSEC = \sqrt{SEC^2 - Bias^2}$	Measure of calibration accuracy using the calibration set data
Root mean standard error of prediction (RMSEP) (same unit as the reference value)	$RMSEP = \sqrt{SEP^2 - Bias^2}$	Measure of prediction accuracy using the independent validation set data

CHAPTER 5: RESULTS AND DISCUSSION

5.1. RESULTS

5.1.1. Sample Description

5.1.1.1. Summary of the wet chemistry analyses

Values obtained through reference methods for total glucosinolates, crude protein and ether extract contents performed on the samples used for both prediction and validation of Model I, Model II, and Model III are presented in Tables 5 and 6. Descriptive data presented in Tables 5-7 are reported in percentage on dry matter basis for crude protein and ether extract, and in μmol per gram on dry matter basis for glucosinolate contents. Glucosinolate contents are reported on a molecular basis due to the high range of molecular weights for individual glucosinolates.

Table 6. Summarized reference results for canola meal and expeller cold-pressed canola samples used for prediction and validation sets of Model I

	Glucosinolates (μmol g^{-1} DM)	Crude Protein (% DM)	Ether Extract (% DM)
Prediction set			
n	227	258	254
Mean	7.76	41.01	5.19
Standard Deviation	5.25	2.04	3.24
Minimum	1.03	38.60	1.16
Median	7.37	41.39	3.82
Maximum	28.55	44.18	16.02
Validation set (canola meal)			
n	50		
Mean	2.01		
Standard Deviation	0.37		
Minimum	1.33		
Median	2.03		
Maximum	2.87		
Validation set (expeller cold-pressed canola)			
n	16	16	16
Mean	7.85	38.03	18.43
Standard Deviation	2.67	1.64	2.57
Minimum	4.10	35.99	15.31
Median	7.22	37.43	17.98
Maximum	12.74	40.79	24.16

Table 7. summarizes the reference values obtained through wet chemistry for glucosinolates content, crude protein and ether extract performed on the samples that were used in prediction and validation of Model II, constructed using mash canola meal.

Table 7. Summarized reference results for the analyses of mash canola meal samples used for prediction and validation of Model II

	Glucosinolates (μmol g^{-1} DM)	Crude Protein (% DM)	Ether Extract (% DM)
Prediction set			
n	278	196	199
Mean	3.55	41.95	3.45
Standard Deviation	2.76	1.83	0.75
Minimum	0.09	36.80	1.56
Median	2.57	41.77	3.49
Maximum	12.76	46.72	5.63
Validation set			
n	101	62	58
Mean	3.13	41.71	3.27
Standard Deviation	2.51	2.13	0.64
Minimum	0.40	37.00	2.03
Median	2.31	41.59	3.25
Maximum	12.56	45.86	4.84

Overview of results of glucosinolates, crude protein and ether extract of pellet canola meal samples utilized for both the prediction and validation sets within Model III is presented in Table 8.

Table 8. Summarized reference results for the analyses of pellet canola meal samples used for prediction and validation sets of Model III

	Glucosinolates (μmol g^{-1} DM)	Crude Protein (% DM)	Ether Extract (% DM)
Prediction set			
n	149	148	145
Mean	4.18	41.39	3.43
Standard Deviation	2.63	1.28	0.74
Minimum	0.16	37.91	1.16
Median	3.34	41.36	3.42
Maximum	11.50	44.96	5.89
Validation set			
n	46	47	47
Mean	4.02	41.27	3.53
Standard Deviation	2.57	1.25	0.72
Minimum	0.53	38.02	1.61
Median	3.18	41.05	3.47
Maximum	10.55	44.82	5.41

5.1.1.2. Relationships of chemical components in canola meal

Correlations were assessed between glucosinolates content, neutral detergent insoluble crude protein, neutral detergent fibre, and total dietary fibre for 249 of the 271 canola meal samples collected at the beginning of the study. Tabulated statistics are summarized in Tables 8 and 9 and the correlations are illustrated in linear graphs (Figs. 6 to 12). Glucosinolates formed moderate to strong correlations with NDICP, NDF and TDF (R^2 = 0.61, 0.82, and 0.95, respectively). Neutral detergent insoluble crude protein formed moderate correlations with NDF and TDF (R^2 = 0.73 and 0.61, respectively).

Table 9. Glucosinolates, neutral detergent insoluble crude protein, neutral detergent fibre, and total dietary fibre contents of canola meal samples (n = 249).

n = 249	Glucosinolates (µmol g^{-1} DM)	NDICP[1] (% DM)	NDF[2] (% DM)	TDF[3] (% DM)
Mean	3.54	5.31	29.60	38.51
St. Dev	1.72	1.63	2.77	3.19
Minimum	0.00	2.24	22.68	28.22
Median	2.97	5.09	29.42	38.44
Maximum	12.76	14.19	38.79	48.92

[1]Neutral detergent insoluble crude protein; [2]neutral detergent fibre; [3]total dietary fibre

Table 10. Glucosinolates, neutral detergent insoluble crude protein, neutral detergent fibre and total dietary fibre contents of expeller-cold pressed canola samples (n = 16).

n = 16	Glucosinolates (µmol g^{-1} DM)	NDICP[1] (% DM)	NDF[2] (% DM)	TDF[3] (% DM)
Mean	7.88	4.15	24.08	31.70
St. Dev	2.99	1.51	2.36	2.86
Minimum	4.86	2.23	21.16	28.14
Median	7.15	3.46	23.31	31.43
Maximum	15.69	7.63	29.94	37.12

[1]Neutral detergent insoluble crude protein; [2]neutral detergent fibre; [3]total dietary fibre

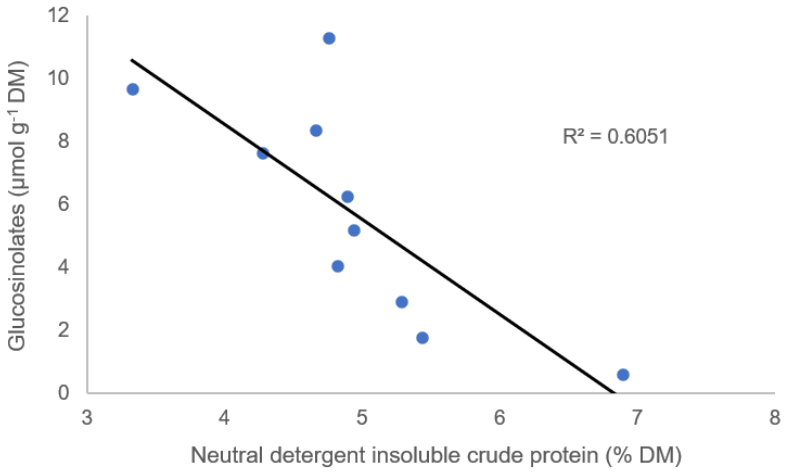

Fig 6. Correlation between total glucosinolate content (µmol g^{-1} DM) content and neutral detergent insoluble crude protein (% DM) for canola meal

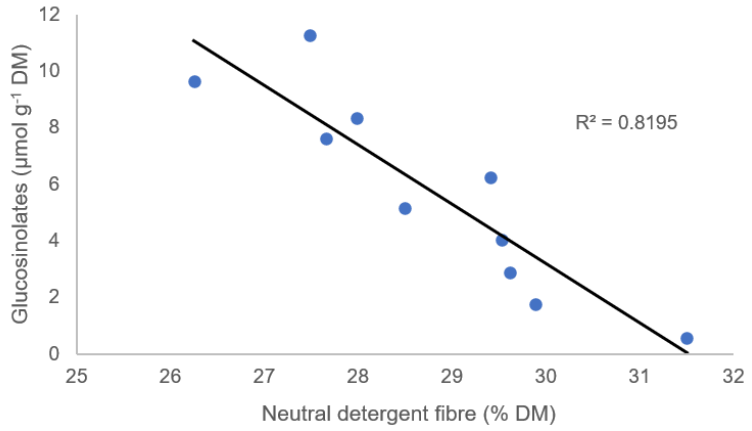

Fig 7. Correlation between total glucosinolate content (µmol g^{-1}) DM and neutral detergent fibre (% DM) for canola meal

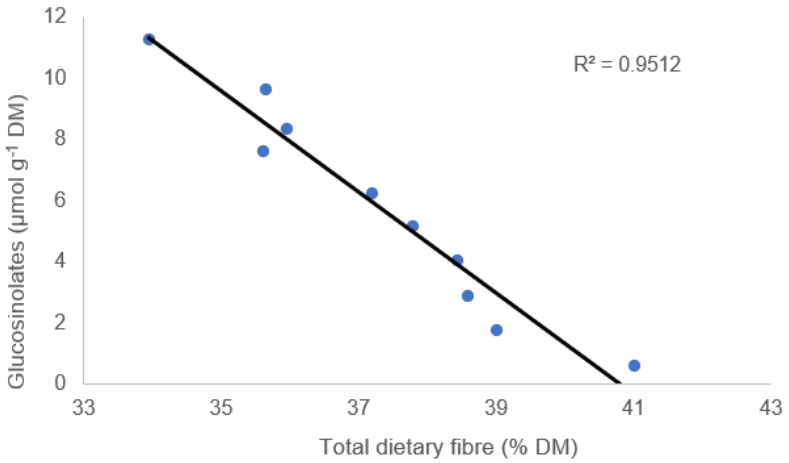

Fig 8. Correlation between total glucosinolate contents (µmol g^{-1} DM) and total dietary fibre (% DM) for canola meal

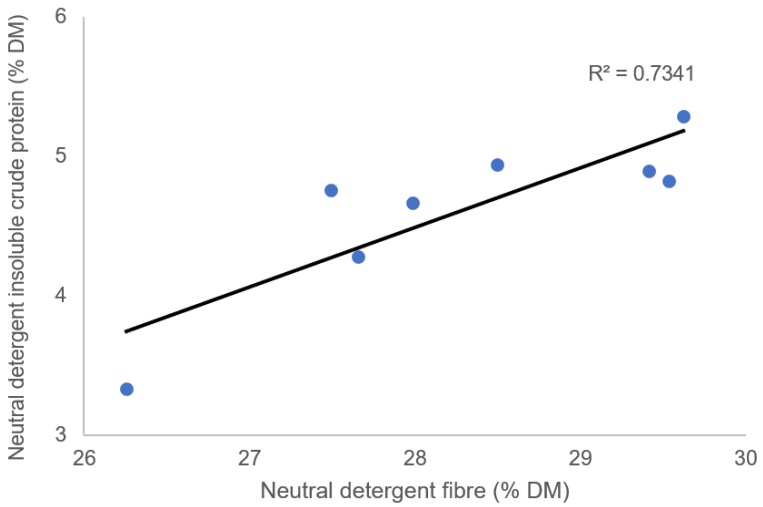

Fig 9. Correlation between neutral detergent insoluble crude protein (% DM) and neutral detergent fibre (% DM) for canola meal

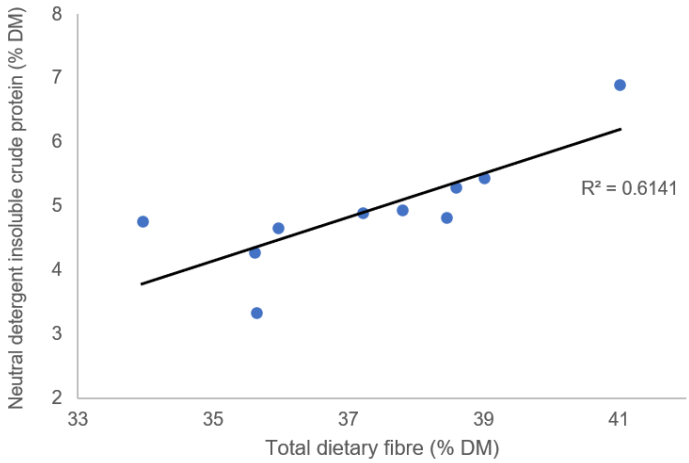

Fig 10. Correlation between neutral detergent insoluble crude protein (% DM) and total dietary fibre (% DM) for canola meal

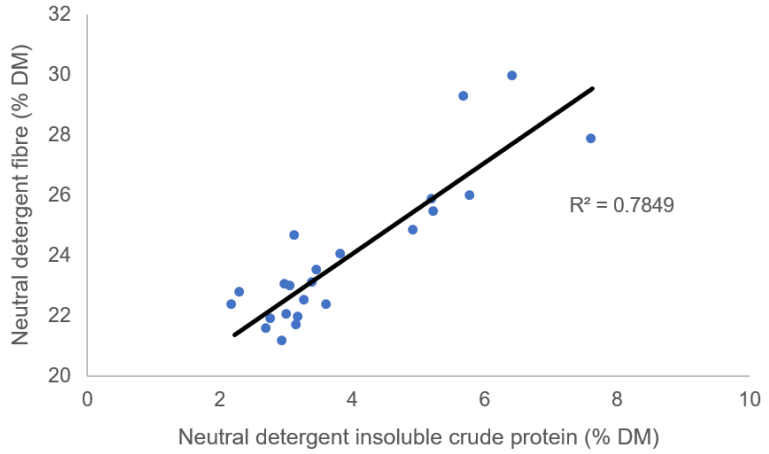

Fig 11. Correlation between neutral detergent fibre (% DM) and neutral detergent insoluble crude protein (% DM) in expeller-cold pressed canola (ECPC)

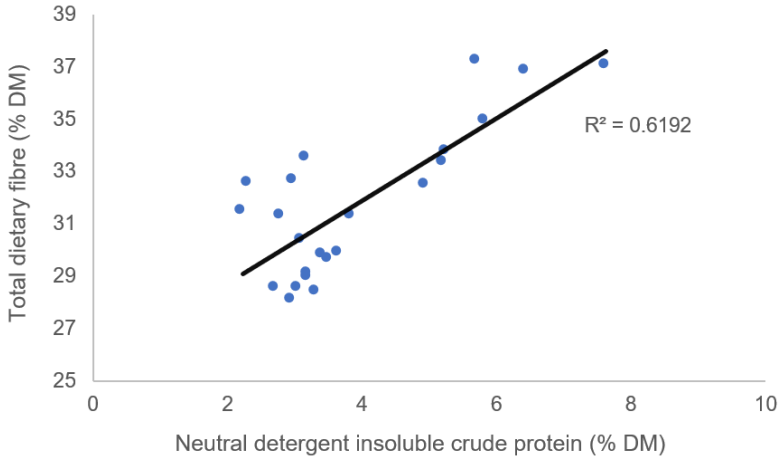

Fig 12. Correlation between total dietary fibre (% DM) and neutral detergent insoluble crude protein (% DM) in expeller-cold pressed canola (ECPC)

5.1.2. Prediction Models

Three prediction models (Models I, II and III) were developed for canola meals to predict glucosinolates, protein, and ether extract contents by correlating reference values with the sample spectral data. Model I was validated for all types of canola meals for glucosinolates and then expeller cold-pressed canola meals for glucosinolates, protein and ether extract. Validation of Model I produced unsatisfactory results when tested with expeller cold-pressed canola meals, prompting the developments of models specific for the canola meal type; (i.e., Model II for mash canola meal and Model III for pellet canola).

5.1.2.1. Prediction Model I

Model I generated a high correlation in the prediction of calibration. Contents of glucosinolates, CP, and EE generated R^2 = 0.90, 0.91 and 0.97, respectively, with low standard error of calibration (SEC) and good to excellent relative percent difference (RPD) (Table 11.). Calibration statistics are also illustrated in linear equations (Figs 13-15). Validation with ECPC for glucosinolates, protein, and fat generated R^2 = 0.49, 0.61, and 0.95, respectively, featuring moderate standard error of prediction (SEP) and low RPD, except for ether extract (Table 11.). Validation statistics are also illustrated in linear equations (Figs. 16-19). Validation with the canola meal set was only performed for glucosinolates, generating R^2 = 0.0005 (Table 12.), low SEP, good RPD, and is illustrated in a linear equation (Fig 16.).

Table 11. Cross-validation and validation set statistics for glucosinolates (μmol g^{-1} DM), crude protein (% DM) and ether extract (% DM) for Model I

	Glucosinolates	Crude Protein	Ether Extract
Cross-validation using the calibration sample set			
R^2	0.90	0.91	0.97
SEC[a]	1.65	0.62	0.54
Number of factors in the model	20	20	20
Validation set performed with an independent canola meal sample set (2012-2019)			
R^2	0.0005		
$\bar{e}$[b]	1.76		
SEP[c]	0.57		
RMSEP[d]	N/A		
RPD[e]	3.51		
Validation Set performed with an independent expeller cold-pressed canola meal sample set (2020-2023)			
R^2	0.49	0.61	0.95
$\bar{e}$[b]	-0.31	-2.28	0.42
SEP[c]	1.98	1.06	0.58
RMSEP[d]	1.96	2.51	0.40
RPD[e]	1.35	1.55	4.46

[a] standard error of calibration; [b] bias; [c] standard error of prediction; [d] root means square error of prediction; [e] ratio of performance to deviation

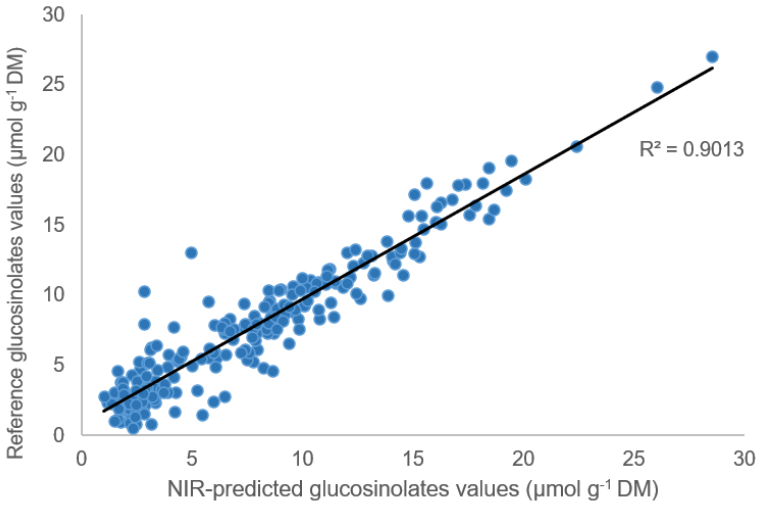

Fig 13. Correlation between reference and NIR-predicted values for calibration sample set for total glucosinolates content (µmol g^{-1} DM) of Model I

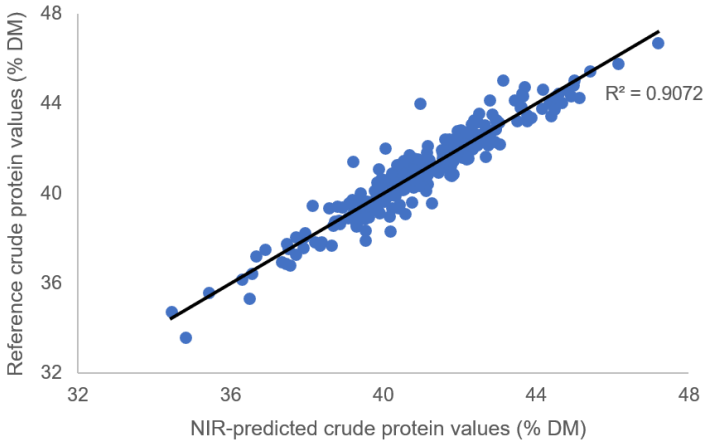

Fig 14. Correlation between reference and NIR-predicted values for calibration sample set for crude protein (% DM) of Model I

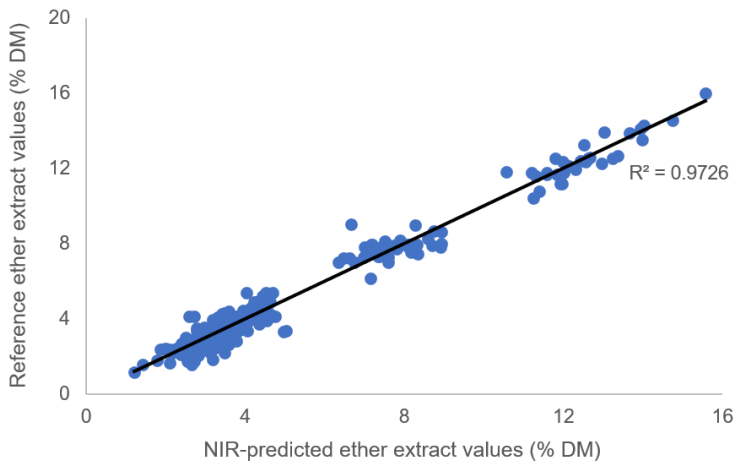

Fig 15. Correlation between reference and NIR-predicted values for calibration sample set for ether extract (% DM) of Model I

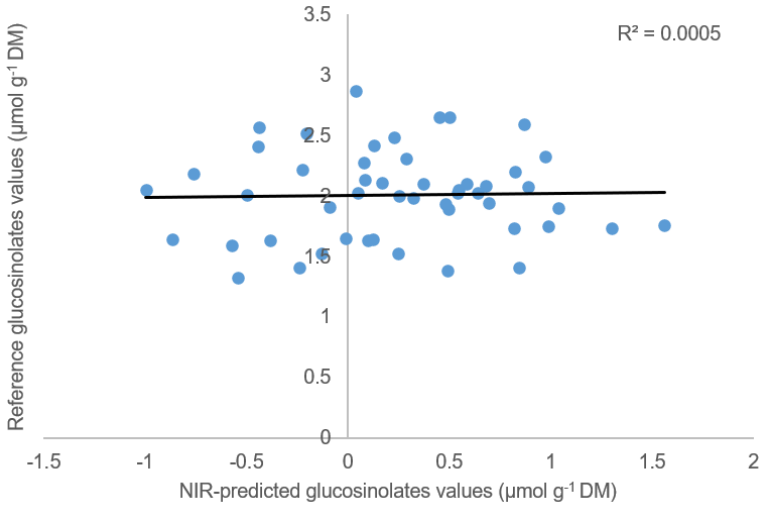

Fig 16. Correlation between reference and NIR-predicted values for validation sample set for total glucosinolates content (µmol g^{-1} DM) with canola meal of Model I

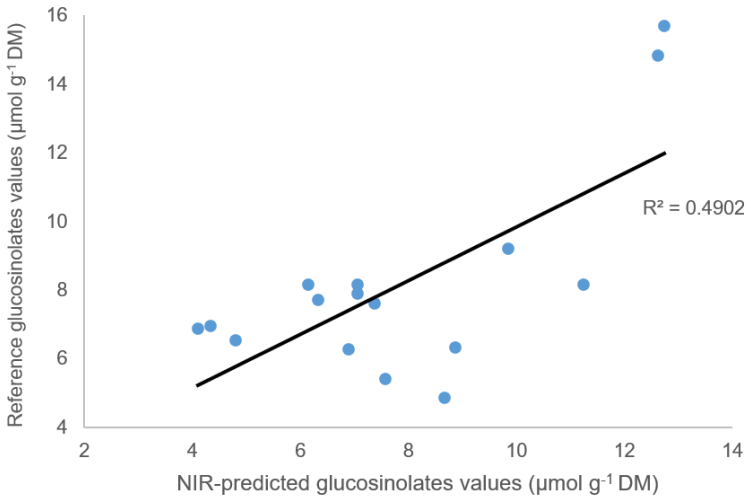

Fig 17. Correlation between reference and NIR-predicted values for validation sample set for total glucosinolates content (µmol g^{-1} DM) with expeller-cold pressed canola of Model I

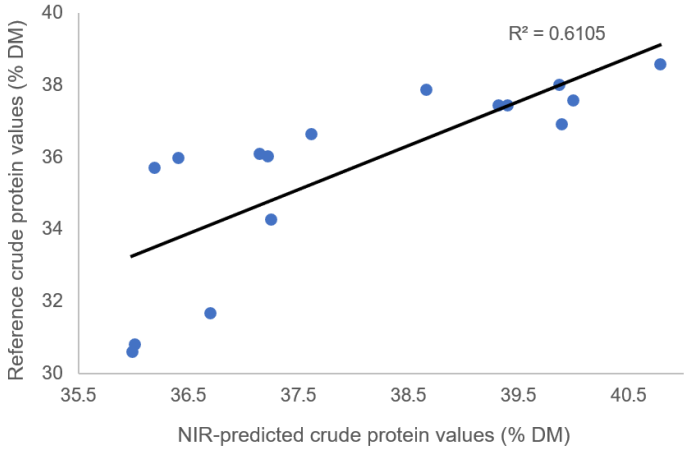

Fig 18. Correlation between reference and NIR-predicted values for validation sample set for crude protein (% DM) with expeller-cold pressed canola of Model I

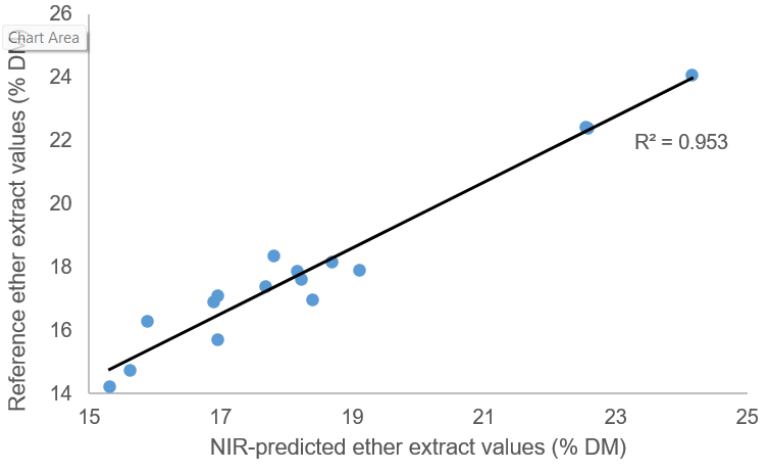

Fig 19. Correlation between reference and NIR-predicted values for validation sample set for ether extract (% DM) with expeller-cold pressed canola of Model I

5.1.2.2. Prediction Model II

Model II generated high correlations when tested on the validation set. Contents of glucosinolates, crude protein and ether extract generated R^2 = 0.89, 0.93 and 0.91, respectively, with low standard error of calibration and good to excellent RPD (Table 12.). Calibration statistics are also illustrated by the linear regressions presented in Figures 20 to 22. The canola meal validation statistics for glucosinolates, crude protein and ether extract, generated R^2 = 0.87, 0.91, and 0.74, respectively, featuring moderate SEP and low to moderate RPD (Table 12.), the linear graphic representations are presented in Figures 23 to 25).

Table 12. Cross-validation and validation set statistics for glucosinolates (µmol g^{-1} DM), crude protein (% DM) and ether extract (% DM) for Model II, developed with mash canola meal

	Glucosinolates	Crude Protein	Ether Extract
Cross-validation using the calibration sample set			
R^2	0.89	0.93	0.91
SEC[a]	0.93	0.49	0.22
Validation set with independent sample set			
R^2	0.87	0.91	0.74
$\bar{e}$[b]	0.00	-0.09	-0.02
SEP[c]	0.90	0.66	0.33
RMSEP[d]	0.90	0.67	0.33
RPD[e]	2.78	3.25	1.93

[a] standard error of calibration; [b] bias; [c] standard error of prediction; [d] root means square error of prediction; [e] ratio of performance to deviation

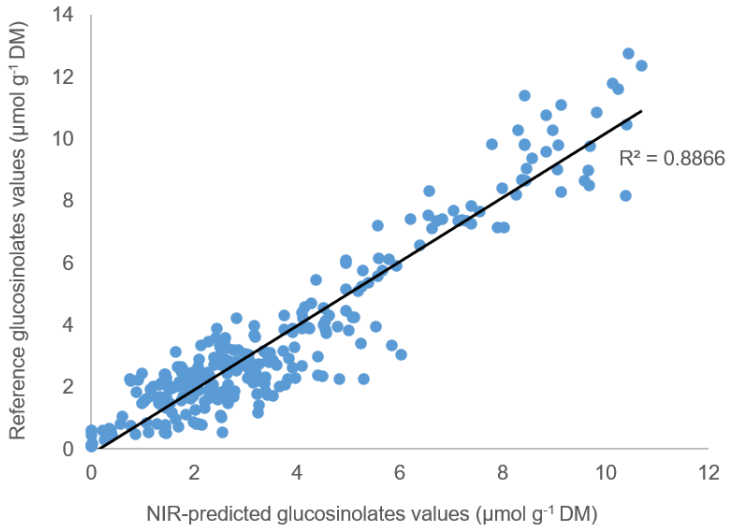

Fig 20. Correlation between reference and NIR-predicted values for calibration sample set for total glucosinolates contents (µmol g^{-1} DM) of Model II

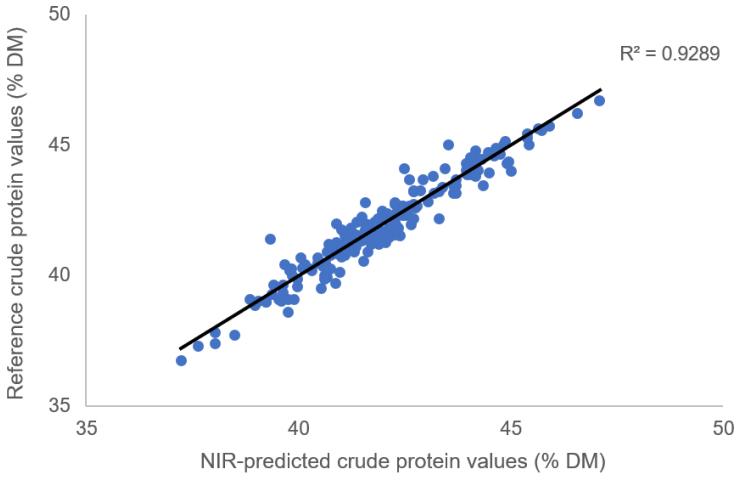

Fig 21. Correlation between reference and NIR-predicted values for calibration sample set for crude protein (% DM) of Model II

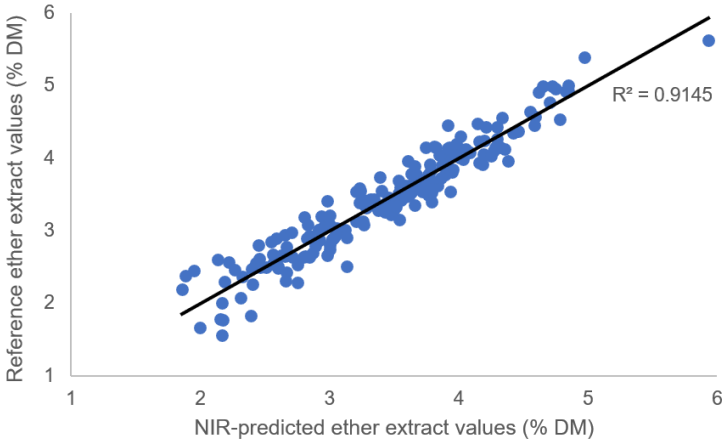

Fig 22. Correlation between reference and NIR-predicted values for calibration sample set for ether extract (% DM) of Model II

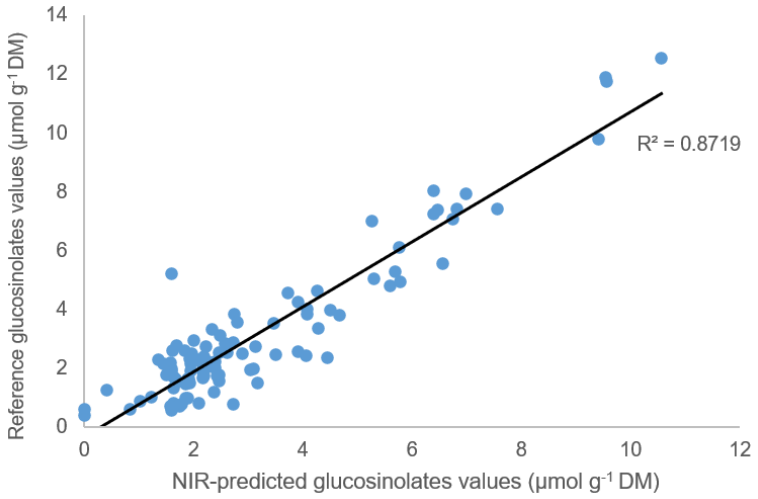

Fig 23. Correlation between reference and NIR-predicted glucosinolates values for validation sample set for total glucosinolates content (µmol g^{-1} DM) of Model II

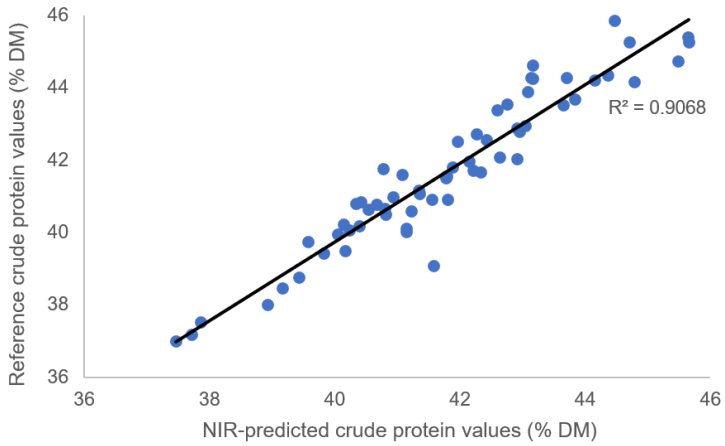

Fig 24. Correlation between reference and NIR-predicted values for validation sample set for crude protein (% DM) of Model II

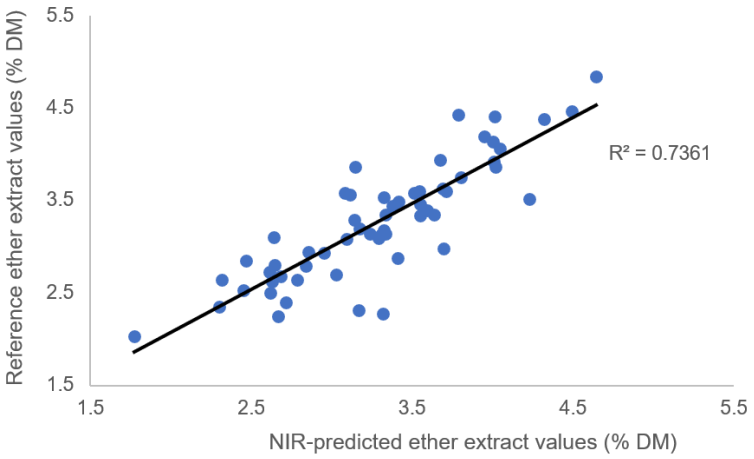

Fig 25. Correlation between reference and NIR-predicted values for validation sample set for ether extract (% DM) of Model II

5.1.2.3. Prediction Model III

Model III generated a high correlation in internal cross-validation. Contents of glucosinolates, crude protein and ether extract generated R^2 = 0.91, 0.90 and 0.95, respectively with low standard error of calibration and good to excellent RPD (Table 13.). Validation with canola meal sample set for glucosinolates, protein, and ether extract generated R^2 = 0.91, 0.83, and 0.82, respectively, featuring low SEP and moderate to good RPD. The linear regressions of the predicted NIR results versus the reference data for both the calibration samples and the verification samples are presented in Figures 26 to 31.

Table 13. Cross-validation and validation set statistics for glucosinolates (μmol g^{-1} DM), Crude Protein (% DM) and Ether Extract (% DM) for Model II, developed with pellet canola meal

	Glucosinolates	Crude Protein	Ether Extract
Cross-validation using the calibration sample set			
R^2	0.91	0.90	0.95
SEC[a]	0.80	0.40	0.16
Validation set with independent sample set			
R^2	0.91	0.83	0.82
$\bar{e}$[b]	0.07	-0.11	0.12
SEP[c]	0.79	0.52	0.31
RMSEP[d]	0.79	0.53	0.33
RPD[e]	3.25	2.40	2.36

[a] standard error of calibration; [b] bias; [c] standard error of prediction; [d] root means square error of prediction; [e] ratio of performance to deviation

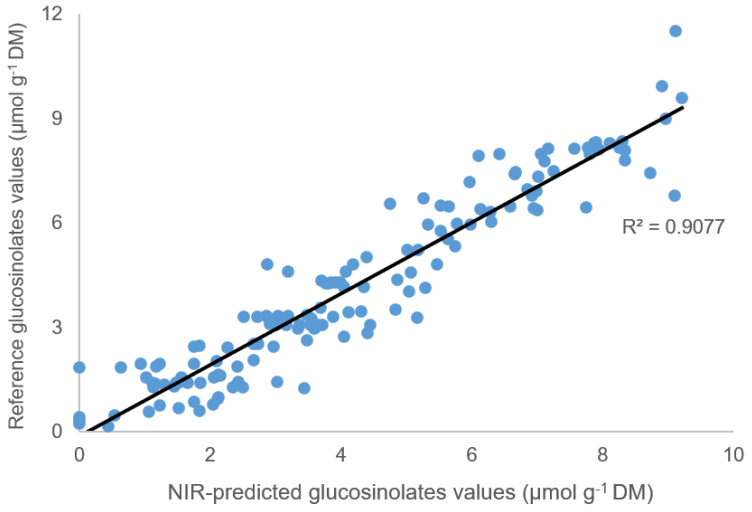

Fig 26. Correlation between reference and NIR-predicted values for calibration sample set for total glucosinolate content (µmol g^{-1} DM) of Model III

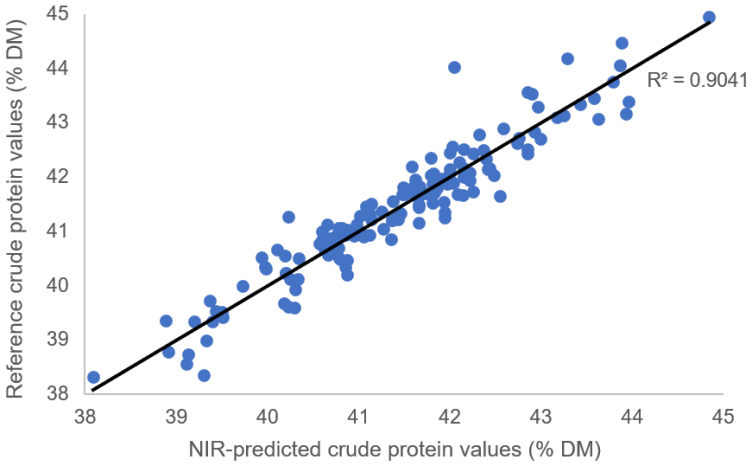

Fig 27. Correlation between reference and NIR-predicted values for calibration sample set for crude protein (% DM) of Model III

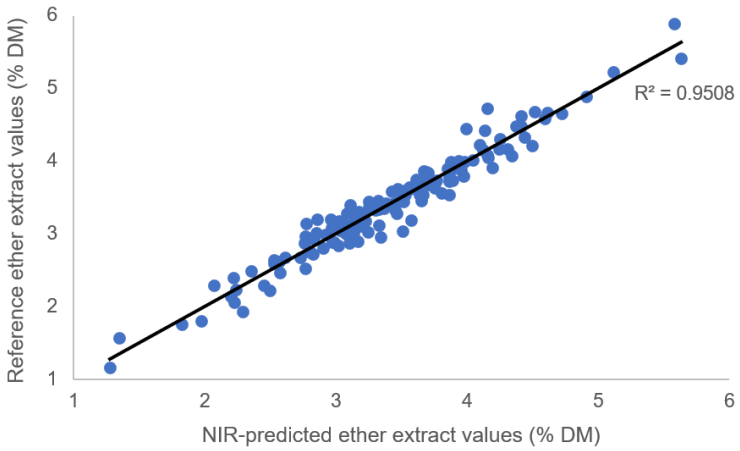

Fig 28. Correlation between reference and NIR-predicted values for calibration sample set for ether extract (% DM) of Model III

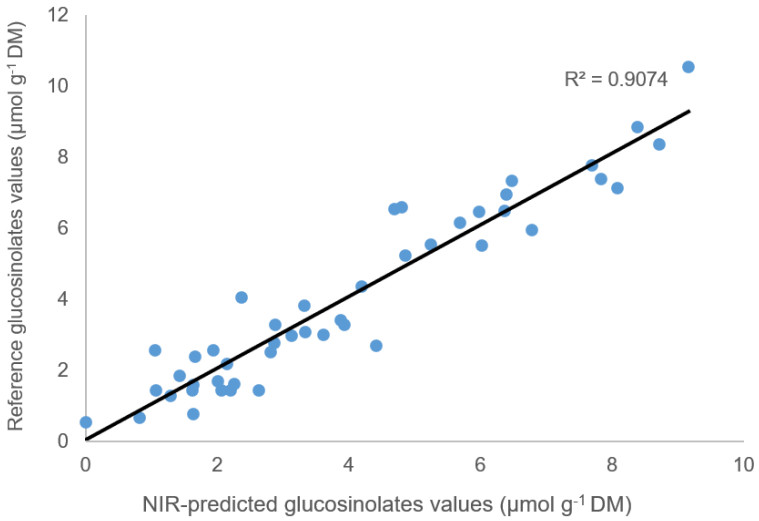

Fig 29. Correlation between reference and NIR-predicted values for validation sample set for total glucosinolate contents (µmol g^{-1} DM) of Model III

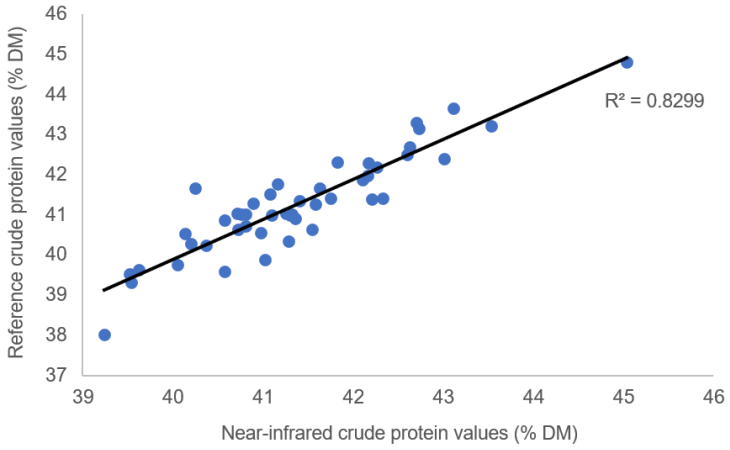

Fig 30. Correlation between reference and NIR-predicted values for validation sample set for crude protein (% DM) of Model III

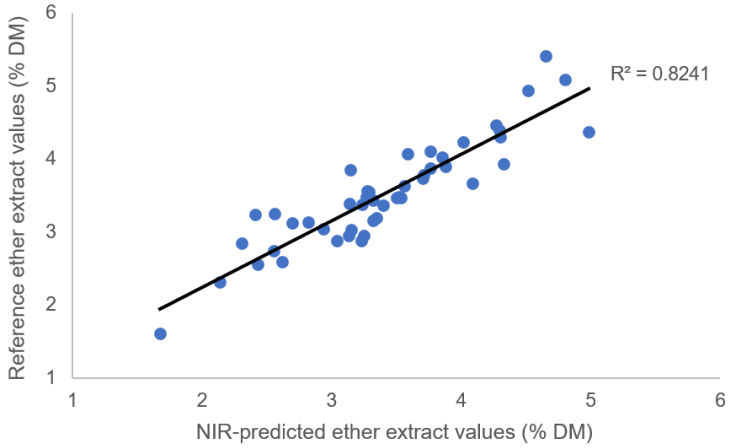

Fig 31. Correlation between reference and NIR-predicted values for validation sample set for ether extract (% DM) of Model III

5.2. DISCUSSION

5.2.1. Notable relationships between canola meal components

One of the aims of the study was to utilize the prediction of glucosinolate contents from NIR spectroscopy as an indirect indicator of protein quality and nutritive value in canola meal. Protein with a well-balanced amino acid composition is a crucial nutrient in canola meal. When comparing canola meal to soybean meal (SBM), it is important to note that canola meal contains less lysine and arginine but more methionine and cystine. Thus, it is advisable in monogastric animal nutrition to use them in combination for diet formulation as they both can be complementary, and to reduce feed costs (Khajali & Slominski, 2012). While amino acid contents are considered, they are not the sole indicators of canola meal quality. Instead, emphasis is placed on their digestibility, which can be influenced by the processing conditions during oil extraction. Excessive heat treatment during the pre-press solvent extraction of canola seeds might diminish the digestibility of amino acids, leading to variations in the nutritive value of canola meal. Pre-press solvent extraction has several steps involving heat such as pre-conditioning, cooking, pressing, solvent extraction, and desolventizing/toasting, all of which contribute to alterations in the chemical composition and nutritive value of the meal. The most significant impact of heat is observed during the desolventizing/toasting phase, where high temperatures are used for a prolonged duration to remove hexane from meal. The intensity and duration of heat applied during this processing step can lead to the damage of heat-sensitive amino acids, which may become biologically unavailable, or unreactive. Lysine is the most sensitive to heat treatment (Almeida et al., 2014); therefore, mild conditions should be possibly maintained to preserve its availability. The cold expelling process is deemed beneficial for the meal protein quality due to exposure to relatively low heat for a short period of time. It needs to be emphasized that lysine is one of the

most important amino acids for animals because it serves key metabolic roles in swine and poultry (Liao et al., 2015). As an indispensable amino acid, lysine is a regulator in muscle accretion and fat deposition in pigs and poultry (Wang et al., 2017). Losses in available lysine, arginine, threonine, aspartic acid, or glutamic acid due to early and advanced Maillard reactions are common in processed feed ingredients, and this phenomenon is not exclusive to canola meal but is also observed in DDGS, SBM and complete diets subjected to pelleting (Pahm et al., 2009). For example, Zhu et al., (2018) demonstrated a substantial variation in the standardized ileal digestibility of lysine within DDGS for broiler chickens, ranging from 29.00 to 84.83%. This diversity was mainly caused by the variability in processing conditions during DDGS production.

As the MRP formation following heat treatment indicated protein damage, it can be concluded in the chemical composition of canola meal as the MR contributed to increased dietary fiber, NDF, and lignin contents. Such relation was demonstrated in research conducted by Slominski (1997) that highlighted the effect of *in vitro* moist heat treatment on canola meal quality. Temperatures exceeding 105°C significantly increased the neutral detergent insoluble protein, which contained a highly indigestible fraction of proteins. Such increase contributed to the rise in TDF content, indicating protein damages due to overheating. The impact of the formation of NDICP in canola meal for amino acids availability was presented by Newkirk et al. (2000). They concluded that NDICP values below 10% of total crude protein indicated canola meal with more than 85% lysine availability for poultry. In the study by Adewole et al. (2016), the variation in chemical composition, fibre components, (i.e., NDF, TDF and NDICP) was observed among canola meals obtained from Canadian canola crushing facilities and it was attributed to differences in processing conditions. They demonstrated strong positive correlations between NDF and TDF ($R^2 = 0.97$), as well as between TDF and NDICP contents ($R^2 = 0.79$), which could corroborate

with the strong correlation between NDF and NDICP in this study ($R^2 = 0.73$). Moreover, they also connected the effect of processing on glucosinolates contents. The glucosinolates contents of canola meal from different crushing plants across Canada ranged from 2.0 to 10.1 μmol g^{-1} DM, which was similar to the concentration range of the sample set used to develop the NIR prediction models in this study. A very strong correlation between glucosinolates contents and both TDF ($R^2 = 0.95$) and NDF ($R^2 = 0.82$) observed in this study indicated a direct relationship; overheating of canola meal led to glucosinolates decomposition and a fiber content increase. The same relationship was observed by Adewole et al. (2016). The lysine damage due to heat treatment during desolventizing/toasting agreed with the negative relationships between lysine and NDF ($R^2 = 0.53$), lysine and TDF ($R^2 = 0.64$), and lysine and NDICP ($R^2 = 0.48$) allowing them to develop equations to predict the available lysine in canola meal and estimate the degree of protein damage. These results allowed them to correlate the glucosinolates and lysine availability as positive relationships were observed between their contents ($R^2 = 0.54$).

In the past, diets for monogastric animals were formulated based on the total amino acid requirement. For more precise nutrition, digestible amino acid contents are now preferred for this purpose. This approach allows for more efficient nutrient utilization, thus, reducing nitrogen excretion, and the cost of feed. However, measuring amino acid digestibility or relative bioavailability in any given feed ingredients requires extensive metabolic research with animals and involves laborious and costly analytical procedures. While *in vitro* methods may be used to estimate bioavailable lysine, they are limited to quantifying reactive lysine (Mauron & Bujard, 1964; Hurrell & Carpenter, 1981). However, these procedures are more relevant to evaluating the chemical properties of the feed ingredient rather than assessing their nutritive value. Additionally, in the *in vitro* evaluation, the quality of other amino acids would be overlooked. Therefore,

evaluation protein quality based on other chemical components like TDF, NDF, NDICP or glucosinolates could be used in assessment of potential protein damage resulting from heat treatment during feedstuff processing, including canola meal. Such relationship could be integrated into the prediction equations, with variables like glucosinolates or NDF, possibly estimated using NIR technology.

5.2.2. Near-infrared prediction models

5.2.2.1. Evaluation of prediction models

In this study, NIR prediction models were constructed to predict glucosinolate contents, crude protein, and ether extract in mash and pellet canola meal. Model I was developed using Canadian canola meal, RS meal of non-Canadian origin, and ECPC, however, its poor performance resulted in the subsequent development of Model II specific for mash canola meal, and Model III specific for pellet canola meal, respectively. Model I showed poor statistical performances for the total glucosinolates predictions for canola meal, and total glucosinolates and crude protein predictions of ECPC (R^2_{CM} = 0.0005 for glucosinolates and R^2_{ECPC} = 0.49 for glucosinolates, and R^2_{ECPC} = 0.61 for crude protein). On the other hand, the validation prediction data for the ECPC ether extract showed excellent performance (R^2 = 0.95, SEP = 0.58) as presented in Table 11. The poor validation statistical results associated with the good cross-validation statistical data (R^2 = 0.90, 0.91, and 0.97 for glucosinolates, crude protein and ether extract) indicated potential issues with the validation sample set (n = 50) made of only canola meal. The ratio of performance to deviation or RPD is a parameter that indicates how the model can predict future samples based on variation (SEP) of the model and the sample set (Table 14.). Table 11 showed a RPD of 3.51 for the 2012-2019 canola meal samples for the glucosinolate model. It decreased to 1.35 when tested

with the verification sample set made only of ECPC. The RPD for crude protein and ether extract were 1.55, and 4.46 for crude protein and ether extract, respectively, when the model was tested with the ECPC. At values below 2.3, the prediction equations developed for glucosinolates and crude protein were considered to have poor prediction potential (Table 11.). These findings were consistent with the poor correlation obtained for the linear relationship for glucosinolates and crude protein with R^2 of 0.49 and 0.61, respectively. Validation for glucosinolates with canola meal resulted in RPD = 3.51, indicating fair prediction of future samples. However, its extremely poor R^2 = (0.0005) ultimately implies Model I cannot be used to accurately predict glucosinolates content.

With R^2 = 0.95 and RPD = 4.46 (indicating fair quality), Model I can only be used to effectively predict ether extract content of ECPC. Achieving such high correlation for validation of Model I with ECPC was particularly surprising since most of the samples in the validation dataset fell outside the range of the concentrations of the calibration sample set. Model I's calibration samples were majority canola meal and contained few ECPC samples. It was reasonable to expect that the canola meal validation prediction to outperform the one of ECPC. However, Model I produced better predictability statistics with ECPC than with canola meal as shown by the validation statistics obtained for glucosinolate contents predictions; R^2 was higher with lower SEP, and RMSEP lower for ECPC than with canola meal (see Table 11.).

Expeller cold-pressed canola meal (ECPC) contains a much higher quantity of residual oil than the conventional canola meal. The second verification set was made essentially of ECPC samples, but Model I's predictability indicated a promising future for the use of Model I to predict ether extract for ECPC samples. It is also an example of a situation where verification data could fall outside the range of a prediction dataset due to environmental factors (Barthet et al., 2020).

Barthet et al., (2020) reported a similar situation when their models developed for intact canola samples collected over a period were tested on an independent sample set collected over a different time period. As what was observed with Model I, the various models were robust enough that they could be used to predict canola quality even if the verification data felt outside of the range of the data used in the calibration sample set. However, that it is not often the case and is not a recommended practice.

The ability to predict values outside the prediction dataset range gave Model I an interesting advantage to analyze future unknown samples as many factors (environmental, processing conditions, etc.) could potentially contribute to high variation in chemical components, possibly pushing the minima and maxima of the unknown samples outside the range of the ether extract prediction dataset. Model I will need to be verified periodically to confirm the ether extract predictions and/or improve its performances to maintain its reliability and robustness.

Sahamishirazi et al. (2017) also reported poor statistical results for their model to predict glucosinolate contents in broccoli suggesting the high variation across a smaller number of samples of the calibration set was the issue; the cross-calibration showed R^2 ranging from 0.25 to 0.76) while the R^2 for the independent verification sample set ranged from 0.03 to 0.55. Model I showed an excellent linear relationship with $R^2 > 0.90$ for the cross-calibration but poor statistical results when tested on the completely independent verification sample set (Table 11). This study also applied preprocessing to the spectra data to develop the prediction models, unlike that of Sahamishirazi et al., (2017), whose lack of preprocessing might have contributed to the poor results for their models. Model I's excellent prediction for the cross-calibration yet poor verification prediction statistics for the independent verification set (containing mostly canola meal samples not included in the calibration sample set) were likely due to the type of samples used in the

verification than the prediction models (i.e., Chinese RS meal). These "out-of-the norm" samples were eliminated from further model development and testing as there was not enough of them to include them as they could be considered as outliers, (i.e., rapeseed instead of canola).

Model I's poor prediction results with the verification sample set suggested that prediction models could be improved by gaining a better understanding of the sample type effect on the models. Although good results were obtained for ether extract prediction in ECPC, the objective of the project (i.e., developing a robust prediction model for glucosinolate contents in canola meal), was not achieved. A principal component analysis (PCA) allowed the detection of spectral similarities between samples while identifying the presence of outliers. The PCA showed two distinct groupings for the calibration and verification samples based on their extraction process, cold pressed meal, mash canola meal and pellet canola meal. In the canola industry, canola meal is formed into pellets to preserve nutrient quality, increase bulk density, reduce storage size as well as decreasing transportation costs and maximizing durability during transport (Huang et al., 2015). Pelleting canola meal reduced potential bacteria contamination to meet trade requirements for *Salmonella*-free commodity (Mathews Jr. et al., 2001). Pelleting also introduced additional heat and transformation to the canola meal, physically altering the meal particles (Huang et al., 2015). Model II was developed to predict the quality of mash canola meal while Model III was developed to predict the quality of pellet canola meal. The development of these two models proved to be successful, as their performance displayed drastic improvement in prediction when compared to the statistical results obtained for Model I. Improvements in prediction observed following the verification of Models II and III with their appropriate independent samples demonstrated that when proper steps were taken robust prediction models could be developed. Removing ECPC and non-Canadian samples, use of spectral averaging (section 5.2.2.3.), and segregating samples to

mash and pellet canola meal to develop two models were strategies to increase the sample homogeneity while achieving a continuous distribution of the various component concentrations of the canola meal samples for each of the parameters in Models II and III.

Model II provided excellent prediction results for glucosinolate contents, crude protein, and ether extract with R^2 of 0.89, 0.93 and 0.91, respectively and low standard error of calibration (SEC) of 0.93, 0.49 and 0.22, respectively for the cross-calibration (Table 12.). When the model was tested with the verification sample set, the statistical analysis showed a good linear relationship for glucosinolate and crude protein contents (R^2 = 0.87 and 0.91, respectively; Table 12.). The good correlation combined with minimal bias (0.00 and -0.09 for glucosinolates and protein, respectively) and low standard error prediction (0.90 and 0.66, for glucosinolates and protein, respectively) showed that the prediction model could be useful for the industry to predict glucosinolate to and crude protein contents of mash canola meal (Table 12.). Ether extract exhibited lower correlation (R^2 = 0.74) and very poor RPD (1.93) that could reduce the use of Model II for industrial application to predict ether extract content of mash canola meal. However, despite the discouraging RPD, bias was minimal (-0.02) and standard error of prediction was low (0.33), suggesting that the ether extract prediction for Model II could be improved in the future. Such results also implied the model for ether extract could be used in screening future samples where accuracy and precision may not be considered as crucial as in qualitative analysis.

Model III provided good statistical results when tested during the cross-calibration test for glucosinolates, crude protein, and ether extract contents (R^2 = 0.91, 0.90, and 0.95 respectively) and low standard error of calibration (SEC = 0.80, 0.40, and 0.16, respectively). When tested on the verification sample set with pellet canola meals gave good linear correlation for glucosinolate content (R^2 = 0.91; Figure 26.). Slightly lower linear correlations were observed for crude protein

and ether extract (R^2 = 0.83 and 0.82, respectively; Figures 26, and 27). The good linear correlation obtained the glucosinolates verification combined with low bias (0.07) and low SEP (0.79) indicated that Model III could be used for prediction of total glucosinolates in pellet canola meals. Predictions of crude protein and ether extract were not as good as for glucosinolates, however, models showed minimal bias (-0.11 and 0.12, respectively) and low standard error of prediction (SEP = 0.52 and 0.31, respectively) when tested on the verification set. Model III could be used to predict glucosinolates, crude protein and ether extract in pellet canola meal (Table 13.).

Although statistical evaluation (R^2 and RPD) is commonly used to indicate the overall quality of a prediction model and performance on future samples, respectively, Ali Redha et al., (2023) argued that assessing the underlying details about the samples and instrument are required by the operator to maximize the efficiency of the calibration and the operation of NIRS. Such details include the standard error (SE) of the reference values, limit of detection (LOD) and limit of quantification (LOQ) of the reference method, standard error of calibration (SEC) and prediction (SEP) for the prediction models as well as the characteristics of the sample (breed, processing, origin, form, etc.). In addition to the parameters described previously and in Table 5, other statistics could be used to evaluate prediction models. Although not as commonly used in literature, evaluating the SEC or SEP against the standard error of laboratory/reference values (SEL) could help to determine the model's precision. According to Shenk & Westerhaus (1996), models with a SEP/SEL ratio 1.0-1.5 could be considered of excellent precision while higher SEP/SEL ratios could indicate a decrease in precision quality (Table 15).

Table 14. Evaluation of prediction model and potential applications based on RPD values according to Williams & Norris (2001)

RPD	Classification	Potential Application
0.0 – 2.3	Very poor	Not recommended
2.4 – 3.0	Poor	Very rough screening
3.1 – 4.9	Fair	Screening
5.0 – 6.4	Good	Quality control
6.5 – 8.0	Very good	Process control
8.1 - ∞	Excellent	Any application

Table 15. Prediction model precision rankings based on SEP/SEL ratio according to Shenk & Westerman (2018)

SEP/SEL ratio	Precision
1.0 – 1.5	Excellent
2.0 – 3.0	Good
4.0 – 5.0	Moderate
5.0 - ∞	Poor

Table 16. Calculations of SEP/SEL ratio for glucosinolates, crude protein and ether extract in Models II and II from standard error of prediction and standard error of laboratory values

Model	Component	Standard error of prediction (SEP)	Standard error of laboratory (SEL)	SEP/SEL ratio
Model II (Mash)	Glucosinolates	0.90	0.25	3.6
	Crude Protein	0.66	0.27	2.4
	Ether extract	0.33	0.08	4.1
Model III (Pellet)	Glucosinolates	0.79	0.37	2.1
	Crude Protein	0.52	0.18	2.9
	Ether extract	0.31	0.10	3.1

In this project, the prediction model for protein of the mash canola meal, and the models for all components (glucosinolates, crude protein and ether extract) of the pellet canola meal exhibited good precision SEP/SEL ratio (2.4, 2.1, 2.9, and 3.1, respectively). Glucosinolates and ether extract prediction models for the mash canola meal showed moderate precision SEP/SEL ratio (3.6 and 4.1, respectively) (Table 16.). The precision determined by SEP/SEL ratio was in agreement with the statistical results showing the robustness of the developed models (Table 16.). In Model II, crude protein exhibited the highest R^2 and RPD for cross-validation and the verification of the model, followed by glucosinolates and ether extract. The SEP/SEL ratio also followed in this pattern, as crude protein showed the lowest value, indicating highest precision of all the developed equations, followed by glucosinolates and ether extract. A similar Model III showed a similar pattern, with the best SEP/SEL ratio observed for the components showing the highest R^2 and RPD for the verification for glucosinolates, crude protein and ether extract (Tables 13.). These observations suggested that the SEP/SEL ratio could be an additional statistical factor to evaluate the data relevant to NIR model quality as some statistical parameters (i.e., R^2 and RPD) could give contradictory indications for model prediction quality.

5.2.2.2. Effect of spectral data selection and characteristic spectral absorbance

Spectral data (or wavelength) selection can contribute to the improvement of the NIR prediction models as specific spectral regions can be characteristic or fundamental to identify, qualify and/or quantify compounds. On the other hand, spectral regions unrelated to the compounds may be at best uninformative or worse, generate noise when used in the prediction model. During the development of the models, spectral regions with information related to glucosinolates, crude protein, and ether extract were retained for their prediction models whereas irrelevant spectral data were neglected during the optimization of the scan processing.

Table 17. Spectral regions for each component retained for the prediction models developed for Models II and III.

Component	Model II (mash canola meal)		Model III (pellet canola meal)	
	cm^{-1}	nm	cm^{-1}	nm
Protein	9400 – 5448 4600 – 4248	1064 – 1835 2174 – 2354	9404 – 6094 5454 – 4243	1063 – 1641 1833 – 2357
Ether Extract	9400 – 4600	1064 – 2174	9404 – 5446 4605 – 4243	1063 – 1836 2171 – 2357
Glucosinolates	7504 – 5448	1333 – 1836	9404 – 4243	1063 – 2357

Table 17. summarizes the spectral regions retained to develop the various models. Spectral selection was made based on the absorption regions of the common bonds within each component. Protein is rich in nitrogen and possesses a complex structure. The spectral bands associated with proteins are located throughout the entire NIR range, therefore, the excluded NIR regions when developing the protein models were minimal (Burns & Ciurczak, 2008). These wavelength regions depend on the vibrations (usually stretching or bending) of the nitrogenous bonds of the protein structure (amide, carboxylic group, aromatic hydrocarbons, etc.). The regions used Models II and III are consistent with that of Daszykowski et al., (2008) used to predict crude protein in rapeseed meal. Their study identified several regions between 1200 and 2500 nm considered relevant in modelling for protein, including the 1410-1500 nm region where key absorbances were located. One of the most important is the 1500 nm region associated with the N-H stretch second overtone correlated to the peptide backbone. It is regarded as the one of the most identifiable and characteristic bands for measuring the content of protein with NIRS (Daszykowski et al., 2008). At 1900 nm onward, the response linked to the nitrogenous groups decreased while the response linked to the hydrocarbons and carboxyl groups increased. The mash model excluded the 1836 -

2174 nm region, while the pellet model excluded the 1641-1834 nm region. Models II and III used other wavelengths considered important to predict protein content, including the 1640-1680 nm and 2168-2180 nm regions associated with aromatic protein structures, and N-H bend second overtones, respectively (Burns & Ciurczak, 2008).

Palmitic (C16:0), oleic (C18:1), linoleic (C18:2), and linolenic (C18:3) acids are long-chain fatty acids that make up majority of the fatty acid composition in canola meal (Zambiazi et al., 2007; Semwogerere et al., 2019). Near-infrared absorption regions for the atoms that compose the hydrocarbon chains are located consistently throughout the entire spectral region (700 nm – 2500 nm). Majority of the spectrum was retained (1064 – 2174 nm) for the mash model's ether extract prediction equation, while the pellet model's ether extraction prediction equation was optimized when the region between 1836 and 2171 nm was removed. Since this region did not contain information on the C-H bonds, its removal from the model likely eliminated undesired noise from the ether extract prediction model for pellet canola meal. This agreed with Daszykowski et al. (2008)'s prediction model for fat content. Both the mash and pellet canola meal models retained the wavelengths around 1740 nm and 2300 nm. The absorption region around 1740 nm was of high importance in ether extract prediction models as it is correlated with symmetric and asymmetric C-H stretching first overtone of the methylene groups of the fatty acid chain.

Even if they are small chemical components, glucosinolates possess a very complex structure. Variability in the side chain greatly contributes structural diversity of the glucosinolates structure (Font et al., 2005). The complex network and diversity of chemical bonds in its structure likely contributed to challenges often encountered when developing prediction models for glucosinolate contents. Several authors agreed upon the existence of key bond vibrations (i.e., stretching or bending modes of O-H, N-H, C-H and S-H) within the functional groups of the

glucosinolates structures, yet the wavelengths where such specific vibrations occurred in the spectra are often debated (Ali Redha et al., 2023). Prediction models for glucosinolates in several studies have selected different wavelength ranges for specific chemical bonds in its structure (Font et al., 2005; Toledo-Martin et al., 2018; Renner & Fritz, 2020; Ali Redha et al., 2023). The optimal absorption regions for glucosinolates for Kumar et al., (2010) were between 1333-1837 nm and 2173-2355 nm. These regions capture the O-H stretching (1430-1440 nm and ~1921 nm), the N-H stretching of the indole moiety (1450 nm), the S-H stretch first overtone (~1695 nm), the C-H stretch first overtone of the methyl groups, and the C-H stretching by methylene groups (1730-1764 nm). Other studies also mentioned that relevant spectral regions for glucosinolates included the 950-1650 nm, 1450 nm, and 2000-2500 nm regions (Renner & Fritz, 2020; Ali Redha et al., 2023). Quantification of glucobrassicin has been performed using the 950-1650 nm region – the first overtone of O-H stretching (1420 – 1425 nm) and first overtone of indolyl N-H asymmetric stretching (1450 nm) – generating successful prediction models (Renner & Fritz, 2020). Hernandez-Hierro et al., 2012 demonstrated that the 2000-2500 nm region consisted of intricate characteristics linked to the stretching of N-H, O-H and C-C, and was most effective in quantifying glucobrassicin, neoglucobrassicin and total glucosinolate contents. Although Models II and III shared many common regions reported in these publications, the selection regions showed some differences. Model III included the 2000-2500 nm region used by Hernandez-Hierro et al., (2012) and Kumar et al. (2010) used to predict total glucosinolates. Model II's region did not exceed 1836 nm and still managed to generate a good correlation ($R^2 = 0.87$, Figure 23.) when the model was tested on the independent verification sample set. Attempts were made to select several spectral regions of shorter wavelength ranges consisting of prominent peaks, however, this strategy

resulted in subpar calibration prediction as too many overlapping bands were selected, resulting in unclear spectral information for the prediction model to predict with high accuracy.

5.2.2.3. Effect of spectral averaging

The improvement in prediction from Model I to Models II and III were achieved based on reducing the variability within Model I. As the prediction dataset contained pellet, mash and expeller cold-pressed canola, there may have been spectral variability present and a lack of even distribution for each parameter in Model I. For example, very little samples of expeller-cold pressed canola in Model I contributed to uneven distributions of samples for glucosinolates and ether extract. Expeller cold-pressed canola had higher glucosinolate contents due to lack of heat treatment during processing when compared to conventional canola meal. Pressing of canola seed also generated higher residual oil contents in expeller-cold pressed canola meals when compared to the conventional prepress solvent extraction group. Distribution of prediction set samples for glucosinolates (Figure 13.) in Model I were skewed toward the lower tail of the glucosinolate contents. The same unbalanced distribution was also present in the samples when considering the ether extract contents of the various canola meal as shown in Figure 15. Spectral averaging in Models II and III increased the sample size, increased the underrepresented values in the prediction datasets and made the concentration distribution for each component more homogenous. Model II for oil, protein and glucosinolates were developed using 199 scans made of 107 sample scans plus 92 averaged spectra), 196 scans made of 109 sample scans plus 87 averaged spectra, and 278 scans made of 184 sample scans plus 94 averaged spectra, respectively. Model III for oil, protein and glucosinolates were developed using 145 scans made of 77 sample scans plus 68 averaged spectra, 148 scans made of 80 sample scans plus 68 average spectra, and 149 scans made of 81 sample scans plus 68 averaged spectra, respectively. This strategy of adding artificial samples of

less represented values by averaging both spectra data and reference values of extreme samples, contributed to improved prediction models for glucosinolates, crude protein and ether extract, as indicated by the statistical data obtained during the verification (Tables 11-13.).

5.2.2.4. Challenges to overcome

Although many obstacles were overcome throughout the procedure of the project, it is reasonable to face similar challenges when developing a robust prediction model. One of the most troublesome obstacles in prediction model development is sampling error, particularly related to sample homogeneity. Sampling effect may present issues in spectral acquisition but has a greater influence in the wet chemistry use to acquire the reference values. Grinding to the same particle size and sample homogenization could reduce the sample effect when acquiring the spectral data. High errors in reference data analysis could induce greater sampling effects leading into issues with accuracy and precision for the prediction model. This emphasized the importance of thoroughly mixing the sample before the analysis and executing good laboratory practices (GLP) to minimize errors. Specific to glucosinolate analysis, one of the major difficulties experienced in this project was the limit of detection (LOD) and limit of quantification (LOQ) related to samples of low concentrations. Reference methods with chromatographic techniques offered sufficient sensitivity to identify and quantify total glucosinolates as low as 0.10 μmol g^{-1}. However, predictions arising during prediction model development may not reflect the reference values since the LOD of NIR was 1.0 μmol g^{-1} for glucosinolates. Ali Redha et al. (2023) proposed to overcome this issue by identifying individual glucosinolates that existed in low quantities and calibrating for their total concentration (i.e., summing all indole glucosinolates). However, this may only be effective for rapeseed meal as low-glucosinolates canola often lacks the high quantities of indole glucosinolates. It was important to recognize that some individual glucosinolates and other

chemical components (i.e., NDICP) may not be accurately reflected by NIRS. As presented in section 5.2.2.2., there were numerous spectral regions consisting of important bands correlated with glucosinolates that were not used in developing the model. However, publications related to prediction model development did not have a standard set of wavelengths that should be included in the developed prediction models. Ali Redha et al., (2023) mentioned the formation of glucosinolates from their amino acid precursors that could lead to poor prediction ability as a result of indirect correlations that were observed. Provided there were more standardized wavelengths that could correlate with the characteristic bands of glucosinolates in NIR spectra, prediction model development could be made easier with respect to the spectral regions. Likewise, the lack of standardized or optimal numbers of samples that are required to develop prediction models and for the validation sample set was also an issue. These numbers are often ambiguous and relies on judgement of the researcher leading to variation in sample size across studies. Furthermore, there is no "universal recipe" that exists for preprocessing the spectral data (Jiao et al., 2020). While the application of preprocessing is largely dependent on data at hand, there are no standardized method based on type of instrument, or type or form of sample. This lack of standardization could contribute to variability in spectral data of the same sample selected for the models and complicate tasks such as transferring calibration models between instruments.

Studies are available in literature related to NIR calibration of glucosinolates in rapeseed meal, however, the same studies for canola meal are limited. This is possibly due to the higher content of glucosinolates occurring in rapeseed meal (as high as 120 μmol g^{-1}) compared to that of canola meal (by definition, below 30 μmol g^{-1}). Challenges begin to surface in a prediction model for those samples with limited quantities of analytes. An important consideration in NIR prediction model development is for the dataset to be large and with a wide range and even

distribution for the analytes amongst all samples. This study provided 432 total samples that was available to develop and validate the prediction equations, however, distribution was an issue especially for glucosinolates content. Majority of the sample set consisted of low-glucosinolates canola meal, with a smaller number of high-glucosinolate RS meal and ECPC samples; the component distributions were skewed toward the lower end for glucosinolates and ethyl ether. While this was rectified using spectral averaging, it highlighted the difficulty in developing NIR prediction models for glucosinolate contents due the low quantities in canola meal. The NIR limit of detection was also an issue for glucosinolate prediction. We found that the models with the Bruker Matrix-I instrument had a limit of detection of 1.0 μmol g^{-1} for glucosinolate content; all samples with content lower than 1.0 μmol g^{-1} were predicted as zero. Future model development should avoid such samples in the prediction dataset at the expense of sample size. Once again, this issue could be corrected through spectral averaging. However, this will pose a major challenge when it comes to predict unknown canola meal samples with glucosinolates lower than 1.0 μmol g^{-1}.

While the general procedure to develop a prediction model is basic and straightforward (see Figure 4.), it lacks a "universal recipe" that includes techniques which played a vital role in model development. Principal component analysis (PCA) can be used to detect outliers. PCA confirmed the presence of RS meal, ECPC and canola meal in a dataset, and identified canola meal samples processed as mash or pellet. The Bruker software did not allow the PCA, it has to be run using other statistical software. The PCA analysis of the canola meal provided the option of developing two models to optimize the goal of predicting canola meal quality. With such an important role in this study, it is highly recommended that all projects include a PCA to develop of prediction models to (i) detect outliers at an earlier stage, (ii) group potential homogenous

samples that would be suitable in a prediction dataset, and (iii) identify any outstanding differences that may negatively influence the model.

5.2.3. Indirect use of near-infrared predicted data to assess canola meal quality

As of the 1990s, Bell (1993) reported a shift away from glucosinolates to other nutrients, potentially impacting utilization of canola meal in animal nutrition. The concept of using glucosinolates content as an indirect indicator of canola meal quality is a novel, and there is limited literature related to the use glucosinolates to assess and quantify other chemical components (Adewole, 2017). Nevertheless, the link between low levels of glucosinolates in canola meal and improved apparent metabolizable energy (AME_n) and nutrient digestibility has been earlier suggested (Bell, 1993). It has to be emphasized that determining of the nutrient digestibility requires a laborious and expensive studies with poultry or swine, making obtaining such data challenging. Associations between glucosinolates and other chemical components have been proposed, giving light to using the concept of glucosinolates as an indicator of canola meal quality (Slominski et al., 1999; Newkirk, 2002). Although the conventional seed processing proved to be successful in degrading glucosinolate contents, lysine bioavailability dropped by 8-15% with thermal treatment at 105°C and higher (Keith & Bell, 1984). Such results were consistent with the findings of Almeida et al., (2014), in which autoclaving canola meal at temperatures above 130°C for 45 minutes gave a consistent decrease in reactive lysine and reactive lysine-to-CP ratio and increased in NDF and insoluble nitrogen. Additionally, complete degradation of glucosinolates was observed after 20 minutes. Thus, glucosinolates' relationship with other quality parameters were established as part of this project. Research conducted Slominski (1997) also demonstrated that excessive heat treatment of canola meal significantly increased its dietary fibre content by bounding the protein.

Relationships were established between glucosinolates and NDICP, NDF and TDF, as well as between NDICP and NDF and TDF in canola meal, respectively. The formation of NDICP or glycoproteins subsequently increases the NDF and TDF content in canola meal. Thus, a relationship between NDICP and NDF, and TDN could be anticipated. The relationship can also be explained in the mash and pellet canola meal, and ECPC that were used to construct Models I-III. Neutral detergent insoluble crude protein content was lower in ECPC (4.15%) compared to mash (5.12%) and pellet solvent-extracted canola meal (5.52%), supporting the concept that protein damage due to heat and moisture in solvent extracted canola meal contributed to the MRP formation, and subsequent inflation in NDF (24.08% in ECPC, 29.14% in mash, and 30.20% in pellet canola meal) and TDF (31.70% in ECPC, 38.20% in mash, and 38.88% in pellet canola meal). The stronger correlation ($R^2 = 0.73$) between NDICP and NDF could explain the increase in crude protein within the NDF residue from MR. Ideally, these correlations should be much higher, as both NDF and TDF increased directly due to NDICP formation from the Maillard reactions. The robust correlation between glucosinolates and TDF ($R^2 = 0.95$), and glucosinolates and NDF strongly suggests that glucosinolate content could be an indirect indicator of both NDF and TDF in canola meal.

The negative correlations demonstrate the decrease in glucosinolates content shared a linear relationship with increasing NDF and TDF. The heat and moisture during canola processing led to both glucosinolates degradation and the Maillard reactions to generate NDICP formation and subsequently increased NDF and TDF. The degradation rate was important for glucosinolates content, it was higher for mash (3.37 μmol g^{-1} DM) and pellet canola meal (3.80 μmol g^{-1} DM) than ECPC (7.88 μmol g^{-1} DM). However, the correlation between glucosinolates and NDICP was only moderate ($R^2 = 0.61$) and should have been stronger if glucosinolates content was directly

related to NDICP content via heat and moisture degradation. The R^2 (0.61) suggested that this regression equation would not be recommended to predict NDICP from total glucosinolate contents. Despite this, most of the data points fell within an acceptable range of the trend line with only one point appearing as a potential outlier in the linear relationship between glucosinolates and NDICP (Figure 6.). Based on this observation, there is a potential to use glucosinolates as an indirect indicator of NDICP in future models to canola meal quality for the feed industry.

Similar results were obtained when establishing chemical component relationships among 24 ECPC, with 16 of them used in validating Model I. Neutral detergent insoluble crude protein demonstrated a moderate correlation with TDN ($R^2 = 0.62$) but strong correlation with NDF ($R^2 = 0.79$).

5.2.4. Potential prediction model for individual glucosinolates

Developing prediction models for individual glucosinolates was considered as an option to improve Model I. Gohain et al., (2021) reported excellent correlation for gluconapin ($R^2 = 0.95$), hydroxy-brassicin and sinigrin ($R^2 = 0.99$) for *Brassica juncea* seed calibration model on individual glucosinolates using NIRS. Similarly, Chen et al., (2014) achieved high correlation for glucobrassicin ($R^2 = 0.93$), neoglucobrassicin ($R^2 = 90$), and gluconapin ($R^2 = 0.82$) in their investigation involving Chinese kale. However, the canola meal samples used in this study originated from low-glucosinolates canola seed by-product, containing neoglucobrassicin, glucobrassicin, and gluconapoleiferin at concentrations well under 1.0 μmol g^{-1}. This highlighted detection limits inherent in NIRS technology. In addition, the individual glucosinolates profile can vary among Brassica species which acquired a canola meal quality. Considering these factors, the exploration of calibrating for individual glucosinolate species was not pursued.

CHAPTER 6: SUMMARY, CONCLUSIONS AND FUTURE RESEARCH

Three NIR prediction models were constructed for glucosinolates, crude protein and ether extract. Model I, constructed with Canadian canola meal, non-Canadian rapeseed meal and ECPC, was developed for prediction of glucosinolates (n = 227), crude protein (n = 258) and ether extract (n = 254). While it displayed a poor prediction for glucosinolates and crude protein in ECPC, as well as glucosinolates in canola meal, it demonstrated excellent predictive ability for ether extract in ECPC during validation. Therefore, Model I could be utilized solely in predicting ether extract in ECPC. Its inadequacy in predicting glucosinolates and crude protein makes it less suitable for industrial use, however, could motivate efforts to further refinement of the model. Model II, constructed with mash canola meal, was developed for glucosinolates (n = 278), crude protein (n = 196), and ether extract (n = 199), offering good predictions for glucosinolates and crude protein, but only adequate predictability for ether extract. Model III, constructed with pellet canola meal, was developed for glucosinolates (n = 149), crude protein (n = 148), and ether extract (n = 145), demonstrating good predicting performance for all three components. Noteworthy correlations were established in the chemical components of canola meal, particularly an excellent correlation between glucosinolates and TDF. Additionally, good correlations were observed between glucosinolates and NDF, as well as NDICP and NDF.

The improved predictability of the two models, calibrated using both mash and pellet forms of canola meal, strongly indicated the promising application of this technology in the livestock feed industry. Model I could be recommended for its excellent ability to predict fat in ECPC. However, the development of Models II and III, which resulted from disappointing predictions in glucosinolates and crude protein contents by Model I, has shown advancement in predicting these compounds. As a result, these enhanced models could be effectively employed for predictions by

the industry. The division of Model I into Models II and III underscored the needs of designing specific prediction models for meals destined for further processing like pelleting or expelling. This emphasizes the necessity to evaluate both the type and degree of processing applied to canola meal samples before construction a NIR prediction model. Failure to address the excessive variation may lead to insufficient predictability within the model.

Accurate assessment of protein quality in canola meal is critical for its effective utilization as a protein source in animal nutrition. To enhance the efficacy of prediction model, future research should focus on strengthening the calibration by adding more samples. This would aid more balanced distribution, reducing the need for spectral averaging. Given the continuous variation in chemical composition of canola, incorporating newer samples into the prediction model will diminish discrepancies between the prediction dataset and the samples used for analysis. Setting reasonable performance objectives for the future, the goal should be to reach $R^2 > 0.90$ and RPD > 5.0 for all components in both mash and pellet canola meal prediction models.

The goal of developing robust predictions for glucosinolates, crude protein and ether extract was achieved through developing two models based on the post-processing of canola meal. Future research efforts should prioritize refining Model I to improve its predictive accuracy specifically for canola meal. Additionally, employing a single model could be advantageous at the feed mill level compared to managing multiple models. Using a singular model also eliminates a need for additional sample information, such as identifying whether the samples are in the mash or pellet form, simplifying the loading of the appropriate model.

Additionally, it is advisable to assess the transferability of the developed models. In scenarios where a new instrument replaces the existing one, having a transferable model offers a

significant advantage. This allows users to avoid the necessity of developing an entirely new prediction model.

NIRS technology has already significantly enhanced canola analysis in Canada and will continue to benefit the agri-food industry. The ongoing enhancement of prediction models will substantially benefit the efficiency of assessing the quality of Canada's most important oilseed by-product for animal nutrition.

CHAPTER 7: LITERATURE CITED

Adewole DI, Rogiewicz A, Dyck B, Slominski BA. 2016. Chemical and nutritive characteristics of canola meal from Canadian processing facilities. Anim Feed Sci and Tech. 222:17-30

Adewole DI. 2017. The effect of pre-press solvent extraction conditions on the chemical composition and nutritive value of canola meal for broilers chickens and pigs. University of Manitoba Faculty of Graduate Studies. Winnipeg (MB): University of Manitoba.

Agussabti R, Satriyo P, Munawar AA. 2020. Data analysis on near infrared spectroscopy as a part of technology adoption for cocoa farmers in Aceh province, Indonesia. Data in Brief 29:105251

Ali Redha A, Torquati L, Langston F, Nash GR, Gidley MJ Cozzolino. 2023. Determination of glucosinolates and isothiocyanates in glucosinolate-rich vegetables and oilseeds using infrared spectroscopy: A systematic review. Crit Rev in Food Sci and Nut.;1-17

Aljahdali N & Carbonero F. 2019. Impact of Maillard reaction products on nutrition and health: current knowledge and need to understand their fate in the human digestive system. Crit Rev in Food Sci and Nut. 59:(3)474-487

Almeida FN, Htoo JK, Thomson J, Stein HH. 2014. Effects of heat treatment on the apparent and standardized ileal digestibility of amino acids in canola meal fed to growing pigs. Ani Feed Sci and Tech. 187:44-52

Anderssen HR and Sorensen H. 1985. Double low rapeseed meal in diets of young bulls. Advances in the production and utilization of cruciferous crops. In Proceedings of the Seminar in the CEC Programme of Research on Plant Improvement:208-217

Anderson-Hafermann JC, Zhang Y, Parsons CM. 1993. Effects of processing on the nutritional quality of canola meal. Poultry Sci. 72:(2)326-333

AOAC International. 2005. Official Methods of Analysis of AOAC International. 18th ed. Gaithersburg (MD); AOAC International

AOAC International. 2023. Official Methods of Analysis of AOAC International. 22nd ed. New York (NY); AOAC International

Aripin NFB and Surugau N. 2016. Effects of temperature and pH on myrosinase activity and gluconasturtiin hydrolysis products in watercress. Trans on Sci and Tech. 3:(2)449-454

Barthet VJ, Petryk MWP, Siemens B. 2020. Rapid nondestructive analysis of intact canola seeds using a handheld near-infrared spectrometer. J Am Oil Chem Soc. 97:(6)577-589 Bell JM. 1993. Factors affecting the nutritional value of canola meal: A review. Can J Anim Sci. 73:679-697

Bell JM and Keith MO. 1991. A survey of variation in the chemical composition of commercial canola meal produced in Western Canadian crushing plants. Can J Anim Sci. 71:469-480

Bell JM, Keith MO, Hutcheson DS. 1991. Nutritional evaluation of very low glucosinolate canola meal. Can J Anim Sci. 71:(2)497-506

Bell JM. 1993. Factors affecting the nutritional value of canola meal: A review. Can J Anim Sci. 73:679-697

Berzaghi P and Riovanto R. 2009. Near-infrared spectroscopy in animal science production: principles and applications. Ital J Anim Sci. 8:(3)39-62

Bin J, Yan X, Liu R, Kang C, Chen Y. 2023. Calibration transfer for near-infrared (NIR) spectroscopy based on local preserving projection. Vib Spec. 126:103538

Brown D. 2018. Glucosinolates (Goitrogenic Glycosides). Ithica (NY). Cornell University College of Agriculture and Life Sciences; [Assessed 2023 May]. https://poisonousplants.ansci.cornell.edu/toxicagents/glucosin.html

Burns DA and Ciurczak EW. 2008. Handbook of near-infrared analysis. 3rd ed. Boca Raton (FLA): CRC Press

Canadian Grain Commission. 2022. Quality of western Canadian canola 2021. Winnipeg (MB): Canadian Grain Commission research and data for grain harvest and export quality; [assessed 2023 June]. https://www.grainscanada.gc.ca/en/grain-research/export-quality/oilseeds/canola/2021/08-glucosinolate.html

Canola Council of Canada. 2019a. Canola meal feed guide: industry feed guide. 2nd ed. Winnipeg (MB): Canola Council of Canada

Canola Council of Canada. 2019b. Our plan to grow the Canadian canola industry. Winnipeg (MB): Canola Council of Canada; [accessed 2023 June].https://www.canolacouncil.org/about-us/strategic-plan/

Canola Council of Canada. 2022. Canola industry: at a glance. Winnipeg (MB): Canola Council of Canada's market & statistics; [accessed 2023 June]. https://www.canolacouncil.org/markets-stats/

Canola Council of Canada. 2023. Explore our 50-year history. Winnipeg (MB): Canola Council of Canada's interactive timeline; [accessed 2023 February]. https://www.canolacouncil.org/canola-history/

Cherian G. 2020. A guide to the principles of animal nutrition. Corvallis (OR): Pressbooks

Chen J, Li L, Wang S, Tao X, Wang Y, Sun A. 2014. Assessment of Glucosinolates in Chinese kale by Near-Infrared Spectroscopy. Intl J of Food Prop. 14:(8)1668-1679

Comer JP. 1956. Semi-quantitative specific test paper for glucose in urine. Analytical chem. 28:1748-1750

Daxenbichler ME and Van Etten CH. 1977. Glucosinolates and derived products in cruciferous vegetables. Identification of organic nitriles from cabbage. J Ag and Food Chem. 25:(1)121-12

Daszykowski M, Wrobel MS, Czarnik-Matusewicz H, Walczak B. 2008. Near-infrared reflectance spectroscopy and multivariate calibration techniques applied to modelling the crude protein, fibre and fat content in rapeseed meal. Analyst. 133:(11)1523-1531

Davies T. 2011. Happy 90th birthday to Karl Norris, Father of NIR Technology. NIR News. 22:(4)3-16

De Jonge LH and Jackman FS. 2013. The feed analysis laboratory: establishment and quality control. Setting up a feed analysis laboratory and implementing a quality assurance

system compliant with ISO/IEC. Rome (ITA). Food and Agriculture Organization of the United Nations

Englyst HN and Cummings JH. 1984. Simplified method for measurement of total non-starch polysaccharides by gas-liquid chromatography of constituent sugars as alditol acetates. Analyst 109:(7)937-942

Fearn, T. 2001. Standardisation and calibration transfer for near infrared instrumentals: a review. J Near Infrared Spec 9:229-244

Feng DY and Zuo JJ. 2015. Nutritional and anti-nutritional composition of rapeseed meal and its utilization as a feed ingredient for animal. Feed and Industrial Raw Material: Feed. 73:(4)689-697

Folch-Fotuny A, Vitale R, de Noord OE, Ferrer A. 2017. Calibration transfer between NIR spectrometers: new proposals and a comparative study. J Chem. 31:(3)2874

Font R, del Rio-Celestino M, Cartea E, de Haro-Bailón A. 2005. Quantification of glucosinolates in leaves of leaf rape (*Brassica napus* ssp. *Pabularia*) by near-infrared spectroscopy. Phytochem. 66:(2)175-185

Franco Nogueira FR, Sakomura NK, Balbino Leme B, de Paula Reis M, Kochenborger Fernandes JB, da Silva Viana G. 2022. Lysine and arginine requirements of male and female broiler in the starter, grower and finisher phase. Ani Feed Sci and Tech. 283:115174

Gohain B, Kumar P, Malhotra B, Augustine R, Pradhan A, Bisht NC. 2021. A comprehensive Vis-NIR equation for rapid quantification of seed glucosinolate content and composition for diverse Brassica oilseed chemotypes. Food Chem. 354:129527

Gunzler H and Williams, A. 2001. Handbook of analytical techniques. 1st ed. Weinheim (GER): Wiley-VCH

Hall MB. 2014. Feed analyses and their interpretation. Vet Clin Food Anal. 30:(3)487-505

Heberger, K. 2008. Chapter 7 – Chemoinformatics – multivariate mathematical statistical methods for data evaluation. Med App of Mass Spec. 1st ed. Amsterdam (NED): Elsevier

Hernández-Hierro JM, Valvarde S, Villacreces S, Reilly K, Gaffney M, Gonzalez-Miret ML, Heredia FJ, Downey G. 2012. Feasible study on the use of visible-near-infrared Spectroscopy for the screening of the individual and total glucosinolates contents in broccoli. J Ag and Food Chem. 60:(30)7352-7358

Herschel W. 1800. Experiments on the refrangibility of the invisible rays of the sun. Phil Trans of the Royal Soc of London. 90:284-292

Huang X, Khan NA, Zhang X, Yu P. 2015. Effects of canola meal pellet conditioning temperature and time on ruminal and intestinal digestion, hourly effective degradation ratio, and potential nitrogen to energy synchronization in dairy cows. J Dairy Sci. 98:(12)8836-8845

Hurrell RF and Carpenter KJ. 1981. The estimation of available lysine foodstuffs after Maillard reactions. Prog. Food Nutr Sci 5:159-176

International Organization for Standardization 12099:2017. Animal feeding stuffs: Guidelines for the application of near infrared spectroscopy.

Jiao Y, Li Z, Chen X, Fei S. 2020. Preprocessing methods for near-infrared spectrum. J of Chemo. 34:(11)1-19

Karbasi M and Madadlou A. 2018. Interface-related attributes of the Maillard reaction-born glycoproteins. Crit Rev in Food Sci and Nut. 58:(10)1595-1603

Khajali F and Slominski BA. 2012. Factors that affect the nutritive value of canola meal for poultry. Poultry Sci. 91:(10)2564-2575

Killner MHM, Rohwedder JJR, Pasquini, C. 2011. A PLS regression model for on-line monitoring of the biodiesel production reaction. Fuel. 90:(11)3268-3273

Keith MO and Bell JM. 1984. Effects of ammoniation of canola (low glucosinolate rapeseed) meal on its nutritional value for the rat. Can J Anim Sci. 64:(4)997-1004

Khwatenge CN, Kimathi BM, Taylor-Bowden T, Nahashon SN. 2020. Expression of lysine-mediated neuropeptide hormones controlling satiety and appetite in broiler chickens. Poult Sci. 99:1409-1420

Kim BG, Kil DY, Zhang Y, Stein HH. 2012. Concentrations of analyzed or reactive lysine, but not crude protein, may predict the concentration of digestible lysine in distillers dried grains with solubles fed to pigs. Amer Soc of Anim Sci. 90:(3)3798-3808

Kumar S, Chauhan JS, Kumar A. 2010. Screening for erucic acid and glucosinolate content in rapeseed-mustard seed using near infrared reflectance spectroscopy. J Food Sci Technol. 47:(6)690-692

Lee JW, Kim IH, Woyengo TA. 2020. Toxicity of canola-derived glucosinolate degradation products in pigs – a review. Anim. 10:(12)2337

Lee LC, Liong CY, Jemain AA. 2017. A contemporary review on data preprocessing (DP) practice strategy in ATR-FTIR spectrum. Chemo and Intel Lab Sys. 163:64-75

Leeson S, Atteh JO, Summers JD. 1987. The replacement value of canola meal for soybean meal in poultry diets. Can J Anim Sci. 67:(1)151-158

Leiva SF, Sandoval JL, Abascal-Panciano GA, Flees JJ, Calderon AJ, Pacheco WJ, Starkey CW. 2022. Improper sample preparation negatively affects near infrared reflectance spectroscopy (NIRS) nutrient analysis of ground corn. Ani Feed Sci and Tech. 293:115472

Liao S, Wang T, Regmi N. 2015. Lysine nutrition in swine and the related monogastric animals: muscle protein biosynthesis and beyond. SpringerPlus. 4:1-12

Linting M, Meulman J, Groenen P, van der Kooij A. 2007. Nonlinear principal components analysis: introduction and application. Psy Meth. 12:(3)336-368

LMC International. 2020. The economic impact of canola on the Canadian economy. 2020 update. New York (NY): LMC International

LUMEX Instruments. 2018. SpectraLUM/Pro User's Manual. Mission (BC): LUMEX Instruments

Luque de Castro MD and Priego-Capote F. 2010. Soxhlet extraction: Past and present panacea. J Chromatogr. 1217:(16)2383-2389

Marangos A, Hill R, Laws BM, Muschamp D. 1974. The influence of three rapeseed meals and a mustard seed meal on egg and broiler production. Brit Poultry Sci. 15:(4)404-414

Mark H and Workman Jr. J. 2013. Calibration transfer. Spectroscopy. 28:(2)24-37

Martineau R, Ouellet DR, Lapierre H. 2013. Feeding canola meal to dairy cows: A meta-analysis on locational responses. J Dairy Sci. 96:(3)1701-1714

Martineau R, Ouellet DR, Lapierre H. 2014. The effect of feeding canola meal on concentrations of plasma amino acids. J Dairy Sci. 97:(3)1603-1610

Matthäus B and Luftmann H. Glucosinolates in the members of the family Brassicaceae: separation and identification by LC/ESI-MS-MS. J Agric Food Chem. 48:(6)2234-2239

Mathews Jr. KH, Bernstein J, Buzby JC. 2001. Chapter 4: Internatoinal Trade of Meat/Poultry Products and Food Safety Issues. Intl Trade and Food Safety. Agricultural Economic Report 828:48-73

Mauron J and Buiard E. 1964. Guanidination, an alternative approach to the determination of available lysine in foods. Mills AR, Fassmore R. (Eds.). Proceedings 6th Int Nutr Congr. Livingstone, Edinburgh, UK:489-490

McGregor DI and Downey RK. 1975. A rapid and simple assay for identifying low glucosinolate rapeseed. Can J Pl Sci. 55:(1)191-196

Newkirk RW, Classen HL, Scott TA, Edney MJ, Declercq DR, Siemens AJ. 2000. Prediction of lysine availability in canola meal. World's Poult. Sci. J. Abstracts and Proceeding: 56,60

Newkirk RW. 2002. The effects of processing on the nutritional value of canola meal for broilers chickens. University of Saskatchewan Department of Animal and Poultry. Saskatoon (SK): University of Saskatchewan

Newkirk RW, Classen HL, Edney MJ. 2003. Effects of prepress-solvent extraction on the nutritional value of canola meal for broiler chickens. Anim Feed Sci and Tech.104:(1-4)111-119

Niu X, Rogiewicz A, Patterson R, Slominski BA. 2023. Enhancing the nutritive value of canola meal for broiler chickens through enzymatic modifications. J Ani Sci. 101:1-10

Noel SJ, Jørgensen HJH, Knudsen KEB. 2022. The use of near-infrared spectroscopy (NIRS) to determine the energy value of individual feedstuffs and mixed diets for pigs. Anim Feed Sci and Tech. 283:1-17Oerlemans K, Barrett DM, Suade CM, Verkerk R, Dekker M. 2006. Thermal degradation of glucosinolates. Food Chem. 95:(1)19-29

Opalka M, Dusza L, Koziorowski M, Staszkiewicz J, Lipinski K, Tywonczuk, J. 2001. Effect of long-term feeding with graded levels of low glucosinolates rapeseed meal on endocrine status gilts and their piglets. Live Prod Sci. 69:(3)233-243

Ozaki Y, Huck C, Tsuchikawa, Engelsen SB. 2001. Near-infrared spectroscopy: theory, spectral analysis, instrumentation, and applications. 1st ed. Singapore: Springer Nature Singapore

Pahm AA, Pedersen C, Stein HH. 2008. Application of the reactive lysine procedure to estimate lysine digestibility in distillers dried grains with solubles fed to growing pigs. J Agric Food Chem. 56:(20)9441-9446

Pahm AA, Scherer CS, Pettigrew JE, Baker DH, Parsons CM, Stein HH. 2009. Standardized amino acids digestibility in cecostomized roosters and lysine bioavailability in chicks fed distillers dried grains with solubles. Poult. Sci. 88:571-578

Prieto MA, Lopez CJ, Simal-Gandara J. 2019. Glucosinolates: molecular structure, breakdown, generic, bioavailability, properties and healthy and adverse effects. Adv in Food and Nut Res. 90:305-350

Rad-Spice M, Rogiewicz A, Jankowski J, Slominski, BA. 2018. Yellow-seeded *B. napus* and *B. juncea* canola. Part 1. Nutritive value of the meal for broiler chickens. Ani Feed Sci and Tech. 240:66-77

Radfar M, Rogiewicz A, Slominski BA. 2017. Chemical composition and nutritive value of canola-quality *Brassica juncea* meal for poultry and the effect of enzyme supplementation. Anim Feed and Nut. 225:97-108

Renner IE and Fritz VA. 2020. Using near-infrared reflectance spectroscopy (NIRS) to predict glucobrassicin concentrations in cabbage and brussels sprout leaf tissue. Plant Methods. 16:(136)1-12

Ring, EFJ. 2000. The discovery of infrared radiation in 1800. The Image Sci J. 48:(1)1-8

Roberts CA, Workman Jr. J, Reeves JB. 2004. Near-infrared spectroscopy in agriculture. Madison (WIS): American Society of Agronomy Inc.; Crop Science Society of America Inc.; Soil Science Society of America Inc.

Roth-Mailer DA, Böhmer, Roth FX. 2004. Effects of feeding canola meal and sweet lupin (*L. luteus*, *L. angustifoluis*) in amino acids balance diets on growth performance and carcass characteristics of growing-finishing pigs. Anim Res. 53:(1)21-34

Ruan D, Wang H, Cheng F. 2018. The Maillard reaction in food chemistry: current technology and applications. Cham (SUI): Springer

Sáez-Plaza P, Michałowski T, Navas MJ, Asuero AG, Wybraniec S. 2013. An overview of Kjeldahl method of nitrogen determination. Part 1. Early history, chemistry of the procedure, and titrimetric finish. Crit Rev in Anal Chem. 43:(4)224-272

Sahamishirazi S, Zikeli S, Fleck M, Claupein W, Graeff-Hoenninger, S. 2017. Development of a near-infrared spectroscopy method (NIRS) for fast analysis of total, indolic, aliphatic and individual glucosinolates in new bred open pollinating genotypes of broccoli (*Brassica oleracea. botrytis var. italica)*. Food Chem. 232:272-277

Salazar-Villanea S, Astúa-Ureña, Masía-Montoya A, Herrera-Muñoz JI, Salas-Durán. 2022. Effect of protease supplementation on apparent ileal crude protein and amino acid digestibility of over-processed soybean meal in broilers. J Anim Sci and Biotech. 13:(74)1-8

Semwogerere F, Neethling J, Muchenje V, Hoffman LC. 2019. Meat quality, fatty acid profile, and sensory attributes of spent laying hens fed expeller press canola meal or a conventional diet. Poult Sci. 98:(9)3557-3570

Shenk JS and Westerhaus MO. 1996. Calibration the ISI way. Near inf spec: The future waves.:196-202

Siesler HW, Ozaki Y, Kawata S, Heise HM. 2001. Near-infrared spectroscopy: principles, instruments, applications. 1st ed. Weinheim (GER): Wiley-VCH

Simonne AH, Simonne EH, Eitenmiller R, Mills HA, Presman III CP. 1997. Could the Dumas method replace the Kjeldahl digestion for nitrogen and crude protein determinations in foods?. J Sci Food Agric. 73:(1)39-45

Singh, R. 2016. Harmonic and anharmonic oscillator models for pure vibrational infrared spectroscopy of diatomic molecules. Int J Adv Sci Eng. 8:(1)2016-2024

Slominski BA. 1997. Developments in the breeding of low fibre rapeseed/canola. J of Anim and Feed Sci. 6:(3)303-317

Slominski BA. 2018. Advances in understanding the dietary fibre and its components in relation to the use of alternative feed ingredients in modern poultry and livestock production. Proceedings of the 2018 Animal Nutrition Conference of Canada. 2018 May 2-3:107-130

Slominski BA and Campbell LD. 1987. Gas chromatographic determination of indole glucosinolates – a re-examination. J Sci Food and Ag. 40:(2)131-43

Slominski BA and Campbell LD. 1990. Non-starch polysaccharides of canola meal: quantification, digestibility in poultry and potential benefit of dietary enzyme supplementation. J Sci Food Ag. 53:(2)175-184

Slominski BA and Rogiewicz A. 2019. Low-glucosinolate rapeseed meal as a valuable source of protein for poultry. Proc 22nd Eur Sympo Gdansk, Poland:15-19

Slominski BA, Jia W, Rogiewicz A, Nyachoti CM, Hickling D. 2012. Low-fiber canola. Part 1. Chemical and nutritive composition of the meal. J of Ag and Food Chem. 60:(50)12225-12230

Slominski BA, Campbell LD, Guenter W. 1994. Carbohydrates and Dietary Fibre Components of Yellow-and Brown-Seeded Canola. J Agric Food Chem. 42:(3)704-707

Slominski BA, Simbaya J, Campbell LD, Rakow G, Guenter W. 1999. Nutritive value of broilers of meals derived from newly developed varieties of yellow-seeded canola. Ani Feed Sci and Tech. 78:(3)249-262

Smith BC. 1999. Infrared spectral interpretation. 1st ed. Boca Raton (FL): CRC Press

Solana M, Mirofci S, Bertucco A. 2016. Production of phenolic and glucosinolate extracts from rocket salad by supercritical fluid extraction: process design and cost benefits analysis. J Food Eng. 168:35-41

Thiex NJ, Anderson S, Gildemeister B, and Collaborators: Adcock W, Boedigheimer K, Bogren E, Coffin R, Conway K, DeBaker A, Frankenius E, Gramse M, Hogan P, Knese T, MacDonald JM, Iler J, Royle R, Russell M, Shafiee F, Shreve B, Sieh J, Spann M, Töpler E, Watts M. 2003. Crude fat, diethyl ether extraction, in feed, cereal grain, and forward (Randall/Soxhlet/submersion method): collaborative study. J of AOAC Intl. 86:(5)888-898

Thomas VM, Katz RJ, Auld DL, Peterson CL. 1984. Value of mechanically extract rape and safflower oilseed meals as supplement for growing lambs. Ani Feed Sci and Tech. 11(4):269-277

Thompson M, Owen L, Wilkinson K, Wood R, Damant A. 2002. A comparison of Kjeldahl and Dumas methods for the determination of protein in foods, using data from a proficiency testing scheme. 127:(2)1666-1668

Toledo-Martín EM, García-García MDC, Font R, Moreno-Rojas JM, Salinas-Navarro M, GómezP, del Río-Celestino M. 2018. Quantification of total phenolic and carotenoid content in blackberries (*Rubus Fructicosus* L.) using near infrared spectroscopy (NIRS) and multivariate analysis. Mol. 23:199-259

Toghyani M, Mohammedsalehi A, Gheisari A, Tabeidian SA. 2009. The effect of low-glucosinolate rapeseed meal in diets with multi-enzymes supplement on performance and protein digestibility in broiler chicks. J Anim and Feed Sci. 18:(2)313-321

Trindade Neto MAD, Opepaju FO, Slominski BA, Nyachoti CM. 2012. Ileal amino acid digestibility in canola meals from yellow and black-seeded Brassica napus and Brassica juncea fed to growing pigs. J Anim. Sci. 90:(10)3477-3484

Van Etten CH, McGrew CE, Dazenbichler ME. 1974. Glucosinolate determination in cruciferous seeds and meal by measurement of enzymatically release glucose. J Ag and Food Chem. 22(3)483-487

Velasco L and Becker HC. 1998. Analysis of total glucosinolate content and individual glucosinolates in *Brassica* spp. by near-infrared reflectance spectroscopy. Plant Breeding. 117:(2)371-382

Varmuza K and Filzmoser P. 2009. Introduction to multivariate statistical analysis in chemometrics. 1st ed. Boca Raton (FL). CRC Press

Waldern DE. 1973. Rapeseed meal versus soybean meal as the only protein supplement for lactating cows fed corn silage roughage rations. Can J Anim Sci. 53:(1)107-122

Wang T, Crenshaw MA, Regmi N, Rude BJ, Shamimul Hasan M, Sukumaran AT, Dinh T, Liao SF. 2017. Effets of dietary lysine level on content and fatty acid composition of intramuscular fat in late-stage finishing pigs. Can J Anim Sci. 98:(2)241-249

Wiedemann SCC, Hansen WG, Sneider M, Wortel VAL. 1998. NIR calibration in practice. Analusis Magazine. 26(4):38-43

Williams P and Norris KH. 2001. Near-infrared technology: in the agricultural and food industries. 2nd ed. St. Paul (MN): American Association of Cereal Chemists

Williams P, Antoniszyn A, Manley M. 2019. Near-infrared technology: getting the best out of light. Stellenbosch (SA): African Sun Media

Workman Jr. J and Weyer L. 2012. Practical guide and spectral atlas for interpretive near-infrared spectroscopy. 2nd ed. Baton Rouge (LA): CRC Press

Workman Jr. J. 2014. An introduction to Near Infrared Spectroscopy. Waltham (Mass): Wiley Analytical Science Magazine; [accessed May 2023]. https://analyticalscience.wiley.com/content/article-do/introduction-near-infrared-spectroscopy

Workman Jr. J. 2018. A review of calibration transfer practices and instrument differences in spectroscopy. App Spec. 72:(3)340-365

Wu X, Huang H, Childs H, Wu Y, Yu L, Pehrsson PR. 2021. Glucosinolates in *Brassica* vegetables: characterization and factors that influence distribution, content and intake. Anim Review of Food Sci and Tech. 12: 485-511

Zambiazi RC, Przybylski R, Zambiazi MW, Mendonca CB. 2007. Fatty acid composition of vegetable oils and fats. Boletim do Centro de Pesquisa de Processamento de Alimentos 25:(1)

Zhu JL, Zeng JK, Shurson GC, Urriloa PE. 2018. A meta-analysis to predict the concentration of standard ileal digestible amino acids in distillers dried grains with solubles for poultry. Poult Sci. 97:4359-4366