Fresh forage analysis by near infrared spectroscopy

Pierre Dardenne, Richard Agneessens and Georges Sinnaeve
Station de Haute Belgique - CRAGx, 100, rue de Serpont, 6800 Libramont, Belgium.

Introduction

During the last decade the Station de Haute Belgique (Centre de Recherches Agronomiques de Gembloux) has developed robust calibrations for forages: fresh grass, hay, grass silage, fresh maize, maize silage etc.The parameters are those used to determine the feeding value: ash, protein, crude fibre and organic matter digestibility. By selecting each year new samples the databases become very wide by integrating all the possible sources of variation that can be expected: varieties, cutting dates, locations, climates, sample preparations etc. Thanks to the standardisation of the spectra between instruments, the data collected over several years and different instruments can be kept in the same calibration database. These calibrations have been developed on dried and ground samples. New hardware developments and specially the transport mechanism built by NIRSystems (Silver Spring, MD, USA) attached to a full scanning 6500 monochromator and the "coarse" cell developed by Infrasoft International (Port Mathilda, PA, USA) allow us to measure intact forage samples with a minimum of preparation: drying and grinding are avoided. But the main cost of a calibration development is still the wet chemistry. The idea is to use already existing robust equations on dried and ground samples to get the "reference" values which are used to compute the calibrations for the fresh material. This article describes the procedure on trials carried out in 1993 in dehydration plants in France to analyse fresh alfalfa.

Material and methods

Samples

The alfalfa fresh samples were collected during the 1993 and 1994 harvests in three plants located in the north of France. The Champagne–Ardennes region is the most important producer of dehydrated products with more than 1 MT whereas the total amount for the European Community reaches 4.5 MT. The days of sampling were chosen at the beginning and at the end of each cut to cover the whole variability of the product composition.

NIR measurements and sample handling

The alfalfa samples are directly collected from the trucks when they are weighed at the plant and brought in a van which is equipped with a NIRSystems 6500 spectrometer and the transport attachment. Three sub-samplings of approximately 100 g are used to fill the coarse cell (Infrasoft International Inc.). The spectrum of each sub-sample is the average of 32 scans obtained when the cell comes down once. The individual spectra are recorded on the PC's hard disk.

Directly after the near infrared (NIR) measurements, the three sub-samples are merged, weighed, labelled and placed in a refrigerator. In the evening, the samples of the day are dried in an air oven (48 h at 70°C) located in the laboratory. After the drying and the second weighing, the

samples are ground on a hammer mill (2 mm sieve) and afterwards on a cyclotec (1 mm sieve). The powders are then re-measured on another NIRSystems 5000 spectrometer equipped with an auto sampler. Only one spectrum is taken per sample.

Results and discussion

Sample selection

Among the 515 samples measured, 65 samples were selected on the basis on their spectral characteristic by means of the SELECT program of the ISI package. The average neighbourhood distance between the closest pairs was 0.91 for the 65 selected samples.

The 65 selected samples were analysed by the reference methods to obtain residual moisture, protein, crude fibre, ash and organic matter digestibility (enzymatic method). The 65 samples were merged with the existing database of fresh lucern already available and the results of the new calibration are reported in Table 1. The math treatments are SNV–Detrend[1] and a second derivative with a gap of 15 data points and a smoothing segment of five data points. The regression method is the modified PLS of ISI software using a cross-validation with four groups.[2] Table 2 shows the results of the prediction of the 65 samples re-predicted by these equations.

The chemical values of the remaining 450 samples are determined by using these equations and the predicted values are entered vs the averaged spectra obtained on the fresh material.

For each parameter, an automatic procedure, built by the option "Teach Automatic Sequence" of the ISI calibration routine, is applied to test 60 different math treatments. There are six pre-treatments of the spectra: 1: None, 2: SNV & Detrend,[1] 3: SNV, 4: Detrend, 5: MSC (Multivariate Scatter Correction) and 6: WMSC (Weighted MSC). Then, for each of these spectral correction, 10 derivatives are performed: 0-0-1, 0-0-5, 1-5-5, 1-10-5, 1-15-5, 1-20-5, 2-5-5, 2-10-5, 2-15-5 and 2-20-5 [the first number is the derivative order, the second one is the gap for the subtraction and the third one is the length of the segment for the smoothing (data points)]. The calibration program runs overnight to compute the 60 equations for each parameter using modified PLS algorithm and four group cross-validations with one outlier pass. Moreover, four wavelength ranges are tested: 1: 400–2500, 2: 800–2500, 3: 800–1850 and 2050–2500 and 4: 800–1850.

Among the 240 equations by parameter, the best equation is chosen by the evaluation of these models on the 65 selected samples not included in the calibration set. Table 3 gives the results of the calibrations made by NIR values and Table 4 reports the performances of the calibrations on

Table 1. Statistical results of the alfalfa calibrations of dried and ground alfalfa samples.

Variable	*N*	Min.	Max.	*SECV*	R^2CV
Dry Matter	282	90	96	0.37	0.86
Protein	287	14	27	0.56	0.96
Crude fibre	248	18	41	1.08	0.96
Ash	288	8	16	0.50	0.83
OMD	287	50	83	1.38	0.96

N: number of samples; Min.: minimum value; Max.: maximum value; *SECV*: standard error of cross-validation (four groups); R^2CV: coefficient of determination of cross-validation.

Table 2. Statistical results of the 65 selected samples predicted by the previous equations.

Variable	*N*	Min.	Max.	*SEP*	R^2P
Dry Matter	65	91	96	0.77	0.64
Protein	65	14	25	0.55	0.96
Crude fibre	65	18	41	1.24	0.96
Ash	65	8	16	0.82	0.77
OMD	65	50	83	2.49	0.94

Rem.: 65 samples included in the calibration set.

the 65 samples. The best math pre-treatment is “none” for all the parameters. It seems that the scatter correction algorithms (SNV, Detrend or MSC) are not suitable for spectra of high moisture samples like alfalfa. The influence of the large water bands is too important and disturbs the corrections. Only a second derivative with a large gap is efficient to improve the models.

We observe in the previous table that the slopes are different from one. An example is given in Figure 1 for the protein determination. The standard deviations of the predicted values are smaller than the lab values. This is due to the fact that the 450 samples show normal distribution (Figure 2) with many samples around the mean and consequently the R^2 of calibration are smaller than usual. The second reason is that the original calibrations on dried and ground alfalfa produce already predicted values with a standard deviation smaller than the standard deviation of the reference values, by definition of the least square linear regression, where R = SDyest/SDyref when the slope is equal to one and this is the case for the calibration data set.

To overcome this problem two methods have been used. i) The predicted NIR values coming from the “powder” calibrations are modified to make their standard deviations equal to the standard deviations of the laboratory values. We compute for each parameter the slope and the intercept by the least rectangle linear regression (Yref = a + b.Ynir) on the 288 dried and ground samples and we correct the 450 predicted values by applying these equations. As the slope is always higher than one, the variances of the predicted values are increased. iii) As the distribution of the 450 samples presents many samples around the mean, the number of samples for calibration

Table 3. Statistical results of the fresh alfalfa calibrations based on NIR reference values.

Variable	*N*	Min.	Max.	*SECV*	R^2CV	Math
Dry Matter (6)	439	11	44	0.85	0.98	D2,20,5
Protein (13)	447	13	24	0.83	0.78	D2, 5,5
Crude fibre (12)	445	18	43	1.61	0.84	D2,20,5
Ash (12)	442	9	15	0.43	0.79	D2,20,5
OMD (13)	443	52	83	1.94	0.83	D2,20,5

( ): number of PLS terms.

Table 4. Statistical results of the prediction of the 65 fresh selected samples.

Variable	*SEP*	Bias	*SEP*(c)	Slope	R^2P	*SDR*
Dry Matter	1.32	-.09	1.33	1.09	0.97	1.22
DM (64)	1.05	-.20	1.03	1.03	0.98	1.03
Protein	1.00	-.07	1.07	1.17	0.89	0.94
Crude fibre	2.44	0.42	2.42	1.12	0.87	2.34
CF(63)	1.99	0.44	1.94	1.13	0.96	1.81
Ash	1.11	0.26	1.09	1.25	0.58	1.07
OMD	3.54	-1.07	3.40	1.30	0.91	2.74

( ) Number of samples used: 1 deleted for DM, 2 deleted for CF.

is reduced by selecting first the 50 samples which are the most representative of the whole set by using the method of the neighbourhood distance included in the ISI package (Select program). Second, from the 400 remaining samples, the 75 smallest values and the 75 highest values are merged with the first 50 samples. This way, we ensure we cover the spectral variation and the chemical variation with a bimodal distribution (Figure 3) able to produce more robust calibrations with higher R2.

This method leads us to build a specific data set for each parameter from which new PLS models are computed. Table 5 reports the prediction results of the 65 samples (ref values vs NIR values) obtained by the models built from the 200 samples. We observe better *SEP*s than those reported in Table 4 and the slopes are very close to one. From these results, we can observe that it is possible to develop precise and robust calibrations with NIR results.

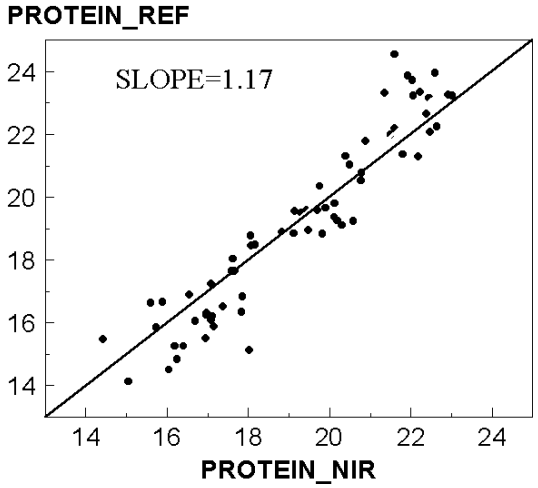

Figure 1. Comparison between reference and predicted protein values.

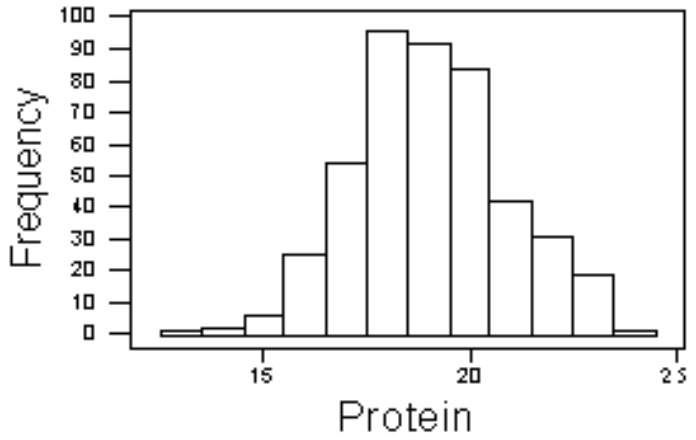

Figure 2. Frequency distribution of the 450 protein values predicted on the dried and ground samples.

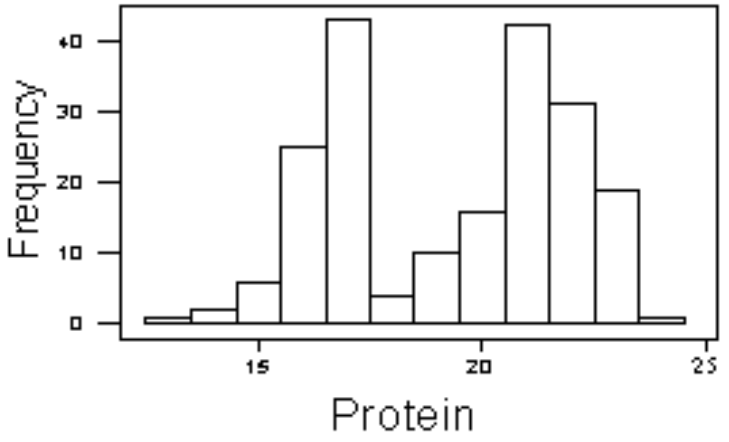

Figure 3. Frequency distribution of the 200 selected samples of the protein data set: 50 from spectral selection, 150 from chemical selection.

Table 5. Statistical results of the prediction of the 65 samples on the models built from specific data set of 200 samples with NIR corrected values.

Variable	*SEP*	Bias	*SEP*(c)	Slope	R^2P
Protein (13)	0.85	-.08	0.85	1.00	0.91
Crude fibre (11)	1.61	0.34	1.59	0.99	0.94
Ash (11)	1.10	0.14	1.09	0.92	0.55
OMD (13)	2.89	-0.99	2.74	1.02	0.92

( ): number of PLS terms.

Conclusion

The experiment shows that it is possible to develop NIR calibrations from reference values coming from NIR. The predicted values on the samples scanned after drying and grinding are used as reference values to compute models on the intact material. The calibration performances tested with some samples analysed by the original wet chemistry show that the procedure provides models as good as those we could have developed from actual laboratory values. The method can be expanded for any calibration transfer, when it is possible to measure a set of samples on two instruments: one already calibrated and the other one to be calibrated. Some precautions have to be taken concerning the spectral variation and the chemical distributions.

References

1. R.J. Barnes, M.S. Dhanoa and J.S. Lister, *Appl. Spectrosc.* **39(5),** 772 (1989).
2. J.S. Shenk "Routine Operation, Calibration and Network System Management Software for Near Infrared Instruments". ++NIRS3++, vers. 3.00, ISI software manual (1994).