Research Article

Journal of CHEMOMETRICS

Received: 23 October 2013, Revised: 3 January 2014, Accepted: 7 February 2014, Published online in Wiley Online Library: 27 February 2014

(wileyonlinelibrary.com) DOI: 10.1002/cem.2610

# On the type II error in SIMCA method

Alexey L. Pomerantsev[a,b]* and Oxana Ye. Rodionova[a]

**A novel method for theoretical calculation of the type II ($\beta$) error in soft independent modeling by class analogy is proposed. It can be used to compare tentatively predicted and empirically observed results of classification. Such an approach can better characterize model quality and thus improve its validation. Method efficiency is demonstrated on the famous Fisher Iris dataset and on a real-world example of quality control of packed raw materials in pharmaceutical industry. Copyright © 2014 John Wiley & Sons, Ltd.**

**Keywords:** SIMCA; one-class classifier; type II error; noncentral chi-squared distribution

## 1. INTRODUCTION

One-class classifiers (OCCs) are a special collection of methods within the group of pattern recognition tools. A typical OCC feature is that these methods try to distinguish objects of one particular class, also called *target class*, from all other objects and classes. The OCC model is established using a training set that contains only target objects. An overview of various OCCs can be found elsewhere [1,2].

Soft independent modeling by class analogy (SIMCA) is an OCC that was initially proposed in its simplest version [3]. Afterwards, it underwent several modifications [4] and was made more robust [5]. Nowadays, SIMCA is designed as an approach that consists of a (robust) principal component analysis model development, followed by the calculation of the orthogonal and score distances with the subsequent determination of their cut-off levels [6]. One of the unique SIMCA features is its ability to calculate the errors of misclassification theoretically. The development of various decision areas, both for a regular dataset and in the presence of outliers, is described in [7,8]. Such a modification of SIMCA is called DD-SIMCA.

An essential feature of OCC is the absence of an *alternative class* at the stage of OCC training. Subsequently, the $\beta$-error for OCC cannot be determined *a priori*. However, in case an alternative class is presented later, a particular $\beta$-error can be evaluated. So far, this was performed empirically by calculating a number of wrong acceptances in the alternative class [9–11]. In the current paper, we present a theoretical method for the type II error calculation, which can be used to compare the tentatively expected and empirically observed results. We think that such an approach better characterizes model quality.

This paper should be considered as an addition to the study presented in [8], where the SIMCA approach review and the discussion regarding the degrees of freedom (DoF) are presented. At the same time, some basic formulas are repeated in Section 2.2 for the purpose of clarity.

## 2. THEORY

### 2.1. Classification errors

None of classification models are complete without validation of the model quality, which is primarily associated with the expected errors of misclassification. The *type I error*, $\alpha$, is the rate of false rejections (false alarm), that is, the share of objects from the target class, which are misclassified as aliens. The *type II error*, $\beta$, is the rate of false acceptances (miss) [9], that is, the share of alien objects that are misclassified as the members of the target class. In general, the type II error for OCC is equal to $1-\alpha$ as a full OCC alternative includes all conceivable objects, which do not belong to the target class.

When alternative classes are presented, the $\beta$-error can be calculated for a given $\alpha$-error as shown in Figure 1. The $\alpha$-error is equal to the area under curve 1 to the right of line 4. The $\beta$-error is equal to the area under curve 2 to the left of line 4. Moving critical level 4, we can change the risks of wrong rejection ($\alpha$) and wrong acceptance ($\beta$) decisions.

### 2.2. DD-SIMCA

Let $\mathbf{X}$ be the $(I\times J)$ matrix, which originates from the initial raw matrix $\mathbf{X}_{\text{raw}}$ representing the target class data. Matrix $\mathbf{X}$ is the result of matrix $\mathbf{X}_{\text{raw}}$ preprocessing. The principal component analysis decomposition of matrix $\mathbf{X}$ is given by

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

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 (PCs). Matrix $\mathbf{\Lambda}=\mathbf{T}^t\mathbf{T}=\text{diag}(\lambda_1, \ldots, \lambda_A)$ 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 [5].

The number of PCs ($A$) in the SIMCA classification is selected using a parsimonious criterion, the minimal number at which the goal is achieved. Certainly, the goal could be diverse, so

* Correspondence to: Alexey L. Pomerantsev, Semenov Institute of Chemical Physics RAS, Kosygin str. 4, 119991, Moscow, Russia.
E-mail: forecast@chph.ras.ru

a A. L. Pomerantsev, O. Y. Rodionova
Semenov Institute of Chemical Physics RAS, Kosygin str. 4, 119991, Moscow, Russia

b A. L. Pomerantsev
Institute of Natural and Technical Systems RAS, Kurortny pr. 99/18, 354024, Sochi, Russia

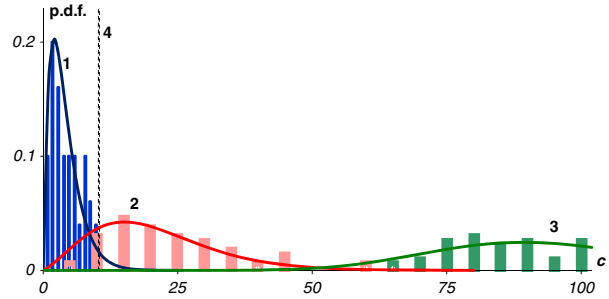


**Figure 1**. Fisher's iris dataset. Probability density distributions of statistics $c$ and $c'$ in case *Versicolor* (1) is the target class, while *Virginica* (2) and *Setosa* (3) are the alternative classes. Line 4 represents the critical cut-off value.

the number can change. For a pure one-class problem, the goal is to attribute the training and test samples properly, that is, to make the amount of extreme objects agree with the type I error level. If an alternative class is presented, the addition goal should be achieved, which is to adjust the number of aliens to the type II error. This could give a rise to the number of PCs.

Two statistics are important for interpretation and characterization of each target class object [4]. The first one is the score distance,

$$h_i = \mathbf{t}_i^{\mathrm{t}}(\mathbf{T}^{\mathrm{t}}\mathbf{T})^{-1}\mathbf{t}_i = \sum_{a=1}^{A}\frac{t_{ia}^2}{\lambda_a}, i = 1,\ldots,I \tag{2}$$

and the other one is the orthogonal distance,

$$v_i = \sum_{j=1}^{J} e_{ij}^2 = \sum_{a=A+1}^{K} t_{ia}^2, i = 1,\ldots,I \tag{3}$$

where $K \le \min(I, J)$ is the rank of matrix $\mathbf{X}$.

Let us suppose that the majority of data samples $\mathbf{X}$ follow the normal distribution $N$, and a minor part comes from the other distribution $D$, that is,

$$R_\mu = (1-\mu)N + \mu D, \quad \text{where } 0 \le \mu << 1 \tag{4}$$

In [7], it is shown that in this case, the distributions of both distances are well approximated by the scaled chi-squared distribution

$$N_h\frac{h}{h_0} \propto \chi^2(N_h) \qquad N_v\frac{v}{v_0} \propto \chi^2(N_v) \tag{5}$$

where $v_0$ and $h_0$ are the scaling factors, and $N_h$ and $N_v$ are the numbers of the DoF. We consider these parameters unknown *a priori* and suggest to estimate them using the distance samples $(v_i, h_i)$, $i = 1, \ldots, I$, obtained from the training set by a method of moments as explained in [8].

The fact that both distances follow the scaled chi-squared distributions provides a possibility to introduce a new statistics that can be called the *total distance*, $c$. It is calculated as a weighted sum of the score distance variable, $h$, and the orthogonal distance variable, $v$,

$$c = N_h\frac{h}{h_0} + N_v\frac{v}{v_0} \propto \chi^2(N_h + N_v) \tag{6}$$

It is clear that $c$ has the chi-squared distribution with $N_h + N_v$ DoF. Therefore, given the type I error $\alpha$, the acceptance area is determined by

$$c \le c_{\mathrm{crit}}(\alpha) \tag{7}$$

where

$$c_{\mathrm{crit}} = \chi^{-2}(1-\alpha, N_h + N_v) \tag{8}$$

is the $(1-\alpha)$ quantile of the chi-squared distribution with $N_v + N_h$ DoF. The MATLAB code for DD-SIMCA can be downloaded from [12].

### 2.3. The type II error

To distinguish between symbols that stand for the target and alternative classes, we will apply a stroke character for the alternative class symbols, for example, $I$ stands for the number of objects in the target class, and $I'$ stands for the number of objects in the alternative class, and so on.

Let us consider the $(I' \times J)$ matrix $\mathbf{X}'$, which is constituted of $I'$ objects from the alternative class. Matrix $\mathbf{X}'$ is assumed to be preprocessed in the same way as the target class matrix $\mathbf{X}$. The projection of $\mathbf{X}'$ on the PC space formed by the target class is determined by the equations

$$\mathbf{T}' = \mathbf{X}'\mathbf{P}\ \mathbf{E}' = \mathbf{X}' - \mathbf{T}'\mathbf{P}^{\mathrm{t}} \tag{9}$$

and the score and orthogonal distances for the alternative class objects are calculated by

$$\begin{aligned} h_i' &= \mathbf{t}_i'^{\mathrm{t}}(\mathbf{T}^{\mathrm{t}}\mathbf{T})^{-1}\mathbf{t}_i' = \sum_{a=1}^{A}\frac{t_{ia}'^2}{\lambda_a}; \quad SD_i' = \frac{h_i'}{h_0} \\ v_i' &= \sum_{j=1}^{J} e_{ij}'^2 = \sum_{a=A+1}^{K} t_{ia}'^2; \qquad OD_i' = \frac{v_i'}{v_0} \end{aligned} \tag{10}$$

The total distance $c'$ is calculated by

$$c_i' = N_h\frac{h_i'}{h_0} + N_v\frac{v_i'}{v_0}, \quad i = 1,\ldots,I' \tag{11}$$

where parameters $h_0$, $v_0$, $N_h$, and $N_v$ have been estimated using the target class data. Then, the type II error, $\beta$, is defined by

$$\beta = \Pr\{c' \leqslant c_{\mathrm{crit}}\} \tag{12}$$

To calculate this value, we will make two assumptions. The first one is that statistics $c'$ has the following distribution:

$$\frac{c'}{c_0'} \propto \chi'^2(k,s) \tag{13}$$

where $c_0'$ is a scaling factor, and $\chi'^2(k,s)$ is the noncentral chi-squared distribution [13]. This is a generalization of a well-known chi-squared distribution. The noncentral $\chi'^2(k,s)$ depends on two parameters: $k$ that specifies the number of DoF and $s$ that is the noncentrality parameter. The second assumption is that $k = N_h + N_v$, that is, we suppose that distributions in Equations 6 and 13 have the same number of DoF.

The ground for these assumptions is as follows. In case the alternative class objects are distributed in line with Equation (4) around their own center, their projections on the PC space formed by the target class also have a distribution that agrees with Equation (4).

Let us consider the sum $U = u_1^2 + \cdots + u_n^2$, where $u_1, \ldots, u_n$ are independent variables with expectations $m_1, \ldots, m_n$ and variances $v_1, \ldots, v_n$. Then,

(1) In case $\mu = 0$, and $m_1 = \ldots = m_n = 0$, and $v_1 = \ldots = v_n = 1$, $U$ has the chi-squared distribution with $n$ DoF—by definition [13];

(2) In case $\mu = 0$, and $v_1 = \cdots = v_n = 1$, $U$ has the noncentral chi-squared distribution with $n$ DoF and the noncentrality parameter $s = m_1^2 + \cdots + m_n^2$—by definition [13];
(3) In case $\mu \neq 0$, and $m_1 = \cdots = m_n = 0$, $U/A$ can be approximated by the chi-squared distribution with $k$ DoF, where $A$ and $k$ can be estimated in various ways—see [8].

Therefore, one can expect that

In general case, $U/A$ can be approximated by the noncentral chi-squared distribution with $k$ DoF and noncentrality parameter $S$. Parameters $A$, $k$, and $s$ are unknown and have to be estimated from given $u_1, \ldots, u_n$.

The direct calculations by the method of moments show that for a dataset of a reasonable size ($n < 1000$), parameters $A$ and $k$ are highly correlated, and thus only one of them can be found. Fixing $k = N_h + N_v$, we gain an evident advantage because the distribution of statistics $c'$ becomes similar to the distribution of statistics $c$ excepting a non-zero center.

Estimating two unknown parameters $c_0^{'}$ and $s$ in Equation 13 by the method of moments, we should use the right trimmed sample mean and variance. The right trimming is necessary because sometimes an alternative class consists of several subclasses that should be separated. This case is studied in Section 4. Additionally, trimming removes outliers. Fortunately, the type II error is determined by the left-side samples that are close to the acceptance area given by Equation 7.

Finally, the type II error is calculated by

$$\beta = \Pr\left\{\chi^{'2}(k,s) < \frac{c_{\text{crit}}}{c_0^{'}}\right\} \tag{14}$$

where $c_{\text{crit}}$ is defined in Equation (8). For calculation of the noncentral chi-squared distribution, an approximation presented in [14] can be used.

## 3. EXAMPLE I. IRIS DATA

To illustrate the proposed approach, a famous Fisher Iris dataset is used [15]. These data are traditionally employed in chemometrics to present a new classification approach [16,17]. The dataset depends on four variables and consists of three classes of 50 samples [18]. Each class corresponds to one of Iris species: *Iris Setosa*, *Iris Versicolour*, and *Iris Virginica*.

Let us use *Versicolor* as a target class and establish the DD-SIMCA model with two PCs. The total distances, $c$, for both the target and alternative classes are shown in Figure 1. The curves present density functions for the theoretical distributions, while the data-driven histograms are shown by columns. Classification results are presented in Table I.

The table is divided into three horizontal sections. Each section corresponds to an Iris class, considered as the target class, while the others are used as the alternative classes. The first column contains various $\alpha$-values; the numbers of tentatively expected and practically observed wrong rejection decisions are presented in columns 2 and 3. The $\beta$-values calculated for each alternative class are given in columns 4 and 7. The numbers of tentatively expected and observed wrong acceptance decisions are presented in columns 5 and 6 (the first alternative class) and columns 8 and 9 (the second alternative class). The $\beta$-values depend on $\alpha$-value in line with Equation 14 via Equation 8. Consider target class *Versicolor* (the upper part of Table I). The first alternative class *Virginica* partly overlaps with the target class (Figure 1), so the $\beta$-values are rather large. The second alternative class (*Setosa*) is located very far from the target class; therefore, the $\beta$-values are close to zero.

From Figure 1, it can be seen that class *Setosa* is well separated from other classes, and classes *Virginica* and *Versicolor* are very similar. The second and third parts of the table confirm this. The results presented in Figure 1 and Table I demonstrate a good agreement between the theory and practice.

**Table I.** Fisher's Iris dataset

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| Target class wrong rejection | | | First alternative class wrong acceptance | | | Second alternative class wrong acceptance | | |
| $\alpha$ | Expected | Observed | $\beta$ | Expected | Observed | $\beta$ | Expected | Observed |
| *Versicolor* | | | *Virginica* | | | *Setosa* | | |
| 0.1 | 5 | 7 | 0.109 | 5 | 5 | $2 \cdot 10^{-11}$ | 0 | 0 |
| 0.05 | 3 | 1 | 0.157 | 8 | 6 | $1 \cdot 10^{-10}$ | 0 | 0 |
| 0.01 | 1 | 0 | 0.285 | 14 | 14 | $4 \cdot 10^{-9}$ | 0 | 0 |
| 0.005 | 0 | 0 | 0.344 | 17 | 18 | $1 \cdot 10^{-8}$ | 0 | 0 |
| *Setosa* | | | *Versicolor* | | | *Virginica* | | |
| 0.1 | 5 | 7 | $3 \cdot 10^{-9}$ | 0 | 0 | $1 \cdot 10^{-8}$ | 0 | 0 |
| 0.05 | 3 | 3 | $6 \cdot 10^{-9}$ | 0 | 0 | $2 \cdot 10^{-8}$ | 0 | 0 |
| 0.01 | 1 | 0 | $2 \cdot 10^{-8}$ | 0 | 0 | $4 \cdot 10^{-8}$ | 0 | 0 |
| 0.005 | 0 | 0 | $4 \cdot 10^{-8}$ | 0 | 0 | $5 \cdot 10^{-8}$ | 0 | 0 |
| *Virginica* | | | *Setosa* | | | *Versicolor* | | |
| 0.1 | 5 | 7 | $6 \cdot 10^{-20}$ | 0 | 0 | 0.074 | 4 | 3 |
| 0.05 | 3 | 3 | $5 \cdot 10^{-19}$ | 0 | 0 | 0.119 | 6 | 6 |
| 0.01 | 1 | 1 | $3 \cdot 10^{-17}$ | 0 | 0 | 0.259 | 13 | 11 |
| 0.005 | 0 | 0 | $1 \cdot 10^{-16}$ | 0 | 0 | 0.329 | 16 | 17 |

SIMCA classification, three options (each class is considered as target).





  

## 4. EXAMPLE II. CONTROL OF SUBSTANCES

This real-world example has been originally presented in [19]. The task is a near infrared-based incoming inspection of a pharmaceutical substance taurine packed in closed polyethylene (PE) bags. Two hundred forty-six near infrared spectra were recorded using a handheld diffuse reflectance fiber optic probe for 82 bags with the substance. For each bag, the measurements were taken three times at different places. Additional experiments that measured the substance directly in open PE bags and a detailed data analysis [8,19] showed that the whole dataset represents a mix of two groups. The first group consists of 200 spectra of fine measurements through a single PE layer. The second group consists of 46 readings where the main substance peaks are distorted by the varying thickness of PE bags caused by the bags' folds. In this case study, we will consider the first group as a target class and the second group as an alternative class in order to calculate theoretically and afterwards validate empirically the type II errors. Here, we apply the results [8,19] of data preprocessing and the target class complexity (three PCs) yielded previously.

A thorough analysis of the alternative class samples showed that they did not comprise a uniform class but broke down into four alternative subclasses—AC1, AC2, AC3, and AC4—as presented in Table II.

This partitioning can be performed by means of a sequential right trimming of the total distances $\left(c_1', \ldots, c_l'\right)$. On the other hand, we can suggest a solid explanation of this fact. Figure 2 demonstrates typical spectra related to our example. Spectrum SUB represents the pure substance, and spectrum PE stands for polyethylene. TC is a typical spectrum of the target class samples, while AC1–AC4 are characteristic spectra of samples from the alternative subclasses AC1, AC2, AC3, and AC4, respectively.

It can be seen that two absorbance bands of PE (5780 and 5663 $cm^{-1}$) primarily influence the separation of the alternative class. At the same time, a strong substance band at 5840 $cm^{-1}$ is poorly visible in the alternative classes' spectra. Therefore, a straightforward reason of the alternative class partition is the influence of PE, that is, the target class samples were acquired though a single PE layer, the AC1 samples through two layers, and so on. The last AC4 spectra were obtained through numerous PE folds. Certainly, we do not insist on knowing the exact number of layers behind each alternative subclass; this is just a trend, not a strict rule.

Figure 3, the layout of which is similar to that of Figure 1, demonstrates the probability density distributions of the total distance statistics $c$ and $c'$ for the target class (TC), and for the alternative classes (AC1 and AC2). Other alternative subclasses are located far to the right, and they are not shown.

The separation of alternative subclasses AC1 and AC2 is clearly seen. We can also conclude that only AC1 provides a significant $\beta$ error.

Figure 4 presents the type II errors calculated by Equation 14 for alternative class AC1. It displays the numbers of wrongly accepted alternative samples in dependence on the type I error $\alpha$. Theoretically expected numbers are shown by dots with 0.90 tolerance limits, while empirically observed numbers are represented by squares. One can see that theory agrees well with experimental findings in this plot.

Thus, this real-world example confirms that the suggested theoretical approach to the type II error prediction could be used in practice.

**Table II.** Alternative subclasses

| Class | Members | $c'$ range | $c'$ mean |
|---|---|---|---|
| AC1 | 12 | 18–77 | 41 |
| AC2 | 22 | 111–364 | 204 |
| AC3 | 8 | 610–1240 | 906 |
| AC4 | 4 | 1980–4000 | 2636 |

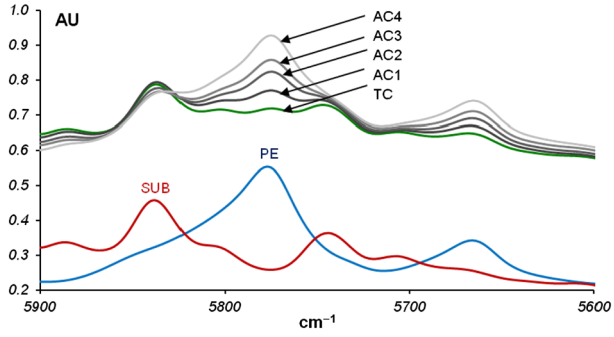


**Figure 2.** Typical spectra of packed substances. Clear substance (SUB), polyethylene (PE), the target class (TC), and the alternative classes (AC1–AC4). Spectra of SUB and PE are shifted downward for convenience.

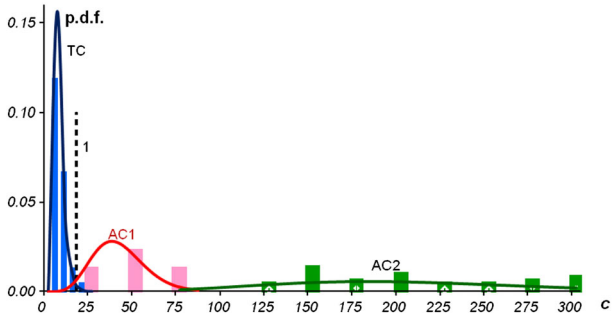


**Figure 3.** Substances' dataset. Probability density distributions of statistics $c$ and $c'$ for the target class (TC) and for the alternative classes (AC1 and AC2). Line 1 represents the critical cut-off value.

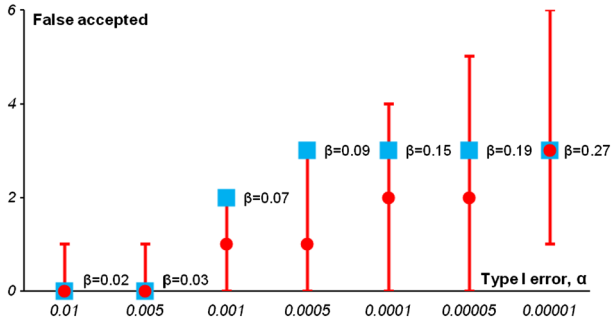


**Figure 4.** The numbers of false accepted samples from alternative class AC1 in dependence on the type I error $\alpha$. Theoretically expected numbers are shown by dots with 0.90 tolerance limits, while empirically observed numbers are represented by squares. Calculated values of the type II errors $\beta$ are indicated near the markers.

## 5. CONCLUSION

It was shown that in the presence of an alternative class, the type II error can be calculated theoretically for the SIMCA classifier. In this approach, the $\beta$-error can be obtained for a given $\alpha$-error, and, vice versa, the $\alpha$-error can be found for a given $\beta$-error. Two examples confirmed a good agreement between the theory and practice.

The ability to calculate the errors of misclassification theoretically is certainly an advantage over conventional empirical calculation. Consider, for example, the area of the quality control that employs the three main concepts. They are process analytical technology, Quality by Design, and risk management. The main point that unites all these issues is a science-based approach. One can find in literature a number of interesting papers, which successfully employed process analytical technology and Quality by Design [20]. As to risk management, the authors often confine themselves to such words as "understanding of potential risks" or "risk-based framework". At the same time, every quality control procedure must result in making a data-driven decision that relied on a quantitative risk assessment. The latter should be based on the rule that no decision rule is perfect, and inadvertent errors ($\alpha$ and $\beta$) can always be expected and accounted to develop a decision that is optimal with respect to the customer demands. It is evident that only a theoretical relationship between $\alpha$ and $\beta$ is able to give such a decision. This approach has been successfully applied to a real-world example of classification of drugs with identical active pharmaceutical ingredients content [21,22].

Similar applications could be expected for the food products authentication, in the process control, and in many other areas where a quantitative risk assessment is necessary.

## REFERENCES

1. Tax D. One-class classification: concept-learning in the absence of counter-examples. Doctoral Dissertation, University of Delft, The Netherlands, 2001.
2. Mazhelis O. One-class classifiers: a review and analysis of suitability in the context of mobile-masquerader detection. *ARIMA/SACJ* 2006; **36**: 29–48.
3. Wold S, Sjostrom M. SIMCA: a method for analyzing chemical data in terms of similarity and analogy. In Kowalski, B.R. (Ed) Chemometrics Theory and Application, American Chemical Society Symposium Series 52, Wash., D.C., American Chemical Society, 1977; 243–282.
4. Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics* 1995; **37**: 41–59.
5. Hubert M, Rousseeuw PJ, Vanden Branden K. ROBPCA: a new approach to robust principal component analysis. *Technometrics* 2005; **47**: 64–79.
6. Durante C, Bro R, Cocchi M. A classification tool for N-way array based on SIMCA methodology. *Chemom. Intell. Lab. Syst.* 2011; **106**: 73–85.
7. Pomerantsev AL. Acceptance areas for multivariate classification derived by projection methods. *J. Chemom.* 2008; **22**: 601–609.
8. Pomerantsev AL, Rodionova OYe. Concept and role of extreme objects in PCA/SIMCA. *J. Chemom.* 2013. DOI: 10.1002/cem.2506
9. Cho JH, Gemperline PJ. Pattern recognition analysis of near-infrared spectra by robust distance method. *J. Chemom.* 1995; **9**: 169–178.
10. Candolfi A, De Maesschalck R, Massart DL, Hailey PA, Harrington ACE. Identification of pharmaceutical excipients using NIR spectroscopy and SIMCA. *J. Pharm. Biomed. Anal.* 1999; **19**: 923–935.
11. Forina M, Oliveri P, Casale M, Lanteri S. Multivariate range modeling, a new technique for multivariate class modeling. The uncertainty of the estimates of sensitivity and specificity. *Anal. Chim. Acta* 2008; **622**: 85–93.
12. http://rcs.chemometrics.ru/SIMCA/DDSIMCA.zip [accessed Dec 30, 2013]
13. Muirhead R. *Aspects of Multivariate Statistical Theory* (2nd edn). Wiley, Hoboken, New Jersey, 2005.
14. Sankaran M. Approximations to the non-central chi-squared distribution. *Biometrika* 1963; **50**: 199–204.
15. Fisher RA. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 1936; **7**: 179–188.
16. Kiers HAL. Discrimination by means of components that are orthogonal in the data space. *J. Chemom.* 1997; **11**: 533–545.
17. Pérez NF, Ferré J, Boqué R. Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chemom. Intell. Lab. Syst.* 2009; **95**: 122–128.
18. http://archive.ics.uci.edu/ml/datasets/Iris [accessed Dec 30, 2013]
19. Rodionova OYe, Sokovikov YV, Pomerantsev AL. Quality control of packed raw materials in pharmaceutical industry. *Anal. Chim. Acta* 2009; **642**: 222–227.
20. Pomerantsev AL, Rodionova OYe. Process analytical technology: a critical view of the chemometricians. *J. Chemom.* 2012; **26**: 299–310
21. Rodionova OYe, Izmaylova NG, Balyklova KS, Titova AV, Pomerantsev AL. Feasibility of 3D-SIMCA application for explorative analysis and classification of drugs with identical API content. International Conference on Near Infrared Spectroscopy, France, 2013.
22. http://rcs.chemometrics.ru/present/ICNIRS13Poster.pdf [accessed Dec 30, 2013]