A meta-analysis approach for evaluating the accuracy of NIR calibration equations for forages

J. Fuentes-Pila[a], C. Pérez Hugalde[a], S. Martínez[b], I. Ferrero[b], B. de la Roza[c], V. Jimeno[d]

[a] *Department of Statistics and Management Sciences in Agriculture, University Polytechnic of Madrid. Ciudad Universitaria s/n, 28040 Madrid, Spain*

[b] *NIRS Laboratory, Feed Manufacturing Company, NUTEGA, S. L. c/ Marconi nº 9, Coslada, Spain.*

[c] *Animal Nutrition, Grasslands and Forages. Regional Institute for Research and Agro-Food Development. Apdo. 13, 33300, Villaviciosa, Spain*

[d] *Department of Animal Production, University Polytechnic of Madrid. Ciudad Universitaria s/n, 28040 Madrid, Spain*

Introduction

Near infrared (NIR) spectroscopy is becoming a technology widely used by the feed industry for rapid determination of chemical composition of forages. The authors of this study are participating in a R&D project sponsored by NUTEGA, Inc., a Spanish feed company that in addition to its products provides its clients with nutrition advice and quality control services. The main objective of this project is the development of validated calibration equations for the NIR equipment at the NUTEGA laboratory in order to evaluate the nutritive value of the forages more commonly used in Spanish dairy farms.

Multiple linear regression (MLR), principal component regression (PCR) and partial least squares (PLS) regression have been extensively used for statistical calibration of NIR instruments. Because severe collinearity is usually present in spectral data, MLR often has a limited ability to provide reliable information on the relative importance of the spectral data included in the regression model. Severe collinearity problems can also reduce the prediction accuracy of equations estimated by applying MLR.[1] If the model estimation data sets are affected by severe collinearity problems, more accurate prediction equations may be developed by using PCR, PLS or ridge regression (RR) techniques. However, the empirical comparisons of the accuracy of these statistical calibration techniques in the chemometric literature suggest that the superiority of one method over another for a particular data set cannot be generalized to other data sets.[2] For this reason, the widespread practice of validating the calibration equations with a single independent data set may lead spectroscopists to discard calibration equations that could be more robust with other data sets.

This work focuses on making available an integrated statistical description of the published information on the accuracy of the calibration equations developed for NIR analysis of forages. This type of study is known as a meta-analysis study. A meta-analysis study consist of selecting relevant empirical research studies available in the literature, coding the appropriate information about their characteristics and quantitative findings, including all this information in a data set, and analysing and describing the information contained in this data set.[3] This paper shows the results of applying

Near Infrared Spectroscopy: Proceedings of the 11th International Conference.

statistical process control techniques to the analysis of a data set containing published information on the prediction accuracy of different calibration equations developed to predict the crude protein content of various forages. The control charts obtained in this study allow a better understanding of the intrinsic variation in the average accuracy of these calibration equations. Likewise, they provide a support for the decision making in the calibration process.

Materials and methods

A data set was compiled from eight published studies that reported the results of validating with independent data several NIR calibration equations for forages. Table 1 describes the sources and the major characteristics of the data used for this preliminary meta-analysis study.

Table 1. Studies selected from the chemometric literature for compiling the dataset used in this meta-analysis study.

Reference	Number of observations compiled	Forage type	Dried?
4	3	Hay, haylage, corn silage	YES
5	2	Hay, haylage	YES
6	1	Grass silage	YES
7	1	Grass silage	NO
8	1	Corn silage	NO
9	12	Grass silage	YES
10	1	Legume and grass silages	YES
11	6	Alfalfa	YES

When being evaluated with independent datasets, the mean square prediction error (*MSPE*) as well as other statistics related to the *MSPE* are the measures most often used for evaluating the prediction accuracy of equations estimated by regression techniques.[1,12] The *MSPE* is defined as the mean of the squared prediction errors. The positive root square of the *MSPE*, denoted either as the root mean square error of prediction,[11] either as the standard error of prediction[6–8] or the standard error of validation,[9] has been extensively used to evaluate the accuracy of NIR calibration equations. In the chemometric literature, standard error of prediction (*SEP*) is the term most frequently utilised for this statistic. For this reason, we are going to use this notation in this work. The *SEP* can be mathematically expressed as:

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (1)$$

where n is the number of observations in the validation data set.

The *SEP* values found in the different studies selected were used to construct control charts for individual measurements. This type of control charts were considered as a statistical process control tool suitable for these data because the strategy followed in the selected studies was evaluating the accuracy of each calibration equation with a single validation data set. From a statistical process control perspective, it is obvious that the validation of each calibration equation with several independent data sets that could be assumed as drawn from the same population should have provided more accurate estimates of the population mean and variance of the *SEP*.[13] The greater the

size of the statistically homogenous lot of measurements used to estimate sample statistics, the grater the chance of detecting the presence of "assignable causes of variation" that should be identified in order to eliminate them or to incorporate them to the process.[13] The use of several validation data sets also provides some information on the robustness of prediction equations. But in this case, individual measurement charts is the type of Shewhart control charts that seems more adequate to the characteristics of the validation studies available in the chemometric literature. Since the Central Limit Theorem is not applicable for these control charts, a test of normality of the data should be performed.[13] The normality tests available in the STATGRAPHICS PLUS Version 5 statistical software were used in order to test the null hypothesis that *SEP* values in the meta-analysis data set come from a normal distribution.

The *SEP* values were randomly sorted before developing the individual measurement control charts by using the STATGRAPHICS 5.0 statistical software. The Shewhart control charts for individual charts are built by using the moving range of two successive observations as the basis of estimating the process variability.[14] The moving range is defined as:

$$MR_i = |x_i - x_{i-1}| \tag{2}$$

Because each *SEP* value is the result of validating with an independent data set each one of the different calibration equations, sorting at random seems appropriate in order to reduce possible biases due to the arbitrary sequence established for the data incorporated to the meta-analysis dataset.

Results and discussion

Figures 1 and 2 show, respectively, the individual measurement control chart and the moving range control chart for the *SEP* values obtained in the validation of the NIR calibration equations developed for crude protein (CP).

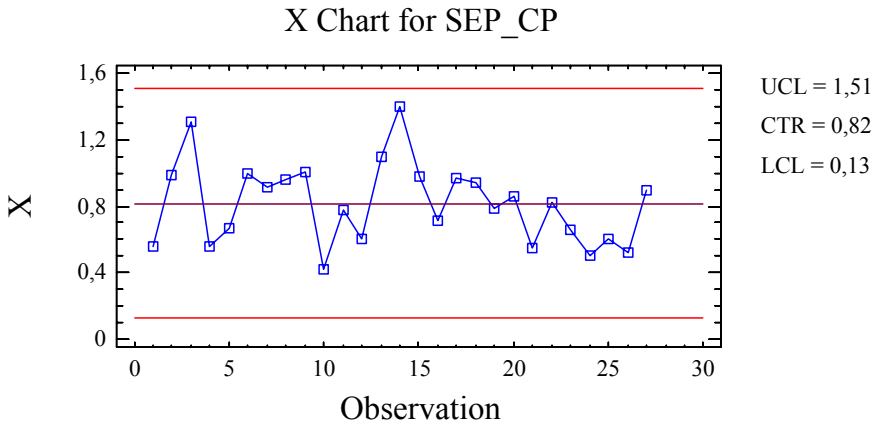

Figure 1. Individual measurement control chart for the 27 values of the standard error of in the meta-analysis dataset. Such values were obtained in the validation of NIR calibration equations for the CP content of different forages.

These control charts suggest that the *SEP* for CP (*SEP*_CP) can be considered as a quality measurement of a process (the development and validation of the NIR calibration equations) under statistical control. The tests of normality performed does not allow to reject the null hypothesis of the *SEP*_CP variable comes from a normal distribution with mean 0.82 and standard deviation 0.23. This information can be used for evaluating the *SEP* values obtained in the validation of new NIR calibration equations.

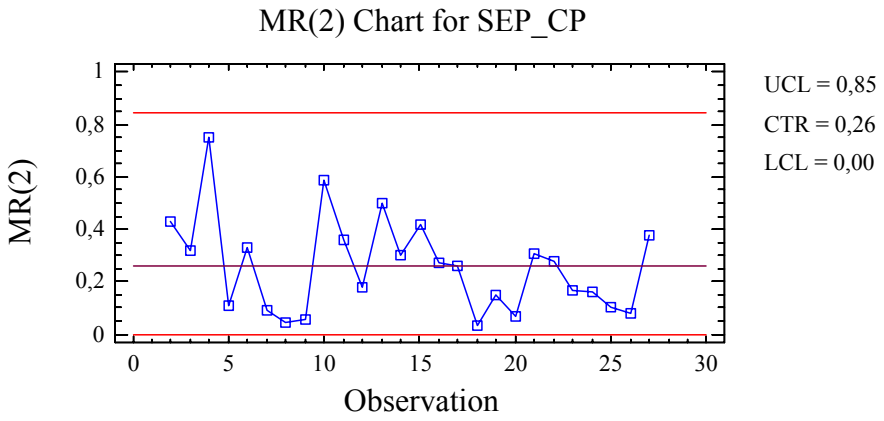

Figure 2. Moving range control chart for the 27 values of the standard error of in the meta-analysis dataset. Such values were obtained in the validation of NIR calibration equations for the CP content of different forages.

Conclusions

This study illustrates the potential of meta-analysis studies to provide Shewhart control charts that could be used to evaluate the accuracy of the new calibration equations. This approach also allows integrating the NIR calibration process in a statistical quality control and continuous improvement framework. The use of the ISO 9001:2000 standard for third-part certification of laboratory quality systems and the future versions of the ISO 17025 standard for laboratory accreditation are focusing on the adoption of this type of frameworks for quality control and improvement.

References

1. J.O. Rawlings, *Applied Regression Analysis. A Research Tool.* The Wadsworth & Brooks / Cole Advanced Books & Software, Pacific Grove, CA, USA (1988).
2. J. Huang, D. Brennan, L. Sattler, J. Alderman, B. Lane and C. O'Mathuna. *Chemom. Intell. Lab. Syst.* **62,** 25 (2002).
3. M.W. Lipsey and D.B. Wilson, *Practical Meta-Analysis*. Sage Publications, Thousand Oaks, CA, USA (2001).
4. J.S. Shenk and M.O. Westerhaus, *Crop Sci.* **25,** 1120 (1985).
5. J.S. Shenk and M.O. Westerhaus, *Crop Sci.* **31,** 1548 (1991).
6. B. de la Roza and A. Martinez, in *Making Light Work: Advances in Near Infrared Spectroscopy*, Ed by I. Murray and I.A. Cowe. VCH Publishers, Weinheim, Germany (1992).

7. J.L. De Boever, B.G. Cottyn, D.L. De Brabander, J.M. Vanacker and Ch.V. Boucqué, *Anim. Feed Sci. Technol.* **60,** 103 (1996).
8. J.L. De Boever, B.G. Cottyn, D.L. De Brabander, J.M. Vanacker and Ch.V. Boucqué, *Anim. Feed Sci. Technol.* **66,** 211 (1997).
9. B. Villamarin, "Calibración, mediante NIR, y Validación de Parámetros Analíticos para la Estimación del Valor Nutritivo de Ensilados de Pradera de Galicia", Ph. D. Thesis, University of Vigo, Vigo, Spain (1998).
10. P.C. Hoffman, N.M. Brehm, L.M. Bauman, J.B. Peters and D.J. Undersander, *J. Dairy Sci.* **82,** 764 (1999).
11. I. Ruisánchez, F.X. Rius, S. Maspoch, J. Coello, T. Azzouz, R. Tauler, L. Sarabia, M.C. Ortiz, J.A. Fernández, D. Massart, A. Puigdomènech and C. García, *Chemometr. Intell. Lab. Syst.* **64,** 169 (2002).
12. J. Fuentes-Pila, M.A. DeLorenzo, D.K. Beede, C.R. Staples and J.B. Holter, *J. Dairy Sci.* **79,** 1562 (1996).
13. J.R. Thompson and J. Koronacki, *Statistical Process Control. The Deming Paradigm and Beyond.* Chapman & Hall / CRC, Boca Raton, FL, USA (2002).
14. D.C. Montgomery, *Introduction to Statistical Quality Control*. John Wiley & Sons, New York, NY, USA (2001).