Trends in Analytical Chemistry 78 (2016) 17–22

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Discriminant analysis is an inappropriate method of authentication

Oxana Ye. Rodionova [a,b,*], Anna V. Titova [b,c], Alexey L. Pomerantsev [a]

[a] Semenov Institute of Chemical Physics RAS, Kosygin str. 4, 119991 Moscow, Russia
[b] Information and Methodological Center for Expertise, Stocktaking and Analysis of Circulation of Medical Products, Slavyanskaya sq., 4-1, 109074 Moscow, Russia
[c] Pirogov Russian National Research Medical University, Ostrovityanov str., 1, 117997 Moscow, Russia

ARTICLE INFO

Keywords:
Authentication
Class modeling
One-class classifier
PLS-DA
DD-SIMCA

ABSTRACT

Authentication is the process of determining whether an object is, in fact, what it is declared to be. In practice, this problem is often solved by using discriminant methods. In this paper, we explain that such techniques do a poor authentication job. The main drawback of discriminant methods is their inability of proper classification of new samples, which do not belong to any of the predefined classes. Our considerations are illustrated by a real-world example and a comparison of the results provided by the following two methods: Partial Least Squares- Discriminant Analysis, PLS-DA, and Data Driven Soft Independent Modeling of Class Analogy, DD-SIMCA.

Contents

1. Introduction

Authentication is the process of determining whether an object is, in fact, what it is declared to be. In some cases, the answer is found by means of direct chemical analysis, which confirms that the product quality meets technical/regularity documentation. As a rule, these analyses are time and labor consuming. Another approach is to conduct some quick, relatively cheap, and often non-destructive measurements with subsequent data processing by means of chemometrics. Typical authentication problems, relevant for analytical chemistry and chemometrics, are counterfeit drug detection [1,2], food adulteration detection [3–7], identification of illegal additives in fuels [8,9], and confirmation of geographical origin of products [10,11].

When claiming authentication as a goal, analysts often substitute authentication task with solving discrimination problems. In a recently published review [12], 42 food authentication studies are presented. Out of 56 chemometric methods applied, in total, there

* Corresponding author. Tel.:+7(495)9397483; fax: +7(495)9397483.
E-mail address: rcs@chph.ras.ru (O. Ye Rodionova).

http://dx.doi.org/10.1016/j.trac.2016.01.010
0165-9936/

were 46 various discrimination methods, 5 SIMCA (Soft Independent Modeling of Class Analogy), and 5 unsupervised PCA (Principal Component Analysis). Discriminant analysis perfectly suits a task of separation of samples originated from two different regions (Region 1 and Region 2). In this case, we have to identify the most important features that distinguish the two sets of samples. Confirmation that a specific sample truly originated from Region 1 is a different task. In this case, we have to identify general properties that characterize a set of samples originated from Region 1 independently of other classes and regions.

The term 'classification' is very often used as a synonym of discriminant analysis methods, because they assign objects to predefined classes. In its turn, methods used for solving authentication problems comprise a separate class among pattern recognition/classification techniques. These methods are called one-class classifiers [13], or class modeling [14,15]. The main properties of the latter group of methods and their distinctions from discrimination techniques are analyzed in this study. A real world example illustrates the differences in the results of application of these two approaches using PLS-DA (Partial Least Squares- Discriminant Analysis) and DD-SIMCA (Data Driven SIMCA). It is worth mentioning that this work is directed towards a proper treatment of the authentication problem, and by no means against discriminat analysis, in general, and PLS-DA, in particular.

2. Theory

2.1. The main steps of the authentication problem

2.1.1. Definition of a target class

General requirement of any authentication procedure is that a genuine class has to be known. We refer to this class as a target class. The target class is always unique for a given authentication problem. Any other objects, or classes of objects, that are not members of the target class are considered aliens, or, depending on their specific task, as counterfeits, forgeries, frauds, etc. Aliens very often belong to different classes, which are referred to as alternative classes, or do not belong to any specific class at all. Unlike discriminant analysis, aliens are not used for target class modeling. For example, genuine roasted coffee (target class) could be contaminated by a mixture of cheaper products such as twigs, coffee berry skin and parchment, spent coffee grounds, roasted barley, corn and other grains [6]. Each way of adulteration (say, coffee berry skin plus roasted barley) forms a separate class that differs from other alternative classes. Moreover, each new adulterant (e.g. roasted acorn) immediately generates a series of new alternative classes. The alternative class membership is far beyond the objectives of authentication.

Target class is denoted by the properties of its representative members. These properties, also called fingerprints, are multivariate analytical signals acquired by means of spectroscopy [1,12], chromatography [7], electro-analytical [8], or other analytical techniques. The results of fingerprints collection can be presented in data matrices (or tensors). The main matrix $\mathbf{X}_0$ is a set of data obtained using target class samples. Thus, the first step of the authentication problem is collection of representative data, which undoubtedly belong to the target class. These data are divided into training and validation sets.

2.1.2. Data processing

The analysis of these datasets is performed by means of chemometrics. The most popular methods of analysis are UNEQ (Unequal Dispersed Classes) [16], SIMCA [17] and its later modifications such as robust SIMCA [18] and data driven SIMCA [19], as well as a machine learning method SVDD (Support Vector Domain Description) [20]. Certainly, there are many other methods developed to solve specific authentication problems.

Table 1
Fundamental differences between authentication and discrimination problems

Authentication problems	Discrimination problems
The goal	
Determination whether an object is, in fact, what it is declared to be	Determination of a membership of an object to one of the predefined classes
Data sets	
Objects that represent a target class	Several sets of objects that represent predefined classes
Statistical/Chemometric methods	
UNECO, SIMCA, SVDD, etc...	LDA, QDA, PLS-DA, SVM, etc ...
Result of data modeling/ Decision rule development	
Decision rule for a given α value	Boundaries/delineators between classes
Figures of merits	
Sensitivity is given a priori. Specificity can be found theoretically when an alternative class is given.	Sensitivity and specificity are found empirically post factum

In the context of authenticity, whatever chemometric technique is employed, it has to develop a decision rule, which helps answering the main question – whether a new sample belongs to the target class or not. The decision rule may take the form of an acceptance area and/or values of thresholds. Undoubtedly, the established rule should be carefully trained using collected fingerprint data, and it has to be suspiciously validated against new genuine objects. If available, this rule should be tested against alien samples.

2.1.3. Figures of merit

Even the intensive training and validation measures cannot prevent us from unavoidable decision errors [21]. Ordinarily, the results of classification are described in terms of 'sensitivity' and 'specificity'. Sensitivity denotes a share of correctly identified samples of the target class. Specificity is a portion of objects of an alternative class that were correctly identified as members of that alternative class. Definitions of sensitivity and specificity are often based on notations like 'true positive', 'true negative', etc., that have different meanings in various application domains such as clinical trials, medical diagnostics and chemical analyses. We prefer to use traditional statistical terms as the type I error, α, and the type II error, β. The first error, α, is the rate of wrong rejections of the target samples, while β is the rate of wrong acceptances of aliens as target objects. Both errors are harmful and, generally, for a given dataset, an effort to reduce one type of error results in an increase in the other type of error [22,23]. Following statistical terminology, sensitivity can be defined as $100\,(1\text{-}\alpha)\%$ and specificity as $100\,(1\text{-}\beta)\%$. It should be mentioned that these are statistical measures of the performance of a binary classification test. As to one-class classifiers, theoretically, alien objects could be as close to a target class as possible. In this case, the β value tends to $1\text{-}\alpha$. Often, the value of α error is evaluated empirically, subsequent to the model development. Sometimes, e.g. in DD-SIMCA, it is established theoretically, using data driven distributions [22]. In some special cases of authentication problems, β error can be evaluated for a predefined alternative class. If several alternative classes are available, the β error is calculated regarding each alternative class separately.

The main differences between authentication and discrimination problems are summarized in Table 1.

2.2. PLS-DA

PLS-DA is an effective technique frequently applied in chemometrics, its detailed description may be found elsewhere

[24,25]. It is a conventional PLS regression method, where the $(I \times J)$ matrix $\mathbf{X}$ is a predictor matrix, and the $(I \times K)$ response matrix $\mathbf{Y}$ comprises categorical (dummy) variables that describe class memberships. K is equal to the number of classes. If only two classes are considered, matrix $\mathbf{Y}$ is reduced to a $(I \times 1)$ vector $\mathbf{y}$. When PLS regression is developed, the response value $\mathbf{Y}_{pred}$ is predicted for a new sample. The decision is based on the comparison of $\mathbf{Y}_{pred}$ with given categorical variables in $\mathbf{Y}$. The sample is attributed to the class, which has minimal distance between $\mathbf{Y}$ and $\mathbf{Y}_{pred}$.

2.3. DD-SIMCA

A modification of the well-known SIMCA method [17,26] called DD-SIMCA [22,23] is used as authentication technique. The method consists of two steps. At a first step, the Principal Component Analysis (PCA) [25] is applied to the training data from the target class. The $(I \times J)$ data matrix $\mathbf{X}$ (duly preprocessed, e.g. centered) is decomposed

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E} \tag{1}$$

where $\mathbf{T} = \{t_{ia}\}$ is the $(I \times A)$ scores matrix; $\mathbf{P} = \{p_{ja}\}$ is the $(J \times A)$ loadings matrix; $\mathbf{E} = \{e_{ij}\}$ is the $(I \times J)$ matrix of residuals; and A is the number of principal components (PC). Matrix $\mathbf{T}^t\mathbf{T} = \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_A)$ is a diagonal with elements $\lambda_a = \sum_{i=1}^{I} t_{ia}^2$, which are the eigenvalues of matrix $\mathbf{X}^t\mathbf{X}$ ranked in descending order.

At the second step, for each object $i = 1, \ldots, I$ from the training set, two distances are calculated. They are the score distance (SD), h_i, and the orthogonal distance (OD), v_i:

$$h_i = \mathbf{t}_i^t (\mathbf{T}^t\mathbf{T})^{-1} \mathbf{t}_i = \sum_{a=1}^{A} \frac{t_{ia}^2}{\lambda_a}, \quad v_i = \sum_{j=1}^{J} e_{ij}^2 \tag{2}$$

The SD represents the position of a sample within the score space, and the OD characterizes the distance of the sample to the score space. DD-SIMCA adds the possibility of estimation the data-driven distribution parameters, which are the mean values v_0 and h_0, and the numbers of the degrees of freedom N_h and N_v for the SD $\{h_i\}$ and OD $\{v_i\}$ respectively. Thus, we can develop an acceptance area /decision rule for a given value α [22].

Optionally, in case an alternative class is available, DD-SIMCA provides the possibility to calculate the type II β error and construct the corresponding extended acceptance area, which guarantees that the risk of accepting a sample from the alternative class is not greater than β [23].

3. Experimental data

To illustrate the differences between discrimination and authentication problems we use a part of a big dataset presented in [21]. The data contain the NIR spectra acquired for uncoated tablets of Amlodipine, calcium channel blockers, produced by 3 different manufacturers, denoted as A3, A4, A7. The names of the producers correspond to those used in [21]. The aim of the study is proper authentication of medicines performed in the framework of counterfeit drug detection.

The decision rule for each medicine should be as general as possible taking into account natural variations in the course of manufacturing. At the same time, it is very important to reveal even ‘high quality fakes’, i.e. samples that are very similar to the genuine ones but not produced by a specific manufacturer. In the case no fakes are available, it has been proposed to collect similar drugs of various producers with identical active pharmaceutical ingredient (API) and similar composition of excipients [21]. Each producer is presented by a set of batches ranging from five to ten. Each batch consists of 10 tablets. In total, we analyzed 220 tablets. The summary of subsets is presented in Table 2. Different batches comprise the training and validation subsets. Further, we employ these training and validation sets to construct various models. In case a dataset is not involved in modeling, it is used as a “new sample” for prediction.

Table 2
Data description

Name	Marker	Number of training objects	Number of validation objects	Tablet mass, (mg)
A3	▲	50	20	200
A4	■	30	20	180
A7	●	80	20	200

Spectra are acquired in the interval 4000–12500 cm^{-1} with a resolution of 8 cm^{-1} using the FT-NIR spectrometer (MPA by Bruker Optics) equipped with a handheld fiber-optic probe. Measurements are carried out in a diffuse reflectance mode through PVC blisters. To control reproducibility, we use triplicate readings which are averaged for data analysis. In order to remove artifacts caused by the PVC blister and the fiber probe [27] all spectra are preprocessed by the second order Savitzky-Golay differentiation with a 21 point window and third order polynomial. All calculations were conducted using standard Excel functions and Chemometrics add-in for Excel [28].

4. Results and discussion

4.1. Data overview

PCA applied to all data sets together shows an overall disposition of the subsets (Fig. 1).

Subsets A4 and A7 are rather similar and may cause difficulties in the course of classification. Subset A3 differs materially from the other two subsets.

4.2. Discriminant analysis

Let our goal be determining the authenticity of Amlodipine tablets produced by manufacturer A4 and we choose to use discriminate analysis. Considering subsets A4 as the target, and A3 as the alternative class we apply PLS-DA with two PLS components and develop

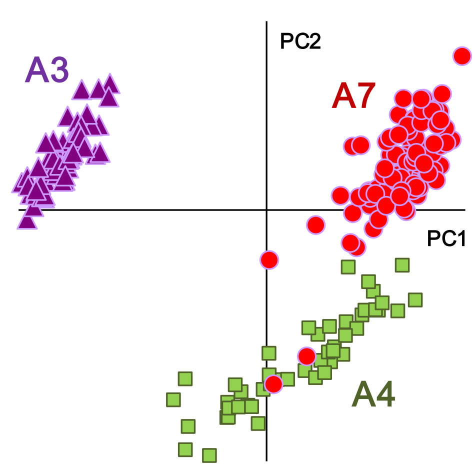

Fig. 1. Joint PCA analysis using all data. Scores plot PC1 vs. PC2.

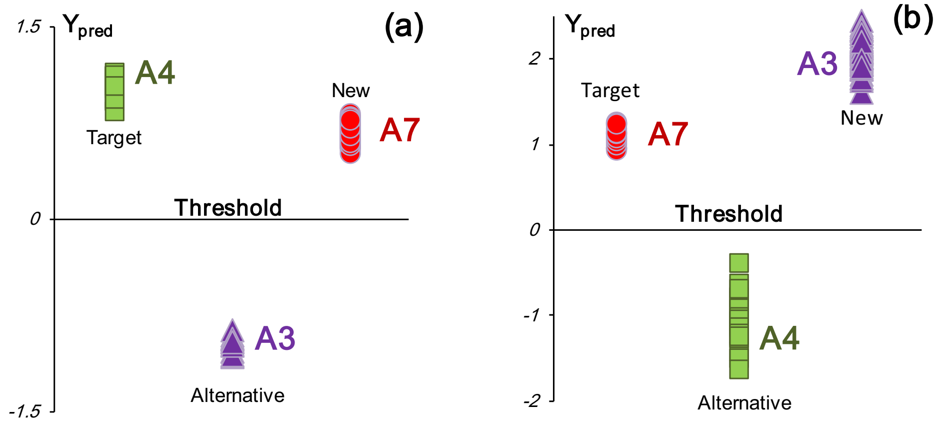

Fig. 2. Application of PLS-DA for authentication. Plot (a): A4 is the target class, A3 is an alternative class, A7 is a new class. Plot (b): A7 is the target class, A4 is an alternative class, A3 is a new class.

a decision rule, which correctly assigns all objects from the training sets to their own classes. Objects from the corresponding validation sets are also properly attributed. Method selectivity and sensitivity are equal to 100% and we could be satisfied with the obtained results. At the same time, employing the discrimination rule to objects from a new subset A7 we find out that all these aliens are attributed to the target class A4 (Fig. 2a). In this case, we see that the discrimination model provides a wrong answer to the question of authenticity.

Are sets A4 and A7 so similar indeed? Now, we consider subset A7 as a target class and subset A4 as an alternative one, and apply PLS-DA to these objects. Again, two PLS-components are enough for a 100% separation of these classes and a proper attribution of samples from the validation sets. As to the objects from the new subset A3, they are wrongly attributed to the target class A7. Thus we yield good results for the discrimination between classes A7 and A4 but unable to classify new objects (Fig. 2b).

All possible combinations of the target/alternative/new classes are shown in Table 3. It can be seen that the prediction of class membership for new objects is acceptable only in a half of cases.

Now, we develop a PLS2-DA model for all three classes simultaneously. Classes are well separated with three PLS components. The results are presented in Table 4. Diagonal elements represent sensitivity, and other elements show specificity. As specificity is a feature of binary classification, its values are calculated for each possible pairs of classes. In Table 4, as well as in similar tables thereafter, the rows stand for a target class and columns represent alternative classes. For example, element A4/A7, equal to 97%, is the specificity obtained when A4 is the target, and A7 is the alternative class.

These results could be considered as satisfactory insofar as there are no alien objects, i.e. counterfeited, or produced by some other manufacturer, analyzed in future. Obviously, all new objects will have to be attributed to one of the predefined classes and classified as authentic for this class. The results of application of other discriminant techniques may vary in details, but in general, they cannot provide a correct decision. The reason for the poor results is the absence of the exhaustive representation of the alternative class/classes.

Table 3
Results of PLS-DA analysis (2 PLS-components)

Target	Alternative	New	Sensitivity	Specificity Target/Alternative	Specificity Target/New
A3	A4	A7	100%	100%	100%
A3	A7	A4	100%	100%	98%
A4	A3	A7	100%	100%	0%
A4	A7	A3	100%	100%	100%
A7	A3	A4	100%	100%	2%
A7	A4	A3	100%	100%	0%

Table 4
Sensitivity/specificity obtained by PLS2-DA analysis with 3 PLS-components

	A3	A4	A7
A3	100%	100%	100%
A4	100%	100%	97%
A7	100%	100%	96%

4.3. One-class classifier

In this section, we illustrate an application of DD-SIMCA for authentication. Firstly, we consider A4 as the target class. 30 spectra from 3 batches are used as training samples. The PCA model with 2 PCs is sufficient for proper modeling. Acceptance areas established for the given α-values are shown in Fig. 3a. All objects located inside the area are considered as members of the target class.

The dotted line corresponds to $\alpha = 0.05$, meaning that one or two objects out of 30 can be misclassified as aliens. In practice, we have two such objects. Applying $\alpha = 0.01$ (solid line), we extend the acceptance area, and all training objects become properly classified. Model validation confirms the results (Fig. 3b). It is important that, unlike methods of discriminant analysis, we develop the acceptance area and estimate the quality of classification without using any information regarding other objects than target. Of course, now we can check class membership for objects from the alternative sets A7 and A3. Fig. 3b shows that all of them are classified as aliens, though the A7 objects are located closer to the acceptance area than A3 ones. Using the threshold constructed for $\alpha = 0.01$, and the SD and OD values calculated for the A7 objects, we can estimate the type II error, β [23], for class A7. It is $3.7 \cdot 10^{-5}$, therefore the chance of wrong acceptance is very small. First misclassifications are expected when α value is decreased to 10^{-5}. In this case, the corresponding β becomes 0.028 and some 2–3 aliens can penetrate into the acceptance area. Indeed, in practice, we observe that 2 objects from the A7 set are wrongly accepted.

Similar results, not shown here, were yielded when DD-SIMCA was applied to A3 or A7 objects as target classes. The summary of DD-SIMCA application is presented in Table 5.

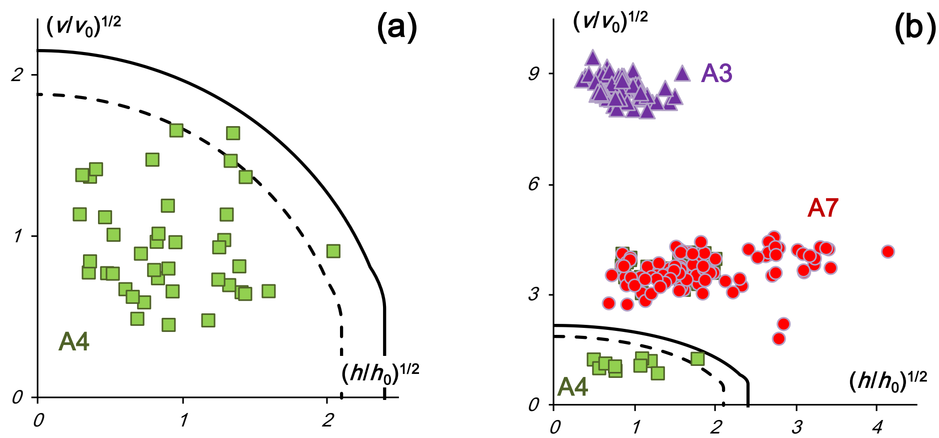

Fig. 3. Application of DD-SIMCA for objects authentication. A4 is the target class. PCA model with 2 PCs. Dotted line is the threshold for α = 0.05, solid line is the threshold for α = 0.01. Plot (a): training set, plot (b): validation set from A4, and full classes A3 and A7.

This table actually presents two sensitivity/specificity tables obtained for α = 0.05, and for α = 0.01. Each cell demonstrates two values: theoretically predicted/empirically obtained sensitivity (diagonal), or specificity (non-diagonal).

The outcomes of the one-class technique are very promising. From decreasing the risk of wrong rejection (α) we can obtain perfect outcomes 100/100% in each cell. However, keeping in mind a possibility of new high quality forgeries, we can tune α and β values depending of the specific problem.

5. Conclusions

Pattern recognition encloses a big variety of different methods and techniques. Each type of problem requires an application of relevant methods. A well constructed discrimination method will perfectly classify a new sample only if this sample is a member of one of the predefined classes [29]. However, in case the new sample does not belong to any of such classes, the discriminant analysis is unable to properly define the membership of the sample. Thus, discrimination methods are inappropriate for solving authentication problems. Class-modeling methods develop the acceptance area around the target class, and, thus, delimit the target objects from any other objects and classes. This is the reason why only one-class classifiers should be used for authentication.

There are many papers devoted to the comparison of various classification methods [30,31]. In our opinion, it is not consistent to compare methods that employ various amounts of modeling information. Methods of discriminant analysis include information regarding several classes in their algorithms. In its turn, class modeling methods do not know anything about existence of alternative classes or samples. At the same time, there is a wide-spread opinion among chemometricians that PLS-DA better separates various classes than SIMCA, as PLS-DA “may go further than the classical SIMCA classification method that works more on the reassignment of units to pre-defined classes.” [32]. This notion should be taken with care as the results are always problem dependent. For example, PLS-DA is successfully used in metabolomics [33], genomics [34] and in other ‘omics’ applications. As to the authentication problems, SIMCA shows more reliable results. Of course, the ‘best’ classification method does not exist. Every task at hand requires an application of a pertinent chemometric method best suited to answer the posed question.

Table 5
Sensitivity/specificity obtained by DD-SIMCA for two α values

	α = 0.05			α = 0.01		
	A3	A4	A7	A3	A4	A7
A3	95/96%	100/100%	100/100%	99/97%	100/100%	100/100%
A4	100/100%	95/95%	100/100%	100/100%	99/100%	100/100%
A7	100/100%	100/100%	95/99%	100/100%	100/100%	99/99%

References

[1] O. Ye Rodionova, A.L. Pomerantsev, NIR based approach to counterfeit-drug detection, Trends Anal. Chem. 29 (2010) 781–938.
[2] P.-Y. Sacré, E. Deconinck, T. De Beer, P. Courselle, R. Vancauwenberghe, P. Chiap, et al., Comparison and combination of spectroscopic techniques for the detection of counterfeit, J. Pharm. Biomed. Anal. 53 (2010) 445–453.
[3] F. Guimet, J. Ferre, R. Boque, Rapid detection of olive–pomace oil adulteration in extra virgin olive oils from the protected denomination of origin “Siurana” using excitation–emission fluorescence spectroscopy and three-way methods of analysis, Anal. Chim. Acta 544 (2005) 143–152.
[4] L. Vaclavik, T. Cajka, V. Hrbek, J. Hajslova, Ambient mass spectrometry employing direct analysis in real time (DART) ion source for olive oil quality and authenticity assessment, Anal. Chim. Acta 645 (2009) 56–63.
[5] P. Oliveri, G. Downey, Multivariate class modeling for the verification of food-authenticity claims, Trends Anal. Chem. 35 (2012) 74–86.
[6] N. Reis, A.S. Franca, L.S. Oliveira, Performance of diffuse reflectance infrared Fourier transform spectroscopy and chemometrics for detection of multiple adulterants in roasted and ground coffee, LWT – Food Sci. Technol. 53 (2013) 395–401.
[7] C. Cordella, J.S.L.T. Militao, M.-C. Clement, P. Drajnudel, D. Cabrol-Bass, Detection and quantification of honey adulteration via direct incorporation of sugar syrups or bee-feeding: preliminary study using high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) and chemometrics, Anal. Chim. Acta 531 (2005) 239–248.
[8] A. Camilo Silva, J.E. Matos Paz, L.F.B. Lira Pontes, S. Guimarães Lemos, M.J. Coelho Pontes, An electroanalytical method to detect adulteration of ethanol fuel by using multivariate analysis, Electrochim. Acta 111 (2013) 160–164.
[9] G. Mendes, P.J.S. Barbeira, Detection and quantification of adulterants in gasoline using distillation curves and multivariate methods, Fuel 112 (2013) 163–171.
[10] P. Wlasiuk, A. Martyna, G. Zadora, A likelihood ratio model for the determination of the geographical origin of olive oil, Anal. Chim. Acta 853 (2015) 187–199.
[11] A.M. Peres, P. Baptista, R. Malheiro, L.G. Dias, A. Bento, J.A. Pereira, Chemometric classification of several olive cultivars from Trás-os-Montes region (northeast of Portugal) using artificial neural networks, Chemom. Intell. Lab. Syst. 105 (2011) 65–73.
[12] J. Riedl, S. Esslinger, C. Fauhl-Hassek, Review of validation and reporting of non-targeted fingerprinting approaches for food authentication, Anal. Chim. Acta 885 (2015) 17–32.
[13] D. Tax, R. Duin, Outlier detection using classifier instability, Lect. Notes Comput. Sci. 1451 (1998) 593–601.
[14] M.P. Derde, D.L. Massart, Comparison of the performance of the class modelling techniques UNEQ, SIMCA and PRIMA, Chemom. Intell. Lab. Syst. 4 (1988) 65–93.
[15] P. Forina, S. Oliveri, Class-modeling techniques, classic and new, for old and new problems, Chemom. Intell. Lab. Syst. 93 (2008) 132–148.
[16] M.P. Derde, D.L. Massart, UNEQ: a disjoint modelling technique for pattern recognition based on normal distribution, Anal. Chim. Acta 184 (1986) 33–51.
[17] S. Wold, M. Sjostrom, SIMCA: a method for analyzing chemical data in terms of similarity and analogy, in: B.R. Kowalski (Editor), Chemometrics Theory and

Application, American Chemical Society Symposium Series 52, American Chemical Society, Washington, DC, 1977, pp. 243–282.
[18] K. Vanden Branden, M. Hubert, Robust classification in high dimensions based on the SIMCA method, Chemom. Intell. Lab. Syst. 79 (2005) 10–21.
[19] A. Pomerantsev, Acceptance areas for multivariate classification derived by projection methods, J. Chemom. 22 (2008) 601–609.
[20] D.M.J. Tax, Support vector data description, Mach. Learn. 54 (2004) 45–46.
[21] O. Ye Rodionova, K.S. Balyklova, A.V. Titova, A.L. Pomerantsev, Quantitative risk assessment in classification of drugs with identical API content, J. Pharm. Biomed. Anal. 98 (2014) 186–192.
[22] A.L. Pomerantsev, O. Ye Rodionova, Concept and role of extreme objects in PCA/SIMCA, J. Chemom. 28 (2014) 429–438.
[23] A.L. Pomerantsev, O. Ye Rodionova, On the type II error in SIMCA method, J. Chemom. 28 (2014) 518–522.
[24] L. Stahle, S. Wold, Partial least squares analysis with cross-validation for the two-class problem: a Monte Carlo study, J. Chemom. 1 (1987) 185–196.
[25] T. Naes, T. Isaksson, T. Fearn, T. Davies, Multivariate Calibration and Classification, Wiley, Christerer, 2002.
[26] G.R. Flåten, B. Grung, O.M. Kvalheim, A method for validation of reference sets in SIMCA modeling, Chemom. Intell. Lab. Syst. 72 (2004) 101–109.
[27] O. Ye Rodionova, K.S. Balyklova, A.V. Titova, A.L. Pomerantsev, The influence of fiber-probe accessories application on the results of near-infrared (NIR) measurements, Appl. Spectrosc. 67 (12) (2013) 1401–1407.
[28] A.L. Pomerantsev, Chemometrics in Excel, John Wiley & Sons, Hoboken, NJ, 2014.
[29] J.M. Amigo, H. Babamoradi, S. Elcoroaristizabal, Hyperspectral image analysis. A tutorial, Anal. Chim. Acta 896 (2015) 34–52.
[30] O. Preisner, J.A. Lopes, J.C. Menezes, Uncertainty assessment in FT-IR spectroscopy based bacteria classification models, Chemom. Intell. Lab. Syst. 94 (2008) 33–42.
[31] N.F. Pérez, J. Ferré, R. Boqué, Multi-class classification with probabilistic discriminant partial least squares (p-DPLS), Anal. Chim. Acta 664 (2010) 27–33.
[32] PLS-DA, <http://www.camo.com/resources/pls-da.html> (accessed 30.01.2016).
[33] A.-L. Boulesteix, K. Strimmer, Partial least squares: a versatile tool for the analysis of high-dimensional genomic data, Brief. Bioinform. 8 (2007) 32–44.
[34] S. Bijlsma, I. Bobeldijk, E.R. Verheij, R. Ramaker, S. Kochhar, I.A. Macdonald, et al., Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation, Anal. Chem. 78 (2006) 567–574.