

UDK: 502/504

*Originalni naučni rad
Original scientific paper*

DEVELOPMENT OF MULTIVARIATE REGRESSION MODEL FOR QUANTIFICATION OF PROXIMATE CONTENT IN *VIGNA RADIATA* USING FOURIER TRANSFORM –NIR SPECTROSCOPY

Ravi Pandiselvam^{*1}, S. Sunoj², D. Uma³

¹ICAR –Central Plantation Crops Research Institute,
Physiology, Biochemistry and Post Harvest Technology Division, Kerala, India

²North Dakota State University,
Department of Agricultural and Biosystems Engineering, Fargo, ND-58102, USA

³Tamil Nadu Agricultural University,
Department of Biochemistry, Coimbatore . India

Abstract: The Fourier Transform Near Infrared (*FT-NIR*) absorbance spectra (12800-3600 cm⁻¹) of 222 green gram samples was used to build calibration models for the determination of the content of protein, fat and carbohydrate. The samples that comprised the dataset had an average composition of 22.18% of protein, 1.30% fat, and 50.72% carbohydrate. Multivariate regression was used to develop the quantitative models for protein, fat and carbohydrate compounds. The root mean square error of cross validation (*RMSECV*) was 0.191 ($R^2 = 91.52$) for protein, 0.0271 ($R^2 = 88.54$) for fat and 0.765 ($R^2 = 93.62$) for carbohydrate. A fast, simple and accurate method to quantify the proximate content of green gram was developed by using *FT-NIR* spectroscopy. The results showed that *FT-NIR* spectroscopy could support chemical analysis methods.

Key words: *FT-NIR Spectroscopy, green gram, protein, first derivative, calibration*

* Corresponding author. E-mail: anbupandi1989@yahoo.co.in

The authors would like to acknowledge Shri.S.Kumaravel, Assistant professor and Bannu Priya, Research Scholar, Indian Institute of Crop Processing Technology, Thanjavur, for providing chemicals and support during analysis. The special thanks to Dr.V.Thirupathi, Professor and V.Arun Prasath, Research Scholar for his valuable support.

INTRODUCTION

Green gram (*Vigna radiata*) is one of the important pulse crops in India. It is a protein rich staple food [1]. It contains about 24 percent protein, which is almost three times that of cereals. It supplies protein requirement of vegetarian population of the country. It is consumed in the form of split pulse as well as whole pulse, which is an essential supplement of cereal based diet. The biological value improves greatly, when wheat or rice is combined with green gram because of the complementary relationship of the essential amino acids. The biochemical composition changes the structure of the grains. It influences the engineering properties of grains reported by [2-4]. Most common methods reported for the determination of protein content are kjeldahl and lowery method. But these methods are tedious, time-consuming, destructive, not economically viable and as they require highly skilled operators.

Fourier transform near-infrared (*FT-NIR*) spectroscopy is an analytical technique that has gained popularity in recent years for analyzing a wide variety of samples used in used in nutritional, pharmaceutical, petrochemical, textile and agricultural industries. The major strength of *FT-NIR* spectroscopy is a rapid technique, nondestructive, accurate and that can be employed as a replacement for time-consuming chemical methods. It is a non-invasive method for the characterization of fat, nitrogen and moisture content in cocoa powder [5] and protein and moisture in fishmeal [6]. Indistinctly, many authors have used different ranges of wavelengths, which extend from the visible spectrum to the NIR in order to estimate multiple properties and they have established different wavelengths ranges. But until now there is not consent between researchers which is the best wavelengths range to study each grain parameters due to different grain nature, type, characteristic and influence of grain grow environments [7]. [5] conclude that the second derivative of *NIR* is the recommended procedure to quantify fat, nitrogen, and moisture content in cocoa powders by infrared spectroscopy and also showed that no single wavelengths were strongly correlated with the protein, fat and carbohydrate content of cocoa powder, which indicates the difficulty of using selected wavelengths or bans to accurately predict the protein, fat and carbohydrate content of the selected product.

On the other hand, India still lags behind the countries with a well-developed of grain sector, especially in processing technology, quality and taste improvement, variety improvement and storage techniques. The grain chemical composition also changes during ozone fumigation [20, 21]. The aim of the present study was to evaluate the potential of *NIR*-Spectroscopy, as a rapid and non-destructive method to predict the protein, fat and carbohydrate content of the green gram. On the other hand, *NIR* models were developed based on partial least square (*PLS*) technique.

MATERIAL AND METHODS

Raw material. Green gram was procured from Department of Pulses, Tamil Nadu Agricultural University, Coimbatore of Tamil Nadu, India and used for the study. The green gram was cleaned manually to remove all foreign materials.

Sample preparation. The initial moisture content of the samples was determined by hot air oven drying at $103 \pm 2^\circ\text{C}$ for 5 hours [8]. The initial moisture content of green

gram was found to be 11.18% (db). In order to maintain the moisture content of green gram for this study, the samples were kept in a refrigerator at $4\pm 2^\circ\text{C}$. The required quantity of green gram sample was withdrawn and equilibrated at room temperature ($29\pm 3^\circ\text{C}$) before conducting different tests [9].

Protein, Fat and Carbohydrate determination. Protein content was estimated using Kjeldahl method. The fat content of the green gram was estimated by the method described by [10]. Soxhlet extraction with petroleum ether was used for determination of the fat content. The percentage of carbohydrate content in the green gram sample was determined by the method reported by [11].

FT-NIR spectroscopy. The *FT-NIR* spectra were collected on multipurpose analyzer (*MPA*) (Bruker Optics, Ettlingen, Germany) equipped with an integrated Michelson interferometer combined with *OPUS* software (v. 7.2 Bruker Optics, Ettlingen, Germany). It was used for spectral acquisition and instrumental control. For the current study spectra's were recorded in diffuse reflectance mode on the green gram sample with sphere macro sample integrating sphere measurement channel over the range $12800\text{--}3600\text{ cm}^{-1}$ at room temperature. For each calibration standard, the spectrum was attained by averaging three spectral scans. Each spectrum was the average spectrum of 64 scans.

Chemometrics. Multivariate analysis was used for quantitative and qualitative analysis. In the present study the in built software (*OPUS/Quant 7.2*) with the instrument was used for multivariate calibration which exclusively uses the PLS algorithm for the calibration and is designed for the quantitative analysis of spectra consisting of bands showing considerable overlap. It correlates more spectral information using larger spectral range with reference value of calibration set. This leads to a higher degree of precision with reduced chance error [12]. Partial Least Square algorithm (PLS), which was proven to be effective in many quantitative applications [13], was used in the present study. The *OPUS 7.2* software was used for PLS analysis. The samples (222) were divided in the two sets, viz., calibration sample and validation sample set randomly, each set containing 111 samples. The quality of the calibration models for prediction of protein, fat and carbohydrate content was checked by cross validation of the models. It has been reported that the number of samples to develop the calibration model should cover the desired quantification range for the specific analyses, with a minimum of 20–50 samples depending on the complexity of the problem [14].

Data analysis. The spectral data were analyzed using PLS regression with various preprocessing techniques. In this study three spectral preprocessing methods were applied comparatively; it includes first derivative, vector normalization and first derivative plus vector normalization. Vector normalization normalizes a spectrum by first calculating the average intensity value and subsequent subtraction of this value from the spectrum. This method is used to account for different samples thickness [15].

Model accuracy. The performance of final PLS model was evaluated in terms of coefficient of determination (R^2) and root mean square error of cross validation (*RMSECV*). The accuracy of the validation models is obtained according to the highest values of R^2 and *RPD* and lowest *RMSECV*.

$$SSE = \sum [\text{Residual}]^2 \quad (1)$$

where:

Residual [-] - experimental value-Predicted value

The coefficient of determination (R^2) gives the proportion of variability of the property that is described by the model.

$$R^2 = \left(1 - \frac{SSE}{\sum (y_i - y_m)^2} \right) \times 100 \quad (2)$$

where:

$y_i = i^{th}$ [-] - observation of experimental value,
 y_m [-] - mean of the reference results for all samples.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (\bar{y}_i - y_i)^2}{n}} \quad (3)$$

where:

n [-] - number of samples in the validation set
 y_i, \bar{y}_i [-] - measured and predicted value of the i^{th} observation in the test set.

The number of *PLS* factors included in the model is chosen according to the lowest *RMSECV*.

RESULTS AND DISCUSSION

Distribution of data sets. The 222 samples of green gram that comprised the dataset under study had an average composition of 22.18% of protein, 1.30% of fat and 50.72% of carbohydrate. The protein content in the dataset ranged from 20.5 to 23.4%, the fat from 1.16 to 1.48% and carbohydrate from 45.1 to 55.5%. The samples can be split into three groups according to their protein content. The group with high protein content included 98 samples and had an average protein content of 22.8%, ranging from 22.4 to 23.4%; the intermediate protein content group included 104 samples and had an average protein content of 21.8%, ranging from 21.2 to 22.3% and the group with lower protein content included 20 samples and had an average protein content of 21%, ranging from 20.5 to 21.1%. The uneven distribution of the data leads to better prediction by the model and can be used for wide range of samples [16].

Spectra analyzes. Fig. 1 shows the *FT-NIR* spectra of green gram samples which have major peaks at absorbance units (wave numbers) of 4721.1, 5168.5, 5662.2, 6780.8, 8346.8 and 10028.6 cm^{-1} . The peak and depression in the spectra showed the strong and weak absorbance characteristics of green gram within the region of study. Almost spectra of all samples are parallel (Fig. 1), which means the response of the detector for the sample is linear within the range of study and thus may give better results [17].

Major peaks at wave numbers of 4721.1 and 5168.5 cm^{-1} may be due to the stretching vibrations of *NH* and $2 \times \text{C}=\text{O}$ (esters) bonds of protein. Peaks at 5662.2 and 6780.8 cm^{-1} may be due to first overtone of $-\text{CH}$ and ArNH_2 bonds of amine (NH_2) groups. Peaks at 8346.8 and 10028.6 cm^{-1} may be due to second overtone of symmetric stretching of $-\text{CH}_3$ bonds of methyl groups and second overtone of ArNH_2 bonds of amine (NH_2) groups. The vibration of the *NH*, $-\text{CH}$, ArNH_2 and $-\text{CH}_3$ molecules are caused by ingredients such as protein, fat and carbohydrate.

The *NIR* region contains several bands that often overlap, making it difficult to extract spectral parameters of the individual bands [13]. Multivariate analysis with

partial least square technique has provided a way of overcoming these problems through empirical models. Despite the lack of distinct peaks, it has been shown the PLS can extract relevant information for quantitative determinations [18].

Whole range of wave-number was split a fixed interval to know the group of the most effective wave-number for prediction of protein, fat and carbohydrate. After several pre-processing were choice the best model using each multivariate analysis method (*PLS*), the results to each parameter analyzed are summarized in Tab. 1. In the application of *PLS* algorithm, it is generally known that the spectral pre-processing methods and the number of *PLS* factors are critical parameters [13]. The main advantages of *PLS* is that the resulting spectral vectors are directly related to the constituents of interest; also when analyzing systems that have constituent concentrations that are widely varied and number of samples are not very large *PLS* offers satisfactory results [19]. The performance of the final *PLS* factor was evaluated in terms of correlation coefficient of determination (R), root mean square error of cross-validation (*RMSECV*) and the root mean square error of prediction (*RMSEP*). The optimum number of factors is determined by the highest R^2 and lowest value for *RMSECV* and *RMSEP*.

Fig. 2 (a) and (b) shows the R^2 and *RMSECV* values plotted as a function of *PLS* factors for determining protein, fat and carbohydrate content with first derivative and vector normalization method as the pre-processing technique. Seen from figure, R^2 value increased up to certain limit reached a maximum value and thereafter maintained the value for quantitative model for carbohydrate content but there is no significant change in protein and fat content model. Similarly quantitative model for carbohydrate content the *RMSECV* value decreases sharply with initial factors and maintain the value as *PLS* factor increases.

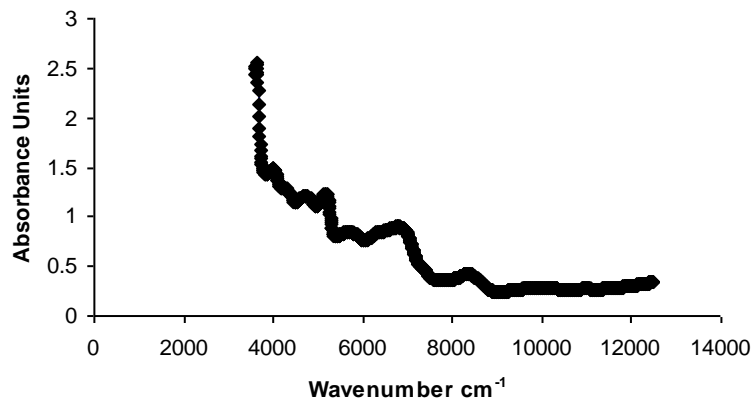


Figure 1. FT-NIR spectra of green gram

The linear regression plot of the validation data sets for the best model showing measured versus predicted protein, fat and carbohydrate content is presented in Fig. 3 (a), (b) and (c) respectively. The equation of the straight line for this cross validation plots of the calibration data sets is represented as $y = 0.9134x + 1.9237$ ($R^2 = 91.08$) for protein, $y = 0.8945x + 0.1373$ ($R^2 = 88.11$) for fat and $y = 0.9366x + 3.2316$ ($R^2 =$

93.18) for carbohydrate showing good performance by this model in predicting protein, fat and carbohydrate content of the green gram samples. The best model was selected based on high value of correlation coefficient and low *RMSECV* values (Tab. 1). The offset and the slope for the equation of the regression line for this set were 1.916 and 0.914, 0.137 and 0.895, 3.234 and 0.937 for protein, fat and carbohydrate respectively. The results of this study clearly demonstrated the efficiency of *FT-NIR* for this application.

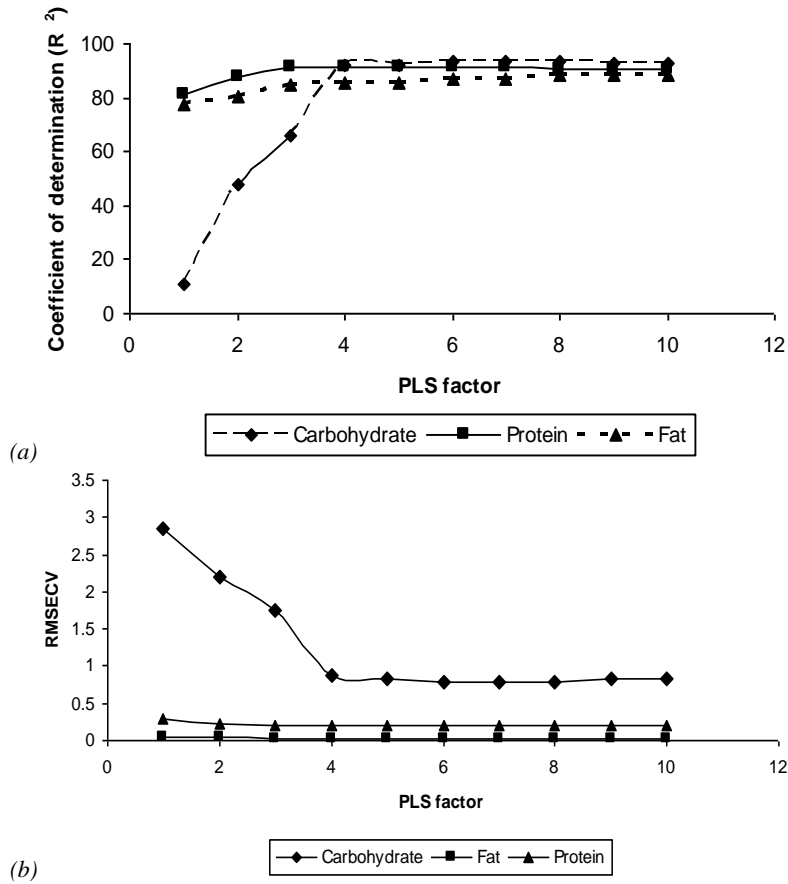
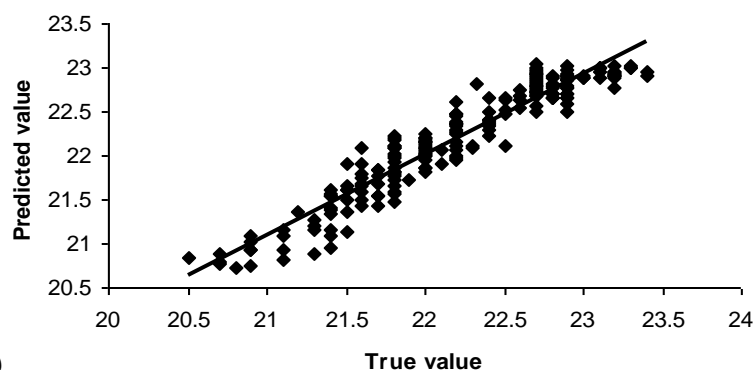


Figure 2. Effects of number of PLS factors on R^2 (a) and *RMSECV* (b) for the validation model

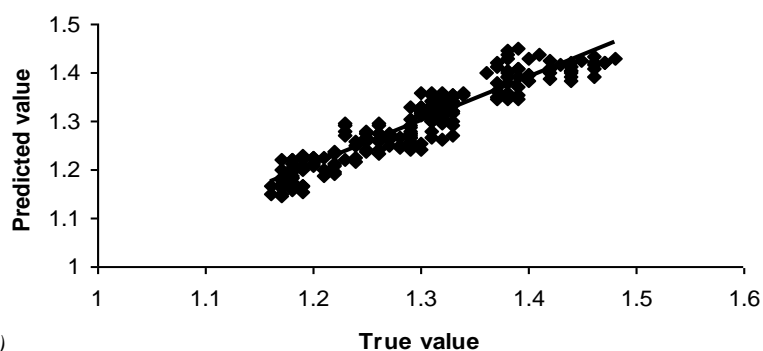
Table 1. R^2 , *RMSECV* and *RMSEP* values corresponding to PLS factor for determining protein, fat and carbohydrate content with different spectral pre-processing methods

Protein					
Pre-processing technique	PLS factors	R^2 (Validation)	<i>RMSECV</i>	R^2 (Calibration)	<i>RMSEP</i>
No pre-processing	7	91.24	0.350	92.34	0.185

Vector normalization	3	90.54	0.202	91.21	0.196
First derivative	5	91.52	0.191	92.46	0.183
First derivative plus Vector normalization	3	91.01	0.196	91.72	0.191
<i>Fat</i>					
Pre-processing technique	PLS factors	R^2 (Validation)	RMSECV	R^2 (Calibration)	RMSEP
No pre-processing	7	86.88	0.029	88.23	0.028
Vector normalization	8	87.85	0.0279	89.95	0.0259
First derivative	9	88.54	0.0271	91.06	0.0245
First derivative plus Vector normalization	8	88.09	0.0276	90.54	0.0251
<i>Carbohydrate</i>					
Pre-processing technique	PLS factors	R^2 (Validation)	RMSECV	R^2 (Calibration)	RMSEP
No pre-processing	6	93.29	0.785	94.27	0.737
Vector normalization	7	93.62	0.765	94.18	0.745
First derivative	6	93.37	0.780	94.09	0.749
First derivative plus Vector normalization	5	93.47	0.774	93.96	0.755



(a)



(b)

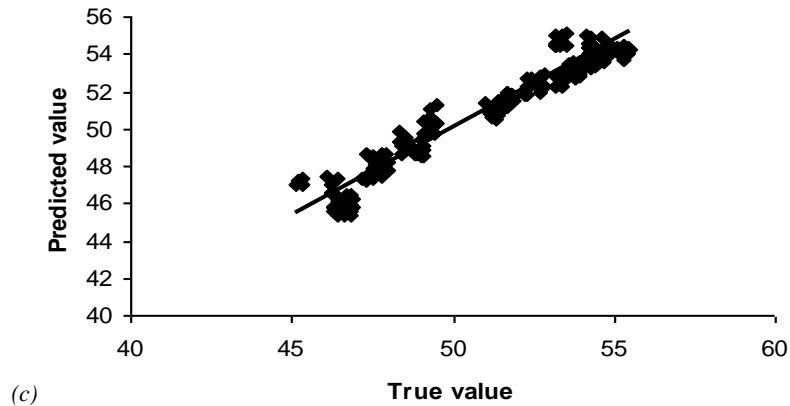


Figure 3. Scatter plot of the experimental Vs NIR predicted protein (a), fat (b) and carbohydrate (c) content of the validation data set

CONCLUSIONS

A rapid and simple *FT-NIR* procedure to estimate protein, fat and carbohydrate content in green gram was developed using a calibration model. The model was developed using the spectral region $3600 - 12800 \text{ cm}^{-1}$. Lower values of *RMSECV* and relatively higher values of R^2 showed that *NIR* spectroscopy has potential to predict the quality of green gram nondestructively with almost same accuracy as that of laboratory method. The results presented in this work show that *FT-NIR* can be used as quick, simple, specific and easy automatic tool to predict the content of protein, fat and carbohydrate in green gram samples. It might be an application for green gram quality monitoring in the grain processing industry and various green gram research stations using *FT-NIR* spectroscopy.

BIBLIOGRAPHY

- [1] Pandiselvam, R., Thirupathi, V. 2015. Reaction kinetics of Ozone gas in Green gram (Vigna radiate). *Ozone: Science & Engineering*, 37, 1-7.
- [2] Ravi, P., Venkatachalam, T. 2014. Important engineering properties of paddy. *Poljoprivredna tehnika*, 39(4), 73-83.
- [3] Pandiselvam, R., Kailappan, R., Pragalyaashree, M. M., Smith, D. 2013. Frictional, Mechanical and Aerodynamic Properties of Onion Seeds. In *International Journal of Engineering Research and Technology* (ESRSA Publications), 2(10), 2647-2657.
- [4] Pandiselvam, R., Thirupathi, V., Mohan, S. 2015. Engineering properties of rice. *Poljoprivredna tehnika*, 40(3), pp.69-78.
- [5] Vesela, A., Antonio, S., Barros, A., Synytsyaa, I., Delgadillo, J., Copikova, M.A., Coimbra, 2007. Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder, *Analytica chimica acta*, 601, 77-86.

- [6] Masoum, S., Alishahi, A., Farahmand, H., shekarchi, M., Prieto, N. 2012. Determination of Protein and Moisture in Fishmeal by Near-Infrared Reflectance Spectroscopy and Multivariate Regression Based on Partial Least Squares, *Iran. J. Chem. Chem. Eng.*, 31(3), 51-59.
- [7] Gomez, A.H., He, Y., Pereira, A.G. 2006. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. *J. of Food Eng.*, 77, 313–319.
- [8] *Official Methods of Analysis*. 2002. Association of Official Analytical Che., Washington DC, USA.
- [9] Pandiselvam, R., Pragalyaashree, M.M., Kailappan, R., Thirupathi, V., Krishnakumar, P. 2014. Moisture Dependent Engineering Properties of Onion Seeds, *Journal of Agricultural Engineering*, 51 (2), 36-43.
- [10] Cohen, E. H. 1917. Association of Official Analytical Chemists, 54, 212.
- [11] Onyeike, E.N., Olungwe, T., Uwakwe, A.A. 1995. Effect of heat treatment and defatting on the proximate composition of some Nigerian local soup thickeners, *Food Che.*, 53, 173-175.
- [12] Tripathi, S., Patel, K.G., Bafna, A.M. 2010. Nondestructive determination of curcuminoids from turmeric powder using FT-NIR, *J. Food Science Tec.*, 47(6), 678–681.
- [13] Sinija, V.R., Mishra, H.N. 2009. FT-NIR spectroscopy for caffeine estimation in instant green tea powder and granules, *LWT - Food Science and Tec.*, 42, 998–1002.
- [14] Chen, C., Cai, W., Shao, X. 2007. An Adaptive for selecting representative calibration samples in the continuous wavelet domain for near-infrared spectral analysis, *Analytical and Bio analytical Che.*, 387, 1041–1048.
- [15] Ravi, P., Venkatachalam, T., Palanisamy, V. 2014. Fourier transform near-infrared spectroscopy for nondestructive and rapid measurement of moisture content of paddy. *Poljoprivredna tehnika*, 39(3), 31-40.
- [16] Tripathi, S., Mishra, H.N. 2009. Rapid FT-NIR method for estimation of aflatoxin B1 in red chili powder, *Food Control*, 20, 840–846.
- [17] Jha, S. N. 2007. Nondestructive Determination of Sweetness of Orange Juice using Portable Near Infrared Spectroscopy, *J. of Agricultural Eng.*, 44(3), 10-14.
- [18] Mcshane, M. J., Cote, G. L. 1998. Near-infrared spectroscopy for determination of glucose lactate, and ammonia in cell culture media, *Applied Spectroscopy*, 52, 1073–1078.
- [19] Cramer, R. D. 1993. Partial least squares (PLS): Its strengths and limitations”, *Perspectives in Drug Discovery and Design*, 1(2), 269–278.
- [20] Ravi, P., Venkatachalam, T., Rajamani, M. 2015. *Decay Rate Kinetics of Ozone Gas in Rice Grains*. *Ozone: Science & Engineering*, 37(5), pp.450-455.
- [21] Pandiselvam, R., Thirupathi, V., Anandakumar, S. 2015. Reaction Kinetics of Ozone Gas in Paddy Grains. *Journal of Food Process Engineering*, 38(6), pp.594-600.

**RAZVOJ MULTIVARIJANTNOG MODELA REGRESIJE ZA
APROKSIMATIVNU PROCENU SADRŽAJA VIGNA RADIATA
KORIŠĆENJEM FURIJEOVE TRANSFORMACIJE – NIR SPEKTROSKOPIJE**

Ravi Pandiselvam¹, Shajahan Sunoj¹, Doraiswamy Uma³

¹ICAR –Centralni institut za istraživanje plantažnih kultura,
Odelek za fiziologiju, biohemiju i postžetvenu tehnologiju, Kerala, India

²Državni univerzitet Severne Dakote,
Institut za poljoprivrednu tehniku i biosisteme, Fargo, ND-58102, USA

³Poljoprivredni univerzitet Tamil Nadu,
Institut za biohemiju, Coimbatore. India

Sažetak: Furijeova transformacija blizu infracrvenog (*FT-NIR*) absorbovanog spektra (12800-3600 cm⁻¹) kod 222 uzorka zelenog pasulja korišćena je da se izrade kalibracioni modeli za određivanje sadržaja proteina, masti i ugljenih hidrata. Uzorci koji su činili skup podataka imali su srednji sadržaj od 22.18% proteina, 1.30% masti i 50.72% ugljenih hidrata. Multivarijantnom regresijom su razvijeni kvantitativni modeli proteina, masti i ugljenih hidrata. Srednje kvadratno odstupanje unakrsne procene (*RMSECV*) bilo je 0.191 ($R^2 = 91.52$) za proteine, 0.0271 ($R^2 = 88.54$) za masti i 0.765 ($R^2 = 93.62$) za ugljene hidrate. Tako je razvijen brz, jednostavan i precizan metod za kvantitativnu procenu sastava zelenog pasulja upotrebom *FT-NIR* spektroskopiju. Rezultati su pokazali da *FT-NIR* spektroskopija može da podrži metode hemijske analize.

Ključne reči: *FT-NIR* spektroskopija, zeleno zrno, protein, pasulj, prvi derivat, kalibracija

Prijavljen: 04.09.2015.
Submitted:
Ispravljen:
Revised:
Prihvaćen: 11.04.2016.
Accepted: