RAPID SINGLE-KERNEL NIR MEASUREMENT OF GRAIN AND OIL-SEED ATTRIBUTES

P. R. Armstrong

ABSTRACT. A single-kernel near infrared (SKNIR) instrument was designed and tested for rapid measurement of corn and soybean attributes. The design was centered on achieving a spectral collection rate of 10 kernels/s, which limited integration times of the spectrometer to 30 ms. A spectrum of an individual kernel was collected as it slid along the length of a glass tube and was illuminated by multiple lamps. PLS regression models, developed to predict constituents from spectra, resulted in models with standard errors of cross validation (SECV) of 0.93% dry basis moisture content (MCdb) for corn, 0.32% MCdb for soybean moisture content, and 0.99% for soybean protein content. RPD values for these models were 4.4 for corn moisture content, 7.3 for soybean moisture content, and 4.9 for soybean protein content. RPD values were defined as the ratio of the standard deviation of the reference data to the SECV for each model. Multiplicative scatter correction improved predictions for soybean moisture and protein content but not for corn moisture content. These results indicate that reasonable predictions can be made at fast NIR scan rates.

Keywords. Near infrared, Grain, Spectroscopy, Single-kernel.

Use of near-infrared reflectance (NIR) or transmis-
sion (NIT) spectroscopy within the grain industry is well established for bulk sample measurement of major constituents. More recent work in NIR spectroscopy has focused, in part, on single-kernel NIR (SKNIR) measurement where detection of attributes at this level can provide more usable information in some applications. This is particularly true for determining the variability within grain samples (Maghirang and Dowell 2003), sorting for particular attributes, or detecting attributes which may only be present in a small percentage of the kernels (Maghirang et al., 2003).

Previous research encompassed a broad range of SKNIR uses. Delwiche and Massie (1996) identified wheat class with good distinction between red and white classes. Delwiche (1998) generally had good prediction ability of single-kernel protein content for different wheat classes, with a standard error of calibration ranging from 0.513% to 0.648%. Campbell et al. (2000) concluded that corn classes with unique starch or amylose content could be classified using near-infrared transmittance spectroscopy. In other applications apart from constituent measurement, Maghirang et al. (2003) classified sound wheat kernels and kernels containing live insects at pupal, large, medium-sized, and small larval stages with 94%, 93%, 84%, and 62% accuracy, respectively. Fumonisin levels greater than 100 ppm and less than 10 ppm in corn could be accurately classed as fumonisin positive or negative (Dowell et al., 2002). Wang et al. (2003) used SKNIR to classify fungal damage in soybeans. Classification accuracy of healthy beans and those infected with Phomopsis spp, C. kikuchii, soybean mosaic virus and Peronospora manshurica (downy mildew) were 100%, 99%, 84%, 94%, and 96%, respectively. NIR and NIT spectroscopy were used by Pearson et al. (2001) to distinguish aflatoxin in single corn kernels.

High-speed optical sorters can perform some of the functions of multi-spectral instruments, but the gain in speed generally results in a loss of precision and sorting accuracy. Pasikatan and Dowell (2004) used a ScanMasterII 200 high-volume color sorter (Satake-USA, Houston, Tex.) to segregate wheat by protein. This particular sorter is limited to a binary sort using two wavelengths, one in the visible and one in the NIR region, but it has a capacity of several hundred kilograms per hour. Current SKNIR technology has measurement rates of about one kernel per second. At this rate, measuring and sorting 1 kg of wheat (approximately 30 × 10³ kernels) requires more than 8 h. The limitation on speed is caused by spectra collection times and the materials handling system. Improving speed would enhance the use of SKNIR and allow its use for more routine quality analysis and screening in a broader range of processing applications.

The purpose of this study was to design and test a SKNIR instrument that has the ability to collect spectra at a rate of 10 kernels/s and assess the prediction ability for common constituents. This provides a benchmark for the prediction of other constituents which may be more difficult to measure. This work reports on the methods used to collect spectra and the results of the developed prediction equations for corn moisture content and soybean moisture and protein content.

Submitted for review in October 2005 as manuscript number IET 6144; approved for publication by the Information & Electrical Technologies Division of ASABE in June 2006.

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

The author is Paul R. Armstrong, ASABE Member, Agricultural Engineer, USDA-ARS, Grain Marketing Production Research Center, 1515 College Ave, Manhattan, Kansas 66502; phone: 785-776-2728; fax: 785-537-5550; e-mail: Paul.Armstrong@gmprc.ksu.edu.
MATERIALS AND METHODS

SINGLE-KERNEL NIR INSTRUMENT

The SKNIR system used in this study was designed for corn, soybeans, and similarly sized biological specimens. Considerations in the design were presentation of kernels for spectral viewing and methods to illuminate and scan at a rate of 10 kernels/s. The physical configuration is shown in figure 1. Primary components consisted of a light tube used for illumination of kernels during spectra collection, electronics used to detect kernels and implement sorting controls, and a spectrometer for spectra collection. The operational sequence was as follows. The signal from the photo-switch pulsed as a kernel fell through it. The signal was fed to an electronic hardware delay, which then triggered the spectrometer to collect a spectrum. The spectrum was collected as the kernel traveled down the light tube and was transferred from the spectrometer to a PC by a USB interface where it was displayed and stored. Software was developed to incorporate various spectrometer controls and hardware to sort into five constituent categories using developed prediction equations and constituent thresholds entered by the user. An automated single-kernel feeding and sorting mechanism is currently being developed.

SPECTROMETER

The spectrometer model used was a CDI 256L-1.7T1 (Control Development Inc., South Bend, Ind.) and incorporates a 256-element InGaAs photo-diode array, coupled with a 16-bit A/D converter that spans the spectral range from 906 to 1683 nm. The diode array is thermo-electrically cooled for stability. Selection of this particular spectrometer was, in part, due to its ability to interface with a PC and provide close control with a set of function libraries via the PC. Other criteria were that a photo-diode array provides a relatively simple and fast method to collect spectra and covers a broad range of the NIR spectrum.

LIGHT TUBE

The light tube was designed to provide a bright and reasonably constant light source throughout the length of the tube. Forty-eight miniature lamps were placed in eight rows along the circumference of an aluminum tube. Rows were equally spaced radially and lamps were 8 mm apart within a row. The lamps have emission in the visible to infrared range and are rated at 5 VDC and 0.6 W each (Gilway Technical Lamp, part 1150, Peabody, Mass.). A glass tube (12-mm internal diameter), positioned concentric to the light tube, was used to convey kernels by gravity through the light tube. Reflectance measurements were collected through a 400-micron fiber-optic cable (2-m length) positioned at the lower end of the light tube, with the other end of the cable attached to the spectrometer input. A 10-mm, circular, boro-silicate glass window was positioned in front of the fiber end to protect it from falling kernels.

EXPERIMENTAL PROCEDURES

Kernel residence time in the tube and illumination intensity of the kernel were considered primary factors in achieving a measurement rate on 10 kernels/s. Measurements of residence time, using photo-switches positioned at each end of the tube, revealed that the residence time ranged from approximately 50 to 70 ms for kernels hand fed at the entrance of the tube. This allowed a maximum of 30 ms of dead time between a kernel exiting the tube and the next kernel to enter the tube, and a maximum of 50 ms for spectral readings. The selection of 48 lamps and their placement to achieve adequate illumination for spectral measurements, was primarily experimental. A fundamental need was enough illumination to take fast spectral measurements. Spectrometer integration time was set by experimentally determining integration times that resulted in saturation of the detector and then reducing it until there was no saturation. The purpose of this procedure was to provide as much signal as possible to the detector without saturation. Typically, a 40-ms integration time caused saturation for some corn kernels in the middle region of their spectra, while no saturation was observed for 30 ms. This was also observed for soybeans. Spectrometer settings thus used an integration time of 30 ms with no averaged spectra, i.e. single-spectrum readings were used. A single spectrum was used because multiple spectra collection would require too much time considering the residence time of the kernel in the light tube. It was possible to set the spectrometer with lower integration times and take multiple spectra but it was not known if the latency of the system between spectrum measurements would have an adverse effect or if multiple spectra would actually improve spectral measurements. Maximum kernel velocities were estimated at approximately 1.0 m/s based on free-fall velocity.

Prior to collection of spectra of kernels, a background spectrum was taken by blocking the fiber-optic input. The background spectrum was automatically subtracted from the measurement spectrum and accounted for dark current in the spectrometer. Background collection is transparent to the user as dark current levels are stored internally within the spectrometer. Background measurements also initiate discrete wavelength correction by the spectrometer and each spectral point is guaranteed to be within 0.5 nm of the stated value. Reference spectra were taken by inserting a piece of cylindrical Spectralon™ (Labsphere, North Sutton, N.H.) material, slightly smaller than the diameter of the glass tube, into the light tube. The Spectralon™ piece was stationary for this operation and was positioned at the midpoint of the light tube. The reference spectrum represented the 100% light level for a particular physical configuration and integration time. The reference spectrum minus the background was
used by the spectrometer to calibrate individual photodiodes. Background and reference spectra were automatically stored in the spectrometer.

**Corn Moisture Content**

Yellow-dent corn kernels of a single hybrid were conditioned to predetermined moisture contents by determining their existing moisture content (ASAE Standards, 2003) and adding the appropriate amount of distilled water to elevate kernel moisture, or by drying at 35°C to achieve a weight corresponding to a given dry basis moisture content (MCdb). Water was added in increments with half the required water initially sprayed onto the kernels, and the remaining half the following day. Samples were stored at 3°C for five days to allow equilibration of water within the kernel. The moisture content levels used were 8%, 10%, 12%, 14%, 16%, and 18% with 15 kernels at each level.

Corn kernels were individually hand-fed into the SKNIR instrument with specific orientation of dent-end or tip-cap end first. Sample size was 90 kernels for each orientation. Different kernels were used for the two orientations.

**Soybean Moisture Content**

A single variety of soybeans was conditioned to moisture contents of 6%, 9%, 12%, 14%, 16%, and 18% MCdb with 20 beans at each level (total n = 120) using a procedure similar to that used for corn. Five of the 20 beans at each moisture level were hand selected for their obvious discoloration due to either natural heat or fungal damage. Severely wrinkled, discolored, and split beans were not used. Unlike corn, soybeans tumbled down the tube and orientation was random during spectral measurements. Reference moisture content was determined similar to corn.

**Soybean Protein Content**

Spectra for soybean protein content were collected from three varieties of foundation seed soybeans (KS5502N, KS4103SP, KS4402SP) obtained from the Agronomy Department at Kansas State University. One hundred beans were randomly selected from each variety and placed into pill boxes for identification. Samples were very clean and split or discolored beans were almost nonexistent in these samples. Soybeans were hand-fed into the spectrometer and returned to the pill boxes. Reference values for protein (N × 6.25) was determined by combustion using a Leco model FP-528 nitrogen analyzer (St. Joseph, Mich.) on individual whole beans using AACC Method 46-30 (2000). Protein was adjusted to 13% moisture basis MC using the average MC for each variety. Protein ranged from 29% to 55% with a standard deviation of 4.88% for combined varieties. Average protein for the varieties KS5502N, KS4103SP, and KS4402SP was 36.8%, 46.3%, and 39.7%, respectively.

**RESULTS AND DISCUSSION**

Partial least-squares (PLS) regression was used to develop prediction models for all constituents using GRAMS AI software (Thermo Galactic Industries, Salem, N.H.). These models represent the systems potential ability to measure constituents and are not considered a calibration model for measurement. Factor levels used for model selection were based on the F-ratio which equals the predicted residual sum of squares (PRESS) at a specific factor level divided by the minimum PRESS value. Williams (2001) suggested that RPD values of 2.5 to 3 were suitable for rough screening; a value of 5 to 8 could be used for quality control, while an RPD of 8 or higher was excellent. Combined orientation data predictions were similar to that of tip-cap predictions and multiplicative scatter correction (MSC) pretreatment of spectra were performed in all analyses. MSC was examined as a method to reduce spectral variance that was not attributable to constituents but was caused by light scattering resulting from kernel movement or illumination variance as the kernel travels down the tube. Visual inspection of model prediction residuals did not reveal any spectral outliers.

**Corn Moisture Content Prediction**

PLS regression yielded prediction statistics for corn moisture content shown in Table 1. Corn moisture prediction statistics for the dent-end orientation were better than the tip-cap orientation, as indicated by the SECV and the RPD value. RPD is the ratio of the standard deviation, SD, of the reference data to the SECV. Williams (2001) suggested that RPD values of 2.5 to 3 were suitable for rough screening; a value of 5 to 8 could be used for quality control, while an RPD of 8 or higher was excellent. Combined orientation data predictions were similar to that of tip-cap predictions and multiplicative scatter correction adversely affected predictions for each orientation. Beta (regression) coefficients from PLS regression (fig. 2) were fairly random or noisy with no pronounced spectral regions influenced by water. Typically one would expect larger positive or negative coefficients

<table>
<thead>
<tr>
<th>Orientation</th>
<th>SECV</th>
<th>Factors</th>
<th>R²</th>
<th>Sample SD</th>
<th>RPD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn moisture, %db[a]</td>
<td>0.76</td>
<td>12</td>
<td>0.97</td>
<td>4.17</td>
<td>5.5</td>
<td>90</td>
</tr>
<tr>
<td>Corn moisture, %db[b]</td>
<td>0.97</td>
<td>5</td>
<td>0.94</td>
<td>4.17</td>
<td>4.3</td>
<td>90</td>
</tr>
<tr>
<td>Corn moisture, %db[c]</td>
<td>1.04</td>
<td>9</td>
<td>0.94</td>
<td>4.07</td>
<td>3.9</td>
<td>90</td>
</tr>
<tr>
<td>Corn moisture, %db[d]</td>
<td>1.37</td>
<td>7</td>
<td>0.91</td>
<td>4.07</td>
<td>3.1</td>
<td>90</td>
</tr>
<tr>
<td>Corn moisture, %db[e]</td>
<td>0.93</td>
<td>11</td>
<td>0.94</td>
<td>4.11</td>
<td>4.4</td>
<td>180</td>
</tr>
<tr>
<td>Corn moisture, %db[f]</td>
<td>1.33</td>
<td>12</td>
<td>0.91</td>
<td>4.11</td>
<td>3.1</td>
<td>180</td>
</tr>
</tbody>
</table>

[a] Dent-end first.
[b] Multiplicative scatter correction applied.
[c] Tip-cap first.
[d] Combined dent and tip-cap.
[e] Figure 2 shows beta coefficients from the prediction model.
in spectral regions of strong water absorption. The worst-case consequences of this are that the prediction model may work adequately on the sample set used to derive it but may not extend as well for other samples.

**SOYBEAN MOISTURE CONTENT PREDICTION**

PLS regression statistics (table 2) using all beans (n = 140) and excluding the five discolored beans at each moisture content (n = 105), showed slight improvements when discolored beans were removed. As opposed to the case for corn moisture, MSC improved prediction accuracy. SECV values ranged from 0.32% to 0.47% MCdb depending on the sample and spectral pre-treatment. Lamb and Hurburgh (1991) found standard errors of prediction (SEP) for wet-basis soybean moisture content (MCwb) prediction to be between 0.65% and 0.69% for spectra obtained by transmittance. It is difficult to directly compare this work with the previously cited work as samples were entirely different. The instruments for spectra collection also used different wavelength ranges and may also have had significant noise and sensitivity differences. It is apparent, however, that the prediction accuracy was of the same order. Beta coefficients had more pronounced absorption regions compared to corn results (fig. 3). Regions such as 950 to 975 nm and 1400 to 1450 nm are associated with water absorption.

**SOYBEAN PROTEIN CONTENT PREDICTION**

Predictions statistics from PLS regression (table 3) showed reasonable predictive ability for soybean protein for both individual varieties and using combined varietal data. Some spectra had irregular absorption behavior in the region of 906 to 920 nm. Predictive statistics were modestly improved when this region was removed from the PLS analysis.

---

**Table 2. Prediction statistics for soybean moisture.**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>SECV</th>
<th>Factors</th>
<th>R^2</th>
<th>Sample SD</th>
<th>RPD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soybean moisture, %db[^a][^c]</td>
<td>0.47</td>
<td>12</td>
<td>0.98</td>
<td>3.38</td>
<td>7.20</td>
<td>140</td>
</tr>
<tr>
<td>Soybean moisture, %db[^b][c]</td>
<td>0.34</td>
<td>8</td>
<td>0.99</td>
<td>3.38</td>
<td>9.99</td>
<td>140</td>
</tr>
<tr>
<td>Soybean moisture, %db[^b]</td>
<td>0.46</td>
<td>12</td>
<td>0.98</td>
<td>3.38</td>
<td>7.30</td>
<td>105</td>
</tr>
<tr>
<td>Soybean moisture, %db[^b][d]</td>
<td>0.32</td>
<td>8</td>
<td>0.99</td>
<td>3.38</td>
<td>10.5</td>
<td>105</td>
</tr>
</tbody>
</table>

[^a] All spectra were mean-centered.
[^b] Multiplicative scatter correction applied.
[^c] Discolored seeds included.
[^d] Figure 3 shows beta coefficients from the prediction model.

---

**Table 3. Prediction statistics for soybean protein from PLS regression analysis.**

<table>
<thead>
<tr>
<th>Variety, Data Treatment</th>
<th>SECV</th>
<th>Factors</th>
<th>R^2</th>
<th>Sample SD</th>
<th>RPD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS5502N[^a]</td>
<td>1.03</td>
<td>8</td>
<td>0.86</td>
<td>2.87</td>
<td>2.8</td>
<td>100</td>
</tr>
<tr>
<td>KS5502N[^b]</td>
<td>0.81</td>
<td>9</td>
<td>0.91</td>
<td>2.87</td>
<td>3.5</td>
<td>100</td>
</tr>
<tr>
<td>KS4103SP[^b]</td>
<td>1.25</td>
<td>16</td>
<td>0.86</td>
<td>3.35</td>
<td>2.7</td>
<td>100</td>
</tr>
<tr>
<td>KS4103SP[^b]</td>
<td>1.04</td>
<td>10</td>
<td>0.90</td>
<td>3.35</td>
<td>3.2</td>
<td>100</td>
</tr>
<tr>
<td>KS4402SP</td>
<td>0.98</td>
<td>7</td>
<td>0.83</td>
<td>2.22</td>
<td>2.3</td>
<td>100</td>
</tr>
<tr>
<td>KS4402SP[^b]</td>
<td>0.79</td>
<td>8</td>
<td>0.88</td>
<td>2.22</td>
<td>2.8</td>
<td>100</td>
</tr>
<tr>
<td>Combined data</td>
<td>1.46</td>
<td>8</td>
<td>0.91</td>
<td>4.88</td>
<td>3.3</td>
<td>300</td>
</tr>
<tr>
<td>Combined data[^b][^c]</td>
<td>1.01</td>
<td>10</td>
<td>0.96</td>
<td>4.88</td>
<td>4.8</td>
<td>300</td>
</tr>
<tr>
<td>Combined data[^b][^d]</td>
<td>0.99</td>
<td>10</td>
<td>0.96</td>
<td>4.88</td>
<td>4.9</td>
<td>300</td>
</tr>
</tbody>
</table>

[^a] All spectra were mean-centered.
[^b] Multiplicative scatter correction applied.
[^c] 910-1683 nm spectral range used for model development.
[^d] Figures 4, 5, and 6 show original and MSC spectra and beta coefficients from the prediction model.
Figure 4. Absorbance spectra for soybean protein predictions before multiplicative scatter correction applied. Absorbance calculated from reflectance (log (1/R)).

Figure 5. Absorbance spectra for soybean protein predictions after multiplicative scatter correction applied. Absorbance calculated from reflectance (log (1/R)).

Figure 6. Prediction model beta coefficients for soybean protein content using 10 PLS factors. Mean centering and multiplicative scattering applied.

from this study is of the same order as cited literature. In other work, Velasco et al. (1999) obtained an SECV value of 0.94% for protein content in rapeseed which had a sample mean protein content of 21.1 ± 3%. Delwiche (1998) obtained SEP values for wheat protein content of 0.46% to 0.72%.

It was generally believed that this SKNIR system would perform better for soybeans than for corn, because spectra from the beans are a composite signal from various views as they tumble down the tube. Corn, however, maintains orientation as it travels down the tube, and spectra are believed to be comprised primarily of views of the tip-cap or dent-end. This belief seems justified from the results when comparing moisture predictions between soybeans and corn.

For corn moisture predictions, the tip-cap end can contain various fractions that can have significantly different chemical compositions. The predominant portion of the tip-cap is the pedicel, which is the attachment point to the cob. This contains the hilar layer, which can become brown or black depending on kernel maturity and could cause much variation in spectra. The dent-end, however, is predominantly soft endosperm and should be more consistent in chemical composition. For corn, predictions could probably be improved by collecting a spectrum from areas of the corn kernel which are of constituent interest or from gathering a composite spectrum. This might be achieved by placement of multiple fibers along the length of the light tube.

It is anticipated that the long-term performance of the instrument will be affected most by lamp degradation and contamination of the glass tube surface by dust and abrasion. The degree to which this occurs and its affects need to be addressed. The instrument physical configuration and materials can be modified which can affect optics and kernel drop speed. The effects of these will cause changes in spectra. For applications such as sorting for specific trait characteristics in grain and oil-seed breeding programs, long-term stability may not be as critical as predictions models are often developed on the same material to be sorted and sorting completed in a short time frame. However, the use of the instrument for quality control in industry would require stable instrument performance.

CONCLUSIONS

The system’s predictive performance proved to be adequate for screening of corn moisture content with RPD values ranging from 3.1 to 5.5, and for soybean protein content with RPD values from 2.3 to 4.9. Soybean moisture prediction content was more quantitative with RPD values ranging from 7.2 to 10.5. Multiplicative scatter correction of spectra improved prediction of soybean moisture and protein content but not corn moisture content. Different methods for spectral measurement of corn kernels should be examined to improve corn moisture results by reconfiguration of the fiber optics. Spectral measurement at 10 kernels/s appears to be feasible based on these prediction statistics, but material handling methods for delivery and sorting of kernels need to be developed.

ACKNOWLEDGEMENTS

The author is appreciative of the soybean protein content measurements performed by Rhonda Lyne of the Grain Quality Structure Unit at the Grain Marketing Production Research Center, Manhattan, Kansas.

REFERENCES


