Beiträge zur Tabakforschung International · Volume 11 · No. 4 · August 1982 DOI: 10.2478/cttr-2013-0516

Rapid Spectrophotometric Analysis of the Chemical Composition of Tobacco

Part 3: Polyphenols *

by

W. F. McClure

Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, North Carolina, U.S.A.

and

R. E. Williamson

Tobacco Research Laboratory, Agricultural Research Service, United States Department of Agriculture, Oxford, North Carolina, U.S.A.

INTRODUCTION

Tobacco breeders are interested in developing varieties with desirable smoke characteristics suitable for the production of cigarettes. In future breeding programs, tobaccos will have to be analyzed and routinely screened for more chemical constituents than at present. Near-infrared spectrophotometry (NIR) can be used to increase the rapidity with which the level of chemical constituents in tobacco can be determined (1, 3, 5). It offers the advantage of allowing several chemical constituents to be determined simultaneously, whereas most automated colorimetric systems can only be used to measure a few chemical constituents at the same time with the same set-up.

The need for a rapid method for determining polyphenols in tobacco is essential as these are important indicators of leaf composition which may be precursors of biologically active smoke constituents, such as phenols and related compounds. The major polyphenols in tobacco are chlorogenic acid and rutin (7) with chlorogenic acid being 75 %–95 % of the total. *Zane* and *Wender* (9) have shown that catechol is a pyrolytic product of chlorogenic acid. This study was conducted to determine the potential of computerized NIR spectroscopy for determining total polyphenols in tobacco.

* Paper No. 6614 of the Journal Series of the North Carolina Agricultural Experiment Station, Raleigh, N.C. Mention of a trademark, proprietary product or vendor does not constitute a guarantee or warranty of the product by the N.C. Agricultural Experiment Station or the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products or vendors that may also be suitable.

Received: 17th October 1980 – accepted: 14th December 1981.

MATERIALS AND METHODS

Instrumentation

The spectral data were recorded and analyzed on the computerized spectrophotometric system (COMP/SPEC) in the Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, North Carolina (4). A simplified schematic of this facility is shown in Figure 1. The system, custom designed and constructed in-house, incorporates a Cary 17 monochromator with a 31 × 23 × 23 cm^3 sample compartment and is controlled by a Nova 2/10 minicomputer with 64 K bytes of core memory. Recorded spectra may be displayed on the Tektronix Model 611 storage dis-

Figure 1. Block diagram of computerized spectrophotometer for reflectance measurements.

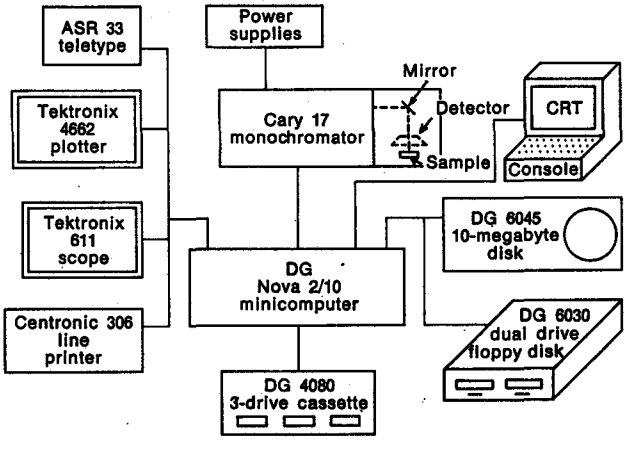


Table 1. Total polyphenol content by Williamson's method (8) of samples in data files.

File No.	Number of samples	Range (%)			Mid-range	Standard deviation	Year grown	Date of chemical analysis	Date of near-infra-red scan
		min.	max.	mean					
1	59	0.050	7.200	3.435	3.625	1.804	1975	1976	1979
2	58	0.180	6.950	3.434	3.565	1.760	1975	1976	1979
3	62	2.000	7.500	5.050	4.750	1.835	1976	1976	1979
4	53	1.270	6.600	3.940	3.935	1.741	1976	1978	1979
5 *	173	0.050	7.500	3.644	4.000	1.855			

* Combined files 2, 3, and 4.

play monitor; hard copy spectra are recorded on a Tektronix 4662 plotter. Computer peripherals consist of: [A] a Data General Dasher console and a Model ASR 33 teletype, [B] magnetic media consisting of Data General's Model 6045 10-megabyte disk, a Model 6030 two-drive floppy disk and a Model 4080 three-drive cassette unit, and [C] a Centronic Model 306 line printer.

Samples and Chemical Analysis

The samples were collected over a period of four years from 1973 through 1976. Seventeen cultivars of flue-cured tobacco from eight locations in Virginia, North Carolina, South Carolina and Georgia were used in these investigations. Leaves were selected from three stalk positions: bottom, middle, and top. After removing the midribs, the lamina was dried to 2%–3% moisture and ground in a Wiley mill with a 1 mm mesh screen. The total polyphenol contents of 238 samples were determined with a Technicon AutoAnalyzer by the calorimetric method of *Williamson* (8). In addition, 65 samples (35 from 1975 and 30 from 1976) were analyzed a second time at a later date to determine the standard error of *Williamson*'s method.

Table 1 shows the statistical distribution of polyphenols within each file. Files 1 and 2 were from tobacco grown in 1975, analyzed chemically for polyphenols in 1976, and scanned with the COMP/SPEC in 1979. Tobacco used to establish file 3 was grown in 1976, analyzed chemically the same year, and scanned in 1979. File 4 was made up of tobacco grown in 1976, analyzed chemically in 1978, and scanned in 1979. File 5 was a combined file consisting of NIR scan data of the 173 samples from files 2, 3, and 4.

In addition to the total polyphenols in tobacco, chlorogenic acid, rutin and caffeic acid (obtained from Sigma Chemical Co.) were scanned to show the relationships of the spectra of pure polyphenols and tobacco at wavelengths selected by the stepwise multiple linear regression (SMLR) model for predicting levels of total polyphenols.

Near-Infrared Spectrophotometry

Two grams of dry, ground tobacco, as used for wet chemistry analysis, were placed in a Technicon solid-sample holder for scanning on the computerized NIR spectrophotometer. Each sample was scanned in the reflectance (R) mode and the spectrum was stored on floppy disks. Each reflectance spectrum was made up of 1700 discrete data points, the points spaced one nanometer apart along the wavelength axis from 0.9 to 2.6 micrometers. Furthermore, each data point was the

Table 2. Preparation of spectral data for statistical analyses.*

	Smoothing function		
	$\frac{\log(1/R)}{21\ \text{MPA}^{**}}$	$\frac{dR/R\,d\lambda}{21\ \text{MPA}}$	$\frac{d^2(\log(1/R))/d\lambda^2}{21\ \text{MPA}}$
1st derivative	–	21 PS+	–
2nd derivative	–	–	21 PS
Shrink factor	1/2	1/2	1/2
Points per spectrum	840	830	830

* Statistical analyses consisted of development of the multiple linear regression prediction model and predicting unknown samples.

** MPA = moving-point average smoothing.

+ PS = a derivative technique achieved by truncation of the Taylor Series expansion of the spectral function.

average of 2000 15-bit analog-to-digital conversions, a technique employed to reduce random electronic noise. Three alternate representations, $\log(1/R)$, $dR/R\,d\lambda$, and $d^2(\log(1/R))/d\lambda^2$, were computed for each spectrum and treated as shown in Table 2. All three representations were computed from the 21-point moving-point-average (MPA) smoothed reflectance spectra (1). The first and second derivative spectra were computed by means of a truncated Taylor Series representation of the spectral data function (6).

A stepwise multiple linear regression program was developed that could use up to 2000 independent variables in a memory-limited environment (2). This program was used to develop the relationship between the polyphenol content of the samples and the optical parameters. The program was designed so that at each step in the regression process, the variable with the highest partial correlation will enter the regression equation. Before a variable is accepted, the variables already in the equation are tested and the variable which causes an insignificant increase in the error sum of squares is deleted from the regression. The stepwise process is terminated when [a] there is no variable left which causes a statistically significant improvement in the regression equation, [b] the variable entering the equation is the one just eliminated, or [c] when the desired number of steps has been executed. An equation of the following form is produced:

$$P = \alpha + \beta_1 V_1 + \beta_2 V_2 + \ldots + \beta_n V_n$$

where

P = polyphenol content in percent,

α = intercept,

V_n = optical variable at wavelength n, and

β_n = regression coefficient for V_n.

Files 1 and 2 (Table 1) were obtained from a single file of 117 samples. The odd and even samples were split alternately into file 1 and file 2, respectively. Calibration of the system was based on the file 1 and performance of the COMP/SPEC was based on the use of this calibration to predict the other four files, including combined file 5.

RESULTS AND DISCUSSION

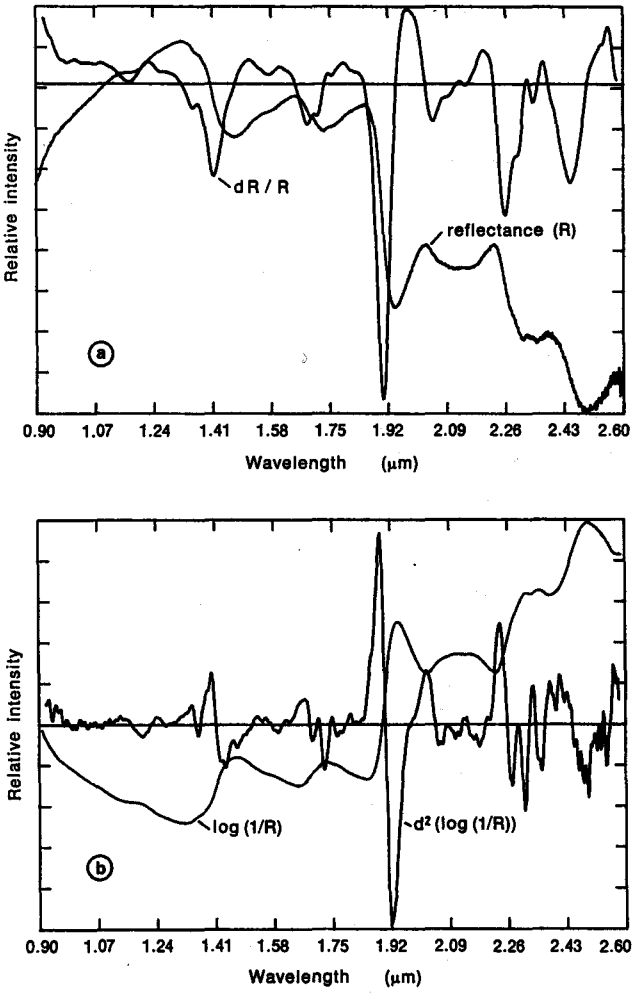
Polyphenol Content of Samples

Table 1 shows the distribution of the polyphenols for each file. This information is essential in evaluating the NIR method. For example, neither the standard error of calibration (SE_c) nor the coefficient of determination (r^2) can be the only criterion of performance of the multiple linear regression for predicting levels of chemical constituents. A lower case r is used here to be distinguished from the reflectance R. Since the NIR method is a secondary method (i.e. it is calibrated against chemical methods of analysis) a high r^2 and a low SE_c may be achieved by calibrating the NIR method for a given constituent against a set of samples having a narrow range of chemical composition. It is always advisable to calibrate with a set of samples that encompasses, and is distributed across, the range of concentrations expected in the unknowns.

Spectral Characteristics

Figure 2 shows a plot of reflectance R, $\log(1/R)$, $dR/R\,d\lambda$, and $d^2(\log(1/R))/d\lambda^2$ which are proportional to the solute in a non-scattering absorbing medium. Therefore they were all considered in this particular study. The log (1/R) spectra, with absorption shown as peaks or shoulders, have few definitive characteristics. Water, with absorption peaks at 1.45 μm and 1.94 μm, is the best documented of all the absorbers in tobacco. Enhancement of the R data by computation of $dR/R\,d\lambda$ and $d^2(\log(1/R))/d\lambda^2$ reveals the presence of subtle differences which are not visually perceptible in the R curves. Note that the absorption bands appear as positive-going zero crossings in the $dR/R\,d\lambda$ spectrum and as negative peaks in the $d^2(\log(1/R))/d\lambda^2$. The $d^2(\log(1/R))/d\lambda^2$ spectrum

Figure 2. Reflectance R, log(1/R), $dR/R\,d\lambda$, and $d^2(\log(1/R))/d\lambda^2$ spectra of a tobacco sample.



shows major absorbers at 1.21, 1.44, 1.725, 1.91, 2.26, 2.298, 2.345, and 2.476 μm. In addition, there are a number of other less prominent peaks throughout the spectrum.

The most precise mathematical representation investigated for predicting levels of polyphenols was $d^2(\log(1/R))/d\lambda^2$; $dR/R\,d\lambda$ was reasonably accurate; and log (1/R) was less precise than the others. For file 1, a 6-step equation (data from 6 wavelengths), calibration equations based on optical parameters log (1/R), $dR/R\,d\lambda$, and $d^2(\log(1/R))/d\lambda^2$ gave r^2 values of 0.902, 0.936, and 0.959, respectively. Subsequent data in this study are for the optical parameter $d^2(\log(1/R))/d\lambda^2$.

Calibration and Analysis

As the number of steps used in the prediction equation increased from 1 to 15, the standard error of prediction (SE_p) decreased up to the sixth step, then gradually increased. Six to 9 steps in the prediction equation were about equally accurate. For several other chemical constituents in tobacco with which we have worked, six steps in the prediction equation have been adequate. Calibration and prediction data, based on the optical parameter $d^2(\log(1/R))/d\lambda^2$, are shown in Table 3.

Figure 3a gives the calibration plot for the NIR COMP/SPEC based on file 1. The multiple linear regression equation of the line in Figure 3a is:

$$P = -2.731 + 275.877\,D^2_{2.220} + 237.165\,D^2_{2.440} - 435.536\,D^2_{2.156} - 456.954\,D^2_{1.742} + 197.555\,D^2_{2.346} + 180.790\,D^2_{2.210}$$

where

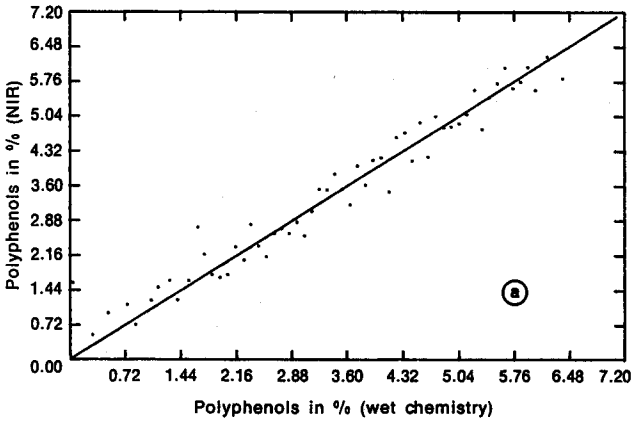
P = polyphenols in %,

D^2_i = 2nd derivative parameter,

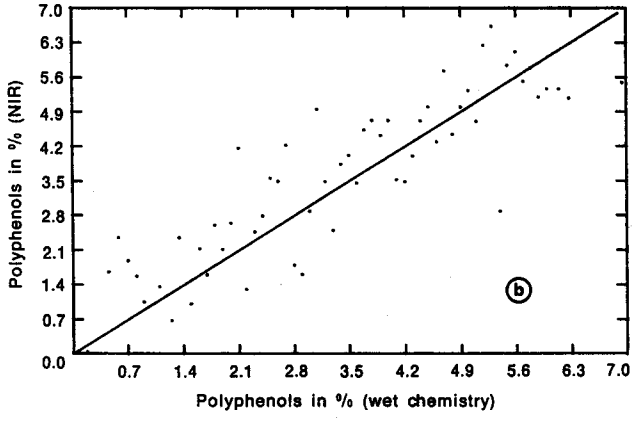
i = wavelength in micrometers at which the parameter is measured.

Figure 3. Relationship between the chemically determined total polyphenols and the values predicted by infrared reflectance data.

a) Calibration plot.



b) Prediction plot from a second group of samples.



The coefficient of determination r^2 was 0.959 and the standard error of calibration was ± 0.379 %.

Chemical methods are usually calibrated daily by analyzing samples with known concentrations at regular intervals. It is unlikely that the COMP/SPEC would be used without routine checks on calibration. The pre-

Table 3. Calibration and prediction data for total polyphenols using a six-step equation. *

File No.	Number of Samples	Calibration			Prediction	
		r^2	SE_c	CV_c	SE_p	CV_p
1 **	59	0.959	0.378	11.0	–	–
2	58				0.747	21.75
3	62				1.353	26.79
4	53				0.885	22.46
5	173				1.022	30.04

* r^2, coefficient of determination; SE_c, standard error of calibration; CV_c, coefficient of variation for calibration; SE_p, standard error of prediction; CV_p, coefficient of variation for prediction.

** Calibration equation:

$$P = -2.731 + 275.877\,D^2_{2.220} + 237.165\,D^2_{2.440} - 435.536\,D^2_{2.156} - 456.954\,D^2_{1.742} + 197.555\,D^2_{2.346} + 180.790\,D^2_{2.210}$$

diction plot of the COMP/SPEC for file 2 is shown in Figure 3b. In looking at this figure, remember that these are data based on 58 even-numbered samples (file 2, Table 1) from the same set of samples from which the calibration samples (file 1, Table 1) were taken. Here, the standard error of prediction SE_p was ± 0.747% with a coefficient of variation of 21.75%. The chemical method, against which the COMP/SPEC was calibrated, had a SE_p of ± 0.70% with a CV_p of 23.0% when computed the same way the NIR variance was computed.* Thus, it appears that the variance of the NIR method approached that of the wet chemical method.

If the set of calibration samples were large enough and highly representative of the population of the unknowns to be predicted, it is likely that the calibration could be used for any set of similar samples over the years without recalibration. Only instrument response checks would need to be made. Although file 1 does not represent the population of unknowns, it is interesting to see in Figure 4 how the system performed in predicting files from different years.

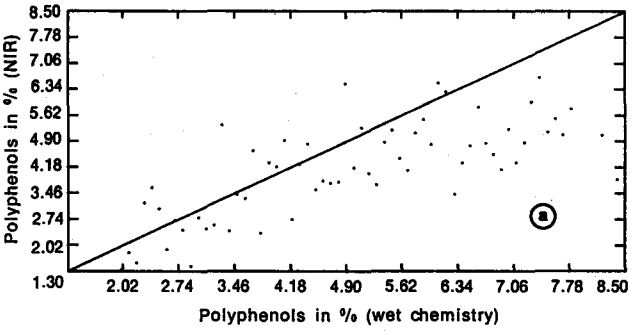
Figure 4a gives the prediction on samples from file 3, grown in 1976. SE_p was ± 1.353% and CV_p was 26.79%. Obviously the system underestimated the concentration of polyphenols in many of the samples. Figure 4b shows the performance of the NIR COMP/SPEC for predicting file 4 (Table 1). With SE_p of ± 0.885% and CV_p of 22.46%, prediction of polyphenols in these samples approaches the calibration results. The polyphenols concentrations in some of these samples were overestimated. Figure 4c gives the prediction results on the three combined files 2, 3 and 4. SE_p was ± 1.022% and CV_p was 30.04%.

Figure 5 shows the spectra of pure polyphenols. Figure 5a is the second derivative spectrum of chlorogenic acid. Tobacco chemists generally consider chlorogenic acid to be the predominant polyphenol in tobacco, making up 75% to 95% of the total polyphenols. The vertical lines in this figure mark the wavelengths selected and the numbers indicate the order in which the wavelengths were selected by the regression model. Wavelengths 1 and 6 correspond to a two-peak absorption band (negative peaks in the second derivative spectra) at 2.210 µm and 2.220 µm, respectively. Wavelength 2 corresponds to a very strong absorption at 2.440 µm. On the other hand, in Figure 5b (caffeic acid), the wavelengths 1 and 6 fall very close to a strong single absorber, while wavelength 2 lies near a very weak absorber. In Figure 5c, the derivative spectrum of rutin, neither the pair of wavelengths 1 and 6 nor the wavelength 2 falls at absorption bands of rutin. It is surprising that rutin has no absorption bands at wavelengths 1 and 2 while caffeic acid does, especially since rutin, in many cases, constitutes approximately 5% to 25% of the total polyphenols in tobacco. Wavelengths 3, 4 and 5 appear to have less evidence to support their selection.

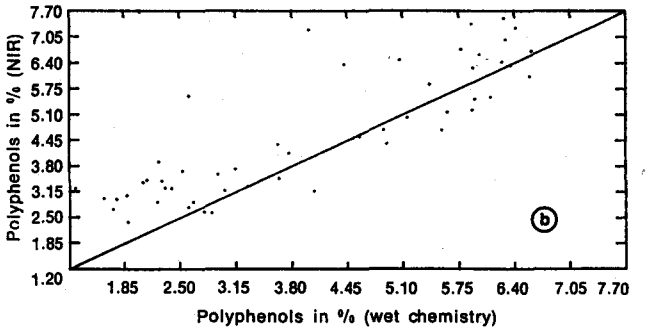
The basis for the selection of the first wavelength in the multiple linear regression model is based primarily on the correlation of spectral data with polyphenols. For

Figure 4. Predicting total polyphenols from independent data. The line represents the calibration curve from samples in file 1.

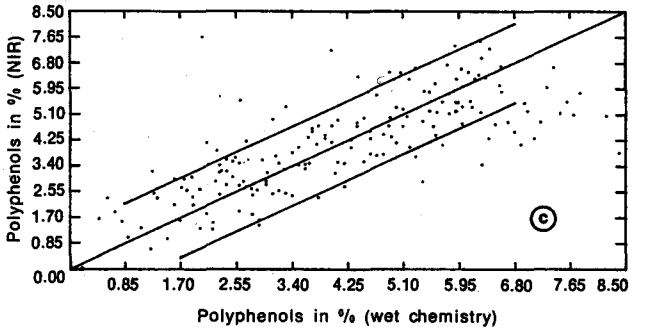
a) Data points of the predicted values for file 3.



b) Data points of the predicted values for file 4.



c) Data points of the predicted values for file 5.



---

*

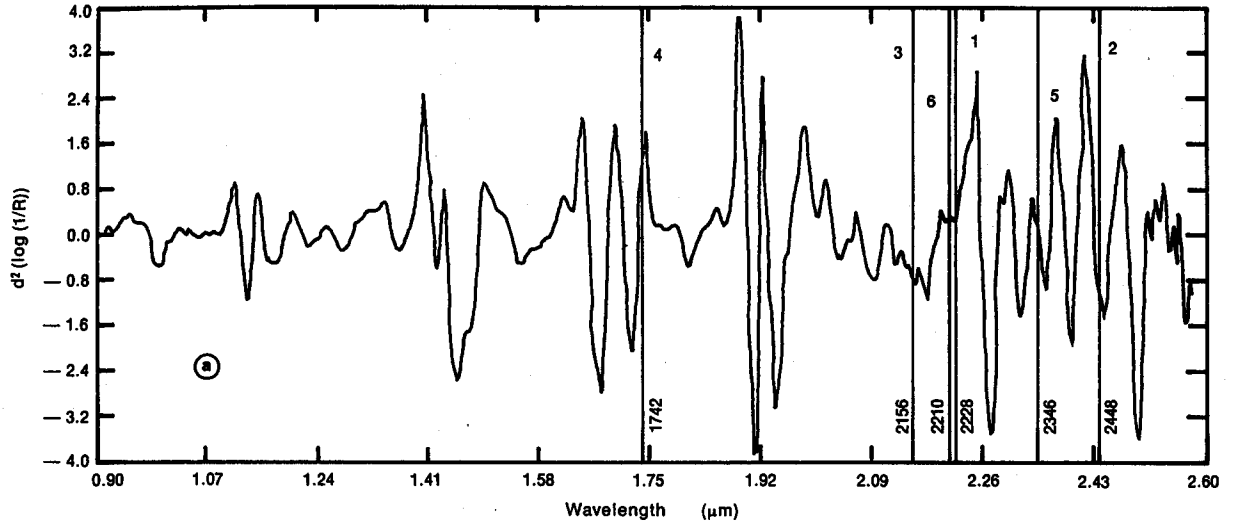
$$SE_p = \left(\frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} (x_{ib} - x_{ia})^2 \right) \right)^{1/2} ,$$

$$CV = \frac{SE_p}{\frac{1}{n} \cdot \sum_{i=1}^{n} x_{ia}} \cdot 100 ,$$

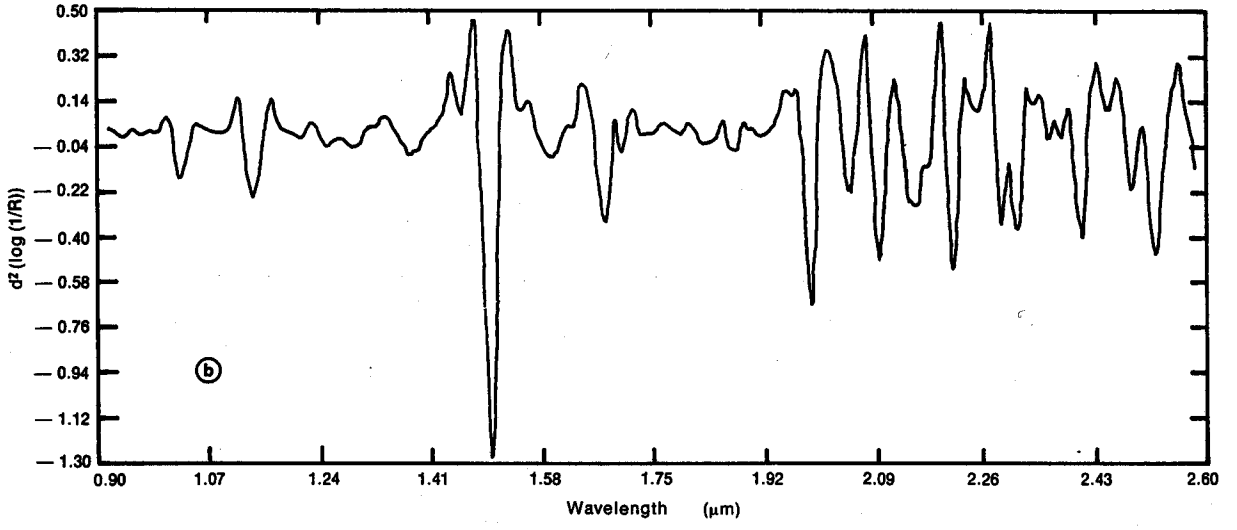
where x_{ib} = the known concentration from a previous analysis, x_{ia} = the new determination, and n = number of samples.

Figure 5. $d^2(\log(1/R)) / d\lambda^2$ **spectra of three polyphenols.**

a) Chlorogenic acid (the numbers on the vertical lines indicate the order in which the wavelengths were chosen by the SMLR program for predicting the total polyphenol content of tobacco).



b) Caffeic acid.



c) Rutin.

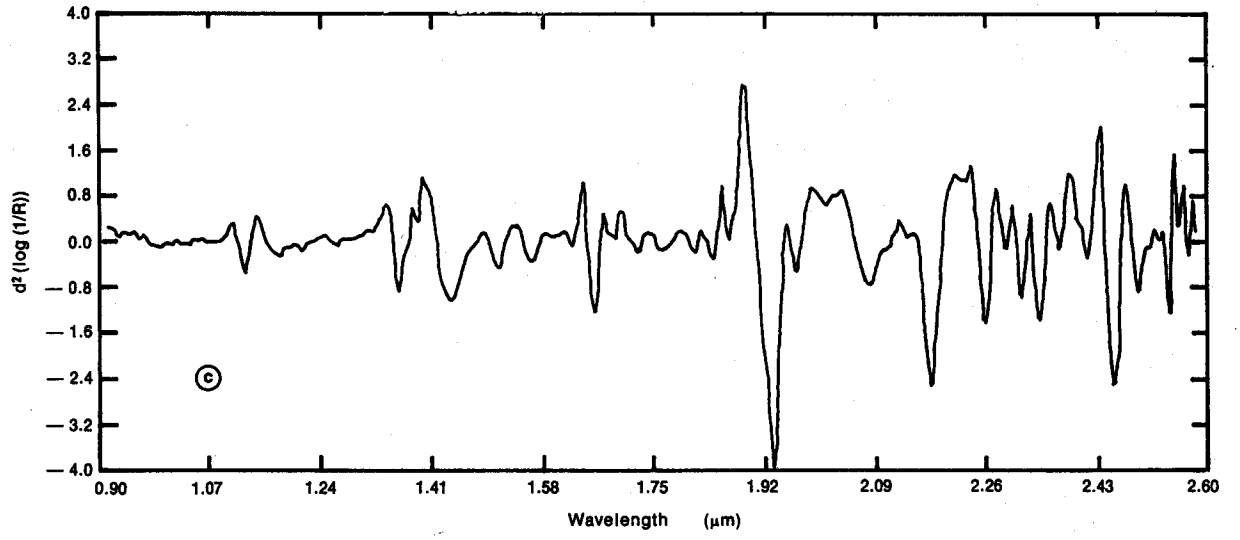
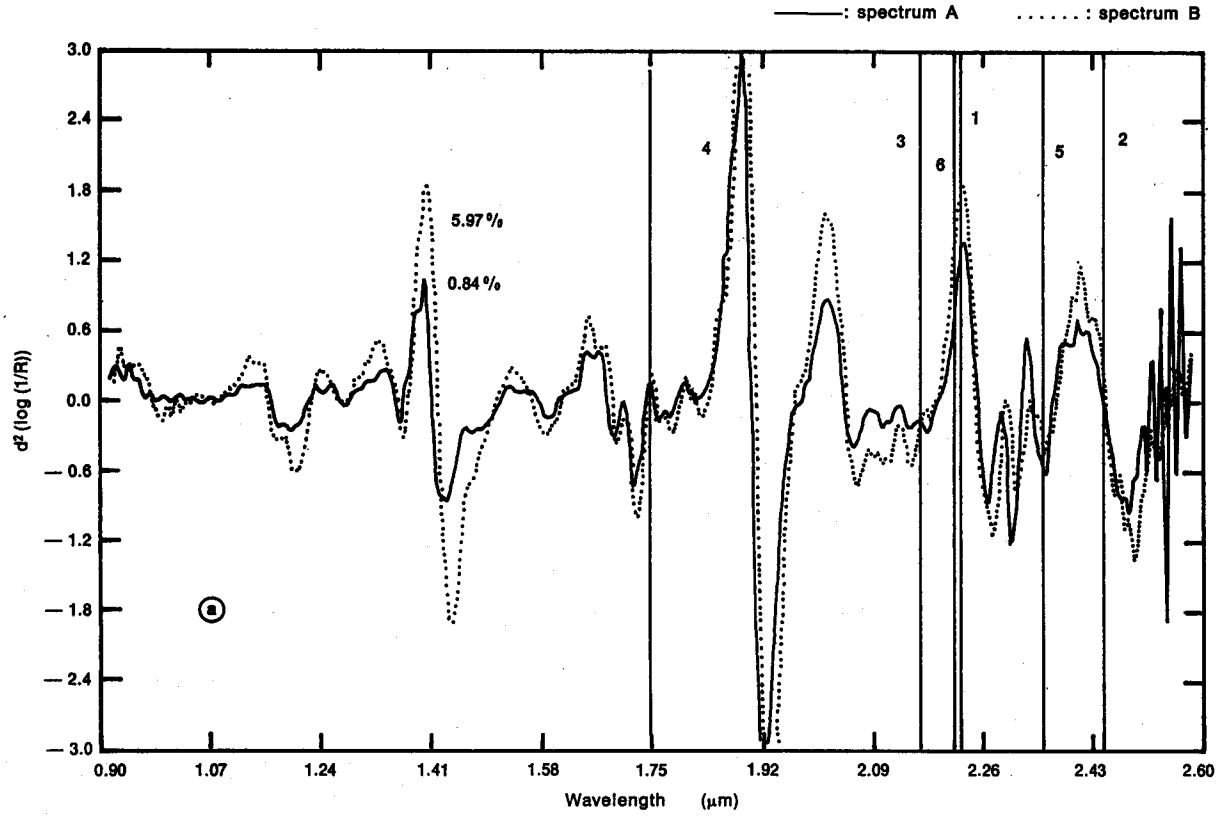
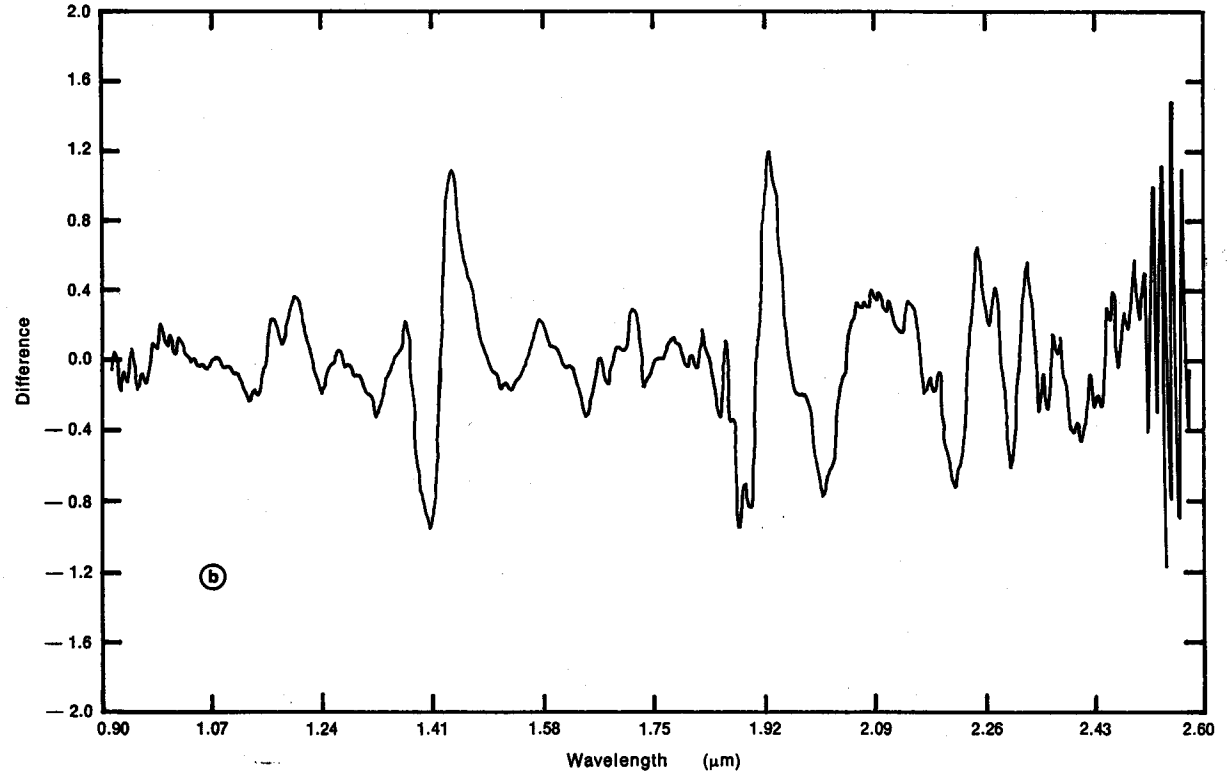


Figure 6. Average $d^2(\log(1/R)) / d\lambda^2$ spectra of ten tobacco samples each at two levels of total polyphenols.

a) Spectra for two levels of total polyphenols (vertical lines indicate the same as in Fig. 5).



b) Difference spectrum of spectra in Fig. 6a.



example, when the moisture content is highly correlated to the total polyphenol level, the moisture peak may be picked first. It is quite likely, since the algorithm makes intercorrelation checks, that this first choice will be deleted when other wavelengths are added. If it is not, a further test of the first choice can be made by programming the stepwise multiple linear regression model to select the first wavelength again, taking into consideration that the other 5 wavelengths have been added to the model. The first wavelength, 2.220 μm, was subjected to both of the above tests and in both cases the wavelength 2.220 μm was again included in the model. Thus, the model, and especially the first wavelength of the model, was considered to be valid.

Figure 6a shows two average spectra. Spectrum A (solid line) is the average of 10 spectra of tobacco with an average polyphenol content of 0.84%. Spectrum B (dotted line) is the average of 10 spectra; the average polyphenol content of the 10 samples used to produce the 10 spectra was 5.97%. Only a cursory scan of the spectra shows that the two spectra have significant differences. Figure 6b is the difference spectrum obtained by subtracting data for spectrum B in Figure 6a from spectrum A. It is interesting to note that wavelength 1 falls at a point where there is a large difference. Thus, the first wavelength in the model is further substantiated.

SUMMARY

Two hundred thirty-eight ground samples of tobacco were scanned with a computerized near-infrared (NIR) spectrophotometer to study the relationship of NIR spectra to the polyphenol content of the samples. A multiple linear regression model was used to select the most appropriate wavelengths for making the measurements.

The equation

$$P = -2.731 + 275.877\,D^2_{2.220} + 237.165\,D^2_{2.440} - 435.536\,D^2_{2.156} - 456.954\,D^2_{1.742} + 197.555\,D^2_{2.346} + 180.790\,D^2_{2.210}$$

is a valid equation for predicting polyphenols in tobacco. If the coefficients are validated on the same kind of tobacco and in the same year, the standard error of prediction of the NIR method (± 0.747%) approaches that of the wet chemistry method (± 0.70%) with coefficients of variation of 21.75% and 23.0%, respectively. The largest standard error of prediction across years was ± 1.353%.

ZUSAMMENFASSUNG

Zur Aufklärung des Zusammenhanges zwischen dem Polyphenolgehalt des Tabaks und dessen Infrarotspektrum untersuchten die Autoren 238 Proben gemahlenen Tabaks unter Verwendung eines computergestützten Spektrophotometers für den nahen Infrarotbereich (NIR). Die für die Messung am besten geeigneten Wellenlängen wurden mit Hilfe eines multiplen linearen Regressionsmodells ermittelt.

Durch die Gleichung

$$P = -2{,}731 + 275{,}877\,D^2_{2,220} + 237{,}165\,D^2_{2,440} - 435{,}536\,D^2_{2,156} - 456{,}954\,D^2_{1,742} + 197{,}555\,D^2_{2,346} + 180{,}790\,D^2_{2,210}$$

wird der Polyphenolgehalt des Tabaks gut vorhergesagt. Wenn der Bestimmung der Regressionskoeffizienten dieselbe Tabaksorte und dasselbe Erntejahr zugrunde liegen, nähert sich der Standardvoraussagefehler der NIR-Methode (± 0,747 %) dem der naßchemischen Methode (± 0,70 %); die entsprechenden Variationskoeffizienten beliefen sich auf 21,75 % bzw. 23,0 %. Der größte Vorhersagefehler über die Jahre hinweg betrug ± 1,353 %.

RÉSUMÉ

Deux cent trente-huit échantillons de tabac en poudre ont été examinés au moyen d'un spectrophotomètre dans le proche infrarouge (NIR) relié à un calculateur, afin d'établir la relation entre le spectre NIR et la teneur en polyphénols des échantillons. On a utilisé un modèle de régression linéaire multiple pour choisir les longueurs d'onde les mieux adaptées aux mesures.

L'équation

$$P = -2{,}731 + 275{,}877\,D^2_{2,220} + 237{,}165\,D^2_{2,440} - 435{,}536\,D^2_{2,156} - 456{,}954\,D^2_{1,742} + 197{,}555\,D^2_{2,346} + 180{,}790\,D^2_{2,210}$$

est une formule adéquate pour la prédiction des polyphénols dans le tabac. Si les coefficients sont établis pour un même type de tabac et une même année de récolte, l'erreur-standard de prévision (écart-type) obtenu par la méthode NIR (± 0,747 %) est voisine de celle qu'on observe par les méthodes de la chimie analytique classique (± 0,70 %) avec des coefficients de variation respectifs de 21,75 % et 23,0 %. L'erreur-standard maximale de prédiction au cours des années successives atteint ± 1,353 %.

REFERENCES

1. Hamid, A., W. F. McClure and W. W. Weeks: Rapid spectrophotometric analysis of the chemical composition of tobacco, Part 2: Total alkaloids; Beitr. Tabakforsch. Int. 9 (1978) 267–274.
2. Hamid, A., W. F. McClure and T. B. Whitaker: NIR analysis of food products, Part II: Stepwise linear regression in memory-limited environments; Amer. Lab. 13 (1981) 108–121.

3. McClure, W. F., K. H. Norris and W. W. Weeks: Rapid spectrophotometric analysis of the chemical composition of tobacco, Part 1: Total reducing sugars; Beitr. Tabakforsch. 9 (1977) 13–18.
4. McClure, W. F., and A. Hamid: Rapid NIR measurement of the chemical composition of foods and food products, Part 1: Hardware; Amer. Lab. 12 (1980) 57–69.
5. Pandeya, R. S., N. Rosa, F. H. White and J. M. Elliot: Rapid estimation of some flue-cured tobacco chemical characteristics by infrared-reflectance spectroscopy; Tob. Sci. 22 (1978) 27–31.
6. Sakolnikoff, I. S.: Advanced calculus; McGraw-Hill Book Company, Inc., New York, p. 206ff., 1939.
7. Tso, T. C.: Physiology and biochemistry of tobacco plants; Dowden, Hutchinson & Ross, Inc., Stroudsburg, Pa., p. 259ff., 1972.
8. Williamson, R. E.: Automated colorimetric determination of polyphenols in tobacco leaf; 29th Tobacco Chemists' Research Conference, Beltsville, Maryland, 1975, abstracts, p. 23.
9. Zane, A., and S. H. Wender: Pyrolysis products of rutin, quercetin, and chlorogenic acid; Tob. Sci. 7 (1963) 21–23.

Authors' address:

Department of Biological and
Agricultural Engineering,
School of Agriculture and Life Sciences,
North Carolina State University,
P.O. Box 5906,
Raleigh, North Carolina, 27650, U.S.A.