Analytica Chimica Acta 653 (2009) 121–130

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis

Paul Williams[a], Paul Geladi[b], Glen Fox[a,c], Marena Manley[a,*]

[a] Department of Food Science, Stellenbosch University, Private Bag X1, Matieland (Stellenbosch), 7602, South Africa
[b] Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, KBC Huset, Linnaeus vaeg 6, SE 901 87 Umeå, Sweden
[c] Department of Primary Industries and Fisheries, Queensland Grains Research Laboratory, PO Box 2282, Toowoomba, Queensland 4350, Australia

ARTICLE INFO

Article history:
Received 15 May 2009
Received in revised form 29 August 2009
Accepted 2 September 2009
Available online 6 September 2009

Keywords:
Near infrared hyperspectral imaging
Hyperspectral image analysis
Maize hardness
Principal component analysis
Partial least squares discriminant analysis

ABSTRACT

The use of near infrared (NIR) hyperspectral imaging and hyperspectral image analysis for distinguishing between hard, intermediate and soft maize kernels from inbred lines was evaluated. NIR hyperspectral images of two sets (12 and 24 kernels) of whole maize kernels were acquired using a Spectral Dimensions MatrixNIR camera with a spectral range of 960–1662 nm and a sisuChema SWIR (short wave infrared) hyperspectral pushbroom imaging system with a spectral range of 1000–2498 nm. Exploratory principal component analysis (PCA) was used on absorbance images to remove background, bad pixels and shading. On the cleaned images, PCA could be used effectively to find histological classes including glassy (hard) and floury (soft) endosperm. PCA illustrated a distinct difference between glassy and floury endosperm along principal component (PC) three on the MatrixNIR and PC two on the sisuChema with two distinguishable clusters. Subsequently partial least squares discriminant analysis (PLS-DA) was applied to build a classification model. The PLS-DA model from the MatrixNIR image (12 kernels) resulted in root mean square error of prediction (RMSEP) value of 0.18. This was repeated on the MatrixNIR image of the 24 kernels which resulted in RMSEP of 0.18. The sisuChema image yielded RMSEP value of 0.29. The reproducible results obtained with the different data sets indicate that the method proposed in this paper has a real potential for future classification uses.

1. Introduction

Maize kernel hardness is principally a genetic expression, but environment and post-harvest handling (e.g. transportation, drying and storage) also have an influence on hardness properties [1]. Hardness is related to kernel density, bulk density, breakage susceptibility caused by drying, storage, handling and processing. Maize kernels constitute both glassy and floury endosperm which are found within a single kernel in a particular ratio [1–3]. It is this ratio that determines the inherent hardness of the kernel [3]. Hard kernels have predominantly glassy endosperm; soft kernels consist principally of floury endosperm while intermediate kernels are expected to possess approximately equal quantities of both. Maize protein does play a strong role in influencing the mechanical strength of endosperm, dependent presumably on its ability to pack between the starch granules, and to adhere to the granule surfaces. The glassy endosperm is tightly compacted with few or no air spaces. The starch granules are held together by the protein matrix and protein bodies are found on the starch granules [2,4]. The floury endosperm on the other hand comprises spherical starch granules that are covered with a protein matrix without zein (maize storage protein) bodies.

Maize hardness is important to producers and processors in the grain trade [1,5,6] since it greatly influences end-use processing performance, including dry-milling yield and power requirements; dust formation during processing; and processing of maize grits into certain foods.

Most methods, assessed to date, to determine maize hardness require destruction of the sample. Near infrared (NIR) spectroscopy [5,7,8] can be used to estimate maize hardness. Other destructive methods include density [5,9]; particle size index (PSI) or particle size analysis (PSA) [5]; the Stenvert hardness tester [10]; the tangential abrasive dehulling device (TADD) [8,9]; and the rapid visco analyser (RVA) [9]. These methods have been reviewed comprehensively [11].

Using some of the aforementioned techniques as reference methods, NIR instruments can be calibrated to predict dry-milling quality characteristics of maize with a reliability suitable for rough screening [8]. Additionally, wet-milling starch yield, with a bias-corrected standard error of prediction of 1.41%, was observed using NIR spectroscopy [12]. In another study, the classification and prediction of maize hardness properties into hardness clusters using

* Corresponding author. Tel.: +27 21 808 3511 fax: +27 21 808 3510.
E-mail address: mman@sun.ac.za (M. Manley).

0003-2670/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2009.09.005

discriminant analysis showed a classification rate of 87% [6]. The problem with selecting the most appropriate method to determine maize hardness, either as a reference method for NIR calibration development or as a standalone method, remains unresolved. It is important that the method to be selected satisfy all potential users.

NIR hyperspectral imaging is a powerful spectroscopic technique which is capable of capturing images at many wavelengths in the NIR region [13–18] and has been used in a number of applications in wheat [19–23]. Reasonable moisture content calibrations have been obtained from images of single maize kernels measured in transmission mode [18]. Using diffuse reflection mode, oil and oleic acid could be quantified in individual whole maize kernels with the prediction of oil content more accurate than that of oleic acid [24].

A NIR hyperspectral imaging system collects image data by arranging it into a three-way data matrix, known as a hypercube [13,18]. The first two axes (x and y) of the matrix are the vertical and horizontal pixel coordinates while the third (z) axis depicts the spectral dimension. Hundreds of single channel black and white (grayscale) images are stacked on top of each other to produce hyperspectral images (or hypercubes) [13]. Each of these grayscale images represents a single band of spectral wavelength. Typically, a commercial NIR imaging instrument produces hypercubes with dimensions $256 \times 320 \times 118$, *i.e.* 118 single channel images each with 256×320 pixels.

Despite posing data extraction challenges, this vast accumulation of data creates novel possibilities. NIR hyperspectral imaging has been evaluated for the classification of sound and stained wheat grains with an overall accuracy of 95% over the 420–2500 nm range, as well as for the reduced ranges of 420–1000 nm and 420–700 nm [25]. In another study, NIR hyperspectral imaging revealed sprouting for single wheat kernels with sensitivity greater than the human eye [23]. The discrimination between vitreous and non-vitreous wheat kernels with a classification rate of 94% has been achieved using NIR hyperspectral imaging with a wavelength range of 650–1100 nm [19] and single kernel maize analysis by hyperspectral transmittance in the range of 750–1090 nm was carried out for the development of predictive calibrations for moisture (SECV 1.20% and RPD 2.74) and oil (SECV 1.38% and RPD 1.45) content [18].

Multivariate image analysis [26] or, lately, hyperspectral image analysis [27] is helpful for image data exploration, classification and quantification. As with single-point NIR spectroscopy, chemometrics techniques are applied to the data (image data set in this case) to decompose the data, pre-process the data and perform regression or classification analyses. Principal component analysis (PCA) was evaluated for the investigation of visible and NIR multispectral images of works of art [28]. In a recent study NIR hyperspectral imaging combined with PCA has been evaluated for bruise damage detection on white mushrooms [29]. Partial least squares discriminant analysis (PLS-DA) has been assessed in multivariate image analysis to discriminate food products of different natures with classification rates of 75%, 83%, 98% and 89% for maize, pea, soya bean meal and wheat respectively [30]. In a NIR hyperspectral imaging study evaluating commercial maize hybrids, three distinct clusters related to endosperm texture were observed [31]. This was in contrast to literature indicating that endosperm of maize kernels of different hardness categories comprise different ratios of only two types of endosperm, i.e. glassy and floury. The question arose if the endosperm texture of inbred lines, used in breeding programmes would also comprise these three endosperm types or only two as expected.

The aim of this study was to determine whether NIR hyperspectral imaging could distinguish between whole maize kernels (inbred lines) of varying hardness by evaluating the use of PCA and the interpretation of principal component (PC) loading line plots to explain the chemical variation within and between the maize kernels of different hardness; and PLS-DA as a possible chemometrics classification technique.

2. Experimental

2.1. Samples

Maize kernels of varying degrees of hardness were provided by Pioneer Hi-Bred Research RSA (P.O. Box 699 Delmas, 2210 South Africa). These whole kernels were randomly selected from yellow maize inbred lines grown in breeding trials. The maize samples consisted of three categories of hardness, i.e. hard, intermediate and soft. These descriptions of the hardness categories were provided by experienced maize breeders.

2.2. NIR hyperspectral imaging systems

NIR hyperspectral images were acquired using a Spectral Dimensions MatrixNIR camera (Malvern Instruments Ltd., Malvern, Worcestershire, UK) and a sisuChema SWIR (short wave infrared) hyperspectral imaging system (Specim, Spectral Imaging Ltd., Oulu, Finland).

The Spectral Dimensions MatrixNIR camera comprised an InGaAs diode array detector with LCTF bandpass filter enabling wavelength selection. Four current controlled quartz-halogen lamps provided illumination (Gilway 12 V 75 W). The image field of view was 50 mm × 62 mm. The spectral range recorded was 960–1662 nm at increments of 6 nm, producing hyperspectral images with 256×320 pixels at 118 wavelength channels.

The sisuChema comprised an imaging spectrograph coupled to a 2-D array HgCdTe detector. Individual images were acquired with a spectral range of 1000–2498 nm with 6–7 nm resolution. When stacked, the individual images created a $301 \times 220 \times 239$ hypercube. The physical image size was 60 mm × 44 mm.

2.3. Image acquisition

Twelve yellow maize kernels, four kernels randomly selected from each of the three hardness categories (hard, intermediate and soft) were positioned on a 70 mm × 70 mm silicon carbide (SiC) sandpaper as illustrated in Fig. 1. In addition 24 yellow maize kernels, eight randomly selected from the three hardness categories were also positioned on SiC sandpaper as depicted in Fig. 2. Duplicate kernels of each category were within each block within the grid. Samples were presented on the sample holders of both systems for image acquisition.

Image acquisition of reflectance standards (Labsphere, USA) necessary for image calibration and to correct for pixel-to-pixel variances due to camera inconsistencies and variation in sample illumination [32], was captured before and after image acquisition of samples. Standards of 2% (dark reference) and 75% (white reference) reflectance were used for MatrixNIR images. Internal dark and white reference standards were used for the sisuChema.

2.4. Hyperspectral image analysis (HIA)

2.4.1. Image correction

Sample images ($I_{\lambda,n}$) acquired with the Spectral Dimensions MatrixNIR camera were converted from instrument measurement values, i.e. A/D (analogue to digital) counts to absorbance values according to Eq. (1) [33]. The transformation involved correcting spectra for detector dark current (dark reference subtracted from sample spectra) and dividing by a similarly corrected

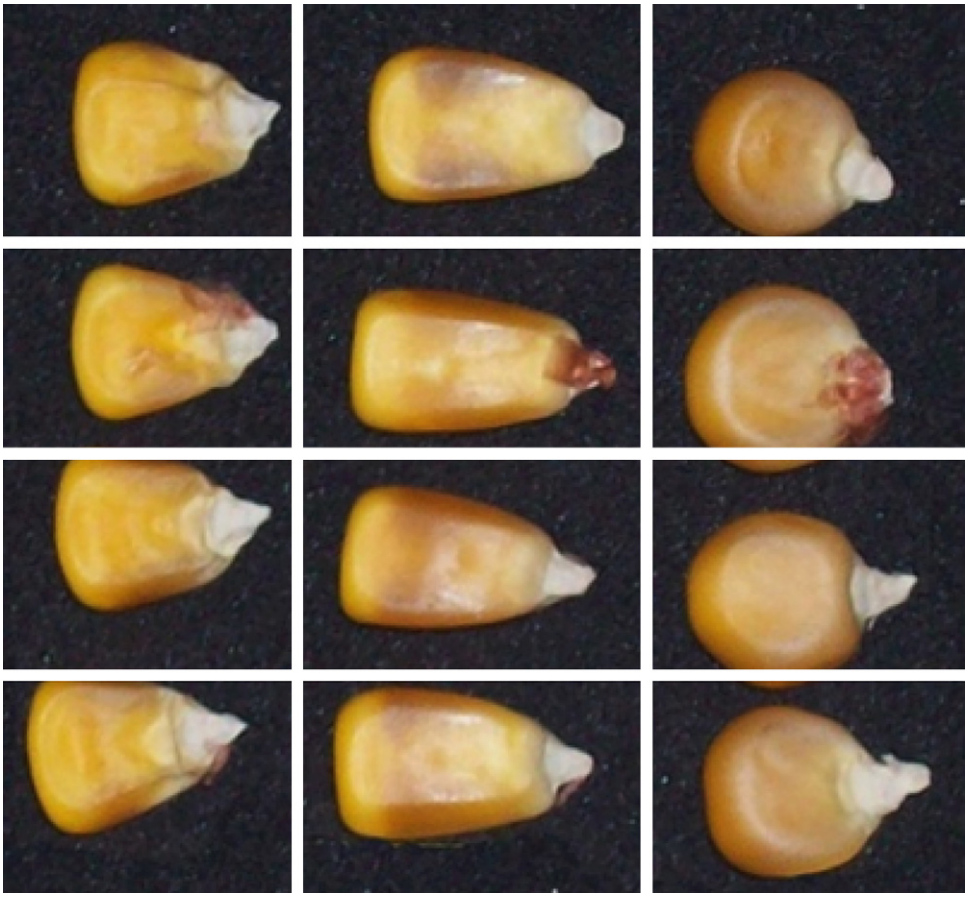

Fig. 1. Digital image illustrating sample presentation of the 12 yellow maize kernels for image acquisition with the Spectral Dimensions MatrixNIR (hard (right); intermediate (centre); soft (left)). The white lines were added for clarity and are not part of the analysed image.

total reflectance spectrum; 75% in this case (dark reference subtracted from white reference). Image acquisition was realised using MatrixAcquire software (Malvern Instruments Ltd, Malvern, Worcestershire, UK) while transformation to absorbance was performed using ISys 4.0 (Malvern Instruments Ltd, Malvern, Worcestershire, UK). In addition, absorbance images were converted to Matlab files and exported to Evince 2.020 (UmBio AB, Umeå, Sweden) for further image analysis.

$$I_{\lambda,n} = -\log\left[\left(\frac{S_{\lambda,n} - B_{\lambda,n}}{W_{\lambda,n} - B_{\lambda,n}}\right) \times 0.75\right] \quad (1)$$

WHERE n = pixel index variable ($n = 1..N$); $I_{\lambda,n}$ = standardised absorbance intensity, pixel n, at wavelength λ; $S_{\lambda,n}$ = sample image, pixel n, at wavelength λ; $B_{\lambda,n}$ = dark reference image, pixel n, at wavelength λ; $W_{\lambda,n}$ = white reference image, pixel n, at wavelength λ; 0.75 = total reflectance

Images acquired with the sisuChema using the ChemaDAQ software programme (Specim, Oulu, Finland) were automatically converted to pseudo-absorbance in the Evince 2.020 hyperspectral image analysis software package using the same principles as above.

2.4.2. Image cleaning

Absorbance images were subjected to PCA. PCA score plots as well as the PCA score images were used to identify and classify outliers, bad pixels, illumination errors and shading errors, edge effects and background. After the unwanted classes were removed,

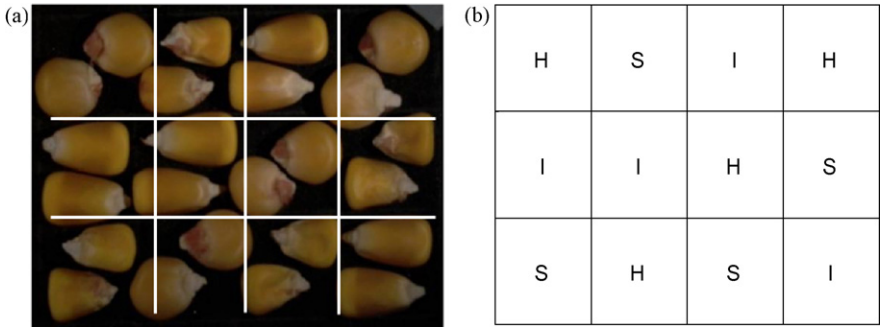

Fig. 2. (a) Digital image illustrating sample presentation of the 24 yellow maize kernels for image acquisition with the Spectral Dimensions MatrixNIR camera as well as the sisuChema SWIR imaging system and (b) grid illustrating the hardness class of each maize kernel (H = hard; I = intermediate; S = soft). The white lines were added to the digital image for clarity and are not part of the analysed image.

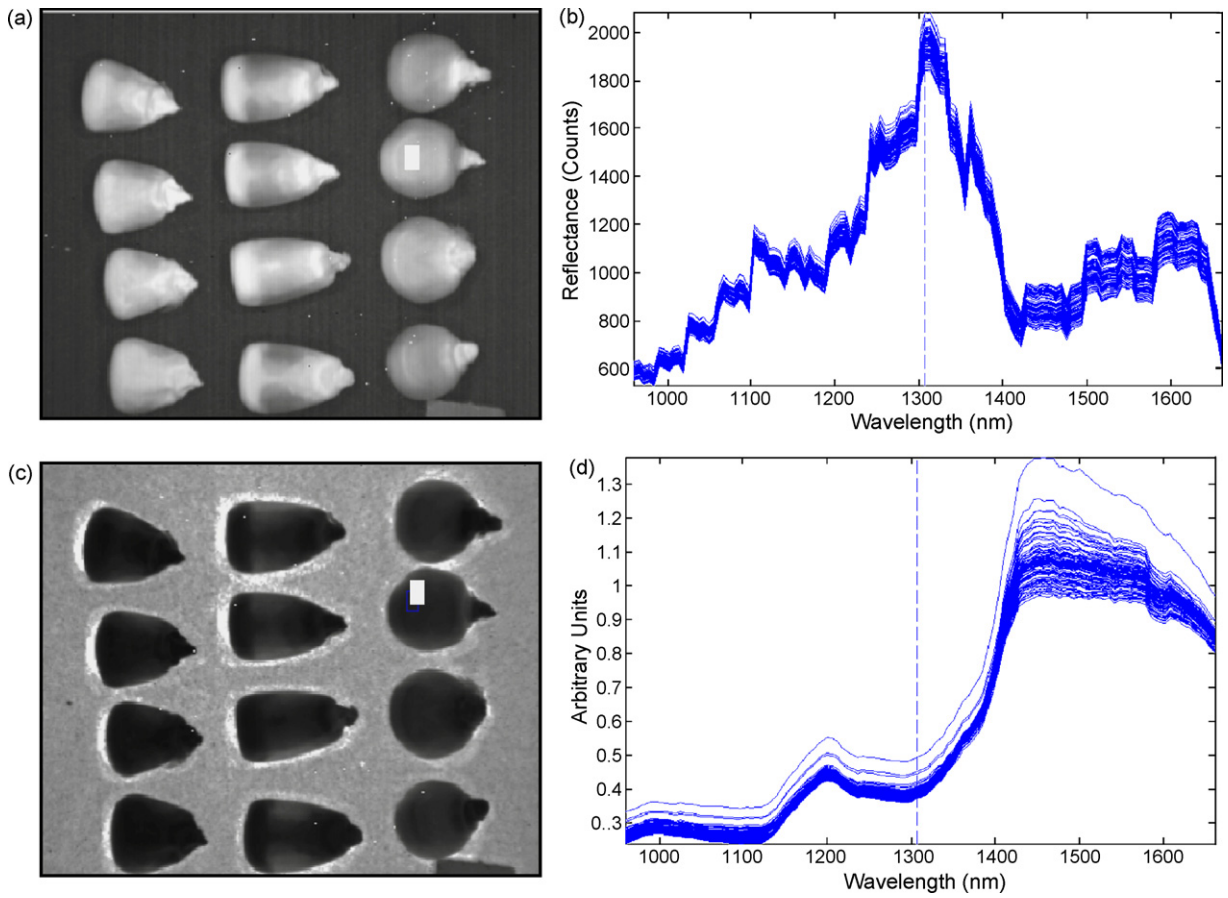

Fig. 3. (a) Raw NIR hyperspectral image at 1310 nm in detector counts and (b) spectra of the selected region; indicated by white square in (a). (c) Absorbance corrected image at 1310 nm and (d) absorbance spectra of selected region; indicated by white square in (c).

PCA was recalculated with additional components obtaining a cleaned image. The cleaned image was used in subsequent analysis.

2.4.3. Image analysis of cleaned image

PCA score plots and the accompanied score images were used to identify and classify clusters to the various hardness classes (glassy and floury). These were projected onto PCA score images. Various data pre-treatments such as multiplicative scatter correction (MSC), standard normal variate (SNV) and derivatives were applied to the mean-centered data and evaluated.

2.4.4. Partial least squares discriminant analysis

PLS-DA is a technique for dimension reduction and discrimination [34]. It facilitates this based on the notion that group membership is to be predicted. A dummy matrix (Y–endosperm type in this case) that registers the membership is paired with the training set (X–spectral data). PLS is then calculated in the usual way and the model tested on the test set.

PLS-DA models were calculated to distinguish between the two types of endosperm. Each object in the two classes (glassy and floury endosperm) was assigned to a dummy variable, zero or one, *i.e.* floury endosperm or glassy endosperm with a threshold of 0.5. The images were divided into training and test sets by assigning approximately two thirds of the image to the training set and the remainder to the test set. The pedicle of the maize kernels was excluded before performing PLS-DA.

3. Results and discussion

3.1. Hyperspectral image analysis–MatrixNIR (12 kernels)

3.1.1. Image correction

An example of the difference between raw and corrected images at 1310 nm is shown in Fig. 3. Similar arguments are valid for other wavelength bands, but not shown here. Fig. 3a shows an image at 1310 nm in A/D converter counts. A representative collection of spectra from this image is shown in Fig. 3b. The shape of the spectra shows only the combined effect of the sensitivity of the InGaAs diode array and the distribution of the incoming radiation. Therefore the use of background (dark) and bright images is needed for correction. Fig. 3c and d show the corrected image and spectra, now in absorbance. These spectra look like normal NIR absorbance spectra.

3.1.2. Image cleaning

PCA with mean-centering was carried out and three components accounted for 91.74% of the explained variance (% sum of squares (SS)). The PCA score plot (PC1 versus PC2) (Fig. 4a) of the MatrixNIR acquired image (12 maize kernels) comprised of 81 920 pixels. This plot was used for exploration and identification of any special features (clusters or groupings) within the data set. It is known that if mean-centered data are used, PC1 typically contains little chemical information and usually includes variation related to physical differences useful for cleaning an image [35]. Used in conjunction with the score image (a technique called brushing) (Fig. 4b), it was possible to identify various clusters within the score

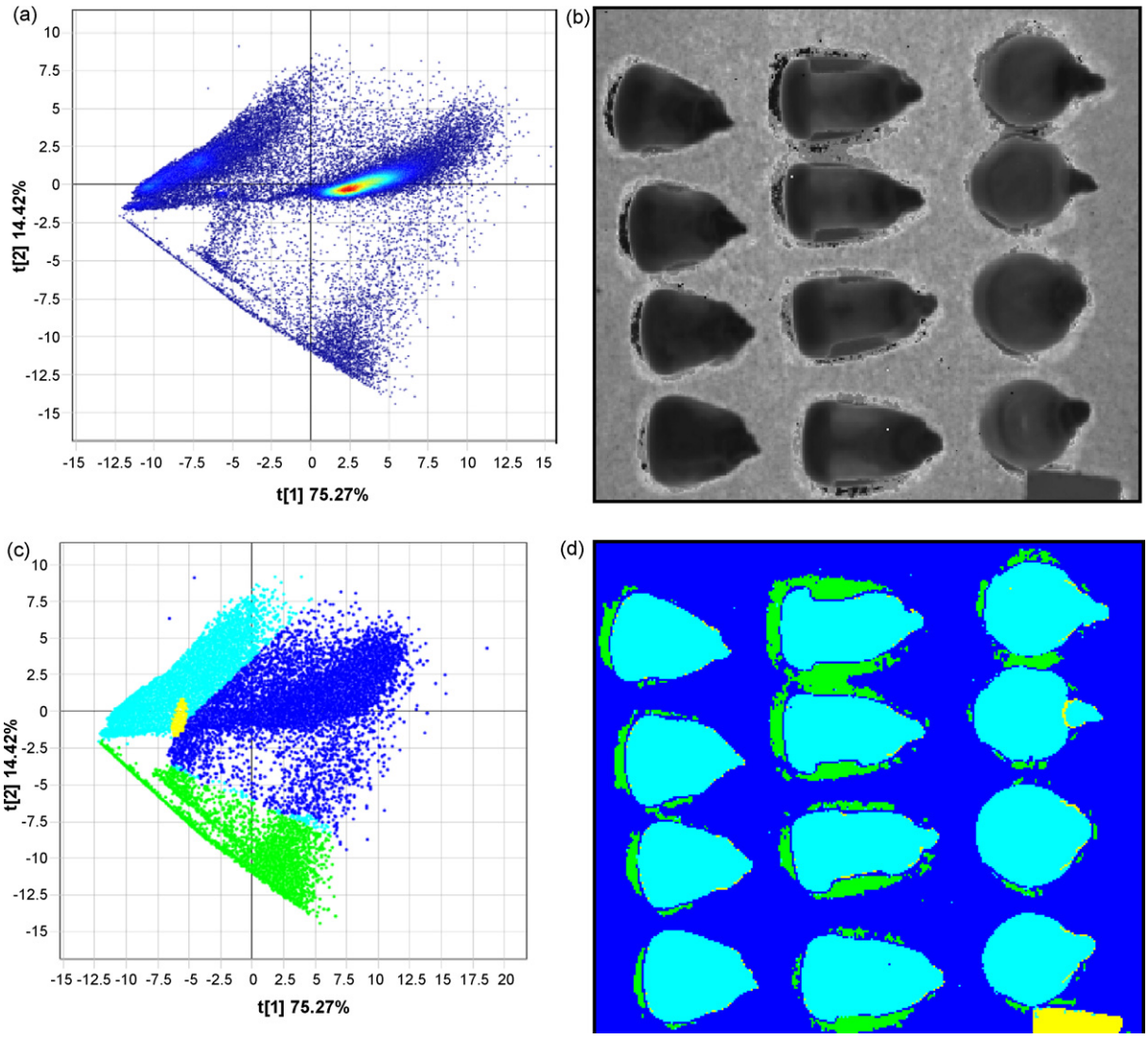

Fig. 4. (a) PCA score plot of absorbance image (PC1 versus PC2) obtained with the MatrixNIR (81 920 pixels) and (b) score image of PC1 used together by brushing to identify and locate unwanted pixels; (c) PCA score plot with selected clusters projected onto (b) to obtain (d) classification image (blue = sandpaper, bad pixels and edge effects; cyan = maize kernels; green = shading; yellow = plastic marker) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

plot that were associated with different components in the image. The score plot with selected clusters (Fig. 4c) was subsequently projected on the score image (Fig. 4b) which resulted in a classification image (Fig. 4d). This allowed for the identification of the respective clusters such as SiC sandpaper, bad pixels and edge effects, shading, maize kernels and plastic marker [36]. The removal of the unwanted clusters, as indicated in the score plot (Fig. 4c) and classification image (Fig. 4d) resulted in a cleaned score image (Fig. 5a) that was used in subsequent analyses.

3.1.3. Image analysis of cleaned images

The score plot of PC2 versus PC3 (Fig. 5a) showed three clusters. The respective clusters in the score plot could be associated with specific parts of the maize kernels in the score image interactively by brushing. Employing this information in addition to histological knowledge of a maize kernel [2], the clusters could be assigned to classes with the resulting classification plot (Fig. 5b). The latter was projected onto the score image (Fig. 5c) resulting in a classification image (Fig. 5d) comprising three classes, i.e. glassy endosperm (blue), floury endosperm (green) and the pedicle and hull (yellow). This classification was possible as after removal of the irrelevant clusters the cleaned score image (Fig. 5c) was more interpretable. Other combinations of PCs could not improve on the classification obtained from the score plot of PC2 versus PC3.

From these results the MatrixNIR with wavelength range 960–1660 nm showed it was possible to distinguish between the different endosperm types, i.e. glassy and floury. Each kernel, albeit hard, intermediate or soft, contained both glassy and floury endosperm but in different ratios. This means the clustering is based on internal composition of the kernels, i.e. endosperm texture and not on whole kernels. These inbred lines thus comprised only two types of endosperm; hard maize kernels constituting mainly of glassy endosperm, soft kernels mainly of floury endosperm and intermediate kernels of different ratios of glassy and floury endosperm. These observations were in contrast to a recent study where it was shown that commercial maize hybrids of intermediate hardness also included a third type of endosperm [31]. It was suggested that this third class of endosperm consists of a mixture or layers of floury and vitreous endosperm.

For the spectral interpretation of the difference between glassy and floury endosperm, spectra from both classes were extracted

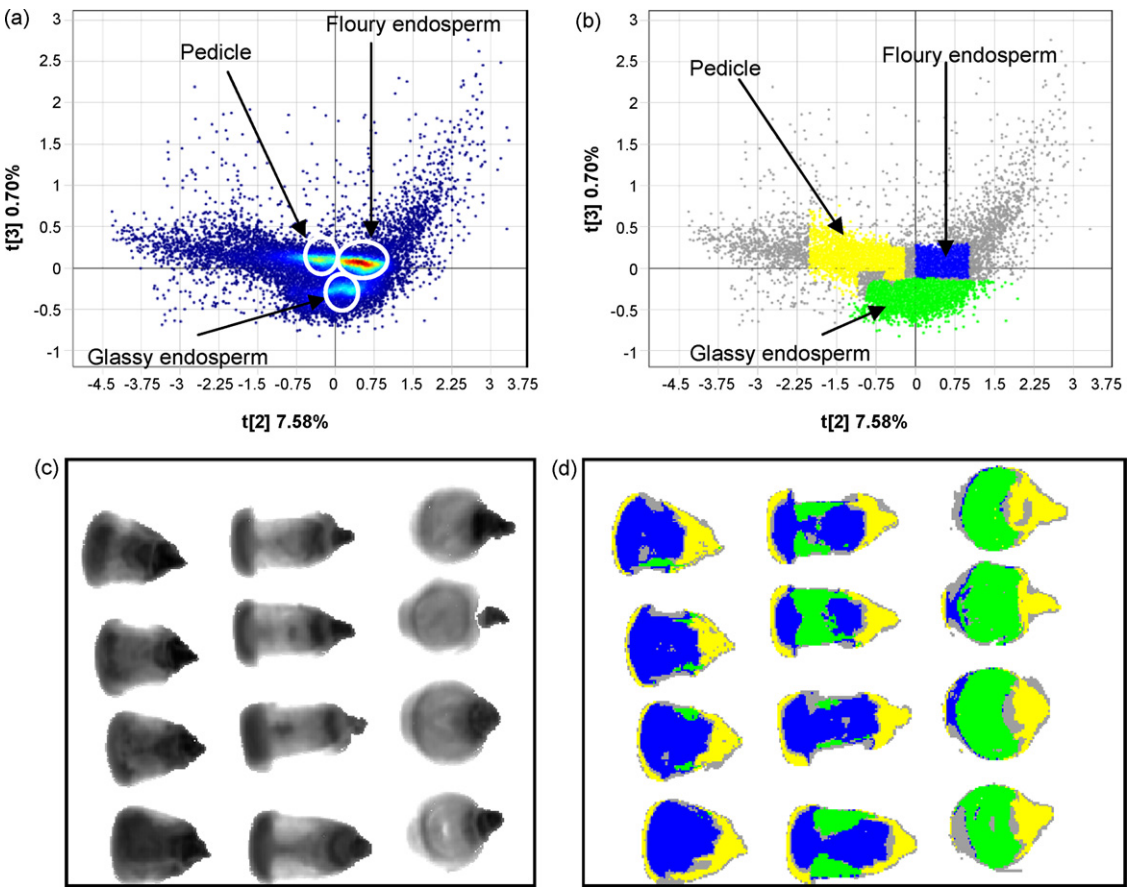

Fig. 5. PCA score plots of PC2 (7.58%) versus PC3 (0.70%) depicting (a) three distinct density clusters and (b) with the three clusters assigned to classes which was then projected onto (c) the score image (PC1) resulting in (d) the classification image (green = glassy endosperm; blue = floury endosperm; yellow = pedicle) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

and analysed together. The matrices were 5699×118 for glassy endosperm and 8993×118 for soft endosperm making a combined data matrix of size $14\,692 \times 118$. It was found that there was a difference in scattering properties. From the different pre-treatments evaluated, standard normal variate (SNV) correction was the most efficient pre-treatment. The score plots obtained showed more distinct clusters. After a PCA analysis on the SNV corrected and mean-centered data, the first PC (83% SS) showed the difference between glassy and floury endosperm very clearly in the score values (Fig. 6a). In Fig. 6a positive values are related to glassy endosperm and negative values to floury endosperm. This meant that the loading for the first PC (Fig. 6b) would allow a chemical interpretation of this difference. The most prominent feature in the PC loading line plot is the starch-moisture contrast with positive values below 1400 nm for starch and negative values above 1400 nm for moisture.

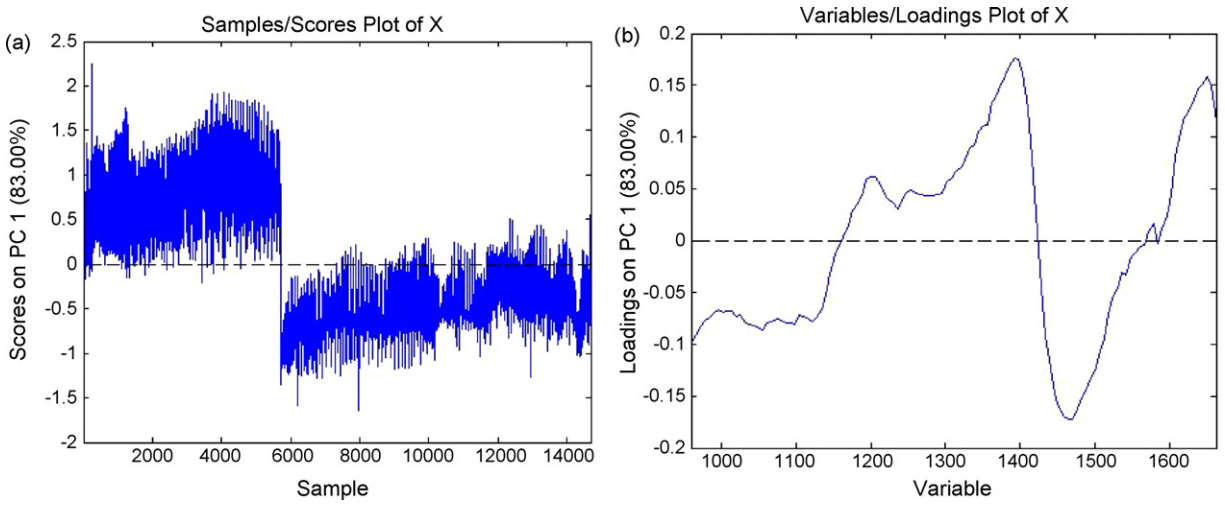

Fig. 6. (a) Score plot and (b) loading line plot of PC1 for MatrixNIR (12 kernels).

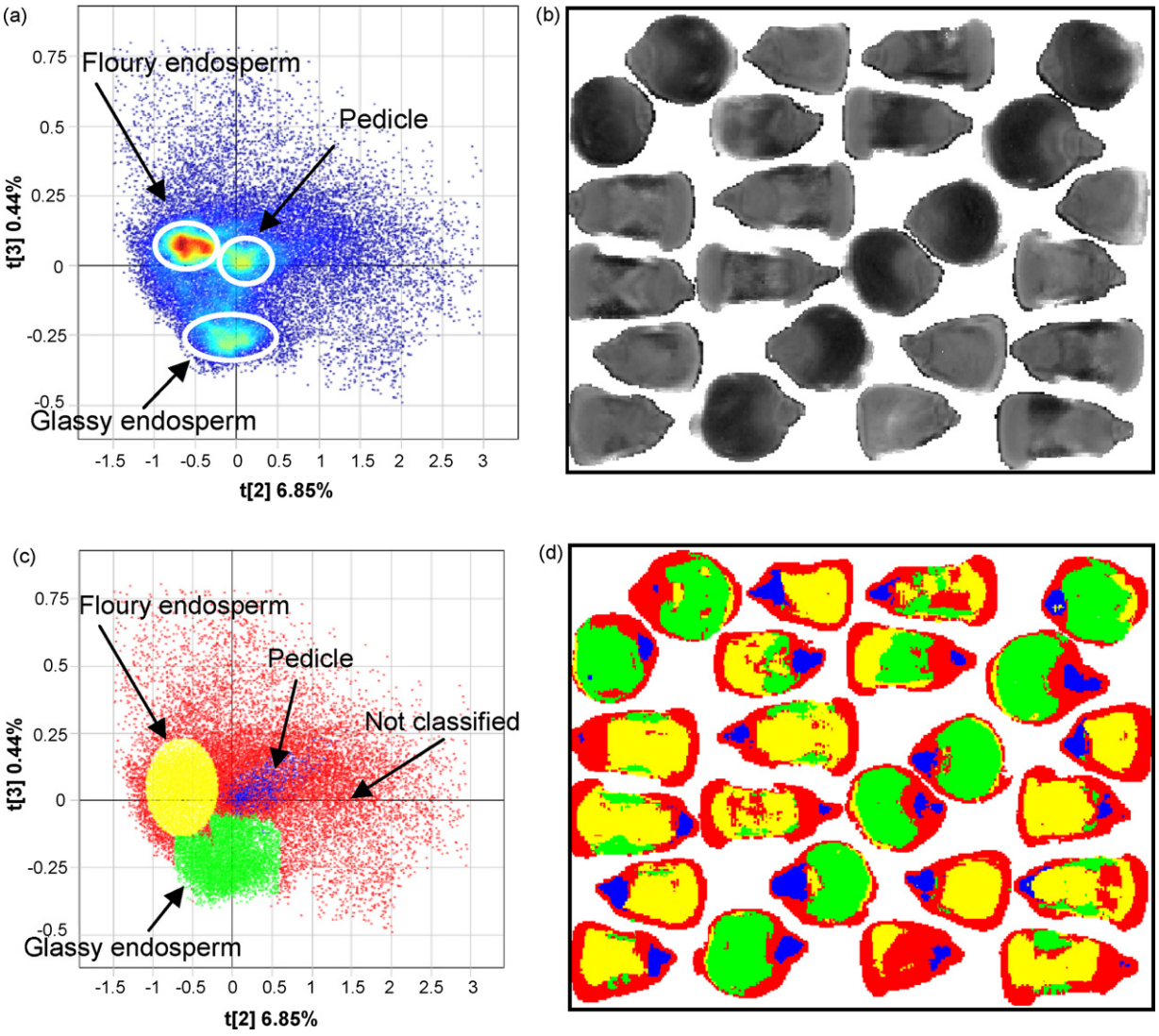

Fig. 7. PCA score plots of PC2 (6.85%) versus PC3 (0.44%) for MatrixNIR 24-kernel image depicting (a) three distinct density clusters and (b) the score image (PC2) onto which (c) the three clusters assigned to classes was then projected resulting in (d) the classification image (green = glassy endosperm; blue = pedicle; yellow = floury endosperm; red = not classified) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

3.2. Hyperspectral image analysis–MatrixNIR and sisuChema (24 kernels)

3.2.1. Image correction and image cleaning

The raw images of the 24 maize kernels acquired with the MatrixNIR and sisuChema cameras were corrected and cleaned following the same approach as described earlier.

3.2.2. Image analysis of cleaned images–MatrixNIR

The first image collected on the MatrixNIR included 12 kernels positioned according to hardness category in the field of view. To remove possible lighting and edge effects a second image was acquired with 24 kernels randomly positioned. After removal of the irrelevant clusters, the PC score plot of PC2 (6.85%) versus PC3 (0.44%) showed similar clustering of pixels (Fig. 7a) as the previous MatrixNIR score plot (Fig. 5a). Already in the score image of PC2 (Fig. 7b) differences between glassy and floury endosperm regions were evident and with the knowledge gained from the previous analyses, it was clear which kernels were predominantly hard, intermediate and soft. As previously discussed, the score plot (Fig. 7a) was used to identify clusters and generated a classified score plot (Fig. 7c) which was further used together with the score image (Fig. 7b) to create a classification image (Fig. 7d). The matrices were 11 649 × 118 for glassy endosperm and 15 079 × 118 for soft endosperm making a combined data matrix of size 26 728 × 118. The classification results were similar to those in Fig. 5d since the chemical constituents remained the same. This confirms the possibility of distinguishing between different endosperm types using NIR hyperspectral imaging. The average difference spectra for the glassy and floury endosperm classes show the difference in starch content between the two classes (Fig. 8).

3.2.3. Image analysis of cleaned images–sisuChema

The potential of using NIR hyperspectral imaging to distinguish between maize kernels of different hardness categories was also confirmed on a line-scan system operating at a longer wavelength range (1000–2498 nm) compared to the MatrixNIR. A score plot of PC2 (4.61%) versus PC3 (0.60%), was obtained subsequent to the removal of the irrelevant pixels which included the pedicle in this case (Fig. 9a). Even though the clusters were not apparent, two clusters representing glassy and floury could be identified which were subsequently assigned to classes in the score plot (Fig. 9b) and projected onto the score image (Fig. 9c) to generate the classification image (Fig. 9d). The matrices were 11 308 × 239 for glassy

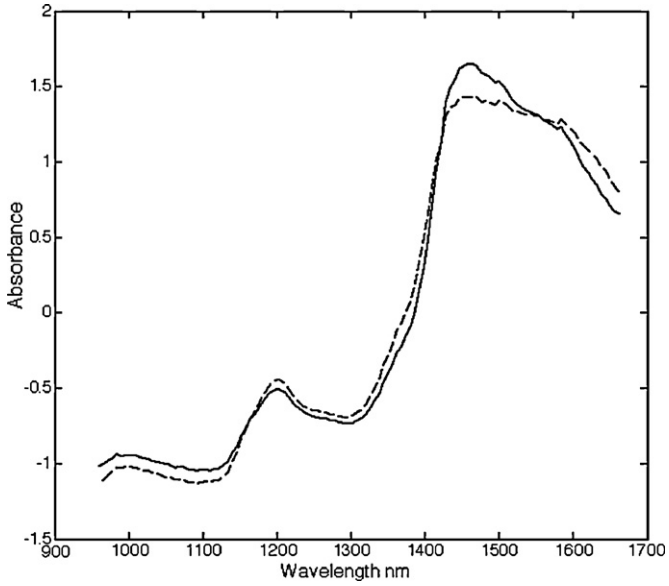

Fig. 8. Standard normal variate (SNV) corrected average spectra of glassy and floury endosperm classes with the main difference at 1450 nm indicating difference in starch for the MatrixNIR 24 kernel image.

endosperm and 14 915 × 239 for soft endosperm making a combined data matrix of size 26 223 × 118.

3.3. Partial least squares discriminant analysis–MatrixNIR (12 kernels)

The PLS-DA model, to distinguish between glassy and floury endosperm, had a coefficient of determination of 88% for 10 PLS components, but already with 3 to 6 components 84–86% was explained (floury endosperm = 2364 pixels and glassy endosperm = 1215 pixels). PLS-DA was used to (1) determine whether it was possible to discriminate between the classes and (2) to build a model that could be used to predict future images. More than 99% correct prediction was obtained. Simply making a regression model can, however, lead to overfitting and a better test of prediction ability had to be carried out. This was done by making separate training and test sets from the image. The large number of available pixels made this acceptable. The training set was made up of three rows and the test set of one row. The regression model was tested in all possible combinations of training and test sets and similar results obtained each time. The results shown was when the training set was made up of the upper three rows (see Fig. 1) and the test set of the bottom row. A four PLS component model had a coefficient of determination of 85% and a root mean square error of prediction (RMSEP) of 0.18. The calibration and prediction images (Fig. 10) illustrate the localisation of the modelled and predicted classes, i.e. glassy and floury endosperm. It has to be noted that the

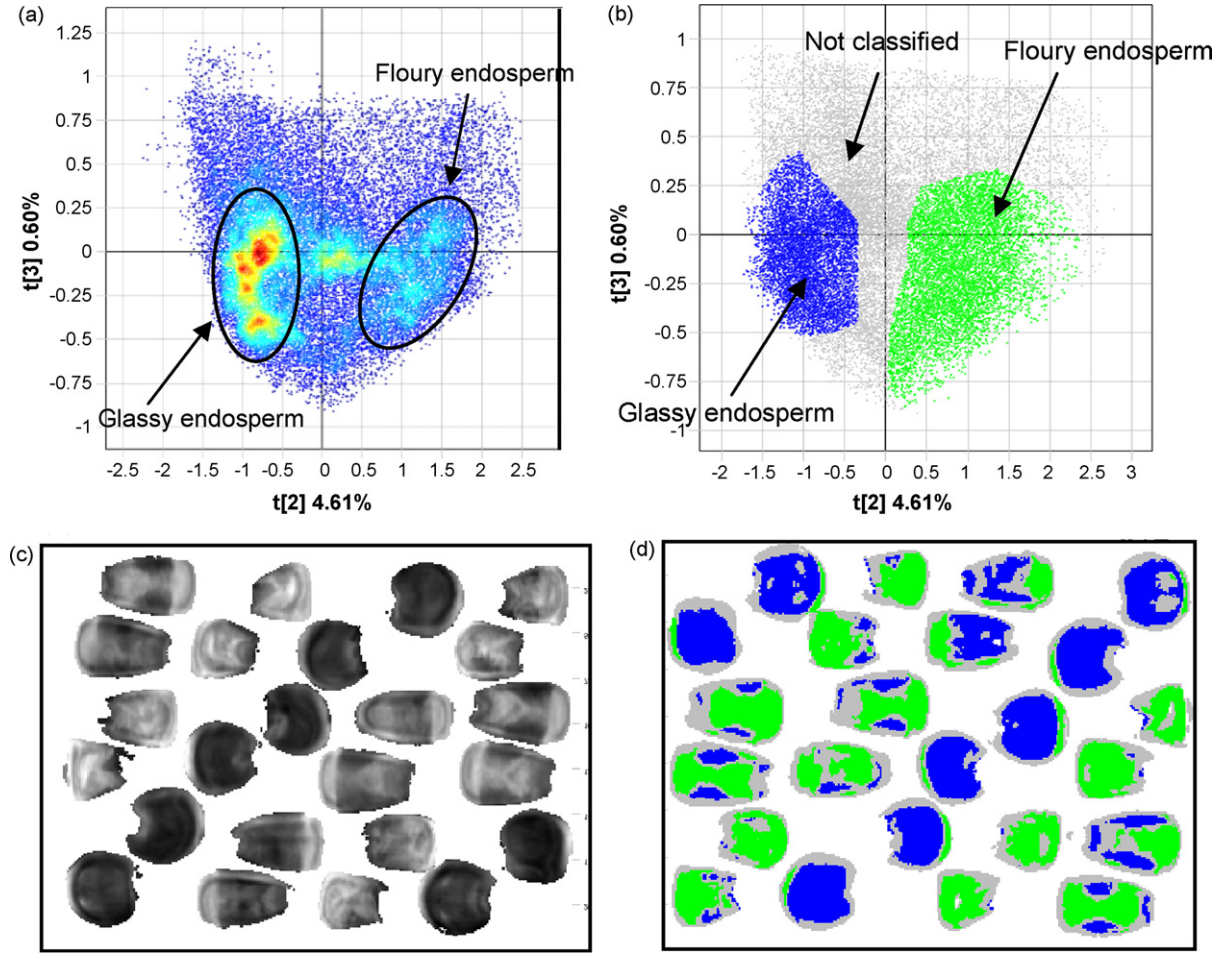

Fig. 9. PCA score plots of PC2 (4.61%) versus PC3 (0.60%) for sisuChema depicting (a) two density clusters and (b) the two clusters assigned to classes which was then projected onto (c) the score image (PC2) resulting in (d) the classification image (blue = glassy endosperm; green = floury endosperm; grey = not classified) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

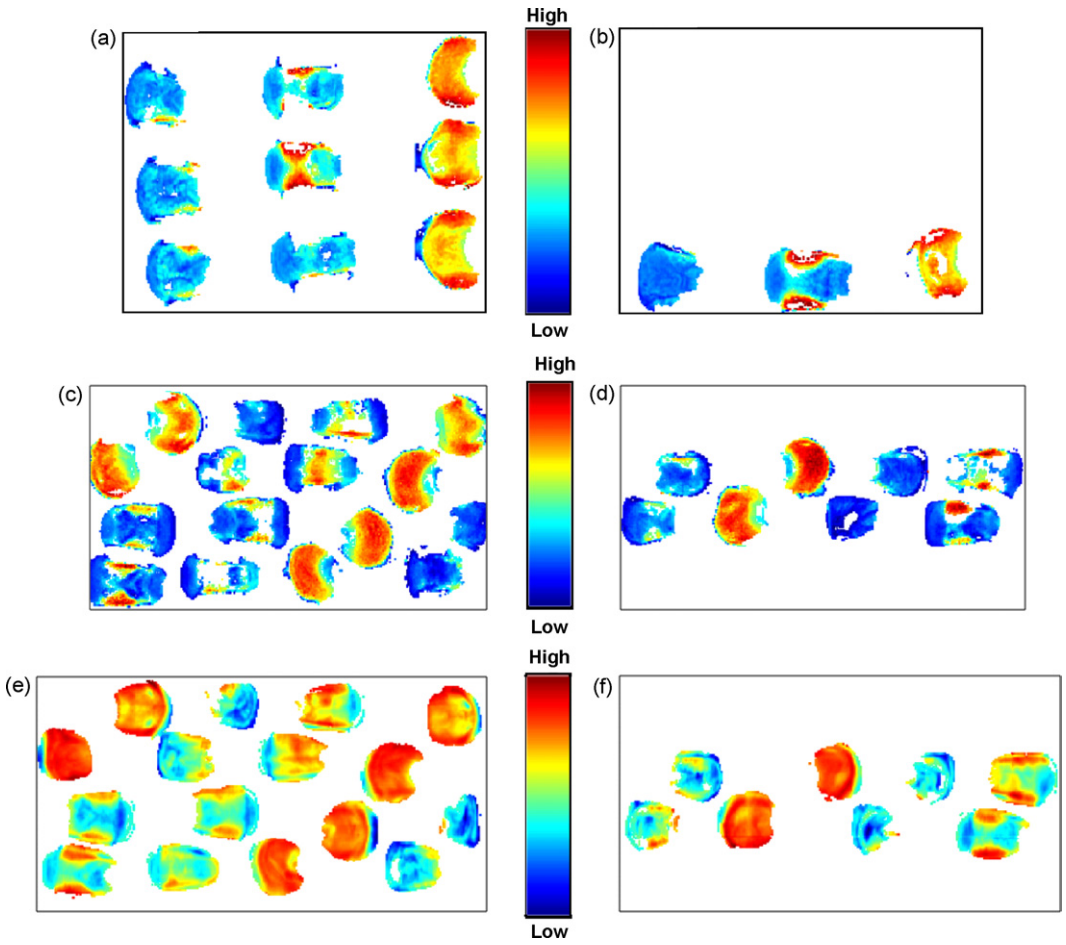

Fig. 10. (a) Calibration and (b) prediction images of the MatrixNIR (12 kernels) after four PLS components; (c) calibration and (d) prediction images of the MatrixNIR (24 kernels) after three PLS factors and (e) calibration and (f) prediction images of the sisuChema (24 kernels) after three PLS components for glassy endosperm (red = glassiness high, blue = glassiness low, i.e. flouriness high) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

scale in the legend (Fig. 10) is continuous; unlike that of the classification images (Figs. 5, 7 and 9). From Fig. 10 it is clear the soft maize kernels (left column in Fig. 10a) were high in floury endosperm and the hard maize kernels (right column in Fig. 10a) were high in glassy endosperm. The remaining intermediate kernels (middle column in Fig. 10a) had a mixture of glassy and floury endosperm. Fig. 10b shows the same levels of high and low glassiness occurring in the prediction image.

3.4. *Partial least squares discriminant analysis–MatrixNIR (24 kernels)*

After three PLS factors the coefficient of determination, to distinguish between glassy and floury endosperm, was 84.9%. Correct classification rates of more than 98% were obtained (RMSEP = 0.181). The calibration and prediction images are shown in Fig. 10c and d. The top four rows (see Fig. 2) were selected for the training set and the bottom two rows for the test set. Again different combinations of training and test sets were evaluated; only one is shown. The calibration image (Fig. 10c) illustrated the ability of the model to discriminate between glassy and floury endosperm. Fig. 10d gives an indication of the model's ability to predict glassiness and flouriness of the endosperm. From the ratio of glassy and floury endosperm present in each kernel its hardness could be inferred. The conclusions of the previous section (MatrixNIR 12 kernels) still hold and are thus corroborated.

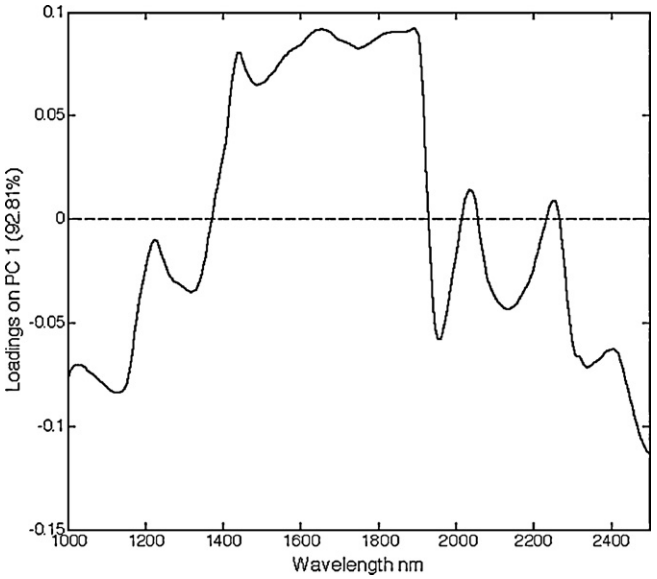

Fig. 11. First PC loading of the SNV corrected data from the sisuChema. The corresponding score plot (not shown) showed a marked difference between glassy (high score) and floury (low score) endosperm.

3.5. Partial least squares discriminant analysis–sisuChema (24 kernels)

After three PLS components the coefficient of determination, to distinguish between glassy and floury endosperm, was 76.3% and further components gave no substantial improvement of the model. Similar to the other two images, where different combinations of training and test sets were evaluated, only the results of the subdivision as in Fig. 10c and d are shown. The calibration and prediction images shown in Fig. 10e and f confirms the ability of the model to predict glassiness and flouriness correctly (RMSEP = 0.294).

The first PC loading line plot on the SNV corrected and mean-centered data are shown in Fig. 11. Again the first component showed the difference between glassy and floury endosperm very clearly in the score values (not shown) with positive values related to glassy endosperm and negative values to floury endosperm. This meant the loading for the first PC (Fig. 6b) would allow a chemical interpretation of this difference with starch and moisture being prominent.

4. Conclusion

With both data sets (12 and 24 kernels) and both instruments a number of hyperspectral images were made. It is very important to remove all parasitic signals from any hyperspectral image. These signals comprise background, bad pixels, edge effects and specular reflection. In all cleaned images analysed, it was possible to detect glassy and floury endosperm inside the kernels, using exploratory PCA analysis with only two to three principal components. It was found that no kernels consisted of only one endosperm type. Because of this it is not possible to make models based solely on whole kernels. The hardness of the kernels could, however, be deduced from the ratio of the glassy and floury endosperm present in the maize kernels. The PCA classifications observed, corresponded well with histological knowledge of maize kernels and indicated starch-moisture contrast as the main feature in the PC loading line plot. With the observed classes, PLS-DA models could be made with reasonably high coefficients of determination. The number of test set kernels was not very high, but the results were reproducible between data sets and between instruments, indicating that the method proposed in this paper has a real potential for future classification uses. The benefit of NIR hyperspectral imaging compared to existing hardness testing methods is speed of analysis and measurement of single kernels non-destructively. In addition the two types of endosperm could be distinguished using hyperspectral imaging, not possible using existing methods.

Acknowledgements

Julian White (Specim, Spectral Imaging Ltd, Oulu, Finland) for use of the sisuChema instrument, Oskar Jonsson (Umbio AB, Umeå, Sweden) for use of Evince software, Pioneer Seed (Delmas, South Africa) for supplying and labelling samples, 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 NRF, South Africa for funding (FA2006032900007).

References

[1] S.A. Watson, in: S.A. Watson, P.E. Ramstad (Eds.), Corn: Chemistry and Technology, American Association of Cereal Chemists, Inc, St Paul, Minnesota, USA, 1987, p. 125.
[2] R.C. Hoseney, Principles of Cereal Science and Technology, American Association of Cereal Chemists, Inc, St. Paul, Minnesota, 1994.
[3] S.A. Watson, in: S.A. Watson, P.E. Ramstad (Eds.), Corn: Chemistry and Technology, American Association of Cereal Chemists, Inc, St. Paul, Minnesota, USA, 1987, p. 53.
[4] K.M. Lee, S.R. Bean, S. Alavi, T.J. Herrman, R.D. Waniska, J. Agric. Food Chem. 54 (2006) 4260.
[5] Y. Pomeranz, C.R. Martin, D.D. Traylor, F.S. Lai, Cereal Chem. 61 (1984) 147.
[6] K.-M. Lee, T.J. Herrman, J. Lingenfelser, D.S. Jackson, J. Cereal Sci. 41 (2005) 85.
[7] G. Eyherabide, J. Robutti, F. Borras, Cereal Chem. 73 (1996) 775.
[8] R.L. Wehling, D.S. Jackson, B.R. Hamaker, Cereal Chem. 73 (1996) 543.
[9] H.D. Almeida-Dominguez, E.L. Suhendro, L.W. Rooney, J. Cereal Sci. 25 (1997) 93.
[10] Y. Pomeranz, Z. Czuchajowska, C. Martin, F. Lai, Cereal Chem. 62 (1985) 108.
[11] G.P. Fox, M. Manley, J. Agric. Food Chem. 57 (2009) 5647.
[12] R.L. Wehling, D.S. Jackson, D.G. Hooper, A.R. Ghaedian, Cereal Chem. 70 (1993) 720.
[13] J. Burger, P. Geladi, Analyst 131 (2006) 1152.
[14] A. Peirs, N. Scheerlinck, B.M. Nicolai, J. De Baerdemaeker, Proceedings of the International Conference Postharvest Unlimited, 2003, p. 315.
[15] J.G. Tallada, M. Nagata, T. Kobayashi, Environ. Control Biol. 44 (2006) 245.
[16] P. Geladi, J. Burger, T. Lestander, Chemom. Intell. Lab. Syst. 72 (2004) 209.
[17] P. Tatzer, M. Wolf, T. Panner, RTI 11 (2005) 99.
[18] R.P. Cogdill, C.R. Hurburgh, G.R. Rippke, Trans. ASAE 47 (2004) 311.
[19] N. Gorretta, J.M. Roger, M. Aubert, V. Bellon-Maurel, F. Campan, P. Roumet, J. Near Infrared Spectrosc. 14 (2006) 231.
[20] C.B. Singh, D.S. Jayas, J. Paliwal, N.D.G. White, Trans. ASABE 50 (2007) 2171.
[21] H. Zhang, J. Paliwal, D.S. Jayas, N.D.G. White, Trans. ASABE 50 (2007) 1779.
[22] H. Koç, V.W. Smail, D.L. Wetzel, J. Cereal Sci. 48 (2008) 394.
[23] V.W. Smail, A.K. Fritz, D.L. Wetzel, Vib. Spectrosc. 42 (2006) 215.
[24] B.A. Weinstock, J. Janni, L. Hagen, S. Wright, Appl. Spectrosc. 60 (2006) 9.
[25] M. Berman, P.M. Connor, L.B. Whitbourn, D.A. Coward, B.G. Osborne, M.D. Southan, J. Near Infrared Spectrosc. 15 (2007) 351.
[26] P. Geladi, H.F. Grahn, Multivariate Image Analysis, John Wiley & Sons Ltd, Chichester, West Sussex, 1996.
[27] P. Geladi, H.F. Grahn, J. Burger, in: H.F. Grahn, P. Geladi (Eds.), Techniques and Applications of Hyperspectral Image Analysis, John Wiley & Sons Ltd, Chichester, West Sussex, 2007, p. p1.
[28] S. Baronti, A. Casini, F. Lotti, S. Porcinai, Chemom. Intell. Lab. Syst. 39 (1997) 103.
[29] A.A. Gowen, C.P. O'Donnell, M. Taghizadeh, P.J. Cullen, J.M. Frias, G. Downey, J. Chemom. 22 (2008) 259.
[30] S. Chevallier, D. Bertrand, A. Kohler, P. Courcoux, J. Chemom. 20 (2006) 221.
[31] M. Manley, P. Williams, D. Nilsson, P. Geladi, J. Agric. Food Chem. (2009) doi:10.1021/jf9018323.
[32] J. Burger, P. Geladi, J. Chemom. 19 (2005) 355.
[33] N. Gat, Imaging spectroscopy using tunable filters: a review, in: H.H. Szu, M. Vetterli, W.J. Campbell, J.R. Buss (Eds.), Wavelet Applications VII, SPIE, Orlando, FL, USA, 2000, p. 50.
[34] M. Barker, W. Rayens, J. Chemom. 17 (2003) 166.
[35] K.H. Esbensen, T.T. Lied, in: H.F. Grahn, P. Geladi (Eds.), Techniques and Applications of Hyperspectral Image Analysis, John Wiley & Sons, Ltd, Chichester, West Sussex, 2007, p. 17.
[36] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, Trends Food Sci. Technol. 18 (2007) 590.