ARTICLES =

Application of the Curve Resolution Method to the Preprocessing Spectral Data in Two-Layer Systems

O. Ye. Rodionova^{a, *} and A. L. Pomerantsev^{a, b}

^aInstitute of Chemical Physics, Russian Academy of Sciences, ul. Kosygina 4, Moscow, 119991 Russia ^bInstitute of Natural and Technical Systems, Russian Academy of Sciences, Kurortnyi pr. 99/18, Sochi, 354024 Russia *e-mail: rcs@chph.ras.ru

Received October 21, 2014; in final form, July 2, 2015

Abstract—An approach is proposed for recovering individual spectra of components in two-layer systems using the method of multivariate curve resolution—alternating least squares. The method is applicable to recording spectra in the near infrared region in both transmission and diffuse reflectance modes. The efficiency of the method is tested on a model system polymer film—polymer film and on real samples polymer packaging film—infinite layer of powdered product.

Keywords: chemometrics, near infrared spectroscopy, signal separation, polyethylene, taurine **DOI:** 10.1134/S1061934816010111

The method of near infrared (NIR) spectroscopy (12 500–4000 cm⁻¹) in conjunction with chemometric treatment, called a NIR analysis, gives a way to control the chemical and physical properties of pharmaceutical products, both medicines and substances, without sample preparation [1, 2]. Spectra can be recorded directly through the packaging, such as polyethylene (PE) bags [3], glass ampoules [4], plastic bottles [5], or polyvinylchloride blisters [6, 7]. In the cases of solid and powdered samples, NIR spectra are recorded predominantly in a diffuse reflectance mode. The application of fiber-optic probes enables measurements in not only the laboratory but also right at warehouse.

After building a chemometric model based on a spectral library, the confirmation of the quality of new samples takes a few minutes. In particular, this enables an operator to check each package of input raw materials, which comply with modern standards of analytical control of pharmaceutical products [8, 9]. It should be considered that plastic packaging materials have their own absorption bands in the NIR region, so the recorded spectra are a complex superposition of the absorption bands of a substance and the packing. A random variation in the thickness of package or its layers affects the recorded spectrum, which can lead to erroneous test results.

The goal is to develop a procedure to evaluate the contribution of packing material in the recorded spectrum and, where possible, to separate the sample spectrum and the spectrum of the packing. The need for such a procedure is demonstrated on a real example of assessing the authenticity of incoming raw materials in the warehouse of a pharmaceutical company. In order to assess quickly the quality of the substance without opening the polyethylene package, it was proposed to apply NIR analysis [3] and to use the spectra accumulated in the library for solving the problem of authentication.

From a mathematical point of view, authentication is the task of building of a one-class classifier [10, 11], which helps to answer the question whether the new sample belongs to the target class (the class of authentic samples of the substance under study) or not. To develop the classifier, one needs to collect spectra of samples that knowingly belong to the target class. This set, called a training set, should take into account natural variations both inside the same batch of substance and between different series. Based on this training set, a model of authentication is built and the risks of taking wrong decisions, that is, the errors of the first type (share of wrongly rejected samples of the target class) and the second type (share of alien samples incorrectly assigned to the target class), is assessed [11, 12].

The quality of the training set substantially affects the final results. In the following example [3], a preliminary analysis showed that the set of spectra, collected based on knowingly quality samples, is not uniform. The cause of the spread is that the spectra of substance are distorted by recording them through a layer of polyethylene, the thickness of which or, more precisely, the number of layers of packing material, is not the same in different measurements. We have developed an iterative procedure for separating samples into groups depending on the number of polyethylene layers. In the first approximation, we identified two groups: a group of samples A0, the spectra of which are not distorted, that are then used to build the classifica-

JOURNAL OF ANALYTICAL CHEMISTRY Vol. 71 2016 No. 1

tion model, and a group of "suspicious samples" Al [3, 12]. Subsequent analysis [11] showed that group A1 is divided into four subgroups A1,..., A4, determined, apparently, by the number of polyethylene layers, through which the measurements were made. Detailed analysis of samples not included in the target class A0, helped to estimate the errors of the first and second types. This procedure used for the separation of the samples into classes was rather time-consuming, so we need to find a simpler tool for analyzing data for widespread use.

In the present work, the method of decomposition of multivariate signals [13], called multivariate curve resolution-alternating least squares (MCR-ALS), is used as a mathematical tool for preliminary analysis of the spectra of samples. This method is widely used to study the kinetics of different processes, because it offers the evaluation of the spectra of intermediates formed during reactions [14]. Updating the algorithm with kinetic equations transfers the method from the category of a formal approach to meaningful physical and chemical simulation [15]. Recently, this approach has been widely used for the quantitative analysis when it is required to determine the concentrations of the components of complex mixtures by spectral data. For this purpose, additional correlation constraints are introduced into the algorithm [16]. The application of MCR-ALS, in contrast to the widely used method of projection to latent structures, enables not only the evaluation of the concentrations of components but also the obtaining of their "pure" spectra [17]. Recently, it is shown that the nonlinear version of MCR-ALS is capable of solving the problems of calibration even when the spectra of mixtures have saturation in certain regions [18].

CHEMOMETRIC METHODS

Principal component analysis (PCA) [19, 20] is used as an independent method for the primary data analysis. Additionally, PCA is widely used in chemometrics as an integral part of other methods: classification, curve resolution, calibration, etc.

Consider a $(I \times J)$ matrix X, which is obtained after routine preprocessing of the original matrix of spectra X_{raw} . The matrix X consists of the spectra of I samples obtained for J wavelengths. The decomposition of matrix X using PCA is determined by equation

$$X = TP^t + E, (1)$$

where $T = \{t_{ia}\}$ is the score matrix of dimension $(I \times A)$, P = { p_{ja} } is the loading matrix of dimension ($J \times A$), E = { e_{ij} } is the error matrix of dimension ($I \times J$), and Ais the number of principal components. The decomposition proceeds so that the matrix $\Lambda = T^{t}T = \text{diag}(\lambda_{1})$ \dots, λ_A is a diagonal $(A \times A)$ matrix, where the diagonal

elements $\lambda_a = \sum_{i=1}^{I} t_{ia}^2$ are the eigenvalues of matrix X^tX, arranged in a descending order.

Classification. The one-class classification method, namely, data driven soft independent modeling of class analogy (DD-SIMCA) [11, 12, 21], involves the following steps:

(1) Construction of a PCA model of the target class, describing the totality of the samples.

(2) Construction of the decision area separating the samples of the target class from all other samples.

(3) Risk assessment of making wrong decisions, that are the errors of the first and second type.

In DD-SIMCA, two values are used for each sample, characterizing both the objects of the training set and new samples.

The first value is the distance within the PCA model

$$h_i = t_i^t (T^t T)^{-1} t_i = \sum_{a=1}^A \frac{t_{ia}^2}{\lambda_a}, i = 1, ..., I.$$
 (2)

The second value is the orthogonal distance to the model.

$$v_i = \sum_{j=1}^{J} e_{ij}^2, i = 1,..., I.$$
 (3)

The total distance q_i is introduced based on these two distances [11, 12], which is computed as a weighted sum of the distances described by Eqs. (2) and (3) by equation

$$q_i = N_h \frac{h_i}{h_0} + N_v \frac{V_i}{V_0}, i = 1, ..., I.$$
(4)

Given that the values of h, v, and q obey the chisquare distribution with the degrees of freedom of, respectively, N_{v} , N_{h} , $N_{v} + N_{h}$, the belonging to a class is determined by equation

$$q \le \chi^{-2}(1-\alpha, N_h + N_v),$$
 (5)

where α is the error of the first type.

Method of curve resolution. MCR-ALS is one of the widely used methods for processing the experimental results [20, 22]. The method is based on the bilinear model

$$X = CS^t + E, (6)$$

wherein X is an $(I \times J)$ matrix containing the spectra of I samples at J wavelengths, C is an $(I \times N)$ concentration matrix of pure components, and S is a $(J \times N)$ matrix of the spectra of pure components. N is the number of pure components in the system, and E is an $(I \times J)$ matrix of errors.

From a computational point of view, MCR-ALS is an iterative procedure, in which, using the least squares method (LSM)

Sample	Number of PE layers	Conventional concentration	
		PS	PE
PS0	0	1.01	0.03
PS1	1	1.00	0.98
PS3	3	0.99	3.00
PS5	5	0.96	5.01
PS7	7	1.00	7.05

Polystyrene with several layers of polyethylene film

minimize
$$\left\| \mathbf{X} - \mathbf{CS}^{t} \right\|^{2}$$
 (7)

matrices S and C are alternately determined. At the Cstep, the value of $S \equiv S_{hat}$ is fixed, and the next approximation for the concentration profile, that is, matrix C_{in} , is calculated

$$C_{in} = XS_{hat} (S_{hat}^{t}S_{hat})^{-1}.$$
 (8)

At the S-step, $C \equiv C_{hat}$ is fixed, and the LSM estimation of the matrix of spectra is found, that is,

$$\mathbf{S}_{\text{in}} = \mathbf{X}^{\mathsf{t}} \mathbf{C}_{\text{hat}} (\mathbf{C}_{\text{hat}}^{\mathsf{t}} \mathbf{C}_{\text{hat}})^{-1}.$$
(9)

The procedure is performed until the process converges, which depends on how successful were selected initial approximations, and how great is the norm of matrix E.

It is important to note that the solution of Eq. (6) is not unique, and there are rotational and scale ambiguities [20]. To get rid of rotational ambiguity and to obtain estimates of concentrations and spectra with a physical sense, various constraints are included in an iterative procedure, for example, the non-negativity of concentrations, non-negativity of spectra, etc.; the scale uncertainty remain the same.

The MCR–ALS procedure requires a preliminary stage, that is, initialization. At this stage, the number of pure components in the mixture is estimated, and initial approximations for either the matrix S of spectra or the matrix C of concentrations are found. If there is no additional information, an initial approximation for the matrix S is generally searched by any suitable method, for example, via SIMPLISMA [22].

In this case, an averaged spectrum of polyethylene was used as an initial approximation for one of components, and a difference between the spectrum of polyethylene and an average spectrum of the whole system was taken for another component. MCR–ALS is an absolute method, as it does not require knowledge of the spectra of the initial components. However, the inclusion of such information into the procedure simplifies calculations. All calculations were performed using a Chemometrics Add-In software for Excel [20].

EXPERIMENTAL

Experiment 1. Samples of multilayer objects consisting of a single layer of polystyrene (PS) film with a thickness of 3 mm and several coating layers of low-density polyethylene film with a thickness of 0.087 mm. In total, five samples are studied (table).

Measurements were performed using a Nicolet 6700 Fourier transform spectrometer (ThermoFisher) in a transmission mode in the range of 12 500–4000 cm⁻¹ with increments of 8 cm⁻¹. An empty sample compartment was measured as a background. A polystyrene film involved in the experiment is often used as a reference sample for verifying the accuracy of the wave scale of NIR spectrometers in the transmission mode.

Experiment 2. Test samples were pharmaceutical substance taurine (2-aminoethanesulfonic acid, $C_2H_7NO_3S$) in the form of a white crystalline powder, prepacked in 20-kg polyethylene bags. Each packet is stored in a cardboard cylinder with a lid. For the measurement, a Spectrum 100N Fourier transform spectrometer (PerkinElmer) was used, equipped with a fiber-optic probe. NIR spectra were recorded in a diffuse reflectance mode in the spectral range of 10 000- 4000 cm^{-1} with an increment of 2 cm^{-1} . The measurements were mainly performed directly through the polyethylene packaging. We tested bags in 82 cylinders: at least three spectra were recorded for each bag. In total, 246 spectra were obtained. Three bags were opened, and five repetitive measurements of substance without packaging were conducted. We also obtained spectra of 18 empty polyethylene packages, which originally contained the substance. A detailed description of the experiment is given in [3].

RESULTS AND DISCUSSION

Experiment 1. In this case, the recorded spectrum is the superposition of the spectra of layers, at least at a small change in the thickness of each layer; that is, the recorded spectrum is the sum of the spectra of polyethylene and polystyrene. By increasing the number of layers of polyethylene, its absorption peaks increase proportionally to the total thickness of polyethylene layer. The heights of the characteristic peaks of polystyrene (8720-8430 and 5000-4480 cm⁻¹) do not change, if they are located in the transparency windows of polyethylene, and are combined with the polyethylene peaks (5900–5600 cm⁻¹) if they overlap (Fig. 1). This is consistent with the Lambert–Beer law, considering that the concentration of components in the test system do not change, while the optical path length varies according to the number of polyethylene layers.



Fig. 1. Experimental NIR spectra of polystyrene film coated with several layers of polyethylene film; number of film layers: (PS0) 0, (PS3) 3, and (PS5) 5 (table).

Using the MCR-ALS procedure for curve resolution, we can obtain the spectra of pure components and estimate the number of layers of the polyethylene film. For this purpose, a matrix of NIR spectra with I = 5, J = 2360, N = 2 was used in Eq. (6) as matrix X. The spectra of pure components (matrix S) are shown in Fig. 2.

The concentration matrix C, presented in the last columns of the table, consists of five rows (number of samples) and two columns (arbitrary concentrations). The first column corresponds to the concentration of polystyrene, and all these values are close to 1. The values in the second column, corresponding to polyethylene, grow linearly and approximate the number of layers. In this regard, it is important to clarify the following significant issue. The term of "concentration matrix," usually applied in the problems of the curve resolution, in this approach is used conditionally, as the concentrations themselves do not change, but the optical path length changes, which affects the amount of absorption. Hereafter, we will use the term of "pseudo concentration."

Thus, in the simplest case, the possibility of recovery of pure spectra of two-component layered system is shown when the thickness of one layer changes. We emphasize that for restoration of spectra, we do not use information about one or more components.

Experiment 2. A two-component system, more complex than in the first case, is considered. The samples under study were a polyethylene film ~ 0.060 mm in thickness and a layer of powdery substance, which can be considered infinite, since its thickness is much greater than the penetration depth of the source beam. The substance is contained in a closed bag; therefore, the actual thickness of the polyethylene film trapped under the sensor probe is unknown beforehand because of the folds of polyethylene, even when the same bag is measured in different parts of the package.



Fig. 2. (*1*, thin solid line) Measured and (*2*, thick gray line) recovered spectra of (a) polyethylene and (b) polystyrene.

The spectra of the test samples after the MSC correction [19] are presented in Fig. 3. In the spectra, there are the absorption bands of the substance (approximately 8510 and 4720 cm⁻¹), the absorption band of polyethylene (approximately 8100 and 5770 cm⁻¹), and the overlapping peaks of both components.

In distinction from the recording of the spectra in the transmission mode, when operating in the diffuse reflectance mode, not only the spectrum of polyethylene overlaps the spectrum of substance, but also the signal of the substance beneath a layer of polyethylene is attenuated (Fig. 4). The study of the spectra showed that the effect of polyethylene, that is, the effective thickness of the polyethylene layer is different for different samples.



60

Fig. 3. Spectra in Experiment 2 after MSC-correction; characteristic absorption bands of taurine and polyethylene are marked.



Fig. 4. Characteristic absorption spectra; (A0) a sample with a light effect of polyethylene, (A3) a sample of strong effect of polyethylene, (Taurine) a substance from an open package, and (PE) empty polyethylene bag folded in several layers.

In order to separate the spectra measured through one layer of polyethylene from the spectra poorly recorded through several layers and thereby to select the training set, we applied the MCR–ALS procedure. In Eq. (6), matrix X is the matrix of the spectra of all samples in bags, matrix C is the matrix of pseudo concentrations, and matrix S is the matrix of the spectra of pure components. The spectra of polyethylene and the substance are called the spectra of pure components here. Term "pseudo concentrations" refers to the ratio between the contributions from the spectra of the substance and polyethylene. The number of samples is I =246; the number of wavelengths is J = 2301. The working spectral range is 9000–4450 cm⁻¹. The number of components N is 2.

The MCR-ALS procedure enabled us to estimate the value of C_h that is a pseudo concentration of poly-



Fig. 5. Analysis of samples in Experiment 2: (a) relation between the thickness of polyethylene (C_h) and the division into groups by DD-SIMCA (q) (values in the axes are represented in arbitrary units) and (b) absorption spectra of the substance; (1, thin solid line) measured and (2, thick gray line) recovered spectra.

ethylene in samples, in other words, to determine the relative thickness of the polyethylene film, which was our goal. Now, the samples can be arranged in ascending C_h (the number of polyethylene layers), and the results can be compared with the previously conducted analysis [11]. In Fig. 5a, the abscissa represents the values of the first column of matrix C, that is, the relative thickness of the polyethylene layer C_h . The ordinate axis is the logarithm of distances q_i , calcu-

- Shpak, A.V., and Shpigun, O.A., Anal. Bioanal. Chem., 2010, vol. 397, no. 5, p. 1927.
- 1. Roggo, Y., Chalus, P., Lene Maurer, L., Lema-Martinez, C., Edmond, A., and Jent, N., J. Pharm. Biomed.
- Anal., 2007, vol. 44, no. 3, p. 683.
- Anal. Chem., 2010, vol. 29, no. 8, p. 795.
- 3. Rodionova, Sokovikov, O.Ye., Y.V., and Pomerantsev, A.L., Anal. Chim. Acta, 2009, vol. 642, nos. 1–2, p. 222.
- 4. Rodionova, O.Ye., Pomerantsev, A.L., Houmuller, L.,

lated in solving the problem of classification by Eq. (4). Each series corresponds to the group of sam-

ples, studied earlier [11] in the solution of the problem

previously developed division into groups is in good

agreement with an estimate of the thickness of poly-

ethylene laver. In addition, MCR–ALS offers a rather

precise reconstruction of the pure spectrum of sub-

stance (Fig. 5b). Since the packaging should not be

violated in the routine monitoring, it is important to

obtain this spectrum to make a decision about the

film-polymer film and polymer film-infinite powder

sample can be simply reconstructed using the MCR-

ALS method. The procedure is applicable both to

recording NIR spectra in the transmission mode and

to measurements in the diffuse reflectance mode. The

relative thickness of the polymer film can also be

found. Since the beginning of the procedure for the

curve resolution requires only a first approximation

rather that measurements of the spectra of pure com-

ponents, the procedure can be classified as absolute.

This approach considerably simplifies the analysis of

the results performed through polymer films.

Thus, the spectra of two-layer systems polymer

quality of the substance.

- 2. Rodionova, O.Ye. and Pomerantsev, A.L., TrAC, Trends

- REFERENCES
- 2006, vol. 36, nos. 3–4, p. 163. 15. Diewok, J., de Juan, A., Maeder, M., Tauler, R., and
- Lendl, B., Anal. Chem., 2003, vol. 75, no. 3, p. 641.
- 16. de Oliveira, R.R., de Lima, K.M.G., Tauler, R., and
- de Juan, A., Talanta, 2014, vol. 125, p. 233.
- 17. Azzouz, T. and Tauler, R., Talanta, 2008, vol. 74, p. 1201.
- 18. Pomerantsev, A.L., Zontov, Y.V., and Rodiono
 - va, O.Ye., J. Chemom., 2014, vol. 28, no. 10, p. 740. 19. Esbensen, K.H., Multivariate Data Analysis-in Prac-
 - tice. An Introduction to Multivariate Data Analysis and Experimental Design, Oslo: CAMO, 2001, 5th ed.
 - 20. Pomerantsev, A.L., Chemometrics in Excel, Hoboken: Wiley, 2014.
 - 21. Pomerantsev, A.L., J. Chemom., 2008, vol. 22, nos. 11-12, p. 601.
 - 22. Juan, A., Rutan, S.C., Maeder, M., and Tauler, R., in Comprehensive Chemometrics, vol. 2, Brown, D., Tauler, R., and Walczak, B., Eds., Amsterdam: Elsevier, 2009, p. 325.

Translated by O. Zhukova

- 5. Broad, N.W., Jee, R.D., Moffat, A.C., Eaves, M.J., Mann, W.C., and Dziki, W., Analyst, 2000, vol. 125, no. 11, p. 2054.
 - 6. Rodionova, O.Ye., Balyklova, K.S., Titova, A.V., and Pomerantsev, A.L., J. Pharm. Biomed. Anal., 2014, vol. 98, no. 9, p. 186.
 - 7. Guoa, Q., Wub, W., Massart, D.L., Bouconc, C., and de Jong, S., Anal. Chim. Acta, 2001, vol. 446, nos. 1-2, p. 85.
 - 8. U.S. Food and Drug Administration, Guidance for Industry, PAT: a Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assur-
 - ance, September 2004.
 - 9. Pomerantsev, A.L. and Rodionova, O.Ye., J. Chemom., 2012, vol. 26, no. 6, p. 299.

 - 10. Tax, D., Doctoral Dissertation, Delft, the Netherlands: Univ. of Delft, 2001.
 - 11. Pomerantsev, A.L. and Rodionova, O.Ye., J. Chemom.,
 - 2014, vol. 28, no. 6, p. 518.

 - 12. Pomerantsev, A.L. and Rodionova, O.Ye., J. Chemom., 2014, vol. 28, no. 5, p. 429.
 - 13. Monakhova, Yu.B., Astakhov, S.A., Mushtakova, S.P.,
 - p. 351.
 - and Gribov, L.A., J. Anal. Chem., 2011, vol. 66, no. 4, 14. de Juan, A. and Tauler, R., CRC Crit. Rev. Anal. Chem.,

This article is protected by the copyright law. You may copy and distribute this article for your personal use only. Other uses are only allowed with written permission by the copyright holder