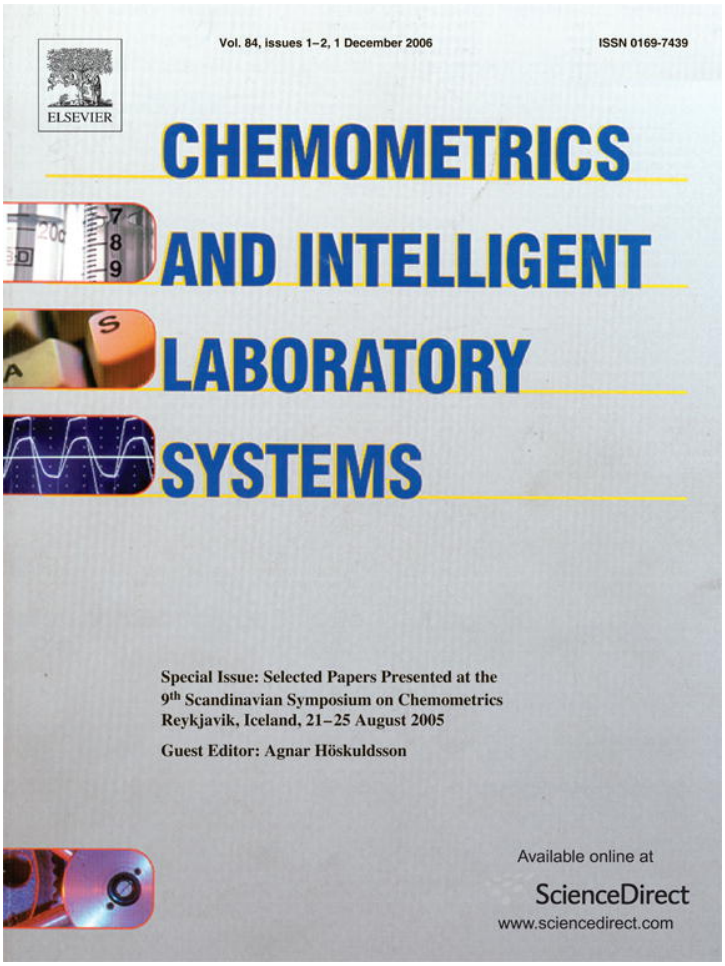

Available online at www.sciencedirect.com
ScienceDirect
Chemometrics and Intelligent Laboratory Systems 84 (2006) 114–118
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

Screening of compound feeds using NIR hyperspectral data

J.A. Fernández Pierna *, V. Baeten, P. Dardenne

Walloon Agricultural Research Centre (CRA-W), Quality of Agricultural Products Department, Chaussée de Namur no. 24, 5030 Gembloux, Belgium

Received 12 January 2006; received in revised form 3 March 2006; accepted 15 March 2006
Available online 7 July 2006

Abstract

Recent developments in spectroscopy have led to the use of spectroscopic imaging instruments for the control and monitoring of food and feed products. This kind of instruments offers the possibility of collecting thousands of spectra of particles being the result of the grinding of compound feedstuffs. The major advantages are that the recognition of feed ingredients is independent on the expertise of the analyst and that it is possible to automate all procedures and to analyse more samples per unit of time than classical microscopy or NIR microscopy. The objective of this study is the development of a new method for a rapid, precise and reliable screening of compound feeds. For that, a classification tree was built by sorting the particles in a dichotomist way where each node constitutes a discriminating step. These steps are completed by discriminant equations created from the hyperspectral databases obtained with a near infrared (NIR) camera for each class of raw materials. Discriminant equations were constructed using Support Vector Machines (SVM). For a new sample the aim is to determine its composition by using the classification tree. As general conclusion, hyperspectral data in combination with SVM as classification technique is a promised methodology for the determination of open formulations.

Keywords: Compound feeds; Screening; NIR camera; Chemometrics; SVM

1. Introduction

Animals farmed for human consumption should find in their feeding a daily supply on energy, proteins, vitamins, minerals and fibres. One part of these nutritious substances comes from the fodder (grass, hay, corn silage). These feed ingredients, rich in fibres, are eaten by the ruminant, which have the faculty to digest and to extract from them a large part of their needs [1]. But the fodder itself is not enough. The ruminants need a complement of proteins, oligo-elements and vitamins justified by the fact that the animals are bred in an intensive system. At European level, the fodder represents half of the total volume of feed. The other part of the volume is composed of a large diversity of products as cereals, peas and also co-products of the foodstuff industries. An example of co-product is the oilcake, i.e. the rest after the extraction of oil from the colza, sunflower or soya beans [2].

The composition of a compound feed is adapted depending on the kind of animal in such a way that the daily consumption should give a balance between the most important nutritional elements. Feed can contain more than 20 different feed ingredients, all of them known and officially authorised and it can be modified according to the market data. The European Commission, in the Council Directive 96/25/EC, defines compound feedingstuffs as «mixtures of feed materials which are intended for oral animal feeding as complete or complementary feedingstuffs». Feed materials (or feed ingredients) are defined as «various products of vegetable or animal origin, in their natural state, fresh or preserved, and products derived from the industrial processing thereof, and organic or inorganic substances, whether or not containing additives, which are intended for use (…) in the preparation of compound feedingstuffs or as carriers of pre-mixtures». Directive 2002/2/EC says that «a quantitative detailed information can improve the trace back of the components in order to reassure the stock breeders and the consumers». In this directive the Commission underlines the advantages of the labelling previsions of compound feedstuffs for production animals in order to facilitate the trace

* Corresponding author.
E-mail address: fernandez@cra.wallonie.be (J.A. Fernández Pierna).

0169-7439/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2006.03.012

back of compound feed. From January 2004, all the components used for making a compound feedingstuff have to be indicated together with their exact percentage [3]. However it has to be mentioned that this Directive has not yet been implemented in all countries.

NIR spectroscopy can give a very accurate global characterization of the samples but it is not normally suitable to study the distribution of different constituents in the products [4–6]. In order to characterize such distributions, it is necessary to develop devices including new technologies allowing to obtain a large number of information in a few minutes. This is why hyperspectral techniques were developed in the last decades [7–10]. Hyperspectral techniques consist in the use of images in order to distinguish or resolve closely spaced objects in an image. Spectral resolution is defined by the number and widths of wavelength bands that are collected by a detector. The detector used in this study is an array detector that gathers spectral and spatial data simultaneously by recording sequential images of a pre-defined sample; each image plane is collected at a single wavelength band. In the present study, reflectance images were collected in the near infrared range of 900–1700 nm window, with an increment of 10 nm. Many organic products have a unique spectral signature in that range that can be used for identification.

Then, in order to ensure the trace back of the feed we decided to work on the development of a hyperspectral technique: the NIR imaging technique. This technique allows the analysis of several hundreds of particles resulting from the grinding of compound feedstuffs. The major advantages of this technique are: (1) the recognition is not dependent on the expertise of the analyst and (2) it is possible to automate all procedures and to analyse more samples per unit of time than the classical microscopy or NIR microscopy [11,12]. NIR imaging spectroscopy has been presented as a more efficient method than those currently available, which tend to be time-consuming and to require significant analytical expertise [7–10].

The large quantity of data obtained with NIR imaging techniques needs the use of Chemometrics. Chemometrics has been defined by Wold as 'how to get chemically relevant information out of measured chemical data, how to represent and display information, and how to get such information into data'. He also explains that the only reasonable way to extract and represent the chemical information is in terms of models [13]. Then, in order to perform the trace back of the materials, discriminant models between the different feed ingredients have to be created. In this study, due to its powerful discrimination ability compared to other methods, Support Vector Machines (SVM) is used to construct those models [8,14,15].

2. Theory

2.1. Compound feed composition

The different feed ingredients making up compound feeds are normally organised depending on their botanical characteristics [16]. Besides the utilisation of the different vegetables in their whole shape, are the by-products widely used for the compound feeds. Among others we can cite are: corn, barley, wheat, rapeseed, flax, soya bean, sunflower, lucerne, pea, coconut, manioc, beet and palm by-products.

2.2. Support vector machines

Developed in the framework of Statistical Learning Theory (SLT), the objective of an SVM classifier is to derive a function, $f(x)$, based on a high-dimensional space that describes the decision boundary or hyperplane which optimally separates two classes of points. The problem of calculating an SVM model can be solved by finding an optimal solution for the quadratic programming problem [8]:

$$\min\left(\frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i\right) \tag{1}$$

subject to

$$y_i(w^T x_i) + b \geq 1 - \xi_i$$
$$\xi_i \geq 0, \text{ for } i = 1, \ldots, n$$

where n is the number of training data points, y_i is the reference or group value (± 1 for training points x_i), C is a penalty that has to be added in order to take into account the samples that cannot be properly separated and ξ_i represents the variables that measure the error made at a certain point (x_i, y_i).

The Lagrangian multipliers (α_i) technique is used for solving this equation and support vectors are the points for which α_i is positive and correspond to the nearest points to the boundary. In order to solve the problem of the dimensionality the kernel (K) function is introduced. The Kernel function maps the data into some dot product space (*feature space*) via a non-linear map and in such a way it constructs an optimal hyperplane in a high-dimensional space and is returned to the original space as a non-linear decision frontier. The Kernel function used in this work is the Radial Basis Function (RBF) described by:

$$K(x, x_i) = \exp\left(-\frac{||x - x_i||^2}{2\sigma^2}\right) \tag{2}$$

where σ represents the width of the Gaussian function and reflects the degree of generalization.

Finally the separating surface $f(x)$ that maximizes the distance between the two groups of cases is found of the form:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n_{SV}} \alpha_i y_i k(x_i, x) + b\right) \tag{3}$$

$0 \leq \alpha_i \leq C$.

3. Experimental

Since the main principle of the NIR camera is detailed elsewhere [7] only the main results are given here. The instrument used is a Matrix NIR™ instrument (Spectral Dimensions Inc., Olney, USA). This instrument includes an

InGas array detector (240×320 pixels) active in the 900–1700 nm region of the electromagnetic spectrum. The study concerns the discrimination between the different vegetal origin of the ingredients present in the compound feeds. The whole data set consists of 2,764,800 spectra (36 images collected using the NIR camera containing 76,800 spectra each). Then a first selection of the most representative spectra for each image has to be performed. The procedure includes a mask in order to remove the background and a Kennard and Stone selection [17] applied on the rest of the pixels in order to select the most representative spectra. The final data bank contains 94,140 spectra and includes spectra representative of 13 different vegetal ingredients: barley, coconut, corn, flax, lucerne, manioc, palm, rapeseed, soya bean, sugar beet, sunflower, sweet peas and wheat. Typical spectra of vegetal ingredients obtained with the NIR camera are presented in Fig. 1.

In order to perform the trace back of the different materials, a classification tree has been built by sorting the 13 species in a dichotomist way. This dichotomist tree should allow the discrimination of each ingredient participating in the composition of the compound feed because each node of this tree includes a discrimination model constructed using the spectral data bank of the vegetal ingredients (see Fig. 2). All these models aim the separation of particles of interest labelled as +1 (left part of the figure) from the rest labelled as −1 (right part of the figure).

The tree has been constructed after a detailed exploratory study of the data (not shown) including PCA analysis and individual PLS models in order to decide the order of discrimination of the feed ingredients. Different PLS models were first constructed by applying the 'one versus one' and 'one versus the rest' techniques. For both cases all the ingredients were tested. For instance: palm versus coconut, palm versus sugar beet, …and palm versus all the vegetables together. These PLS models were tested by cross-validation. Even if PLS could easily discriminate some ingredients (as palm or coconut) it could not do the same with others like the cereals for instance, while SVM does. That is the reason why PCA and PLS are only used as extra information and not for final model construction and SVM is used for that.

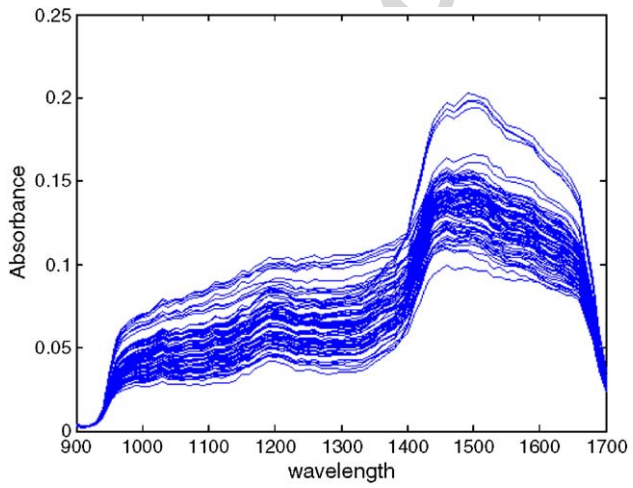

Fig. 1. Typical NIR spectra of vegetal ingredients obtained with the NIR camera.

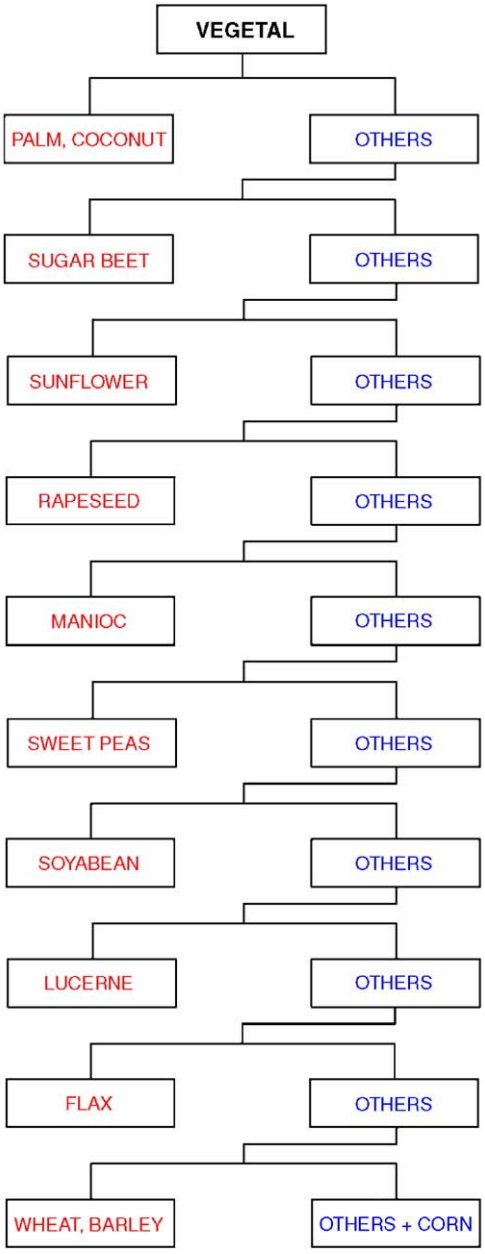

Fig. 2. Classification tree built by sorting the 13 species in a dichotomist way in order to perform the trace back of the different materials.

Then, the first step in the tree is the discrimination of palm and coconut from the rest of feed ingredients due to the fact that both ingredients were easily separated when looking at the principal components or when applying a cross-validated PLS model in our first exploratory analysis. The second node consists in the discrimination between sugar beet and the rest (except palm and coconut that have been already discriminated) and so on. Cereals go to the last part of the tree due to the difficulties to be separated.

SVM with an RBF Gaussian kernel is used as chemometric method in order to create these models. By analysing the results of each discriminant model, it should be possible to know

whether a feed ingredient is present or not. It should be also possible, if the feed ingredient is present, to know its percentage in the compound feed.

The main problem when working with SVM is the parameter optimisation, i.e. C and σ. When C increases, the errors are minimized. On the other hand, when σ increases the generalization also increases. It is important to find a good compromise between these two extremes. For this, an optimisation has been developed in which the data set has been split into two subsets: a calibration set containing 75% of the objects from each group of ingredients. This subset will be useful for the construction of the SVM models; and a test set, containing 25% of the objects from each class, will be useful for the validation of the constructed model. This optimisation works by performing both calibration and prediction using different combinations of C and σ. The optimisation program keeps the model whose classification percentages for both calibration and validation are higher with the minimum error. More than 36 different combinations of C and σ have been tested and the best models for each equation are shown in Table 1.

This table shows the results of classification for both the calibration and the test set and as it can be observed a correct classification of 99–100% in the calibration step is obtained for almost all the equations; in the case of the test set this percentage reaches 95.3%.

Once the different models have been constructed, they can be used for the prediction of new samples. For this reason, a data set has been collected in order to validate the different equations and to estimate their composition. This data set consists of compound feeds containing different feed ingredients with different percentages. This data set contains not only feed ingredients used in the construction of the discriminating tree but also feed ingredients not taken into account during the building of the models. Most of these new feed ingredients have similar characteristics as some of the feed ingredients used for the tree and, therefore, they should be detected. The data set consists on 270 spectra containing four different feed ingredients: wheat, oats, spelt and a hybrid between wheat and rye. None of the three last feed ingredients were included in the calibration set but they belong to the family of cereals that are included in the last step of the tree (wheat and barley).

Table 1
Results of classification for both the calibration and the test set with the best set of parameters (C and σ) for each node of the tree

Tree node	C	σ	% Correct calibration	% Correct validation
Palm, coconut versus rest	10,000	0.1	0.99	0.99
Sugar beet versus rest	10,000	0.01	0.99	0.92
Sunflower versus rest	10,000	0.1	0.99	0.96
Rapeseed versus rest	10,000	0.1	0.99	0.95
Manioc versus rest	100,000	0.1	1	0.98
Sweet peas versus rest	100,000	0.001	0.99	0.94
Soya bean versus rest	100,000	0.1	0.99	0.96
Lucerne versus rest	100,000	0.01	0.99	0.96
Flax versus rest	100,000	0.1	1	0.99
Wheat, barley versus rest	100,000	0.1	1	0.88

Table 2
Results after the application of the different steps of the tree in the validation data set

Tree node	Number of spectra detected as +1	Percent of spectra detected as +1
Palm, coconut versus rest	0	0
Sugar beet versus rest	2	0.74
Sunflower versus rest	3	1.11
Rapeseed versus rest	2	0.74
Manioc versus rest	3	1.11
Sweet peas versus rest	5	1.84
Soya bean versus rest	5	1.84
Lucerne versus rest	5	1.84
Flax versus rest	6	2.21
Wheat, barley versus rest	227	83.77
Rest	12	4.79

Table 2 shows the results after the application of the different steps of the tree. In the first nine steps a small percentage of spectra is misclassified. More than 83.7% of the spectra, though, is classified as being part of the cereal group. This result shows the good performance of the tree and the close relationship between the different cereal spectra. Step 10 could be considered as a discriminant equation between cereals versus the rest.

Other tests have been performed in order to validate the methodology proposed. A large number of mixtures have been applied through the tree. As example, for a mixture containing 72.60% of cereals and 7.5% of soybean and small quantities of other ingredients, more than 63.13% of the spectra have been classified in the group of cereals and almost 5.7% as soybean.

4. Conclusion

This study was a first tentative to contribute to the development of techniques for rapid, precise and reliable screening of compound feed. The combination explained in this paper, i.e. hyperspectral techniques applied in NIR spectroscopy (NIR imaging) and SVM as chemometric classification technique, allows to envisage a complete screening of feedstuffs at the same time that it provides an attractive solution for the characterization of feed mixtures. SVM performed adequately in analysing the training data and when analysing the data of spectral types which were represented within the training data set. Also SVM shows good results in generalization ability when unmodelled data were encountered.

Quantification of single feed ingredient could be conceivable and expressed by taking the number of correctly classified pixels of one ingredient divided by the total number of pixels.

The use of hyperspectral imaging in the agricultural science is increasing and in this study it is showed that it should be possible to perform a complete screening of feedstuffs to detect and quantify the presence of the different feed ingredients.

Acknowledgments

The authors thank Géraud Didelez and Isabelle Fissiaux from the CRA-W for their enthusiastic and organised participation in the realisation of this work.

References

[1] Y. Montecot, L'alimentation des animaux, Séminaire FEFAC-CEJA, 1–9 (2003).
[2] AFSSA, Inventaire des pratiques de l'alimentation animale, 18–73 (2000).
[3] AFSSA, Le rôle de la réglementation de l'alimentation animale dans la sécurité sanitaire, 74–111 (2000).
[4] D. Bertrand, La spectroscopie proche infrarouge et ses applications dans les industries de l'alimentation animale, Prod. Anim. 15 (3) (2002) 209–219.
[5] B.G. Osborne, 'Near Infrared Spectroscopy in Food Analysis', Text Encyclopaedia of Analytical Chemistry (Meyers, R.A.), John Wiley & Sons Ltd., Chichester, 2000.
[6] M.D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero, J. Zegers, A. Swinkels, L. Moya, J. Lizaso, Ad hoc labelling of compound feedingstuffs: the feed industry bets by near infrared spectroscopy', Proceedings of the NIR2003, NIR Publications, Córdoba, Spain, 2003.
[7] V. Baeten, J.A. Fernández Pierna, A. Michotte Renier, P. Dardenne, Imagerie infrarouge: analyse de l'alimentation animale, Techniques de l'Ingénieur (2005) (3) RE 34-1-8.
[8] J.A. Fernández Pierna, V. Baeten, A. Michotte Renier, R.P. Cogdill, P. Dardenne, Combination of Support Vector Machines (SVM) and Near Infrared (NIR) imaging spectroscopy for the detection of meat and bone meat (MBM) in compound feeds, J. Chemom. 18 (2004) 341–349.
[9] V. Baeten, J.A. Fernández Pierna, P. Dardenne, Hyperspectral imaging techniques: an attractive solution for the analysis of biological and agricultural materials, in: Hans F. Grahn, Paul Geladi (Eds.), Techniques and Applications of Hyperspectral Image Analysis, (in press).
[10] V. Baeten, P. Dardenne, Applications of near-infrared imaging for monitoring agricultural food and feed products', in Rohit Bhargava, Ira Levin (Eds.), Spectrochemical analysis using infrared multichannel detectors, (2005).
[11] M.J. De la Haba, J.A. Fernández Pierna, O. Fumière, A. Garrido-Varo, J.E. Guerrero, D.C. Pérez-Marín, P. Dardenne, V. Baeten, Discrimination of the class origin of bones present in the sediment fraction of animal by-products using Near Infrared Microscopy (NIRM), Proceedings of the NIR2005, NIR Publications, New Zealand, 2005.
[12] V. Baeten, C. von Holst, A. Garrido, J. Vancutsem, A. Michotte Renier, P. Dardenne, Detection of banned meat and bone meal in feedstuffs by near-infrared microscopic analysis of the dense sediment fraction, Anal. Bioanal. Chem. 382 (2005) 149–157.
[13] S. Wold, Chemometrics; what do we mean with it, and what do we want from it? Chemom. Intell. Lab. Syst., Lab. Inf. Manag. 30 (1995) 109–115.
[14] J.A. Fernández Pierna, P. Volery, R. Besson, V. Baeten, P. Dardenne, Classification of modified starches by FTIR spectroscopy using support vector machines, J. Agric. Food Chem. 53 (17) (2005) 6581–6585.
[15] S. Zomer, http://www.acc.umu.se/~tnkjtg/chemometrics/index.html.
[16] http://www.alp.admin.ch/en/fuetterung/fuetterung.php.
[17] R.W. Kennard, L.A. Stone, Technometrics 11 (1) (1969).