Application of Near Infrared handheld spectrometers to predict semolina quality

Cristina Cecchini[a*], Francesca Antonucci[b], Corrado Costa[b], Alessandra Marti[c], Paolo Menesatti[b]

[a] Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA) - Centro di ricerca Ingegneria e Trasformazioni agroalimentari - Via Manziana 30, 00189 Roma, Italy

[b] Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA) - Centro di ricerca Ingegneria e Trasformazioni agroalimentari - Via della Pascolare 16, 00015 Monterotondo (Roma), Italy

[c] Department of Food Science, Environmental, and Nutritional Sciences, DeFENS, Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy

* Corresponding author: cristina.cecchini@crea.gov.it

ABSTRACT

Background: Durum wheat semolina is the best raw material for pasta production and its protein content and gluten strength are essential for the cooking quality. The need of finding rapid methods to speed up quality control makes Near Infrared spectroscopy (NIR) a useful method and widely accepted in cereal sector. In this study two non-destructive and rapid technologies, a low-cost sensor providing a short wavelength NIR range (swNIR: 700-1100 nm) and a handheld spectrometer providing a classical NIR range (cNIR: 1600-2400 nm), were employed to evaluate semolina quality parameters.

Results: Semolina samples were firstly characterized by the most used reference methods (protein content, Gluten Index, Alveograph® and Sedimentation test) and more recent one (GlutoPeak®). The spectra data were correlated with the chemical and rheological parameters. Partial Least Squares (PLS) model was used to compare the efficacy of swNIR or cNIR. The protein content is the reference parameter better correlated to the spectra data and showed the best regression model (r model = 0.9788 for cNIR and 0.9561 for swNIR). GlutoPeak indices also were well correlated with spectral data, particularly with swNIR spectra. Furthermore, the application of a provisional multivariate model (SIMCA) was used to classify quality of a semolina sample by means of its spectrum, obtaining a better modelling efficiency for swNIR.

Conclusion: The results have highlighted the applicability of pocket-sized low cost sensor (swNIR) easy to use directly to the sample source, compared to laboratory instruments or more expensive portable device.

Keywords: semolina, quality, Near Infrared spectroscopy, handheld devices

1. INTRODUCTION

The worldwide success of the Italian dried pasta is due to the use of durum wheat semolina as raw material, as well as to a tradition of pasta-making combined with years of research and experimentation. For pasta production, high protein content and gluten strenght, as indicator of its visco-elasticity, are essential to transform the semolina into a product able to guarantee an excellent cooking quality, expressed by low stickiness and bulkiness, and good firmness at optimal cooking time and/or overcooking (D'Egidio et al., 1993).

The performance of the raw material for pasta making is usually assessed through the total protein content and the rheological tests whose indices allow to predict the viscoelastic characteristics of gluten that are correlated to the pasta cooking quality (D'Egidio et al., 1990). Some of the methods used to assess raw material and pasta quality are well known standard procedures,while others are emerging and still in the phase of evaluation and comparison with the standards methods (Marti et al., 2014).

The needs of all actors in the supply chain always remain those of finding rapid methods to improve and speed up quality control at all stages of production. In this context, Near Infrared spectroscopy (NIR) is a rapid and non-destructive technique widely used in the agri-food sector (Cortés et al., 2019). In the case of durum wheat, NIR technique is accepted as a useful method to determine moisture and protein content (method EN 15948:2015). This technology has also been applied as predictive test to evaluate the quality (i.e., test weight, hardness, semolina yield and yellow pigment) of grain in early generations in breeding program (Sissons et al., 2006), to classify vitreous and nonvitreous kernels (Dowell, 2000), and to quantify the degree of adulteration of durum wheat flour with common bread wheat flour (Cocchi et al., 2006, Vermeulen et al., 2018). In the last decade, the applications of NIR have been focused to predict semolina technological quality (Sinelli et al., 2011, Firmani et al., 2020) and the technique has been proposed for in-line determination of moisture content in pasta immediately after the extrusion process (De Temmerman et al., 2007).

While defining the qualitative characteristics of a sample by NIR technique, calibration models are required to extract information from spectral data (Porep et al., 2015). Multivariate calibration techniques are often employed to relate the concentration of a certain analyte to the spectral data collected from that sample. Menesatti et al. (2014), for example, applied multivariate provisional soft independent modeling of class analogy (SIMCA) to distinguish between the use or not of organic wheat analyzed by a rapid and non-destructive method based on hyperspectral imaging. In addition, Partial Least Squares (PLS) model was used to compare the efficacy of NIR vs. mid-infrared (MIR) to determinate the nutritional properties in wheat bran samples (Hell et al., 2016).

Recently, to perform a direct and rapid detection, various manual NIR devices have been developed that have already found application in the food industry (Ayvaz et al., 2015) and in the cereal sector to control the sprouting process of wheat (Grassi et al., 2018). The portable NIR analyzers allow the instrument to be taken directly to the sample source, eliminating the time and protocols required to move samples to the lab. Taking into consideration that, to the best of our knowledge, no studies have been carried out on durum wheat semolina, in this study, two non-destructive and rapid technologies, a low-cost sensor providing a short wavelength NIR range (swNIR) and a handheld spectrometer providing a classical NIR range (cNIR), were employed to evaluate some semolina quality parameters. In addition, to determine the correspondence between the spectral data and some reference quality variables different multivariate statistical analyses were performed.

2. MATERIAL AND METHODS

2.1. Materials

The study was performed on 64 durum wheat varieties obtained from experimental trials of the Italian network realized during the growing season 2016/2017. The samples were representative of three different agro-climatic areas: Po valley (11 samples); Adriatic coast (26 samples) and Sicilian insular (27 samples).
All durum wheat grains were conditioned to 17% for about 20 h and milled by pilot milling plant Buhler MLU 202 (Bühler, Switzerland). Then semolina was passed twice to the purifier (Namad, Italy) for further refinement. The semolina obtained from each sample has an ash content between 0.80 and 0.90% d.b. maximum limit defined by Italian legislation for the production and marketing of durum wheat semolina pasta (Italian law 580/67 and subsequent amendments).

2.2. Methods

2.2.1. Reference quality tests on semolina

Firstly, semolina samples were characterized by means of standard methods. Protein content was determined by Dumas combustion method (ICC method n. 167) with automatic instrument Leco FP 528 (Leco Corp., USA). The conversion factor used was N x 5.7. Gluten content was determined according to EN ISO 21415 method and Gluten Index by ICC method n. 158 using Glutomatic System (Perten, Sweden). The alveograph test (Chopin Co., France) was conducted according to UNI 10453 method for durum wheat semolina.
Gluten quality was also evaluated by nonconventional test, such as GlutoPeak devices (Brabender GmbH and Co., Germany). GlutoPeak test was performed according to Marti et al. (2014), with some modifications. In particular, 9 g of semolina and 9 g of distilled water were used, adjusting the quantity of semolina to 14% humidity. The speed of the rotating element was set at 2750 rpm while the temperature at 36 °C. The main indices considered, automatically evaluated by the software, were i) Maximum consistency (BEM) (expressed in GlutoPeak Units, GPU), corresponding to the peak occurring as gluten aggregation; ii) Total energy equivalent to the area under the peak (from 0 to 15 s after the maximum peak) expressed in GlutoPeak Equivalents (GPE).
All the samples were also characterized by the sedimentation test in Sodium-Dodecyl-Sulphate (SDS test, ICC method No. 151) carried out on whole wheat flour, obtained by grinding with a Cyclotec mill (FOSS AB Analytical, Sweden) equipped with a 1 mm sieve.

2.2.2. NIR spectroscopy

Semolina samples were analyzed using a NIR handheld spectrometer with a short wavelength range (swNIR) and one with a classic NIR range (cNIR).
For both NIR analysis, semolina was placed in a plastic capsule and covered by a low reflectance glass plate. The measurements were carried out at three different points and repeated for three more fills, obtaining 9 data per sample.
The short-wavelength spectra were recorded by SCiO (ConsumerPhysics Inc®), a pocket-sized device, with a reflectance range of 700-1100 nm. The spectral data were transferred to a smartphone via Bluetooth wireless technology and recorded in the cloud. The data in a CSV format were transferred to an Excel spreadsheet for analysis.
The classic-wavelength spectra were collected by MicroPHAZIR RX analyzer (Thermo fisher scientific®), a handheld NIR instrument for on-site material identification, with a spectral range of

1600-2400 nm. Spectral data were transferred to a PC via a USB cable in a TXT file and transferred to an Excel spreadsheet for analysis.

2.2.3. Statistical analyses

Partial Least Square Regression (PLS)

The results obtained on the semolina samples were further processed through a multivariate regression using the PLS method to observe the predictive capacity of spectral data matrices (swNIR or cNIR; X-blocks). The predicted reference quality variables (Y-block) were: Protein content, Gluten content, Gluten Index, Sedimentation value, Alveograph parameters (W and P/L) and GlutoPeak parametrs (BEM and Total Energy). The PLS procedure (Wold et al 2001) was elaborated using the PLS Toolbox in MATLAB V7.0 R14 (The Math Works, Natick, MA, USA) and included the following steps: 1) extraction of raw spectra dataset, (X-block variables); 2) creation of measured values dataset to be used as reference or response variable (Y variable); 3) data fusion of the two dataset (Y and X-block) in one analysis dataset (ADs); 4) analysis dataset partitioning into model set (MS=80% of ADs) and external validation test set (TS=20% of ADs) by means of sample set partitioning based on joint x-y distances (SPXY) algorithm (Harrop Galvao et al, 2005). This method employs a partitioning algorithm that takes into account the variability in both x- and y-spaces; 5) application of different pre-processing algorithms to X-block and Y (none, Log 1/R, diff1, mean centre, autoscale, median centre, baseline) - the matrices were pre-processed using the autoscale Matlab algorithm; 6) application of chemometric technique: modelling and testing; 7) calculation of efficiency parameter of prediction.

The performances of the model were estimated by evaluating the coefficient of correlation (r) between observed and predicted values, Standard Error of Prevision (SEP), Root-Mean-Square Error of Calibration (RMSEC) and bias calculated as the average of the differences between predicted and measured. Residual Predictive Deviation (RPD), defined as the ratio of the standard deviation of the laboratory measured (reference) data to the RMSE (Williams, 1987), was used to verify the accuracy of the model. RPD values between 2.0 and 2.5 indicate very good, quantitative model and/or predictions; RPD values major than 2.5 indicate excellent model and/or predictions (Viscarra Rossel et a., 2007; Febbi et al., 2015).

The model accuracy and precision were evaluated according to the highest r, minimum SEP, maximum RPD and bias value very close to zero.

Soft Independent Modeling of Class Analogy (SIMCA)

A different processing approach was applied to evaluate the possibility to find a model able to perform a classification of semolina based on the NIR spectra. About that the 64 semolina samples were also classified using the quality ranges of the technological parameters reported in the UNI method for classification of semolina for pasta making (UNI 10940: 2001). The UNI method includes 3 quality grades (A, B, C) for the following parameters: protein content, gluten content, gluten index, alveographic parameters W and P/L. In this work the samples were grouped into three classes according to the scheme shown in Table 1.

Table 1

In order to search for an optimal classification model for semolina quality (as reported in Table 1) a SIMCA (Wold and Sjostrom, 1977) was applied. Two models, one for each spectral data (swNIR or cNIR), were built (single class modelling approach; Forina et al., 2008). SIMCA, computed with the software V-Parvus 2010, is a collection of Principal Component Analysis (PCA) models [Nonlinear

Iterative vartial Least Squares (NIPALS) algorithm], one for each class of dataset (one in this case), after a separate category autoscaling. SIMCA cross validates the PCA model of each class (training set), splitting the data (evaluation set) into four contiguous groups (cross validation groups). In this case, the modified model with expanded range was used substituting the one first introduced by Wold and Sjöström (1977). The unweighted augmented SIMCA distance was considered in building the models. For each class, the number of significant components of the inner space was estimated considering four Principal Components (PC) (lowest noise found). For each class, a critical square distance based on the F-distribution was calculated using a confidence interval (95%). The class boundary was determined according to the confidence interval. An observation is attributed to the model class when its residual distance from the model has a value below the statistical limit for the class. SIMCA allows both the modelling and classification analysis. In the classification phase, all the observations should be attributed to one of the pre-defined classes. The efficiency was evaluated by classification (training set) and prediction (evaluation set) matrices, which reported the percentage of correct classification for each considered class. SIMCA also expressed the statistical parameters indicating the modelling efficiency. Unknown objects could be either classified into the class or recognized as outliers. The modelling efficiency was indicated by sensitivity. This is the measure of how well the model correctly identifies the cases really belonging to the class. The modelling power for each variable, which represents the influence of that variable in defining of the model, was expressed. In order to express a metric index for semolina quality based on spectral reflectance data, square SIMCA distances were linearized converting the values into a logarithmic scale and then translating them by adding a certain value in order to have all positive values. To avoid overfitting, only 8 out of 10 best samples (Table 1) were used to construct and cross-validate each SIMCA model. The remaining 2 samples together with all the other classes samples has been used to test the performance of each SIMCA models. The partitioning of the artificial datasets is optimally chosen with Euclidean distances, based on the Kennard and Stone (1969) algorithm that selects objects without a priori knowledge of a regression model (*i.e.*, the hypothesis is that a flat distribution of the data is preferable for a regression model).

3. RESULTS AND DISCUSSION

3.1. Reference quality tests on semolina

Table 2 showed the results in terms of average, standard deviation, minimum and maximum value, obtained with the reference quality tests on semolina samples. The methods used express different aspects of the characteristics of the raw material, specifically of gluten, and all together they contribute to providing a broader qualitative evaluation.

The samples considered in this study cover a wide variability range for each parameter, above all for those related to the protein content and gluten quality based on which semolina is generally classified for pasta making (UNI method 10940). In this study the sample variability is important to allow a better comparison between different analysis approach and to be able to evaluate and predict semolina properties.

Table 2

3.2. NIR spectroscopy

The PLS regression was performed to make a quantitative prediction and to find the best relationship between the set of reference variables and the set of spectral data. The results of the models obtained for swNIR and cNIR are reported in Table 3. For the variables not shown, the models reported low performance in regressing quality variables.

Table 3

Generally, a good predictive model should have high values of r and low values for RMSEC and low SEP (Liu et al. 2014). According to these considerations, the protein content is the reference parameter better correlated to the spectra data. In particular, the best regression model (*r* model = 0.9788) was obtained with the cNIR spectra, but good correlation (*r* model = 0.9561) also occurred with swNIR spectra. In addition, the RMSEC was very low for both models (swNIR = 0.2903 and cNIR = 0.2028) as well as the SEP value (swNIR = 0.4899 and cNIR = 0.3263). The model robustness for protein content was validated by RPDtest, precisely swNIR = 2.4036 which indicate very good, quantitative model and/or predictions and cNIR = 3.9405 denote excellent model and/or predictions. Moreover, the systematic error in the predictive values (bias) of these models were also very small (swNIR = -0.0006 and cNIR = 0.0004). These results confirm the applicability of the NIR technique for measuring the protein content as widely reported in the literature (Sinelli et al., 2011, Dowell et al., 2006, Delwiche and Hruschka, 2000).

As for the other reference parameters (particularly SDS test and Alveographic parameters), the models obtained are not very good, the regression models are not enough high as well as the RPD values. A better correlation was found between the swNIR spectra and the Maximum consistency (BEM) and Total energy (TE) (*r* model 0.9245 and 0.9390 respectively) obtained by the GlutoPeak test. The NIR prediction of the qualitative parameters, based on rheology or viscosity measures, can be traced to the relationship between physical properties and chemical constituents (proteins, starch contained in water etc) (William, 2007). The GlutoPeak test measures the aggregation kinetics of gluten proteins. It has been showed that flours with similar protein content can show different gluten aggregation profiles which comes from the way gluten proteins interact forming the gluten network. In winter wheat varieties, a correlation between maximum torque and gliadin content was found, whereas the area under the entire GlutoPeak profile was correlated to the amount of glutenins and to the insoluble fractions of the glutenins (Marti et al., 2015).

The robustness of the models could be influenced by the fact that in this preliminary study the data set used is not particularly large.

The performance of the models was represented graphically with the scatter plot in which the estimated variable is a function of the measured variable (Figure 1). In the case of perfect regression, the points relative to the samples used as tests should be placed along the bisector. The graphics confirmed the very good model performance for protein and good for GlutoPeak parameters.

Figure 1.

SIMCA was instead applied to spectral data (swNIR or cNIR) to find a classification model for semolina quality. Semolina samples were grouped into three classes: Best, Good and Sufficient basing on criteria reported in Table 1. Through the application of the SIMCA model the semolina samples were classified according to the collected spectrum. The models have been developed based on best samples, the rest (good and enough samples) have been used as an external test to verify the goodness of the quality metric scale obtained (Forina et al., 2008). The swNIR SIMCA model, shown in Figure 2A, presented a square critical distance equal to 1.62 indicating that a semolina sample with a SIMCA distance lower than the critical distance (*i.e.*, 95% confidence interval) was considered having a best quality by the model. The modelling efficiency, indicated by sensitivity value, was equal to 70% (3 best observations out of 10 outside the model). The samples belonging to the Good class were much closer than those Sufficient to the model and 2 Good samples were included inside the model (values lower than the critical distance). This was also highlighted by the average values for the normal distributions: Best = 1.35, Good = 4.20, Sufficient = 5.60. This result underlined that the obtained SIMCA model based on swNIR spectra data was efficient to identify semolina quality; the translated log squared SIMCA distance was a good metric indicator for semolina quality. It must be underlined that good and enough samples were not included in the model construction.

Figure 2B reported the same approach but based on cNIR spectral data. The obtained model returned a square critical distance equal to 2.27. The percentage of sensitivity was higher than the swNIR one and equal to 60% (4 best objects out of 10 outside the model). The distance between the average Good and Sufficient observed samples was not well outlined. In fact, the values of the averages of normal distributions were close to each other and inverted (Best = 2.00, Good = 3.20, Sufficient = 3.00). The obtained model base on cNIR spectral data resulted not able to obtain a metric indicator for semolina quality.

Figure 2.

4. CONCLUSIONS

As a whole, this study showed the possibility of using handheld NIR spectrometers to predict some chemical and rheological characteristics of semolina samples. The results, combined with multivariate statistical analyses, confirmed the use of NIR technology to evaluate protein content, a fundamental parameter to define the commercial class of semolina. The calibration models resulted to be good with high accuracy ($r = 0.9561$, SEP = 0.4899 for swNIR and $r = 0.9788$, SEP = 0.3263 for cNIR). Furthermore, the application of a provisional multivariate model (SIMCA) appeared to be efficient in distinguishing the class quality of a semolina sample by means of its spectrum. In particular, the results showed a better performance of a short wavelength NIR sensor (swNIR), also obtaining an application model based on an immediately applicable metric indicator. Although the application on these devices required optimization of model robustness, the preliminary results highlighted the applicability of short wavelength tool for a commercial characterization (protein content) of semolina in very short time. The innovation and advantage of swNIR device were due to a pocket-sized low cost sensor, to be taken directly to the sample source and ready for use, compared to laboratory instruments or more expensive portable device. Further applications of these devices on final product will be developed in future.

REFERENCE

1. D'Egidio MG, Mariani BM, Novaro P, Influence of raw material characteristics and drying technologies on pasta cooking quality: a review of our results. *Italian Food & Beverage Technology* **1**: 29-32 (1993).
2. D'Egidio MG, Mariani BM, Nardi S, Novaro P, Cubadda R, Chemical and technological variables and their relationships: a predictive equation for pasta cooking quality. *Cereal Chemistry* **38**: 67, 275–281(1990).

3. Marti A, Cecchini C, D'Egidio MG, Dreisoerner J, Pagani MA, Characterization of durum wheat semolina by means of a rapid shear-based method. *Cereal Chemistry* **91:** 542–547 (2014). https://doi.org/10.1094/CCHEM-10-13-0224-R
4. Cortés V, Blasco J, Aleixos N, Cubero S, Talensa P, Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. *Trends in food science & technology* **85**: 138-148 (2019).
5. Sissons M, Osborne B, Sissons S, Application of near infrared reflectance spectroscopy to a durum wheat breeding programme. *Journal of Near Infrared Spectroscopy* **14:** 17-25 (2006).
6. Dowell FE, Differentiating vitreous and non-vitreous durum wheat kernels by using near-infrared spectroscopy. *Cereal Chemistry* **77:** 155-158 (2000).
7. Cocchi M, Durante C, Foca G, Marchetti A, Tassi L, Ulrici A, Durum wheat adulteration detection by NIR spectroscopy multivariate calibration. *Talanta* **68**: 1505-1511 (2006).
8. Vermeulen P, Suman M, Pierna JAF, Baeten V, Discrimination between durum and common wheat kernels using near infrared hyperspectral imaging. *Journal of Cereal Science* **84**: 74-82 (2018).
9. Sinelli N, Pagani MA, Lucisano M, D'Egidio MG, Mariotti M, Prediction of semolina technological quality by FT-NIR spectroscopy. *Journal of Cereal Science* **54:** 218-223 (2011).
10. Firmani P, Nardecchia A, Nocente F, Gazza L, Marini F, Biancolillo A, Multi-block classification of Italian semolina based on Near Infrared Spectroscopy (NIR) analysis and alveographic indices. *Food Chemistry* **309**: 125677 (2020).
11. De Temmerman J, Saeys W, Nicolaï B, Ramon H, Near infrared reflectance spectroscopy as a tool for the in-line determination of the moisture concentration in extruded semolina pasta. *Biosystems Engineering* **97:** 313-321 (2007).
12. Porep JU, Kammerer DR, Carle R, On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science & Technology* **46:** 211–230 (2015).
13. Menesatti P, Antonucci F, Pallottino F, Bucarelli FM, Costa C, Spectrophotometric qualification of Italian pasta produced by traditional or industrial production parameters. *Food and Bioprocess Technology* **7:** 1364-1370 (2014).
14. Hell J, Prückler M, Danner L, Henniges U, Apprich S, Rosenau T, Kneifel W, Böhmdorfer S, A comparison between near-infrared (NIR) and mid-infrared (ATR-FTIR) spectroscopy for the multivariate determination of compositional properties in wheat bran samples. *Food control* **60:** 365-369 (2016).
15. Ayvaz H, Rodriguez-Saona LE, Application of handheld and portable spectrometers for screening acrylamide content in commercial potato chips. *Food Chem.* **174**: 154–162 (2015).
16. Grassi S, Cardone G, Bigagnoli D, Marti A, Monitoring the sprouting process of wheat by non-conventional approaches. *Journal of Cereal Science* **83:** 180-187 (2018).
17. Wold S, Sjostrom M, Erikssonn L, PLS-regression: A basic tool of chemometrics. Chemometr, *Intell. Lab. Syst.* **58:** 109-130 (2001).
18. Harrop Galvao RK, Ugulino Araujo MC, Emidio Jose G, Coelho Pontes MJ, Cirino Silva E, Bezerra Saldanha TC, A method for calibration and validation subset partitioning. *Talanta* **67:** 736-740 (2005).
19. Williams P, Variables affecting near-infrared reflectance spectroscopic analysis, In: *Near-infrared technology in the agricultural and food industries*, ed. Williams P, Norris K. St Paul, Minnesota: American Association of Cereal Chemists. pp. 143-166 (1987).

20. Viscarra Rossel RA, Taylor HJ, McBratney AB, Multivariate calibration of hyperspectral gamma-ray energy spectra for proximal soil sensing. *Eur J Soil Sci* **58:** 343-353 (2007).
21. Febbi P, Menesatti P, Costa C, Pari L, Cecchini M, Automated determination of poplar chip size distribution based on combined image and multivariate analyses. *Biomass & Bioenergy* **73:** 1-10 (2015).
22. Wold S and Sjostrom M, SIMCA: A method for analyzing chemical data in terms of similarity and analogy, In *Chemometrics: Theory and Application*, ed. Kowalski BR. American Chemical Society Symposium Series 52, Wash., D.C., pp. 243-282 (1977).
23. Forina M, Oliveri P, Lanteri S, Casale M, Class-modeling techniques, classic and new, for old and new problems. *Chemometrics and Intelligent Laboratory Systems* **93**: 132-148 (2008).
24. Kennard RW, Stone LA, Computer aided design of experiments. *Technometrics* **11:** 137–148 (1969).
25. Lin C, Chen X, Jian L, Shi C, Jin X, Zhang G, Determination of grain protein content by near-infrared spectrometry and multivariate calibration in barley. *Food Chemistry* **162**: 10-15 (2014)
26. Delwiche SR, Hruschka WR, Protein content of bulk wheat from near-infrared reflectance of individual kernels. *Cereal Chemistry* **77:** 86-88 (2000).
27. Williams P, Grains and seeds, In: *Near-Infrared Spectroscopy in Food Science and Technology*, ed. Ozaki Y, McClure WF, Christy AA. John Wiley & Sons, Inc., pp. 165-217, (2007). ISBN: 978-0-471-67201-2, Hoboken.
28. Marti A, Augst E, Cox S, Koehler P, Correlations between gluten aggregation properties defined by the GlutoPeak test and content of quality-related protein fractions of winter wheat flour. *Journal of Cereal Science* **66:** 89-95 (2015).

Table 1. Quality classification of semolina samples (n = 64)

Table 2. Quality characteristics of semolina samples (n = 64)

Table 3. Characteristics and principal results of the Partial Least Squares (PLS) regression models in estimating the principal reference quality variables from spectral data (swNIR or cNIR).
In particular: LVs = Latent Vectors; SEP = Standard Error of Prevision; RMSEC = Root-Mean-Square Error of Calibration; RPD = Residual Predictive Deviation.

Figure 1 – Partial Least Squares (PLS) scatter plots of the observed versus predicted principal reference quality variables from spectral data (swNIR or cNIR) for both validation (80%) and test (20%) datasets. Note: Line represented the bisectrix (*i.e.*, perfect attribution). Black circles indicated the model set samples, meanwhile white circles the test set samples.

Figure 2 – Soft Independent Modeling of Class Analogy (SIMCA) histogram by frequency class of the translated log squared values for A) swNIR and B) cNIR datasets built on 10 best samples of semolina. The three qualitative classes (reported in Table 1) were plotted with different colors. The dashed line represented the critical value (*i.e.*, model boundary).