Mathematical modelling of NIR spectral data and wavelength selection for determination of olive oil mixtures

Andrzej Miękina and Roman Z. Morawski
Warsaw University of Technology, Faculty of Electronics and Information Technology, Institute of Radioelectronics, Warsaw, Poland
E-mail: r.morawski@ire.pw.edu.pl

Abstract. The spectrophotometric analysis of oil mixtures, containing olive oil, is the subject of this paper. Its objective is to propose and evaluate a new method for wavelength selection aimed at optimisation of the numerical conditioning of the problem of determination of a selected component of such a mixture. The performance of the proposed methodology is assessed using semi-synthetic data and a criterion related to measurement uncertainty.

1. Introduction
The quality and purity of olive oil, extensively used in the food industry, is of significant commercial importance. According to the EU regulations, in force since 2002, a manufacturer of products, being based on or containing olive oil, must either indicate the share of olive oil in the total weight of the product or the percentage of olive oil as percentage of the total fat. That's why an increased interest in the methods for olive oil analysis has been observed for the last five years. Near-infrared (NIR) spectrophotometry, when combined with sophisticated procedures for spectrophotometric data processing, has turned out to be a simple and very flexible tool for this application; consult, for example, [1–5] for more details. Numerous existing methods of estimation are potentially suitable for this application [6], in particular – least-squares-type methods most frequently used in the domain of chemometrics. In [7], the authors have compared performance of six of them, when applied for estimation of the concentration of a selected component of an oil mixture, on the basis of the data representative of the NIR spectrum of this mixture, viz.: the ordinary least-squares estimator, the generalized least-squares estimator, the ridge least-squares estimator, the robust least-squares estimator, the total least-squares estimator and the partial least-squares estimator. Following the results of this comparison, the authors decided to focus on the ridge least-squares estimator, and to develop a methodology for its efficient application [8]. In this paper, after thorough analysis of the results presented in [7] and [8], taking into account relatively weak effectiveness of various mechanisms of regularisation studied there, the authors have returned to a very fundamental problem of data pre-processing, viz. wavelength selection.

For each sample, the spectrum is measured for many wavelength values (e.g. 1000). By selecting only most informative spectrum values (called predictors), one may improve the numerical conditioning of the problem of concentration estimation. Many techniques have been developed for this purpose; they can be classified into four categories:

---

1 To whom any correspondence should be addressed.
− the subset selection techniques, e.g. by means of the stepwise ordinary least squares (OLS) or of genetic algorithms coupled with OLS or with partial least squares (PLS);
− the dimension-wise selection techniques which work on a single dimension (principal component of or latent variable) of the regression technique;
− the model-wise elimination techniques which require repeated identification of the model on the basis of all the predictors but with only a fraction of the available samples or on the basis of all samples but only a fraction of the predictors;
− the interval selection techniques which target the wavelength interval rather than isolated wavelength values.

Exemplary implementations of various techniques may be found in [9–17]; in [10], nine techniques are compared, viz.: the stepwise OLS, the interactive stepwise elimination, the iterative predictors weighting, the elimination of uninformative variables, the generating optimal linear PLS estimation, the Martens uncertainty test, the Höskuldsson technique of most correlated predictors, the genetic algorithm-based ordinary least-squares, and the least absolute shrinkage and selection operator. Unfortunately, they do not yield satisfactory improvement in the considered case of olive-oil mixtures due to the extreme ill-conditioning of this problem; hence an attempt to develop a method of semi-exhaustive search described in this paper.

The following general rules are consistently used for generation of the mathematical symbols throughout this paper: \( x, y, \ldots \) are real-valued scalar variables; \( \hat{x}, \hat{y}, \ldots \) are exact values of \( x, y, \ldots \); \( \tilde{x}, \tilde{y}, \ldots \) are estimated values of \( x, y, \ldots \); \( \tilde{x}, \tilde{y}, \ldots \) are error-corrupted values of \( x, y, \ldots \). The diacritical signs, whose meaning has been explained above in reference to scalar variables, are applied in an analogous way with respect to vectors \( (x, y, \ldots) \) and matrices \( (X, Y, \ldots) \) of real-valued variables.

2. Problem formulation
Like in [7] and [8], it is assumed here that an oil mixture to be analysed is composed of 3 known components, and that the vectors of exact data \( s_j \) \((j = 1, 2, 3 < M)\), representative of the absorbance spectra of all those components, are available. According to the Lambert-Beer's law, the vector of exact absorbance data \( s \), representative of the spectrum of the mixture, satisfy the equation:

\[
\tilde{s} = c_1 \cdot \tilde{s}_1 + c_2 \cdot \tilde{s}_2 + c_3 \cdot \tilde{s}_3
\] (1)

where \( c = [c_1, c_2, c_3]^T \) is the vector of concentrations of all components, subject to the following constraints: \( c_1, c_2, c_3 \in [0, 1] \) and \( c_1 + c_2 + c_3 = 1 \). It is assumed that the real-world absorbance data \( \tilde{s} \), representative of the spectrum of a mixture, are corrupted by errors \( \Delta \tilde{s} \) resulting both from inaccurate preparation of the mixture and imperfections of the spectrophotometer. The estimation problem, addressed in this paper, consists in determination of the concentration of only one component of the mixture, viz. \( c_1 \), by means of a linear operator of the form:

\[
\tilde{c}_1 = \tilde{s}^T \cdot p_1
\] (2)

where \( p_1 = [p_{1,1} \ldots p_{1,M}]^T \) is a vector of parameters to be determined on the basis of a set of calibration data:

\[
\tilde{y}_{\text{cal}} = \left\{ \tilde{s}_{n,\text{cal}}, \tilde{c}_{1,n,\text{cal}}, \tilde{c}_{2,n,\text{cal}} \right\}_{n = 1, \ldots, N}^\text{cal}
\] (3)

which – for the sake of convenience of numerical manipulations – are organised in a matrix and two vectors:

\[
\tilde{S}_{\text{cal}} = \left[ \tilde{s}_{1,\text{cal}} \ldots \tilde{s}_{N,\text{cal}} \right]^T, \quad \tilde{c}_{1,\text{cal}} = \left[ \tilde{c}_{1,1,\text{cal}} \ldots \tilde{c}_{1,N,\text{cal}} \right]^T \quad \text{and} \quad \tilde{c}_{2,\text{cal}} = \left[ \tilde{c}_{2,1,\text{cal}} \ldots \tilde{c}_{2,N,\text{cal}} \right]^T
\] (4)
A methodology for wavelength selection, proposed in Section 4, has been tested under an assumption that the vector of parameters $\mathbf{p}_1$ is obtained during calibration by solving the overdetermined system of linear algebraic equations:

$$\widetilde{\mathbf{S}}_{\text{cal}} \cdot \mathbf{p}_1 \cong \mathbf{c}^{\text{val}}_1$$  \hspace{1cm} (5)

by means of the OLS method. Its performance is assessed on the basis of an indicator of worst-case uncertainty of the final result of measurement, using a set of validation data:

$$\mathbf{c}^{\text{val}}_n \cong \mathbf{S}_{\text{cal}} \cdot \mathbf{p}_1 \quad n = 1, \ldots, N$$  \hspace{1cm} (6)

which – for the sake of convenience of numerical manipulations – are organized in a matrix and two vectors:

$$\mathbf{S}^{\text{val}} = \begin{bmatrix} \mathbf{s}^{\text{val}}_1 & \ldots & \mathbf{s}^{\text{val}}_N \end{bmatrix}^T, \quad \mathbf{c}^{\text{val}}_1 = \begin{bmatrix} \mathbf{c}^{\text{val}}_{1,1} & \ldots & \mathbf{c}^{\text{val}}_{1,N} \end{bmatrix}^T$$ and $$\mathbf{c}^{\text{val}}_2 = \begin{bmatrix} \mathbf{c}^{\text{val}}_{2,1} & \ldots & \mathbf{c}^{\text{val}}_{2,N} \end{bmatrix}^T$$  \hspace{1cm} (7)

The study presented in this paper is based on semi-synthetic data representative of NIR spectra of trinary mixtures of edible oils.

3. Synthesis of data

The study has been based on the data generated using the real-world data representative of nut oil, corn oil and olive oil. The sequences of the latter data, each containing $N = 501$ points, are shown in Fig. 1. The data for calibration and validation have been synthesised according to the methodology proposed in [8], i.e. in a way imitating the laboratory procedure used for obtaining the real-world data. First, the reference (exact) values of the concentrations of nut oil ($\hat{c}_1$) and of corn oil ($\hat{c}_2$) have been selected, and the value of the concentration of olive oil has been calculated: $\hat{c}_3 = 1 - \hat{c}_1 - \hat{c}_2$. Next, the error-corrupted values of all concentrations have been determined by emulation of the process of sample preparation which consists in mixing the measured-out volumes of the components.

![Figure 1. The denoised and baseline-corrected spectrophotometric data used for experimentation.](image)

The error-free values of concentrations are related to the error-free values of those volumes in the following way:
\[ \hat{c}_j = \frac{\hat{V}_j}{\hat{V}} \quad \text{for } j = 1, 2, 3 \quad (8) \]

where \( \hat{V} = \hat{V}_1 + \hat{V}_2 + \hat{V}_3 \). Their error-corrupted values of concentrations are related to the error-corrupted values of those volumes in the following way:

\[ \bar{c}_j = \frac{\bar{V}_j}{\bar{V}} \quad \text{for } j = 1, 2, 3 \quad (9) \]

where: \( \bar{V} = \bar{V}_1 + \bar{V}_2 + \bar{V}_3 \) and \( \bar{V}_j = \bar{V}_j (1 + \bar{\vartheta}_j) \) with \( \bar{\vartheta}_j \) being relative errors of measuring out the volumes. The resulting concentrations \( \bar{c}_j \) are subject to relative errors \( \bar{\vartheta}_j \), which are related to the errors in the volume errors \( \bar{\vartheta}_j \) in the following way:

\[ \bar{c}_j = \hat{c}_j (1 + \bar{\vartheta}_j) = \left( \frac{\bar{V}_j}{V_1 + V_2 + V_3} \right) (1 + \bar{\vartheta}_j) = \hat{c}_j (1 + \vartheta_j) + \hat{c}_j (1 + \vartheta_j) + \hat{c}_j (1 + \vartheta_j) \quad (10) \]

The spectral data \( \tilde{s} = [\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n]^T \), corresponding to a given triplet of concentrations \( c_1, c_2, c_3 \) determined after Eq.(10) – have been generated after the formulae:

\[ \tilde{s} = c_1 \tilde{s}_1 + c_2 \tilde{s}_2 + c_3 \tilde{s}_3 + z \quad \text{and} \quad \tilde{s}_n = \tilde{s}_n + z_n \quad \text{for } n = 1, 2, \ldots \quad (11) \]

where \( \tilde{s}_1, \tilde{s}_2, \) and \( \tilde{s}_3 \) are the vectors of denoised and baseline-corrected real-world data representative of nut oil, corn oil and olive oil – respectively; \( z = [z_1, z_2, \ldots, z_n]^T \) is the vector of realisations of random variables modelling additive errors of spectrum measurement. For generation of the relative errors of the volumes (\( \vartheta_j \)) uncorrelated pseudorandom numbers following the zero-mean normal distribution with the standard deviation \( \sigma_{\vartheta} = 2.20 \cdot 10^{-3} \), truncated outside of the interval \([-3\sigma_{\vartheta}, 3\sigma_{\vartheta}]\), have been used; for generation of the errors of spectrum measurement (\( z_n \)) – uncorrelated pseudorandom numbers following the zero-mean normal distribution with the standard deviation \( \sigma_z = 3.33 \cdot 10^{-5} \), truncated outside of the interval \([-3\sigma_z, 3\sigma_z]\). The set of data for calibration \( \tilde{y}_{val} \) has been assumed to contain all the pairs \( N = 121 \) of the following values of concentrations:

\[ c_1^{cal}, c_2^{cal} \in \left\{ k \cdot 0.01 \mid k = 0, 1, \ldots, 10 \right\} \quad (12) \]

and the corresponding spectral data. The set of data for validation \( \tilde{y}_{val} \) has been assumed to contain all the pairs \( N' = 100 \) of the following values of concentrations:

\[ c_1^{val}, c_2^{val} \in \left\{ 0.005 + k \cdot 0.01 \mid k = 0, 1, \ldots, 9 \right\} \quad (13) \]

and the corresponding spectral data. Thus, the validation has been carried out over an area of the \( c_1 - c_2 \) plane, slightly smaller that the area covered by the calibration data. In this way, the impact of border effects has been mitigated. Taking into account that the three largest singular values of the matrix \( \tilde{S}_{121 \times 501}^{val} \) (composed of all spectral data available for calibration) are 64.16, 0.088 and 0.023 – while the forth one and others are of order \( 10^{-14} \) – three wavelength values have been assumed sufficient for this study. They have been selected using a methodology proposed in the next section.

4. Proposed methodology for wavelength selection

The proposed procedure for wavelength selection consists of two steps: the pre-selection of wavelength values, and the selection of optimum triplets of wavelengths values. The first step is resulting in a subset of 30 most promising wavelength values; the second one consists in reviewing all the triplets of those values and selection of the best of them using a criterion directly related to the accuracy of concentration estimation.
Two methods have been alternatively used for the pre-selection of wavelength values, viz.:

- PM1 based on three normalised eigenvectors of the matrix \( \tilde{S}^{cal} \)
- PM2 using correlations of the vector \( \tilde{\epsilon}^{cal}_1 \) with all the columns of the full matrix \( \tilde{S}^{cal} \).

The normalisation of the eigenvectors used in PM1 is based on the following representation of the "exact" OLS estimate of the parameters:

\[
\hat{\mathbf{p}}_1 = (\tilde{\mathbf{S}}^{cal})^T \cdot \tilde{\mathbf{c}}^{cal}_1 = \mathbf{V} \cdot \Sigma^+ \cdot \mathbf{U}^T \cdot \tilde{\mathbf{c}}^{cal}_1 = \sum_{n=1}^{N} \frac{\mathbf{u}^T_n \cdot \tilde{\mathbf{c}}^{cal}_1}{\sigma_n} \cdot \tilde{\mathbf{v}}_n \tag{14}
\]

where \( \mathbf{V}, \Sigma^+ \) and \( \mathbf{U} \) are matrices resulting from singular-value decomposition of the matrix \( \tilde{S}^{cal} \). This formula suggests that the eigenvectors \( \tilde{\mathbf{v}}_n \) with weights \( \frac{\mathbf{u}^T_n \cdot \tilde{\mathbf{c}}^{cal}_1}{\sigma_n} \) should be compared rather than the eigenvectors \( \tilde{\mathbf{v}}_n \) itself. Since those weights for \( \tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2 \) and \( \tilde{\mathbf{v}}_3 \) are (respectively) –0.0078, 2.95 and 5.4, the vector \( \tilde{\mathbf{v}}_1 \) may be excluded from considerations. The correspondence of the wavelength values and elements of the normalised vectors \( \tilde{\mathbf{v}}_2 \) and \( \tilde{\mathbf{v}}_3 \) is shown in Fig. 2. The correlations of the vector \( \tilde{\mathbf{c}}^{cal}_1 \) with the columns of the matrix \( \tilde{S}^{cal} \), used in PM2 have been computed using the MATLAB function \texttt{corrcoef}; the result is shown in Fig. 3. The weighted values of \( \tilde{\mathbf{v}}_2 \) and \( \tilde{\mathbf{v}}_3 \), corresponding to 30 most promising wavelength values, selected using PM1, are larger than 0.4. The values of the correlation coefficient, corresponding to 30 most promising wavelength values, selected using PM2, are larger than 0.8.

![Figure 2](image_url)  
**Figure 2.** The correspondence of the wavelength values and elements of the normalised vectors \( \tilde{\mathbf{v}}_2 \) and \( \tilde{\mathbf{v}}_3 \) (with the nut oil spectrum in the background).

The second step of the proposed methodology consists in ordering the triplets of wavelength values \( (\lambda_1, \lambda_2, \lambda_3) \) using the following performance criterion:
Figure 3. The correlations of the vector $\hat{c}^{\text{val}}$ with the columns of the matrix $\tilde{S}^{\text{cal}}$ (black line) with the nut oil spectrum in the background (blue line).

$$\tilde{\Lambda}_i = \left\| \tilde{\Lambda} \tilde{c}^{\text{val}} \right\|_\infty = \left\| \tilde{S}^{\text{cal}} \cdot \tilde{S}^{\text{cal}} \cdot \tilde{c}^{\text{cal}} - \hat{c}^{\text{val}} \right\|_\infty$$ for $i = 1, \ldots, I$ \hspace{1cm} (15)

In this formula: $\tilde{S}^{\text{cal}}_{121\times3}$ ($i = 1, \ldots, I$) is a calibration matrix corresponding to an evaluated triplet, $\tilde{S}^{\text{val}}_{121\times3}$ ($i = 1, \ldots, I$) is a validation matrix corresponding to that triplet, and $\tilde{S}^{\text{cal}} \cdot \tilde{c}^{\text{cal}}$ is an estimate of the vector $p_i$, resulting from calibration based on the matrix $\tilde{S}^{\text{cal}}$ and the vector $\tilde{c}^{\text{cal}}$; consequently, $\tilde{S}^{\text{val}} \cdot \tilde{S}^{\text{val}} \cdot \tilde{c}^{\text{val}}$ is the corresponding estimate of the vector of concentrations used for validation $\hat{c}^{\text{val}}$, and $\tilde{S}^{\text{cal}} \cdot \tilde{c}^{\text{cal}} - \hat{c}^{\text{val}}$ is the error of validation.

5. Performance of the proposed methodology for wavelength selection

The performance of the above described methodology of wavelength selection has been assessed using $R = 100$ versions of the set of data $\widetilde{D}^{\text{cal}}$ and $R' = 100$ versions of the set of data $\widetilde{D}^{\text{val}}$, each corresponding to a different realisation of errors. For each version $\widetilde{D}^{\text{cal}}(r)$, $r = 1, \ldots, R$, of the set $\widetilde{D}^{\text{cal}}$, the full calibration has been performed, and its results have been validated by means of $R$ versions $\widetilde{D}^{\text{val}}(r')$ of $\widetilde{D}^{\text{val}}$, $r' = 1, \ldots, R'$. The worst-case error of estimation, calculated for each concentration value used for validation:

$$\overline{\Delta}_i = \sup \left\{ \left| \hat{c}^{\text{val}}(r, r') - \hat{c}^{\text{val}}_i \right| : \ r = 1, \ldots, R, \ r' = 1, \ldots, R' \right\}$$ for $i = 1, \ldots, I$ \hspace{1cm} (16)

has been used as a performance indicator of a triplet under study.

The exhaustive evaluation of all triplets of wavelength values pre-selected by means of PM1 has resulted in identification of the best triplet $\{\bar{\lambda}_1 = 1710 \text{ nm}, \bar{\lambda}_2 = 1766 \text{ nm}, \bar{\lambda}_3 = 2306 \text{ nm}\}$ for which $\overline{\Delta}_i = 0.00497$, while the exhaustive evaluation of all triplets of wavelength values pre-selected by
means of PM2 has indicated the triplet \( \langle \lambda_1 = 1710 \text{ nm}, \lambda_2 = 2120 \text{ nm}, \lambda_3 = 2306 \text{ nm} \rangle \) for which \( \bar{\Delta}_i = 0.00527 \).

The best wavelength triplets, obtained using PM1 for pre-selection and the criterion \( \hat{\Delta}_i \) for triplet selection, are shown in Table 1; the best wavelength triplets, obtained using PM2 and \( \hat{\Delta}_i \), are shown in Table 2. The correspondence of the indicator \( \bar{\Delta}_i \) with the selection criterion \( \hat{\Delta}_i \) is demonstrated in the last two rows of both tables. The coefficients of correlation of the sequences \( \{\bar{\Delta}_i \}_{i=1,...,30} \) and \( \{\hat{\Delta}_i \}_{i=1,...,30} \), computed for PM1 and PM2, are \( \rho_{\text{PM1}} = 0.51 \) and \( \rho_{\text{PM2}} = 0.56 \), respectively. The values of \( \bar{\Delta}_i \), presented in both tables, have been compared with the value of this indicator obtained for the wavelength triplet used for experimentation in [8], viz.: \( \langle \lambda_1 = 1762 \text{ nm}, \lambda_2 = 2130 \text{ nm}, \lambda_3 = 2136 \text{ nm} \rangle \). Each of the triplets indicated in both tables is guaranteeing at least 30\% reduction of measurement uncertainty.

### Table 1. The best triplets of wavelength values selected on the basis of PM1-based pre-selection.

| Triplet #1 | Triplet #2 | Triplet #3 | Triplet #4 | Triplet #5 | Triplet #6 | Triplet #7 | Triplet #8 | Triplet #9 | Triplet #10 |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|
| \( \lambda_1 \) [nm] | 1706       | 1710       | 1710       | 1702       | 1706       | 1706       | 1766       | 1840       | 1710        | 1706        |
| \( \lambda_2 \) [nm] | 1766       | 1800       | 1766       | 1766       | 1798       | 1766       | 2306       | 2306       | 1766        | 1766        |
| \( \lambda_3 \) [nm] | 2306       | 2306       | 2306       | 2306       | 2310       | 2324       | 2324       | 2308       | 2308        | 2308        |
| \( \bar{\Delta}_i \) | 0.00345    | 0.00349    | 0.00349    | 0.00352    | 0.00380    | 0.00383    | 0.00389    | 0.00395    | 0.00396     | 0.00401     |
| \( \bar{\Delta}_i \) | 0.00497    | 0.00590    | 0.00561    | 0.00674    | 0.00665    | 0.00600    | 0.00762    | 0.00654    | 0.00654     | 0.00654     |

### Table 2. The best triplets of wavelength values selected on the basis of PM2-based pre-selection.

| Triplet #1 | Triplet #2 | Triplet #3 | Triplet #4 | Triplet #5 | Triplet #6 | Triplet #7 | Triplet #8 | Triplet #9 | Triplet #10 |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|
| \( \lambda_1 \) [nm] | 1704       | 1712       | 1700       | 1704       | 1704       | 1700       | 1712       | 1712       | 1708        | 1700        |
| \( \lambda_2 \) [nm] | 2306       | 2306       | 2306       | 2306       | 2306       | 2306       | 2306       | 2306       | 2306        | 2306        |
| \( \lambda_3 \) [nm] | 2348       | 2346       | 2346       | 2346       | 2352       | 2348       | 2352       | 2348       | 2346        | 2352        |
| \( \bar{\Delta}_i \) | 0.00294    | 0.00297    | 0.00301    | 0.00316    | 0.00321    | 0.00331    | 0.00344    | 0.00347    | 0.00354     | 0.00354     |
| \( \bar{\Delta}_i \) | 0.00544    | 0.00575    | 0.00576    | 0.00559    | 0.00595    | 0.00540    | 0.00624    | 0.00550    | 0.00585     | 0.00582     |

6. **Conclusions**

The main contribution of this paper to the development of methods for solving difficult numerical problems, related to the determination of the mixtures of edible oils on the basis of NIR spectrophotometric data, consists in significant reduction of the estimation uncertainty (by ca. 30\% in comparison to [8]) via a novel methodology of wavelength selection. Two alternative, numerically simple but reliable, procedures have been proposed for this purpose. They have turned out comparable both in terms of their potential for reduction of measurement uncertainty and in terms of numerical complexity. Both of them depend of the method used of calibration, but the adaptation of their version, developed and presented here for the OLS estimator, affects only the form of the performance criterion, defined by Eq. (15).

**Acknowledgements**

The study, presented in this paper, has been supported by the Ministry of Science and Higher Education in Poland (grant No. N N505 464832). The authors express their sincere gratitude to Dr. Grażyna Zofia Żukowska from the Faculty of Chemistry, Warsaw University of Technology, for the acquisition of data used for numerical experimentation reported in this paper.
References

[1] Tay A, Singh R K, Krishnan S S and Gore J P 2002 Authentication of olive oil adulterated with vegetable oils using Fourier transform infrared spectroscopy, Lebensm.-Wiss. u.-Technol. 35, pp 99–103

[2] Christy A A, Kasemsumran S, Du Y and Ozaki Y 2004 The detection and quantification of adulteration in olive oil by near-infrared spectroscopy and chemometrics, Anal. Sci. 20, pp 935–940

[3] Armenta S, Garrigues S and de la Guardia M 2007 Determination of edible oil parameters by near infrared spectrometry, Anal. Chim. Acta 596, pp 330–337

[4] Özdemir D and Öztürk B 2007 Near infrared spectroscopic determination of olive oil adulteration with sunflower and corn oil, J. Food & Drug Anal. 15, pp 40–47

[5] Sinelli N, Casiraghi E, Tura D and Downey G 2007 Characterization and classification of Italian virgin olive oils by near and medium infrared spectroscopy, Proc. 13th Int. Conf. Near Infrared Spectroscopy (Umeå-Vasa, Sweden & Finland, June 15-21, 2007), paper #4-6

[6] Morawski R Z 2006 Spectrophotometric applications of digital signal processing, Meas. Sci. Technol. 17, pp R117–R144

[7] Latała A and Morawski R Z 2008 Comparison of LS-type methods for determination of olive oil mixtures on the basis of NIR spectral data, Metrol.&Meas.Syst. XV, pp 409–420

[8] Miękina A and Morawski R Z 2009 A calibration method, based on piecewise ridge LS estimator, designed for determination of olive oil mixtures on the basis of NIR spectral data, Proc. XIX IMEKO World Congress (Lisbon, Portugal, September 6–11, 2009), pp 2559–2563 (CD-ROM)

[9] Alonso-Lomillo M A, Dominguez-Renedo O and Arcos-Martinez M J 2004 Resolution of binary mixtures of rifamycin SV and rifampicin by UV/VIS spectroscopy and partial least-squares method (PLS), Chem. & Biodiv. 1, pp 1336–1343

[10] Baskir I M and Drozd A V 2003 New Matlab software for wavelength selection, Chemometr.& Intell. Lab. Syst. 66, 2003, pp 89–91

[11] Du Y P, Liang Y Z, Jiang J H, Berry R J and Ozaki Y 2004 Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares, Anal. Chim. Acta 501, pp 183–191

[12] Hiroaki I, Toyonori N and Eiji T 2002 Measurement of pesticide residues in food based on diffuse reflectance IR spectroscopy, IEEE Trans. Instrum. & Meas. 51, pp 886–890

[13] Kasemsumran S, Du Y P, Murayama K, Huehne M and Ozaki Y 2004 Near-infrared spectroscopic determination of human serum albumin, gamma-globulin, and glucose in a control serum solution with searching combination moving window partial least squares*, Anal. Chim. Acta 512, pp 223–230

[14] McShane M J, Coté G L and Spiegelman C 1997 Variable selection in multivariate calibration of a spectroscopic glucose sensor, Appl. Spectrosc. 51, pp 1559–1564

[15] Udelhoven T, Novozhilov M and Schmitt J 2003 The NeuroDeveloperR: a tool for modular neural classification of spectroscopic data*, Chemometr. & Intell. Lab. Syst. 66, pp 219–226

[16] Forina M, Lanteri S, Cerrato Oliveros M C and Pizarro Millan C 2004 Selection of useful predictors in multivariate calibration, Anal. & Bioanal. Chem. 380, pp 397–418

[17] Igne B and Hurburgh C R 2010 Local chemometrics for samples and variables: optimizing calibration and standardization processes, Journal of Chemometrics, 24, pp 75–86