# DEVELOPMENT OF A CALIBRATION MODEL TO ESTIMATE QUALITY TRAITS IN WHEAT FLOUR USING NIR (NEAR INFRARED REFLECTANCE) SPECTROSCOPY

**F. KAHRIMAN, C. O. EGESEL**

***Çanakkale Onsekiz Mart University, Faculty of Agriculture, Field Crops Department***
***Terzioğlu Campus, 17100, Çanakkale, TURKEY***
***E-mail: fkahriman@hotmail.com***

***Abstract:*** *Introduction of fast, cheap and reliable methods in the analysis of agricultural products is of great importance for the breeding programs targeting to enhance quality traits. The objective of this study was to develop a NIRs (Near Infrared Reflectance Spectroscopy) calibration to enable a fast evaluation of quality traits in wheat flour. We used 120 flour samples from 40 bread wheat varieties as plant material. We developed calibrations based on Partial Least Squares (PLS) regression model for moisture, protein ratio, gluten ratio, gluten index value, sedimentation value, and modified sedimentation value. In the generated calibrations, it was found that the correlation coefficients between the reference analyses and NIR readings were higher than 0.90 for all variables (r=0.92 for moisture ratio, r=0.90 for protein ratio, r=0.94 for gluten ratio, r=0.93 for sedimentation value and r=0.92 for modified sedimentation value) except gluten index value(r=0.87). It was determined that NIR estimates of moisture, protein and gluten were in acceptable levels when the standard error of estimate (SEE), root mean square error of prediction (RMSEP) and Root-Mean-Square Error of Cross-Validation (RMSECV) in the developed calibrations were taken into account. Relative deviation prediction (RPD) values indicated that calibration model for gluten ratio had the potential to be used for screening (RPD=3.25). The other parameters (RPD values between 2 and 3) except gluten index value (RPD=1.90) had rough screening potential. Regression coefficients ($R^2$) of calibration parameters were above 0.80, except for gluten index value. Furthermore, as a result of the correlation analysis performed to determine the relationships of the scanned interval (1200-2400 nm) with the investigated traits, it was found that the variables except gluten index value and moisture level had high correlation values in the similar absorption regions. Ten wavelength regions had similar direction of correlation coefficient between the 1200-2400 nm for these traits. General results of this study suggest that the calibration models developed here could successfully estimate most of the wheat flour quality traits, whereas better models are needed for more robust estimations on some other traits, such as gluten index value.*

***Key words*****:** *Protein*; *Gluten; Triticum aestivum*

## INTRODUCTION

Quality evaluations of bread wheat varieties and their products have important effect on both the quality of improved wheat varieties and wheat products in scientific and industrial areas. However, currently used tests for determination of the quality include time consuming, laborious and expensive methods. Therefore, there is a need for faster, easier and cheaper techniques. NIR technology allows the use of such advantages in analyzing many different traits of plant material in breeding studies as well as food industry.

NIRs (Near Infrared Reflectance Spectroscopy) has improved since 1960's and it has been used for quite a few purposes in agriculture, medicine and industrial areas (PASQUINI, 2003). Analysis of biological materials with NIRs is based on the differences of C-H, O-H, N-H and S-H functional groups in their structure (SANDORFY ET AL. 2007). In order to estimate the value of a trait by means of NIR spectra, a "calibration" is used. Calibration is a mathematical model developed to make NIR estimations by relating measured NIR spectra to reference analysis results. Currently, NIR calibrations are generated by different statistical methods such as Multiple Linear Regression (MLR), Partial Least Square Regression (PLSR), Principal Component Analysis (PCA) (LU ET AL. 2000). The most important advantage of using these techniques in generating NIR calibrations is the quick attainment of the results about several variables on the same sample. Limited variation in the data set used for calibration, and lack of precision of the reference analysis are the chief disadvantages.

Scientific literature on the NIR calibrations used to estimate wheat quality traits focused on the development of “quantitative” type of calibrations. HRUSKOVA AND FAMERA (2003) generated a NIR calibration to estimate moisture, protein, ash, and wet gluten ratio in wheat flour; and they reported that this technique could be effectively used. DOWELL ET AL (2006) generated calibrations with different NIR instruments on 186 different grain, flour, dough, and bread quality traits using 198 bread wheat samples. In the result of this study, they reported that NIR technique could be used in process control for protein and moisture ratio and flour color. In a different study, calibrations were developed for 7 different quality traits using the flour of 228 different bread wheat varieties, and it was concluded that protein ratio and Zeleny sedimentation value could be efficiently predicted by NIR technique (JIRSA ET AL. 2008). Literature contains more studies targeting the development of calibrations to estimate different quality variables in other crops, such as rapeseed, maize, common bean, and safflower (VELASCO AND BECKER, 1998; SPIELBAUER ET AL. 2009; HACISALIHOGLI ET AL. 2010; ELFADL ET AL. 2010).

The intensity of the literature on the NIR techniques indicates that a great deal of effort has been spent to integrate these techniques with the applied scientific areas. The objective of this study was to generate a PLSR calibration for the use of routine quality analyses in bread wheat variety-yield testing trials. We also intended to relate the individual spectral areas with calibration parameters.

## MATERIAL AND METHODS

*Wheat Samples and Laboratory Analyses*

A total of 120 wheat samples belonging to 40 different bread wheat genotypes were used for calibration development. Approximately two hundred grams grain samples of each genotype were milled with laboratory mill (Perten 3100, Sweeden) with a 0.1 mm sieve mesh for laboratory analysis and NIR measurements. Prior to the laboratory analyses, NIR spectra between 1200-2400 nm were obtained and saved for calibration development. Moisture contents of the samples were determined by oven drying method. Protein ratios were calculated by use of Kjeldahl method (ANNONYMOUS, 1980). The protein quality tests, namely, gluten ratio (ANNONYMOUS, 1994a), gluten index value (ANNONYMOUS, 1994b), sedimentation value (ANNONYMOUS, 1994c) and modified sedimentation (ATLI ET AL. 1988) value were done utilizing internationally accepted standard methods in the scientific literature.

*Data Pretreatment and Calibration Development*

Prior to the chemical analyses of the samples, NIR spectra were measured and saved within 1200-2400 nm range with 1 nm interval in Spectrastar 2400D NIR instrument. Twenty four spectra were taken for each wheat sample and saved in DEFAULT file in the instrument. Afterwards, these spectra were combined into a single file using Infostar data converter program. The combined spectrum file was transferred into Sensologic Calibration Workshop (CWS) software and spectra were transformed to absorbance (1/logR), First Derivative (Segment:4, Gap:3) and Standard Normal Variate (Figure 1).

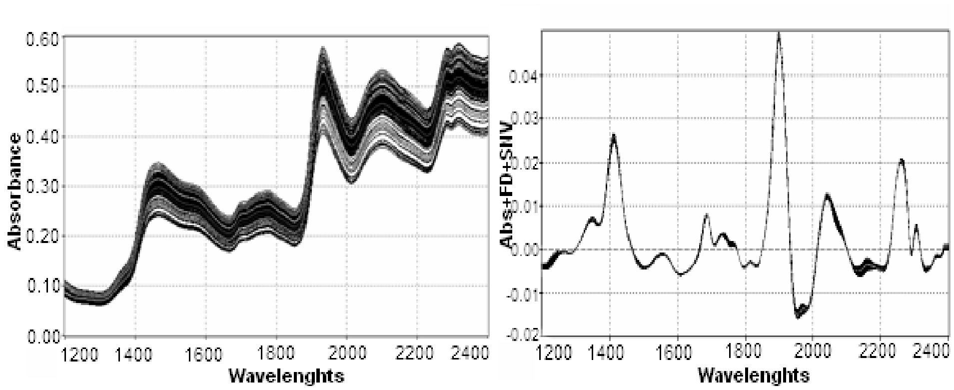


Figure 1. Absorbance values obtained from calibration set and their status after first derivative and SNV transformation

The individual evaluation parameters for PLSR calibration robustness, namely, RMSEP, SEE (CWS MANUAL, 2003) and RPD (BAILLERES ET AL. 2002) were computed by the following equations (Equation 1, 2, 3, 4).

$$RMSEP = \sqrt{\frac{\sum (Ypred - Yref)^2}{n}} \qquad (1)$$

$$Bias = \bar{Y}pred - \bar{Y}ref \qquad (2)$$

$$SEE = \sqrt{\frac{N}{N-1}(RMSEP^2 - Bias^2)} \qquad (3)$$

$$RPD = \frac{STD_{ref}}{SE_{pred}} \qquad (4)$$

where; *RMSEP*: root mean square error of prediction, *Ypred*: predicted value, *Yref*: reference value with standard analysis, *n*: number of observations, *Bias*: total differences between predicted and reference values, *STDref*: standard deviation of reference analysis, and *SEpred*: standard deviation of prediction, *RPD*: relative prediction deviation. Due to the lack of other statistics (RPD, Min-Max, Mean values of reference analysis) in the CWS software, they were computed in SAS V8 statistical program (SAS INST., 1999).

## RESULTS AND DISCUSSIONS

*Evaluation of NIR calibrations*

Descriptive statistics of reference analysis methods were summarized in Table 1. It seems that there was enough variation for calibration development for all variables. Gluten ratio, gluten index value, sedimentation and modified sedimentation values showed relatively higher variation. Also, standard deviations of gluten index, sedimentation and modified sedimentation values were higher than the other calibration parameters. The evaluation parameters of PLSR calibration in this research were shown in Table 2.

*Table 1.*

Descriptive statistics for the parameters used in calibration development

| Calibration Parameter | Mean | Standard deviation | Min- Max |
|---|---|---|---|
| Moisture ratio (%) | 12.1 | 0.73 | 11.3-14.4 |
| Protein ratio (%) | 10.7 | 1.30 | 8.4-13.6 |
| Gluten ratio (%) | 35.2 | 4.71 | 19.9-46.9 |
| Gluten index value (%) | 75.4 | 14.9 | 43.3-99.4 |
| Sedimentation value (ml) | 54.3 | 10.9 | 28.0-71.0 |
| M. Sedimentation value (ml) | 54.1 | 11.1 | 28.0-71.0 |

*Table 2.*

Evaluation parameters for the generated calibration models

| Calibration Parameter | FS | SEE | RMSECV | RMSEP | $R^2$ | RPD |
|---|---|---|---|---|---|---|
| Moisture ratio (%) | 11 | 0.30 | 0.47 | 0.27 | 0.85 | 2.43 |
| Protein ratio (%) | 5 | 0.60 | 0.68 | 0.58 | 0.81 | 2.17 |
| Gluten ratio (%) | 1 | 1.45 | 1.50 | 1.59 | 0.88 | 3.25 |
| Gluten index value (%) | 10 | 7.84 | 11.6 | 7.43 | 0.76 | 1.90 |
| Sedimentation value (ml) | 6 | 4.27 | 4.88 | 4.11 | 0.87 | 2.55 |
| M. Sedimentation value (ml) | 6 | 4.52 | 5.13 | 4.35 | 0.84 | 2.46 |

SEE: Standard error of prediciton, FS: Factor number in regression equation, RMSEP: Root mean square error of prediciton, RMSECV: Root mean square error of cross validaiton, $R^2$: Regression coefficient, RPD: Relative deviation prediction.

The lowest correlation between NIR prediction and reference analysis was determined in gluten index value (r=0.87). Other variables had correlation coefficients of at least 0.90 (Figure 2). A NIR estimation could be considered "robust" as long as the correlation coefficient between the results of NIR readings and the reference method is above 0.75 (CWS MANUAL, 2003). In this point of view, we can say that there was an acceptable relation for all calibration parameters between observed and predicted values. Regression coefficient ($R^2$) is another evaluation parameter of calibrations. ELFADL ET AL. (2010) reported that this value should be at least 0.83. Except gluten index value and protein ratio, all variables had $R^2$ values above this limit (Table 2). Another important evaluation parameter is standard error of estimation, which is desired to be as small as possible (CWS MANUAL, 2003). Protein ratio (0.60%), moisture ratio (0.30%), and gluten ratio (1.54%) showed desirable values for standard error of estimate. However, standard errors of estimates were relatively higher in gluten index value (7.79%), sedimentation value (4.27 ml) and modified sedimentation value (4.52 ml) (Table 2).

In a similar manner; protein, moisture and gluten ratios yielded acceptable values for root mean square error of cross validation (RMSECV) and root mean square error of prediction (RMSEP) and regression coefficient ($R^2$), while those values were relatively higher in gluten index, sedimentation and modified sedimentation (Table 2). There were similar results to our findings for sedimentation and gluten index value in the literature (DOWELL ET AL 2006). Modified sedimentation has similar results as sedimentation and gluten index value but there is a lack of scientific literature on NIR calibration development about this trait. In addition protein and gluten ratio have positive correlation because of gluten is main protein in wheat flour. Therefore, NIR calibration gives similar results for gluten and protein ratio in out study as the other scientific literature. Another important evaluation parameter is the relative performance in prediction (RPD).

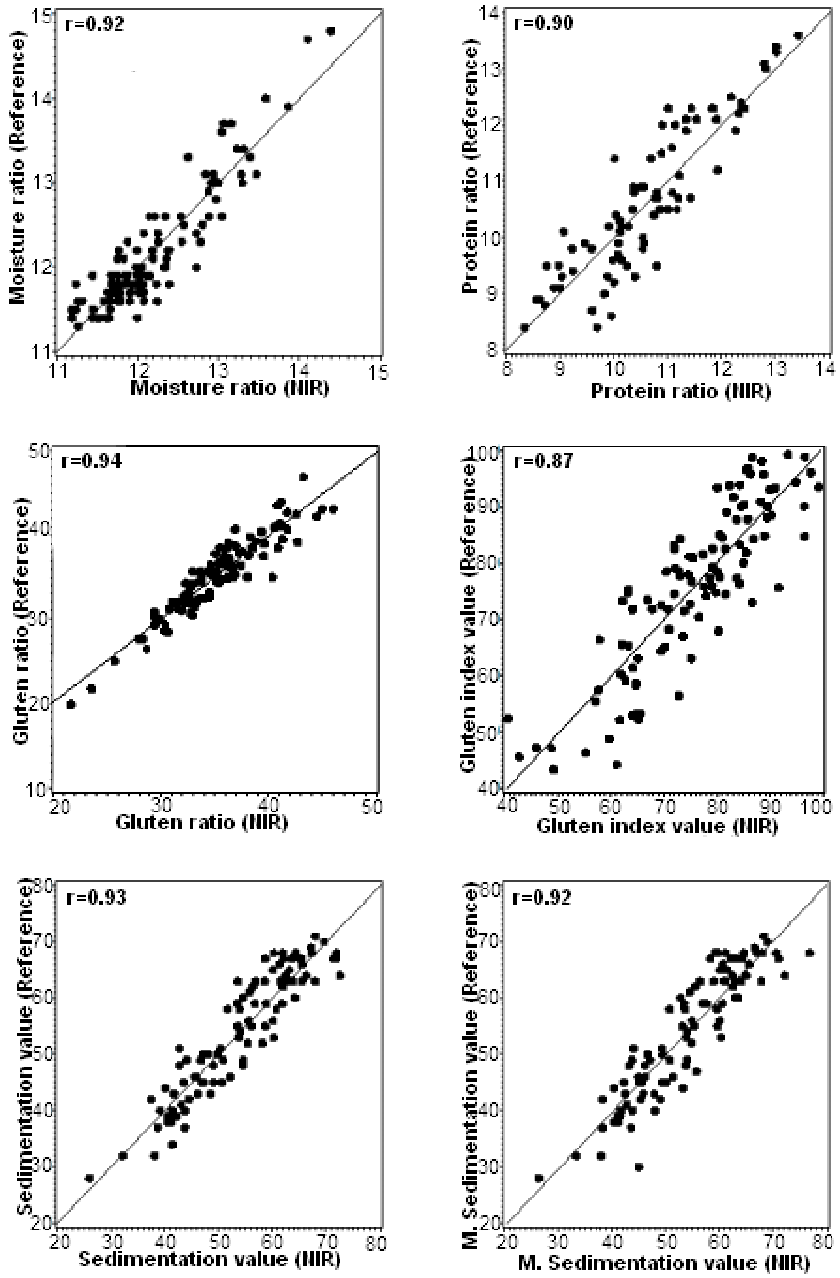


Figure 2. The relationships between NIR predictions and reference methods in the generated calibration model

It was reported that generated calibration model is robust if the RPD value is between two and three it is suitable for rough screening, between three and five; model has screening potential, five and eight; model could be used in quality control analysis, above the eight; model is suitable for analytical application (BAILLERES ET AL. 2002).

In this study, the lowest RPD value (1.90) was determined for gluten index, while the RPD values of other variables were above two (Table 2). Among the evaluated traits, moisture, protein and gluten ratio could be predicted by NIR instrument via current PLS regression models.

*The relationships between the screened wavelengths and calibration parameters*

The relationships of each wavelength between 1200-2400 nm with calibration parameters in the generated NIR calibrations were shown Figure 3. These graphs indicate that the correlation coefficients of gluten index and moisture ratio with the screened wavelengths were lower compared to those of other calibration parameters. Moisture ratio had negative correlation with transformed spectra between the wavelengths 1466-1479 nm, while the correlation coefficients were positive at the 1856 and 1767 nm. The correlation coefficients varied between -0.42 and 0.38 for moisture ratio and screened spectrum region. This range was between -0.46 and 0.47 for protein ratio. Positive correlation was found on the wavelengths between 1664-1679 nm and 1714-1728 nm, while the correlation coefficients were negative within 1536-1558 nm and 1748-1760 nm ranges for this trait. Negative correlations were found on different wavelengths within 1536-1553 nm and 1748-1761 nm ranges, while positive correlations were found on different regions in 1666-1679 nm and 1715-1725 nm ranges with gluten ratio. Correlation values between the gluten ratio and screened region varied from -0.68 to 0.78. The correlation coefficients between gluten index values and the wavelengths changed between -0.32 to 0.38. The highest correlations of gluten index with the wavelengths were in 1852 nm and 1466 nm in the negative and positive directions, respectively. The correlation values between wavelengths and sedimentation value changed from -0.48 to 0.60. The positive correlations were determined between the wavelengths in the 1665-1679 nm, 1715-1729 nm and 1927-1933 nm regions, while negative correlations were determined in 1538-1550 nm and 1747-1763 nm. Looking at the correlation graph between modified sedimentation value and scanned wavelengths, there were positive correlations with the wavelengths within 1744-1768 nm while the direction was negative within 1662-1670 nm ranges. The correlation values varied between -0.52 and 0.51 for this trait.

The main objective of NIR calibration development is not putting forward the relationships between screened wavelengths and calibration parameters. Indeed, the scientific literature about this subject is not abundant. However, NIR estimations are made based on different reactions of chemical bonds resulted by the energy changes in the NIR electromagnetic spectrum. Therefore, uncovering such relationships may enable us to reach different conclusions. For example, as shown in the correlation graphs, it can be said that the traits associated with protein structure in wheat flour except the gluten index value were to be associated with the similar spectral regions (Figure 3).

That is, it is possible to estimate protein quality traits based on the differences of similar wavelengths. Indeed, COZZOLINO ET AL. (2006) emphasized that the regions related with N-H and O-H bounds in the electromagnetic spectrum were correlated with gluten ratio and sedimentation value. They argued that these regions could be utilized in the development of calibrations for these traits. In addition, it was shown that moisture ratio can be identified in four different regions of electromagnetic spectra, and of those regions, the 1400-1500 nm area was found to be highly correlated with moisture content in our study. This approach used in our study was proposed to determine the associations between the scanned NIR wavelengths and composition of fatty acids in rapeseed by VELASCO AND BECKER (1998).

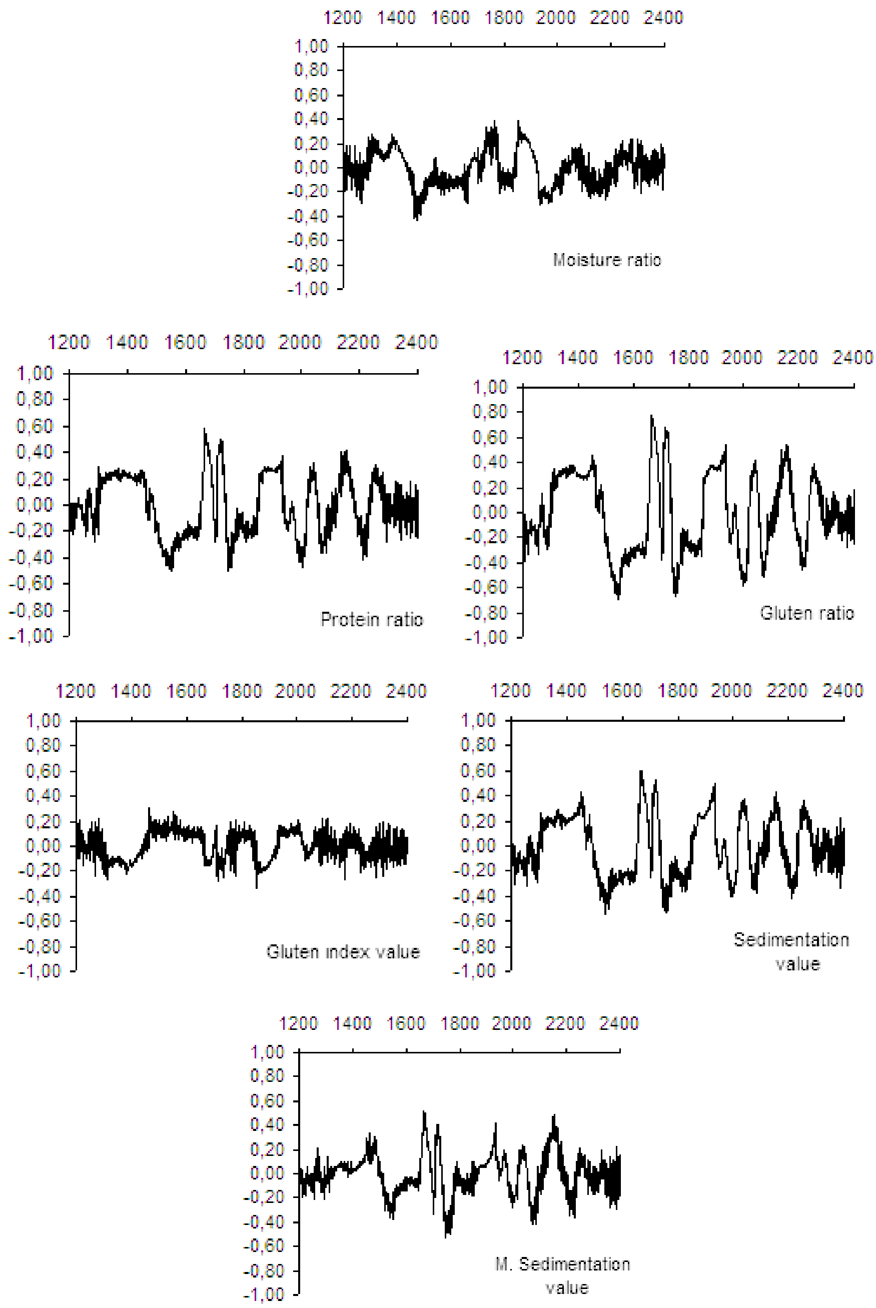


Figure 3. Graphs for the correlations between quality traits and transformed spectral reflectance values within 1200-2400 nm in wheat flour

## CONCLUSIONS

Based on the calibration evaluation parameters, namely, SEE, RMSEP, RMSECV, RPD and $R^2$, protein, moisture and gluten ratio were the best estimated traits, among the others. Although, the correlation coefficients between the reference values and NIR measurements for gluten index and sedimentation values were acceptable, the robustness of the estimations was not high due to rather large standard error terms. The reasons for this may be the high standard deviations associated with these traits in the reference analyses. Moisture content and gluten index value seemed to be associated with the different wavelengths than those with the other traits. Data suggested that similar regions of electromagnetic spectra could be used for the prediction of protein ratio, sedimentation and modified sedimentation values. Use of more sensitive reference analysis instruments and/or methods could enhance the results of NIR estimates. Improvement of the results may also be possible by sample (e.g., drying samples, extracting out non-target components) or data pretreatments (e.g., ordering of derivatives) in the future researches.

## BIBLIOGRAFY


1. ANNONYMOUS (1980). ICC Standard No: 105/1. Method for the Determinations of Crude Protein in Cereals and Cereal Products for Food and for Feed. Standard Methods of the International Association for Cereal Chemistry (ICC). Verlag Moritz Schafer. Detmold.
2. ANNONYMOUS (1994a). ICC Standard No: 137. Mechanical Determinations of the Wet Gluten Content of Wheat Flour (Glutomatic). Standard Methods of the International Association for Cereal Chemistry (ICC). Verlag Moritz Schafer. Detmold.
3. ANNONYMOUS (1994b). ICC Standard No: 155. Determination of Wet Gluten Quantity and Quality (Gluten Index ac. to Perten) of Whole Wheat Meal and Wheat Flour (Triticum aestivum). Standard Methods of the International Association for Cereal Chemistry (ICC). Verlag Moritz Schafer. Detmold.
4. ANNONYMOUS (1994c). ICC Standard No: 116. Determination of the Sedimentation Value (According to Zeleny) as an Approximate Measure of Baking Quality. Standard Methods of the International Association for Cereal Chemistry (ICC). Verlag Moritz Schafer. Detmold.
5. ATLI A., KÖKSEL H., DAĞ A., (1988). Unda süne ve kımıl zararının belirlenmesi için geliştirilen yöntemler ve bu yöntemlerin uygulanabilirliği üzerine araştırmalar. Tarla Bitkileri Merkez Araştırma Enstitüsü Yayınları. Genel Yayın No: 1988/3, Araştırma Yayın No:1988/2, Tarım Matbaası.
6. BAILLERES H., DAVRIEUX F., HAM-PICHAVANT F., (2002). Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program. Ann. for Sci 59:479–490.
7. COZZLINO D., DELLUCHI I., KHOLI M., VAZQUEZ D., (2006). Use of Near Infrared Reflectance Spectroscopy to Evaluate Quality Characteristics in Whole-Wheat Grain, Agric. Téc. 66(4):370-375.
8. CWS MANUAL, (2003). Sensologic Calibration Workshop Version 2.02, Sensologic Gmbh, Germany.
9. DOWELL F. E., MAGHIRANG E.B., XIE F., LOOKHART G.L., PIERCE R.O., SEABURN B.W., BEAN S.R., WILSON J.D., CHUNG O.K., (2006). Predicting Wheat Quality Characteristics and Functionality Using Near-Infrared Spectroscopy. Cereal Chem. 83(5):529–536
10. ELFADL E., REINBRECHT C., CLAUPEIN W., (2010). Development of near infrared reflectance spectroscopy (NIRS) calibration model for estimation of oil content in a worldwide safflower germplasm collection. International Journal of Plant Production, 4(4):259-270.
11. HACISALIHOGLU G., LARBI B., SETTLES M., (2010). Near-Infrared Reflectance Spectroscopy Predicts Protein, Starch, and Seed Weight in Intact Seeds of Common Bean (*Phaseolus vulgaris* L.). Journal of Agric. Food Chem., 58:702–706.

12. HRUSKOVA M., FAMERA O., (2003). Prediction of wheat and flour Zeleny sedimentation value using NIR technique. Czech J. Food Sci. 21:91-96.
13. JIRSA O., HRUSKOVA M., SVEC I., (2008). Near-infrared prediction of milling and baking parameters of wheat varieties. Journal of Food Engineering. 87:21–25
14. LU W., YANG H., XU G., QING D., (2000). Mod. NIR Spectrosc. Anal. Technol.. Chinese Shihua Publishing Company, Beijing.
15. PASQUINI C., (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. J. Braz. Chem. Soc. 14(2):198-219.
16. SANDORFY C., BUCHET R., LACHENAL G., (2007). Principles of molecular vibrations for near-infrared spectroscopy. In Near-Infrared Spectroscopy in Food Science and Technology; Ozaki, Y., McClure, W. F., Christy, A. A., Eds.; John Wiley and Sons, Inc.: Hoboken, NJ, pp: 11-46.
17. SAS INST., 1999. SAS V8 User Manual, Cary, NC.
18. SPIELBAUER G., ARMSTRONG P., BAIER J.W., ALLEN W.B., RICHARDSON K., SHEN B., SETTLES A.M., (2009). High-Throughput Near-Infrared Reflectance Spectroscopy for Predicting Quantitative and Qualitative Composition Phenotypes of Individual Maize Kernels. Cereal Chem. 86(5):556-564.
19. VELASCO L., BECKER C.H., (1998). Estimating the fatty acid composition of the oil in intactseed rapeseed (*Brassica napus* L.) by near infrared reflectance spectroscopy. Euphytica, 101:221-230