Univerzitet u Beogradu
Poljoprivredni fakultet
Institut za poljoprivrednu tehniku
Naučni časopis
POLJOPRIVREDNA TEHNIKA
Godina XL
Broj 4, 2015.
Strane: 27 - 34

University of Belgrade
Faculty of Agriculture
Institute of Agricultural Engineering
Scientific Journal
AGRICULTURAL ENGINEERING
Year XL
No. 4, 2015.
pp: 27 - 34

UDK: 637.52

Originalni naučni rad
Original scientific paper

DEVELOPMENT OF *PLS* MODEL FOR RAPID ESTIMATION OF PROTEIN CONTENT OF RICE USING FOURIER TRANSFORM - NEAR INFRARED SPECTROSCOPY

Ravi Pandiselvam[*1], Venkatachalam Thirupathi [1], Striramasarma Mohan [2], Doraiswamy Uma[3]

[1]Tamil Nadu Agricultural University,
Department of Food and Agricultural Process Engineering, Coimbatore. India
[2] Tamil Nadu Agricultural University,
Department of Agricultural Entomology, Coimbatore. India
[3]Tamil Nadu Agricultural University,
Department of Bio chemistry, Coimbatore. India

Abstract: This study attempted the feasibility to use near infrared spectroscopy as a nondestructive analysis method to qualitative and quantitative assessment of rice quality of Central Warehousing Corporation, India. A *PLS* model were developed using rice standards of different concentrations in the near-infrared region (4.000–12.000 cm^{-1}). The developed models ware authenticated using test validation technique. *FT-NIR* spectroscopy with chemometrics, using the *PLS*–first derivative plus vector normalization method could predict the protein content of stored rice samples accurately up to an correlation coefficient (R^2) and residual predictive deviation (*RPD*) values were 0,98 and 7,21, respectively. The error values such as root mean square error of cross validation (*RMSECV*) and root mean square error of estimation (*RMSEE*) were 0,28 and 0,25, respectively, with 11 factors in the prediction model. The developed model was applied to predict protein content in rice samples within 15 seconds. The developed procedure was further validated by recovery studies by comparing with micro Kjeldahl method of protein determination. These results show that *NIR* spectroscopy could hold up traditional techniques in studying qualitative assessment of rice.

Key words*: NIR spectroscopy, rice, protein, wave number, chemometrics, PLS model*

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

INTRODUCTION

Rice (*Oryza sativa*) is one of the most important cereal crops in India. It is staple food in most of the Indian states. Eighty per cent of the rice produced in the country is stored at farm level, Food Corporation of India (*FCI*) and Central Warehousing Corporation (*CWC*). In *CWC* and *FCI* are grade the rice based on quality. During procurement, storage and issue time to assess the quality of rice is difficult task and time consuming process.

Proteins are amino acids. In most cereals, the types and amounts of protein significantly affect the end-use of the grain. Proteins in rice grains influence cooking properties. The structural modifications of protein and starch gels may enhance the hardness of the cooked rice prepared from the aged samples [1]. Types of proteins that accumulate in the grain depend on the genotype of the plant, but high temperatures and different nutritional conditions affect the ability of the grain to express the expected suite of storage proteins [2]. Food scientists are interested in knowing the concentration, molecular structure and functional properties of the proteins in foods. The conventional methods used in order to determine the protein content of rice is Kjeldahl method. However, this method is also tedious, required toxic chemicals, destructive, relatively expensive and time-consuming, as well as it require highly skilled operators.

Near infrared spectroscopy (*NIRS*) technique as an advance technology has come to stay in food and petrochemical industries. The NIR spectrum of an organic material gives a global signature of composition based on the assessment of the organic chemical structures containing O-H, N-H and C-H bonds [3, 4]. This technology coupled with the development of chemometric techniques has become a powerful, fast, reliable and non-destructive analytical tool for the measurement of qualitative and quantitative properties in organic materials [5]. The objective of this study was to investigate the capability of using NIR spectroscopy to estimation of protein content in rice.

MATERIAL AND METHODS

Rice samples obtained from Central warehousing corporation, Trichy, India was sorted manually to remove the foreign and undesirable materials then it was used for this study. The required quantity of sample was withdrawn from refrigerator and equilibrated at room temperature (31±2°C) before conducting different tests [6].

FT-NIR spectra were recorded on multipurpose analyzer (*MPA*) (Bruker Optics, Germany) equipped with a quartz beam splitter and highly sensitive lead sulfide detector (12.000–4.000 cm^{-1}) combined with opus 7.2 software. The spectra were acquired in reflectance mode directly on the rice, over the range 12.000–4.000 cm^{-1} [7]. For each sampling, 10 g of rice was analyzed at room temperature and the average spectra were used for further evaluations.

Spectra recorded samples were manually analyzed for creating calibration library of *FT-NIR* spectroscopy. The protein content present in the sample was estimated following Micro-Kjeldahl method as represented by [8] using a laboratory kjel plus equipment (Pelican equipments, model-*REC 22238-A2*, Chennai).

In order to check the accuracy of newly developed *FT-NIR* method, recovery study was conducted by artificially spiking the rice samples and back estimating its amount through the new method.

The standard values were fed into *NIR* library and remaining samples were validated by using suitable chemometric method. The spectral data were analyzed using *PLS*

regression with various preprocessing techniques. The *OPUS* 7.2 software package was used for processing the spectral data and *FT-NIR* models were developed with the full calibration data set. In this study three spectral preprocessing methods were applied comparatively; it includes vector normalization, first derivative and first derivative plus vector normalization. The performance of final *PLS* model was evaluated in terms of root mean square error of cross validation (*RMSECV*), root mean square error of estimation (*RMSEE*), residual predictive deviation (*RPD*) and correlation coefficient of determination (R^2). Ratio of standard deviation to standard error of prediction gives *RPD* value (*SD/SEP*). The accuracy of the calibration models is obtained according to the largest values of R^2 and *RPD* and smallest values obtained for *RMSECV* and *RMSEE* values for cross validation.

$$SSE = \sum \left[\mathrm{Re}\, sidual \right]^2 \tag{1}$$

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 (Eq. 2).

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

where:
y_i - i^{ih} observation of experimental value,
y_m - mean of the reference results for all samples.

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

where:
n - number of samples in the validation set,
$\bar{y}_i$ - measured and predicted value of the ith observation in the test set, respectively.

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

RESULTS AND DISCUSSION

A calibration model was developed using protein content standards of varying concentrations in the near-infrared region (4.000–12.500 cm^{-1}) is shown in Fig 1. From the figure it is seen that almost spectra of all samples are parallel. This means that response of the detector for the sample is linear within the range of study and thus may give better results [9].

Fig 2 shows the preprocessed *FT-NIR* spectra of stored rice which has major peaks at absorbance bands (wave numbers) of 3610.2862, 5145.4293, 5623.7151, 6811.7153, 8316.0013 and 10028.5729 cm^{-1}. The peak and depression in the spectra showed the strong and weak absorbance characteristics of rice within the region of study. These true peaks of spectra were selected after smoothing the spectrum to avoid interference due to

noise. The fundamental vibrations in the 4.000-3.500 cm^{-1} region are generally due to O–H, C–H and N–H stretching. The presence of hydrogen bonding is of great importance in a range of molecules [10].

Protein molecule has amine group and carboxyl group; hence, the peak may play an important role in the estimation of protein. From the Fig.2, which is observed that the major peaks at absorbance bands or wave numbers of 5145.4293, 5623.7151 and 6811.7153 cm^{-1} may be due to the C=O stretching, first overtone of –CH and first overtone of NH_2 of protein molecule, respectively. Peaks at 8316.0013 and 10028.5729 cm^{-1} may be due to second overtone of symmetric stretching of carboxyl group (–CH) and second overtone of primary amines (-NH_2) group, respectively.

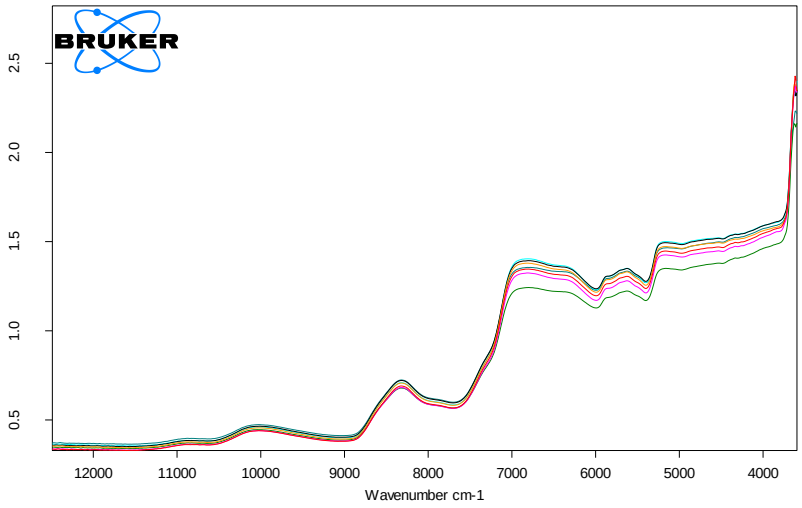

Figure 1. Spectra of rice samples

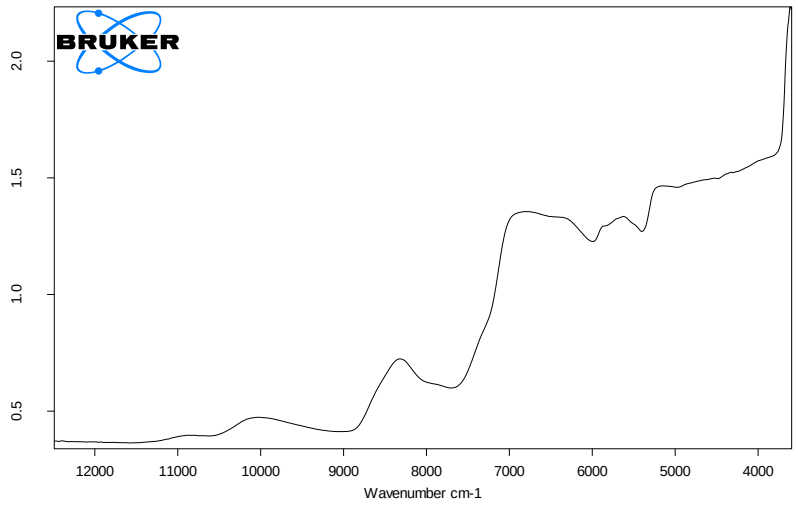

Figure 2. Preprocessed FT- NIR Spectra of rice sample

Protein has a defined amount of energy. When infrared radiation energy is falls on the sample, an energy exchange occurs between the molecules. The most intensive band in the spectrum belong to the vibration and stretching of the carbonyl group (5145.4293 cm^{-1}), followed by the -NH_2 (6811.7153 cm^{-1}), -CH (5623.7151 cm^{-1}) and -CH overtone (8316.0013 cm^{-1}). The vibration of the C=O, -NH_2 and -CH are caused by ingredients such as protein and starch compounds. Some minor peaks observed in the rice spectrum may be due to unknown bond vibrations.

The *NIR* region contains bands that often overlap, making it difficult to extract spectral parameters of the individual bands. Chemometrics have provided a way of overcoming these problems through empirical models that relate the multiple spectral intensities to known analyses in the samples. As the spectra show similar basic *FT-NIR* spectral patterns, mathematical transformations were required to use the *FT-NIR* data for quantitative analysis. Despite the lack of distinct peaks, it has been shown the PLS can extract relevant information for quantitative determinations [11, 12]. PLS algorithm is generally known that the spectral pre-processing methods and the number of *PLS* factors are critical parameters [12].

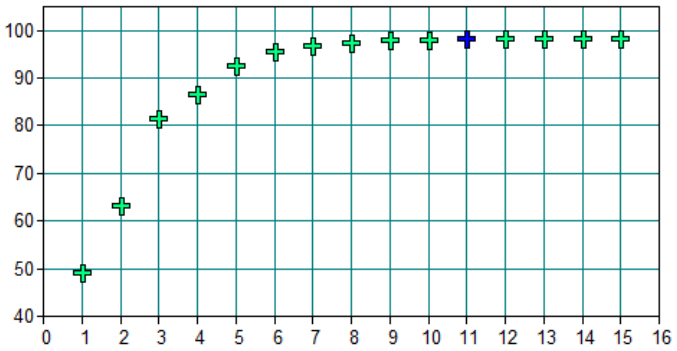

Figure 3. R^2 value as a function of PLS factor

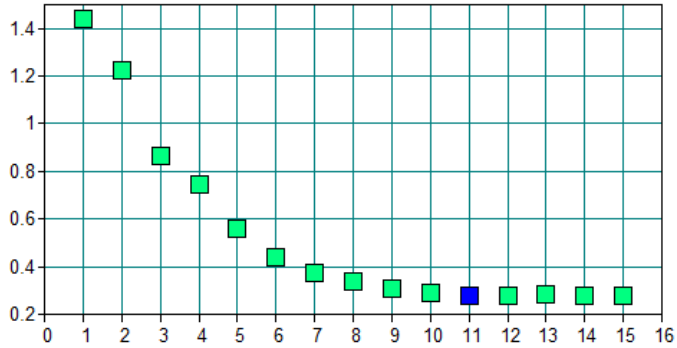

Figure 4. RMSECV as a function of PLS factor

Fig 3, 4 and 5 shows the R^2, *RMSECV* and *RMSEE* values plotted as a function of *PLS* factors for determining protein content with first derivative plus vector normalization method as the pre-processing technique. Seen from figure, *RMSECV* and

RMSEE value decreases sharply with initial factors and maintain the constant value as PLS factor increases from 11 to 15. R^2 value increased up to a *PLS* factor of 11 and maintain the constant value. From the figures, understood that *PLS* factor 11 give satisfactory results for estimation of protein content in rice. From the figures, it is observed that the optimum number of factors is determined by the highest values for R^2 and lowest values of *RMSECV* and *RMSEE*. *PLS* regression method gave R^2, *RMSECV* and *RMSEE* values of 0,98, 0,28 and 0,25, respectively. The results of this study clearly indicated the efficiency of *FT-NIR* for this application.

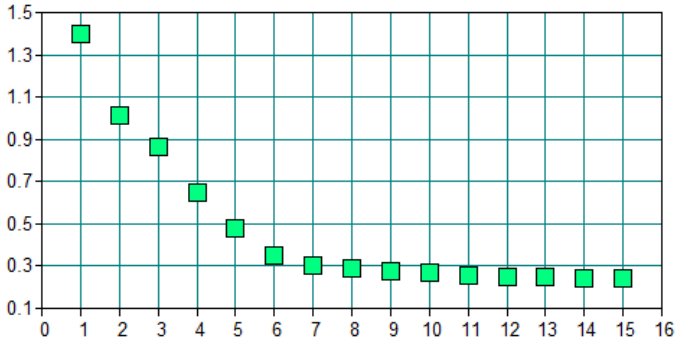

Figure 5. RMSEE as a function of PLS factor

Cross validation was done to check the calibrated values are shown in Fig 6. From the figure, it is seen that a comparison of scatter plots of predicted versus *NIR* true values for validation sets also showed that the *NIR* method gave close results. The developed *NIR* model thus may be able to accurately determine the protein content of rice samples.

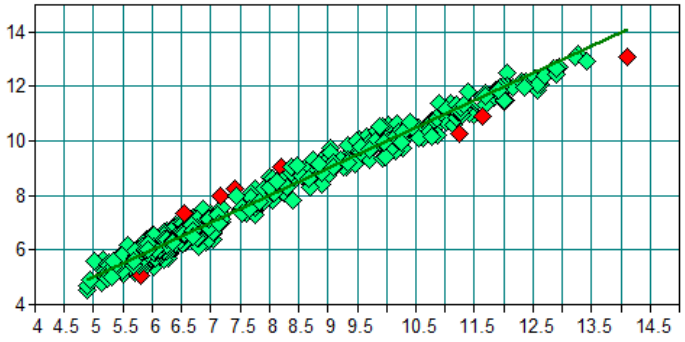

Figure 6. Cross validation of rice samples of actual and NIR predicted

Table 1. Range, mean and standard deviation of the crude protein content of rice samples

Chemical data	*Range*	*Mean*	*SD*
Calibration set (n=318)	*7,25-9,28*	*8,26*	*1,435*
Validation set (n=46)	*7,43-9,22*	*8,325*	*1,26*

SD - Standard deviation

Rice samples prepared were analyzed by *FT-NIR* spectroscopy and the previously developed chemometric method was applied to quantify protein content in rice samples. Results obtained from *FT-NIR* spectroscopy were compared with that of the laboratory methods are shown in Tab. 1. Results obtained from *FT-NIR* method were found to be approximately equal to Kjeldahl method.

CONCLUSIONS

The feasibility of measuring protein content in rice was investigated by using Fourier transform near-infrared (*FT-NIR*) spectroscopy with suitable chemometric techniques. Results of this study show that *NIR* spectroscopy could support chemical analysis in studying the quantitative assessment of rice. *NIR* is very promising tool to detect protein content in rice because, it is fast, nondestructive, accurate and reliable. Lower values of *RMSECV* and *RMSEE* and relatively higher values of R^2 showed that *NIR* spectroscopy has potential to predict the protein content of rice nondestructively with almost same accuracy as that of laboratory method. Furthermore, this developed model allowed for analyze the quality of rice at the time of procurement and fortnight observation of *FCI* (Food Corporation of India) and *CWC* (Centre Warehousing Corporation) with a very simple sample preparation method.

BIBLIOGRAPHY

[1] Tananuwong, K., Malila, Y. 2011. Changes in physicochemical properties of organic hulled rice during storage under different conditions. *Food Chemistry* 125, pp.179–185.

[2] Martin, M., Fitzgerald, M.A. 2002. Proteins in Rice Grains Infuence Cooking Properties. *Journal of Cereal Science* 36, pp.285- 294.

[3] Murray, I., Cowe, I. 2004. Sample preparation. In: C.A. Roberts, J. Workman, J.B. Reeves, (eds.), *Near Infrared Spectroscopy in Agriculture*, American Society of Agronomy, Crop Science Society of America, Soil Science Society of America. Madison, Wisconsin, USA. pp.75-85.

[4] Cozzolino, D., Cynkar, W., Shah, N., Smith, P. 2011. Quantitative analysis of minerals and electric conductivity of red grape homogenates by near infrared reflectance spectroscopy. *Computers and Electronics in Agriculture* 77, pp.81–85.

[5] Tripathi, S., Mishra, H. 2009. A rapid FT-NIR method for estimation of aflatoxin B_1 in red chili powder. *Food Control* 20(9), pp.840-846.

[6] Ravi, P., Venkatachalam, T. 2014. Important Engineering Properties of Paddy. *Agricultural Engineering* 4, pp.73-83.

[7] Ravi, P., Venkatachalam, T., Palanisamy, V. 2015. Fourier Transform Near-Infrared Spectroscopy for Nondestructive and Rapid Measurement of Moisture Content of Paddy. *Agricultural Engineering* 3, pp.31-40.

[8] Sadasivam, S., Manickam, A. 1992. *Biochemical methods*, New age International Publishers, New Delhi. pp 34-37.

[9] Pettersson, H., Aberg, L. 2003. Near infrared spectroscopy for determination of mycotoxins in cereals. *Food Control*, 14(4), pp.229-232.

[10] Stuart, B. 2004. *Infrared Spectroscopy* -Fundamentals and Applications. John Wiley & Sons, Ltd. pp 14-16

[11] McShane, M.J., Cote, G.L. 1998. Near-infrared spectroscopy for determination of glucose lactate, and ammonia in cell culture media. *Applied Spectroscopy* 52, pp.1073–1078.

[12] Sinija, V.R., Mishra, H.N. 2009. FT-NIR spectroscopy for caffeine estimation in instant green tea powder and granules. *Food Science and Technology* 42: pp.998-1002.

RAZVOJ *PLS* MODELA ZA BRZO ODREĐIVANJE SADRŽAJA PROTEINA U PIRINČU UPOTREBOM FURIJEOVE TRANSFORMACIJE – BLISKA INFRACRVENA SPEKTROSKOPIJA

Ravi Pandiselvam[1], Venkatachalam Thirupathi [1], Striramasarma Mohan [2], Doraiswamy Uma[3]

[1]Poljoprivredni univerzitet Tamil Nadu,
Institut za inženjering hrane i poljoprivrednih procesa, Coimbatore. India
[2] Poljoprivredni univerzitet Tamil Nadu,
Institut za poljoprivrednu entomologiju, Coimbatore. India
[3] Poljoprivredni univerzitet Tamil Nadu,
Institut za biohemiju, Coimbatore. India

Sažetak: Ovim istraživanjem je primenjena mogućnost upotrebe bliske infracrvene spektroskopije kao nedestruktivne analitičke metode za kvalitativnu i kvantitativnu procenu kvaliteta pirinča. *PLS* model je razvijen upotrebom pirinča sa različitim koncentracijama u bliskoj infracrvenoj oblasti (4.000–12.000 cm^{-1}). Razvijeni modeli su provereni tehnikom validacionog testa. *FT-NIR* spektroskopija sa hemometrijom, korišćenjem metoda *PLS*-prvog derivate plus vektroske normalizacije, može predvideti sadržaj proteina uskladištenog pirinča, sa koeficijentom korelacije (R^2) i rezidualnom prediktivnom devijacijom (*RPD*) od 0.98 i 7.21, redom. Mere greške, kao srednja kvadratna greška unakrsne validacije (*RMSECV*) i srednja kvadratna greška procene (*RMSEE*) bile su 0.28 i 0.25, redom. Razvijeni model je bio primenjen za procenu sadržaja protein u uzorku pirinča u roku od 15 sekundi. Postupak je dalje ocenjivan poređenjem sa mikro Kjeldahl metodom određivanja proteina. Ovi rezultati pokazuju da *NIR* spektroskopija može da podrži tradicionalne tehnike kvalitativne analize pirinča.

Ključne reči*: NIR spektroskopija, pirinač, protein, broj talasa, hemometrija, PLS model*

Prijavljen: *Submitted*: 08.04.2015.
Ispravljen: *Revised*:
Prihvaćen: *Accepted*: 04.10.2015.