Near-Infrared Reflectance Measurement of Moisture, Protein and Oil Content of Ground Crambe Seed

R.A. Hartwig* and **Charles R. Hurburgh, Jr.**
Agricultural Engineering Department, Iowa State University, Ames, Iowa 50011

Crambe moisture, protein and oil percentages were predicted by fixed-filter near-infrared reflectance with standard errors of prediction (SEP) of 0.26, 0.59 and 0.86 percentage points, respectively. Crambe had a large range of protein and oil percentages, 17.4%–25.0% and 17.7%–36.4% respectively. Calibration samples were selected on the basis of relative spectral data, with no advance knowledge of protein and oil content. This procedure selected samples representing the full range of constituent values, and resulted in calibrations that had lower SEP's than standard errors of calibration.

Crambe (*Crambe abyssinica* Hochst. Ex. R.E. Fries) is regarded as a promising new industrial oil seed. Crambe oil is about 55–60% erucic acid (*cis*-13-docosenoic) (1). When treated with ozone, erucic acid gives brassylic and pelargonic acids. Brassylic acid can be used for polyesters, plasticizers, alkyd resins, lubricants, rubber additives and surface-active agents. Pelargonic acid is used for plasticizers, alkyd resins, vinyl stabilizers, hydrotropic salts, pharmaceuticals, synthetic flavors and odors, flotation agents and insect repellents (2).

The oil-free, hull-free meal contains 46%–58% crude protein with a good amino-acid balance for feed. The meal will contain thioglucosides, which places some restrictions on its use as a feedstuff.

Agronomic research in crambe breeding gave rise to a need for nutrient analysis of crambe seed. Previous success with soybeans (3) and other oilseeds (4) suggested that near-infrared reflectance (NIR) could be used in lieu of wet chemical methods. However, crambe seed is light and small, conditions certain to complicate grinding.

The multiple-linear-regression calibration of fixed-wavelength near-infrared reflectance analyzers has been reported extensively in the literature. Little has been written about the actual selection process of calibration and prediction samples. Typically, a number of samples are collected and randomly divided into two groups—one for calibration and one for prediction (5,6), or samples are compared against those of previously known chemical composition (4,7). Barton and Cavanagh (8) used two stepwise linear-regression analysis programs, which allowed every i[th] sample to be saved for prediction verification.

Selection of calibration samples by spectra alone has been reported for near-infrared monochramotors (9) but not for fixed-filter units.

The range of protein and oil values for crambe was unknown. Instead of running wet chemistry on all our available samples, we wanted to select a spectrally representative calibration set by using only reflectance

values from the 10 filters. This would save considerable effort and laboratory expense, but such a procedure has not been previously reported.

Therefore, the research had two objectives: i) to develop and validate a near-infrared calibration for crambeseed, and ii) to test a calibration sample selection procedure based on spectral information rather than on previous estimates of composition.

MATERIALS AND METHODS

A group of 165 samples of unknown varieties was available for calibration. Samples had been grown in Iowa in the 1986 growing season. Initially, all 165 samples were ground in a Magic Mill III[+] flour mill. Spectral data from a 10-filter Dickey-john Instalab 800 NIR were collected for three subsamples of the ground material. Triplicate air-oven moisture determinations were made on each ground sample. The moisture determinations were done by drying for 1 hr at 130°C, the official method for soybeans (10). There is no standard air-oven moisture method directly applicable to crambe.

The NIR output consisted of 10 "log numbers," L0 to L9 (415.25 * log [1/R]). The average log values of the three subsamples were used for calibration.

Selection of calibration samples. The calibration sample selection was based upon two sorting criteria, the moisture content determined at the time of the NIR reading and spectral significance as measured by relative log (1/R) values. The average ($\bar{x}$) and standard deviation (sd) of the moisture content was 5.24% and 0.48 percentage points, respectively.

High, medium and low moisture categories were determined on the basis of 0.5 sd being added to or subtracted from the average value. This placed an approximately equal number of samples in each moisture category. If prior estimates of composition (protein and oil) had been available, a complete matrix of samples could have been generated, but we had no such prior estimates, only spectral data.

Across the near-infrared spectrum, there are wavelengths typically unaffected by composition. Their main source of variation is from particle size differences. Filter 5 (1680 nm) is such a wavelength. The average L5 value was 100.87, with a standard deviation of 8.55. Again, high, medium, and low categories, this time of L5, were determined based upon x ± .5 sd. Because the Magic Mill grinder has no screens, there was no constraint on fineness of grind. In a multiple-linear calibration, it is important to include the maximum anticipated particle size variation in the calibration set (11).

Next, a wavelength associated with oil (2230 nm), and a wavelength associated with protein (2310 nm) were identified on the basis of our current soybean calibration (3). Log values in the two wavelengths (L_0 and L_1) were regressed against 1680 nm (L_5) values by

*To whom correspondence should be addressed.

sample. This yielded the two equations:

$$L_0 = 150.8311 + 1.0818 * L_5 \qquad [1]$$

$$R^2 = .447$$

$$RMSE* = 3.971$$

$$L_1 = 108.7248 + 0.6875 * L_5 \qquad [2]$$

$$R^2 = .486$$

$$RMSE* = 3.406$$

*Root mean square error.

The important spectral information was now contained in the residuals (deviations) from Equations [1] and [2]. Even though both equations were significant at the 0.001 probability level, the low R^2 and high RMSE showed that there was considerable information within these residuals. Assuming that the largest positive and negative residuals would indicate the extremes of composition, the five samples with the highest positive sum, and the five with the lowest negative sum of 2310 nm and 2230 nm residuals were chosen within each combination of high, medium and low of 1680 nm and moisture. This gave 60 samples for chemical analysis. The practical consequence of this selection method was to identify spectrally significant samples at all levels of moisture and grinding consistency.

Chemical analysis. We ran triplicate ground grain moisture determinations, duplicate Kjeldahl nitrogen determinations and duplicate Butt-tube ether oil extractions on each calibration sample. The Kjeldahl procedure, AACC 46-11A, (12), was a 70-min digestion with 2 CT-50 Kjeltab catalyst tablets (5.0 g K_2SO_4 and 0.15 g $CuSO_4$) (Alfie Packers, Inc., Omaha, Nebraska). A 4% aqueous boric acid solution (instead of the 0.1N H_2SO_4) was used to collect ammonia in the receiving flask. The AACC method was used to avoid the mercury catalyst specified in AOCS methods.

The Goldfisch oil extraction method, AOAC 14.084-14.085 (13), was used with a Labconco Butt-tube extractor (Labconco Corp., Kansas City, Missouri). This procedure uses a 5-hr petroleum ether (35°C to 60°C boiling range) extraction. Moisture determinations (10) were made at the time of crude protein and crude fat analyses so that chemical values could be adjusted to the moisture determined at the time of NIR analysis.

Statistical analysis and validation. A multiple linear regression with maximum R^2 improvement was used to determine constants for the calibration equation. No filter was included that did not have significance above the 0.05 probability level. The R^2 and the standard error of calibration (SEC, the root mean square error of the calibration regression equation) were calculated for all possible filter combinations.

Calibrations were validated by analyzing an additional 25 samples. Bias and standard error of prediction (SEP) were calculated.

RESULTS AND DISCUSSION

Table 1 summarizes the calibration and validation results.

The ranges for the protein and oil constituents, 7.66 and 18.63 percentage points, respectively, were considerably larger than originally expected on the basis of our experience with soybeans. The large spread led to a relatively large SEC in both protein (0.70 percentage points) and oil (1.62 percentage points).

The biases for moisture, protein and oil from the validation set were nonsignificant. The standard er-

TABLE 1

Crambe Calibration for a Dickey-john Instalab 800 Analyzer

Wavelength, nm	Constituent		
	Moisture	Protein[a]	Oil[a]
	1940, 1680, 1759	2180, 2100, 1940	2310, 2100, 1940, 1680, 1759
Calibration samples			
Number used	60	60	60
Average (%)	6.16	21.38	28.50
Maximum (%)	8.21	25.01	36.37
Minimum (%)	5.05	17.35	17.74
Calibration regression			
R^2 (%)	91.1	89.3	93.3
SEC (percentage points)	0.15	0.7	1.62
Coefficient of variation (%)	2.9	3.3	5.7
Validation samples			
Number	25	25	25
Bias (percentage points)	0.01[b]	0.08[b]	0.39[b]
SEP (percentage points)	0.26	0.59	0.86

[a]"As is" moisture basis.
[b]NS = not significantly different from 0.0 (P=0.05).

rors of prediction (SEP) were 0.26, 0.59 and 0.86 for moisture, protein and oil, respectively. No slope was evident. Protein and oil had larger SEP's than expected, from our experience with soybeans, but considering the wide range of values, the SEP's were acceptable.

The SEPs were also lower than the SECs. This shows that the spectral selection procedure gave a robust calibration. It also points out the danger of being overly concerned with calibration equation statistics. Performance on unknown samples is the key indicator of calibration effectiveness.

Clearly the spectral selection procedure gave an acceptable calibration. It also distributed the calibration samples over the full range of constituent percentages. A 3 × 3 × 3 matrix of high, medium and low categories for moisture, protein and oil was set up by using the NIR-predicted values for all 160 samples. The 60 calibration samples represented all 27 cells in this matrix.

In summary, the study suggests the following conclusions: i) Crambe moisture, protein and oil contents can be predicted by near-infrared reflectance, ii) Crambe has a larger range of constituent percentages than most other grains which created relatively high SEPs of 0.59 and 0.86 percentage points for protein and oil, respectively; and iii) Selection of calibration samples from spectral data was successful because the calibration had no bias in validation, the SEP was less than the SEC, and the full range of constituent values was encompassed.

REFERENCES

1. Lessman, K.J., and W.P. Anderson, *Crambe. New Sources of Fats and Oils*, edited by E.H. Pryde, L.H. Princen and K.D. Muherjee, American Oil Chemists' Society, Champaign, IL, 1981, pp. 223–246.
2. Bruun, J.H., and J.R. Matchett, *J. Am. Oil Chem. Soc. 40*:1 (1963).
3. Hurburgh, C.R., Jr., L.N. Paynter and S.G. Schmitt, *Appl. Eng. Agric. 3(2)*:159 (1987).
4. Panford, J.A., P.C. Williams and T.M. deMan, *J. Am. Oil Chem. Soc. 65*:1627 (1988).
5. Rubentmaler, G.L., and Y. Pomeranz, *Cereal Chem. 64*:407 (1987).
6. Tkachuk, R., F.D. Kuzina and R.D. Reichert, *Ibid. 64*:418 (1987).
7. McGuire, C.F., *Ibid. 63*:155 (1986).
8. Barton, F.E., II, and A.C. Cavanagh, *J. Am. Oil Chem. Soc. 65*:768 (1988).
9. Schenk, J., M.O. Westerhaus and W.C. Templeton, Jr., *Crop Science 25*:159 (1985).
10. *Federal Grain Inspection Service Moisture Handbook*, United States Department of Agriculture, Washington, D.C. (1986).
11. Hurburgh, C.R., Jr., *Cereal Foods World 33(6)*:503 (1988).
12. *American Association of Cereal Chemists Approved Methods*, AACC, 1983, Method 46-11A.
13. *Official Chemists Official Methods*, AOAC, 1984, Method 14.084-14.085.