Measurement of Soybean Fatty Acids by Near-Infrared Spectroscopy: Linear and Nonlinear Calibration Methods

Igor V. Kovalenko*, Glen R. Rippke, and Charles R. Hurburgh

Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa 50011

ABSTRACT: A key element of successful development of new soybean cultivars is availability of inexpensive and rapid methods for measurement of FA in seeds. Published research demonstrated applicability of NIR spectroscopy for FA profiling in oilseeds. The objectives of this study were to investigate the applicability of NIR spectroscopy for measurement of FA in whole soybeans and compare performance of calibration methods. Equations were developed using partial least squares (PLS), artificial neural networks (ANN), and support vector machines (SVM) regression methods. Validation results demonstrated that (i) equations for total saturates had the highest predictive ability (r^2 = 0.91–0.94) and were usable for quality assurance applications, (ii) palmitic acid models (r^2 = 0.80–0.84) were usable for certain research applications, and (iii) equations for stearic (r^2 = 0.49–0.68), oleic (r^2 = 0.76–0.81), linoleic (r^2 = 0.73–0.76), and linolenic (r^2 = 0.67–0.74) acids could be used for sample screening. The SVM models produced significantly more accurate predictions than those developed with PLS. ANN calibrations were not different from the other two methods. Reduction in the number of calibration samples reduced predictive ability of all equations. The rate of performance degradation of SVM models with sample reduction was the lowest.

Paper no. J11261 in *JAOCS 83*, 421–427 (May 2006).

KEY WORDS: Artificial neural networks (ANN), chemometrics, fatty acids, near-infrared (NIR) spectroscopy, partial least squares (PLS), soybeans, support vector machines (SVM).

Improvement of nutritional and/or functional properties of soybean oil by modification of soy FA composition is a current objective of plant breeders (1). Several directions have been taken to meet needs of different end users. Reduction of levels of PUFA (particularly linolenic acid) and increase of oleic acid concentration improves oxidative stability of soybean oil during storage and processing. This avoids the hydrogenation process that results in increased concentrations of unhealthy *trans*-FA (2,3). In contrast, soybean varieties with high levels of saturated FA (palmitic and stearic) can be important for production of margarine and shortening (4,5).

Regardless of the target, identification and tracking of traits is a major element of plant breeding. Therefore, availability of inexpensive and rapid methods for determination of FA composition of seed samples is a key element of success for development of new grain cultivars.

Research papers published over the last decade demonstrated applicability of NIR spectroscopy for FA profiling in oilseeds. Calibration models for single rapeseeds developed by Velasco *et al*. (6) demonstrated comparatively close relationships between GLC measurements and those of NIR spectroscopy for oleic ($r^2 = 0.85$, calibration set size $n = 530$) and erucic ($r^2 = 0.88$, $n = 219$) FA. However, no reliable correlation existed for linoleic ($r^2 = 0.56$, $n = 530$) and linolenic ($r^2 = 0.53$, $n = 530$) acids. An earlier experiment with bulk rapeseeds conducted by Velasco and Becker (7) resulted in excellent cross-validation results for oleic, linoleic, linolenic, and erucic acids (r^2 = 0.95–0.98, $n = 220$). In contrast, determination coefficients for palmitic, stearic, and eicosenoic acids in bulk rapeseeds were not as high: 0.76, 0.62, and 0.69, respectively (all $n = 220$). Studies by Sato *et al*. (8,9), Velasco *et al*. (10), and Perez-Vich *et al*. (11) provide other examples of application of NIR spectroscopy for determination of FA concentrations in oil-bearing crops such as rapeseeds and sunflower seeds.

In soybeans, the predictive ability of NIR spectroscopy for FA analysis is not well documented. Dyer and Feng (12) reported SE of performance of 2.2% for oleic acid and 1.8% for stearic acid calibrations for relative concentrations. Pazdernik *et al*. (13) reported determination coefficients (validation) of 0.38–0.71 and 0.18–0.56 ($n = 90$) for five FA of ground and whole soybean samples, respectively, using one NIR spectrometer, FOSS NIRSystems 6500. The objectives of this study were (i) to investigate further the applicability of readily available NIR units for analysis of FA composition in whole soybeans, and (ii) to compare one linear and two nonlinear calibration methods for this application.

EXPERIMENTAL PROCEDURES

Raw data. A pool of approximately 1,400 soybean samples (U.S. crops of 1991, 1993–1998, and 2003) with FA profiles was used in this study. Whole soy samples were scanned on three FOSS Infratec spectrometers, Infratec Grain Analyzers 1225, 1229, and 1241 (FOSS Group, www.foss.dk). A common calibration database consisting of 4,144 scans (≈1,400 samples times 3 spectrometers) was created. It was important to focus on instrumentation already in use because there was a short-term need to support breeding programs for reduced linolenic acid. The Infratec was used because the majority of commercial units currently availiable are Infratecs, and this instrument was as effective as any other unit in a previous amino acid study (14). Relative concentrations of total saturates

---

*To whom correspondence should be addressed at Department of Agricultural and Biosystems Engineering, Iowa State University, 1551 Food Sciences Bldg., Ames, IA 50011. E-mail: igorkov@iastate.edu.

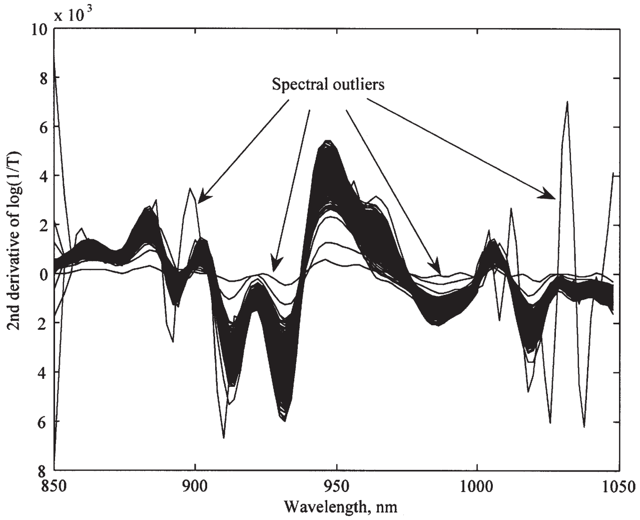

FIG. 1. Detection of spectral outliers: Samples with variation of the second derivative of log(1/*T*) beyond ±3 SD at 940 and 1020 nm were removed from calibration and validation data sets.

(palmitic plus stearic), palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2), and linolenic (C18:3) acids were determined in Department of Agronomy at Iowa State University using the GC method described by Hammond (15). FA composition was expressed in relative concentrations (as opposed to absolute) because that is what the marketplace currently requires.

Selection of samples and data preprocessing. To reduce computational load and remove redundant information from the plane of reference values, the original calibration database was resampled and a new, approximately uniformly distributed (based on FA concentrations) data subset was created for each FA. All instruments received approximately equal representation in these new data sets. In addition, samples with abnormally low or high variation of the second-derivative of the NIR signal were considered spectral outliers (Fig. 1) and were excluded from further calculations. Reference data statistics for the six final calibration/validation subsets are provided in Table 1. Seventy-five percent of the samples (randomly selected) from each subset were used for calibration and the other 25% were reserved for model validation.

NIR spectra were corrected for scatter effects by estimating their second derivative using Savitzky-Golay algorithm (5-point window and 3rd-order polynomial). In addition, samples' spectral and reference data (rows) were normalized to have zero mean and unity SD.

Calibration procedures: Theory. One linear (partial least squares) and two nonlinear (artificial neural networks and support vector machines) regression methods were used. A brief description is provided below.

(i) Partial least squares (PLS). The principle behind the PLS algorithm is to extract the important information from variation of both optical (***X***) and reference chemical composition (***y***) data and compress it in a set of new independent latent variables. The prediction equation becomes

$$\hat{y} = f(\mathbf{w}, \mathbf{s}) = w_0 + w_1 s_1 + w_2 s_2 + \cdots + w_{(p-1)} s_{(p-1)} + w_p s_p \qquad [1]$$

where $\hat{y}$ is predicted concentration, ***w*** is a vector of weights (regression coefficients), ***s*** is a vector of new independent variables, and *p* is the number of latent variables. The elements of ***s*** are defined as successive linear combinations of those original variables (wavelengths) that have the greatest covariance with optical data. The optimal number of latent variables is usually found by minimizing cross-validation standard error.

(ii) Artificial neural networks (ANN). The ANN modeling technique was inspired by attempts to imitate biological neural systems that are capable of learning on examples. A neural network is a set of interconnected neurons that establishes a relationship between optical properties of the material and its chemical composition from a set of examples and then uses it for future predictions. A trained network is a function described by the number of hidden layers, the number of neurons at each layer (with their transfer functions), and a set of weights (including bias terms) assigned to links connecting the neurons. For example, the equation for a neural network with *D* inputs, *K* neurons in one hidden layer, and transfer (activation) function s in both output and hidden layers takes the form

$$\hat{y} = \sigma_1 \left[\sum_{j=1}^{K} \sigma_2 \left(\sum_{i=1}^{D} w_{ij} x_i + b_j \right) v_j + b_0 \right] \qquad [2]$$

where x_i is *i*th input variable, w_{ij} is the weight of the connection from *i*th input to *j*th neuron of the hidden layer (number of *w*-weights is equal to *D* for each hidden layer neuron); v_j is the weight of the connection from *j*th neuron of the hidden layer to

TABLE 1
Reference Data Statistics for Soybean Samples Used for Calibration and Model Validation

FA	Correlation with oil[a]	Number of samples[b]	Mean relative concentration, %	Concentration range,%	SD
Total saturates	0.00	721	16.4	5.3–37.3	7.46
Palmitic C16:0	0.00	616	11.0	2.8–32.3	6.94
Stearic C18:0	0.01	619	4.7	2.0–8.4	1.34
Oleic C18:1	0.15	771	27.7	11.8–51.0	8.71
Linoleic C18:2	0.03	758	51.4	32.4–69.4	7.27
Linolenic C18:3	0.33	976	6.5	1.2–13.3	3.01

[a] Determination coefficient describing correlation of relative FA concentration with the total oil content.
[b] Different number of samples was required to create a uniformly distributed set for each FA.

output neuron (number of v-weights is equal to K); b_j is bias of jth neuron of the hidden layer; b_0 is bias of the output neuron; s_1 and s_2 are functions defined, for example, as

$$\sigma_1(x) = \sigma_2(x) = \frac{1}{1+\exp(-x)} \qquad [3]$$

The main limiting factor of this method is a sufficient number of training samples. More complicated networks require more training examples to perform adequately during prediction. When ANN is used with NIR spectral data where the number of input variables (wavelengths) is usually large (on the order of tens, hundreds, or even thousands) and the number of training samples is limited, it is practical to reduce the number of dimensions of the input space. The optimal number of new inputs and number of neurons in hidden layer(s), as in case with PLS, is found by minimizing cross-validation SE. For more details on the ANN method refer to Haykin (16), Borggaard (17), and Næs *et al.* (18).

(iii) Support vector machines (SVM). The SVM method is based on principles of statistical learning theory developed by Vapnik (19) and was intended for solving classification problems. Later, this technique was adapted for linear and nonlinear function estimation (20).

In the SVM regression approach, data from the original input space is transformed using a mapping function $\varphi(x)$ into a high-dimensional feature space where linear regression is performed. This problem is formulated as constrained quadratic optimization in high-dimensional space. The solution of this problem using the Least Squares SVM regression (LS-SVM) algorithm implemented by Suykens *et al.* (21) is given by the model

$$\hat{y} = \sum_{k=1}^{N} \alpha_k K(\boldsymbol{x}, \boldsymbol{x}_k) + b \qquad [4]$$

where vector $\boldsymbol{x}$ represents new sample, $\boldsymbol{x}_k$ is kth training sample, α_k is Lagrangian multiplier for kth training sample, b is bias term, N is number of training samples, $K(\boldsymbol{x}, \boldsymbol{x}_k)$ is a kernel function defined as

$$K(\boldsymbol{x}, \boldsymbol{x}_k) = \varphi(\boldsymbol{x})'\varphi(\boldsymbol{x}_k) \qquad [5]$$

An SVM model contains information about the relevance of each training sample for calculation of $\hat{y}$ and makes predictions based on relative comparison of new (unknown) sample spectra to the spectra of the k training samples. SVM training is computationally intensive if k is large. More information on SVM may be found in Vapnik *et al.* (20), Suykens *et al.* (21), Smola and Scholkopf (22), and Cogdill and Dardenne (23).

Calibration procedures: Application. (i) PLS. PLS_Toolbox 3.0 (Eigenvector Research Inc., www.eigenvector.com) for MATLAB (The MathWorks Inc., www.mathworks.com) was used for PLS modeling. The number of latent variables was selected using 5-block cross-validation on the training set. (*Note*: In 5-block cross-validation, the training set is divided into 5 equal blocks, 4 of which are used for training and 1 for validation during 5 training iterations.)

(ii) ANN. MATLAB/Neural Network Toolbox (The MathWorks Inc., www.mathworks.com) was used for development of ANN calibration models. Feedforward backpropagation networks were trained on 80% of the calibration samples available for each FA. The other 20% (randomly chosen) of the calibration samples were used as an early stopping set to prevent overfitting during the training process. Input dimensionality was reduced from 100 to 25 by taking every fourth wavelength of the NIR spectra. (*Note*: A preliminary study demonstrated that this simple resampling resulted in ANN calibrations superior to those developed on data compressed with principal component analysis.) The best number of neurons in one hidden layer was determined by 5-block cross-validation on the training set. A tangent sigmoid function and linear function were used as activation functions of hidden layer neurons and an output neuron, respectively.

(iii) LS-SVM. LS-SVMlab1.5 toolbox for MATLAB developed by Suykens *et al.* (21) was used for this part of the experiment. Radial basis function (RBF)

$$K(\mathbf{x}, \mathbf{x_k}) = \exp(-\|\mathbf{x} - \mathbf{x_k}\|^2/\sigma^2) \qquad [6]$$

where σ^2 is the RBF bandwidth, was used as a kernel function. The best pair of complexity regularization parameter and RBF bandwidth for every FA calibration model was determined by 5-block cross-validation on the training set (same training set as for PLS and ANN calibration).

Comparison of calibration methods. Validation sets (25% of samples not used for calibration) were applied to all calibration models to compute the following: coefficient of determination r^2; SE of prediction corrected for bias, SEP; bias or mean difference between NIR-predicted and reference concentrations d; and relative predictive determinant, RPD. Definitions of these parameters can be found in Williams and Norris (24).

To establish the relative significance of the calibration methods, RPD coefficients that characterize overall predictive ability of calibrations were compared using ANOVA. In addition, the effect of calibration set size on performance of the regression methods was studied. Calibration sets for total saturated (palmitic plus stearic) and linolenic FA were reduced to smaller data sets that ranged from 50 to 5% of the original size (for example, calibrations for saturates were developed on sets consisting of 271, 181, 136, 109, 91, 78, 68, 61, 55, 50, 46, 37, 31, and 28 samples). Models for the two constituents developed with the three regression methods on all calibration sets were tested on the same validation sets.

RESULTS AND DISCUSSION

Overall results. Validation results of calibration models developed with PLS, ANN, and LS-SVM regression methods for six FA in whole soybeans are shown in Table 2 (actual vs. predicted concentration plots for saturates and linolenic acid models are provided in Fig. 2). Determination coefficients (r^2) of models ranged from 0.49 to 0.68 for stearic acid and from 0.91

TABLE 2
Validation Statistics of PLS, ANN, and LS-SVM Calibration Equations Developed for Estimation of Relative Fatty Acid Composition in Whole Soybeans by FOSS Infratec

FA	Test statistics[a] and model parameters[b]	PLS calibration model	ANN calibration model	LS-SVM calibration model
Saturates (C16:0 + C18:0)	r^2	0.91	0.92	0.94
	SEP	2.23	2.13	1.80
	d	0.01	−0.31	−0.02
	RPD	3.3	3.5	4.2
	Model param.	16	25/3	1413/7450
Palmitic (C16:0)	r^2	0.80	0.84	0.82
	SEP	3.16	2.79	2.94
	d	−0.33	−0.44	−0.39
	RPD	2.2	2.5	2.4
	Model param.	15	25/4	8154/38888
Stearic (C18:0)	r^2	0.49	0.64	0.68
	SEP	0.97	0.82	0.77
	d	−0.08	−0.01	−0.06
	RPD	1.4	1.7	1.8
	Model param.	9	25/3	26/208
Oleic (C18:1)	r^2	0.76	0.80	0.81
	SEP	4.27	3.93	3.88
	d	−0.63	−0.63	−0.28
	RPD	2.1	2.2	2.3
	Model param.	9	25/3	19/236
Linoleic (C18:2)	r^2	0.73	0.74	0.76
	SEP	3.77	3.67	3.56
	d	0.18	0.19	2.27
	RPD	1.9	2.0	2.0
	Model param.	9	25/3	117/1348
Linolenic (C18:3)	r^2	0.67	0.73	0.74
	SEP	1.74	1.56	1.53
	d	−0.09	−0.02	−0.13
	RPD	1.7	1.9	2.0
	Model param.	13	25/3	32/397

[a] r^2 is determination coefficient, SEP is SE of prediction corrected for bias, *d* is bias, RPD is relative predictive determinant, PLS is partial least squares, ANN is artifical neural networks, and LS-SVM is Least Squares support vector machines.
[b] Model parameters provide number of latent variables for PLS, number of inputs and neurons in a hidden layer for ANN, and radial basis function bandwidth and complexity regularization parameter for LS-SVM.

to 0.94 for total saturates. In terms of RPD (RPD = SD_y/SEP, where SD_y is SD of reference values in the test set), predictive ability varied from 1.4 to 1.8 for stearic acid and from 3.3 to 4.2 for total saturates. Based on guidelines for interpretation of r^2 coefficients outlined by Williams and Norris (24), NIR calibration equations for saturated FA were usable for quality assurance applications, whereas those for palmitic acid were "usable with caution for most applications, including research." Models for the other four FA had lower predictive power. However, they could still be used for sample screening, which is an important task in seed breeding programs. Validation results also demonstrated that predictive ability of NIR calibration equations was not dependent on correlation between total oil and relative FA concentration (refer to Table 1 for determination coefficients describing relationship of individual FA with total oil content). For example, validation r^2 values of NIR calibration models (PLS, ANN, and LS-SVM) for saturates were high (0.91–0.94), whereas correlation between oil content and saturates was practically zero (0.003). In addition, predictive ability of the calibration models was independent of error of the reference GC method. This suggests that NIR spectroscopy and calibration methods used in this study could have using information from individual FA absorption bands, not from wider total fat absorption bands.

Most of the variation (79%) in predictive ability of NIR calibration models could be explained by the SD of reference data in calibration sets (graphs of r^2 or RPD vs. SD of reference data are not shown). Thus, the introduction of a larger number of samples with extremely low and high values of stearic and linolenic acids into corresponding calibration data sets may improve the predictive ability of NIR spectroscopy for these constituents.

A previous study by Pazdernik *et al*. (13) on the applicability of NIR spectroscopy for determination of FA composition in soybeans (NIRSystems 6500 spectrometer and PLS regression) resulted in validation r^2 values of 0.38 (palmitic), 0.66 (stearic), 0.68 (oleic), 0.70 (linoleic), and 0.71 (linolenic) for models developed on ground samples and 0.18 (palmitic), 0.54 (stearic), 0.38 (oleic), 0.52 (linoleic), and 0.56 (linolenic) for those of the whole-seed samples. The r^2 coefficients of our ex-

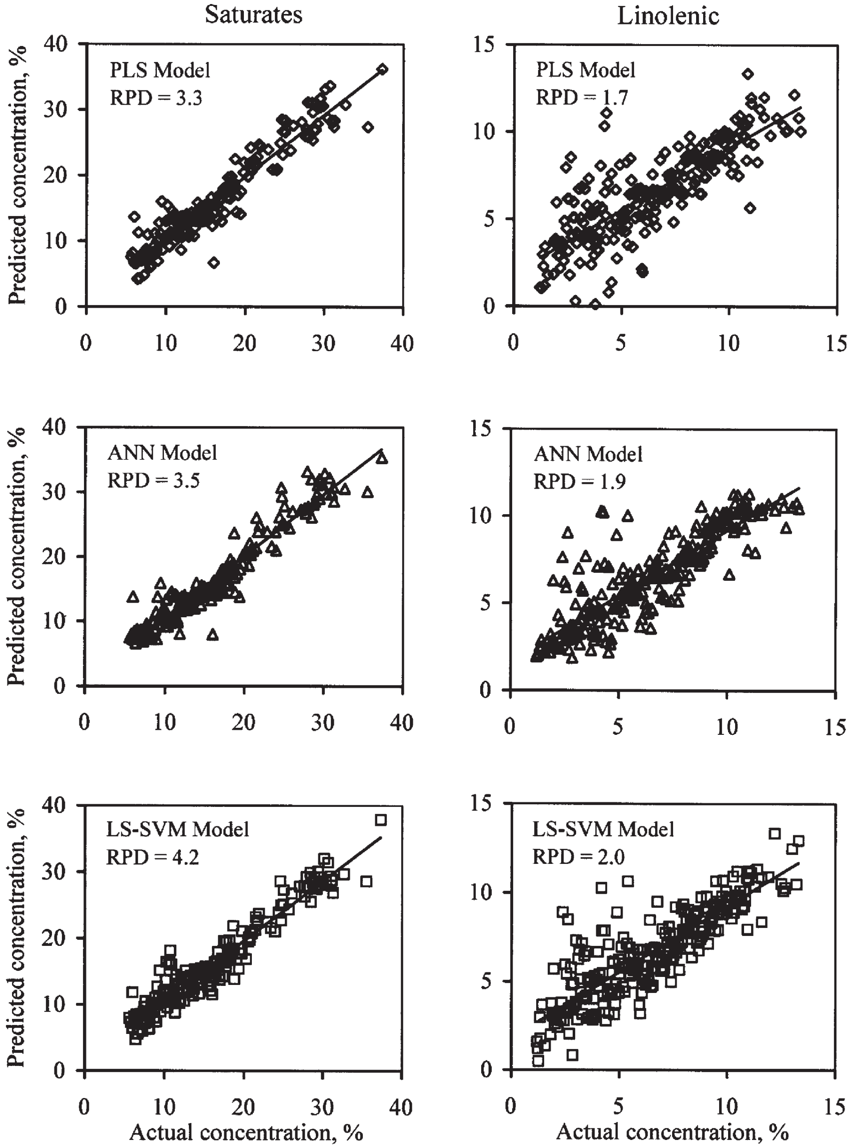

FIG. 2. Actual vs. predicted concentration plots for saturates and linolenic FA calibrations. Models were tested on sets of 180 (saturates) and 244 (linolenic) samples. The solid line on each plot represents the regression line.

periment (Table 2) were higher than both sets of results reported by Pazdernik *et al.*, which suggests that satisfactory accuracy of NIR predictions may be achieved without grinding the seed samples.

Comparison of calibration methods. Interpretation and comparison of the validation results in the previous section was based on r^2 values; however, because of the mathematical relationship between r^2 and RPD, the same conclusions regarding predictive ability of calibration models could be drawn from analysis of RPD values. For the benefit of RPD-accustomed researchers, further comparison of calibration methods is based on the aforementioned statistic.

Validation RPD ratios for 18 calibration models are illustrated in Figure 3. Visual analysis of the bar chart suggested superior performance of nonlinear regression methods. In order to confirm this, ANOVA modeling of the form

$$\mathrm{RPD} = \mathrm{FA} + \mathrm{M} + \mathrm{error} \qquad [7]$$

where FA is the FA factor and M is the calibration method factor, was performed. The interaction FA*M was used as the error term. Mean values of M factor were compared using Tukey's test at a = 0.05. The results of statistical analysis demonstrated that both FA and M factors had significant effect on the RPD coefficient ($P < 0.0001$ and $P = 0.017$, respectively). The mean RPD of LS-SVM equations (mean = 2.4, SE of mean based on six replicates = 0.064) was significantly better than that of PLS equations (mean = 2.1, SE = 0.064). How-

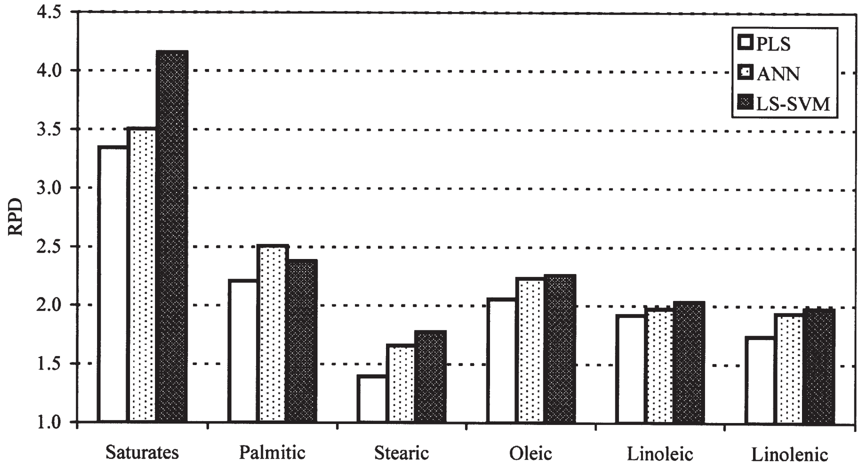

FIG. 3. Validation relative predictive determinant (RPD) of partial least squares (PLS), artificial neural networks (ANN), and the Least Square support vector machines (LS-SVM) calibration equations developed for determination of FA composition in whole soybeans by FOSS Infratec spectrometers.

ever, mean RPD of ANN calibrations (mean = 2.3, SE = 0.064) was not significantly different from the other two methods.

Calibration models for saturated and linolenic FA were developed using reduced calibration data sets as described in the last paragraph of the Experimental Procedures section. The results, RPD coefficient as a function of calibration set size, are shown in Figure 4. As expected, the predictive ability of calibration equations dropped as the number of calibration (training) samples decreased, regardless of the regression method or type of predicted constituent. However, the rate of performance degradation was dependent on calibration method and constituent. ANN models had the highest rate of performance degradation for both saturated and linolenic FA. These equations became unusable (RPD < 1.4 or $r^2 < 0.5$) when they were

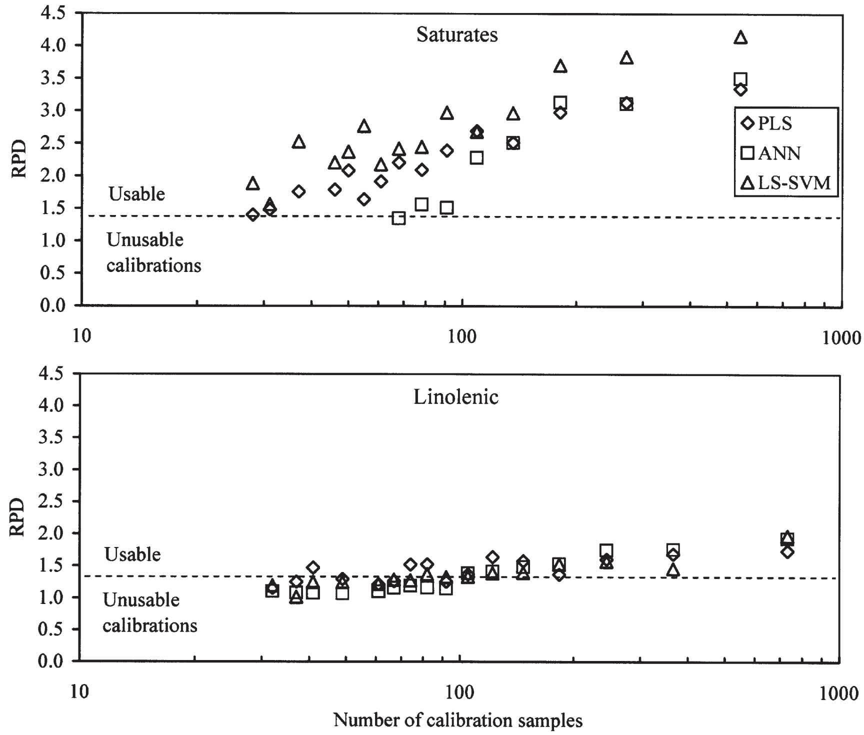

FIG. 4. Validation RPD of PLS, ANN, and LS-SVM equations as a function of calibration set size for saturated and linolenic FA in soybeans. For abbreviations see Figure 3.

developed using calibration sets of fewer than 70 samples for saturates and 120 samples for linolenic acids. LS-SVM equations for saturates displayed the best tolerance to reduction of number of calibration samples. LS-SVM calibrations for linolenic acid demonstrated behavior similar to PLS and ANN. Another important observation about saturates models was that variation of RPD of LS-SVM equations developed on calibration sets of fewer than 100–150 samples was higher than that of PLS models. This suggests a strong sensitivity of this nonlinear regression method to outliers and/or unusual samples in small calibration sets.

ACKNOWLEDGMENTS

This journal paper of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, was supported by Hatch Act, State of Iowa, and United Soybean Board funds.

REFERENCES

1. Durham, D., The United Soybean Board's Better Bean Initiative: Building United States Soybean Competitiveness from the Inside Out, *AgBioForum 6*:23–26 (2003) (netlink: http://www.agbioforum.org/v6n12/v6n12a07-durham.htm).
2. Knowlton, S., S.K.B. Ellis, and E.F. Kelly, Performance Characteristics of High Oleic Soybean Oil: An Alternative to Hydrogenated Fats, presented at the 87th American Oil Chemists' Society Annual Meeting & Expo, paper No. 29-O (1996).
3. Streit, L., W. Fehr, G. Welke, E. Hammond, and S. Cianzio, Family and Line Selection for Reduced Palmitate, Saturates, and Linolenate of Soybean, *Crop Sci. 41*:63–67 (2001).
4. Fehr, W.R., G.A. Welke, E.G. Hammond, D.N. Duvick, and S.R. Cianzio, Inheritance of Elevated Palmitic Acid Content in Soybean Seed Oil, *Ibid. 31*:1522–1524 (1991).
5. List, G.R., T.L. Mounts, F. Orthoefer, and W.E. Neff, Potential Margarine Oils from Genetically Modified Soybeans, *J. Am. Oil Chem. Soc. 73*:729–732 (1996).
6. Velasco, L., C. Moellers, and H.C. Becker, Estimation of Seed Weight, Oil Content and Fatty Acid Composition in Intact Single Seeds of Rapeseed (*Brassica napus* L.) by Near-Infrared Reflectance Spectroscopy, *Euphytica 106*:79–85 (1999).
7. Velasco, L., and H.C. Becker, Estimating the Fatty Acid Composition of the Oil in Intact-Seed Rapeseed (*Brassica napus* L.) by Near-Infrared Reflectance Spectroscopy, *Ibid. 101*:221–230 (1998).
8. Sato, T., I. Uezono, T. Morishita, and T. Tetsuka, Nondestructive Estimation of Fatty Acid Composition in Seeds of *Brassica napus* L. by Near-Infrared Spectroscopy, *J. Am. Oil Chem. Soc. 75*:1877–1881 (1998).
9. Sato, T., Y. Takahata, T. Noda, T. Yanagisawa, T. Morishita, and S. Sakai, Nondestructive Determination of Fatty Acid Composition of Husked Sunflower (*Helianthus annua* L.) Seeds by Near-Infrared Spectroscopy, *Ibid. 72*:1177–1183 (1995).
10. Velasco, L., A. Schierholt, and H.C. Becker, Performance of Near-Infrared Reflectance Spectroscopy (NIRS) in Routine Analysis of C18 Unsaturated Fatty Acids in Intact Rapeseed, *Fett/Lipid 100*:44–48 (1998).
11. Perez-Vich, B., L. Velasco, and J.M. Fernandez-Martinez, Determination of Seed Oil Content and Fatty Acid Composition in Sunflower Through the Analysis of Intact Seeds, Husked Seeds, Meal and Oil by Near-Infrared Reflectance Spectroscopy, *J. Am. Oil Chem. Soc. 75*:547–555 (1998).
12. Dyer, D.J., and P. Feng, Near Infrared Applications in the Development of Genetically Altered Grains, in *Near Infrared Spectroscopy: The Future Waves*, edited by A.M.C. Davies and P. Williams, NIR Publications, Chichester, United Kingdom, 1995, pp. 490–493.
13. Pazdernik, D.L., A.S. Killam, and J.H. Orf, Analysis of Amino and Fatty Acid Composition in Soybean Seed, Using Near Infrared Reflectance Spectroscopy, *Agron. J. 89*:679–685 (1997).
14. Kovalenko, I.V., Determination of Amino and Fatty Acid Composition of Soybeans Using Near-Infrared Spectroscopy, Doctoral Dissertation, Iowa State University, Ames, Iowa, 2005, pp. 20–48.
15. Hammond, E.G., Organization of Rapid Analysis of Lipids in Many Individual Plants, in *Modern Methods of Plant Analysis, New Series, Volume 12, Essential Oils and Waxes*, edited by H.F. Linskens and J.F. Jackson, Springer-Verlag, Berlin, 1991, pp. 321–330.
16. Haykin, S., *Neural Networks: A Comprehensive Foundation*, 2nd edn., Prentice Hall, New Jersey (1999).
17. Borggaard, C., Neural Networks in Near-Infrared Spectroscopy, in *Near-Infrared Technology in the Agricultural and Food Industries*, edited by P. Williams and K. Norris, 2nd edn., AACC Inc., St. Paul, Minnesota, 2001, pp. 101–107.
18. Næs, T., T. Isaksson, T. Fearn, and T. Davies, *A User-Friendly Guide to Multivariate Calibration and Classification*, NIR Publications, Chichester, United Kingdom, 2002.
19. Vapnik, V., *The Nature of Statistical Learning Theory*, Springer, New York, New York, 1995.
20. Vapnik, V., S. Golowich, and A. Smola, Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing, in *Advances in Neural Information Processing Systems 9*, edited by M. Mozer, M. Jordan, and T. Petsche, MIT Press, Cambridge, Massachusetts, 1997, pp. 281–287.
21. Suykens, J.A.K., T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002.
22. Smola, A.J., and B. Scholkopf, *A Tutorial on Support Sector Regression*, NeuroCOLT2 Technical Report Series, NC-TR-98-030, Royal Holloway College, University of London, United Kingdom, 1998.
23. Cogdill, R.P., and P. Dardenne, Least-Squares Support Vector Machines for Chemometrics: An Introduction and Evaluation, *J. Near Infrared Spectrosc. 12*:93–100 (2004).
24. Williams, P., and K. Norris (eds.), *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd edn., AACC Inc., St. Paul, Minnesota, 2001.

[Received October 17, 2005; accepted February 24, 2006]