*Received: 10 October 2014* ■ *Revised: 23 March 2015* ■ *Accepted: 24 March 2015* ■ *Publication: 21 April 2015*

JNIRS JOURNAL OF NEAR INFRARED SPECTROSCOPY

# Prediction of starch content and ethanol yields of sorghum grain using near infrared spectroscopy

**Junhui Li,[a] Mary-Grace C. Danao,[b]* Shih-Fang Chen,[b] Song Li,[b] Vijay Singh[b] and Patrick J. Brown[c]**

[a]Department of Electronic Engineering, China Agricultural University, Beijing, 100083, PR China. E-mail: gdanao@illinois.edu

[b]Department of Agricultural and Biological Engineering, University of Illinois, Urbana, IL 61801, USA

[c]Department of Crop Sciences, University of Illinois, Urbana, IL 61801, USA

**A rapid quantification method was developed and validated for non-destructive measurement of starch content, theoretical ethanol yield and actual ethanol yield of 48 cultivars of sorghum grain using Fourier transform near infrared (FT-NIR) spectroscopy in diffuse reflectance mode. Multiplicative scatter correction, Savitzky–Golay derivative smoothing and mean centring were used for processing the spectra of ground sorghum grain. The processed spectra were correlated with starch content, theoretical ethanol yield and ethanol produced through simultaneous saccharification and fermentation using partial least-squares regression (PLSR). The spectral range and number of factors were optimised for the low number of factors, high coefficients of determination for calibration ($R^2$) and validation ($r^2$), low root mean square error of prediction (*RMSEP*), high ratio of performance to deviation (*RPD*) and high ratio of the standard error of prediction to the range (*RER*). The best PLSR model for starch content utilised the 4000–6000 $cm^{-1}$ wavebands and had the following values: $R^2 = r^2 = 0.97$, *RMSEP* = 5.5 g $kg^{-1}$ grain, *RPD* = 5.9 and *RER* = 15. Likewise, the model for theoretical ethanol yield utilised the 4000–8000 $cm^{-1}$ wavebands and had $R^2$ and $r^2$ values of >0.90, *RMSEP* = 4.9 g $kg^{-1}$ grain, *RPD* = 4.47 and *RER* = 12.8. It was more difficult to predict actual ethanol yield using FT-NIR spectroscopy given the small data set, and spectra were collected prior to the fermentation step. Resulting PLSR models had $R^2$ and $r^2$ values of <0.60, *RMSEP* = 11.2–21.4 g $kg^{-1}$, *RPD* < 3 and *RER* < 6. These results demonstrated that FT-NIR spectroscopy may be a practical method for rough screening of sorghum cultivars for desirable starch content and theoretical ethanol yield. The models may be improved by including more cultivars in the model and additional compositional information, such as tannin and free amino nitrogen contents, in the chemometric analysis and using FT-NIR scans of the fermentation products to predict actual ethanol yields.**

*Keywords*: Fourier transform near-infrared spectroscopy, sorghum, starch, ethanol, partial least-squares regression

## Introduction

The increasing costs of energy, finite oil and gas reserves have created a worldwide need to improve energy production, interest in alternative transportation fuel resources and, specifically, biofuels from biomass. Currently, the majority of biofuel production in the United States is ethanol derived primarily from starch from corn grain, with sorghum grain as a second source.[1,2] Sorghum has a composition that is similar to corn,[3] and most sorghum starches, on a 1 kg basis, contain 200–300 g of amylose and 700–800 g of amylopectin, with waxy varieties containing 0–150 g of amylose and 850–1000 g of amylopectin.[4] While starch content varies as a function of many factors including plant genetics, growth environment,

ISSN: 0967-0335
doi: 10.1255/jnirs.1146

harvesting method and storage, it may be used as a predictor of ethanol yield.[5,6] These sources of compositional variance are difficult to control. A rapid analysis tool to monitor starch and potential ethanol yields at different stages of bioconversion, as well as to provide critical information for screening sorghum varieties being developed in breeding programmes, would be useful.

Rapid, nondestructive, inexpensive and suitable for online analysis, near infrared (NIR) diffuse reflectance spectroscopy has been widely and successfully applied in many fields, including rapid quantitative and qualitative analysis of composition of cereal grains and biomass for fuel ethanol production.[6–13] Starch-content models for maize or other cereal grains using NIR spectroscopy have been reported by several researchers.[14–18] Compared with NIR studies on corn-quality attributes, there have been relatively few studies on sorghum grain composition using NIR spectroscopy. For feed quality and seed viability, amylose, crude protein, amino acids, lipid and polyphenols contents, as well as gross calorific values, have been estimated using NIR spectroscopy and chemometrics.[19–23] In bioenergy applications, NIR models have been developed to estimate cellulose, hemicellulose, lignin, extractives and ash contents, as well as free sugars in leaves and stalks.[24,25] Thus far, no study has reported using NIR spectroscopy to estimate starch content and potential ethanol yield of sorghum grain. Therefore, in this study, the starch contents and ethanol yields of 48 cultivars of sorghum grain were determined and correlated to their NIR spectra using partial least-squares regression (PLSR) to rapidly characterise different sorghum varieties in breeding programmes.

# Experimental

## Sorghum samples

Sorghum grain from 48 cultivars of dwarf sorghum containing bicolour, caudatum, durra, kafir and guinea races of sorghum was harvested from the Energy Biosciences Institute farm at the University of Illinois in Urbana, IL, USA in November 2012. The samples were cleaned and ground using a hammer mill (Model MHM4, 1100 W, Glen Mills, Clifton, NJ) using a 0.5 mm sieve. A two-stage convection oven was used to determine the moisture content of the ground sorghum following American Association of Cereal Chemists International (AACCI) Method 44-19.01.[26]

## Starch assay and ethanol fermentation

Triplicate samples of ground sorghum (1 g dry solids) were placed in 100 mL glass bottles, mixed with 50 mL of 0.4 M HCl and sterilised in an autoclave (Model SN550, Yamato, Tomy Seiko, Tokyo, Japan) at 126°C for 1 h. After sterilisation, the samples were cooled at room temperature for 10 min. A 2 mL aliquot of the hydrolysate was transferred into a microcentrifuge tube and mixed with 0.22 mL of $NaCO_3$ solution (21.2% w/v) to neutralize the sample. The mixture was then centrifuged at 17,760 × *g* (Durafuge 100, Precision, Winchester, VA) for 5 min. A 1 mL aliquot of the supernatant was transferred into a tube for glucose content ($G_{sample}$, g $L^{-1}$) using a Multiparameter Bioanalytical System (Model YSI 7100MBS, YSI, Yellow Springs, OH). The wet-basis moisture content ($MC_{wb}$, %) of each sample was determined using AACCI Method 44-19.01.[26] Starch recovery (SR) was calculated as

$$\mathrm{SR}\left(\frac{g}{100\,g}\right)=\frac{G_{\mathrm{standard}}}{1.11} \quad (1)$$

where $G_{standard}$ was the amount (g) of glucose derived from 1 g of starch standard used in the YSI, and the value 1.11 was the theoretical conversion factor of starch into glucose. Starch content, *S*, was then calculated as

$$S\left(\frac{g}{100\,g}\right)=0.9\times\frac{0.05G_{sample}}{1-\frac{MC_{wb}}{100}}\times\frac{100}{SR} \quad (2)$$

where the coefficient 0.9 was the correction factor for converting glucose estimates to starch content,[27] and 0.05 was the amount of dilute acid solution (in litres) used to dissolve 1 g of dry solids.

For ethanol fermentation, ground sorghum samples (30 g dry mass) were mixed with 70 mL of distilled water at 35°C to obtain slurries with 30 g solids per 100 g slurry. Slurry pH was adjusted to 6.0 using 5.0 N $H_2SO_4$ solution. Afterwards, the slurry was mixed with 8 mL of Spezyme CL (15,225 AAU $g^{-1}$, 1.17 g $mL^{-1}$, Dupont Industrial Biosciences, Palo Alto, CA), an enzyme containing thermostable α-amylase and 0.1 g of $KH_2PO_4$ solution, and then liquefied in a Labomat reactor (BFA 12, Werner Mathis AG, Concord, NC) at 85°C for 2 h with continuous agitation at 180 rpm. After 2 h, the slurry temperature was decreased to 30°C, and its pH was adjusted to 4.2 with 5.0 N $H_2SO_4$ solution. *Saccharomyces cerevisiae* yeast culture was prepared by dispersing 5 g of yeast in 25 mL of distilled water at 30°C for 20 min with shaking.

The simultaneous saccharification and fermentation (SSF) process was performed by adding 20 mL of Distillase® SSF (380 GAU $g^{-1}$, 1.14 g $mL^{-1}$, Dupont Industrial Biosciences, Palo Alto, CA), 1 mL of activated *S. cerevisiae*, 0.3 g of yeast extract (Fisher Scientific, Fair Lawn, NJ) and 10 mL of Fermgen (1000 SAPU $g^{-1}$, 1.14 g $mL^{-1}$, Dupont Industrial Biosciences, Palo Alto, CA), an acid proteolytic enzyme, into the mash. The mash was maintained at 30°C for 72 h with constant agitation at 150 rpm.

A 5 mL fermentation broth sample was collected at 0 and 72 h for high-performance liquid chromatography (HPLC) measurement. Each sample was centrifuged at 2348 × *g*. The supernatant was passed through a 0.2 µm syringe filter and transferred into 1 mL vials. The filtered supernatant was analysed using HPLC with an ion-exclusion column (Aminex HPX-87H, Bio-Rad, Hercules, CA) maintained at 50°C, and data were processed using HPLC software (Waters Corporation, Milford, MA). The ethanol conversion efficiency was calculated from a theoretical yield of 56.72 g of ethanol produced from 100 g of dry starch based on the assumption that 1 g of starch

could be hydrolysed into 1.11 g of glucose, and each gram of glucose could generate 0.511 g of ethanol.

An independent samples *t*-test was conducted using the Data Analysis Tool Pack in Microsoft Excel®[28] to compare the starch contents and ethanol yields of waxy and non-waxy sorghum grains. The null hypothesis was tested at $\alpha = 0.05$.

## NIR spectroscopy

FT-NIR spectra of dry ground sorghum grain were collected using a Spectrum™ Software One Near Infrared Testing System (NTS) (Perkin Elmer, Waltham, PA) using a spinner accessory, which held approximately 25 g of ground grain. Each sample was loaded into a near infrared reflectance accessory sample cup and levelled with a spatula. The spectrophotometer was set to record an average spectrum from 64 scans for the wavelength range 4000–10,000 $cm^{-1}$ at 4 $cm^{-1}$ resolution.

## Data processing and chemometrics

All spectral data were imported into Unscrambler® X.[29] The 48 samples were divided into a calibration set ($n = 39$) and validation set ($n = 9$) by first creating a histogram of the component to be modelled and randomly selecting nine samples that covered the component range. Because the number of sorghum cultivars available for this study was limited, the chemometric analysis was repeated two more times, each time with a different set of nine samples used for validation.

The spectral data were preprocessed with either multiplicative scatter correction (MSC)—first-order or second-order derivative with a Savitzky–Golay (SG) smoothing technique—or SG followed by MSC. The purpose of preprocessing the data was to remove multiplicative and additive effects owing to instrument settings or variations from sample and environmental conditions. Afterwards, PLSR models were developed for select regions of the FT-NIR spectra. The number of factors used in the PLSR model was determined by internal cross-validation. Resulting prediction models were validated and evaluated based on number of factors used in the model, coefficient of determination of calibration and validation, $R^2$ and $r^2$, respectively; root mean square error of prediction, *RMSEP*; ratio of deviation to performance, *RPD*; and ratio of the standard error of prediction to the range, *RER*. Models were evaluated for a low number of factors, high $R^2$ and $r^2$ values, low *RMSEP*, and high *RPD* and *RER* values. PLSR models with *RPD* values from 2.4 to 3.0 are adequate for rough screening purposes: 3.1–4.9, for screening, and 5.0–6.4, for quality control.[30] *RER* values from 7 to 12 also show that the model may be used for rough screening; 13–20, for screening, and 21–30, for quality control.[30]

# Results and discussion

## Starch content and ethanol yield

The starch content and ethanol yield of the 48 sorghum samples tested were comparable with those reported by Yan *et al.*[3] Starch content ranged from 614 g $kg^{-1}$ to 765 g $kg^{-1}$ with a mean and standard deviation of 697 ± 33 g $kg^{-1}$ (Table 1). Ethanol yield ranged from 263 g $kg^{-1}$ to 367 g $kg^{-1}$ with an average and

**Table 1. Sample information, starch content, ethanol (EtOH) yields and fermentation conversion efficiency of the 48 sorghum grain samples.**

| Accession no. | Waxy (+) or non-waxy | $MC_{wb}$ (%) | Starch (g $kg^{-1}$) | Theoretical EtOH yield (g $kg^{-1}$) | EtOH from HPLC (mL $L^{-1}$) | EtOH yield (g $kg^{-1}$) | Conversion efficiency (%, w/w) |
|---|---|---|---|---|---|---|---|
| 0006 | | 11.35 | 681.2 | 386.4 | 150.1 | 291.5 | 75.4 |
| 0015 | + | 9.52 | 613.8 | 348.1 | 148.6 | 286.0 | 82.1 |
| 0023 | | 11.29 | 710.1 | 402.8 | 159.6 | 309.8 | 76.9 |
| 0056 | | 10.37 | 673.8 | 382.2 | 176.5 | 341.1 | 89.2 |
| 0063 | + | 10.61 | 683.6 | 387.7 | 151.7 | 293.5 | 75.7 |
| 0079 | | 11.04 | 709.5 | 402.5 | 158.2 | 306.8 | 76.2 |
| 0099 | + | 10.61 | 703.5 | 399.0 | 161.2 | 312.0 | 78.2 |
| 0103 | | 9.98 | 682.7 | 387.2 | 158.6 | 305.9 | 79.0 |
| 0170 | | 11.58 | 665.6 | 377.5 | 16.30 | 316.8 | 83.9 |
| 0200 | | 11.01 | 739.9 | 419.7 | 185.1 | 358.9 | 85.5 |
| 0214 | | 10.28 | 654.7 | 371.3 | 166.3 | 321.2 | 86.5 |
| 0230 | + | 9.91 | 692.7 | 392.9 | 161.4 | 311.2 | 79.2 |
| 0258 | + | 10.66 | 743.6 | 421.8 | 162.4 | 314.3 | 74.5 |
| 0283 | | 11.00 | 702.6 | 398.5 | 158.6 | 307.5 | 77.2 |
| 0301 | | 10.80 | 678.4 | 384.8 | 176.2 | 341.2 | 88.7 |
| 0303 | | 10.88 | 676.4 | 383.6 | 167.3 | 324.1 | 84.5 |

Table 1. Continued

| Accession no. | Waxy (+) or non-waxy | $MC_{wb}$ (%) | Starch ($g\,kg^{-1}$) | Theoretical EtOH yield ($g\,kg^{-1}$) | EtOH from HPLC ($mL\,L^{-1}$) | EtOH yield ($g\,kg^{-1}$) | Conversion efficiency (%, w/w) |
|---|---|---|---|---|---|---|---|
| 0305 | + | 10.23 | 716.3 | 406.3 | 154.3 | 297.9 | 73.3 |
| 0370 | + | 10.07 | 741.4 | 420.5 | 183.6 | 354.1 | 84.2 |
| 0372 | | 10.33 | 764.7 | 433.7 | 189.8 | 366.6 | 84.5 |
| 0373 | + | 10.28 | 714.1 | 405.0 | 158.7 | 306.5 | 75.7 |
| 0382 | | 10.11 | 657.3 | 372.8 | 167.3 | 322.7 | 86.6 |
| 0387 | + | 11.26 | 692.6 | 392.8 | 156.9 | 304.6 | 77.5 |
| 0391 | | 10.00 | 697.4 | 395.6 | 164.8 | 317.8 | 80.3 |
| 0396 | + | 9.88 | 692.5 | 392.8 | 160.6 | 309.5 | 78.8 |
| 0397 | + | 10.59 | 678.1 | 384.6 | 161.3 | 312.0 | 81.1 |
| 0400 | + | 10.33 | 719.8 | 408.3 | 160.0 | 309.1 | 75.7 |
| 0414 | | 10.78 | 667.6 | 378.7 | 173.1 | 335.1 | 88.5 |
| 0551 | + | 10.68 | 737.1 | 418.1 | 153.2 | 296.5 | 70.9 |
| 0574 | + | 10.59 | 659.6 | 374.1 | 143.4 | 277.4 | 74.1 |
| 0577 | + | 10.58 | 641.0 | 363.6 | 135.9 | 262.8 | 72.3 |
| 0578 | + | 10.78 | 694.1 | 393.7 | 158.0 | 306.0 | 77.7 |
| 0615 | + | 10.80 | 655.8 | 372.0 | 143.7 | 278.4 | 74.8 |
| 0627 | | 10.56 | 708.4 | 401.8 | 179.1 | 346.5 | 86.2 |
| 0721 | + | 10.12 | 744.4 | 422.2 | 148.3 | 286.1 | 67.8 |
| 0747 | + | 10.46 | 685.1 | 388.6 | 158.2 | 305.9 | 78.7 |
| 0827 | + | 10.26 | 715.2 | 405.7 | 158.7 | 306.4 | 75.5 |
| 0834 | + | 9.56 | 631.1 | 358.0 | 146.4 | 281.7 | 78.7 |
| 0929 | | 9.89 | 688.6 | 390.6 | 155.5 | 299.8 | 76.8 |
| 0964 | | 10.53 | 665.8 | 377.6 | 143.6 | 277.7 | 73.5 |
| 1038 | | 10.56 | 736.2 | 417.6 | 169.5 | 327.8 | 78.5 |
| 1047 | | 10.70 | 720.1 | 408.4 | 165.8 | 320.8 | 78.5 |
| 1055 | | 11.02 | 737.3 | 418.2 | 166.2 | 322.1 | 77.0 |
| 1103 | | 10.15 | 709.1 | 402.2 | 164.8 | 318.1 | 79.1 |
| 1124 | | 11.08 | 755.1 | 428.3 | 163.6 | 317.3 | 74.1 |
| 1203 | | 9.76 | 674.5 | 382.6 | 170.3 | 328.0 | 85.7 |
| 1205 | + | 9.64 | 716.9 | 406.6 | 179.4 | 345.4 | 84.9 |
| 1277 | | 10.85 | 708.0 | 401.6 | 155.6 | 301.4 | 75.1 |
| 1333 | + | 10.57 | 697.9 | 395.8 | 150.2 | 290.5 | 73.4 |

standard deviation of $312 \pm 22\,g\,kg^{-1}$. The starch contents of the waxy grain samples ($694 \pm 36\,g\,kg^{-1}$) were not different from that of the non-waxy grain ($699 \pm 31\,g\,kg^{-1}$) ($p > 0.05$). However, the average ethanol yields were different for waxy grain ($302 \pm 20\,g\,kg^{-1}$) and non-waxy grain ($321 \pm 20\,g\,kg^{-1}$) ($p < 0.05$) despite having both mean values within one standard deviation of each other. As other researchers have reported for corn starch-to-ethanol yield relationships,[31,32] there was a low correlation between sorghum starch content and ethanol yield (Table 1 and Figure 1). The lack of correlation may be due to a number of factors such as non-extractable starch getting converted into ethanol; physical separation (milling) or thermal treatment; or the presence of tannins retarding the liquefaction process.[33] Nevertheless, other researchers have

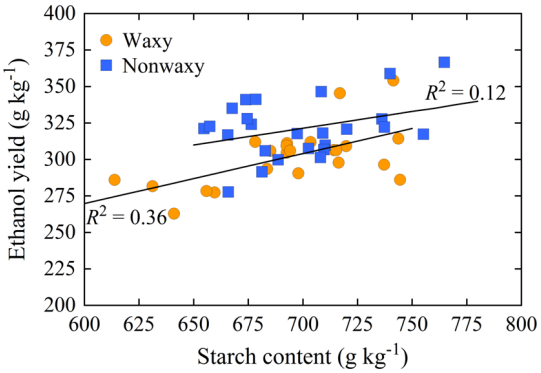


**Figure 1. Relationship between starch content and ethanol yield.**

demonstrated that NIR spectra can be correlated to ethanol yields.[34]

## NIR model development and validation

The FT-NIR spectra of the sorghum grain samples showed prominent peaks in the 4250–5500 cm$^{-1}$ region [Figure 2(a)]. The region around 5250 cm$^{-1}$ has been noted for O–H stretching and O–H deformation vibrations, which are normally observed in the 5180–5250 cm$^{-1}$ region. Lindberg and Kaila reported from their investigation of the adhesion and gelation mechanism of starch[16] that both amylose and amylopectin had strong absorption bands in this region and that this region could be used to measure the stability of amylose and amylopectin gel structures. Other researchers also attributed the amylose and amylopectin content to this region. In FT-NIR spectra of dry ground corn grain samples, the region around 5250 cm$^{-1}$ was also noted as an important waveband in predicting unreacted starch content.[34] Near this region, the 4695–4808 cm$^{-1}$ waveband and 4367–4396 cm$^{-1}$ waveband had been reported to be sensitive to increasing amylose levels and changes in the amylose–amylopectin ratio, respectively. Hence, the prominent peaks in the 4250–5500 cm$^{-1}$ are critical for classification of waxy vs. non-waxy grain[35] (Figure 2(b) and 2(c)] and quantification of starch content.

In general, PLSR models for starch content based on MSC-pretreated spectra had low $R^2$, $r^2$, *RPD* and *RER* values, and high root mean square error of cross-validation (*RMSECV*) and *RMSEP* values. First derivative-based models performed better with $R^2 > 0.90$, $0.80 < r^2 < 0.90$, *RMSECV* = 12.5–13.8 g kg$^{-1}$; *RMSEP* = 11.8–13.6 g kg$^{-1}$. Resulting models are useful for rough screening of sorghum cultivars as *RPD* values ranged from 2.53 to 2.90, and *RER* values ranged from 7 to 9. While most second-derivative-based models developed had performances comparable with those of first derivative-based models, one model focusing solely on the 4000–6000 cm$^{-1}$ region, pretreated with a combination of SG with 49 points and MSC, and utilising three factors had the following values: $R^2 = r^2 = 0.97$, *RMSECV* = 28.7 g kg$^{-1}$; *RMSEP* = 5.5 g kg$^{-1}$; *RPD* = 2.90 and *RER* = 15 (Table 2). In some screening applications, theoretical ethanol yields of different cultivars may be more useful to know than starch content. When the spectral information was calibrated against theoretical ethanol yields, the best model developed utilised the 4000–8000 cm$^{-1}$ wavebands, pretreated with first-derivative SG with 49 points followed by MSC. Both PLSR models for starch content and theoretical ethanol yield had high regression coefficients in the important wavebands of 4000–5000 cm$^{-1}$ and 5180–5250 cm$^{-1}$ (Figure 3).

Predicting actual ethanol yields was more difficult than predicting starch content and theoretical yield. Most models developed, regardless of spectral region and pretreatment used, had $R^2$ and $r^2$ values of <0.60, *RMSEP* = 11.2–21.4 g kg$^{-1}$, *RPD* < 3 and *RER* < 6. One model, based on 4000–6000 cm$^{-1}$ wavebands pretreated with first-derivative SG followed by MSC, yielded an *RPD* = 3.2 and *RER* = 11 (Table 2). This model barely passed the criteria for rough screening applications. Poor model performance was likely due to the ground sorghum samples having been scanned prior to fermentation. Any variation in fermentation conditions, such as yeast and enzyme activity, would have affected actual ethanol yields but was not included in the FT-NIR spectra. This is why other researchers have found a low correlation between starch content (obtained before fermentation) and actual ethanol yield (obtained after fermentation).[31,32] Hao *et al.*[6] reported PLSR models with *RPD*

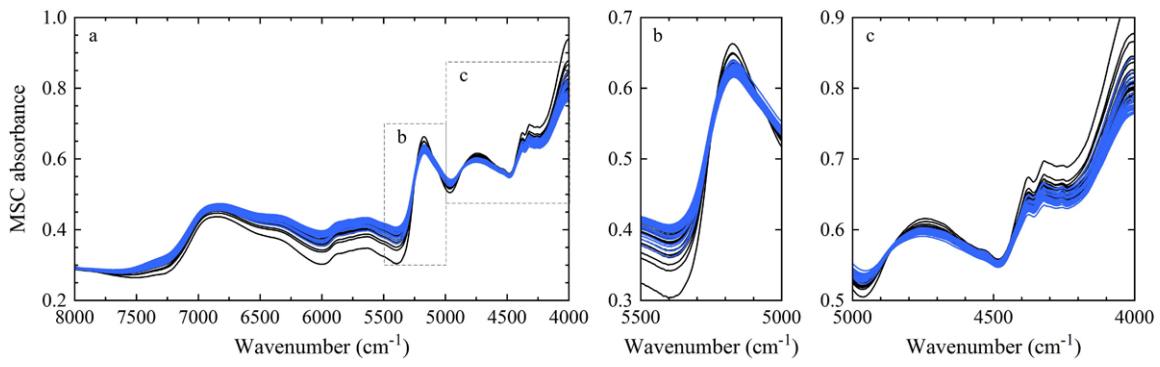


**Figure 2. (a) FT-NIR spectra of waxy grain (black curves) and non-waxy grain (blue curves). Waxy grain had higher absorbances in (b) the 5000–5500 cm$^{-1}$ region and (c) the combination region, 4000–5000 cm$^{-1}$.**

**Table 2. PLSR models of starch content, theoretical and actual ethanol yields.**

| | Starch content (g kg$^{-1}$) | Theoretical ethanol yield (g kg$^{-1}$) | Actual ethanol yield (g kg$^{-1}$) |
|---|---|---|---|
| Spectral range (cm$^{-1}$) | 4000–6000 | 4000–8000 | 4000–6000 |
| Data preprocessing[a] | SG-2 + MSC | SG-1 + MSC | SG-1 + MSC |
| Calibration set (*n* = 39) | | | |
| Range (g kg$^{-1}$) | 150.9 | 85.6 | 103.8 |
| *SD* (g kg$^{-1}$) | 32.9 | 18. 8 | 22.2 |
| Validation set (*n* = 9) | | | |
| Range (g kg$^{-1}$) | 87.1 | 55.5 | 81.5 |
| *SD* (g kg$^{-1}$) | 34.3 | 19.3 | 23.2 |
| No. of factors | 3 | 5 | 7 |
| $R^2$ | 0.97 | 0.91 | 0.75 |
| $r^2$ | 0.97 | 0.96 | 0.90 |
| *RMSECV* (g kg$^{-1}$) | 28.7 | 7.7 | 17.0 |
| *RMSEP* (g kg$^{-1}$) | 5.5 | 5.6 | 12.0 |
| *SEP* (g kg$^{-1}$) | 5.8 | 4.3 | 7.2 |
| Bias (g kg$^{-1}$) | 0.0 | -3.9 | -9.9 |
| *RPD* | 5.9 | 4.5 | 3.2 |
| *RER* | 15 | 13 | 11 |

[a]SG-1 and SG-2 derivative transformation was based on a second-order polynomial with 49 data points.

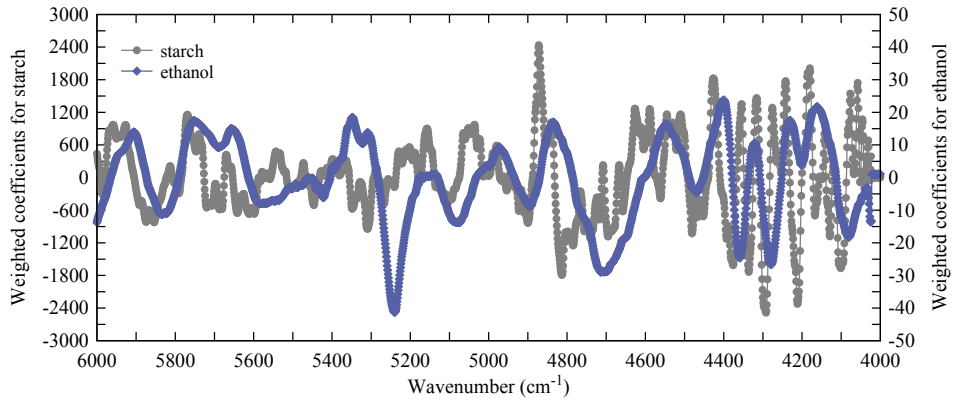


**Figure 3. Weighted regression coefficients for starch and theoretical ethanol PLSR models. Outside the combination region, the starch model had high regression coefficients at 5960 cm$^{-1}$, 5763 cm$^{-1}$, 5228 cm$^{-1}$ and 5188 cm$^{-1}$, while the ethanol model had high regression coefficients at 5905 cm$^{-1}$, 5753 cm$^{-1}$, 5657 cm$^{-1}$, 5348 cm$^{-1}$, 5240 cm$^{-1}$, 4973 cm$^{-1}$, 4835 cm$^{-1}$ and 4625 cm$^{-1}$.**

values between 1.1 and 1.2 when they calibrated NIR spectra of ground corn to ethanol yields. Rathore *et al.*[34] reported PLSR models with *RPD* = 6.0 when they calibrated NIR spectra of the supernatants from fermentation flasks to the ethanol yield of corn samples. In addition, the use of more replications of ethanol yield measurements and incorporating other factors such as free amino nitrogen (FAN) and tannin contents into the model will help improve the modelling of actual ethanol yields. High levels of FAN and tannin usually retard the liquefaction process,[34] so their concentrations will need to be accounted for when predicting ethanol yield.

# Conclusion

The PLSR models developed in this study show that sorghum grain may be screened for its starch content and theoretical ethanol yield using FT-NIR spectroscopy. The starch model used important wavebands in the 4000–6000 cm$^{-1}$ region while the theoretical ethanol yield model was based on a wider 4000–8000 cm$^{-1}$ region. Since starch is not often strongly correlated with actual ethanol yield, additional information regarding composition and fermentation conditions may be added to the models and enhance their prediction performance.

## Acknowledgements

The authors acknowledge Ming-Hsu Chen and Wei Liu for their technical assistance. We thank the Center for Applied Bioenergy Research (CABER) and the Illinois Plant Breeding Center (IPBC) for providing funding support of this research project. We also thank Camo Software Inc. for providing the multivariate data analysis software, Unscrambler®, used in this study.

## References

1. RFA (Renewable Fuels Association), *2012 Ethanol Industry Outlook: Accelerating Industry Innovation*. Renewable Fuels Association, Washington, DC (2012).
2. W.L. Rooney, J. Blumenthal, B. Bean and J.E. Mullet, "Designing sorghum as a dedicated bioenergy feedstock", *Bioprod. Bioref.* **1,** 147 (2007). doi: http://dx.doi.org/10.1002/bbb.15
3. S. Yan, X. Wu, S.R. Bean, J.F. Pedersen, T. Tesso, Y. Chen and D. Wang, "Evaluation of waxy grain sorghum for ethanol production", *Cereal Chem.* **88,** 589 (2011). doi: http://dx.doi.org/10.1094/CCHEM-04-11-0056
4. L.W. Rooney and S.O. Serna-Saldivar, "Sorghum", in *Handbook of Cereal Science and Technology*, 2nd Edn, Ed by K. Pulp and J.G. Ponte Jr. Marcel Dekker, New York, pp. 149–176 (2000).
5. J.A. Lacerenza, J.M. Martin, L.E. Talbert, S.P. Lanning and M.J. Giroux, "Relationship of ethanol yield to agronomic and seed quality characteristics of small grains", *Cereal Chem*. **85,** 322 (2008). doi: http://dx.doi.org/10.1094/CCHEM-85-3-0322
6. X. Hao, K. Thelen and J. Gao, "Prediction of the ethanol yield of dry-grind maize grain using near infrared spectroscopy", *Biosyst. Eng.* **112,** 161 (2012). doi: http://dx.doi.org/10.1016/j.biosystemseng.2012.03.007
7. P. Williams and K. Norris, *Near Infrared Technology: In the Agricultural and Food Industries*, 2nd Edn. American Association of Cereal Chemists, St Paul, MN (2001).
8. B.R. Hames, S.R. Thomas, A.D. Sluiter, C.J. Roth and D.W. Templeton, "Rapid biomass analysis", *Appl. Biochem. Biotech.* **105–108**, 5 (2003). doi: http://dx.doi.org/10.1007/978-1-4612-0057-4_1
9. G. Spielbauer, P. Armstrong, J.W. Baier, W.B. Allen, K. Richardson, B. Shen and A.M. Settles, "High-throughput near-infrared reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels", *Cereal Chem.* **86,** 556 (2009). doi: http://dx.doi.org/10.1094/CCHEM-86-5-0556
10. K.P. Vogel, B.S. Dien, H.G. Jung, M.D. Casler, S.D. Masterson and R.B. Mitchell, "Quantifying actual and theoretical ethanol yields for switchgrass strains using NIRS Analyses", *Bioenerg. Res.* **4,** 96 (2011). doi: http://dx.doi.org/10.1007/s12155-010-9104-4
11. L. Liu, P. Ye, A.R. Womac and S. Sokhansanj, "Variability of biomass chemical composition and rapid analysis using FT-NIR techniques", *Carbohyd. Polym.* **81,** 820 (2010). doi: http://dx.doi.org/10.1016/j.carbpol.2010.03.058
12. D.W. Templeton, A.D. Sluiter, T.K. Hayward, B.R. Hames and S.R. Thomas, "Assessing corn stover composition and sources of variability via NIRS", *Cellulose* **16,** 621 (2009). doi: http://dx.doi.org/10.1007/s10570-009-9325-x
13. M. Sohn, D.S. Himmelsbach, F.E. Barton, C.A. Griffey, W. Brooks and K.B. Hicks, "Near-infrared analysis of ground barley for use as a feedstock for fuel ethanol production", *Appl. Spectrosc.* **61,** 1178 (2007). doi: http://dx.doi.org/10.1366/000370207782597148
14. T.M. Baye, Pearson, T.C. and A.M. Settles, "Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy", *J. Cereal Sci.* **43,** 236 (2006). doi: http://dx.doi.org/10.1016/j.jcs.2005.11.003
15. M. Hódsági, S. Gergely, T. Gelencsér and A. Salgó, "Investigations of native and resistant starches and their mixtures using near-infrared spectroscopy", *Food Bioprocess. Tech.* **5,** 401 (2012). doi: http://dx.doi.org/10.1007/s11947-010-0491-5
16. J.J. Lindberg, and T. Kaila, "NIR and spin probe studies on amylose and amylopectin", *Acta Chem. Scand. B* **34,** 757 (1980). doi: http://dx.doi.org/10.3891/acta.chem.scand.34b-0757
17. M.R. Paulsen, S.W. Mbuvi, A.E. Haken, B. Ye and R.K. Stewart, "Extractable starch as a quality measurement of dried corn", *Appl. Eng. Agric.* **19,** 211 (2003). doi: http://dx.doi.org/10.13031/2013.13097
18. Paulsen, M.R. and M. Singh, "Calibration of a near-infrared transmission grain analyzer for extractable starch in maize", *Biosyst. Eng.* **89,** 79 (2004). doi: http://dx.doi.org/10.1016/j.biosystemseng.2004.05.009
19. J. Fontaine, B. Schirmer and J. Horr, "Near-infrared reflectance spectroscopy (NIRS) enables the fast and accurate prediction of essential amino acid contents. 2. Results for wheat, barley, corn, triticale, wheat bran/middlings, rice bran and sorghum", *J. Agr. Food Chem.* **50,** 3902 (2002). doi: http://dx.doi.org/10.1021/jf011637k
20. C.M. McGoverin, P. Elgelbrecht, P. Geladi and M. Manley, "Characterization of non-viable whole barley, wheat and sorghum grains using near-infrared hyperspectral data and chemometrics", *Anal. Bioanal. Chem.* **401,** 2283 (2011). doi: http://dx.doi.org/10.1007/s00216-011-5291-x
21. M.X. Liu, Y.W. Wang, and J.G. Han, "Determination of polyphenols in sorghum grains by near infrared spectroscopy", *Chin. J. Anal. Chem.* **37,** 1275 (2009).
22. L.F. Alencar Figeuiredo, F. Davrieux, G. Fliedel, J.F. Rami, J. Chantereau, M. Deu, B. Courtois and C. Mestres, "Development of NIRS equations for food grain quality traits through exploitation of a core collection of cultivated sorghum", *J. Agr. Food Chem.* **54,** 8501 (2006). doi: http://dx.doi.org/10.1021/jf061054g

**23.** C.A. Roberts, J.H. Houx III and F.B. Fritschi, "Near-infrared analysis of sweet sorghum bagasse", *Crop Sci.* **51,** 2284 (2011). doi: http://dx.doi.org/10.2135/cropsci2010.12.0691

**24.** E. Wolfrum, C. Payne, T. Stefaniak, W. Rooney, N. Dighe, B. Bean and J. Dahlberg, Multivariate calibration models for sorghum composition using near-infrared spectroscopy, Technical Report NREL/TP-5100-56838, National Renewable Energy Laboratory (NREL), Golden, CO (2013).

**25.** S.F. Chen, M.G.C. Danao, V. Singh and P.J. Brown, "Determining sucrose and glucose levels in dual-purpose sorghum stalks by Fourier transform near infrared (FT-NIR) spectroscopy", *J. Sci. Food Agr* **94,** 2569–2576. doi: http://dx.doi.org/10.1002/jsfa.6606 (2014)

**26.** AACC International (AACCI), "Moisture—air-oven method, drying at 135°C, method 44-19.01", in *Approved Methods of* Analysis, 11th Edn. The American Association of Cereal Chemists, St Paul, MN (2010).

**27.** J. Holm, I. Björk, A. Drews and N.G. Asp, "A rapid method for the analysis of starch", *Starch/Stärke* **38,** 224 (1986). doi: http://dx.doi.org/10.1002/star.19860380704

**28.** Microsoft Excel®, 2013. Microsoft Corporation, Redmond, WA.

**29.** Unscrambler®, 2011, Version 10.1. Camo Software, Woodbridge, NJ.

**30.** P. Williams, *Near Infrared Technology—Getting the Best Out of Light. A Short Course in the Practical Implementation of Near-Infrared Spectroscopy for the User.* Ed. 5.3. PDK Projects, Nanaimo, Canada (2008).

**31.** D. Haefele, D. Sevenich, D. Jones, J. Janni and S. Wright, "NIR-based analytical systems enable development of corn seed products and corn grain markets that are specific for dry-grind ethanol production", *Int. Sugar J.* **109,** 154 (2007).

**32.** V. Singh and J.V. Graeber, "Effect of corn hybrid variability and planting location on ethanol yields", *Trans. ASABE* **48,** 709 (2005). doi: http://dx.doi.org/10.13031/2013.18301

**33.** X. Wu, R. Zhao, S.R. Bean, P.A. Seib, J.S. McLaren, R.L. Madl, M. Tuinstra, M.C. Lenz, and D. Wang, "Factors impacting ethanol production from grain sorghum in the dry-grind process", *Cereal Chem.* **84,** 130 (2007). doi: http://dx.doi.org/10.1094/CCHEM-84-2-0130

**34.** S.S.S. Rathore, M.R. Paulsen, V. Sharma and V. Singh, "Use of near-infrared spectroscopy for monitoring fermentation in a corn dry grind ethanol process", *Trans. ASABE* **50,** 2337 (2007). doi: http://dx.doi.org/10.13031/2013.24086

**35.** B.M. Plumier, M.G.C. Danao, V. Singh and K.D. Rausch, "Analysis and prediction of unreacted starch content in corn using FT-NIR spectroscopy", *Trans. ASABE* **56,** 1877 (2013). doi: http://dx.doi.org/10.13031/trans.56.10