

Non-Contact Measurement of Cereal Quality by Image Sensing and Numerical Regression Techniques

W. B. Hussein¹, A. A. Moaty², K. A. Nagaty¹ and M. A. Hussein²

¹Faculty of Informatics and Computer Science, The British University in Egypt, Cairo, Egypt

²Department of Life Science Engineering, Technische Universität München, Freising, Germany

Abstract. In this paper, digital image processing techniques are applied to measure some of the quality parameters of the durum wheat semolina. One of these parameters is the semolina colour value in the lab colour space $L^*a^*b^*$, which is the commonly employed colour space in food field. Several numerical methods are developed and analysed for mapping the RGB digital images to $L^*a^*b^*$. These methods are direct, polynomial regression, and neural network methods. The accuracy of each method is obtained with respect to the measured $L^*a^*b^*$ values captured with a Chroma-Meter instrument. The numerical models outcomes showed lowest colour deviations of 0.72. The results also demonstrated a significant effect of the training data set on the numerical $L^*a^*b^*$ outputs. Moreover, a partial least-squares regression model was developed to numerically predict the β -carotene content in semolina, as another important quality parameter. The model proved a correlation coefficient of 0.94 between numerical predictions and experimental measurements according to the ICC standard method 152 for extracting the durum carotenoids, thus bears a high potential for facilitating carotene detection in durum.

1 Introduction

Food colour is the primary quality parameter checked by consumers, who use it as a tool to accept or reject the product. In addition, the colour has been widely demonstrated that it correlates with physical, chemical and sensorial indicators of product quality [1]. Therefore, many researches indicated that the observation of colours drives a potential stream for detection of several defects that the food items may have [2]. For example, the colour in durum is directly related to its trade value; at which high colour durum preserve higher value.

The analysis of the surface colour of a food product is normally performed by obtaining colour distribution and average values for the tested samples [3]. The colour measuring instruments (colorimeters) are commonly used since they are measuring the colours values mapped on a device-independent colour space with the coordinate L^* for the lightness and a^* and b^* the colour-opponent dimensions. Some of these instruments require measuring samples that have extremely smooth and flat surfaces to avoid the dissipation of light in/out the measured area under the device prop. Within the food products, the flatness and smoothness of the tested samples is not guaranteed as well the accuracy of the colour measurements.

As an alternative solution, the RGB digital cameras can offer a reasonable solution, which can obtain information about the colour of the whole surface of the food product in terms of pixels.

It is always a challenge to drive computational models capable to convert the RGB digital images to $L^*a^*b^*$ colour space with reasonable accuracy [4], [5]. As a result, a room has been made for researchers to develop and improve several conversion models, and to implement them as quality measurement criteria [6].

In this work, a novel quantitative analysis of the (*Triticum durum*) semolina colour was performed. According to the expectations of consumers to have uniform, amber-yellow coloured pasta without shades of grey or red, the yellowness of the semolina was used to evaluate the quality of the semolina [7]. Therefore, the b^* value was used as an indicator of the yellowness since it represents the colours variation from blue to yellow [8]. Currently, the b^* value is determined by gathering some measurement using the colorimeter devices. Due to the previously illustrated drawbacks of this technique, the RGB digital images were selected to be used as an alternative solution to get the b^* value of semolina in a pixels wise approach. Three different numerical colour calibration methods are proposed to obtain accurate information about the b^* value. The first method is the direct transformation from RGB to $L^*a^*b^*$, by mapping the RGB on the XYZ colour space then transform the XYZ values to $L^*a^*b^*$ [9]. The second method is based on polynomial regression functions [10]. 1st, 2nd and 3rd order polynomials are investigated in order to find the proper function that accurately relates the RGB values to the $L^*a^*b^*$ space. The third method is using artificial neural network to design a nonlinear transformation

model. Finally, since studies indicate that the b^* value is correlated with the carotene content [11], a numerical model implementing the partial least-squares regression method (PLS) is constructed to estimate the β -carotene content (yellow pigments) in semolina by having the colour (b^* value) and the percentage humidity as input variables.

2 Material and methods

Sixty-four semolina samples of 0.5mm particle size were randomly selected from different locations. Each sample was placed in measuring cup with 5cm diameter to measure the $L^*a^*b^*$ colour components using a Minolta CR-200 Chroma-Meter instrument. For each sample, five enclosed regions of 1 cm² area were marked as shown in Figure 1.

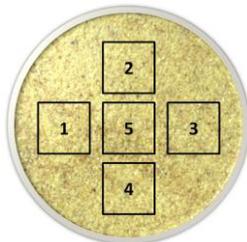


Figure 1. One semolina sample showing the selected five locations for measuring the colour values.

The mean RGB values with their corresponding $L^*a^*b^*$ measurements were calculated for each enclosed region. A data set out of 320 mean RGB values with their corresponding $L^*a^*b^*$ measurements were constructed to calibrate and validate the numerical colour calibration models proposed in this paper. The β -carotene content in these samples were extracted according to the ICC-152 standard method for yellow pigment extraction. The β -carotene content was calculated using equation (1).

$$\beta - \text{carotene (mg)} = \left(\frac{a * 5 * 10}{1000} \right) \left(\frac{100}{100 - b} \right) \quad (1)$$

where a is the mg of β -Carotene in 10 ml of extract and b is the percentage relative humidity. After constructing the acquisition system, a set of RGB- $L^*a^*b^*$ data was prepared using several colour charts. The best calibration process of the acquisition setup was achieved when 17 colour charts were employed, as in Figure 2-a.

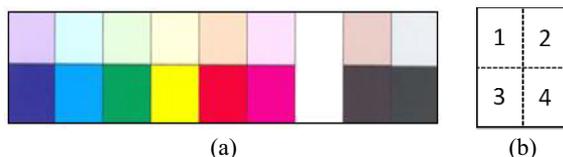


Figure 2. (a) Colour chart to generate the calibration data set. (b) The dividing method applied on each coloured area in order to increase the number of points.

Each colour was divided into 4 regions as in Figure 2-b. In each region, the $L^*a^*b^*$ values were measured using Minolta CR-200 Chroma-meter instrument. The

mean of the colour values in each region was measured using equation (2).

$$R, G, B_{\text{mean}} = \frac{\sum_{i=1}^{i_{\text{max}}} \sum_{j=1}^{j_{\text{max}}} P_{R,G,B}}{N_p} \quad (2)$$

where $P_{R,G,B}$ is the pixel R , G and B values; the N_p is the number of pixels enclosed in each region; i, j is the row and column index respectively

2.1 Image acquisition

An industrial colour camera type Basler Pilot-piA2400 CCD camera with a resolution of 5 Mega Pixels (2456 x 2058) and pixel size (3.45 x 3.45) μm . The illumination was achieved by using 2 Philips day light fluorescent lamps with colour temperature of 6500 K approximately.

The RGB- $L^*a^*b^*$ transformation consists of two phases: first the background is eliminated using histogram analysis techniques as shown in Figure 3-(b).

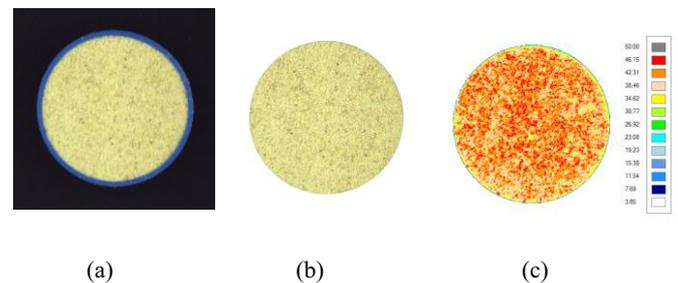


Figure 3. (a) Original RGB image, (b) the digital image after eliminating the background noise, (c) colour contoured image for representing the numerically calculated b^* value

The RGB- $L^*a^*b^*$ colour conversion took place using the thresholded image with different numerical colour calibration models. Finally, the numerically calculated $L^*a^*b^*$ values were arranged over the pixels location and exported in a proper format such as colour contoured images as shown in Figure 3-(c).

Three types of colour calibration models were used to perform the RGB- $L^*a^*b^*$ transformation: the direct model, polynomial regression model and neural network model. The b^* value in the colour space ($L^*a^*b^*$) which represents the chromaticity coordinate for yellow-blue ($+b^*$ = yellow direction; $-b^*$ = blue direction) is used as an indicator of the yellowness of the tested sample since it represents the colours variation from blue to yellow while ignoring the L^* and a^* values. A special colour difference formula was used to evaluate the accuracy of the different models. This colour difference formula is based only on the Δb^* as in equation (3).

$$\Delta E_{b^*} = \sqrt{\Delta b^{*2}} \quad (3)$$

The accuracy of each model will be evaluated with the mean and the maximum value of the ΔE_{b^*} difference calculated for all the tested samples.

2.2 Direct model

This calibration model performs the RGB-L*a*b* by firstly using equation (4).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [M]_{3 \times 3} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4)$$

where matrix M depends on the chromaticity coordinates of the used RGB system and its reference white point. Convert the obtained XYZ values to L*a*b* colour space according to equation (5).

$$\begin{aligned} L^* &= \begin{cases} 116 \left(\sqrt[3]{\frac{Y}{Y_r}} \right) - 16 & \text{if } \frac{Y}{Y_r} > \varepsilon \\ 903.3 \left(\frac{Y}{Y_r} \right) & \text{if } \frac{Y}{Y_r} \leq \varepsilon \end{cases} \\ a^* &= 500 \left[\left(\sqrt[3]{\frac{X}{X_r}} \right) - \left(\sqrt[3]{\frac{Y}{Y_r}} \right) \right] \\ b^* &= 200 \left[\left(\sqrt[3]{\frac{Y}{Y_r}} \right) - \left(\sqrt[3]{\frac{Z}{Z_r}} \right) \right] \end{aligned} \quad (5)$$

Secondly convert the obtained XYZ values to the L*a*b* colour space as shown in equation (5). The equation for mapping relative luminance (Y/Y_r) into lightness L^* is composed of two separate functions, $f(Y/Y_r)$ and $g(Y/Y_r)$. The two functions are grafted together at junction point $\varepsilon = 0.008856$ according to the CIE standard. X_r, Y_r, Z_r are the XYZ tri-stimulus values of the reference white point which equal to (0.95047, 1.0, 1.08883) according to the CIE standard illumination D65 (colour temperature of 6500 K approximately).

2.3 Polynomial regression model (PRM)

Polynomial functions of different orders (i.e., from order 1 to order 3) were investigated. For each polynomial function, the relation between the RGB and the L*a*b* colour components were assumed to follow a polynomial function of the RGB colour values and some other calibration parameters M , equation (6).

$$\begin{bmatrix} L^* \\ a^* \\ b^* \end{bmatrix} = \begin{bmatrix} f_L(M_L, R, G, B) \\ f_a(M_a, R, G, B) \\ f_b(M_b, R, G, B) \end{bmatrix} \quad (6)$$

where $M_{L,a,b} = [M_1 \ M_2 \ \dots \ M_n]^T_{L,a,b}$ are 3 vectors of the model calibration parameters for L^* , a^* and b^* , respectively. Using the PRM, the RGB-L*a*b* transformation process is performed by firstly, using a training data to calculate the model calibration parameters M based on the minimization of the mean absolute error between the estimated L*a*b* variables (model output) and the L*a*b* variables measured by the colorimeter device [2]. Secondly, after calculating the M parameters a test data is used to evaluate the accuracy of the PRM model. The numerical L*a*b* values are compared with

$$\begin{aligned} 3^{rd} PRM &= a_{19}R + a_{18}G + a_{17}B + a_{16}RG + a_{15}RB + a_{14}GB + a_{13}R^2 + a_{12}G^2 + a_{11}B^2 + a_{10}RGB + a_9R^2G + a_8R^2B + a_7G^2R \\ &\quad + a_6G^2B + a_5B^2R + a_4B^2G + a_3R^3 + a_2G^3 + a_1B^3 + a_0 \end{aligned} \quad (13)$$

the measured L*a*b* by calculating the colour differences using equation (3) for each tested point. The model calibration parameters (M) are calculated in equation (7), as a function of some RGB and L*a*b* inputs.

$$M = [X^T X]^{-1} X^T Y \quad (7)$$

where f_L, f_a, f_b are the constructed regression functions of 1st, 2nd, or 3rd order. X and Y are the input matrices which carry (n) measurements of (RGB) and (L*a*b*) inputs respectively. The structure of the X input matrix is formed according to the order and number of terms in the polynomial functions. For a first order polynomial regression model, the relation between the RGB and the L*a*b* is shown in equation (8).

$$\begin{bmatrix} L^* = a_3R + a_2G + a_1B + a_0 \\ a^* = b_3R + b_2G + b_1B + b_0 \\ b^* = c_3R + c_2G + c_1B + c_0 \end{bmatrix} \quad (8)$$

Therefore, by setting the RGB inputs in a matrix form, the X matrix for n input points is given in equation (9).

$$X = \begin{bmatrix} R_1 & G_1 & B_1 & 1 \\ R_2 & G_2 & B_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ R_n & G_n & B_n & 1 \end{bmatrix} \quad (9)$$

For the second order polynomial shown in equation (10), the X matrix of $10 * n$ elements as shown in equation (11) will be used to calculate the calibration parameters M .

$$\begin{aligned} 2^{nd} PRM &= a_9R + a_8G + a_7B + a_6RG + a_5RB + a_4GB \\ &\quad + a_3R^2 + a_2G^2 + a_1B^2 + a_0 \end{aligned} \quad (10)$$

$$X = \begin{bmatrix} R_1 & G_1 & B_1 & RG_1 & RB_1 & GB_1 & R^2_1 & G^2_1 & B^2_1 & 1 \\ R_2 & G_2 & B_2 & RG_2 & RB_2 & GB_2 & R^2_2 & G^2_2 & B^2_2 & 1 \\ \vdots & \vdots \\ R_n & G_n & B_n & RG_n & RB_n & GB_n & R^2_n & G^2_n & B^2_n & 1 \end{bmatrix} \quad (11)$$

The second order polynomial in equation (11) will be enriched with one additional RGB terms in order to investigate the influence of the combined third order term. In this case, the X matrix structure will have a structure of $11 * n$ terms as seen in equation (12).

$$X = \begin{bmatrix} R_1 & G_1 & B_1 & RG_1 & RB_1 & GB_1 & R^2_1 & G^2_1 & B^2_1 & RGB_1 \\ R_2 & G_2 & B_2 & RG_2 & RB_2 & GB_2 & R^2_2 & G^2_2 & B^2_2 & RGB_2 \\ \vdots & \vdots \\ R_n & G_n & B_n & RG_n & RB_n & GB_n & R^2_n & G^2_n & B^2_n & RGB_n \end{bmatrix} \quad (12)$$

The same procedure is used for the third order polynomial, equation (13). The X matrix will follow the same structure like above but with 20 different terms in each row.

For the $L^*a^*b^*$ input data, the Y matrix of $3 \times n$ elements (the L^* , a^* and b^* components for different n points), equation (14), will be used for all the polynomial regression models.

$$Y = \begin{bmatrix} L_1 & a_1 & b_1 \\ L_2 & a_2 & b_2 \\ \vdots & \vdots & \vdots \\ L_n & a_n & b_n \end{bmatrix} \quad (14)$$

Finally, the test data set (RGB input and $L^*a^*b^*$ measurements) is combined with the previously calculated (M) parameters and supplied to model. The model will transform the inputs RGB points to $L^*a^*b^*$ then calculate the colour differences between the numerical outcomes and the $L^*a^*b^*$ input measurements. The RGB- $L^*a^*b^*$ transformation is performed through equation (15).

$$[L \ A \ B] = [M]^T [R \ G \ B] \quad (15)$$

where $[L \ A \ B]$ is a vector consists of three elements L^* , a^* and b^* and The $[RGB]$ is a vector of R, G and B terms having the same structure and order of the implemented polynomial function.

2.4 Neural network model (NNM)

A feed forward back propagation multi-layer perceptron network was designed. The input layer contains three neurons for the three colour values (R, G, and B), 8 hidden layer neurons and the output layer has three neurons for the developed values of (L^* , a^* , and b^*).

2.5 Partial least squares regression model

It is based on principal component scores of both the independent X variables (b^* value and the percentage of humidity of a sample) and dependent Y variable (β -carotene content in semolina) to develop a regression model. Using singular value decomposition (SVD), matrix X is decomposed into score matrix T and loading matrix P plus error matrix E , eq. (16). And matrix Y is decomposed into U and Q plus error term F , eq. (17).

$$X = T.P^T + E \quad (16)$$

$$Y = U.Q^T + F \quad (17)$$

where P and Q are orthogonal matrices. The PLS algorithm minimizes $\|F\|$ while maintain correlation between X and Y . The PLS algorithm computes the factor score matrix $T=XW$ for an appropriate weight matrix W . Once Q matrix is computed, the regression model is equivalent to $Y=XB+E$, where $B=WQ$, which can be used as a predictive regression model. More details about the PLS algorithm can be found in [12].

3 Results and discussion

The different numerical colour calibration models were tested using 64 available samples. 80% of the samples (51 samples) were used for training purposes while (20% = 17 samples) were used for validation. The accuracy of the four polynomial models (1st order, 2nd order, 3rd order), direct model and the neural network model were investigated. By comparing the calculated L^* , a^* and b^* with respect to the measured values at the same points, the results show that the 1st order model has higher fluctuations compared to other models. Moreover, the 3rd order model provided the highest regression accuracy, as presented in Table 1.

Table 1. The error in the $L^*a^*b^*$ Error calculations along 17 coloured samples using different polynomial regression models

| Model | Max. error |
|----------------------------------|------------|
| 1 st order | 30.29 |
| 2 nd Order + RGB Term | 7.11 |
| 3 rd Order | 2.90 |

The deviations between the measured $L^*a^*b^*$ data and the numerically calculated values obtained from the 3rd order PRM and the NNM is displayed in Figure 4.

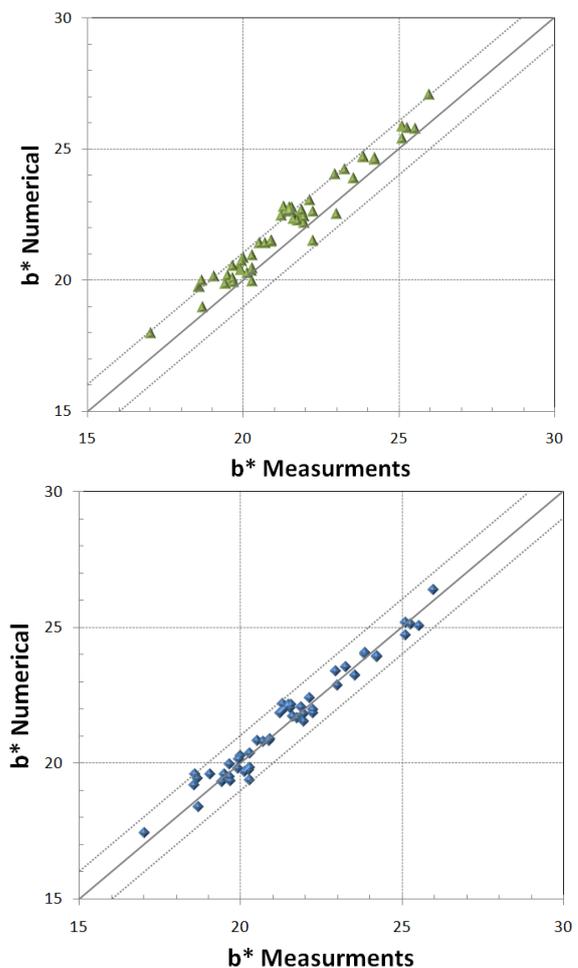


Figure 4. (a) the 3rd order polynomial regression model (b) the neural network model

Concerning the neural network model, a network of three neurons in both input and output layers was designed (3 input neurons for RGB values and 3 output

neurons for the $L^*a^*b^*$ values). A parametric study was performed to determine the number of neurons in the hidden layer. The best performance was achieved with 10 hidden layer neurons that had the lower mean square error (MSE) of 2.32×10^{-5} .

A data set of 50 semolina samples was divided into 80% (40 samples) for training the regression model and 20% (10 samples) for validating the model outputs. Figure 5 shows the comparison between the measured and predicted β -carotene content in 10 samples. The obtained results show that the predicted β -carotene values are highly correlated with the measured values with a correlation coefficient ($r=0.94$).

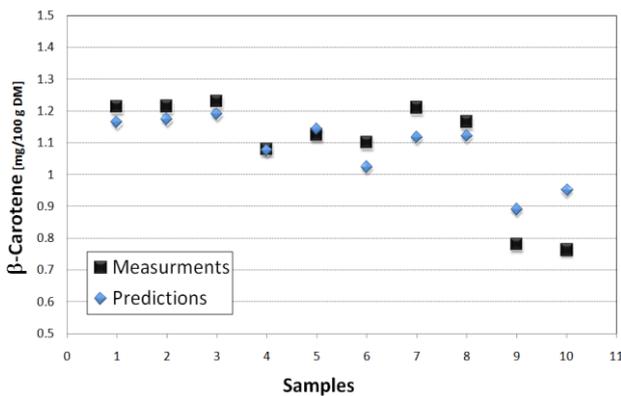


Figure 5. The measured and the predicted β -carotene content

4 Conclusion

The $L^*a^*b^*$ colour space was selected in this study as the standard device-independent colour space. The direct, polynomial regression and neural network models were developed to perform the RGB- $L^*a^*b^*$ transformation. Overall, the obtained results show that the lower error is achieved by the 3rd order polynomial model and the neural network models. Meanwhile, the PLS regression model for predicting the β -carotene shows a 0.94 correlation coefficient. The results indicate that the developed models are promising to present a consistent and reliable schemes for measuring $L^*a^*b^*$ and β -carotene as main quality measures of the semolina.

By increasing the order of the polynomial regression models enhanced the accuracy of the output results but also increased the computational time (from 2nd to 3rd order nearly one order of magnitude time increases).

Improving the imaging environment by using such a LED lamps will produces more stable environment without flickering as in the florescent lamps. The optimization of the lamps position will produce more uniform illumination over the tested sample and reduces the unrealistic variations of the RGB values induced by light which tends to more accurate $L^*a^*b^*$ predictions. Therefore, further parametric study will be performed to simulate the light distribution inside the acquisition setup

by solving numerically the differential equation of the light diffusional motion.

References

1. F. Mendoza, P. Dejmek and M. Aguilera: Calibrated color measurements of agricultural foods using image analysis. *Postharvest Biology and Technology* Volume **41**, Issue 3, (2006), Pages 285-295
2. K. León, D. Mery, F. Pedreschi: Color measurement in $L^*a^*b^*$ units from RGB digital images. *Food Research International* **39** (10): 1084-1091 (2006)
3. L. Kit, E. Spyridon: A simple digital imaging method for measuring and analyzing color of food surfaces. *Journal of Food Engineering*. Volume **61**, Issue 1, (2004), Pages 137-142
4. S. Basvaraj, C. Burkpallyi: Color Based Identification and Classification of Boiled Food Grain Images. *International Journal of Food Engineering*, Volume **5**, Issue 5 (2009) Article 7
5. A. Danti, M. Madgi, B. Anami: A Neural Network Based Recognition and Classification of Commonly Used Indian Non Leafy Vegetables. *International Journal of Image, Graphics and Signal Processing*. Vol. **6**, No. 10, pp. 62-68, (2014)
6. A. Danti, M. Madgi, S. Anami: Mean and Range Color Features Based Identification of Common Indian Leafy Vegetables. *International Journal of Signal processing, Image Processing and Pattern Recognition*, **5**, No. 3, September, (2012), pp. 151-160
7. Landi A.: Durum Wheat, semolina and pasta quality characteristics for an Italian food company. *CIHEAM-IAMZ*, (1995). p. 33-42
8. Y. Rharrabti, C. Royo, D. Villegas, N. Aparicio and L. F. Garcia del Moral: Durum wheat quality in Mediterranean environments: I. Quality expression under different zones, latitudes and water regimes across Spain. *Field Crops Research*, Volume **80**, Issue 2, 20, 123-131 (2003)
9. Hunt, R.W.G. (1991): *Measuring Color*. 2nd Edition. Ellis Horwood. New York
10. Alvin G. Wee, Delwin T. Lindsey, Shanglun Kuo, and William M. Johnston: Color accuracy of commercial digital cameras for use in dentistry. *Dent Mater.* (2006); **22**(6): 553-559.
11. V. Hentschel, K. Kranl, J. Hollmann, G. Meinolf: Spectrophotometric Determination of Yellow Pigment Content and Evaluation of Carotenoids by High-Performance Liquid Chromatography in Durum Wheat Grain. *J. Agric. Food Chem.* (2002), **50**, 6663-6668.
12. P. Geladi and B. R. Kowalski, "Partial Least Squares Regression: A Tutorial," *Analytica Chimica Acta*, vol. **185**, pp. 1-17, (1986)