CrossMark

A Proof of Concept Study for the Parameters of Corn Grains Using Digital Images and a Multivariate Regression Model

Vanessa Rodrigues de Camargo[1] · Lucas Janoni dos Santos[1] ·
Fabíola Manhas Verbi Pereira[1]

Abstract In this method, a numerical matrix comprised of ten color scales (RGB, HSV, L, and rgb) as independent variables from digitalized images was used as a proof of concept for the prediction of the mass, apparent volume, and bulk density parameters of grains for quality control considering post-harvest purposes. The goal was to develop a high throughput multivariate regression model using partial least squares (PLS) combined with the information from color images to assess the raw product. The data set of external samples was successfully evaluated with standard error of cross-validation (SECV) values of 1.23 g (16.4–28.9), 2.03 cm^3 (20.5–40.5), and 0.018 g cm^{-3} (0.68–0.85) for the mass, apparent volume, and bulk density, respectively.

Introduction

Image analysis converted into mathematical arrays can be performed with the aid of several chemometric tools (Russ, 1992; Wojnar, 1999; Liu & MacGregor, 2007; Pereira & Bueno, 2007). In this study, the principle will be an emphasis on the

✉ Fabíola Manhas Verbi Pereira
fabiola@iq.unesp.br

[1] Instituto de Química de Araraquara, Universidade Estadual Paulista (UNESP), Rua Professor Francisco Degni, 55, Araraquara, SP 14800-060, Brazil

partial least squares (PLS) regression method (Geladi et al., 2004).

Walker and Panozzo (Walker & Panozzo, 2012) evaluated three different mathematical models for measuring the volume and bulk density of a barley grain using an ellipsoid approximation from a two-dimensional digital image. They achieved correlations to grain features of 0.97 and 0.63 for the volume and bulk density, respectively. However, it is not reported how the PLS is able to perform for the prediction of these properties using a single model.

By means of the red, green, and blue (RGB) scale and chemometrics, Borin et al. (Borin et al., 2007) accomplished the quantification of *lactobacilli* colonies in commercial fermented milk using digital images provided by a household scanner. The authors applied one-dimensional vector data based on the frequency distribution of values of curves (histograms) for the three colors (RGB). Two models have been developed: (i) the first was nonlinear using LS-SVM (support vector machine least squares) and (ii) the second was linear based on PLS. Both models were used to calculate the number of *lactobacilli* colonies. The relative error for the quantification was approximately 10%. This fact indicated that the proposed method can be used for automated counting of this type of colony.

The application of the RGB scale to images scanned using a conventional digital camera (CCD, common digital camera) in food samples was investigated by Antonelli et al. (Antonelli et al., 2004). The authors used these images to obtain colorgrams for the three colors (red, green, and blue). A recognition model was obtained using the colorgrams from the pesto sauce samples and an algorithm based on the wavelet transformation. The authors reported having achieved an accuracy of 100% for their ratings with this method.

Santos et al. (Santos et al., 2012) and Santos and Pereira-Filho (Santos & Pereira-Filho, 2013) have reported two

studies that employed digital images obtained in a conventional scanner for the identification of adulterated cow milk samples. The authors used the following ten color parameters: R, G, B, H, S, V (value), r, g, b, and L. The r, g, and b color parameters are related, that is, they are each a primary color (R, G, B) divided by the sum of the colors (L). The authors investigated the presence of many types of adulterants that are commonly added to cow milk, including water, synthetic milk, cow urine, and hydrogen peroxide, among others. In the case of regression models, they used the PLS technique, and prediction errors were approximately 6%. As for the classification models, hit rates were obtained in the range of 90 to 100%.

The goal of this study was the development of a method for determining the mass, apparent volume, and bulk density of corn grains for the assessment of the raw product of the most widely grown crop in the Americas.

This method was developed with the use of images acquired through a conventional household scanner, thus enabling important information to be obtained without subjectivity. Information of the color histograms from the RGB (red, green, and blue), HSV (hue, saturation, and value), L (luminosity), and rgb (relative RGB) of digital images was organized as a numerical matrix. This evaluation was performed with the use of chemometric tools that were able to better understand the generated data.

With the proof of concept described in this study, it can already be inferred that the multivariate method has potential for predicting the mass, apparent volume, and bulk density in corn grains.

Materials and Methods

Samples and Instrument

Two types of samples of corn grain were purchased locally. The identifying characteristic of these grains was their size, with the largest ones measuring approximately 10 by 10 mm, being the corn as type 1 which is employed to feed animals, and with the smallest ones measuring approximately 5 by 5 mm, which are those that are specific to popcorn (type 2). For sampling of the material, the entire grains, all of regular size and no defects, were distributed in a 10 cm by 6 cm transparent plastic package so that the contents of each had a total of 100 entire grains.

The images were obtained by placing the plastic packaging directly on the glass of a conventional scanner (HP Color LaserJet Pro MFP 200 M276nw, Brazil). For the acquisition of images, the scanner automatic settings were disabled, such as the brightness and contrast adjusts, and the resolution was 300 dpi (dots *per* inch). The cover of the scanner was closed to avoid any influence of external light.

Physical Trait Measurements

The mass (m) was estimated using an analytical electronic balance (FA-2104N, EQUIPAR, Curitiba/PR, Brazil) with a precision of ± 0.0001 g. The apparent volume (V) was measured with the help of a 50-mL graduated cylinder (± 0.5 mL). The relationship between the values of mass and apparent volume was used to estimate the bulk density (g cm^{-3}), where $\rho = m/V$.

Image Evaluation

The data were treated using computer programs such as Matlab R2015b (The MathWorks, Natick, USA) and Pirouette 4.5 rev. 2 (Infometrix, Bothel, USA). The original images were converted to the color gamut RGB [red (R), green (G), and blue (B)], HSV [hue (H), saturation (S), and value (V)], the relative value scale for RGB colors (rgb), and L (luminosity, light) by means of a computer routine available as Supplementary content, which can be calculated with Matlab. The size of all images was standardized to be 500 × 850 pixels (width × height). Samples were placed in training and validation sets by means of the Kennard-Stone algorithm considering 80 and 20% of the samples, respectively (Daszykowski et al., 2002).

Results and Discussion

Corn type 1 represents the suitable target for grain production considering that the size, mass, and geometry are more homogeneous, and corn type 2 typifies the features not proper for the raw product. With the aid of Fig. 1, examples of images of samples packaged in plastic can be verified, since these were used to develop the proposed method.

The images, such as those shown in Fig. 1, have been previously converted into averages of color histograms. The independent variables were the colors from ten scales (RGB, HSV, rgb, and L), and the response was how often they are repeated. As shown in the pictorial of Fig. 2, the advantage of the average of color histograms is the size of the matrix with 10 columns and 1 numerical response for each color. In other words, each sample was represented by 10 averaged values *per* color (50 samples × 10 variables). In this case, the data can be computed faster since each original colorgram has 256 variables for each color (total of 2560 independent variables).

By the visual inspection of images in Fig. 1, it is possible to differentiate the size of grains for the homogeneous composition with 100 larger grains (Fig. 1a) or even 100 smaller grains (Fig. 1b), but for the image in Fig. 1c, it is very hard to differentiate that it represents the composition of 55 larger grains and 45 smaller ones. To overcome this difficulty, the image can be

Fig. 1 Examples of scanned images of representative samples. **a** With larger grains. **b** With smaller grains. **c** With the proportion 55:45 for larger and smaller grains, respectively
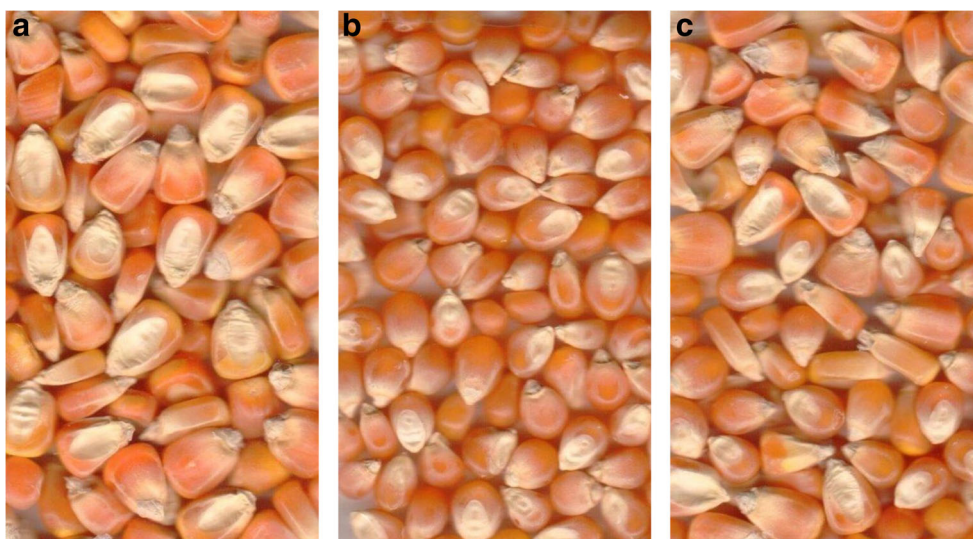
converted into a numerical matrix using the color scale as parameters, as previously described.

The computational routine to process the calculations in the image and different magnitudes may be checked as follows: (i) the scale is represented by RGB values between 0 and 255; (ii) for HSV and the relative colors of RGB (rgb), the magnitude range is between 0 and 1; and (iii) the luminosity is the sum of the RGB colors.

For example, the differences among the scales were on the order of hundreds, and as described, the minimum and maximum values were as follows: R (196–217), G (122–145), B (75–100), H (0.06–0.07), S (0.55–0.64), V (0.77–085), r (0.49–0.53), g (0.30–0.31), b (0.17–0.20), and L (397–461).

The luminosity (L) has the highest values and the largest range of variation. The other colors showed faint differences, and in the case of HSV and rgb, they had ranges with lower magnitudes. Another piece of relevant information from the color images is that the blue scale showed correlation with the parameters being investigated. This is strong evidence that the colors from the images can detect variations related to the shape and can differentiate questions about interferences such as leaves or soil. The color yellow is the most predominant nuance for corn grains visualized with the naked eye. To support this statement, the blue color is the complementary color of the yellow in visible radiation (Boynton, 1960).

A high variation of luminosity corroborates with the information extracted from the images, which is in fact related to the size of the grains, promoting the possibility of detecting variations between big grains mixed with small ones or even in the presence of undesirable components, such as leaves, soil, or other residues from the harvest.

The suggestion of a multivariate model can be interesting because the mass and apparent volume are extensive properties of the materials and they vary with different compositions, as shown in Table 1. PLS is the chemometric tool that is more robust and disseminated for predicting these parameters, and it is possible to use more than one variable to predict, since the data of this study did not show high correlation with only one independent variable (Geladi et al., 2004; Santos et al., 2012; Santos & Pereira-Filho, 2013).

For analysis with PLS, the matrix was organized with the samples in the rows and the independent variables in columns

Fig. 2 Picture of the arrangement of the matrix color scale from the corn grain images shown in Fig. 1

Independent variables								

Sample a (100 larger grains)

R	G	B	H	S	V	r	g	b	L
214	139	92	0.1	0.6	0.8	0.5	0.3	0.2	445

Sample b (100 smaller grains)

R	G	B	H	S	V	r	g	b	L
205	129	86	0.1	0.6	0.8	0.5	0.3	0.2	419

Sample c (55 larger and 45 smaller grains)

R	G	B	H	S	V	r	g	b	L
207	135	92	0.1	0.6	0.8	0.5	0.3	0.2	434

Table 1 Physical traits estimated for corn grain samples. The square symbols represent the type of sampling on the graphs from the developed model

Number of samples (plastic bags)	Type 1 (larger grains) per bag	Type 2 (smaller grains) per bag	Mass/g ($\bar{x} \pm$ sd)	Apparent volume/cm³ ($\bar{x} \pm$ sd)	Bulk density/g cm⁻³ ($\bar{x} \pm$ sd)
10 ■	100	0	25.4 ± 0.5	36.1 ± 0.7	0.70 ± 0.01
10 ■	0	100	17.0 ± 0.3	21.1 ± 0.5	0.80 ± 0.01
3 □	5	95	18.6 ± 0.8	23.2 ± 1.2	0.80 ± 0.02
3 □	15	85	19.6 ± 1.5	25.2 ± 1.5	0.78 ± 0.02
3 □	25	75	20.4 ± 1.2	26.2 ± 1.5	0.78 ± 0.02
3 □	35	65	21.6 ± 1.8	27.8 ± 2.5	0.78 ± 0.01
3 □	45	55	23.4 ± 2.2	29.5 ± 1.0	0.79 ± 0.05
3 ■	55	45	23.5 ± 1.6	32.5 ± 2.0	0.72 ± 0.01
3 ■	65	35	24.6 ± 1.6	33.8 ± 3.2	0.73 ± 0.03
3 ■	75	35	24.8 ± 2.6	34.2 ± 3.1	0.73 ± 0.01
3 ■	85	15	25.7 ± 2.9	35.2 ± 3.8	0.73 ± 0.01
3 ■	95	5	26.4 ± 2.2	36.8 ± 3.2	0.72 ± 0.01

corresponding to the color scales. To better analyze this data, the independent variables were auto-scaled (Geladi et al., 2004) so that all passes had a zero mean and standard deviation equal to one. The goal is for all measured variables to contribute the same weight, even if the magnitudes between the scales are very different. The dependent variables were the mass, apparent volume, and bulk density of the grains measured earlier.
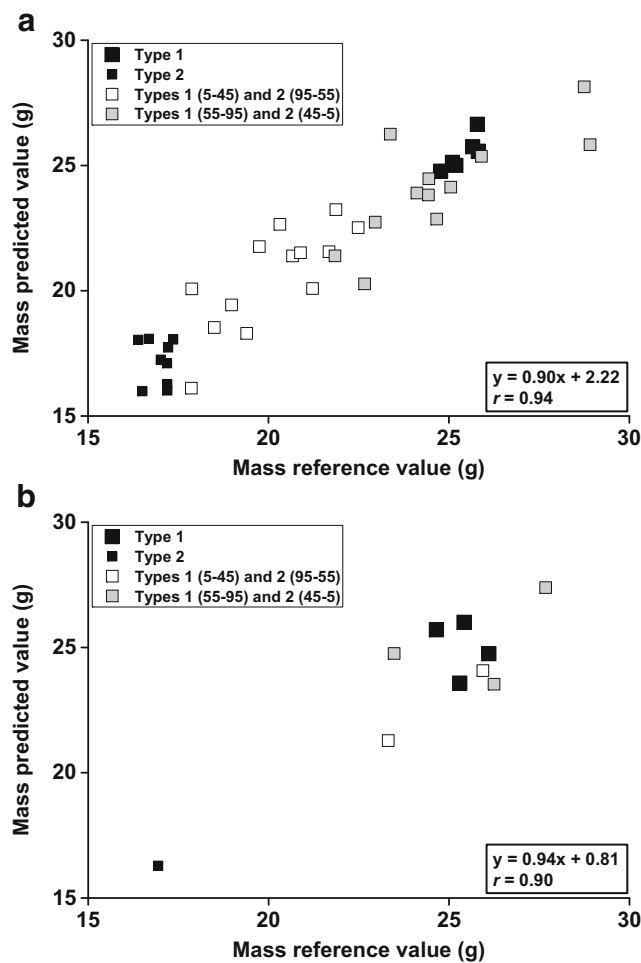

Fig. 3 Predicted values for the PLS model for the mass of grains using the data of the image colors for the **a** training and **b** validation sets
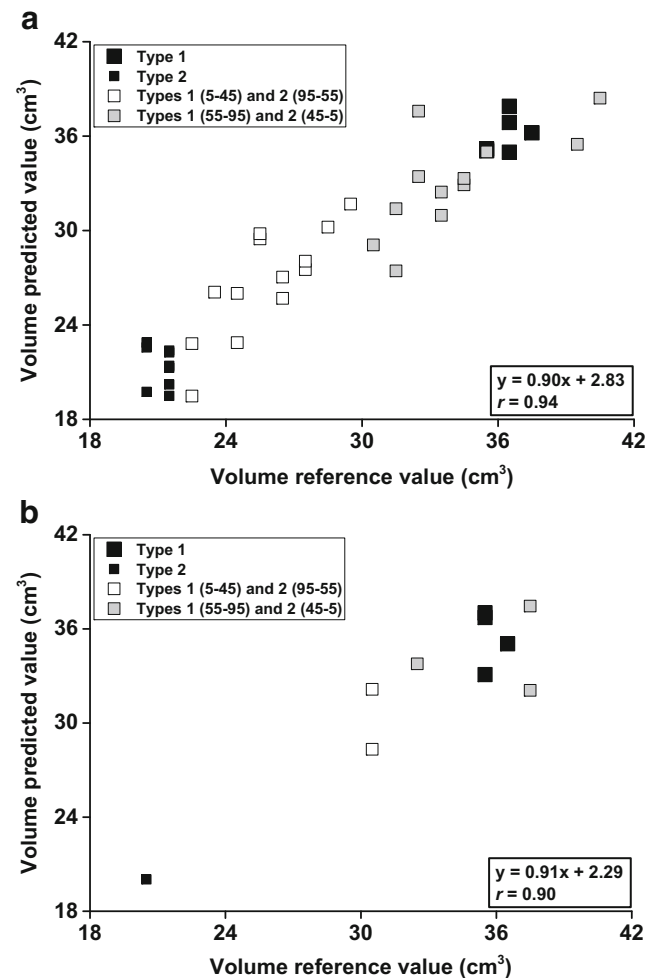

Fig. 4 Predicted values obtained by the PLS model for the apparent volume of grains using the data of the image colors for the **a** training and **b** validation sets

The standard error of cross-validation (SECV) can assess the predictive ability of the models. In this case, using five latent variables (LV) with a total explained variance of approximately 100%, the RMSECV values were 1.23 g, 2.03 cm^3, and 0.018 g/cm^3 for the mass, apparent volume, and bulk density, respectively.

Figures 3, 4, and 5 show the good linear correlation between the reference values for the parameters measured for the corn grains and those predicted by the PLS prediction model. The numbers between brackets in the legend of these figures denote the range of the proportion for larger and smaller grains, as shown in Table 1. For all plots, it is possible to verify the spreading of the data along the x-coordinate, and this information confirmed that the sampling has the variation necessary and the color scales from images detected these differences among them. In this case, the mass and apparent volume suffered variations according to the levels of the undesirable corn (type 2).
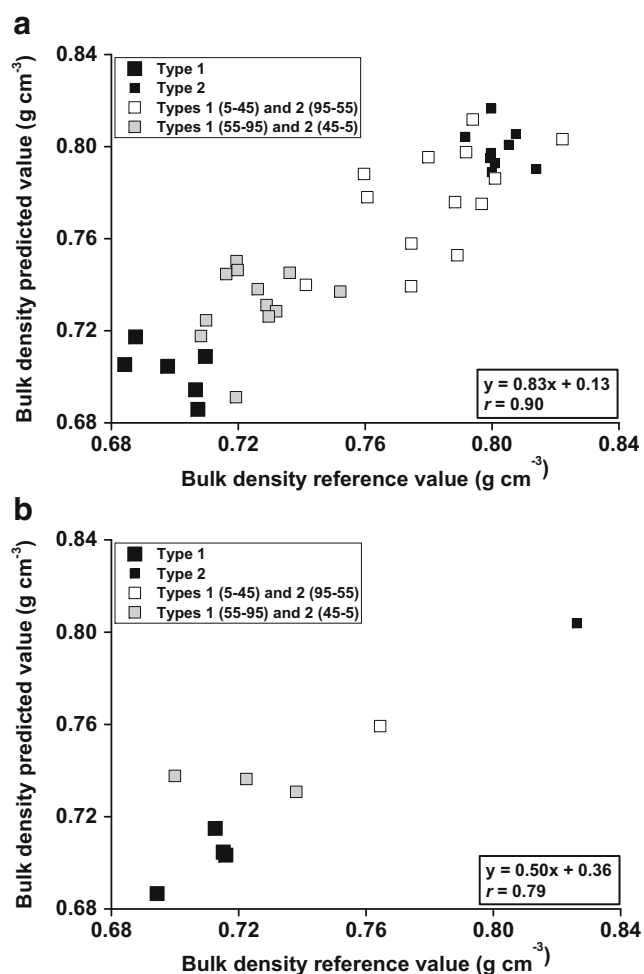
Conclusions

This study can generate fast and accurate methods for predicting the mass, apparent volume, and bulk density of grains as an alternative method to the existing gravimetric and geometric methods. The color scales from digital images provide information that proved be useful to detect variations between grains for post-harvest purposes with some advantages, such as its practicality, economy, safety, efficiency, speed, and accuracy.

Compliance with Ethical Standards

Conflict of Interest Fabíola Manhas Verbi Pereira declares that she has no conflict of interest. Vanessa Rodrigues de Camargo declares that she has no conflict of interest. Lucas Janoni dos Santos declares that he has no conflict of interest.

Ethics Approval This article does not contain any studies with human or animal subjects performed by any of the authors.

Consent for Publication Publication has been approved by all individual participants.

Fig. 5 Predicted values obtained by the PLS model for the bulk density of grains using the data of the image colors for the **a** training and **b** validation sets

References

Russ JC (1992) Image processing handbook. CRC Press, New York

Wojnar L (1999) Image analysis: applications in materials engineering. CRC Press, Boca Raton

Liu JJ, MacGregor JF (2007) On the extraction of spectral and spatial information from images. Chemom Intell Lab Syst 85:119–130

Pereira FMV, Bueno MIMS (2007) Image evaluation with chemometric strategies for quality control of paints. Anal Chim Acta 588:184–191

Geladi P, Sethson B, Nyström J, Lillhonga T, Torbjörn L, Burger J (2004) Chemometrics in spectroscopy. Part 2. Examples. Spectrochim Acta B 59:1347–1357

Walker CK, Panozzo JF (2012) Measuring volume and density of a barley grain using ellipsoid approximation from a 2-D digital image. J Cereal Sci 55:61–68

Borin A, Ferrão MF, Mello C, Cordi L, Pataca LCM, Durán N, Poppi RJ (2007) Quantification of *Lactobacillus* in fermented milk by multivariate image analysis with least-squares support-vector machines. Anal Bioanal Chem 387:1105–1112

Antonelli A, Cocchi M, Fava P, Foca G, Franchini GC, Manzini D, Ulrici A (2004) Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-based classification algorithm. Anal Chim Acta 515:3–13

Santos PM, Wentzell PD, Pereira-Filho ER (2012) Scanner digital images combined with color parameters: a case study to detect adulterations in liquid cow's milk. Food Anal Methods 5:89–95

Santos PM, Pereira-Filho ER (2013) Digital image analysis—an alternative tool for monitoring milk authenticity. Anal Methods 5:3669–3674

Daszykowski M, Walczak B, Massart DL (2002) Representative subset selection. Anal Chim Acta 468:91–103

Boynton RM (1960) J Opt Soc Am 50:929–944