JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

J. Agric. Food Chem. 2009, 57, 8761–8769 8761
DOI:10.1021/jf9018323

Near Infrared Hyperspectral Imaging for the Evaluation of Endosperm Texture in Whole Yellow Maize (*Zea maize* L.) Kernels

Marena Manley,*,† Paul Williams,† David Nilsson,‡,§ and Paul Geladi§

†Department of Food Science, Stellenbosch University, Private Bag X1, Matieland (Stellenbosch) 7602, South Africa, ‡UmBio AB, Tvistevägen 48, Umeå, Sweden, and §Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, KBC huset, Linnaeus väg 6, SE 90187, Umeå, Sweden

Near infrared hyperspectral images (HSI) were recorded for whole yellow maize kernels (commercial hybrids) defined as either hard, intermediate, or soft by experienced maize breeders. The images were acquired with a linescan (pushbroom) instrument using a HgCdTe detector. The final image size was 570 × 219 pixels in 239 wavelength bands from 1000 to 2498 nm in steps of approximately 6.5 nm. Multivariate image cleaning was used to remove background and optical errors, in which about two-thirds of all pixels were removed. The cleaned image was used to calculate a principal component analysis (PCA) model after multiplicative scatter correction (MSC) and mean-centering were applied. It was possible to find clusters representing vitreous and floury endosperm (different types of endosperm present in varying ratios in hard and soft kernels) as well as a third type of endosperm by interactively delineating polygon based clusters in the score plot of the second and fourth principal components and projecting the results on the image space. Chemical interpretation of the loading line plots shows the effect of starch density and the protein matrix. The vitreous and floury endosperm clusters were used to make a partial least-squares discriminant analysis (PLS-DA) model, using four components, with a coefficient of determination (R^2) for the *y* data (kernel hardness category) for the training set of over 85%. This PLS-DA model could be used for prediction in a test set. We show how the prediction images can be interpreted, thus confirming the validity of the PCA classification. The technique presented here is very powerful for laboratory studies of small cereal samples in order to produce localized information.

KEYWORDS: Maize; kernel hardness; near-infrared; hyperspectral imaging; maize endosperm; commercial hybrids

INTRODUCTION

Maize (*Zea mays* L.) kernel hardness is principally a genetic expression, but environment and postharvest handling also have an influence on hardness properties (*1*). Maize is different from wheat in that both glassy (hard) and floury (soft) endosperm are found within a single kernel in a particular ratio (*1*). It is this ratio that determines whether the kernel is hard, soft, or intermediate (*1*, *2*). Hard kernels have predominantly vitreous (glassy) endosperm; soft kernels consist principally of floury endosperm, while intermediate kernels are expected to possess approximately equal quantities of both. The floury endosperm is usually localized toward the center of the kernel and close to the hull, whereas the vitreous endosperm is situated toward the sides of the kernel. The vitreous endosperm is tightly compacted with few or no air spaces. The starch granules are held together by a protein matrix, and protein bodies (zein) are found on the starch granules (*2*, *3*). The floury endosperm, however, comprises spherical starch granules that are covered with a protein matrix without zein bodies. Maize hardness is important to producers and processors in the grain trade (*1*, *4*, *5*) since it greatly influences end-use processing performance and processing of maize grits into certain foods.

Maize hardness has been assessed to date using kernel density (*4*, *5*); particle size index (PSI) (*4*); the Stenvert hardness tester (*7*); the tangential abrasive dehulling device (TADD) (*6*, *8*); and the rapid visco analyzer (RVA) (*6*). Most of these methods require the destruction of the sample.

Using some of the aforementioned techniques together with bulk near-infrared (NIR) spectroscopy and multivariate data analysis, predictions of maize hardness (*4*, *5*, *9*), dry-milling quality (*8*), and wet-milling starch yield were possible (*10*).

The problem of selecting the most appropriate method to determine maize hardness nondestructively either as a reference method for NIR calibration development or as a standalone method remains unresolved. The selected method should satisfy both maize breeders and industry.

A likely method to be employed could be NIR hyperspectral imaging that produces both localization information and a

*To whom correspondence should be addressed. Tel: +27 21 808 3511. Fax: +27 21 808 3511. E-mail: mman@sun.ac.za.

Published on Web 09/03/2009

Table 1. Comparison between NIR Hyperspectral Imaging and Bulk NIR Spectroscopy

instrument setup	NIR hyperspectral imaging	bulk NIR spectroscopy
sample preparation	ideally none	often grinding, homogenizing
sample holder	flat surface with dark or reflecting background	cup, petri dish, cuvette
illumination	as homogeneous as possible	may be heterogeneous
detector	Si, InGaAs, HgCdTe, InSb	Si, InGaAs, PbS
wavelength range	VIS-1000 nm; 900−1700 nm; 1000−2500 nm	VIS-2500 nm
depth of penetration	1−2 mm preferred	0.1 to tens of mm
scanning mode	reflection; fluorescence; transmission	reflection; transmission; transflection
measurement	focal plane; line scan; point scan	integrating; rotating cup; liquid cell
desired property	irregularities in samples, i.e., local damages; local infections; concentration gradients	concentration

complete spectrum in each pixel in the NIR wavelength region. It combines the strengths of imaging with those of bulk NIR spectroscopy. A hyperspectral image can be regarded as a stack of gray level images taken at different wavelengths or as an image where each pixel is a NIR spectrum (*11*−*13*). The objective of this study was thus to evaluate the potential of NIR hyperspectral imaging to distinguish between maize kernels (commercial hybrids) of different kernel hardness categories, i.e., hard, intermediate, or soft. Advantages of using this system successfully would be that, depending on the system being used, images of whole kernels can be acquired nondestructively in as little as 8 s, and although images of multiple kernels are collected, the results of single kernels can be obtained.

Hyperspectral images (HSI) are large in size and provide an excellent means of applying chemometrics techniques. A single hyperspectral image can consist of close to 100 000 spectra, and the only way of dealing with such large data sets is by applying multivariate data analysis (chemometrics) techniques. Data reduction is a good means of controlling the large amounts of data. In many cases, a principal component (PCA) model can reduce a hyperpectral image with >100 wavelength bands to less than 10 principal component (PC) score images. The importance of local models is evident from the fact that HSI have approximately 10^5 pixels. This means that even a subset of less than 1% of the images can become a data set with over 1000 spectra. A comparison between NIR hyperspectral imaging and bulk NIR spectroscopy is given in **Table 1**.

A number of reviews exist for the application of NIR hyperspectral imaging in the literature (*14*−*16*). The main applications include fruits and vegetables, meats and fish, and cereals for the region 400−1100 nm. In some cases, the InGaAs region from 900 to 1700 nm is used with a few applications using the mercury−cadmium−telluride (HgCdTe) wavelength range (1000−2500 nm). Linescan (pushbroom) is becoming the most important mode of imaging because of its speed and versatility, especially if real-time chemometrics algorithms are integrated in the solution.

HSI are three-way arrays with two pixel coordinate ways and one wavelength way (*11*−*13*). It is customary to reorganize these arrays into large matrices with pixels × wavelengths as the ways. One such matrix may be called **X**. For the example in this article, the three-way array is of the size $570 \times 219 \times 239$ and is reorganized into a $124\,830 \times 239$ matrix. When background, optical errors, and bad pixels are removed, the matrix becomes smaller than $124\,830 \times 239$, but the principles of the analysis shown below still remain valid. The same goes for the selection of regions of interest (ROI) or classes (*17*).

The matrix **X** can be subjected to PCA for the purpose of data reduction, usually after a spectral correction such as multiplicative scatter correction (MSC) (*18*) or standard normal variate (SNV) (*19*) and mean-centering. The equation for a PCA on a data matrix is:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad (1)$$

where **X** = (IxK) the matrix **X** after spectral pretreatment and mean-centering; **T** = (IxA) the score matrix with the score vectors as columns; **P** = (KxA) the loading matrix with the loading vectors as columns (**P**′ is the transpose of **P**); and **E** = (IxK) the residual matrix. The score vectors in **T** can be reorganized into images to form score images. They can also be shown as score plots with density contours.

Regression on images uses a matrix $\mathbf{X}_{train}$ and a vector of responses **y** for the training set. Usually, mean-centering of $\mathbf{X}_{train}$ and **y** is used.

$$\mathbf{y} = \mathbf{X}_{train}\mathbf{b} + \mathbf{f} \quad (2)$$

In eq 2, **y** = (Ix1) the vector of the response values, mean-centered; $\mathbf{X}_{train}$ = (IxK) the matrix version of the training image or subset, preprocessed and mean-centered; **b** = (Kx1) the vector of regression coefficients; and **f** = (Ix1) the vector of residuals.

The goal of any good regression model is to have a large enough proportion of the sum of squares (SS) of **y** in the model part **Xb** and a small part in the residual **f**. The coefficient of determination (R^2) for the training set is given as follows:

$$R^2 y = \mathrm{SS}(\mathbf{Xb})/\mathrm{SS}(\mathbf{y}) \quad (3)$$

Once a satisfactory model is found, it can be tested on a test set $\mathbf{X}_{test}$ as follows:

$$\mathbf{y}_{hat} = \mathbf{X}_{test}\mathbf{b} \quad (4)$$

where $\mathbf{y}_{hat}$ = predicted values for the test set, and $\mathbf{X}_{test}$ = the test set image or image subset reorganized into a matrix.

If the responses $\mathbf{y}_{test}$ for the test set are known, a residual can be calculated as follows:

$$\mathbf{f}_t = \mathbf{y}_{test} - \mathbf{y}_{hat} \quad (5)$$

where $\mathbf{f}_t$ = residual for the test set, and $\mathbf{y}_{test}$ = measured values for the test set.

$\mathbf{f}_t$ can be used to construct the root mean square error of prediction (RMSEP) as follows:

$$\mathrm{RMSEP} = [\mathbf{f}_t'\mathbf{f}_t/J]^{\frac{1}{2}} \quad (6)$$

where J is the number of test objects.

MATERIALS AND METHODS

Samples. Two sets of 18 whole yellow maize kernels, six each of three different genotypes and hardness categories, were randomly selected from three commercial hybrids, i.e., PHB 36K66 (soft); CRN 8010 (intermediate); and PHB 30F40 (hard).

Instrumentation and Image Acquisition. A sisuChema SWIR (short wave infrared) linescan imaging system, giving 231 pixels per line for 239 wavelength bands, was used (Specim, Spectral Imaging Ltd., Oulu, Finland). The sisuChema is an imaging spectrograph coupled to a 2-D array HgCdTe detector able to acquire a spectral range of 1000−2500 nm

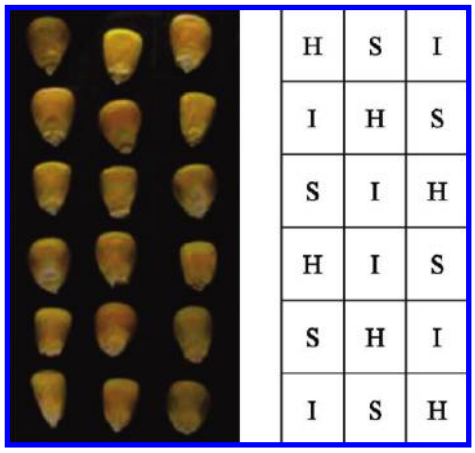

Figure 1. Digital image of maize kernels of different hardness (H = hard; I = intermediate; S = soft) showing a typical Latin Square layout (2nd sample set).

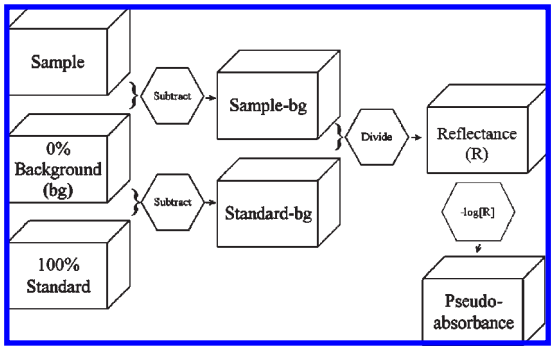

Figure 2. Schematic illustration of the conversion of reflection intensity to pseudoabsorbance (0% background = black reference; 100% standard = white reference).

with ca. 6.5 nm spacing. The HgCdTe camera (*f*/2.0; focal length = 31 mm) and electronics are fast, and data for a line can be collected in < 10 ms. The stepper motor mechanism of the linescan is the limiting factor needing about 15 ms per line. The number of lines is chosen before the scan and differs for each sample.

The two sets of 18 maize kernels were positioned flat, embryo down on silicon carbide sandpaper in a Latin Square (Sudoku) type design (**Figure 1**) and imaged separately. This type of pattern avoids incorrect classification because of uneven illumination. Eighteen was the maximum number of maize kernels that could fit effectively into the field of view (ca. 110×43 mm) of the sisuChema. However, due to the large number of spectra associated with each HSI (ca. 100 000), six maize kernels of each hardness type were adequate to allow the making of training and test sets. There are two sampling aspects here: 18 kernels are not adequate to technically describe the content of, e.g., a whole maize silo, truck load, or even 100 kg bag, but 18 kernels are adequate for studying differences between breeding lines on a laboratory scale.

Before each sample scan, black (with the camera shutter closed) and white (using a Spectralon reference) reference images were acquired. The images acquired with the sisuChema were converted to pseudoabsorbance using the black (0% reflectance standard or background) and white (100% reflectance standard) references, and a schematic illustration can be seen in **Figure 2**. The original image cube was 618 × 231 × 239, but after cropping to remove empty rows and columns, a 570 × 219 × 239 image was obtained. The pixel size was approximately 0.2 mm × 0.2 mm giving an image area of 114 mm × 43.8 mm. An image of the second set of 18 maize kernels was acquired in a manner similar to that used with the first one in order to test repeatability of the analysis process. This image was 679 × 229 × 239 after cropping.

Multivariate Image Analysis. Multivariate data analysis was carried out in Evince 2.020 (UmBio AB, Umeå, Sweden). This included the removal of dead pixels (elements of the detector array that give false or inadequate responses), calculation of absorbances, image cleaning, and removal of background in a PCA model (after mean-centering has been applied), and visual interactive classification in a multivariate model. The selection of pixels to be removed was made in the score plot. Since these plots can be used interactively, the score image was thus used to confirm that pixels from the maize kernels were not removed. After cleaning, MSC (mean spectrum of image used as target spectrum) and SNV preprocessing were applied to mean-centered data. **Figure 3** illustrates how a sequence of multivariate analysis operations combined with visual interactive selection leads to the finally selected classes. More details can be found in Geladi and Grahn (*20*) and Grahn and Geladi (*12*). Some calculations were carried out in MATLAB, version 7.5 (The MathWorks, Natick, MA) and using the PLS_Toolbox, version 5.0 (eigenvector Technologies, Wenatchee, WA). Finally, a partial least-squares discriminant analysis (PLS-DA) model for validation was built. Again, the analysis was performed after mean-centering, and MSC was applied to the raw data.

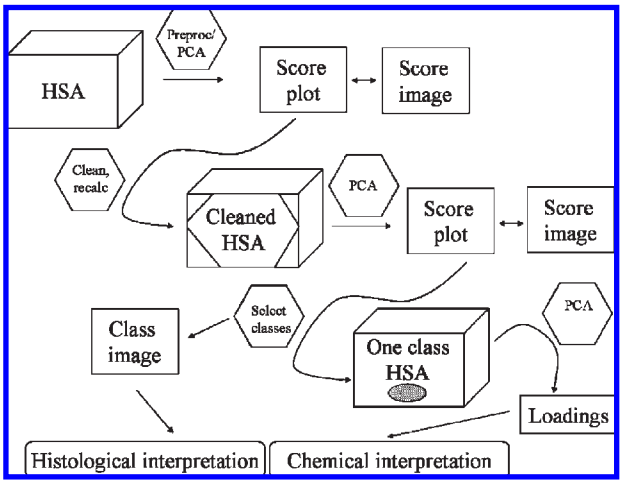

Figure 3. Sequence of multivariate data analysis operations used during the analysis of NIR hyperspectral image data of whole maize kernels (HSA = hyperspectral array; preproc = preprocessing; PCA = principal component analysis; clean = image cleaning; recalc = recalculate; class image = classification image).

RESULTS AND DISCUSSION

HSI are never error-free, and it was necessary to remove background, shading and geometrical errors, bad pixels, and specular reflection. The background and errors dominate in the HSI and would use up extra components if an analysis by PCA was attempted (*21*). Therefore, they had to be removed. Background removal was performed by making a three component PCA model (eq 1) and interactively identifying background and error components by going back and forth between score plot and score image. About two-thirds of all pixels were removed. **Figure 4** shows the image of the first sample set at 1102 nm with background included and after background removal. Errors due to shading, geometrical errors, bad pixels, and specular reflection are often shown as outliers in the score plot. Using the interactive nature between the score plot and score image, these outliers could be identified, as, for example, shading in the score image, and were removed.

The following nomenclature will be used when referring to maize kernels of different hardness categories and the different types of endosperm. Maize kernels of different hardness categories will be referred to as hard, intermediate, and soft, as labeled by the maize breeders. The endosperm types will be referred to as vitreous (glassy) and floury indicating hard and soft endosperm, respectively. Theoretically, it is expected that hard kernels would contain predominantly vitreous endosperm, soft

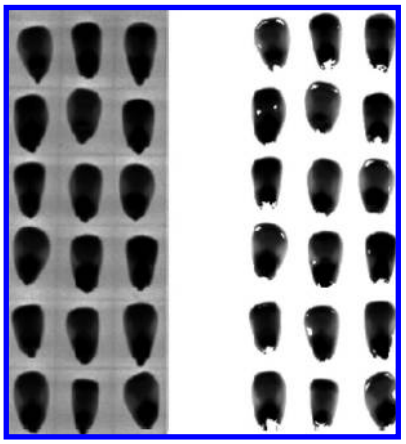

Figure 4. Image at 1102 nm with background (left) and cleaned image at 1102 nm (right). The physical size is 11.4 mm × 4.4 mm.

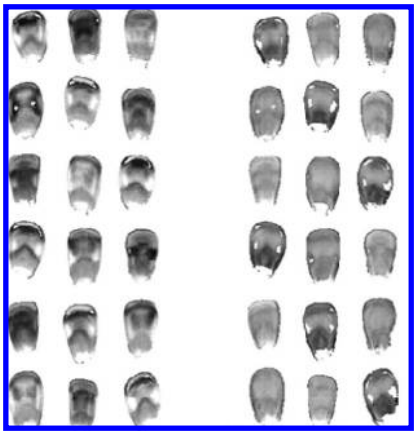

Figure 5. Score images t2 (left) and t4 (right) after removal of background and other disturbances such as geometrical errors and shadows (32% of the total imaged area was left after cleaning). Kernels with similar gray scale intensities would be similar in kernel hardness or endosperm texture.

kernels predominantly floury endosperm, and intermediate kernels the two types of endosperm in approximately equal amounts (*1*, *2*).

Some preliminary results obtained in the present study, without preprocessing of the spectra, were not optimal, and therefore, preprocessing (*22*) was applied. Success of the preprocessing methods was based on most distinct clusters being observed in the score plot. Both MSC and SNV preprocessing gave excellent results. Because of the similarity of the results obtained, it was not deemed necessary to show both sets of results. The results of either preprocessing method could have been shown; MSC was chosen for no particular reason. After MSC, to remove the effect of scattering differences, a five component PCA model (eq 1) was made on the cleaned HSI. Principal component (PC) sizes expressed as % SS were 91.3, 2.7, 1.9, 1.3, and 0.04, for the first image. Further components were too noisy to be included. Of the PCs calculated, the second and fourth PCs looked most promising. **Figure 5** shows the score images for t2 and t4 after background and shading removal. Similar score values or score values that would overlap in the score plot are shown with similar grayscale intensities in the score image. If a color heat map is used, similar score values will be observed in similar colors in the score image. Classification in score images can thus be observed on the basis of similar grayscale or similar color intensities. Score image t2 shows differences between the kernels of different hardness,

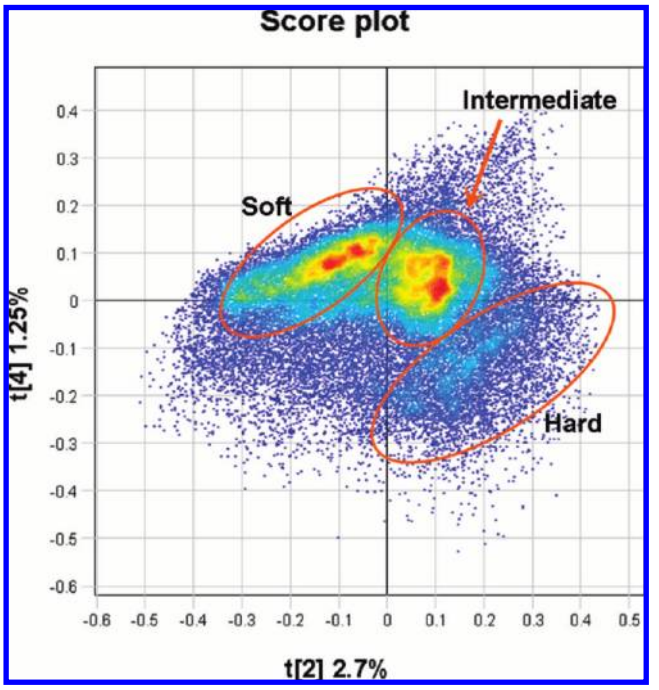

Figure 6. PCA score plot (t2 vs t4) of the first hyperspectral image (first sample set). The ellipses indicate clusters that are of the potentially soft, intermediate, and hard classes. A heat map is used to indicate the density of similar score values with blue indicating lowest density and red highest density.

while score image t4 shows darker areas for the kernels with predominantly vitreous endosperm. No differences in grayscale intensities based on hardness categories were noticeable in the other PC score images. It is, however, not optimal to determine such differences on score images alone. The use of score plots is more optimal (**Figure 6**). The score plot for t2 and t4 (for the first samples set) shows a rough delineation of three clusters potentially indicating the different hardness classes; these are indicated with ellipses in **Figure 6**.

A more correct description of the clusters is given in **Figure 7** where the classes were made by polygon marking on the score plot with interactive projection of the classes on the score image. By interactive evaluation between the score plot and score image, we identified two clusters as being vitreous and floury endosperm present in hard and soft kernels, respectively. The third cluster turned out to be endosperm predominantly present in the maize kernels identified as being of intermediate hardness. **Figure 7** shows the final selected classes in different colors in the score plot and projected on the score image giving a classification image, while unclassified pixels are shown in gray.

Figure 8 shows the score plot and classification image for the image of the second set of samples, showing similar results after PCA was applied, independent from the first image, to this second image. For the second image, PCs 2 and 4 were also selected, and the % SS values were 3.8% and 1.6%, respectively, for the two PCs. The agreement between the two sample sets indicates that the selection of classes was reasonably robust.

In the classification images (**Figures 7** and **8**), large green areas are observed in the kernels originally labeled as hard. These may be assumed to represent vitreous endosperm. Similarly, large blue areas in the kernels orginally labeled as soft may represent floury endosperm. This is a histological observation based on a priori knowledge from cereal science. This shows that hard and soft maize kernels constitute different ratios of vitreous and floury endosperm. Large red areas are seen mostly in the maize kernels

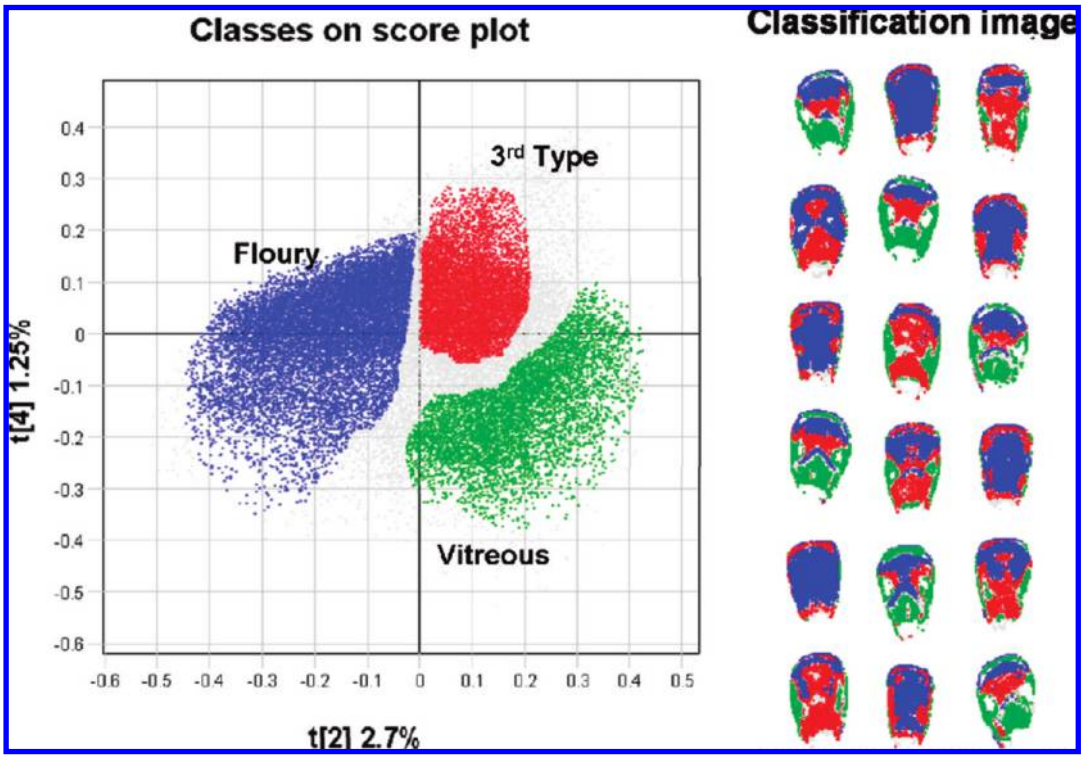

Figure 7. PCA score plot (left) of the first hyperspectral image (first sample set) with classification (green = vitreous; red = 3rd type endosperm; blue = floury) and the corresponding classification projected onto a score image (right). Vitreous = 15.6%, 3rd type = 28.6%, and floury = 39.9%.

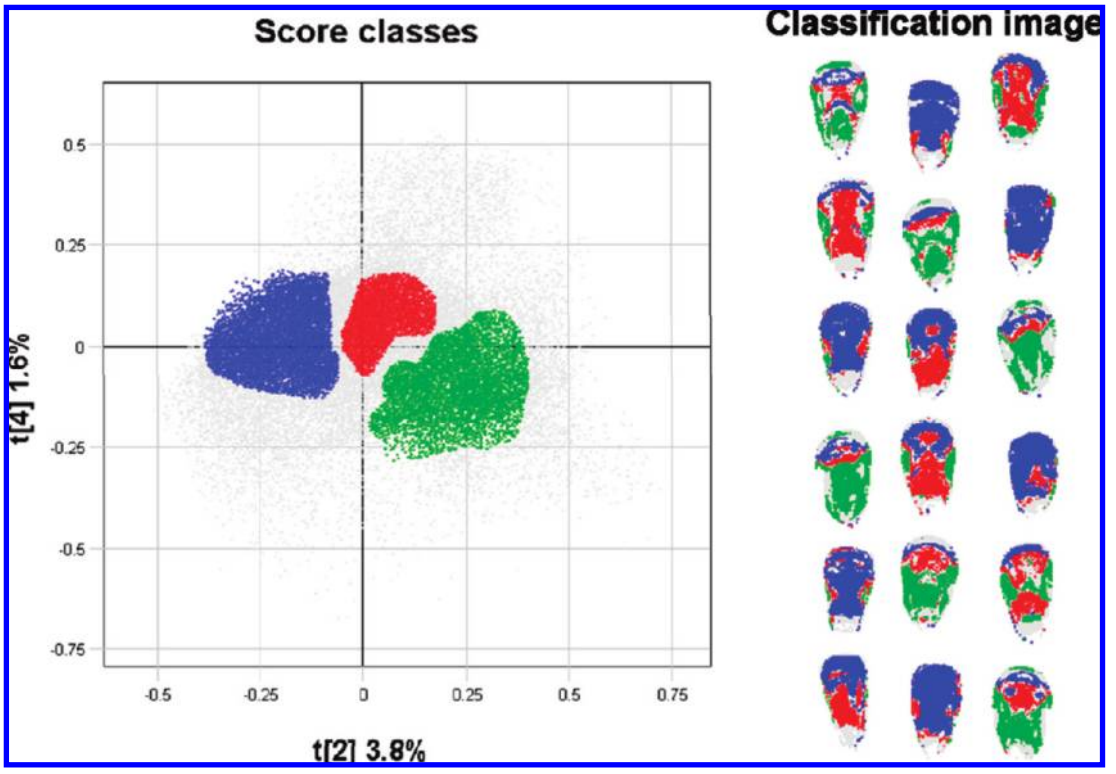

Figure 8. PCA score plot (left) of the second hyperspectral image (second sample set) with classification (green = vitreous; red = 3rd type endosperm; blue = floury) and the corresponding classification projected onto a score image (right).

labeled as being of intermediate hardness. This indicates the presence of a third type of endosperm different in physical properties and chemical composition from vitreous and floury endosperm. It is the first time a third endosperm type has been observed and reported, which was possible because of the combined spatial and spectral information offered by NIR hyperspectral imaging.

An important aspect of NIR hyperspectral image analysis is spectral interpretation. This interpretation is done by looking at PCA loadings, but it should not be done until pure classes are

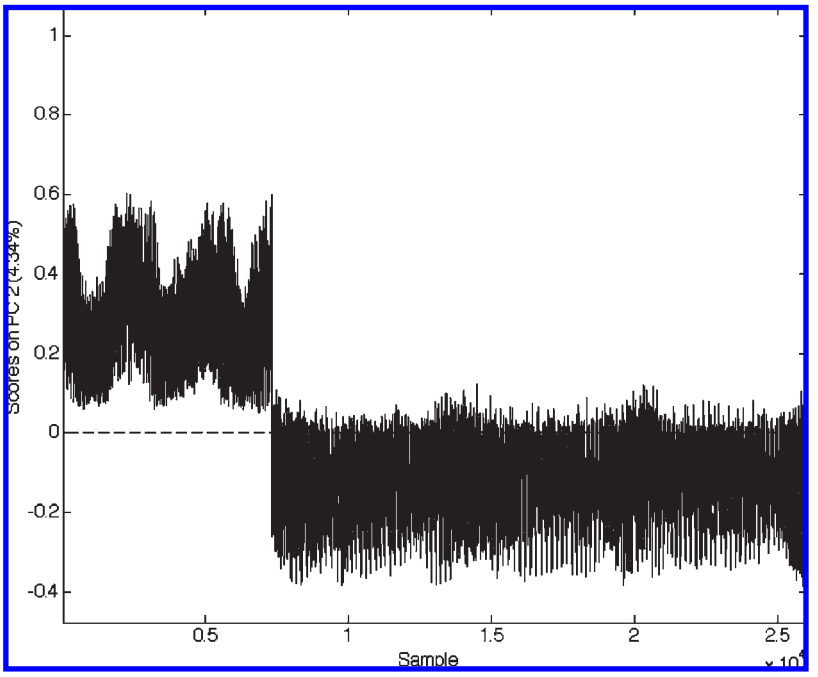

Figure 9. Score plot of principal component 2 for vitreous and floury endosperm classes of the first sample set after MSC correction. Score values for vitreous endosperm are shown in approximately the first third of the plot with the remainder of the plot showing the score values of the floury endosperm.

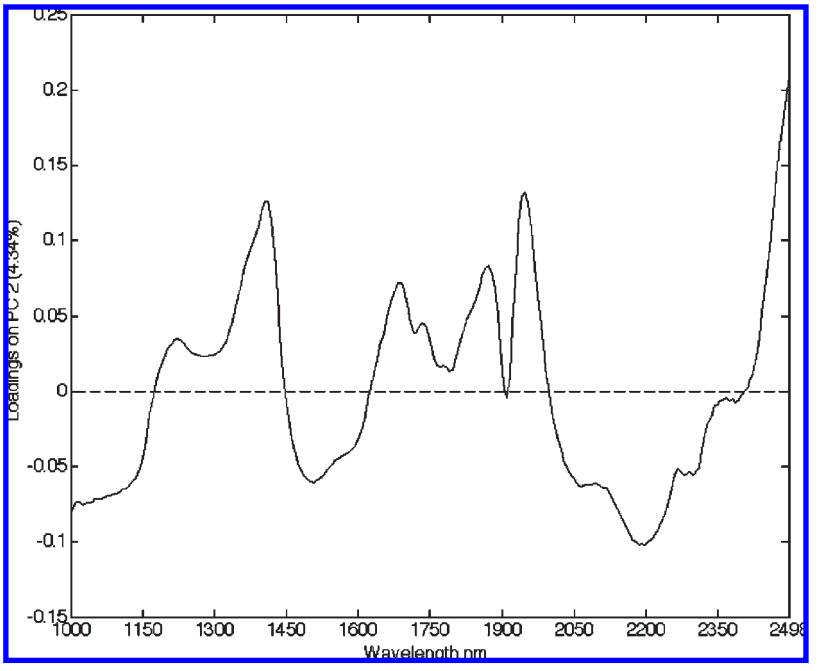

Figure 10. Loading line plot of principal component 2 for vitreous and floury endosperm classes of samples set one after MSC correction.

found. The loadings would simply become too complicated when more than three classes are present in the same model.

The interpretation of the loadings was, however, possible once the most optimal vitreous and floury classes were made. A data set was made of 7302 spectra of vitreous endosperm and 18 742 spectra of floury endosperm. A preliminary PCA analysis, after only mean-centering, showed that the first three components were based on scattering effects and that interesting details related to chemical composition only came in higher components. After MSC correction and mean-centering, a PCA model was developed giving 88.1% SS for the first and 4.3% SS for the second component. The second PC score shows almost perfectly separated vitreous and floury pixels (**Figure 9**). Because of this, the corresponding loading line plot in **Figure 10** gives a chemical interpretation of why these two pixel classes are different. The peaks at 1220, 1405, 1690, and 1870 nm indicate variation in endosperm texture due to differences in starch composition (*23*). These peaks are positive for the vitreous endosperm class, as is the moisture peak at 1944 nm which is clearly visible. A negative peak at 2195 nm indicates amino acids (*23*) related to the protein content and relates to the floury endosperm class. The presence of these peaks was expected as it is known that starch granules are held together by the protein matrix differently for vitreous and floury endosperm.

A PLS-DA regression model (eq 2) for the properties vitreous and floury endosperm was developed by setting dummy values of 1 for vitreous and 0 for floury endosperm with 0.5 as the cutoff value. Because of the clear clusters, this cutoff value would be suitable. In order to get a simple model, all pixels except those representing vitreous and floury endosperm were removed. This is an extreme form of ROI selection. Because of the expected correlation between wavelengths, it was necessary to use a partial least-squares (PLS) regression model (*24*). There were enough kernels to allow the selection of a training set, and a partition was made with 4 hard, intermediate, and soft kernels each in the training set, and 2 hard, intermediate, and soft kernels each in the test set. This small number of kernels was adequate since it gave 17 813 pixels (68.4%) in the training set and 8230 (31.6%) in the test set for the first image and 17 210 (49.1%) and 17 829 (50.9%) for the training and test sets, respectively for the second image.

As a rule of thumb, a PLS-DA model should give a coefficient of determination (R^2y) (eq 3) of 66% or more with less than 5 PLS components. The following R^2y values were obtained for the first 6 PLS components: 17.1%, 81.9%, 85.3%, 85.9%, 87.0%, and 87.3% after the PLS-DA model was developed for the first image. However, in order to avoid overfitting and on the basis of the classification observed in the score images of PC 4 between the kernels of different hardness, a model with 4 PLS components would suffice. This calibration model should also work well in the test set. This means that vitreous endosperm pixels in the test set are predicted as 1 and that floury endosperm pixels are predicted as 0. In reality, this becomes a distribution around 1 for vitreous and a distribution around 0 for floury. The region around 0.5 should remain empty. The R^2y values obtained for the first 6 PLS components of the second image were 90.5%.

Figure 11 shows the prediction image results $\mathbf{X}_{train}\mathbf{b}$ (training set) (eq 2) and $\mathbf{y}_{hat}$ (test set) (eq 4) for the 4 PLS component model (first sample set). We found that more than 4 PLS components gave no visual improvement and would therefore constitute an overfit.

The *y*-variable is color coded in the prediction image, making 1 (the dummy variable for vitreous endosperm) cyan or gray and 0 (the dummy variable for floury endosperm) magenta or black. For the left image (**Figure 11a**), the training set is in color and the test set in grayscale, while for the image at the right (**Figure 11b**), the test set is in color and the training set in grayscale. The presence of these endosperm types in the prediction images is thus indicated by large areas of cyan/gray or magenta/black.

From **Figure 11**, it is clear that floury endosperm (magenta or black) covers the largest area of the endosperm of kernels classified as soft, followed by the intermediate kernels with the hard kernels constituting almost no floury endosperm. Within the intermediate kernels, some floury endosperm tends to be localized in the center of the kernel with small areas of vitreous endosperm (cyan or gray) noticed on the sides of the kernels. This agrees with standard cereal literature (*1*, *2*). The vitreous endosperm (cyan or gray) areas in the hard kernels are localized at the bottom of the kernel above the germ (on the basis of imaging position). On the basis of the chemical information from the image, these areas are similar to the areas of vitreous endosperm noticed on the sides of the kernels. The prediction image, however, only explains the composition of the kernels in terms of vitreous and floury endosperm types.

In order to clarify the presence of the third endosperm type observed in the score plots and score images, this class was added to the test set and its $\mathbf{y}_{hat}$ also calculated. **Figure 12** shows the results of this operation for the images of both the samples sets. In **Figure 11**, only magenta/black and cyan/gray areas were observed referring to floury and vitreous endosperm, respectively. A third

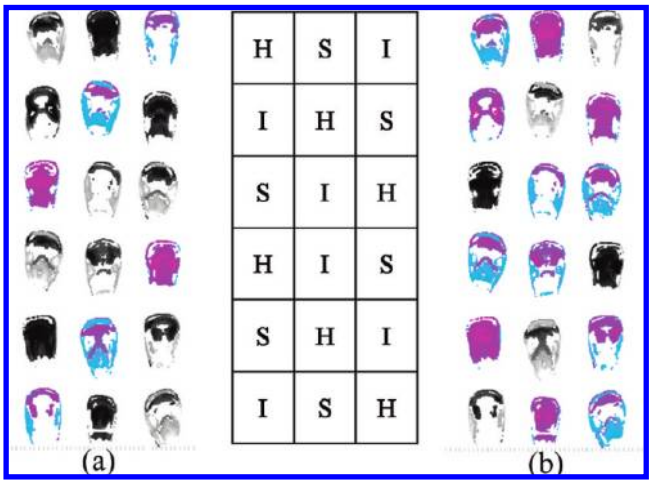

Figure 11. (**a**) Training set in color with test set in grayscale and (**b**) test set in color with training set in grayscale of a four component PLS-DA model for classification between vitreous and floury endosperm. These are the 18 maize kernels of only the first sample set. The pixels for the third type of endosperm observed are not included in the test set. Identification of the kernels is given in the table. Legend: magenta or black = floury; cyan or gray = vitreous.

color, i.e., purple now emerges (**Figure 12**) in both the training and test sets, representing the third endosperm type, thus indicating this class is neither vitreous nor floury. This confirms what was illustrated earlier in the PCA analysis, i.e., the presence of a third type of endosperm.

This example also shows additional information about penetration depth. It was possible to see the endosperm through the pericarp, and it shows that penetration depth is down to the endosperm; the fact the embryo below was not seen shows that the penetration depth is not that deep. The pericarp thickness is 0.1 mm, and the distance to the embryo is 1−2 mm. From this, it can be assumed that the observed endosperm types as seen in **Figures 7**, **8**, and **12** are up to about 1−2 mm under the surface of the kernel.

It is essential that NIR HSI are cleaned efficiently to ensure analysis of a high quality. This makes the removal of background, bad pixels, and any geometrical, physical, and/or optical disturbances imperative. The interactive choice of clusters/classes between the score plots and score images is an extremely versatile tool. Although it is subjective and not completely reproducible, it has been found to be acceptable if performed by a trained researcher. Because of the huge number of pixels in a score plot, it is almost impossible to select the same classes repeatedly. Chemometrics models change when clusters are chosen differently, but they are still robust if histological background information is taken into account. It was possible to clearly delineate vitreous and floury endosperm classes. These classes are present in all kernels but in different ratios. A surprising third endosperm class was observed. This type of endosperm has physical and chemical properties different from those of the vitreous and floury endosperm. It is possible that this third class of endosperm consists of a mixture or layers of floury and vitreous endosperm. The exact properties of this third class still need to be confirmed.

The interpretation of loading line plots allows for chemical interpretation of the differences between vitreous and floury endosperm. These seem to relate to the density of the starch and the associated protein matrix.

The PLS-DA model with a high R^2y (> 85%) confirms that the correct classes for vitreous and floury endosperm were chosen. The PLS prediction ($\mathbf{y}_{hat}$) image could as easily be interpreted,

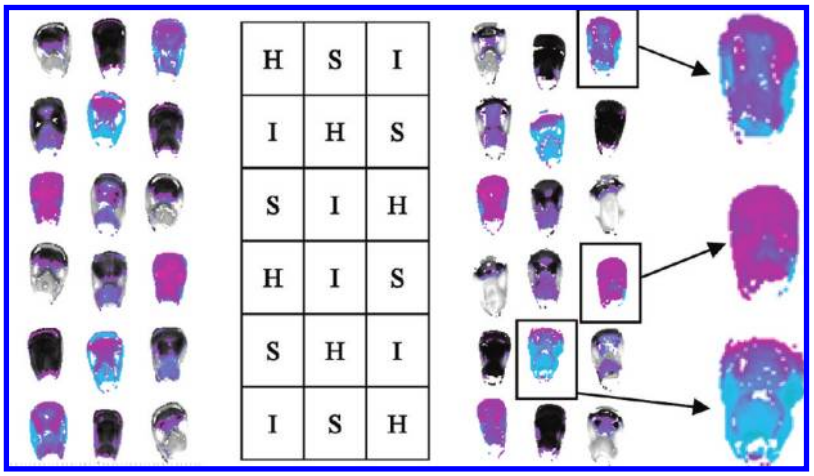

Figure 12. Prediction results (left = 1st sample set and right = 2nd sample set) for the vitreous and floury PLS-DA model. Training set in color with test set in grayscale are shown. Legend: magenta or black = floury; blue or gray = vitreous; purple = third endosperm type. The labels of the kernels are indicated in the table. Note that the pixels for the third type of endosperm observed (purple) are included in the training as well as the test set. Enlargements of three kernels (soft, intermediate, and hard) of the test set have been included to more clearly illustrate the difference in colors.

from the known distribution of vitreous and floury endosperm within a maize kernel, as the PCA classification. The presence of the third endosperm class was also confirmed in the PLS-DA prediction images.

Penetration depth in this study was estimated to be 1−2 mm, as the germ below was not noticeable in the image, but will differ for other food products. For future studies, a better understanding of the three-dimensional structure of the maize kernels would be advantageous. This should include studying kernels differing in shape and thickness. Currently, NIR hyperspectral image analysis, as applied in this study, is only suitable for laboratory studies. It is, however, a powerful technique that could be used for cereal grains in order to produce localized information. Because of its high speed of analysis, it would be of great value when thousands of samples need to be analyzed for endosperm texture in a maize breeding program.

ABBREVIATIONS USED

HSI, hyperspectral images; MCT, mercury−cadmium−telluride; MSC, multiplicative scatter correction; NIR, near-infrared; PC, principal component; PCA, principal component analysis; PLS, partial least-squares; PLS-DA, partial least-squares discriminant analysis; PSI, particle size index; ROI, regions of interest; RVA, rapid visco analyzer; SWIR, short wave infrared; SNV, standard normal variate; TADD, tangential abrasive dehulling device.

ACKNOWLEDGMENT

We thank Julian White (Specim, Spectral Imaging Ltd, Oulu, Finland) for the use of the sisuChema imaging system, Oskar Jonsson (UmBio AB, Umeå, Sweden) for the use of Evince software, and Pioneer Seed (Delmas, South Africa) for supplying and labelling samples.

LITERATURE CITED

(1) Watson, S. A. Structure and Composition. In *Corn: Chemistry and Technology*; Watson, S. A., Ramstad, P. E., Eds.; American Association of Cereal Chemists, Inc.: St. Paul, MN, 1987; pp 53−82.

(2) Hoseney, R. C. *Principles of Cereal Acience and Technology*; American Association of Cereal Chemists, Inc.: St. Paul, MN, 1994; p 378.

(3) Lee, K. M.; Bean, S. R.; Alavi, S.; Herrman, T. J.; Waniska, R. D. Physical and biochemical properties of maize hardness and extrudates of selected hybrids. *J. Agric. Food. Chem.* **2006**, *54*, 4260–4269.

(4) Pomeranz, Y.; Martin, C. R.; Taylor, D. D.; Lai, F. S. Corn hardness determination. *Cereal Chem.* **1984**, *61*, 147–150.

(5) Lee, K. M.; Herrman, T. J.; Lingenfelser, J.; Jackson, D. S. Classification and prediction of maize hardness-associated properties using multivariate statistical analyses. *J. Cereal Sci.* **2005**, *41*, 85–93.

(6) Almeida-Dominguez, H. D.; Suhendro, E. L.; Rooney, L. W. Factors affecting rapid visco analyser curves for the determination of maize kernel hardness. *J. Cereal Sci.* **1997**, *25*, 93–102.

(7) Pomeranz, Y.; Czuchajowska, Z.; Martin, C.; Lai, F. Determination of maize hardness by Stenvert hardness tester. *Cereal Chem.* **1985**, *62*, 108–112.

(8) Wehling, R. L.; Jackson, D. S.; Hamaker, B. R. Prediction of corn dry-milling quality by near-infrared spectroscopy. *Cereal Chem.* **1996**, *73*, 543–546.

(9) Eyherabide, G.; Robutti, J.; Borras, F. Effect of near-infrared transmission-based selection on maize hardness and the composition of zeins. *Cereal Chem.* **1996**, *73*, 775–778.

(10) Wehling, R. L.; Jackson, D. S.; Hooper, D. G.; Ghaedian, A. R. Prediction of wet-milling starch yield from corn by near-infrared spectroscopy. *Cereal Chem.* **1993**, *70*, 720–723.

(11) Goetz, A.; Curtiss, B. Hyperspectral imaging of the earth: Remote analytical chemistry in an uncontrolled environment. *Field Anal. Chem. Technol.* **1996**, *1*, 67–76.

(12) Grahn H. F.; Geladi P. *Techniques and Applications of Hyperspectral Image Analysis*; Grahn H. F., Geladi, P., Eds.; John Wiley & Sons Ltd.: Chichester, U.K., 2007; p 368.

(13) Geladi, P.; Burger, J.; Lestander, T. Hyperspectral imaging: Calibration problems and solutions. *Chemom. Intell. Lab. Syst.* **2004**, *72*, 209–217.

(14) Du, C.; Sun, D. Recent developments in the applications of image processing techniques for food quality evaluation. *Trends Food Sci. Technol.* **2004**, *15*, 230–249.

(15) Gowen, A. A.; O'Donnell, C. P.; Cullen, P. J.; Downey, G.; Frias, J. M. Hyperspectral imaging: An emerging process analytical tool for food quality and safety control. *Trends Food Sci. Technol.* **2007**, *18*, 590–598.

(16) Wang, W.; Paliwal, J. Near-infrared spectroscopy and imaging in food quality and safety. *Sens. Instrum. Food Qual. Saf.* **2007**, *1*, 193–207.

(17) Burger, J.; Geladi, P. Hyperspectral NIR image regression part I: Calibration and correction. *J. Chemom.* **2005**, *19*, 355–363.

(18) Geladi, P.; Macdougall, D.; Martens, H. Linearizaton and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* **1985**, *39*, 491–500.

(19) Barnes, R. J.; Dhanoa, M. A.; Lister, S. J. Standard normal variate transformatoin and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777.

(20) Geladi, P.; Grahn, H. F. *Multivariate Image Analysis*; John Wiley & Sons Ltd: Chichester, U.K., 1996; p 316.
(21) Burger, J.; Geladi, P. Hyperspectral NIR image regression Part II: Dataset preprocessing diagnostics. *J. Chemom*. **2006**, *20*, 106–119.
(22) Burger, J.; Geladi, P. Spectral pre-treatments of hyperspectral near infrared images: Analysis of diffuse reflectance scattering. *J. Near Infrared Spectrosc*. **2007**, *15*, 29–38.
(23) Osborne, B. G.; Fearn, T.; Hindle, P. H. *Practical NIR Spectroscopy with Practical Applications in Food and Beverage Analysis*, 2nd ed.; Longman Scientific and Technical: Harlow, U.K., 1993; p 220.
(24) Esbensen, K. *Multivariate Data Analysis in Practice*; Camo: Oslo, Norway, 1994; p 598.

Received May 29, 2009. Revised manuscript received August 5, 2009. Accepted August 22, 2009. The South African-Swedish Research Partnership Programme Bilateral Agreement, National Research Foundation (NRF), South Africa (UID 60958 & VR 348-2006-6715) provided funding for exchange of researchers, and the NRF, South Africa provided funding for running costs (FA2006032900007).