Simultaneous Spectrophotometric and Chemometric Determination of Oleic, Linoleic, and Linolenic Fatty Acids in Vegetable Oils

Gerard G. Dumancas, Mary Muriuki, Neil Purdie, and Lisa Reilly

Abstract-**Simultaneous spectrophotometric methods are described for the determination of oleic, linoleic, and linolenic fatty acids in vegetable oil samples using neural network (NN), principal component regression (PCR), partial least squares (PLS1 and PLS2), and K-matrix (KM) algorithms. The assay used to obtain the absorbance spectrum unique for each fatty acid is selective to the -CH=CH-CH_2 that reaches spectral maturity after 15 minutes. Results show that the root mean square error of prediction (RMSEP) compared quite equally well for PCR, PLS1, and PLS2 algorithms for the three components, with these algorithms outperforming NN and KM. In sunflower and vegetable oil unknown samples, PLS2 mostly yielded a better performance than PLS1 and PCR algorithms when validated with the USDA database. This paper shows how the novel assay coupled with chemometric algorithms might provide faster and cheaper methods for simultaneously quantitating oleic, linoleic, and linolenic fatty acids in vegetable oil samples.**

Index Terms-**Partial least squares, mixture models, chemometrics, vegetable oils, spectrophotometry**

I. INTRODUCTION

Vegetable oils are a group of natural products consisting largely of triacylglycerols (TAGs) that find widespread uses as sources of edible oils and surfactants. High performance liquid chromatography (HPLC), gas chromatography (GC), or hyphenated methods such as HPLC/mass spectrometry (MS), and GC/MS are classical methods of determining the fatty acid composition and levels in vegetable oils [1], [2]. For the determination of the fatty acid composition, the TAGs are transesterified to give the methyl esters prior to analysis because the esters are less polar than the corresponding fatty acids, and, thus, are more compatible with the various chromatographic systems [3]. Though these procedures have been successfully used in various chemical analyses, they, however, do suffer from the disadvantages of being time, labor, and resources consuming.

Manuscript received January 25, 2011; revised January 29, 2011. This work was supported in part by the Oklahoma State University (OSU) Technology Business Assessment Group (TBAG).

Gerard G. Dumancas is with the OSU Chemistry Department, 002 Physical Science Building, Stillwater, OK 74078, USA (Phone: 405-744-5948; fax: 405-744-6007; email: gerard.dumancas@okstate.edu).

Mary Muriuki is with the OSU Chemistry Department, 002 Physical Science Building, Stillwater, OK 74078, USA (email: marywm@okstate.edu).

Neil Purdie is with the OSU Chemistry Department, 018 Physical Science Building, Stillwater, OK 74078, USA (email: neil.purdie@okstate.edu)

Lisa Reilly is with the Department of Physical Sciences, Bethany College, Bethany, WV 26032, USA (email: LReilly@bethanywv.edu)

The analysis of fatty acid levels is a continuing concern for workers in the lipid field. Originally carried out by titration, the need for speed and/or increased sensitivity led to the development of a number of procedures for the determination of such analytes [4]. Oleic, linoleic, and linolenic fatty acids are among the most abundant fatty acid analytes found in vegetable oils.

The degree of oil's unsaturation determines its stability. This is of particular importance in determining whether oils are of good quality. Oils that are more unsaturated are oxidized more quickly than less unsaturated oils, and, thus, are of low quality [5]. As the degree of unsaturation increases, both the rate of formation and the amount of primary oxidation compounds accumulated at the end of the induction period increase [6]. Oxidation of unsaturated lipids not only produces offensive odors and flavors but can also decrease the nutritional quality and safety by the formation of secondary reaction products in foods after cooking and processing [7].

This study is aimed at simultaneously quantitating oleic, linoleic, and linolenic fatty acids in vegetable oil samples using faster and cheaper methods. The methods are based on the application of a mature, patented reagent system selective to the -CH=CH-CH_2- group in various fatty acids. Various chemometric models consisting of K-matrix (KM), neural network (NN), principal component regression (PCR), and partial least squares (PLS1 and PLS2) were utilized for the deconvolution of the spectrophotometric data. The most robust chemometric models were then compared against each and their regression coefficients applied for the molar concentration determination of oleic, linoleic, and linolenic fatty acids in olive and sunflower oil samples. The obtained concentrations were then validated with the existing USDA database concentrations [8].

Chemometric algorithms

Theoretical background of the different chemometric techniques are discussed below.

The KM approach can be expressed in a matrix notation:

$$A = CK \qquad (1)$$

where A is the n x p matrix of absorbances, C is the n x m matrix of concentrations of constituents, K is the m x p matrix of absorptivities, n (=128) is the number of samples, p (=101) is the number of wavelengths, and m (=7) is the number of components. Calibration is based on a set of n samples of known concentrations for which the spectra are measured. By means of the calibration sample set, estimation of absorptivities is possible by solving for the matrix K according to the general least squares solution:

ISBN: 978-988-19251-5-2 (revised on 20 September 2011)
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$K = (C^T C)C^T A \quad (2)$$

The analysis is based on the spectrum a_0 (1 x p) of the unknown sample by use of:

$$c_0 = a_0 K^T (KK^T)^{-1} \quad (3)$$

where c_0 is the (1 x m) vector of sought-for concentrations [9].

PCR is best performed by means of SVD (singular value decomposition). This method involves the decomposition of the absorbance matrix A into two orthogonal matrices U and V joined by a diagonal matrix W of singular values:

$$A = UWV^T \quad (4)$$

Estimation of the matrix of regression coefficients B is performed column-wise by use of:

$$b = A^+ c \quad (5)$$

with A^+ being the pseudo-inverse of the absorbance matrix A [9].

Details of the PLS method can be referred to Otto [9]. It involves the decomposition of A and C according to:

$$A = TP^T + E \quad (6)$$

$$C = UQ^T + F \quad (7)$$

where T and U are the n x d scores matrices containing orthogonal rows; P are the p x d loadings of the A matrix; E is the n x p error (residual) matrix of A matrix; Q is the m x d loading matrix of the C matrix; and F is the n x m error (residual) matrix for the C matrix.

Computation of the B-coefficients for the general model gives:

$$B = W(P^T W)^{-1} Q^T \quad (8)$$

with W as d x p matrix of PLS-weights.

Neural network on the, other hand, is divided into three layers comprised of input, hidden layers, and output. The input parameters are the absorbance at specified wavelengths. The parameters are connected to neurons in the hidden layer. The number of hidden layers and the number of neurons in each layer is flexible and is determined by the examination of errors in the results in the output layer in terms of concentrations. A basic network design is show in Fig. 1. More detailed information about neural network can be found in Hagan *et al* [10].

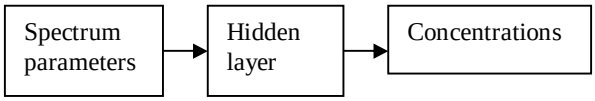

Fig. 1. Example of a basic neural network design.

The theories behind the various chemometric algorithms will not be discussed further in this study and can be referred in various bibliographic references [11]-[16].

Table I lists the fatty acid composition for the TAG oils of interest. Monounsaturated fatty acids such as palmitoleic (C16:1), gadoleic (C20:1), and erucic (C22:1) exist in traces, and, thus, are not taken into account in this study [17]-[19].

Oleic, linoleic, and linolenic fatty acids, are known to give unique absorbance spectra allowing the possibility of deconvolution by using various chemometric models [20].

TABLE I. FATTY ACID COMPOSITION RANGES FOR THE TAG OILS UNDER STUDY (% WEIGHT COMPOSITION)

Fatty Acid	Sun-flower	Soy-bean	Saf-flower	Corn	Flax seed
18:1 (oleic)	14-65	19-30	8.4-21.3	19-49	19
18:2 (linoleic)	20-75	44-62	67.8-83.2	34-52	24.1
18:3 (linolenic)	<0.7	4.0-11	0-0.1	Trace	47.4

II. MATERIALS AND METHODS

Training, prediction, and unknown sets

Oleic, linoleic, and linoleic fatty acid methyl esters should be as much as possible in concentration ranges of 0.0025 to 0.02 M in chloroform solutions in order to maintain the absorbance units from 0 to 1.2. It should be noted that the linolenic acid discussed all throughout in this paper refers to the alpha form. Tables II and III show the actual molar concentrations of the training and prediction set standards prepared. All samples were obtained from Sigma-Aldrich.

TABLE II. OLEIC, LINOLEIC, AND LINOLENIC FATTY ACID METHYL ESTERS (FAME) TRAINING MATRIX BY CENTRAL COMPOSITE AND SIMPLEX LATTICE DESIGNS IN CHLOROFORM SOLUTIONS.

	Oleic	Linoleic	Linolenic
Mixture 1	8.31E-03	2.42E-03	1.92E-02
Mixture 2	1.91E-02	1.90E-02	2.33E-03
Mixture 3	8.31E-03	1.90E-02	2.33E-03
Mixture 4	8.31E-03	2.42E-03	2.33E-03
Mixture 5	1.91E-02	2.42E-03	2.33E-03
Mixture 6	1.72E-02	1.71E-02	1.74E-02
Mixture 7	8.31E-03	2.42E-03	1.92E-02
Mixture 8	8.31E-03	1.90E-02	1.92E-02
Mixture 9	1.37E-02	1.07E-02	2.33E-03
Mixture 10	1.37E-02	1.07E-02	1.92E-02
Mixture 11	1.37E-02	1.90E-02	1.08E-02
Mixture 12	1.37E-02	2.42E-03	1.08E-02
Mixture 13	1.91E-02	1.07E-02	1.08E-02
Mixture 14	8.31E-03	1.07E-02	1.08E-02
Mixture 15	1.37E-02	1.07E-02	1.08E-02
Mixture 16	1.37E-02	1.07E-02	1.08E-02
Mixture 17	1.37E-02	1.07E-02	1.08E-02
Mixture 18	1.37E-02	1.07E-02	1.08E-02
Mixture 19	1.37E-02	1.07E-02	1.08E-02
Mixture 20	1.37E-02	1.07E-02	1.08E-02
Mixture 21	1.43E-02	2.42E-03	2.33E-03
Mixture 22	9.55E-03	1.90E-02	2.33E-03
Mixture 23	9.55E-03	2.42E-03	1.92E-02
Mixture 24	1.24E-02	6.67E-03	2.33E-03
Mixture 25	8.31E-03	1.24E-02	2.33E-03
Mixture 26	1.24E-02	2.42E-03	6.65E-03
Mixture 27	8.31E-03	2.42E-03	1.24E-02
Mixture 28	8.31E-03	1.24E-02	6.65E-03
Mixture 29	8.31E-03	6.67E-03	1.24E-02
Mixture 30	8.31E-03	6.67E-03	2.33E-03

ISBN: 978-988-19251-5-2 (revised on 20 September 2011) WCE 2011
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE III. OLEIC, LINOLEIC, AND LINOLENIC FATTY ACID METHYL ESTERS PREDICTION MATRIX BY AN INDEPENDENT SIMPLEX LATTICE DESIGN IN CHLOROFORM SOLUTIONS.

	Oleic	Linoleic	Linolenic
Mixture 1	1.24E-02	3.45E-03	3.50E-03
Mixture 2	8.42E-03	1.73E-02	3.50E-03
Mixture 3	8.42E-03	3.45E-03	1.75E-02
Mixture 4	1.40E-02	7.48E-03	3.50E-03
Mixture 5	1.01E-02	1.44E-02	3.50E-03
Mixture 6	1.40E-02	3.45E-03	8.16E-03
Mixture 7	1.01E-02	3.45E-03	1.34E-02
Mixture 8	1.01E-02	1.44E-02	8.16E-03
Mixture 9	1.01E-02	7.48E-03	1.34E-02
Mixture 10	1.01E-02	7.48E-03	3.50E-03

Analysis of FAME standards and vegetable oils using the assay and validation

The procedure for the acetyl chloride/perchloric acid (AC/PA) color assay reaction is conceptually simple. It entails placing a 10 microL aliquot of the fatty acid methyl esters (FAME) standards or vegetable oils into a 13 x 100 mm borosilicate disposable test tube followed immediately by the careful addition of 1.0 mL AC then 40 microL of PA. The test tube is sealed tightly with parafilm and gently shaken for 20 seconds. The supernate is then transferred by pipette to a 10 mm pathlength optical glass cuvette and placed in the sample holder of a diode-array spectrophotometer (HP8452A). Analysis is done after 15 minutes from 350-550 nm at every 2 nm and 5 s integration time. Using the developed calibration matrix, the FAME standard concentrations in prediction sets and vegetable oils will be determined using the chemometric techniques mentioned in the first project.

The obtained chemometric molar concentrations were then validated with the existing USDA database concentrations [8].

Chemometric techniques

The training, prediction, and unknown set spectra were deconvoluted using various chemometric algorithms. KM, NN, PCR, and PLS algorithms were utilized in this study. Mean centering was performed prior to the chemometric analyses. Chemometric analyses were performed in MATLAB using *Chemometric Toolbox* [21]. Neural network was performed using the *JMP Software Package* [22].

Determining the number of factors (rank) to be used in the calibration is a key step in both PCR and PLS. To select the number of factors for PLS and PCR methods, the cross validation, leaving out one sample at a time, was used. This process was repeated 29 times, until each sample had been left out once. The Predicted Residual Error Sum of Squares (PRESS) was used to determine the optimum number of factors in both algorithms. To calculate the PRESS we computed the errors between the expected and predicted concentrations for all of the samples, square them, and sum them together as given by Eq. (9) [21]:

$$PRESS = \sum_{i=1}^{N} (y_i - y_i')^2 \qquad (9)$$

where y and y' are the predicted and actual concentrations and N is the number of samples. The plot of the PRESS values as a function of the number of factors indicates the rank to be used in the calibration. The root mean square error (RMSE) is also calculated for each algorithm. The general equation is:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - y_i')^2}{N}} \qquad (10)$$

The model with the minimum values for the RMSE indicated the appropriate model.

III. RESULTS AND DISCUSSIONS

The molar absorbance spectra for oleic, linoleic, and linolenic fatty acid methyl esters obtained using the training set are shown in Fig. 2. Central composite design and simplex lattice design training sets were used because they have demonstrated to be a useful method in formulations of experiments, fits nicely into the sequential experimentation that is involved with the experimental design, requires fewer experiments, and provides convenience and high accuracy [23]-[25]. It is readily apparent that linoleic and linolenic molar absorbance spectra are six times greater than that of the oleic. Oleic is characterized by two smooth valley type peaks found at 368 and 442 nm. Linoleic, on the other hand has maxima that both occur at 376 and 426 nm. Linolenic has two maximum peaks, also occurring at 376 and 426 nm, with the latter peak about 1000 molar absorbance more than the 426 nm peak of the linoleic. It is also apparent from Fig. 2, a small shoulder found at 444 nm for linolenic.

Oleic fatty acid is a monounsaturated fatty acid with a double bond occurring at carbon 9 relative to the -COOH terminal. Linoleic, on the other hand has two double bonds occurring at carbons 9 and 12; while linolenic has three double bonds found at carbons 9, 12, and 15 all relative to the -COOH terminal (Fig. 3). The most probable reason why the molar absorbance of oleic is buried under that of linoleic and linolenic fatty acids is due to its monounsaturated property.

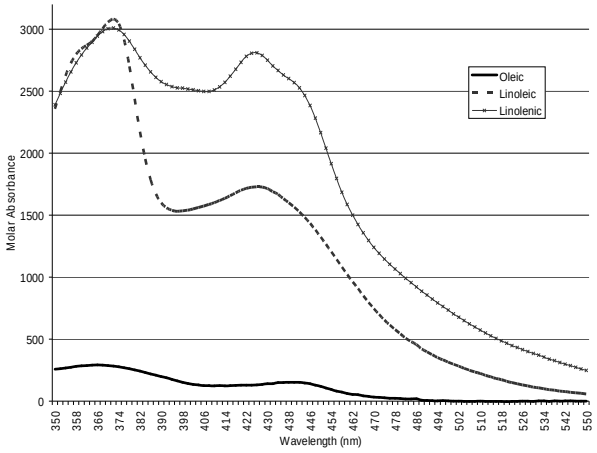

Fig. 2. Molar absorbance spectra of oleic, linoleic, and linolenic fatty acids obtained by the K-matrix chemometric approach.

ISBN: 978-988-19251-5-2 (revised on 20 September 2011)
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

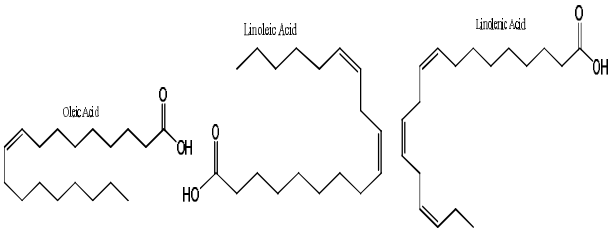

Fig. 3. Structures of oleic, linoleic, and linolenic fatty acids.

The first attempt of deconvoluting the spectra is by the KM model. However, the KM approach yielded RMSEP high enough that the regression coefficients may yield high errors in the actual sample sets. The RMSEP is especially useful in comparing the prediction errors of the different regression models [26]. A high RMSEP (greater than 10 μM) as in this study simply means the regression model will give high errors in the unknown set samples. It is observed that there are less samples (n=30) in the training sets than the number of variables (p=101). In such case, the KM calibration model has limited applicability, yielding high RMSEP [27]. Although KM approach offers the advantage of estimating the true constituent spectra of the components in the training sets, it has, however, the disadvantage of requiring the knowledge of the concentrations of all interfering chemical constituents with a spectral profile in the training set and that the calibration and analysis are connected to the inversion of the matrix [28], [29].

NN was next attempted in the study using 3 hidden nodes and 200 maximum iterations. NN improved the RMSEP over the KM approach yet not low enough to be used for the unknown sets (Fig. 4). NN offers the disadvantage of requiring a large amount of data to ensure that the results are statistically accurate and the networks adapt their analysis of data in response to the training which is connected to the network [30].

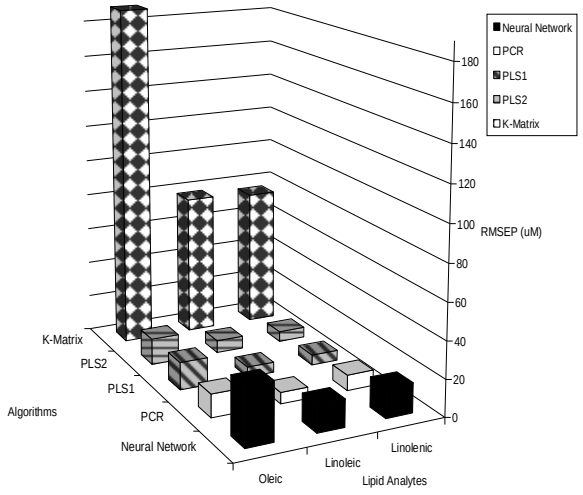

Fig. 4. RMSEP of oleic, linoleic, and linolenic fatty acid methyl esters compared against the different algorithms.

PCR and PLS algorithms were then attempted to obtain their RMSEPs, and the results are satisfying over the KM and NN models. A quite equal performance for PCR, PLS2, and PLS1 algorithms were obtained for their respective RMSEPs (Fig. 4). PLS differs from PCR in that it uses the concentration data from the training set and the spectral data in modeling, whereas PCR only uses the spectral data [31]. However, the solutions and hence the performance of PLS and PCR tend to be quite similar in most situations, largely because they are applied to problems involving high collinearity [29].

A decision about the number of suitable PLS/PCR factors is necessary. Too few factors lead to underfitting leading to inadequate predictions, since the information extracted by the model is not enough to explain the data. On the other hand, too many factors leads to overfitting, that is, the model cannot be generalized to new data that did not contribute to the model construction [32].

For PCR, 6 factors were used for the model while 18 factors were used for PLS2. For PLS1, 5, 6, and 12 factors were used for oleic, linoleic, and linolenic, respectively (Fig. 5). These factors were chosen based on the plot of PRESS vs number of factors/rank chart as shown in Fig. 5.

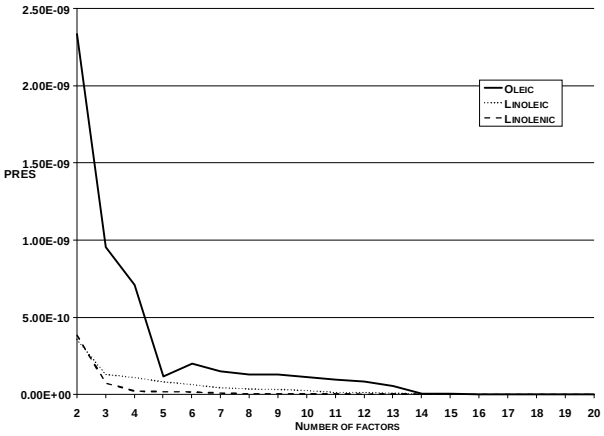

Fig. 5. PRESS chart for oleic, linoleic, and linolenic fatty acid methyl esters. 5, 6, and 12 factors were chosen for oleic, linoleic, and linolenic, respectively in PLS1.

The plot with the lowest PRESS indicated the number of factors to be used for PLS1. After choosing the number of factors for each algorithm, the RMSEP was calculated to indicate the appropriate model. The model(s) with the lowest RMSEP indicated to be the appropriate model to be tested for the unknown samples.

After choosing the number of factors for PCR, PLS2, and PLS1 algorithms (Fig. 5), the obtained regression coefficients were applied to determine the molar concentrations of olive and sunflower oil samples. In Table IV, PLS2 yielded the lowest % error for oleic but not so much difference with PLS1 and PCR. For linoleic, PLS2 also yielded the lowest % error compared to PLS1 and PCR algorithms. In Table V, PLS2 also yielded the lowest % error for oleic while PCR yielded the lowest % error for linoleic. Based on the obtained results, PLS2 mostly yielded a better performance than PLS1 and PCR algorithms.

Normally we expect PLS1 to give a better model than PLS2. However, PLS2 gives better results than PLS1 especially if the analyte concentrations are strongly correlated [27]. Noticeable zero concentrations were obtained for linolenic fatty acid using the USDA database for the primary reason that it exists in low quantities in

ISBN: 978-988-19251-5-2 (revised on 20 September 2011) WCE 2011
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

vegetable oil samples relative to both oleic and linoleic [8], [17]-[19].

Expanding the training and prediction sets and testing the PLS and PCR algorithms to other types of vegetable oil samples would probably improve the differentiation as to which algorithm would be the most appropriate one to be employed in this study.

TABLE IV

TABLE IV. MOLAR CONCENTRATIONS OF OLEIC LINOLEIC, AND LINOLENIC FATTY ACID METHYL ESTERS IN OLIVE OIL SAMPLES COMPARED USING THE THREE MOST ROBUST ALGORITHMS.

Component	PLS2	Database	% Error	PLS1	Database	% Error	PCR	Database	% Error
Oleic	1.50E-02	2.37E-02	36.7	1.49E-02	2.37E-02	37.2	1.46E-02	2.37E-02	38.3
Linoleic	1.06E-03	1.05E-03	-1.3	9.17E-04	1.05E-03	12.7	9.99E-04	1.05E-03	4.9
Linolenic	1.50E-03	0	-	1.47E-03	0	-	1.50E-03	0	-

TABLE V. MOLAR CONCENTRATIONS OF OLEIC LINOLEIC, AND LINOLENIC FATTY ACID METHYL ESTERS IN SUNFLOWER OIL SAMPLES COMPARED USING THE THREE MOST ROBUST ALGORITHMS.

Component	PLS2	Database	% Error	PLS1	Database	% Error	PCR	Database	% Error
Oleic	2.08E-03	2.30E-03	9.6	6.88E-03	6.10E-3	-12.7	1.09E-03	2.30E-03	52.5
Linoleic	4.41E-03	3.84E-03	-14.9	1.27E-02	1.06E-2	-19.9	4.20E-03	3.84E-03	-9.4
Linolenic	8.59E-04	0	-	1.11E-03	0	-	5.85E-04	0	-

IV. CONCLUSIONS

The most important aspect of this work is the possibility of simultaneous determination of oleic, linoleic, and linolenic fatty acids in vegetable oil samples using the patented assay developed. No extraction step is required, and hence the use of organic solvents for separation, which are generally toxic pollutants, is avoided. It has been shown in this study that PCR, PLS2, and PLS1 algorithms compared quite equally well in the prediction sets and that PLS2 mostly yielded a better performance than PLS1 and PCR algorithms in the unknown samples. Compared to most other existing methods, the proposed methods are very simple, cheap, rapid and especially selective.

REFERENCES

[1]. G.P. Savage, P.C. Dutta, D.L. McNeil, "Fatty Acid and Tocopherol Contents and Oxidative Stability of Walnut Oils," *Journal of the American Oil Chemists' Society*, vol. 76, no 9, pp. 1059-1063, 1999.

[2]. F.O. Ayorinde, J. Clifton Jr., O.A. Afolabi, R.L. Shepard, "Rapid transesterification and mass spectrometric approach to seed oil analysis," *Journal of the American Oil Chemists' Society*, vol. 65, no. 6, pp. 942-947, 1988.

[3]. B.H. Jennings, C.C. Akoh, "Enzymatic modification of triacylglycerols of high seicosapentaenoic and docosahexaenoic acids content to produce structured lipids," *Journal of the American Oil Chemists'* Society, vol. 7, no. 10, pp. 1133-1137, 1999.

[4]. R.R. Lowry, I.J. Tinsley, "Rapid colorimetric determination of free fatty acids," *Journal of the American Oil Chemists' Society*, vol. 53, no. 7, pp. 470-472, 1969.

[5]. T.D. Parker, D.A. Adams, K. Zhou, M. Harris, L. Yu, "Fatty acid composition and oxidative stability of cold-pressed edible seed oils," *Journal of Food Science*, vol. 68, no. 4, pp. 1240–1243, 2003.

[6]. M. Martin-Polvillo, G. Marquez-Ruiz, M.C. Dobarganes, "Oxidative stability of sunflower oils differing in unsaturation degree during long-term storage at room temperature," *Journal of the American Oil Chemists' Society*, vol. 81, no. 6, pp. 577–583, 2004.

[7]. E.N. Frankel, "Lipid oxidation," *Prog. Lipid Res.* vol. 19, pp. 1-22, 1980.

[8]. USDA. (2010, November). National Nutrient Database for Standard Reference. *Nutrient Data Laboratory*. [Online] . Available: http://www.nal.usda.gov/fnic/foodcomp/search.

[9]. M. Otto, *Chemometrics: Statistics and Computer Application in Analytical Chemistry*. Wiley-VCH, 1999, pp. 201-203.

[10]. M.T. Hagan, H.B. Demuth, M.H. Beale, *Neural network design*. PNS Publishing*:* Boston, MA, 1996.

[11]. R. Kramer, *Chemometric Techniques for Quantitative Analysis*. Marcel Dekker, Inc., 270 Madison Avenue, New York, NY 10016, 1998.

[12]. P.C. Jurs, B.R. Kowalski, T.L. Isenhour, C.N. Reilly, "Computerized learning machines applied to chemical problems: Multicategory pattern classification by least squares," *Analytical Chemistry*, vol. 41, no. 6, pp. 695-700, 1969.

[13]. C. Jochum, P. Jochum, B.R. Kowalski, "Error propagation and optimal performance in multicomponent analysis," *Analytical Chemistry*, vol. 53, no. 1, pp. 85-92, 1981.

[14]. G.J. Newell, B. Lee, "Ridge regression: an alternative to multiple linear regression for highly correlated data," *Journal of Food Science*, vol. 46, no. 3, pp. 968-969, 2006.

[15]. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Data Handling in Science and Technology Volume 20 Handbook of Chemometrics and Qualimetrics: Part A.* Elsevier Science, 1997.

[16]. A.E. Hoerl, R.W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55-67, 1970.

[17]. F. Gunstone, R. Hamilton, *Oleochemical manufacture and applications*. Sheffield Academic Press, Sheffield S11 9AS, England, 2001.

[18]. J.V. Gerpen, B. Shanks, R. Pruszko, D. Clements, G. Knothe, , *Biodiesel production technology*. National Renewable Energy Laboratory Subcontractor Report NREL/SR-510-36244, 2004, ch.1, p 1.

[19]. A.G. Vereshagin, G.V. Novitskaya, "The triglyceride composition of linseed oil," *Journal of the American Oil Chemists' Society*, vol. 42, no. 11, pp. 970-974, 1965.

[20]. J. Studer, N. Purdie, J. Krouse, "Exploitation of friedel-crafts acylation as a color derivatizing reaction for terpene and lipids analyses," *Applied Spectroscopy*, vol .57, no. 7, pp. 784-790, 2003.

[21]. R. Kramer, *Chemometrics Toolbox for use with MATLAB*. The Mathworks, Inc.: Natick, Mass. 1994.

[22]. SAS. *JMP Statistical Discovery Software*, 8.0.2. SAS Institute, Inc.: Cary, NC, USA, 2009.

[23]. B. Rotthauser, G. Kraus, P.C. Schmidt, "Optimization of an effervescent tablet formulation containing spray dried L-leucine and polyethylene glycol 6000 as lubricants using a central composite design," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 46, pp. 85–94, 1998.

[24]. R.L. Tranter, R.L. (edtr), *Design and analysis in chemical research*.

ISBN: 978-988-19251-5-2 (revised on 20 September 2011) WCE 2011
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Sheffield Academic Press, 2000, pp. 250-251. [25]. L. Yan-Jie, X. Rong-Wu, "Application of central composite design/response surface methodology in pharmacy experiment design," *Chinese Journal of Modern Applied Pharmacy*, pp. 6-7, 2007

[26]. N.M. Faber, "Estimating the uncertainty in estimates of root mean square error of prediction: application to determining the size of an adequate test set in multivariate calibration," *Chemometrics and Intelligent Laboratory Systems*, vol. 49 , pp. 79–89, 1999.

[27]. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Data Handling in Science and Technology Volume 20 Handbook of Chemometrics and Qualimetrics: Part B*. Elsevier Science, 1997, p. 352.

[28]. D.A. Burns, E.W. Ciurczak, E.W. (edtrs), *Practical spectroscopy series volume 35: Handbook of near-infrared analysis 3rd edition*, CRC Press, 2008, p. 214.

[29]. I.E. Frank, J.H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, pp. 109-135, 1990

[30]. J. Cannady, "Artificial neural networks for misuse detection," in *National Information Systems Security Conference*, 1998, pp. 368–81

[31]. A. Garrido Frenich, M. Martinez Galera, J.L. Martinez Vidal, M.D. Gil Garcia, "Partial least-squares and principal component regression of multi-analyte high-performance liquid chromatography with diode-array detection," *Journal of Chromatography A*, vol. 727, pp. 27-38, 1996.

[32]. M.H. Zhang, Q.S. Xu, D.L. Massart, "Boosting partial least squares," *Anal. Chem.*, vol. 77 no. 5, pp. 1423-1431, 2005.

Date of modification: September 6, 2011
Brief description of changes: Address and email of Dr. Neil Purdie changed to OSU Chemistry Department, 018 Physical Science Building, Stillwater, OK 74078, USA (email: neil.purdie@okstate.edu)