

# Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra

R. J. BARNES,\* M. S. DHANOA, and SUSAN J. LISTER

AFRC Institute for Grassland and Animal Production, Hurley, Maidenhead, Berks SL6 5LR, U.K.

Particle size, scatter, and multi-collinearity are long-standing problems encountered in diffuse reflectance spectrometry. Multiplicative combinations of these effects are the major factor inhibiting the interpretation of near-infrared diffuse reflectance spectra. Sample particle size accounts for the majority of the variance, while variance due to chemical composition is small. Procedures are presented whereby physical and chemical variance can be separated. Mathematical transformations—standard normal variate (SNV) and de-trending (DT)—applicable to individual NIR diffuse reflectance spectra are presented. The standard normal variate approach effectively removes the multiplicative interferences of scatter and particle size. De-trending accounts for the variation in baseline shift and curvilinearity, generally found in the reflectance spectra of powdered or densely packed samples, with the use of a second-degree polynomial regression. NIR diffuse reflectance spectra transposed by these methods are free from multi-collinearity and are not confused by the complexity of shape encountered with the use of derivative spectroscopy.

Index Headings: De-trending of NIR spectra.

## INTRODUCTION

Variation within individual near-infrared diffuse reflectance spectra recorded as the reciprocal log of the reflected energy is the result of three main sources, i.e.,

- (1) nonspecific scatter of radiation at the surface of particles;
- (2) variable spectral pathlength through the sample; and
- (3) chemical composition of the sample.

Scatter is dependent on the physical nature of the sample particles, and spectral pathlength is largely dependent on sample particle size. There is a high degree of collinearity between data points in the log 1/R spectra, which is a function to some extent of scatter and variable pathlength. The multiplicative combination of these effects is unique to any one spectrum, and any corrections for these interferences should be made on the same basis.

Principal component analysis<sup>1</sup> of near-infrared diffuse reflectance spectra shows a nearly linear slope for the first principal component when principal component weights are plotted against wavelength. For most agricultural materials, the first principal component can account for greater than 90% of the variance within the spectrum. The slope factor varies with particle size and packing density of the material. When the spectra of individual samples of similar composition are examined, the change of slope between different presentations is evident.

Various procedures have been reported to minimize the within-sample-set divergence of spectra.<sup>2,3</sup> These sample-set-dependent corrections are based on centering the underlying linear slope of each individual sample spectrum to the slope of the mean spectrum of the sample set. Cowe *et al.*<sup>4</sup> have shown, by using a spectral standardization procedure,<sup>2</sup> that the linear slope of the first principal component is eliminated. Second-derivative transformation of the log 1/R data will remove both parallel and slope changes and is not dependent on a mean spectrum. However, derivatized log 1/R spectra are complex and difficult to interpret and are dependent on operator choice of segment and gap size.

## STANDARD NORMAL VARIATE TRANSFORMATION

Mathematical transformation of the log 1/R spectra by calculation of the standard normal variation<sup>5</sup> at each wavelength removes slope variation on an individual sample basis by the use of the following calculation:

$$SNV_{(1-700)} = (y_{(1-700)} - \bar{y}) / \sqrt{\frac{\sum (y_{(1-700)} - \bar{y})^2}{n - 1}}$$

where  $SNV_{(1-700)}$  are the individual standard normal variations for 700 wavelengths,  $y$  is the 700-wavelength log 1/R values, and  $\bar{y}$  is the mean of the 700-wavelength log 1/R values.

## EXAMPLES

Six samples of sucrose obtained from different sources (four crystalline and two finely powdered forms) were

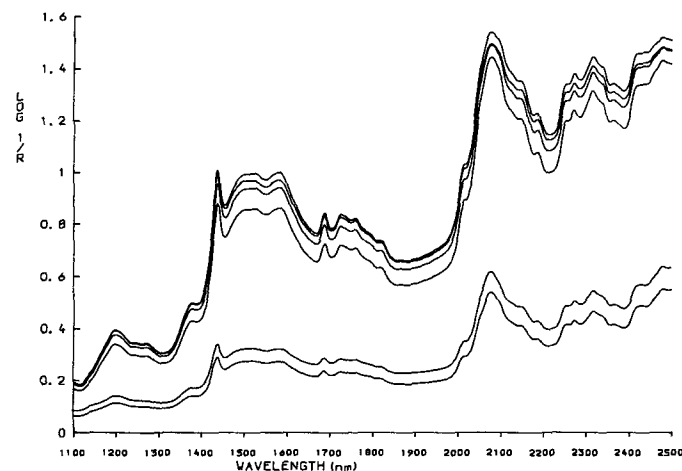


FIG. 1. Log 1/R spectra of sucrose. (Four upper = crystalline, two lower = powdered.)

Received 28 December 1988; revision received 10 February 1989.

\* Author to whom correspondence should be sent. Present address: Perstorp Analytical Ltd., Cooper Road, Thornbury, Bristol BS12 2UW U.K.

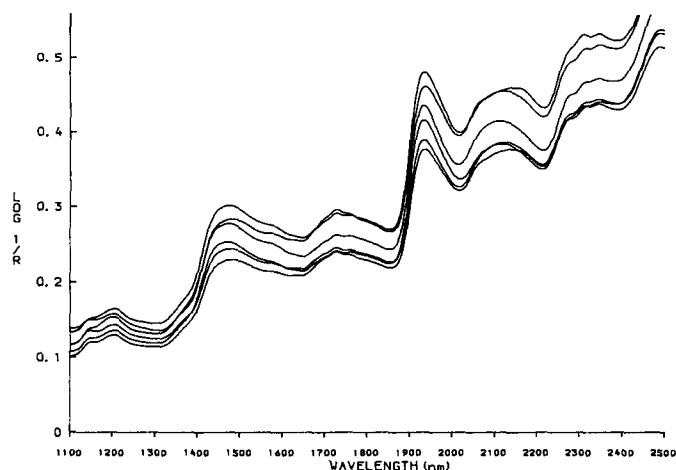


FIG. 2. Log  $1/R$  spectra of a series of grass hays.

scanned on a PSCo Model 6250 monochromator. Their near-infrared spectra are presented in Fig. 1, the four upper spectra being the crystalline sucrose samples. Both the underlying slope and divergence of that slope due to variation in particle size are evident, the powdered and hence more densely packed form showing much lower absorbances than the crystalline forms. Figure 2 shows the same divergence effect on a set of grass hays.

Graphical plots of the SNV transformed log  $1/R$  data vs. wavelength (SNV spectra) are the same shape as the log  $1/R$  spectra, but there is no divergence between similar samples with varying particle size. Figure 3 shows the SNV spectra of the six sucrose samples, and Fig. 4 the SNV spectra of the hay samples. The slopes of the SNV spectra for all six sucrose samples are similar, although differences in shape are apparent. The shape difference is most distinct between the powdered and the crystalline forms.

To test consistency of this mathematical transformation, we transformed a further set of straw spectra. Figure 5 shows the divergence of the log  $1/R$  spectra of the straw samples, and Fig. 6 the absence of divergence in the SNV spectra.

Principal component analysis of both log  $1/R$  and SNV spectra of the hay and straw sets were carried out. For both the hay and the straw samples the variance due to

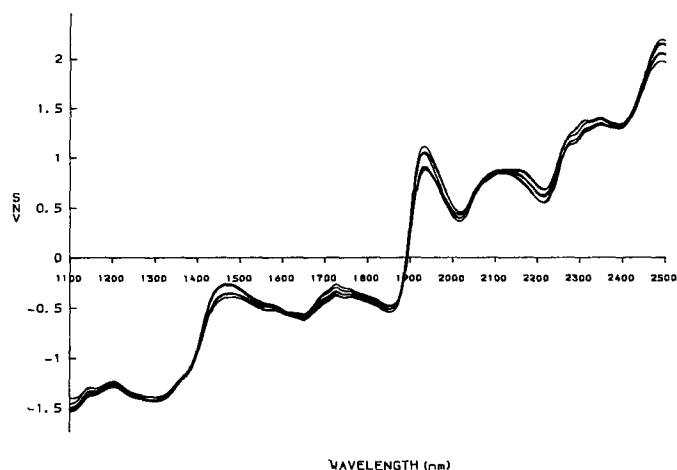


FIG. 4. Standard normal variate (SNV) spectra of grass hays.

the first principal component was always less for the SNV spectra. Subsequent principal components contributed more to the variance for the SNV spectra than the log  $1/R$  spectra (Table I).

Figures 7A–7C show the plots of the principal component weights derived from log  $1/R$  data vs. wavelength for the first three components on the straw samples. Note that the first principal component weight plot has little structure and is virtually a linear slope. The second and third components have structure. Figures 8A–8C are the principal component plots derived from the SNV straw data. Now component number one has structure, and the underlying slope does not appear to contribute to the variance within the spectra.

Table II shows the collinearity between chemically unrelated spectral wavelengths at 1700 and 2100 nm and between chemically related wavelengths at 1420 and 1932 nm.

Collinearity at the water wavelengths is expected on a chemical basis and is most strong in the log  $1/R$  form. A proportion of the collinearity must still be attributable to particle size and scatter effects. The relationship is stronger in the straw spectra than the hay samples, which had been dried at 135°C. Drying at this high temperature results in very low correlation between spectral data at 1420 nm and the residual dry matter data for the sam-

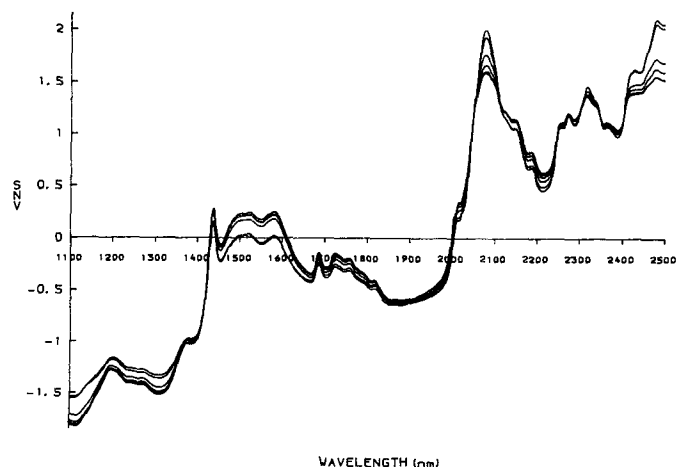


FIG. 3. Standard normal variate (SNV) spectra of sucrose.

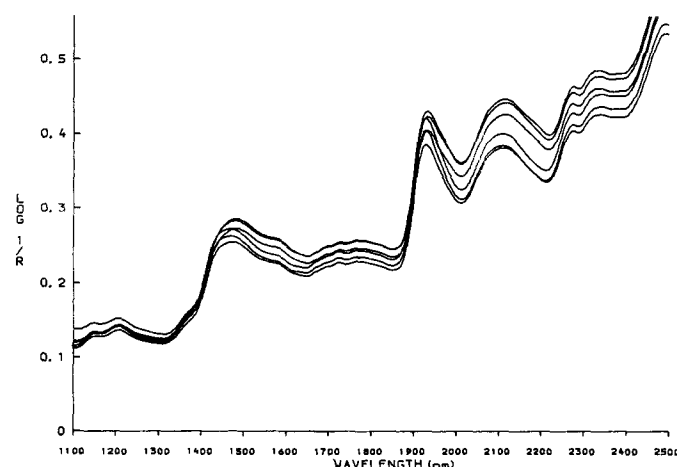


FIG. 5. Log  $1/R$  spectra of a series of straw.

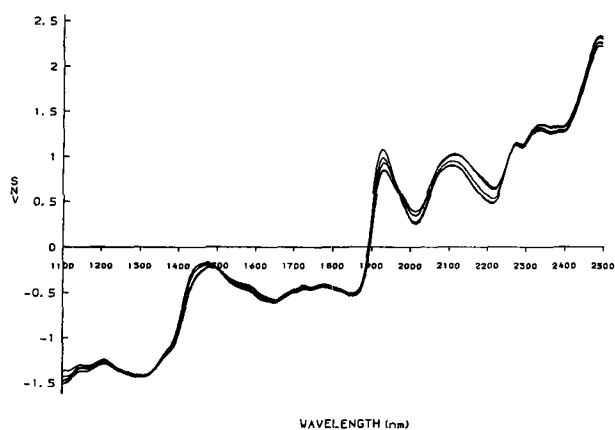


FIG. 6. Standard normal variate (SNV) spectra of straw.

ples. Figures 9A and 9B show the correlation between laboratory nitrogen analysis of the straw samples and the near-infrared reflectance values over the wavelength range considered for both the log 1/R and SNV transformed data. The enhanced correlation of the laboratory-determined nitrogen values to the SNV transformed data can be seen.

Slopes of the log 1/R and SNV spectra of the sucrose samples were calculated and examined for the presence of curvature. The slopes and intercepts of the log 1/R spectra were variable, whereas those of the SNV spectra were nearly constant. Overlaid SNV spectra did not match exactly, the larger differences in the overlays being between the finely powdered and crystalline sucrose spectra. Examination of both the log 1/R and SNV spectral slopes for linearity revealed significant curvature in the spectra of the finely powdered forms of sucrose (see Table III). This curvature was observed to varying extent in the straw and hay spectra.

Clearly, standardization of spectra using SNV achieves a scaling effect. Libraries of differing organic compounds can be observed on an equalized reflectance basis. Figures 10A and 10B show the log 1/R and the SNV spectra of cellulose, urea, glycine, and tri-palmitin. The slopes of the log 1/R spectra can be seen to be divergent and the reflectance ranges different for each compound, whereas the SNV spectra are all on an equalized basis.

As the major proportion of the variance in a log 1/R spectrum is due to sample presentation effects, then it follows that difference spectra will consist mainly of variance due to presentation effects. This effect will be even more apparent on samples that have undergone a physical or chemical change where it is required to produce a difference spectrum related to that change. Figures 11A

TABLE I. Percent variance accounted for by principal-component (PC) analysis of log 1/R and SNV spectra.

Samples	Math	% Variance				
		PC1	PC2	PC3	...	PC10
Hay	log 1/R	79.5	16.0	2.4		0.012
	SNV	65.6	21.0	10.5		0.062
Straw	log 1/R	93	4.5	1.4		0.003
	SNV	76.7	15.3	5.4		0.020

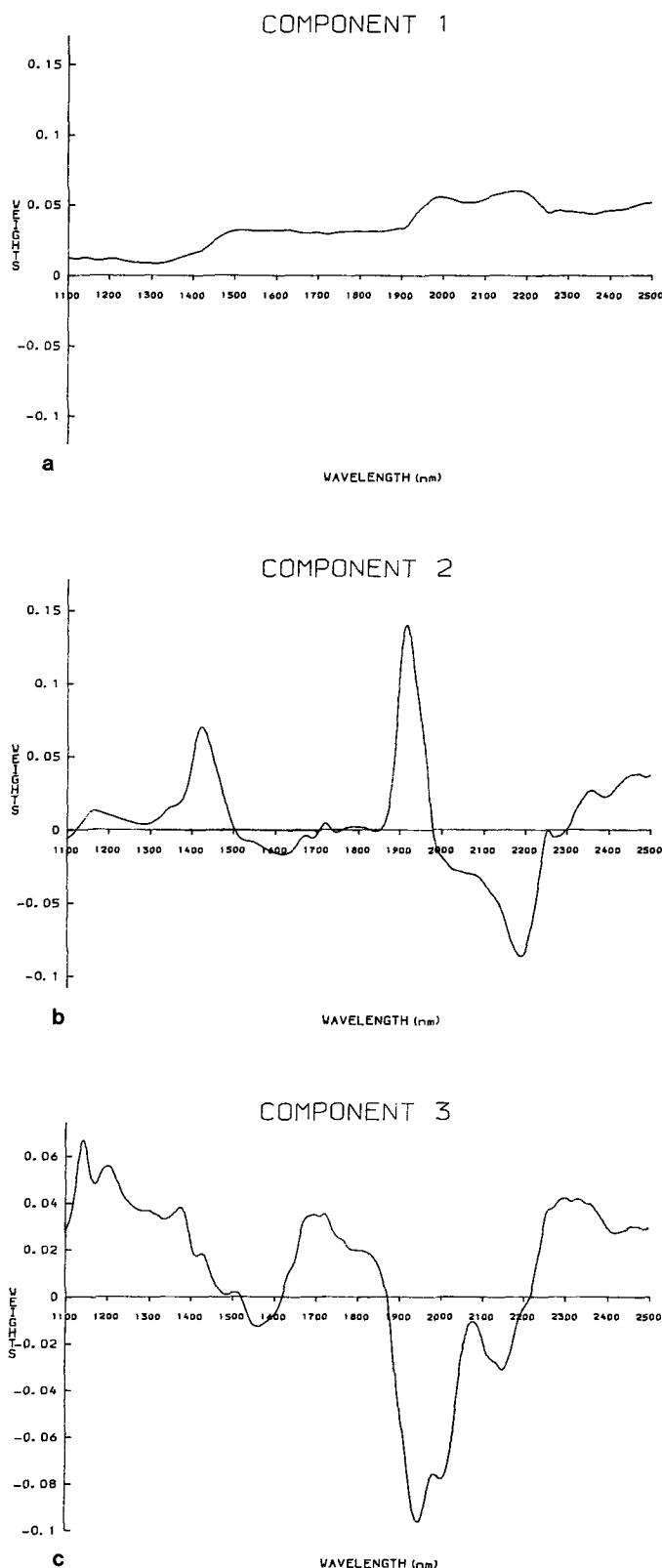


FIG. 7. Principal component weights of log 1/R NIR spectra of straw.

and 11B show the log 1/R and SNV difference spectra between the neutral detergent fiber (NDF)<sup>6</sup> and the acid detergent fiber (ADF)<sup>6</sup> fractions of four straw samples. The effects of curvature are apparent in the SNV difference spectra.

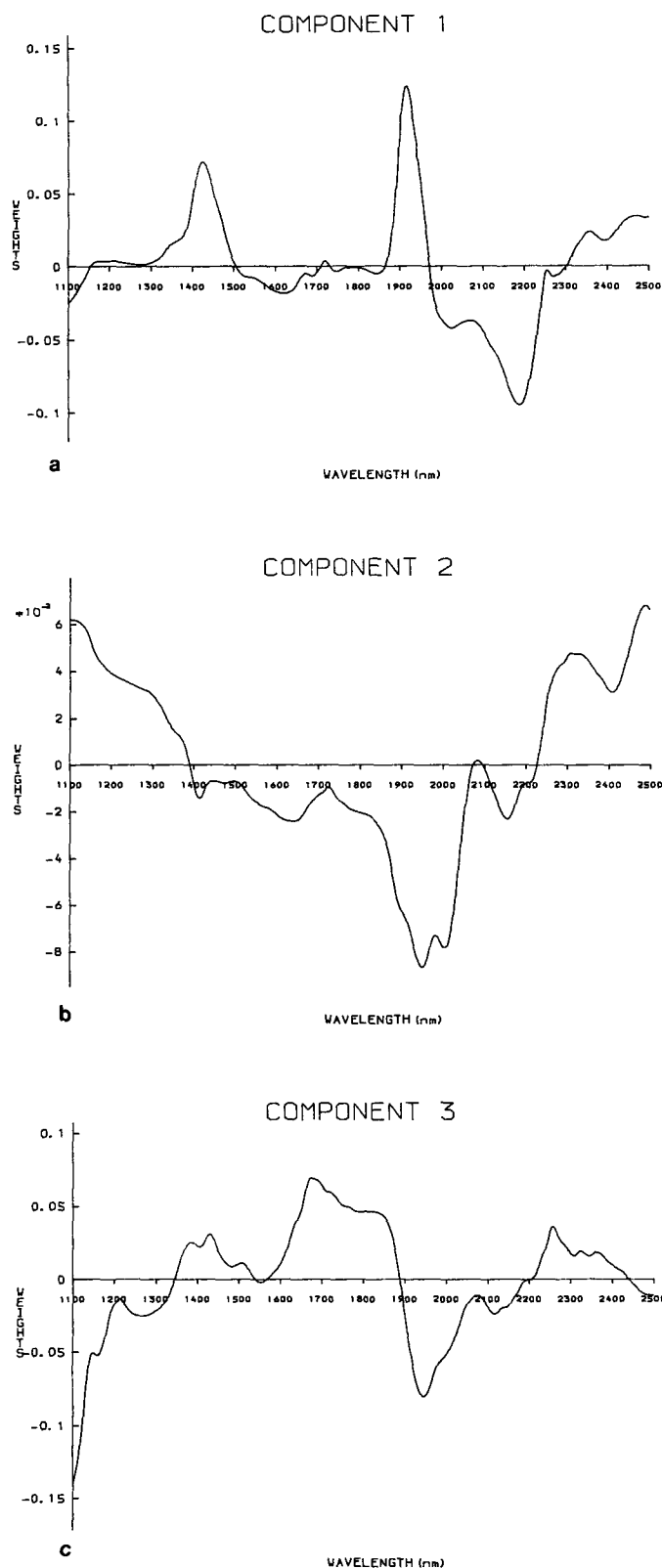


FIG. 8. Principal component weights of SNV spectra of straw.

## DE-TRENDING

The dominant feature of NIR diffuse reflectance spectra is the increasing level of the  $\log 1/R$  reflectance values over the range 1100 to 2500 nm. This trend is generally

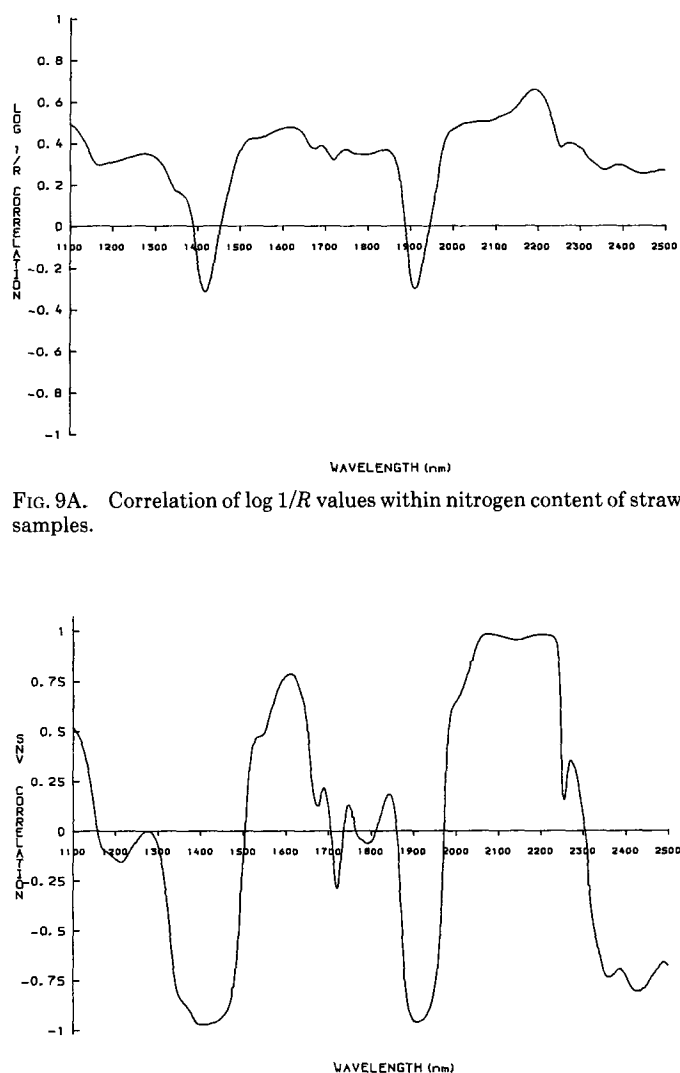


FIG. 9A. Correlation of  $\log 1/R$  values within nitrogen content of straw samples.

FIG. 9B. Correlation of SNV values with nitrogen content of straw samples.

linear, but it becomes curvilinear for the spectra of densely packed samples, e.g., powdered sucrose (see Table III). The linear slope of the SNV spectra of a given material is near constant, while the curvature varies with particle size and packing density. Most solids normally show absorbance spectra which are more intense at the longer wavelengths as the combination bands become more probable when the wavelength region approaches the fundamental vibration. A second-degree polynomial was

TABLE II. Collinearity between chemically unrelated and related wavelengths in NIR spectra.

Math	Wave-length 1 nm	Wave-length 2 nm	Straw $R^2$	Hay $R^2$
$\log 1/R$	1700	2100	0.937	0.965
D2 $\log 1/R$	1700	2100	0.371	0.059
SNV	1700	2100	0.010	0.029
$\log 1/R$	1420	1932	0.945	0.890
D2 $\log 1/R$	1420	1932	0.759	0.593
SNV	1420	1932	0.878	0.568

TABLE III. Curvilinearity in NIR spectra.

Spectrum	Maths	Linear slope	Linear intercept	Quad t	Variance	
					Linear fit	Curve fit
Sucrose 1 <sup>a</sup>	log 1/R	0.000865	-0.655	ns	77.5	77.5
2 <sup>a</sup>		0.000830	-0.635	ns	77.7	77.7
3 <sup>a</sup>		0.000802	-0.646	ns	77.5	77.5
4 <sup>a</sup>		0.000838	-0.624	ns	78.0	78.0
5 <sup>b</sup>		0.000292	-0.255	5.81	77.0	78.1
6 <sup>b</sup>		0.000336	-0.283	5.46	78.5	79.4
Sucrose 1 <sup>a</sup>	SNV	0.002178	-3.92	ns	77.5	77.5
2 <sup>a</sup>		0.002181	-3.92	ns	77.7	77.7
3 <sup>a</sup>		0.002177	-3.92	ns	77.5	77.5
4 <sup>a</sup>		0.002184	-3.93	ns	78.0	78.0
5 <sup>b</sup>		0.002171	-3.91	5.81	77.0	78.1
6 <sup>b</sup>		0.002193	-3.94	5.46	78.5	79.4
Straw 1	SNV	0.002339	-4.20	8.82	89.4	90.5
2		0.002343	-4.21	7.77	89.7	90.5
3		0.002349	-4.22	9.02	90.2	91.2
4		0.002364	-4.25	6.58	91.4	91.9
5		0.002364	-4.25	8.37	91.4	92.2
6		0.002360	-4.24	9.63	91.1	92.2
Hay 1	SNV	0.002361	-4.25	6.32	91.1	91.6
2		0.002396	-4.31	9.80	93.9	94.7
3		0.002383	-4.29	10.61	92.9	93.8
4		0.002357	-4.24	8.40	90.8	91.7
5		0.002371	-4.27	5.39	92.0	92.3
6		0.002398	-4.31	7.11	94.0	94.4

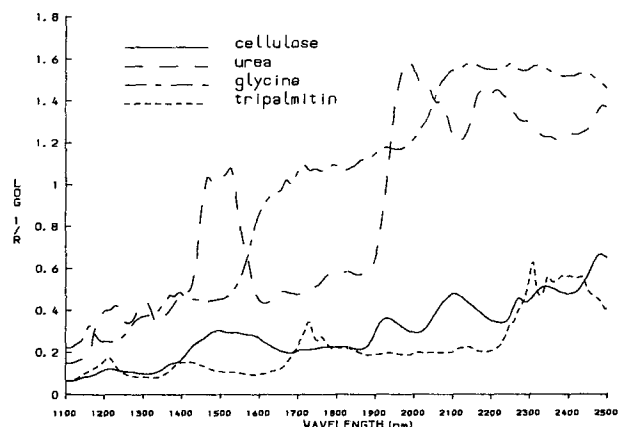
<sup>a</sup> Crystalline.<sup>b</sup> Powdered.

FIG. 10A. Log 1/R NIR spectra of various library compounds.

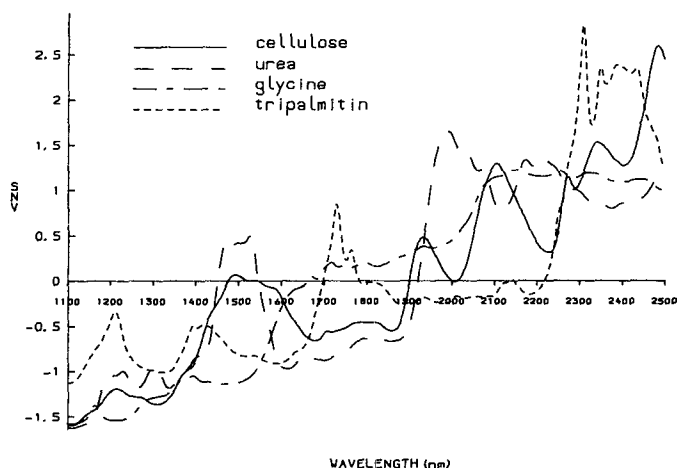


FIG. 10B. SNV NIR spectra of various library compounds.

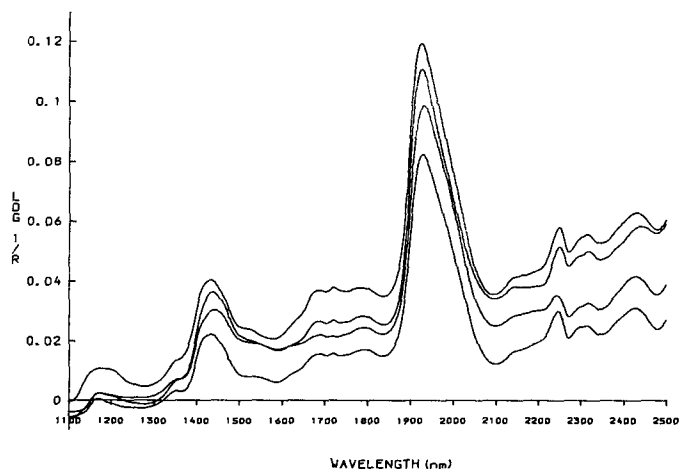


FIG. 11A. Log 1/R difference NIR spectra. Straw NDF minus straw ADF.

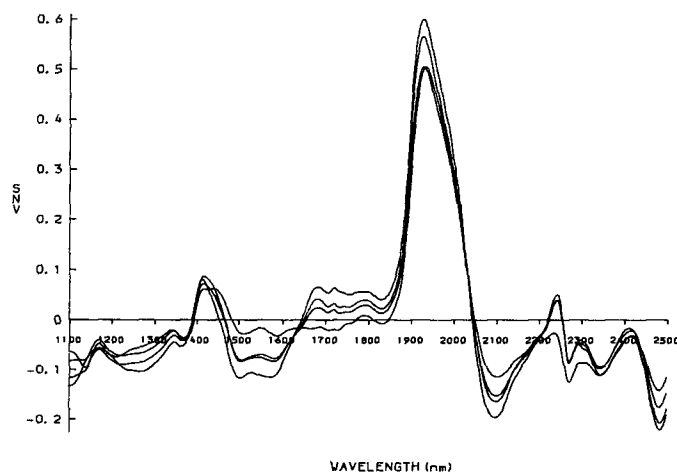


FIG. 11B. SNV difference NIR spectra. Straw NDF minus straw ADF.

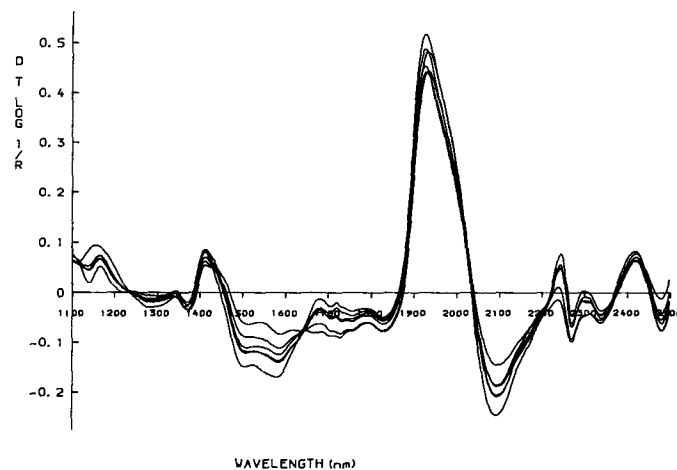


FIG. 11C. SNV and de-trended NIR difference spectra of straw NDF minus straw ADF.

used to standardize the variation in curvilinearity and proved to be an adequate model for de-trending of NIR spectra.

De-trending includes SNV transformation, but if there is interest in the shape differences in spectra then de-trending may be carried out without SNV. Figure 12

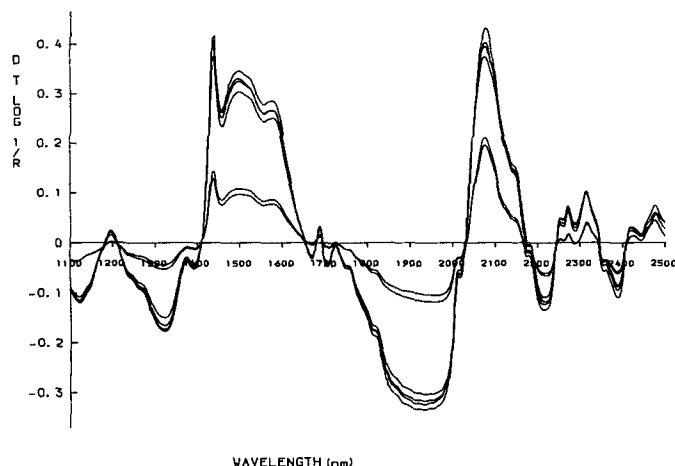


FIG. 12. De-trended NIR spectra of sucrose. No SNV transformation.

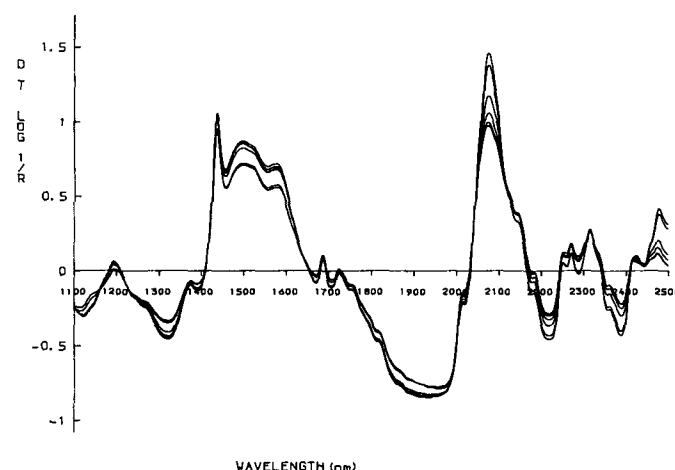


FIG. 13. De-trended NIR spectra of sucrose. With SNV transformation.

shows the set of sucrose spectra after de-trending without the use of SNV. The curvilinear trend is removed, but the distinction between crystalline and powdered forms remains.

When the underlying curvature trend is removed from the difference spectra of the neutral and acid detergent fiber residues of straw, the de-trended difference spectra appear on an apparently flat baseline (Fig. 11C).

De-trended sucrose spectra are shown in Fig. 13, and de-trended straw spectra in Fig. 14. The powdered forms of sucrose have the same absorbance bands as the crystalline form, but their extent varies. For the crystalline and powdered forms of sucrose, indications are that in a log  $1/R$  calibration situation the two forms could be regarded as separate sample populations, whereas application of SNV or DT transformations would bring both forms into the same population.

Linear regression was applied to the laboratory nitrogen data of the whole straw samples, against the log  $1/R$ , SNV, and de-trended NIR spectra. The regression statistics are given in Table IV.

Both SNV and DT spectra achieve calibration correlation statistics superior to those obtained with log  $1/R$

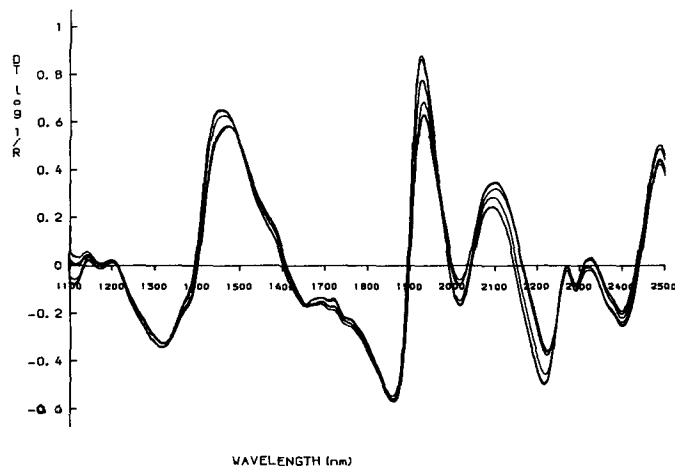


FIG. 14. De-trended NIR spectra of straw. With SNV transformation.

TABLE IV. Calibration regression statistics for log  $1/R$ , derivative, SNV, and DT transforms.

Spectra	Maths	Mean % N	SD	n	$R^2$	SEC
log $1/R$	0	1.044	0.448	48	0.423	0.340
	D1 <sup>a</sup>				0.957	0.093
	D2 <sup>b</sup>				0.940	0.109
SNV	0	1.044	0.448	48	0.964	0.085
	D1 <sup>a</sup>				0.957	0.092
	D2 <sup>b</sup>				0.955	0.096
De-trend	0	1.044	0.448	48	0.960	0.089
	D1 <sup>a</sup>				0.963	0.087
	D2 <sup>b</sup>				0.955	0.095

<sup>a</sup> First derivative.

<sup>b</sup> Second derivative.

spectra, and are equal to or better than those for the first- or second-derivative transformations of log  $1/R$  spectra. Derivatization of SNV and/or DT produces calibration statistics which are better than those from derivatized log  $1/R$  spectra, but are not significantly better than SNV and/or DT transformations.

#### ACKNOWLEDGMENTS

The authors wish to thank Ian Cowe and J. McNicol of the Scottish Crops Research Institute for the use of their principal-component software and helpful discussion; Dr. V. C. Mason of the AFRC Institute for Grassland and Animal Production for making the straw data available; and Prof. H. Nursten of the University of Reading for a useful discussion. This work was partly funded by the Ministry for Agriculture, Fisheries, and Food.

1. I. A. Cowe and J. W. McNicol, *Appl. Spectrosc.* **39**, 257 (1985).
2. I. Murray and P. Hall, *Anal. Proc.* **20**, 75 (1983).
3. H. Martens, S. A. Jensen, and P. Geladi, in *Proceedings of the Nordic Symposium on Applied Statistics* (Stokkand Forlag, Stavanger, Norway, 1983), pp. 235-268.
4. I. A. Cowe, J. W. McNicol, and D. Clifford Cuthbertson, *Analyst* **110**, 1227 (1985).
5. M. G. Kendall and Stewart, in *The Advanced Theory of Statistics*, (Charles Griffin and Company Limited London, 1977), Vol. 1, p. 49.
6. P. J. Van Soest and R. H. Wine, *J. Assoc. Official Anal. Chem.* **50**, 50 (1967).