Non-destructive prediction of chemical composition in sunflower seeds by near infrared spectroscopy

A. Fassio[a] and D. Cozzolino,[b,c]

[a]*Summer Crop Program. INIA La Estanzuela. Colonia - Uruguay*
[b]*Animal Nutrition and NIRS Lab. INIA La Estanzuela*
[c]*Actual address: The Australian Wine Research Institute. PO Box 197. Glen Osmond, 5064. South Australia. E-mail: Daniel.Cozzolino@awri.com.au*

Introduction

Sunflower (*Helianthus annuus* L.) is one of the most widely cultivated oil crops in the world.[1] Although seed oil of standard cultivated sunflower is considered to be of good quality for edible purposes, the development of cultivars with high oil content and high oleic acid concentration is an important breeding objective for this crop.[1] Plant improvement for quality purposes depends on the ability to evaluate large number of individuals where the only way to marriage the increased variability is with more frequent analysis and sampling.[2–4] Usually such ability requires chemical and physical methods of analysis that are destructive, slow and expensive in chemical reagents and labour.[4] One of the mayor concerns of oil chemists is the time consuming method available for routine analysis of grains and seeds.[5] The objectives in sunflower breeding vary with specific programs but generally emphasise high seed yield and oil content.[6] The oil extracted from the seed contributes about 80 per cent of the total value of the oilseed sunflower crop, depending on both the percentages of hull and oil concentration in the kernel. About two thirds of the increase in achene oil content from breeding and selection has resulted from a reduction in hull percentage, and about one-third from an increase in kernel oil content. The rate of increase in improving oil content has been reduced in recent years and some concern has been expressed that oil content in seeds may be approaching a biological limit. However, it appears that it is feasible to develop lines and hybrids possessing over 550 to 600 g kg^{-1} oil and most breeders believe that selecting for higher oil content is still a very important and realistic objective.[7] Other breeding objectives for oil are related to oil quality, especially an increase in oleic and linoleic acid content, in order to improve the nutritional value of sunflower oil. The availability of rapid, non-destructive methods to evaluate seed quality traits is one of the most important factors determining the success of plant breeding projects.[7] The near infrared (NIR) region spans the wavelength range 780–2500 nm, in which absorption bands correspond mainly to overtones and combinations of fundamental vibrations.[8,9] NIR relies on calibrations which utilise absorbances at several, even many wavelengths, to predict the composition of a samples.[8] Reports were found in the literature related to the determination of both oil content and fatty acid composition in sunflower seeds by NIR.[7,10] The objective of this study was to evaluate the potential of NIRS to predict oil, protein and moisture content as no-destructive technique in pre-screening of sunflower seed samples for breeding purposes in Uruguay.

Near Infrared Spectroscopy: Proceedings of the 11th International Conference.

Materials and Methods

Three hundred (n: 300) sunflower (*Helianthus annuus* L.) samples obtained from several breeding lines (individual plants) from the Summer Crop Breeding Program at INIA La Estanzuela were used for the development of the NIRS calibration models. All the samples were grown at INIA La Estanzuela (Uruguay, South America) in different years from 1997 to 2000. Previous to oil analysis samples were cleaned, oven-dried (40°C for 4 h) and ground in a Cyclotec mill 1 mm (Foss Tecator, Sweden). For oil analysis, samples were taken immediately after grinding and analysed by Soxhlet extraction using petroleum ether as solvent for 6 h and corrected for moisture.[11] Moisture was determinate by oven drying the sample at 105°C for 16 hours to constant weight.[11] Nitrogen was determined using a semi-micro automated Kjeldhal method (Tecator, Sweden) and converted to CP using the factor 6.25.[11] Oil, protein and moisture content were expressed as g kg^{-1} on dry weight basis and performed in duplicate. Intact samples (about six to ten achenes) were scanned in the visible and near infrared region in reflectance mode (400–2500 nm) on a scanning monochromator instrument NIRS 6500 (NIRSystems, Silver Spring, MD, USA) and placed in a small ring cup (50 mm diameter; 20 mm depth) (Part number IH-0307, NIRSystems, USA). Reflectance data were stored as the logarithm of reciprocal of reflectance [log (1/*R*)] at 2 nm intervals, collecting 1050 data points. Spectra data collection, manipulation and calibrations were developed using the NIRS 2 version 3.0 (Infrasoft international, Port Matilda, USA). Equations were developed using modified partial least squares (MPLS) regression with internal cross-validation.[12] Two scatter corrections were used standard normal variate (SNV) and detrend and no scatter correction (log 1/*R*).[12,13] The SNV and detrend scatter correction is designed to remove additive baseline and multiplicative signal effects resulting in a spectrum with zero mean and a variance equal to one. These algorithm enhance the differences in spectra related to the chemical composition of samples by reducing differences in spectra related to physical characteristics of the sample (primarily particle size). Two mathematical treatments were applied 1,5,5 and 2,5,5. The first number indicates the order of derivative (one is first derivative of log 1/*R*), the second number is the gap in data points over which the derivative is calculated; the third number is the number of data points used in the first smoothing. Calibration statistics calculated include the standard error of calibration (*SEC*), the coefficient of determination in calibration (R^2_{cal}), the standard error of cross-validation (*SECV*) and the coefficient of determination in cross validation (R^2_{val}).[12] The optimum calibrations were selected on the basis of minimising the *SECV*. This error was calculated by an internal validation of 33 per cent of samples randomly taken by the software routine, which was predicted by an equation based on a calibration with the remaining 66 per cent of all samples. The outlier elimination pass was set to allow the computer program to remove outliers twice before completing the final calibration. Two outlier detection provided by the ISI software were applied: t and H. The t statistics outliers, which have residuals from the reference analysis of greater than 2.5 times *SEC*, are samples whose reference analysis is in doubt. These should be re-analysed by the reference method. The H statistics outliers are samples whose spectra are atypical of all the others that make up the calibration set. They may not belong in the population. The *SD*/*SECV* relationship (standard deviation of the constituent data / standard error of cross validation) was calculated to evaluate the performance of the calibrations.[9] Each calibration was tested using an independent set of samples (n = 50). The performance of the validation set was tested based on the standard error of prediction (*SEP*) and the coefficient of correlation (r).

Results and Discussion

Table 1 and 2 show the NIR calibration and cross-validation statistics for oil and protein respectively. Table 3 shows the NIR prediction statistics for oil in the intact sunflower samples

Table 1. Near infrared calibration statistics for oil in whole sunflower seeds (g kg^{-1} DM).

	Wavelength segment nm	*N*	*SD*	R^2_{cal}	*SEC*	R^2_{val}	*SECV*
1,5,5 -SNV	400–2500	216	42.8	0.82	18.2	0.70	24.8
1,5,5 -SNV	400–750	231	44.0	0.50	32.7	0.30	38.3
1,5,5 -SNV	1100–2500	211	40.2	0.77	19.0	0.65	24.7
2,5,5 -SNV	400–2500	219	43.5	0.90	15.4	0.80	22.3
2,5,5 -SNV	400–750	218	40.2	0.60	26.5	0.50	30.3
2,5,5 -SNV	1100–2500	220	42.8	0.82	18.1	0.75	23.6
1,5,5 None	400–2500	225	43.3	0.76	21.2	0.60	27.7
2,5,5 None	400–2500	217	42.5	0.86	16.0	0.75	22.3

SD: standard deviation; R^2_{cal}. coefficient of determination in calibration; *SEC*: standard error in calibration; R^2_{val}: coefficient of determination in cross-validation; *n*: number the samples used to perform the calibration; *SECV*: standard error of cross validation

Table 2. Near infrared calibration statistics for protein in whole sunflower seeds (g kg^{-1} DM).

	Wavelength segment nm	*N*	*SD*	R^2_{cal}	*SEC*	R^2_{val}	*SECV*
1,5,5 -SNV	400–2500	141	35.1	0.90	11.4	0.82	15.0
1,5,5 -SNV	400–750	144	35.5	0.65	20.9	0.60	23.3
1,5,5 -SNV	1100–2500	137	34.6	0.90	11.1	0.82	14.8
2,5,5 -SNV	400–2500	132	34.6	0.96	6.6	0.90	13.1
2,5,5 -SNV	400–750	144	37.3	0.77	18.3	0.62	23.1
2,5,5 -SNV	1100–2500	144	35.5	0.90	11.5	0.81	15.3
1,5,5 None	400–2500	138	35.0	0.88	12.1	0.70	19.0
2,5,5 None	400–2500	136	35.0	0.92	9.8	0.85	14.9

R^2_{cal}. coefficient of determination in calibration; *SEC*: standard error in calibration; 1-*VR*: coefficient of determination in cross-validation; *n*: number the samples used to perform the calibration; *SECV*: standard error of cross validation, *SEP*: standard error of prediction

Table 3. Near infrared prediction statistics for oil in whole sunflower seeds (g kg^{-1} DM).

	Wavelength segment nm	*r*	*SEP*	Slope	Bias
1,5,5-SNV	400–2500	0.34	44.0	0.75	0.45
1,5,5-SNV	400–750	0.40	41.0	1.20	–0.30
1,5,5-SNV	1100–2500	0.30	46.1	0.76	0.43
2,5,5-SNV	400–2500	0.75	26.2	1.06	0.02
2,5,5-SNV	400–750	0.33	44.2	1.18	–0.34
2,5,5-SNV	1100–2500	0.60	28.5	0.96	0.60
1,5,5 None	400–2500	0.60	27.2	1.03	–0.27
2,5,5 None	400–2500	0.60	28.1	1.04	–0.11

r: coefficient of correlation; *SEP*: standard error of prediction

using two mathematical treatments and two scatter corrections. The performance of the calibration for the chemical constituents analysed were evaluated using the *SD*/*SECV* ratio. When the error in calibration (*SEC* or *SECV*) exceeds one third of the SD of the population, regression models can be misleading.[9] In decreasing order of performance the results were for protein, moisture and oil, respectively. The concentration of oil, protein and moisture in all the samples were found to be comparable to reported data in the literature.[1] Mainly the CH overtones and combinations bonds associated with the fatty acid carbon chain and *cis*-unsaturation in sunflower were observed in the NIR mean spectrum (data not presented).[14] The inclusion of both the visible and NIR regions improved the calibration statistics, but did not give good calibrations if they were used alone. The best coefficient of determination in calibration for moisture (R^2_{cal}) was 0.95 (*SECV*: 3.3) on g kg^{-1}; for protein, the best R^2_{cal} was 0.96 *(SECV*: 13.1) and for oil, the best R^2_{cal} was 0.90 (*SECV*: 22.3) in g kg^{-1} on a dry weight basis. From the results obtained it is clearly that NIR could be sensitive to variations in the colour of the individual achenes (from black to different strips). This could affect the calibration statistics and suggested by previous authors where high NIR correlation was obtained for husked samples compared with those obtained for intact samples.[10] The results in this study also showed that high and low oil content absolute values were not well predicted by the NIR calibration models. High oil content as well as old samples (harvested in 1997 and 1998) had a different spectral position compared with the rest of the samples (data not presented). This indicated that not only the amount of oil in the sample (high or low) affects the calibration statistics for this parameter, but also the storage conditions of the sample and the year of growth (for example, humidity and temperature). Factors such as oven temperature, time of drying and vaporisation of substances other than water and biochemical reactions are critical factors that influence the accuracy and precision of the predictions of chemical composition in grains. Storage of the sample without either control of temperature or humidity explained the poorest prediction statistics for moisture on the intact sunflower seeds. The advantages of evaluating either individual or groups of seeds in sunflower breeding for oil concentration outweigh the losses of in accuracy.[15] NIR has been criticised because there are errors associated with the predicted value for chemical composition. Errors are defined as the difference between the measured values and the true value. The truth is that all method have errors associated with them, including the traditional methods used for reference.[4] Since NIR is a predictive method, it inherently has built-in errors associated with the reference method plus other errors that could occur (sample presentation, sampling, scatter effects). NIR has proved that it is capable of producing repeatable results for breeding purposes. Although not high coefficients of determination were obtained for oil ($R^2_{cal} < 0.90$), reliable equations were developed relating to breeding objectives. The level of accuracy obtained in this study is probably adequate for the selection of genotypes according to low, medium and high oil content, independently of the absolute value. Our results show that NIR can be used to estimate oil, protein and moisture content in intact sunflower seed samples. This has special relevance to plant breeding programmes, where the analysis of a large number of entries (varieties) is often necessary in a short period of time. Although the *SEP* obtained for oil are considered high, the use of NIR for intact sunflower samples could be used for pre-screening purposes to pre-select different varieties to be further analysed by reference methods if the absolute value is needed. NIR would reduce the number of more expensive, destructive and time-consuming analysis such as Soxhlet extraction.

Conclusions

The results of this work demonstrate that NIR analysis of intact sunflower seeds for oil, protein and moisture content were sufficiently reliable for screening sunflower samples breeding purposes. NIR is a cost effective technique for plant breeding since it is very versatile, usually requires

minimal sample preparation, permits a large number of samples to be analysed per day and may produce results for many traits simultaneously. Re-calibration may be necessary for moisture and oil over the next years, especially as these parameters are susceptible to change with climatic or agronomic conditions.

References

1. Z. Flagella, T. Rotunno, E. Tarantino, R. Di Caterina, and A. De Caro, *Eur. J. Agron*. **17,** 221 (2002).
2. J. Fernandez-Martinez, J. Muñoz and J. Gomez-Arnau, *Crop. Sci*. **33,** 1158 (1993).
1. I. Murray, in *The value of traditional analytical methods and near infrared spectroscopy to the feed industry. In: Recent Advances in Animal Nutrition*. Ed by P.C. Gansworthy, J. Wiseman and W. Haresign. Butterworth Publishers, UK, 294 pp. (1996).
2. J.A. Panford and J.M. de Man, *J. Am. Oil. Chem. Soc*. 67, 473 (1990).
3. G.N. Fick and J.F. Miller, "Sunflower breeding", in, *Sunflower Technology and Production*, Agronomy Monograph No.35, Ed by Albert A. Schneiter. A.A. ASA; CSSA; SSSA. Madison, USA, pp. 395–439 (1997).
4. L. Velasco, B. Perez-Vich and J.M. Fernandez-Martinez, *Crop. Sci*. **39,** 219 (1999).
4. G.D. Batten, *Aust. J. Exp. Agr*. **38,** 697 (1998).
5. I. Murray, "Forage analysis by near infrared spectroscopy, in *Sward Herbage Measurement Handbook*, Ed by A. Davies, R.D. Baker, S.A. Grant and A.S. Laidlaw. British Grassland Society, Reading, UK, pp. 285–312 (1993).
6. B. Perez-Vich, L. Velasco and J.M. Fernandez-Martinez, *J. Am. Oil. Chem. Soc*. **75,** 545 (1998).
7. *Official methods of analysis of the Association of Official Analytical Chemists*, 15th edition, Ed by K. Helrich. Official Analytical Chemists, Inc., Arlington, Virginia, USA (1990).
8. J.S. Shenk and M.O. Westerhaus, *Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy*. Infrasoft International, Port Matilda, PA, USA (1993).
9. R.J. Barnes, M.S. Dhanoa and S.J. Lister, *Appl. Spectrosc*. **43,** 772 (1989).
10. I. Murray, "The NIR spectra of homologous series of organic compounds", in *Proceedings of NIR/NIT Conference*, Ed by J. Hollo, K.J. Kaffka and J.L. Gonczy. Akademiai Kiado, Budapest, Hungary, pp. 13–28 (1986).