Soybean seed quality determination for small sample sizes through the use of a parabolic mirror cup and diode array spectrometer

S.L. Naeve,[a] J. Orf,[a] A. Killam,[a] W. Shadow[b] and D. Honigs[b]
[a]*University of Minnesota, 411 Borlaug Hall, St. Paul, MN 55108 USA.*
E-mail: naeve002@umn.edu
[b]*Perten Instruments Inc., 6444 S. Sixth Road, Springfield, IL 62712 USA*

Introduction

Soybean seed is highly prized worldwide due to its valuable primary constituents, protein and oil; however, increasingly many purchasers are interested in secondary constituents such as individual fatty acids, amino acids, and other components such as isoflavones and soluble carbohydrates. In an effort to breed and produce soybeans with greater value to the primary purchaser and the final consumer, it is essential to have the means to quickly and cheaply evaluate breeding lines and other soybean samples for quality characteristics. However, individual soybean plants may hold only 100–200 seeds. Because whole seed must be retained for planting, a very limited number of seeds are available to be ground and analysed. In addition, thousands of breeding lines must be evaluated annually, making traditional chemical analysis too costly and slow. A rapid, inexpensive, and nondestructive analytical method that is capable of analysing very small seed lots is essential to screen soybean germplasm quickly for appropriate quality traits. Therefore, we have been investigating the utility of a prototype mirrored sample cup [Micro Mirror Module (Perten Instruments, Stockholm Sweden)].

Materials and methods

Five unique seed lots (designated M1-M5) were identified that differed in protein, oil, and linolenic acid concentrations, seed size, and pedigree (Table 1).

A sample of approximately 90 g of each seed lot was subjected to near infrared spectroscopy analysis for protein and oil with a Perten DA 7200 Feed Analyser (Perten Instruments, Stockholm, Sweden) using a standard small metal cup with a 75 mm diameter. Each sample was scanned and repacked 20 times to provide average reference quality profiles. Samples were subsequently divided into 12–15 subsamples for analysis on the Micro Mirror Module. Subsamples consisted of 8 to 14 seeds depending on seed size. Each subsample was scanned and repacked 15 times. Data were analysed using the GLM and MIXED procedures of SAS v9.1.[1] Standard deviation and error

Table 1. Identities, seed size, and special characteristics of five distinct soybean seed lots used in this experiment.

Seed Lot	Identity	Feature	Seeds 100 grams^{-1}	mg seed^{-1}
M1	MN1806SP	large seed, high protein	388	258
M2	MN1702SP	low linolenic oil	673	149
M3	M129	non-nodulating	782	128
M4	MN1410	normal seed size	516	194
M5	MN1203SP	small seed size	851	118

variance terms were then used to investigate error at three main levels; 1) scan to scan error of the 90 g sample in the standard small cup, 2) sampling error as estimated by the 12–15 sub-samples per seed lot, and 3) scan to scan error through repacks and scans of the same 8–14 seed subsamples. Results are provided in Tables 2–5.

Results

Mean and standard deviation values were determined for nine soybean seed constituents of five distinct seed lots (Table 2).

Table 2. Mean and standard deviation values for nine soybean seed constituents of five distinct seed lots consisting of approximately 90 g of whole seed, and scanned using the small metal cup and repacked 20 times each.

Protein				Oil		Fibre		Ash		Palmitic	
Seed lot	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
M1	20	48.12	0.22	15.55	0.23	4.88	0.10	5.26	0.02	9.17	0.49
M2	20	41.68	0.23	19.76	0.22	5.44	0.07	5.16	0.02	10.22	0.39
M3	20	28.59	0.41	23.02	0.25	6.41	0.10	5.07	0.03	12.84	0.54
M4	20	42.76	0.33	18.54	0.25	4.95	0.05	5.13	0.02	11.68	0.36
M5	20	39.42	0.37	19.50	0.19	5.93	0.07	5.12	0.02	13.31	0.38
		Stearic		Oleic		Linoleic		Linolenic			
Seed Lot	N	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.		
M1	20	3.64	0.05	23.09	0.87	56.01	0.69	7.62	0.55		
M2	20	4.12	0.09	23.94	0.81	58.34	0.76	3.07	0.69		
M3	20	4.57	0.06	15.55	0.70	54.36	0.65	11.76	0.61		
M4	20	4.60	0.05	17.48	0.65	55.51	0.77	11.18	0.57		
M5	20	4.40	0.04	15.48	0.56	55.75	0.64	10.76	0.70		

Each consisted of approximately 90 g of whole seed, and was scanned using the small metal cup and repacked 20 times each. The error terms represent the sum of error caused by sample heterogeneity, analytical, and machine error. Sample heterogeneity is a result of seed to seed variation in each of the constituents[2,3] as well as a non-uniform distribution of these constituents in each seed.[4] Standard deviation values varied by seed lot and constituent; however, large differences among constituent indicates that there are differences in levels of either heterogeneity within seed lots or in analytical precision between constituents.

Mean and standard deviation values were determined for nine soybean seed constituents of the five distinct seed lots using multiple scans of distinct subsamples (Table 3).

Sub-samples of 8–14 seeds each were scanned using the Micro Mirror Cup 20 times with repacks. In this instance, means and standard deviation values represent averages of within sub-samples and across sub-samples. Large within sub-sample standard deviation values indicate that re-orienting seeds within the Mirror Cup affects the result. This is likely due to the non-uniform distribution of soybean compositional factors within each seed. Large standard deviation values indicate large sample heterogeneity and therefore large sampling error. Individual 8–14 seed sub-samples are not likely to precisely represent a larger sample of soybean seed.

In order to partition error associated with sub-sampling and scanning, the percent of total variance attributed to either sub-sampling error or scan error within five soybean seed lots was determined (Table 4).

When examined across subsamples, the error associated with sub-sampling and scanning can be evaluated relative to total variation among the five samples. Table 4 presents the percent of total variance attributed to the sample itself, sub-samples within samples, or scans within sub-samples across five soybean seed lots scanned using the Mirror Cup. The data indicate that among a diverse set of samples, variation among samples is by far the greatest source of variation for all constituents except linoleic acid. Examination of variance attributed to either sub-samples (sampling error) or scanning error indicates that sampling error plays a very large role in affecting total variance.

The Micro Mirror Cup is useful for very small samples, yet accuracy similar to that of the standard small cup is often required. Table 5 provides estimates of the number of samples and repeated scans of each sample with the Micro Mirror Cup required to match the precision of the standard small cup.

Confidence intervals are based on 95% probability of accuracy within provided ranges. Minimum samples and scans for the Micro Mirror Cup, to meet the precision of the small cup, are provided based on two criteria; 1) combinations of samples and scans of each sample that results in the lowest number of total scans, and 2) combinations of samples and scans resulting in the lowest number of samples required. While additional samples are required to increase the precision of the Mirror Cup for estimating protein, multiple scans of the same sample greatly increase the precision of the instrument to predict linolenic acid concentrations.

Table 3. Mean and standard deviation values for nine soybean constituents of five distinct seed lots. Sub-samples of 8-14 seeds each were scanned using the Micro Mirror Cup 20 times with repacks. Means and standard deviation values here represent averages of within sub-samples and across sub-samples.

Seed Lot	Sub-sample	Scans	Protein			Oil		
			Mean	Std Dev. (within)†	Std Dev. (across)†	Mean	Std Dev. (within)†	Std Dev. (across)†
M1	15	20	48.94	0.48	0.62	16.57	0.46	0.31
M2	15	20	42.22	0.67	0.63	20.85	0.50	0.31
M3	14	20	28.85	0.67	0.93	24.07	0.48	0.23
M4	12	20	43.48	0.37	0.92	19.21	0.38	0.51
M5	14	20	40.16	0.35	0.70	20.67	0.37	0.23
Seed Lot	Sub-sample	Scans	Fiber			Ash		
			Mean	Std Dev. (within)	Std Dev. (across)	Mean	Std Dev. (within)	Std Dev. (across)
M1	15	20	4.57	0.17	0.16	5.39	0.04	0.04
M2	15	20	5.26	0.15	0.09	5.26	0.04	0.05
M3	14	20	6.42	0.15	0.11	5.16	0.05	0.05
M4	12	20	4.97	0.10	0.08	5.28	0.03	0.05
M5	14	20	5.95	0.12	0.16	5.23	0.04	0.03
Seed Lot	Sub-sample	Scans	Palmitic			Stearic		
			Mean	Std Dev. (within)	Std Dev. (across)	Mean	Std Dev. (within)	Std Dev. (across)
M1	15	20	9.31	0.84	0.33	3.70	0.12	0.05
M2	15	20	12.66	1.00	1.00	4.78	0.19	0.32
M3	14	20	12.90	0.83	0.60	4.65	0.11	0.03
M4	12	20	12.96	0.63	0.44	4.63	0.09	0.07
M5	14	20	14.58	0.70	0.65	4.46	0.09	0.06
Seed Lot	Sub-sample	Scans	Oleic			Linoleic		
			Mean	Std Dev. (within)	Std Dev. (across)	Mean	Std Dev. (within)	Std Dev. (across)
M1	15	20	21.86	1.56	1.09	56.38	1.37	0.84
M2	15	20	18.39	1.84	1.63	57.43	1.50	1.30
M3	14	20	14.06	1.59	1.15	55.50	1.32	0.43
M4	12	20	15.28	1.11	0.73	55.03	1.08	0.67
M5	14	20	11.68	1.20	1.57	57.05	1.17	0.70

Table 3. (continued)

Seed Lot	Sub-sample	Scans	Linolenic					
			Mean	Std Dev. (within)	Std Dev. (across)			
M1	15	20	8.69	1.42	0.45			
M2	15	20	7.13	1.40	1.46			
M3	14	20	12.41	1.22	0.50			
M4	12	20	12.76	1.04	0.61			
M5	14	20	11.94	1.11	0.86			

† = within or across sub-samples

Table 4. Percent of total variance attributed to either sub-sampling error or scan error within five soybean seed lots or percent of total variance attributed to the sample itself, sub-samples within samples, or scans within sub-samples across five soybean seed lots scanned using the Mirror Cup.

Seed		Protein	Oil	Fiber	Ash	Palmitic	Stearic	Oleic	Linoleic	Linolenic
Lot	Covariance Parameter	Percent of Total Variance								
M1	Split	59.8	27.6	41.5	44.5	9.0	12.4	28.6	23.6	4.7
	Scan	40.2	72.4	58.5	55.5	91.0	87.6	71.4	76.4	95.3
M2	Split	43.5	23.8	22.4	52.1	47.5	70.3	40.9	39.2	48.8
	Scan	56.5	76.2	77.6	47.9	52.5	29.7	59.1	60.8	51.2
M3	Split	58.4	13.9	29.4	46.4	31.3	3.8	30.8	4.9	9.8
	Scan	41.6	86.1	70.6	53.6	68.7	96.2	69.2	95.1	90.2
M4	Split	83.1	61.0	35.1	65.2	29.2	36.0	27.3	24.6	22.6
	Scan	16.9	39.0	64.9	34.8	70.8	64.0	72.7	75.4	77.4
M5	Split	78.0	23.2	59.7	39.8	39.3	24.2	58.3	23.9	29.8
	Scan	22.0	76.8	40.3	60.2	60.7	75.8	41.7	76.1	70.2
	Seed Lot	98.4	96.0	94.1	67.2	77.0	81.3	79.7	28.6	72.8
	Split (Sample)	1.0	1.2	2.3	16.3	8.1	10.6	7.8	18.8	7.7
	Scan	0.6	2.8	3.6	16.6	14.9	8.1	12.4	52.6	19.5

Table 5. Number of samples and scans of each sample required to meet various confidence intervals for selected soybean constituents for both the standard small cup and the Micro Mirror Cup. Confidence intervals are based on 95% probability of accuracy within provided ranges.

		Protein				Oil				Linolenic			
Seed Lot	Cup	±0.5%		±1.0%		±0.5%		±1.0%		±1.0%		±1.5%	
		n†	# Sc§	*n*	# Sc	*n*	# Sc	*n*	# Sc	*n*	# Sc	*n*	# Sc
M1	Small	1	1	1	1	1	1	1	1	2	1	1	1
	Mirror - Smallest # Total Scans	10	1	3	1	5	1	2	1	5	1	2	2
	Mirror - Smallest # Samples	8	2	2	2	2	6	1	2	3	3	1	5
M2	Small	1	1	1	1	1	1	1	1	2	1	1	1
	Mirror - Smallest # Total Scans	10	2	3	2	6	1	2	1	10	4	7	1
	Mirror - Smallest # Samples	8	3	2	3	4	2	1	2	10	4	5	3
M3	Small	3	1	1	1	1	1	1	1	2	1	1	1
	Mirror - Smallest # Total Scans	–	–	6	1	5	1	2	1	4	2	3	1
	Mirror - Smallest # Samples	–	–	5	2	3	2	1	2	2	5	1	4
M4	Small	2	1	1	1	1	1	1	1	2	1	1	1
	Mirror - Smallest # Total Scans	–	–	4	1	6	2	2	1	4	2	3	1
	Mirror - Smallest # Samples	–	–	4	1	5	3	2	1	2	8	1	4
M5	Small	3	1	1	1	1	1	1	1	2	1	1	1
	Mirror - Smallest # Total Scans	10	1	3	1	4	1	1	1	8	1	4	1
	Mirror - Smallest # Samples	8	5	2	3	3	2	1	1	5	2	2	3

† *n* = number of samples; § # *Sc* = number of scans

Conclusions

- This study examines the precision of the Perten Micro Mirror Cup, which can be used to evaluate very small soybean seed lots for primary constituents relative to the standard small cup.

- The study does not evaluate accuracy of either cup relative to traditional wet chemical analysis.
- We found that the Mirror Cup provides a confidence interval of about 2x that of the standard small cup.
- We found significant error between 8–14 seed sub-samples. This sampling error is likely due, in large part, to seed-to seed heterogeneity within seed lots.
- For well-characterised constituents (protein and oil), total error variance for sample analysis with the Mirror Cup was 4% or less. Ninety-six percent or more of the error variance for these constituents was due to variation between seed lots.
- Variance due to sampling (sampling error) is a major source of error when analysing seed lots using a small number of seeds. This error can be overcome by analysis of multiple sub-samples from the same seed lot.
- For protein estimation, 2–4 samples are required to meet a $\pm$1% confidence interval when using the Mirror Cup. For oil, a $\pm$1% confidence interval can be met with only 1 or two samples.
- We have demonstrated that the Micro Mirror Cup can be used to evaluate a very small number of whole soybean seeds for common quality parameters with a level of precision that nears that of traditional techniques utilising large samples. However, we determined that multiple sub-samples may be required to achieve a necessary level of precision in certain circumstances.

References

1. SAS Institute, *The SAS system for Windows. 9.1*., SAS Inst., Cary, NC, USA (2002).
2. R.A. Illipronti, Jr, W.J.M. Lommen, C.J. Langerak and P.C. Struik, *Neth. J. Agric. Sci*. **48,** 165 (2000).
3. F.I. Collins and J.L. Cartter, *Agron. J.* **48,** 216 (1956).
4. J.B. Carlson and N.R. Lersten, in *Soybeans: Improvement, Production, and Uses.* 3rd Edn, Ed by H.R. Boerma and J.E. Specht. ASA and SSSA, Madison, WI, USA, p. 59 (2004).