Alternative common basis for wavelet compression of NIR spectra and its application to food samples

M. Casale,[a,*] P. Oliveri,[a] N. Sinelli[b] and M. Forina[a]

[a]*Dipartimento di Chimica e Tecnologie Farmaceutiche ed Alimentari, Università degli Studi di Genova*

[b]*Dipartimento di Scienze e Tecnologie Alimentari e Microbiologiche, Università degli Studi di Milano*

Keywords: Wavelet compression, alternative common basis, Fisher weight, table olives, olive oils, SIMCA

Introduction

In the last 2 decades there has been an exponential increase in the application of NIR spectroscopy in many fields of food science. Considering the huge amounts of data acquired in short time, there is a need for advanced and efficient chemometrics strategies for the analysis of data. In this context wavelet transform compression has been proposed as a valid tool to reduce the data dimension, thus reducing also the noise. Wavelet compression replaces the original signal with a series of coefficients called the basis, but a common basis for all signals has to be found, when a matrix of signals is present instead of a single signal.

In the present state-of-the-art, the common basis is obtained from the variance spectrum or from the variance tree,[1,2] but despite this, sometimes the maximum variance does not correspond to the maximum useful information. In this study, a new strategy[3] was evaluated as an alternative to the ones described in the literature, with the objective of choosing the "best common basis" for classification issues. An alternative common basis, computed using the Fisher weight spectrum was developed, and its ability in classification problems was tested.

In particular two interesting applications to the food field were considered:

(1) the identification of the Taggiasca cultivar, starting from the analysis of table olives;
(2) the discrimination between monovarietal extra virgin olive oils, on the basis of the cultivar.

Materials and methods

Samples - Data set 1

Ninety olive samples were analysed by an FT Near-Infrared Spectrometer, based on a Polarisation Interferometer (NIRFlex N-500, Buchi), in the 4000–10000 cm^{-1} range with a 4 cm^{-1} resolution.

Of these 46 were olives from Taggiasca cultivar and 44 were olives from other cultivars. The samples were obtained from the 2007–2008 olive crop by collecting them directly from farmers. Olives of the Taggiasca cultivar, produced exclusively in Liguria, a region of Italy, are famous all over the world for their extraordinary quality.

These olives are principally used to obtain extra virgin olive oil, but a minor part of Taggiasca crop is used to prepare table olives of high commercial value that can be easily mixed with olives having less appreciated sensory features.

Samples - Data set 2

Fifty-five mono-varietal extra virgin olive oil samples (27 from cultivar Casaliva and 28 from Leccino) were analysed by an FT-NIR spectrometer (MPA, Bruker Optics) in the 4500–12500 cm^{-1} range with a 8 cm^{-1} resolution.

The mono-varietal extra virgin olive oil samples were obtained from 6-10 kg of olives, harvested in the harvest period of 2005-2006 in three different sites representative of Italian olive growing regions.

Multivariate data analysis

All spectra were pre-treated by Standard Normal Variate (SNV)[4] and first derivative, and then compressed by Wavelet transform.[2] Soft Independent Modelling of Class Analogy (SIMCA)[5] was used as a class-modelling technique.

Results and discussion

Alternative common basis procedure

Because we worked with matrices of spectra and not with a single spectrum, the application of the wavelet transform required a common basis, equal for all the objects.

The variance spectrum procedure computes the variance for the *N* variables. The spectrum of the variance is transformed by means of wavelets, and its compressed best basis is then used to transform and to compress all the objects.

This procedure has the objective to retain in the compressed data a large fraction of the variance of the original data. This is a very reasonable objective in many cases; however, in some cases, the maximum variance doesn't correspond to the maximum useful information. For this reason we used an alternative procedure to obtain a common best basis with the maximum information useful for classification. In this procedure the Fisher weights spectrum, instead of the variance spectrum, was considered.

Samples - Data set 1

To provide a benchmark against which to evaluate the merits of the alternative procedure based on the Fisher weight spectrum, the data compression was also performed using the variance spectrum procedure. Then, SIMCA was applied on the compressed spectra in order to build class models for the table olives, and the results were compared.

Table 1. Data set 1, SIMCA results.

	Classification ability (%)	Prediction ability (%)	Mean Sensibility (%)	Mean Specificity (%)
VARIANCE SPECTRUM	93.3	89.1	81.1	77.7
FISHER WEIGHT SPECTRUM	98.2	89.1	85.5	96.4

Table 2. Data set 2, SIMCA results.

	Classification ability (%)	Prediction ability (%)	Mean Sensibility (%)	Mean Specificity (%)
VARIANCE SPECTRUM	78.2	61.8	70.9	43.5
FISHER WEIGHT SPECTRUM	96.7	91.1	78.9	91.1

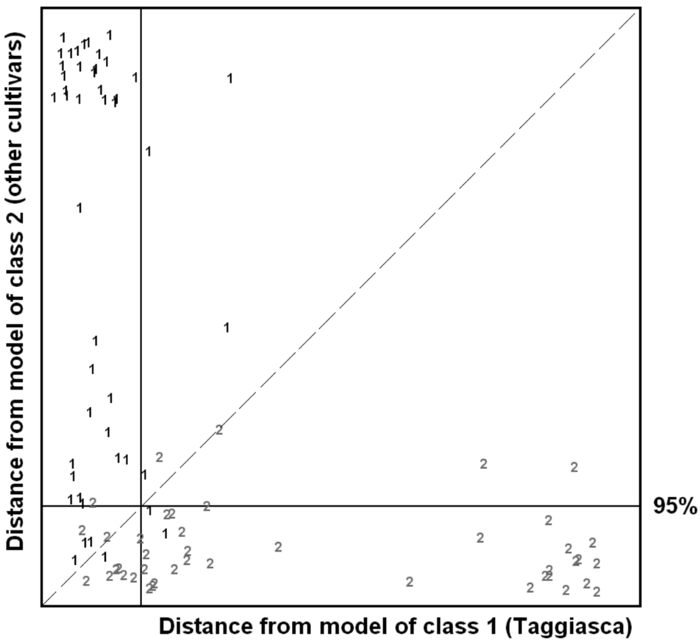

Figure 1. Table olive data set: Coomans plot for the SIMCA models after data compression by Wavelet transform (Fisher weight spectrum). Samples are represented by their class index.

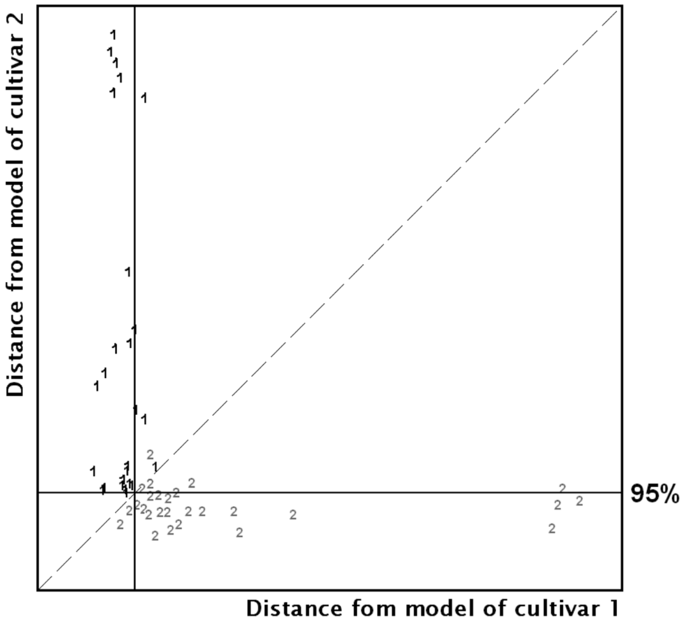

Figure 2. Extra virgin olive oil data set: Coomans plot for the SIMCA models after data compression by Wavelet transform (Fisher weight spectrum). Samples are represented by their class index.

The percentages of correct classification and prediction, and the sensitivities and specificities are listed in Table 1. These results showed that the alternative basis computed, starting from the Fisher weight spectrum, was more effective than the one based on the variance spectrum.

Samples - Data set 2

The same elaboration was performed on the olive oil data. The results are shown in Table 2.

The class models built for the 2 data sets using the Wavelet compression based on the Fisher weight spectrum are shown in the Coomans[6] plots in Figures 1 and 2, respectively.

Conclusion

In this paper, an alternative common basis for wavelet compression of NIR data, obtained not from the variance spectrum, but from the Fisher weight spectrum, was presented. The ability of this new procedure in classification and authentication problems of food samples was investigated and two studies were performed.

The alternative common basis proved very promising as a tool to compress NIR spectra, extracting the useful information, thus reducing also the noise.

SIMCA was applied on the compressed NIR spectra and the results indicated that the information retained after compression in the wavelets space using the Fisher weight spectrum was more efficient than that retained with a common basis, obtained by the variance spectrum.

References

1. B. Walczak and D.L. Massart, *Chemometr. Intell. Lab. Syst.* **38,** 39 (1997).
2. B. Walczak and D.L. Massart in *Wavelets in Chemistry*, Ed by B. Walczak. Elsevier, Amsterdam, The Netherlands, p. 165 (2000).
3. M. Forina, P. Oliveri, M. Casale and R. Boggia, *Chemometr. Intell. Lab. Syst.*, submitted.
4. R.J. Barnes, M.S Dhanoa and S.J. Lister, *Appl. Spectrosc.* **43,** 772 (1989).
5. S. Wold and M. Sjostrom in *Chemometrics, Theory and Application*, Ed by B.R. Kowalski, ACS Symposium Series 52. American Chemical Society Washington, DC, USA (1977).
6. Y. Mallet, D. Coomans and O. de Vel in *Wavelets in Chemistry*, Ed by B. Walczak. Elsevier, Amsterdam, The Netherlands, p. 151 (2000).