ELSEVIER

Journal of Cereal Science 43 (2006) 236–243

Journal of CEREAL SCIENCE

www.elsevier.com/locate/jnlabr/yjcrs

Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy

Tesfaye M. Baye [a,1], Tom C. Pearson [b], A. Mark Settles [a,*]

[a] *Horticultural Sciences Department, University of Florida, P.O. Box 110690, Gainesville, FL 32611-0690, USA*
[b] *USDA-ARS Grain Marketing and Production Research Center, 1515 College Avenue, Manhattan, KS 66502, USA*

Received 8 August 2005; received in revised form 10 November 2005; accepted 16 November 2005

Abstract

The relative composition of protein, oil, and starch in the maize kernel has a large genetic component. Predictions of kernel composition based on single-kernel near infrared spectroscopy would enable rapid selection of individual seed with desired traits. To determine if single-kernel near infrared spectroscopy can be used to accurately predict internal kernel composition, near infrared reflectance (NIR) and near infrared transmittance (NIT) spectra were collected from 2160 maize kernels of different genotypes grown in several environments. A validation set of an additional 480 kernels was analyzed in parallel. Constituents were determined analytically by pooling kernels of the same genotype grown in the same environment. The NIT spectra had high levels of noise and were not suitable for predicting kernel composition. Partial least squares regression was used to develop predictive models from the NIR spectra for the composition results. Calibrations developed from the absolute amount of each constituent on a per kernel basis gave good predictive power, while models based on the percent composition of constituents in the meal gave poor predictions. These data suggest that single kernel NIR spectra are reporting an absolute amount of each component in the kernel.

Keywords: Maize; Near infrared spectroscopy; Seed composition; Single kernel

1. Introduction

Cereal grains contribute to over 60% of the total world food production (Lásztity, 1999). Cereals are predominantly composed of carbohydrates, mostly in the form of starch, with considerable amounts of protein as well as some lipids, vitamins, and minerals. Both genetic and environmental effects create significant variation in the amount and quality of each of these constituents. Multiple methods have been developed to help breeders screen crops for various seed composition traits (e.g. Baenziger et al., 2001; Dunlap et al., 1995). Chemical analysis procedures are the most widely accepted reference methods for determining seed composition. However, these methods frequently are destructive and require large samples of grain.

Near infrared spectroscopy provides an alternative, non-destructive technology for measuring constituents of biological materials. Organic molecules have specific absorption patterns in the near infrared region that can report the chemical composition of the material being analyzed (Williams and Norris, 2001). Near infrared spectra can be collected either from the reflectance (NIR) of a sample or the transmittance (NIT) through a sample (Delwiche, 1995; Williams, 1979). Both NIR and NIT measurements allow the simultaneous determination of multiple constituents in a sample and are commonly used to predict the composition of bulk whole grain samples in maize (Orman and Schumann, 1991). Bulk whole-grain samples can be screened rapidly, require no sample preparation, and preserve the kernels after the measurement for further analysis or for propagation (Baye and Becker, 2004; Velasco et al., 1999). However, use of whole grain samples does not allow the identification of individual kernels that deviate significantly from the mean composition within

Abbreviations: bt1, *brittle1*; *bt2*, *brittle2*; *dek*, *defective kernel*; NIR, near infrared reflectance; NIT, near infrared transmittance; PLS, partial least squares regression; *rgh3*, *rough endosperm3*; *sh2*, *shrunken2*; *su1*, *sugary1*.

* Corresponding author. Tel.: +1 352 392 7571; fax: +1 352 392 5653.
E-mail address: settles@ufl.edu (A.M. Settles).
[1] Present Address: Department of Biostatistics, University of Alabama at Birmingham, 1665 University Blvd, RPHB 445B, Birmingham, AL 35294, USA.

0733-5210/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jcs.2005.11.003

a population. In addition, bulk samples give no indication of an abnormal distribution of kernels within the sample, such as a few kernels contaminated by a fungus or a segregating population of kernels with differing composition (Dowell et al., 2002).

Non-destructive analysis of single kernel composition is valuable for identifying outlying individuals both for breeding and for industrial seed sorting applications. In maize, single-kernel nuclear magnetic resonance (NMR) technology has been used to select oil traits (Alexander et al., 1967), and it has been demonstrated that the rate of improvement for maize oil traits can be enhanced over bulk sample analysis through single-kernel selection (Silvela et al., 1989). However, NMR is limited to detecting liquid constituents such as oil and moisture in maize seeds. NIR and NIT spectroscopy give the possibility for rapid screening of individual kernels for multiple chemical constituents. Single kernel NIT spectra between 920 and 950 nm were found to correlate very well with moisture content, with a 2% standard error of prediction of moisture content (Finney and Norris, 1978). More recently, single kernel NIR and NIT spectroscopy has been used to sort individual maize kernels for different types of fungal infections (Dowell et al., 2002; Pearson et al., 2001, 2004) or to identify genetically modified kernels (Munck et al., 2001).

The present study was conducted to determine if single kernel NIR or NIT spectroscopy could also be used to predict kernel composition in maize. These predictions will help geneticists and breeders to screen large numbers of samples and then select and propagate single seeds with desirable composition traits. Single kernel NIR data have been used to develop a predictive model for wheat protein content (Delwiche and Hruschka, 2000). However, maize kernels have a much less uniform internal structure with the maize embryo comprising a larger proportion of the seed than in wheat. Indeed, attempts to develop calibrations for maize oil content using NIT data suggest that single kernel predictive models would be difficult to develop for maize (Cogdill et al., 2004; Orman and Schumann, 1992).

Here, we report the development calibration equations to predict accurately individual maize kernel constituents using NIR technology. Near infrared predictive models are best developed with samples that display a large range of compositional variation (reviewed in Willimas and Norris, 1987). In maize, *defective kernel* (*dek*) mutants have large effects on seed size and can effectively delete major constituents of the kernel by affecting starch biosynthesis (reviewed in Boyer and Hannah, 2001), storage protein accumulation (reviewed in Gibbon and Larkins, 2005), or by aborting embryo growth early in seed development (Magnard et al., 2004). In the mature kernel, the embryo contains the highest proportion of oil and removal of the embryo causes a reduction in total oil content. We reasoned that *dek* mutants would provide a large range of different maize kernel compositional variants for developing predictive models from near infrared spectra and focused on these mutants in this study.

2. Materials and methods

2.1. Seed stocks

All maize seeds used were obtained from plants grown at the University of Florida, Plant Science Research and Education Unit (Citra, Florida). Eight inbred lines commonly used for genetic studies were included as reference stocks. These lines included: W22, W23, Mo17, B73, A632, W64A as well as color-converted W22 and A632 stocks ($W22^{ACR}$ and $A632^{ACR}$). All inbred seeds were from self-pollinated ears. To develop the calibration and external validation sets, maize kernel mutants and their normal sibling seeds were used as a source of large variance in seed composition. The majority of the seed mutants were derived from the UniformMu transposon-tagging population (McCarty et al., 2005). The UniformMu mutants were selected based on visible defects in embryo or endosperm development, which are likely to cause altered kernel composition.

Seed for each of 24 UniformMu mutants were grown in two field seasons and self-pollinated to identify segregating ears. In each field season, there were variations in soil type, average temperature, and day length. A pair of segregating ears for each mutant was selected to include both environmental and genetic variation. In addition, six mutants with strong effects on starch accumulation were used as controls including: *shrunken2* (*sh2*) in a W22 background, *brittle1* (*bt1*) in a W23 background, and *sugary1* (*su1*), *bt1*, *brittle2* (*bt2*), and *sh2* in a W64A background. Two ears of *su1* in W64A and a single ear of each of the other mutants were included in the study for a total of 55 ears of corn segregating for 30 different seed mutant loci. A mutant and normal sample of 24 kernels each was selected from each ear for single-kernel near infrared spectroscopy and analytical determination of composition. A total of 110 samples were used to develop and test the calibrations. Fig. 1A shows examples of the mutant and corresponding normal kernels for eight of the ears used in this study, and a schematic of the overall study design is shown in Fig. 1B.

2.2. Near infrared data collection and pretreatments

Near infrared spectra were collected from 24 mutant and 24 normal sibling kernels from each of the 55 ears selected for the calibration development. Prior to collecting the spectra, the kernels were equilibrated to ambient humidity for 2 or more weeks in a controlled laboratory environment. NIR and NIT spectra were collected from the abgerminal side of individual kernels. The NIR spectra were measured on maize kernels placed manually onto the quartz window of a bifurcated interactance fiber-optic bundle (Fig. 2) attached to a diode-array spectrometer and light source (DA7000, Perten Instruments, Springfield, IL). The viewing area was 17 mm diameter, the illumination bundle was a 7 mm diameter ring, and the reflectance probe bundle was 2 mm diameter. The light source was chopped at a frequency of 30 Hz and the integration time for the spectrometer was set at 33 ms. A total of 15 spectra were acquired for each kernel without repositioning the kernel

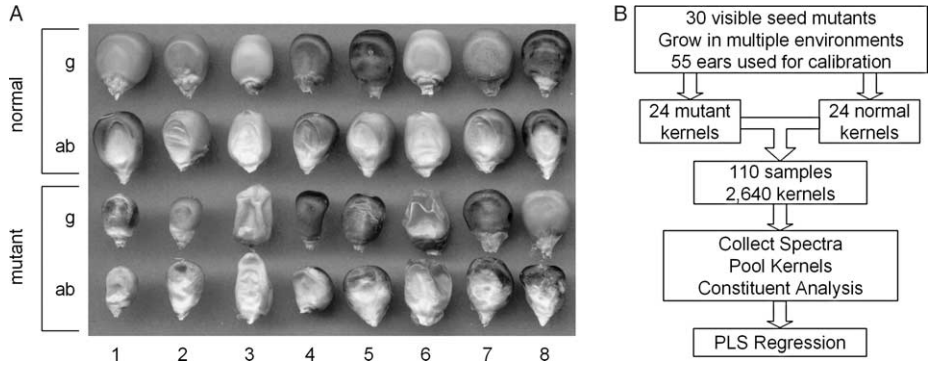

Fig. 1. (A) Examples of the mutant and normal sibling seeds used in this study. The top two rows show normal kernels, and the bottom two rows show sibling mutant kernels. Each column of four kernels derives from different ears of corn segregating for a mutant, and the germinal (g) and abgerminal (ab) faces of the kernels are shown. Columns three and six show examples of known low starch mutants, *bt1* and *su1*, respectively. (B) Flow diagram of the experimental methods used to develop the PLS regression models.

on the fiber optic. These spectra were averaged and saved for analysis. Prior to scanning batches of 48 kernels, a Spectralon reference measurement was taken. Reflectance (R) values were collected at 5 nm intervals for wavelengths between 890 and 1700 nm, and absorbance values were calculated as $\log(1/R)$. First derivative absorbance spectra were calculated in Microsoft Excel by subtracting the absorbance values 10 nm above and below each wavelength from 900 to 1690 nm. The derivative spectra were transformed with the standard normal variant (SNV) transformation prior to regression against the analytical constituent data (Barnes et al., 1989).

The NIT spectra were collected with a fiber optic InGaAs diode array spectrometer ((OSC/256L-1.7T1-250A/0.9-1.7/3.2, Control Development, SouthBend, IN, USA) from 900 to 1500 nm at 3 nm intervals. This was accomplished by positioning the InGaAs spectrometer fiber optic 2 cm above the kernels on the quartz window of the interactance probe.

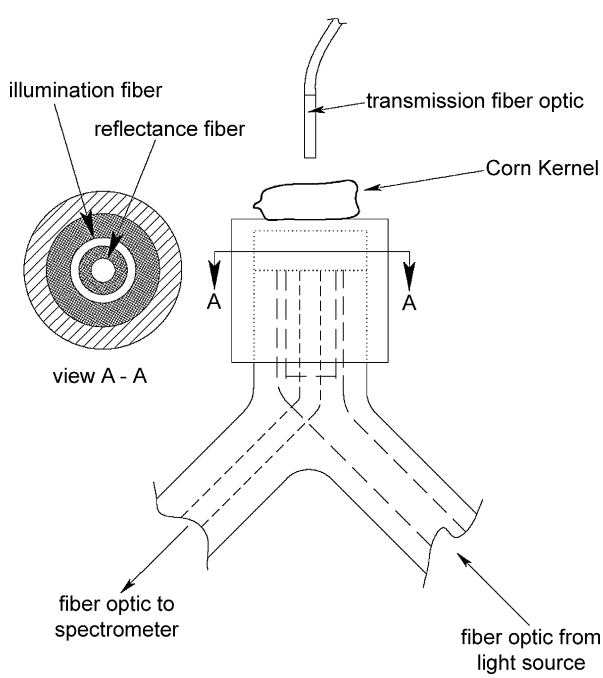

Fig. 2. Schematic of the combined near infrared reflectance and transmittance sampling apparatus.

The fiber optic consisted of a single low OH fiber, 400 μm in diameter and 61 cm long. The fiber captured light in a circle of about 3 mm diameter centered on the corn kernel. This arrangement allowed simultaneous acquisition of both transmission and reflectance spectra and minimized sample handling (Fig. 2). Prior to measurement of each set of 48 kernels, a 0.5 in. thick piece of Teflon was used to take a reference transmittance measurement. The NIT spectra contained substantial noise due to light scattering and attenuation through the whole corn kernel. A variety of integration times were tested with the best signal-to-noise ratio observed for an integration time of 300 ms per kernel. This lengthy integration time led to detector saturation for a significant fraction of the kernels. The correlations between measured constituents and the NIT spectra were poor and not studied further.

2.3. Constituent analysis

The kernels were pooled into samples of 24 mutant or 24 normal kernels from individual ears to obtain sufficient meal for the analysis of multiple constituents. An average seed weight was determined for each pool, and the seeds were ground to a fine meal using a UDY Cyclone sample mill (UDY Co., Fort Collins, CO). For the inbred lines, a large sample of ~200 kernels was ground for multiple replicates of each assay. The meal was divided into aliquots for assays of gross energy, protein, starch, and relative fatty acid content using methods described below. All reference measurements were tested with four replicate assays from the eight reference inbred stocks and showed a coefficient of variation $<3\%$ (see Section 3). For the calibrations, laboratory measurements were completed in duplicate, and the average of the two measurements was used for partial least squares regression.

Gross energy was measured with an isoperibol bomb calorimeter standardized with benzoic acid as described by Miller and Payne (1959). Briefly, 100 mg of dried meal was placed in a Parr Model 1261 bomb calorimeter (Parr Instrument Co., Moline, IL) and ignited in a pressurized, highly pure oxygen environment. Gross energy was calculated based on the heat released according to the manufacturer's instructions.

Protein content ($N \times 6.25$) was determined indirectly from total nitrogen content using a fully automated Elantech NCS 2500 C/N combustion analyzer (Angelino et al., 1997; AOAC International, 2000; Fiedler et al., 1973). Total nitrogen was measured on a 10 mg sample of meal following the manufacturer's instructions.

Starch was determined from an enzymatic digestion of the maize meal followed with a colorimetric assay for glucose content (Holm et al., 1986; Karkalas, 1985). Briefly, 100 mg of sample meal was hydrated in 5 ml H_2O and then digested with 0.1 ml heat stable *alpha*-amylase (21,390 units/ml; Sigma-Aldrich A-3403) by heating to 93 °C for 1 h. A 0.1 ml aliquot of the digested sample was mixed with 0.9 ml H_2O, 1 ml 0.1 M sodium acetate (pH 4.5), and 50 μl amyloglucosidase (1244 units/ml; Sigma-Aldrich A-1602), and the extract incubated at 60 °C for 30 min. The enzyme-treated extract was diluted 1:250 in H_2O and 0.5 ml of the diluted extract was mixed with 2.5 ml glucose oxidase–peroxidase reagent [0.91% Na_2HPO_4, 0.5% KH_2PO_4, 0.1% phenol, 0.015% 4-aminoantipyrine, 7 units/ml glucose oxidase (Sigma-Aldrich G-6641), 7 units/ml peroxidase (Sigma-Aldrich P-8415)]. The reaction was incubated at 37 °C for 45 min, cooled to room temperature, and the absorbance measured at 505 nm using a Beckman DU-50 series spectrophotometer. A standard curve of glucose was used to determine the glucose content in the extracts, which was then converted to starch with a multiplication factor of 0.9. A parallel extraction of the maize meal was completed without the addition of polysaccharide degrading enzymes, and soluble sugar content for the maize samples was estimated with the same assay. The quantity of soluble sugars in the meal was lower than could be accurately determined with a coefficient of variation of 14% for the eight reference stocks (data not shown), and the total saccharides were reported as starch content.

Relative fatty acid contents were determined by gas chromatography (GC) after base catalyzed release, as fatty acid methyl esters (Thies, 1971). About 10 mg meal was mixed with 1 ml 0.5 M sodium methoxide in methanol and transmethylated by incubating at 20 °C for 20 min. Iso-Octane (500 μl) and 5% $NaHSO_4$ (200 μl) was then added and the mixture centrifuged at 300 *g* for 5 min to separate the organic phase. The upper phase was transferred into a septum vial and analyzed for fatty acid methyl esters on a Hewlett–Packard model 5890 Series II gas chromatograph (Avondale, PA) equipped with a split/split-less injector, a flame-ionization detector, an automatic sampling device, and a 15-M Durabond -23 capillary column (J&W Scientific, Deerfield, IL), which was 0.25 mm i.d. with a film thickness of 0.25 μm. The column temperature was programmed to increase from 180 to 220 °C at 15 °C/min, and the injector and detector ports were set at 250 and 300 °C, respectively. The carrier gas was helium with a flow rate of 2.2 ml/min. Individual fatty acids were identified on the basis of peak areas of injections and measured with a Hewlett–Packard 3390A reporting integrator and expressed as a percentage of the total fatty acids (area/area), including minor fatty acids.

2.4. Partial least squares regression

A partial least squares (PLS) regression was used to determine if predictive models could be developed for the NIR spectra, NIT spectra, and the different expressions of kernel composition. The spectra and analytical data were randomly separated into a 90-sample training set (calibration set) and a 20 sample external validation set (prediction set). For each model, the descriptor data set (X matrix) consisted of the transformed spectral data, which was averaged for each 24 kernel sample. The response data set (Y matrix) consisted of the analytical data for the kernel constituents. Each analytical value represents the average composition for each 24 kernel sample. Absolute amounts of the constituents were calculated by multiplying the average kernel weight by the percent values. The same calculation was completed for relative fatty acid levels to give a normalized fatty acid composition.

The PLS models were developed on mean-centered data sets, and the calibration was made using SIMCA-P software (Umetrics AB, 2005). The number of significant PLS factors to build the model for each compositional trait was determined by cross-validation. Cross-validation was completed by removing one-seventh of the training set data, and a calibration model was developed with the remaining samples. The removed samples were then predicted, and the entire procedure was repeated until all 90 samples were removed and predicted. An optimal calibration was then selected using the number of significant factors recommended by the default settings of the software. A factor was considered significant if the Q^2 statistic of cross-validation for the factor was >0.05 and the factor explained $>3\%$ of the variance for the response data set ($R^2Y>0.03$). Q^2 was calculated as: $Q^2 = 1.0\text{-PRESS/SS}$, where PRESS is the prediction error sum of squares ($\text{PRESS} = \Sigma[Y_{\text{actual}} - Y_{\text{predicted}}]^2$) and SS is the residual sum of squares for the previous factors. The optimal computed models were used to predict the samples in the external validation set, and the correlation coefficient of this prediction was reported as Q^2ev.

3. Results

The single-kernel NIT spectra showed high levels of noise (data not shown) as previously reported by Cogdill et al. (2004) and were not analyzed further. In contrast, the NIR spectra showed clear differences that appear to relate to the internal chemical composition of the kernels. Fig. 3 shows individual NIR spectra for three samples of 24 kernels. Based on the analytical results (discussed below), these spectra represent individuals with low protein (UniformMu *rough endosperm3, rgh3*), low starch (*sugary1*, *su1*), or average protein and starch (UniformMu, normal kernels). The *su1* kernels are characterized by a flat overall absorbance, which is reflected in the first derivative as lower magnitude spectra. The low protein, *rgh3*, mutants showed a lower absorbance over the entire NIR spectra with a sharper upward slope in absorbance between 1400 and 1450 nm. This caused the peak in the first derivative to occur at slightly longer wavelengths as can be seen in Fig. 3B. These observations suggest that single-kernel NIR spectra can detect

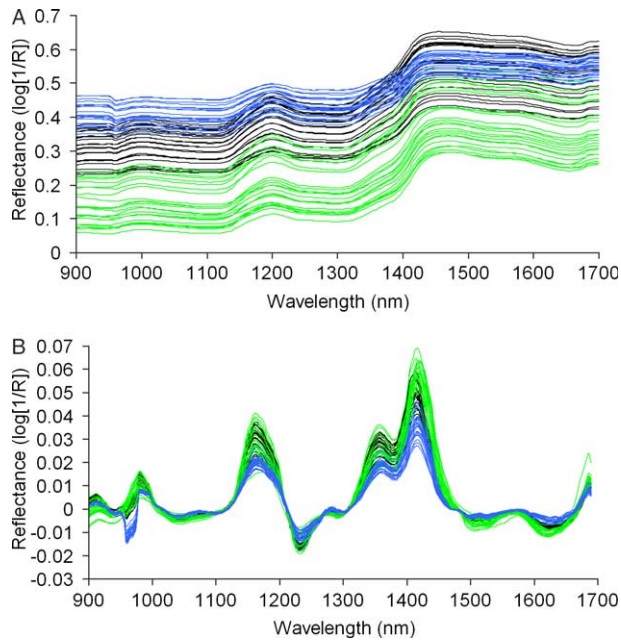

Fig. 3. Examples of the direct NIR spectra (A) and first derivative NIR spectra (B) used in the study. Each panel shows the 24 spectra for each of three kernel samples including a low protein mutant (green or light gray lines, *rgh3*), a low starch mutant (blue or dark gray lines, *su1*), and a sample with average protein and starch (black lines, UniformMu normal kernels)

clear differences in relative kernel composition. However, the NIR spectrum contains information from all the chemical constituents of the sample and direct interpretation of the absorbance values is difficult for complex mixtures such as an intact kernel (Williams and Norris, 2001).

Standard analytical techniques were performed on a series of eight maize inbred lines to ensure that they gave accurate composition data with low levels of variance due to measurement errors (Table 1). These methods were then used to collect composition data from mutant and normal kernels. The mutants showed lower average levels of the major kernel components, as well as higher SD relative to the means, suggesting that the mutants add a large amount of composition variance to the overall data set (Table 2). The descriptive statistics in Table 3 show that the 90 samples assigned to the calibration data set and the 20 samples assigned to the prediction data set had similar ranges and variation in kernel composition regardless of whether the composition values were expressed as relative composition of the meal or absolute quantities within the kernel. These data indicate that the two sets contain enough variation to develop calibration equations and that the external validation set is representative of the range of variation found in the overall seed samples.

PLS models were developed based on relative kernel composition (e.g. % protein or % starch) as well as absolute amounts of kernel constituents (e.g. mg protein/kernel or cal/kernel). Using models based on the relative levels of each constituent in the meal, it was not possible to predict the composition of the external validation set (Table 4). The relative kernel composition models explained $<56\%$ of the spectral variation (R^2X) and $<41\%$ of the composition variation (R^2Y). The best relative prediction was for linoleic acid with a correlation coefficient of 0.42 for the predicted versus analytically determined values of the external validation samples (Q^2ev).

In contrast, PLS models using NIR data and absolute levels of the kernel constituents are predictive. The most accurate models were for predictions of the major components of the kernel including protein, starch, and calorie content as well as seed weight (Table 4 and Fig. 4). These PLS equations for protein, starch, calories, and kernel weight all explain $\geq 85\%$ of the variation found in the NIR spectra. Based on cross-validation statistics, one or two significant PLS factors (A) were selected for each calibration equation, and these equations explained $\geq 82\%$ of the variation existing in the analytically determined components. The models have a good predictive power based on internal cross-validation (Q^2cv = 0.81–0.91), external cross-validation (Q^2ev = 0.85–0.90), and the ratio of the standard error of prediction (SEP) to the SD of the prediction set (SEP/SD = 0.32–0.41). Scatter plots of the analytically determined and PLS predicted values indicate that the models show similar levels of prediction error over the full range of values measured (Fig. 4). In addition, all models showed a slight bias with a trend to over-estimation of the constituents in the kernels with the lowest absolute levels and under-estimating kernels with high constituent levels. Interestingly, PLS models could also be developed for the major fatty acids within the kernel when the relative levels were normalized for kernel weight (Table 4 and Fig. 3E and F). These data indicate that models accurately describing

Table 1
Maize meal composition for inbred lines

Inbred line	Protein[a] (%) ± SE	Starch[a] (%) ± SE	Palmitic, 16:0[b] (%) ± SE	Oleic, 18:1[b] (%) ± SE	Linoleic, 18:2[b] (%) ± SE	Calorie (cal/g) ± SE
W22^{ACR}	15.6 ± 0.41	68.9 ± 0.32	17.2 ± 0.46	18.5 ± 0.20	54.0 ± 0.49	4005 ± 22
W22	12.4 ± 0.22	70.3 ± 0.43	13.8 ± 0.26	28.8 ± 0.35	48.1 ± 0.84	3774 ± 21
W23	10.7 ± 0.20	76.4 ± 0.78	15.7 ± 0.08	29.1 ± 0.22	42.5 ± 0.43	4017 ± 14
Mo17	14.5 ± 0.09	71.1 ± 0.46	14.2 ± 0.14	21.9 ± 0.25	56.7 ± 0.57	4122 ± 24
B73	10.7 ± 0.06	74.3 ± 0.42	14.8 ± 0.08	27.3 ± 0.77	46.3 ± 1.50	4049 ± 19
A632	14.5 ± 0.32	70.3 ± 0.39	15.6 ± 0.42	20.8 ± 0.61	56.4 ± 0.58	3977 ± 43
A632^{ACR}	11.7 ± 0.19	70.7 ± 0.68	13.3 ± 0.08	28.1 ± 0.16	50.8 ± 0.37	4121 ± 35
W64A	11.6 ± 0.18	77.2 ± 0.89	11.5 ± 0.16	25.6 ± 0.14	56.1 ± 0.51	4011 ± 20

[a] % of maize meal.
[b] % of extracted oil.

Table 2
Comparison of average normal and mutant kernel composition

Constituent (±SD)	Normal ($n=55$)	Mutant ($n=55$)
Protein (mg/k)[a]	28.6±4.0	18.0±6.0
Starch (mg/k)[a]	144±21	81±33
Energy (cal/k)[a]	921±133	559±174
Kernel weight (mg/k)[a]	229±31	139±40
Palmitic acid (normalized)[b]	27.4±4	19±6
Oleic acid (normalized)[b]	65±12	42±14
Linoleic acid (normalized)[b]	123±15	70±23

[a] k=kernel.
[b] Fatty acid content was normalized by multiplying average seed weight by the relative levels.

individual kernel composition can be developed using single-kernel NIR. However, the results suggest that the NIR spectra are reporting an absolute amount of each component in the kernel rather relative levels.

4. Discussion

We tested the possibility of developing single-kernel calibration models for the major components of the maize seed. Our results suggest that these models will be a challenge to develop, but with an appropriate strategy, accurate compositional predictions for single kernels are certainly possible. However, the standard paradigms for developing PLS or principal component regression predictive models do not appear to be suitable for single kernel predictions in maize kernels. First, single-kernel NIT data contain excessive noise and cannot be used as an equivalent to NIR data for predicting individual kernel composition. Similar results have been reported previously for individual maize kernels (Cogdill et al., 2004; Orman and Schumann, 1992). Potentially, the NIT spectra are more sensitive to the density or total mass of the kernel, because maize kernels are relatively large and longer wavelength light does not penetrate individual kernels. However, the NIT data cannot be used to predict kernel weight (data not shown) suggesting that the noise from single kernel NIT data results from more complex sources than simply seed mass differences.

Second, the standard approach of regressing against composition data expressed as a percentage does not yield accurate predictive models for single maize kernels. Instead, NIR data regressed against kernel composition expressed as an absolute or normalized amount yield the most accurate predictions. This latter approach takes into account variations in kernel weight and allows for independent changes of individual constituents. The models based on absolute or normalized amounts are predictive and could be used to make selections of individual kernels with altered composition. Overall, these observations indicate that single kernel near infrared spectroscopy reports an absolute amount of the kernel constituents.

This conclusion provides insight into the reasons for the poor prediction of the relative composition PLS models. A change in the absolute amount of a single constituent will cause percentage changes in all of the constituents of the kernel, and two kernels with identical absolute amounts of one constituent would have different relative levels of this constituent if any other component within the kernel changes. As an example, consider two kernels that each contain 25 mg of protein. One kernel has slightly more total mass than the other, 185 vs. 175 mg, which could be caused by a different accumulation of starch or oil. These kernels will have different relative levels of protein, 13.5 vs. 14.2%, and the kernel with less total mass will appear to have a 6% gain in protein content. However, our results suggest that the underlying NIR spectral fingerprint will still report that both kernels have 25 mg of protein. Thus, any regression based on percent protein will be inherently inaccurate, because the percentage differences in relative protein for this example are not translated into differences within the NIR spectra.

Table 3
Range of composition variation in the samples used to develop PLS models

	Calibration set ($n=90$)				Prediction set ($n=20$)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Relative								
Protein (% meal)	12.3	0.9	10.8	16.0	12.1	0.8	10.9	15.0
Starch (% meal)	60.2	9.2	33.1	75.9	59.4	9.5	39.5	71.8
Energy (cal/g)	4027	245	2874	4536	3999	316	3121	4396
Palmitic acid (% oil)	13.3	2.2	9.6	24.9	13.5	3.2	10.0	19.3
Oleic acid (% oil)	29.6	3.6	20.5	41.4	27.1	5.3	11.9	35.8
Linoleic acid (% oil)	51.7	4.7	40.5	65.6	53.8	4.2	45.4	63.7
Absolute value								
Kernel weight (mg/k)	183	59	39	287	187	56	90	271
Protein (mg/k)	23.1	7.2	5.0	37.2	24.3	7.1	10.7	37.0
Starch (mg/k)	119	42	21	193	114	43	36	181
Energy (cal/k)	738	241	144	1183	745	231	353	1136
Palmitic acid[a]	23.4	6.9	5.3	37.2	24.5	6.3	13.8	33.7
Oleic acid[a]	52.7	17.9	10.7	88.0	53.3	18.2	11.9	76.6
Linoleic acid[a]	95.8	33.8	20.8	135.2	100.9	31.7	40.9	130.7

[a] Fatty acid content was normalized by multiplying average seed weight by the relative levels.

Table 4
PLS model statistics based on relative or absolute values for kernel constituent data

Constituent	Calibration statistics				Prediction statistics		
	A	R^2X	R^2Y	Q^2cv	Q^2ev	SEP	SEP/SD
Relative							
Protein (% meal)	3	0.56	0.37	0.16	0.23	1.7	2.12
Starch (% meal)	1	0.29	0.29	0.23	0.24	11.5	1.21
Energy (cal/g)	2	0.30	0.26	0.18	0.41	183	0.58
Palmitic acid (% oil)	2	0.49	0.41	0.30	0.38	2.2	0.69
Oleic acid (% oil)	1	0.26	0.30	0.20	0.39	4.2	0.79
Linoleic acid (% oil)	1	0.40	0.34	0.27	0.42	4.2	1.00
Absolute							
Protein (mg/k)	2	0.90	0.92	0.91	0.90	2.3	0.32
Starch (mg/k)	2	0.86	0.89	0.88	0.87	17.8	0.41
Energy (cal/k)	1	0.85	0.82	0.81	0.85	93.9	0.41
Palmitic acid[a]	2	0.66	0.71	0.67	0.77	3.2	0.51
Oleic acid[a]	2	0.68	0.79	0.75	0.79	9.6	0.53
Linoleic acid[a]	1	0.85	0.87	0.86	0.84	13.3	0.42

A = number of PLS factors used to build the model; R^2X = the explained spectral variation, (1-SS(E)/SS(X)); SS(E) = sums of squares of the X residuals; SS(X) = total sums of squares for the spectral variation. R^2Y = the explained variation in seed composition, (1-SS(F)/SS(Y)); SS(F) = sums of squares of the Y residuals and SS(Y) = total sums of squares for the constituent variation. Q^2cv = the predictive power of the model according to cross-validation (1-Π(PRESS/SS)); Q^2ev = correlation coefficient for the external validation set; SEP = standard error of prediction for the external validation set; SD = standard deviation of the external validation set.

[a] Units are normalized fatty acid content.

Furthermore, these results suggest that current approaches to develop calibrations for commercial seed sorting instruments are limited. Accurate seed composition predictions are probably possible using relative composition data. However, these predictions will be limited to a narrow range of seed weight, because seed weight changes can cause dramatic relative composition changes that are not necessarily reflected in near infrared spectra. More robust calibrations need to account for variations in seed weight to allow for accurate maize kernel composition predictions and sorting. Finally, it is important to note that the calibrations developed here rely on analytical data derived from pooled kernel samples. More accurate calibrations are likely to be possible by obtaining absolute quantities of constituents from single kernels.

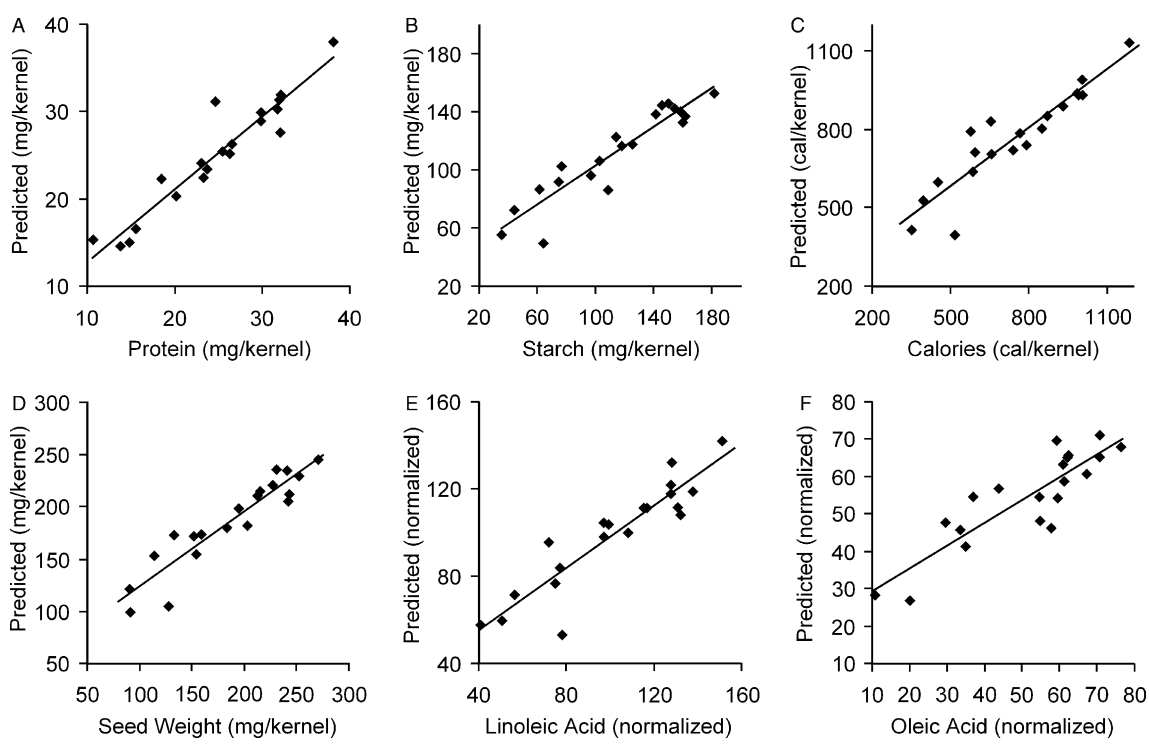

Fig. 4. Scatter plots of the analytically determined and predicted values for the PLS models using NIR and absolute kernel composition data. Each plot shows the values for the prediction set of 20 samples and a linear regression trend line. Plots are given for (A) protein, (B) starch, (C) calorie content, (D) seed weight, (E) normalized linoleic acid, and (F) normalized oleic acid.

Acknowledgements

This research was supported with funding from the NSF Plant Genome Research Program (Award (#0404560), the Vasil–Monsanto Endowment, the USDA-ARS Project entitled, 'Objective Grading and End-Use Property Assessment of Single Kernels and Bulk Grain Samples', and the Florida Agricultural Experiment Station. We thank Tiffany Bispham, Augustine Obour, Carrie Schwartz, Adrienne Berry, Andrew Rosander, Elane Walker, Amanda Gleason, and James Morrill for their technical assistance as well as the anonymous reviewers for their helpful comments.

References

Alexander, D.E., Silvela, L., Collins, S.F.I., Rodgers, R.C., 1967. Analysis of oil content in maize by wide-line NMR. Journal of the American Oil Chemists' Society 44, 555–558.

Angelino, S.A.G.F., van Laarhoven, H.P.M., van Westerop, J.J.M., Broekhuijse, B.M., Mocking, H.C.M., 1997. Total nitrogen content in single kernel malting barley samples. Journal of the Institute of Brewing 103, 41–46.

AOAC International, 2000. Official Methods of Analysis of AOAC International, 17th ed. AOAC International, Arlington, VA.

Baenziger, P.S., Shelton, D.R., Shipman, M.J., Graybosch, R.A., 2001. Breeding for end-use quality: reflections on the Nebraska experience. Euphytica 119, 95–100.

Barnes, B.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra. Applied Spectroscopy 43, 772–777.

Baye, T., Becker, H.C., 2004. Analyzing seed weight, fatty acid composition, oil, and protein contents in *Vernonia galamensis* germplasm by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists Society 81, 641–645.

Boyer, C.D., Hannah, L.C., 2001. Kernel mutants in corn. In: Hallauer, A.R. (Ed.), Specialty Corns. CRC Press, Boca Raton, FL, pp. 1–32.

Cogdill, R.P., Hurburgh, C.R., Rippke, G.R., 2004. Single-kernel maize analysis by near-infrared hyperspectral imaging. Transactions of the ASAE 47, 311–320.

Delwiche, S.R., 1995. Single wheat kernel analysis by near-infrared transmittance—protein content. Cereal Chemistry 72, 11–16.

Delwiche, S.R., Hruschka, W.R., 2000. Protein content of bulk wheat from near-infrared reflectance of individual kernels. Cereal Chemistry 77, 86–88.

Dowell, F.E., Pearson, T.C., Maghirang, E.B., Xie, F., Wicklow, D.T., 2002. Reflectance and transmittance spectroscopy applied to detecting fumonisin in single corn kernels infected with *Fusarium verticillioides*. Cereal Chemistry 79, 222–226.

Dunlap, F.G., White, P.J., Pollak, L.M., Brumm, T.J., 1995. Fatty acid composition of oil from adapted, elite corn breeding materials. Journal of the American Oil Chemists' Society 72, 981–987.

Fiedler, R., Proksch, G., Koepf, A., 1973. The determination of total nitrogen in plant material with an automatic nitrogen analyzer. Analytica Chimica Acta 63, 435–443.

Finney, E.E., Norris, K.H., 1978. Determination of moisture in corn kernels by near-infrared transmittance measurements. Transactions of the American Society of Agricultural Engineers 21, 581–584.

Gibbon, B.C., Larkins, B.A., 2005. Molecular genetic approaches to developing quality protein maize. Trends in Genetics 21, 227–233.

Holm, J., Björck, L., Drews, A., Asp, N.-G., 1986. A rapid method for the analysis of starch. Starch/Die Starke 7, 224–226.

Karkalas, J.J., 1985. An improved enzymatic method for the determination of native and modified starch. Journal of the Science of Food and Agriculture 36, 1019–1027.

Lásztity, R., 1999. Cereal Chemistry. Akadémiai Kiadó, Budapest, Hungary, pp. 11–51.

Magnard, J.L., Heckel, T., Massonneau, A., Wisniewski, J.P., Cordelier, S., Lassagne, H., Perez, P., Dumas, C., Rogowsky, P.M., 2004. Morphogenesis of maize embryos requires ZmPRPL35-1 encoding a plastid ribosomal protein. Plant Physiology 134, 649–663.

McCarty, D.R., Settles, A.M., Suzuki, M., Tan, B.C., Latshaw, S., Porch, T., Robin, K., Baier, J., Avigne, W., Lai, J., Messing, J., Koch, K.E., Hannah, L.C., 2005. Steady-state transposon mutagenesis in maize. The Plant Journal 44, 52–61.

Miller, D.S., Payne, P.R., 1959. A ballistic bomb calorimeter. British Journal of Nutrition 13, 501–508.

Munck, L., Nielsen, J.P., Møller, B., Jacobsen, S., Søndergaard, I., Engelsen, S.B., Nørgaard, L., Bro, R., 2001. Exploring the phenotypic expression of a regulatory proteome-altering gene by spectroscopy and chemometrics. Analytica Chimica Acta 446, 171–186.

Orman, B.A., Schumann, R.A., 1991. Comparison of near-infrared spectroscopy calibration methods for the prediction of protein, oil, and starch in maize grain. Journal of Agricutural and Food Chemistry 39, 883–886.

Orman, B.A., Schumann, R.A., 1992. Nondestructive single-kernel oil determination of maize by near-infrared transmission spectroscopy. Journal of the American Oil Chemists' Society 69, 1036–1038.

Pearson, T.C., Wicklow, D.T., Maghirang, E.B., Xie, F., Dowell, F.E., 2001. Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy. Transactions of the American Society of Agricultural Engineers 44, 1247–1254.

Pearson, T.C., Wicklow, D.T., Pasikatan, M.C., 2004. Reduction of aflatoxin and fumonisin contamination in yellow corn by high-speed dual-wavelength sorting. Cereal Chemistry 81, 490–498.

Silvela, L., Rodgers, R., Barrera, A., Alexander, D.E., 1989. Effect of selection intensity and population size on percent oil in maize, *Zea mays* L. Theoretical and Applied Genetics 78, 298–304.

Thies, W., 1971. Schnelle und einfache Analysen der Fettsaeurezusammensetzung in einzelnen Raps-Kotyledonen I. Gaschromatographische und papierchromatographische Methoden. Zeitschrift für Pflanzenzuechtung 65, 181–202.

Umetrics AB, 2005. SIMCA-P 8.0 User Guide and Tutorial. Umetrics AB, Umea, Sweden. http://www.umetrics.com/software_simcap.asp

Velasco, L., Goffman, F.D., Becker, H.C., 1999. Development of calibration equations to predict oil content and fatty acid composition in *Brassicaceae* germplasm by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists' Society 76, 25–30.

Williams, P.C., 1979. Screening wheat for protein and hardness by near-infrared spectroscopy. Cereal Chemistry 56, 169–172.

Willimas, P., Norris, K., 1987. Near-Infrared Technology in the Agriculture and Food Industries. American Association of Cereal Chemists, St Paul, MN.

Williams, P., Norris, K., 2001. Near-Infrared Technology in the Agricultural and Food Industries, second ed. American Association of Cereal Chemists Inc., St Paul, MN.