PAN

POLSKA AKADEMIA NAUK

M&MS

MEASUREMENT DATA PROCESSING IN SPECTROPHOTOMETRIC ANALYSERS OF FOOD
Review paper

Roman Z. Morawski

Warsaw University of Technology, Faculty of Electronics and Information Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland
(✉ r.morawski@ire.pw.edu.pl, +48 22 234 7721)

Abstract

Spectrometry, especially spectrophotometry, is getting more and more often the method of choice not only in laboratory analysis of (bio)chemical substances, but also in the off-laboratory identification and testing of physical properties of various products, in particular – of various organic mixtures including food products and ingredients. Specialised spectrophotometers, called spectrophotometric analysers, are designed for such applications. This paper is on the state of the art in the domain of data processing in spectrophotometric analysers of food (including beverages). The following issues are covered: methodological background of food analysis, physical and metrological principles of spectrophotometry, the role of measurement data processing in spectrophotometry. General considerations are illustrated with examples, predominantly related to wine and olive oil analysis.

Keywords: spectrophotometry, chemometrics, spectral data processing, food analysis, wine analysis.

1. Introduction

The Committee on Higher Education of the Institute of Food Technologists – the largest international, non-profit professional organisation involved in the advancement of food science and technology – has provided the following definition of food science: *"Food science is the discipline in which the engineering, biological, and physical sciences are used to study the nature of foods, the causes of deterioration, the principles underlying food processing, and the improvement of foods for the consuming public."* [1]. This definition is emphasising the empirical nature of food science, and – consequently – its strong affinity to measurement and instrumentation. That is the aspect of food science to be discussed in this review paper. A range of measurement methods and techniques, that are currently applied in food science and technology, is encompassing both simple tools and procedures for measuring mass or temperature, and very sophisticated ones – such as computer-based measuring systems combining liquid chromatography (separation of mixture compounds) with mass spectrometry (elemental analysis).

This paper is devoted to the methods for processing measurement data in spectrophotometric analysers of food (including beverages). The rapid development of such instruments, observed since the last decades of the XXth century, is implied both by a spectacular technological progress in the domain of spectrophotometry (spectrophotometric transducers, instruments and systems) and by growing demand for adequate methods of food analysis. The latter is driven by several factors; the most important among them is the growing awareness of consumers concerning the nutritional value of food products, which – in particular – is fuelling public debates over genetically modified food ingredients and

organic food products as well as on food scandals which, despite the legislative efforts, aimed at protecting consumers and ensuring fair trade, are reported quite frequently. In most cases, their origin is related to pesticide residues, veterinary drug residues, endocrine disruptors, processing contaminants, packaging materials and natural toxins.

A successful fusion of two powerful technologies: near-infrared spectrophotometry and measurement data processing, has provided a quite universal solution for the analysis and testing of many alimentary products, especially outside analytical laboratories. Spectrophotometric analysers for grain and fruits, milk and beer, chocolate and cheese, *etc.*, are today manufactured by numerous companies all over the world. Their development is possible due to the intensive research on chemometric methods dedicated to the analysis of such raw materials and products as grape fruits and wine [2-26], olive fruits and olive oil [27-58], other fruits and juices [59-75], vinegars [76-79], cheeses and other milk products [80, Chapter 4], [81-90], eggs [80, Chapter 5], [91-92], meat [93-103], honey [80, Chapter 6], [104-107], chocolate [108], and tea [109-110].

This paper contains an outline of a broad engineering and application background of that technology and a more detailed review of the state of the art in the domain of measurement data processing dedicated to spectrophotometric analysers of food. It is focused on a review of problems rather than of literature; the latter is referred to rather selectively – only to provide illustrations of reviewed problems; an exhaustive listing and analysis of the literature devoted to spectrophotometric analysers of food would require a book rather than a journal paper.

The paper is a significantly restructured, updated and extended version of the author's conference paper of 2008 [111]; it is reporting – in particular – the progress in the domain of spectrophotometric data processing, accomplished during last four years.

2. Spectrophotometric analysers of food

Spectrophotometric tools, and – consequently – the methods for interpretation of spectrophotometric data, are of increasing importance for analytical laboratories, as well as for environmental, biomedical and industrial monitoring. On the one hand, the development of corresponding applications is driven by a growing demand for this kind of tools, the demand implied *i.a.* by the advancement of standards related to environment protection, health care, individual and collective security, as well as by the widespread use of optical means for inspection of industrial processes. On the other hand, however, this development is due to the market availability of miniature spectrophotometers: mini- and micro-spectrophotometers [112].

The name *spectrophotometric analyser* is used for various spectrophotometric sensors, devices, instruments, probes and testers designed for measuring physical and/or chemical parameters characterising a pre-defined class of chemical or biochemical substances. In each version, an analyser must contain a spectrophotometric transducer (ST) converting an optical signal into a sequence of raw data $\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{y}_1 \ ... \ \tilde{y}_N \end{bmatrix}^T$ representative of the spectrum $x(\lambda)$ of that signal, where λ is wavelength, x is light intensity, N is the number of data, and the tilde placed over $y_1, ..., y_N$ is to indicate that the data are subject to various disturbances of external and internal origin. The wavelength values may cover a broader or narrower subrange $[\lambda_{\min}, \lambda_{\max}]$ of one of the following standard ranges: 200-300 nm – middle-ultraviolet radiation (MUV), 300-380 nm – near-ultraviolet radiation (UV), 380-750 nm – visible radiation (Vis), 750-2,500 nm – near-infrared radiation (NIR) or 2.5-10 μm – middle-infrared radiation (MIR). Several physical principles and corresponding devices may be used for the design of STs:

− a dispersive element (a grating or a linear variable filter) that enables separation of spectral

components in space;
− a tuneable filter that enables separation of spectral components in time;
− an optical heterodyne that enables shifting the spectrum to a wavelength range where its analysis is easier;
− an interferometer providing the data whose Fourier transform is representative of the spectrum.

The first of them, a dispersive element, is applied today most frequently. Referring to Fig. 1, one may explain its functioning as follows:
− An optical signal to be converted into a digital representation of its spectrum is separated into N optical signals, corresponding to narrow subranges of wavelength λ by a dispersive element, *e.g.* a grating.
− Each of those signals is reaching a corresponding photodetector (PD) converting it into a current proportional to its intensity.
− The output currents of the photodetector matrix are scanned by an analogue-to-digital converter and, one by one, coded.
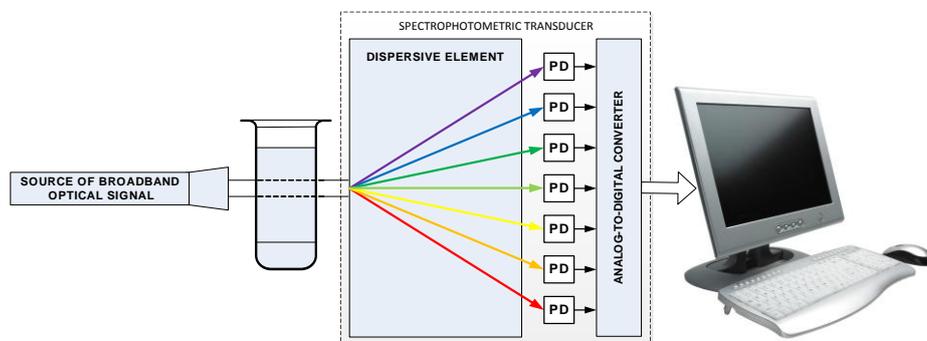

Fig. 1. A spectrophotometer based on the ST with a dispersive element:
the principle of functioning.

Depending on application, one of the following modes of acquisition of spectral data may be applied in an analyser: transmittance mode, interactance mode, transflectance mode, diffuse transmittance mode or diffuse reflectance mode. Further considerations will be limited here to the first of them which is most frequently used in SAFs. This is sufficient because the mode of data acquisition has a very limited impact on the methods and algorithms of their processing. A functional diagram of a spectrophotometer, based on the ST with a dispersive element and working in the transmittance mode, is shown in Fig. 1. Its operation may be explained as follows:
− First, the source optical signal is converted into the data $\tilde{\mathbf{y}}_0$ representative of its intensity spectrum $x_0(\lambda)$.
− Next, the same source optical signal is passed through a cuvette, containing a sample of the substance to be analysed, and converted into the data $\tilde{\mathbf{y}}_1$ representative of its intensity spectrum, modified by a sample, $x_1(\lambda)$.
− Finally, the intensity data $\tilde{\mathbf{y}}_0$ and $\tilde{\mathbf{y}}_1$ are processed by a digital processor (a controller, a microprocessor, a digital signal processor or a personal computer) in order to obtain an estimate of the target parameters of the sample, such as concentrations $c_1, ..., c_J$ of selected compounds of that sample.

The estimation of target parameters is, as a rule, an ill-conditioned numerical problem of *measurand reconstruction* [113-114]. Therefore, its solution may require not only simple

conversion of codes and estimation of the transmittance or absorbance spectrum on the basis of two intensity spectra, but also some more sophisticated operations, aimed at extraction of information hidden in the data, such as estimation of the optical noise level, estimation of the spectrum with a better resolution than the resolution of the ST or estimation of concentrations on the basis of preprocessed data. In some cases, estimation of special intermediate or auxiliary parameters, such as positions and magnitudes of spectral peaks, may be necessary, because the information on the positions of peaks in the absorbance spectrum may facilitate identification of the compounds which are present in the sample; and the information on their magnitudes may facilitate estimation of the concentrations of the identified compounds.

The complexity of spectrophotometric data processing significantly depends on the technological and functional imperfections of the ST. Their external manifestations – such as noise-type disturbances in the data or blurring and overlapping of spectral peaks – must be taken into account during data processing. Therefore, the corresponding algorithms, as a rule, require considerable amount of *a priori* information on the mathematical model of the ST (or on the mathematical model of the data at its output). This information is acquired during an operation, most frequently performed off-line, called *calibration of the ST*; it is comprising not only scaling of wavelength axis and intensity axis, but also the identification of a forward or inverse mathematical model of the ST on the basis of the ST data corresponding to some standards of the spectrum or of a class of substances the analyser is designed for.

The interpretation of the intensity spectra $x_0(\lambda)$ and $x_1(\lambda)$, in terms of concentrations (or other physical or (bio)chemical parameters) characterising a sample under study, may be simplified if those spectra are transformed into the transmittance spectrum defined as:

$$x^{Tr}(\lambda) = x_1(\lambda)/x_0(\lambda). \tag{1}$$

The corresponding data $\tilde{\mathbf{y}}_0$ and $\tilde{\mathbf{y}}_1$ should be then converted into the transmittance data representative of $x^{Tr}(\lambda)$ according to the formula:

$$\tilde{y}_n^{Tr} = \frac{\tilde{y}_{1,n} - b_n}{\tilde{y}_{0,n} - b_n} \quad \text{for } n = 1, \ldots, N, \tag{2}$$

where $\mathbf{b} = [b_1 \ldots b_N]^T$ is the ST response to the zero-intensity optical signal. The interpretability of spectra may be further improved by using the absorbance scale defined as follows:

$$x^{Ab}(\lambda) = -\log_{10}\big(x^{Tr}(\lambda)\big) \tag{3}$$

and the corresponding absorbance data:

$$\tilde{y}_n^{Ab} = -\log_{10}\big(\tilde{y}_n^{Tr}\big) \quad \text{for } n = 1, \ldots, N. \tag{4}$$
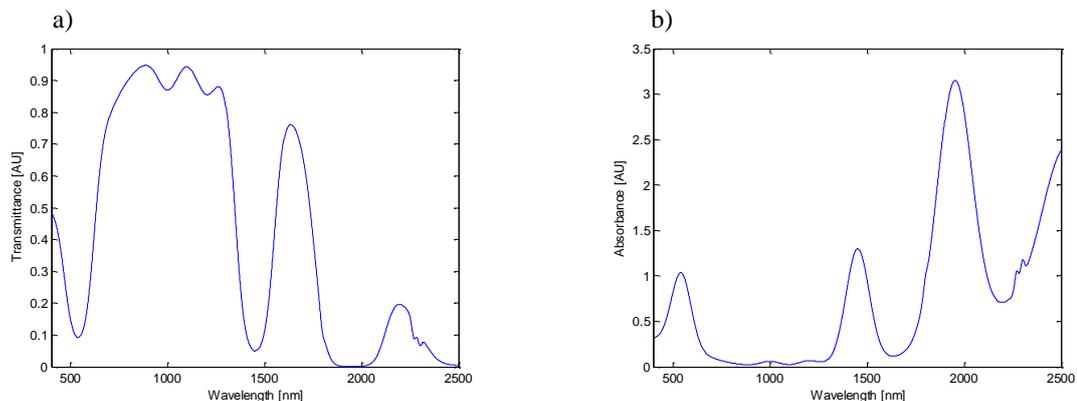

Fig. 2. The transmittance (a) and absorbance (b) data representative of the spectrum of red wine.

An example of transmittance data is shown in Fig. 2a, and the result of their transformation into absorbance data – in Fig. 2b. The absorbance data emphasise the most informative parts of the spectrum, *viz.* absorption peaks whose parameters, positions and magnitudes, depend on the qualitative and quantitative composition of a sample under study. Their graphical representation is therefore a finger-print enabling the identification of that sample.

The name *spectrophotometric analyser of food* (SAF) is used for various spectrophotometric sensors, devices, instruments, probes and testers dedicated to measuring physical and (bio)chemical parameters characterising food products (including beverages). The NIR spectrophotometry is of particular usefulness for food analysis because spectra of organic samples comprise broad bands arising from overlapping of absorption peaks corresponding to the C-H, O-H and N-H chemical bonds. The main advantage of NIR spectrophotometry, when applied for routine analysis of food, is its simplicity and speed: usually no sample preparation is necessary, and the time of analysis is not greater than 1 minute.

Another advantage of NIR spectrophotometry is that it allows several constituents to be identified concurrently. Finally, the relatively weak absorption due to water enables one to analyse high-moisture food products and ingredients. Depending on the category of food to be analysed, the range of analysis may be extended towards the UV-Vis range or the MIR range.

The driving forces behind the development of SAFs may be easily recognised on the basis of the general structure of food business, shown in Fig. 3 and of the specification of needs of its main actors, provided in Table 1.

All those actors are potential users of SAFs, but the measurement needs are today better defined, and the use of SAFs is more common, among the food producers and supervisors of food business than among the providers of raw materials, food distributors, food vendors and individual consumers. It may be predicted, however, that the significant lowering of the prices of SAFs will quickly increase the demand among the members of the latter group.

The main reasons for using SAFs in the food business may be summarised as follows: checking the quality of food, monitoring of the food production process, providing data necessary for production control, specification of food products necessary for their labelling, and precise classification of food products enabling their better pricing. There are more-or-less evident economic benefits behind each of them. Precise classification, for example, enables one to minimise losses due to aging of food by replacing a worst-case approach with a realistic-case approach since the selection and grading of food may be based on the objective measurement results rather than on the "best-before" date.
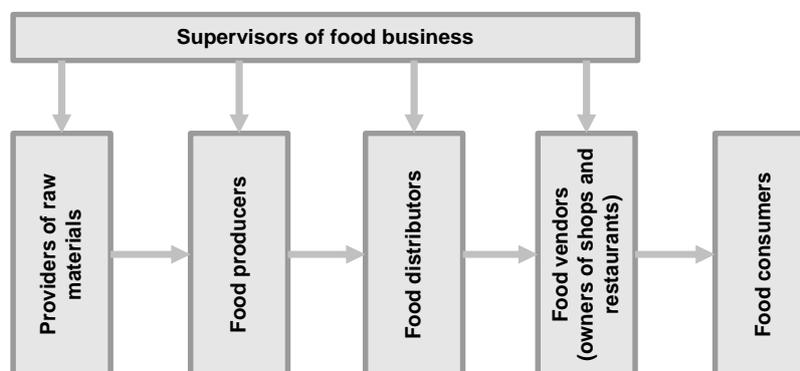

Fig. 3. The general structure of the food business.

Table 1. The motivation for using SAFs by the main actors of food business.

	Specific examples of SAFs	Reasons to use SAFs
Providers of raw materials	analysers of fruits ripeness, moisture content in grains, meat tenderness, raw milk;	checking the quality of raw materials, specification of raw materials enabling their labelling, precise classification of raw materials enabling their better pricing;
Food producers	analysers of olive oil: chemical contents, olive oil: geographical origin, chocolate products, cheese;	monitoring of the production process from the verification and selection of raw materials to the final products, providing data necessary for manufacturing control, specification of final products enabling their labelling, precise classification of final products enabling their better pricing;
Food distributors	analysers of flour, honey: chemical contents, honey: floral origin, coffee grains;	checking the quality of products to be distributed, monitoring of the quality of stored products, precise classification of products to be distributed, enabling their better pricing, increased productivity;
Food vendors	analysers of water (esp. for restaurants), bread, dairy products, fruits (aging);	checking the quality of products to be sold, monitoring of the quality of stored products, precise classification of products for sale, enabling their better pricing,
Food consumers	analysers of mineral water, wine, beer, tee;	checking the quality of purchased products, monitoring of the quality of stored products, improved selectivity in purchasing food products (based on better quality price evaluation);
Supervisors	analysers for microbiologial evaluation of food, toxicological evaluation of food.	checking the quality of food products *in situ*, increased effectiveness of supervision.

3. Mathematical modelling of spectrophotometric data

The operation principle of any SAF is based on the sensitivity of the spectrum $x(\lambda)$ to a measurand being a vector – let's denote it with the symbol $\mathbf{c}$ – containing quantities characterising a sample of food to be analysed: concentrations of some compounds of that sample, and – optionally – some quantities specific of the category of food the SAF is designed for. A good example of such a quantity is the indicator of acidity pH in case of wine [26], fruit vinegars [77], yogurt [115], cheese [81], pork [102], kiwi fruit [71] or tomato [62]. In this Section, the methodology for mathematical modelling of the relationship $\mathbf{c} \leftrightarrow x(\lambda)$ is outlined under an assumption that all components of the measurand are concentrations. This methodology may be generalised on the measurands containing some other quantities, provided the corresponding physical, chemical or purely numerical models of the sensitivity of the spectrum $x(\lambda)$ to those quantities are available.

The contents of this and three following sections at many points refer to the author's review paper of 2006 [116], where many more algorithmic details may be found.

3.1. Forward models of intensity data

The intensity data, provided by an ST, may be modelled using a white-box approach, a black-box approach, or a grey-box approach combining some advantages of white-box and black-box approaches [117]. The experience behind modelling of spectrophotometric data, accumulated since the appearance of a seminal paper published by Pieter Hendrik van Cittert in 1931 [118], seems to prove that the approximation power of the so-called *Wiener operator* (a superposition of a linear integral operator with a nonlinear algebraic operator) is sufficient for adequate modelling of the relationship between the intensity spectrum $x(\lambda)$ and the corresponding raw spectral data $\tilde{\mathbf{y}}$. Let the variable y_n denote the mathematical model of the "noise-free" version of the datum $\tilde{y}_n$; then this operator may be given the following form:

$$y_n = F\left(\int_{-\infty}^{+\infty} g_n(\lambda_n - \lambda)\, x(\lambda)\, d\lambda; \boldsymbol{\alpha}_n \right) \quad \text{for } n = 1, \ldots, N, \tag{5}$$

where: $g_n(\lambda)$ is the response of the ST, measured at the output of the nth photodiode, to a tuneable monochromator producing an optical signal whose spectrum is close to $\delta(\lambda - l)$ with l moving from $\lambda_{\min}$ to $\lambda_{\max}$; $F(\bullet; \boldsymbol{\alpha}_n)$ is an *a priori* known function (*e.g.* an algebraic polynomial, a cubic spline) whose parameters, $\boldsymbol{\alpha}_n = [\alpha_{n,1}\ \alpha_{n,2}\ ...]^T$, have to be determined during the ST calibration.

3.2. Forward models of absorbance data

The reasoning presented above for the intensity data may be repeated for the transmittance data. As a rule, due to the compensation of some irregularities by the division of corresponding intensity data, the function $F(\bullet; \boldsymbol{\alpha}_n)$ may be less complex in this case, and the variability of $\boldsymbol{\alpha}_n$ and $g_n(\lambda)$ along the wavelength axis – less important. Consequently, quite often this variability is neglected, and the corresponding model of the transmittance data is simplified by fixing: $\boldsymbol{\alpha}_n \equiv \boldsymbol{\alpha}$ and $g_n(\lambda) \equiv g(\lambda)$ for $n = 1, ..., N$. The following mathematical model of the relationship between the transmittance spectrum $x^{Tr}(\lambda)$ and the corresponding data $\tilde{\mathbf{y}}^{Tr} = [\tilde{y}_1^{Tr} ... \tilde{y}_N^{Tr}]^T$ is then applied:

$$\tilde{y}_n^{Tr} = \int_{-\infty}^{+\infty} g(\lambda_n - \lambda) x^{Tr}(\lambda) d\lambda + \eta_n \quad \text{for } n = 1, \ldots, N, \tag{6}$$

where the additive residuals η_n represent the total uncertainty of data modelling.

For modelling the dependence of the data on concentrations, the Lambert-Beer laws of absorption are used most frequently, *viz.* the relationship between $x^{Tr}(\lambda)$ in (6) and the vector of concentrations $\mathbf{c} \equiv [c_1 ... c_J]^T$ is established, taking into account that the absorbance of a solution of a single compound is proportional to its concentration, and the absorbance spectrum of a solution containing J compounds:

$$x^{Ab}(\lambda) \equiv -\log_{10}[x^{Tr}(\lambda)] \tag{7}$$

is equal to the linear combination of the normalised absorbance spectra of those compounds $x_1^{Ab}(\lambda), ..., x_J^{Ab}(\lambda)$:

$$x^{Ab}(\lambda) = c_1 x_1^{Ab}(\lambda) + ... + c_J x_J^{Ab}(\lambda). \tag{8}$$

The last three equations – (6), (7) and (8) – have to be discretised to become a useful basis for the development of numerical methods for estimation of concentrations. Thus, (6) is replaced with:

$$\tilde{\mathbf{y}}^{Tr} = \mathbf{G} \cdot \mathbf{x}^{Tr} + \boldsymbol{\eta}, \tag{9}$$

where: $\boldsymbol{\eta} \equiv [\eta_1 ... \eta_N]^T$, and:

$$\mathbf{x}^{Tr} \equiv [x^{Tr}(\lambda_1') ... x^{Tr}(\lambda_M')]^T \quad \text{with } \lambda_m' \equiv \lambda_{\min} + (m-1)\frac{\lambda_{\max} - \lambda_{\min}}{M-1} \quad \text{for } m = 1, \ldots, M, \tag{10}$$

and $\mathbf{G}$ is a rectangular matrix whose values depend on the values of the function $g(\lambda)$ and on the chosen method of numerical integration. The discretisation of (7) and (8), consistent with (9), yields:

$$\mathbf{x}^{Ab} = \left[-\log_{10}\left(x^{Tr}(\lambda_1')\right) \ldots -\log_{10}\left(x^{Tr}(\lambda_M')\right)\right]^T = c_1 \mathbf{x}_1^{Ab} + \ldots + c_J \mathbf{x}_J^{Ab} = \mathbf{X}^{Ab}\mathbf{c}, \qquad (11)$$

where:

$$\mathbf{x}_j^{Ab} \equiv \left[x_j^{Ab}(\lambda_1') \ldots x_j^{Ab}(\lambda_M')\right]^T \quad \text{for } j = 1, \ldots, J \quad \text{and} \quad \mathbf{X}^{Ab} \equiv \left[\mathbf{x}_1^{Ab} \ldots \mathbf{x}_J^{Ab}\right]^T. \qquad (12)$$

The above-developed model of the relationship $\mathbf{c} \to \tilde{\mathbf{y}}^{Tr}$ characterises this relationship for one particular pair of $\mathbf{c}$ and $\tilde{\mathbf{y}}^{Tr}$. It should be generalised on the populations of possible vectors $\mathbf{c}$ and $\tilde{\mathbf{y}}^{Tr}$ to make feasible the use of probabilistic tools of data processing. This is usually achieved by introducing random vectors (denoted hereinafter with underlined symbols) for modelling unknowns. The randomisation of the vector $\mathbf{\eta}$, representative of modelling uncertainty, yields:

$$\underline{\tilde{\mathbf{y}}}^{Tr} = \mathbf{G} \cdot \mathbf{x}^{Tr} + \underline{\mathbf{\eta}}, \qquad (13)$$

where $\underline{\mathbf{\eta}} \equiv \left[\underline{\eta}_1 \ldots \underline{\eta}_N\right]^T$. The model of the relationship between concentrations and absorbance spectrum, corresponding to (11), takes on the form:

$$\underline{\mathbf{x}}^{Ab} = \mathbf{X}^{Ab} \cdot \underline{\mathbf{c}} + \underline{\mathbf{\epsilon}}, \qquad (14)$$

where $\underline{\mathbf{c}} \equiv \left[\underline{c}_1 \ldots \underline{c}_J\right]^T$ is the random vector modelling the concentrations to be estimated, $\underline{\mathbf{\epsilon}} \equiv \left[\underline{\epsilon}_1 \ldots \underline{\epsilon}_M\right]^T$ is the random vector modelling the residual spectrum corresponding to unexpected or neglected compounds, and CAT is the operator of absorbance-to-transmittance conversion:

$$\underline{\mathbf{x}}^{Tr} \equiv \left[\underline{x}_1^{Tr} \ldots \underline{x}_M^{Tr}\right]^T = \mathrm{CAT}\left[\underline{\mathbf{x}}^{Ab}\right] = \left[10^{-\underline{x}_1^{Ab}} \ldots 10^{-\underline{x}_M^{Ab}}\right]^T. \qquad (15)$$

The substitution of (14) to (15), and then of (15) to (13) yields a compact-form, discretised and randomised model of the data:

$$\underline{\tilde{\mathbf{y}}}^{Tr} = \mathbf{G} \cdot \mathrm{CAT}\left[\mathbf{X}^{Ab} \cdot \underline{\mathbf{c}} + \underline{\mathbf{\epsilon}}\right] + \underline{\mathbf{\eta}}. \qquad (16)$$

3.3. Inverse models of absorbance data

The model of absorbance data, corresponding to (14) has the form:

$$\underline{\tilde{\mathbf{y}}}^{Ab} = \mathrm{CAT}^{-1}\left[\mathbf{G} \cdot \mathrm{CAT}\left[\mathbf{X}^{Ab} \cdot \underline{\mathbf{c}} + \underline{\mathbf{\epsilon}}\right] + \underline{\mathbf{\eta}}\right]. \qquad (17)$$

Its inversion with respect to $\underline{\mathbf{c}}$ yields a nonlinear function $\underline{\mathbf{c}} = \Phi\left(\underline{\tilde{\mathbf{y}}}^{Ab}, \underline{\mathbf{\epsilon}}, \underline{\mathbf{\eta}}\right)$ which hardly can be used for spectrophotometric data processing. Fortunately, in the majority of practically important cases, it may be approximated by a known function linear with respect to unknown parameters $\mathbf{P}$:

$$\underline{\mathbf{c}} = \hat{\Phi}\left(\underline{\tilde{\mathbf{y}}}^{Ab}, \underline{\mathbf{\epsilon}}, \underline{\mathbf{\eta}}; \mathbf{P}\right) \qquad (18)$$

or even with respect to $\underline{\tilde{\mathbf{y}}}^{Ab}$ and to a residual vector $\rho(\underline{\mathbf{\epsilon}}, \underline{\mathbf{\eta}})$ whose impact on $\underline{\mathbf{c}}$ is – according to some criteria – equivalent to that of $\underline{\mathbf{\epsilon}}$ and $\underline{\mathbf{\eta}}$:

www.czasopisma.pan.pl PAN www.journals.pan.pl
POLSKA AKADEMIA NAUK

Metrol. Meas. Syst., Vol. XIX (2012), No. 4, pp. 623-652.

$$\underline{\mathbf{c}} = \mathbf{P} \cdot \underline{\widetilde{\mathbf{y}}}^{Ab} + \rho(\underline{\boldsymbol{\varepsilon}}, \underline{\boldsymbol{\eta}}). \tag{19}$$

The matrix $\mathbf{P}\big|_{J \times N}$ contains parameters which should be estimated during calibration of the ST.

4. Spectrophotometric data interpretation

4.1. Two principal approaches

Any procedure of spectrophotometric data interpretation, aimed at estimation of concentrations, refers to a more or less complex mathematical model of the ST and related model of those data [119-120]. The identification of that model is the main objective of the ST calibration. There are two basic approaches of the problem of spectrophotometric data interpretation that imply substantially different solutions for calibration: the forward-model-based approach (FMA) and the inverse-model-based approach (IMA). The first of them consists in identification of a forward model of the ST, *i.e.* in estimation of the parameters of an operator mapping the space of real-valued functions $x(\lambda; \mathbf{c})$, or the space of vectors $\mathbf{c}$, into the space of vectors $\widetilde{\mathbf{y}}^{Ab}$. The IMA is aimed at identification of an inverse model of the ST, *i.e.* at estimation of the parameters of an operator, mapping the space of vectors $\widetilde{\mathbf{y}}^{Ab}$ into the space of real-valued functions $x(\lambda; \mathbf{c})$ or the space of vectors $\mathbf{c}$. In both cases, a set of reference data:

$$\widetilde{\mathsf{D}}^{cal} = \left\{ \widetilde{\mathbf{c}}_n^{cal}, \widetilde{\mathbf{y}}_n^{Ab,cal} \mid n = 1, 2, \ldots, N^{cal} \right\} \tag{20}$$

is necessary to attain the desired validity of the model. It should contain concentrations $\widetilde{\mathbf{c}}_n^{cal}$, characterising reference samples used for calibration and measured by reference instrumentation, and the corresponding spectral data $\widetilde{\mathbf{y}}_n^{Ab,cal}$ acquired by means of the calibrated ST.

The estimation of the measurand $\mathbf{c}$ according to the FMA consists in the direct use of the forward model defined by (17) or – much more frequently – of its linearised version. The corresponding calibration procedure is in this case limited to the estimation of the matrices $\mathbf{G}$ or $\mathbf{H}$ and $\mathbf{X}^{Ab}$ on the basis of the reference data acquired by means of the ST subject to calibration and by means of an ST of significantly higher resolution; if necessary, some *a priori* information (*e.g.* covariance matrices) characterising the random vectors $\underline{\boldsymbol{\varepsilon}}$ and $\underline{\boldsymbol{\eta}}$ or $\mathbf{H} \cdot \underline{\boldsymbol{\varepsilon}} + \underline{\boldsymbol{\eta}}$ is also extracted from those data. An estimate of the measurand $\mathbf{c}$, corresponding to a vector of raw spectral data $\widetilde{\mathbf{y}}^{Ab}$ is obtained by solving the following system of algebraic equations:

$$\mathrm{CAT}^{-1}\Big[\hat{\mathbf{G}} \cdot \mathrm{CAT}\big[\hat{\mathbf{X}}^{Ab} \cdot \mathbf{c}\big]\Big] \cong \widetilde{\mathbf{y}}^{Ab} \tag{21}$$

or:

$$\hat{\mathbf{H}} \cdot \hat{\mathbf{X}}^{Ab} \cdot \mathbf{c} \cong \widetilde{\mathbf{y}}^{Ab}, \tag{22}$$

where $\hat{\mathbf{G}}$, $\hat{\mathbf{H}}$ and $\hat{\mathbf{X}}^{Ab}$ are the estimates of $\mathbf{G}$, $\mathbf{H}$ and $\mathbf{X}^{Ab}$, obtained during calibration. The *a priori* information on $\underline{\boldsymbol{\varepsilon}}$ and $\underline{\boldsymbol{\eta}}$ or $\mathbf{H} \cdot \underline{\boldsymbol{\varepsilon}} + \underline{\boldsymbol{\eta}}$, acquired during calibration, may be used for regularisation of that numerical problem.

The estimation of the measurand $\mathbf{c}$ according to the linear IMA consists in execution of the following operation:

$$\hat{\mathbf{c}} = \hat{\mathbf{P}} \cdot \tilde{\mathbf{y}}^{Ab}, \qquad (23)$$

where $\hat{\mathbf{P}}$ is an estimate of the matrix $\mathbf{P}$, obtained during calibration as a regularised solution to the system of linear algebraic equations:

$$\mathbf{P} \cdot \tilde{\mathbf{y}}_n^{Ab,cal} \cong \tilde{\mathbf{c}}_n^{cal} \quad \text{for } n = 1, 2, \ldots, N^{cal}. \qquad (24)$$

If the concentrations of all compounds are to be estimated, then this system is given the form:

$$\tilde{\mathbf{S}}^{cal} \cdot \mathbf{P}^T \cong \tilde{\mathbf{C}}^{cal}, \qquad (25)$$

where:

$$\tilde{\mathbf{S}}^{cal}\Big|_{N^{cal} \times N} \equiv \begin{bmatrix} \tilde{\mathbf{y}}_1^{Ab,cal} & \ldots & \tilde{\mathbf{y}}_{N^{cal}}^{Ab,cal} \end{bmatrix}^T \quad \text{and} \quad \tilde{\mathbf{C}}^{cal}\Big|_{N^{cal} \times J} \equiv \begin{bmatrix} \tilde{\mathbf{c}}_1^{cal} & \ldots & \tilde{\mathbf{c}}_{N^{cal}}^{cal} \end{bmatrix}^T. \qquad (26)$$

If the concentration of a selected compound, *e.g.* of the j th compound, is to be estimated, then that system of equations may be reduced to the form:

$$\tilde{\mathbf{S}}^{cal} \cdot \text{col}_j\big[\mathbf{P}^T\big] \cong \text{col}_j\big[\tilde{\mathbf{C}}^{cal}\big], \qquad (27)$$

where $\text{col}_j[\circ]$ is the operator of extraction of the j th column of a matrix.

4.2. Methodological considerations

From computational point of view, the key numerical problem – common for both approaches – is the necessity to solve a system of linear algebraic equations: $\tilde{\mathbf{A}} \cdot \mathbf{x} \cong \tilde{\mathbf{b}}$, whose elements are defined by (22) in case of the FMA, and by (25) or (27) in case of the IMA. The uncertainty of the solution in both cases depends on the structure and properties of the matrix $\tilde{\mathbf{A}}$ and on the level of errors in the data $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ – the errors defined with respect to their hypothetical exact versions $\dot{\mathbf{A}}$ and $\dot{\mathbf{b}}$.

In case of the FMA, the system is as a rule overdetermined because $\dim[\tilde{\mathbf{A}}] = N \times J$ and the number of data representative of a single spectrum (N) is on the whole greater than the number of concentrations to be estimated (J). Since each column of the matrix $\tilde{\mathbf{A}}$ contains in this case the data representative of the spectrum of a single compound, the conditioning number of this matrix depends on the similarity of compound spectra: the more they differ, the smaller it is. The level of errors in the data $\tilde{\mathbf{A}}$ can be made negligible by repeated measurements of compound spectra and appropriate averaging included in the calibration procedure.

In case of the IMA, the system is as a rule underdetermined because $\dim[\tilde{\mathbf{A}}] = N^{cal} \times N$ and the number of data representative of a single spectrum (N) is on the whole greater than the number of reference samples used for calibration (N^{cal}) due to the high costs of their preparation. For the same reason, the reduction of errors in the data is more problematic than in case of the FMA. Consequently, both source errors and coefficients of their amplification are larger than in case of the FMA unless regularisation measures are applied. Moreover, the final result of measurand estimation is in case of the IMA subject to uncertainties inherited from the estimate $\hat{\mathbf{P}}$ of the parameters $\mathbf{P}$ and from the raw data $\tilde{\mathbf{y}}^{Ab}$. Despite all those apparent drawbacks, the IMA is more and more frequently applied due to its better robustness to the impact of deviations of the mathematical models – described in the previous sections – "from reality". For this reason, the considerations to follow will be focused on the IMA-based

www.czasopisma.pan.pl   PAN   www.journals.pan.pl
POLSKA AKADEMIA NAUK

Metrol. Meas. Syst., Vol. XIX (2012), No. 4, pp. 623-652.

methods of spectrometric data processing.

In most cases, the dimension of the vector $\mathbf{c}$ (1-25) is significantly smaller than the number of the data in the vector $\tilde{\mathbf{y}}^{Ab}$ (200-1000) and the data corresponding to the neighbour wavelength values are highly correlated; they may be also correlated from one reference sample to another. This serial correlation decreases the apparent information in the data; but – if a method applied takes into account this serial correlation – it is possible to turn the correlation to a benefit by taking advantage of the redundancy. Therefore, a methodology for solving those problems, as a rule, comprises two steps: the compression of the data $\tilde{\mathbf{y}}^{Ab}$, *i.e.* their transformation into an estimate $\hat{\mathbf{z}}_y$ of a low-dimensional vector of informative parameters $\mathbf{z}_y$:

$$\tilde{\mathbf{y}}^{Ab} \rightarrow \hat{\mathbf{z}}_y \qquad (28)$$

and subsequent estimation of the concentrations on the basis of $\hat{\mathbf{z}}_y$:

$$\hat{\mathbf{z}}_y \rightarrow \hat{\mathbf{c}}. \qquad (29)$$

The complexity of both steps depends on the number of informative parameters: the greater it is, the larger is the time required both for the ST calibration and for estimation of concentrations. The ST calibration, performed on the basis of the reference data $\tilde{D}^{cal}$, is aimed at estimation of the parameters of the operator of data compression, defined by (28), and of the operator of measurand estimation, defined by (29), provided the structure of both operators is chosen in advance. A large variety of algorithms can be generated by combining various techniques of data compression with various techniques of estimation. The compression consists, as a rule, in the selection of the most informative among the parameters computed according to various principles, such as: projection of the data $\tilde{\mathbf{y}}^{Ab}$ on a set of linearly independent vectors; parameterisation of the data $\tilde{\mathbf{y}}^{Ab}$, *i.e.* their approximation by means of a known function $\hat{y}(\lambda_n; \mathbf{z}_y)$ with unknown parameters $\mathbf{z}_y$; estimation of the moments of $\tilde{\mathbf{y}}^{Ab}$; the estimation of latent variables, such as principal components. The first approach may be exemplified with the author's recent paper [121], where a very simple and effective method of compression has been proposed, *viz.* the method which consists in multiplication of the vector $\tilde{\mathbf{y}}^{Ab}$ by a matrix $\mathbf{W}|_{K \times N}$ ($K < N$) whose rows contain Walsh functions.

The most self-imposing tool for estimation of concentrations is a neural network being a universal approximator [122-123], but – for the reason of efficiency – multidimensional B-splines and wavelet transforms are viable alternatives in many practical cases. Chemometric examples of the methods for selection of the most informative variables or elimination of non-informative variables are provided in the references [124-127].

5. Calibration of spectrometric transducers

5.1. Methodology of inverse-model-based calibration

The univariate calibration is, by definition, related to modelling the dependence of the concentration of a single compound on a single spectral datum [128], [129, Section 5.2], while multivariate calibration might involve determining the measurand using the spectral data acquired for tens or hundreds of wavelength values. A general scheme of multivariate calibration, applied in practice, comprises the following steps:

- selection of the reference samples (mixtures) to be used for identification of the model;
- acquisition and visual, or computer-aided [130], evaluation of the spectral data used for calibration, before and after preprocessing;
- a first modelling trial to decide whether it is possible to attain the expected quality of the model and whether non-linearity should be introduced in this model;
- iterative refinement of the model, *e.g.* by considering elimination of possible outliers with respect to the model, selecting the model complexity;
- final validation of the model.

The optimum complexity of the model, resulting from calibration, is a key issue. As the complexity of the model is increasing, the prediction error assessed for the data set $\tilde{\mathsf{D}}^{cal}$ is diminishing. To determine the optimum number of significant components in a series of spectral data, one may look at how the error is decreasing when that number is growing [131-134]. By increasing the complexity of the model, one may reduce this error almost to zero, but at the same time – make the model produce artefacts when it is applied to unknown samples. The final validation of the model is an operation that consists in testing the prediction power of the chosen model: an additional set of data $\tilde{\mathsf{D}}^{val}$, independent of $\tilde{\mathsf{D}}^{cal}$ but having the same structure, should be used for this purpose:

$$\tilde{\mathsf{D}}^{val} = \left\{ \tilde{\mathbf{c}}_n^{val}, \tilde{\mathbf{y}}_n^{val} \mid n = 1, \dots, N^{val} \right\}. \tag{30}$$

As already mentioned, the problem of estimation of the parameter matrix $\mathbf{P}$ is numerically ill-conditioned because of the collinearity in the data matrix $\tilde{\mathbf{A}}$ [135]. Two kinds of measures are applied to remediate this difficulty, *viz.*: selection of samples and wavelength values used for calibration (which are discussed in the next subsection), and so-called *soft modelling*. The first step of the latter consists in expressing the data in terms of a new set of axes based on the different contributions to variation in the data. Most frequently, it is based on covariance in the data $\tilde{\mathbf{A}}$: a set of orthogonal components, the so-called latent variables, being linear combinations of the independent spectral data, is created to describe independent sources of the observed variations in those data. They are, as a rule, ordered in terms of the magnitude of the independent sources of variation that they explain: the first latent variable explains the largest independent source of variance in the data, the second latent variable – the second largest, *etc.* The second step in soft modelling is the elimination of non-informative latent variables. Generally, re-expression of the data $\tilde{\mathbf{A}}$ in terms of latent variables is not very useful unless a decision is made on the number of latent variables that are sufficient to adequately explain the systematic variation in the data $\tilde{\mathbf{A}}$. There are numerous references containing detailed reviews of the methodology and practice of soft modelling, both for experts and for beginners [136-138].

No single method for solving linear algebraic equations has turned out the best in all spectrophotometric applications. Sometimes, the simplest tool, *i.e.* the ordinary least squares (OLS) estimator is sufficient, much more frequently, however – more sophisticated one must be applied. Therefore, several alternative methods have been developed; among them the partial-least-squares (PLS) estimator is probably the most commonly acknowledged. It has gained much popularity among the researchers in chemometrics, and – consequently – has been enhanced and modified in various ways, including the incorporation of some nonlinearity and constraints, such as the positivity of spectral data and the positivity of concentrations of compounds [5-6, 9-10, 12-13, 21-22, 24, 26, 28, 35, 37-38, 42, 44-46, 52-56, 61, 67-68, 71, 73, 77-79, 82, 84, 86, 88, 92-93, 101, 105-106, 108, 139-180]. The PLS estimator is a relatively complex (sophisticated) mathematical tool, both from logical and algorithmic point of view – see, for example, its traditional description in the handbooks [129,

Section 5.5], [181, Section 3.5], [182, Section 4.7] or its probabilistic interpretation in the paper [183]. Its widespread use is most probably motivated by the availability of numerous library procedures of PLS – such as those in the *PLS_Toolbox* by Eigenvector Research, Inc. [184] – rather than by the real need. As shown in a recent comparative study of five least-squares estimators of concentrations [185], in many cases much simpler tools – such as ridge least-squares (RiLS) estimator, the total-least-squares (TLS) estimator or even the OLS estimator – may provide results of comparable metrological quality as those generated by the PLS estimator. The RiLS estimator has turned out in this study to be a real alternative for the PLS estimator since it provided better results than the latter in the middle-range values of concentrations. This is a one more confirmation for an earlier conclusion that it may be an effective remedy for numerical ill-conditioning of the parameter estimation problems [186–187]; this conclusion applies even more to the generalised RiLS estimator where each variable gets a slightly different regularisation parameter [188].

It should be noted that the linear-model approach may fail when considerably nonlinear relationships have to be taken into account. Numerous linearisation techniques have been proposed to deal with such situations: some of them work on the data and others work on the model or on some model parameters. In general, the artificial neural networks are natural tools for dealing with nonlinear models. They are used more and more frequently in spectrophotometric data processing, as a rule in combination with various data compression techniques, such as principal component analysis (PCA), to avoid excessive overfitting [189].

5.2. Selection of samples and wavelength values

The multivariate calibration requires standards, *i.e.* samples for which the estimates of concentrations, obtained by a reference method, are known. For the cost and/or time reasons, the number of such standards cannot be too large (usually not greater than 50 or 100). Since the model has to be used for the prediction on new samples, all possible sources of variation that can be encountered later must be included in the calibration data $\tilde{D}^{cal}$. This means that the chemical compounds present in the samples to be analysed must be included in the samples used for calibration, and the range of variation in their concentrations should be at least as wide as that expected of the samples to be analysed. There is, however, a practical limit on what is available. It is, therefore, necessary to achieve a compromise between the number of samples to be analysed and the prediction error that can be attained. When it is possible to artificially generate a number of samples, experimental design can and should be used to decide on the composition of the calibration samples. In most cases only real samples are available, so that an experimental design is not possible. This is the case for the analysis of food products and ingredients.

There are several strategies available for selection of the calibration samples representative of the problem to be solved. The simplest of them is random selection, but it is open to the possibility that some sources of variation may be lost. Another strategy refers to *a priori* knowledge about the problem under study: if all the sources of variation are known, then the calibration samples can be selected on the basis of that knowledge. A more detailed information on selection of samples may be found in the literature [34, 63, 190-191]. For each sample, the spectral data are recorded for many wavelength values. An important opportunity to improve the numerical conditioning of the estimation problem is the proper selection of those values. It can be based on *a priori* knowledge of the most informative wavelength values, derived from previous experience, or may be performed by means of special techniques of selection such as the stepwise OLS estimator. A more detailed information on wavelength selection may be found in the abundant literature [24, 35, 67, 107-108, 159-160, 166-167, 175, 178, 180], [182, Section 4.5], [192-198].

6. Preprocessing of spectrophotometric data

6.1. Objectives of spectrophotometric data preprocessing

The estimation of the concentrations, directly from the acquired spectrophotometric data, is in many practically important cases problematic due to instrumental imperfections of the ST and to some quantum phenomena in the analysed substance that produce blurring in the data representative of the measured spectrum. Both effects may be significantly reduced by means of the algorithms of deconvolution or generalised deconvolution. As a rule, those algorithms are designed and assessed using the criteria of the quality of deconvolution, which are not specific of spectrophotometry, such as the root-mean-square error of approximation. Consequently, the estimates of the concentrations, determined on the basis of the results produced by those algorithms, may turn out to be poor. On the other hand, the variational algorithms, used for estimation of concentrations, may be very inefficient if not provided with a good initial guess of the sought-for estimates [199].

A significant improvement may be attained by taking into account that the positions of spectral peaks are carrying information on the compounds of the analysed substance, and the magnitudes of peaks − on their concentrations. In a series of the author's publications [200-202], a methodology for estimation of the concentrations, on the basis of spectrophotometric data, has been developed − the methodology comprising the compression of spectral data to the estimates of the positions and magnitudes of spectral peaks, and qualitative identification of the compounds on the basis of those estimates. The ratio of compression attained in this way may be very high (100 or more); for example: 1024 data points representative of the Vis-NIR spectrum of *Satsuma mandarin* may be replaced with 6 pairs of the coordinates of the most informative absorption peaks [69]. Not only parameters of spectral peaks, but also some other morphological features of the spectrum of an analyte may be used for identification of its contents. In the paper [203], for example, the concentrations are estimated using a procedure based on an assumption that the list of possible compounds is known *a priori*: first, the spectral data are twice differentiated and partitioned into segments along the wavelength axis; next, the presence or absence of the compounds from the predefined list is checked on the basis of the binary codes of the segments, using a neural network; finally, the concentrations are estimated on the basis of the data corresponding to the most informative parts of the spectrum, using another neural network.

The main aims of spectrophotometric data preprocessing may be summarised as follows:
− the elimination or suppression of those features of the spectral data that are not related to the concentrations;
− the enhancement of those features of the spectral data that are carrying information on the concentrations;
− the improvement of the resolution.

Some methods of calibration, *e.g.* the PCA-based compression followed by the OLS parameter estimation, are derived or designed under an assumption that the mean of spectral data is zero and their standard deviation is equal to one. The application of those methods must be preceded by centring and scaling of spectral data. It should be, however, noted that, as shown in the author's recent paper [185], this standardisation may occur useless in other cases.

6.2. Suppression of undesirable features of spectrophotometric data

The first objective of data preprocessing is traditionally attained by reduction of noise present in the data, by elimination of the non-informative part of the spectrum called *baseline*

www.czasopisma.pan.pl   PAN   www.journals.pan.pl
POLSKA AKADEMIA NAUK

Metrol. Meas. Syst., Vol. XIX (2012), No. 4, pp. 623-652.

or *background*, and by elimination of outliers. More recent are methods for elimination of systematic variations in the spectral data, being not correlated to the concentrations, regardless of the nature of those variations: orthogonal signal correction [204] and pre-whitening of spectral data by their covariance-weighted filtering [205].

The noise in spectral data is most frequently reduced by means of moving-window averaging filters [206], in particular – by means of the Savitzky-Golay filter [166, 175], [182, Section 7.2], [195]. Alternatively, it may be suppressed with the filters defined in the domain of the discrete Fourier transform or in the domain of wavelet transform [207-208]. Another way of reducing noise is the repeated acquisition of the data and their averaging [206]; the signal-to-noise ratio increases then with the square root of the number of data vectors. The tutorial information on the noise reduction in spectral data may be found in a series of internet papers [209], more advanced – in the Ph.D. thesis [210].

The baseline may be informally defined as a component of the data, slowly-changing with wavelength, whose subtraction from the data leaves spectral peaks placed at the zero level of intensity or absorbance. There are numerous heuristic methods for baseline subtraction, referring to this informal definition. For example: constant baseline differences can be eliminated by using offset correction (each vector of spectral data is corrected by subtracting either the absorbance value at a selected wavelength or the mean value in a selected wavelength range); the constant and linear components of the baseline may be removed by double differentiation of the data. Further examples may be found in the literature [211-215]. Outliers are extreme, very large or very small, measurement results. If they are retained in the data, one may falsely conclude that they do not follow a normal distribution. Among various ways for overcoming this problem, the simplest consist in removing the outliers, more advanced – in using special criteria for fitting a model.

6.3. Enhancement of informative features of spectrophotometric data

The second aim of data preprocessing is attained mainly by their numerical differentiation which is enhancing spectral differences. Both first and second derivatives are used, but the second derivative seems to be applied more frequently [182, Section 7.2]. An important drawback of the use of derivatives is the decrease of the signal-to-noise ratio implied by the amplification of noise (for that reason, smoothing is needed before differentiation). Another drawback is that models, obtained as a result of calibration, based on spectral data preprocessed by differentiation, are sometimes less robust to instrumental changes, such as wavelength shifts which may occur over time. The tutorial information on the methods of data differentiation may be found in the references [206, 216], while reviews of recent developments – in [210-211], and examples of applications – in [5, 45, 55, 66, 68, 71, 101-102, 106, 166, 168, 170, 217-222].

6.4. Improvement of spectral data resolution

If a method, used for estimation of the measurand $\mathbf{c}$, is ignoring the limitations of the optical resolution of the ST, then the preprocessing of spectral data, increasing their resolution, may reduce the uncertainty of the estimates $\hat{\mathbf{c}}$. This operation, called spectrum reconstruction, consists in estimation of the spectrum $x(\lambda)$ on the basis of the data $\tilde{\mathbf{y}}$. The convolution-type model of spectrometric data, defined by (6), and – consequently – the methods of deconvolution, may be used for this purpose. In the majority of practically interesting cases, the problem of spectrum estimation is numerically ill-conditioned, *i.e.* oversensitive to the errors in the data. Consequently, the direct numerical approach of (6) yields, as a rule, solutions buried in the noise "inherited" from the data with an amplification

coefficient reaching the values of hundreds, thousands or millions. Thus, the problem of spectrum reconstruction, although very simple from the logical point of view, is not trivial and far from being definitively solved in practice. Numerous sophisticated algorithms for dealing with it have been developed since 1931 when a first iterative algorithm of deconvolution for improving resolution of spectrometric measurements was proposed [118]. None of them, nevertheless, has turned out to be the best in all applications under all possible measurement conditions. Consequently, a complex methodology for solving problems of spectrum reconstruction has been developed, and comparative studies of the methods have been continually carried out [223-227]. This methodology refers to the idea of a numerical mechanism, called regularisation, that consists in modification of the initial problem in such a way as to make it less sensitive to the errors in the data.

There are various general-purpose methods of regularisation – all consist in constraining the set of admissible solutions. They may be classified into the following groups: direct methods, variational methods, probabilistic methods, iterative methods, parametric methods, and transformation-based methods [113]. As a rule, the methods of reconstruction, met in practice, are combinations of some of these elementary methods of regularisation. The rapid increase in computing power available for dealing with practical tasks of measurement data processing has radically changed views on the applicability of various mathematical methods and algorithms for solving problems of spectrum reconstruction. The speed and accuracy of computing – offered by general-purpose computers, digital signal processors or even application-specific processors – make possible implementation of the algorithms which for decades have been considered too complex for practical applications. First of all, optimisation-based approaches of reconstruction problems have become very common. They produce efficient variational algorithms of calibration or reconstruction – due to the strong regularising properties, the possibility of combining all elementary mechanisms of regularisation, and the convenience of incorporating additional constraints of the set of feasible solutions, deduced from *a priori* information on the solution $x(\lambda)$ and on the disturbances in the data $\tilde{\mathbf{y}}$. A variational algorithm of spectrum reconstruction is a numerical implementation of the operation:

$$\hat{x}(\lambda) = \arg_{x(\lambda)} \text{opt}\{ \mathsf{J} \ [x(\lambda); \hat{\mathbf{p}}] \,|\, x(\lambda) \in \mathsf{X} \}, \tag{31}$$

where J is a criterion of optimisation, and X is a set of admissible solutions, defined on the basis of *a priori* information derived from physical or technical context of the measurement task under consideration. A key problem to be solved in this case is the choice of the criterion J in such a way as to mathematically express our expectations concerning good spectrum reconstruction.

7. Example: spectrophotometric analysis of wine

Instruments, designed for measurement of ethanol concentration and sugar concentration in wine, were used by wine makers already in the first half of the XIXth century, and started to be their common tools after 1857 when Louis Pasteur explained the biochemical nature of fermentation. However, more complete analysis of wine had to wait *ca.* 100 years for the development of such modern analytical techniques as liquid chromatography and mass spectrometry. The time for spectrophotometry came by the end of the 1980's. In 2006, this measurement technique was applied by a Japanese research team for prototyping a robot able to check whether the content of a bottle of wine corresponds to its label [228]. Automatic analysers of beverages, based on various measurement principles, had been known long before 2006, but that achievement opened a new chapter in their development since wine

testing is a task much more complex than testing of any other kind of alimentary product, especially when it is not limited to the differentiation between white and red wine or between French and Chilean *Cabernet Sauvignon...* A robot for wine testing is expected to confirm, with sufficiently low uncertainty, that the content of a bottle has all the features promised by its label, including the geographical origin and vintage of wine, the sort of grapes it has been made of, and the type of barrels it has been aged in. Of course, still much information will remain behind the scene: sensory properties of wine, identified by keepers using organoleptic means, and its chemical properties, identified by oenology expert equipped with sophisticated analytical instruments such as systems of liquid chromatography combined with mass spectrometry.

For obvious economic reasons, the research progress in the domain of wine analysis is quite spectacular: the robot for checking bottles of wine may soon become an effective tool for fighting the fraud on the wine markets. A simple example is sufficient to show the financial dimensions of this social issue. Let's assume that 900 bottles of wine, priced EUR 20 per bottle, has been mixed with 100 bottles of wine, priced EUR 5. The sale of 1000 bottles of this mixture, at the price of EUR 20 per bottle, will generate the gain of EUR 1 500! The detection of fraud by means of a robot will be much easier, quicker and cheaper than by means of the today's procedures referring to expensive and time-consuming laboratory analyses and experts' services.

From chemical point of view, wine is a water solution of *ca.* 1000 organic and non-organic substances. The most important of them are the following: glycerol (*ca.* 7000 mg/l), fixed acids (*ca.* 6000 mg/l), phenols (*ca.* 1800 mg/l), minerals (*ca.* 1200 mg/l), sugars (*ca.* 750 mg/l), amino acids (*ca.* 550 mg/l), higher alcohols (*ca.* 500 mg/l), volatile acids (*ca.* 400 mg/l), sorbitol & mannitol (*ca.* 300 mg/l), sulfites (*ca.* 80 mg/l), acetaldehydes (*ca.* 70 mg/l) and esters (*ca.* 60 mg/l). Even this incomplete list of wine compounds enables one to imagine in how many parameters (variables) wines may differ, depending on their geographical origin, sort of grapes and technology used for their fabrication. The task to be performed by a spectrophotometric wine analyser (estimation of those parameters or recognition of a pattern they belong to) is extremely difficult because the spectra of very different wines may differ insignificantly. It may be successfully carried out only if very high requirements, concerning both metrological features of the ST and the in-built capacity to process the raw data, are satisfied.

The progress in this domain is, thus, conditioned by parallel advancements of micro- and optoelectronics, on the one hand, and methods for measurement data processing, on the other. The corresponding research-and-development works are carried out by numerous academic and industrial centres employing biochemists and ecologists, as well as computer and measurements experts. As a rule, the industrial works are oriented on prototyping of complete wine analysers. Technical specifications of numerous examples of commercially available spectrophotometric wine analysers may be found in the internet [229]. The research works carried out by universities and other non-industrial institutions are, as a rule, oriented on solving selected problems related to partial analysis of wine rather than on the design of a complete analyser [2-26, 142, 230, 231-245]. A typical objective of such research works is to attain – by means of spectrophotometry – similar metrological outcomes as those attainable by means of expensive systems of liquid chromatography combined with mass spectrometry.

The research works of this type are carried out by numerous institutions in the countries, traditionally involved in wine making and distribution (France, Greece, Italy, Spain, Portugal), as well as in Australia which has declared wine production as their national industry. China – after spectacular successes in some other domains of economy – is currently joining the league of mass producers of wine and investing in related research activities. Although the objectives of research works, carried out in various regions of the world are

similar, their motivation may differ. In the countries, traditionally associated with wine culture, the improvement of laboratory-dedicated techniques of analysis seems to be the main driving force. In the countries where industrial wine production is aimed rather than wine making (Australia, Brazil, China), the motivation of research works is related to the demand for automated monitoring of wine production at all its stages. The common motive is the need to counteract the adulteration of wine. The development of the analysers (robots) for this purpose may be based on two approaches:

−   the translation of spectral data into the language of wine contents, followed by the translation of wine contents into the language of wine labels;
−   the direct identification of wine on the basis of spectral data and appropriately compiled libraries.

The latter approach requires concerted efforts of interdisciplinary teams of scientists (biochemists, mathematicians, metrologists) and practitioners (engineers and sommeliers) to "teach" the analysers (robots) the skills which are up to now reserved for human beings. Despite the fact, that many wine experts, especially in Europe, look with scepticism at all the efforts aimed at automatic wine testing, the intensive research works are going on and delivering many spectacular results.

8. Conclusions

The research problems related to the spectrophotometric analysis and testing of food ingredients and products − such as olive oil, cheese, grains, bread, meat, milk or chocolate − are approached in a similar way as those related to wine analysers and robots. The existing instruments are rather bulky (both in terms of external dimensions and mass), but their miniaturisation seems to be imminent due to the rapid advancements of micro-technologies applicable in spectrophotometry. The number of commercially-available miniature and relatively cheap spectrophotometric transducers has been growing quickly. Their dimensions are in the range of millimetres and − consequently − their metrological performance is on the whole far below that of laboratory instruments, but this deficiency may be at least partially compensated by sophisticated algorithms of measurement data processing.

It seems very probable that the on-going progress in the domain of microspectrophotometry and measurement data processing will lead, in the perspective of a decade, to the widespread use of hand-held spectrophotometric analysers in food manufacturing and distribution, next − in restaurants, bars and private kitchens.

The overview of the methods for spectrophotometric analysis of food, presented in this paper, is demonstrating richness and diversity of approaches and techniques used for solving relevant problems of food analysis. Although they have been studied mathematically for a long time, only recently they can find mass applications. Hence the motivation for further development of the methods and techniques for measurement data processing, dedicated to food analysis, including calibration of spectrophotometric transducers.

The main trends involve generalisation of the existing methods towards problems described by more complex models of the object of measurement, spectral data and their relationship. Apart from the increased dimensionality, which is of crucial importance, the nonlinearity and nonstationarity of the models are to be considered. When looking to the future, one should also mention the methods of measurement data processing, which are already in use, but whose potential of applicability seems to be far from being exhausted, *viz.*: methods of time-frequency analysis, including wavelet transforms [32, 42, 77, 102, 107, 158, 164, 198, 246], and Bayesian methods [180, 247-249] as means for preprocessing of spectral data and for calibration of spectrophotometric transducers.

Acknowledgements

References

[1] Heldman, D.R. (2006). President's message: IFT and the Food Science Profession. *Food Technol. Biotech.,* 60(10), 11.

[2] Application of spectroscopy for non-destructive 'in-bottle' measurement of wine composition and quality. *Report of The Australian Wine Research Institute*, June 2009, *www.awri.com.au* [2012.09.22].

[3] Casale, M., Oliveri, P., Armanino, C., Lanteri, S., Forina, M. (2010). NIR and UV-Vis spectroscopy, artificial nose and tongue: comparison of four fingerprinting techniques for the characterisation of Italian red wines. *Anal. Chim. Acta*, 668(2), 143–148.

[4] Cosme, F., Ricardo-Da-Silva, J.M., Laureano, O. (2009). Tannin profiles of *Vitis vinifera* L. cv. red grapes growing in Lisbon and from their monovarietal wines. *Food Chem.,* 112, 197-204.

[5] Cozzolino, D., Kwiatkowski, M.J., Dambergs, R.G., Cynkat, W.U., Janik, L.J., Skouroumounis, G., Gishen, M. (2008). Analysis of elements in wine using near infrared spectroscopy and partial least squares regression. *Talanta*, 74, 711-716.

[6] Cozzolino, D., Cynkar, W.U., Dambergs, R.G., Mercurio, M.D., Smith, P.A. (2006). Measurement of Condensed Tannins and Dry Matter in Red Grape Homogenates Using Near Infrared Spectroscopy and Partial Least Squares. *J. Agr. Food Chem.*, 56(20), 7631-7636.

[7] Cozzolino, D., Cowey, G., Lattey, K.A., Godden, P., Cynkar, W.U., Dambergs, R.G., Janik, L., Gishen, M. (2008). Relationship between wine scores and visible-near-infrared spectra of Australian red wines. *Anal. Bioanal. Chem.,* 391, 975-981.

[8] Cozzolino, D., Cynkar, W. U., Shah, N., Dambergs, R. G., Smith, P.A. (2009). A brief introduction to multivariate methods in grape and wine analysis. *Int. J. Wine Res.*, 1, 123-130.

[9] Cozzolino, D., Cynkar, W.U., Shah, N., Smith, P.A. (2011). Can spectroscopy geographically classify *Sauvignon Blanc* wines from Australia and New Zealand. *Food. Chem.*, 126(2), 673-678.

[10] Cozzolino, D., Cynkar, W., Shah, N., Smith, P. (2011). Quantitative analysis of minerals and electric conductivity of red grape homogenates by near infrared reflectance spectroscopy. *Comput. Electron. Agr.*, 77(1), 81-85.

[11] Cozzolino, D., Cynkar, W., Shah, N., Smith, P. (2011). Technical solutions for analysis of grape juice, must, and wine: the role of infrared spectroscopy and chemometrics. *Anal. Bioanal Chem.*, 401(5), 1475-1484.

[12] Cynkar, W., Dambergs, R., Smith, P., Cozzolino, D. (2010). Classification of *Tempranillo* wines according to geographic origin: Combination of mass spectrometry based electronic nose and chemometrics. *Anal. Chim. Acta*, 660, 227-231.

[13] Fernández-Novales, J., López, M.-I., Sánchez, M.-T., Morales, J., González-Caballero, V. (2009). Shortwave-near infrared spectroscopy for determination of reducing sugar content during grape ripening, winemaking, and aging of white and red wines. *Food Res. Int.*, 42, 285-291.

[14] Fernández-Novales, J., Sánchez, M.-T., López, M.-I., García-Mesa, J.-A., Ramírez, P. (2011). Feasibility of using a miniature fiber optic UV-Vis-NIR spectrometer to assess total polyphenol index, color intensity and volumic mass in red wine fermentations. *J. Food Process Eng.*, 34, 1028-1045.

[15] Ferreira, M.L., Costa, a. M., Ribeiro, N., Simões, T., Barros, P. (2009). Quality control in FTIR wine analysis – acceptance of analytical results. *Ciência e Técnica Vitivinícola*, 24(1), 47-53.

[16] Ferrer-Gallego, R., Hernández-Hierro, J.M., Rivas-Gonzalo, J.C., Escribano-Bailón, M.T. (2011). Determination of phenolic compounds of grape skins during ripening by NIR spectroscopy. *Food Sci. Technol. – LEB*, 44(4), 847-853.

[17] Geraudie, V., Roger, J.M., Ferrandis, J.L., Gialis, J.M., Barbe, P., Bellon Maurel, V., Pellenc, R. (2009). A revolutionary device for predicting grape maturity based on NIR spectrometry. In *Proc. 8th Fruit Nut and Vegetable Production Engineering Symposium*, Concepcion, Chile.

[18] González-Caballero, V., Sánchez, M.-T., López, M.-I., Pérez-Marín, D. (2010). First steps towards the

development of a non-destructive technique for the quality control of wine grapes during on-vine ripening and on arrival at the winery. *J. Food Eng.*, 101(2), 158-165.

[19] Li, H., Wang, X., Li, Y., Li, P., Wang, H. (2009). Polyphenolic compounds and antioxidant properties of selected China wines. *Food Chem.*, 112, 454-460.

[20] Kruzlicova, D., Mocak, J., Balla, B., Petka, J., Farkova, M., Havel, J. (2009). Classification of Slovak white wines using artificial neural networks and discriminant techniques. *Food Chem.*, 112, 1046-1052.

[21] Larrain, M., Guesalaga, A.R., Agosin, E. (2008). A Multipurpose Portable Instrument for Determining Ripeness in Wine Grapes Using NIR Spectroscopy. *IEEE T. Instrum. Meas.,* 57(2), 294-302.

[22] Lorenzo, C., Garde-Cerdán, T., Pedroza, M.A., Alonso, G.L., Salinas, M.R. (2009). Determination of fermentative volatile compounds in aged red wines by near-infrared spectroscopy. *Food Res. Int.*, 42(9), 1281-1286.

[23] Versari, A., Boulton, R.B., Parpinello, G.P. (2008). A comparison of analytical methods for measuring the color components of red wines. *Food Chem.*, 106, 397-402.

[24] Niu, X., Shen, F., Yu, Y., Yan, Z., Xu, K., Yu, H., Ying, Y. (2008). Analysis of Sugars in Chinese Rice Wine by Fourier Transform Near-Infrared Spectroscopy with Partial Least-Squares Regression. *J. Agr. Food Chem.*, 56(16), 7271-7278.

[25] Yanez, L., Saavedra, J., Martınez, C., Cordova, A., Ganga, M.A. (2012). Chemometric Analysis for the Detection of Biogenic Amines in Chilean *Cabernet Sauvignon* Wines: A Comparative Study between Organic and Nonorganic Production. *J. Food Sci.*, 77(8), T143-T150.

[26] Yu, H.Y., Niu, X.Y., Lin, H.J., Ying, Y.B., Li, B.B., Pan, X.X. (2009). A feasibility study on on-line determination of rice wine composition by Vis-NIR spectroscopy and least-squares support vector machines. *Food Chem.*, 113, 291-296.

[27] de O. Alves, J., Neto, W. B., Mitsutake, H., Alves, P.S.P., Augusti, R. (2010). Extra virgin (EV) and ordinary (ON) olive oils: distinction and detection of adulteration (EV with ON) as determined by direct infusion electrospray ionization mass spectrometry and chemometric approaches. *Rapid Commun. Mass Sp.*, 24(13), 1875-1880.

[28] Barros, A.S., Nunes, A., Martins, J., Delgadillo, I. (2009). Determination of oil and water in olive and olive pomace by NIR and multivariate analysis. *Sensing Instrum. Food Qual. Safety*, 3(3), 180-186.

[29] Bellincontro, A., Taticchi, A., Servili, M., Esposto, S., Farinelli, D., Mencarelli, F. (2012). Feasible Application of a Portable NIR-AOTF Tool for On-Field Prediction of Phenolic Compounds during the Ripening of Olives for Oil Production. *J. Agr. Food Chem.*, 60, 2665-2673.

[30] Cajka, T., Riddellova, K., Klimankova, E., Cerna, M., Pudil, F., Hajslova, J. (2010). Traceability of olive oil based on volatiles pattern and multivariate analysis. *Food Chem.*, 121(1), 282-289.

[31] Calvano, C.D., Aresta, A., Zambonin, C.G. (2010). Detection of hazelnut oil in extra-virgin olive oil by analysis of polar components by micro-solid phase extraction based on hydrophilic liquid chromatography and MALDI-ToF mass spectrometry. *J. Mass Spectrom.*, DOI: 10.1002/jms.1753.

[32] Casale, M., Sinelli, N., Oliveri, P., Di Egidio, V., Lanteri, S. (2010). Chemometrical strategies for feature selection and data compression applied to NIR and MIR spectra of extra virgin olive oils for cultivar identification. *Talanta,* 80(5), 1832-1837.

[33] Casale, M., Zunin, P., Cosulich, M.E., Pistarino, E., Perego, P., Lanteri, S. (2010). Characterisation of table olive cultivar by NIR spectroscopy. *Food Chem.*, 122(4), 1261-1265.

[34] Casale, M., Casolino, C., Oliveri, P., Forina, M. (2010). The potential of coupling information using three analytical techniques for identifying the geographical origin of Liguria extra virgin olive oil. *Food Chem.*, 118(1), 163–170.

[35] Casale, M., Oliveri, P., Casolinoa, C., Sinelli, N., Zunina, P., Armaninoa, C., Forina, M., Lanteria, S. (2012). Characterisation of PDO olive oil *Chianti Classico* by non-selective (UV–visible, NIR and MIR spectroscopy) and selective (fatty acid composition) analytical techniques. *Anal. Chim. Acta*, 712, 56-63.

[36] Concha-Herrera, V., Lerma-García, M.a.J.s., Herrero-Martínez, J.M., Simó-Alfonso, E.F. (2009). Prediction of the Genetic Variety of Extra Virgin Olive Oils Produced at La Comunitat Valenciana, Spain, by Fourier Transform Infrared Spectroscopy. *J. Agr. Food Chem.*, 57(21), 9985-9989.

[37] Dupuy, N., Galtier, O., Le Dréau, Y., Pinatel, C., Kister, J., Artaud, J. (2010). Chemometric analysis of combined NIR and MIR spectra to characterize French olives. *Eur. J. Lipid Sci. Tech.*, 112(4), 463-475.

[38] Dupuy, N., Galtier, O., Ollivier, D., Vanloot, P., Artaud, J. (2010). Comparison between NIR, MIR,

www.czasopisma.pan.pl   PAN   www.journals.pan.pl
POLSKA AKADEMIA NAUK

Metrol. Meas. Syst., Vol. XIX (2012), No. 4, pp. 623-652.

concatenated NIR and MIR analysis and hierarchical PLS model. Application to virgin olive oil analysis. *Anal. Chim. Acta*, 666(1-2), 23-31.

[39] Frankel, E.N. (2010). Chemistry of Extra Virgin Olive Oil: Adulteration, Oxidative Stability, and Antioxidants. *J. Agr. Food Chem.*, 58(10), 5991-6006.

[40] Galtier, O., Abbas, O., Le Dréau, Y., Rebufa, C., Kister, J., Artaud, J., Dupuy, N. (2011). Comparison of PLS1-DA, PLS2-DA and SIMCA for classification by origin of crude petroleum oils by MIR and virgin olive oils by NIR for different spectral regions. *Vib. Spectrosc.*, 55, 132-140.

[41] Gómez-Robledo, L., Melgosa, M., Huertas, R., Roa, R., Moyano, M.J., Heredia, F.J. (2008). Virgin-Olive-Oil Color in Relation to Sample Thickness and the Measurement Method. *J. Am. Oil Chem. Soc.*, 85(11), 1063-1071.

[42] Gurdeniz, G. Ozen, B. (2009). Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data. *Food Chem.*, 116, 519-525.

[43] Gurdeniz, G., Ozen, B., Tokatli, F. (2010). Comparison of fatty acid profiles and mid-infrared spectral data for classification of olive oils. *Eur. J. Lipid Sci. Tech.*, 112(2), 218-226.

[44] Harrington, P.d.B., Kister, J., Artaud, J., Dupuy, N. (2009). Automated Principal Component-Based Orthogonal Signal Correction Applied to Fused NI-MIR Spectra of French Olive Oils. *Anal. Chem.*, 81(17), 7160-7169.

[45] Hennessy, S., Downey, G., O'Donnell, C.P. (2009). Confirmation of Food Origin Claims by Fourier Transform Infrared Spectroscopy and Chemometrics: Extra Virgin Olive Oil from Liguria. *J. Agr. Food Chem.*, 57, 1735-1741.

[46] Kavdir, I., Buyukcan, M.B., Lu, R., Kocabiyik, H., Seker, M. (2009). Prediction of olive quality using FT-NIR spectroscopy in reflectance and transmittance modes. *Biosystems Eng.*, 103(3), 304-312.

[47] Kružlicová, D., Mrázová, V., Šnuderl, K., Mocák, J., Lankmayr, E. (2008). Chemometric classification of edible oils. *Acta Chim. Slovaca*, 1(1), 165-174.

[48] Lerma-García, M.J., Ramis-Ramos, G., Herrero-Martínez, J.M., Simó-Alfonso, E.F. (2010). Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy. *Food Chem.*, 118(1), 78-83.

[49] Lerma-García, M.J., Simó-Alfonso, E.F., Bendini, A., Cerretani, L. (2011). Rapid evaluation of oxidised fatty acid concentration in virgin olive oil using Fourier-transform infrared spectroscopy and multiple linear regression. *Food Chem.*, 124(2), 679-684.

[50] Maggio, R.M., Valli, E., Bendini, A., Gómez-Caravaca, A.M., Toschi, T.G., Cerretani, L. (2011). A spectroscopic and chemometric study of virgin olive oils subjected to thermal stress. *Food Chem.*, 127, 216-221.

[51] Morales-Sillero, A., Fernández-Cabanás, V.-M., Casanova, L., Jiménez, M.-R., Suárez, M.-P., Rallo, P. (2011). Feasibility of NIR spectroscopy for non-destructive characterization of table olive traits. *J. Food Eng.*, 107(1), 99-106.

[52] Nunes, A., Martins, J., Barros, A.S., Galvis-Sánchez, A.C., Delgadillo, I. (2009). Estimation of olive oil acidity using FT-IR and partial least squares regression. *Sensing Instrum. Food Qual. Safety*, 3(3), 187-191.

[53] Obeidat, S.M., Khanfar, M.S., Obeidat, W.M. (2009). Classification of Edible Oils and Uncovering Adulteration of Virgin Olive Oil Using FTIR with the Aid of Chemometrics. *Aust. J. Basic Appl. Sci.*, 3(3), 2048-2053.

[54] Rohman, A., Che Man, Y.B., Ismail, A., Hashim, P. (2010). Application of FTIR Spectroscopy for the Determination of Virgin Coconut Oil in Binary Mixtures with Olive Oil and Palm Oil. *J. Am. Oil Chem. Soc.*, 87(6), 601-606.

[55] Rohman, A. Man, Y.B.C. (2010). Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. *Food Res. Int.*, 43(3), 886-892.

[56] Sinelli, N., Cerretani, L., Egidio, V.D., Bendini, A., Casiraghi, E. (2010). Application of near (NIR) infrared and mid- (MIR) infrared spectroscopy as a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity. *Food Res. Int.,* 43(1), 369-375.

[57] Sinelli, N., Casale, M., Di Egidio, V., Oliveri, P., Bassi, D., Tura, D., Casiraghi, E. (2010). Varietal discrimination of extra virgin olive oils by near- and mid-infrared spectroscopy. *Food Res. Int.*, 43, 2126-2131.

[58] Torrecilla, J.S., Rojo, E., Domínguez, J.C., Rodríguez, F. (2010). A Novel Method To Quantify the

Adulteration of Extra Virgin Olive Oil with Low-Grade Olive Oils by UV−Vis. *J. Agr. Food Chem.*, 58(3), 1679-1684.

[59] Bobelyn, E., Serban, A.-S., Nicu, M., Lammertyn, J., Nicolai, B.M., Saeys, W. (2010). Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of biological variability on spectra and model performance. *Postharvest Biol. Tech.*, 55(3), 133-143.

[60] Bureau, S., Ruiz, D., Reich, M., Gouble, B., Bertrand, D., Audergon, J.-M., Renard, C.M.G.C. (2009). Rapid and non-destructive analysis of apricot fruit quality using FT-near-infrared spectroscopy. *Food Chem.*, 113, 1323-1328.

[61] Cayuela, J.A. Weiland, C. (2010). Intact orange quality prediction with two portable NIR spectrometers. *Postharvest Biol. Tec.*, 58(2), 113-120.

[62] Clement, A., Dorais, M., Vernon, M. (2008). Nondestructive Measurement of Fresh Tomato Lycopene Content and Other Physicochemical Characteristics Using Visible-NIR Spectroscopy. *J. Agr. Food Chem.*, 56, 9813-9818.

[63] Cozzolino, D., Cynkar, W.U., Shah, N., Smith, P. (2011). Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. *Food Res. Int.*, 44(7), 1888-1896.

[64] Hongjian, L., Yibin, Y. (2009). Theory and application of near infrared spectroscopy in assessment of fruit quality: a review. *Sensing Instrum. Food Qual. Safety,* 3, 130-141.

[65] Judd, M.J., Meyer, D.H., Meekings, J.S., Richardson, A.C., Walton, E.F. (2010). An FTIR study of the induction and release of kiwifruit buds from dormancy. *J. Sci. Food Agr.*, 90, 1071-1080.

[66] Lijuan, X., Xingqian, Y., Donghong, L., Yibin, Y. (2009). Quantification of glucose, fructose and sucrose in bayberry juice by NIR and PLS. *Food Chem.*, 114, 1135-1140.

[67] Lijuan, X., Yibin, Y., Tiejin, Y. (2009). Rapid determination of ethylene content in tomatoes using visible and short-wave near-infrared spectroscopy and wavelength selection. *Chemometr. Intell. Lab.*, 97, 141-145.

[68] Lijuan, X., Xingqian, Y., Donghong, L., Yibin, Y. (2011). Prediction of titratable acidity, malic acid, and citric acid in bayberry fruit by near-infrared spectroscopy. *Food Res. Int.*, 44(7), 2198-2204.

[69] Magwaza, L.S., Opara, U.L., Nieuwoudt, H., Cronje, P.J.R., Saeys, W., Nicolaï, B. (2012). NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit – A Review. *Food Bioprocess Technol.*, 5, 425–444.

[70] Makino, Y., Ichimura, M., Oshita, S., Kawagoe, Y., Yamanaka, H. (2010). Estimation of oxygen uptake rate of tomato (Lycopersicon esculentum Mill.) fruits by artificial neural networks modelled using near-infrared spectral absorbance and fruit mass. *Food Chem.*, 121(2), 533-539.

[71] Moghimi, A., Aghkhani, M.H., Sazgarnia, A., Sarmad, M. (2010). Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. *Biosystems Eng.*, 106(3), 295-302.

[72] Pérez-Marín, D., Paz, P., Guerrero, J.-E., Garrido-Varo, A., Sánchez, M.-T. (2010). Miniature handheld NIR sensor for the on-site non-destructive assessment of post-harvest quality and refrigerated storage behavior in plums. *J. Food Eng.*, 99, 294-302.

[73] Tewari, J.C., Dixit, V., Cho, B.-K., Malik, K.A. (2008). Determination of origin and sugars of citrus fruits using genetic algorithm, correspondence analysis and partial least square combined with fiber optic NIR spectroscopy. *Spectrochim. Acta A*, 71, 1119-1127.

[74] Tong, S., Kang, H., Huirong, X., Yibin, Y. (2010). Research advances in nondestructive determination of internal quality in watermelon/melon: A review. *J. Food Eng.*, 100(4), 569-577.

[75] Neo, Y.P., Ariffin, A., Tan, C.P., Tan, Y.A. (2008). Determination of oil palm fruit phenolic compounds and their antioxidant activities using spectrophotometric methods. *International J. Food Sci. & Technology*, 43(10), 1832-1837.

[76] Fei, L., Li, W., Yong, H. (2008). Determination of acetic acid of fruit vinegars using near infrared spectroscopy and least squares-support vector machine. In *Proc. of 7th International Conference on Machine Learning and Cybernetics*, Kunming, China, 1232-1237.

[77] Fei, L., Yong, H., Li, W., Guangming. S. (2009). Detection of Organic Acids and pH of Fruit Vinegars Using Near-Infrared Spectroscopy and Multivariate Calibration. *Food Bioprocess Technol.*, 4(8), 1331-1340.

[78] Guerrero, E.D., Mejías, R.C., Marín, R.N., Lovillo, M.P., Barroso, C.G. (2010). A new FT-IR method

www.czasopisma.pan.pl   PAN   www.journals.pan.pl
POLSKA AKADEMIA NAUK

Metrol. Meas. Syst., Vol. XIX (2012), No. 4, pp. 623-652.

combined with multivariate analysis for the classification of vinegars from different raw materials and production processes. *J. Sci. Food Agr.*, 90, 712-718.

[79] Zunyi, W., Fei, Li., Yong, H. (2009). Comparison and Determination of Acetic Acid of Plum Vinegar Using Vis/Near Infrared Spectroscopy and Multivariate Calibration. In *Proc. World Congress on Computer Science and Information Engineering*, Los Angeles, USA, 201-204.

[80] Posudin, Y.I. (2007). *Practical spectroscopy in agriculture and food science*. Enfield, NH, USA: Science Publishers.

[81] Andersen, C.M., Frøst, M.B., Viereck, N. (2010). Spectroscopic characterization of low- and non-fat cream cheeses. *Int. Dairy J.,* 20(1), 32-39.

[82] González-Martín, I., Hernández-Hierro, J.M., Vivar-Quintana, A., Revilla, I., González-Pérez, C. (2009). The application of near infrared spectroscopy technology and a remote reflectance fibre-optic probe for the determination of peptides in cheeses (cow's, ewe's and goat's) with different ripening times. *Food Chem.*, 114, 1564-1569.

[83] González-Martín, I., Hernández-Hierro, J.M., Salvador-Esteban, J., González-Pérez, C., Revillab, I., Vivar-Quintanab, A. (2011). Discrimination of seasonality in cheeses by near-infrared technology. *J. Sci. Food Agr.*, 91, 1064-1069.

[84] González-Martín, M.I., Severiano-Pérez, P., Revilla, I., Vivar-Quintana, A.M., Hernández-Hierro, J.M., González-Pérez, C., Lobos-Ortega, I.A. (2011). Prediction of sensory attributes of cheese by near-infrared spectroscopy. *Food Chem.*, 127(1), 256-263.

[85] Subramanian, A., Harper, W.J., Rodriguez-Saona, L.E. (2009). Rapid Prediction of Composition and Flavor Quality of Cheddar Cheese Using ATR–FTIR Spectroscopy. *J. Food Sci.*, 74(3), 292-297.

[86] Balabin, R.M. Smirnov, S.V. (2011). Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. *Talanta*, 85(1), 562-568.

[87] Carleos Artime, C.E., de la Fuente, J.A.B., Garcia, M.A.P., Vega, R.M., Blanco, N.C. (2008). On-line Estimation of Fresh Milk Composition by means of VIS-NIR Spectrometry and Partial Least Squares Method (PLS). In *Proc. IEEE Instrumentation and Measurement Technology Conference*, Victoria, Vancouver Island, Canada.

[88] Coppa, M., Ferlay, A., Leroux, C., Jestin, M., Chilliard, Y., Martin, B., Andueza, D. (2010). Prediction of milk fatty acid composition by near infrared reflectance spectroscopy. *Int. Dairy J*., 20(3), 182-189.

[89] Muniz, R., Pérez, M.A., de la Torre, C., Carleos, C.E., Corral, N., Baro, J.A. (2009). Comparison of Principal Component Regression (PCR) and Partial Least Square (PLS) methods in prediction of raw milk composition by Vis-NIR spectrometry. application to development of on-line sensors for fat, protein and lactose contents. In *Proc. XIX IMEKO World Congress 'Fundamental and Applied Metrology'*, Lisbon, Portugal, 2564-2568.

[90] Woodcock, T., Fagan, C.C., O'Donnell, C.P., Downey, G. (2008). Application of Near and Mid-Infrared Spectroscopy to Determine Cheese Quality and Authenticity. *Food Bioprocess Technol.*, 1, 117-129.

[91] Hao, L., Jiewen, Z., Li, S., Quansheng, C., Fang, Z. (2011). Freshness measurement of eggs using near infrared (NIR) spectroscopy and multivariate data analysis. *Innov Food Sci. Emerging Technol.*, 12(2), 182-186.

[92] Jiewen, Z., Hao, L., Quansheng, C., Xingyi, H., Zongbao, S., Fang, Z. (2010). Identification of egg's freshness using NIR and support vector data description. *J. Food Eng*., 98(4), 408-414.

[93] González-Martín, M.I., Bermejo, C.F., Hierro, J.M.H., González, C.I.S. (2009). Determination of hydroxyproline in cured pork sausages and dry cured beef products by NIRS technology employing a fibre–optic probe. *Food Control*, 20(8), 752-755.

[94] Guy, F., Prache, S., Thomas, A., Bauchart, D., Andueza, D. (2011). Prediction of lamb meat fatty acid composition using near-infrared reflectance spectroscopy (NIRS). *Food Chem.*, 127(3), 1280-1286.

[95] Haibo, H., Haiyan, Y., Huirong, X., Yibin, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *J. Food Eng.*, 87, 303-313.

[96] Kapper, C., Klont, R.E., Verdonk, J.M.A.J., Urlings, H.A.P. (2012). Prediction of pork quality with near infrared spectroscopy (NIRS) − 1. Feasibility and robustness of NIRS measurements at laboratory scale. *Meat Sci.*, 91, 294-299.

[97] Kapper, C., Klont, R.E., Verdonk, J.M.A.J., Urlings, H.A.P. (2012). Prediction of pork quality with near infrared spectroscopy (NIRS) − 2. Feasibility and robustness of NIRS measurements under production

plant conditions. *Meat Sci.*, 91, 300-305.

[98] del Moral, F.G., Guillén, A., del Moral, L.G., O'Valle, F., Martinez, L., del Moral, R.G. (2009). Duroc and Iberian pork neural network classification by visible and near infrared reflectance spectroscopy. *J. Food Eng.*, 90, 540-547.

[99] Pérez-Marín, D., Fearn, T., Guerrero, J.E., Garrido-Varo, A. (2010). Robustness in pig fat NIRS calibrations by orthogonal projection. *Chemometr. Intell. Lab*., 100(1), 36-40.

[100] Quansheng, C., Jianrong, C., Xinmin, W., Jiewen, Z. (2011). Application of linear/non-linear classification algorithms in discrimination of pork storage time using Fourier transform near infrared (FT-NIR) spectroscopy. *Food Sci. Technol. – LEB*, 44(10), 2053-2058.

[101] Ritthiruangdej, P., Ritthiron, R., Shinzawa, H., Ozaki, Y. (2011). Non-destructive and rapid analysis of chemical compositions in Thai steamed pork sausages by near-infrared spectroscopy. *Food Chem.*, 129(2), 684-692.

[102] Liao, Y.T., Fan, Y.X., Cheng, F. (2010). On-line prediction of fresh pork quality using visible/near-infrared reflectance spectroscopy. *Meat Sci.*, 86(4), 901-907.

[103] Alexandrakis, D., Downey, G., Scannell, A.G.M. (2011). Detection and identification of selected bacteria, inoculated on chicken breast, using near infrared spectroscopy and chemometrics. *Sensing Instrum. Food Qual. Safety*, 5(2), 57-62.

[104] Etzold, E. Lichtenberg-Kraag, B. (2007). Determination of the botanical origin of honey by Fourier-transformed infrared spectroscopy: an approach for routine analysis. *Eur. Food Res. Technol.,* 227(2), 579-586.

[105] Gallardo-Velázquez, T., Osorio-Revilla, G., Zuňiga-de Loa, M., Rivera-Espinoza, Y. (2009). Application of FTIR-HATR spectroscopy and multivariate analysis to the quantification of adulterants in Mexican honeys. *Food Res. Int*., 42, 313-318.

[106] Chen, L., Xue, X., Ye, Z., Zhou, J., Chen, F., Zhao J. (2011). Determination of Chinese honey adulterated with high fructose corn syrup by near infrared spectroscopy. *Food Chem.*, 128(4), 1110-1114.

[107] Xiangrong, Z., Shuifang, L., Yang, S., Zhuoyong, Z., Gaoyang, L., Donglin, S., Feng, L. (2010). Detection of adulterants such as sweeteners materials in honey using near-infrared spectroscopy and chemometrics. *J. Food Eng*., 101(1), 92-97.

[108] da Costa-Filho, P.A. (2009). Rapid determination of sucrose in chocolate mass using near-infrared spectroscopy. *Anal. Chim Acta*, 631, 206-211.

[109] Chena, Q., JZhao, J., Liu, M., Cai, J., Liu, J. (2008). Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms. *J. Pharamaceut. Biomed*., 46, 568-573.

[110] Quansheng Chen, Jiewen Zhao, Sumpunchaitep, Zhiming Guo (2009). Simultaneous analysis of main catechins contents in green tea (*Camellia sinensis* (L.)) by Fourier transform near-infrared reflectance (FT-NIR) spectroscopy. *Food Chem.*, 113, 1272-1277.

[111] Morawski, R.Z. (2008). On food, spectrophotometry, and measurement data processing. *In Proc. of 12th IMEKO TC1 & TC7 Joint Symposium 'Man Science & Measurement'*, Annecy, France, 7-20.

[112] Examples of mini- and micro-spectrophotometers*: http://www.oceanoptics.com*; *http://www.controldevelopment.com*; *http://www.arcoptix.com/Spectrometers.htm; http://www.horiba. com*; *http://www.microspectralanalysis.com*; *http:// www.stellarnet.us* [2012.09.22].

[113] Morawski, R.Z. (1994). Unified Approach to Measurand Reconstruction. *IEEE T. Instrum. Meas*., 43(2), 226-231.

[114] Mroczka, J. Szczuczyński, D. (2009). Inverse problems formulated in terms of first-kind Fredholm integral equations in indirect measurements. *Metrol. Meas. Syst.*, 26(3), 333-357.

[115] Yongni, S., Yong, H. (2009). Measurement of Soluble Solids Content and pH of Yogurt Using Visible/Near Infrared Spectroscopy and Chemometrics. *Food Bioprocess. Technol*., 2(2), 229-233.

[116] Morawski, R.Z. (2006). Spectrophotometric applications of digital signal processing. *Meas. Sci. Technol*., 17, R.117-R.144.

[117] Wisniewski, M.P., Morawski, R.Z., Barwicz, A. (1999). Modeling the Spectrometric Microtransducer. *IEEE T. Instrum. Meas*., 48(3), 747-752.

[118] van Cittert, P.H. (1931). Zum Einfluss der Spaltbreit auf die Intensitätsverteilung in Spectrallinien II. *Z. Physik*, 69, 298-311.

[119] Kalivas, J. (1999). Interrelationships of multivariate regression methods using eigenvector basis sets. *J.*

Chemometr., (13), 111-132.

[120] Geladi, P., Dabakk, E. (1999). *Computational methods of analysis and chemometrics in near-infrared spectroscopy. Encyclopedia of Spectroscopy and Spectrometry*, Lindon, J*., et al.*, Eds., London: Academic Press.

[121] Morawski, R.Z., Miękina, A., Wagner, J. (2012). A method of weighing matrix for spectrophotometric analysis of oil mixtures. In *Advanced Mathematical and Computationol Tools in Metrology and Testing IX*, Pavese, F*., et al.*, Eds., Singapore-Hackensack-London: World Scientific Publishing Company.

[122] Cardaliaguet, P., Euvrard, G. (1992). Approximation of a Function and its Derivative with a Neural Network. *Neural Networks*, 5, 207-220.

[123] Geva, S., Sitte, J. (1992). A Constructive Method for Multivariate Function Approximation by Multilayer Perceptrons. *IEEE T. Neural Networ.*, 3(4), 621-624.

[124] Höskuldsson, A. (Nov. 2003). Selection of subsets of variables in linear regression. *Homepage of Chemometrics, http://www.acc.umu.se/~tnkjtg/Chemometrics/Editorial* [2012.09.20].

[125] Swierenga, H., Wülfert, F., deNoord, O.E., de Weijer, A.P., Smilde, A.K., Buydens, L.M.C. (2000). Development of robust calibration models in near infra-red spectrometric applications. *Anal Chim. Acta*, 411, 121-135.

[126] Wold, S. (Jan. 2004). PAC, PAT and Variable Selection. *Homepage of Chemometrics, http://www.acc.umu.se/~tnkjtg/Chemometrics/Editorial* [2012.09.20].

[127] Xueguang Shao, Fang Wang, Da Chen, Qingde Su (2004). A method for near-infrared spectral calibration of complex plant samples with wavelet transform and elimination of uninformative variables. *Anal. Bioanal. Chem.*, 378(5), 1382-1387.

[128] Francois, N., Govaerts, B., Boulanger, B. (2004). Optimal designs for inverse prediction in univariate nonlinear calibration models. *Chemometr. Intell. Lab.*, 74, 283-292.

[129] Brereton, R.G. (2005). *Chemometrics – Data Analysis for the Laboratory and Chemical Plant*. Chichester, J. Wiley & Sons.

[130] Kruse, R., M. Steinbrecher, M. (2010). Visual data analysis with computational intelligence methods. *Bull. Pol. Acad. Sci.-Te.*, 58(4), 393-401.

[131] Alciaturi, C.E., Escobar, M.E., Esteves, I. (2005). The use of the autocorrelation function in modeling of multivariate data. *Anal. Chim. Acta*, 553, 134-140.

[132] Benoudjit, N., Francois, D., Meurens, M., Verleysen, M. (2004). Spectrophotometric variable selection by mutual information. *Chemometr. Intell. Lab.*, 74, 243-251.

[133] Benoudjit, N., Cools, E., Meurens, M., Verleysen, M. (2004). Chemometric Calibration of Infrared Spectrometers: Selection and Validation of Variables by Non-Linear Models. *Chemometr. Intell. Lab.*, 70, 47-53.

[134] Pontes, M.J.C., Galvao, R.K.H., Araujo, M.C.N., Moriera, P.N.T., Neto, O.D.P., Jose, G.E., Saldanha, T.C.B. (2005). The Successive projections algorithm for spectral variable selection in classification problems. *Chemometr. Intell. Lab.*, 78, 11-18.

[135] Naes, T., Mevik, B.-H. (2001). Understanding the collinearity problem in regression and discriminant analysis. *J. Chromatometr.*, 15, 413-426.

[136] De Juan, A., Casassas, E., Tauler, R. (2000). *Soft modeling of analytical data*. In *Encyclopedia of Anal. Chem.*, Myers, R. A., Ed., Chichester, Wiley and Sons.

[137] Jackson, J.E. (1991). *A User's Guide To Principal Components*. New York – Singapore: John Wiley & Sons, Inc.

[138] Lavine, B.K. (2000). Chemometrics: Fundamental review. *Anal. Chem.*, 72, 91R-98R.

[139] Bro, R. (1998). *Multi-way Analysis in the Food Industry – Models, Algorithms, and Applications*. Ph.D. Thesis, Universiteit van Amsterdam, Amsterdam, The Netherlands.

[140] Lomillo, M.A.A., Renedo, O.D., Martinez, M.J.A. (2004). Resolution of Binary Mixtures of Rifamycin SV and Rifampicin by UV/Vis Spectroscopy and Partial Least-Squares Method (PLS). *Chem. Biodivers.*, 1, 1336-1343.

[141] Azzouz, T., Puigdoménech, A., Aragay, M., Tauler, R. (2003). Comparison between different data pre-treatment methods in the analysis of forage samples using near-infrared diffuse reflectance spectroscopy and partial least-squares multivariate calibration method. *Anal. Chim. Acta*, 484(1), 121-134.

[142] Cozzolino, D., Kwiatkowski, M.J., Parker, M., Cynkar, W.U., Dambergs, R.G., Gishen, M., Herdrich, M.J. (2004). Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy. *Anal. Chim. Acta*, 513, 73-80.

[143] Du, Y.P., Liang, Y.Z., Jiang, J.H., Berry, R.J., 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(2), 183-191.

[144] Fernandez-Pierna, J.A., Jin, L., Wahl, F., Faber, N.M., Massart, D.L. (2003). Estimation of partial least squares regression prediction uncertainty when the reference values carry a sizeable measurement error. *Chemometr. Intell. Lab.*, 65, 281-291.

[145] Ghasemi, J., Shahabadi, N., Seraji, H.R. (2004). Spectrophotometric simultaneous determination of cobalt, copper and nickel using nitroso-R-salt in alloys by partial least squares. *Anal. Chim. Acta*, 510(1), 121-126.

[146] Hiroaki, I., Toyonori, N., Eiji, T. (2002). Measurement of Pesticide Residues in Food Based on Diffuse Reflectance IR Spectroscopy. *IEEE T. Instrum. Meas*., 51(5), 886-890.

[147] Huang, J., Brennan, D., Sattler, L., Alderman, J., Lane, B., O'Mathuna, C. (2002). A Comparison of Calibration Methods Based on Calibration Data Size and Robustness. *Chemometr. Intell. Lab*., 62, 25-35.

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

[149] Khajehsharifi, H., Mousavi, M.F., Ghasemi, J., Shamsipur, M. (2004). Kinetic spectrophotometric method for simultaneous determination of selenium and tellurium using partial least squares calibration. *Anal. Chim. Acta*, 512(2), 369-373.

[150] Muik, B., Lendl, B., Molina-Diaz, A., Villarejo, L.P., Ayora-Canada, M.J. (2004). Determination of oil and water content in olive pomace using near infrared and Raman. *Anal. Bioanal. Chem*., 379, 35-41.

[151] Pizarro, C., Esteban-Díez, I., Nistal, A.-J., González-Sáiz, J.-M. (2004). Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Anal. Chim. Acta*, 509(2), 217-227.

[152] Shen, Q., Jiang, J.H., Shen, G.L., Yu, R.Q. (2003). Variable selection by an evolution algorithm using modified Cp based on MLR and PLS modeling: QSAR studies of carcinogenicity of aromatic amines. *Anal. Bioanal. Chem*., 375(2), 248-254.

[153] Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *WIREs Comput. Stat*., 2(1), 97-106.

[154] Ashok, P.C., Praveen, B.B., Dholakia, K. (2011). Near infrared spectroscopic analysis of single malt Scotch whisky on an optofluidic chip. *Opt. Express*, 19(23).

[155] Balabin, R.M., Safieva, R.Z. (2011). Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data. *Anal. Chim. Acta,* 689(2), 190-197.

[156] Brunt, K., Drost, W.C. (2010). Design, Construction, and Testing of an Automated NIR In-line Analysis System for Potatoes. Part I: Off-line NIR Feasibility Study for the Characterization of Potato Composition. *Potato Res*., 53(1), 25-39.

[157] Brunt, K., Smits, B., Holthuis, H. (2010). Design, Construction, and Testing of an Automated NIR In-line Analysis System for Potatoes. Part II. Development and Testing of the Automated Semi-industrial System with In-line NIR for the Characterization of Potatoes. *Potato Res*., 53(1), 41-60.

[158] Cai, C.B., Han, Q.J., Tang, L.J., Nie, J.F., Ouyang, L.Q., Yu, R.Q. (2008). Treating NIR data with orthogonal discrete wavelet transform: Predicting concentrations of a multi-component system through a small-scale calibration set. *Talanta*, 77, 822-826.

[159] Costa Filho, P.A.D. (2009). Rapid determination of sucrose in chocolate mass using near infrared spectroscopy. *Anal. Chim. Acta*, (631), 206-211.

[160] Costa-Pereira, A.F., Coelho Pontes, M.J., Gambarra Neto, F.F., Bezerra Santos, S.R., Harrop Galvao, R.K., Ugulino Araújo, M.C. (2008). NIR spectrometric determination of quality parameters in vegetable oils using *i*PLS and variable selection (review). *Food Res. Int.*, 41, 341-348.

[161] Cozzolino, D., Chree, A., Murray, I., Scaife, J.R. (2009). Usefulness of near infrared spectroscopy to monitor the extent of heat treatment in fish meal. *Int. J. Food Sci. Technol.*, 44(8), 1579-1584.

[162] Cozzolino, D., Restaino, E., Fassio, A. (2010). Discrimination of yerba mate (Ilex paraguayensis St. Hil.) samples according to their geographical origin by means of near infrared spectroscopy and multivariate

www.czasopisma.pan.pl   PAN   www.journals.pan.pl
POLSKA AKADEMIA NAUK

Metrol. Meas. Syst., Vol. XIX (2012), No. 4, pp. 623-652.

analysis. *Sensing Instrum. Food Qual. Safety*, 4(2), 67-72.

[163] Di Egidio, V., Sinelli, N., Giovanelli, G., Moles, A., Casiraghi, E. (2010). NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation. *Eur. Food Res. Technol.*, 230(6), 947-955.

[164] Fei, L., Yong, H., Li, W., Guangming, S. (2009). Detection of Organic Acids and pH of Fruit Vinegars Using Near-Infrared Spectroscopy and Multivariate Calibration. *Food Bioprocess Technol.*, 14(8), 1331-1340.

[165] Horikawa, Y., Imai, T., Takada, R., Watanabe, T., Takabe, K., Kobayashi, Y., Sugiyama, J. (2011). Near-infrared chemometric approach to exhaustive analysis of rice straw pretreated for bioethanol conversion. *Appl. Biochem. Biotech.*, 164(2), 194-203.

[166] Huazhou, C., Tao, P., Jiemei, C., Qipeng, L. (2011). Waveband selection for NIR spectroscopy analysis of soil organic matter based on SG smoothing and MWPLS methods. *Chemometr. Intell. Lab.*, 107, 139-146.

[167] Hyonho, C., Keles, S. (2010). Sparse PLS regression for simultaneous dimension reduction and variable selection. *J. Roy. Stat. Soc. B Met.*, 72(Part 1), 3-25.

[168] Jin, H.L., Choung, M.G. (2011). Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. *Food Chem.*, 126(1), 368-373.

[169] Xie, L.J., Ying, J.B. (2009). Use of near-infrared spectroscopy and least-squares support vector machine to determine quality change of tomato juice. *J. Zhejiang Univ. Sci. B*, 10(6), 465-71.

[170] Moros, J., Llorca, I., Cervera, M.L., Pastor, A., Garrigues, S., de la Guardia, M. (2008). Chemometric determination of arsenic and lead in untreated powdered red paprika by diffuse reflectance near-infrared spectroscopy. *Anal. Chim. Acta*, 613, 196-206.

[171] Pojić, M., Mastilović, J., Palić, D., Pestorić, M. (2010). The development of near-infrared spectroscopy (NIRS) calibration for prediction of ash content in legumes on the basis of two different reference methods. *Food Chem.*, 123(3), 800-805.

[172] Quansheng, C., Pei, J., Jiewen, Z. (2010). Measurement of total flavone content in snow lotus (Saussurea involucrate) using near infrared spectroscopy combined with interval PLS and genetic algorithm. *Spectrochim. Acta. A*, 76(1), 50-55.

[173] Queji, M.D., Wosiacki, G., Cordeiro, G.A., Peralta-Zamora, P.G., Nagata, N. (2010). Determination of simple sugars, malic acid and total phenolic compounds in apple pomace by infrared spectroscopy and PLSR. *Int. J. Food Sci. Technol.*, 45, 602-609.

[174] Sherazi, S.T.H., Kandhro, A., Mahesar, S.A., Bhanger, M.I., Younis Talpur, M., Arain, S. (2009). Application of transmission FT-IR spectroscopy for the *trans* fat determination in the industrially processed edible oils. *Food Chem.*, 114, 323-327.

[175] Shetty, N., Gislum, R. (2011). Quantification of fructan concentration in grasses using NIR spectroscopy and PLSR. *Field Crops Res.*, 120(1), 31-37.

[176] Shiroma, C., Rodriguez-Saona, L. (2009). Application of NIR and MIR spectroscopy in quality control of potato chips. *J. Food Compos. Anal.*, 22(6), 596-605.

[177] Smyth, H.E., Cozzolino, D., Cynkar, W.U., Dambergs, R.G., Sefton, M., Gishen, M. (2008). Near infrared spectroscopy as a rapid tool to measure volatile aroma compounds in Riesling wine: possibilities and limits. *Anal. Bioanal. Chem.*, 390(7), 1911-1916.

[178] Soares, I.P., Rezende, T.F., Silva, R.C., Castro, E.V.R., Fortes, I.C.P. (2008). Multivariate Calibration by Variable Selection for Blends of Raw Soybean Oil/Biodiesel from Different Sources Using Fourier Transform Infrared Spectroscopy (FTIR) Spectra Data. *Energ. Fuels*, 22, 2079-2083.

[179] Ulissi, V., Antonucci, F., Benincasa, P., Farneselli, M., Tosti, G., Guiducci, M., Tei, F., Costa, C., Pallottino, F., Pari, L., Menesatti, P. (2011). Nitrogen Concentration Estimation in Tomato Leaves by VIS-NIR Non-Destructive Spectroscopy. *Sensors*, 11(6), 6411-6424.

[180] Zou Xiaobo, Zhao Jiewen, Povey, M. J., Holmes, M., Mao Hanpin (2010). Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta*, 667(1-2), 14-32.

[181] Martens, H., Naes, T. (1989). *Multivariate Calibration*. Chichester – Singapore, J.Wiley&Sons.

[182] Varmuza, K., Filzmoser, P. (2009). *Introduction to Multivariate Statistical Analysis in Chemometrics*. Boca Raton: CRC Press – Taylor & Francis Group.

[183] Gustafsson, M.G. (2001). A Probabilistic Derivation of the Partial Least-Squares Algorithm. *J. Chem. Inf. Comp. Sci.*, 41, 288-294.

[184] Website of Eigenvector Research, Inc., *http://www.eigenvector.com/software/pls_toolbox.htm*

[2012.09.22].

[185] Morawski, R.Z., Miękina, A. (2012). A comparative study of forty algorithms for spectrophotometric analysis of edible oil mixtures. In *Proc. of XXth IMEKO World Congress 'Metrology for Green Growth'*, Busan, Republic of Korea.

[186] Frank, I., Friedman, J. (1993). A statistical view of some chemometric regression tools. *Technometrics*, 35, 109-148.

[187] Sundberg, R. (1993). Continuum Regression and Ridge Regression. *Food Qual. Prefer.*, 55(3), 653-659.

[188] Vigneau, E., Devaux, M., Qannari, M., Robert, P. (1997). Principal component regression, ridge regression and ridge principal component regression in spectroscopy calibration. *J. Chemometr.*, (11), 239-249.

[189] Li-Xian, S., Reddy, A.M., Matsuda, N., Takatsu, A., Kato, K., Okada, T. (2003). Simultaneous determination of methylene blue and new methylene blue by slab optical waveguide spectroscopy and artificial neural networks. *Anal. Chim. Acta*, 487(1), 109-116.

[190] Petersen, L., Minkkinen, P., Esbensen, K.H. (2005). Representative sampling for reliable data analysis: Theory of Sampling. *Chemometr. Intell. Lab.*, 77, 261-277.

[191] Serneels, S., Croux, C., van Espen, P.J. (2004). Influence properties of partial least squares regression. *Chemometr. Intell. Lab.*, 71, 13–20.

[192] Corona, F., Reinikainen, S.-P., Aaljoki, K., Perkiö, A., Liitiäinen, E., Baratti, R., Simula, O., Lendasse, A. (2008). Wavelength selection using the measure of topological relevance on the self-organizing map. *J. Chemometr.*, 22, 610-620.

[193] Heng, X., Zhichao, L., Wensheng, C., Xueguang, S. (2009). A wavelength selection method based on randomization test for near-infrared spectral analysis. *Chemometr. Intell. Lab.*, 97(2), 189-193.

[194] Igne, B., Hurburgh, C. R. (2010). Local chemometrics for samples and variables: optimizing calibration and standardization processes. *J. Chemometr.*, 24, 75-86.

[195] McLeod, G., Clelland, K., Tapp, H., Kemsley, E.K., Wilson, R.H., Poulter, G., Coombs, D., Hewitt, C.J. (2009). A comparison of variate pre-selection methods for use in partial least squares regression: A case study on NIR spectroscopy applied to monitoring beer fermentation. *J. Food Eng.*, 90, 300-307.

[196] Han, Q.J., Wu, H.L., Cai, C.B., Xu, L., Yu, R.Q. (2008). An ensemble of Monte Carlo uninformative variable elimination for wavelength selection. *Anal. Chim. Acta*, 612, 121-125.

[197] Roger, J.M., Palagos, B., Bertrand, D. (2009). Variable selection for highly multivariate and multi-response calibration. In *Proc. of 14th International Conference on Near Infrared Spectroscopy*, Bangkok, Thailand.

[198] Wensheng, C., Yankun, L., Xueguang, S. (2008). A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemometr. Intell. Lab.*, 90, 188-194.

[199] Gilbert, A. (1991). *The Resolution of Bands in Spectroscopy*. In *Analytical Applications of Spectroscopy II*, Davies, A.M.C. and Creaser, C.S., Eds., Cambridge, Royal Soc. Chem.

[200] Morawski, R.Z., Miękina, A. (1996). Combined Use of Tikhonov Deconvolution and Curve Fitting for Spectrogram Interpretation. *Instrum. Sci. Technolog.*, 24(3), 155-167.

[201] Miękina, A., Morawski, R.Z., Barwicz, A. (1997). The Use of Deconvolution and Iterative Optimization for Spectrogram Interpretation. *IEEE T. Instrum. Meas.*, 46(4), 1049-1053.

[202] Miękina, A., Morawski, R.Z. (1998). Industrial Application of Algorithms for Interpretation of Spectrometric Data. In *Proc. 6th IMEKO Symp. 'Metrology for Quality Control in Production'*, Vienna, Austria, 441-447.

[203] Benjathapanun, N., Boyle, W.J.O., Grattan, K.T.V. (1997). Binary Encoded 2nd-Differential Spectrometry Using UV-Vis Spectral Data and neural Networks in teh Estimation of Spieces Type and Concentration. *IEE P.- Sci. Meas. Tech.*, 144(2), 73-80.

[204] Trygg, J. (Mar. 2002). Everything you need to know about Orthogonal Signal Correction (OSC) filters – and how they can improve interpretation of your data. *Homepage of Chemometrics, http://www.acc.umu.se/~tnkjtg/Chemometrics/Editorial* [2012/09.20].

[205] Hoy, M. (2002). *Building simple, reliable and relevant multivariate data-analysis tools*. Ph.D. Thesis, NTNU, Trondheim, Norway.

[206] Morton, S. (2006). Lecture course covering all practical, and interpretive aspects of electron spectroscopy, *http://www.uksaf.org/tutorials.html* [2011.09.11].

[207] Trygg, J. (May 2002). Wavelets in Chemometrics - compression, denoising and feature extraction. *Homepage of Chemometrics, http://www.acc.umu.se/~tnkjtg/Chemometrics/Editorial* [2012/09.20].

[208] Xiuqi, Z., Jiye, J., Jianbin, Z., Hong, G. (2003). Genetic algorithms based on wavelet transform for resolving simulated overlapped spectra. *Anal. Bioanal. Chem.*, 377(7-8), 1153-1158.

[209] Mark, H., Workman, J. (2000–2003). Chemometrics in Spectroscopy – Analysis of Noise, I-XIV. *Spectroscopy on line, www.spectroscopyonline.com* [2011.08.23].

[210] Brown, C.D. (2000). *Rational Approaches to Data Preprocessing in Multivariate Calibration*, Ph.D. Thesis, Dalhousie University, Halifax, Canada.

[211] Bosch-Ojeda, C., Sanchez-Rojas, F. (2004). Recent developments in derivative ultraviolet/visible absorption spectrophotometry. *Anal Chim. Acta,* 518(1-2), 1-24.

[212] Da, C., Xueguang, S., Bin, H., Qingde, S. (2004). A background and noise elimination method for quantitative calibration of near infrared spectra. *Anal. Chim. Acta*, 511(1), 37-45.

[213] Mazet, V., Carteret, C., Brie, D., Idier, J., Humbert, B. (2005). Background removal from spectra by designing and minimizing a non-quadratic cost function. *Chemometr. Intell. Lab.*, 76, 121-133.

[214] Qu, H.B., Ou, D.L, Cheng, Y.Y. (2005). Background correction in near-infrared spectra of plant extracts by orthogonal signal correction. *J. Zhejiang Univ. Sci. B*, 6(8), 838-843.

[215] Morawski, R.Z., Miękina, A. (2009). Improving Absorbance Spectrum Reconstruction via Spectral Data Decomposition and Pseudo-Baseline Optimization. *IEEE T. Instrum. Meas.*, 58(3), 691-697.

[216] Mark, H., Workman Jr., J. (I, 18(4), 2003; II, 18, (9), 2003; III, 18(12), 2003; 19(1), 2004). Chemometrics – Derivatives in Spectroscopy. *Spectroscopy on line*, *www.spectroscopyonline.com* [2011.08.23]

[217] Badea, I., Moja, D., Vladescu, L. (2002). Determination of para-aminobenzoic acid, a degradation product of procaine hydrochloride, by zero-crossing first-derivative spectrometry. *Anal. Bioanal. Chem.*, 374(1), 51-53.

[218] Ergun, E., Demirata, B., Gumus, G., and Apak, R. (2004). Simultaneous determination of chlorophyll a and chlorophyll b by derivative spectrophotometry. *Anal. Bioanal. Chem.*, 379(5-6), 803-811.

[219] Kharintsev, S. S. and Salakhov, M.K. (2004). A simple method to extract spectral parameters using fractional derivative spectrometry. *Spectrochim. Acta A*, 60, 2125-2133.

[220] Moldovan, Z., Alexandrescu, L. (2002). Derivative spectrophotometry in the determination of phenyl-Î²-naphthylamine used as an antioxidant in rubber mixtures. *Anal. Bioanal. Chem.*, 374(1), 46-50.

[221] Ait Kaddour, A., Cuq, B. (2009). In line monitoring of wet agglomeration of wheat flour using near infrared spectroscopy. *Powder Technol.*, 190(1-2), 10-18.

[222] Yande, L., Xudong, S., Jianmin, Z., Hailiang, Z., Chao, Y. (2010). Linear and nonlinear multivariate regressions for determination sugar content of intact Gannan navel orange by Vis–NIR diffuse reflectance spectroscopy. *Math. Comp. Model.*, 51(11-12), 1438-1443.

[223] Jansson, P.A. (1997). *Deconvolution of Images and Spectra*. London: Academic Press.

[224] Crilly, P.B. (1991). A Quantitative Evaluation of Various Iterative Deconvolution Algorithms. *IEEE T. Instrum. Meas.,* 40(3), 558-562.

[225] Milewski, M.S., Morawski, R.Z., Szczeciński, L. (1993). Comparative Study of Some Iterative and Variational Algorithms of Measurand Reconstruction. In *Proc. IMEKO-TC1&TC7 Colloq.* London, UK, 305-310.

[226] Miękina, A., Morawski, R.Z., Podgórski, A. (1994). Dynamic Reconstrtuction of Measurands and Calibration of Measuring Systems. *Measurement*, (14), 63-72.

[227] Morawski, R.Z., Szczeciński, L., Barwicz, A. (1995). Deconvolution Algorithms for Instrumental Applications: A Comparative Study. *J. Chemometr.*, 9, 3-20.

[228] (2006). Introducing a Personal Robot with a Sense of Taste. *NEC System Technologies Ltd. News*, *http://www.necst.co.jp/english/news/20060801/index.htm* [2012.09.20].

[229] Examples of spectrophotometric wine analysers: *http://www.foss.dk*; *http://www.randoxfooddiagnostics.com*; *http://www.mikochem.gr*; *http://www.brukeroptics.com; http://www.tridentinstrumentation.co.za* [2012.09.22].

[230] *Application of spectroscopy for non-destructive 'in-bottle' measurement of wine composition and quality*. CRCV Technology Application Note, April 2007. *www.crcv.com.au/viticare/vitinotes/Viti-Notes/* [2012.10.11].

[231] Cozzolino, D., Cynkar, W.U., Dambergs, R.G., Janik, L., Gishen, M. (2007). Near infrared spectroscopy in the Australian grape and wine industry. *CRCV Technology Application Note*, *www.crcv.com.au/viticare/vitinotes/Viti-Notes/* [2012.10.11].

[232] Araújo, A.N., Lima, J.F.L.C., Rangel, A.O.S.S., Segundo, M.A. (2000). Sequential injection system for the spectrophotometric determination of reducing sugars in wines. *Talanta*, 52, 59-66.

[233] Bevin, C.,J., Dambergs, R.G., Fergusson, A.J., Cozzolino, D. (2008). Varietal discrimination of Australian wines by means of mid-infrared spectroscopy and multivariate analysis. *Anal Chim Acta*, (621), 19-23.

[234] Cozzalino, D., Parker, M., Dambergs, R.G., Herderich, M., Gishen, M. (2006). Chemometrics and Visible-Near Infrared Spectroscopic Monitoring of Red Wine Fermentation in a Pilot Scale. *Biotechnology and Bioengineering*, 95(6), 1101-1107.

[235] Cozzolino, D. (2007). Non-Destructive Analysis by Vis-NIR Spectroscopy of Fluid (S) in its Original Container, International Patent No WO2007006099.

[236] Fei, L., Yong, H., Li, W., Hongming, P. (2007). Feasibility of the use of visible and near infrared spectroscopy to assess soluble solids content and pH of rice wines. *J. Food Eng.*, 83, 430-435.

[237] Ferreira, A.P., Alves, T.P., Menezes, J.C. (2005). Monitoring Complex Media Fermentations with Near-Infrared Spectroscopy: Comparison of Different Variable Selection Methods. *Biotechnol. Bioeng*, 91(4), 474-481.

[238] Herrera, J., Guesalaga, A., Agosin, E. (2003). Shortwave-near infrared spectroscopy for nondestructive determination of maturity of wine grapes. *Meas. Sci. Technol.*, 14, 689-697.

[239] Liu, L., Cozzolino, D., Cynkar, W.U., Dambergs, R.G., Janik, L., O'Neill, B.K., Colby, C.B., Gishen, M. (2008). Preliminary study on the application of visible–near infrared spectroscopy and chemometrics to classify Riesling wines from di.erent countries. *Food Chem.*, 106, 781-786.

[240] Moreira, J.L. Santos, L. (2004). Spectroscopic interferences in Fourier transform infrared wine analysis. *Anal. Chim Acta*, 513, 263-268.

[241] Riganakos, K.A., Veltsistas, P.G. (2003). Comparative spectrophotometric determination of the total iron content in various white and red Greek wines. *Food Chem.*, 82, 637-643.

[242] Sáiz-Abajo, M.J., González-Sáiz, J.M., Pizarro, C. (2006). Prediction of organic acids and other quality parameters of wine vinegar by near-infrared spectroscopy. A feasibility study. *Food Chem.*, 99, 615-621.

[243] Serapinas, P., Venskutonis, P.R., Aninkevičius, V., Ezerinskis, Ž., Galdikas, A., Juzikině, V. (2008). Step by step approach to multi-element data analysis in testing the provenance of wines. *Food Chem.,* 107, 1652-1660.

[244] Skogerson, K., Downey, M., Mazza, M., Boulton, R. (2007). Rapid Determination of Phenolic Components in Red Wines from UV-Visible Spectra and the Method of Partial Least Squares. *Am. J. Enol. Viticult.*, 58(3), 318-325.

[245] Soriano, A., Perez-Juan, P.M., González, J.M., Perez-Coello, M.S. (2007). Determination of anthocyanins in red wine using a newly developed method based on Fourier transform infrared spectroscopy. *Food Chem.*, 104, 1295-1303.

[246] Foca, G., Cocchi, M., Li Vigni, M., Caramanico, R., Corbellini, M., Ulrici, A. (2009). Different feature selection strategies in the wavelet domain applied to NIR-based quality classification models of bread wheat flours. *Chemometr. Intell. Lab.*, 99, 91-100.

[247] Niedziński, C., Morawski, R.Z. (2004). Estimation of low concentrations in presence of high concentrations using Bayesian algorithms for interpretation of spectrophotometric data. *J. Chemometr.*, 18, 217-230.

[248] Morawski, R. Z., Niedziński, C. (2008). Application of a Bayesian estimator for identification of edible oils on the basis of spectrophotometric data. *Metrol. Meas. Syst.*, 15(3), 247-266.

[249] Pérez-Marín, D., Fearn, T., Guerrero, J. E., Garrido-Varo, A. (2012). Improving NIRS predictions of ingredient composition in compound feedingstuffs using Bayesian non-parametric calibrations. *Chemometr. Intell. Lab.*, 110, 108-112.