

2021

Maximizing Processing Value with Selective Handling Strategies: An Analysis of Soybeans Received at Iowa Elevators

Bennett E. Barr
Iowa State University

Charles R. Hurburgh
Iowa State University, tetry@iastate.edu

Gretchen A. Mosher
Iowa State University, gamosher@iastate.edu

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



Part of the [Agriculture Commons](#), and the [Bioresource and Agricultural Engineering Commons](#)

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/abe_eng_pubs/1235. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the Agricultural and Biosystems Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Agricultural and Biosystems Engineering Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Maximizing Processing Value with Selective Handling Strategies: An Analysis of Soybeans Received at Iowa Elevators

Abstract

A better ability to understand and use geographic variations in protein and oil is one way to maximize the value potential of soybeans for handlers and processors. An Iowa cooperative had been sourcing soybeans for processing from nearby elevator locations and wanted to know whether this strategy was maximizing the net processing value of the soybeans. Random and systematic errors from testing and measurement instruments also impact marketing decisions and were investigated as part of this project. 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 soybean processing plant. Samples were analyzed using near-infrared spectroscopy (NIR), and protein and oil content data were entered into an Estimated Processing Value (EPV) model to determine value differences of soybeans among elevator locations. Results showed substantial variability among locations that represented a \$0.23/bushel EPV spread. No significant variation was found in soybean quality over the harvest season, suggesting that marketing decisions can be made at the beginning of the season. To determine the incidence of random errors, a simulated Excel-based model was used with three test cases. The introduction of random error lowered value gaps between locations, which made the discrimination of high-value locations from average or low-value locations difficult. Although protein and oil measurement with the NIR instrument was feasible even on busy harvest season days, the validity of marketing decisions using these data depended highly on the error involved in sample analysis. Future studies should identify specific sources of error and attempt to eliminate them. Specifically, one of the largest sources of error in a commodity-based market system is in the measuring units. The ability to isolate and quantify measurement error will improve the capability of the commodity-based soybean market system to focus trade decisions on end use traits, maximizing soybean value and providing incentive for improvement.

Keywords

Grain elevators, soybean value, selective handling, soybean processing, error analysis

Disciplines

Agriculture | Bioresource and Agricultural Engineering

Comments

This is a manuscript of an article published as Barr, Bennett E., Charles R. Hurburgh, and Gretchen A. Mosher. "Maximizing Processing Value with Selective Handling Strategies: An Analysis of Soybeans Received at Iowa Elevators." *Applied Engineering in Agriculture* (2021). DOI: [10.13031/aea.14534](https://doi.org/10.13031/aea.14534). Posted with permission.

MAXIMIZING PROCESSING VALUE WITH SELECTIVE HANDLING STRATEGIES: AN ANALYSIS OF SOYBEANS RECEIVED AT IOWA ELEVATORS

Bennett E. Barr¹, Charles R. Hurburgh¹, Gretchen A. Mosher^{1,*}

The authors are **Bennett E. Barr**, Former Graduate Student, and **Charles R. Hurburgh**, Professor,

¹ Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa, USA.

* Correspondence: gamosher@iastate.edu.

Highlights

- Levels of protein and oil influence the marketability of soybeans
- Geographic differences in soybeans have economic implications for grain elevators
- Random errors in the receiving and analysis process mask these differences
- For soybeans to be marketed on protein and oil levels, errors in handling and analysis must be addressed

ABSTRACT.

A better ability to understand and use geographic variations in protein and oil is one way to maximize the value potential of soybeans for handlers and processors. An Iowa cooperative had been sourcing soybeans for processing from nearby elevator locations and wanted to know whether this strategy was maximizing the net processing value of the soybeans. Random and systematic errors from testing and measurement instruments also impact marketing decisions and were investigated as part of this project. 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 soybean processing plant. Samples were analyzed using near-infrared spectroscopy (NIR), and protein and oil content data were entered into an Estimated Processing Value (EPV) model to determine value differences of soybeans among elevator locations. Results showed substantial variability among locations

that represented a \$0.23/bushel EPV spread. No significant variation was found in soybean quality over the harvest season, suggesting that marketing decisions can be made at the beginning of the season.

To determine the incidence of random errors, a simulated Excel-based model was used with three test cases. The introduction of random error lowered value gaps between locations, which made the discrimination of high-value locations from average or low-value locations difficult. Although protein and oil measurement with the NIR instrument was feasible even on busy harvest season days, the validity of marketing decisions using these data depended highly on the error involved in sample analysis. Future studies should identify specific sources of error and attempt to eliminate them. Specifically, one of the largest sources of error in a commodity-based market system is in the measuring units. The ability to isolate and quantify measurement error will improve the capability of the commodity-based soybean market system to focus trade decisions on end use traits, maximizing soybean value and providing incentive for improvement.

Keywords.

Grain elevators, soybean value, selective handling, soybean processing, error analysis

INTRODUCTION

Increasing profit from soybean production and processing is a major issue in the agricultural community. Choices in soybean varieties influence every user in the supply chain from the producer forward. As stated by the United Soybean Board, while producers often judge varietal selections by yield first, “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). Presently, protein and oil contents are included only indirectly in the U.S. soybean pricing system (Updaw, Bullock, and Nichols, 1976).

For this reason, increased knowledge of soybean traits is useful for producers growing soybeans, for grain handlers who buy and sell soybeans into the larger grain market, and for processors who transform them into an assortment of products (United Soybean Board, 2017). There is a current disconnect in the supply chain. The typical trade between producers and elevators and elevators and processors are stand-alone decisions not involving composition. Elevators and processors pay based on U.S. Grades of test weight, moisture, and foreign material, not on protein and oil content (Hurburgh, Paynter, and Schmitt, 1987, Laux, Mosher, and Hurburgh, 2015). This procedure takes place despite the importance of protein and oil content for end users (Medic, Atkinson, and Hurburgh, 2014).

The setting for this research was an Iowa-based cooperative that, in addition to its elevator locations, owns a soybean

processing plant. The company purchases soybeans from its elevators and then either sells them to other processors or transports the beans to their own facility to make soybean meal and soybean oil. Lower value beans that are not suitable for processing are directed to another end use. The current strategy is to source soybeans from locations closest to the processing plant to save on transportation costs. The primary question of this study was whether the strategy was maximizing the net processing value for the company. Processors have traditionally believed that localized patterns of protein and oil content do exist, but little empirical evidence exists in the public domain (Hurburgh & Brumm, 1990; Barr & Hurburgh, 2019). Table 1 summarizes previous studies on protein and oil patterns in soybeans at specific delivery points.

Table 1. Historical data on geographic protein and oil variations.

Reference	States/Regions Covered	(Number of Locations) Number of Samples Tested	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%
Average Values (% at 13% moisture)			35.1%	1.0%	18.9%	0.6%

The cooperative wanted to update and expand the information on geographic soybean variations, then assess the economic potential of identifying beans for the plant by value rather than distance. Measurement error is often cited as a primary reason for not considering testing. The effect of potential analytical errors on the location rankings is an important consideration in ensuring that location differences are statistically “different enough” to make segregation worth the cost and effort. Therefore, an error analysis was also performed to validate the measured differences.

In this study, data collected 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 percentages at 13% moisture basis. SUM values in soybeans

are closely tied to their estimated processing values (EPVs) because the relative values of protein and oil are proportional to the market prices for soybean meal and oil. Hurburgh (1994) tested the validity of using the SUM as a ranking criterion and found that differentiation in the market was possible. 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; Hurburgh, Neal, McVea, and Baumel, 1994). Because the relationship between SUM and EPV is linear, a dollar value can be estimated for every percentage point of increase or decrease in SUM. SUM values provide a quick valuation estimate for making decisions when time is limited or if a full EPV analysis is not feasible. Because SUM and EPV are so closely related, errors in component measurement create error in the soybean valuation.

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

- Systematic:
 - Consistent bias in the measuring instrument (i.e. one analyzer reading a constant amount too high on each measurement) would cause a shift across a group of unit measurements by that specific NIRS unit.
 - Variances in instrument standardization (individual instrument alignment with standards to compensate for differences among copies) would cause instruments to read consistent differences among the testing locations.
- Random:
 - User error: analyzers operated by largely untrained personnel that were different over time and across locations. Any differences in NIR operating technique would result in an inconsistent user operating bias.
 - Environmental effects: temperature fluctuations were different each day in the scale houses where the analyzers were housed and thus could affect each measurement to a different extent.
 - Equipment deterioration: power inconsistencies causing dimming or brightening of the internal light source in the NIR affected each measurement differently. Individual pixels in a diode array may change with use, requiring more frequent recalibration.

It is also worth noting that errors designated as random may also be systematic. Differences in users, environmental effects such as temperature, and equipment deterioration could potentially be based on error that has a non-random pattern.

All of these sources of error combine to form the uncertainty of individual tests, which then affects the SUM calculation. Greater uncertainty reduces the statistical ability to separate high and low value soybeans.

PROJECT RESEARCH OBJECTIVES

This project included two components: the first examined the feasibility of identifying high levels of protein and oil in a commodity soybean handling environment and whether these levels showed regional patterns. The second component addressed the role of measurement errors on the component value ranking of soybean samples and harvest locations in a typical trade area of a soybean processing plant. The error analysis estimated the degree to which analytic errors must be controlled to make location selection economically feasible for a processor.

PART 1 – DATA COLLECTION AND RAW DATA ANALYSIS

Accordingly, the following questions guided data collection and analysis:

1. Do levels of protein and oil content in soybeans received by different receiving locations owned by an Iowa grain cooperative differ?
2. Is it economically feasible to preferentially source soybeans with high levels of protein, oil, and processing value?
3. Can simulating random and systematic errors to NIRS-predicted soybean composition data influence protein and oil SUM ranking of soybeans from individual elevator locations?

MATERIALS AND METHODS

PART 1 DATA COLLECTION AND ANALYSIS

Location Identification

Thirty-two locations were identified by the company as testing sites for determination of protein and oil content of soybeans delivered to their facilities. All locations were within 50 miles of the processing plant, typical trucking distance of the plant. Of the 32, 3 of the larger-volume locations were designated as near-infrared spectroscopy (NIRS) 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 3 testing locations were renamed and labeled as locations A*, B*, and C* in this paper. The tributaries were labeled with their corresponding main location letter and then a number (e.g. the tributaries of location A are called A1, A2, A3, and so on). Locations A* and B* each had 13 tributary locations, and location C* (a much smaller facility) had 3 tributary locations. A geographic representation of sampling locations are shown in Figure 1, with sample

testing locations indicated by stars and the tributary locations connected to the respective testing site. All locations are located in western and southwestern Iowa.

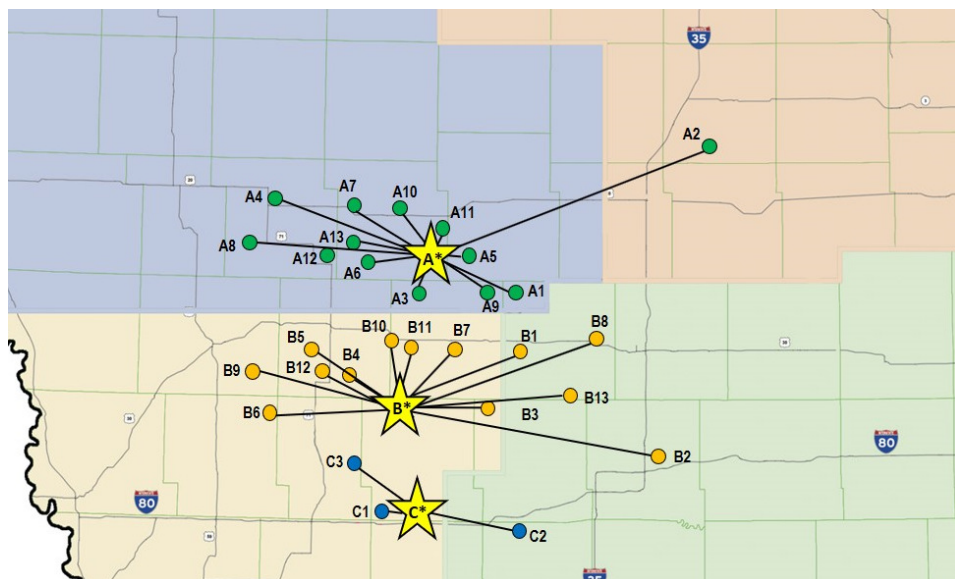


Figure 1. Geographic locations of sampling locations in SW, NW, and North Central Iowa

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-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 through A13 sent their composite samples to A* to be tested and similarly for B* and C*). Therefore, the tributary samples were themselves mixtures of undivided loads.

Near Infrared Analysis

Infratec 1241 near infrared transmission (NIT) analyzers (Foss North America, Eden Prairie, MN) were provided to the A*, B*, and C* locations. The units were able to transmit data to a laptop for Excel data storage, were user-friendly, and took the same size sample (~500 g) that was already used for the capacitance moisture meter, which were used to collect data on moisture content. The three analyzers were loaned to the locations from Iowa State University, and employee training was provided. The NIR machines had been calibrated for soybean composition analysis before delivery by the Iowa State University Grain Quality Laboratory. The NIRS technology offered a quick and non-destructive method of measuring protein and oil levels

(Shenk, Workman, and Westerhaus, 2008). The time required to run each sample and process data was approximately 2.5 minutes from start to finish.

Protein, oil, and fiber concentrations were determined by the Infratec units, using calibrations developed at the Iowa State University Grain Quality Laboratory. The Iowa State University calibration process was described by Rippe 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 (Mills, 2017). The last crop year represented in the calibration data set was from 2013 (Hurburgh, 2015). The validation data for the 2014 and 2015 crop years are displayed in Table 2.

Table 0. Infratec validation data for two years after calibration creation

Factor	2014 Crop (n = 91 samples used)		2015 Crop (n = 96 samples)	
	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

SEP = standard error of prediction; RPD = standard deviation of data/SEP

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. The procedure was then repeated for the afternoon. The resulting two samples per day were sent to the corresponding testing location. The number of truck loads in each composite was not recorded but was estimated to be between 10 and 25, depending on receiving rates for the day.

Verification of NIT Analyzers

Randomly selected samples from A*, B*, and C* were transported to the Grain Quality Lab at Iowa State University once per week for testing on laboratory NIR analyzers. The same sample was tested on the A*, B*, C*, and ISU analyzers. The results were compared to ensure readings were consistent 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 completed using the SPROC 3.0 soybean processing model in Microsoft Excel (Wagner, Hurburgh, and Brumm, 2017). Figure 2 displays a typical solvent extraction flow diagram, which was adapted to become an expeller-press process scenario in Step 1. Figure 2 from Wagner (2017) shows the user input sheet of the SPROC 3.0 model.

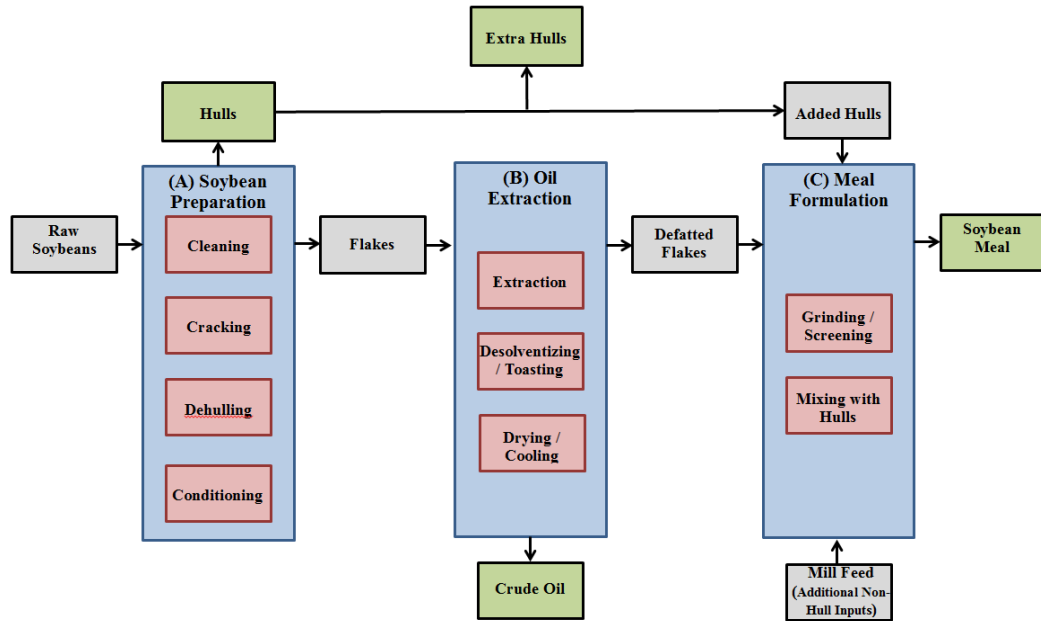


Figure 2. Soybean solvent extraction flow diagram.

Table 3 displays the model inputs and assumptions. The current prices at the end of the market day on December 12th 2018 were used. The following parameters were used for all model iterations.

1. The soybean preparation section (A Parameters in Figure 1, System A Parameters in Figure 2) was turned off, because no hulls were removed in the cooperative’s process. The model simulated an expeller-press scenario instead of a solvent-extraction scenario. All weight of incoming beans 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 was sold “as is” for protein and fiber.
3. Amino acid content and prices were not considered as they were not necessary for the meal and oil valuations. This meal was used for dairy cattle feed, where amino acids are not limiting.

Table 3. SPROC processing parameters and 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 ^[e]
	% 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; no hulls were added 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 1994 study, Hurburgh 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 R² 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 were able to provide quick valuation estimates for decision-making when time was limited or if a full EPV analysis was not feasible.

PART 2 ERROR ANALYSIS

Error Simulation

The same data that were collected during the Fall 2018 soybean harvest from multiple grain elevator locations of an Iowa grain cooperative were used for the error analysis. The simulation of random and systemic errors was performed using the NORMINV() function in Microsoft Excel. This function was used to create a set of numbers to model errors in the data set from the three instruments. Random error components were assumed to be normally distributed without any systemic bias between and among the three NIR analyzers. Bias was simulated by preferentially adding a bias element to one instrument's data. The syntax for the function was NORMINV(RAND(), Mean, Standard Deviation). Excel created 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 based on prior analysis of the data. Brumm and Hurburgh (1990) found 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 initial estimates in the simulation model. This gave the simulated analysis errors in the dataset equal to the approximate variation in the whole data set. 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 resulted in a different data result each time the simulation was run. A mean of 0 indicated that the overall averages remained the same.

To represent a potential distribution of random errors, the individual sample data for each location was inputted through the simulation 10 times each in three test cases. Figure 3 shows a flowchart of the simulation operations for the three test cases.

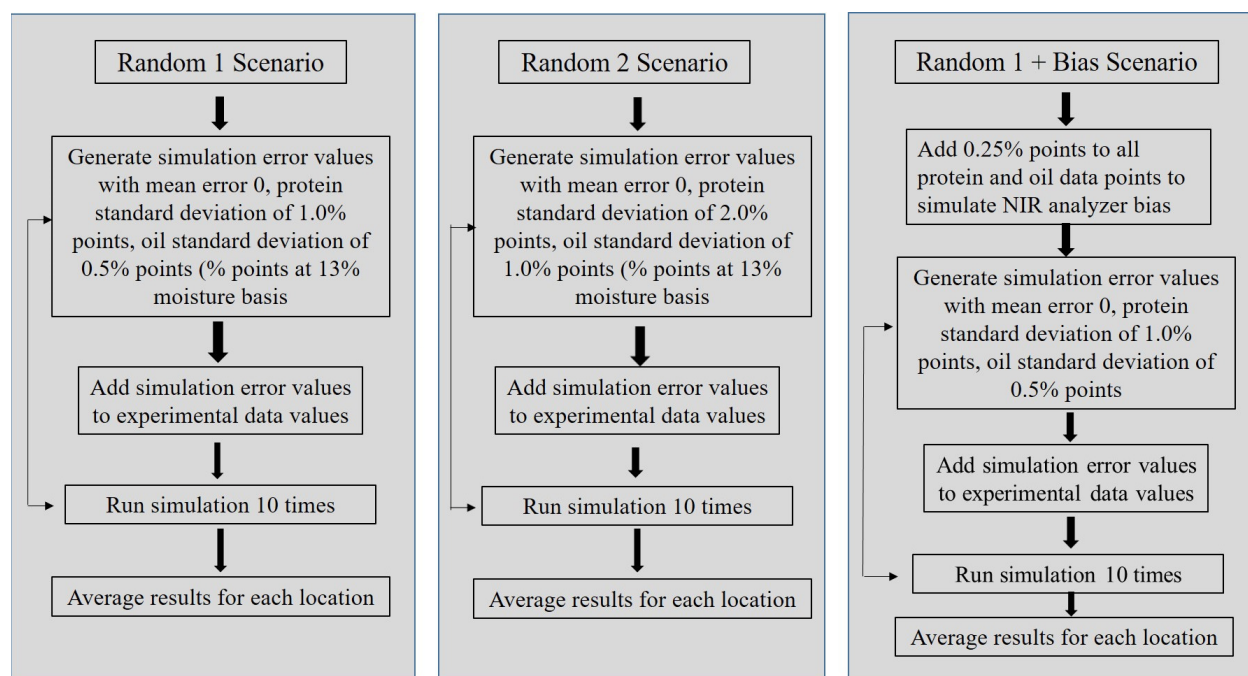


Figure 0. Flowchart of operations for random error simulation.

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

DATA ANALYSIS

The statistics for the overall data set collected during the Fall 2018 soybean harvest are listed below in Table 4.

Table 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. This decision was determined by graphing the number of samples versus the standard error of the mean ($\frac{SD}{\sqrt{\# \text{ of samples}}}$) for the tributaries. Figure 4 shows that the standard error of the mean stabilizes after approximately 15 samples. These data were used to justify the decision to only use samples greater than 15 in the analysis.

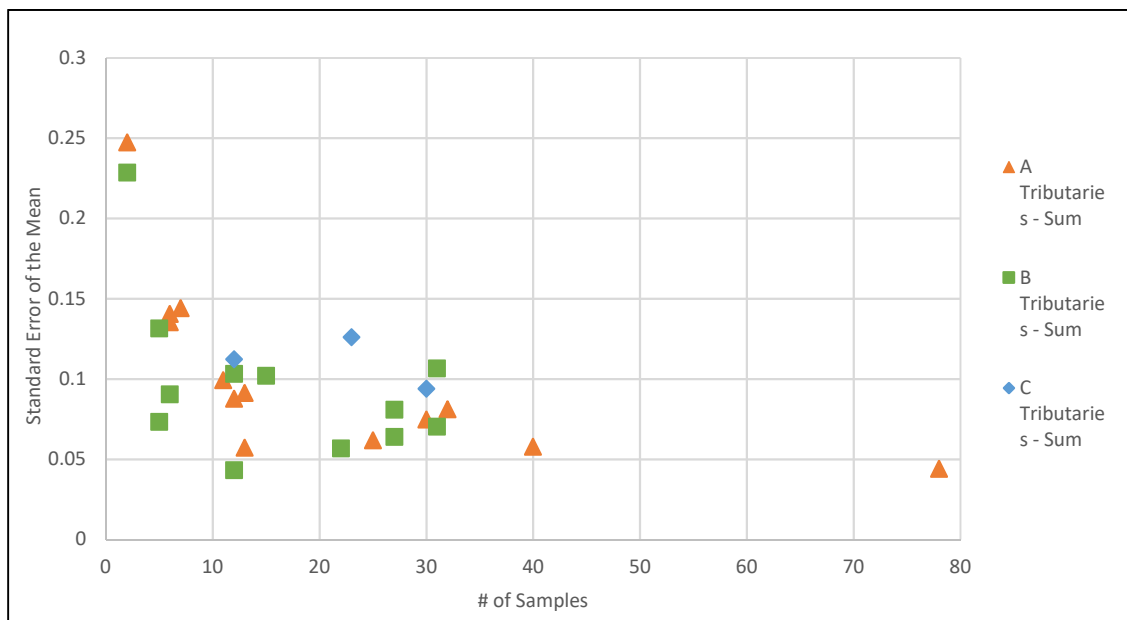


Figure 4. A, B, and C Tributaries: Number of samples versus standard error of the mean

Tables 5, 6, and 7 summarize the percentage points protein + percentage points oil (SUM) averages for each testing location and its tributaries, addressing research question 1 concerning the protein and oil levels and the SUM values. All protein, oil, and SUM values are reported on a percentage at 13% moisture basis. Locations reporting less than 15 samples are listed in the table footnotes.

Table 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 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 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 8.

Table 1. Weekly verification samples from three dates throughout harvest.

Date/Analyzer Location	Protein				Oil			
	<i>ISU Grain Lab</i>	A*	B*	C*	<i>ISU Grain Lab</i>	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. Values from all days showed no significant differences from the laboratory reference analyzer or from each instrument, which demonstrated that the analyzers were all producing accurate results throughout the course of the experiment.

A goal of the project was to determine if this processor could identify locations within its trade area that had higher value soybeans, and direct those beans to its own processing plant, with the other beans sold to other processors who were not testing. For this reason, this processor is the only party who will benefit. Other processors will experience some loss. Yet, the methodology provides a method of sending market “signals” to producers to influence choices in genetics selection. 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. Ten data points were selected to reflect the range of the data and values were entered into the SPROC model.

The resulting EPVs (Estimated Processed Values) are shown in Figure 5 below with a corresponding trend line. This line is specific to the price combinations on December 12th, 2018. The same process of using fewer points was used by Brumm and Hurburgh (1990) to demonstrate the high correlation of SUM and EPV. The slope of SUM on EPV is a function of the relative price combinations for soybean meal and oil, which can make a graph look very busy. A chart with fewer points was less cluttered, illustrated the point that SUM and EPV have a relationship, and clearly demonstrated that SUM is an adequate proxy for ranking.

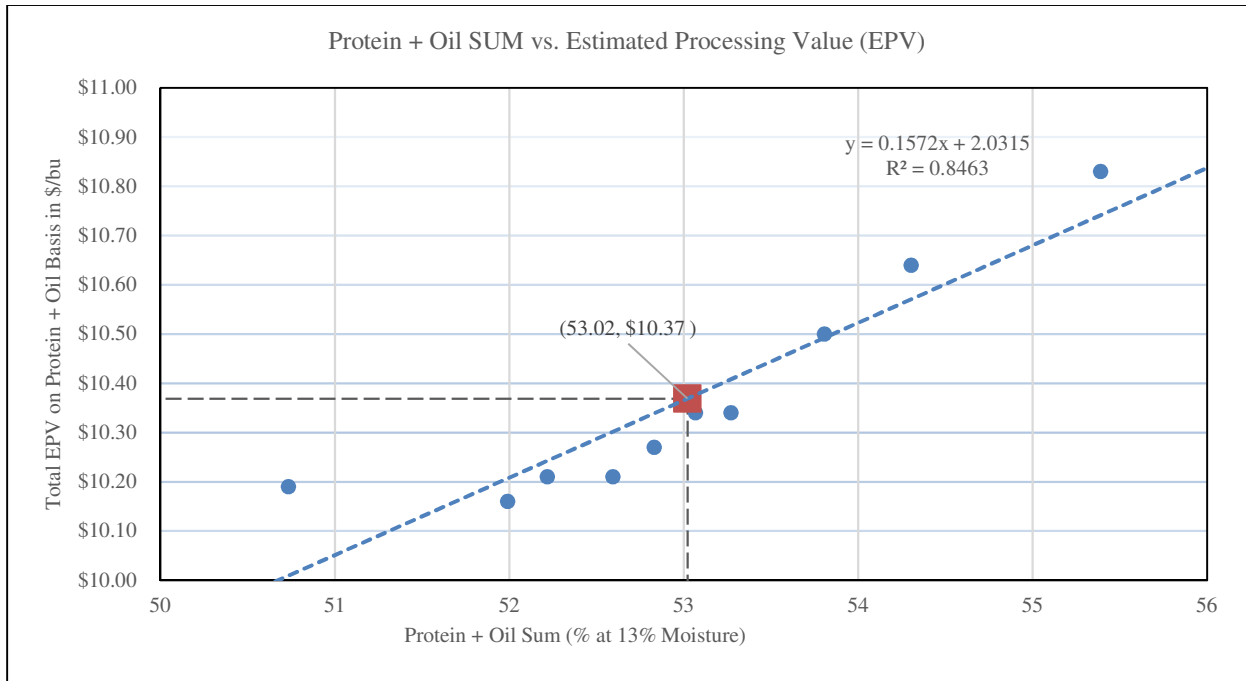


Figure 5. SUM vs. Estimated Processing Value (EPV) for 10 randomly selected data points.

The approximate economic worth of a change in the protein + oil sum is represented by the slope of the trend line. For every percentage point increase or decrease in the protein + oil sum, the EPV per bushel will accordingly increase or decrease by \$0.1572, or about 16 cents per bushel per 1% increase or decrease of SUM. This was applicable for the given combination of prices. The square point represents the average sum and average EPV values. The cooperative was receiving soybeans grown in locations close to the processing plant. However, Figure 5 shows that there is value in sorting based on bean characteristics such as protein or oil content and not just proximity to the processor. Even with the challenging weather patterns and sporadic harvest during the 2018 growing season, clear value differences were evident between locations. 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 potentially be reduced.

Currently, processors attempt to identify regional areas that have a higher SUM and bid preferentially (although not openly) for these soybeans. The lower value beans default to users who are either unaware of the value or are out of an economic transportation position to acquire the higher value soybeans. These users tend to respond by lowering their bids to sellers. Yet, the reason for the lower bid is not fully transparent from a market or economic perspective.

Average protein and oil values for each location were entered into the SPROC model and graphed. These results can be seen in Figure 6, which shows the EPVs by location.

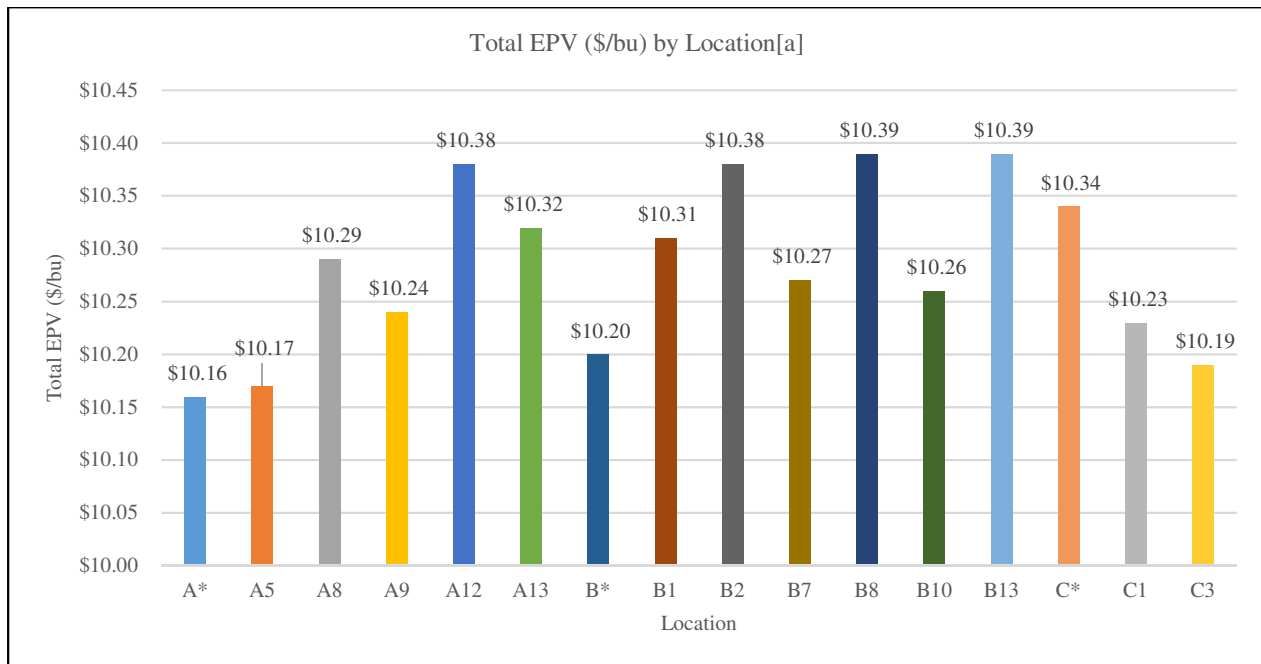


Figure 6. EPVs in \$/bushel by location for locations with 15 or more samples.

^[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 6 builds on the conclusions from Figure 5 in terms of separating and sourcing beans based on value and not purely on transportation costs. The data in Figure 6 represent 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 being routed to the processing plant. 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 of 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 greater 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 the results do not reveal potential contributing factors for these differences. Possible explanations include, but are not limited to, differences in soil types, weather patterns, soybean varieties grown, and individual farming practices. Contributing factors resulting in geographic differences in soybean protein and oil 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 would facilitate more efficient processing and direct end-use decisions.

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 the 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 valid marketing decisions as early into harvest as possible. They wanted data to confirm that bean quality at a given location varied minimally across harvest. Accordingly, 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 7) against the numerical day of harvest on which the sample was tested.

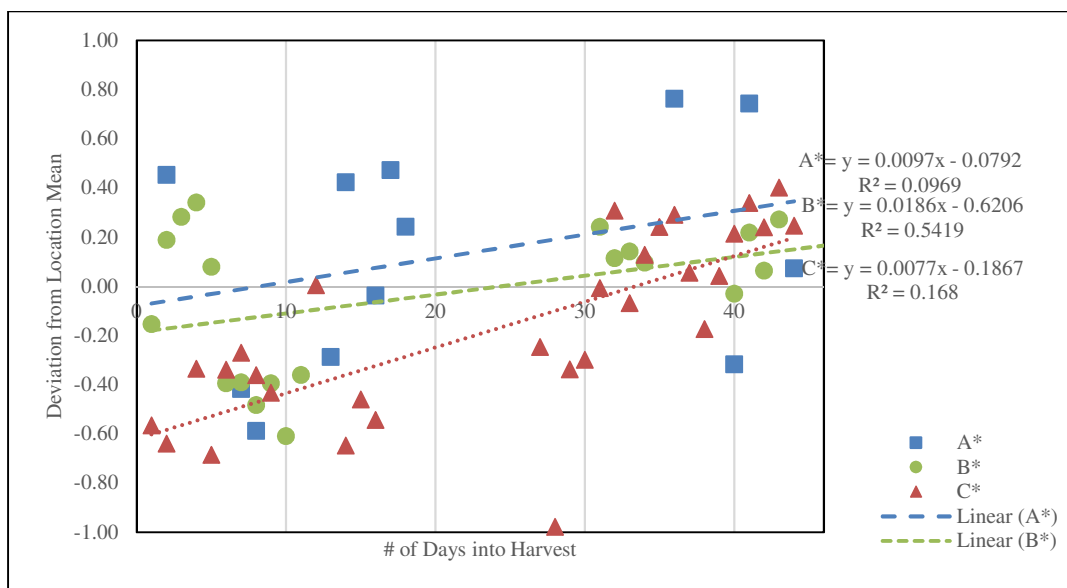


Figure 7. Variability of soybeans over harvest at A*, B*, and C* NIR testing locations.

As shown in Figure 7, all daily data points were contained within a $\pm 1\%$ -point range of the 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, with 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 was a large gap in the middle of the graph – this was due to a three-week rainy period when harvest stopped, and no samples were taken. Figure 7 confirms that bean quality was consistent throughout harvest, and for this reason, it was not necessary to measure every load that comes into a facility over the entire harvest season. The atypical pattern of C* indicates, however, that there could be situations specific to only one part of a trade area, that could impact characteristics at harvest.

Past studies (Hurburgh, 1994) suggested that two weeks of sampling was sufficient to characterize the harvest data, which was also true in this experiment. Additionally, a total count of 15 samples 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. A confirmation follow-up mid or late harvest would identify any changes in trends. 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.

PART 2 – ERROR ANALYSIS

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. This confirmation was crucial to any further application of the data, because one of the largest hypothesized sources of error in a commodity-based market system is inconsistency of measuring units with each other. Table 9 shows the averages by location of the 10 simulation runs for each test case.

Table 9. 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
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

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 10. 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. The ranking by the Random 1 SUMs negated value differences between the locations seen in the raw data analysis. 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 10. 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 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 11.

Table 11. 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 and equivalence would be crucial to substantiate value differences.

CONCLUSIONS

Several findings can be drawn from this research. In addressing the first research question which asked if levels of protein and oil content in soybeans received by different locations differed, no significant differences were found between locations or among different dates of collection. Soybean protein and oil levels aligned with 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. Yet, within each region, geographic variability was present, although the level of variation was not high, this raised the question of whether it made economic sense to preferentially source soybeans with high levels of protein, oil, and processing value.

To determine if preferential sourcing is feasible economically, a representative set of data points were entered into the SPROC model, resulting in the Estimate Processed Values (EPVs) on protein and oil basis in dollars per bushel. ANOVA analysis of the data indicated a significant difference ($p < 0.05$) in the EPVs between locations. Findings suggest preferential sourcing of soybeans for processing makes economic sense for the cooperative. Although the analysis used data from one unusual crop year, given the historic robustness of soybean composition (Hurburgh, Brumm, Guinn, and Hartwig, 1990; Thakur and Hurburgh, 2007; Medic, Atkinson, and Hurburgh, 2014), use of the SUM approach (Hurburgh, 1994), and the application of the Estimated Processed Values model in previous research studies (Wagner, 2017; Brumm and Hurburgh, 1990), it is believed the output of the model would not change substantially when using soybeans from other crop years. The 2018 crop year included unusual weather, which influenced the quality of soybeans. If the selection strategy was effective in modeling under these somewhat extreme conditions, the ability of the model to provide a valid prediction in more stable crop years is high.

Individual producer loads of low value soybeans were not rejected based on the SUM data. A goal of the project was to identify delivery location averages rather than to pay individual producers more or less. There was no plan to reject any beans or to pay any producer differently than any other relative to the company's posted bid on a given day. Influencing producers to modify their genetics choices was not within the scope of this project, but is considered the next step. Before this policy could be adopted, some demonstration was needed that elevator locations had higher or lower averages. Selling on location rather than paying producers by testing at every location is simpler and easier, as long as the differences between locations are large enough to support the effort. In this case, the differences were large enough to warrant the effort.

The role of sample analysis error was also hypothesized to have economic implications. Geographic variability across the cooperative's locations was evident, and the variability potentially corresponded 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 is highly dependent on the sample analysis error. Findings suggest that controlling error is critical to substantiating geographic value differences in soybeans.

Random and systematic errors clearly impacted the ability to differentiate high value soybeans for processing in this case. Several sources of random error were evident in the process. The NIR analyzers were operated by staff at the locations,

many of whom had limited experience performing the tasks required, save a short training session. The beans were not always tested as soon they were collected, especially in the case of composite samples from tributary locations. It was important not to interfere with the harvest schedule of the cooperative locations, yet the decision to remain distant from the schedule resulted in potential random and systematic errors, some of which may not even be known to the researchers. As an example, 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. While the delay may have affected the moisture content of the beans, it would not have affected the percentages of oil and protein on a constant moisture basis. Scale houses, where the instruments were located, are not a controlled or sterile environment. Changes in temperature, relative humidity, and the amount of dust in the air could have influenced the analyzer results, potentially introducing random or systematic error. Specifically, systematic patterns of temperature, relative humidity, and dust could have had stronger effects at some locations than others. All of these aspects would contribute to unknown sources of random and systematic error. Several potential systematic errors were not addressed in this study, including the experience of the employee, genetic variety of soybeans, and measurement biases resulting from individual instruments and employees. Although the specific characterization of errors within the receiving and sourcing process was not the primary focus on this project, the implementation of statistical process control (SPC) provides one method of characterizing systematic and random errors at receipt, as described in previous research (Laux, Mosher, and Hurburgh, 2015). Future studies and resources should focus on identifying specific sources of error and attempt to eliminate or mitigate them. Maximizing potential value of soybeans, by geographic or any other observed differences, will not be possible unless the value differences between locations are characterized as precisely as possible.

ACKNOWLEDGEMENTS

The authors are grateful to Tim Sullivan, Glen Rippe and Alan Kuuttila for their assistance with the details of this project, the cooperative staff involved in testing for their patience and cooperation, and for Wyatt Runksmeier for assistance in troubleshooting the data collection process at individual locations. Authors are also grateful for the feedback given by the three reviewers of this manuscript. Funding from the United States Soybean Board is also gratefully acknowledged.

REFERENCES

AACC. (1999). AACC method 39-00: Near-infrared methods: Guidelines for model development and maintenance. In *Approved Methods of the American Association of Cereal Chemists*, 10th ed. St. Paul, Minnesota: AACC.

- Baker, S. Herrman, T., & Fairchild, F. (1997, January). *Capability of Kansas grain elevators to segregate wheat during harvest* (Report No. 781).
- Brumm, T. J., & Hurburgh, C. R., Jr. (1990). Estimating the processed value of soybeans. *Journal of the American Oil Chemists' Society*, 67, 302-307.
- 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., Brumm, T.J., Guinn, J.M., and Hartwig, R.A. (1990). Protein and oil patterns in U.S. and world soybean markets. *Journal of the American Oil Chemists' Society*, 67, 966-973. DOI: 10.1007/BF02541859
- 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.
- Hurburgh, C. R., Jr., Paynter, L. N., & Schmitt, S. G. (1987). Quality characteristics of Midwestern soybeans. *Applied Engineering in Agriculture*, 3, 159-165.
- Laux, C.M., Mosher, G.A. and Hurburgh, C.R. (2015). Application of quality management systems to grain handling: An inventory management case study. *Applied Engineering*, 31(2), 313-321.
- Medic, J., Atkinson, C., and Hurburgh, C.R. 2014. Current knowledge in soybean composition. *Journal of the American Oil Chemists' Society*, 91, 363-384. DOI: 10.1007/s11746-013-2407-9
- 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>
- R: The R Project for Statistical Computing [Computer software]. (2019).
- Ripke, G.R., Hardy, C.L., Hurburgh, C.R., & Brumm, T.J. (1995). *Calibration and field standardization of Tecator Infratec analyzers for corn and soybeans*. Proceedings from the 7th International Conference on Near Infrared Spectroscopy, August 6-11, Montreal, Canada.
- 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.).

Thakur, M. and Hurburgh, C.R. (2007). Quality of US soybean meal compared to the quality of soybean meal from other origins. *Journal of the American Oil Chemist's Society*, 84, 835-843. DOI: 10.1007/s11746-007-1107-8

Updaw, N. J., Bullock, J. B., & Nichols, T. E., Jr. (1976). Pricing soybeans on the basis of protein and oil content. *Southern Journal of Agricultural Economics*, 129-132.

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->

United States Soybean Quality: Annual Report 2015. 2016. Retrieved from: <https://ussec.org/wp-content/uploads/2016/01/2015-Final-US-Commodity-Soybean-Quality-Report.pdf>

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

Wagner, K. P. (2017). A mass balance based model to evaluate the impact of amino acid profiles on the feeding and processed value of soybeans. *Graduate Theses and Dissertations*. 16233. DOI: <https://doi.org/10.31274/etd-180810-5862>.

Author 1

First name or initial	Middle name or initial	Surname	Suffix (Jr., III, etc.)	Role (job title, etc.)	Email (and phone for contact author)	Contact author? yes or no
Bennett	E.	Barr		Former Research Assistant		No

Affiliation for Author1

Organization	Address	Country	URL or other info.
Iowa State University	3144 NSRIC, 1029 N. University Blvd., Ames, IA 50011	USA	

Author 2

First name or initial	Middle name or initial	Surname	Suffix (Jr., III, etc.)	Role (job title, etc.)	Email (and phone for contact author)	Contact author? yes or no
Charles	R.	Hurburgh		Professor	tatry@iastate.edu	No

Affiliation for Author 2

Organization	Address	Country	URL or other info.
Iowa State University, Department of Agricultural & Biosystems Engineering	3167 NSRIC, 1029 N. University Blvd., Ames, IA 50011	USA	

Author 3 (repeat Author and Affiliation tables for each author)

First name or initial	Middle name or initial	Surname	Suffix (Jr., III, etc.)	Role (job title, etc.)	Email (and phone for contact author)	Contact author? yes or no

Gretchen	A.	Mosher		Associate Professor	gamosher@iastate.edu	Yes
----------	----	--------	--	------------------------	----------------------	-----

Affiliation for Author 3

Organization	Address	Country	URL or other info.
Iowa State University, Department of Agricultural & Biosystems Engineering	3331 Elings Hall, 605 Bissel Road, Ames, IA 50011	USA	