LOCALLY WEIGHTED NEURAL NETWORKS FOR AN ANALYSIS OF THE BIOSENSOR RESPONSE

Romas Baronas, Feliksas Ivanauskas, Romualdas Maslovskis, Marijus Radavičius and Pranas Vaitkus

This paper presents a semi-global mathematical model for an analysis of a signal of amperometric biosensors. Artificial neural networks were applied to an analysis of the biosensor response to multi-component mixtures. A large amount of the learning and test data was synthesized using computer simulation of the biosensor response. The biosensor signal was analyzed with respect to the concentration of each component of the mixture. The paradigm of locally weighted linear regression was used for retraining the neural networks. The application of locally weighted regression significantly improved the quality of the prediction of the concentrations.

Keywords: locally weighted regression, artificial neural network, modelling, biosensor
AMS Subject Classification: 62J12, 68T05, 92C45

1. INTRODUCTION

Biosensors convert the physico-chemical change of a biological sensing element, usually an enzyme, resulting from the interaction with analyte into an output concentration dependent signal [22, 24]. Biosensors are classified according to the nature of the physical transducer. Amperometric biosensors measure changes of the current on a working indicator electrode due to direct oxidation or reduction of the products of the biochemical reaction. In this case, the potential at the electrode is held constant while the current flow is measured. Amperometric biosensors are reliable, relatively cheap and highly acceptable for environment, clinical and industrial purposes [16, 26].

Traditionally, changing biological process variables have been treated separately and systematically [15]. In recent years, analytical multivariate methods have introduced refined ways to handle complex signal inputs and to interpret their relations to selected observations [1, 10]. Methods such as artificial neural networks (ANN) [8, 14] become powerful tools for experimental data analysis to improve sensitivity and selectivity of sensor systems [5, 13, 27].

This paper presents a semi-global mathematical model for the multivariate calibration of signals of amperometric biosensors. Artificial neural networks were applied to an analysis of the biosensor response to multi-component mixtures. The

output signal was analyzed with respect to the concentration of each component of the mixture. The calibration data was allocated into two different data sets: a learning set, with which the neural network as global model was trained, and a test set, which was used to improve the results of the calibration by the locally weighted linear regression method [12, 17, 25]. The application of locally weighted regression significantly improved the quality of the prediction of the concentrations.

This paper is organized as follows. Section 2 describes the data sets generated using computer simulation and briefly discusses the pecularities of the biosensor response. In Section 3, a locally weighted neural network is applied to an analysis of the biosensor response. Section 4 discusses the results of the analysis. Finally, Section 5 presents our conclusions.

2. THE GENERATION OF DATA SETS

While an artificial neural network provides a nonlinear approach that needs no *a priori* knowledge of functional dependencies, it requires training. Training is based upon cumulative experimental data.

An accurate and reliable calibration of the system as well as a proper test of the methods of chemometrics requires a lot of experimental data. A mathematical model of amperometric biosensors was used to synthesize experimental biosensor responses to mixtures [3, 5]. The model is based on the reaction-diffusion equations containing a nonlinear term related to Michaelis-Menten kinetics of the enzymatic reaction. The digital simulation was carried out using the finite difference technique [19, 23]. Assuming good enough adequacy of a mathematical model to the physical phenomena, data synthesized using computer simulation was employed instead of experimental data.

A computer simulation of the physical experiment is usually more affordable and faster than actual experimentation. A computer simulation is especially reasonable when the biosensors to be used in practice are in a stage of development. Then the development of smart biosensors and the development of effective methods of data analysis may be carried out in parallel.

Both modes of analyte analysis, bath (BA) and injection (IA), were supported [18]. In the bath analysis, a biosensor remains immersed in an analyte throughout the analysis, while in the injection analysis, the biosensor contacts the substrate for only a short time.

The current is measured as a response of a biosensor in a physical experiment. The overall biosensor response to a mixture is represented as the total sum of individual responses to each constituent substrate S_k ($k = 1, \ldots, K$). In the mathematical model each component S_k of the mixture was characterized by the individual maximal enzymatic rate $V_{\max}^{(k)}$ [5].

The biosensor response is known to be under mass transport control if the enzymatic reaction is faster than the mass transport in the enzyme layer [22, 23, 24]. The dimensionless diffusion modulus σ_k^2 essentially compares the rate of the enzymatic reaction ($V_{\max}^{(k)}/K_M$) with the diffusion ($D_S^{(k)}/d^2$) of the component S_k through the

enzyme layer

$$\sigma_k^2 = \frac{V_{\max}^{(k)} d^2}{D_S^{(k)} K_M}, \tag{1}$$

where d is the thickness of the enzyme layer, $D_S^{(k)}$ is the diffusion coefficient and K_M is the Michaelis constant, $k = 1, \ldots, K$. The biosensor response is under the diffusion control when $\sigma_k^2 \gg 1$, while enzyme kinetics controls the response when $\sigma_k^2 \ll 1$. Assuming the diffusion coefficient $D_S^{(k)}$ and the Michaelis constant K_M to be constant, the diffusion modulus σ_k^2 is a function of the maximal enzymatic rate $V_{\max}^{(k)}$ and the membrane thickness d.

Let $\vec{c} = (c_1, \ldots, c_K)$ be a vector of concentrations of K components S_1, S_2, ..., S_K of a mixture and $\vec{z} = \vec{z}(\vec{c}) = (z_1(\vec{c}), \ldots, z_N(\vec{c}))$ be a vector of the biosensor current densities at times t_1, $\ldots$, t_N calculated by the computer simulation. Thus, $\vec{z}$ defines a response of the biosensor to the mixture of K components of the concentrations $\vec{c}$. Note that $\vec{z}$ implicitly depends also on the diffusion moduli σ_k^2, $k = 1$, $\ldots$, K, and hence on the membrane thickness d. Our goal is to define a nonlinear map $\mathbf{N}$, such that $\mathbf{N}(\vec{z}) = \vec{c}$.

The data vector $\vec{z}$ has a very large dimension N and contains a lot of redundant information. Therefore the principal component analysis (PCA) [11] was applied to reduce the dimensionality of the vector of the input data. As the result of data pre-processing with PCA, the vector $\vec{x} = (x_1, \ldots, x_J)$ of J principal components of the original data vector $\vec{z}$ was obtained. The dimensionality J of the resulting vector $\vec{x}$ is usually significantly less than the dimensionality of $\vec{z}$, $J \ll N$. There exist some rules of thumb on how many dimensions to use, such as keeping all dimensions whose contribution to the total variation exceeds 80 % . The PCA resulting vector $\vec{x}$ is passed to a neural network.

3. LOCALLY WEIGHTED NEURAL NETWORKS

Let $\vec{c}_x = (c_1, \ldots, c_K)$ be a vector of concentrations of K components of a mixture and $\vec{x} = (x_1, \ldots, x_J)$ be the data vector of the biosensor signal after the pre-processing. $\vec{c}_x$ is the target concentrations for the input data $\vec{x}$.

The kth component of the quested nonlinear mapping from $\vec{x}$ to $\vec{c}$ can be expressed by an artificial neural network

$$N_k(\vec{x}) = \beta_{0k} + \sum_{s=1}^{p} \beta_{sk} \varphi(\langle \vec{\alpha}_s, \vec{x} \rangle + \alpha_{0s}), \quad k = 1, \ldots, K, \tag{2}$$

where $N_k(\vec{x})$ is the output of the kth output node expressing the concentration of the kth component S_k of the mixture, p is the number of nodes in the hidden layer, $\vec{\beta}_k = (\beta_{0k}, \beta_{1k}, \ldots, \beta_{pk})$, α_{0s}, $\vec{\alpha}_s$ are the parameters, and φ is the nonlinear activation function. The sigmoid (logistic) function was employed as the activation function φ, $\varphi(u) = 1/(1 + \exp(-u))$. The number p of nodes in the hidden layer was chosen according to a rule of thumb. Having L observations (elements) in the learning set, the degrees of freedom in the neural network should not exceed $0.1 \times L$, i.e. $(p + 1) \times K + p(J + 1) < 0.1 \times L$.

Figure 1 shows the overall architecture of the network used for the analysis of the biosensor response. The data analysis showed that one hidden layer is enough to achieve sufficiently good results in concentration estimation [5].
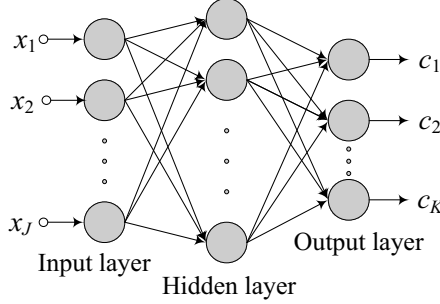

Fig. 1. Schematic diagram of three-layer feedforward artificial neural network, where $x_1, \ldots, x_J$ are values of the biosensor current and $c_1, \ldots, c_K$ are the determined concentrations of mixture components.

The neural network is usually trained by a supervised batch learning procedure that requires a set of examples for which the desired network response is known. An advanced variant of the back-propagation algorithm called Levenberg–Marquardt was used to optimize the process of learning [6, 14]. After the first phase of learning the estimated values α_{0s}^*, $\vec{\alpha}_s^*$ of α_{0s}, $\vec{\alpha}_s$, $s = 1, \ldots, p$, were kept fixed. In the second phase, we used the locally weighted regression method to re-estimate the value of $\vec{\beta}_k$ [2, 7, 20, 21].

Let Ω_1 be the learning set and Ω_2 be the test set of examples. Ω_1 is used in the first phase of learning, while data of Ω_2 is employed for retraining the neural network in the second phase.

For the fixed coefficients $\alpha_{0s} = \alpha_{0s}^*$, $\vec{\alpha}_s = \vec{\alpha}_s^*$, $s = 1, \ldots, p$, the neural network can be expressed as a linear regression $\vec{\beta}_k^T \vec{\Psi}(\vec{x})$, where

$$\vec{\Psi}(\vec{x}) = \left(1, \varphi(<\vec{\alpha}_1^*, \vec{x}> + \alpha_{01}^*), \ldots, \varphi(<\vec{\alpha}_p^*, \vec{x}> + \alpha_{0p}^*)\right)^T. \tag{3}$$

Given a test vector $\vec{x}_q \in \Omega_2$ and a weight function $w(\vec{x}, \vec{x}_q)$, the locally weighted loss functions are defined as follows:

$$\sum_{\vec{x} \in \Omega_1} w^2(\vec{x}, \vec{x}_q) \left(c_k(x) - \vec{\beta}_k^T \vec{\Psi}(\vec{x})\right)^2, \quad k = 1, \ldots, K. \tag{4}$$

Let $\Omega_1 = (\vec{x}_1, \ldots, \vec{x}_{L_1})$. Then the parameter vectors $\vec{\beta}_k^* = \vec{\beta}_k^*(\vec{x}_q)$ minimizing loss functions (4) are given by

$$\vec{\beta}_k^* = \left[(W\Psi)^T(W\Psi)\right]^{-1}(W\Psi)^T W C, \tag{5}$$

where Ψ is a $L_1 \times (p+1)$ (design) matrix with the elements $\Psi_{ls} = (\Psi_s(\vec{x}_l))$, $l = 1, \ldots, L_1$, $s = 0, \ldots, p$; W is a $L_1 \times L_1$ diagonal matrix of the local weights with the

diagonal elements $w_{ll} = (w(\vec{x}_l, \vec{x}_q))$, $l = 1, \ldots, L_1$; $C = (c_k(\vec{x}_1), \ldots, c_k(\vec{x}_{L_1}))^T$. In (5) we use a pseudo-inverse. Here for brevity we omit the subscript k.

From (2), (3) and (5) we obtain the final estimate of $N_k(\vec{x}_q)$:

$$N_k^*(\vec{x}_q) = \beta_{0k}^*(\vec{x}_q) + \sum_{s=1}^{p} \beta_{sk}^*(\vec{x}_q)\varphi(<\vec{\alpha}_s^*, \vec{x}_q> + \alpha_{0s}^*), k = 1, \ldots, K. \qquad (6)$$

Locally weighted learning systems require a measure of relevance. The major assumption that locally weighted learning rests upon is that the relevance can be measured using the distance between data points. Nearby training points are more relevant. A weighting or kernel function is used to calculate the weight for a given distance between the two points. A typical weighting function is Gaussian

$$w(\vec{x}_n, \vec{x}_q) = \exp\left\{-\|\vec{x}_n - \vec{x}_q\|^2/(2h^2)\right\}, \qquad (7)$$

where h is the bandwidth or the kernel width. It determines the range over which the generalization is performed.

The forecasting quality of concentration c_k of each component S_k, $k = 1, \ldots, K$, is estimated by the percentage of true interval predictions [5]

$$Q_k = \frac{1}{L_2} \sum_{i=1}^{L_2} Ind\left(N_k^*(\vec{x}_i) \in \Delta y\right) \times Ind\left(c_k(\vec{x}_i) = y\right) \times 100\,\%, \qquad (8)$$

where the indicator function $Ind\left(N_k^*(\vec{x}_i) \in \Delta y\right)$ equals unity when the kth output of the network (after locally weighted training for $\vec{x}_i$ as a query point) belongs to the interval $(y - \delta_{1,y}, y + \delta_{2,y})$ of the target concentration $c_k(\vec{x}_i) = y$, and zero otherwise. L_2 is the number of observations in the test set Ω_2.

When using locally weighted retraining, it is necessary to retrain the network each time a new point is evaluated. This operation is quite time-consuming. Another disadvantage is the need to keep the original training set. However, the improvement in accuracy in concentration prediction offsets these disadvantages, especially in such applications as the detection of toxins in waste water [9].

4. RESULTS AND DISCUSSION

The modelling biosensors were calibrated for mixtures of four $(K = 4)$ components. Each component of eight $(M = 8)$ different concentrations was employed in the calibration to have the biosensor response to a wide range of substrate concentrations [3].

The total set of full factorial of $M^K = 8^4 = 4096$ responses was split randomly into Ω_1 (learning) and Ω_2 (test) sets having approximately the same number of responses. Two thousand response curves were chosen independently as the Ω_2 set. The remaining 2096 samples were accepted as the Ω_1 set, $L = 2096$.

The following M concentrations for each of the K components $S_1, \ldots, S_K$ of the mixture were used:

$$c_k \in \left\{s_{k,m} : s_{k,m} = \gamma_m \text{ nmol/cm}^3, \ m = 1, \ldots, M\right\}, \ k = 1, \ldots, K,$$

$$K \quad = 4, \; M = 8;$$
$$\gamma_1 \quad = 1, \gamma_2 = 2, \gamma_3 = 4, \gamma_4 = 8, \gamma_5 = 12, \gamma_6 = 16, \gamma_7 = 32, \gamma_8 = 64.$$

The biosensor response depends upon the concentration c_k of the component S_k ($k = 1, \ldots, K$), the mode of analyte analysis (BA or IA), and the membrane thickness d. The response of the first biosensor, having a membrane of thickness $d = 0.2\,\text{mm}$, was under dual control: by diffusion (for three components S_1, S_2, S_3) and by enzyme kinetics (for component S_4). Another thickness $d = 0.4\,\text{mm}$ was chosen so, that the response was controlled by the diffusion for all four components.

The biosensor responses were simulated until a steady state was reached. During the computer simulation, values $\vec{z} = (z_1, \ldots, z_N)$ of the density of the biosensor current were stored every second of the biosensor operation, i. e. the current density at time $t_i = i\,\text{s}$ was accepted as the z_i, $i = 1, \ldots, N$. The moment of the occurrence of the steady state depends on the mode of analysis and the thickness of the enzyme membrane [4, 22, 24].

The application of PCA to the Ω_1 sets resulted in 6 principal components in all three cases of the biosensor operation: BA at membrane thickness $d = 0.2\,\text{mm}$ ($N = 300$), BA at $d = 0.4\,\text{mm}$ ($N = 300$), and IA at $d = 0.2\,\text{mm}$ ($N = 150$). Due to the PCA the neural networks having $J = 6$ nodes in the input layer were employed. Since the mixtures to be analysed consist of four components, the networks have $K = 4$ nodes in the output layer. A single hidden layer feedforward neural network with sigmoid activation in the hidden layer has universal approximation capabilities [8, 14]. Twelve nodes in a single hidden layer were used in the case of BA when membrane thickness d is $0.2\,\text{mm}$. In two other cases, BA at $d = 0.4\,\text{mm}$ and IA at $d = 0.2\,\text{mm}$, eight nodes in a single hidden layer were enough.

The forecasting quality Q_k of concentration $y = s_{k,m}$ of the component S_k of a mixture was estimated using quality measure (8) with the tolerance intervals Δy presented in Table 1.

Table 1. The accuracy intervals Δy for prediction
of the concentration y using forecasting quality (8).

y	Δy	y	Δy
1	[0, 1.5)	12	[11, 13)
2	[1.5, 2.9]	16	[15, 17)
4	[3.1, 5)	32	[31, 33)
8	[7, 9)	64	[63, 65)

In the case of injection analysis, the prediction quality achieves a suitably high classification accuracy of over 99 % .

In the case of bath analysis, the quality of prediction depends on whether the biosensor response is under diffusion or enzyme kinetics control. When biosensor response was under enzyme kinetics control, we achieved a prediction quality of 76.75 % . In all other cases the forecasting quality was over 99 % . Let us remind the reader that the biosensor response was under diffusion control when predicting

only the component S_4 with the biosensor having an enzyme membrane thickness of 0.2 mm.

The prediction quality of 76.75 % is fairly low. However, this value is significantly higher than the corresponding value of the forecasting quality (57.4 %) achieved using a standard ANN [5].

Table 2. The forecasting quality Q_k of the concentration prediction in bath analysis at the test set and membrane thickness of 0.2 mm.

Component, k	Standard ANN	Locally weighted ANN
1	100	100
2	100	100
3	96	99.55
4	57.4	76.75

Figure 2 shows values of the forecasting accuracy Q_k versus values of the bandwidth h for all components S_1, ..., S_4 of the mixtures. One can see notable non-monotonous behaviour of the function Q_k when analysing components S_3 and S_4. The quality 76.75 % of the prediction of the component S_4 has been achieved at h = 0.25. The best prediction quality (99.55 %) of S_3 has been achieved at a slightly different value (0.4) of h. Consequently, the maximizing value of the bandwidth h depends on the activity of the enzyme to the concrete component.
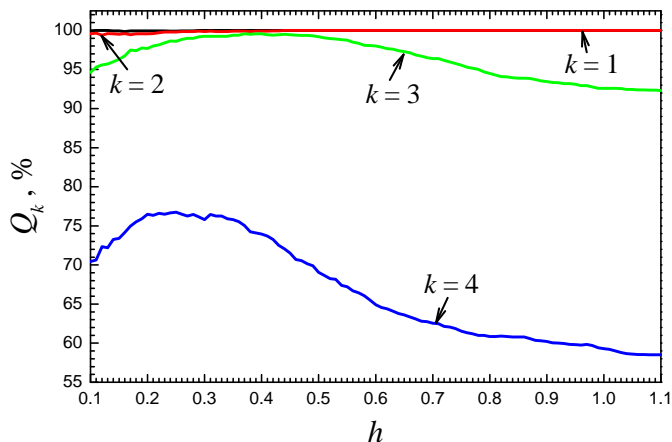

Fig. 2. The forecasting accuracy Q_k versus the bandwidth h for the mixture component S_k, $k = 1, \ldots, K$.

5. CONCLUSIONS

Artificial neural networks can be successfully used to discriminate components of mixtures and to estimate the concentration of each component from the biosensor response data.

In the case of bath analysis, the prediction quality depends on whether the biosensor response is under diffusion or enzyme kinetics control. The concentration of component S_k is predicted more accurately when the biosensor response is under diffusion control, i. e. when the diffusion modulus σ_k^2 is greater than one. Because of this, the enzyme membrane thickness as one of the factors determining the diffusion modulus is of crucial importance for the detection limit of the biosensor.

In the case of injection analysis, the prediction quality is much less sensitive to the ratio of the enzyme reaction rate to the rate of the mass transport through the enzyme layer. An application of the injection analysis instead of the bath one can increase prediction quality.

Prediction quality can also be significantly increased by the application of a locally weighted linear regression for retraining the neural networks (Table 2).

Work is now in progress to increase the quality of phenol detection in waste water [9].

ACKNOWLEDGEMENT

REFERENCES

[1] T. Artursson, T. Eklöv, I. Lundström, P. Mårtensson, M. Sjöström, and M. Holmberg: Drift correction for gas sensors using multivariate methods. J. Chemometrics *14* (2000), 711–723.

[2] C. G. Atkeson, A. W. Moore, and S. Schaal: Locally weighted learning. Artificial Intelligence Rev. *11* (1997), 11–73.

[3] R. Baronas, J. Christensen, F. Ivanauskas, and J. Kulys: Computer simulation of amperometric biosensor response to mixtures of compounds. Nonlinear Anal. Model. Control *7* (2002), 3–14.

[4] R. Baronas, F. Ivanauskas, and J. Kulys: The influence of the enzyme membrane thickness on the response of amperometric biosensors. Sensors *3* (2003), 248–262.

[5] R. Baronas, F. Ivanauskas, R. Maslovskis, and P. Vaitkus: An analysis of mixtures using amperometric biosensors and artificial neural networks. J. Math. Chem. *36* (2004), 281–297.

[6] L. W. Chan and C. C. Szeto: Training recurrent network with block-diagonal approximated Levenberg–Marquardt algorithm. In: Proc. IEEE Internat. Joint Conference on Neural Networks, IJCNN '99, pp. 1521–1526, 1999.

[7] L. Devroye, L. Gyorfi, and G. Lugosi: A Probabilistic Theory of Pattern Recognition. Springer–Verlag, New York 1996.

[8] S. Haykin: Neural Networks: A Comprehensive Foundation. Second edition. Prentice Hall, New York 1999.

[9] INTELLISENS: Intelligent Signal Processing of Biosensor Arrays Using Pattern Recognition for Characterisation of Wastewater: Aiming Towards Alarm Systems. EC RTD project. 2000 – 2003.

[10] P. Malkavaara, R. Alén, and E. Kolehmainen: Chemometrics: an important tool for the modern chemist, an example from wood-processing chemistry. J. Chem. Inf. Comput. Sci. *40* (2000), 438–441.

[11] H. Martens and T. Næs: Multivariate Calibration. Wiley, Chichester 1989.

[12] A. W. Moore, J. G. Schneider, and K. Deng: Efficient Locally Weighted Polynomial Regression Predictions. In: Proc. Fourteenth International Conference on Machine Learning, pp. 236–244, 1997.

[13] T. Nakamoto and H. Hiramatsu: Study of odor recorder for dynamical change of odor using QCM sensors and neural network. Sens. Actuators B *85* (2002), 98–105.

[14] D. Patterson: Artificial Neural Networks, Theory and Applications. Prentice Hall, Upper Saddle River 1996.

[15] C. R. Rao: Linear Statistical Inference and its Application. Wiley, New York 1973.

[16] K. R. Rogers: Biosensors for environmental applications. Biosens. Biolectron. *10* (1995), 533–541.

[17] D. Ruppert and M. P. Wand: Multivariate locally weighted least squares regression. Ann. Statist. *22* (1994), 1346–1370.

[18] J. Ruzicka and E. H. Hansen: Flow Injection Analysis. Wiley, New York 1988.

[19] A. A. Samarskii: The Theory of Difference Schemes. Marcel Dekker, New York – Basel 2001.

[20] S. Schaal and C. G. Atkeson: Assessing the quality of learned local models, In: Advances in Neural Information Processing Systems 6 (J. Cowan, G. Tesauro, J. Alspector, eds.), Morgan Kaufmann 1994, pp. 160–167.

[21] S. Schaal and C. G. Atkeson: Constructive incremental learning from only local information. Neural Comput. *10* (1998), 2047–2084.

[22] F. Scheller and F. Schubert: Biosensors, Vol. 7. Elsevier, Amsterdam 1992.

[23] T. Schulmeister: Mathematical modelling of the dynamics of amperometric enzyme electrodes. Selective Electrode Rev. *12* (1990), 260–303.

[24] A. P. F. Turner, I. Karube, and G. S. Wilson: Biosensors: Fundamentals and Applications. Oxford University Press, Oxford 1987.

[25] Z. Wang, T. Isaksson, and B. R. Kowalski: New approach for distance measurement in locally weighted regression. Anal. Chem. *66* (1994), 249–260.

[26] U. Wollenberger, F. Lisdat, and F. W. Scheller: Frontiers in Biosensorics 2: Practical Applications. Birkhauser Verlag, Basel 1997.

[27] C. Ziegler, W. Göpel, H. Hämmerle, H. Hatt, G. Jung, L. Laxhuber, H.-L. Schmidt, S. Schütz, F. Vögtle, and A. Zell: Bioelectronic noses: A status report. Part II. Biosens. Bioelectron. *13* (1998), 539–571.

Romas Baronas, Department of Software Engineering, Faculty of Mathematics and Informatics, Vilnius University, Naugarduko 24, LT-03225 Vilnius. Lithuania.
e-mail: romas.baronas@maf.vu.lt

Feliksas Ivanauskas, Institute of Mathematics and Informatics, Akademijos 4, LT-08663 Vilnius, Lithuania; Faculty of Mathematics and Informatics, Vilnius University, Naugarduko 24, LT-03225 Vilnius, Lithuania.
e-mail: feliksas.ivanauskas@maf.vu.lt

Romualdas Maslovskis and Pranas Vaitkus, Department of Mathematical Statistics, Faculty of Mathematics and Informatics, Vilnius University, Naugarduko 24, LT-03225 Vilnius. Lithuania.
e-mails: romualdasmaslovskis@yahoo.com, vaitkuspranas@yahoo.com

Marijus Radavičius, Applied Statistics Department, Institute of Mathematics and Informatics, Akademijos 4, LT-08663 Vilnius. Lithuania.
e-mail: marijausr@gmail.com