A study of soybean processing value maximization using selective handling strategies based on the analysis of soybeans received at Iowa elevators

by

Bennett Elizabeth Barr

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Agricultural and Biosystems Engineering

Program of Study Committee:
Charles R. Hurburgh, Jr., Major Professor
Gretchen A. Mosher
Bobby J. Martens

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

Iowa State University

Ames, Iowa

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my major professor, Dr. Charlie Hurburgh, and my committee members, Dr. Gretchen Mosher and Dr. Bobby Martens, for their guidance, advice, and flexibility over the past year. Dr. Hurburgh, thank you for teaching me that most of the lessons of graduate school are learned outside the classroom, and for always being willing to answer my questions about agriculture and Iowa. I would also like to thank Dr. Erin Bowers and Mrs. Kris Bell– I would not have made it through this process without their support and friendship. Thank you also to Glen Rippke and Alan Kuuttila for their NIR and technology assistance, and to all the cooperative staff who tested soybean samples.

In addition, I would like to thank my family and boyfriend for their love and constant encouragement. I appreciate and love you all more than I can express.

ABSTRACT

During the Fall 2018 soybean harvest, soybean samples were collected from 32 country elevator locations belonging to one Iowa-based cooperative which has its own elevator locations and processing plant. This was done to update historical data about geographic variations in protein and oil content of Iowa soybeans, and to assist the cooperative in making more informed decisions about their soybeans to maximize value potential. These samples were analyzed using near-infrared spectroscopy (NIR) to determine protein and oil contents. The data were accumulated and sorted to look for geographic variations in protein and oil content of soybeans throughout Iowa. The data were run through an Estimated Processing Value (EPV) model to determine value differences of soybeans between elevator locations. The cooperative source soybeans for processing from the elevator locations closest to the plant to mitigate trucking costs. They wanted to know whether this strategy was maximizing their net processing value. Results showed that significant variability between locations did exist, which represented a $0.23/bushel EPV spread. Additionally, it was found that 15 samples were needed to accurately represent an elevator location, and that two weeks was a sufficient period to characterize the data to be able to make a marketing decision. Lastly, the soybean quality was not found to vary significantly over the course of harvest, so marketing decisions can made at the beginning of the season.

An error analysis was also performed to find the effects of potential error on location separation, because errors would reduce the certainty of any marketing decisions based on measured value differences. Both random and systematic errors were possible with the use of NIR analyzers. Random errors were simulated using an Excel-based model that created random values with a specified standard deviation and mean, which were then added to the original data points. This simulation was performed for three test cases – one with typical standard deviations

for protein and oil contents, one with higher-than-average standard deviations, and one with typical standard deviations but with a bias element added to a subset of the locations. The introduction of random error made any value gaps between locations smaller, which made discrimination of high-value locations from average or low-value locations difficult. These results showed the importance of having standards for measuring instruments if the soybean supply chain is ever to move to a protein and oil pricing basis, because one of the largest sources of error in a commodity-based market system is inconsistency of measuring units with each other.

Overall, geographic variability across the cooperative's locations was evident, and testing inbound loads with an NIR analyzer, even during busy harvest days, was feasible to characterize soybean protein and oil content. However, the validity of marketing decisions made using the resulting data depends highly on the amount of error involved in sample analysis. Future studies should identify specific sources of error and attempt to eliminate them, because maximizing potential value capture will not be possible unless the value differences between locations are characterized as precisely as possible.

CHAPTER 1. GENERAL INTRODUCTION

Increasing profit from soybean farming is a current topic of interest in the agricultural community. This discussion begins with the soybean varieties farmers choose to plant each year, as their choices mark the beginning of the supply chain and affect every user from that point forward. The United Soybean Board states that "farmers often look to yield first" for varietal selections, but that "the future of profitability lies in meeting end-user composition needs" because "end users don't need *soybeans*. They need the protein and oil that come from soybeans" ("Measuring Beyond," 2017). These facts of the industry are important for farmers growing soybeans, for the grain handlers who both buy soybeans and sell them into the larger grain market, and for the processors who transform them into an assortment of products. Currently, the trades between farmers and elevator and elevator and processor are standalone decisions not involving composition – processors pay for weight and moisture, not protein and oil content. However, as Hurburgh and Brumm write, "processors have traditionally believed that localized patterns of protein and oil content do exist" but that "there is little such data in the public domain" (Hurburgh & Brumm, 1990).

More data is needed to confirm that protein and oil patterns are present, but that data is difficult to acquire, as it must be collected from individual elevator locations. An Iowa-based cooperative group wanted to collect samples from their locations to characterize the soybeans they were receiving to determine their value. This group is unique in that it has its own elevators and processing plant, which allows them more flexibility in their marketing and handling strategies than other cooperatives, who are accountable to external processors. This company purchases beans from farmers and then either sells them to other processors or transports them to their own facility, which makes soybean meal and oil (which is then made into biodiesel). They currently allocate their beans geographically, meaning that they move beans from their closest locations to their processing plant to save on transportation costs. This strategy may not be allowing them to capture the full value of the soybeans they handle. However, most elevators still consider segregation of soybeans into high and low value groups at an elevator to be too costly and inefficient, especially during the busiest days of harvest, though this has been disproven in previous studies. This cooperative group wanted to know whether a separation strategy for their soybeans to identify the highest-value beans for processing would be a) feasible and b) economically worthwhile. The goal of this thesis is to answer that question through analysis of soybean samples collected during the Fall 2018 harvest in Iowa.

Literature Review

Introduction

Soybeans have been an important part of the American poultry and livestock feed markets since the early 1950s and have become a key part of the US's presence in worldwide agricultural commodities trading. For example, the United States exported an estimated 2.13 billion bushels of soybeans in 2018 and are projected to export another 1.9 billion bushels in 2019 (Soy Meal Info Center, 2018). The United Soybean Board states that "soybeans are extremely important to the U.S. farm economy, valued at about $15 billion dollars annually" (Johnson & Smith, 2018). Soybeans are one of the most important crops grown in the Midwest and throughout the country, accounting for 90% of oilseed production in the United States (USDA-ERS, 2018). According to the USDA-ERS, over 90 million acres of soybeans were planted in 2017 and in 2018, for the first time since 1983, there were more acres of soybeans planted than acres of corn. Figure 1.1 illustrates the typical nutritional composition of a soybean (United Soybean Board, 2015). The key values are the 36% protein and 19% oil, which make soybeans valuable for a variety of nutritional and industrial uses.

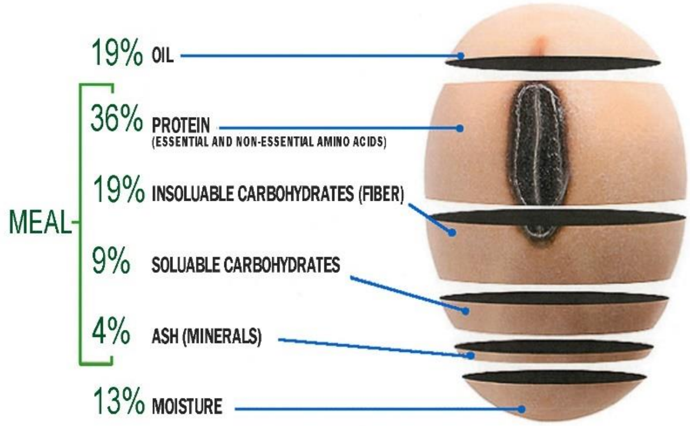

Figure 1.1. Soybean composition infographic.

Soybean meal is easily manufactured from these beans and has a high protein content that makes it digestible. It is an ideal component of animal food diets. Additionally, soybean oil can be extracted and used in processes ranging from cooking to biodiesel manufacturing. These two processes influence each other, because as meal production increases, so does oil production to keep up with the increased volume of soybeans being processed. The U.S. Soybean Export Council (USSEC) writes that "increased use of soybean oil in the production of biodiesel in the U.S. will result in a greater supply of soybean meal availability for animal feed rations" They also point out that "numerous feeding studies around the world have compared soybean meal from other major soybean producing countries with U.S. dehulled soybean meal" and that "animals perform better when their feed rations" contain soybean meal produced in the U.S. because of its "quality, consistency, and reliability". They point out that "it all comes down to economics" – that U.S. meal adds value to animal diets which then make more profits for the producer. (*U.S. Soybean*, 2015)

Use of NIR Technology for Soybean Characterization

Near infrared spectroscopy (NIR) can characterize soybeans in terms of protein and oil content. Karl Norris, a USDA employee, responded to the “growing demands for fast, quantitative determinations of moisture, protein, and oil” by developing NIR technology (Hindle, 2008). The first application was for measuring protein and oil in wheat. Norris suggested that NIR could be used to measure protein, oil, and moisture in soybeans with the use of specific wavelengths. Agricultural products selectively absorb NIR radiation which then gives information about the molecular bonds within the material being measured (Shenk, Workman, & Westerhaus, 2008). By the 1970s, companies such as Dickey-John and Technicon were emerging as manufacturers of laboratory-scale NIR equipment. The technology has continued to evolve and is now a primary measurement method for many industries. NIR technology is especially useful in the agricultural industry because it can perform rapid low-cost analysis of nutrient composition so that nutritionists can formulate diets to more efficiently feed livestock (Shenk, Workman, & Westerhaus, 2008).

Knowledge of the protein and oil contents through NIR analysis can enable more targeted end usage of the soybeans, which can then maximize profits for all involved. Dr. Roy Brister, the director of nutrition and feed milling at Tyson foods, states that “having data on soy protein and oil content spreads the general awareness of the components that hold value for end users” (United Soybean Board, 2017). If farmers know that end users are looking for high protein beans, for example, they can select varieties that will balance their yield and desired protein levels, as long as market incentives support doing so. In turn, if an elevator knows incoming bean loads have high protein contents, they could choose to send those loads to a meal processor that would pay a premium for higher protein percentages. This leads to increased value at all levels of soybean processing and helps the commodity market shift towards more strategically managed soybean supply chains. However, the market does not currently offer premiums for high protein or oil contents in soybeans, which discourages farmers from varietal selection based on quality. Instead, yield drives producer decisions and little motivation exists for elevators to separate soybeans into value groups.

Estimated Processing Value (EPV) Calculation

In order to offer more information about the actual economics of this process, models have been developed that can use characteristics such as protein, oil, and moisture content, to predict an estimated processing value per bushel (EPVB) of those soybeans in a meal and oil production scenario. The earliest of these models was written by Updaw, Bullock, and Nichols in 1976, but was limited in its ability to adjust for soybeans with variable composition (Updaw, Bullock, & Nichols, 1976). Brumm and Hurburgh upgraded this model in 1990. Their new model enabled the user to adjust for marketing practices such as protein premiums or fiber content limitations and to account for dehulling and adding hulls back into mill feed (Brumm & Hurburgh, 1990). Wagner, Hurburgh, and Brumm updated the model again in 2017 to evaluate soybeans based not only on protein and oil content, but also amino acids, carbohydrates, and fatty acids (Wagner, 2017).

Past Studies on Soybean Composition

Soybean breeders and growers have, for the most part, focused primarily on yield instead of composition in choosing which traits to select and varieties to plant. This is because the supply chain participants have chosen not to invest in the testing and handling infrastructure that would preserve component values. However, Hurburgh and Brumm write, “processors have traditionally believed that localized patterns of protein and oil content do exist” but that “there is little such data in the public domain" (Hurburgh & Brumm, 1990). Table 1.1 summarizes previous studies that have produced data on average protein and oil contents of soybeans across the country.

Table 1.1. Past studies of geographic variations in soybean protein and oil content.

Reference	States/Regions Covered	# Samples Tested (Brumm and Hurburgh 2006) or # Locations Used	Protein (% at 13% Moisture)		Oil (% at 13% Moisture)	
			Average	Standard Deviation of Data (% pts)	Average	Standard Deviation of Data (% pts)
Brumm and Hurburgh, 2006 (U.S. Soybean Protein and Oil Survey Data 1994-2004)	IA, KS, MN, MO, NE, ND, SD	7963	34.9%	0.8%	18.6%	0.5%
	IL, IN, MI, OH, WI	6722	35.7%	0.7%	18.6%	0.7%
	AR, KY, LA, MS, OK, TN, TX	1535	35.9%	0.9%	18.9%	0.7%
	AL, FL, GA, NC, SC	274	36.4%	1.1%	18.8%	0.8%
	DE, MD, NJ, NY, PA, VA	274	36.4%	1.1%	18.8%	0.8%
Hurburgh, 1994	IA	1 (1989)	34.9%	1.0%	18.5%	0.7%
		1 (1990)	35.6%	0.9%	19.1%	0.6%
		1 (1991)	35.3%	1.2%	18.4%	0.7%
Hurburgh and Brumm, 1990	IA	9 (1985)	33.8%	1.0%	19.8%	0.5%
		12 (1986)	35.3%		18.6%	
		12 (1987)	34.5%		18.9%	
Hurburgh, Paynter, and Schmitt, 1987	IA, IL, OH, MN	11 (1983)	33.9%	1.0%	19.7%	0.5%
		11 (1984)	34.2%	1.0%	19.1%	0.5%

The standard deviation represents the variability of the data points relative to the average. The overall average standard deviation for these studies was 1.0% for protein and 0.6% for oil.

This forms an advance estimate of what an Iowa farm might expect in its assessment of variability in its beans, which in turn offers information about the potential value spread represented.

Past Studies on Segregation Feasibility in Other Grains

The industry has a long-standing belief that it would be too difficult and costly to separate grains into higher-quality and lower-quality categories at the elevator level. Separation is believed to cause more trouble than the gain in economic benefits would be worth. This is especially of concern during harvest, when grain elevators have farmers lined up to unload trucks and return to harvesting as quickly as possible. However, on soybeans and wheat, studies have shown that material segregation would not be impossible as long as the elevator has enough viable storage space and a quick method (usually NIR) available to determine key characteristics. Below is an analysis of previous research about the capability of the grain industry to segregate other grains and then soybeans specifically.

There have been many studies done on the feasibility of segregation, some of which have focused on modeling either the capability of elevators to segregate or the cost involved. Sivaraman, Lyford, and Brorsen (2002) focused on an extension of an already-existing grain blending and segregation model created by Hennessy and Wahl (1997). Hennessy and Wahl's model was modified to handle non-linearly separable grain attributes such as protein. Attributes are considered linearly separable when they can be divided easily into groups. Dockage is an example of a linearly separable attribute because one hundred bushels of grain with 1% dockage can be separated into 99 bushels of grain and 1 bushel of dockage. This cannot be done with protein, because it is an internal property of each piece of grain. Nonlinear optimization was used to achieve desired segregation levels. This was applied to sorting hard red winter wheat by protein content. They found that a large elevator could capture most of the "benefits of segregation" (protein premiums paid by the flour mill) with just two bins (for high and low protein), but that this could be cost-prohibitive for smaller elevators (Sivaraman, Lyford, and Brorsen, 2002).

Hurburgh, Neal, McVea, and Baumel (1994) found similar results with an engineering-economic model created to estimate the costs of grain segregation by composition. They looked at data from 50 country elevators in three counties in Iowa to test their hypothesis that "country elevators hold the key to successful differentiation" as these locations experience the maximum variability of quality in grain. According to their model, 50% of the elevators representing 75% of the total available storage capacity would be able segregate grain for less than $0.03/bushel. In contrast to this, elevators representing less than 10% of the total storage capacity (the smallest locations) would have segregation costs of more than $0.04/bushel. They also pointed out that while the larger locations generally had lower costs, small elevators could overcome this if they had multiple dump pits. The majority of costs were related to operations and financial management, which indicate that staff training and optimum facility management would be more beneficial than new storage or handling equipment (Hurburgh, Neal, McVea, & Baumel, 1994). These results contradicted the general industry belief that higher capacity and throughput meant higher costs for differentiation.

Maier and Berruto (2001) analyzed how different grain types (e.g. corn, soybeans, wheat) are received at a country elevator. Because each elevator has different receiving and storage capabilities, the model was individualized for each elevator configuration. The basis of this study was the creation of queuing methods (how trucks line up at the elevator to deliver grain). The more traditional FIFO (first in, first out) model was comparable to a batch delivery, which involved unloading all trucks with the same grain type at the same time to avoid pit, bucket elevator, and conveyor cleanouts and reduce waiting times for farmers. They found that if the elevator was operating at receiving rates of less than 72% capacity, the FIFO method had shorter average waiting times for each farmer. However, if the elevator was operating at receiving rates of more than 72% of capacity (most likely during harvest), the batch method decreased average waiting times by up to 27%. The primary issue in implementing the conclusons of Maier and Berruto (2001) was how to incentivize farmers to accept a new organization scheme that reduced average wait times for most but sometimes increased wait times for a few. Receiving was a key element in grain segregation of all types (e.g. high oil or protein vs. low oil or protein or genetically modified vs. not genetically modified), so determining a case-specific optimal trucking pattern was important to ensure successful and timely differentiation and farmer cooperation (Berruto & Maier, 2001).

Baker, Herrman, and Fairchild (1997) analyzed the grain receiving systems for wheat at twenty country elevators in Kansas. They agreed with previous studies that segregation at the first point of collection was key in the "transition from a commodity-based to a quality-based marketing system". They examined each part of the receiving system (bucket elevators, belt conveyors, screw conveyors, and drag conveyors) to determine the efficiency of each portion. The bucket elevator was most often responsible for bottlenecks, and low operating efficiencies were also caused by inexperienced help, large amounts of dockage and foreign material in the wheat, and slow delivery rates. However, they determined that the capacity of the receiving system is not a major barrier to segregation, and that constraints were more likely caused by personnel or inefficient organization of the receiving driveway. Although quality-testing and grading of wheat only took about two minutes per sample, it was sometimes done incorrectly. This demonstrated a need for more personnel training, because segregation is impossible if the loads are not correctly characterized. Their overall conclusion was that even during a harvest rush of deliveries, segregation of wheat at these country elevators was possible as long as the personnel were qualified and traffic was managed efficiently (Baker, Herrman, & Fairchild, 1997).

Past Studies on Segregation of Soybeans

With soybeans specifically, the current goal is to separate on the basis of protein and oil content into higher-value and lower-value categories. There must be easily identifiable differences on which the elevator can segregate the soybeans. Two studies have demonstrated that these variations do exist in Iowa soybeans. Hurburgh and Brumm (1990) analyzed data collected from one large Iowa grain elevator and eight smaller locations owned by the same company during the 1985, 1986, and 1987 soybean harvests. Three other high-capacity elevators were also included in the 1986 and 1987 data collection for a total of twelve locations. Protein and oil differences between locations were present. Some, but not all, of these differences were consistent from year to year. Of the total variations, 70% of the total variability was due to differences between lots at a single location, while 30% was due to consistent geographic variations. Standard deviations were 1.0% and 0.5% respectively for protein and oil contents of soybeans delivered to individual elevators regardless of the average values.

Hurburgh and Brumm (1990) found that the near-infrared analyzer was capable of operating in an elevator environment, though it did require commitments from the elevator personnel to prepare and grind samples. At the time, NIR units required ground samples. Whole grain analyzers were not available until 1989. Hurburgh and Brumm (1990) agreed with other studies that the composition of the grain must be identified at the country elevator level. They pointed out that, in their study, the potential for soybean segregation was much greater at an individual elevator than at a processing plant or an export terminal buying beans from many country elevators (Hurburgh & Brumm, 1990).

In the 1989-1991 harvest seasons, Hurburgh (1994) measured the protein and oil contents at one Iowa country elevator and put soybean segregation into practice. Data for the 1989 and 1990 harvests was collected with a whole-grain NIR analyzer located in the elevator's office. In 1991 beans were separated into high and low value groups. The SUM of protein and oil components (both in % at a 13% moisture basis) could be used as an effective segregation metric because it was directly correlated to the Estimated Processing Values (EPVs) calculated by the Brumm and Hurburgh (1990) model. The top 23% of incoming loads were separated from the bottom 77% with a 20% error rate. This separation rate was chosen to match available bin space for the high-value beans. This study was able to successfully segregate soybeans into high and low value groupings for 2-3 cents/bushel, but few elevators have adopted the techniques used because soybeans are still traded on yield, not a protein and oil basis. In the Hurburgh (1994) study, no nearby processors were willing to pay more for the high-value soybeans. They contended that all the beans would end up at their facilities anyway because of transportation costs.

In summary, separation of soybeans into high and low value groups is feasible. However, Hurburgh (1994) showed that processors did not need to pay premiums for higher value beans to obtain them – the elevators were already sending beans to them because of their proximity to mitigate transportation costs. The cooperative used for data collection in this study, because it has its own soybean meal plant, does not have to manage external processors like the elevators in the Hurburgh (1994) study. They have the flexibility to experiment with sorting and transportation because any value gained would remain within the cooperative. Therefore, it is to their benefit to evaluate their soybean handling practices.

The cooperative needs to determine whether sorting their beans by value instead of geographically would add value to their processing. Judging by the previous studies reported in this literature review, regional differences between their locations, and thus potential for additional value capture in their processing plant, likely exist.

References

Baker, S., Herrman, T., & Fairchild, F. (1997, January). *Capability of Kansas grain elevators to segregate wheat during harvest* (Report No. 781).

Berruto, R., & Maier, D. E. (2001). Analyzing the receiving operation of different grain types in a single-pit country elevator. *Transactions of the ASAE*, *44*(3), 631-638.

Brumm, T. J., & Hurburgh, C. R., Jr. (n.d.). Project 2339: Estimating the processed value of soybeans. *Iowa Agriculture and Home Economics Experiment Station*, 11-32.

Brumm, T. J., & Hurburgh, C. R., Jr. (1990). Estimating the processed value of soybeans. *Journal of the American Oil Chemists Society*, *67*, 302-307.

Brumm, T. J., & Hurburgh, C. R., Jr. (2006). [Letter to the editor]. *Journal of the American Oil Chemists Society*, *83*, 981-983.

Hindle, P. H. (2008). Historical development. In D. A. Burns & E. W. Ciurczak (Eds.), *Handbook of Near-Infrared Analysis* (3rd ed., pp. 3-6).

The history of U.S. soy [Infographic]. (2017, August 18). Retrieved from https://ussoy.org/wp-content/uploads/2017/08/USSEC-Timeline.pdf

Hurburgh, C. R., Jr. (1994). Identification and segregation of high-value soybeans at a country elevator. *Journal of the American Oil Chemists Society*, *71*, 1073-1078.

Hurburgh, C. R., Jr., & Brumm, T. J. (1990). Protein and oil content of soybeans received at country elevators. *Applied Engineering in Agriculture*, *6*, 65-68.

Hurburgh, C. R., Jr., Neal, J. L., McVea, M. L., & Baumel, P. (1994). *The capability of elevators to segregate grain by intrinsic quality*. Paper presented at the summer meeting of the American Society of Agricultural Engineers.

Johnson, L., & Smith, K. (n.d.). *Soybean processing*. Retrieved from United Soybean Board: Soybean Meal Info Center website: https://www.soymeal.org/wp-content/uploads/2018/04/soybean_processing.pdf

Mills, R. (2017, December 1). NIR technology for routine analysis of food and agricultural products. Retrieved from Foss Analytics website: https://www.fossanalytics.com/en/news-articles/technologies/nir-technology

Shenk, J. S., Workman, J. J., Jr., & Westerhaus, M. O. (2008). Application of NIR spectroscopy to agricultural products. In D. A. Burns & E. W. Ciurczak (Eds.), *Handbook of Near-Infrared Analysis* (3rd ed.).

Sivaraman, E., Lyford, C. P., & Brorsen, B. W. (2002). A general framework for grain blending and segregation. *Journal of Agribusiness*, *20*(2), 155-161.

United Soybean Board. (2015). Soybean composition [Infographic]. Retrieved from https://www.nopa.org/resources/datafacts/soybean-composition/

United Soybean Board. (2017, September 20). Measuring beyond the bushel: Helping farmers know more about their soybeans can help them capture more value. Retrieved from https://unitedsoybean.org/article/measuring-beyond-the-bushel-1

USDA-ERS. (n.d.). Oil crops sector at a glance. Retrieved from United States Department of Agriculture Economics Research Service website: https://www.ers.usda.gov/topics/crops/soybeans-oil-crops/oil-crops-sector-at-a-glance/

US soybean meal: All soybean meal is not created equal. (n.d.). Retrieved from U.S. Soybean Export Council website: https://ussec.org/wp-content/uploads/2015/10/US-Soybean-Meal-Information.pdf

U.S. 2018 soybean crop statistics [Fact sheet]. (2018, December). Retrieved from Soy Meal Info Center website: https://www.soymeal.org/soy-meal-articles/u-s-2018-soybean-crop-statistics/

Wagner, K. P. "A mass balance based model to evaluate the impact of amino acid profiles on the feeding and processed value of soybeans" (2017). *Graduate Theses and Dissertations.* 16233.

CHAPTER 2. STRATEGIES FOR SELECTIVE HANDLING OF SOYBEANS TO MAXIMIZE SOYBEAN MEAL PROTEIN AND YIELD

Bennett Barr and Charles Hurburgh

Modified from a manuscript to be submitted to the Applied Engineering in Agriculture Journal of the American Society of Agricultural and Biological Engineers (ASABE)

Abstract

During the Fall 2018 soybean harvest, soybean samples were collected from 32 country elevator locations belonging to one Iowa-based cooperative which has its own elevator locations and processing plant. This was done to update historical data about geographic variations in protein and oil content of Iowa soybeans, and to assist the cooperative in making more informed decisions about their soybeans to maximize value potential. These samples were analyzed using near-infrared spectroscopy (NIR) to determine protein and oil contents. The data were accumulated and sorted to look for geographic variations in protein and oil content of soybeans throughout Iowa. The data were entered into an Estimated Processing Value (EPV) model to determine value differences between elevator locations. The cooperative sources soybeans for processing from the elevator locations closest to the plant to mitigate trucking costs. They wanted to know whether this strategy was maximizing their net processing value. Results showed that significant variability between locations did exist, which represented a $0.23/bushel EPV spread. Additionally, it was found that 15 samples were needed to accurately represent an elevator location, and that two weeks was a sufficient period to characterize the data to be able to make a marketing decision. Lastly, the soybean quality was not found to vary significantly over the course of harvest, so marketing decisions can made at the beginning of the season.

Introduction

Increasing the profit from soybean farming is a current topic of interest in the agricultural community. This discussion begins with the soybean varieties farmers choose to plant each year, as their choices mark the beginning of the supply chain and affect every user from that point forward. The United Soybean Board (2017) states that farmers often judge varietal selections by yield first, but that “the future of profitability lies in meeting end-user composition needs” because “end users don’t need *soybeans*. They need the protein and oil that come from soybeans” (United Soybean Board, 2017).

These facts of the industry are important for producers growing soybeans, for the grain handlers who both buy soybeans and sell them into the larger grain market, and for the processors who transform them into an assortment of products. However, there is a disconnect in the supply chain because typically the trades between farmers and elevator and elevator and processor are standalone decisions not involving composition. Processors pay based on weight, moisture, and foreign material, not protein and oil content.

The setting for this research is an Iowa-based cooperative group that is unique in that it has producer members, elevators, and its own processing plant. This company purchases beans at its elevators and then either sells them to other processors or transports them to their own facility, which makes soybean meal and soybean oil. They currently allocate their beans geographically, meaning that beans are moved from the closest locations to the processing plant to save on transportation costs. The question of this study is whether or not this strategy was maximizing the net value for processing for the company as a whole. Processors have traditionally believed that localized patterns of protein and oil content do exist, but there is little such data in the public domain (Hurburgh & Brumm, 1990). Table 2.1 summarizes studies that have produced historical data on protein and oil patterns at delivery points.

Table 2.1. Historical data on geographic protein and oil variations.

<table>
<tr><th rowspan="2">Reference</th><th rowspan="2">States/Regions Covered</th><th rowspan="2">Number of Locations Used (or Number of Samples Tested*)</th><th colspan="2">Protein (% at 13% Moisture)</th><th colspan="2">Oil (% at 13% Moisture)</th></tr>
<tr><th>Average</th><th>Standard Deviation of Data (% pts)</th><th>Average</th><th>Standard Deviation of Data (% pts)</th></tr>
<tr><td rowspan="5">Brumm and Hurburgh, 2006 (U.S. Soybean Protein and Oil Survey Data 1994-2004)</td><td>IA, KS, MN, MO, NE, ND, SD</td><td>7963*</td><td>34.9%</td><td>0.8%</td><td>18.6%</td><td>0.5%</td></tr>
<tr><td>IL, IN, MI, OH, WI</td><td>6722*</td><td>35.7%</td><td>0.7%</td><td>18.6%</td><td>0.7%</td></tr>
<tr><td>AR, KY, LA, MS, OK, TN, TX</td><td>1535*</td><td>35.9%</td><td>0.9%</td><td>18.9%</td><td>0.7%</td></tr>
<tr><td>AL, FL, GA, NC, SC</td><td>274*</td><td>36.4%</td><td>1.1%</td><td>18.8%</td><td>0.8%</td></tr>
<tr><td>DE, MD, NJ, NY, PA, VA</td><td>274*</td><td>36.4%</td><td>1.1%</td><td>18.8%</td><td>0.8%</td></tr>
<tr><td rowspan="3">Hurburgh, 1994</td><td rowspan="3">IA</td><td>1 (1989)</td><td>34.9%</td><td>1.0%</td><td>18.5%</td><td>0.7%</td></tr>
<tr><td>1 (1990)</td><td>35.6%</td><td>0.9%</td><td>19.1%</td><td>0.6%</td></tr>
<tr><td>1 (1991)</td><td>35.3%</td><td>1.2%</td><td>18.4%</td><td>0.7%</td></tr>
<tr><td rowspan="3">Hurburgh and Brumm, 1990</td><td rowspan="3">IA</td><td>9 (1985)</td><td>33.8%</td><td rowspan="3">1.0%</td><td>19.8%</td><td rowspan="3">0.5%</td></tr>
<tr><td>12 (1986)</td><td>35.3%</td><td>18.6%</td></tr>
<tr><td>12 (1987)</td><td>34.5%</td><td>18.9%</td></tr>
<tr><td rowspan="2">Hurburgh, Paynter, and Schmitt, 1987</td><td rowspan="2">IA, IL, OH, MN</td><td>11 (1983)</td><td>33.9%</td><td>1.0%</td><td>19.7%</td><td>0.5%</td></tr>
<tr><td>11 (1984)</td><td>34.2%</td><td>1.0%</td><td>19.1%</td><td>0.5%</td></tr>
<tr><td colspan="3">Average Values (% at 13% moisture)</td><td>35.1%</td><td>1.0%</td><td>18.9%</td><td>0.6%</td></tr>
</table>

The cooperative wanted to update and expand this information to cover their locations, then assess the economic potential of identifying beans for the plant by value rather than distance. This paper addresses the present results of that study, which began during the Fall 2018 harvest. This paper will focus primarily on the methods of sample collection, analysis of samples using NIR technology, and patterns observed in the first data set.

Objectives

The objectives of this project were to:

1. Determine the protein and oil content of soybeans received by an Iowa grain cooperative at multiple receiving locations owned by the company.
2. Identify potential high protein, high oil, and high processing value locations.
3. Estimate whether preferentially sourcing soybeans by some combination of protein and oil values could be a feasible economic decision.

Materials and Methods

Project Development

Location Identification

The cooperative identified 32 locations (country elevators) as testing sites for determination of protein and oil content of soybeans delivered to their facilities. These were locations within reasonable trucking distance (< 50 miles) of the plant. Of these 32, 3 of the larger-volume locations were designated as near-infrared spectroscopy (NIR) testing centers because each had an NIR analyzer and sufficient office space to operate it. The other 29 were designated as "tributary" locations. To protect the confidentiality of the cooperative, the locations were renamed – the 3 testing centers are labeled as locations A*, B*, and C* in this paper, and the tributaries are labeled with their corresponding main location letter and then a number (e.g. the tributaries of location A are called A1, A2, A3, etc.). Locations A* and B* each had 13 tributary locations, and location C* (a much smaller facility) had 3.

Sample Collection

The sample collection procedures differed between testing locations A*, B*, and C* and their corresponding tributaries. A*, B*, and C* were asked to test each load coming into their respective facilities as well as the composite samples collected by their tributaries. The tributary locations were asked to collect two 1000 g composite samples per day, one in the morning and one in the afternoon. Each composite sample was then delivered to the corresponding NIR testing location to be analyzed (i.e. locations A1-A13 sent their composite samples to A* to be tested and similarly for B* and C*). Therefore, the tributary samples were themselves mixtures of loads.

NIR Analysis

Infratec 1241 near infrared transmission analyzers (Foss North America, Eden Prairie, MN) were provided to the A*, B*, and C* locations. These units can transmit data to a laptop in Excel, are user-friendly, and take the same size sample (~500 g) that is already used for the capacitance moisture meter. The three analyzers were loaned to the locations from Iowa State University and employee training was provided. These machines had been calibrated for soybean composition analysis before delivery by the Iowa Grain Quality Lab at Iowa State University. The NIR technology offered a quick method of finding protein and oil levels. The time required to run a sample was approximately 2.5 minutes from start to finish.

Oil, protein and fiber concentrations were determined in whole grain samples by near-infrared spectroscopy using calibrations developed at the Iowa State University Grain Quality Laboratory. The calibrations apply to the Foss Infratec transmission analyzers (Foss North America, 7682 Executive Drive, Eden Prairie, MN 55344, www.fossnorthamerica.com). The Iowa State calibration process was described by Rippke et. al. (1995) and was subsequently the basis for the standard method of the American Association of Cereal Chemistry (AACC, 1999). The present calibrations are based the Artificial Neural Network (ANN) algorithm as adapted for Infratec analyzers by Foss (Buchmann, et. al., 2001). The last crop year represented in the calibration data set is 2013 (Hurburgh, 2015). The validation data for the 2014 and 2015 crop years are displayed in Table 2.2.

Table 2.2. Infratec validation data for two years after calibration creation. SEP = Standard Error of Prediction. RPD = Standard Deviation of Data/SEP

	2014 Crop (n = 91 samples used)		2015 Crop (n = 96 samples)	
Factor	RPD	SEP	RPD	SEP
Moisture	10.1	0.28	17.0	0.26
Protein	5.6	0.55	6.4	0.55
Oil	5.3	0.40	3.6	0.51
Fiber	5.0	0.10	3.2	0.09

Sample Testing

The three main locations tested as many inbound loads as possible. The staff at each location were already testing samples with a moisture meter for each load. The samples were tested with the NIR instrument after the moisture test. Location staff were responsible for entering an ID number for each sample, then completing the test cycle. The data were electronically transferred to Excel spreadsheets on attached computers. Files were downloaded weekly from locations A*, B*, and C*. ISU personnel reviewed the data and corrected obvious errors, identifications, and other issues.

Composite samples were collected twice a day at each tributary location by collecting all of the morning moisture meter samples in a bucket, then taking a 1,000-gram sample from that composite. This was then repeated for the afternoon. The resulting two samples per day were sent to the corresponding testing location. The number of trucks in each composite was not recorded but estimated to be between 10 and 25, depending on receiving rates for the day.

Verification of NIR Analyzers

Weekly randomly selected samples from A*, B*, and C* were transported to the Grain Quality Lab at Iowa State University for testing on their laboratory NIR analyzers. The same sample was tested on the A*, B*, C*, and ISU analyzers and the results were compared to make sure the three NIR units at cooperative locations were reading consistently with the Grain Lab and with each other. These were a mix of individual truck samples and tributary composites.

Estimated Processing Value (EPV) Calculation

All EPV calculations were done using the SPROC 3.0 soybean processing model in Microsoft Excel (Wagner, Hurburgh, and Brumm, 2017). Figure 2.1 from Wagner (2017) displays a typical solvent extraction flow diagram, which was adapted to become an expeller-press scenario in Step 1 below. Figure 2.2 from Wagner (2017) shows the user input sheet of the SPROC 3.0 model.

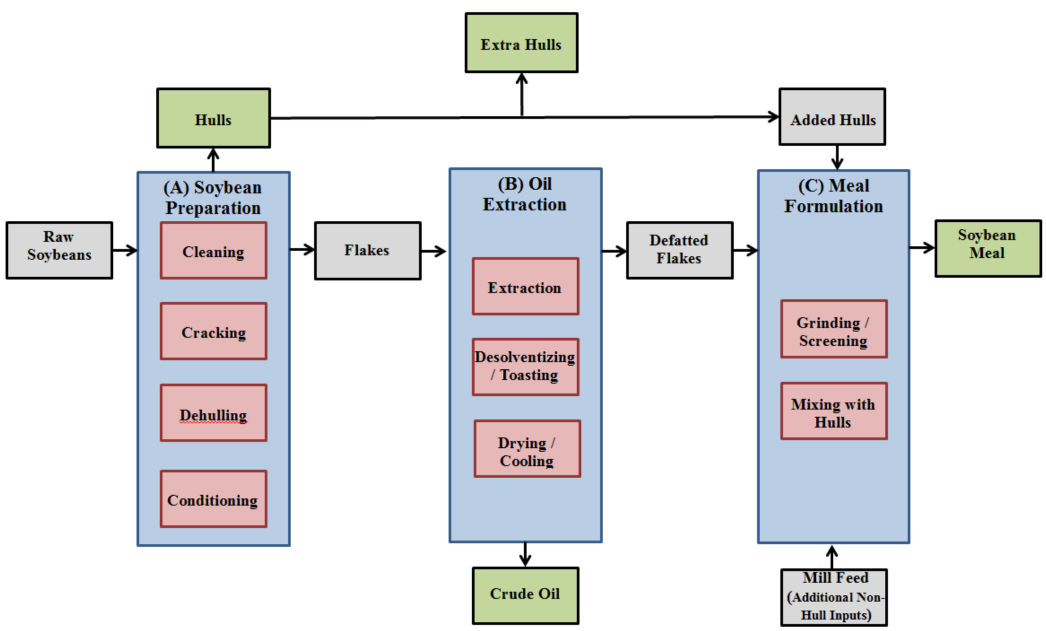

Figure 2.1. Soybean solvent extraction flow diagram.

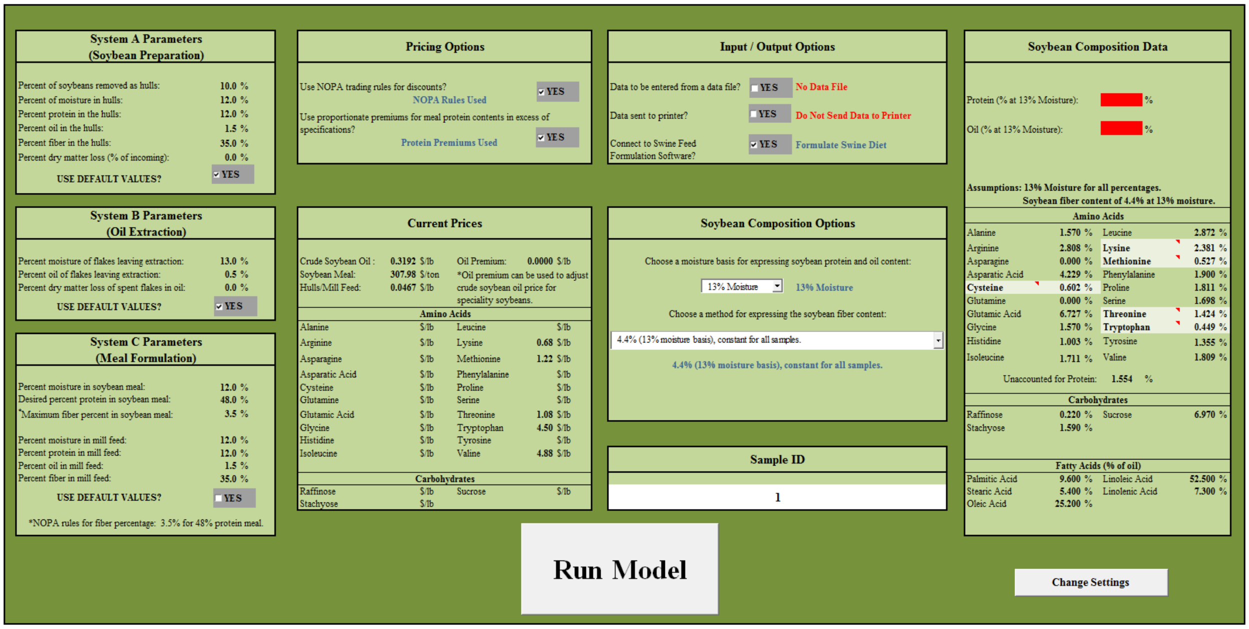

Figure 2.2. SPROC 3.0 user inputs sheet.

Table 2.2 displays the inputs used for the model. All prices were the current prices at the market close on December 12th 2018. The following parameters were used for all model iterations.

1. The soybean preparation section (A Parameters in Figure 2.1, System A Parameters in Figure 2.2) was essentially turned off, because no hulls were removed in the cooperative's specific process. The model was then running an expeller-press scenario instead of a solvent-extraction scenario. All the weight of incoming beans that was not extracted as oil or moisture loss was assumed to be meal.
2. No National Oilseed Processors' Association (NOPA) trading rules were used for meal protein discounts, and premiums were not used for meal proteins in excess of the user specifications. The expeller meal is sold at whatever protein content is created.
3. Amino acid prices were not taken into account as they were not necessary for the meal and

oil valuations. This meal is used for dairy cows.

Table 2.3. SPROC processing parameter assumptions.

System A Parameters (Soybean Preparation)	% soybeans removed as hulls	0.0%
	% moisture in hulls	N/A[a]
	% protein in hulls	N/A[b]
	% oil in hulls	N/A[c]
	% fiber in hulls	N/A[d]
	% dry matter loss (% of incoming)	N/A[e]
System B Parameters (Oil Extraction)	% moisture of flakes leaving extraction.	13.0%
	% oil of flakes leaving extraction	6.0%
	% dry matter loss of spent flakes in oil	1.0%
System C Parameters (Meal Formulation)	% moisture in soybean meal	12.0%
	Desired % protein in soybean meal	40.0%
	Maximum fiber % in soybean meal	3.5%
	% moisture in mill feed	N/A[f]
	% protein in mill feed	N/A[g]
	% oil in mill feed	N/A[h]
	% fiber in mill feed	N/A[i]
Pricing Options	Use NOPA trading rules for discounts?	NO
	Use proportionate premiums for meal protein contents in excess of specifications?	NO
Current Prices	Crude soybean oil	0.2880 $/lb
	Soybean meal	310.80 $/ton
	Hulls/mill feed	N/A[j]

[a-j] These categories were deactivated in the model by specifying 0.0% of weight removed as hulls and therefore not adding any hulls back as mill feed.

Use of SUM as a Rapid Estimator of Value

Protein and oil levels in soybeans are closely tied to their EPVs because the relative values of protein and oil tend to be proportional to the price combinations for soybean meal and oil. In his earlier study, Hurburgh (1994) tested using the sum of protein and oil components (SUM) as a ranking criterion. Linear regression equations between SUM and EPV for 1989, 1990, and 1991 harvests had R^2 values of 97.8, 87.8, and 99.6 respectively (Hurburgh, 1994). Because the relationship between SUM and EPV is linear, a dollar value can be estimated for every 1%-point increase or decrease in SUM. Therefore, SUM values can provide quick valuation estimates for making decisions when time is limited or if a full EPV analysis is not feasible.

Results and Discussion

The overall collection statistics for the data collected during the Fall 2018 soybean harvest are listed below in Table 2.3.

Table 2.4. Overall data collection results by location.

	A*	A Tributaries (13)	B*	B Tributaries (13)	C*	C Tributaries (3)
No. of Samples Expected[a]	856	1144	679	1196	1403	264
No. of Samples Collected	164	275	377	196	1292	65
% of Expected Samples Collected	19.1%	24.0%	55.5%	16.4%	92.1%	24.6%

[a] This was determined for the testing locations with predictions from the cooperative about the number of loads to be delivered. For the tributary locations, this was calculated after harvest had finished by taking the actual number of days of harvest for each tributary location, multiplying by two (for one morning and one afternoon sample expected) and adding up the resulting totals.

Any location that reported less than 15 samples was removed from further analysis, as it was not considered to be properly representative. This was decided by graphing the number of samples vs. the standard error of the mean ($\frac{SD}{\sqrt{\# \: of \: samples}}$) for the A and B tributaries (Figure 2.3). The figure shows that that the standard error of the mean levels off after ~15 samples.

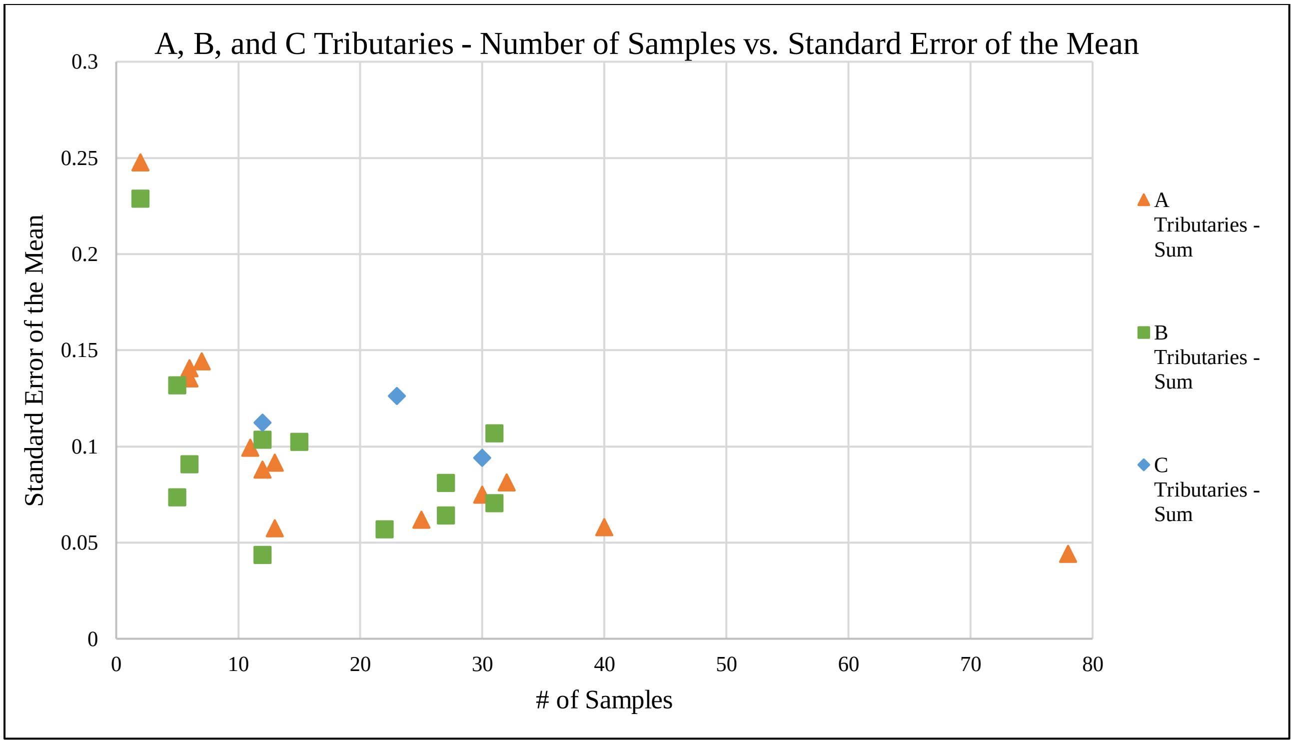

Figure 2.3. Justification for use of only locations with >15 samples in analysis.

Tables 2.5 - 2.7 give a summary of the % points protein + % points oil (SUM) averages for each testing location and its tributaries. All protein, oil, and SUM values are reported on a % at 13% moisture basis. Locations reporting less than 15 samples are listed in the table footnotes. These data satisfy Objective 1 of the project to determine the protein and oil content of soybeans at the elevator locations.

Table 2.5. Protein and oil SUM averages by location - A* and tributaries.

Location	% protein + % oil = SUM (% at 13% moisture basis)	Standard Deviation of the Data (% points)		
		Protein	Oil	SUM
A*	31.97 + 20.23 = 52.20	0.99	0.43	0.78
A5	32.41 + 20.08 = 52.49	0.52	0.33	0.37
A8	33.38 + 19.53 = 52.91	0.58	0.30	0.39
A9	32.81 + 19.96 = 52.76	0.56	0.25	0.41
A12	33.83 + 19.46 = 53.29	0.47	0.32	0.31
A13	33.51 + 19.50 = 53.01	0.54	0.27	0.46

[a] Locations A1, A2, A3, A4, A6, A7, A10, and A11 reported less than 15 samples.

Table 2.6. Protein and oil SUM averages by location - B* and tributaries.

Location	% protein + % oil = SUM (% at 13% moisture basis)	Standard Deviation of the Data (% points)		
		Protein	Oil	SUM
B*	32.33 + 20.34 = 52.67	1.16	0.76	0.80
B1	33.22 + 19.84 = 53.06	0.34	0.17	0.33
B2	33.23 + 20.17 = 53.40	0.61	0.41	0.40
B7	32.97 + 19.95 = 52.92	0.53	0.38	0.33
B8	33.76 + 19.56 = 53.31	0.47	0.28	0.39
B10	32.80 + 20.09 = 52.88	0.63	0.44	0.42
B13	33.47 + 19.96 = 53.43	0.93	0.66	0.59

[a] Locations B3, B4, B5, B6, B9, B11, and B12 reported less than 15 samples.

Table 2.7. Protein and oil SUM averages by location - C* and tributaries.

Location	% protein + % oil = SUM (% at 13% moisture basis)	Standard Deviation of the Data (% points)		
		Protein	Oil	SUM
C*	33.10 + 20.15 = 53.24	1.31	0.93	0.72
C1	32.58 + 20.18 = 52.76	1.08	0.66	0.60
C3	32.52 + 20.05 = 52.57	0.92	0.53	0.52

[a] Location C2 reported less than 15 samples.

The averages of the weekly verification samples from the check dates are presented in Table 2.8.

Table 2.8. Weekly verification samples from three dates throughout harvest.

Date/Analyzer Location	Protein				Oil			
	ISU Grain Lab	*A**	*B**	*C**	*ISU Grain Lab*	*A**	*B**	*C**
9/24/18	33.2	33.2	33.2	32.9	19.9	19.9	20.0	20.0
10/12/18	33.3	33.0	33.0	33.3	19.7	19.7	19.7	19.6
11/2/18	33.0	33.0	33.0	33.2	19.8	19.6	19.8	19.6

An ANOVA analysis was performed on three dates (9/24, 10/12, and 11/2) from the weekly verification samples to ensure that the analyzers were reading within expected variance levels of each other and to the reference analyzer at the ISU Grain Lab. The ANOVA analysis produced p-values of 0.7138, 0.7669, and 0.5551 for the 9/24, 10/12, and 11/2 samples respectively. None of these values were statistically significant ($p < 0.05$), which showed that the analyzers were all producing accurate results (with respect to the reference unit at the ISU Grain Lab) throughout the course of the experiment.

After characterization of the data was complete and check samples were verified, the data were combined and sorted by SUM values from low to high. 10 data points were randomly selected from this spread and the values were entered into the SPROC model as described in the "Materials and Methods" section. The resulting EPVs (Estimated Processed Values) are shown in Figure 2.2 below with a corresponding trend line. This line is specific to the price combinations on December 12th, 2018.

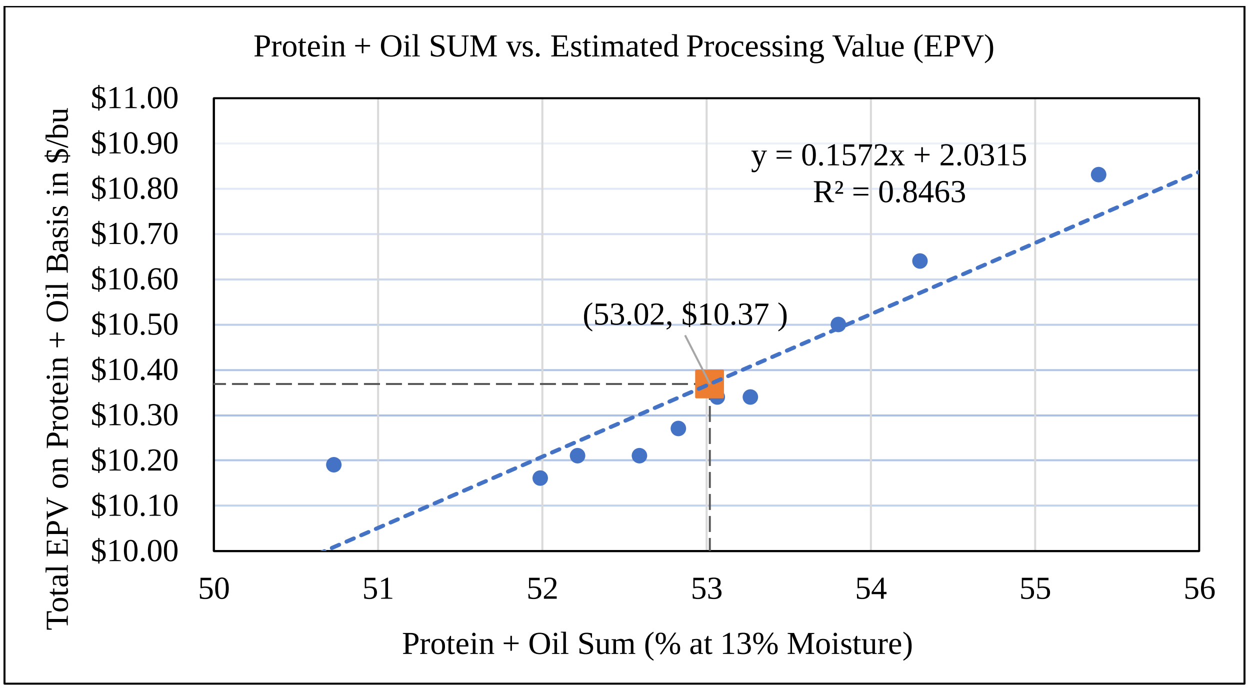

Figure 2.4. SUM vs. Estimated Processing Value (EPV) for 10 randomly selected data points.

This analysis was performed to determine the approximate economic worth of a change in protein + oil sum. The slope of the trend line indicates that for every percentage point increase or decrease in protein + oil sum, the EPV per bushel will accordingly increase or decrease by

$0.1572, or about 16 cents per bushel per 1% point increase or decrease of SUM. This is applicable for the given combination of prices. For example, a bushel from soybeans with a SUM value of 54% would be worth 16 cents more than a bushel from soybeans with a SUM value of 53%. The square point represents the average sum and average EPV values. The cooperative was pulling beans into their processing plant based on geographic location (i.e. the beans grown or brought to locations closest to the plant were processed). However, Figure 2.2 shows that there is value in sorting based on bean characteristics such as protein or oil content and not just proximity. Even with the rough weather and a sporadic harvest during the 2018 growing season, clear value differences were evident between locations, and many of the lower-value locations were those closest to the processing plant. If the cooperative chooses to sort their beans in the future, a more typical harvest could provide further future improved profit margins. The economic impact of future substandard harvests such as 2018, which had a lower protein average than normal (~33% as opposed to ~35%), could be lessened.

The average protein and oil values for each location were entered into the SPROC model and graphed. These results can be seen in Figure 2.5 , which shows the EPVs by location.

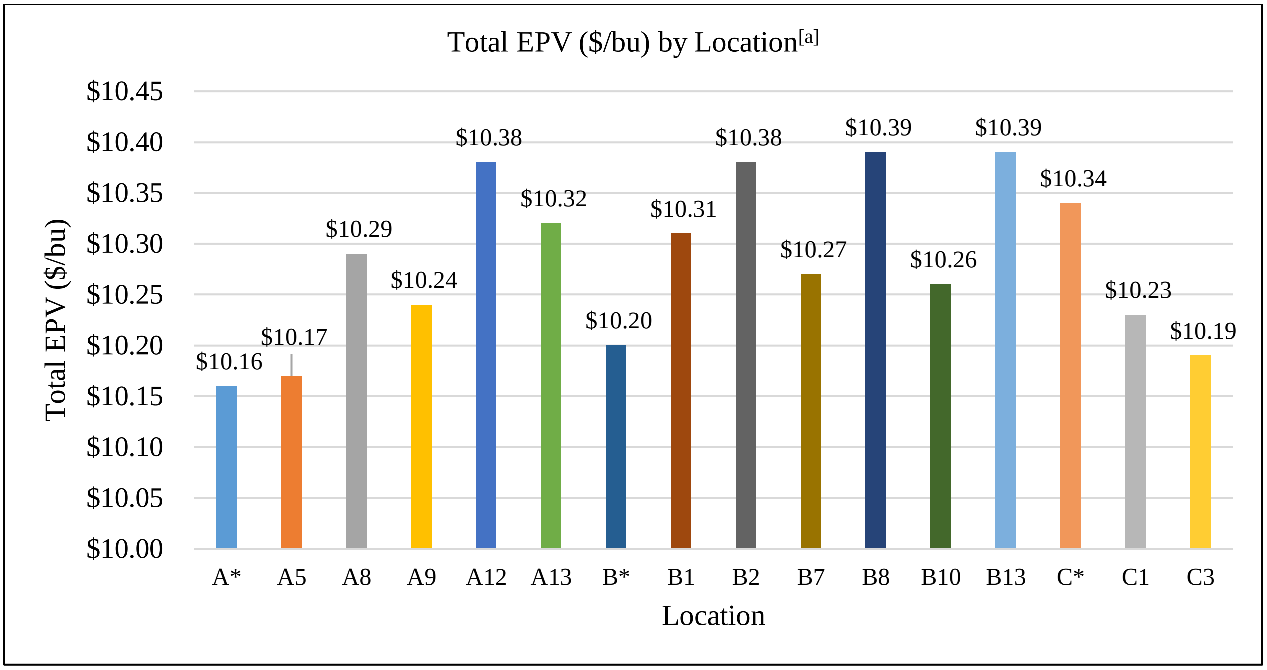

Figure 2.5. EPVs in $/bushel by location for locations with 15 or more samples.

[a] All EPV calculations were performed using the prices of 0.2880 $/lb for crude soybean oil and 310.80 $/ton for soybean meal.

The information in Figure 2.5 builds on the conclusions from Figure 2.4 in terms of separating and sourcing beans based on value and not purely on transportation costs. The data in Figure 2.5 represents an EPV spread of $0.23/bushel (from $10.16/bushel to $10.39/bushel), which indicates that the cooperative may not be maximizing their revenue. Some of the highest-value beans are not the ones being routed to the processing plant, though transportation costs should be included for a final decision. The locations in the graph are significantly different, which was confirmed by an ANOVA analysis of the data. The daily average SUM values throughout harvest of all locations graphed above were compared and produced a p-value of 2.2×10^{-16}, which indicated a significant ($p < 0.05$) difference between the locations.

The daily average SUM values also align with historical soybean reports in Iowa. This alignment was confirmed by performing a comparison between the standard deviations at the A*, B*, and C* locations gathered in this study and the standard deviation values from Hurburgh's 1994 study at country elevators from 1989-1991.. The elevators studied in 1994 were among the same ones involved in this study. The standard deviations were not significantly different ($p < 0.05$) from those measured in the 1994 study (Hurburgh, 1994). The 2018 data do not have any more sample-to-sample variation than the 1989-1991 data. This shows that although protein and oil averages change, the variation of producer-delivered beans for a given year has not changed substantially across a nearly 30-year period.

These data indicate localized geographic differences in soybean protein and oil, but it does not provide potential causes for these differences. Possible explanations include, but are not limited to, differences in soil types, weather patterns, soybean varieties grown, and individual farming practices. This should be further investigated to improve information for farmers about varietal selection and farming methods. From the cooperative's perspective, the ability to plan logistics based on soybean properties for a specific crop year is the goal.

The cooperative also wanted an estimate on the number of samples needed from each load to ensure a representation sample and how long during the harvest season this number of samples was needed. Testing the numbers of loads collected for this data set required significant time input from the location staff. The cooperative was concerned about a) keeping wait times short for farmers and b) being able to make a valid marketing decision as early into harvest as possible. They wanted data to confirm that the bean quality at a given location varied minimally across harvest. The data were normalized by subtracting the respective means of SUM for A*, B*, and C* from each data point for A*, B*, and C*. These differences were averaged by date and then graphed (Figure 2.6) against the numerical day of harvest on which the sample was tested.

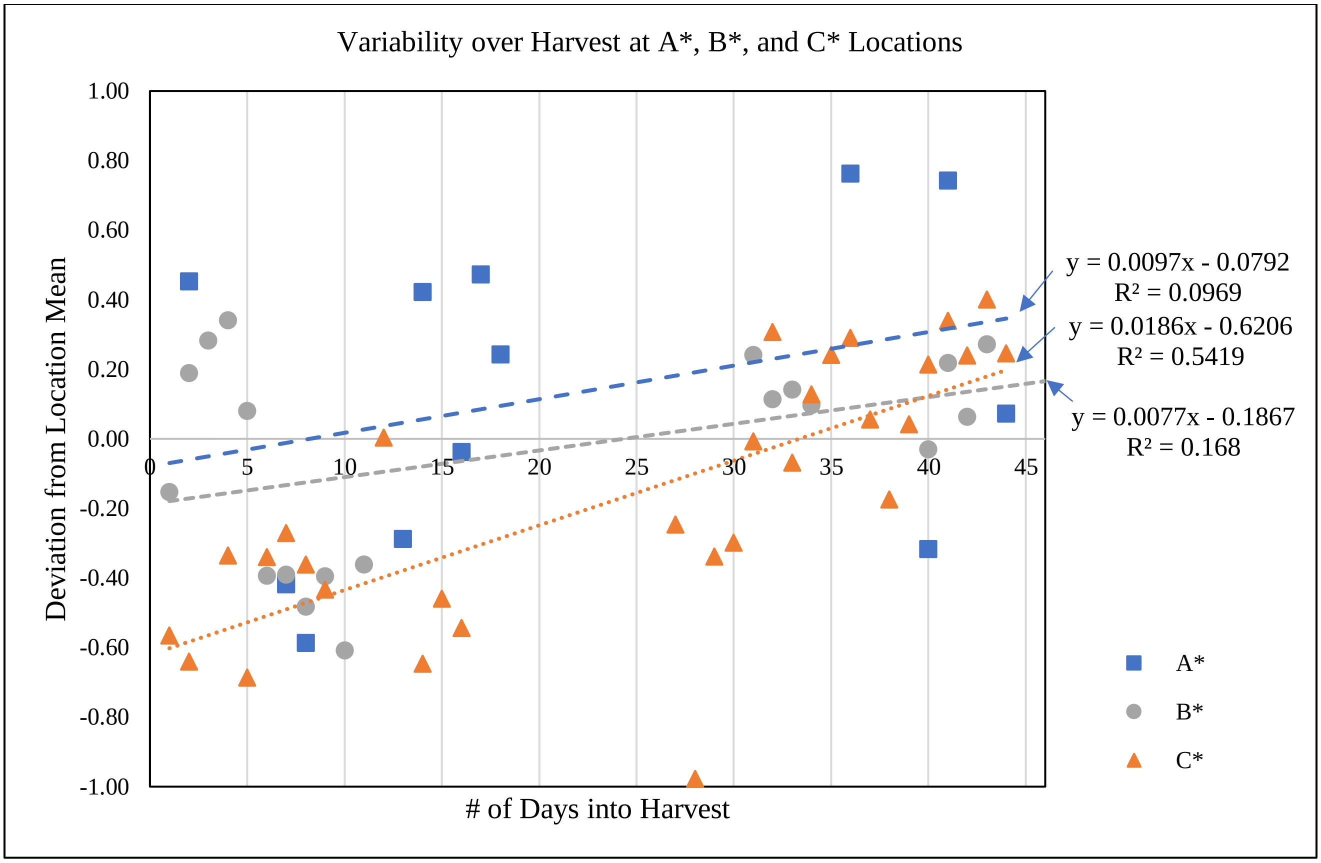

Figure 2.6. Variability over harvest at NIR testing locations.

As shown in Figure 2.6, all daily data points were contained within a ±1% point range of protein + oil sum relative to the location means. The linear regression lines show negligible slopes (0.0097 and 0.0077 respectively) for the A* and B* locations, meaning there were no noticeable upward or downward trends over the length of harvest. The slope for the C* regression line was somewhat stronger at 0.0186 with a higher R^2 value. There is a large gap in the middle of the graph – this was due to a two-to-three-week rainy period where harvest stopped and no samples were taken. Figure 2.6 confirms that bean quality is reasonably consistent throughout harvest, and as such it is not necessary to measure every load that comes into a facility.

Past studies (Hurburgh, 1994) have indicated that two weeks of sampling was sufficient to characterize the data, which also held true in this experiment. Additionally, a total count of 15 samples was found to be sufficient to accurately represent the location. These values could be used to develop a sampling protocol for the locations – for example, two samples a day for two weeks would give enough data to accurately represent the locations. This would help the cooperative determine segregation and selective handling logistics as quickly as possible for the year, though the slight trend at the C* location suggests that there should be an ongoing evaluation protocol to update or verify the estimates. Furthermore, if the relative positions of the locations were constant over harvest (even if the absolute values were not), the correct decisions would be made.

Conclusions

There are five main conclusions that can be drawn from this study.

1. It is feasible to use an NIR instrument to measure protein and oil contents of soybean loads delivered to country elevators, even during peak harvest season.
2. Significant localized geographic differences in protein and oil content exist between the cooperative's locations. High and low-value locations from the 2018 harvest were identified. The cooperative can do further analysis with transportation costs to determine the economic feasibility of sourcing beans preferentially from the high-value locations.
3. Soybean loads do not vary significantly over the harvest period – bean quality at each location was consistent from the beginning to the end of harvest.
4. The variations of the 2018 data were not significantly different from the variations in previous studies. Standard deviations of protein and oil contents remained at approximately 1 and 0.5 percentage points respectively.
5. 15 samples were sufficient to representatively characterize each location.

Acknowledgements

The authors are grateful to Tim Sullivan for his guidance, Glen Rippke and Alan Kuuttila from ISU, the cooperative staff involved in testing for their patience and cooperation, and Wyatt Runksmeier for assistance in troubleshooting and collection of data from locations.

References

Brumm, T. J., & Hurburgh, C. R., Jr. (1990). Estimating the processed value of soybeans. *Journal of the American Oil Chemists Society*, *67*, 302-307.

Brumm, T. J., & Hurburgh, C. R., Jr. (2006). [Letter to the editor]. *Journal of the American Oil Chemists Society*, *83*, 981-983.

Hurburgh, C. R., Jr. (1994). Identification and segregation of high-value soybeans at a country elevator. *Journal of the American Oil Chemists Society*, *71*, 1073-1078.

Hurburgh, C. R., Jr. (2015). *Iowa State University Grain Quality Laboratory calibration descriptions with citations: Soybeans and corn, Infratec 12xx series, calibrations 20130x*.

Hurburgh, C. R., Jr., & Brumm, T. J. (1990). Protein and oil content of soybeans received at country elevators. *Applied Engineering in Agriculture*, *6*, 65-68.

Hurburgh, C. R., Jr., Paynter, L. N., & Schmitt, S. G. (1987). Quality characteristics of Midwestern soybeans. *Applied Engineering in Agriculture*, *3*, 159-165.

Measuring beyond the bushel: Helping farmers know more about their soybeans can help them capture more value. (2017). Retrieved from https://unitedsoybean.org/article/measuring-beyond-the-bushel-1

Wagner, K., Hurburgh, C. R., Jr., & Brumm, T. J. (2017). SPROC 3.0 [Computer program].

Wagner, K. P. "A mass balance based model to evaluate the impact of amino acid profiles on the feeding and processed value of soybeans" (2017). *Graduate Theses and Dissertations.* 16233.

CHAPTER 3. VARIABILITY ANALYSIS OF SOYBEANS DELIVERED TO IOWA ELEVATORS TO DETERMINE LOCATION RANKING FEASIBILITY

Bennett Barr and Charles Hurburgh

Modified from a manuscript to be submitted to the Applied Engineering in Agriculture Journal of the American Society of Agricultural and Biological Engineers (ASABE)

Abstract

Soybean processors have long believed that there are geographical differences in protein and oil content, and that careful source identification would offer opportunities for greater value capture. In order to identify high and low value areas, soybean sample data from grain elevators throughout Iowa were analyzed with near-infrared spectroscopy (NIR) to determine protein and oil contents. Geographic and economic variability between the elevator locations were found. An error analysis was performed to find the effects of potential error on location separation, because errors would reduce the certainty of any marketing decisions based on measured value differences. Both random and systematic errors were possible with the use of NIR analyzers. Random errors were simulated using an Excel-based model that created random values with a specified standard deviation and mean, which were then added to the original data points. This simulation was performed for three test cases – one with typical standard deviations for protein and oil contents, one with higher-than-average standard deviations, and one with typical standard deviations but with a bias element added to a subset of the locations. The introduction of random error made value gaps between locations smaller, which made discrimination of high-value locations from average or low-value locations difficult. These results showed the importance of having standards for measuring instruments if the soybean supply chain is ever to move to a

protein and oil pricing basis, because one of the largest sources of error in a commodity-based market system is inconsistency of measuring units with each other.

Introduction

Protein and oil contents of soybeans determine the potential value of the products made in soybean processing, yet soybeans are not marketed based on this information (Hurburgh, 1994). Processors have long believed that there are geographical differences in protein and oil content, and that careful source identification would offer opportunities for greater value capture. Concerns about additional operating and transportation costs had been the limiting factor to development of more managed soybean supply chains. Several studies have been done to estimate any potential value available for capture in terms of high protein or oil contents and are summarized in Table 3.1.

Table 3.1. Historical data on geographic protein and oil variability.

Reference	States Covered	Number of Samples Used	Protein (% points at 13% Moisture basis)		Oil (% points at 13% Moisture basis)	
			Average Value	Standard Deviation of the Data	Average Value	Standard Deviation of the Data
Brumm and Hurburgh 2006 (U.S. Soybean Protein and Oil Survey Data 1994-2004)	IA, KS, MN, MO, NE, ND, SD	7963	34.9	0.8	18.6	0.5
	IL, IN, MI, OH, WI	6722	35.7	0.7	18.6	0.7
	AR, KY, LA, MS, OK, TN, TX	1535	35.9	0.9	18.9	0.7
	AL, FL, GA, NC, SC	274	36.4	1.1	18.8	0.8
	DE, MD, NJ, NY, PA, VA	274	36.4	1.1	18.8	0.8

Table 3.1, continued, with modification from “Number of Samples Used” to “Number of Locations Used”.

Reference	States Covered	Number of Locations Used	Protein (% points at 13% Moisture basis)		Oil (% points at 13% Moisture basis)	
			Average Value	Standard Deviation of the Data	Average Value	Standard Deviation of the Data
Hurburgh, 1994	IA	1 (1989)	34.9	1.0	18.5	0.7
		1 (1990)	35.6	0.9	19.1	0.6
		1 (1991)	35.3	1.2	18.4	0.7
Hurburgh and Brumm 1990	IA	9 (1985)	33.8	1.0	19.8	0.5
		12 (1986)	35.3		18.6	
		12 (1987)	34.5		18.9	
Hurburgh, Paynter, and Schmitt 1987	IA, IL, OH, MN	11 (1983)	33.9	1.0	19.7	0.5
		11 (1984)	34.2	1.0	19.1	0.5
Barr and Hurburgh 2019	IA	32	32.8	1.2	20.1	0.7

These data represent historical patterns of soybean protein and oil contents across the country. However, the last study published was in 1994, and updated data were needed to make sure those patterns had not changed. The most recent analysis of regional differences in protein and oil contents is detailed in Barr and Hurburgh (2019). Data from 32 country elevators showed that some locations received significantly higher-value soybeans than others. The marketplace does not currently consider protein and oil, and measurement error is often cited as the primary reason for not testing. In concept, measurement variability would reduce the certainty of decisions based on measured value differences. However, Barr and Hurburgh’s 2019 study did not analyze the effect of potential errors on the location rankings, which is an important consideration in making sure that locations are actually statistically different enough to make segregation worth the cost and effort.

In Barr and Hurburgh (2019), 32 elevator locations were identified by the cooperative as testing sites because of their proximity to its processing plant. Three of the larger-volume locations were designated as near-infrared spectroscopy (NIR) testing centers because each had an NIR analyzer and sufficient office space to operate it. The other 29 were designated as "tributary" locations. The 3 testing centers are labeled as locations A*, B*, and C* in this paper, and the tributaries are labeled with their corresponding main location letter and then a number (e.g. the tributaries of location A are called A1, A2, A3, etc.). Locations A* and B* each had 13 tributary locations, and location C* (a much smaller facility) had 3. A*, B*, and C* were asked to test each load coming into their respective facilities as well as the composite samples collected by their tributaries. The tributary locations were asked to collect two 1000-gram composite samples per day, one in the morning and one in the afternoon. Each composite sample was then delivered to the corresponding NIR testing location to be analyzed (i.e. locations A1-A13 sent their composite samples to A* to be tested and similarly for B* and C*).

In this study, the data from Barr and Hurburgh (2019) were used to rank locations by SUM values. The SUM value is the addition of the protein component and the oil component, both expressed on a percentage of points at 13% moisture basis. SUM values in soybeans are closely tied to their EPVs because the relative values of protein and oil tend to be proportional to the price combinations for soybean meal and oil. In his previous study, Hurburgh (1994) tested the validity of using the SUM as a ranking criterion. Linear regression equations between SUM and EPV for 1989, 1990, and 1991 harvests had R^2 values of 97.8, 87.8, and 99.6 respectively (Hurburgh, 1994). Because the relationship between SUM and EPV is linear, a dollar value can be estimated for every 1%-point increase or decrease in SUM. Barr and Hurburgh found this to be $0.16 per bushel per % point of SUM change with the price combination used in the study (i.e. a bushel from soybeans with a SUM value of 54% would be worth 16 cents more than a bushel from soybeans with a SUM value of 53%). SUM values can provide quick valuation estimates for making decisions when time is limited or if a full EPV analysis is not feasible. Because SUM and EPV are so closely tied, any error in either component measurement created error in the soybean valuation.

Both systematic errors (which represent a consistent error from the true value, either too high or too low) and random errors (which represent inconsistent errors in both directions, high and low, from the true value) are possible with the use of near-infrared analyzers (which are an indirect measurement technique) to determine protein and oil contents. Possible sources of error with these instruments included:

1. Systematic:
    a. Consistent bias in the measuring instrument (i.e. one analyzer reading 0.25% too high on each measurement) would cause a shift in all of a location's (including its tributaries) measurements.
    b. Variances in instrument standardization would cause instruments to consistently read differently between the three testing locations.
2. Random:
    a. User error: analyzers were run by cooperative personnel who had not participated in any similar experiments besides a short training session, and a few different employees operated the analyzer at each testing location. Any

personnel differences in NIR operating technique would cause a user operating bias inconsistency.

b. Environmental effects: temperature fluctuations would be different each day in the scale houses where the analyzers were housed and thus affect each measurement to a different extent.

c. Equipment deterioration: power inconsistencies causing dimming or brightening of the internal light source in the NIR would affect each measurement differently.

All of these sources of error combine to the uncertainty of the individual tests performed in the study, which then affects the SUM calculation.

Objective

The specific objective of this research was to evaluate the impact of measurement errors on the component value ranking of soybean samples and harvest locations in a typical trade area of a soybean processing plant. This was accomplished through the following steps:

1. Simulate random and systemic errors to NIRS-predicted soybean composition data collected at Iowa grain elevators in the fall of 2018.
2. Determine the impact of simulated errors on the protein and oil SUM ranking of soybeans from the individual elevator locations.
3. Identify the key sources of error in the NIRS testing program as it was performed at the elevators.

The larger meaning of this objective is to determine the degree to which analytic errors should and can be controlled if component value marketing is to be successful.

Materials and Methods

Data Source

All data were collected during the Fall 2018 soybean harvest from multiple grain elevator locations of an Iowa grain cooperative. This firm also owns a soybean processing plant that crushes about 50% of the soybeans that the firm receives from producers at its elevator locations. The rest are merchandized to other processors. The business question is whether or not there is economic value in processing soybeans taken selectively from certain locations and not others.

The cooperative identified 32 of their locations (country elevators) as testing sites for determination of protein and oil content of soybeans delivered to their facilities. Of these 32, 3 of the larger-volume locations were designated as NIR testing centers, and the other 29 were designated as "tributary" locations. The 3 testing centers are labeled as locations A*, B*, and C* and the tributaries are labeled with their corresponding main location letter and then a number (e.g. the tributaries of location A are called A1, A2, A3, etc.). Locations A* and B* each had 13 tributary locations, and location C* (a much smaller facility) had 3. As discussed in Barr and Hurburgh (2019), only tributary locations reporting 15 or more samples were considered for analysis. The standard error of the mean for the locations leveled off after approximately 15 samples. All three testing locations (A*, B*, and C*) collected far more than 15 samples.

Excel Simulation

The simulation of random and systemic errors was performed by using the NORMINV() function in Microsoft Excel. This function was used to create a set of numbers which modeled errors in the data set from the three instruments. Random error components were assumed to be normally distributed without any systemic bias between the three NIR analyzers. Bias was simulated by preferentially adding a bias element to one instrument's data. The syntax for the function is NORMINV(RAND(), Mean, Standard Deviation). Excel creates a set of random numbers with a chosen mean and standard deviation that are normally distributed (Kuo, 2016). For this simulation, a mean of 0 and standard deviations of 1 and 0.5 respectively were used for the protein and oil simulations. These were selected from prior analysis of the data – in a 1990 study, Brumm and Hurburgh found that the average standard deviations of protein and oil contents at country elevators across Iowa were 1.0 and 0.5 percentage points, respectively (Hurburgh & Brumm, 1990). Barr and Hurburgh (2019) found no significant differences between these standard deviations and the standard deviations for the 2018 dataset, so 1.0 and 0.5 were used as estimates in the simulation model. Running the simulation with a mean of 0 and the selected standard deviations created a set of random "error" values which were then added to the corresponding existing data points. The random number generation component of the NORMINV() function caused the data result to be different each time the simulation was run. The mean of 0 meant that the overall averages remained the same.

To represent a potential spread of random errors, the individual sample data for each location was run through the simulation 10 times each in three test cases. Figure 4.1 shows a flowchart of the simulation operations for the three test cases.

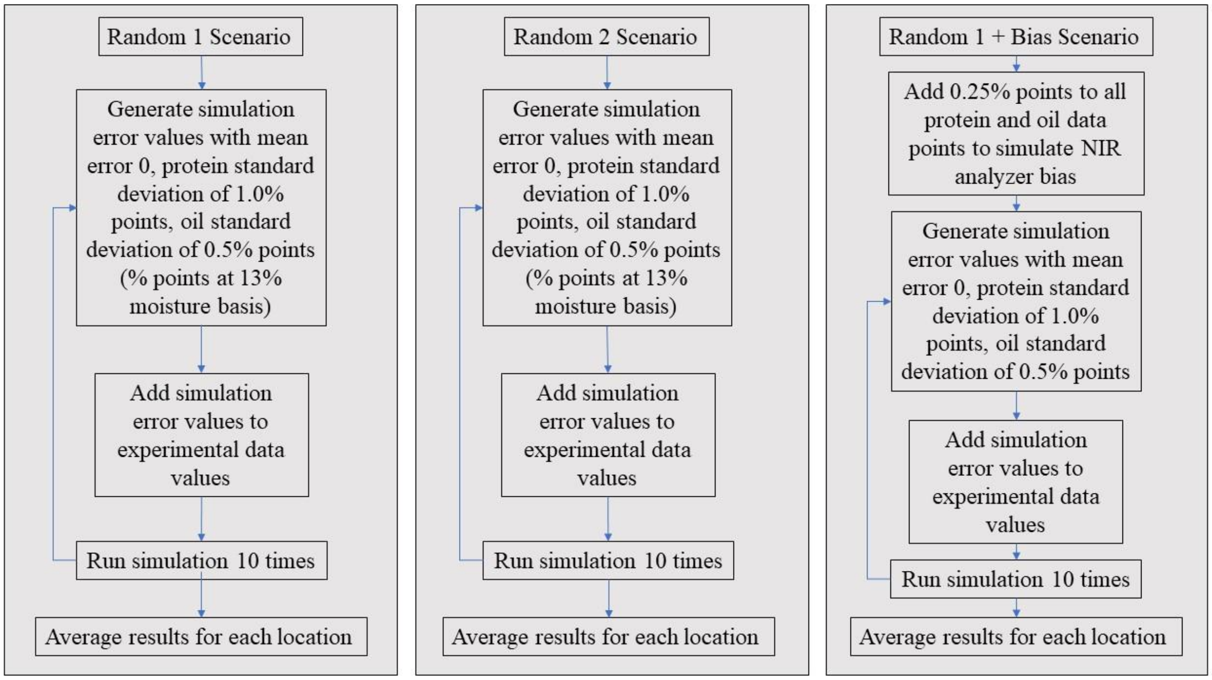

Figure 3.1. Flowchart of operations for random error simulation.

The SUM values for each location for each simulation were compared using an ANOVA test in R, a statistical computing software (R, 2019), to determine the least significant differences between the locations. This analysis included each location that reported greater than 15 samples including the A*, B*, and C* locations which tested individual trucks.

Results and Discussion

Table 3.2 shows the averages by location of the 10 simulation runs for each test case.

Table 3.2. Averages by location of simulation runs for three test cases.

Location	# of Samples	SUM, Original Value (% at 13% moisture)	Simulated Values		
			Random 1 SUM 1.0, 0.5[a] (% at 13% moisture)	Random 2 SUM 2.0, 1.0[a] (% at 13% moisture)	Random 1 SUM + Bias for A Locations (+ 0.25% points[b]) (% at 13% moisture)
A*	164	52.20	52.25	52.20	52.75

Table 3.2, continued.

Location	# of Samples	SUM, Original Value (% at 13% moisture)	Simulated Values		
			Random 1 SUM 1.0, 0.5[a] (% at 13% moisture)	Random 2 SUM 2.0, 1.0[a] (% at 13% moisture)	Random 1 SUM + Bias for A Locations (+ 0.25% points[b]) (% at 13% moisture)
A5	40	52.49	52.54	52.60	52.94
A8	78	52.91	52.89	52.92	53.45
A9	30	52.76	52.86	52.72	53.34
A12	25	53.29	53.13	53.06	53.70
A13	32	53.01	52.99	53.12	53.44
B*	377	52.67	52.68	52.66	52.68
B1	22	53.06	52.98	53.10	52.98
B2	15	53.40	53.45	53.17	53.45
B7	27	52.92	52.71	52.90	52.71
B8	31	53.31	53.29	53.14	53.29
B10	27	52.88	52.91	52.85	52.91
B13	31	53.43	53.40	53.43	53.40
C*	1292	53.24	53.24	53.24	53.24
C1	23	52.76	52.68	52.74	52.68
C3	30	52.57	52.55	52.68	52.55

[a]Simulated standard deviations of protein and oil data.
[b]Simulated constant bias of unit at A*.

The introduction of a random error element was intended to simulate a combination of sources of error in the experiment. With errors included, gaps between locations were smaller. Barr and Hurburgh (2019) found no significant differences between the three NIR analyzers with respect to each other and to the ISU Grain Lab reference instrument over the harvest period. Table 3.3 shows the deviation of each analyzer from the reference instrument mean.

Table 3.3. Deviation of three test site analyzers from reference instrument.

ISU Grain Lab Reference Analyzer Mean (n = 18)		A* Analyzer Deviation from Mean		B* Analyzer Deviation from Mean		C* Analyzer Deviation from Mean	
Protein	Oil	Protein	Oil	Protein	Oil	Protein	Oil
32.94	19.92	-0.19	-0.02	-0.19	0.08	-0.16	0.08

The three analyzers averaged -0.18 on protein deviation and +0.05 on oil deviation. This deviation was fairly even across all three – one analyzer did not contribute more to error than another. Therefore, the NIR analyzers themselves were not biased relative to each other, though they still contributed to random error. This confirmation was crucial to any further application of the data, because one of the largest sources of error in a commodity-based market system is inconsistency of measuring units with each other.

However, the analyzers were operated by staff at the locations, many of whom had not performed any such experiment before and only had a short training session. The beans were not always tested as soon they were collected, especially for the composite samples from tributary locations. It was important not to interfere with the harvest schedule of the cooperative locations. During busy periods, testing was deferred until later in the day or sometimes later in the week. The samples from the tributary locations were usually not brought to a testing location until a few days after they were collected. This could have affected the moisture content of the beans, but would not affect the percentages of oil and protein on a constant moisture basis, so likely had no impact on the composition results. Scale houses, where the instruments were located, are not a controlled or sterile environment – changes in temperature, relative humidity, and amount of dust in the air could have influenced the analyzer results, likely introducing random error. All of these in practice would contribute to the unknown source portion of the random error.

Including random error in the measurements made it more difficult to tell which location had the “best” beans. Rankings based on the average values from the 10 Random 1 simulation runs for each location are displayed in Table 3.4. SUM was used as the proxy for value, with each point of increase or decrease in SUM value representing a $0.16 per bushel increase or decrease in value. The "Estimated Value Difference from Mean" column shows the effects of random error on pricing – ranking by the Random 1 SUMs essentially negated any value differences between the locations seen in Barr and Hurburgh (2019). The rankings by Random 1 SUM did not correlate to a greater positive value difference from the mean Random 1 SUM because the error introduced in the simulation for both protein and oil components was compounded once by calculating the SUM, and again by calculating prices.

Table 3.4. Location rankings by average values of Random 1 SUMs.

Ranking by Random 1 Sum	Location	Location Random 1 Sum – Mean Random 1 Sum	Estimated Value Difference from Mean (+1% point SUM = +$0.16/bushel)
1	B2	-0.66	-$0.10
2	B13	-0.37	-$0.06
3	B8	-0.02	$0.00
4	C*	-0.05	-$0.01
5	A12	0.22	+$0.03
6	B10	0.08	+$0.01
7	A8	-0.23	-$0.04
8	A13	0.07	+$0.01
9	A9	0.54	+$0.08
10	C1	-0.20	-$0.03
11	B*	0.38	+$0.06
12	B1	0.00	$0.00
13	A5	0.49	+$0.08
14	C3	0.33	+$0.05
15	B7	-0.23	-$0.04
16	A*	-0.36	-$0.06

The Least Significant Difference (LSD) between the locations was calculated for each simulation. This was found by comparing all the locations in R for each test case to find the Mean Squared Error (MSE) for each simulation. The LSD was then calculated with the following formula: $LSD = t_{critical} * \sqrt{MSE * (\frac{1}{n_1} + \frac{1}{n_2})}$, where n_1 and n_2 are the sample sizes for the groups being compared. $t_{critical}$ for 144 degrees of freedom (160 observations – 16 locations) was found from a standard t-table to be 1.645. The results for the Random 1 and Random 2 test cases are displayed in Table 3.5.

Table 3.5. Least significant differences between locations for Random 1 and Random 2 simulation cases.

	Random 1 Simulation	Random 2 Simulation
Mean Squared Error (MSE)	0.038	0.199
Least Significant Difference (LSD) between Locations	0.144	0.328
Maximum Location SUM	53.45	53.43
# of Locations within 1 LSD of Top Location	1	6

The MSE increased by a factor of approximately 5, and the LSD more than doubled. The number of locations within 1 LSD of the top location increased from 1 to 6, showing that as variability increases, differentiation between locations becomes more difficult. If variability continued to increase, separating locations based on value would become impossible because the range from the maximum SUM to 1 LSD from the maximum would encompass all of the locations. In a larger market network with many NIR units of more than one make and model, error control would be crucial to substantiate value differences.

Conclusions

The introduction of random error made any value gaps between locations smaller, which made discrimination of high-value locations from average or low-value locations difficult. These results show the importance of having standards for measuring instruments if the soybean supply chain is ever to move to a protein and oil pricing basis, because inconsistency of measuring units with each other is a major source of error in any commodity-based marketing system. Unknown but systemic biases complicate decisions even further. Therefore, the validity of marketing decisions made using data collected from elevators depends highly on the amount of error involved in sample analysis. Future studies should identify specific sources of error and develop protocols to minimize or eliminate them, because maximizing potential value capture will not be possible unless the value differences between locations are characterized as precisely as possible.

References

Barr, B. E., & Hurburgh, C. R., Jr. (2019). *Strategies for selective handling of soybeans to maximize soybean meal protein and yield*. (Manuscript in preparation for publication).

Brumm, T. J., & Hurburgh, C. R., Jr. (2006). [Letter to the editor]. *Journal of the American Oil Chemists Society*, *83*, 981-983.

Hurburgh, C. R., Jr. (1994). Identification and segregation of high-value soybeans at a country elevator. *Journal of the American Oil Chemists Society*, *71*, 1073-1078.

Hurburgh, C. R., Jr., & Brumm, T. J. (1990). Protein and oil content of soybeans received at country elevators. *Applied Engineering in Agriculture*, *6*, 65-68.

Hurburgh, C. R., Jr., Paynter, L. N., & Schmitt, S. G. (1987). Quality characteristics of Midwestern soybeans. *Applied Engineering in Agriculture*, *3*, 159-165.

Kuo, M. (2016, January 10). How to create a normally distributed set of random numbers in Excel. Retrieved from MBAExcel website: http://www.mbaexcel.com/excel/how-to-create-a-normally-distributed-set-of-random-numbers-in-excel/

R: The R Project for Statistical Computing [Computer software]. (2019).

Wagner, K., Hurburgh, C. R., Jr., & Brumm, T. J. (2017). SPROC 3.0 [Computer program].

CHAPTER 4. GENERAL CONCLUSIONS

Through these studies, many conclusions can be made about segregation of soybeans into value groupings at country elevators. Soybean protein and oil levels matched historical trends, meaning that the overall variability of soybeans in Iowa is not changing significantly. The quality of bean deliveries over the harvest season was consistent, meaning that a cooperative could theoretically test only the 15 samples over the two-week period deemed necessary to adequately represent a location in order to make an informed marketing decision. Geographic variability across the cooperative's locations was evident, and this variability, in theory, corresponds to lost profits if the cooperative does not source beans from the higher-value locations. However, the validity of marketing decisions made using data collected from elevators depends highly on the amount of error involved in sample analysis. Future studies should identify specific sources of error and attempt to eliminate them, because maximizing potential value capture will not be possible unless the value differences between locations are characterized as precisely as possible.