Near-Infrared Reflectance for Analysis of Cottonseed for Gossypol[1]

G. S. BIRTH[2] and H. H. RAMEY, Jr.[3]

Cereal Chem. 59(6):516-519

ABSTRACT

Gossypol, a toxic yellow pigment in glanded cottonseed, is a major obstacle to the use of cottonseed for human consumption. Near-infrared reflectance spectroscopy (NIR) was investigated as a means of rapidly determining the gossypol content of cottonseed. Cottonseed samples were delinted and ground. Reflectance spectra from 1.0 to 2.38 μm were recorded followed by derivative data processing. The reflectance data and data obtained from chemical analysis of the cottonseed samples were used in a stepwise multiple regression analysis to develop an equation to express the gossypol content of cottonseed based on reflectance spectra alone. The results, using 30 samples, yielded a multiple correlation coefficient of 0.987 for a three-term equation, and a standard error of 0.074. Comparable results were obtained when the analysis was repeated with an independent group of samples.

Cottonseed (*Gossypium* spp.) meal is an excellent protein concentrate for cattle and sheep and has potential as a protein source for humans and other animals if the toxic gossypol can be removed (Martinez 1977). Gossypol, a bright yellow pigment occurring in small bulb-shaped glands in cottonseed, can cause death to nonruminant animals by reducing the oxygen-carrying capacity of the blood (Abou-Donia et al 1969). Although glandless cotton varieties exist that have a negligible gossypol content (<0.1%), most cottonseed produced in the United States is of the glanded type, in which the gossypol varies from 0.4 to 1.7% of the moisture-free meats (Hoffpauir and Pons 1953).

Considerable attention has been given to developing procedures for removing gossypol from cottonseed. Gardner et al (1976) reported on a liquid cyclone process capable of consistently removing pigment glands from cottonseed to produce a high-protein, edible flour. According to U.S. Food and Drug regulations (Federal Register 1974), cottonseed protein products for human consumption must contain no more than 0.045% "free" gossypol.

Conventional procedures for gossypol analysis consist of an extraction procedure followed by colorimetric determination of the the concentration (AOCS 1978). Because the extraction procedure is too slow for quality control purposes, a more rapid method for determining gossypol content of cottonseed is needed.

Near-infrared reflectance (NIR) offers a possiblity for measuring gossypol content more rapidly. This technique has been used in many applications for rapid composition analysis (Norris et al 1976). The oil and protein contents of corn, soybean, and oat seeds were measured on an instrument that used six wavelengths (Hymowitz et al 1974).

The physical definitions of the reflection process can be confusing. In optically clear nonmetals, the reflectance (about 5% of the incident radiation) is limited to radiation reflected at the surface. For optically clear materials, reflectance is not influenced by absorbers in the material. Absorbing constituents affect only the transmitted radiation. Most nonmetals are not optically clear but include large numbers of microscopic internal optical interfaces because of the cellular or granular nature of the material. The transmitted radiation (about 95% of the incident radiation) becomes diffuse under that condition, is distributed in all directions, and is subject to absorption. The radiation that is not absorbed leaves the material close to the illuminated area and thus contributes to the reflectance. This portion of the reflectance, referred to as body reflectance, is the major factor in determining the NIR of all organic materials (Birth 1976).

Body reflectance has the same relation to the concentration of absorbers as transmittance has in conventional spectrophotometry. For diffuse body reflectance, the effective path length is an unknown function of the mean distance between the internal optical interfaces and the relative index of refraction at each interface. The gross effect is referred to as light scattering. Current NIR procedures aim to maintain uniform particle size so that light scattering will be constant.

Use of NIR technology to measure seed quality is affected by distribution of constituents in the seed and by the many constituents in the sample. Each constituent has a unique spectral absorbance curve with absorption peaks that may overlap those of other constituents. The constituent distribution problem can be resolved by grinding the seed to create a nearly homogeneous sample. Grinding may not be essential for many applications (Tkachuk 1981). The multiple constituent problem is resolved by wavelength analysis and mathematical treatment of the data.

Current procedures for measuring constituent concentration with NIR are empirical. Many problems have been resolved by trial and error, and the procedures giving the best results provide guidelines for subsequent research. Reflectance spectra of several samples are recorded over as wide a wavelength range as possible and with automatic digital data acquisition. Derivative computation (O'Haver and Green 1975) is frequently used as the mathematical treatment. Data smoothing is usually incorporated in the processing to reduce the effects of noise generated in the analog circuits. Finally, a multiple stepwise regression analysis is implemented. The additional proof necessary to provide adequate confidence that the relation is valid can be obtained by repeating the experiment with an independent group of samples or by applying the regression equation to an alternate group of samples.

MATERIALS AND METHODS

Sample Preparation

For the 1978 tests, 139 samples of 55 varieties and breeding lines were screened for free gossypol content. Thirty of these were selected for gossypol levels from 0.1 to 1.4%. These samples were dehulled.

For the 1979 tests, 46 samples were obtained from the 1977 National Cotton Variety Testing Program, and two of the 30 samples from 1978 were included. Gossypol levels ranged from 0.1 to 1.46%. The samples were acid-delinted to remove the fuzz fibers from the seed coat. All samples were ground with a Moulinex model MX-228 electric grinder. A 5-sec grinding period was followed by a 30-sec interval during which the mill was inverted to prevent packing, and then by another 5-sec grinding period.

Chemical Analysis

The cottonseed was chemically analyzed in two commercial laboratories and in our laboratory with the AOCS method for free gossypol.

NIR Procedures

A computer-controlled single-beam spectrophotometer

[1] Mention of a company or product name does not imply its approval or recommendation by the USDA to the exclusion of others that may be suitable.

[2] Research agricultural engineer, Quality Evaluation Laboratory, USDA, Russell Research Center, Athens, GA 30613.

[3] Research geneticist, Southern Regional Research Center, New Orleans, LA 70179.

(Spectro/Computer) as described by Rosenthal et al (1976) was used to record reflectance spectra of the ground cottonseed samples. Experimental procedures similar to those described by Norris et al (1976) were used. The wavelength range was 1.00–2.38 μm. The slit width was set at 2.0 mm, giving a spectral bandpass of 0.008 μm. The computer program for data acquisition was based on sampling data at 300 points (0.0046 μm apart) for each reflectance curve.

Samples for recording spectra were held in Neotec sample holders containing 10 g of ground cottonseed. The linear signal from the detectors was digitized, sent to the computer, and then recorded on magnetic tape. A reflectance standard consisting of powdered Teflon was used to record a curve representing the nonabsorbing condition. Subsequently, the data were processed into a normalized derivative of reflectance; ie, the derivative of reflectance with respect to wavelength divided by the reflectance for the same wavelength. The computation of first derivative is the difference in reflectance between two adjacent wavelengths. In practice, the computation involves the averages for a group of data points covering a short wavelength region. The more data points, the smoother the resulting derivative curve; however, excessive averaging would result in loss of significant information. In this experiment, six data points were used for each wavelength band.

A program for linear regression analysis was used to compute a correlation between the chemical analysis of gossypol and the reflectance data at each wavelength. This defines the point on the wavelength scale where the lowest standard error for predicting gossypol content is obtained and thus establishes the first term in the regression equation. The process is repeated to add more terms to the regression equation for improving the correlation.

As the regression equation is developed by adding more terms, it is important to ascertain whether the improvement is due to chance or, with a specific level of confidence, the improvement is real. The program computes 300 correlations, one at each wavelength. If these correlations were computed with random numbers, the true correlation should be zero. Most of the computed correlation coefficients will not be zero but will be distributed around zero. The program selects the highest correlation, which may be much different from zero. Consequently, a reference point is necessary for evaluating the regression computations. Because r is limited to values between −1.0 and +1.0, normal curve statistics cannot be applied directly. If we make the transformation using the hyperbolic tangent $r' = \tanh^{-1} r$, then normal curve statistics can be applied (Morrison 1976). The standard deviation for the distribution of correlations is $S = 1/\sqrt{N-3}$, where N is the number of samples used in the experiment. If confidence limits are selected so that the total area in the tails of the distribution is equal to 1/300, then we can predict the value of the correlation that will be selected when the true correlation is zero. The normal variable, z, for this criterion is $z = 2.93$. For 30 samples, the standard deviation of r' is $S = 1/\sqrt{30-3} = 0.192$. Then the highest correlation for these conditions will be $r' = 2.93 \times 0.192 = 0.562$ and $r = \tanh 0.562 = 0.510$. This approach was used to test correlations for significant improvement at the 99% confidence level, for which the normal variable is 2.326. To have a 99% level of confidence that the true correlation is not zero, the correlation must exceed 0.785.

The program is designed so that additional terms are added based upon using the term that gives the greatest reduction in the standard error. The group of samples used in an experiment must represent the entire population of cottonseed. This requirement is difficult to satisfy, so the regression equation must be evaluated with an independent group of samples. In our work, the experiment was repeated to evaluate the equation. Terms that are not repeated in subsequent regression equations should not be considered a significant part of the equation without further evaluation.

RESULTS AND DISCUSSION

The regression analysis using first derivative processing resulted in a correlation r = 0.9674. The regression equation involved the computation of the difference between the first derivatives at two adjacent wavelengths. This produced an equation for computing the second derivative. Consequently, in all subsequent processing, second derivative was used. The choice of second derivative data processing agrees with the procedure followed by Norris et al (1976). The results are shown in Fig. 1. The correlation with the first term alone was 0.8988, which meets the requirement for significance. With three terms, the correlation was 0.9869, which was significantly better than the results using first derivative.

The results for three regression equations are listed in Table I. Comparison of the results of the two regression equations for the 1978 test shows that the wavelengths were repeated for the first two terms but not for the third. The standard error of the chemical analysis performed in this laboratory was 0.056, which accounts for about one half of the standard error for the regression computation. For both the 1978 and 1979 experiments, the second wavelength term improved the correlation significantly, but the improvement from adding the third wavelength term was not significant.

The results for the 1979 test are shown in Fig. 2. Even though that test did not use dehulled cottonseed, the results were very similar to the 1978 data. The small differences in wavelength were caused by a difference in programming for the wavelength scanning. The wavelengths for the first two terms, but not the third, were in agreement between the 1978 and 1979 tests.

Sampling is probably the major source of error for three reasons. The NIR and the chemical analysis are not necessarily determined on exactly the same material; ie, both measurements were

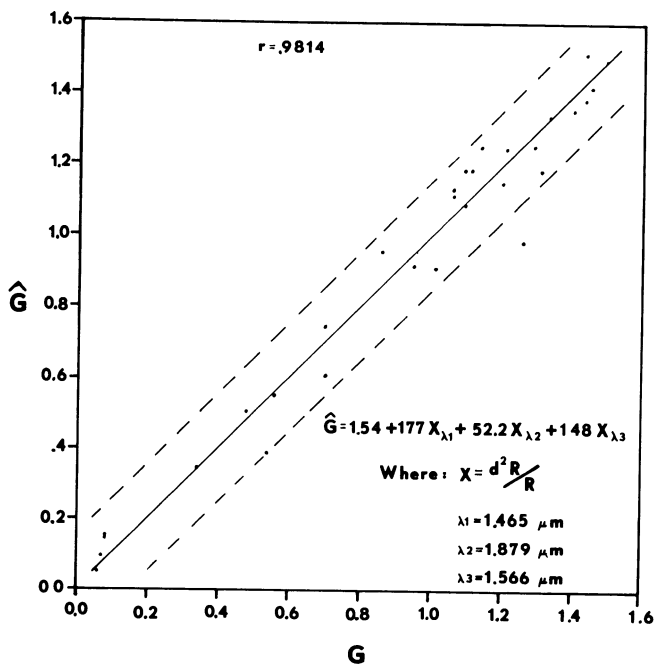

Fig. 1. Correlation between the gossypol content of cottonseed determined by chemical analysis (G) and the predicted content based on reflectance measurements ($\hat{G}$) for 1978 data. The standard error, not including the one outlying point (G = 1.26), was 0.0745% gossypol. The 30 samples were delinted, dehulled, and ground.

TABLE I
Results of Regression Analysis for Prediciting the Gossypol Content of Cottonseed

Test Year	Analysis Laboratory	Wavelength, μm 1	2	3	Correlation Coefficient	Standard Error
1978	BA[a]	1.469	1.883	1.530	0.987	0.074
1978	QE[b]	1.469	1.883	1.322	0.973	0.116
1979	PT[c]	1.464	1.880	2.004	0.973	0.078

[a] Barrow-Agee, Memphis, TN.
[b] Quality Evaluation Laboratory, Russell Research Center, Athens, GA.
[c] Pope Testing Laboratories, Dallas, TX.

determined on the basis of sampling from a larger volume of material. The area illuminated for the NIR measurement is about 3 mm wide and 10 mm long, so the volume of material actually involved in the reflectance measurement is a small fraction of the material in the sample holder. The gossypol is contained in discrete packages (glands) in the seed, and these glands must be distributed uniformly through the ground material to produce a reliable measurement.

For low gossypol levels, the distribution of the pigment glands is likely to become a serious problem. If the cottonseed were treated to release the gossypol from the glands so that the pigment becomes distributed throughout the sample, then even though the concentration may be very low the gossypol would still be highly detectable.

The reflectance (log 1/R) of extracted gossypol is shown in Fig. 3. The first wavelength (1.46 μm) selected in the regression computations corresponds with an absorption maximum for gossypol, which verifies the significance of that wavelength. The additional curve for water shows that measurements at 1.46 μm are subject to error because of differences in moisture content of the samples.

The regression equations (Figs. 1 and 2) involved the addition of the derivatives for the first two wavelength-dependent terms; however, the second derivative at 1.46 μm becomes more negative as either gossypol or water content increases, whereas the second derivative at 1.880 μm becomes more positive as the water content increases (Fig. 3). This results in cancelling the effects of differences in water content between samples. Another concern is that measurements in spectral regions corresponding with water absorption can introduce a temperature dependence (Collins 1925).

Specifications for a filter to be used in a tilting-filter instrument were defined by computing several regression equations with 20 samples selected for a range of gossypol content. These computations were used to determine the optimum bandpass of the monochromator, a more precisely defined wavelength for measuring gossypol, and the optimum data processing. This experiment was conducted with a recently developed computer program in which the initial data processing computes log 1/R.

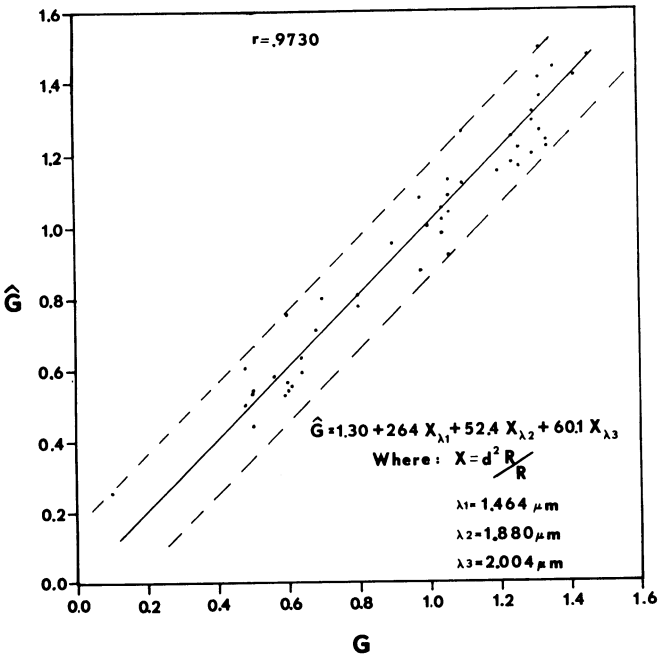

Fig. 2. Correlation, using 1979 data, between gossypol content (G) and the predicted content based on reflectance measurements ($\hat{G}$). Forty samples of cottonseed were delinted and ground. The standard error was 0.0777% gossypol.

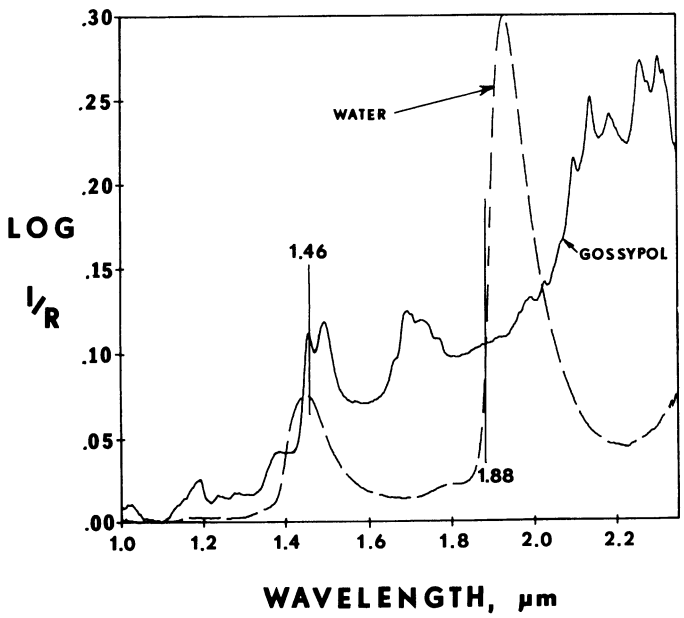

Fig. 3. Reflectance (log 1/R) spectra of extracted gossypol. The marked wavelengths identify the wavelength of the first two wavelength-dependent terms of the regression equation. The water was measured with transmittance geometry using a thickness of 0.0026 cm.

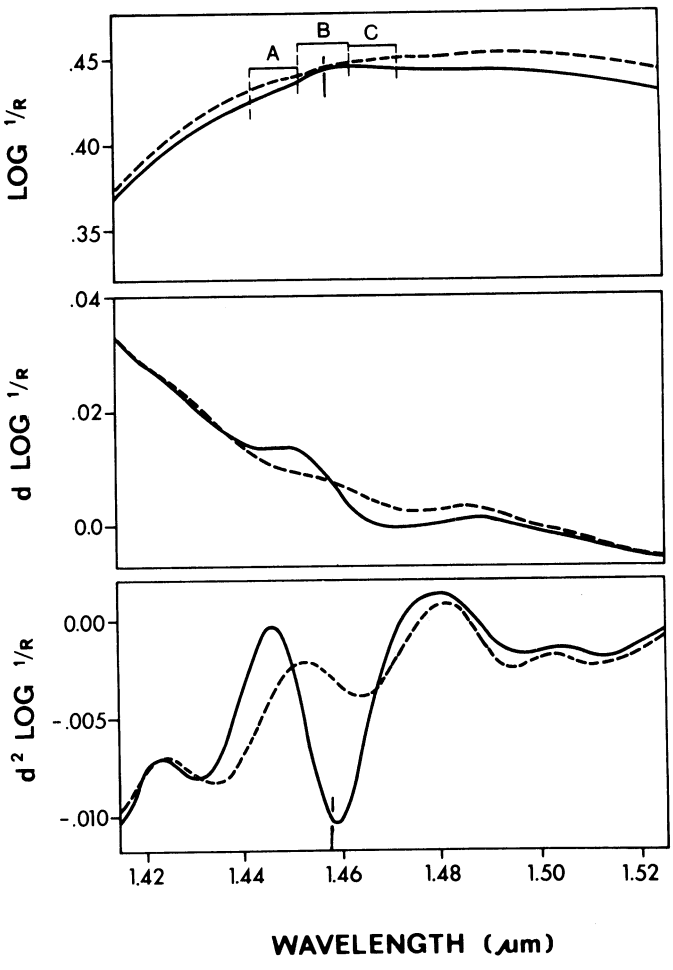

Fig. 4. Examples of the primary procedures for processing near-infrared reflectance data: log 1/R, first derivative and second derivative. The first derivative at a specific wavelength is equal to B − A and second derivative is A + C − 2B where each (A, B, or C) is the average of all the data recorded in a 0.01-μm wavelength range within the context shown at the top of the figure. The dashed line is for a sample having 0.06% gossypol, and the solid line is for a sample having 1.49% gossypol.

Data files having as many as 1,000 numbers could be accumulated. Slit widths of 1.0, 2.0, and 3.0 mm were used. Data were obtained at 0.0007-μm intervals. For data processing, the second derivative was computed using data averaging, starting with nine data points and increasing in increments of six to 63 data points.

The results showed that the optimum conditions for measuring gossypol were: slit width, 2.0 mm; wavelength, 1.457 μm; and for data processing, 45 points to compute second derivative. This corresponds to an overall wavelength range of 0.0315 μm for computing second derivative, which can be rounded to 0.03 μm. With the optimum conditions, a correlation of 0.9720 and standard error of 0.09 were computed with one wavelength term. Examples of derivative data are illustrated in Fig. 4 for two samples of cottonseed: one with essentially no gossypol and one with 1.4% gossypol. The data were recorded as log 1/R. Examples are shown for the two samples at the top of the figure. For computing second derivative, all of the data recorded in each of three 0.01-μm wavelength bands were averaged. The horizontal lines in Fig. 4 define these bands, and the quantities A, B, and C represent the averages. Two averages were used to compute first derivative: $d \log 1/R = B - A$. The second derivative used all three averages: $d^2 \log 1/R = A + C - 2B$. The result is identified with the center wavelength of the entire wavelength range of the data used for computation. Subsequently, the process was repeated after advancing one data point, and this continued until the entire data set for one sample was processed. The more data used to compute the average, the smoother the derivative spectra. The results will be valid if the wavelength constraints defined above are maintained. The justification for using second derivative is that better results are obtained as judged by either the correlation coefficient or the standard error.

The Spectro/Computer as used in this experiment is a research instrument and was used to develop the equation relating reflectance of ground cottonseed to gossypol concentration. For the routine measurements of many samples an instrument is used in which the required wavelength isolation is obtained with filters instead of a monochromator. The bandpass of such an instrument should be 0.01 μm or less; ie, it should be equivalent to the wavelength range covered by one data group, as shown in Fig. 4. Several such instruments are on the market. Generally the filter instruments are independently calibrated with several samples and acquire the data in much less time than the Spectro/Computer, so data averaging improved sampling procedures, and other techniques can be used more extensively to reduce errors. Therefore, the results given should not be interpreted as the limit in the precision of the measurement. For determining the gossypol content at very low concentrations, as is required for regulatory purposes, modification of the procedures may be required to provide the necessary sensitivity.

LITERATURE CITED

ABOU-DONIA, M. B., LYMAN, C. M., and DIECKERT, J. W. 1969. Metabolic fate of gossypol: The metabolism of ^{14}C-gossypol in rats. Lipids 5(11):938.

AMERICAN OIL CHEMISTS' SOCIETY. 1978. Official and tentative Methods of the AOCS. 3rd ed. Official Method Ba 7-58, revised 1969. The Association, Champaign, IL.

BIRTH, G. S. 1976. How light interacts with foods. Page 6 in: Quality Detection in Foods. J. J. Gaffney, ed. Am. Soc. Agric. Eng., St. Joseph, MI.

COLLINS, J. R. 1925. Change in the infrared absorption spectrum of water with temperature. Phys. Rev. 26:771.

FEDERAL REGISTER. 1974. Modified cottonseed products. 39(177):32735.

GARDNER, H. K., JR., HRON, R. J., SR., and VIX, H. L. E. 1976. Removal of pigment glands (gossypol) from cottonseed. Cereal Chem. 53(4):549.

HOFFPAUIR, C. L., and PONS, W. A., JR. 1953. Review of the properties of gossypol and methods of its estimation. J. Assoc. Off. Agric. Chem. 36:1108.

HYMOWITZ, T., DUDLEY, J. W., COLLINS, F. I., and BROWN, C. M. 1974. Estimations of protein and oil concentration in corn, soybean, and oat seed by near-infrared light reflectance. Crop. Sci. 14:713.

MARTINEZ, W. H. 1977. Other antinutritional factors of practical importance. Page 304 in: Evaluation of Proteins for Humans. C. E. Bodwell, ed. Avi Publ. Co., Westport, CT.

MORRISON, D. F. 1976. Multivariate Statistical Methods. 2nd ed. McGraw-Hill, New York.

NORRIS, K. H., BARNES, R. F., MOORE, J. E., and SHENK, J. S. 1976. Predicting forage quality by infrared reflectance spectroscopy. J. Anim. Sci. 43:889.

O'HAVER, T. C., and GREEN, G. L. 1975. Derivative spectroscopy. Am. Lab. 7(3):15.

ROSENTHAL, D., HAYDEN, R., and ENGLER, S. 1976. A highly flexible 9.0 OD light transmittance/reflectance research spectrophotometer. Page 12 in: Quality Detection in Foods. J. J. Gaffney, ed. Am. Soc. Agric. Eng., St. Joseph, MI.

TKACHUK, R. 1981. Protein analysis of whole wheat kernels by near infrared reflectance. Cereal Foods World 26(10):584.

[Received June 22, 1981. Accepted May 15, 1982]