CO-EXTRA

GM and non-GM supply chains: their CO-EXistence and TRAceability

Project number: 007158

Integrated project
Sixth Framework Programme
Priority 5

Food Quality and Safety

Deliverable D5.4.

Title: Report specifying the applicability of hyperspectral NIR in GMO detection on different levels including the economic level

Authors: J.A. Fernández Pierna, Ph. Vermeulen, Th. Buhigiro, G. Berben, P. Dardenne & V. Baeten.

Walloon Agricultural Research Centre (CRA-W), Quality of Agricultural Products Department, Chaussée de Namur n°24, 5030 Gembloux, Belgium

Due date of deliverable: M18

Actual submission date: M 32

Start date of the project: April 1st, 2005 **Duration**: 48 months

Organisation name of lead contractor: CRA-W

Revision: vFINAL

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)	
Dissemination Level	
PU Public	
PP Restricted to other programme participants (including the Commission Services)	
RE Restricted to a group specified by the consortium (including the Commission Services)	**X**
CO Confidential, only for members of the consortium (including the Commission Services)	

Table of contents

Summary

In order to analyse the technological quality of kernels and to detect and quantify GMO events, several analytical methods have been developed.

Here a recent and more sophisticated technique is applied: the NIR imaging method, which has been described as a powerful approach for remote sensing in agronomical applications. The major advantages of this technique are that it is non-invasive, the recognition is not dependent on the expertise of the analyst; it is possible to automate all procedures and consequently to analyse a large number of samples.

The objective of this project is to propose a fast and reliable method for the detection and the quantification of GMO in grains, essential for establishing an efficient system for analytical traceability.

The first results of the research undertaken at CRA-W in the framework of the CO-EXTRA project allow to define the potential ability in GMO detection. As general conclusion it is evident that the kernel study can be successfully done by using NIR and NIR hyperspectral imaging spectrometers and that possibilities for the application of this method to the GMO detection become clear.

The NIR imaging protocol used to analyse the samples is presented in Annex 2. More details about the analytical protocol and software tools used can be available upon request.

1 Introduction

Amongst the different edible parts of a plant, the kernel is the most important product used in food and feed throughout the world. The kernel production is issued from two crop categories: the temporary crops and the permanent crops (FAO 2001).

Temporary crops. The most important is the cereal production. With a high content in carbohydrates, mainly starch, cereals produce the main food energy source. Major cereals include wheat, durum wheat, barley, oat, rye, spelt but also rice and corn. Minor cereals are sorghum, millet, buckwheat. Rice is the most important cereal since it is the base food for more than 50% of the world population and more than 20% of food energy is coming from this plant (FAO 2005). Another important temporary crop concerns the pulse production which are the kernels produced by legumes. They reach a high content in proteins and are used mainly in feed but also in food in complement to cereals. Major legumes as crops are beans, peas, broad beans, lentils and lupines. Another category of temporary crop producing kernels with interest for food and feed are oilseeds crops. Their industrial potential is in their high content in oil (17% for soybean to 50% for sesame). Major oil bearing crops are sesame, soybean, peanut, rapeseed, sunflower, cotton, flax and hemp. A fourth type of kernel-producing temporary crop extends to some vegetables. Their kernels are consumed while being not completely mature and therefore contain high water contents (70 to 95 %). Major vegetables of that category are green maize and green peas.

Permanent crops. The main category concerns the nuts, which are dry fruits or kernels produced by trees or bushes. These kernels are characterised by a hard outer husk and a high content of oil. Major nut producing crops are hazelnut, chestnut, walnut and acorn. Another category covers spices when plants produce kernels containing strongly flavoured and aromatic substances used as condiments. (e.g. pepper, mustard). Stimulant crops are still another source of kernels which contain alkaloids as caffeine (e.g. coffee beans). Within this category some kernels also have a high content in fat as cocoa beans.

As the NIR analyses are performed on a kernel basis, beans and grains for food and feed are not the sole possible materials to be analysed, all kernels used as seeds are also in the scope of the analyses.

1.1 Economic use of the kernel

The kernel has a double economic use: as seed and as food or industrial resource.
Seed. The seed sector concerns the farm seeds and the certified seeds. Even if farm seeds are yet largely used throughout the world, its use is decreasing. The certified seed is issued from a breeding process. First of all, seed research companies create new varieties by cross, by single seed descent, by breeding or by using different genetic methods to improve yield, pests and disease resistance, environmental adaptation or technological quality. Before to be used by farmers, seeds have to be registered according to specific criteria. The next step is the multiplication of seeds and the conservative breeding to keep the registered variety characters. Then, after cleaning, germination test, disease treatment, the seeds are packaged and stored before to be available to the farmers through the commercial chain. At each step of the seed creation and production, the grains are analysed to control the characters (yield, quality, purity..)

Agro Food industry. **The kernel can be used for different outlets. The main outlets are the food and feed sectors and then the cosmetic and energy sectors.**

In the food sector, some grains are entirely eaten without any transformation. So, 95% of the rice production and 6% of the corn production are sold directly as they are.

Other kernels need processing to become food/feed products:
- the flour industry is in charge to mill the grain (wheat, rye, ..) before to deliver the flour to the bakery sector to produce bread and cakes by fermentation or to the pasta industry (durum wheat);
- the brewing and distilling industry with grain fermentation to produce beer (2 row barley), alcohol as Whisky (2 row barley) or Sake (rice);
- the oil industry with oil extraction from kernels (rapeseed, sunflower, corn germ, sesame, soybean). With the importance of omega 3, 6, 9 and fatty acids in food, some secondary crops receive new outlets as lupine, flax, hemp;
- the roasting industry (coffee, cocoa);
- the condiment industry (mustard, pepper);
- the feed industry (6 rows barley, feed wheat, corn, oat, millet, soybean, rapeseed) to produce feed for cattle.

Other Industry. Other sectors can also use grains to make products not dedicated for food and feed. One example of such industry, at the moment in high evolution, is the bio-energy sector with the production of fuel made from renewable resources such as vegetable oils (rapeseed, soybean, wheat). Besides the natural properties of some plant kernels, pharmaceutical sector found new resources with the production of GMO lines selected for their capacity to produce specific proteins dedicated to make drugs. Regarding the cosmetic sector, some cosmetics are produced from rice powder and corn. A common sector to the three previously mentioned industrial sectors is the handling sector which is very important to dispatch productions and products from farm to industry and from industry to the consumer. This sector includes the transport, the packaging and the storage in silo.

Analyses are needed all along the product transformation chain in order to 1) check the quality of the product, 2) detect any undesirable product, and 3) identify and assess the seed being planted but the also harvested linked to it and its utilization in further transformation.

1.2 Kernel quality criteria

The quality of a kernel can be affected by a large number of factors: insect infestation, water stress,… Any loss in the quality can lead to financial losses for producers and retailers if the contamination or the deficiency is not detected on time.
The criteria used to determine the quality of a kernel are strongly related to the destination: food or seed.

In the **food sector**, some quality criteria are visual as the specific purity (no contamination by other crop seeds, weed seeds or inert material), the uniformity of size, shape and colour or the absence of external damage.
Other criteria can not be determined by only looking at the grain as some insect damage (weevil), some toxin contamination (mycotoxins), some stress linked to the storage conditions (temperature, humidity) or composition content (starch, protein, oil, humidity) or genetic origin (GMO presence).

In the **seed sector**, besides the quality criteria for food, other criteria are very important as the variety purity (no GMO seeds) and the seed germination.
For seed registration, the main criteria are DHS and VAT:
- D for Distinction against existing varieties to guarantee to the farmer a variety with innovation and to the breeder a patent;

- H for Homogeneity of characters through the seed multiplication;
- S for Stability of characters through the years;
- VAT for technological and agronomical value to guarantee a better yield or quality.

1.3 Grain analytical methods

To analyse the technological quality of kernels and to identify and assess the varieties, a large number of methods have been developed.

The first set of methods are **visual examinations** of the kernel (colour, size, shape) taking into consideration a reference description. Normally these quality assessment procedures are carried out by trained personnel. This manual assessment is very subjective, which makes the task inefficient and unreliable. The kernel industry wants tools based on consistent and objective criteria to do quality evaluation in real-time. To meet those requirements, recent research has shown that the **video camera imaging method** has the potential to become a viable tool for kernel quality inspection (Huckerby, Winter et al. 2005). Efforts have been made to assess damaged kernels and identify kernel types under controlled conditions (on single kernels, thus without overlapping between kernels) but also in situation where kernel touch or overlap each other (Wang, Paliwal et al. 2004).

The second set of methods are **simple laboratory tests and measures** (yield, 1000 Kernels Weight, Specific Weight, Phenol test). Those methods are based on simple instruments and give quantitative information.

The third set of methods are more **elaborated and longer lasting methods** as protein analysis or DNA analysis by gel electrophoresis or PCR.

The fourth set of methods are **non destructive and rapid methods**: the near infrared spectroscopy (NIRS) methods.

In the 80's, NIRS began to be used in the Agriculture sector as rapid method compared to traditional methods in general.. The first NIRS protocols proposed were dedicated to the analysis of flours. Then new NIRS protocols were developed allowing the direct analysis of grain without the need to grind the samples previously to the analysis. To meet the quality product specifications required by the world grain markets and by the agro-food industries, analytical methods have been adapted for analysis at the kernel level. A recent and more sophisticated technique is the ***NIR imaging*** method, which is able to perform single kernel analysis in order to detect small contamination. This technology has been described as a powerful approach for remote sensing and agronomical applications as for instance in precision agriculture and mineralogy among others. The major advantages of this technique are that the recognition is not dependent 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 NIRS. The first studies demonstrated the potential of this technique to detect weevil larvae in grain, ergot in wheat and to predict moisture and oil in maize kernel (Baeten and Dardenne 2005; Baeten, Fernandez Pierna et al. 2006). The development of NIR methods on a single kernel involves the analysis by chemistry reference methods on one kernel as well. A limitation of the NIRS method could be the non existence of those reference methods.

2 GMO context

Legislation enacted worldwide to regulate the presence of genetically modified organisms (GMOs) in crops, foods and ingredients, requires the development of reliable and sensitive

methods for GMO detection (Ahmed 2002). In the present work, the aim is to investigate the potential of the NIRS and the NIR imaging methodology together with chemometric tools in order to discriminate soybean and barley kernels for the detection and quantification of GMO. In its review of GMO detection and quantification techniques the European Commission presents the NIR spectroscopy as a possible alternative technique for GMO analysis. In this review, based on the work of Hurburgh *et al.* (2000), NIR spectroscopy is shown as a promising technique for the detection of the genetic modifications that may alter the fibre structure in plants. Roussel *et al.* (2001; 2004) have also demonstrated the utility of NIRS as rapid and inexpensive test for distinguishing GMO soybean (Roundup Ready ^{TM}) from non-GMO grain in inbound deliveries. They have applied chemometric tools as Partial Least Squares (PLS), Locally Weighted Regression (LWR) or Artificial Neural Networks (ANN) using a database of more than 8000 samples showing a classification accuracy of more than 90%.

3 Methodology

The objective of this project is to propose a fast and reliable method for the detection and the quantification of GMO in grains, essential for establishing an efficient system for traceability . This objective will be reached by:

1) **Developing a sample bank.** This sample bank should contain different varieties of soybean and barley coming from different origins and some of them being transgenic.

2) **Developing a spectral data base.** A spectral data base has to be developed during the project. This spectral data base contains classical NIR spectra and images/spectra coming from the NIR camera.

3) **Developing models to discriminate GMO.** The collected spectra are fingerprints that are based on the chemical composition of the analysed kernels. They should allow a good identification of kernel origin as well as the presence of GMO by using a statistical model constructed using SVM. This model allows the discrimination between pixels corresponding to GM and non GM kernels. Before the construction of these models, an exploratory analysis has to be performed using Principal Component Analysis (PCA) and wavelength selection (Massart *et al.* 1988). For the discriminant models, Partial Least Squares Discriminant Analysis (PLS-DA) (Barker and Rayens 2003) and Support Vector Machines (SVM) (Vapnik 2000; Fernandez Pierna *et al.* 2006) techniques can be used to construct the models.

4) **Testing and validating the models.** The models developed have to be tested with independent data sets in order to validate them. With this aim, the spectral data base has to be split into two groups, one for model construction and the other for validation.

3.1 Infrared Spectroscopy

Near-infrared spectroscopy (NIRS) is a branch of vibrational spectroscopy that provides information about the vibrations of molecules.

Figure 1 shows the location of the infrared (MIR and NIR) spectroscopy in the electromagnetic spectrum. The information is contained within the wavelength or frequency spectrum of absorbed intensity. From a chemical point of view, infrared spectroscopy is based on the vibrational transitions occurring in the ground electronic state of the molecules.

Infrared absorption requires a change of the intrinsic dipole moment with the molecular vibration.

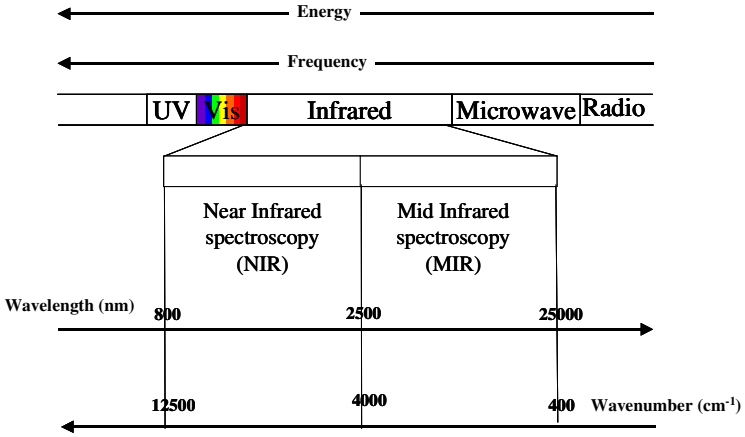

Figure 1: NIR and MIR spectroscopy in the electromagnetic spectrum

The near Infrared technique is based on the measurement of the intensity of the absorption of infrared radiation (800-2500 nm) by a sample. The energy in this range may excite molecular vibration to higher energy levels.

NIR radiation can only be absorbed by bonds within a molecule if the radiation has exactly the right energy to induce a vibration of the bond. This is the reason why only specific frequencies are absorbed. Absorption occurs at wavelengths that match the frequency of the molecular vibration. The spectrum generated is made up of “Absorption bands”. Moreover, the absorption is characteristic of chemical and biochemical bonds of the product, therefore the study of the absorption bands provide information about the composition of the product. Near infrared spectroscopy is commonly used to chemically analyse kernel due to its rapid and non-destructive characteristics.

3.2 Near Infrared Camera - Imaging

Recent developments in NIR focal plane array (FPA) technology led to the creation of imaging spectroscopy, which combines the advantages of spectroscopic and microscopic methods, along with much faster sample analysis since the spectral data are acquired in parallel. An imaging spectrometer 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. The instrument used at the CRA-W (Fernandez Pierna *et al.* 2005) is a MatrixNIR^{TM} Chemical Imaging System (Spectral Dimensions Inc., Olney, USA). The principle of the instrument is represented in Figure 2a and is as follows: Four dedicated lamps for NIR analysis (polymachromatic source) illuminate the surface where the sample is spread. The imaging spectrometer utilises an InGaAs focal plane array (FPA) with 240 x 320 pixels (76 800 spectra per scan), along with a liquid crystal tuneable filter (LCTF) for wavelength selection. The effective field of view (FOW) covers approximately 5 cm² allowing simultaneous analysis of 10-15 kernels. Reflectance images are collected in the 900 - 1700 nm window, with an increment of 10 nm. The image planes are stacked to form a three-sided matrix or spectral volume (see Figure 2b), where the first two axes (x and y) define the image plane (field of view- FOV), and the third (z axis) corresponds to the spectrum at each pixel location in the FOV. Figure 3 shows a few spectra of soybean kernels obtained with the NIR camera.

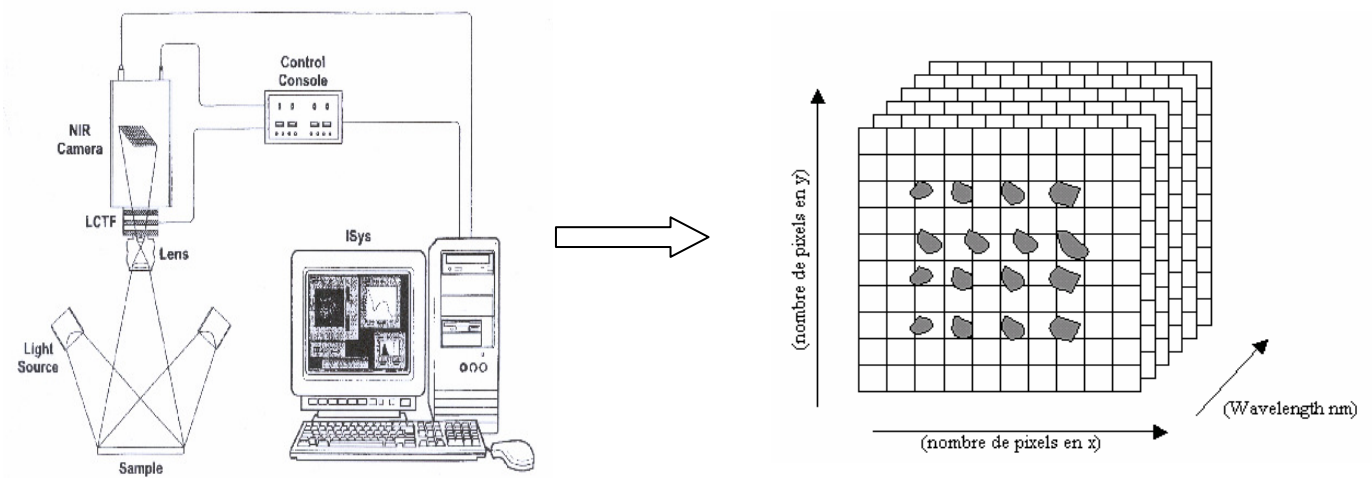

Figure 2: NIR imaging system and spectral volume obtained with it.

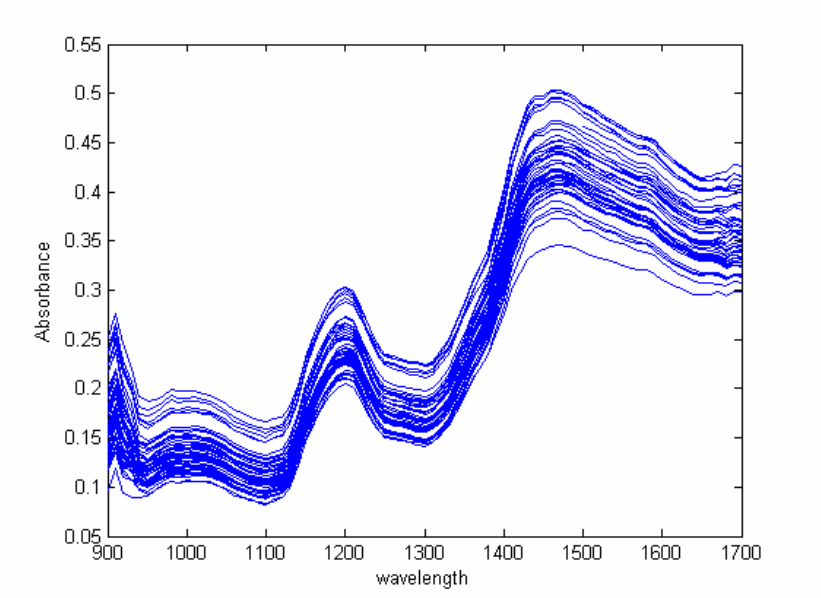

Figure 3: NIR spectra of soybean kernels obtained with the NIR camera. Each spectrum is the mean of the spectra from the pixels area covered by the grain.

The success of NIR imaging can be considered as a combination of different factors: high-performance and uncooled NIR sensitive focal plane array detector, digitally-tuneable infrared optical filters, the drastic increase in computer speed and the capacity of laboratory computing platforms. The integration of these elements shows promising results in the determination of quality parameters for complex matrices. Indeed, it allows to obtain at the same time spatial and spectral (and therefore chemical) information characterising the samples.

3.3 Chemometrics

Due to the large quantity of data obtained with the NIR camera, the use of Chemometrics becomes mandatory in order to get chemically relevant information out of measured physical data and to represent and display this information.

3.3.1 PCA

Principal Component Analysis is used to reduce the data dimensionality. PCA creates new orthogonal variables (scores) that are linear combinations of the original measured variables (absorbances at different wavelengths). The new variables are orthogonal, moreover, we assume that the first new variables or PCs, accounting for the majority of the variance of the original data, contain meaningful information, while the last ones, which account for a small amount of variance, only contain noise and can be ignored. PCA is performed into the data in order to make a first homogeneity study. The scores can highlight clustering, trends and outliers in the sample space in the data matrix.

3.3.2 PLS

PLS (Partial Least Squares) is used to construct models in order to discriminate between varieties or between GMO and non GMO samples. The programme is working by giving to the spectra dummy variables: 1 when the spectrum belongs to a certain group (i.e. GMO), 0 when it does not belong to this group. PLS searches for the directions where the gravity centres of the groups are the most separated, but with the constraint that the PLS factors and scores must be able to reconstruct the original spectra.

3.3.3 SVM

The foundations of Support Vector Machines (SVM) have been developed by Vapnik (Vapnik 2000), and the formulation embodies the Structural Risk Minimisation (SRM) principle. SRM minimises an upper bound on the generalisation error, as opposed to ERM (Empirical Risk Minimisation) which minimises the error on the training data. It is this difference which equips SVM with a greater potential to generalise, which is our goal in statistical learning.
In very simple terms, a SVM corresponds to a linear method in a very high dimensional feature space that is nonlinearly related to the input space. Even though we think of it as a linear algorithm in a high dimensional feature space, in practice, it does not involve any computations in that high dimensional space. By the use of kernels, all necessary computations are performed directly in input space.

4 Experimental and results

4.1 Samples and Images

Because the camera has a wavelength-dependent dark current, a dark current image has been recorded for each wavelength. Also a reference standard (back) of close to 100% reflectance has been used to make a reference image for each wavelength. Then the reflectance image at a certain wavelength has been calculated as:

$$R = \frac{sample - dark}{black - dark}$$

Each kernel was measured twice (both sides of the kernel) in order to include a large part of information. A large number of spectra has been collected coming from different origins, different varieties and some of them being transgenic. Once the kernels have been measured with the camera, two approaches can be applied: 1) working with all the pixels for each kernel (each kernel contains +/- 3500 pixels); or 2) working only with the mean spectrum of each kernel considering this mean as representative of the characteristics of the kernel. In the second case, in order to determine the mean of each kernel a mask has been constructed by a process of erosion in order to determine the contour of each grain. This mask is based on the different intensities found in the image. Then, the mean reflectance spectrum of a kernel has been supplied by averaging reflectance of the remaining pixels for that grain. Doing so a spectrum for each single kernel included in the image is obtained.

4.2 Analysis

In order to study the potential of NIR imaging in the GMO problematic, the CRA-W has analysed three different data sets:

Dataset 1: Roundup Ready soybean data measured on material coming from ILVO (Melle, Belgium) compared to non GM soybean kernels which to the best of our knowledge should be isogenic with the GM kernels,

Dataset 2: Barley data set coming from two different origins: a seed of five transgenic barley lines with their corresponding isogenic non GM lines coming from UK (Dr. Wendi Harwood/JIC/the Biotechnology and Biological Sciences Council (BBSRC)) and a group of barley samples coming from Ir. Jean-Luc Herman from the Department of vegetal production of the CRA-W, Gembloux.

Dataset 3: Roundup Ready soybean data set coming from the KelDa (Kernel Lot Distribution Assessment) project, a JRC-led project concerning the distribution of GM material in kernel lots imported within EU Member States (Paoletti *et al*., 2005) the evaluation of sampling strategies for the detection of GM materials in lots of bulk raw materials, and providing recommendations for implementing sampling strategies.
All datasets have been measured using classical Near Infrared Spectroscopy (NIRS) and NIR imaging. In the case of NIR imaging, for all datasets the mean of each kernel present in the images is used to construct the final databases.

4.2.1 Data set 1

Figure 5a shows an example of a NIR image of 24 soybean kernels. In this image two groups have been separated according to GM (left part of the image) and non GM samples (right part of the image). Figure 5b shows the NIR spectra for the whole dataset and a certain difference can be observed between both groups (GM and non GM in different colors in the figure). In total more than 160 spectra of soybean have been collected with the NIR camera coming from isogenic transgenic and non transgenic soybeans (GTS-40-3-2 event).

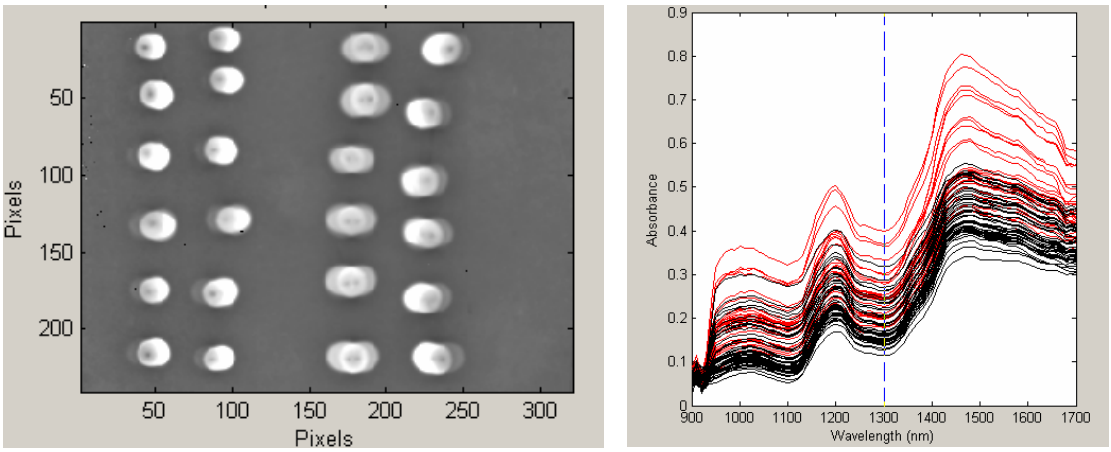

Figure 5: **a**) Example of a NIR image of soybean kernels obtained with the NIR camera and **b**) the whole mean spectra for dataset 1. Black spectra correspond to non GM samples and red spectra to GM samples.

In a first step and in order to visualise the data, Principal Component Analysis (PCA) has been performed on the data. PCA is used to reduce the data dimensionality. Figure 6 show the potential of PCA to indicate the possible discrimination between GM and non GM samples for a reduced set.

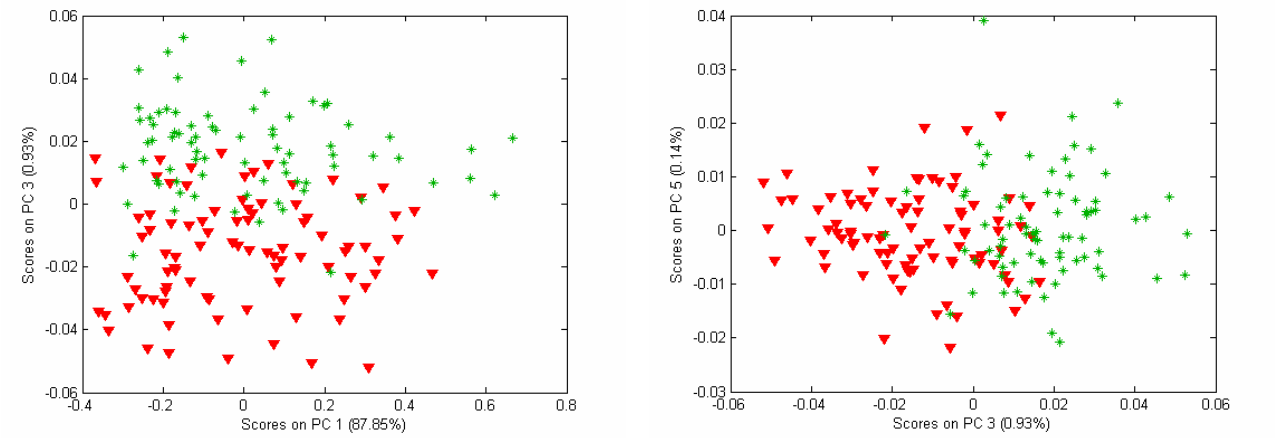

Figure 6: a) PC1 vs. PC2 for the soybean kernels; **b)** PC3 vs. PC5. In both figures the triangles represent the GM samples and the + the non GM samples.

Supervised techniques as PLS-DA have been also applied in a first tentative study to discriminate between GM and non GM samples. Discrimination models are tested using leave-one-out cross-validation (LOOCV) and an average error of 2% is obtained, which correspond to 98% of correct classification as average for each class. Also very low level of false positive results and false negative results are obtained (around 7%) which shows the potential of this technique for discrimination in a reduced set between GM and non GM samples. Figure 7 shows the LOOCV results.

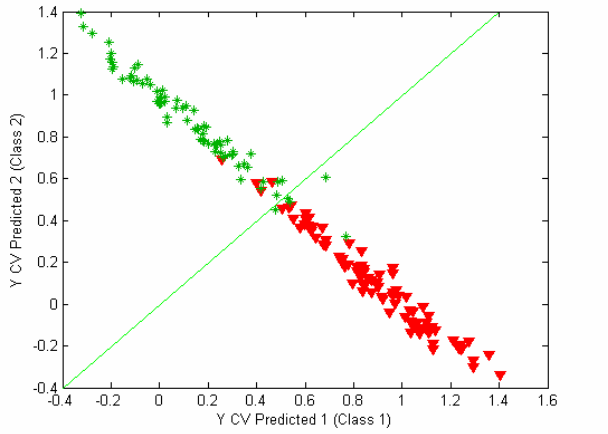

Figure 7: Cross-validation predicted value for the kernels belonging to the GM group (Y CV predicted 1 (Class 1)) vs. cross-validation predicted value for the kernels belonging to the non-GM group (Y CV predicted 1 (Class 2)). 7% of the samples are considered as false positive.

4.2.2 Data set 2

This work has been undertaken on material obtained from Dr. Wendi Harwood from the Biotechnology and Biological Sciences Council (BBSRC), UK. The data set consists on barley kernels measured with the NIR camera. In total more than 640 spectra have been collected coming from different origins, different varieties and some of them being transgenic (5 varieties including transgenic events with always their isogenic non GM counterpart). Figure 8a shows an image of 10 kernels of barley obtained with the NIR camera at a certain wavelength and Figure 8b is a mask performed on the previous figure. After the application of the mask, 10 NIR spectra by image are obtained (Figure 9).

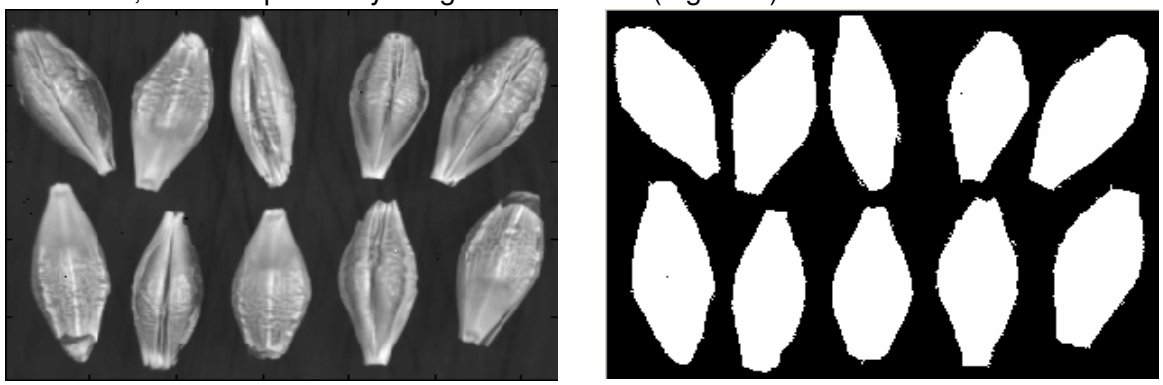

Figure 8: **a**) NIR image of barley grains obtained with the NIR camera, **b**) mask from the

previous image.

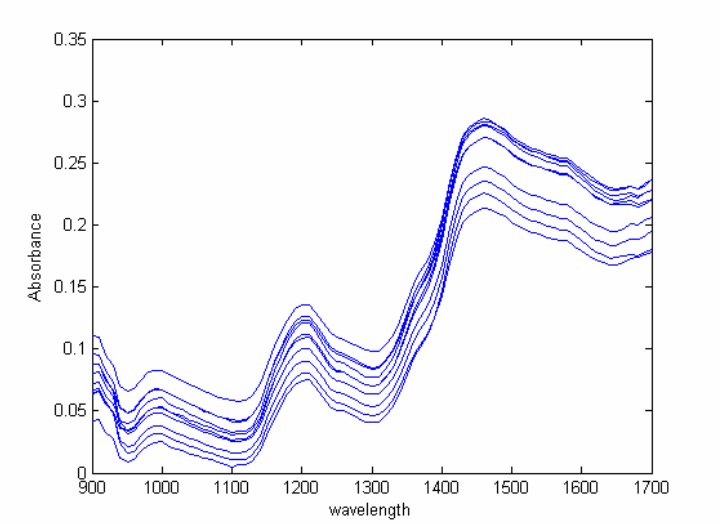

Figure 9: Mean NIR spectra obtained for each of the 10 grains represented in Figure 8a

As in the case of the soybean, PCA can be applied in order to visualise the data. Figure 10 shows the PC1 vs. PC6 for the whole data set. Samples marked as triangles represent the GM samples and the circles the non GM samples. No clear separation can be found but these figures puts in evidence that one variety (class A) is different from the rest.

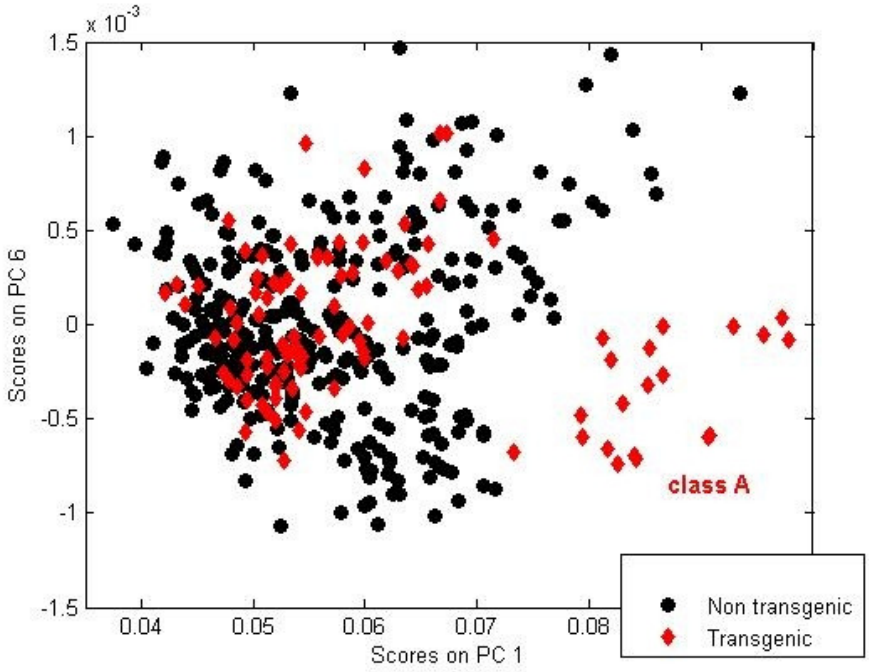

Figure 10: Classication of the barley kernels according to the first principal component (PC1) explaining 91.4 % of the variance and the sixth principal component (PC6) explaining 0.01 % of the variance observed in the data set.

Figures 11 and 12 show the results when working with the different varieties of barley samples. In these figures PCA models have been used to distinguish between the GM and

not GM samples for class B and class C. In both cases a clear separation is found between GM and not GM samples. The same observation is true for the remaining varieties.

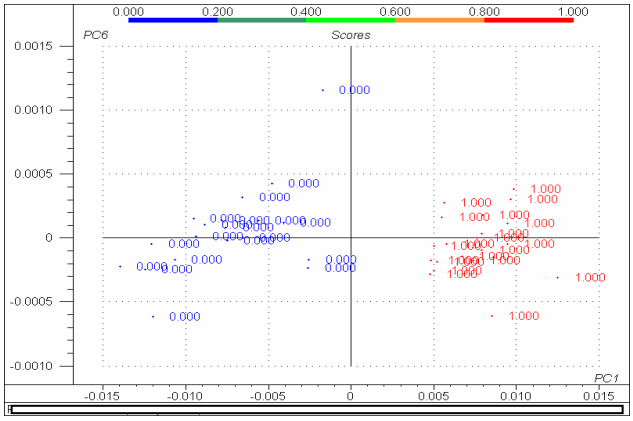

Figure 11: class B samples : first derivative 1: GM samples, 0: non GM samples

Figure 12: class C samples: 1: GM samples, 5: non GM samples

4.2.3 Data set 3

The third data set consists on more than 300 single kernel spectra of GM soybean coming from three different ships and measured with classical NIR. Figure 13 shows the PCA plot where the three different clusters corresponding to the different ships can be observed.

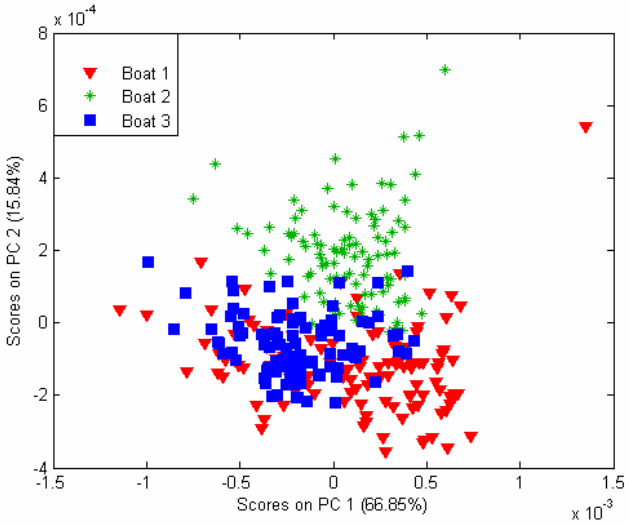

Figure 13: Classication of the KelDa soybean kernels according to the first principal component (PC1) explaining 66.85 % of the variance and the second principal component (PC2) explaining 15.84 % of the variance observed in the data set.

A first tentative study of calibration using the reference values for each ship has been performed. For that, an individual linear model (Multiple Linear Regression, MLR) for each ship has been constructed showing a correlation (R^2) of 0.5 as average.

5 Potential of the approach and perspectives

Taking all the results together, it is evident that the kernel study can be successfully done by using NIR and NIR hyperspectral imaging spectrometers and possiblities for the application of the method in the GMO detection become clear, at least for some traits or events. Briefly the studies undertaken by the CRA-W brought to the fore several points:

- It has been demonstrated that the high sensitivity and fast running ability of the NIR hyperspectral imaging spectrometer make it suitable for the kernel study. The data obtained with this imaging instrument agree well with those obtained by NIRS. The imaging spectrometer thus, possess all advantages of conventional NIR spectrometers. It has, however, additional features which conventional NIR spectrometers cannot offer, namely, its ability to provide spectra at different positions within a sample, i.e. to supply spatial information.
- NIRS and NIR hyperspectral systems present a fast answer and a good repeatability.
- The use of image processing techniques as edge or contour selection have permitted us to extract the maximum of information from the images as the determination of the mean spectra for each grain thanks to the creation of a mask based on that image.
- It is clear that the spectral information obtained can be of interest for the discrimination of samples not only for the variety discrimination but also for the presence of GMO on a reduced data set. This study shows that a separation between GM and non GM samples are possible. In the case of barley when working using individual varieties a clear separation between GM and non GM samples is obtained for almost every variety, that is not the case when working with all varieties

together showing that the 'within-groups' variability is higher than the variability between classes.
- The first tentative study using PLS-DA for discrimination between GM and non GM soybean kernels showed a high correct classification rate in cross-validation.
- When reference values are available calibration models can be constructed. The first model constructed in the case of dataset 3 showed that a small correlation between some spectral data and the GM level of an increment is obtained. In their paper, Roussel *et al.* (2001), explained that soybean composition has long been known to be environmentally variable; the component responsible for the GM effect is probably also environmentally variable. Because the magnitude of differences was not consistent throughout the sample set, a linear regression model might not accurately describe those differences. A more detailed study will be undertaken in this direction.
- In all data sets the first results have shown that a complete discrimination could be performed according to the variety and the presence of GM, and from the pattern recognition point of view more interesting approaches in order to make estimations of the statistical properties based on the images combined with the spectral information can be studied. This possibility will be the subject of a larger and more in depth study.

The first results of the research undertaken by the CRA-W in the framework of the CO-EXTRA project allow to define the potential ability in GMO detection:
- for reduced data sets as confirmatory method, i.e. to confirm the membership of a certain sample to a certain known variety.
- for larger data sets, like the soybean coming from different ships, as target method for sample selection.
- at the level of the control laboratory. This study has brought to the fore the difficulties to conclude if the variability observed between GM and non GM lines is higher than variability between varieties.

The CRA-W will continue the research in these three domains.

Annex 1

Important definitions

To resolve ambiguity between some biological terms like nucleus, seed, grain, kernel, or fruit, their definitions can be found hereafter:

Seed – From a botanic point of view it is the embryo of a plant. It contains stored food rich in oil or starch and protein and is wrapped in a more or less thick and hard seed coat. The seeds of land plants are generally contained in a hard and/or fleshy structure called a fruit. From a legal point of view a seed is a material meant for propagation of a plant variety (it may differ from the botanic definition for instance a potato tuber is also a seed in this sense while botanically it is not a seed). In this sense seed is considered as not being part of the food chain, it is the step just before.

Grain - It is the fruit of a cereal crop, technically called a caryopsis. Grain can also refer to other types of small seeds, though this is technically less correct.

Kernel - Kernel is used here as a generic term for grains and seeds.

Fruit – From a botanic point of view, we can distinguish 2 types of fruit: fleshy fruits and dried fruits.
For the **fleshy fruits**, the food part is the flesh part around the stone (drupe) or the pip (berry) including the seeds which have very often only a reproduction function. Two main exceptions are the fruit of the cocoa, the pod, which is a type of berry containing several cocoa grains or cocoa beans and the fruit of the coffee, the drupe, containing 2 coffee grains. On another hand, the **dried fruits** are widely used in food or feed. In the dehiscent fruit, the fruit opens spontaneously and discharge seeds following several processes (follicle, legume, silique, capsule). In the non dehiscent fruits (achene, caryopsis), the seed is closed to the outer layer.

In the **leguminous** family, including pulses crop (beans, peas, broad beans, lentils, lupines, vetch) but also some oil bearing crops (soybean, arachid), the fruit is a legume. The legume opens along a seam on 2 sides and contains several seeds with high content in protein (25%).

In the **brassicaceae** family, also called cruciferae family, including some oil bearing crops (rapeseed, canola) and spices crops (mustard), the fruit is a silique.

In the **pedaliaceae** family, also called sesame family, including some oil bearing crops (sesame), the fruit is a capsule.

In the **graminous** family, including cereal and some vegetables (green maize), the fruit (caryopsis), popularly named grain, is characterised by an outer layer joined to the seed. The caryopsis consist of an outer layer, the bran, including carbohydrate, vitamins and minerals, an underlying layer, the aleurone layer, including proteins (sources of lysine and tryptophan as amino acids) and phosphorus, an albumen , 80% of the grain size, including starch (90%) and gluten (10%) and a germ with high content in lipids.
Some plant families have achenes as fruits, it is characterized by an outer layer not joined to the seed. It contains one seed. It is for instance the case of the sunflower and the hazelnut.

6 Bibliography

- Ahmed, F. E. (2002). "Detection of genetically modified organisms in foods." Trends biotechnol 20(5): 215-223.
- Baeten, V. and P. Dardenne (2005). Application of near-infrared imaging for monitoring agricultural food and feed products. Spectrochemical analysis using infrared multichannel detectors. R. Bhargava and I. W. Levin, Blackwell Publishing: 30.
- Baeten, V., J. A. Fernandez Pierna and Dardenne, P. (2006). 'Hyperspectral imaging techniques : an attractive solution for the analysis of biological and agricultural materials. Techniques and Applications of Hyperspectral Image Analysis. H. F. G. P. G. Editors.
- Barker, M. and W. Rayens (2003). "Partial least quares for discrimination." Journal of Chemometrics 17: 166-173.
- FAO (2001). CONCEPTS, DEFINITIONS AND CLASSIFICATION: Crop statistics. Food Balance Sheets. Rome. Annex II: 12.
- FAO (2005). "Rice is life": International Rice Commission meets in Peru.
- Fernandez Pierna, J. A., V. Baeten, Michotte Renier, A., Cogdill, R. P. and Dardenne, P. (2005). "Combination of Support Vector Machines (SVM) and Near Infrared (NIR) imaging spectroscopy for the detection of meat and bone meat (MBM) in compound feeds." Journal of Chemometrics 18(7-8): 341-349.
- Fernandez Pierna, J. A., P. Volery, Besson, R., Baeten, V., Claustriaux, J.-J. and Dardenne, P. (2006). "Classification of modified starches by FTIR spectroscopy using support vector machines." Journal of Agricultural and Food Chemistry 53(17): 6581-6585.
- Huckerby, S., P. Winter and Symons, S. (2005). Système d'évaluation de la couleur des lentilles.
- Hurburgh, C. R., G. R. Rippke, Heithoff, C., Roussel, S. and Hardy, C., L. (2000). Detection of genetically modified grains by near infrared spectroscopy. PITTCON 2000 - Science for the 21st century, New Orleans, LA.
- Massart, D. L., B. G. M. Vandeginste, Deming, S. N., Michotte, Y. and Kaufman, L. (1988). Chemometrics: a textbook. Amsterdam, Elsevier Science Publishers B.V.
- Paoletti, C., Heissenberger, A., Mazzara, M., Larcher, S., Grazioli, E., Corbisier, P., Hess, N., Berben, G., Lübeck, P. S., De Loose, M., Moran, G., Henry, C., Brera, C., Folch, I., Ovesna, J., and Van den Eede, G. (2006). Kernel Lot Distribution Assessment (KeLDA): a study on the distribution of GMO in large soybean shipments. European Food Research and Technology in press, 11 pages.
- Roussel, S. and R. P. Cogdill (2004). NIR methods for GMO detection in food. Testing Genetically Modified organisms in Food.
- Roussel, S., C. Hardy, L., L., Hurburgh, C. R. and Rippke, G., R. (2001). "Detection of Roundup Ready ™ Soybeans by Near-Infrared Spectroscopy." Applied Specroscopy 55(10): 1425-1430.
- Vapnik, V. (2000). The nature of statistical learning theory. New York, Springer.
- Wang, W., J. Paliwal and Jayas, D. S. (2004). "Separation and Classification of Touching Grain Kernels Using Morphological Features." The Society for engineering in agricultural, food, and biological systems MB04-204: 9.