2011

Development of Single-Seed Near-Infrared Spectroscopic Predictions of Corn and Soybean Constituents using Bulk Reference Values and Mean Spectra

Paul R. Armstrong
United States Department of Agriculture

Jasper G. Tallada
United States Department of Agriculture

Charles R. Hurburgh Jr.
Iowa State University, tatry@iastate.edu

David F. Hildebrand
University of Kentucky

James V. Spencer
University of Nebraska–Lincoln

Development of Single-Seed Near-Infrared Spectroscopic Predictions of Corn and Soybean Constituents using Bulk Reference Values and Mean Spectra

Abstract
Rapid, non-destructive single-seed compositional analyses are useful for many areas of crop science, including breeding and genetics. Seeds are sometimes unique and require preservation due to small samples, which necessitates development of methods for total non-destructive measurement. Near-infrared reflectance spectroscopy (NIRS) can be used for non-destructive single-seed composition prediction, but the reference methods used to develop prediction models are usually destructive. Reference methods are costly, and extensive sets of seeds must be used to obtain prediction models for multiple constituents. In this research, single-seed NIRS prediction models were developed for common constituents of soybeans and corn using composition values from bulk reference measurement and respective averaged single-seed spectra as opposed to single-seed reference and spectra. The bulk reference model and a true single-seed model for soybean protein were also compared to determine how well the bulk model performs in predicting single-seed protein. This provided a basis for evaluating bulk model performance for other constituents. Bulk model statistics indicated that bulk models should perform well for soybean protein and oil, but not well for fiber; corn bulk models should perform well for protein, oil, starch, and seed density. Bulk model predictions of single-seed soybean reference protein show, at best, that bulk models work reasonably well, with a standard error of prediction (SEP) = 1.82%) compared to an SEP of 0.97% for a true single-seed protein model. Bias correction may be needed, though, depending how the bulk model is developed. Overall, the bulk models should be useful for selecting single seeds in breeding programs targeting specific composition traits and segregating individual samples based on composition.

Keywords
Corn, Near-infrared spectroscopy, Single seed, Soybean

Disciplines
Agriculture | Bioresource and Agricultural Engineering

Comments
This article is from *Transactions of the ASABE* 54 (2011): 1529–1535. Posted with permission.

Development of Single-Seed Near-Infrared Spectroscopic Predictions of Corn and Soybean Constituents using Bulk Reference Values and Mean Spectra

P. R. Armstrong, J. G. Tallada, C. Hurburgh, D. F. Hildebrand, J. E. Specht

Abstract. *Rapid, non-destructive single-seed compositional analyses are useful for many areas of crop science, including breeding and genetics. Seeds are sometimes unique and require preservation due to small samples, which necessitates development of methods for total non-destructive measurement. Near-infrared reflectance spectroscopy (NIRS) can be used for non-destructive single-seed composition prediction, but the reference methods used to develop prediction models are usually destructive. Reference methods are costly, and extensive sets of seeds must be used to obtain prediction models for multiple constituents. In this research, single-seed NIRS prediction models were developed for common constituents of soybeans and corn using composition values from bulk reference measurement and respective averaged single-seed spectra as opposed to single-seed reference and spectra. The bulk reference model and a true single-seed model for soybean protein were also compared to determine how well the bulk model performs in predicting single-seed protein. This provided a basis for evaluating bulk model performance for other constituents. Bulk model statistics indicated that bulk models should perform well for soybean protein and oil, but not well for fiber; corn bulk models should perform well for protein, oil, starch, and seed density. Bulk model predictions of single-seed soybean reference protein show, at best, that bulk models work reasonably well, with a standard error of prediction (SEP) = 1.82%) compared to an SEP of 0.97% for a true single-seed protein model. Bias correction may be needed, though, depending how the bulk model is developed. Overall, the bulk models should be useful for selecting single seeds in breeding programs targeting specific composition traits and segregating individual samples based on composition.*

Keywords. *Corn, Near-infrared spectroscopy, Single seed, Soybean.*

Whole-grain near-infrared spectroscopy has been successfully used in routine analysis of seed chemical composition of corn and soybeans using both near-infrared reflectance spectroscopy (NIRS) and near-infrared transmittance spectroscopy (NITS) methods (Rippke et al., 1995; Hardy et al., 1995). The instrumental methods are useful because they are non-destructive, fast, accurate, and repeatable. In addition, these methods require little to no sample preparation and simultaneously estimate constituent values such as moisture, protein, oil, starch, and fiber contents from a single measurement. However, in plant breeding or genetic applications, bulk analysis procedures cannot recognize individual seeds that contain special constituent levels that deviate from normal values (Baye et al., 2006). The ability to select these types of seeds allows further assessment and potential advancement to successive stages of breeding work.

Single-seed spectroscopy has been advanced to hasten varietal development by identifying seeds that possess specific quality traits in a segregating population. Substantial work has been done for wheat, corn, soybeans, and other seeds. Delwiche and Massie (1996) achieved good distinction between red and white classes of wheat using multiple linear regression and partial least-squares models from two spectral regions. In other work by Delwiche (1998), prediction models for single-seed protein contents of several classes of wheat attained standard errors of prediction (SEP) ranging from 0.462% to 0.720% and were obtained in models using the spectral range from 1100 to 1400 nm. Dowell et al. (2006) used NIRS to sort wheat seeds by protein content and hardness, and proso millet into amylose-bearing and amylose-free fractions.

Velasco et al. (1999) studied oil and fatty acid composition of rapeseed and found reasonably good correlation between NIRS of single seeds and reference measurements for oleic, erucic, linoleic, and linolenic constituents. Abe et al. (1996) reported SEP values ranging from 0.68% to 1.74% for soybean and wheat protein, respectively, using multiple linear regression models. To support their research on quantitative trait loci analysis for recombinant soybean inbred lines, Ta-

Submitted for review in April 2010 as manuscript number IET 8514; approved for publication by the Information & Electrical Technologies Division of ASABE in July 2011.

Mention of trademark or proprietary product does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products that may also be suitable.

The authors are **Paul R. Armstrong, ASABE Member,** Research Engineer, and **Jasper G. Tallada,** Post-Doctoral Researcher, USDA-ARS Engineering and Wind Erosion Research Unit, Center for Grain and Animal Health Research, Manhattan, Kansas; **Charles Hurburgh, ASABE Member,** Professor, Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa; **David F. Hildebrand,** Professor, Department of Plant and Soil Science, University of Kentucky, Lexington, Kentucky; and **James E. Specht**, Professor, Department of Agronomy and Horticulture, University of Nebraska, Lincoln, Nebraska. **Corresponding author:** Paul R. Armstrong, USDA-ARS Engineering and Wind Erosion Research Unit, Center for Grain and Animal Health Research, 1515 College Ave. Manhattan, KS 66502; phone: 785-776-2728; e-mail: paul.armstrong@ars.usda.gov.

juddin et al. (2002) developed single-seed soybean calibration models for prediction of protein and lipid contents by NITS in the range of 700 to 1100 nm. F_8 and F_9 generation seed samples were divided into two size groups (<6 mm and ≥6 mm seed diameter). The researchers obtained standard errors for predicting protein contents of 1.32% and 1.57% and correlation coefficients of 0.88 and 0.87 for the large and small seed classes, respectively, using four-term linear regression models. Baye et al. (2006) developed calibration models to predict seed composition of normal and mutant seeds from eight inbred lines of corn using reflectance and transmittance spectroscopy. They found that better predictive models were obtained from the absolute contents data than from the relative content data. Cogdill et al. (2004) concluded that moisture content was easier to predict than oil content in single-seed transmittance hyperspectral imaging for corn. Armstrong (2006) developed a single-seed NIR sorting instrument and showed its usefulness for prediction of moisture in corn and soybeans and protein content in soybeans. Standard error of cross-validation of models for protein content in three varieties of soybeans ranged from 0.79% to 1.46%, with a corresponding range of coefficients of determination of 0.83 to 0.96. The instrument was designed to handle large seeds at substantially higher measurement rates than what has previously been developed. Janni et al. (2008) also developed an NIRS system with good predictions for corn oil content. Spectra were collected on air-tumbled seeds in this system using a 12 s scan time.

NIRS typically uses single-seed constituent measurements coupled with single-seed spectra for prediction model development. Unfortunately, many reference measurements of constituents are difficult, expensive, or imprecise for single seeds. An alternative approach uses bulk reference measurements matched to the mean spectra of a sample obtained from averaging single-seed spectra. This method was proposed by Shadow and Carrasco (2000) and is implied indirectly by Delwiche and Hruschka (2000). The method was also used successfully by Dowell et al. (2006) for wheat and millet. Benefits of this approach are that prediction models are quicker to develop, less costly, suitable for small breeder samples that must be retained for planting, and self-prediction models can be developed.

The focus of this research is to examine the single-seed prediction accuracy of NIRS models developed from bulk reference analysis. This work is important for the development of NIRS prediction of small sample lines, sometimes consisting of less than a hundred seeds. This is particularly important for soybean protein and oil measurement and the subsequent selection of seeds from genetic lines produced at the University of Kentucky. This work is focused on increasing both protein and oil levels in soybeans. Corn samples are included in this study, as results should be applicable to these seed types. This work also addresses use of a single-seed NIRS system that can collect spectra on single seeds at a high rate, which would facilitate quick development of bulk models at a lower cost. Previous research has used NIRS systems that could only measure seeds at rates of a few per minute, as opposed to seconds.

Objectives

This work is intended to provide a better understanding of NIRS model accuracies attainable when using mean sample spectra derived from single seeds and bulk sample reference values in lieu of single-seed spectra and reference values. Objectives of this study were to develop and examine the prediction accuracies for common constituents of corn and soybeans using bulk models to predict single-seed constituents. A secondary objective was to determine an effective number of seeds that were adequate for deriving the mean sample spectra used in bulk model development.

Materials and Methods

Corn and Soybean Samples

Forty hybrids of commercial yellow corn and 40 varieties of soybeans were provided by Iowa State University. Bulk analysis values were provided for protein, oil, and starch for corn, and for protein, oil, and fiber for soybeans. Twenty-one soybean samples with bulk reference analysis for protein and oil were also provided by the University of Kentucky. Samples are summarized in table 1. Reference measurements for 35 of the Iowa samples were obtained using the following standards: AOAC 990.03 (protein) and 920.39 (oil), Corn Refiners Association A-20, (starch), and AOCS Ba 6-84 (fiber). Five samples were measured using a grain analyzer (Infratec 1241, Foss, Eden Prairie Minn.), calibrated and maintained by Iowa State University. The SEP values for their calibrations were 0.53%, 0.32%, 0.74%, and 0.02 for corn protein, oil, starch, and density, respectively, and 0.55%, 0.35%, and 0.08% for soybean protein, oil, and fiber, respectively. Seed density was measured using a pycnometer (AccuPyc 1330, Micromeritics Instrument Corp., Norcross, Ga.).

Kentucky samples were measured using an NIR analysis system (DA7200, Perten Instruments, Springfield, Ill.), maintained by the University of Kentucky. Twelve of the samples were obtained from the Soybean Quality Traits program administered by the American Oil Chemists Society and the United Soybean Board. Four samples were obtained from the Department of Agronomy and Horticulture at the University of Nebraska due to their high protein content. Four samples were obtained from the USDA-ARS (Raleigh, North Carolina), representing two high and two low oil samples. One sample represents a control commonly used to evaluate other soybean lines by the University of Kentucky. The SEP values for the DA7200 calibrations were 0.32% and 0.38% for soybean protein and oil, respectively. The DA7200 measurements for 14 samples were cross-checked with analytical methods, and the difference between measurements was

Table 1. Statistical profile of the sample sets.

Sample	Mean	SD	Range	*N*
Kentucky soybeans[a]				
Protein (%)	42.32	5.42	35.40-53.80	21
Oil (%)	20.94	2.70	15.50-26.10	21
Iowa soybeans[a]				
Protein (%)	42.66	2.99	37.06-52.23	40
Oil (%)	21.31	1.25	17.33-23.10	40
Fiber (%)	5.26	0.30	4.17-5.77	40
Iowa corn[b]				
Protein (%)	8.04	1.75	5.47-13.00	40
Oil (%)	4.36	1.43	2.80-8.37	40
Starch (%)	59.55	2.24	54.15-62.63	35
Seed density (g cm^{-3})	1.27	0.04	1.20-1.34	35

[a] Protein and oil 0% moisture basis.
[b] Protein, oil, and starch are reported on a 15% moisture basis.

found to be at most 0.51% for protein and 0.36% for oil. All samples were selected, as much as possible, to have a broad range of constituent values that were evenly distributed across their constituent ranges.

Single-Seed NIR Instrument

The NIRS instrument used for single-seed spectra collection was principally designed to rapidly measure and sort corn and soybean seeds by composition, such as protein and oil content. Measurement rates are three seeds per second. The instrument's main components are an NIR spectrometer, fiberoptic bundle, light tube assembly, control circuit, and computer. The spectrometer (model NIR256-1.7T1, Control Development, Inc., South Bend, Ind.) has a thermoelectrically cooled InGaAs diode array with a spectral range of 904 to 1685 nm. The light tube assembly has 48 miniature tungsten light bulbs (part 1150, 5 V, 0.115 A, Gilway Technical Lamp, Woburn, Mass.) arranged equidistantly in six rows along the tube periphery. The lights are housed in an aluminum tube. A glass borosilicate tube, 12 mm internal diameter, runs through the center and length of the aluminum tube. A spectrum of a seed was taken as it slid down the glass tube. The conceptual drawing of the light tube is shown in figure 1. A 2 m long, bifurcated optical fiber assembly with a 600 μm core diameter (Ocean Optics, Dunedin, Fla.) is attached at both ends of the light tube and connected to the spectrometer.

The photo-detector (model D12DAB6FP, Banner Engineering Corp., Minneapolis, Minn.) was used to detect the seed and trigger the spectrometer. The integration time for the spectrometer was set at 43 ms. Spectral data were automatically sent to a controlling PC via a USB interface. A Microsoft Visual C++ program using the CDI software library was used to save the spectral data. A description of the construction and operation of an earlier assembly is provided by Armstrong (2006). One difference between the earlier instrument and the present instrument is that a bifurcated fiber collects spectra from both ends of the tube rather than the single fiber previously used at the top end.

Prior to spectra collection, the instrument was allowed to warm up for at least an hour to stabilize the lights and spectrometer. A background dark current and reference reflectance spectra from white Spectralon (99% diffuse reflectance) were collected at constant time intervals (1/2 h) during the spectral data collection. The white reference measurements were made by inserting a slice of Spectralon at the midsection of the light tube. The glass tube is comprised of two tube sections with a small gap to allow insertion of the Spectralon.

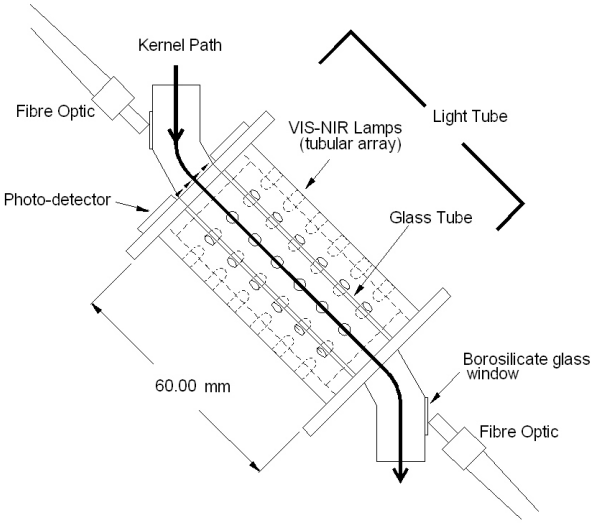

Figure 1. Component assembly used for spectral measurements.

Bulk Reference NIRS Models

In developing prediction models, expected variations in the samples caused by factors such as genetic variability, seasonal changes, location, and cultural practices had to be accounted for. Armstrong (2006) achieved good model prediction statistics for rapid single-seed NIRS of moisture content of corn and soybeans, and protein content for soybeans. The soybean dataset consisted of three varieties having low, medium, and high protein contents. The current study worked on a wider array of hybrids and varieties, creating greater diversity of genetic variation in the prediction models for both corn and soybeans. The constituent values of these samples were also selected to represent a broad and reasonably even range of values. Sample selection did result in a sample set having a higher number of samples closer to the sample-set average and is the result of selecting for multiple constituents.

Seed spectra were collected from 48 randomly selected seeds from each sample and seed type. Each seed was scanned three times, and the mean seed spectrum was computed after mean centering. The mean sample spectrum was then computed, from the mean seed spectrum, for five combinations of 10, 20, 30, 40, and 47 seeds. The seed spectra used in these combinations were randomly selected from the 48 seeds with replacement. The mean sample spectrum was also computed for all 48 seeds. The five combinations for each seed number were used to develop better trends of the effect of seed number on model statistics. Samples were randomly sequenced when scanning seeds.

Partial least squares regression (PLS1) with cross-validation was performed between the mean sample spectra determined from the seed combinations defined above and bulk reference values. Bulk prediction models were developed separately for the respective seed sets from Iowa and Kentucky. A combined Iowa-Kentucky model was also developed for soybean protein. Spectral pretreatments were mean centering (MC) or MC plus standard normal variate (MC-SNV). ParLeS software (Viscarra Rossel, 2008) was used to develop models, and no validation sets were used due to the low sample number. The number of factors determined as optimal for each model was based on the minimum root mean square error (RMSE) value obtained from the PLS1 modeling after examination of RMSE versus the factor levels.

Single-Seed NIRS Model for Soybeans

The goal for selecting seeds used for single-seed prediction model development was to obtain a broad and equal distribution of seed protein content, as suggested by Williams (2001). The initial step was to select seeds from the 40 Iowa soybean samples that would be individually analyzed for protein content; Kentucky soybeans were not used as they were needed for future work. The seeds selected should not only compose a wide span of constituent values but each level should be equally represented. To achieve this, single-seed

protein content was predicted using the bulk NIRS protein prediction model developed from the Kentucky samples. Single-seed protein values predicted by the model yielded no obvious outliers. A normal distribution plot was then produced from predicted protein and used to select 15 seeds at each binned protein levels of 1%, for a total of 360 seeds. Predicted protein content ranged from 29% to 52%. Seeds were placed in labeled cell-bind trays for identity preservation.

Spectra from these seeds were obtained by the same methods described previously and then sent to the Soil Testing Laboratory in the Department of Agronomy, Kansas State University, for protein analysis (FP-528, Leco Corp., St. Joseph, Mich.) using AACC Method 46-30 (AACC, 2000). Moisture content determination was performed on several representative samples prior to protein analysis according to *ASABE Standards* (2006) and used for moisture adjustment. Single-seed PLS1 modeling was also developed using ParLeS software (Viscarra Rossel, 2008). Comparisons were then made with bulk model predictions from single-seed spectra and reference values.

Results and Discussion

Bulk Model Predictions

Bulk model prediction statistics are shown in table 2. Mean spectra were determined using seed counts of 10, 20, 30, 40, and 47 seeds subsampled from the pool of 48 seeds. Five different models were developed for each of the seed counts to provide the standard deviation of model statistics but are not shown. Additional models used all of the 48 seeds available and included a protein model for combined Iowa and Kentucky soybeans. Factor levels selected for soybean protein models were selected from plots of RMSE versus factor level, as shown in figure 2. Similar selections were made for other models. At the selected factor levels, model regression coefficients generally corresponded to known constituent absorption bands with pronounced coefficient peaks occurring close to these wavelengths (data not shown).

Results in table 2 show that the mean and standard deviation of the modeling statistics (R^2, SECV, and RPD) generally improved with an increase in seed number. Part of the reason for this was that only 48 seeds were available from the sampling pool, and models with higher seed numbers will

Table 2. PLS1 prediction statistics for mean spectra regressed with bulk reference values. The number of seeds varied from 10 to 48 for determining the mean spectra. Five different models were determined for each number of seed levels with mean regression statistics reported.[a]

No. of Seeds	R^2	SECV	RPD	F	R^2	SECV	RPD	F
	Iowa Corn Protein, Mean, *n* = 40				Iowa Corn Oil, Mean, *n* = 40			
10	0.75	0.86	2.0	6	0.73	0.73	1.9	6
20	0.75	0.86	2.0	6	0.84	0.58	2.6	6
30	0.77	0.83	2.1	6	0.87	0.52	2.7	6
40	0.78	0.82	2.1	6	0.87	0.50	2.9	6
47	0.78	0.81	2.2	6	0.87	0.51	2.8	6
48	0.79	0.78	2.2	6	0.87	0.50	2.8	6
	Iowa Corn Starch, Mean, *n* = 35				Iowa Corn Density, Mean, *n* = 35			
10	0.68	1.26	1.8	5	0.88	0.012	3.0	3
20	0.79	1.13	2.3	5	0.91	0.011	2.9	3
30	0.80	1.01	2.3	5	0.863	0.013	2.7	3
40	0.79	1.01	2.2	5	0.88	0.012	3.0	3
47	0.80	1.00	2.3	5	0.88	0.012	2.9	3
48	0.80	.99	2.3	5	0.88	0.012	3.0	3
	Kentucky Soybean Protein, Mean, *n* = 21				Iowa Soybean Protein, Mean, *n* = 40			
10	0.86	1.99	2.7	4	0.79	1.18	2.2	6
20	0.88	1.76	3.1	4	0.82	1.06	2.4	6
30	0.88	1.77	3.1	4	0.83	1.07	2.4	6
40	0.89	1.76	3.1	4	0.84	1.04	2.5	6
47	0.89	1.68	3.2	4	0.84	1.02	2.6	6
48	0.90	1.66	3.3	4	0.84	1.02	2.6	6
	Kentucky Soybean Oil, Mean, *n* = 21				Iowa Soybean Oil, Mean, *n* = 40			
10	0.83	1.09	2.5	4	0.77	1.91	2.1	6
20	0.84	1.06	2.5	4	0.81	1.28	2.2	6
30	0.84	1.06	2.6	4	0.81	1.24	2.5	6
40	0.87	.922	2.7	4	0.84	1.26	2.5	6
47	0.87	.977	2.8	4	0.84	1.26	2.5	6
48	0.87	.966	2.8	4	0.84	1.26	2.5	6
	Iowa Soybean Fiber, Mean, *n* = 40				Iowa-Kentucky Combined Protein, *n* = 61			
10	0.44	0.19	1.3	6	NA	--	--	--
20	0.54	0.17	1.5	6	NA	--	--	--
30	0.54	0.18	1.5	6	NA	--	--	--
40	0.57	0.17	1.5	6	NA	--	--	--
47	0.54	0.18	1.5	6	NA	--	--	--
48	0.54	0.18	1.5	6	.89	1.00	3.93	7

[a] R^2 = coefficient determination for the cross-validation model, SECV = standard error of the cross-validation model, RPD = ratio of the standard deviation to the standard error for the cross-validation model, F = factor level used for PLSR model, and NA = not applicable.

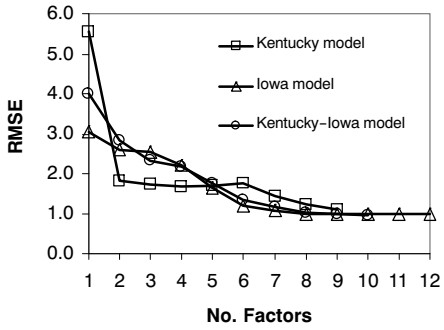

Figure 2. RMSE versus factor levels for bulk protein models using 48-seed mean spectra.

undoubtedly use the same seeds; thus, model statistics at these levels should not vary as much. Regardless, the model statistics showed good improvement when going from 10 seeds to 20 seeds, with improvements less pronounced as additional seeds were included. However, subsamples were more independent of each other at lower seed counts. Delwiche and Hruschka (2000) showed significant model improvements for wheat protein prediction as the number of seeds used for mean spectra increased from 1 to 100. This study differed in that single-seed protein was used to determine the bulk sample protein.

Overall, the predictive ability of bulk reference models could be described as qualitative for many constituents and thus adequate for sorting seeds into course segregations. Models using MC-SNV were as good, or better, than models using MC (not shown) alone. Iowa soybean fiber was not predicted well, and model statistics did not improve with increasing seed numbers.

Single-Seed and Bulk Model Protein Predictions

Predictions of single-seed protein reference values from single-seed spectra are shown in figure 3, and RMSE versus factor levels are shown in figure 4. The single-seed model used MC-SNV as a spectral pretreatment, and the statistics shown are for cross-validation only. Statistics derived from MC spectra were good but poorer compared to MC-SNV. The single-seed model had good quantitative prediction ability (SECV = 0.98%, RPD = 5.56) at a factor level of 8.

Iowa, Kentucky, and Iowa-Kentucky bulk protein models were used to predict Iowa single-seed soybean reference protein from single-seed spectra. Most bulk model predictions of protein from single-seed spectra were reasonable but not as good as the single-seed model predictions. The Iowa and Iowa-Kentucky bulk model SEP was 1.82% and was about twice that of the single-seed model (0.97%) with little bias present. Bias was calculated as the mean reference value minus the mean predicted value. Statistics were basically identical for both of these bulk models at the factor levels used (figs. 5 and 6). The Kentucky model, however, had poor predictions of Iowa soybean protein at the original selected factor level of 4 (fig. 7), with significant bias and much greater SEP (3.94%). When the factor level was increased to 8 (fig. 8), predictions were much better (SEP = 2.65%, RPD = 2.1) but not as good as the Iowa and Iowa-Kentucky predictions. While a factor level of 8 could be regarded as an overfit of the data, the model was predicting an independent Iowa data set and did so approaching the performance of the self-predicting Iowa bulk model.

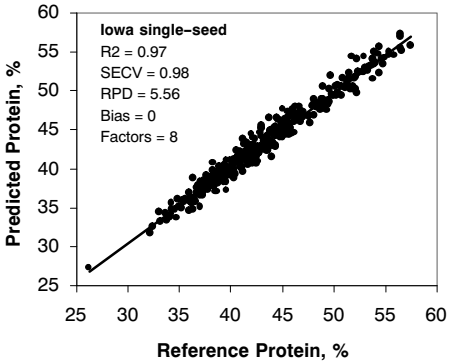

Figure 3. Protein prediction for Iowa single soybean seeds using a single-seed model.

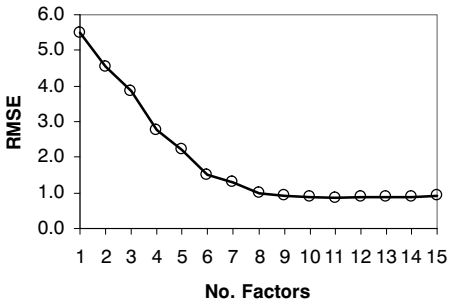

Figure 4. RMSE versus number of factors for the single-seed protein model developed from Iowa soybeans.

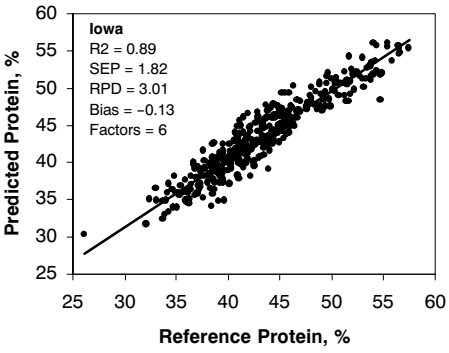

Figure 5. Iowa bulk model prediction of single protein from single-seed spectra.

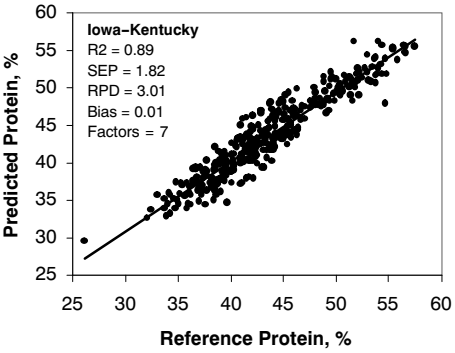

Figure 6. Iowa-Kentucky bulk model prediction of single-seed protein from single seed spectra.

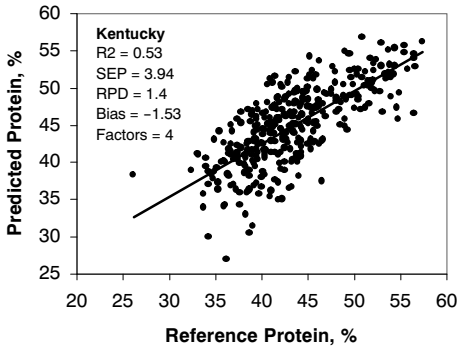

Figure 7. Kentucky bulk model prediction of Iowa single-seed protein from single seed spectra at a factor level of 4.

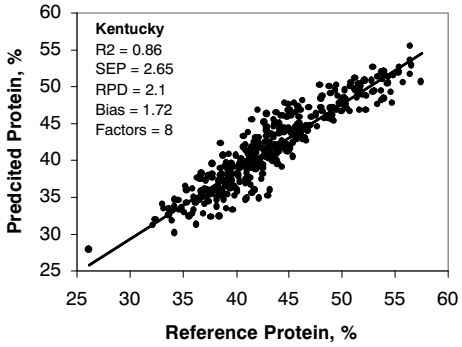

Figure 8. Kentucky bulk model prediction of Iowa single seed protein from single-seed spectra at a factor level of 8.

Results indicate that bulk models can provide rough segregation of samples for common constituents such as protein and oil. The instrumentation used in this work allows seeds to be scanned quickly and easily so that models can be developed from bulk samples and reference values as opposed to single seeds. This can reduce the time and expense of developing single-seed models and allow development of self-prediction models, as was the case for predicting single-seed protein using the Iowa bulk model. In cases where seeds need to be retained and cannot be used for destructive reference analysis, the prediction of Iowa single-seed protein from the Kentucky bulk model indicates that it is possible to get a rough constituent ranking of seeds using an independent prediction model. In this case, though, the Kentucky bulk models developed were very sensitive to model factor levels and should be viewed with caution or somehow validated with external reference samples. Future work should address this issue, and a possible solution would be development of reference standards to check and correct model predictions.

Conclusions

Development of single-seed NIRS prediction models using bulk sample references and mean single-seed spectra for corn and soybeans should provide a means to sort seeds into different levels of protein and oil and starch (corn). It should also be possible to sort corn seeds by seed density. Bulk model predictions for soybean fiber were poor and are unlikely to segregate seeds based on fiber. Statistics show that the number of seeds used to derive the mean spectra for a sample should be at least 30. Single-seed soybean protein values predicted by bulk NIRS models and a single-seed protein model compared favorably with each other, although bulk model performance was not considered quantitative when compared to a single-seed model. Self-prediction bulk models performed much better in measuring single-seed protein than independent bulk models. The instrumentation used in this work can facilitate quick development of bulk models for seed sorting, but verification of sorted fractions would seem advisable.

References

AACC. 2000. Method 46-30: Crude protein - Combustion method. Approved methods of the AACC. St. Paul, Minn.: American Association of Cereal Chemists.

Abe, H., T. Kusama, S. Kawano, and M. Iwamoto. 1996. Non-destructive determination of protein content in a single-seed of wheat and soybean by near-infrared spectroscopy. In *Near-Infrared Spectroscopy: The Future Waves. Proc. 7th Intl. Conf. on Near-Infrared Spectroscopy*, 457-461. A. Davies and P Williams, eds. Chichester, U.K.: NIR Publications.

Armstrong, P. R. 2006. Rapid single-seed NIR measurement of grain and oil-seed attributes. *Applied Eng. in Agric.* 22(5): 767-772.

ASABE Standards. 2006. Standard S352.2. Moisture measurement unground grain and seeds. St. Joseph, Mich.: ASABE.

Baye, T. M., T. C. Pearson, and A. M. Settles. 2006. Development of a calibration to predict maize seed composition using single-seed near-infrared spectroscopy. *J. Cereal Sci.* 43(2): 236-243.

Cogdill, R. P., C. R. Hurburgh, G. R. Rippke, S. J. Bajic, R. W. Jones, J. F. McClelland, T. C. Jensen, and J. Liu. 2004. Single-seed maize analysis by near-infrared hyperspectral imaging. *Trans. ASAE* 47(1): 311-320.

Delwiche, S. R. 1998. Protein content of single wheat seeds of wheat by near-infrared spectroscopy. *J. Cereal Sci.* 27(3): 241-254.

Delwiche, S. R., and W. R. Hruschka. 2000. Protein content of bulk wheat from near-infrared reflectance of individual seeds. *Cereal Chem*. 77(1): 86-88.

Delwiche, S. R., and D. R. Massie. 1996. Classification of wheat by visible and near-infrared reflectance from single seeds. *Cereal Chem*. 73(3): 399-405.

Dowell, F. E., E. B. Maghirang, R. A. Graybosch, P. S. Baenziger, D. D. Baltensperger, and L. E. Hansen. 2006. An automated near-infrared system for selecting individual seeds based on specific quality characteristics. *Cereal Chem.* 83(5): 537-543.

Hardy, C. L., G. R. Rippke, and C. R. Hurburgh Jr. 1995. Calibration and field standardization of Foss Grainspec analyzers for corn and soybeans. In *Near-Infrared Spectroscopy: The Future Waves. Proc. 7th Intl. Conf. on Near-Infrared Spectroscopy*, 132-141. A. Davies and P. Williams, eds. Charlton, U.K.: NIR Publications.

Janni, J. B., A. Weinstock, L. Hagen, and S. Wright. 2008. Novel near-infrared sampling apparatus for single-seed analysis of oil content in maize. *Applied Spectrosc.* 62(4): 423-426.

Rippke, G. R., C. L. Hardy, and C. R. Hurburgh Jr. 1995. Calibration and field standardization of Tecator Infratec analyzers for corn and soybeans. In *Near-Infrared Spectroscopy: The Future Waves. Proc. 7th Intl. Conf. on Near-Infrared Spectroscopy*, 122-131. A. Davies and P. Williams, eds. Charlton, U.K.: NIR Publications.

Shadow, W., and A. Carrasco, 2000. Practical single-seed NIR/visible analysis for small grains. *Cereal Foods World* 45(1): 16-18.

Tajuddin, T., S. Watanabe, R. Masuda, K. Haruda, and S. Kawano, 2002. Application of near-infrared spectroscopy to the estimation of protein and lipid contents in single seeds of soybean recombinant inbred lines for quantitative trait loci analysis. *J. Near-Infrared Spectrosc.* 10(4): 315-325.

Viscarra Rossel, R. A. 2008. ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and Intelligent Lab. Systems* 90(1): 72-83.

Velasco, L., C. Möllers, and H. C. Becker. 1999. Estimation of seed weight, oil content, and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. *Euphytica* 106(1): 79-85.

Williams, P. C. 2001. Implementation of near-infrared technology. In *Near-Infrared Technology in the Agricultural and Food Industries*, 150-153. 2nd ed. P. Williams and K. Norris, eds. St. Paul, Minn.: American Association of Cereal Chemists.