BARLEY VARIETIES DISCRIMINATED BY THE NEAR INFRARED HYPERSPECTRAL IMAGING TECHNIQUE

Ph. Vermeulen, J. A. Fernández Pierna, Th. Buhigiro, J.L. Herman, B. Lecler, A. Mouteau, G. Sinnaeve, V. Baeten & P. Dardenne

Walloon Agricultural Research Centre (CRA-W), Quality of Agricultural products Department,
Chaussée de Namur, 24, 5030 Gembloux, Belgium
vermeulen@cra.wallonie.be

INTRODUCTION

In order to analyse the technological quality of kernels as well as to identify and assess the barley varieties, a large number of analytical methods have been developed: visual examination of the kernel morphology (colour, size, shape, texture); simple laboratory tests and measures (yield, 1000 Kernels Weight (TKW), Specific Weight, calibration, aleurone colour test, germination analyses); more elaborated and slower methods as protein detection or DNA detection using the Polymerase Chain Reaction technique for variety fingerprinting; and non destructive and rapid methods as near infrared spectroscopy.
The objective of this study is to propose a fast and reliable method for the discrimination of varieties using the near infrared hyperspectral imaging technique. This is essential for establishing an efficient system for the traceability and quality control required in the seed sector as well as in the food and feed sectors.

EXPERIMENTAL - BARLEY

The trials for barley registration on the Belgium catalog studied by the department of plant production at the CRA-W offered a real opportunity for a variety discrimination study. Three hundred ninety-eight barley samples are available. Those samples are issued from different trials planned on three years (2004-2005-2006) and from seven Belgian locations. The varieties tested are winter and spring barley, 2 and 6 rows barley including malting and feed grade barleys. These samples have been selected in such a way that any variation due to climate, geographical location or agronomy in Belgium is included. Amongst the 32 varieties, a set of 6 varieties, 6 rows winter barley (Nikel, Seychelles, Palmyra, Jolival Mandy and Pelican), were studied on the three years in the same 6 locations. This set of 108 samples, i.e. 18 samples by variety, was selected for the study proposed in this poster.
All those varieties were first observed in the field and then the samples were analysed at the laboratory. Information regarding yield and component yield (TKW, Specific weight, Humidity), have been collected as well as technological analyses by using reference methods (size <2.2mm and >2.5mm) and spectroscopic methods (protein). Beside those classical analyses, a recent and more sophisticated technique was also applied: the kernel by kernel analysis using NIR hyperspectral imaging technique.

HYPERSPECTRAL IMAGING

The near infrared camera used here is a MatrixNIR™ Chemical Imaging System (Malvern Instruments Ltd - UK). The characteristics of the NIR camera are:

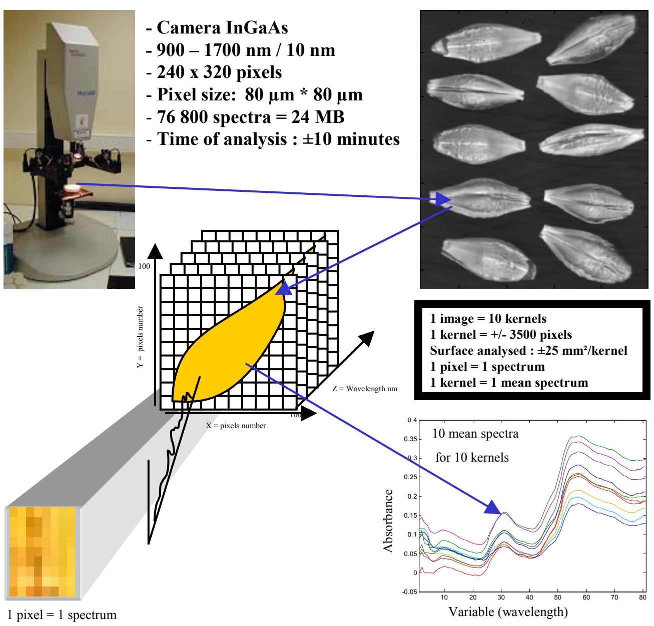
- Camera InGaAs
- 900 – 1700 nm / 10 nm
- 240 x 320 pixels
- Pixel size: 80 µm * 80 µm
- 76 800 spectra = 24 MB
- Time of analysis : ±10 minutes
1 image = 10 kernels
1 kernel = +/- 3500 pixels
Surface analysed : ±25 mm²/kernel
1 pixel = 1 spectrum
1 kernel = 1 mean spectrum
Y = pixels number
X = pixels number
Z = Wavelength nm
10 mean spectra
for 10 kernels
Absorbance
Variable (wavelength)
1 pixel = 1 spectrum

Figure 1. NIR camera and its characteristics.

For each barley sample, 10 kernels were analysed. The spectrum of each kernel is calculated from the average of the spectra acquired on the full surface of the kernel (see figure 1).

This work is partly undertaken in the framework of the FP6 **Co-Extra** – 007158 project (GM and non-GM supply chains: their **CO-Ex**istence and **TRA**ceability
The authors wish to thank department of plant production at the CRA-W for the supplying of the barley samples.

RESULTS

The first part of the study concerns the discrimination between the 6 varieties based on the field and technological data. For this, a set of 108 values (6 varieties X 6 locations X 3 years) was constructed for each variable: yield, humidity, size <2.2mm, size >2.5mm, TKW, specific weight and proteins content.
The most used statistical tool by the breeders are the univariate statistics with the analyse of variance (Anova) and the multiple comparison methods. The **one-way Anova** shows for which variable the differences between varieties are significant or not. The **multiple comparison** shows if the mean of one variety is significantly different from the mean of another variety (see table 1).

	Anova1		Means comparison						
Variable	Prob>F		Nikel	Seychelles	Palmyra	Jolival	Mandy	Pelican	
Yield	0.0209	*	98.3333	99.0556	99	99	105.5	104.0556	NS
Humidity	0.7317	NS	13.4905	13.3851	13.5649	13.2143	13.5067	13.0966	NS
Cal<2.2	0.0007	***	3.7088	6.1907	1.7786	6.6448	7.1919	4.0598	*
Cal>2.5	4.05E-09	***	82.9889	78.5172	92.9084	78.0516	71.13	85.0733	*
1000K weight	4.11E-11	***	46.325	44.1633	46.6991	43.5182	41.1796	47.8815	*
Specific Weight	3.05E-12	***	63.0912	63.4225	65.8316	63.4702	67.9098	63.8509	*
Protein	0.0169	*	12.1174	11.5426	11.8728	12.1178	11.8827	10.8131	*

Table 1. Anova and Means multiple comparison on field and technological data.

To extract the maximum of information from the data, multivariate tools as principal component analysis (**PCA**) can be applied. By analysing the scores (relationship between samples) together with the loadings (relationship between variables), one can distinguish which variables are responsible for the differences between varieties. For example, the variety PELICAN is clearly described as a variety with big kernels (Size >2.5) and high TKW but low specific weight, according to the PC1 and as a variety with high yield potential but low level of proteins according to the PC2 (see figure 2).

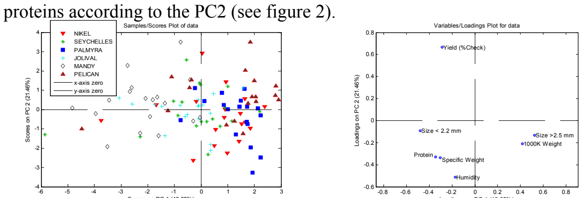
Samples/Scores Plot of data
NIKEL
SEYCHELLES
PALMYRA
JOLIVAL
MANDY
PELICAN
x-axis zero
y-axis zero
Scores on PC 2 (21.46%)
Scores on PC 1 (46.65%)
Variables/Loadings Plot for data
Yield (%Check)
Size < 2.2 mm
Size >2.5 mm
1000K Weight
Protein
Specific Weight
Humidity
Loadings on PC 2 (21.46%)
Loadings on PC 1 (46.65%)

Figure 2. PCA on field and technological data: scores and loadings graphs.

For the seed control laboratories, the main issue is to be able to classify the varieties in order to identify an unknown variety or to check the variety label. Supervised chemometric tools as **PLS-DA** allow to build classification models.
For the validation of the model, a calibration set (76 samples) and a prediction set (36 samples) have been built by using the Kennard-Stone method. The data were normalized and autoscaled. Table 2 shows the sensitivity and the specificity for each of the 6 varieties in calibration, leave one out cross-validation and prediction. The classification errors in prediction vary from 2% (Mandy) to 57% (Seychelles).

PLSDA	Nikel	Seychelles	Palmyra	Jolival	Mandy	Pelican
Sensitivity (Cal):	0.769	0.615	1.000	0.923	1.000	0.833
Specificity (Cal):	0.525	0.458	0.867	0.542	0.984	0.850
Sensitivity (CV):	0.769	0.615	1.000	0.846	1.000	0.667
Specificity (CV):	0.492	0.441	0.850	0.542	0.952	0.867
Sensitivity (Pred):	0.800	0.400	0.833	1.000	1.000	0.667
Specificity (Pred):	0.645	0.452	0.867	0.355	0.963	0.767
Class. Err (Cal):	0.353	0.463	0.067	0.267	0.008	0.158
Class. Err (CV):	0.370	0.472	0.075	0.306	0.024	0.233
Class. Err (Pred):	0.277	0.574	0.150	0.323	0.019	0.283
RMSEC:	0.411	0.422	0.342	0.418	0.245	0.357
RMSEP:	0.363	0.376	0.412	0.376	0.296	0.393

Table 2. PLSDA on field and technological data.

The second part of the study concerns the discrimination between the 6 varieties based on the spectral data acquired with the NIR-camera. A set of 1080 spectra (10 kernels X 6 varieties X 6 locations X 3 years) was constructed. The same chemometric tool, PLS-DA, was used. For the validation of the model a calibration set (992 spectra) and a validation set (88 spectra) have been built by using the Duplex method. The data were preprocessed by the Standard Normal Variate transform followed by 1st derivative Savitzky-Golay treatment (15,2,1).

PLSDA	Nikel	Seychelles	Palmyra	Jolival	Mandy	Pelican
Sensitivity (Cal):	0.608	0.782	0.800	0.717	0.780	0.801
Specificity (Cal):	0.799	0.554	0.738	0.694	0.739	0.746
Sensitivity (CV):	0.585	0.721	0.776	0.675	0.767	0.795
Specificity (CV):	0.790	0.542	0.728	0.692	0.733	0.742
Sensitivity (Pred):	0.889	0.867	0.867	0.714	0.810	0.714
Specificity (Pred):	0.722	0.603	0.699	0.581	0.806	0.797
Class. Err (Cal):	0.296	0.332	0.231	0.295	0.240	0.227
Class. Err (CV):	0.312	0.369	0.248	0.316	0.250	0.231
Class. Err (Pred):	0.195	0.265	0.217	0.352	0.192	0.244
RMSEC:	0.349	0.361	0.332	0.353	0.330	0.332
RMSEP:	0.318	0.381	0.380	0.342	0.367	0.388

Table 3 . PLSDA on spectral data.

Table 3 shows the sensitivity and the specificity for each of the 6 varieties in calibration, leave one out cross-validation and prediction. The classification errors in prediction vary from 19% (Mandy) to 35% (Jolival).

CONCLUSION

The first results of the study allow to define the potential ability in variety discrimination by analysing kernel by kernel using NIR hyperspectral imaging spectrometers combined with chemometrics methods. PLS-DA models applied on spectral data give 81% of correct classification (sensitivity) in average against 78 % for the same model applied on field and technological data. The major advantage of the NIR camera technique is that for a limited number of kernels, it is possible to analyse several variables in the same time and on a large number of samples by unit of time.

REFERENCES

[1] Fernandez Pierna, J.A., Vermeulen, P., Buhigiro, T., Baeten, V. & Dardenne, P. (2006). *Les graines sous l'œil de l'imagerie hyperspectrale proche infrarouge.* Chimiométrie 2006, Paris, 30/11-01/12/2006.
[2] Baeten, V., Fernandez Pierna, J.A. & Dardenne, P. (2007). Hyperspectral imaging techniques: an attractive solution for the analysis of biological and agricultural materials. In: *Techniques and Applications of Hyperspectral Image Analysis,* Editors, Hans F. Grahn & Paul Geladi.

Walloon Agricultural Research Centre