

Influence of mathematical pre-treatments on calibrations for ground and unground compound feedingstuffs

V.M. Fernández-Cabanás,^a A. Garrido-Varo,^b M.D. Pérez-Marín^b and P. Dardenne^c

^a*Escuela Universitaria de Ingeniería Técnica Agrícola. University of Seville. Ctra. Utrera Km. 1. 41013 Sevilla. Spain.*

^b*Escuela Técnica Superior de Ingenieros Agrónomos y de Montes. University of Córdoba. Apdo. 3048. 14080 Córdoba. Spain.*

^c*Centre de Recherches Agronomiques de Gembloux. Chaussée de Namur, 24. 5030 Gembloux. Belgium.*

Introduction

NIR analysis is nowadays conceived as a synthesis of spectroscopy, mathematics, statistics and instrumentation, being the progress in this technology closely related to the advances produced in these fields.¹ Chemometrics procedures^{2,3} are usually applied to NIR spectra in order to extract relevant information, as the spectra are influenced by features like particle size, texture, light scattering and different sources of error.^{4,5} Precision in NIR analysis is subjected to the availability of algorithms or spectral data pre-treatments which have been developed to reduce the above mentioned effects, particularly the scatter effect, one of the most relevant factors affecting to the spectra of agro-food products.^{6,7}

At present time, there is a great proliferation of chemometrics software packages, which allow the application of several pre-treatments (derivatives, MSC, SNV, DT, OSC, etc.), despite the lack of a sound understanding of their theoretical basis or application requirements, to develop a NIR model for a given constituent and product.

This work was designed to study the effect of different combinations of spectral pre-treatments, on the performance of calibrations developed on intact and ground compound feedingstuffs.

Material and Methods

Samples and reference data

A total of 354 compound feedingstuffs samples for different animal species, and with reference data for crude protein (CP) and crude fibre (CF) were used in this study. Global set was randomly divided into a calibration and a validation set. Chemical composition for both sets is displayed in Table 1.

NIRS hardware

All the samples were scanned twice: intact and after grinding through a cyclone mill (1 mm screen).

Reflectance spectra for ground samples were obtained on a Foss NIRSystems 6500 SY-I monochromator, from 400 to 2498 nm, every 2 nm. Analysis was performed using a spinning module and samples were scanned on standard ring cells. Unground samples were scanned using a Foss NIRSystems 6500 SY-II monochromator, from 400 to 2498 nm, every 2 nm, provided with a transport module, and using a “*coarse cell*” cup.

Table 1. Composition of calibration and validation sets

	Total		Calibration		Validation	
	CP	CF	CP	CF	CP	CF
n	354	351	324	321	30	30
Mean	17.2	7.6	17.2	7.5	17.1	8.0
Minimum	11.9	1.4	11.9	1.4	12.7	3.2
Maximum	32.5	25.3	32.5	25.3	21.1	18.1
sd	2.28	4.4	2.3	4.3	2.1	4.9

Table 2. Combinations of pre-treatments used in calibration

Equation	Scatter correction	Derivative	Equation	Scatter correction	Derivative
1	None	None	29	SDT	1,10,10,1
2	SNV	None	30	SDT	2,5,5,1
3	DT	None	31	SDT	2,10,5,1
4	SDT	None	32	SDT	2,10,10,1
5	DTS	None	33	DTS	1,5,5,1
6	Standard MSC	None	34	DTS	1,10,5,1
7	Weighted MSC	None	35	DTS	1,10,10,1
8	Inverse MSC	None	36	DTS	2,5,5,1
9	None	1,5,5,1	37	DTS	2,10,5,1
10	None	1,10,5,1	38	DTS	2,10,10,1
11	None	1,10,10,1	39	Standard MSC	1,5,5,1
12	None	2,5,5,1	40	Standard MSC	1,10,5,1
13	None	2,10,5,1	41	Standard MSC	1,10,10,1
14	None	2,10,10,1	42	Standard MSC	2,5,5,1
15	SNV	1,5,5,1	43	Standard MSC	2,10,5,1
16	SNV	1,10,5,1	44	Standard MSC	2,10,10,1
17	SNV	1,10,10,1	45	Weighted MSC	1,5,5,1
18	SNV	2,5,5,1	46	Weighted MSC	1,10,5,1
19	SNV	2,10,5,1	47	Weighted MSC	1,10,10,1
20	SNV	2,10,10,1	48	Weighted MSC	2,5,5,1
21	DT	1,5,5,1	49	Weighted MSC	2,10,5,1
22	DT	1,10,5,1	50	Weighted MSC	2,10,10,1
23	DT	1,10,10,1	51	Inverse MSC	1,5,5,1
24	DT	2,5,5,1	52	Inverse MSC	1,10,5,1
25	DT	2,10,5,1	53	Inverse MSC	1,10,10,1
26	DT	2,10,10,1	54	Inverse MSC	2,5,5,1
27	SDT	1,5,5,1	55	Inverse MSC	2,10,5,1
28	SDT	1,10,5,1	56	Inverse MSC	2,10,10,1

NIRS software and chemometric treatments

All spectra were manipulated and processed, and all calibration equations were obtained, using ISI software NIRS3 ver. 4.0 and WINISI ver. 1.5 (Infrasoft International, Port Matilda, PA).

A total of 56 combination of pre-treatments (Table 2) were applied to obtain Modified Partial Least Squares (MPLS) calibration equations.^{8,9} The mathematical pre-treatments used were Standard Normal Variate (SNV), Detrending (DT),¹⁰ SNV and DT (SDT), DT and SNV (DTS) the three versions of Multiplicative Scatter Correction¹¹ included in ISI software (Normal MSC, Weighted MSC and Inverse MSC) and four different derivative math treatments. The derivative math treatments are referred to by a four-digit notation (a,b,c,d).¹² All calibrations were obtained for the spectral range 1100-2498 nm, for each chemical parameter, milling status, and maximum number of outliers elimination passes none, 2 or 9. Equations were obtained using an automatic routine ("Teach automatic sequence") included in ISI software.

NIRS equations were evaluated, by examining the statistical values obtained for calibration and validation procedures. Main calibration statistics are 1-VR (determination coefficient for cross validation) and SECV (standard error of cross validation), while validation was evaluated by SEP(C) (standard error of prediction corrected for bias) and R² (determination coefficient for validation).

Equation errors were compared using two test. First test sets, an arbitrary limit of 20% for significative differences with minimum error,¹³ and it is used for calibration and validation errors. Second test is based on a Fisher test,^{14,15} and defines a confidence interval (Equation 1) for errors with no significative differences with minimum error (Error_{min})

$$(Error_{min}, Error_{min} \sqrt{F_{1-\alpha, n-1, n-1}}) \quad (1)$$

where α is the significance level (5% in this study) and (n-1) the degrees of freedom. This test was used to compare validation errors.

Results and Discussion

The influence of the data pre-treatments used in calibration is displayed in Table 3, where it can be observed that remarkable differences appears between minimum and maximum SECV values. This effect is also evident with SEP(C) values, but in this case it can be observed that statistics for unground and ground products are more similar for each level of elimination passes.

Table 3: Maximum, minimum and 20% limit for SECV values for CP and CF calibrations with 0, 2 or 9 maximum outliers elimination passes. Grey background indicates significative differences found with minimum error.

	Elim. passes	Unground samples						Ground samples					
		0		2		9		0		2		9	
		SECV	eqa	SECV	eqa	SECV	eqa	SECV	eqa	SECV	eqa	SECV	eqa
CP	Minimum	1.22	44	0.88	19	0.61	32	0.91	17	0.61	49	0.52	44
	Maximum	1.82	1	0.98	7	0.92	37	0.98	14	0.70	1	0.70	1
	20% limit	1.46		1.06		0.73		1.09		0.73		0.62	
CF	Minimum	1.57	36	0.60	30	0.47	36	0.94	19	0.60	27	0.51	20
	Maximum	1.95	14	0.85	2	0.68	7	1.10	1	0.80	1	0.71	25
	20% limit	1.88		0.72		0.56		1.13		0.72		0.61	

Differences in SECV are significant for unground samples, but not for ground samples, unless a proper number of outliers elimination passes is applied. Significant differences are more evident for crude fibre than for crude protein.

A similar trend is found for SEP(C) values (Table 4), where no significant differences are found for CP in ground samples. In the case of errors for CF, in spite that differences found are significant for both criteria used, the error values are very close to maximum values.

For any error comparison, F limit appears as more restrictive than 20% limit.

Table 4: Maximum, minimum, 20% limit and F limit for SEP(C) values for CP and CF calibrations with 0, 2 or 9 maximum outliers elimination passes. Grey background indicates significant differences found with minimum error.

	Elim. passes	Unground samples						Ground samples					
		0		2		9		0		2		9	
		SEP(C)	eqa	SEP(C)	eqa	SEP(C)	eqa	SEP(C)	eqa	SEP(C)	eqa	SEP(C)	eqa
CP	Minimum	0.78	45	0.64	1	0.61	9	0.75	49	0.69	36	0.68	36
	Maximum	1.27	1	0.90	37	0.86	12	0.86	3	0.82	6	0.82	52
	F limit	1.06		0.87		0.83		1.02		0.94		0.92	
	20% limit	0.94		0.77		0.73		0.90		0.83		0.82	
CF	Minimum	0.85	24	0.78	21	0.74	11	0.83	48	0.74	36	0.75	48
	Maximum	1.33	21	0.95	17	1.02	2	1.14	3	1.09	3	1.07	14
	F limit	1.16		1.06		1.01		1.13		1.01		1.02	
	20% limit	1.02		0.94		0.89		1.00		0.89		0.90	

Conclusions

There is not a clear distribution pattern for results with the different data pre-treatments evaluated, but as scatter is the main source of spectral variability in this type of products, most of the scatter correction algorithms tested improved the predictive ability of the equations developed.

The best pre-treatments in calibration and in prediction usually differ. So that, when making evaluation of pre-treatments it would be desirable to use some automatic software routines that permit an efficient evaluation of different validation sets and/or the execution of a large number of cross validation passes.

Acknowledgements

This work was carried out using NIR hardware and software at the NIR/MIR Unit of the SCAI (University of Córdoba). The authors are grateful to SAPROGAL feed company for providing the samples and to the Spanish DGICYT for providing financial support for this study as part of the Project CICYT-FEDER IFD1997-0990. Thanks are also given to Mr. Antonio López, Mr. Alberto Sánchez de Puerta and Ms. Isabel Leiva for technical assistance.

References

1. A. Garrido, A. Gómez, J.E. Guerrero, y V. Fernández, in *Avances en Alimentación y Nutrición Animal*, Ed by C. De Blas, G.G. Mateos, P. and Rebollar, FEDNA. Madrid, España. p. 275 (1996).
2. D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte and L. Kaufman, *Chemometrics: A textbook*. Elsevier Science Pub., Amsterdam, The Netherlands. (1988).
3. P. Geladi and E. Dăbakk, E. J. *Near Infrared Spectrosc.* **3**,119 (1995)

4. P.C. Williams and K.H. Norris, In Near-Infrared technology in the agricultural and food industries, Ed by P.C. Williams and K.H. Norris, p. 171 (2001).
5. J.S. Shenk, J. Workman and M. Westerhaus, In Handbook of Near-Infrared Analysis, Ed by D.A. Burns and E.M. Ciurczsak, E.M., Practical Spectroscopy series. 13. Marcel Dekker, New York (USA). p. 383 (1992).
6. M.S. Dhanoa, S.J. Lister, R. Sanderson and R.J. Barnes, J. Near Infrared Spectrosc. **2**, 43 (1994).
7. D. Bertrand, in La spectroscopies infrarouge et ses applications analytiques, Ed by D. Bertrand, and E. Dufour, Technique & Documentation, p. 351 (2000).
8. J.S. Shenk, and M.O. Westerhaus, Crop. Sci. **31**, 1694 (1991).
9. H. Martens and T. Naes, Multivariate calibration. John Wiley & Sons. Chichester, UK. (1989).
10. R.J. Barnes, M.S. Dhanoa and S.J. Lister, Appl. Spectrosc. **43**, 772 (1989).
11. H. Martens, S.A. Jensen and P. Geladi,. Proceedings of the Nordic Symposium on Applied Statistics, Stokkand Forlag Publishers, Stavanger, Norway. p 205 (1983).
12. J.S. Shenk M.O. and Westerhaus, Routine operation, calibration, development and network system management manual, NIRSystems, Inc., 12101 Tech Road, Silver Spring. MD 20904. USA. (1995).
13. W.R. Windham, D.R. Mertens and F.E. Barton II, in Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality. Agriculture Handbook, N° 643, Ed by G.C. Martens, J.S. Shenk and F.E. Barton II, USDA-ARS. US Government Printing Office, Washintong, D.C. p. 96 (1989).
14. T. Fearn, NIR News. **7(5)**, 5 (1996).
15. Y. Roggo, L. Duponchel, B. Noe and J.P. Huvenne, J. Near Infrared Spectrosc. **10**, 137 (2002).