Study of Soybeans of Other Colors and the Impact on End-Use Functionality in 2021-2022 Market Samples

Yun Tao and Timothy Norden

United States Department of Agriculture
Agricultural Marketing Service
Federal Grain Inspection Service
Technology and Science Division

November 8, 2022

Executive Summary

One-hundred officially-graded yellow soybean samples containing Soybeans of Other Colors (SBOC) were collected in July-August of 2022. The number of samples requested from each official agency was set to be representative of soybean production and geographically diverse. The samples contained original inspection results ranging from 2.0% – 6.9% SBOC that equate to grades of U.S. number 3 and U.S. number 4 yellow soybeans for all but 7 samples that were U.S. number 2. These samples were analyzed for protein and oil on the FGIS master near infrared transmission (NIRT) instruments both with and without the SBOC portions. A statistical comparison of the differences for the 100 sample pairs resulted in the finding that no significant differences in official protein or oil results could be found.

Study of Soybeans of Other Colors and the Impact on End-Use Functionality in 2021-2022 Market Samples

Introduction

The Grain Inspection Advisory Committee asked the USDA Agricultural Marketing Service (AMS) Federal Grain Inspection Service (FGIS) to provide independent research to determine if functionality is impacted by seed coat discoloration associated with current commercially available genetically engineered soybeans. To answer this question, FGIS gathered market samples representing officially inspected soybean lots from the 2021 crop year collected in July-August 2022 and compared the protein and oil results with and without the Soybeans of Other Colors (SBOC) portion. The study assumes that end-use functionality can be assessed by analyzing for protein and oil content, which are key nutritional components that establish the market value of soybeans.

In July-August of 2022, one-hundred market samples were collected by the Domestic Inspection Operations Office by requesting the file samples for officially graded yellow soybean samples that contained SBOC levels greater than 2.0%. Samples collected would grade U.S. number 3 or greater based on the requested SBOC percentage. SBOC portions from the samples collected were expected to range from 2.1% to 10%, which translates to a minimum of 21 g and a maximum of 100 g based upon a 1000 g file sample. The actual samples received ranged from 2.0% to 6.9% SBOC. Samples with greater than 10% SBOC would be graded "sample grade" and no samples in this range were received for this study. The target number of samples requested from each official agency was set to be representative of soybean production and geographically diverse.

Since the approved near infrared transmission (NIRT) instruments represent the standard for official protein and oil measurement for soybeans, FGIS will investigate any differences that are directly related to the SBOC. If no differences in protein or oil are found in the 2021-2022 market samples using the official NIRT instruments, it can be concluded that SBOC at levels of 2% to 7% do not influence the protein or oil results in yellow soybeans.

Statistical Approach

One hundred market samples from the 2021 crop year were analyzed on the two FGIS master FOSS 1241 NIRT instruments (A & B). The weights of the market samples were, on average, 1154 grams. The market samples were evaluated by the FGIS Board of Appeals and Review (BAR) and SBOC was removed to create a separated sound portion for each market sample. The market samples had, on average, 27.91 grams of SBOC (2.55%). The separated sound portions (Picked), as well as the same sound portions with 30 grams randomly taken out (Picked30), are derived from the market samples. The comparison between the latter two (Picked vs Picked30) serves as the control for the comparison between the first two (Market vs Picked), because the Picked sample is 27.91 grams on average less than the market samples and the NIRT results may be sensitive to the sample mass if the protein and oil results are heterogeneous. The two master NIRT instruments (A and B) were used for all measurements. A total of 1200 data points

were collected (100 samples × 3 sample types (Market, Picked & Picked30) × 2 instruments (A & B) × 2 factors (Protein & Oil)).

The statistical analysis aims to answer two questions:

(1) Whether there are statistically detectable differences for protein and oil between the Market and Picked samples?
(2) If any, are the differences practically meaningful?

We first apply Linear Mixed Model (LMM) to get a general evaluation of the datasets of protein and oil (each with 600 data points) to assess whether Sample Types (Market, Picked and Picked30) and Instruments (A and B) have any statistically significant effects. Because the factors Sample Type and Instrument are nested within Sample ID, that is, the protein and oil measurements are obtained and analyzed in blocks of size 1 sample, conventional techniques of *a posteriori* tests such as Tukey HSD are not applicable. As for follow-up analysis we will apply paired *t*-tests to make various direct pairwise comparisons, particularly, between the Market and Picked samples. Whenever multiple non-independent comparisons are being made, it is necessary to correct the critical p values that will determine the statistical significance. In this case, we will use the correction proposed by Bonferroni in the statistical analyses.

Analytic Results and Interpretation

1. Data Overview

   The set of measurement results of 600 data points, each for protein and oil, from 100 sets of samples (Market, Picked and Picked30 on instruments A and B) are plotted for visual inspection (Figures 1-2). The samples are ordered by their mean values of measurements and there are no recognizable patterns emerging from the visual presentations, thus giving the impression that no systematic differences will be found among the three sample types (Market, Picked and Picked30) or between the two instruments (A and B).

   In Figures 3-4 the differences between the Market and Picked samples are plotted against the mean protein and oil result. The average of the two instruments (A and B) were used to calculate the differences. From these plots it is clear there is no trend or pattern across the range of protein and oil results studied. This suggests that no systematic differences will be found between the Market and Picked samples.

2. ANOVA & VCA Results

   A Linear Mixed Model (LMM) is set up as below:

   Substance [Protein or Oil] ~ (SampID)/SampleType + (SampID)/Instrument (1),

where factor enclosed in round brackets is of random effect and that not enclosed is of fixed effect.

The *lm* function from the R/Stats package and the *anovaMM* function from the R/VCA package are used for estimating variance components and *F* tests and the respective results are combined into Tables 1 and 2.

For protein (Table 1), no statistically significant effect of Sample Type and Instrument on the measurement are detected, albeit a strong effect from Samples (SampID) is found. Quantitatively speaking, the predominant (93.58%) variance component is attributed to Samples, while that from Sample Type and Instrument, 0.54% and 0.14%, respectively, are only minuscule. In comparison, the error term (residuals) contributes 5.74% variance component, suggesting that the measurement error from NIRT is at least one order of magnitude larger than that resulting from the two factors Sample Type and Instrument.

For oil (Table 2), similar interpretation can be made as for protein but there is a distinctive difference, which is the statistically significant effect from Instrument, suggesting that the two instruments have some different results on the same samples for oil, though still the error term appears to be a bigger contributor to variance than instrument.

Both results imply that no significant difference in protein and oil is found between Market vs Picked vs Picked30 samples.

Table 1. ANOVA and VCA of NIRT Results of Protein

	DF	Sum Square	Mean Square	Est. Var. Comp.		sd	F value	P (>F)
SampID	99	163.8648	1.6552	0.2724	93.58%	0.5219	99.0841	0.0000***
SampID:SampleType	200	3.9653	0.0198	0.0016	0.54%	0.0395	1.1869	0.1133
SampID:instrument	100	1.7960	0.0180	0.0004	0.14%	0.0205	1.0751	0.3305
Residuals	200	3.3410	0.0167	0.0167	5.74%	0.1292	NA	NA

***P < 0.001

Table 2. ANOVA and VCA of NIRT Results of Oil

	DF	Sum Square	Mean Square	Est. Var. Comp		sd	F value	P (>F)
SampID	99	92.5237	0.9346	0.1540	94.48%	0.3924	170.1283	0.0000***
SampID:SampleType	200	1.1708	0.0059	0.0002	0.11%	0.0134	1.0657	0.3267
SampID:instrument	100	1.0329	0.0103	0.0016	1.00%	0.0401	1.8802	0.0001***
Residuals	200	1.0987	0.0055	0.0055	3.41%	0.0741	NA	NA

***P < 0.001

3. Paired t-test Results.

3.1. Basic statistical formulas

Define d_i as the difference of paired values of x_{ki} ($k = 1,2; i = 1,2, \dots, n$ for n samples), the following metrics can be calculated:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \quad (2),$$

$$Var(d_i) = s_{d_i}^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} d_i^2 - n \times \bar{d}^2\right) \quad (3).$$

The 95% confidence interval for $\bar{d}$ can be calculated with formula

$$CI = \bar{d} \pm t_{0.975,n-1} \times \frac{s_d}{\sqrt{n}} \quad (4).$$

For any individual pair of measurement (not the mean of d_i discussed in eqs. 2-4), the expected width of this comparison can be calculated using prediction interval (eq. 5):

$$PI = \bar{d} \pm t_{0.975,n-1} \times s_d \times \sqrt{1 + \frac{1}{n}} \quad (5),$$

which is roughly equivalent to two commonly used formulas (eqs. 6 & 7) for the same purpose (Eq. 7 is often used when n is large):

$$CI = \bar{d} \pm t_{0.975,n-1} \times s_d \quad (6),$$

$$CI = \bar{d} \pm z_{0.975} \times s_d \quad (7).$$

3.2. Interpretation of Paired t-tests of Protein and Oil

Various paired t-tests have been done to "zoom in" some details further from *lm* and *anovaMM* analyses (Tables 3 & 4). Since there is a total of $k = 9$ paired t-tests, so a Bonferroni correction is made:

$$\alpha' = \frac{\alpha}{k} = 0.0055 \text{ if } \alpha = 0.05 \quad (8).$$

Keep in mind that Bonferroni correction is well known for its conservative nature—that is, test with Bonferroni tends to reject the null hypothesis less often, so we will also flag and discuss cases where p values are slightly larger than 0.0055 (marginally significant).

We did find there are significant differences in 3 cases and 1 marginal difference between instruments A and B: one Market, and one Picked (marginal) protein comparison (Table 3), as well as one Picked, and one Picked30 oil comparison (Table 4). These 4 cases are highlighted in yellow in Tables 3 & 4. These cases provide support for the significant contribution of variance from instrument found in the ANOVA and VCA analyses.

However, there is no statistically significant difference found between Market and Picked samples (our main question), nor between Picked vs Picked30 samples (control), for measurements made on instrument A or B alone, or the average of the two instruments. This finding suggests that instrument itself is less precise/accurate than that required to measure the subtle difference between Market and Picked samples, even with 100 samples in the experimental setup.

There is only one somewhat marginally significant difference between Market and Picked as measured by Instrument B for protein (p =0.060, without Bonferroni correction). The two types of samples have a mean difference 0.035% (Market has less protein than Picked) and the 95% confidence interval (CI) is (-0.070, 0.001), with CI Width=0.074. With Bonferroni correction (eq. 8), $t_{0.997,n-1}$ will replace $t_{0.975,n-1}$ in Eq. 5 and the newly calculated CI will be -0.086 – 0.017, and the CI width is now 0.103. Either way, the mean difference is very subtle and practically almost nonexistent, and the CI is very narrow. The experiment setup based on 100 samples can detect, at most, ~ 0.1% difference in protein or oil between two similar sample types.

If one sample is measured on two instruments and compared, the expected width of this comparison (on individual samples) can be calculated using prediction interval (eq. 5).

From Table 3 we can see the widths of PI range from 0.525 to 0.780, one order of magnitude larger than the marginally significant CI width 0.074. In other words, for individual samples, the NIRT instruments do not have the resolution power to detect the subtle difference (even it exists) between Market and Picked samples.

A similar argument can be made for oil.

Conclusion

Based on 100 market samples with SBOC content from 2.0% - 6.9%, the mean differences of protein and oil from the 100 market samples are not statistically different from zero.

Table 3. Paired t-test results for various comparisons of protein

Comparison		Mean Diff ($\bar{d}$)	sd	PI (95%)			Paired t-test and CI for $\breve{d}$				
				Low	Upper	Width	t	p	Low	Upper	Width
Average of Instruments A & B	Market vs Picked	-0.017	0.145	-0.307	0.273	0.580	-1.176	0.243	-0.046	0.012	0.058
	Picked vs Picked30	0.006	0.132	-0.257	0.268	0.525	0.418	0.677	-0.021	0.032	0.052
Instrument A	Market vs Picked	0.000	0.196	-0.390	0.390	0.780	0.015	0.988	-0.039	0.039	0.078
	Picked vs Picked30	0.013	0.184	-0.355	0.380	0.735	0.678	0.499	-0.024	0.049	0.073
Instruments B	Market vs Picked	-0.035	0.181	-0.396	0.327	0.723	-1.903	0.060	-0.070	0.001	0.072
	Picked vs Picked30	-0.002	0.187	-0.373	0.370	0.743	-0.081	0.936	-0.038	0.035	0.074
Instruments A vs B	Market vs Market	0.079	0.180	-0.281	0.439	0.720	4.389	0.000*	0.043	0.115	0.072
	Picked vs Picked	0.044	0.169	-0.292	0.381	0.672	2.634	0.010^	0.011	0.078	0.067
	Picked30 vs Picked30	0.030	0.183	-0.335	0.396	0.730	1.660	0.100	-0.006	0.067	0.073

*Significant after Bonferroni correction; ^Marginally significant

Table 4. Paired t-test results for various comparisons of oil

Comparison		Mean Diff ($\bar{d}$)	sd	PI (95%)			Paired t-test and CI for $\breve{d}$				
				Low	Upper	Width	t	p	Low	Upper	Width
Average of Instruments A & B	Market vs Picked	-0.011	0.076	-0.161	0.140	0.301	-1.421	0.159	-0.026	0.004	0.030
	Picked vs Picked30	0.001	0.075	-0.148	0.150	0.298	0.181	0.857	-0.013	0.016	0.030
Instrument A	Market vs Picked	-0.017	0.114	-0.244	0.210	0.454	-1.497	0.138	-0.040	0.006	0.045
	Picked vs Picked30	-0.006	0.107	-0.219	0.208	0.428	-0.513	0.609	-0.027	0.016	0.043
Instrument B	Market vs Picked	-0.004	0.097	-0.197	0.188	0.385	-0.455	0.650	-0.024	0.015	0.038
	Picked vs Picked30	0.008	0.098	-0.188	0.204	0.392	0.834	0.406	-0.011	0.028	0.039
Instruments A vs B	Market vs Market	0.024	0.126	-0.228	0.276	0.504	1.878	0.063	-0.001	0.049	0.050
	Picked vs Picked	0.036	0.104	-0.171	0.243	0.414	3.506	0.001*	0.016	0.057	0.041
	Picked30 vs Picked30	0.050	0.109	-0.167	0.267	0.434	4.601	0.000*	0.028	0.072	0.043

*Significant after Bonferroni correction

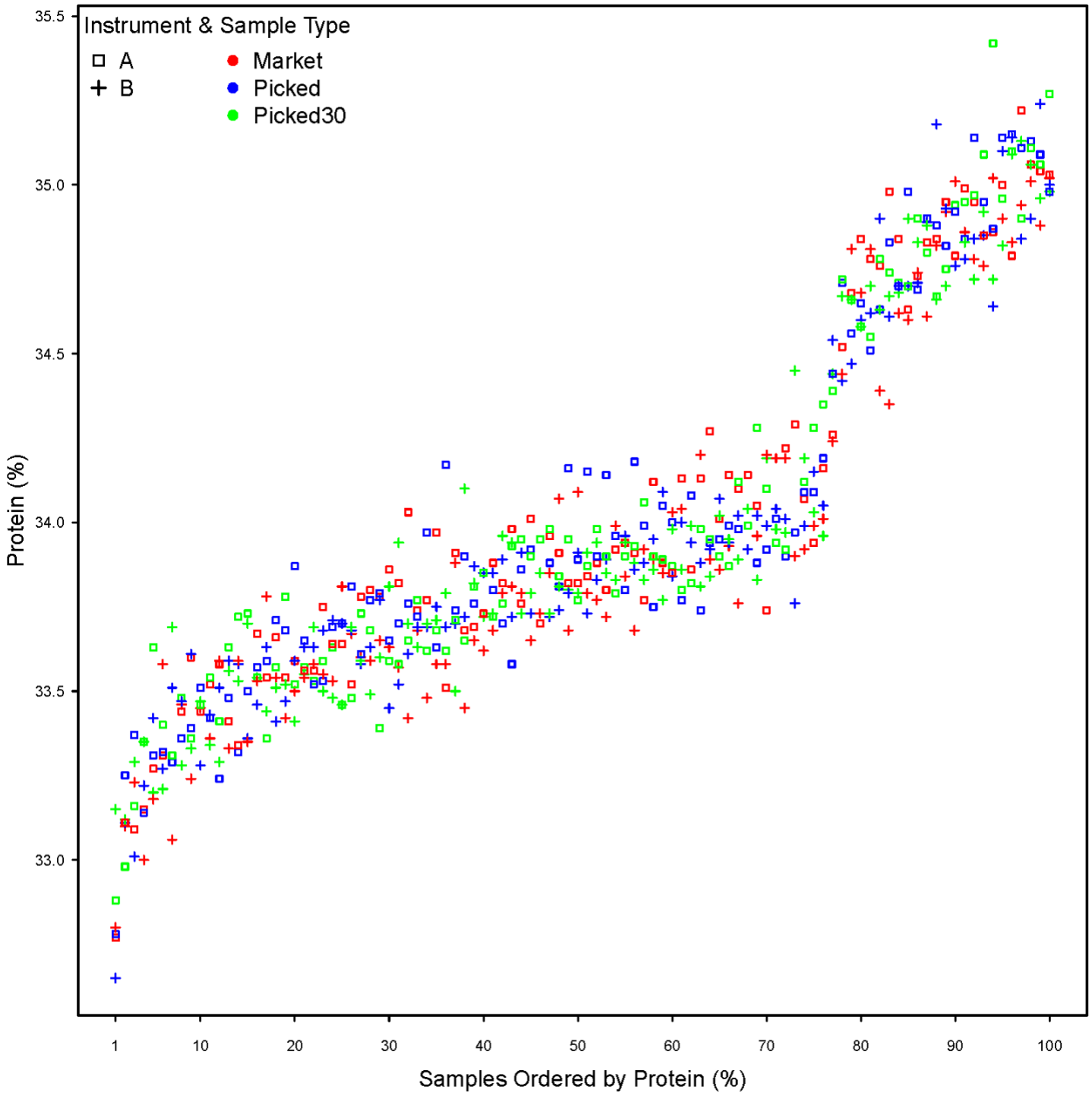

Figure 1. Graphic Display of the 600 Data Points from NIRT Protein Results. The 100 samples are ordered by average protein (%) of 6 NIRT results, which are from 3 Sample Types (Marked, Picked and Picked30) times 2 Instruments (A and B) and aligned on one of the 100 vertical lines (invisible) corresponding to the 100 ordered samples. No noticeable pattern can be seen for the NIRT results associated with Sample Type and Instrument.

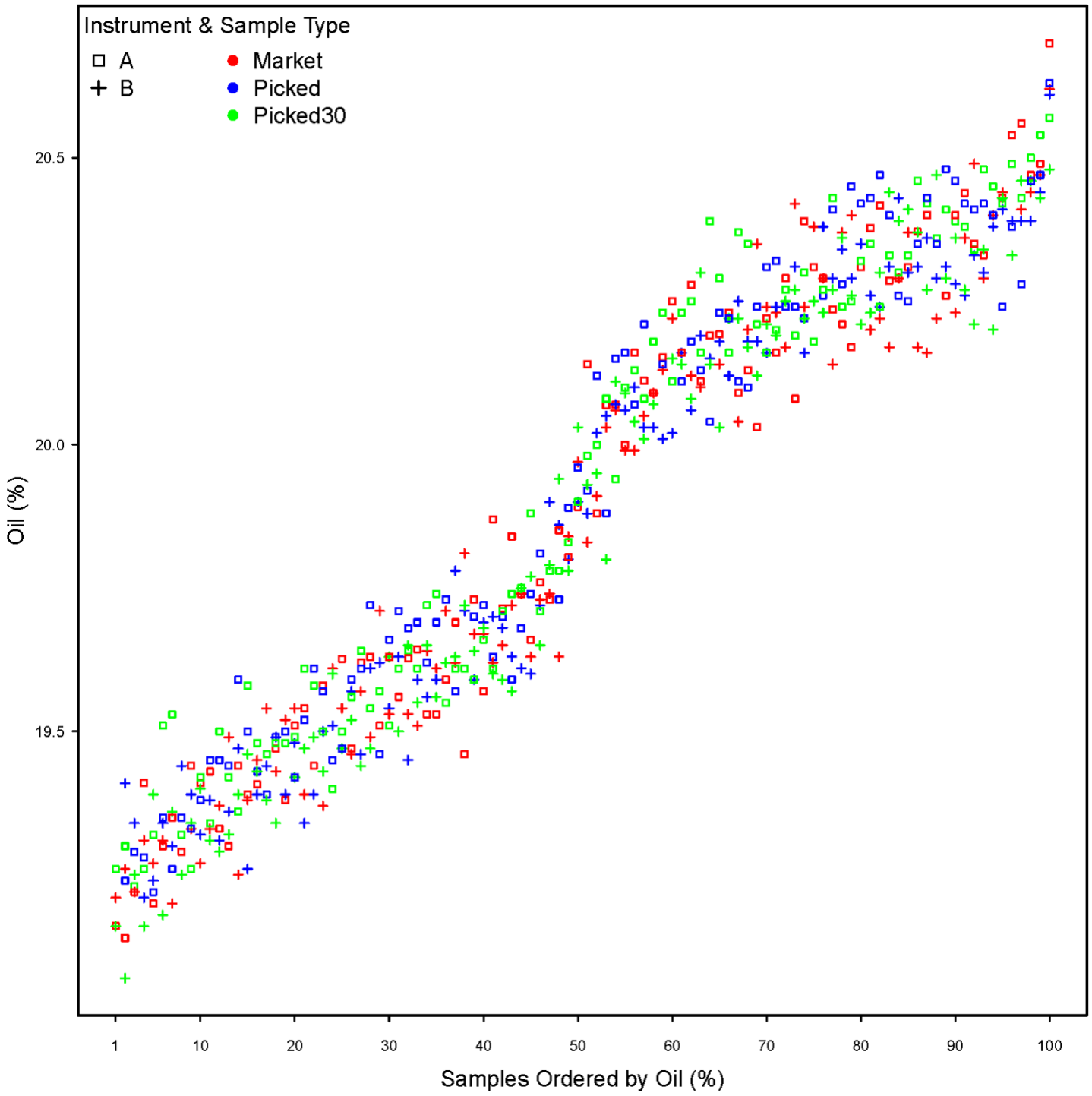

Figure 2. Graphic Display of the 600 Data Points from NIRT Oil Results. See Figure 1 for figure description.

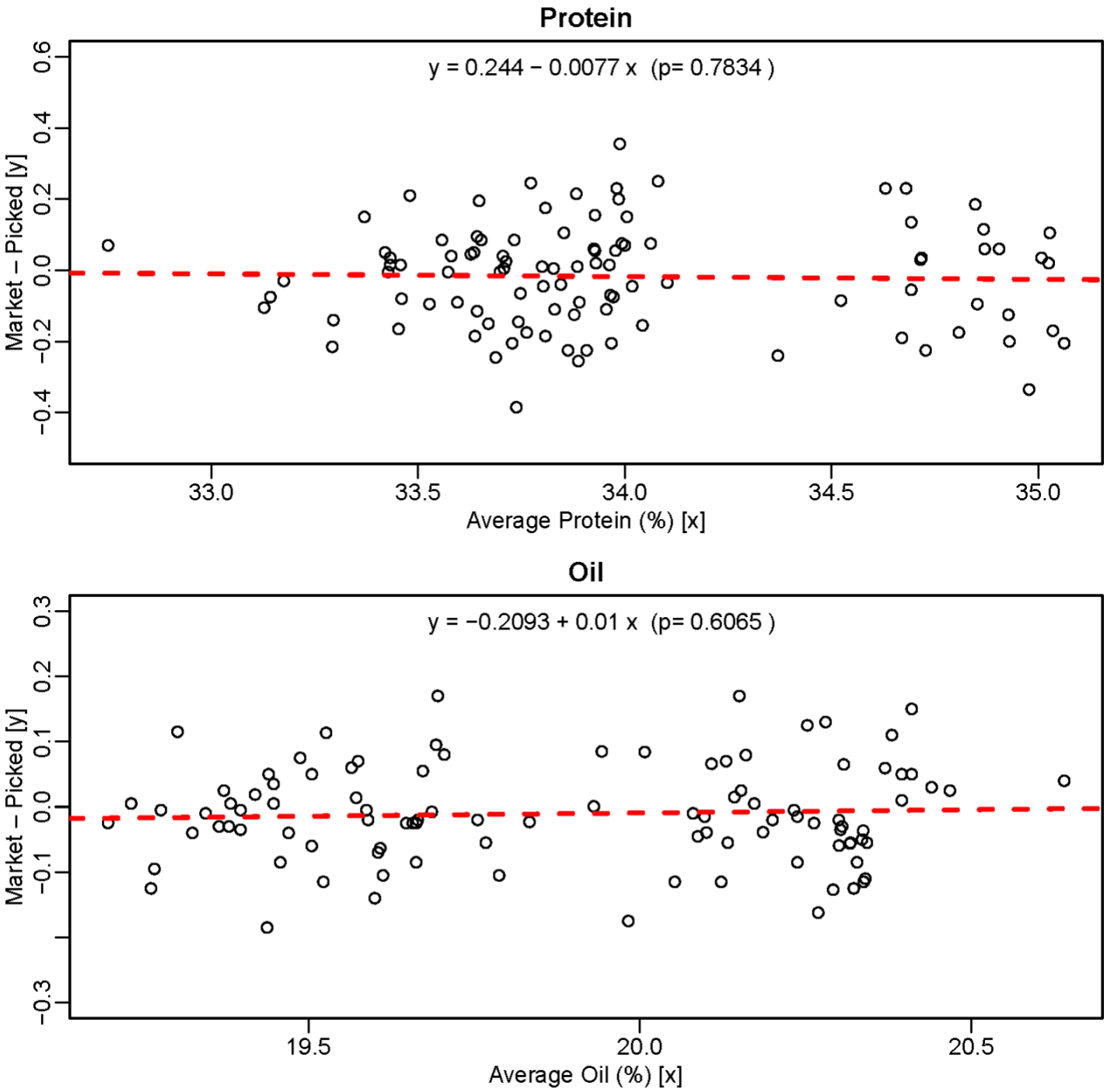

Figure 3. Bland-Altman Plot of the Protein (Upper Panel) and Oil (Lower Panel) Comparisons between Market and Picked Samples. Each individual value is an average from instruments A and B. The differences in protein or oil between Market and Picked samples (y) are regressed on the averages of the two samples (x) and the resulting regression functions are shown in the panels and are also drawn (red dashed lines). The p values in the regressions are for the slopes, both of which are not significantly different from 0.

Acknowledgments

The authors thank the Domestic Inspection Operations Office for collecting the samples, the Inspection Instrumentation Branch for running the NIRT analyses and the Board of Appeals and Review for providing the SBOC separations.