PRELIMINARY ANALYSIS OF RGB IMAGES FOR THE IDENTIFICATION OF DEFECTIVE MAIZE KERNELS

Giorgia Orlandi, Rosalba Calvini, Giorgia Foca, Alessandro Ulrici

Department of Life Sciences and Interdepartmental Research Centre BIOGEST-SITEIA, University of Modena and Reggio Emilia, Padiglione Besta, Via Amendola 2, 42122 Reggio Emilia, Italy, giorgia.orlandi@unimore.it, rosalba.calvini@unimore.it, giorgia.foca@unimore.it, alessandro.ulrici@unimore.it

Abstract – In order to investigate the effectiveness of multivariate image analysis for the evaluation of maize defects, RGB images of maize samples containing different percentages of defective kernels were acquired and then converted into *colourgrams*, i.e., signals codifying colour-related features. Multivariate analysis of the colourgrams matrix showed a distribution of the acquired samples according to the amount of defective kernels.

Keywords: maize kernels; defects identification; RGB image; multivariate analysis.

1. INTRODUCTION

The quality of maize is strictly connected with the presence of stained, dark or rotten kernels, which are always a sign of deterioration and of high possibility of contamination by fungi belonging to the genera *Aspergillus*, *Fusarium* and *Penicillium* [1]. These filamentous fungi produce mycotoxins as secondary metabolites, which represent a major risk for human health.

The most common analyses to ensure food safety and to estimate the concentration of mycotoxins in maize are based on immunoassay and chromatographic methods. Although these techniques allow gaining high performance, an interesting issue is the possibility to develop faster, cheaper and non-destructive methods in order to speed up the transfer phases of the crop to the warehouse and, at the same time, to further increase the quality level of the final product. In this context, multivariate image analysis could represent an opportunity for the identification of maize kernels defects.

In the present work, RGB images of maize samples containing different percentages of defective maize kernels were acquired using both a flatbed scanner and a digital camera, in a manner to compare different acquisition modalities. Then, each image was converted into the corresponding signal, named colourgram, which objectively codifies the colour properties of the RGB image [2].

In order to test the applicability of the method and to delineate the correct steps for image analysis, the colourgrams matrix was firstly evaluated by means of Principal Component Analysis (PCA), then calibration models were developed using Partial Least Squares (PLS).

2. MATERIALS AND METHODS

2.1. Maize samples and images acquisition

In this study, two different types of maize kernels were considered: dry and wet maize. For both types, expert assessors manually separated the maize kernels into defective and non-defective kernels, distinguishing the stained, dark or rotten kernels from those with uniform yellow pericarp.

The RGB images were acquired using: i) a Canon 9000F MarkII scanner with a resolution of 300 DPI; ii) a Panasonic DMC-225 digital camera mounted on a box with white inner surface. The lighting system for camera acquisition consisted of a white LED strip assembled on a metallic support and directed upwards; in this manner the sample was only illuminated by diffused light, to avoid the presence of undesired shadows or reflection effects. All the images were saved in JPEG format with a spatial resolution of 4000x3000 pixels.

The image acquisition scene consisted of a white conveyor belt as background, where eight standard colour references were included to correct possible variations of the lighting conditions.

For image acquisition, 13 different mixtures were prepared containing the following percentages by weight (w/w) of defective maize kernels: 0%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 100%. Each mixture consisted of a total amount of maize kernels equal to 150 g.

For dry maize two replicate samples of each mixture were prepared, while for wet maize only one replicate of each mixture was considered. Therefore, 39 samples (13 mixtures of dry maize × 2 replicate + 13 mixtures of wet maize) were obtained. A detailed diagram of the considered samples is reported in Fig. 1. Then, two images of each sample were acquired, shuffling the maize kernels between the two acquisitions. Furthermore, the same procedure was repeated in a second measurement session in a different day to check the day-to-day variability. Therefore, 156 images (39 samples × 2 repeated acquisitions × 2 measurement sessions) were acquired for each acquisition method (i.e., scanner and camera).

2.2. Images correction and conversion into colourgrams

For both the acquisition methods, the key steps of image analysis are summarized in Fig. 2.

First of all, the acquired images were corrected by means of the standard colour references contained in the image scene, in order to minimize the differences among images due to variations of the acquisition system or of the light source.

The algorithm for image correction is based on the following procedure: an image is chosen as reference (master image), then each other image is corrected by subtracting from its RGB values the difference between the mean RGB values of the colour reference of the master image and the corresponding values of the image to be corrected.

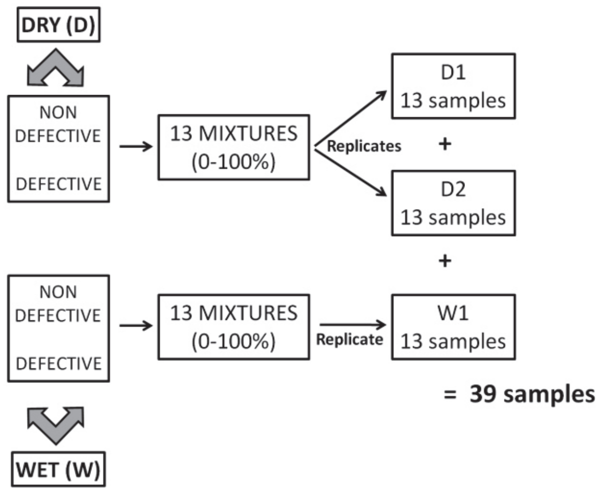

Fig. 1. Preparation of the samples used for image acquisition.

After the correction phase, the images of the samples were converted into colourgrams. In order to verify that image correction does not remove useful information, the colourgrams were also calculated considering the colour references cropped from each image.

In particular, the colourgrams approach is based on the idea to calculate the frequency distribution curves of a series of colour-related parameters and to merge them in sequence to give a 4900 points-long one-dimensional signal. The colourgram can be therefore considered as a "fingerprint" of the colour content of an RGB image. The matrix of

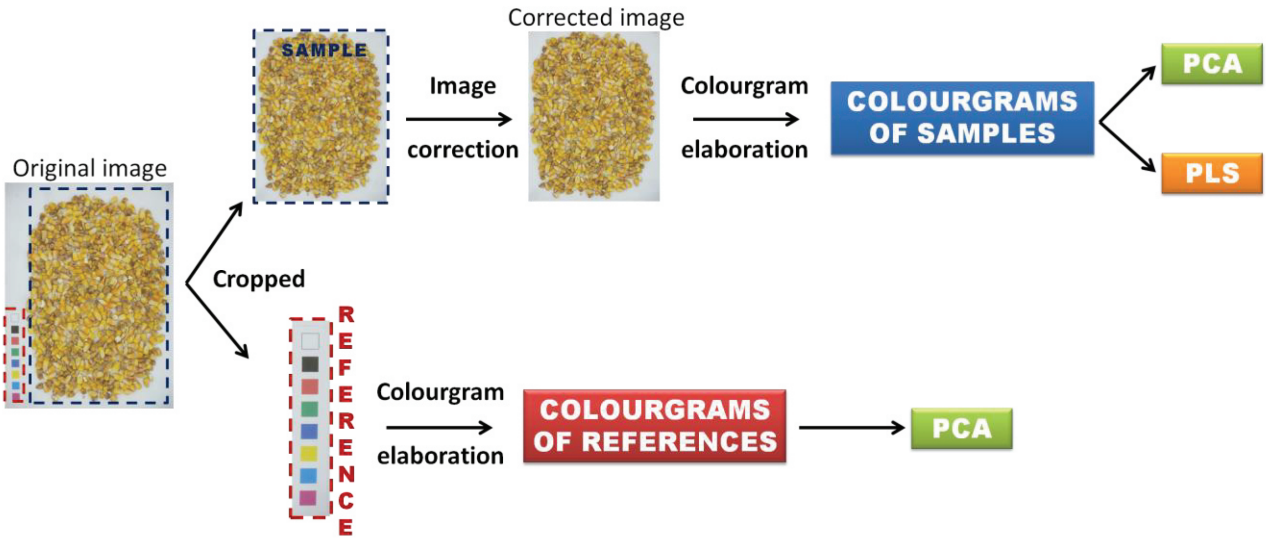

Fig. 2. Key steps of the image analysis followed for both considered acquisition methods (scanner and camera).

colourgrams can be subsequently elaborated by means of proper multivariate analysis techniques, like data exploration, calibration or classification. More details about the calculation of colourgrams can be found in [2].

2.3. Analysis of colourgrams

In this work, four data matrices of 156 rows (number of images/colourgrams) and 4900 columns (points of the colourgrams) were obtained:

1. Colourgrams of samples (scanner);
2. Colourgrams of references (scanner);
3. Colourgrams of samples (camera);
4. Colourgrams of references (camera).

In order to obtain a first overview of the dataset structure and to identify possible outlier images, PCA models were calculated on each colourgram matrix, which was previously preprocessed by autoscaling or meancentering.

Before calculating calibration models with the aim of predicting the percentage of defective kernels in each image, the 156 colourgrams were split into a training set of 84 signals and an external test set of 72 signals. In particular, PLS [3] was used as calibration method, considering venetian-blinds cross-validation with 4 deletion groups, were the replicate images were kept together. The optimal number of Latent Variables to include in the model was identified according to the minimum value of Root Mean Square Error in Cross-Validation (RMSECV).

3. RESULTS AND DISCUSSION

First of all, the colourgrams of standard references collected using both scanner and digital camera were investigated by means of PCA (Fig. 3). In particular for the images acquired using the digital camera, it was observed that the references follow a trend along PC1 according to the amount of defective maize of the corresponding sample. This effect is probably due to the light reflected by the sample contained in the image scene; in fact, 100% defective samples are globally darker than 0% defective samples. Nevertheless, in the analysis of the colourgrams of the samples, the importance of this effect is quite negligible compared to the effect of the percentage of defective maize kernels. However, this observation suggests that for future image acquisitions, the position of reference should be optimized in order to limit the effect of reflected light.

Subsequently, PCA models were calculated on the colourgrams of maize samples considering both autoscaling and meancentering as signal preprocessing methods. For both the acquisition methods, the corresponding PCA models (Fig. 4) showed that the samples are distributed along PC1 according to the percentage of defective kernels.

Furthermore, the PCA models calculated on autoscaled colourgrams allowed separating the samples according to the type of maize along PC2 and PC3. In particular dry and wet non-defective samples are separated each other along PC2 for

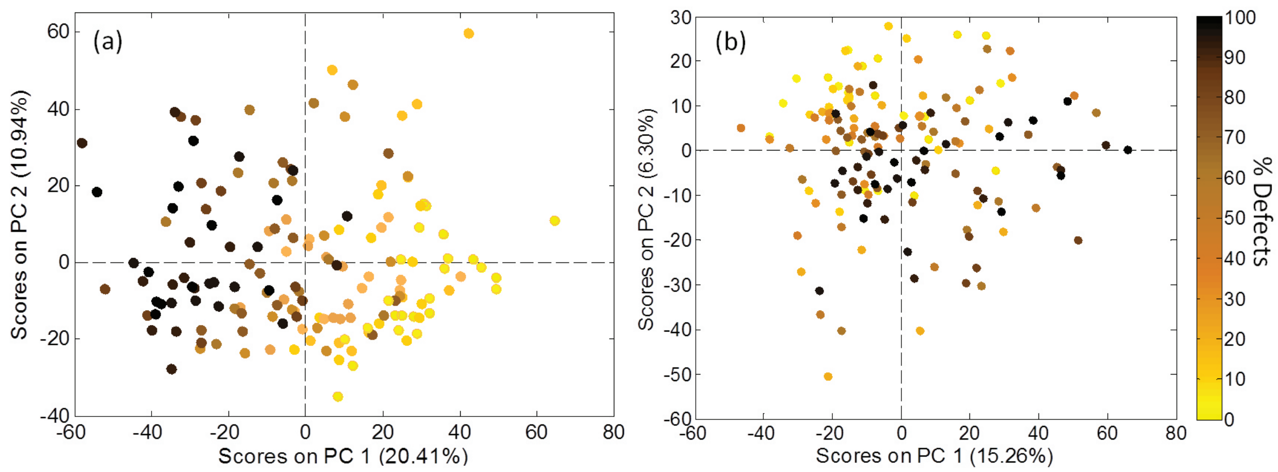

Fig. 3. PC1 vs PC2 score plots of the PCA models calculated on autoscaled colourgrams of camera (a) and scanner (b) references, coloured according to the concentration of defective maize kernels of the corresponding samples.

images acquired using the camera and along PC3 for images acquired using the scanner, while the opposite is observed for defective samples, which are separated each other along PC3 for images acquired using the camera and along PC2 for images acquired using the scanner.

The score plots of the PCA models calculated on autoscaled colourgrams of camera and scanner samples are reported in Fig. 4a and 4c, respectively, where the samples are coloured according to the percentage of defective maize, and in Fig. 4b and 4d, where the samples are coloured according to the type of maize.

The difference observed between dry and wet maize is ascribable to their different specific weight. Indeed, the kernels of wet maize are heavier, therefore the corresponding images include a lower number of kernels and more pixels related to the white background compared to the images of dry maize, even if the total weight of the sample is the same. Nevertheless, the difference between dry and wet maize is orthogonal to the percentage of defective maize. Therefore, the type of maize had a limited influence on the development of calibration models for the prediction of the percentage of defective maize kernels.

The best PLS model, highlighted in yellow colour in Table 1, was obtained considering the digital camera images and using autoscale as signal preprocessing method. This model led to a Root Mean Square Error in Prediction (RMSEP) equal to 4.8 %, estimated on the external test set.

Figure 5 shows the plot of the actual values of percentage of defective maize kernels versus the predicted values with the best PLS model.

This preliminary result is satisfactory, also taking into account the number and the variability of the analysed samples.

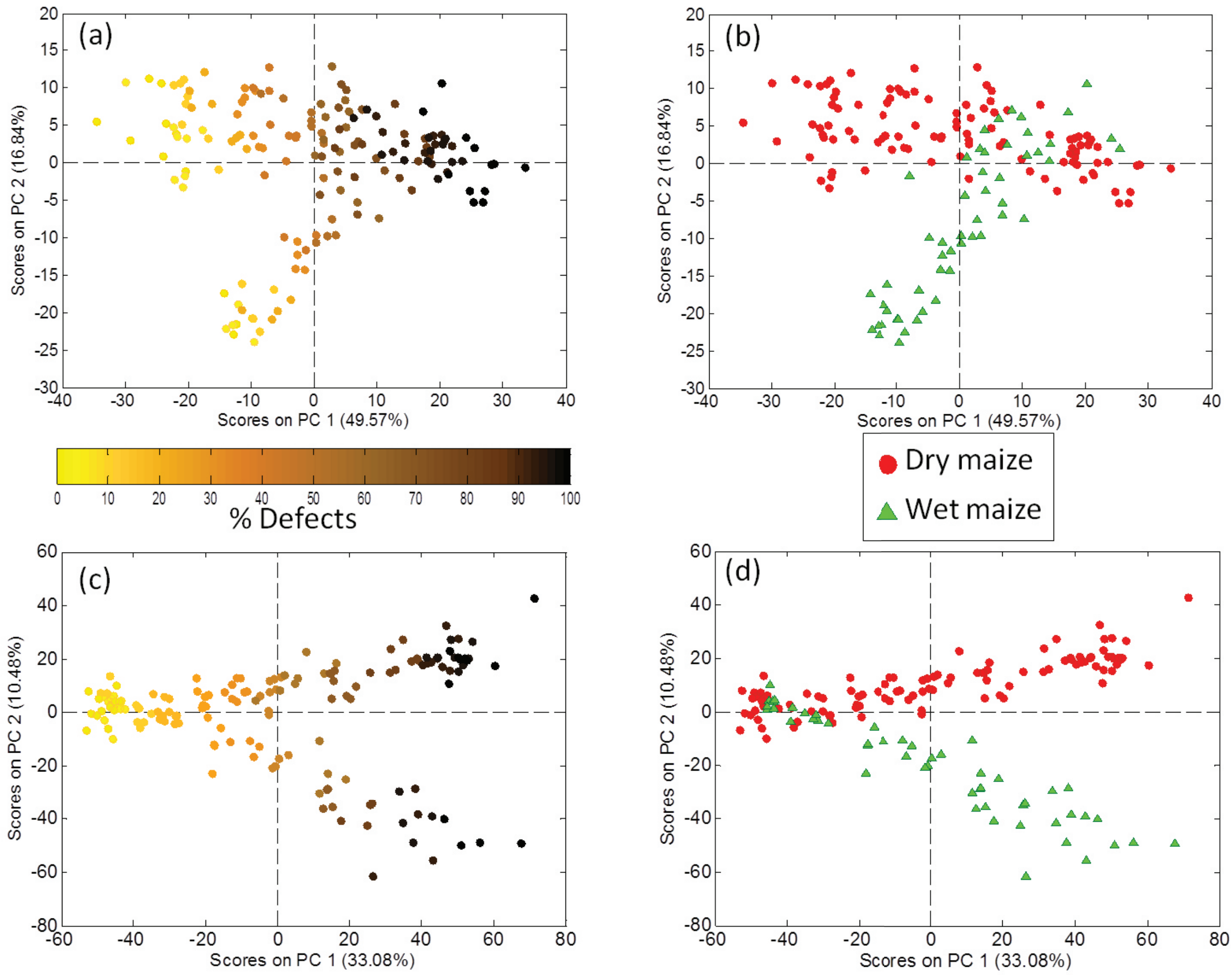

Fig. 4. PC1-PC2 score plots of the PCA models calculated on colourgrams of the images acquired using camera, (a) and (b), and scanner, (c) and (d). In (a) and (c) the samples are coloured according to the concentration of defective kernels, while in (b) and (d) the samples are coloured according to maize type.

Table 1. Results of the PLS calibration models. The best calibration model is highlighted in yellow colour.

Acquisition modality	Preproces.	LVs	RMSEC	RMSECV	RMSEP	R^2_{Cal}	R^2_{CV}	R^2_{Pred}
Digital camera	Autoscale	3	2.6	5.9	4.8	0.995	0.973	0.977
	Mean center	3	4.7	8.1	5.1	0.983	0.949	0.974
Scanner	Autoscale	2	3.4	7.5	4.8	0.991	0.956	0.977
	Mean center	3	4.1	7.9	4.9	0.987	0.952	0.976

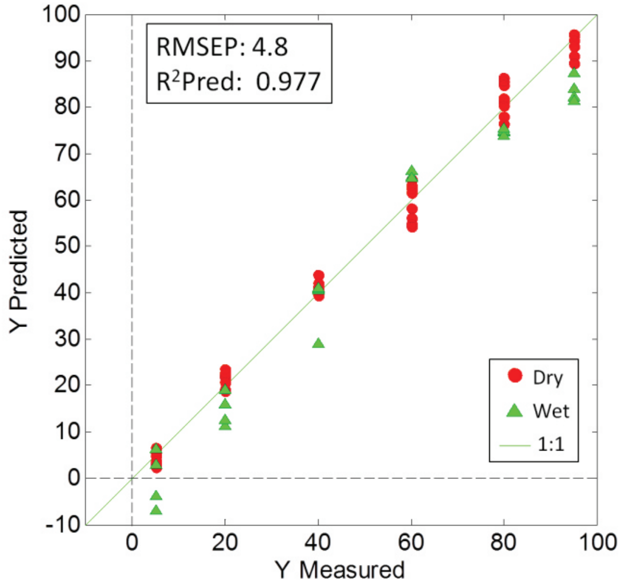

Fig.5. Actual percentage of defective kernels (Y Measured) vs. predicted percentage (Y Predicted) resulting from the best PLS calibration model applied to the test set.

4. CONCLUSIONS

The present work has supplied useful preliminary information for the possible development of an automated system, based on the analysis of RGB images, for the detection of defective maize kernels.

In particular, the obtained results allowed to define the best modality for image acquisition, to evaluate possible limitations in the instrumental setup, and to verify the effectiveness of processing the images by means of multivariate techniques.

ACKNOWLEDGMENTS

The authors wish to thank Freeray S.r.l. and Fornasier Tiziano & C. S.a.s. for technical and financial support.

REFERENCES

[1] M. Z. Hossain and T. Goto, “Near-and mid-infrared spectroscopy as efficient tools for detection of fungal mycotoxin contamination in agricultural commodities”, *World Mycotoxin Journal*, Vol. 7, nº. 4, pp. 507-515, June 2014.

[2] A. Antonelli, M. Cocchi, P. Fava, G. Foca, G.C. Franchini, D. Manzini and A. Ulrici, “Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-based classification algorithm”, *Analytica Chimica Acta*, nº. 515, pp. 3-13, February 2004

[3] P. Geladi and B. R. Kowalski, “Partial least-squares regression: a tutorial”, *Analytica Chimica Acta*, nº. 185, pp. 1-17, 1986.