Standardization of the fourier transform near-infrared reflectance spectroscopy for estimation of some oil quality parameters in mustard (*Brassica* spp.)

J. Singh, P.C. Sharma, S.K. Sharma, A. Kumar

Central Soil Salinity Research Institute, Karnal, Haryana, India

ABSTRACT

The possibility of the application of the fourier transform near-infrared reflectance spectroscopy (NIRS) to the analysis of the selected quality parameters in the mustard oil was followed to determine oil, protein, erucic acid and crude fibre content at the Central Soil Salinity Research Institute, Karnal, India. The samples were analysed by reference methods and by the fourier transformed near infrared (FT-NIR) spectroscope at integrating sphere within reflectance mode in the wavelength range 10 000–4000/cm (1000–2500 nm) with 32 scans. To develop the calibration model for the examined components, the partial least square was used and this model was validated by full cross validation. The coefficients of determination (r^2) for intact seeds were 0.907, 0.922, 0.902 and 0.903 for oil, protein, erucic acid and crude fibre content, respectively thus showing that NIRS calibrations are applicable for the estimation of seed quality parameters which is highly desirable in *Brassica* breeding programs for a quick and non-destructive analysis of oil, protein, erucic acid and crude fibre contents in intact seed.

Keywords: mustard; oil content; near-infrared spectroscopy; calibration; validation

Brassica oilseed species now hold the third position among oilseed crops and are an important source of vegetable oil. The most common *Brassica* oil-seed crops grown for commercial purposes are rape seeds, (*Brassica campestris* L. and *B. napus* L.) and Indian mustards (*B. juncea* (L.) Czern. & Coss. and *B. carinata* A. Br.). Indian mustard is an important cash crop as well as a source of edible oil in worldwide diet especially in Eastern and North-Western India. Improved genotypes of mustard along with consumer's acceptance of good oil quality are required for obtaining optimum yield and expansion of cultivated area. However, significant inter- and intraspecific variation for these traits exists within *Brassicas*, which can be exploited through selection through screening and breeding for enhancing oil quality of the crops. These concerns prompted an intensive breeding program by researchers, as mustard had become a major oil seed crop.

Oil of the Indian mustard varieties exhibits quite high content of erucic acid (Chauhan et al. 2007). High amount of erucic acid in edible oils was reported to impair myocardial conductance, causes lipidosis in children and increases blood cholesterol (Ackman et al. 1977) with applications in the lubricants industry. Nowadays edible oils are also being utilized as a biofuel substitute for the traditional fossil diesel fuel with the potential to reduce greenhouse gas emissions (Cardone et al. 2003). Because of the adverse effects of high erucic acid content in oil of Indian mustard varieties, the varietal improvement programme in India aims at reducing erucic acid level up to internationally accepted norms, which necessitates non-destructive screening of a large number of samples with limited seed availability especially in the potential germplasm.

High fibre content (12–13%) in the seed meal reflects lower value of the metabolisable energy and it may negatively influence the protein digestibility and bioavailability of minerals such as manganese and zinc (Ahuja and Bajaj 1999, Chauhan et al. 2002). Thus information on the nutritional and

anti-nutritional make-up of mustard oil would be quite useful for the breeders in the quality improvement programme.

In order to utilize or develop genotypes with high content of oil and protein, low content of fibre and erucic acid as well as fast screening of the existing genotypes is required.

Plant breeding programs usually involve extensive evaluations of the quality components of interest. Thus, large numbers of screenings by the standard analytical methods of germplasm lines are usually performed, in order to detect target genotypes. Currently, chemical analysis methods are generally used to estimate oil, erucic acid, protein and crude fibre contents. Although the standard analytical techniques usually offer a high level of accuracy and precision, these methods are expensive, time consuming, and require destruction of seed samples which could be a limitation in case of the valuable and scarce materials.

Recently, the development of low cost, non-destructive, high throughput equipment featuring improved electronic and optical components, along with the advent of computers capable of effectively processing information contained in the spectra, and development of powerful chemometric applications facilitated the expansion of spectroscopic techniques in increasing number of fields, thus allowing efficient management of spectral and chemical data. These screening and selection techniques to measure the parameters would increase breeding efficiency.

Newer technological advances have brought a rapid, lower cost analytical technique termed near infrared reflectance (NIR) spectroscopy. The use of NIR spectroscopy was already reported for the non-destructive screening of oil, fatty acids, protein, amino acid, and individual and total glucosinolate content of rapeseed mustard seeds (Petisco et al. 2010, Chen et al. 2011) in large breeding populations.

Moreover, NIR spectroscopy is a fast, accurate, and non-destructive technique which requires minimal or no sample preparation, and can be used as a replacement for the conventional time-consuming chemical methods. Keeping in view the potential advantages of NIR over chemical methods, the present study was undertaken to develop calibration models for estimation of oil quality parameters of Indian mustard genotypes and to explore its applicability in identifying variability for these traits.

MATERIAL AND METHODS

Standard sample collection. Sixty nine seed samples of mustard genotypes (50 of *Brassica juncea*, 15 of *Brassica napus* and 4 of *Brassica carinata*) were obtained from the Indian Agricultural Research Institute, New Delhi, Directorate of Rapeseed and Mustard Research, Bharatpur and Punjab Agricultural University, Ludhiana, India during 2012–2013. These samples were pre-analysed in laboratory for oil, protein, erucic acid and crude fibre content with chemical methods. The oil content (%) was estimated using nuclear magnetic resonance (NMR), according to the protocol of the AOCS (1980). The protein content (%) was estimated by determining the nitrogen content using the Kjeldahl analysis. Protein content was estimated by multiplying with a factor 6.25 (AOAC 1990). The erucic acid content was determined through gas-liquid chromatograph (Nucon Model 5765, New Delhi, India) equipped with SP 2300 + 2310 SS column, following the procedure of fatty acid methyl esters developed by Thies (1971). Crude fibre content (%) in seed meal was estimated using the modified AOAC method (Ahuja and Bajaj 1999).

All the samples represented the spectral and chemical variability in the mustard in the calibration and validation groups used for preparation of library and standardization of fourier transformed near infrared spectroscopy (FT-NIR).

FT-NIR spectroscopy. For the FT-NIR measurement, the seed samples were poured into glass vials and set into sample holder for the spectral acquisition. Seed samples were analyzed as intact. NIR spectra were recorded in reflectance mode by using a FT-NIR spectrometer (Perkin Elmer, Massachusetts, USA) equipped with an integrative sphere, over the range 10 000–4000/cm (1000–2500 nm) at 1 nm interval and were stored. The spectrum of each sample was the average of 32 scans. Spectrum10 software (Perkin Elmer, Massachusetts, USA) was used for spectral acquisition and instrumental control.

Data pre-processing. Data pre-treatment using mathematical transformation (e.g., derivatives, multiple scatter correction, smoothing) of the NIR spectra were applied to enhance spectral features and/or remove or reduce the unwanted sources of variation. The spectral datasets were correlated with oil, protein, erucic acid and crude fibre content by using the partial least squares

(PLS) regression algorithm. Calibrations were performed by using the Spectrum Quant+ software (v.4 60, Perkin Elmer, Massachusetts, USA). To evaluate the calibration performance of the developed models, an external validation procedure was carried out to determine the accuracy and precision of the equations obtained in the calibration for each component in each species. To evaluate the accuracy of the equations, different statistics were used, namely the coefficient of determination (r^2) (Williams 1987); the RPD, which is the ratio of the standard deviation (SD) for the validation samples to the standard error of prediction or performance (SEP) (Bailleres et al. 2002). The mathematical expressions of these statistics are as follows:

$$r^2 = \left(\sum_{i=1}^{n}(y-\bar{y})^2\right) / \left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right)$$

$$\text{RMSEP} = \sqrt{\frac{\sum\left(Y_{\text{pred}} - Y_{\text{ref}}\right)^2}{n}}$$

$$\text{Bias} = \bar{Y}_{\text{pred}} - \bar{Y}_{\text{ref}}$$

$$\text{SEE} = \sqrt{\frac{n}{n-1}}\left(\text{RMSEP}^2 - \text{Bias}^2\right)$$

Where: y – NIR measured value; y – mean 'y' value for all samples; y_i – lab reference value for the i^{th} sample; RMSEP – root mean square error of prediction; Y_{pred} – predicted value; Y_{ref} – reference value with standard analysis, n – number of samples; Bias – total differences between predicted and reference values.

$$\text{RPD} = \text{SD}/\text{SEP}$$

Where: RPD – relative prediction deviation; SD – standard deviation; SEP – standard error of prediction or performance.

Statistical analysis. The mean, standard deviation and coefficient of variability for different characters among quality characters were worked out following the SAS 9.2 software (SAS Institute Inc., Cary, USA).

RESULTS

FT-NIR spectra. To perform the NIR calibration model for prediction oil, protein, erucic acid and crude fibre content of mustard seeds spectra were collected in available whole NIR range spectral domain 10 000–4000/cm (Figure 1). It is clear from the figure that spectral patterns of all the samples were found to be similar across the whole wavelength range along the X-axis, however along the Y-axis changes among different samples were observed. The NIR spectrum does not only depend on the chemical composition of samples but also on the physical characteristics of the samples, which are usually observed as the background and noise in the spectrum (Chen et al. 2007).

Development of the NIR calibration models. During the process of development of the calibration model and its validation, certain number of samples had to be excluded in order to obtain the most reliable model possible. Therefore, the number of samples used for developing calibration were (Table 1) lower than the initial number of samples. To develop the calibration models, calibration was initially started with 69 samples (50 of *Brassica juncea*, 15 of *Brassica napus* and 4 of *Brassica carinata*), gradually outliers were

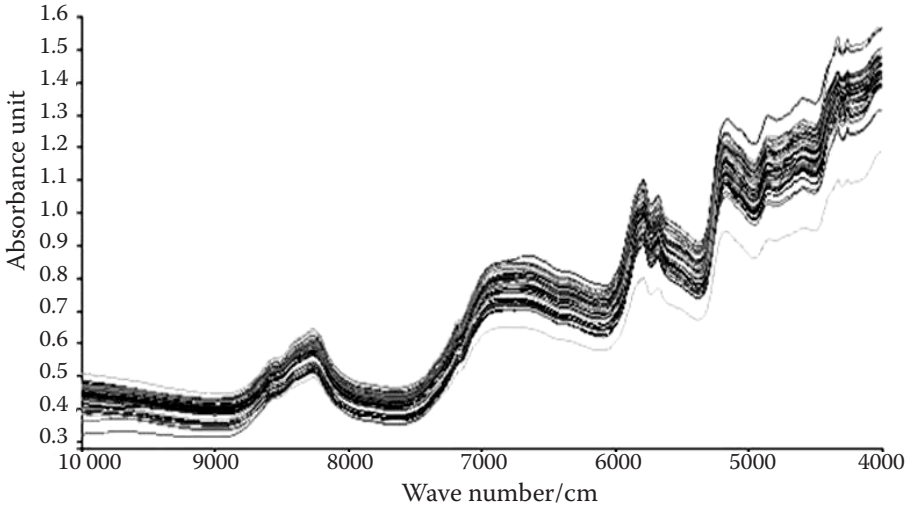

Figure 1. Near infrared spectra of intact seed samples of mustard in the whole NIR range

Table 1. Calibration and validation statics in fourier transformed near infrared spectroscopy (FT-NIR) models for estimation of oil (%), protein (%), erucic acid (%) and crude fibre (%) content in mustard

	Oil		Protein		Erucic acid		Crude fibre	
	calibration $n = 69$	validation $n = 32$	calibration $n = 69$	validation $n = 32$	calibration $n = 69$	validation $n = 32$	calibration $n = 69$	validation $n = 32$
Mean	39.240	39.200	19.450	19.300	39.610	39.100	10.910	10.500
Range	37.0–41.5	37.1–41.3	17.6–20.2	17.0–20.0	0.0–57.30	0.02–57.0	6.3–16.7	6.1–17.1
SD	4.720	4.750	3.200	3.100	3.720	3.720	2.040	2.100
CV	12.029	12.117	16.452	16.062	9.392	9.514	18.698	20.000
r^2	0.900	0.907	0.910	0.922	0.910	0.902	0.910	0.903
SEE	0.70	–	0.50	–	0.73	–	0.36	–
SEP	–	1.01	–	0.68	–	0.80	–	0.43
RPD	–	4.67	–	4.71	–	4.65	–	4.74

SD – standard deviation; CV – coefficient of variation; r^2 – coefficient of determination; SEE – standard error of estimation or calibration; SEP – standard error of performance or cross validation; RPD – relative prediction deviation

removed and finally a set of 32 samples (19 of *Brassica juncea*, 11 of *Brassica napus* and 2 of *Brassica carinata*) was used for cross validation.

The results of the statistics related to PLS calibration model using full cross validation obtained by the FT-NIR technology for the studied traits

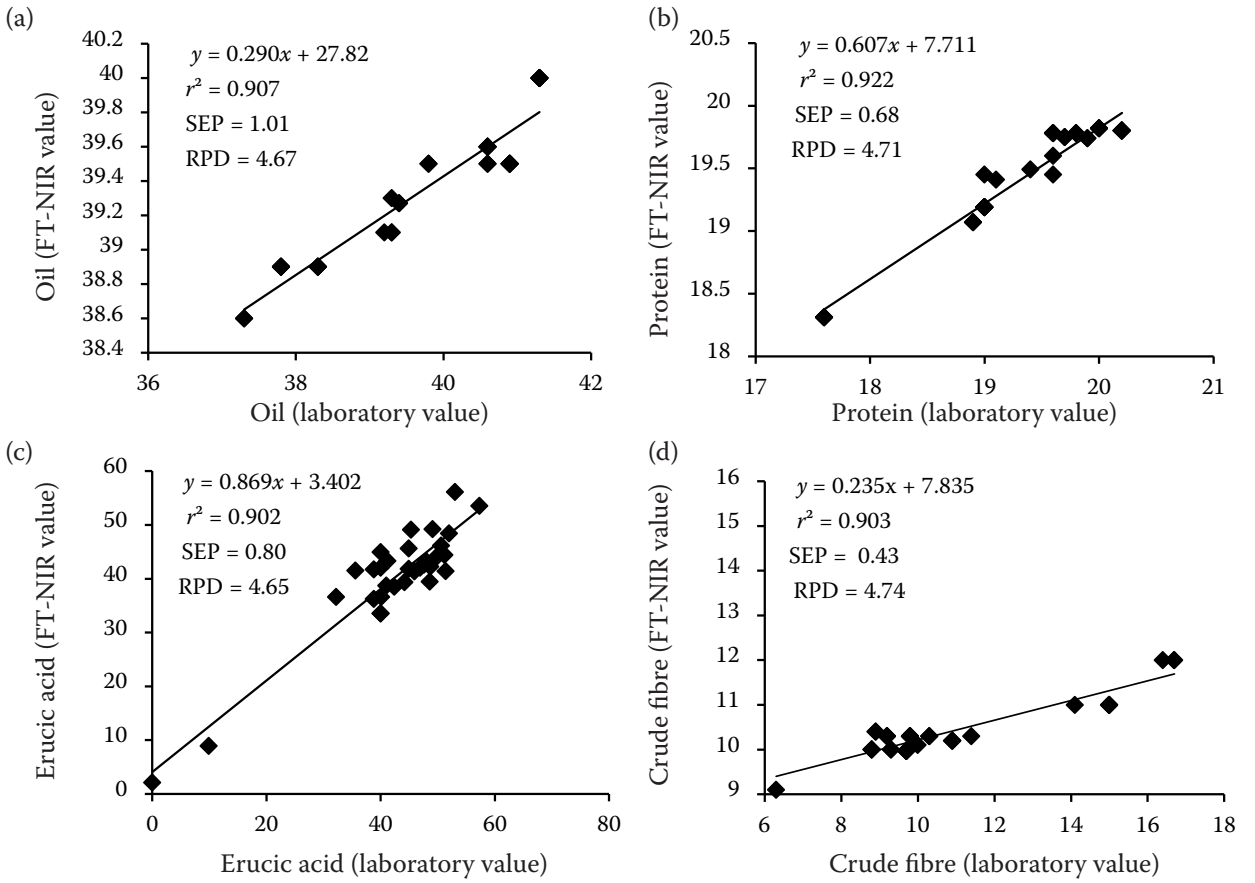

Figure 2. External validation scatter plot for (a) oil (%); (b) protein (%); (c) erucic acid (%) and (d) crude fibre (%) of laboratory vs. predicted data by FT-NIRS in mustard. SEP – standard error of performance or cross validation; RPD – relative prediction deviation

showed that for the developed calibration models standard error of performance (SEP) using the combined spectral data of the three species were 1.01, 0.68, 0.80 and 0.43 for oil content, protein content, erucic acid content and crude fibre content, respectively. Whereas, the coefficient of determination (r^2) were 0.907, 0.922, 0.902 and 0.903 for oil, protein, erucic acid and crude fibre content, respectively (Table 1).

Among the genotypes used in this calibration, the desirable direction of traits revealed that; gp-55 has the highest oil content (41.5%) and protein (20.20%); Jhumka was genotype containing the highest erucic acid (57.30%), whereas the lowest erucic acid content was found in gp-46 (0.00%) and the lowest crude fibre containing genotypes was JT 1 (6.30%).

DISCUSSION

Based upon the cross validation statistics, we observed that the calibrations derived using the combined spectral data of the three species were reliable for the estimation of oil, protein, erucic acid and crude fibre content. Individual spectral data of the three species showed a lower value of coefficient of determination ($r^2 < 0.10$) and RPD value lower than 3, hence these could not be considered fair, were not recommended for screening purposes and were not good for quality control. Similar type of recommendation was given by earlier researchers Williams and Norris (2002).

Equations developed for *Brassica* seed oil showed sufficient accuracy for using this technique as a valuable tool for the analysis of the studied component (Figure 2a). The r^2 (0.907) shown by the equations for oil content determination in mustard seed indicated excellent quantitative information (Shenk and Westerhaus 1996). On the other hand, on the basis of the RPD statistics, the equation was higher than 3 (Table 1), thus being useful for screening (Williams and Sobering 1996).

Protein in mustard was predicted by NIRS with a high accuracy (Figure 2b). The r^2 and RPD shown by this equation (Table 1), together with those for oil in *Brassica*, were the highest of all the components studied. These statistics verify accurate analysis for this component (Williams and Norris 1987). The high r^2 (0.922) and also the RPD statistics (4.71), indicated a high prediction ability (Daun et al. 1994).

Similar to the oil content estimation, the calibration equation developed for erucic acid content was observed to be highly accurate $r^2 = 0.902$. However, on the basis of the RPD statistics, the equation was higher than 3 (Table 1, Figure 2c), thus suggesting its utility as a tool for screening (Williams and Sobering 1996).

Fibre is a component of high interest in mustard, as it was demonstrated to be negatively correlated with the oil and protein content in seed and with meal digestibility (Simbaya et al. 1995). The equation obtained for crude fibre determination resulted in an r^2 value (0.903) in the cross validation that was indicative of good quantitative information equations (Shenk and Westerhaus 1996) (Table 1, Figure 2d). Yet, RPD obtained (4.74) was above the cut-off point of 3, which is recommended by Williams and Sobering (1996) and Williams and Norris (2002) for using the equation for screening purposes.

It could be summarised from the present investigation that FT-NIR is a highly accurate and powerful technique that could be utilized successfully for rapid mass screening in potential germplasm for selecting high oil and protein, low crude fibre and erucic acid containing Indian mustard lines and thus to enhance the effectiveness of quality breeding programme aiming at developing canola quality mustard.

Acknowledgements

Author is grateful to Dr. D.K. Yadav, Division of Genetics, IARI, New Delhi, Head, Division of Genetics and Plant Breeding, PAU Ludhiana and Director, DRMR, Bharatpur, India for providing the samples of *Brassica* spp. used in this study. Thanks are also due to Director, CSSRI, Karnal, for encouragement.

REFERENCES

Ackman R.G., Eaton C.A., Sipos J.C., Loew F.M., Hancock D. (1977): A comparison of fatty acids from high levels of docosenoic acids of rapeseed oil (erucic acid) and of partially hydrogenated fish oil (primarily cetoleic acid) in a non-human primate species in a short-term exploratory study. Bibliotheca Nutrition et Dieta, *25*: 170–185.

Ahuja K.L., Bajaj K.L. (1999): Colorimetric determination of crude fibre in cruciferous oilseeds. Cruciferae Newsletter, *21*: 61–62.

AOAC (1990): Official Methods of Analysis. Association of Official Analytical Chemists, Arlington.

AOCS (1980): Official Methods and Recommended Practices of the American Oil Chemist's Society. The American Oil Chemists' Society, Champaign.

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. Annals of Forest Science, *59*: 479–490.

Cardone M., Mazzoncini M., Menini S., Rocco V., Senatore A., Seggiani M., Vitolo S. (2003): *Brassica carinata* as an alternative oil crop for the production of biodiesel in Italy: Agronomic evaluation, fuel production by transesterification and characterization. Biomass and Bioenergy, *25*: 623–636.

Chauhan J.S., Tyagi M.K., Kumar P.R., Tyagi P., Singh M., Kumar S. (2002): Breeding for oil and seed meal quality in rapeseed-mustard in India – A review. Agricultural Reviews, *23*: 71–92.

Chauhan J.S., Bhadauria V.P.S., Singh M., Singh K.H., Kumar A. (2007): Quality characteristics and their interrelationship in Indian rapeseed-mustard (*Brassica* sp) varieties. The Indian Journal of Agricultural Sciences, *77*: 616–620.

Chen D., Cai W., Shao X. (2007): An adaptive strategy for selecting representative calibration samples in the continuous wavelet domain for near-infrared spectral analysis. Analytical and Bioanalytical Chemistry, *387*: 1041–1048.

Chen G.L., Zhang B., Wu J.G., Shi C.H. (2011): Nondestructive assessment of amino acid composition in rapeseed meal based on intact seeds by near-infrared reflectance spectroscopy. Animal Feed Science and Technology, *165*: 111–119.

Daun J.K., Clear K.M., Williams P. (1994): Comparison of three whole seed near-infrared analyzers for measuring quality components of canola seed. Journal of the American Oil Chemists' Society, *71*: 1063–1068.

Metternicht G.I., Zinck J.A. (2003): Remote sensing of soil salinity: Potentials and constraints. Remote Sensing of Environment, *85*: 1–20.

Petisco C., García-Criado B., Vázquez-de-Aldana B.R., de Haro A., García-Ciudad A. (2010): Measurement of quality parameters in intact seeds of *Brassica* species using visible and near-infrared spectroscopy. Industrial Crops and Products, *32*: 139–146.

Shenk J.S., Westerhaus M.O. (1996): Calibration the ISI way. In: Davies A.M.C., Williams P.C. (eds.): Near Infrared Spectroscopy: The Future Waves. NIR Publications, Chichester, 198–202.

Simbaya J., Slominski B.A., Rakow G., Campbell L.D., Downey R.K., Bell J.M. (1995): Quality characteristics of yellow-seeded *Brassica* seed meals: Protein, carbohydrate, and dietary fiber components. Journal of Agricultural and Food Chemistry, *43*: 2062–2066.

Thies W. (1971): Schnelle und einfache Analysen der Fettsaurzusammensetzung in einzelnen RapsKotyledonen. I. Gaschromatographische und papierchromatographische Methoden. Zeischrift für Pflanzenzücht, *65*: 181–202.

Williams P., Norris K. (2002): Variables affecting near-infrared reflectance spectroscopy analysis. In: Williams P., Norris K. (eds.): Near-Infrared Technology in the Agricultural and Food Industries. Volume 2. American Association of Cereal Chemist, St. Paul.

Williams P.C., Norris K. (1987): Near-Infrared Technology in the Agricultural and Food Industries. AACC, Inc., USA.

Williams P.C., Sobering D.C. (1996): How do we do it: A brief summary of the methods we use in developing near infrared calibrations? In: Davies A.M.C.,Williams P.C. (eds.): Near Infrared Spectroscopy: The Future Waves. NIR Publications, Chichester, 185–188.

Received on May 28, 2013
Accepted on August 26, 2013

Corresponding author:

Dr. Jogendra Singh, Ph.D., Central Soil Salinity Research Institute, Karnal, Haryana, India
e-mails: jogendra82@gmail.com, jogendra@cssri.ernet.in