Calibration pixel selection for hyperspectral discrimination of ruminant and fish animal by-products

C. Riccioli,[a,*] W. Saeys,[b] P. Vallesquino-Laguna,[a]
D. Pérez-Marín,[a] J.E. Guerrero-Ginel[a] and A. Garrido-Varo[a]

[a]*Department of Animal Production, Faculty of Agricultural and Forestry Engineering, University of Córdoba, 14071, Córdoba, Spain.*
E-mail: z62riric@uco.es
[b]*BIOSYST-MeBioS, Faculty of Bioscience Engineering, K.U. Leuven, Kasteelpark Arenberg 30, B-3001 Leuven, Belgium*

Introduction

The ban on using animal by-products as feed ingredients for production animals is an important measure aimed at preventing the spread of Bovine Spongiform Encephalopathy (BSE) and is regulated by the EC Regulation 999/2001, amended by EC Regulation 1234/2003. This legislation bans the use of meat and bone meal (MBM) and allows the use of fishmeal only for non-ruminants. Reliable methods for identifying MBM of different species may help justify some modification of the total MBM ruling towards lifting it on fish meal as feed ingredient for ruminants. Traditional near infrared (NIR) spectroscopy has been utilized to determine the percentage of meat and bone meal in compound feeding-stuffs and also to detect the presence of animal species in animal protein by-products.[1,2] However, the models obtained are still not sufficiently accurate to justify the use of NIR spectroscopy for detecting—at forensic level—the presence of MBM from a given animal species in compound feeding stuffs.

Hyperspectral Imaging (HI), combining spectral and spatial information, is one of the candidate methods for rapid species identification in MBM.[3] Compared to traditional NIR spectroscopy, HI has the advantage of providing the opportunity to investigate localized micro-domains within a product. A physical or chemical abnormality that makes a minimal contribution to the bulk sample may go undetected by traditional NIR spectroscopy. The same abnormality could dominate a micro-domain and be detected in the NIR image.[4]

Since the hyperspectral image of each MBM sample consists of several thousand pixel spectra, a set of calibration pixels has to be selected for training the pixel classifier. For this purpose, all possible sources of variation that may be encountered later must be included in the calibration set, even though an excessive number of samples may result in redundant information,[5] which will

greatly increase the interference and complexity of multivariate calibration. In this study, several distance measures have been tested for use in the Kennard-Stone algorithm[6] for selecting pixels from pure samples of fish meal and ruminant meal to train a partial least squares discriminant analysis (PLSDA) classifier.

Materials and methods

13 samples from ruminant meal and 65 samples from fishmeal were analyzed with a NIR camera (MatrixNIR, Malvern Instruments, Maryland, USA). The camera had an Indium Gallium Arsenide (InGaAs) focal-plane array detector that could acquire images of 240 × 320 pixels with a resolution of 97.8 µm/pixel in the NIR region of 900 – 1750 nm. The spectral data were transformed by Standard Normal Variate (SNV) followed by two different treatments: a) Orthogonal Signal Correction (OSC) and Mean Centre and b) only Mean Centre. Basically, the OSC is a preprocessing technique used for correction of instrumental drift, bias and scatter in near-infrared spectra.[7] Outliers in spectral data were detected by studying score plots from a Robust PCA model, by the Hotelling's T^2-statistic and by the Q-statistic.[8] Afterwards, four different pixel selection strategies based on Kennard–Stone algorithm were evaluated to obtain a data reduction of 1/800 from the original spectra, Euclidean & Mahalanobis distance calculated on the original

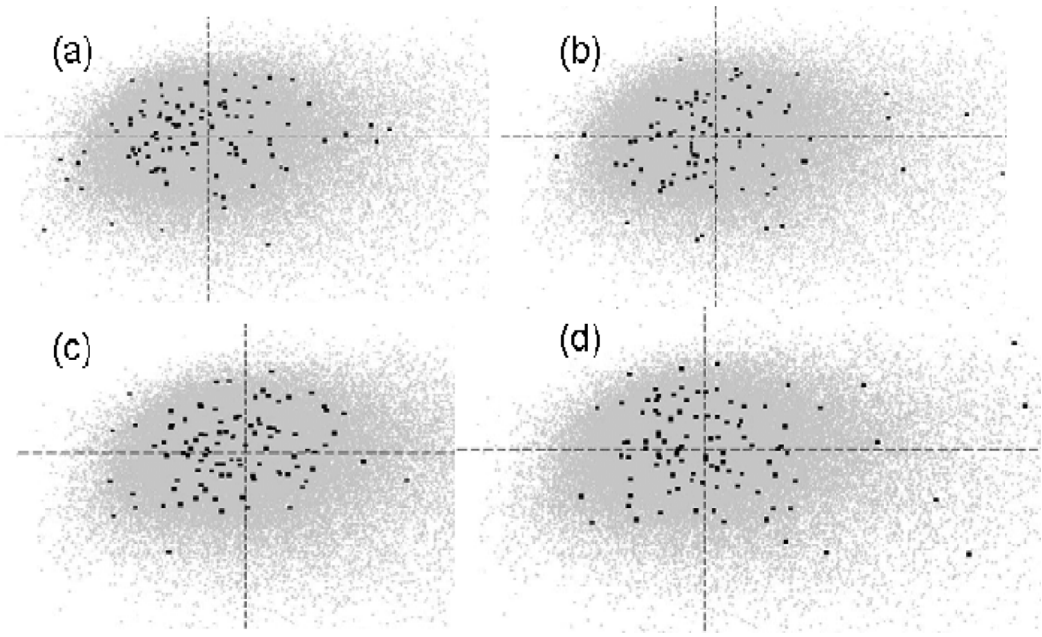

Figure 1. Pixels selected by the Kennard-Stone algorithm (■) based on: (a) Euclidean distance on original spectra, (b) Euclidean distance on PCs, (c) Mahalanobis distance on original spectra and (d) Mahalanobis distance on PCs.

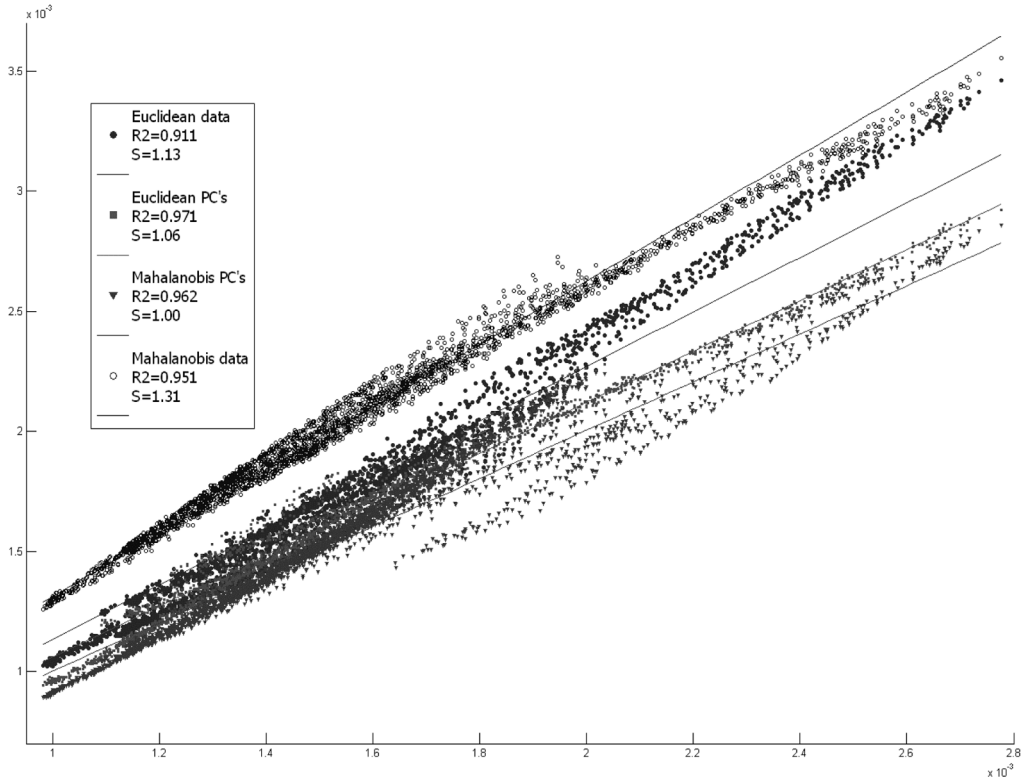

Figure 2. Linear regression from the covariance matrices of the original data and the reduced dataset.

spectra, and on the first nine PCs. Predictive models were developed by PLSDA. For the model validation, more than 6 million pixels of the original 78 pure samples were classified as ruminant or fishmeal. Although this validation set will give an over-optimistic image of the discrimination potential that can be obtained, it is very interesting to evaluate how good the different pixel selection strategies have been able to capture the variation present in more than 6 million pixels in a calibration set of 7800 pixels.

Table 1. Predictions of validation set (mean values) calculated applying PLSDA models

	Euclidean distance data (R^2=0.911 S=1.13)		**Mahalanobis distance data (R^2=0.951 S=1.31)**		**Euclidean distance PCA (R^2=0.971 S=1.06)**		**Mahalanobis distance PCA (R^2=0.962 S=1.00)**	
	OSC	No OSC	OSC	No OSC	OSC	No OSC	OSC	No OSC
% pixels correctly classified	95.34	93.91	95.41	94.23	94.86	93.92	95.10	93.98

Results and discussion

Figure 1 illustrates pixels selected from a fish-meal sample. The suitability of the new subset was evaluated as a first step through a corrected coefficient of determination (R^2) and the fitting slope value S obtained by linear regression from the covariance matrices of the original data and the new data set (Figure 2). PLSDA prediction results shown in Table 1 highlight the great accuracy of the models. According to the research undertaken, the use of Kennard–Stone algorithm based on Mahalanobis distance on original spectra and OSC after SNV have been identified as the most successful procedure. Nevertheless, for practical purposes, there was essentially no significant difference among any of the approaches.

Looking at R^2 and S values, it can be observed that pixels selected based on their first nine PCs have the highest ability to conserve the original variability of the whole data. A combination of both results points out the suitability of Kennard–Stone method applied using Mahalanobis distance on the first nine PCs.

Conclusions

The study demonstrated the good accuracy of the models created with a limited number of pixels/spectra for each sample of ruminant and fish meals. The best prediction model was obtained applying Mahalanobis distance on original spectra and treating spectra with OSC after SNV. Looking at R^2 and S, the most suitable values were obtained using indifferently Mahalanobis or Euclidean distances on the first nine PC's. Results show the suitability of the Kennard–Stone pixels selection method, applied using Mahalanobis distance instead of Euclidean distance.

References

1. A. Garrido-Varo, M.D. Pérez-Marín, J.E. Guerrero-Ginel, A. Gómez-Cabrera, M.J. de la Haba, J. Bautista, A. Soldado, F. Vicente, A. Martnez, B. de la Roza-Delgado and S. Termes, *Biotechnol. Agron. Soc. Environ*. **9(1),** 3–9 (2005).
2. M.J. De la Haba, A. Garrido-Varo, D. Pérez-Marín and J.E. Guerrero, *J. Near Infrared Spectrosc*. **15,** 237–245 (2007).
3. V. Baeten and P. Dardenne, in *Spectrochemical Analysis Using Infrared Multichannel Detectors*, Ed by Rohit Bhargava and Ira W. Levin). Blackwell Publishing, Oxford, UK, pp. 283–301 (2005).
4. E.N. Lewis, L.H. Kidder, E. Lee and K.S. Haber, in *Spectrochemical Analysis Using Infrared Multichannel Detectors*, Chapter 2, Ed by R. Bhargava and I.W. Levin. Blackwell Publishing, Oxford, UK, pp. 25–55 (2005).
5. A. Lorber and B.R. Kowalski, *J. Chemometr.* **2,** 67–79 (1988).
6. R.W. Kennard and L.A. Stone, *Technometrics* **11,** 137–148 (1969).
7. N.G. Yee and G.G. Coghill, *Chemometr. Intell. Lab. Syst*. **67(2),** 145–156. (2003).
8. Eigenvector Research, Inc., 830 Wapato Lake Road, Manson, WA 98831, USA.