Economic and chemometric studies to supplement food-grade soybean variety development in the Mid-Atlantic region

Nick Lord

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
in
Crop & Soil Environmental Science

Bo Zhang, Chair
Clinton Neill
Yun Yin
Song Li

December 16, 2020
Blacksburg, VA

Keywords: edamame, vegetable soybean, production feasibility, willingness-to-pay, near-infrared reflectance spectroscopy (NIRS), soybean sugar profiling

Economic and chemometric studies to supplement food-grade soybean variety development in the Mid-Atlantic region

Nick Lord

ACADEMIC ABSTRACT

Sustainability of the soybean industry relies on the growth of new industries and the continued improvement of seeds for utilization. Grower adoption and growth of the edamame industry has been slow in part due to insufficient information on its potential profitability and marketability. As such, the first and second objectives of this thesis aimed at 1) determining production costs of hand-harvested fresh edamame enterprise and 2) exploring consumer willingness-to-pay (WTP) for fresh, local, organic, and "on-the-stalk" marketed edamame. Sucrose, raffinose, and stachyose sugars hold tremendous implications for utilization of soybean seeds in livestock, soyfood, and probiotics industries. Current sugar phenotyping methods using high-performance liquid chromatography (HPLC) are costly and inefficient. Therefore, the third objective of this study was to develop calibrations to predict sugar content using near-infrared reflectance spectroscopy (NIRS). Results showed that labor accounted for 72% of production costs for edamame pods, which largely limits its profit potential. Mean WTP for fresh and local edamame exceeded their frozen and non-local counterparts by 94 and 88 cents, respectively. In addition, mean WTP for organic edamame exceeded non-GMO edamame by 33 cents. Pro-environmental attitudes appeared to be a consistent driver of WTP these three attributes. Meanwhile, a 40-cent discount for "on-the-stalk" edamame compared to pods indicates convenience may also be a factor in edamame marketability. Calibration development for sucrose and stachyose was successful, with R^2_{cal}, R^2_{cv}, RMSEC, and RMSECV of 0.901, 0.869, 0.516, and 0.596, and 0.911, 0.891, 0.361, and 0.405, respectively. Alternative methods should be investigated for quantification of raffinose.

Economic and chemometric studies to supplement food-grade soybean variety development in the Mid-Atlantic region

Nick Lord

GENERAL ABSTRACT

Sustainability of the soybean industry relies on the growth of new industries and the continued improvement of seeds for utilization. Grower adoption and growth of the edamame industry has been slow in part due to insufficient information on its potential profitability and marketability. As such, the first and second objectives of this thesis aimed at 1) determining production costs of hand-harvested fresh edamame and 2) exploring relative marketing potential of fresh, local, organic, and “on-the-stalk” edamame. Sucrose, raffinose, and stachyose sugars hold tremendous implications for utilization of soybean seeds in livestock, soyfood, and probiotics industries. Current methods of quantifying sugar are costly and time inefficient. Therefore, the third objective of this study was to develop prediction models to estimate sugar content using near-infrared reflectance spectroscopy (NIRS). Results showed that labor accounted for 72% of production costs for edamame pods, which largely limits its profit potential. Fresh and local edamame showed considerable marketing potential over frozen and non-local edamame. In addition, organic edamame showed marginal marketing benefit over non-GMO. Pro-environmental attitudes were an important driver of these results. Meanwhile, on-the-stalk edamame shows poor marketing potential, likely due to preference for convenience of food preparation in the U.S. NIRS prediction models for sucrose and stachyose showed strong predictive accuracy and low error, suggesting potential for implementation. The prediction model for raffinose, however, remained poor.

Dedication

I would like to dedicate this thesis to my parents, Sharmini & Eugene Lord. While I can give myself some credit for conducting all the research in this thesis and for somehow managing to assemble this big, fancy document, my parents are the ones who deserve the real praise. For the past 25 years, they have tirelessly pushed me to study hard, stay focused, and to always strive for excellence. They have always been there to provide me with sound financial, career, and personal advice when I've needed it. They programmed me to value my education and – perhaps what I am most grateful for – they never let me take it lightly.

Ammi and Thathi - if it were not for the solid foundation that you spent years helping me build, I am not sure I would have had the wherewithal to complete this Master's degree.

Acknowledgements

Firstly, I would like to thank my primary graduate advisor, Dr. Bo Zhang, for all of her support and guidance. I am very fortunate to have been able to have Dr. Bo as a mentor during my time in graduate school. Over the past three years, I have learned so much about soybeans, research, and leadership from being in her program. What I am most appreciative to her for is that she helped me see the potential in myself. If it weren't for her constantly pushing me out of my comfort zone and challenging me to increase my productivity level, I may have never realized what I was truly capable of. I am extremely indebted to her for helping me experience such profound personal and professional growth during my time as her student, and hope I can make her proud as I continue to progress through my career.

I would like to thank all my committee members, Dr. Yun Yin, Dr. Song Li, and Dr. Clinton Neill, who have served as constant sources of encouragement throughout my time in graduate school and on the USDA Edamame project. In regards to my projects, I would like to especially thank Dr. Neill for spending extra time to help bring me up to speed on the numerous economic concepts and methods involved in my thesis research given my limited background in economics. Chemometrics can also be a very dense topic to learn. Fortunately, I had a couple of extremely helpful individuals – David Honigs and Chao Shang – who were literally and figuratively "instrumental" in guiding me through how to operate and troubleshoot the NIR and HPLC analytical instruments as well as the spectral and wet methodologies involved for my Chapter 4 project. I am very grateful to both of them for going out of their way to help me when I needed it – and believe me, I very often did.

Last but certainly not least, I would like to thank the rest of my colleagues both formerly and currently in the soybean breeding program at Virginia Tech including: Tom Pridgen, Andy Jensen, Ben Averitt, Diana Escamilla, Matthew Colson, Luciana “Lab Mama” Rosso, Muliang Peng, Will Singer, Zach Shea, Qian Zhu, Xingbo Wu, Xiaoying Li, Jessica Wilbur, Zhibo Wu, and the numerous undergrads who have helped me harvest edamame plants, pick pods, grind seeds, or scan samples in the NIR at one time or another. Aside from helping me complete my projects, I am enormously thankful to you all for your friendship, camaraderie, teamwork, and most of all, for making me feel at home in this program.

Attributions

Many subject matter experts assisted with the research projects described in this thesis. Their names, departments, and description of their roles in each project can be found below.

Chapter 2: Assessing labor and production costs of fresh market edamame production in Southwest Virginia

Nick Lord, Master's Degree Candidate in the School of Plant & Environmental Sciences, Virginia Tech. Mr. Lord conducted the field trials and collected all experimental data, made all calculations, and prepared the manuscript presented in this chapter.

Bo Zhang, Assistant Professor in the School of Plant & Environmental Sciences, Virginia Tech. Dr. Zhang assisted with designing the experiment and experimental objectives. She also helped edit the manuscript.

Clinton Neill, Assistant Professor in the Charles H. Dyson School of Applied Economics and Management at Cornell University. Dr. Neill provided guidance on data collection, calculation of labor costs, and interpretation of results. Dr. Neill also provided considerable guidance in preparing the manuscript, as well as editing it.

Chapter 3: Investigating consumer demand and willingness-to-pay (WTP) for edamame labeled for various external attributes

Nick Lord, Master's Degree Candidate in the School of Plant & Environmental Sciences, Virginia Tech. Mr. Lord was responsible for devising survey questions and content, administering surveys, performing data analysis, and preparing the manuscript for this chapter.

Bo Zhang, Assistant Professor in the School of Plant & Environmental Sciences, Virginia Tech. Dr. Zhang assisted with experimental objectives and helped edit the manuscript.

Clinton Neill, Assistant Professor in the Charles H. Dyson School of Applied Economics and Management at Cornell University. Dr. Neill provided guidance on survey design and format, data collection procedures, data analysis, and helped edit the manuscript.

Chapter 4: Development of Near-Infrared Reflectance Spectroscopy (NIRS) calibration for sucrose, stachyose, and raffinose in ground soybean seed using a Perten DA7250 analyzer

Nick Lord, Master's Degree Candidate in the School of Plant & Environmental Sciences, Virginia Tech. Mr. Lord was responsible for assembling calibration set materials, collecting all spectral data and reference chemistry data, performing data analysis, and preparing the manuscript.

Bo Zhang, Assistant Professor in the School of Plant & Environmental Sciences, Virginia Tech. Dr. Zhang assisted with experimental objectives and helped edit the manuscript.

Chao Shang, Senior Research Associate in the School of Plant & Environmental Sciences, Virginia Tech. Dr. Shang provided guidance on collection of reference chemistry data and how to safely operate and maintain the high-performance liquid chromatography equipment. He also helped edit the manuscript.

Luciana Rosso, Faculty Research Associate in the School of Plant & Environmental Sciences, Virginia Tech. Dr. Rosso provided guidance on laboratory protocol for extraction of sugars from ground soybean seed and also assisted with operating the high-performance liquid chromatography instrument. She also helped to edit the manuscript.

Table of Contents

List of Tables

List of Figures

Introduction

Cultivated soybean (Glycine Max (L.) Merril) is an annual legume species in the Fabaceae family related to nuts, peas, and lentils. It originated in East Asia, where it has been used as a source of food and animal feed for centuries (Shurtleff & Aoyagi, 2009). The relatively high protein (40%) and oil (20%) content found in soybean seeds can be exploited for a broad range of agricultural and industrial applications, which differentiates it from most other commodities. Over the past century, tremendous advances in plant breeding have significantly expanded the yield potential of soybean, especially in the U.S. The combination of its valuable seed composition and high productivity has allowed soybean to become one of the most highly traded and heavily utilized commodities in the world (Wilcox, 2004).

While many industries rely on the successful utilization of soybean for production of vegetable oil, pharmaceutical products, natural emulsifiers, protective coatings, and more, the vast majority of soybeans produced in the U.S. are devoted to soybean meal used for animal feed (*USDA Coexistence Factsheets - Soybeans,* 2015). According to SoyStats, Monogastric animals such as poultry and swine are the primary beneficiaries of soybean meal, which together comprised approximately 80% of soybean meal consumption in 2018. A growing portion of soybean acreage annually is also being produced for human consumption as soy food products, such as natto, tofu, soy milk, edamame, and more, which have been historically consumed in East Asia and have recently begun to gain traction in the U.S. marketplace (Mayta et al., 2014).

The continued success of the U.S. soybean industry is vital to support the global food system and economy. Therefore, it is essential for plant breeders to not only continue to enhance soybean seed composition for improved utilization in established industries, but also to develop seed inputs that support growth of new industries that can further strengthen soybean's importance in the global marketplace. In this thesis, I sought to use economic and chemometric studies to support breeding programs in pursuit of these goals. In the first two studies, I aim to use field trials and contingent valuation (CV) survey methodology to foster development of enterprise budgets and marketing strategy that can support growth of the commercial industry for vegetable-type soybean – which has shown increasing promise in domestic markets over the past few decades – in the U.S. In my final study, I also describe the development of near-infrared reflectance spectroscopy (NIRS) calibrations that allow for more rapid prediction of the relative concentrations of sucrose, raffinose, and stachyose – the three most abundant sugars in soybean seed which hold major implications for its utilization in soymeal and soyfood industries – to enhance phenotyping efficiency in breeding programs.

References

Mayta, J., Chen, P., Popp, M. P., Dong, D., Wu, C. J., Zhang, B., Smith, S. F., & Scaboo, A. M. (2014). *Break-even profitability for food-grade specialty soybeans*. *2*(2), 01–11. https://doi.org/10.12735/as.v2i2p01

Shurtleff, W., & Aoyagi, A. (2009). *History of Edamame, Green Vegetable Soybeans, and Vegetable-Type Soybeans (1275-2009): Extensively Annotated Bibliography and Sourcebook*. Soyinfo Center.

Wilcox, J. R. (2004). World Distribution and Trade of Soybean. In *Soybeans: Improvement, Production, and Uses* (3rd ed., pp. 7–8). American Society of Agronomy, Crop Science Society of America, Soil Science Society of America. https://doi.org/10.2134/agronmonogr16.3ed.c1

Chapter 1: Review of Literature

Fresh Market Edamame as an Alternative Crop

Virginia has been a historically important state for tobacco since the 1600's and currently ranks third in the US in total tobacco production. Up until deregulation in 2004, revenue of smallholder tobacco producers in the US was secured by government regulation which restricted production and guaranteed an artificially high price for growers. Since then, increased foreign competition and expiration of Tobacco Transition Payment Program (TTPP) payouts in 2014 has left many smallholder farmers unable to compete at the current market price and in search of alternative crops to produce (Bomey, 2015). Fresh market edamame is a similarly low-acreage, high-value crop that presents an ideal alternative for tobacco production. Many research and extension efforts have been directed at educating growers on production, marketing, and consumer demand for edamame. Adoption, however, remains low.

For an acre of Flue-cured tobacco grown in Virginia, it is estimated that a price of $2.01 (0.34 cent profit margin) is needed for a desired net income of $1,000 above total costs assuming an upper-end yield of 3,000 lb yield (Eberly, 2014). The lower-end of yield estimates for edamame are 6,000 lb per acre, with potential to go as high as 10,000 lb per acre (Kaiser & Ernst, 2020). A 2011 marketing study performed in Richmond, VA revealed a $3 per pound price for wholesale edamame and a $4 per pound price for direct-sale edamame. Given these estimates, fresh market edamame may have the potential for per acre revenue to be $18,000-30,000 for wholesale and $24,000-$40,000 for direct-sale. However, there remains general uncertainty regarding the profitability of fresh market edamame enterprise, given the general lack of information available about its production costs, especially in regards to labor.

Enterprise budgets are tools commonly used by cooperative extension to provide detailed information regarding costs, inputs, yield, and revenue of an enterprise to growers (Curtis et. al., 2005). These budgets, which can be customized by growers according to their specific operations, greatly simplify estimation of profitability. They can also greatly facilitate comparison of two separate enterprises for growers interested in adding or replacing their current enterprise with an alternative crop. Enterprise budgets can be developed using cost estimates from similarly produced crops and through conducting field trials where production is simulated and costs are carefully recorded. Currently, no enterprise budgets are available to estimate production costs and potential profitability of edamame. Given edamame's relatively unique production, field trials that can allow for better estimates of its required labor hours and production costs are needed for development of enterprise budgets that can be used by tobacco and other specialty crop producers interested in adopting an edamame enterprise.

Interest & Demand for Edamame in the U.S.

Relatively few studies have been conducted on consumer interest and demand for edamame and related products in the US to date, however the few studies that were conducted have been promising. Simonne et al. (2000) reported that frozen immature soybean seeds performed favorably in sensory evaluations compared to other legume products such as frozen peas and lima beans, and that freeze-dried soybean snacks (F87) may be a potential product use for immature soybean. Young et al. (2000) conducted a sensory panel comparing 31 green soybean genotypes among which considerable variability was observed in regards to sensory attributes. Several of these attributes were found to be significant for overall consumer acceptability of certain genotypes over others, suggesting this may be an important consideration in breeding efforts. A series of studies in the metro-Philadelphia area in the mid-2000's by Kelley & Sánchez (2005) and D.

Montri et al. (2006) showed generally positive results for consumer familiarity and interest in edamame as well as for edamame-based products. Findings from a more recent study by Sciarappa et al. (2016) suggest that growing consumer interest in edamame may be explained by the relatively rapid growth of Asian/Indian populations in the US which may create increased market opportunities for domestic growers to produce higher-value, specialty Asian crops. In addition, the numerous purported health benefits of soy (Ruiz-Larrea et al., 1997; Taku et al., 2007; Zaheer & Akhtar, 2017) combined with its high quality protein that contains all essential amino acids (Rizzo & Baroni, 2018) give edamame a distinguished nutritional profile over many other vegetables in the vegetable marketplace which may also increase its potential as a plant-based protein source.

Consumer preferences and willingness-to-pay (WTP)

The selection and quality of fresh vegetables available to consumers in the domestic marketplace has drastically improved over the past few decades (Pollack, 2001). Consumer preference for fresh produce can be largely attributed to perceived losses in nutritional quality from blanching and freezing that is associated with frozen produce. Numerous studies have suggested that the nutritional quality of fresh and frozen produce is comparable, and that fresh produce are just as likely to undergo losses in nutritional quality as frozen produce (Heinrichs, 2016; Rickman, Barrett, et al., 2007; Rickman, Bruhn, et al., 2007). Nevertheless, an extensive consumer study by Heinrichs (2016) revealed that even after providing participants with ample information on the comparable nutritional quality of fresh and frozen produce items, they still showed a considerably stronger inclination towards fresh. In the case of edamame, several studies have suggested relatively low production costs and the potential profitability of fresh market edamame (Garber & Neill, 2019; Lord, Neill, et al., 2019; Sharma, 2013; Shockley et al., 2011). However, it is unknown whether U.S. consumers would be willing to pay considerably more for a fresh product when a frozen product – which consumers may already be accustomed to and is available year-round – is also present.

Growing consumer inclination for convenience has become increasingly important in consumer food purchasing decisions in recent years (Brunner et al., 2010; Pollack, 2001). This has led to the growing popularity of fresh-cut or minimally processed food items in the domestic marketplace, which are supplied to consumers in forms that are relatively quick and easy to prepare (Sillani & Nassivera, 2015). The term convenience encompasses time, physical, and mental effort associated with meal preparation and has been shown to be positively correlated with a number of socioeconomic and attitudinal variables including household size, working status, and cooking enjoyment among others (Brunner et al., 2010; Candel, 2001). Lifestyle variables such as time pressure, value for money, and avoiding waste have also been shown to contribute to increased consumer preference for convenience (Buckley et al., 2007). Interestingly, Lockie et al. (2004) found that concern for convenience was negatively correlated with consumption of organic food, of which naturalness of the product was a particularly strong predictor. These findings were later corroborated by Brunner et al. (2010), who found that naturalness was a consistently strong, negative predictor for consumption of highly processed, minimally processed, and single-component food categories.

Many of the perceived improvements to quality associated with purchase of fresh produce are also associated with local food. For example, results from a web-based survey conducted in 2010 investigating consumer purchasing motivations and behavior for local food revealed that respondents considered local produce superior to domestically grown, non-local produce in freshness, eating quality, food safety, and nutritional value (Onozaka et al., 2010). Evidence suggests that local food need not confer any actual benefits to earn a premium, according to a recent study by Fan et al. (2019) who demonstrated that inclusion of locally grown information alone is capable of increasing consumer WTP. Aside from perceived gains in quality, consumers also associate benefits to the regional economy with the purchase of local products (Darby et al., 2008). As consumers increasingly desire local products, studies have shown that chefs, restaurants, and supermarkets have also expressed interest in sourcing ingredients locally and increasing availability of locally sourced food items to meet this demand (D. N. Montri et al., 2006; Reynolds-Allie & Fields, 2011).

The purchase of organic vegetables has been drastically risen in the U.S. for the past few decades, with its consumer base becoming increasingly diverse (Dettmann & Dimitri, 2009; Stevens-Garmon et al., 2007). In 2003 the USDA established the National Organic Program (NOP), which regulates organic marketing by certifying that products are produced in accordance with a set of approved inputs and substances that are consistent with regenerative and sustainable agricultural principles. As a result of the myriad provisions guaranteed by the NOP and the additional cost to obtain certification, organic produce often commands substantial premiums in the marketplace. Interestingly, evidence suggests that many consumers may not necessarily value all provisions of the USDA Certified Organic label equally. Some, for example, are only willing-to-pay more for the guaranteed use of non-GMO seed inputs but are not necessarily concerned about which synthetic fertilizers and pesticides are used during production. To test this hypothesis, Bernard, Zhang, & Gifford (2006) conducted a study intended to mimic market conditions by observing subjects in an experimental auction setting where they were asked to indicate their willingness to pay for various food items that fell under conventional, organic, and non-GMO categories. While overall results showed the highest WTP estimates for food items in the organic category, further analysis revealed that beyond the non-GM attribute, subjects did not appear willing to pay significantly extra for the remaining attributes of the organic category. A recent study by Wolfe et al. (2018) recently demonstrated that non-GMO edamame may already hold some appeal to consumers with risk aversion to GMO products, and that this may even compensate for shortcomings in sensory quality.

The inclination of consumers to purchase eco-labelled products can be described by three main value bases: egoistic, social-altruistic, and biospheric (Stern & Dietz, 1994). Egoistic valuation refers to some perceived benefit that an individual expects to realize for themselves, whereas social-altruistic and biospheric values refer to a perceived benefit either to society as a whole or to the earth, respectively. Because social-altruistic and biospheric values involve the same personal moral norms and reflect similarly broad societal benefit, they are often considered jointly (Stern & Dietz, 1994). Studies have shown that such social-altruistic and biospheric values have been translated to higher WTP for regional and organic products (Rahman & Reynolds, 2017; Shin et al., 2017; Umberger et al., 2009). In addition to the three value bases outlined by Stern & Dietz, results from Onozaka et al. (2011) suggest that shopping locale may also influence consumer valuation for various labeling campaigns by sorting consumers through different sustainability claims made in-store or within the market venue.

Soluble Carbohydrates in Soybean Seed

Carbohydrates are the third most abundant group of chemical constituents in soybean seed behind protein and oil. Carbohydrates comprise approximately 35% of soybean seed dry weight, and 40% of these carbohydrates are soluble (Zeng et al., 2014). Soluble carbohydrates found in soybean seeds include: sucrose, raffinose, stachyose, glucose, and fructose. The formation of disaccharide sucrose results from the enzymatic combination of D-glucose with D-fructose (Hassid & Doudoropf, 1950). Raffinose and stachyose, collectively referred to as raffinose family oligosaccharides (RFO's), are in turn derived from disaccharide sucrose. Raffinose and stachyose contain either one or two additional galactosyl residues, respectively, connected to the glucose moiety of sucrose via α-1,6 glycosidic linkages (Meyer et al., 2015). In soybean seed, sucrose is present in the largest concentration, followed by stachyose and raffinose, with glucose and fructose only found in trace amounts (Y. Wang et al., 2014). Given their larger relative concentrations in seed, sucrose, raffinose, and stachyose are primarily targeted by breeding programs to improve the utility of feed- and food-grade soybeans.

Variability studies have found that sucrose, raffinose, and stachyose concentration in soybean seeds can naturally range from 3.0-10.2%, 0.4-1.8%, and 1.2-3.8% of soybean seed dry weight, respectively (Hymowitz & Collins, 1974). Several factors may influence the concentration of sucrose, raffinose, and stachyose in soybean seed. Firstly, sugar content in soybeans can be largely impacted by temperature. Wolf et al. (1982) studied changes in soybean seed constituents in response to changes in temperature, and found that sucrose content greatly decreased as temperature increased. This decrease appeared to be inversely correlated with oil content. In a more recent study, Kumar et al. (2010) also observed that soybean cultivars grown at colder locations tended to have much higher sucrose content. Several studies have suggested that protein content may also influence sucrose content in soybean seeds. In an experiment aimed at understanding the relationship of seed protein to soluble sugars in soybean, Krober & Cartter (1962) found that high protein soybean cultivars contained the lowest total sugar content. Numerous studies have been conducted on randomly selected groups of elite and high protein breeding lines and germplasm to study interrelationships between soybean seed constituents. In all three experiments, a strong inverse relationship between protein and sucrose content across genotypes was observed (Hartwig et al., 1997; Hymowitz et al., 1972; Wilcox & Shibles, 2001).

The desired concentration of sucrose, raffinose, and stachyose sugars in soybean varieties greatly varies depending on market needs. In soymeal and most soyfood products, sucrose is desired in high concentrations for its role in conferring sweetness (Kumar et al., 2011; Rackis, 1975), while lower sucrose content is desired in fermented soybean (natto) varieties for increased processing efficiency (Zeng et al., 2014). Humans and monogastric animals can readily break down sucrose, however they lack the α-1,6 glycosidase enzyme capable of breaking down RFO's in the gastrointestinal tract (Rackis, 1975). This results in increased incidence of flatulence in humans and reduced feed efficiency in animals (Coon et al., 1990; Kumar et al., 2010; Suarez et al., 1999). To this end, RFO's have been desired in very low concentrations historically. More recent evidence suggests that RFO's may actually be a useful prebiotic substance (Wongputtisin et al., 2015) for humans which may encourage development of RFO-rich soybean cultivars. Studies have found that the inability of RFO's to be digested in the gastrointestinal tract allows them to reach the colon intact, where beneficial bifidogenic microorganisms are capable of cleaving α-1,6 glycosidic linkages to break down RFO's and utilize them as a food source (Espinosa-Martos & Rupérez, 2006; Švejstil et al., 2015).

Quantification of Soybean Seed Sugars

Black & Bagley (1978) were the first to describe a method to determine sugar content in ground soybean seed (soy flour) using high performance liquid chromatography (HPLC). The method proposed by Black & Bagley, which incorporated use of a known internal standard into the experimental analysis, allowed for a considerably more accurate and reproducible means of quantitatively determining soybean sugar concentration compared to previously proposed methods using paper chromatography (Lineback & Ke, 1975), gas chromatography (Hymowitz et al., 1972), and colorimetric methods (Kawamura, 1967). Today, the HPLC method described by Black &

Bagley (1978) remains the most widely used to determine sugar content in soybean samples (Giannoccaro et al., 2008).

Despite its reliability, the time and per sample cost required for sugar determination via HPLC are major limitations for its use, especially in breeding programs (Maughan et al., 2000). In addition, the refractive index (RI) detection system that is most commonly coupled with HPLC instruments decreases in sensitivity at appreciably low concentrations. As a result, many studies have sought to improve upon the method described by Black & Bagley (1978) using more recently developed instruments, columns, and detection systems that offer increased functionality and/or sensitivity. For example, Kennedy et al. (1985) proposed an HPLC method using a Waters Dextropak column and observed comparable accuracy and reproducibility while reducing sample preparation time. Giannoccaro et al. (2008) compared the performance of more traditional HPLC system with a high-performance anion exchange chromatography coupled with pulsed-amperometric detection (HPAEC-PAD) and enzymatic methods to determine which would be most suitable for breeding programs. While he observed similarly accurate and reproducible results across all three methods, the HPAEC-PAD system was able to reduce time and quantify glucose and fructose sugars that the standard HPLC system could not. Enzymatic methods for sugar quantification have been described by Kumar et al. (2010), Stitt et al. (1989), and Teixeira et al. (2012). While these methods prevent the need to purchase expensive HPLC instruments, they can still be cost-prohibitive for breeding programs and also lack the ability to quantify soybean RFO's separately.

Near-infrared Reflectance Spectroscopy (NIRS)

The near-infrared light occupies the segment of the electromagnetic spectrum that falls just outside of the visible region, from approximately 780 to 2500nm. Near-infrared wavelengths of light are differentially absorbed by objects based on their specific chemical make-up. Absorbance of wavelengths by functional groups such as C-H, N-H, and O-H bonds form a spectral fingerprint based on the relative amounts of fat, protein, moisture, and other important chemical constituents present in each sample (Stark et al., 1986). Near-infrared reflectance spectroscopy (NIRS) refers to the practice of exploiting these differential patterns of near-infrared wavelength absorbance in order to predict concentration of various chemical constituents by modeling differences in spectral absorbance to a set of previously obtained reference chemistry values. Development of these prediction models, more commonly referred to as calibrations or calibration equations, has been performed for many products in commercial industry, especially in food and agriculture (Baianu & Guo, 2011).

NIRS holds many advantages that make it an ideal candidate to replace HPLC for sugar profiling. Firstly, NIRS instruments are capable of collecting data non-destructively, considerably faster, and with much improved cost-efficiency as compared to HPLC and other laboratory-based quantification methods. Secondly, NIRS offers estimates of chemical constituents that contain high predictive accuracy and reproducibility. Thirdly, samples need only minimal sample preparation before scanning and do not require chemicals and extraction such as is needed to prepare samples for sugar analysis via HPLC (Baianu & Guo, 2011). Another major advantage of NIRS is its ease of operation as opposed to HPLC, where experienced personnel are needed to operate instruments. Today, NIRS instruments are regularly employed to collect information on protein, oil, fatty acids, and moisture for a wide array of crops. More recently, calibrations have begun to be developed for numerous value-added parameters in some crops and products, such as fatty acids in soybean oil (Karn et al., 2017), beta-glucan content in barley (Ringsted et al., 2017), and tryptophan content in quality maize (Shiferaw et al., 2017). NIRS has also been successfully applied to predict carbohydrates in numerous fruits and vegetables (Katayama et al., 1996; Kawano, 1994; López et al., 2017; Walsh et al., 2004; Yan-de & Yi-bin, 2004).

There has been considerable interest in using NIRS to predict soluble carbohydrates in soybean seeds. NIRS for determination of sugars in soybean has been used previously in the literature to expedite data collection (Baianu & Guo, 2011; Bellaloui et al., 2010; Wilcox & Shibles, 2001). However, only few studies have aimed to develop calibrations for widescale use in breeding programs, where prediction accuracy and representativeness at the upper and lower ends of the naturally observable concentration range are particularly important. Choung (2010) was the first to attempt calibration development for sucrose in soybean seed. While his prediction model for sucrose showed 92% accuracy, the calibration set lacked genotypes with greater than 7% sucrose, limiting its utility. Baianu & Guo (2011) sought to develop NIR calibrations for a series of soy food components including soybean seed sugars, and found that even after using 12,000+ samples over four years, the overall prediction range remained narrow. Sato et al. (2012) were later able to extend the upper bound of the prediction range for sucrose by artificially spiking samples in lab. By supplementing their calibration set with these artificially spiked samples, predictive accuracy was considerably observed, thereby demonstrating the importance of concentration range for model creation. Aside from sucrose, calibration development for soybean RFO’s has been seldomly attempted hitherto with the few conducted attempts offering prediction accuracy too poor for application in commercial and breeding operations (Choung, 2010).

References

Baianu, I., & Guo, J. (2011). NIR Calibrations for Soybean Seeds and Soy Food Composition Analysis: Total Carbohydrates, Oil, Proteins and Water Contents. *Nature Precedings*. https://doi.org/10.1038/npre.2011.6611.1

Bellaloui, N., Smith, J. R., Gillen, A. M., & Ray, J. D. (2010). Effect of Maturity on Seed Sugars as Measured on Near-Isogenic Soybean (Glycine max) Lines. *Crop Science*, *50*(5), 1978–1987. https://doi.org/10.2135/cropsci2009.10.0596

Bernard, J. C., Zhang, C., & Gifford, K. (2006). An Experimental Investigation of Consumer Willingness to Pay for Non-GM Foods When an Organic Option Is Present. *Agricultural and Resource Economics Review*, *35*(2), 1–12.

Black, L. T., & Bagley, E. B. (1978). Determination of oligosaccharides in soybeans by high pressure liquid chromatography using an internal standard. *Journal of the American Oil Chemists' Society*, *55*(2), 228–232.

Bomey, N. (2015, September 2). *Thousands of farmers stopped growing tobacco after deregulation payouts*. USA TODAY. https://www.usatoday.com/story/money/2015/09/02/thousands-farmers-stopped-growing-tobacco-after-deregulation-payouts/32115163/

Brunner, T. A., van der Horst, K., & Siegrist, M. (2010). Convenience food products. Drivers for consumption. *Appetite*, *55*(3), 498–506. https://doi.org/10.1016/j.appet.2010.08.017

Buckley, M., Cowan, C., & McCarthy, M. (2007). The convenience food market in Great Britain: Convenience food lifestyle (CFL) segments. *Appetite*, *49*(3), 600–617. https://doi.org/10.1016/j.appet.2007.03.226

Candel, M. (2001). Consumers' convenience orientation towards meal preparation: Conceptualization and measurement. *Appetite*, *36*(1), 15–28. https://doi.org/10.1006/appe.2000.0364

Choung, M.-G. (2010). Determination of Sucrose Content in Soybean Using Near-infrared Reflectance Spectroscopy. *Journal of the Korean Society for Applied Biological Chemistry*, *53*(4), 478–484. https://doi.org/10.3839/jksabc.2010.073

Coon, C. N., Leske, K. L., Akavanichan, O., & Cheng, T. K. (1990). Effect of Oligosaccharide-Free Soybean Meal on True Metabolizable Energy and Fiber Digestion in Adult Roosters1. *Poultry Science*, *69*(5), 787–793. https://doi.org/10.3382/ps.0690787

Curtis, K., Harris, T., Riggs, W. (2005). Importance & use of enterprise budgets in agricultural operations. University of Nevada Cooperative Extension. Retrieved from: https://extension.unr.edu/publication.aspx?PubID=2383

Darby, K., Batte, M. T., Ernst, S. C., & Roe, B. E. (2006). *Willingness to pay for locally produced foods: A customer intercept study of direct market and grocery store shoppers* (Selected Paper No. 379-2016–21958). American Agricultural Economics Association Annual Meeting, Long Beach, California. https://doi.org/10.22004/ag.econ.21336

Dettmann, R. L., & Dimitri, C. (2009). Who's Buying Organic Vegetables? Demographic Characteristics of U.S. Consumers. *Journal of Food Products Marketing*, *16*(1), 79–91. https://doi.org/10.1080/10454440903415709

Eberly, E. (2014). Flue-cured tobacco budget information. *Farm Business Management*. https://www.arec.vaes.vt.edu/content/dam/arec_vaes_vt_edu/southern-piedmont/Documents/tobacco-budget-2014.pdf

Espinosa-Martos, I., & Rupérez, P. (2006). Soybean oligosaccharides. Potential as new ingredients in functional food. *Nutricion Hospitalaria*, *21*(1), 92–96.

Fan, X., Gómez, M. I., & Coles, P. S. (2019). Willingness to Pay, Quality Perception, and Local Foods: The Case of Broccoli. *Agricultural and Resource Economics Review*, *48*(3), 414–432.

Garber, B., & Neill, C. (2019). Edamame: Costs, Revenues, and Profitability. *No. AAEC-189P*.

Giannoccaro, E., Wang, Y.-J., & Chen, P. (2008). Comparison of two HPLC systems and an enzymatic method for quantification of soybean sugars. *Food Chemistry*, *106*(1), 324–330. https://doi.org/10.1016/j.foodchem.2007.04.065

Hartwig, E. E., Kuo, T. M., & Kenty, M. M. (1997). Seed Protein and its Relationship to Soluble Sugars in Soybean. *Crop Science*, *37*(3), cropsci1997.0011183X003700030013x. https://doi.org/10.2135/cropsci1997.0011183X003700030013x

Hassid, W. Z., & Doudoropf, M. (1950). Enzymatic Synthesis of Sucrose and Other Disaccharides. In C. S. Hudson & S. M. Cantor (Eds.), *Advances in Carbohydrate Chemistry* (Vol. 5, pp. 29–48). Academic Press. https://doi.org/10.1016/S0096-5332(08)60333-1

Heinrichs, P. A. (2016). *Consumer preferences and willingness to pay for fresh and frozen vegetables* [PhD Thesis, University of Illinois at Urbana-Champaign]. https://pdfs.semanticscholar.org/02fa/c09206b0db3db090197ed217d2ba47842c5c.pdf

Hymowitz, T., & Collins, F. I. (1974). Variability of Sugar Content in Seed of Glycine max (L.) Merrill and G. soja Sieb. And Zucc.1. *Agronomy Journal*, *66*, 239–240. https://doi.org/10.2134/agronj1974.00021962006600020017x

Hymowitz, T., Collins, F. I., Panczner, J., & Walker, W. M. (1972). Relationship Between the Content of Oil, Protein, and Sugar in Soybean Seed1. *Agronomy Journal*, *64*(5), 613–616. https://doi.org/10.2134/agronj1972.00021962006400050019x

Kaiser, C., & Ernst, M. (2020). Center for Crop Diversification Profile: Edamame (CCD-CP-94). *CCD-CP-94*.

Karn, A., Heim, C., Flint-Garcia, S., Bilyeu, K., & Gillman, J. (2017). Development of Rigorous Fatty Acid Near-Infrared Spectroscopy Quantitation Methods in Support of Soybean Oil Improvement.

Journal of the American Oil Chemists' Society, *94*(1), 69–76. https://doi.org/10.1007/s11746-016-2916-4

Katayama, K., Komaki, K., & Tamiya, S. (1996). Prediction of Starch, Moisture, and Sugar in Sweetpotato by Near Infrared Transmittance. *HortScience*, *31*(6), 1003–1006. https://doi.org/10.21273/HORTSCI.31.6.1003

Kawamura, S. (1967). Quantitative paper chromatography of sugars of the cotyledon, hull, and hypocotyl of soybeans of selected varieties. *Technical Bulletin of Faculty of Agriculture*, *18*(2).

Kawano, S. (1994). Non-Destructive NIR Quality Evaluation of Fruits and Vegetables in Japan. *NIR News*, *5*(6), 10–12. https://doi.org/10.1255/nirn.278

Kelley, K. M., & Sanchez, E. S. (2005). Accessing and understanding consumer awareness of and potential demand for edamame. *Hortscience: A Publication of the American Society for Hortcultural Science*, *40*(5), 1347–1353.

Kennedy, I. R., Mwandemele, O. D., & McWhirter, K. S. (1985). Estimation of sucrose, raffinose and stachyose in soybean seeds. *Food Chemistry*, *17*(2), 85–93. https://doi.org/10.1016/0308-8146(85)90077-9

Krober, O. A., & Cartter, J. L. (1962). Quantitative Interrelations of Protein and Nonprotein Constituents of Soybeans1. *Crop Science*, *2*(2), cropsci1962.0011183X000200020028x. https://doi.org/10.2135/cropsci1962.0011183X000200020028x

Kumar, V., Rani, A., Goyal, L., Dixit, A. K., Manjaya, J. G., Dev, J., & Swamy, M. (2010). Sucrose and Raffinose Family Oligosaccharides (RFOs) in Soybean Seeds As Influenced by Genotype and Growing Location. *Journal of Agricultural and Food Chemistry*, *58*(8), 5081–5085. https://doi.org/10.1021/jf903141s

Kumar, V., Rani, A., Goyal, L., Pratap, D., Billore, S. D., & Chauhan, G. S. (2011). Evaluation of Vegetable-Type Soybean for Sucrose, Taste-Related Amino Acids, and Isoflavones Contents. *International Journal of Food Properties*, *14*(5), 1142–1151. https://doi.org/10.1080/10942911003592761

Lineback, D. R., & Ke, C. H. (1975). Starches and low-molecular-weight carbohydrates from chick pea and horse bean flours. *Cereal Chemistry*, *52*, 334–347. Scopus.

Lockie, S., Lyons, K., Lawrence, G., & Grice, J. (2004). Choosing organics: A path analysis of factors underlying the selection of organic food among Australian consumers. *Appetite*, *43*(2), 135–146. https://doi.org/10.1016/j.appet.2004.02.004

Lord, N., Neill, C., & Zhang, B. (2019). Production and Economic Considerations for Fresh Market Edamame in Southwest Virginia. *AAEC-188P*.

Lord, N., Zhang, B., Kuhar, T., Carneiro, R., Sutton, K., & Yu, D. (2019). USDA Edamame Project. *SPES-104NP*. https://www.pubs.ext.vt.edu/content/dam/pubs_ext_vt_edu/spes/SPES-104/SPES-104.pdf

Maughan, P. J., Maroof, M. A. S., & Buss, G. R. (2000). Identification of quantitative trait loci controlling sucrose content in soybean (Glycine max). *Molecular Breeding*, *6*(1), 105–111. https://doi.org/10.1023/A:1009628614988

Meyer, T. S. M., Miguel, Â. S. M., Fernández, D. E. R., & Ortiz, G. M. D. (2015). Biotechnological Production of Oligosaccharides—Applications in the Food Industry. In *Food Production and Industry* (pp. 26–78). IntechOpen. https://www.intechopen.com/books/food-production-and-industry/biotechnological-production-of-oligosaccharides-applications-in-the-food-industry

Montri, D. N., Kelley, K. M., & Sánchez, E. S. (2006a). Direct marketing edamame (glycine max [L.] merrill) to professional chefs. *Journal of Extension*, *44*(1), 183–194.

Montri, D. N., Kelley, K. M., & Sánchez, E. S. (2006b). Consumer Interest in Fresh, In-shell Edamame and Acceptance of Edamame-based Patties. *HortScience*, *41*(7), 1616–1622. https://doi.org/10.21273/HORTSCI.41.7.1616

Onozaka, Y., Nurse, G., & McFadden, D. T. (2010). Local Food Consumers: How Motivations and Perceptions Translate to Buying Behavior. *Choices*, *25*(1). JSTOR. https://www.jstor.org/stable/choices.25.1.03

Onozaka, Y., Nurse, G., & McFadden, D. T. (2011). Defining Sustainable Food Market Segments: Do Motivations and Values Vary by Shopping Locale? *American Journal of Agricultural Economics*, *93*(2), 583–589. JSTOR.

Pollack, S. L. (2001). Consumer demand for fruit and vegetables: The U.S. example. In *Changing structure of global food consumption and trade* (pp. 49–54). DIANE Publishing.

Rackis, J. J. (1975). Oligosaccharides of Food Legumes: Alpha-Galactosidase Activity and the Flatus Problem. In *Physiological Effects of Food Carbohydrates* (Vol. 15, pp. 207–222). AMERICAN CHEMICAL SOCIETY. https://doi.org/10.1021/bk-1975-0015.ch013

Rahman, I., & Reynolds, D. (2017). Organic Wine: The Influence of Biospheric, Altruistic, and Egoistic Values on Purchase Intention, Willingness to Pay More, and Willingness to Sacrifice. *International Journal of Hospitality Beverage Management*, *1*(1). https://doi.org/10.34051/j/2019.1

Reynolds-Allie, K., & Fields, D. (2011, February 5). *Alabama Restaurant Preferences and Willingness to Pay for Local Food: A Choice Based Approach* (Selected Paper No. 1371-2016–108892). American Agricultural Economics Association Annual Meeting, Corpus Christi, Texas. https://doi.org/10.22004/ag.econ.98822

Rickman, J. C., Barrett, D. M., & Bruhn, C. M. (2007). Nutritional comparison of fresh, frozen and canned fruits and vegetables. Part 1. Vitamins C and B and phenolic compounds. *Journal of the Science of Food and Agriculture*, *87*(6), 930–944. https://doi.org/10.1002/jsfa.2825

Rickman, J. C., Bruhn, C. M., & Barrett, D. M. (2007). Nutritional comparison of fresh, frozen, and canned fruits and vegetables II. Vitamin A and carotenoids, vitamin E, minerals and fiber. *Journal of the Science of Food and Agriculture*, *87*(7), 1185–1196. https://doi.org/10.1002/jsfa.2824

Ringsted, T., Ramsay, J., Jespersen, B. M., Keiding, S. R., & Engelsen, S. B. (2017). Long wavelength near-infrared transmission spectroscopy of barley seeds using a supercontinuum laser: Prediction of mixed-linkage beta-glucan content. *Analytica Chimica Acta*, *986*, 101–108. https://doi.org/10.1016/j.aca.2017.07.008

Rizzo, G., & Baroni, L. (2018). Soy, Soy Foods and Their Role in Vegetarian Diets. *Nutrients*, *10*(1). https://doi.org/10.3390/nu10010043

Ruiz-Larrea, M. B., Mohan, A. R., Paganga, G., Miller, N. J., Bolwell, G. P., & Rice-Evans, C. A. (1997). Antioxidant Activity of Phytoestrogenic Isoflavones. *Free Radical Research*, *26*(1), 63–70. https://doi.org/10.3109/10715769709097785

Sato, T., Zahlner, V., Berghofer, E., Lošák, T., & Vollmann, J. (2012). Near-infrared reflectance calibrations for determining sucrose content in soybean breeding using artificial reference samples. *Plant Breeding*, *131*(4), 531–534. https://doi.org/10.1111/j.1439-0523.2012.01975.x

Sciarappa, W. J., Simon, J., Govindasamy, R., Kelley, K., Mangan, F., Zhang, S., Arumugam, S., Nitzsche, P., Vranken, R. V., Komar, S., Ayeni, A., McAvoy, G., Park, C., Reichert, W., Byrnes, D., Wu, Q., Schilling, B., & Orellana, R. (2016). Asian Crops Overview: Consumer Preference and Cultivar Growth on the East Coast of the United States. *HortScience*, *51*(11), 1344–1350. https://doi.org/10.21273/HORTSCI11040-16

Sharma, K. P. (2013). *Varietal yield stability and appropriate management for quality organic edamame production* (pp. 10–24). The Sharing Farm Society. https://certifiedorganic.bc.ca/programs/osdp/I-166_Edamame2_Final_Report.pdf

Shiferaw, L., Tsegay, G., & Asamenew, G. (2017). Near-Infrared Reflectance Spectroscopy (NIRs) for Determination of Tryptophan Content in Quality Protein Maize (QPM). *Science Journal of Analytical Chemistry*, *5*(1), 8. https://doi.org/10.11648/j.sjac.20170501.12

Shin, Y. H., Moon, H., Jung, S. E., & Severt, K. (2017). The effect of environmental values and attitudes on consumer willingness to pay more for organic menus: A value-attitude-behavior approach. *Journal of Hospitality and Tourism Management*, *33*, 113–121. https://doi.org/10.1016/j.jhtm.2017.10.010

Sillani, S., & Nassivera, F. (2015). Consumer behavior in choice of minimally processed vegetables and implications for marketing strategies. *Trends in Food Science & Technology*, *46*(2, Part B), 339–345. https://doi.org/10.1016/j.tifs.2015.07.004

Simonne, A. H., Weaver, D. B., & Wei, C. (2000). Immature soybean seeds as a vegetable or snack food: Acceptability by American consumers. *Innovative Food Science & Emerging Technologies*, *1*(4), 289–296. https://doi.org/10.1016/S1466-8564(00)00021-7

Stark, E., Luchter, K., & Margoshes, M. (1986). Near-Infrared Analysis (NIRA): A Technology for Quantitative and Qualitative Analysis. *Applied Spectroscopy Reviews*, *22*(4), 335–399. https://doi.org/10.1080/05704928608060440

Stern, P. C., & Dietz, T. (1994). The Value Basis of Environmental Concern. *Journal of Social Issues*, *50*(3), 65–84. https://doi.org/10.1111/j.1540-4560.1994.tb02420.x

Stevens-Garmon, J., Huang, C. L., & Lin, B.-H. (2007). Organic Demand: A Profile of Consumers in the Fresh Produce Market. *Choices*, *22*(2), 109–115. JSTOR.

Stitt, M., Lilley, R. McC., Gerhardt, R., & Heldt, H. W. (1989). [32] Metabolite levels in specific cells and subcellular compartments of plant leaves. In *Methods in Enzymology* (Vol. 174, pp. 518–552). Academic Press. https://doi.org/10.1016/0076-6879(89)74035-0

Suarez, F. L., Springfield, J., Furne, J. K., Lohrmann, T. T., Kerr, P. S., & Levitt, M. D. (1999). Gas production in humans ingesting a soybean flour derived from beans naturally low in oligosaccharides. *The American Journal of Clinical Nutrition*, *69*(1), 135–139. https://doi.org/10.1093/ajcn/69.1.135

Švejstil, R., Musilová, Š., & Rada, V. (2015). Raffinose-Series Oligosaccharides in Soybean Products. *Scientia Agriculturae Bohemica*, *46*(2), 73–77. https://doi.org/10.1515/sab-2015-0019

Taku, K., Umegaki, K., Sato, Y., Taki, Y., Endoh, K., & Watanabe, S. (2007). Soy isoflavones lower serum total and LDL cholesterol in humans: A meta-analysis of 11 randomized controlled trials. *The American Journal of Clinical Nutrition*, *85*(4), 1148–1156. https://doi.org/10.1093/ajcn/85.4.1148

Teixeira, A. I., Ribeiro, L. F., Rezende, S. T., Barros, E. G., & Moreira, M. A. (2012). Development of a method to quantify sucrose in soybean grains. *Food Chemistry*, *130*(4), 1134–1136. https://doi.org/10.1016/j.foodchem.2011.07.128

Umberger, W. J., McFadden, D. D. T., & Smith, A. R. (2009). Does altruism play a role in determining U.S. consumer preferences and willingness to pay for natural and regionally produced beef? *Agribusiness*, *25*(2), 268–285. https://doi.org/10.1002/agr.20194

Walsh, K. B., Golic, M., & Greensill, C. V. (2004). Sorting of Fruit Using near Infrared Spectroscopy: Application to a Range of Fruit and Vegetables for Soluble Solids and Dry Matter Content. *Journal of Near Infrared Spectroscopy*, *12*(3), 141–148.

Wang, Y., Chen, P., & Zhang, B. (2014). Quantitative trait loci analysis of soluble sugar contents in soybean. *Plant Breeding*, *133*(4), 493–498. https://doi.org/10.1111/pbr.12178

Wilcox, J. R., & Shibles, R. M. (2001). Interrelationships among Seed Quality Attributes in Soybean. *Crop Science*, *41*(1), 11–14. https://doi.org/10.2135/cropsci2001.41111x

Wolf, R. B., Cavins, J. F., Kleiman, R., & Black, L. T. (1982). Effect of temperature on soybean seed constituents: Oil, protein, moisture, fatty acids, amino acids and sugars. *Journal of the American Oil Chemists' Society*, *59*(5), 230–232. https://doi.org/10.1007/BF02582182

Wolfe, E., Popp, M., Bazzani, C., Nayga Jr, R. M., Danforth, D., Popp, J., Chen, P., & Seo, H.-S. (2018). Consumers' willingness to pay for edamame with a genetically modified label. *Agribusiness*, *34*(2), 283–299.

Wongputtisin, P., Ramaraj, R., Unpaprom, Y., Kawaree, R., & Pongtrakul, N. (2015). Raffinose family oligosaccharides in seed of Glycine max cv. Chiang Mai60 and potential source of prebiotic substances. *International Journal of Food Science & Technology*, *50*(8), 1750–1756. https://doi.org/10.1111/ijfs.12842

Yan-de, L., & Yi-bin, Y. (2004). Measurement of sugar content in Fuji apples by FT-NIR spectroscopy. *Journal of Zhejiang University-SCIENCE A*, *5*(6), 651–655. https://doi.org/10.1007/BF02840975

Young, G., Mebrahtu, T., & Johnson, J. (2000). Acceptability of green soybeans as a vegetable entity. *Plant Foods for Human Nutrition*, *55*(4), 323–333. https://doi.org/10.1023/A:1008164925103

Zaheer, K., & Akhtar, M. H. (2017). An updated review of dietary isoflavones: Nutrition, processing, bioavailability and impacts on human health. *Critical Reviews in Food Science and Nutrition*, *57*(6), 1280–1293. https://doi.org/10.1080/10408398.2014.989958

Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B., Orazaly, M., Florez-Palacios, L., Brye, K., Song, Q., & Cregan, P. (2014). Identification of Quantitative Trait Loci for Sucrose Content in Soybean Seed. *Crop Science*, *54*(2), 554–564. https://doi.org/10.2135/cropsci2013.01.0036

Chapter 2: Assessing feasibility of fresh market edamame production

Nick Lord[1], Clinton Neill[2,3], Bo Zhang[1]

[1]School of Plant & Environmental Sciences, Virginia Tech

[2]Department of Agricultural & Applied Economics, Virginia Tech

[3]Dyson School of Applied Economics & Management, Cornell University

This chapter was published by Virginia Cooperative Extension on July 15, 2019:

Lord, N., Neill, C., & Zhang, B. (2019). Production and Economic Considerations for Fresh Market Edamame in Southwest Virginia. *AAEC-188P*. https://digitalpubs.ext.vt.edu/vcedigitalpubs/3279487597799699/MobilePagedReplica.action?pm=2&folio=1#pg1

Introduction

Vegetable soybean [Glycine max (L.) *Merril*] is a type of food-grade soybean originating from East Asia. Edamame, pronounced "eh-dah-mah-may," is an immature soybean, harvested between growth stages R6 and R7 when pods are about 85%-90% filled and bright green. It is consumed by sucking the fleshy, immature beans, often bearing a smooth texture and sweet, nutty flavor, out of the pods. Most often served as a snack or appetizer, edamame is widely heralded for its potentially beneficial roles in cancer prevention, reduction of LDL ("bad") cholesterol, and cardiovascular health (Taku et al., 2007). It is also highly valued for its complete protein, containing all essential amino acids, and plant-promoting phytohormones called isoflavones that have been associated with cancer prevention (Ju, 2016). This myriad of health and nutritional benefits, as well as consumers' widespread assessment of edamame as being tasty, versatile, and easy to prepare, has generated considerably increased domestic demand for the product since its introduction to U.S. markets nearly two decades ago.

Currently, edamame production in the U.S. is primarily dependent on the availability of nearby commercial processing facilities that buy edamame from growers, process it, and either export or distribute it to grocery stores where it is predominately sold as a frozen product available to consumers year-round. Unfortunately, with the edamame industry in the U.S. still in its infancy, an overall lack of these processing facilities around the country has largely limited production (Bennett, 2016). This has forced the vast majority of domestic demand to be filled via foreign imports (Roseboro, 2012), highlighting the need to find ways to boost domestic production and allow U.S. growers to capitalize on edamame's premium market position.

Some reports conclude that processing methods used by commercial processing facilities may reduce the nutritional quality of edamame (Simonne, Smith, et al., 2000). Incidence of various foodborne pathogens such as *Listeria monocytogenes* in processing plant machinery has also been reported (Aguado et al., 2004). Growers who produce edamame for the fresh market sidestep these issues and forego the need for nearby processing facilities, but fresh market edamame production is time-sensitive and labor-intensive. Growers interested in producing fresh market edamame require tools that can help them make more calculated and informed production decisions to maximize use of time and inputs, boosting profits.

As with any new crop, growers must first determine if edamame is feasible to produce by estimating its production costs and potential expected returns (Hofstrand et al., 2020). To obtain information about production expenses, fresh market edamame production was simulated using field trials conducted during May to October 2017. This study is the first to present production and economic considerations for fresh market edamame in the state of Virginia and can be useful to growers considering its production. To evaluate the economic costs of producing and marketing edamame in Virginia, fresh edamame was grown in research plots at the Kentland Farms and sold to local distributors in Blacksburg, Virginia. We estimated costs of production using information from research plots, including labor hours associated with harvesting and post-harvest activities, as well as revenue from sales. Growers can use this information as a guideline to develop planting and crop rotation decisions. This will aid in assessing the economic cost of producing and marketing of these edamame products.

Materials & Methods

Plant Material

Genotypes V09-4192, V10-3637, and V13-1644, from the Virginia Tech Soybean Breeding Programs's large-seeded germplasm are specifically bred and adapted to Virginia's local climate. These genotypes contain large bean size and high sucrose content that are characteristic of edamame varieties. In addition, each of the three genotypes differed in relative maturity in order to stagger harvest and allow for a prolonged supply window. These lines were used in place of commercial edamame seed inputs to observe potential yield that can be obtained using locally adapted edamame varieties. Each genotype was planted in four-row plots, each 18 feet in length with 30 inch spacing between rows. Plots were planted at Kentland farms in Blacksburg, VA throughout May 2017 and harvested during the month of October.

Input Costs

Quantity of pre-plant fertilizer, pre-emergent herbicide, and fuel for machine planting was recorded during planting. To calculate total pre-plant costs, quantity of inputs used was multiplied by their unit price. Cost of seed for large-seeded, food-grade soybean varieties was determined from online sources. Assumptions were made for fixed costs of machinery and equipment, general overhead, and refrigeration.

Harvest & Post-Harvest Processing

Plants were harvested approximately halfway between soybean growth stage R6-R7, when beans filled out approximately 80-90% of the pod. Plants were harvested both by hand and using a simple gas mower. For hand-harvest, sheers were used to cut edamame plants at the base of the stalk. The gas mower contained rapidly moving blades that cut plants at the same position so as not to cause any additional yield loss. Once plants were harvested, they were immediately transferred to the Virginia Tech Agronomy Farm for post-harvest processing. Harvested edamame plants were processed into one of two different edamame end-products described in the literature: pods (Konovsky et al., 1994) and on-the-stalk bouquets (Miles et al., 2000). Pods were picked from branches by hand. On-the-stalk edamame bouquets were assembled by removing leaves from stalks to expose pods, and bunching stalks into bouquets weighing approximately 650 grams each. Immediately following post-harvest processing, pods and on-the-stalk bouquets were weighed in order to determine marketable yield. Yellow pods were not included in yield as per current market standards. Yield was calculated based on row length, spacing, and data on per row weight calculated following post-harvest processing and extrapolated to the acre.

Labor Costs

Number of workers and total hours worked were used to estimate cost of in-season labor required for spot spraying weeds. Time of harvest, number of workers, and quantity of plant rows harvested were recorded during harvest and used to estimate cost of harvest labor. Post-harvest labor cost was estimated in a similar fashion, where number of workers, quantity of rows processed, and time of post-harvest processing were all used for calculations. A weighted average was used to account for differences in number of workers.

Revenue Information

Pods were supplied in bulk to VT Dining Services, while bouquets were supplied to Oasis International Supermarket and Blacksburg Farmer’s Market, in exchange for information on total revenue and pricing quantity. Assumptions were made for marketing charge and variable refrigeration costs (such as electricity and maintenance). Pricing quantity was used to set unit price for revenue calculations since no unit price for edamame has been published in the literature hitherto. To determine overall revenue, marketable yield was multiplied by the unit price determined from pricing quantity data.

Results & Discussion

Pre-harvest input costs

Per acre costs for pre-plant fertilizer, pre-emergence herbicides, fuel were relatively low ($98.31). Cost of commercial edamame seed comprised the major portion of the pre-harvest variable costs. Commercial edamame seed is relatively expensive compared to conventional or grain type soybean seeds. Edamame seed can cost anywhere from $25-$50 per kg (Dhaliwal & Williams, 2020). In our economic analysis, we used the popular commercial edamame variety Chiba Green, which is supplied by both Johnny's Seeds and Wannamaker Seeds, to estimate seed cost. At the 87 lb per acre seeding rate for Chiba Green recommended on the Johnny's Seeds website, per acre seed cost was estimated to be roughly $1,473.

Food-grade soybean seeds may offer a cheaper alternative to commercial edamame seeds inputs. Food-grade soybeans, which are bred for tofu or soy milk production, contains many similar characteristics to edamame seeds such as relatively large bean size, non-GMO, and superior sensory and nutritive quality than conventional, grain-type soybeans. At an equivalent 100 lb quantity, food-grade soybean seed inputs can potentially save growers over $1,200 in seed cost compared to commercial edamame seeds. Moreover, consumers did not appear to tell the difference between the food-grade soybean varieties used in this study and typical commercial edamame varieties.

Growers may also be able to reduce cost of seed inputs by using reduced seeding rates. Seeding rates for commercial edamame seeds often follow recommendations for grain type soybeans, however a recent study by Dhaliwal & Williams (2020) suggest that these seeding rate recommendations are too high for edamame and can even decrease pod yield per plant. Using the optimal seeding rates recommended from their study can further reduce seed input costs while potentially increasing marketable yield.

Labor costs

While pre-harvest labor required for in-season spot spraying of weeds only accounted for $51.43, labor costs required for hand-harvest and post-harvest processing accounted for the most significant portion of overall production costs in this study. Labor hours required for harvesting edamame plants by hand was estimated to be approximately 886 hours, which equates to $13,290 assuming a $15 per hour wage rate. Post-harvest labor hours for pods and edamame bouquets was roughly 398 and 430, respectively, resulting in an additional $5,970 or $6,450 cost for growers depending on end-product.

Yield and revenue

Yield of edamame bouquets was calculated by extrapolating approximate per row yield to the acre. Total estimated yield was 11,592 lb of fresh edamame bouquets per acre and 8,023 lb of fresh edamame pods per acre. A unit price of $3.58/650g ($2.39/lb) was calculated for edamame bouquets using information on pricing quantity provided by distributors. Virginia Tech Dining Services sold pods at a retail price of $2.95. Assuming a 10-cent per unit cost of packaging and a 70% retail markup, we calculated the price returned to the farmer to be approximately $0.79 per 4oz ($3.16/lb). Total expenses for hand-harvested edamame production was $27,303 for bouquets and $26,208 for pods. Pre-harvest variable input costs accounted for $1,622.74, while fixed costs such as machinery and equipment, general overhead, and refrigeration accounted for another $2,688. Total cost of harvest, post-harvest handling, marketing, and refrigeration costs for bouquets was $22,993 and $22,525 for bouquets and pods, respectively. Using information on yield and production cost, we calculated a break-even price of 3.32/lb for pods, and $2.35/lb for bouquets.

Conclusions

Overall revenue was relatively similar for edamame bouquets ($27,666) and pods ($25,352). However, when compared to total production expenses, returns to land, capital, and management for bouquets ($362.31) and pods (-$1,143.49) suggest poor feasibility and potential profitability. This can be largely attributed to labor hours for harvest and post-harvest activities, which together accounted for over 70% of total production expenses of production in this study for both end-products. Ultimately, regardless of end-product supplied, the excessive labor costs associated with a hand-harvested edamame enterprise prevent it from being profitable. As such, in order to produce fresh edamame profitably, growers should consider mechanized harvest which offers improvements in cost- and time-efficiency during harvest that are needed to increase profit margins.

References

Aguado, V., Vitas, A. I., & García-Jalón, I. (2004). Characterization of Listeria monocytogenes and Listeria innocua from a vegetable processing plant by RAPD and REA. *International Journal of Food Microbiology*, *90*(3), 341–347. https://doi.org/10.1016/S0168-1605(03)00313-1

Bennett, C. (2016). Edamame Waits in the Wings. *AgWeb*. https://www.agweb.com/article/edamame-waits-in-the-wings-NAA-chris-bennett

Dhaliwal, D. S., & Williams, M. M. (2020). Economically Optimal Plant Density for Machine-harvested Edamame. *HortScience*, *55*(3), 368–373. https://doi.org/10.21273/HORTSCI14642-19

Hofstrand, D., Holz-Clause, M., & Wright, G. (2020, July). *When to Do and How to Use a Feasibility Study | Ag Decision Maker*. Iowa State University Extension & Outreach. https://www.extension.iastate.edu/agdm/wholefarm/html/c5-64.html

Ju, Y. H. (2016). To Soy or Not to Soy: Effects of Soybeans on Breast Cancer, Menopause and Heart Disease. *HNFE-339P*. https://www.pubs.ext.vt.edu/content/dam/pubs_ext_vt_edu/HNFE/HNFE-339/HNFE-339-PDF.pdf

Konovsky, J., Lumpkin, T. A., & McClary, D. (1994). Edamame: The vegetable soybean. In *Understanding the Japanese food and agrimarket: A multifaceted opportunity* (pp. 173–181). Haworth Press. https://coolbean.info/pdf/soybean_research/library/forage_and_food_production/EDAMAME_%20THE%20VEGETABLE%20SOYBEAN.pdf

Miles, C. A., Lumpkin, T. A., & Zenz, L. (2000). Edamame. *PNW0525*, 1–8.

Roseboro, K. (2012, April 4). *Edamame offers good non-GMO opportunities to US farmers*. The Organic & Non-Gmo Report. https://non-gmoreport.com/articles/april2012/edamame-non-gmo-us-farmers.php

Simonne, A. H., Smith, M., Weaver, D. B., Vail, T., Barnes, S., & Wei, C. I. (2000). Retention and Changes of Soy Isoflavones and Carotenoids in Immature Soybean Seeds (Edamame) during Processing. *Journal of Agricultural and Food Chemistry*, *48*(12), 6061–6069. https://doi.org/10.1021/jf000247f

Taku, K., Umegaki, K., Sato, Y., Taki, Y., Endoh, K., & Watanabe, S. (2007). Soy isoflavones lower serum total and LDL cholesterol in humans: A meta-analysis of 11 randomized controlled trials. *The American Journal of Clinical Nutrition*, *85*(4), 1148–1156. https://doi.org/10.1093/ajcn/85.4.1148

Chapter 3: Investigating consumer demand and willingness-to-pay (WTP) for fresh, local, certified organic, & on-the-stalk edamame

Nick Lord[1], Clinton Neill[2,3], Bo Zhang[1]

[1] School of Plant & Environmental Sciences, Virginia Tech

[2]Department of Agricultural & Applied Economics, Virginia Tech

[3]Dyson School of Applied Economics, Cornell University

This chapter was submitted to the Agricultural & Resource Economics Review on July 9, 2019

Abstract

Introduction of locally-adapted, commercially-viable edamame varieties can allow it to be marketed as fresh, local, organic, or on-the-stalk. Here, we utilized contingent valuation (CV) methodology to estimate mean WTP for these external attributes in relation to a vector of explanatory variables. Results showed 94, 88, and 33-cent premiums for fresh, local, and organic edamame. Pro-environmental attitudes drove WTP for all three of these attributes, while shopping locale significantly increased mean WTP for fresh and organic attributes specifically. A 40-cent price discount was observed for the "on-the-stalk" attribute, suggesting that convenience also plays an important role in marketing edamame.

Introduction

As Americans become more health- and/or environmentally conscious, there has been a growing desire among consumers to reduce meat consumption which has led to noticeable growth in demand for alternative, plant-based sources of protein (Bashi et al., 2019). In addition, disproportionately large and rapid growth of Asian populations in the U.S. promise to create new market opportunities for specialty Asian produce (Govindasamy et al., 2010). Vegetable soybean, more commonly referred to as edamame (pronounced "eh-duh-MAH-may"), has quickly emerged as a prime candidate to capitalize on both of these domestic trends, appearing more and more frequently in salad bars and sushi restaurants nationwide.

Edamame has a long history of consumption in East Asia (Aoyagi, 2009) where the industry has become well-established. It was not until the turn of the 21st century that it began to be imported to U.S. markets. In order to meet domestic demand, roughly 25,000 to 30,000 tons of edamame are consumed annually (Associated Press, 2013), which is predominantly met through frozen imports sold to consumers year-round in grocery stores. In the U.S., edamame is most commonly supplied as pods, where consumers suck the beans out of the pod. Edamame can also be supplied as shelled beans that have already been removed from the pods.

Soybean proteins contain all essential amino acids that are required for human growth and development, which has earned it designation as a high-quality protein source (Michelfelder, 2009). In addition, several studies have suggested that soy consumption can reduce levels of low-density lipoprotein (LDL) cholesterol, or "bad" cholesterol, in the body which can build up in arteries and increase risk of cardiovascular disease (Hasler, 2002; Taku et al., 2007). As legumes, soybeans also exploit biological nitrogen fixation which reduces the need for synthetic nitrogen application and is beneficial for soil health (Beyan et al., 2018). This unique combination of nutritional and environmentally friendly characteristics has allowed edamame to carve out a distinguished position in the domestic marketplace for vegetables.

Edamame's growing popularity has sparked interest in bolstering U.S. edamame production to reduce reliance on imports and allow domestic growers to capitalize on edamame's premium market position. In the coming years, plant breeders are set to release seed inputs adapted to U.S. production regions with higher yield potential, enhanced sensory quality, and increased suitability for mechanized harvest in hopes of catalyzing domestic production (Lord, Zhang, et al., 2019). The provision of locally-adapted seed inputs can lead to many new edamame product options in the domestic marketplace. With more edamame grown on U.S. soil, edamame producers and distributors can seek to access premiums associated with quality-differentiated external attributes such as fresh supplied, locally-grown, and USDA certified organic, or by marketing it as an alternative end-product. As such, it is essential to understand what marketing advantage, if any, that these attributes hold over the frozen, imported, and non-GMO product that are currently available to consumers today.

Materials & Methods

Contingent valuation (CV) is a flexible survey technique used to estimate WTP for non-market goods and services (Lopez-Feldman, 2013). It has been frequently exploited in the literature to forecast success or demand for various agricultural products, ecolabels, and marketing trends (Carpio & Isengildina-Massa, 2009; Haghiri et al., 2009; Neill & Williams, 2016; Owusu & Owusu Anifori, 2013). CV methodology can be particularly useful in exploring price premiums and WTP for products or product attributes that have yet to appear in the marketplace. This is achieved by presenting participants with a hypothetical scenario through which to observe their purchasing decisions when both the "status-quo" and newly proposed product or attribute of interest are present.

During the month of October 2018, consumer-intercept CV surveys were distributed at three primary locations in Blacksburg, VA – Oasis International Supermarket, the Blacksburg Farmer's Market, and the Virginia Tech Squires Student Center. Surveys consisted of 20 questions each. The first 16 questions pertained to demographic information and characteristics of participants relating to purchasing habits, environmental attitudes, dietary habits, and familiarity with edamame (Figure 1). Location of survey was included as a categorical variable in order to also observe the effect of market venue. Responses to these questions were used to identify important explanatory variables of mean WTP. The remaining four questions were used to elicit WTP using dichotomous choice (DC) questions. Dichotomous choice (DC) questions are the most adequate, reliable, and heavily utilized CV elicitation technique used in the literature given their increased statistical efficiency and reduced response bias over other CV elicitation methods such as open-ended questions and payment cards (Lopez-Feldman, 2012; Venkatachalam, 2004). As opposed to DC questions, single-bounded questions consisting of a single "take-it-or-leave-it" question simply ask the individual whether they would accept ($y_i = 1$) or reject ($y_i = 0$) a bid for the product/attribute of interest. To increase the statistical efficiency of the single bound estimation, Hanemann et al. (1991) later proposed the addition of a follow-up question dependent on the response to the initial question. Responses to initial and follow-up questions can be captured by dichotomous variables y_i^1 and y_i^2, respectively. If the individual selects "no" to the initial bid ($y_i^1 = 0$), a lower bid is offered whereas if the individual selects "yes" ($y_i^1 = 1$), then a higher bid is offered. Despite its increased statistical efficiency, discrepancies observed between the initial and follow-up questions of DB format have caused concern over its consistency and reproducibility. One and one-half bound (OOHB) question format was thus developed to address this response bias while maintaining statistical efficiency (Cooper et al., 2002). In OOHB-DC questions, respondents are only asked to answer a follow-up if they select "yes" for the bid listed in the initial question. Responses to OOHB-DC questions can therefore result in one of following three scenarios: no-no, yes-no, or yes-yes.

An example of the OOHB-DC questions can be seen in Figure 2**.** In these questions, respondents must choose between purchasing the "status-quo" option which is always listed at the "fair" price (in this case $3.50 per 10 ounces of pods) or the new or alternatively marketed product of the same quantity when it is offered at a premium. In order to get information on multiple premiums, three survey versions were randomly administered, each only differing in the extent of the premium for the alternative product. One of the three versions also offered the alternative product at a lower price than the status quo to see if participants actually discounted the alternative product. In total, premiums of 0.25, 0.50, 1.00, and a discount of 0.25 were observed.

An interval data model was used to estimate individual WTP as a function of explanatory variables and error. The probability that the WTP falls between a specific minimum and maximum bound can be estimated using a log-likelihood function. To understand the econometric estimation behind the probabilities for each of the aforementioned scenarios, let us consider the following set of equations, taken from Neill & Williams (2016):

(1) $$WTP_i(z_i, u_i) = z_i\beta + u_i$$

(2) $$\Pr(y_i^1 = 1, y_i^2 = 0 \mid z_i) = \Pr(s, n)$$

In equation 1, we assume WTP is a function of an individual's characteristics (demographics, environmental attitudes, diet, etc.) plus error, where z_i represents a vector of explanatory variables, β represents a vector of coefficients for the explanatory variables, and u_i represents an error term. Equation 2 represents a simplified notation for the probability of an individual answering "yes" to the initial question, and "no" to the follow-up, dependent on the vector of explanatory variables. Under the assumptions of Equation 1 and that our data is normally distributed, we can obtain:

(3) $$\Pr(s, n) = \Pr(t^1 \leq WTP < t^2)$$

(4) $$= \Pr(t^1 \leq z_i'\beta + u_i < t^2)$$

(5) $$= \Pr\left(\frac{t^1 - z_i'\beta}{\sigma} \leq \frac{u_i}{\sigma} < \frac{t^2 - z_i'\beta}{\sigma}\right)$$

(6) $$= \varphi\left(\frac{t^2 - z_i'\beta}{\sigma}\right) - \varphi\left(\frac{t^1 - z_i'\beta}{\sigma}\right)$$

where t^1 represents the suggested bid for the initial question and t^2 represents the suggested bid for the follow-up question. Given that Equation 6 follows $\Pr(a \leq X < b) = F(b) - F(a)$, using symmetry of the normal distribution, we can rearrange once more to get:

$$\Pr(s,n) = \varphi\left(z_i' \frac{\beta}{\sigma} - \frac{t^1}{\sigma}\right) - \varphi\left(z_i' \frac{\beta}{\sigma} - \frac{t^2}{\sigma}\right) \tag{7}$$

The other two conditions can be similarly derived by replacing $t^1 \leq WTP < t^2$ with $t^2 \leq WTP < \infty$ for the yes-yes scenario, and $0 \leq WTP < t^1$ for the no scenario. Taken together, these scenarios comprise the censored likelihood function shown below:

$$\sum_{i=1}^{N} \left[d_i^{sn} ln\left(\varphi\left(z_i' \frac{\beta}{\sigma} - \frac{t^1}{\sigma}\right) - \varphi\left(z_i' \frac{\beta}{\sigma} - \frac{t^2}{\sigma}\right)\right) + d_i^{ss} \ln\left(\varphi\left(z_i' \frac{\beta}{\sigma} - \frac{t^2}{\sigma}\right)\right) + d_i^{nn} ln\left(1 - \varphi\left(z_i' \frac{\beta}{\sigma} - \frac{t^2}{\sigma}\right)\right)\right] \tag{8}$$

where d_i^{sn}, d_i^{ss}, and d_i^{nn} represent indicator variables for yes-no, yes-yes, and no-no, respectively. Depending on how the individual responds to the DC questions, the non-relevant indicator variables will take the value of zero, allowing only the relevant case to contribute to the likelihood function at any given time.

Censored regression analysis to estimate mean WTP for each of the four external attributes was conducted in R using the DCchoice package. The DCchoice package contains an oohbchoice function which was specifically developed to execute maximum likelihood estimation on OOHB-DC data based on a number of required and optional arguments that the researcher specifies, such as the formula, distribution, and omission of missing data (Nakatani et al., 2020). The output is similar to that which is generated from the LIFEREG procedure on SAS v. 9.3, where parameter estimates can be directly interpreted as changes in the marginal WTP (Neill & Williams, 2016). Survey locations were coded in the model as categorical variables, where the locations not relevant took the value of zero in order to prevent singularity in the model.

A total of 222 surveys were collected. For the OOHB-DC questions, only complete sets of responses (yes-no, yes-yes, or no-no) were used for the interval regression. A complete table of summary statistics can be seen in Table 1. Approximately 69% of the respondents were between the age of 18-30 and 34% of respondents reported household income below $20,000, which can be attributed to the survey being distributed in a small college-town where students are more prevalent in the community. The vast majority of respondents had received at least some level of secondary schooling, with 42% having taken some school, 25% having received a Bachelor's degree, and 26% having received some sort of advanced or post-graduate degree. About 48% of the population were female. The majority of respondents were either Caucasian (57%) or Asian (33%). Approximately 44% of respondents indicated that they had a desire to reduce their meat consumption, of which 78% indicated health as a reason why and 56% indicated environment as a reason why. In regards to familiarity with edamame, approximately 55% indicated that they were mostly (20%) to extremely familiar (35%) while another 42% indicated that they consume edamame at least once per week.

Results & Discussion

WTP estimates derived from the interval regression for all four attributes can be seen in Table 2. The interval regression estimated a mean WTP of approximately $4.44 per 10 ounces for the fresh attribute. In other words, survey respondents indicated that they were willing to pay up to 93 cents more for edamame pods available fresh as opposed to frozen on average. The variables *Female* and *Likelihood to shop local*, significant at the 1% and 5% levels (respectively), both exhibited a negative effect on mean WTP for the fresh attribute. If the survey respondent was female, the average discount associated with fresh edamame is about 25 cents. This observation may be related to public uncertainty surrounding the safety of phytohormones in soy products for women's health (Bar-El & Reifen, 2010; Cederroth et al., 2012; Duffy et al., 2007; White et al., 2000), which are present in lower levels in frozen edamame products (Simonne et al., 2000). Meanwhile, the more likely a respondent was to shop local, the less they were willing to pay for a fresh edamame product. In other words, for a one-point increase in likelihood to shop local respondents indicated via likert scale, consumers discount the fresh edamame product by approximately 25 cents.

A considerable price discount was observed for edamame supplied as an the on-the-stalk product. When given the choice between edamame supplied as pods versus on-the-stalk, survey respondents were not willing to pay any more than $3.10 per 10 ounces. The lone variable that was statistically significant was familiarity with edamame. One may think that the more familiar an individual is with edamame, the more likely they are to value the traditional nature of an on-the-stalk product and thus be willing to pay more. However, our results suggested that greater familiarity with edamame further drove down how much a respondent was willing to pay for an on-the-stalk product. Specifically, for each unit increase in familiarity with edamame a respondent indicated on the likert scale, average WTP dropped by 20 cents.

For the certified organic attribute, only a small premium was observed; survey respondents indicated a maximum premium of 33 cents more for an edamame product labeled as USDA certified organic if an equivalent non-GMO product was also available. That is to say, beyond the non-GMO guarantee of their edamame product, respondents were not willing to pay appreciably more for the other provisions encompassed under the USDA certified label.

The mean WTP estimate for the local attribute was similar to the estimate for the fresh attribute. When given a choice between local and non-local edamame pods, survey respondents indicated that they would be willing to pay an average of up to 88 cents more for the local option. It should be noted that no formal definition of "local" regarding vicinity of production, state, or region, was provided as part of the study. Nevertheless, a high mean WTP was observed. Moreover, with the exception of the *Likelihood to purchase eco-labeled products* variable, no other variables appeared to have a significant affect on mean WTP for the local attribute suggesting relatively broad interest in a local edamame product among our survey sample.

The variable *Likelihood to purchase eco-labeled products* was used in this study to gauge pro-environmental attitudes of the survey respondents. By including it as a parameter, our intention was to determine if pro-environmental attitudes had an impact on consumer WTP for the various attributes. In our results, we observed that *Likelihood to purchase eco-labeled products* consistently appeared to be a significant driver of mean WTP for three out of the four attributes studied. For each unit increase in pro-environmental attitudes indicated on the likert scale, mean WTP increased by 32 cents for fresh edamame, 27 cents for local edamame, and 33 cents for

USDA certified organic edamame. The *Likelihood to purchase eco-labeled products* did not show significance for the on-the-stalk attribute, which may be due to the fact that the on-the-stalk attribute is more related to physical appearance and as a result, does not hold as many environmental implications as the other three attributes. These results follow a stream of literature suggesting that pro-environmental attitudes and beliefs continue to play an increasingly important role in consumer WTP for products (Lusk et al., 2014; Neill & Williams, 2016; Rahman & Reynolds, 2017; Shin et al., 2017; Umberger et al., 2009).

Survey location was a notable predictor of mean WTP observed for the fresh and organic attributes in this study. In total, all surveys collected fell into one of four categories: those collected at the farmer's market, the local supermarket, the student center, and those collected from miscellaneous individuals around campus (henceforth referred to as "Misc group"). We included the two purchasing-locations (farmer's market and supermarket) and one non-purchasing location (student center) in our analysis, and interpreted the results as they related to the Misc group. Results showed that if a respondent completed the survey at the farmer's market, mean WTP for the fresh attribute was 89 cents higher as compared to Misc group. Meanwhile, mean WTP for surveys completed at the international supermarket were $1.29 higher for the fresh attribute and 64 cents higher for the organic attribute. For all attributes studied, mean WTP for surveys collected at the student center – where no produce is sold and no sustainability claims are therefore made – failed to show significance. Here, we can clearly see that whether or not survey were collected from a purchasing locations largely influenced the premiums observed for both the fresh and organic attributes. Furthermore, the type of purchasing location (farmer's market versus local supermarket) appeared to drastically impact the extent of the premiums observed for each of these attributes.

Conclusions

With the hope of increased domestic production of edamame slowly becoming a reality, it is important to understand what factors can favor successful marketing efforts in the U.S. In this study, we used CV methodology to estimate mean WTP for edamame marketed as fresh, local, certified organic, or as an on-the-stalk end-product and then related these estimates to information on demographic information, dietary habits, and personal beliefs. Our results showed that fresh and local edamame held a significant marketing advantage over frozen products and non-local products, respectively. We also observed that despite its myriad of environmental provisions, edamame labeled as USDA certified organic may offer only limited marketing benefit to distributors if consumers already know that the edamame product they are purchasing is non-GMO. In regards to end-products, the convenience of pre-stripped edamame pods appears to take priority over the naturalness and freshness offered by an on-the-stalk or whole plant edamame product.

Several factors are important for understanding the domestic potential of alternatively marketed edamame. Firstly, WTP for fresh edamame appears to be particularly influenced by demographic factors, especially gender. Meanwhile, personal beliefs such as pro-environmental attitudes appear to more consistently drive premiums for fresh, local, and organic edamame. This study also provides major evidence for the sorting effect that market venues have on consumer valuation of external attributes, especially for fresh and organic edamame. These findings also highlight the importance of accounting for potential bias that can occur when collecting consumer intercept surveys from purchasing and non-purchasing locations.

The results from this study present important preliminary findings regarding the potential of alternatively marketed edamame in the domestic marketplace. Overall, these results support the continued growth of the domestic edamame industry. As with many WTP studies, our assertions are based on a single geographic location using a limited sample population which may not be generally true in other states or regions across the U.S. Therefore, future research should aim to substantiate these findings in other locations using different populations. Given the hypothetical bias inherent in CVM, non-hypothetical valuation methods may also serve to improve upon the results of this study by more accurately modeling consumer valuation of edamame based on real consequences. As edamame variety development continues, there may be additional opportunities to study consumer WTP for edamame on the basis of sensory characteristics such as quality, flavor profile, and appearance as well.

References

Associated Press. (2013, March 29). *Edamame: Is the future of American soy farmers' profits in edible soy beans for people?* CBS News. https://www.cbsnews.com/news/edamame-is-the-future-of-american-soy-farmers-profits-in-edible-soy-beans-for-people/

Bar-El, D. S., & Reifen, R. (2010). Soy as an Endocrine Disruptor: Cause for Caution? *Journal of Pediatric Endocrinology and Metabolism*, *23*(9), 855–861. https://doi.org/10.1515/jpem.2010.138

Bashi, Z., McCullough, R., Ong, L., & Ramirez, M. (2019). *The market for alternative protein: Pea protein, cultured meat, and more | McKinsey*. McKinsey & Company. https://www.mckinsey.com/industries/agriculture/our-insights/alternative-proteins-the-race-for-market-share-is-on

Beyan, S. M., Wolde-meskel, E., & Dakora, F. D. (2018). An assessment of plant growth and N2 fixation in soybean genotypes grown in uninoculated soils collected from different locations in Ethiopia. *Symbiosis (Philadelphia, Pa.)*, *75*(3), 189–203. https://doi.org/10.1007/s13199-018-0540-9

Carpio, C. E., & Isengildina-Massa, O. (2009). Consumer willingness to pay for locally grown products: The case of South Carolina. *Agribusiness*, *25*(3), 412–426. https://doi.org/10.1002/agr.20210

Cederroth, C. R., Zimmermann, C., & Nef, S. (2012). Soy, phytoestrogens and their impact on reproductive health. *Molecular and Cellular Endocrinology*, *355*(2), 192–200.

Cooper, J. C., Hanemann, M., & Signorello, G. (2002). One-and-One-Half-Bound Dichotomous Choice Contingent Valuation. *The Review of Economics and Statistics*, *84*(4), 742–750. JSTOR.

Duffy, C., Perez, K., & Partridge, A. (2007). Implications of Phytoestrogen Intake for Breast Cancer. *CA: A Cancer Journal for Clinicians*, *57*(5), 260–277. https://doi.org/10.3322/CA.57.5.260

Govindasamy, R., Van Vranken, R., Sciarappa, W., Ayeni, A., Puduri, V. S., Pappas, K., Simon, J. E., Mangan, F., Lamberts, M., & McAvoy, G. (2010). Ethnic crop opportunities for growers on the east coast: A demand assessment. *J. Ext*, *48*, 1–9.

Haghiri, M., Hobbs, J. E., & McNamara, M. L. (2009). Assessing Consumer Preferences for Organically Grown Fresh Fruit and Vegetables in Eastern New Brunswick. *International Food and Agribusiness Management Review*, *12*(4), 1–20. https://doi.org/10.22004/ag.econ.92556

Hanemann, M., Loomis, J., & Kanninen, B. (1991). Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics*, *73*(4), 1255–1263. https://doi.org/10.2307/1242453

Lopez-Feldman, A. (2013). Introducción a la valoración contingente utilizando Stata. In *Aplicaciones en Economía y Ciencias Sociales con Stata*. Stata Press; Munich Personal RePEc Archive. https://mpra.ub.uni-muenchen.de/41018/

Lord, N., Zhang, B., Kuhar, T., Carneiro, R., Sutton, K., & Yu, D. (2019). USDA Edamame Project. *SPES-104NP*. https://www.pubs.ext.vt.edu/content/dam/pubs_ext_vt_edu/spes/SPES-104/SPES-104.pdf

Lusk, J. L., Schroeder, T. C., & Tonsor, G. T. (2014). Distinguishing beliefs from preferences in food choice. *European Review of Agricultural Economics*, *41*(4), 627–655. https://doi.org/10.1093/erae/jbt035

Nakatani, T., Aizaki, H., & Sato, K. (2020). Oohbchoice. *Package "DCchoice,"* 23–26.

Neill, C. L., & Williams, R. B. (2016). CONSUMER PREFERENCE FOR ALTERNATIVE MILK PACKAGING: THE CASE OF AN INFERRED ENVIRONMENTAL ATTRIBUTE. *Journal of Agricultural and Applied Economics*, *48*(3), 241–256. https://doi.org/10.1017/aae.2016.17

Owusu, V., & Owusu Anifori, M. (2013). Consumer Willingness to Pay a Premium for Organic Fruit and Vegetable in Ghana. *International Food and Agribusiness Management Review*, *16*(1), 67–86. https://doi.org/ISSN #: 1559-2448

Rahman, I., & Reynolds, D. (2017). Organic Wine: The Influence of Biospheric, Altruistic, and Egoistic Values on Purchase Intention, Willingness to Pay More, and Willingness to Sacrifice. *International Journal of Hospitality Beverage Management*, *1*(1), 1–47. https://doi.org/10.34051/j/2019.1

Shin, Y. H., Moon, H., Jung, S. E., & Severt, K. (2017). The effect of environmental values and attitudes on consumer willingness to pay more for organic menus: A value-attitude-behavior approach. *Journal of Hospitality and Tourism Management*, *33*, 113–121. https://doi.org/10.1016/j.jhtm.2017.10.010

Shurtleff, W., & Aoyagi, A. (2009). *History of Edamame, Green Vegetable Soybeans, and Vegetable-Type Soybeans (1275-2009): Extensively Annotated Bibliography and Sourcebook*. Soyinfo Center.

Simonne, A. H., Smith, M., Weaver, D. B., Vail, T., Barnes, S., & Wei, C. I. (2000). Retention and Changes of Soy Isoflavones and Carotenoids in Immature Soybean Seeds (Edamame) during Processing. *Journal of Agricultural and Food Chemistry*, *48*(12), 6061–6069. https://doi.org/10.1021/jf000247f

Taku, K., Umegaki, K., Sato, Y., Taki, Y., Endoh, K., & Watanabe, S. (2007). Soy isoflavones lower serum total and LDL cholesterol in humans: A meta-analysis of 11 randomized controlled trials.

The American Journal of Clinical Nutrition, *85*(4), 1148–1156. https://doi.org/10.1093/ajcn/85.4.1148

Umberger, W. J., McFadden, D. D. T., & Smith, A. R. (2009). Does altruism play a role in determining U.S. consumer preferences and willingness to pay for natural and regionally produced beef? *Agribusiness*, *25*(2), 268–285. https://doi.org/10.1002/agr.20194

Venkatachalam, L. (2004). The contingent valuation method: A review. *Environmental Impact Assessment Review*, *24*(1), 89–124. https://doi.org/10.1016/S0195-9255(03)00138-0

White, L. R., Petrovitch, H., Ross, G. W., Masaki, K., Hardman, J., Nelson, J., Davis, D., & Markesbery, W. (2000). Brain Aging and Midlife Tofu Consumption. *Journal of the American College of Nutrition*, *19*(2), 242–255. https://doi.org/10.1080/07315724.2000.10718923

Tables & Figures

Table 1: Summary statistics for survey respondents.

Variable	Description	Percentage of Occurrence	Mean	Standard Deviation
Age	Age of participant:			
	1 = Under 18	0.90%	2.6347	1.213
	2 = 18-30	67.57%		
	3 = 30 - 45	13.06%		
	4 = 46 - 50	4.50%		
	5 = 51 - 60	6.31%		
	6 = Older than 60	5.86%		
Gender	Dummy variable;			
	0 = Male	50.45%	0.4886	0.501
	1 = Female	48.20%		
Ethnicity	Categorical variable:			
	1 = Caucasian	56.68%		
	2 = African American	2.76%		
	3 = Asian	33.18%		
	4 = Hispanic	4.15%		
	5 = Other	3.23%		
Main shopper for household	Dummy variable;			
	0 = No	36.92%	0.6308	0.4837
	1 = Yes	63.08%		
Household size	Number of people currently living in household:			
	1 = 1	32.57%	2.6284	1.4669
	2 = 2	20.18%		
	3 = 3	13.76%		
	4 = 4	18.81%		
	5 = More than 4	14.68%		
Home food consumption	Number of meals consumed at home per week:			
	1 = 1 to 5	16.89%	2.5707	0.9992
	2 = 6 to 10	29.68%		

	3 = 11 to 15	32.88%		
	4 = 16 or more	20.55%		
Education	Highest level of education completed:			
	1 = Some school	1.39%	3.6497	1.0031
	2 = High school diploma	6.48%		
	3 = Some college	41.67%		
	4 = Bachelor's degree	25.00%		
	5 = Advanced degree	25.46%		
Household Income	Household income levels:			
	1 = Less than $20,000	34.91%	3.1173	2.0375
	2 = $20,000 - $35,000	16.04%		
	3 = $35,001 - $50,000	9.43%		
	4 = $50,001 - $70,000	6.13%		
	5 = $70,001 - $100,000	10.85%		
	6 = More than $100,000	22.64%		
Meat consumption	How many days per week meat is consumed:			
	0 = 0	10.60%	4.7981	4.4932
	1 = 1	2.76%		
	2 = 2	8.29%		
	3 = 3	10.14%		
	4 = 4	13.36%		
	5 = 5	10.14%		
	6 = 6	12.90%		
	7= 7	31.80%		
Desire to reduce meat consumption	Dummy variable;			
	0 = No	56.19%	0.238	0.4364
	1 = Yes	43.81%		
Reason for desire to reduce meat consumption	Health	78.26%		
	Religion	3.26%		
	Environment	56.52%		
	Other	8.70%		
Fully/Partially Vegan or Vegetarian	Dummy Variable;			

	0 = No	79.53%	0.2047	0.4044
	1 = Yes	20.47%		
Environmental value	How often the consumer purchases eco-labeled products (Likert scale 1-5):			
	1 = Never	9.30%	3.1116	1.1983
	2 = Rarely	22.79%		
	3 = Some times	31.16%		
	4 = Somewhat frequently	20.93%		
	5 = Always	15.81%		
Edamame familiarity	Familiarity with edamame (Likert scale 1-5):			
	1 = Not familiar at all	14.02%	3.4672	1.4716
	2 = Slightly familiar	17.29%		
	3 = Somewhat familiar	12.62%		
	4 = Mostly familiar	20.09%		
	5 = Extremely familiar	35.98%		
Edamame consumption	Frequency of edamame consumption (Likert scale 1-5):			
	1 = Never	27.44%	2.8372	1.2367
	2 = Once a month	3.26%		
	3 = Twice a month	27.44%		
	4 = Once a week	41.86%		
	5 = Multiple times per week	0.00%		
Buying local	Likeliness to purchase local (Likert scale 1-5):			
	1 = Never	11.57%	2.9167	1.1743
	2 = Rarely	27.31%		
	3 = Somewhat often	30.09%		
	4 = Very often	19.91%		
	5 = Always	11.11%		

Table 2. WTP estimates for edamame marketed as fresh, local, organic, and on-the-stalk.

Parameter	Fresh	Local	Organic	On-the-stalk
Mean WTP (per 10 oz.)	4.44	4.37	3.83	3.10
Intercept	3.640**	5.107**	4.851**	4.645**
Location – Farmer's Market	0.895**	0.456	-0.041	-0.109
Location – Int'l Supermarket	1.291**	0.133	0.639*	0.170
Location – Student Center	0.314	-0.188	-0.283	-0.366
Age	0.008	0.192	0.094	-0.029
Female	-0.678**	0.144	-0.039	-0.370
Ethnicity	0.380	-0.094	-0.177	-0.080
Shopper	-0.399	-0.047	0.300	0.258
Household size	0.737	-0.023	0.041	0.081
Meals consumed at home	0.196	-0.107	0.131	0.050
Education	0.120	0.064	-0.016	-0.077
Household Income	-0.109	0.031	-0.019	-0.029
Freq. of meat consumption	0.007	-0.002	0.017	-0.113
Desire to decr. meat consump.	-0.040	0.159	0.173	0.044
Vegan/Vegetarian	-0.053	0.152	-0.358	-0.755
Purchase of eco-labeled prod.	0.315**	0.268*	0.331**	0.092
Familiarity	-0.020	-0.043	-0.096	-0.197*
Freq. of edamame consump.	-0.071	-0.022	0.101	0.131
Likelihood to shop local	-0.245*	0.118	0.088	-0.056
BID	-0.955**	-1.412**	-1.738**	-1.091**

Note: Asterisks '**', and '*', indicate statistical significance at 1% and 5% levels, respectively.

Figure 1. Example of demographic questions on first page of survey instrument.

1. **What is your current age?** Under 18 18 to 30 30 to 45 46 to 50 Older than 60
2. **What gender do you identify with?** Male Female Other
3. **What is your ethnicity?**

   Caucasian African American Asian Hispanic Other
4. **Are you the main shopper for your household?** Yes No
5. **What is your current household family size, including yourself?**

   1 2 3 4 More than 4
6. **How many meals do you eat at home per week?**

   1 to 5 6 to 10 11 to 15 16 or more
7. **What is your current level of education?**

   Some school High school diploma Some college Bachelor’s degree Advanced degree
8. **What is your household income?**

Less than 20,000 20,000-35,000 35,001-50,000 50,001-70,000 70,001-100,000 100,000+

9. **How many days a week do you consume meat (including fish)?**

   0 1 2 3 4 5 6 7
10. **Do you have any desire to reduce your meat consumption?** Yes No

10a. If you answered “yes” to question 10, please check which reasons you desire to reduce meat consumption: ___ health ___ religious reasons ___reduce environmental impact ___other

11. **Are you full/partially vegan or vegetarian?** Yes No

12. **On a scale of 1 to 5 with 1 being never and 5 being very often/always, how often do you purchase food that is labeled organic, free range/cage free, Rainforest Alliance, Wild Caught, natural, or non-GMO?** 1 2 3 4 5

Figure 2. Example of one and one-half bound dichotomous choice (OOHB-DC) questions found on survey instrument.

17. Under the following prices, which edamame product would you choose?

Non-GMO edamame pods at $3.50 (10 oz. bag) USDA organic edamame pods at $3.75 (10 oz. bag)

17a. If you answered USDA organic edamame pods, choose which edamame product you would rather purchase:

Non-GMO edamame pods at $3.50 (10 oz. bag)

USDA organic edamame pods at $4.00 (10 oz. bag)

17b. If you answered Non-GMO edamame pods, would you be willing to pay anything above $3.50 for USDA Organic edamame pods? Yes No

Chapter 4: Development of near-infrared reflectance spectroscopy (NIRS) calibrations for soybean seed sugars using a Perten DA7250 analyzer

Nick Lord[1], Chao Shang[1], Luciana Rosso[1], Bo Zhang[1]

[1]School of Plant & Environmental Sciences, Virginia Tech

Accepted to Crop Science for publication on September 29, 2020:
https://doi.org/10.1002/csc2.20358

Abstract

There is growing interest in developing calibrations for soybean seed sugars on near-infrared reflectance spectroscopy (NIRS) instruments in order to increase the efficiency of sugar profiling. In this study, a set of 253 samples from Virginia Tech's soybean germplasm with wide range of sugar content were used to create prediction models for sucrose, raffinose, and stachyose in ground soybean seed on the Perten DA7250 NIRS instrument. Following acquisition of spectral data, seed sugars were extracted from ground samples and analyzed using high-performance liquid chromatography (HPLC) to obtain reference data. Spectral and HPLC data were modeled using PLS regression on CAMO Unscrambler® X software, and internally cross-validated using the same software. Resulting calibrations showed high quantitative accuracy with R^2_{cal} = 0.901, R^2_{cv} = 0.869, RMSEC = 0.516, and RMSECV of 0.596 for sucrose, R^2_{cal} = 0.911, R^2_{cv} = 0.891, RMSEC = 0.361, and RMSECV of 0.405 for stachyose. These calibrations appear suitable for use in breeding operations. Meanwhile, performance of the raffinose calibration remained poor with R^2_{cal} = 0.568, R^2_{cv} = 0.476, RMSEC = 0.129, and RMSECV of 0.143. Alternative methods for more accurate and rapid quantification of raffinose concentration in soybean seed should be investigated.

Keywords: Near-Infrared Reflectance Spectroscopy, high-performance liquid chromatography, calibration development, soybean seed sugars, soluble carbohydrates, raffinose family oligosaccharides

Introduction

For its relatively high protein and oil content, soybean (*Glycine max*) has become one of the most heavily utilized commodities in the world. According to the USDA Coexistence Fact Sheet (https://www.usda.gov/sites/default/files/documents/coexistence-soybeans-factsheet.pdf), about 70% of domestically grown soybeans are devoted to production of animal feed. As soy's distinguished nutritional profile becomes more well known to consumers, an increasing proportion of soybeans are also used for production of soyfoods such as natto, tofu, sprouts, and soymilk (Mayta et al., 2014). To support these industries, improving the quality of both feed- and food-grade soybeans have become a major goal of many soybean breeding programs.

Soluble carbohydrates in soybeans seeds represent one of the most significant breeding targets for feed- and food-grade cultivar development. Di-, tri-, and tetrasaccharide sugars, better known as sucrose, raffinose, and stachyose (respectively), are the most abundant soluble carbohydrates in soybean seeds and thus hold the greatest implications for soybean quality (Jo, 2016). Sucrose is typically desired in high concentrations (> 6%) in soybean seeds to increase both sweetness in soyfood products and metabolizable energy in soybean meal. Raffinose and stachyose are classified as Raffinose Family Oligosaccharides (RFO). Unlike sucrose, RFO's contain alpha-(1,6) linkages that are only cleavable by alpha-(1,6) galactosidase which humans and monogastric animals lack (Švejstil et al., 2015). This results in decreased feed efficiency in monogastric animals – the primary beneficiaries of soybean meal – and increased flatulence and discomfort in humans (Rackis, 1975; Suarez et al., 1999). Not surprisingly, breeders have long selected for low-RFO content as a result. This trend may change, however, as recent evidence suggesting potential for soybean RFO's as a prebiotic substance (Wongputtisin et al., 2015) may encourage some breeders to begin selecting for high-RFO's instead.

Given the particularly strong influence of environmental factors on soybean soluble carbohydrates (Jo et al., 2018; Kumar et al., 2010; Maupin et al., 2011), it is crucial to quantify sugar content in breeding lines on a yearly basis. Currently, sugar quantification in soybean is primarily performed using laboratory extraction followed by High-Performance Liquid Chromatography (HPLC) (Black & Bagley, 1978). Despite providing highly accurate and reliable results, HPLC commands a relatively high cost and time requirement per sample compared to other methods (Kumar et al., 2010; Stitt et al., 1989; Teixeira et al., 2012) making it inconvenient for screening large volumes of samples. In addition, a lab technician or specialist with a degree of expertise is often needed to operate and maintain HPLC equipment. This combination of high cost, time, and personnel can be widely limiting for many soybean breeding programs and even meal producers, highlighting the need to identify more rapid and cost-efficient sugar quantification methods that can provide similarly accurate and robust results.

Near-Infrared Reflectance Spectroscopy (NIRS) is an analytical technique that exploits absorbance of light from the near-infrared region (700-2500nm) by various C-O, C-N, and C-H chemical bonds to estimate sample concentration for a given chemical constituent of interest (Stark et al., 1986). In NIRS, a light beam of near-infrared wavelengths, varying in useful width between instruments of different manufacturers, is emitted through a sample. Upon contact, the near-infrared light of various wavelengths is differentially absorbed based on the sample's distinct chemical composition. The wavelengths that are not absorbed by the sample are reflected back to a detector on the NIR instrument, resulting in a unique fingerprint for each sample. These spectral fingerprints are then modeled against predetermined reference chemistry values of the same samples to develop algorithms, called calibrations or calibration curves, that are capable of predicting sample concentration in minutes or seconds (Givens & Deaville, 1999).

NIRS holds many advantages over the aforementioned sugar quantification methods that make it ideal to replace HPLC for soybean sugar profiling. Such advantages include considerably faster data acquisition, high reproducibility, greater sample flexibility, and inexpensive per sample cost (Baianu & Guo, 2011). Unlike HPLC, instrumentation is simple to maintain, does not require experienced personnel to operate, and involves minimal sample preparation which avoids the need for laboratory extraction altogether. NIR instruments are also capable of making multiple chemical measurements from a single NIR spectrum. Moreover, if spectral data is stored, calibrations for other traits of interest that are developed can be back-applied ex post facto. For these reasons, NIR instruments are already common in most breeding programs where they are routinely employed to collect information on protein, oil, and numerous other economically important chemical constituents in agricultural crops and products (Karn et al., 2017; Ringsted et al., 2017; Shiferaw et al., 2017; Xie et al., 2009).

The reliability of NIRS calibrations are directly related to the representativeness of the calibration set used for model creation (Mark, 1991). Calibration sets must contain hundreds to thousands of samples to account for the variable and unpredictable nature of samples from year to year, especially in breeding populations. A few studies have attempted to develop NIRS calibrations for soybean sugars using relatively large calibration sets previously. Choung (2010) was the first to attempt NIRS calibration development for soybean seed sugars using 155 samples. While his prediction model showed capability of predicting sucrose within 0.5% error, the calibration set lacked genotypes that exceeded 7% sucrose. Another study conducted by Baianu & Guo (2011), which sought to develop NIR calibrations for a series of soy food components including soybean seed sugars, found that even after using 12,000+ samples over four years, concentration range remained narrow. The only group that developed a prediction model with prediction range greater than 7% was Sato et al. (2012), who were only able to do so by supplementing their calibration set with samples containing artificially spiked levels of sucrose. As these studies demonstrate, naturally-occurring genotypes with appreciably high sucrose or RFO's are not easily ascertained in programs wishing to develop NIRS calibrations. Fortunately, the soybean breeding program at Virginia Tech has had an extensive focus on soybean sugars over the past two decades which has resulted in breeding germplasm exhibiting wide variation in sugar content and which contains multiple genotypes with high sucrose content and low RFO content (Maroof & Buss, 2011; Maughan et al., 2000; Maupin & Rainey, 2011; Skoneczka et al., 2009). This germplasm, which spans the entire 2 to 10% natural concentration range for sucrose, is ideal for NIRS calibration development for soybean seed sucrose. It also holds increased potential to develop improved prediction models for raffinose and stachyose which have been largely unsuccessful hitherto.

The DA7250 NIR instrument is the third generation of Perten's signature diode-array technology which exploits absorption along the 950-1,650nm wavelength range to predict sample concentrations in under 10 seconds (Perten Industries, n.d.). To our knowledge, no calibration studies have been reported for sucrose, raffinose, and stachyose content in ground soybean for this instrument. If accurate calibrations can be developed, the DA7250 can considerably reduce data collection time and effort needed for soybean sugar profiling. As such, the objectives of this study were to 1) exploit the wide variation observed in Virginia Tech's soybean germplasm to develop calibrations for sucrose, stachyose, and raffinose in ground soybean on the DA7250 instrument and 2) assess the potential that these calibrations hold to allow the DA7250 to replace HPLC for more rapid soybean sugar profiling.

Materials & Methods

Plant Material

A total of 268 samples for calibration development were selected from 184 different genotypes harvested from 2017 and 2018 yield trials in a triple replicated, randomized complete block design (RCBD) in five environments in Virginia: Blacksburg, Orange, Mount Holly, Warsaw full-season and Warsaw double-crop. These genotypes are advanced breeding lines developed by Virginia Tech soybean breeding program from maturity groups 4 and 5 for conventional, herbicide-tolerant, and specialty food-grade soybean markets with relatively diverse genetic background. Because the same genotype may produce different sugar content in different environments (location and year), we used some samples that are the same genotypes but from different environment in order to establish a calibration set with sufficient sample numbers. Upon harvest, subsamples of each replication of one sample are pooled into one composite sample for each sample at a particular location. These composite samples are subsequently transferred to lab for trait analysis. All samples that comprised the calibration set in this study were selected from these materials. Among them, 52 samples containing high, mid, and low concentrations of each sugar that showed increased potential for calibration development. These samples were deliberately included in the calibration set to establish a baseline for model creation with improved representativeness across natural sugar concentration ranges for each sugar. The remaining 216 samples were randomly selected from available breeding materials.

Spectral Methodology

Prior to NIRS scanning, approximately 20 grams of soybean seeds from each sample were ground in a temperature-controlled FOSS Knifetec 1095 grinder (FOSS Analytical, Hillerod, Capital Region, Denmark). All 268 samples were scanned twice (two repacks) on the Perten

DA7250 NIRS instrument (Perten Instruments, Springfield, IL) using a small breeder's cup (Perten Instruments, Springfield, IL) which is intended for use with flour-type products, to obtain two replications of spectral data. The resulting spectral dataset consisted of 141 datapoints collected between the 950 and 1,650 nm range. A dataset with averaged spectral absorbance values was subsequently exported for model creation. Following spectral data acquisition, ground samples were retained for immediate sugar extraction and HPLC analysis.

Sugar Quantification by HPLC

Sugar was extracted from the 268 ground samples using a micro-sugar sample extraction protocol optimized for our HPLC instrument. Extraction was performed twice for each sample in order to produce a technical replicate. In brief, 0.1 g of ground sample and 1 mL of HPLC-grade water were vortexed in a 2-mL centrifuge tubes and samples were put on a rocker shaker for 15 min. at 400 strokes per min. Following extraction, samples were centrifuged for 15 min. at 13.2 rpm, and 0.5 mL of resulting supernatant was transferred to a separate 2.0 mL centrifuge tube containing 0.7 mL HPLC-grade acetonitrile (ACN) bringing samples to 1.2 mL total volume and 58% ACN. The tubes were inverted several times and left to sit at room temperature for 1 hr. Samples were then centrifuged at 12.3 rpm for 15 min., and 100 µL of resulting supernatant was then mixed with 900 µL of 65% ACN and filtered into 2 mL HPLC vials for analysis using 0.2 µm membranes. The moisture content of ground samples was determined by oven dry method using a representative sample of the calibration set.

The HPLC instrumentation for the analysis includes an Agilent 1260 Infinity series (Agilent Technologies, Santa Clara, CA) equipped with an apHera^{TM} NH_2 Polymer, 5 µm analytical column and a 1260 Infinity ELSD detector (Agilent Technologies, Santa Clara, CA). The elution solvent was acetonitrile:water (65:35, v/v) with a flow rate of 1.0 mL/min. Final concentrations for sucrose, raffinose, and stachyose were averaged between technical replicates and reported in percent g per 100g seed on a dry weight basis. Standard deviation of replicates was 0.258% for sucrose, 0.054% for raffinose, and 0.198% for stachyose, while the mean difference between technical replicates for each sample was 0.29% for sucrose, 0.06% for raffinose, and 0.18% for stachyose. For quality control of the reference data, samples which contained a difference in concentration between technical replicates that exceeded three standard deviations of the mean difference between replicates were not considered for calibration development. The correlation between sucrose, raffinose, and stachyose was calculated using JMP®, 4.0 (SAS Institute Inc., Cary, NC).

Model creation, cross-validation, and statistical analysis

Spectroscopic data pretreatment, model creation, and internal cross-validation were performed using CAMO Unscrambler® X software (CAMO Analytics AS, Oslo, Norway). Pretreatment of spectra consisted of transformation with standard normal variate (SNV) and detrending to correct for particle size and light scatter. Final models were created using Partial Least Squares Regression (PLSR) preceded by mean centering. PLSR is an ideal multivariate method to use for calibration development since it imposes no restrictions on the number of terms, in this case wavelengths, that can be included in the prediction model which ultimately allows it to extract the maximum amount of information from spectral data (Hollung et al., 2005). PLSR has been widely cited in the literature for NIRS calibration development hitherto (Cozzolino et al., 2005; Rambla et al., 1997; Wang et al., 2006; Wu et al., 2012; Xie et al., 2009). The optimal number of PLSR components for model creation was selected based on the number of factors that minimized the PRESS statistic following cross-validation given a maximum of 20 components. Cross-validation consisted of randomly dividing the total number of samples in the calibration set into twenty segments. Each segment contained roughly eleven to twelve samples. During cross-validation, samples in a given segment were held out of the model and the model was recalibrated to the remaining samples. Then, the concentrations of the held-out samples were predicted based on the recalibrated model. This process is repeated for several iterations until all samples in the calibration set were held out. The CAMO Unscrambler X software can also be used to examine sample outliers. The limit of outliers removed during model creation was set to not exceed 15% of the total sample size. The resulting R^2 of cross-validation (R^2_{cv}) and RMSE of cross-validation (RMSECV) were used to evaluate the calibration model's prediction accuracy. R^2 explains how well the predicted sugar concentration, determined by the spectral-derived model, matches the actual sugar concentration obtained from HPLC, while RMSE reveals the average magnitude of prediction error that is expected to be observed in concentration percent for each sugar.

Results & Discussion

Of the 268 total samples from which HPLC data were collected, 15 were excluded from further analysis due to high standard deviation of difference (SDD) between replicates. Descriptive statistics for reference chemistry can be found in **Table 3**. The frequency graphs depicted in **Figures 3a-c.** illustrate how well the concentration range for each of the sugars was covered and represented by the 253 samples remaining after quality control of reference chemistry data.

The samples showed a considerably wide concentration range for sucrose content (2.06-10.37%) that closely mirrored the 2-10% sucrose concentration range reported in previous soybean sugar variation studies (Hou et al., 2009; Hymowitz & Collins, 1974) (**Figure 3a**). Here, the deliberate inclusion of the aforementioned 52 "baseline" set samples allowed our calibration set to exceed 7% sucrose without the need to artificially spike samples. Moreover, a considerable portion of samples contained between 2 and 3%, which also greatly improved representativeness at the lower end of prediction range. These results suggest that the baseline set samples used here can be broadly useful to other soybean breeding programs in allowing them to more conveniently develop in-house sucrose calibration equations on other NIRS instruments.

For stachyose, the vast majority of samples fell between approximately 2.4 and 4.5% (**Figure 3b**). Given that commercial soybean varieties typically contain between 3-4% stachyose, this was to be expected. While calibration set failed to contain samples with 0.7 to 2.4% stachyose concentration, many samples did contain ultra-low stachyose concentration (< 0.7%). This may be due to the tremendous advances that have been made in identifying low stachyose soybean mutants and their associated genomic regions over the past decade (Maroof & Buss, 2011; Skoneczka et al., 2009). One of the more important genes identified for soybean stachyose is the *mips1* gene, for which a mutation-specific marker has been shown to explain approximately 88-94% of the phenotypic variation for soybean seed stachyose content (Skoneczka et al., 2009). This *mips1* mutant has been introduced into breeding populations and repeatedly selected for over time in our breeding program, which may well explain the relatively high occurrence of ultra-low stachyose samples (< 0.7%) observed in our calibration set. Unlike ultra-low stachyose genotypes, genotypes with 1 to 2% stachyose are much less coveted by plant breeders making them considerably more difficult to obtain. The most recent sugar variation study of worldwide soybean germplasm collections by Hou et al. (2009) showed very low occurrence of genotypes with stachyose content between 0 and 2%, suggesting that even naturally such genotypes rarely occur.

The complete set of samples also covered a raffinose concentration range of 0.14-1.22% (**Figure 3c**), which is the widest concentration range reported for raffinose calibration development hitherto. Encompassed in this range is the 0.85-1.13% range that is typically observed in commercial soybean cultivars (Dierking & Bilyeu, 2009). The wide range observed here resulted from inclusion of numerous samples with low (< 0.7%) and ultra-low (< 0.4%) raffinose concentration, which allowed us to extend the lower end of our calibration set concentration range. The high occurrence of low and ultra-low raffinose genotypes reflects the successful integration of low raffinose genes into elite soybean germplasm over time, which are attributed to development of numerous molecular markers that have greatly enhanced breeding selection efficiency for low-raffinose individuals in breeding populations (Dierking & Bilyeu, 2008; Yang et al., 2014).

Correlational analysis of the reference values showed that raffinose and stachyose showed a significant ($p<0.05$) positive correlation with each other ($r = 0.64$). The positive correlation between raffinose and stachyose has been observed in extensive variation and compositional studies of soybean seeds hitherto (Hou et al., 2009; Y. Wang et al., 2014). Raffinose (-0.45) and stachyose (-0.83) both also showed significant negative correlations with sucrose. While significant positive correlations have been previously reported between sucrose and raffinose (Hartwig et al., 1997; Hymowitz et al., 1972), the negative correlation observed here is likely due to strong breeding selection for sucrose and against raffinose during food-grade soybean cultivar development at Virginia Tech over the years.

Final models for sucrose, stachyose, and raffinose were developed using 11 PLSR variables. During model creation, several outliers from the complete set of 253 samples were identified. These outliers all showed particularly high Y variance, possibly resulting from errors in HPLC automatic peak integration. These outliers were omitted from final model creation in order to improve the accuracy of the prediction models. Ultimately, 34 of the 253 samples were omitted from the sucrose prediction model, and the raffinose and stachyose prediction models each omitted 23 samples (**Table 4**).

The predicted vs. reference plots for sucrose, stachyose, and raffinose models shown in **Figures 4a-c** provide a visual representation of the accuracy of each model. Accuracy of cross-validation is described by determination coefficient of cross-validation (R^2_{cv}) and root mean square error of cross-validation (RMSECV) which indicate how closely the predicted values from the model align with actual lab values obtained via HPLC. A summary of calibration statistics obtained for the three prediction models and their cross-validations can be found in **Table 4**. Despite observing a considerably narrower wavelength region than previous calibration attempts, prediction models with high predictive accuracy were developed for sucrose and stachyose. For sucrose, the final model showed capability of predicting sucrose concentration within 0.6% of the reference value during cross validation (RMSECV = 0.596) with a considerably wide prediction range. For stachyose, a similarly accurate prediction model capable of predicting stachyose within approximately 0.4% (RMSECV = 0.405) was also developed. For raffinose however, development of an accurate prediction model was not successful. Even after removal of outliers, the model poorly explained variation in concentration (R^2_{cv} = 0.477) and showed high RMSECV (0.144) within to its narrow concentration range.

To better understand how each prediction model is constructed, we can graphically represent the contribution of the regression coefficient for each wavelength — larger values indicate greater variance is explained, while directionality indicates whether the prediction is positive or negative. In **Figure 5a-c**, the peak observed near the 1,520nm wavelength region is notably larger than surrounding wavelengths for all three sugars. In addition, large peaks are observed near the 1,220nm wavelength region for sucrose and stachyose. The relatively high influence of the 1,220nm and 1,520nm wavelength regions for sucrose calibration are consistent with previous findings by Choung (2010). In their study, the 1,212nm and 1,518nm wavelengths were listed among the main absorption bands for sucrose prediction over a considerably wider NIR wavelength range (400-2,500nm). The 1,220nm and 1,520nm wavelength regions were identified using WinISI II software 4.0 (Foss and Infrasoft International LLC, USA) to corresponded to the C-H stretching second overtone (-CH_2) of carbohydrates and N-H stretching first overtone of protein, respectively. Considering this, the positive influence of the 1,220nm wavelength and concomitant negative influence of 1,520nm wavelength reflects an inverse relationship between sugar and protein content in soybean seed (Wilcox & Shibles, 2001; Hartwig et. al., 1997). Though the 1,520nm wavelength region is associated with protein content and does not directly reflect changes in sugar content, our results suggest that the highly interdependent nature of protein and sugar content in soybeans allow it to be nevertheless useful in soybean sugar NIRS calibration development.

A major difference in wavelength contribution was observed in the raffinose prediction model compared to sucrose and stachyose prediction models, which may have contributed to its poor performance. Namely, the 1,220nm wavelength region – which sucrose and stachyose prediction were highly dependent on – showed relatively low influence for prediction of raffinose content. The considerably lower relative contribution observed for the 1,220nm wavelength here may be explained by intrinsic differences in the sugar molecular structure of raffinose (trisaccharide) compared to sucrose (disaccharide) and stachyose (tetrasaccharide). Such differences may diminish absorbance at useful wavelengths, ultimately reducing the number of diagnostic wavelengths that can be more heavily relied upon by the calibration. Another factor likely contributing to the poor model performance for raffinose is due to the natural variation in raffinose being below the sensitivity of the measurement device. It maybe also because of the relative low variation of raffinose content, arising from its considerably narrower concentration range compared to other sugars. Small variation within the calibration set are known to exacerbate the effect of slight variability between samples at similar points along the concentration range on the overall regression (Mark, 1991). This has been cited in previous calibration development attempts for soybean RFOs, especially for raffinose (Choung, 2010). Here, it is possible that more diagnostic absorption bands which more effectively capture variation in raffinose content lie outside of the 950-1,650 nm wavelength range of the DA7250 instrument. Raffinose calibration development was also found to be problematic for Choung (2010), even when using 1,050 data points collected from an NIRS instrument that spanned a considerably wider wavelength range (400-2,500nm). In addition, even though ultra-high raffinose genotypes (< 3%) can be obtained for calibration development through artificial spiking or mutation, the R^2 value might increase, but

RMSECV may not be changed by much. Given these results, NIRS instruments may not be a viable tool to replace HPLC for determination of raffinose content.

As with any chemometric study, the utility of prediction models largely depends on the genetic diversity used for calibration, and the reported models can always be improved in subsequent years by using new genotypes grown in different environments which can continually improve representativeness and potentially further extend prediction range. It is crucial to note that in our study, genetic diversity was limited to germplasm from the Virginia Tech soybean breeding program. However, because we have been incorporating diverse germplasm from other countries and other breeding programs into our gene pool, the genetic base of this set is relatively large with broad pedigree. In addition, the 52 baseline samples used for calibration which spanned the entire 2-10% sucrose concentration range have been retained and are available upon request with a material transfer agreement. These samples can be used by other soybean breeding programs to develop their own in-house calibrations using whichever NIRS instrumentation they have available.

Conclusions

Here we report the development of accurate prediction models for sucrose and stachyose on the Perten DA7250 NIRS instrument using soybean germplasm originating at Virginia Tech that exhibited exceptionally wide, naturally-occurring variation in sugar content. Having developed these models using the DA7250, we demonstrated that NIR instruments need not cover the entire 700-2,500nm NIR wavelength range for successful calibration development for these two sugars, provided that spectral data can be collected from the 1,220nm and 1,520nm wavelength regions. In regards to raffinose calibration development, the poor results observed in our study suggest that NIRS instruments may have limited potential to accurately predict raffinose concentration. Instead, future studies should investigate the use of other cost-effective, high-throughput analytical instruments capable of capturing information from other regions of the electromagnetic spectrum that can more effectively capture natural variation in raffinose content that can be exploited for prediction.

References

Baianu, I., & Guo, J. (2011). NIR Calibrations for Soybean Seeds and Soy Food Composition Analysis: Total Carbohydrates, Oil, Proteins and Water Contents. *Nature Precedings*. https://doi.org/10.1038/npre.2011.6611.1

Black, L. T., & Bagley, E. B. (1978). Determination of oligosaccharides in soybeans by high pressure liquid chromatography using an internal standard. *Journal of the American Oil Chemists' Society*, *55*(2), 228–232.

Choung, M.-G. (2010). Determination of Sucrose Content in Soybean Using Near-infrared Reflectance Spectroscopy. *Journal of the Korean Society for Applied Biological Chemistry*, *53*(4), 478–484. https://doi.org/10.3839/jksabc.2010.073

Cozzolino, D., Murray, I., Chree, A., & Scaife, J. R. (2005). Multivariate determination of free fatty acids and moisture in fish oils by partial least-squares regression and near-infrared spectroscopy. *LWT - Food Science and Technology*, *38*(8), 821–828. https://doi.org/10.1016/j.lwt.2004.10.007

Dierking, E. C., & Bilyeu, K. D. (2008). Association of a Soybean Raffinose Synthase Gene with Low Raffinose and Stachyose Seed Phenotype. *The Plant Genome*, *1*(2), 135–145. https://doi.org/10.3835/plantgenome2008.06.0321

Dierking, E. C., & Bilyeu, K. D. (2009). Raffinose and stachyose metabolism are not required for efficient soybean seed germination. *Journal of Plant Physiology*, *166*(12), 1329–1335. https://doi.org/10.1016/j.jplph.2009.01.008

Givens, D. I., & Deaville, E. R. (1999). The current and future role of near infrared reflectance spectroscopy in animal nutrition: A review. *Australian Journal of Agricultural Research*, *50*(7), 1131–1145. https://doi.org/10.1071/ar98014

Hollung, K., Øverland, M., Hrustić, M., Sekulić, P., Miladinović, J., Martens, H., Narum, B., Sahlstrøm, S., Sørensen, M., Storebakken, T., & Skrede, A. (2005). Evaluation of Nonstarch Polysaccharides and Oligosaccharide Content of Different Soybean Varieties (*Glycine max*) by Near-Infrared Spectroscopy and Proteomics. *Journal of Agricultural and Food Chemistry*, *53*(23), 9112–9121. https://doi.org/10.1021/jf051438r

Hou, A., Chen, P., Alloatti, J., Li, D., Mozzoni, L., Zhang, B., & Shi, A. (2009). Genetic Variability of Seed Sugar Content in Worldwide Soybean Germplasm Collections. *Crop Science*, *49*(3), 903–912. https://doi.org/10.2135/cropsci2008.05.0256

Hymowitz, T., & Collins, F. I. (1974). Variability of Sugar Content in Seed of Glycine max (L.) Merrill and G. soja Sieb. And Zucc.1. *Agronomy Journal*, *66*(2), 239–240. https://doi.org/10.2134/agronj1974.00021962006600020017x

Jo, H., Lee, J.-D., & Bilyeu, K. D. (2018). Environmental Stability of Carbohydrate Profiles in Different Soybean Genotypes. *Crop Science*, *58*(2), 773–782. https://doi.org/10.2135/cropsci2017.08.0497

Jo, H. (Scientist). (2016). *Understanding sucrose and raffinose family of oligosaccharides in soybean seed for human and animal utilization* [Thesis, University of Missouri--Columbia]. https://doi.org/10.32469/10355/57188

Karn, A., Heim, C., Flint-Garcia, S., Bilyeu, K., & Gillman, J. (2017). Development of Rigorous Fatty Acid Near-Infrared Spectroscopy Quantitation Methods in Support of Soybean Oil Improvement. *Journal of the American Oil Chemists' Society*, *94*(1), 69–76. https://doi.org/10.1007/s11746-016-2916-4

Kumar, V., Rani, A., Goyal, L., Dixit, A. K., Manjaya, J. G., Dev, J., & Swamy, M. (2010). Sucrose and Raffinose Family Oligosaccharides (RFOs) in Soybean Seeds As Influenced by Genotype and Growing Location. *Journal of Agricultural and Food Chemistry*, *58*(8), 5081–5085. https://doi.org/10.1021/jf903141s

Mark, H. (1991). *Principles and Practice of Spectroscopic Calibration*. John Wiley & Sons.

Maroof, M. A. S., & Buss, G. R. (2011). *Low phytic acid, low stachyose, high sucrose soybean lines* (United States Patent No. US8003856B2). https://patents.google.com/patent/US8003856B2/en

Maughan, P. J., Maroof, M. A. S., & Buss, G. R. (2000). Identification of quantitative trait loci controlling sucrose content in soybean (Glycine max). *Molecular Breeding*, *6*(1), 105–111. https://doi.org/10.1023/A:1009628614988

Maupin, L. M., & Rainey, K. M. (2011). Improving emergence of modified phosphorus composition soybeans: Genotypes, germplasm, environments, and selection. *Crop Science*, *51*(5), 1946–1955.

Maupin, L. M., Rosso, M. L., & Rainey, K. M. (2011). Environmental Effects on Soybean with Modified Phosphorus and Sugar Composition. *Crop Science*, *51*(2), 642–650. https://doi.org/10.2135/cropsci2010.07.0396

Mayta, J., Chen, P., Popp, M. P., Dong, D., Wu, C. J., Zhang, B., Smith, S. F., & Scaboo, A. M. (2014). *Break-even profitability for food-grade specialty soybeans*. *2*(2), 01–11.

Perten Industries. (n.d.). *DA 7250^{TM} NIR instrument overview | Perten Instruments*. https://www.perten.com/Products/DA-7250-NIR-analyzer/Instrument-overview/

Rackis, J. J. (1975). Oligosaccharides of Food Legumes: Alpha-Galactosidase Activity and the Flatus Problem. In *Physiological Effects of Food Carbohydrates* (Vol. 15, pp. 207–222). AMERICAN CHEMICAL SOCIETY. https://doi.org/10.1021/bk-1975-0015.ch013

Rambla, F. J., Garrigues, S., & de la Guardia, M. (1997). PLS-NIR determination of total sugar, glucose, fructose and sucrose in aqueous solutions of fruit juices. *Analytica Chimica Acta*, *344*(1), 41–53. https://doi.org/10.1016/S0003-2670(97)00032-9

Ringsted, T., Ramsay, J., Jespersen, B. M., Keiding, S. R., & Engelsen, S. B. (2017). Long wavelength near-infrared transmission spectroscopy of barley seeds using a supercontinuum laser: Prediction of mixed-linkage beta-glucan content. *Analytica Chimica Acta*, *986*, 101–108. https://doi.org/10.1016/j.aca.2017.07.008

Sato, T., Zahlner, V., Berghofer, E., Lošák, T., & Vollmann, J. (2012). Near-infrared reflectance calibrations for determining sucrose content in soybean breeding using artificial reference samples. *Plant Breeding*, *131*(4), 531–534. https://doi.org/10.1111/j.1439-0523.2012.01975.x

Shiferaw, L., Tsegay, G., & Asamenew, G. (2017). Near-Infrared Reflectance Spectroscopy (NIRs) for Determination of Tryptophan Content in Quality Protein Maize (QPM). *Science Journal of Analytical Chemistry*, *5*(1), 8. https://doi.org/10.11648/j.sjac.20170501.12

Skoneczka, J. A., Maroof, M. A. S., Shang, C., & Buss, G. R. (2009). Identification of Candidate Gene Mutation Associated With Low Stachyose Phenotype in Soybean Line PI200508. *Crop Science*, *49*(1), 247–255. https://doi.org/10.2135/cropsci2008.07.0403

Stark, E., Luchter, K., & Margoshes, M. (1986). Near-Infrared Analysis (NIRA): A Technology for Quantitative and Qualitative Analysis. *Applied Spectroscopy Reviews*, *22*(4), 335–399. https://doi.org/10.1080/05704928608060440

Stitt, M., Lilley, R. McC., Gerhardt, R., & Heldt, H. W. (1989). [32] Metabolite levels in specific cells and subcellular compartments of plant leaves. In *Methods in Enzymology* (Vol. 174, pp. 518–552). Academic Press. https://doi.org/10.1016/0076-6879(89)74035-0

Suarez, F. L., Springfield, J., Furne, J. K., Lohrmann, T. T., Kerr, P. S., & Levitt, M. D. (1999). Gas production in humans ingesting a soybean flour derived from beans naturally low in oligosaccharides. *The American Journal of Clinical Nutrition*, *69*(1), 135–139. https://doi.org/10.1093/ajcn/69.1.135

Švejstil, R., Musilová, Š., & Rada, V. (2015). Raffinose-Series Oligosaccharides in Soybean Products. *Scientia Agriculturae Bohemica*, *46*(2), 73–77. https://doi.org/10.1515/sab-2015-0019

Teixeira, A. I., Ribeiro, L. F., Rezende, S. T., Barros, E. G., & Moreira, M. A. (2012). Development of a method to quantify sucrose in soybean grains. *Food Chemistry*, *130*(4), 1134–1136. https://doi.org/10.1016/j.foodchem.2011.07.128

Wang, L., Lee, F. S. C., Wang, X., & He, Y. (2006). Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR. *Food Chemistry*, *95*(3), 529–536. https://doi.org/10.1016/j.foodchem.2005.04.015

Wongputtisin, P., Ramaraj, R., Unpaprom, Y., Kawaree, R., & Pongtrakul, N. (2015). Raffinose family oligosaccharides in seed of Glycine max cv. Chiang Mai60 and potential source of prebiotic substances. *International Journal of Food Science & Technology*, *50*(8), 1750–1756. https://doi.org/10.1111/ijfs.12842

Wu, Z., Xu, B., Du, M., Sui, C., Shi, X., & Qiao, Y. (2012). Validation of a NIR quantification method for the determination of chlorogenic acid in Lonicera japonica solution in ethanol precipitation process. *Journal of Pharmaceutical and Biomedical Analysis*, *62*, 1–6. https://doi.org/10.1016/j.jpba.2011.12.005

Xie, L., Ye, X., Liu, D., & Ying, Y. (2009). Quantification of glucose, fructose and sucrose in bayberry juice by NIR and PLS. *Food Chemistry*, *114*(3), 1135–1140. https://doi.org/10.1016/j.foodchem.2008.10.076

Yang, K., Ko, J.-M., Ha, T. J., Lee, Y.-H., Baek, I.-Y., Yang, T.-J., & Nou, and I.-S. (2014). Development of Molecular Markers for Low Raffinose and Stachyose in Korean Soybean Cultivars. *Plant Breeding and Biotechnology*, *2*(2), 151–157. https://doi.org/10.9787/PBB.2014.2.2.151

Tables & Figures

Table 3. Descriptive statistics of sugar content of 253 entries.

Sugars	Mean (%)*	Range (%)	Standard Dev. (%)
Sucrose	4.33	2.06-10.37	1.77
Stachyose	3.01	0.16-4.38	1.24
Raffinose	0.66	0.14-1.22	0.21

*Mean, range, and standard deviation are reported in g per 100g seed (%).

Table 4. Summary of calibration and cross-validation statistics for three prediction models on sugar content.

Sugars	PLSR Variables	Sample number	Calibration		Cross-validation	
			R^2_C	RMSEC	R^2_{CV}	RMSECV
Sucrose	11	219	0.901	0.516	0.869	0.596
Stachyose	11	231	0.911	0.361	0.891	0.405
Raffinose	11	231	0.568	0.129	0.476	0.143

Note: PLSR, Partial Least Squares Regression; R^2c, coefficient of determination; RMSEC, Root Mean Squared Error of Calibration; R^2cv, coefficient of determination of Cross-Validation; RMSECV, Root Mean Squared Error of Cross-Validation.

Figure 3. Frequency distribution of the 253 entries for sucrose, stachyose, and raffinose concentration. In Fig. 3A, bin width = 0.5%, starting with 2.1-2.6% ending with >10% sucrose. In Fig. 3B, bin width is 0.5%, starting with 0.2-0.7% and ending with >4% stachyose. In Fig. 3C, bin width is 0.1%, starting with 0.1-0.2% and ending at 1.1-1.2% raffinose.

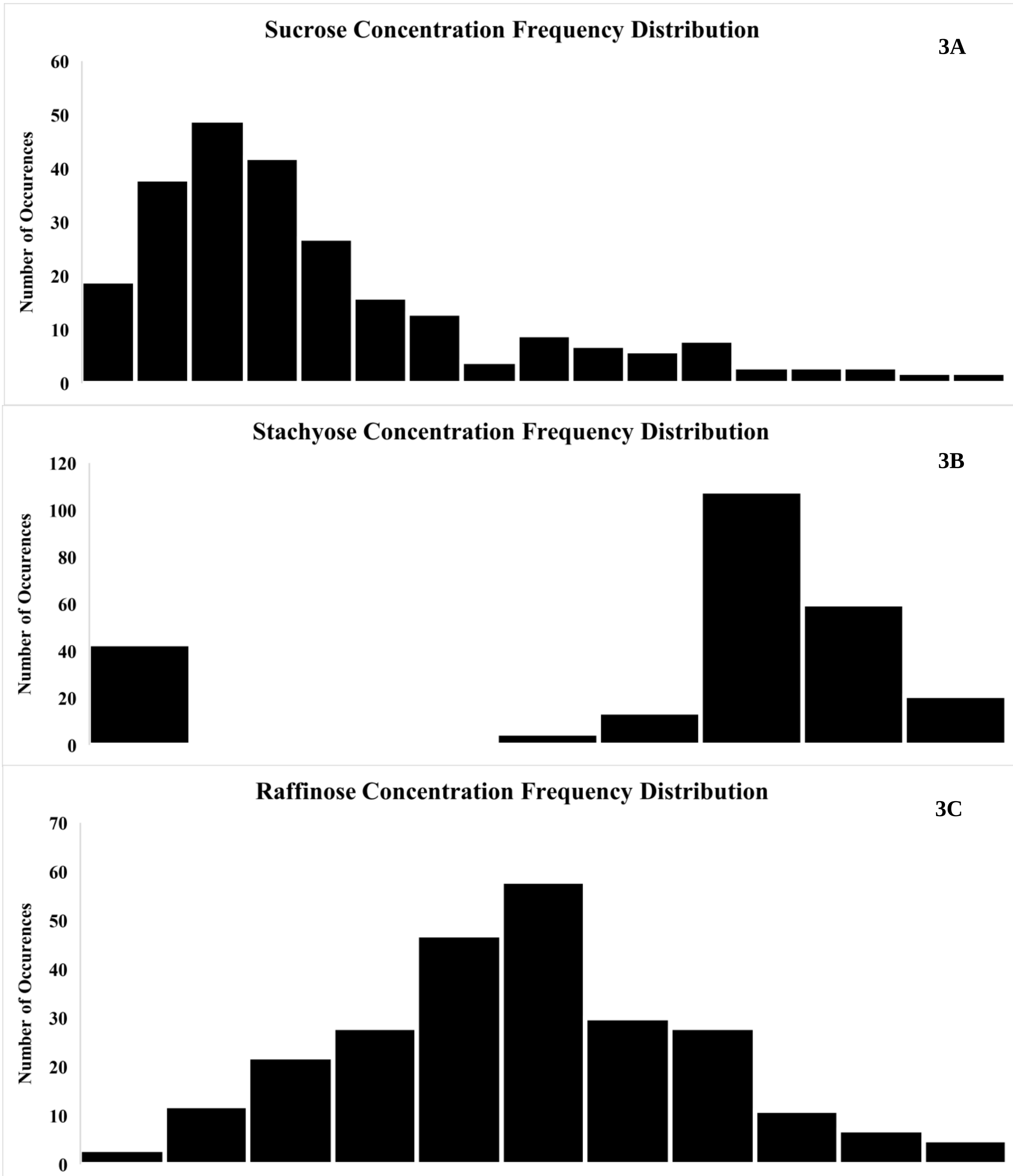


Figure 4. Predicted vs. Reference graphs for sucrose (4A), stachyose (4B), and raffinose (4C) prediction models after cross-validation.

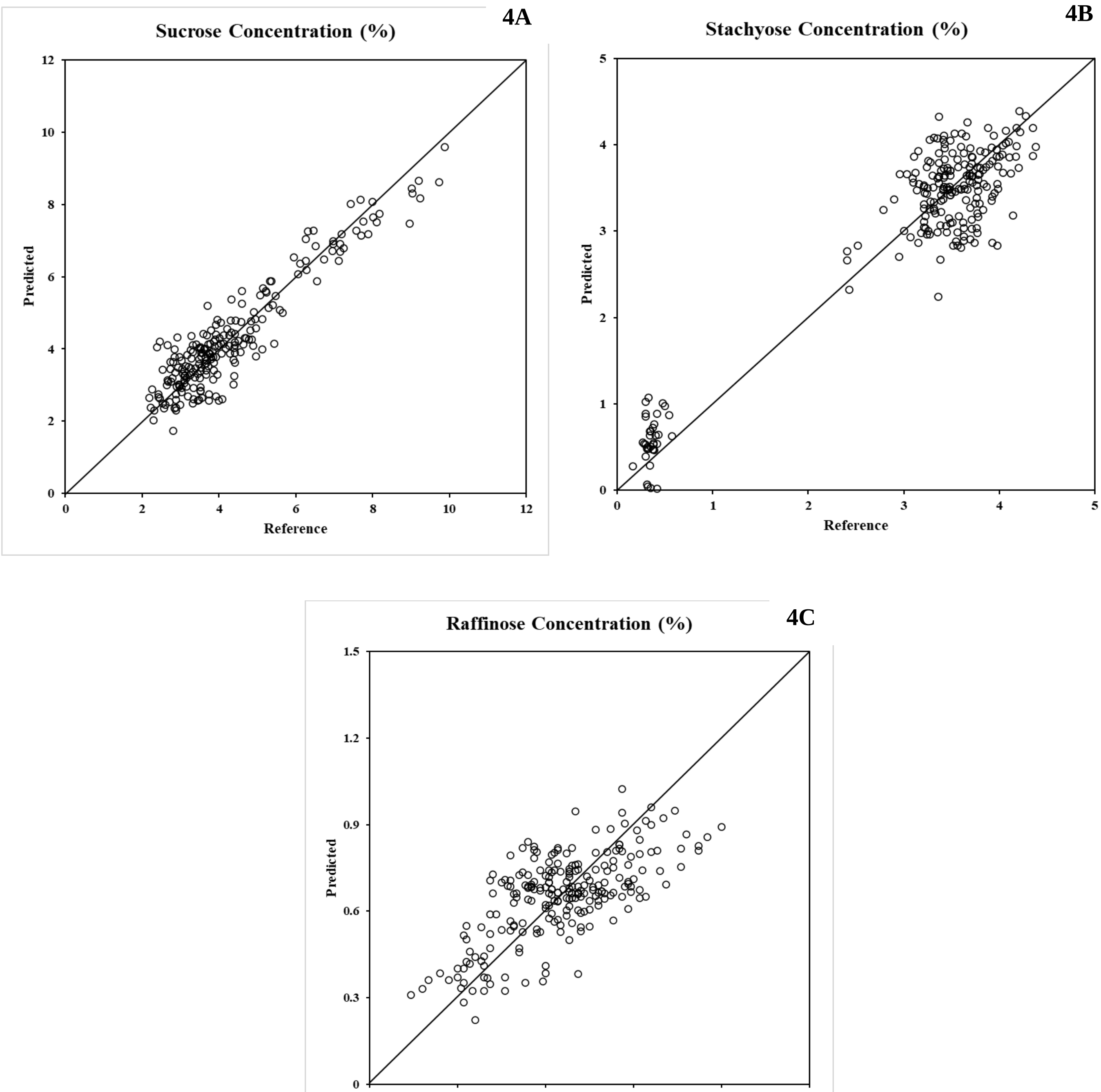
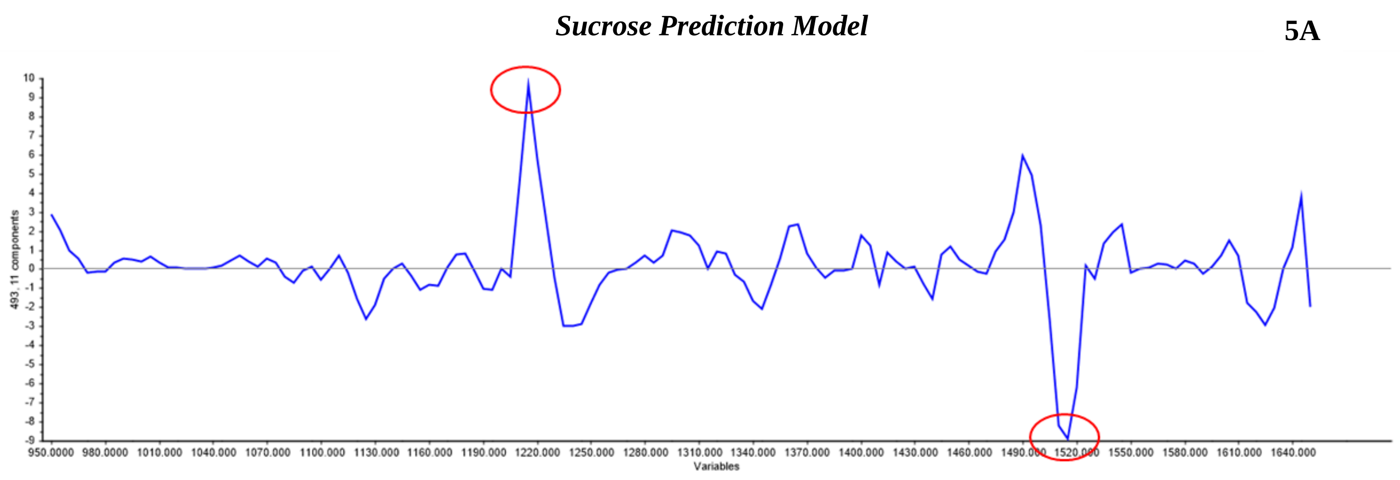
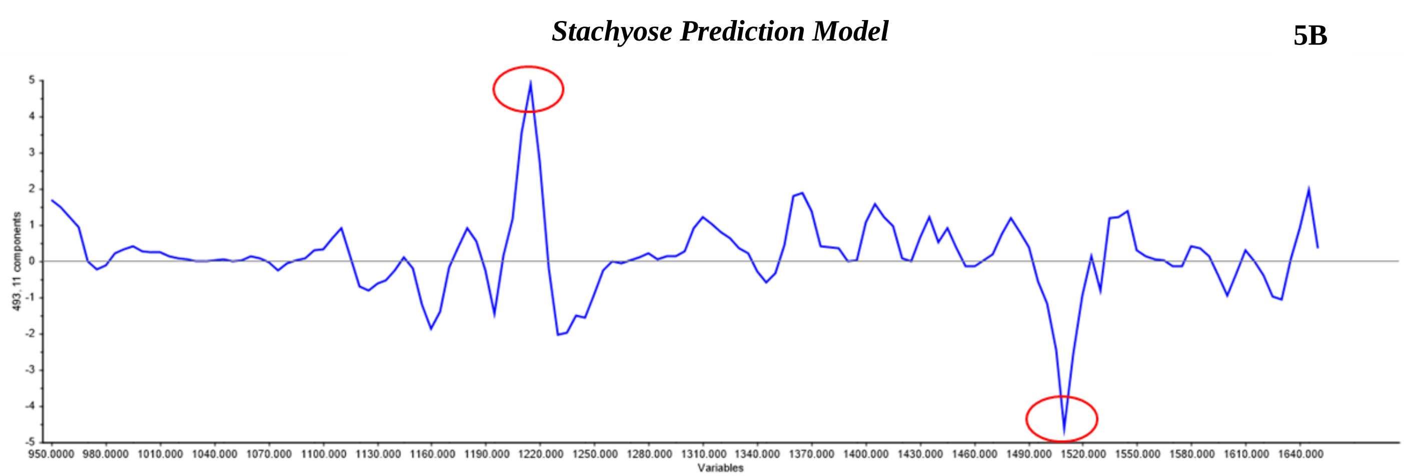
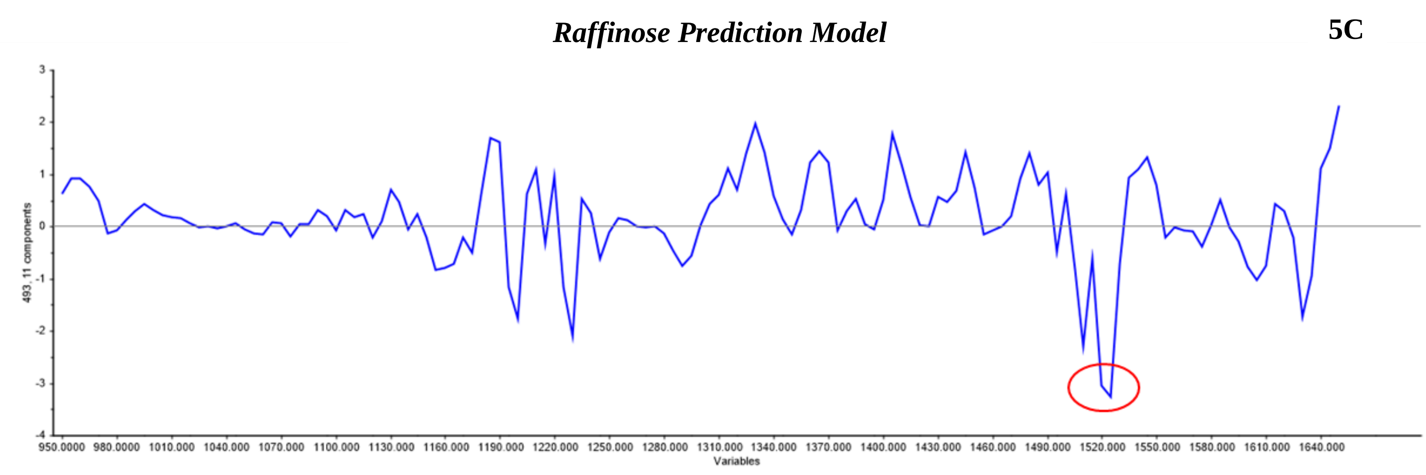


Figure 5. Relative contribution of wavelengths between 950 and 1,650nm to prediction models for sucrose (5A), stachyose (5B), and raffinose (5C) prediction models.







Final Conclusions

Sustainability of the soybean industry can be enhanced by increasing the amount of acreage devoted to soy production for specialty products. Vegetable-type soybean, more commonly known as edamame, has surged in consumer demand and interest over the past two decades. Domestic production of edamame has been hindered in part due to a lack of strong economic evidence to demonstrate its potential profitability. Here, we employed a pair of economic studies – looking at production cost and consumer demand – to address this issue for producers in the Mid-Atlantic region. Observations from our feasibility study suggested that despite observing higher yield potential from the use of locally-adapted edamame cultivars, the substantial labor costs associated with hand-harvested edamame still largely offset or exceed any potential gains in revenue. Future research should therefore focus on production costs associated with mechanized harvest, including yield loss from harvesters and the viability of using commercial bean harvesters for other crops such as green beans and lima beans to harvest edamame. Determining feasibility and production costs associated with shelled edamame, which involves the added expense of shelling equipment and bean sorters, would also be of considerable value in clarifying the economic potential of edamame in the region.

In our willingness-to-pay study, we sought to better understand the future outlook for domestically produced edamame by estimating the premiums and discounts that consumers placed on fresh, local, organic, and on-the-stalk edamame while available alongside their currently available, status-quo counterparts (frozen, non-local, non-GMO, pods) under a hypothetical, one-and-one-half bound contingent valuation framework and related these estimates to a vector of explanatory variables. Our results revealed that survey participants strongly preferred fresh or locally produced edamame rather than frozen or non-local edamame by an average margin of 94 and 87 cents (respectively). For fresh, this estimate appeared to be largely driven by demographic factors, especially gender, and shopping locale, while local showed more broad interest across explanatory variables. For both fresh and local, environmental value was also a significant driving factor. Taken together, these results suggest that fresh and/or locally supplied edamame may hold a strong marketing advantage over currently available products in the domestic marketplace.

The final study presented in this thesis sought to use chemometric methods to address the burdensome task of quantifying soybean seed sugars that hold implications for taste, digestibility, and feed efficacy of food-grade soybean. In our study, we demonstrated that by using germplasm that exhibited a relatively wide range of sucrose and stachyose content, highly accurate and robust NIRS prediction models for these major seed sugars could be developed on the DA7250 NIRS Analyzer. Ultimately, this can save money, time, resources, and personnel for soybean breeding programs and other stakeholders by allowing them to estimate relative sugar content more rapidly and efficiently in order to meet industry needs. Despite the success observed for sucrose and stachyose, calibration development for raffinose remains problematic. As such, higher resolution NIRS instruments or alternative methods to more accurately and efficiently quantify raffinose in soybean seeds must be investigated.