The *Journal of Cotton Science* 2:23-26 (1998)
http://journal.cotton.org, © The Cotton Foundation 1998

COTTON IMPROVEMENT

Evaluation of Near Infrared Reflectance for Oil Content of Cottonseed

R.J. Kohel*

INTERPRETIVE SUMMARY

Cotton improvement is based on the production and sale of fiber, but cottonseed products are an important secondary product. Cottonseed is a major oilseed in domestic and international markets. Currently, no ready methods exist that can be used in cotton breeding and testing programs to routinely evaluate oil content of cottonseed. We are reporting the development and evaluation of methods calibrating near infrared reflectance spectroscopy (NIRS) instrumentation for the measurement of oil content in cottonseed. Six cotton lines were grown at two locations for three years to generate prediction equations for oil content of cottonseed. The equations were generated with NIRS based on seed-oil contents from continuous wave nuclear reflectance spectroscopy (NMR) analysis. The prediction equations were evaluated in the test material and the National Cotton Variety Test entries. The prediction equations were capable of identifying differences in seed-oil content within our test populations, but a universal prediction of seed-oil content will require a larger sample base to eliminate spectral variability that interferes with calibration.

ABSTRACT

The status of ongoing research to use near infrared reflectance spectroscopy (NIRS) to measure cottonseed oil content is summarized. A data set of 36 seed samples consisting of a single replication of six entries grown at two locations for 3 yr was used to generate prediction equations for oil content of cottonseed (*Gossypium hirsutum* L.). The equations were generated with NIRS based on seed-oil contents from continuous wave nuclear magnetic resonance (NMR) analysis. The results in predicting the remaining four replications in the tests, 144 samples, demonstrate the ability to develop prediction equations capable of identifying differences in seed-oil content within specified populations (r^2 = 0.89-–0.92). However, these prediction equations were not as successful when tested on the National Cotton Variety Test entries (r^2 = 0.00-0.84). A universal prediction of seed-oil content will require a larger sample base to eliminate spectral variability that might interfere with calibration.

Plant improvement depends on the ability to evaluate large numbers of individuals. Usually such ability requires methods of evaluation that are rapid and inexpensive, yet precise and accurate enough to make the required separations among genetic segregants. In the genetic selection of oil content of seed, the NMR instrument is a valuable tool. This instrument has the advantage of being nondestructive of the sample being evaluated. In investigations of ways to genetically improve oil seeds, the sample consists of viable seeds. In the case of cottonseed, alternate probe sizes can accommodate approximately 1 or 10 g seed samples that are suitable for different types of experimentation. The disadvantage of the NMR instrument is that it is limited to the measurement of oil content, so expansion of the number of constituents under investigation requires other methods. Although additional methods are needed, the NMR instrument probably will not be replaced because of its nondestructive nature; furthermore, the sizes of sample that it can accommodate make it desirable for many experiments involving oil content of seeds. A type of instrumentation that has been introduced in many quality evaluation programs is that utilizing NIRS.

These NIRS instruments have the advantage of expanding the range of constituents evaluated.

USDA-ARS, Crop Germplasm Research Unit, 2765 F&B Road, College Station, TX 77845. Contribution from USDA-ARS, College Station, TX, in cooperation with the Texas Agric. Exp. Stn. Received 15 Aug. 1997. *Corresponding author (rjk0339@acs.tamu.edu).

Abbreviations: ANOVA, analysis of variance; NIRS, near infrared reflectance spectroscopy; NMR, continuous wave nuclear magnetic resonance.

Although NIRS instruments are nondestructive of the sample, current practice that measures reflectance from the seed sample requires grinding of the material to be evaluated.

The cottonseed quality improvement program at College Station, TX, was limited to measurement of oil content by the NMR method. This method has been very useful in our genetic investigations because it measures the total oil content of the seed, not just that extracted by solvent procedures. The method's nondestructive nature allowed us to grow the seed evaluated (Kohel, 1980). In this paper the methods and results of calibrating NIRS instrumentation for the measurement of oil content in cottonseed are presented.

MATERIALS AND METHODS

Since the cottonseed quality improvement program at College Station, TX, has previously used NMR as its standard measurement for oil content, it was chosen as the standard on which to calibrate the NIRS spectra. Cottonseed samples prepared for oil measurement by NMR are routinely acid delinted, forced air dried at 38°C for 24 h, and equilibrated to room temperature before measurement. The usual sample size is 10 g. The NMR integration period is approximately 30 s. The instrument is calibrated to display oil content of the 10 g sample. A standard of 2.5 g cottonseed oil mixed with glass beads is periodically measured to check the instrument. The average of two readings for each sample is recorded.

Samples for NIRS analysis were measured in the same manner with the NMR, except that a 2 min integration period was used to obtain greater precision. The following procedures were adopted for preparation of samples for the NIRS instrument. A 5 g aliquot of the 10 g NMR sample was ground in a coffee bean grinder. A special top for the grinder was fabricated from plexiglass to reduce the chamber size. This modification was made to enable the use of a 5 g sample and to improve mixing of the sample during grinding. A 5 g cottonseed sample size was selected because this size was the amount needed to fill the sample cup chamber and limit waste. Grinding consisted of three 10 s grindings. The chamber was shaken during grinding to increase mixing, and between grindings the sample was mixed to minimize caking and separation by particle size. The NIRS system and software were those described by Shenk et al. (1981). The only change was the modification to provide a rotating sample cup.

Samples selected for calibration were from a test with five replications of six entries grown in 1979, 1980, and 1982 at two locations on the Texas A&M University Brazos River Farm. One replication was used for calibration (one replication by six entries by two locations by 3 yr = 36 samples) to predict the remaining four replications (144 samples). The entries included: 'Tamcot SP-37'; Lyman, GN-2, and DOR-S from L. S. Bird; Texas Marker-1 (Kohel et al., 1970); and 'Stoneville 213'. Lyman and GN-2 are glandless lines and the remainder are normally glanded. These entries were selected because they represented a range in seed-oil content, gossypol content, seed size and maturity.

RESULTS AND DISCUSSION

Spectral data were collected and processed using the NIRS system previously described (Shenk et al., 1981). Mathematical treatment of the spectral data included first and second derivative options and averaging over various wavelength width options. The treatment that provided the set of prediction equations that best fit the data was the first derivative with averaging over 4 nm segments. It was possible to create equations with up to seven wavelengths from the 36 samples (Table 1). The R^2 value for fit of the equations to the calibration data ranged from 0.92 to 0.98 for the one- to seven-

Table 1. Prediction equations for oil content of cottonseed with one to seven wavelengths from the 3 yr of data.†

$y1 = 13.32 - 1549.39\ (X_{2343})$
$y2 = 10.70 - 1681.37\ (X_{2343}) + 1118.34\ (X_{1747})$
$y3 = 15.16 - 2770.08\ (X_{2343}) + 1109.87\ (X_{2303}) + 979.80\ (X_{1487})$
$y4 = 19.20 + 826.88\ (X_{2239}) + 1797.07\ (X_{1739}) + 1130.63\ (X_{1487}) + 253.29\ (X_{1527})$
$y5 = 19.27 + 722.33\ (X_{2239}) + 1929.93\ (X_{1739}) + 1092\ (X_{1487}) + 272.07\ (X_{1527}) - 921.67\ (X_{1811})$
$y6 = 19.95 + 856.65\ (X_{2239}) + 1854.17\ (X_{1739}) + 1401.67\ (X_{1487}) + 245.50\ (X_{1527}) - 1898.47\ (X_{1807}) + 805.49\ (X_{1215})$
$y7 = 21.41 + 697.73\ (X_{2239}) + 1095.37\ (X_{1739}) + 1399.35\ (X_{1491}) + 703.29\ (X_{1527}) - 1430.88\ (X_{1811}) + 1709.73\ (X_{1219}) + 1023.03\ (X_{1371})$

† **Where y is the predicted seed-oil content and X_{2343}, etc., is the first derivative of log (1/reflectance value) at the wavelength of 2343 nm.**

Table 2. Calibration and prediction statistics associated with near infrared reflectance spectroscopy (NIRS) analysis of oil content of cottonseed for prediction equations with one to seven wavelengths and the associated SE of the mean seed-oil content.

Wavelengths	Calibration samples		Prediction samples	
	R^2	SE	r^2	SE
no.		g kg^{-1}		g kg^{-1}
1	0.916	0.77	0.886	0.86
2	0.930	0.71	0.892	0.88
3	0.939	0.66	0.905	0.80
4	0.976	0.42	0.919	0.71
5	0.977	0.40	0.923	0.70
6	0.981	0.36	0.917	0.74
7	0.985	0.32	0.917	0.73

wavelength equations, respectively (Table 2). These seven prediction equations were used to predict seed-oil content in the remaining 144 samples and to compare the actual values (NMR) with the seven sets of NIRS predicted values. The equation with one wavelength had the lowest (0.89) and the equation with five wavelengths had the highest (0.92) r^2 value between NMR and NIRS values (Table 2). The predicted values for one and five wavelengths and the NMR values for the prediction samples were analyzed by analysis of variance (ANOVA). The results of the analyses and means for the six entries are shown in Table 3. All three methods produced similar results in separating entry, year, and location effects.

Experience has shown us that seed-oil content is subject to large environmental variation (Cherry et al., 1978, 1981; Kohel, 1980). The repetitive growing of this specific set of entries was done to include a range of environmental variability. The entries selected for a range in seed-oil content, gossypol content, seed size, and maturity should represent a desirable group for the calibration of the NIRS spectra because they include several factors that influence seed-oil content. The diversity of materials and environments should include NIRS spectral variability that will facilitate selection of wavelengths most suited to the prediction of seed-oil content.

Remnant seed of the National Cotton Variety Test seed evaluations was supplied by H.H. Ramey. The seven prediction equations were used to predict seed-oil content in these samples. Four national standards were predicted for locations within each of the six regions. The one-wavelength and five-wavelength equations were the best predictors of the laboratory values, however, the r^2 value between predicted and actual values ranged from 0.00 to 0.84 and 0.00 to 0.81 for the one- and five-wavelength equations, respectively, over the six regions. The one-wavelength equation was as good as or better than the equation with five wavelengths for these data, average r^2 values were 0.48 and 0.46, respectively.

We have experimented with several data sets for calibration of the NIRS spectra. The most successful calibrations (large r^2 values between actual and predicted measurements) are within a single year. When calibrations are extended over years and locations, they are less satisfactory, as observed in the data reported in this paper for 3 yr and two locations and the National Cotton Variety Test. However, the purposes for which the seed-oil contents are to be used influence the utility of prediction equations. The data in Table 3 demonstrate that the one- and five-wavelength equations are capable of separating differences in

Table 3. Analysis of variance for the three main effects and the residual error term and means of seed-oil contents from actual nuclear magnetic resonance (NMR) and the one and five wavelength near infrared reflectance spectroscopy (NIRS) prediction equations for cottonseed.

Source (DF)		Actual	Predicted	
			One wavelength	Five wavelengths
		----------Mean squares ----------		
Entries (5)		11.67**	17.28**	14.04**
Years (2)		287.64**	296.07**	288.86**
Locations(1)		1.87	9.20*	7.02*
Residual (105)		0.54	0.69	0.76
		---------- Means, g kg^{-1} ----------		
Entries				
SP-37	1	27.3	27.3	27.5
DOR-S	2	27.2	27.6	27.7
GN-2	3	27.7	28.2	27.7
Lyman	4	28.1	28.4	28.2
S-213	5	26.2	26.1	26.1
TM-1	6	26.7	26.8	26.8
Years				
	1	28.5	29.3	29.0
	2	28.8	28.3	28.5
	3	24.4	24.6	24.5
Locations				
	1	27.3	27.7	27.6
	2	27.1	27.2	27.1

*,** Significance at the 0.05 and 0.01 levels of probability.

our sources of variation. For a genetic and breeding program, the major need is to separate the sources of variation. For other uses such as marketing and processing, a more precise estimation of seed-oil content is needed. The continued inclusion of materials in the calibration database should provide greater accuracy in predicting seed-oil content. For research purposes it is difficult to compare seed-oil content unless a standard procedure is used. At this stage of development, using cottonseed samples that are dried and acid delinted is recommended. Use of such seed allows comparability between NMR and NIRS instrumentation. The use of whole fuzzy seed introduces several sources of variability that influence seed-oil content, that is, differences in ginning that change fuzz content or genetic differences in fuzz amounts.

REFERENCES

Cherry, J.P., R.J. Kohel, L.A. Jones, and W.H. Powell. 1981. Cottonseed quality: Factors affecting food and feed uses. p. 266-283. *In* J.M. Brown (ed.) Proc. Beltwide Cotton Prod. Res. Conf., New Orleans, LA. 4–8 Jan. 1981. Natl. Cotton Council Am., Memphis, TN.

Cherry, J.P., J.G. Simmons, and R.J. Kohel. 1978. Potential for improving cottonseed quality by genetic and agronomic practices. p. 343-364. *In* M. Friedman (ed.) Nutritional improvement of food and feed proteins. Plenum Publ. Co., New York.

Kohel, R.J. 1980. Genetic studies of seed oil in cotton. Crop Sci. 20:784-787.

Kohel, R.J., T.R. Richmond, and C.F. Lewis. 1970. Texas Marker-1. Description of a genetic standard for *Gossypium hirsutum* L. Crop Sci. 10:670-671.

Shenk, J.S., I. Landa, M.R. Hoover, and M.O. Westerhaus. 1981. Description and evaluation of a near infrared reflectance spectro-computer for forage and grain analysis. Crop Sci. 21:355-358.