

# **Comparison of multivariate calibration methods for prediction of feeding value by near infrared reflectance spectroscopy**

P. W. GOEDHART

Agricultural Mathematics Group, P.O.Box 100, NL 6700 AC Wageningen, Netherlands /  
Research Institute for Livestock Feeding and Nutrition, P.O.Box 160, NL 8200 AD Lelystad,  
Netherlands

Received 20 November 1989; accepted 8 March 1990

## **Abstract**

Routine prediction of feeding value of feeds by *in vitro* methods is time-consuming and expensive. Recently, near infrared spectroscopy (NIR) measurements are used to predict the feeding value. NIR instruments measure absorbance of near infrared light of a sample at several wavelengths. The spectrum of absorbance measurements contains indirect, non-specific information about the feeding value of the sample and can be used to predict this value. To this end a linear calibration model was estimated from experimental data. The model was used to predict unknown *in vitro* values with measured spectra in future samples. Multicollinearity in the NIR measurements occurs frequently. Inclusion of all absorbances in a calibration model gives overparameterisation and large prediction errors. To overcome multicollinearity several methods were proposed. The methods were described and employed to data in which *in vitro* digestibility of organic matter of maize for cattle was predicted by means of absorbances measured at 351 wavelengths. Comparison of methods showed that for these data Partial Least Squares is the best method. Multiplicative scatter correction of the spectra prior to estimation gave better predictions for all methods.

*Keywords:* calibration, near infrared spectroscopy, digestible organic matter, multiple linear regression, principal components regression, partial least squares, generalized least squares, best linear prediction

## **Introduction**

For adequate feeding of livestock, farmers need information about the feeding value of available feedstuffs. Knowledge of the content of digestible organic matter ( $d_{OM}$ ) is very important for energy evaluation of feeds, in particular for forages. Accordingly there is great practical interest in a rapid, accurate and cheap method to predict  $d_{OM}$  for forages. This interest is evidenced by the over 80 000 samples of forages a year that are analysed by the Routine Laboratories for Forage Estimation in the Netherlands.

Until recently the chemical composition (crude fibre, crude protein, crude ash) of these samples was routinely determined by classical chemical methods and this composition was used to predict  $d_{OM}$ . This method relies on pre-established relationships between  $d_{OM}$  and chemical composition. Criticism of this method focused on using crude fibre as a predictor (van Es, 1986). An alternative method to estimate  $d_{OM}$  was developed by Tilly & Terry (1963) who used an in vitro method with rumen fluid and pepsin. This method can be impractical as it requires regular supply of rumen fluid. Therefore replacement of rumen fluid by commercially available enzymes was studied, resulting in a new in vitro method which uses cellulase (van der Meer, 1988). However, both in vitro methods are laborious and are considered too slow to be used for routine analysis of large numbers of forage samples.

Almost parallel to the new in vitro methods a physically based method for determining chemical composition was introduced. This method is based on absorption of near infrared light (NIR) at wavelengths in the region 1100-2500 nm. Since chemical composition determines the amount of absorbed near infrared light, a NIR spectrum contains information on this composition. The information in a NIR spectrum is however non-selective because of interference of strongly overlapping constituents and because of light scatter variation. Interpretation of differences in spectra is therefore not straightforward. Hence, calibration of a NIR instrument has to be done indirectly. Calibration equations are derived by relating spectral information to chemical information for a large number of samples. In most cases the statistical technique that is used to derive prediction equations, is some form of variable selection in which a few absorbances are selected to predict the parameter of interest. There are however a number of statistical methods which employ all of the information in the spectrum.

The object of this study is to compare statistical methods for prediction of  $d_{OM}$  of forage maize for ruminants feeding by means of NIR spectroscopy. This study was part of a larger scientific program, carried out at the Research Institute of Livestock Feeding and Nutrition, which resulted in the introduction in 1989 of NIR spectroscopy as a means for routinely predicting  $d_{OM}$  of forage maize in the Netherlands.

## **Material and methods**

### *Samples*

A total of 125 samples of maize products was selected from a large group of samples, gathered at different locations in different years and consisting of whole maize plants (fresh or ensiled), as well as parts of maize plants (ears or stems). The selection was done in such a way that almost all variation in age, dry matter, ash, crude protein and crude fibre was represented. The samples were thus not randomly drawn from some well defined population. The set of 125 samples was used to derive calibration equations and will be referred to as the calibration set. A wide calibration set is important because it results in a general prediction equation with small bias. Very wide calibration sets can however increase the standard error of predic-

tion (Wetherill & Murray, 1987).

Furthermore, a set of 96 forage maize samples was collected at random in 1987. This set was used to evaluate the calibration equations and will be referred to as the evaluation set. The evaluation set was less diverse than the calibration set since the evaluation set was drawn at random and the calibration set was specifically selected to cover a wide range of variation.

### *Chemical analysis*

The content of digestible organic matter was estimated by an *in vitro* method which uses cellulase. Samples were ground and 0.5 gram was incubated in a pepsin-HCl solution at 39 °C for 24 hours and then placed in a waterbath at 80 °C during 30 minutes. After washing and filtering they were incubated in a cellulase-amylase buffer solution, again at 39 °C for 24 hours. After filtration samples were dried and incinerated. Digestible organic matter was calculated as the dry matter loss after incineration, expressed as a percentage of the initial dry matter. Together with each series of samples a number of standard samples of similar material with known *in vitro*  $d_{OM}$  values was analysed. From the results of these samples a regression formula was calculated which was used to predict  $d_{OM}$  for the unknown samples. The method has been described fully by van der Meer (1988).

The results of the chemical analyses are given in Table 1.

### *NIR measurements*

NIR spectra of samples were recorded with a Technicon InfraAlyzer 500C spectrophotometer between 1100 and 2500 nm, at 4 nm intervals. So, absorbances (spec-

Table 1. Crude protein, crude ash, crude fibre and *in vitro* digestible organic matter of forage maize samples.

Component	Range	Mean	Standard deviation
<i>Calibration samples (n = 125)</i>			
Crude ash (g kg <sup>-1</sup> )	13 -239	69	34
Crude protein (g kg <sup>-1</sup> )	57 -119	86	13
Crude fibre (g kg <sup>-1</sup> )	54 -347	214	54
Digestible organic matter (%)	45.7- 78.1	59.1	6.3
<i>Evaluation samples (n = 90)</i>			
Crude ash (g kg <sup>-1</sup> )	30 - 77	49	7
Crude protein (g kg <sup>-1</sup> )	71 -101	86	5
Crude fibre (g kg <sup>-1</sup> )	170 -263	213	20
Digestible organic matter (%)	57.9- 67.6	63.1	2.1

tral  $\log(I/R)$  values) measured at 351 wavelengths were available for every sample. Each spectrum was obtained by a single scan.

With 351 variables some of the calibration methods can not be used due to computational problems. This is especially so for the method of all combinations search multiple linear regression. Other methods require large amounts of computing time with 351 variables or give very poor numerical results due to rounding errors. Therefore 18 wavelengths, covering the whole range between 1100 and 2500 nm, were selected to compare all calibration methods. Some of these 18 wavelengths were selected because it was known from prior work that they were informative. The selection was thus done on a rather ad-hoc basis. Some methods were also compared for 351 wavelengths.

### *Transformations*

It is well known that specular reflection, particle size and structure of a sample can cause an overall shift in a spectrum unrelated to chemical composition. Consequently, variation in spectra due to differences in chemical composition can be snowed under by variation due to differences in structure. To eliminate these effects a Multiplicative Scatter Correction (MSC), as proposed by Geladi et al. (1985), was employed. For every sample a linear regression of the spectrum upon the average spectrum in the calibration set was calculated according to the model:

$$y_j = \alpha + \beta x_j + \epsilon_j$$

in which  $y_j$  is the absorption of the individual sample and  $x_j$  is the average absorption in the calibration set, both measured at wavelength  $j$ . In this model chemical information is in the residuals  $\epsilon_j$  and information on light scatter effects, particle size and structure is in the parameters  $\alpha$  and  $\beta$ . The latter information is eliminated by the transformation:

$$z_j = (y_j - \alpha)/\beta$$

in which  $z_j$  is the transformed spectrum and  $\alpha$  and  $\beta$  are replaced by estimates. This correction is successfully used by Naes et al. (1986). A similar transformation is used by Robert et al. (1986).

After the MSC, spectral data were centered and scaled. Centering is convenient but not essential. For most calibration techniques, non-scaled data would increase the relative importance of absorbances having a large variance. In order to give every absorbance an equal weight, spectral data were scaled.

### *Statistical methods for multivariate calibration*

The major difficulty with multivariate calibration of NIR instruments is the problem of multicollinearity which occurs because of very high intercorrelation between absorbances. This causes very large variances of estimated regression coeffi-

cients, leading to unstable estimation of the regression equation. To overcome this problem, various approaches have been proposed, all of which try to reduce the number of parameters to be estimated. Usually the data are described better by including more parameters in a model to minimize model errors. However, at the same time, the precision with which parameters are estimated decreases and so estimation errors increase. The error of prediction depends both on model errors and estimation errors and these have to be balanced against each other. This is illustrated in Figure 1.

All proposed methods are based on linear regression, see e.g. Montgomery & Peck (1982). So called inverse calibration methods are based on regressing  $d_{OM}$  upon a few absorbances or upon a few variables derived from the absorbances. In this paper three inverse calibration methods are used: All combinations search multiple linear regression (ALL), Principal Components Regression (PCR) and Partial Least Squares (PLS). So called classical calibration methods are based on multivariate regression of absorbances upon  $d_{OM}$ . Both classical methods that are used in this paper employ an approximation of the residual variance-covariance matrix of

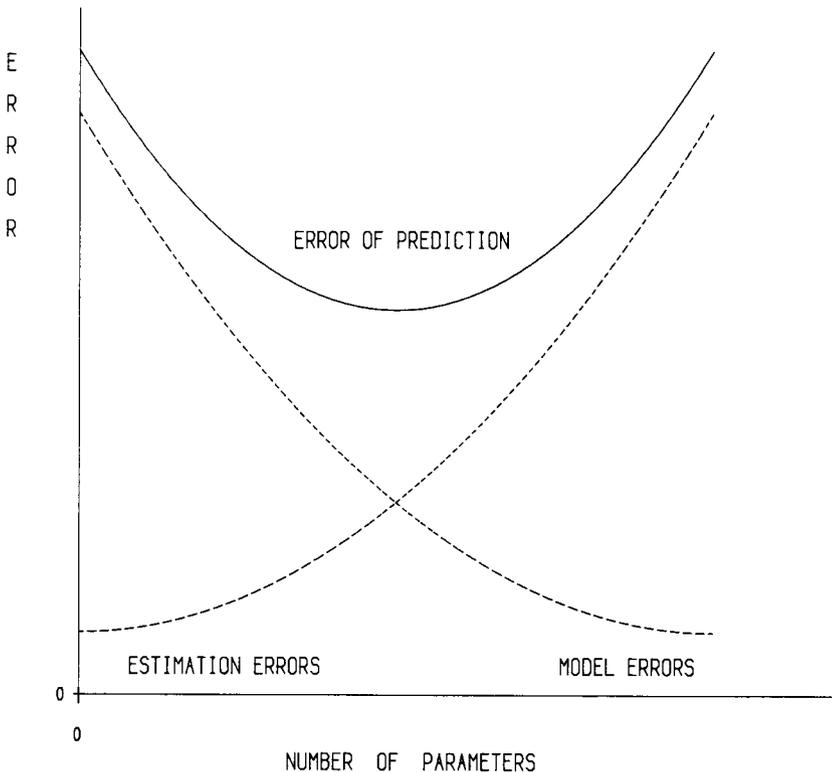


Fig. 1. Dependence of the error of prediction on both model errors and estimation errors.

the absorbances. In their original form, without this approximation, these methods are known as Generalized Least Squares (GLS) and Best Linear Prediction (BLP).

In this paper the description of the methods is mainly intuitive. References can be consulted for precise formulation of statistical models and for computational techniques.

#### *All combinations search multiple linear regression*

All combinations search multiple linear regression (ALL) essentially fits all possible models that can be obtained from the full model (with all absorbances) by deleting one or more absorbances. With 18 absorbances 262144 different models can be fitted, although intelligent numerical procedures only fit a fraction of this number. Which of these models is the 'best' is not very clear cut. Most criteria that are proposed, see e.g. Draper & Smith (1981), have in common that with a fixed number of variables in the model, a model is better when its residual variance is smaller or, equivalently, its coefficient of determination  $R^2$  is larger. Here for every fixed number of absorbances in the model the three models with the largest  $R^2$  are selected.

Other variable selection methods such as forward selection or backward elimination can be used as well. Berk (1978) constructed an artificial example with 4 variables, in which forward selection and backward elimination selected the same subset of 2 variables, missing a subset of the same size which gave a residual variance equal to 1/90th of that for the selected subset. These methods can thus fail to find the best model and therefore ALL is the preferred method.

The use of any selection method which uses the data, implies that the standard least squares properties are not applicable because they only hold if the subset of variables has been selected without reference to the data. Hence, the residual variance will be underestimated by the residual mean square of the selected regression. It is therefore important to test the regression equation on an independent set of data.

#### *Principal components regression*

Principal Components Regression (PCR) uses most of the available information in the spectra by deriving new variables from the absorbances. Then  $d_{OM}$  is regressed on these new variables, which are called principal components. The principal components are linear combinations of the absorbances. The first principal component is that linear combination that describes most of the variation in the spectra. The second linear combination, constrained to be orthogonal to the first, describes most of the remaining variation and so on. It is hoped that all the relevant information in the spectra is represented in the first few principal components.

Since the first principal components contain most of the variation in the spectra it seems appropriate to regress  $d_{OM}$  on the first components. However, as noted by Jolliffe (1982), the first components are not necessarily good predictors. Therefore two approaches were followed here. One in which the first components were used and one in which method ALL was applied to the first 10 components. For computational methods and related techniques see Jolliffe (1986).

*Partial least squares*

PCR is essentially a two-stage method in which the data compression stage is done independently of the regression stage, such that the principal components account for most of the variation in the NIR spectra alone. In Partial Least Squares (PLS) these two stages are combined. PLS forms linear combinations that potentially have high covariance with  $d_{OM}$  and thus have large predictive power. The first PLS-component is a linear combination of absorbances in which every absorbance is weighted by its covariance with  $d_{OM}$ . So absorbances with high covariance with  $d_{OM}$  play an important role in the first PLS-component and absorbances with low covariance play a small role. In the next step a regression is performed of  $d_{OM}$  on the first PLS-component which gives residuals of non-explained variation in  $d_{OM}$ . The second PLS-component again is a sum of absorbances, but now weighted by the covariances between residuals and absorbances. Regression of residuals on the second component gives new residuals and in an analogous way new components are formed. In this way the variation in  $d_{OM}$  is peeled off by regressing on successively new components, formed to have high covariance with  $d_{OM}$ .

PLS is computationally much simpler than PCR. There are a number of equivalent algorithms for PLS. Naes et al. (1986) give two algorithms, one with, the other without orthogonal components. Wold et al. (1984) show that PLS essentially is a conjugate gradient algorithm for linear least squares problems and such algorithms are reviewed by Paige & Saunders (1982). Here the algorithm from Naes et al. (1986) with orthogonal components is used. Theoretical properties of PLS can be found in Höskuldsson (1988).

*Generalized least squares*

Generalized Least Squares (GLS) performs a multivariate regression of the spectrum on  $d_{OM}$  (Naes, 1985a). This corresponds to simple univariate regressions of each absorbance on  $d_{OM}$ . Along with the regression coefficients, the inverse of the variance-covariance matrix of the residuals of the univariate regressions also enters the prediction equation. The covariance matrix is however nearly singular for NIR measurements, leading to unstable predictions. Naes (1985b) proposes a linear factor model in order to give a parsimonious description of the covariance matrix, the rationale being that the number of independent variation sources in the spectrum is smaller than the number of absorbances. Naes (1985b) claims that the prediction error benefits from the linear factor model.

*Best linear prediction*

Best linear prediction (BLP) is based on the same model as GLS with the same approximation of the residual variance-covariance matrix. The only difference is that, in order to improve the predictor, prior information about the variable which needs to be predicted is incorporated. If it is known that a sample is drawn at random from a population with known distribution of  $d_{OM}$ , then the GLS predictor can be improved (Naes, 1985b). This clearly needs an accurate definition of the population under consideration and knowledge of the distribution of  $d_{OM}$  within the population. Often the calibration set is considered as a random sample from the population

and the mean and variance of  $d_{OM}$  in the population are estimated from this sample. This procedure was also followed here, although it is clear that the calibration set was in fact not a random sample.

## Results

Calibration methods were compared in two ways. Firstly methods were compared in the way they describe the calibration set. Secondly the resulting prediction equations were employed for the 96 samples in the evaluation set and the predicted  $d_{OM}$  was compared with the measured  $d_{OM}$  for these samples. The criterion that was used in both comparisons was the Root Mean Squared Error (RMSE) with

$$RMSE^2 = \frac{1}{n} \sum_{\text{set}} (d_{OM\text{predicted}} - d_{OM\text{measured}})^2$$

where  $n$  is the number of samples in the set. A rough prediction interval for every individual sample is given by  $d_{OM\text{predicted}} \pm 2 \times RMSE$ , where RMSE is calculated in the evaluation set.

A dimension has to be selected for every method. For ALL the number of variables, for PCR and PLS the number of components and for GLS and BLP the number of independent sources of variation has to be chosen. In this study the dimension was varied systematically. In practice it would be selected on the basis of the RMSE in an appropriate evaluation set because that is the most straightforward validation of a prediction equation.

Most calculations were done with the statistical package Genstat 5, Payne et al. (1987).

### *Description of the calibration set*

The RMSE in the calibration set for each of the methods and for dimensions up to 10 is given in Table 2. In this table, ALL-1, ALL-2 and ALL-3 are the best, the second best and the third best subset respectively, for every dimension. PCR-1 is based on regressing  $d_{OM}$  on the first principal components, while PCR-A selects the best components out of the first 10. The other methods are as described above.

From Table 2 it appears that ALL describes the calibration set best. From dimension 4 onwards there are very small differences between the 3 models selected by ALL. As expected, PLS describes the data better than both PCR methods, the PCR-A method being better than the PCR-1 method. For higher dimensions PLS and PCR-A are approximately equivalent to ALL with respect to RMSE. The classical calibration methods GLS and BLP describe the calibration set less well, especially for higher dimensions.

### *Prediction of the evaluation set*

The RMSE in the evaluation set is given in Table 3. From Table 3 it appears that, although the ALL models with the same number of variables describe the calibra-

CALIBRATION METHODS FOR NEAR INFRARED REFLECTANCE SPECTROSCOPY

Table 2. Root Mean Squared Error (RMSE) for samples in the calibration set for the different calibration methods and for dimensions up to 10; RMSE is calculated for the percentage of digestible organic matter of forage maize, using 18 wavelengths.

Dim.	ALL-1	ALL-2	ALL-3	PCR-1	PCR-A	PLS	GLS	BLP
1	3.95	4.09	4.28	4.60	4.60	4.21	5.20	4.80
2	2.97	3.12	3.65	4.40	3.83	3.66	4.00	3.67
3	2.79	2.86	2.86	3.59	3.54	3.37	3.77	3.33
4	2.54	2.59	2.66	3.58	3.26	2.86	3.23	2.97
5	2.41	2.46	2.49	3.39	2.98	2.75	3.50	3.24
6	2.25	2.28	2.33	3.10	2.75	2.53	3.01	2.80
7	2.21	2.22	2.22	2.93	2.55	2.40	2.95	2.76
8	2.13	2.15	2.16	2.76	2.36	2.24	2.93	2.77
9	2.11	2.12	2.12	2.76	2.35	2.17	2.75	2.60
10	2.10	2.10	2.10	2.34	2.34	2.09	2.76	2.59

Dim.: Number of variables in the model or dimension of linear factor model.  
 ALL-1: Best subset by all combinations search multiple regression.  
 ALL-2: Second best subset by all combinations search multiple regression.  
 ALL-3: Third best subset by all combinations search multiple regression.  
 PCR-1: Principal components regression with first components.  
 PCR-A: Principal components regression with best components out of first 10.  
 PLS: Partial least squares.  
 GLS: Generalized least squares with a linear factor model.  
 BLP: Best linear prediction with a linear factor model.

Table 3. Root Mean Squared Error (RMSE) for samples in the evaluation set for the different calibration methods and for dimensions up to 10; RMSE is calculated for the percentage of digestible organic matter of forage maize, using 18 wavelengths. The smallest RMSE is bold in every column.

Dim.	ALL-1	ALL-2	ALL-3	PCR-1	PCR-A	PLS	GLS	BLP
1	2.18	2.95	1.72	3.87	3.87	2.94	4.60	4.02
2	1.70	1.75	<b>1.52</b>	2.75	3.23	2.00	2.30	2.09
3	1.79	<b>1.71</b>	1.82	2.04	3.63	1.89	2.98	2.12
4	<b>1.50</b>	1.85	1.67	1.94	3.94	1.50	<b>1.88</b>	<b>1.77</b>
5	1.60	1.90	1.80	1.75	2.64	1.72	1.97	1.88
6	1.84	1.97	2.15	1.85	2.47	<b>1.47</b>	2.18	1.82
7	1.84	2.09	2.07	1.64	2.19	1.57	3.01	2.56
8	2.04	2.02	2.09	1.53	1.83	1.88	2.99	2.64
9	2.06	2.08	2.12	<b>1.52</b>	1.84	1.93	2.85	2.70
10	2.04	2.09	2.07	1.81	<b>1.81</b>	2.05	3.17	3.08

Dim.: Number of variables in the model or dimension of linear factor model.  
 ALL-1: Best subset by all combinations search multiple regression.  
 ALL-2: Second best subset by all combinations search multiple regression.  
 ALL-3: Third best subset by all combinations search multiple regression.  
 PCR-1: Principal components regression with first components.  
 PCR-A: Principal components regression with best components out of first 10.  
 PLS: Partial least squares.  
 GLS: Generalized least squares with a linear factor model.  
 BLP: Best linear prediction with a linear factor model.

tion set equally well, there is quite some variation in the RMSE for ALL models in the evaluation set, most prominently so for dimensions 5 and 6. This can be explained by the fact that ALL chooses among a large number of models and can thus accidentally choose a model with less predictive ability in the evaluation set. Accordingly, ALL is a data dependent technique and thus not very robust.

Furthermore, although PCR-A describes the calibration set better than PCR-1, the latter predicts the evaluation set better.

PLS seems to be preferable compared to PCR because it needs less components to reach a smaller RMSE. Compared to ALL however, PLS needs more variables in the model to reach a slightly lower RMSE.

The classical regression methods GLS and BLP both give larger RMSE than the inverse methods. For this set of data the inverse methods are clearly superior to the classical methods.

*PLS and PCR with all wavelengths*

Application of both PCR and PLS with all wavelengths is relatively easy. The RMSE in the evaluation set, using all 351 wavelengths, are given in Table 4. Also in Table 4 PLS with 18 wavelengths with and without Multiplicative Scatter Correction can be compared. The latter comparison shows that Multiplicative Scatter Correction improves the predictions considerably. This is not only true for MSC in combination with PLS, but also in combination with the other methods. Using MSC gives models with fewer parameters giving better predictions.

Table 4. Root Mean Squared Error (RMSE) for samples in the evaluation set for PLS with and without Multiplicative Scatter Correction (MSC) using 18 wavelengths, and for PCR and PLS using all 351 wavelengths; RMSE is calculated for the percentage of digestible organic matter of forage maize. The smallest RMSE is bold in every column.

Dimension	18 Wavelengths		351 Wavelengths		
	PLS, no MSC	PLS, MSC	PCR-1, MSC	PCR-A, MSC	PLS, MSC
1	4.43	2.94	3.49	3.49	2.89
2	3.35	2.00	2.72	2.46	2.05
3	2.11	1.89	1.95	2.00	1.83
4	1.93	1.50	2.00	2.19	1.65
5	1.95	1.72	1.65	1.76	<b>1.50</b>
6	2.01	<b>1.47</b>	1.78	2.21	1.67
7	1.83	1.57	1.61	2.11	1.82
8	1.70	1.88	1.61	<b>1.71</b>	1.91
9	<b>1.64</b>	1.93	<b>1.57</b>	1.73	2.02
10	1.74	2.05	1.75	1.75	1.88

Dimension: Number of variables in the model.

PCR-1: Principal components regression with first components.

PCR-A: Principal components regression with best components out of first 10.

PLS: Partial least squares.

With 351 wavelengths PLS is better than either form of PCR. Also with 351 wavelengths the path of the RMSE is very smooth for PLS, indicating robustness. Furthermore, there are hardly any differences between PLS with 18 and with 351 wavelengths. The 18 selected wavelengths are apparently representative.

## Discussion

For NIR applications subset selection is compared with PCR-1 by Robert et al. (1986), Bertrand et al. (1987) and Downey et al. (1987). They found that prediction equations based on principal components are as accurate as conventional multiple regression methods (among which ALL and forward selection). They also found that the choice of number of components does not seem to be very critical. Furthermore, the principal components can be used to identify samples which contribute significantly to each component. Also with PCR a single calibration equation can be generated in relatively short time.

Naes et al. (1986) compare PCR-1 with PLS for NIR calibration with only 6 wavelengths. They found that PLS is a serious competitor to PCR and they highly recommended PLS for practical use in NIR. They also noted that PLS lacks a firm theoretical basis.

For the set of data at hand, PLS gives better predictions with fewer components than PCR. PLS thus compresses the information in the spectra more successfully than PCR. Moreover, PLS is computationally simpler than PCR. Classical calibration methods GLS and BLP perform less well. ALL gives good prediction results with 18 wavelengths. With 351 wavelengths ALL is practically impossible and a prior selection of wavelengths has to be made. This requires knowledge of the problem at hand and it is always doubtful whether the right selection has been made. Furthermore, ALL is not very robust as it can find a model which accidentally describes the calibration set best, but is a bad predictor.

Partial Least Squares seems to be the method of choice. However, this conclusion is based on one set of data only. Currently calibration methods are compared for other material. Partial Least Squares is now used for routine prediction of DOM of forage maize in the Netherlands.

## Acknowledgements

It is a pleasure to record many helpful discussions on the subject of this paper with Arend Oosterhoorn, Herman Vedder, Pieter Vereijken and Hilko van der Voet.

## References

- Berk, K. N., 1978. Comparing subset regression procedures. *Technometrics* 20: 1-6.  
 Bertrand, D., M. Lila, V. Furtoss & P. Robert, 1987. Application of principal component analysis to the prediction of lucerne forage protein content and in vitro dry matter digestibility by NIR spectroscopy. *Journal of the Science of Food and Agriculture* 41: 299-307.

- Downey, G., P. Robert, D. Bertrand & M. F. Devaux, 1987. Near infra-red analysis of grass silage by principal component analysis of transformed reflectance data. *Journal of the Science of Food and Agriculture* 41: 219-229.
- Draper, N. & H. Smith, 1981. Applied regression analysis (2nd edition). Wiley, New York, 709 pp.
- Es, A. J. H. van, 1986. Energy values of feeds for livestock and their prediction. *Netherlands Journal of Agricultural Science* 34: 405-412.
- Geladi, P., D. MacDougall & H. Martens, 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy* 39: 491-500.
- Höskuldsson, A., 1988. PLS regression methods. *Journal of Chemometrics* 2: 211-228.
- Jolliffe, I. T., 1982. A note on the use of principal components in regression. *Applied Statistics* 31: 300-303.
- Jolliffe, I. T., 1986. Principal component analysis. Springer Verlag, New York, 271 pp.
- Meer, J. M. van der, 1988. Prediction of organic matter digestibility by the EC-cellulase method. In: Chr. Paul (Ed.), Forage quality analysis by near infrared reflectance spectroscopy. Commission of the European Communities, Luxembourg, Report VI/3137/88 EN (8828): 88-96.
- Montgomery, D. C. & E. A. Peck, 1982. Introduction to linear regression analysis. Wiley, New York, 504 pp.
- Naes, T., 1985a. Comparison of approaches to multivariate linear calibration. *Biometrical Journal* 27: 265-275.
- Naes, T., 1985b. Multivariate calibration when the error covariance matrix is structured. *Technometrics* 27: 301-311.
- Naes, T., C. Irgens & H. Martens, 1986. Comparison of linear statistical methods for calibration of NIR instruments. *Applied Statistics* 35: 195-206.
- Paige, C. C. & M. A. Saunders, 1982. A bidiagonalization algorithm for sparse linear equations and least squares problems. *ACM Transactions on Mathematical Software* 8: 43-71.
- Payne, R. W., P. W. Lane, A. E. Ainsley, K. E. Bickell, P. G. N. Digby, S. A. Harding, P. K. Leech, H. R. Simpson, A. D. Todd, P. J. Verrier & P. W. White, 1987. GENSTAT 5 Reference Manual. Oxford Science Press, Oxford, 749 pp.
- Robert, P., D. Bertrand & C. Demarquilly, 1986. Prediction of forage digestibility by principal component analysis of near infrared reflectance spectra. *Animal Feed Science and Technology* 16: 215-224.
- Tilley, J. M. A. & R. A. Terry, 1963. A two-stage technique for the in vitro digestion of forage crops. *Journal of the British Grassland Society* 18: 104-111.
- Wetherill, G. Z. & I. Murray, 1987. The spread of the calibration set in near-infrared reflectance spectroscopy. *Journal of Agricultural Science (Cambridge)* 109: 539-544.
- Wold, S., A. Ruhe, W. Wold & W. J. Dunn, 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal of Scientific and Statistical Computing* 5: 735-743.