NIR hyperspectral imaging for animal feed ingredient applications

Princess Tiffany Galaura Dantes

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Recommended Citation
Dantes, Princess Tiffany Galaura, "NIR hyperspectral imaging for animal feed ingredient applications" (2020). Graduate Theses and Dissertations. 18113.
https://lib.dr.iastate.edu/etd/18113

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
NIR hyperspectral imaging for animal feed ingredient applications

by

Princess Tiffany Galaura Dantes

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Agricultural and Biosystems Engineering

Program of Study Committee:
Charles R. Hurburgh Jr., Major Professor
Thomas J. Brumm
Matthew J. Darr
Dirk E. Maier
Ranjan Maitra

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University
Ames, Iowa
2020

Copyright © Princess Tiffany Galaura Dantes, 2020. All rights reserved.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>1</td>
<td>CHAPTER 1. GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Dissertation Organization</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>CHAPTER 2. APPLICATIONS OF NIR HYPERSPECTRAL IMAGING FOR ANIMAL FEED</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>QUALITY AND SAFETY – A REVIEW</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Abstract</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>HSI Applications in Feed Safety and Quality</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Future Trends and Conclusion</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>CHAPTER 3. ANALYSIS OF PROTEIN AND OIL IN SOYBEAN MEAL USING NIR</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>HYPERSONTICAL IMAGING</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Abstract</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Methodology</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Results and Discussion</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>62</td>
</tr>
<tr>
<td>4</td>
<td>CHAPTER 4. NEAR-INFRARED HYPERSONTICAL IMAGING FOR</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>DETERMINATION OF LYSINE CONCENTRATION IN ANIMAL FEED</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>INGREDIENTS</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Abstract</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Materials and Methods</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Results and Discussion</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>86</td>
</tr>
<tr>
<td>5</td>
<td>CHAPTER 5. GENERAL CONCLUSION</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Guideline Method of Model Calibration for NIR HSI.</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>General Findings</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Recommendations for Future Work</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Appendix: Near-Infrared Methods—Guidelines for Model Development,</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Instrument Management and Maintenance (A Draft)</td>
<td></td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my major professor, Dr. Charles R. Hurburgh Jr., for his invaluable guidance, consistent support, and encouragement throughout my Ph.D. program and professional development.

My sincere thanks to my committee members, Dr. Thomas Brumm, Dr. Matthew Darr, Dr. Steven Hoff, Dr. Dirk Maier, and Dr. Ranjan Maitra for their valuable feedback and contributions to my research and program of study. I would also like to offer my appreciation to Corning Optics for the NIR hyperspectral imaging instrument. To our laboratory manager, Glen Rippke, I very much appreciate his assistance and guidance in my research.

I would like to thank Connie Hardy, Ria Gasaway, my friends and colleagues at the Grain Quality Laboratory, and my Filipino friends in Ames for being my family here in the US and making my stay at Iowa State University worthwhile. My parents and brother have been supportive in my endeavor despite the distance, my special thanks to them. I would also like to give my husband, Querriel Arvy Mendoza, warm gratitude for his patience, understanding, and constant support in every step of the way.

Above all, I would like to thank God for providing me the wisdom and strength to complete my dissertation and doctoral program and for making all these things possible.
ABSTRACT

Considering its wide application in the food industry, near-infrared hyperspectral imaging (NIR HSI) was explored for animal feed applications. Its ability to provide chemical composition of the sample at the pixel level provides an advantage over the typical NIR spectroscopy. In this dissertation, a literature review was presented that highlights the applications of NIR HSI on grains, oilseeds, and animal feed ingredients. In our first study, a Corning NIR HSI instrument was used to predict protein and oil content in soybean meal and visualize predicted protein distribution over the entire soybean meal sample. Preprocessing by standard normal variate and Savitzky-Golay derivative was effective in improving calibration model performance. The NIR HSI instrument was also compared with two commercially available single-point NIR spectrometers which are typically used in the grain and feed industry. Absorbance spectra from the NIR HSI instrument were relatively close to those from the two NIR instruments in most of the wavelengths. Regression coefficients from soybean meal protein model calibration highlighted the similarities in the contributing variables of the three instruments. In our second study, lysine concentration was determined in soybean meal and dried distillers’ grains with solubles (DDGS) using NIR HSI in combination with partial least squares regression or spectral angle mapper (SAM) classification. Score plots from principal component analysis separated pure lysine spectra from soybean meal and DDGS. Increasing the SAM maximum angle also increased the model calibration performance. Overall, both PLS regression and SAM classification obtained promising results thereby indicating the potential of this technology to be used in evaluating amino acid concentration in animal feeds.
CHAPTER 1. GENERAL INTRODUCTION

Soybean meal and corn dried distillers’ grains with solubles (DDGS) are among the top animal feed ingredients for livestock and poultry in the United States (U.S.) (American Feed Industry Association [AFIA], 2019). Corn constitutes over 95% of the overall feed grain produced and utilized in the country (U.S. Department of Agriculture [USDA], 2018). Soybean meal contains 44 to 50% crude protein content and a balanced amino acid profile which makes it the major source of plant protein for livestock (Balasreri et al., 2016; Food and Agriculture Organization [FAO], 2003; Shen et al., 2016; Zhou et al., 2015). DDGS are considered a low-cost alternative to soybean meal with a protein content that has been reported to enhance productivity of animals (Zhou et al., 2015).

In the animal feed industry, quality and safety of animal feed are of utmost importance. The nutritional quality of feed ingredients is vital in effectively matching diet specifications to animal requirements (Van Barneveld et al., 2018). Rations are now being balanced by diet formulation methods based on the amounts of amino acids that can limit the feed’s nutritional efficiency. Accurate information of the amino acid contents and the use of rapid analysis methods are therefore critical for a precise and cost-effective feed formulation (Van Barneveld et al., 2018; Fontaine et al., 2001; Kovalenko et al., 2006). Wet chemistry methods for amino acid analysis, such as high-pressure liquid chromatography (HPLC), are destructive, costly, labor-intensive, slow, and unsuitable for routine use, especially when a large sample size needs to be tested. For these reasons, it is a common practice for feed nutritionists to approximate the
amino acid contents by utilizing published regression equations based on crude protein content (De La Haba et al., 2006).

Being a proven successful tool in the rapid and accurate analysis of moisture and protein of cereals, grains and oilseeds, near-infrared spectroscopy (NIRS) has been used in various studies to measure amino acid contents, such as lysine since 1978 (Fontaine et al., 2001). These studies have shown the superior performance of NIRS models in determining amino acid contents, specifically for lysine in soybean meal, as compared to estimating from crude protein regressions (Fontaine et al., 2002). However, it was also observed that NIRS was able to measure amino acid concentration due to its correlation with its protein content, and that the calibrations were actually predicting protein instead of amino acids (Fontaine et al., 2001; Kovalenko et al., 2006). As such, the NIRS measurement of amino acid content is a function of the errors of protein prediction and amino acid derivation from predicted protein in soybeans (Kovalenko et al., 2006).

One of the recent developments in NIRS is near-infrared hyperspectral imaging (NIR HSI), a combination of spectroscopy and machine vision. NIR HSI provides more information by giving a full spectrum at each pixel of the sample. Most of the studies done on NIR HIS for grain analyses are focused on detecting contaminants, classifying grain hardness, quantifying protein of whole grains and discriminating materials (Caporaso et al., 2018). The potential of this technique in quantifying amino acid contents in ground feed ingredients has not been explored. Considering the variability and low concentrations of amino acids in feed ingredients, NIR HSI can possibly offer more advantages than conventional NIRS because of the former’s ability to show the
distribution of the property within the sample and spot areas with high or low concentration.

In this study, Chapter 2 highlights the applications of NIR hyperspectral on animal feed quality and safety. Chapter 3 focuses on the application of NIR hyperspectral imaging for the analysis of protein content in soybean meal. The performance of reflectance NIR HSI was analyzed and compared with two existing reflectance NIR instruments. The main objective of Chapter 4 was to develop model calibration to determine lysine content in soybean meal and DDGS using NIR HSI.

This project applied near-infrared hyperspectral imaging technology for prediction of protein and lysine content in ground feed ingredients. The specific amino acid being measured was lysine. Dried distillers’ grains with solubles and soybean meal were the feed ingredients used as samples for amino acid prediction, while only soybean meal samples were utilized for protein prediction.

**Dissertation Organization**

To address the general objective of this research, the dissertation was organized into the following sequence: 1) general introduction, 2) applications of NIR hyperspectral imaging for animal feed quality and safety – a review, 3) analysis of protein and oil in soybean meal, 4) near-infrared hyperspectral imaging for determination of lysine concentration in animal feed ingredients, and 5) general conclusion. The general conclusion section covers three subtopics: method guideline in NIR hyperspectral imaging from instrumentation to data analysis, general findings, and recommendations for future work. An appendix of the guideline method for NIR model development was also attached at the end of chapter 5.
References


CHAPTER 2. APPLICATIONS OF NIR HYPERSPECTRAL IMAGING FOR ANIMAL FEED QUALITY AND SAFETY – A REVIEW

Princess Tiffany Dantes¹, Charles R. Hurburgh², Thomas Brumm³, Matthew John Darr², Dirk E. Maier², Ranjan Maitra²

¹Graduate research assistant, Iowa State University, ²Professor, Iowa State University, ³Associate Professor, Iowa State University

Modified from a manuscript to be submitted to Transactions of the ASABE

Abstract

The increasing need for safe and quality animal-derived food products and the limitations of slow and destructive traditional screening methods have driven the development of fast, accurate and reliable testing methods for evaluation of animal feed safety and quality parameters. One of the methods being applied recently is near-infrared hyperspectral imaging (NIR HSI), an analytical technique that provides chemical and spatial information of the sample. It has been applied to various food and agricultural products and is gaining acceptance for commercial use in the food industry. This review presents research progress with respect to applications of NIR HSI for safety and quality evaluation of animal feed ingredients, specifically in screening illegal adulteration and contamination of feed materials, detecting biological contamination and damage, measuring constituent concentrations and other quality properties, and evaluating feed processing parameters. Major results, multivariate analysis and hyperspectral image processing algorithms used, and limitations of the different studies being reviewed were highlighted. With its potential uses and applications on animal feed, NIR HSI is a promising breakthrough in the feed industry.
Introduction

Recently, there is a growing public concern on the quality and safety of animal-derived food products as a result of emerging problems from these foods, such as contamination and outbreaks of food-borne bacterial infections which inadvertently pose health risk to humans (Food and Agriculture Organization[FAO], 2017). In 2016, the U.S. Food and Drug Administration (FDA) reported 18 out of 201 food and feed safety incidents that were related to animal feed and pet food products (U.S. FDA, 2016). It has been recognized that the prerequisite for human food safety is animal feed safety (Pinotti and Dell’ Orto, 2011).

These increasing issues on human and animal food quality and safety have prompted the food and feed industry to have close consideration of all aspects of food production and of the feed supply chain, and to ensure conformity to quality and traceability standards (FAO, 2017; Pinotti and Dell’ Orto, 2011).

These have also encouraged more research on the development of technologies and processes that will help address the associated problems. Moreover, regulations and protocols are now being implemented towards prevention and control to lessen the contamination of animal feed and related food safety and quality risks. For example, the FDA Food Safety Modernization Act (FSMA) Preventive Controls for Animal Food requires animal food facilities to comply with the Current Good Manufacturing Practice (cGMP) requirements. In plant operations, testing procedures for identification of sanitation failures or possible animal food contamination must be ensured (21 CFR 507.25(a)(6)), and raw materials and other ingredients must be properly evaluated (21 CFR 507.25(b)) (U.S. FDA, 2017).
In feed processing, there are techniques and machines available that will ensure that the final feed product is uniform, not adulterated, free from contamination, and has the desired amino acids and nutrient composition. For quality assurance, a sample product that goes out of the feed mill is usually sent to an outside laboratory and it may take days for the quality analysis to be available. Thus, alongside with these processing technologies, non-destructive, fast, and accurate testing methods that can detect minor components at low concentrations need to be available for monitoring feed safety and quality especially during feed mill operations.

Near infrared hyperspectral imaging (NIR HSI) is a recent technique that integrates machine vision and near infrared spectroscopy. It provides both spectral (composition) and spatial (location) information of the product. It can be applied for in-, at- or on-line quality and safety monitoring. Some studies on the application of NIR HSI on food and agricultural products have already been published, and there are companies who are using NIR HSI in their actual operations.

This technology has a lot of potential in the feed industry. It can be used as a primary screening method in feed mills and as quality control in feed processing to solve problems related to mixer uniformity, contamination, and ingredient quality (de Jong et al., 2016). It can provide an efficient automatic quality control system and minimize possible recalls and economic losses.

Published studies and reviews on NIR HSI are mostly focused on its application with human food. So far, there are applications of hyperspectral imaging specifically for animal feed. Thus, there is a need for a review that will summarize the methods and applications that have been used for feed safety and quality. This review will encourage
more researchers to explore the potential of HSI in feed processing so that industry users will adopt this technology.

**Principle of near-infrared hyperspectral imaging**

Spectrometers, such as near infrared spectrometers, work by vibrational spectroscopy which is based on optical technology – the interaction between incident light and molecules (Feng and Sun, 2012). The sensitivity of different molecules to light with varying wavelengths (according to light absorption or scattering) thereby allows the resultant spectra to record the information of molecules corresponding to the wavelengths (Feng & Sun, 2012). From these, the physical, chemical and biological details within the samples are determined based on the spectral data (Feng & Sun, 2012). As such, the spectra may not always be the representative for the whole sample and it does not provide spatial (location) information (Feng & Sun, 2012).

On the other hand, computer vision uses the color bands of red, green, and blue to determine the sample’s characteristics such as shape, color, size and texture (Feng & Sun, 2012). As opposed to NIRS, there are not many reports on computer vision being able to measure chemical and biological characteristics of the sample (Feng & Sun, 2012). Despite their extensive range of applications, both optical sensing methods have limits and drawbacks.

A spectral image is composed of three-dimensional (two for spatial and one for spectral) images of the same object which are stacked at various spectral bands, as shown in Figure 1 (Lind & Murhed, 2012; Sendin et al., 2018). Each pixel in the image has its specific wavelength. Spectral images can be categorized as hyperspectral imaging, multispectral imaging or ultra-spectral (Lind & Murhed, 2012; Wu & Sun, 2013). In multispectral, images are acquired with very few (typically less than 10)
discrete wavebands wherein a set of isolated data points is produced for each pixel (ElMasry & Sun, 2010; Qin et al., 2013). For a very fine spectral resolution, an ultraspectral imaging is the commonly utilized system (ElMasry & Sun, 2010). In hyperspectral imaging, tens to hundreds of images having contiguous wavelengths are acquired. Every pixel in the acquired image will have a full spectrum (Lind & Murhed, 2012; Wu & Sun, 2013).

Hyperspectral cubes, also called hypercubes, can be acquired using point-scan, line-scan, and area-scan methods (Lind & Murhed, 2012; Wu & Sun, 2013). Point-scan and line-scan are spatial-scan methods, while area-scan is classified as spectral-scan method (ElMasry & Sun, 2010; Qin et al., 2013).

In the point-scan method, also referred to as whisk-broom method, a single spectrum is measured for each spatial location (pixel) in the scene at a time; then the sample or detector is moved point by point (pixel by pixel) to capture another spectrum until the entire scene has been completely scanned (Qin et al., 2013; ElMasry & Sun, 2010; Amigo & Grassi, 2020). This method stores hypercubes in Band Interleaved by
Pixel (BIP) format (ElMasry & Sun, 2010). It takes time to acquire data, but it ensures a higher spectral resolution. (Amigo & Grassi, 2020).

The line-scan method or the push-broom method is faster than point-scan because of its simultaneous acquisition of full spectral data for each spatial point in the linear field of view (FOV) of the sample (Qin et al., 2013; ElMasry & Sun, 2010; Amigo & Grassi, 2020). In this method, hypercubes are stored in Band Interleaved by Line (BIL) format (ElMasry & Sun, 2010). With its ability to capture images of moving samples, line-scan method is among the preferred technologies for online food inspection in industrial applications since food products typically have linear movement, such as in conveyor belt systems along a production line (Qin et al., 2013; Amigo & Grassi, 2020; ElMasry & Sun, 2010).

In area-scan method, hyperspectral images are acquired per wavelength basis with full spatial data at once and are stored in Band Sequential (BSQ) format (Qin et al., 2013; ElMasry & Sun, 2010). Unlike the first two methods, there is no movement needed for the sample or detector in area-scan method (Qin et al., 2013; ElMasry & Sun, 2010).

Similar to near infrared spectrometers, HSI systems also have reflectance, transmittance and interactance as the most commonly used measurement modes (Wu & Sun, 2013). The difference among these sensing modes lies in the configuration of light source, camera, spectrograph and lens, and thus, they also vary in application (Wu & Sun, 2013). Shown in Figure 2 are the different elements commonly found in a hyperspectral imaging system.
With their large data size, analyzing hyperspectral images requires hypercube pre-processing (spectral and/or spatial), chemometric algorithms and image processing techniques to extract the needed quantitative and qualitative information (Lind & Murhead, 2012; Wu & Sun, 2013; Amigo et al., 2013). Pre-processing is needed to remove undesirable phenomena in the image that can have negative effect in the multivariate models to be developed later (Amigo et al., 2013). Different software tools can be used to process hyperspectral images, such as Environment for Visualizing Images (ENVI) software, Unscrambler and MATLAB, and to facilitate model development and calibration (Lind & Murhead, 2012; Wu & Sun, 2013).

![Components in a hyperspectral imaging system](Sendin et al., 2018)

Spatial pre-processing involves image compression, background removal, selection of region of interest, image segmentation (segments), thresholding (maximum or minimum limits), and removal of dead pixels, among others (Amigo et al., 2013). Dead pixels, which usually result from detector anomalies, may appear in the form of missing, zero or infinite values (Amigo et al., 2013; Amigo & Santos, 2020).
Spectral preprocessing, such as standard normal variate, multiplicative scatter correction, Savitzky-Golay derivatives and smoothing, and detrending, that are commonly used in NIR spectroscopy can be also applied to reduce scattering effects and undesirable spectral variability (Amigo et al., 2013). They can also improve signal-to-noise ratio of the spectra (Lin et al., 2014; Zhao et al., 2018; Li et al., 2016). Preprocessed spectra will then be used as input to develop calibration model that will teach the instrument in measuring the desired properties of a product. Model performance is typically evaluated by statistical parameters (i.e., $R^2$, standard error, root mean square error, or classification accuracy). Principal component analysis, partial least squares discriminant analysis, partial least squares regression and support vector machine are among the widely used multivariate techniques in hyperspectral imaging.

**HSI Applications in Feed Safety and Quality**

**Adulteration and contamination**

Various controls are used to ensure feed quality and safety, and these are mostly performed at the laboratory level (Jong et al., 2016). The usual focus of analytical methods used is on a specific type of adulterant or contaminant which limits their capacity to spot unpredicted problems (Jong et al., 2016; Shen et al., 2020). Recently, official controls failed to promptly detect the illegal melamine adulterations of feed materials used as protein sources with melamine and the contamination of oil with technical fats containing dioxins (Jong et al., 2016). The melamine contamination in pet food and animal feed caused death and illness of pet dogs and cats in North America in 2007 (Shen et al., 2020). These issues happened because only few samples were usually tested for contaminants and melamine was not part of the typically monitored chemical compounds (Jong et al., 2016). Hence, there is an urgent need to have a
different detection technique that is capable of rapid and accurate screening of any unusual spectra to manage future problems in adulteration and contamination (Shen et al., 2020).

Fernández Pierna et al. explored the use of near infrared hyperspectral imaging (1118-2424 nm) in combination with principal component analysis (PCA) and partial least-squares discriminant analysis (PLS-DA) to detect and quantify melamine and cyanuric acid contamination in soybean meal (Fernández Pierna et al., 2014). Melamine and cyanuric acid contents from 0.5% to 5% were added to pure soybean meal after obtaining spectra of pure products (Fernández Pierna et al., 2014). The addition of melamine and cyanuric acid to pure soybean meal is an example of spiking the mixture which is done to produce laboratory-scale samples that can enhance the calibration signal-to-noise ratio and contribute to a robust quantitative calibration (Cogdill and Anderson, 2005). The spectra and PCA plots in Figure 3 show that it was possible to characterize soybean meal and segregate the adulterants (Fernández Pierna et al., 2014). Moreover, the high $R^2$ of regression between laboratory % concentration and number of detected pixels, 0.89 for melamine and 0.94 for cyanuric acid, implied that both contaminants could be identified at levels as low as 0.5% concentration (Fernández Pierna et al., 2014). The authors suggested for more research to be done on quantitative applications to examine lower concentration levels of melamine or other adulterants (Fernández Pierna et al., 2014). A similar study indicated that PLS-DA and partial least-squares regression (with variable selection using least angle regression – least absolute shrinkage and selection operator methods) could be used to determine at least 25 ppm of melamine concentration in soybean meal (Li et al., 2016).
Shen et al. also developed methods for non-targeted screening of various contaminants (melamine, cyanuric acid, urea, di-ammonium phosphate, biuret, and mono-ammonium phosphate) in soybean meal using a Fourier transform near-infrared (FT-NIR) imaging system (1282-2500 nm), PCA, PLS-DA and global H (GH) value discrimination method (Shen et al., 2016). Since PLS-DA is a supervised method which needs prior information of the data, the authors reported that it could not be useful if the adulterants were unknown and that GH method could be applied instead for non-targeted purposes (Shen et al., 2016). In 2020, Shen et al. used the same FT-NIR as in 2016 and a push-broom NIR hyperspectral imaging (1100-2400) to detect those six contaminants in soybean meal (Shen et al., 2020). Data analysis was done by applying local anomaly detection (LAD), PLS-DA, and regression of reference concentration with number of pixels detected (Shen et al., 2020). This study showed how LAD method could be used effectively for untargeted screening as low as 0.5% for single material and 3% for multiple adulterants in soybean meal (Shen et al., 2020).
In Europe, the crisis on bovine spongiform encephalopathy (BSE) had led to strict regulations and measures assuring animal feed quality and safety, specifically prohibiting processed animal proteins (PAPs) or meat and bone-meal (MBM) in feed ingredients (Fernández Pierna et al., 2004; Fernández Pierna et al., 2010). Due to the high diversity of animal feed composition and the limitations of current quality control procedures in monitoring animal feed for PAPs fragments or other contaminants, NIR hyperspectral imaging was being explored along with other analytical techniques to facilitate the enforcement of regulations. Several authors performed studies on discriminating PAPs in compound feeds and on screening the composition of feedstuffs and identifying impurities (Fernández Pierna et al., 2004; Fernández Pierna et al., 2010; Wilcox et al., 2014; Garrido-Novell et al., 2018; Riccioli et al., 2018; Fernández Pierna et al., 2006; Fernández Pierna et al., 2012). Six of these studies used HSI systems in the 900-1700 wavelength range and analyzed using one or more methods: support vector machines (SVM), artificial neural network (ANN), and PLS-DA (Fernández Pierna et al., 2004; Fernández Pierna et al., 2010; Garrido-Novell et al., 2018; Riccioli et al., 2018; Fernández Pierna et al., 2006; Fernández Pierna et al., 2012). The SVM technique was found to provide the best results, giving fewer false negatives (and fewer false positives) and good results (Fernández Pierna et al., 2004; Fernández Pierna et al., 2006; Riccioli et al., 2018). In the study on in-house validation of NIR hyperspectral imaging, Fernández Pierna et al. reported that this technology could be used to detect PAPs in compound feeds qualitatively with 0.1% limit of detection (Fernández Pierna et al., 2010).
Recent studies conducted on wheat and cereal flours used NIR hyperspectral imaging for detection of talcum powder (Fu et al., 2020), benzoyl peroxide (Fu et al., 2020), peanuts (Laborde et al., 2020; Zhao et al., 2018; Mishra et al., 2016; Mishra et al., 2015) and tree nuts (Zhao et al., 2018; Mishra et al., 2015), all of which may cause serious health risks to humans. In one of these studies, a SWIR push-broom HSI system (1000 to 2500 nm spectral range) was used along with preprocessing and independent component analysis (ICA) (Mishra et al., 2016). The authors found that ICA performed better than PCA, thereby concluding that HSI combine with ICA has the potential to quantitatively predict chemical properties of samples (Mishra et al., 2016).

One of the challenges reported was the influence of sample depth and particle size on classification accuracy (Fu et al., 2020). The size of the pixel could be bigger than the particle size in powder samples like the wheat flour and the pixel could be composed of more than one pure chemical component (Laborde et al., 2020). This challenge was addressed in the study of Laborde et al. with the use of Linear Mixing Model (LMM) and by applying Matched Subspace Detector (MSD) algorithm which enabled the subpixel detection of peanut at the minimum of 0.2% adulteration in wheat flour (Laborde et al., 2020). In LMM, there is an assumption that the radiance acquired by a pixel is the “sum of the chemical radiances weighted by their surface contribution in the pixel field of view” (Laborde et al., 2020).

Overall, most of these studies suggested to conduct further research on particle size and investigate inclusion of neighboring pixels in pixelated spectra (Fu et al., 2020; Laborde et al., 2020).
Biological contamination and damage

Aflatoxin (AFB₁), fumonisins (FUM), and deoxynivalenol (DON) are among the highly important mycotoxins in corn, soybean, wheat, distillers dried grains, and finished feeds. AFB₁ is considered as the strongest carcinogen (Y. Zhu et al., 2016; Chu et al., 2017). Conventional methods for testing of AFB₁ (i.e., high-performance liquid chromatography) typically involve laborious extraction and destructive sampling procedures (Chu et al., 2017). Near-infrared spectroscopic techniques have been used in cereals for mycotoxin detection, classification of kernels according to the levels of mycotoxin contamination, and discrimination of contaminated cereals (Chu et al., 2017). The studies done have shown the potential of NIRS in identifying mycotoxins, but the information obtained are usually from a limited region on the samples and no spatial data are given (Chu et al., 2017). This is a problem since there is non-uniform distribution of mycotoxin in a grain sample or in a pile (Chu et al., 2017). Some authors also stressed that it is more practical to eliminate the portion of grains which are contaminated with mycotoxins than to reject the whole batch of grains (Chu et al., 2017). For these reasons, the use of near-infrared hyperspectral imaging has been explored as a non-destructive mycotoxin detection method for the past decade (Chu et al., 2017).

Using a system constructed at the USDA–ARS, NIR HSI was used to identify and detect AFB₁ from pre-inoculated whole single maize kernels and then to create an algorithm that will match toxin concentration with pixel distribution (Wang et al., 2014; Wang et al., 2015). Images were acquired with the wavelength range from 1000 to 2500 nm (Wang et al., 2015). Wavelength dependent offset, masking method (for noise reduction and extraction of regions of interest of spectral image), and PCA were used in
processing and analyzing spectral data (Wang et al., 2015). It was found that AFB$_1$ were exclusively indicated in 1729 and 2344 nm (Wang et al., 2015). The results from this study, which specified detection accuracy for verification at 92.3%, show that NIR HSI is useful in aflatoxin detection of maize kernels that were artificially inoculated in the field (Wang et al., 2015). The images in Figure 4 show how PCA as a multivariate technique can separate healthy kernels from contaminated ones and how the level of contamination varies in each kernel.

In a different study but adopting the same HSI system, PCA was applied with SVM and support vector regression (SVR) (Chu et al., 2017). SVM classification of AFB$_1$ contamination levels in single corn kernels obtained overall accuracy of 82.5% for validation, while SVR results showed $R^2$ of 0.70 and relative predictive determinant (RPD) of 1.4 for validation (Chu et al., 2017). There was a wide range of AFB$_1$ content in the sample set and a highly skewed distribution of contaminated samples. The visualization maps showed the non-uniform distribution of AB1 within the single corn kernels (Chu et al., 2017). Chu et al. also published a study in 2020 which used reflectance NIR HSI (900-1700 nm), PCA, successive projections algorithm (SPA) and SVM modeling with pixel-wise and object-wise sampling strategies to classify healthy and fungi-infected corn kernels (Chu et al., 2020). They achieved classification accuracy of 92 to 100% and indicated that classification by pixel-wise performed better than object-wise in showing the distribution of infection within the kernels (Chu et al., 2020).
In contrast to the usual reflectance mode, a different study used fluorescence in the visible HSI (400 to 600 nm wavelength range) in determining the levels of aflatoxin in single maize kernels (Yao et al., 2010). Multiple linear regression, multivariate analysis of variance and two-class schema (20 or 100 mg/kg threshold) results gave classification accuracy of 0.84 to 0.91 (Yao et al., 2010). This study suggested that aflatoxin content of artificially contaminated single maize kernels can also be predicted by fluorescence HIS (Yao et al., 2010). After this, another research was done which involved both fluorescence and reflectance visible NIR HSI for the same objective (F. Zhu et al., 2016; Xing et al., 2017). Least squares support vector machines model provided classification accuracy of 90 – 95.33% from the integrated fluorescence and reflectance spectra (F. Zhu et al., 2016; Xing et al., 2017).

In wheat, various hyperspectral imaging studies in the 400-1700 nm range aimed at sorting healthy from fusarium-contaminated wheat kernels (Xing et al., 2017). PCA and PLS-DA were the commonly used methods for data analysis which led to 86 to 95% classification accuracy (Xing et al., 2017). In addition to fusarium damage, wheat
kernels also suffer from sprout damage which reduces milling yield and degrades flour quality (Xing et al., 2010). Grading system is usually performed by visual inspection and the entire batch of bulk wheat sample will be discarded even when sprouting is found in very few samples (Xing et al., 2010). Thus, a VNIR hyperspectral imaging system in the 400-1000 nm range was proposed to evaluate sprout damage in wheat kernels (Xing et al., 2010). The authors observed a useful index to classify sound from sprouted kernels by obtaining the ratio of reflectance at 878 nm to that of 728 nm (Xing et al., 2010). Data analysis by PCA identified three spatial and four spectral features that could be used to detect sprouting and classify sound, sprouted and severely sprouted kernels (Xing et al., 2010).

On the other hand, Chelladurai et al. explored the use of NIR HSI (900-1700 nm) technique and soft x-ray, combined with linear and quadratic discriminant analysis to detect weevil infestation in soybeans (Chelladurai et al., 2014). They obtained classification accuracies of 87% and 79% for uninfected soybeans from linear and quadratic discriminant analysis, respectively, and higher performance when NIR HSI and soft x-ray were combined (Chelladurai et al., 2014). Then in cereal flour, NIR hyperspectral imaging (1118-2425 nm) was assessed to detect and quantify ergot bodies by applying PLS-DA (Vermeulen et al., 2017). Although the spectra of ergot and wheat flour could be distinguished from one another, the authors found that it was not possible for their developed models to differentiate wheat flours when low levels of ergot particles were added (Vermeulen et al., 2017).

In general, the performance of the models, as the studies suggested, indicated that NIR HSI could be applied in the industry for sorting and screening purposes to
enhance safety controls (Xing et al., 2017; Xing et al., 2010). More work still needs to be done to make the models more robust and reflective of natural contamination rather than spiked contamination. It was also recommended to use the selected wavebands in a particular application to create a portable multispectral imaging tool for on-line inspection in grain handling and feed processing facilities (Xing et al., 2010; Chelladurai et al., 2014).

**Constituent concentrations and other quality parameters**

In grain and feed processing, materials other than grain (MOG) (such as chaff, straw, stones, and other types of dockage and foreign materials), contribute to reducing the product quality and to increasing wear and tear of processing equipment. Thus, the amount of MOG is considered as one of the significant quality parameters in grain products (Wallays et al., 2009). Usual quality control procedure is by using sieves, dockage tester or grain cleaner, but these are mostly done before or after and rarely during grain processing. An alternative to the manual process is the use of machine vision system which has a camera that will allow screening based on physical features, such as color, area, length, width and location. This will not give chemical identification of the product. This system can be prone to misclassification especially if the grain has the same physical characteristics with the foreign material. NIRS has also been used for classification and discrimination by proving the chemical information of the samples. However, since NIRS will only give the average measurement, it can possibly lead to rejecting the entire product, instead of just eliminating the unwanted materials within the product.

Being able to identify the presence of MOG during feed mill processing and locate them, if possible, will be beneficial to segregate faster, to remove the unwanted
materials immediately and to prevent the possible quality and equipment issues. Since machine vision and NIRS have both shown how they can be used in evaluating grain cleanness online, Wallays et al. (2009) combined these two techniques in a study on NIR hyperspectral imaging (400 to 950 nm wavelength) in the reflectance mode to: a) select the combination of five hyperspectral wavebands, b) discriminate wheat kernels from MOG, and c) evaluate the proportion of MOG in the sample based on the number of pixels as illustrated in Figure 5 (Wallows et al., 2009). PLS-DA was used as a multivariate classification technique (Wallows et al., 2009). Results obtained showed that the selected wavebands were 465 – 475 nm, 522 – 532 nm, 676 – 705 nm, 849 – 858 nm, and 906 – 945 nm, which will be utilized for further study to develop a multispectral vision sensor that will be used in actual operations (Wallows et al., 2009).

Likewise, Ravikanth et al. developed a method using NIR HSI (1000-1600 nm) to separate wheat from MOG or referred to as contaminants (foreign materials, dockage, and animal waste) in their study (Ravikanth et al., 2015). Spectra were preprocessed (SNV, first derivative, second derivative, Savitzky-Golay smoothing and derivative, multiplicative scatter correction (MSC), and standard normal variate (SNV)) prior to classification by SVM, Naïve Bayes (NB), and k-nearest neighbors (k-NN) (Ravikanth et al., 2015). A combination of SNV and k-NN obtained the best classification results having accuracy as high as 99.9% (Ravikanth et al., 2015). Broken wheat kernels were found to have the lowest classification accuracy (Ravikanth et al., 2015). This result was expected due to the inherent property of broken wheat, most likely sharing the same spectra as clean whole wheat samples. Image preprocessing and thresholding could be useful in segregating broken wheat samples.
Near-infrared spectroscopy has long been used to predict constituents of grain and grain products. Because of its limitation in capturing only a small part of grain samples, Cogdill et al. used NIR HSI in the transmittance mode in the 750 to 1090 nm wavelength range to predict moisture and oil contents of maize (Cogdill et al., 2004). Results from the spectral obtained, along with various multivariate processing techniques, achieved standard error of cross-validation at 1.20% and 1.38% for moisture and oil, respectively (Cogdill et al., 2004). RPD values were 2.74 for moisture and 1.45 for oil, indicating that models developed for moisture were much better than for oil (Cogdill et al., 2004). The main contributing factor in the poor results for oil calibration was the reference chemistry measurements for single kernels resulting from a) balance and human errors, and b) the destructive nature of the reference method (Cogdill et al., 2004). Still, these results suggested that hyperspectral imaging could be useful in predicting moisture content of maize kernels and that further research would be needed to make this technology the standard method in grain quality analysis (Cogdill et al.,

Figure 5. Wheat kernels with materials other than grain (MOG) and (b) MOG in red pixels subtracted from the black background (Wallays et al., 2009).
Meanwhile in reflectance NIR HSI, oil of individual corn kernels was also predicted together with oleic acid (Weinstock et al., 2006). Obtained $R^2$ of up to 0.76 and root mean square error of prediction of at least 0.74 were considered by the authors as poor performance (Weinstock et al., 2006). The authors proposed that the models could be improved by increasing the sample size, removing outliers, improving lighting design, and selecting the necessary wavelengths (Weinstock et al., 2006).

In using hyperspectral imaging to predict protein content in wheat, studies showed that PLS regression performed better than principal component regression, and that SNV + first derivative spectral preprocessing could further improve the calibration performance (Caporaso et al., 2018; Mahesh et al., 2015). Although the calibration error in NIR HSI in these studies were higher than in the usual single-point NIR spectrometers, the models and procedure would serve as starting point for the potential application of hyperspectral imaging in breeding and screening purposes in grain and feed (Caporaso et al., 2018). These studies were suggesting that NIR HSI could also be used for quantitative predictions aside from the usual classification and detection applications.

In corn, one of the valuable parameters that have an effect on processing and final product quality is kernel hardness, which relates to bulk density and breakage susceptibility (Williams et al., 2009; Williams et al., 2016). During dry milling, soft kernels produce fewer large grits than hard kernels, while extremely hard kernels need more power input and are more susceptible to stress cracks and breakage – which are the reasons why grain producers and processors are concerned about corn hardness (Williams et al., 2016). To address the disadvantages of destructive hardness
measurement methods and the limitation in spatial dimension in NIR spectroscopy, hyperspectral imaging was utilized to classify hard (glassy), intermediate and soft (floury) corn kernels by applying spectral preprocessing, PCA and PLS-DA (Williams et al., 2009). NIR hyperspectral images were obtained from two HSI cameras: one with InGaAs diode array detector MatrixNIR camera in the 960-1662 nm spectral range, and one SisuCHEMA 2-D array HgCdTe detector short wave infrared (SWIR) push-broom system with 1000-2498 nm spectral range (Williams et al., 2009). PLS-DA models obtained from both cameras produced root mean square error of prediction of 0.18 – 0.29 (Williams et al., 2009).

Using the second instrument (SisuCHEMA SWIR) only, Manley et al. (2009) evaluated endosperm texture of whole yellow corn kernels. Aside from classifying the endosperm as vitreous or floury, they also found a third classification using PCA (Manley et al., 2009). The PCA score plot in the left side of Figure 6 shows that each hardness classification type was separated within one another which also agreed to the classification image of the kernels on the right (Manley et al., 2009). The success of both studies had encouraged another research to examine two different techniques for hardness classification: pixel-wise and object-wise, which were also reported to be useful in other applications (Williams et al., 2016). The authors were able to classify soft, medium, and hard kernels at 0.87 to 0.96 classification accuracies for both techniques (Williams et al., 2016). They reported that their system had an advantage of being suitable to either bulk or individual kernels which would be applicable in industry which requires bulk sample analysis and individual kernel identification (Williams et al., 2016).
Figure 6. Principal Component Analysis of grain hardness from hyperspectral imaging data (left) and the resulting classification image (right) (Manley et al., 2009).

**Processing parameters**

Pelleting is one of the processes done in feed manufacturing that directly affects feed quality, aside from contributing to energy consumption and throughput. In a study on biomass feedstocks, hyperspectral imaging (880-1720 nm spectral range) system was used to predict the specific energy required in pelleting, the moisture content of the agricultural feedstocks, and the feeding rate of the feedstocks into the pellet die (Gillespie et al., 2016). For moisture content, the PLSR model developed obtained $R^2$ of 0.94, RPD of 4.14, RMSECV of 1.11% and latent variables of 7 and these statistics were considered by Gillespie et al. (2016) as indication of a robust developed model which could be applied in any purpose. In prediction of electric consumption, the model, which produced $R^2$ of 0.64, RPD of 0.91 and RMSECV of 0.12 kWh/kg, was reported as suitable to use for screening purposes only (Gillespie et al., 2016). Then for the prediction of feed rate to pellet die, the developed model had $R^2$ of 0.70, RMSCEV of
0.20 kg/min and RPD of 1.81 (Gillespie et al., 2016). Despite the lower performance (lower R² and RPD) of the models for specific energy and feed rate than that of moisture content, these were still considered useful in screening and approximating predictions (Gillespie et al., 2016; Williams, 2014).

Another important operation in feed processing is mixing. The goal is to have a uniform mixture of ingredients that are accurately proportioned and as cost-effective as possible (Martin, 2005). The quality of mixing process, which may have an effect on the physico-chemical, rheological and nutritional properties of the product, is measured based on its homogeneity that is usually done using sample probes (Martin, 2005; Achata et al., 2018). This kind of invasive testing introduces sampling bias and may cause issues from agglomeration and segregation of powdered samples during mixing (Achata et al., 2018). For these reasons, Achata et al. explored the use of near infrared (880-1720 nm) and raman hyperspectral imaging combined to evaluate binary mixture in food powders (corn flour, icing sugar and mixtures of both) (Achata et al., 2018). Spectra with first derivative and SNV preprocessing produced the best PLSR model calibration statistics for prediction of sample concentrations: 0.7% RMSECV, 1.0% RMSEP, and 3 latent variables (Achata et al., 2018). Using the best NIR HSI model, prediction of corn flour concentration was performed on each pixel in the image for samples scanned after 0 s to 30 s of mixing (Achata et al., 2018). Corn flour mixture quality obtained a coefficient of variation (CV) lower than 6% after 10 seconds of mixing (Achata et al., 2018). This value is better than the feed industry index of 10% CV for a good mix (Martin, 2005). The visualization maps for prediction indicated that sample uniformity improves as mixing time increases (Achata et al., 2018). In this study, the
authors had proven that NIR HSI is slightly superior to raman HSI in this kind of application possibly due to bigger scan area in the former (Achata et al., 2018). For this technology to be used in the industry, additional validation and application in the commercial scale need to be done in the future (Achata et al., 2018).

Other published HSI applications related to post-harvest handling and primary processing of grains and oilseeds could also be applied to animal feed applications, such as the evaluation of milling quality in wheat flour (Delwiche et al., 2013), monitoring the diffusion of conditioning water in wheat kernels (Manley & Geladi, 2011), and classifying sound and heat-damaged soybeans during transportation and storage (Liu et al., 2020).

**Future Trends and Conclusion**

The presented studies have shown how near infrared hyperspectral imaging can be used non-destructively to assess quality and safety, determine properties, discriminate contaminants, and characterize animal feed products not only for safety and quality control but also for process optimization and improvement of operational efficiencies in a faster and less-laborious way than traditional analytical methods. Table 1 presents a summary of reviewed studies, highlighting the wavelengths covered, data analysis used and the obtained $R^2$ and classification accuracies.

Unlike near infrared spectroscopy, HSI is not yet considered a standard analytical tool due to its limitations and disadvantages, such as increased data processing load which adds to slow measurement speed, instrument limitations (i.e., noise), high price of the instrument itself, and data redundancy in the entire data set. These disadvantages have contributed to lower prediction accuracy and robustness of models, along with difficulty in pixel interpretation.
Table 1. Hyperspectral imaging applications for animal feed.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Product</th>
<th>Wavelength</th>
<th>Data analysis</th>
<th>$R^2$</th>
<th>% Accuracy*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adulteration and contamination</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fernández Pierna et al., 2004</td>
<td>compound feeds</td>
<td>900-1700</td>
<td>SVM, PLS, ANN</td>
<td>-</td>
<td>3-78</td>
</tr>
<tr>
<td>Fernández Pierna et al., 2006</td>
<td>compound feeds</td>
<td>900-1700</td>
<td>SVM</td>
<td>-</td>
<td>88-99</td>
</tr>
<tr>
<td>Fernández Pierna et al., 2012</td>
<td>cereals</td>
<td>900-1700</td>
<td>Quadratic SVM</td>
<td>-</td>
<td>&gt;95</td>
</tr>
<tr>
<td>Fernández Pierna et al., 2014</td>
<td>soybean meal</td>
<td>1118-2424</td>
<td>PCA, PLSDA</td>
<td>0.89-0.94</td>
<td>-</td>
</tr>
<tr>
<td>Garrido-Novell et al., 2018</td>
<td>protein meals</td>
<td>900-1700</td>
<td>PLS-DA, classification trees</td>
<td>-</td>
<td>90-92</td>
</tr>
<tr>
<td>Li et al., 2016</td>
<td>soybean meal</td>
<td>1282-2500</td>
<td>PLS-DA, PLSR</td>
<td>0.96-0.98</td>
<td>-</td>
</tr>
<tr>
<td>Riccioli et al., 2018</td>
<td>animal meals</td>
<td>900-1750</td>
<td>RF, SVM, PLS-DA, subspace discriminant</td>
<td>-</td>
<td>83-94</td>
</tr>
<tr>
<td>Shen et al., 2020</td>
<td>soybean meal</td>
<td>1100-2400</td>
<td>LAD, PLS-DA</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td>Zhao et al., 2018</td>
<td>wheat flour</td>
<td>936-1720</td>
<td>PLSR</td>
<td>0.96-0.99</td>
<td>-</td>
</tr>
<tr>
<td><strong>Biological contamination and damage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chelladurai, et al., 2014</td>
<td>soybean</td>
<td>900-1700</td>
<td>LDA, QDA</td>
<td>-</td>
<td>79-87</td>
</tr>
<tr>
<td>Chu et al., 2017</td>
<td>corn</td>
<td>850-2800</td>
<td>SVM, SVR</td>
<td>0.70-0.77</td>
<td>83-84</td>
</tr>
<tr>
<td>Chu et al., 2020</td>
<td>corn</td>
<td>900-1700</td>
<td>PCA, SPA, SVM</td>
<td>-</td>
<td>92-99</td>
</tr>
<tr>
<td>F. Zhu et al., 2016</td>
<td>corn</td>
<td>461-877</td>
<td>LS-SVM, KNN</td>
<td>-</td>
<td>95</td>
</tr>
<tr>
<td>J. Xing, et al., 2010</td>
<td>wheat</td>
<td>400-1000</td>
<td>PCA</td>
<td>-</td>
<td>94-100</td>
</tr>
<tr>
<td>Wang et al., 2014</td>
<td>corn</td>
<td>1000-2500</td>
<td>stepwise FDA</td>
<td>-</td>
<td>&gt;88</td>
</tr>
<tr>
<td>Wang et al., 2015</td>
<td>corn</td>
<td>1000-2500</td>
<td>PCA</td>
<td>-</td>
<td>92</td>
</tr>
<tr>
<td>Yao et al., 2010</td>
<td>corn</td>
<td>400-600</td>
<td>MLR, MANOVA</td>
<td>-</td>
<td>89-91</td>
</tr>
<tr>
<td>** Constituent concentrations and other quality parameters**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caporaso et al., 2018</td>
<td>wheat</td>
<td>980-2500</td>
<td>PLSR</td>
<td>0.62-0.79</td>
<td>-</td>
</tr>
<tr>
<td>Cogdill et al., 2004</td>
<td>corn</td>
<td>750-1090</td>
<td>PLSR, PCR, GA+MLR</td>
<td>0.56-0.87</td>
<td>-</td>
</tr>
<tr>
<td>Mahesh et al., 2015</td>
<td>wheat</td>
<td>960-1700</td>
<td>PLSR, PCR</td>
<td>0.57-0.88</td>
<td>-</td>
</tr>
<tr>
<td>Manley et al., 2009</td>
<td>corn</td>
<td>1000-2500</td>
<td>PLS-DA</td>
<td>&gt;0.85</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 1 Continued

<table>
<thead>
<tr>
<th>Reference</th>
<th>Product</th>
<th>Wavelength</th>
<th>Data analysis</th>
<th>$R^2$</th>
<th>% Accuracy*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ravikanth et al., 2015</td>
<td>wheat</td>
<td>1000-1600</td>
<td>NB, SVM, kNN</td>
<td>-</td>
<td>61-100</td>
</tr>
<tr>
<td>Vermeulen et al., 2017</td>
<td>cereal flour</td>
<td>1100-2400</td>
<td>PLS-DA, SVM</td>
<td>&gt;0.99</td>
<td>-</td>
</tr>
<tr>
<td>Weinstock et al., 2006</td>
<td>corn</td>
<td>950-1700</td>
<td>PLSR</td>
<td>0.68-0.76</td>
<td>-</td>
</tr>
<tr>
<td>Williams &amp; Kucheryavskiy, 2016</td>
<td>corn</td>
<td>975-2570</td>
<td>PCA, PLS-DA</td>
<td>-</td>
<td>89-96</td>
</tr>
<tr>
<td>Williams et al., 2009</td>
<td>corn</td>
<td>1000-2498</td>
<td>PLS-DA</td>
<td>0.76</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Processing parameters</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Achata et al., 2018</td>
<td>corn flour</td>
<td>880-1720</td>
<td>PCA, PLSR</td>
<td>0.997-1.000</td>
<td>-</td>
</tr>
<tr>
<td>Delwiche et al., 2013</td>
<td>wheat</td>
<td>1000-1700</td>
<td>MLR</td>
<td>&lt;0.60</td>
<td>-</td>
</tr>
<tr>
<td>Gillespie et al., 2016</td>
<td>feedstocks</td>
<td>880-1720</td>
<td>PCA, PLSR</td>
<td>0.64-0.94</td>
<td>-</td>
</tr>
<tr>
<td>Liu et al., 2020</td>
<td>soybeans</td>
<td>400-1000</td>
<td>ELM, RF</td>
<td>-</td>
<td>100</td>
</tr>
</tbody>
</table>

Abbreviations: LAD - local anomaly detection, PLS-DA - partial least squares discriminant analysis, PAPs - processed animal proteins, SPA - successive projection algorithm, MLR - multiple linear regression, MANOVA - multivariate analysis of variance, LDA - linear discriminant analysis, QDA - quadratic discriminant analysis, NB - Naïve Bayes (NB), SVM - Support vector machines, kNN - k-nearest neighbors, SCM - spectral correlation measurement, ICA - independent components analysis, FDA - factorial discriminant analysis, PCR - principal component regression, GA - genetic algorithm, ELM - extreme learning machine, RF - random forest

*Accuracy - classification, detection or prediction

Near-infrared hyperspectral imaging offers advantages with its ability to: measure low concentrations, evaluate the entire sample (instead of providing the average data), assess multi-constituent information, and map distribution of components over the surface of a sample. These will be useful for animal feed ingredient analysis and feed quality and safety control that are vital in feed processing operations. Therefore, there is a need to explore and maximize the potential of this technology in various applications, especially in the postharvest management and processing practices in the feed mill and livestock industries, to ensure safe and quality feed. Suggested future research areas are on instrument development, chemometric methods, and image analysis across
various products and processing operations. With its potential uses and applications on animal feed, near infrared hyperspectral imaging technique is a promising development in the feed industry.

References


CHAPTER 3. ANALYSIS OF PROTEIN AND OIL IN SOYBEAN MEAL USING NIR HYPERSPECTRAL IMAGING

Princess Tiffany Dantes\textsuperscript{1}, Charles R. Hurburgh\textsuperscript{2}, Thomas Brumm\textsuperscript{3}, Matthew John Darr\textsuperscript{2}, Dirk E. Maier\textsuperscript{2}, Ranjan Maitra\textsuperscript{2}

\textsuperscript{1}Graduate research assistant, Iowa State University, \textsuperscript{2}Professor, Iowa State University, \textsuperscript{3}Associate Professor, Iowa State University

Modified from a manuscript to be submitted to Journal of Spectral Imaging

Abstract

Soybean meal is the second most consumed animal feed ingredient next to corn. Protein and amino acid contents are the most important quality factors. Protein correlates well but not perfectly with amino acid composition. Non-uniform distribution and varying amino acid profiles have made protein content, and by inference amino acid contents, hard to measure. In this study, near-infrared (NIR) reflectance hyperspectral imaging (HSI) was used to predict protein and oil content of soybean meal and to show the distribution of protein content in soybean meal samples. Soybean meal samples (n=189) were scanned using a reflectance NIR HSI system in the 850 to 1700 nm wavelength. Preprocessed mean spectra and reference measurements were the input data in the partial least squares (PLS) regression for model calibration. The final model for protein achieved $R^2$, standard error of cross-validation (SECV) and relative predictive determinant (RPD) of 0.89, 0.65\% and 3.04, respectively. For oil, the final model obtained $R^2$ of 0.89, SECV of 0.46\%, and RPD of 2.95. Prediction results were satisfactory for protein content only. The final model was then used to visualize protein content distribution in a sample. The performance of the NIR HSI for soybean meal protein model validation/prediction was comparable with the two NIR single-point
spectrometers being widely used today. This study serves as a precursor to the use of imaging technology to measure amino acid composition (i.e., lysine, methionine, and cysteine) in soybean meal and other animal feed ingredients.

**Introduction**

Soybean meal (SBM), a by-product of solvent-extraction of oil in soybean, has an average protein content of 44-48% (as fed basis) (Balastreri et al., 2016; de Oliveira Silva & Perrone, 2015; Shen et al., 2016). It ranks as the highest in world protein meal production, imports and exports (USDA Foreign Agricultural Service, 2020). SBM is used as a major source of protein in animal feed for swine and poultry because of its high-quality protein and balanced amino acid profile, aside from being relatively inexpensive and readily available (Balastreri et al., 2016; de Oliveira Silva & Perrone, 2015; Krishnan & Jez, 2018). Protein content in SBM is an important quality factor used as a rapid estimator for amino acid composition in animal diet formulation.

Over the past few decades, near-infrared spectroscopy (NIRS) has been extensively used as a non-destructive, accurate and rapid routine tool to assess the quality parameters (such as moisture, protein and oil) of feed ingredients, and to substitute time-consuming and costly traditional laboratory methods (Blasco et al., 2020; ElMasry & Sun, 2010; Siesler, 2008). As a single-point spectroscopy, NIRS measures the average spectral data of the sample and the measured value may not always be the representative of the whole sample (Feng & Sun, 2012). It does not also provide spatial information – the composition at a specific location within the sample (Blasco et al., 2020; ElMasry & Sun, 2010).

Hyperspectral imaging (HSI) integrates spectroscopic (chemical composition) and spatial (imaging features) components in each point of the image, with full spectral
information (Blasco et al., 2020; ElMasry & Sun, 2010; Fu & Ying, 2014). Hyperspectral imaging produces a spectral image which is composed of three-dimensional (two for spatial and one for spectral) images of the same object which are stacked at various spectral bands (Fu & Ying, 2014). In HSI, ten to hundred images having contiguous wavelengths of large quantities are acquired, and every pixel in the acquired image can have a full spectrum that serves as the pixel’s fingerprint (Blasco et al., 2020; Wu & Sun, 2013). In near-infrared hyperspectral imaging (NIR HSI), overlapping occurs between broad bands (which describe the images) and other bands originating from various compounds in the same sample, thereby the pixels typically comprised of mixtures of varying compounds and overlapping spectral signatures (Marini & Amigo, 2020). The combination of chemometrics and image processing techniques are then used to analyzed NIR hyperspectral images.

NIR hyperspectral imaging has been used for different grain and animal feed applications: moisture and oil in single maize kernels (Cogdill et al., 2004), protein content in single wheat kernels (Caporaso et al., 2018b); oil and oleic acid in individual corn kernels (Weinstock et al., 2006); variety, texture and hardness classification of corn (Feng et al., 2019); variety and classes classification of rice, soybean and wheat (Feng et al., 2019); viability prediction in corn, wheat, barley and sorghum; insect damage detection in soybean and wheat (Feng et al., 2019); protein in wheat flour (Baeten et al., 2019); sprout damage detection in wheat kernels (Barbedo et al., 2018); and, melamine detection in soybean meal (Fernandez Pierna et al., 2014; Haughey et al., 2015). However, most of these
applications are focused on qualitative or quantitative for single kernels. Only few studies involve the use of NIR HSI on ground feed ingredients such as soybean meal.

Variability in soybean meal nutrient composition and digestibility (quality) occurs among samples due to differences in soybean varieties, geographic origins, production and postharvest conditions, and soybean meal processing methods, which may incur important effects in practical feed formulation (de Coca-Sinova et al., 2010; van Eys, 2015). As reported, the differences in processing of soybean meal is regarded as major influence on soybean meal quality (van Eys, 2015).

Hence, there is a need to explore the use of hyperspectral imaging in ground ingredients, which is highly variable in composition across the sample, to contribute to having an effective quality control and animal diet formulation.

The objective of this study was to develop an HSI calibration model for soybean meal protein and oil content, then show the distribution of protein content within the soybean meal. It also compared the performance of NIR hyperspectral imaging instrument to two single-point NIR spectrometers, used on the same samples.

**Methodology**

**Samples**

The soybean meal quickset (n = 184) of the Iowa State University (ISU) Grain Quality Laboratory were used for this study. Samples were divided into calibration set (nc = 164) and prediction set (np = 20). Soybean meal quickset is composed of samples collected from 21 processing plants from different locations (mostly in Iowa) in the U.S. between 2007 and 2012, with protein content (13% moisture basis) and oil (13% moisture basis) reference measurements from Eurofins laboratory (Des Moines, IA, U.S.A). Laboratory reproducibility is 0.44% and 0.21% for protein and oil, respectively.
Instrumentation

Hyperspectral NIR images were captured using a reflectance push-broom laboratory type SWIR microHSI™ hyperspectral imaging sensor from Corning Optics (Keene, NH, USA) with InGaAs detector array, frame rate of 86 Hz, spectral range from 850 to 1700 nm, 45 pixels/inch (~18 pixels/cm), and 9 nm spectral resolution. Each image contains 101 bands (wavelengths), 220 lines (rows), 512 samples (columns). The sample holder (cup) was placed on Zaber linear stage (Vancouver, British Columbia, Canada) beneath the camera. Image acquisition was controlled using the HyperC+ software (Version 3.03) by Corning (Keene, NH, USA). Image calibration was performed prior to the sample scans by scanning a dark image (recording data while covering the illumination source with its cap) and a gray Spectralon (reference material with 50% reflectance) from Avian Technologies (New London, NH, USA). Single-point reflectance NIR spectrometers, FOSS DS2500 (420–2500 nm) and Perten DA7200 (950–1500 nm), were used for comparison with the hyperspectral imaging.

Data analysis

Raw hyperspectral images (with raw spectral values) were acquired from HyperC+ and processed in ENVI 5.2 software (Harris Geospatial Solutions, Broomfield, CO, USA). Images were resized to form a square area due to the stretched output. Using the region of interest (ROI) tool in ENVI, one round (35-pixel diameter) region at the center of the scanned sample area was selected for each image to ensure that noise and shadows around the edges were not considered for analysis. Since the reference data available were for the entire sample, mean raw spectra were exported for each ROI. To convert these spectra to reflectance, raw spectral values were divided
by reflectance scale factor of 1000 (a value obtained from the header file of the hyperspectral image).

Absorbance values were then computed as the log10 of (1/reflectance) and statistically analyzed using The Unscrambler 11.0 software (Camo Analytics, Norway). Principal component analysis (PCA) was used for initial evaluation and identifying noise in the first and last few bands (wavelengths).

Absorbance spectral data were pre-processed in eight different ways: 1) standard normal variate (SNV), 2) multiplicative scatter correction (MSC), 3) first derivative Savitzky-Golay (SG1) (polynomial order 3, window points 5), 4) second derivative Savitzky-Golay (SG2) (polynomial order 3, smoothing points 5), 5) SNV + SG1, 6) SNV + SG2, 7) MSC + SG1, and 8) MSC + SG2. Derivatives are used for removal of baseline and enhancement of peaks, while SNV and MSC are used for normalization (AOCS, 2009).

Raw and pre-processed spectral data and their respective protein and oil reference measurements were the input data for protein and oil model calibration and validation. Partial least squares (PLS) regression was used in the analysis with the non-linear iterative partial least squares (NIPALS) algorithm. NIPALS, one of the methods used in providing the parameter estimates of the models in PLS regression, was originally made for PCA and is widely utilized for incomplete data (Nengsih et al., 2019). A step-by-step computation for NIPALS algorithm could be found from the study of Stott et al. (2017).

We used random cross-validation with 20 segments and 8-9 samples per segment. The optimum number of factors were determined based on the statistical
results: coefficient of determination – R², root mean square error - RMSE, standard error – SE, and relative predictive determinant - RPD. RPD is measured by dividing the standard deviation (SD) of reference data to the SE (of cross-validation [SECV] or prediction [SEP]). RPD for forages, feeds, soils, functionality factors, and other materials with complex physical properties was interpreted using Table 1 (Williams, 2014).

Table 1. Interpretation of RPD statistic (Williams, 2014).

<table>
<thead>
<tr>
<th>RPD value</th>
<th>Classification</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0–1.9</td>
<td>Very poor</td>
<td>Not recommended</td>
</tr>
<tr>
<td>2.0–2.9</td>
<td>Poor</td>
<td>Rouch screening</td>
</tr>
<tr>
<td>2.5–2.9</td>
<td>Fair</td>
<td>Screening</td>
</tr>
<tr>
<td>3.0–3.4</td>
<td>Good</td>
<td>Quality control</td>
</tr>
<tr>
<td>3.5–4.0</td>
<td>Very good</td>
<td>Process control</td>
</tr>
<tr>
<td>4.1+</td>
<td>Excellent</td>
<td>Any applications to this type of material</td>
</tr>
</tbody>
</table>

After model calibration and validation, regression coefficients were used to predict protein and oil content for a new set of soybean meal samples. Instead of using next year’s samples as prediction set, 20 soybean samples were randomly selected from the entire set to cover wide range of values.

Visualization of predicted protein content distribution was performed in Breeze 2.7.10 (Prediktera AB, Sweden) by applying the model to each pixel in the image. To enhance image clarity, a low level of smoothing for post-processing was applied on the pixels. Distribution maps were generated which comprised of the images and the corresponding colormap of the property being measured.
Comparison of protein PLS regression statistics (validation and prediction) among Corning NIR HSI, FOSS DS2500 and Perten DA7200 instruments was performed using (a) each instrument’s full wavelength range and (b) the common wavelength range of the three. Regression coefficient plots and scatterplots of predicted vs. reference were also created.

**Results and Discussion**

Reference measurements for protein content (12% moisture basis) of the samples were from 39.75–50.01% with a mean of 46.33% and standard deviation of 1.98% points. Oil reference data were in the range of 0.39–9.03% (12% moisture basis) with 1.63% points and 1.34% points as the mean and standard deviation, respectively.

**Spectral data analysis**

Mean spectra extracted from 164 hyperspectral images (obtained from the 164 calibration samples) of soybean meal are shown in a line plot in Figure 1a. The spectral lines followed a pattern typical for a soybean meal sample, such that absorption bands around 1200 nm and 1500 nm were found similar to those reported by Fernández Pierna et al. (2014) in their study. Offset and scattering effects between samples can be observed in the figure, while possible instrument noise is evident in the first and last wavelengths.

One sample that had the highest absorbance values looks like an outlier. This possible outlier and other observed differences were matched with the results in principal component analysis (PCA) and spectral pre-processing. Extreme values may indicate that it is a new sample that needs to be included in the model to cover wider range and predict such values in the future. This particular sample was from a processing plant in Des Moines, Iowa and it had relatively lower % oil than the other
samples from the same plant location and soybean origin in 2007, which might suggest an error in reference chemistry.

Figure 1. (a) Raw absorbance spectra (849-1704 nm) of soybean meal samples and its (b) PCA loadings (PC 1 vs. PC 2).

The loadings plot (PC 1 vs. PC 2) from PCA results shows that PC 2 divided the wavelengths into two groups as seen in Figure 1b. At the upper part of PC 2, data points with lower values (< 0.1) of PC 1 were the wavelengths of 849 nm to 866 nm. These were the first three wavelengths in Figure 1a that show upward trend (a break could be observed) to the leftmost part of the plot. The data points with higher PC 2 loadings (> 0.1) were from 1653 nm to 1704 nm wavelengths which were the last seven wavelengths in the spectra. Unlike the first three wavelengths, these last seven wavelengths could be easily noticed because of their abrupt shift downward. These unusual patterns could be due to instrument noise: detector saturation or blackout. The
10 wavelengths were removed. The spectral data to be used for further analysis were reduced to 91 wavelengths instead of 101 as seen in Figure 2a.

Figure 2. (a) Raw absorbance data with reduced wavelength range, and the results of pre-processing: (b) standard normal variate (SNV), (c) multiplicative scatter correction, (d) first derivative Savitzky-Golay (SG1), (e) second derivative Savitzky-Golay (SG2), (f) SNV and SG1, (g) SNV and SG2, (h) MSC and SG1, and (i) MSC and SG2.
Effect of spectral pre-processing

Line plots of preprocessed spectra are shown in Figure 2(b-i) where visual comparison can be made upon checking the raw (no preprocessing) spectral plot in Figure 2a as reference. The line that had the highest absorbance and looked like an outlier in the raw spectra could not be distinguished after preprocessing. Standard normal variate (SNV) (Figure 2b) and multiplicative scatter correction (Figure 2c), which are two commonly used preprocessing techniques in reduction of spectral distortions due to light scattering, seemed to have similar shape and suppressed scatter (Agelet & Hurburgh, 2010; Forchetti & Poppi, 2017; Jia et al., 2017). They resembled the pattern in the raw spectra but with smoothing and less noise. MSC is reported to be used in most diffuse reflectance applications for powder samples such as milk (Forchetti & Poppi, 2017). Both first and second derivative Savitzky-Golay (polynomial order: 3; smoothing points: 5) line plots reveal specific peaks of the spectral data that pertained to overtones and combination of molecular bands comprising the properties of soybean meal. In NIRS, overlapping peaks is a common problem and this is usually resolved with the use of derivatives (Manley, 2014). Partial least squares (PLS) regression statistics for each pretreatment are presented in Table 2.

Protein content prediction

In this study, the PLS regressions models for protein using different preprocessing methods obtained $R^2$ of 0.86 to 0.89, standard error of cross-validation (SECV) of 0.65 to 0.74%, and RPD (ratio of standard deviation of the reference values to the standard error of cross-validation or prediction) of 2.69 to 3.04. Single spot NIR models for soybean meal protein in other studies (Balasterri et al., 2016; Fontaine et al., 2001) typically have $R^2 > 0.95$ and SECV < 0.50%. Aside from being measured using
NIR hyperspectral imaging, the difference between this study and other published data could also be due to origin of samples, range of measured values, variability in soybean meal composition, and particle size distribution. Although not mentioned in their studies, Balasterri et al. (2016) and Fontaine et al. (2001) might have obtained their samples from Brazil and Germany, respectively, where their laboratories were based. In our study, soybean samples were from different processing plants in the U.S. In general, the models produced in this study were of relatively better performance than other hyperspectral imaging PLS regression models for grain protein (Caporaso et al., 2018b; Cheng et al., 2018).

Table 2. Partial least squares (PLS) regression validation statistics for protein and oil according to different preprocessing techniques.

<table>
<thead>
<tr>
<th>Preprocessing</th>
<th>None</th>
<th>SNV</th>
<th>MSC</th>
<th>SG1</th>
<th>SG2</th>
<th>SNV-SG1</th>
<th>SNV-SG2</th>
<th>MSC-SG1</th>
<th>MSC-SG2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protein</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.88</td>
<td>0.89</td>
<td>0.89</td>
<td>0.87</td>
<td>0.86</td>
<td><strong>0.89</strong></td>
<td>0.88</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>SECV</td>
<td>0.70</td>
<td>0.67</td>
<td>0.67</td>
<td>0.72</td>
<td>0.74</td>
<td><strong>0.65</strong></td>
<td>0.68</td>
<td>0.66</td>
<td>0.69</td>
</tr>
<tr>
<td>RPD</td>
<td>2.83</td>
<td>2.97</td>
<td>2.97</td>
<td>2.76</td>
<td>2.69</td>
<td><strong>3.04</strong></td>
<td>2.91</td>
<td>3.01</td>
<td>2.89</td>
</tr>
<tr>
<td>factors</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td><strong>Oil</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.86</td>
<td>0.88</td>
<td>0.88</td>
<td>0.86</td>
<td>0.87</td>
<td>0.88</td>
<td><strong>0.89</strong></td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>SECV</td>
<td>0.51</td>
<td>0.47</td>
<td>0.47</td>
<td>0.51</td>
<td>0.49</td>
<td>0.47</td>
<td><strong>0.46</strong></td>
<td>0.47</td>
<td>0.46</td>
</tr>
<tr>
<td>RPD</td>
<td>2.62</td>
<td>2.83</td>
<td>2.85</td>
<td>2.65</td>
<td>2.77</td>
<td>2.84</td>
<td><strong>2.95</strong></td>
<td>2.86</td>
<td>2.93</td>
</tr>
<tr>
<td>factors</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

*Values in bold indicate they are from the chosen pre-processing method

*SECV – standard error of cross-validation, %; RPD – relative predictive determinant; $R^2$ – determination coefficient for cross-validation; SNV – standard normal variate; MSC - multiplicative scatter correction; SG1 - first derivative Savitzky-Golay; SG2 - second derivative Savitzky-Golay.

Among the four preprocessing methods, SNV obtained the best performance in PLS regression with $R^2$ of 0.89, RMSECV of 0.66% and RPD of 2.97. MSC had almost the same values. This result coincides with the visual assessment and with other NIRS
reports that MSC and SNV usually produced similar outputs. SG1 and SG2 did not improve the model statistics except for the factors used. In this case, SNV and MSC were able to reduce variability within samples due to scattering effect. They are also designed to remove multiplicative interferences of particle size (Xu & Sun, 2018). One reason that the SG derivatives did not work well could be that SG derivatives are effective in reducing additive effects but the soybean meal spectra in this study had more multiplicative effects instead of additive (Xu & Sun, 2018). Additive and multiplicative scattering effects on spectra could be caused by physical factors such as non-uniform lighting, sample particle size and presentation angle, and surface inhomogeneities (Gowen & Burger, 2008; Caporaso et al., 2018a). In our study, scattering could be due to only one light source used for illumination and the varying particle sizes of the samples.

When used with Savitzky-Golay first (SG1) and second (SG2) derivatives, SNV combined with SG1 improved the validation statistics more than SNV alone or than SNV-SG2. The SNV-SG1 combination produced a lower SECV of 0.65%, an increased RPD of 3.04 and a reduced number of factors of 7. Therefore, SNV-SG1 was chosen as the optimum pre-processing method.

In the study of HSI on determination of protein in single wheat kernels, the SNV and SG1 combination also produced the best results against the other preprocessing methods used (normalisation, MSC, SNV, SNV–de-trend, and SG2) (Caporaso et al., 2018b). Interestingly, SNV–SG1 combination was also the most effective in the study of NIR HSI on packaging film but the order of applying preprocessing made a difference, such that SG1 before SNV was better than SG1 after SNV (Gowen et al., 2010). In their
study, SG1 after SNV led to overlapped clusters and undesirable removal of variability caused by inherent sample differences (Gowen et al., 2010).

In terms of outliers, no samples were removed for protein content calibration. Some suspected outliers, as suggested by Hotelling’s $T^2$ limits and leverage plots, had low residuals and when removed from the model, they contributed to no improvement to the prediction model. Samples with relatively high residuals had low leverage and were not considered outliers in the influence plot. These were the samples with relatively high protein values but they still follow the regression line. The scatterplot and regression statistics for the final protein model are illustrated in Figure 3.

Using the regression coefficients and the preprocessing methods in this model, the protein content values of 20 soybean samples were predicted with $R^2$ of 0.87, standard error of prediction (SEP) of 0.59% with seven factors (Figure 4). The final RPD in cross-validation of 3.04 can be classified as good and can be used for quality control applications (Williams, 2014).

**Oil content prediction**

Results for oil model calibration in soybean meal were of comparable performance with those of protein. Using the same set of pre-processing techniques, the validation statistics (Table 2) indicated 0.86–0.89 $R^2$, 0.46–0.51% SECV, 2.62–2.95 RPD, and 4–9 factors. With respect to pre-processing, SNV and MSC were still better than SG1 and SG2. One notable difference in results was the model performance of SNV - SG2 pre-processing combination which provided the highest $R^2$, lowest SECV, highest RPD, and least number of factors. The inconsistency with the pre-processing results between protein and oil models is typical in most NIRS and NIR HSI.
Figure 3. (a) Scatter plot of final PLS regression model for soybean meal protein and (b) the regression coefficients.

(a) 
\[ R^2 = 0.89 \]
\[ SEP = 0.65 \]
\[ RPD = 3.04 \]
\[ n = 164 \]
\[ factors = 7 \]
Two samples (sample 10 and sample 62) were considered as suspected outliers and checked for leverage, influence, and residuals. They performed well in the protein model and neither of the two were initially considered outlier in the raw spectra. They had high residuals and one of them (sample 10) had high leverage and high influence. Upon checking origin and reference chemistry, these two samples had the highest oil content among the rest of the samples from their respective processing plant and collection point of the same year. This could mean that the problem was the percent oil reference measurements of these two samples. After removing them from the calibration model, the performance had improved in both the cross-validation and prediction statistics as presented in Table 3. The final oil calibration model (Figure 5) had $R^2$ of 0.90, SECV of 0.42%, and RPD of 3.15 using 4 factors and 162 samples, which were much better than the protein model statistics. For the oil prediction, removal
of outliers led to increased $R^2$ from 0.43 to 0.50, reduced SEP from 0.35 to 0.33% and increased RPD from 1.34 to 1.39. Still, despite the improvement with outlier removal, the model performance in predicting oil content (Figure 6) in soybean meal was not as good as that of protein.

In studies involving NIR hyperspectral imaging and pre-processing methods to predict oil content in single maize kernels, one study using a transmittance NIR HSI instrument (750 to 1090 nm) reported $R^2 < 0.56$, SECV > 1.09% and RPD < 1.46 (Cogdill et al., 2004) and indicated that the poor model calibration results obtained could be attributed to the reference method. On the other hand in reflectance mode, another study used NIR HSI (950 – 1700 nm) for oil prediction and obtained $R^2 < 0.76$, RMSECV > 0.58% and RMSEP > 0.74% (Weinstock et al., 2006). Possible sources of errors could be instrument design, specifically sample area and lighting (Weinstock et al., 2006). In our study, the set-up uses only one lighting source that contributes shadow to one side of the image.

Table 3. Cross-validation and prediction statistics before and after the removal of outliers for the soybean meal oil PLS regression model preprocessed with standard normal variate and second derivative Savitzky-Golay.

<table>
<thead>
<tr>
<th></th>
<th>Without outlier removal</th>
<th>With outlier removal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cross-validation</td>
<td>Prediction</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.89</td>
<td>0.43</td>
</tr>
<tr>
<td>SE</td>
<td>0.46</td>
<td>0.35</td>
</tr>
<tr>
<td>RPD</td>
<td>2.94</td>
<td>1.34</td>
</tr>
<tr>
<td>factors</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>n</td>
<td>164</td>
<td>20</td>
</tr>
</tbody>
</table>

*R$^2$ – coefficient of determination; SE – standard error, %; RPD – relative predictive determinant; n – number of samples.*
Figure 5. (a) Scatterplot of predicted and reference oil content in soybean meal and (b) the corresponding regression coefficients of the final calibration model.

Figure 6. Predicted vs. reference % oil content (12% moisture basis) of soybean meal samples from the prediction (test/validation) set.

**Visualization of predicted protein distribution**

Figure 7 shows hyperspectral images of five soybean meal samples in (a) raw form, with (b) visualization of protein at the pixel level with the predicted % protein value, (c) with low level of smoothing of images in (b), and the corresponding % protein
color map that interprets the colors in (b) and (c). The color map indicates that spots close to red are high in protein (upper limit of 50%) and those close to blue are low in protein (lower limit of 40%). These distributions maps were created to show the spatial variability (pixel basis) of protein contents due to differences in soybean meal processing, origin, and particle size, among others, which could not be visible using the conventional NIR spectroscopy.

These images show the uneven distribution of protein over the entire sample. A sample which is high in protein may contain spots of very low protein that can affect blend uniformity and influence the values in animal diet formulation.

Figure 7. Distribution maps of predicted protein content in five soybean meal samples.
Comparison with commercial NIR instruments

Plotting the spectra of soybean meal sample 1 from three instruments, NIR HSI and two commercially available NIR single-point spectrometers, using their respective full wavelength range shows how each instrument was similar or different with the others (Baeten et al., 2019). It can be seen in Figure 8a that the Corning NIR HSI had wider wavelength range (850 – 1700 nm) than the Perten DA7200 (950 – 1500 nm) but shorter than the FOSS DS2500 (420 – 2500 nm). A slight gap can also be observed between Corning and the other two at around 1500 nm peak. Upon looking at another SBM sample's spectra from the three instruments and considering only their common wavelength range (Figure 8b), the Corning spectra was almost in line with the FOSS spectra at 1150 nm and higher than the Perten spectra. At the 1500 nm, all three instruments were separated such that FOSS had the highest absorbance, then followed by Perten and the lowest one was from Corning.

PLS regression results of the three instruments using full wavelength range and the common wavelength range are reported in Table 4 and Table 5, respectively. The NIR HSI instrument used in this study performed comparably to existing single-point NIR spectrometers used in the industry. When full wavelength range is used, the Corning microHSI™ had validation and prediction statistics around the same as the FOSS DS2500 and lower in performance than the Perten DA7200. When the common wavelength range (950 to 1650 nm) was used, validation statistics of the Corning instrument did not improve except for the number of factors from 12 to 10, but the prediction were the lowest in performance among the three.
Figure 8. Comparison of spectra from three instruments (Corning microHSI™, FOSS DS2500 and Perten DA7200) using (a) full wavelength of each instrument in sample 1, and (b) common wavelength range of three instruments in sample 2.

Table 4. PLS regression for soybean meal protein model cross-validation and prediction statistics of three instruments (Corning microHSI™, FOSS DS2500 and Perten DA7200) using their full wavelength range.

<table>
<thead>
<tr>
<th>Cross-validation</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Corning microHSI™</td>
</tr>
<tr>
<td></td>
<td>Corning microHSI™</td>
</tr>
<tr>
<td>R²</td>
<td>0.87</td>
</tr>
<tr>
<td>SE</td>
<td>0.73</td>
</tr>
<tr>
<td>RPD</td>
<td>2.71</td>
</tr>
<tr>
<td>factors</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 5. PLS regression for soybean meal protein model cross-validation and prediction statistics of three instruments (Corning microHSI™, FOSS DS2500 and Perten DA7200) using their common wavelength range.

<table>
<thead>
<tr>
<th>Cross-validation</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Corning microHSI™</td>
</tr>
<tr>
<td></td>
<td>Corning microHSI™</td>
</tr>
<tr>
<td>R²</td>
<td>0.87</td>
</tr>
<tr>
<td>SE</td>
<td>0.72</td>
</tr>
<tr>
<td>RPD</td>
<td>2.76</td>
</tr>
<tr>
<td>factors</td>
<td>10</td>
</tr>
</tbody>
</table>

*R² – coefficient of determination; SE – standard error, %; RPD – relative predictive determinant; n – number of samples.
Regression coefficients of the three instruments in the common wavelength range as shown in plots in

Figure 9 reveal that all three models have some similarities in the contributing variables, specifically at 1200 nm and 1500 nm. One reason that Perten DA7200 is the superior in model performance here could be the wavelength range and the fewer number of variables used by the instrument. The Corning NIR HSI instrument could be improved to limit the range or increase the number of bands or to have a bigger pixel group depending on the application to optimize its performance.

**Conclusion**

From the results of this study, we were able to conclude that: a) near-infrared hyperspectral imaging could be applied to ground animal feed ingredients, particularly soybean meal and DDGS, for prediction of protein content and possibly other quality parameters; b) the findings from this study may also suggest that these models for protein and oil content prediction still need to be improved whether in terms of spatial and image preprocessing, dimensionality reduction due to large and redundant data in hyperspectral images, hyperspectral data analysis or sample type consistency; c) the differences in $R^2$ and standard error among the different preprocessing methods may be small and might not have much effect. In a bigger scale such as in industrial application with large batches of samples being analyzed continuously, these small differences in standard error may lead to high losses in quality and profits; and, d) Hyperspectral imaging technique, when applied in the animal feed industry for feed processing operations, could be used in achieving accurate amount of animal feed ingredients that could save cost, minimize resources and contribute to optimizing feed efficiency and animal nutrition.
Figure 9. Regression coefficients of Corning microHSTM, FOSS DS2500 and Perten DA7200 for soybean meal protein content using the common wavelength range (950 - 1650 nm) and with no spectral pre-processing.
References


CHAPTER 4. NEAR-INFRARED HYPERSPECTRAL IMAGING FOR DETERMINATION OF LYSINE CONCENTRATION IN ANIMAL FEED INGREDIENTS

Princess Tiffany Dantes¹, Charles R. Hurburgh², Thomas Brumm³, Matthew John Darr², Dirk E. Maier², Ranjan Maitra²

¹Graduate research assistant, Iowa State University, ²Professor, Iowa State University, ³Associate Professor, Iowa State University

Modified from a manuscript to be submitted to Cereal Chemistry

Abstract

Knowing the amount of available amino acids in animal feed will increase formulation efficiency in animal diets and will contribute to sustainable use of feed materials, effective quality control, reduced cost in production, and appropriate pricing for feed ingredients. In addition to the labor-intensive and slow chromatographic methods, non-destructive methods, such as near-infrared spectroscopy (NIRS) and regression equations based on crude protein, are currently used to estimate amino acid concentration. Both alternative methods are reported to depend on the protein content and actually estimate protein instead of amino acids. Near-infrared hyperspectral imaging (NIR HSI) is a combination of NIRS and machine vision. It is being used in food and feed quality and safety applications, particularly in detection and classification. The ability of NIR HSI to produce spectra on a pixel by pixel basis provides an advantage over the conventional NIRS, specifically in products which are highly variable and of low concentration such as amino acids in feed ingredients. Lysine concentration was determined in two different major feed ingredients (soybean meal and dried distiller’s grains with solubles [DDGS]) using NIR hyperspectral imaging combined with either partial least squares (PLS) regression or spectral angle mapper (SAM) classification.
Different concentrations of pure lysine were added to the samples and calibration models were developed. PLS regression obtained $R^2$ of 0.98 (soybean meal) to 0.99 (DDGS) and standard error of cross-validation (SECV) of 0.05% (DDGS) to 0.21% (soybean meal). Using SAM classification, $R^2$ values were 0.99 for soybean meal and 0.95 for DDGS at SAM maximum angle threshold of 0.10 rad, and 0.59 (soybean meal) and 0.38 (DDGS) at SAM maximum angle threshold of 0.05 rad. Standard error of prediction (SEP) using both methods were 0.52–0.55%. Results from this study showed that NIR hyperspectral imaging could be used to predict lysine concentration in animal feed ingredients.

**Introduction**

With the increasing global human population and the subsequent demand for animal-source foods, animal production has been continuously growing along with the need for high quality protein-rich animal feed ingredients (Jansman, 2016). Animal feed contains protein as needed by species. Its constituent essential amino acids are needed by the animals for growth, health, performance, and quality protein of food products (Jansman, 2016).

Most animal diet formulations are based on crude protein and amino acid contents, and the restrictions are on the most limiting indispensable amino acid (de Coca-Sinova et al., 2010; Siquera et al., 2013). The amino acid that is utilized as the reference in expressing the ideal protein concept is lysine, which is nearly related with body protein deposition potential (de Coca-Sinova et al., 2010; Siquera et al., 2013). In this case, the concentrations of the rest of the essential amino acids will be modified if there is a change in the lysine concentration. Thus, it is essential that the amino acid requirement, particularly lysine in a swine feed formulation is provided accurately (de
Coca-Sinova et al., 2010; Siquera et al., 2013). Knowing the amount of available amino acids in the feed ingredients and obtaining rapid reliable values contribute not only to increased formulation efficiency in animal diets but also to sustainable use of feed materials, effective quality control, reduced cost in production, and appropriate pricing for feed ingredients (Bryden and Li, 2010; Cemin et al., 2017; Ravindran et al., 2017; Chen et al., 2011; van Kempen, 1996).

Determination of amino acid concentration is typically done by chromatographic methods, which involve oxidation, hydrolysis, and ion exchange chromatography (Siquera et al., 2013; Fontaine et al., 2001). Such methods require skilled personnel and complicated sample preparation, may take days to produce results, and can be costly when used in routine feed formulations (Fontaine et al., 2001; Kovalenko, et al., 2006). Two methods that are fast and non-destructive are now being used by feed companies to estimate amino acid concentrations, (a) regression equations based on total crude protein and (b) prediction by near-infrared spectroscopy.

Near-infrared spectroscopy (NIRS), an analytical technique based on the interaction of light with the molecules of the sample composition, has long been successfully used for rapid and accurate analysis of animal feed, mostly moisture and protein contents. It requires no sample preparation and can be applied for at-, on-, or in-line feed processing operations. NIRS calibration models for prediction of amino acid contents in wheat and barley were first developed in 1978 and 1979 (Fontaine, 2003). Since then, the application of NIRS on amino acid contents has continued and expanded to cereals, grains, oilseeds, bran, meals and other feed raw materials for the purpose of replacing the wet chemistry procedures (Fontaine, 2003). Many studies were
able to produce satisfactory results for lysine prediction such as in rapeseed meal (Chen et al., 2011), ground processed animal proteins (de la Haba et al., 2006), DDGS (Zhou et al., 2012), mixed poultry feeds (Bastianelli et al., 2005), soybeans (Kovalenko et al., 2006), soybeans and soybean meal (Fontaine et al., 2001), wheat (Fontaine et al., 2002), barley (Fontaine et al., 2002), and corn (Fontaine et al., 2002) as presented in Table 1.

Table 1. Studies using near-infrared spectroscopy for lysine prediction in feed ingredients.

<table>
<thead>
<tr>
<th>Product</th>
<th>1-VR</th>
<th>R²</th>
<th>SECV, %</th>
<th>RPD</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapeseed meal</td>
<td>0.96</td>
<td>-</td>
<td>0.119</td>
<td>4.86</td>
<td>Chen et al., 2011</td>
</tr>
<tr>
<td>Ground processed animal proteins</td>
<td>0.93</td>
<td>-</td>
<td>0.110</td>
<td>3.82</td>
<td>de la Haba et al., 2006</td>
</tr>
<tr>
<td>DDGS</td>
<td>-</td>
<td>0.82</td>
<td>0.044</td>
<td>2.35</td>
<td>Zhou et al., 2012</td>
</tr>
<tr>
<td>Mixed poultry feeds</td>
<td>-</td>
<td>0.92</td>
<td>0.066</td>
<td>2.58</td>
<td>Bastianelli et al., 2005</td>
</tr>
<tr>
<td>Soybeans</td>
<td>-</td>
<td>-</td>
<td>0.070 – 0.090ᵃ</td>
<td>2.45 – 2.96</td>
<td>Kovalenko, Rippke &amp; Hurburgh, 2006</td>
</tr>
<tr>
<td>Soybeans and soybean meal</td>
<td>0.93</td>
<td>-</td>
<td>0.083</td>
<td>-</td>
<td>Fontaine et al., 2001</td>
</tr>
<tr>
<td>Wheat</td>
<td>0.84</td>
<td>-</td>
<td>0.015</td>
<td>-</td>
<td>Fontaine et al., 2002</td>
</tr>
<tr>
<td>Barley</td>
<td>0.86</td>
<td>-</td>
<td>0.015</td>
<td>-</td>
<td>Fontaine et al., 2002</td>
</tr>
<tr>
<td>Corn</td>
<td>0.77</td>
<td>-</td>
<td>0.013</td>
<td>-</td>
<td>Fontaine et al., 2002</td>
</tr>
</tbody>
</table>

*1-VR – fraction of explained variance for cross-validation; R² – coefficient of determination; SECV – standard error of cross-validation; RPD – relative predictive determinant; DB – dry basis
ᵃStandard error of prediction

However, at present, NIRS is still not considered a standard tool for amino acid analysis. The ability of NIR models to predict has been dependent on the correlation of the amino acid to protein. The calibrations effectively predicted protein and not amino acid itself (Kovalenko et al., 2006). As reported, the major difficulty in evaluating amino acid contents in soybean and grains using NIRS lies in exceeding the amino acid and protein correlation or in gathering sets of the samples that would disrupt the correlation (Kovalenko et al., 2006). In the study of Zhou et al. (2012) on DDGS, they reported that protein was weakly correlated with some amino acids and that the NIRS models...
developed performed better than the crude protein regression, but still the models for lysine and other major amino acids showed unsatisfactory results, probably due to comparatively low lysine concentrations. Moreover, NIRS, being considered as a single-point spectroscopy, gives the average value of the quality parameter and the spectra may not always be a representative of the entire sample (Riccioli et al., 2018; Feng & Sun, 2012). Due to processing, grain origin and raw material variation, amino acid composition in a feed sample is not evenly distributed. It may have low concentration and in wide variability all throughout the sample.

Raw grains such as whole soybeans and whole corn kernels may have uniform composition specific to their inherent material properties. Processed grain products (i.e., DDGS and soybean meal) may be sourced from grains with the same protein level but with varying composition. Extrusion, heat-treatment, and other grain processing operations could change the protein content and digestibility of grain products (Beloshapka et al., 2016). These may lead to the animal feed ingredients being more variable in composition as opposed to raw whole grains.

Near-infrared hyperspectral imaging (NIR HSI), a combination of NIRS and machine vision, provides spectral and spatial characteristics of a sample (Lu & Park, 2015). One of its advantages over normal NIRS is its ability to produce spectra (information) at each pixel in the image, in a way that the spatial distribution of the various chemical compounds in the sample can be visualized simultaneously (ElMasry & Sun, 2010). NIR HSI has already been applied to evaluate quality and safety parameters in different food and agricultural products.
In grains and oilseeds, several researchers have used NIR HSI for applications related to classification and prediction (Caporaso et al., 2018). In spite of the different studies already done, there has been no study on the use of NIR hyperspectral imaging in quantifying lysine or any amino acid concentration in animal feed. The ability of NIR HSI to produce spectra on a pixel by pixel basis provides an advantage over the conventional NIR, specifically in products which are highly variable and of low concentration such as amino acids in feed ingredients. Therefore, the aim of this study was to determine lysine concentration in two major animal feed ingredients (soybean meal and dried distiller's grains with solubles) using NIR hyperspectral imaging.

**Materials and Methods**

**Sample preparation**

Soybean meal (SBM) and DDGS, which were pooled from ISU Grain Quality Laboratory standardization samples, were passed through 0.5 mm sieves to obtain uniform particle size, homogenize the mixture, and minimize unfavorable spectral scattering due to particle size variation. Lysine (≥ 98%, Sigma-Aldrich, Atlanta, GA, USA), 5 g SBM and 5 g DDGS were prepared for analysis. For lysine-SBM mixtures, lysine was added in 2.0%, 2.5%, 3.0%, 3.5%, and 4.0% concentrations. For lysine-DDGS mixtures, lysine was added in 0.25%, 0.50%, 0.75%, 1.00%, and 1.25% concentrations. The addition of pure amino acids to animal feed ingredients was pointed out by Bastianelli et al. (2005) as a strategy to have calibration models that were not based on the relationship between protein and amino acids and that would resist qualitative differences among samples.

Reference % lysine (w/w) concentration (at 12% moisture) of soybean samples in the prediction set were from a University of Missouri laboratory while reference %
protein (at 12% moisture) were from Eurofins Laboratory (Des Moines, IA, USA). There were 10 soybean meal samples in the prediction set and no DDGS samples due to unavailability of reference chemistry data.

**Acquisition of hyperspectral data**

Three sets of samples were scanned and analyzed in this study. The first set was composed of pure lysine, SBM and DDGS which were placed in one sampling cup and scanned. Their spectra were saved in the spectral library for reference in the later analysis. The second set was the SBM-lysine mixtures and DDGS-lysine mixtures where each concentration was mixed and scanned three times. The third set consisted of the 10 soybean meal samples used for model prediction.

In our study, a reflectance push-broom laboratory type SWIR microHSI™ hyperspectral imaging sensor from Corning Optics (Keene, NH, USA) was utilized to capture hyperspectral images from 850 to 1700 nm (with 101 bands or wavelengths). The instrument has frame rate of 86 Hz, 45 pixels/inch (~18 pixels/cm), 9 nm spectral resolution, and 1000 reflectance scale factor. It is equipped with a camera that has 512-by 220-pixel Indium Gallium Arsenide array detector and a linear stage (Zaber, Vancouver, British Columbia, Canada) which holds the sampling cup. The HyperC+ software (Version 3.03) by Corning Optics (Keene, NH, USA) controlled the image calibration, image acquisition and camera settings. Before scanning the samples, image calibration was done using a dark reference (captured by covering the light source with its cap) and a gray Spectralon (Avian Technologies, New London, NH, USA) (reference material with 50% reflectance).
Data analysis

Hyperspectral images were processed in ENVI (Environment for Visualizing Images) 5.2 software (Harris Geospatial Solutions, Broomfield, CO, USA). Two methods of data analysis were used: spectral angle mapper classification and partial least squares (PLS) regression. Before proceeding with SAM or PLS regression, spectral data of pure lysine, SBM and DDGS were analyzed by extracting 1 pixel and 15 pixels from each material. Spectra were examined in line plots and in principal component analysis (PCA) score plots.

Spectral angle mapping classification

Spectral angle mapper (SAM) is one of the most effective supervised classification methods (Cucci & Casini, 2020). In SAM algorithm (Figure 1), the spectra of individual pixels serve as vectors in the n-dimensional space of wavebands, where n represents the number of wavelengths (ElMasry & Nakauchi, 2016). These spectra were grouped by computing their difference in angle relative to the reference spectra of known materials (Cucci & Casini, 2020). The smaller the angle, the closer is the match between the spectra and the reference (Park et al., 2007). To analyze using SAM, the target detection tool in ENVI was used to measure the angle between reference spectra (pure lysine spectra obtained from first set) and predicted spectra (SBM-lysine mixture or DDGS-lysine mixture). The tool counts the number of pixels in the sample which correspond to the reference spectra. The maximum angle which serves as the threshold could be adjusted and two angles were used in this study: 0.05 and 0.10 rad. Spectral data was reduced to limit spectra from 875 nm to 1644 nm since the first 3 and last 7 wavelengths were previously identified to contain noise. Linear regression equations were built using the number of detected pixels and theoretical % concentration.
Coefficient of determination $R^2$ and correlation coefficient $r$ were calculated. For model prediction, the detected number of lysine pixels from the prediction sample set were used in the developed regression equation. Residuals were calculated to evaluate the model.

Partial least squares regression

In NIR spectroscopy, partial least squares (PLS) regression has been proven superior in performance particularly in quantifying purposes. For PLS regression in this study, a region of interest (ROI) (45-pixel by 90-pixel ellipse) was selected in each hyperspectral image and mean spectra was measured. To obtain reflectance values, raw spectral values were divided by the reflectance scale factor of the instrument which is 1000. Absorbance values were calculated as the log of $(1/\text{reflectance})$. Absorbance spectral data were preprocessed with standard normal variate (SNV), the preprocessing method that is effective in minimizing scattering effects in most published studies and was most useful earlier. Moving average was also used for visualization purposes. SNV
preprocessed mean spectra of different concentrations and the reference concentrations were used as input data for the PLS regression. Similar to what was done in SAM, only the data within the 875 nm to 1644 nm wavelength range were used in the analysis. To evaluate PLS regression results, \( R^2 \) and standard error of cross-validation were examined. Analysis of spectral data, PLS regression and model prediction were performed in The Unscrambler 11.0 (Camo Analytics, Norway). Figure 2 summarizes the major steps done in this study.

Figure 2. Schematic diagram summarizing the methodology for prediction of % lysine concentration in feed ingredients spiked with pure lysine.
Results and Discussion

NIR reflectance data characteristics

From the first set of scanned samples, a spectrum of each material, SBM, DDGS and lysine, is presented in Figure 3 to visually compare how the spectra deviate from one another. It shows that SBM and DDGS followed the same trend, although DDGS seemed to have higher absorbance values than SBM. This is in agreement with the results of Balastreri et al. (2016) who analyzed soybean meal and maize bran (similar with DDGS that originated from corn) and explained that the result could be attributed to differences in processing.

Among the three materials, lysine had the lowest absorbance values. One thing in common was the peak of the three samples at around 1200 nm. This common peak became pronounced when the spectra were preprocessed with standard normal variate and moving average smoothing as shown in Figure 3a. At certain wavelengths, data points for SBM and DDGS were almost the same. However, the three lines could easily be distinguished from 874 to 1070 nm range. In the PCA score plot (Figure 4) of 15 pixels from each sample, lysine could be easily distinguished from other animal feed ingredients using two principal components (PCs). PC1 explained 98% and 95% of the variation without preprocessing and with preprocessing, respectively. These score plots in Figure 4 thus demonstrate that PCA could be used for initial screening and identification of lysine and possibly other amino acids. This would be beneficial when receiving raw materials and feed ingredients in the feed mill.
Figure 3. Absorbance spectra of soybean meal (SBM), dried distillers' grains with solubles (DDGS) and lysine (a) without preprocessing and (b) with preprocessing by standard normal variate and moving average smoothing.
Figure 4. Principal component (PC) scores plot (PC 1 vs PC 2) of DDGS, lysine and soybean meal spectra (a) without preprocessing and (b) with preprocessing by SNV and moving average smoothing.
Partial least squares regression

Calibration

Results for PLS regression of reference against predicted % added lysine concentration in soybean meal-lysine mixture are presented in Figure 5. The scatterplot in Figure 3a indicated high $R^2$ of 0.98 and low SECV of 0.21% at factor 2. These results are also presented in Table 1. These $R^2$ and SECV were higher than what were reported by Fontaine et al., in 2001 using NIRS on soybeans and soybean meal (0.93 1-VR and 0.083% SECV) with naturally occurring lysine in the samples. With respect to regression coefficients, the plot in Figure 5b shows that lower wavelengths (leftmost side) had higher absolute coefficients than the rest. There was also a peak at 1533 nm which seems to resemble the spectrum of lysine in Figure 3a and Figure 3b.

Table 1. Summary of reference data and results for PLSR and SAM in predicting % lysine in soybean meal and DDGS.

<table>
<thead>
<tr>
<th>Product</th>
<th>Reference</th>
<th>PLSR</th>
<th>SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Protein, %</td>
<td>Lysine, %</td>
<td>$R^2$</td>
</tr>
<tr>
<td>Soybean meal</td>
<td>45.44</td>
<td>0.98</td>
<td>0.21</td>
</tr>
<tr>
<td>DDGS</td>
<td>1.018</td>
<td>0.99</td>
<td>0.05</td>
</tr>
</tbody>
</table>

*SECV – standard error of cross-validation, SEP – standard error of prediction; PLSR – partial least squares regression; SAM – spectral angle mapper; DDGS – dried distillers grains with solubles
Figure 5. (a) PLS regression scatterplot of % added lysine concentration in soybean meal-lysine mixture and (b) the regression coefficients at factor 2.

Figure 6. (a) PLS regression scatterplot of % lysine concentration in DDGS-lysine mixture and (b) the regression coefficients at factor 3.

**Prediction**

For model prediction, 10 new soybean meal samples were scanned “as is” and without sieving. A bar chart is presented in Figure 7 that compares predicted against
reference % lysine concentration. In this data set, root mean square error of prediction (RMSEP) was 0.72% and SEP was 0.55%. This SEP was lower than the SEP (0.79%) of Bastianelli et al. (2005) in predicting lysine in poultry feeds using conventional NIR spectroscopy and with the addition of pure amino acids in the feed ingredients.

Among the 10 samples, sample 5 had the highest residual which was very high relative to the others that it seemed like an outlier. Samples 6 and 10 also have large variation with reference data. Possible causes of discrepancy could be particle size, reference data, or changes in chemistry during storage which may be caused by unintentional mixing. Difference in particle sizes and surface roughness could contribute to scattering effects and noise in the spectral data. Due to the narrow range of % lysine in the available samples, the availability of the model to predict low values of lysine in soybean meal could not be evaluated. Future work needs to involve soybean samples that cover a wide range of % lysine concentration.

Figure 7. Comparison of predicted and reference % lysine concentration in soybean meal samples using the calibration model from partial least squares regression.
Spectral angle mapper classification

Calibration

Spectral angle mapper classification results from ENVI target detection gave output in terms of number of pixels detected. The relationship of the mean number of lysine pixels detected to the % theoretical lysine was determined using linear regression and correlation. In this study, at SAM maximum angle of 0.10 rad (Figure 8), number of lysine pixels detected was highly related with the % theoretical lysine in soybean meal-lysine mixture having $R^2$ of 0.993. Likewise, in DDGS-lysine mixture, high positive relationship was found with $R^2$ of 0.950. However, at SAM angle of 0.05 rad (Figure 9), both soybean meal-lysine and DDGS-lysine mixtures obtained poor results with correlation of 0.768 and 0.616, respectively. These results agreed with the study of Park et al. (2007) which stated that at increased spectral angle, the mean accuracy of SAM classifiers was improved, such that 7.37% and 75.31% accuracies were achieved at spectral angles of 0.05 and 0.10 rad, respectively. Kim et al. (2014) also produced the same trend of increasing accuracy from 82.5% to 97.6% at threshold value (angle) of 0.06 to 0.09 rad. In our study, increasing the angle further than 0.10 rad would lead to false positives, while in the report of Kim et al. (2014), they had reduced accuracy starting at angle of 0.10 rad.

Prediction

To predict % lysine concentration for 10 new soybean meal samples, SAM target detection tool was done at 0.05 and 0.10 maximum angles. SAM detected zero pixel in all the samples at maximum angle of 0.05. At 0.10 maximum angle, SAM detected 435 to 754 pixels which were used as input data to the regression equations. Utilizing the equation in Figure 8a ($y = 0.0006x - 0.4341$), the predicted % lysine concentration in
soybean meal samples were in the range of -0.173 to 0.018 which contained mostly negative values and were way lower than reference data (2.735 – 3.152%). When the equation in Figure 9a \( y = 0.0047x + 0.7153 \) was used, predicted % lysine values were from 2.760 to 4.260% with SEP of 0.52%. Interestingly, the regression equation developed using SAM maximum angle of 0.05, which produced lower regression statistics, performed better in the prediction of % lysine in new samples.

These prediction results were illustrated in a bar chart in Figure 10. It could be observed that sample 6 had the highest residual, then followed by sample 10, sample 9 and sample 5. All of which also incurred high residuals in PLS regression results as shown in Figure 7. Upon visual observation of samples 9 and 10, the large particle size of the samples relative to the others was noticeable.

![Figure 8](image_url)  
**Figure 8.** Relating number of pixels of lysine detected to the theoretical % lysine concentration using SAM classification at 0.10 maximum angle for (a) soybean meal-lysine and (b) DDGS-lysine mixtures.
Comparison of different methods

Prediction results from SAM and PLS regression were compared with estimation using % crude protein for soybean meal (%Lys = 0.0577CP + 0.273) developed from ISU database 1995-2009 (R²=0.634; n=760). Protein values were obtained by scanning the samples in FOSS DS2500 NIR reflectance instrument. In Figure 11, predictions
using protein regression were closer to the lysine from reference method. NIR hyperspectral imaging, specifically PLS regression and SAM, had higher residuals than those based on protein particularly for samples 5, 6, and 10. Since the samples with high residuals were common for both NIR hyperspectral imaging methods, it might be that these identified samples were the issue or hyperspectral imaging captured extreme values that were not covered by single-point NIR spectroscopy. It is therefore suggested to include more samples in the prediction set to cover wide range of variabilities and to make models robust.

![Bar chart showing residuals (% point difference in lysine between predicted and reference method) from spectral angle mapper (SAM), partial least squares (PLS) regression, and regression based on protein content.]

**Conclusion**

In this study, the possibility of utilizing NIR hyperspectral imaging along with hyperspectral imaging techniques and multivariate analysis to quantify lysine concentration in soybean meal and dried distillers’ grain with solubles was shown. The calibration and prediction results from both PLS regression and SAM classification methods were promising and could be further improved in the future. Considering the usually published model performance on NIR hyperspectral imaging, the results in this
study appeared to be satisfactory. The prediction results revealed that predicted values of some samples had high residuals as compared to the reference data.

Spectral angle mapper classification could be improved in the future by optimizing angle thresholds and applying imaging processing techniques. Examining neighboring pixels and evaluating the level of spectral matching at the pixel level might help improve classification results. PLS regression and preprocessing of spectra still proved to be effective in this study as how it was in other quantification studies.

Samples that had lysine added were found to previously contain lysine in small amounts. This might have also affected the results and it could be prevented in the future by measuring the actual lysine concentration of the samples selected. Animal feed ingredients such as soybean meal usually contain lysine inherently as opposed to others which need to be supplemented. In addition, since samples may vary widely especially for feed ingredients, variation in processing, geographical origin, and particle size, among others, may have influenced the results.

For further study, it is suggested to involve the use of samples with wider range of lysine concentration for model prediction, examine the effect of particle size in samples in the prediction set, explore other hyperspectral image processing algorithms, and include other amino acids in the study.

References


CHAPTER 5. GENERAL CONCLUSION

Guideline Method of Model Calibration for NIR HSI

This section serves as a guide on the methods utilized in this study and other typical methods used in near-infrared hyperspectral imaging (NIR HSI) analysis with emphasis on quantitative applications. Specific areas covered were instrumentation, sample preparation, data collection, data preprocessing, and data analysis. This guide is intended to help future users of NIR HSI instruments and researchers who want to explore other applications of NIR hyperspectral imaging. A more detailed guideline method for NIR spectroscopy model development, instrument management and model maintenance is presented in the Appendix of this chapter and it will be submitted for the approval process of standard methods in Cereals & Grains Association.

Instrumentation

NIR HSI are of different types depending on the application. The commonly used approach in agriculture and food applications is the push-broom type which sweeps the image line-by-line thereby generating full spectrum for each pixel in the image. It is also called line scan. The other types are area scan and point scan.

The wavelength range also differs among instruments. The most common are visible near-infrared (400-800 nm), short wave infrared (1000-2500 nm), and near-infrared (800-1000 nm). The choice of wavelength range will also depend on the application. There are instruments which cover longer wavelength range as compared to the others, but the performance could be affected depending on the sample. In Chapter 3, the performance of different instruments was compared when the full wavelength range was used and when the wavelengths were reduced.
Similar with single-point NIR spectroscopy, NIR HSI also measures in reflectance or transmittance mode. The choice of which mode to use lies on the type and characteristics of the sample and the property to be analyzed (Pojic & Mastilovic, 2013; Lohumi et al., 2015). In reflectance instruments, one side of sample is illuminated and the reflected light from the sample surface is measured (American Association of Cereal Chemists [AACC], 1999; The American Oil Chemists’ Society [AOCS], 2009; Pojic & Mastilovic, 2013). Most studies involving grain and flour products use NIR diffuse reflectance instruments (Sileoni et al., 2015). On the other hand, in transmittance instrument, light passes through the sample and composition is measured through the fixed thickness of the sample material (AACC, 1999; AOCS, 2009). This type is suitable for measuring the internal and external qualities of liquid and some solid (grains, meat or dairy) or gaseous samples (Lohumi et al., 2015; Qu et al., 2015).

In setting-up the instrument, the protocol of the instrument manufacturer needs to be followed in the installation prior to data collection, such as warm-up time, operating conditions, and performance tests, and during the operation (American Society for Testing and Materials [ASTM], 2016; AOCS, 2009). In NIR HSI, camera height with respect to the sample, the camera focus and other settings could be adjusted depending on the instrument and manufacturer’s specifications. In this study, there were some adjustments made to cater the instrument to our application since the instrument was originally designed for remote sensing applications. These were the camera settings, camera height, and sampling cup.

Sample preparation

In reflectance analysis, it is usually required for samples to be homogenous. For samples in powder form such as flours and those used in this study, they can be used
as-is. To ensure that calibration samples are of uniform particle size, a grinder with a specific particle size or sieves can be utilized for this purpose. Calibration samples refer to the set of reference samples, which have known composition or properties as measured by a reference or laboratory analytical method, used in developing a calibration model.

Calibration samples should be representative of the entire population, such that all chemical and physical variability that are typically encountered will be covered. The number of samples to be included in the calibration set depends on the application and the type of algorithm to be used. In Chapter 3 of our study, the ISU soybean meal calibration quick set is composed of samples from different years and different origins. The set is being used across all NIR instruments in the Grain Quality Laboratory. In Chapter 4, the range of concentrations used for lysine was based on the typical range of lysine concentration found in soybean meal and DDGS. Lastly, it is important for all laboratory samples to be kept under stable conditions in order to avoid changes in the sample composition prior to calibration process.

Data collection

Prior to scanning samples, dark background was captured by covering the cap of light source and then image calibration was performed using a spectralon (50% reflectance) as white reference material. These would be used for image calibration. In the instrument we used, the dark reference and white reference images were automatically included when image calibration process was performed. In most published studies, these two images are used as input to the formula to obtain the pixel-based relative reflectance, \( I = (I_0 - D)/(W - D) \), where \( I_0 \), \( D \) and \( W \) stand for the raw reflectance, dark reference and white reference images, respectively (ElMasry & Sun,
2010). This image calibration setting should ideally be used in all the samples to be scanned from the set to ensure uniformity.

In this study, all samples were scanned on as-is basis. Reference measurements (for protein and lysine content) were obtained from outside laboratories. Reference method is evaluated in terms of precision, commonly referred to as the standard error of the laboratory (SEL) (AOCS, 2009). The error of prediction model in NIR hyperspectral imaging and NIR spectroscopy, in general, will depend on the SEL, particularly in supervised methods that require reference data (AOCS, 2009).

**Data preprocessing**

Some of the common image preprocessing steps to select the relevant information are background removal, selection of region of interest (ROI), removal of bad pixels, and resizing. In this study, selection of ROI was done to limit the area to be analyzed and resizing since the image was stretched after saving the files. Both are tools in ENVI which are easy to use. In selection of ROI, an ellipse shape was used, and the area was uniformly set for the entire sample set. Other options for ROI selection are by pixel size and by setting minimum/maximum threshold values. Resizing was done using nearest neighbor algorithm and by setting the desired dimensions. In MATLAB, these can be done using hyperspectral imaging functions and tools which can be used readily or can be put into codes.

Spectral preprocessing, also referred to as spectral pretreatment, can be applied on the extracted spectral data from hyperspectral images for purposes of enhancing, smoothing, correcting, and/or scaling (AOCS, 2009). Among the usual methods being used are standard normal variate (SNV) and multiplicative scatter correction (MSC) for normalization; offset correction, weighted least squares baseline correction, detrend,
smoothing, and derivative for noise and baseline effect; and, autoscaling and mean centering for variable scaling (AOCS, 2009). Depending on the application and the characteristics of the spectra, one method can be used or a combination of two to three from each type of function, as performed in published studies (A. Gowen et al., 2010).

Data analysis

Qualitative data analysis can be performed using multivariate statistics such as cluster analysis, discriminant analysis and principal component analysis (PCA) (ElMasry & Sun, 2010). For example, in Chapter 4, PCA was used to analyze spectral data of pure lysine, soybean meal and DDGS. The PCA score plots were able to identify pure lysine from the other samples.

In model development, various regression and classification methods are being used. One of the most widely used methods for quantitative prediction is partial least squares (PLS) analysis which was used in the two studies here. The performance of PLS regression models are evaluated using model performance statistics, such as coefficient of determination $R^2$, standard error (SE), root mean square error (RMSE) and relative predictive determinant (RPD) (ratio of standard deviation of reference values in the set to the standard error) (AOCS, 2010). Models performed better when they have higher $R^2$, higher RPD, lower SE and lower RMSE. Model validation can be done using 20% of the samples or by leave-one-out cross-validation as in this study. Ideally, model prediction needs to be performed using a new set of samples.

In Chapter 4, spectral angle mapper (SAM) classification was used in the analysis which is a hyperspectral image processing algorithm. It uses the reference spectra (of pure lysine as in Chapter 4) and measures the angle between the reference spectra and the target spectra. This method is useful in target detection mostly for
remote sensing applications. SAM classification gives the output in terms of pixels detected. The number of pixels would be used for regression with the reference data. The same procedure of SAM classification will be done in prediction.

In a number of studies, the authors employ methods for dimensionality reduction and wavelength selection using spectral data or hyperspectral images (ElMasry & Sun, 2010). Image post-processing algorithms can also be used depending on the application. Visualization of predicted data is one of the benefits of hyperspectral imaging (ElMasry and Sun, 2010). In Chapter 3, the calibration model was applied on the spectra of the sample per pixel basis which resulted to a distribution map of protein content across the sample’s image.

**General Findings**

The ability of NIR hyperspectral imaging to predict protein content in soybean meal and lysine content in soybean meal and dried distillers’ grains with solubles (DDGS) had been demonstrated and evaluated in this study. Multivariate analysis, particularly PCA and PLS regression, made it possible for NIR hyperspectral imaging to achieve the objectives.

Spectral preprocessing by SNV combined with Savitzky-Golay derivative was found to reduce scattering effects and provide the best validation and prediction statistics: highest $R^2$, lowest standard error, and highest RPD. Principal component analysis revealed information that could not be visually observed in the spectral line plot, such as potential outliers and wavelengths that had more noise than the others. Removal of outliers and reduction of wavelength range also improved the performance of the models. The results for protein content prediction were satisfactory and
comparable to those from single-point NIR spectroscopy instruments, unlike with oil content which did not produce good prediction results.

In lysine content prediction in soybean meal and DDGS samples, both PLS regression and SAM classification showed promising output in model calibration. The choice of which SAM maximum angle (0.05 or 0.10) to use varied depending on application. Angles lower or higher than this range would lead to more false positives or false negatives and thereby reduced calibration or prediction performance. Among the possible sources of error for low prediction performance were particle size, reference data, and the limited number of samples in the prediction set.

**Recommendations for Future Work**

To date, there has been no published paper on the use of NIR hyperspectral imaging on prediction of amino acid concentration in animal feed ingredients. This study serves as a precursor to more applications of NIR HSI in the grain and feed industry where quality is a critical issue, rapid and accurate testing is a need, and the products are highly variable.

To improve the model calibrations and apply the technology in the commercial scale, it is recommended to include more samples in the calibration and prediction set, apply data dimensionality reduction and image processing techniques, and, predict other amino acids such as methionine and cystine. Suggested future research topics on NIR HSI are prediction of aflatoxin contamination in inoculated ground animal feed ingredients, and identification of specific wavelengths to be included in a portable NIR HSI tool that can be used for remote applications in the field.
References


This draft guideline is a combination of different standard methods from AACC (American Association of Cereal Chemists), AOCS (The American Oil Chemists’ Society), ASTM (American Society for Testing and Materials) International, and ISO (International Organization for Standardization). This is being prepared for submission to Cereals & Grains Association (formerly AACC International).

Objective

This guideline provides general information on near-infrared (NIR) model development, instrument management, and evaluation of instrument calibration. NIR spectroscopy, applied to the quantitative analysis of constituents in cereals, is a popular secondary procedure for reasons of simplicity of operation, throughput, objectivity, and accuracy. The absorption response by overtone and combination frequencies of O-H, C-H, and N-H molecular vibrations, which are abundant in biological matter, allows samples to be analyzed. These signals are discernible by photometric detectors, which inherently have a very high signal-to-noise response. With the aid of a computer for post-processing and statistical tools such as multivariate analysis, the composition of cereals can be determined from NIR spectra, in terms of the contents of protein, carbohydrate, moisture, lipid, and other constituents. Linear and nonlinear regression modeling is used to relate NIR readings to chemical concentrations determined by conventional reference methods. As such, the robustness and accuracy of an NIR model are strongly dependent on the calibration samples, the quality of the reference methods, the statistical methods used to develop the prediction model, and the calibration procedure itself.

Scope

Model development in this guideline includes sample preparation, data preprocessing, model calibration, validation and performance evaluation, and calibration transfer to instruments of like brand. Instrument management involves instrument set-up, maintenance, and performance evaluation. Evaluation of instrument calibration covers tests for slope of regression line, bias, precision, and accuracy, and assumes that reference and instrument values have been collected from a common set of samples by an appropriate method. This guideline can also be used for agricultural applications other than grains and oilseeds.

Apparatus

1. NIR instruments that are available commercially are in the NIR wavelength region of 750 – 2500 nm. Measurements are performed with reflectance or transmittance instruments (Appendix) that are available from several companies. See Supplier Index for Method 39-10.

2. The average is often obtained for multiple measurements of the same sample to improve the signal to noise ratio (S/N), which is proportional to the square root of the number of subsamples.

3. Reflectance analysis usually require samples to be homogenous and therefore need milling device or grinder. Different types of grinders may give different values on the same instrument due to variations in particle size, shape, and distribution. The same type of grinder and grinding procedure used in calibration development should be used for testing ongoing samples. Grinders capable of grinding samples to less than 1-mm particle size are recommended for cereals (2-mm for oilseeds). It should be noted that NIR measurements can be influenced by variations in grinding conditions. If samples are already in fine powder form, then these can be used as-is.

4. Calculator or, preferably, a personal computer loaded with an application program for spreadsheet analysis capable of performing simple linear regression, such as Microsoft Excel, can be used for statistical assessment of instrument’s calibration. A computer software statistical program that can perform multivariate analysis is recommended for model calibration.

Calibration Process

Instrument performance

1. Follow the manufacturer’s procedure to perform checks on the internal functioning of the spectrometer before starting tests, such as warm-up time period and daily performance quality
tests. This is recommended to guarantee that correct and safe operation as well as optimum accuracy are achieved before data collection.

2. The instrument must be stored and operated within the environmental conditions (temperature and relative humidity) specified by the manufacturer.

3. Instrument accuracy, which measures how close a measurement is from the accepted value, should be tested on a minimum of 20 samples and 3 replicates for each sample. It is measured by the standard error of instrumental performance (SEIP) and comparisons are made with the reference measurement for each sample with the first replicate for each sample.

\[
SEIP = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n - 1}}
\]

where
- \( y_i = x_i - r_i \) represents the difference between the prediction for the first replicate and the reference value for sample \( i \)
- \( \bar{y} \) represents the mean of \( y_i \)
- \( n \) represents the number of samples

4. Precision, or repeatability, which defines how close each measurements of the same sample are from each other, should be tested on a minimum of 20 representative samples and 3 replicates for each sample. It is determined by calculating the standard deviation (SD) of the replicates across the entire sample population.

\[
SD = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{3}(P_{ij} - \bar{P}_i)^2}{2n}}
\]

where
- \( P_{ij} \) represents the predicted value for sample \( i \) and replicate \( j \)
- \( \bar{P}_i \) represents the average of the three predicted values for sample \( i \)
- \( n \) represents the number of samples

5. Daily check could be implemented on samples that are usually analyzed by the lab (usually in duplicate), in addition to instrument self-checks and in substitution to accuracy and precision tests. This allows a daily evaluation of instrument performance, evaluation of instrumental variations due to environmental changes and hardware evolution, and detection of any malfunction.
   a. A minimum of one control sample should be tested at least once daily. The control sample should be stable for longer storage duration and should resemble the samples to be analyzed.
   b. Control charts with the plotted daily recorded data should be examined for significant patterns or trends.

6. Checking the instrument noise and the precision and accuracy of wavelength or wavenumber should be done in scanning spectrophotometers at least once in a weekly basis or as frequent as advised by the manufacturer. Comparison of the data from these checks should be compared to the instrument specifications and requirements. For instruments in a network, the recommendations of the manufacturer will be particularly considered in the standardization of the instruments.

Sample preparation
1. Calibration samples should be representative of the population of samples that will be predicted by the model.
   a. All chemical and physical variability that are typically encountered for routine analysis for the desired application should be well-represented.
   b. Ideally, a calibration set will include samples
      i. Which contain examples of all chemical components to be predicted by the model,
ii. For which the concentrations of the chemical components: have range of variation that exceeds those of the predicted samples, and are following a uniform distribution over the entire range of variation,

iii. Which are sufficient in number to statistically define the relationships between the reference data and spectral data

2. If prior knowledge of reference data is available, the range of values can be included in a factorial design across constituents. Manufacturers may have a spectral selection procedure that identifies samples most suitable for use in calibration from spectral data. The constituent ranges should still be covered as uniformly as possible. Each manufacturing company will recommend a minimum number of samples, usually at least 100 of varying origin and background is the standard for a grain calibration. Expect to collect spectra on more samples than are actually used for calibration purposes.

3. Approximation of the number of samples depends on the application, the complexity of the samples to be analyzed and the type of algorithm, and whether for concept test, in controlled conditions, or uncontrolled conditions. Commonly used algorithm or calibration techniques are multiple linear regression (MLR), principal component regression (PCR), partial least squares (PLS) regression, artificial neural networks (ANN), support vector regression (SVR), and locally weighted regression (LWR).

<table>
<thead>
<tr>
<th>Algorithm*</th>
<th>Concept test</th>
<th>Use in controlled conditions</th>
<th>Use in uncontrolled conditions (i.e., temperature changes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR, PCR, PLS</td>
<td>100</td>
<td>300</td>
<td>500</td>
</tr>
<tr>
<td>ANN</td>
<td>500</td>
<td>1000</td>
<td>+1000</td>
</tr>
<tr>
<td>SVR</td>
<td>50</td>
<td>100</td>
<td>n/a</td>
</tr>
<tr>
<td>LWR</td>
<td>300</td>
<td>500</td>
<td>1000</td>
</tr>
</tbody>
</table>

4. Before calibration, clean samples in the same way as would be done in practice. If uncleaned routine samples must be run, do not clean calibration samples, other than removing obviously large pieces of foreign material. If grinding is required, grinder should be vacuum cleaned between samples unless it is self-cleaning. The process of sample preparation must be standardized all throughout for calibration, validation and routine samples.

5. Keep all laboratory samples under stable conditions that will prevent changing of the sample composition from the time of sampling up to the time of beginning the calibration procedure.

**Data collection**

1. Collect spectral data. Moisture must be determined by an approved method at the time of scanning. Other constituent references may be analyzed later, even after drying. If the moisture range is very high, e.g., 10% or more, more accurate results are obtained by NIR if calibrations are based on “as-is” moisture data. Subsequent moisture analysis is needed to correct the reference data back to the original moisture level, if the material has to be dried during sample preparation. Test all samples in at least duplicate by reference laboratory methods.

2. Sample temperature and instrument temperature may cause wavelength shifts in NIR spectra. When temperature conditions for samples are different from those of the calibration samples, follow manufacturer’s recommended methods for adjustments. These methods include adjusting spectral data to emulate constant temperature, including warm and cold samples empirically in the calibration files, including samples over the range of the calibration whose temperature has been adjusted to 3 to 5 temperatures over the future range of use, and using temperature as an independent variable in multivariate equations. Frozen samples may produce such large spectral variations that they must be eliminated from the calibration process.
3. Reference measurements should be performed, on the samples kept in the calibration set, as soon as possible after collection of spectral data. Some elements such as moisture and other volatile constituents can vary over time. For other elements not sensible to environmental conditions, analysis can be done later if needed.

a) Reference methods to measure moisture, protein, fat, and other constituents and parameters should be internationally accepted.

b) Precision of reference method should also be known, and it can be measured by using the standard error of the laboratory (SEL):

\[
SEL = \sqrt{\frac{\sum_{i=1}^{n} \left( \sum_{j=1}^{r} (y_{ij} - \bar{y}_i)^2 \right)}{n}}
\]

where

- \(y_{ij}\) represents the concentration measured for sample and replicate
- \(\bar{y}_i\) represents the average values of the different replicates for sample
- \(n\) represents the number of samples
- \(r\) represents the number of replicates

A representative block of samples (\(n > 20\)) should be assayed repeatedly to establish and update SEL.

c) NIR spectroscopy is an indirect method and the error of the prediction model will be a function of the SEL. A calibration model can be, at most, as good as the reference method.

**Data preprocessing**

Various preprocessing or pretreatment methods could be applied on collected spectral data for enhancing, smoothing, correcting, and/or scaling. These are the most common preprocessing methods:

1. Noise and baseline effect
   a. Offset correction
   b. Weighted least squares baseline correction
   c. Detrend
   d. Smoothing
   e. Derivative

2. Normalization
   a. Standard normal variate
   b. Multiplicative scatter correction

3. Variable scaling
   a. Autoscaling
   b. Mean centering

Consult manufacturer for the list with the various preprocessing methods that are compatible with their instrument software. Details on the definition and when to use which type of preprocessing/pretreatment technique are discussed in Appendix.
**Model calibration**

1. For cases where the relationship is linear and the collinearity among wavelength is low, MLR can be used.
2. For highly correlated spectral data, PCR and PLS regression methods have been developed.
3. When dealing with non-linear relationships, learning-based methods such as ANN, LWR and support vector SVR methods can be employed.
4. Most instruments can easily handle calibration models developed using MLR, PCR, and PLS. ANN and SVR require an additional prediction engine because they are not included in most chemometrics suite.

**Validation**

1. Calibrations must be validated on samples not included in the calibration. If the manufacturer's procedure suggests separating validation samples from the same pool that generated the calibration set, recognize that this is not an independent validation. Independent validation can be achieved by validating the calibration on a new data set (new batch samples or next year samples) which should be representative of the samples to be analyzed in the future. The true model performance will be estimated by the validation parameters if a completely independent validation set will be used. To perform independent validation: select 30–50 samples, analyze them by NIR and reference methods, determine residuals, and evaluate model performance.

   Validity of calibration models is limited to the variations used in the validation, that is, they cannot be used in the range they have not been validated. It is recommended to cover variations such as a) combinations and composition ranges of major and minor sample components; b) seasonal, geographic and genetic effects; c) processing techniques and conditions; d) storage conditions; e) sample and instrument temperature; and, f) instrument variations.

2. Cross-validation is performed on the original sample set, using a single-sample elimination procedure or also known as leave-one-out cross-validation. Compute the calibration without one sample and use the calibration to compute the NIR results. Reinstall the sample and remove another sample. Repeat the process until all samples have been removed and reinstated and the NIR results are computed for each sample, using a calibration in which it was not included. Report the overall coefficient of correlation and the standard error of performance as the validation statistics.

3. If temperature is a major concern, analyze validation samples at high and low temperatures (or samples at various temperatures). Take care to avoid condensation.

4. Model performance can be evaluated by:
   a. Standard error for precision prediction
      \[
      SE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_{ip})^2 - \left(\sum_{i=1}^{n}(y_i - \bar{y})\right)^2}{n - 1}}
      \]
      where
      - \(\hat{y}_{ip}\) represents the predicted value of the sample i in prediction scenario
      - \(y_i\) represents the reference value for the sample i
      - \(n\) represents the number of samples

   Note that this is a bias adjusted formula. The root mean square error, including the bias is presented later.

   i. During calibration
      The standard error of calibration (SEC) will provide a first evaluation of the predictive ability of the model. It is calculated by predicting all calibration samples with the model. It generally provides over-optimistic statistics.

   ii. During cross-validation
The standard error of cross validation (SECV) is calculated by withholding predetermined groups of samples from the calibration set, developing new models from the remaining samples, and finally predicting left apart samples. The operations is repeated by including withheld samples in the calibration set and predicting a second set of samples isolated for validation. Prediction statistics are combined to output cross-validation error. SECV is usually a better indicator of method precision than SEC. Different cross-validation methods can be used based on the nature of the data.

iii. During operation

The standard error of prediction (SEP) is calculated by predicting “unknown” samples. SEP is the best estimate of model precision of a model given the variability encountered in validation is present in calibration.

b. Relative predictive determinant (RPD) is another common statistic for precision. It is the ratio of the standard deviation of the reference values of the set to predict to the standard error (SEC/SECV/SEP), when applied to the same set. RPD values can be interpreted using the following scale:

<table>
<thead>
<tr>
<th>RPD values</th>
<th>Classification</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0–2.3</td>
<td>Very poor</td>
<td>Not recommended</td>
</tr>
<tr>
<td>2.4–3.0</td>
<td>Poor</td>
<td>Very rough screening</td>
</tr>
<tr>
<td>3.1–4.9</td>
<td>Fair</td>
<td>Screening</td>
</tr>
<tr>
<td>5.0–6.4</td>
<td>Good</td>
<td>Quality control</td>
</tr>
<tr>
<td>6.5–8.0</td>
<td>Very good</td>
<td>Process control</td>
</tr>
<tr>
<td>8.1+</td>
<td>Excellent</td>
<td>Any application</td>
</tr>
</tbody>
</table>

c. Accuracy of prediction model in validation situations is often characterized by bias along with SEP.

\[ Bias = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)}{n} \]

where

\( \hat{y}_i \) represents the predicted value of the sample \( i \)

\( y_i \) represents the reference value for the sample \( i \)

\( n \) represents the number of samples

Provided the primary (destructive) method is used to calibrate the secondary (non-destructive) method, the bias generally should be very close to zero. If not, it is time to recalibrate.

d. Root mean square error (RMSE) is another statistic that combines both precision and accuracy. It is also applied for calibration, cross-validation and/or prediction.

\[ RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - 1}} \]
where

\[ \hat{y}_{ip} \] represents the predicted value of the sample \( i \) prediction scenario
\[ y_i \] represents the reference value for the sample \( i \)
\[ n \] represents the number of samples

5. Impact of other sources of error. As mentioned earlier, the final error of a calibration model is a function of many parameters, the most common being the error of the laboratory, the sampling error, and the instrumental error.

**Calibration transfer (standardization)**

1. Calibration transfer is the process of aligning slave units with the calibration master to obtain equivalent results. Developing a model on the slave unit (or secondary unit) will cost higher than performing a calibration transfer. It is important to correct variation in instrumental properties (absorption, S/N, or wavelength alignment) that may occur.

2. Most of the standardization methods require standardization samples run on both master and secondary units. Standardization samples must be representative of the constituent range and well predicted by the calibration. It is possible that a different standardization set exist for different parameters on the same product. A minimum of 20 samples should be used.

3. In general, there are two methods: optical alignment of spectral data before calculation of predicted values (e.g., direct standardization, piecewise direct standardization, single wavelength standardization, etc.) and slope-intercept of calculated values.
   a. Optical alignment methods are easier to maintain, more stable, and cheaper if several units are to use the same calibration. They are based on the modification of the spectra collected on the secondary unit to match the spectra on the master unit. They calculate for each wavelength a slope and a bias that are applied to the spectra of the secondary units.
      For optical alignment, select a set of at least 10 varied samples to use as check samples. Check and standardization samples should be known to give repeatable readings in the master unit, with agreement to reference within ± 1 standard error of performance on each factor measured, and to be representative of the data set.
   b. Slope-intercept method, also referred to as post-regression correction, uses a slope and a bias which are calculated to match predictions of the standardization set on both master and secondary units. To apply this method, select a set of 20 or more varied samples to use as standardization samples.

4. Test check samples in the master instrument and slave(s) before standardization. Test standardization samples and compute standardization parameters by the manufacturer's procedures.

5. Install standardization. Retest check samples.

6. Verify that the average differences between master and slave(s) are not statistically significant \((P = 0.05)\). Compute standard deviation between master and slave instruments.

7. Development of robust calibration models. It is possible to include in the calibration model the spectra collected on the secondary unit. The variability is modeled by the regression algorithm and no further standardization is required. This can be complex for large numbers of secondary units.

8. A post-standardization set of a minimum of 20 samples (independent from the calibration, validation, and standardization sets) should be used to validate the standardization method.

**Routine tests**

1. Before tests each day, follow manufacturer’s recommended daily check (warm-up time and self-check) procedures. Run daily checks and maintain a control chart over time. Manage the evolution of the instrument and sample over time.

2. To confirm that the instrument is functioning well, perform frequent instrument check performances of accuracy and repeatability.

3. Prepare samples in the same way as calibration samples were prepared, mix thoroughly, place in test cell, and take readings.
Outlier detection
1. Outliers are samples that are statistically different from the population of interest. This difference can come from hardware failures, sample degradation, or poor reference values.
2. The detection of such samples is typically done with a Hotelling’s $T^2$ test—a multivariate method that tests the membership of an observation to a group—at a given confidence interval, or by detecting samples in the calibration process that present an abnormally large residual value.
3. The leverage statistic is also used to identify outliers during calibration, detect samples that are extrapolation of the model, and give an estimate of an estimated value's uncertainty. In other software packages available commercially, this is referred to as the Mahalanobis Distance or the hat matrix.
4. Obvious outliers must be removed. However, “abnormal” samples can be representative of a new variability that should be included in the calibration set. There is no substitute for an understanding of the physical or biological system.

Moisture basis
1. Percent constituents can be expressed either on an “as-is” (moisture content as received) or constant-moisture basis. Most units are capable of making the conversion from “as-is” percentages used for calibration to a constant-moisture basis, using the NIRS-determined moisture.
2. Calibrations can be done directly to constant-moisture basis. The reference data are converted to the moisture basis before entering in the data file(s). In this case, a moisture calibration is implicitly included in the constituent calibration. Direct moisture-basis calibrations generally produce unacceptable variability if the moisture range of included samples is greater than 10 percentage points. Anticipated moisture range must be included in direct moisture calibration sets.

Calibration of correlated elements
1. Often, the parameter to measure will be highly correlated with other constituents that are commonly predicted by NIR.
2. In some cases, using a simple model, based on common predicted parameters to predict the correlated elements, will provide accurate results and be more economically viable than developing an NIR calibration. However, using indirect measurements such as NIR predictions to predict another element with a simple model can quickly become unstable because of the multiplication of error sources.
3. It is possible that some samples will not follow a general pattern of correlation with one or more easily predicted constituents and will be miss-predicted. A strong validation strategy must be employed to ensure the validity of these third level models.

References
Definition of terms

_NIR reflectance instrument_ – capable of measuring sample composition by illuminating one side of the sample and detecting the reflected light. In this set-up, the light generally penetrates in the first few millimeters of samples. This makes reflectance analysis to normally require homogeneous samples of ground material to minimize scattering differences, but some manufacturers apply reflectance to whole grains as well.

_NIR transmittance instrument_ – capable of measuring sample composition through a fixed thickness of material (commodity dependent). In NIR transmittance instruments (as with reflectance), subsample readings are averaged to reduce sampling errors, increase the signal to noise ratio (S/N) and minimize spectral variation caused by the random positioning of particles or kernels.

_Calibration_ – a process used in creating a model that will relate spectral data to reference data (known component concentrations or properties) for calibration samples. The calibration model will be applied to the spectra of future samples of unknown composition.

_Model validation_ – the process of evaluating the performance of a calibration model using validation samples to compute for bias between the predicted values of the model and the reference method.

Data pre-processing

<table>
<thead>
<tr>
<th>Preprocessing method</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise and baseline effect</td>
<td></td>
</tr>
<tr>
<td>Offset correction</td>
<td>Remove constant baseline</td>
</tr>
<tr>
<td>Weighted least squares baseline correction</td>
<td>Remove varying baseline</td>
</tr>
<tr>
<td>Detrend</td>
<td>Remove constant, linear, or curves baseline</td>
</tr>
<tr>
<td>Smoothing</td>
<td>Low pass filter removing high frequency noise</td>
</tr>
<tr>
<td>Derivative</td>
<td>Remove baseline and enhance peaks</td>
</tr>
<tr>
<td>Normalization</td>
<td></td>
</tr>
<tr>
<td>Standard Normal variate (SNV)</td>
<td>Normalize sample to unit standard deviation across variables</td>
</tr>
<tr>
<td>Multiplicative Scatter Correction (MSC)</td>
<td>Remove scattering effects</td>
</tr>
<tr>
<td>Variable scaling</td>
<td></td>
</tr>
<tr>
<td>Autoscaling</td>
<td>Standardize samples to mean zero and unit standard deviation across samples</td>
</tr>
<tr>
<td>Mean centering</td>
<td>Subtract each sample by the mean sample of the set</td>
</tr>
</tbody>
</table>