Application of near-infrared reflectance spectroscopy for the estimation of protein and L-canavanine contents in seeds of one-flowered vetch (*Vicia articulata* Hornem.)

R. Sánchez Vioque*, M. F. Rodríguez Conde and M. de los Mozos Pascual
Departamento de Cultivos Herbáceos, CIA Albaladejito. Ctra. Toledo-Cuenca, Km. 174. 16194 Cuenca, Spain.

Abstract

In this work it has been evaluated the effectiveness of near-infrared reflectance spectroscopy (NIRS) for the estimation of two constituents affecting the nutritional quality of the seeds of one-flowered vetch (*Vicia articulata* Hornem.): protein and L-canavanine, a toxic non-protein amino acid. The NIRS calibrations showed good statistics with coefficients of multiple determination (R^2) of 0.97 for total protein and 0.95 for L-canavanine. The developed equations were further used to estimate the contents of protein and L-canavanine of a set of unknown samples (external validation). The equation for total protein was able to predict with an accuracy similar to that of the reference method showing a correlation (r^2) between the reference and predicted values of 0.95. In the case of L-canavanine, r^2 was only 0.72 and the equation was only effective to discriminate samples into groups of low, medium and high contents.

Additional key words: animal feeding, antinutritional factors, legumes, neglected crops, NIRS, plant breeding.

Resumen

Aplicación de la espectroscopía de reflectancia en el infrarrojo cercano para la estimación de los contenidos de proteína y L-canavanina en semillas de algarroba (*Vicia articulata* Hornem.)

En este trabajo se ha evaluado la eficacia de la espectroscopía de reflectancia en el infrarrojo cercano (NIRS) en la cuantificación de dos componentes que afectan a la calidad nutricional de las semillas de algarroba (*Vicia articulata* Hornem.): proteína y L-canavanina, un aminoácido tóxico no proteico. Las calibraciones NIRS mostraron unos buenos estadísticos, con coeficientes de determinación múltiple (R^2) de 0,97 para la proteína total y 0,95 para la L-canavanina. Las ecuaciones desarrolladas fueron posteriormente utilizadas para estimar los contenidos en proteína y L-canavanina de un grupo de muestras desconocidas (validación externa). La ecuación para proteína total fue capaz de predecir con una exactitud similar a la del método de referencia, mostrando una correlación (r^2) entre los valores de referencia y los predichos de 0,95. En el caso de la L-canavanina, r^2 fue sólo de 0,72 y la ecuación sólo sirvió para discriminar las muestras en grupos de bajo, medio y alto contenidos.

Palabras clave adicionales: alimentación animal, cultivos abandonados, factores antinutricionales, leguminosas, mejora genética vegetal, NIRS.

Introduction

A combination of advantages such as speed, versatility, reliability, low cost, the non-destructive character of the technique, etc, has favoured the use of near-infrared reflectance spectroscopy (NIRS) in many research laboratories and industries in a variety of fields like petrochemicals, food and feed, agriculture, pharmaceuticals and cosmetics, biological and medical applications, etc (Williams, 2001). Protein, fat, starch, humidity and fibre contents are nowadays routinely analysed in agriculture and food science, and the number of new applications of NIRS continuously increases. The technique has become very useful for the analysis of the large number of samples issued from the plant breeding programs or from agronomic studies due to its speed and to the fact that reagents are not required.

One-flowered vetch (*Vicia articulata* Hornem.) has been grown in a large scale in several Mediterranean countries for animal feeding and as green manure. It has been the most important feed legume in Spain but in recent years its area has been reduced up to the nearly disappearance. Main reasons of this situation are the marginalization of the crop from the agricultural subsidies and its scarce domestication, which results in an absence of commercial varieties of one-flowered vetch in Spain. Despite of this neglect, one-flowered vetch is very suitable to the regions of south Europe because of its frost and drought tolerance, good adaptation to infertile soils, and adequate growth against prostrate weeds (López Bellido, 1994; Franco Jubete, 1996). The potential of this species and of other underutilized crops as feed has attracted the attention of researchers to enhance some characteristics of these plants like harvestability, seed production, improved quality and reduced toxicity, etc (Francis *et al*., 1999; Janick, 1999).

The nutritional quality of seeds of one-flowered vetch largely depends on its contents of total protein and L-canavanine. As in many other crops, the content of protein is one of the most important traits to be considered for animal feeding. One-flowered vetch seems adequate to compete with other feed legumes as source of protein although there is a wide range of protein percentage among accessions that points to the interest of selecting the seeds with higher contents (Sánchez Vioque *et al*., 2008). On the other hand, L-canavanine is an undesirable compound in animal diets, responsible for both toxic effects and the reduction of feed intake observed mainly in non-ruminants. In the plant this non-protein amino acid is assumed to be a source of nitrogen for embryo growth and its toxicity is related to the competition with arginine in many enzymatic reactions because of their similar structure. (Rosenthal, 1991; Enneking *et al*., 1993).

The objective of this work is to evaluate the effectiveness of NIRS for the estimation of total protein and L-canavanine in seeds of one-flowered vetch. To our knowledge, this is the first work on the use of NIRS for the estimation of the chemical composition of this species and, as mentioned earlier, can largely facilitate such a task especially when the number of samples to analyse is high.

Material and methods

Plant material. Accessions of one-flowered vetch presently preserved in the germplasm collection of the Banco de Germoplasma Vegetal de Cuenca (BGV-Cuenca, Spain) have been used. This collection has been established by means of donations of other institutions (see acknowledgments). Around 600 seeds of each accession were ground (0.08 mm mesh) in a Ultra Centrifugal ZM 1000 mill (Retsch, Haan, Germany) and kept in a dessicator until analysis for reference methods and NIRS scanning.

Reagents. L-canavanine and sodium amminepentacyanoferrate (II) hydrate were obtained from Sigma-Aldrich (St. Louis, MO, USA), and potassium peroxodisulphate from Panreac (Barcelona, Spain). All other chemicals were of analytical grade.

Reference method for the analysis of protein. The nitrogen content was determined in a Kjeldahl system (Gerhardt, Königswinter, Germany) and expressed in terms of protein content.

Reference method for the analysis of L-canavanine. The L-canavanine content was determined according to Cacho *et al*. (1989) with some modifications as follows. Samples (400-500 mg) were extracted twice by stirring with 5 mL of 0.1 N HCl for 1h. The slurries were centrifuged at 5,000 rpm for 30 min, the supernatants recovered and the volume completed up to 10 mL. One hundred and fifty microlitres of each extract were added with 350 µL of distilled water, 2.1 mL of 0.2 M sodium phosphate pH= 7.0, 300 µL of potassium peroxodisulphate (10 g L^{-1}) and 150 µL of sodium amminepentacyanoferrate (II) hydrate (10 g L^{-1}). This mixture was incubated at room temperature in the dark for 1 h and the absorbance measured at 520 nm in a Beckman Coulter DU 640 spectrophotometer (Fullerton, CA, USA).

The amount of L-canavanine was calculated from a calibrating curve made with the standard product.

NIRS scanning. Ground seed samples were placed in a ring cup and scanned in reflectance mode in a Foss NIRSystems model 6500 equipped with the ISIscan software (Foss Tecator AB, Höganäs, Sweden). Spectral data were recorded as log 1/R from 400 to 2500 nm at 2 nm intervals. One sample from each accession was scanned once for protein calibration and three times for L-canavanine by rotating the position of the ring cup in the sample cup holder, which resulted in 96 spectra for protein calibration and 240 for L-canavanine (*i.e.* 80 accessions, three times). A standard of L-canavanine was also scanned to help us in the selection of the more appropriate range of wavelength to calibrate. Calibration equations for protein and L-canavanine were validated using respectively 25 and 37 accessions not included in the calibration sets (external validation).

NIRS calibration and validation. Calibrations were carried out using the WinISI III 1.50e software (Foss Tecator AB, Höganäs, Sweden). The populations used for the NIRS calibrations and validations were selected randomly with no outlier or any other selection algorithm. Predictive equations were developed using modified partial least square (MPLS) regression and cross validation. In the cross validation method each sample in the calibration is predicted using the remaining samples of the set and it is used to prevent overfitting of the equations. Data pretreatments as multiplicative scatter correction (MSC) and standard normal variate and detrend (SNVD) have been used mainly to overcome problems associated with radiation scattering by a solid sample. MSC calculates the average spectrum from all the data in the training set and uses it as an "ideal" spectrum from which linearizes each spectrum. SNVD is actually two separate algorithms that are usually applied together. SNV is applied first to correct for the effects of the multiplicative interferences of scatter and particle size, similar to MSC. Detrending usually follows to attempt to remove the additional variations in baseline shift and curve linearity typically present in diffuse reflectance spectra. First and second derivatives of the original spectra have been assayed in this work and combined with a gap over which the derivative is calculated from 1 to 12, with smoothings ranging from 1 to 10, and with a second smoothing always set at 1. First derivative of a spectrum is a measure of the slope of the spectral curve at every point and is used to remove baseline offsets among samples. Second derivative is simply obtained by calculating the slopes from the first derivative. Calibration equations were evaluated on the basis of their coefficients of multiple determination (R^2), standard errors of calibration (SEC), coefficients of determination of cross validation (1-VR) and standard errors of cross validation (SECV). SEC is the standard deviation of the differences between the reference laboratory data and the predicted values of the samples within the calibration set whereas SECV accounts for the standard deviation of the differences between the reference laboratory data of a sample and the predicted value generated during the cross validation method. Because the predicted values of a sample used in the SECV are calculated from a regression curve generated without using such sample, SECV is higher than SEC but gives a better estimation than this latter on the robustness of the equation. The value of 1-VR corresponds to the coefficient of determination obtained in the cross validation method.

The external validation process was evaluated by means of the coefficient of determination (r^2), standard error of prediction (SEP) and the ratio between the SD and the SEP (RPD). SEP calculation is similar to those of SEC or SECV but in this case the samples predicted are those of the external validation set.

Results

The distribution of protein and L-canavanine contents in the populations used for the NIRS calibrations and validations are shown in Table 1. For calibrations protein content ranged from 184.7 to 283.1 g kg^{-1} with a mean value of 234.9 g kg^{-1} (SD=13.26), whereas the content of L-canavanine ranged from 2.8 to 5.8 g kg^{-1} with a mean of 4.2 g kg^{-1} (SD=0.61). External validations were carried out using a set of samples that ranged from 172.3 to 281.2 g kg^{-1} with a mean value of 226.4 g kg^{-1} (SD=27.93) for protein, and from 3.0 to 5.5 g kg^{-1} with a mean value of 4.2 g kg^{-1} (SD=0.62) for L-canavanine.

Best calibration equation for protein was obtained using the spectral information from 1108 to 2492 nm with MSC and 1,8,1,1 as mathematical treatment, that is, first derivative (digit 1), a gap over which the derivative is calculated of 8 (digit 8), a smoothing of points of 1 (digit 1) and a second smoothing of 1 (digit 1). As a standard product, L-canavanine showed the main absorptions within the ranges of 408-800 nm and 1480-2476 nm and this range was used to develop an equation for the estimation of the contents of this compound by

Table 1. Distribution of protein and L-canavanine contents (g kg^{-1}) in the populations used for the NIRS calibrations and validations

		n	Mean	SD	Range (min-max)
Protein	Calibration	96	234.9	13.26	184.7 – 283.1
	Validation	25	226.4	27.93	172.3 – 281.2
L-canavanine	Calibration	240	4.2	0.61	2.8 – 5.8
	Validation	37	4.2	0.62	3.0 - 5.5

using a mathematical treatment of 1, 4, 4, 1 (*i.e.* first derivative, gap 4, smooth 4 and second smooth 1) (Figure 1).

In the calibration process, the coefficients of multiple determination (R^2) were 0.97 and 0.95 for protein and L-canavanine, respectively, and were also close to 1 in cross validation (1- VR): 0.96 for protein and 0.91 for L-canavanine. The standard error of calibration (SEC) and standard error of cross validation (SECV) were, respectively, 2.09 and 2.70 for protein, and 0.14 and 0.19 for L-canavanine (Table 2 and Figure 2).

Statistics of the external validations of the equations for the estimation of protein and L-canavanine are shown in Table 2 and Figure 3. Equation for protein showed a coefficient of determination (r^2) of 0.95 between the reference and predicted values and a RPD

of 2.8, and in the case of L-canavanine the r^2= 0.72 and RPD= 1.9, respectively.

Discussion

The applicability and accuracy of a NIRS calibration depend on the population used to calibrate. Some factors such as growing season and growing location result in different environmental factors that may affect the composition of the samples and decrease the accuracy of the calibrations. These influences are well known in the case of grass mixtures where the composition of samples may change from one season to another depending on the percentage of each grass (García Ciudad *et al*., 1999; Gislum *et al*., 2004). In the case of grains, it has been also observed an influence of the environmental conditions on the developing of NIRS calibrations in triticale (Igne *et al*., 2007) and mustard (Velasco *et al*., 1997). A mean to avoid the influence of such factors is to select calibration sets as homogeneous as possible by using samples grown during the same season and location to minimize the environmental influence. Nevertheless, this approach usually results in models with a narrow applicability due to the scarce plant variation included in the calibration set and that consequently fail when they attempt to predict samples belonging to different natural populations (Shenk and Westerhaus, 1991). That is, there is a balance between accuracy and applicability (Locher *et al*., 2005). In this work, accessions from a plant germplasm bank have been directly used without any previous multiplication step of the material in the field and consequently, the seeds were not homogeneous neither in growing season nor location. A priori, this should be advantageous to the application of the developed calibrations in further agronomic studies where the resulting seeds are usually grown under different environmental conditions. Samples of calibration sets showed a rather wide range of variability in both protein and L-
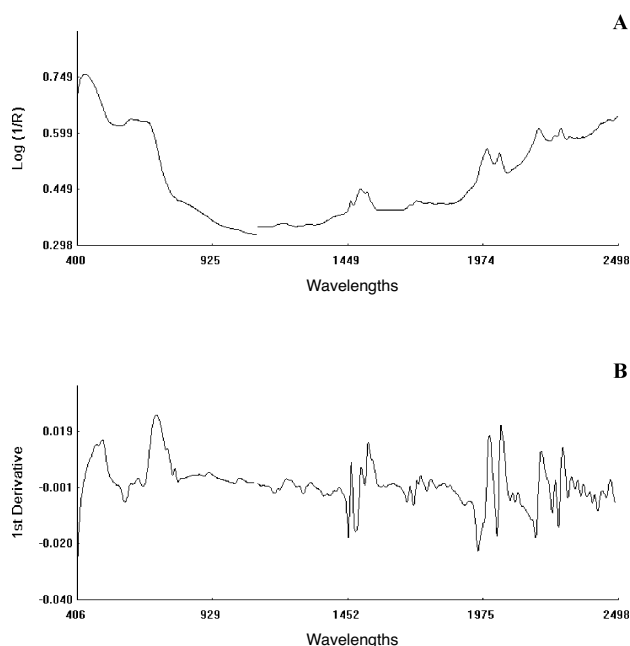

Figure 1. Visible and near-infrared spectrum of standard L-canavanine: (A) Log 1/R, (B) Corrected using mathematical treatment 1,4,4,1 and multiplicative scatter correction (MSC).

Table 2. Calibration and validation statistics for the analysis of protein and L-canavanine by NIRS. Coefficient of multiple determination (R^2), standard error of calibration (SEC), standard error of cross validation (SECV), coefficient of determination of cross validation (1- VR), coefficient of determination (r^2), standard error of prediction (SEP) and ratio between the standard deviation and the SEP (RPD)

	Calibration				Validation		
	R^2	SEC	SECV	1-VR	r^2	SEP	RPD
Protein	0.97	2.09	2.70	0.96	0.95	10.09	2.8
L-canavanine	0.95	0.14	0.19	0.91	0.72	0.33	1.9

canavanine but as a consequence of the natural genetic diversity of seeds the values were not evenly distributed and the samples having extreme values of protein and L-canavanine were scarce. Nevertheless, similarities between the ranges of calibration and validation sets suggest a good depiction of the variation existing in both constituents and a wide applicability of the calibrations developed (Table 1).
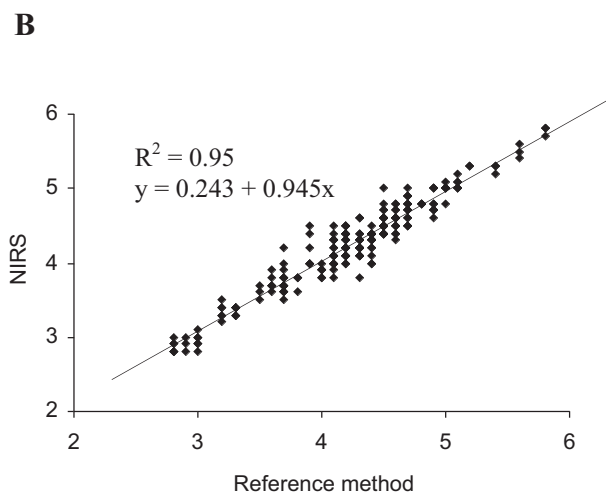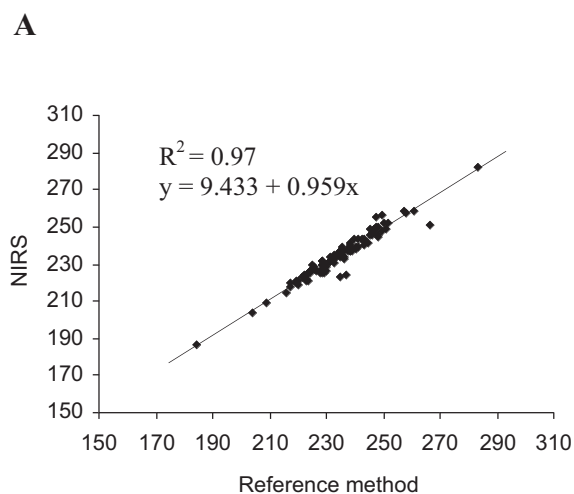
A

$R^2 = 0.97$
$y = 9.433 + 0.959x$

A

$r^2 = 0.95$
$y = 16.202 + 0.964x$

B

$R^2 = 0.95$
$y = 0.243 + 0.945x$

B

$r^2 = 0.72$
$y = 0.971 + 0.766x$

Figure 2. NIRS predicted versus reference method values (g kg^{-1}) of protein (A) and L-canavanine (B) for the calibration sets.
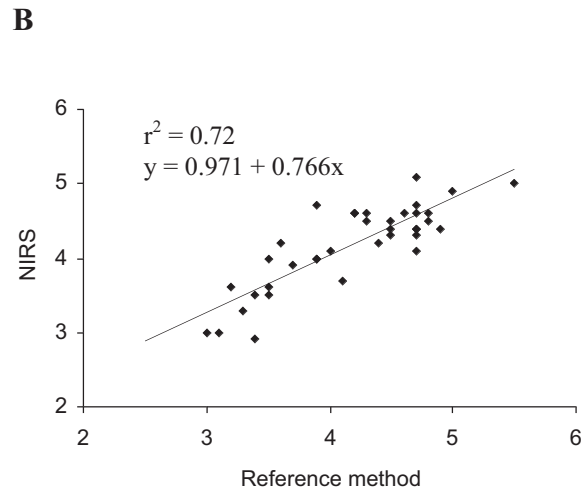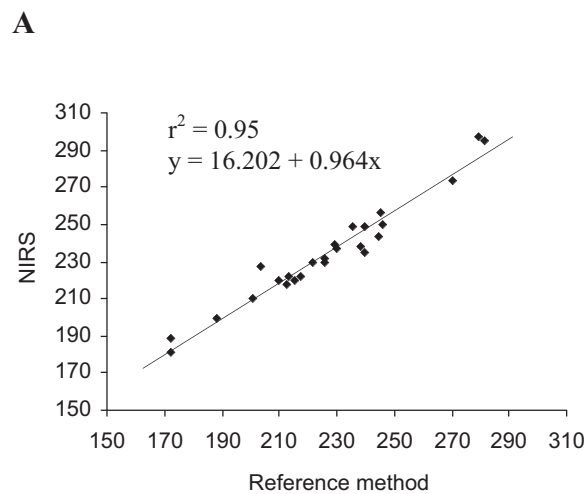
Figure 3. NIRS predicted versus reference method values (g kg^{-1}) of total protein (A) and L-canavanine (B) for the external validation sets.

The best equation for the estimation of total protein by NIRS was obtained without the visible part of the spectral information. Some authors have also discarded the visible region for the estimation of protein contents in seeds of sorghum or rapeseed claiming an increase of noise of the models when it is included (Velasco and Möllers, 2002; De Alencar Figueiredo *et al.*, 2006). For the estimation of L-canavanine the best equation was obtained by selecting to calibrate the regions where the standard L-canavanine displayed the main bands of absorption, which included both visible and infrared regions (Figure 1).

The statistics of the calibrations for the analysis of total protein and L-canavanine pointed to a good applicability of the equations developed on the basis of the values of R^2, SEC, SECV and 1-VR (Table 2 and Figure 2). The statistics R^2 and SEC were not very different from their equivalents statistics in cross validation 1-VR and SECV, respectively, which suggests a good robustness of a calibration.

An external validation by using a set of samples independent of those used for calibrating has been carried out to test the accuracy of the calibration equations. The external validation for protein estimation showed a good coefficient of determination ($r^2 = 0.95$) between the reference and the predicted values and the RPD ratio was close to 3, which indicates that the equation is able to predict with a satisfactory accuracy (Williams, 2001; Locher *et al.*, 2005). In contrast, the model for L-canavanine showed a limited accuracy for the prediction of the external validation set as can be deduced from the values of r^2 (0.72) and the RPD ratio (1.9) (Table 2 and Figure 3). The increases of the SEP values in comparison with the SECV can be attributed to the fact that unlike SECV, the predicted samples in the SEP calculation are not included in the calibration set. In fact, cross validation may be too optimistic to test the quality of a NIRS calibration and performing an external validation is the best way to obtain a more realistic idea of its accuracy and applicability (Fontaine *et al.*, 2001; Igne *et al.*, 2007).

The results shown in this work indicate that NIRS technique could be used to estimate the content of protein and L-canavanine in one-flowered vetch. Nevertheless, some remarks must be done. The model for protein was much more accurate than that of L-canavanine as observed from a better statistics of the external validation. In contrast, the prediction of L-canavanine contents was poor and according to the classification of the r^2 of the equations by Shenk and Westerhaus (1996), the model is only effective to grade samples into groups of low, medium and high content of L-canavanine. The accuracy of the equation could be limited by the low contents of L-canavanine of the samples which probably affects the sensitivity of NIRS (Pasquini, 2003). In summary, protein calibration can be used for quantitative purposes and L-canavanine calibration for qualitative purposes for example in plant breeding programs to improve the nutritional quality of the seeds by selecting those with a low content of L-canavanine. It is possible that the accuracy of the calibration model of L-canavanine can be improved by adding new samples to increase the plant variability and by using the algorithms of the software to select calibration sets resulting in models giving better predictions.

Acknowledgments

References

CACHO J., GARCÍA M.A., FERRANDO I., 1989. Selective spectrophotometric determination of L-canavanine. Analyst 114, 965-968. doi: 10.1039/an9891400965.

DE ALENCAR FIGUEIREDO L.F., DAVRIEUX F., FLIEDEL G., RAMI J.F., CHANTEREAU J., DEU M., COURTOIS B., MESTRES C., 2006. Development of NIRS equations for food grain quality traits through exploitation of a core collection of cultivated sorghum. J Agric Food Chem 54, 8501-8509. doi: 10.1021/jf061054g.

ENNEKING D., GILES L.C., TATE M.E., DAVIES R.L., 1993. L-canavanine: A natural feed intake inhibitor for pigs (isolation, identification and significance). J Sci Food Agric 61, 315-325. doi: 10.1002/jsfa.2740610305.

FONTAINE J., HÖRR J., SCHIRMER B., 2001. Near-infrared reflectance spectroscopy enables the fast and accurate prediction of the essential amino acid contents in soy, rapeseed meal, sunflower meal, peas, fishmeal, meat meal products, and poultry meal. J Agric Food Chem 49, 57-66. doi: 10.1021/jf000946s.

FRANCIS C.M., ENNEKING D., ABD EL MONEIM A.M., 1999. When and where will vetches have an impact as grain legumes? In: Linking research and marketing opportunities for pulses in the 21st century. Proc. Third International Food Legume Research Conference, Adelaide 1997. Current Plant Science and Biotechnology in Agriculture (Knight R., ed). Kluwer Academic Publishers, Dordrecht/ Boston/ London. vol. 34 pp. 671-683.

FRANCO JUBETE F., 1996. La algarroba. In: El cultivo de las leguminosas de grano en Castilla y León (Franco Jubete F., Ramos Monreal A., eds). Consejería de Agricultura y Ganadería de la Junta de Castilla y León, Valladolid, Spain. pp. 247-251. [In Spanish].

GARCÍA CIUDAD A., RUANO A., BECERRO F., ZABALGOGEAZCOA I., VÁZQUEZ DE ALDANA B.R., GARCÍA CRIADO B., 1999. Assessment of the potential of NIR spectroscopy for the estimation of nitrogen content in grasses from semiarid grasslands. Anim Feed Sci Technol 77, 91-98. doi: 10.1016/S0377-8401(98)00237-5.

GISLUM R., MICKLANDER E., NIELSEN J.P., 2004. Quantification of nitrogen concentration in perennial ryegrass and red fescue using near-infrared reflectance spectroscopy (NIRS) and chemometrics. Field Crop Res 88, 269-277. doi: 10.1016/j.fcr.2004.01.021.

IGNE B., GIBSON L.R., RIPPKE G.R., SCHWARTE A., HURBURGH Jr C.R., 2007. Triticale moisture and protein content prediction by near-infrared spectroscopy (NIRS). Cereal Chem 84, 328-330. doi: 10.1094/CCHEM-84-4-0328.

JANICK J., 1999. New crops and the search for new food resources. In: Perspectives on new crops and new uses (Janick J., ed). ASHS Press, Alexandria, VA, USA. pp. 104-110.

LOCHER F., HEUWINKEL H., GUTSER R., SCHMIDHALTER U., 2005. The legume content in multispecies mixtures as estimated with near infrared reflectance spectroscopy: method validation. Agron J 97, 18-25.

LÓPEZ BELLIDO L., 1994. Grain legumes for animal feeding. In: Neglected crops: 1492 from a different perspective (Hernándo Bermejo J.E., León J., eds). FAO Plant Production and Protection Series, Rome, Italy. vol. 26 pp. 273-288.

PASQUINI C., 2003. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. J Braz Chem Soc 14, 198-219. doi: 10.1590/S0103-50532003000200006.

ROSENTHAL G.A., 1991. Nonprotein amino acids as protective chemicals. In: Hervibores: their interaction with secondary plant metabolites. The chemical participants. Academic Press, San Diego, CA. vol. I pp. 1-34.

SÁNCHEZ VIOQUE R., CANTÓN PRADO I., FLORES GIL F., GIMÉNEZ ALVEAR M.J., DE LOS MOZOS PASCUAL M., RODRÍGUEZ CONDE M.F., 2008. Contents of total protein, L-canavanine and condensed tannins of the one-flowered vetch (*Vicia articulata* Hornem.) collection of the Bank of Plant Germplasm of Cuenca (Spain). Genet Resour Crop Evol. 55, 949-957. doi: 10.1007/s10722-007-9302-x.

SHENK J.S., WESTERHAUS M.O., 1991. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. Crop Sci 31, 469-474.

SHENK J.S., WESTERHAUS M.O., 1996. Calibration the ISI way. In: Near infrared spectroscopy: the future waves (Davies A.M.C., Williams P., eds). NIR Publications, Chichester, UK. pp. 198-202.

VELASCO L., MÖLLERS C., 2002. Nondestructive assessment of protein content in single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Euphytica 123, 89-93. doi: 10.1023/A:1014452700465.

VELASCO L., FERNÁNDEZ MARTÍNEZ J.M., DE HARO A., 1997. Determination of the fatty acid composition of the oil in intact-seed mustard by near-infrared reflectance spectroscopy. J Am Oil Chem Soc 74, 1595-1602. doi: 10.1007/s11746-997-0083-3.

WILLIAMS P.C., 2001. Implementation of near-infrared technology. In: Near infrared technology in the agricultural and food industries (Williams P., Norris K., eds) Ed American Association of Cereal Chemists, St. Paul, MN. pp. 145-169.