Computer Communication & Collaboration (Vol. 5, Issue, 2017)
ISSN 2292-1028(Print) 2292-1036(Online)
Submitted on May 26, 2017
DOIC: 2292-1036-2017-03-002-89

Quantitative "Reading" of the Score Matrix Components in the PCA: New General Tool for Electrochemical Data Analysis

***Raoul R. Nigmatullin*[1)]** (Correspondence author)

[1)] Radioelectronic and Informative-Measurements Technics Department,

Kazan National Research Technical University (KNRTU-KAI),

K. Marx str. 10., 420111, Kazan, Tatarstan, Russian Federation

E-mail: renigmat@gmail.com

Herman K. Budnikov [2)]

[2)] Institute of Chemistry, Kazan Federal University

Tatarstan, Russian Federation

Artem V. Sidelnikov[3)], ***Elza I. Maksyutova*** [3)]

[3)] Chemistry Department, Bashkir State University,

Ufa, Russian Federation

e-mail:artsid2000@gmail.com

Abstract: In this paper, we consider the new quantitative features that can be attached to the conventional PCA. It is based on some peculiarities of the Fourier spectra obtained from the vectors forming the score matrix. These peculiarities allow parameterizing the score vectors and fitting them accurately by the segment of the Fourier series. This key moment can be generalized and used for any vector having the leading frequencies in the Fourier spectrum. As a criterion of this specific behavior the neighboring score vectors should form the Lissajous figures that are very well known in the theory of mechanical vibrations. In addition, we found a new source of information related to analysis of the eigenvalues of the diagonal matrix that are appeared in the singular values decomposition. As an example we considered the region of the remnant currents (that was considered before as spurious) of the electrochemical cell background and found the definite differences between two regimes of the glassy carbon electrode usage. We do hope that these new peculiarities will find a proper place in chemometrics and practical electrochemistry for detection of traces and other micro-components on the background of the given macro-components presenting in some electrochemical cell. In addition, this observation will be useful for the PCA at whole; the found parameterization will help "to read" quantitatively the score matrix vectors and, thereby, makes this analysis more informative and efficient.

Keywords: PCA, Electrochemical Data Analysis, Electronic Tongues

Abbreviation list:

AFR – the amplitude-frequency response

CFFT –complex fast Fourier transform

DEL – double electric layer

GCE – the glassy carbon electrode

LLSM – the linear least square method

PCA – principal component analysis

SVD – singular values decomposition

VAGs – voltammogram(s)

1. Introduction and formulation of the problem

If we print on a keyboard in some searching system (Google, Yandex and etc.) the abbreviation "PCA" then we obtain more than 330 million references! The principal component analysis (PCA) was invented by K. Pearson in 1901 year and received the successful continuation in many applied sciences. As it is known the PCA defined as is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (or sometimes, principal modes of variation). We are not going to explain this transformation and possible modifications because they are well known among researches that are trying to improve this method and apply it for solution of many practical problems. However, the attentive analysis of the PCA shows that in the result of its application we obtain a set of *qualitative* pictures of independent correlations between possible statistical variables. On the intuitive level, the problem can be formulated as follows: Is it possible to transform a part of the probabilistic information to the deterministic one and express the components of the score matrices *analytically* in the form of some analytical/fitting curves? We found a *positive* answer on this posed question in important application region of the PCA as electrochemistry. In order to understand better the original part of the paper we should outline a situation that exists with application of the PCA in electrochemistry.

Starting from the 90th years of the XX-th century in electro-analytical chemistry, the effective methods of analysis were suggested. They were based on the usage of the multisensor systems (defined as a battery of sensors) of the "electron tongue" type allowing to detect the presence of micro- and macro-components in the research objects and to solve problems of evaluation the nonparametric/qualitative properties of the complex liquids as the taste, tartness and etc. These sensor systems are important when it is necessary (without the detailed chemical analysis) to detect quickly the unknown liquid. The literature review in this region shows the wide application of the designed voltammetric multisensor systems for the quality control of foods [1–3], drags [4, 5], sewage and polluted waters [6, 7], etc.

Voltammetric systems use the phenomenon of the electric current passing under the applied voltage between the working and auxiliary electrodes, whereas the potential is measured between the working and reference electrodes [8]. As a massive of analytic signals the voltammetric curve is used i.e. the dependence of the current passing through the investigated solution against the applied

potential. Voltammetric cell having one electrode only one can consider as a multisensor detector if in the process of voltammograms registration some chemical compounds are oxidized/restored. Besides, the voltammetry with the usage of various forms of the polarized voltage (cyclic, square-wave, impulse and etc.) allows to vary the sensitivity and selectivity of the used sensors and extract the information about the structure of the double-electric layer (DEL), ion diffusion coefficients and etc. However, even having the good conditions for the extraction of the well-distinct signals (the oxidation/restoration of the molecular peaks) the registered voltammograms (VAGs) always contain the background current component, which affects the value of the obtained analytical signal and frequently distorts its form. This effect takes place at analysis of the substances with small concentrations (micro-components on the macro-components background) when the dispersion of the capacity current (arising because of the DEL charging) becomes comparable with faraday current (arising because of oxidation/reduction micro-component). Besides this obstacle at working with solid electrodes, one takes place the time-drift of the registered signal at constant functioning of a sensor system. It appears as the displacement of the potential peaks or decreasing of the sensor sensitivity because of inevitable electrode surface changing under the influence of electric current (oxidation of the electrode material, adsorption of the electrochemical reaction products, the roughness surface changing etc.) [9]. One of the technical solution of this problem is [8, 10] the regeneration of the sensor layer. Nevertheless, in the conditions of the constant functioning of the sensors massive it is necessary to apply the mathematical transformation methods for separation of useful signals from random components (known as a noise) without interference to the used sensor block. The exiting mathematical methods used in electrochemistry cannot solve the analytical problem of detection/determination of the hidden micro-component when the height of its oxidation/regeneration peak becomes close to the standard deviation of the remnant (background) current or the "annihilation" of the micro-component peak that takes place in the presence of the macro-component. Therefore, the actual problem is the quantitative description of voltammetric micro-components behavior when their intensity becomes comparable with the level of noise. One of the effective methods used in chemometrics and electrochemistry is the principal component analysis (PCA). The PCA is widely used practically in every experiment associated with construction of "electronic tongues".

As it has been mentioned above, the measured VAGs contain always the remnant currents alongside with the conventional faraday currents. However, in the most papers the researchers use only the faraday region of potentials, where the oxidation/restoration peaks are clearly seen and can be interpreted. The remnant current region is always considered as undesirable (interfering) and all improvements of analytical methods were related presumably with elimination of these "spurious" data [11].

Our early investigations related to the usage of the PCA [12–14] showed that in the region of the remnant currents the chemical information presents *always* especially near the liquids contacting with electrodes. However, the PCA in its classical application does not allow describing the measured VAGs quantitatively in the intermediate region and establishing the factors that can modify the form of the registered VAGs and extract the desired signals. Here one can remind another inevitable problem. The numerous experiments show that the reiterative electrochemical actions on the registered sensor form a specific electrode "memory". This effect is confirmed in papers [15–19]. The influence of the measurements to each other (it can be considered as partial correlations between measurements) leads to data that do not obey to the normal distribution and not form on the score plots the conventional clusters as rings or chairs used for some qualitative interpretation. In the result

of application of the usual PCA algorithms, it is *impossible* to explain the observed phenomena for *all* regions of currents. From the mathematical point of view, one can formulate the basic defect of the PCA technics viz.: the absence of the general and *quantitative* tool for description of the dependencies of the score matrix components.

The remarks saying above one can refer to the PCA overall. The PCA uses presumably only three initial and leading components and, therefore, all analysis is used for detection of possible clusters and correlations expressed in the form of the segments of the straight lines or uniformly filled ellipses or circles that are interpreted as the correlated clusters. Actually, two important problems are solved in this paper. We formulate them in the form of two questions, viz :

1. Is it possible to read the columns (vectors) of the score matrix *quantitatively* for more detailed detection of the different qualitative factors that always present in any electrochemical experiment?

2. Is it possible to read *quantitatively* the distribution of the eigenvalues of the score matrix for more accurate analysis of the different external/internal factors?

2. Mathematical section: Basic formulas and the proposed algorithm

In this section, we remind to a potential reader the basic expressions of the PCA for better understanding of the original part. We consider the rectangle matrix $M_{i,j}$ (i=1,2,…,N_r, j=1,2,…,N_c), where indexes i, j define all rows (N_r) and columns ($N_c \le N_r$), respectively. In accordance with the general algorithm accepted in the PCA the initial matrix should be presented in the form

$$M_{i,j} = T \cdot P^T + E \equiv \sum_{a=1}^{A} T_i^{(a)} P_j^{(a)} + E_{i,j}, \qquad A << N_r, N_c. \qquad (1)$$

Here T determines the score matrix (having components $T_j^{(a)}$) and P^{T} is the load transposed matrix with components $P_j^{(a)}$. The remnant matrix E (with components $E_{i,j}$) determines as the error matrix. The total number of the common components A coinciding for the both matrices should be much less in comparison with the numbers N_r, N_c ($A << N_{c,r}$). This representation (it is valid at condition $N_r \ge N_c$) can be realized easily with the help of the singular values decomposition (SVD) procedure. If we present the initial matrix M in the form

$$M = USP^T,\ S(\sigma_1 \ge \sigma_2 \ge ...\sigma_{N_r}), \qquad (2)$$

and compare it with expression (1) then we obtain that $US \equiv T$ and P is associated with the load matrix. From this definition, it follows that

$$TT^T = USS^*U^T = S^2, \qquad (3)$$

and this expression can be used as an *additional* test for numerical verification of the eigenvalues of the diagonal matrix S in (2). The PCA as the method is considered as the *reduction* method because it allows using small number of components a in decomposition (1) and, therefore, substitute approximately the initial matrix M and evaluate the arising error E. Usually, the researchers use the leading components of the score matrix T (a=1, 2, 3) in order to detect possible correlations between

the components/vectors $T^{(a)}$ (for simplicity we omit the index i) expressed in the form of a segment of the straight line or compact circle. Other components of $T^{(a)}$ are associated with error matrix E and not considered.

In this paper we show *how* to consider a large set of $T^{(a)}$ (a >3) and prove that the further components being plotted with respect to each other and looking initially as random distribution of the data points have actually the certain *deterministic* nature. Therefore, one can restore more information from analysis of these plots if they will be presented by another way. Apart from this analytical presentation of the score matrix T components, we want to show how to fit the distribution of the eigenvalues of the S matrix. These new modifications of the conventional PCA solve actually two problems formulated above and allow increasing the limits of the application of the PCA. Below, for illustration of new presentation associated with analysis of VAGs it is convenient to use the *dimensionless* potential (we normalize simply the applied potential U to 1V, i.e. $u = U$(V)/1V and use the uniform scale

$$u_i = \min(u) + \frac{i}{N_r}\big(\max(u) - \min(u)\big), \quad i = 0,1,..,N_r. \tag{4}$$

Then we use the normalized matrix G determined below by (5) (associated with the measurements of the background by means of the cycling potential U) for glassy carbon electrode (GCE) and covering 4 regions: (1) the cathode region that is associated with reduction of water; (2) the region of the remnant currents; (3) the faraday currents region associated with oxidation of different organic compounds (in the case of their presence in solution); (4) the region of faraday currents associated also with oxidation of water. The basic problem is to find discrepancies between these regions and express them *quantitatively* with respect to each other. Other details related to these experiments are given in the next section. We should notice that the number of data points associated with applied U equaled N_r = 425, whereas the number of columns is measured in numbers of the successive experiments equaled N_c = 100. Duration of each experiment acting with electrode is equaled to 25 sec. It includes in itself the potential cycling procedure (15 sec) and the measuring VAG (5 sec). See Fig.1 for details. The plot of these experiments for the normalized matrix

$$G_s = \frac{M_s - mean(M_s)}{Stdev\big(M_s - mean(M_s)\big)},\ s = 1,2,...,N_r \tag{5}$$

and s=1, 10, 20, 50, 100 and mean measurement averaged over all experiments is given by Figs.2(a,b). We should stress here that in Fig. 2(a) we demonstrate the derivatives dJ/dU as the most *sensitive* components in comparison with integral curves as $J(U)$ depicted in Fig.2(b). However, the integral plot for the cumulative currents has also the oscillating character. Differentiation of these curves with respect to U allows eliminating the undesirable envelopes and expressing the oscillating effect more clearly. Then after application of the SVD procedure we present the second component of the score matrix $T^{(2)}$ as the function of the vector $T^{(1)}$. This plot is shown in Fig.3. Taking the Fourier-transform from the vectors $T^{(1,2,3,4)}$ we obtain the key picture that is shown in Fig.4. One can obtain the accurate fit for all 4 initial score vectors. This observation follows from the fact that the Fourier transform of these vectors contains a couple of the leading frequencies that are shown in Fig. 6. From analysis of this figure one can notice two leading frequencies with predominant amplitudes satisfying approximately to condition $\omega_0^{(1)} + \omega_1^{(1)} \cong 2\pi$ and other frequencies with unnoticeable amplitudes that can be omitted in the first stage.

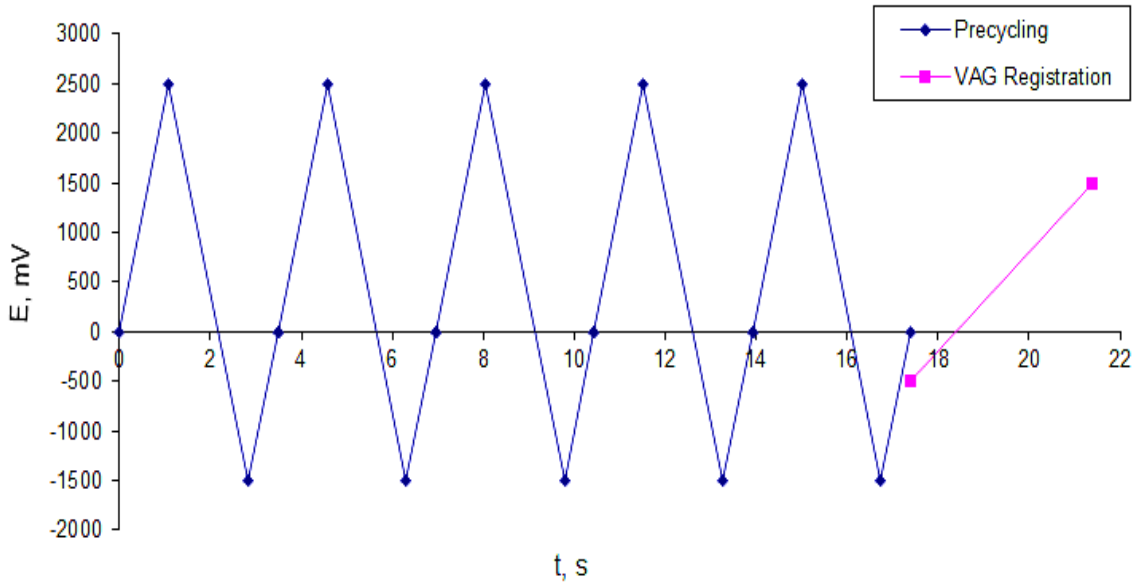

Figure 1. Here we demonstrate the precycling regime for the potential U evolution (5 cycles for each VAG registration) and then the registration of the desired VAG (shown in the end of each cycle). One measured cycle covers 25 sec. The period of the VAG registration occupies 5 sec. Then this similar procedure is continued for receiving the desired number 100-200 registered VAGs.

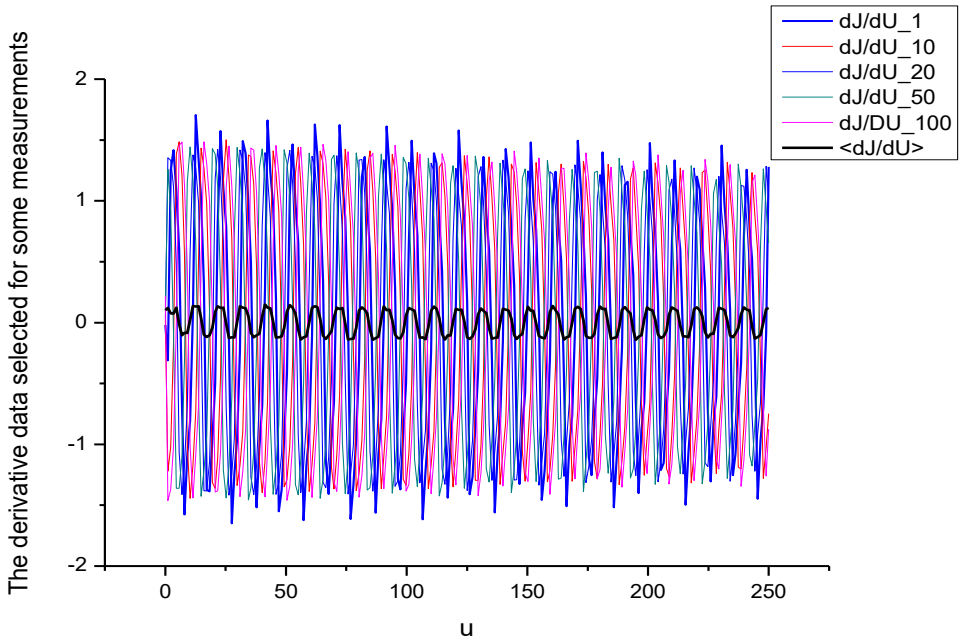

Figure 2(a). Here we show the initial data for selected measurements (s = 1,10,20,50,100) together with the mean measurement <dJ/dU> averaged over all experiments S=100 for the cycling regime (the region of the remnant currents). This oscillating plot is obtained by differentiation of the dependence $J(U)$ depicted for the same measurements in Fig.2(b) below.

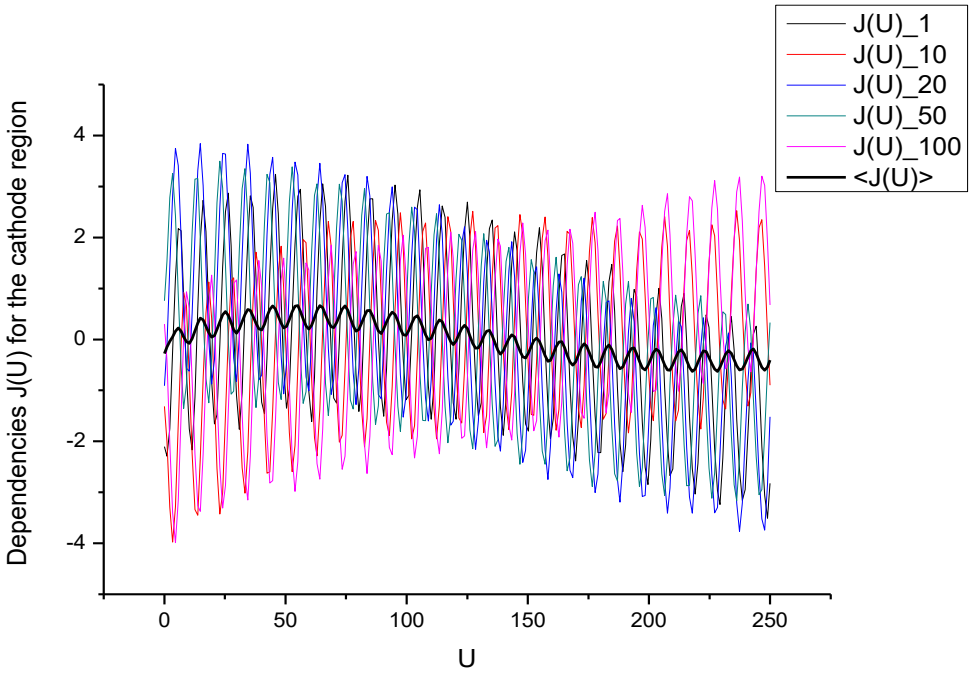

Figure 2(b). The dependencies $J(U)$ have also the oscillating character. We demonstrate here the plots obtained for the same selected measurements (s = 1,10,20,50,100) together with the mean measurement $<J(U)>$ averaged over all experiments for the cycling regime. The scheme of this regime is shown in Fig.1.

Mathematically, the dependence of the vector $T^{(1)}(u)$ can be expressed approximately as

$$T^{(1)}(u) = A_0^{(1)} + Ac_m^{(1)} \cos\left(\omega_0^{(1)} u - \varphi_0^{(1)}\right), \qquad (6)$$

where the parameters $A_0^{(1)}, Ac_m^{(1)}, \omega_0^{(1)}, \varphi_0^{(1)}$ determine the initial vibration point, the initial amplitude, frequency and phase, correspondingly. The contribution of the second frequency $\omega_1^{(1)} = 2\pi - \omega_0^{(1)}$ keeps expression (6) invariant and, therefore, initially is not considered. If we apply the same transformation with respect to the vector $T^{(2)}(u)$ we obtain the Fourier transform similar to Fig.6 with the same leading frequencies $\omega_1^{(1)} = 2\pi - \omega_0^{(1)}$ but other parameters can be different. So, one can write approximately

$$\begin{aligned} &T^{(2)}(u) = A_0^{(2)} + Ac_m^{(2)} \cos\left(\omega_0^{(2)} u - \varphi_0^{(2)}\right), \\ &\omega_0^{(1)} = \omega_0^{(2)}, \ \ \omega_0^{(2)} + \omega_1^{(2)} \cong 2\pi. \end{aligned} \qquad (7)$$

Therefore, the plot $T^{(2)} = F(T^{(1)})$ depicted in Fig. 3 expresses the different Lissajous figures that are accepted in the conventional theory of mechanical vibrations.

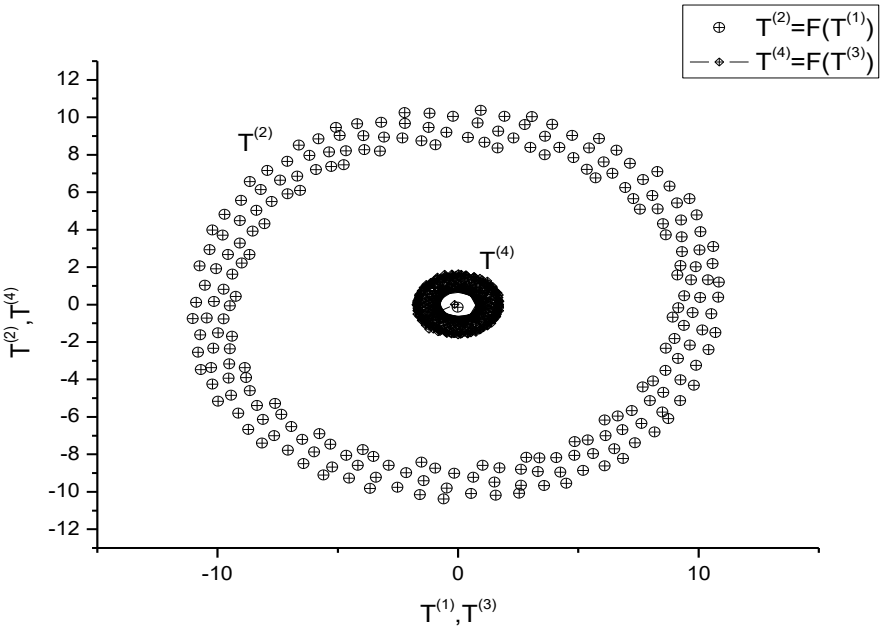

Figure 3. After realization of the SVD procedure we can plot $T^{(2,4)} = F(T^{(1,3)})$ corresponding to the conventional analysis. In order to explain these plots we apply the parameterization of these functions in accordance with expression (8). The fit of these functions is shown in the next figure.

In particular, for the case $\varphi_0^{(1)} = \varphi_0^{(2)}$ we obtain the straight line; for $\varphi_0^{(2)} = \varphi_0^{(1)} + \frac{\pi}{2}$ we obtain the classical ellipse; for $\varphi_0^{(2)} = \varphi_0^{(1)} + \alpha$ we obtain the ellipse rotated counterclockwise on the angle α. For $\omega_0^{(1)} \neq \omega_0^{(2)}$, $\varphi_0^{(2)} - \varphi_0^{(1)} = \alpha$ we obtain the different Lissajous figures that can be observed in the case of clusterization of different data points forming the initial vectors $T^{(1,2)}$. As a prove in Fig.3 we show the dependence $T^{(2)} = F(T^{(1)})$ for the real data described above that forms an ellipse. The envelope of the ellipse is slightly destroyed by the influence of other small frequencies presenting in the decomposition of the both Fourier series forming the vectors $T^{(1,2)}$. This simple observation prompts to use for *quantitative* characterization and parameterization the fragments of the Fourier series belonging to the vectors of the score matrix.

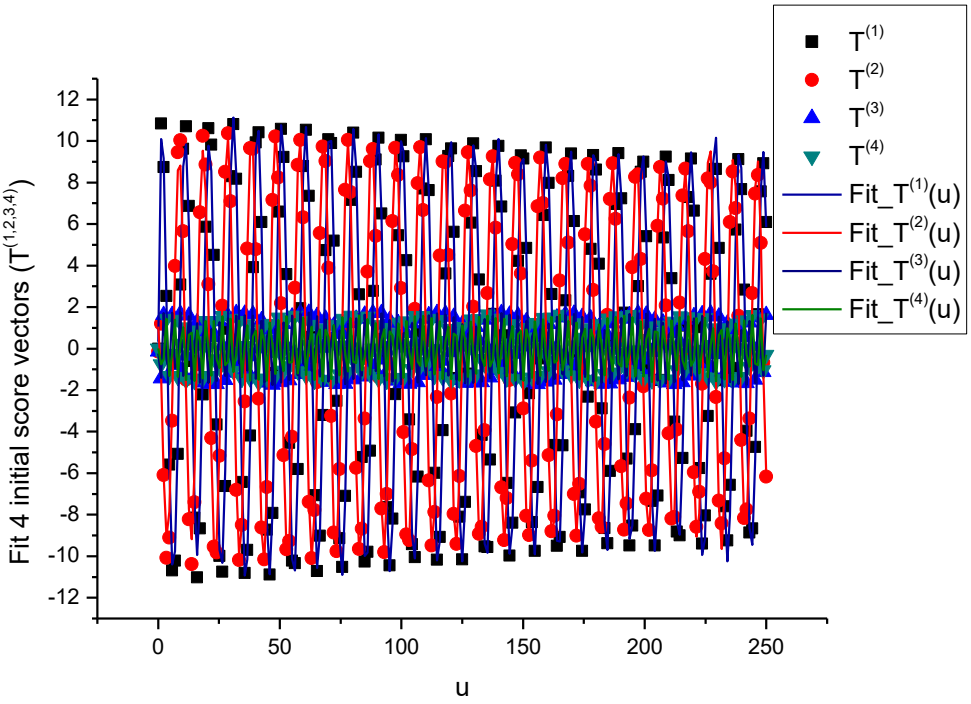

Figure 4. Here we show the fit of the first 4 vectors $T^{(1,2,3,4)}$ of the score matrix. The fitting error defined by expression (13) for these 4 vectors is located in the interval (1% < RelErr <8%).

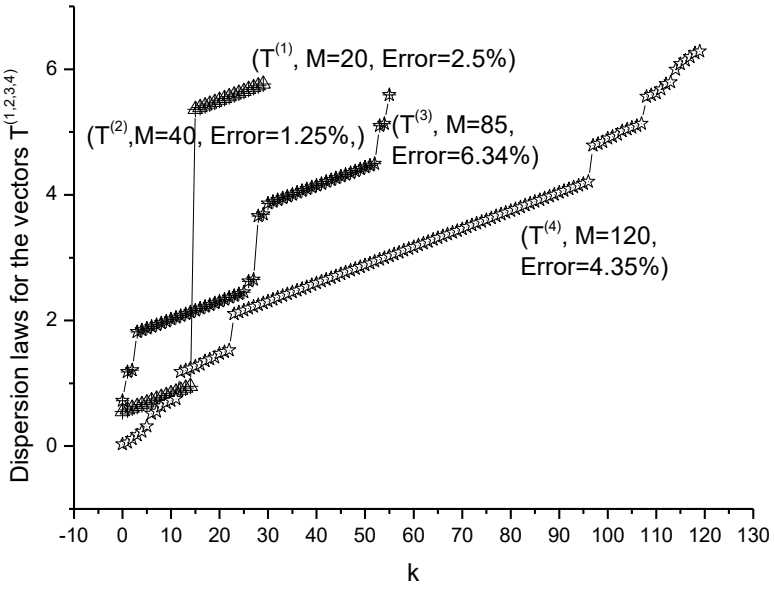

Figure 5. The dispersion laws for the selected 4 score vectors $T^{(1,2,3,4)}$. In order to keep the fitting error in the interval (1% < RelErr < 8%) we should adjust the parameter *M* figuring in expression (12). It is shown for each dispersion curve in the corresponding parenthesis together with the value of the fitting error.

$$T^{(s)}(u) = A_0^{(s)} + \sum_{k=1}^{K}\left(Ac_k^{(s)}\cos\left(\omega_k^{(s)}u\right) + As_k^{(s)}\sin\left(\omega_k^{(s)}u\right)\right) \equiv$$

$$\equiv A_0^{(s)} + \sum_{k=1}^{K} Am_k^{(s)}\cos\left(\omega_k^{(s)}u - \varphi_k^{(s)}\right), \qquad (8)$$

$$Am_k^{(s)} = \sqrt{\left(Ac_k^{(s)}\right)^2 + \left(As_k^{(s)}\right)^2},\ \tan\left(\varphi_k^{(s)}\right) = \frac{As_k^{(s)}}{Ac_k^{(s)}},\ s = 1, 2, ..., N_r.$$

We should stress here the following fact. If $\omega_{k+1}^{(s)} - \omega_k^{(s)} = a^{(s)}, a^{(s)} = \frac{2\pi}{\tau^{(s)}}$ and $\tau^{(s)}$ coincides with the period of vibration of the vector $T^{(s)}$ then the fragment (8) coincides completely with the Fourier decomposition. Otherwise, when $\omega_{k+1}^{(s)} - \omega_k^{(s)} \neq a^{(s)}$ then this decomposition recovers the NAFASS approach [20, 21], where the set of the frequencies follows to another dispersion law. These dispersion laws for the first 4 vectors are shown in Fig.5. Concluding this section one can say that the behavior $T^{(s)}(u)$ can be characterized naturally by its AFR (amplitude-frequency response) $(\omega_k^{(s)}, Am_k^{(s)})$ and, therefore, the dependencies $T^{(s+1)}(u) = F(T^{(s)}(u))$ looking initially as random dependencies receive their natural explanation. Actually, we found the *parameterization* of the score vectors using for this purpose the dimensionless potential u. Now as a criterion of neglecting by a possible dependence $T^{(s)}(u)$ for relatively large s one can use the range of this vector. It follows from the definition of $T^{(s)}(u)$ ($US \equiv T$) when this vector is multiplied by the eigenvalue σ_s decreasing its range with the growth of the order of s. The *main* problem of this behavior lies in finding of the adequate *chemical interpretation* of the AFR, especially for the spectrum $\{\omega_k^{(s)}\}$ belonging to each component of the score matrix $T^{(s)}$. Nevertheless, the first problem formulated above is solved. We found the adequate *parameterization* (8) for a wide set of the vectors that describe the experimental data expressed in the form of the normalized matrix (5). Coming back to distribution of the eigenvalues of the matrix $S(\sigma_1 \geq \sigma_2 \geq ...\sigma_{N_r})$ located in the descending order one can realize the following procedure. Firstly, it is necessary to calculate the mean value and deviations $\Delta\sigma_i$ of this set

$$\langle\sigma\rangle = \frac{1}{N_r}\sum_{i=1}^{N_r}\sigma_i,\ \Delta\sigma_i = \sigma_i - \langle\sigma\rangle. \qquad (9)$$

Then integrating these deviations $\Delta\sigma_i$ with respect to the variable u_i we obtain the bell-like curve that can be fitted with high accuracy (2% < RelError < 10%) by means of expression

$$Bd(u) = A\left(u - \min(u)\right)^{\alpha}\cdot\left(\max(u) - u\right)^{\beta} + B,$$

$$\mathrm{Re}\,lErr = \left(\frac{Stdev(J(u) - Bd(u))}{\left|mean(J(u)\right|}\right)\cdot 100\%. \qquad (10)$$

If the eigenvalues are decreased essentially then one can take the natural logarithm from the initial distribution and then apply expression (10) again for the fitting of the desired bell-like curve. As one can see from expression (10) only 4 fitting parameters (A, B, α, β) can characterize quantitatively the distribution of the eigenvalues that it makes convenient for quantitative analysis of different data, as well. Therefore, the conventional PCA gets the quantitative part that allows comparing one

normalized matrix (5) with another one if this new matrix of measurements is subjected by some *qualitative* modifications. These modifications or influence of the random factors can be detected and expressed in terms of AFR $(\omega_k^{(s)}, Am_k^{(s)})$ belonging to the set of the fitting vectors $T^{(s)}(u)$. The number of s is determined by the range of the given vector $T^{(s)}(u)$. If the range of the vector $T^{(s)}(u)$ is small then the contribution of this vector will be ignored.

3. Experimental details

All voltammetric measurements were realized with the usage of three-electrode cell and potentiostat "Elins-P30S" (Chernogolovka, Russia, www.elins.su). The glassy carbon electrode (GCE) with the area of the working surface S=7 mm^2 was used as the working electrode. The glassy carbon bar (S=282 mm^2) and chloride-silver electrode (Ag / AgCl (3.5 M KCl)) were used as auxiliary and comparison electrodes, accordingly. As an background electrolyte the standard buffer solution with pH=6,86 (the mixture Na_2HPO_4 and KH_2PO_4) was served. For preparation of the desired solutions, the bidistilled water was used.

The massive of data were formed by continuous registrations of VAGs in the linear conditions with the potential scanning rate 500 mV/sec in the range of -500 – 1500 mV. We used two regimes of the VAGs registration:

Regime-1 – the regime with the preliminary fifth-fold electrochemical regeneration of the GCE and before every registration of the VAG the potential with high scanning rate 2300 mV/sec in the range –1500, +2500 mV was applied. The scheme of the potentials application was shown in Fig.1.

Regime-2 – this regime does not use the scheme of the potential regeneration. In this regime, only the registration of the desired VAGs takes place.

Each matrix of data includes S=100-200 registered VAGs. The obtained VAGs before PCA were differentiated and the first derivatives $dJ^{(s)}/dU$ (s=1,2,…,S) (this operation eliminates the undesirable envelopes from the curves $J^{(s)}(U)$) were analyzed. All experiments were performed in the constant stirred solution at room temperature.

4. The treatment of experimental data

General notice

For application of the fitting procedure described in section 2 we notice that after application of the conventional program CFFT (vector) (abbreviation CFFT means "Complex Fast Fourier Transform" and applicable for a vector having an arbitrary length) we obtain the set of frequencies located in the range [0, 2π], because the frequencies exceeding this range and containing the part 2πn can be omitted in accordance with relationship (11)

$$\omega_{i,k}^{(s)} = \left(\omega_k^{(s)} + 2\pi n\right) \cdot \frac{i}{N_r}, \; i = 0,1,...,N_r, \; \frac{i \cdot n}{N_r} \geq 1. \tag{11}$$

Taking into account the limit

$$\frac{\max\left(\omega_k^{(s)}\right)}{M} \leq \omega_k^{(s)} \leq \max\left(\omega_k^{(s)}\right), \tag{12}$$

where the factor M is chosen from the condition that the value of the relative error

$$\mathrm{Re}\,lError(M) = \left(\frac{Stdev(T_i^{(s)} - T^{(s)}(u_i, \omega_k^{(s)}(M))}{mean\left|T_i^{(s)}\right|} \right) \cdot 100\%, \qquad (13)$$

between $T_i^{(s)}$ and its fitting expression (8) should not exceed the value of (1-10)%. In realization of the fitting procedure we try to keep the value of the final mode $K(M)$ in (8) satisfying to the condition $K << N_r$. This procedure sounds reasonable because it helps to eliminate the set of "useless" frequencies with small amplitudes that are located out of the interval (12). In order not to overload the text of the paper by large number of figures we choose only 4 components of the vector $T^{(s)}$ for each regime and try to compare them with each other. In this paper, we deliberately choose the region of the *remnant* currents (that previously was never considered seriously) for two regimes of the continuous voltammetric measurements explained above:

Regime 1 ((pre)cycling regime with high scanning potential rate-simply defined below as simply cycling)

Regime 2 (it does not include preliminary electrochemical handling of the GCE-defined below and considered below as linear)

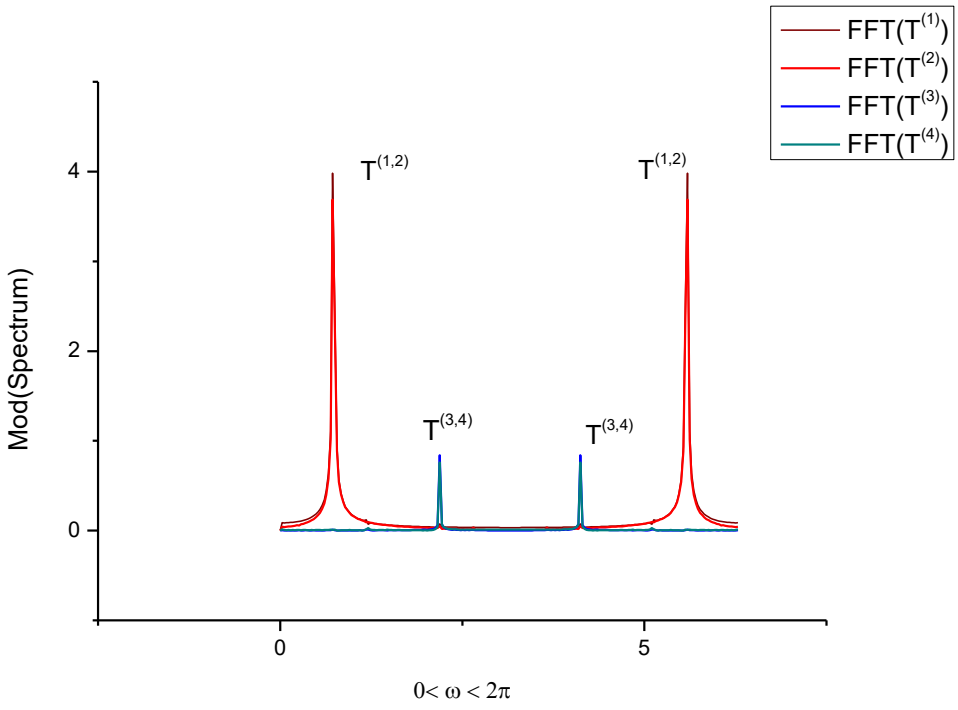

Figure 6. These plots of the CFFT applied to the first 4 vectors demonstrate their similarity of the first two vectors $T^{(1,2)}$ and $T^{(3.4)}$, accordingly. They have the similar leading frequencies but other frequencies having small amplitudes are slightly different. They can be important in the accurate fitting of these vectors to the function (8).

In the both regimes, the registered VAGs were obtained in the conditions of the linear potential scanning. Data obtained for other regions (cathode, faraday, faraday with water oxidations) were *not* considered because their analysis requires large number of figures. Their consideration merits a separate research.

4.1 The (pre)cycling (fast) regime (the region of the remnant currents)

Part of the figures (Figs.1-6) has been explained above. In Figs 7-8 we exhibit the AFR for the vectors $T^{(1,2,3,4)}$ playing the key role in analysis of each regime. Because of the character of vibrations of the initial data and vectors $T^{(1,2)}$ are not known properly (see Fig.2 and Fig.3) then one can concentrate on the analysis of the AFR for these vectors. It is interesting to notice that accurate fit of these initial vibrations leads to considerable increasing of the amplitudes (see Figs. 7-8). This presentation plays a role of a specific "magnifying glass" that allows to notice some hidden details. This phenomenon probably arises from the fact that real fit with small relative error (2% < RelError < 10%) is secured by the differences of the amplitudes. These plots can fix all qualitative differences that are realized in this first regime.

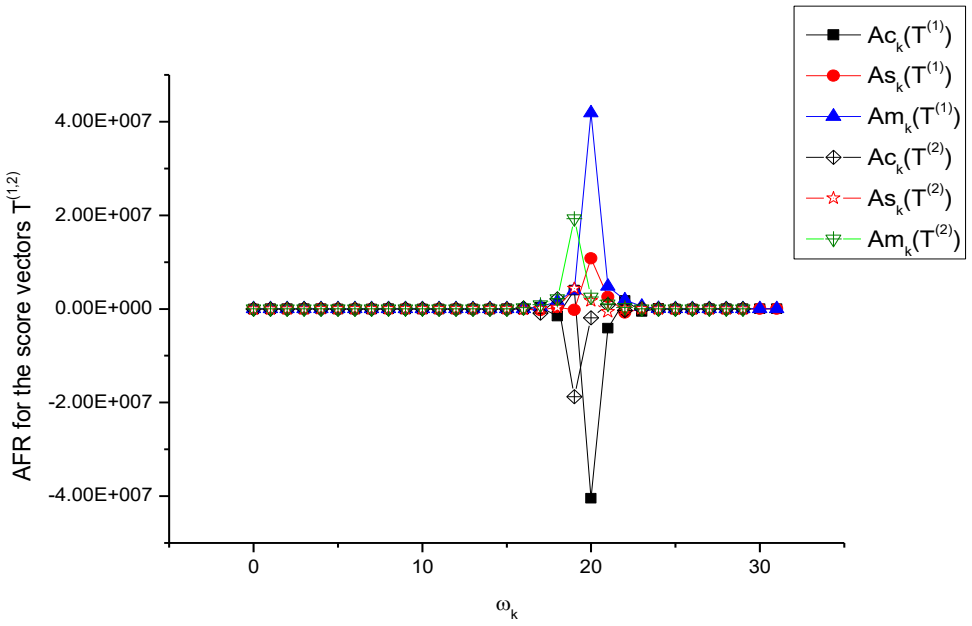

Figure 7. This plot shows the AFR for the first two vectors T(1,2). All large amplitudes are located near the value ωk = 20.

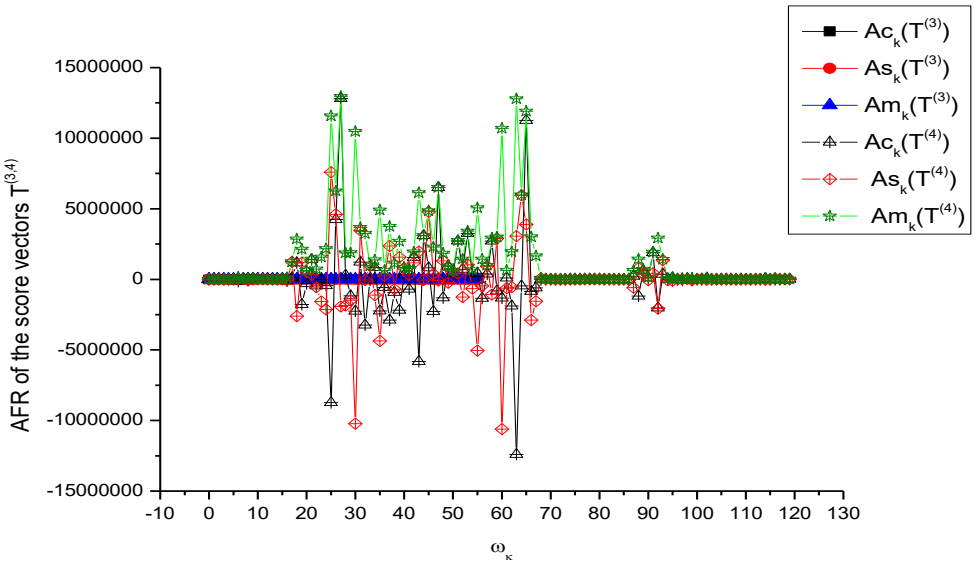

Figure 8. The AFR for the vectors T(3,4). They have more rich spectra and occupy more wide frequency range in comparison with the AFR for the vectors T(1,2) depicted in the previous figure.

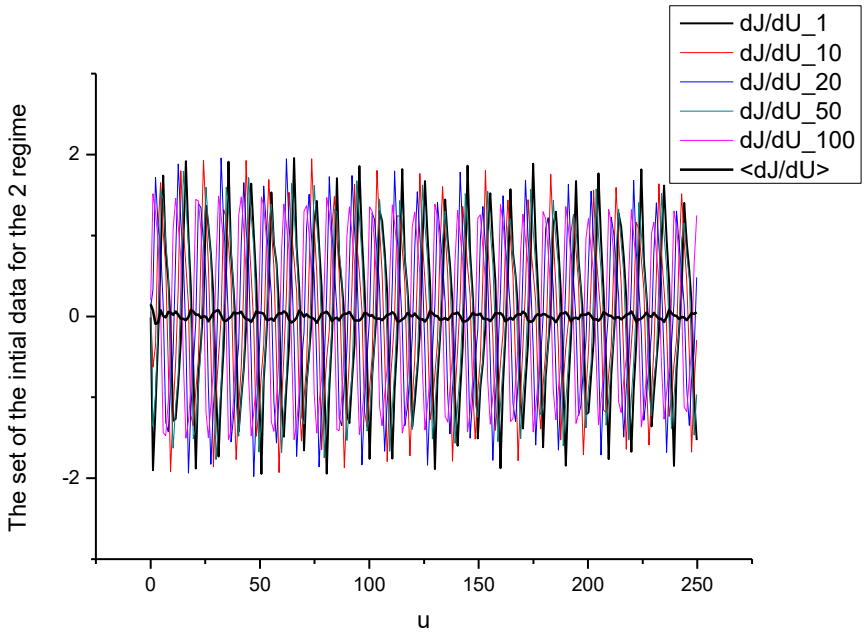

Figure 9. Here we show the initial data for selected measurements (*s* = 1,10,20,50,100) together with the mean measurement <*dJ*/*dU*> averaged over all experiments for the 2-nd linear regime (the region of the remnant currents). If we compare these initial data with Fig. 2 depicted for the cycling regime then the significant changes are not noticeable.

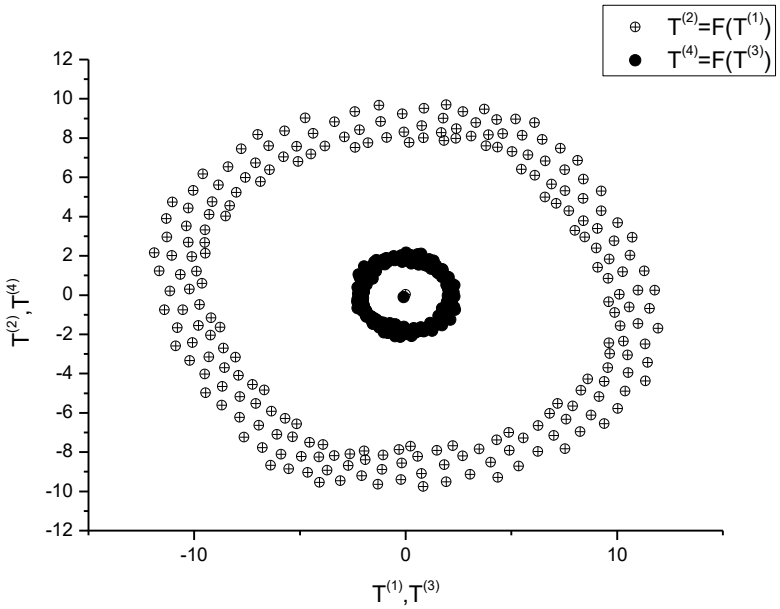

Figure 10. After realization of the SVD procedure one can plot $T^{(2,4)} = F(T^{(1,3)})$ corresponding to the conventional analysis. In order to explain these plots we apply the parameterization of these functions in accordance with expression (8). The fit of these functions is shown in the next figure. If one compares these plots with depicted ones in Fig.3. one can notice some small discrepancies. But they can be compared only with the help of expression (8) that differs these two regimes.

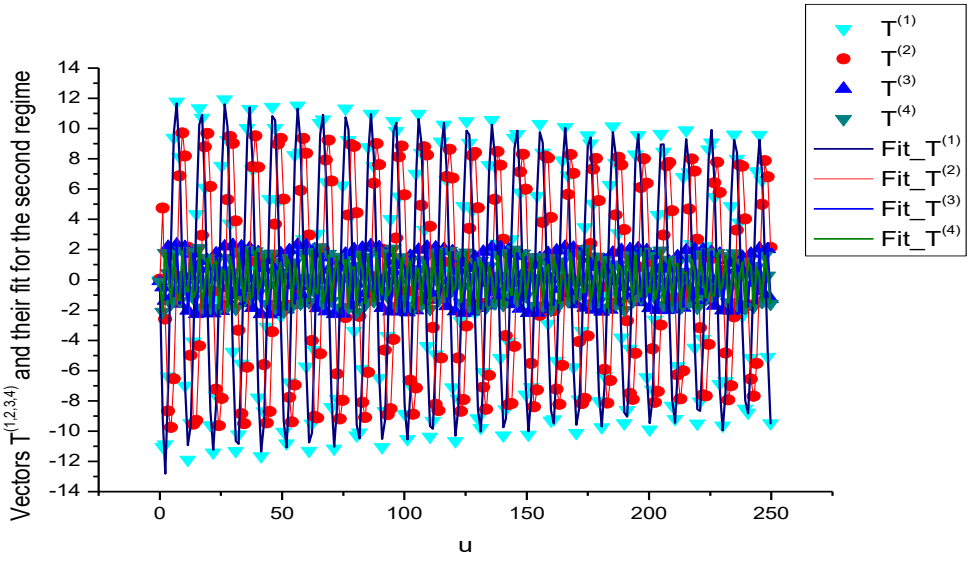

Figure 11. Here we show the fit of the first 4 vectors $T^{(1,2,3,4)}$ of the score matrix for the linear (2-nd) regime. The fitting error defined by expression (13) for these 4 vectors is located in the interval (1% < RelErr <8%). How to compare the vectors shown in Fig.4 for the cycling regime with the vectors corresponding to the linear regime? The answer is given in Figs.12(a,b) below.

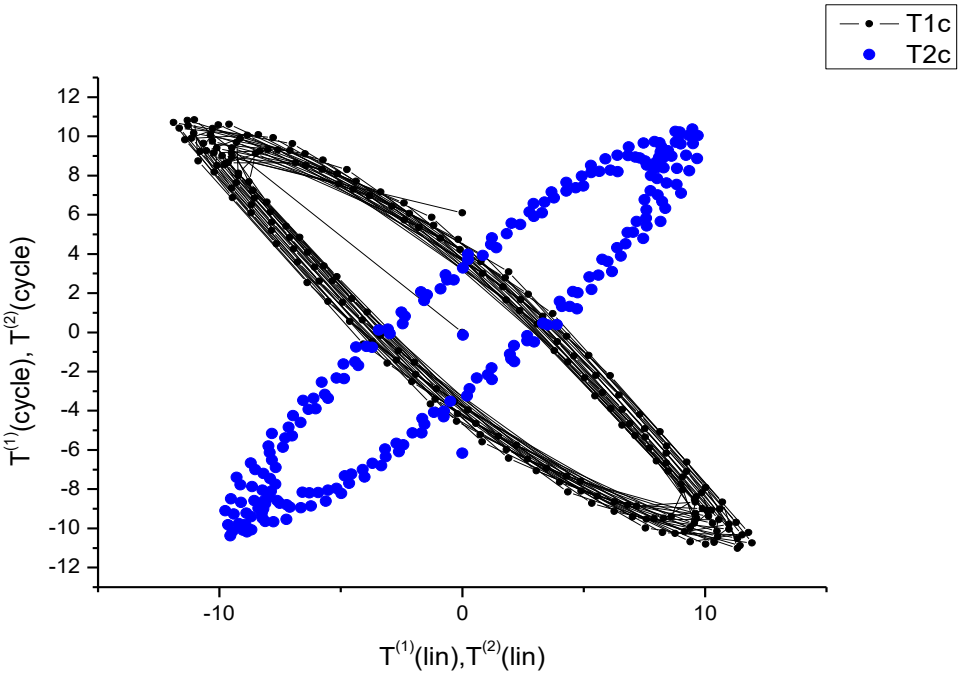

Figure 12(a). Being plotted with respect to each other the vectors $T^{(1,2)}$ corresponding two different regimes exhibit their clear difference. Actually they present approximately two ellipses rotated relatively each other on the angle 90^0.

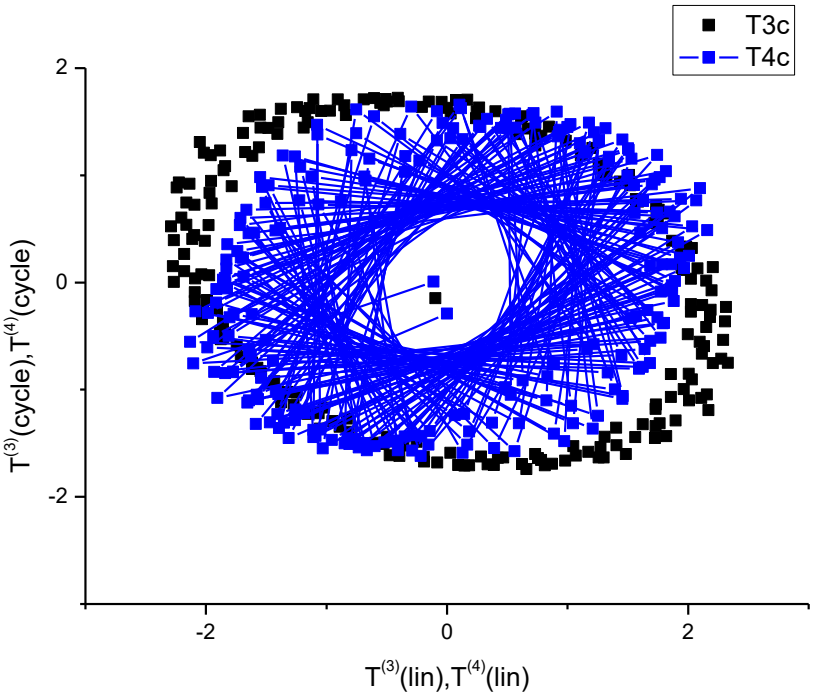

Figure 12(b). The same presentation for the vectors $T^{(3,4)}$ corresponding for two regimes exhibits a similar behavior. We see again approximately two deformed ellipses but the angle of rotation does not equal 90^0.

4.2 The linear (2-nd) regime (the region of the remnant currents)

In the same manner, one can consider another regime associated only with slow and linear changing of the potential *u*. The analysis of this regime is presented by the set of the Figs. 9-16. In Fig.9, we select the initial data corresponding to this case. The visual comparison of this figure with Fig.2. does not contain some significant information. One can notice some visual changes comparing Fig. 10 with Fig.3. However, these changes become more distinct from analysis of the Figs. 12 (a, b). One can read all visual changes and expresses them quantitatively with the usage of expression (8). The desired fit is given in Fig.11. One can notice the differences between Fourier spectra that are shown in Fig. 14. The dispersion laws are different and asymmetrical that it follows from comparison of Fig. 4 with Fig. 13. The significant discrepancies are clearly seen from Figs. 7-8 if we compare them with Figs. 15-16. The AFRs contains some important information that can be related with these two regimes. Now we cannot decode this new language properly based on some microscopic model but we show at least a way for researches working in the region of electrochemistry for more accurate analysis of the measured data that earlier were never analyzed in specific coordinates as "potential – time".

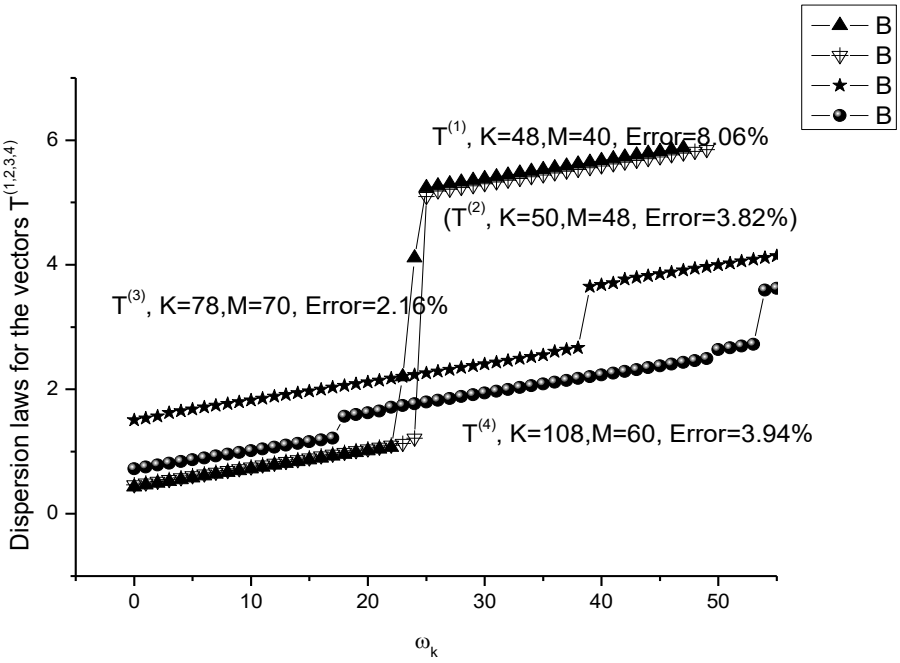

Figure 13. Here we show the dispersion laws for the selected 4 score vectors $T^{(1,2,3,4)}$ corresponding to the linear (2-nd) regime. In order to keep the fitting error in the interval (1 < RelErr < 8%) we should adjust the parameter *M* figuring in expression (12). It is shown for each dispersion curve in parenthesis together with the value of the final mode *K* and fitting error.

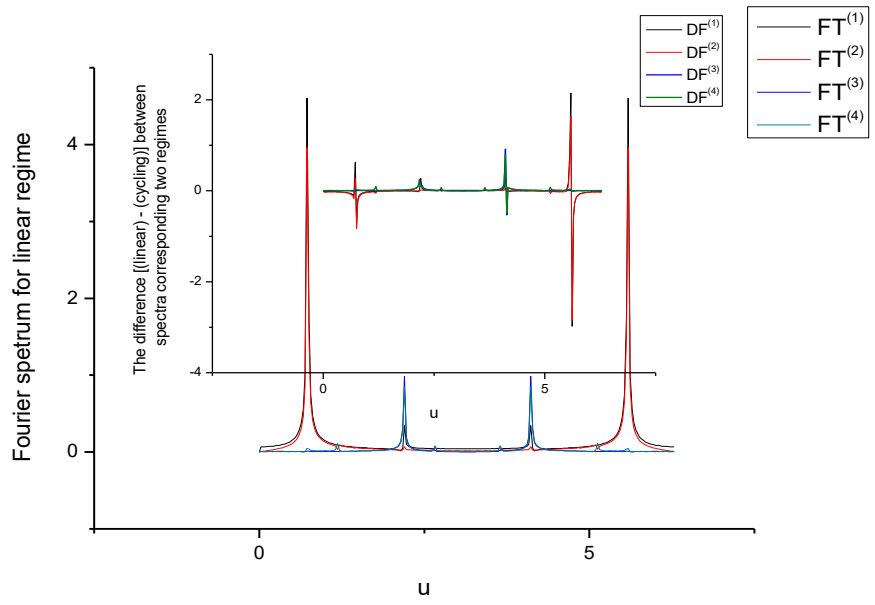

Figure 14. Here we demonstrate the Fourier spectra for 4 vectors corresponding to the linear(2-nd) regime. The difference between spectra (linear –cycling) regimes is shown in the small figure above. Therefore, these vectors reflect all possible changing associated with the given regime.

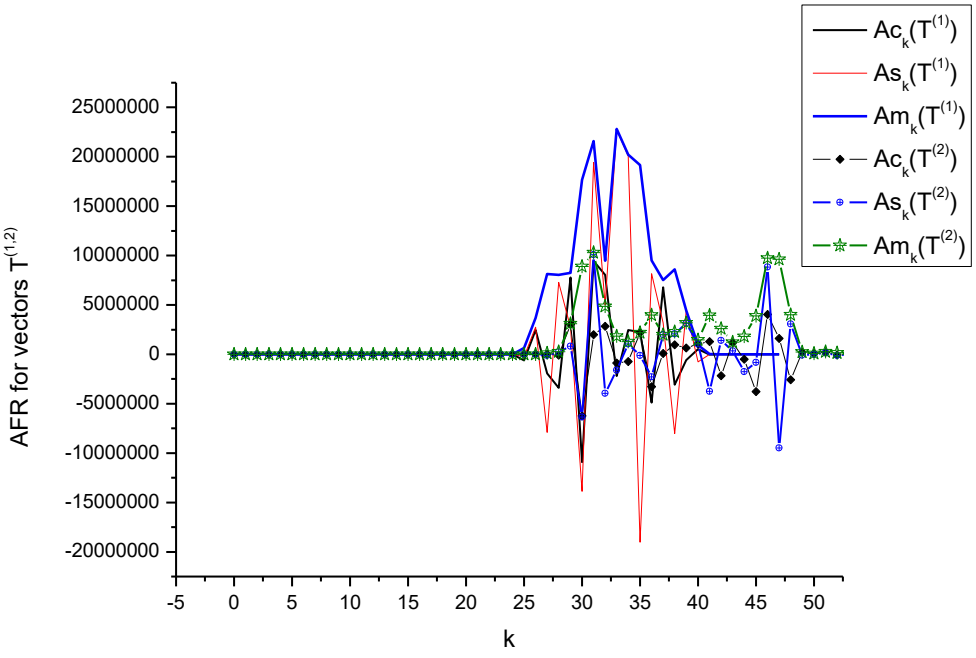

Figure 15. The AFR for the vectors $T^{(1,2)}$. As we see the fitting organized with the help of the LLSM gives more accurate description of the AFR based on expression (8) and dispersion laws for each regime depicted on Figs. 6 and 12 that can be asymmetric in general case.

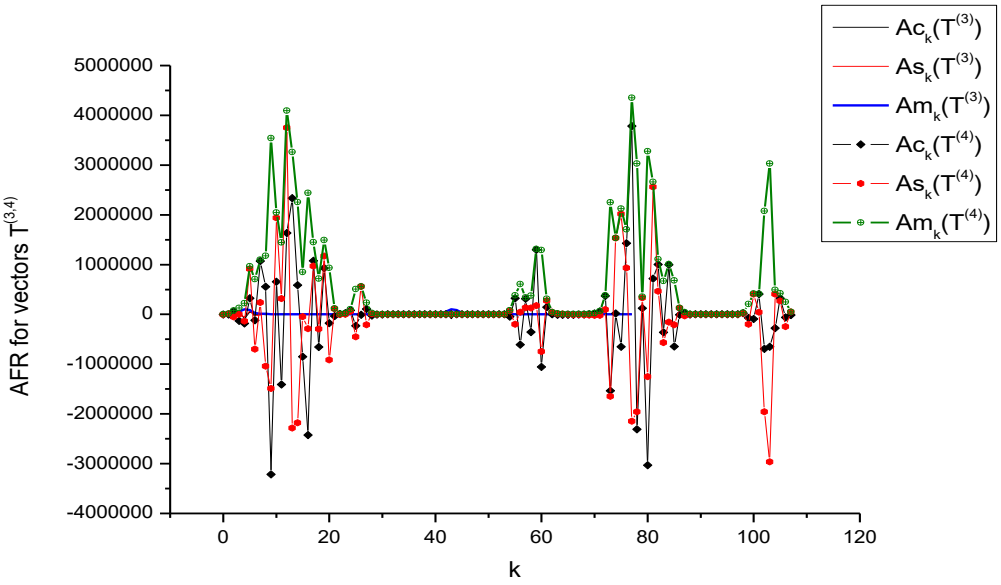

Figure 16. Here we show the AFR for the score vectors $T^{(3,4)}$. If we compare these AFRs with Fig. 7 one can notice the big difference between them. It implies that these AFR are the most sensitive for analysis of possible qualitative changings that are appeared in different regimes.

4.3 Analysis of the bell-like curves

Now one can use sources of quantitative information that are given by expressions (9) and (10). We show only how to realize the relative comparison of the bell-like curves that correspond to selected regimes. This relative comparison is explained by the Figs 17-18.

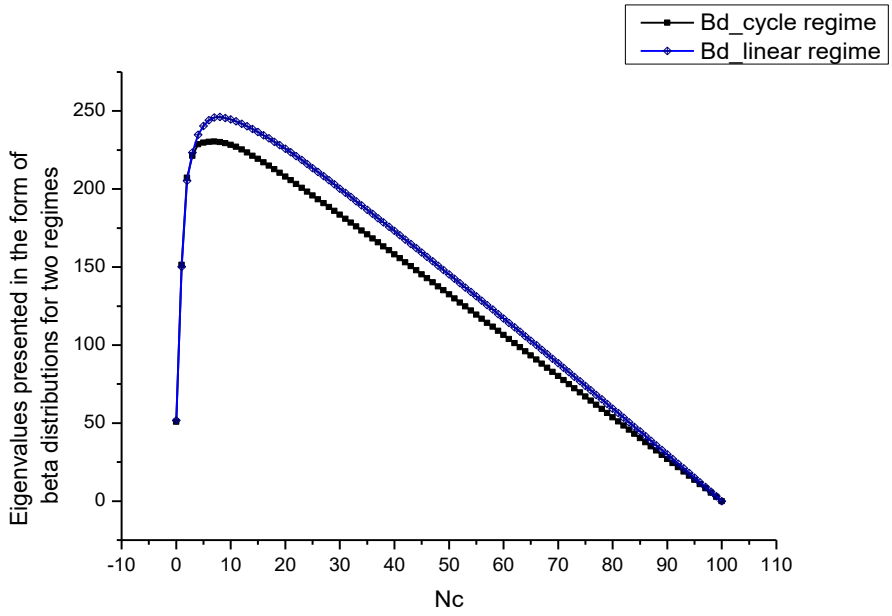

Figure 17. Comparison of the eigenvalues presented in the form of bell-like curves for two regimes. One can notice the differences between two regimes. In order to express them quantitatively one can use expression (10) or relative comparison if we present them in the form of the segments of the straight lines.

5. Results and discussion

In this paper we show how to solve two problems mentioned above. These new elements can be added to the conventional PCA that makes it more informative and flexible in analysis of many similar data. One can pose the following question: where the key element of the PCA generalization is located to make it more efficient? From our point of view, this element is contained in expressions (6)-(8). These expressions, in turn, are followed from the peculiarities of the Fourier spectrum depicted in Fig 6 and Fig.14. The Fourier spectra should contain the leading frequencies with large amplitudes and another set of frequencies with small amplitudes. Another criterion (as a conclusion) lies in Figs (3, 10). If we observe the slightly distorted ellipses rotated relatively to each other on the angle α then one concludes that we have the Fourier spectrum with the leading frequencies. At these conditions one can apply the parameterization of the components of the score matrices $T^{(s)}$ given by expression (8). In addition, we suggest new source of quantitative information that are given by expressions (9) and (10).

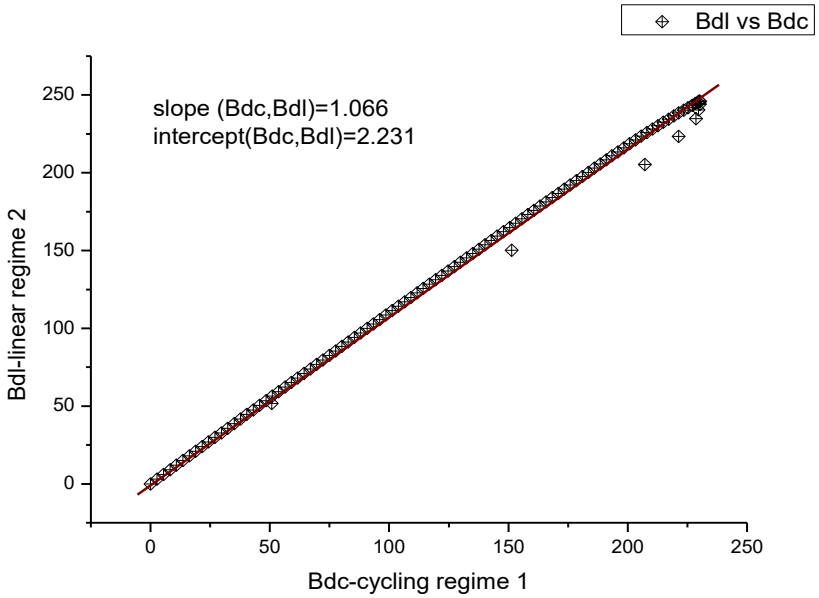

Figure 18. The relative comparison of the eigenvalues corresponding to regimes 1 and 2 expressed in the form of the bell-like curves.

But one important problem remains open, especially for researches, working in electrochemistry. What are the microscopic reasons that evoke this oscillating behavior for the dependence $J(U)$? In order to see this behavior properly we use the first derivative dJ/dU then test the dependencies $T^{(s+1)}=L(T^{(s)})$ (s=1,2,…,S) and extend all measurements for the long period of observation ($S \geq 100$). For more clear observation of this effect, we used the "quantized" time measured in number of independent VAGs. It is easy to transform these successive measurements in time multiplying the parameter $s\cdot\Delta\tau$ (where s = 1,2,…,S), $\Delta\tau$=5sec (for the experiments considered above). This phenomenon allows fitting the components of the score matrix and considering all regions (cathode, remnant, faraday etc.) different currents including also different types of electrodes and solutions. These new possibilities merit further research. We deliberately analyzed in this paper the region of the remnant currents that was considered before as spurious and useless.

References

[1]. M. Gutiérrez-Capitán, J.L. Santiago, J. Vila-Planas, A. Llobera, S. Boso, P. Gago, M.C. Mart ínez, C. Jiménez-Jorquera, (2013), *J. Agric. Food. Chem.* 61, 9325–9332.

[2]. Z. Wei, J. Wang, (2014), *Comput.Electron. Agric.* 108, 112–122.

[3]. Z. Wei, J. Wang, X. Zhang, (2013), *Electrochim. Acta.* 88, 231–239.

[4]. A.V. Sidel'nikov, V.N. Maistrenko, R.A. Zil'berg, Yu. A. Yarkaeva, E.M. Khamitov, (2017), *J. Anal. Chem*. 72(5), 575-581.

[5]. R.A. Zil'berg, Yu. A. Yarkaeva, E.I. Maksyutova, A.V. Sideln'ikov, V.N. Maistrenko, (2017), *J. Anal. Chem*. 72(4), 402–409.

[6]. I. Campos, M. Alca ñiz, D. Aguado, R. Barat, J. Ferrer, L. Gil, M. Marrakchi, R. Mart ínez-M áñez, J. Soto, J.L. Vivancos, (2012), *J. Water Research*, 46, 2605-2614.

[7]. I. Campos, A. Sangrador, R. Bataller, D. Aguado, R. Barat, J. Soto, R. Mart ínez-M áñez, (2014), *Electroanalysis*, 26, 588-595.

[8]. F. Scholz (Ed.), (2002), *Electroanalytical Methods, Guide to Experiments and Applications*, Springer-Verlag Berlin Heidelberg.

[9]. S. Holmin, C. Krantz-Rulcker, I. Lundstrom, F. Winquist, (2001), *Meas. Sci. Technol*, 12(8), 1348–1354.

[10]. R.G. Compton, C.E. Banks, (2011), *Understanding voltammetry*, Second edition, Imperial College Press.

[11]. G. Inzelt "Kinetics of Electrochemical Reactions", in: F. Scholz (Ed.), (2002), *Electroanalytical Methods, Guide to Experiments and Applications*, Springer-Verlag Berlin Heidelberg, 29–48.

[12]. A.V. Sidelnikov, D.M. Bikmeev, F.Kh. Kudasheva, V.N. Maystrenko, (2013), *J. Anal. Chem*. 68(2), 140–147.

[13]. A.V. Sidel'nikov, D. M. Bikmeev, D.I. Dubrovskii, F.Kh. Kudasheva, V.N. Maistrenko, (2015), *J. Anal. Chem*. 70(7), 837–842.

[14]. A.V. Sidel'nikov, D.I. Dubrovskii, F.Kh. Kudasheva, V.N. Maistrenko, (2016), *J. Anal. Chem*. 71(11), 1109–1114.

[15]. J. Wang, J.R. Fernandas, L.T. Kubota, (1998), *Anal. Chem.* 70(17), 3699–3702.

[16]. J. Wang, A.-N. Kawde, (2001), *Anal. Chim. Acta.* 431(2), 219–224.

[17]. N. Diab, A. Abuzuhri, W. Schuhmann, (2003), *Bioelectrochemistry*, 61 (1-2), 57–63.

[18]. R.E. Moss, R.E. Perez-Roa, M.A. Anderson, (2013), *Electrochim. Acta.* 104, 314–321.

[19]. Battistel, S. Daniele, (2013), *J. Solid State Electrochem*. 17(6), 1509–1516.

[20]. R.R. Nigmatullin, W. Zhang, (2013), *Communications in Nonlinear Science and Numerical Simulation*. 18, 547-558.

[21]. R.R. Nigmatullin, S.I.Osokin and V.A. Toboev. (2011), *Chaos, Solitons and Fractals*. 44, 226-240.