



Applicability of NIR spectroscopy to determine oil and other physicochemical parameters in Rosa mosqueta and Chilean hazelnut

Franco, Daniel; Núñez, María José; Pinelo-Jiménez, Manuel; Sineiro, Jorge

Published in:
European Food Research and Technology

Link to article, DOI:
[10.1007/s00217-005-0084-0](https://doi.org/10.1007/s00217-005-0084-0)

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Franco, D., Núñez, M. J., Pinelo-Jiménez, M., & Sineiro, J. (2006). Applicability of NIR spectroscopy to determine oil and other physicochemical parameters in Rosa mosqueta and Chilean hazelnut. *European Food Research and Technology*, 222(3-4), 443-450. <https://doi.org/10.1007/s00217-005-0084-0>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Daniel Franco · María José Núñez · Manuel Pinelo ·
Jorge Sineiro

Applicability of NIR spectroscopy to determine oil and other physicochemical parameters in *Rosa mosqueta* and Chilean hazelnut

Received: 17 February 2005 / Revised: 30 May 2005 / Published online: 26 October 2005
© Springer-Verlag 2005

Abstract Due to the increasing interest in certain components, specially the oil, from non-conventional seeds as *Rosa mosqueta* (*Rosa rubiginosa*) and Chilean hazelnut (*Gevuina avellana*), quick determinations of oil and other parameters were carried out by using near-infrared (NIR) spectroscopy. Moisture, oil, fiber (as acid detergent fiber) and protein from solid samples of the seeds as mentioned, along with those of soybean (*Glycine max*), already analyzed by NIR and for serving as control for the variability of the method, were studied. Sample interactions to NIR radiations were processed using the multivariate regression algorithm Partial Least Squared (PLS) to build a calibration model. Standard error of cross-validation (SECV) was used to estimate the prediction error. Moisture of *Rosa mosqueta*, Chilean hazelnut presscake and soybean meal (in the ranges 10–15, 10–15, 8–10%, respectively), acid detergent fiber (60–68, 12–16, 10–15%, respectively), oil (1–4, 14–20, 5–13%, respectively) and protein (1–5, 8–15, 27–45%, respectively) were previously determined by wet analysis using standard methods, so creating a library. The possibility to analyze parameters from very different oilseeds with an acceptable uncertainty was also established. Standard errors of cross-validation were between 1.25 and 2.99%, being the oil content the best predicted parameter.

Keywords Near-infrared · *Gevuina avellana* · *Rosa rubiginosa* · *Glycine max* · Oil · Moisture · Acid detergent fiber · Protein

Introduction

Safety, quality and composition of products in food industry are daily supported by a great number of analyses. The

classical chemical methods are often slow and expensive, so being important to search alternative procedures. Near-infrared (NIR) spectroscopy is a non-destructive technique, characterized by its quick and simple preparation and analysis of samples; it is commonly used to check the quality range of food products. However, the achievement of an optimal calibration is an essential step for applying this technique. The calibration emerges from previous analysis of a lot of samples, which must be representative of the expected variability in unknown samples.

In the last years, a great number of scientific reports showed studies where NIR technique is used to determine some significant variables in food industry, for instance oil content in materials so different as cheese, trout muscle or sunflower [1–3].

Near-infrared spectroscopy was also a valuable tool to discriminate among oils, and to detect possible adulterations [4, 5].

Moisture content is one of the most widely common parameters measured using near-infrared technique since a long time [6]. The zones of the infrared spectrum used to carry out the calibration for moisture content was 7,300–6,500 cm^{-1} . Nowadays, full-spectrum methods are used, instead of several discrete wavelengths. Spectra are processed by multivariate methods such as principal component analysis (PCA) or the PCA-derived methods as Partial Least Square (PLS) to improve calibration, reducing at time the effect of noise in quantitative analysis.

Other recently reported parameters, successfully measured by near-infrared spectroscopy, were sugars in starch [7], total or soluble dietary fiber in cereals [8, 9], or protein in wheat flour [10].

The authors of this paper have been working through last years with non-conventional oilseeds Chilean hazelnut and *Rosa mosqueta*. The main feature of the former is that its oil has a high content of palmitoleic acid (C16:1), about 24%, which penetrates into skin faster than other fatty acids, and is rapidly adsorbed, so being useful as a carrier for pharmaceutical and cosmetic products; often it is added to creams as solar UV filter [11]. Hazelnuts also have a high content in Vitamins A and E. Its protein presents a good quality and

D. Franco · M. J. Núñez (✉) · M. Pinelo · J. Sineiro
Departamento de Ingeniería Química, Escuela Técnica Superior de Ingeniería, Universidad de Santiago de Compostela,
Lope Gómez de Marzoa, 15782 Santiago de Compostela, Spain
e-mail: eqmajose@usc.es
Tel.: +34-981-56-31-00
Fax: +34-981-52-80-50

it is employed for *gevuina* butter, high-protein flour and other products [12]. Regarding *Rosa mosqueta*, the unique properties of its oil were scientifically discovered through a well-planned projects which began in 1983 in Chile, although from 1970s several studies had already started [13]. The secret of the astonishing regenerative effects lies in the content of essential fatty acids (linoleic and linolelic are in the greatest known concentration), so as the content in trans-retinoic acid. The therapeutic oil attenuates scars and helps to prevent the signs of ageing [14]. The content in Vitamin C (up to 0.5% in weight) becomes useful for jams, drinks, cosmetics, etc.

Extracts from the seeds of *mosqueta* and of the hulls of hazelnuts showed also a high antioxidant activity and the former offered a great potential as a natural source of carotenoids for use in food technology [15, 16]. No literature was found on the use of NIR-spectroscopy in the characterization of this material.

The scope of this work was the assessment of methods to determine some physicochemical parameters referred to different oil seeds, using the near-infrared method. These parameters were moisture (% wet basis), oil (% dry basis), acid detergent fiber (adf) (% dry basis) and protein (% dry basis). Samples of soybean meal were included in this work because of the objective of studying a multi-calibration model with samples of different nature. Soybean contributes to variability (high-protein, medium-oil and low-fiber content), and at the same time it is a previously studied seed, which acts as a reference to check the goodness of the employed technique.

The model could be useful for other vegetal materials, at the price of a lower precision than calibrations with an only material. The study also shows the differences among several methods of data pretreatment in multivariate analysis.

Materials and methods

Sample preparation

Rosa rubiginosa seeds and semi-defatted presscake from *Gevuina avellana* were supplied by the School of Biochemical Engineering of Valparaíso (Chile). Soybean meal samples were kindly supplied by Moyresa (Coruna, Spain). The maximum particle size in hazelnut was 0.6 mm, whereas *mosqueta* seeds were ground in an agate mill and then sieve at three different particle sizes (1–2, 0.6–1, and <0.6 mm). The storage temperature was 4 °C.

Analytical determinations

Moisture

Samples were kept in stove at 105–108 °C until constant weight. Moisture value corresponds to the weight loss.

Oil content

Samples were subjected to a liquid–solid extraction using petroleum ether (Panreac, Barcelona, Spain) in a Soxhlet extractor for 8 h. Solvent was vacuum-evaporated. Oil yield was referred to the sample weight, in dry basis.

Acid detergent fiber (adf)

This parameter was assayed by the Van Soest method [17].

Protein content

Kjeldahl method was used. Sample was subjected to reaction with sulfuric acid (copper sulfate was employed as a catalyst). Organic nitrogen was transformed to ammonium sulfate, which was distilled in alkaline conditions and absorbed onto 100 ml of boric acid solution. The reaction gives ammonium borate in solution, which was titrated against 0.1 M sulfuric acid.

Protein content was obtained applying the universal factor 6.25 to nitrogen content.

NIR analysis

Spectra recording

A Bomem MB-160D FT-NIR spectrometer, equipped with Difuss-*ir* accessory to acquire diffuse reflectance spectra, and a high-sensitivity Peltier cell-cooled InAs detector was used. The recording range was 3,500–10,000 cm⁻¹ (2,857–1,000 nm). The Difuss-*ir* from Bomem was optimized to improve the results of NIR analysis with coarse and matrix-complex solid materials, by minimizing the specular reflection. The spectrometer send the interferometer data to PC, where they are transformed to spectral values using Fourier transformers.

NIR calibration

Calibration quality depends on having a representative set of samples. A good calibration contains data of a considerable number of samples, whose spectra are correlated to wet chemical (laboratory) analysis. NIR calibration was performed by multivariate Partial Least Squares regression. For this reason, principal component analysis is used as a spectral data compression technique. PCA is a widely used mathematical tool that led to the Partial Least Squared method to allow quantitative measurements from PCA-processed data. This method takes into account the values of the samples concentration during the spectra decomposition. In this way, the most concentrated

components possess a major weight. A different group of factors and scores are calculated for each component in the version of the method used in this work (PLS-1). It is also possible to apply some mathematical corrections to spectrum data to minimize some errors derived from the dispersion or the different “weight” of data. Although the diffuse reflectance equipment used in this work produces a high signal/noise ratio, two mathematical pre-treatments were used in order to improve the calibration process.

- (1) *Scattering correction with mean centering.* Samples that are measured using diffuse reflectance often exhibit significant differences in the spectra due to the non-homogeneous distribution of the particles. The degree of scattering depends on wavelength, being longer wavelengths more adversely affected. Therefore, the scattering is not uniform throughout the spectrum. This method attempts to remove the effects of scattering by linearizing each spectrum to some ‘ideal’ spectrum of the sample. Mean centering is commonly applied on any multivariate calibration model, which involves calculating the average spectrum of all the spectra in the training set and then subtracting the result from each spectrum. In addition, the mean concentration value for each constituent is calculated and subtracted from the concentrations of every sample.
- (2) *Scattering correction with variance scaling.* It is used to emphasize small variations in the data by giving equal weight to all values. Variance scaling is calculated by dividing the response at each spectral data point by the standard deviation of the responses of all training spectra at that point.
- (3) These data processing procedures are extensively described in Galactic homepage (www.galactic.com).

Statistical analysis

The errors in calibration were expressed as a function of the standard error for cross-validation (SECV), defined as the mean square value of the error. The software BGRAMS32[®] and PLS-IQ[®], from Galactic Corporation (Salem, NH, USA) performs a cross-validation by the following method: in a calculation cycle, a sample is extracted, making a calibration set with the remaining data; value of this sample is calculated and stored as “predicted”. Next calculation cycle repeats this protocol until the procedure is applied to all the samples.

$$\text{SECV} = \sqrt{\frac{\sum_{i=1}^n (Y_{k_i} - Y_{p_i})^2}{n}}$$

where Y_k is the concentration value obtained by wet chemical analysis, Y_p the predicted concentration when the sample is out of the calibration set, and n the number of samples in the training/validation set.

Results and discussion

Figure 1 shows the spectra of the materials used in this work with the calibration zones in grey. In these zones a high correlation between spectral and chemical data variance was observed, the higher value of correlation coefficients r being the criteria to select them (in absolute values, because it can be either positive or negative). These zones corresponded to some of those reported in literature for other materials, although in the past the chemical content was frequently correlated with spectral data using only a wavelength. The modern multivariate algorithms allow us to rebuild the spectra by principal component analysis and to correlate them with chemical data (PLS algorithms). So, wide zones of the spectra can be used for determination of chemical composition, although we can characterize each spectral zone by an absorbance with a known meaning. In fact, 5,787 and 8,264 cm^{-1} corresponds to the first and second overtone of the C–H bonds of CH_2 groups (it was therefore used to fiber and oil determinations). The value 6,896 cm^{-1} corresponds to the second overtone of O–H bonds (used for moisture content determination). Murray and Williams [18] used this wavelength interval to determine moisture in several cereals with good results. The value 4,255 cm^{-1} corresponds to bonds and combination bands of C–O bond (oil) and 6,250 and 9,000 cm^{-1} correspond to the first and second overtone of N–H bonds (protein) [19].

Table 1 shows a summary of the statistical data corresponding to each determination. Number of samples was different depending on the parameter subjected to measurement. Moisture required of a minor amount of samples to obtain a good calibration; by contrast, a higher number of samples was used to determine oil content, although this parameter was obtained with a lower SECV. The samples include different sizes obtained after grinding mosqueta rosehips. For example, the number of total samples used for moisture content determination was 164 (79 soy samples, 10 of grounded hazelnut (size <0.6 mm), 21 of rosehip with size <0.06 mm, 22 with size in the range 0.6–1 mm and 32 with size in the range 1–2 mm).

Moisture

Figure 2a shows the predicted value of moisture using NIR versus the value obtained by the chemical method, and it can be observed that the major amount of samples was in the range 8–14%. The points with the lowest moisture values were those from soybean. Hazelnut presscake and mosqueta rosehips were reported to have moisture values between 10 and 15% [20]. The different nature of the materials did not result in clearly different samples, because moisture content values covered all the range homogeneously. Moistening properties of the mosqueta seeds were not dependent on the size, on the contrary that their oil content (see “Oil” section). Figure 2b and c shows the residuals of

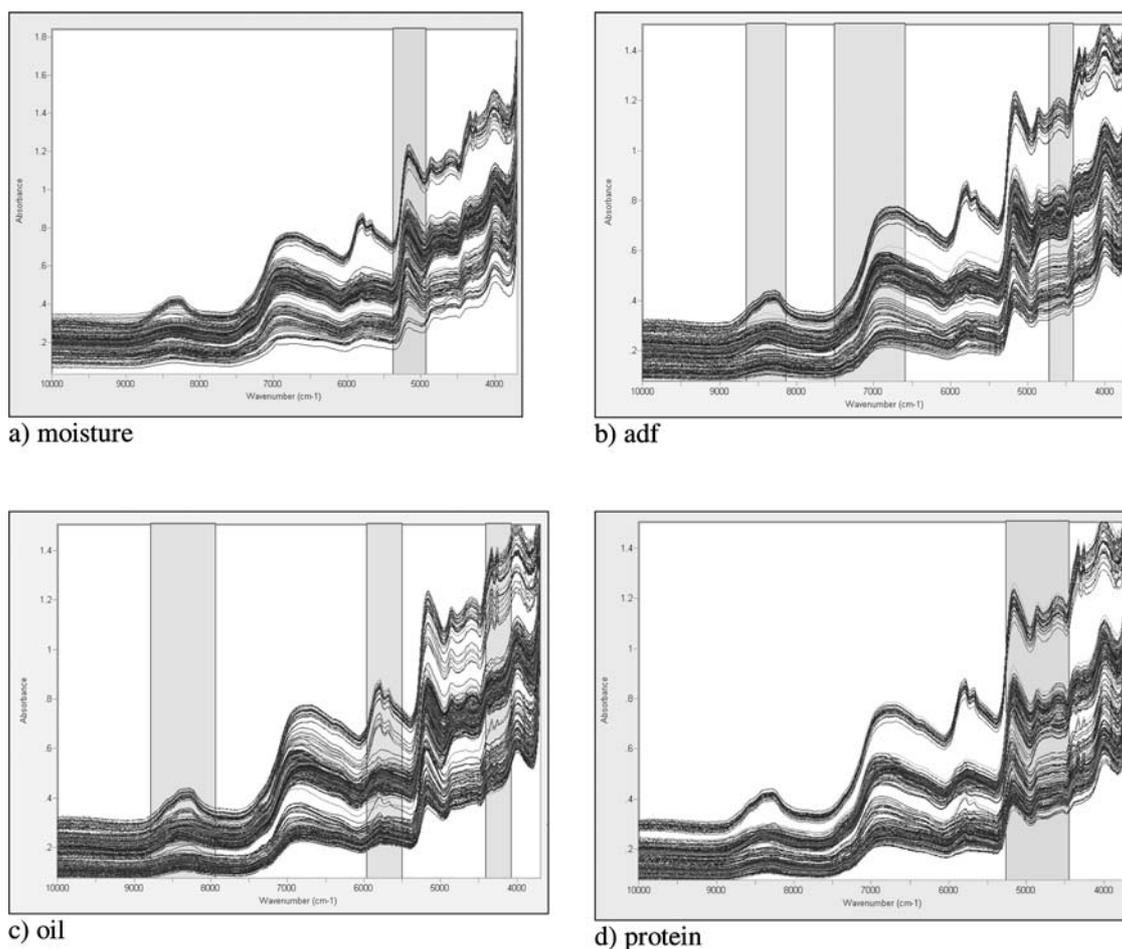


Fig. 1 Spectral zones used for calibration due to the maximum correlation coefficient between chemical and spectral data, as can be seen in software BGRAMS32[®]

both spectral and concentration values, respectively. Residual of concentration values are the difference between the NIR predicted value and that obtained after doing the calibration, excluding this sample. This is useful to refute the samples out of range, improving in this way the quality of the calibration.

Figure 2d shows the value of SECV versus the number of factors, which was chosen to minimize the error. The optimum number is usually so low that while adding a new factor, it does not significantly diminishes the SECV. Therefore, calibration was carried out with seven factors with SECV=1.63% with mathe-

tical pretreatment 1, as described in materials and methods.

Fiber content

The zones of the spectral region used to determine the adf values were 8,648–8,150; 7,498–6,599 and 4,700–4,100 cm^{-1} , where a correlation between spectral variance and chemical data variance was calculated by the software BGRAMS32[®]. A good correlation was observed by absorption due to O–H groups ($7,062 \text{ cm}^{-1}$) and the

Table 1 Statistical values related to the determination of different parameters in *Gevuina avellana* presscake, *Rosa rubiginosa* seeds and soybean meal

Parameter	Number of samples	PLS factors	SECV ^a	R ^a	Mathematical pretreatment	Interval (%) ^b
Moisture	164	7	1.63	0.928	1	5–35
ADF ^c	236	9	2.81	0.986	1	5–75
Oil	280	5	1.25	0.956	2	0–22
Protein	220	7	2.99	0.955	2	0–45

^aStandard error of cross-validation

^bValues in dry basis

^cAcid detergent fiber

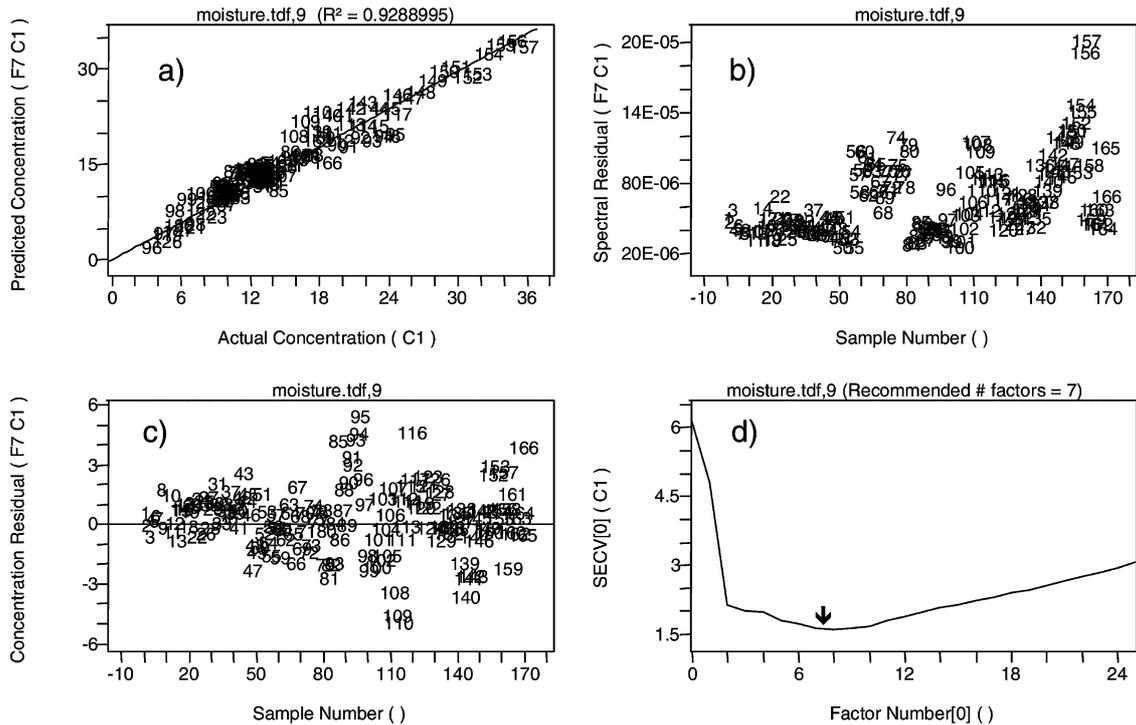


Fig. 2 Determination of moisture for *Gevuina avellana*, *Rosa rubiginosa* and *Glycine max*. **a** Multivariable regression of predicted concentration versus concentration determined by chemical meth-

ods. **b, c** Residual values of spectra and concentration values, respectively. **d** Number of factors that minimize the standard error of cross-validation

carbohydrate band ($4,800\text{ cm}^{-1}$) [18]. C–H bond absorption also promotes differences at $4,420$ and $4,230\text{ cm}^{-1}$. Therefore, the changes in fiber concentration were indicated mainly by the C–H and O–H bonds. The best calibration was obtained applying pretreatment 1, with a SECV value of 2.81, and with nine factors. Two different zones can be observed in Fig. 3a. The first zone is around 15% of the fiber content, corresponding to the hazelnut presscake and soybean meal, while another zone about 65% corresponded to mosqueta seeds. The existence of three subgroups in the latter zone show the differences in the content of lignin due to the size differentiation, because the lignaceous outer fraction of the seeds were not distributed equally among the fragments sizes. So, higher the size, higher the lignin content was. Font et al. [21] measured acid detergent fiber in brassicas, but their results could not be easily compared to those published here because they do not report the number of PLS factor used. The range was from 6 to 16%, more homogeneous than ours because it was mono-variety calibration. SECV of the intact seeds was in the range 1.81–2.34 with three outliers and of the ground seeds it was between 2.59 and 2.94 (with no outliers), similar to ours. Another way to express the fiber content is “dietary fiber”, “soluble fiber” and “non-soluble fiber”. Kays and Barton [9] used the range 1,104–2,494 nm, preprocessing data with detrending and standard normal variate (SNV), to achieve a SECV between 0.84 and 1.33% for soluble fiber, and 1.13% for insoluble fiber, using 12 PLS factors. This error was lower than that achieved in this work. Archibald and Kays [22] determined total dietary fiber on cereal food

samples, with Savitsky-Golay mathematical treatment, using the entire spectrum, except a zone between 1,370 and 1,480 nm. Their SECV was slightly lower than that obtained in this work, 2.1% using 11 factors and 1.84% with 13 factors. These results can be considered similar to ours, because there is no significant difference. Summarizing, fibers can be measured with SECV between 1.2 and 2% in mono-variety calibrations and the price paid by extending the calibration to several materials was to have SECV nearly 3%, maintaining a robust calibration and not increasing the PLS factors to avoid introducing noise in the calibration.

Oil content

Three different regions were used to carry out the calibration of oil content, as can be observed in Fig. 4. Chosen intervals were $8,775\text{--}7,949$, $5,951\text{--}5,508$ and $4,401\text{--}4,081\text{ cm}^{-1}$. Similar ranges were used to predict oil content in rapeseeds by Reinhardt et al. [23]. The minimum error (SECV=1.25) was obtained selecting pre-treatment 2 with so few factors as five. Several well-defined zones appear in Fig. 4a; the lower values corresponding to rosehip samples (up to 4%), followed by soy samples (5–13%) and lastly, the hazelnut presscake samples were those that showed the higher oil content (values near 20%). These studies have been used to classify different types of oil, related to chain length and unsaturation levels of oil acids [24, 4]. Oil content could be measured with a

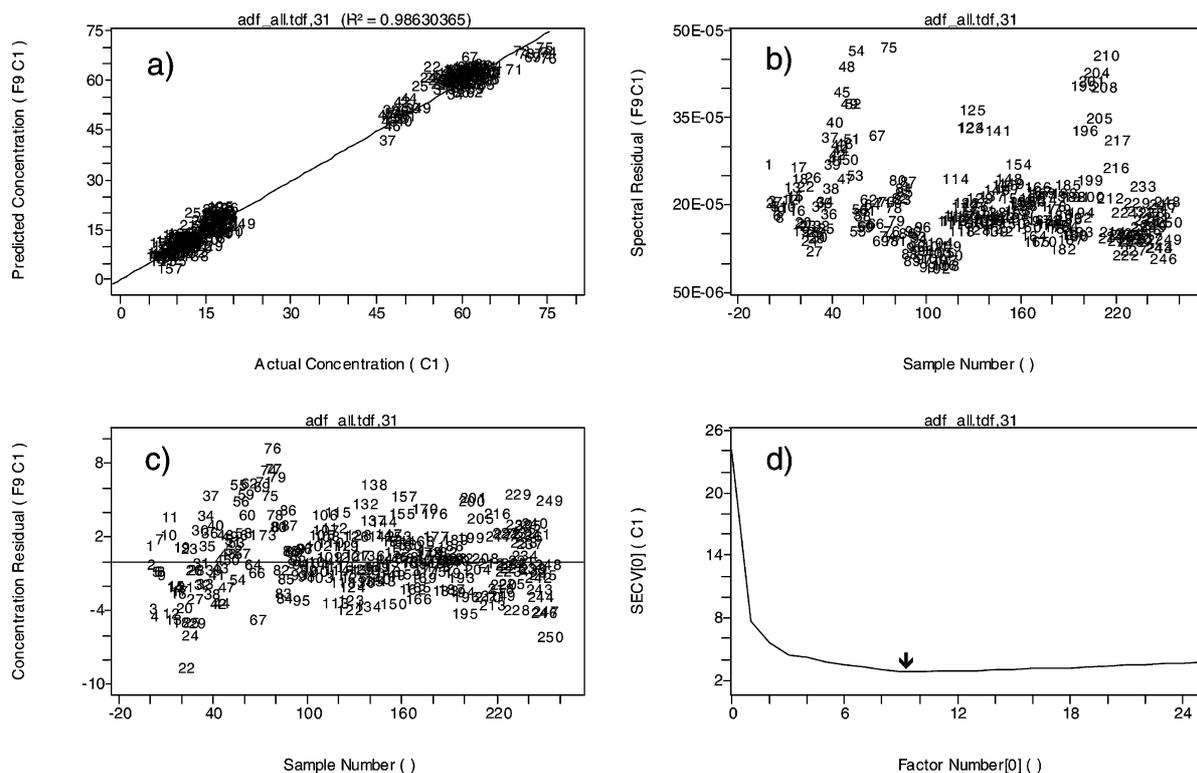


Fig. 3 Determination of acid detergent fiber (adf) of *Gevuina avellana*, *Rosa rubiginosa* and *Glycine max*. **a** Multivariable regression of predicted concentration versus actual concentration determined by chem-

ical methods. **b, c** Residual values of spectra and concentration, respectively. **d** Number of factors that minimize the standard error of cross-validation

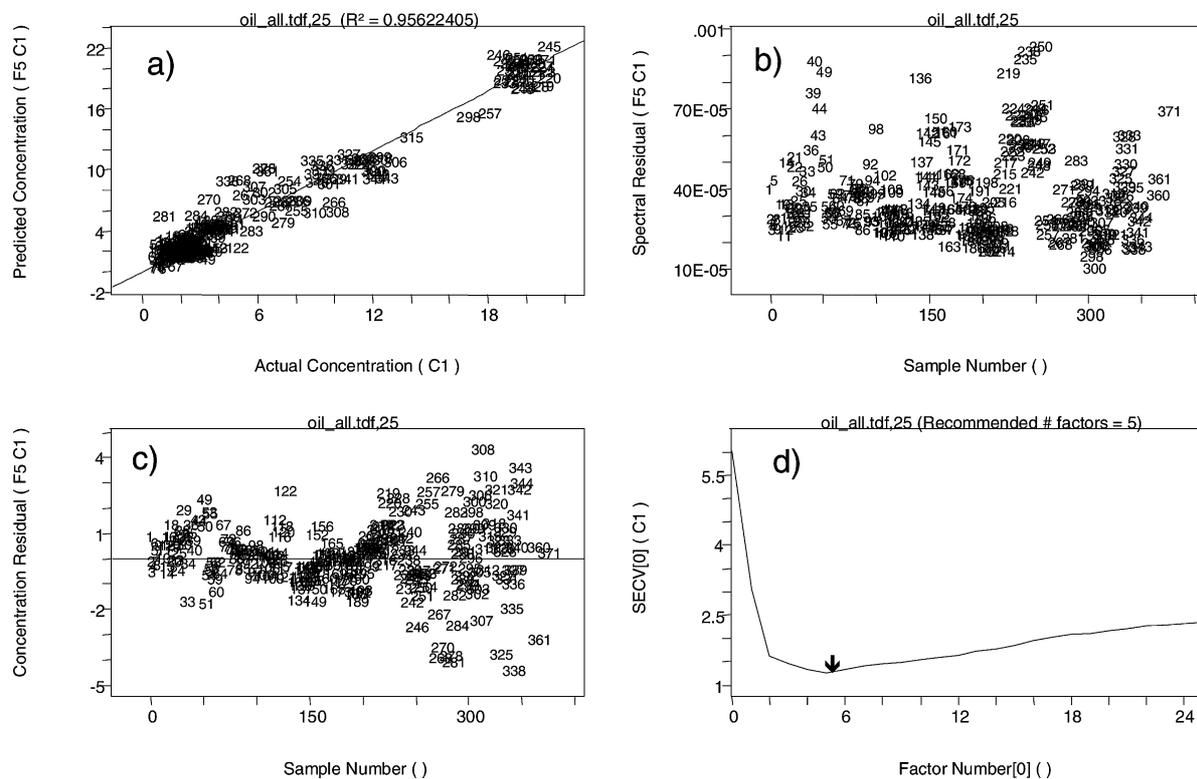


Fig. 4 Determination of oil content of *Gevuina avellana*, *Rosa rubiginosa* and *Glycine max*. **a** Multivariable regression of predicted concentration versus actual concentration determined by chemical meth-

ods. **b, c** Residual values of spectra and concentration values, respectively. **d** Number of factors that minimize the standard error of cross-validation

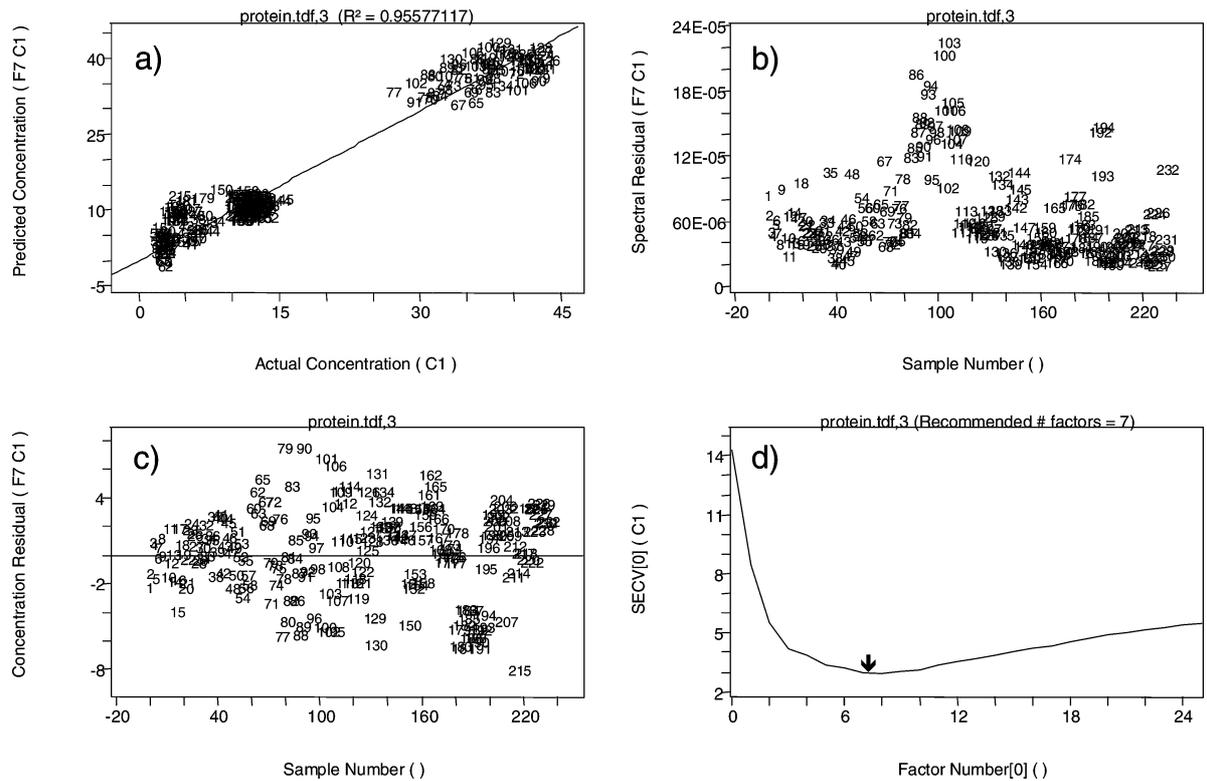


Fig. 5 Determination of protein content of *Gevuina avellana*, *Rosa rubiginosa* and *Glycine max*. **a** Multivariable regression of predicted concentration versus concentration determined by chemical meth-

ods. **b, c** Residual values of spectra and concentration values, respectively. **d** Number of factors that minimize the standard error of cross-validation

low SECV although using samples from a very different nature.

Protein

Chosen wavelength intervals to determine protein content were $5,253\text{--}4,459\text{ cm}^{-1}$, $10,000\text{--}9,000\text{ cm}^{-1}$ (corresponding to second overtone of N–H bond) and $6,896\text{--}6,250\text{ cm}^{-1}$ (first overtone of N–H) [18]. In this case, three different zones of protein content were observed (Fig. 5). Lower values corresponded to mosqueta rosehips (about 5%), followed by hazelnut (about 12%) and soy (40%). Similar results were obtained using both pretreatments, reaching a SECV of 2.99 with seven factors. This was the most difficult parameter to be correlated, mainly because nitrogen is not only present as proteins. Alkaloids, inorganic nitrogen and other molecules can be measured as total nitrogen by Kjeldahl method, giving different response in NIR. It is important to remark that the factor 6.25 is not a representative of all the proteins and the use of the same value for different species can introduce an error due to the different aminoacid profile. This effect should be negligible when using samples of only a vegetal specie. So, Azzouz et al. [25], applied NIR techniques for forages found errors 2–3 times higher for proteins than for moisture. Errors were about 0.80, but with samples of an only material (alfalfa).

Conclusions

Near-infrared diffuse reflectance spectroscopy is a valid, precise and fast technique to determine some compositional parameters in natural products. In this work, the efficiency in the determination of moisture, oil, fiber and protein content in presscake Chilean hazelnut, *Rosa mosqueta* seeds and soy meal has been proved. Among the results obtained in this study, the high fiber content of *Rosa rubiginosa* could be remarkable (around 60%) and the 40% of protein content detected in soy. SECV values obtained for a calibration with different vegetal species was below 3%, oil being the best predicted parameter and protein the most difficult, because of the presence of inorganic nitrogen and perhaps alkaloids. Fiber and moisture contents can be measured with a low SECV even using samples from a very different nature.

The use of near-infrared technique requires the utilization of specific equipment and the performance of a previous calibration before analyzing the real samples. For these reasons, it is mainly a suitable technique to carry out routine determinations in food industry. The response is much faster than for wet analysis, being especially suitable for the screening of input materials. The use in characterization of final products is limited by the standard error of determination.

Future studies could guide the establishment of a relationship between the parameters of calibration and the

structural characteristics of each new material subjected to this technique.

Acknowledgements This work was funded by the Spanish ‘Ministerio de Ciencia y Tecnología’, PPQ2000-0688-C0501.

References

1. Lee MH, Cavinato AG, Mayes DM, Rasco B (1992) *J Agric Food Chem* 40:2176–2182
2. Rodríguez-Otero JL, Hermida M, Cepeda A (1995) *J AOAC Int* 78:802–806
3. Pérez-Vich B, Velasco L, Fernández JM (1998) *J Am Oil Chemist’ Soc* 75:547–555
4. Bewig K, Clarke A, Roberts C, Unklesbay N (1994) *J Am Oil Chemist’ Soc* 71:195–200
5. Ozen BF, Mauer LJ (2002) *J Agric Food Chem* 50:3898–3901
6. Wählby U, Skjöldebrand C (2001) *J Food Eng* 47:303–312
7. Chung HM, Arnold A (1995) *Appl Spectrosc* 49:1997–1102
8. Kays SE, Windham WR, Barton I (1996) *J Agric Food Chem* 46:2266–2271
9. Kays SE, Windham WR, Barton I (2002) *J Agric Food Chem* 50:3024–3029
10. Ferrao MF, Carvalho CW, Muller EI, Davanzo CU (2004) *Cienc Tecnol Alimentos* 24:333–340
11. Franco D, Moure A, Sineiro J, Domínguez H, Núñez MJ Natural extract of *Gevuina avellana* hulls as antioxidant/UV filter for food and cosmetic uses. Spanish Patent P9902683
12. Moure A, Domínguez H, Zúñiga ME, Soto C, Chamy R (2002) *Food Chem* 74:327–339
13. Pareja B, Kehl H (1990) *Anal Real Acad Farm Chile* 56:283–293
14. Palma M, Sandoval C (1986) *Anal Real Acad Farm Chile* 52:597–612
15. Moure A, Franco D, Sineiro J, Domínguez H, Núñez MJ, Lema JM (2000) *J Agric Food Chem* 48:3890–3897
16. Hornero-Méndez D, Mínguez-Mosquera MI (2000) *J Agric Food Chem* 48:825–828
17. Van Soest PJ, Wine RH (1968) *J Assoc Off Agric Chem* 51:780–785
18. Murray I, Williams P (1987) In: Williams P, Norris K (eds) *Near-infrared technology in the agricultural and food industries*. American Association of Cereal Chemists, St Paul, Minnesota, Chapter 2, p 17
19. Osborne BG, Fearn T (1986) *Near infrared spectroscopy in food analysis*. Longman Scientific and Technical, Harlow, Essex Chapter 2, pp 37–39
20. Franco D (2002) *Processing of oilseeds: ethanolic extraction of Chilean hazelnut and Rosa mosqueta*. PhD Thesis, University Santiago de Compostela
21. Font R, del Río M, Fernández JM, de Haro A (2003) *J Agric Food Chem* 51:2917–2922
22. Archibald D, Kays S (2000) *J Agric Food Chem* 48:4477–4486
23. Reinhardt TC, Paul C, Röbbelen G (1992) In: Murray I, Cowe I (eds) *Making light work. Advances in near-infrared spectroscopy*, London, pp 1380–1384
24. Sato T (1996) *J Am Oil Chemist’ Soc* 71:293–298
25. Azzouz T, Puigdomenech A, Aragay M, Tauler R (2003) *Anal Chim Acta* 484:121–134