Optimization of Mathematical Treatments of Raw Near-Infrared Signal in the Measurement of Protein in Hard Red Spring Wheat. I. Influence of Particle Size

K. H. NORRIS[1] and P. C. WILLIAMS[2]

Cereal Chem. 61(2):158-165

ABSTRACT

Diffuse reflectance does not vary linearly with absorber concentration. Therefore, if linear correlation techniques are used for near-infrared reflectance spectroscopy, mathematical treatments of the reflectance data are required. Data treatments found to give a linear correlation with protein content of wheat include log 1/R, $dR/_R$, $\Delta(\log 1/R)$, $d(\log 1/R)$, $d^2(\log 1/R)$, and $(1-R)^2/2R$. These data treatments were evaluated for performance in predicting protein content for wheat samples ground with different grinders to give a wide range of particle sizes. The spectral data were also normalized by dividing by the same data treatment at a reference wavelength. Particle size had a marked effect on the near-infrared spectra, and this effect carried through to produce large errors in protein prediction without normalization of spectral data. Normalization of data effectively removed the particle-size effect such that second derivative divided by second derivative at optimum wavelengths gave an average bias error of −0.02% with a standard deviation of bias of 0.12% protein for six sample lots with mean particle size varying from 161 to 327 μm. Normalized second-derivative treatment of data gave the best performance in predicting protein content of samples varying widely in particle size. Inclusion of particle-size variation in the calibration samples improved the performance of all the other data treatments, making several of them equal to the normalized second derivative.

The most important factors affecting the accuracy of near-infrared reflectance spectroscopy (NIRS) for the analysis of hard red spring (HRS) wheat are the mean particle size (MPS) and particle size distribution (PSD) of the ground wheat (Williams 1975, Williams and Thompson 1978), sample temperature (Williams and Norris 1982), the moisture and protein content relative to that of the samples used in calibration (Williams and Thompson 1978, Watson et al 1977), and the growing environment of wheat (Watson et al 1977). The moisture and protein content of the wheat as well as the growing environment affect the MPS, PSD, and bulk density of the sample. Bulk density, MPS, and PSD affect the packing characteristics of ground wheat and the nature of the surface presented to the instrument. This influences the penetration of radiation into the sample and the reflectance from the sample. Sample-induced errors in NIR analysis can be caused by suboptimum wavelength selection and by increases in the variance of the NIR signal received as a result of factors associated with the sample surface. The incidence of such errors could be minimized by improving the methods of utilizing the energy signal received by the instrument.

Modern NIR instruments normally receive the raw energy signal from sample surfaces and transform this signal into the log 1/R, where R is the sample reflectance. The original source of the energy is usually a tungsten-filament lamp. The energy utilized by the

[1]Chief, Instrumentation Research Laboratory, Beltsville Agricultural Research Center, ARS, USDA, Beltsville, MD 20705.
[2]Chemist-in-charge, Grain Research Laboratory, Canadian Grain Commission, Winnipeg, Manitoba, Canada R3C3G9.

instrument is restricted to that which is reflected at selected wavelengths in the NIR region, which corresponds to the optimum wavelengths selected for the measurement of protein, water, oil, or other constituents. The log 1/R signals are translated into predicted values for protein, moisture, and oil content by analog circuitry or digital microprocessors in the instrument. The log 1/R signals can be used in the computation of the percentage of protein and other constituents, either directly or after preliminary mathematical processing. Finally, each wavelength data point, recognized by the computing circuitry of commercial NIR instruments, corresponds to the summed signals taken over a range of several units, usually 8–15 nm. System noise is controlled by summation of data taken from several adjacent points (smoothing) and by the summation of repeated readings taken over the same wavelength points. The mathematical treatment of log 1/R data by the instrument incorporates both manipulations of the data itself and the method of summation of data over short ranges of wavelength points.

The Technicon/Dickey-john InfraAlyzer (model 2.5), the Technicon InfraAlyzer (models 400, 300, and 200), and the Dickey-john Grain Analysis Computer (model III [GAC III]) use the log 1/R signals from six wavelength points without preliminary treatment. This algorithm or mathematical model is referred to as the log 1/R treatment. The specific wavelengths are preselected by six narrow-band interference filters. Four of these filters coincide with the wavelengths of peak absorption for oil, protein, carbohydrates, and water. The other two filters are at reference wavelengths where these four components have minimum absorption.

The Neotec Grain Quality Analyzer, model 31 (GQA 31), performs a mathematical pretreatment of the log 1/R signals at three wavelength points, corresponding to the selected optima for protein, water, and oil. At each of these three points, the log 1/R signal of a reference wavelength is subtracted. The differences in log 1/R for the three data points are referred to as delta (Δ) values, and the mathematical treatment is known as the Δ(log 1/R) or Δ OD (optical density) treatment. The three Δ(log 1/R) values in the GQA 31 are each multiplied by correction factors specific to the actual filters used in individual instruments. The three (Δ[log 1/R] × correction factors) constitute the familiar C values used in calibration of the GQA 31.

The Neotec GQA model 41 uses diffuse reflectance, R, in making the computation. The signal is further treated by taking a derivative of the R value at selected wavelength points, and dividing the result through by the R value of the original data point. This is known as the dR/R treatment.

The Δ(log 1/R) treatment becomes a first derivative of log 1/R, d(log 1/R) when the spacing between the two wavelength points approaches zero. The optimum distance between the two points can be established during the design of the instrument by the use of regression analysis, but it is usually within 1–10 wavelength points of the original; ie, within about 2–20 nm. An alternative treatment is to compute the second derivative of the log 1/R signal at selected wavelength points, and incorporate the second derivative into an algorithm for the final computation of constituent percentage. This is referred to as the *second derivative* or $d^2(log\ 1/R)$ treatment. Fourth and eighth derivatives of the log 1/R signal have also been used in the computation of protein, moisture, and other constituents by NIRS, although system noise increases markedly at higher derivatives and the second derivative appears to be optimum. These mathematical treatments of the data are shown in Fig. 1.

Kubelka and Munk (1931) developed a treatment of the signal that involved reconversion of the signal to the Kubelka-Munk function (K/S). Their function was as follows:

$$\frac{K}{S} = \frac{(1-R)^2}{2R}$$

where R is the reflectance, K is the absorption coefficient, and S is the scatter coefficient.

The above mathematical treaments were applied to the analysis of HRS wheat in the present study, both alone and after normalization. Normalization involved the selection of two sets of wavelengths. The first predicted protein in the usual way, and the second served as reference wavelength. The signal received at the first, or prediction, wavelength was treated by one of the mathematical treatments referred to above, and the normalization was achieved by dividing the result by the result obtained by performing the same mathematical treatment of the signal received at the second, or reference, wavelength.

MATERIALS AND METHODS

HRS wheat samples were selected at the Grain Research Laboratory, Winnipeg, Canada. These included sets of 50 samples of grades 1 and 2 Canadian Western red spring wheat. Also included were four sets of the same 40 samples ground to different particle sizes by using different grinding procedures. Set A was ground on the Udy cyclone grinder using a 1.0-mm screen. Set B was ground on the Falling Number, KT-30, burr mill set at its finest setting (0). Sets C and E were ground on a Retsch centrifugal grinder, using, respectively, a 0.35- and a 0.6-mm screen. MPS of these four sets were, respectively, 191, 327, 161, and 248 μm. A combined file was compiled containing 85 samples (20–22 samples from each of the four "particle-size" series). A further series of 50 Canadian HRS wheat samples (W50) were used from a previous study set up to examine the influence of grade on the selection of wavelengths and the accuracy of measurement of protein and moisture. These samples were also ground on the cyclone grinder, using a 1.0-mm screen. This series of samples was assumed to have the normal variance with an MPS of 198 μm ±9, in agreement with

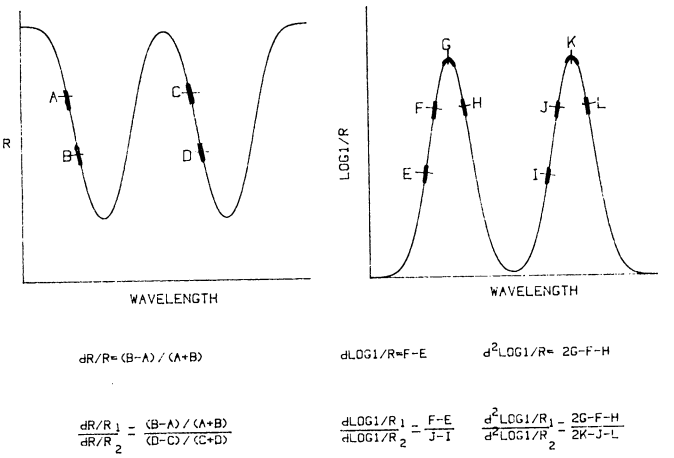

Fig. 1. Computation of dR/R, d(log 1/R₁)/d(log 1/R₂) and d²(log 1/R₁)/d²(log 1/R₂). The broad lines indicate the data points to be averaged in the computations.

TABLE I
Details of Files (sets) of Hard Red Spring Wheat Samples Used in Experiments A and B

Name	Grinder	Screen Aperture (mm)	No. of Samples	Particle Size Mean (μm)	Variability[a] (μm)
27A	Cyclone	1.0	27	191	9.2
27B	KT-30	···	27	327	8.9
27C	Retsch	0.35	27	161	10.3
27E	Retsch	0.60	27	248	18.5
69AB	All	···	69	233	66.4
85AB	All	···	85	230	63.1
W50	Cyclone	1.0	50	ND[b]	···
LG2	Cyclone	1.0	50	ND	···
WM95	Cyclone	1.0	95	ND	···

[a] Standard deviation of mean particle size of samples in file.
[b] Not determined on these samples, long-term average mean particle size of hard red spring wheat ground on cyclone grinder with 1.0 mm screen is 198 ± 9 μm.

the long-term MPS data for cyclone-ground HRS wheat. Finally, a series of 95 samples of HRS wheat was prepared that was to serve as the basis for a comparative study of physicochemical methods for the determination of protein in wheat (Williams et al 1978). These samples were also used as calibration samples in part of the study. Tables I and II summarize the makeup of the files and the wavelengths selected for prediction.

The Beltsville Universal Computerized Spectrophotometer (BUCS) consists of two Cary model 14 grating/prism monochromators which, by incorporating several different detection devices, enable studies to be done in the ultraviolet, visible, and near-infrared regions of the spectrum. The visible/UV spectrophotometer has both reflectance and transmission attachments. Both spectrophotometers are interfaced to a Nova model 2 computer with a memory capacity of 32,000 words. Data can be recorded and stored on a nine-track magnetic tape or hard disk, and a Tektronix model 613 CRT display enables the visual examination of spectra and correlation plots. Original raw spectral data from samples are recorded either as R, T, log 1/R, or log 1/T, and different mathematical transformations are applied by the computer for subsequent analytical work.

In this study, the monochromator optimized for infrared reflectance was used, and all samples were scanned from 1,000 to 2,640 nm, with readings taken at intervals of every 0.2 nm. The near-infrared readings measured as log 1/R were smoothed by averaging 21 adjacent points and reinstating the result to the original wavelength point. Finally, the array of 8,192 wavelength points was shrunk to 1,024 data points by averaging each point with the four points on either side. Then, every eighth point was selected so that each data point represented a wavelength interval of 1.6 nm. Mathematical treatments were compared by treating the arrays of log 1/R data at the 1,024 wavelength points by the representative simple mathematical processes described above, and regressing the resultant values against protein for each sample. The wavelength at which the highest correlation is obtained is selected for the prediction of the required constituent. The program used in this study yielded the best overall correlation of the treated log 1/R data against protein, and the standard error of the resultant estimate. The veracity of the prediction equation was tested by using the wavelength/mathematical treatment combination to predict protein in "unknown" files of samples.

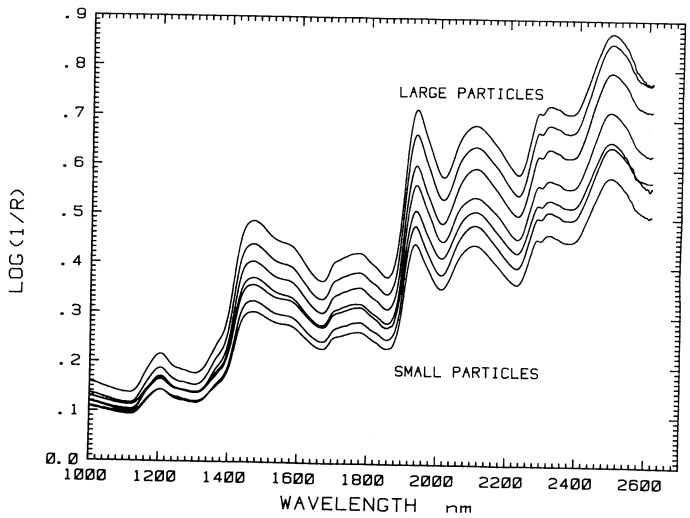

Fig. 2. Typical log 1/R spectral data for ground wheat samples with MPS varying from 150 to 335 μm.

Two sets of comparisons were made. In experiment A, calibrations were performed by all mathematical treatments, using 145 samples of HRS wheat (files LG2 and WM95). All of these samples were ground on the cyclone grinder and were of normal variance in MPS. From these calibrations, protein was predicted in the files of samples of wheat that had been ground on different grinders to determine the interactions between variations in MPS and mathematical treatment on the accuracy of prediction of protein. The objective of the second experiment (B) was to determine the manner in which the accuracy of protein prediction on ground wheat of different MPS would be improved by the inclusion in calibration of wheat displaying more variance in MPS. For this experiment, calibrations by all mathematical treatments were based on the file containing 85 samples with maximum variance in MPS (Table I). The overall weighted-mean standard deviation of difference between NIRS and chemical data was taken to indicate the optimum mathematical treatment of the raw log 1/R data for the prediction of the required constituent. High correlations were obtained in several different regions of the spectrum, and each of the possible high correlations was tested by prediction to verify the selection of the optimum point.

For the normalization procedure, a program was written to select the optimum denominator wavelength using the optimum numerator wavelength as a starting point for the selection. The procedure was thereafter repeated using the selected denominator wavelength to reselect the numerator wavelength. The process was repeated using up to 12 alternative reference wavelengths to verify that the selection of paired numerator/denominator wavelengths were truly the wavelength combination that resulted in the closest prediction of the required constituent. In each case, the numerator

TABLE II
Wavelengths (nm) Used in Prediction of Protein in Experiments A and B

	No. of Terms	Sums (nm)	1	2	3	4	5	6
Sim. Inf. 2.5	3	8	2,179	2,101	1,680			
Sim. Inf. 2.5	6	8	2,179	2,101	1,941	2,310	1,680	2,230
Sim. GQA 31	3	8	(2,162	– 2,107)	(1,931	– 1,869)	(2,304	– 2,264)
Sim. GQA 41	4	14	2,149	1,890	2,291	2,080	GAP = 18 nm	
log 1/R	3	5	2,171	2,139	2,466			
log 1/R	5	5	2,171	2,139	2,466	2,371	1,840	
d(log 1/R)	2	5	2,160	2,197	GAP = 11 nm			
dR/R	2	5	2,160	2,197	GAP = 11 nm			
d(log 1/R)	2	8	2,173	2,126	GAP = 29 nm			
KM	3	5	2,211	2,152	2,171			
KM	5	5	2,211	2,152	2,171	2,189	1,715	
$\frac{d(\log 1/R_1)}{d(\log 1/R_2)}$	1	0	$\frac{2,160}{2,262}$	GAP = 21 nm				
$\frac{d^2(\log 1/R_1)}{d^2(\log 1/R_2)}$	1	18	$\frac{2,178}{1,771}$	GAP = 24 nm				
$\frac{KM_1}{KM_2}$	1	11	$\frac{2,167}{2,157}$					

and denominator wavelength points were further optimized by varying the number of data points summed. In addition to the mathematical treatments described above, results were also predicted using, as closely as possible, simulations of the treatments of the log 1/R signal used by the commercially available NIR instruments.

RESULTS AND DISCUSSION

Visual Evaluation

The log 1/R values are affected by particle size as shown in Fig. 2 for typical spectra of HRS wheat samples with MPS ranging from 150 to 335 μm. The coarser samples have higher absorption and higher log 1/R values. The particle size effect is greater at longer wavelengths but, as shown in Fig. 3, the effect is not consistent with wavelength. The lines of Fig. 3 are the regression lines relating log 1/R to particle size at selected wavelengths, using the 69 samples of file 69AB. The correlation coefficients for these regressions varied from 0.86 to 0.96. The lines of Fig. 3 have been normalized by subtracting the log 1/R value at the MPS value of 100 to permit better comparison of the slopes, which represent the particle-size effect. The same type of data treatment was applied for the lines of Fig. 4, showing the particle-size effect at different log 1/R levels. These plots show that the particle-size effect increases consistently with log 1/R level, indicating that the primary relationship is to log 1/R rather than to wavelength.

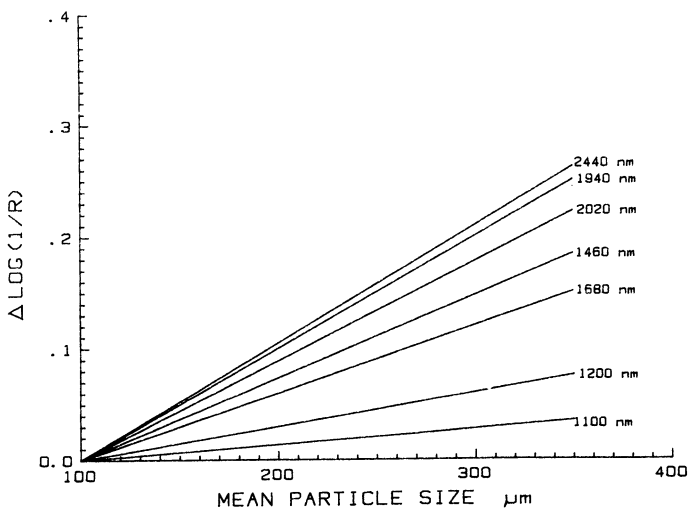

Fig. 3. Relationship between the change in log 1/R and particle size at selected wavelengths.

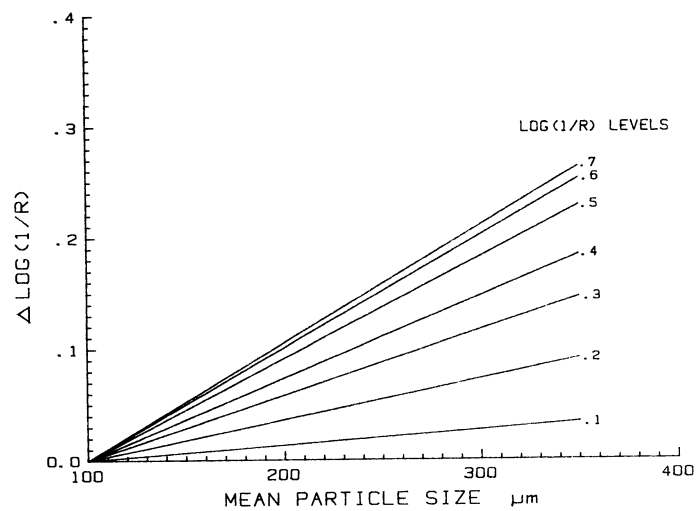

Fig. 4. Relationship between the change in log 1/R and particle size at different log 1/R levels.

The effect of a 5% difference in protein content is illustrated in Fig. 5 for two samples having nearly the same particle size. The protein absorption band at 2,180 nm is difficult to distinguish on the log 1/R traces, but the first derivative curves clearly show the difference at 2,160 nm. Less defined deviations also occur at approximately 1,560, 1,680, 1,760, and 2,520 nm for the derivative traces. Wavelength point A (2,160 nm) is the wavelength found by the regression procedure for the numerator with wavelength point B (2,262 nm) for the denominator to best correlate to as-is protein content. A plot of the correlation coefficients as a function of numerator wavelength (Fig. 6) shows high correlations at several wavelengths, with point A being the highest. The denominator correlation coefficients as a function of wavelength (Fig. 7) shows high correlations at many wavelength regions. The correlations are essentially equal at 2,200 nm and 2,262 nm, although one is positive and the other negative. Wavelength point B (2,262 nm) was chosen for the denominator or reference wavelength for the first-derivative regression treatment.

The dR/R traces for the two wheat samples discussed above are shown in Fig. 8. The dR/R traces are essentially the same as the negative value of the d(log 1/R) traces. They also show the greatest difference at 2,160 nm for these two wheat samples, and this was the wavelength found by the regression procedure for the numerator wavelength. Wavelength point B (2,262 nm) was found to provide

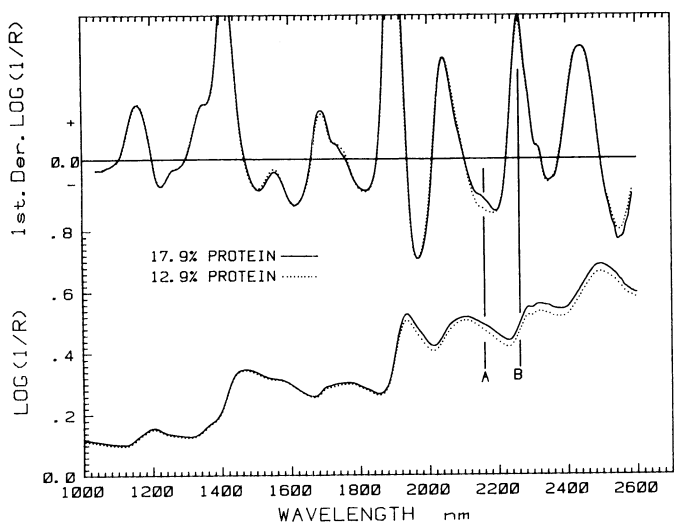

Fig. 5. Sharpening of spectral character with first derivative for two wheat spectra with 5% protein difference.

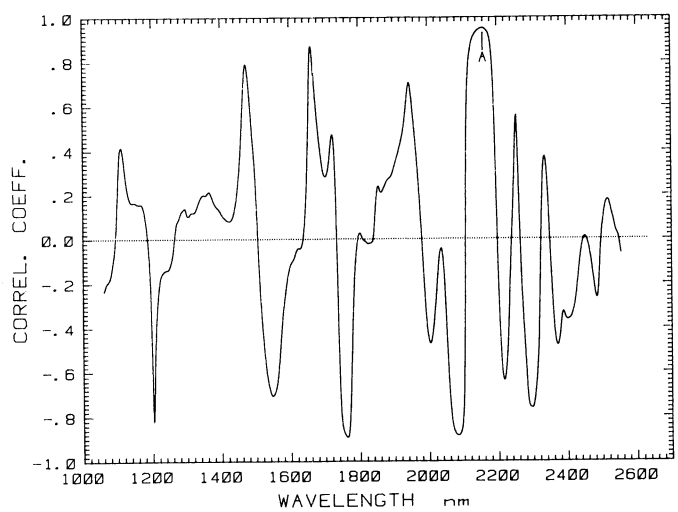

Fig. 6. Correlation coefficient versus wavelength when scanning numerator using $d(\log 1/R_1)/d(\log 1/R_2)$ to correlation with protein content of wheat. Denominator wavelength at 2,262 nm.

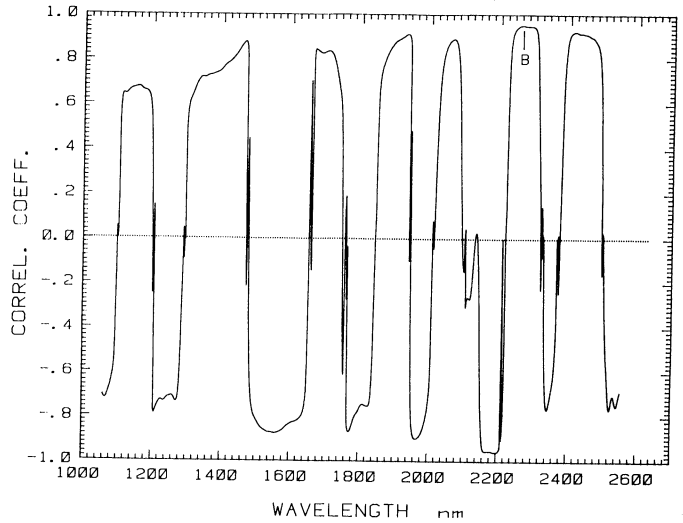

Fig. 7. Correlation coefficient versus wavelength when scanning denominator using $d(\log 1/R_1)/d(\log 1/R_2)$ to correlate with protein content of wheat. Numerator wavelength at 2,160 nm.

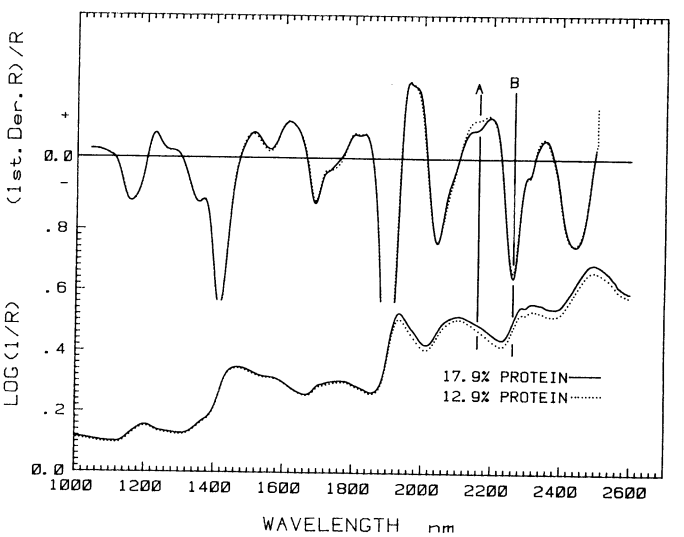

Fig. 8. Sharpening of spectral character with dR/R treatment for two wheat spectra with 5% protein difference.

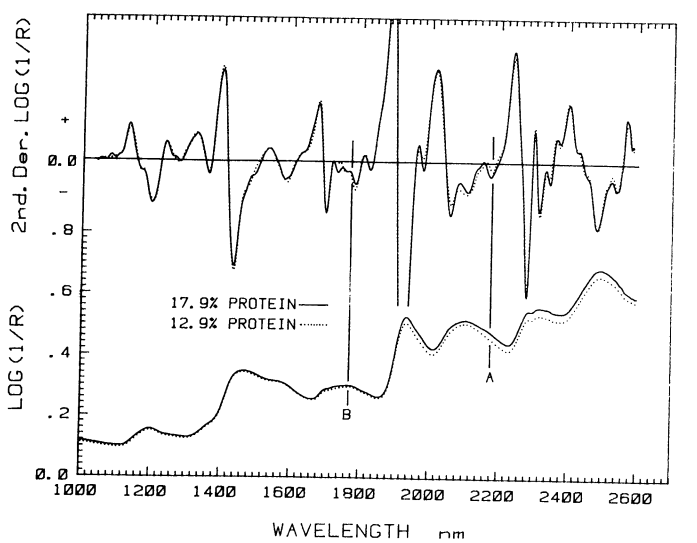

Fig. 9. Sharpening of spectral character with second derivative for two wheat spectra with 5% protein difference.

the best denominator or reference wavelength. Figure 9 illustrates the log 1/R traces and the second-derivative traces for the same two wheat samples. In this case, the maximum deviation between the derivative curves occurred at 2,060 nm with smaller deviations at 1,500, 1,730, 1,780, 2,080, 2,170, and 2,530 nm. However, the regression procedure gave the highest correlation at 2,178 nm for the numerator and 1,771 nm for the denominator.

The computed K/S for the two samples previously discussed are shown in Fig. 10. The curves are also shown for the same wheat, but ground to provide a much larger particle size. The particle-size effect is more pronounced on the K/S traces than for the log 1/R traces. Wavelength point A (2,163 nm) was found to provide the best numerator and B (2,158 nm) the best denominator for correlating to as-is protein content.

The influence of the MPS of ground wheat upon the log 1/R traces has been referred to above. Figure 11 gives the log 1/R traces of the two grinds of a sample at 15.7% protein and illustrates the differences in absolute reflectance values induced by the differences in MPS. The d(log 1/R) plots for the same two samples are compared with the log 1/R spectra in Fig. 12. The particle-size effect also influences the derivative traces as illustrated in Fig. 12 for the first derivative traces. At wavelength point A, the derivative treatment gives nearly the same value for these two samples of the same protein content but different particle size. The values at point B are also close together but, in addition, the difference between the

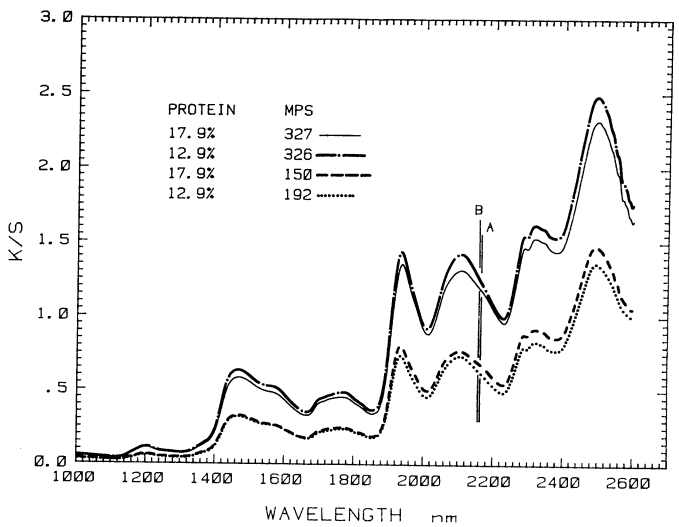

Fig. 10. Kubelka-Munk function, K/S, for two wheat samples with 5% protein difference, and large difference in particle size.

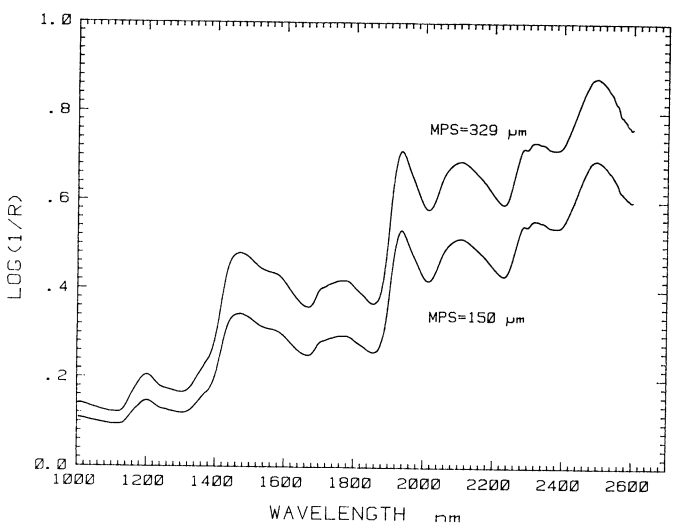

Fig. 11. Effect of grinder (particle size) on the log 1/R spectrum for the same wheat sample.

two derivative curves at point B are in the opposite direction from those at point A. Therefore, dividing the respective derivative values at A by the derivative values at B will give an even smaller difference between the values for these two samples. The dR/R curves for these two samples are not shown because they are the same as the negative of the d(log 1/R) curves as noted for Figs. 5 and 8.

The effects of particle size on second derivative traces are shown in Fig. 13. Again, the derivative curves show a small difference at wavelength point A, the wavelength providing the best numerator for calibration for protein. A small difference is also apparent at wavelength point B, but most of the particle-size effect has been eliminated by the second derivative at these two wavelengths.

The normalization effect of dividing by the derivative value at a reference wavelength is illustrated in Fig. 14 for the first derivative. Only a limited region of the spectrum is shown to permit a more expanded scale. The four log 1/R traces represent two samples of different protein content ground with two different grinders to provide a big difference in particle size. The four derivative curves coincide at wavelength point B because of the normalization at this wavelength. At wavelength point A, the derivative curves coincide for the same protein level in spite of the big differences in particle size. At this wavelength, the protein difference is also apparent. The corresponding plots for the normalized second derivatives (Fig. 15) show the same cancellation of particle-size effects while retaining the sensitivity to protein differences.

It should be noted that the normalized dR/R and d(log 1/R) treatments requires four wavelengths, whereas the normalized d^2(log 1/R) treatment requires six. In terms of commercial instrumentation, the two additional wavelengths per constituent could represent a disadvantage in design and cost.

Statistical Comparisons

Experiment A: Normal MPS variance in calibration samples. Table III summarizes the results of this study. Calibration to as-is protein was performed with 145 samples of files LG2 and WM95, all having normal variance in MPS. Thirteen different mathematical treatments were used, including four simulations of commercial instruments. Protein was predicted in six different files of HRS wheat, which displayed both a wide overall variation in MPS and big differences in the variability of MPS within individual files. The normalized treatment of the second derivative of the log 1/R data proved to be the most satisfactory for the prediction of protein in wheat varying widely in MPS, when the calibration was based on wheat with normal variation in MPS. The average standard deviation from Kjeldahl was 0.30% with an overall bias of only −0.02%. All other mathematical treatments displayed considerably more variance in the results of prediction of

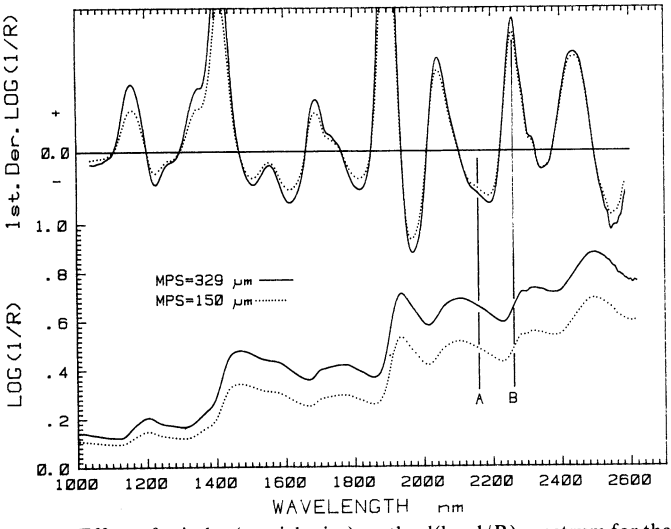

Fig. 12. Effect of grinder (particle size) on the d(log 1/R) spectrum for the same wheat sample.

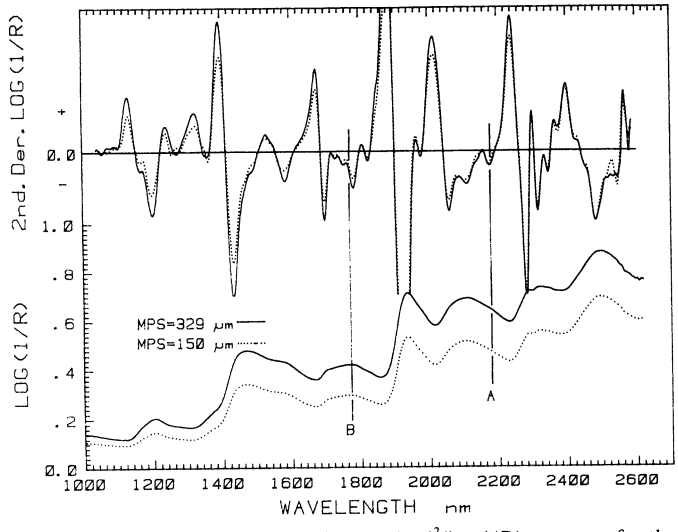

Fig. 13. Effect of grinder (particle size) on the d^2(log 1/R) spectrum for the same wheat sample.

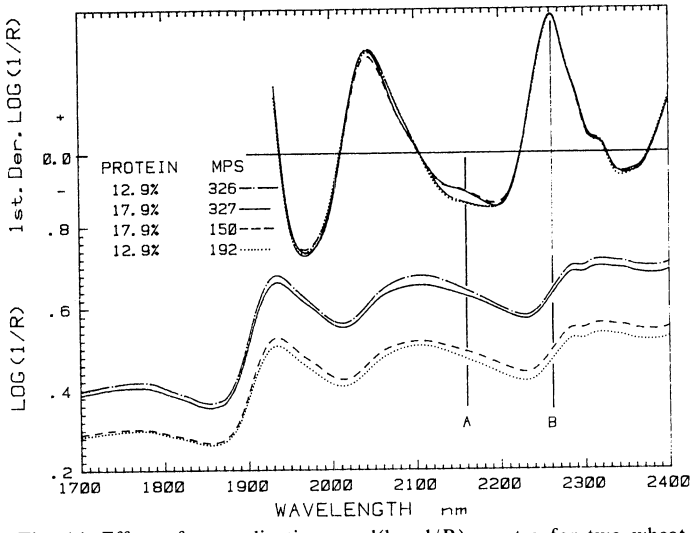

Fig. 14. Effect of normalization on d(log 1/R) spectra for two wheat samples ground on two different grinders.

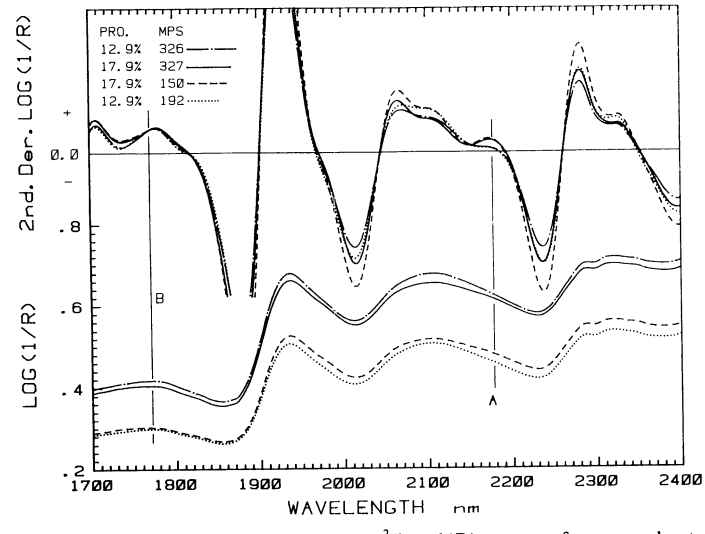

Fig. 15. Effect of normalization on d^2(log 1/R) spectra for two wheat samples ground on two different grinders.

wheat of different MPS. This was particularly apparent in the average bias from Kjeldahl for individual files. The last column of the table gives the standard deviation of these biases. Every mathematical treatment other than the normalized second derivative treatment biased at least two of the six files by more than 0.20% protein. Of the simulated instruments, the GQA 41 appeared to provide the most stable predictions of protein in the presence of large variations in MPS. The five specially selected log 1/R wavelengths also provided a stable series of results with a low overall standard deviation of difference, and it is possible that incorporation of these wavelengths into instruments such as the Technicon InfraAlyzer or the Dickey-john GAC series would lead to improvements in the accuracy of protein determinations.

Experiment B: High MPS variance in calibration samples. It was expected that differences in the reaction of NIR instruments to ground wheat of different MPS could be minimized by the incorporation of maximum variance in MPS in the calibration. File 85AB was used for this calibration. Protein was predicted in the same six files as in experiment A, as shown in Table IV. In most instances, the performance of individual mathematical treatments improved over their performance in experiment A. The normalized second-derivative treatment still gave the best overall predictions

TABLE III
Influence of Mathematical Treatment of Reflectance Data on Accuracy of Protein Prediction in Hard Red Spring Wheat of Different MPS, Using Low-Variance MPS Calibration

Simulated Instruments	No. of Terms	27A		27B		27C		27E		69AB		W50		Overall		
		d[a]	SD[b]	d[a]	SD[b]	d[a]	SD[b]	d[a]	SD[b]	d[a]	SD[b]	d[a]	SD[b]	d[a]	SD[b]	SD Bias
INF 2.5[c]	3	0.34	0.28	0.63	0.43	0.46	0.37	0.76	0.38	0.49	0.39	−0.41	0.24	0.32	0.35	0.41
INF 2.5[c]	6	−0.58	0.43	−2.46	0.68	0.95	0.58	−0.42	0.41	−0.72	1.33	−0.43	0.27	−0.61	0.83	1.09
GQA-31	3	0.23	0.45	0.00	0.59	−0.09	0.56	−0.88	0.64	−0.32	0.85	−0.07	0.27	−0.20	0.52	0.38
GQA-41	4	0.04	0.25	0.24	0.40	0.05	0.33	−0.46	0.43	−0.09	0.38	−0.28	0.22	−0.10	0.34	0.25
Math Treatments																
log 1/R	3	−0.31	0.20	−0.80	0.42	−0.09	0.31	−0.22	0.37	−0.39	0.39	−0.17	0.31	−0.32	0.35	0.25
log 1/R	5	0.22	0.25	0.14	0.35	0.41	0.23	0.13	0.33	0.04	0.36	−0.09	0.30	0.00	0.32	0.23
d(log 1/R)	2	0.01	0.20	0.28	0.35	0.13	0.19	−0.27	0.32	0.08	0.32	−0.15	0.39	0.01	0.32	0.20
d^2(log 1/R)	2	0.04	0.23	0.30	0.41	0.17	0.27	−0.18	0.35	0.14	0.38	−0.34	0.31	0.01	0.34	0.24
KM	3	−0.04	0.35	0.38	0.95	0.19	0.29	−0.13	0.60	0.21	0.49	−0.25	0.52	0.06	0.56	0.24
KM	5	−0.02	0.29	0.71	0.87	0.15	0.26	0.25	0.51	0.37	0.49	−0.38	0.32	0.16	0.49	0.37
$\frac{d(\log 1/R_1)}{d(\log 1/R_2)}$	1	−0.13	0.20	0.23	0.34	0.49	0.24	0.42	0.34	0.31	0.36	−0.07	0.29	0.21	0.30	0.26
$\frac{d^2(\log 1/R_1)}{d^2(\log 1/R_2)}$	1	−0.06	0.23	0.11	0.27	0.09	0.25	−0.18	0.36	0.04	0.29	−0.14	0.34	−0.02	0.30	0.12
$\frac{KM_1}{KM_2}$	1	−0.01	0.32	1.48	0.51	0.08	0.38	0.54	0.38	0.62	0.67	0.22	0.52	0.48	0.53	0.54

[a]d = bias between near-infrared reflectance protein prediction for individual files versus Kjeldahl protein.
[b]SD = standard deviation of difference between near-infrared reflectance protein prediction for individual files versus Kjeldahl.
[c]Also includes Dickey-john GAC III and 660 and Technicon InfraAlyzer 400, 300, and 200.

TABLE IV
Influence of Mathematical Treatment of Reflectance Data on Accuracy of Protein Prediction in Hard Red Spring Wheat of Different MPS, Using High-Variance MPS Calibration

Simulated Instruments	No. of Terms	27A		27B		27C		27E		69AB		W50		Overall		
		d[a]	SD[b]	d[a]	SD[b]	d[a]	SD[b]	d[a]	SD[b]	d[a]	SD[b]	d[a]	SD[b]	d[a]	SD[b]	SD Bias
INF 2.5[c]	3	−0.27	0.28	0.08	0.43	−0.17	0.35	0.10	0.36	−0.12	0.38	−0.98	0.26	−0.28	0.35	0.39
INF 2.5[c]	6	−0.12	0.29	0.06	0.37	−0.03	0.23	−0.10	0.30	−0.02	0.33	−0.50	0.26	−0.10	0.30	0.22
GQA-31	3	0.35	0.33	0.23	0.38	−0.08	0.36	−0.12	0.36	−0.03	0.41	−0.63	0.26	−0.16	0.36	0.32
GQA-41	4	−0.24	0.25	0.35	0.32	−0.02	0.33	−0.13	0.36	−0.02	0.36	−0.41	0.29	−0.10	0.32	0.26
Math Treatments																
log 1/R	3	−0.21	0.21	0.11	0.30	−0.08	0.25	−0.11	0.34	−0.02	0.29	−0.03	0.32	−0.02	0.29	0.14
log 1/R	5	0.08	0.19	0.10	0.30	0.16	0.24	0.10	0.28	−0.01	0.26	−0.34	0.28	−0.08	0.26	0.15
d(log 1/R)	2	−0.05	0.22	0.36	0.34	0.06	0.21	−0.18	0.33	0.10	0.33	−0.20	0.41	0.02	0.30	0.21
d^2(log 1/R)	2	−0.05	0.26	0.28	0.36	0.07	0.29	−0.17	0.32	0.09	0.39	−0.38	0.33	−0.04	0.34	0.23
KM	3	−0.20	0.47	0.30	0.68	0.02	0.41	−0.01	0.49	0.14	0.49	−0.37	0.54	−0.03	0.52	0.24
KM	5	−0.26	0.44	0.22	0.64	−0.01	0.39	0.07	0.45	0.10	0.44	−0.54	0.47	−0.09	0.47	0.28
$\frac{d^2(\log 1/R_1)}{d^2(\log 1/R_2)}$	1	−0.35	0.24	−0.03	0.26	0.36	0.21	0.15	0.32	0.08	0.37	−0.19	0.31	−0.01	0.31	0.25
$\frac{d(\log 1/R_1)}{d(\log 1/R_2}$	1	−0.02	0.23	0.14	0.27	0.12	0.25	−0.15	0.36	0.07	0.29	−0.10	0.34	−0.01	0.30	0.12
$\frac{KM_1}{KM_2}$	1	−0.41	0.34	0.86	0.39	−0.33	0.32	0.06	0.37	0.09	0.58	−0.16	0.46	0.01	0.46	0.46

[a]d = bias between near-infrared reflectance protein prediction for individual files versus Kjeldahl protein.
[b]SD = standard deviation of difference between near-infrared reflectance protein prediction for individual files versus Kjeldahl.
[c]Also includes Dickey-john GAC III and 660 and Technicon InfraAlyzer 400, 300, and 200.

on all files, but the specially selected 3- and 5-wavelength log 1/R treatment also provided excellent data. An interesting contrast occurred between the results of the simulated InfraAlyzer/GAC results for normal- and high-variance calibrations. In the case of normal variance calibrations, the six wavelength series gave rather poor results and the three wavelength series, in each case, showed significant improvement. When high variance was included in the calibration, the six wavelength series gave significantly better results than the three wavelength series. These observations also contrasted with the results obtained for the three- and five-wavelength series of specially selected log 1/R wavelengths, where the five wavelength series gave the better performance with normal MPS variance in calibration, and with high MPS variance there was very little difference between the two sets.

SUMMARY AND CONCLUSIONS

The experiments described illustrate the influence of variations in MPS upon the log 1/R signal from ground HRS wheat. The influence of MPS after several different mathematical treatments of the original reflected signal was also demonstrated. Differences between spectral traces of HRS wheat samples of different protein content were greater at wavelength points selected by the computer as possible operation points for the measurement of protein. Normalization of the signal by dividing throughout by the signal at a reference wavelength reduced the differences between samples of the same wheat ground to whole meals of different MPS and effectively increased the differences between samples of different protein content at specific wavelength points. It was shown that: log 1/R treatment of the reflectance signal leads to differences in the accuracy of prediction of protein when the MPS is variable; normalization of the reflectance signal by dividing through by the signal at a reference wavelength improves the accuracy of analysis regardless of the mathematical treatment; statistical comparisons based on measurements using equivalent numbers of wavelengths confirmed that the most satisfactory mathematical treatment was that of taking the second derivative of the reflectance signal at a wavelength of 2,178 nm and dividing by the signal at the reference wavelength of 1,771 nm; and inclusion of maximum variance in MPS on the samples used in calibration greatly improved the accuracy of prediction of protein in HRS wheat in the presence of variations in MPS. The improvement was such that there was no significant difference between several mathematical treatments, provided that equivalent numbers of wavelength points were used, although the normalized second-derivative treatment remained slightly superior to all others in terms of minimum bias and minimum variability of bias.

ACKNOWLEDGMENTS

The authors acknowledge with gratitude the technical assistance of S. G. Stevenson and H. Cordeiro in the preparation and analysis of wheat samples, and J. Shaffer in the processing of data.

LITERATURE CITED

KUBELKA, P., and MUNK, F. 1931. Ein Beitrag zur Optik der Farbanstriche. Z. Tech. Physik. 12:593.

WATSON, C. A., SHUEY, W. C., BARIASIK, O. J., and DICK, J. W. 1977. Effect of wheat class on near-infrared reflectance. Cereal Chem. 54:1264.

WILLIAMS, P. C. 1975. Application of near-infrared reflectance spectroscopy to analysis of cereal grains and oilseeds. Cereal Chem. 52:561.

WILLIAMS, P. C., and NORRIS, K. H. 1982. Influence of temperature on estimation of protein and moisture in wheat by near-infrared reflectance. Cereal Chem. 59:473.

WILLIAMS, P. C., and THOMPSON, B. N. 1978. Influence of whole meal granularity on analysis of HRS wheat for protein and moisture by near-infrared reflectance spectroscopy (NRS). Cereal Chem. 55:1014.

WILLIAMS, P. C., NORRIS, K. H., JOHNSON, R. L., STANDING, K., FRICIONI, R., MacAFFREY, D., and MERCIER, R. 1978. A cooperative study of modern physicochemical methods for the measurement of total nitrogen in hard red spring wheat. Cereal Foods World 23:544.

[Received October 15, 1981. Accepted November 8, 1983]